CHAPTER 3

Amplitude (Linear) Modulation

Inside this Chapter

3.1. Introduction

As discussed earlier in Chapter 1, Modulation is a fundamental requirement of a communication system. Modulation may be defined as a process by which some characteristic of a signal known as carrier is varied according to the instantaneous value of another signal known as modulating signal. The signals containing intelligence or information to be transmitted are called modulating signals. These modulating signals containing information are also called baseband signals. Also the carrier frequency is greater than the modulating frequencies and the signal which results from the process of modulation is known as modulated signal.

Modulation may be classified as continuous wave modulation and pulse modulation. If the carrier waveform is continuous in nature then the modulation process is called as continuous-wave (CW) modulation. The examples of this type of modulation are amplitude modulation and angle modulation. On the other hand, if the carrier waveform is a pulse-type waveform, then the modulation process is called as pulse modulation. The examples of this type of modulation are Pulse-Amplitude Modulation (PAM), Pulse-Width Modulation (PWM), Pulse Code Modulation (PCM) etc.

Amplitude Modulation and Angle Modulation are the two families of continuous-wave (CW) modulation systems. In amplitude modulation, the amplitude of a sinusoidal carrier wave is varied in accordance with the baseband (modulating) signal. On the other hand, in angle modulation, the angle of the sinusoidal carrier wave is varied in accordance with baseband signal.

Amplitude modulation is discussed in this chapter, whereas the angle modulation will be discussed in next chapter.

3.2. Multiplexing

Multiplexing is a technique in which several message signals are combined into a composite signal for transmission over a common channel. In order to transmit a number of these signals over the same channel, the signals must be kept apart so that they do not interfere with each other, and hence they can be separated easily at the receiver end.

Basically, multiplexing is of two types as under:

(*i*) frequency division multiplexing (FDM)
(*ii*) time-division multiplexing (TDM)

(*i*) Frequency Division Multiplexing (FDM) *(Important)*

The FDM scheme is illustrated in figure 3.1 with the simultaneous transmission of three message or baseband signals. The spectra of the message signals and the sum of the modulated carriers are indicated in the figure. DSB modulation is used in illustrating the spectra of figure 3.1. Any type of modulation can be used in FDM as long as the carrier spacing is sufficient to avoid spectral overlapping. However, the most widely used method of modulation is SSB modulation. At the receiving end of the channel the three modulated signals are separated by bandpass filters (BPFs) and then demodulated.

FDM is used in telephone system, telemetry, commercial broadcast, television, and communication networks. Commercial AM broadcast stations use carrier frequency spaced 10 kHz apart in the frequency range from 540 to 1640 kHz. This separation is not sufficient to avoid spectral overlap for AM with a reasonably high fidelity (50 Hz to 15 kHz) audio signal. Therefore, AM stations on adjacent carrier frequencies are placed geographically far apart to minimize interference. Commercial FM (Frequency-Modulation) broadcast uses carrier frequencies spaced 200 kHz apart. In a long-distance telephone system, up to 600 or more voice signals (200 Hz to 3.2 kHz) are transmitted over a coaxial cable or microwave links by using SSB modulation with carrier frequencies spaced 4 kHz apart. In practice, the composite signal formed by spacing several signals in frequency may, in turn, be modulated by using another carrier frequency. In this case, the first carrier frequencies are often called ***sub-carriers***.

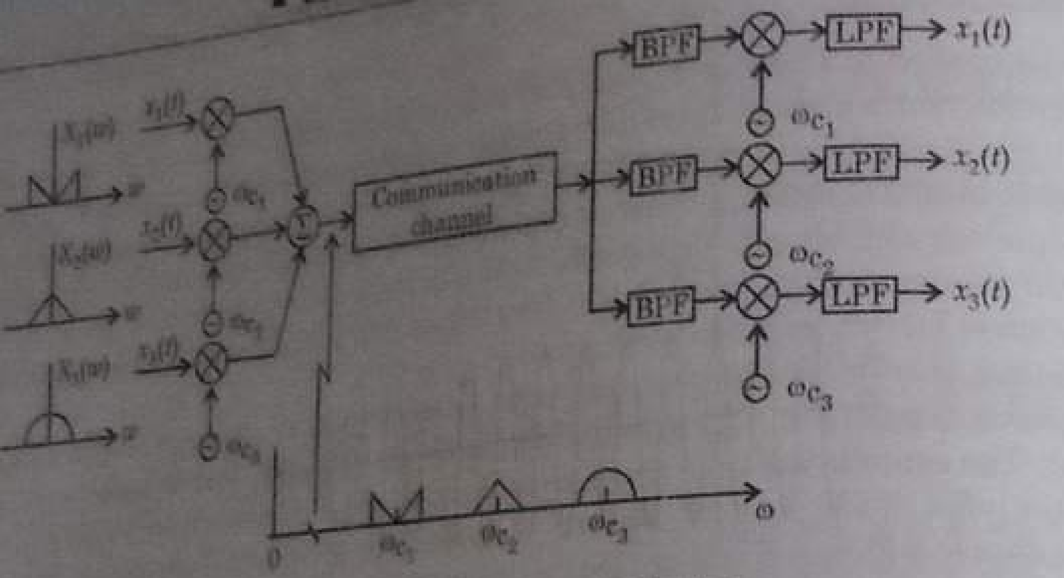


Fig. 3.1. *Frequency-division multiplexing.*

(ii) Time Division Multiplexing (TDM)

In case of Time Division Multiplexing (TDM), the complete channel bandwidth is allotted to one user for a fixed time slot. As an example, if there are ten users, then every user can be given the time slot of one second. Thus, complete channel can be used by each user for one second time in every ten seconds. This techniques is suitable for digital signals. Because digital signals are transmitted intermittently and the time spacing between two successive digital codewords can be utilized by other signals. There is possibility of crosstalk in FDM whereas Intersymbol Interference is possible in TDM. These problems can be overcome by some special cares.

3.3. Amplitude Modulation

Amplitude Modulation may be defined as a system in which the maximum amplitude of the carrier wave is made proportional to the instantaneous value (amplitude) of the modulating or baseband signal.

Let us consider a sinusoidal carrier wave $c(t)$ given as

$$c(t) = A \cos \omega_c t \quad ...(3.1)$$

Here A is the maximum amplitude of the carrier wave and ω_c is the carrier frequency. For simplicity here we have assumed that the phase of the carrier wave is zero in equation (3.1).

Let $x(t)$ denotes the modulating or baseband signal, then according to amplitude modulation, the maximum amplitude A of the carrier will have to be made proportional to the instantaneous amplitude of modulating signal $x(t)$.

The standard equation for amplitude modulated (AM) wave may be expressed as

$$s(t) = x(t) \cos \omega_c t + A \cos \omega_c t \quad ...(3.2)$$

or

$$s(t) = [A + x(t)] \cos \omega_c t \quad ...(3.3)$$

Figure 3.2 shows the modulating signal or baseband signal, carrier signal and modulated signal.

Few Points about Amplitude Modulation (AM):

(i) It may be observed that equation (3.2) or equation (3.3) describes the time-domain behaviour of amplitude-modulated signal.

(ii) It may be noted that carrier signal [*i.e.* $c(t) = A \cos \omega_c t$] is a fixed frequency signal having frequency ω_c. The modulating or baseband signal $x(t)$ contains the information or intelligence to be transmitted. In the process of amplitude modulation, this information is superimposed upon the carrier signal in the form of amplitude variations of the carrier signal. This means that the information to be transmitted, is now, contained in the amplitude variations of the carrier signal. In other words, in amplitude modulation, the information is transmitted in the form of amplitude variations of the carrier signal.

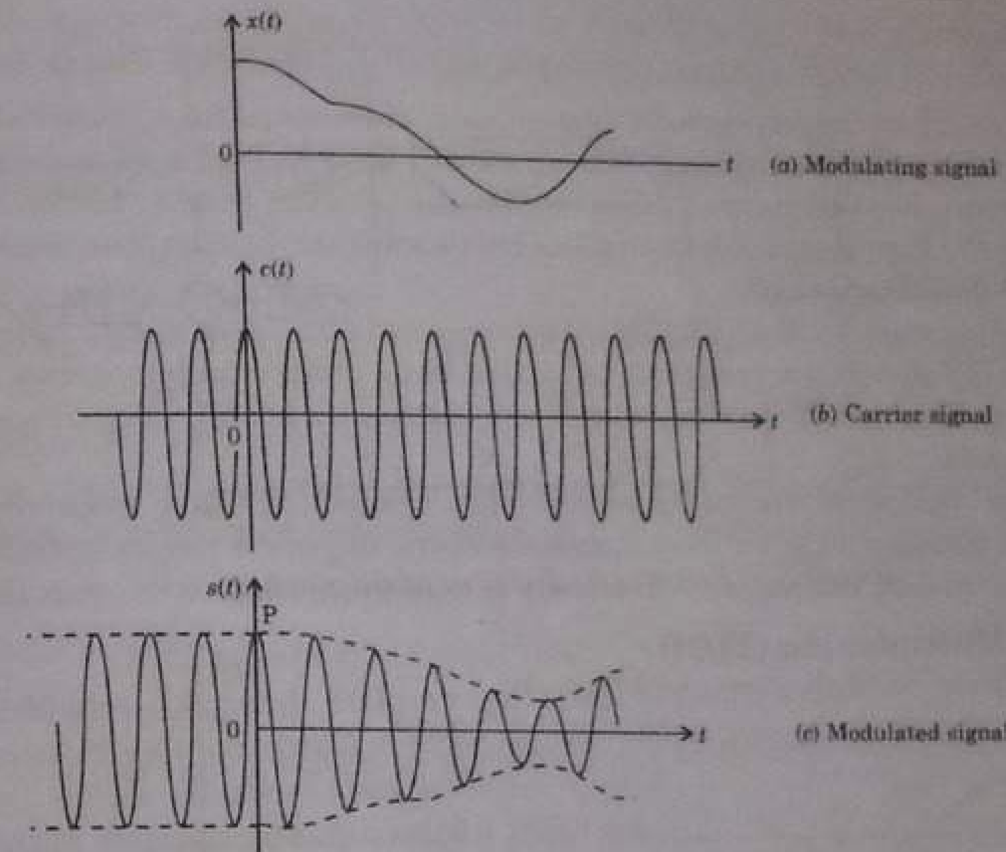


Fig. 3.2. *Illustration of amplitude modulation.*

(iii) The resulting signal from the process of amplitude-modulation is called amplitude modulated signal or simply AM wave.

(iv) In the process of amplitude modulation, the frequency and phase of the carrier remain constant whereas the maximum amplitude varies according to the instantaneous value of the information signal.

(v) Equation (3.3) represents an amplitude modulated wave. This wave has a constant frequency ω_c and amplitude $A + x(t)$. This implies that the amplitude of the wave is changing around A in accordance with the value of the modulating signal $x(t)$. The frequency of the AM signal remains unchanged and is equal to ω_c.

(vi) Figure 3.2 (c) illustrates the process of amplitude modulation. It may be observed that upto point P modulating signal is not applied so there is no modulation and maximum amplitude of the carrier remains constant at A. Now, at point P, the modulating signal is applied. Therefore, after point P, the amplitude modulation occurs. This means that the maximum amplitude A of the carrier now varies in accordance with the instantaneous value of the modulating signal $x(t)$.

(vii) The AM wave has a time-varying amplitude called as the envelope of the AM wave. Figure 3.2(c) shows that the envelope of AM wave consists of the modulating or baseband signal. This means that the unique property of AM is that the envelope of the modulated carrier has the same shape as the message signal or baseband signal.

We know that the expression for AM wave is

$$s(t) = [A + x(t)] \cos \omega_c t = E(t) \cos \omega_c t \quad ...(3.4)$$

where $E(t) = A + x(t)$

$E(t)$ is called the envelope of AM wave. This envelope consists of the baseband signal $x(t)$. Hence, the modulating or baseband signal may be recovered from an AM wave by detecting the envelope.

3.3.1. Spectrum of *AM* Wave or Frequency-Domain Representation

If $x(t)$ is modulating signal and carrier signal is given by the expression

$$c(t) = A \cos \omega_c t \quad ...(3.5)$$

Then the equation for AM wave will be

$$s(t) = x(t)\cos\omega_c t + A\cos\omega_c t \qquad ...(3.6)$$

This equation describes the AM wave in time-domain. However, if we want to know the frequency description or frequencies present in AM wave, we will have to find its spectrum or frequency-domain representation. For this purpose, first we have to take the Fourier transform of AM wave.

Let $S(\omega)$ denote the Fourier transform of $s(t)$, $C(\omega)$ denotes the Fourier transform of $c(t)$ and $X(\omega)$ denotes the Fourier transform of $x(t)$.

Let the modulating signal $x(t)$ and its Fourier transform $X(\omega)$ be as shown in figure 3.3(a).

Let the modulating signal or message signal $x(t)$ be band limited to the interval $-\omega_m \le \omega \le \omega_m$ as shown in figure 3.3(a). This means that the modulating signal does not have any frequency component outside the interval $(-\omega_m, \omega_m)$.

It may be noted at this point that in figure 3.3(a) the modulating signal frequency ranges extend from $-\omega_m$ to ω_m i.e. it includes negative frequencies also from $-\omega_m$ to 0. Practically there is no meaning of negative frequency. In fact, the negative frequency is used for mathematical convenience only.

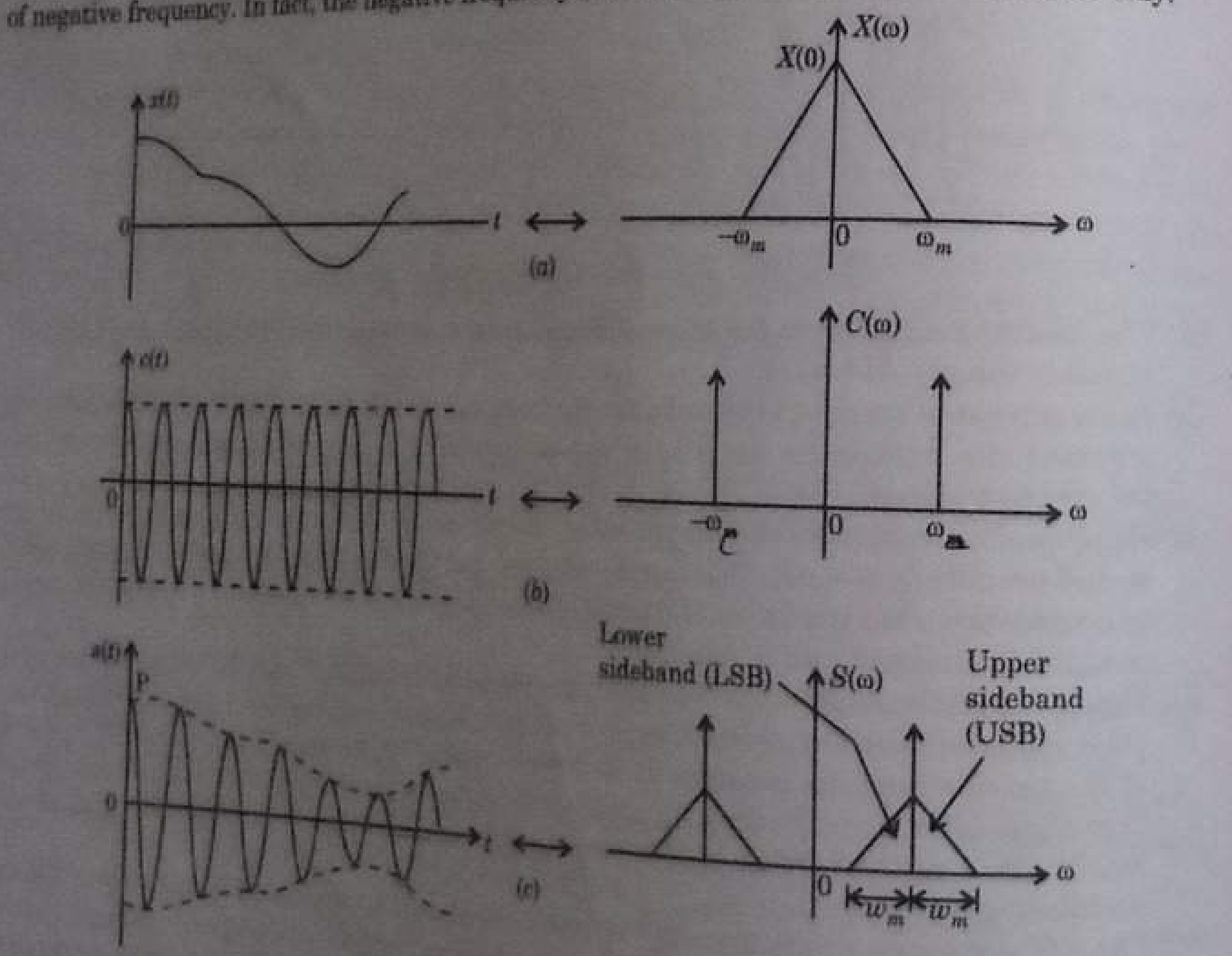


Fig. 3.3.

Hence, we can say that the modulating signal contains frequencies from 0 to ω_m or simply the bandwidth of modulating signal is ω_m. We know that the Fourier transform of a cosine signal $\cos\omega_c t$ consists of two impulse at ω_c and $-\omega_c$ as

$$\cos\omega_c t \leftrightarrow \pi[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)] \qquad ...(3.7)$$

Since the carrier signal is $c(t) = A\cos\omega_c t$, therefore

$$A\cos\omega_c t \leftrightarrow \pi A[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)] \qquad ...(3.8)$$

Figure 3.3(b) shows the carrier signal $A\cos\omega_c t$ and its Fourier transform.

Now, the AM wave is given as

$$s(t) = x(t)\cos\omega_c t + A\cos\omega_c t \qquad ...(3.9)$$

To find the spectrum of AM wave, we take its Fourier transform.

The Fourier transform of $s(t)$ may be found by considering the two factors $x(t)\cos\omega_c t$ and $A\cos\omega_c t$ separately as follow:

To find the Fourier transform of $x(t)\cos\omega_c t$, we first note the frequency-shifting theorem of Fourier transform as

If $\quad x(t) \leftrightarrow X(\omega)$

then $\quad e^{j\omega_c t}x(t) \leftrightarrow X(\omega-\omega_c) \qquad ...(3.10)$

This property states that if a signal $x(t)$ is multiplied by $e^{j\omega_c t}$ in time-domain, then its spectrum $X(\omega)$ in frequency-domain is shifted by an amount ω_c.

Similarly, $e^{-j\omega_c t}x(t) \leftrightarrow X(\omega+\omega_c) \qquad ...(3.11)$

But, since $e^{j\omega_c t}$ is not a real function and cannot be generated practically, therefore frequency shifting in practice is achieved by multiplying $x(t)$ by a sinusoid such as $\cos\omega_c t$.

Therefore,

$$x(t)\cos\omega_c t = x(t)\left[\frac{1}{2}\left(e^{j\omega_c t}+e^{-j\omega_c t}\right)\right] \qquad ...(3.12)$$

$$x(t)\cos\omega_c t = \frac{1}{2}x(t)e^{j\omega_c t}+\frac{1}{2}x(t)e^{-j\omega_c t} \qquad ...(3.13)$$

Hence, using equations (3.10) and (3.11), we get

$$x(t)\cos\omega_c t \leftrightarrow \frac{1}{2}\left[X(\omega-\omega_c)+X(\omega+\omega_c)\right] \qquad ...(3.14)$$

This means that the multiplication of a signal $x(t)$ by a sinusoid of frequency ω_c shifts the spectrum $X(\omega)$ by $\pm\omega_c$.

The Fourier transform of the second factor $A\cos\omega_c t$ will be as in equation (3.8)

$$A\cos\omega_c t \leftrightarrow \pi A[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)] \qquad ...(3.15)$$

This means that Fourier transform of $A\cos\omega_c t$ consists of two impulses at $\pm\omega_c$.

Therefore, the Fourier transform of AM wave given by equation (3.6) will be the sum of equations (3.14) and (3.15).

Thus, $$S(\omega) = \frac{1}{2}\left[X(\omega-\omega_c)+X(\omega+\omega_c)\right]+\pi A\left[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)\right] \qquad ...(3.16)$$

Above equation for Amplitude Modulated Wave contains two factors. The first factor given as $1/2\ [X(\omega-\omega_c)+X(\omega+\omega_c)]$ represents the spectrum of original or baseband signal shifted in the positive as well as in the negative direction by the factor ω_c. On the other hand, the second factor given as $\pi A[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)]$ represents the presence of carrier signal i.e., two impulses each having strength equal to πA.

Thus the spectrum of modulated signal contains shifted spectrum of modulating signal and the spectrum of carrier signal as shown in figure 3.3(c).

From figure 3.3(c) following points may be observed:

(i) For positive frequencies, a portion of the spectrum of AM wave is lying above the carrier frequency ω_c. This band of frequency which is lying above the carrier frequency ω_c is known as the **upper sideband (USB)** whereas the symmetrical portion below carrier frequency ω_c is known as the **lower sideband (LSB)**. For negative frequencies, the upper sideband (USB) is represented by the portion of the spectrum below $-\omega_c$ and the lower sideband by the symmetrical portion above $-\omega_c$. Two sidebands have been shown in figure 3.3(c).

(ii) We generally keep $\omega_c > \omega_m$ which ensures that the two sidebands do not overlap each other.

(iii) At this point, it may be noted that the negative frequencies appeared in spectrum analysis is due to the exponential representation of Fourier transform. These negative frequencies are used for mathematical convenience. For a real function, these negative frequencies have no meaning. For general purpose, it is sufficient to consider only the positive frequency region while treating the negative frequency region as a replica of the positive frequencies. This means that to calculate the BW of a signal, we consider only positive side.

(ii) From figure, it is obvious that for positive side, the highest frequency component present in the spectrum of AM wave is $\omega_c + \omega_m$ and the lowest frequency component is $\omega_c - \omega_m$. Infact, the difference between these two extreme frequencies is equal to the bandwidth of the AM wave.

Therefore, Bandwidth $B = (\omega_c + \omega_m) - (\omega_c - \omega_m)$

$$B = 2\omega_m$$

Thus, it is clear that the bandwidth of the amplitude modulated wave is twice the highest frequency present in the baseband or modulating signal.

3.4. Modulation Index

In AM system the modulation index is defined as the measure of extent of amplitude variation about an unmodulated maximum carrier. It is represented by m_a.

Mathematically,

$$\text{Modulation Index} \quad m_a = \frac{|x(t)|_{max}}{\text{Maximum carrier amplitude}}$$

$$\text{or Modulation Index} \quad m_a = \frac{|x(t)|_{max}}{A} \qquad ...(3.17)$$

where $|x(t)|_{max}$ represents the maximum amplitude of modulating signal and A represents the maximum amplitude of carrier signal.

The modulation index is also known as depth of modulation, degree of modulation or modulation factor.

Also, the absolute value of m_a multiplied by 100 is known as **percentage modulation.**

3.4.1. Overmodulation

We know that modulation index is given as

$$m_a = \frac{|x(t)|_{max}}{A} \qquad ...(3.18)$$

where 'A' is the maximum amplitude of the carrier signal.

The baseband or modulating signal will be preserved in the envelope of the *AM* signal only if we have

$$|x(t)|_{max} \leq A$$

i.e. modulation index m_a is less than or equal to unity.

In other words, the modulating signal is preserved in the envelope of *AM* signal only if the percentage modulation is less than or equal to 100 per cent.

On the other hand, if $m_a > 1$ or the per centage modulation is greater than 100, the baseband signal is not preserved in the envelope. It means that in this case, the baseband signal recovered from the envelope will be distorted. This type of distortion is called envelope distortion and the AM signal with $m_a > 1$ or $m_a > 100\%$ is called overmodulated signal.

Let us consider a baseband signal $x(t)$ and a carrier signal $A \cos \omega_c t$ as shown in figure 3.4(*a*) and (*b*). On Amplitude modulation, following two cases arise depending upon the amplitude of the baseband signal relative to the maximum amplitude of the carrier signal.

(i) Amplitude Modulation with $m_a < 1$

In this case, the waveform of the AM signal is as shown in figure 3.4(*c*). Here the maximum amplitude of baseband signal is less than maximum carrier amplitude *A*, *i.e.*,

$$|x(t)|_{max} < A$$

From figure, it may be observed that the envelope is not reaching the zero-amplitude axis of the waveform and so the baseband signal may be fully recovered from the envelope of AM wave.

(ii) Amplitude Modulation with $m_a > 1$

In this case, the waveform of the *AM* signal is as shown in figure 3.4(*d*). Here the amplitude of a baseband signal exceeds maximum carrier amplitude *i.e.*

$$|x(t)|_{max} > A$$

In this case, the modulation index m_a is more than 1 or 100%.

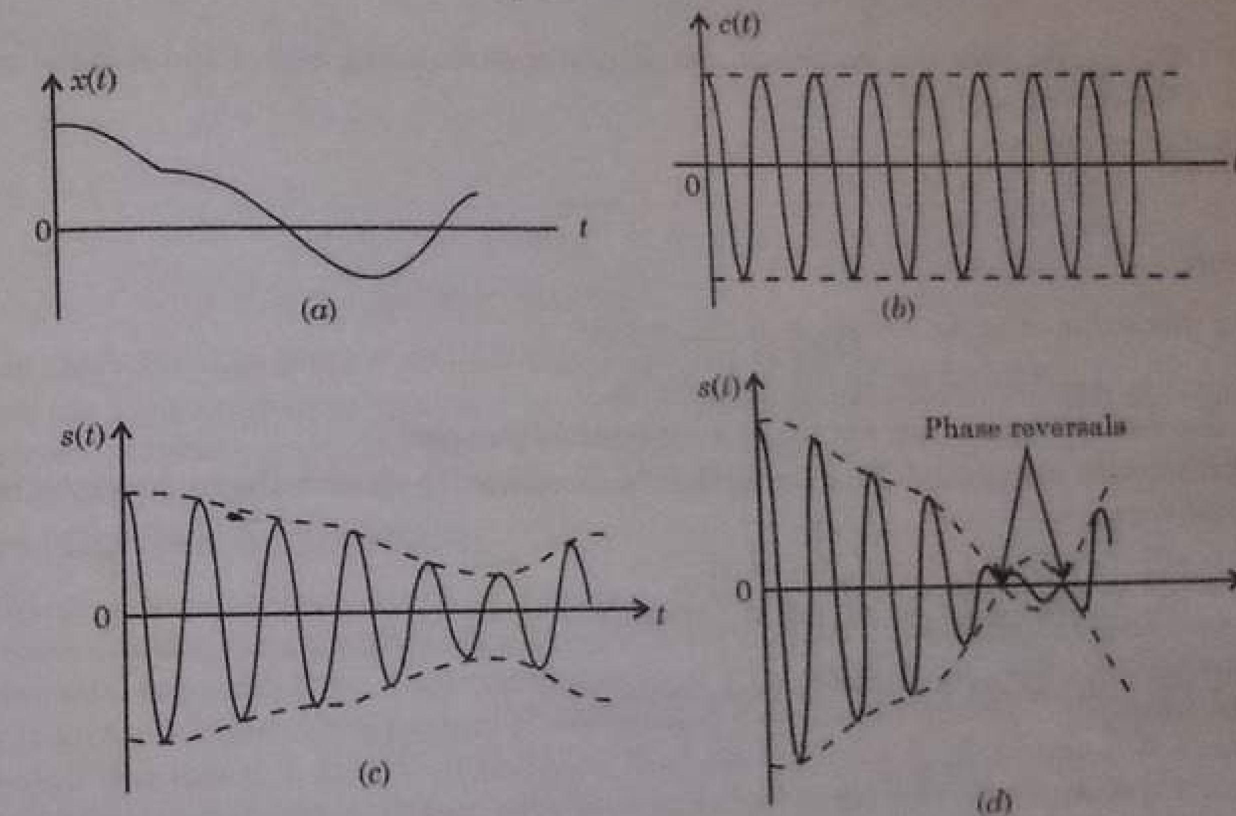


Fig. 3.4.

3.5. Single Tone Amplitude Modulation (*AM*) *(Important)*

Till now, we discussed amplitude modulation in which we assumed that baseband or modulating signal is a random signal which contains a large number of frequency components. This means that a carrier signal (fixed frequency signal) is modulated by a large number of frequency components.

In this section, we shall discuss amplitude modulation in which the modulating or baseband signal consists of only one (single) frequency *i.e.* modulation is done by a single frequency or tone. This type of amplitude modulation is known as **single tone amplitude modulation.**

Let us consider a single tone modulating signal as

$$x(t) = V_m \cos \omega_m t^* \qquad ...(3.19)$$

which contains a single frequency ω_m.

Let the carrier signal be

$$c(t) = A \cos \omega_c t \qquad ...(3.20)$$

We know that the general expression for *AM* signal is

$$s(t) = [A + x(t)] \cos \omega_c t \qquad ...(3.21)$$

or

$$s(t) = A \cos \omega_c t + x(t) \cos \omega_c t$$

Putting the value of $x(t)$, we get

$$s(t) = A \cos \omega_c t + V_m \cos \omega_m t \cos \omega_c t$$

or

$$s(t) = A \cos \omega_c t + V_m \cos \omega_c t \cos \omega_m t$$

* Signal $x(t)$ may be a voltage signal or a current signal. Here, we have assumed that $x(t)$ is a voltage signal with maximum amplitude equal to V_m.

or
$$s(t) = A \cos\omega_c t\left[1+\frac{V_m}{A}\cos\omega_m t\right] \quad ...(3.22)$$

But we know that for *AM* the modulation index m_a is given as

$$m_a = \frac{|x(t)|_{max}}{A}$$

where $|x(t)|_{max}$ denotes the maximum amplitude of modulating signal and A is the maximum amplitude of the carrier signal.

In this case, we have

$$|x(t)|_{max} = V_m$$

Therefore,
$$m_a = \frac{V_m}{A}$$

Putting this value of m_a in equation (3.22), we get

$$s(t) = A\cos\omega_c t\,[1 + m_a \cos\omega_m t] \quad ...(3.23)$$

This is the desired expression for single-tone modulating signal.

The expression in equation (3.23) may further be simplified to observe the frequency components present in *AM* signal.

$$s(t) = A\cos\omega_c t\,[1 + m_a\cos\omega_m t]$$

or
$$s(t) = A\cos\omega_c t + A\cdot m_a\cos\omega_c t\cos\omega_m t$$

or
$$s(t) = A\cos\omega_c t + \frac{A\cdot m_a}{2}[2\cos\omega_c t\cos\omega_m t]$$

$$= A\cos\omega_c t + \frac{A\cdot m_a}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$$

or
$$s(t) = A\cos\omega_c t + \frac{A\cdot m_a}{2}\cos(\omega_c+\omega_m)t + \frac{A\cdot m_a}{2}\cos(\omega_c-\omega_m)t \quad ...(3.24)$$

Equation (3.24) reveals that the *AM* signal has three frequency components as follow:

(i) carrier frequency ω_c having amplitude A

(ii) upper sideband $(\omega_c + \omega_m)$ having amplitude $\frac{m_a\cdot A}{2}$,

(iii) lower sideband $(\omega_c - \omega_m)$ having amplitude $\frac{m_a\cdot A}{2}$

With the help of these frequency components, we can plot the frequency-spectrum of single-tone amplitude modulated (AM) wave. Figure 3.4a shows the one-sided frequency spectrum of single-tone AM wave.

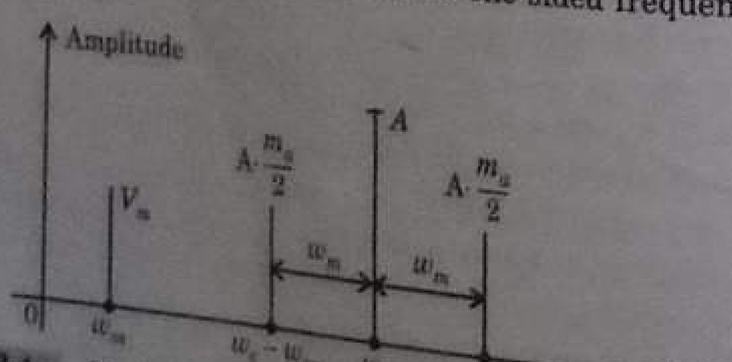


Fig. 3.4a. *Single-sided frequency spectrum of single-tone AM wave.*

Example 3.1. The tuned-circuit of the oscillator in an AM transmitter uses a 50 µH coil and a 1 nF capacitor. Now, if the oscillator output is modulated by audio frequencies upto 8 kHz, then find the frequency range occupied by the sidebands.

Solution: The oscillator in AM transmitter is used to generate high carrier frequency. Hence, the resonance frequency of the oscillator will be the carrier frequency.

Therefore,

Carrier frequency, $f_c = \frac{1}{2\pi\sqrt{LC}}$

Here given that $L = 50\ \mu H$

$L = 50 \times 10^{-6}$ H

and $C = 1\ nF = 1 \times 10^{-9}$ F

Thus,
$$f_c = \frac{1}{2\pi\sqrt{50\times10^{-6}\times1\times10^{-9}}} = \frac{1}{2\pi\sqrt{5\times10^{-14}}} = \frac{1}{2\pi\times10^{-7}\times\sqrt{5}}$$

$$f_c = 7.12 \times 10^5 \text{ Hz} = 712 \text{ kHz}$$

Now, it is given that the highest modulating frequency is 8 kHz.

Therefore, the frequency range occupied by the sidebands will range from 8 kHz above to 8 kHz below the carrier frequency, extending from 7 to 4 kHz to 720 kHz. **Ans.**

3.6. Power Content in AM Wave

It may be observed from the expression of *AM* wave that the carrier component of the amplitude modulated wave has the same amplitude as unmodulated carrier. In addition to carrier component, the modulated wave consists of two sideband components. It means that the modulated wave contains more power than the unmodulated carrier. However, since the amplitudes of two sidebands depend upon the modulation index, it may be anticipated that the total power of the amplitude modulated wave would depend upon the modulation index also. In this section, we shall find the power contents of the carrier and the sidebands.

We know that the general expression of *AM* wave is given as

$$s(t) = A\cos\omega_c t + x(t)\cos\omega_c t \quad ...(3.25)$$

The total power P of the *AM* wave is the sum of the carrier power P_c and sideband power P_s.

Carrier Power

The carrier power P_t is equal to the mean-square (*ms*) value of the carrier term $A\cos\omega_c t$ *i.e.*

$$P_c = \text{mean square value of } A\cos\omega_c t$$

$$P_c = [A\cos\omega_c t]^2 = \frac{1}{2\pi}\int_0^{2\pi} A^2\cos^2\omega_c t.\,dt^* = \frac{A^2}{2} \quad ...(3.26)$$

Sideband Power

The sideband power P_s is equal to the mean square value of the sideband term $x(t)\cos\omega_c t$, *i.e.*

$$P_s = \text{mean square value of } x(t)\cos\omega_c t$$

$$P_s = [x(t)\cos\omega_c t]^2 = \frac{1}{2\pi}\int_0^{2\pi} x^2(t)\cos^2\omega_c t\,dt$$

$$P_s = \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\left[2\cos^2\omega_c t\right]x^2(t)\,dt$$

or
$$P_s = \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}x^2(t)\,dt + \frac{1}{2\pi}\int_0^{2\pi}x^2(t)\cos 2\omega_c t\,dt \quad ...(3.27)$$

* Since period of the signal $A\cos\omega_c t$ is 2π.

In AM generation, a bandpass filter (BPF) or a tuned circuit tuned to carrier frequency ω_c is used to filter out the second integral term.

Therefore,

$$P_s = \frac{1}{2\pi}\int_0^{2\pi}\left[\frac{1}{2}x^2(t)\right]dt$$

or $$P_s = \text{mean square } (ms) \text{ value of } \frac{1}{2}x^2(t)$$

or $$P_s = \frac{1}{2}\overline{x^2(t)} \quad ...(3.28)$$

However, the total sideband power P_s is due to the equal contributions of the upper and lower sidebands. Hence, the power carried by the upper and the lower sidebands will be

$$P_{s(LSB)} = P_{s(USB)} = \frac{P_s}{2} = \frac{1}{4}\overline{x^2(t)} \quad ...(3.29)$$

Therefore, the total power P_t of the AM signal is the sum of the carrier power P_c and sideband power P_s.

$$P_t = P_c + P_s = \frac{1}{2}A^2 + \frac{1}{2}\overline{x^2(t)}$$

$$P_t = \frac{1}{2}[A^2 + \overline{x^2(t)}] \quad ...(3.30)$$

3.7. Transmission Efficiency of Amplitude Modulated Signal

We know that the total modulated power of an AM signal is expressed as

$$P_t = P_c + P_s = \frac{1}{2}[A^2 + \overline{x^2(t)}] \quad ...(3.31)$$

Out of this total power P_t, the useful message or baseband power is the power carried by the sidebands, *i.e.* P_s. The large carrier power P_c is a waste from the transmission point of view because it does not carry any information or message. This large carrier power P_c is transmitted alongwith the sideband power only for the convenient and cheap detection. Hence, P_s is the only useful message power present in the AM wave.

In AM wave, the amount of useful message power P_s may be expressed by a term known as **transmission efficiency η.**

Hence transmission efficiency of AM wave may be defined as the per centage of total power contributed by the sidebands.

Mathematically,

Transmission efficiency, $$\eta = \frac{P_s}{P_t}\times 100 \quad ...(3.32)$$

or $$\eta = \frac{\frac{1}{2}\overline{x^2(t)}}{\frac{1}{2}A^2 + \frac{1}{2}\overline{x^2(t)}}\times 100$$

or $$\eta = \frac{100\,\overline{x^2(t)}}{A^2 + \overline{x^2(t)}} \quad ...(3.33)$$

The maximum transmission efficiency of the AM is only 33.33%. This implies that only one-third of the total power is carried by the sidebands and the rest two-third is wasted.

3.8. Power of a Single-Tone Amplitude-Modulated (AM) Signal

In article 3.6, we have found the power content of the AM signal when modulating-signal is any random signal and may consist of several frequency components. Likewise we can find power content of single-tone Amplitude Modulated (AM) signal.

Let us consider that a carrier signal $A\cos\omega_c t$ is amplitude-modulated by a single-tone modulating signal $x(t) = V_m\cos\omega_m t$.

Then the unmodulated or carrier power

$$P_c = \text{mean square } (ms \text{ value})$$

$$P_c = \overline{(A\cos\omega_c t)^2} = \frac{A^2}{2}$$

The sideband power $$P_s = \frac{1}{2}\overline{x^2(t)} = \frac{1}{2}\overline{(V_m\cos\omega_m t)^2}$$

$$P_s = \frac{1}{2}\frac{V_m^{\,2}}{2} = \frac{1}{4}V_m^{\,2} \quad ...(3.34)$$

We know that the total modulated power P_t is the sum of P_c and P_s.

Therefore

$$P_t = P_c + P_s = \frac{A^2}{2} + \frac{1}{4}\cdot V_m^{\,2}$$

or $$P_t = \frac{A^2}{2}\left[1 + \frac{1}{2}\left(\frac{V_m}{A}\right)^2\right]$$

But $$\frac{V_m}{A} = \frac{\text{Maximum baseband amplitude}}{\text{Maximum carrier amplitude}} = m_a$$

$$= \text{modulation index for } AM$$

Hence

$$P_t = \frac{A^2}{2}\left[1 + \frac{1}{2}\cdot m_a^{\,2}\right]$$

But $$\frac{A^2}{2} = P_c = \text{carrier power}$$

Therefore, $$P_t = P_c\left(1 + \frac{m_a^{\,2}}{2}\right) \quad ...(3.3$$

Example 3.2. A 400 watts carrier is modulated to a depth of 75 per cent. Find the total pow in the amplitude-modulated wave. Assume the modulating signal to be a sinusoidal one.

Solution: We know that for a sinusoidal modulating signal, the total power is expressed as

$$P_t = P_c\left(1 + \frac{m_a^{\,2}}{2}\right)$$

where
P_t = total power or modulated power
P_c = carrier power or unmodulated power
m_a = modulation index

Given that,
P_c = 400 watts
m_a = 75 per cent = 0.75

Therefore, $P_t = P_c\left(1+\frac{m_a^2}{2}\right) = 400\left(1+\frac{0.75^2}{2}\right)$

$P_t = 512.5$ watts **Ans.**

Example 3.3. An AM broadcast radio transfer radiates 10 K watts of power if modulation per centage is 60. Calculate how much of this is the carrier power.

Solution: We know that the total power is expressed as

$$P_t = P_c\left(1+\frac{m_a^2}{2}\right) \quad ...(i)$$

where P_t = total power or modulated power
P_c = carrier power or unmodulated power
m_a = modulation index

Given that, $P_t = 10$ K watts
$m_a = 60$ per cent $= 0.6$

From equation (i), we get

$$P_c = \frac{P_t}{1+\frac{m_a^2}{2}} = \frac{10}{1+\frac{0.6^2}{2}} = \frac{10}{1.18}$$

$P_c = 8.47$ kW **Ans.**

3.9. Current Calculation for Single-Tone AM

In AM, it is generally more convenient to measure the *AM* transmitter current than the power. In this case, the modulation index may be calculated from the values of unmodulated and modulated currents in the AM transmitter.

Let I_c be the r.m.s. value of the carrier or unmodulated current and I_t be the r.m.s. value of the total or modulated current of an *AM* transmitter. Let *R* be the antenna resistance through which these currents flow.

Now, we know that for a single-tone modulation the power relation is expressed as

$$P_t = P_c\left(1+\frac{m_a^2}{2}\right) \quad ...(3.36)$$

where P_t = total or modulated power
P_c = carrier or unmodulated power
m_a = modulation index

From equation (3.36), we may write

$$\frac{P_t}{P_c} = 1+\frac{m_a^2}{2}$$

or $$\frac{I_t^2 \cdot R}{I_c^2 \cdot R} = 1+\frac{m_a^2}{2}$$

or $$\frac{I_t}{I_c} = \sqrt{1+\frac{m_a^2}{2}}$$

or $$I_t = I_c\sqrt{1+\frac{m_a^2}{2}} \quad ...(3.37)$$

Example 3.4. The antenna current of an AM transmitter is 8 A, if only the carrier is sent, but it increases to 8.93 A, if the carrier is modulated by a single sinusoidal wave. Determine the per centage modulation. Also find the antenna current if the per cent of modulation changes to 0.8. *(U.P. Tech. Sem., Exam., 2004-05) (06 marks)*

Solution: (*i*) The current relation for a single-tone amplitude modulation is expressed as

$$I_t = I_c\sqrt{1+\frac{m_a^2}{2}} \quad ...(i)$$

where I_t = total or modulated current
I_c = carrier or unmodulated current
m_a = modulation index

Using equation (*i*), we get

$$\frac{I_t}{I_c} = \sqrt{1+\frac{m_a^2}{2}}$$

or $$\left(\frac{I_t}{I_c}\right)^2 = 1+\frac{m_a^2}{2}$$

or $$\frac{m_a^2}{2} = \left(\frac{I_t}{I_c}\right)^2 - 1$$

or $$m_a^2 = 2\left[\left(\frac{I_t}{I_c}\right)^2 - 1\right]$$

or $$m_a = \sqrt{2\left[\left(\frac{I_t}{I_c}\right)^2 - 1\right]}$$

Putting all the given values, we have

$$m_a = \sqrt{2\left[\left(\frac{8.93}{8}\right)^2 - 1\right]} = \sqrt{2[(1.116)^2 - 1]}$$

$$m_a = \sqrt{2(1.246-1)} = \sqrt{0.492} = 0.701 = 70.1\%$$

(*ii*) Since $I_t = I_c\sqrt{1+\frac{m_a^2}{2}}$

Here, $I_c = 8$ A and $m_a = 0.8$

Therefore, $I_t = 8\times\sqrt{1+\frac{0.8^2}{2}} = 8\sqrt{1+\frac{0.64}{2}}$

or $I_t = 8\sqrt{1.32} = 8\times 1.149 = 9.19$ A **Ans.**

3.10. Power Content in Multiple-Tone Amplitude Modulation (AM)

A multiple-tone amplitude modulation is that type of modulation in which the modulating signal consists of more than one frequency components.

Let us consider that a carrier signal $A\cos\omega_c t$ is modulated by a baseband or modulating signal $x(t)$ which is expressed as

$$x(t) = V_1\cos\omega_1 t + V_2\cos\omega_2 t + V_3\cos\omega_3 t \quad ...(3.38)$$

We know that the general expression for AM wave is

$$s(t) = A\cos\omega_c t + x(t)\cos\omega_c t \quad ...(3.39)$$

Putting the value of $x(t)$, we get

$$s(t) = A\cos\omega_c t + [V_1\cos\omega_1 t + V_2\cos\omega_2 t + V_3\cos\omega_3 t]\cos\omega_c t$$

or

$$s(t) = A\left[1 + \frac{V_1}{A}\cos\omega_1 t + \frac{V_2}{A}\cos\omega_2 t + \frac{V_3}{A}\cos\omega_3 t\right]\cos\omega_c t \quad ...(3.40)$$

But we know that

$$\frac{V}{A} = \frac{\text{Maximum amplitude of modulating signal}}{\text{Maximum amplitude of carrier signal}} = \text{modulation index } m_a$$

Therefore,

$$s(t) = A[1 + m_1\cos\omega_1 t + m_2\cos\omega_2 t + m_3\cos\omega_3 t]\cos\omega_c t \quad ...(3.41)$$

where

$m_1 = \frac{V_1}{A}$, $m_2 = \frac{V_2}{A}$ and $m_3 = \frac{V_3}{A}$, are the modulation indexes of the corresponding frequency components.

The expression for AM wave in equation (3.41) may further be expanded as

$$s(t) = A\cos\omega_c t + m_1 A\cos\omega_c t\cos\omega_1 t + m_2 A\cos\omega_c t\cos\omega_2 t + m_3 A\cos\omega_c t\cos\omega_3 t$$

Now we know that the total power in *AM* is given as

$$P_t = \text{carrier power} + \text{sideband power}$$

$$P_t = P_c + P_s \quad ...(3.42)$$

The carrier power P_c is given as

$$P_c = \overline{(A\cos\omega_c t)^2} = \frac{A^2}{2}$$

Similarly, the sideband power is given as

$$P_s = \frac{1}{2}\overline{x^2(t)}$$

$$P_s = \frac{1}{2}[\overline{(V_1\cos\omega_1 t)^2} + \overline{(V_2\cos\omega_2 t)^2} + \overline{(V_3\cos\omega_3 t)^2}] \quad ...(3.43)$$

But we know that

$$m_1 = \frac{V_1}{A} \text{ so that } V_2 = m_1 A$$

$$m_2 = \frac{V_2}{A} \text{ so that } V_2 = m_2 A$$

and

$$m_3 = \frac{V_3}{A} \text{ so that } V_3 = m_3 A$$

Putting the values of V_1, V_2 and V_3 in equation (3.43), we get

$$P_s = \frac{1}{2}[\overline{(m_1 A\cos\omega_1 t)^2} + \overline{(m_1 A\cos\omega_1 t)^2} + \overline{(m_3 A\cos\omega_3 t)^2}]$$

$$P_s = \frac{1}{2}\left[\frac{m_1 A^2}{2} + \frac{m_2 A^2}{2} + \frac{m_3 A^2}{2}\right]$$

or

$$P_s = \frac{1}{4}A^2[m_1^2 + m_2^2 + m_3^2] \quad ...(3.44)$$

Now putting the value of P_c and P_s in equation (3.41), we get

$$P_t = P_c + P_s = \frac{A^2}{2} + \frac{1}{4}A^2[m_1^2 + m_2^2 + m_3^2]$$

or

$$P_t = \frac{A^2}{2}\left[1 + \frac{1}{2}(m_1^2 + m_2^2 + m_3^2)\right] = P_c\left[1 + \frac{m_1^2 + m_2^2 + m_3^2}{2}\right]$$

or

$$P_t = P_c\left[1 + \frac{m_1^2}{2} + \frac{m_2^2}{2} + \frac{m_3^2}{2}\right]$$

This expression may be extended upto to n-modulating terms *i.e.*

$$P_t = P_c\left[1 + \frac{m_1^2}{2} + \frac{m_2^2}{2} + \frac{m_3^2}{2} + ... + \frac{m_n^2}{2}\right] \quad ...(3.45)$$

3.10.1. Total or Net Modulation Index for Multiple-Tone Modulation

Let us consider that m_t is the total or net modulation indexes for a multiple-tone modulation.

We know that for a multiple-tone modulation, the total power is expressed as

$$P_t = P_c\left[1 + \frac{m_1^2}{2} + \frac{m_2^2}{2} + \frac{m_3^2}{2} + ... + \frac{m_n^2}{2}\right] \quad ...(3.46)$$

where m_1, m_2 .... m_n are the modulation indexes for different modulating signals.

The power for *AM* wave is also expressed as

$$P_t = P_c\left(1 + \frac{m_t^2}{2}\right) \quad ...(3.47)$$

Comparing equations (3.46) and (3.47), we get

$$m_t^2 = m_1^2 + m_2^2 + m_3^2 + ... + m_n^2$$

or

$$m_t = \sqrt{m_1^2 + m_2^2 + m_3^2 + ... + m_n^2} \quad ...(3.4$$

This is the desired expression for the total or net modulation index.

Example 3.5. An AM transmitter radiates 9 K watts of power when the carrier is unmodulat and 10.125 K watts when the carrier is sinusoidally modulated. Find the modulation ind per centage of modulation. Now, if another sine wave, corresponding to 40 per cent modulati is transmitted simultaneously, then calculate the total radiated power. *(Very Importa*

Solution: (*i*) We know that for a single-tone sinusoidal amplitude-modulation the total powe expressed as

$$P_t = P_c\left(1 + \frac{m_a^2}{2}\right)$$

where
P_t = modulated or total power
P_c = unmodulated or carrier power
m_a = modulation index

Given that, P_t = 10.125 kW

and P_c = 9 kW

Using equation (*i*), we get

$$1 + \frac{m_a^2}{2} = \frac{P_t}{P_c}$$

$$\frac{m_2^2}{2} = \frac{P_t}{P_c} - 1$$

or
$$\frac{m_2^2}{2} = \frac{10.125}{9} - 1 = 1.125 - 1 = 0.125$$

or
$$m_2^2 = 0.125 \times 2 = 0.250 = 0.50$$

or

(*ii*) We know that in case of modulation by two sinusoidal waves, the total modulation index m_t is expressed as

$$m_t = \sqrt{m_1^2 + m_2^2}$$

Let $m_1 = m_a = 0.5$

Given that $m_2 = 0.4$

Therefore, $m_t = \sqrt{(0.5)^2 + (0.4)^2} = \sqrt{0.25 + 0.16} = \sqrt{0.41} = 0.64$

The total radiated power in this case will be

$$P_t = P_c\left(1 + \frac{m_t^2}{2}\right) = 9 \times \left(1 + \frac{0.64^2}{2}\right)$$

$$P_t = 9\,(1 + 0.205) = 10.84 \text{ kW}$$ **Ans.**

3.11. Generation of Amplitude Modulation (AM)

The device which is used to generate an amplitude modulated (AM) wave is known as Amplitude Modulator.

The methods of AM Generation may be broadly classified as follow:

(*i*) Low-level AM Modulation

(*ii*) High-level AM Modulation

3.11.1. Low Level Amplitude Modulation

Figure 3.5(*a*) shows the block diagram of a low level AM modulation system. In a low-level amplitude modulation system, the modulation is done at low power level. At low power levels, a very small power is associated with the carrier signal and the modulating signal. Because of this, the output power of modulation is low. Therefore, the power amplifiers are required to boost the amplitude-modulated signals upto the desired output level. From block diagram in figure 3.5(*a*), it is clear that modulation is done at low power level. After this, the amplitude-modulated signal (*i.e.* a signal containing a carrier and two sidebands) is applied to a wideband power amplifier. A wideband power amplifier is used just to preserve the sidebands of the modulated signal. Amplitude modulated systems, employing modulation at low power levels are also called **low-level amplitude modulation transmitters.**

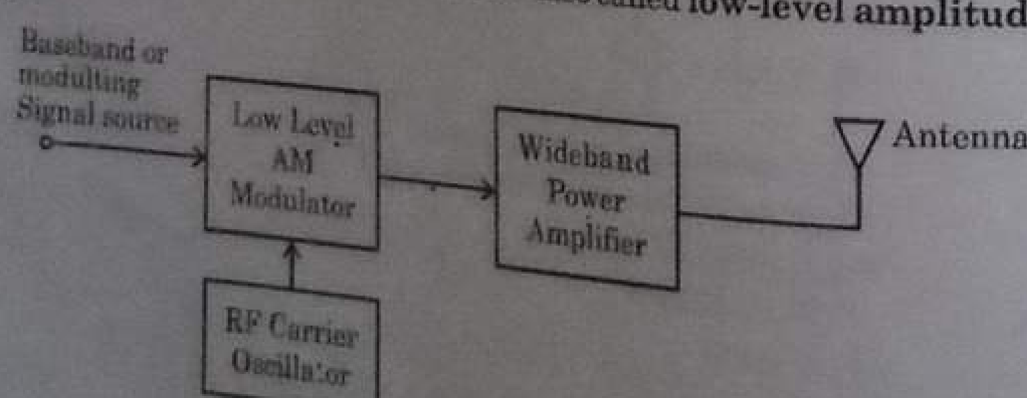


Fig. 3.5. *(a) Block Diagram for Low level AM Modulation.*

Square-low diode modulation and switching modulation are ...

3.11.2. High Level Amplitude Modulation

Figure 3.5(*b*) shows the block diagram of a high level AM modulation system. In a high-level amplitude modulation system, the modulation is done at high power level. Therefore, to produce amplitude-modulation at these high power levels, the baseband signal and the carrier signal must be at high power levels. In block diagram of figure 3.5(*b*) the modulating signal and carrier signal are first power amplified and then applied to AM high-level modulator. For modulating signal, the wide-band power amplifier is required just to preserve all the frequency components present in modulating signal. On the other hand, for carrier signal, the narrow band power amplifier is required because it is a fixed-frequency signal. The collector modulation method is the example of high-level modulation.

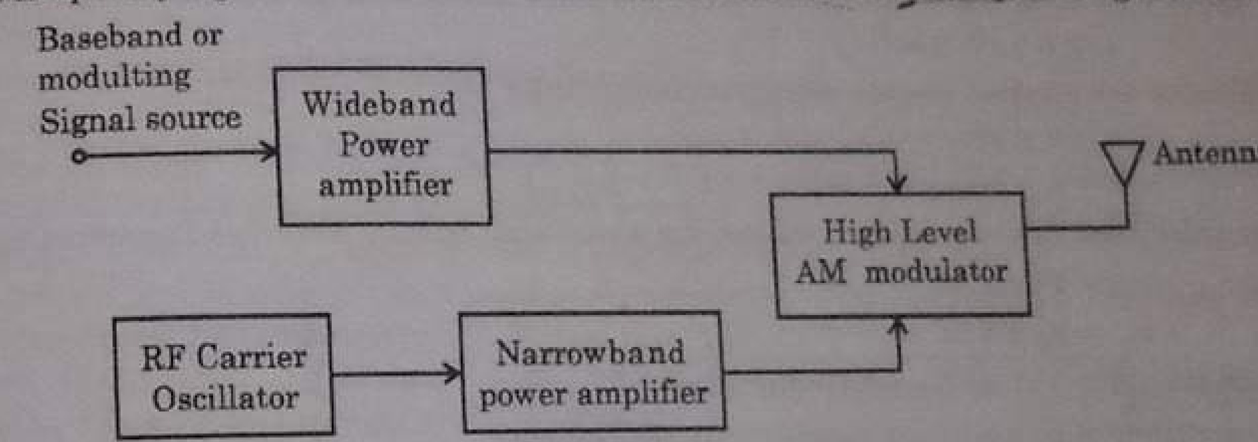


Fig. 3.5. *(b) Block diagram for high level AM Modulation.*

Before we discuss low level and high level modulation methods in detail, we shall establish the fact that a non-linear resistance of a non-linear device can be made to produce Amplitude Modulation when two different frequencies are passed together through it.

3.11.3. Non-linear resistance or Non-linear Circuits

We know that the relationship between voltage and current in a linear resistance is expressed as

$$i = bv \qquad \text{...(3.49)}$$

where

v = voltage across the linear resistance

i = current through linear resistance

and b = any constant of proportionality

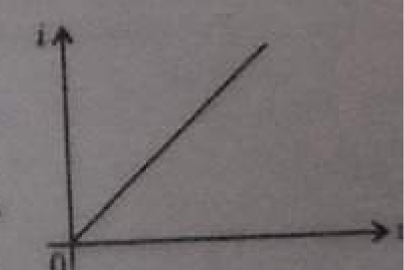


Fig. 3.6. *Linear resistance characteristics*

If equation (3.49) is applied to a resistor, then constant b is clearly its conductance.

Also, if equation (3.49) is applied to the linear portion of the transistor characteristic then i is the collector current and v is the voltage applied to the base.

As more general, equation (3.49), may be written as

$$i = a + bv \qquad \text{...(3.50)}$$

where a is the d.c. component of the current.

Now, let us consider a non-linear resistance. For a non-linear resistance the current-voltage characteristics will be non-linear as shown in figure 3.7.

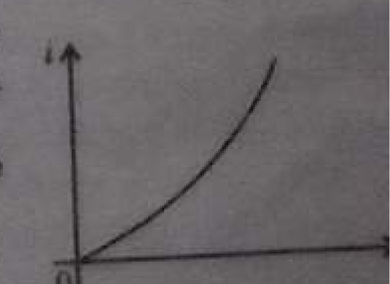


Fig. 3.7. *Non-linear resistance characteristics*

The non-linear relationship between voltage and current may be expressed as

$$i = a + bv + cv^2 + dv^3 + \ldots \qquad \text{...(3.51)}$$

This means that due to non-linearity in the v-i characteristics of a non-linear resistance, the current becomes proportional not only to voltage but also to the square, cube and higher powers of the voltage.

For simplicity, neglecting the higher terms in equation (3.51), we have

$$i = a + bv + cv^2 \quad ...(3.52)$$

Devices or circuits having non-linear v-i characteristics can be treated as a non-linear resistance. For example, devices like diodes, transistors, *FET* etc. exhibit non-linear characteristics and hence work as a non-linear resistance.

Equation (3.52), may be used in relating the output current to the input voltage of a non-linear resistance.

We can apply this equation to the gate-voltage-drain-current characteristics of a *FET*, then

$$i = a + bv + cv^2$$

If two voltages are applied at gate simultaneously, then

$$i = a + b(v_1 + v_2) + c(v_1 + v_2)^2$$

$$i = a + b(v_1 + v_2) + c(v_1^2 + v_2^2 + 2\,v_1 v_2) \quad ...(3.53)$$

Let us consider that the two input voltages are sinusoidal having different frequencies, then,

$$v_1 = V_1 \sin \omega t$$

and $$v_2 = V_2 \sin pt$$

Here ω and p are the two different frequencies.

Putting the values of v_1 and v_2 in equation (3.53), we get

$$i = a + b(V_1 \sin \omega t + V_2 \sin pt) + c(V_1^2 \sin^2 \omega t + C_2^2 \sin^2 pt + 2V_1 V_2 \sin \omega t \sin pt)$$

or $$i = a + bV_1 \sin \omega t + bV_2 \sin pt + cV_1^2 \sin^2 \omega t + cV_2^2 \sin^2 pt + 2cV_1 V_2 \sin \omega t \sin pt$$

or $$i = a + bV_1 \sin \omega t + bV_2 \sin pt + \frac{1}{2} cV_1^2 (2\sin^2 \omega t) + \frac{1}{2} cV_2^2 (2\sin^2 pt) + cV_1 V_2 (2\sin \omega t \sin pt)$$

or $$i = a + bV_1 \sin \omega t + bV_2 \sin pt + \frac{1}{2} cV_1^2 (1 - \cos 2\omega t) + \frac{1}{2} cV_2^2 (1 - \cos 2pt) + cV_1 V_2 [\cos(\omega - p)t - \cos(\omega + p)t$$

or $$i = a + bV_1 \sin \omega t + bV_2 \sin pt + \frac{1}{2} cV_1^2 - \frac{1}{2} cV_1^2 \cos 2\omega t + \frac{1}{2} cV_2^2 - \frac{1}{2} cV_2^2 \cos 2pt + cV_1 V_2 \cos(\omega - p)t - cV_1 V_2 \cos(\omega + p)t$$

or $$i = \underbrace{\left(a + \frac{1}{2} cV_1^2 + \frac{1}{2} cV_2^2\right)}_{(1)} + \underbrace{bV_1 \sin \omega t}_{(2)} + \underbrace{bV_2 \sin pt}_{(3)} - \underbrace{\left(\frac{1}{2} cV_1^2 \cos 2\omega t + \frac{1}{2} cV_2^2 \cos 2pt\right)}_{(4)} + \underbrace{cV_1 V_2 \cos(\omega - p)t}_{(5)} \underbrace{- cV_1 V_2 \cos(w + p)\,t}_{(6)} \quad ...(3.54)$$

In equation (3.54), if ω is assumed as the carrier frequency and p is assumed as modulating frequency then we can identify the different terms as follow:

term (1) is the d.c. component
term (2) is the carrier signal
term (3) is the modulating signal
term (4) contains harmonics of the carrier signal and the modulating signal
term (5) is the lower sideband
term (6) is the upper sideband

Since, the process of amplitude modulation consists of carrier signal, lower sideband and upper sideband, therefore we can conclude from above that when two voltages of different frequencies are passed through a non-linear resistance, the Amplitude Modulati...

In a practical modulation circuit, the frequencies other than the carrier and two sideband frequencies are rejected with the help of a tuned circuit.

3.11.4. Square Law Diode Modulation

Square law diode modulation circuit make use of nonlinear current-voltage characteristics of diode.* This method is suited at low voltage levels because of the fact that current-voltage chara-cteristic of a diode is highly nonlinear particularly in the low voltage region as shown in figure 3.8.

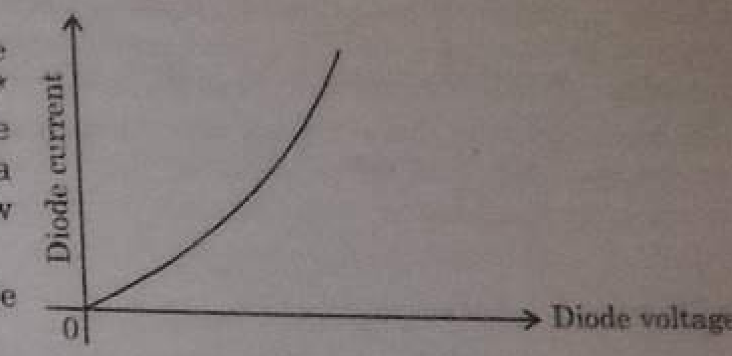


Fig. 3.8. *Current-voltage characteristic of a diode.*

Figure 3.9 shows the circuit of square law diode modulation.

It may be observed from the figure 3.9, that carrier and modulating signals are applied across the diode. A d.c. battery V_{cc} is connected across the diode to get a fixed operating point on the v–i characteristics of diode. The working of this circuit may be explained by considering the fact when two different frequencies are passed through a non-linear device, the process of amplitude modulation takes place. Hence, when carrier and modulating frequencies are applied at the input of diode, then different frequency terms appear at the output of diode. These different frequency terms are applied across a tuned circuit which is tuned to the carrier frequency and has a narrow bandwidth just to pass two sidebands alongwith the carrier and reject other frequencies. Hence at the output of tuned circuit, carrier and two sidebands are obtained *i.e.*, Amplitude Modulated (AM) wave is produced.

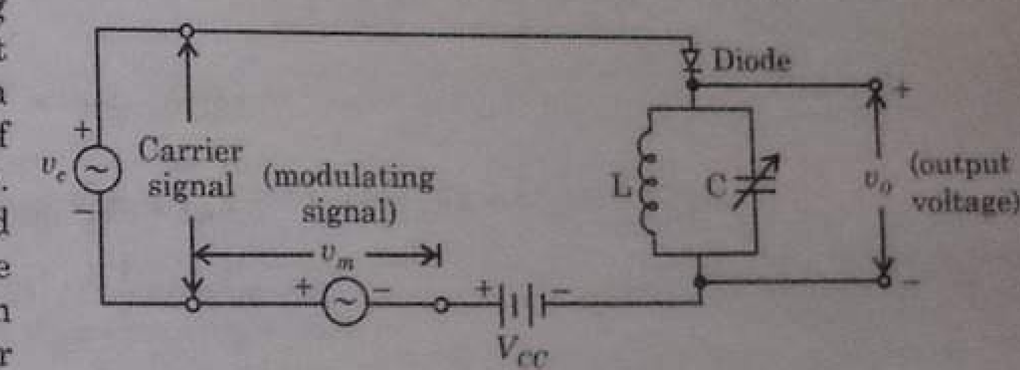


Fig. 3.9. *Square-low diode modulation.*

Mathematical Analysis

Let us consider that carrier voltage is expressed as

$$v_c = V_c \cos \omega_c t \quad ...(3.55)$$

where ω_c is the carrier frequency.

Let the modulating voltage be expressed as

$$v_m = V_m \cos \omega_m t \quad ...(3.56)$$

where w_m is the modulating frequency.

The total a.c. voltage across the diode is given as

$$v_s = v_c + v_m \quad ...(3.57)$$

$$v_s = V_c \cos \omega_c t + V_m \cos \omega_m t \quad ...(3.58)$$

The non-linear relationship between voltage and current for a diode is expressed as

$$i = a + bv_s + cv_s^2 \quad ...(3.59)$$

where a, b, and c are constants

i = current through the diode

v_s = voltage across the diode

* Explain the working of a modulator for generating AM wave. *(U.P. Tech, Sem. Exam., 2003-04) (05 marks)*

Putting the value of v_s from equation (3.58) in equation (3.59), we get

$$i = a + bv_s + cv_s^2 = a + b(V_c \cos \omega_c t + V_m \cos \omega_m t) + c(V_c \cos \omega_c t + V_m \cos \omega_m t)^2$$

$$i = a + bV_c \cos \omega_c t + bV_m \cos \omega_m t + c(V_c^2 \cos^2 \omega_c t + V_m^2 \cos^2 \omega_m t + 2V_c V_m \cos \omega_c t \cos \omega_m t)$$

or $$i = a + bV_c \cos \omega_c t + bV_m \cos \omega_m t + cV_c^2 \cos^2 \omega_c t + cV_m^2 \cos^2 \omega_m t + 2c.V_c.V_m \cos \omega_c t \cos \omega_m t$$

or $$i = a + bV_c \cos \omega_c t + bV_m \cos \omega_m t + \frac{1}{2} cV_c^2 (2\cos^2 \omega_c t)$$

or $$i = a + bV_c \cos \omega_c t + bV_m \cos \omega_m t + \frac{1}{2} cV_c^2 (2\cos^2 \omega_c t) + \frac{1}{2} cV_m^2 (2\cos^2 \omega_m t) + cV_c V_m (2 \cos \omega_c t \cos \omega_m t)$$

or $$i = a + bV_c \cos \omega_c t + bV_m \cos \omega_m t + \frac{1}{2} cV_c^2 (1 + \cos^2 \omega_c t) + \frac{1}{2} cV_m^2 (1 + \cos^2 \omega_m t) + cV_c V_m [\cos (\omega_c + \omega_m) t + \cos (\omega_c - \omega_m) t]$$

or $$i = a + bV_c \cos \omega_c t + bV_m \cos \omega_m t + \frac{1}{2} cV_c^2 + \frac{1}{2} cV_c^2 \cos^2 \omega_c t + \frac{1}{2} cV_m^2 + \frac{1}{2} cV_m^2 \cos 2\omega_m t + cV_c V_m \cos (\omega_c + \omega_m)t + cV_c V_m \cos (\omega_c - \omega_m)t$$

or $$i = \underbrace{\left(a + \frac{1}{2} cV_c^2 + \frac{1}{2} cV_m^2\right)}_{(1)} + \underbrace{bV_c \cos \omega_c t}_{(2)} + \underbrace{bV_m \cos \omega_m t}_{(3)} + \underbrace{\left(\frac{1}{2} cV_c \cos 2\omega_c t + \frac{1}{2} cV_m^2 \cos 2\omega_m t\right)}_{(4)} + \underbrace{cV_c V_m \cos (\omega_c + \omega_m)t}_{(5)} + \underbrace{cV_c V_m \cos (\omega_c - \omega_m)t}_{(6)} \quad ...(3.60)$$

Equation (3.60), consists of six terms in all as follow:

term (1) is the d.c. term

term (2) is the carrier signal

term (3) is the modulating signal

term (4) consists of harmonics of carrier and modulating signals

term (5) represents the upper sideband

term (6) represents the lower sideband

As discussed earlier, in the diode modulation circuit, the load impedance is a tuned circuit which is tuned to the carrier frequency ω_c. Therefore, this tuned circuit responds to a narrowband of frequencies centred about the carrier frequency ω_c. Thus the frequency components which are actually developed in the output are terms of frequency ω_c, $(\omega_c + \omega_m)$ and $(\omega_c - \omega_m)$. The rest of the frequency components are rejected by the tuned circuit.

Therefore, the required expression of output current will be

$$i_0 = bV_c \cos \omega_c t + c V_c V_m \cos (\omega_c + \omega_m)t + cV_c V_m \cos (\omega_c - \omega_m)t$$

or $$i_0 = bV_c \cos \omega_c t + cV_c V_m [\cos (\omega_c + \omega_m)t + \cos (\omega_c - \omega_m)t]$$

or $$i_0 = bV_c \cos \omega_c t + 2\, cV_c V_m \cos \omega_c t \cos \omega_m t$$

or $$i_0 = bV_c \left(1 + \frac{2cV_m}{b} \cos \omega_m t\right) \cos \omega_c t$$

$$i_0 = bV_c (1 + m_a \cos \omega_m t) \cos \omega_c t \quad ...(3.61)$$

where $m_a = \frac{2cV_m}{b}$ is the modulation index. Equation (3.61) is the required expression for AM current.

3.12. Collector Modulation Method

Collector modulation method is a very popular method for AM generation. Figure 3.10 illustrates the circuit diagram of a collector modulation method. Here, the transistor T_1 makes a radio frequency (*RF*) class-*C* amplifier. At the base of T_1, the carrier signal is applied. V_{cc} makes the collector supply used for biasing purpose. Also, the transistor T_2 makes a class-*B* amplifier which is used to amplify the audio or modulating signal. The baseband or modulating signal appears across the modulation transformer after amplification. This amplified baseband or modulating signal appears in series with the collector supply V_{cc}. The function of the capacitor *C* is to offer low impedance path for the high-frequency carrier signal and hence the carrier signal is prevented from flowing through the modulation transformer.

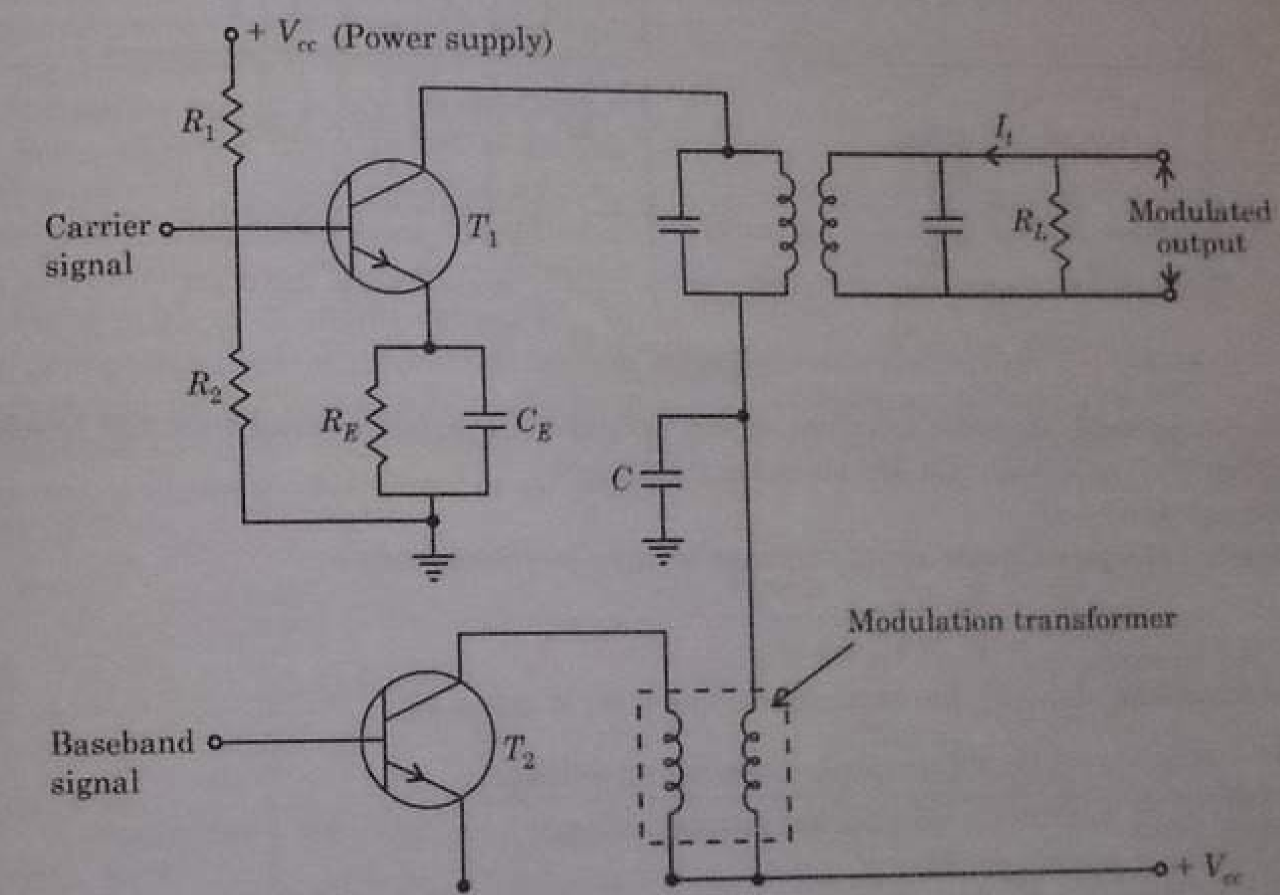


Fig. 3.10. *Illustration of collector modulation method.*

Operating Principle

It is a known fact that in a class-*C* amplifier, the magnitude of the output voltage is a definite fraction of or at the most equal to the supply voltage V_{cc}. In addition to this, a linear relationship exists between the output tank current I_t and the variable supply Voltage V_c (*i.e.*, here we have assumed that the supply voltage V_{cc} is a varying quantity and its varying value is denoted by V_c). This means that in class-C amplifier, the output voltage will be an exact replica of the input voltage waveform and the magnitude of the output voltage will be approximately equal to the carrier supply voltage V_{cc}.

Now if *R* is the resistance of the output tank circuit at resonance, then the magnitude of the output voltage is given as

$$RI_t \cong V_{cc}$$

Therefore, the unmodulated carrier is amplified by class-*C* modulated amplifier using transistor T_1 and its magnitude will remain constant at V_{cc} since there appears no voltage across the modulating transformer in the absence of baseband or modulating voltage.

But now if a baseband or modulating voltage $v_m = V_m \cos \omega_m t$ appears across the modulating transformer, this signal will be added to the carrier supply voltage V_{cc}. This results in a quite slow variation in carrier supply voltage V_{cc}. This type of slow variation in carrier supply voltage changes the magnitude of the carrier signal voltage at the output of the modulated class-C amplifier as shown in figure 3.11.

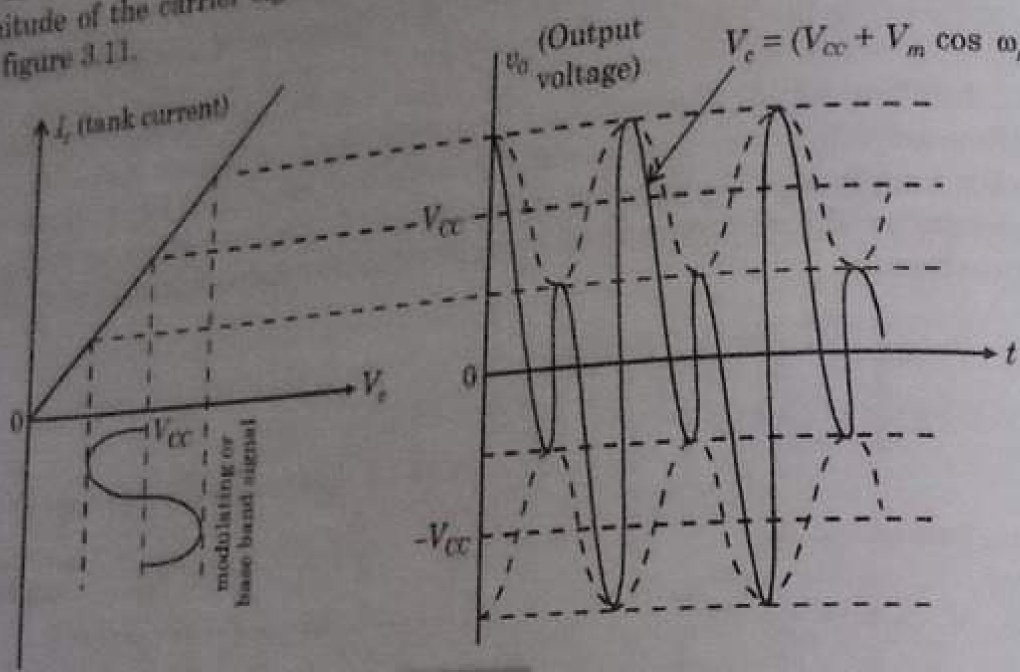


Fig. 3.11.

It may be observed that the envelope of the output voltage is identical with the baseband or modulating voltage and hence an AM signal is generated.

Mathematical Analysis

The slowly changing carrier supply voltage V_c may be expressed as

$$V_c = V_{cc} + v_m$$

or

$$V_c = V_{cc} + V_m \cos \omega_m t \quad ...(3.62)$$

But we know that for AM, the modulation index m_a is given by

$$m_a = \frac{\text{Maximum modulating voltage}}{\text{Maximum carrier voltage}}$$

In this case,

$$m_a = \frac{V_m}{V_{cc}} \quad ...(3.63)$$

Therefore,

$$V_c = V_{cc} + m_a V_{cc} \cos \omega_m t = V_{cc}(1 + m_a \cos \omega_m t) \quad ...(3.64)$$

Let the carrier voltage be

$$v_c = V_{cc} \cos \omega_c t$$

Then the modulated output voltage will be

$$v_0 = V_c \cos \omega_c t \quad ...(3.65)$$

Putting the value of V_c in above equation from equation (3.64), we get

$$v_0 = V_{cc}(1 + m_a \cos \omega_m t) \cos \omega_c t$$

which is the required expression for AM wave.

3.13. Demodulation of AM Waves

The process of extracting a modulating or baseband signal from the modulated signal is called **demodulation or detection**. In other words, demodulation or detection is the process by which the message is recovered from the modulated signal at receiver. The devices used for demodulation or detection are called demodulators or detectors. For amplitude modulation, detectors or demodulators are categorized as:

(i) Square-Law detectors

(ii) Envelope detectors

AM signal with large carrier are detected by using the envelope detector. The envelope detector uses the circuit which extracts the envelope of the AM wave. Infact, the envelope of the AM wave is the baseband or modulating signal. But a low-level amplitude modulated signal can only be detected by using square-law detectors in which a device operating in the non-linear region is used to detect the modulating signal.

3.13.1. Square-Law Detector

(U.P. Tech. Sem., Exam., 2003-04) (05 marks)

The square law detector circuit is used for detecting modulated signal of small magnitude (i.e. below 1 volt) so that the operating region may be restricted to the non-linear portion of the *V-I* characteristics of the device. Figure 3.12 shows the circuit of a square-law detector.* It may be observed that the circuit is very similar to the square-law modulator. The only difference lies in the filter circuit. In a square law modulator, the filter used is a bandpass filter whereas in a square law detector, a low-pass filter is used.

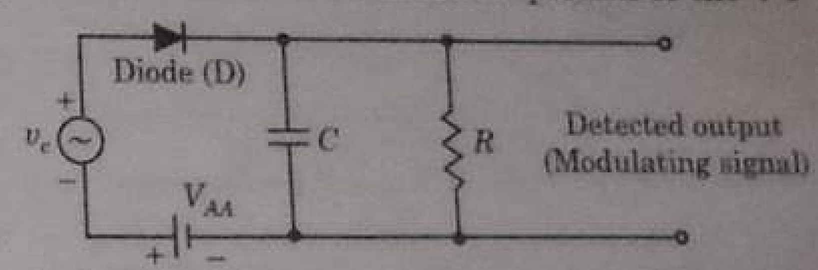


Fig. 3.12. *Basic circuit of square law diode detector.*

In the circuit, the dc supply voltage V_{AA} is used to get the fixed operating point in the non-linear portion of the diode *V-I* characteristic. Since, the operation is limited to the non-linear region of the diode characteristics, the lower half-portion of the modulated waveform is compressed. This produces envelope applied distortion. Due to this, the average value of the diode-current is no longer constant, rather it varies with time as shown in figure 3.13.

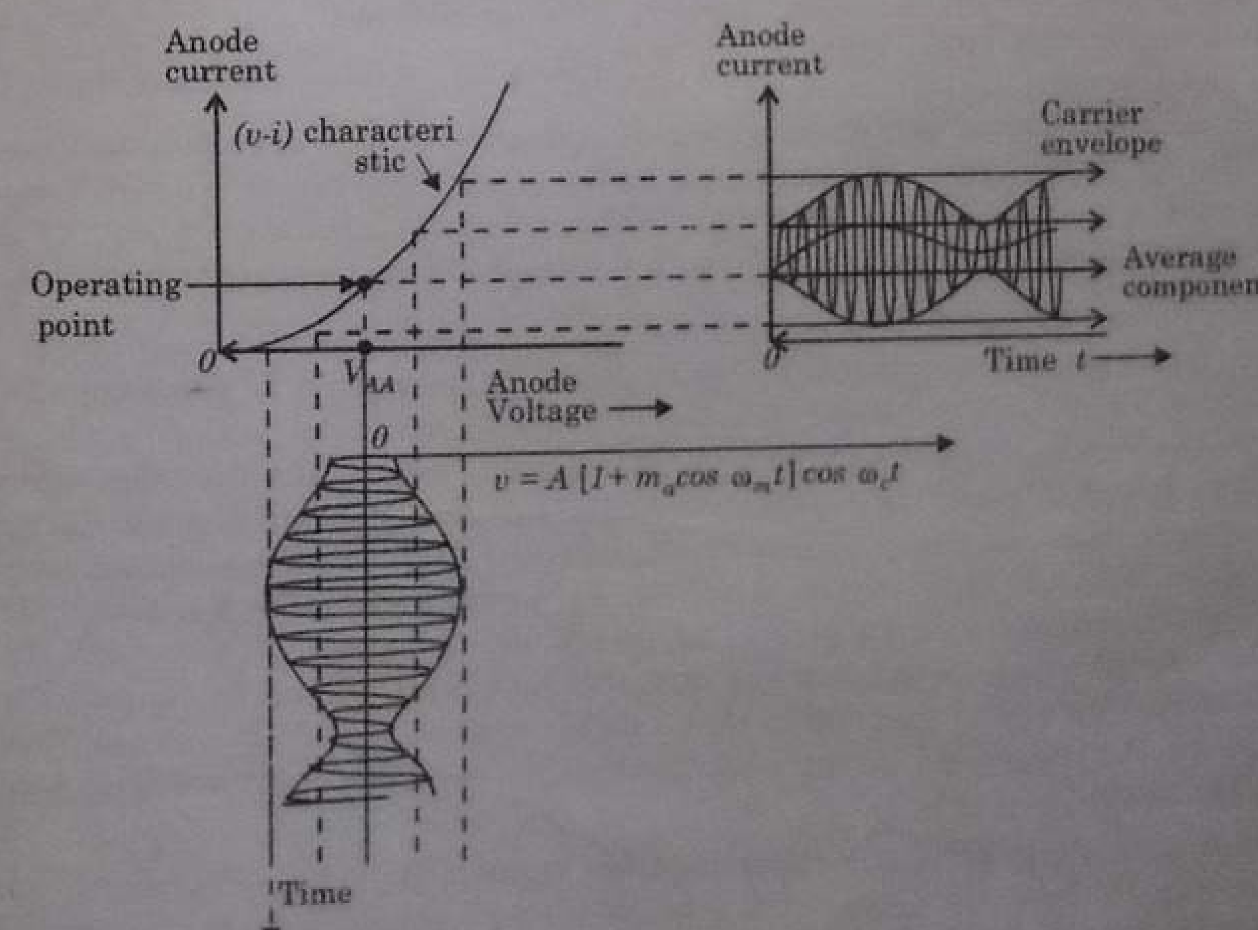


Fig. 3.13.

This distorted output diode current is expressed by the non-linear *v-i* relationship (i.e. square law) as

* Explain the working of square law demodulator for detection of AM wave.
(U.P. Tech. Sem. Exam., 2003-04) (05 marks)

$$i = av + bv^2 \quad \text{...(3.66)}$$

Here, v is the input modulated voltage.

We know that AM wave is expressed as

$$v = A(1 + m_a \cos \omega_m t) \cos \omega_c t \quad \text{...(3.67)}$$

Substituting, the value of v in equation (3.66), we get

$$i = a[A(1 + m_a \cos \omega_m t) \cos \omega_c t] + b[A(1 + m_a \cos \omega_m t) \cos \omega_c t]^2 \quad \text{...(3.68)}$$

Now, if above expression is expanded, then we may observe the presence of terms of frequencies like $2\omega_c$, $2(\omega_c \pm \omega_m)$, ω_m and $2\omega_m$ besides the input frequency terms.

Hence, this diode current i containing all these frequency terms is passed through a low-pass filter which allows to pass the frequencies below or upto modulating frequency ω_m and rejects the other higher frequency components. Therefore, the modulating or baseband signal with frequency ω_m is recovered form the input modulated signal.

3.13.2. Linear Diode or Envelope Detector

It is a known fact that a diode operating in a linear region of its *V-I* characteristics can extract the envelope of an AM wave. This type of detector is known as **envelope detector** or a **linear detector.** Envelope detector is most popular in commercial receiver circuits since it is very simple and is not expensive, also at the sametime, it gives satisfactory performance for the reception of broadcasting programmes. Figure 3.14 shows the circuit diagram of a linear diode detector or envelope detector. In the input portion of the circuit, the tuned transformer provides perfect tuning at the desired carrier frequency. *R-C* network is the time-constant network. If the magnitude of the modulated signal at the input of the detector is 1 volt or more, the operation takes place in the linear portion of the V-I characteristics of diode. Figure 3.15 shows the idealized linear characteristics of the diode along with the input voltage and output current waveforms.

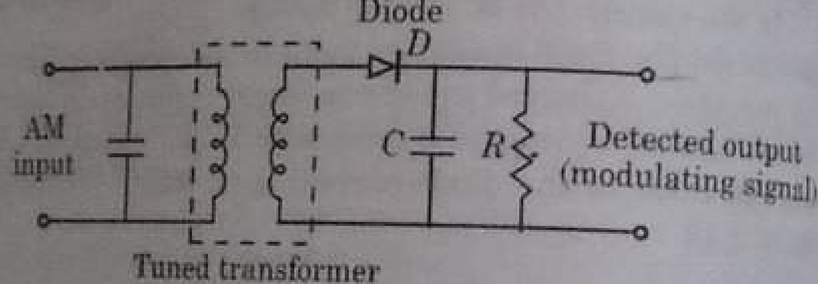


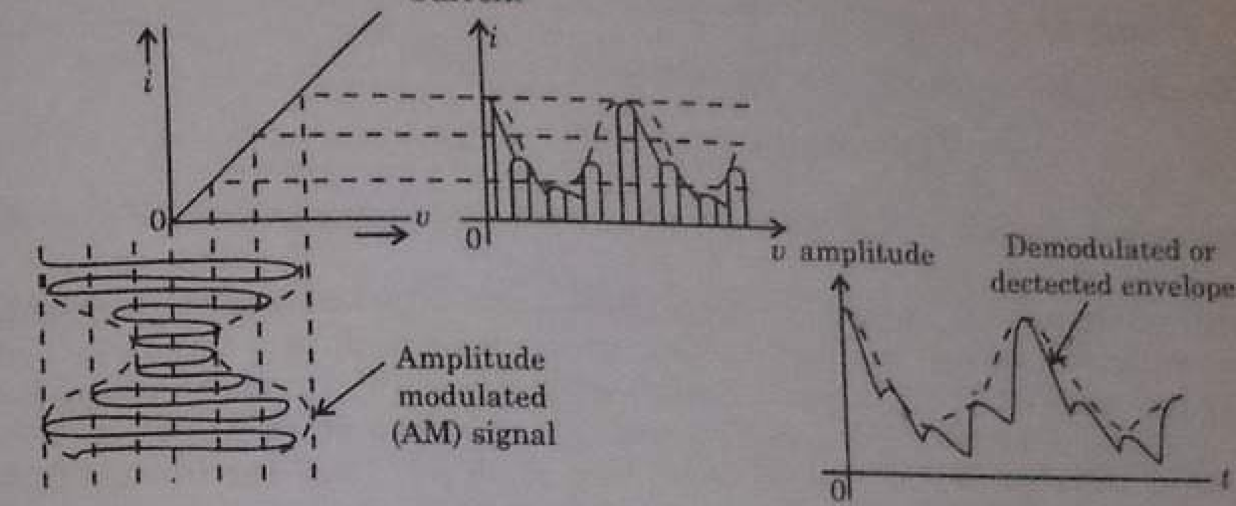
Fig. 3.14. *A linear diode detector.*

Operating Principle

First of all, let us assume that the capacitor is absent in the circuit. In this case, the detector circuit will work as a half-wave rectifier. Therefore, the output waveform would be a half-rectified modulated signal as shown in figure 3.15. Now let us consider that the capacitor is introduced in the circuit. For the positive half-cycle, the diode conducts and the capacitor is charged to the peak value of the carrier voltage. However, for a negative half-cycle, the diode is reverse-biased and does not conduct. This means that the input carrier voltage is disconnected from the *R-C* circuit. Therefore, the capacitor starts discharging through the resistance *R* with a time-constant $\tau = RC$. If the time-constant $\tau = RC$ is suitably chosen, the voltage across the capacitor *C* will not fall appreciably during the small period of negative half-cycle, and by that time the next positive cycle appears. This positive cycle again charges the capacitor *C* to the peak value of the carrier voltage and thus this process repeats again and again.

Hence, the output voltage across the capacitor *C* is a spiky modulating or baseband signal. This means that the voltage across the capacitor *C* is same as envelope of the modulated carrier signal, however, spikes are introduced because of charging and discharging of the capacitor *C*.

Figure 3.15(*b*) illustrates the resulting detected modulating or baseband signal. We can reduce discharges negligibly amount by keeping the time constant *RC* large so that the capacitor *C* discharges negligible small. However, the large value of $\tau = RC$ produces another problem known as diagonal clipping. Thus, we cannot increase the time-constant beyond a certain limit. Therefore, the time-constant is an important consideration for envelope detector.



(a) Characteristics of linear diode detector *(b) Detected output*

Fig. 3.15.

3.14. Double Sideband Suppressed Carrier (DSB-SC) System

The equation of AM wave in its simplest form *i.e.*, single-tone modulation, is expressed as

$$s(t) = A \cos w_c t + A.\frac{m_a}{2} \cos (w_c + w_m)t + A.\frac{n_a}{2} \cos (w_c - w_m)t \quad \text{...(3.68)}$$

From this equation, it is obvious that the carrier component in AM wave remains constant in amplitude and frequency. This means that the carrier of amplitude modulated wave does not convey any information. In power calculation of AM signal, it has been observed that for single-tone sinusoidal modulation, the ratio of the total power to the carrier power is $\left(1 + \frac{m_a^2}{2}\right)$, m_a being the modulation index. Thus for 100% modulation about 67% of the total power is required for transmitting the carrier which does not contain any information. Hence, if the carrier is suppressed, only the sidebands remain and in this way a saving of two-third power may be achieved at 100% modulation. This type of suppression of carrier does not affect the baseband signal in any way. The resulting signal obtained by suppressing the carrier from the modulated wave is called **Double sideband suppressed carrier (DSB-SC) system.**

Thus, in a double-sideband suppressed carrier modulation there is no carrier signal only sidebands are present.

We know that the frequency-shifting property of Fourier transform is given as

If $\quad x(t) \longleftrightarrow X(w)$

then $\quad e^{jw_c t} x(t) \longleftrightarrow X(w - w_c) \quad \text{...(3.69)}$

This property states the if a signal $x(t)$ is multiplied by $e^{jw_c t}$ in time-domain then its spectrum $X(w)$ in frequency-domain is shifted by an amount w_c.

Similarly,

$$e^{-jw_c t} x(t) \longleftrightarrow X(w + w_c) \quad \text{...(3.70)}$$

But, since $e^{jw_c t}$ is not a real function and cannot be generated practically, therefore, frequency shifting in practice is achieved by multiplying $x(t)$ by a sinusoid such as $\cos w_c t$.

Hence, $\quad x(t) \cos w_c t = x(t) \cdot \frac{1}{2}(e^{jw_c t} + e^{-jw_c t})$

$$x(t)\cos w_c t = \frac{1}{2}x(t)e^{jw_c t} + \frac{1}{2}x(t)e^{-jw_c t} \quad ...(3.71)$$

Using frequency-shifting property in equation (3.71), we get

$$x(t)\cos w_c t \longleftrightarrow \frac{1}{2}[X(w - w_c) + X(w + w_c)] \quad ...(3.72)$$

This means that the multiplication of a signal $x(t)$ by a sinusoid of frequency w_c shifts the spectrum $X(w)$ by $\pm w_c$.

Now if $x(t)$ is taken as modulating or baseband signal and $\cos w_c t$ is taken as carrier signal, then $x(t) \cos w_c t$ represents the modulated signal. Further, the Fourier transform of this modulated signal is given by the equation (3.72). This equation shows that the spectrum of modulated signal contains only shifted spectrum of signal and there is no carrier components. However, we know that the modulated signal, which contains no carrier but two sidebands is called **Double-Sideband Suppressed Carrier (DSB-SC) modulation.**

This means that the term $x(t)\cos w_c t$ represents a DSB-SC signal.

Therefore, a DSB-SC signal is obtained by simply multiplying modulating signal $x(t)$ with carrier signal $\cos w_c t$. This is achieved by a product modulator. The block diagram of a product modulator is shown in figure 3.16.

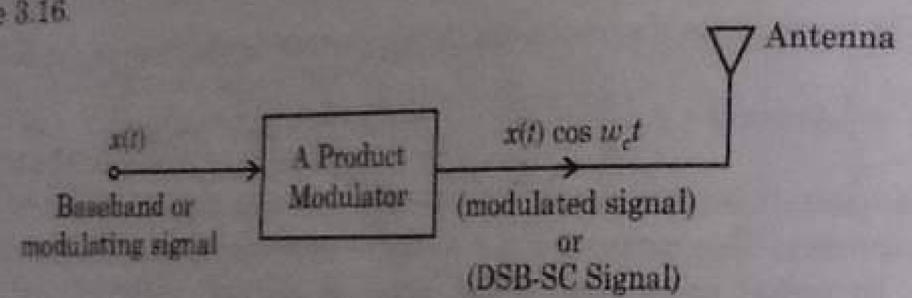


Fig. 3.16.

Figure 3.17 shows the modulating signal and its Fourier transform, carrier signal and its Fourier transform and DSB-SC signal and its frequency spectrum.

Few Points

(i) It is obvious from the figure 3.17, that the DSB-SC signal exhibits phase-reversal at zero crossings, *i.e.*, whenever the baseband signal $x(t)$ crosses zero. Because of this, the envelope of a DSB-SC modulated signal is different from the message signal. This is unlike the case of an AM wave.

(ii) From figure 3.17, it is also clear that the impulses at $\pm w_c$ are missing which means that the carrier term is suppressed in the spectrum and only two sideband terms, USB and LSB are left. Therefore, it is called double sideband suppressed carrier (DSB-SC) system.

(iii) In figure 3.17, considering only positive side the upperside band frequency is $w_c + w_m$ whereas the lowerside band frequency is $w_c - w_m$. The difference of these two is equal to the **transmission bandwidth of a DSB-SC signal,** *i.e.*,

Bandwidth $B = (w_c + w_m) - (w_c - w_m) = 2\,w_m$

It is obvious that the bandwidth of DSB-SC modulation is same as that of general AM wave.

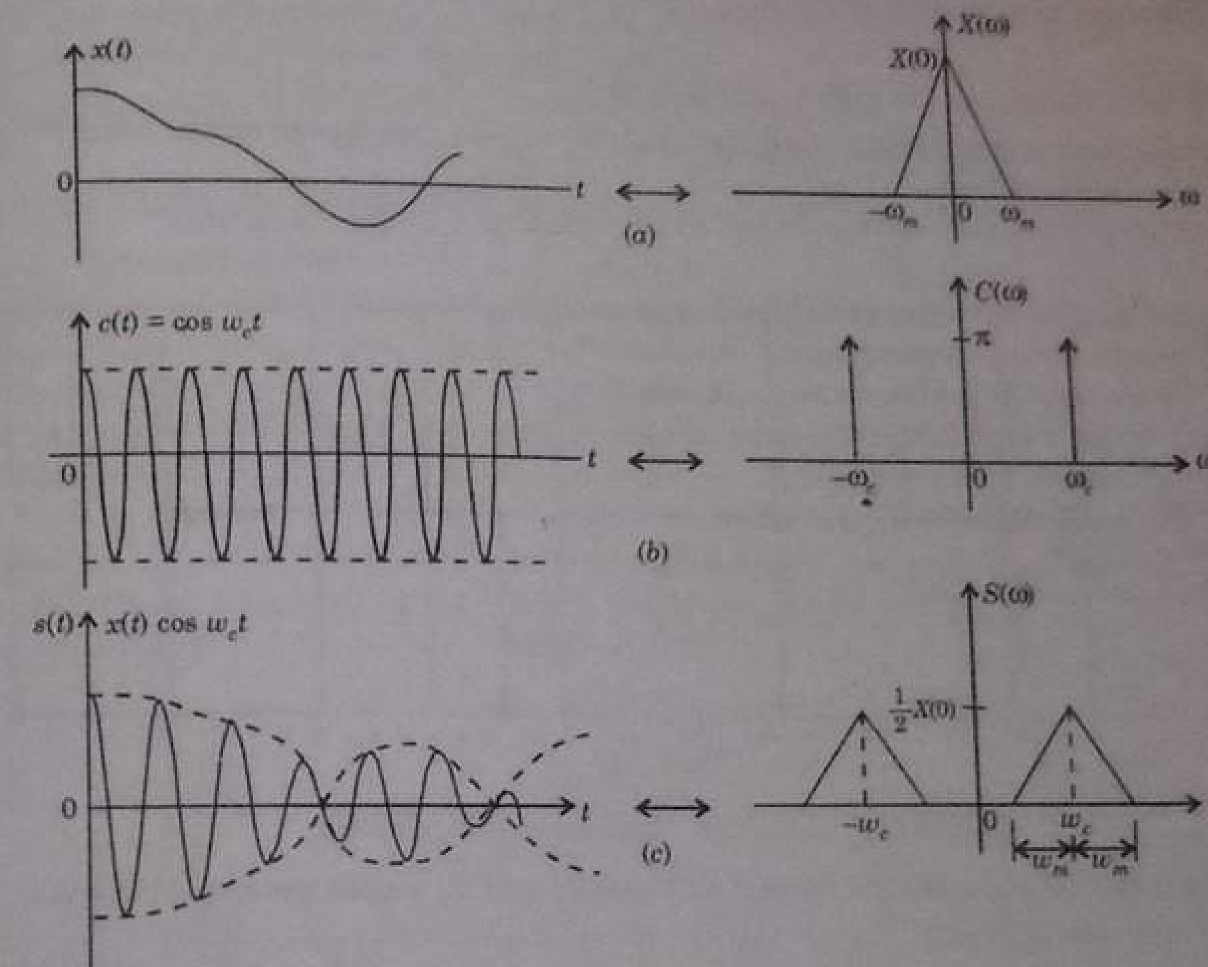


Fig. 3.17.

3.15. Generation of DSB-SC Signal

The expression for DSB-SC signal is given as

$$s(t) = x(t)\cos w_c t$$

where $x(t)$ = Baseband signal

$\cos w_c t$ = Carrier signal

From this expression, we see that a DSB-SC signal is basically the product of the modulating or baseband signal and the carrier signal. Unfortunately, a single electronic device cannot generate a DSB-SC signal. As discussed earlier, a circuit to achieve the generation of a DSB-SC signal is called a product modulator. In this section, we shall discuss two types of product modulator, namely the **Balanced Modulator** and the **Ring Modulator.**

3.15.1. The Balanced Modulator

We know that a non-linear resistance or a non-linear device may be used to produce Amplitude Modulation, *i.e.*, one carrier and two sidebands. However, a DSB-SC signal contains only two sidebands. Thus, if two non-linear devices such as diodes, transistors etc. are connected in a balanced mode so as to suppress the carriers of each other, then only sidebands are left, *i.e.*, a DSB-SC signal is generated.

Therefore, a balanced modulator may be defined as a circuit in which two non-linear devices are connected in a balanced mode to produce a DSB-SC signal.

In the following section, we shall discuss a balanced modulator circuit using diodes.

Figure 3.18 shows a balanced modulator circuit using two diodes. A modulating signal $x(t)$ is applied to the two diodes through a centre-tapped transformer with the carrier signal $\cos w_c t$.

A non-linear v-i relationship is given as

$$i = av + bv^2 \quad \text{...(3.73)}$$

Here, we have neglected the higher power terms.

In above expression, v is the input voltage applied across a non-linear device and i is the current through the non-linear device.

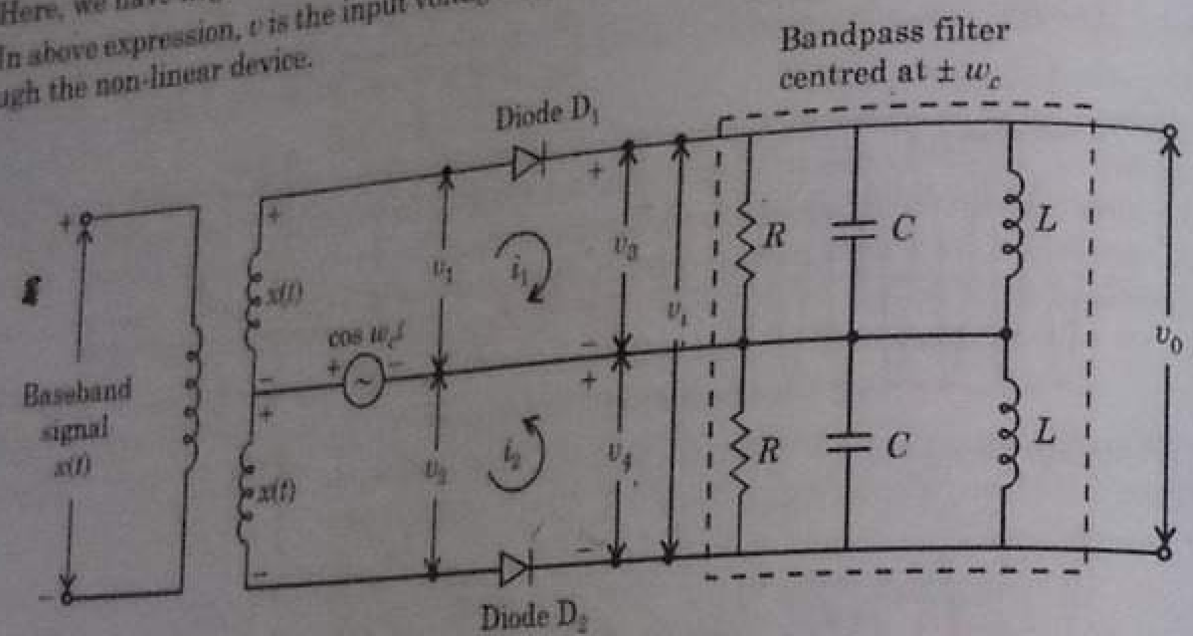


Fig. 3.18. *Balanced Modulator using Diodes.*

From figure 3.18, we write the two input voltages v_1 and v_2 across the two diodes as

$$v_1 = \cos w_c t + x(t) \quad \text{...(3.74)}$$

$$v_2 = \cos w_c t - x(t) \quad \text{...(3.75)}$$

For diode D_1, the nonlinear v-i relationship becomes

$$i_1 = av_1 + bv_1^2 \quad \text{...(3.76)}$$

Similarly, for diode D_2, the nonlinear v-i relationship becomes

$$i_2 = av_2 + bv_2^2 \quad \text{...(3.77)}$$

In the expression of current i_1, substituting the value of v_1, we get

$$i_1 = a\,[\cos w_c t + x(t)] + b\,[\cos w_c t + x(t)]^2$$

or $$i_1 = a \cos w_c t + ax(t) + b\,[\cos^2 w_c t + x^2(t) + 2x(t) \cos w_c t]$$

or $$i_1 = a \cos w_c t + ax(t) + b \cos^2 w_c t + bx^2(t) + 2bx(t) \cos w_c t \quad \text{...(3.78)}$$

Similarly, in the expression of current i_2, substituting the value of v_2, we get

$$i_2 = a\,[\cos w_c t - x(t)] + b[\cos w_c t - x(t)]^2$$

or $$i_2 = a \cos w_c t - ax(t) + b\cos^2 w_c t + bx^2(t) - 2bx(t) \cos w_c t \quad \text{...(3.79)}$$

Due to currents i_1 and i_2, the net voltage v_i at the input of bandpass filter is expressed as

$$v_i = v_3 - v_4 \quad \text{...(3.80)}$$

But from figure, we have

$$v_3 = i_1 R$$

and $$v_4 = i_2 R$$

Therefore, $$v_i = i_1 R - i_2 R$$

or $$v_i = R(i_1 - i_2)$$

In above equations, substituting the values of i_1 and i_2 from equations (3.78) and (3.79), we get

$$v_i = R[2ax(t) + 4bx(t) \cos w_c t]$$

or $$v_i = 2R[ax(t) + 2bx(t) \cos w_c t]$$

This voltage v_i is the input to the bandpass filter centered around ...

A bandpass filter is that type of filter which allows to pass a band of frequencies. Here, since the bandpass filter is centred around $\pm w_c$, it will pass a narrowband of frequencies centred at $\pm w_c$ with a small bandwidth of $2\,w_m$ to preserve the sidebands.

Therefore, the output of BPF centred around $\pm w_c$ is given by

$$v_0 = 4b\,Rx(t) \cos w_c t = K\,x(t) \cos w_c t$$

which is the expression for a DSB-SC signal

3.15.2. Ring Modulator

Ring modulator is another product modulator, which is used to generate DSB-SC signal.* Figure 3.19 shows the circuit diagram of a ring modulator. In a ring modulator circuit, four diodes are connected in the form of a ring in which all four diodes point in the same manner. All the four diodes in ring are controlled by a square wave carrier signal $c(t)$ of frequency f_c applied through a centre-tapped transformer.

In case, when diodes are ideal and transformer are perfectly balanced, the two outer diodes are switched on if the carrier signal is positive whereas the two inner diodes are switched off and thus presenting very high impedance as shown in figure 3.20(a). Under this condition, the modulator multiplies the modulating signal $x(t)$ by +1.

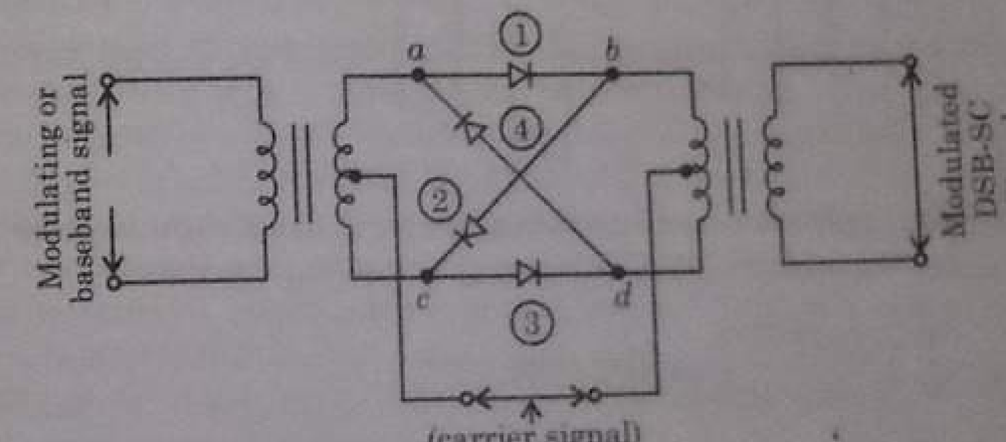


Fig. 3.19. *Circuit diagram of a ring modulator.*

Now, in case when carrier signal is negative, the situation becomes reversed as shown in figure 3.20(b). In this case, the modulator multiplies the modulating signal by −1.

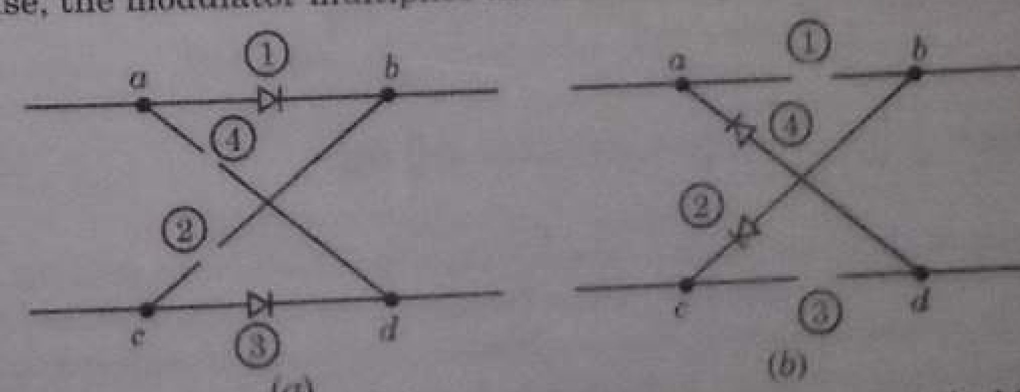


Fig. 3.20. (a) *Illustration of condition when the diodes (1) and (3) are switched ON and diodes (2) and (4) are switched OFF.*
(b) *Illustration of condition when the diodes (2) and (4) are switched ON and diodes (1) and (3) switched OFF.*

Hence, the ring modulator is a product modulator for a square wave carrier and modulating signal.

Figure 3.21 illustrates the modulating signal (assuming a sinusoidal signal), square wave carrier signal and the modulated (DSB-SC) signal.

* What is a DSB-SC modulator? Explain the working principle of a ring modulator. *(U.P. Tech Sem. Exam., 2003-04) (05 marks)*

The square wave carrier may be represented in Fourier series as under:

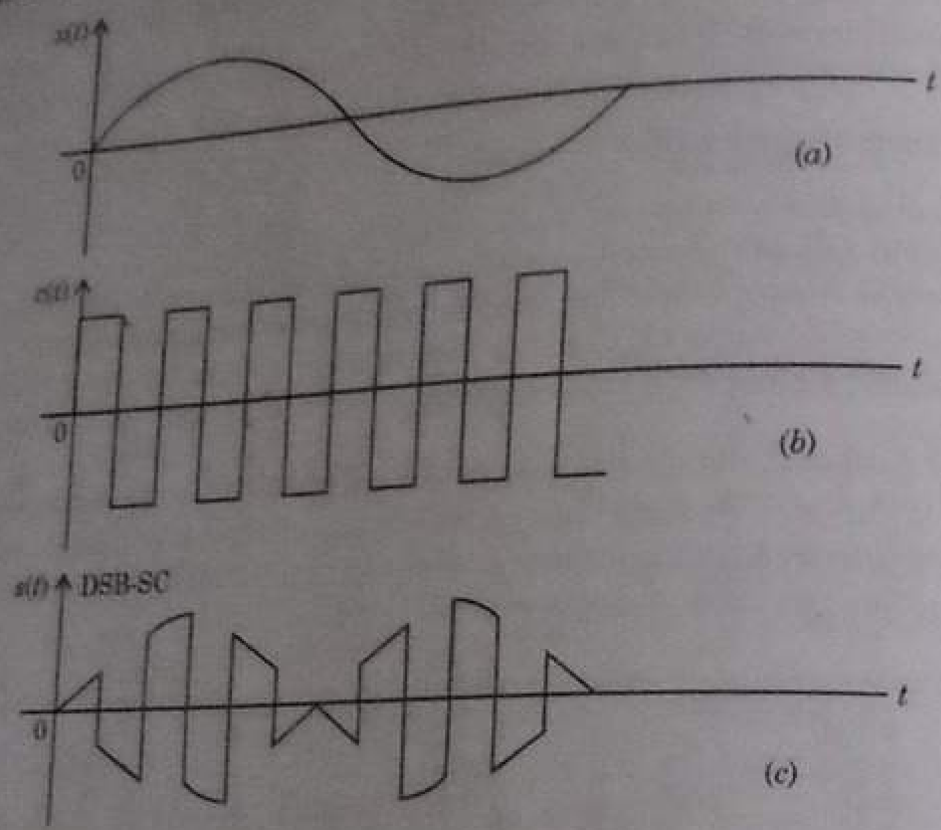


Fig. 3.21. *(a) Sinusoidal modulating wave (b) Square wave carrier (c) Modulated wave (DSB-SC) output of the Ring Modulator.*

$$c(t) = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{(2n-1)}\{\cos[2\pi f_c t\,(2n-1)]\}$$

We have $s(t) = x(t)\,c(t)$

Substituting the value of $c(t)$, we have

or $$s(t) = x(t)\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{(2n-1)}\{\cos[2\pi f_c t\,(2n-1)]\}$$

or $$s(t) = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{(2n-1)}\{\cos[2\pi f_c t(2n-1)]\}x(t)$$

hus, with the help of above mathematical analysis, it may be verified that the output from ring lator does not have any component at carrier frequency. Hence, the modulated output does not ny component at carrier frequency. Hence, the modulated output contains only product terms. ring modulator is also known as a **double-balanced modulator** since it is balanced with t to the baseband signal as well as the square wave carrier signal.

e frequency spectrum of the ring modulator output contains sidebands around each of the odd nics of the square wave carrier signal as illustrated in figure 3.22. We have assumed that the ting or baseband signal is bandlimited to $-f_m \le f \le f_m$. The desired sideband around the frequency f_c may be selected by using a bandpass filter (BPF) having center frequency w_c and dth $2f_m$.

m figure 3.22, it may be observed that to avoid overlapping of sidebands we must have

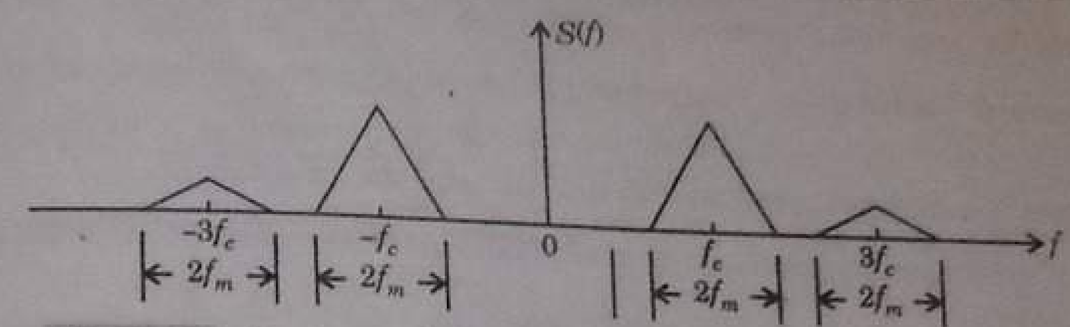


Fig. 3.22. *Spectrum of the DSB-SC output of the ring modulator.*

3.16. Demodulation of DSB-SC Signals

The DSB-SC signal may be demodulated by following two methods:

(i) Synchronous detection method

(ii) Using envelope detector after carrier reinsertion.

3.16.1. Synchronous Detection Method

We know that the DSB-SC system is used at the transmitter end to shift the modulating signal (having maximum frequency w_m) to a higher carrier frequency $\pm w_c$. Now, this modulated (DSB-SC) signal is transmitted from the transmitter and it reaches the receiver through a transmission medium. At the receiver end, the original modulating signal $x(t)$ is recovered from the modulated (DSB-SC) signal. This can be achieved by simply retranslating the baseband or modulating signal from a higher spectrum, centered at $\pm w_c$, to the original spectrum. This process of retranslation is called **demodulation** or **detection**. Hence, the original or baseband signal is recovered from the modulated signal by the detection process.

A method of DSB-SC detection is known as **synchronous detection.** Figure 3.23 shows the block diagram of synchronous detection method.

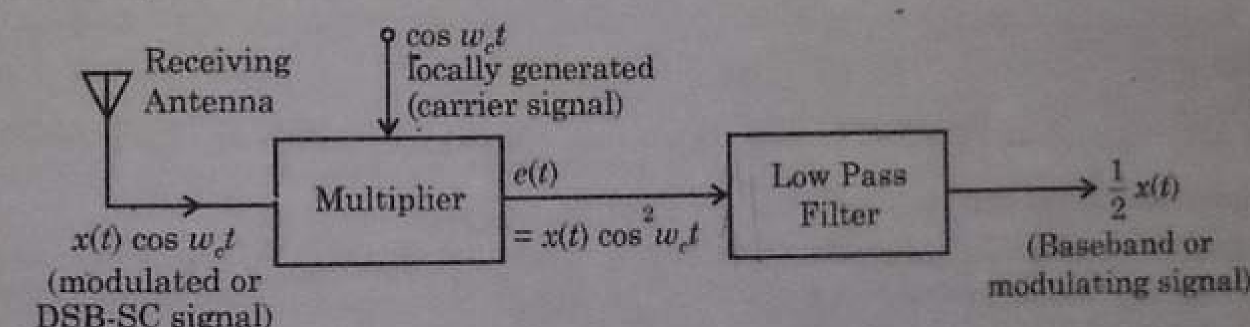


Fig. 3.23. *Synchronous detection method.*

Working Principle

In synchronous detection method, the received modulated or DSB-SC signal is first multiplied with a locally generated carrier signal $\cos \omega_c t$ and then passed through a low-pass filer (LPF). At the output of a low-pass filter (LPF), the original modulating signal is recovered.

Mathematically,

$$e(t) = \underbrace{x(t)\cos w_c t}_{\text{DSB-SC signal}} \cdot \underbrace{\cos w_c t}_{\text{Locally generated carrier signal}} \quad ...(3.81)$$

or $$e(t) = x(t)\cos^2 w_c t = \frac{1}{2}x(t)\left[2\cos^2 w_c t\right]$$

or $$e(t) = \frac{1}{2}x(t)[1+\cos 2w_c t] = \frac{1}{2}x(t) + \frac{1}{2}x(t)\cos 2w_c t \quad ...(3.82)$$

Now, it may be observed that when multiplied signal $e(t)$ is passed through a low-pass filter (LPF), then the term $\frac{1}{2}x(t)\cos 2w_c t$, centred at $\pm 2w_c$ is suppressed by low-pass filter and thus at the output of low-pass filter, the original modulating signal $\frac{1}{2}x(t)$ is obtained.

We may get the frequency-spectrum of multiplied signal $e(t)$ using Fourier Transform as

$$x(t)\cos^2 w_c t \longleftrightarrow \frac{1}{2}X(w) + \frac{1}{4}[X(w+2w_c) + X(w - 2w_c)] \quad ...(3.83)$$

Figure 3.24 shows this frequency spectrum

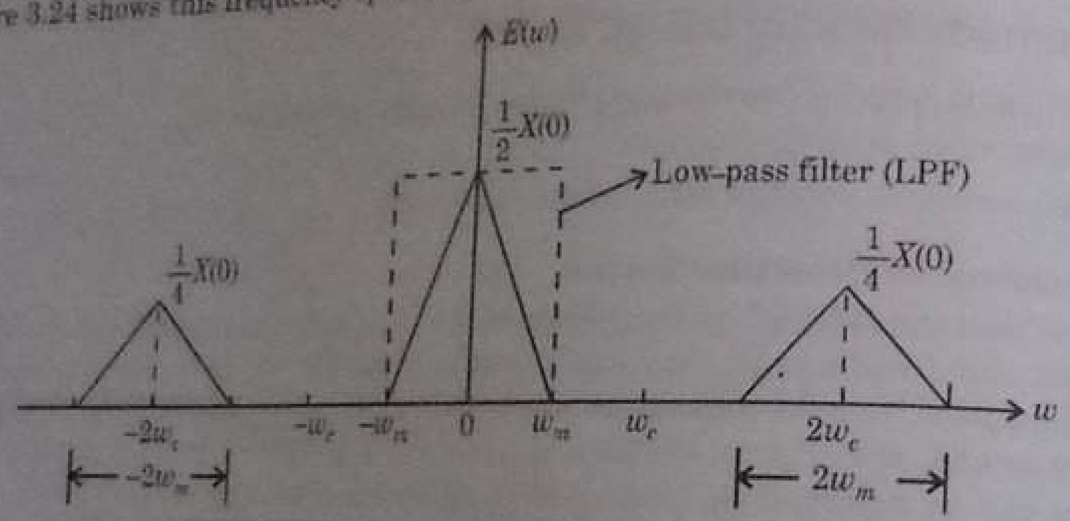


Fig. 3.24. *Frequency spectrum of $x(t)\cos^2 w_c t$.*

Note: *The spectrum reveals the fact that the original baseband or modulating signal (0 to w_m) is present alongwith another frequency spectrum centered around $\pm w_c$ when this signal i.e., $x(t)\cos 2w_c t$ is passed through a low pass filter (LPF) with cut-off frequency w_m, the original baseband signal or modulating signal appears at the output of low-pass filter. It may be noted $w_c >> w_m$ and $2w_c$ is still greater than w_m and thus is easily filtered out.*

Hence, the original modulating or message signal $x(t)$ is recovered from the DSB-SC signal.

Synchronous or Coherent Detection

We have observed in last article that the detection process for DSB-SC requires a local oscillator at the receiver end. The frequency and phase of the locally generated carrier signal and the carrier signal at the transmitter carrier must be identical. This means that the local oscillator signal must be exactly coherent or synchronized with the carrier signal at the transmitter, both in frequency and phase, otherwise the detected signal would get distorted. Therefore, this method of recovery is called **synchronous detection** or **coherent detection.** Thus, the demerits of the synchronous detection is that if requires an additional system at the receiver to ensure that the locally generated carrier is synchronized with the transmitter carrier making the receiver complex and costly.

3.16.2. Envelope Detection after Suitable Carrier Re-insertion

The other possible method of demodulating DSB-SC signal is by inserting a carrier generated at the receiver end with the help of a local oscillator. However, the phase and the frequency of the re-inserted carrier must be properly synchronized with those at the transmitter end in order to avoid distortion. We know that if we insert a sufficient carrier of same frequency and phase to DSB-SC signal, it converts DSB-SC signal into a conventional AM wave. Now, this AM wave is demodulated by an envelope detector. However, phase and frequency errors will result in similar type of distortion as obtained in coherent detection.

Let us consider that the received DSB-SC signal is expressed by

$$s(t) = c(t) \cdot x(t)$$

or $$s(t) = A\cos(2\pi f_c t)\, x(t) \quad ...(3.84)$$

Let us assume $A = 1$ in equation (3.84). Then equation (3.84) can be written as

$$s(t) = \cos(2\pi f_c t)\, x(t) \quad ...(3.85)$$

The inserted carrier at the receiver will be

$$c'(t) = A\cos(2\pi f_c t + \phi)$$

where ϕ = amount of phase discrepancy.

Then the resulting signal will be

$$r(t) = s(t) + \text{re-inserted carrier at the receiver}$$

or $$r(t) = s(t) + c'(t) \quad ...(3.86)$$

Substituting equations (3.84) and (3.85) in equation (3.86), we get

$$r(t) = s(t) + c'(t) = \cos(2\pi f_c t)x(t) + A\cos(2\pi f_c t + \phi)$$

or $$r(t) = x(t)\cos(2\pi f_c t) + A\cos(2\pi f_c t)\cos\phi - A\sin(2\pi f_c t)\sin\phi$$

or $$r(t) = [x(t) + A\cos\phi]\cos(2\pi f_c t) - (A\sin\phi)\sin(2\pi f_c t)$$

or $$r(t) = e(t)\cos[(2\pi f_c t) + \theta(t)] \quad ...(3.87)$$

where $$e(t) = \sqrt{[A + x(t)]^2 - 2A\,x(t)[1-\cos\phi]}$$

and $$\theta(t) = \tan^{-1}\left[\frac{A\sin\phi}{x(t) + A\cos\phi}\right]$$

Now, from the expression

$$r(t) = e(t)\{\cos[(2\pi f_c t) + \theta(t)]\}$$

it may be observed that $e(t)$ is the envelope of the resulting signal $r(t)$.

Also, if we take $\phi = 0$, then envelope will be given by

$$e(t) = A + x(t)$$

Hence, modulating signal $x(t)$ can be recovered from $r(t)$ using an envelope detector since the $r(t)$ is basically a conventional AM wave given by

$$r(t) = [A + x(t)]\cos 2\pi f_c t \quad ...(3.88)$$

This is however possible only when $[A + x(t)] > 0$ for all values of t.

It is possible only when the modulation index m is less than unity.

If $\phi \neq 0$, then the phase error exists between the two carriers. It is given as

$$e(t) = A\left[1 + \frac{2x(t)}{A}\cos\phi + \left\{\frac{x(t)}{A}\right\}^2\right]^{\frac{1}{2}} \quad ...(3.89)$$

If $A >> |x(t)|$, then, we have

$$e(t) \cong A + x(t)\cos\phi \quad ...(3.90)$$

The desired signal output will thus be $x(t)\cos\phi$. If $\phi = 0$ and there is a difference in frequency Δf between the two oscillators, then the envelope of the resulting signal $r(t)$ will be given by

$$e(t) = A + x(t)\cos[2\pi\Delta f t] \quad \text{for } A >> |x(t)| \quad ...(3.91)$$

The envelope can be detected by an envelope detector but here distortion will be identical to coherent detection.

3.17. Quadrature-Amplitude Modulation (QAM)

Definition

This modulation scheme is also called **Quadrature Carrier Multiplexing.** In fact, this modulation scheme enables two DSB-SC modulated signals to occupy the same transmission bandwidth and therefore it allows for the separation of the two message signals at the receiver output. It is therefore known as a **bandwidth-conservation scheme.**

Figure 3.25 shows a block diagram of the Quadrature Amplitude Modulation (QAM) system. The QAM transmitter has been shown in figure 3.26(*a*) and QAM receiver has been shown in figure 3.25(*b*). The QAM transmitter consists of two separate balanced modulators (BM) which are supplied with two carrier waves of the same frequency but differing in phase by 90°. The output of the two balanced modulators are added in the adder and transmitted. The transmitted signal is thus given by

$$s(t) = x_1(t)\, A \cos(2\pi f_c t) + x_2(t)\, A \sin(2\pi f_c t)$$

or

$$s(t) = A\, x_1(t) \cos(2\pi f_c t) + A\, x_2(t) \sin(2\pi f_c t) \quad ...(3.92)$$

where, $x_1(t)$ and $x_2(t)$ are the two different message signals applied to the product modulators. Both $x_1(t)$ and $x_2(t)$ are band-limited in the interval $-f_m \le f \le f_m$, then $s(t)$ will occupy a bandwidth of $2f_m$. This bandwidth $2f_m$ is centred at the carrier frequency f_c, where f_m is the bandwidth of message signal $x_1(t)$ or $x_2(t)$.

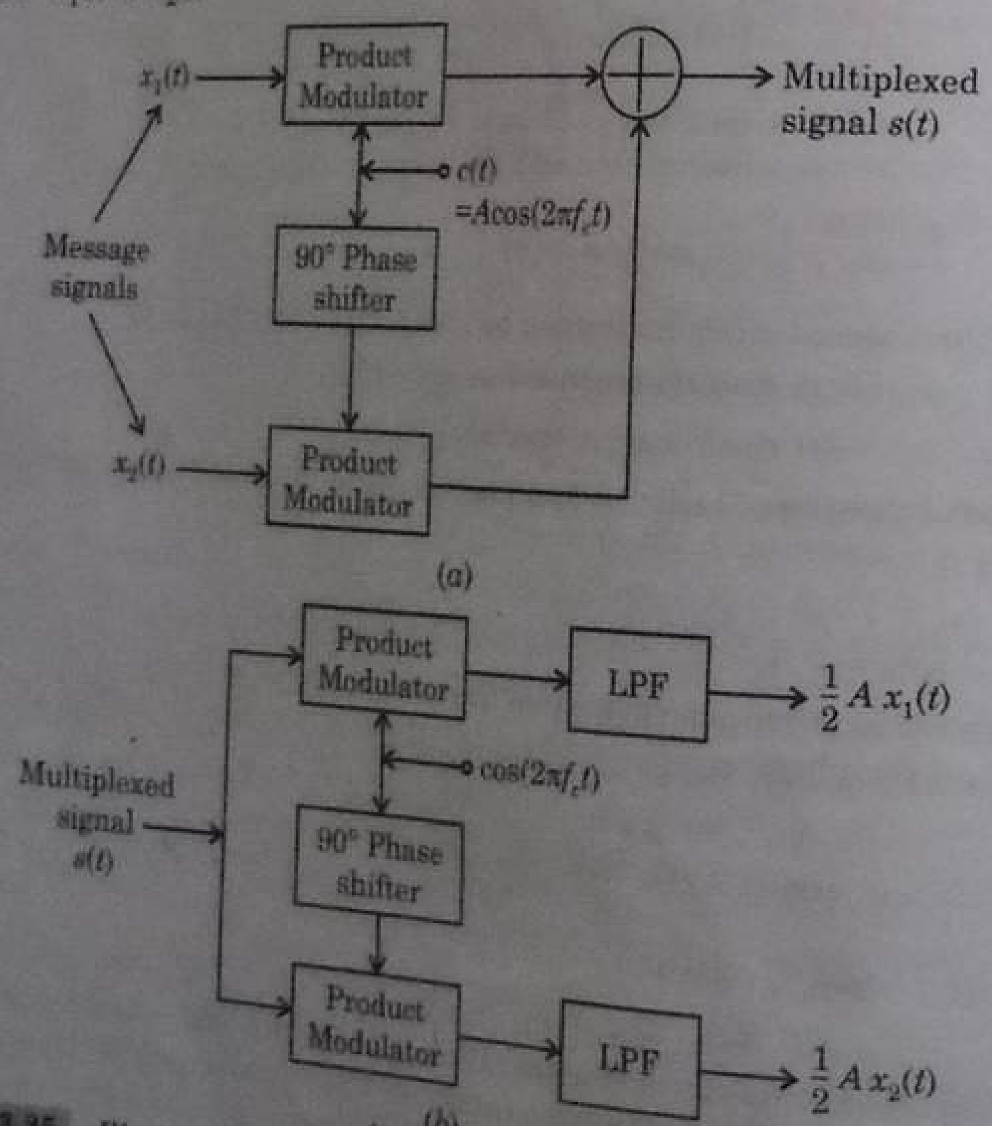


Fig. 3.25. *Illustration of Quadrature Amplitude Modulation (QAM) system (a) QAM transmitter (b) QAM receiver.*

Hence, the multiplexed signal consists of the in-phase component $A\, x_1(t)$ and the quadrature component $-A\, x_2(t)$.

The multiplexed signal $s(t)$ from QAM transmitter is applied simultaneously to two separate coherent detectors that are supplied with two local carriers of the same frequency, but differing in phase by 90°. The output of one detector is $1/2\, A\, x_1(t)$, whereas the output of the second detector is $1/2\, A\, x_2(t)$. For satisfactory operation of the coherent detector, it is essential to maintain coherent phase and frequency relationship between the oscillators used in the QAM transmitter and receiver parts of the system. The Quadrature Amplitude Modulation (QAM) finds application in colour television (CTV).

3.18. Effect of Phase and Frequency Errors in Synchronous Detection

As discussed earlier, the frequency and phase of the local oscillator signal in coherent detection method at the receiver end, must be identical to the transmitted carrier signal. Any kind of discrepancy in frequency or phase produces a distortion in the detected output at the receiver end.

In this subsection, let us examine the nature of distortion produced by a frequency or phase discrepancy.

For this, let us assume that the modulated signal reaching the receiver is denoted as $x(t) \cos \omega_c t$. Considering, a locally generated signal with frequency and phase error equal to $\Delta\omega$ and ϕ respectively, the product of the two signals in the synchronous detector provides

$$e_d(t) = x(t) \cos \omega_c t \, . \cos [(\omega_c + \Delta\omega)\, t + \phi]$$

or

$$e_d(t) = \frac{1}{2} x(t) \{\cos [(\Delta\omega)\, t + \phi] + \cos [(2\omega_c t + \Delta\omega) t + \phi]\} \quad ...(3.93)$$

Now, when this signal is allowed to pass through a low-pass filter (LPF) having a cut-off frequency ω_m, the terms centered around $\pm 2\omega_c$ are filtered out and the filter output is given as

$$e_0(t) = \frac{1}{2} x(t) \cos [(\Delta\omega) t + \phi] \quad ...(3.94)$$

The baseband signal $x(t)$ is multiplied by a slow-time varying function $\cos[(\Delta\omega)\, t + \phi]$ which distorts the message signal $x(t)$.

Now, let us consider the following special cases:

(*i*) When the frequency error $\Delta\omega$ and phase error ϕ are both zero, then last equation (3.94) gives

$$e_0(t) = \frac{1}{2} x(t)$$

This means that there is no distortion in the detected output signal.

(*ii*) When there is only the phase error *i.e.*,

$$\Delta\omega = 0 \quad \text{but} \quad \phi \neq 0$$

In this case, equation (3.94) gives

$$e_0(t) = \frac{1}{2} x(t) \cos\phi$$

This shows that the output signal is multiplied by $\cos\phi$. When ϕ is time-independent, there is no distortion, rather, there is only attenuation. The output will be maximum when $\phi = 0$ and minimum when $\phi = 90°$. But, in general ϕ randomly varies with respect to the time due to random variation of propagation media (*i.e.*, iono sphere). This causes undesirable distortion in the detected output.

It may be noted that the detected output is zero when $\phi = 90°$. This is known as a **quadrature null effect** since the signal is zero when the local carrier is in-phase quadrature with the transmitted carrier signal.

(*iii*) When there is only the frequency error, *i.e.*,

$$\Delta\omega \neq 0 \quad \text{and} \quad \phi = 0$$

In this case, equation (3.94) gives

$$e_0(t) = \frac{1}{2} x(t) \cos(\Delta\omega) t$$

Here, the multiplying factor $\cos(\Delta\omega)t$ is time-dependent and produces distortion in the detected output signal. The error $\Delta\omega$ is usually small and thus a message signal $x(t)$ is multiplied by a slow varying sinusoidal signal. This is a more serious distortion. Therefore, a frequency error must be avoided.

(*iv*) When both errors are non-zero, *i.e.*,

$$\Delta\omega \neq 0 \quad \text{and} \quad \phi \neq 0$$

In this case equation (3.94) itself provides the detected output signal. Also, in this case, the constant phase error provides attenuation and the frequency error produces distortion in the detected output signal.

Hence, we get an attenuated and distorted output signal at the receiver end.

3.19. Carrier Acquisition in DSB-SC System or Synchronization Techniques in DSB-SC System

As discussed in the last subsection, the phase and the frequency of the locally generated carrier signal in synchronous detector is very critical. Precision phase and frequency control of the local carrier requires an expensive and a complex circuitry at the receiver end. Some important synchronization techniques are given as under:

(*i*) Pilot Carrier

A small amount of carrier signal known as **pilot carrier** is transmitted alongwith the modulated signal from the transmitter. This small amount of carrier signal is called **pilot carrier**. This pilot carrier, separated at the receiver by an appropriate filter, is amplified, and is used to phase lock the locally generated carrier signal at the receiver. The phase locking provides synchronization. This system, where a weak carrier is transmitted alongwith the DSB-SC signal is also referred to as **partially suppressed carrier system** as the carrier is not totally suppressed. The process in which a large carrier is transmitted alongwith DSB-SC signal is known as amplitude modulation. This has been already discussed. The large carrier simplifies the reception system. The DSB-SC with partially suppressed carrier is equivalent to an over modulated AM signal.

(*ii*) Costa's Receiver

This system used for synchronous detection of DSB-SC signal has been shown in figure 3.26.

This system has two synchronous detectors–one detector is fed with a locally generated carrier ignal which is in phase with the transmitted carrier signal. This detector circuit is called inphase oherent detector or I-channel. The other synchronous detector employs a local carrier which is in hase quadrature with the transmitted carrier signal and is called Quadrature phase coherent etector or Q-channel. On combining, the two detectors constitute a negative feedback system hich synchronizes the local carrier signal with the transmitted carrier signal.

Figure 3.26 shows a costa's receiver.

perating Principle

To start with, let us assume that the local carrier signal is synchronized with the transmitted rier signal and $\phi = 0$. As shown in figure 3.26, the output of the I-channel is the desired modulating nal (since $\cos\phi = 1$), but the output of the Q-channel is zero (since $\sin\phi = 0$) because of the drature null effect. Now, assuming that the local oscillator frequency drifts slightly *i.e.*, ϕ is a small non-zero quantity, I-channel output will be almost unchanged, but Q-channel output

now is not a zero, rather some signal would appear at its output and is proportional to $\sin\phi$. Thus, the output of the Q-channel,

(*i*) is proportional to ϕ (since $\sin\phi = \phi$ for small ϕ)

(*ii*) would have a polarity same as the I-channel for one direction of phase shift in local oscillator, whereas, the polarity would be opposite to I-channel for the other direction of phase shift.

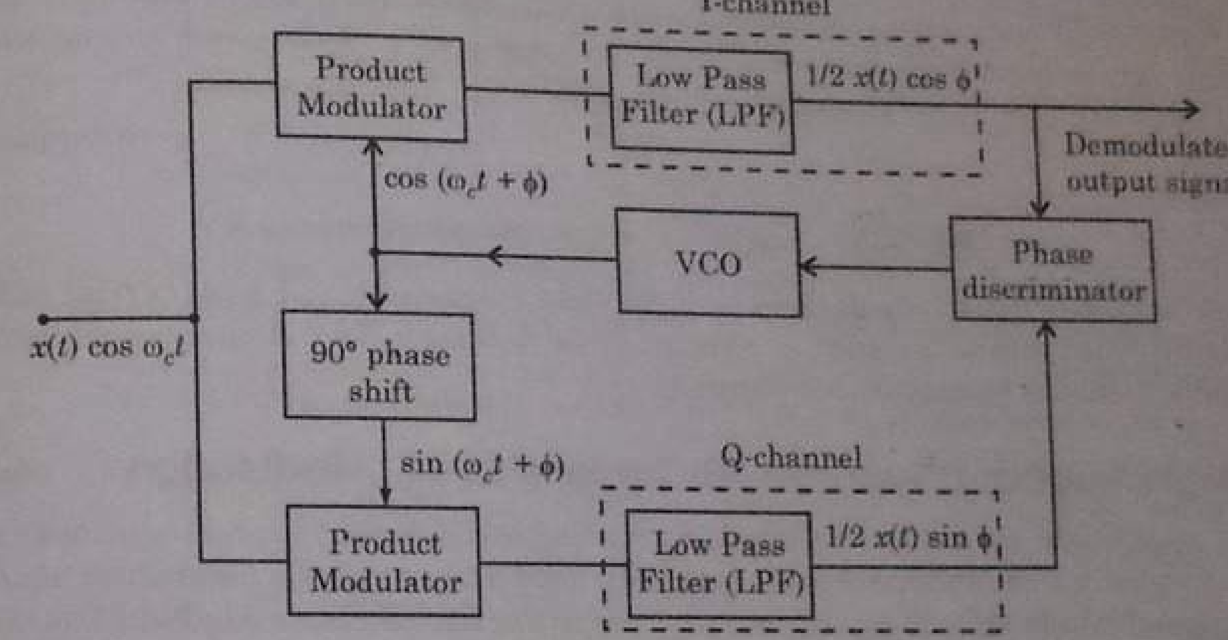


Fig. 3.26. *A Costa's receiver.*

The phase discriminator provides a d.c. control signal which may be used to correct local oscillator phase error. The local oscillator is a voltage controlled oscillator (VCO). Its frequency may be adjusted by an error control d.c. signal.

Limitation

The costa's receiver cases phase control when there is no modulation *i.e.*, $x(t) = 0$. The phase control reestablishes itself on the reappearance of modulation. However, the reestablishment is so fast that distortion is not perceptible in case of voice communication.

(*iii*) Squaring Loop

In this method, the received signal is squared by a squaring circuit as shown in figure 3.27.

The output of the squarer will be given as

$$[A.\, x(t) \cos \omega_c t]^2 = A^2 x^2(t) \cos^2 \omega_c t$$

For simplicity let us assume that $x(t)$ is a single tone sinusoid denoted as $\cos \omega_m t$ *i.e.*,

$$x(t) = \cos \omega_m t$$

then the output of the squarer becomes

$$[A \cos \omega_c t.\, \cos \omega_m t]^2 = A^2 \cos^2 \omega_m t \cos^2 \omega_c t$$

$$= \frac{A^2}{4} (1 + \cos 2\omega_m)(1 + \cos 2\omega_c t)$$

$$= \frac{A^2}{4} [1 + \cos 2\omega_m t + \cos 2\omega_c t + \cos 2\omega_c t \cos 2\omega_m t] \quad ...(3.95)$$

The term $\cos 2\omega_c t$ can be obtained by using a narrowband filter centred at $\pm 2\omega_c$. This frequency $\pm 2\omega_c$ is kept constant by tracking through a phase locked loop (PLL). The PLL uses negative feedback technique to provide a constant frequency signal, $\cos 2\omega_c t$.

Any drift in frequency is corrected by an error signals $e(t)$, generated at the output of the lowpass filter of PLL as depicted in figure 3.27. PLL will be discussed later on.

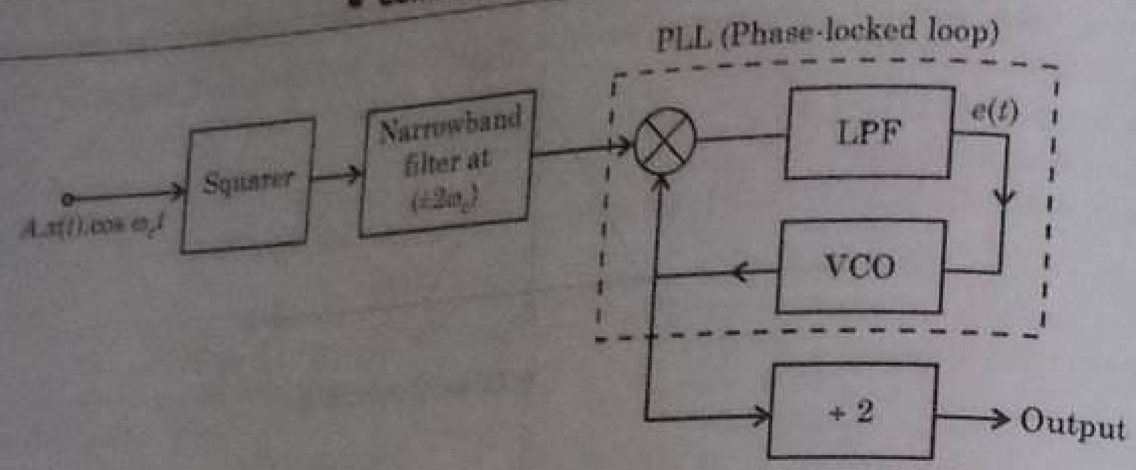


Fig. 3.27. *A squaring circuit for synchronization.*

The VCO output is frequency divided by 2, to yield a synchronized local carrier of frequency ω_c. This local carrier signal is used in synchronous detector. The frequency division can be accomplished by using a bistable multivibrator.

3.20. Single Sideband Suppressed-Carrier (SSB-SC) Modulation *(Important)*

As discussed earlier that amplitude modulation and double-sideband suppressed-carrier (DSB-SC) modulation are wasteful of bandwidth since they both need a transmission bandwidth equal to twice the message signal bandwidth. In either case one half of the transmission bandwidth is occupied by the upper sideband of the modulated signal whereas the other half is occupied by the lower sideband. However, the lower and upper sidebands are uniquely related to each other by virtue of their symmetry about the carrier frequency, *i.e.*, if amplitude and phase spectra of either sideband is given, we can uniquely determine the other. This means that as far as the transmission of information is concerned, only one sideband is necessary. Thus, if the carrier and one of the two sidebands are suppressed at the transmitter, no information is lost. Modulation of this type which provides a single sideband with suppressed carrier is known as single sideband suppressed carrier (SSB-SC) system. Thus, SSB-SC system reduces the transmission bandwidth by half. This means that in a given frequency band we can accommodate twice the number of channels by using a single sideband in place of both the sidebands.

Figure 3.28 further illustrated the concept of single sideband modulation with the help of different frequency spectrums.

Figure 3.28(*a*) shows the frequency spectrum of modulating or baseband signal. It contains ω_m as the maximum frequency component.

Figure 3.28(*b*) shows the frequency spectrum of a DSB-SC modulation which contains no carrier but two sidebands, *i.e.*, lower sideband and upper sideband.

Figure 3.28(*c*) shows the frequency spectrum of single sideband suppressed carrier modulation onsisting of upper sideband only, *i.e.*, lower sideband and carrier signal are suppressed.

Figure 3.28(*d*) shows the frequency spectrum of single sideband suppressed carrier modulation nsisting of lower sideband only, *i.e.*, upper sideband and carrier signal are suppressed.

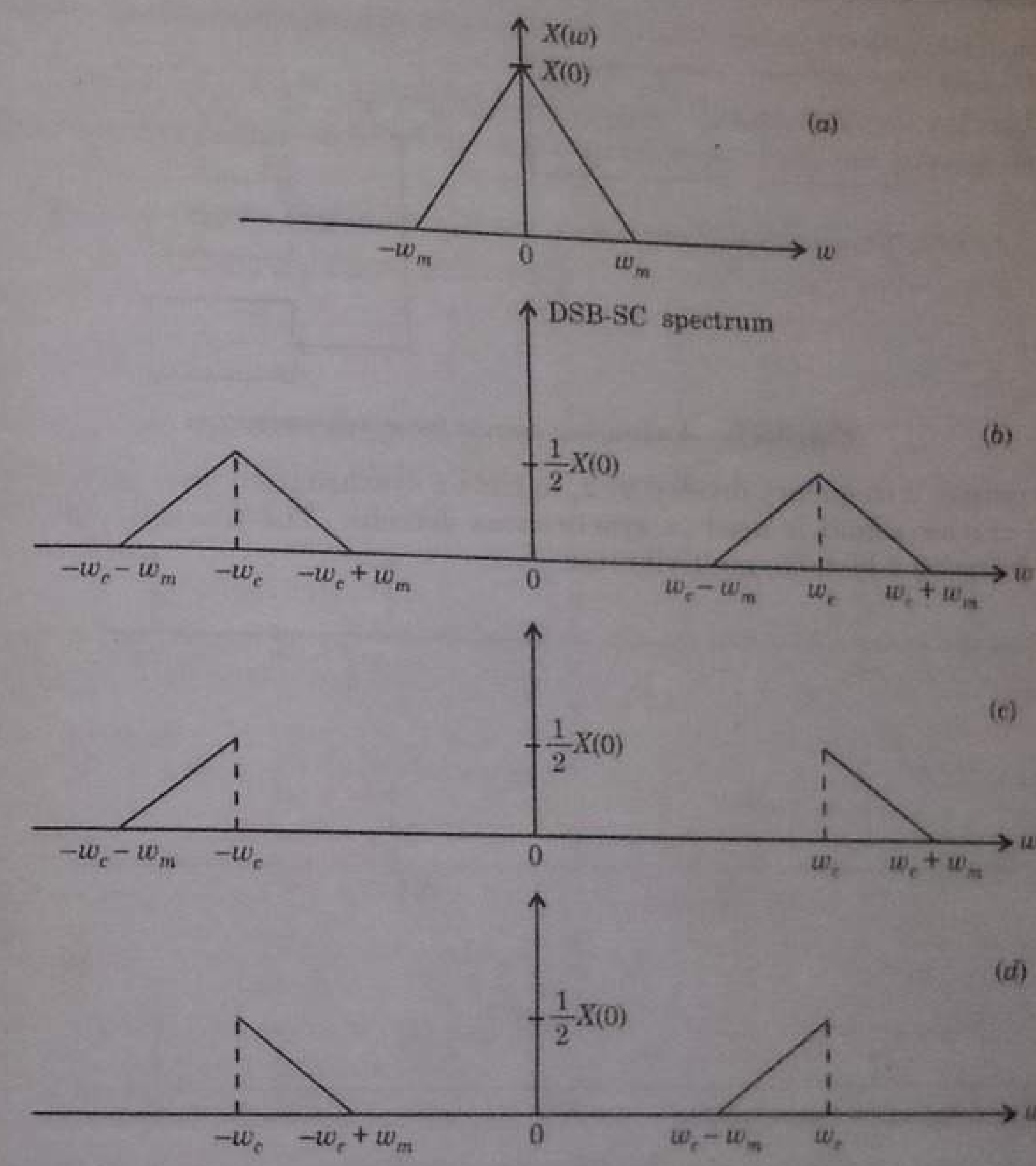


Fig. 3.28. *(a) Spectrum of baseband signal*
(b) Spectrum of DSB-SC wave
(c) Spectrum of SSB-SC wave with the upper sideband transmitted
(d) Spectrum of SSB-SC wave with the lower sideband transmitted.

3.20.1. Time-Domain Description of the SSB-SC Wave

We can derive an expression to represent the SSB wave in time-domain considering a special case of single-tone modulating signal. We can derive a general expression for SSB wave using the concept of a pre-envelope

SSB-SC wave with single-tone modulating signal

Let us consider a single-tone modulating signal as

$$x(t) = \cos w_m t \quad ...(3.96)$$

The frequency spectrum of this modulating signal consist of two impulses located at $w = \pm w_c$ as shown in figure 3.29(*a*). If this modulating signal, modulates a carrier signal $\cos w_c t$, then the resulting spectrum of DSB-SC signal will be as shown in figure 3.29(*b*).

To get the SSB-SC waveform, we will have to eliminate one of the two sidebands.

Figure 3.29(c) shows the spectrum of SSB-SC wave with lower sidebands. From this figure, it is clear that this spectrum corresponds to a time-domain signal $\cos (w_c - w_m)t$, because the frequency spectrum of cosine function contains two impulses in its frequency domain. This means that a SSB-SC wave with lower sideband may be expressed as

$$\cos(w_c - w_m)t = \cos w_m t \cos w_c t + \sin w_m t \sin w_c t \quad ...(3.97)$$

In the same manner, the expression for the single-tone SSB-SC wave with upper sideband may be expressed as

$$\cos(w_c + w_m)t = \cos w_m t \cos w_c t - \sin w_m t \sin w_c t \quad ...(3.98)$$

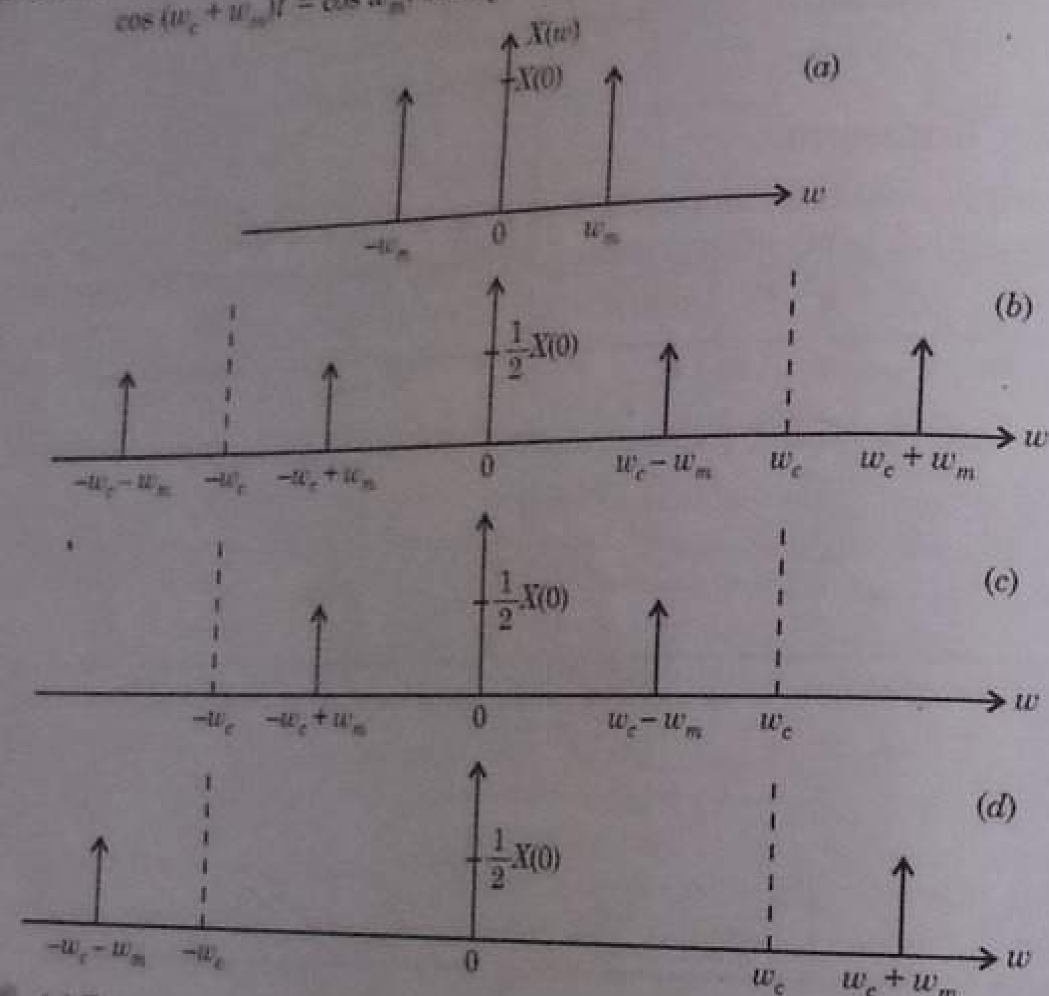


Fig. 3.29. *(a) Frequency spectrum of modulating signal (b) Frequency spectrum of DSB-SC wave (c) Spectrum of SSB-SC with lower sideband, (d) Spectrum of SSB-SC with upper sideband.*

Both these equations (3.97) and (3.98) may be combined as

$$s(t)_{SSB} = \cos w_m t \cos w_c t \pm \sin w_m t \sin w_c t \quad ...(3.99)$$

Here, the (+) sign represents the lower sideband and (–) sign represented the upper sideband.

We may write the terms $\sin w_c t$ and $\sin w_c t$ as

$$\sin w_m t = \cos\left(w_m t - \frac{\pi}{2}\right) \quad ...(3.100)$$

$$\sin w_c t = \cos\left(w_c t - \frac{\pi}{2}\right) \quad ...(3.101)$$

This means that the sine terms may be obtained using the corresponding cosine terms simply by giving a phase shift of $(-\pi/2)$.

The expression in equation (3.101) represents the SSB-SC wave for the case of single-tone modulation, however, it provides a way to get a general expression for SSB-SC wave.

In the expression of equation (3.99), the term $\sin w_m t$ is obtained by giving a phase shift of $(-\pi/2)$ to the modulating frequency $\cos w_m t$.

Similarly, in a general modulating signal $x(t)$, if all the frequency components are shifted by $(-\pi/2)$, it may lead to a general expression of SSB-SC signal.

As this point, it may be noted that $x(t)$ may be expressed as a continuous sum of sinusoidal signals.

Hence, expression in equation (3.99) may be extended for a SSB-SC signal modulated by a general modulating signal $x(t)$ as expressed below:

$$s(t)_{SSB} = x(t) \cos w_c t \pm x_h(t) \sin w_c t \quad ...(3.102)$$

where $x_h(t)$ is a signal obtained by shifting the phase of every component present in $x(t)$ by $(-\pi/2)$

Similar to equation (3.99), (+) signal corresponds to the lower sideband and (–) sign corresponds to the upper sideband.

3.21. Hilbert Transform

(Very Important)

It may be observed that the function $x_h(t)$ obtained by providing $(-\pi/2)$ phase shift to every frequency component present in $x(t)$, actually represents the **Hilbert transform** of $x(t)$. This means that $x_h(t)$ is the Hilbert transform of $x(t)$ defined as

$$x_h(t) = \frac{1}{\pi} x(t) \otimes \frac{1}{t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad ...(3.103)$$

Also, the inverse Hilbert transform is defined as

$$x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x_h(\tau)}{t-\tau} d\tau \quad ...(3.104)$$

Example 3.7. Show that if every frequency component of a signal $x(t)$ is shifted by an amount $\pi/2$, then the resultant signal $x_h(t)$ is the Hilbert transform of $x(t)$.

Solution: The given situation may be considered as though the signal $x(t)$ is passed through a phase shifting system having transfer function $H(\omega)$ and the output is $x_h(t)$ as shown in figure 3.30.

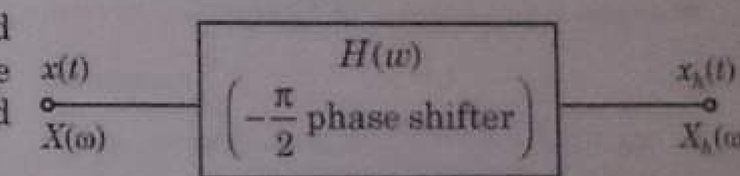


Fig. 3.30. *A phase shifting system.*

The characteristics of this system may be specified as under:

(*i*) The magnitude of the frequency components present in $x(t)$ remains unchanged when it is passed through the system. This means that $H(\omega) = 1$, and

(*ii*) The phase of the positive frequency components is shifted by $-\frac{\pi}{2}$. Now, since the phase spectrum $\theta(\omega)$ has an odd symmetry, the phase of the negative frequency components is shifted by $+\frac{\pi}{2}$. The spectrums $H(\omega)$ and $\theta(\omega)$ have been plotted in figure 3.31.

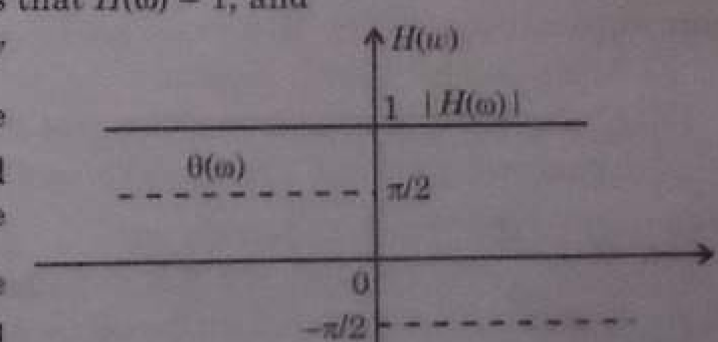


Fig. 3.31. *Transfer function of $-\pi/2$ phase shifter.*

The transfer function is expressed as

$$H(\omega) = |H(\omega)| e^{j\theta(\omega)} \quad ...(3.105)$$

or $\quad H(\omega) = 1.e^{j\theta}(\omega)$

From figure 3.31, it may be observed that

$$\theta(\omega) = \begin{cases} +\frac{\pi}{2} & \text{for } \omega < 0 \quad (i.e., \text{negtive frequencies}) \\ -\frac{\pi}{2} & \text{for } \omega > 0 \quad (i.e., \text{positive frequencies}) \end{cases} \quad ...(3.106)$$

Therefore equation (3.105) may be modified as

$$H(\omega) = \begin{cases} e^{j\pi/2} & \text{for } \omega < 0 \\ e^{-j\pi/2} & \text{for } \omega > 0 \end{cases} \quad ...(3.107)$$

Also, we know that $e^{j\frac{\pi}{2}} = \cos\frac{\pi}{2} + j\sin\frac{\pi}{2} = j$

and $e^{-j\frac{\pi}{2}} = \cos\left(-\frac{\pi}{2}\right) + j\sin\left(-\frac{\pi}{2}\right) = -j$

Thus, $H(\omega)$ becomes

$$\frac{H(\omega)}{j} = \begin{cases} 1 & \text{for } \omega < 0 \\ -1 & \text{for } \omega > 0 \end{cases} = -\,sgn\,(\omega)$$

or
$$H(\omega) = -j\,sgn\,(\omega) \quad ...(3.108)$$

The response $X_h(\omega)$ of the phase shifting system is related to the input $X(\omega)$ as

$$X_h(\omega) = X(\omega) \cdot H(\omega) \quad ...(3.109)$$

where $x(t) \longleftrightarrow X(\omega)$

and $x_h(t) \longleftrightarrow X_h(\omega)$

Now, substituting the value of $H(\omega)$ in equation (3.109) from equation (3.108), we get

$$X_h(\omega) = -jX(\omega)\,sgn\,(\omega)$$

Taking the inverse Fourier transform of both sides of last equation, we get

$$x_h(t) = F^{-1}[-j\,X(\omega)\,sgn(\omega)]$$

Also the time-domain of $sgn(\omega)$ is given as

$$\frac{1}{\pi t} \leftrightarrow sgn(\omega)$$

Using time convolution theorem, we get

$$x_h(t) = \frac{1}{\pi}\left[x(t) \otimes \frac{1}{t}\right] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}\,d\tau$$

which is the Hilbert transform of $x(t)$.

Some applications of the Hilbert transform may be listed as under:

(*i*) for generation of SSB signals,

(*ii*) for designing of minimum phase type filters,

(*iii*) for representation of bandpass signals.

Properties of the Hilbert Transform

Following are the properties of Hilbert transform:

(*i*) A signal $x(t)$ and its Hilbert transform $x_h(t)$ have the same energy density spectrum.

(*ii*) A signal $x(t)$ and its Hilbert transform $x_h(t)$ have the same autocorrelation function.

(*iii*) A signal $x(t)$ and its Hilbert transform $x_h(t)$ are mutually orthogonal. Mathematically,

$$\int_{-\infty}^{\infty} x(t)x_h(t)\,dt = 0$$

(*iv*) If $x_h(t)$ is a Hilbert transform of $x(t)$, then the Hilbert transform of $x_h(t)$ is $-x(t)$, i.e., ...(3.110)

If $H[x(t)] = x_h(t)$

then $H[x_h(t)] = -x(t)$

Here 'H' denotes the Hilbert transform.

The Pre-envelope or Analytic Signal

The concept of pre-envelope, also called as the analytic function is quite useful in deriving the general expression of the SSB-SC signal.

The pre-envelope of a real-valued signal $x(t)$ is defined as

$$x_p(t) = x(t) + jx_h(t) \quad ...(3.111)$$

where $x_h(t)$ is the Hilbert transform of signal $x(t)$. Clearly the pre-envelope $x_p(t)$ is a complex-valued signal. The real part of $x_p(t)$ is $x(t)$, and the imaginary part is its Hilbert transform $x_h(t)$. The complex conjugate of the pre-envelope denoted by $x_p(t)$ is expressed as

$$x_p^* = x(t) - jx_h(t) \quad ...(3.111a)$$

3.22. SSB-SC for a General Modulating Signal

In previous section, we have already derived an expression for a generalized SSB-SC signal by extending the concept of SSB-SC for a signal-tone modulation. In this article let us derive the same expression (*i.e.*, equation) by using the concept of pre-envelope or analytic signal.

As we know the pre-envelope of a function $x(t)$ is defined as

$$x_p(t) = x(t) + jx_h(t)$$

The Fourier transform of $x_p(t)$ is the sum of the Fourier transforms of $x(t)$ and $x_h(t)$ *i.e.*,

$$F[x_p(t)] = F[x(t)] + jF[x_h(t)]$$

or
$$X_p(\omega) = X(\omega) + j[-j\,X(\omega)\,sgn\,(\omega)]$$
$$= X(\omega) + X(\omega)\,sgn\,(\omega) \quad ...(3.112)$$

We know that

$$sgn\,(\omega) = \begin{cases} 1 & \text{for } \omega > 0 \\ -1 & \text{for } \omega < 0 \end{cases}$$

Therefore, we have

$$X_p(\omega) = \begin{cases} 2X(\omega) & \text{for } \omega > 0 \\ 0 & \text{for } \omega < 0 \end{cases} \quad ...(3.113)$$

Figure 3.32(*a*) and (*b*) show $X(\omega)$ and $X_p(\omega)$ respectively. It is obvious from figure 3.32(*b*) that $X_p(\omega)$ vanishes for negative frequencies.

Similarly, we can find the Fourier transform of $x_p^*(t)$ defined by equation (3.111)

$$X_p^*(\omega) = X(\omega) - j[-j \cdot X(\omega)\,sgn\,(\omega)]$$

or
$$X_p^*(\omega) = X(\omega) - X(\omega)\,sgn\,(\omega)$$

which may be written as

$$X_p^*(\omega) = \begin{cases} 0 & \text{for } \omega > 0 \\ 2X(\omega) & \text{for } \omega < 0 \end{cases}$$

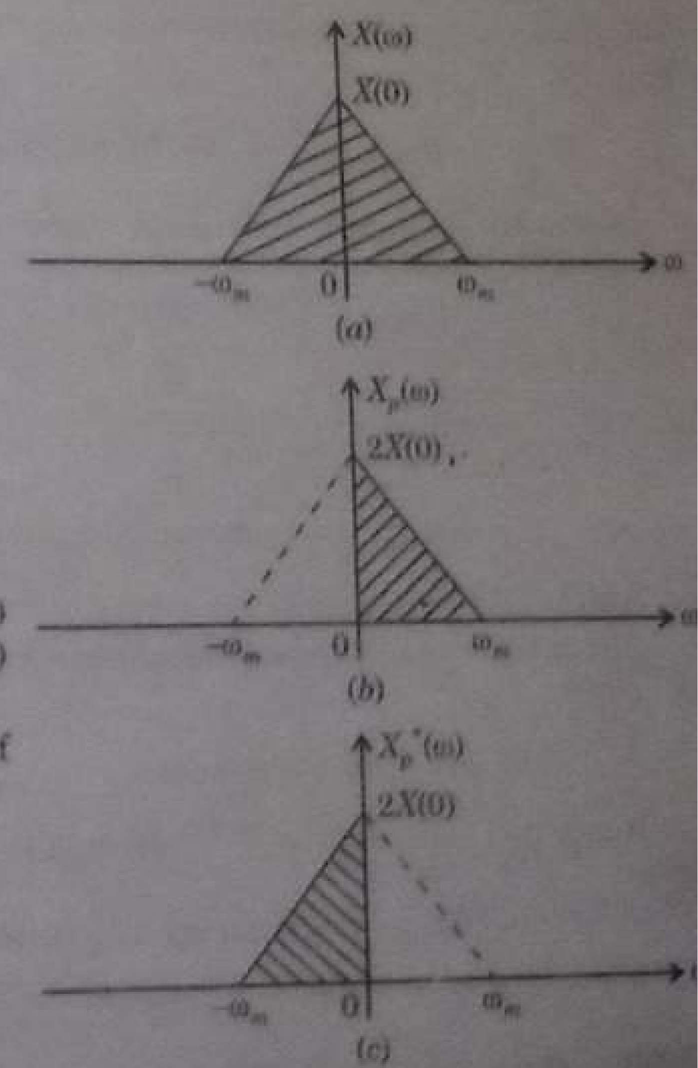


Fig. 3.32.

This has been plotted in figure 3.32(*c*). Clearly, $X_p^*(\omega)$ vanishes for positive frequencies.

Now, let us take an SSB-SC wave consisting of only the lower sidebands of a general modulating signal $x(t)$ as shown in figure 3.33.

Following points may be observed from figure 3.33.

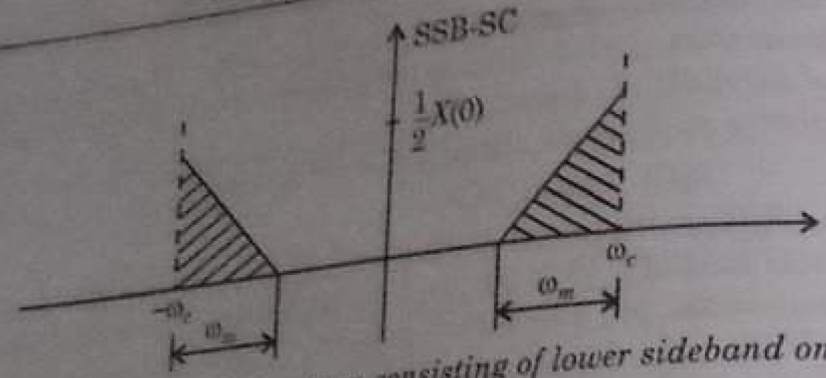


Fig. 3.33. *A SSB-SC spectrum consisting of lower sideband only.*

(i) The right-hand portion of the figure represents a spectrum of $\frac{1}{4}x_p^*(t)\,e^{j\omega_c t}$. This is equivalent to shifting the spectrum $X_p^*(\omega)$ towards right by ω_c.

(ii) Similarly, the left-hand portion of figure 3.34 represents the spectrum of $\frac{1}{4}x_p(t)\,e^{-j\omega_c t}$.

Thus, figure 3.33 represents the spectrum of a combined signal $x_p(t)\,e^{+j\omega_c t} + x_p(t)\,e^{-j\omega_c t}$.

Therefore, the time-domain representation of the SSB-SC spectrum shown in figure 3.34 is expressed as

$$s(t)_{SSB} = \frac{1}{4}[x_p^*(t)\,e^{j\omega_c t} + x_p(t)\,e^{-j\omega_c t}]$$

Substituting in terms of Hilbert transform from equations (3.110) and (3.111), we get

$$s(t)_{SSB} = \frac{1}{4}[x(t) - jx_p(t)]\,e^{j\omega_c t} + \frac{1}{2}[x(t) + jx_h(t)]\,e^{-j\omega_c t}$$

or
$$s(t)_{SSB} = \frac{1}{2}x(t)\left[\frac{e^{j\omega_c t} + e^{-j\omega_c t}}{2}\right] + \frac{j}{2}x_h(t)\left[\frac{e^{-j\omega_c t} - e^{j\omega_c t}}{2}\right]$$

or
$$s(t)_{SSB} = \frac{1}{2}[x(t)\cos\omega_c t + x_h(t)\sin\omega_c t] \quad ...(3.114)$$

which is the time-domain description of the SSB-SC signal consisting of only the lower sidebands.

Similarly, we can derive an expression for an SSB-SC signal consisting of the upper sidebands as under:

$$s(t)_{SSB} = \frac{1}{2}[x(t)\cos\omega_c t - x_h(t)\sin\omega_c t] \quad ...(3.115)$$

Hence, the time-domain description of the SSB-SC wave is represented by the expression as under:

$$s(t)_{SSB} = x(t)\cos\omega_c t \pm x_h(t)\sin\omega_c t \quad ...(3.116)$$

where '+' and '–' signs correspond to the lower sidebands and upper sidebands respectively.

3.23. Generation of SSB-SC Signal

SSB-SC signals may be generated by two methods as under:

(i) Frequency discrimination method or filter method

(ii) Phase discrimination method or phase-shift method

3.23.1. Frequency Discrimination Method

In a frequency discrimination method, firstly, a DSB-SC signal is generated simply by using an ordinary product modulator or a balanced modulator. After this, from the DSB-SC, signal one of the

two sidebands is filtered out by a suitable bandpass filter (BPF). The schematic diagram for this method is shown in figure 3.34.

Infact, the design of bandpass filter is quite critical and thus puts same limitations on the modulating or baseband and carrier frequencies

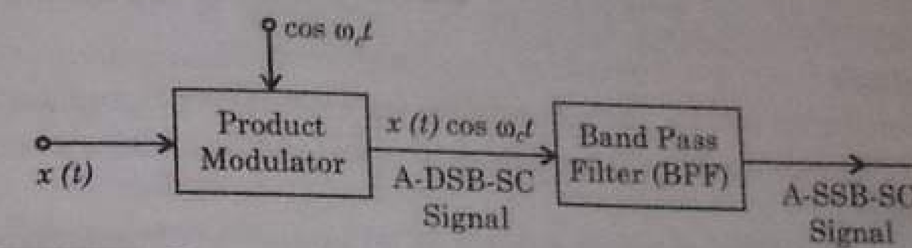


Fig. 3.34. *Frequency-discrimination method for SSB-SC generation.*

Following are the limitations for frequency discrimination method:

(i) The frequency-discrimination method is useful only if the baseband signal is restricted at its lower edge due to which the upper and lower sidebands are non-overlapping. For example, the filter method is used for speech communication where lowest spectral component is 70 Hz and it may be taken as 300 Hz without affecting the intelligibility of the speech signal. However, the system is not useful for video communication where the baseband signal starts from d.c.

(ii) The another restriction of the frequency discrimination method is that the baseband signal must be appropriately related to the carrier frequency. Infact, the design of the bandpass filter (BPF) becomes difficult if the carrier frequency is quite higher than the bandwidth of the baseband signal.

3.23.2. Phase-Shift Method

The phase-shift method avoids filter. This method makes use of two balanced modulators and two phase-shifting networks as shown in figure 3.35.*

In figure 3.35, one of the modulators, M_1 receives the carrier voltage shifted by 90° and the modulating voltage, whereas another balanced modulator M_2 receives the modulating voltage shifted by 90° and the carrier voltage.

Both balanced modulators produce an output consisting only of sidebands.

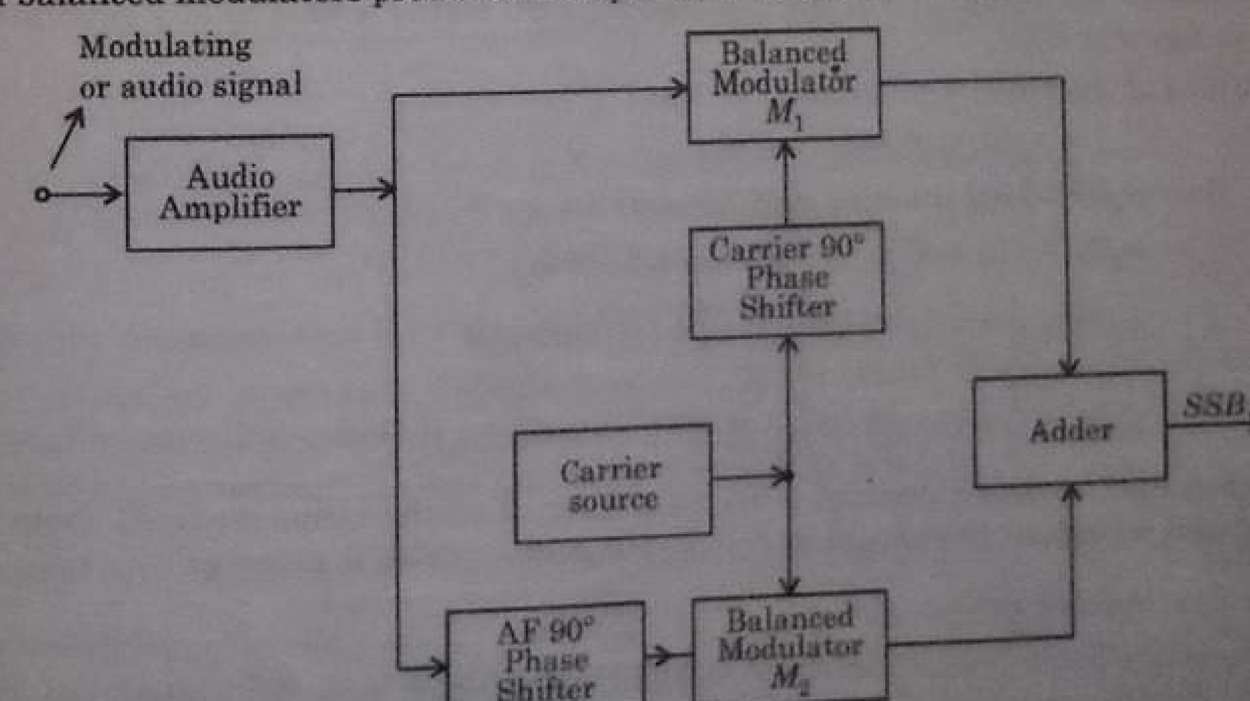


Fig. 3.35. *Phase-shift method for SSB-SC generation.*

* Draw the block diagram of phase cancellation SSB generation and explain how the carrier and the unwanted sideband are suppressed. (U.P. Tech. Sem. Exam., 2004-05)

Both the upper sidebands leads the input carrier voltage by 90°. One of the lower sidebands leads the reference voltage by 90° and the other lags it by 90°. The two lower sidebands are thus out of phase, and when combined together in the adder, they cancel each other. The upper sidebands are in phase at the adder and hence they add producing SSB in which the lower sideband has been cancelled.

Mathematical Analysis

Let the expression for carrier be sin $w_c t$ and that for modulating signal sin $w_m t$.

Now, the balanced modulator M_1 will receive sin $w_m t$ and sin $(w_c t + 90°)$ whereas M_2 will receive sin $(w_m t + 90°)$ and sin $w_c t$.

We know that the output of balanced modulator M_1 will contain sum and difference frequencies.

Hence,
$$v_1 = \cos[(w_c t + 90°) - w_m t] - \cos[(w_c t + 90°) + w_m t]$$

or
$$v_1 = \underset{\text{(LSB)}}{\cos(w_c t - w_m t + 90°)} - \underset{\text{(USB)}}{\cos(w_c t + w_m t + 90°)} \quad ...(3.117)$$

Similarly, the output of balanced modulator M_2 will contain

$$v_2 = \cos[w_c t - (w_m t + 90°)] - \cos[(w_c t + (w_m t + 90°)]$$

or
$$v_2 = \cos(w_c t - w_m t - 90°) - \cos(w_c t + w_m t + 90°) \quad ...(3.118)$$

Therefore, the ouput of the adder will be

$$v_o = v_1 + v_2 = 2\cos(w_c t + w_m t + 90°) \quad ...(3.119)$$

which is the expression for SSB-SC.

3.24. Demodulation of SSB-SC Signals

The baseband or modulating signal $x(t)$ can be recovered from the SSB-SC signal by using the synchronous detection technique as already discussed for DSB-SC signals. With the help of synchronous detection method the spectrum of an SSB-SC signal centered about $\omega = \pm\,\omega_c$, is retranslated to the baseband spectrum which is centered about $\omega = 0$. The process of synchronous detection involves multiplication of the received SSB-SC signal with a locally generated carrier as shown in figure 3.36(a).

The output of the multiplier will be

$$e_d(t) = s(t)_{SSB} \cdot \cos\omega_c t$$

or
$$e_d(t) = [x(t)\cos\omega_c t \pm x_h(t)\sin\omega_c t]\cos\omega_c t$$

or
$$e_d(t) = x(t)\cos^2\omega_c t \pm x_h(t)\sin\omega_c t\cos\omega_c t$$

or
$$e_d(t) = \frac{1}{2}x(t)[1+\cos 2\omega_c t] \pm \frac{1}{2}x_h(t)\sin 2w_c t$$

or
$$e_d(t) = \frac{1}{2}x(t) + \frac{1}{2}[x(t)\cos 2\omega_c t \pm x_h(t)\sin 2\omega_c t] \quad ...(3.120)$$

Now when $e_d(t)$ is passed through a low-pass filter, then the terms centered about $\pm\,2\omega_c$ are filtered out and we get, at the output of detector, signal e_0 which is given as

$$e_0(t) = \frac{1}{2}x(t) \quad ...(3.121)$$

The frequency-domain explanation is evident from figure 3.36. Multiplication of the SSB signal with cos $\omega_c t$ in the time-domain is equivalent to the convolution of their spectra as shown in figure 3.36(*d*). The component centred around $\pm\,2\omega_c$ is filtered out where as the message signal centered at $\omega = 0$ appears at the output. The synchronous detection can be achieved by a ring demodulator circuit or a modulator circuit using non-linear devices.

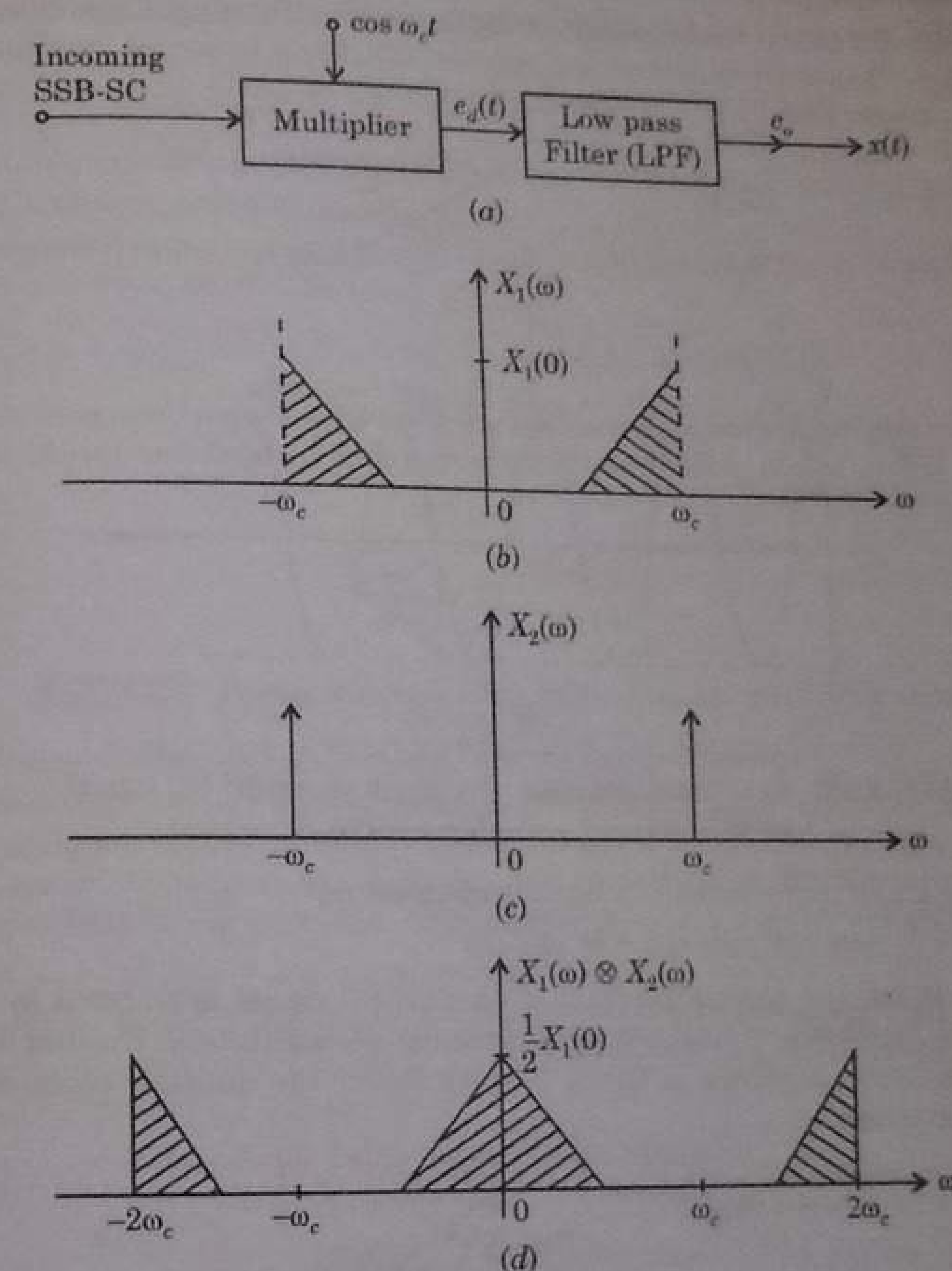


Fig. 3.36. *Demodulation of SSB-SC signal (a) Synchronous detector (b) SSB-SC (lower sideband) (c) Spectrum of cos $\omega_c t$ (locally generated carrier) (d) Convolution of figures (b) and (c).*

3.24.1. Phasor Diagram and Waveform of SSB-SC Signals

In this article, let us consider a single tone modulating signal $V_1\cos\omega_1 t$ which modulates a carrier signal $V_c\cos\omega_c t$ to generate an SSB-SC signal.

The resulting signal considering lower sideband, will be given as

$$s(t)_{SSB} = \frac{1}{2}V_1 V_c\cos(\omega_c - \omega_1)t \quad ...(3.122)$$

The signal represents a single sinusoid with an amplitude $\frac{1}{2}V_1 V_C$ and a frequency $(\omega_c - \omega_1)$.

Therefore the phasor diagram consists of a single phasor and there will be no amplitude fluctuation in the modulated wave. This is obvious from the phasor diagram in figure 3.37(*a*) and the waveform drawn in figure 3.37(*b*). It is also clear from the waveform that both amplitude and frequency of the wave are constant. This means that there is no amplitude modulation.

Hence, an SSB-SC signal modulated by a signal tone modulating signal does not carry any useful information. However, the baseband signals appearing in actual systems are multiple tone rather than single tone.

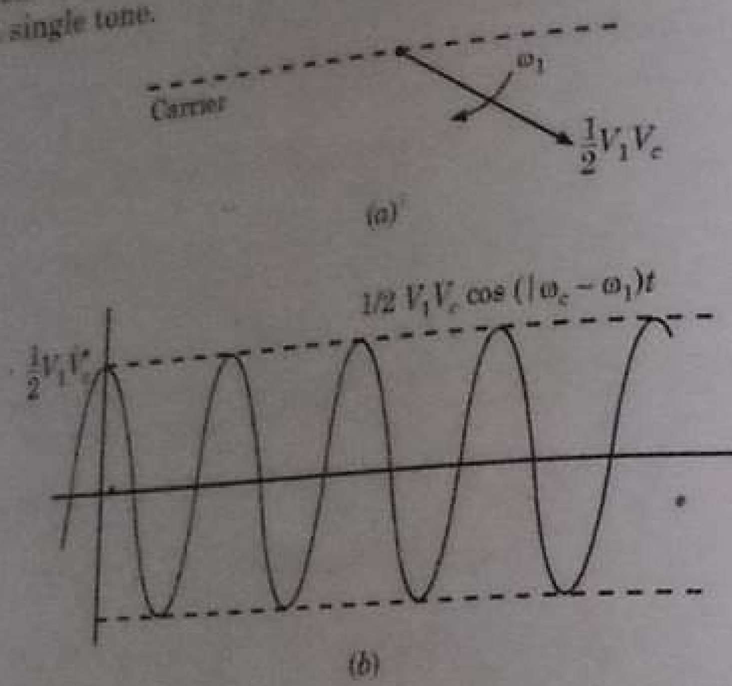


Fig. 3.37. *(a) Phasor diagram of a single tone SSB-SC signal*
(b) Waveform of a single tone SSB-SC signal

Let us consider a two tones baseband signal expressed as

$$x(t) = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t$$

Figure 3.38(*a*) shows the phasor diagram of an SSB-SC signal modulated by the baseband signal $x(t)$. The resultant phasor varies with time as the two phasors rotate. The amplitude variation in the resulting waveform is shown in figure 3.38(*b*). Hence the multiple tones SSB-SC signal shows amplitude variation.

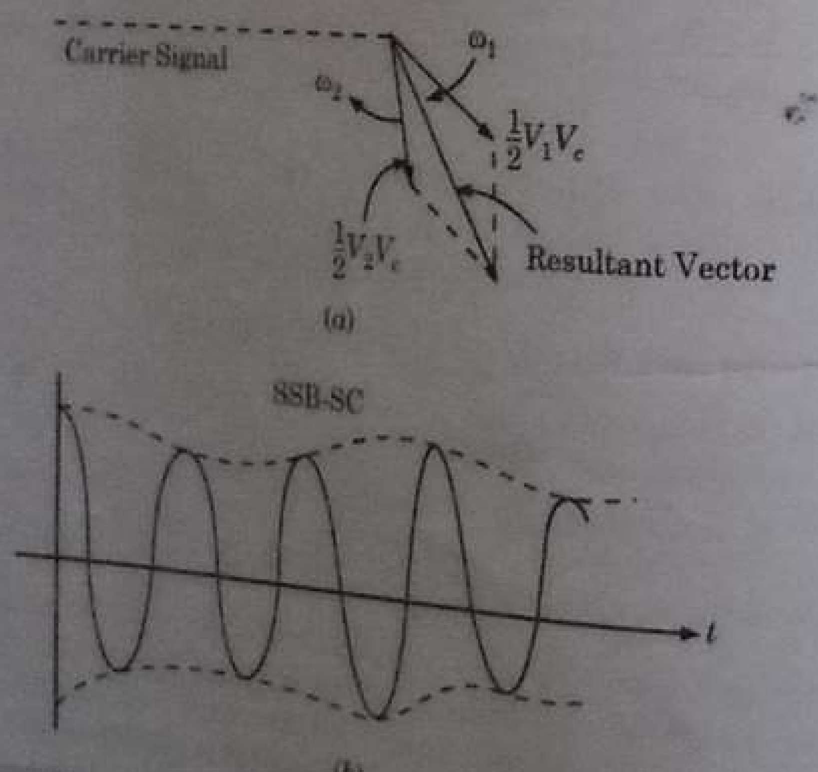


Fig. 3.38. *(a) Phasor diagram of a two-tones SSB-SC signal.*
(b) Waveform of a two-tones SSB-SC signal.

3.24.2. Waveform of SSB-SC with Large Carrier Signal

In this article, let us consider an SSB-SC signal in which a large carrier is also present. This carrier signal can be introduced at the transmitter as in simple AM, or can be added at the receiver end. When a large carrier is added to the SSB-SC signal the demodulation becomes quite easy. When a large carrier signal is present, the SSB-SC waveform shows amplitude variations even for a single-tone modulating signal.

The expression for such a wave may be obtained by adding a large carrier terms in the expression of SSB-SC given in equation (3.122) *i.e.*,

$$s(t)_{SSB} = \frac{1}{2} V_1 V_c \cos(\omega_c - \omega_1)t + V_c \cos \omega_c t \text{ for } V_c > V_1$$

The phasor diagram has been shown in figure 3.39. The SSB-SC waveform is formed by the super position of two waveforms (*i.e.*, lower sideband and large carrier term).

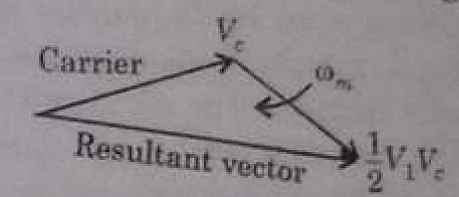


Fig. 3.39. *Phasor diagram of an SSB-SC signal with large carrier.*

3.24.3. Detection of the SSB-SC Signal having Large Carrier

Whenever a large carrier is introduced, a synchronous detection is not essential to recover the modulating signal. In this situation, an envelope detector provides the approximate modulating signal. An envelope detector is quite simpler and cheaper than synchronous detector. Thus, an expression for a SSB-SC signal with a large carrier $A \cos \omega_c t$ may be written as

$$s(t)_{SSB} = x(t) \cos \omega_c t + x_h(t) \sin \omega_c t + A \cos \omega_c t \quad \text{...(3.12}$$

Here, it may be noted that the carrier is added to an SSB-SC signal (*i.e.*, not multiplied with it as in synchronous detection). This technique is also known as **carrier re-insertion technique.**

Equation (3.123) may be written as

$$s(t)_{SSB} = [A + x(t)] \cos \omega_c t + x_h(t) \sin \omega_c t$$

or

$$s(t)_{SSB} = e(t) \cos(\omega_c t + \theta)$$

where

$$e(t) = \sqrt{[A + x(t)]^2 + x_h^2(t)}$$

is the envelope of the wave and

$$\theta = -\tan^{-1}\left[\frac{x_h(t)}{A + x(t)}\right]$$

is the phase of the wave. When this SSB-SC signal is applied to an envelope detector, the output of the detector will be the envelope $e(t)$ *i.e.*,

$$e(t) = \left[\{A + x(t)\}^2 + x_h^2(t)\right]^{\frac{1}{2}}$$

Simplifying, we get

$$e(t) = A\left[1 + \frac{2x(t)}{A} + \frac{x^2(t)}{A^2} + \frac{x_h^2(t)}{A^2}\right]^{\frac{1}{2}}$$

which may be expanded in the form of a Binomial series, gives as

$$e(t) = A\left[1 + \frac{2x(t)}{2A} + \text{higher terms}\right]$$

But since $A >> |x(t)|$ and $|x_h(t)|$ and higher terms may be neglected, therefore

$$e(t) \cong A + x(t)$$

Hence, the output of the envelope detector is close to the desired modulating signal $x(t)$. The carrier re-insertion technique is also useful for the detection of DSB-SC signals as discussed already.

This large carrier may be added either at the transmitter or at the receiver. When the carrier signal is added at the receiver, the synchronization problem remains as it is and this technique has no advantage. When carrier is added at the transmitter than this mode has the advantage of both SSB-SC and AM systems. It requires only half the bandwidth, and at the same time, detection becomes quite simple.

3.24.4. Compatible Single Sideband Signal

An SSB-SC signal may be generated in which the carrier is suppressed, even then it can be detected with an envelope detector (*i.e.*, simple diode detector). Such a signal is compatible for reception using a commercial AM radio receiver. The signal with this characteristics is termed as **compatible single sideband (CSSB)**. However, such a signal involves complex signal processing and the system is presently impractical for commercial applications.

3.25. Vestigial Sideband (VSB) Modulation Systems *(Important)*

A VSB modulation system is actually a compromise between DSB-SC and SSB modulation system. In other words, we can say that it is an optimum choice in which the advantages of DSB-SC and SSB modulation systems have been exploited.

As a matter of fact, the generation of VSB modulation signals is easier than other modulated signals such as conventional AM, DSB-SC and SSB signals. Its bandwidth is only slightly higher (approximately 25%) than SSB signals but considerable less than DSB-SC signals. SSB modulation is rather most suited for the transmission of voice signals because of the energy gap that exists in the frequency spectrum of the vioce signals between zero and few hundred hertz. On the other hand, when signals contain frequency components at extremely low frequencies (as in telegraph signals) the USB and LSB of the translated signal tend to meet at the carrier frequency. Under such circumstances, it becomes very difficult to isolate one sideband from the other. Hence, SSB scheme becomes unsuitable for handling such types of signals.

This difficulty has been overcome in a scheme known as **vestigial sideband (VSB) modulation.** In VSB modulation instead of rejecting one sideband completely as in SSB modulation scheme, a gradual cut-off of one sideband is allowed. This gradual cut is compensated by a vestige or portion of the other sideband.

This technique has been illustrated in figure 3.40 which shows the frequency spectrum of the modulating signal and corresponding DSB-SC, SSB and VSB signals.

Obviously, the bandwidth of VSB signal is given by

$$BW = f_c + f_v - f_c + f_m$$

$$BW = f_m + f_v$$

where f_m = Bandwidth of the message signal, and
f_v = width of the VSB.

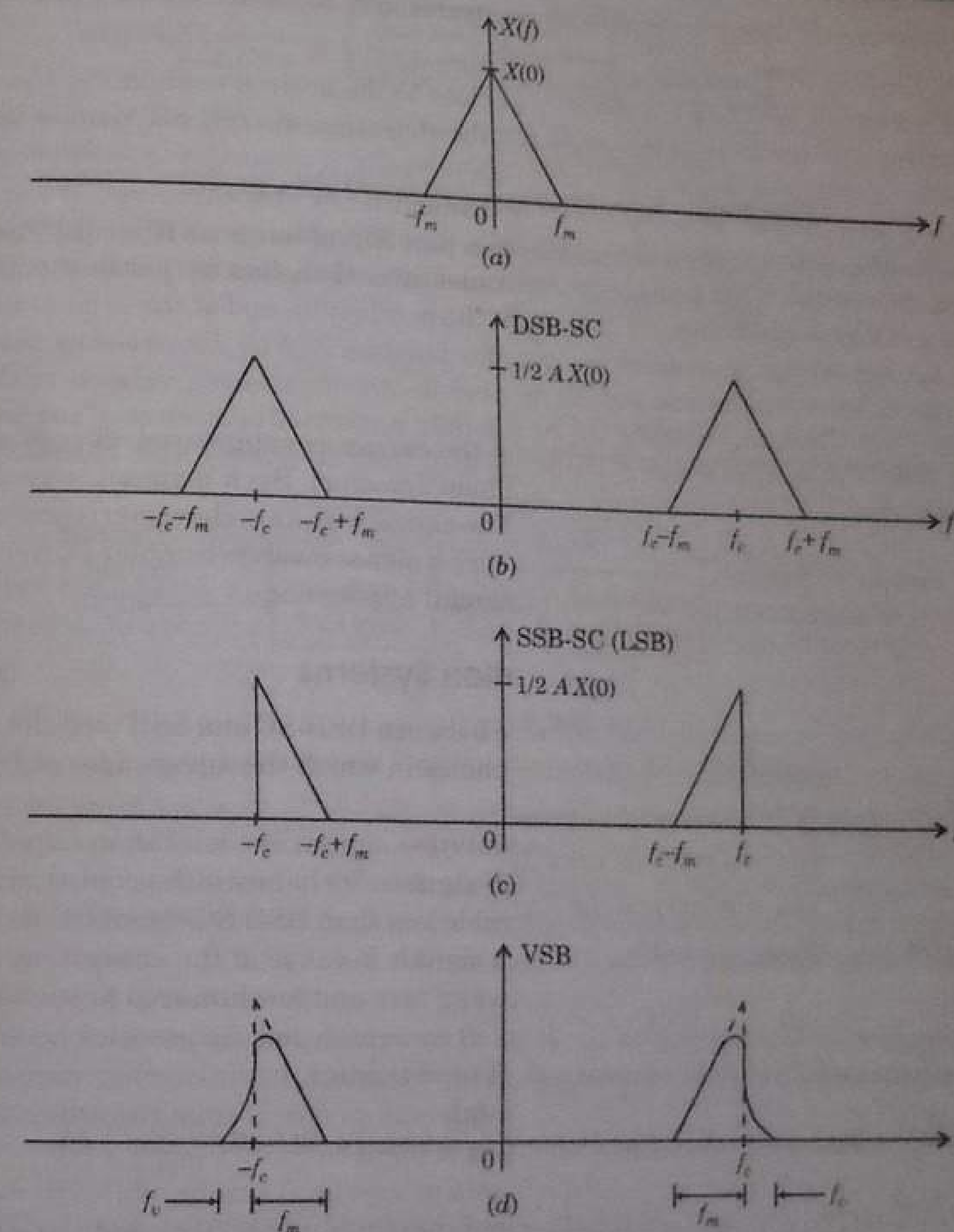


Fig. 3.40. *Illustration of frequency spectrum of VSB signal.*
(a) Frequency spectrum of Baseband signal x(t)
(b) Frequency spectrum of DSB-SC signal
(c) Frequency spectrum of SSB signal
(d) Frequency spectrum of VSB signal.

3.26. Generation of VSB Signals

Basically, VSB signal can be generated by passing a DSB-SC signal through an appropriate filter having transfer function $H(f)$. It has been shown in figure 3.41. The frequency spectrum of VSB signal, $S(f)$ is therefore, given by

$$S(f) = \text{Fourier transform } \{s(t)\}$$

or

$$S(f) = \frac{A}{2}[X(f-f_c) + X(f+f_c)]\,H(f) \quad ...(3.124)$$

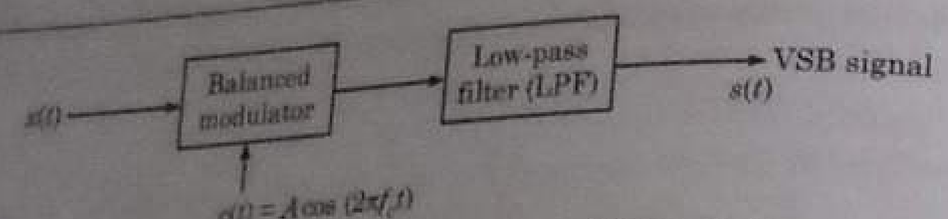


Fig. 3.41. *Scheme for generation of VSB signal.*

Now, let us find the specification of transfer function $H(f)$ of low-pass filter (LPF) correspond to the spectrum $S(f)$ of the VSB signal $s(t)$. This can be established by passing $s(t)$ through a coherent detector without distortion.

We shall, now, derive the expression for transfer function $H(f)$ by determining the necessary conditions for the coherent detection output to provide an undistorted version of the original modulating signal $x(t)$. Thus, multiplying $s(t)$ by a locally generated sine wave $A' \cos 2\pi f_c t$), which is synchronous with the carrier wave $A \cos(2\pi f_c t)$ in both frequency and phase. It has been shown in figure 3.42.

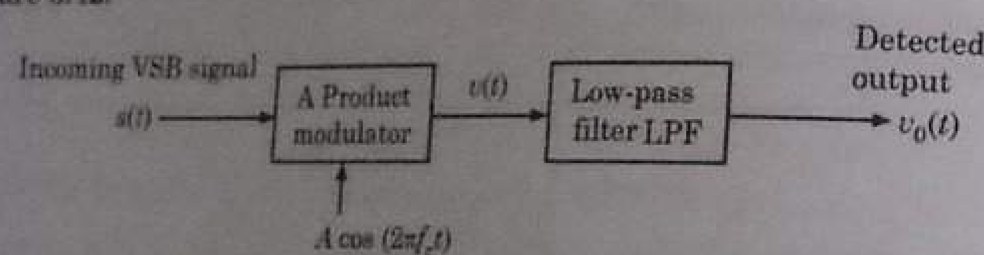


Fig. 3.42. *Block diagram of VSB demodulator.*

The output of the product modulator is given by

$$v(t) = s(t)\, A' \cos (2\pi f_c t)$$

or

$$v(t) = A' s(t) \cos (2\pi f_c t)$$

Taking Fourier Transform, we have ...(3.125)

$$V(f) = \frac{A'}{2}[S(f-f_c)+S(f+f_c)]$$

Substituting equation (3.124) in equation (3.126), we obtain ...(3.126)

$$V(f) = \frac{AA'}{4}[X(f-2f_c)+X(f)]H(f-f_c)+\frac{AA'}{4}[X(f)+X(f+2f_c)]H(f+f_c)$$

or

$$V(f) = \frac{AA'}{4}[H(f-f_c)+H(f+f_c)]X(f)+\frac{AA'}{4}[X(f-2f_c)H(f-f_c)+X(f+2f_c)H(f+f_c)]$$

...(3.127)

The spectrum of $V(f)$ has been shown in figure 3.43. The first term of equation (3.127) corresponds the frequency spectrum of the modulating signal. The second term of $V(f)$ corresponds to the quency spectrum of the VSB signal having carrier frequency $2f_c$. The second term can be removed using low-pass filter (LPF).

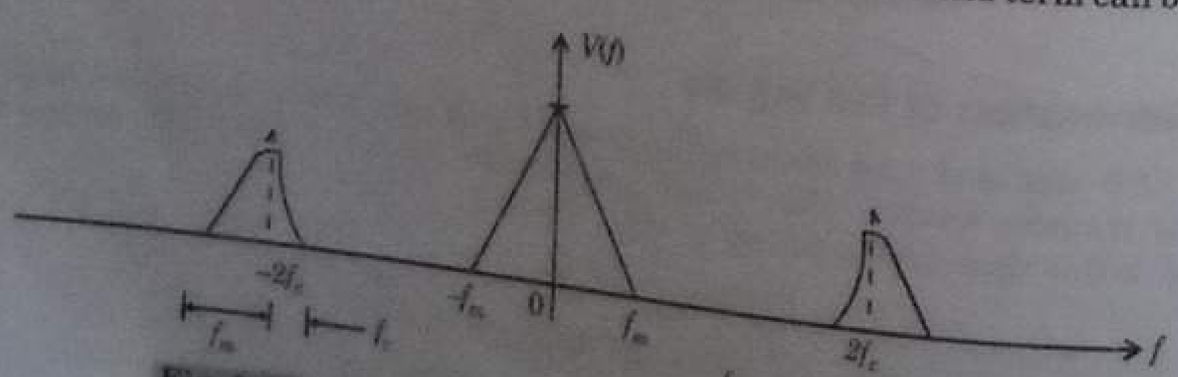


Fig. 3.43. *Spectrum of the product modulator ...*

The frequency spectrum of the signal $v_0(t)$ at the output of the LPF is given by

$$V_0(f) = \frac{AA'}{4}[H(f-f_c)+H(f+f_c)]X(f) \quad ...(3.128)$$

For a distortionless reproduction of the original modulating signal $x(t)$ at the output of the coherent detector, we must have $V_0(f)$ as a scaled version of $X(f)$. It has been shown in figure 3.44.

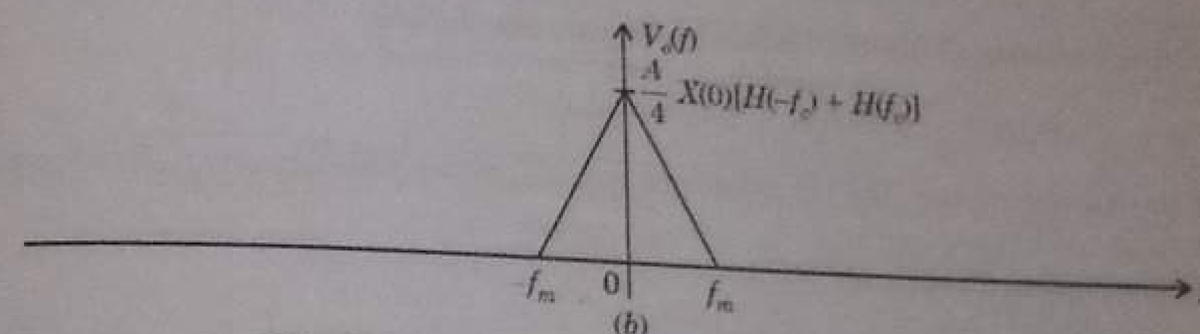


Fig. 3.44. *Spectrum of the demodulated signal $v_0(t)$.*

Thus, for distortionless reception by the VSB modulation scheme, it is necessary that the transfer function of the filter $H(f)$ must satisfy the condition given as under:

$$H(f-f_c)+H(f+f_c) = 2H(f_c) \quad ...(3.129)$$

where $H(f_c)$ is the transfer function of LPF at carrier frequency f_c. It is a constant value.

Now, when the spectrum $X(f)$ is zero outside the interval $-f_m \le f \le f_m$, equation (3.141) is required to be satisfied only in the above range. This requirement can be fulfilled by using a filter having a normalised amplitude response as shown in figure 3.45 for positive frequencies. The magnitude of normalised $|H(f)|$ should be one-half at the carrier frequency f_c. The sum of the values of $|H(f)|$ at any two frequencies equally spaced above and below f_c, should be unity in the transition band $f_c - f_v \le f \le f_c + f_v$. The design of such a filter is much more simpler than the design of filters needed in generation of SSB signals.

In order to preserve the frequency spectrum of $x(t)$ properly, the phase spectrum of the signal should be linear in the interval $f_c - f_v \le |f| \le f_c + f_m$ and its value should be equal to zero at f_c or an integral multiple of 2π.

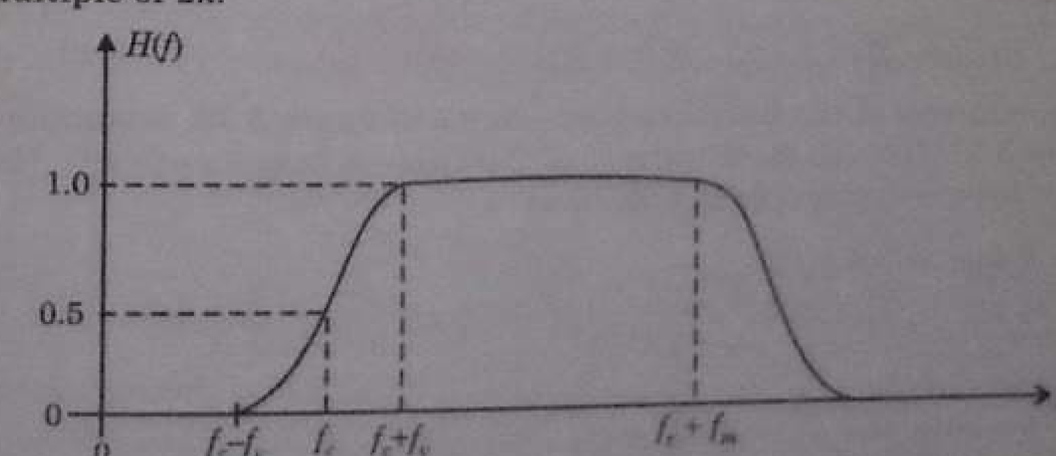


Fig. 3.45. *Illustration of amplitude response of VSB filter.*

3.26.1. The Time-Domain Representation of the VSB Signal

The time-domain representation of the VSB signal can be obtained by using the canonical form of the bandpass signal given by

$$s(t) = s_c(t)\cos(2\pi f_c t) - s_s(t)\sin(2\pi f_c t) \quad ...(3.129a)$$

Frequency spectrum of VSB signal will be

$$S(f) = \frac{A}{2}[X(f-f_c) + X(f+f_c)]\,H(f) \qquad ...(3.129b)$$

Fourier transform of the in phase component $s_c(t)$ will be given by

$$S_c(f) = \begin{cases} S(f-f_c) + S(f+f_c), & \text{for } -f_m \le f \le f_m \\ 0, & \text{otherwise} \end{cases} \qquad ...(3.130)$$

Substituting equation (3.129b) in equation (3.130), we obtain

$$S_c(f) = \frac{1}{2}A[H(f-f_c) + H(f+f_c)]X(f) \qquad ...(3.131)$$

Here we have assumed that the transfer function $H(f)$ of the VSB filter satisfies the equation (3.129) with $H(f_c) = 0.5$

Thus using equation (3.131), we obtain

$$S_c(f) = \frac{1}{2}A\,X(f) \qquad ...(3.132)$$

Taking the inverse Fourier transform of both sides of equation (3.132), we get the in-phase component of the VSB signal as under:

$$s_c(t) = \frac{1}{2}A\,x(t) \qquad ...(3.133)$$

The Fourier transform of the quadrature components $s_s(t)$ can be obtained by substituting equation (3.131) in equation given below:

$$S_s(f) = \begin{cases} j[S(f-f_c) + S(f+f_c)], & \text{for } -f_m \le f \le f_m \\ 0, & \text{otherwise} \end{cases}$$

$$S_s(f) = \frac{j}{2}A[H(f-f_c) - H(f+f_c)]X(f) \qquad ...(3.134)$$

From equation (3.134), it may be observed that we can generate the quadrature component $s_s(t)$, except for a scaling factor, by passing the message signal through a filter whose transfer function is given by

$$H_s(f) = j[H(f-f_c) - H(f+f_c)] \qquad ...(3.135)$$

The frequency response of the filter has been shown in figure 3.46, assuming that $H(f)$ is defined as in figure 3.46. Let $x_s(t)$ be the output of this filter in response to $x(t)$. Hence, we can express the quadrature component of the VSB signal as

$$s_s(t) = \frac{1}{2}A\,x_s(t) \qquad ...(3.136)$$

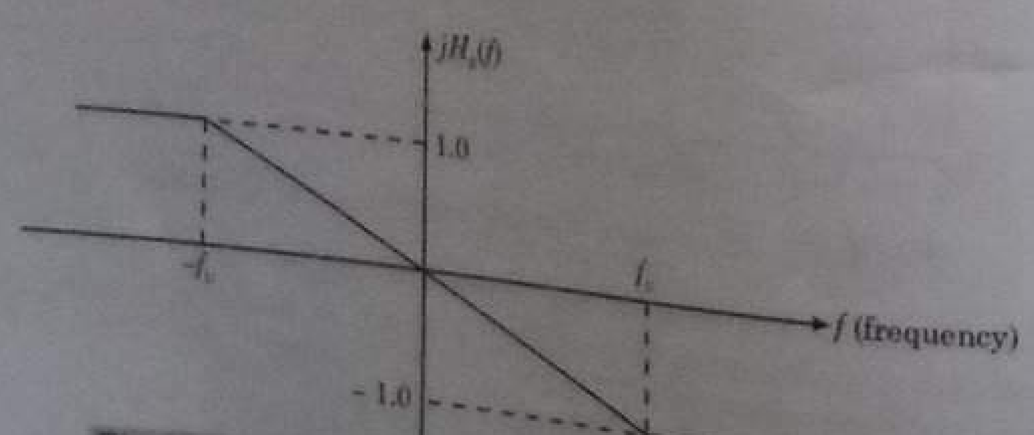


Fig. 3.46. *Frequency response of the filter to produce $s_s(t)$.*

Substituting equations (3.133) and (3.136) in equation (3.129a), we get the time-domain representation of VSB signal as under:

$$s(t) = \frac{A}{2}[x(t)\cos(2\pi f_c t) - x_s(t)\sin(2\pi f_c t)] \qquad ...(3.137)$$

A scheme for generating VSB signal on the basis of the time-domain representation has been illustrated in figure 3.47.

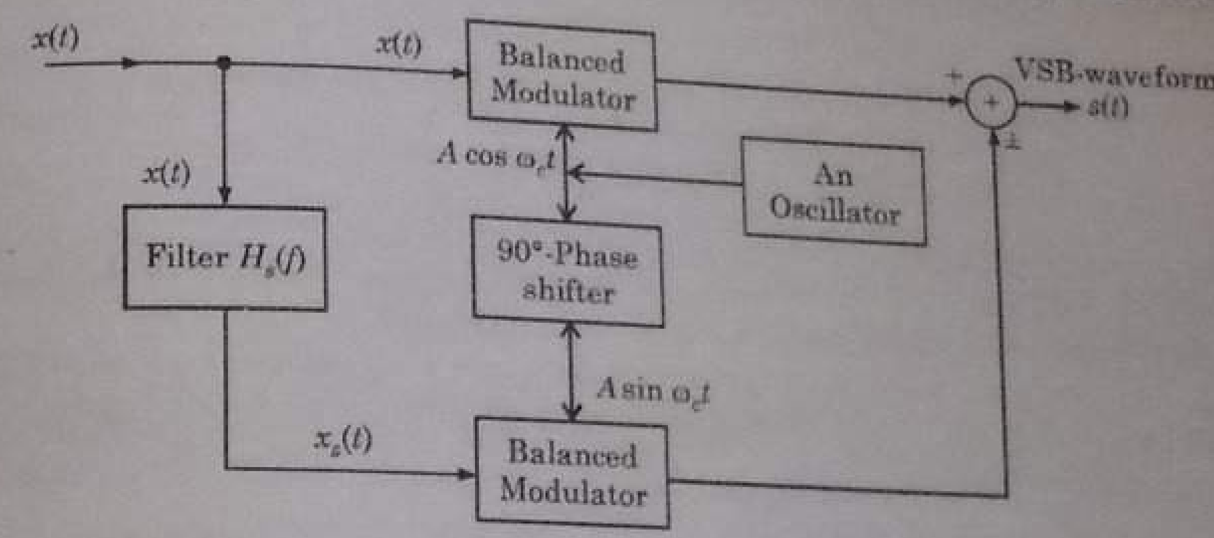


Fig. 3.47. *Illustration of Block diagram of phase discrimination method for generating VSB signals.*

Note: *By following a procedure similar to the above, the time-domain representation of VSB signal containing a vestige of the LSB may also be obtained in a similar form as in equation (3.137) except that a plus sign appears on the RHS in place of minus sign. Both types of VSB can be obtained by using the generalised scheme shown in figure 3.46, taking care of adding and subtracting the outputs of the two product modulators.*

3.27. Demodulation of VSB Signals *(Expected)*

VSB signals can be detected using coherent detection or synchronous detection technique. However, we use another scheme in which an envelope detector is used for demodulating VSB signal containing a large carrier. In commercial television broadcasting, a sizeable carrier is usually transmitted with the video signal which is sent using VSB. We can represent VSB + carrier using equation given as follows:

$$s(t) = \frac{A}{2}[x(t)\cos(2\pi f_c t) - x_s(t)\sin(2\pi f_c t)] \qquad ...(3.138)$$

as

$$s(t) = \frac{A}{2}\left[1 + \frac{k_a}{2}x(t)\right]\cos(2\pi f_c t) - \frac{k_a x_s(t)}{2}\sin(2\pi f_c t)$$

where k_a is the scaling factor.*

The last equation can be written as under:

$$s(t) = e(t)\cos(2\pi f_c t + \theta) \qquad ...(3.139)$$

where

$$e(t) = \sqrt{A^2\left[1 + \frac{k_a x(t)}{2}\right]^2 + A^2\left[\frac{k_a x_s(t)}{2}\right]^2} \qquad ...(3.140)$$

* Scaling factor k_a determines the percentage modulation.

and
$$\theta = \tan^{-1}\left[\frac{\frac{1}{2}k_a x_s(t)}{1+\frac{1}{2}k_a x(t)}\right] \quad ...(3.141)$$

The envelope of $s(t)$ can thus be written as under:

$$e(t) = A\left[1+\frac{1}{2}k_a x(t)\right]\left[1+\left\{\frac{\frac{1}{2}k_a x_s(t)}{1+\frac{1}{2}k_a x(t)}\right\}^2\right]^{\frac{1}{2}} \quad ...(3.142)$$

Note: *It can be observed from equation (3.142) that the envelope consists of the modulating signal $x(t)$ and hence can be detected using an envelope detector. However, distortion is introduced by the quadrature component $x_s(t)$ of the incoming VSB signal. This distortion can be reduced by reducing the modulation index and by increasing the width of VSB and thus reducing $x_s(t)$.*

3.28. Comparison of Various AM Systems

In this section, let us compare conventional AM system, DSB-SC system, SSB system and VSB system of Amplitude modulation. We know that the conventional AM signal contains the two sidebands as well as carrier. In suppressed carrier schemes, the carrier is removed before transmission process. The relative merits and demerits of various forms of AM can be summarized below:

(*i*) The demodulation or detection of AM signal is simpler than DSB-SC and SSB systems. The conventional AM can be demodulated by rectifier or envelope detector. Detection of DSB-SC and SSB is rather difficult and expensive alsop. Furthermore, it is quite easier to generate conventional AM signals at high power levels as compared to DSB-SC and SSB signals. For this reason, conventional AM systems are used for broadcasting purpose.

(*ii*) The advantage of DSB-SC and SSB systems over conventional AM system is that the former requires lesser power to transmit the same information. For sinusoidal modulation the carrier consumes about 2/3 of the total power for 100% modulation in conventional AM. However, only 1/3 of the total power is carried by the sidebands which carry the information. This makes AM transmitters less coefficient. On the other hand, the receivers of DSB-SC and SSB systmes, though efficient, are much more complex and expensive too. Due to this reason, DSB-SC and SSB systems only find applications in point-to-point communication. In point-to-point communication only a few receives and one transmitter are needed. In public broadcast system one transmitter caters to millions of receivers. It is obviously needed to the simpler and cheaper.

(*iii*) SSB scheme needs only one-half of the bandwidth required in DSB-SC system and less than that required in VSB also. Thus, we can say that SSB modulation scheme is the most-efficient scheme among DSB-SC and VSB schemes. SSB modulation scheme is used for long distance transmission of voice signals because it allows longer spacing between repeaters.

3.29. Frequency Translation and Mixing

In the processing of signals in communication systems, it is often required to translate or shift the modulated signal to a new frequency band. As an example, in most commercial AM radio receivers, the received radio-frequency (RF) signal [540 to 1640 kilohertz (kHz)] is shifted to the intermediate-

frequency (IF) (455 kHz) band for processing. The received signal, now translated to a fixed IF, can easily be amplified, filtered, and demodulated.

A device that performs the frequency translation of a modulated signal is known as a *frequency mixer*. The operation is often called ***frequency mixing, frequency conversion,*** or ***heterodyning.***

A common problem associated with frequency mixing is the presence of the *image frequency*. For example, in an AM superheterodyne receiver, the locally generated frequency is chosen to be 455 kHz higher than the incoming signal. Suppose that the reception of an AM station at 600 kHz is desired. The locally generated signal is at 1055 kHz. Now if there is another station at 1510 kHz, it will also be received (note that 1510 kHz – 155 kHz = 455 kHz). The second frequency, 1510 kHz = 600 kHz + 2(455 kHz), is called the image frequency of the first, and after the heterodyning operation it is impossible to distinguish the two. Note that the image frequency is separated from the desired signal by exactly twice the IF. Usually, the image frequency signal is attenuated by a selective RF amplifier placed before the mixer.

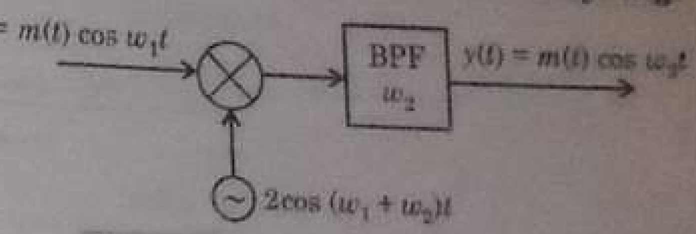


Fig. 3.48. *A Frequency mixer.*

MISCELLANEOUS SOLVED EXAMPLES

Example 3.8. An audio signal given as 15 sin 2 π(1500*t*) amplitude modulates a carrier given as 60 sin 2 π (100,000*t*) determine the following:

(*a*) Sketch the audio signal.

(*b*) Sketch the carrier signal.

(*c*) Construct the modulated wave.

(*d*) Determine the modulation index and per cent modulation.

(*e*) What are the frequencies of the audio signal and the carrier?

(*f*) What frequencies would present in a spectrum analysis of the modulated wave?

Solution: Given that Audio signal = $15 \sin 2\pi(1500t)$

Carrier = $60 \sin 2\pi(100{,}000t)$

(*a*) audio signal or modulating signal is sketched in figure 3.49

(*b*) The carrier signal is sketched in figure 3.49

(*c*) To construct the modulated wave, first let us develop the envelope of the modulated wave in the following two steps:

(*i*) Locate the amplitude of the carrier (dashed line).

(*ii*) Using the amplitude of the carrier as an axis, lay in the audio signal.

Now that the envelope has been determined, a signal having an amplitude defined by the envelope found above and having a frequency of the carrier is laid in within the envelope as shown in figure 3.49 (*c*).

(*d*) The modulation index is given by

$$m_a = \frac{\text{Maximum audio amplitude}}{\text{Maximum carrier amplitude}} = \frac{15}{60} = \frac{1}{4} = 0.25 \quad \textbf{Ans.}$$

Further, converting modulation index to per cent modulation, we have

$$M = m_a \times 100 = 0.25 \times 100 = 25\% \quad \textbf{Ans.}$$

(*e*) Since the expression for audio signal is given by

$$v_m = V_m \sin 2\pi f_m t = 15 \sin 2\pi(1500t)$$

Hence $f_m = 1500$ Hz **Ans.**

Since the expression for carrier signal is given by

$$v_c = V_c \sin 2\pi f_c t = 60 \sin 2\pi (100{,}000)$$

Therefore, $f_c = 100{,}000$ Hz **Ans.**

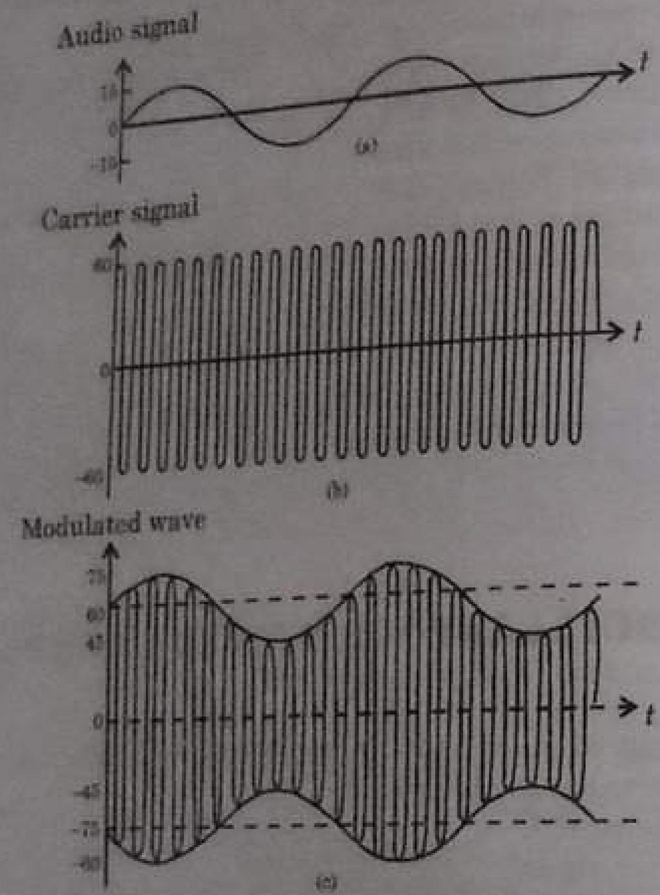


Fig. 3.49. *Waveforms for example 3.8.*

(f) We know that the frequency spectrum of an amplitude-modulated wave consists of

$$f_c,\ f_c + f_m \text{ and } f_c - f_m$$

$$f_c = 100{,}000 \text{ Hz}$$

$$f_c + f_m = 100{,}000 + 1500 = 101{,}500 \text{ Hz}$$

$$f_c - f_n = 100{,}00 - 1500 = 98{,}500 \text{ Hz}$$

Therefore, the frequency content of the modulated wave will be

100,000 Hz
101,500 Hz
98,500 Hz **Ans.**

Example 3.9. The maximum power efficiency of an AM modulator is

(a) 25% **(b) 50%**
(c) 75% **(d) 100%**

(GATE Examination-1998)

Solution: We know that $P_t = P_c\left(1+\frac{m^2}{2}\right) = \frac{A_c^2}{R}\left(1+\frac{m^2}{2}\right)$

Also, useful power $P_u = \frac{m^2}{2}\frac{A_c^2}{R}$

and $P_c = \frac{A_c^2}{R}$

Hence, efficiency $\eta = \frac{P_u}{P_c} = \frac{m^2}{2}$

But $m \le 1$, therefore $\eta_{max} = \frac{1}{2}$ *i.e.,* 50%

Thus, option (b) is correct.

Example 3.10. A 75 MHz carrier signal having an amplitude of 50 V is modulated by a 3 kHz audio signal having an amplitude of 20 V.

(a) Sketch the audio signal.
(b) Sketch the carrier signal
(c) Construct the modulated wave.
(d) Determine the modulation index and per cent modulation.
(e) What frequencies would be there in a spectrum analysis of the modulated wave?
(f) Write trigonometric equation for the carrier and the modulating waves.

Solution: Given that:

$$f_c = 75 \text{ MHz};\quad V_c = 50 \text{ V};\quad f_m = 3 \text{ kHz};\quad V_m = 2\ V$$

(a) Audio signal or modulating signal is sketched in figure 3.50(a).

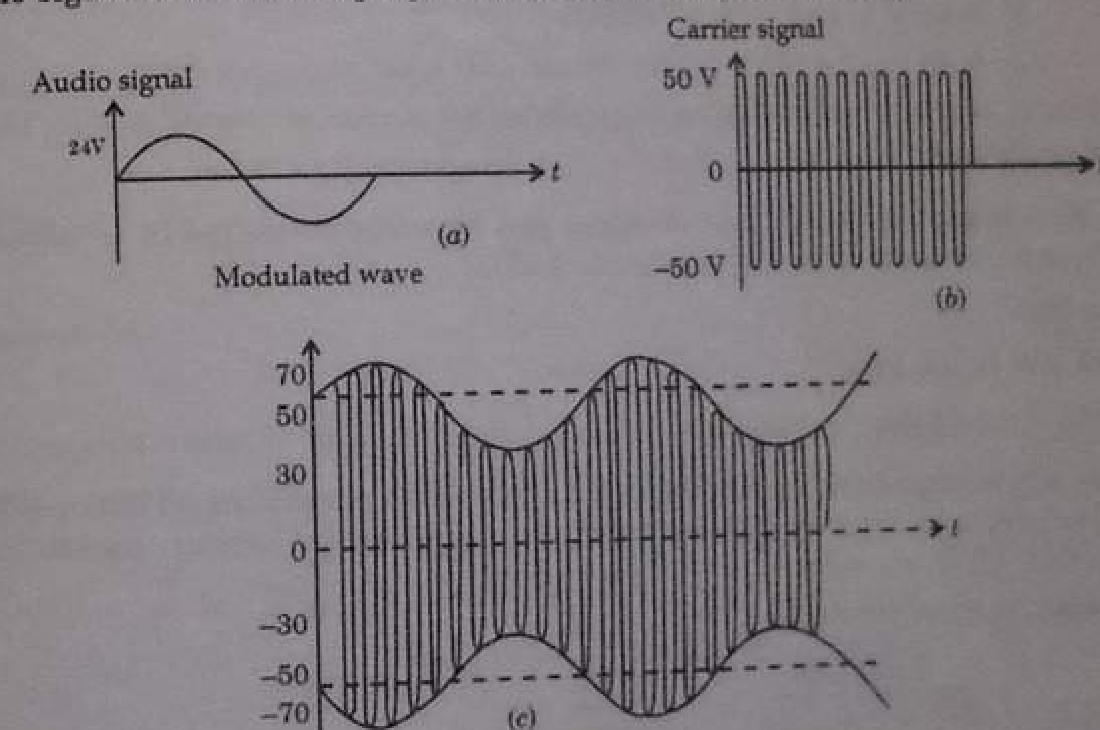


Fig. 3.50.

(b) The carrier signal is sketched in figure 3.50(b)

(c) The envelope of the carrier is fist developed by drawing the horizontal dashed line at the unmodulated carrier amplitude, both positive and negative. The audio signal is now sketched around the dashed line, providing the envelope within which the radio-frequency signal can be laid as shown in figure 3.50(c)

(d) Modulation index is given as

$$m_a = \frac{V_m}{V_c} = \frac{20}{50} = 0.4 \quad \textbf{Ans.}$$

Per cent modulation may now be determined by multiplying the modulation index by 100, *i.e.,*

$$M = m_a \times 100 = 0.4 \times 100 = 40\% \quad \textbf{Ans.}$$

(e) The frequency content of an AM signal consists of the carrier frequency and the sideband frequencies which result from adding the audio frequency to the carrier and from subtracting the audio frequency from the carrier frequency as under:

$$f_c = 75 \text{ MHz}$$
$$f_c + f_m = 75 \text{ MHz} + 3 \text{ kHz}$$
$$= 75{,}000 \text{ kHz} + 3 \text{ kHz} = 75{,}003 \text{ kHz}$$
$$f_c - f_m = 75{,}000 \text{ kHz} - 3 \text{ kHz} = 74{,}997 \text{ kHz}$$

Thus, the frequency content of the AM wave will be

75.000 MHz

75.003 MHz

75.997 MHz **Ans.**

(*f*) The expression for modulating signal is given by

$$v_m = V_m \sin 2\pi f_m t$$

where $V_m = 20$ V and $f_m = 3000$ Hz. **Ans.**

Therefore, $v_m = 20 \sin 2\pi (3000)t = 20 \sin 6000\,\pi t$

The expression for carrier signal is given by

$$v_c = V_c \sin 2\pi f_c t$$

where $V_c = 50$ V and $f_c = 75$ MHz

$$v_c = 50 \sin 2\pi (75 \times 10^6)t = 50 \sin 150 \times 10^6 \pi t \quad \textbf{Ans.}$$

Remember that v_c represents the maximum amplitude of the carrier signal and v_m the maximum amplitude of the modulating signal.

Example 3.11. How many AM broadcast stations can be accommodated in 10 kHz bandwidth if the highest frequency modulating a carrier is 5 kHz?

Solution: Given that:

$$\text{Total BW} = 100 \text{ kHz}$$
$$f_{m_{max}} = 5 \text{ kHz}$$

We know that, any station being modulated by a 5 kHz signal will produce an upper-side frequency 5 kHz above its carrier and a lower-side frequency 5 kHz below its carrier, thereby requiring a bandwidth of 10 kHz. Thus, we have

Number of stations accommodated

$$= \frac{\text{Total BW}}{\text{BW per station}} = \frac{100 \times 10^3}{10 \times 10^3}$$

Hence, number of stations accommodated = 10 stations. **Ans.**

Example 3.12. A bandwidth of 20 MHz is to be considered for the transmission of AM signals. If the highest audio frequencies used to modulate the carriers are not to exceed 3 kHz, how many station could broadcast within this band simultaneously without interfering with one another?

Solution: Given that:

$$\text{Total BW} = 20 \text{ MHz}$$
$$f_{m_{max}} = 3 \text{ kHz}$$

We know that the maximum bandwith of each AM station is determined by the maximum frequency of the modulating signal

$$\text{Station BW} = 2 f_{m_{max}}$$
$$= 2 \times 3 \times 10^3 = 6 \times 10^3 = 6 \text{ kHz}$$

Thus, the number of stations that can be broadcasted simultaneously without interfering with one another will be

$$\frac{20 \times 10^6}{6 \times 10^3} = 3.333 \times 10^3$$

Number of station = 3333 **Ans.**

Example 3.13. The total power content of an AM signal is 1000 W. Determine the power being transmitted at the carrier frequency and at each of the sidebands when the per cent modulation is 100%.

Solution: Given that

$$P_t = 1000 \text{ W}$$
$$m = 100\% = 1$$

We know that the total power consists of the power at the carrier frequency, at the upper sideband frequency and at the lower sideband frequency, *i.e.*,

$$P_t = P_c + P_{USB} + P_{LSB}$$

Form the equation for total power, we have

$$P_t = P_c + \frac{m_a^2 P_c}{4} + \frac{m_a^2 P_c}{4} = P_c + \frac{m^2 P_c}{2}$$

Substituting, given values, we get

$$1000 = P_c + \frac{(1.0)^2 P_c}{2} = P_c + 0.5\,P_c = 1.5 P_c$$

or

$$\frac{1000}{1.5} = P_c$$

Solving, we get

$$P_c = 666.67 \text{ W} \quad \textbf{Ans.}$$

The leaves 1000 – 666.67 = 333.33 watts to be shared equally between upper and lower sidebands.

i.e, $P_{USB} + P_{LSB} = 333.33$ W

But, $P_{USB} = P_{LSB}$

Thus, $2P_{LSB} = 333.33$

or $P_{LSB} = P_{USB} = \frac{333.33}{2} = 166.66$ **Ans.**

Example 3.14. Determine the power content of the carrier and each of the sidebands for an AM signal having a per cent modulation of 80% and a total power of 2500 W.

Solution: Given that

$$M = 80\%; \quad m_a = 0.8$$
$$P_t = 2500 \text{ W}$$

We have to find P_c, P_{USB}, and P_{LSB}.

We know that the total power of an AM signal is the sum of the power at the carrier frequency and the power contained in the sidebands, *i.e.*

$$P_t = P_c + P_{USB} + P_{LSB}$$

Using the equation for total power, we write

$$P_t = P_c + \frac{m_a^2 P_c}{4} + \frac{m_a^2 P_c}{4}$$

But

$$\frac{m_a^2 P_c}{4} + \frac{m_a^2 P_c}{4} = \frac{m_a^2 P_c}{2}$$

So,

$$P_t = P_c + \frac{m_a^2 P_c}{2}$$

$$2500 = P_c + \frac{(0.8)^2 P_c}{2} = P_c + \frac{0.64}{2} P_c = 1.32 P_c = 1.32\ P_c$$

$$P_c = \frac{2500}{1.32} = 1893.9\ \text{W} \quad \textbf{Ans.}$$

The power in the two sidebands is the difference between the total power and the carrier power,

i.e.,

$$P_{USB} + P_{LSB} = 2500 - 1893.9$$

or

$$P_{USB} + P_{LSB} = 606.1\ \text{W}$$

or

$$P_{USB} = P_{LSB} = \frac{606.1}{2}\ \text{W} = 303.50\ \text{W} \quad \textbf{Ans.}$$

Example 3.15. The power content of the carrier of an AM wave is 5 kilowatts. Determine the power content of each of the sidebands and the total power transmitted when the carrier is modulated upto 75%.

Solution: Given that

$$P_c = 5\ \text{kW}$$

$$M = 75\%;\ m_a = 0.75$$

We have to find P_{USB}, P_{LSB} and P_t

Since, in an AM wave, the power in each of the sidebands is equal, therefore, we have

$$P_{USB} = P_{LSB} = \frac{m_a^2 P_c}{4} = \frac{(0.75)^2 (5000)}{4}$$

$$P_{USB} = P_{LSB} = 703.13\ \text{W} \quad \textbf{Ans.}$$

Now, the total power is the sum of the carrier power and the power in the two sidebands.

Thus, we have

$$P_t = P_c + P_{USB} + P_{LSB}$$

or

$$P_t = 5000 + 703.13 + 703.13 = 6406.26\ \text{W} \quad \textbf{Ans.}$$

Example 3.16. Which of the following demodulators can be used for demodulating the signal $x(t) = 5(1 + 2\cos 2000\pi t)\cos 2000\pi t$

(a) Envelope demodulator

(b) Square-law demodulator

(c) Synchronous demodulator

(d) None of the above.

Solution: The given signal is *(GATE Examination-1999)*

$$x(t) = 5(1 + 2\cos 2000\pi t)\cos 2000\pi t$$

Note that this is essentially an amplitude modulation (AM) double-sideband signal which can be demodulated by any of the 3 demodulators (a), (b) or (c), though envelope detection using a square law device (such as a diode) is most commonly used.

Hence, answers (a), (b) and (c) are correct.

Example 3.17. An amplitude-modulated wave has a power content of 800 W at its carrier frequency. Determine the power content of each of the sidebands for a 90% modulation.

Solution: Given that

$$P_c = 800\ \text{W};\quad M = 90\%;\quad m = 0.90$$

We have to find P_{USB}, P_{USB}

Since, the power in each of the sidebands is equal to $\frac{m_a^2 P_c}{4}$, therefore, we get

$$P_{LSB} = P_{USB} = \frac{m_a^2 P_c}{4} = \frac{(0.9)^2\,800}{4}$$

or

$$P_{LSB} = P_{USB} = 162\ \text{W} \quad \textbf{Ans.}$$

Example 3.18. Determine the per cent modulation of an amplitude-modulated wave which has a power content at the carrier of 8 kW and 2 kW in each of its sidebands when the carrier is modulated by a simple audio tone.

Solution: Given that

$$P_c = 8\ \text{kW};\quad P_{USB} = P_{LSB} = 2\ \text{kW}$$

We have to find M.

Knowing the power content of the sidebands and the carrier, the relationship of sideband power can be used to determine the modulation index. Once the modulation index is known, merely multiplying it by 100 provides per cent modulation.

$$P_{USB} = P_{LSB} = \frac{m_a^2 P_c}{4}$$

$$2 \times 10^3 = \frac{m_a^2 (8 \times 10^3)}{4}$$

or

$$m_a^2 = \frac{4 \times 2 \times 10^3}{8 \times 10^3} = 1.0$$

or

$$m_a = 1.0$$

Also,

$$M = m_a \times 100 = 100\% \quad \textbf{Ans.}$$

Example 3.19. The total power content of an AM wave is 600 W. Determine the per cent modulation of the signal if each of the sidebands contains 75 W.

Solution: Given that

$$P_t = 600\ \text{W};\quad P_{USB} = P_{LSB} = 75\ \text{W}$$

We have to find M.

In order to determine the per cent modulation, the power contained at the carrier frequency is first determined. Once P_c is known, the relationship between P_c and the sideband power will provide a means of determining the modulation index, from which the per cent modulation is easily found.

Carrier power can be determined as follows:

$$P_t = P_c + P_{USB} + P_{LSB}$$

$$600 = P_c + 75 + 75$$

or

$$P_c = 6 - 150 = 450\ \text{watts} \quad \textbf{Ans.}$$

Now, using the relationship between sideband power and carrier power, we have

$$P_{USB} = P_{LSB} = \frac{m_a^2 P_c}{4}$$

$$75 = \frac{m_a^2 (450)}{4}$$

or

$$m_a^2 = \frac{4(75)}{450} = 0.667$$

or

$$m_a = 0.816 \quad \textbf{Ans.}$$

Converting modulation index to per cent modulation, we get

$$M = m_a \times 100 = 0.816 \times 100$$

$$M = 81.6\% \quad \textbf{Ans.}$$

or

Example 3.20. Find the per cent modulation of an AM wave whose total power content is 2500 W and whose sidebands each contains 400 W.

Solution: Given that

$$P_t = 2500 \text{ W}; \quad P_{USB} = P_{LSB} = 400 \text{ W}$$

We have to find M.

First find the power contained at the carrier frequency. Then use the relationship between sideband power and carrier power to determine the modulation index. Once the modulation index is known, the per cent modulation can easily be found merely by multiplying by 100.

The power at the carrier frequency can be determined as follows:

$$P_t = P_c + P_{USB} + P_{LSB}$$

$$2500 = P_c + 400 + 400$$

$$P_c = 2500 - 800 = 1700 \text{ W}$$

$$P_{USB} = P_{LSB} = \frac{m_a^2 P_c}{4}$$

$$400 = \frac{m_a^2(1700)}{4}$$

or $$m_a^2 = \frac{400(4)}{1700} = \frac{1600}{1700} = 0.941$$

or $$m_a = 0.970$$

Thus, $$M = 0.970 \times 100 = 97\% \quad \textbf{Ans.}$$

Example 3.21. Determine the power content of each of the sidebands and of the carrier of an AM signal that has a per cent modulation of 85% and contains 1200 W of total power.

Solution: Given that

$$M = 85\%; \quad m_a = 0.85; \quad P_t = 1200 \text{ W}$$

We have to find P_c, P_{USB}, P_{LSB}

Using the expression which relates the total power to carrier power, we have

$$P_t = P_c\left(1 + \frac{m_a^2}{2}\right)$$

Substituting, $$1200 = P_c\left[1 + \frac{(0.85)^2}{2}\right] = P_c\left[1 + \frac{0.7225}{2}\right] = P_c\,[1+0.3613] = 1200 = 1.363P_c$$

or $$P_c = \frac{1200}{1.3613} = 881.5 \text{ W} \quad \textbf{Ans.}$$

The sum of carrier power and sideband power is equal to total power, *i.e.*,

$$P_c + P_{SB} = P_t$$

$$881.5 + P_{SB} = 1200$$

$$P_{SB} = 1200 - 881.5 = 318.5 \text{ watts}$$

The total sideband power is made up equally of upper sideband power and lower sideband power

i.e., $$P_{USB} = P_{LSB} = \frac{P_{SB}}{2} = \frac{318.5}{2} = 159.25 \quad \textbf{Ans.}$$

Example 3.22. An AM signal in which the carrier is modulated upto 70% contains 1500 W at the carrier frequency. Determine the power content of the upper and lower sidebands for this per cent modulation. Calculate the power at the carrier and the power content of each of the sidebands when the per cent modulation drops to 50%.

Solution: Given that

$$M = 70\%; \quad m_a = 0.70; \quad P_{c70} = 1500 \text{ W}$$

We have to find P_{USB70}, P_{LSB70}, P_{c50}, P_{USB50}, P_{LSB50}

We know that the power content of each of the sidebands is equal to $m_a^2 P_c/4$, thus, we have

$$P_{USB70} = P_{LSB70} = \frac{m_a^2 P_c}{4} = \frac{(0.7)^2 1500}{4} = 183.75$$

$$P_{USB70} = P_{LSB70} = 183.75 \text{ W} \quad \textbf{Ans.}$$

Since in standard AM transmission, carrier power remains the same, regardless of per cent modulation, thus, we have

$$P_{c50} = P_{c70} = 1500 \text{ W}$$

$$P_{USB50} = P_{LSB50} = \frac{m_a^2 P_c}{4} = \frac{(0.5)^2 1500}{4}$$

$$P_{c50} = 93.75 \text{ W} \quad \textbf{Ans.}$$

Example 3.23. The per cent modulation of an AM wave changes from 40% to 60%. Originally, the power content at the carrier frequency was 900 W. Determine the power content at the carrier frequency and within each of the sidebands after the per cent modulation has risen to 60%.

Solution: Given that

$$M_1 = 40\%; \quad m_1 = 0.40; \quad M_2 = 60\%; \quad m_2 = 0.60$$

$$= P_{c40} = 900 \text{ watts}$$

We have to find P_{c60}, P_{USB60}, P_{LSB60}.

The power content of the carrier of an AM signal remains the same regardless of per cent modulation. Thus, we have

$$P_{c60} = P_{c40} = 900 \text{ W}$$

The power content of each of the sidebands is equal to $m_a^2 P_c/4$.

Thus, we have $$P_{USB60} = P_{LSB40} = \frac{m^2 P_c}{4} = \frac{(0.60)^2(900)}{4}$$

$$P_{USB60} = P_{LSB60} = 81.0 \text{W} \quad \textbf{Ans.}$$

Example 3.24. A signal sideband (SSB) signal contains 1 kW. How much power is contained in the sidebands and how much at the carrier frequency?

Solution: Given that

$$P_{SSB} = 1 \text{ kW}$$

We have to find P_{SB}, P_c

We know that in a signal sideband transmission, the carrier and one of the two sidebands are eliminated. Therefore, all the transmitted power is transmitted at one of the sidebands regardless of per cent modulation.

Thus, $$P_{SB} = 1 \text{ kW}$$

and $$P_c = 0 \text{ W} \quad \textbf{Ans.}$$

Example 3.25. An SSB transmission contains 10 kW. This transmission is to be replaced by a standard amplitude modulated signal with the same power content. Determine the power content of the carrier and each of the sidebands when the per cent modulation is 80%.

Solution: Given that

$$P_{SSB} = 10 \text{ kW}; \quad M = 80\%; \; m_a = 0.80$$

We have to find P_c, P_{LSB}, P_{USB}

Since the total power content of the new AM signal is to be the same as the total power content of the SSB signal, therefore, we have

$$P_t = P_{SSB} = 10 \text{ kW}$$

Solving for power contained at the carrier frequency, we get

$$P_t = P_c + P_{LSB} + P_{USB} = P_c + \frac{m_a^2 P_c}{4} + \frac{m_a^2 P_c}{4}$$

$$10{,}000 = P_c + \frac{(0.8)^2 P_c}{4} + \frac{(0.8)^2 P_c}{4} = P_c + \frac{0.64 P_c}{2} = 1.32\, P_c$$

$$\frac{10{,}000}{1.32} = P_c$$

$$P_c = 7575.76 \text{ W} \quad \textbf{Ans.}$$

The power content of the sidebands is equal to the difference between the total power and the carrier power, *i.e.*,

$$P_{SB} = P_t - P_c$$

The power content of the upper and the lower sidebands is equal, *i.e.*,

$$P_{LSB} + P_{USB} = 10{,}000 - 7575.76 = 242.24$$

$$P_{LSB} = P_{LSB} = \frac{2424.24}{2} = 1212.12 \text{ W}$$

Thus,
$$P_c = 7575.76 \text{ W}$$
$$P_{LSB} = P_{USB} = 1212.12 \text{ W} \quad \textbf{Ans.}$$

Example 3.26. Determine the modulation index and per cent modulation of the signal shown in figure 3.51.

Solution: Given that an AM signal which is shown in figure 3.51.

We have to find m_a and M

Using the equation relating maximum peak-to-peak amplitude and minimum peak-to-peak amplitude to modulation index, we have

$$m = \frac{\max p\text{-}p - \min p\text{-}p}{\max p\text{-}p + \min p\text{-}p}$$

From, figure 3.51, we get

$$\max p\text{-}p = 2(80) = 160$$
$$\min p\text{-}p = 2(20) = 40$$

Thus,
$$m_a = \frac{160-40}{160+40} = \frac{120}{200} = 0.6$$
$$m_a = 0.6$$

Hence,
$$M = m_a \times 100 = 0.6 \times 100$$
$$M = 60\% \quad \textbf{Ans.}$$

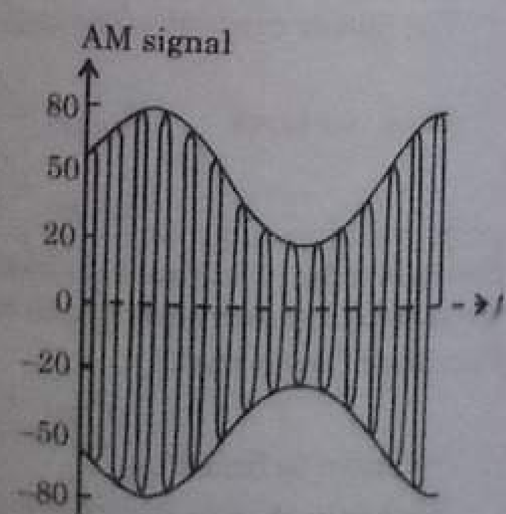


Fig. 3.51.

Example 3.27. Find the modulation index and the per cent modulation of the signal shown in figure 3.52.

Solution: Given that AM signal as shown in figure 3.52.

We have to find m_a and M

Using
$$m_a = \frac{\max p\text{-}p - \min p\text{-}p}{\max p\text{-}p + \min p\text{-}p}$$

and values from figure 3.52, we get

$$\max p\text{-}p = 2(50) = 100$$
$$\min p\text{-}p = 2(15) = 30$$

Thus,
$$m_a = \frac{100-30}{100+30} = \frac{70}{130} = 0.538$$

Hence,
$$M = m_a \times 100 = 53.8\% \quad \textbf{Ans.}$$

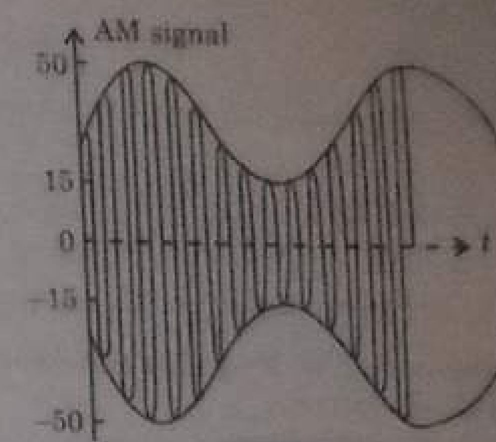


Fig. 3.52.

Example 3.28. Verify that the message signal $m(t)$ is recovered from a modulated DSB signal by first multiplying it by a local sinusoidal carrier and then passing the resultant signal through a low-pass filter (LPF), as shown in figure 3.53 (*a*) in the time domain and (*b*) in the frequency domain.

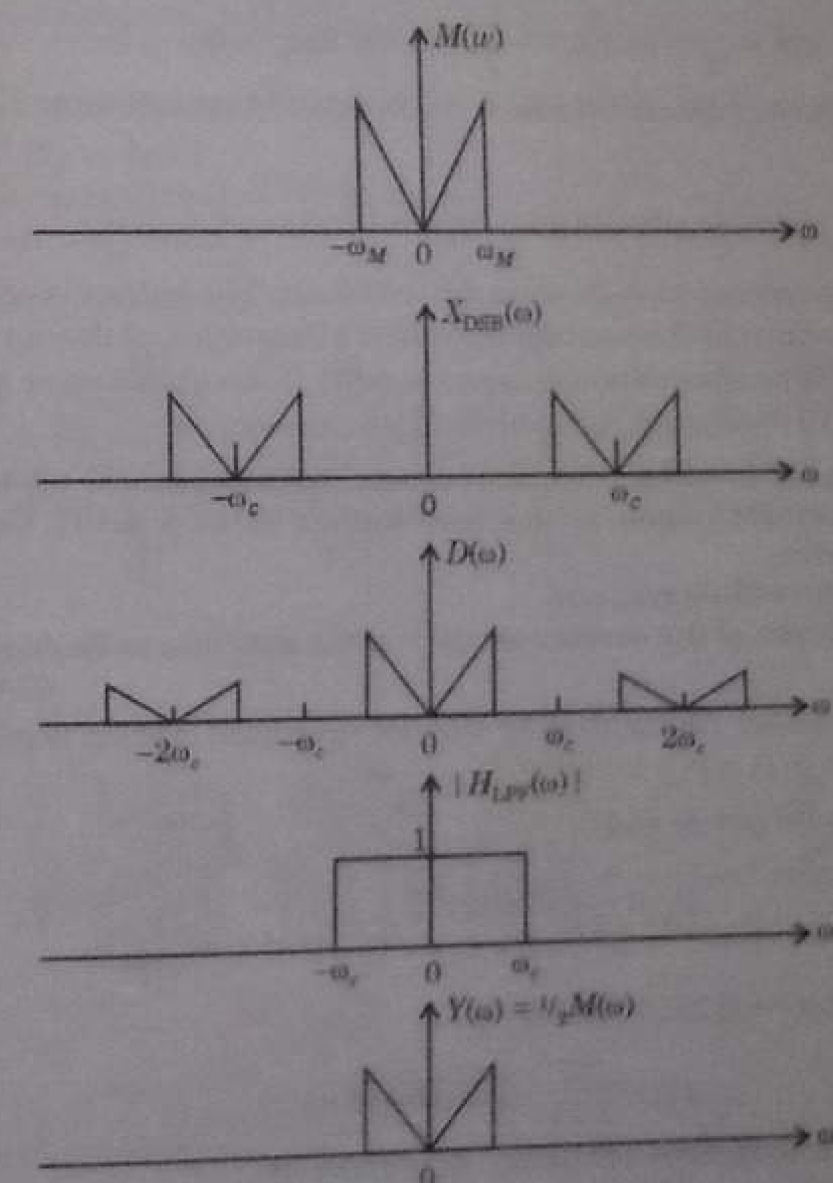


Fig. 3.53. *Demodulation of a DSB signal.*

Solution: The output of the multiplier is given by

$$d(t) = x_{DSB}(t) \cos w_c t = [m(t) \cos w_c t] \cos w_c t$$

$$d(t) = m(t)\cos^2 \omega_c t$$

or

$$d(t) = \frac{1}{2}m(t) + \frac{1}{2}m(t)\cos 2\omega_c t$$

or

After low-pass filtering of signal $d(t)$, we get

$$y(t) = \frac{1}{2}m(t)$$

Thus, by proper amplification (multiplying by 2), we can recover the message signal $m(t)$.

(b) Demodulation of signal $x_{DSB}(t)$ in frequency domain has been illustrated in figure 3.53.

Example 3.29. Evaluate the effect of a phase error in the local oscillator on synchronous DSB demodulation.

Solution: Let the phase error of the local oscillator be ϕ. Then the local carrier is expressed as $\cos(\omega_c t + \phi)$. Now, we have

$$x_{DSB}(t) = m(t)\cos \omega_c t$$

and

$$d(t) = [m(t)\cos \omega_c t]\cos(\omega_c t + \phi)$$

$$= \frac{1}{2}m(t)[\cos\phi + \cos(2\omega_c t + \phi)]$$

or

$$d(t) = \frac{1}{2}m(t)\cos\phi + \frac{1}{2}m(t)\cos(2\omega_c t + \phi)$$

The second term on the right-hand side is filtered out by the low-pass filter (LPF), and thus we obtain

$$y(t) = \frac{1}{2}m(t)\cos\phi$$

This output is proportional to $m(t)$ when ϕ is constant. The output is completely lost when $\phi = \pm\pi/2$. Thus, the phase error in the local carrier cause attenuation of the output signal without any distortion as long as ϕ is constant and not equal to $\pm\pi/2$. If the phase error ϕ varies randomly with time, then the output also will vary randomly and is undersirable.

Example 3.30. A given AM broadcast station transmits a total power of 50 kW when the carrier is modulated by a sinusoidal signal with a modulation index of 0.707. Calculate:

(i) the carrier power

(ii) the transmission efficiency, and

(iii) the peak amplitude of the carrier assuming the antenna to be represented by a (50 + j0) Ω load.

(GATE Examination, 1998)

Solution: (i) The total power transmitted by the AM broadcast station is given by

$$P_c(1 + P_x) = 50 \text{ kW}$$

where P_c is the carrier power and

$$P_x = \frac{1}{2}\times(0.707)^2 = 0.25$$

Thus, we write

$$P_c(1 + 0.25) = 50 \text{ kW}$$

or

$$P_c = \frac{50}{1.25} = 40 \text{ kW} \quad \textbf{Ans.}$$

(ii) We know that transmission efficiency, η, is given by

$$\eta = \frac{P_c P_x}{P_c + P_c P_x} = \frac{P_x}{1 + P_x}$$

or

$$\eta = \frac{0.25}{1.25} = 0.2 \text{ or } 20\%$$

(iii) Also, carrier power, $P_c = \frac{A_c^2}{2\times 50} = 40\times 10^3$ W

Therefore, peak carrier amplitude,

$$A_c = \sqrt{2\times 50\times 40\times 10^3} = 2000 \text{ Volts or } 2 \text{ kV} \quad \textbf{Ans.}$$

Example 3.31. Evaluate the effect of a small frequency error in the local oscillator on synchronous DSB demodulation.

Solution: Let the frequency error of the local oscillator be $\Delta\omega$. The local carrier is then expressed as $\cos(\omega_c + \Delta\omega)t$.

Then, we have

$$d(t) = m(t)\cos\omega_c t \cos(\omega_c + \Delta\omega)t$$

or

$$d(t) = \frac{1}{2}m(t)\cos(\Delta\omega)t + \frac{1}{2}m(t)\cos 2\omega_c t$$

Thus,

$$y(t) = \frac{1}{2}m(t)\cos(\Delta\omega)t$$

Note: *The output is the signal m (t) multiplied by a low-frequency sinsuoid. This is a* ***"beating"*** *effect and is a very undersirable distortion.*

Example 3.32. Sketch the ordinary AM signal for a single-tone modulation with modulation indices of $m_a = 0.5$ and $m_a = 1$.

Solution: For a single-tone modulation, we have

$$m(t) = a_m \cos\omega_m t$$

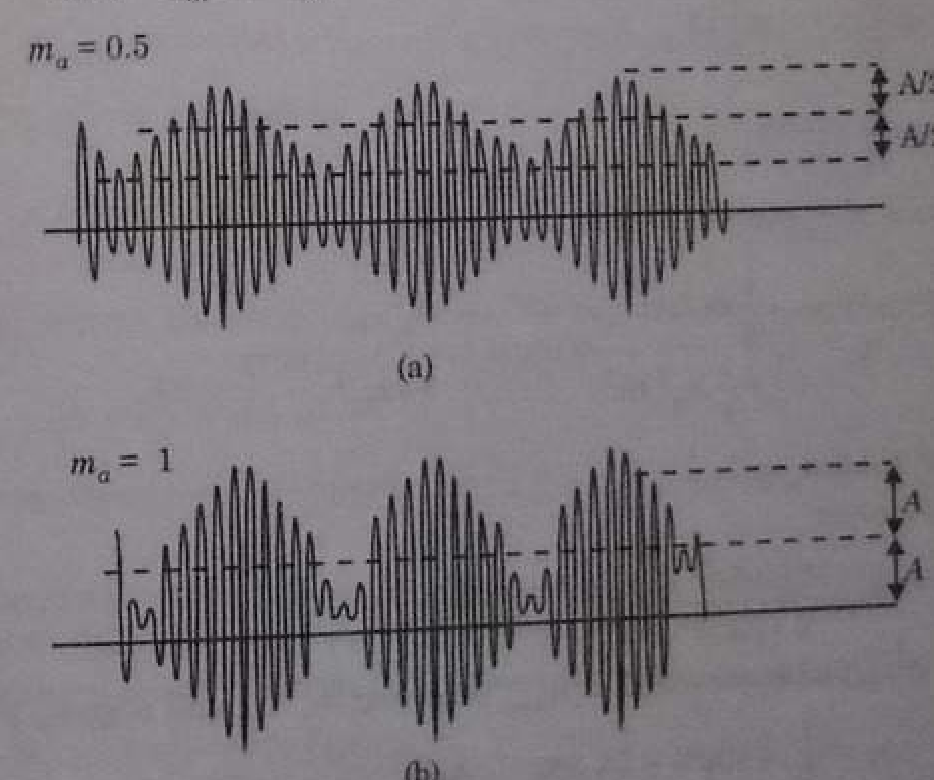


Fig. 3.54.

Then, the modulation index is

$$m_a = \frac{|\max\{m(t)\}|}{A} = \frac{a_m}{A}$$

Hence, $m(t) = a_m\cos\omega_m t = m_a A\cos\omega_m t$

Also, $x_{AM}(t) = [A + m(t)]\cos\omega_c t = A[1 + m_a\cos\omega_m t]\cos\omega_c t$

Figure 3.54(a) and (b) show the ordinary AM signal corresponding to $m_a = 0.5$ and $m_a = 1$, respectively.

Example 3.33. The efficiency η of ordinary AM is defined as the per centage of the total power carried by the sidebands, that is,

$$\eta = \frac{P_s}{P_t} \times 100\%$$

where P_s is the power carried by the sidebands and P_t is the total power of the AM signal.

(i) Find η for $m_a = 0.5$ (50 per cent modulation).

(ii) Show that for a single-tone AM, η_{max} is 33.3 percent at $m_a = 1$.

Solution: A single-tone AM signal can be expressed as under:

$$x_{AM}(t) = A\cos\omega_c t + m_a A\cos\omega_m t\cos\omega_c t$$

$$= A\cos\omega_c t + \frac{1}{2} m_a A\cos(\omega_c - \omega_m)t + \frac{1}{2} m_a A\cos(\omega_c + \omega_m)t$$

P_c = carrier power = $\frac{1}{2}A^2$

P_s = sideband power

$$P_s = \frac{1}{2}\left[\left(\frac{1}{2}m_a A\right)^2 + \left(\frac{1}{2}m_a A\right)^2\right] = \frac{1}{4}m_a^2 A^2 \quad ...(i)$$

The total power P_t is given by

$$P_t = P_c + P_s = \frac{1}{2}A^2 + \frac{1}{4}m_a^2 A^2$$

$$P_t = \frac{1}{2}\left(1 + \frac{1}{2}m_a^2\right)A^2 \quad ...(ii)$$

Thus $$\eta = \frac{P_s}{P_t} \times 100\% \quad ...(iii)$$

Substituting values of P_s and P_t in equation (iii), we get

or $$\eta = \frac{\frac{1}{4}m_a^2 A^2}{\left(\frac{1}{2} + \frac{1}{4}m_a^2\right)A^2} \times 100\% = \frac{m_a^2}{2 + m_a^2} 100\%$$

with the condition that $m_a \le 1$.

(i) For $m_a = 0.5$, we have

$$\eta = \frac{(0.5)^2}{2 + (0.5)^2} \times 100\% = 11.1\%$$

(ii) Since $m_a \le 1$, it can be observed that η_{max} occurs at $m_a = 1$ and is given by

$$\eta = \frac{1}{3} \times 100\% = 33.3\% \quad \textbf{Ans.}$$

Example 3.34. Show that a synchronous demodulator shown in figure 3.55(a) can demodulate an AM signal $x_{AM}(t) = [A + m(t)]\cos\omega_c t$ regardless of the value of A.

From figure 3.55(a), we have

$$d(t) = x_{AM}(t)\cos\omega_c t$$

But given that

$$x_{AM}(t) = [A + m(t)]\cos\omega_c t$$

Therefore, we write

$$d(t) = [A + m_a(t)]\cos^2\omega_c t$$

Applying $\cos^2\theta = \frac{1}{2}[1 + \cos 2\theta]$

$$d(t) = \frac{1}{2}[A + m(t)] + \frac{1}{2}[A + m(t)]\cos^2\omega_c t$$

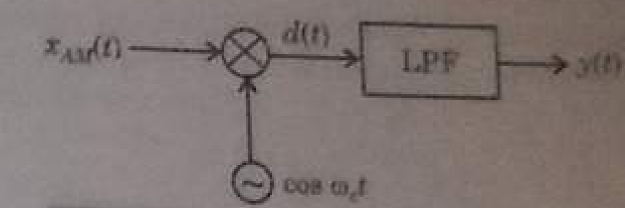


Fig. 3.55. (a) Synchronous detector for Example 3.34.

Now, this signal is passed through a low-pass filter (LPF).

Hence, after low-pass filtering (LPF), we obtain

$$y(t) = \frac{1}{2}[A + m(t)] = \frac{1}{2}m(t) + \frac{1}{2}A$$

A blocking capacitor will suppress the direct-current (dc) term $\frac{1}{2}$ A, yielding the output $\frac{1}{2}$ m(t).

Hence Proved

Example 3.35. Show that an AM signal with large carrier can be demodulated by squaring it and then passing the resulting signal through a low-pass filter (LPF), as shown in figure 3.55(b). This type of detector is known as a square-law detector.

We know that AM signal $x_{AM}(t)$ is given by

$$x_{AM}(t) = [A + m(t)]\cos\omega_c t$$

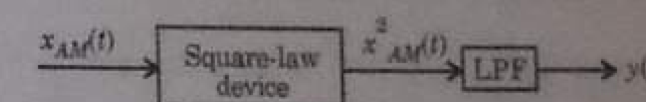


Fig. 3.55. (b) A square-law detector.

Squaring the above signal, we get

$$x^2_{AM}(t) = [A + m(t)]^2\cos^2\omega_c t$$

Simplifying, we get

$$x^2_{AM}(t) = \frac{1}{2}[A^2 + 2.A.m(t) + m^2(t)](1 + \cos 2\omega_c t)$$

The low-pass filter output $y(t)$ will be

$$y(t) = \frac{A^2}{2}\left\{1 + 2\frac{m(t)}{A} + \left[\frac{m(t)}{A}\right]^2\right\}$$

Now, with a large carrier, the term $[m(t)/A]^2$ can be neglected and we shall have

$$y(t) = \frac{A^2}{2} + Am(t)$$

Further, a blocking capacitor will suppress the d.c. term $A^2/2$, yielding the output $Am(t)$.

Hence Proved

Example 3.36. The input to an envelope detector is a single-tone AM signal $x_{AM}(t) = A(1 + m_a\cos\omega_m t)\cos\omega_c t$, where m_a is a constant, $0 < m_a < 1$, and $\omega_c >> \omega_m$.

(i) Show that if the detector output is to follow the envelope of $x_{AM}(t)$, it is required that at any time t_0

$$\frac{1}{RC} \ge \omega_m\left(\frac{m_a\sin\omega_m t_0}{1 + m_a\cos\omega_m t_0}\right) \quad ...(i)$$

(ii) Show that if the detector output is to follow the envelope at all times, it is required that

$$RC \le \frac{1}{\omega_m}\frac{\sqrt{1 - m_a^2}}{m_a} \quad ...(ii)$$

Solution: (i) Figure 3.56 shows the envelope of $x_{AM}(t)$ and the output of the detector (*i.e.*, the voltage across the capacitor). Let us assume that the capacitor discharges from the peak value $E_0 = A(1 + m_a \cos \omega_m t_0)$ at $t_0 = 0$. Then the voltage $v_c(t)$ across the capacitor can be given by

$$v_c(t) = E_0 e^{-t/RC} \qquad ...(iii)$$

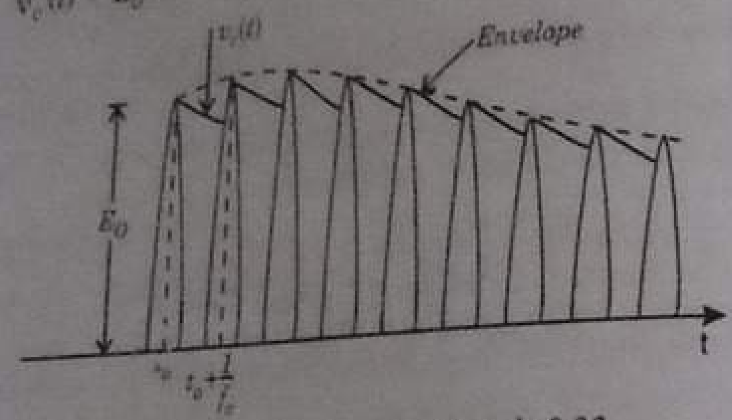


Fig. 3.56. *Waveform for example 3.36.*

The interval between two successive carrier peaks is $1/f_c = 2\pi/\omega_c$ and $RC >> 1/\omega_c$. This means that the time constant RC is much larger than the interval between two successive carrier peaks.

Therefore, $v_c(t)$ can be approximated as

$$v_c(t) \cong E_0\left(1-\frac{t}{RC}\right) \qquad ...(iv)$$

Thus, if the voltage $v_c(t)$ is to follow the envelope of $x_{AM}(t)$, it is required that at any time t_0

$$(1 + m_a \cos \omega_m t_0)\left(1-\frac{1}{RCf_c}\right) \le 1 + m_a \cos\omega_m\left(t_0+\frac{1}{f_c}\right)$$

Now if $\omega_m << \omega_c$, then, we have

$$1+m_a \cos \omega_m\left(t_0+\frac{1}{f_c}\right) = 1 + m_a \cos\left(\omega_m t_0+\frac{\omega_m}{f_c}\right)$$

$$= 1 + m_a \cos \omega_m t_0 \frac{\omega_m}{f_c} - m_a \sin \omega_m t_0 \sin \frac{\omega_m}{f_c}$$

or $$1+m_a \cos\omega_m\left(t_0+\frac{1}{f_c}\right) \cong 1 + m_a \cos \omega_m t_0 - m_a \frac{\omega_m}{f_c} \sin \omega_m t_0$$

Hence $$(1 + m_a \cos \omega_m t_0)\left(\frac{1}{RCf_c}\right) \ge \frac{m_a\omega_m}{f_c} \sin\omega_m t_0$$

or $$\frac{1}{RC} \ge \omega_m\left(\frac{m_a \sin\omega_m t_0}{1+m_a \cos\omega_m t_0}\right)$$

Hence Proved

(ii) Now, rewriting equation (i), we have

$$\frac{1}{RC}+\frac{m_a}{RC}\cos\omega_m t_0 \ge m_a\omega_m \sin\omega_m t_0$$

or $$m_a\left(\omega_m \sin\omega_m t_0 - \frac{1}{RC}\cos\omega_m t_0\right) \le \frac{1}{RC}$$

or $$m_a\sqrt{\omega_m^2+\left(\frac{1}{RC}\right)^2}\sin\left(\omega_m t_0 - \tan^{-1}\frac{1}{\omega_m RC}\right) \le \frac{1}{RC}$$

Since this inequality must hold for every t_0, we must have

$$m_a\sqrt{\omega_m^2+\left(\frac{1}{RC}\right)^2} \le \frac{1}{RC}$$

or $$m_a^2\left[\omega_m^2+\left(\frac{1}{RC}\right)^2\right] \le \left(\frac{1}{RC}\right)^2$$

From which we obtain

$$RC \le \frac{1}{\omega_m}\frac{\sqrt{1-m_a^2}}{m_a}$$

Hence Proved

Example 3.37. Using the single-tone modulating signal $\cos \omega_m t$, verify that the output of the SSB generator is indeed an SSB signal, and show that an upper-sideband (USB) or a lower-sideband (LSB) signal results from subraction or addition at the summation junction.

Solution: Referring to figure 3.57, we have

$$m(t) = \cos \omega_m t$$

$$\cos\left(\omega_c t - \frac{\pi}{2}\right) = \sin \omega_c t$$

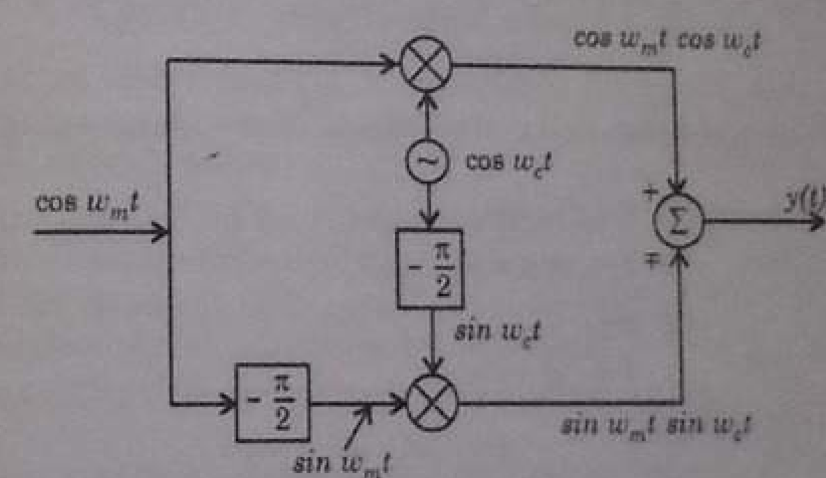


Fig. 3.57. *Block diagram for example 3.37.*

Therefore, $$\hat{m}(t) = \cos\left(\omega_m t - \frac{\pi}{2}\right) = \sin \omega_m t$$

Hence, we have $$y(t) = \cos\omega_m t \cos\omega_c t \mp \sin\omega_m t \sin\omega_c t = \cos(\omega_c \pm \omega_m)t$$

Thus, with subtraction, we have

$$y(t) = x_{USB}(t) = \cos(\omega_c + \omega_m)t$$

and with addition, we shall have

$$y(t) = x_{LSB}(t) = \cos(\omega_c - \omega_m)t$$

Hence Proved

Example 3.38. Show that if the output of a phase-shift modulator is an SSB signal, (i) the difference of the signals at the summing junction produces the upper-sideband SSB signal and

(ii) the sum produces the lower-sideband SSB signal. That is,

$$x_c(t) = x_{USB}(t) = m(t) \cos w_c t - \hat{m}(t) \sin w_c t \qquad ...(i)$$

is an upper-sideband SSB signal, and

$$x_c(t) = x_{LSB}(t) = m(t) \cos w_c t + \hat{m}(t) \sin w_c t \qquad ...(ii)$$

is a lower-sideband SSB signal.

Solution: (*i*) Let $m(t) \leftrightarrow M(w)$

and $\hat{m}(t) \leftrightarrow \hat{M}(w)$

Now, let us apply the modulation theorem or the frequency-shifting property of Fourier transform, we have

$$m(t)\cos w_c t \leftrightarrow \frac{1}{2} M(w-w_c) + \frac{1}{2} M(w+w_c)$$

and

$$\hat{m}(t)\sin w_c t \leftrightarrow \frac{1}{2j} \hat{M}(w-w_c) - \frac{1}{2j} \hat{M}(w+w_c)$$

Taking the Fourier transform of equation (i), we have

$$X_c(w) = \frac{1}{2} M(w-w_c) + \frac{1}{2} M(w+w) - \left[\frac{1}{2j}\hat{M}(w-w_c) - \frac{1}{2j}\hat{M}(w+w_c)\right]$$

we also have

$$\hat{M}(w-w_c) = -j\, sgn\,(w-w_c)\, M(w-w_c)$$

and $$\hat{M}(w+w_c) = -j\, sgn\,(w+w_c)\, M(w+w_c)$$

Thus $$X_c(w) = \frac{1}{2} M(w-w_c) + \frac{1}{2} M(w+w_c) - \left[-\frac{1}{2} sgn(w-w_c) M(w-w_c) + \frac{1}{2} sgn(w+w_c) M(w+w_c)\right] \quad ...(iii)$$

or $$X_c(w) = \frac{1}{2} M(w-w_c)[1+sgn(w-w_c)] + \frac{1}{2} M(w+w_c)[1-sgn(w+w_c)]$$

Since $$1 + sgn\,(w-w_c) = \begin{cases} 2 & \text{for } w > w_c \\ 0 & \text{for } w < w_c \end{cases}$$

and $$1 - sgn\,(w-w_c) = \begin{cases} 2 & \text{for } w < -w_c \\ 0 & \text{for } w > -w_c \end{cases}$$

we have $$X_c(w) = \begin{cases} 0 & \text{for } |w| < w_c \\ M(w+w_c) & \text{for } w < -w_c \\ M(w-w_c) & \text{for } w > w_c \end{cases} \quad ...(iv)$$

which has been sketched in figure 3.58(*b*). We see that $x_c(t)$ is an upper sideband SSB signal.

Hence Proved

(*ii*) In a similar manner, taking the Fourier transform of equation (*ii*), we have

$$X_c(w) = \frac{1}{2} M(w-w_c)[1 - sgn\,(w-w_c)] + \frac{1}{2} M(w+w_c)[1 + sgn\,(w+w_c)] \quad ...(v)$$

Now, since $$1 - sgn\,(w-w_c) = \begin{cases} 2 & \text{for } w < w_c \\ 0 & \text{for } w > w_c \end{cases}$$

and $$1 + sgn(w+w_c) = \begin{cases} 2 & \text{for } w > -w_c \\ 0 & \text{for } w < -w_c \end{cases}$$

we have $$X_c(w) = \begin{cases} 0 & \text{for } |w| > w_c \\ M(w-w_c) & \text{for } w < w_c \\ M(w+w_c) & \text{for } w > -w_c \end{cases} \quad ...(vi)$$

ch has been sketched in figure 3.58(*c*). We see that $x_c(t)$ is a lower-sideband SSB signal.

Hence Proved

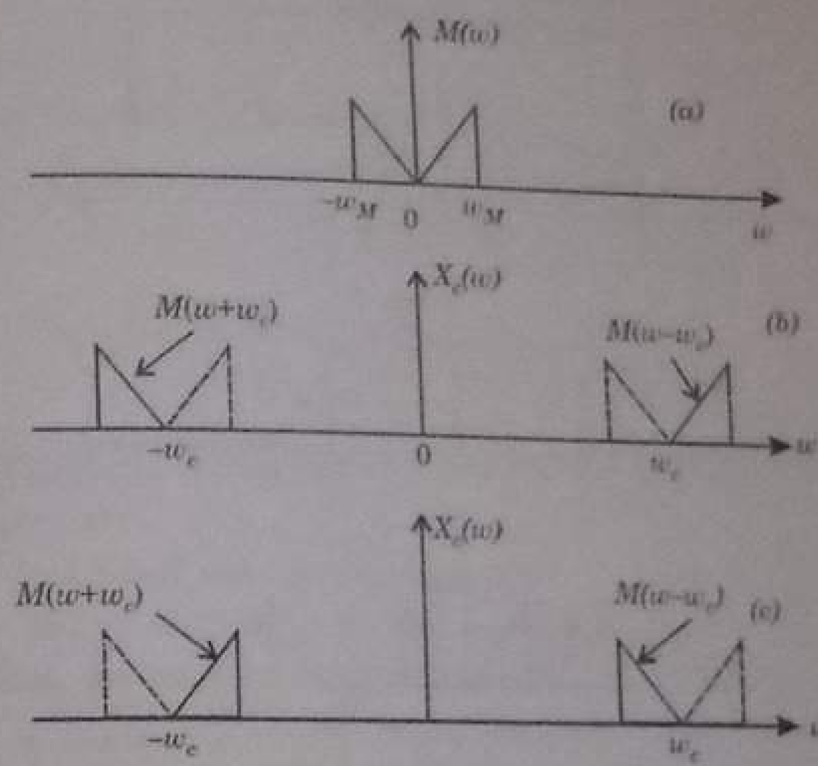


Fig. 3.58. *Waveforms for example 3.38.*

Example 3.39. Show that an SSB signal can be demodulated by the synchronous detector of figure 3.59 by sketching the spectrum of the signal at each point and (*ii*) by the time-domain expression of the signals at each point.

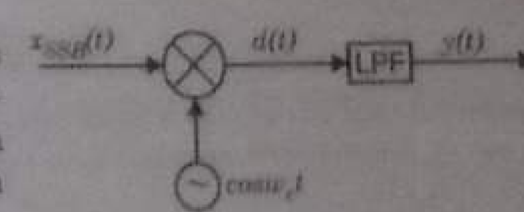


Fig. 3.59. *Synchronous detector for example 3.39.*

Solution: (*i*) Let $M(w)$, the spectrum of the message $m(t)$, be as shown in figure 3.60(*a*). Also let us assume that $x_{SSB}(t)$ is a lower-sideband SSB signal and its spectrum is $X_{SSB}(\omega)$, as shown in figure 3.60(*b*). Multiplication by cos $\omega_c t$ shifts the spectrum $X_{SSB}(\omega)$ to $\pm\omega_c$ and we obtain $D(\omega)$, the spectrum of $d(t)$ as shown in figure 3.60 (*d*) After low-pass filtering, we obtain $Y(w) = \frac{1}{2} M(w)$, the spectrum of $y(t)$ as shown in figure 3.60(*c*).

Thus, we obtain $y(t) = \frac{1}{2} m(t)$ which is proportional to message signal $m(t)$. **Hence Proved.**

(*ii*) We know that $x_{SSB}(t)$ can be expressed as under: ...(*i*)

$$x_{SSB}(t) = m(t)\cos w_c t \mp \hat{m}(t)\sin w_c t$$

Thus, $$d(t) = x_{SSB}(t)\cos w_c t$$

Using equation (*i*), we get

$$d(t) = m(t)\cos^2 w_c t \mp \hat{m}(t)\sin w_c t \cos w_c t$$

$$= \frac{1}{2} m(t)(1 + \cos 2w_c t) \mp \frac{1}{2}\hat{m}(t)\sin 2w_c t$$

or $$d(t) = \frac{1}{2} m(t) + \frac{1}{2} m(t)\cos 2w_c t \mp \frac{1}{2}\hat{m}(t)\sin 2w_c t$$

Hence, after low-pass filtering, we obtain

$$y(t) = \frac{1}{2} m(t)$$

Hence Proved.

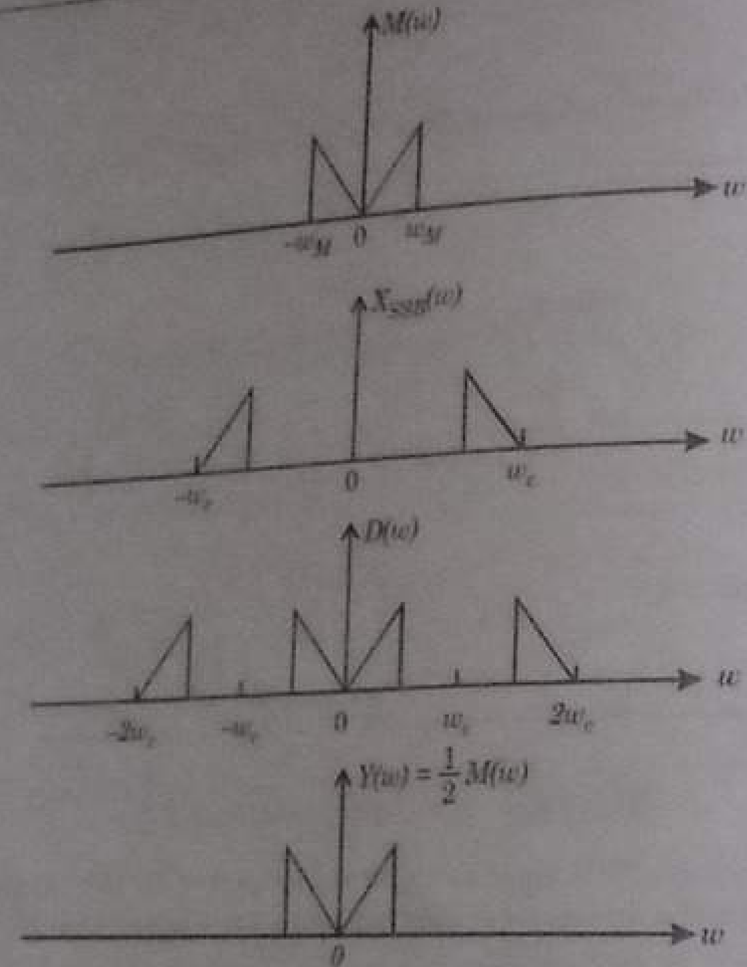


Fig. 3.60. *Waveforms for example 3.39.*

Example 3.40. A carrier signal of 1.0 Volt amplitude and a sinusoidal modulating signal of 0.5 V, put in series, are applied to a square law modulator of characteristics,

$$i'_o = 10 + kV_i + k'V_i^2 \text{ mA}$$

where V_i is input in volts,

$$k = 2 \text{ mA/V and } k' = 0.2 \text{ mA/V}^2$$

considering only the frequency components of the AM signal corresponding to the carrier frequency, find the depth of modulation in the resulting AM signal. *(GATE Examination-1999)*

Solution: We have

$$V_i(t) = \cos(w_c t) + 0.5\cos(w_m t)$$

and

$$i_o = 10 + kV_i + k'V_i^2$$

or

$$i_o = 10 + 2\times10^{-3}[\cos(w_c t) + 0.5\cos(w_m t)] + 0.2\times10^{-3}[\cos(w_c t) + 0.5\cos(w_m t)]^2$$

Further, considering only the carrier terms, we have

$$i_o = 2\times10^{-3}\cos(w_c t) + \frac{0.2\times10^{-3}}{0.5}\times 0.5\cos(w_c t)\cos(w_m t)$$

Now, modulation depth will be

$$m = \frac{0.2\times10^{-3}}{0.5} = 0.4\times10^{-3}$$ **Ans.**

Example 3.41. Show that the system shown in figure 3.61 can be used to demodulate an SSB signal.

Solution: We know that the upper-sideband SSB signal $x_c(t)$ is given by

$$x_c(t) = m(t)\cos w_c t - \hat{m}(t)\sin w_c t$$

we also have,

$$\hat{x}_c(t) = m(t)\sin w_c t + \hat{m}(t)\cos w_c t$$

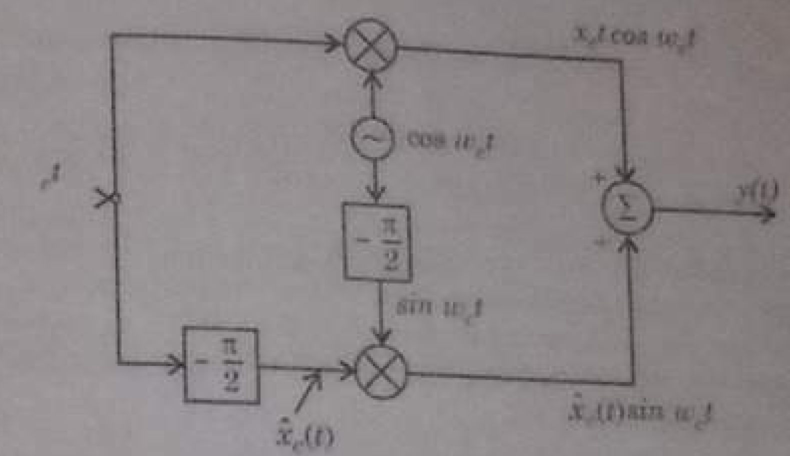


Fig. 3.61. *Phase-shift SSB demodulator for example 3.41.*

From figure 3.61, we have

$$y(t) = x_c(t)\cos w_c t + \hat{x}_c(t)\sin w_c t$$

or

$$y(t) = m(t)(\cos^2 w_c t + \sin^2 w_c t) = m(t)$$

In a similar manner, the lower-sideband SSB signal $x_c(t)$ is given by

$$x_c(t) = m(t)\cos w_c t + \hat{m}(t)\sin w_c t$$

and

$$\hat{x}_c(t) = m(t)\sin w_c t - \hat{m}(t)\cos w_c t$$

Again, from figure 3.63, we obtain

$$y(t) = x_c(t)\cos w_c t + \hat{x}_c(t)\sin w_c t$$
$$= m(t)(\cos^2 w_c t + \sin^2 w_c t) = m(t)$$

Hence Proved

Example 3.42. Determine the per centage power saving when the carrier wave and one of the sidebands are suppressed in an AM wave modulated to a depth of (*i*) 100% and (*ii*) 50%.

Solution: (*i*) Total power in *AM* wave is given by

$$P_t = P_c\left(1 + \frac{m^2}{2}\right) \quad ...(i)$$

Given that modulation index m = 100% = 1.0

Equation (*i*) can be written as

$$P_t = P_c\left(1 + \frac{(1)^2}{2}\right) = 1.5P_c \quad ...(ii)$$

The power in one sideband is given by

$$P_{SB} = P_c\frac{m^2}{4} = P_c\frac{(1)^2}{4} = \frac{P_c}{4} = 0.25P_c \quad ...(iii)$$

$$\text{Power saving} = \frac{P_t - P_{SB}}{P_t} = \frac{1.5P_c - 0.25P_c}{1.5P_c} = \frac{1.25}{1.50} = \frac{5}{6} = 0.833 = 83.3\%$$

(*ii*) Given that $m = 50\% = 0.50$ and $P_t = P_c\left(1 + \frac{m^2}{2}\right)$

Putting the values of m, we have

$$P_t = P_c\left(1 + \frac{(0.5)^2}{2}\right) = 1.125\,P_c \quad ...(iv)$$

$$P_{SB} = P_c \frac{m^2}{4} = P_c \frac{(0.5)^2}{4} = 0.0625\, P_c \qquad ...(v)$$

Hence, power saving $= \frac{P_t - P_{SB}}{P_t} = \frac{1.125P_c - 0.0625P_c}{1.125P_c} = 0.944 = 94.4\%$ **Ans.**

Example 3.43. An AM signal $s(t) = [1 + 0.2\cos\{2\pi(f_m/3)t\}]\cos 2\pi f_c t$ is detected using a square-law demodulator. This square-law modulator has the characteristic $v_0(t) = [s(t) + 2]^2$. The output is then filtered by an ideal low-pass filter (LPF) with a cut-off frequency at f_m Hertz. Obtain the frequency spectrum of the demodulated signal. Also plot the frequency spectrum in the frequency range $-f_m \le f \le f_m$.

Solution: The output of the demodulator is given as

$$v_0(t) = [s(t) + 2]^2 = s^2(t) + 4 + 4s(t) \qquad ...(i)$$

here $s(t)$ is input to the modulator.

Given an AM modulated signal as under :

$$s(t) = \left[1 + 0.2\cos\left\{2\pi\left(\frac{f_m}{3}\right)t\right\}\right]\cos(2\pi f_c t) \qquad ...(ii)$$

Substituting equation (ii) in equation (i), we get

$$v_0(t) = s^2(t) + 4 + 4s(t)$$

or $$v_0(t) = \left\{\left[1 + 0.2\cos\left\{2\pi\left(\frac{f_m}{3}\right)t\right\}\right]\cos(2\pi f_c t)\right\}^2 + 4\left[1 + 0.2\cos\left\{2\pi\left(\frac{f_m}{3}\right)t\right\}\right]\cos(2\pi f_c t)$$

or $$v_0(t) = \left[1 + 0.2\cos\left\{2\pi\left(\frac{f_m}{3}\right)t\right\}\right]^2\cos^2(2\pi f_c t) + 4\left[1 + 0.2\cos\left\{2\pi\left(\frac{f_m}{3}\right)t\right\}\right]\cos(2\pi f_c t) \qquad ...(iii)$$

Simplifying last equation, we get

$$v_0(t) = 4.51 + 0.2\cos\left[2\pi\left(\frac{f_m}{3}\right)t\right] + 0.01\cos\left[2\pi\left(\frac{2f_m}{3}\right)t\right] + 1.02\cos(4\pi f_c t)$$
$$+ 0.2\cos\left[2\pi\left(2f_c + \frac{f_m}{3}\right)t\right]\left[2\pi\left(2f_c - \frac{f_m}{3}\right)t\right] + 0.01\cos\left[2\pi\left(2f_c + \frac{2f_m}{3}\right)t\right]$$
$$+ 0.01\cos\left[2\pi\left(2f_c - \frac{2f_m}{3}\right)t\right] + 4\cos(2\pi f_c t) + 0.4\cos\left[2\pi\left(f_c + \frac{f_m}{3}\right)t\right] + 0.4\cos\left[2\pi\left(f_c - \frac{f_m}{3}\right)t\right] \qquad ...(iv)$$

Also, the frequency spectrum of the signal $v_0(t)$ can be obtained by taking Fourier transform of both sides of equation (iv), as under:

$$V_o(f) = 4.51\,\delta(f) + 0.1\left[\delta\left(f - \frac{f_m}{3}\right) + \delta\left(f + \frac{f_m}{3}\right)\right] + 0.005\left[\delta\left(f - \frac{2f_m}{3}\right) + \delta\left(f + \frac{2f_m}{3}\right)\right]$$
$$+ 0.51\,[\delta(f - 2f_c) + \delta(f + 2f_c)] + 0.01\left[\delta\left(f - 2f_c - \frac{f_m}{3}\right) + \delta\left(f + 2f_c + \frac{f_m}{3}\right)\right] + 0.005$$
$$+ 0.01\left[\delta\left(f - 2f_c - \frac{2f_m}{3}\right) + \delta\left(f + 2f_c + \frac{2f_m}{3}\right)\right] + 0.005\left[\delta\left(f - 2f_c + \frac{2f_m}{3}\right) + \delta\left(f + 2f_c - \frac{2f_m}{3}\right)\right]$$
$$+ 2[\delta(f - f_c) + \delta(f + f_c)] + 0.2\left[\delta\left(f - f_c - \frac{f_m}{3}\right) + \delta\left(f + f_c + \frac{f_m}{3}\right)\right] + 0.2\left[\delta\left(f - f_c + \frac{f_m}{3}\right) + \delta\left(f + f_c - \frac{f_m}{3}\right)\right]$$

The spectrum of the low-pass filter LPF output has been shown in figure 3.62.

The low-pass filter with a cut-off frequency equal to f_m, will only allow the d.c. component and the components having frequencies $f_m/3$ and $2f_m/3$ and reject all other frequency components present in the output $V_o(f)$ of the square-law modulator.

Hence, the spectrum of the filter output is given by

$$S_{LPF}(f) = 4.51\,\delta(f) + 0.1\left[\delta\left(f - \frac{f_m}{3}\right) + \delta\left(f + \frac{f_m}{3}\right)\right]$$
$$+ 0.005\left[\delta\left(f - \frac{2f_m}{3}\right) + \delta\left(f + \frac{2f_m}{3}\right)\right] \quad \textbf{Ans.}$$

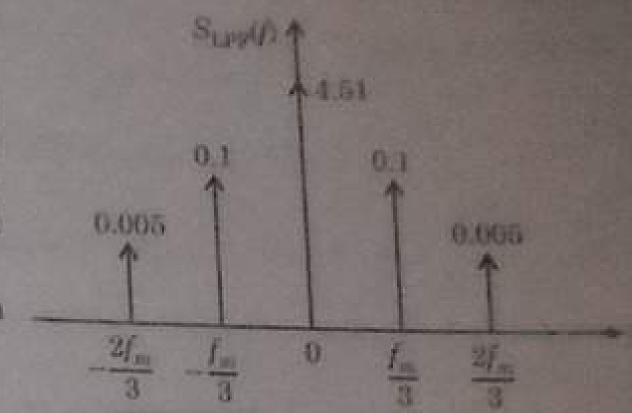


Fig. 3.62. *Showing spectrum of the LPF output.*

Example 3.44. The amplitude modulated (AM) waveform $s(t) = [1 + m\cos(2\pi f_m t)]\cos(2\pi f_c t)$ is to be demodulated by an envelope detector. Find the maximum permissible value of the modulation index m for which such demodulation is possible. Also sketch the output of the envelope detector for $m = 2$.

Solution: The envelope of the AM signal $s(t) = [1 + m\cos(2\pi f_m t)]\cos(2\pi f_c t)$ corresponds to the modulating, *i.e.*, message signal so long as

$$|m\cos(2\pi f_m t)| < 1 \qquad ...(i)$$

This means that the per centage of modulation is less than 100%. Hence, it is possible to recover the message signal $x(t)$ from AM signal $s(t)$ by an envelope detection so long as per centage modulation is less than 100%.

For modulation index, $m = 2$, the AM wave or signal will appear like one shown in figure 3.63(*a*). The corresponding demodulated signal is shown in figure 3.63(*b*) Here, it can be noted that the demodulator cannot detect the modulating signal $x(t)$ properly since it can follow only the positive cycles.

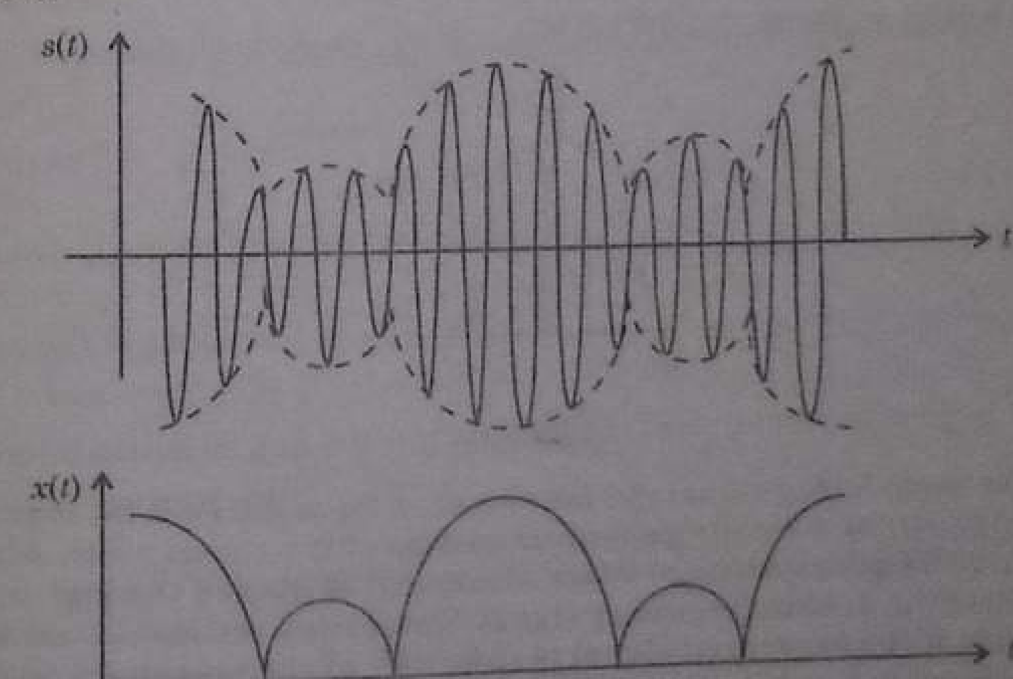


Fig. 3.63. *(a) AM signal for m = 2, (b) Corresponding demodulated signal.*

Example 3.45. A signal $v(t) = [1 + m(t)]\cos(\omega_c t)$ is detected using a square law detector, having the characteristic $v_o = v^2$. If the Fourier transform of $m(t)$ is constant, M_o, extending from $-f_m$ to $+f_m$, sketch the Fourier transform of $v_o(t)$ in the frequency range $-f_m < f < f_m$.

(GATE Examination-1996)

Solution: Given

$$v(t) = [1 + m(t)] \cos w_c t$$

$$v_o = v^2 = [1 + m^2(t) + 2m(t)] \cos^2 w_c t = \frac{1}{2}[1 + m^2 + 2m][1 + \cos 2 w_c t]$$

$$v_o = \frac{1}{2}[1 + m^2 + 2m + \cos 2w_c t + m^2 \cos 2w_c t + 2m \cos 2w_c t]$$

or $$v_o(t) = \frac{1}{2}[1 + m^2 + 2m + \cos 2 w_c t + m^2 \cos 2 w_c t + 2m \cos 2w_c t]$$

Also, $$V_o(f) = \frac{1}{2}[\delta(f) + M(f) \otimes M(f) + 2M(f) + \frac{\partial(f - 2f_c)}{2} + \frac{\partial(f + 2f_c)}{2}$$

$$\frac{M_c(f - 2f_c)}{2} + \frac{M_c^2(f + 2f_c)}{2} + M(f - 2f_c) + M(f + 2f_c)$$

where $M_c(f) = M(f) \otimes M(f)$

In the frequency range, $-f_m \le f \le f_m$, we have

$$V_m(f) = \frac{1}{2}[\delta(f) + M(f) \otimes M(f) + 2M(f)]$$

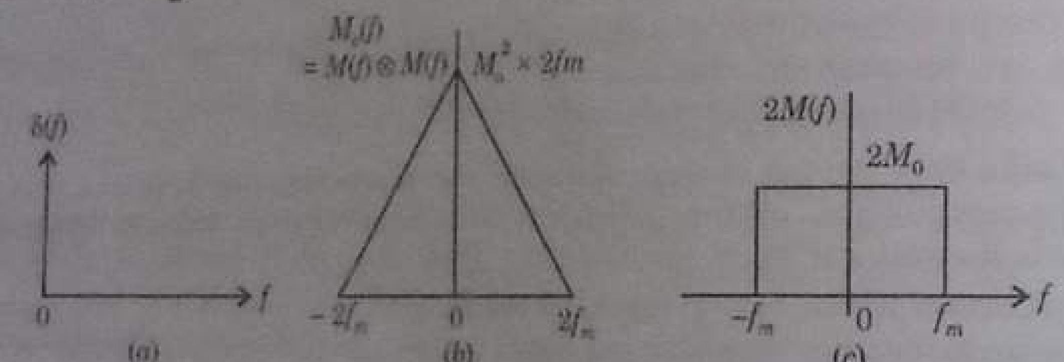


Fig. 3.64.

Hence $V_m(f)$ will be as shown in figure 3.65.

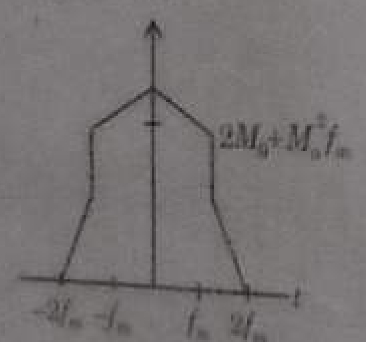


Fig. 3.65.

Exmaple 3.46. The modulating signal $x(t)$ is recovered from the DSB-SC signal $s(t) = x(t) \cos(\omega_c t)$ by multiplying $s(t)$ by a locally generated carrier $c'(t) = \cos(\omega_c t + \phi)$, where $\omega_c = 2\pi f_c$, is the angular carrier frequency. The product of $s(t)\, c'(t)$ is passed through a low-pass filter (LPF) which rejects the double frequency signal. Determine the maximum allowable value for the phase angle if the recovered signal is to be 95% of the maximum possible output. If the modulating signal is band limited to 10 kHz, determine the minimum value of carrier frequency f_c for which $x(t)$ can be recovered by filtering.

Solution: The product signal is given by

$$v(t) = s(t) \cos(2\pi f_c t + \phi)$$

or $$v(t) = x(t) \cos(2\pi f_c t) \cos(2\pi f_c t + \phi)$$

or $$v(t) = x(t)\left[\frac{\cos(4\pi f_c t + \phi) + \cos\phi}{2}\right]$$

or $$v(t) = \frac{x(t)}{2}\cos(4\pi f_c t + \phi) + \frac{x(t)}{2}\cos\phi$$

The low pass filter (LPF) will reject the first term.

The maximum allowable value of phase angle ϕ can be found as under:

$$\cos(\phi_{max}) = \frac{\left[\frac{x(t)}{2}\cos\phi\right]}{\max\left[\frac{x(t)}{2}\cos\phi\right]} = 0.95$$

i.e., $$\phi_{max} = \cos^{-1}(0.95) = 18°.19$$

In order to recover $x(t)$ from $v(t)$ using filter method, it is essential that the lowest frequency contained in the first term of $v(t)$ must be greater than the highest frequency contained in the second term, *i.e.*,

$$2f_c - 10 \text{ kHz} > 10 \text{ kHz}$$

or $$f_c > 10 \text{ kHz}$$

Hence, the minimum value of f_c will be

$$f_c = 10 \text{ kHz} \quad \textbf{Ans.}$$

Example 3.47. (*i*) The modulating signal $x(t)$ is to be recovered from the DSB-SC signal given by $s(t) = x(t)\cos(2\pi f_c t)$. The available local carrier is a periodic signal $p(t)$ with a time period $1/f_c$. Prove that the modulating signal $x(t)$ can be recovered from the product $v(t) = s(t)\, p(t)$ by using appropriate filter.

(*ii*) Show that $x(t)$ can be recovered as well if the periodic waveform has a period n/f_c, where n is an integer.

(*iii*) If modulating signal $x(t)$ is band-limited to the frequency range from 0 to 5 kHz, find the largest value of n which will allow $x(t)$ to recover.

Given f_c = 1 MHz.

Solution : (*i*) The periodic waveform can be expanded in exponential Fourier series as under:

$$p(t) = \sum_{n=-\infty}^{\infty} C_n e^{(j2\pi n f_c t)} \quad ...(i)$$

where C_n's are complex Fourier coefficient.

Thus, $$C_{-n} = C_n^*$$

The product $v(t)$ is expressed as ...(*ii*)

$$v(t) = s(t)\, p(t)$$

Substituting equation (*i*) and $s(t) = x(t)\cos(2\pi f_c t)$ in equation (*ii*), we get ...(*iii*)

$$v(t) = p(t)\, v(t)$$

or $$v(t) = \sum_{n=-\infty}^{\infty} C_n e^{(j2\pi n f_c t)} x(t)\cos(2\pi f_c t) = \sum_{n=-\infty}^{\infty} C_n x(t)\cos(2\pi f_c t)\, e^{(j2\pi n f_c t)}$$

or $$v(t) = x(t)\sum_{n=-\infty}^{\infty}\left[C_n\left\{\frac{e^{j2\pi f_c t} + e^{-j2\pi n f_c t}}{2}\right\} e^{(j2\pi n f_c t)}\right]$$

or $$v(t) = \frac{x(t)}{2}\sum_{n=-\infty}^{\infty} C_n e^{[j2\pi(n+1)f_c t]} + \frac{x(t)}{2}\sum_{n=-\infty}^{\infty} C_n e^{[j2\pi(n-1)f_c t]} \quad ...(iv)$$

The signal $x(t)$ can be recovered from product $v(t)$ by passing it through a low-pass filter (LPF) which rejects double frequency signal. The output of the filter is then given by

$$v_0(t) = \frac{x(t)}{2}(C_1 + C_{-1}) = \frac{x(t)}{2}[2R_e(C_1)]$$

or
$$v_0(t) = [R_e(C_1)]\, x(t) \quad ...(i)$$

(ii) Substituting f_c/n in place of f_c in Fourier series expansion of $p(t)$, we obtain the product signal $v(t)$ as

$$v(t) = \frac{x(t)}{2}\sum_{n=-\infty}^{\infty} C_n e^{\left[j2\pi\left(\frac{1}{n}+1\right)f_c t\right]} + \sum_{n=-\infty}^{\infty} C_n e^{\left[j2\pi\left(\frac{1}{n}-1\right)f_c t\right]}$$

The original modulating signal $x(t)$ can be recovered from $v(t)$ only when the largest frequency contained in $x(t)$ is less than f_c/n.

Under this condition, we have

$$v_0(t) = [R_e(C_n)]\, x(t)$$

(iii) For the given band-limited modulating signal $x(t)$, highest frequency component contained will be

$$f_m = 5 \text{ kHz}$$

The carrier frequency $f_c = 1$ MHz (given)

The largest value of n can be found from

$$\frac{f_c}{n_{max}} = f_m$$

i.e.,
$$n_{max} = \frac{f_c}{f_m} = \frac{1 \times 10^6 \text{ Hz}}{5 \times 10^3 \text{ kHz}} = 200 \quad \textbf{Ans.}$$

Example 3.48. (i) Prove that the signal

$$s(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t + \phi_i) - \sin(\omega_c t)\sin(\omega_i t + \phi_i)]$$

is an SSB signal ($f_c >> f_N$), where $\omega_c = 2\pi f_c$, carrier angular frequency and $\omega_i = 2\pi f_i$. Identify the sideband.

(ii) Obtain an expression for missing sideband.

(iii) Obtain an expression for the total DSB-SC signal.

Solution: (i) The given signal can be modified using trigonometric expression as

$$s(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t + \phi_i) - \sin(\omega_c t)\sin(\omega_i t + \phi_i)]$$

or
$$s(t) = \sum_{i=1}^{N}\left[\frac{1}{2}\cos\{(\omega_c + \omega_i)t + \phi_i\} + \frac{1}{2}\cos\{(\omega_c - \omega_i)t - \phi_i\}\right] - \sum_{i=1}^{N}\left[\frac{1}{2}\cos\{(\omega_c - \omega_i)t - \phi_i\} + \frac{1}{2}\cos\{(\omega_c + \omega_i)t + \phi_i\}\right]$$

or
$$s(t) = \sum_{i=1}^{N}\cos\{(\omega_c + \omega_i)t + \phi_i\} = s_U(t) \quad \text{[Let us say]} \quad ...(i)$$

Hence, $s(t)$ contains frequency components higher than the carrier frequency f_c and thus SSB signal with upper sideband (USB).

(ii) The lower sideband (LSB) will be expressed as

$$s(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t + \phi_i) + \sin(\omega_c t)\sin(\omega_i t + \phi_i)] \quad ...(ii)$$

Equation (ii) can be modified using trigonometric expression and we obtain

$$s(t) = \sum_{i=1}^{N}\cos\{(\omega_c - \omega_i)t - \phi_i\} = s_L(t) \quad \text{[Let us say]} \quad ...(iii)$$

(iii) The total DSB-SC signal can be written as under:

$$s_{\text{DSB-SC}}(t) = s_U(t) + s_L(t)$$

Substituting equations (i) and (iii) in equation (iv), we get

$$s_{\text{DSB-SC}}(t) = \sum_{i=1}^{N}\cos\{(\omega_c + \omega_i)t + \phi_i\} + \sum_{i=1}^{N}\cos\{(\omega_c - \omega_i)t - \phi_i\}$$

or
$$s_{\text{DSB-SC}}(t) = \sum_{i=1}^{N}\cos\{(w_c + w_i)t + \phi_i\} + \cos\{(\omega_c - \omega_i)t - \phi_i\}$$

On modifying last expression, we get

$$s_{\text{DSB-SC}}(t) = 2\sum_{i=1}^{N}\cos(\omega_c t)\cos(\omega_i t + \phi_i)$$

or
$$s_{\text{DSB-SC}}(t) = \left[2\sum_{i=1}^{N}\cos(\omega_i t + \phi_i)\right]\cos(\omega_c t) \quad \textbf{Ans.}$$

Exmaple 3.49. In SSB-SC signal generation using phase discrimination method, the carrier phase shift network produces a phase shift which differs from $\pi/2$ by a small angle α. Obtain the output waveform and also verify that the output no longer corresponds to SSB-SC signal. The modulating signal $x(t)$ may be considered to be a single tone sinusoidal signal 1.0 cos ($2\pi f_m t$).

Solution: Given that single-tone sinusoidal signal is

$$x(t) = 1.0\cos(2\pi f_m t) \quad \text{and} \quad \hat{x}(t) = 1.0\sin(2\pi f_m t + \alpha) \quad ...(i)$$

The output of the SSB generating system will be

$$s(t) = x(t)\cos(2\pi f_c t) \pm \hat{x}(t)\sin(2\pi f_c t + \alpha) \quad ...(ii)$$

Substituting equation (i) in equation (ii), we get

$$s(t) = 1.0\cos(2\pi f_m t)\cos(2\pi f_c t) \pm 1.0\sin(2\pi f_m t + \alpha)\sin(2\pi f_c t + \alpha)$$

or
$$s(t) = \cos(2\pi f_m t)\cos(2\pi f_c t) \pm \sin(2\pi f_m t + \alpha)\sin(2\pi f_c t + \alpha) \quad ...(iii)$$

Modifying equation (iii), we get

$$s(t) = \frac{1}{2}[\cos\{2\pi(f_c + f_m)t\} + \cos\{2\pi(f_c - f_m)t\}] \pm \frac{1}{2}[\cos\{2\pi(f_c - f_m)t - \alpha\} - \cos\{2\pi(f_c + f_m)t + \alpha\}]$$

or
$$s(t) = \frac{1}{2}[\cos\{2\pi(f_c + f_m)t\} \mp \cos\{2\pi(f_c + f_m)t + \alpha\}] + [\cos\{2\pi(f_c + f_m)t\} \pm \cos\{2\pi(f_c - f_m)t - \alpha\}]$$

Hence, we can say that the output waveform is no longer an SSB-SC waveform.

Hence Proved

Example 3.50. A synchronous detection of SSB-SC signal shows phase and frequency discrepancy. Consider the synchronous detection of SSB signal given by

$$s(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t+\phi_i) - \sin(\omega_c t)\sin(\omega_i t+\phi_i)]$$

The signal is multiplied by the locally generated carrier $\cos \omega_c t$ and then low-pass filtered.

(i) Prove that the modulating signal can be completely recovered if the cut-off frequency of the filter is $f_N < f_o < 2f_c$.

(ii) Determine the recovered signal when the multiplying signal is $\cos[\omega_c t + \phi]$.

(iii) Determine the recovered signal when the multiplying signal is $\cos[(\omega_c + \Delta\omega)\, t]$.

Given $\Delta f << f_i$ where $\omega_c = 2\pi f_c$, $\Delta\omega = 2\pi\Delta f$

Solution: (i) The product signal is

$$v(t) = s(t)\cos(\omega_c t) \quad ...(i)$$

Substituting value of $s(t)$ in equation (i), we get

$$v(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t+\phi_i) - \sin(\omega_c t)\sin(\omega_i t+\phi_i)]\cos(\omega_c t)$$

$$= \sum_{i=1}^{N}\cos[(\omega_c + \omega_i)t + \phi_i]\cos(\omega_c t) \quad ...(ii)$$

Modifying equation (ii), we get

$$s(t) = \sum_{i=1}^{N}\left[\frac{1}{2}\cos\{(2\omega_c + \omega_i)t + \phi_i\} + \frac{1}{2}\cos(\omega_i t + \phi_i)\right]$$

or
$$s(t) = \frac{1}{2}\sum_{i=1}^{N}\cos[(\omega_c + \omega_i)t + \phi_i] + \frac{1}{2}\sum_{i=1}^{N}\cos(\omega_i t + \phi_i) \quad ...(iii)$$

The output waveform of the low-pass filter (LPF) is thus given by

$$v_0(t) = \frac{1}{2}\sum_{i=1}^{N}\cos(\omega_i t + \phi_i) \quad ...(iv)$$

when the cut-off frequency, f_o of the LPF is in the range $f_N < f_o < 2f_c$.

(ii) when the multiplying signal is $\cos(\omega_c t + \phi)$, the product signal will be given by

$$v(t) = s(t)\cos(\omega_c t + \phi) \quad ...(v)$$

Substituting the value of $s(t)$ in equation (v), we get

$$v(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t+\phi_i) - \sin(\omega_c t)\sin(\omega_i t+\phi)]\cos(\omega_c t + \phi)$$

or
$$v(t) = \sum_{i=1}^{N}\cos\left[(\omega_c + \omega_i)\phi_i\right]\cos(\omega_c t + \phi) \quad ...(vi)$$

Modifying equation (vi), we get

$$v(t) = \sum_{i=1}^{N}\left[\frac{1}{2}\cos\{(2\omega_c + \omega_i)t + (\phi_i + \phi)\} + \frac{1}{2}\cos\{\omega_i t + (\phi_i - \phi)\}\right]$$

$$v(t) = \frac{1}{2}\sum_{i=1}^{N}\left[\frac{1}{2}\cos\{(2\omega_c + \omega_i)t + (\phi_i + \phi)\} + \frac{1}{2}\sum_{i=1}^{N}\cos\{\omega_i t + (\phi_i - \phi)\}\right] \quad ...(vii)$$

The output of the low-pass filter (LPF) will be

$$v_0(t) = \frac{1}{2}\sum_{i=1}^{N}\cos[\omega_i t + (\phi_i - \phi)] \quad ...(viii)$$

Therefore, the demodulated signal shows phase discrepancy.

(iii) when we multiply $s(t)$ by $\cos[(\omega_c + \Delta\omega)\, t]$ then the product signal will be expressed as

$$v(t) = s(t)\cos[(\omega_c + \Delta\omega)t] \quad ...(ix)$$

Substituting the value of $s(t)$ in equation (ix), we get

$$v(t) = s(t)\cos[(\omega_c + \Delta\omega)\, t]$$

or
$$v(t) = \sum_{i=1}^{N}[\cos(\omega_c t)\cos(\omega_i t+\phi_i) - \sin(\omega_c t)\sin(\omega_i t+\phi_i)]\cos[(\omega_c + \Delta\omega)t]$$

or
$$v(t) = \sum_{i=1}^{N}\cos[(\omega_c + \omega_i)\, t + \phi_i]\cos[(\omega_c + \Delta\omega)\, t] \quad ...(x)$$

On modifying equation (x), we get

$$v(t) = \sum_{i=1}^{N}\left[\frac{1}{2}\cos\{(2\omega_c + \omega_i + \Delta\omega)\, t + \phi_i\} + \frac{1}{2}\cos\{(\omega_i - \Delta\omega)\, t\}\right]$$

or
$$v(t) = \frac{1}{2}\sum_{i=1}^{N}\cos\{(2\omega_c + \omega_i + \Delta\omega)\, t + \phi_i\} + \frac{1}{2}\sum_{i=1}^{N}\cos\{(\omega_i - \Delta\omega)\, t\}$$

The output of the low-pass filter (LPF) will then be given by

$$v_0(t) = \frac{1}{2}\sum_{i=1}^{N}\cos\{(\omega_i - \Delta\omega)\, t\}$$

Hence, the demodulated signal shows frequency error. **Ans.**

Example 3.51. A received single-tone sinusoidally modulated SSB-SC signal $\cos\{(\omega_c + \omega_m)t\}$ has a normalised power of 0.5 volt^2. The signal is to be detected by carrier reinsertion technique. Find the amplitude of the carrier to be reinserted so that the power in the recovered signal at the demodulator output is 90% of the normalised power. The d.c. component can be neglected and $\omega_c = 2\pi f_c$ and $\omega_n = 2\pi f_m$.

Solution: For a single-tone SSB-SC signal, the waveform after carrier insertion becomes

$$s'(t) = s(t) + c(t)$$

or
$$s'(t) = \cos\{(\omega_c + \omega_m)t\} + A\cos(\omega_c t)$$

Using trigonometric relation, we get

$$s'(t) = \cos(\omega_c t)\cos(\omega_m t) - \sin(\omega_c t)\sin(\omega_m t) + A\cos(\omega_c t)$$

or
$$s'(t) = [A + \cos(\omega_m t)]\cos(\omega_c t) - \sin(\omega_c t)\sin(\omega_m t) \quad ...(i)$$

The output of the demodulator is given by

$$v(t) = \sqrt{[A + \cos(\omega_m t)]^2 + [\sin(\omega_m t)]^2}$$

$$v(t) = \sqrt{A^2 + \cos^2(\omega_m t) + 2A\cos(\omega_m t) + \sin^2(\omega_m t)}$$

or $$v(t) = \sqrt{A^2 + 1 + 2A\cos(\omega_m t)} = [A^2 + 1 + 2A\cos(\omega_m t)]^{1/2}$$

or $$v(t) = \sqrt{A^2+1}\left[1 + \frac{2A}{A^2+1}\cos(\omega_m t)\right]^{1/2} \cong \sqrt{A^2+1}\left[1 + \frac{A}{A^2+1}\cos(\omega_m t)\right]$$

or $$v(t) \cong \sqrt{A^2+1} + \frac{A}{\sqrt{A^2+1}}\cos(\omega_m t) \quad ...(ii)$$

Neglecting the d.c. component, the normalised power of the detected signal will be

$$P_d = \frac{1}{2}\left(\frac{A^2}{A^2+1}\right)$$

For the recovered signal at the demodulator output to be 90% of this normalised power, we must have

Detector output = 90% of normalised power of SSB modulated waveform

i.e., $P_d = 90\% \times P_n$

Substituting values, we get

$$\frac{1}{2}\left(\frac{A^2}{A^2+1}\right) = 0.90 \times 0.5 = 0.45$$

or $$A = 0.9\,(A^2 + 1)$$

or $$0.9\,A^2 - A^2 + 1 = 0$$

or $$-0.1\,A^2 = -1$$

or $$A^2 = 10 \quad \text{or} \quad A \cong 3$$

Thus, amplitude of the reinserted carrier must be almost 3 volt. **Ans.**

Example 3.52. A sinusoidally modulated ordinary AM waveform is shown in figure 3.66.

(i) Determine modulation index.

(ii) Compute the efficiency.

(iii) Also, find the amplitude of the carrier which must be added to attain a modulation index of 0.1.

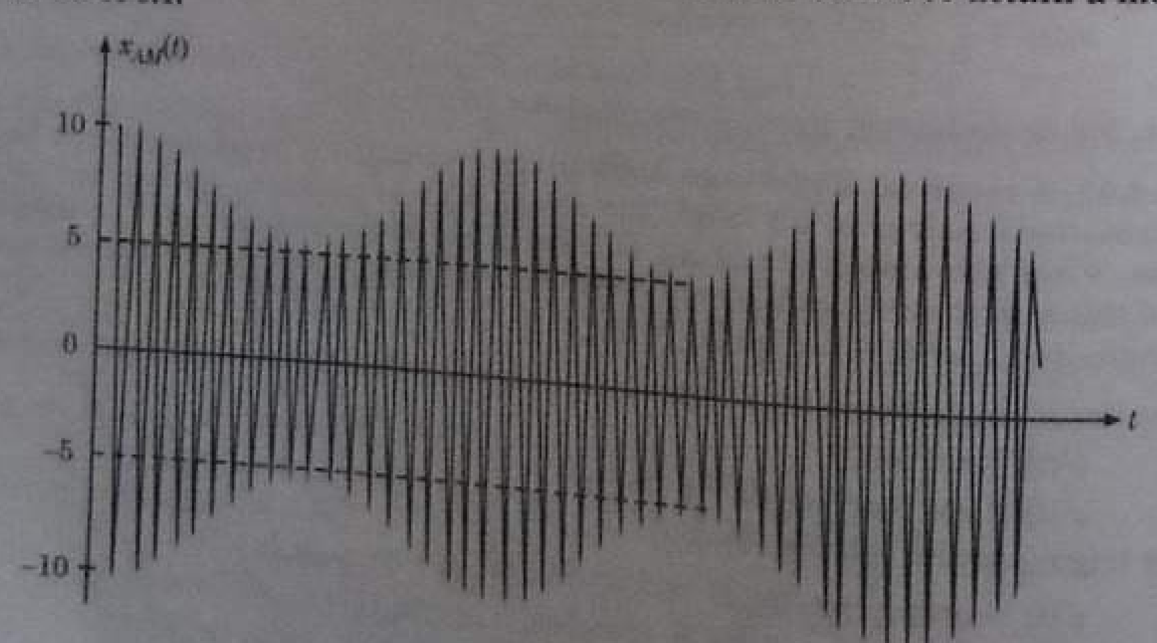


Fig. 3.66.

Solution: (i) Modulation index is given by

$$m_a = \frac{A_{max} - A_{min}}{A_{max} + A_{min}}$$

From figure, on substituting values, we get

$$m_a = \frac{10-5}{10+5} = \frac{1}{3}$$

(ii) Modulated AM signal is expressed as,

$$x_{AM}(t) = A\cos\omega_c t + m_a A\cos\omega_m t\cos\omega_c t \quad ...(i)$$

or $$x_{AM}(t) = A\cos\omega_c t + \frac{1}{2}m_a A[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t]$$

Now, P_c = carrier power = $\frac{1}{2}A^2$

P_s = sideband power

or $$P_s = \frac{1}{2}\left[\left(\frac{1}{2}m_a A\right)^2 + \left(\frac{1}{2}m_a A\right)^2\right] = \frac{1}{4}m_a^2 A^2$$

Total power, $$P_t = P_c + P_s = \frac{1}{2}\left(1 + \frac{m_a^2}{2}\right)A^2$$

Thus, efficiency $$\eta = \frac{P_s}{P_t} \times 100\% = \frac{\frac{1}{4}m_a^2}{\left[\frac{1}{2} + \frac{m_a^2}{4}\right]A^2} \times 100\% = \frac{m_a^2}{2 + m_a^2} \times 100\%$$

Substituting $m_a = \frac{1}{3}$, we get

Efficiency, $$\eta = \frac{\left(\frac{1}{3}\right)^2}{2 + \left(\frac{1}{3}\right)^2} \times 100 = 5.26 \text{ per cent}$$

(iii) From equation (i), we have

$$A_{max} = A_1 + m_1 A_1 \quad [\because \text{Initially amplitude of carrier wave is } A_1]$$

and $$A_{min} = A_1 - m_1 A_1$$

then $$m_1 A_1 = \frac{A_{max} - A_{min}}{2}$$

or $$A_1 = \frac{5}{2 \times \frac{1}{3}} = 7.5 \text{ V}$$

Let, new amplitude of the carrier be A_2.

Then, we have

$$m_2 A_2 = \frac{A_{max} - A_{min}}{2}$$

$$A_2 = \frac{5}{2 \times 0.1} = 25 \text{ V}$$

Thus, amplitude to be added = $A_2 - A_1$ = (25 – 7.5) V = 17.5 V **Ans.**

Example 3.53. Given a real signal $m(t)$, define a signal

$$M_+(t) = m(t) + j\hat{m}(t)$$

where $\hat{m}(t)$ is the Hilbert transform of $m(t)$. And $m_+(t)$ is called an analytic signal.

(*i*) Show that

$$F[m_+(t)] = M_+(\omega) = \begin{cases} 2M(\omega) & \text{for } \omega > 0 \\ 0 & \text{for } \omega < 0 \end{cases}$$

(*ii*) Show that $Re[m_+(t)\, e^{j\omega_c t}]$ is an upper-sideband SSB signal and $Re[m_+(t)\, e^{-j\omega_c t}]$ is a lower lower-sideband SSB signal.

Solution: (*i*) Given, $M_+(t) = m(t) + j\hat{m}(t)$

Taking Fourier transform of both sides, we get

$$F[m_+(t)] = F[m(t)] = jF[\hat{m}(t)]$$
$$= M(\omega) + j[-j\, sgn(\omega)\, M(\omega)] = M(\omega) + sgn(\omega)\, M(\omega)$$
$$= \begin{matrix} 2M(\omega), & \omega > 0 \\ 0, & \omega < 0 \end{matrix} \quad \begin{bmatrix} \text{Using} \\ \text{sgn}(\omega) = 1, \omega > 0 \\ = -1, \omega < 0 \end{bmatrix}$$

(*ii*)
$$\text{Re}[m_+(t)e^{j\omega_c t}] = \text{Re}[\{m(t) + j\hat{m}(t)\}(e^{j\omega_c t})]$$
$$= \text{Re}[(m(t) + j\,\hat{m}(t))][(\cos\omega_c t + j\sin\omega_c t)]$$
$$= m(t)\cos\omega_c t - \hat{m}(t)\sin\omega_c t$$

which is an upper SSB-signal.

Now, $\text{Re}[m_+(t)e^{-j\omega_c t}]$

$$= \text{Re}[\{m(t) + j\,\hat{m}(t)\}.e^{-j\omega_c t})]$$
$$= \text{Re}[\{m(t) + j\,\hat{m}(t)\}.(\cos\omega_c t - j\sin\omega_c t)]$$
$$= m(t)\cos\omega_c t + \hat{m}(t).\sin\omega_c(t)$$

which is an lower SSB-signal.

Example 3.54. The frequency multiplier is a nonlinear device followed by a bandpass filter, as shown in the figure 3.67. Suppose that the nonlinear device is an ideal square-law device with input-output characteristics:

$$e_o(t) = a\, e_i^2(t)$$

Determine the output $y(t)$ if the input is an FM signal given by

$$e_i(t) = A\cos(\omega_c t + \beta\sin\omega_m t)$$

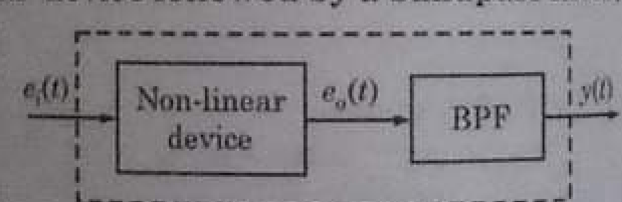


Fig. 3.67.

Solution: $y(t) = A'\cos(2\omega_c t + 2\beta\sin\omega_m t)$

where $A' = 1/2aA^2$

This result shows that a square-law device can be used as a frequency doubler.

Now, output of the non-linear device will be

$$e_o(t) = ae_i^2(t) = aA^2\cos^2(\omega_c t + \beta\sin\omega_m t) = \frac{aA^2}{2}[1 + \cos(2\omega_c t + 2\beta\sin\omega_m t)]$$

After bandpass filtering, the output will be

$$y(t) = \frac{aA^2}{2}\cos(2\omega_c t + 2\beta\sin\omega_m t) = A'\cos(2\omega_c t + 2\beta\sin\omega_m t)$$

where $A' = \dfrac{aA^2}{2}$

Hence Proved.

Example 3.55. The given figure 3.68 shows the circuit diagram of a square-law modulator. The signal applied to the nonlinear device is relatively weak, such that it can be represented by the following square law:

$$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t)$$

where a_1 and a_2 are constants, $v_1(t)$ is the input voltage, and $v_2(t)$ is the output voltage. The input voltage is defined as

$$v_1(t) = A_c\cos(2\pi f_c t) + m(t)$$

where $m(t)$ is a message signal and $A_c\cos(2\pi f_c t)$ is the carrier wave.

(*i*) Evaluate the output voltage $v_2(t)$

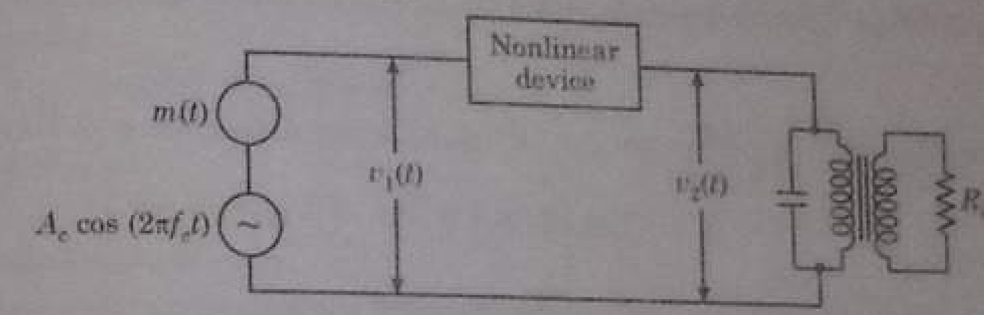


Fig. 3.68.

(*ii*) Specify the frequency response that the tuned circuit in the given figure must satisfy in order to generate and AM signal with f_c as the carrier frequency.

(*iii*) What will be the amplitude sensitivity of this AM signal?

Solution: The input voltage $v_1(t)$ is given by

$$v_1(t) = A_c\cos(2\pi f_c t) + m(t) \quad ...(i)$$

Then, the output voltage will be

$$v_2(t) = a_1[A_c\cos(2\pi f_c t) + m(t)] + a_2[A_c\cos(2\pi f_c t) + m(t)]^2 \quad ...(ii)$$

(*i*)
$$v_2(t) = a_1 A_c\left[1 + \frac{2a_2}{a_1}m(t)\right]\cos(2\pi f_c t) + a_1 m(t) + a_2 m^2(t) + a_2 A_c^2\cos^2(2\pi f_c t) \quad ...(iii)$$

Note that the required term for the AM signal is term $a_1 A_c\left[1 + \frac{2a_2}{a_1}m(t)\right]\cos(2\pi f_c t)$.

Then, the amplitude sensitivity, will be $k_a = \dfrac{2a_2}{a_1}$.

(*ii*) The remaining terms after AM signal can be removed by filtering. For this tuned filter, at the modulator output, to be designed such that it has a mid-band frequency f_c and bandwidth $2f_m$.

Example 3.56. Given the AM signal

$$s(t) = A_c[1 + m_a\cos(2\pi f_m t)]\cos(2\pi f_c t)$$

produced by a sinusoidal modulating signal of frequency f_m. Assume that the modulation factor is $m_a = 2$, and the carrier frequency f_c is much greater than f_m. The AM signal $s(t)$ is applied to an ideal envelope detector, producing the output $v(t)$.

(*i*) Find the Fourier series representation of $v(t)$.

(*ii*) What will be the ratio of second-harmonic amplitude of fundamental amplitude in $v(t)$?

Solution: (*i*) Envelope detector output is given by

$$v(t) = A_c|1 + m_a\cos(2\pi f_m t)|$$

which has been illustrated in figure 3.69, for the case when $m_a = 2$.

From the figure, it may be observed that, $v(t)$ is periodic with a period equal to f_m, and even function of t, and so only cosine terms will be present in its Fourier series representation i.e.,

$$v(t) = a_0 + 2\sum_{n=1}^{\infty} a_n \cos(2\pi f_m t) \qquad ...(i)$$

where $$a_0 = 2f_m \int_0^{\frac{1}{2f_m}} v(t)\, dt$$

or $$a_0 = 2A_c f_m \int_0^{\frac{1}{3f_m}} [1 + 2\cos(2n\pi f_m t)]\, dt + 2A_c f_m \int_{\frac{1}{3f_m}}^{\frac{1}{2f_m}} [-1 - 2\cos(2\pi f_m t)]\, dt$$

or $$a_0 = \frac{A_c}{3} + \frac{4A_c}{\pi}\sin\left(\frac{2\pi}{3}\right)$$

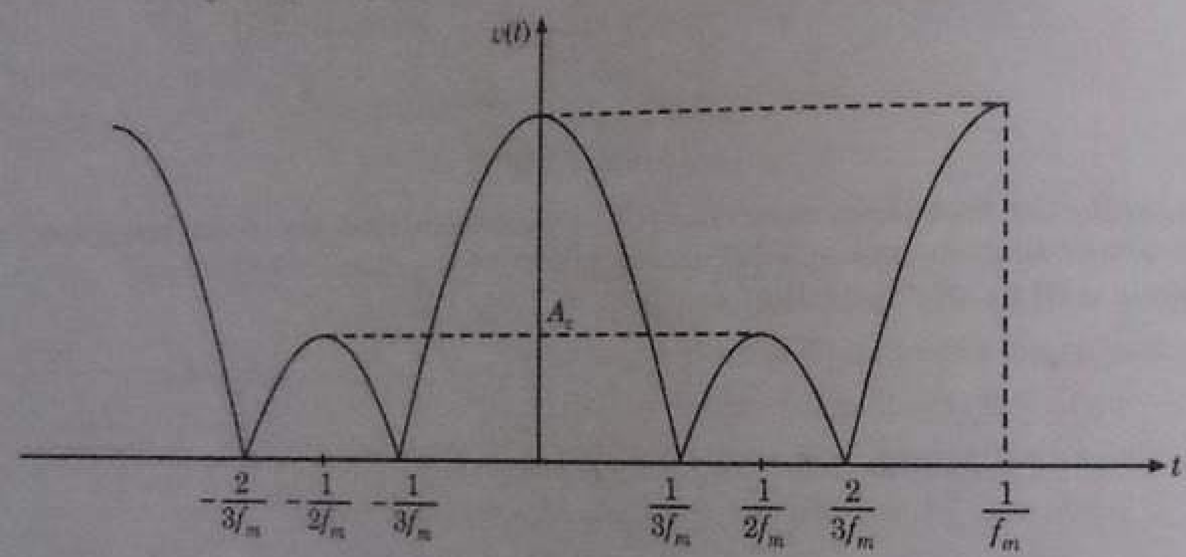


Fig. 3.69.

Now, we have

$$a_n = 2f_m \int_0^{\frac{1}{2f_m}} v(t)\cos(2n\pi f_m t)\, dt$$

or $$a_n = 2A_c f_m \int_0^{\frac{1}{3f_m}} [1 + 2\cos(2n\pi f_m t)\cos(2\pi f_m t)\, dt + 2A_c f_m \int_{\frac{1}{3f_m}}^{\frac{1}{2f_m}} [1 - 2\cos(2n\pi f_m t)\cos(2n\pi f_m t)\, dt$$

or $$a_n = \frac{A_c}{n\pi}\left[2\sin\left(\frac{2n\pi}{3}\right) - \sin(n\pi)\right] + \frac{A_c}{(n+1)\pi}\left\{2\sin\left[\frac{2\pi}{3}(n+1)\right] - \sin[\pi(n+1)]\right\}$$
$$+ \frac{A_c}{(n-1)\pi}\left\{2\sin\left[\frac{2\pi}{3}(n-1)\right] - \sin[\pi(n-1)]\right\} \qquad ...(ii)$$

For $n = 0$, equation (*ii*) reduces to that shown in equation (*i*).

(*ii*) For $n = 1$, equation (*ii*) yields

$$a_1 = A_c\left(\frac{\sqrt{3}}{2\pi} + \frac{1}{3}\right)$$

For $n = 2$, it yields

$$a_2 = \frac{A_c\sqrt{3}}{2\pi}$$

Therefore, the required ratio of second-harmonic amplitude to fundamental amplitude in $v(t)$ is

$$\frac{a_2}{a_1} = \frac{3\sqrt{3}}{2\pi + 3\sqrt{3}} = 0.452 \qquad \textbf{Ans.}$$

Example 3.57. Given a square-law detector, using a nonlinear device whose transfer characteristic is defined by

$$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t)$$

where a_1 and a_2 are constants, $v_1(t)$ is the input, and $v_2(t)$ is the output. The input consists of the AM wave

$$v_1(t) = A_c[1 + k_a\, m(t)]\cos(2\pi f_c t)$$

(*i*) Compute the output $v_2(t)$.

(*ii*) Determine the conditions for which the message signal $m(t)$ may be recovered from output $v_2(t)$.

Solution: Given,

$$v_1(t) = A_c[1 + k_a m(t)]\cos 2\pi f_c t$$

Then, the output of the square-law device will be

(*i*) $$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t)$$
$$= a_1 A_c[1 + k_a m(t)]\cos 2\pi f_c t + \frac{1}{2}a_2 A_c^2[1 + k_a m(t) + a + k_a^2 m^2(t)][1 + \cos(4\pi f_c t)] \qquad \textbf{Ans.}$$

(*ii*) The required signal $aA_c^2 k_a m(t)$ can be extracted by a low-pass filter. Also the term $\frac{1}{2}a_2 A_c^2 k_a^2 m^2(t)$ will give rise to plurality of similar frequency components within the baseband spectrum. Thus, ratio of wanted signal to distortion is $\frac{2}{k_a m(t)}$. To make this ratio large, $k_a m(t)$ should be small compared to unity. **Ans.**

Example 3.58. The following AM signal

$$s(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$$

is applied to the system in the figure 3.70. Assuming that $|k_a m(t)| < 1$ for all t and message signal $m(t)$ is limited to the interval $-f_m \le f \le f_m$, and that the carrier frequency $f_c > 2f_m$, show that $m(t)$ can be obtained from the square-rooter output $v_3(t)$.

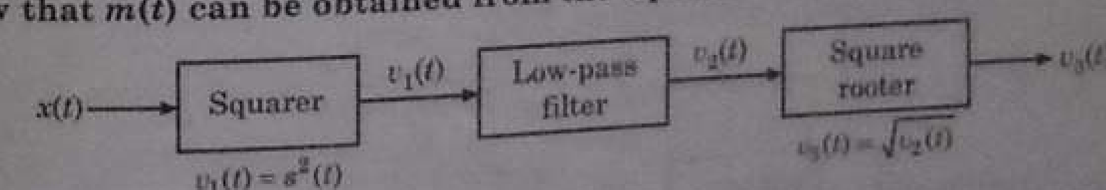


Fig. 3.70

Solution: Squarer output is given by

$$v_1(t) = s_1^2(t)$$

or $$v_1(t) = A_c^2[1 + k_a m(t)]^2\cos^2(2\pi f_c t)$$

or $$v_1(t) = \frac{A_c^2}{2}[1 + 2k_a m(t)] + m^2(t)[1 + \cos(4\pi f_c t)]$$

The amplitude spectrum of $v_1(t)$ is, therefore, as shown in figure 3.71, assuming that $m(t)$ is limited to the interval $-f_m \leq f \leq f_m$.

Since $f_c > 2f_m$, then $2f_c - 2f_m > 2f_m$. Therefore, by choosing the cutoff frequency of the low-pass filter greater than $2f_m$, but less than $2f_c - 2f_m$. The output will be

$$v_2(t) = \frac{A_c^2}{2}[1 + k_a m(t)]^2$$

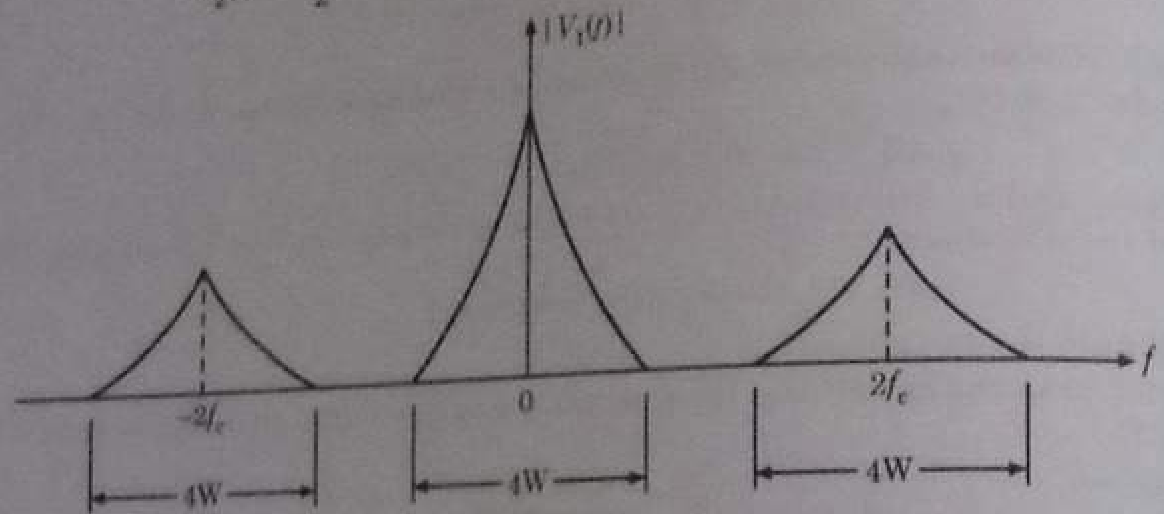


Fig. 3.71.

∴ Output of square rooter is

$$v_3(t) = \sqrt{v_2(t)} = \frac{A_c}{\sqrt{2}}[1 + k_a m(t)]$$

Now, the blocking capacitor will suppress the dc term present, thus, we have

$$v_0(t) = \frac{A_c}{\sqrt{2}} k_a m(t)$$

Hence, the output is proportional to the message signal $m(t)$. **Hence Proved.**

Example 3.59. A DSB-SC modulated signal is demodulated by applying it to a coherent detector.

(i) Compute the effect of a frequency error Δf in the local carrier frequency of the detector, measured with respect to the carrier frequency of the incoming DSB-SC signal.

(ii) For the case a sinusoidal modulating wave, show that because of this frequency error, the demodulated signal exhibits beats at the error frequency. Give your answer with a sketch of this demodulated signal.

Solution: (i) We know that the output of a product modulator is given by

$$v(t) = A_c m(t) \cos(2\pi f_c t) \cos[2\pi(f_c + \Delta f)t]$$

$$= \frac{A_c}{2} m(t) \{\cos(2\pi \Delta f t) + \cos[2\pi(2f_c + \Delta f)t]\}$$

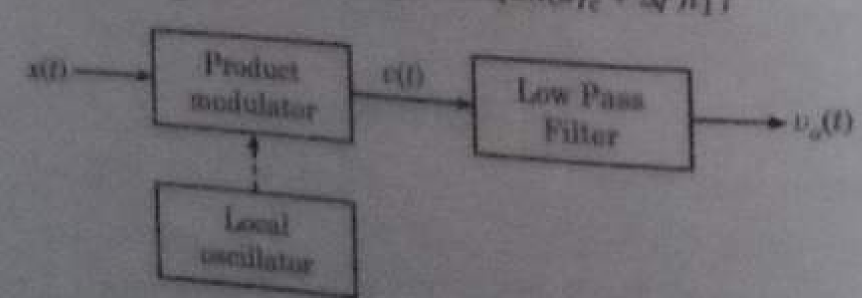


Fig. 3.72.

Then after low pass filtering, we have

$$v_0(t) = \frac{A_c}{2} m(t) \cos(2\pi f_c t)$$

Hence, the output signal is the message signal modulated by a sinusoid of frequency Δf.

(ii) If $\quad m(t) = \cos(2\pi f_m t)$

then $\quad s_2(t) = \frac{A_c}{2}\cos(2\pi f_m t)\cos(2\pi \Delta f t)$

or $\quad s_2(t) = \frac{A_c}{4}\{\cos[2\pi(f_m + \Delta f)t] + \cos[2\pi(f_m - \Delta f)t]\}$

the waveform of $s_2(t)$ will be as shown in figure 3.73.

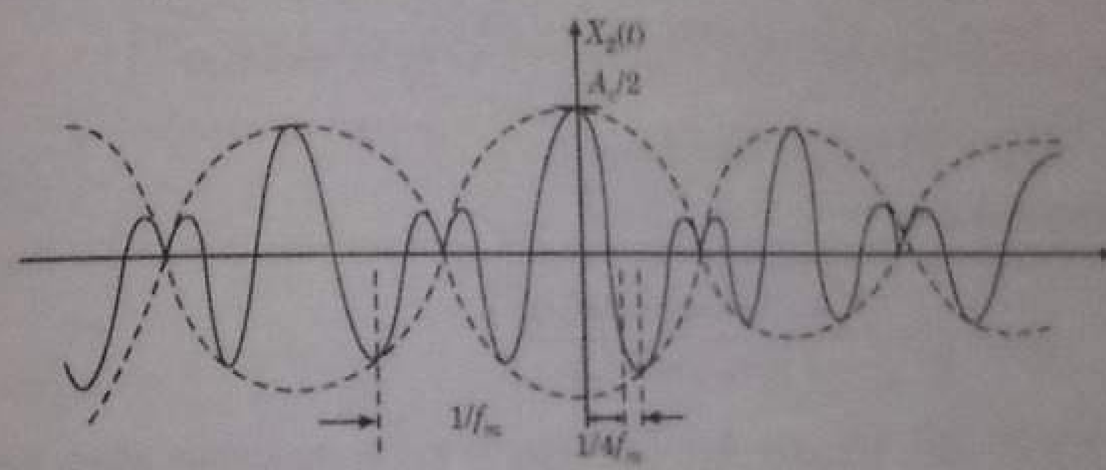


Fig. 3.73.

Example 3.60. The signal tone modulating $m(t) = A_m \cos(2\pi f_m t)$ is used to generate the following VSB signal:

$$s(t) = \frac{1}{2} a A_m A_c \cos 2\pi(f_c + f_m)t] + \frac{1}{2} A_m A_c (1 - a) \cos [2\pi(f_c - f_m)t]$$

where a is constant, less than unity, representing the attenuation of the upper side frequency.

(i) Determine the quadrature component of the VSB signal $s(t)$.

(ii) The VSB signal, plus the carrier $A_c \cos(2\pi f_c t)$, is passed through an envelope detector. Determine the distortion produced by the quadrature component.

(iii) What will be the value of constant a for which this distortion reaches its worst possible condition ? *(GATE Examination, 1991)*

Solution: (i) We know that any signal $s(t)$ can be expressed in terms of in phase and quardative components as under:

$$s(t) = s_I(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t) \qquad ...(i)$$

Now, expanding the given signal $s(t)$, we have

$$s(t) = \frac{1}{2} a A_m A_c \cos(2\pi f_c t)\cos(2\pi f_m t) - \frac{1}{2} a A_m A_c \sin(2\pi f_c t)\sin(2\pi f_m t)$$
$$+ \frac{1}{2}(1 - a) A_c A_m \cos(2\pi f_c t)\cos(2\pi f_m t) + \frac{1}{2}(1 - a) A_m A_c \sin(2\pi f_c t)\sin(2\pi f_m t)$$

or $$s(t) = \frac{1}{2} A_m A_c \cos(2\pi f_c t)\cos(2\pi f_m t) + \frac{1}{2} A_m A_c (1 - 2a)\sin(2\pi f_c t)\sin(2\pi f_m t) \qquad ...(ii)$$

Now, comparing equation (ii) with equation (i), we have the quadrative component.

$$s_Q(t) = -\frac{1}{2} A_c A_m (1-2a)\sin(2\pi f_m t)$$

(ii) Further, the VSB signal after adding carrier signal $A_c \cos(2\pi f_c t)$, will be

$$s(t) = A_c\left[1+\frac{A_m}{2}\cos(2\pi f_m t)\right]\cos(2\pi f_c t) + \frac{1}{2}A_c A_m (1-2a)\sin(2\pi f_m t)\sin(2\pi f_c t)$$

Then, envelope will be equal to

$$a(t) = \sqrt{s_Q^2(t) + s_I^2(t)} = A_c\sqrt{\left[1+\frac{1}{2}A_m\cos(2\pi f_m t)\right]^2 + \left[\frac{1}{2}A_m(1-2a)\sin(2\pi f_m t)\right]^2}$$

or
$$a(t) = A_c\left[1+\frac{1}{2}A_m\cos(2\pi f_m t)\right]\sqrt{1+\left[\frac{\frac{1}{2}A_m(1-2a)\sin(2\pi f_m t)}{1+\frac{1}{2}A_m\cos(2\pi f_m t)}\right]^2}$$

or
$$a(t) = A_c[1+\tfrac{1}{2}A_m\cos(2\pi f_m t)]\,d(t)$$

where distortion is given by

$$d(t) = \sqrt{1+\left[\frac{\frac{1}{2}A_m(1-2a)\sin(2\pi f_m t)}{1+\frac{1}{2}A_m\cos(2\pi f_m t)}\right]^2} \qquad \text{...(iii)}$$

(iii) From equation (iii), it is clear that the distortion will be maximum when, $a = 0$. **Ans.**

Example 3.61. Given the following modulated wave

$$s(t) = A_c\cos(2\pi f_c t) + m(t)\cos(2\pi f_c t) - \hat{m}(t)\sin(2\pi f_c t)$$

which represents a carrier plus an SSB signal, with $m(t)$ denoting the message signal and $\hat{m}(t)$ its Hilbert transform. Find the conditions for which an ideal envelope detector, with $s(t)$ as input, would produce a good approximation to the message signal $m(t)$.

Solution: Given modulated wave is

$$s(t) = [A_c + m(t)]\cos 2\pi f_c(t) - \hat{m}(t)\sin(2\pi f_c t)$$

then, in phase component

$$s_I(t) = A_c + m(t)$$

and, quadrative component

$$s_Q(t) = +\hat{m}(t)$$

then, envelope detector output will be

$$a(t) = \sqrt{s_I^2(t) + s_Q^2(t)} = \sqrt{[A_c + m(t)]^2 + [\hat{m}(t)]^2}$$

If carrier amplitude is such that,

$A_c >> |m(t)|$ and $A_c >> |\hat{m}(t)|$, for all values of t,

then, we have
$$a(t) \cong \sqrt{A_c^2 + 2A_c m(t)} \cong A_c\sqrt{1+\frac{2}{A_c}m(t)}$$

or
$$a(t) \cong A_c\left[1+\frac{m(t)}{A_c}\right] = A_c + m(t)$$

It may be observed that except for dc bias A_c, the envelope $a(t)$ is approximately equal to $m(t)$. **Ans.**

Example 3.62. Given the SSB wave

$$s(t) = m(t)\cos(2\pi f_c t) - \hat{m}(t)\sin(2\pi f_c t)$$

where f_c is the carrier frequency, $m(t)$ is the message signal and $\hat{m}(t)$ it its Hilbert transform. This modulated wave is applied to a square-law device characterized by

$$y(t) = s^2(t)$$

Prove that the output $y(t)$ contains a frequency components at twice the carrier frequency but that it has a time-varying phase, which makes it impractical to recover the carrier by squaring.

Solution: The square law device output, with SSB wave $s(t)$ as input, is given by

$$y(t) = s^2(t) = [m(t)\cos(2\pi f_c t) - \hat{m}(t)\sin(2\pi f_c t)]^2$$
$$= m^2(t)\cos^2(2\pi f_c t) + \hat{m}^2(t)\sin^2(2\pi f_c t) - 2m(t)\,\hat{m}(t)\cos(2\pi f_c t)\sin(2\pi f_c t)$$
$$= \frac{1}{2}m^2(t)[1+\cos(4\pi f_c t)] + \frac{1}{2}\hat{m}^2(t)[1-\cos(4\pi f_c t)] - m(t)\,\hat{m}(t)\sin(4\pi f_c t)$$

$$y(t) = \frac{1}{2}[m^2(t) + \hat{m}^2(t)] + \left\{\frac{1}{4}[m^2(t) + \hat{m}^2(t)]^2 + m^2(t)\,\hat{m}^2(t)\right\}^{1/2} - \cos[4\pi f_c t + Q(t)]$$

where $Q(t) = \tan^{-1}\left[\dfrac{2m(t)\,\hat{m}(t)}{m^2(t) + \hat{m}^2(t)}\right]$

Since, y(t) is a mixed combination of amplitude and frequency modulation. The AM can be eliminated by filtering but the phase modulation is unaffected by such an operation, which makes it impractical to recover the carrier wave by squaring. **Hence Proved.**

Example 3.63. Single-sideband modulation can be viewed as a hybrid form of amplitude modulation and frequency modulation. Compute the envelope and instantaneous frequency of an SSB wave for the following two cases:

(i) When only the upper sideband is transmitted.

(ii) When only the lower sideband is transmitted.

Solution: (i) For SSB modulation, the modulated wave is given by

$$s(t) = \frac{A_c}{2}[m(t)\cos(2\pi f_c t) \pm \hat{m}(t)\sin(2\pi f_c t)]$$

where +ve and –ve sign corresponds to upper and lower sideband of the modulated wave respectively. Now, the envelope for the SSB wave will be given by,

$$a(t) = \frac{A_c}{2}\sqrt{m^2(t) + \hat{m}^2(t)}$$

which is same for both transmitted upper sideband and lower sideband.

For upper sideband transmission, the angle, is given by

$$Q_i(t) = 2\pi f_c t + \tan^{-1}\frac{\hat{m}(t)}{m(t)}$$

and the instantaneous frequency, will be

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt}$$

Thus, we have

or
$$f_i(t) = f_c + \frac{m(t)\,\hat{m}'(t) - \hat{m}(t)\,m'(t)}{2\pi[m^2(t) + \hat{m}^2(t)]}$$

here ' ′ ' denotes time derivative.

(*ii*) For lower sideband transmission, the angle is given by

$$\theta_i(t) = 2\pi f_c t + \tan^{-1}\left(-\frac{\hat{m}(t)}{m(t)}\right)$$

and, the instantaneous frequency will be

$$f_i(t) = f_c + \frac{\hat{m}(t)\,m'(t) - m(t)\hat{m}'(t)}{2\pi[m^2(t) + \hat{m}^2(t)]} \quad \textbf{Ans.}$$

Example 3.64. An SSB signal is demodulated by using a synchronous demodulator. However, the locally arranged carrier has a phase error θ. Determine the effect of the error on demodulation. What will be the effect of this error if the input is DSB-SC in place of SSB?

(GATE Examination, 1998)

Solution: We know that an SSB signal is expressed by,

$$x_{SSB}(t) = m(t).\cos(\omega_c t) \mp \hat{m}(t).\sin(\omega_c t)$$

where $\hat{m}(t)$ is the Hilbert transform of message signal $m(t)$.

In a synchronous demodulator shown in figure 3.74, an SSB signal is present at the input.

Then, we have $\quad d(t) = [m(t).\cos(\omega_c t) \mp \hat{m}(t).\sin(\omega_c t)]\cos[\omega_c t + \theta]$

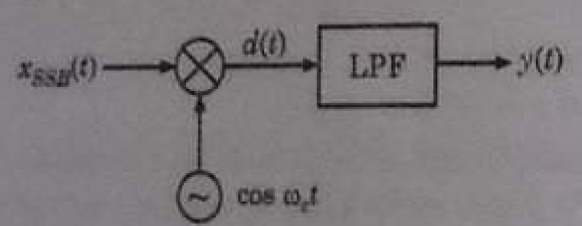


Fig. 3.74.

or $\quad d(t) = \frac{m(t)}{2}.[\cos(2\omega_c t + \theta) + \cos\theta] \mp \hat{m}(t).[\sin(2\omega_c t + \theta) - \sin\theta)]$

After low-pass filtering, we get

Output $\quad y(t) = \frac{m(t)}{2}.\cos\theta \mp \hat{m}(t).\sin\theta$

Hence, there is no attenuation but only phase distortion by angle θ.

Now, when the input is DSB-SC

then, we have $\quad x_{DSB}(t) = m(t).\cos(\omega_c t + \theta)$

and, $\quad d(t) = [m(t).\cos(\omega_c t)].\cos[\omega_c t + \theta] = \frac{1}{2}m(t)[\cos\theta + \cos(2\omega_c t + \theta)]$

after low-pass filtering, we get the output

$$y(t) = \frac{1}{2}m(t).\cos\theta$$

Hence, output is proportional to $m(t)$, when θ is is constant. But if θ varies randomly with time, then the output will also vary which is undesirable.

Example 3.65. Derive an expression for the signal $v_3(t)$ in figure 3.75 for $v_1(t) = 10\cos(2000\pi t) + 4\sin(200\pi t)$. Assume that $v_2(t) = v_1(t) + 0.1\,v_1^2(t)$ and that the BPF is an ideal unity gain filter with passband from 800 Hz to 1200 Hz.

(GATE Examination-1998)

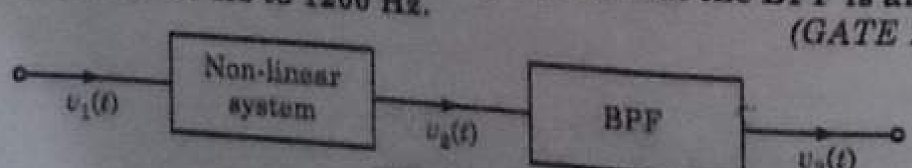


Fig. 3.75.

Solution: From the given figure, we obtain

Output at non-linear system,

$$v_2(t) = v_1(t) + 0.1v_1^2(t)$$

or $\quad v_2(t) = [10\cos(2000\pi t) + 4\sin(200\pi t)] + [10\cos(2000\pi t) + 4\sin(200\pi t)]^2 \times 0.1$

or $\quad v_2(t) = 10\cos(2000\pi t) + 4\sin(200\pi t) + 0.1 \times [100\cos^2(200\pi t) + 16\sin^2(200\pi t) + 80\cos(2000\pi t)\sin(200\pi t)]$

or $\quad v_2(t) = 10\cos(2000\pi t) + 4\sin(200\pi t) + \left[100\left(\frac{1+\cos(4000\pi t)}{2}\right) + 16\,\frac{[1-\cos(400\pi t)]}{2}\right] \times 0.1 + 40\,[\sin(2200\pi t) - \sin(1800\pi t)] \times 0.1$

or $\quad v_2(t) = 10.\cos(2000\pi t) + 0.5\cos(4000\pi t) - 0.8\cos(400\pi t) + 4\sin(200\pi t) + 4\sin(2200\pi t) - 4\sin(1800\pi t) + 5.8$

Because, BPF has a pass-band from 800 Hz-1200 Hz, therefore the output after filtering will be

$$v_3(t) = 10\cos(2000\pi t) + 4\sin(2200\pi t) - 4\sin(1800\pi t) \quad \textbf{Ans.}$$

Example 3.66. A given AM broadcast station transmits a total power of 5 kW when the carrier is modulated by a sinusoidal signal with a modulation index of 0.7071. Compute

(*i*) the carrier power

(*ii*) the transmission efficiency, and

(*iii*) the peak amplitude of the carrier assuming the antenna to be represented by a (50 + *j*0) Ω load.

Solution: (*i*) An amplitude modulated signal can be expressed as under:

$$x_{AM}(t) = A\cos\omega_c t + mA\cos\omega_m t\cos\omega_c t$$

or $\quad x_{AM}(t) = A\cos\omega_c t + \frac{1}{2}mA\cos(\omega_c - \omega_m)t + \frac{mA}{2}.\cos(\omega_c + \omega_m)t \quad$ ...(*i*)

Now, $\quad P_c = \text{carrier power} = \frac{A^2}{2}$

$$P_s = \text{sideband power} = \frac{1}{2}\left[\left(\frac{1}{2}mA\right)^2 + \left(\frac{mA}{2}\right)^2\right] = \frac{m^2A^2}{4} \quad \text{...(ii)}$$

We know that the total power is given by

$$P_t = \frac{A^2}{2}\left(1 + \frac{m^2}{2}\right) = P_c\left(1 + \frac{m^2}{2}\right)$$

Given $\quad P_t = 50 \times 10^3$ watt,

then, $\quad P_c = \frac{P_t}{1 + \frac{m^2}{2}} = \frac{50 \times 10^3}{1.25} = 40$ kW. **Ans.**

(*ii*) Further, we know that the transmission efficiency is given by

$$\eta = \frac{P_s}{P_t} \times 100\% = \frac{\frac{1}{4}m^2A^2}{\left(\frac{1}{2} + \frac{m^2}{4}\right)A^2} \times 100\%$$

or $\quad \eta = \frac{m^2}{2 + m^2} \times 100\% = \frac{0.50}{2 + 0.50} \times 100\% = 20\%$ **Ans.**

(iii) Power delivered to the load will be

or $$A_c = \frac{A_c^2}{2R} = 40 \times 10^3 = \sqrt{2 \times 40 \times 10^3 \times 50} = 2 \text{ kV} \quad \textbf{Ans.}$$

Example 3.67. Show that in quadrature amplitude moduation (QAM) each signal can be recovered by synchronous detection of the received signal by using two local carriers of same frequency but in phase quadrature.

Solution: A QAM scheme has been shown in figure 3.76.

In the figure, the boxes labelled $\frac{-\pi}{2}$ are phase shifters. If the two baseband signals to be transmitted are $m_1(t)$ and $m_2(t)$, then, we will get

$$x_{QAM}(t) = m_1(t) \cos \omega_c t + m_2(t). \sin \omega_c t$$

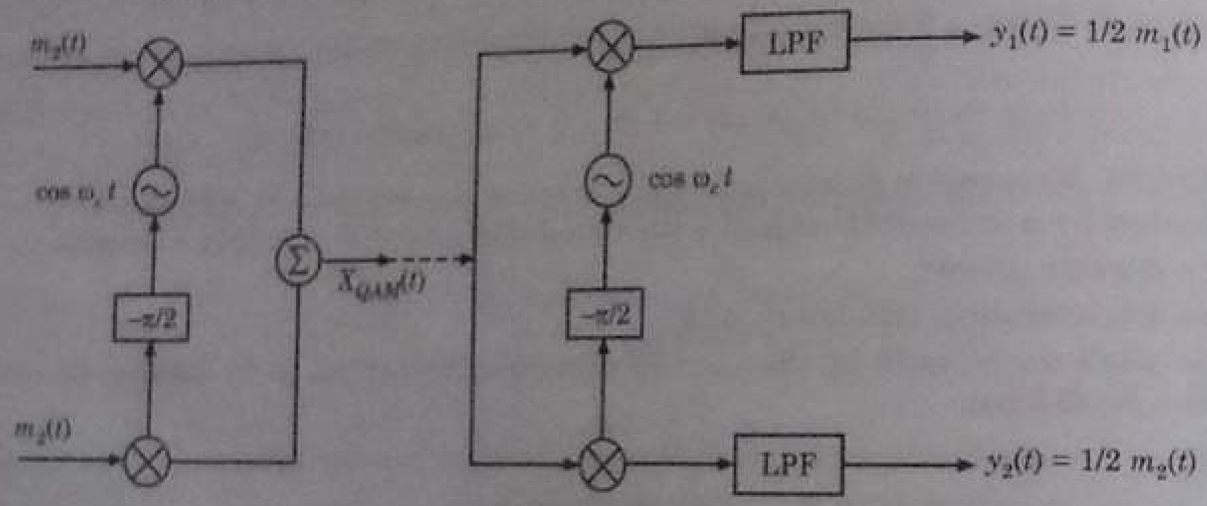


Fig. 3.76.

The output $x_1(t)$ of the upper multiplier will be

$$x_1(t) = 2x_{QAM}(t). \cos \omega_c t = 2[m_1(t) \cos \omega_c t + m_2(t) . \sin \omega_c t]. \cos \omega_c t$$
$$= m_1(t) + m_1(t) \cos 2\omega_c t + m_2(t). \sin 2\omega_c t$$

After low-pass filtering, we have

$$x_1(t) = m_1(t)$$

Now, in a similar fashion, we get

$$x_2(t) = 2x_{QAM}(t). \sin \omega_c t$$
$$= 2[m_1(t). \cos \omega_c t + m_2(t).\sin \omega_c t]. \sin \omega_c t$$

or $$x_2(t) = m_1(t) \sin 2\omega_c t + m_2(t) - m_2(t). \cos 2\omega_c t$$

Again, low-pass filtering yields

$$x_2(t) = m_2(t)$$

It may noted that Q_{AM} is an efficient method of transmitting two message signals within the same bandwidth.

Example 3.68. Explain the Quadrature null effect in a coherent detector for DSB-SC signals.

Solution: A coherent detector has been depicted in figure 3.77. Here, it is assumed that the local oscillator signal is exactly coherent or synchronized, in both frequency and phase, with the carrier wave. However, the more general demodulation process is using a local oscillator having same frequency but arbitrary phase difference ϕ with respect to carrier wave. Hence, expressing local oscillator signal as, $A_c' \cos(2\pi f_c t + \phi)$, the product modulator output will be

$$v_0(t) = A'_c \cos(2\pi f_c t + \phi) . s(t) = A_c' A_c \cos(2\pi f_c t + \phi). \cos(2\pi f_c t)\, m(t) \left[A_c \cos 2\pi = \frac{1}{2}\right]$$

[Here, $s(t) = m(t) = m(t). A_c \cos 2\pi f_c t$]

or $$v_0(t) = \frac{1}{2} A_c A'_c (\cos 4\pi f_c t + \phi(t)\, m(t) + \frac{1}{2} A_c A'_c \cos\phi\, m(t)$$

The first term is removed by low-pass filtering having cut-off frequency of this filter is greater than W but less than $2f_c - W$.

Then, filter output would be

$$v_0(t) = \frac{1}{2} A_c A'_c \cos\phi\, m(t)$$

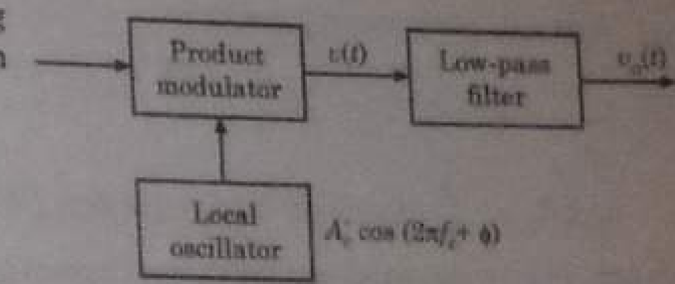


Fig. 3.77.

The demodulated signal $v_0(t)$ will therefore proportional to $m(t)$ if ϕ is constant.

The amplitude goes maximum for $\phi = 0$, and minimum for $\phi = \pm\frac{\pi}{2}$. The zero demodulator signal occurring at $\phi = \pm\frac{\pi}{2}$ represents the **quadrature null effect** of synchronous demodulator or **coherent detector**.

Example 3.69. In an AM system, the modulating signal is sinusoidal with frequency f_m Hz. If 80% modulation is used, then find the ratio of total sideband power in the modulation signal, to the total power.

Solution: The AM signal is expressed as

$$x_{AM}(t) = A \cos \omega_c t + mA \cos \omega_m t \cos \omega_c t$$
$$= A \cos\omega_c t + \frac{mA}{2} [\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t]$$

$$P_c \text{(carrier power)} = \frac{1}{2} A^2$$

$$P_s \text{(sideband power)} = \frac{1}{2}\left[\left(\frac{1}{2}mA\right)^2 + \left(\frac{1}{2}mA\right)^2\right] = \frac{1}{4} m^2 A^2$$

Total power is given by

$$P_s = P_c + P_s = \frac{1}{2}\left(1 + \frac{m^2}{2}\right) A^2$$

$$\text{Ratio} = \frac{\text{Sideband power}}{\text{Total power}} = \frac{1/4 m^2 A^2}{1/4(2+m^2)A^2} = \frac{m^2}{2+m^2}$$

$$= \frac{(0.80)^2}{2 + (0.80)^2} = \frac{0.64}{2.64} = \mathbf{0.242.} \quad \textbf{Ans.}$$

Example 3.70. In an envelope detector, the input is an AM signal which is expressed as

$$x_{AM}(t) = A(1 + m \cos \omega_m t). \cos \omega_c t$$

where m is a constant, $0 < m < 1$, and $\omega_c >> \omega_m$. Then prove that

(i) At any time t_0

$$\frac{1}{RC} \geq \omega_m \left(\frac{m \sin \omega_m t_0}{1 + m \cos \omega_m t_0}\right) \text{ and}$$

(ii) $$\frac{1}{RC} \geq \frac{m\omega_m}{\sqrt{1-m^2}}$$

Solution: Figure 3.78 shows envelope detector and envelope of $x_{AM}(t)$.

Voltage across capacitor C is given by

$$v_c(t) = E_0\, e^{-t/RC}$$

Interval between two peaks is

$$= \frac{1}{f_c} = \frac{2\pi}{\omega_c}$$

and $R_C >> \frac{1}{\omega_c}$, then $v_c(t)$ discharges exponentially for a short time compared to its time constant.

Hence, the exponential can be approximated by a straight line, *i.e.*,

$$v_c(t) \cong E\left(1-\frac{t}{RC}\right)$$

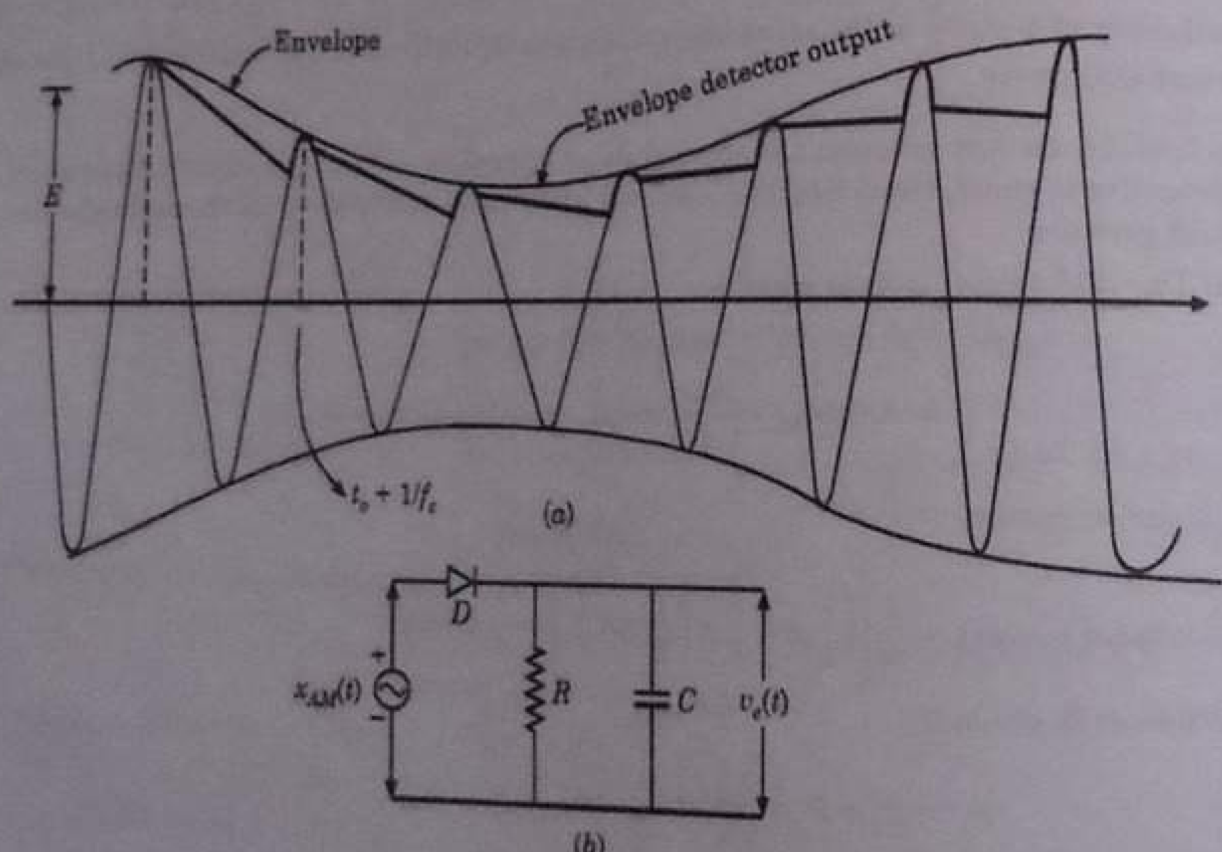


Fig. 3.78.

Because, in the envelope detector, the output of the detector follows the envelope of modulated signal, it is prerequisite at any time t_0,

$$(1+m\cos\omega_m t_0)\left(1-\frac{1}{RC\,f_c}\right) \le 1+m\cos\omega_m\left(t_0+\frac{1}{f_c}\right)$$

If $\omega_m << \omega_c$, we shall have

$$1+m\cos\omega_m\left(t_0+\frac{1}{f_c}\right) = 1+m\cos\left(\omega_m t_0+\frac{\omega_m}{f_c}\right)$$

$$\cong 1+m\cos\omega_m t_0-\frac{m\omega_m}{f_c}.\sin\omega_m t_0$$

or $$(1+m\cos\omega_m t_0)\left(\frac{1}{RC\,f_c}\right) \ge \frac{m\omega_m}{f_c}\sin\omega_m t_0$$

or $$\frac{1}{RC} \ge \omega_m\cdot\left(\frac{m\sin\omega_m t_0}{1+m\cos\omega_m t_0}\right)$$

(*ii*) Now from equation (*i*), we get

$$\frac{1}{RC}+\frac{m}{RC}\cos\omega_m t_0 \ge m\,\omega_m\sin\omega_m t_0$$

or $$m\left(\omega_m\sin\omega_m t_0-\frac{1}{RC}\cos\omega_m t_0\right) \le \frac{1}{RC}$$

or $$m\sqrt{\omega_m^2+\frac{1}{(RC)^2}}\,.\sin\left(\omega_m t_0-\tan^{-1}\frac{1}{\omega_m R_C}\right) \ge \frac{1}{RC}$$

Because $\sin\theta \le 2$

Therefore, we have $\frac{1}{RC} \ge m\sqrt{\omega_m^2+\left(\frac{1}{RC}\right)^2}$

or $$m^2\left[\omega_m^2+\left(\frac{1}{RC}\right)^2\right] \le \left(\frac{1}{RC}\right)^2$$

or $$RC \le \frac{1}{\omega_m}\frac{\sqrt{1-m^2}}{m}$$ **Hence Proved.**

Example 3.71. Evaluate the condition for distortionless demodulation of a VSB signal, initially generated by passing a PSB signal through a vestigial filter, using synchronous detector.

Solution:

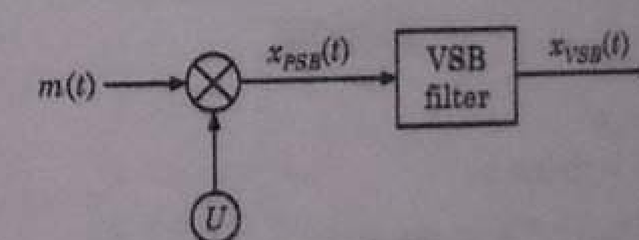


Fig. 3.79.

We have $x_{DSB}(t) = 2m(t)\cos\omega_c t$

Its spectrum is given by

$$X_{DSB}(\omega) = M(\omega-\omega_c)+M(\omega+\omega_c)$$

then, we have

$$X_{VSB}(\omega) = X_{DSB}(\omega).H(\omega)$$

where $H(\omega)$ is frequency response of the VSB filter ...(*i*)

$$X_{VSB}(\omega) = [M(\omega-\omega_c)+M(\omega+\omega_c)]\,H(\omega)$$

Now, for demodulation using synchronous demodulator, we have

$$d(t) = x_{VSB}(t).\cos\omega_c t$$

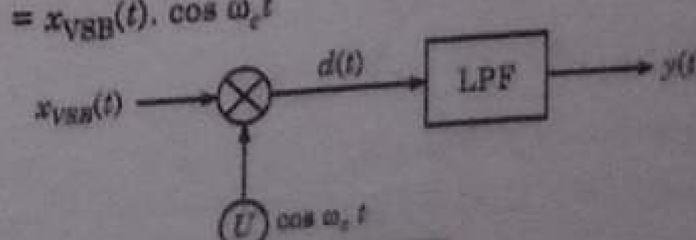


Fig. 3.80.

the Fourier transform is given by ...(*ii*)

$$D(\omega) = \frac{1}{2}[X_{VSB}(\omega-\omega_c)+X_{VSB}(\omega+\omega_c)]$$

Using equations (i) and (ii), and eliminately the spectra at $\pm 2\omega_c$ by LPF, we get

$$y(t) \leftrightarrow \frac{1}{2} M(\omega)[H(\omega+\omega_c)+H(\omega-\omega_c)]$$

Now for distortionless detection, we must have

$$y(t) = km(t) \leftrightarrow kM(\omega)$$

where k is constant.

Hence, for distortionless demodulation, we have

$$H(\omega+\omega_c) + H(\omega-\omega_c) = \text{constant}, \ |\omega| \le \omega_M$$

Example 3.72. Prove that the output of the phase-shift modulator is an SSB and verify that upper-sideband (USB) or a lower sideband (LSB) signal results from subtraction or addition at the summer.

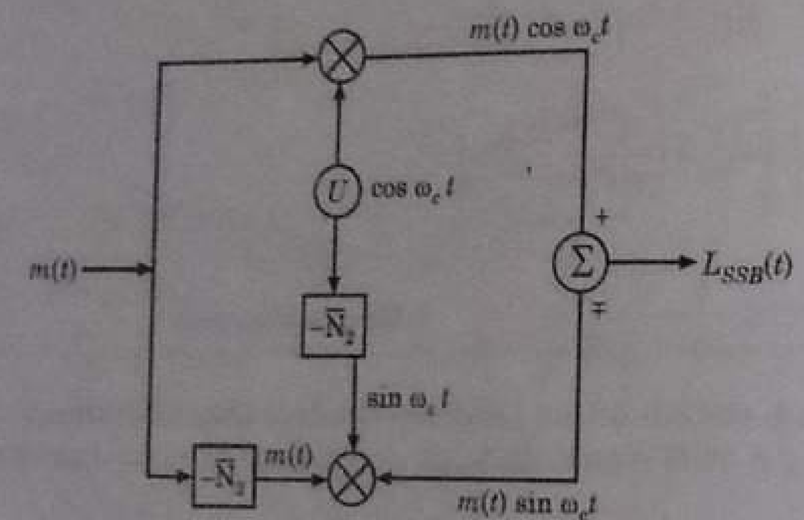


Fig. 3.81.

Solution:

Let $m(t) = \cos\omega_m t$

Hilbert transform of $m(t)$ is given by

$$\hat{m}(t) = \cos|(\omega_m t - \pi/2)| = \sin\omega_m t$$

Then, $x_{SSB}(t) = \cos\omega_m t \cos\omega_c t \mp \sin\omega_m t . \sin\omega_c t = \cos(\omega_c \pm \omega_m)t$

Now, with positive sign, we have

$$x_{SSB}(t) = \cos(\omega_c+\omega_m)t = m(t)\cos\omega_c t - \hat{m}(t).\sin\omega_c t \quad \ldots(i)$$

and $$x_{SSB}(t) = \sin(\omega_c-\omega_m)t = m(t)\cos_c t + \hat{m}(t).\sin\omega_c t \quad \ldots(ii)$$

Now, tacking Fourier transform of equation (i), we get

$$X_{SSB}(\omega) = \frac{1}{2}M(\omega-\omega_c)+\frac{1}{2}M(\omega+\omega_c) - \frac{1}{2j}[\hat{M}(\omega-\omega_c)-\hat{M}(\omega+\omega_c)]$$

Using $$m(t).\cos\omega_c t \leftrightarrow \frac{1}{2}[M(\omega-\omega_c)+M(\omega+\omega_c)]$$

and $$\hat{m}(t).\sin\omega_c t \leftrightarrow \frac{1}{2j}[\hat{M}(\omega-\omega_c)-\hat{M}(\omega+\omega_c)]$$

Now, $\hat{M}(\omega-\omega_c) = -j\,\text{sgn}(\omega-\omega_c) . M(\omega-\omega_c)$

$\hat{M}(\omega+\omega_c) = -j\text{sgn}(\omega+\omega_c) . M(\omega+\omega_c)$

We have

$$X_{SSB}(\omega) = \frac{1}{2}M(\omega-\omega_c)+\frac{1}{2}M(\omega+\omega_c) - \left[-\frac{1}{2}\sin(\omega-\omega_c).M(\omega-\omega_c) + \frac{1}{2}sgn(\omega+\omega_c)M(\omega+\omega_c)\right]$$

Since, $$1+\text{sgn}(\omega-\omega_c) = \begin{cases} 2 & \text{for } \omega > \omega_c \\ 0 & \text{for } \omega < \omega_c \end{cases}$$

and $$1-\text{sgn}(\omega-\omega_c) = \begin{cases} 2 & \text{for } \omega < -\omega_c \\ 0 & \text{for } \omega > -\omega_c \end{cases}$$

Then, $$X_{SSB}(\omega) = \begin{cases} 0 & \text{for } |\omega| < \omega_c \\ M(\omega+\omega_c) & \text{for } \omega < -\omega_c \\ M(\omega-\omega_c) & \text{for } \omega > \omega_c \end{cases}$$

Its sketch has been shown in figure 3.82.

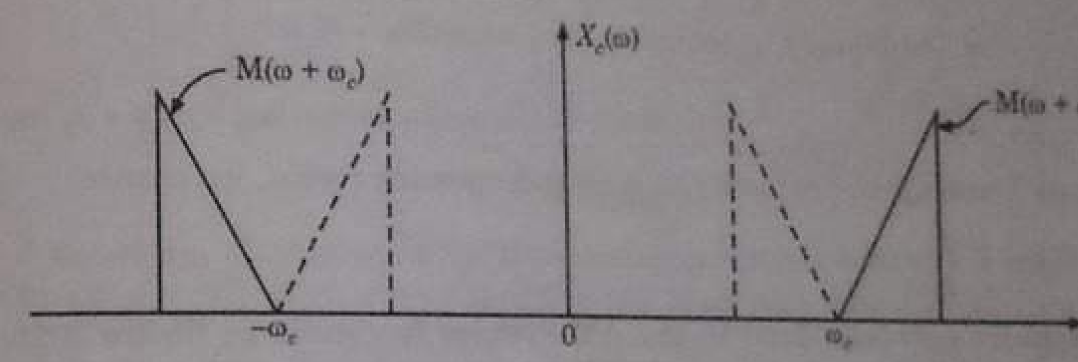


Fig. 3.82.

Its spectrum shows that X_{SSB} is an USB – SSB signal. In a similar fashion, the Fourie transform of equation (ii) yields

$$X_{SSB}(\omega) = \frac{1}{2}M(\omega-\omega_c)[1-sgn(\omega-\omega_c)] + \frac{1}{2}M(\omega+\omega_c)[1+sgn(\omega+\omega_c)]$$

As, $$1-\text{sgn}(\omega-\omega_c) = \begin{cases} 2 & \text{for } \omega < \omega_c \\ 0 & \text{for } \omega > \omega_c \end{cases}$$

and $$1+\text{sgn}(\omega+\omega_c) = \begin{cases} 2 & \text{for } \omega > -\omega_c \\ 0 & \text{for } \omega < -\omega_c \end{cases}$$

Example 3.73. The frequency response $H(\omega)$ of a VSB filter has been shown in figure 3.83.

(a) Find the VSB signal $x_{VSB}(t)$ when

$$m(t) = a_1\cos\omega_1 t + a_2\cos\omega_2 t$$

(b) Show that $x_{VSB}(t)$ can be demodulated by a synchronous demodulator.

(Vinayaka University, 200

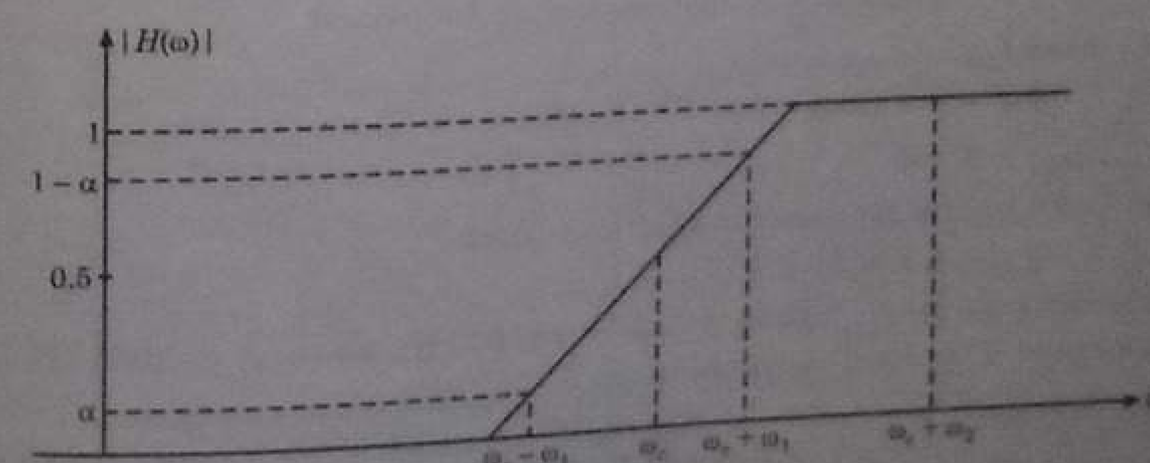


Fig. 3.83.

Solution: (a) We have

(i) $x_{DSB}(t) = m(t)\cos\omega_c t = (a_1\cos\omega_1 t + a_2\cos\omega_2 t)\cos\omega_c t$

$$\text{or}\quad x_{DSB}(t) = \frac{1}{2}a_1\cos(\omega_c - \omega_1)t + \frac{1}{2}a_1\cos(\omega_c + \omega_1)t + \frac{1}{2}a_2\cos(\omega_c - \omega_2)t + \frac{1}{2}a_2\cos(\omega_c - \omega_2)t$$

These sinusoids are transmitted through $H(\omega)$, shown in figure 3.85, which has gains of 0, α, $1-\alpha$, and 1 at $\omega_c - \omega_2$, $\omega_c - \omega_1$, $\omega_c + \omega_1$, and $\omega_c + \omega_2$, respectively. Thus, the VSB filter output $x_{VSB}(t)$ is

$$x_{VSB}(t) = \frac{1}{2}a_1\alpha\cos(\omega_c - \omega_1)t + \frac{1}{2}a_1(1-\alpha)\cos(\omega_c + \omega_1)t + \frac{1}{2}a_2\cos(\omega_c - \omega_2)t \quad ...(i)$$

(*b*) Referring to figure 3.85, we get

$$d(t) = x_{VSB}(t)\cos\omega_c t$$

$$= \frac{1}{4}a_1\cos\omega_1 t + a_2\cos\omega_2 t) + \frac{1}{4}[a_1\,\alpha\cos(2\omega_c - \omega_1)t$$

$$+ a_1(1-\alpha)\cos(2\omega_c + \omega_1)t + a_2\cos(2\omega_c + \omega_2)t]$$

Using low-pass filtering to eliminate the double-frequency terms, we obtain

$$y(t) = \frac{1}{4}(a_1\cos\omega_1 t + a_2\cos\omega_2 t) = \frac{1}{4}m(t)$$

Example 3.74. A radio receiver used in the AM system is shown in figure 3.84. The mixer translates the frequency f_c to a fixed. If 455 kHz by using a local oscillator of frequency f_{LO}. The broadcase-band frequencies range from 540 to 1600 kHz.

(*a*) Determine the range of tuning that must be provided in the local oscillator (*i*) when f_{LO} is higher than f_c (superheterodyne receiver) and (*ii*) when f_{LO} is lower than f_c.

(*b*) Explain why the usual AM radio receiver uses a superheterodyne system.

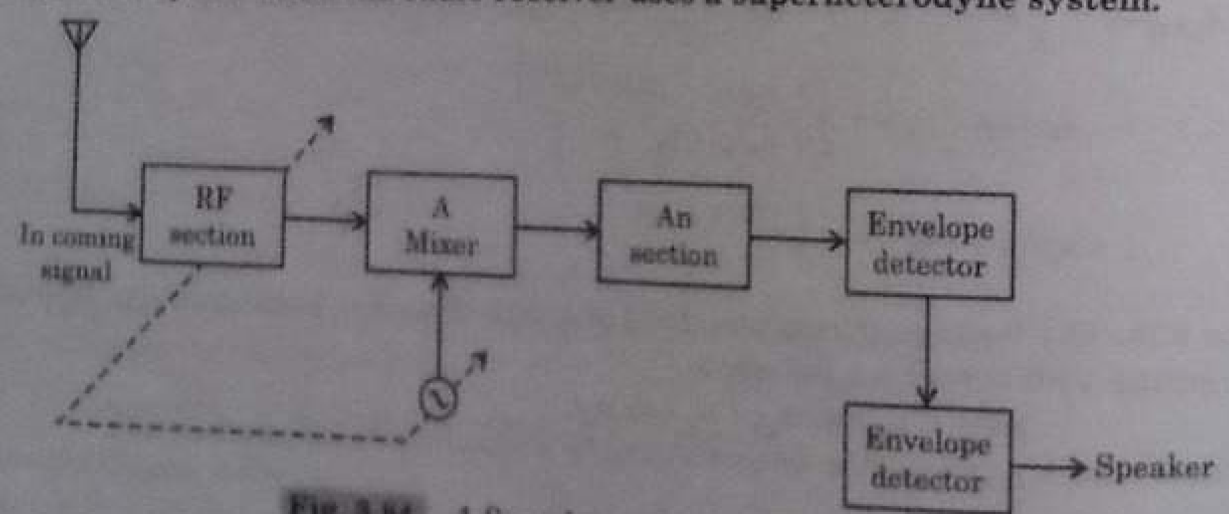


Fig. 3.84. *A Superheterodyne AM receiver.*

Solution: (*a*) (*i*) When $f_{LO} > f_c$

$$540 < f_c < 1600$$

$$f_{LO} - f_c = 455$$

where both f_c and f_{LO} are expressed in kilohertz. Thus

$$f_{LO} = f_c + 455$$

When $f_c = 540$ kHz, we get $f_{LO} = 995$ kHz; and $f_c = 1600$ kHz, we get $f_{LO} = 2055$ kHz. Thus, the required tuning range of the local oscillator is 995 – 2055 kHz.

(*ii*) When $f_{LO} < f_c$

$$f_{LO} = f_c - 455$$

When $f_c = 540$ kHz, we get $f_{LO} = 85$ kHz; and when $f_c = 1600$ kHz, we get $f_{LO} = 1145$ kHz. Thus, the required tuning range of the local oscillator for this case is 85 – 1145 kHz.

(*b*) The frequency ratio, that is, the ratio of the highest f_{LO} to the lowest f_{LO} is 2.07 for case (*i*) and 13.47 for case (*ii*). It is much easier to design an oscillator that is tunable over a smaller frequency ratio, that is the reason why the usual AM radio receiver uses the superheterodyne system.

Example 3.75. The spectrum of a message signal $m(t)$ is shown in figure 3.85(*a*). The ensure communication privacy, this signal is applied to a system (known as a scrambler) shown in figure 3.85(*b*). Analyze the system and sketch the spectrum of the output $x(t)$.

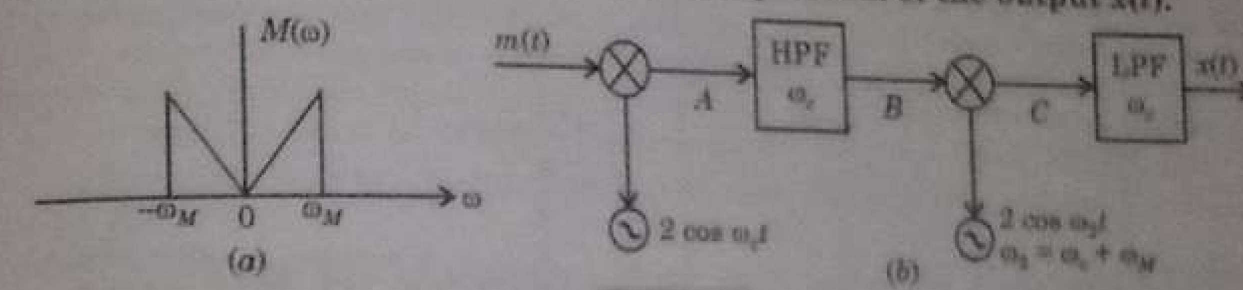


Fig. 3.85.

Solution: The spectrum of the signal at each point is shown in figure 3.86. We see that the spectrum of the output $x(t)$, $X(\omega)$, consists of the two reversed lobes of $M(\omega)$.

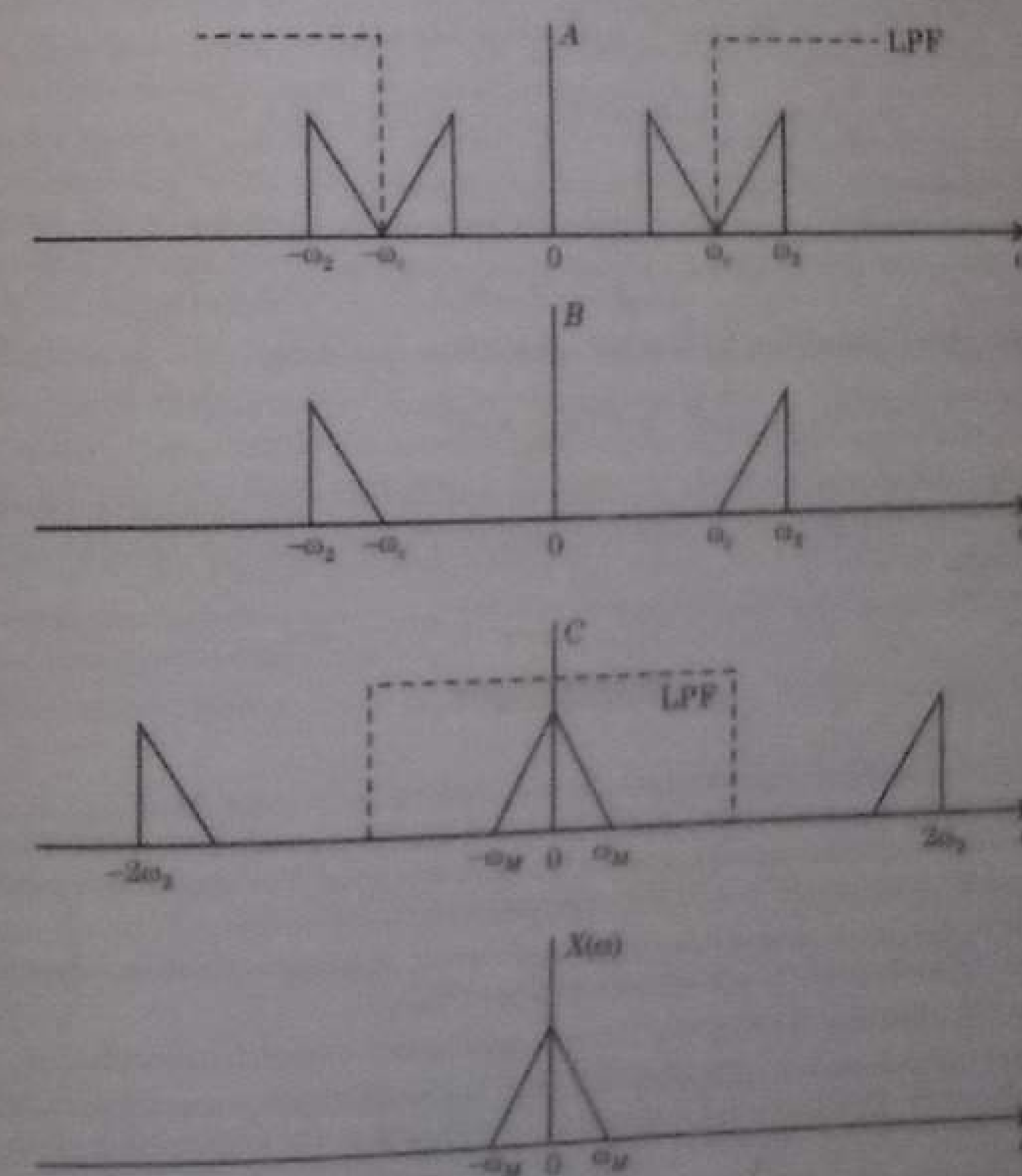


Fig. 3.86

Example 3.76. Using the orthogonality of sine and cosine makes it possible to transmit and receive two different signals simultaneously on the same carrier frequency. A scheme for doing this, known as *quadrature multiplexing*, or *quadrature, amplitude modulation* (QAM).

is shown in figure 3.87. Show that each signal can be recovered by synchronous detection of the received signal by using two local carriers of same frequency but in phase quadrature.

(U.P.S.C., I.E.S., Examination, 1998)

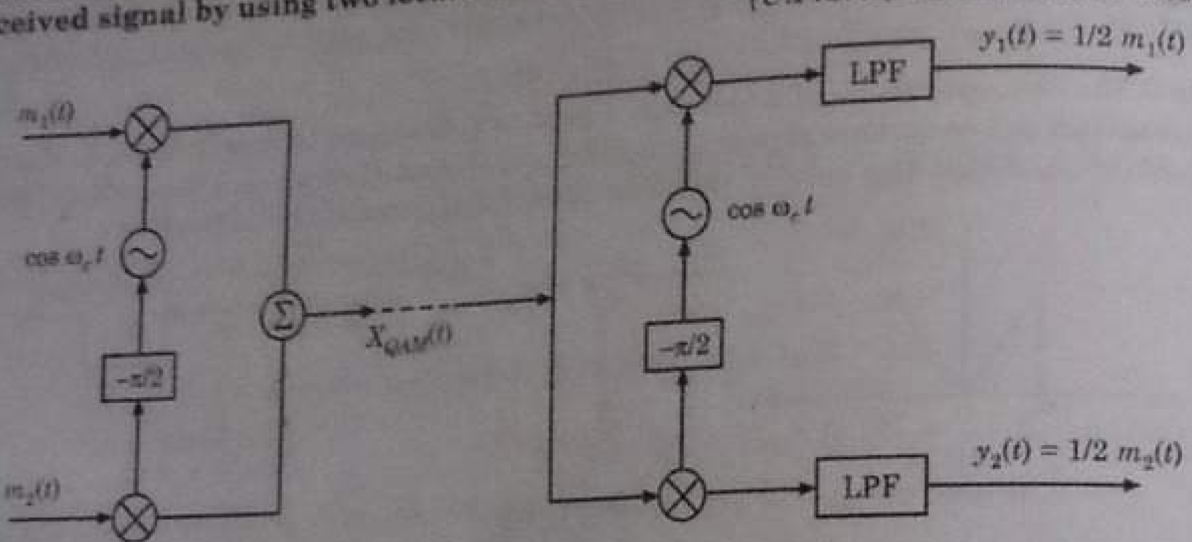


Fig. 3.87. *Quadrature multiplexing system.*

$$x_{QAM}(t) = m_1(t) \cos \omega_c t + m_2(t) \sin \omega_c t$$

$$x_{QAM}(t) \cos \omega_c t = m_1(t) \cos^2 \omega_c t + m_2(t) \sin \omega_c t \cos \omega_c t$$

$$= \frac{1}{2} m_1(t) + \frac{1}{2} m_1(t) \cos 2\omega_c t + \frac{1}{2} m_2(t) \sin 2\omega_c t$$

$$x_{QAM}(t) \sin\omega_c t = m_1(t) \cos \omega_c t \sin \omega_c t + m_2(t) \sin^2 \omega_c t$$

$$= \frac{1}{2} m_1(t) \sin 2\omega_c t + \frac{1}{2} m_2(t) - \frac{1}{2} m_2(t) \cos 2\omega_c t$$

All terms at $2\omega_c$ are filtered out by the low-pass filter, yielding

$$y_1(t) = \frac{1}{2} m_1(t) \quad \text{and} \quad y_2(t) = \frac{1}{2} m_2(t)$$

Note that quadrature multiplexing is an efficient method of transmitting two messages signals within the same bandwidth. It is used in the transmission of colour information signals in commercial television broadcasts.

SUMMARY

1. Modulation is a fundamental requirement of a communication system. Modulation may be defined as a process by which some characteristic of a signal known as carrier is varied according to the instantaneous value of another signal known as modulating signal. The signals containing intelligence or information to be transmitted are called modulating signals.
2. Also the carrier frequency is greater than the modulating frequencies and the signal which results from the process of modulation is known as modulated signal.
3. Modulation may be classified as continuous wave modulation and pulse modulation.
4. If the carrier waveform is continuous in nature then the modulation process is called as continuous-wave (*CW*) modulation. The examples of this type of modulation are amplitude modulation and angle modulation.
5. On the other hand, if the carrier waveform is a pulse-type waveform, then the modulation process is called as pulse modulation. The examples of this type of modulation are Pulse Amplitude Modulation (*PAM*), Pulse Width Modulation (*PWM*), Pulse Code Modulation (*PCM*) etc.
6. Amplitude Modulation and Angle Modulation are the two families of continuous-wave (*CW*) modulation systems.

7. In amplitude modulation, the amplitude of a sinusoidal carrier wave is varied in accordance with the baseband (modulating) signal. On the other hand, in angle modulation, the angle of the sinusoidal carrier wave is varied in accordance with baseband signal.
8. Amplitude Modulation may be defined as a system in which the maximum amplitude of the carrier wave is made proportional to the instantaneous value (amplitude) of the modulating or baseband signal.
9. Carrier signal [*i.e.*, $c(t) = A \cos \omega_c t$] is a fixed frequency signal having frequency ω_c. The modulating or baseband signal $x(t)$ contains the information or intelligence to be transmitted. In the process of amplitude modulation, this information is superimposed upon the carrier signal in the form of amplitude variations of the carrier signal. This means that the information to be transmitted, is now, contained in the amplitude variations of the carrier signal. In other words, in amplitude modulation, the information is transmitted in the form of amplitude variations of the carrier signal.
10. The resulting signal from the process of amplitude-modulation is called amplitude modulated signal or simply AM wave.
11. In the process of amplitude modulation, the frequency and phase of the carrier remain constant whereas the maximum amplitude varies according to the instantaneous value of the information signal.
12. The *AM* wave has a time-varying amplitude called as the envelope of the *AM* wave. Figure 3.1(c) shows that the envelope of AM wave consists of the modulating or baseband signal.
13. This means that the unique property of AM is that the envelope of the modulated carrier has the same shape as the message signal or baseband signal.
14. For positive frequencies, a portion of the spectrum of *AM* wave is lying above the carrier frequency w_c. This band of frequency which is lying above the carrier frequency w_c is known as the upper sideband (USB) whereas the symmetrical portion below carrier frequency w_c is known as the lower sideband (LSB). For negative frequencies, the upper sideband (USB) is represented by the portion of the spectrum below-w_c and the lower sideband by the symmetrical portion above $-w_c$.
15. We generally keep $w_c > w_m$ which ensures that the two sidebands do not overlap each other.
16. The negative frequencies appeared in spectrum analysis is due to the exponential representation of Fourier Transform. These negative frequencies are used for mathematical convenience. For a real function, these negative frequencies have no meaning. For general purpose, it is sufficient to consider only the positive frequency region while treating the negative frequency region as a replica of the positive frequencies. This means that to calculate the BW of a signal, we consider only positive side.
17. For positive side, the highest frequency component present in the spectrum of AM wave is $w_c + w_m$ and the lowest frequency component is $w_c - w_m$.
18. The difference between these two extreme frequencies is equal to the bandwidth of the AM wave. Therefore,

    Bandwidth $\qquad B = (w_c + w_m) - (w_c - w_m) = 2w_m$
19. The bandwidth of the amplitude modulated wave is twice the highest frequency present in the baseband or modulating signal.
20. In *AM* system the modulation index is defined as the measure of extent of amplitude variation about an unmodulated maximum carrier. It is represented by m_a.

    Mathematically,

    Modulation Index $\qquad m_a = \dfrac{|x(t)|_{max}}{\text{maximum carrier amplitude}}$

    or Modulation Index $\qquad m_a = \dfrac{|x(t)|_{max}}{A}$
21. The baseband or modulating signal will be preserved in the envelope of the AM signal only if we have

    $$|x(t)|_{max} \leq A$$

    i.e. modulation index m_a is less than or equal to unity.

22. If $m_a > 1$ or the per centage modulation is greater than 100, the baseband signal is not preserved in the envelope. It means that in this case, the baseband signal recovered from the envelope will be distorted. This type of distortion is called envelope distortion and the AM signal with $m_a > 1$ or $m_a > 100\%$ is called overmodulated signal.
23. In which the modulating or baseband signal consists of only one (single) frequency *i.e.* modulation is done by a single frequency or tone. This type of amplitude modulation is known as single tone amplitude modulation.
24. The total power *P* of the AM wave is the sum of the carrier power P_c and sideband power P_s.
25. The total modulated power of an *AM* signal is expressed as

$$P_t = P_c + P_s = \frac{1}{2}[A^2 + \overline{x^2(t)}]$$

26. In AM wave, the amount of useful message power P_s may be expressed by a term known as transmission efficiency η.
27. Hence transmission efficiency of *AM* wave may be defined as the per centage of total power contributed by the sidebands.
28. Mathematically,

$$\text{Transmission Efficiency,} \quad \eta = \frac{P_s}{P_t} \times 100 = \frac{\frac{1}{2}\overline{x^2(t)}}{\frac{1}{2}A^2 + \frac{1}{2}\overline{x^2(t)}} \times 100 = \frac{100\,\overline{x^2(t)}}{A^2 + \overline{x^2(t)}}$$

29. The maximum transmission efficiency of the AM is only 33.33%. This implies that only one-third of the total power is carried by the sidebands and the rest two-third is wasted.
30. In *AM*, it is generally more convenient to measure the AM transmitter current than the power. In this case, the modulation index may be calculated from the values of unmodulated and modulated currents in the *AM* transmitter.
31. The device which is used to generate an Amplitude-modulated (AM) wave is known as Amplitude Modulator.

    The methods of AM generation may be broadly classified as follow:
    (*i*) Low-level AM Modulation
    (*ii*) High-level AM Modulation
32. Square-low diode modulation and switching modulation are examples of low-level modulation.
33. In a practical modulation circuit, the frequencies other than the carrier and two sideband frequencies are rejected with the help of a tuned circuit.
34. Square law diode modulation circuit make use of nonlinear current-voltage characteristics of diode. This method is suited at low voltage levels because of the fact that current-voltage characteristic of a diode is highly nonlinear particularly in the low voltage region.
35. The process of extracting a modulating or baseband signal from the modulated signal is called demodulation or detection. In other words, demodulation or detection is the process by which the message is recovered from the modulated signal at receiver. The devices used for demodulation or detection are called demodulators or detectors.
36. For amplitude modulation, detectors or demodulators are categorized as:
    (*i*) Square-Law detectors
    (*ii*) Envelope detectors
37. AM signal with large carrier are detected by using the envelope detector. The envelope detector uses the circuit which extracts the envelope of the AM wave. Infact, the envelope of the AM wave is the baseband or modulating signal. But a low-level amplitude modulated signal can only be detected by using square-law detectors in which a device operating in the non-linear region is used to detect the modulating signal.

38. The square law detector circuit is used for detecting modulated signal of small magnitude (*i.e.*, below 1 volt) so that the operating region may be restricted to the non-linear portion of the *V-I* characteristics of the device.
39. It is a known fact that a diode operating in a linear region of its *V-I* characteristics can extract the envelope of an AM wave. This type of detector is known as envelope detector or a linear detector. Envelope detector is most popular in commercial receiver circuits since it is very simple and is not expensive, also at the same time, it gives satisfactory performance for the reception of broadcast programmes.
40. The time constant RC cannot be chosen either too high or too low. If the time constant RC is quite low, the discharge curve during the non-conducting period is almost vertical which results in large fluctuations in a output voltage. Whereas, if the time constant RC is very large, the discharge curve is almost horizontal and it then misses several peaks of the rectified output voltage during negative peaks.
41. One source of distortion in linear diode detector has already been discussed, namely distortion due to improper selection of time constant RC. If RC is too low, removal of radio frequency components is inadequate and the output voltage waveform is spiky in nature. On the other hand, if RC is too large, distortion may result due to clipping during the negative peaks of the modulation wave.
42. The modulated signal, which contains no carrier but two sidebands is called Double-Sideband Suppressed Carrier (DSB-SC) modulation.
43. A balanced modulator may be defined as a circuit in which two non-linear devices are connected in a balanced mode to produce a DSB-SC signal.
44. Ring modulator is another product modulator, which is used to generate DSB-SC signal.
45. The DSB-SC signal may be demodulated by following two methods:
    (*i*) Synchronous detection method
    (*ii*) Using envelope detector after carrier re-insertion.
46. A small amount of carrier signal known as pilot carrier is transmitted alongwith the modulated signal from the transmitter. This small amount of carrier signal is called pilot carrier.
47. Following are the properties of Hilbert transform:
    (*i*) A signal *x*(*t*) and its Hilbert transform $x_h(t)$ have the same energy density spectrum.
    (*ii*) A signal *x*(*t*) and its Hilbert transform $x_h(t)$ have the same autocorrelation function.
    (*iii*) A signal *x*(*t*) and its Hilbert transform $x_h(t)$ are mutually orthogonal.

    $$\text{Mathematically,} \quad \int_{-\infty}^{\infty} x(t)x_h(t)\,dt = 0$$

    (*iv*) If $x_h(t)$ is a Hilbert transform of *x*(*t*), then the Hilbert transform of $x_h(t)$ is $-x(t)$, *i.e.*,

    $$\text{If} \quad H[x(t)] = x_h(t) \quad \text{then} \quad H[x_h(t)] = -x(t)$$

    Here '*H*' denotes the Hilbert transform.
48. SSB-SC signals may be generated by two methods as under:
    (*i*) Frequency discrimination method or filter method
    (*ii*) Phase discrimination method or phase-shift method
49. The phase-shift method avoids filter. This method makes use of two balanced modulators and two phase-shifting networks.
50. The demodulation or detection of AM signal is simpler than DSB-SC and SSB systems. The conventional AM can be demodulated by rectifier or envelope detector. Detection of DSB-SC and SSB is rather difficult and expensive alsop. Furthermore, it is quite easier to generate conventional AM signals at high power levels as compared to DSB-SC and SSB signals. For this reason, conventional AM systems are used for broadcasting purpose.

51. The advantage of DSB-SC and SSB systems over conventional AM system is that the former requires lesser power to transmit the same information.
52. SSB scheme needs only one-half of the bandwidth required in DSB-SC system and less than that required in VSB also. Thus, we can say that SSB modulation scheme is the most efficient scheme among DSB-SC and VSB schemes. SSB modulation scheme is used for long distance transmission of voice signals because it allows longer spacing between repeaters.
53. A device that performs the frequency translation of a modulated signal is known as a *frequency mixer*. The operation is often called *frequency mixing, frequency conversion*, or *heterodyning*.
54. Multiplexing is a technique in which several message signals are combined into a composite signal for transmission over a common channel. In order to transmit a number of these signals over the same channel, the signals must be kept apart so that they do not interfere with each other, and hence they can be separated easily at the receiver end.

REVIEW QUESTIONS

1. What is meant by the term amplitude modulation.
2. Define the term modulation index for AM.
3. Derive an expression for single-tone amplitude-modulated wave.
4. Draw the single-sided frequency spectrum for a single tone amplitude modulated signal.
5. Derive the power relations for single-tone amplitude-modulated wave.
6. Derive the current relations for single-tone amplitude modulated wave.
7. Derive an expression for the transmission efficiency of AM wave.
8. Explain the low-level and high-level AM modulation methods with help of figures.
9. Explain the square-law diode modulation method for AM generation.

SHORT QUESTIONS WITH ANSWERS

Q.1. Write the expression for AM wave?

Ans. The standard equation for amplitude modulated (AM) wave may be expressed as

$$s(t) = x(t)\cos\omega_c t + A\cos\omega_c t$$

or

$$s(t) = [A + x(t)]\cos\omega_c t$$

Q.2. What is the envelope of AM wave?

Ans. The expression for *AM* wave is

$$s(t) = [A + x(t)]\cos\omega_c t$$

or

$$s(t) = E(t)\cos\omega_c t$$

where

$$E(t) = A + x(t)$$

$E(t)$ is called the envelope of *AM* wave. This envelope consists of the baseband signal $x(t)$. Hence, the modulating or baseband signal may be recovered from an AM wave by detecting the envelope.

Q.3. What is the BW for AM wave?

Ans. The difference between these two extreme frequencies is equal to the bandwidth of the AM wave.

Therefore,

Bandwidth

$$B = (w_c + w_m) - (w_c - w_m)$$

$$B = 2w_m$$

Q.4. Define modulation index for AM wave in AM system?

Ans. The modulation index is defined as the measure of extent of amplitude variation about an unmodulated maximum carrier. It is represented by m_a.

Mathematically,

Modulation Index $$m_a = \frac{|x(t)|_{max}}{\text{maximum carrier amplitude}}$$

or Modulation Index $$m_a = \frac{|x(t)|_{max}}{A}$$

where $|x(t)|_{max}$ represents the maximum amplitude of modulating signal and A represents the maximum amplitude of carrier signal.

The modulation index is also known as depth of modulation, degree of modulation or modulation factor.

Also, the absolute value of m_a multiplied by 100 is known as per centage modulation.

Q.5. What are the frequency components in an AM wave?

Ans. The AM signal has three frequency components as follow:

(*i*) carrier frequency ω_c having amplitude A

(*ii*) upper sideband $(\omega_c + \omega_m)$ having amplitude $\frac{m_a \cdot A}{2}$

(*iii*) lower sideband $(\omega_c - \omega_m)$ having amplitude $\frac{m_a \cdot A}{2}$

With the help of these frequency components, we can plot the frequency-spectrum of single-tone amplittude modulated (AM) wave. Figure 3.88 shows the one-sided frequency spectrum of single-tone AM wave.

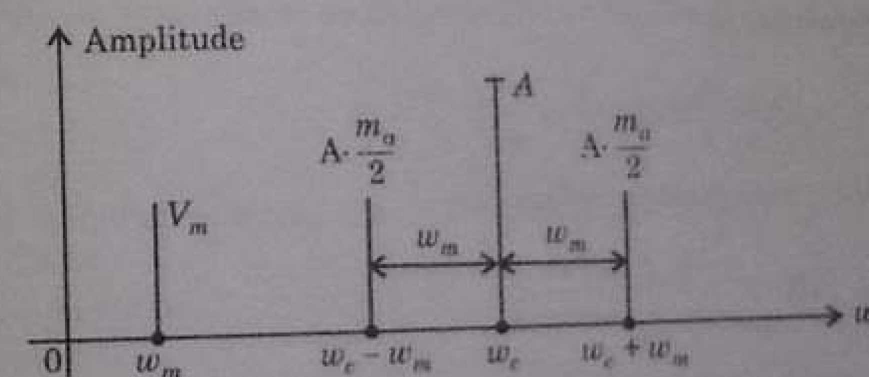


Fig. 3.88. *Single-sided frequency spectrum of single-tone AM wave.*

Q.6. Define transmission efficiency in AM wave?

Ans. In AM wave, the amount of useful message power P_s may be expressed by a term known as transmission efficiency η.

Q.7. What is the current relationship for AM wave?

Ans. $$I_t = I_c\sqrt{1 + \frac{m_a^2}{2}}$$

Q.8. What is demodulation?

Ans. The process of extracting a modulating or baseband signal from the modulated signal is called demodulation or detection. In other words, demodulation or detection is the process by which the message is recovered from the modulated signal at receiver. The devices used for demodulation or detection are called demodulators or detectors. For amplitude modulation, detectors or demodulators are categorized as:

(*i*) Square-Law detectors, (*ii*) Envelope detectors.

Q.9. How can you obtain a DSB-SC signal?

Ans. A DSB-SC signal is obtained by simply multiplying modulating signal $x(t)$ with carrier signal $\cos\omega_c t$. This is achieved by a product modulator. The block diagram of a product modulator is shown in figure 3.89.

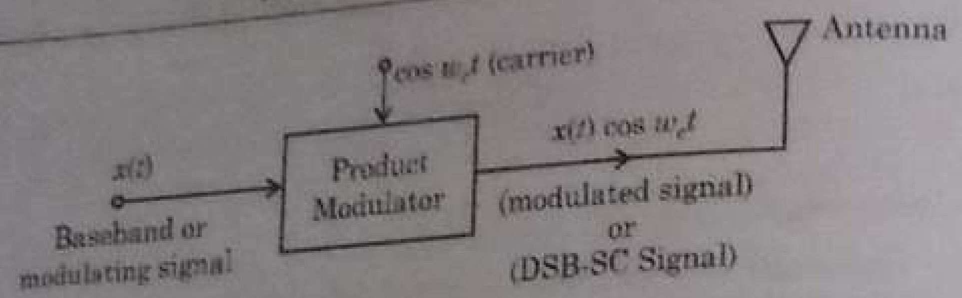


Fig. 3.89.

Q.10. What is the BW of DSB-SC signal

Ans. Bandwidth $B = (w_c + w_m) - (w_c - w_m)$

$= 2\,w_m$

It is obvious that the bandwidth of DSB-SC modulation is same as that of general AM wave.

Q.11. What are the demodulation methods for DSB-SC signal.

Ans. The DSB-SC signal may be demodulated by following two methods:

(*i*) Synchronous detection method

(*ii*) Using envelope detector after carrier reinsertion.

Q.12. What do you mean by Hilbert transfer and inverse Hilbert transform.

(Anna University, Chennai-1999)

Ans. It may be observed that the function $x_h(t)$ obtained by providing $(-\pi/2)$ phase shift to every frequency component present in $x(t)$, actually represents the Hilbert transform of $x(t)$. This means that $x_h(t)$ is the Hilbert transform of $x(t)$ defined as

$$x_h(t) = \frac{1}{\pi}x(t) \otimes \frac{1}{t} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}\,d\tau$$

Also, the inverse Hilbert transform is defined as

$$x(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x_h(\tau)}{t-\tau}\,d\tau$$

Q.13. Write few applications of Hilbert transform.

Ans. (*i*) for generation of SSB signals,

(*ii*) for designing of minimum phase type filters,

(*iii*) for representation of bandpass signals.

Q.14. Write few properties of Hilbert transform.

Ans. Following are the properties of Hilbert transform:

(*i*) A signal $x(t)$ and its Hilbert transform $x_h(t)$ have the same energy density spectrum.

(*ii*) A signal $x(t)$ and its Hilbert transform $x_h(t)$ have the same autocorrelation function.

(*iii*) A signal $x(t)$ and its Hilbert transform $x_h(t)$ are mutually orthogonal.

Mathematically,

$$\int_{-\infty}^{\infty} x(t)\,x_h(t)\,dt = 0$$

(*iv*) If $x_h(t)$ is a Hilbert transform of $x(t)$, then the Hilbert transform of $x_h(t)$ is $-x(t)$, *i.e.*,

If $\quad H[x(t)] = x_h(t)$

then $\quad H[x_h(t)] = -x(t)$

Here 'H' denotes the Hilbert transform.

Q.15. What are the generating methods for SSB-SC signal.

Ans. SSB-SC signals may be generated by two methods as under:

(*i*) Frequency discrimination method or filter method

(*ii*) Phase discrimination method or phase-shift method

Q.16. What do you mean by multiplexing? *(Anna University, Chennai-2000)*

Ans. Multiplexing is a technique in which several message signals are combined into a composite signal for transmission over a common channel. In order to transmit a number of these signals over the same channel, the signals must be kept apart so that they do not interfere with each other, and hence they can be separated easily at the receiver end.

NUMERICAL PROBLEMS

1. An audio signal whose mathematical description is 25 sin (2π 1000t) modulates a carrier described as 75 sin (2π 150,000 t)
   (*a*) Sketch the audio signal.
   (*b*) Sketch the carrier.
   (*c*) Construct the modulated wave showing all amplitude magnitudes.
   (*d*) Calculate the modulation factor and per cent modulation.
   (*e*) What is the frequency of the audio signal of the carrier?
   (*f*) What frequencies would show up in a spectrum analysis of the modulated wave?
   (JNTU, Hyderabad-1998)
   [**Ans.** (*d*) 0.333, 33.33%, (*e*) 1000 Hz, 150,000 Hz, (*f*) 149,000 Hz, 150,000 Hz, 151,000 Hz]

2. An audio signal described as 30 sin (2π 2500 t) amplitude modulates a carrier which is described as 65 sin (2π 250, 000 t)
   (*a*) Sketch the audio signal
   (*b*) Sketch the carrier
   (*c*) Construct the modulated wave
   (*d*) What is the modulation factor and per cent modulation?
   (*e*) What is frequency of the audio signal of the signal?
   (*f*) What frequencies would show up in a spectrum analysis of the modulated wave?
   [**Ans.** (*d*) 0.4615, 46.15%, (*e*) 2500 Hz, 250,000 Hz, (*f*) 247, 500 Hz, 252,500 Hz]

3. A 2000 Hz audio signal having an amplitude of 15 V modulates a 100 kHz carrier which has a peak value of 25 V when not modulated.
   (*a*) Sketch the audio signal to scale.
   (*b*) Sketch the carrier to scale.
   (*c*) Construct the modulation factor and per cent modulation of the modulated wave.
   (*d*) Calculate the modulation factor and per cent modulation of the modulated wave? *(Pune University-1999)*
   (*e*) What frequencies appear in a spectrum analysis of the modulated wave?
   [**Ans.** (*d*) 0.60, 60% ; (*e*) 98 kHz, 100 kHz, 102 kHz]

4. An 1800 Hz signal which has an amplitude of 30 V amplitude modulates a 50 MHz carrier which when the unmodulated has an amplitude of 65 V.
   (*a*) Sketch the modulating signal.
   (*b*) Sketch the carrier.
   (*c*) Construct the modulated wave.
   (*d*) Calculate the modulation factor and per cent modulation.

(e) What frequencies would show up in a spectrum analysis of the AM wave?

(f) Write the trigonometric equations for the carrier and for the audio signal.

[Ans. (d) 0.4614, 46.15% ; (e) 49.9982 MHz, 50.0000 MHz, 50.0018 MHz ; (f) 65 sin $[2\pi (50 \times 10^6)t]$

How many AM broadcast station can be accommodated in a 6 MHz bandwidth if each station transmits a signal which was modulated by an audio signal having a maximum frequency of 5 kHz? [Ans. 600 stations]

A bandwidth of 12 MHz becomes available for assignment. If assigned for TV broadcast service, only ...wo channels could be accommodated. Determine the number of AM stations that could broadcast ...imultaneously if the maximum modulating frequency is limited to 5 kHz. [Ans. 1200 stations]

...90 kHz bandwidth is to accommodate six AM broadcasts simultaneously. What maximum modulating ...equency must each station be limited to? [Ans. 7500 Hz]

...antenna transmits an AM signal having a total power content of 15 kW. Determine the power ...ing transmitted at the carrier frequency and at each of the sidebands when the per cent modulation ...85%. [Ans. 11,019 W, 1990 W]

...lculate the power content of the carrier and of each of the sidebands of an AM signal whose total ...adcast power is 40 kW when the per cent modulation is 60%. [Ans. 33,898, 3050 W]

...500 Hz audio tone amplitude modulates a 200 kHz carrier resulting in a modulated signal having ...er cent modulation of 85%. The total power being transmitted is 15 kW.

What frequencies would appear in a spectrum analysis of the modulated wave?

Determine the power content at each of the frequencies that appear in a spectrum analysis of the modulated wave. [Ans. (a) 196,500 Hz, 200,000 Hz, 203,500 Hz; (b) 11,019 W, 1990 W]

...ermine the power contained at the carrier frequency and within each of the sidebands for an AM ...al whose total power content is 15 kW when the modulation factor is 0.70.

[Ans. 12,048, 1475 W]

...mplitude-modulated signal contains a total of 6 kW. Calculate the power being transmitted at ...arrier frequency and at each of the sidebands when the per cent modulation is 100%.

[Ans. 4000 W, 1000 W]

...M wave has a power content of 1800 W at its carrier frequency. What is the power content of ...of the sidebands when the carrier is modulated 85% ? [Ans. 325 W]

...M signal contains 500 W at its carrier frequency and 100 W in each of its sidebands.

...etermine the per cent modulation of the AM signal.

...ind the allocation of power if the per cent modulation is changed to 60%.

The power content of the carrier of an AM wave does not vary with per cent modulation.

[Ans. (a) 89.44% (b) 500 W, 45 W, 45 W]

...V is contained at the carrier frequency of an AM signal. Determine the power content of each of ...ebands for each of the following per cent modulations: (a) 40%, (b) 50%, (c) 75%, (d) 100%.

The power content of the carrier of an AM wave does not vary with per cent modulation.

[Ans. (a) 89.44%, (b) 75 W, (c) 168.75 W, (d) 300 W]

□□□

CHAPTER 4

Angle Modulation

Inside this Chapter

4.1. Introduction

In last chapter, we discussed the Amplitude (linear) Modulation in which we varied the amplitude of the carrier according to the message signal. Also the transmission bandwidth of AM signal never exceeds twice the maximum frequency component present in AM signal. Linear or amplitude modulation has another important property that the signal-to-noise ratio (SNR) at the destination is no better than one which does not involve modulation (*i.e.*, baseband modulation). The SNR can be improved only be increasing the transmitted power. There is another significant method of modulation used for message transmission known as *angle modulation*. In angle modulation, the frequency or phase of the carrier signal is varied according to the message signal. In this method of modulation, the amplitude of the carrier is kept constant.

4.2. Concept of Angle Modulation Basic Definition

Angle modulation may be defined as the process in which the total phase angle of a carrier wave is varied in accordance with the instantaneous value of the modulating or message signal while keeping the amplitude of the carrier constant.

Mathematical Representation

Let us consider that an unmodulated carrier signal is expressed as

$$c(t) = A\cos(\omega_c t + \theta_0) \quad ...(4.1)$$

where A = amplitude of carrier

ω_c = carrier frequency

θ_0 = some phase angle

Substituting $\omega_c t + \theta_0 = \phi$, we get

$$c(t) = A\cos\phi \quad ...(4.2)$$

where $\phi = (\omega_c t + \theta_0)$ ...(4.3)

is the total phase angle of the carrier signal.

We may consider equation (4.2) as the real part of rotating phasor $A\,e^{j\phi}$. Let us denote this rotating phasor by C, so that

$$C = A\,e^{j\phi} \quad ...(4.4)$$

Also $c(t) = Re\,(A\,e^{j\phi}) = A \cdot Re\,(e^{j\phi}) = A\cos\phi$ ...(4.5)

This phasor C rotates at a constant angular velocity ω_c. Clearly θ_0 is the phase angle of the unmodulated carrier at $t = 0$.

From equation (4.3), the constant angular velocity ω_c of the phasor C is related to its total phase angle 'ϕ' as

$$\phi = \omega_c t + \theta_0$$

Differentiating both sides of this equation with respect to t, we have

$$\omega_c = \frac{d\phi}{dt} \quad ...(4.6)$$

It may be noted at this point that for unmodulated carrier, the derivative $\frac{d\phi}{dt}$ is constant. However, this derivative $\frac{d\phi}{dt}$ may not be a constant with time, in general *i.e.* this may vary with time. Therefore, the angular velocity of the phasor C would also vary with time. Hence, this time-dependent angular velocity or angular frequency is known as **instantaneous angular velocity or instantaneous angular frequency.** This instantaneous angular frequency is denoted by w_i.

Hence, for this case, equation (4.6) becomes

$$\frac{d\phi}{dt} = \omega_i$$

where angular frequency ω_i is time-dependent. ...(4.7)

From equation (4.7), we get

$$\phi = \int \omega_i\, dt \quad ...(4.8)$$

Again writing equation (4.1) of unmodulated carrier as

$$c(t) = A\cos(\omega_c t + \theta_0) = A\cos\phi$$

where ϕ is the total phase angle of the unmodulated carrier expressed as

$$\phi = \omega_c t + \theta_0$$

Now, if this angle ϕ is varied according to the instantaneous value of the message or modulating signal, the carrier signal is then said to be angle modulated.

Types of Angle Modulation

We can vary this phase angle ϕ in two ways and thus there are two types of angle modulation as under:

(*i*) Phase Modulation (PM)

(*ii*) Frequency Modulation (FM)

4.2.1. Phase Modulation (PM)

Definition

Phase modulation (PM) is that type of angle modulation in which the phase angle ϕ is varied linearly with a baseband or modulating signal $x(t)$ about an unmodulated phase angle $(\omega_c t + \theta_0)$. This means that in Phase Modulation, the instantaneous value of the phase angle is equal to the phase angle of the unmodulated carrier $(\omega_c t + \theta_0)$ plus a time-varying component which is proportional to modulating signal $x(t)$.

Mathematical Representation

We know that unmodulated carrier signal is expressed as

$$c(t) = A\cos(\omega_c t + \theta_0) = A\cos\phi$$

where $\phi = \omega_c t + \theta_0$

Neglecting θ_0, we get

total phase angle of unmodulated carrier is

$$\phi = \omega_c t \quad ...(4.9)$$

Now, according to Phase Modulation, this phase angle 'ϕ' is varied linearly with a baseband or modulating signal $x(t)$.

Let the instantaneous value of phase angle be denoted by ϕ_i.

Therefore,

$$\phi_i = \omega_c t + k_p \cdot x(t) \quad ...(4.10)$$

where k_p is the proportionality constant and is known as **phase sensitivity** of the modulator. This is expressed in radians/volts.

Since, the expression for unmodulated carrier wave is

$$c(t) = A\cos\phi$$

Therefore, the expression for phase modulated wave will be

$$s(t) = A\cos\phi_i \quad ...(4.11)$$

Putting the value of ϕ_i in equation (4.11) from equation (4.10), we get

$$s(t) = A\cos[\omega_c t + k_p \cdot x(t)] \quad ...(4.12)$$

which is the required mathematical expression for a phase modulated wave.

4.2.2. Frequency Modulation

Definition *(U.P. Tech., Semester, Exam., 2004-05)* (4 marks)

Frequency modulation is that type of angle modulation in which the instantaneous frequency ω_i is varied linearly with a message or baseband signal $x(t)$ about an unmodulated carrier frequency ω_c. This means that the instantaneous value of the angular frequency ω_i will be equal to the carrier frequency ω_c plus a time-varying component proportional to the baseband signal $x(t)$.

Mathematical, Representation

We know that the instantaneous frequency is given by

$$\omega_i = \omega_c + k_f \cdot x(t) \quad ...(4.13)$$

where k_f is proportionality constant and is known as the **frequency sensitivity** of the modulator. This is expressed in Hz/volt.

Now, let the expression for unmodulated carrier signal be

$$c(t) = A\cos(\omega_c t + \theta_0) \quad ...(4.14)$$

or

$$c(t) = A\cos\phi \quad ...(4.15)$$

where

$$\phi = \omega_c t + \theta_0 \quad ...(4.16)$$

'ϕ' is the total phase angle of the unmodulated carrier.

Let ϕ_i be the instantaneous phase angle of the modulated signal.

From equation (4.15), the equation for unmodulated carrier is

$$c(t) = A\cos\phi$$

On frequency modulation amplitude A must remain constant and only angle 'ϕ' will change.

Hence, the expression for frequency modulated wave will be

$$s(t) = A\cos\phi_i \quad ...(4.17)$$

where ϕ_i = instantaneous phase angle

From equation (4.16)

$$\phi = \omega_c t + \theta_0$$

On differentiation, we get

$$\frac{d\phi}{dt} = \omega_c$$

or

$$\phi = \int \omega_c dt \quad ...(4.18)$$

Based on equation (4.18), we may write the expression for instantaneous phase angle ϕ_i as

$$\phi_i = \int \omega_i\, dt \quad ...(4.19)$$

where ω_i = instantaneous frequency of frequency modulated wave.

Putting the value of w_i in equation (4.19) form equation (4.13), we get

$$\phi_i = \int [\omega_c + k_f x(t)]\, dt = \omega_c t + k_f \int x(t)\, dt \quad ...(4.20)$$

Putting this value of ϕ_i in equation (4.17), we get the expression for frequency modulated wave will be

$$s(t) = A\cos\left[\omega_c t + k_f \int x(t)\, dt\right] \quad ...(4.21)$$

Now, if the phase angle of the unmodulated carrier is taken at $t = 0$, then the limit of integration in equation (4.21) will be 0 to t.

In this case the expression for FM wave will be

$$s(t) = A\cos\left[\omega_c t + k_f \int_0^t x(t)\, dt\right]$$

which is the required **general expression for FM wave.**

4.3. Frequency Deviation

(U.P. Tech., Semester, Exam., 2003-04) (05 marks)

We know that the instantaneous frequency of FM wave is given as

$$\omega_i = \omega_c t + k_f \cdot x(t) \quad ...(4.22)$$

The instantaneous frequency of FM signal varies with time around the carrier frequency ω. This means that the instantaneous frequency of FM signal varies according to the modulating signal.

The maximum change in instantaneous frequency form the average frequency ω_c is called **frequency deviation.***

* The frequency deviation is a useful parameter for determining the bandwidth of FM signals.

This maximum change in instantaneous frequency ω_i from the average or carrier frequency ω_c depends upon the magnitude and sign of $k_f \cdot x(t)$. This means that the frequency deviation would be either positive or negative depending upon the sign of $k_f \cdot x(t)$. However, the amount of frequency deviation in both these cases is given by the maximum magnitude i.e. $|k_f \cdot x(t)|_{max}$.

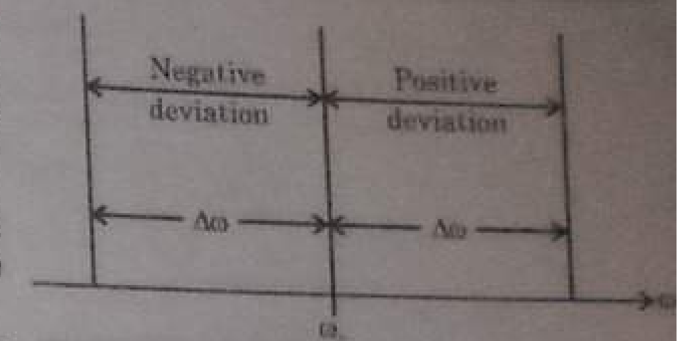


Fig. 4.1. *Illustration of frequency deviation*

Maximum frequency deviation is generally denoted by $\Delta\omega$.

Therefore,

Frequency deviation $\Delta\omega = |k_f \cdot x(t)|_{max}$

Figure 4.1 further illustrates the concept of frequency deviation.

4.4. Relationship between Phase Modulation (PM) and Frequency Modulation (FM)

(U.P. Tech., Semester, Exam., 2004-05) (06 marks)

We know that an angle modulated wave is given as

$$s(t) = A\cos\phi_i \quad ...(4.2$$

where A = Amplitude

ϕ_i = Instantaneous total phase angle of the angle modulated wave.

The expression for phase modulated (PM) wave is

$$s(t) = A\cos[\omega_c t + k_p x(t)] \quad ...(4.2$$

Similarly, the expression for frequency modulated (FM) wave is

$$s(t) = A\left[\cos\omega_c t + k_f \int_0^t x(t)\, dt\right] \quad ...(4.2$$

It may be observed from above equations that Phase Modulation (PM) and Frequency Modulati (FM) are closely related to each other because in both the cases there is a variation in the total pha angle. In Phase Modulation (PM), the phase angle varies linearly with baseband signal $x(t)$ whereas case of Frequency Modulation (FM), the phase angle varies linearly with the integral of baseba signal $x(t)$. This means that FM wave may be obtained by using PM. Conversely, PM wave may obtained by using FM.

To get FM by using PM, we first integrate the baseband signal and then apply to the pha modulator. This process is illustrated with the help of a block diagram shown in figure 4.2.

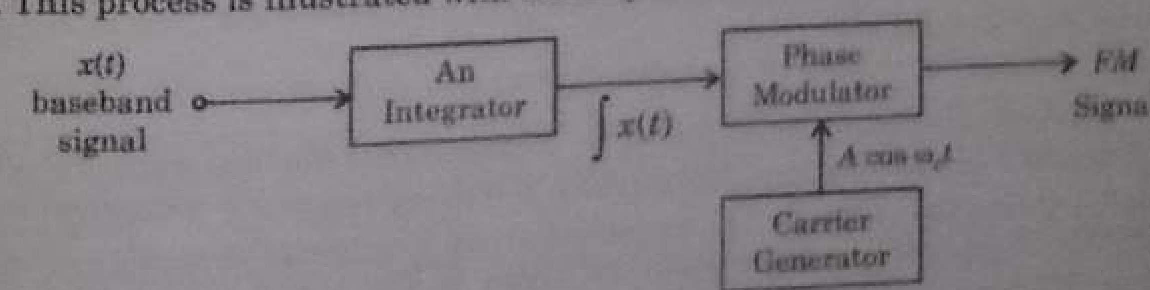


Fig. 4.2. *Generation of FM using phase modulator.*

Similarly, PM wave may be generated by using frequency modulator by first differentiati modulating or baseband signal $x(t)$ and then applying to the frequency modulator. This process illustrated with the help of a block diagram shown in figure 4.3.

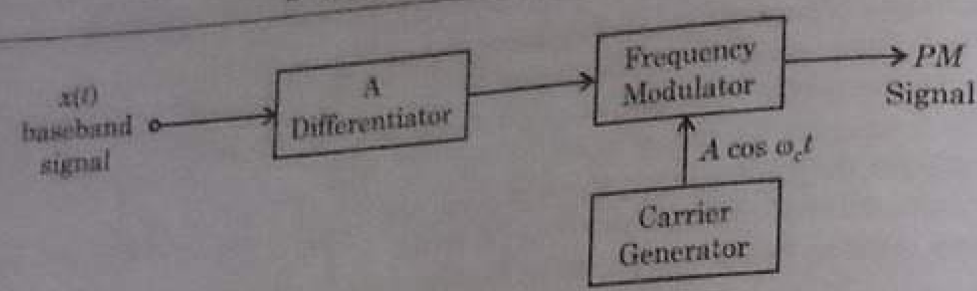


Fig. 4.3. *Generation of PM using frequency modulator.*

4.5. Single Tone Frequency Modulation

Definition

We know that the general expression for FM wave is given as

$$s(t) = A\cos\left[\omega_c t + k_f \int_0^t x(t)\,dt\right] \quad \text{...(4.26)}$$

From this equation, it is clear that the FM wave $s(t)$ is a nonlinear function of the baseband or modulating signal $x(t)$. This means that frequency modulation is a **nonlinear modulation process.** Due to this reason, unlike amplitude modulation, the spectrum of an FM wave is related in a complex manner with the modulating or baseband signal.

Therefore, to study the spectrum or spectral properties of FM signal, we shall consider the simplest case of FM wave known as **single tone frequency modulation.**

In general expression of FM wave equation (4.26), we consider a modulating or baseband signal $x(t)$ which may consist of any number of frequency components.

On the other hand, a single tone frequency modulation is that type of frequency modulation (*FM*) in which the modulating or baseband signal contains a single frequency.

Let us consider a carrier signal as

$$c(t) = A\cos\omega_c t$$

Let the modulating signal be

$$x(t) = V_m \cos\omega_m t^*$$

where V_m = maximum amplitude of modulating signal

ω_m = frequency of modulating signal

Then according to the definition of *FM*, the frequency of carrier signal is varied according to the instantaneous value of the modulating or baseband signal, $x(t)$. Figure 4.4 illustrates the process of single tone frequency modulation.

Few Points

(*i*) Figure 4.4 shows the *a* single-tone FM wave. It is clear that the frequency of carrier is varied according to the modulating signal keeping the maximum amplitude of the carrier as constant.

(*ii*) Figure 4.4 shows the variation of instantaneous carrier frequency ω_i with time around the unmodulated carrier frequency or centre frequency ω_c. This variation of instantaneous carrier frequency is proportional to the modulating signal. This means that the instantaneous carrier frequency varies according to the shape of modulating or baseband signal.

(*iii*) From figure 4.4 it may also be observed that when the modulating or message signal is applied, the instantaneous carrier frequency deviates up and down from its centre frequency ω_c. This change or shift either above or below the resting or centre frequency is called frequency deviation denoted by $\Delta\omega$.

* Note that here we have considered the modulating signal $x(t)$ as a voltage signal $v(t)$.

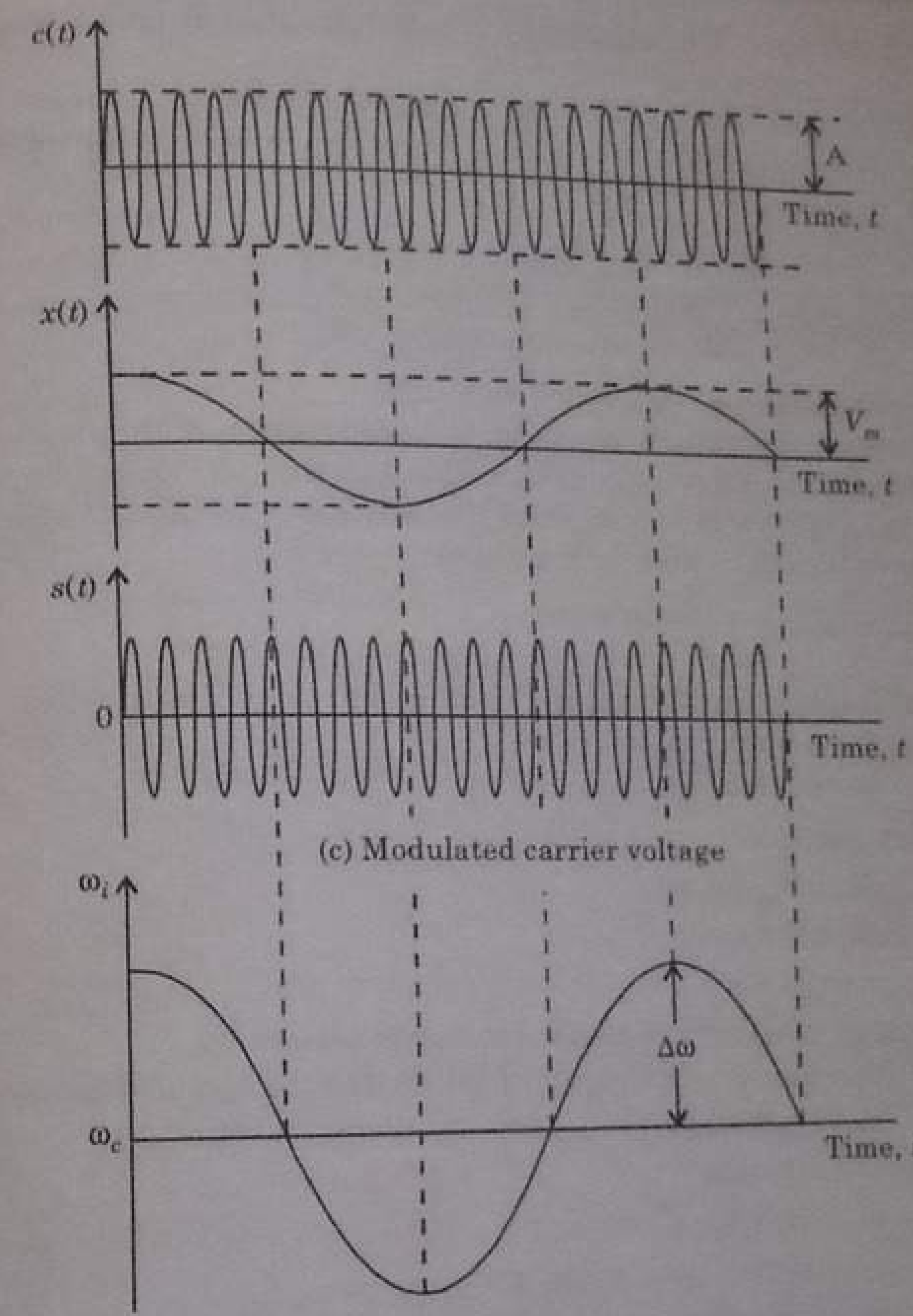


Fig. 4.4. *Illustration of FM (a) Unmodulated carrier (b) Modulating signal (c) FM signal (d) Instantaneous carrier frequency.*

(*iv*) The total variation in frequency from the lowest to the highest point is called carrier swing. Obviously

The carrier swing $= 2 \times$ frequency deviation $= 2 \times \Delta\omega$

(*v*) The amount of frequency deviation or variation depends upon the amplitude (loudness) of the modulating (audio) signal. This means that louder the sound, greater the frequency deviation and vice versa.

However, in FM broadcast, it has been internationally agreed to restrict maximum deviation to 75 kHz on each side of the centre frequency for sounds of maximum loudness. This means that the sounds of lesser loudness are permitted proportionally less frequency deviation. A maximum frequency deviation of 25 kHz is permitted for the sound portion of television broadcasts.

(*vi*) As stated earlier, the frequency deviation is useful in determining the FM signal bandwidth. Since, a maximum frequency deviation of 75 kHz is allowed for commercial FM broadcast stations using a band of 88 MHz to 108 MHz, therefore approximate FM channel width is $2 \times 75 = 150$ kHz. Allowing a 25 kHz guard band on either side, the channel width becomes

$2(75 + 25) = 200$ kHz. This guardband is meant to prevent interference between adjacent channels.

In FM broadcast, the highest audio frequency transmitted is 15 kHz.

(vii) For FM, the modulation index is defined as the ratio of frequency deviation to the modulating frequency.

Mathematically,

$$\text{Modulation index, } m_f = \frac{\text{Frequency deviation}}{\text{Modulation frequency}} = \frac{\Delta\omega}{\omega_m} \qquad ...(4.27)$$

This modulation index may be greater than unity.

(viii) The term "per cent modulation" as it is used in reference to FM refers to the ratio of actual frequency deviation to the maximum allowable frequency deviation. Thus 100% modulation corresponds to 75 kHz for the commercial FM broadcast band and 25 kHz for the commercial FM broadcast band and 25 kHz for television.

$$\text{Per cent modulation } M = \frac{\Delta f_{actual}}{\Delta f_{max}} \times 100$$

4.5.1. Mathematical Expression for Single-Tone Frequency Modulation

Let the expression for a carrier signal be

$$c(t) = A \cos \omega_c t$$

Let the modulating signal be

$$x(t) = V_m \cos \omega_m t$$

Let the expression for FM wave be

$$s(t) = A \cos \phi_i \qquad ...(4.28)$$

where ϕ_i is the instantaneous phase angle of the modulated wave.

We know that the instantaneous frequency of the modulated signal is given as

$$\omega_i = \omega_c + k_f \cdot x(t) \qquad ...(4.29)$$

Putting the value of $x(t)$, we get

$$\omega_i = \omega_c + k_f \cdot V_m \cos \omega_m t$$

But we know that frequency deviation is given as

$$\Delta\omega = |k_f \cdot x(t)|_{max} = k_f |x(t)|_{max} = k_f \cdot V_m$$

Therefore,

$$\omega_i = \omega_c + \Delta\omega . \cos \omega_m t \qquad ...(4.30)$$

The total phase angle ϕ_i of the modulated wave is given as

$$\phi_i = \int \omega_i \, dt$$

Putting the value of ω_i from equation (4.30), we get

$$\phi_i = \int [\omega_c + \Delta\omega \cos \omega_m t] \, dt = \omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t \qquad ...(4.31)$$

But, modulation index m_f is given as

$$m_f = \frac{\text{Frequency deviation}}{\text{Modulating frequency}} = \frac{\Delta\omega}{\omega_m}$$

Therefore, putting the value of m_f in equation (4.31), we obtain

$$\phi_i = \omega_c t + m_f \sin \omega_m t$$

Substituting this value of ϕ_i in equation (4.28), we get the expression for single-tone FM wave

$$s(t) = A \cos \phi_i = A \cos [\omega_c t + m_f \sin \omega_m t] \qquad ...(4.32)$$

which is the required mathematical expression for single tone FM wave.

Example 4.1. A single-tone FM is represented by the voltage equation as:

$$v(t) = 12 \cos (6 \times 10^8 t + 5 \sin 1250 t)$$

Determine the following:

(i) carrier frequency

(ii) modulating frequency

(iii) the modulation index

(iv) maximum deviation

(v) what power will this FM wave dissipate in 10 Ω resistors.

Solution: We know that the standard expression for a single-tone FM wave is given as

$$v(t) = A \cos (w_c t + m_f \sin w_m t) \qquad ...(i)$$

The given expression is

$$v(t) = 12 \cos (6 \times 10^8 t + 5 \sin 1250 t) \qquad ...(ii)$$

Comparing equations (i) and (ii), we get

(i) carrier frequency

$$w_c = 6 \times 10^8 \text{ rad/sec.}$$

or

$$f_c = \frac{6 \times 10^8}{2\pi} = 95.5 \text{ MHz} \qquad \textbf{Ans.}$$

(ii) modulating frequency

$$w_m = 1250 \text{ rad/sec.}$$

or

$$f_m = \frac{1250}{2\pi} = 199 \text{ Hz} \qquad \textbf{Ans.}$$

(iii)

$$m_f = 5 \qquad \textbf{Ans.}$$

(iv) maximum frequency deviation is given as

$$m_f = \frac{\Delta w}{w_m} = \frac{\Delta_f}{f_m}$$

or

$$\Delta_f = m_f \cdot f_m$$

$$= 5 \times 199 = 995 \text{ Hz} \qquad \textbf{Ans.}$$

(v) The power dissipated is

$$P = \frac{V_{rms}^2}{R} = \frac{(12/\sqrt{2})^2}{10}$$

or

$$P = \frac{72}{10} = 7.2 \text{ watts} \qquad \textbf{Ans.}$$

Example 4.2. A 107. 6 MHz carrier signal is frequency modulated by a 7 kHz since wave. The resultant FM signal has a frequency deviation of 50 kHz. Determine the following:

(i) the carrier swing of the FM signal.

(ii) the highest and the lowest frequencies attained by the modulated signal.

(iii) the modulation index of the FM wave.

Solution: Given that

$$f_c = 107.6 \text{ MHz}$$

$$f_m = 7 \text{ kHz}$$

$$\Delta_f = 50 \text{ kHz}$$

(i) Carrier swing $= 2 \times$ frequency deviation

$$= 2 \times 50 = 100 \text{ kHz} \qquad \textbf{Ans.}$$

(ii) The highest frequency attained by the modulated signal is equal to the resting or carrier frequency plus the frequency deviation, *i.e.*

$$f_H = f_c + \Delta_f = 107.6 \times 10^6 + 50 \times 10^3$$
$$= (107600 \times 10^3) + (50 \times 10^3)$$
or
$$f_H = 107650 \times 10^3 \text{ Hz} = 107.65 \text{ MHz} \quad \textbf{Ans.}$$

Similarly the lowest frequency attained by the modulated signal is equal to the resting or carrier frequency minus the frequency deviation i.e.

$$f_L = f_c - \Delta_f = 107.6 \times 10^6 - 50 \times 10^3)$$
$$= (107600 \times 10^3) - (50 \times 10^3)$$
or
$$f_L = 107550 \times 10^3 \text{ Hz} = 107.55 \text{ MHz} \quad \textbf{Ans.}$$

(*iii*) The modulation index is given as

$$m_f = \frac{\text{frequency deviation}}{\text{modulating frequency}} = \frac{\Delta_f}{f_m}$$

or

$$m_f = \frac{50 \times 10^3}{7 \times 10^3} = 7.143 \quad \textbf{Ans.}$$

Example 4.3. Determine the frequency deviation and carrier swing for a frequency-modulated (FM) signal which has a resting frequency of 105.00 MHz and whose upper frequency is 105.007 MHz when modulated by a particular wave. Find the lowest frequency reached by the FM wave.

Solution: Given that

Carrier frequency $f_c = 105.000$ MHz

$$f_H = 105.007 \text{ MHz}$$

We know that frequency deviation Δ_f is defined as the maximum change in frequency of the modulated signal away from the resting or carrier frequency *i.e.*,

$$\Delta_f = f_H - f_c = (105.007 - 105.000)\, 10^6$$
or
$$\Delta_f = 0.007 \times 10^6 \text{ Hz}$$
or
$$\Delta_f = 7000 \text{ Hz} = 7 \text{ kHz} \quad \textbf{Ans.}$$

Carrier swing is given by

carrier swing = 2 × frequency deviation

$$= 2 \times \Delta_f = 2 \times 7 = 14 \text{ kHz.} \quad \textbf{Ans.}$$

Lowest frequency reached by the modulated wave is given as

$$f_L = f_c - \Delta_f = (105.000 - 0.007) \text{ MHz.}$$
$$= 104.993 \text{ MHz.} \quad \textbf{Ans.}$$

Example 4.4. What is the modulation index of an FM signal having a carrier swing of 100 kHz when the modulating signal has a frequency of 8 kHz?

Solution: Given that

carrier swing = 100 kHz

modulating frequency $f_m = 8$ kHz

Modulation index is given as

$$m_f = \frac{\text{frequency deviation}}{\text{modulating frequency}} = \frac{\Delta_f}{f_m} \quad \text{...(i)}$$

But we know that

carrier swing = $2 \times \Delta_f$

$$\Delta_f = \frac{\text{carrier swing}}{2} = \frac{100}{2} = 50 \text{ kHz}$$

Using equation (*i*) we get

$$m_f = \frac{50}{8} = 6.25 \quad \textbf{Ans.}$$

Example 4.5. An FM transmission has a frequency deviation of 20 kHz.

(*i*) determine the per cent modulation of this signal if it is broadcasted in the 88–108 MHz band.

(*ii*) Calculate the per cent modulation if this signal is broadcasted as the audio portion of a television broadcast.

Solution: Given that

$$\Delta_f = 20 \text{ kHz}$$

(*i*) Per cent modulation for an FM wave is defined as

$$M = \frac{\Delta f_{actual}}{\Delta f_{max}} \times 100$$

Δf_{actual} is given = 20 kHz

The maximum frequency deviation Δf_{max} permitted in the FM broadcast band is 75 kHz.

Thus,
$$M = \frac{20 \times 10^3}{75 \times 10^3} \times 100 = 26.67\% \quad \textbf{Ans.}$$

(*ii*)
$$M = \frac{\Delta f_{actual}}{\Delta f_{max}} \times 100$$
$$\Delta f_{actual} = 20 \text{ kHz}$$

The maximum frequency deviation Δf_{max} permitted for the FM audio portion of a TV broadcast is 25 kHz.

Thus
$$M = \frac{20 \times 10^3}{25 \times 10^3} \times 100 = 80\% \quad \textbf{Ans.}$$

4.6. Phasor Representation of Angle Modulated (*i.e.* FM and PM Waves)

The expression for FM wave is

$$s(t)_{FM} = A \cos\left[w_c t + k_f \int_0^t x(t)\, dt\right] \quad \text{...(4.33)}$$

Similarly, the expression for PM wave is

$$s(t)_{PM} = A \cos[w_c t + k_p \cdot x(t)] \quad \text{...(4.34)}$$

We may consider the signa ls $s(t)_{FM}$ and $s(t)_{PM}$ as real parts of phasors C_{FM} and C_{PM} respectively expressed as

$$C_{FM} = Ae^{j[(w_c t + k_f \int x(t)\, dt)]}$$

and
$$C_{PM} = Ae^{j[w_c t + k_p\, x(t)]}$$

Assuming
$$\int x(t)\, dt = y(t) \quad \text{...(4.35)}$$

So that
$$C_{FM} = Ae^{j[w_c t + k_f\, y(t)]}$$

4.7. Types of Frequency Modulation (FM)

(U.P. Tech., Semester, Exam., 2003-04) (04 marks)

In article 4.3, we discussed that the bandwidth of an *FM* signal depends upon the frequency deviation $\Delta w = k_f\, x(t)$.

Thus, as a matter of fact, if the frequency deviation is high, the bandwidth will be large. Similarly, if the frequency deviation is low, the bandwidth will be small.

Since deviation is given by $\Delta w = k_f\, x(t)$, it means that for any given $x(t)$, the frequency deviation will depend upon frequency sensitivity k_f. Hence, when k_f is quite small then the bandwidth will be narrow and when k_f is large then the bandwidth will be large.

Thus, depending upon the value of frequency sensitivity k_f, *FM* may be divided as under –

(i) Narrowband FM: In this case, k_f is small and hence the bandwidth of *FM* is narrow.*

(ii) Wideband FM: In this case, k_f is large and hence the *FM* signal has a wide bandwidth.**

4.7.1. Narrowband FM

We know that the general expression for FM wave in the phasor form is given by

$$C_{FM} = Ae^{j[w_c t + k_f\, y(t)]} \quad ...(4.36)$$

However, for a narrowband FM, we have

$k_f \times y(t) << 1$ for all the values of t

Thus $e^{jk_f \cdot y(t)} \cong 1 + jk_f \cdot y(t)$

Hence, the FM phasor expression, now, becomes

$$C_{FM} = Ae^{jw_c t} \cdot e^{jk_f \cdot y(t)} = A[1 + jk_f \cdot y(t)]\, e^{jw_c t}$$

Also, since the *FM* signal is the real part of its phasor representation, therefore, we write

$$s(t)_{FM} = Re[C_{FM}] = A \cos w_c t - Ak_f\, y(t) \sin w_c t \quad ...(4.37)$$

In the same way, on the other hand, the expression for narrowband PM is given as

$$s(t)_{PM} = A \cos w_c t - Ak_p\, x(t) \sin w_c t \quad ...(4.38)$$

Thus, the expressions for narrowband FM and PM are quite similar to the expression for AM signal with a little modification. Therefore, the bandwidth of a narrowband FM is almost same as that of the AM signal.

Now, we can also show that the signals $y(t)$ and $x(t)$ have the same bandwidth.

Since

$$y(t) = \int_0^t x(t)\, dt$$

Therefore,

$$F[y(t)] = F\left[\int_0^t x(t)\, dt\right] = \frac{1}{jw} \cdot X(w)$$

or

$$Y(w) = \frac{1}{jw} \cdot X(w) \quad ...(4.39)$$

Conclusion

Hence, from above expression it is clear that if $X(w)$ is band limited to w_m, then $Y(w)$ is also band-limited to w_m. Both FM and PM wave expressions have carrier and sideband terms which are identical to the AM expression except with differing phase relations between carrier and sideband terms.***

4.7.2. Generation of Narrowband FM

We know that the expressions for narrowband FM and narrowband PM are given by

$$s(t)_{FM} = A \cos w_c t - Ak_f \cdot y(t) \sin w_c t \quad ...(4.40)$$

and

$$s(t)_{PM} = A \cos w_c t - Ak_p\, x(t) \sin w_c t \quad ...(4.41)$$

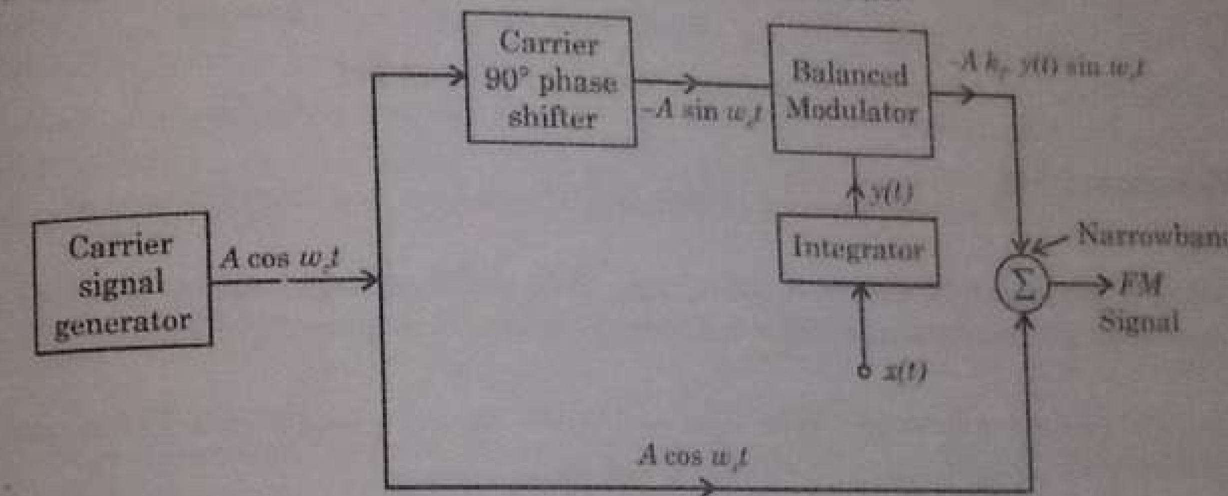
* We shall observe, later on, that the BW of a narrowband FM is almost same as that of AM.

** We shall study, later on, that unlike AM, the bandwidth of wideband FM is too large.

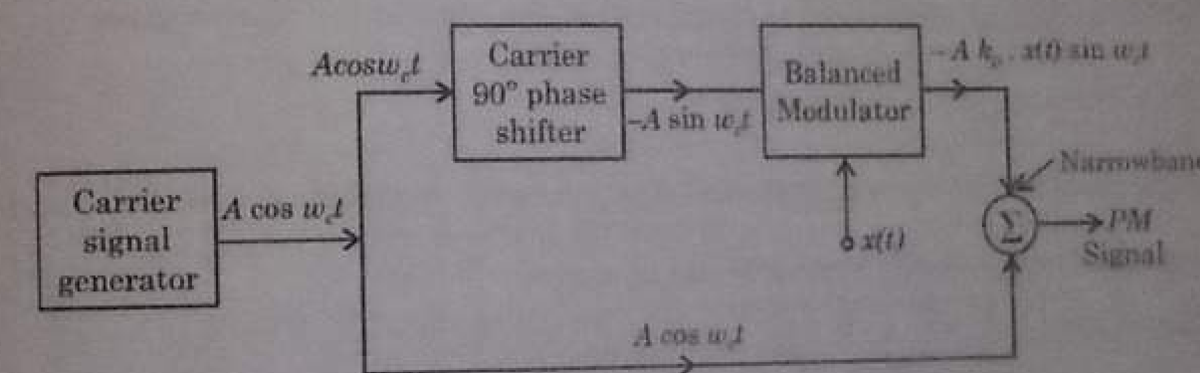
*** In fact, this is the difference that makes the amplitude of FM constant.

The equations (4.40) and (4.41) provide an idea to generate narrowband FM and narrowband PM respectively. This means that the sideband terms are generated by a balanced modulator, as in a DSB-SC system, and then the carrier term is added to sideband terms.

Based upon equations (4.40) and (4.41), figure 4.5 shows the methods to generate narrowband FM and narrowband *PM*. It may be observed that the block diagrams shown in figure 4.5 satisfy the corresponding equations for narrowband *FM* and narrowband *PM*.



(a) Narrowband FM Generator



(b) Narrowband PM Generator.

Fig. 4.5. *Generation schemes for narrowband FM and PM*

4.7.3. Expression for Single-tone Narrowband FM

We know that the general expression for narrowband FM is given as

$$s(t)_{FM} = A \cos w_c t - Ak_f\, y(t) \sin w_c t \quad ...(4.42)$$

Here $y(t) = \int x(t)\, dt$

where $x(t)$ is any modulating signal

and $A \cos w_c t$ is a carrier signal.

For a single-tone narrowband *FM*, we take modulating signal as consisting of a single frequency *i.e.*,

$$x(t) = V_m \cos w_m t^* \quad ...(4.43)$$

where w_m is the single modulating frequency.

Thus

$$y(t) = \int x(t)\, dt$$

$$y(t) = \int V_m \cos w_m t = \frac{V_m}{w_m} \sin w_m t$$

Putting this value of $y(t)$ in the equation (4.42), we get

* Here, we have taken the signal $x(t)$ as a voltage signal.

$$s(t)_{FM} = A\cos w_c t - Ak_f \cdot \frac{V_m}{w_m}\sin w_m t.\sin w_c t$$

But we know that

$$\frac{k_f V_m}{w_m} = m_f = \text{modulation index}$$

Hence $$s(t)_{FM} = A\cos w_c t - Am_f \sin w_m t \sin w_c t \quad ...(4.44)$$

which is the required expression for a single-tone narrowband FM.

4.8. Wideband FM

Definition

When the value of modulation index m_f is quite large, then in FM, a large number of sidebands are produced and hence the bandwidth of FM is sufficiently large. This type of FM system is known as wideband FM.

Now, we shall analyse wideband FM by considering the case of a single-tone FM system.

The expression for a single-tone FM wave is given as

$$s(t) = A\cos(w_c t + m_f \sin w_m t) \quad ...(4.45)$$

This expression may be considered as a real part of the exponential phasor given by

$$C_{FM}(t) = Ae^{j(w_c t + m_f \sin w_m t)}$$

$$C_{FM}(t) = Ae^{jw_c t}\, e^{jm_f \sin w_m t} \quad ...(4.46)$$

In above expression, the second exponential is a periodic function of period $\frac{1}{f_m}$ and may be expanded in the form of a complex Fourier series as under:

$$e^{jm_f \sin w_m t} = \sum_{n=-\infty}^{\infty} C_n e^{jnw_m t} \quad \text{for } -\frac{1}{2f_m} \le t \le \frac{1}{2f_m} \quad ...(4.47)$$

Here, the coefficient C_n is given by

$$C_n = f_m \int_{-\pi/w_m}^{\pi/w_m} e^{j(m_f \sin w_m t)} e^{-jnw_m t}\, dt$$

Substituting $x = w_m t$, we have

$$C_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(m_f \sin x - nx)} dx \quad ...(4.48)$$

In above expression, the integral on the right hand side is the n^{th} **order Bessel function of the first kind and argument m_f**. This function is represented by $J_n(m_f)$.

Thus

$$C_n = J_n(m_f) \quad ...(4.49)$$

Putting the value of C_n in equation (4.47), we get

$$e^{jm_f \sin w_m t} = \sum_{n=-\infty}^{\infty} J_n(m_f) e^{jnw_m t}$$

Substituting above expression in equation (4.46), we have

$$C_{FM}(t) = Ae^{jw_c t}\sum_{n=-\infty}^{\infty} J_n(m_f) e^{jnw_m t}$$

$$C_{FM}(t) = A\sum_{n=-\infty}^{\infty} J_n(m_f) e^{j(w_c + nw_m)t} \quad ...(4.50)$$

In above expression, the real part of the right hand side provides the expression for FM signal i.e.,

$$s(t) = A\sum_{n=-\infty}^{\infty} J_n(m_f)\cos(w_c + nw_m t) \quad ...(4.51)$$

Thus, we have converted original single-tone FM expression into modified form which consists of Bessel function.

Now, the Bessel function $J_n(m_f)$ may be expanded in a power series given by

$$J_n(m_f) = \sum_{m=0}^{\infty} \frac{(-1)^m\left(\frac{1}{2}m_f\right)^{n+2m}}{\angle m \ \angle n+m} \quad ...(4.52)$$

The above expression i.e., the Bessel function $J_n(m_f)$ with respect to m_f has been plotted in figure 4.6 for various positive integer values of n.

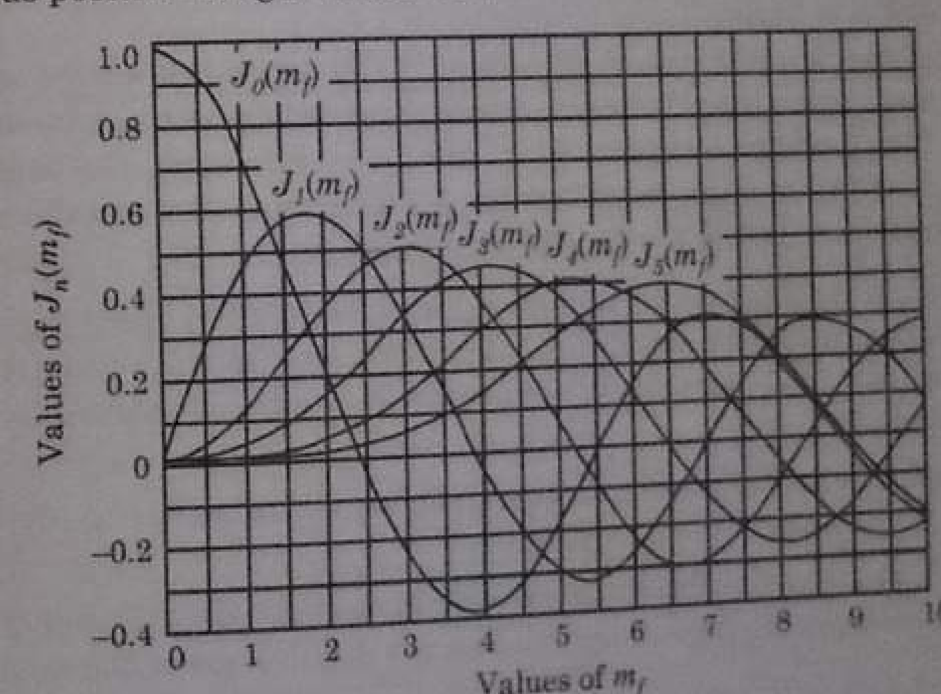


Fig. 4.6. *Bessel functions.*

Few important properties of the Bessel function may be summarized as follows:

(i) $J_n(m_f) = J_{-n}(m_f)$ for even n

$J_n(m_f) = -J_{-n}(m_f)$ for odd n

(ii) For a small value of m_f, we have

$J_0(m_f) \cong 1$

$J_1(m_f) \cong \frac{m_f}{2}$

$J_n(m_f) \cong 0$ for $n > 1$

(iii) $\sum_{n=-\infty}^{\infty} J_n^2(m_f) = 1$

Making use of first property, the equation (4.52) may be expressed as

$$s(t) = A\,J_o(m_f)\cos w_c t + AJ_1(m_f)\,[\cos(w_c + w_m)t - \cos(w_c - w_m)t]$$
$$+ AJ_2(m_f)\,[\cos(w_c + 2w_m)\,t + \cos(w_c - 2w_m)t]$$
$$+ AJ_3(m_f)\,[\cos(w_c + 3w_m)t - \cos(w_c - 3\,w_m)t] + ... \quad ...(4.53)$$

In order to evaluate the value of a given pair of sidebands or the value of the carrier, it is necessary to know the value of the corresponding Bessel function. Table 4.1 provides this type of information.

We may observe the following points from above expression:

(i) The expression contains a carrier term $\cos w_c t$ having magnitude $AJ_o(m_f)$. This means that the magnitude of the carrier term is reduced by a factor $J_o(m_f)$. Thus in FM, unlike in AM, the amplitude of the carrier component does not remain constant.

(ii) From expression in equation (4.53) it may be noted that in FM, theoretically an infinite number of sidebands are produced and the amplitude of each sideband is determined by the corresponding Bessel function $J_n(m_f)$. Thus, the presence of an infinite number of sideband components makes the ideal bandwidth for FM signal infinite. But practically, the distant sidebands with small or insignificant amplitudes are ignored and sidebands with significant amplitudes are considered in calculating the bandwidth for FM signal. Thus the practical bandwidth for FM is finite.

The number of significant sidebands generated in FM depends upon the value of modulation index m_f i.e. modulation index determines how many sideband components have significant amplitudes. The modulation index m_f in turn, depends upon the maximum frequency deviation (Δw) and modulating frequency (w_m). This means that practical bandwidth of FM system depends upon the value of modulation index.

(iv) From Table 4.1, it is clear that for a small value of modulation index m_f (*i.e.*, less than 0.6), only $J_o(m_f)$ and $J_1(m_f)$ have significant values and the higher terms are negligible. Thus, for small values of m_f, FM has a carrier term and only one pair of sidebands. This is equivalent to narrowband FM.

(v) Since the amplitude of FM signal remains unchanged, the power of the FM signal will be same as that of the unmodulated carrier.

(vi) From point (v), we know that the total power of FM remains same as that of the unmodulated carrier which is unlike AM where the total power depends upon the modulation index m_f.

Actually in FM, out of the total power $\frac{A^2}{2}$, the power carried by the carrier term depends upon the value of coefficient $J_o(m_f)$ and the power carried by a sideband depends upon the value of corresponding $J_o(m_f)$. Now if m_f is adjusted to have a value of 2.4 or 5.52 so that $J_o(m_f) = 0$, then the power carried by the carrier term in FM would be zero. Thus, all the power is carried by the sidebands and thus giving a 100 per cent transmission efficiency. However, for other values of m_f, some amount of power is carried by the carrier also and hence the transmission efficiency is less than 100 per cent. This means that by choosing the value of m_f, one can get FM efficiency much more than AM which is 33% and may approach equal to the efficiency of DSB-SC (100%). Hence the efficiency of FM is between AM and DSB-SC.

(viii) There are a number of international regulations for frequency modulation prescribed by CCIR (*i.e.* Consultative Committee for International Radio). These regulations must be followed by all the commercial FM broadcast stations to avoid interference problem. They are as follows:

Table 4.1. Bessel Functions of the First Kind

J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}	J_{15}	J_{16}	J_{16}
0.00	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0.25	0.98	0.12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.94	0.24	0.03	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1.0	0.77	0.44	0.11	0.02	—	—	—	—	—	—	—	—	—	—	—	—	—
1.5	0.51	0.56	0.23	0.06	0.01	—	—	—	—	—	—	—	—	—	—	—	—
2.0	0.22	0.58	0.35	0.13	0.03	—	—	—	—	—	—	—	—	—	—	—	—
2.5	–0.05	0.50	0.45	0.22	0.07	0.02	—	—	—	—	—	—	—	—	—	—	—
3.0	–0.26	0.34	0.49	0.31	0.13	0.04	0.01	—	—	—	—	—	—	—	—	—	—
4.0	–0.40	–0.07	0.36	0.43	0.28	0.13	0.05	0.02	—	—	—	—	—	—	—	—	—
5.0	–0.18	–0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02	—	—	—	—	—	—	—	—
6.0	0.15	–0.28	–0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	—	—	—	—	—	—	—
7.0	0.30	0.00	–0.30	–0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02	—	—	—	—	—	—
8.0	0.17	0.23	–0.11	–0.29	–0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03	—	—	—	—	—
9.0	–0.09	0.24	0.14	–0.18	–0.27	–0.06	0.20	0.33	0.30	0.21	0.12	0.06	0.03	0.01	—	—	—
10.0	–0.25	0.04	0.25	0.06	–0.22	–0.23	–0.01	0.22	0.31	0.29	0.20	0.12	0.06	0.03	0.01	—	—
12.0	0.05	–0.22	–0.08	0.20	0.18	–0.07	–0.24	–0.17	0.05	0.23	0.30	0.27	0.20	0.12	0.07	0.03	0.01
15.0	–0.01	0.21	0.04	–0.19	–0.12	–0.13	0.21	0.03	–0.17	–0.22	–0.09	0.10	0.24	0.28	0.25	0.18	0.12

(a) Maximum modulating frequency is 15 kHz
(b) Maximum frequency deviation is ± 75 kHz
(c) Frequency stability of the carrier is ± 2 kHz
(d) Allowable bandwidth per channel = 200 kHz.

4.9. Effect of Variation of Modulation Index m_f on the Spectrum of FM Signal

As discussed in previous section, the number of sidebands produced in FM increases with increase in the value of modulation index m_f.

The modulation index m_f is given as

$$m_f = \frac{\text{frequency deviation}}{\text{modulating frequency}}$$

or

$$m_f = \frac{\Delta w}{w_m} = \frac{k_f V_m}{w_m}$$

Thus according to above expression m_f may be increased

(i) by reducing w_m and keeping amplitude of modulating signal V_m constant.
(ii) by increasing amplitude of modulating signal V_m and keeping modulating frequency w_m constant.

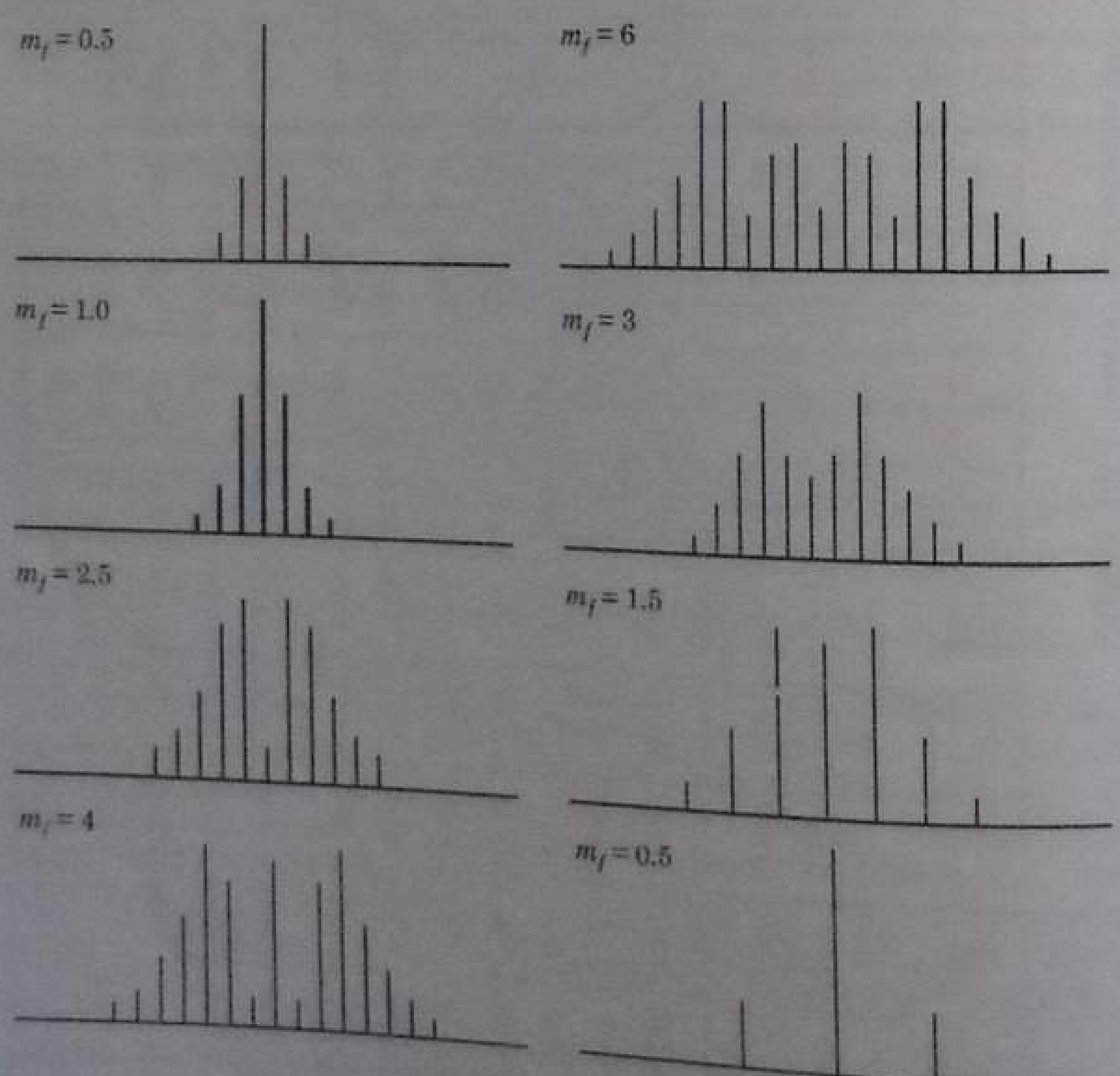


Fig. 4.7.

Infact, it is possible to evaluate the size of the carrier and each sideband for each specific value [of] the modulation index. When this is done, the frequency spectrum of the FM wave for that particular

value of modulation index m_f may be plotted. This has been done in figure 4.7 which shows these frequency spectrums first for increasing deviation and f_m constant and then for decreasing modulating frequency f_m and frequency deviation Δf constant. From figure 4.7, it may be observed that as modulation depth increases, so does the bandwidth and also that reduction in modulating frequency f_m increases the number of sidebands, though not necessarily the bandwidth. Another point which may be observed is that although the number of sideband components is theoretically infinite, but in practice a lot of the higher sidebands have insignificant relative amplitudes and this is why they are not shown in the spectrums.

4.10. Transmission Bandwidth of FM Signal

As discussed in previous article, the bandwidth of FM signal depends upon the value of modulation index m_f. With the increase of modulation index m_f, more and more number of sidebands acquire significant amplitudes and thus bandwidth is increased.

For a particular value of modulation index m_f, we can note the number of significant sideband[s] from Table 4.1.

Basically the effective bandwidth is the separation between the two extreme significant sid[e] frequencies on either side of the carrier.

Fourier analysis reveals that the number of side frequencies which contain a significant amoun[t] of power and thus the effective bandwidth of the FM signal is dependent on the modulation index [of] the modulated wave.

For practical purpose, a side frequency is considered to be significant if its amplitude is at lea[st] one per cent of the unmodulated carrier amplitude.

Let us consider that in the spectrum of FM wave the number of significant sidebands is n. Sin[ce] the upper sidebands are separated by w_m, therefore, they form a frequency band of nw_m.

A similar frequency band of nw_m is formed by the lower sidebands.

Thus, for n sidebands the bandwidth of FM wave is given by

$$BW = 2\,nw_m \text{ radians/sec.} = 2nf_m \text{ Hz.} \qquad ...(4.55)$$

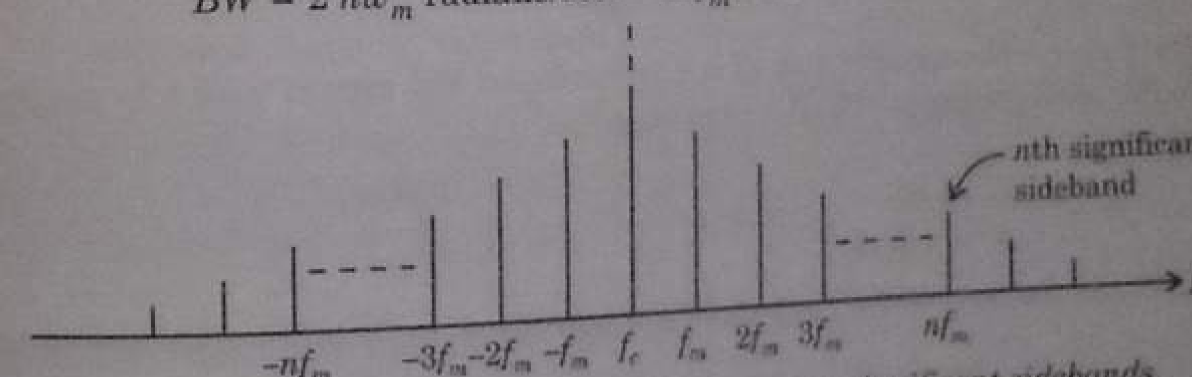


Fig. 4.8. *Frequency spectrum of FM wave for n significant sidebands.*

Universal Curve

Schwartz developed a graph for determining the bandwidth of an FM signal if the modulation index is known. This graph has been shown in figure 4.9. Schwartz uses as his criterion the rule of thumb that any frequency component with a signal strength (voltage) less than 1% of that of the unmodulated carrier will be considered too small to be significant. This graph is also called **universal curve** which shows the variation of the Bandwidth B normalised with respect to Δf against m_f.

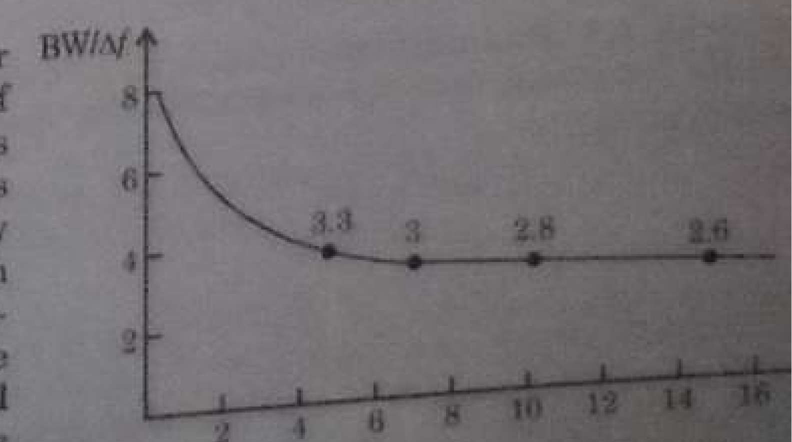


Fig. 4.9. *Schwartz's universal curve.*

4.10.1. Carson's Rule

Carson's rule provides a thumb formula to calculate the bandwidth of a single-tone wideband FM. According to this rule, the FM bandwidth is given as twice the sum of the frequency deviation and the highest modulating frequency. However, it must be remembered that this rule is just an approximation.

Mathematically,

$$BW = 2\,(\Delta w + w_m)$$

But $\quad m_f = \dfrac{\Delta w}{w_m}$

Therefore, $\quad BW = 2\,(m_f\, w_m + w_m) = 2(1 + m_f)w_m \qquad ...(4.56)$

Regarding Carson's rule, we can consider two special cases:

(i) $\because \quad BW = 2(1 + m_f)\, w_m$

If $\Delta w << w_m$ i.e. $m_f << 1$ as is the case for narrowband FM

then $\quad BW = 2\, w_m$* $\qquad ...(4.57)$

(ii) $\because \quad BW = 2(1 + m_f)\, w_m$

If $\Delta w >> w_m$ i.e. $m_f >> 1$ as is the case for wideband FM

then $\quad BW = 2(m_f)w_m$

But $\quad m_f\, w_m = \Delta w$

Thus $\quad BW = 2\Delta w$** $\qquad ...(4.58)$

Example 4.6. Find the bandwidth of a commercial FM transmission if frequency deviation $\Delta f = 75$ kHz and modulating frequency $f_m = 15$ kHz.

Solution: According to Carson's rule, the bandwidth is given by

$$BW = 2(\Delta_f + f_m) = 2(75 + 15)$$

$$BW = 2 \times 90 = 180 \text{ kHz} \qquad \textbf{Ans.}$$

Example 4.7. Determine the bandwidth of a narrowband FM signal which is generated by a 4 kHz audio signal modulating a 125 MHz carrier.

Solution: Given that

$$f_m = 4 \text{ kHz}$$

and $\quad f_c = 125$ MHz

Since this is a narrowband FM signal the bandwidth is found merely by doubling modulating frequency as

$$BW = 2f_m = 2 \times 4 \times 10^3$$

$$BW = 8 \text{ kHz} \qquad \textbf{Ans.}$$

Example 4.8. A 2 kHz audio signal modulates a 50 MHz carrier causing a frequency deviation of 2.5 kHz. Determine the bandwidth of the FM signal.

Solution: Given that

$$\Delta f = 2 \text{ kHz}$$

$$f_c = 50 \text{ MHz}$$

$$\Delta f = 2.5 \text{ kHz}$$

Modulation index m_f is given by

* This case is equivalent to AM.

** For large values of m_f, this bandwidth relation has a very small error and hence can be assumed to be true bandwidth for all practical purposes.

$$m_f = \frac{\Delta_f}{f_m} = \frac{2.5}{2} = 1.25$$

Since the value of m_f is less than $\pi/2$, therefore we are dealing with a narrowband signal. Thus,

$$BW = 2\, f_m = 2 \times 2 \times 10^3 = 4 \text{ kHz} \qquad \textbf{Ans.}$$

Example 4.9. The maximum deviation allowed in an FM broadcast system is 75 kHz. If the modulating signal is a single-tone sinusoid of 8 kHz, determine the bandwidth of FM signal. What will be the bandwidth when modulating signal amplitude is doubled?

Solution: Given that

$$\Delta f = 75 \text{ kHz}$$

$$f_m = 8 \text{ kHz}$$

Bandwidth is given by

$$BW = 2(\Delta f + f_m) = 2\,(75 + 8) = 166 \text{ kHz}$$

Now when the modulating signal amplitude is doubled, the frequency deviation Δf becomes $2 \times 75 = 150$ kHz.

Therefore the bandwidth becomes

$$BW = 2\,(\Delta_f + f_m) = 2(150 + 8) = 316 \text{ kHz} \qquad \textbf{Ans.}$$

4.11. FM Bandwidth for an Arbitrary Modulating Signal x(t)

Till now, we have discussed the bandwidth for a single-tone FM. Now let us consider that the modulating signal $x(t)$ is an arbitrary signal *i.e.* it may consists of large number of frequency components. In this case to calculate the bandwidth, we first find a parameter known as deviation ratio D defined as

$$D = \frac{\text{Peak frequency deviation corresponding to the maximum possible amplitude of } x(t)}{\text{Maximum frequency component present in the modulating signal } x(t)} \qquad ...(4.59)$$

Note: *The deviation ratio D has same meaning to estimate the bandwidth as m_f does for the single-tone FM. These for this case assuming deviation ratio D as equivalent to m_f, we can calculate the transmission bandwidth of FM using Carson's rule.*

4.12. Narrowband FM Versus Wideband FM

An examination of the Schwartz bandwidth curve of figure 4.10 shows that at high values of m_f the curve tends towards a horizontal asymptote and at low values of m_f it tends toward the vertical. Detailed mathematical study shows that the bandwidth of an FM signal for which m_f is less than $\pi/2$ is dependent mainly upon the frequency of the modulating signal and is quite independent of frequency deviation. Further analysis shows that the bandwidth of an FM signal for which m_f is less than $\pi/2$ is equal to twice the modulating frequency.

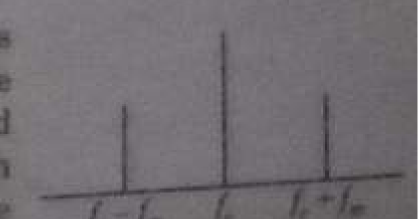


Fig. 4.10

$$\text{Bandwidth} = 2f_m \quad \text{for } m_f < \pi/2$$

Just as with AM, and unlike the situation in which $m_f > \pi/2$, two side frequencies show up for each modulating frequency, one above and one below the frequency of the carrier, each spaced f_m away from the carrier frequency. Because of the limited bandwidth of FM signals with $m_f < \pi/2$, such modulations are known as narrowband FM, and FM signals with $m_f > \pi/2$ are known as *wideband FM*.

Though the spectrum for an AM signal and a narrowband FM signal appear to be the same, a Fourier analysis shows that the magnitude and phase relationships for AM and FM are quite different. See Figure 4.10 for the frequency spectrum of a narrowband FM signal.

Many of the advantages obtained with wideband FM, such as noise reduction, are not available with narrowband FM. Why, then, would one want to use narrowband FM rather than AM? One reason is that with narrowband FM (as well as with wideband FM) the power content at the carrier frequency decreases as the modulation increases so that we have desirable situation of putting the power where the information is.

Example 4.10. A 5 kHz audio tone is used to modulate a 50 MHz carrier causing a frequency deviation of 2 kHz. Determine (*i*) the modulation index and (*ii*) the bandwidth of the FM signal.

Solution : Given that $f_m = 5$ kHz,

$f_c = 50.0$ MHz,

$\Delta f = 20$ kHz

(*i*) Modulation index is defined as

$$m_f = \frac{\Delta f}{f_m} = \frac{20 \times 10^3}{5 \times 10^3} = 4 \quad \textbf{Ans.}$$

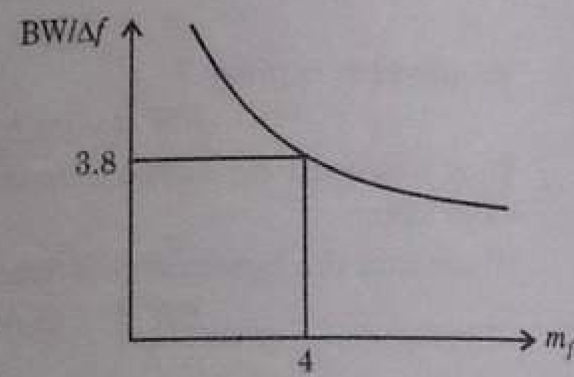


Fig. 4.11.

(*ii*) Referring to the Schwartz bandwidth curve, Figure 4.11, and entering on the horizontal axis with $m_f = 4$, it is found that

$$\frac{BW}{\Delta f} = 3.8$$

This is shown in figure 4.11. Substituting 20×10^3 for Δf as given,

$$\frac{BW}{20 \times 10^3} = 3.8$$

Solving for BW, we have

$$BW = 3.8 \times 20 \times 10^3 = 76 \times 10^3 = 76 \text{ kHz} \quad \textbf{Ans.}$$

Example 4.11. Determine the frequency of the modulating signal which is producing an FM signal having a bandwidth of 50 kHz when the frequency deviation of the FM signal is 10 kHz.

Solution: Given that BW = 50 kHz;

$\Delta f = 1$ kHZ

In order to find f_m, reference must be made to the Schwartz bandwidth curve, *i.e.*, figure 4.12. In order to enter this curve, we have to determine BW/Δf as

$$\frac{BW}{\Delta f} = \frac{50 \times 10^3}{10 \times 10^3} = 5$$

From figure 4.12, we have

$$m_f = 2 = \frac{\Delta f}{f_m}$$

Therefore,
$$2 = \frac{10 \times 10^3}{f_m}$$

$$f_m = \frac{10 \times 10^3}{2} = 5 \text{ kHz} \quad \textbf{Ans.}$$

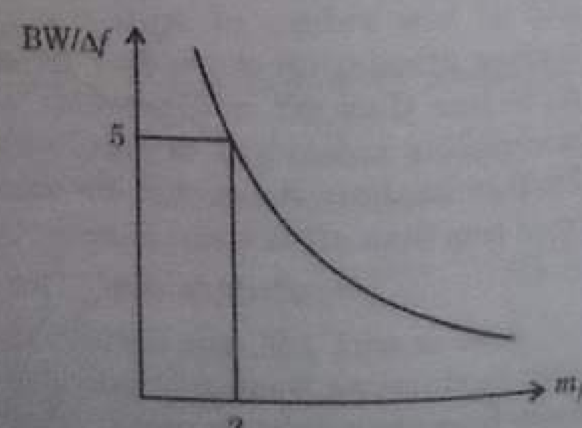


Fig. 4.12.

4.13. Multiple Frequency Modulation

In article 4.5, we discussed a specific case of a single frequency-modulating signal. Now, let us extend these results to the case of multiple frequencies. First only two frequencies will be considered. It can then be generalized to any number of frequencies.

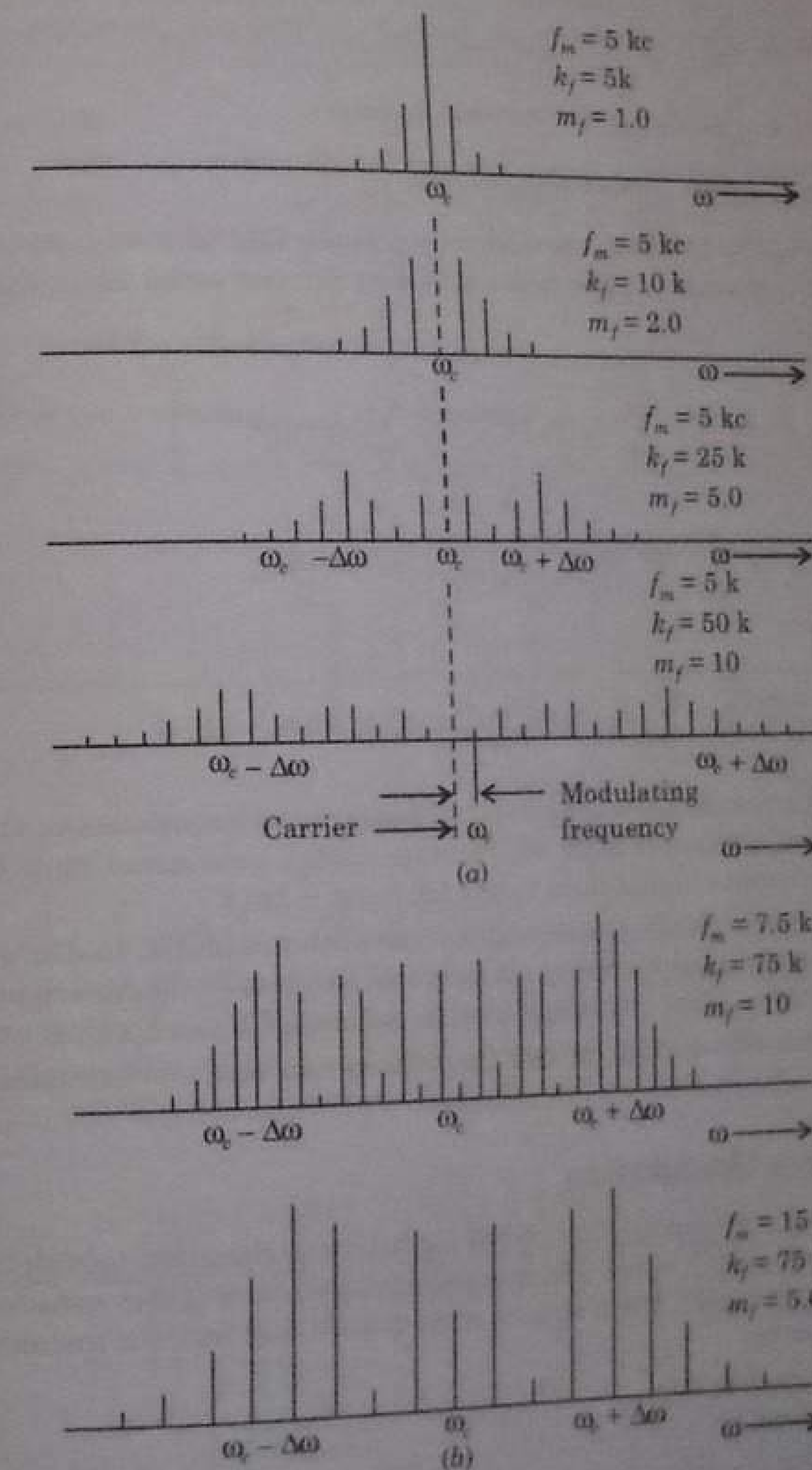


Fig. 4.13.

Let us consider

$$x(t) = a_1 \cos \omega_1 t + a_2 \cos \omega_2 t$$

$$\omega_i = \omega_c + k_f x(t) = \omega_c + k_f (a_1 \cos \omega_1 t + a_2 \cos \omega_2 t)$$

The maximum frequency deviation will be

$$\Delta\omega = (a_1 + a_2) k_f$$

and
$$\phi = \int \omega_i \, dt = \omega_c t + \frac{a_1 k_f}{\omega_1} \sin \omega_1 t + \frac{a_2 k_f}{\omega_2} \sin \omega_2 t$$

or
$$\phi = \omega_c t + m_1 \sin \omega_1 t + m_2 \sin \omega_2 t$$

where
$$m_1 = \frac{a_1 k_f}{\omega_1} \quad \text{and} \quad m_2 = \frac{a_2 k_f}{\omega_2}$$

Also,
$$\hat{s}(t) = Ae^{j\phi(t)} = Ae^{j(\omega_c t + m_1 \sin \omega_1 t + m_2 \sin \omega_2 t)} *$$

or
$$\hat{s}(t) = Ae^{j\omega_c t} \left(e^{jm_1 \sin \omega_1 t}\right)\left(e^{jm_2 \sin \omega_2 t}\right) \quad ...(4.60)$$

The exponentials in the brackets are obviously periodic functions with periods $2\pi/\omega_1$ and $2\pi/\omega_2$ respectively. These exponentials can be represented by Fourier series using Bessel functions.

Hence, we have

$$\hat{s}(t) = Ae^{j\omega_c t}\left[\sum_{n=-\infty}^{\infty} J_n(m_1) e^{jn\omega_1 t}\right]\left[\sum_{k=-\infty}^{\infty} J_k(m_2) e^{jk\omega_2 t}\right]$$

or
$$\hat{s}(t) = A \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} J_n(m_1) J_k(m_2) e^{j(\omega_c + n\omega_1 + k\omega_2)t} \quad ...(4.61)$$

and
$$s(t) = A \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} J_n(m_1) J_k(m_2) \cos[(\omega_c + n\omega_1 + k\omega_2)t] \quad ...(4.62)$$

It is evident from this result that when $x(t)$ is composed of two frequencies ω_1 and ω_2, the spectrum of the FM signal contains sidebands $(\omega_c \pm n\omega_1)$ and $(\omega_c \pm k\omega_2)$, corresponding to frequencies ω_1 and ω_2. In addition, there are cross modulation terms $(\omega_c \pm n\omega_1 \pm k\omega_2)$.

It may be noted that this behaviour contrasts to that observed in AM. In AM, each new frequency added to the modulating signal gives rise to its own side bands only. There are no cross modulation terms. For this reason, AM is called linear modulation, whereas FM is a nonlinear type of modulation.

The results derived in this section for two frequencies can easily be extended to any number of frequencies.

4.14. Square Wave Modulation

Now, let us consider one more special case of FM modulation where the modulating signal $x(t)$ is a square wave as shown in figure 4.14 (a). The method discussed here is very general and is applicable to any periodic modulating signal. For a square wave (figure 4.14 (a)), the instantaneous frequency of the FM carrier is given by

$$\omega_i = \omega_c + k_f x(t)$$

and the phase ϕ is given by

$$\phi = \int \omega_i \, dt = \omega_c t + k_f \int x(t) \, dt$$

* Here, we are using the notation $\hat{s}(t)$ for the exponential representation of $s(t)$.
For example, if
$$s(t) = A \cos \theta(t)$$
then, $\hat{s}(t) = Ae^{j\theta(t)}$
and $s(t) = Re\ \hat{s}(t)$

or
$$\phi = \omega_c t + \psi(t)$$

where
$$\psi(t) = k_f \int x(t) \, dt$$

is shown in figure 4.14. This is a triangular periodic function with period T.

It may be noted that the maximum carrier frequency deviation $\Delta\omega$ in this case is k_f since $|x(t)|_{max} = 1$. Hence,

$$k_f = \Delta\omega$$

and
$$\psi(t) = \begin{cases} (\Delta\omega)t & \text{for } -\frac{T}{4} < t < \frac{T}{4} \\ (\Delta\omega)\left[\frac{T}{2} - t\right] & \text{for } \frac{T}{4} < t < \frac{3T}{4} \end{cases} \quad ...(4.63)$$

and
$$\psi(t) = \psi(t \pm nT)$$

Also
$$\hat{s}(t) = Ae^{j\theta(t)} = Ae^{j\psi(t)} e^{j\omega_c t}$$

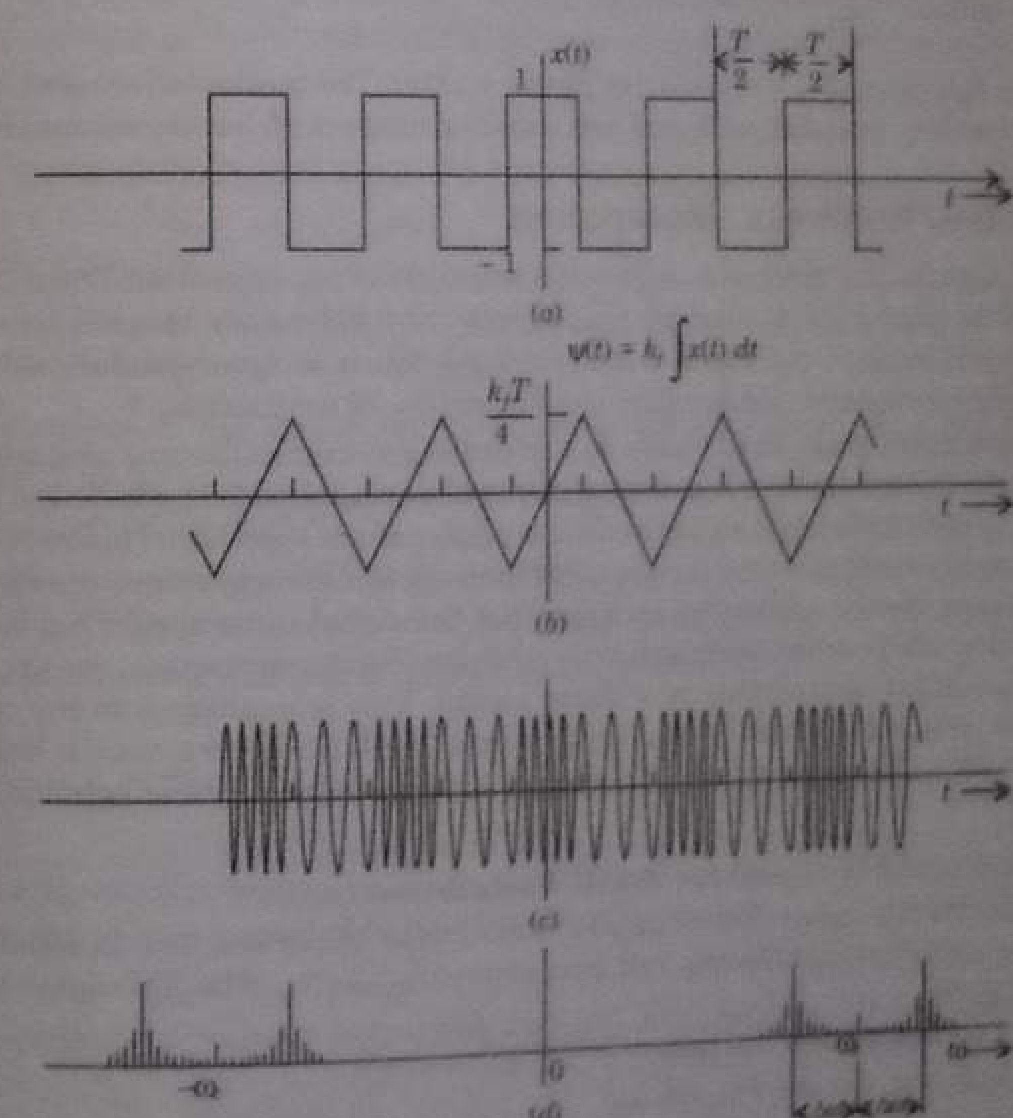


Fig. 4.14

The function $e^{j\psi(t)}$ is itself a periodic function of period T, and it can be expressed by Fourier series as

$$e^{j\psi(t)} = \sum_{n=-\infty}^{\infty} \alpha_n e^{jn\omega_0 t}$$

$$\omega_s = \frac{2\pi}{T}$$

where $$\alpha_n = \frac{1}{T}\int_{-T/4}^{3T/4} e^{j\psi(t)}e^{-jn\omega_s t}dt$$

Substitution of equation (4.63) in above equation and the subsequent integration will yield

$$\alpha_n = \frac{1}{2}\left\{Sa\left[\frac{\pi}{2}(m_f - n)\right] + (-1)^n Sa\left[\frac{\pi}{2}(m_f + n)\right]\right\} \quad ...(4.64)$$

where $$\beta = \frac{\Delta\omega}{\omega_s}$$

Hence, $$\hat{s}(t) = Ae^{j\psi(t)}e^{j\omega_c t} = A\sum_{n=-\infty}^{\infty} \alpha_n e^{j(\omega_c + n\omega_s)t}$$

and $$s(t) = A\sum_{n=-\infty}^{\infty} \alpha_n \cos(\omega_c + n\omega_s)t$$

The frequency spectrum of $s(t)$ is shown in figure 4.14(d). The method discussed here is general and can be applied to any modulating signal $x(t)$ that is periodic and has a zero mean.

4.15. Linear and Nonlinear Modulation

In the case of AM signals, the sidebands follow the principle of superposition. Thus, if signals $x_1(t)$ and $x_2(t)$ give rise to sidebands φ_1 and φ_2, respectively, the sidebands created by the composite signal $x_1(t) + x_2(t)$ will be $\varphi_1 + \varphi_2$. There is no intermodulation or cross-product sidebands as we observed in FM. For this reason, AM is called the linear type of modulation.*

The linear modulation lends itself easily to mathematical manipulations and generalizations. The spectrum of a modulated signal due to sum of two modulating signals can be found by calculating the spectrum due to each individual signal and then adding them together. This proves very useful in noise calculations in communication systems. For linear modulation systems, the effect of additive noise over the channel can be calculated by assuming the signal to be zero. This is not true of a nonlinear modulation where cross-modulation terms arise. For these reasons, we are interested in approximating a nonlinear modulation by a linear model. This is analogous to the case of system analysis where one can approximate a nonlinear system by a linear one over a limited range of signal amplitudes. Now, let us show that FM closely approximates a linear behaviour for a small modulation index.

4.15.1. Linearization of FM Signal for Small Modulation Index

For a small modulating index, FM closely exhibits linear behaviour. Let us consider again the case of modulating signal $x(t)$ containing two frequencies ω_1 and ω_2. The FM signal in this case is given by equation (4.60).

If m_1 and $m_2 < 1$, then for $x(t)$, we have

$$x(t) = a_1 \cos \omega_1 t + a_2 \cos \omega_2 t$$

$$\hat{s}(t) \cong A(1 + jm_1 \sin \omega_1 t)(1 + jm_2 \sin \omega_2 t)e^{j\omega_c t} \cong A(1 + jm_1 \sin \omega_1 t + jm_2 \sin \omega_2 t)e^{j\omega_c t}$$

* The modulated signal is a function of the modulating signal $x(t)$. Let $\varphi[x(t)]$ be the modulated signal. Then, the modulation is linear if $(d/d[x(t)])\{\phi[x(t)]\}$ is independent of $x(t)$. Otherwise it is a nonlinear modulation. The reader can easily verify, according to this definition that AM is linear, whereas FM is not.

Note that if

$$x_1(t) = a_1 \cos \omega_1 t \text{ and } m_1 << 1$$

then $$\hat{s}(t) = Ae^{jm_1 \sin \omega_1 t}e^{j\omega_c t}$$

or $$\hat{s}(t) \cong A[1 + jm_1 \sin \omega_1 t]e^{j\omega_c t}$$

If $$x_2(t) = a_2 \cos \omega_2 t \quad \text{and } m_2 << 1$$

then $$\hat{s}(t) \cong A(1 + jm_2 \sin \omega_2 t)e^{j\omega_c t}$$

It is easy to see that under the conditions $m_1, m_2 << 1$, the sidebands due to modulating signa $x_1(t) + x_2(t)$ are the sum of sidebands due to individually. Hence, FM can be assumed to be a linea modulation for a small modulation index. The cross-modulation terms under this assumption can b ignored.

4.16. Phase Modulation: An Analytic View

As discussed earlier, phase modulation may be analysed in a similar way as the FM signal. The PI has some distinct features as compared to FM. The main distinct feature of phase-modulation that the deviation in the carrier frequency w_c in PM is linearly proportional to the baseband modulating frequency w_m. However in FM, the deviation is independent of the baseband or modulatir frequency.

Let us find an expression for frequency deviation for a PM wave.

We know that the total phase angle of a PM wave is expressed as

$$\phi_i = w_c t + k_p . x(t) \quad ...(4.6$$

For a single-tone modulating signal, we have

$$x(t) = V_m \cos w_m t \quad ...(4.6$$

Thus $$\phi_i = w_c t + k_p . V_m \cos w_m t \quad ...(4.6$$

In above expression, the maximum departure in the phase is $k_p . V_m$. This is called as pha deviation denoted by θ_d.

Hence phase deviation, $\theta_d = k_p . V_m$ ...(4.

Now, the expression for the PM wave becomes

$$s(t)_{PM} = A \cos \phi_i = A \cos[w_c t + \theta_d \cos w_m t] \quad ...(4.$$

Also, the instantaneous frequency related to ϕ_i is expressed as

$$w_i = \frac{d\phi_i}{dt} = \frac{d}{dt}[w_c t + k_p V_m \cos w_m t] \quad ...(4.$$

or $$w_i = w_c - k_p V_m w_m \sin w_m t$$

Thus, the maximum departure in the frequency from w_c is $k_p . V_m . w_m$.

Hence, for PM, the frequency deviation is given by

$$\Delta w_{PM} = k_p . V_m . w_m \quad ...(4.$$

which depends upon the modulating frequency w_m. On the other hand, the frequency deviat for FM is

$$\Delta w_{FM} = k_f . V_m \quad ...(4.$$

Thus we conclude that for an equal bandwidth in FM and PM, we have

$$k_f = k_p . w_m \quad ...(4.$$

4.16.1. Transmission Bandwidth of PM

We may estimate the transmission bandwidth for PM using Carson's rule as

$$BW_{PM} \cong 2(\Delta w) \cong 2k_p . V_m . w_m \quad ...(4.$$

Hence, the bandwidth of the PM wave varies fastly with the variation in the modulating frequency w_m. On the other hand, the FM bandwidth varies slowly with modulating frequency w_m.

For PM, the modulation index m_p will be same as the deviation θ_d and will be expressed as

$$m_p = k_p \cdot V_m = \theta_d \qquad ...(4.76)$$

Example 4.12. A baseband or modulating signal $x(t) = 5 \cos 2\pi\, 15 \times 10^3\, t$ angle modulates a carrier signal $A \cos w_c t$.

(i) Determine the modulation index and bandwidth for

(a) FM system (b) PM system

(ii) Find the change in the bandwidth and modulation index for both FM and PM if modulating frequency f_m is reduced to 5 kHz.

Assume $k_f = k_p = 15$ **kHz/volt**

Solution: (i) Given that

$$V_m = 5\ V$$
$$f_m = 15 \text{ kHz}$$

(a) For FM system

Frequency deviation $\Delta f = k_f \cdot V_m$

$$m_f = 15 \times 10^3 \times 5 = 75 \text{ kHz}$$

Hence, modulation index

$$m_f = \frac{\Delta f}{m_f} = \frac{75}{15} = 5$$

Thus, according to Carson's rule the bandwidth is

$$BW = 2(\Delta f + f_m) = 2(75 + 15) = 2 \times 90$$

or $$BW = 180 \text{ kHz}$$

(b) For PM system

Frequency deviation $\Delta f = k_p \cdot V_m \cdot f_m = 15 \times 10^3 \times 5 \times 15 \times 10^3$

$$\Delta f = 1125 \text{ MHz}$$

Then the bandwidth will be

$$BW = 2(\Delta f + f_m) \cong 2\Delta f = 2 \times (1125 \text{ MHz})$$
$$BW = 2250 \text{ MHz}$$

It is clear that the bandwidth is quite large as compared to FM. This is due to the fact that the modulation index in PM is quite large as

$$m_p = k_p \cdot V_m = 15 \times 10^3 \times 5 = 75{,}000$$

This very large value of modulation index m_p produces a large number of sidebands and thus this is the reason for very large bandwidth

(ii) In this case $f_m = 5$ kHz

(a) For FM

Since the frequency deviation Δf is independent of f_m and hence remains 75 kHz.

Thus $$m_f = \frac{\Delta f}{f_m} = \frac{75}{5} = 15$$

In this case the bandwidth will be

$$BW = 2(\Delta f + f_m) = 2(75 + 5) = 2 \times 80 = 160 \text{ kHz}$$

This means that in FM, the modulation index changes sufficiently with a change in f_m, however, the bandwidth changes only slightly.

(b) For PM

In this case the frequency-deviation Δf is dependent on f_m and is given as

$$\Delta f = k_p \cdot V_m \cdot f_m = 15 \times 10^3 \times 5 \times 5 \times 10^3$$
$$\Delta f = 375 \text{ MHz}$$

Thus $$BW = 2(\Delta f + f_m) \cong 2\Delta f = 2 \times 375 = 750 \text{ MHz}$$

Hence, the bandwidth has changed considerably as compared to previous one i.e. 2250 MHz.

The modulation index is independent of f_m and is given by

$$m_p = k_p \cdot V_m = 15 \times 10^3 \times 5$$
$$m_p = 75 \text{ kHz}$$

which is same as the previous one *i.e.* 75 kHz.

4.17. Comparison of Angle Modulated Waves and Amplitude Modulated Wave

Figure 4.15 shows a single-tone modulating signal, a carrier signal, amplitude-modulated (AM) wave and angle-modulated (*i.e.* FM and PM) waves. On comparison, we note the following differences between two techniques:

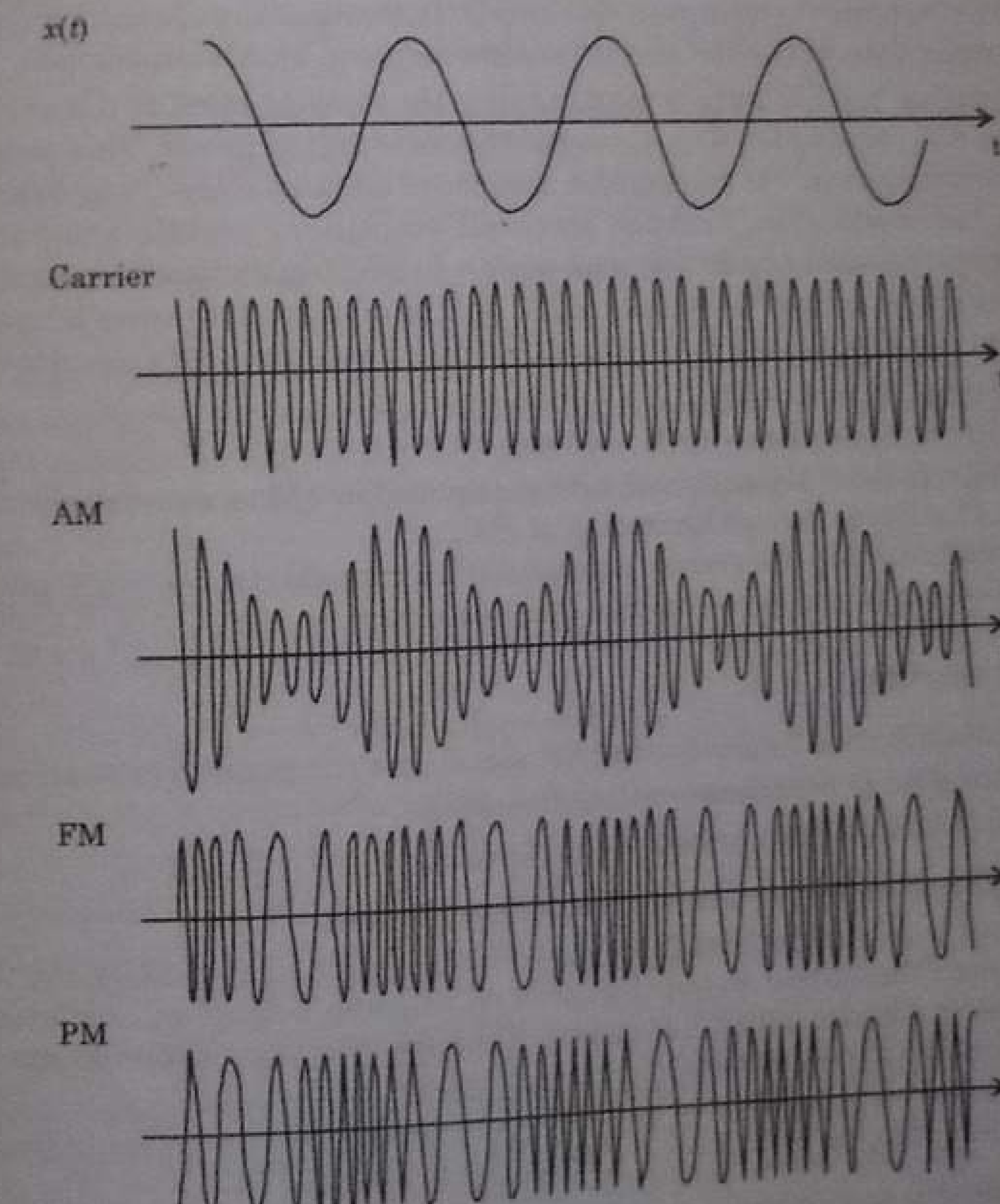


Fig. 4.15. *AM FM, and PM waveforms.*

(i) The envelope of FM wave or PM wave is constant and is equal to the unmodulated carrier amplitude. On the other hand, the envelope of AM wave is dependent on the modulating signal $x(t)$.

Hence, the bandwidth of the PM wave varies fastly with the variation in the modulating frequency w_m. On the other hand, the FM bandwidth varies slowly with modulating frequency w_m.

For PM, the modulation index m_p will be same as the deviation θ_d and will be expressed as

$$m_p = k_p \cdot V_m = \theta_d \quad ...(4.76)$$

Example 4.12. **A baseband or modulating signal $x(t) = 5 \cos 2\pi\, 15 \times 10^3\, t$ angle modulates a carrier signal $A \cos w_c t$.**

(i) Determine the modulation index and bandwidth for

(a) FM system (b) PM system

(ii) Find the change in the bandwidth and modulation index for both FM and PM if modulating frequency f_m is reduced to 5 kHz.

Assume $k_f = k_p = 15$ **kHz/volt**

Solution: (i) Given that

$$V_m = 5\ V$$
$$f_m = 15 \text{ kHz}$$

(a) For FM system

Frequency deviation $\Delta f = k_f \cdot V_m$

$$m_f = 15 \times 10^3 \times 5 = 75 \text{ kHz}$$

Hence, modulation index

$$m_f = \frac{\Delta f}{m_f} = \frac{75}{15} = 5$$

Thus, according to Carson's rule the bandwidth is

$$BW = 2(\Delta f + f_m) = 2(75 + 15) = 2 \times 90$$

or $BW = 180$ kHz

(b) For PM system

Frequency deviation $\Delta f = k_p \cdot V_m \cdot f_m = 15 \times 10^3 \times 5 \times 15 \times 10^3$

$$\Delta f = 1125 \text{ MHz}$$

Then the bandwidth will be

$$BW = 2(\Delta f + f_m) \cong 2\Delta f = 2 \times (1125 \text{ MHz})$$
$$BW = 2250 \text{ MHz}$$

It is clear that the bandwidth is quite large as compared to FM. This is due to the fact that the modulation index in PM is quite large as

$$m_p = k_p \cdot V_m = 15 \times 10^3 \times 5 = 75{,}000$$

This very large value of modulation index m_p produces a large number of sidebands and thus this is the reason for very large bandwidth

(ii) In this case $f_m = 5$ kHz

(a) For FM

Since the frequency deviation Δf is independent of f_m and hence remains 75 kHz.

Thus $$m_f = \frac{\Delta f}{f_m} = \frac{75}{5} = 15$$

In this case the bandwidth will be

$$BW = 2(\Delta f + f_m) = 2(75 + 5) = 2 \times 80 = 160 \text{ kHz}$$

This means that in FM, the modulation index changes sufficiently with a change in f_m, however the bandwidth changes only slightly.

(b) For PM

In this case the frequency-deviation Δf is dependent on f_m and is given as

$$\Delta f = k_p \cdot V_m \cdot f_m = 15 \times 10^3 \times 5 \times 5 \times 10^3$$
$$\Delta f = 375 \text{ MHz}$$

Thus $$BW = 2(\Delta f + f_m) \cong 2\Delta f = 2 \times 375 = 750 \text{ MHz}$$

Hence, the bandwidth has changed considerably as compared to previous one *i.e.* 2250 MHz.

The modulation index is independent of f_m and is given by

$$m_p = k_p \cdot V_m = 15 \times 10^3 \times 5$$
$$m_p = 75 \text{ kHz}$$

which is same as the previous one *i.e.* 75 kHz.

4.17. Comparison of Angle Modulated Waves and Amplitude Modulated Wave

Figure 4.15 shows a single-tone modulating signal, a carrier signal, amplitude-modulated (AM) wave and angle-modulated (*i.e.* FM and PM) waves. On comparison, we note the following differences between two techniques:

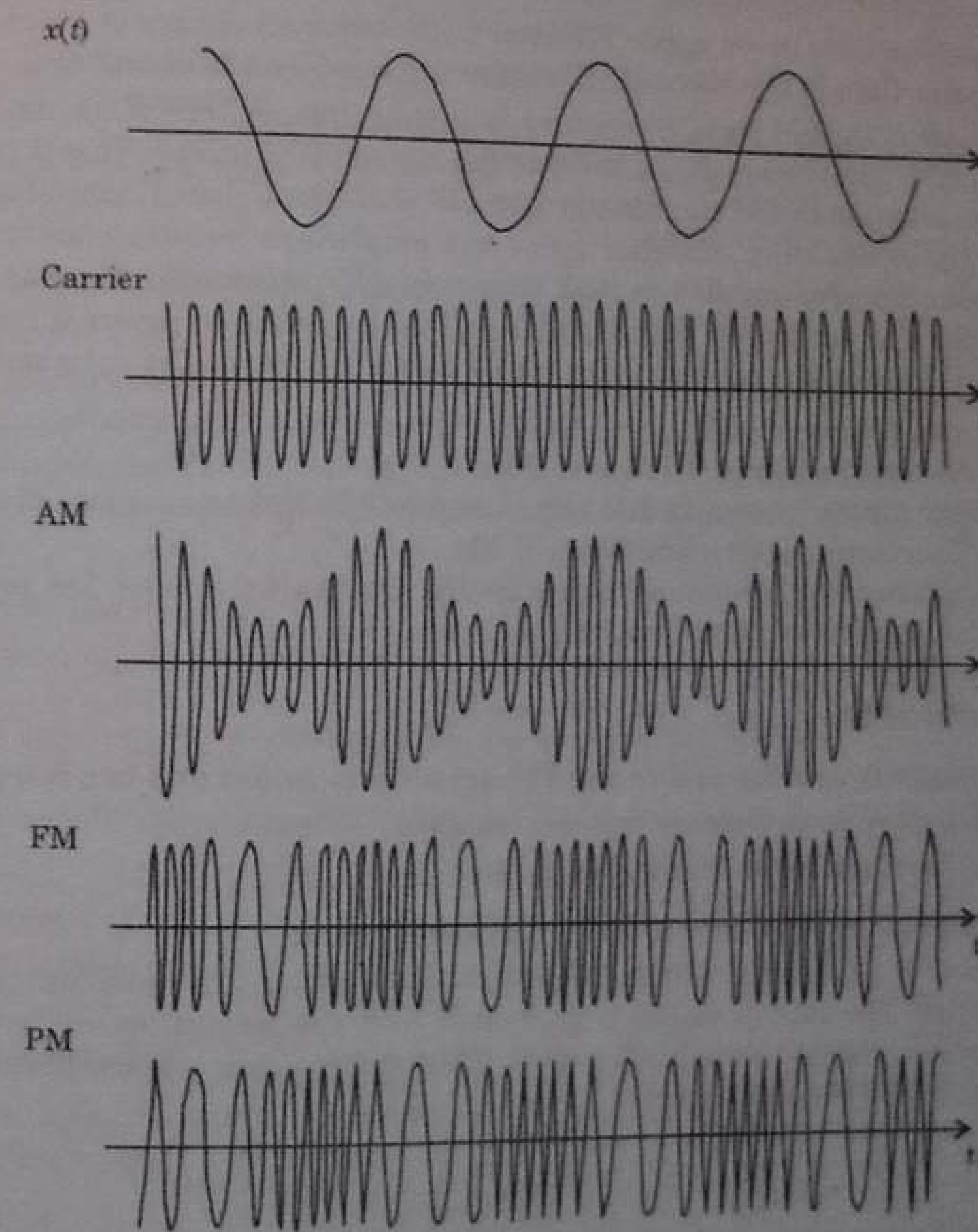


Fig. 4.15. *AM FM, and PM waveforms.*

(i) The envelope of FM wave or PM wave is constant and is equal to the unmodulated carrier amplitude. On the other hand, the envelope of AM wave is dependent on the modulating signal $x(t)$.

(ii) The zero crossings (i.e. the instants of time at which the waveform changes from negative to a positive value or vice-versa) of a FM wave or a PM wave no longer exhibit a perfect regularity in their spacing like AM wave. Thus this makes the instantaneous frequency of the angle modulated wave depend upon time.

4.18. Comparison of Frequency Modulation and Amplitude Modulation

(A) Advantages of FM over AM

The frequency modulation (FM) has the following advantages over AM:

(i) FM receivers may be fitted with amplitude limiters to remove the amplitude variations caused by noise.

This makes FM reception a good deal more immune to noise than AM reception.

(ii) It is possible to reduce noise still further by increasing the frequency-deviation. This is a feature which AM does not have because it is not possible to exceed 100 per cent modulation without causing severe distortion.

(iii) Standard Frequency Allocations provide a guard band between commercial FM stations. Due to this, there is less adjacent-channel interference than in AM.

(iv) FM broadcasts operate in the upper VHF and UHF frequency ranges at which there happens to be less noise than in the MF and HF ranges occupied by AM broadcasts.

(v) The amplitude of the FM wave is constant. It is thus independent of the modulation depth, whereas in AM, modulation depth governs the transmitted power. This permits the use of low-level modulation in FM transmitter and use of efficient class C amplifiers in all stages following the modulator. Further since all amplifiers handle constant power, the average power handled equals the peak power. In AM transmitter the maximum power is four times the average power. Finally in FM, all the transmitted power is useful whereas in AM, most of the power is carrier power which does not contain any information.

(B) Disadvantages of FM over AM

Following are the disadvantages of FM over AM:

(i) A much wider channel typically 200 kHz is required in FM as against only 10 kHz in AM broadcast. This forms serious limitation of FM.

(ii) FM transmitting and receiving equipments particularly used for modulation and demodulation tend to be more complex and hence more costly.

4.19. FM Generation

The FM modulator circuits used for generating FM signals may be put into two categories as under:

(i) The direct method or parameter variation method.

(ii) The indirect method or the Armstrong method.

4.19.1. The Direct Method or Parameter Variation Method

In direct method or parameter variation method, the baseband or modulating signal directly modulates the carrier. The carrier signal is generated with the help of an oscillator circuit. This oscillator circuit uses a parallel tuned L-C circuit. Thus the frequency of oscillation of the carrier generation is governed by the expression

$$\omega_c = \frac{1}{\sqrt{LC}}$$

Now, we can make the carrier frequency w_c to vary in accordance with the baseband or modulating signal $x(t)$ if L or C is varied according to $x(t)$. An oscillator circuit whose frequency is controlled by a modulating voltage is called voltage controlled oscillator (VCO). The frequency of VCO is varied according to the modulating signal simply by putting a shunt voltage variable capacitor with its tuned circuit. This voltage variable capacitor is called varactor or varicap. This type of property is

exhibited by reverse biased semiconductor diodes. Also the capacitance of bipolar transistors (BJT) and field-effect transistors (FET is varied by the Miller-effect. This miller capacitance may be utilized for frequency modulation.

In addition to this, the electron tubes may also provide variable reactance (either it is inductive or capacitive) which is proportional to modulating or baseband signal. This type of tubes are called reactance tubes and may be used for FM generation.

The inductance L of the tuned circuit may also be varied in accordance with the baseband or modulating signal $x(t)$. The FM circuit using such inductors is called saturable reactor modulator. Frequency modulation can also be achieved from voltage controlled devices such as PIN diode, Klystron oscillators and multivibrators.

4.19.2. Varactor Diode Method for FM Generation

The varactor diode is a semiconductor diode whose junction capacitance changes with d.c. bias voltage. This varactor diode is connected in shunt with the tuned circuit of the carrier oscillator. This arrangement is shown in figure 4.16.

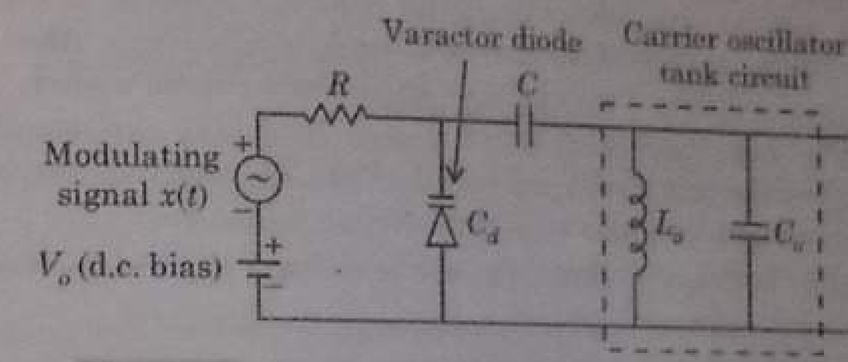


Fig. 4.16. *Varactor diode method of FM generation.*

In varactor diode FM generation arrangement, the capacitor C is made much smaller than the varactor diode capacitance C_d so that the radio frequency (RF) voltage from oscillator across the diode is small as compared to reverse bias d.c. voltage across the varactor diode. In addition to this, the reactance of the capacitor C at the highest modulating frequency is made large enough compared to resistor R so that the shunting of the baseband or modulating signal through the tuned circuit may be checked.

Mathematical Analysis

The capacitance C_d of the varactor diode is expressed as

$$C_d = \frac{k}{\sqrt{v_D}} = k(v_D)^{-1/2} \quad ...(4.77)$$

Here, v_D is the total instantaneous voltage across the varactor diode and is given by

$$v_D = V_o + x(t) \quad ...(4.78)$$

Also, k is a constant of proportionality.

The oscillation frequency is given as

$$\omega_c = \frac{1}{\sqrt{LC}} \quad ...(4.79)$$

Now, the total capacitance of the oscillator tank circuit will be $C_o + C_d$ and thus the instantaneous frequency of oscillation w_i is expressed as

$$\omega_i = \frac{1}{\sqrt{L_o(C_o + C_d)}} \quad ...(4.80)$$

In above equation, substituting the value of C_d from equation, we have

$$\omega_i = \frac{1}{\sqrt{L_o(C_o + kv_D^{-1/2})}}$$

We conclude that the instantaneous frequency w_i of FM signal depends upon v_D which in turn depends upon the value of the modulating signal $x(t)$. Thus, the instantaneous oscillator frequency w_i also depends upon the baseband or modulating signal $x(t)$ and hence frequency modulation is generated.

4.20. Drawbacks of Direct Method for FM Generation

Following are the drawbacks of the direct method:

(i) In direct method of FM generation, it is not easy to get a high order stability in carrier frequency. This is due to the fact that generation of carrier signal is directly affected by the baseband or modulating signal. The baseband signal directly controls the tank circuit of the carrier generator and thus a stable oscillator circuit (*i.e.* crystal oscillator) cannot be used. This means that the carrier generation cannot be of high stability which is a necessary requirement.

A solution to this problem is the indirect method *i.e.* Armstrong method of FM generation. In fact, in this method, the carrier oscillator does not respond to the modulating signal directly rather the generation of the carrier is made isolated from other parts of the circuit. Thus stable crystal oscillators may be used to generate carrier signal.

(ii) The non-linearity of the varactor diode produces a frequency variation due to harmonics of the modulating or baseband signal and therefore the FM signal is distorted. We will have to take the proper cares to keep this type of distortion minimum.

Despite all above drawbacks, the direct method is utilized for high power FM generation in several applications.

4.21. The Indirect or Armstrong Method of FM Generation

(U.P. Tech. Semester, Exam. 2003-04)

In Armstrong method of FM generation, we can get very high frequency stability since in this case the crystal oscillator may be used as a carrier frequency generator. The working principle of Armstrong method is to generate a narrowband FM (NBFM) indirectly by utilizing the phase-modulation technique and then changing this narrowband FM into a wideband FM as shown in figure 4.17. Since in narrowband FM the modulation index is small, therefore the distortion is very low in narrowband FM. Here we prefer phase modulation technique because its generation is easy. The multiplier circuit apart from multiplying the carrier frequency also increases the frequency deviation and hence the narrowband FM (with small frequency deviation) is converted into wideband FM (with large frequency deviation).

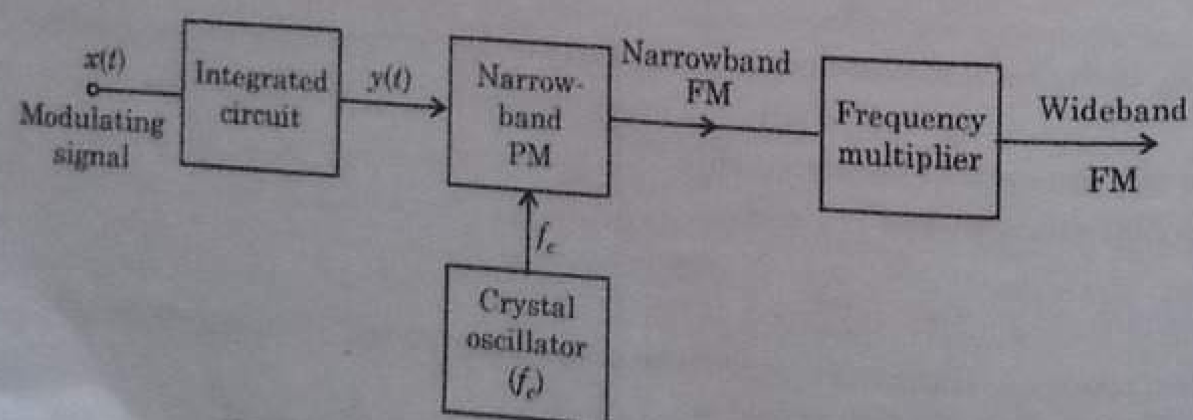


Fig. 4.17. *Armstrong method for FM generation.*

4.22. Practical Armstrong Method for FM Generation

[In] this section, we shall discuss a practical method of generating FM signal. Figure 4.18 shows a [si]mplified block diagram of a commercial FM generation system using Armstrong Method.

We know that in commercial use we require to transmit audio signal consisting frequencies in the range of 50 Hz to 15 kHz and the value of Δf = 75 kHz. Let us assume that the final carrier frequency of the FM required is f_c = 100 MHz.

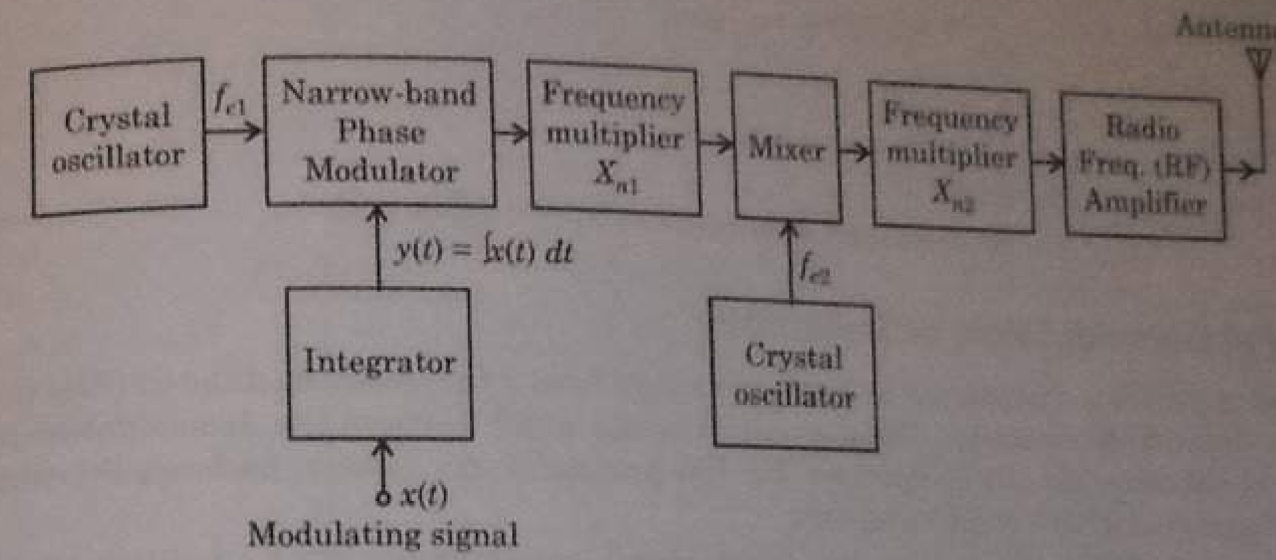


Fig. 4.18.

Let us begin with narrowband FM having a carrier frequency f_{c1} = 100 kHz generated by a crystal oscillator. To check the harmonic distortion produced by narrowband phase modulator, the modulation index m_f to a maximum value of 0.3 radians.

Let us assume m_{f1} = 0.2 radian. The lowest modulation frequency of 100 Hz produces a frequency deviation of $\Delta_{f1} = 0.2 \times 50 = 10$Hz at the narrowband phase modulator output whereas the largest modulating frequency 15 kHz produces a frequency deviation of

$$\Delta_{f2} = 0.2 \times 15 \text{ kHz} = 3 \text{ kHz} \quad ...(4.81)$$

Thus, the lowest modulating frequency is of immediate concern. We take the value of Δ_{f1} = 10 Hz, so that at the highest modulating frequency, m_f becomes even less.

To produce a frequency deviation of Δf = 75 kHz at the output, a frequency multiplication is required. As an example Δf_1 = 10 Hz and the required deviation is Δf = 75 kHz. We thus require a total frequency multiplication by a factor

$$n = \frac{75000}{10} = 7500 \quad ...(4.82)$$

However, a straight forward frequency multiplication equal to this value leads to a very high value of carrier frequency than the required 100 MHz. To get the required frequency-deviation and carrier frequency we use a two-stage frequency multiplier as shown in figure 4.18. This arrangement makes use of two multipliers and a mixer. With the help of a mixer, the carrier frequency is translated suitably without altering frequency-deviation Δf. The final stage multiplier provides the required carrier frequency and deviation.

Let us assume that n_1 and n_2 are the frequency multiplication factors for the two multipliers, so that

$$n = n_1 \cdot n_2 \quad ...(4.83)$$

$$= \frac{\Delta f}{\Delta f_1} = \frac{75000}{10} = 7500 \quad ...(4.84)$$

The carrier frequency at the output of first multiplier is translated downwards to frequency $(f_2 - nf_1)$ by mixing it with a carrier wave of frequency f_2 which is produced by another oscillator.

The carrier frequency at the input of the second multiplier is $\frac{f_c}{n_2}$.

Hence, $f_2 - n_1 f_1 = \frac{f_C}{n_2}$

Thus with $f_1 = 0.1$ MHz

and $f_2 = 8.5$ MHz, we have

$$8.5 - 0.1\, n_1 = \frac{100}{n_2}$$

Using equation (4.78), we have

$$n_1 = 100 \text{ and } n_2 = 75$$

4.23. FM Demodulators or Detectors

The process of getting a modulating or baseband signal from a frequency modulated (FM) signal is called demodulation or detection. The electronic circuits which perform the demodulation process are known as FM demodulators or detectors. The FM demodulator or detector performs the extraction of modulating signal in two steps as follows:

(i) It converts the frequency-modulated (FM) signal into a corresponding amplitude modulated (AM) signal with the help of frequency dependent circuits, *i.e.*, the circuits whose output voltage depends upon the input frequency. These circuits are generally known as **frequency discriminators.**

(ii) The original modulating or baseband signal is recovered from this AM signal with the help of the linear diode detector or envelope detector.

A simpel R-L circuit may be used as a discriminator, but this circuit has a very poor sensitivity as compared to a tuned L-C circuit. Therefore, L-C circuits are generally used as frequency-discriminators.

4.24. Types of FM Demodulators

FM demodulators are of following types:

1. Slope detector

The principle of operation of slope detectors depends upon the slope of the frequency response characteristics of a frequency selective network. The two main FM detectors, which use detuned resonant circuits come under this category, are as under:

(i) Single-tuned detector circuit or simple slope detector.

(ii) Stagger-tuned detector circuit or balanced slope detector.

2. Phase Difference Detectors

The following two circuits come under this category:

(i) Foster-Seeley detector

(ii) Ratio detector

.24.1. Slope Detector

Figure 4.19 shows the circuit diagram of a slope detector.

The circuit consists of a tuned circuit which is slightly detuned from the carrier frequency w_c. In her words, the circuits uses two tuned circuits which are tuned to two different freqencies. First e is tuned to the incoming FM carrier frequency w_c whereas the second one is tuned to a frequency htly different from the carrier frequency w_c. Therefore, this portion of the circuit which contains tuned circuits tuned to different frequencies, is called discriminator.

This circuit converts the FM singal into an AM signal as shown in the slope detector characteristic e. The another portion of the circuit is envelope detector. The AM signal from the output of the iminator is applied at the input of envelope detector.

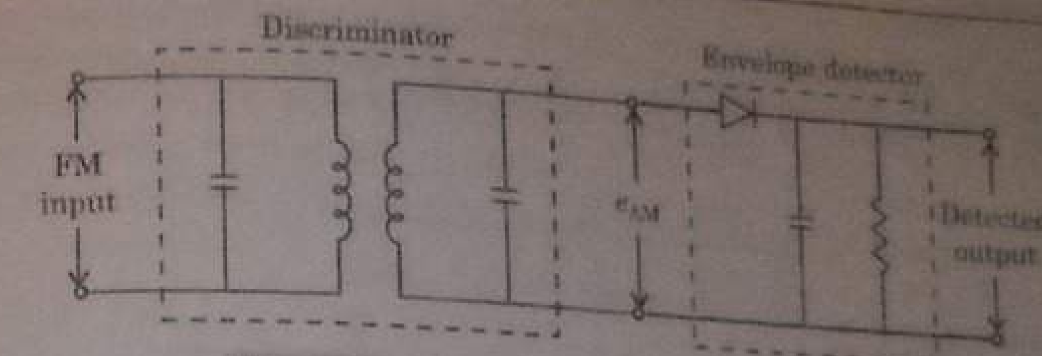


Fig. 4.19. *Circuit diagram of a slope detector.*

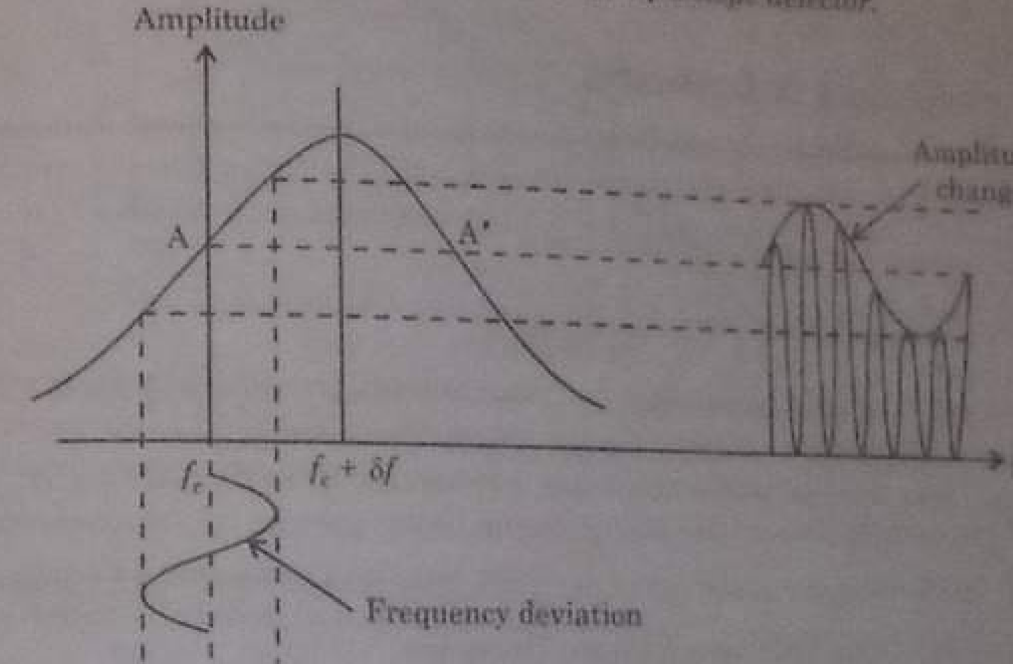


Fig. 4.20. *Slope detector characteristic curve.*

At the output of envelope detector, the original modulating or baseband signal is obtained. Although this circuit is simple and inexpensive, it has following drawbacks.

(i) The circuit's non-linear characteristic produces a harmonic distortion. The non-linearity is obvious from the fact that the slope is not the same at each point of the characteristics.

(ii) The circuit does not eliminate the amplitude variations and the output is sensitive to any amplitude variations in the input FM signal which is obviously not a desirable feature. A good discriminator circuit must respond only to frequency variations and not to amplitude variations.

4.24.2. Balanced Slope Detector

(U.P. Tech., Semester, Exam., 2004-05) (05 marks)

The balanced slope detector is an improvement over the simple slope detector. This circuit of balanced slope detector consists of two L-C circuits as shown in the figure 4.21. The frequency response for the circuit has been shown in figure 4.22.

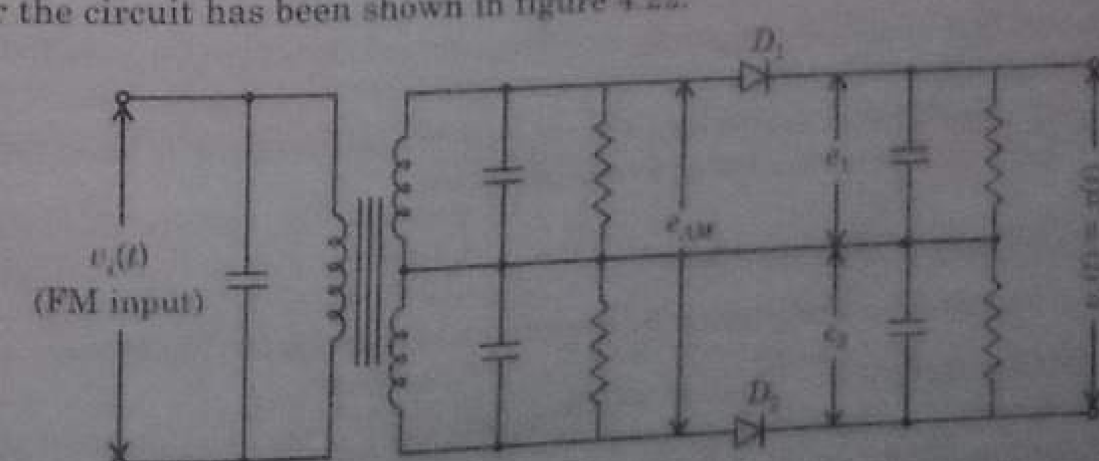


Fig. 4.21. *Circuit diagram of a balanced slope detector.*

In this circuit, the two tuned circuits are used in the stagger-tuned mode *i.e.*, one tuned circuit is tuned above the carrier frequency ω_c (curve e_1) and another tuned circuit is tuned below ω_c (curve e_2). The resultant curve $(e_1 + e_2)$ is linear as depicted by the dotted line in the figure 4.22.

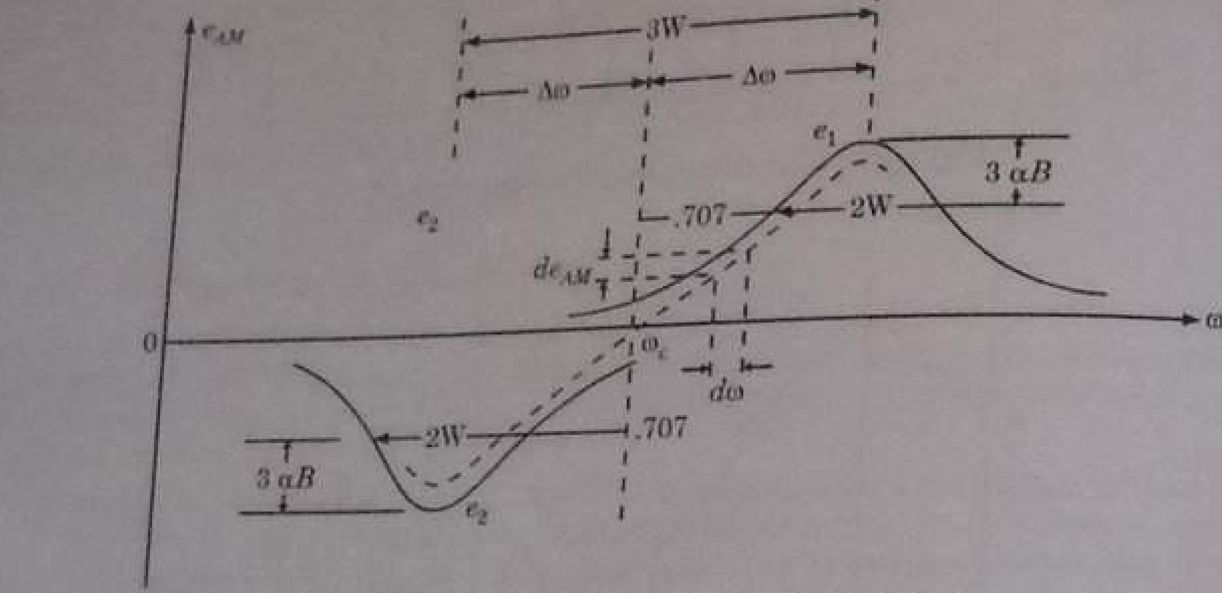


Fig. 4.22. *Balanced slope detector-characteristics.*

The slope $\frac{de_{AM}}{d\omega}$ *is identical at all the points on the linear portion of the characteristics curve. Hence, no harmonic distortion is caused when the operation is restricted to the linear region.*

4.24.3. Disadvantages of the Balanced Slope Detector

The balanced slope detector has few disadvantages as under:

(*i*) The linear characteristics is limited to a small frequency deviation $\Delta\omega$. The frequency deviation, $\Delta\omega$, for which the resultant characteristics is linear, depends upon the 3 dB bandwidth of each tuned circuit which is equal to 2W. The linearity is limited to a range of 3 W as shown in figure 4.22. Thus, the deviation $\Delta\omega$, for which the curve is linear and gives a satisfactory output is given as

$$2(\Delta\omega) = 3\ W$$

or

$$\Delta\omega = 1.5\ W$$

if the frequency deviation of an input FM wave is more than the one expressed by above equation, the distortion will occur due to the non-linearity of the characteristics. Hence, the operation of this detector is limited to small deviation only.

(*ii*) The discriminator characteristics depends critically upon the amount of detuning of the resonant circuits.

(*iii*) The tuned circuit output is not purely bandlimited and thus the low pass RC filter of the envelope detector introduces distortion.

However, in spite of all these drawbacks, this circuit can be designed to keep the distortions within tolerable limits and provide a satisfactory operation.

4.25. Phase Difference Detectors

In this article, we shall discuss Foster-Seeley detector and ratio detector one by one.

4.25.1. Foster-Seeley Detector

This type of detector is most widely used. Figure 4.23 shows the circuit arrangement for Foster-Seeley detector. This circuit consists of an inductively coupled double-tuned circuit in

which both primary and secondary coils are tuned to the same frequency (intermediate frequency). The centre of the secondary coil is connected to the top of the primary (collector end) through a capacitor C. This capacitor C perform the following functions:

(*i*) It blocks the d.c. from primary to secondary.

(*ii*) It couples the signal frequency from primary to centre tapping of the secondary.

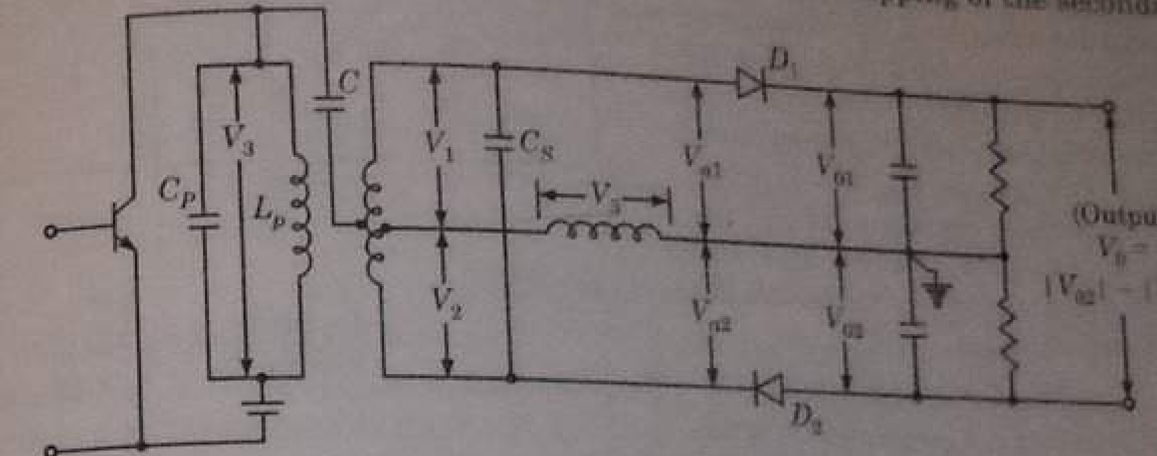


Fig. 4.23.

The primary voltage V_3 (*i.e.*, signal voltage) thus appears across the indicator L. Nearly entire voltage V_3 appears across inductor L except a small drop across the capacitor C. However by a suitable choice of C and L, the drop across the capacitor C can be kept negligible.

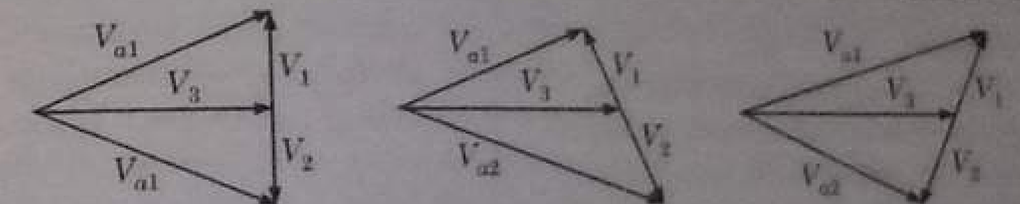


Fig. 4.24. *Phasor diagram.*

The centre-tapping of the secondary coil has an equal and opposite voltage across each half winding. Hence, V_1 and V_2 are equal in magnitude but opposite in phase. The radio frequency voltages V_{a1} and V_{a2} applied to the diodes D_1 and D_2 are expressed as

$$V_{a1} = V_3 + V_1$$

and

$$V_{a2} = V_3 - V_2$$

Voltages V_{a1} and V_{a2} depend upon the phasor relations between V_1, V_2 and V_3. The phasor diagrams for different frequencies have been shown in figure 4.24. The phasor diagrams V_1 and V_2 are always equal and are in phase opposition. However, the phase position of V_1 and V_2 relative to V_3 would depend upon the tuned secondary coil at the resonance or off the resonance as discussed under:

(*i*) At resonance: When an input voltage has a frequency equal to the resonant frequency of the tuned secondary, V_3 is in phase quadrature (90° out of phase) with V_1 and V_2. This has been shown in the first phasor diagram in figure 4.25. The resultant voltages V_{a1} and V_{a2} are equal in magnitude.

(*ii*) Off resonance: When the input voltage is off the resonant frequency f_{if} of the tuned secondary, the phase position of V_1 and V_2 relative to V_3 will be different from 90°. Let Q_s be the quantity factor of the tuned secondary coil. When an input signal frequency is above the resonant frequency f_{if} by an amount $(f_{if}/2Q_S)$ the phase difference between V_3 and V_1 is 45° as shown in the second phasor diagram. Because V_2 is in phase opposition of V_1 the phase difference between V_3 and V_2 is 135°. The phasor diagram reveals that V_{a1} is reduced where as V_{a2} is increased. The situation is reversed when the input voltage has a frequency below f_{if} which is evident from third phasor diagram. Hence the amplitude of the voltage V_{a1} and V_{a2} will vary with the instantaneous frequency f in the manner shown in figure 4.25.

The RF voltage V_{a1} and V_{a2} are separately rectified by the diodes D_1 and D_2 respectively to produce voltage V_{01} and V_{02}. The RF components are by passed by the capacitors leaving only modulating frequency component and a d.c. term. The voltage V_{01} and V_{02} then represent the amplitude variations of V_{a1} and V_{a2} respectively. The diodes are so arranged that the output voltage V_0 is equal to the arithmetic difference $|V_{02}| - |V_{01}|$.

i.e., $V_0 = |V_{02}| - |V_{01}|$

Hence, the voltage V_0 will vary with instantaneous frequency in accordance with the difference $|V_{a2}| - |V_{a1}|$, as shown by the dotted curve in figure 4.25(c). This dotted curve is known as discriminator characteristic in figure 4.25.

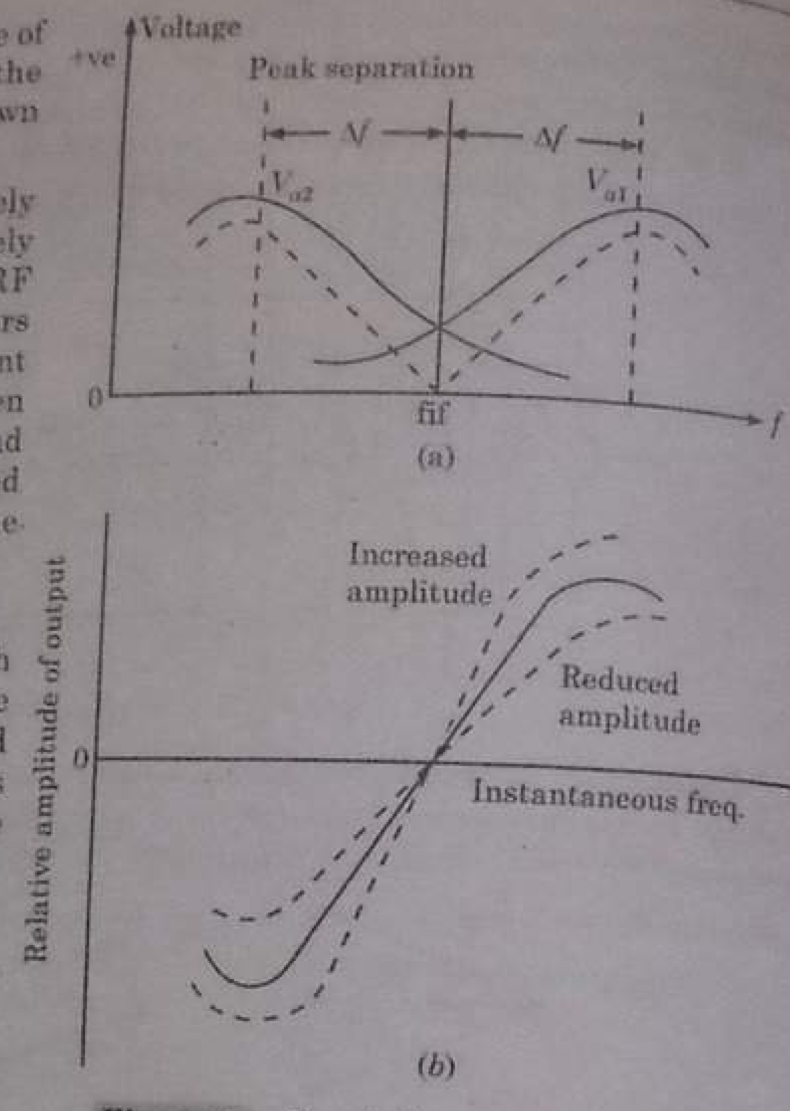


Fig. 4.25. *Discriminator characteristics.*

4.25.2. Ratio Detector

Ratio detector is an improvement over the Foster-Seeley discriminator and is most widely used. Since it does not respond to amplitude variations, therefore limiter is not needed. The circuit diagram of a ratio detector has been shown in figure 4.26. The circuit is similar to the circuit of Foster-Seeley discriminator (figure 4.26), except the following:

(i) The polarity of diode D_2 has been reversed.

(ii) The output V_o is taken from the center tap of a resistor R which shunts the load impedance of the two diodes. The output voltage varies with the input signal frequency (FM) exactly in the same way as it does in the Foster-Seeley discriminator, however its magnitude is reduced to half. This can be explained as under:

For instant, we shall ignore the two resistances (both denoted by R) and the capacitance C. The voltage V_{o1} and V_{o2} have the same magnitude as in the case of a Foster-Seeley discriminator, but V_{o2} is now reversed in polarity. Therefore, the voltage V_R is now sum of V_{o1} and V_{o2} unlike in the Foster-Seeley case, where it was the difference of V_{o2} and V_{o1}. Hence,

$$V_R = |V_{o1}| + |V_{o2}| \quad ...(4.85)$$

The output voltage V_0 is taken across the terminal t_1 and t_2

From the circuit diagram, we have

$$V_o = V_{t1, t2} = |V_{t, t1}| - |V_{t, t2}| = |V_{o2}| - \left|\frac{V_R}{2}\right|$$

But
$$V_R = |V_{o1}| + |V_{o2}|$$

Thus,
$$V_o = V_{o2} - \frac{V_{o1} - V_{o2}}{2} = \frac{V_{o2} - V_{o1}}{2}$$

Thus, the ratio detector has exactly same behaviour except that its output is reduced.

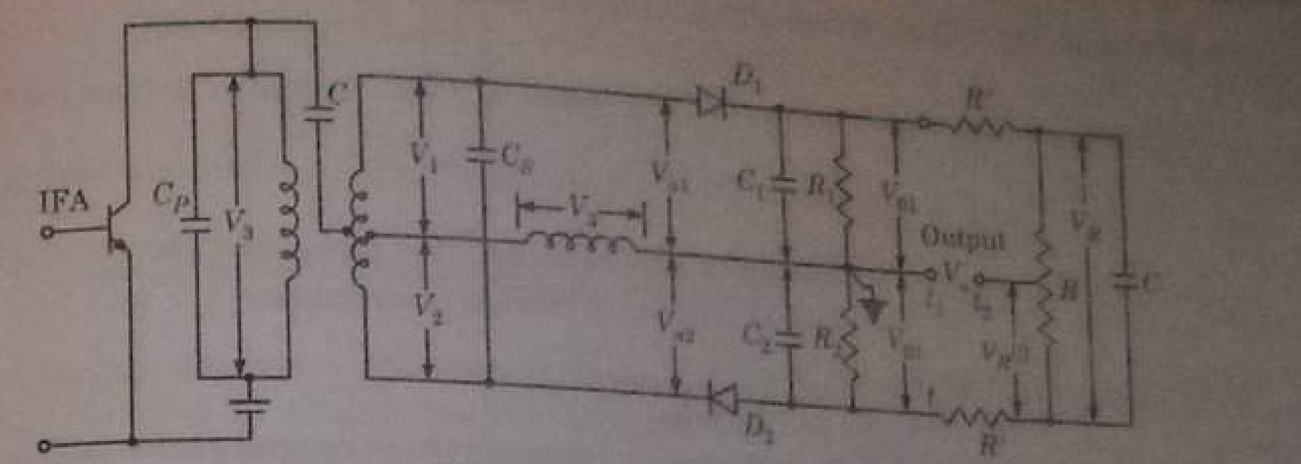


Fig. 4.26. *Ratio detector.*

4.25.3. PLL-FM Demodulator

A phase-locked loop (PLL) is primarily used in tracking the phase and frequency of the carrier component of an incoming FM signal. PLL is also useful for synchronous demodulation of AM-SC (*i.e.*, Amplitude Modulation with Suppressed Carrier) signals or signals with few cycles of pilot carrier. Further, PLL is also useful for demodulating FM signals in presence of large noise and low signal power.

This means that, PLL is most suitable for use in space vehicle-to-earth data links or where the loss along the transmission line or path is quite large. Recently, it has found application in commercial FM receivers.

A Phase-Locked Loop (PLL) is basically a negative feedback system. It consists of three major components. These components are multiplier, a loop filter and a voltage controlled oscillator (VCO) connected together in the from of a feedback loop. A VCO is a sine wave generator whose frequency is determined by the voltage applied to it from an external source.

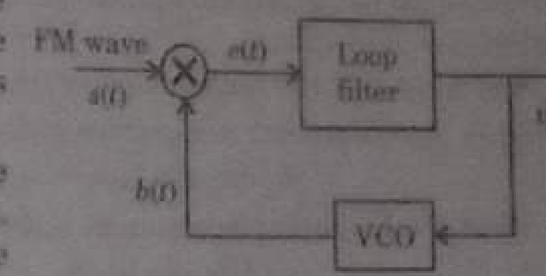


Fig. 4.27. *The block diagram of a Phase-Locked Loop (PLL).*

It means that any frequency modulator can work as a VCO.

Operation

The operation of a PLL is similar to any other feedback system. In any feedback system, th[e] feedback signal tends to follow the input signal. If the signal fedback is not equal to the inpu[t] signal, the error signal will change the value of the fedback signal until it is equal to the inpu[t] signal. The difference signal between $s(t)$ and $b(t)$ is called an **error signal**. A PLL operates on [a] similar principle except for the fact that the quantity fedback is not the amplitude, but a generalise[d] phase $\phi(t)$.

The error signal or difference signal $e(t)$ is utilised to adjust the VCO frequency in such a wa[y] that the instantaneous phase angle comes close to the angle of the incoming signal $s(t)$. At th[is] point, the two signals $s(t)$ and $b(t)$ are in synchronism and the PLL is locked to the incomin[g] signal $s(t)$.

The following mathematical steps would help us understand how FM demodulation of detectio[n] can be performed by using PLL.

Here, we have assumed that the VCO is adjusted initially so that when the control voltag[e] comes to zero, the following two conditions are satisfied:

(i) *The frequency of the VCO is precisely set at the unmodulated carrier frequency f_c and*

(ii) *The VCO output has a 90° phase-shift w.r.t. the unmodulated carrier wave.*

Let the input signal applied to the PLL be an FM wave. It is defined as

$$s(t) = A \sin[\omega_c t + \phi_1(t)] \quad ...(4.86)$$

where A is the unmodulated carrier amplitude and $\omega_c = 2\pi f_c$ = Angular carrier frequency

and $$\phi_1(t) = 2\pi k_f \int_0^t x(t)\, dt \quad ...(4.87)$$

where $x(t)$ is the message or baseband signal or modulating signal and

k_f = frequency sensitivity of frequency modulator.

Let the VCO output be defined by

$$b(t) = A_v \cos[\omega_c t + \phi_2(t)] \quad ...(4.88)$$

where A_v = Amplitude of VCO output when the control voltage applied to the VCO is denoted by $v(t)$, then, we have

$$\phi_2(t) = 2\pi k_v \int_0^t v(t)\, dt \quad ...(4.89)$$

Here, k_v is the frequency sensitivity of VCO, measured in Hertz/volt.

It may be observed from equations (4.86) and (4.88) that the VCO output and the incoming signals are 90° out of phase, while the VCO frequency in absence of $v(t)$ is precisely equal to the unmodulated frequency of the FM signal. The incoming FM have $s(t)$ and the VCO output $b(t)$ are applied to a multiplier.

The output of the multiplier has the following components:

(i) A high frequency component represented by

$$k_m A A_v \sin[2\omega_c t + \phi_1(t) + \phi_2(t)] \quad ...(4.90)$$

(ii) A low frequency component represented by

$$k_m A A_v \sin[\phi_1(t) - \phi_2(t)] \quad ...(4.91)$$

where k_m = Multiplier Gain measured in per volt.

The high frequency component can be eliminated by using a filter. Hence, discarding the high frequency component, the effective input to the low pass filter (LPF) will be given by

$$e(t) = k_m A A_v \sin[\phi_1(t) - \phi_2(t)]$$

or $$e(t) = k_m A A_v \sin[\phi_e(t)] \quad ...(4.92)$$

where $\phi_e(t)$ is the phase error and is expressed as

$$\phi_e(t) = \phi_1(t) - \phi_2(t)$$

This means that

$$\phi_e(t) = \phi_1(t) - 2\pi k_v \int_0^t v(t)\, dt \quad ...(4.93)$$

The loop filter operates on error signal $e(t)$ to produce the output $v(t)$. It is given by

$$v(t) = \int_{-\infty}^{\infty} e(\tau)\, h(t-\tau)\, d\tau \quad ...(4.94)$$

where $h(t)$ = Impulse response of the low-pass filter (LPF).

Using equations (4.92), (4.93) and (4.94), we get

$$\phi_e(t) = \phi_1(t) - 2\pi k_m k_v A A_v \int_0^t \int_{-\infty}^{\infty} \sin[\phi_e(\tau)]\, h(t-\tau)\, d\tau\, dt$$

or $$\phi_e(t) = \phi_1(t) - 2\pi k_o \int_0^t \int_{-\infty}^{\infty} \sin[\phi_e(\tau)]\, h(t-\tau)\, d\tau\, dt \quad ...(4.95)$$

where $$k_o = k_m k_v A A_v \quad ...(4.96)$$

Now, differentiating both sides of equation (4.95), we obtain

$$\frac{d\phi_e(t)}{dt} = \frac{d\phi_1(t)}{dt} - 2\pi k_o \int_{-\infty}^{\infty} \sin[\phi_e(\tau)]\, h(t-\tau)\, d\tau \quad ...(4.97)$$

Here k_o has the dimension of frequency. On the basis of equation (4.97), we can construct and equivalent model of PLL. It has been shown in figure 4.28.

In this model, $v(t)$ and $e(t)$ are also included utilising the relationship between them as given in equations $e(t) = k_m A A_v \sin[\phi_e(t)]$ and

$$v(t) = \int_{-\infty}^{\infty} e(\tau)]\, h(t-\tau)\, d\tau.$$

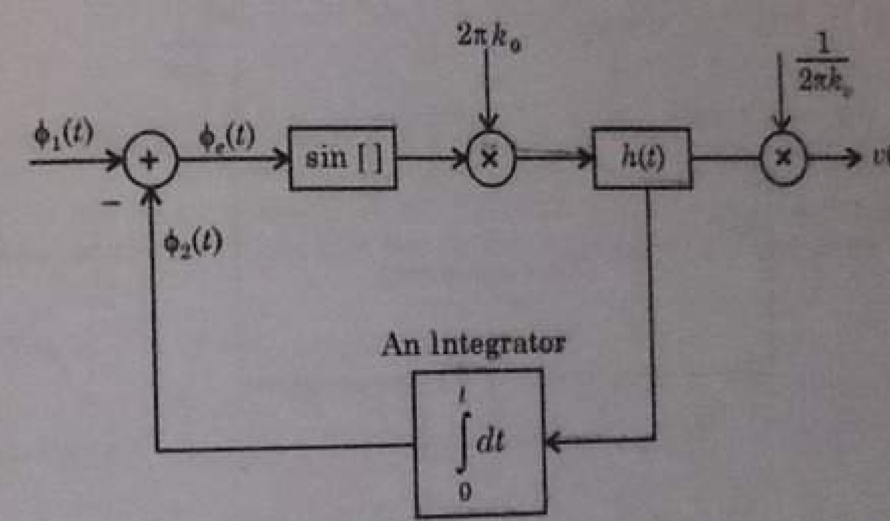


Fig. 4.28. *A non-linear equivalent model of PLL.*

A comparison of figures 4.27 and 4.28 reveals the fact that they are similar except for the fact that the multiplier in the equivalent model has been replaced by a subtractor and a sinusoidal non-linearity and the VCO by an integrator.

When the phase error $\phi_e(t)$ is zero, then PLL is said to be phase-locked. When the phase error $\phi_e(t)$ at all times is small compared to 1 radian, then we can approximate $\sin[\phi_e(t)]$ as $\phi_e(t)$, *i.e.*,

$$\sin[\phi_e(t)] \equiv \phi_e(t) \quad ...(4.98)$$

It is fairly accurate as long as $\phi_e(t)$ is less than 0.5 radian. In this case, PLL is said to be Near-Lock Condition and the sinusoidal non-linearity can be disregarded. The linearised model of PLL is valid under above-mentioned condition as shown in figure 4.29a. In this model, phase error $\phi_e(t)$ is related to the input phase $\phi_1(t)$ by the Integro-differential equation. It is expressed as

$$\frac{d\phi_e(t)}{dt} + 2\pi k_o \int_{-\infty}^{\infty} \phi_e(\tau)\, h(t-\tau)\, d\tau = \frac{d\phi_1(t)}{dt} \quad ...(4.99)$$

Taking the Fourier transform of both sides of equation (4.99), we obtain

$$\Phi_e(f) = \frac{1}{1 + k_o \frac{H(f)}{jf}} \Phi_1(f) \quad ...(4.100)$$

where $\Phi_e(f)$ and $\Phi_1(f)$ are the Fourier transform of $\phi_e(t)$ and $\phi_1(t)$, respectively and $H(f)$ is the Fourier transform of impulse response $h(t)$ and is known as transfer function of the loop filter.

The quantity $k_o H(f)/jf$ is called the open loop transfer function of the PLL.

$$L(f) = \frac{k_0 H(f)}{jf} \quad \text{...(4.100(a))}$$

Substituting equation (4.100(a)) in equation (4.100), we obtain

$$\Phi_e(f) = \frac{1}{1+L(f)}\Phi_1(f) \quad \text{...(4.101)}$$

Now, let us consider that for all values of frequency f inside the baseband signal, we make the magnitude of $L(f)$ very large compared to unity.

Thus, from equation (4.101), we obtain

$$\Phi_e(f) \to 0 \text{ as } L(f) >> 1 \quad \text{...(4.102)}$$

Under above-mentioned condition, the phase of the VCO becomes asymptotically equal to the phase of the incoming wave and the phase lock is thereby established.

From figure 4.29(a), we observe that $V(f)$ is related to $\Phi_e(f)$ by

$$V(f) = \frac{k_o}{k_v} H(f)\Phi_e(f) \quad \text{...(4.103)}$$

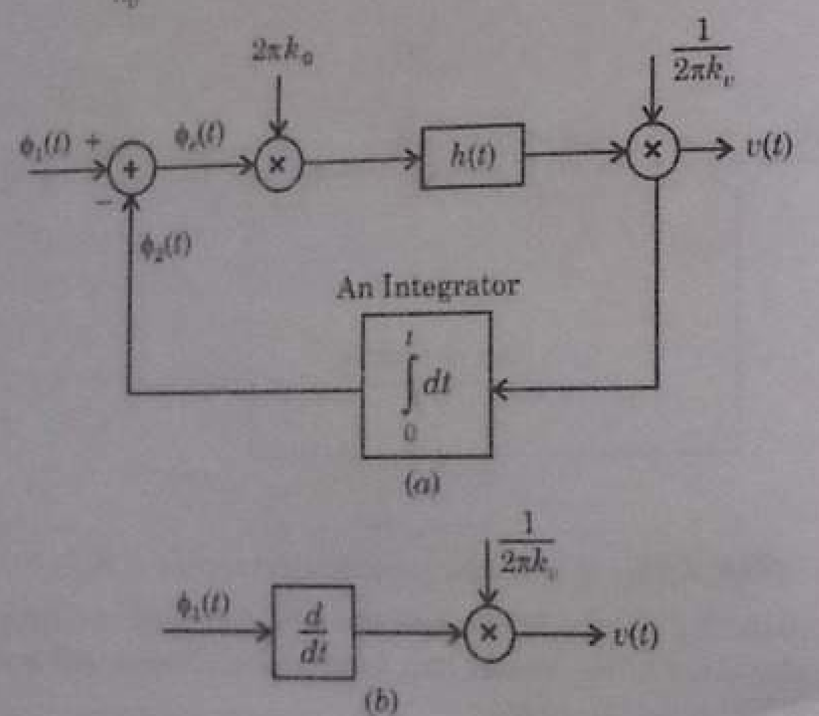


Fig. 4.29. *Illustration of an equivalent model of PLL.*

Since, we have

$$L(f) = k_o \frac{H(f)}{jf} \quad \text{...(4.104)}$$

Above equation (4.104), may be obtained by taking Fourier transform of both sides of equation given below:

$$v(t) = \int_{-\infty}^{\infty} e(\tau)\, h(t-\tau)\, d\tau$$

and using the approximate of equation, $\sin[\phi_e(t)] \cong \phi_e(t)$.

Now, substituting the values of $H(f)$ from equation (4.104) in equation (4.103), we obtain

$$V(f) = \frac{k_o}{k_v}\left[\frac{jf}{k_o}L(f)\right]\Phi_e(f) = \frac{jf}{k_v}L(f)\Phi_e(f) \quad \text{...(4.105)}$$

Substituting the values of $\Phi_e(f)$ from equation (4.101) in equation (4.105), we have

$$V(f) = \frac{jf}{k_v}L(f)\Phi_e(f) = \frac{jf}{k_v}L(f)\frac{1}{1+L(f)}\Phi_1(f)$$

or

$$V(f) = \left(\frac{jf}{k_v}\right)\left(\frac{L(f)}{1+L(f)}\right)\Phi_1(f) \quad \text{...(4.106)}$$

If $|L(f) >> 1|$ then equation (4.106) may be approximated as

$$V(f) = \left(\frac{jf}{k_v}\right)[1]\Phi_1(f) = \left(\frac{jf}{k_v}\right)\Phi_1(f) \quad \text{...(4.107)}$$

The corresponding time-domain representation of equation (4.107) can be obtained by taking inverse Fourier transform of both sides of equation (4.107). Hence, we have

$$\text{Inverse FT }[V(f)] = \text{Inverse } FT\left[\left(\frac{jf}{k_v}\right)\Phi_1(f)\right]$$

or

$$v(t) = \frac{1}{2\pi k_v}\frac{d\phi_1(t)}{dt} \quad \text{...(4.108)}$$

Conclusion

Here, it can be concluded that if the magnitude of $L(f)$ is very large for all frequencies of interest, then PLL may be modeled as a differentiator with its output scaled by a factor $1/2\pi K_v$. It has been illustrated in figure 4.29(b). The simplified model of PLL shown in figure 4.29(b) may be used as an FM demodulator. This can be easily verified by substituting the value of $\phi_1(t)$ from equation $\phi_1(t) = 2\pi k_f \int_0^t x(t)\, dt$ into equation (4.108).

Therefore, we have

$$v(t) = \frac{1}{2\pi k_v}\frac{d\phi_1 t}{dt} = \frac{1}{2\pi k_v}\frac{d}{dt}\left[2\pi k_f \int_0^t x(t)\, dt\right]$$

or

$$v(t) = \frac{2\pi k_f}{2\pi k_v}\frac{d}{dt}\left[\int_0^t x(t)\, dt\right] = \frac{k_f}{k_v}x(t) \quad \text{...(4.109)}$$

Hence, we can say that output $v(t)$ of PLL is approximately same except for a scaling factor k_f/k_v, as the original baseband or modulating signal $x(t)$ and the frequency demodulation is performed.

Note: *It may be noted that the incoming FM wave can have much wider bandwidth than that of the loop filter with transfer function H(f). Here H(f) is restricted to baseband. The control signal of the VCO has a bandwidth of the baseband signal whereas the output of the VCO is a wide band FM wave. The instantaneous frequency of the WBFM wave tracks the incoming FM. The complexity of a PLL is obtained by the transfer function H(f) of the loop filter. If H(f) = 1 then PLL is called simplest PLL. This means that there is no loop filter and the PLL is referred to as first-order PLL. The order of the PLL is determined by the order of the denominator polynomial of the closed loop transfer function. The transfer function H(f) determines the output Fourier transform V(f) in terms of input Fourier transform $\Phi_1(f)$. It is given by equation (4.107).*

MISCELLANEOUS SOLVED EXAMPLES

Example 4.13. Determine the permissible range in maximum modulation index for

(i) Commerical FM which has 30 Hz to 15 kHz modulating frequencies.

(ii) narrowband FM system which allows maximum deviation of 10 kHz and 100 Hz to 3 kHz modulating frequencies.

Solution: (i) The maximum deviation in commerical FM is given as

$$\Delta f = 75 \text{ kHz}$$

Modulation index in FM is

$$m_f = \frac{\Delta f}{f_m} \quad \text{...(i)}$$

Modulation index for commerical FM at $f_m = 30$ Hz is

$$m_f = \frac{\Delta f}{f_m} = \frac{75 \times 10^3}{30} = 2500$$

Modulation index for commerical FM at $f_m = 15$ kHz is

$$m_f = \frac{\Delta f}{f_m} = \frac{75 \times 10^3}{15 \times 10^3} = 5$$

Hence, the modulation index for commerical FM varies between 2500 and 5.

(ii) For a given Narrowband FM system, the maximum frequency deviation is given as $\Delta f =$ 10 kHz.

Hence, modulation index for a given NBFM system varies between

$$m_f = \frac{\Delta f}{f_m} = \frac{10 \times 10^3}{100} = 100$$

and

$$m_f = \frac{\Delta f}{f_m} = \frac{10 \times 10^3}{3 \times 10^3} = 3.33 \quad \textbf{Ans.}$$

Example 4.14. A 100 MHz carrier wave has a peak voltage of 5 volts. The carrier is frequency modulated (FM) by a sinusoidal modulating signal or waveform of frequency 2 kHz such that the frequency deviation Δf is 75 kHz. The modulated waveform passes through zero and is increasing at $t = 0$. Determine the expression for the modulated carrier waveform.

Solution: Because the frequency modulated carrier waveform passes through zero and is increasing at $t = 0$, therefore, the FM signal must be sine wave (signal). Thus

$$s(t) = A\sin\left[2\pi f_c t + m_f \sin(2\pi f_m t)\right] \quad \text{...(i)}$$

where m_f = modulation index of FM wave = $\frac{\Delta f}{f_m}$

Given that f_c = Carrier wave frequency = $100 \times 10^6 = 10^8$ hertz

Δf = frequency deviation = 75 kHz = 75×10^3 hertz

f_m = modulating frequency = 2 kHz = 2×10^3 hertz

A = peak voltage of carrier wave = 5 volt

Now,

$$m_f = \frac{\Delta f}{f_m} = \frac{75 \times 10^3}{2 \times 10^3} = 37.5$$

Substituting all the above values in equation (i), we get

$$s(t) = 5\sin\left[2\pi \times 10^8 t + 37.5\sin(2\pi \times 2 \times 10^3 t)\right]$$

or

$$s(t) = 5\sin\left[2\pi \times 10^8 t + 37.5\sin(4\pi \times 10^3 t)\right] \quad \textbf{Ans.}$$

Example 4.15. A carrier wave of frequency 1 GHz and amplitude 3 volts is frequency modulated (FM) by a sinusoidal modulating signal frequency of 500 Hz and of peak amplitude 1 volt. The frequency deviation Δf is 1 kHz. The level of the modulating waveform (signal) is changed to 5 volt peak and the modulating frequency is changed to 2 kHz. Obtain the expression for the new modulated waveform (FM).

Solution: We know that the FM wave is given by the expression

$$s(t) = A\cos\left[2\pi f_c t + m_f \sin(2\pi f_m t)\right]$$

where m_f = Modulation index of FM wave = $\frac{\Delta f}{f_m}$

and Δf = frequency deviation = $k_f A_m$

k_f = Sensitivity of frequency modulator

A_m = Amplitude of the modulating signal

Given that $f_c = 1 \text{ MHz} = 1 \times 10^6$ Hz

$A = 3$ volt

$A_m = 1$ volt

and $\Delta f = 1$ kHz

Therefore, k_f can be found as

$$k_f = \frac{\Delta f}{f_m} = \frac{1 \times 10^3}{1} = 10^3 \text{ Hz/volt}$$

Now, for the second case, we have

when, $A_m = 5$ volt and $f_m = 2$ kHz

Modulation Index will be

$$m_f = \frac{\Delta f}{f_m} = \frac{10^3 \times 5}{2 \times 10^3} = 2.5$$

The desired FM signal can be expressed by

$$s(t) = A\cos\left[2\pi f_c t + m_f \sin(2\pi f_m t)\right]$$

Substituting all the values, we get

$$s(t) = 3\cos\left[2\pi 10^6 t + 2.5\sin(2\pi \times 2 \times 10^3 t)\right]$$

or

$$s(t) = 3\cos\left[2\pi 10^6 t + 2.5\sin(4\pi \times 10^3 t)\right] \quad \textbf{Ans.}$$

Example 4.16. Given a signal

$$s(t) = \cos(2\pi f_c t) + 0.2\cos(2\pi f_m t)\sin(2\pi f_c t)$$

(i) Prove that $s(t)$ is a combination of AM-FM signal

(ii) Draw the phasor diagram at $t = 0$. *(U.P. Tech., Sem., Exam., 2004-05) (05 marks)*

Solution: (i) The given signal $s(t)$ can be modified in the following form:

$$s(t) = \cos(2\pi f_c t) + 0.2\cos(2\pi f_m t)\sin(2\pi f_c t)$$

$$s(t) = \left[1 + \{0.2\cos(2\pi f_m t)\}^2\right]^{1/2} \times \cos\left[\{2\pi f_c t - \tan^{-1}(0.2\cos(2\pi f_m t))\}\right]$$

$$s(t) = \left[1 + \frac{0.04}{2}\cos^2(2\pi f_m t)\right]\cos\left[(2\pi f_c t) - 0.2\cos(2\pi f_m t)\right]$$

or $$s(t) = \left[1 + 0.01\cos(4\pi f_m t)\right]\cos\left[(2\pi f_c t) - 0.2\cos(2\pi f_m t)\right] \quad ...(i)$$

From equation (i), we know that the amplitude as well as the instantaneous phase angle of the given signal changes in accordance with the modulating or message signal. Hence, the given signal s(t) is a combination of AM-FM signal.

(*ii*) For the construction of phasor diagram, we can express the signal $s(t)$ as under:

$$s(t) = \cos(2\pi f_c t) + 0.2\cos(2\pi f_m t)\sin(2\pi f_c t)$$

or $$s(t) = \cos(2\pi f_c t) + \frac{0.2}{2}\sin\left[2\pi(f_c + f_m)t\right] + \frac{0.2}{2}\sin\left[2\pi(f_c - f_m)t\right] \quad ...(ii)$$

Phasor diagram for equation (*ii*) has been shown in figure 4.30.

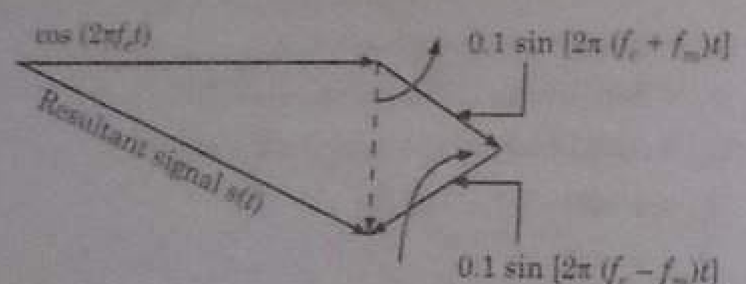


Fig. 4.30. *Illustration of Phasor diagram for example 4.19.*

Example 4.17. (*i*) **Show with a phasor diagram that the signal $s(t)$ expressed by**

$$s(t) = \cos(2\pi\times 10^6 t) + 0.02\cos\left[2\pi(10^6 + 10^3)t\right]$$

represents a carrier wave which is modulated both in amplitude and frequency.

(*ii*) **Represent $s(t)$ in the form**

$$s(t) = \left[1 + m\cos(2\pi\times 10^6 t)\right]\cos\left[2\pi\times 10^6 t + m_f\sin(2\pi\times 10^3 t)\right]$$

Find the value of m and m_f. Also, write an expression for instantaneous frequency as a function of time t. Verify that both amplitude and frequency vary approximately sinusoidally with frequency 1 kHz.

Solution: (*i*) The phasor diagram of the signal

$$s(t) = \cos(2\pi\times 10^6 t) + 0.02\cos\left[2\pi(10^6 + 10^3)t\right]$$

has been shown in figure 4.31.

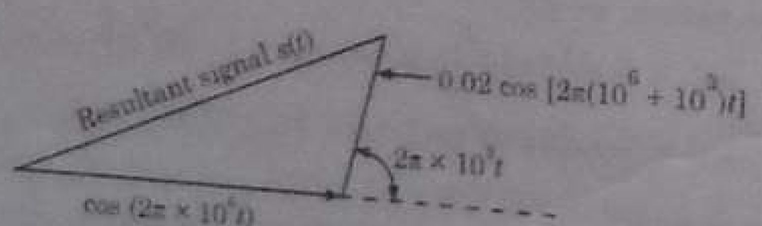


Fig. 4.31. *Phasor diagram.*

Observation

It may be easily observed from this figure that the resultant amplitude differs from the carrier ampltude A and the resultant is not in phase with the carrier. Hence, the given signal is modulated both in amplitude and phase.

Hence Proved.

(*ii*) The given signal $s(t)$ can be expressed as

$$s(t) = \cos(2\pi\times 10^6 t) + 0.02\cos\left[2\pi(10^6 + 10^3)t\right]$$

or $$s(t) = \cos(2\pi\times 10^6 t) + 0.02\cos(2\pi\times 10^6 t)\cos(2\pi\times 10^3 t)$$

$$-0.02\sin(2\pi\times 10^6 t)\sin(2\pi\times 10^3 t)$$

or $$s(t) = \left[1 + 0.02\cos(2\pi\times 10^3 t)\right]\cos(2\pi\times 10^6 t) - 0.02\sin(2\pi\times 10^6 t)\sin(2\pi\times 10^3 t)$$

$$s(t) = \left\{\left[1 + 0.02\cos(2\pi\times 10^3 t)\right]^2 + \left[0.02\sin(2\pi\times 10^3 t)\right]^2\right\}^{1/2} \times$$

$$\cos\left\{2\pi\times 10^6 t + \tan^{-1}\frac{0.02\sin(2\pi\times 10^3 t)}{1 + 0.02\cos(2\pi\times 10^3 t)}\right\}$$

or $$s(t) \cong \left[1 + 0.02\cos(2\pi\times 10^3 t)\right]\cos\left[2\pi\times 10^6 t + 0.02\sin(2\pi\times 10^3 t)\right]$$

or $$s(t) \cong \left[\left[1 + m\cos(2\pi\times 10^3 t)\right]\cos\left[2\pi\times 10^6 t + m_f\sin(2\pi\times 10^3 t)\right]\right]$$

Hence, Modulation index of AM part, $m = 0.02$ and Modulation Index of FM part, $m_f = 0.02$

The instantaneous frequency is found as

$$f_i = \frac{1}{2\pi}\frac{d}{dt}\left[\theta_i(t)\right] = \frac{1}{2\pi}\frac{d}{dt}\left[2\pi\times 10^6 t + 0.02\sin(2\pi\times 10^3 t)\right]$$

or $$f_i = 10^6 + 20\cos(2\pi\times 10^3 t)$$

It can be observed that the amplitude and instantaneous frequency of modulated signal vary with cos $(2\pi\times 10^3 t)$. **Ans.**

Example 4.18. Given a Narrowband Frequency Modulated (NBFM) signal as

$$s(t) = \cos\left[\omega_c t + \phi(t)\right]$$

where ω_c = **angular carrier frequency** $= 2\pi f_c$

and $\phi(t) = m_1\sin(2\pi f_1 t) + m_2\sin(2\pi f_2 t)$

so that $|\phi(t)| << \frac{\pi}{2}$

Show that the superposition principle is satisfied by NBFM.

Solution: The Narrowband FM (NBFM) is given by

$$s(t) = \cos\left[\omega_c t + \phi(t)\right] \quad ...(i)$$

where $\omega_c = 2\pi f_c$

This signal can be considered to be simultaneously modulated by two modulating signals of frequency f_1 and f_2 and thus can be written as under:

$$s(t) = \cos(\omega_c t)\cos\left[\phi(t)\right] - \sin(\omega_c t)\sin\left[\phi(t)\right] \quad ...(ii)$$

For $|\phi(t)| << \frac{\pi}{2}$, we have

$$\cos\left[\phi(t)\right] \cong 1 \quad \text{and} \quad ...(iii)$$

$$\sin\left[\phi(t)\right] \cong \phi(t)$$

Substituting equation (*iii*) in equation (*ii*), we get

$$s(t) = \cos(\omega_c t)1 - \sin(\omega_c t)\,\phi(t)$$

or $$s(t) = \cos(\omega_c t) - \sin(\omega_c t)\,\phi(t) \quad ...(iv)$$

or $$s(t) = \cos(2\pi f_c t) - \sin(2\pi f_c t)\,\phi(t)$$

Now, substituting equation (i) in equation (iv), we get

$$s(t) = \cos(2\pi f_c t) - \sin(2\pi f_c t)\,\phi(t)$$

or
$$s(t) = \cos(2\pi f_c t) - \sin(2\pi f_c t)\left[m_1 \sin(2\pi f_1 t) + m_2 \sin(2\pi f_2 t)\right] \quad ...(v)$$

Therefore, superposition principle holds for this NBFM. In other words, NBFM can be considered approximately, as a linear modulation scheme.

Example 4.19. A sinusoidal carrier wave of frequency of f_c = 100 MHz is exponentially modulated by a modulating signal $x(t)$. It has been shown in figure 4.32. Find the bandwidth when the resulting signal is

(i) FM with sensitivity $k_f = 10^5$ Hertz/volt.

(ii) PM with sensitivity $k_p = 10\pi$ radians/volt.

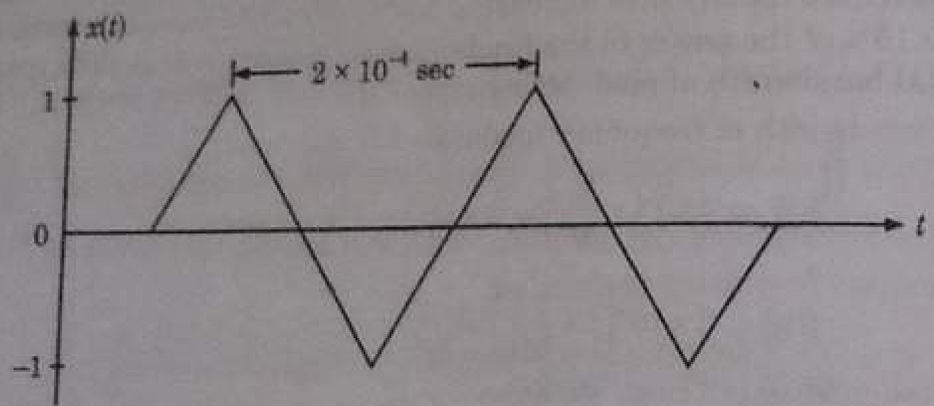


Fig. 4.32. *For example 4.19.*

Solution: We know that for FM, the instantaneous frequency is given by

$$f_i = f_c + k_f x(t) = f_c + 10^5\, x(t) \quad ...(i)$$

or
$$[f_i]_{max} = f_c + 10^5 [x(t)]_{max} \quad ...(ii)$$

From figure 4.30, we have $x(t)_{max} = 1$.

Substituting $[x(t)]_{max} = 1$ in equation (ii), we get

$$[f_i]_{max} = f_c + 10^5\,[x(t)]_{max} = f_c + 10^5 \times 1 = f_c + 10^5 = 100 \times 10^6 + 10^5$$

or
$$[f_i]_{max} = 100.1 \times 10^6 \text{ Hertz} = 100.1 \text{ MHz}$$

Again, from figure 4.30, $[x(t)]_{min} = -1$

Substituting the $[x(t)]_{min} = -1$ in equation (i), we get

$$[f_i]_{min} = 100 \times 10^6 + 10^5(-1) = 100 \times 10^6 - 0.1 \times 10^6$$
$$= 99.9 \times 10^6 \text{ Hertz} = 99.9 \text{ MHz}$$

Frequency deviation Δf can be found as

$$\Delta f = [f_i]_{max} - f_c = (101.1 - 100) \text{ MHz}$$
$$= 0.1 \text{ MHz} = 10^5 \text{ Hertz}$$

The bandwidth of the resulting signal may be calculated using Carson's rule as under:

$$BW = 2\Delta f\left(1 + \frac{f_m}{\Delta f}\right) = 2\Delta f\left(1 + \frac{1}{D}\right) \quad ...(iii)$$

where $D = \frac{\Delta f}{f_m}$

f_m = Essential bandwidth of the modulating signal $x(t)$. This can be determined by expanding $x(t)$ shown in figure 4.30 in Fourier series.

The modulating signal $x(t)$ can be expanded as

$$x(t) = \sum_{n=-\infty}^{\infty} C_n \cos(2\pi n f_o t) \quad ...(iv)$$

where,
$$f_o = \frac{1}{T_o} = \frac{1}{2 \times 10^{-4}} = 5000 \text{ Hz} = 5 \text{ kHz}$$

and Fourier series coefficient

$$C_n = \begin{cases} \frac{8}{\pi^2 n^2}; \text{ for odd values of } n \\ 0 \;\; ; \text{ for even values of } n \end{cases} \quad ...(v)$$

Now, it may be observed that the amplitudes of the various harmonic frequency components present in $x(t)$ decreases rapidly with increasing n. The powers of third harmonic and fifth harmonic are 1.21% and 0.16% of the power of the fundamental frequency. It is thus reasonable to assume that the essential bandwidth of modulating signal $x(t)$ is $3f_0 = 3 \times 5 \times 10^3 = 1.5 \times 10^4$ Hertz.

Hence, the bandwidth of frequency modulated (FM) wave is given by

$$BW = 2\Delta f\left(1 + \frac{f_m}{\Delta f}\right) = 2 \times 10^5\left(1 + \frac{1.5 \times 10^4}{2 \times 10^5}\right)$$

or
$$BW = 2 \times 10^5\left(1 + \frac{1.5}{20}\right) = 230 \times 10^3 \text{ Hz} = 230 \text{ kHz}$$

For the phase modulated case, we have

$$\theta_i(t) = 2\pi f_c t + k_p x(t)$$

Instantaneous frequency is found as

$$f_i = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} = \frac{1}{2\pi}\frac{d}{dt}\left[2\pi f_c t + k_p x(t)\right]$$

or
$$f_i = f_c + \frac{k_p}{2\pi}\frac{d}{dt}[x(t)] \quad ...(vi)$$

Maximum and minimum values of f_i will be

$$[f_i]_{max} = 100 \times 10^6 + \frac{10\pi}{2\pi}\frac{d}{dt}[x(t)]\Big|_{max}$$

or
$$[f_i]_{max} = 100 \times 10^6 + 5 \times 20{,}000 = 100 \times 10^6 + 0.1 \times 10^6$$

or
$$[f_i]_{max} = 100.1 \times 10^6 \text{ Hertz} = 100.1 \text{ MHz}$$

Further, $[f_i]_{max} = 99.9$ MHz

Hence, frequency deviation will be

$$\Delta f = [f_i]_{max} - f_i = [100.1 - 100] \text{ MHz} = 0.1 \text{ MHz} = 10^5 \text{ Hertz}.$$

Note: *The values of* $\frac{d[m(t)]}{dt}\Big|_{max}$ *has been obtained by noting that* $\frac{d[m(t)]}{dt}$ *switches back and forth between −20,000 and 20,000.*

The bandwidth of the PM signal is given by

$$BW_{PM} = 2\Delta f\left(1 + \frac{f_m}{\Delta f}\right) = 2 \times 10^5\left(1 + \frac{1.5 \times 10^4}{2 \times 10^5}\right) = 230 \text{ kHz} \quad \textbf{Ans.}$$

Example 4.20. Find the bandwidth of commercial FM transmission assuming frequency deviation Δf = 75 kHz and bandwidth of modulating signal $x(t)$, f_m = 15 kHz.

Solution: The deviation ratio for commerical FM transmission is given by

$$D = \frac{\Delta f}{f_m} = \frac{75 \text{ kHz}}{15 \text{ kHz}} = 5$$

Bandwidth of FM signal can be calculated by using Carson's rule as

$$BW = 2\Delta f\left(1+\frac{f_m}{\Delta f}\right) = 2\Delta f\left(1+\frac{1}{D}\right)$$

or $$BW = 2\times 75\times 10^3\left(1+\frac{1}{5}\right) = 150\times 10^3\left(\frac{6}{5}\right)$$

or $$BW = 180 \times 10^3 \text{ Hertz} = 180 \text{ kHz}$$

Also, using universal curve, replacing m_f by D, we get

$$BW = 3.2 \times \Delta f \text{ for } D = 5 = 3.2 \times 75 \times 10^3$$
$$= 240 \times 10^3 \text{ Hertz} = 240 \text{ kHz}$$

Now, we can find per centage of underestimation of bandwidth by using Carson's rule as % under estimation of bandwidth will be

$$= \frac{240-180}{240}\times 100 = \frac{60}{240}\times 100 = 25\%$$

It means that Carson's rule underestimates the bandwidth by 25% as compared with the result obtained from the universal curve. **Ans.**

Example 4.21. A rule of bandwidth for FM signal is sometimes used as

$$BW = (2m_f + 1)f_m$$

Find the fraction of the signal power that is included in that frequency band. Assume that $m_f = 1$ and 10.

Solution: The bandwidth for FM signal can be calculated on the basis of 98% power requirement given by Carson's rule as

$$BW = 2\Delta f\left(1+\frac{1}{m_f}\right) = 2(m_f + 1)f_m \quad ...(i)$$

The fraction of signal power included in the frequency band B is

$$P = \frac{B}{BW}\times\left(\frac{98}{100}\right) \quad ...(ii)$$

For $m_f = 1$ $$P = \frac{B}{BW}\times\left(\frac{98}{100}\right) = \frac{B}{BW}\times 0.98$$

or $$P = \frac{(2m_f+1)f_m}{2(m_f+1)f_m}\times 0.98 = \frac{2\times 1+1}{2(1+1)}\times 0.98$$

or $$P = \frac{2\times 1+1}{2(1+1)}\times 0.98 = \frac{3}{4}\times 0.98 = 73.5\%$$

For $m_f = 10$ $$P = \frac{B}{BW}\times 0.98 = \frac{(2m_f+1)f_m}{2(m_f+1)f_m}\times 0.98$$

or $$P = \frac{(2\times 10+1)}{2(10+1)}\times 0.98$$

or $$P = \frac{21}{22}\times 0.98 = 97.6\% \quad \textbf{Ans.}$$

Example 4.22. A carrier is frequency modulated (FM) by a sinusoidal modulating signal x(t) of frequency 2 kHz, it results in a frequency deviation Δf of 5 kHz. Find the bandwidth occupied by the FM waveform. The amplitude of the modulating sinusoid is increased by a factor of 3 and its frequency lowered by 1 kHz. Find the new bandwidth.

Solution: Given that

$$f_m = 2 \text{ kHz and}$$
$$\Delta f = 5 \text{ kHz}$$

Hence, the bandwidth of the FM signal will be

$$BW = 2(\Delta f + f_m) = 2(5+2)$$

or $$BW = 14 \text{ kHz}$$

When the amplitude of the modulating signal is trippled, then the frequency deviation increases three times.

Therefore, $$\Delta f = 3 \times 5 \text{ kHz} = 15 \text{ kHz}$$

Also, $$f_m = 1 \text{ kHz}$$

The new bandwidth will be

$$BW' = 2(\Delta f + f_m) = 2(15+1)$$

or $$BW' = 32 \text{ kHz} \quad \textbf{Ans.}$$

Example 4.23. Determine the relative power of the carrier wave and side frequencies when modulation index $m_f = 0.20$ for 10 kW FM transmitter.

Solution: For $$m_f = 0.20$$
$$J_0(m_f) = 0.99$$

(See Table 4.1 of Bessel Functions)

Thus, only significant side frequency pair is $f_c \pm f_m$ with relative amplitude,

$$J_1(m_f) = 0.099$$

Hence, the carrier power will be

$$P_c = J_0^2(m_f) \times \text{Power of FM transmitter}$$
$$P_c = (0.99)^2 \times 10 \times 10^3 = 9.8 \times 10^3 \text{ watt}$$
$$P_c = 9.8\ K \text{ watt} \quad \textbf{Ans.}$$

and the power in each side frequency is expressed as

$$P_{s1} = P_{s2}$$
$$= J_1^2(m_f) \times \text{Power of FM transmitter}$$
$$= (0.099)^2 \times 10 \times 10^3 = 98 \text{ watt.} \quad \textbf{Ans.}$$

Example 4.24. In an FM, system a 7 kHz modulating (or baseband) signal modulates 107.6 MHz carrier wave so that the frequency deviation is 50 kHz. Find

(i) carrier swing in the FM signal and modulating index m_f

(ii) the highest and lowest frequencies attained by the FM signal.

Solution: (i) Given that frequency deviation

$$\Delta f = 50 \text{ kHz} = 50 \times 10^3 \text{ Hertz}$$

Carrier swing in FM signal will be

$$= 2\Delta f = 2 \times 50 \times 10^3 \text{ Hz} = 100 \text{ kHz}$$

Modulation Index of FM wave,

$$m_f = \frac{\Delta f}{f_m} = \frac{50\times 10^3 \text{ Hertz}}{7\times 10^3 \text{ Hertz}} = 7.143$$

(ii) The upper or highest frequency attained by FM signal will be

$$= f_c + \Delta f$$
$$= 107.6 \times 10^6 + 50 \times 10^3 = 107.65 \text{ MHz}$$

The lower or lowest frequency attained by FM signal will be

$$= f_c - \Delta f = 107.6 \times 10^6 - 50 \times 10^3$$
$$= 107.55 \text{ MHz} \quad \textbf{Ans.}$$

Example 4.25. Determine the frequency deviation Δf and carrier swing for an FM signal which has a carrier frequency of 100 MHz and whose upper frequency is 100.007 MHz when modulated by a particular modulating signal or wave. Also find the lowest frequency reached by the FM wave.

Solution: We know that frequency deviation Δf is defined as the maximum change in frequency of the modulated signal away from the carrier frequency f_c.

This means that

i.e. $$\Delta f = f_u - f_c = [100.007 - 100.000] \text{ MHz}$$
$$= 0.007 \text{ MHz} = 7 \times 10^3 \text{ Hz} = 7 \text{ kHz}$$

$$\text{Carrier swing} = 2\Delta f = 2 \times 7 = 14 \text{ kHz}$$

The lowest frequency f_L reached by the modulated FM wave is equal to the difference of the frequency deviation from the carrier frequency.

$$f_L = f_c - \Delta f = (100.00 - 0.007) \text{ MHz}$$

or $$f_L = 99.993 \text{ MHz} \quad \textbf{Ans.}$$

Example 4.26. Determine the modulation index m_f of an FM signal which is being broadcast in the 88 – 108 MHz band. This FM wave has a carrier swing of 125 kHz.

Solution: We know that the frequency deviation is given by

$$\Delta f = \frac{\text{Carrier swing}}{2} = \frac{125 \times 10^3}{2} \text{ Hz}$$

or $$\Delta f = 62.5 \times 10^3 \text{ Hz} = 62.5 \text{ kHz}$$

Since, maximum frequency deviation for the FM broadcast band is 75 kHz, therefore

$$m_f = \text{modulation index}$$

$$m_f = \frac{\Delta f}{f_m} = \frac{62.5}{75} = 83.3\% \quad \textbf{Ans.}$$

Example 4.27. The modulating signal in an FM wave is 500 Hz with amplitude 3.2 volt and frequency deviation is 6.4 kHz. If the audio frequency voltage is now increased to 8.4 volt, determine the new frequency deviation and modulation index. If the audio frequency voltage is raised to 20 volts while the audio frequency is dropped to 200 Hz, determine the frequency deviation and modulation index.

Solution: Given that

$$\Delta f = 64 \text{ kHz}$$
$$f_m = 500 \text{ Hz}$$
$$V_m = 3.2 \text{ volt}$$

We know that,

$$\text{Frequency sensitivity, } k_f = \frac{\Delta f_1}{V_{m1}} = \frac{6.4 \text{ kHz}}{3.2 \text{ volt}} = 2 \text{ kHz/volt}$$

Frequency deviation for $V_m = 8.4$ volt will be determined as

$$\Delta f_2 = k_f V_{m2} = 2 \times 8.4 = 16.8 \text{ kHz.}$$

Frequency deviation for $V_m = 20$ volt will be expressed as

$$\Delta f_3 = k_f V_{m3} = 2 \times 20 = 40 \text{ kHz}$$

It may be observed that the change in modulating frequency made no difference to the frequency deviation because it is independent of the modulating frequency.

The modulation indices can be calculated as under:

$$m_1 = \frac{\Delta f_1}{f_{m1}} = \frac{6.4 \text{ kHz}}{0.5 \text{ kHz}} = 12.8$$

$$m_2 = \frac{\Delta f_2}{f_{m2}} = \frac{16.8 \text{ kHz}}{0.5 \text{ kHz}} = 33.6$$

$$m_3 = \frac{\Delta f_3}{f_{m3}} = \frac{40 \text{ kHz}}{0.2 \text{ kHz}} = 200 \quad \textbf{Ans.}$$

Example 4.28. An FM wave is given by

$$s(t) = 20 \sin (6 \times 10^8 t + 7 \sin 1250t). \textbf{ Determine}$$

(i) The carrier and modulating frequencies, the modulation index, and the maximum deviation.

(ii) Power dissipated by this FM wave in a 100 ohm resistor.

Solution: (a) The standard expression for FM is

$$s(t) = A \sin \left[\omega_c t + m_f \sin (\omega_m t)\right] \quad ...(i)$$

Given expression is

$$s(t) = 20 \sin \left[6 \times 10^8 t + 7 \sin 1250t\right] \quad ...(ii)$$

On comparing equations (i) and (ii), we obtain

$$f_c = \frac{\omega_c}{2\pi} = \frac{6 \times 10^8}{2\pi} = 95.5 \text{ MHz}$$

$$f_m = \frac{\omega_m}{2\pi} = \frac{1250}{2\pi} = 199 \text{ Hertz}$$

and $$m_f = 7, \Delta f = m_f f_m = 7 \times 199 = 1393 \text{ Hz} \quad \textbf{Ans.}$$

(ii) Power dissipated by the given FM wave in 100 ohm resistor can be calculated as under:

$$P = \frac{\left(\frac{A}{\sqrt{2}}\right)^2}{R} = \frac{\left(\frac{20}{\sqrt{2}}\right)^2}{100} = 2 \text{ watt.} \quad \textbf{Ans.}$$

Example 4.29. Find the instantaneous frequency in hertz of each of the following signals:

(i) $10 \cos\left(200\pi t + \frac{\pi}{3}\right)$

(ii) $10 \cos(20\pi t + \pi t^2)$

(iii) $\cos 200\pi t \cos(5 \sin 2\pi t) + \sin 200\pi t \sin(5 \sin 2\pi t)$

Solution: (i) We have

$$\theta(t) = 200\pi t + \frac{\pi}{3}$$

$$\omega_i = \frac{d\theta}{dt} = 200\pi = 2\pi(100)$$

The instantaneous frequency of the signal is 100 Hz, which is constant.

(ii)
$$\theta(t) = 20\pi t + \pi t^2$$

$$\omega_i = \frac{d\theta}{dt} = 20\pi + 2\pi t = 2\pi(10 + t)$$

The instantaneous frequency of the signal is 10 Hz at $t = 0$ and increases linearly at a rate of 1 Hz/s.

(iii) $\cos 200\pi t \cos(5\sin 2\pi t) + \sin 200\pi t \sin(5 \sin 2\pi t) = \cos(200\pi t - 5\sin 2\pi t)$

$$\theta(t) = 200\pi t - 5 \sin 2\pi t$$

$$\omega_i = \frac{d\theta}{dt} = 200\pi - 10\pi \cos 2\pi t = 2\pi(100 - 5 \cos 2\pi t)$$

The instantaneous frequency of the signal will be 95 Hz at $t = 0$ and oscillate sinusoidally between 95 and 105 Hz.

Example 4.30. Given an angle-modulated signal

$$x_c(t) = 10 \cos[(10^8)\pi t + 5\sin 2\pi(10^3)t]$$

Determine the maximum phase deviation and the maximum frequency deviation.

Solution: Comparing the given $x_c(t)$ with standard FM wave equation, we have

$$\theta(t) = \omega_c t + \phi(t) = (10^8)\pi t + 5 \sin 2\pi(10^3)t$$

and $\quad \phi(t) = 5 \sin 2\pi (10^3)t$

Now, $\quad \phi'(t) = 5(2\pi)(10^3) \cos 2\pi (10^3)t$

Therefore, the maximum phase deviation will be

$$|\phi(t)|_{max} = 5 \text{ rad}$$

and the maximum frequency deviation will be

$$\Delta\omega = |\phi'(t)|_{max} = 5(2\pi)(10^3) \text{ rad/s}$$

or $\quad \Delta f = 5 \text{ kHz}$

Example 4.31. An angle-modulated signal is described by

$$x_c(t) = 10\cos[2\pi(10^6)t + 0.1 \sin(10^3)\pi t]$$

(i) considering $x_c(t)$ as a PM signal wich $k_p = 10$, obtain $m(t)$

(ii) considering $x_c(t)$ as an FM signal with $k_f = 10\pi$, find $m(t)$,

Solution: (i) We have

$$x_{PM}(t) = A \cos [\omega_c t + k_p m(t)] = 10\cos [2\pi(10^6)t + 10m(t)]$$
$$= 10 \cos [2\pi(10^6)t + 0.1 \sin (10^3) \pi t]$$

Hence, $\quad m(t) = 0.01 \sin(10^3) \pi t \qquad$ **Ans.**

(ii) We have $\quad x_{FM}(t) = A\cos\left[\omega_c t + k_f \int_{-\infty}^{t} m(\lambda)\, d\lambda\right] = 10\cos [2\pi(10^6)t + 0.1 \sin(10^3)\pi t]$

Assuming

$$m(t) = a_m \cos (10^3) \pi t$$

we obtain

$$10\pi \int_{-\infty}^{t} m(\lambda)\, d\lambda = 10\pi a_m \int_{-\infty}^{t} \cos(10^3)\pi\lambda\, d\lambda = \frac{a_m}{100} \sin(10^3)\pi t = 0.1 \sin(10^3)\pi t$$

Hence, $\quad a_m = 10$, and

$$m(t) = 10 \cos (10^3) \pi t$$

Example 4.32. Let $m_1(t)$ and $m_2(t)$ be two message signals, and let $x_{c1}(t)$ and $x_{c2}(t)$ be the modulated signals corresponding to $m_1(t)$ and $m_2(t)$, respectively.

(i) Prove that if the modulation is DSB(AM), then $m_1(t) + m_2(t)$ will produce a modulated signal equal to $x_{c1}(t) + x_{c2}(t)$. (This is why AM is sometimes referred to as a linear Modulation.)

(ii) Prove that if the modulation is PM, then the modulated signal produced by $m_1(t) + m_2(t)$ will not be $x_{c1}(t) + x_{c2}(t)$, that is, superposition does not apply to angle-modulated signal. (This is why angle modulation is sometimes referred to as a nonlinear modulation.)

Solution: (i) For DSB (AM), we have

$$m_1(t) \rightarrow x_{c1}(t) = m_1(t) \cos \omega_c t$$
$$m_2(t) \rightarrow x_{c2}(t) = m_2(t) \cos \omega_c t$$

Now,

$$m_1(t) + m_2(t) \rightarrow x_c(t) = [m_1(t) + m_2(t)] \cos \omega_c t$$
$$= m_1(t) \cos \omega_c t + m_2(t) \cos \omega_c(t) = x_{c1}(t) + x_{c2}(t)$$

Therefore, DSB (AM) modulation is a linear modulation.

(ii) For PM, we have

$$m_1(t) \rightarrow x_{c1}(t) = A \cos [\omega_c t + k_p m_1(t)]$$
$$m_2(t) \rightarrow x_{c2}(t) = A \cos [\omega_c t + k_p m_2(t)]$$

Also, $\quad m_1(t) + m_2(t) \rightarrow x_c(t) = A \cos [\omega_c t + k_p[m_1(t) + m_2(t)]]$

$$\neq x_{c1}(t) + x_{c2}(t)$$

Therefore, PM is a nonlinear modulation.

Example 4.33. A carrier signal is angle-modulated by the sum of two sinusoids as under:

$$x_c(t) = A \cos (\omega_c t + \beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t) \quad ...(i)$$

where ω_1 and ω_2 are not harmonically related. Determine the spectrum of $x_c(t)$.

Solution: The given signal $x_c(t)$ can be expressed as under:

$$x_c(t) = A \operatorname{Re}[e^{j\omega_c t} e^{j(\beta_1 \sin\omega_1 t + \beta_2 \sin\omega_2 t)}]$$
$$= A \operatorname{Re}(e^{j\omega_c t} e^{j\beta_1 \sin\omega_1 t} e^{j\beta_2 \sin\omega_2 t}) \quad ...(ii)$$

We have

$$e^{j\beta_1 \sin\omega_1 t} = \sum_{n=-\infty}^{\infty} J_n(\beta_1)\, e^{jn\omega_1 t}$$

$$e^{j\beta_2 \sin\omega_2 t} = \sum_{m=-\infty}^{\infty} J_m(\beta_1)\, e^{jm\omega_2 t}$$

Substituting these expressions into equation (ii) and taking the real part, we get

$$x_c(t) = A \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} J_n(\beta_1) J_m(\beta_2) \cos(\omega_c + n\omega_1 + m\omega_2)t \quad ...(iii)$$

Form equation (iii) it maybe observed that the spectrum of $x_c(t)$ consists of four categories:

(1) the carrier line

(2) sideband lines at $\omega_c + n\omega_1$ due to one tone alone

(3) sideband lines at $\omega_c \pm m\omega_2$ due to the other tone alone

(4) sideband lines at $\omega_c \pm n\omega_1 \pm m\omega_2$ because of the and nonlinear property of angle modulation

Example 4.34. In a tone-modulated angle modulation, the modulated signal $x_c(t)$ is given by

$$x_c(t) = A\cos(\omega_c t + \beta \sin \omega_m t)$$

when $\beta << 1$, we have NB angle modulation.

(i) Determine the spectrum of this NB angle-modulated signal.

(ii) Compare the result with that of a tone-modulated AM signal.

(iii) Discuss the similarities and difference by drawing their phasor representations.

Solution: (i) Given $x_c(t) = A\cos(\omega_c t + \beta \sin \omega_m t)$

$$= A\cos\omega_c t \cos(\beta \sin \omega_m t) - A \sin \omega_c t \sin(\beta \sin\omega_m t)$$

Now, when $\beta<<1$, we can write

$$\cos(\beta \sin \omega_m t) \cong 1$$

$$\sin(\beta \sin\omega_m t) \cong \beta \sin \omega_m t$$

Then the NB signal may be approximated as order:

$$x_{NBc}(t) \cong A\cos\omega_c t - \beta A \sin \omega_m t \sin \omega_c t$$

$$= A\cos\omega_c t - \frac{\beta A}{2}\cos(\omega_c - \omega_m)t + \frac{\beta A}{2}\cos(\omega_c + \omega_m)t \quad \text{...(i)}$$

From equation (i) it may be observed that the spectrum of $x_{NBc}(t)$ consists of carrier line and a pair of side lines at $\omega_c \pm \omega_m$.

(ii) The above result is almost identical to the situation for a tone-modulated AM signal

$$x_{AM}(t) = A\cos\omega_c t + mA \cos\omega_m t \cos\omega_c t$$

$$= A\cos\omega_c t + \frac{mA}{2}\cos(\omega_c - \omega_m)t + \frac{mA}{2}\cos(\omega_c + \omega_m)t \quad \text{...(ii)}$$

where m is the modulation index for AM.

Now, comparisons of equations (i) and (ii) reveals that the main difference between NB angle modulation and AM is the phase reversal of the lower sideband component.

(iii) By using

$$e^{j\beta \sin\omega_m t} \cong 1 + j\beta \sin \omega_m t \text{ for } \beta <<1$$

Equation (i) can be written in phasor form as under:

$$x_{NBc}(t) = \text{Re}[A e^{j\omega_c t}(1 + j\beta \sin\omega_m t)]$$

$$= \text{Re}\left[Ae^{j\omega_c t}\left(1 + \frac{\beta}{2}e^{j\omega_m t} - \frac{\beta}{2}e^{-j\omega_m t}\right)\right] \quad \text{...(iii)}$$

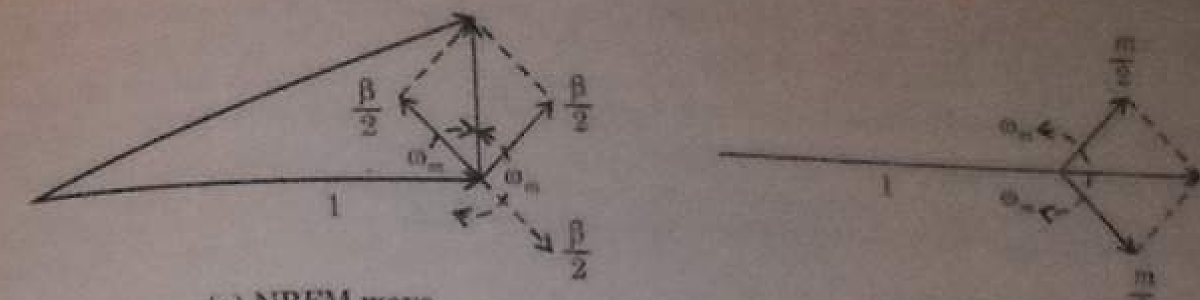
Similarly, equation (ii) can be written in phasor form as under:

$$x_{AM}(t) = \text{Re}[A e^{j\omega_c t}(1 + m\cos\omega_m t)]$$

or

$$x_{AM}(t) = \text{Re}\left[A e^{j\omega_c t}\left(1 + \frac{m}{2}e^{j\omega_m t} + \frac{m}{2}e^{-j\omega_m t}\right)\right] \quad \text{...(iv)}$$

By taking the term $Ae^{j\omega_c t}$ as the reference, the phasor representations of equations (iii) and (iv) have been shown in figure 4.33. From the figure the difference between equations (iii) and (iv) is quite obvious. In narrowband angle modulation, the modulation is added in quadrature with the carrier, which results in phase fluctuation with little amplitude change. In the AM case, the modulation is added in phase with the carrier, producing amplitude fluctuation with no phase deviation.



(a) NBFM wave (b) AM wave

Fig. 4.33. *Phasor Representation.*

Example 4.35. Given the angle-modulated signal

$$x_c(t) = 10\cos(2\pi 10^8 t + 200\cos 2\pi 10^3 t)$$

what will be bandwidth of this angle modulated signal?

Solution: The instantaneous frequency is given by

$$\omega_i = 2\pi(10^8) - 4\pi(10^5)\sin 2\pi(10^3)t$$

So $\Delta\omega = 4\pi(10^5)$, $\omega_m = 2\pi(10^3)$, and

$$\beta = \frac{\Delta\omega}{\omega_m} = \frac{4\pi(10^5)}{2\pi(10^3)} = 200$$

Also, we have

Bandwidth $W_B = 2(\beta + 1)\omega_m = 8.04\pi(10^5)$ rad/s

Since $\beta >> 1$, therefore

$$W_B \cong 2\Delta\omega = 8\pi(10^5) \text{ rad/s}$$

or $f_B = 400$ kHz **Ans.**

Example 4.36. A 20 megahertz (MHz) carrier is frequency-modulated by a sinusoidal signal such that the maximum frequency deviation is 100 kHz. Determine the modulation index and the approximate bandwidth of the FM signal if the frequency of the modulating signal is

(i) 2 kHz (ii) 100 kHz, and (iii) 500 kHz.

Given $\Delta f = 100$ kHz, $f_c = 20$ MHz $>> f_m$

Solution: For sinusoidal modulation, $\beta = \Delta f/f_m$.

(i) With $f_m = 1$ kHz, $\beta = 100$. This is a WBFM signal, and $f_B \cong 2\Delta f = 200$ kHz.

(ii) With $f_m = 100$ kHz, $\beta = 1$. Thus, we have

$$f_B \cong 2(\beta + 1)f_m = 400 \text{ kHz}$$

(iii) With $f_m = 500$ kHz, $\beta = 0.2$. This is an NBFM signal, and $f_B \cong 2f_m = 1000$ kHz $= 1$ MHz.

Example 4.37. Given an angle-modulated signal

$$x_c(t) = 10\cos(\omega_c t + 3\sin\omega_m t)$$

Assume PM and $f_m = 1$ kHz. Calculate the modulation index and find the bandwidth when (i) f_m is double and (ii) f_m is decreased by one-half.

Solution: Given $x_{PM}(t) = A\cos[\omega_c t + k_p m(t)] = 10\cos(\omega_c t + 3\sin\omega_m t)$

Thus, $m(t) = a_m \sin\omega_m t$, and

$$x_{PM}(t) = 10\cos(\omega_c t + k_p a_m \sin\omega_m t)$$

We have $\beta = k_p a_m = 3$

It may be observed that the value of β is independent of f_m. Now, when $f_m = 1$ kHz, we have

$$f_B = 2(\beta + 1)f_m = 8 \text{ kHz}$$

(i) When f_m is doubled, $\beta = 3$, $f_m = 2$kHz, and

$$f_B = 2(3+1)2 = 16 \text{ kHz}$$

(ii) When f_m is decreased by one-half, $\beta = 3$, $f_m = 0.5$ kHz, and

$$f_B = 2(3+1)(0.5) = 4 \text{ kHz}$$

Example 4.38. Repeat problem (4.40) when FM is assumed *i.e.*,

$$x_{FM}(t) = A\cos\left[\omega_c t + k_f \int_{-\infty}^{t} m(\lambda)\,d\lambda\right] = 10\cos(\omega_c t + 3\sin\omega_m t)$$

Solution: Thus $\quad m(t) = a_m \cos\omega_m t$ and

$$x_{FM}(t) = 10\cos\left(\omega_c t + \frac{a_m k_f}{\omega_m}\sin\omega_m t\right)$$

We have $\quad \beta = \dfrac{a_m k_f}{\omega_m} = \dfrac{a_m k_f}{2\pi f_m} = \dfrac{a_m k_f}{2\pi(10^3)} = 3$

It may be observed that the value of β is inversely proportional to f_m. Thus, when $f_m = 1$ kHz, we have

$$f_B = 2(\beta+1)f_m = 2(3+1)1 = 8 \text{ kHz}$$

(i) When f_m is doubled, $\beta = 3/2$, $f_m = 2$kHz, and

$$f_B = 2(\beta+1)f_m = 2\left(\frac{3}{2}+1\right)2 = 10 \text{ kHz}$$

(ii) When f_m is decreased by one-half, $\beta = 6$, $f_m = 0.5$ kHz, and

$$f_B = 2(\beta+1)f_m = 2(6+1)(0.5) = 7 \text{ kHz} \qquad \textbf{Ans.}$$

Example 4.39. A carrier signal is frequency-modulated with a sinusoidal signal of 2 kHz resulting in a maximum frequency deviation of 5 kHz.

(i) Find the bandwidth of the modulated signal.

(ii) The amplitude of the modulating sinusoid signal is increased by a factor of 3, and its frequency is lowered to 1 kHz. Determine the maximum frequency deviation and the bandwidth of the new modulated signal.

Solution: (i) We have

$$\beta = \frac{k_f a_m}{\omega_m} = \frac{\Delta f}{f_m} = \frac{5(10^3)}{2(10^3)} = 2.5$$

Then, the bandwidth will be

$$f_B = 2(\beta+1)f_m = 2(2.5+1)2 = 14 \text{ kHz}$$

(ii) Let β_1 be the new modulation index. Then, we have

$$\beta_1 = \frac{k_f 3a_m}{\frac{1}{2}\omega_m} = 6\frac{k_f a_m}{\omega_m} = 6\beta = 6(2.5) = 15$$

Therefore, we have

$$\Delta f = \beta_1 f_m t = (15)(1) = 15 \text{ kHz}$$

$$f_B = 2(\beta_1+1)f_{m1} = 2(15+1)(1) = 32 \text{ kHz}$$

Example 4.40. As a matter of fact, in addition to Carson's rule, the following formula is often used to estimate bandwidth of an FM signal:

$$W_B \equiv 2(D+2)\omega_m \qquad \text{for } D > 2$$

where $\omega_M = 2\pi f_M$ and f_M is the highest frequency of the signal in hertz. Calculate the bandwidth, using formula, and compare it to the bandwidth, using Carson's rule for the FM signal with $\Delta f = 75$ kHz and $f_M = 15$ kHz.

Note that commercial FM broadcast stations in the United States are limited to maximum frequency deviation of 75 kHz, and modulating frequencies typically cover 50 Hz to 15 kHz.

Solution: We have $\omega_m = 2\pi f_M$, where $f_M = 15$ kHz, we get

$$D = \frac{\Delta f}{f_M} = \frac{75(10^3)}{15(10^3)} = 5$$

and by using the given formula, the bandwidth will be

$$f_B = 2(D+2)f_M = 210 \text{ kHz}$$

Using Carson's rule, observe that the bandwidth will be

$$f_B = 2(D+1)f_M = 180 \text{ kHz}$$

Note: *High-quality FM radio require bandwidth of at least 200 kHz. Hence, it seems that Carson's rule underestimates the bandwidth.*

Example 4.41. Given a frequency multiplier circuit and an NBFM signal

$$x_{NBFM}(t) = A\cos(\omega_c t + \beta\sin\omega_m t)$$

With $\beta < 0.5$ and $f_c = 200$ kHz. Let f_m range from 50 Hz to 15 kHz, and let the maximum frequency deviation Δf at the output be 75 kHz. Determine the required frequency multiplication n and the maximum allowed frequency deviation at the input.

Solution: From the expression

$$\beta = \Delta f / f_m, \text{ we have}$$

$$\beta_{min} = \frac{75(10^3)}{15(10^3)} = 5$$

and $\quad \beta_{max} = \dfrac{75(10^3)}{50} = 1500$

If $\beta_1 = 0.5$, where β_1 is the input β, then the required frequency multiplication will be

$$n = \frac{\beta_{max}}{\beta_1} = \frac{1500}{0.5} = 3000$$

The maximum allowed frequency deviation at the input, denoted Δ_{f1}, will be

$$\Delta f_1 = \frac{\Delta f}{n} = \frac{75(10^3)}{3000} = 25 \text{ Hz} \qquad \textbf{Ans.}$$

Example 4.42. A block diagram of an indirect (Armstrong) FM transmitter has been shown in figure 4.34. Calculate the maximum frequency deviation Δf of the output of the FM transmitter and the carrier frequency f_c if $f_1 = 200$ kHz, $f_{LO} = 10.8$ MHz, $\Delta f_1 = 25$ Hz, $n_1 = 64$, and $n_2 = 48$.

Solution: We have $\quad \Delta f = (\Delta f_1)(n_1)(n_2) = (25)(64)(48)\text{ Hz} = 76.8 \text{ kHz}$

$$f_2 = n_1 f_1 = (64)(200)(10^3) = 12.8(10^6)\text{ Hz} = 12.8 \text{ MHz}$$

$$f_3 = f_2 \pm f_{LO} = (12.8 \pm 10.8)(10^6)\text{ Hz} = \begin{cases} 23.6 \text{ MHz} \\ 2.0 \text{ MHz} \end{cases}$$

Thus, when $f_3 = 23.6$ MHz, then, we have

$$f_c = n_2 f_3 = (48)(23.6) = 1132.8 \text{ MHz}$$

When $\quad f_3 = 2$ MHz, then, we have

$$f_c = n_2 f_3 = (48)(2) = 96 \text{ MHz}$$

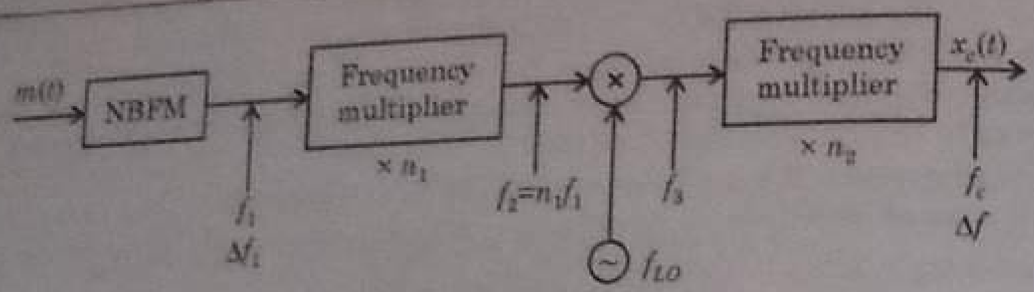


Fig. 4.34. *Block diagram of an indirect FM transmitter.*

Example 4.43. In an Armstrong-type FM generator of figure 4.32, the crystal oscillator frequency is 200 kHz. The maximum phase deviation is limited to 0.2 to avoid distortion. Let f_m range from 50 Hz to 15 kHz. The carrier frequency at the output is 108 MHz, and the maximum frequency deviation is 75 kHz. Select multiplier and mixer oscillator frequencies.

Solution: We have

$$\Delta f_1 = \beta f_m\ (0.2)(50) = 10\ \text{Hz}$$

and
$$\frac{\Delta f}{\Delta f_1} = \frac{75(10^3)}{10} = 7500 = n_1 n_2$$

$$f_2 = n_1 f_1 = n_1 (2)(10^5)\ \text{Hz}$$

Assuming down conversion, we have

$$f_2 - f_{LO} = \frac{f_c}{n_2}$$

Thus,
$$f_{LO} = n_1 f_1 - \frac{f_c}{n_2} = \frac{7500(2)(10^5) - 108(10^6)}{n_2} = \frac{1392}{n_2}(10^6)\ \text{Hz}$$

Letting $n_2 = 150$, we get

$$n_1 = 50 \text{ and } f_{LO} = 9.28\ \text{MHz} \qquad \textbf{Ans.}$$

Example 4.44. A given modulated signal has maximum frequency deviation of 50 Hz for an input sinusoid of unit amplitude and a frequency of 120 Hz. Find the required frequency multiplication factor n to produce a maximum frequency deviation of 20 kHz when the input sinusoid has unit amplitude and a frequency of 240 Hz and the angle modulation used is (*i*) PM and (*ii*) FM.

Solution: (*i*) It may be observed that in sinusoidal PM, the maximum frequency deviation Δf is proportional to f_m. Thus, we have

$$\Delta f_1 = \left(\frac{240}{120}\right)(50) = 100\ \text{Hz}$$

Therefore
$$n = \frac{\Delta f_2}{\Delta f_1} = \frac{20(10^3)}{100} = 200$$

(*ii*) Again we observe that in sinusoidal FM, the maximum frequency deviation Δf is independent of f_m. Thus, we have

$$n = \frac{\Delta f_2}{\Delta f_1} = \frac{20(10^3)}{50} = 400 \qquad \textbf{Ans.}$$

Example 4.45. At low carrier frequencies, it may be possible to generate an FM signal by varying the capacitance of a parallel resonant circuit. Prove that the output $x_c(t)$ of the tuned circuit shown in figure 4.35 is an FM signal if the capacitance has a time dependence of the form

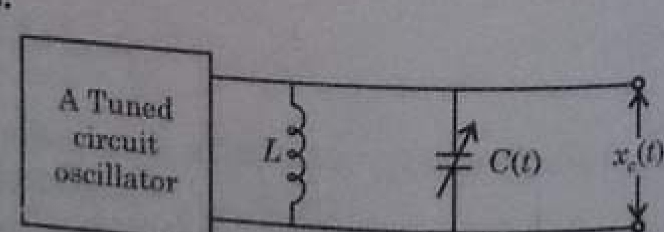


Fig. 4.35.

$$C(t) = C_0 - km(t)$$

and
$$\left|\frac{k}{C_0} m(t)\right| << 1$$

Solution: If we assume $km(t)$ is small and slowly varying, then the output frequency ω_i of the oscillator will be given by

$$\omega_i = \frac{1}{\sqrt{LC(t)}} = \frac{1}{\sqrt{L[C_0 - km(t)]}} = \frac{1}{\sqrt{LC_0}}\left[1 - \frac{k}{C_0}m(t)\right]^{-1/2}$$

Because $|(k/C_0)m(t)| << 1$, then, we can use the following approximation:

$$(1 - z)^{-1/2} \cong 1 + \frac{1}{2}z$$

and obtain
$$\omega_i \cong \omega_c\left[1 + \frac{1}{2}\frac{k}{C_0}m(t)\right] = \omega_c + k_f m(t)$$

where
$$\omega_c = \frac{1}{\sqrt{LC_0}} \quad \text{and } k_f = \frac{1}{2}\frac{\omega_c k}{C_0}$$

Thus, $x_c(t)$ is an FM signal **Hence Proved.**

Example 4.46. An FM signal

$$x_{FM}(t) = A\cos\left[\omega_c t + k_f \int_{-\infty}^{t} m(\lambda)\, d\lambda\right]$$

is applied to a system consisting of a high-pass RC filter and an envelope detector. Assu[me] that $\omega RC << 1$ in the frequency band occupied by $x_{FM}(t)$. Determine the output signal y[…] assuming that $k_f|m(t)| < \omega_c$ for all t.

Solution: Taking the Fourier transform of both sides of equation (*i*) yields

$$Y(\omega) = \frac{1}{\tau}[X(\omega) - e^{-j\omega\tau}X(\omega)] = \frac{1}{\tau}X(\omega)(1 - e^{-j\omega\tau})$$

If $\omega\tau << 1$, then $1 - e^{-j\omega\tau} \cong j\omega\tau$ and

$$Y(\omega) \cong j\omega X(\omega)$$

which indicates that $y(t)$ is approximately equal to the derivative of $x(t)$ and τ must satisfy following condition:

$$\tau << \frac{1}{\omega} = \frac{1}{\omega_c + \Delta\omega} \qquad \textbf{Ans.}$$

Example 4.47. The sinusoidal modulating wave

$$m(t) = A_m \cos(2\pi f_m t)$$

is applied to a phase modulator with phase sensitivity k_p. The unmodulated carrier w[ave] has frequency f_c and amplitude A_c.

(*i*) Determine the spectrum of the resulting phase-modulated signal, assuming that maximum phasor deviation $\beta_p = k_p A_m$ does not exceed 0.3 radians.

(*ii*) Construct a phase diagram for this modulated signal, and compare it with that of corresponding narrowband FM signal. *(U.P.S.C., I.E.S., Examination, […]*

Solution: (*i*) A phase-modulated wave is expressed as,

$$s(t) = A_c \cos[\omega_c t + k_p m(t)]$$

where, $m(t) = A_m \cos(2\pi f_m t)$

then, $s(t) = A_c \cos[2\pi f_c t + k_p A_m \cos(2\pi f_m t)]$

$= A_c \cos[2\pi f_c t + \beta_p \cos(2\pi f_m t)]$

where $\beta_p = k_p A_m$ known as modulation index for PM.

$= A_c \cos(2\pi f_c t).\cos[\beta_p \cos(2\pi f_m t) - A_c \sin(2\pi f_c t) + \sin[\beta_p \cos(2\pi f_m t)]$

Now, if $\beta_p \leq 0.5$, then, we have

$\cos[\beta_p \cos(2\pi f_m t)] \cong 1$

$\sin[\beta_p \cos(2\pi f_m t)] \cong \beta_p \cos(2\pi f_m t)$

Then, we can write equation (i) as,

$s(t) \cong A_c \cos(2\pi f_c t) - \beta_p A_c \sin(2\pi f_c t)\cos(2\pi f_m t)$

$$= A_c \cos(2\pi f_c t) - \frac{1}{2}\beta_p A_c \sin[2\pi(f_c + f_m)t] - \frac{1}{2}\beta_p A_c \sin[2\pi(f_c - f_m)t] \quad ...(ii)$$

Therefore, spectrum of $s(t)$, will be

$$\cos(2\pi f_c t) \Leftrightarrow \frac{1}{2}[\delta(f - f_c) + \delta(f + f_c)]$$

and $$\sin 2\pi f_c t \Leftrightarrow \frac{1}{2j}[\delta(f - f_c) + \delta(f + f_c)]$$

Using Fourier transform pairs, we have

$$S(f) \cong \frac{1}{2}A_c[\delta(f - f_c) + \delta(f + f_c)] - \frac{1}{4j}\beta_p A_c[\delta(f - f_c - f_m) - \delta(f + f_c + f_m)]$$

$$-\frac{1}{4}\beta_p A_c[\delta(f - f_c + f_m) - \delta(f + f_c - f_m)]$$

(ii) The phasor diagram for $s(t)$ is deduced from equation (ii), as follows:

Comparing these two phasor diagrams in figure 4.35 and figure 4.36, we see that except for a phase difference the narrowband PM and FM waves have the same form.

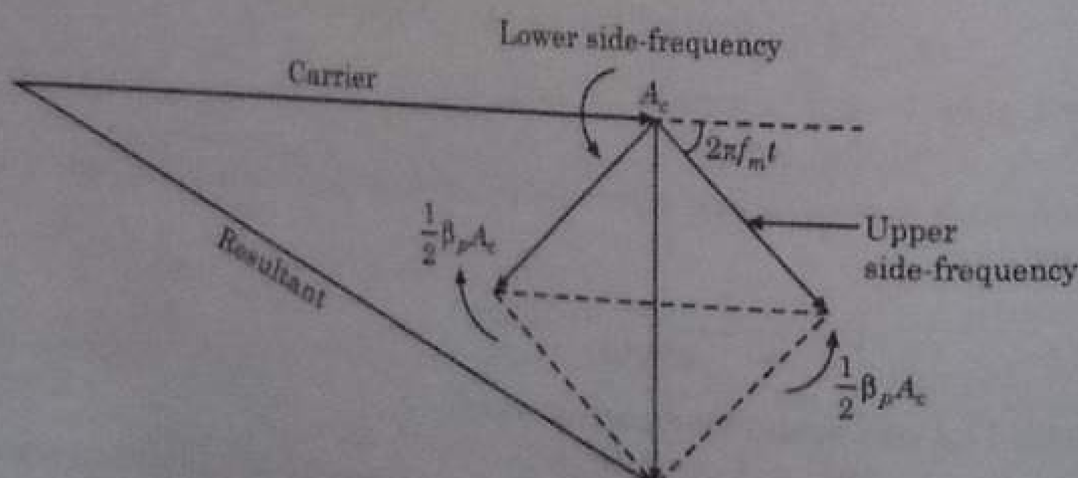


Fig. 4.36.

Phasor diagram for narrow FM has been shown in figure 4.37.

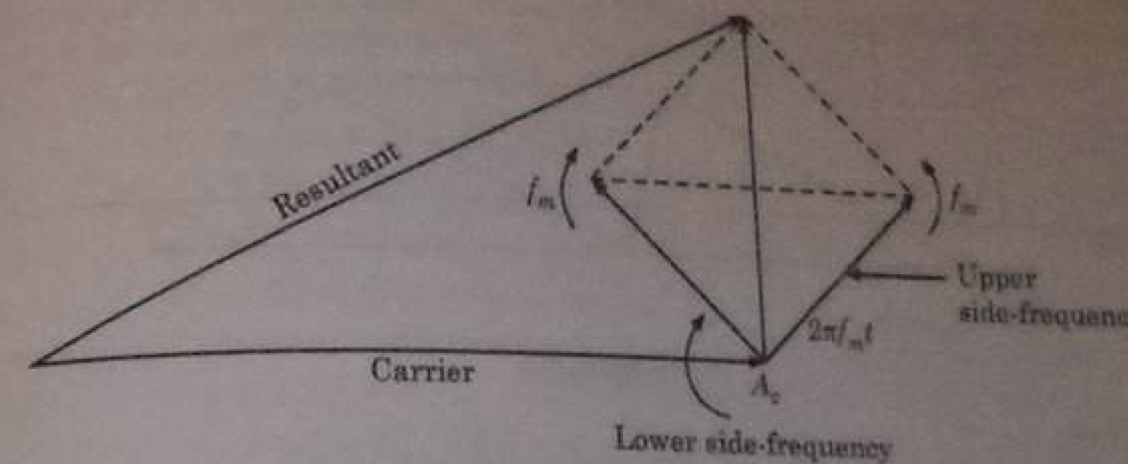


Fig. 4.37.

Example 4.48. An angle-modulated signal is given by

$$x_c(t) = 5\cos[2\pi(10^6)t + 0.2\pi t]$$

Can you identity whether $x_c(t)$ is a PM or and FM signal?

Solution: For angle modulation, the modulated carrier is represented by,

$$x_c(t) = A\cos[\omega_c t + \phi(f)]$$

In case of PM, $\phi(t) = k_p m(y)$,

where $m(t)$ is the message signal and k_p is phase duration constant.

and for FM, $$\frac{d\phi(t)}{dt} = k_f m(t)$$

where k_f is frequency deviation constant.

Given $x_c(t) = 5\cos[2\pi(10^6 t) + 0.2\cos 200\pi t)$

Now, if $m(t) = a_m \cos(200\pi t)$

then, $\phi(t) = K_p a_m \cos(200\pi t) = 0.2\cos(200\pi t)$

and then $x_c(t)$ = PM signal.

But, if $m(t) = a_m \sin \omega_m t$

then $$\phi(t) = k_f \int_{t_0}^{t} m(\lambda)\,d\lambda + \phi(t_0) = k_f \int_{-\infty}^{t} m(\lambda)\,d\lambda + \phi(-\infty)$$

or $$\phi(t) = k_f \int_{-\infty}^{t} m(\lambda)\,d\lambda \;[\text{At } t_0 = -\infty, \phi(t_0) =]$$

or $$\phi(t) = \frac{k_f a_m}{\omega_m}\cos\omega_m t$$

and then $x_c(t)$ is a FM signal. Hence it can be either PM or an FM signal.

Example 4.49. A baseband signal $m(t)$ modulates a carrier to produce the following angle modulated signal.

$A_c \cos[2\pi \times 10^8 t + k_p m(t)]$, where $m(t)$ is shown in the figure 4.38.

(GATE Examination-1999)

Determine the value of k_p so that the peak-to-peak frequency deviation of the carrier is 100 kHz.

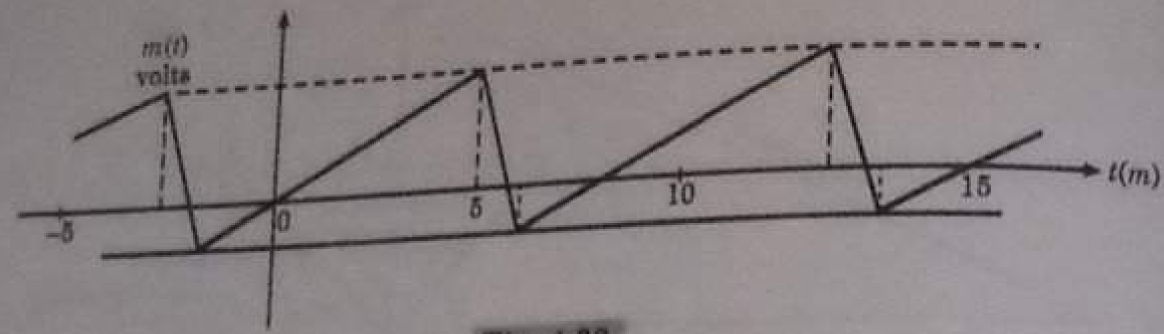


Fig. 4.38.

Solution: Given

$x(t) = A_c \cos[2\pi \times 10^8 t + k_p m(t)]$ as an angle modulated signal.

Since, phase deviation $\phi(t) = k_p m(t)$ *i.e.*, proportional to message signal, therefore, it is an phase-modulated signal.

Now, instantaneous frequency is given by

$$\omega_i = \omega_c + k_p \frac{dm(t)}{dt}$$

Then, maximum frequency deviation will be

$$|\omega_i - \omega_c|_{max} = k_p \left|\frac{dm(t)}{dt}\right|_{max}$$

or
$$100 \times 2\pi = k_p \left|\frac{dm(t)}{dt}\right|_{max} \quad ...(i)$$

from the given figure, we have

$$\left|\frac{dm(t)}{dt}\right|_{max} = \frac{17.5}{7} = 2.5$$

then, from equation (*i*), we get

$$k_p = \frac{100 \times 2\pi}{2.5} = 80\,\pi \quad \textbf{Ans.}$$

Example 4.50. An angle modulated signal with carrier frequency $\omega_c = 2 \times 10^6$ rad/sec is given by

$$s(t) = \cos 2\pi(2 \times 10^6 t + 30 \sin 150t + 40 \cos 150t)$$

then,

(*i*) Determine the maximum frequency deviation

(*ii*) Find maximum phase deviation, and

(*iii*) Find the bandwidth of $s(t)$.

Solution: (*i*) We know that instantaneous frequency is given by

$$\omega_i = \frac{d}{dt}\theta(t) = \frac{d}{dt}[2\pi(2 \times 10^6 t + 30 \sin 150t + 40 \cos 150t)]$$

$$= 2\pi \times 10^6 + 60\pi \times 150 \,.\, \cos 150t - 80\pi \times 150 \,.\, \sin 150t$$

or
$$\Delta\omega = \omega_i - \omega_c = 60\pi \times 150.\ \cos 150t - 80\pi \times 150.\ \sin 150t$$

$$= 3000\pi\,[3 \cos 150t - 4 \sin 150t]$$

$$\Delta\omega = 15000\pi \,.\, \cos(150t + \alpha),$$

Example 4.51. An angle-modulated signal is given as below:

$$s(t) = 20 \cos(\omega_c t + 4 \sin \omega_m t)$$

Assuming as a PM signal and $f_m = 2$ kHz, calculate the modulation index and bandwidth when (*i*) f_m is increased two times (*ii*) f_m is decreased by half.

Solution: A PM signal is given as,

$$s(t) = A_m \cos[\omega_c t + k_p m(t)]$$

or
$$s(t) = 20 \cos[\omega_c t + 4 \sin \omega_m t]$$

Hence,
$$m(t) = a_m \sin \omega_m t \quad ...(i)$$

then,
$$s(t) = 20 \cos[\omega_c t + k_p a_m \sin \omega_m t) \quad ...(ii)$$

Comparing equations (*i*) and (*ii*), we get

$$\beta = k_p a_m = 4$$

Also,
$$f_B = 2(\beta + 1)f_m = 2(4 + 1).2 = 20 \text{ kHz}$$

(*i*) When,
$$f_{m1} = 2f_m = 4\text{kHz}$$

$$f_B = 2(4 + 1) \cdot 4 = 40 \text{ kHz} \quad \textbf{Ans.}$$

(*ii*) When,
$$f_{m2} = \frac{f_m}{2} = 1 \text{ kHz}$$

$$f_B = 2(\beta + 1) \cdot f_{m2} = 2(4 + 1) \,.\, 1 = 10 \text{ kHz} \quad \textbf{Ans.}$$

SUMMARY

1. Angle modulation may be defined as the process in which the total phase angle of a carrier wave is varied in accordance with the instantaneous value of the modulating or message signal while keeping the amplitude of the carrier constant.
2. We can vary this phase angle ϕ in two ways and thus there are two types of angle modulation as under:
   (*i*) Phase Modulation (PM)
   (*ii*) Frequency Modulation (FM).
3. Phase modulation (PM) is that type of angle modulation in which the phase angle ϕ is varied linearl with a baseband or modulating signal $x(t)$ about an unmodulated phase angle $(\omega_c t + \theta_0)$.
4. This means that in Phase Modulation, the instantaneous value of the phase angle is equal to th phase angle of the unmodulated carrier $(\omega_c t + \theta_0)$ plus a time-varying component which is proportion to modulating signal $x(t)$.
5. Frequency modulation is that type of angle modulation in which the instantaneous frequency ω_i varied linearly with a message or baseband signal $x(t)$ about an unmodulated carrier frequency ω. This means that the instantaneous value of the angular frequency ω_i will be equal to the carri frequency ω_c plus a time-varying component proportional to the baseband signal $x(t)$.
6. $s(t) = A \cos\left[\omega_c t + k_f \int_0^t x(t)\,dt\right]$ which is the required general expression for FM wave.
7. The maximum change in instantaneous frequency form the average frequency ω_c is called frequen deviation.
8. This maximum change in instantaneous frequency ω_i from the average or carrier frequency ω_c depen upon the magnitude and sign of $k_f \cdot x(t)$.
9. Therefore,
   Frequency deviation $\Delta\omega = |k_f \cdot x(t)|_{max}$
10. To get FM by using PM, we first integrate the baseband signal and then apply to the phase modulat

11. The total variation in frequency from the lowest to the highest point is called carrier swing. Obviously

The carrier swing = 2 × frequency deviation

= 2 × $\Delta\omega$

12. The amount of frequency deviation or variation depends upon the amplitude (loudness) of the modulating (audio) signal. This means that louder the sound, greater the frequency deviation and vice versa.

13. The frequency deviation is useful in determining the FM signal bandwidth. Since, a maximum frequency deviation of 75 *KHz* is allowed for commercial FM broadcast stations using a band of 88 MHz to 108 MHz, therefore approximate FM channel width is 2 × 75 = 150 kHz. Allowing a 25 kHz guardband on either side, the channel width becomes 2(75 + 25) = 200 kHz. This guardband is meant to prevent interference between adjacent channels.

14. In FM broadcast, the highest audio frequency transmitted is 15 *KHz*.

15. For FM, the modulation index is defined as the ratio of frequency deviation to the modulating frequency.

16. Modulation index, $m_f = \dfrac{\text{Frequency deviation}}{\text{Modulation frequency}}$

or $m_f = \dfrac{\Delta\omega}{\omega_m}$

This modulation index may be greater than unity.

19. The term "per cent modulation" as it is used in reference to FM refers to the ratio of actual frequency deviation to the maximum allowable frequency deviation. Thus 100% modulation corresponds to 75 kHz for the commercial FM broadcast band and 25 kHz for the commercial FM broadcast band and 25 kHz for television.

Per cent modulation $M = \dfrac{\Delta f_{actual}}{\Delta f_{max}} \times 100$.

20. When the value of modulation index m_f is quite large, then in FM a large number of sidebands are produced and hence the bandwidth of FM is sufficiently large. This type of FM system is known as wideband FM.

21. Since the amplitude of FM signal remains unchanged, the power of the FM signal will be same as that of the unmodulated carrier.

22. There are a number of international regulations for frequency modulation prescribed by CCIR (*i.e.* Consultative Committee for International Radio). These regulations must be followed by all the commercial FM broadcast stations to avoid interference problem. They are as follows:
    (*i*) Maximum modulating frequency is 15 kHz
    (*ii*) Maximum frequency deviation is ± 75 kHz
    (*iii*) Frequency stability of the carrier is ± 2 kHz
    (*iv*) Allowable bandwidth per channel = 200 kHz.

23. The modulation index m_f is given as

$$m_f = \frac{\text{Frequency deviation}}{\text{Modulating frequency}} = \frac{\Delta w}{w_m} = \frac{k_f . V_m}{w_m}.$$

24. Basically the effective bandwidth is the separation between the two extreme significant side frequencies on either side of the carrier.

25. Fourier analysis reveals that the number of side frequencies which contain a significant amount of power and thus the effective bandwidth of the FM signal is dependent on the modulation index of the modulated wave.

26. For practical purpose, a side frequency is considered to be significant if its amplitude is at least one per cent of the unmodulated carrier amplitude.

27. Carson's rule provides a thumb formula to calculate the bandwidth of a single-tone wideband FM. According to this rule, the FM bandwidth is given as twice the sum of the frequency deviation and the highest modulating frequency.

28. The transmission bandwidth for PM using Carson's rule as

$$BW_{PM} \cong 2(\Delta w)$$
$$\cong 2k_p . V_m . w_m$$

29. For PM, the modulation index m_p will be same as the deviation θ_d is expressed as

$$m_p = k_p . V_m = \theta_d$$

30. The envelope of FM wave or PM wave is constant and is equal to the unmodulated carrier amplitude. On the other hand, the envelope of AM wave is dependent on the modulating signal $x(t)$.

31. The zero crossings (*i.e.* the instants of time at which the waveform changes from negative to a positive value or vice-versa) of a FM wave or a PM wave no longer exhibit a perfect regularity in their spacing like AM wave. Thus this makes the instantaneous frequency of the angle modulated wave depend upon time.

32. FM receivers may be fitted with amplitude limiters to remove the amplitude variations caused by noise.

33. It is possible to reduce noise still further by increasing the frequency-deviation. This is a feature which AM does not have because it is not possible.

34. Standard Frequency Allocations provide a guardband between commercial FM stations. Due to this there is less adjacent-channel interference than in AM.

35. FM broadcasts operate in the upper VHF and UHF frequency ranges at which there happens to be less noise than in the MF and HF ranges occupied by AM broadcasts.

36. The amplitude of the FM wave is constant. It is thus independent of the modulation depth, whereas in AM, modulation depth governs the transmitted power. This permits the use of low-level modulation in FM transmitter and use of efficient class *C* amplifiers in all stages following the modulator.

37. A much wider channel typically 200 kHz is required in FM as against only 10 kHz in AM broadcast. This forms serious limitation of FM.

38. FM transmitting and receiving equipments particularly used for modulation and demodulation tend to be more complex and hence more costly.

39. The FM modulator circuits used for generating FM signals may be put into two categories as under
    (*i*) The direct method or parameter variation method.
    (*ii*) The indirect method or the Armstrong method.

40. In direct method or parameter variation method, the baseband or modulating signal directly modula[tes] the carrier. The carrier signal is generated with the help of an oscillator circuit. This oscillator circ[uit] uses a parallel tuned L-C circuit. Thus the frequency of oscillation of the carrier generation is gover[ned] by the expression

$$w_c = \frac{1}{\sqrt{LC}}$$

41. The varactor diode is a semiconductor diode whose junction capacitance changes with d.c. bias volta[ge]. This varactor diode is connected in shunt with the tuned circuit of the carrier oscillator.

42. In direct method of FM generation, it is not easy to get a high order stability in carrier freque[ncy]. This is due to the fact that generation of carrier signal is directly affected by the baseband or modula[ting] signal. The baseband signal directly controls the tank circuit of the carrier generator and th[us] stable oscillator circuit (*i.e.* crystal oscillator) cannot be used. This means that the carrier genera[tor] cannot be of high stability which is a necessary requirement.

43. The non-linearity of the varactor diode produces a frequency variation due to harmonics of modulating or baseband signal and therefore the FM signal is distorted. We will have to take proper cares to keep this type of distortion minimum.

44. In Armstrong method of FM generation, we can get very high frequency stability since in this case the crystal oscillator may be used as a carrier frequency generator.
45. The working principle of Armstrong method is to generate a narrowband FM (NBFM) indirectly by utilizing the phase-modulation technique and then changing this narrowband FM into a wideband FM.
46. It converts the frequency-modulated (FM) signal into a corresponding amplitude modulated (AM) signal with the help of frequency dependent circuits, *i.e.*, the circuits whose output voltage depends upon the input frequency. These circuits are generally known as **frequency discriminators.**
47. The original modulating or baseband signal is recovered from this AM signal with the help of the linear diode detector or envelope detector.
48. The principle of operation of slope detectors depends upon the slope of the frequency response characteristics of a frequency selective network. The two main FM detectors, which use detuned resonant circuits come under this category, are as under:
    (i) Single-tuned detector circuit or simple slope detector.
    (ii) Stagger-tuned detector circuit or balanced slope detector.
49. The following two circuits come under this category:
    (i) Foster-Seeley detector
    (ii) Ratio detector.

SHORT QUESTIONS WITH ANSWERS

Q.1. What do you mean by angle modulations?

Ans. Angle modulation may be defined as the process in which the total phase angle of a carrier wave is varied in accordance with the instantaneous value of the modulating or message signal while keeping the amplitude of the carrier constant.

Q.2. What are the types of angle modulation?

Ans. There are two types of angle modulation as under:
(i) Phase Modulation (PM)
(ii) Frequency Modulation (FM)

Q.3. Define phase modulation?

Ans. Phase modulation (PM) is that type of angle modulation in which the phase angle ϕ is varied linearly with a baseband or modulating signal $x(t)$ about an unmodulated phase angle $(\omega_c t + \theta_0)$. This means that in Phase Modulation, the instantaneous value of the phase angle is equal to the phase angle of the unmodulated carrier $(\omega_c t + \theta_0)$ plus a time-varying component which is proportional to modulating signal $x(t)$.

Q.4. Define frequency modulation?

Ans. Frequency modulation is that type of angle modulation in which the instantaneous frequency ω_i is varied linearly with a message or baseband signal $x(t)$ about an unmodulated carrier frequency ω_c. This means that the instantaneous value of the angular frequency ω_i will be equal to the carrier frequency ω_c plus a time-varying component proportional to the baseband signal $x(t)$.

Q.5. What is frequency deviation?

Ans. The instantaneous frequency of FM signal varies with time around the carrier frequency ω_c. This means that the instantaneous frequency of FM signal varies according to the modulating signal.

The maximum change in instantaneous frequency form the average frequency ω_c is called **frequency deviation**.

Q.6. What is carrier signal?

Ans. The total variation in frequency from the lowest to the highest point is called carrier swing. Obviously

The carrier swing = 2 × frequency deviation
= 2 × $\Delta\omega$

Q.7. Define modulation index for FM?

Ans. For *FM*, the modulation index is defined as the ratio of frequency deviation to the modulating frequency.

Mathematically,

$$\text{Modulation index, } m_f = \frac{\text{Frequency deviation}}{\text{Modulation frequency}} = \frac{\Delta\omega}{\omega_m}$$

This modulation index may be greater than unity.

Q.8. What is per cent modulation an for FM?

Ans. The term "per cent modulation" as it is used in reference to FM refers to the ratio of actual frequency deviation to the maximum allowable frequency deviation. Thus 100% modulation corresponds to 75 kHz for the commercial FM broacast band and 25 kHz for the commercial FM broadcast band and 25 kHz for television.

$$\text{Per cent modulation } M = \frac{\Delta f_{actual}}{\Delta f_{max}} \times 100$$

Q.9. What is sideband FM?

Ans. When the value of modulation index m_f is quite large, then in FM a large number of sidebands are produced and hence the bandwidth of FM is sufficiently large. This type of FM system is known as wideband FM.

Q.10. What is transmission BW for FM?

Ans. For n sidebands the bandwidth of FM wave is given by

$$BW = 2\,nw_m \text{ radians/sec.} = 2nf_m \text{ Hz.}$$

Q.11. What is Carson's rule?

Ans. Carson's rule provides a thumb formula to calculate the bandwidth of a single-tone wideband FM. According to this rule, the FM bandwidth is given as twice the sum of the frequency deviation and the highest modulating frequency. However, it must be remembered that this rule is just an approximation.

Mathematically,

$$BW = 2\,(\Delta w + w_m)$$

Q.12. What are disadvantages of FM order?

Ans. A much wider channel typically 200 kHz is required in FM as against only 10 kHz in AM broadcast. This forms serious limitation of FM.

FM transmitting and receiving equipments particularly used for modulation and demodulation tend to be more complex and hence more costly.

REVIEW QUESTIONS

1. Explain the term instantaneous frequency in case of angle-modulation.
2. Define the following terms for FM wave:
    (i) Carrier swing
    (ii) Frequency deviation
    (iii) Per cent modulation
3. Derive an expression for a single-tone frequency-modulated wave.
4. Explain the difference between narrowband FM and wideband FM.
5. Derive an expression for a single-tone narrowband frequency-modulated wave.
6. What is meant by the term 'significant sidebands' in frequency-modulated wave.
7. What are the principal merits and limitations of FM.
8. Explain the salient features of wideband FM system.

9. State the typical applications of narrowband FM system.
10. Draw the circuit diagram of varactor-diode modulator and explain its working.
11. Explain the Armstrong method for the generation of wideband FM.

NUMERICAL PROBLEMS

1. A single-tone FM signal is given by
   $v(t) = 10 \sin(16\pi \times 10^6 t + 20 \sin 2\pi \times 10^3 t)$ volts.
   Determine the modulation index, modulating frequency, frequency deviation, carrier frequency and the power of the FM signal.
2. An FM wave, modulated to a depth of 8, generates a bandwidth of 180 kHz. Find the frequency deviation.
3. A carrier signal $10 \cos 8\pi\, 10^6 t$ is angle modulated by a modulating signal $5 \cos 30\pi\, 10^3 t$.
   (i) Find the bandwidth for frequency modulation assuming $k_f = 15$ kHz per volt.
   (ii) Assuming the same bandwidth, find k_p for phase modulation.
   (iii) Find the change in the bandwidth for frequency and phase modulation if the modulating signal becomes $5 \cos 10\pi.\, 10^3 t$.
4. Determine the carrier swing the highest and the lowest frequencies attained and the modulation index of the FM signal generated by frequency modulating a 101.6 MHz carrier with an 8 kHz. Sine wave causing a frequency deviation of 40 kHz. [**Ans.** 80 kHz, 101.64 MHz, 101.56 MHz, 5)]
5. Calculate the frequency deviation and carrier swing of a frequency-modulated wave which was produced by modulating a 50.400 MHz carrier. The highest frequency reached by the FM wave is 50.406 MHz. Then calculate the lowest frequency reached by the FM wave.
6. In an FM system, the frequency-deviation is 4 kHz when the audio modulating frequency is 200 Hz and the audio modulating voltage is 4V. Compute the modulation index. Also compute the frequency deviation and modulating index if
   (i) AF voltage is increased to 8 V and modulating frequency is increased to 800 Hz.
   (ii) AF voltage is increased to 12 V and modulating frequency is decreased to 100 Hz.
7. An angle-modulated signal is given by
   $$x_c(t) = 5 \cos[2\pi(10^6)\, t + 0.2 \cos 200\pi t]$$
   Can you identify whether $x_c(t)$ is a PM or an FM signal ?
   [**Ans.** No. It can be either a PM or an FM signal]
8. The frequency multiplier is a nonlinear device followed by a bandpass filter, as shown in figure 4.39. Suppose that the nonlinear device is an ideal square-law device with input-output characteristics.
   $$e_o(t) = a e_i^2(t)$$
   Find the output $y(t)$ if the input is an FM signal given by
   $$e_i(t) = A \cos(\omega_c t + \beta \sin\omega_m t)$$

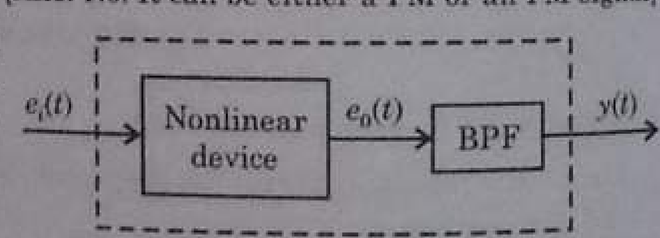


Fig. 4.39.

[**Ans.** $Y(t) = A \cos(2\omega_c t + 2\beta \sin \omega_m t)$ where $A' = \frac{1}{2} a A^2$. This result indicates that a square-law device can be used as a frequency doubler]

9. A given has maximum frequency deviation of 25 Hz for a modulating sinusoid of unit amplitude and a frequency of 100 kHz. Find the required value of frequency multiplication n to produce a maximum frequency deviation of 20 kHz when the modulating sinusoid has unit amplitude and a frequency of 200 Hz. [**Ans.** $n = 800$]
10. Assume that the 1.8 MHz signal in figure 4.40 is derived from the 200 kHz oscillator (multiplication by 4) and that the 200 kHz oscillator drift is 0.1 Hz.

   (i) Find the drift in the 10.8 MHz signal.
   (ii) Find the drift in the carrier of the resulting FM signal. [**Ans.** (a) ± 54 Hz. (b) 48 Hz]

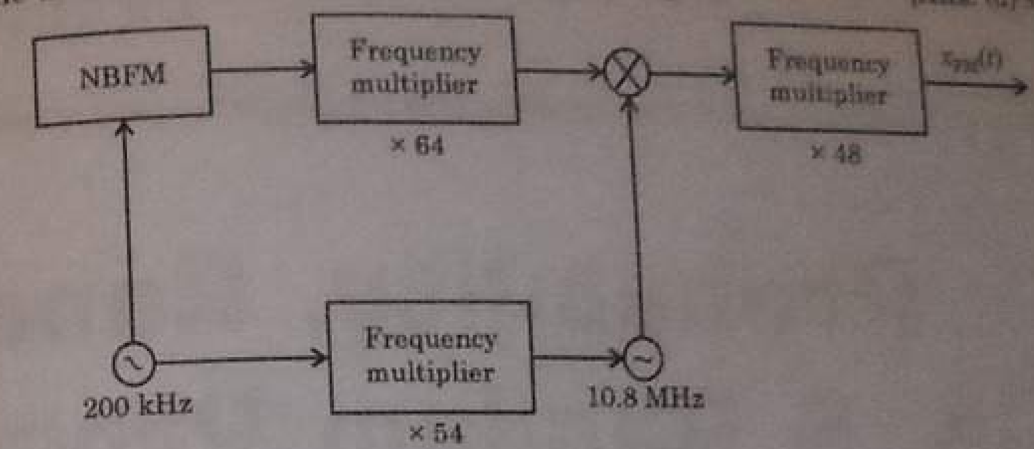


Fig. 4.40.

11. A block diagram of a typical FM receiver, covering the broadcast range of 88 to 108 MHz, is shown in figure 4.41. The amplifier frequency is 10.7 MHz. The limiter is used to remove the amplitude fluctuations caused by channel imperfection. The FM receiver is tuned to a carrier frequency of 100 MHz.

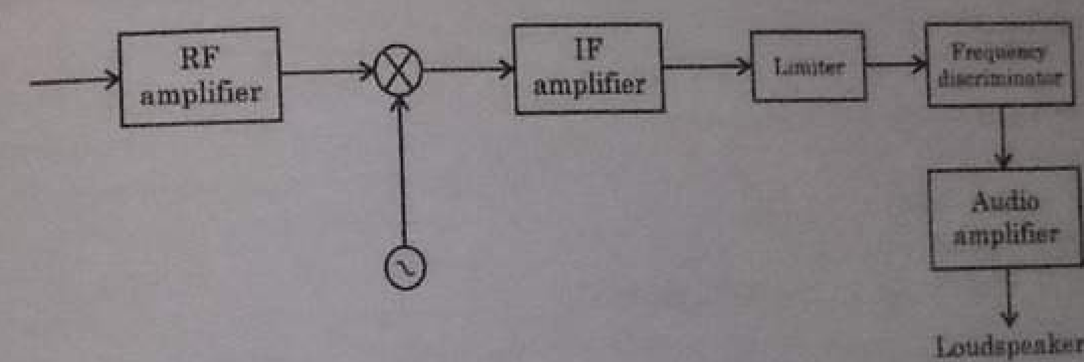


Fig. 4.41. *FM Receiver.*

   (i) A 10 kHz audio signal frequency modulates a 100 MHz carrier, producing $\beta = 0.2$. Find the bandwidths required for the RF and IF amplifiers and for the audio amplifier.
   (ii) Repeat (a) if $\beta = 5$.
   [**Ans.** (i) RF and IF amplifiers: 24 kHz, audio amplifier: 10 kHz
   (ii) RF and IF amplifier: 120 kHz, audio amplifier: 10 kHz]

CHAPTER 5

Probability, Random Signals & Random Process

✍ **Inside this Chapter**

5.1. Introduction

In second chapter, we disscused how to apply Fourier transform to analyze some signals. All these signals were described by some fixed mathematical equations. Such type of signals are called **deterministic signals**. The behaviour of such type of signals as well as processing through linear time invariant (LTI) systems can be determined with the help of mathematical models. These

mathematical models represent the complete behaviour of the signal at every instant of time. Hence, for such deterministic signals, there is no uncertainty about the value at any instant of time. This means that these signals are predictable.

There is one other class of signals, the behaviour of which cannot be predicted. Such type of signals are called **random signals**. These signals are called random signals because the precise value of these signals cannot be predicted in advance before they actually occur. The examples of random signals are the noise interferences in communication systems. This means that the noise interference during transmission is totally unpredictable. In the same way, the noise generated by the receiver itself is random. Even some other signals which are not noise signals are also random signals. These signals cannot be modelled mathematically. Actually the electromagnetic interference is the major source of random noise.

In the receiver, the thermal noise is caused by the random motion of the electrons. Although the random signals are not predictable in advance precisely, they can be described in terms of its statistical properties. It is possible to analyze the random signals statistically with the help of probability theory.

In fact, the probability theory is very essential mathematical tool in the design of digital communication systems.

5.2. Basic Definitions Related to Probability

Before we discuss the concept of probability and its applications in the field of communication, we must discuss few basic terms which are related to probability theory.

5.2.1. Experiment

An experiment is defined as the process which is conducted to get some results. If the same experiment is performed repeatedly under the same conditions, similar results are expected. But there are few experiments which do not produce the same result as we have stated above. The theory of probability mainly deals with such type of experiments. In fact, the theory of probability is applied to such type of experiments to predict the possibility of a particular output. An experiment is sometimes called **trial**. As an example, throw of a coin is an experiment or trial. This trial results in two outcomes namely **Head** and **Tail**. Similarly, drawing a card from a well shuffled pack is the trial or experiment and it results in 52 possible outcomes (*i.e.* cards). Hence, each experiment or a trial has an outcome and the possibility of this outcome can be predicted with the help of probability theory.

The outcomes of an event are called **equally** likely, if any one of them cannot be expected to occur in preference to another. As an example, the tossing of a coin results in two outcomes, Head and Tail. Both have same possibility of 50%. Such type of outcomes are called **equally** likely outcomes. On the otherhand, if the box contains two black and four white balls, then getting a white or black ball is not equally likely. This is possible that we may get white balls sequentially.

5.2.2. Sample Space

A set of all possible outcomes of an experiment *or* trial is called the **sample space** of that experiment. It is generally denoted by 'S'. The total number of outcomes in a sample space is denoted by $n(s)$. As an example, the tossing of a coin has two outcomes. Hence, we may write its sample space as:

$$S = \{H, T\}$$

where, $H \rightarrow$ Head

and $T \rightarrow$ Tail

Now, we consider another example. If three coins are tossed simultaneously, then each experiment has two outcomes namely H and T. Hence, the maxim.um possible number of outcomes will be eight.

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

So that $n(s) = 8$

5.2.3. Event

The expected subset of the sample space or happening is called an event. As an example, let us consider an experiment of throwing a cubic die. In this case, the sample space S will be as

$$S = \{1, 2, 3, 4, 5, 6\}$$

Now, if we want the number '3' to be an outcome or an even number, *i.e.*, $\{2, 4, 6\}$, then this subset is called an **event**. This is denoted by letter 'E'. Hence event E is a subset of the sample space 'S'. If event E has only one outcome, then it is called an elementary event. On the other hand, if event E does not contain any out come, then it is called a null event. If $E = S$, then an event contains all the outcomes. Such as event is called a **certain event**. It always occurs, no matter what so ever is the outcome.

Two events A and B are called independent events if happening of event A has nothing to do with happening of B. As an example, let us consider an experiment of tossing a coin two times. In this case, occurrence of head in the first throw has nothing to do with the occurrence of head in the second throw. Hence, these two events are independent events since their outcomes are independent. On the other hand, if the outcome of one event is affected by other, then these events are called **dependent events**.

5.3. Probability

Probability may be defined as the study of random experiments. In any random experiment, there is always an uncertainty that a particular event will occur or not. As a measure of probability of occurrence of an event, a number between 0 to 1 is assigned. If it is sure that an event will occur, then we can say that its probability is 100% or 1. If it is not sure that an event will occur, then we can say that its probability is 0% or 0. If it is not sure whether the event will occur or not, then its probability is between 0 and 1.

Let us consider an example. The probability of occurrence of 28th February in a year is 1 since it is certain to occur every year. On the other hand, the probability of occurrence of 30th February in a year is 0 since it never comes. Again, the probability of occurrence of 29th February in a year is neither 0 not 1. It is always between 0 and 1. Actually, it is 1/4 since it occurs every leap year *i.e.*, once in four years.

Therefore, from above discussion, we can write a mathematical expression for probability as Probability of an event A,

$$P(A) = \frac{\text{Number of possible favourable outcomes}}{\text{Total number of outcomes}}$$

For example, let us consider the probability of getting an even number in tossing of a die. In tossing of a die, an even number can occur in 3 ways out of 6 equally likely ways.

Therefore, $$P(A) = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$$

5.4. Properties of Probability

As discussed earlier, the sample space S contains all possible outcomes of an experiment. Also, an event is the subset of the sample space. Two events are said to be mutually exclusive if the occurrence of one of them precludes the occurrence of other. For example, in tossing of a coin, events Head and Tail are mutually exclusive. In throw of a die, the occurrence of number '4' will automatically exclude the occurrence of numbers 1, 2, 3 ,5 and 6. If an event contains all the outcomes then it is called certain event. Then, probability of this event is unity, *i.e.*,

$$P(A) = P(s) = 1$$

The properties of probability may be listed as under :

***Property* 1:** The probability of a certain event is unity *i.e.*,

$$P(A) = 1 \qquad ...(5.1)$$

***Property* 2:** The probability of any event is always less than or equal to 1 and non-negative. Mathematically,

$$0 \le P(A) \le 1 \qquad ...(5.2)$$

***Property* 3:** If A and B are two mutually exclusive events, then

$$P(A + B) = P(A) + P(B) \qquad ...(5.3)$$

***Property* 4:** If A is any event, then the probability of not happening of A is

$$P(\overline{A}) = 1 - P(A) \qquad ...(5.4)$$

where $\overline{A}$ represents the complement of event A.

***Property* 5:** If A and B are any two events (not mutually exclussive events), then

$$P(A + B) = P(A) + P(B) - P(AB) \qquad ...(5.5)$$

where $P(AB)$ is called the probability of events A and B both occurring simultaneously. Such an event is called joint event of A and B, and the probability $P(AB)$ is called the joint probability.

Now, if events A and B are mutually exclusive, then the joint probability, $P(AB) = 0$.

5.5. Conditional Probability

The concept of conditional probability is used in conditional occurrences of the events. Let us consider an experiment which involves two events A and B. Now the probability of event B, given that event A has occurred, is represented by $P(B/A)$. Similarly, $P(A/B)$ represents probability of event A given that event B has already occurred. Therefore, $P(B/A)$ and $P(A/B)$ are called **conditional probabilities**. These conditional probabilities may be defined in terms of their independent and joint probabilities as under:

The conditional probability of event B given that event A has already happened.

$$P(B/A) = \frac{P(AB)}{P(A)} \qquad ...(5.6)$$

where $P(AB)$ is the joint probability of A and B.

Similarly, the conditional probability of event A given that event B has already happened.

$$P(A/B) = \frac{P(AB)}{P(B)} \qquad ...(5.7)$$

At this point, it may be noted that the joint probability has commutative property which states that

$$P(AB) = P(BA) \qquad ...(5.8)$$

5.6. Probability of Statistically Independent Events

If A and B are two events in an experiment, and possibility of occurrence of event B does not depend upon occurrence of event A, then these two events A and B are known as statistically independent events. The probability of event B, given that event A has already happened is expressed as

$$P(B/A) = \frac{P(AB)}{P(A)} \qquad ...(5.9)$$

Again, since occurrence of event B does not depend on the occurrence of event A, then the probability of event B will be same as conditional probability $P(B/A)$. Mathematically,

$$P(B/A) = P(B) \qquad ...(5.10)$$

Now, putting the value of $P(B/A)$ from equation (5.10) in equation (5.9), we have

$$P(AB) = P(A)\,P(B) \qquad ...(5.11)$$

The probability of event A, given that event B has already happened, is expressed as:

$$P(A/B) = \frac{P(AB)}{P(B)} \qquad ...(5.12)$$

Again, since events A and B are statistically independent, then the probability of event A will be same as conditional probability $P(A/B)$. Mathematically,

$$P(A/B) = P(A) \quad ...(5.13)$$

Using equations (5.12) and 5.13), we have

$$P(AB) = P(A)\,P(B) \quad ...(5.14)$$

which is same as equation (5.11).

Example 5.1. A box contains 3 red, 4 white and 5 black balls. One ball is drawn at random. Find the probability that it is (*a*) red (*b*) not Black (*c*) black or white.

Solution: Let R, W and B be the events of drawing a red, a white and a black ball respectively.

(*a*)
$$P(R) = \frac{\text{Ways of choosing a red ball}}{\text{Total no. of ways of choosing a ball}}$$

$$P(R) = \frac{3}{3+4+5} = \frac{3}{12} = \frac{1}{4}$$

(*b*) Probability of not getting a black ball is $P(\overline{B})$.

$$P(\overline{B}) = 1 - P(\overline{B})$$

But
$$P(B) = \frac{5}{3+4+5} = \frac{5}{12}$$

Therefore,
$$P(\overline{B}) = 1 - \frac{5}{12} = \frac{7}{12}$$

(*c*) Probability of getting a black or white ball will be $P(B + W)$.

$$\therefore \quad P(B + W) = P(B) + P(W) \quad [\because B \text{ and } W \text{ are two mutually exclusive events}]$$

$$P(B + W) = \frac{5}{12} + \frac{4}{12}$$

$$P(B + W) = \frac{9}{12} = \frac{3}{4}$$

Example 5.2. Two dies are thrown simultaneously. Find the probability of getting a 5.

Solution: We know that each die has six faces. Therefore, the total number of possible outcomes $6 \times 6 = 36$.

Each outcome is having an equal probability of occurrence. Therefore, probability of any outcome is 1/36.

We have 4 different combinations which give a 5. They are (4, 1), (3, 2), (2, 3), (1, 4).

Hence the probability of the overall event may be given as

$$P(A) = P(5) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{4}{36} = \frac{1}{9} \quad \textbf{Ans.}$$

Example 5.3. Two cards are drawn at random from an ordinary deck of 52 playing cards. Obtain the probability that one card is a Diamond and the other card is a Heart.

Solution: Since two cards are drawn, there are $\frac{52 \times 51}{1 \times 2} = 1326$ ways of drawing 2 cards from a deck of 52 cards.

Again, because there are 13 diamonds and 13 Hearts, there are $13 \times 13 = 169$ ways of drawing a Diamond and a Heart. Therefore, the probability of

$$P(A) = \frac{\text{The number of ways of drawing a diamond and a heart}}{\text{The total number of ways of drawing two cards}}$$

$$P(A) = \frac{169}{1326} = \frac{13}{102} \quad \textbf{Ans.}$$

Example 5.4. A box contain 4 white and 3 black balls. Three balls are drawn from the box in succession. Find the probability that the first two balls are white and the third is black.

Solution: Let A_1 be the event that the first ball is a white ball.

$$\therefore \quad P(A_1) = \frac{4}{7}$$

Since there are four white balls in the box out of total, 7 balls.

Let A_2 be the event that the second ball is also a white ball.

$$\therefore \quad P\left(\frac{A_2}{A_1}\right) = \frac{3}{6}$$

Since, after the first drawing, there remain 3 white balls out of a total of 6 balls.

Let A_3 be the event that the third ball is black.

$$\therefore \quad P\left(\frac{A_3}{A_1 A_2}\right) = \frac{3}{5}$$

Since after the second drawing, there are 3 black balls out of a total of 5 balls.

Therefore the required probability is

$$P = P(A_1 + A_2 A_3) = P(A_1)\,P(A_2/A_1)\,P(A_3/A_1A_2) = \frac{4}{7} \times \frac{3}{6} \times \frac{3}{5} = \frac{6}{35} \quad \textbf{Ans.}$$

Example 5.5. A box contains 5 white, 3 red and 2 black balls. Three balls are drawn in succession. Find the probability that the balls will be of different colours.

Solution: Let W, R, B be the events of drawing white, red and black balls, respectively. It may be observed that there are six possible ways of getting the balls of different colours as follow:

$$WRB,\ WBR,\ RWB,\ RBW,\ BWR,\ BRW$$

Now, the probability of first combination WRB will be

$$P(WRB) = \frac{5}{10} \cdot \frac{3}{9} \cdot \frac{2}{8}$$

Here, it may be noted that

$\frac{5}{10}$ is the probability of getting first ball white.

$\frac{3}{9}$ is the probability of getting second ball red.

$\frac{2}{8}$ is the probability of getting third ball black.

Similarly, probabilities for the other combinations may be calculated.

Since all the six combinations (events) are mutually exclusive, the required probability is

$$P = P(WRB + WBR + RWB + RBW + BWR + BRW)$$
$$= P(WRB) + P(WBR) + P(RWB) + P(RBW) + P(BWR) + P(BRW)$$
$$P = \left(\frac{5}{10} \times \frac{3}{9} \times \frac{2}{8}\right) + \left(\frac{5}{10} \times \frac{2}{9} \times \frac{3}{8}\right) + \left(\frac{3}{10} \times \frac{5}{9} \times \frac{2}{8}\right) + \left(\frac{3}{10} \times \frac{2}{9} \times \frac{5}{8}\right)$$
$$+ \left(\frac{2}{10} \times \frac{5}{9} \times \frac{3}{8}\right) + \left(\frac{2}{10} \times \frac{3}{9} \times \frac{5}{8}\right)$$

$$P = \frac{1}{4} \quad \textbf{Ans.}$$

Example 5.6. Three students A, B and C are given a problem in Maths. The probabilities of their solving the problem are $\frac{3}{4}$, $\frac{2}{3}$ and $\frac{1}{4}$ respectively. Determine the probability that the problem is solved if all of them try to solve the problem. (PTU, 1999)

Solution: The probabilities that A, B and C will not be able to solve the problem are $\left(1-\frac{3}{4}\right), \left(1-\frac{2}{3}\right)$ and $\left(1-\frac{1}{4}\right)$ or $\frac{1}{4}, \frac{1}{3}$ and $\frac{3}{4}$ respectively.

Therefore, the probability that no one will solve the problem

$$= \frac{1}{4} \times \frac{1}{3} \times \frac{3}{4} = \frac{3}{48}$$

Hence the probability that the problem will be solved

$$= 1 - \frac{3}{48} = \frac{45}{48} = \frac{15}{16} \quad \textbf{Ans.}$$

Example 5.7. A box contains three white balls W_1, W_2, W_3 and two red balls R_1, R_2. Two balls are drawn in succession at random. Find the probability that the first drawn ball is white and the second ball is red.

Solution: Probability of event W_1 *i.e.* first ball is white will be

$$= \frac{\text{No. of white balls}}{\text{Total no. of balls}} = \frac{3}{5}$$

Now, if a white ball is drawn, there remain two white and two red balls.

Now, the conditional probability of the event of second ball red with first white

$$= P\left(\frac{R_2}{W_1}\right) = \frac{2}{4}$$

Hence, the probability of the event that the second ball is red with first white, will be

$$= P[W_1, R_2] = P\,(\text{first ball white, second red})$$

$$= P\left(\frac{R_2}{W_1}\right) \cdot P(W_1) = \frac{2}{4} \times \frac{3}{5} = \frac{6}{20} = \frac{3}{10} \quad \textbf{Ans.}$$

Example 5.8. In a coin-tossing experiment, if the coin has head, one die is thrown and the result is recorded. But, if the coin has tail, two dice are thrown and their sum is recorded. Find the probability that the recorded number is 2.

Solution: We know that in a coin-tossing experiment, the probability of getting head $= \frac{1}{2}$.

Similarly, the probability of getting tail $= 1 - \frac{1}{2} = \frac{1}{2}$

Now, the probability of getting head with recorded number 2.

$$= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) = \frac{1}{12} \quad \ldots(i)$$

Similarly, the probability of getting tail with recorded number 2, will be

$$= \left(\frac{1}{2}\right)\left(\frac{1}{36}\right) = \frac{1}{72} \quad \ldots(ii)$$

Therefore, the required probability of getting recorded number 2 will be the sum of the equations (i) & (ii)

$$= \frac{1}{12} + \frac{1}{72} = \frac{6+1}{72} = \frac{7}{72} \quad \textbf{Ans.}$$

Example 5.9. Two weak students in programming write a program. Their chance of writing a program correctly are $\frac{1}{8}$ and $\frac{1}{12}$. Now, if the probability of making a common error is $\frac{1}{10001}$ and they get the same answer, then find the probability that their program is correct.

Solution: It is given that the probability of two students A and B writing the program correctly are $\frac{1}{8}$ and $\frac{1}{12}$ respectively *i.e.*, $P(A) = \frac{1}{8}$

and $P(B) = \frac{1}{12}$

Therefore, the probability of making common error $= \frac{1}{10001}$

Hence the required probability = Probability (Program is correct)

$$= \frac{\left(\frac{1}{8}\right)\left(\frac{1}{12}\right)}{\left(\frac{1}{8}\right)\left(\frac{1}{12}\right) + \left(1-\frac{1}{8}\right)\left(1-\frac{1}{12}\right)\left(\frac{1}{10001}\right)} = 0.9924 \quad \textbf{Ans.}$$

5.7. Random Variables

As discussed earlier, an event is the possible outcome of an experiment. The range of all the possible outcomes of an experiment is called the **sample space** 'S'. When a trial *or* experiment is performed, any one sample point in the sample space is the outcome of the trial. This means that a sample point always lies in the sample space 'S'. On the other hand, an event may correspond to a single sample point *or* set of sample points.

As an example, in an experiment of tossing a die, the sample space contains six sample points. Every time, the trial is performed, the outcome is any one sample point (number 1 to 6) in the sample space. In every trial, the outcome (sample point) will occur randomly. This has no fixed output. Hence, the outcome of a trial *or* experiment is a variable which can take values over the set of sample points.

Therefore, from above discussion, we can define a random variable as under:

A function which can take on any value from the sample space and its range is some set of real numbers is called a **random variable** of the experiment. Random variables are denoted by upper case letters such as X, Y etc. and the values taken by them are denoted by lower case letters with subscripts such as x_1, x_2, y_1, y_2 etc.

Random variables may be classified as under:

1. Discrete random variables
2. Continuous random variables.

5.7.1. Discrete Random Variables

A discrete random variable may be defined as the random variable which can take on only finit[e] number of values in a finite observation interval. This means that the discrete random variable ha[s] countable number of distinct values.

For example, let us consider an experiment of tossing three coins simultaneously. In this case there are eight possible outcomes. This will constitute the sample space S. Let the number of head[s] be the random variable X. Sample space S and random variable X may be written as

$$S = \{\, HHH \quad HHT \quad HTH \quad THH \quad HTT \quad THT \quad TTH \quad TTT \,\}$$
$$X = \{\, x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \quad x_8 \,\}$$
$$\{\, 3 \quad 2 \quad 2 \quad 2 \quad 1 \quad 1 \quad 1 \quad 0 \,\}$$

Hence, from above example, the concept of random variable is quite clear. From this example, [i]t is also clear that random variable X takes on only a finite number of values *i.e.*, 8. Therefore, it is [a] discrete random variable.

5.7.2. Continuous Random Variables

A random variable that takes on an infinite number of values is called a continuous rando[m] variable. Actually, there are several physical system (experiments) that generate continuous outpu[t] or outcomes. Such systems generate infinte number of outputs or outcomes within the finite perio[d]. Continuous random variables may be used to define the outputs of such systems.

As an example, the noise voltage generated by an electronic amplifier has a continuous amplitud[e]. This means that sample space S of the noise voltage amplitude is continuous. Therefore, in th[is] case, the random variable X has a continuous range of values.

5.8. Probability Function or Probability Distribution of a Discrete Random Variable

Let X be a discrete random variable and also let $x_1, x_2, x_3, ...$ be the values that X can take.

Then $$P(X = x_j) = f(x_j) \quad ...(5.15)$$

where $j = 1, 2, 3, ...$

will be the probability of x_j.

This $f(x_j)$ or simply $f(x)$ is called the *probability function or probability distribution of the discrete random variable.*

Example 5.10. In an experiment, three coins are tossed simultaneously. If the number of heads is the random variable, find the probability function for this random variable.

Solution: In the three tosses of coins, there are eight possible outcomes. This is the sample space S. Let the random variable, number of heads, be X.

Since, in the given experiment, the random variable X, number of heads, takes only finite values, this is a discrete random variable.

Writing the sample space S and values of discrete random variable X as

$$S = \{HHH \quad HHT \quad HTH \quad THH \quad HTT \quad THT \quad TTH \quad TTT\}$$
$$X = \begin{Bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ 3 & 2 & 2 & 2 & 1 & 1 & 1 & 0 \end{Bmatrix}$$

Since the coin is fair, the probability of each of 8 possible outcomes, will be 1/8.

Now, $$P(X = 0) = P(x_8) = \frac{1}{8}$$

$$P(X = 1) = P(x_5) + P(x_6) + P(x_7)$$

or $$P(X = 1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

$$P(X = 2) = P(x_2) + P(x_3) + P(x_4)$$

or $$P(X = 1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

$$P(X = 3) = P(x_1) = \frac{1}{8}$$

Therefore, the probability function for discrete random variable X is as given below:

X	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

5.9. Cumulative Distribution Function (CDF)

The Comulative Distribution Function (*CDF*) of a random variable 'X' may be defined as the probability hat a random variable 'X' takes a value less than or equal to x. Here x is the dummy variable. In ther words, the comulative distribution function (*CDF*) provides probabilistic description of a random ariable.

Let us consider the probability of the event $X \le x$. The probability of this event may be denoted $P(X \le x)$.

Now, according to the definition, the cumulative distribution function (*CDF*) may be written as

$$CDF: F_X(x) = P(X \le x) \quad ...(5.16)$$

where $F_X(x)$ is called cumulative distribution function (*CDF*) of a random variable X.

From equation (5.16), it may be observed that $F_X(x)$ is a function of dummy variable 'x' i.e. the value taken by random variable X.

It may also be noted that for any value of 'x', the *CDF* represents a probability. The *CDF* may be defined for discrete as well as continuous random variables. Because cumulative distribution function (*CDF*) basically represents the probability of random variable 'X' for event $X \le x$, it is also called probability distribution function of the random variable or simply distribution function of the random variable. The *CDF* is sometimes also called cumulative probability distribution function.

5.9.1. Properties of Cumulative Distribution Function (CDF)

The properties of *CDF* may be listed as under:

***Property* 1:** Since cumulative distribution function (CDF) is the probability distribution function i.e. it is defined as the probability of event $(X \le x)$, its value is always between 0 and 1. This means that *CDF* is bounded between 0 and 1. Mathematically,

$$0 \le F_X(x) \le 1 \quad ...(5.17)$$

***Property* 2:** $$F_X(-\infty) = 0 \quad ...(5.18)$$

and $$F_X(\infty) = 1 \quad ...(5.19)$$

For first case, $x = -\infty$ means no possible event. Due to this fact, $P(X \le -\infty)$ will always be zero. Therefore $F_X(-\infty) = 0$.

For second case, $x = \infty$ means $P(X \le \infty)$. Since $P(X \le \infty)$ includes probability of all possible events and the probability of a certain event is '1' therefore

$$F_X(\infty) = 1$$

***Property* 3:** $$F_X(x_1) \le F_X(x_2) \quad \text{if } x_1 \le x_2 \quad ...(5.20)$$

The above property states that the *CDF*, $F_X(x)$ is a monotone non-decreasing function of x.

5.9.2. Cumulative Distribution Function (CDF) for Discrete Random Variables

If 'X' is a discrete random variable, then it takes on values at discrete points.

The Cumulative Distribution Function (CDF) is expressed as

$$F_X(x) = P(X \le x) \quad ...(5.21)$$

Now, if X is a discrete random variable, then it can take on values as $x_1, x_2, x_3, ..., x_n$. This means that x may be expressed interms of $x_1, x_2, ..., x_n$.

Figure 5.1 shows how random variable can be assigned values.

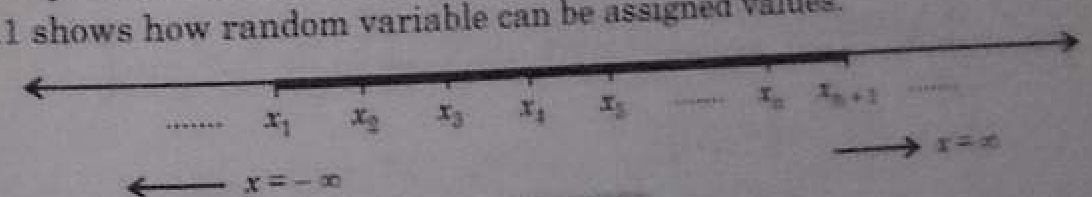


Fig. 5.1.

For a discrete random variable, the Cumulative Distribution Function (CDF) for a complete range of x may be expressed as

$$F_X(x) = \begin{cases} 0 & \text{for } -\infty \le x < x_1 \\ \sum_{j=1}^{n} P(X = x_j) & \text{for } x_1 \le x \le x_n \\ 1 & \text{for } x_n < x < \infty \end{cases} \quad ...(5.22)$$

From above equation, it is clear that *CDF* of a discrete variable at any certain event is equal to the summuation of the probabilities of random variable upto that event.

As x goes from $-\infty$ to ∞, $F_X(x)$ looks like a stair case with upward steps of height $P(X = x_i)$ at each $x = x_i$.

The *CDF* remains constant between the two events or steps.

5.10. Probability Density Function (PDF)

The derivative of cumulative distribution function (CDF) with respect to some dummy variable is known as Probability Density Function (PDF). Probability density function (PDF) is generally denoted by $f_X(x)$. Mathematically, PDF may be expressed as

$$PDF: f_X(x) = \frac{d}{dx} F_X(x) \quad \ldots(5.23)$$

where x is a dummy variable.

Probability density function (*PDF*) is the more convenient representation for continuous random variable.

5.10.1. Properties of Probability Density Function (PDF)

In this article, we shall discuss a number of properties of PDF.

***Property* 1:** Probability Density Function (PDF) is always nonzero for all values of x.

Mathematically,

$$f_X(x) \geq 0 \qquad \text{for all values of } x \quad \ldots(5.24)$$

Proof:

We know that cumulative distribution function (CDF) increases monotonically. Therefore, the derivative of cumulative distribution function (CDF) will always be positive. Since PDF is obtained by taking derivative of CDF, therefore PDF will always be positive.

***Property* 2:** The area under the PDF curve is always equal to unity. Mathematically,

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \quad \ldots(5.25)$$

Proof: The expression for PDF is given as

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Integrating both sides of above equation, we have

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} \left[\frac{d}{dx} F_X(x)\right] dx$$

$$= [F_X(x)]_{-\infty}^{\infty}$$

$$= [F_X(\infty) - F_X(-\infty)] = 1 - 0 = 1$$

***Property* 3:** The cumulative distribution function (CDF) may be obtained by integrating probability density function (PDF). Mathematically,

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx \quad \ldots(5.26)$$

Proof: The expression for Probability Density Function (PDF) is given as

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Integrating both sides of the above equation, we have

$$\int_{-\infty}^{x} f_X(x)\, dx = \int_{-\infty}^{x} \left[\frac{d}{dx} F_X(x)\right] dx$$

or $$\int_{-\infty}^{x} f_X(x)\, dx = [F_X(x)]_{-\infty}^{x} = [F_X(x) - F_X(-\infty)]$$

or $$= F_X(x) - 0 = F_X(x)$$

Hence, we have $$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx$$ **Hence Proved.**

***Property* 4:** Probability of the event $\{x_1 < X \leq x_2\}$ is simply given by the area under the Probability Density Function (PDF) curve in the range $x_1 < X \leq x_2$.

Mathematically,

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\, dx \quad \ldots(5.27)$$

Example 5.11. Probability density function (PDF) is given by the expression $f_X(x) = a\, e^{-b|x|}$. Here X is a random variable whose values lie in the range $x = -\infty$ to $x = +\infty$. Determine the following:

(*i*) The relationship between a and b.

(*ii*) Cumulative Distribution Function (CDF)

(*iii*) The probability that outcome lies between 1 and 2. *(A.M.I.E., Examination, 1997)*

Solution: (*i*) We know that the area under the *PDF* curve is always equal to 1.

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1,$$

Given that $$f_X(x) = a\, e^{-b|x|}$$

Therefore, $$\int_{-\infty}^{\infty} a e^{-b|x|}\, dx = 1$$

or $$\int_{-\infty}^{0} a e^{-b(-x)}\, dx + \int_{0}^{\infty} a e^{-bx}\, dx = 1$$

or $$\int_{-\infty}^{0} a e^{-b(-x)}\, dx + \int_{0}^{\infty} a e^{-bx}\, dx = 1$$

or $$\frac{a}{b}\left[e^{bx}\right]_{-\infty}^{0} - \frac{a}{b}\left[e^{-bx}\right]_{0}^{\infty} = 1$$

or $$\frac{a}{b}\,[e^{0} - e^{-\infty} - e^{-\infty} + e^{0}] = 1$$

or $$\frac{a}{b}\,[1 - 0 - 0 + 1] = 1$$

or $$\frac{a}{b}\,[1 + 1] = 1$$

or $$\frac{2a}{b} = 1$$

or $$b = 2a$$

This is required relationship between a and b.

(*ii*) We know that cumulative distribution function (CDF) may be obtained by integrating PI

CDF: $$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx = \int_{-\infty}^{x} a e^{-b|x|}\, dx$$

Now, we shall break the above expression for $x < 0$, $x = 0$, $x > 0$.

Therefore, $F_X(x) = \underbrace{\int_{-\infty}^{x} a e^{-b(-x)}\, dx}_{\text{This factor for } x<0} + \underbrace{\int_{x=0} a e^{-bx} dx + \int_{0^+}^{x} a e^{-bx}\, dx}_{\text{These two factors for } x \ge 0}$

Hence, for $x < 0$, we have

$$\int_{-\infty}^{x} a e^{bx}\, dx = \frac{a}{b}\left[e^{bx}\right]_{-\infty}^{x} = \frac{a}{b}\,[e^{bx} - e^{-\infty}] = \frac{a}{b}\, e^{bx} = \frac{1}{2}\, e^{bx}$$

Similary, for $x \ge 0$, we have

$$\int_{x=0} a e^{b}\, dx + \int_{0^+}^{x} a e^{-bx}\, dx$$

$$= \frac{a}{b}[e^0] - \frac{a}{b}\left[e^{-bx}\right]_{0^+}^{x}$$

$$= \frac{a}{b} \times 1 - \frac{a}{b}\,[e^{-bx} - e^{0^+}]$$

$$= \frac{a}{b} - \frac{a}{b}\,[e^{-bx} - 1] = \frac{a}{b}[1 - e^{-bx} + 1] \qquad [\because e^{0^+} \cong 1]$$

$$= \frac{a}{b}\,[2 - e^{-bx}] = \frac{a}{2a}\,[2 - e^{-bx}] \qquad (\because b = 2a)$$

$$= \frac{1}{2}\,(2 - e^{-bx}) = 1 - \frac{1}{2}\, e^{-bx}$$

Therefore, the expression for *CDF* will be given by

$$F_X(x) = \begin{cases} \frac{1}{2}\, e^{bx} & \text{for } x < 0 \\ 1 - \frac{1}{2}\, e^{-bx} & \text{for } x \ge 0 \end{cases}$$

Plot of cumulative distribution function (CDF) with respect to x:

We know from previous discussion that for a CDF $F_X(x)$

$F_X(x)$ at $x = -\infty = 0$

and $F_X(x)$ at $x = \infty = 1$

Also, $F_X(x)$ at $x = 0 = 1 - \frac{1}{2}\, e^{-bx}$

$$= 1 - \frac{1}{2}\, e^{-b.0} = 1 - \frac{1}{2}\, e^0 = \frac{1}{2}$$

Hence, from the above information we can have an approximate plot of $F_X(x)$ as:

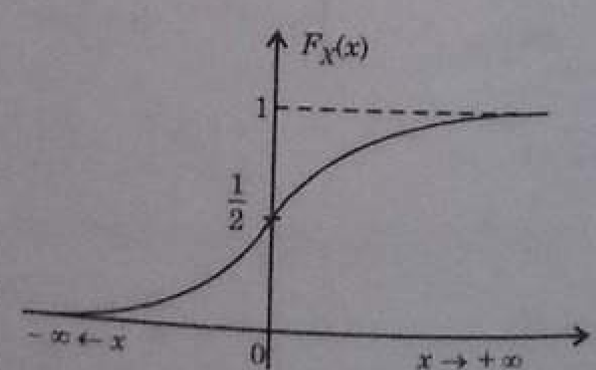


Fig. 5.2. *Plot of CDF*

(*iii*) We know that the probability that outcome lies between x_1 and x_2, is given by the expression

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\, dx$$

Given that

$$x_1 = 1 \text{ and } x_2 = 2$$

Therefore,

$$P(1 < X \le 2) = \int_1^2 a e^{-b|x|}\, dx$$

For the interval 1 to 2, $e^{-b|x|} = e^{-bx}$

Hence,

$$P(1 < X \le 2) = \int_1^2 a e^{-bx}\, dx = -\frac{a}{b}\left[e^{-bx}\right]_1^2 = -\frac{a}{b}\,[e^{-2b} - e^{-b}]$$

or

$$P(1 < X \le 2) = \frac{a}{b}\,[e^{-b} - e^{-2b}]$$

But since

$$2a = b, \text{ So } \frac{a}{b} = \frac{1}{2}$$

$\therefore$

$$P(1 < X \le 2) = \frac{1}{2}\,[e^{-b} - e^{-2b}] \qquad \textbf{Ans.}$$

Example 5.12. The CDF for a certain random variable is given as

$$F_X(x) = \begin{cases} 0 & -\infty < x \le 0 \\ kx^2 & 0 < x \le 10 \\ 100\text{ k} & 10 < x < \infty \end{cases}$$

(*i*) **Find the value of k;**

(*ii*) **Find the value of $P(X \le 5)$;**

(*iii*) **Find the value of $P(5 < X \le 7)$;**

(*iv*) **Find the expression for PDF.**

Solution: (*i*) The given *CDF* is

$$F_X(x) = \begin{cases} 0 & -\infty < x \le 0 \\ kx^2 & 0 < x \le 10 \\ 100\text{k} & 10 < x < \infty \end{cases} \qquad ...(i)$$

We know that the expression for CDF is given as

$$F_X(x) = \begin{cases} 0 & -\infty < x \le x_1 \\ \sum_{j=1}^{n} P(X = x_j) & x_1 < x \le x_n \\ 1 & x_n < x < \infty \end{cases} \qquad ...(ii)$$

Comparing the two equations (*i*) & (*ii*) we have

$$F_X(x) = 100\text{ k} = 1 \qquad \text{for } 10 < x < \infty$$

or

$$k = \frac{1}{100} \qquad \textbf{Ans.}$$

(*ii*) Since CDF is expressed as

$$F_X(x) = P(X \le x)$$

or

$$P(X \le x) = F_X(x)$$

Therefore, for $x = 5$, we have

$$P(X \le 5) = F_X(5)$$

But

$$F_X(x) = \frac{1}{100} \times x^2$$

Therefore,

$$P(X \le 5) = \frac{1}{100}\,(5)^2$$

$$P(X \le 5) = \frac{1}{100} \times 25 = \frac{1}{4} \qquad \textbf{Ans.}$$

(*iii*) The PDF is expressed as

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Since

$$F_X(x) = \begin{cases} 0 & -\infty < x \le 0 \\ \dfrac{x^2}{100} & \text{for } 0 < x \le 10 \\ 1 & \text{for } 10 < x < \infty \end{cases}$$

Now on differentiating $F_X(x)$ with respect to x, we have

PDF, $f_X(x) = 0$ for $-\infty < x \le 0$, $10 < x < \infty$

and

$$f_X(x) = \frac{d}{dx}\left[\frac{x^2}{100}\right] \quad \text{for } 0 < x \le 10$$

$$= \frac{x}{50} \quad \text{for } 0 < x \le 10$$

(*iv*) We know that $P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$

Here, $x_1 = 5, x_2 = 7$

$$P(5 < X \le 7) = \int_5^7 \frac{x}{50}\,dx \qquad \left[\because f_X(x) = \frac{x}{50}, \text{ for } 0 < x \le 10\right]$$

or

$$P(5 < X \le 7) = \frac{1}{50}\left[\frac{x^2}{2}\right]_5^7 = \frac{1}{100}[49 - 25]$$

$$P(5 < X \le 7) = \frac{1}{100} \times 24 = \frac{24}{100} \qquad \textbf{Ans.}$$

xample 5.13. A continuous random variable has a probability density function (PDF) expressed s

$$f_X(x) = 2\,e^{-2x} \qquad \text{for } x \ge 0$$

Determine the probability that it will take a value between 1 and 3.

olution: We know that the relation between probability and PDF is expressed as

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$$

Now the probability that a random variable will be between 1 and 3 is expressed as

$$P(1 < X \le 3) = \int_1^3 f_X(x)\,dx = \int_1^3 2e^{-2x}\,dx$$

or

$$P(1 < X \le 3) = 2\left[\frac{e^{-2x}}{-2}\right]_1^3 = 2\left[-\frac{1}{2}(e^{-6} - e^{-2})\right] = -[-0.1328] = 0.1328 \qquad \textbf{Ans.}$$

mple 5.14. A random process provides measurements x between the values 0 and 1 a probability density function (PDF) given as

$$f_X(x) = 12x^3 - 21x^2 + 10x \quad \text{for } 0 \le x \le 1 = 0, \text{ otherwise}$$

Determine the followings:

(*i*) $P\left[X \le \frac{1}{2}\right]$ **and** $P\left[X > \frac{1}{2}\right]$

(*ii*) Obtain a number k such that $P[X \le k] = \frac{1}{2}$. *(Karnataka University, 1999)*

Solution: (*i*) We know that

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$$

Therefore,

$$P(X \le \tfrac{1}{2}) = \int_0^{1/2} f_X(x)\,dx = \int_0^{1/2} (12x^3 - 21x^2 + 10x)\,dx$$

or

$$P(X \le \tfrac{1}{2}) = \left[\frac{12x^4}{4} - \frac{21x^3}{3} + \frac{10x^2}{2}\right]_0^{1/2} = \left[3x^4 - 7x^3 + 5x^2\right]_0^{1/2}$$

or

$$P(X \le \tfrac{1}{2}) = \left[3\left(\frac{1}{2}\right)^4 - 7\cdot\left(\frac{1}{2}\right)^3 + 5\left(\frac{1}{2}\right)^2 - 0\right] = \frac{9}{16}$$

Also,

$$P(X > \tfrac{1}{2}) = 1 - P(X \le \tfrac{1}{2}) = 1 - \frac{9}{16} = \frac{7}{16}$$

(*ii*) Since

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$$

Therefore, $P(X \le k) = \int_0^k f_X(x)\,dx$

But given that $P(X \le k) = \frac{1}{2}$

Hence,

$$\frac{1}{2} = \int_0^k (12x^3 - 21x^2 + 10x)\,dx = \left[\frac{12x^4}{4} - \frac{21x^3}{3} + \frac{10x^2}{2}\right]_0^k$$

or

$$\frac{1}{2} = \left[3x^4 - 7x^3 + 5x^2\right]_0^k = 3k^4 - 7k^3 + 5k^2 \qquad ...(i)$$

or $6k^4 - 14k^3 + 10k^2 - 1 = 0$

The required answer will be the root of equation (*i*) which lies between 0 and 1.

Therefore, solving equation (*i*) for a root between 0 and 1, we get

$$k = 0.453 \qquad \textbf{Ans.}$$

Example 5.15. Determine whether the function given by expression

$$f_X(x) = \begin{cases} = 0 & \text{for } x < 2 \\ \frac{1}{18}(3+2x) & \text{for } 2 \le x \le 4 \\ = 0 & \text{for } x > 4 \end{cases}$$

(RGP University, MP-1999)

is a density function?

Solution: If any given function is a probability density function (PDF), then it must satisfy following two basic conditions:

(i) $f_X(x) \ge 0$ for every x

(ii) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

The given function is $= 0$ for $x < 2$

$$f_X(x) = \frac{1}{18}(3+2x) \text{ for } 2 \le x \le 4 = 0 \quad \text{for } x > 4$$

From this expression, it may be observed that first condition is satisfied.

For second condition, we have

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{2} 0\, dx + \int_{2}^{4} \frac{1}{18}(3+2x)\, dx + \int_{4}^{\infty} 0\, dx$$

or

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 0 + \frac{1}{18}\left[3x + x^2\right]_2^4 + 0 = \frac{1}{18}\left[\{3(4)+16\} - \{3(2)+4\}\right]$$

$$= \frac{1}{18}[12 + 16 - 6 - 4] = \frac{1}{18}[28 - 10] = \frac{1}{18} \times 18 = 1$$

Therefore, second condition is also satisfied.

This means that the given function is a probability function (PDF).

Example 5.16. Determine the constant k such that the function $f_X(x)$ given by the expression

$$f_X(x) = \begin{cases} \frac{1}{k} & \text{for } a \le x < b \\ = 0 & \text{elsewhere} \end{cases}$$

is a probability density function (PDF). Also, find the commulative distribution function (CDF) of the random variable X if k satisfies the conditions for $f_X(x)$ to be a probability density function (PDF).

Solution: Since it is given that $f_X(x)$ is a density function, therefore, it will satisfy the following requirement:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Putting the value of $f(x)$ in the given limit, we get

$$\int_a^b \frac{1}{k}\, dx = 1 \quad \text{or} \quad \frac{1}{k}\int_a^b dx = 1$$

or

$$\frac{1}{k}[x]_a^b = 1 \quad \text{or} \quad k = [b - a] \qquad \textbf{Ans.}$$

We know that CDF is given as

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx = \int_a^x f_X(x)\, dx = \int_a^x \frac{1}{k}\, dx = \int_a^x \frac{1}{b-a}\, dx$$

or

$$F_X(x) = \frac{1}{b-a}\int_a^x dx = \frac{1}{b-a}[x]_a^x = \frac{1}{b-a} \cdot x - a = \frac{x-a}{b-a} \qquad \textbf{Ans.}$$

Example 5.17. Determine whether the following function is commulative distribution function (CDF):

$$F_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{2}\left(\frac{x}{a} + 1\right) & \text{for } -a \le x \le a \\ 1 & \text{for } x > a \end{cases}$$

(WBTU, Kolkata-2003)

Solution: We know that the probability density function (PDF) is given as

$$f_X(x) = \frac{d}{dx} F_X(x)$$

$$f_X(x) = \begin{cases} \frac{1}{2a} & \text{for } -a \le x \le a \\ 0 & \text{elsewhere} \end{cases}$$

The required condition for a function $f_X(x)$ to be PDF is ...(i)

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Therefore,

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{-a} f_X(x)\, dx + \int_{-a}^{a} f_X(x)\, dx + \int_{a}^{\infty} f_X(x)\, dx$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{-a} 0 \cdot dx + \int_{-a}^{a} \frac{1}{2a}\, dx + \int_{a}^{\infty} 0 \cdot dx = 0 + \frac{1}{2a}[x]_{-a}^{a} + 0 = \frac{1}{2a}[a - (-a)] = 1$$

Hence, the function $f_X(x)$ satisfies the condition of a PDF. This means that $F_X(x)$ is a commulativ[e] distribution function (CDF). **Ans.**

5.11. Joint Cumulative Distribution Function

We may also define the outcome of an experiment by two random variables. In earlier article, w[e] discussed the, cumulative distribution function (CDF) for a single random variable. In this artic[le] we shall discuss the CDF for two random variables X and Y which is known as combined CDF [or] simply joint Distribution Function.

The joint Distribution Function or joint CDF $F_{XY}(x, y)$ of two random variables X and Y [is] defined as the probability that the random variable X is less than *or* equal to a specified value x a[nd] that the random variable y is less than *or* equal to a specified value y.

The joint Cumulative Distribution Function may be defined systematically as:

The joint Cumulative Distribution Function $F_{XY}(x, y)$ may be defined as the probability that the outcome of an experiment will result in a sample point lying inside the range $(-\infty < X \leq x, -\infty < Y \leq y)$ of the joint sample space. The joint sample space is the combined sample space of X and Y. Mathematically,

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \quad \text{...(5.28)}$$

5.11.1. Properties of Joint Cumulative Distribution Function

The Joint Cumulative Distribution Function has the following properties:

***Property* 1:** The Joint Cumulative Distribution Function is a non-negative function. Mathematically,

$$F_{XY}(x, y) \geq 0 \quad \text{...(5.29)}$$

The Joint Cumulative Distribution Function is basically defined as the probability in the Joint sample space of random variables. We know that the probability always lies between 0 and 1. Therefore, the joint Cumulative Distribution Function also lies between 0 and 1 and hence non-negative.

***Property* 2:** The Joint Cumulative Distribution Function is a monotone non-decreasing function of both x and y.

***Property* 3:** The Joint Cumulative Distribution Function is always continuous everywhere in the xy-plane.

5.12. The Joint Probability Density Function

The Joint Probability Density Function *or* simply Joint PDF is the PDF of two or more random variables.

The joint PDF of any two random variables X and Y may be defined as the partial derivative of the joint cumulative distribution function $F_{XY}(x, y)$ with respect to the dummy variables x and y. Mathematically,

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \, \partial y} \quad \text{...(5.30)}$$

Here, we take partial derivative since the differentiation is with respect to two variables x and y.

5.12.1. Properties of Joint PDF

The Joint PDF has the following properties:

***Property* 1:** The Joint PDF is non-negative. Mathematically,

$$f_{XY}(x, y) \geq 0 \quad \text{...(5.31)}$$

Because the joint PDF is a derivative of a non-negative function (Joint CDF), therefore its value is always positive.

***Property* 2:** The total volume under the surface of joint PDF is equal to unity. Mathematically,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1 \quad \text{...(5.32)}$$

Because $f_{XY}(x, y)$ is a two-dimensional function, it represents a surface in the x, y plane. Hence, integration is called as volume under the surface.

***Property* 3:** The Joint PDF is continuous everywhere because joint CDF is continuous.

5.12.2. The Relationship between Joint *PDF* and Probability

The relationship between probability and PDF over a certain interval is expressed as

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\, dx \quad \text{...(5.33)}$$

The above relationship may be extended for two random variables X and Y. Mathematically,

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x, y)\, dx\, dy \quad \text{...(5.34)}$$

In above expression, the double integral represents the volume between x-y palne and the surface $f_{XY}(x, y)$.

Now, if the two random variables X and Y are statistically independent, then joint PDF of these two random variables becomes a product of two separate PDFs. Mathematically,

$$f_{XY}(x, y) = f_X(x)\, f_Y(y) \quad \text{...(5.35)}$$

Putting this value of $f_{XY}(x, y)$ from equation in equation, we have

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_X(x)\, f_Y(y)\, dx\, dy \quad \text{...(5.36)}$$

$$= \int_{y_1}^{y_2} f_Y(y)\, dy + \int_{x_1}^{x_2} f_X(x)\, dx \quad \text{...(5.37)}$$

The above equation provides the relationship between probability PDF of statistically independent random variables X and Y.

Example 5.18. The joint PDF of two random variables is expressed as $F_{XY}(x, y)$. If the random variables X and Y are statistically indendent, then show that

$$f_{XY}(x, y) = F_X(x)\, F_Y(y)$$

Here $f_X(x)$ is the PDF of random variable X and $f_Y(y)$ is the PDF of random variable Y.

Solution: We know that the joint *PDF* may be defined in terms of joint CDF as

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x\, \partial y}$$

Now, if random variables X and Y are statistically independent, then

$$f_{XY}(x, y) = \frac{d}{dx} F_{XY}(x, y)\, \frac{d}{dy} F_{XY}(x, y) = f_X(x)\, f_Y(y)$$

Example 5.19. A joint probability density function of two random variables X and Y is given as

$$f_{XY}(x, y) = \begin{cases} e^{-(x+y)} & \text{for } x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Determine the followings:

(i) $P(X < 1)$

(ii) $P(X > Y)$

(iii) $P(X + Y < 1)$

Solution: Given that joint probability density function is given as

$$f_{XY}(x, y) = e^{-(x+y)} = e^{-x} \cdot e^{-y} = f_X(x) \cdot f_Y(y)$$

Hence, X and Y are independent random variables.

Therefore, $f_X(x) = e^{-x}$ for $x \geq 0$

and $f_Y(y) = e^{-y}$ for $y \geq 0$

Now,

(i) $$P(X<1) = \int_0^1 e^{-x}\,dx = \left[\frac{e^{-x}}{-1}\right]_0^1 = -\left[e^{-x}\right]_0^1$$

or $$P(X<1) = -(e^{-1}-e^{-0}) = -\left(\frac{1}{e}-1\right) = 1-\frac{1}{e} \quad \textbf{Ans.}$$

(ii) $$P(X>Y) = 1-P(X<Y)$$

or $$P(X<Y) = \int_0^\infty\left[\int_0^y f_{XY}(x,y)\,dx\right]dy = \int_0^\infty e^{-y}\left[\frac{e^{-x}}{-1}\right]_0^y dy$$

or $$P(X<Y) = -\int_0^\infty e^{-y}\cdot(e^{-y}-1)\,dy = -\left(\frac{e^{-2y}}{-2}+e^{-y}\right)_0^\infty = \left(\frac{1}{2}e^{-2y}-e^{-y}\right)_0^\infty$$

$$= \left[\left(\frac{1}{2}e^{-\infty}-e^{-\infty}\right)-\left(\frac{1}{2}e^{-0}-e^{-0}\right)\right] = \frac{1}{2}\cdot 0-0-\frac{1}{2}\cdot 1+1$$

or $$P(X<Y) = 0-0-\frac{1}{2}+1 = \frac{1}{2}$$

Hence, $$P(X>Y) = 1-P(X<Y) = 1-\frac{1}{2} = \frac{1}{2}$$

(iii) $$P(X+Y<1) = \int_0^1\left[\int_0^{1-x} e^{-(x+y)}\,dy\right]dx$$

or $$P(X+Y<1) = \int_0^1 e^{-x}\left[\frac{e^{-y}}{-1}\right]_0^{1-x} dx = -\int_0^1 e^{-x}\cdot[e^{-(1-x)}-1)]\,dx$$

or $$P(X+Y<1) = -\int_0^1[e^{-1}-e^{-x}]\,dx = 1-\frac{2}{e} \quad \textbf{Ans.}$$

Example 5.20. The joint probability density function of two random variables X and Y is given as

$$f_{XY}(x,y) = \begin{cases} C(2x+y) & \text{for } 0 \leq x \leq 2,\ 0 \leq y \leq 3 \\ 0 & \text{elsewhere} \end{cases}$$

Determine the value of constant C.

Solution: For a joint PDF, we know that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = 1$$

Putting the value of $f_{XY}(x, y)$, we get

$$\int_0^3\int_0^2 C(2x+y)\,dxdy = 1$$

or $$C\int_0^3\left[\frac{2x^2}{2}+yx\right]_0^2 dy = 1 \quad C\int_0^3(2y+4)\,dy = 1$$

or $$C\left(\frac{2y^2}{2}+4y\right)_0^3 = 1 \quad C(y^2+4y)_0^3 = 1$$

or $$C(9+12-0-0) = 1 \quad 21\,C = 1 = \frac{1}{21} \quad \textbf{Ans.}$$

5.13. Marginal Densities

In case, when the probability density functions $f_X(x)$ and $f_Y(y)$ for any single random variable are obtained from joint PDF, then $f_X(x)$ and $f_Y(y)$ are called marginal PDF or simply marginal densities.

We can obtain the marginal $PDFs$ as follows:

Since, the CDF for a single random variable may be obtained from its PDF as

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx \quad ...(5.38)$$

The above expression may be extended for any two random variable, X and Y as

$$F_{XY}(x,y) = P(X \leq x, Y \leq y) \quad ...(5.39)$$

$$F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(x,y)\,dx\,dy \quad ...(5.40)$$

Now, if we have to find CDF of random variable X, then the value of Y does not matter. Therefore,

$$F_X(x) = P(X \leq x) = P(X \leq x, -\infty < y \leq \infty) \quad ...(5.41)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{XY}(x,y)\,dx\,dy \quad ...(5.42)$$

For a single random variable, the PDF may be obtained by differentiating CDF.

$$f_X(x) = \frac{d}{dx}F_X(x) \quad ...(5.43)$$

In this equation differentiation is with respect to x because y varies from $-\infty$ to $+\infty$.

or $$f_X(x) = \frac{d}{dx}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{XY}(x,y)\,dx\,dy\right] \quad ...(5.44)$$

or $$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy \quad ...(5.45)$$

Here, dx is cancelled due to differentiation. Therefore, the above equation shows that the PDF $f_X(x)$ may be obtained by integrating joint PDF $f_{XY}(x, y)$ with respect to undersired random variable y from $-\infty$ to ∞.

This $f_X(x)$ is then called marginal PDF. In the same way, the marginal PDF for random variable y may be obtained. It is expressed as

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx \qquad ...(5.46)$$

Therefore the marginal densities are:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy \qquad ...(5.47)$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx \qquad ...(5.48)$$

5.14. Conditional Probability Density Function

Out of the two random variables, one variable may take a fixed value. In this case, the PDF is called *conditional*. As an example, out of the two continuous random variables X and Y, let $X = x$.

Then we may find the conditional, *PDF* of Y given that $X = x$ as,

$$f_Y(y/x) = \frac{f_{XY}(x, y)}{f_X(x)} \qquad ...(5.49)$$

where $f_X(x)$ is the marginal density of random variable X.

Similarly, we may find the conditional *PDF* of X given that $Y = y$. Then

$$f_X(x/y) = \frac{f_{XY}(x, y)}{f_Y(y)} \qquad ...(5.50)$$

5.14. 1. Properties of Conditional PDF

The conditional PDF has following properties:

***Property* 1:** A conditional *PDF* is always non-negative function. Mathematically,

$$f_X(x/y) \geq 0 \qquad ...(5.51)$$

and

$$f_Y(y/x) \geq 0 \qquad ...(5.52)$$

The conditional PDF is the ratio of two PDFs which are non-negative. Therefore, conditional PDF is also non-negative.

***Property* 2:** The are under the conditional PDF is equal to unity. Mathematically,

$$\int_{-\infty}^{\infty} f_X(x/y)\, dx = 1 \qquad ...(5.53)$$

and

$$\int_{-\infty}^{\infty} f_Y(y/x)\, dy = 1 \qquad ...(5.54)$$

***Property* 3:** If two random variables X and Y are statistically independent, then the conditional PDF reduces to marginal density.

Mathematically,

$$f_Y(y/x) = f_Y(y) \qquad ...(5.55)$$

and

$$f_X(x/y) = f_X(x) \qquad ...(5.56)$$

5.15. Statistical Averages of Random Variables

We know that the Probability Density Function (PDF) gives some kind of information about the random variable. However, the interpretation of this information is quite complex. There are some other measures or numbers which give more useful and quick information about the random variable. Collectively these characteristic numbers or measures are known as **statistical averages**. Some useful and convenient informations like mean or average, moments, standard deviation, variance etc. may be obtained for random variables. These factors are special characteristics of the Probability Density Function (PDF) and actually they depend on the particular type of PDF.

Now we shall discuss these statistical averages.

5.15.1. Mean or Average

The mean or average of any random variable is expressed by the summation of the values of random variables X weighted by their probabilities. Mean value of a random variable is denoted by m_x

Mean value is also known as expected value of random variable X.

$$m_x = E(X) \qquad ...(5.57)$$

where $E[\ ]$ represents expection operator.

Therefore both m_x and $E(X)$ have the same meaning.

Also, in general term, mean or average is given by

$$\text{Mean or average of } X = \frac{\text{Arithmatic sum of all values of } X}{\text{Total number of values of } X}$$

5.15.2. Mean Value of Discrete Random Variable

Let the discrete random variable X be take the following values

$$X = \{x_1, x_2, x_3 \ldots x_n\} \qquad ...(5.58)$$

Then the mean or average value m_x is expressed as

$$m_x = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) + \ldots + x_nP(x_n) \qquad ...(5.59)$$

Therefore, the mean value of a discrete random variable X is equal to the summation of values of X weighted by probability of that particular value.

Hence, mean value of discrete random variable

$$m_x = E(X) = \overline{X} = \sum_{i=1}^{n} x_i\, P(x_i) \qquad ...(5.60)$$

Here $\overline{X}$ is also the notation for mean values.

5.15.3. Mean Values of Continuous Random Variables

If random variable X becomes continuous, the sample points $x_1, x_2, x_3 \ldots x_n$ becomes quite clos to each other such that $(x_2 - x_1 \cong 0)$. Therefore, summation in equation 5.60 for discrete rando variable converts to an integration over the complete range of $x(-\infty < x < \infty)$. variable converts to an integration over the complete range of x the mean or average value is expressed as

For continuous random variable X, the mean or average value is expressed as

$$m_x = \int_{-\infty}^{\infty} x f_X(x)\, dx \qquad ...(5.6$$

where $f_X(x)$ = Probability density function (PDF)

Also, if a function $g(x)$ transforms random variable X into another random variable, then me or average of $g(x)$ may be expressed as

$$m = \overline{g(x)} = E[g(x)] \quad \text{...(5.62)}$$

$$m = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \quad \text{...(5.63)}$$

Example 5.21. The random variable Z is the function of another random variable X in such a way that $Z = \cos(X)$ and X is uniformly distributed in the interval $(-\pi, \pi)$ *i.e.*,

$$f_X(x) = \begin{cases} \dfrac{1}{2\pi} & \text{for } -\pi < x < \pi \\ 0 & \text{otherwise} \end{cases}$$

Determine the mean value of Z.

Solution: The given random variable is

$$Z = \cos(X)$$

Now, assuming $g(x) = \cos(X)$. The above equation may be written as

$$Z = g(x)$$

We know that the mean value of random variable $Z = g(x)$ is expressed as

$$m = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Putting the values of $g(x)$ and $f_X(x)$, we have

$$m = \int_{-\pi}^{\pi} \cos x \left(\frac{1}{2\pi}\right) dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos x\, dx$$

or

$$m = \frac{1}{2\pi}[-\sin x]_{-\pi}^{\pi} = \frac{1}{2\pi}[0] = 0 \qquad \textbf{Ans.}$$

Thus it may be observed from above that since the given function is $Z = \cos(X)$, it has zero mean value over the period of one cycle.

5.15.4. Moments and Variance

The n^{th} moment of any random variable X may be defined as the mean value of X^n.

i.e.

$$g(x) = X^n \quad \text{...(5.64)}$$

then the mean is written as

$$\text{mean} = \overline{g(x)} \quad \text{...(5.65)}$$

$$\overline{X^n} = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = E(X^n) \quad \text{...(5.66)}$$

$$\overline{X^n} = \int_{-\infty}^{\infty} x^n f_X(x)\, dx$$

Putting $n = 1$, the above equation becomes

$$\overline{X} = E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \quad \text{...(5.67)}$$

or

$$\overline{X} = m_x \quad \text{...(5.68)}$$

Hence, the first moment of random variable X will be same as its mean value.

If $n = 2$, then equation (5.66) becomes

$$\overline{X^2} = E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx \quad \text{...(5.69)}$$

where $\overline{X^2}$ is known as mean square value of random variable X.

Similarly, the central moments are the moments of the difference between random variable X and its mean m_x. Therefore, the nth central moment may be given as

$$E[(X - m_x)^n] = \int_{-\infty}^{\infty} (x - m_x)^n f_X(x)\, dx \quad \text{...(5.70)}$$

The second central moment for $n = 2$, is known as **variance** of random variable X *i.e.*,

$$\text{Variance }[X] = E[(X - m_x)^2] \quad \text{...(5.71)}$$

$$= \int_{-\infty}^{\infty} (x - m_x)^2 f_X(x)\, dx \quad \text{...(5.72)}$$

Hence, it is clear that the variance provides an indication about randomness of the random variable.

Variance is generally represented by σ_x^2 *i.e.*,

$$\sigma_x^2 = \text{Variance }(X) \quad \text{...(5.73)}$$

$$\sigma_x^2 = \text{Var }(X)$$

or

$$\sigma_x^2 = E[(X - m_x)^2] \quad \text{...(5.74)}$$

or

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 f_X(x)\, dx \quad \text{...(5.75)}$$

Expanding equation (5.74), we get

$$\sigma_x^2 = E[X^2 - 2\, m_x X + m_x^2] \quad \text{...(5.7}$$

$$\because (X - m_x)^2 = X^2 + m_x^2 - 2\, m_x$$

The $E[\;]$ operator represents mean or average value, therefore it is linear. Thus,

$$\sigma_x^2 = E(X^2) - 2\, m_x E(X) + m_x^2 \quad \text{...(5.7}$$

or

$$\sigma_x^2 = E(X^2) - 2\, m_x m_x + m_x^2 \quad \text{...(5.}$$

or

$$\sigma_x^2 = E(X^2) - m_x^2$$

or

$$\sigma_x^2 = \text{mean square value} - \text{square of mean value} \quad \text{...(5.}$$

The above equation may also be written as

$$\text{Variance} = \sigma_x^2 \quad \text{...(5}$$

$$= \overline{X^2} - m_x^2$$

The square root of variance is known as **standard deviation** of a random variable X. Standard deviation gives the measure of spread observed over the values of X related to mean value. ...(5

Therefore,

$$\text{Standard deviation} = \sqrt{\text{Variance}} \quad \text{...(}$$

$$= \sqrt{\sigma_x^2} = \sigma_x$$

Thus, standard deviation = σ_x

$$= \sqrt{E(X^2) - m_x^2}$$ ...(5.83)

Example 5.22. The probability density function (PDF) of a continuous random variable is of the form

$$f_X(x) = \frac{1}{2} e^{-|x|} \quad \text{for } -\infty < x < \infty$$

Determine the mean of the random variables.

Solution: We know that for a continuous random variable mean is expressed as

$$m_x = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{2} e^{-|x|}\, dx = \int_{-\infty}^{\infty} \frac{x}{2} e^{-|x|}\, dx$$

Let $\frac{x}{2} e^{-|x|} = f(x)$

Then $m_x = \int_{-\infty}^{\infty} f(x)\, dx$

But $f(x)$ is an odd function since

$$f(-x) = -\frac{x}{2} e^{-|x|} = -f(x)$$

Therefore $\int_{-\infty}^{\infty} f(x)\, dx = 0$

Thus $m_x = 0$ **Ans.**

Example 5.23. The probability density function (PDF) of a continuous random variable X in the range (-3, 3) is defined as follow:

$$f_X(x) = \begin{cases} \frac{1}{16}(3+x)^2 & \text{for } -3 \le x \le -1 \\ \frac{1}{16}(2-6x)^2 & \text{for } -1 \le x \le 1 \\ \frac{1}{16}(3-x)^2 & \text{for } 1 \le x \le 3 \end{cases}$$

Verify that the area under the curve is unity. Also prove that the mean is zero.

(A.M.I.E., Examination, 1998)

Solution: We know that the area under the PDF curve is expressed as

$$\text{area} = \int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-3}^{3} f_X(x)\, dx$$

$$\text{area} = \int_{-3}^{-1} f_X(x)\, dx \int_{-1}^{1} f_X(x)\, dx \int_{1}^{3} f_X(x)\, dx$$

or

$$\text{area} = \int_{-3}^{-1} \frac{1}{16}(3+x)^2\, dx + \int_{-1}^{1} \frac{1}{16}(2-6x^2)\, dx + \int_{1}^{3} \frac{1}{16}(3-x)^2\, dx$$

$$\text{area} = \frac{1}{48}[8+32+8] = \frac{1}{48} \times 48 = 1$$

Therefore, the area under the PDF curve is unity.

We know that for a continuous random variable, the mean is expressed as

$$m_x = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

where $f_X(x)$ = probability density function (PDF)

Now

$$m_x = \int_{-3}^{3} x \cdot f_X(x)\, dx = \int_{-3}^{-1} \frac{1}{16} \cdot x(3+x)^2\, dx + \int_{-1}^{1} \frac{1}{16} x(6-2x^2)\, dx + \int_{1}^{3} \frac{1}{16} x(3-x)^2\, dx$$

$$m_x = \frac{1}{16}\int_{-3}^{-1} (9x+6x^2+x^3)\, dx + \frac{1}{16}\int_{-1}^{1} (6x-2x^3)\, dx + \frac{1}{16}\int_{1}^{3} (9x-6x^2+x^3)\, dx = 0$$

Hence, mean value is zero. **Ans.**

Example 5.24. A random variable X has the uniform distribution given by

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & \text{for } 0 \le x \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$

(BPTU, Orissa-2...)

Determine m_x, $\overline{X^2}$ and σ_x.

Solution: We know that the mean value of a continuous random variable is expressed as

$$m_x = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{0}^{2\pi} x \cdot \frac{1}{2\pi}\, dx$$

$$m_x = \frac{1}{2\pi}\int_{0}^{2\pi} x\, dx = \frac{1}{2\pi}\left[\frac{x^2}{2}\right]_0^{2\pi} = \frac{1}{4\pi}\left[x^2\right]_0^{2\pi} = \frac{1}{4\pi}[4\pi^2 - 0] = \pi \quad \textbf{Ans.}$$

We know that the mean square value of a continuous variable is given as

$$\overline{X^2} = E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \int_{0}^{2\pi} x^2 \frac{1}{2\pi}\, dx = \frac{1}{2\pi}\int_{0}^{2\pi} x^2\, dx$$

or

$$\overline{X^2} = \frac{1}{2\pi}\left[\frac{x^3}{3}\right]_0^{2\pi} = \frac{1}{6\pi}\left[x^3\right]_0^{2\pi} = \frac{1}{6\pi}[8\pi^3 - 0] = \frac{1}{6\pi} \cdot 8\pi^3 = \frac{4}{3}\pi^2 \quad \textbf{Ans.}$$

Variance of a continuous random variable is expressed as

$$\sigma_x^2 = E[X^2] - m_x^2 = \frac{4}{3}\pi^2 - \pi^2 = \frac{1}{3}\pi^2$$

Also the standard deviation σ_x is given as the square root of variance *i.e.*,

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{3}\pi^2} = \frac{\pi}{\sqrt{3}} \quad \text{Ans.}$$

Example 5.25. The probability density function (PDF) of a random variable X is given as $f_X(x)$. A random variable Y is defined as $Y = aX + b$ where $a < 0$. Obtain PDF of Y in terms of the PDF of X.

Solution: We know that the Probability Density Function (PDF) may be expressed as

$$P(x < X \le x + dx) = f_X(x)\, dx$$

Similarly for random variable Y

$$P(y < Y \le y + dy) = f_Y(y)\, dy$$

Because the probabilities of random variables X and Y are same, we may equate the above two equations. Thus we get

$$f_Y(y)\, dy = f_X(x)\, dx = f_X(x) \cdot \frac{dx}{dy} \quad ...(i)$$

Also given that $\quad y = ax + b$

thus $\quad \frac{dy}{dx} = a = \frac{1}{a} \quad$ and $\quad x = \frac{y-b}{a}$

Putting all these values in equation (*i*), we get

$$f_Y(y) = \frac{1}{a} fX\left(\frac{y-b}{a}\right)$$

This is the required expression of PDF of Y in terms of PDF of X.

5.16. Uniform Distribution

The Probability Density Function (PDF) of an uniform distribution is expressed as

Uniform PDF: $f_X(x) = 0 \quad$ for $x < m - \frac{A}{2}$ & $x > m + \frac{A}{2}$

$$= \frac{1}{A} \quad \text{for } \left(m - \frac{A}{2}\right) \le x \le \left(m + \frac{A}{2}\right) \quad ...(5.84)$$

Figure 5.3 shows the PDF of an uniformly distributed random variable.

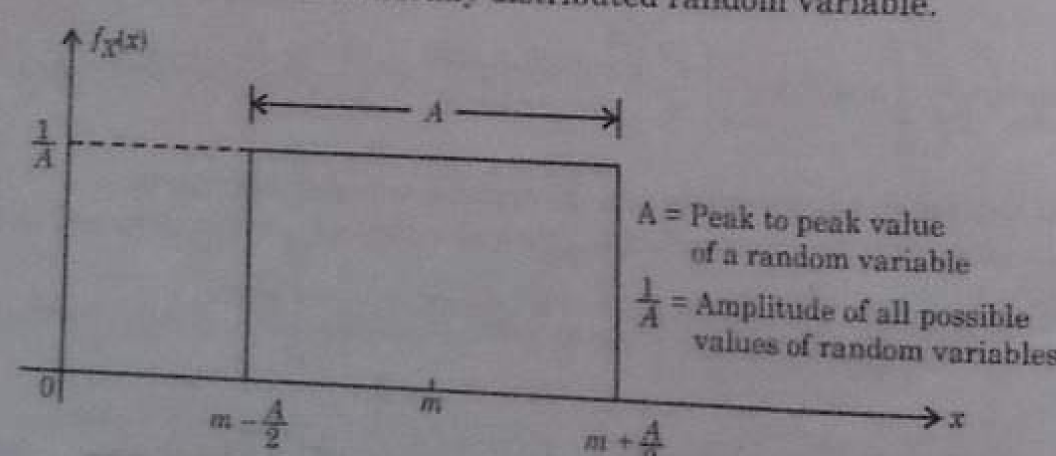


Fig. 5.3. *PDF of an uniformly distributed random variable*

In above figure, the peak to peak value is A and amplitude is uniform *i.e.* A. From figure 5.3, it is also clear that the random variable and its PDF are continuous. Therefore uniform distribution is utilised for continuous random variable.

Also, the value of PDF, $f_X(x)$ is same for all possible values of a random variable. Thus this distribution is known as **Uniform Distribution**.

Example 5.26. Prove that the mean and variance of a random variable X having an unif distribution in the inerval $[a, b]$ are

$$m_x = \frac{a+b}{2} \text{ and } \sigma_x^2 = \frac{(a-b)^2}{12}$$

Solution: An uniform distribution in the interval $[a, b]$ is sketched in the figure (5.4).

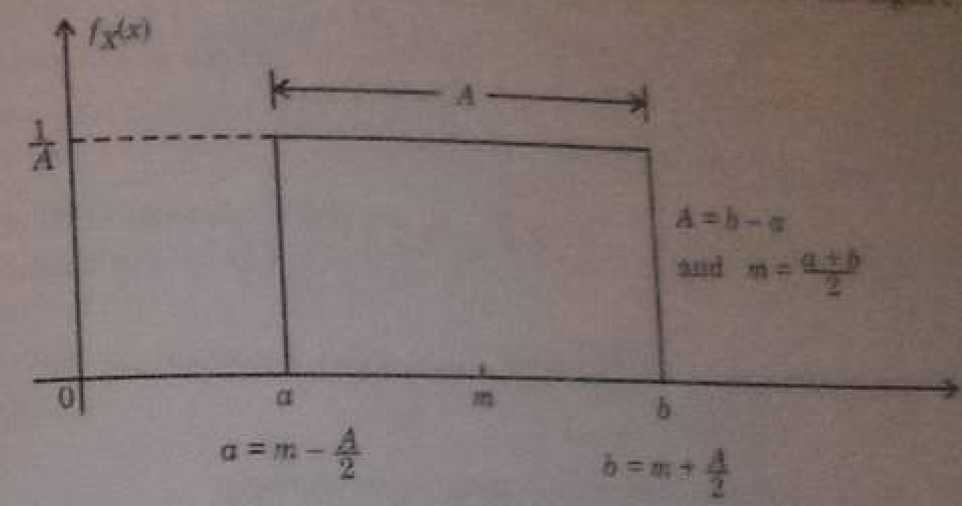


Fig. 5.4. *Uniform distribution having interval [a, b].*

We know that the mean value of a continuous random variable is expressed as

$$m_x = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_a^b x \cdot \frac{1}{A}\, dx$$

From figure 5.4 it may be observed that the value of A will be the difference between up lower limits of interval *i.e.*,

$$A = b - a$$

With this value of A, the mean value will become

$$m_x = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{2(b-a)}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{2(b-a)}\left[x^2\right]_a^b$$

or

$$m_x = \frac{1}{2(b-a)}[b^2 - a^2] = \frac{1}{2(b-a)} \cdot (b-a)(b+a) = \frac{b+a}{2} = \frac{a+b}{2}$$

Also, the variance is expressed as

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 f_X(x)\, dx = \int_a^b (x - m_x)^2 \frac{1}{b-a}\, dx \quad \left[\because f_X(x) = \frac{1}{b-a}\right]$$

Now, let $x - m_x = y$, then $dx = dy$ and limits will become

if $\quad x = a$, then $y = a - m_x$

and if $\quad x = b$, then $y = b - m_x$

So that,

$$\sigma_x^2 = \int_{a-m_x}^{b-m_x} y^2 \cdot \frac{1}{b-a}\, dy = \frac{1}{b-a}\left[\frac{y^3}{3}\right]_{a-m_x}^{b-m_x}$$

or

$$\sigma_x^2 = \frac{1}{3(b-a)}[(b-m_x)^3 - (a-m_x)^3]$$

Putting the value of $m_x = \frac{a+b}{2}$ in above equation, we get

$$\sigma_x^2 = \frac{1}{3(b-a)}\left[\left(b-\frac{a+b}{2}\right)^3-\left(a-\frac{a+b}{2}\right)^3\right]$$

or

$$\sigma_x^2 = \frac{1}{3(b-a)}\left[\left(\frac{b-a}{2}\right)^3-\left(\frac{a-b}{2}\right)^3\right]$$

writing $(b-a) = -(a-b)$, we have

$$\sigma_x^2 = -\frac{1}{3(a-b)}\left[\left(-\frac{a-b}{2}\right)^3-\left(\frac{a-b}{2}\right)^3\right]$$

$$\sigma_x^2 = -\frac{1}{3(a-b)}\cdot-\frac{(a-b)^3}{4} = \frac{(a-b)^2}{12} \quad ...(ii)$$

Because $(a-b)^2 = (b-a)^2$, the above equation may be written as

$$\sigma_x^2 = \frac{(a-b)^2}{12}$$

But $A = b - a.$

Therefore, $\sigma_x^2 = \frac{A^2}{12}$ ...(iii)

From figure 5.4, we have

$$a = m - \frac{A}{2}$$

and

$$b = m + \frac{A}{2}$$

thus

$$a + b = m - \frac{A}{2} + m + \frac{A}{2} = 2m$$

or

$$\frac{a+b}{2} = m$$

Putting this value in equation (i), we get
Mean value of uniform distribution

$$m_x = \frac{a+b}{2} = m$$

Using equations (ii) and (iii), we get
Variance of Uniform Distribution

$$\sigma_x^2 = \frac{(a-b)^2}{12} = \frac{(b-a)^2}{12} = \frac{A^2}{12} \quad \textbf{Ans.}$$

5.17. Gaussian or Normal Distribution

Gaussian Distribution is also known as Normal Distribution. It is defined for a continuous random variable. The PDF for a Gaussian random variable is expressed as

Gaussian PDF

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-m)^2/2\sigma^2} \quad ...(5.85)$$

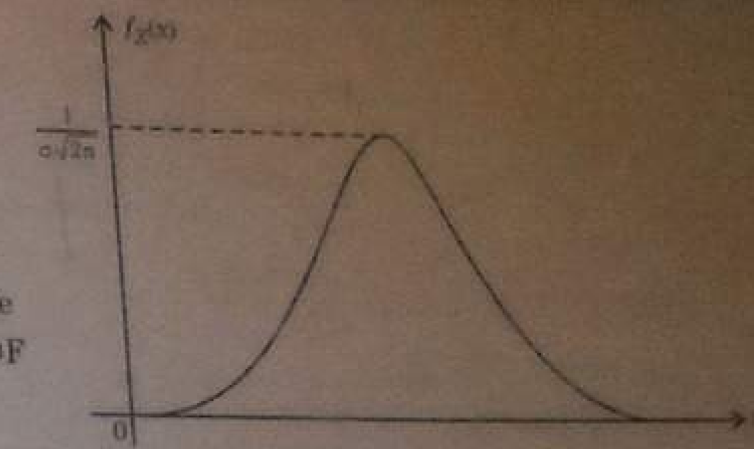


Fig. 5.5. *Plot of Gaussian PDF*

In this equation

m = mean value of the random variable

σ^2 = variance of the random variable

Figure 5.5 shows the plot of Caussian PDF $f_X(x)$ with respect to x.

5.17.1. Properties of Gaussian PDF

The several properties of Gaussian PDF are as under:

Property 1: The peak value occurs at $x = m$ *i.e.* mean value.

Mathematically,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{ at } x = m \text{ (mean value)} \quad ...(5.86)$$

This value may be obtained by putting $x = m$ in equation (5.85).

Property 2: The plot of Gaussian *PDF* exhibit even symmetry around mean value.

Mathematically,

$$f_X(m-\sigma) = f_X(m+\sigma) \quad ...(5.87)$$

Property 3: The area under the PDF curve is 1/2 for all values of x below mean value and 1/2 for all values of x above mean value. Mathematically,

$$P(X \le m) = P(X > m) = \frac{1}{2} \quad ...(5.88)$$

Property 4: As $\sigma \to 0$, the Gaussian function approaches to δ (impulse) function located at $x = m$. This is because the area under the PDF curve is always one (unity). Also the area of impulse function is also unity.

Importance

The Gaussian distribution is used for continuous random variables. We know that the random motion of the thermaly agitated electrons produces thermal noise. This thermal noise has Gaussian distribution. The random errors in the experimental measurements create the measured values to have Gaussian distribution about the true value. The Gaussian distribution is very important in the analysis of communication and statistical systems.

Example 5.27. Find out the Cumulative Distribution Function (CDF) of the Gaussian Random variable.

Solution: We know that the Cumulative Distribution Function (CDF) may be obtained by integrating *PDF* as

...(i)

CDF:
$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$$

But Gaussian PDF is expressed as

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-m)^2/2\sigma^2}$$

Putting this value of $f_X(x)$ in equation (i)

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}\, dx \quad ...(ii)$$

Let $\frac{m-x}{\sigma\sqrt{2}} = z$ in the above equation

then $-\frac{dx}{\sigma\sqrt{2}} = dz$

or $dx = -\sigma\sqrt{2}\, dz$

The limits will be

when $x \to -\infty$; then $z \to +\infty$

when $x \to x$; then $z \to \frac{m-x}{\sigma\sqrt{2}}$

Putting all these values in equation (ii)

$$F_X(x) = \int_{\infty}^{\frac{m-x}{\sigma\sqrt{2}}} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-z^2}\left(-\sigma\sqrt{2}\, dz\right) = -\frac{1}{\sqrt{\pi}} \int_{\infty}^{\frac{m-x}{\sigma\sqrt{2}}} e^{-z^2}\, dz$$

Interchanging the limits, we get

$$= \frac{1}{\sqrt{\pi}} \int_{\frac{m-x}{\sigma\sqrt{2}}}^{\infty} e^{-z^2}\, dz$$

Rearranging, we get

$$F_X(x) = \frac{1}{2} \cdot \frac{2}{\sqrt{\pi}} \int_{\frac{m-x}{\sigma\sqrt{2}}}^{\infty} e^{-z^2}\, dz \quad ...(iii)$$

We know that the error function is given by

$$\text{erfc}\,(u) = \frac{2}{\sqrt{\pi}} \int_{u}^{\infty} e^{-z^2}\, dz \quad ...(iv)$$

Here $u = \frac{m-x}{\sigma\sqrt{2}}$

Using equations (iii) and (iv), we get

Gaussian *CDF*: $F_X(x) = \frac{1}{2}\,\text{erfc}\left(\frac{m-x}{\sigma\sqrt{2}}\right)$

5.18. Rayleigh Distribution

Rayleigh Distribution is used for continuous random variables. It is produced from two Gaussian random variables. Let X and Y be independent Gaussian random variables having mean value zero and variance σ^2. Mathematically,

$$m_x = m_y = 0 \quad ...(5.89)$$

and $$\sigma_x^2 = \sigma^2 \quad ...(5.90)$$

Figure 5.6 shows the rectangular to polar conversion.

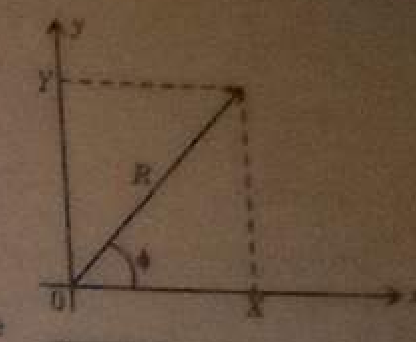


Fig. 5.6. *Rectangular to polar conversion*

From figure 5.6, it may be observed that independent Gaussian random variables X and Y are related to Rayleigh Distribution random variables R and ϕ such that

$$R = \sqrt{X^2 + Y^2} \quad ...(5.91)$$

and $$\phi = \tan^{-1}\left(\frac{Y}{X}\right) \quad ...(5.92)$$

From these equations, it is evident that R will always be positive because the variables X and Y are squared.

Also, the range of ϕ will be form 0 to 2π because it repeats after 2π radians. This means that ϕ completes one circle in figure 5.6 when it goes from 0 to 2π. After transformation of Gaussian random variables X and Y into R, it is found that R is having Rayleigh Probability Density Function (PDF) which is expressed as

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \quad ...(5.93)$$

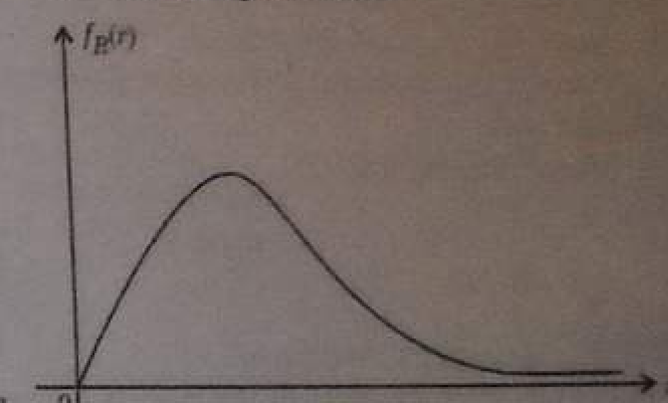


Fig. 5.7.

Figure 5.7 shows the Rayleigh PDF defined by above equation.

Therefore, from *PDF* curve, it is clear that

$$f_R(r) = 0 \quad \text{for } r < 0 \quad ...(5.94)$$

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \quad \text{for } r \geq 0 \quad ...(5.95)$$

This means that R has always positive value and never goes for negative value.

Importance

Rayleigh Distribution is always used for modelling of statistics of signals transmitted through radio channels such as cellular radio.

Example 5.28. If Probability distribution function (PDF) of Rayleigh distribution is given by the expression

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}$$

then above that the cumulative distribution function (CDF) of Rayleigh Distribution will be

$$F_R(r) = P(R \leq r) = \left(1 - e^{-r^2/2\sigma^2}\right) u(r)$$

Solution: We know that *CDF* may be obtained from PDF as

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx$$

where $$F_R(r) = \int_{-\infty}^{x} f_R(r)\, dr$$

Now since r never becomes negative, therefore integration limits will be from 0 to r. ...(i)

Thus, $$F_R(r) = \int_{0}^{r} \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}\, dr$$

Putting $\frac{r^2}{2\sigma^2} = t$

So that $\frac{2r\,dr}{2\sigma^2} = dt$ or $\frac{r\,dr}{\sigma^2} = dt$

when $r = 0$ then $t = 0$

and when $r = r$ then $t = \frac{r^2}{2\sigma^2}$

Putting all these values in equation (i), we get

$$F_R(r) = \int_0^{r^2/2\sigma^2} e^{-t}\,dt = \left[-e^{-t}\right]_0^{r^2/2\sigma^2}$$

$$= \left[-e^{-r^2/2\sigma^2} + e^0\right] = \left[-e^{-r^2/2\sigma^2} + 1\right]$$

or $F_R(r) = 1 - e^{-r^2/2\sigma^2}$

Since we know that $r \geq 0$ always, therefore the above equation may be multiplied by unit step function $u(r)$.

Hence Rayleight CDF:

$$F_R(r) = \left(1 - e^{-r^2/2\sigma^2}\right) u(r)$$

5.19. Random Process

Let there be a random experiment E having outcome λ from the sample sapce S. This means that $\lambda \in S$. Thus every-time an experiment is conducted, the outcome λ will be one of the sample point in sample space. If this outcome λ is associated with time, then a function of λ and time t is formed *i.e.*, $X(\lambda, t)$. Then the function $X(\lambda, t)$ is known as **random process**. Hence, when any random experiment E is given a time dimension, then each outcome appears at some certain time and the random experiment will be converted to Random Process. A random process is the function of two variables λ and t.

Here λ belongs to sample sapce S and $-\infty < t < \infty$. At some specific time say $t = t_0$, $X(\lambda, t_0)$ is a random variable whose value depends upon λ. Also, for a specific outcome λ_i, there is a single time function $X(\lambda_i, t_0)$. This function is known as sample function whereas the collection of all the sample functions is known an **ensemble**.

We can represent random process as $X(t)$ *i.e.* here λ is dropped.

Therefore, a random process $X(t)$ may represent several functions of t and λ, a single time function, a random variable or a single number.

In practical life, there are various physical phenomenon which may be represented with the help of random process.

As an example, let us consider a set of voltage waveforms generated by thermal electron motion in a large number of identical resistors.

We can represent the voltage produced at each instant by the resistor as v. Then the random process $X(v, t)$ denotes the output voltages from the set of resistors.

Figure 5.8 shows the waveforms produced by these resistors randomly and different parameters of random process.

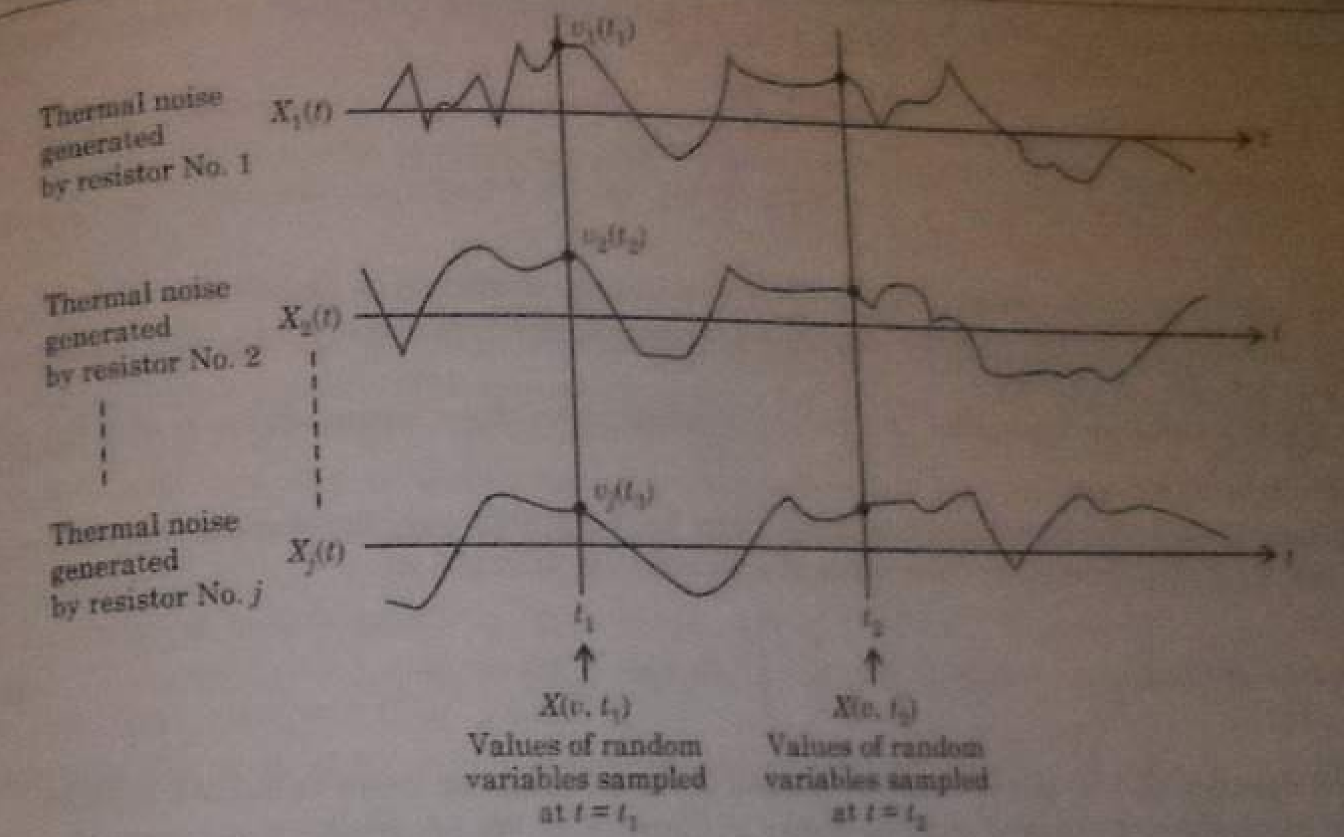


Fig. 5.8.

5.19.1. Ensemble and Time Averages

We know that the random variables may be specified with the help of different probabili distributions. Random proceesses cannot be completely specified with the help of PDFs. General ensemble and time statistics are used to specify the random processes. Most generally, statisti averages such as mean $m_x(t)$ and autocorrelation function $R_x(t_1, t_2)$ are used to describe rand process.

5.19.2. Ensemble Averages

In case of ensemble averages, the average is taken over the ensemble of waveforms, keeping time constant or fixed.

As an example, if we desire ensemble mean value in random process of figure 5.8, then we have to take mean of $v_1(t_1)$, $v_2(t_1)$, $v_3(t_1)$ ..., $v_j(t_1)$.

This ensemble mean is taken at time $t = t_1$.

In the same way, we can take ensemble mean at time $t = t_2$ also. The ensemble mean at $t = t_1$ $t = t_2$ need not be same. Similarly, ensemble mean at other values of t may also be evaluated. T ensemble mean becomes function of time t. Therefore it is represented as $m_x(t)$.

We know that the mean value for a continuous random variable is expressed as

$$m_x = \int_{-\infty}^{\infty} x f_X(x)\,dx \quad ...(5$$

Using this expression, the ensemble mean may be defined as ...(5

$$m_x(t) = E[X(v, t)]$$

Hence, Ensemble mean value ...(5

$$m_x(t) = \int_{-\infty}^{\infty} x f_X(x, t)\,dx$$

In this equation, the time t is treated as constant and the integration is done with respect

The next ensemble average is the autocorrelation function which may be expressed as

$$R_\tau(t_1, t_2) = E[X(v, t_1)\ X(v, t_2)] \quad ...(5.99)$$

$$= E[X(t_1)\ X(t_2)]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{x_1} f_{x_2}(x_1, x_2)\, dx_1\, dx_2 \quad ...(5.100)$$

In this equation, the value of $R_\tau(t_1, t_2)$ represents the similarity of amplitudes at time t_1 and t_2 and is obtained by multiplying the amplitudes of sample functions at t_1 and t_2 and taking mean of this product over an ensemble.

Thus the autocorrelation function gives the information about frequency content of the process.

5.19.3. Time Averages

When the statistical averages are taken along the time; they are known as time averages. As an example, we may define time mean value of a sample function $x(t)$ as

Time mean value:

$$\langle m_x \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t)\, dt \quad ...(5.101)$$

The equation (5.101) is a standard equation to find average or mean value of a function.

Let us consider the random process shown in figure (5.8). We can obtain time mean of waveform $X_1(t)$ by integrating it over a long period. The time mean of $X_1(t)$ and $X_2(t)$ need not be same. Thus the collection of time means becomes a function of the sample function $X_i(t)$.

The autocorrelation function may be expressed using time averaging as

$$\langle R_\tau(\varepsilon) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cdot x(t + \tau)\, dt \quad ...(5.102)$$

5.19.4. Stationary and Non-Stationary Random Process

A random process $X(t)$ is called stationary if its statistics are not affected by any shift in the time origin. We may define stationary process in terms of ensemble averages as:

(*i*) The ensemble mean is independent of time. Mathematically,

$$m_x(t) = m_x(t_1) = m_x(t_2) = m_x(t_3) = ... = \text{constant at all the time instants} \quad ...(5.103)$$

(*ii*) The autocorrelation function $R_\tau(t_1, t_2)$ depends only upon the time difference $t_2 - t_1$.

$$R_\tau(t_1, t_2) = R_\tau(t_1 + t, t_2 + t) \quad ...(5.104)$$

Therefore, the autocorrelation function of any stationary process is a function of time-difference and is given as

$$R_\tau(t_1, t_2) = R_\tau(t_2 - t_1) \quad ...(5.105)$$

$$= R_\tau(\rho)$$

where $\rho = t_2 - t_1$

Thus the system of resistors producing thermal noise voltages represents a stationary process due to the fact that the noise generated by resistor does not depend on time.

5.19.5. Wide Sense Stationary (*i.e.*, Weakly Stationary) Process

The process may not be stationary in strict sense, still the mean and autocorrelation functions are independent of shift of time region. Such type of process is known as **wide sense stationary process.**

All the processes in practice, are non-stationary since every process has some start and end. This means that the statistics of such process are dependent upon time. A true stationary process should start at $t = -\infty$ and should not stop till $t = \infty$. Such type of process is not possible practically.

The process may appear stationary over a certain period of time. Then this will be wide sense stationary process.

5.19.6. Ergodic Process

A random process is known as ergodic process if the time-averages are equal to ensemble averages. Hence for a ergodic process, we have

$$m_x = \langle m_x \rangle \quad ...(5.106)$$

$$R_\tau(t_1, t_2) = \langle R_\tau(\varepsilon) \rangle \quad ...(5.107)$$

For any ergodic process, all ensemble averages will be equal to corresponding time averages for any particular sample function. The time averages are not a function of time. When time and ensemble averages are the same, it implies that ensemble averages also are not a function of time.

Thus an ergodic process is always stationary but converse in not true.

Ergodicity of the process may be defined in terms of some statistical averages like mean and autocorrelation.

The random process is ergodic in the mean if

$$m_x = \langle m_x \rangle$$

and variance of $\langle m_x \rangle \to 0$ as $T \to \infty$

Similarly, the random process is ergodic in the autocorrelation if

$$R_\tau(t_1, t_2) = \langle R_\tau(\varepsilon) \rangle \quad ...(5.108)$$

and variance of $\langle R_\tau(\varepsilon) \rangle \to 0$ as $T \to \infty$

Hence, the Ergodicity of any random process may be determined by evaluating statistical averages of single sample function. This means that a single sample function represents entire random process.

5.19.7. Gaussion Process

Let us consider a random process denoted by $X(t)$ for an interval which starts at a time $t = 0$ and lasts until $t = T$.

Now, let us weigh the random process $X(t)$ by any function $g(t)$ and then integrate the product $X(t)\, g(t)$ over this observation time-interval $(0 \le t \le T)$.

Thus, Gassuion we obtain a random variable Y defined as

$$Y = \int_0^T X(t)\, g(t)\, dt \quad ...(5.109)$$

where Y is a linear function of $X(t)$.

Also, the random variable Y depends upon the argument function $X(t)\, g(t)$ over the observation time interval 0 to T. Here, we assume that the weighing function $g(t)$ is such that the mean square value of the random variable

$$Y = \int_0^T X(t)\, g(t)\, dt \quad \text{is finite.} \quad ...(5.110)$$

Now, if the random variable Y is a Gaussian distributed random variable for each value of function $g(t)$, then the process $X(t)$ is known as a **Gaussian process.** Also, a random variable Y has a Gaussian distribution if its probability density function has the following form:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{\{[-(y - m_Y)^2]/2\sigma_Y\}^2} \quad ...(5.111)$$

Here, m_Y is the mean and σ_Y^2 is the variance of the random variable Y. A Gaussian distributed random variable is completely characterized by spectifying its mean and the variance.

Figure 5.9. exhibits a plot of Gaussian distributed probability density function (pdf) for $m_Y = 0$ and $\sigma_Y^2 = 1$.

It may be noted that a Gaussian process has two main features or advantages:

(i) Firstly, the Gaussian process has many properties which make analytic results possible.

(ii) Secondly, the random process produced by physical phenomena are often such that a Gaussian model is an appropriate one.

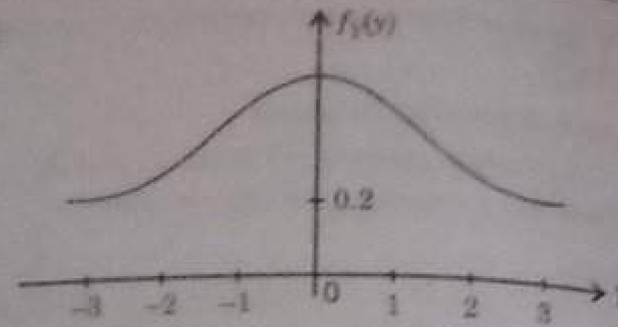


Fig. 5.9. *Normalized Gaussion distributed pdf of a random variably.*

Also the use of a Gaussian model to describe physical phenomena is generally confirmed by the experiments. Thus, the Gaussian process is very important in the study of communication systems.

A Gaussian process has the following important properties:

(i) If a Gaussian process $X(t)$ is made to apply to a stable linear filter, then the random process $Y(t)$ produced at the output of the filter is also Gaussian.

(ii) Let us consider the samples of random process $X(t_1), X(t_2), \ldots, X(t_n)$, obtained by observing a random process $X(t)$ at time $t = t_1, t = t_2, \ldots, t = t_n$.

Now, if the random process $X(t)$ is Gaussian, then this set of random variables are jointly Gaussian for any value of n, with their n-fold joint *pdf* being completely deterimined by specifying the set of means.

Here, means used for more than one mean (or average), *i.e.*

$$m_X(t_i) = E[X(t_i)],\ i = 1, 2, \ldots, n$$

and the set of autocorrelation function

$$C_X(t_j, t_i) = E[\{X(t_j) - m_X(t_j)\}\{X(t_i) - m_X(t_i)\}] \qquad \ldots(5.112)$$

for $\qquad i = j = 1, 2, 3, \ldots, n$

(iii) Also, if a Gaussian process is wide-sense stationary (WSS), then the process is also stationary in the strict sense.

(iv) If the set of random variables $X(t_1), X(t_2), \ldots, X(t_n)$, obtained by sampling a Gaussian process $X(t)$ at times $t = t_1, t = t_2, \ldots, t = t_n$, are not correlated *i.e.*,

$$E[\{X(t_j) - m_X(t_j)\}\{X(t_i) - m_X(t_i)\}] = 0 \qquad \ldots(5.113)$$

for $i \neq j$

then this set of random variables are statistically independent.

5.20. Sum of Random Processes

Now, we shall discuss the sum of two random processes.

Let us consider two WSS processes $X(t)$ and $Y(t)$ with zero means. The sum of random processes another random process denoted as $Z(t)$. Also, the statistical parameters of $Z(t)$ may be determined in terms of random processes $X(t)$ and $Y(t)$.

Thus, the random process $Z(t)$ may be expressed as

$$Z(t) = X(t) + Y(t) \qquad \ldots(5.114)$$

and the autocorrelation function of $Z(t)$ be found as

$$R_Z(\tau) = E[Z(t)\, Z(t-\tau)]$$

$$= E[\{X(t) + Y(t)\}\{X(t-\tau)\, Y(t-\tau)\}]$$

$$= E[X(t)\, X(t-\tau) + X(t)\, Y(t-\tau) + X(t-\tau)\, Y(t) + Y(t)\, Y(t-\tau)]$$

or $\quad R_Z(\tau) = E[X(t)\, X(t-\tau)] + E[X(t)\, Y(t-\tau)] + E[X(t-\tau)\, Y(t)] + E[Y(t)\, Y(t-\tau)]$

or $\quad R_Z(\tau) = R_X(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_Y(\tau) \qquad \ldots(5.115)$

The power spectral density (psd) of random process $Z(t)$ may be obtained by taking Fourier transform of both sides of equation (5.115)

Hence, we have $\quad S_Z(\omega) = S_X(\omega) + S_{XY}(\omega) + S_{YX}(\omega) + S_Y(\omega) \qquad \ldots(5.116)$

The power spectral density (psd) of the sum of two WSS random processes is equal to the sum of the individual power spectral densities (psds) plus the sum of the cross power spectral densities (psds) $S_{XY}(\omega)$ and $S_{YX}(\omega)$.

If random processes $X(t)$ and $Y(t)$ are non-correlated then the cross power spectral densities $S_{XY}(\omega)$ and $S_{YX}(\omega)$ are zero. Now equation (5.116) can be written as

$$S_Z(\omega) = S_X(\omega) + S_Y(\omega)$$

Hence, the power spectral density (psd) of the sum of two uncorrelated WSS random processes is equal to the sum of their individual power spectral densities.

5.21. Correlation Function

The correlation function provides a measure of similarity or coherence between a given signal (process) and a replica of the same signal or other signal (process) by a variable amount.

5.21.1. Autocorrelation Function

Autocorrelation function may be defined as a measure of similarity between a signal or process and its replica by a variable amount.

The autocorrelation function of a stationary process $X(t)$ may be defined as

$$R_{X(t_j - t_i)} = E[X(t_j)\, X(t_i)] \text{ for any } t_j \text{ and } t_i \qquad \ldots(5.117)$$

where $X(t_j)$ and $X(t_i)$ are the random variables obtained by observing the process $X(t)$ at times t_j and t_i respectively.

Also, the autocorrelation function depends only on the time difference $(t_j - t_i)$.

Using $\tau = t_j - t_i$ in equation (5.117), we get

$$R_X(\tau) = E[X(t)\, X(t-\tau)] \qquad \ldots(5.118)$$

The expressions for $X(t)$ and $X(t-\tau)$ are viewed as random variables. The variable τ is known as **time-lag or time-delay parameter**. The autocorrelation function for stationary random process is independent of a shift of time origin.

5.21.2. Properties of Autocorrelation Function of Random Process

The autocorrelation function $R_X(\tau)$ of a wide-sense stationary (WSS) process has the following important properties:

(i) The autocorrelation function of WSS process $X(t)$ is always an even function of time-lag. *i.e.* Autocorrelation function $R_X(\tau)$ satistifies the following mathematical relationship, i.e.

$$R_X(\tau) = R_X(-\tau) \qquad \ldots(5.119)$$

Proof: According to the definition of autocorrelation function of WSS random process, we have

$$R_X(\tau) = E[X(t)\, X(t-\tau)] \qquad \ldots(5.120)$$

Substituting $\tau = -\tau$, in equation (5.120), we get

$$R_X(-\tau) = E[X(t)\, X(t+\tau)] \qquad \ldots(5.121)$$

or $\quad R_X(-\tau) = E[X(t+\tau)\, X(t)] \qquad \ldots(5.122)$

For WSS process, we know that

$$X(t)\, X(t-\tau) = X(t+\tau)\, X(t)$$

From last three equations, we conclude that

$$R_X(\tau) = R_X(-\tau)$$

(ii) The mean square value of a WSS process is equal to the autocorrelation function of the random process for zero time-lag. It may be written as

$$R_X(0) = R_X(\tau)\big|_{\tau=0} = E[X(t)\, X(t-\tau)]\big|_{\tau=0}$$

or $$R_X(0) = E[X(t)\, X(t-0)] = E[X^2(t)]$$

or $$R_X(0) = E[X^2(t)] \quad ...(5.123)$$

(iii) The autocorrelation function of a WSS random process has the maximum magnitude at zero time-lag ($\tau = 0$). This property states that

$$|R_X(\tau)| = \text{magnitude of } R_X(\tau) = R_X(0) \quad ...(5.124)$$

Proof: We know that the mean square value of the difference between $X(t)$ and $X(t-\tau)$ is always non-negative, *i.e.*

$$E[\{X(t) - X(t-\tau)\}^2] \geq 0$$

or $$E[X^2(t) - 2X(t)\, X(t-\tau) + X^2(t-\tau)] \geq 0 \quad ...(5.125)$$

Since expectation operation is a linear operation, we can write above equation as

$$E[X^2(t)] - 2E[X(t)\, X(t-\tau)] + E[X^2(t-\tau)] \geq 0 \quad ...(5.126)$$

Also, for a WSS random process, we have

$$E[X^2(t)] = E[X^2(t-\tau)] = R_X(0) \quad ...(5.127)$$

and $$E[X(t)\, X(t-\tau)] = R_X(\tau) \quad ...(5.128)$$

Substituting equations (5.131) and (5.127) in equation (5.128), we get

$$|R_X(\tau)| \leq R_X(0) \quad ...(5.129)$$

Thus, the autocorrelation functions $R_X(\tau)$ provide means of decreasing the interdependance of two random variables obtained by observing a random process $X(t)$ τ seconds apart. The autocorrelation function is a measure of the rate of the fluctuation of random process. Various types of random processes have been illustrated in figure 5.10.

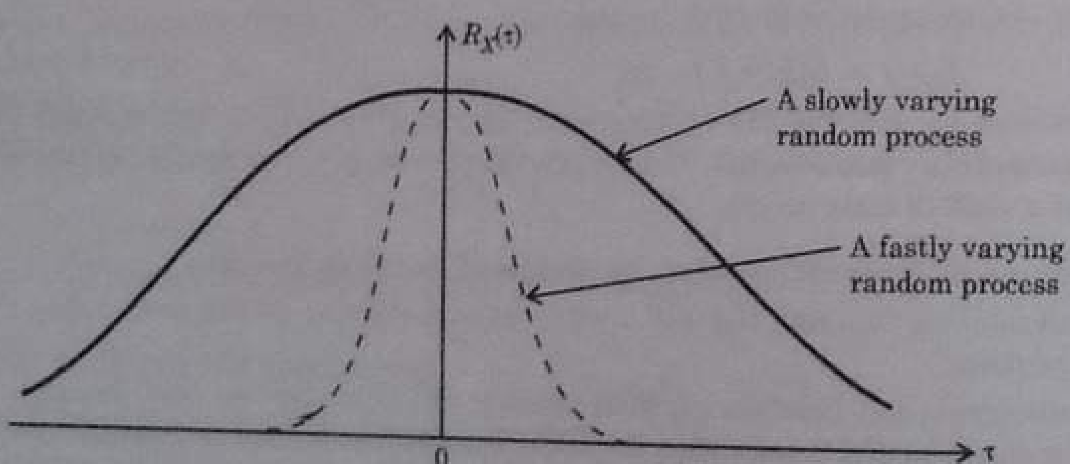


Fig. 5.10. *Illustration of autocorrelation functions of slowly and rapidly varying random processes.*

5.21.3. Cross-Correlation Functions

An autocorrelation function is determined for single random process but cross correlation function is determined for two random processes. Let us consider two random processes $X(t)$ and $X(t)$ with autocorrelation functions $R_X(t, u)$ and $R_Y(t, u)$, respectively. There will be two cross-correlation functions between two random process $X(t)$ and $Y(t)$. These cross-correlations may be defined as

$$R_{XY}(t, u) = E[X(t)\, Y(u)] \quad ...(5.130)$$

and $$R_{YX}(t, u) = E[Y(t)\, X(u)] \quad ...(5.131)$$

where t and u are the two values of time at which the processes are observed. The cross-correlation function of two random processes $X(t)$ and $Y(t)$ may be expressed conveniently in matrix form as

$$R(t, u) = \begin{bmatrix} R_X(t,u) & R_{XY}(t,u) \\ R_{YX}(t,u) & R_Y(t,u) \end{bmatrix} \quad ...(5.132)$$

$R(t, u)$ is called the correlation matrix of random processes $X(t)$ and $Y(t)$.

Equation (5.137) may also be written as

$$R(t, u) = \begin{bmatrix} R_X(t-u) & R_{XY}(t-u) \\ R_{YX}(t-u) & R_Y(t-u) \end{bmatrix} \quad ...(5.133)$$

If both random processes $X(t)$ and $Y(t)$ are WSS processes, then equation (5.132) can be written as equation (5.133). In addition, both random processes $X(t)$ and $Y(t)$ are said to be jointly wide-sense stationary.

It may be noted that in general, the cross-correlation function is not an even function of τ like an autocorrelation function. It is having following symmetry relation

$$R_{XY}(\tau) = R_{YX}(-\tau) \quad ...(5.134)$$

Moreover, cross-correlation does not have a maximum at the origin like the autocorrelation function. The two random processes $X(t)$ and $Y(t)$ are said to be incoherent or orthogonal if cross-correlation function of $X(t)$ and $Y(t)$ is zero.

$$R_{XY}(\tau) = 0 \quad ...(5.135)$$

The two random processes are said to be non-correlated, if their cross-correlation functions $R_{XY}(\tau)$ are equal to the multiplication of mean values.

$$R_{XY}(\tau) = E[X(t)]\, E[Y(t)] \quad ...(5.136)$$

The incoherent or orthogonal processes are non-correlated processes with $E[X(t)]$ and or $E[Y(t)] = 0$.

5.22. Spectral Densities

As a matter of fact, spectral densities are used to represent random process in frequency-domain. In this article, we shall discuss frequency-domain description of random processes. Frequency-domain methods are very powerful tools for analysis of various types of signals. We have already discussed some powerful tools such as Fourier series, Fourier transform, Laplace transform for the analysis of deterministic signals. In this article, let us extend these tools to the stationary random processes.

5.22.1. Power Spectral Density (psd)

Power spectral density $S_X(\omega)$ of a wide-sense stationary (WSS) random process $X(t)$ may be defined as

$$S_X(\omega) = \text{Fourier transform } \{R_X(\tau)\}$$

$$= \int_{-\infty}^{\infty} R_X(\tau)\, e^{-j\omega\tau}\, d\tau \quad ...(5.137)$$

where $R_X(\tau)$ = Autocorrelation function of random process $X(t)$.

Thus, we can say that the Fourier transform of autocorrelation function $R_X(\tau)$ of a random process $X(t)$ is called power spectral density (psd) of random process. Equation (5.137) was intially developed by Albert Einstein. Later it was developed by N. Weiner and Khenchine. Hence, this equation is known as Einstein-Weiner Khenchine (EWK) equation. Power spectral density (psd) is used to measure the energy of the signal in the frequency band $[\omega, \omega + d\omega]$.

Cross Power Spectral Density (CPSD).

The cross power spectral density of two jointly WSS random processes $X(t)$ and $Y(t)$ may be defined as

$$S_{XY}(\omega) = \text{Fourier transform } \{R_{XY}(\tau)\}$$

$$= \int_{-\infty}^{\infty} R_{XY}(\tau)\, e^{-j\omega\tau} d\tau \qquad ...(5.138)$$

where $R_{XY}(\tau)$ = cross-correlation function of random processes $X(t)$ and $Y(t)$.

Properties of Power Spectral Density

Power spectral density (psd) of a random process has the following main properties.

(i) *Power spectral density of a random process is a real function of frequency ω.*

Proof: We know that the autocorrelation function of a random process is an even function, *i.e.*

$$R_X(\tau) = R_X(-\tau)$$

Therefore,
$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)\, e^{-j\omega\tau} d\tau$$

or
$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)\, [\cos \omega\tau - j \sin \omega\tau]\, d\tau$$

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) \cos \omega\tau\, d\tau - j \int_{-\infty}^{\infty} R_X(\tau) \sin \omega\tau\, d\tau \qquad ...(5.139)$$

Now let $\int_{-\infty}^{\infty} R_X(\tau) \cos \omega\tau\, d\tau = I$ ...(5.140)

Substituting $t = -\tau$ in equation (5.140), we get

$$\int_{-\infty}^{\infty} R_X(-t) \sin(-\omega t)(-dt) = I$$

or
$$-\int_{-\infty}^{\infty} R_X(-t) \sin \omega t\, dt = I \qquad ...(5.141)$$

Using the even property of autocorrelation function, we obtain

$$-\int_{-\infty}^{\infty} R_X(\tau) \sin \omega\tau\, d\tau = I \qquad ...(5.142)$$

After adding equations (5.141) and (5.142), we get

$$I = 0 \qquad ...(5.143)$$

Substituting $\int_{-\infty}^{\infty} R_X(\tau) \sin \omega t\, dt = 0$ in equation (5.143), we get

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) \cos \omega t\, d\tau \qquad ...(5.144)$$

Here $S_X(\omega)$ is a real function of ω.

(ii) Power spectral density (psd) of a random process $X(t)$ is an even function of frequency ω, i.e.

$$S_X(\omega) = S_X(-\omega)$$

(iii) Power spectral density (psd) of a random process $X(t)$ is a non-negative function of ω, i.e.,

$$S_X(\omega) \geq 0 \text{ for all } \omega$$

5.22.2. Energy Spectral Density (ESD)

The energy spectral density (psd) $\psi_X(\omega)$ may be defined as a measure of density of the energy contained in random process $X(t)$ in Joules per Hertz. It may be noted that since the amplitude spectrum of a real-valued random process $X(t)$ is an even function of ω, the energy spectral density of such a signal is symmetrical about the vertical axis passing through the origin.

Thus, total energy of the random process $X(t)$ is defined as

$$E = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_X(\omega)\, d\omega \qquad ...(5.145)$$

5.23. Response of Linear Systems to Random Inputs

In this article, we shall discuss about response of linear systems to random inputs. This means that we shall study transmission of random process through linear systems or linear filters.

Suppose that a random process $X(t)$ is applied to the input of a linear time-invariant (LTI) system having an impulse response $h(t)$. Let the input random process $X(t)$ produces a random process $Y(t)$ at the system output. The block diagram of this LTI system has been shown in figure 5.11. Here, we shall determine the mean and the auto-correlation function of the output random process $Y(t)$ in terms of input random process $X(t)$. Also, we are assuming that $X(t)$ is a wide-sense stationary (WSS) process.

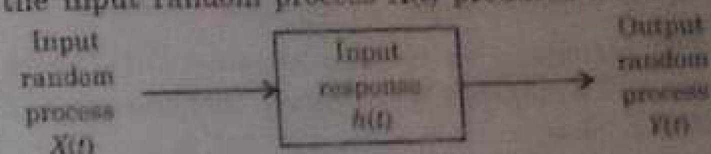


Fig. 5.11. *Determination of response of a linear system to a random input.*

The mean of the output random process is expressed as

$$m_Y(t) = E[Y(t)] = E\left[\int_{-\infty}^{\infty} h(\tau)\, X(t-\tau)\, d\tau\right]$$

or
$$m_Y(t) = \int_{-\infty}^{\infty} h(\tau)\, E[X(t-\tau)]\, d\tau = \int_{-\infty}^{\infty} h(\tau)\, m_X(t-\tau)\, d\tau \qquad ...(5.146)$$

When the input random process $X(t)$ is wide-sense stationary, the mean $m_X(t)$ is a constant m_X. Therefore, equation (5.146) may be written as

$$m_Y = \int_{-\infty}^{\infty} h(\tau)\, m_X\, d\tau = m_X \int_{-\infty}^{\infty} h(\tau)\, d\tau = m_X H(0) \qquad ...(5.147)$$

where $H(0)$ = zero-frequency response of the system.

In other words, we can say that the mean of the output random process of a stable linear time-invariant (LTI) system is equal to the mean of the input random process multiplied by the zero-frequency response of the system.

We shall now determine the autocorrelation function of the output random process $Y(t)$. Let t and u be two values of time at which output process is observed. Autocorrelation function of output random process is defined as

$$R_Y(t, u) = E[Y(t)\, Y(u)] \qquad ...(5.148)$$

By using convolution intergal, equation (5.148) can be written as

$$R_Y(t,u) = E\left[\int_{-\infty}^{\infty} h(\tau_1)\, X(t-\tau_1)\, d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)\, X(u-\tau_2)\, d\tau_2\right] \quad ...(5.149)$$

Assuming $E[X^2(t)]$ is finite for all t and the system is stable, we can interchange the order of the expectation operator and integration w.r.t. τ_1 and τ_2 to get

$$R_Y(t,u) = \int_{-\infty}^{\infty} h(\tau_1)\, d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)\, d\tau_2\; E[X(t-\tau_1)\, X(u-\tau_2)]$$

$$= \int_{-\infty}^{\infty} h(\tau_1)\, d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)\, d\tau_2\; R_X(t-\tau_1, u-\tau_2) \quad ...(5.150)$$

When the input $X(t)$ is a wide-sense stationary process, the autocorrelation function of $X(t)$ is only a function of the difference between the observation time $t-\tau_1$ and $u-\tau_2$. Substituting $\tau = t-u$ in equation (5.150), we get

$$R_Y(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)\, h(\tau_2)\, R_X(\tau-\tau_1+\tau_2)\, d\tau_1\, d\tau_2 \quad ...(5.151)$$

$R_Y(\tau)$ is the function of time lag τ.

By combining the result of equation (5.147) with equation (5.151), we see that if the input to a stable LTI system is a WSS process, then the output of the linear time-invariant (LTI) system is also a WSS process.

Putting $\tau = 0$ in equation (5.147), we get

$$R_Y(0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)\, h(\tau_2)\, R_X(\tau_2-\tau_1)\, d\tau_1\, d\tau_2 \quad ...(5.152)$$

Also we know that

$$R_Y(0) = E[Y^2(t)] \quad ...(5.153)$$

Example 5.29. A random variable X follows Gaussian probability density function with parameters μ and σ. Find its probability distribution function.

Solution: We know that the relationship between probability density function (pdf) and the probability distribution function of a random variable X is expressed as

$$F_X(z) = \int_{-\infty}^{z} f_x(x)\, dx \quad ...(i)$$

Since we have given that the random variable X follows Gaussian probability density function then

$$F_X(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma}}\, e^{-(x-\mu)/2\sigma^2}\, dx \quad ...(ii)$$

Substituting $\frac{x-\mu}{\sigma} = y$ in equation (ii), we get

$$F_x(z) = \int_{-\infty}^{[(Z-\mu)/\sigma]} \frac{1}{\sqrt{2\pi\,\sigma}}\, e^{-y^2/2}(\sigma dy) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{[(z-\mu)/\sigma]} e^{-y^2/2} dy$$

or

$$F_x(z) = \frac{1}{\sqrt{2}}\left[\int_{-\infty}^{0} e^{-y^2/2}\, dy + \int_{0}^{(z-\mu)/2} e^{-y^2/2} dy\right]$$

$$= \left[-\int_{0}^{\infty} e^{-y^2/2} dy + \int_{0}^{(z-\mu)} e^{-y^2/2} dy\right] \quad ...(iii)$$

Also error function is defined as

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-y^2/2}\, dy \quad ...(iv)$$

and

$$erf(\infty) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-y^2/2} dy = \frac{1}{2} \quad ...(v)$$

Putting equations (iv) and (v) in equation (iii), we get

$$F_x(z) = \frac{1}{2} + erf\left(\frac{z-\mu}{\sigma}\right) \quad \text{Ans.}$$

Example 5.30. Prove that for a Gaussian random variable X,

$$P[x_1 \le X \le x_2] = erf\left(\frac{x_2-\mu}{\sigma}\right) - erf\left(\frac{x_1-\mu}{\sigma}\right)$$

here erf (x) = Error function

μ and σ = parameters of Gaussian probability density function and x_1, x_2 = any real numbers.

Solution: We know that

$$P[x_1 \le X \le x_2] = \int_{x_1}^{x_2} f_X(x)\, dx \quad ...(i)$$

here

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{(x-\mu)^2/2\sigma^2} \quad ...(ii)$$

Substituting equation (ii) in equation (i), we get

$$P[x_1 \le X \le x_2] = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\,\sigma}}\, e^{(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi\,\sigma}} \int_{x_1}^{x_2} e^{(x-\mu)^2/2\sigma^2} dx \quad ...(iii)$$

Putting $\left(\frac{x-\mu}{\sigma}\right) = y$ in equation (iii), we get

$$P[x_1 \le X \le x_2] = \frac{1}{\sqrt{2\pi\,\sigma}} \int_{[(x_1-\mu)/\sigma)}^{[(x_2-\mu)/\sigma]} e^{y^2/2}\sigma\, dy = \frac{1}{\sqrt{2\pi}} \int_{[(x_1-\mu)/\sigma]}^{[(x_2-\mu)/\sigma]} -e^{y^2/2}\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\left[-\int_{[(x_1-\mu)/\sigma]}^{0} e^{y^2/2}dy + \int_{0}^{[(x_2-\mu)/\sigma]} e^{y^2/2}dy\right]$$

$$= \frac{1}{\sqrt{2\pi}}\left[-\int_{0}^{(x_1-\mu)/\sigma} e^{y^2/2}dy + \int_{0}^{(x_2-\mu)/\sigma} e^{y^2/2}dy\right]$$

$$= -erf\left(\frac{x_1-\mu}{\sigma}\right) + erf\left(\frac{x^2-\mu}{\sigma}\right) \quad \textbf{Ans.}$$

Example 5.31. The joint probability density function of two random variables X and Y is given by $f_{XY}(x, y) = C(1 - x - y)$ for values of x and y for which (x, y) lies within the triangle as shown in figure 5.12.

Determine:

(i) the value of constant C

(ii) the marginal probability density function $f_X(x)$ and $f_Y(y)$. *(Kerla University, 1991)*

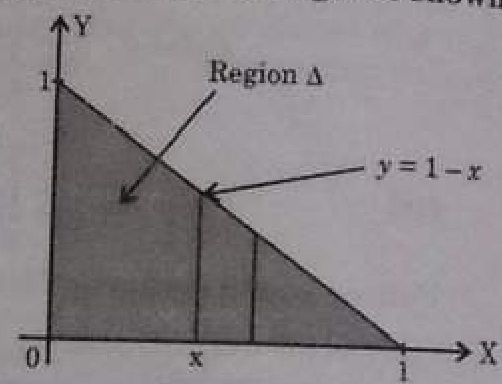


Fig. 5.12.

Solution: (i) The constant C may be found by using following condition

$$\iint_{\Delta} f_{XY}(x, y)\, dx\, dy = 1 \quad ...(i)$$

The triangle is bounded by the region $x = 0, x + y = 1$.

Equation (i) may be written as

$$\int_0^1 \int_0^{1-x} C(1 - x - y)\, dx\, dy = 1$$

or $$\int_0^1 \frac{C(1-x)^2}{2}\, dx = 1$$

or $$\frac{C}{6} = 1 \text{ or } C = 6$$

(ii) The marginal probability density function $f_X(x)$ is expressed as

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dy = \int_0^{1-x} 6(1 - x - y)\, dy = 6\left[-\frac{y^2}{2}\right]_0^{1-x}$$

or $$f_X(x) = \begin{cases} -3(1-x)^2, & \text{for } 0 \le x \le 1 \\ 0, & \text{for } x < 0, x \ge 1 \end{cases}$$

And $$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx = \int_0^{1-y} 6(1 - x - y)\, dx = 6\left[-\frac{x^2}{2}\right]_0^{1-y}$$

or $$f_Y(y) = \begin{cases} -3(1-y)^2, & \text{for } 0 \le y \le 1 \\ 0, & \text{for } y < 0, y \ge 1 \end{cases} \quad \textbf{Ans.}$$

Example 5.32. The marginal probability density functions of two random variables X and Y are given below

$$f_X(x) = \begin{cases} 3(1-x)^2, & \text{for } 0 \le x \le 1 \\ 0, & \text{for } x < 0, x > 1 \end{cases}$$

and $$f_Y(y) = \begin{cases} 3(1-y)^2, & \text{for } 0 \le y \le 1 \\ 0, & \text{for } y < 0, y > 1 \end{cases}$$

Determine:

(i) Mean of random variables X and Y

(ii) Variance of random variables X and Y. *(WBTU, Kolkata-2003)*

Solution: (i) Mean of $$X = \mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

$$E(X) = \int_0^1 x\, 3(1-x)^2\, dx = \frac{1}{4}$$

and the mean of $Y = \mu_Y$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^1 y\, 3(1-y)^2\, dy = \frac{1}{4}$$

(b) Variance of $$X = \sigma_X^2 = E(X^2) - \mu_X^2 \quad ...(i)$$

Now, $E(X^2)$ may be found as

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \int_0^1 x^2\, 3(1-x)^2\, dx = \frac{1}{10}$$

Similarly, we can find $E(Y^2)$ as

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy = \int_0^1 y^2\, 3(1-y)^2\, dy = \frac{1}{10}$$

Substituting the values $E[X^2]$ and μ_X in equation, we get

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \frac{1}{10} - \left(\frac{1}{4}\right)^2 = \frac{8-5}{80} = \frac{3}{80}$$

Similarly, variance of Y is determined as

$$\sigma_Y^2 = E(Y^2) - \mu_Y^2 \quad ...(ii)$$

Putting the values of $E(Y^2)$ and μ_Y in equation (ii), we get

$$\sigma_Y^2 = \frac{1}{10} - \left(\frac{1}{4}\right)^2 = \frac{3}{80}$$

Example 5.33. Given two random processes $X(t)$ and $Y(t)$ as

$$X(t) = Z_1(t) + 3Z_2(t - \tau)$$
$$Y(t) = Z_2(t + \tau) + 3Z_1(t - \tau)$$

here $Z_1(t)$ and $Z_2(t)$ are independent white noise processes each with a variance equal to 0.5. Determine,

(i) autocorrelation functions of $X(t)$ and $Y(t)$.

(ii) cross-correlation function of $X(t)$ and $Y(t)$. *(U.P.S.C., I.E.S., Examination, 1997)*

Solution: (a) Autocorrelation function of $X(t)$ is expressed as

$$R_X(t_1, t_0) = E[X(t_1)\,X(t_0)] = E[\{Z_1(t_1) + 3Z_2(t_1 - \tau)\}\{Z_1(t_0)\,3Z_2(t_0 - \tau)\}]$$

$$= E[Z_1(t_1)\,Z_1(t_0)] + 3E[Z_1(t_1) - Z_2(t_0 - \tau)] + 3E[Z_2(t_1 - \tau)\,Z_1(t_0)] + 9[Z_2(t_1 - \tau)\,Z_2(t_0 - \tau)] \quad ...(i)$$

But we know that

$$E[Z_1(t_1)\,Z_2(t_0 - \tau] = 0$$

Since, $Z_1(t)$ and $Z_2(t)$ are independent.

Similarly, $E[Z_2(t_1 - \tau)\,Z_1(t_0)] = 0$

For $t_1 \neq t_0$, $E[Z_1(t_1)\,Z_1(t_0)] = 0$

If $t_1 = t_0$ $E[Z_1(t_1)\,Z_1(t_0)] = E\,[Z_1^2(t_1)] = \sigma^2 + \mu^2 = 0.5\,\delta(t_1 - t_0)$ ...(ii)

Similarly $E[Z_2(t_1 - \tau)\,Z_2(t_0 - \tau)] = 0$ for $t_1 \neq t_0 = 0.5$ for $t_1 = t_0 = 0.5\,\delta(t_1 - t_0)$ ...(iii)

Substituting equations (ii) and (iii) in equation (i), we get

$$R_X(t_1, t_0) = 0.5\,\delta(t_1 - t_0) + 4.5\,\delta(t_1 - t_0) \quad ...(iv)$$

Similarly, $R_Y(t_1, t_0) = 0.5\,\delta(t_1 - t_0) + 4.5\,\delta(t_1 - t_0)$ ...(v)

From equations (iv) and (v), we see that the random processes $X(t)$ and $Y(t)$ are wide-sense stationary (WSS).

(ii) Cross-correlation function of random processes $X(t)$ and $Y(t)$ is expressed as

$$R_{XY}(t_1, t_0) = E[X(t_1)\,Y(t_0)] = E[\{Z_1(t_1) + 3Z_2(t_1)\}\{Z_2(t_0 + \tau) + 3Z_1(t_0 - \tau)\}]$$

$$= E[Z_1(t_1)\,Z_2(t_0 + \tau)] + 3E[Z_1(t_1)\,Z_1(t_0 - \tau)]$$

$$+ 3E[Z_2(t_1)Z_2(t_0 + t)] + 9E[Z_2(t_1)\,Z_1(t_0 - \tau) \quad ...(vi)$$

Now $E[Z_1(t_1)\,Z_2(t_0)] = 0$ (Since $Z_1(t_1)$ and $Z_2(t_0)$ are independent random variables)

Therefore,
$$E[Z_1(t_1)Z_1(t_0 - \tau)] = \begin{cases} 0, & \text{if } t_1 \neq t_0 - \tau \\ 0.5, & \text{if } t_1 = t_0 - \tau \end{cases}$$

$$E[Z_2(t_1 - \tau)Z_2(t_0)] = \begin{cases} 0, & \text{if } t_0 \neq t_1 - \tau \\ 0.5, & \text{if } t_0 = t_1 - \tau \end{cases}$$

Now, $E[Z_2(t_1 - \tau)\,Z_1(t_0 - \tau)] = 0$

Substituting above values in equation (vi), we get

$$R_{XY}(t_1, t_0) = 1.5\delta(t_1 - t_0 + \tau) + 1.5\,\delta(t_0 - t_1 + \tau) \quad ...(vii)$$

Because cross-correlation depends upon the difference of t_1 and t_0 therefore, random processes $X(t)$ and $Y(t)$ are jointly stationary or simply stationary. **Ans.**

Example 5.34. Stationary random process $X(t)$ has the following autocorrelation function

$$R_X(\tau) = \sigma^2\,e^{-\mu|\tau|}$$

where μ and σ^2 are constants. It is passed through a filter whose impulse response is

$$h(\tau) = \alpha\,e^{-\alpha\tau}\,u(\tau)$$

here α is a contant and $u(\tau)$ is a step function.

(i) Find the power spectral density of random signal $X(t)$.

(ii) Find the power spectral density of the output random signal $Y(t)$.

(GATE Examination, 1998)

Solution: (i) Power spectral density (psd) of the input is expressed as

$$S_X(\omega) = \text{CTFT}\,\{R_X(t)\} = \int_{-\infty}^{\infty} R_X(\tau)\,e^{-j\omega\tau}\,d\tau$$

or
$$S_X(\omega) = \int_{-\infty}^{\infty} \sigma^2\,e^{-\mu|\tau|}\,e^{-j\omega\tau}\,\tau = \sigma^2 \int_{-\infty}^{0} e^{(\mu - j\omega)\tau}\,d\tau + \sigma^2 \int_{0}^{\infty} e^{-(\mu + j\omega)\tau}\,d\tau$$

or
$$S_X(\omega) = \sigma^2\left[\frac{1}{\mu - j\omega} + \frac{1}{\mu + j\omega}\right] = \frac{2\mu\sigma^2}{\mu^2 + \omega^2}$$

(ii) Now, power spectral density of the output is related to the power spectral density of input as

$$S_Y(\omega) = |H(\omega)|^2\,S_X(\omega)$$

and
$$H(e^{j\omega}) = \text{CTFT}\,\{h(\tau)\}$$

or
$$H(e^{j\omega}) = \int_{-\infty}^{\infty} h(\tau)\,e^{-j\omega\tau}\,d\tau = \int_{-\infty}^{\infty} \alpha\,e^{-\alpha\tau}\,u(\tau)\,e^{-j\omega\tau}\,d\tau$$

or
$$H(e^{j\omega}) = \alpha \int_{0}^{\infty} e^{-(\alpha + j\omega)\tau}\,d\tau = \frac{\alpha}{\alpha + \omega}$$

Substituting equation (ii) in equation (i), we obtain

$$S_Y(jw) = \left|\frac{\alpha}{\alpha + j\omega}\right|^2 S_X(\omega) = \frac{\alpha}{\alpha^2 + \omega^2}\,S_X(\omega)$$

or
$$S_Y(jw) = \frac{\alpha}{\alpha^2 + \omega^2}\,\frac{2\mu\sigma^2}{\mu^2 + \omega^2} = \frac{2\mu\alpha^2\sigma^2}{(\alpha^2 + \omega^2)(\mu^2 + \omega^2)} \quad \textbf{Ans.}$$

Example 5.35. A deterministic signal $x(t) = \cos 2\pi t$ is passed through a differentiator as shown in figure 5.13.

(a) Determine the autocorrelation $R_{xx}(\tau)$ and the power spectral density $S_{xx}(f)$.

(b) Find the output power spectral density $s_{yy}(f)$.

(c) Evaluate $R_{xy}(0)$ and and $R_{xy}\left(\frac{1}{4}\right)$.

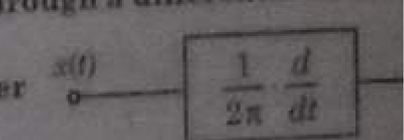


Fig. 5.13.

(GATE Examination

Solution: We know that the autocorrelation is expressed as

$$R_{xx}(\tau) = E[X(t + \tau).X(t)]$$

or
$$R_{xx}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} \cos[2\pi(t + \tau)]\cos 2\pi t\,dt$$

or
$$R_{xx}(\tau) = \frac{1}{2T}\int_{-T/2}^{T/2} [\cos 2\pi(2t + \tau) + \cos 2\pi\tau]\,dt$$

or $$R_{xx}(\tau) = \frac{1}{2T}\left[\frac{\sin 2\pi(2t+\tau)}{4\pi}\Bigg|^{T/2} + (\cos 2\pi\tau).t\Big|_{-T/2}^{T/2}\right]$$

Simplifying, we get

$$R_{xx}(t) = \frac{\cos 2\pi\tau}{2}$$

Further, we know that

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau).e^{-j\omega\tau}d\tau = \frac{1}{2}[\delta(f-1)+\delta(f+1)]$$

We have, using

$$\cos(2\pi f_c t) = \frac{1}{2}[\delta(f-f_c)+\delta(f+f_c)]$$

(b) Now, $$R_{yy}(\tau) = \frac{d}{dt}.R_{xx}(\tau) = \pi\sin(2\pi\tau)$$

Also,

$$S_{yy}(f) = H(j\omega).S_{xx}(f) = j\omega.S_{xx}(f)$$

or $$S_{yy}(f) = (j.2\pi f).\frac{1}{2}[\delta(f-1)+\delta(f+1)]$$

$$= (j\pi f).[\delta(f-1)+\delta(f+1)] \quad \textbf{Ans.}$$

(c) Again, we have

$$R_{xy}(t) = E[X(t).Y(t+\tau)]$$

or $$R_{xy}(t) = \frac{2\pi}{T}\int_{-T/2}^{T/2}[\sin 2\pi(2t+\tau)+\sin(2\pi\tau)].dt$$

or $$R_{xy}(\tau) = \frac{\pi}{T}\left[\frac{\cos 2\pi(2t+\tau)}{4\pi}\Bigg|_{-T/2}^{T/2} + \sin 2\pi\tau.T\right] = \pi.\sin 2\pi\tau$$

then $$R_{xy}(0) = 0$$

$$R_{xy}\left(\frac{1}{4}\right) = \pi.\sin\frac{\pi}{2} = \pi \quad \textbf{Ans.}$$

SUMMARY

1. There is one other class of signals, the behaviour of which cannot be predicted. Such type of signals are called **random signals**.
2. These signals are called random signals because the precise value of these signals cannot be predicted in advance before they actually occur. The examples of random signals are the noise interferences in communication systems.
3. An experiment is defined as the process which is conducted to get some results. If the same experiment is performed repeatedly under the same conditions, similar results are expected.
4. The outcomes of an event are called **equally** likely, if any one of them cannot be expected to occur in preference to another. As an example, the tossing of a coin results in two outcomes, Head and Tail. Both have same possibility of 50%. Such type of outcomes are called **equally** likely outcomes.

Probability may be defined as the study of random experiments. In any random experiment, there is always an uncertainty that a particular event will occur or not.

The concept of conditional probability is used in conditional occurrences of the events.

The conditional probability of event B given that event A has already happened.

$$P(B/A) = \frac{P(AB)}{P(A)}$$

where $P(AB)$ is the joint probability of A and B.

Similarly, the conditional probability of event A given that event B has already happened.

$$P(A/B) = \frac{P(AB)}{P(B)}$$

The joint probability has commutative property which states that

$$P(AB) = P(BA)$$

If A and B are two events in an experiment, and possibility of occurrence of event B does not depend upon occurrence of event A, then these two events A and B are known as statistically independent events.

A function which can take on any value from the sample space and its range is some set of real numbers is called a **random variable** of the experiment.

Random variables may be classified as under:

1. Discrete random variables
2. Continuous random variables.

11. A discrete random variable may be defined as the random variable which can take on only finite number of values in a finite observation interval. This means that the discrete random variable has countable number of distinct values.

12. A random variable that takes on an infinite number of values is called a continuous random variable. Actually, there are several physical system (experiments) that generate continuous outputs or outcomes. Such systems generate infinte number of outputs or outcomes within the finite period.

13. The Comulative Distribution Function (CDF) of a random variable 'X' may be defined as the probability that a random variable 'X' takes a value less than or equal to x.

14. The derivative of cumulative distribution function (CDF) with respect to some dummy variable is known as Probability Density Function (PDF). Probability density function (PDF) is generally denoted by $f_X(x)$. Mathematically, PDF may be expressed as

$$PDF: f_X(x) = \frac{d}{dx}F_X(x)$$

where x is a dummy variable.

Probability density function (PDF) is the more convenient representation for **continuous random variable.**

15. The joint Cumulative Distribution Function may be defined systematically as:

The joint Cumulative Distribution Function $F_{XY}(x, y)$ may be defined as the probability that the outcome of an experiment will result in a sample point lying inside the range $(-\infty < X \le x, -\infty < Y \le$ $y)$ of the joint sample space. The joint sample space is the combined sample space of X and Y. Mathematically,

$$F_{XY}(x, y) = P(X \le x, Y \le y)$$

16. The joint PDF of any two random variables X and Y may be defined as the partial derivative of the joint cumulative distribution function $F_{XY}(x, y)$ with respect to the dummy variables x and y.

$$\text{Uniform PDF } f_X(x) = 0 \quad \text{for } x < m - \frac{A}{2} \;\&\; x > m + \frac{A}{2}$$

$$= \frac{1}{A} \quad \text{for } \left(m - \frac{A}{2}\right) \le x \le \left(m + \frac{A}{2}\right) \qquad ...(5.84)$$

Figure 5.14 shows the PDF of an uniformly distributed random variable.

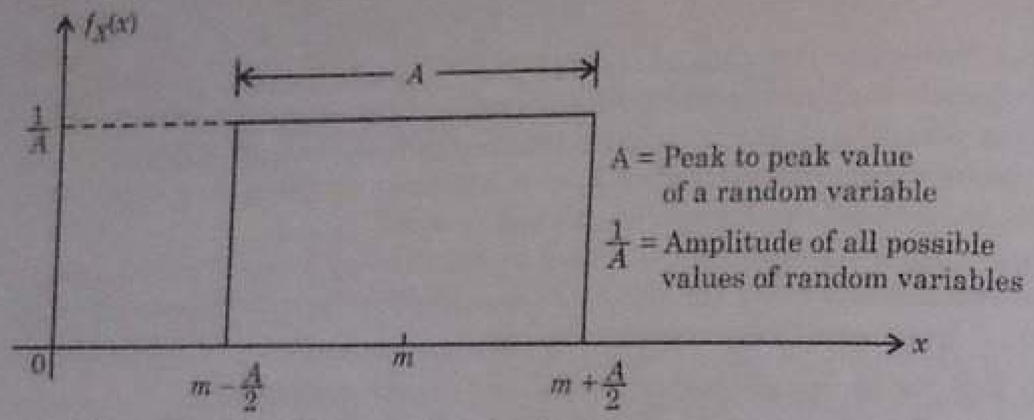


Fig. 5.14. *PDF of an uniformly distributed random variable*

In above figure, the peak to peak value is *A* and amplitude is uniform *i.e. A*. From figure 5.3, it is also clear that the random variable and its PDF are continuous. Therefore uniform distribution is utilised for continuous random variable.

Also, the value of PDF, $f_X(x)$ is same for all possible values of a random variable. Thus this distribution is known as Uniform Distribution.

Q.14. What is importance of Gaussian distribution?

Ans. The Gaussian distribution is used for continuous random variables. We know that the random motion of the thermally agitated electrons produces thermal noise. This thermal noise has Gaussian distribution. The random errors in the experimental measurements create the measured values to have Gaussian distribution about the true value. The Gaussian distribution is very important in the analysis of communication and statistical systems.

Q.15. What do you mean by random process? Explain.

Ans. Let there be a random experiment *E* having outcome λ from the sample sapce *S*. This means that $\lambda \in S$. Thus every time an experiment is conducted, the outcome λ will be one of the sample point in sample space. If this outcome λ is associated with time, then a function of λ and time *t* is formed *i.e.*, $X(\lambda, t)$. Then the function $X(\lambda, t)$ is known as random process. Hence when any random experiment *E* is given a time dimension, then each outcome appears at some certain time and the random experiment will be converted to Random Process. A random process is the function of two variables λ and *t*.

Q.16. What is Ergodic process? Explain.

Ans. A random process is known as ergodic process if the time-averages are equal to ensemble averages. Hence for a ergodic process, we have

$$m_x = \langle m_x \rangle$$
$$R_\tau(t_1, t_2) = \langle R_\tau(\varepsilon) \rangle$$

For any ergodic process, all ensemble averages will be equal to corresponding time averages for any particular sample function. The time averages are not a function of time. When time and ensemble averages are the same, it implies that ensemble averages also are not a function of time.

Thus an ergodic process is always stationary but converse in not true.

Ergodicity of the process may be defined in terms of some statistical averages like mean and autocorrelation.

The random process is ergodic in the mean if

$$m_x = \langle m_x \rangle$$

and variance of $\langle m_x \rangle \to 0$ as $T \to \infty$

Similarly, the random process is ergodic in the autocorrelation if

$$R_\tau(t_1, t_2) = \langle R_\tau(\varepsilon) \rangle$$

and variance of $\langle R_\tau(\varepsilon) \rangle \to 0$ as $T \to \infty$

Hence, the Ergodicity of any random process may be determined by evaluating statistical averages of single sample function. This means that a single sample function represents entire random process.

NUMERICAL PROBLEMS

1. A wheel of chance is divided into three equal segments coloured Green (G), Red (R) and Yellow (Y). The wheel is spinned twice and the outcome is taken as resulting colour sequence. That is GR, RG etc. Let *A* be the event which represents "neither colour is Yellow". Let B be the event which represents "matching colours". Calculate *P*(*A*), *P*(*B*), *P*(*AB*) and *P*(*A* + *B*).

[**Ans.** $P(A) = \frac{4}{9}$, $P(B) = \frac{1}{3}$, $P(AB) = \frac{2}{9}$ and $P(A+B) = \frac{5}{9}$]

2. Three honest coins are tossed simultaneously and the resulting sequence of heads and tails is observed. Let *A* be the event in which "first two tosses match." Let *B* be the event representing "exactly two heads". Find out *P*(*A*), *P*(*B*), *P*(*B*/*A*) and *P*(*A*/*B*).

[**Ans.** $P(A) = \frac{1}{2}$, $P(B) = \frac{3}{8}$, $P(B/A) = \frac{1}{4}$ and $P(A/B) = \frac{1}{3}$]

3. A long binary message contains 1428 binary 1's and 2668 binary 's. What is the probability of obtaining binary 1 in any received bit? [**Ans.** $P(A = 1) = 0.5486$]

4. Which of the following functions satisfy the properties of PDF and Why? Hence state whether the function is PDF or not. *(Kerala University, 1998)*

(*a*) $f_X(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right)$

(*b*) $f_X(x) = |x| \quad$ for $|x| < 1$
$= 0 \quad$ otherwise

(*c*) $f_X(x) = \frac{1}{6}(8 - x) \quad$ for $4 \le x \le 10$
$= 0 \quad$ otherwise

[**Hint:** For all above function apply $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

If this property holds, then it is PDF].

[**Ans.** (*a*) This is valid PDF, (*b*) This is valid PDF, (*c*) This is not PD

5. The PDF of a random variable is given as

$f_X(x) = Ke^{-bx} \quad$ for $x \ge 0$
$= 0 \quad$ for $x < 0 \quad$ and $K, b > 0$

(*i*) Find the value of *K* in terms of *b*. *(A.M.I.E., Examination, 19*

(*ii*) Find m_x and σ^2_x.

6. The PDF of a random variable '*X*' is given as

$f_X(x) = \frac{1}{A^2}(x + A) \quad$ for $x < 0$

4. An AM system with envelope has $(S/N)_o = 30$ dB and tone modulation with $m = 1$ with $B = 8$ kHz. If all bandwidths are increased accordingly while other parameters remain fixed, what is the largest usable value of B? [Ans. 1.2 MHz]
5. Determine the detector gain α_d for an SSB system. [Ans. $\alpha_d = 1$]
6. Determine the output SNR in a PM system for the modulation. [Ans. $\left(\frac{S}{N}\right)_o = \frac{1}{2}\beta^2\gamma$]
7. Determine the detection gain α_d for an FM system with $\beta = 2$. [Ans. $\alpha_d = 36$]
8. Prove that for tone modulation, FM is superior to PM by a factor of 3 from the SNR point of view.
9. For a modulating signal $x(t) = \cos^3\omega_m t$, show that PM is superior to FM by a factor of 2.25 from the output SNR point of view.
10. Consider a communication system with the following characteristics:

$$S_X = E[x^2(t)] = \frac{1}{2} \quad B = 10 \text{ kHz} \quad \frac{\eta}{2} = 10^{-12} \text{ W/Hz}$$

Transmission loss = 70 dB.

Calculate the required transmission power S_T needed to achieve $(S/N)_o = 40$ dB when the modulation is (*i*) SSB, (*ii*) AM with $m = 1$ and $m = 0.5$ (*iii*) PM with $kp = \pi$, (*iv*) FM with $D = 1$ and $D = 5$.

[Ans. (*i*) SSB, $S_T = 1$ kW; (*ii*) AM; $m = 1$, $S_T = 3$ kW; $m = 0.5$, $S_T = 9$ kW; (*iii*) PM, $S_T = 202.6$ W; (*iv*) FM ; $D = 1$, $S_T = 667$ W; $D = 5$, $S_T = 26.7$ W]

□□□

CHAPTER 8

AM and FM Receivers

Inside this Chapter

8.1. Introduction

We know that in a communication system, a radio transmitter radiates or transmits a modulated carrier signal. This modulated carrier signal travels down the channel *i.e.* transmission medium and reaches at the input of radio receiver. This means that the modulated carrier signal is picked up by the antenna of the radio receiver. This modulated signal so received is generally very weak. Therefore, inside the receiver, this weak signal is first amplified in an R.F. (radio frequency) amplifier stage of the radio receiver. Also, since the received modulated signal contains a lot of noise or unwanted signals at adjacent frequencies, it must be selected and the noise must be rejected. Finally, in receiver, the R.F. carrier or modulated signal must be demodulated to get back the original modulating or baseband signal. Further, since the demodulated or detected signal (*i.e.* audio signal in case of broadcast receiver) is generally weak, it has to be amplified in one or more stages of audio amplifier.

From the above discussion, we can summarize the main functions of a radio receiver as under:

(*i*) intercept the incoming modulated signal (*i.e.* electromagnetic waves) by the receiving antenna.

(*ii*) select the desired signal and reject the unwanted signals.

(*iii*) amplify this selected R.F. signal.

(*iv*) detect the modulated signal to get back the original modulating or baseband signal.

(*v*) amplify the modulating frequency signal.

This means that a radio receiver is an electronic equipment which picks up the desired signal, rejects the unwanted signals, amplifies the desired signal, demodulates the modulated signal to get back the original modulating frequency signal.

8.2. Classification of Radio Receivers

We can classify the radio receiver in two ways as under:

(A) Depending upon the applications, the radio receivers may be classified as follows :

(*i*) *Amplitude Modulation (AM) Broadcast Receivers:* These receivers are used to receive the broadcast of speech or music transmitted from amplitude modulation broadcast transmitters which operate on long wave, medium wave or short wavebands.

(*ii*) *Frequency Modulation (FM) Broadcast Receivers:* These receivers are used to receive the broadcast programmes from FM broadcast transmitters which operate in VHF or UHF bands.

(*iii*) *Communication Receivers:* Communication receivers are used for reception of telegraph and short wave telephone signals. This means that communication receivers are used for various purposes other than broadcast services.

(*iv*) *Television Receivers:* Television receivers are used to receive television broadcast in VHF or in UHF bands.

(*v*) *Radar Receivers:* Radar receivers are used to receive radar (*i.e.*, radio detection and ranging) signals.

(B) Depending upon the fundamental aspects, the radio receivers may also be classified as under:

(*i*) Tuned Radio Frequency (TRF) Receivers.

(*ii*) Superheterodyne Receiver.

In fact, various forms of receivers have been proposed from time to time. However, only two of them became popular for commercial applications. These are Tuned Radio Frequency (TRF) receiver and superheterodyne receiver. Presently, the superheterodyne receiver is the most popular and most widely used. The TRF receiver was used earlier in the 1940s. The TRF receiver had some inherent drawbacks which were removed in superheterodyne receiver. Therefore, we shall start our discussion with TRF receiver and then come to the superheterodyne receiver.

8.3. Tuned Radio Frequency (TRF) Receiver

Tuned radio frequency (TRF) receiver is the simplest radio receiver. Figure 8.1 shows the block diagram of a tuned radio frequency receiver. The very first block of this receiver is an RF stage. This stage generally contains two or three RF amplifiers. Actually, these RF (radio frequency) amplifiers

are tuned RF amplifiers *i.e.* they have variable tuned circuit at the input and output sides. At the input of the receiver, there is a receiving antenna as shown in block diagram in figure 8.1. At this antenna signals from different sources (*i.e.* stations) are present. However, with the help of input variable tuned circuit of RF amplifiers the desired signal (*i.e.* station) is selected. But this selected signal is usually very weak of the order of μV. This selected weak signal is amplified by the RF amplifier (*i.e.* RF stage).

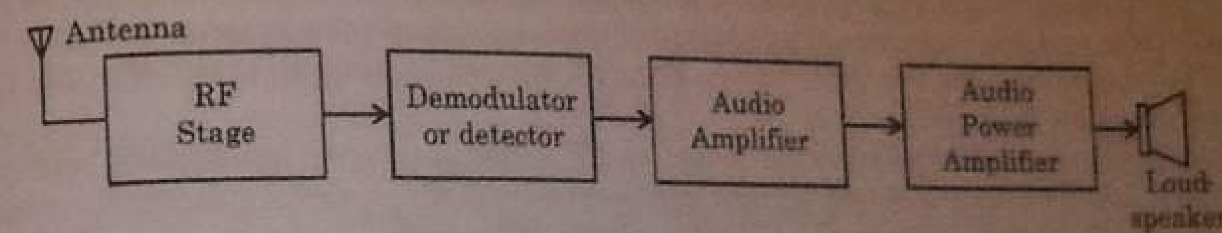


Fig. 8.1. *Block diagram of a tuned radio frequency (TRF) receiver.*

Thus, the function of RF stage is to select the desired signal and amplify it.

After this, the amplified incoming modulated signal is applied to the demodulator. The demodulator or detector demodulates the modulated signal and thus at the output of the demodulator, we get modulating or baseband signal (*i.e.* audio signal). This audio signal is amplified by audio amplifier. After that, this audio signal is further amplified by a power amplifier upto desired power level to drive the loudspeaker. The last stage of this receiver is the loudspeaker. A loudspeaker is a transducer which changes electrical signal into sound signal.

8.3.1. Drawbacks of Tuned Radio Frequency (TRF) Receiver

As discussed above, although TRF receiver is cheaper and the simplest one, it has certain drawbacks as under:

(*i*) The TRF receiver suffers from a tendency to oscillate at higher frequencies from the multistage RF amplifiers with high gain and operating at same frequency. If such an amplifier has a gain of 20,000 then if a small portion of the output leaked back to the input of the RF stage, then positive feedback and oscillation will result. This type of leakage could result from power supply coupling, stray capacitance coupling, radiation coupling or coupling through any other element common to the input and output stages. Definitely, this type of condition is undesirable for a good receiver.

This problem is also termed as instability of the receiver.

(*ii*) The selectivity of a receiver is its ability to distinguish between a desired signal and an undesired signal. The selectivity of TRF receiver is poor. In fact, it is difficult to achieve sufficient selectivity at high frequencies due to the enforced use of single-tuned circuits.

(*iii*) Another problem associated with the TRF receiver is the bandwidth variation over the tuning range. For example, in AM broadcast system, let us consider that a tuned circuit is required to have a bandwidth of 10 kHz at a frequency of 540 kHz.

According to the definition, the Quality factor Q of this tuned circuit must be

$$Q = \frac{\text{Resonance frequency}}{\text{Bandwidth}} = \frac{540}{10} = 54$$

Now, at the other end of this AM broadcast band (*i.e.* 1640 kHz), the Quality factor Q of the coil, according to above equation, must increase by a factor of 1640/535 (*i.e.* 3) to a value of 164. However, in practice due to several losses dependent upon frequency would prevent such a large increase. Thus, practically, the Quality factor Q of this tuned circuit is unlikely to exceed 120 and hence providing a bandwidth of the tuned circuit equal to

$$\Delta f = \frac{f_r}{Q} = \frac{1640}{120} = 13.8 \text{ kHz}$$

Therefore, due to this increased bandwidth of 13.8 kHz in place of a fixed bandwidth of 10 kHz, the receiver would pick up or select adjacent frequencies (i.e. stations) with the desired frequency or station. This means that the bandwidth of the TRF receiver varies with the incoming frequency.

8.4. Superheterodyne Receiver

Figure 8.2 shows the block diagram of a superheterodyne receiver. All the drawbacks in TRF receiver have been removed in a superheterodyne receiver. The basic superheterodyne receiver is most widely used. This means that the superheterodyne principle is used in all types of receiver like television receiver, radar receiver etc.

In a superheterodyne receiver, the incoming RF signal frequency is combined with the local oscillator signal frequency through a mixer and is converted into a signal of lower fixed frequency. This lower fixed frequency is known as **intermediate frequency**.

However, the intermediate frequency signal contains the same modulation as the original signal. This intermediate frequency signal is now amplified and demodulated to reproduce the original signal.

The word heterodyne stands for mixing. Here we have mixed the incoming signal frequency with the local oscillator frequency. Therefore this receiver is called superheterodyne receiver.

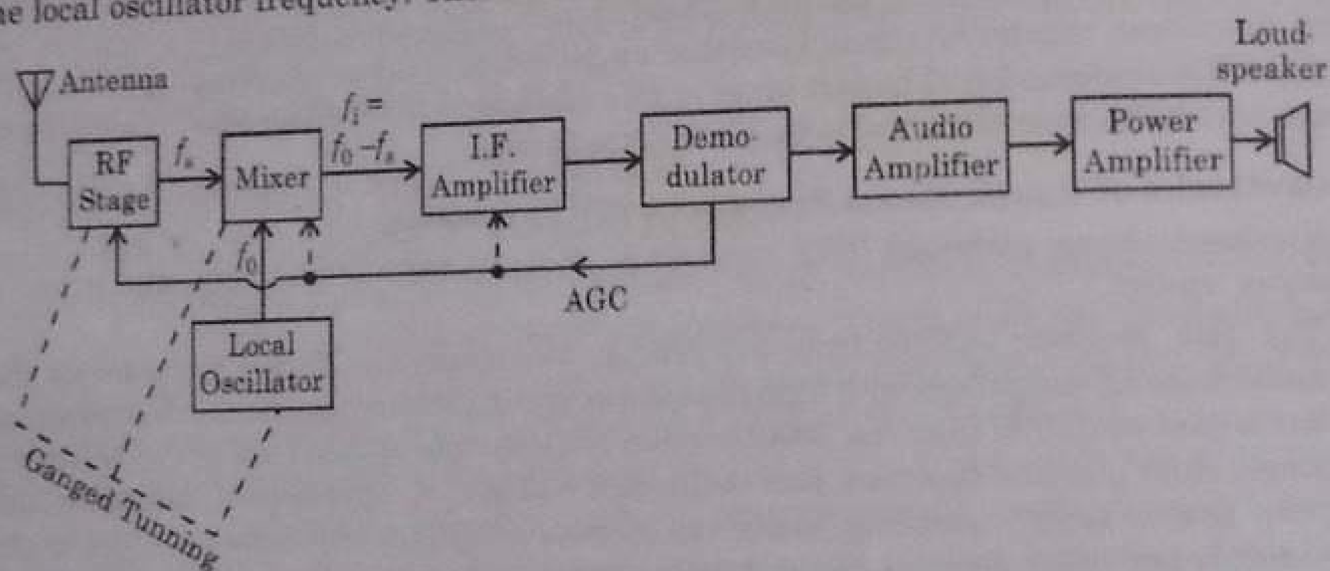


Fig. 8.2. *Block diagram of a superheterodyne receiver.*

Thus, in a superheterodyne receiver, a constant frequency difference is maintained between the local oscillator signal frequency and incoming RF signals frequency through capacitance tuning in which the capacitances are ganged together and operated by a common control knob. The intermediate frequency (IF) amplifier generally contains a number of transformers each consisting of a pair of mutually coupled tuned circuits. Thus, with this large number of double-tuned circuits, operating at a specially chosen frequency, the IF amplifier provides most of the gain (*i.e.* sensitivity) and bandwidth requirements (*i.e.* selectivity) of the receiver. This means that the IF amplifier determines the sensitivity and selectivity of the superheterodyne receiver.

Also, since the characteristics of the IF amplifier are independent of the incoming frequency to which the receiver is tuned, the selectivity and sensitivity of the superheterodyne receiver are quite uniform throughout its tuning range and not subject to the variations like a TRF receiver. Further since the IF amplifier works at a fixed IF frequency, the design of this system is quite easy to provide high gain and constant bandwidth.

Because of its narrow bandwidth, the IF amplifier rejects all other frequencies except intermediate frequency (IF). Actually, this rejection process reduces the risk of interference from other stations or sources. Infact, this selection process is the key to the superheterodyne receiver's exceptional performance.

After the IF amplifier, the signal is applied at the input of demodulator which extracts the original modulating signal (*i.e.* audio signal). This audio signal is amplified by an audio amplifier

to get a particular voltage level. This amplified audio signal is further amplified by a powe[r] to get a specified power level so that it may activate the loudspeaker. The loudspeaker is a [...] which converts this audio electrical signal into audio sound signal and thus the origina[l] reproduced *i.e.* the original transmission is received.

The advantages of the superheterodyne receiver make it the most suitable for the [...] radio receiver applications like AM, FM, communications, single-sideband, televisio[n] radar receiver; all use superheterodyne principle. This means that it can be considere[d] standard form of radio receiver.

8.5. AM Receivers

Because the type of receiver is almost the same for various forms of modulation or system[s] it is generally most convenient to explain the various principles of a superheterody[ne] while dealing with AM receivers. Thus, with the discussion of AM receiver, a basis i[s] the more complex versions of superheterodyne receiver. In this section, let us [...] superheterodyne receiver characteristics.

8.6. Sensitivity

The sensitivity of a radio receiver may be defined as its ability to amplify weak signals. It [...] defined in terms of the voltage which must be applied at receiver input terminals to provi[de] output power measured at the output terminals. For AM broadcast receivers, several releva[nt] have been standardized. A signal modulated by a 400 Hz sine wave and modulation in [...] applied through standard coupling network known as a **dummy antenna**.

In addition to this, the loudspeaker is replaced by an equivalent load resistance. [...] output is measured across this resistance and it must be equal to the standard value [...]

Sensitivity is also expressed in microvolts or in decibels below 1 volt and is meas[ured] [...] points along the tuning range when a production receiver is lined up.

Figure 8.3 shows the sensitivity curve over the tuning band. At 1 MHz, this particular receiver has a sensitivity of 12.7 μ volt or –98 dB volt. Sometimes, the sensitivity defintion is extended, and the manufacture of this receiver may quote it to be, nor merely 12.7 μV, but "12.7 μV for an SNR of 20 dB in the output of the receiver".

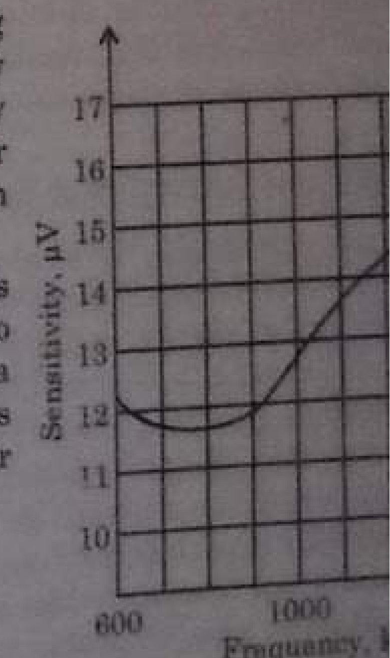


Fig. 8.3. *A typical se[nsitivity curve] for a good domesti[c receiver]*

However, for professional receivers, the sensitivity is generally quoted in terms of signal power required to produce a minimum acceptable output signal with a minimum acceptable output noise level. Few factors determining the sensitivity of a superheterodyne receiver are as under:

(*a*) The gain of the IF amplifiers.

(*b*) The gain of the RF amplifiers.

(*c*) The noise figure of the receiver.

It may be noted that the typical values of sensitivity are 150 μ volt for small broadcast band receivers, and 1 μ volt or below for high quality communication receiver in the HF band.

8.7. Selectivity

The selectivity of a receiver may be defined as the ability to reject unwanted signals. It [...] the attenuation that the receiver offers to signal at frequencies adjacent to the one [...] tuned. It is generally, expressed as a curve as shown in figure 8.4. In selectivity me[...]

frequency of the generators is varied to either side of the frequency to which the receiver is tuned. Naturally, the output of the receiver falls since the input frequency is not incorrect. Thus, the input voltage must be increased until the output is the same as it was originally. The ratio of the voltage required of resonance to the voltage required. When the generator is tuned to the receiver's frequency it is calculated at a number of points and then plotted in decibels to give a curve as shown in figure 8.4. It may be noted that selectivity depends upon the following factors:

(i) Selectivity varies with receiving frequency and becomes somewhat worse when the receiving frequency is raised.

(ii) In general, it is mainly determined by the response of the IF section, with the mixer and RF amplifier input circuits playing a small but significant part.

(iii) Selectivity is the main factor which determines the adjacent channel rejection of a receiver.

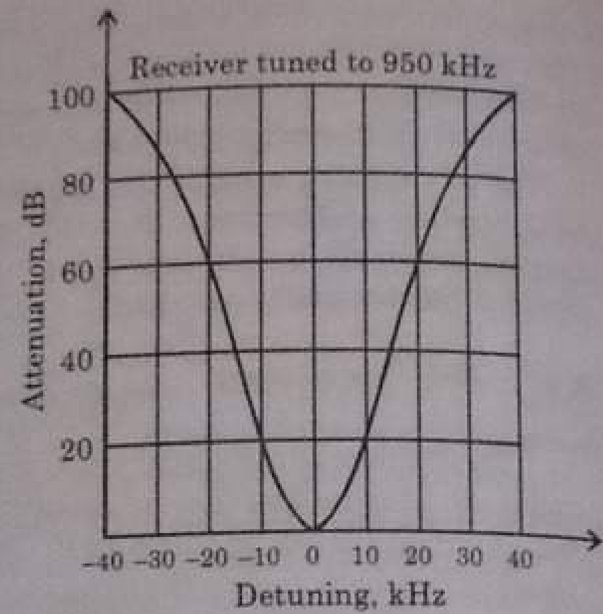


Fig. 8.4. A Typical selectivity curve

8.8. Double Spotting

When a receiver picks up the same short wave station at two nearby points on the receiver dial, the double spotting phenomenon takes place. The main cause for double spotting is poor front-end selectivity, *i.e.*, inadequate image-frequency rejection. The front-end of the receiver does not select different adjacent signal very well. However, fortunately, the IF stage.

The adverse effect of double spotting is that a weak station may be marked by the reception of a nearby strong station at the spurious point on the dial. On the other hand, double spotting may be used to calculate the IF of an unknown receiver. The spurious point on the dial is precisely $2f_i$ below the correct frequency. If image-frequency rejection is improved, then certainly there will be a corresponding decrease in the double spotting occurrence.

8.9. Tracking or Tuning of a Superheterodyne Receiver

In a superheterodyne receiver, the local oscillator frequency is made to track with the tuned circuits which are tuned to the incoming signal frequency in order to make a constant frequency difference at the output of mixer. For general AM broadcast system, the intermediate frequency (IF) is 455 kHz. This indicates that the local oscillator should always be set at a frequency which is 455 kHz above the incoming signal frequency. For purpose the front end of the receiver tuned circuits are made to track together simply by mechanically linked or ganged capacitors. A ganged capacitor has three capacitor sections, one each for the RF amplifier, mixer and the local oscillator. In addition to this, small variable capacitances known as trimmers are connected in parallel with each section. These capacitances can be adjusted for proper operation at highest frequency. However for lowest frequency adjustment, small variable capacitors known as padders are connected in series with the inductor of the tank circuit.

The various tuned circuits are mechanically coupled so that only one tuning control and dial are required. This means that no matter what is the incoming signal frequency, the RF and mixer input tuned circuits must be tuned to it. The local oscillator must simultaneously be tuned to a frequency which is precisely higher than the signal frequency by the intermediate frequency. However, any error that may exist in the frequency difference would result in an incorrect frequency being fed to the intermediate frequency (IF) amplifier. This error must naturally be avoided. Such type of errors are known as tracking errors. These tracking errors result in stations appearing away from their correct position on the dial.

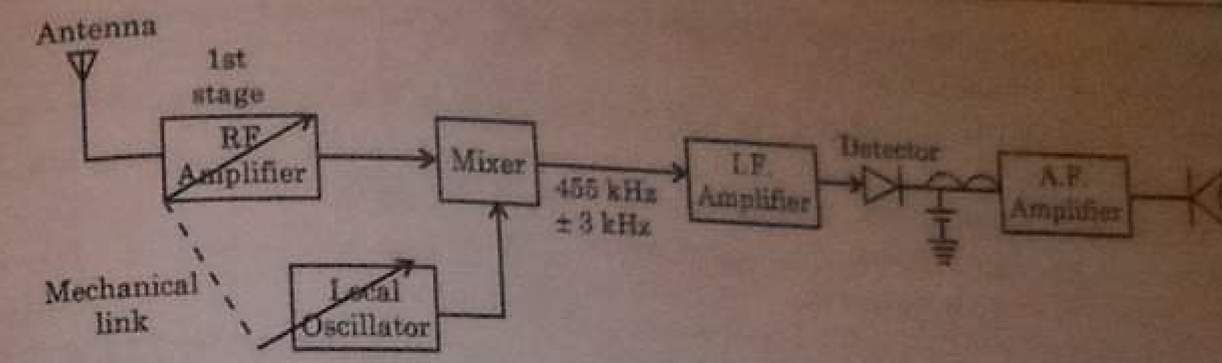


Fig. 8.5. Tuning of a superheterodyne receiver.

The various tuned circuits are mechanically coupled so that only one tuning control and dial are required. This means that no matter what is the incoming signal frequency, the RF and mixer input tuned circuits must be tuned to it. The local oscillator must simultaneously be tuned to a frequency which is precisely higher than the signal frequency by the intermediate frequency. However, any error that may exist in the frequency difference would result in an incorrect frequency being fed to the intermediate frequency (IF) amplifier. This error must naturally be avoided. Such type of errors are known as tracking errors. These tracking errors result in stations appearing away from their correct position on the dial.

It is quite possible to keep the maximum tracking error below 3 kHz. A value as low as this is quite acceptable.

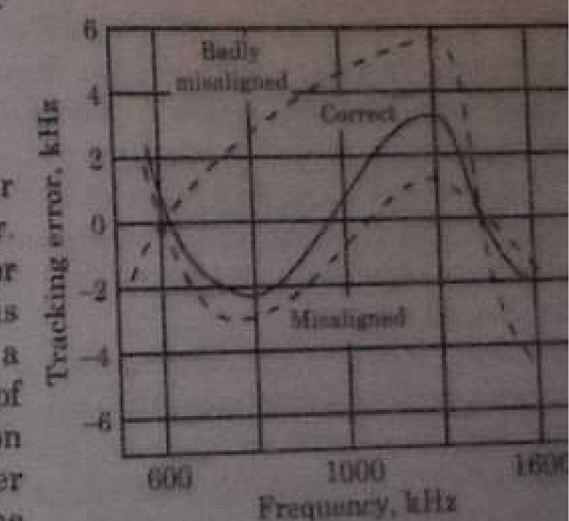


Fig. 8.6. Tracking curves

8.10. Image Frequency and its Rejection

We know that a superheterodyne receiver is a better receiver than a Tuned Radio Frequency (TRF) receiver. However, a superheterodyne receiver suffers from a major drawback known as Image Frequency problem. This problem of image frequency is inherent to a superheterodyne receiver and arises because of the use of heterodyne principle. In fact, the frequency conversion process carried out by the local oscillator and the mixer often allow an undesired frequency in addition to the desired incoming frequency.

In a standard broadcast receiver, the local oscillator frequency is always made higher than the incoming signal frequency. It is kept equal to the signal frequency plus t… intermediate frequency (IF).

Mathematically,

$$f_o = f_s + f_i \quad \text{...(8}$$

where f_o = local oscillator frequency

f_s = desired incoming frequency

f_i = intermediate frequency

From equation (8.1), we have

$$f_i = f_o - f_s$$

Hence, the intermediate frequency is the difference between the local oscillator frequency a… the signal frequency.

Now, if a frequency f_{si} manages to reach the mixer, such that

$$f_{si} = f_o + f_i \quad \text{...(}$$

then this frequency f_{si} would also produce f_i when it is mixed with f_o. This undesired or spuri… intermediate frequency signal will also be amplified by the IF stage and thus would ca…

interference. This has the effect of two sources or stations being received simultaneously. This situation is obviously undesirable.

The term f_{si} is known as the image frequency and is defined as the signal frequency plus twice the intermediate frequency.

Putting the value of f_o in equation (8.2) from equation (8.1), we get

$$f_{si} = f_o + f_i$$

$$f_{si} = f_s + f_i + f_i$$

or $$f_{si} = f_s + 2f_i \quad ...(8.3)$$

Thus this spurious frequency signal cannot be distinguished by the IF stage and hence would be treated in the same manner as the desired frequency signal.

The rejection of an image frequency signal by a single tuned circuit may be defined as the ratio of the gain at the signal frequency to the gain at the image frequency. This is given as

$$\alpha = \sqrt{1+Q^2\rho^2} \quad ...(8.4)$$

Here $$\rho = \frac{f_{si}}{f_s} - \frac{f_s}{f_{si}} \quad ...(8.5)$$

and Q = Quality factor of the tuned circuit in a loaded condition.

If the receiver has an RF stage, then there is only a single tuned circuit and the rejection will be calculated using equation (8.4). However if the receiver has an RF stage then there are two tuned circuits both tuned to f_s. The image frequency rejection of each stage will be calculated by using equation (8.4). The total or overall rejection will be the product of the two.

The image-frequency rejection of the receiver depends upon the front-end selectivity of the receiver. The rejection of image frequency must be achieved before the IF stage. Once an undesired or spurious frequency enters the first IF amplifier, it would become impossible to remove it from the desired signal.

It may be observed that if $\frac{f_{si}}{f_s}$ is large as is the case for AM broadcast band the use of an RF stage is not necessary for good image frequency rejection. However, it would become essential above about 3 MHz.

Example 8.1. For a broadcast superheterodyne AM receiver having no RF amplifier, the loaded Quality factor Q of the antenna coupling circuit is 100. Now if the intermediate frequency is 455 kHz, then determine the following:

(i) the image frequency and its rejection ratio at an incoming frequency of 1000 kHz.

(ii) the image frequency and its rejection ratio at an incoming frequency of 25 MHz.

Solution: Given that

$$Q = 100$$

and $$f_i = 455 \text{ kHz}$$

$$f_s = 1000 \text{ kHz}$$

The image-frequency is given as

$$f_{si} = f_s + 2f_i = 1000 + 2 \times 455$$

or $$f_{si} = 1000 + 910 = 1910 \text{ kHz}$$

Further, $$\rho = \frac{f_{si}}{f_s} - \frac{f_s}{f_{si}} = \frac{1910}{1000} - \frac{1000}{1910}$$

or $$\rho = 1.910 - 0.524 = 1.386$$

Since the given receiver has no RF amplifier, therefore there is only single tuned circuit.

The rejection ratio is given as

$$\alpha = \sqrt{1+Q^2\rho^2} = \sqrt{1+(100)^2 \times (1.386)^2}$$

or $$\alpha = \sqrt{1+(138.6)^2} = 138.6 \quad \textbf{Ans.}$$

(ii) For second case, it is given that

$$Q = 100$$

$$f_i = 455 \text{ kHz} = 0.455 \text{ MHz}$$

and $$f_s = 1000 \text{ kHz}$$

The image frequency is given as

$$f_{si} = f_s + 2f_i = 25 + 2 \times 0.455$$

or $$f_{si} = 25.91 \text{ MHz}$$

$$\rho = \frac{f_{si}}{f_s} - \frac{f_s}{f_{si}} = \frac{25.91}{25} - \frac{25}{25.91}$$

or $$\rho = 1.0364 - 0.9649 = 0.0715$$

The rejection ratio α is given as

$$\alpha = \sqrt{1+Q^2\rho^2} = \sqrt{1+(100)^2 \times (0.0715)^2}$$

$$\alpha = \sqrt{1+(7.15)^2} = 5.22 \quad \textbf{Ans.}$$

8.11. AM Superheterodyne Receiver (Description of Various Blocks)

In this section, let us discuss the principle and working of different blocks used in a superheter[...] AM receiver.

8.12. RF Amplifier

R.F. amplifier is a small signal tuned amplifier with tuned circuits both in the input side an[...] output side. Both these input and output tuned circuits are tuned to the desired incoming c[...] frequency. Accordingly the tuned circuits select the desired carrier frequency and reject all und[...] frequencies including the image frequency. Hence the RF amplifier provides image freq[...] rejection. Also the gain provided by the RF amplifier will result in improved signal/noise ra[...] the output of the receiver. This is due to the fact that the incoming weak signal is raised to a h[...] level with the help of RF amplifier before it is fed at the input of the mixer stage which contr[...] most of the noise generated in the receiver. However, if the incoming weak signal is fed dire[...] the frequency mixer, the signal/noise ratio at the output of the mixer stage is quite poor and[...] any amount of subsequent amplification cannot improve S/N ratio. Thus the one important fu[...] of the RF amplifier is to improve S/N ratio.

There are some cases also where an RF amplifier is not used in the receiver rather its[...] uneconomical there. The best example of this kind of receiver is a domestic receiver used in a[...] signal-strength area like a metropolitan city. Since in a metropolitan city like Delhi, several st[...] are situated and in such places strength is obviously very high and thus, there is no need [...] use of RF amplifier. In such cases the tuned circuit connected to the antenna is the actua[...] circuit of the mixer.

However, a receiver having an RF amplifier is obviously superior in performance to a r[...] without RF amplifier.

We may summarize the advantages of RF amplifier as under:

(i) Greater gain *i.e.* better sensitivity

(ii) Improved rejection of adjacent undesired signals *i.e.* better selectivity

other problem associated with the TRF receiver is the bandwidth variation over the
ing range. For example, in AM broadcast system, let us consider that a tuned circuit is
uired to have a bandwidth of 10 kHz at a frequency of 540 kHz.

do you mean by Sensitivity?

ensitivity of a radio receiver may be defined as its ability to amplify weak signals. It is
efined in terms of the voltage which must be applied at receiver input terminals to provide
utput power measured at the output terminals. For AM broadcast receivers, several relevant
ave been standardized. A signal modulated by a 400 Hz sine wave and modulation index of
ed through standard coupling network known as a **dummy antenna**.
tion to this, the loudspeaker is replaced by an equivalent load resistance. After this the
easured across this resistance and it must be equal to the standard value of 50 mW.
vity is also expressed in microvolts or in decibels below 1 volt and is measured at three
the tuning range when a production receiver is lined up.

Selectivity for a receiver?

lectivity of a receiver may be defined as the ability to reject unwanted signals. It also
he attenuation that the receiver offers to signal at frequencies adjacent to the one to
uned. In selectivity measurement, the frequency of the generators is varied to either
requency to which the receiver is tuned. Naturally, the output of the receiver falls since
equency is not incorrect. Thus, the input voltage must be increased until the output is
it was originally. The ratio of the voltage required of resonance to the voltage required.

the advantages of a R.F. amplifier.

dvantages of RF amplifier as under:
eater gain *i.e.* better sensitivity
roved rejection of adjacent undesired signals *i.e.* better selectivity
roved signal/noise ratio
roved image frequency rejection
roved coupling of the receiver to the antenna
vention of reradiation of the local oscillator voltage through the antenna
vention of spurious frequencies from entering the mixer and heterodyne to produce
erfering frequency equal to IF

REVIEW QUESTIONS

re the salient features of broadcast radio receivers?
e principle of a heterodyne receiver.
e block diagram of a superheterodyne receiver and explain the function of each block.
e circuit of a frequency mixer and local oscillator using transistors and explain its working.
meant by the term 'tracking error'?
typical tracking error curve.
e circuit of an IF amplifier and explain its working.
re the main functions served by an IF amplifier?
ctors govern the choice of intermediate frequency?
meant by the term adjacent channel selectivity?
e circuit of a linear diode detector and explain its working.
ish between simple AGC and delayed AGC.

□□□

CHAPTER 9

Sampling Theory and Pulse Modulation

Inside this Chapter

9.1. Introduction

As we know that broadly, there are two types of signals, continuous time signal and discrete-time
signals. Due to some recent advance development in digital technology over the past few decades,
the inexpensive, light weight, programmable and easily reproducible discrete-time systems are
available. Therefore, the processing of discrete-time signals is more flexible and is also preferable
to processing of continuous-time signals.

This means that in practice, although we have a large number of continuous-time signals,
but we prefer processing of discrete-time signals. For this purpose we should be able to convert a
continuous-time signal into discrete-time signal.

This problem is solved by a fundamental mathematical tool known as sampling theorem. The
sampling theorem is extremely important and useful in signal processing. With the help of sampling
theorem, a continuous-time signal may be completely represented and recovered from the knowledge
of samples taken uniformly. This means that sampling theorem provides a mechanism for representing

a continuous-time signal by a discrete-time signal. Therefore, sampling theorem may be viewed as a bridge between continuous-time signals and discrete-time signals.

The concept of sampling provides a widely used method for using discrete-time system technology to implement continuous-time systems and process the continuous-time signals. We utilize sampling to convert a continuous-time signal to a discrete-time signal, process the discrete-time signal using a discrete-time system and then convert back to continuous-time signals.

9.2. The Sampling Theorem

(U.P. Tech-Semester Exam. 2002-2003)

sampling of the signals is the fundamental operation in signal-processing. A As discussed earlier, continuous time signal is first converted to discrete-time signal by sampling process. The sufficient number of samples of the signal must be taken so that the original signal is represented in its samples completely. Also, it should be possible to recover or reconstruct the original signal completely from its samples. The number of samples to be taken depends on maximum signal frequency present in the signal. Sampling theorem gives the complete idea about the sampling of signals. Different types of samples are also taken like ideal samples, natural samples and flat-top samples. Let us discuss the sampling theorem first and then we shall discuss different types of sampling processes. The statement of sampling theorem can be given in two parts as:

(i) A bandlimited signal of finite energy, which has no frequency-component higher than f_m Hz, is completely described by its sample values at uniform intervals less than or equal to $\frac{1}{2f_m}$ second apart.

(ii) A bandlimited signal of finite energy, which has no frequency components higher than f_m Hz, may be completely recoverd from the knowledge of its samples taken at the rate of 2 f_m samples per second.

The first part represents the representation of the signal in its samples and minimum sampling rate required to represent a continuous-time signal into its samples.

The second part of the theorem represents reconstruction of the original signal from its samples. It gives sampling rate required for satisfactory reconstruction of signal from its samples.

Combining the two parts, the sampling theorem may be stated as under:

"A continuous-time signal may be completely represented in its samples and recovered back if the sampling frequency is $f_s \geq 2f_m$. Here, f_s is the sampling frequency and f_m is the maximum frequency present in the signal".

9.3. Proof of Sampling Theorem

(U.P. Tech-Semester Examination 2002-2003)

To prove the sampling theorem, we shall show that a signal whose spectrum is bandlimited to f_m Hz, can be reconstructed exactly without any error from its samples taken uniformly at a rate $f_s > 2f_m$ Hz.

Let us consider a continuous time signal $x(t)$ whose spectrum is bandlimited to f_mHz. This means that the signal $x(t)$ has no frequency components beyond f_mHz. Therefore, $X(w)$ is zero for $|w| > w_m$ i.e.,

$$X(w) = 0 \text{ for } |w| > w_m$$

where $$w_m = 2\pi f_m$$

Figure 9.1 (*a*) shows this continuous-time signal $x(t)$. Let $X(w)$ represents its Fourier transform or frequency spectrum as shown in figure 9.1(*b*).

Sampling of $x(t)$ at a rate of $f_s Hz$ (f_s samples per second) may be achieved by multiplying $x(t)$ by an impulse train $\delta_{T_s}(t)$. The impulse train $\delta_{T_s}(t)$ consists of unit impulses repeating periodically every T_s seconds, where $T_s = \frac{1}{f_s}$.

Figure 9.1(*c*) shows this impulse train. This multiplication results in the sampled signal g... shown in figure 9.1(*e*).

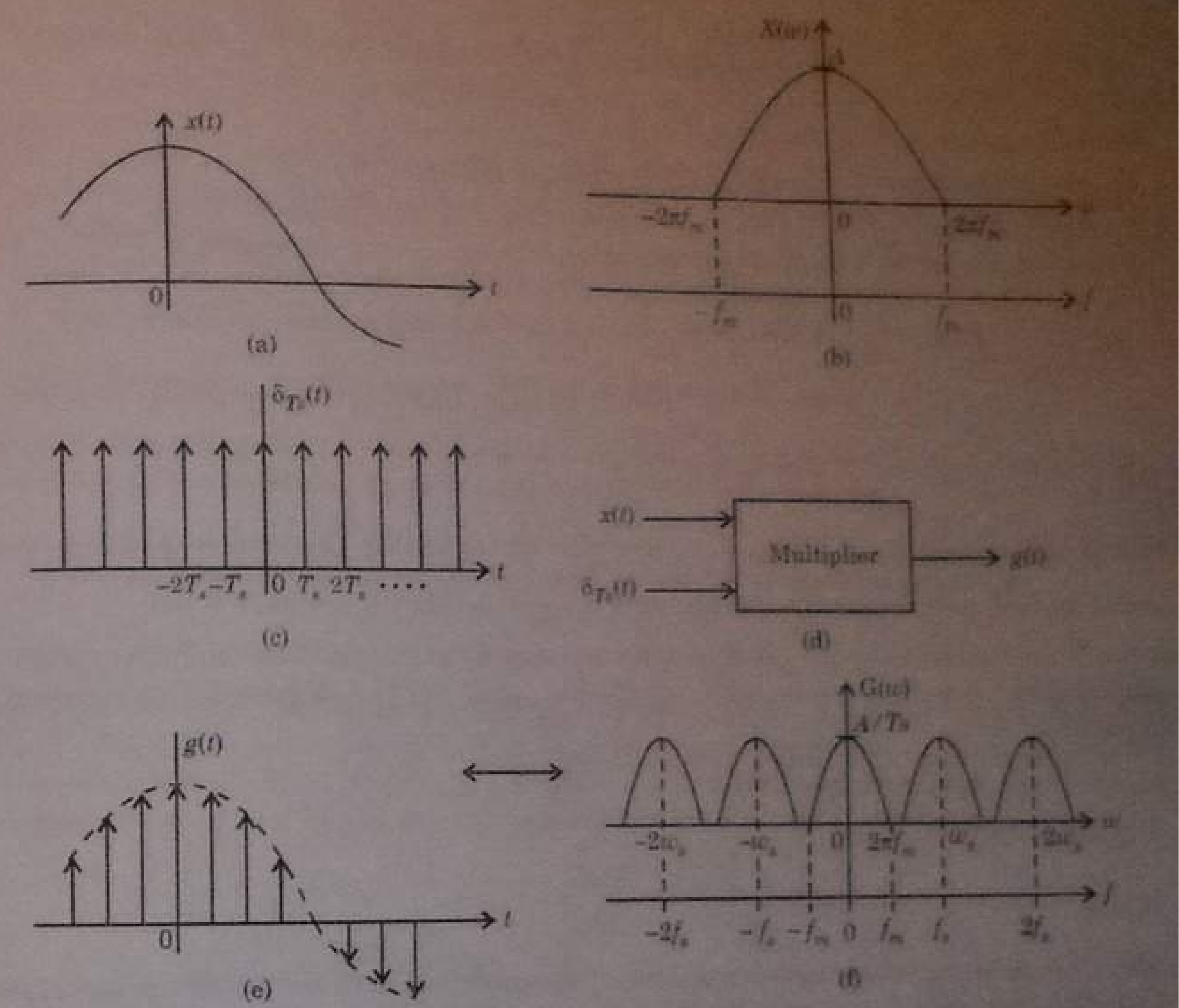


Fig. 9.1. (*a*) *A continuous-time signal.*
(*b*) *Spectrum of continuous-time signal.*
(*c*) *Impulse train as sampling function.*
(*d*) *Multiplier.*
(*e*) *Sampled signal.*
(*f*) *Spectrum of sampled signal.*

This sampled signal consists of impulses spaced every T_s seconds (the sampling interval...

The resulting or sampled signal may be written as

$$g(t) = x(t)\,\delta_{T_s}(t)$$

Again, since the impulse train $\delta_{T_s}(t)$ is a periodic signal of period T_s, it may be expressed ... Fourier series.

The trigonometric Fourier series expansion of impulse-train $\delta_{T_s}(t)$ is expressed as

$$\delta_{T_s}(t) = \frac{1}{T_s}[1 + 2\cos w_s t + 2\cos 2w_s t + 2\cos 3w_s t + ...]$$

Here $$w_s = \frac{2\pi}{T_s} = 2\pi f_s$$

Putting the values of $\delta_{T_s}(t)$ from equation (9.2) in equation (9.1), the sampled signal is

$$g(t) = \frac{1}{T_s}\,[x(t) + 2\,x(t)\cos w_s t + 2x(t)\cos 2\,w_s t + 2x(t)\cos 3\,w_s t + ...] \qquad ...(9.3)$$

Now, to obtain $G(w)$, the Fourier transformation of $g(t)$, we will have to take the Fourier transform of right hand side.

Fourier transform of $x(t)$ is $X(w)$.

Fourier transform of $2x(t)\cos w_s t$ is $[X(w - w_s) + X(w + w_s)]$.

Fourier transform of $2x(t)\cos 2\,w_s t$ is $[X(w - 2w_s) + X(w + 2w_s)]$ and so on.

Therefore, on taking Fourier transformation, the equation (9.3) becomes

$$G(w) = \frac{1}{T_s}\,[X(w) + X(w - w_s) + X(w + w_s) + X(w - 2\,w_s) + X(w + 2\,w_s) + X(w - 3\,w_s) + X(w + 3\,w_s) + ...] \qquad ...(9.4)$$

or

$$G(w) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} X(w - nw_s) \qquad ...(9.5)$$

From equations (9.4) and (9.5), it is clear that the spectrum $G(w)$ consists of $X(w)$ repeating periodically with period $w_s = \frac{2\pi}{T_s}$ rad/sec. or $f_s = \frac{1}{T_s}$ Hz as shown in figure 9.1 (*f*).

Now if have to reconstruct $x(t)$ from $g(t)$, we must be able to recover $X(w)$ from $G(w)$. This is possible if there is no overlap between successive cycles of $G(w)$. Figure 9.1 (*f*) shows that this requires

$$f_s > 2 f_m \qquad ...(9.6)$$

But the sampling interval $T_s = \frac{1}{f_s}$

Hence,
$$T_s < \frac{1}{2 f_m} \qquad ...(9.7)$$

Therefore, as long as the sampling frequency f_s is greater than twice the maximum signal frequency f_m (signal, bandwidth, f_m), $G(w)$ will consist of non-overlapping repetitions of $X(w)$. If this is true, figure 9.1 (f) shows that $x(t)$ can be recovered from its samples $g(t)$ by passing the sampled signal $g(t)$ through an ideal low-pass filter of bandwidth f_m Hz. This proves the sampling theorem.*

9.3.1. Few Points about Sampling Theorem

(*i*) Figure 9.1 (*f*) shows the spectrum of sampled signal. According to the figure, as long as, the signal is sampled at rate $f_s > 2 f_m$, the spectrum $G(w)$ will repeat periodically without overlapping.

(*ii*) The spectrum of sampled signal extends upto infinity and the ideal bandwidth of sampled signal is infinite. But here our purpose is to extract our original spectrum $X(w)$ out of the spectrum $G(w)$.

(*iii*) The original or desired spectrum $X(w)$ is centred at $w = 0$ and is having bandwidth or maximum frequency equal to w_m. The desired spectrum may be recovered by passing the sampled signal with spectrum $G(w)$ through a low pass filter with cut-off frequency w_m. This means that since a low-pass filter allows to pass only low frequencies up to cut-off frequency (w_m) and rejects all other higher frequencies, the original spectrum $X(w)$ extended upto w_m will be selected and all other successive higher frequency cycles in the sampled-spectrum will be rejected. Therefore, in this way, original spectrum $X(w)$ will be extracted out of spectrum $G(w)$. This original spectrum $X(w)$ can now be converted into time-domain signal $x(t)$.

* Maximum signal frequency f_m means that the bandwidth of this signal is simply f_m.

(*iv*) It may also be observed from figure that for the case $f_s > 2f_m$, the successive cycles of G are not overlapping each other. Hence in this case, there is no problem in recovering original spectrum $X(w)$.

(*v*) For the case $f_s = 2\,f_m$, although the successive cycles of $G(w)$ are not overlapping e other, but they are touching each other. In this case also, the original spectrum $X(w)$ be recovered from the sampled spectrum $G(w)$ using a low-pass filter with cu frequency w_m.

(*vi*) For the case $f_s < 2\,f_m$, the successive cycles, of the sampled spectrum will overlap other and hence in this case, the original spectrum $X(w)$ cannot be extracted out o spectrum $G(w)$.

Hence, For reconstruction without distortion, we must have

$$f_s \geq 2 f_m$$

9.4. Nyquist Rate and Nyquist Interval

When the sampling rate becomes exactly equal to $2\,f_m$ samples per second, then it is called Ny rate. Nyquist rate is also called the *minimum* sampling rate. It is given by

$$f_s = 2 f_m$$

Similarly, maximum sampling interval is called *Nyquist interval*. It is given by

$$\text{Nyquist interval } T_s = \frac{1}{2 f_m} \text{ seconds}$$

When the continuous-time bandlimited signal is sampled at Nyquist rate ($f_s = 2\,f_m$), the sa spectrum $G(w)$ contains non-overlapping $G(w)$ repeating periodically. But the successive cy $G(w)$ touch each other as shown in figure 9.2. Therefore, the original spectrum $X(w)$ can be rec from the sampled spectrum by using a low pass filter with a cut-off frequency w_m.

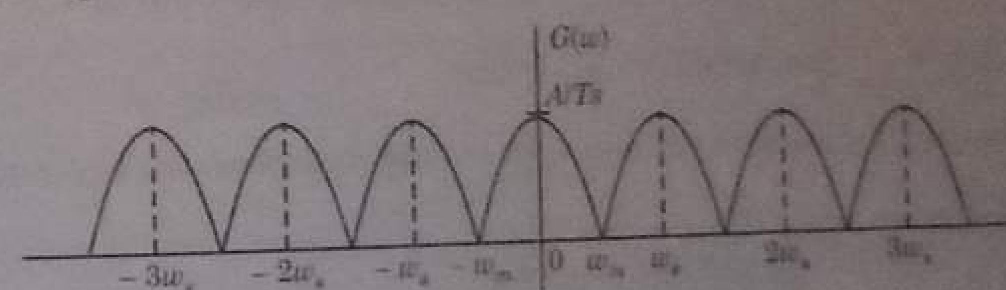


Fig. 9.2. *Sampled spectrum at Nyquist rate.*

Example 9.1. An analog signal is expressed by the equation $x(t) = 3\cos 50\pi t + 10\sin$ … $\cos 100\pi t$. Calculate the Nyquist rate for this signal.

Solution: The given signal is expressed as

$$x(t) = 3\cos 50\pi t + 10\sin 300\pi t - \cos 100\pi t$$

Let three frequencies present be w_1, w_2 and w_3

So that the new equation for signal.

$$x(t) = 3\cos w_1 t + 10\sin w_2 t - \cos w_3 t$$

Comparing equations (i) and (ii) we have

$$w_1 t = 50\,\pi t;\ w_1 = 50\,\pi$$

or
$$2\,\pi f_1 = 50\,\pi \quad \text{or} \quad 2 f_1 = 50$$

$\therefore$
$$f_1 = 25 \text{ Hz}$$

Similarly, for second factor

$$w_2 t = 300\pi t \quad \text{or} \quad w_2 = 300\pi$$

or $2\pi f_2 = 300\pi$ or $2\pi f_2 = 300\pi$

∴ $f_2 = 150$ Hz.

Again, for third factor

$w_3 t = 100\pi t$ or $2\pi f_3 t = 100\pi t$

or $2\pi f_3 = 100\pi$

∴ $f_3 = 50$ Hz

Therefore, the maximum frequency present in $x(t)$ is,

$$f_2 = 150 \text{ Hz}$$

Nyquist rate is given as

$$f_s = 2 f_m$$

where f_m = Maximum frequency present in the signal.

Here $f_m = f_2 = 150$ Hz

Therefore, Nyquist rate

$$f_s = 2 f_2 = 2 \times 150 = 300 \text{ Hz}$$ **Ans.**

Example 9.2. Find the Nyquist rate and the Nyquist interval for the signal

$$x(t) = \frac{1}{2\pi} \cos (4000\pi t) \cos (1000\pi t)$$

Solution: Given signal is

$$x(t) = \frac{1}{2\pi} \cos (4000\pi t) \cos (1000\pi t) = \frac{1}{4\pi} [2 \cos (4000\pi t) \cos 1000\pi t)]$$

or $$x(t) = \frac{1}{4\pi} [\cos (4000\pi t + 1000\pi t) + \cos (4000\pi t - 1000\pi t)]$$

$$[\because 2 \cos A \cos B = \cos (A + B) + \cos (A - B)]$$

or $$x(t) = \frac{1}{4\pi} [\cos 5000\pi t + \cos 3000\pi t] \quad ...(i)$$

Let the two frequencies present in the signal be w_1 and w_2 so that the new equation for the signal will be

$$x(t) = \frac{1}{4\pi} [\cos w_1 t + \cos w_2 t] \quad ...(ii)$$

Comparing equations (i) and (ii), we have

$w_1 t = 5000\pi t$

or $2\pi f_1 t = 5000\pi t$ or $2 f_1 = 5000$

∴ $f_1 = 2500$ Hz

Similarly, for second factor

$w_2 t = 3000\pi t$

or $2\pi f_2 t = 3000\pi t$ or $2\pi f_2 = 3000$

∴ $f_2 = 1500$ Hz

Therefore, the maximum frequency present in $x(t)$ is

$$f_1 = 2500 \text{ Hz}$$

Nyquist rate is given as

$$f_s = 2 f_m$$

where f_m = Maximum frequency present in the signal.

Here $f_m = f_1 = 2500$ Hz

Therefore Nyquist rate

$$f_s = 2 f_m = 2 \times 2500 = 5000 \text{ Hz} = 5 \text{ kHz}$$ **Ans.**

Nyquist interval is given as

$$T_s = \frac{1}{2 f_m} = \frac{1}{2 \times 2500} = \frac{1}{5000}$$

or $$T_s = 0.2 \times 10^{-3} \text{ seconds} = 0.2 \text{ m sec.}$$ **Ans.**

Example 9.3. A continuous-time signal is given below:

$$x(t) = 8 \cos 200 \pi t$$

Determine:

(i) Minimum sampling rate i.e., Nyquist rate required to avoid aliasing.

(ii) If sampling frequency f_s = 400 Hz. What is the discrete-time signal $x[n]$ or $x[nT_s]$ obtained after sampling?

(iii) If sampling frequency f_s = 400 Hz. What is the discrete-time signal $x[n]$ or $x[nT_s]$ obtained after sampling?

(iv) What is the frequency $0 < f < f_s/2$ of siausoidal that yields samples identical to those obtained in part (iii)?

Solution:

(i) The highest frequency component of continuous-time signal is f = 100 Hz. Hence minimum sampling rate required to avoid aliasing is called Nyquist rate and is given as

$$\text{Nyquist rate} = 2f = 2 \times 100 = 200 \text{ Hz}$$ **Ans.**

(ii) The continuous-time signal $x(t)$ is sampled at f_s = 400 Hz. The frequency of the discrete-time signal will be

$$F = \frac{\text{Frequency of continuous-time signal } f}{\text{Sampling frequency, } f_s} = \frac{100}{400} = \frac{1}{4}$$

Then the discrete-time signal will be given as

$$x[n] = 8 \cos 2\pi F n$$

$$= 8 \cos 2\pi \times \frac{1}{4} n = 8 \cos \frac{\pi n}{2}$$ **Ans.**

(iii) The continuous-time signal $x(t)$ is sampled at f_s = 150 Hz. The frequency of discrete-time will be

$$F = \frac{f}{f_s} = \frac{100}{150} = \frac{2}{3}$$

Then, the discrete-time signal will be given as

$$x[n] = 8 \cos 2\pi f n$$

$$= 5 \cos 2\pi \left(\frac{2}{3}\right) n = 8 \cos \frac{4\pi}{3} n$$

$$= 8 \cos \left(2\pi - \frac{2\pi}{3}\right) n$$

$$= 8 \cos \frac{2\pi n}{3} = 8 \cos \frac{2\pi n}{3}$$ **Ans.**

(iv) For sampling rate of f_s = 150 Hz

$$F = \frac{f}{f_s} \text{ or } f = f_s \times F = \frac{1}{3} \times 150 = 50 \text{ Hz}$$

Then, the sinusoidal signal will be

$$y(t) = 8 \cos 2\pi f t = 8 \cos 2\pi \times 50 \times t = 8 \cos 100\,\pi t$$

Sampling at $f_s = 150$ Hz, yields identical samples hence $f = 100$ Hz is an alias of $f = 50$ Hz for sampling rate $f_s = 150$ Hz. **Ans.**

Example 9.4. Determine the Nyquist rate for a continuous-time signal

$$x(t) = 6 \cos 50\,\pi t + 20 \sin 300\,\pi t - 10 \cos 100\,\pi t$$

Solution: In a general form, any continuous-time signal may be expressed as

$$x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t + A_3 \cos \omega_3 t \qquad ...(i)$$

And the given signal is

$$x(t) = 6 \cos 50\pi t + 2 \sin 300\pi t - 10 \cos 100\,\pi t \qquad ...(ii)$$

On comparing given signal (*ii*) with standard form of a signal (*i*), we obtain the frequencies for the given signal as

$$f_1 = \frac{\omega_1}{2\pi} = \frac{50\pi}{2\pi} = 25 \text{ Hz}$$

$$f_2 = \frac{\omega_2}{2\pi} = \frac{300\pi}{2\pi} = 150 \text{ Hz}$$

$$f_3 = \frac{\omega_3}{2\pi} = \frac{100\pi}{2\pi} = 50 \text{ Hz}$$

Thus, the highest frequency component of the given message signal will be

$$f_{max} = 150 \text{ Hz}$$

Therefore, Nyquist rate $= 2 f_{max} = 2 \times 150 = 300$ Hz **Ans.**

9.5. Reconstruction Filter (Low-Pass Filter)

The low-pass filter is used to recover original signal from its samples. This is also known as **interpolation filter**.

A low-pass filter is that type of filter which passes only low frequencies upto a specified cut-off frequency and rejects all other frequencies above cut-off frequency. Figure 9.3 shows the frequency response of low-pass filter.

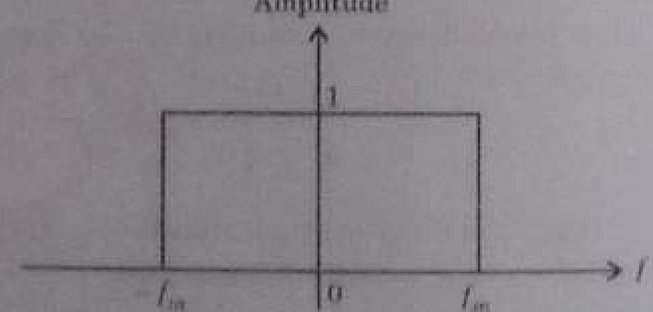


Fig. 9.3. *Ideal low-pass filter (LPF).*

From figure 9.3, it may be observed that in case of low-pass filter, there is sharp-change in response at cut-off frequency, that is amplitude response becomes suddenly zero at cut-off frequency which is not possible practically. This means that an ideal low-pass filter is not physically realizable. In place of ideal-low pass filter, we use practical filter.

Figure 9.4 shows the frequency response of practical low-pass filter. From figure 9.4, it may be observed that in case of practical filter, the amplitude response decreases slowly to become zero. This means that there is a transition band in case of practical filter. Figure 9.5 shows the use of practical low-pass filter in reconstruction of original signal from its sample.

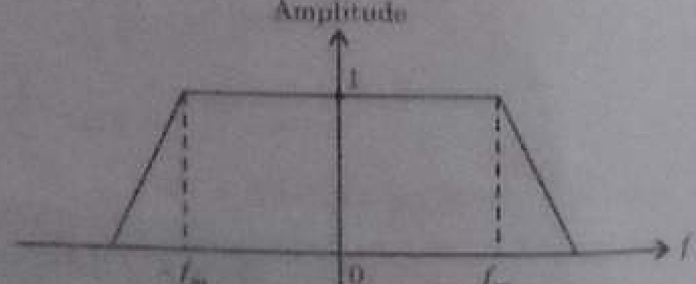


Fig. 9.4. *Practical low-pass filter (LPF).*

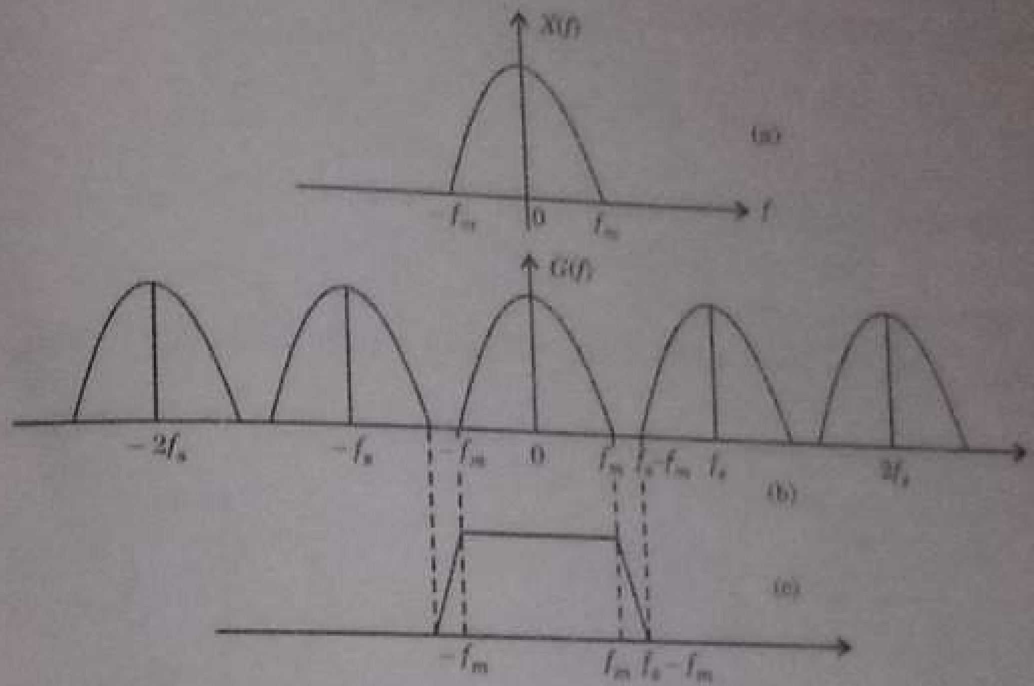


Fig. 9.5. *(a) Spectrum of original signal*
(b) Spectrum of sampled signal
(c) Amplitude response of practical low-pass filter.

9.6. Signal Reconstruction: The Interpolation Formula

The process of reconstructing a continuous-time signal $x(t)$ from its samples is called as **interpolation**.

As discussed earlier, a signal $x(t)$ bandlimited to f_m *Hz* can be reconstructed (interpolated) completely from its samples. This is achieved by passing the sampled signal through an ideal low-pass filter (LPF) of cut-off frequency f_m Hz.

The expression for sampled signal is written as

$$g(t) = x(t).\delta_{T_s}(t) \qquad ...(9.10)$$

$$g(t) = \frac{1}{T_s}\,[x(t) + 2\,x(t) \cos w_s t + 2\,x(t) \cos 2w_s t + ...] \qquad ...(9.11)$$

From above equation, it may be observed that the sampled signal contains a component $\frac{1}{T_s} \times x(t)$.

To recover $x(t)$ or $X(w)$, the sampled signal must be passed through an ideal low-pass filter of bandwidth of f_m Hz and gain T_s.

Therefore, the reconstruction or interpolating filter transfer function may be expressed as

$$H(w) = T_s \times \text{rect}\left(\frac{w}{4\pi f_m}\right) \qquad ...(9.12)$$

The impulse response $h(t)$ of this filter is the inverse Fourier transform of $H(w)$, *i.e.*

$$h(t) = F^{-1}[H(w)] = F^{-1}\left[T_s \,\text{rect}\left(\frac{w}{4\pi f_m}\right)\right]$$

$$h(t) = 2 f_m T_s \,\text{sinc}\,(2\pi f_m t) \qquad ...(9.13)$$

Assuming that sampling is done at Nyquist rate, then

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

So that $2f_m T_s = 1$ ...(9.13a)

Putting this value of $2f_m T_s$ in equation (9.13), we have

$$h(t) = 1 \cdot \text{sinc}(2\pi f_m t) = \text{sinc}(2\pi f_m t) \quad ...(9.14)$$

Figure 9.6(*b*) shows the graph of $h(t)$.

From figure, it may be observed that $h(t) = 0$ at all Nyquist sampling instants $t = \pm n/2f_m$ except at $t = 0$.

Now, when the sampled signal $g(t)$ is applied at the input of this filter, the output will be $x(t)$.

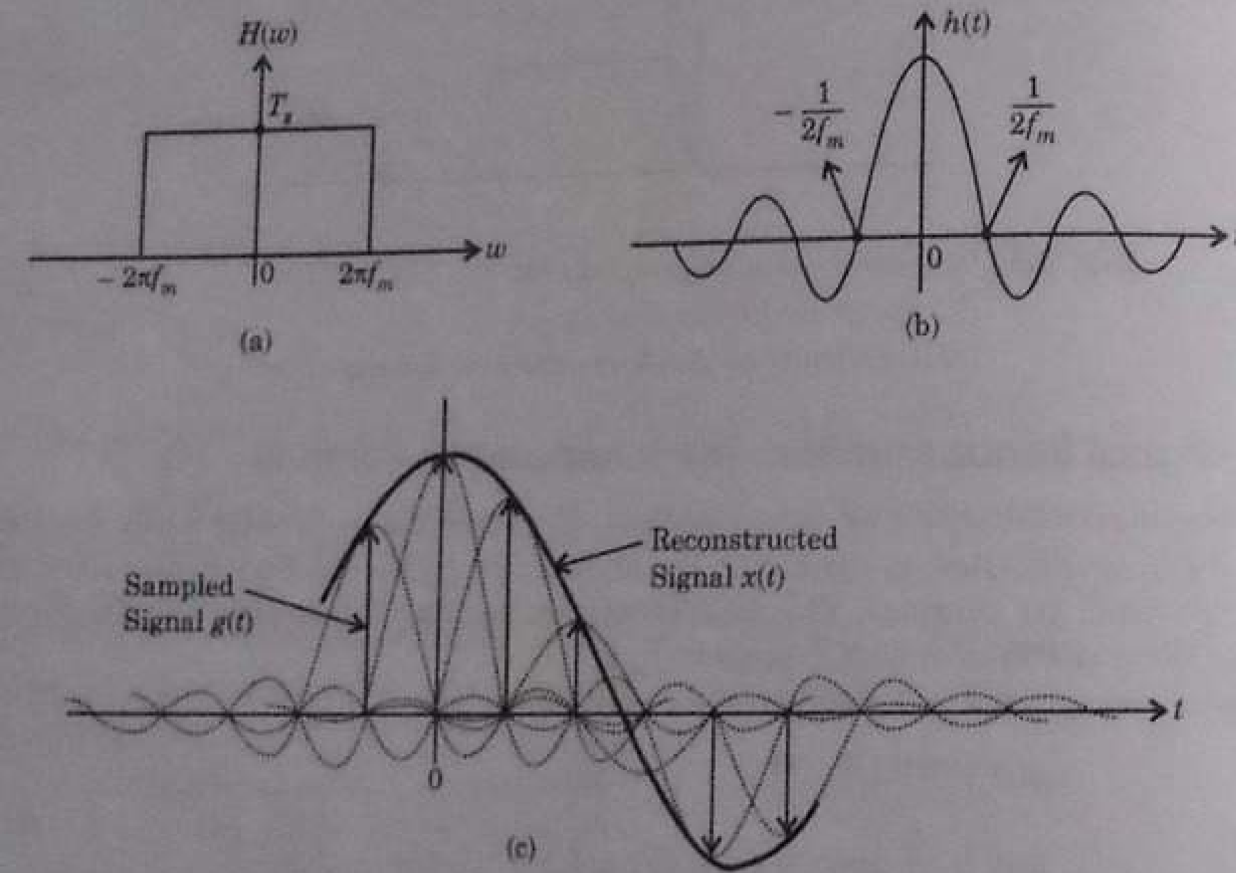


Fig. 9.6.

Each sample in $g(t)$, being an impulse, produces a sinc pulse of height equal to the strength of the sample.

Addition of the sinc pulses produced by all the samples results in $x(t)$.

For instant, the kth sample of the input $g(t)$ is the impulse $x(kT_s)\,\delta(t - kT_s)$.

The filter output of this impulse will be $x(kT_s)\,h(t - kT_s)$.

Therefore, the filter output to $g(t)$, which is $x(t)$, may be expressed as a sum

$$x(t) = \sum_k x(kT_s)\, h(t - kT_s) \quad ...(9.15)$$

$$= \sum_k x(kT_s)\, \text{sinc}[2\pi f_m (t - kT_s)] \quad ...(9.16)$$

$$x(t) = \sum_k x(kT_s)\, \text{sinc}(2\pi f_m - k\pi) \quad \left[\because T_s = \frac{1}{2f_m}\right] \quad ...(9.17)$$

Equation (9.17) is known as the **interpolation formula**, which provides values of $x(t)$ betw samples as a weighted sum of all the sample values.

In the proof of sampling theorem, it is assumed that the signal $x(t)$ is strictly bandlimi But, in general, an information signal may contain a wide range of frequencies and canno strictly bandlimited. This means that the maximum frequency f_m in the signal $x(t)$ canno predictable. Therefore, it is not possible to select suitable sampling frequency f_s.

9.7. Effect of Under Sampling: Aliasing

When a continuous-time bandlimited signal is sampled at a rate lower than Nyquist rate $f_s <$ then the successive cycles of the spectrum $G(\omega)$ of the sampled signal $g(t)$ overlap with each as shown in figure 9.7.

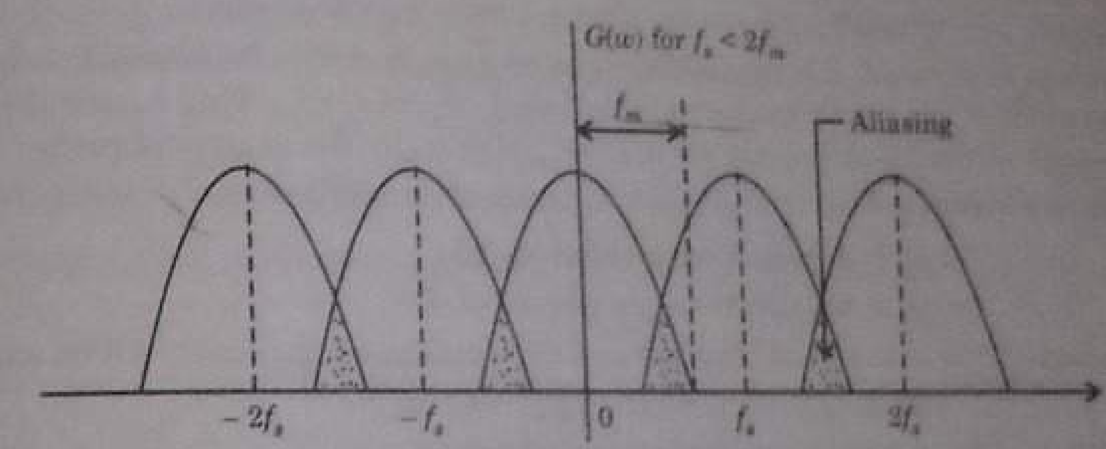


Fig. 9.7. *Spectruym of the sampled signal for the case $f_s < 2f_m$.*

Hence, the signal is under-slampled in this case ($f_s < 2f_m$) and some amount of alia produced in this under-sampling process. In fact, aliasing is the phenomenon in which frequency component in the frequency-spectrum of the signal takes identity of a lower-fre component in the spectrum of the sampled signal.

From figure 9.7, it is obvious that because of the overlap due to aliasing phenomenon, i possible to recover original signal $x(t)$ from sampled signal $g(t)$ by low-pass filtering si spectral components in the overlap regions add and hence the signal is distorted.

Since any information signal contains a large number of frequencies, so, to decide a sa frequency is always a problem. Therefore, a signal is first passed through a low-pass filte frequency is always a problem... low-pass filter blocks all the frequencies which are above f_m Hz. This process is known a low-pass filter blocks all the frequencies which are above f_m Hz. This low-pass filter is called **prelias filter** because it is limiting of the original signal $x(t)$. This low-pass filter is called prelias filter because it is prevent aliasing effect. After bandlimiting, it becomes easy to decide sampling frequenc the maximum frequency is fixed at f_m Hz.

In short, to avoid aliasing:

(*i*) Prealias filter must be used to limit band of frequencies of the signal to f_m Hz.

(*ii*) Sampling frequency 'f_s' must be selected such that

$$f_s > 2f_m$$

9.8. Sampling of Bandpass Signals

In previous sections, we discussed sampling theorem for low-pass signals. However, w given signal is a bandpass signal, then a different criteria must be used to sample the Therefore, the sampling theorem for bandpass signals may be expressed as under.

The bandpass signal $x(t)$ whose maximum bandwidth is $2f_m$ can be completely repr into and recovered from its samples if it is sampled at the minimum rate of twice the ban Here, f_m is the maximum frequency component present in the signal.

Hence if the bandwidth is $2f_m$, then the minimum sampling rate for bandpass signal must be f_m samples per second. Figure 9.8 shows the spectrum of an arbitrary bandpass signal.

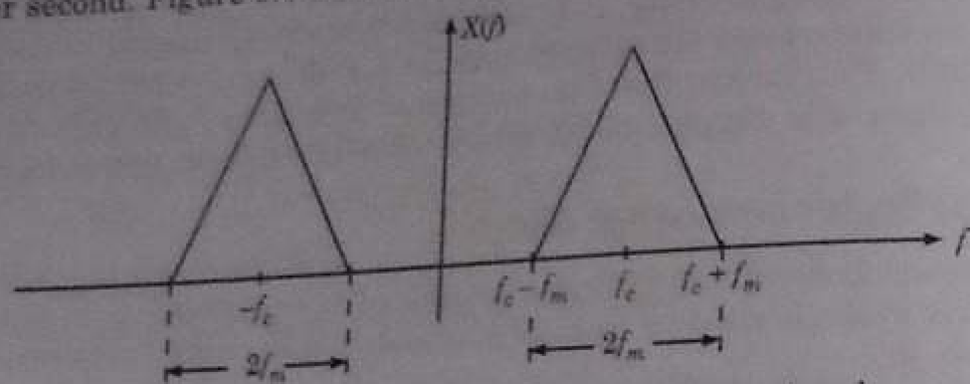


Fig. 9.8. *Spectrum of an orbitrary bandpass signal.*

The spectrum in figure 9.8 is centred around frequency f_c. The bandwidth is $2f_m$. Thus, the quencies present in the bandpass signal are from $f_c - f_m$ to $f_c + f_m$. This means that the highest quency present in the bandpass signal is $f_c + f_m$. Generally the centre frequency $f_c > f_m$.

This bandpass signal is first represented in terms of its inphase and quadrature components

Let $\quad x_I(t)$ = Inphase component of $x(t)$

and $\quad x_Q(t)$ = Quadrature component of $x(t)$

Thus, the signal $x(t)$ in terms of inphase and quadrature components will be expressed as

$$x(t) = x_I(t)\cos(2\pi f_c t) - x_Q(t)\sin(2\pi f_c t) \qquad \text{...(9.18)}$$

The inphase and quadrature components are obtained ultiplying $x(t)$ by $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ and then ressing the sum frequencies by means of low-pass filters. , inphase $x_I(t)$ and quadrature $x_Q(t)$ components contain low frequency components. The spectrum of these onents is limited between $-f_m$ to $+f_m$. This is shown in e 9.9.

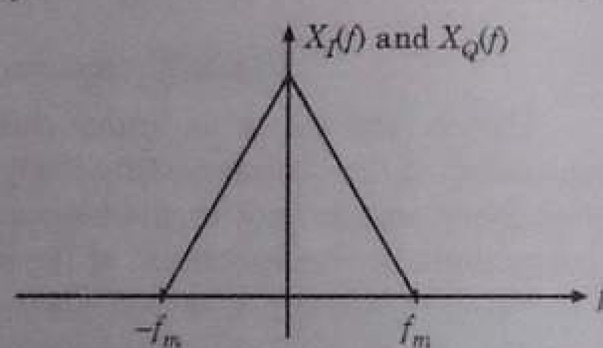


Fig. 9.9. *Spectrum of inphase and quadrature components of bandpass signal x(t).*

After few mathematical manipulation in equation (9.18), otain the reconstruction formula as

$$\sum_{n=-\infty}^{\infty} x\left(\frac{n}{4f_m}\right)\text{sinc}\left(2f_m t - \frac{n}{2}\right)\cos\left[2\pi f_c\left(t - \frac{n}{4f_m}\right)\right] \qquad \text{...(9.19)}$$

omparing this reconstruction formula with that of low-pass signals given in equation (9.17), serve that $x(t)$ is replaced by $x\left(\frac{n}{4f_m}\right)$.

ere, $\quad x\left(\frac{n}{4f_m}\right) = x(nT_s)$ = Sampled version of bandpass signal

d $\quad T_s = \frac{1}{4f_m}$

us, if $4f_m$ samples per second are taken, then the bandpass signal of bandwidth $2f_m$ can be etely recovered from its samples.

nce, for bandpass signals of bandwidth $2f_m$,

Minimum Sampling rate = Twice of bandwidth

$= 4f_m$ samples per second.

Example 9.5. Show that a bandlimited signal of finite energy which has no frequency components higher than f_m Hz is completely described by specifying values of the signals at instants of time separated by $1/2f_m$ seconds and also show that if the instantaneous values of the signal are separated by intervals larger than $1/2\,f_m$ seconds, they fail to describe the signal. A bandpass signal has spectral range that extends from 20 to 82 kHz.

Find the acceptable range of sampling frequency f_s.

Solution: Let $x(t)$ be the bandlimited signal which has no frequency components higher than f_m Hz. Let this signal be sampled by a sampling function given as

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

The sampling function is the train of impulses with T_s as distance between successive impulses. Let $x(nT_s)$ be the instantaneous amplitude of signal $x(t)$ at instant $t = T_s$. The sampled version of $x(t)$ may be given as multiplication of $x(nT_s)$ and $\delta_{T_s}(t)$ i.e.,

$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\,\delta(t - nT_s) \qquad \text{...}(i)$$

Now, Fourier transform of this sampled signal may be obtained as

$$G(f) = FT\{g(t)\} = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)$$

Here, f_s is the sampling rate which is given as

$$f_s = \frac{1}{T_s}$$

and $\quad X(f - nf_s) = X(f)$ at $nf_s = 0, \pm f_s, \pm 2f_s \pm 3f_s + \ldots$

Hence, the same spectrum $X(f)$ appears at $f = 0$,

$$f = \pm f_s,\ f = \pm 2f_s \ \text{ etc.,}$$

This means that a periodic spectrum with period equal to f_s is generated in frequency domai because of sampling $x(t)$ in time-domain.

Thus, equation (i) may be written as

$$G(f) = f_s \times (f) + f_s \times (f \pm f_s) + f_s \times (f \pm 2f_s) + f_s \times (f \pm 3f_s) + f_s \times (f \pm 4f_s) + \ldots \qquad \text{...(ii}$$

or $$G(f) = f_s \times (f) + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} f_s \times (f - nf_s)$$

By definition of Fourier transform, we have

$$X(f) = \int_{-\infty}^{\infty} x(t)\,e^{-j2\pi ft}\,dt$$

For sampled version of $x(t)$, we have $t = nT_s$.

Then the equation (iii) becomes

$$G(f) = \sum_{n=-\infty}^{\infty} x(nT_s)\,e^{-j2\pi fnT_s}$$

Now, given that the signal is bandlimited to f_m Hz and

$$T_s = \frac{1}{2f_m} \text{ seconds}$$

Therefore, $$f_s = \frac{1}{T_s} = 2f_m \quad ...(v)$$

From equation (ii), it may be observed that $G(f)$ is periodic with a period f_s.

Thus the spectrum $X(f)$ and $G(f)$ are shown in figure 9.10.

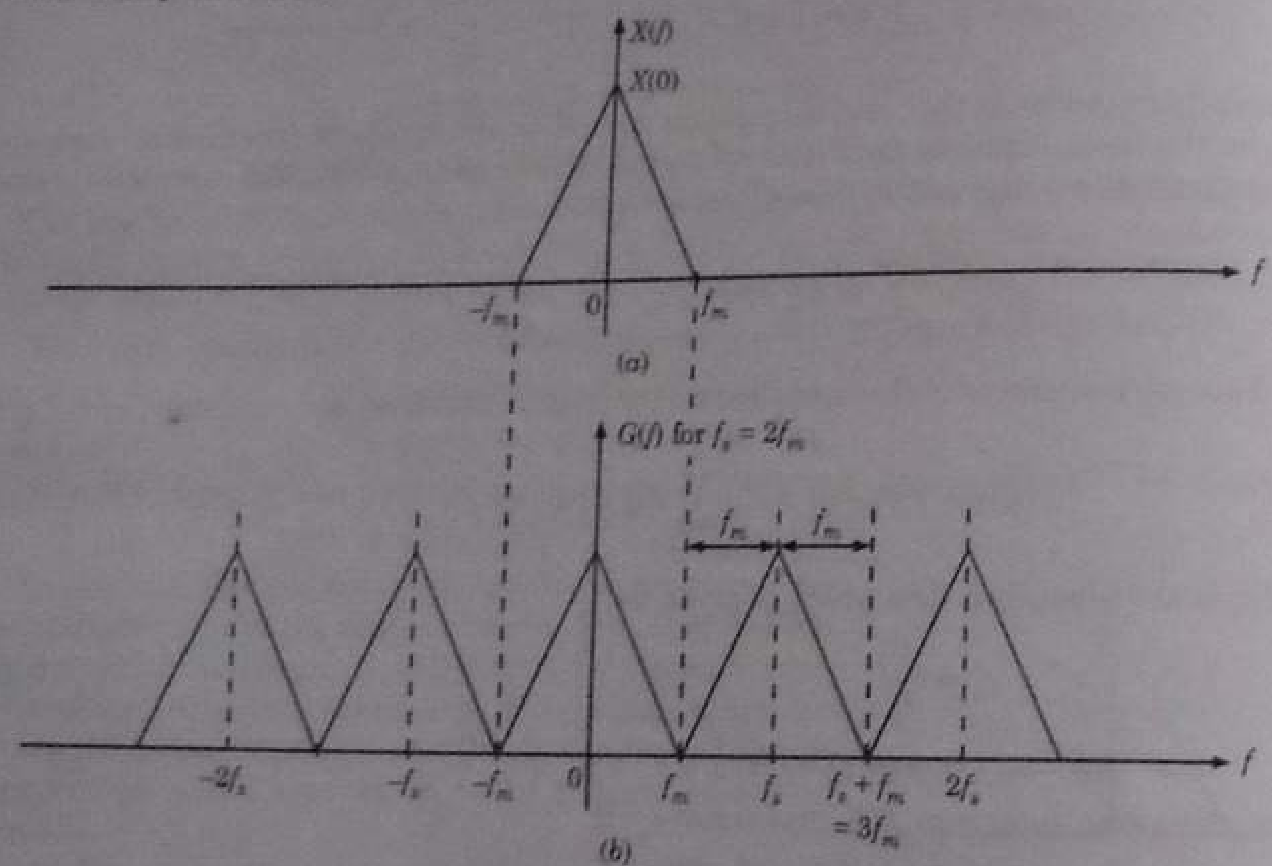


Fig. 9.10. *(a) Spectrum of x(t) (b) Spectrum of g(t) with $f_s = 2f_m$*

Now, since $f_s = 2f_m$

therefore, $f_s - f_m = f_m$ and $f_s + f_m = 3f_m$

Hence, the periodic spectrums $X(f)$ just touch $\pm f_m$, $\pm 3f_m$, $\pm 5f_m$ ... etc.

Thus, there is no aliasing.

Using equation (iii), we write

$$X(f) = \frac{1}{f_s} G(f) - \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} X(f - nf_s) \quad ...(vi)$$

Substituting $f_s = 2f_m$ in equation (vi), we get

$$X(f) = \frac{1}{2f_m} G(f) - \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} X(f - nf_s)$$

or $$X(f) = \frac{1}{2f_m} G(f) \quad \text{for } -f_m \leq f \leq f_m \quad ...(vii)$$

Now putting the value of $G(f)$ from equation (iv) to (vii), we get

$$X(f) = \frac{1}{2f_m} \sum_{n=-\infty}^{\infty} x(nT_s)\, e^{-j2\pi f n T_s}$$

Since $$T_s = \frac{1}{2f_m}$$

i.e., $$X(f) = \frac{1}{2f_m} \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) e^{-j\pi f n/f_m}$$

$x(t)$ may, be recovered from $X(f)$ by taking Inverse Fourier transform of last equation, i.e.

$$x(t) = IFT\left\{\frac{1}{2f_m} \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) e^{-j\pi f_n/f_m}\right\}$$

This equation indicates that $x(t)$ is represented completely by its samples $x\left(\frac{n}{2f_m}\right)$ for $-\infty < \infty$. This means that the sequence $x\left(\frac{n}{2f_m}\right)$ has all the information contained in $x(t)$.

Reconstruction of Signal form Samples

Let us consider equation (ix) as

$$x(t) = IFT\left\{\frac{1}{2f_m} \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) e^{-j\pi f n/f_m}\right\} = \int_{-f_m}^{f_m} \frac{1}{2f_m} \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) e^{-j\pi f n/f_m} e^{j2\pi}$$

Interchanging the order of summation and integration, we get

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) \frac{1}{2f_m} \int_{-f_m}^{f_m} e^{j2\pi f\left(t - \frac{n}{2f_m}\right)} dt$$

or $$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) \frac{\sin(2\pi f_m t - n\pi)}{(2\pi f_m t - n\pi)} = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) \frac{\sin \pi(2f_m t - n\pi)}{\pi(2f_m t - n)}$$

Since, $$\text{sinc}\, \theta = \frac{\sin \pi\theta}{\pi\theta}$$

Therefore, $$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2f_m}\right) \text{sinc}\,(2f_m t - n) \; -\infty < n < \infty$$

Hence, this is the interpolation formula to reconstruct $x(t)$ from its samples $x(nT_s)$. The from all above, it is clear that the signal may be completely represented into and recovere its samples if the spacing between the successive samples is $\frac{1}{2f_m}$ seconds i.e., $f_s = 2f_m$ samp second.

Sampling Frequency for Bandpass Signal

Since the spectral range of the bandpass signal is 20 kHz to 82 kHz

Therefore

$$\text{Bandwidth} = 2f_m = 82\text{ kHz} - 20\text{ kHz} = 62\text{ kHz}$$

Hence, Minimum Sampling rate = 2 × bandwidth = 2 × 62 = 124 kHz

Generally, the range of minimum sampling frequencies is specified for bandpass signals. It lies between $4f_m$ to $8f_m$ samples per second.

Therefore,

Range of minimum sampling frequencies

$$= (2 \times \text{bandwidth}) \text{ to } (4 \times \text{bandwidth})$$

$$= 2 \times 62\text{ kHz to } 4 \times 62\text{ kHz} = 124\text{ kHz to } 248\text{ kHz}$$ **Ans.**

9.9. Sampling Techniques

In the last article, we discussed how sampling of a continuous-time signal is done. This sampling of a signal is done in several ways. Therefore, in this section, we shall discuss different types of sampling *i.e.*, sampling techniques.

Basically, there are three types of sampling techniques as under:

(*i*) Instantaneous sampling

(*ii*) Natural sampling

(*iii*) Flat-top sampling.

Out of these three, instantaneous sampling is called ideal sampling whereas natural sampling and flat-top sampling are called practical sampling methods. Now, let us discuss three different types of sampling techniques in detail.

9.9.1. Ideal Sampling or Instantaneous Sampling or Impulse Sampling

In the proof of sampling theorem, we used ideal or impulse sampling. In this type of sampling, the sampling function is a train of impulses. Figure 9.11(*b*) shows this sampling function.

$x(t)$ is the input signal (*i.e.*, signal to be sampled) as shown in figure 9.11(*a*).

Figure 9.11(*c*) shows a circuit to produce instantaneous or ideal sampling. This circuit is known as the **switching sampler.**

The working principle of this circuit is quite easy. The circuit simply consists of a switch. Now if we assume that the closing time '*t*' of the switch approaches zero, then the output $g(t)$ of this circuit will contain only instantaneous value of the input signal $x(t)$. Since the width of the pulse approaches zero, the instantaneous sampling gives a train of impulses of height equal to the instantaneous value of the input signal $x(t)$ at the sampling instant.

We know that the train of impulses may be represented as

$$\delta T_s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad \text{...(9.20)}$$

This is known as sampling function and its waveform is shown in figure 9.11(*b*).

The sampled signal $g(t)$ is expressed as the multiplication of $x(t)$ ans $\delta_{T_s}(t)$.

Thus,

$$g(t) = x(t) \cdot \delta_{T_s}(t) \quad \text{...(9.21)}$$

$$= x(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad \text{...(9.22)}$$

or

$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \cdot \delta(t - nT_s) \quad \text{...(9.23)}$$

The Fourier transform of the ideally sampled signal given by above equation may be expressed as

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) \quad \text{...(9.24)}$$

Note: *This equation gives the spectrum of ideally sampled signal. It shows that the spectrum $X(f)$ is periodic in f_s and weighted by f_s. However, it may be noted that ideal or instantaneous sampling is possible only in theory since it is impossible to have a pulse whose width approaches zero. Ideal sampling was used in last article to prove sampling theorem. Practically flat-top sampling and nature sampling are used.*

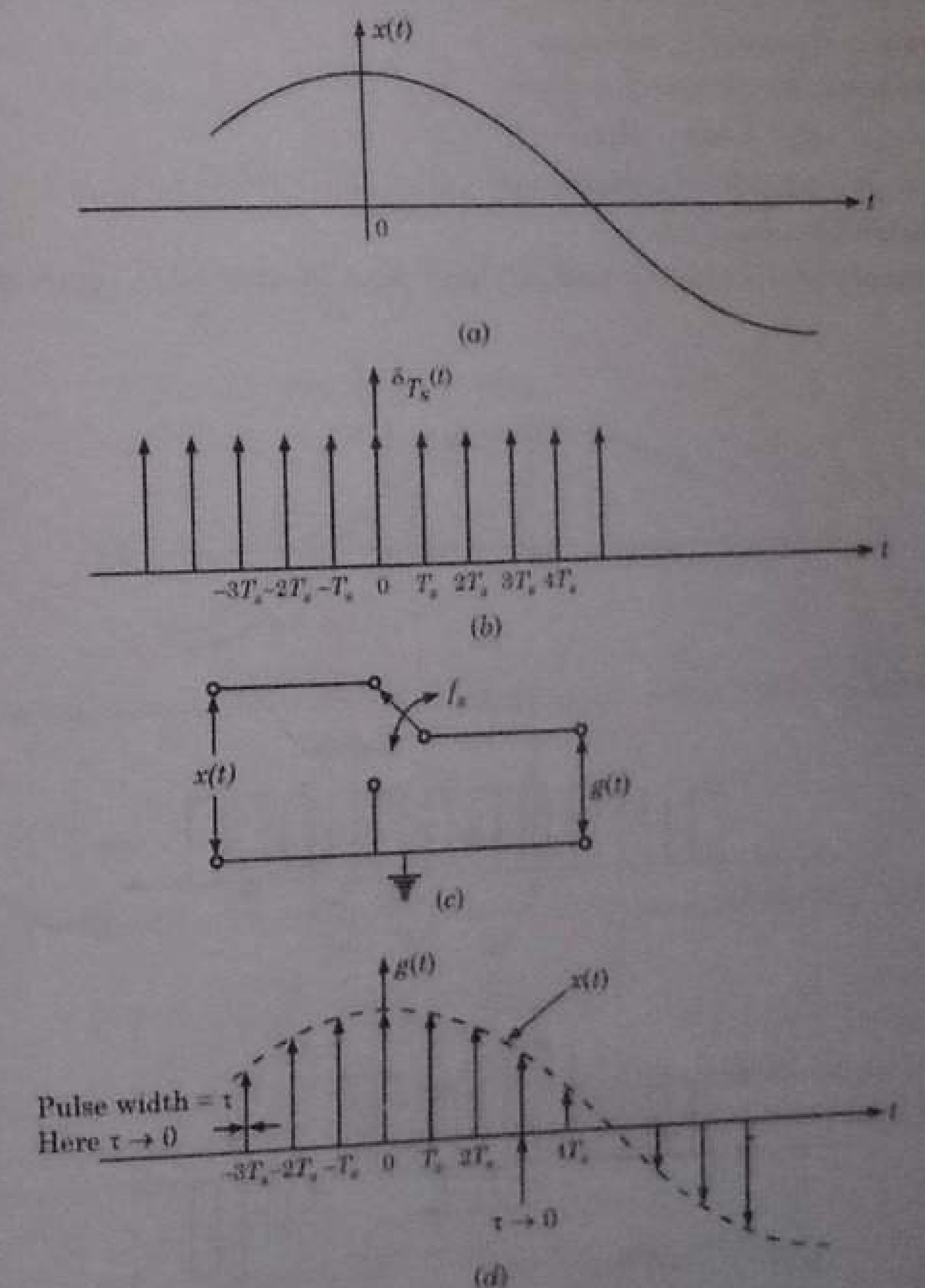


Fig. 9.11. (*a*) *Baseband signal* (*b*) *Impulse train* (*c*) *Functional diagram of a switching sampler* (*d*) *Sampled signal*

9.9.2. Natural Sampling

As discussed in last article, the instantaneous sampling results in the samples whose width τ approaches zero. Due to this, the power content in the instantaneously sampled pulse is negligible. Thus, this method is not suitable for transmission purpose. Natural sampling is a practical method and will be discussed in this section.

In natural sampling the pules has a finite width equal to τ.

Let us consider an analog continuous-time signal $x(t)$ to be sampled at the rate of f_s Hz. Here it is assumed that f_s is higher than Nyquist rate such that simpling theorem is satisfied.

Again, let us consider a sampling function $c(t)$ which is a train of periodic pulses of width τ and frequency equal to f_s Hz.

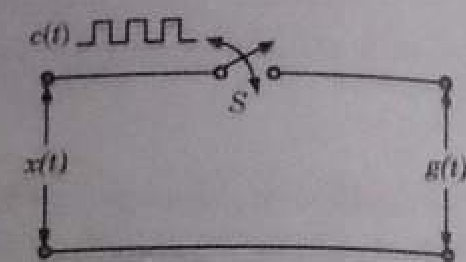


Fig. 9.12. *A functional diagram of a Natural Sampler.*

Figure 9.12 shows a functional diagram of a natural sampler. With the help of this natural sampler, a sampled signal $g(t)$ is obtained by multiplication of sampling function $c(t)$ and input signal $x(t)$.

Now, according to figure 9.12, we have

when $c(t)$ goes high the switch 'S' is closed.

Therefore $\quad g(t) = x(t) \quad$ when $\quad c(t) = A$ ...(9.25)

and $\quad g(t) = 0 \quad$ when $\quad c(t) = 0$ ...(9.26)

where A is the amplitude of $c(t)$.

The waveforms of signals $x(t)$, $c(t)$ and $g(t)$ have been illustrated in figure 9.13(a), (b) and (c) respectively.

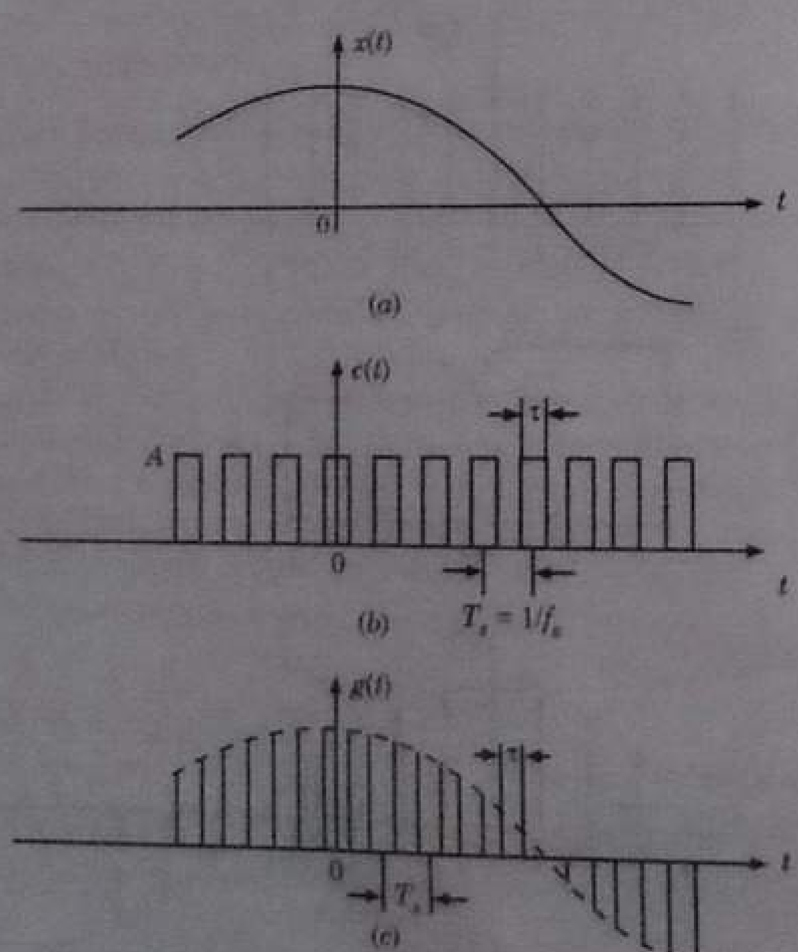


Fig. 9.13. *(a) Continuous time signal x(t).*
(b) Sampling function waveform i.e. periodic pulse train
(c) Naturally sampled signal waveform s(t).

Now, the sampled signal $g(t)$ may also be described mathematically as

$$g(t) = c(t) \quad x(t) \qquad ...(9.27)$$

Here, $c(t)$ is the periodic train of pulse of width τ and frequency f_s.

We know that the Exponential Fourier series for any periodic waveform is expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} C_n e^{j2\pi nt/T_0} \qquad ...(9.28)$$

Also, for the periodic pulse train of $c(t)$, we have

$$T_0 = T_s = \frac{1}{f_s} = \text{period of } c(t)$$

or

$$f_0 = f_s = \frac{1}{T_0} = \frac{1}{T_s} = \text{frequency of } c(t)$$

Therefore according to equation (9.28) for periodic pulse train $c(t)$, we have

$$c(t) = \sum_{n=-\infty}^{\infty} C_n . e^{j2\pi f_s nt} \text{ with } \frac{1}{T_0} = f_s \qquad ...(9.29)$$

Now, it may be noted that since $c(t)$ is a rectangular pulse train, therefore C_n for this waveform will be expressed as

$$C_n = \frac{TA}{T_0} \sin c\,(f_n . T) \qquad ...(9.30)$$

here $\quad T$ = pulse width = τ

and $\quad f_n$ = harmonic frequency

But here, $\quad f_n = nf_s \quad$ or $\quad f_s = \frac{n}{T_0} = nf_0$

Hence, $$C_n = \frac{\tau . A}{T_s} \cdot \sin c(f_n \tau) \qquad ...(9.31)$$

Therefore, using equations (9.29) and (9.30) the Fourier series representation for $c(t)$ will be given as

$$c(t) = \sum_{n=-\infty}^{\infty} \frac{\tau . A}{T_s} \sin c(f_n \tau)\, e^{j2\pi f_s nt} \qquad ...(9.32)$$

Now, substituting the value of $c(t)$ from equation (9.32) to equation (9.27), we get

$$g(t) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \sin c(f_n \tau) . e^{j2\pi f_s nt} . x(t) \qquad ...(9.33$$

This is required time-domain representation for naturally sampled signal $g(t)$.

Now, to get the frequency-domain representation of the naturally sampled signal $g(t)$, let u[s] take its Fourier transform as

$$G(f) = FT[g(t)] = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{si nc}\,(f_n \tau)\, FT[e^{j2\pi f_s nt} x(t)] \qquad ...(9.3$$

Recall the frequency-shifting property of Fourier transform which states that

$$e^{j2\pi f_s n(t)} . x(t) \longleftrightarrow X(f - f_s n) \qquad ...(9.3$$

Therefore,
$$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{sinc}\,(f_s\,\tau)\, X(f - f_s n) \quad ...(9.36)$$

Now, since $f_s = f_s n$ = harmonic frequency

Therefore, equation (9.36) becomes
$$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{sinc}\,(nf_s\,\tau)\, X(f - nf_s) \quad ...(9.37)$$

Hence, we write

Spectrum of Naturally Sampled signal:
$$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{sinc}\,(nf_s\,\tau)\, X(f - nf_s) \quad ...(9.38)$$

This equation shows that the spectra of $x(t)$ *i.e.*, $X(f)$ are periodic in f_s and are weighed by the sin c function.

Figure 9.14. illustrates some arbitray spectra for $x(t)$ and corresponding spectrum $G(f)$.

Note: *Thus form figure 9.14, it may be noted that unlike the spectrum of instantaneously sampled signal shown in figure 9.1(f), the spectrum of a naturally sampled signal is weighted by a sin c function whereas the spectrum of an instantaneously sampled signal (figure 9.1(f)) remains constant throughout the frequency range.*

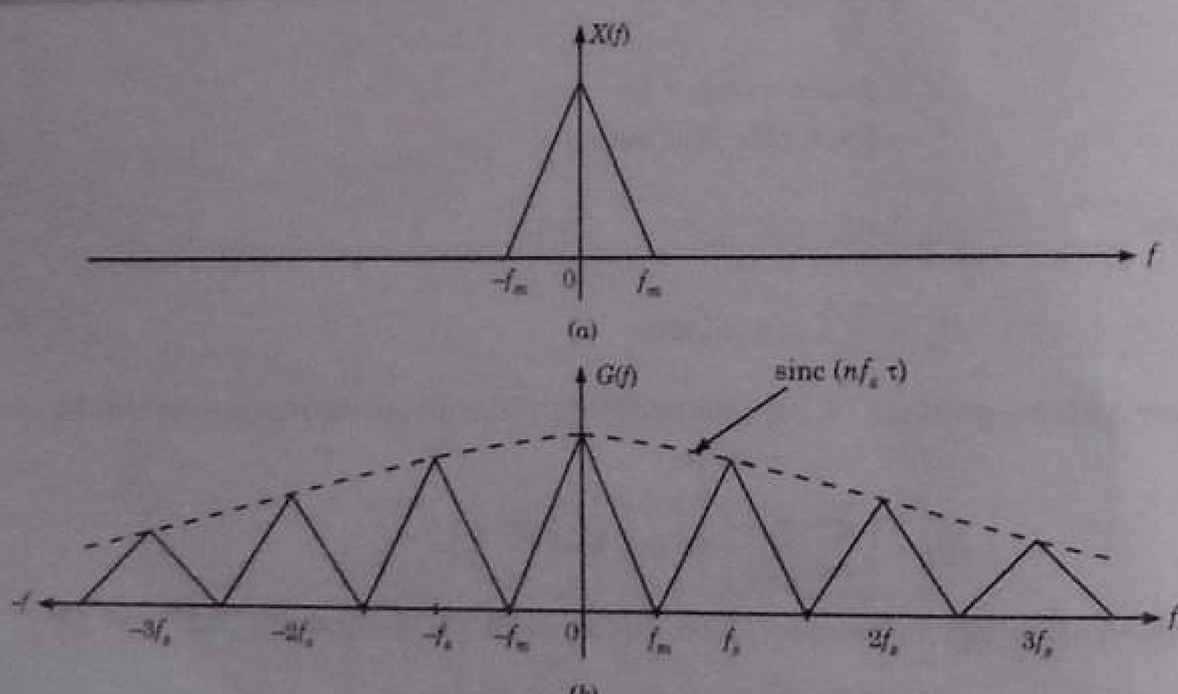


Fig. 9.14. *(a) Spectrum of continuous-time signal x(t).*
(b) Spectrum of naturally sampled signal

9.9.3. Flat Top Sampling or Rectangular Pulse Sampling

(U.P. Tech., Semester Examination, 2003-2004)

Flat top sampling like natural sampling is also a practically possible sampling method. But natural sampling is little complex whereas it is quite easy to get flat top samples.

In flat-top sampling or rectangular pulse sampling, the top of the samples remains constant and is equal to the instantaneous value of the baseband signal $x(t)$ at the start of sampling. The duration or width of each sample is τ and sampling rate is equal to $f_s = \frac{1}{T_s}$. Figure 9.15(a) shows

the functional diagram of a sample and hold circuit which is used to generate the flat top samples. Figure 9.15(b) shows the general waveform of flat top samples.

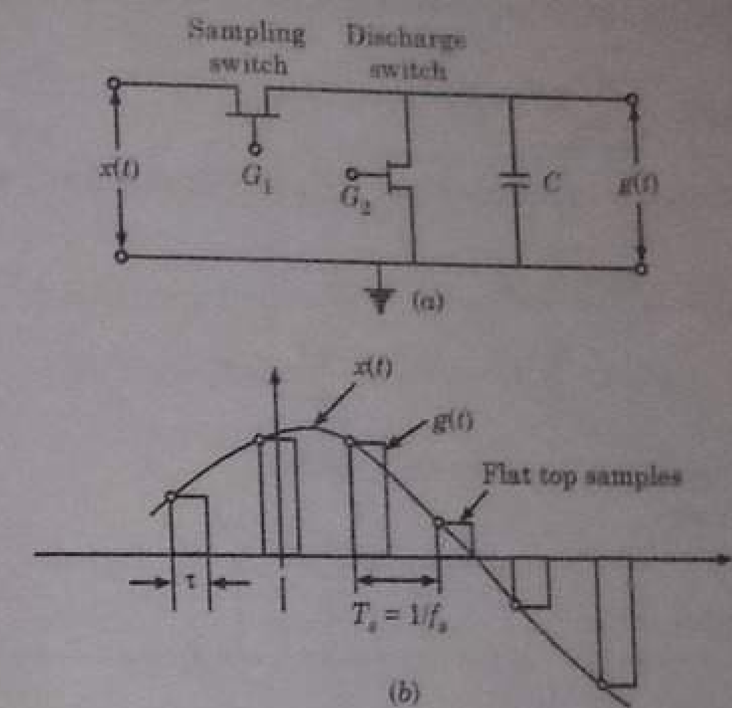


Fig. 9.15. *(a) A sample and hold circuit to generate flat top samples.*
(b) A general waveform of flat top sampling.

From figure 9.15(b), it may be noted that only starting edge of the pulse represents instantaneou value of the baseband signal $x(t)$. Also the flat top pulse of $g(t)$ is mathematically equivalent t the convolution of instantaneous sample and a pulse $h(t)$ as depicted in figure 9.16.

This means that the width of the pulse in $g(t)$ is determined by the width of $h(t)$ and th sampling instant is determined by delta function.

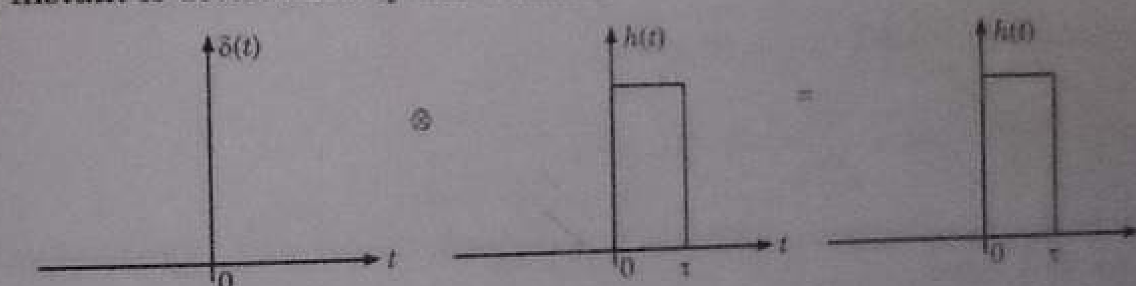


Fig. 9.16. *Convolution of any function with delta function is equal to that function.*

In figure 9.15(b), the starting edge of the pulse represents the point where baseband signal sampled and width is determined by function $h(t)$. Therefore $g(t)$ will be expressed as
$$g(t) = s(t) \otimes h(t) \quad ...(9.3$$

This equation has been explained in figure 9.17.

Now, from the property of delta function, we know that for any function $f(t)$
$$f(t) \otimes \delta(t) = f(t) \quad ...(9.4$$

This property is used to obtain flat top samples. It may be noted that to obtain flat top sampli we are not applying the equation (9.40) directly here *i.e.*, we are applying a modified form equation (9.40). This modified equation is equation (9.39).

Thus, in this modified equation, we are taking $s(t)$ in place of delta function $\delta(t)$. Observe th $\delta(t)$ is a constant amplitude delta function whereas $s(t)$ is a varying amplitude train of impuls This means that we are taking $s(t)$ which is an istantaneously sampled signal and this is convolv with function $h(t)$ as in equation (9.39).

Therefore, on convolution of $s(t)$ and $h(t)$, we get a pulse whose duration is equal to $h(t)$ only but amplitude is defined by $s(t)$.

Now, we know that the train of impulses may be represented mathematically as

$$\delta_{T_s}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad ...(9.41)$$

The signal $s(t)$ is obtained by multiplication of baseband signal $x(t)$ and $\delta_{T_s}(t)$.

Thus,

$$s(t) = x(t) \cdot \delta_{T_s}(t) \quad ...(9.42)$$

$$= x(t) \sum_{n=-\infty}^{\infty} \delta(t - nTs) = \sum_{n=-\infty}^{\infty} x(nT_s) \cdot \delta(t - nT_s) \quad ...(9.43)$$

Now, sampled signal $g(t)$ is given as equation (9.39)

$$g(t) = s(t) \otimes h(t) \quad ...(9.44)$$

or

$$g(t) = \int_{-\infty}^{\infty} s(\tau) h(t - \tau)\, d\tau \quad ...(9.45)$$

or

$$g(t) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT_s)\, \delta(\tau - nT_s) h(t - \tau)\, d\tau = \sum_{n=-\infty}^{\infty} x(nT_s) \int_{-\infty}^{\infty} \delta(\tau - nT_s) h(t - \tau)\, d\tau \quad ...(9.46)$$

According to shifting property of delta function, we know that

$$\int_{-\infty}^{\infty} f(t)\, \delta(t - t_0) = f(t_0) \quad ...(9.47)$$

Using equations (9.46) and (9.47), we get

$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\, h(t - nT_s)$$

This equation represents value of $g(t)$ in terms of sampled value $x(nT_s)$ and function $h(t - nT_s)$ for flat top sampled signal.

Now, again from equation (9.39), we have

$$g(t) = s(t) \otimes h(t)$$

Taking Fourier transform of both sides of above equation, we get

$$G(f) = S(f)\, H(f) \quad ...(9.48)$$

We know that $S(f)$ is given as

$$S(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) \quad ...(9.49)$$

Therefore, equation (9.48) becomes

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)\, H(f) \quad ...(9.50)$$

Thus, spectrum of flat top sampled signal:

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)\, H(f) \quad (9.51)$$

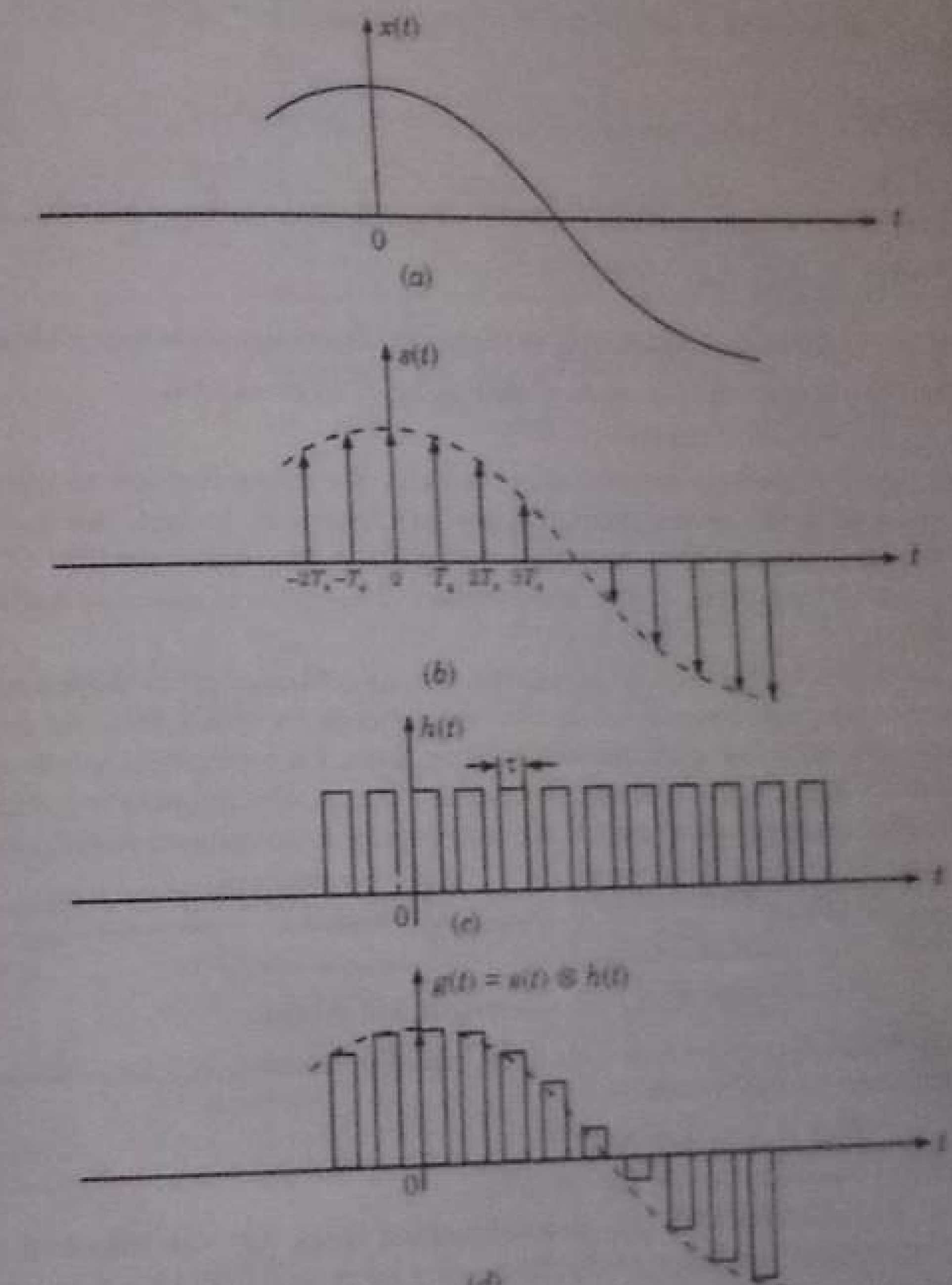


Fig. 9.17. *(a) Baseband signal x(t) (b) Instantaneously sample single s(t) (c) Constant pulse width function h(t) (d) Flat top sampled signal g(t) obtained through convolution of h(t) and s(t)*

9.10. Aperture Effect

The spectrum of flat top sampled signal is exprussed as

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)\, H(f) \quad ...(9.52)$$

This equation shows that the signal $g(t)$ is obtained by passing the signal $s(t)$ through a filter having transfer function $H(f)$. The corresponding impulse response $h(t)$ in time-domain has been shown in figure 9.18(a). This is one pulse of rectangular pulse train shown in figure 9.17(c). Each sample of $x(t)$ [i.e., $s(t)$] is convolved with this pulse. Equation (9.52) represents that the spectrum of this rectangular pulse is multiplied with that of $s(t)$.

Figure 9.18(b) shows the spectrum of one rectangular pulse of $h(t)$.

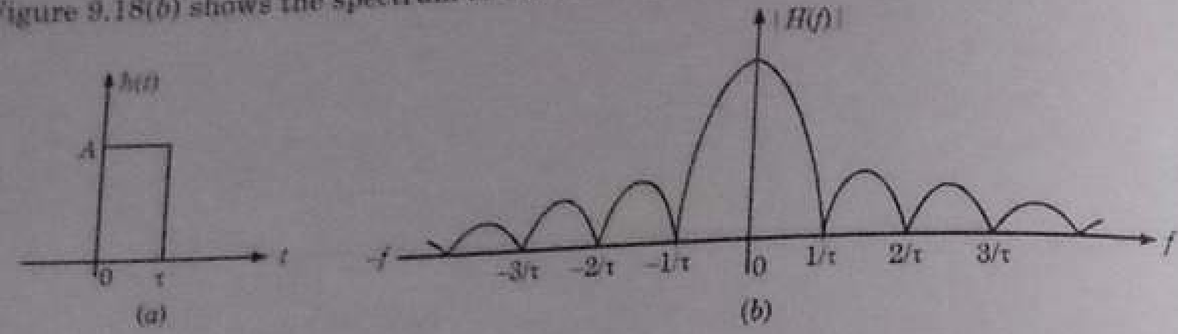


Fig. 9.18. *(a) One pulse of rectangular pulse train (b) Spectrum of the pulse shown in figure (a).*

We know that the spectrum of a rectangular pulse is expressed as

$$H(f) = \tau.\text{sin}\,c(f \cdot t)\, e^{-j\pi f t} \quad [\because A = 1] \qquad ...(9.53)$$

Hence, from figure 9.18(b), it may be observed that by using flat top samples an amplitude distortion is introduced in the reconstructed signal $x(t)$ from $g(t)$. In fact, the high frequency roll-off of $H(f)$ acts like a low-pass filter and thus attenuates the upper portion of message signal spectrum. These high frequencies of $x(t)$ are affected. This type of effect is known as **aperture effect**.

Now, as the duration 'τ' of the pulse increases, the aperture effect is more prominent. Hence, during reconstruction an equalizer is needed to compensate for this effect. As depicted in figure 9.19, the receiver contains a low-pass reconstruction filter with cutoff frequency slightly higher than the maximum frequency present in the message signal. The equalizer compensates for the aperture effect. It also compensates for the attenuation by a low-pass reconstruction filter.

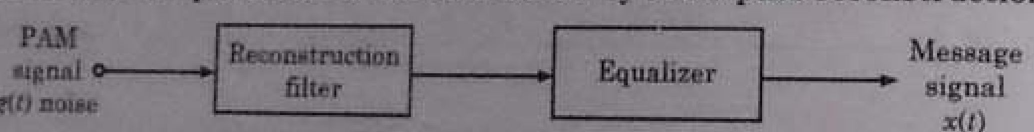


Fig. 9.19. *Recovering x(t) at receiver.*

From equation (9.53), it may be noted that the sample function $h(t)$ acts like a low-pass filter where Fourier transform as expressed as

$$H(f) = \tau.\ \text{sinc}\,(f.\tau)\, e^{-j\pi f t} \qquad ...(9.54)$$

This spectrum has been plotted in figure 9.18.

Equalizer used in casecade with the reconstruction filter has the effect of decreasing the inband loss of the reconstruction filter as the frequency increases in such a way as to compensate for the aperture effect.

Also, the transfer function of the equalizer is expressed as

$$H_{eq}(f) = \frac{K \cdot e^{-j2\pi f t_d}}{H(f)} \qquad ...(9.55)$$

Here 't_d' is known as the delay introduced by low-pass filter which is equal to τ/2.

Therefore,
$$H_{eq}(f) = \frac{K\ e^{-j\pi f \tau}}{\tau \text{si nc}\,(f\tau)\, e^{-j\pi f \tau}} \qquad ...(9.56)$$

or
$$H_{eq}(f) = \frac{K}{\tau\ \text{si nc}\,(f\tau)} \qquad ...(9.57)$$

which is the transfer function of an equalizer.

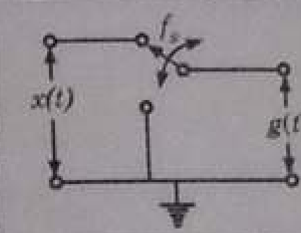
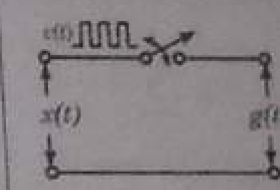
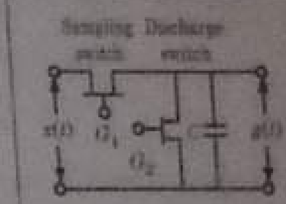
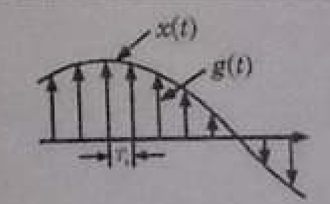
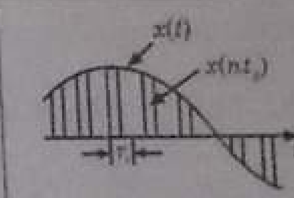
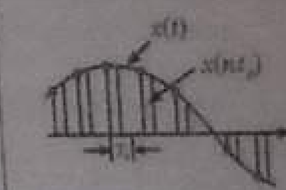
9.11. Comparison of Various Sampling Techniques

We can compare various sampling techniques on the basis of their method, noise interference and spectral properties etc.

The Table 9.1 lists some of the important points of comparison of generation.

Table 9.1. **Comparison of three sampling techniques**

(U.P. Tech., Semester, Examination, 2003-04)

S. No.	*Parameter of comparison*	*Ideal or instantaneous sampling*	*Natural sampling*	*Flat top sampling*
1.	Sampling principle	It uses multiplication	It uses chopping principle	It uses sample and hold circuit
2.	Generation circuit			
3.	Waveforms involved			
4.	Feasibility	This is not a practically possible method	This method is used practically	This method is also used practically
5.	Sampling rate	Sampling rate tends to infinity	Sampling rate satisfies Nyquist criteria	Sampling rate satisfies Nyquist criteria
6.	Noise interference	Noise interference is maximum	Noise interference is minimum	Noise interference is maximum
7.	Time domain representation	$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s)$	$g(t) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} x(t) \text{sin}\,c(nf_s\tau)\, e^{j2\pi n f_s t}$	$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s) h(t - nT_s)$
8.	Frequency domain representation	$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)$	$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{sin}\,c(nf_s\tau) X(f - nf_s)$	$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) H(f)$

Example 9.6. Figure 9.20 shows the spectrum of an arbitary signal $x(t)$. This signal is sampled at the Nyquist rate with a periodic train of rectangular pulses of duration $\frac{50}{3}$ milliseconds.

Determine the spectrum of the sampled signal for frequencies upto 50 Hz giving relevant expression.

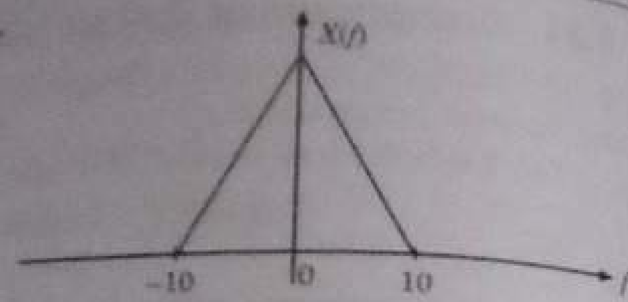


Fig. 9.20.

Solution: From figure 9.20, it may be observed that the signal is bandlimited to 10 Hz

thus $f_m = 10$ Hz

So the Nyquist rate is $= 2 f_m$

$= 2 \times 10 = 20$ Hz

Since the signal is sampled at the Nyquist rate, the sampling frequency would be

$f_s = 20$ Hz

Given that the rectangular pulses are used for sampling *i.e.*, flat top sampling is used. The spectrum of the flat top sampled signal is given as

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) H(f) \quad \text{...(i)}$$

Value of $H(f)$ is expressed as

$$H(f) = \tau \operatorname{sinc}(f\tau)\, e^{-j\pi f\tau} \quad \text{...(ii)}$$

Here τ is the width of the rectangular pulse used for sampling.

The given value of rectangular sampling pulse duration is $\frac{50}{3}$ milliseconds *i.e.*,

$$\tau = \frac{50}{3} \times 10^{-3} = \frac{0.05}{3} \text{ sec}$$

Substituting the value of τ in equation (*ii*), we get

$$H(f) = \frac{0.05}{3} \operatorname{sinc}\left(\frac{0.05f}{3}\right) e^{-j0.05\pi f/3}$$

Again, putting this value of $H(f)$ and f_s in equation (*i*), we get

$$G(f) = 20 \sum_{n=-\infty}^{\infty} X(f - 20n) \frac{0.05}{3} \operatorname{sinc}\left(\frac{0.05f}{3}\right) e^{-j0.05\pi f/3} \quad (\because f_s = 20)$$

$$G(f) = \frac{1}{3} \sum_{n=-3}^{3} X(f - 20n) \operatorname{sinc}\left(\frac{0.05f}{3}\right) e^{-j0.05\pi f/s}$$

This expression gives the spectrum up to 60 Hz (since $n = \pm 3$) for the sampled signal. **Ans.**

ample 9.7. A flat top sampling system samples a signal of maximum frequency 1 Hz with Hz sampling frequency. The duration of the pulse is 0.2 seconds. Compute the amplitude ortion due to aperture effect at the highest signal frequency. Also determine the equalization racteristic.

ution: Given that sampling frequency

$f_s = 2.5$ Hz

Maximum signal frequency,

$f_{max} = 1$ Hz

nd pulse width $\tau = 0.2$ seconds.

e know that the aperture effect is expressed by a transfer function $H(f)$ as

$$H(f) = \tau \operatorname{sinc}(f\tau)\, e^{-j\pi f\tau}$$

The magnitude of this equation will be

$$|H(f)| = \tau \operatorname{sinc}(f\tau)$$

$$|H(f)| = 0.2 \operatorname{sinc}(f \times 0.2) \quad \text{...(i)}$$

Now, aperture effect at the highest frequency will be obtained by putting $f = f_{max} = 1$ Hz in equation (*i*) *i.e.*,

$$|H(1)| = 0.2 \operatorname{sinc}(0.2) = 0.18709$$

or $|H(1)| = 18.70\%$ **Ans.**

Also, the equalizer characteristic is expressed as

$$H_{eq}(f) = \frac{K}{\tau \operatorname{sinc}(f\tau)}$$

Substituting, $\tau = 0.2$ second and assuming $K = 1$, the last equation becomes

$$H_{eq}(f) = \frac{1}{0.2 \operatorname{sinc}(0.2f)}$$

This equation is the plot of $H_{eq}(f)$ versus f and it represents the equalization characteristic to overcome aperture effect.

9.12. Analog Pulse Modulation Methods

We know that in analog modulation systems, some parameter of a sinusoidal carrier is varied according to the instantaneous value of the modulating signal. In pulse modulation methods, the carrier is no longer a continuous signal but consists of a pulse train. Some parameter of which is varied according to the instantaneous value of the modulating signal. There are two types of pulse modulation systems as under:

(*i*) Pulse Amplitude Modulation (PAM)

(*ii*) Pulse Time Modulation (PTM)

In pulse amplitude modulation (PAM), the amplitude of the pulses of the carrier pulse train is varied in accordance with the modulating signal whereas in pulse time modulation (PTM), the timing of the pulses of the carrier pulse train is varied.

They are two types of PTM as under:

(*i*) Pulse width modulation (PWM)*

(*ii*) Pulse position modulation (PPM)

In Pulse width modulation, the width of the pulses of the carrier pulse train is varied in accordance with the modulating signal whereas in Pulse position modulation (PPM), the position of pulses of the carrier pulse train is varied. Figure 9.21 shows three types of pulse analog modulation methods.

According to the sampling theorem, if a modulating signal is bandlimited to f_m Hz,** the sampling frequency must be at least $2f_m$ Hz and, hence the freqiency of the carrier pulse train must also be least $2f_m$ Hz.

At this point, it may be noted that all the above pulse modulation methods (*i.e.*, PAM, PWM and PPM) are called analog Pulse modulation methods because the modulating signal is analog in nature in PAM, PWM and PPM.

* Pulse width modulation is also known as Pulse duration modulation (PDM)

** If a signal is said to be bandlimited to f_m Hz, then it means that the maximum frequency component in this signal is f_m Hz.

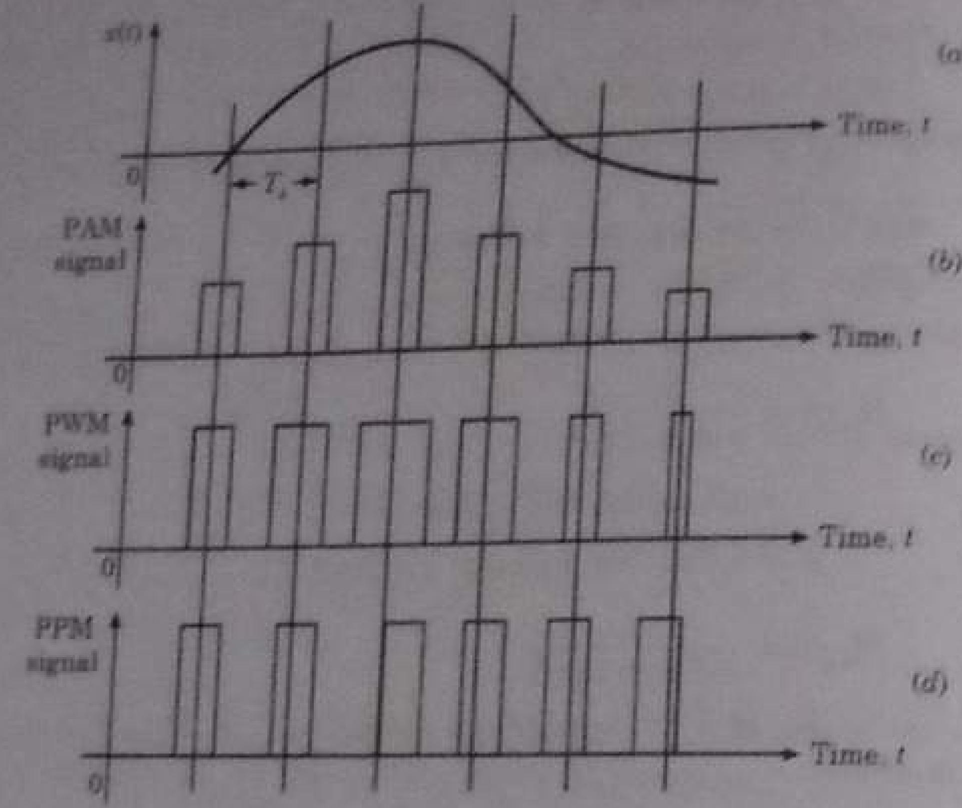


Fig. 9.21. *Different types of pulse analog modulation methods.*

9.13. Pulse Amplitude Modulation (PAM)

Pulse amplitude modulation may be defined as that type of modulation in which the amplitudes of regularly spaced rectangular pulses vary according to instantaneous value of the modulating or message signal. In fact, the pulses in a PAM signal may be of flat top type or natural type or ideal type. Actually all the sampling methods which have been discussed in last sections are basically pulse amplitude modulation methods.

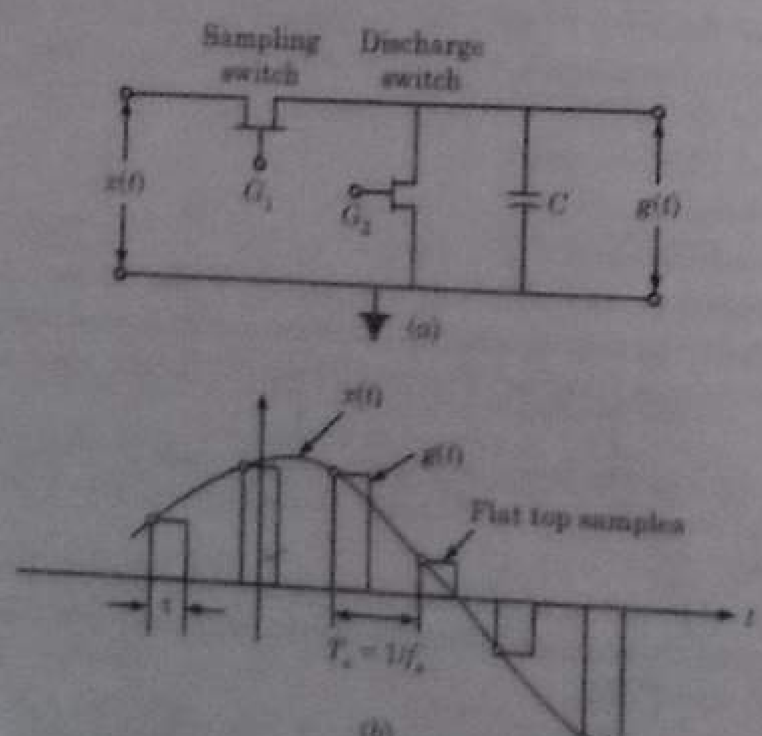


Fig. 9.22. *(a) Sample and hold circuit generating flat top sampled PAM (b) Waveforms of flat top sampled PAM*

Out of these three pulse amplitude modulation methods, the Flat top PAM is most popular and is widely used. The reason for using flat top PAM is that during the transmission, the noise interferes with the top of the transmitted pulses and this noise can be easily removed if the PAM pulse has flat top.

However, in case of natural samples PAM signal, the pulse has varying top in accordance with the signal variation. Now, when such type of pulse is received at the receiver, it is always contaminated by noise. Then it becomes quite difficult to determine the shape of the top of the pulse and thus amplitude detection of the pulse is not exact. Due to this, errors are introduced in the received signal.

Therefore, flat top sampled PAM is widely used.

Figure 9.22 shows the sample and hold circuit to produce flat top sampled PAM and the waveform for flat top sampled PAM.

Working Principle

A sample and hold circuit shown in figure 9.22 is used to produce flat top sampled PAM. The working principle of this circuit is quite easy. The sample and hold (S/H) circuit consists of two field

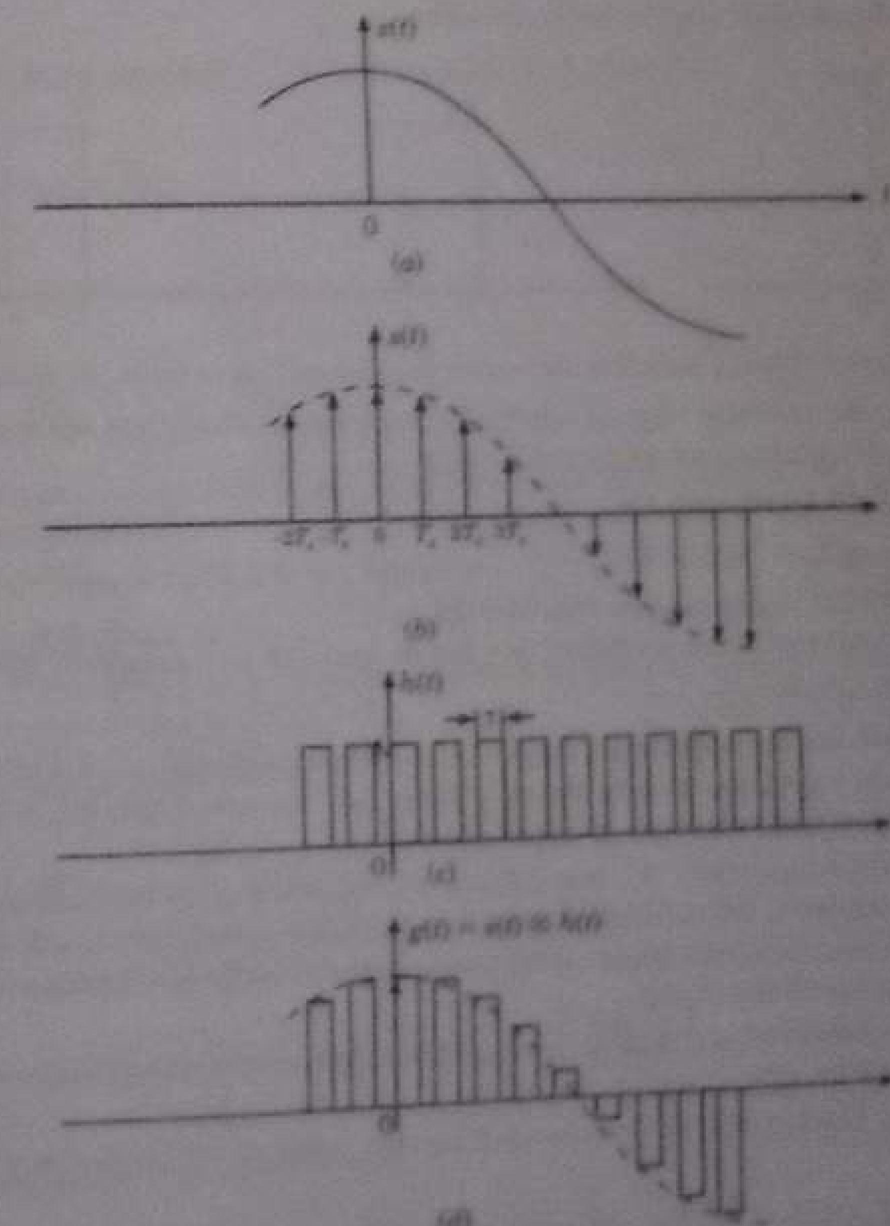


Fig. 9.23. *(a) Baseband signal x(t) (b) Instantaneously sample single s(t) (c) Constant pulse width function h(t) (d) Flat top sampled PAM signal g(t) obtained through convolution of h(t) and s(t)*

effect transistors (FET) switches and a capacitor. The sampling switch is closed for a short duration by a short pulse applied to the gate G_1 of the transistor. During this period, the capacitor 'C' is quickly charged upto a voltage equal to the instantaneous sample value of the incoming signal $x(t)$. Now, the sampling switch is opened and the capacitor 'C' holds the charge. The discharge switch is then closed by a pulse applied to gate G_2 of the other transistor. Due to this, the capacitor 'C' is discharged to zero volts. The discharges switch is then opened and thus capacitor has no voltage.

Hence, the output of the sample and hold circuit consists of a sequence of Flat top samples as shown in figure 9.23.

Mathematical Analysis

In a Flat top PAM, the top of the samples remains constant and is equal to the instantaneous value of the baseband signal $x(t)$ at the start of sampling. The duration or width of each sample is τ and sampling rate is equal to $f_s = \frac{1}{T_s}$. From figure 9.22 (*b*), it may be noted that only starting edge of the pulse represents instantaneous value of the baseband signal $x(t)$. Also, the flat top pulse of $g(t)$ is mathematically equivalent to the convolution of instantaneous sample and a pulse $h(t)$ as depicted in figure 9.24.

This means that the width of the pulse in $g(t)$ is determined by the width of $h(t)$ and the sampling instant is determined by the delta function.

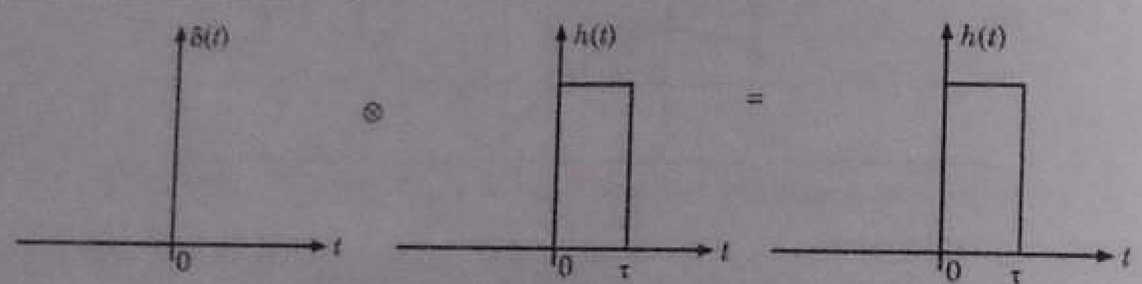


Fig. 9.24. *Convolution of any function with delta function is equal to that function.*

In figure 9.22 (*b*), the starting edge of the pulse represents the point where baseband signal is sampled and width is determined by function $h(t)$.

Therefore, $g(t)$ will be expressed as

$$g(t) = s(t) \otimes h(t) \qquad \ldots(9.58)$$

This equation has been explained in figure 9.23.

Now, from the property of delta function, we know that for any function $f(t)$

$$f(t) \otimes s(t) = f(t) \qquad \ldots(9.59)$$

This property is used to obtain flat top samples. It may be noted that to obtain flat top sampling, we are not applying the equation (9.59) directly here *i.e.*, we are applying a modified from of equation (9.59). This modified equation is equation (9.58).

Thus, in this modified equation, we are taking $s(t)$ in place of delta functions $\delta(t)$. Observe that $\delta(t)$ is a constant amplitude delta function whereas $s(t)$ is a varying amplitude train of impulses. This means that we are taking $s(t)$ which is an instantaneously sampled signal and this is convolved with function $h(t)$, as in equation (9.58).

Therefore, on convolution of $s(t)$ and $h(t)$, we get a pulse whose duration is equal to $h(t)$ only but amplitude is defined by $s(t)$.

Now, we know that the train of impulses may be represented mathematically as

$$\delta_{T_s}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \qquad \ldots(9.60)$$

The signal $s(t)$ is obtained by multiplication of baseband signal $x(t)$ and $\delta_{T_s}(t)$.

Thus,
$$s(t) = x(t) \cdot \delta_{T_s}(t) \qquad \ldots(9.61)$$

or
$$s(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{\infty} x(nT_s)\, \delta(t - nT_s) \qquad \ldots(9.62)$$

Now, sampled signal $g(t)$ is given as equation (9.58)

$$g(t) = s(t) \otimes h(t) \qquad \ldots(9.63)$$

or
$$g(t) = \int_{-\infty}^{\infty} s(\tau)\, h(t-\tau)\, d\tau \qquad \ldots(9.64)$$

or
$$g(t) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT_s)\, \delta(\tau - nT_s)\, h(t-\tau)\, d\tau$$

or
$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \int_{-\infty}^{\infty} \delta(\tau - nT_s)\, h(t-\tau)\, d\tau \qquad \ldots(9.65)$$

According to shifting property of delta function we know that

$$\int_{-\infty}^{\infty} f(t)\, \delta(t - t_0) = f(t_0) \qquad \ldots(9.66)$$

Using equations (9.65) and (9.66), we get

$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\, h(t - nT_s)$$

This equation represents value of $g(t)$ in terms of sampled value $x(nT_s)$ and function $h(t - nT_s)$ for flat top sampled signal.

Now, again from equation (9.58), we have

$$g(t) = s(t) \otimes h(t)$$

Taking Fourier transform of both sides of above equation, we get

$$G(f) = S(f)\, H(f) \qquad \ldots(9.67)$$

We know that $S(f)$ is given as

$$S(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) \qquad \ldots(9.68)$$

Therefore, equation (9.67) will become

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)\, H(f) \qquad \ldots(9.69)$$

Thus, spectrum of flat top PAM signal:

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)\, H(f) \qquad \ldots(9.70)$$

Here, $H(f)$ is the Fourier transform of the rectangular pulse. The spectrum of this rectangular pulse is shown in figure 9.18(*b*). Let the spectrum of $s(t)$ be the rectangular pulse train as shown in figure 9.25(*a*) and the spectrum of $h(t)$ *i.e.*, $H(f)$ is shown in figure 9.25(*b*).

By equation (9.67), we know that

$$G(f) = S(f) \cdot H(f)$$

Thus, according to above equation, we can plot the spectrum $G(f)$ as shown in figure 9.25(*b*).

Note: *It may be observed in figure 9.25(b) that higher frequencies in $S(f)$ are attenuated due to roll-off characteristics of the 'sinc' pulse. This effect is popularly known as* ***aperture effect.***

An equalizer is needed to overcame this effect.

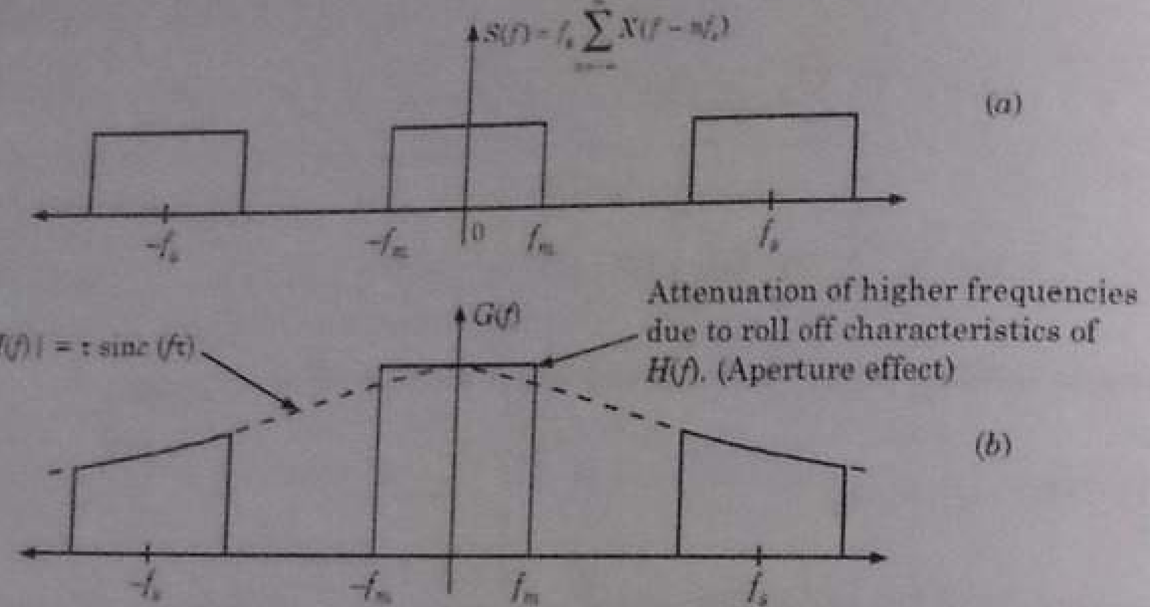


Fig. 9.25. (*a*) *Spectrum of some arbitrary signal. The signal is sampled at f_s and maximum frequency in the signal is f_m,*
(*b*) *Spectrum of flat top signal. The dotted curve is $H(f) = \tau \sin c\,(f\tau)$*

.1. Naturally Sampled Pulse Amplitude Modulated (PAM) Signal

We have discussed natural sampling in article 9.9. This natural sampling is basically pulse litude modulation (PAM). Therefore, it is called naturally sampled PAM signal.

Thus, time-domain representation of a naturally-sampled PAM signal will be given as

$$g(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} x(t).\text{sin}\,c(f_n\tau)\, e^{j2\pi nf_s t} \qquad ...(9.71)$$

nd the frequency-domain representation, *i.e.* frequency-spectrum of a naturally-sampled signal will be given as

$$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{sin}\,c\,(nf_s\tau) \cdot X(f - nf_s) \qquad ...(9.72)$$

. Instantaneous or Idealy Sampled Pulse Amplitude Modulated (PAM) Signal

e have discussed ideal or instantaneous sampling in article 9.9. This instantaneous sampling cally pulse amplitude modulation (PAM). Therefore, it is called ideally or instantaneously d PAM signal.

us, time-domain representation of a ideally or instantaneously sampled PAM signal will n as

$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \cdot \delta(t - nT_s) \qquad ...(9.73)$$

and the frequency-domain representation i.e., frequency-spectrum of a ideally or instantaneously sampled PAM sinal will be given as

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) \qquad ...(9.74)$$

9.13.3. Transmission Bandwidth in Pulse Amplitude Modulation (PAM)

In a pulse amplitude modulated (PAM) signal the pulse duration 'τ' is considered to be very very small in comparison to time period (*i.e.*, sampling period) T_s between any two samples *i.e.*,

$$\tau << T_s \qquad ...(9.75)$$

Now, if the maximum frequency in the modulating signal $x(t)$ is f_m, then according to sampling theorem, the sampling frequency f_s must be equal to or higher than the Nyquist rate, *i.e.*

or $$f_s \geq 2f_m \qquad ...(9.76)$$

or $$\frac{1}{T_s} \geq 2f_m \qquad \left(\because f_s = \frac{1}{T_s}\right)$$

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

But according to equation (9.75), we have

$$\tau << T_s$$

therefore $$\tau << T_s \leq \frac{1}{2f_m} \qquad ...(9.7$$

Now, if the 'ON' and 'OFF' time of the pulse amplitude modulated (PAM) pulse is same as shown in figure 9.26(*a*) then maximum frequency of the PAM pulse will be

$$f_{max} = \frac{1}{\tau + \tau} = \frac{1}{2\tau} \qquad ...(9.78)$$

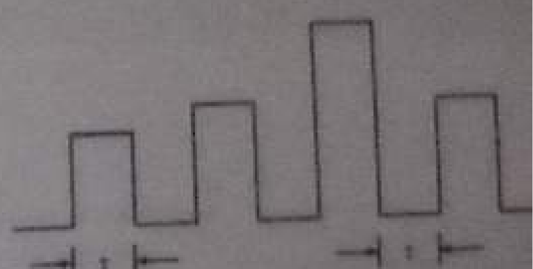


Fig. 9.26. (*a*) *Illustration of maximum frequency in PAM signa*

Therefore, the bandwidth required for the transmission of a PAM signal would be equal to the maximum frequency f_{max} given by the equation (9.78)

Thus, we have

Transmission bandwidth

$$BW \geq f_{max} \qquad ...(9.$$

But $$f_{max} = \frac{1}{2\tau}$$

Hence $$BW \geq \frac{1}{2\tau}$$

Again, since $$\tau << \frac{1}{2f_m} \qquad ...(9.$$

Therefore $$BW \geq \frac{1}{2\tau} >> f_m \qquad ...(9.$$

or $$BW >> f_m$$

9.13.4. Demodulation of PAM Signals

As discussed earlier, demodulation is the reverse process of modulation in which the modulating signal is recovered back from a modulated signal. For pulse-amplitude modulated (PAM) signals, the demodulation is done using a holding circuit. Figure 9.26(*b*) shows the block diagram of a PAM demodulator.

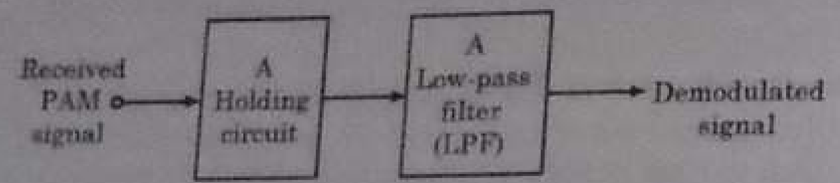


Fig. 9.26. *(b) A block diagram of PAM demodulator*

In this method, the received PAM signal is allowed to pass through a holding circuit and a low ass filter (LPF) as shown in above figure. Now, figure 9.27(*a*) illustrates a very simple holding rcuit. Here the switch '*S*' is closed after the arrival of the pulse and it is opened at the end of the lse. In this way, the capacitor *C* is charged to the pulse amplitude value and it holds this value ring the interval between the two pulses. Hence, the sampled values are held as shown in figure 7(*c*). After this the holding circuit output is smothened in Low Pass filter as shown in figure 7(*c*). It may be observed that some kind of distortion is introduced due to the holding circuit. In the circuit of figure 9.27(*b*) is known as **zero-order holding circuit.** This zero-order holding uit considers only the previous sample to decide the value between the two pulses.

ote: It may be noted the first order hold circuit considers the previous two samples whereas a cond order holding circuit considers the previous three samples and so on. However, as the der of the holding circuit increases, the distortion decreases at the cost of the circuit complexity. fact, the amount of permissible distortion decides the order of the holding circuit.

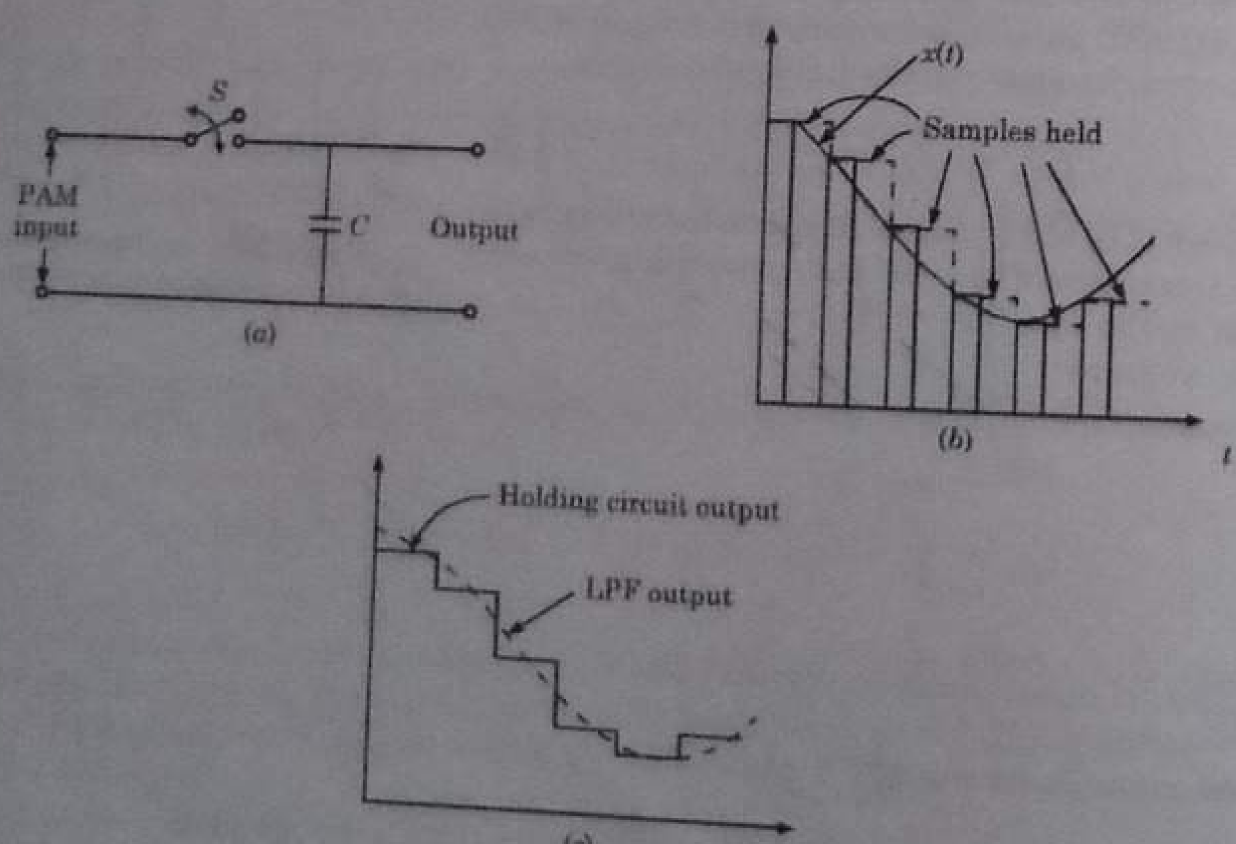


Fig. 9.27. *(a) A zero-order holding circuit (b) The output of holding circuit (c) The output of a low pass filter (LPF)*

9.13.5. Transmission of PAM Signals

If the PAM signals are to be transmitted directly *i.e.*, over a pair of wires then no further signal processing is necessary. However, if they are to be transmitted through the space using an antenna, they must first be amplitude or frequency or phase modulated by a high frequency carrier and only then they can be transmitted. Thus, the overall system will be then known as PAM-AM or PAM-FM or PAM-PM respectively. At the receiving end, AM or FM or PM detection is first employed to get the PAM signal and then the message signal is recovered from it.

Example 9.8. For a pulse-amplitude modulated (PAM) transmission of voice signal having maximum frequency equal to $f_m = 3$ kHz, calculate the transmission bandwidth. It is given that the sampling frequency $f_s = 8$ kHz and the pulse duration $\tau = 0.1 T_s$.

Solution: We know that the sampling period T_s is expressed as

$$T_s = \frac{1}{f_s} = \frac{1}{8\times 10^3} \text{ seconds}$$

$$T_s = 0.125 \times 10^{-3} \text{ seconds} = 125\ \mu \text{ seconds} \quad ...(i)$$

Also, τ is given that

$$\tau = 0.1\ T_s$$

Using (*i*), we get

$$\tau = 0.1 \times 125 = 12.5\ \mu \text{ seconds} \quad ...(ii)$$

Now, we know that the transmission bandwidth for PAM signal is expressed as

$$BW \geq \frac{1}{2\tau}$$

Using equation (*ii*), we get

$$BW \geq \frac{1}{2\times 12.5\times 10^{-6}} \geq \frac{1\times 10^6}{25}$$

$$BW \geq 40 \text{ kHz} \qquad \textbf{Ans.}$$

9.13.6. Drawbacks of Pulse-Amplitude Modulated (PAM) Signal

Following are the drawbacks of a PAM signal:

(*i*) The bandwidth required for the transmission of a PAM signal is very very large in comparison to the maximum frequency present in the modulating signal.

(*ii*) Since the amplitude of the PAM pulses varies in accordance with the modulating signal therefore the interference of noise is maximum in a PAM signal. This noise cannot be removed easily.

(*iii*) Since the amplitude of the PAM signal varies, therefore, this also varies the peak power required by the transmitter with modulating signal.

9.14. Pulse Time Modulation

In pulse time modulation, the signal to be transmitted is sampled as in pulse amplitude modulation (PAM). In pulse time modulation, amplitude of pulse is held constant, whereas position of pulse or width of pulse is made proportional to the amplitude of signal at the sampling instant. There are two types of pulse time modulation, viz. Pulse Width Modulation (PWM) and Pulse Position Modulation (PPM). Because in both PWM and PPM, amplitude is held constant and does not carry any information, therefore amplitude limiters can be used. The amplitude limiters, similar to those used in FM, will clip off the portion of the signal corrupted by noise and hence provide a good degree of noise immunity.

9.14.1. Pulse Width Modulation

Let us first discuss Pulse Width Modulation (PWM). This is also known as Pulse Duration Modulation (PDM). Three variations of pulse width modulation are possible. In one variation, the leading edge of the pulse is held constant and change in pulse width with signal is measured with respect to the leading edge. In other variation, the tail edge is held constant and with respect to it, pulse width is measured. In the third variation, centre of the pulse is held constant and pulse width changes on either side of the centre of the pulse. This has been illustrated in figure 9.28.

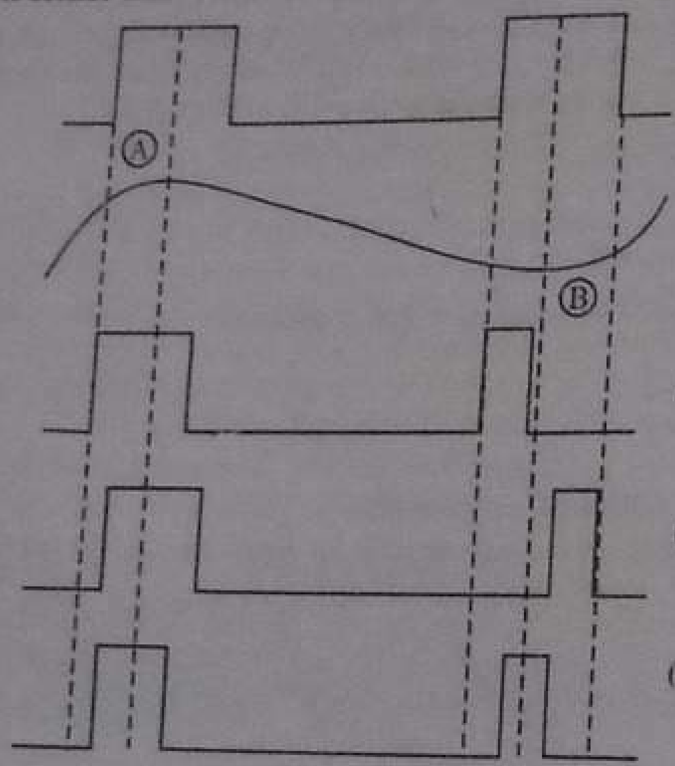


Fig. 9.28. *PWM waveforms*

The modulating signal is at its positive peak at point (A) and at its negative peak at (B). In figure 9.28 (*a*), the leading edge of pulse is kept constant and pulse width is measured from the lead edge. As shown, pulse width is maximum corresponding to point (A), while it is minimum at point (B).

In figure 9.28 (*b*), the tail edge of the pulse is kept constant and pulse width is measured from the tail end of the pulse. As before, pulse width is maximum corresponding to positive peak of the modulating signal and minimum at the negative peak.

As shown in figure 9.28 (*c*), the center of the pulse is kept constant and pulse extends on either side of the center of the pulse, depending upon the modulating signal.

9.14.2. Frequency Spectrum for PWM Wave

With a sinusoidal modulating signal at frequency f_m, the spectrum of PWM signal consists the modulating signal frequency f_m along with several harmonics. This is shown in figure 9.29.

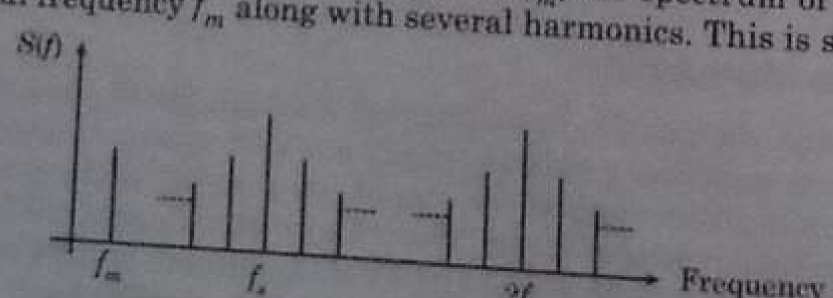
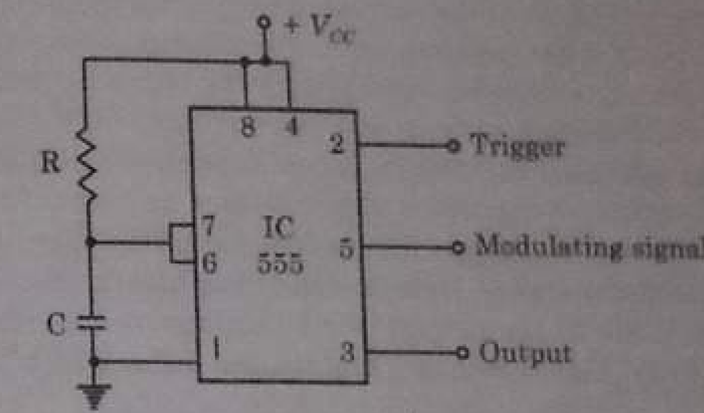


Fig. 9.29. *Spectrum of PWM signal*

To have a better separation with respect to frequency, between highest frequency of baseband signal [in Fig. 9.29, f_m] and lower sidebands of f_s (sampling frequency), a higher sampling frequency which is more than Nyquist rate is used; and pulse width deviation is kept small.

9.14.3. Modulation of PWM Signal or PWM Generation

Figure 9.30 shows pulse width modulator. It is basically a monostable multivibrator with a modulating input signal applied at the control voltage input. Internally, the control voltage is adjusted to the 2/3 V_{CC}. Externally applied modulating signal changes the control voltage, and hence the threshold voltage level. As a result, the time period required to charge the capacitor up to threshold voltage level changes, giving pulse modulated signal at the output, as shown in the figure 9.30 (*b*).



(*a*)

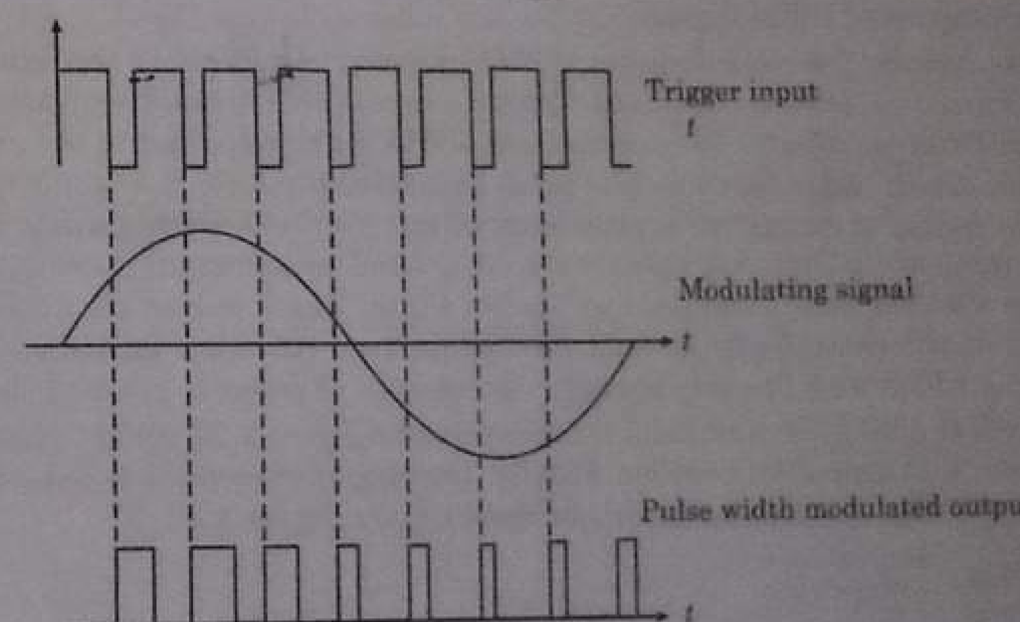


Fig. 9.30. (*b*)

Figure 9.31 illustrates another monostable multivibrator circuit to generate pulse width modulation (PWM).

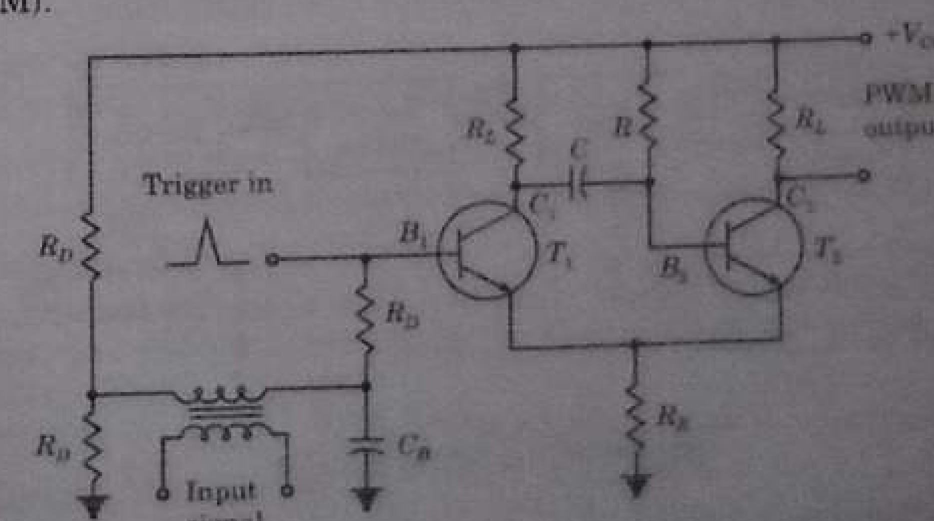


Fig. 9.31. *Monostable multivibrator generating pulse width modulation (PWM).*

The stable state for above circuit is achieved when T_1 is OFF and T_2 is ON. The positive going trigger pulse at B_1 switches T_1 ON. Because of this, the voltage at C_1 falls as T_1 now begins to draw the collector current. As a result, voltage at B_2 also falls and T_2 is switched OFF. C begins to charge up to the collector supply voltage (V_{CC}) through resistor R. After a time determined by the supply voltage and the RC time constant of the charging network, the base of the T_2 becomes sufficiently positive to switch T_2 ON. The transistor T_1 is simultaneously switched OFF by regenerative action and stays OFF until the arrival of the next trigger pulse. To make T_2 ON, the base of the T_2 must be slightly more positive than the voltage across resistor R_E. This voltage depends on the emitter curent I_E which is controlled by the signal voltage applied at the base of transistor T_1. Therefore, the changing voltage necessary to turn OFF transistor T_2 is decided by the signal voltage. If signal voltage is maximum, the voltage that capacitor should charge to turn ON T_2 is also maximum. Therefore, at maximum signal voltage, capacitor has to charge to maximum voltage requiring maximum time to charge. This gives us maximum pulse width at maximum input signal voltage. At minimum signal voltage, capacitor has to charge for minimum voltage and we get minimum pulse width at the output. With this discussion, it can be noted that pulse width is controlled by the input signal voltage, and we get pulse width modulated waveform at the output.

9.14.4. Demodulation of PWM Signal

Figure 9.32 (*a*) shows the block diagram of PWM detector. As shown in the figure 9.32 (*a*), the received PWM signal is applied to the Schmitt trigger circuit. This Schmitt trigger circuit removes the noise in the PWM waveform. The regenerated PWM is then applied to the ramp generator and the synchronization pulse detector. The ramp generator produces ramps for the duration of pulses such that height of ramps are proportional to the widths of PWM pulses. The maximum ramp voltage is retained till the next pulse. On the other hand, synchronous pulse detector produces reference pulses with constant amplitude and pulse width. These pulses are delayed by specific amount of delay as shown in the figure 9.32 (*b*). The delayed reference pulses and the output of ramp generator is added with the help of adder. The output of adder is given to the level shifter. Here, negative offset shifts the waveform as shown in the figure 9.32 (*b*) ⑥. Then the negative part of the waveform is clipped by rectifier. Finally, the output of rectifier is passed through low-pass filter to recover the modulating signal, as shown in the figure 9.32 (*b*).

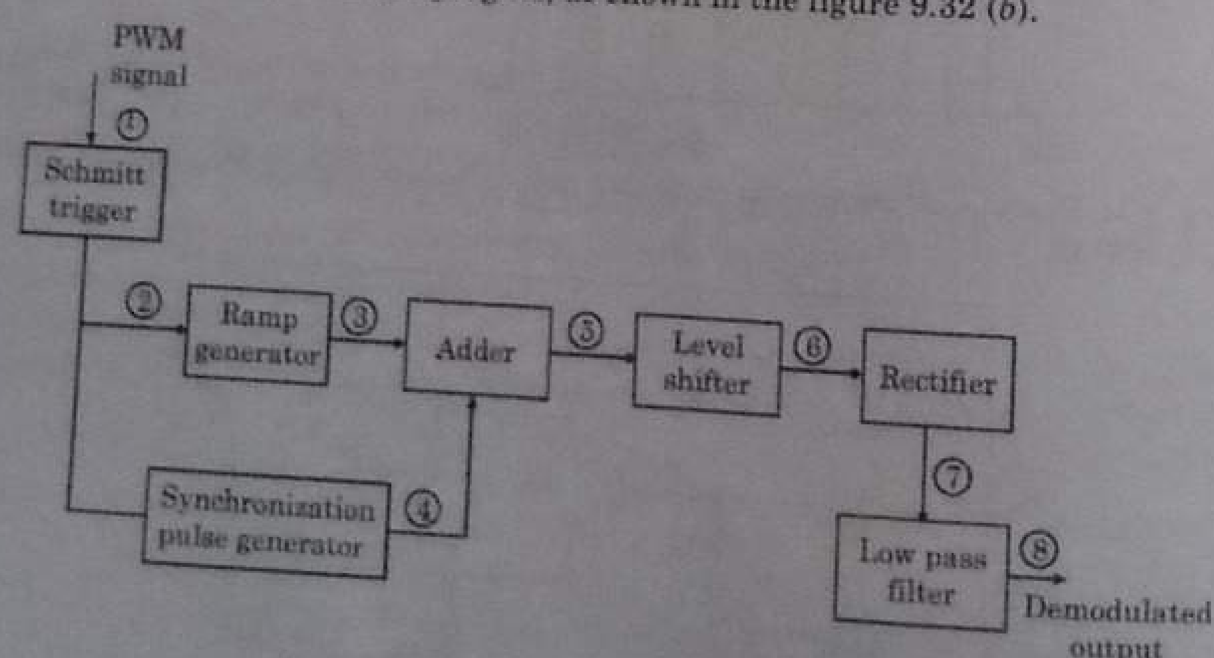


Fig. 9.32. (*a*) *PWM detector*

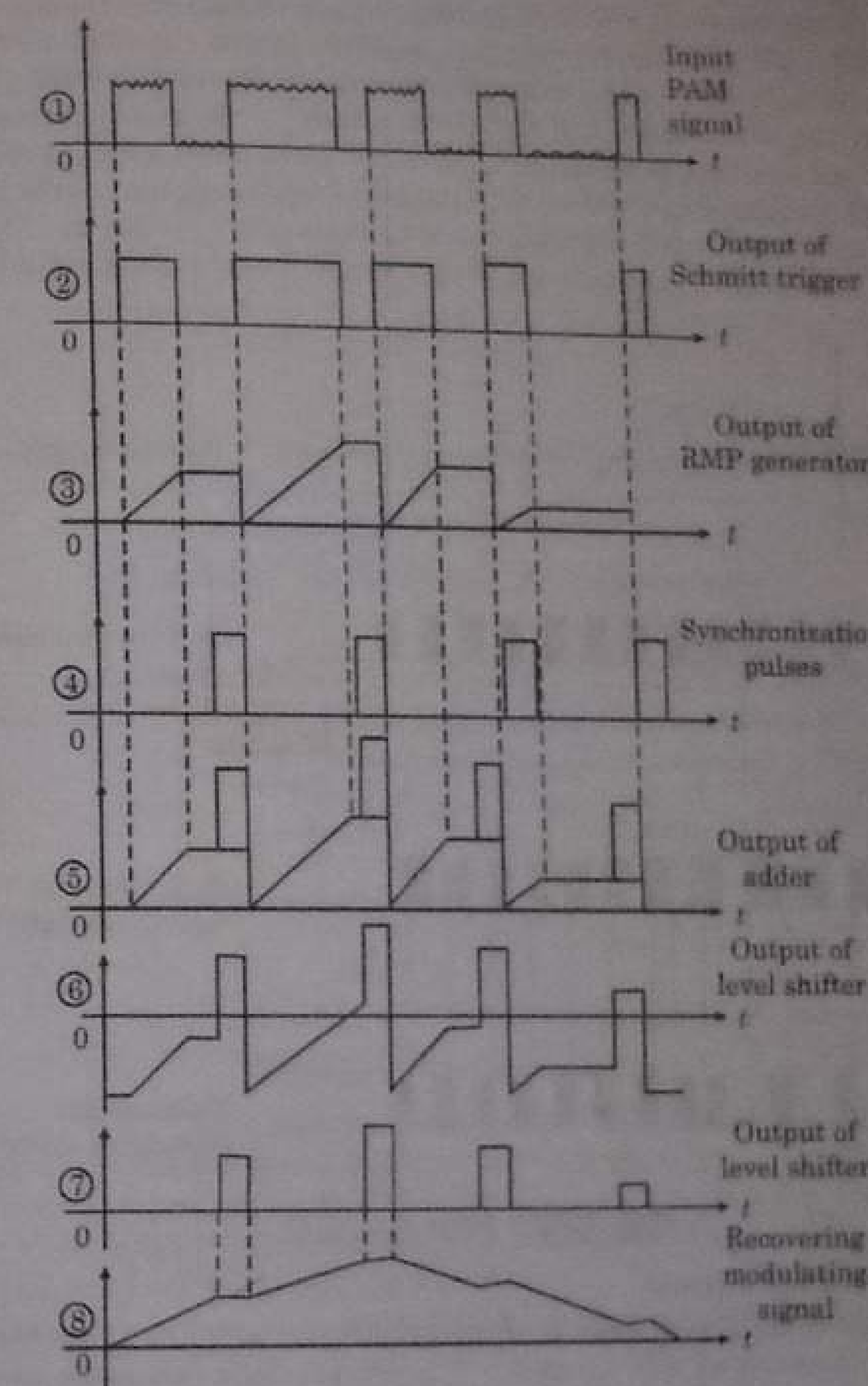


Fig. 9.32. (*b*) *Waveforms for PWM detection circuit.*

9.14.5 Advantages of PWM

(*i*) Unlike, PAM, noise is less, since in PWM, amplitude is held constant.

(*ii*) Signal and noise separation is very easy, as shown in figure 9.32 (*b*) ⑦.

(*iii*) PWM communication does not require synchronization between transmitter and receive

9.14.6. Disadvantages of PWM

(*i*) In PWM, pulses are varying in width and therefore their power contents are variable. Th requires that the transmitter must be able to handle the power contents of the pulse havin maximum pulse width.

(*ii*) Large bandwidth is required for the PWM communication as compared to PAM.

9.14.7. Pulse Position Modulation

In this system, the amplitude and width of the pulses are kept constant, while the position each pulse, with reference to the position of a reference pulse, is changed according to th instantaneous sampled value of the modulating signal. Thus, the transmitter has to send synchronizing pulses to keep the transmitter and receiver in synchronism. As the amplitude and width of the pulses are constant, the transmitter handles constant power output, a definite advantage over the PWM. But the disadvantage of the PPM system is the need for transmitter-receiver synchronization. Pulse position modulation is obtained from pulse width modulation, shown in the figure 9.33. Each trailing edge of the PWM pulse is a starting point of the pulse in the PPM. Therefore, position of the pulse is 1:1 proportional to the width of pulse in PWM and hence it is proportional to the instantaneous amplitude of the sampled modulating signal.

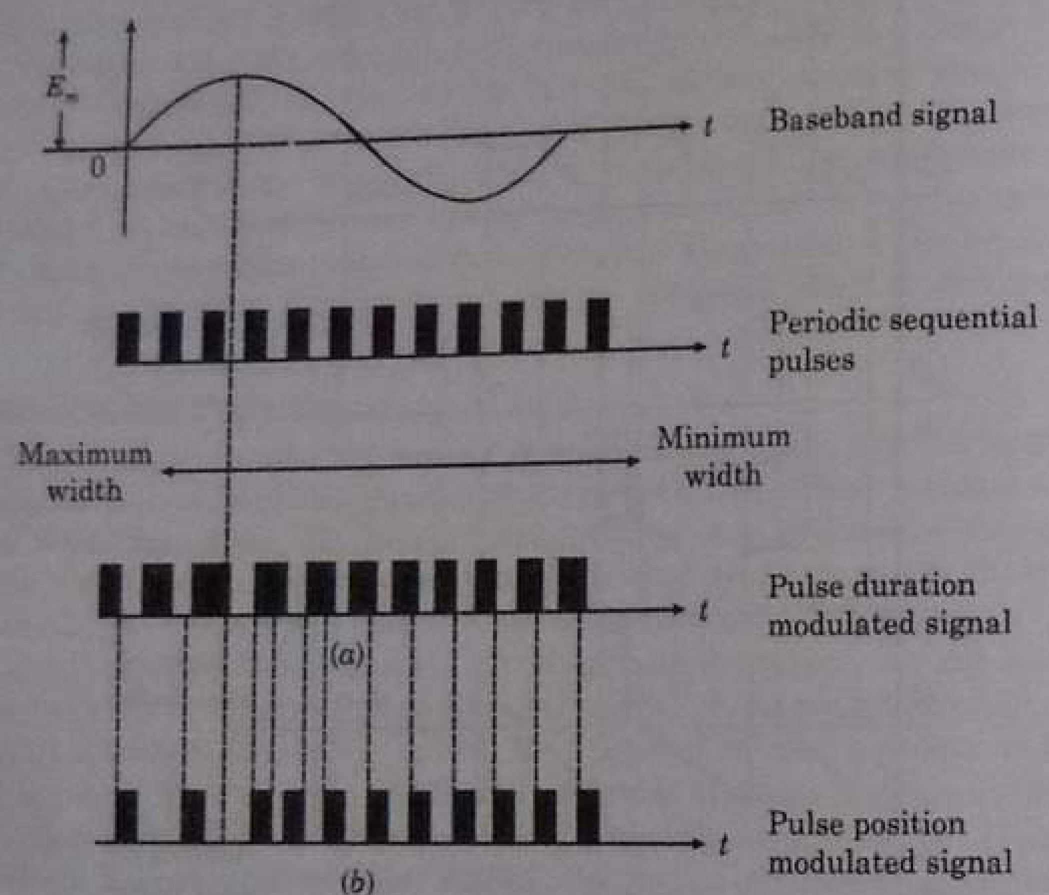


Fig. 9.33. *PPM generator*

9.14.8. Generation of PPM Signal

Figure 9.34 (a) shows the PPM generator. It consists of differentiator and a monostable multivibrator. The input to the differentiator is a PWM waveform. The differentiator generates positive and negative

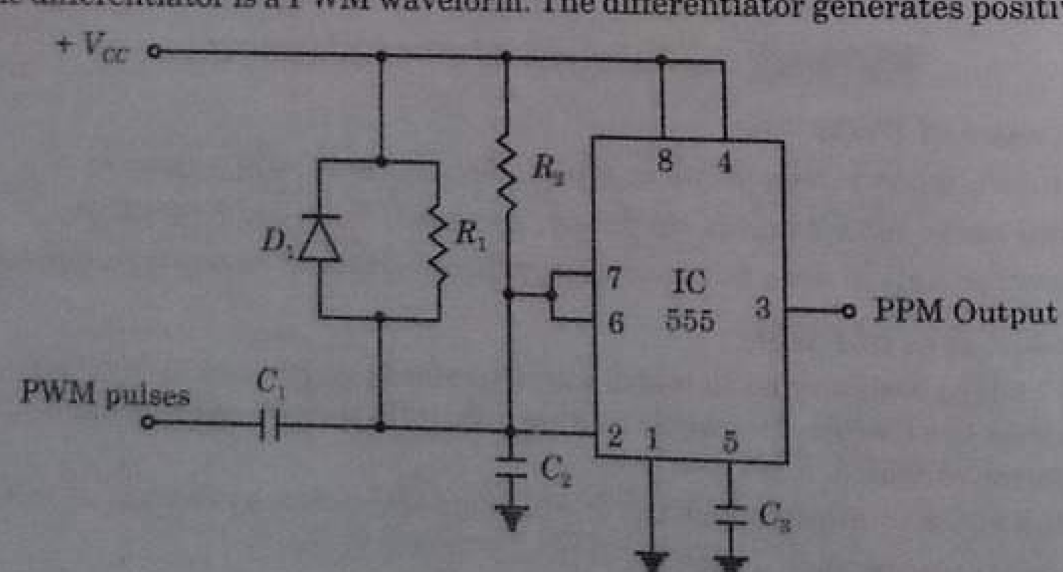


Fig. 9.34. *(a) PPM generator*

spikes corresponding to leading and trailing edges of the PWM waveform. Diode D1 is used to bypass the positive spikes. The negative spikes are used to the trigger monostable multivibrator. The monostable multivibrator then generates the pulses of same width and amplitude with reference to trigger to give pulse position modulated waveform, as shown in figure 9.34 (b).

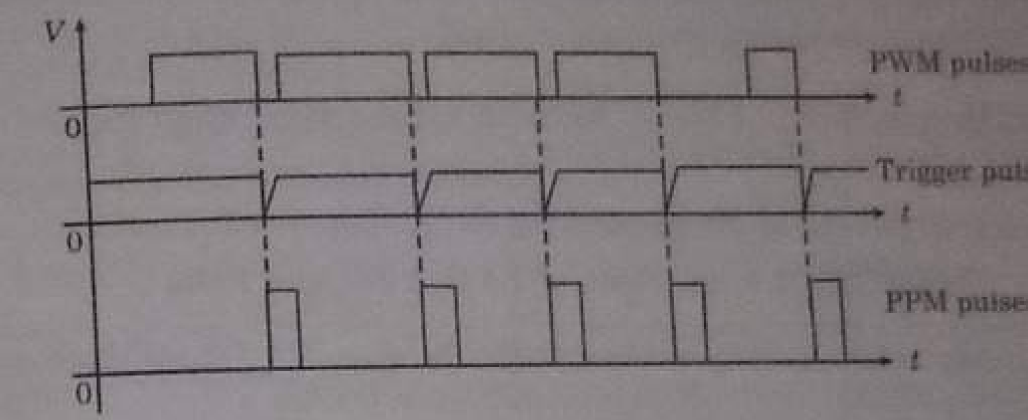


Fig. 9.34. *(b) Waveforms of PPM generator*

9.14.9. Demodulation of PPM

In case of pulse-position modulation, it is customary to convert the received pulses that vary in position to pulses that vary in length. One way to achieve this is illustrated in figure 9.35.

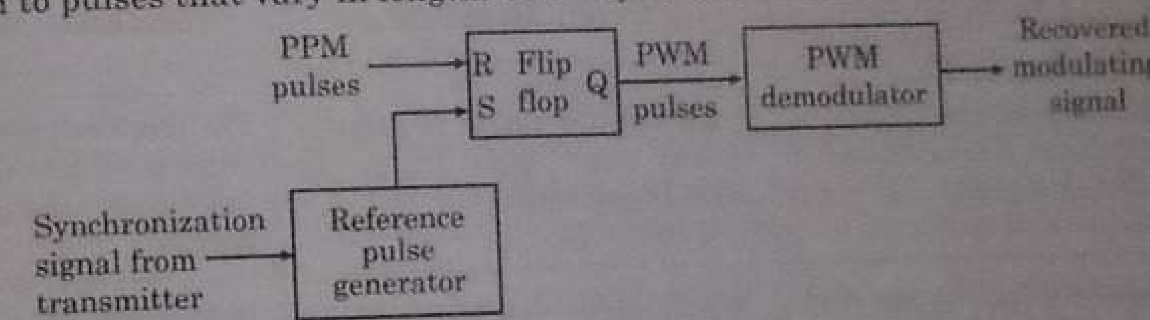


Fig. 9.35. *PPM demodulator*

As shown in figure 9.35, flip-flop circuit is set or turned 'ON' (giving high output) when the reference pulse arrives. This reference pulse is generated by reference pulse generator of the receiver with the synchronization signal from the transmitter. The flip-flop circuit is reset or turned 'OFF' (giving low output) at the leading edge of the position modulated pulse. This repeats and we get PWM pulses at the output of the flip-flop.

The PWM pulses are then demodulated by PWM demodulator to get orginal modulating signal.

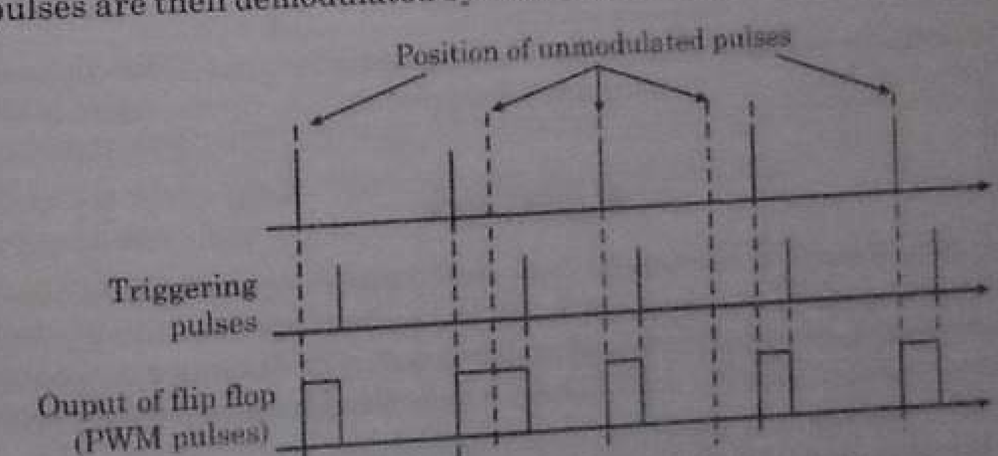


Fig. 9.36. *Demodulation waveform for PPM*

9.14.10. Advantages of PPM

(i) Like PWM, in PPM, amplitude is held constant thus less noise interference.

(ii) Like PPM, signal and noise separation is very easy.

(*iii*) Because of constant pulse widths and amplitudes, transmission power for each pulse is same.

9.14.11. Disadvantages of PPM

(*i*) Synchronization between transmitter and receiver is required.

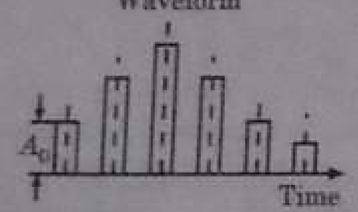
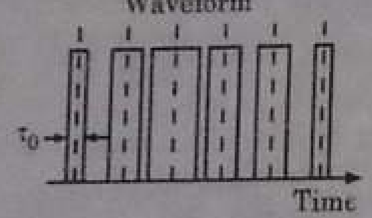
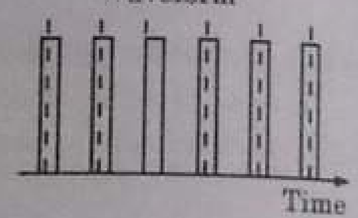
(*ii*) Large bandwidth is required as compared to PAM.

9.15. Comparison of Various Pulse Analog Modulation Methods

(*U.P. Tech, Semester Examination, 2003-04*)

In this section, let us compare PAM, PWM and PPM in the form of a Table 9.2.

Table 9.2. Comparison of PAM, PPM and PDM

S. No.	*Pulse Amplitude Modulation (PAM)*	*Pulse Width/Duration Modulation (PWM) or (PDM)*	*Pulse Position Modulation (PPM)*
1	Waveform 	Waveform 	Waveform 
2	Amplitude of the pulse is proportional to amplitude of modulating signal.	Width of the pulse is proportional to amplitude of modulating signal.	The relative position of the pulse is proprotional to the amplitude of modulating signal.
3	The bandwidth of the transmission channel depends on width of the pulse.	Bandwidth of transmission channel depends on rise time of the pulse.	Bandwidth of transmission channel depends on rising time of the pulse.
4	The instantaneous power of the transmitter varies.	The instantaneous power of the transmitter varies.	The instantaneous power of the transmitter remains constant.
5	Noise interference is high System is complex .	Noise, interference is minimum.	Noise, interference is minimum.
6	Similar to amplitude modulation.	Simple to implement similar to frequency modulation.	Simple to implement similar to phase modulation.

SUMMARY

1. There are two types of signals, continuous time signal and discrete-time signals.
2. Due to some recent advance development in digital technology over the past few decades, the inexpensive, light weight, programmable and easily reproducible discrete-time systems are available. Therefore, the processing of discrete-time signals is more flexible and is also preferable to processing of continuous-time signals.
3. The sampling theorem is extremely important and useful in signal processing.
4. With the help of sampling theorem, a continuous-time signal may be completely represented and recovered from the knowledge of samples taken uniformly.
5. The concept of sampling provides a widely used method for using discrete-time system technology to implement continuous-time systems and process the continuous-time signals.

6. Sampling of the signals is the fundamental operation in signal-processing. A continuous time signal is first converted to discrete-time signal by sampling process.
7. The sufficient number of samples of the signal must be taken so that the original signal is represented in its samples completely. Also, it should be possible to recover or reconstruct the original signal completely from its samples. The number of samples to be taken depends on maximum signal frequency present in the signal.
8. A continuous-time signal may be completely represented in its samples and recovered back if the sampling frequency is $f_s \geq 2 f_m$. Here, f_s is the sampling frequency and f_m is the maximum frequency present in the 'signal.
9. When the sampling rate becomes exactly equal to $2 f_m$ samples per second, then it is called Nyquist rate. Nyquist rate is also called the *minimum* sampling rate. It is given by
$$f_s = 2 f_m$$
10. Maximum sampling interval is called *Nyquist interval*. It is given by

    Nyquist Interval $\quad T_s = \dfrac{1}{2 f_m}$ seconds.
11. The low pass filter is used to recover original signal from its samples. This is also known as interpolation filter.
12. A low-pass filter is that type of filter which passes only low-frequencies upto a specified cut-off frequency and rejects all other frequenices above cut-off frequency.
13. The process of reconstructing a continuous-time signal $x(t)$ from its samples is called as interpolation.
14. Aliasing is the phenomenon in which a high frequency component in the frequency-spectrum of the signal takes identity of a lower-frequency component in the spectrum of the sampled signal.
15. Because of the overlap due to aliasing phenomenon, it is not possible to recover original signal $x(t)$ from sampled signal $g(t)$ by low-pass filtering since the spectral components in the overlap regions add and hence the signal is distorted.
16. Since any information signal contains a large number of frequencies, so, to decide a sampling frequency is always a problem. Therefore, a signal is first passed through a low-pass filter. This low-pass filter blocks all the frequencies which are above f_m Hz. This process is known as band limiting of the original signal $x(t)$. This low-pass filter is called prelias filter because it is used to prevent aliasing effect.
17. To avoid aliasing, we must have:
    (*i*) Preaias filter must be used to limit band of frequencies of the signal to f_m Hz.
    (*ii*) Sampling frequency 'f_s' must be selected such that
$$f_s > 2 f_m$$
18. The bandpass signal $x(t)$ whose maximum bandwidth is $2f_m$ can be completely represented into and recovered from its samples if it is sampled at the minimum rate of twice the bandwidth. Here, f_m is the maximum frequency component present in the signal.
19. Basically, there are three types of sampling techniques as under:
    (*i*) Instantaneous sampling
    (*ii*) Natural sampling
    (*iii*) Flat top sampling.
20. Natural sampling is a practical method and will be discussed in this section.
21. Spectrum of Naturally Sampled signal:
$$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \text{sinc}(n f_s \tau)\, X(f - n f_s)$$
22. Flot top sampling like natural sampling is also a practically possible sampling method. But natural sampling is little complex whereas it is quite easy to get flat top samples.

23. Spectrum of flat top sampled signal:

$$G(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) H(f)$$

24. In analog modulation systems, some parameter of a sinusoidal carrier is varied according to the instantaneous value of the modulating signal. In pulse modulation methods, the carrier is no longer a continuous signal but consists of a pulse train. Some parameter of which is varied according to the instantaneous value of the modulating signal.

25. There are two types of pulse modulation systems as under:
    (i) Pulse Amplitude Modulation (PAM)
    (ii) Pulse Time Modulatin (PTM)

26. In pulse amplitude modulation (PAM), the amplitude of the pulses of the carrier pulse train is varied in accordance with the modulating signal whereas in Pulse time modulation (PTM), the timing of the pulses of the carrier pulse train is varied.

27. They are two types of PTM as under:
    (i) Pulse width modulation (PWM)
    (ii) Pulse position modulation (PPM).

28. In Pulse width modulation, the width of the pulses of the carrier pulse train is varied in accordance with the modulating signal whereas in Pulse position modulation (PPM), the position of pulses of the carrier pulse train is varied.

29. It may be noted that all the above pulse modulation methods (*i.e.*, PAM, PWM and PPM) are called analog Pulse modulation methods because the modulating signal is analog in nature in PAM, PWM and PPM.

30. Pulse amplitude modulation may be defined as that type of modulation in which the amplitudes of regularly spaced rectangular pulses vary according to instantaneous value of the modulating or message signal. In fact, the pulses in a PAM signal may be of flat top type or natural type or ideal type. Actually all the sampling methods which have been discussed in last sections are basically pulse amplitude modulation methods.

31. BW of PAM signal is given by

$$BW \geq f_{max}$$

But $$f_{max} = \frac{1}{2\tau}$$

Hence $$BW \geq \frac{1}{2\tau}$$

Again, since $$\tau << \frac{1}{2f_m}$$

Therefore $$BW \geq \frac{1}{2\tau} >> f_m.$$

32. If the PAM signals are to be transmitted directly *i.e.*, over a pair of wires then no further signal processing is necessary. However, if they are to be transmitted through the space using an antenna, they must first be amplitude or frequency or phase modulated by a high frequency carrier and only then they can be transmitted.

33. The bandwidth required for the transmission of a PAM signal is very very large in comparison to the maximum frequency present in the modulating signal.

34. Since the amplitude of the PAM pulses varies in accordance with the modulating signal therefore the interference of noise is maximum in a PAM signal. This noise cannot be removed easily.

35. Since the amplitude of the PAM signal varies, therefore, this also varies the peak power required by the transmitter with modulating signal.

SHORT QUESTIONS WITH ANSWERS

Q.1. What is sampling?

Ans. The process of converting an analog signal into a discrete signal (or) making an analog (or) continuous signal to occur at a particular interval of time is known as sampling.

Q.2. What do you mean by sampling period (T_s) and sampling rate (f_s)?

Ans. During sampling, the time elapsed between sample to sample (or) time taken by the next sample to occur is known as sampling period. It is denoted as T_s.

Sampling rate is reciprocal of sampling period. It is denoted as $f_s = 1/T_s$

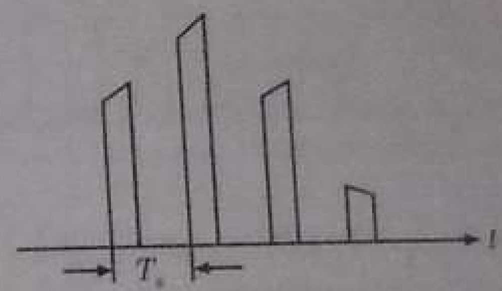


Fig. 9.37.

Q.3. What is Guard band?

Ans. When the sampling rate is chosen much higher than the Nyquist rate then a small space occurs between the samples. This space is said to be **Guard band**. This is the desired one for sampling the signals.

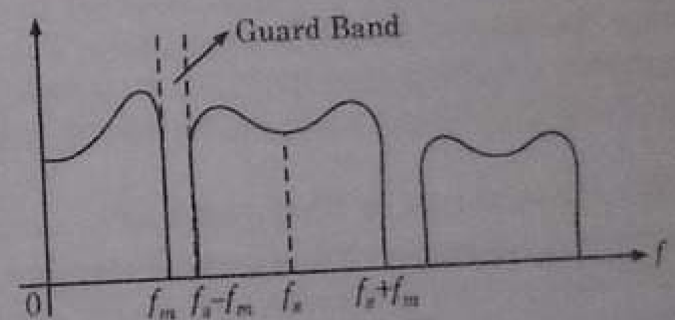


Fig. 9.38.

Q.4. What is Sampling Theorem?

Ans. Let $m(t)$ be a signal which is bandlimited such that its highest frequency spectral component is f_M. Let the values of $m(t)$ be determined at regular intervals separated by times $T_s \leq 1/2f_M$, *i.e.*, the signal is periodically sampled every T_s seconds. Then, these samples $m(nT_s)$ where n is an integer, uniquely determine the signal, and the signal may be reconstructed from these samples with no distortion at the receiver end.

Q.5. What do you mean by aperture effect?

Ans. During flat top sampling, to convert varying amplitudes of pulses to flat top pulses we use a sinc function. Because of this, there would be decrease in the amplitude. This distortion is named as Aperture effect.

This may be eliminated by using an equalizer in cascade with the output low pass filter.

Q.6. What is Natural Sampling?

Ans. It is similar to that of Instantaneous Sampling.

Let us consider a baseband signal $m(t)$ which is to be sampled. To sample $m(t)$, a sampling function $s(t)$ consisting of a train of pulses having duration T and separated by the sampling time T_s is given to one of the input of a product modulator and $m(t)$ to other input. Hence, the output consists of pulses of duration T but with a varying amplitude. This is natural sampling.

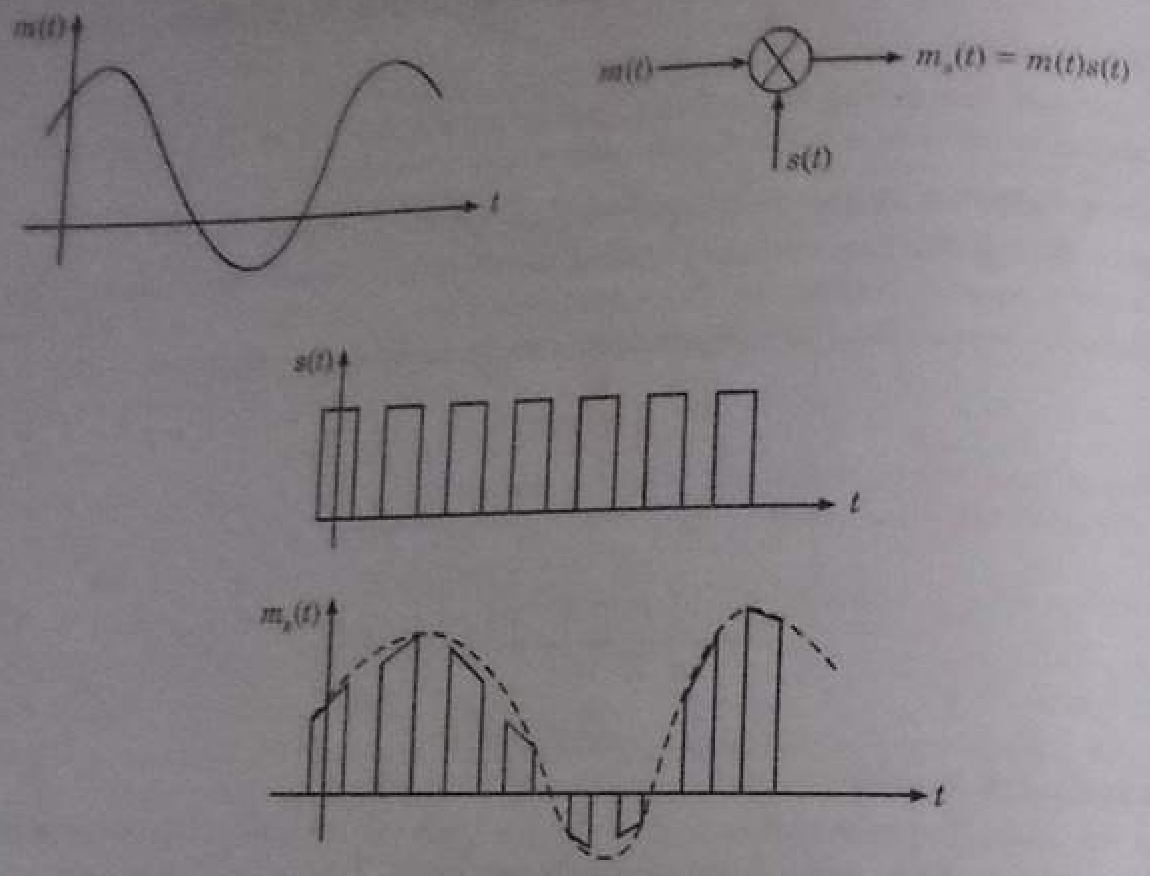


Fig. 9.39.

What is Pulse Amplitude Modulation (PAM)?

The process in which amplitudes of regularly spaced rectangular pulses vary with the instantaneous le values of a continuous message signal in a one to one fashion is known as pulse amplitude lation.

What do you mean by synchronization in PAM systems?

In general, most pulse systems require synchronization of the receiver to the transmitter. There any ways to provide synchronization to the systems.

ne of the method is a start-stop method of synchronization. This method involves transmitting information, in addition to the message-bearing pulses, to serve as a time mark within each interval so that the gates in the receiver are made to open and close at the appropriate instants e.

some cases, the necessary time mark is established by transmitting a distinctive marker pulse me, whereas in some other cases, it is established by omitting a pulse in a particular time slot. ing to be noted is that when markers are used, they should differ from the message-bearing in some recognizable fashion.

a PAM system, the marker pulse may be identified by making its amplitude exceed that of all e message pulses.

us consider the transmission of PAM pluses with marker pulses. This has been shown in .40.

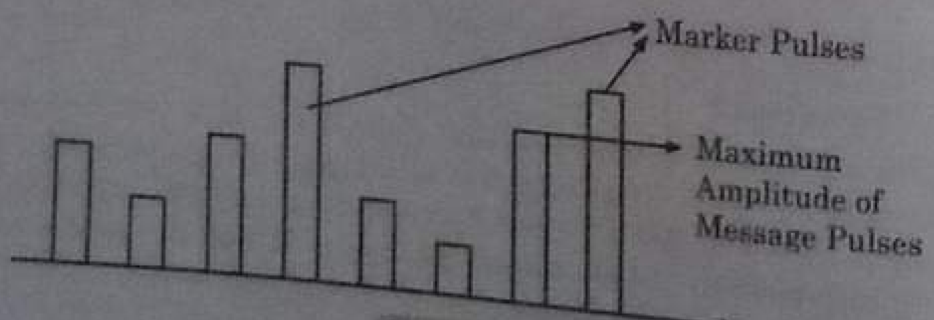


Fig. 9.40.

The PAM signal which is transmitted along with the marker pulses are detected at the receiver by giving these received pulses to a slicer.

The slicer is considered with a slicing level that is just in excess of the maximum amplitude of the message pulses, so that these pulses produces zero output.

An ideal slicer has the property that its output is zero whenever its input is below its slicing level and is constant whenever the input exceeds this level.

Q.9. What do you mean by Pulse-Time Modulation?

Ans. The modulation technique in which the time (or) duration of the pulses is varied in accordance with the amplitude of the message signal keeping the amplitude of the pluses constant is referred to as pulse-time modulation.

Q.10. What are the different types of PTM systems?

Ans. There are two kinds of Pulse-time modulation schemes. They are:

(*i*) Pulse duration (or) pulse width (or) pulse length modulation (PDM (or) PWM (or) PLM).

(*ii*) Pulse position modulation.

Q.11. What is pulse duration modulation (PDM)?

Ans. The method in which the samples of the message signal are used to vary the duration (or) width of the individual pluses. This is referred to as pulse duration modulation.

Q.12. What is pulse position modulation (PPM)?

Ans. In PPM, the position of a pulse relative to its unmodulated time of occurrence is varied in accordance with the message signal.

Q.13. Describe the spectral representation of PDM and PPM waves.

Ans. The spectral analysis of a PDM and PPM wave is complicated.

Let us consider the qualitative description of the spectra of PDM and PPM waves.

Let L_s denotes the time separation between the leading edges of duration modulated pluses obtained by natural sampling.

f_m is sinusoidal modulatin wave of frequency.

Thus the spectrum of a naturally sampled PDM wave consists of the following components:

- Sinusoidal components of frequencies equal to the integer multiplies of $1/T_S$, corresponding to spectral lines at $\pm n/T_S$, where $n = 1, 2, 3,...$ These sinusoidal components as well as the DC components are contributed by the unmodulated pulse train which is regarded as the carrier of the PDM wave.
- A sinusoidal component of frequency f_m and in phase with modulating wave, corresponding to spectral lines at $\pm f_m$.
- Sinusoidal components of frequencies equal to pairs of side-frequencies centred around each spectral line of the unmodulated pulse train, expect the dc component.

It may be noted that these dc components represent the cross-modulation products between the sinusoidal modulation and sampling frequencies.

REVIEW QUESTIONS

1. State and prove sampling theorem in time domain.
2. What is Nyquist rate and Nyquist interval?
3. A bandlimited signal $x(t)$ is sampled by a train of rectangular pulses of width τ and period T.
   (*i*) Find an expression for the sampled signal.
   (*ii*) Determine the spectrum of the sampled signal and sketch it.
4. What is aliasing and how it is reduced?

NUMERICAL PROBLEMS

1. Determine the Nyquist sampling rate and the Nyquist sampling interval for the following singnals:
   (*a*) sinc (100 πt)
   (*b*) $\text{sinc}^2(100\pi t)$
   (*c*) sinc $(100\pi t)$ + sinc $(50\pi t)$
   (*d*) sinc $(100\pi t) + 3\,\text{sinc}^2(60\pi t)$
   (*e*) sinc $(100\pi t)$ sinc $(100\pi t)$
2. A signal $g(t)$ bandlimited to BHz is sampling by a periodic pulse train $pT_s(t)$ made up of a rectangular pulse of with $\frac{1}{8}B$ seconds (centred at the origin) repeating at the Nyquist rate ($2B$ pulses per second). Show that the sampled signal $\overline{g}(t)$ is given by

$$\overline{g}(t) = \frac{1}{4}g(t) + \sum_{n=1}^{\infty}\frac{2}{n\pi}\sin\left(\frac{n\pi}{4}\right)g(t)\cos nu_s t$$

$$\omega_s = 4\pi B$$

   Show that the signal $g(t)$ can recovered by passing $\overline{g}(t)$ through an ideal low-pass filter of bandwidth BHz and a gain of 4.
3. Signal $g_1(t) = 104\,\text{rect}(104t)$ and $g_2(t) = \delta(t)$ are applied at the inputs of ideal low-pass filters $H_1(\omega) = \text{rect}(\omega/40,000\pi)$ and $H_2(\omega) = \text{rect}(\omega/20,000\pi)$ (Figure 9.41). The outputs $y_1(t)$ and $y_2(t)$ of these filters are multiplied to obtain the signal $y(t) = y_1(t)\,y_2(t)$. Find the Nyquist rate of $y_1(t)$, $y_2(t)$, and $y(t)$. *(GATE Examination-1998)*

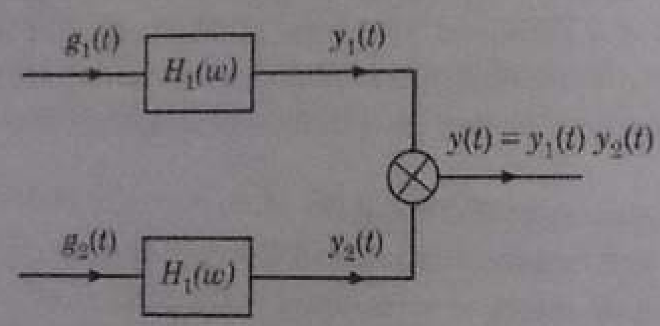


Fig. 9.41.

4. A signal $g(t)\,\text{sinc}^2(5\pi t)$ is sampled (using uniformly spaced impulses) at a rate of : (*i*) 5 Hz; (*ii*) 10 Hz; (*iii*) 20 Hz. Now, for each of the three cases:
   (*a*) Sketch the sampled signal.
   (*b*) Sketch the spectrum of the sampled signal.
   (*c*) Explain whether you can recover the signal $g(t)$ from the sampled signal.
   (*d*) If the sampled signal is passed through an ideal low-pass filter of bandwidth 5 Hz, sketch the spectrum of the output signal.
5. A zero-order hold circuit (Figure 9.42) is often used to reconstruct a signal $g(t)$ from its samples.

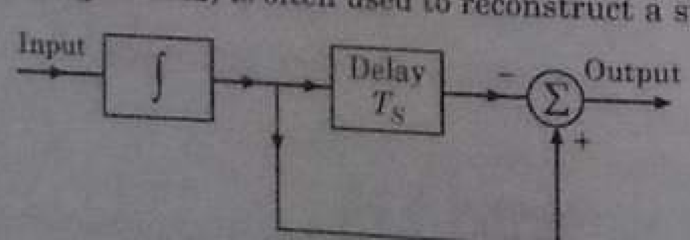


Fig. 9.42.

   (*a*) Find the unit impulse response of this circuit.
   (*b*) Find the transfer function $H(\omega)$ and sketch $|H(\omega)|$.
   (c) Show that when a sampled signal $\overline{g}(t)$ is sampled at the input of this circuit, the output is a staircase approximation of $g(t)$. The sampling interval is T_s. *(Pune University, 1998)*

OBJECTIVE TYPE QUESTIONS

Fill up the Blanks and Multiple Choice Questions

1. ________ is the process in which the analog signal is converted into a corresponding sequence samples that the are uniformly spaced in time.
2. In the process of uniformly sampling, a signal in the time domain results in a periodic spectrum the frequency domain with a period equal to the ________.
3. The sampling rate (f_s) of value 2W samples per second for a signal bandwidth of W Hz is o referred to as ________.
4. Prior to sampling a ________ is used to attenuate the high frequency components of the signal lie outside the band of interest.
5. The bandpass signal is represented as a combination ________ and ________ components
6. The aperture effect in flat top pulses is reduced by using an
   (*a*) Predictor (*b*) Integrator
   (*c*) Equalizer (*d*) Compander
7. The narrow samples produced at the pulse demodulator output are distributed to appropriat pass reconstruction filter by means of
   (*a*) Commutator (*b*) Multiplexer
   (*c*) Decommutator (*d*) none of the above
8. Which of the following method is employed in telephony
   (*a*) Time division multiplexing (*b*) Frequency division multiplexing
   (*c*) both (*a*) and (*b*) (*d*) only (*a*)
9. In ________ the amplitudes of regularly spaced rectangular pulses vary with the instant sample values of a continuous message signal in a one-to-one fashion.
10. The figure of merit of a PPM system is proportional to the square of ________.
11. PPM and PDM systems suffer from a ________ similar to that experienced in FM systems
12. The sampled wave in practical system consists of ________ and ________ rather then imp
13. The term ________ refers to the signal from an adjacent channel spilling over into a desir slot.
14. ________ is the process which involves sampling of the lowpass in-phase and quadrature com of the bandpass signal in accordance with the sampling theorem.
15. In ________ the samples of the message signal are used to vary the duration of the in pulses.
16. Pulse duration modulation is also known as ________ (or) ________.
17. In ________ the position of a pulse relative to its unmodulated time of occurrence is v accordance with the message signal.

ANSWERS

1. Sampling
2. Sampling rate
3. Nyquist rate
4. a lowpass pre-alias filter
5. In-phase, quadrature
6. (c)
7. (c)
8. (c)
9. Pulse amplitude modulation
10. normalized transmission bandwid
11. Threshold effect
12. Finite amplitude, finite duration
13. Crosstalk
14. Quadrature sampling
15. Pulse duration modulation
16. Pulse width (or) pulse length m
17. Pulse position modulation

CHAPTER 10

Waveform Coding Techniques

Inside this Chapter

10.1. Introduction

To transport an information-bearing signal from one point to another point over a communica
channel, we can use digital or analog techniques. As discussed earlier in chapter 1, the us
digital communication offers several important advantages as compared to analog communica
In particular, a digital communication system offers the following highly attractive featur

(*i*) ruggedness to channel noise and external interference unmatched by any an
communication system.

(*ii*) flexible operation of the system.

(*iii*) integration of diverse sources of information into a common format.

(*iv*) security of information in the course of its transmission from source to the destina

In view of above reasons, digital communications have become the dominant for
communication technology in our society.

However, to handle the transmission of analog message signals (*i.e.*, voice and video sign
by digital means, the signal has to undergo an analog-to-digital conversion. We already
about the pulse amplitude modulation (PAM) technique. We also know the disadvanta
using PAM technique. Using the waveform coding technique, we convert the analog PAM s
into a digital signal. This digital signal is in the form of a train or stream of binary digits
1. Thus, with waveform coding techniques, we enter into the world of digital communic
After sampling an analog signal, the next step in its digital transmission is the generati
the "coded version" (digital representation) of the signal. Pulse Code Modulation (PCM) pro
one method to meet such a requirement.

In PCM, the message signal is sampled and amplitude of each sample is approximated (ro
off) to the nearest one of a finite set of discrete levels. This will enable us to represent both
and amplitude in discrete form. Hence, it is possible to transmit the message signal by me
a digital (coded) waveform. Conceptually PCM is simple to understand. It was the first m
which was developed for the digital coding of the waveforms. PCM is the most applied of
digital coding systems in use today. PCM is therefore widely accepted as the standard a
which the other digital coder systems are calibrated. Here, we can note one more point tha
belongs to a class of signal coders known as **waveform coders**. In this chapter, we shall p
a detailed discussion of different digital pulse modulation techniques.

10.2. Discretization in Time and Amplitude

As discussed earlier, pulse modulation may be classified under two heads *i.e.*, pulse analog modu
and pulse digital modulation. In case of pulse analog modulation, only time is expressed
digital form and any one of the pulse parameters (*i.e.*, pulse amplitude, duration or posit
varied in a continuous manner in accordance with the message signal. Pulse amp
modulation (PAM), Pulse duration (width) modulation (PWM) and Pulse position modu
(PPM) are the examples of pulse analog modulation. In these modulation schemes, inforn
transmission is accomplished in an analog form at discrete times. On the other hand,
pulse digital modulation, the time and the pulse parameter (usually the amplitude) oc
discrete form and digital coded form respectively. Pulse digital modulation is therefore ba
a scheme which converts the analog signal to its corresponding digital form. It is for this
that the analog-to-digital conversion is sometimes known as **pulse digital modulation**.
that the analog-to-digital conversion is called pulse code modulation (PCM).

The simplest form of pulse digital modulation is called pulse code modulation (PCM).
system (*i.e.*, PCM), the message signal is first sampled and then amplitude of each san
rounded off to the nearest one of a finite set of allowable values known as quantization lev

* Most of the signals in our suroundings are analog signals.

that both time and amplitude are in the discrete form. This means that in pulse code modulation both parameters *i.e.*, time and amplitude are expressed in discrete form. This process is called **discretization** in time and amplitude.

10.3. Concept of Quantization

In communication systems, sometimes it happens that we are available with analog signal, however, we have to transmit a digital signal for a particular application. In such cases, we have to convert an analog signal into digital signal. This means that we have to convert a continuous time signal in the form of digits. To see how a signal can be converted from analog to digital form, let us consider an analog signal as shown in figure 10.1(*a*). First of all, we get samples of this signal according to sampling theorem. For this purpose, we mark the time-instants t_0, t_1, t_2 and so on, at equal time-intevals along the time axis. At each of these time-instants, the magnitude of the signal is measured and thus samples of the signal are taken. Figure 10.1(*b*) shows a representation of the signal of figure 10.1(*a*) in terms of its samples.

Now, we can say that the signal in figure 10.1(*b*) is defined only at the sampling instants. This neans that it no longer is a continuous function of time, but rather, it is a discrete-time signal. lowever, since the magnitude of each smaple can take any value in a continuous range, the signal n figure 10.1(*b*) is still an analog signal.

This difficulty is neatly resolved by a process known as **quantization**. In quantization, the total mplitude range which the signal may occupy is divided into a number of standard levels.

As shown in figure 10.1(*c*), amplitudes of the signal $x(t)$ lie in the range $(-m_p, m_p)$ which is rtitioned into L intervals, each of magnitude $\Delta v = \frac{2m_p}{L}$. Now, each smaple is approximated or unded off to the nearest quantized level as shown in figure. Since each sample is now approximated one of the L numbers therefore the information is digitized.

The quantized signal is an approximation of the original one. We can improve the accuracy of the antized signal to any desired degree simply by increasing the number of levels L.

.4. Pulse Code Modulation (PCM)

(*U.P. Tech-Semester Exam. 2002-2003*)

se-code modulation is known as a **digital pulse modulation technique**. In fact, the pulse- e modulation (PCM) is quite complex compared to the analog pulse modulation techniques , PAM, PWM and PPM) in the sense that the message signal is subjected to a great number of rations. Figure 10.2 shows the basic elements of a PCM system. It consists of three main s *i.e.*, transmitter, transmission path and receiver. The essential operations in the transmitter PCM system are sampling, quantizing and encoding as shown in figure 10.2. As discussed ier, sampling is the operation in which an analog (*i.e.*, continous-time) signal is sampled rding to the sampling theorem resulting in a discrete-time signal. The quantizing and encoding ations are usually performed in the same circuit which is known as an **analog-to-digital verter (ADC).**

lso, the essential operations in the receiver are regeneration of impaired signals, decoding lemodulation of the train of quantized samples. These operations are usually performed in ame cirucit which is known as a digital-to-analog converter (DAC).

urther, at intermediate points, along the transmission route from the transmitter to the er, regenerative repeaters are used to reconstruct (*i.e.*, regenerate) the transmitted sequence ed pulses in order to combat the accumulated effects of signal distortion and noise.

discussed in article 10.3, the quantization refers to the use of a finite set of amplitude and the selection of a level nearest to a particular sample value of the message signal as resentation for it. In fact, this operation combined with sampling, permits the use of

coded pulses for representing the message signal. *Thus, it is the combined use of quantizing and coding that distinguishes pulse code modulation form analog modulation techniques.*

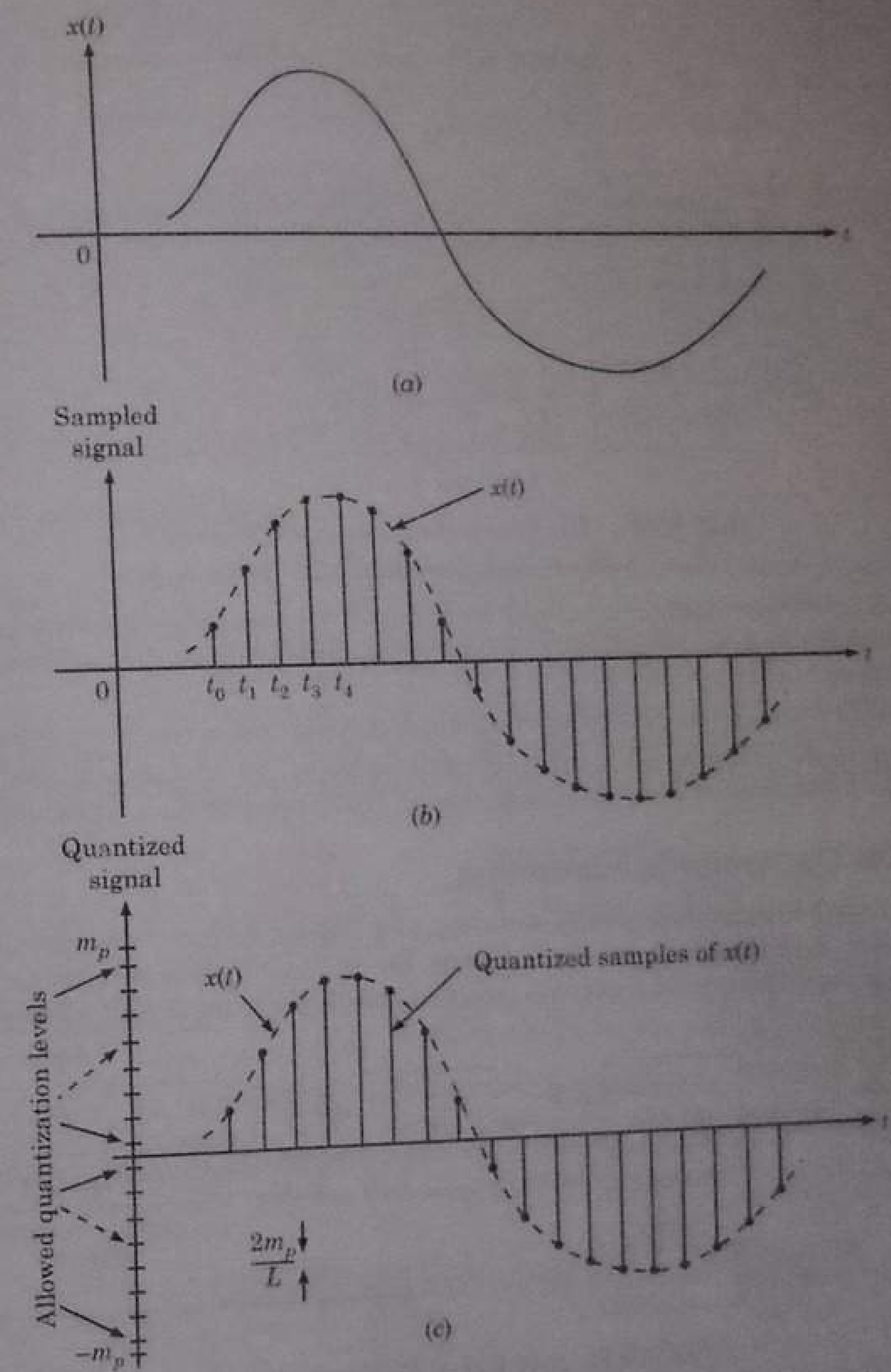


Fig. 10.1. *(a) An analog signal. (b) Samples of analog siagnal, (c) Quantization*

Now, let us summarize PCM in the form of few points as under:

(*i*) PCM is a type of pulse modulation like PAM, PWM or PPM but there is an impor difference between them PAM, PWM or PPM are "analog" pulse modulation sys whereas PCM is a "digital" pulse modulation system.

(*ii*) This means that the PCM output is in the coded digital form. It is in the form of di pulses of constant amplitude, width and position.

(iii) The information is transmitted in the form of "code words". A PCM system consists of a PCM encoder (transmitter) and a PCM decoder (receiver).

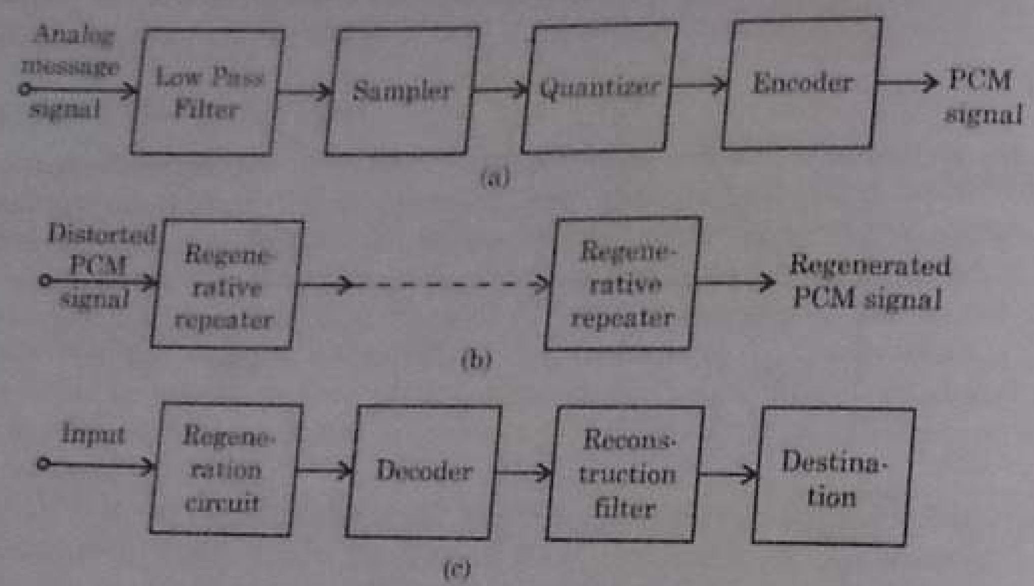


Fig. 10.2. *The basic elements of a PCM system*
(a) Transmitter (b) Transmission path (c) Receiver.

(iv) The essential operations in the PCM transmitter are sampling, quantizing and encoding.

(v) All the operations are usually performed in the same circuit called as **analog-to-digital converter**.

(vi) It should be understood that the PCM is not modulation in the conventional sense.

(vii) Because in modulation, one of the characteristics of the carrier is varied in proportion with the amplitude of the modulating signal. Nothing of that sort happen in PCM.

10.5. A PCM Generator or Transmitter

In the last article, we had an overview of the elements of a PCM system (*i.e.*, transmitter, transmission-path and receiver). In this section, we shall discuss the PCM generator (*i.e.*, transmitter) from a practical point of view. Figure 10.3 shows a practical block diagram of a PCM generator.

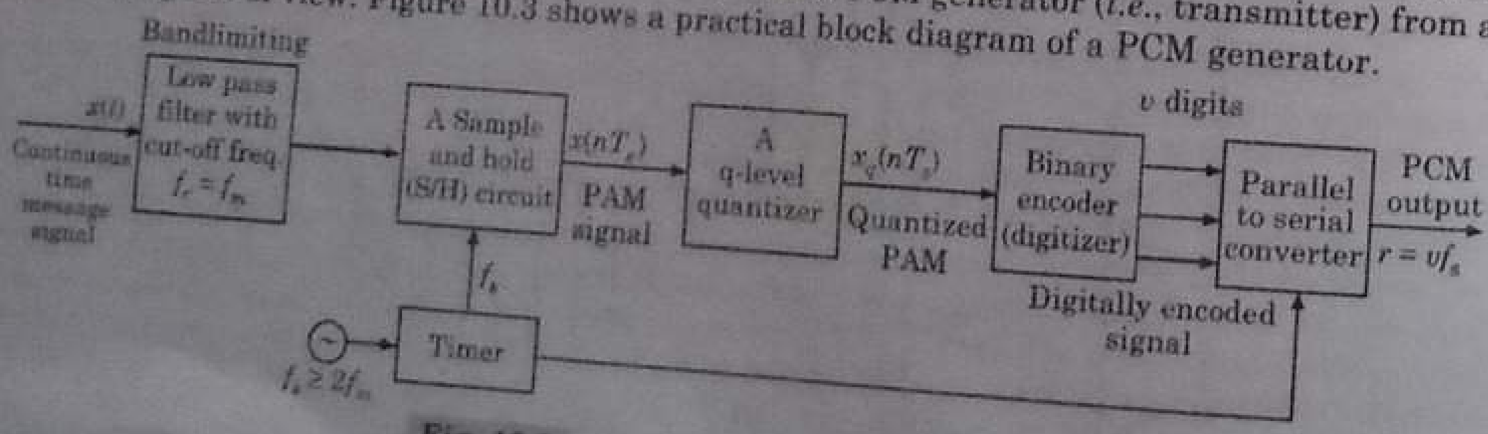


Fig. 10.3. *A practical PCM generator*

In PCM generator of figure 10.3, the signal $x(t)$ is first passed through the low-pass filter of cut-off frequency f_m Hz. This low-pass filter blocks all the frequency components which are lying above f_m Hz.*

This means that now the signal $x(t)$ is bandlimited to f_m Hz. The sample and hold circuit then samples this signal at the rate of f_s. Sampling frequency f_s is selected sufficiently above nyquist rate to aviod aliasing *i.e.*,

$$f_s \geq 2f_m$$

*Recall that this filter is used to avoid aliasing.

In figure 10.3, the output of sample and hold circuit is denoted by $x(nT_s)$. This signal $x(nT_s)$ is discrete in time and continuous in amplitude. A q-level quantizer compares input $x(nT_s)$ with its fixed digital levels. It assigns any one of the digital level to $x(nT_s)$ with its fixed digital level then assigns any one of the digital level to $x(nT_s)$ which results in minimum distortion or error. This error is called **quantization error**. Thus, output of quantizer is a digital level called $x_q(nT_s)$.

Now, the quantized signal level $x_q(nT_s)$ is given to binary encoder. This encoder converts input signal to 'v' digits binary word. Thus $x_q(nT_s)$ is converted to 'v' binary bits. This encoder is also known as digitizer.

Note: *It may be noted that it is not possible to transmit each bit of the binary word separately on transmission line. Therefore 'v' binary digits are converted to serial bit stream to generate single baseband signal. In a parallel to serial converter, usually a shift register does this job. The output of PCM generator is thus a single baseband signal of binary bits.*

Also, an oscillator generates the clocks for sample and hold circuit and parallel to serial converter. In the pulse code modulation generator discussed above, sample and hold, quantizer and encoder combinely form an analog to digital converter (ADC).

10.6. PCM Transmission Path

The Path between the PCM transmitter and PCM receiver over which the PCM signal travel, is called **PCM transmission path** and it is as shown in figure 10.4. The most important feature of PCM system lies in its ability to control the effects of distortion and noise when the PCM wave travels on the channel. PCM accomplishes this capacity by means of using a chain of regenerative repeaters as shown in figure 10.4. Such repeaters are spaced close enough to each other on the transmission path. The regenerator performs three basic operations namely equalization, timing and decision making. Hence, each repeater actually reproduces the clean noise free PCM signal from the PCM signal distorted by the channel. This improves the performance of PCM in presence of noise.

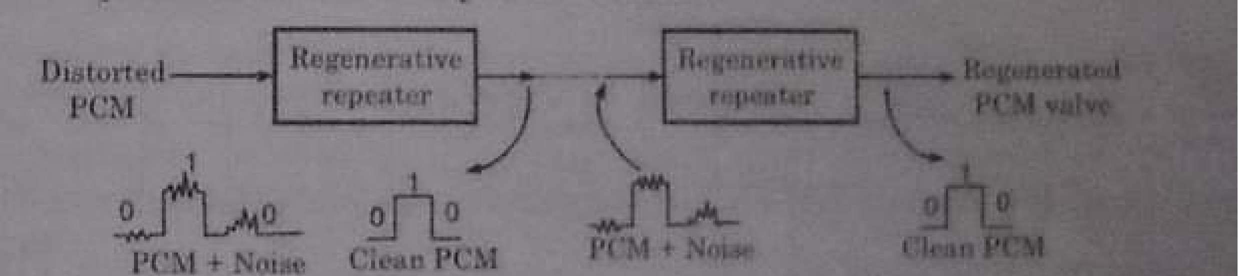


Fig. 10.4. *PCM transmission path.*

10.6.1. Block Diagram of a Repeater

Figure 10.5 shows the block diagram of a regenerative repeater.

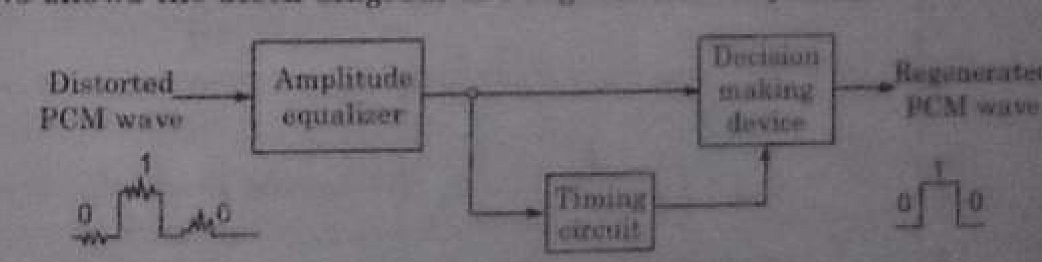


Fig. 10.5. *Block diagram of a regenerative repeater.*

The amplitude equalizer shapes the distorted PCM wave so as to compensate for the effects of amplitude and phase distortions. The timing circuit produces a periodic pulse train which is derived from the input PCM pulses. This pulse train is then applied to the decision making device. The decision making device uses this pulse train for sampling the equalized PCM pulses. The sampling is carried out at the instants where the signal to noise ratio is maximum.

The decision device makes a decision about whether the equalized PCM wave at its input has a 0 value or 1 value at the instant of sampling. Such a decision is made by comparing equalized PCM with a reference level called **decision threshold** as illustrated in figure 10.6a. At the output of the decision device, we get a clean PCM signal without any trace of noise.

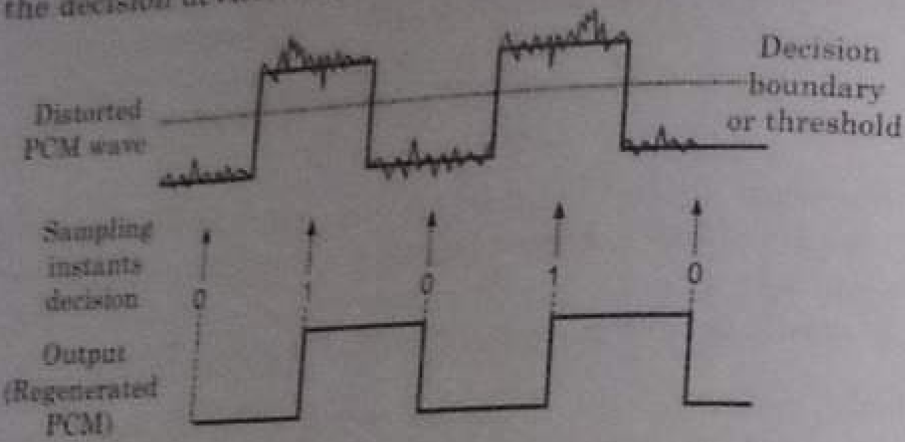


Fig. 10.6a. *Waveforms of a regenerative repeater.*

10.7. PCM Receiver

In this section, we shall discuss a PCM receiver from practical point of view. Figure 10.6(a) shows the block diagram of PCM receiver and figure 10.6 (*b*) shows the reconstructed signal. The regenerator at the start of PCM receiver reshapes the pulse and removes the noise. This signal is then converted to parallel digital words for each sample.

Now, the digital word is converted to its analog value denoted as $x_q(t)$ with the help of a sample and hold circuit. This signal, at the output of sample and hold circuit, is allowed to pass through a lowpass reconstruction filter to get the appropriate original message signal denoted as $y(t)$.

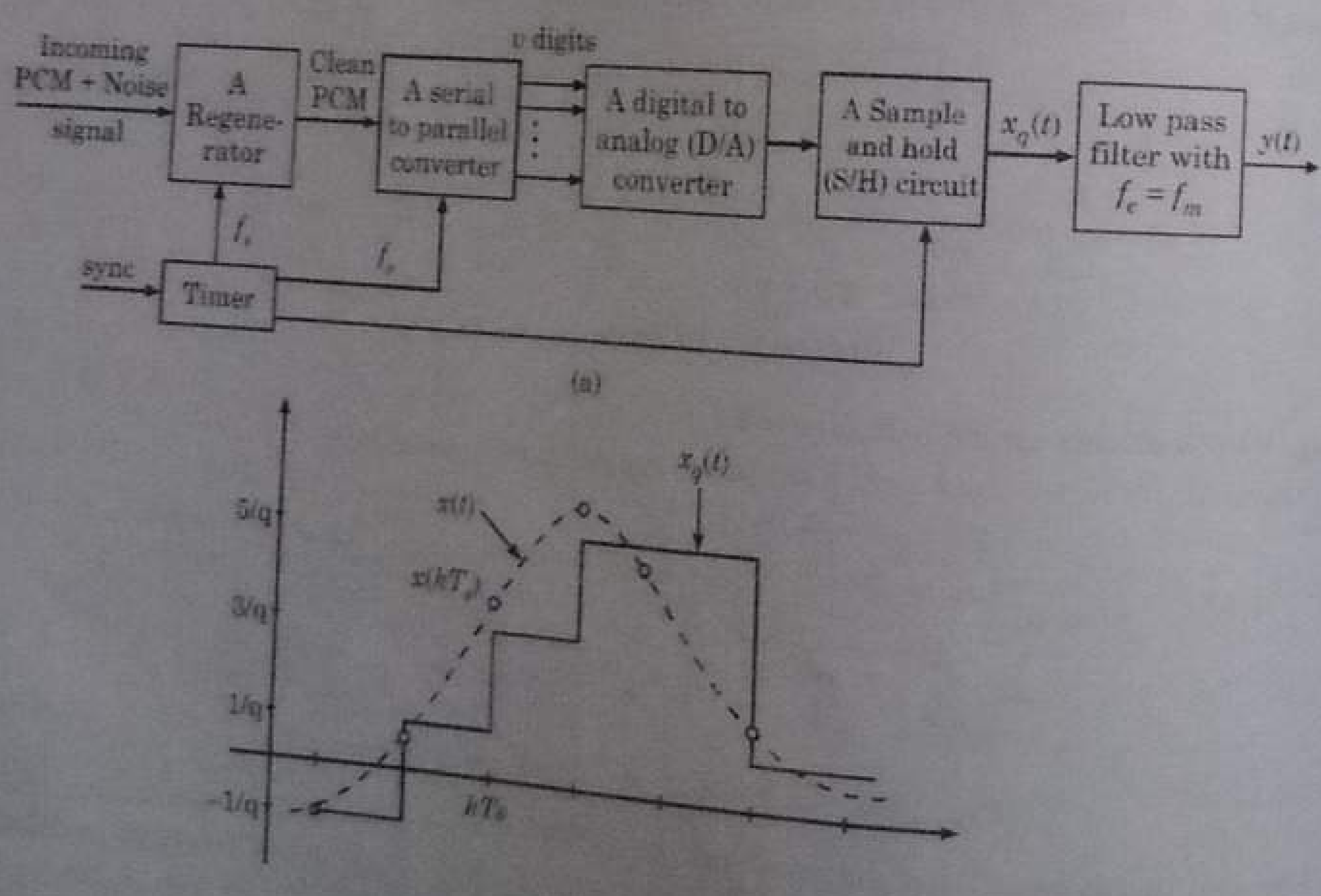


Fig. 10.6. *(a) PCM receiver (b) Reconstructed waveform.*

Note: *As shown in reconstructed signal of figure 10.6 (b), it is impossible to reconstruct exact original signal x(t) because of permanent quantization error introduced during quantization at the transmitter. In fact, this quantization error can be reduced by increasing the binary levels. This is equivalent to increasing binary digits (bits) per sample. But increasing bits 'v' increases the singnaling rate as well as transmission bandwidth as we have observed in last article. Therefore the choice of these parameters is made, in such a manner that noise due to quantization error (i.e., also called as quantization noise) is in tolerable limits.*

10.8. Quantizer

As discussed in article 10.5, a q-level quantizer compares the discrete-time input $x(nT_s)$ with its fixed digital levels. It assigns any one of the digital level to $x(nT_s)$ with its fixed digital levels. It then assigns any one of the digital level to $x(nT_s)$ which results in minimum distortion or error. This error is called quantization error. Thus, the output of a quantizer is a digital level called $x_q(nT_s)$.

10.8.1. Classification of Quantization Process

Figure 10.7 shows the classification of quantization process.

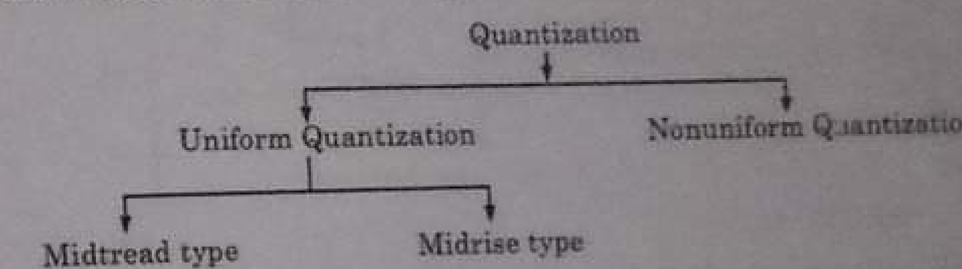


Fig. 10.7. *Classification of quantization process.*

The quantization process can be classified into two types as under:

(*i*) Uniform quantization

(*ii*) Non-uniform quantization.

This classification is based on the step size as defined earlier.

(*i*) Uniform Quantizer

A uniform quantizer is that type of quantizer in which the 'step size' remains same throughout the input range.

(*ii*) Nonuniform Quantizer

A non-uniform quantizer is that type of quantizer in which the 'step-size' varies according to the input signal values.

10.8.2. A Uniform Quantizer

As discussed earlier, a quantizer is called as an uniform quantizer if the step size remains constant throughout the input range.

10.8.2.1. Types of Uniform Quantizer

There are two types of uniform quantizer as under:

(*i*) Symmetric quantizer of the midtread type

(*ii*) Symmetric quantizer of the midrise type

Basically, quantizers can be of a **uniform** or **nonuniform type**. In a uniform quantizer, the representation levels are uniformly spaced; otherwise, the quantizer is nonuniform. Now let us consider only uniform quantizers, nonuniform quantizer shall be considered later on.

The quantizer characteristic can also be *midtread* or *midrise type*. Figure 10.8(*b*) shows the input-output characteristic of a uniform quantizer of the midtread type, which is so called be the origin lies in the middle of a tread of the staircaselike graph. Figure 10.8(*b*) shows the corresponding input-output characteristic of a uniform quantizer of the midrise type, in which the origin lies in the middle of a rising part of the staircaselike graph. It may be noted that both the midtread and midrise types of uniform quantizers illustrated in figure 10.8 are *symmetric* about the origin.

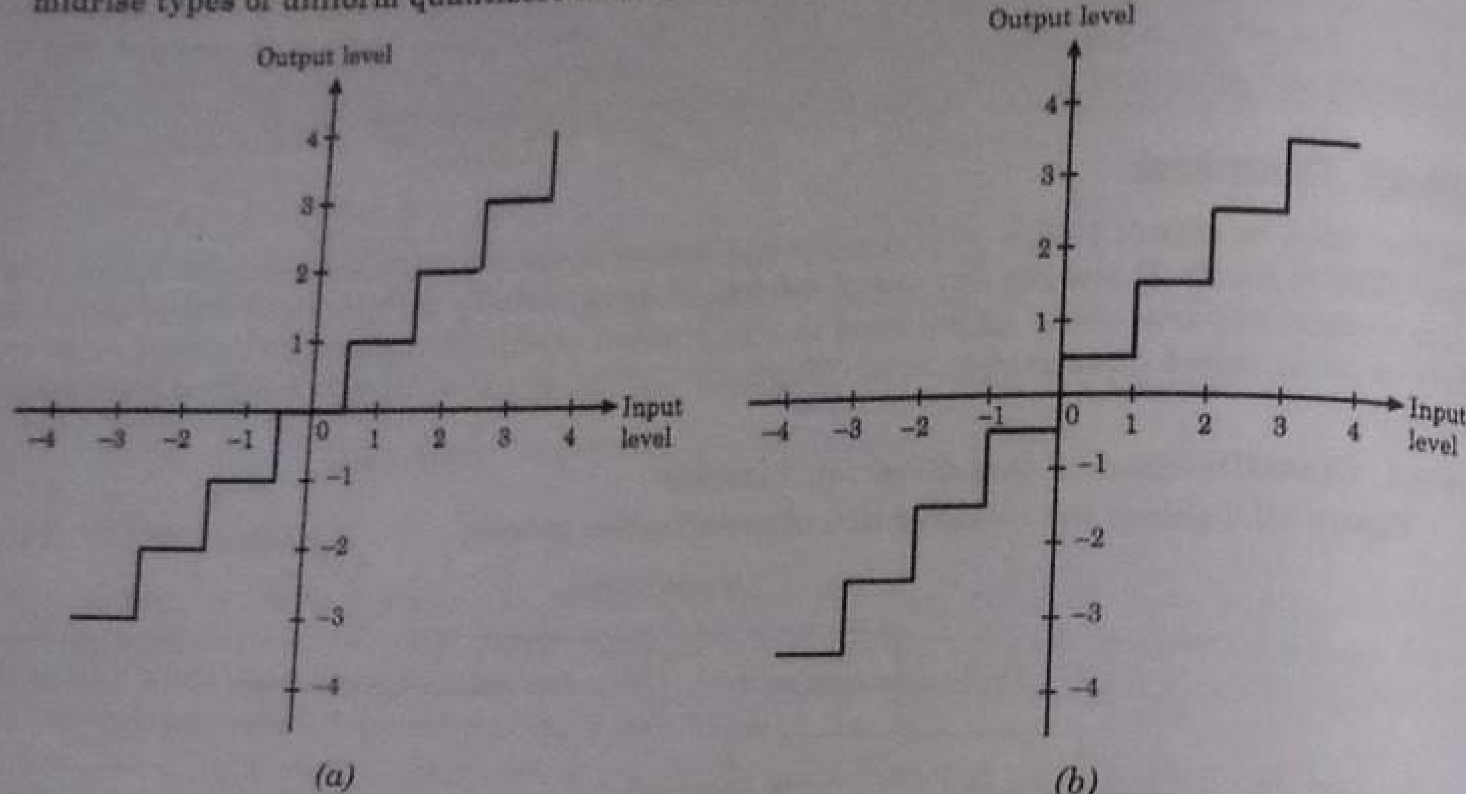


Fig. 10.8. *Two types of Uniform quantization: (a) Midtread, and (b) Midrise.*

10.9. Working Principle of Quantizer

In this section, let us see how uniform quantization takes place. For this purpose, we shall consider uniform quantizer of midrise type. Figure 10.9(*a*) shows the transfer characteristics of a uniform quantizer of midrise type. In figure 10.9(*a*), let us assume that the input to the quantizer $x(nT_s)$ varies from -4Δ to $+4\Delta$. This means that the peak to peak value of $x(nT_s)$ will be between -4Δ to $+4\Delta$. Here 'Δ' is the step size.

Thus, input $x(nT_s)$ can take any value between -4Δ to $+4\Delta$. Now, the fixed digital levels are avilable at $\pm\frac{\Delta}{2}, \pm\frac{3}{2}\Delta, \pm\frac{5}{2}\Delta$ and $\pm\frac{7}{2}\Delta$. These levels are available at quantizer because of its characteristics.

Hence, according to figure 10.9(*a*), we have

If $\quad x(nT_s) = 4\Delta$, then $x_q(nT_s) = \frac{7}{2}\Delta$

and if $\quad x(nT_s) = -4\Delta$, then $x_q(nT_s) = -\frac{7}{2}\Delta$

Thus, it may be observed from figure 10.9(*b*) that maximum quantization error would be $\pm\frac{\Delta}{2}$.

From above, we conclude that quantization error may be expressed as

$$\varepsilon = x_q(nT_s) - x(nT_s) \quad \text{...(10.1)}$$

here 'ε' represents the quantization error

Now, when $x(nT_s) = 0$, quantizer will assign any one of the nearest binary levels *i.e.*, either $\Delta/2$ or $-\Delta/2$. If $\Delta/2$ is assigned, then quantization error will be,

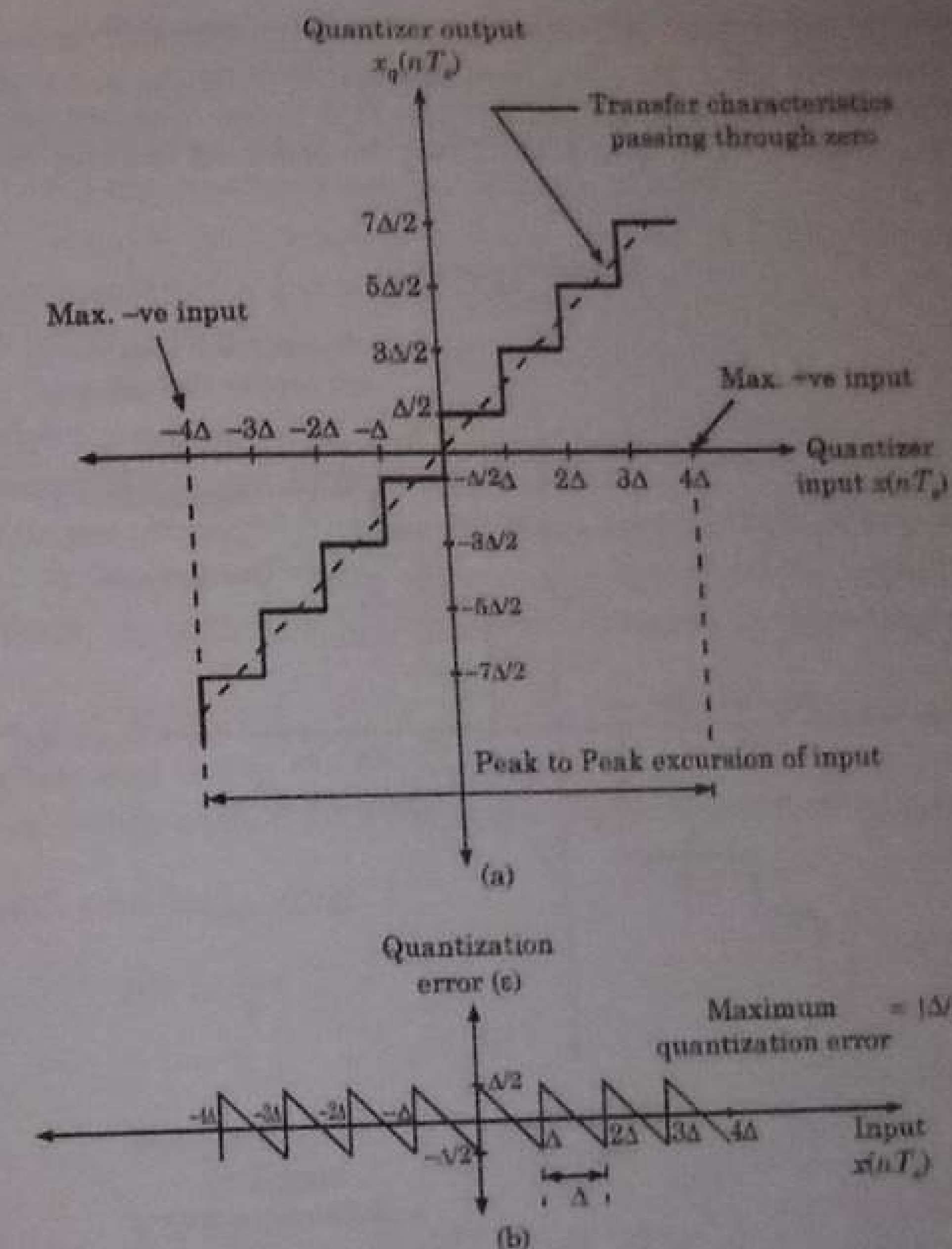


Fig. 10.9. *(a) Transfer characteristic of a quantizer (b) Variation of quantization error with input*

$$\varepsilon = x_q(nT_s) - x(nT_s) = \Delta/2 - 0 = \Delta/2$$

From figure 10.9(*a*), it may also be observed that

for $\quad \Delta < x(nT_s) < 2\Delta, \; x_q(nT_s) = \frac{3}{2}\Delta$

or $\quad -\Delta < x(nT_s) < -2\Delta, \; x_q(nT_s) = -\frac{3}{2}\Delta$

This means that the maximum quantization error will be $\pm\Delta/2$.

In other words, maximum quantization error is given by

$$\varepsilon_{max} = \left|\frac{\Delta}{2}\right|$$

10.10. A Uniform Quantizer with Incorrect Quantization Characteris

The quantizer discussed in last section is known as *uniform quantizer* since the step s same throughout the input range. Also, if step size varies according to the input, the is known as *non uniform quantizer*. The reason for taking the digital levels at $\Delta/2, \pm\frac{3}{2}\Delta, \ldots$ etc. is to reduce the quantization error. This has been illustrated in figure 10.10. Hence, there are two possible characteristics as shown in figure 10.10 (*a*). That is one characteristic 'A' with thick line and second characteristic 'B' with thin line. It may be observed that for characteristic 'A', we have

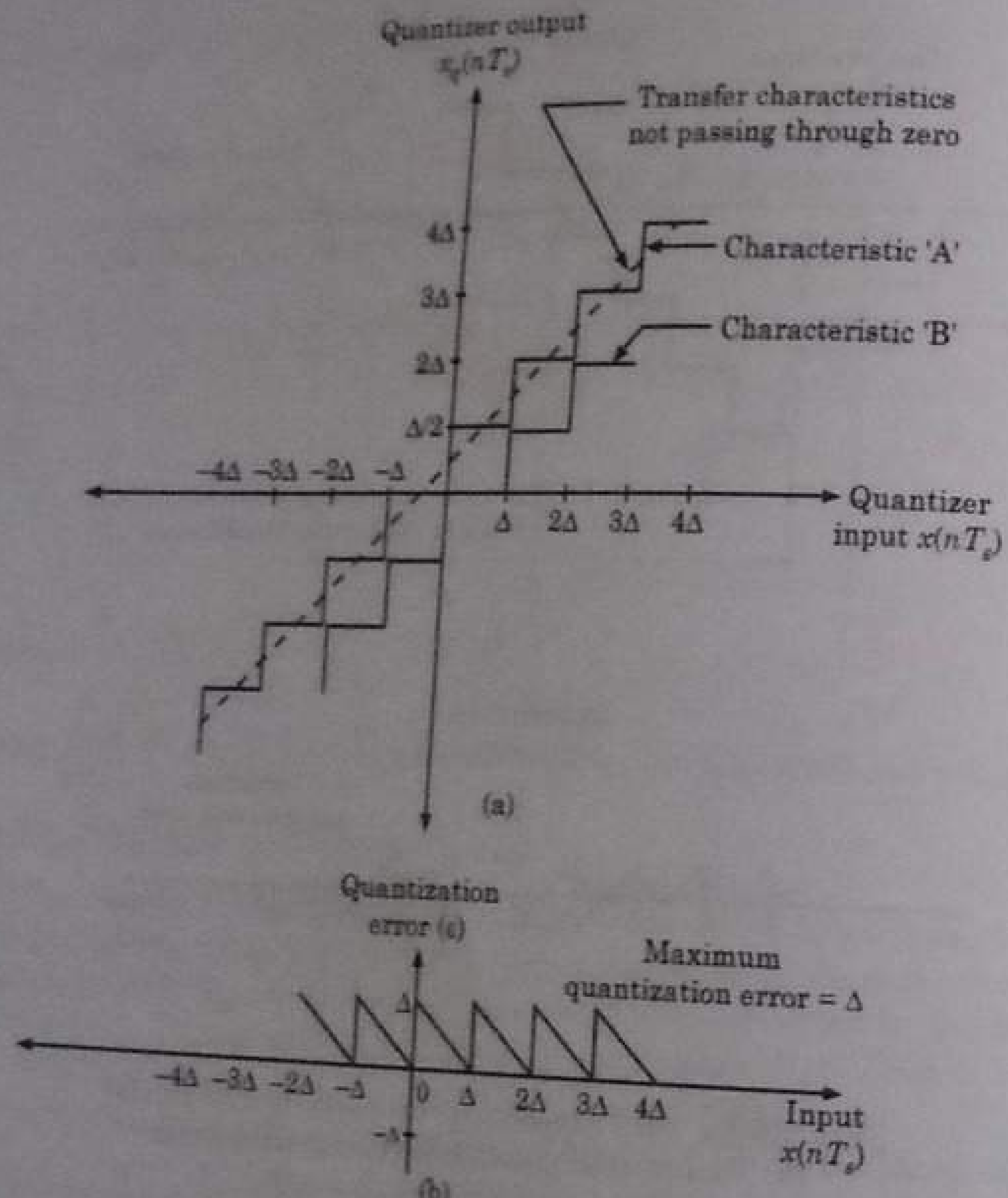


Fig. 10.10. *(a) Incorrect quantization characteristic (b) Increased quantization error.*

If $0 < x(nT_s) < \Delta$; then output $x_q(nT_s) = \Delta$

or $2\Delta < x(nT_s) < 3\Delta$; then output $x_q(nT_s) = 3\Delta$

Therefore, the maximum quantization error will be equal to Δ as shown in figure 10.10 (*b*). imilarly for characteristic '*B*' maximum quantization error is equal to – Δ. The dotted line shows ctual transfer characteristic which passes through origin for characteristic shown in figure 10.10(*a*). n the other hand, it does not pass through characteristic of figure 10.10(*a*). Hence, the digital vels are taken at $\pm\Delta/2, \pm\frac{3}{2}\Delta$ etc. It provides correct quantization characteristic sand reduces antization error.

.11. Transmission Bandwidth in a PCM System

his section, we shall evaluate the transmission bandwidth for PCM system. Let us assume the quantizer use '*v*' number of binary digits to represent each level.

Then, the number of levels that may be represented by '*v*' digits will be,

$$q = 2^v \quad \text{...(10.3)}$$

Here '*q*' represents total number of digital levels of a *q*-level quantizer.

For example, if $v = 4$ bits, the total number of levels will be,

$$q = 2^4 = 16 \text{ levels}$$

Each sample is converted to '*v*' binary bits. *i.e.*,

Number of bits per sample = v.

We know that,

Number of samples per second = f_s

Therefore, Number of bits per second is expressed as

(Number of bits per second) = (Number of bits per samples) × (Number of samples per second)

= v bits per sample × f_s samples per second ...(10.4)

As a matter of fact, the number of bits per second is known as **signaling rate** of PCM and is denoted by '*r*' *i.e.*,

Signaling rate in PCM, $r = v f_s$ ...(10.5)

where $f_s \geq 2 f_m$

Also, since bandwidth needed for PCM transmission is given by half of the signaling rate therefore, we have

Transmission Bandwidth in PCM,

$$BW \geq \frac{1}{2} r \quad \text{...(10.6)}$$

But $r = v f_s$

Therefore, $BW \geq \frac{1}{2} v f_s$ ...(10.7

Again, since $f_s \geq 2 f_m$ ...(10.

Hence, $BW \geq v f_m$

This is the required expression for bandwidth of a PCM system.

10.12. Quantization Noise/Error in PCM

In this section, we shall derive an expression for quantization noise (*i.e.*, error) in a PCM syste for linear quantization or uniform quantization. Because of quantization, inherent errors a introduced in the signal. This error is called **quantization error**. As defined earlier, the quantizati error is given as ...(10.

$$\varepsilon = x_q(nT_s) - x(nT_s)$$

Let us assume that the input $x(nT_s)$ to a linear or uniform quantizer has continuous amplitu in the range $-x_{max}$ to $+x_{max}$.

From figure 10.9(*a*), it may be observed that the total excursion of input $x(nT_s)$ is mapp into '*q*' levels on vertical axis. This means that when input is 4Δ, output is $\frac{7}{2}\Delta$ and when input – 4Δ, output is $-\frac{7}{2}\Delta$. Thus, $+x_{max}$ represents $\frac{7}{2}\Delta$ and $-x_{max}$ represents $-\frac{7}{2}\Delta$. Therefore, total amplitude range becomes,

Total amplitude range = $x_{max} - (-x_{max}) = 2x_{max}$

$$(\text{SNR})_o = \left(\frac{S}{N_q}\right)_0 = \frac{3f_s^3}{8\pi f_m^2 f_M} \quad ...(i)$$

where $f_s = 1/T_s$ is the sampling rate and f_M is the cutoff frequency of a low-pass filter at the output end of the reciver.

Solution: For no-slope-overload, we must have

$$A < \frac{\Delta}{\omega_m T_s} = \frac{\Delta}{2\pi}\left(\frac{f_s}{f_m}\right)$$

Thus, the maximum permissible value of the output signal power equals

$$P_{max} = \frac{A^2}{2} = \frac{\Delta^2 f_s^2}{8\pi^2 f_m^2} \quad ...(ii)$$

We know that the mean quantizing error, or the quantizing noise power, $\langle q_e^2 \rangle = \Delta^2/3$. Let the bandwidth of a postreconstruction low-pass filter at the output end of the reciever be $f_M \geq f_m$ and $f_M << f_s$. Then, assuming that the quantizing noise power P_q is uniformly distributed over the frequency band up to f_s, the output quantizing noise power within the bandwidth f_M is

$$N_q = \left(\frac{\Delta^2}{3}\right)\frac{f_M}{f_s} \quad ...(iii)$$

Combining equation (*ii*) and (*iii*), we see that the maximum output signal-to-quantizing-noise ratio is

$$\left(\frac{S}{N_q}\right)_0 = \frac{P_{max}}{N_q} = \frac{3f_s^3}{8\pi f_m^2 f_M}$$ **Hence Proved.**

Example 10.24. A DM system is designed to operate at 3 times the Nyquist rate for a signal with a 3 kHz bandwidth. The quantizing step size is 250 mV.

(*i*) Determine the maximum amplitude of a 1-kHz input siinusoid for which the delta modulator does not show slope overload.

(*ii*) Determine the posfiltered output signal-to-quantizing-noise ratio for the signal of part (*i*)

Solution: We have

$$m(t) = A\cos\omega_m t = A\cos 2\pi(10^3)t$$

$$\left|\frac{dm(t)}{dt}\right|_{max} = A(2\pi)(10^3)$$

The maximum allowable amplitude of the input sinusoid is

$$A_{max} = \frac{\Delta}{\omega_m T_s} = \frac{\Delta}{\omega_m}$$

$$f_s = \frac{250}{2\pi(10^3)}3(2)(3)(10^3) = 71.2 \text{ mV}$$ **Ans.**

(*ii*) Assuming that the cutoff frequency of the low-pass filter is f_m, we have

$$(\text{SNR})_0 = \left(\frac{S}{N_q}\right)_0 = \frac{3[(3)(6)(10^3)]^3}{8\pi^2(10^3)^3} = 221.6 = 23.5\ dB$$ **Ans.**

Example 10.25. The pulse rate in a DM system is 56,000 per sec. The input signal is 5 cos (2π 1000 *t*) + 2 cos (2π 2000 *t*) *V*, with *t* in sec. Find the minimum value of step size which will avoid slope overload distortion. What would be the disadvantages of choosing a value of larger than the minimum?

(GATE Examination-1998)

Solution: Input signal,

$$m(t) = 5\cos(2000\,\pi t) + 2\cos(4000\,\pi t) = m_1(t) + m_2(t)$$

To avoid slope overloading, we have

$$\left|\frac{dm_1(t)}{dt}\right|_{max} \leq \Delta_1 f_s,$$

where Δ is step size and f_s is sampling rate.

or $$\Delta_1 f_s \geq \left|\frac{dm_1(t)}{dt}\right|_{max}$$

$$\Delta_{1min} \text{ (minimum value of step size)} = \left|\frac{dm_1(t)}{dt}\right|_{max} = \frac{5.(2000\pi)}{56000}$$

$$= \frac{10\pi}{56} = 0.56 \text{ V}$$

Also, $$\Delta_2, f_s \geq \left|\frac{dm_2(t)}{dt}\right|_{max}$$

or $$\Delta_2, \min = \frac{8000\pi}{56000} = \frac{\pi}{7} = 0.45 \text{ V}$$

Hence, larger step size out of two will be the required step size. i.e., = 0.56 V.

If a value larger than the minimum will be choosen, then granular noise will occur

Example 10.26. Bandwidth of the input to pulse code modulator is restricted to 4 kHz. The input varies from -3.8*V* to 3.8 *V* and has the average power of 30 mW, the required signal to quantization noise power ratio is 20 dB. The modulator produces binary output. Assume uniform quantization Calculate the number of bits required per sample.

Solution: Given, $$\left(\frac{S}{N_q}\right)_{odB} = 20 \text{ dB}$$

or $$10\log\left(\frac{S}{N_q}\right)_o = 20 \text{ dB}$$

or $$\left(\frac{S}{N_q}\right)_o = 100$$

Quantizer step size, $$\Delta = \frac{2A}{L}$$

where $L = 2^n$, *n* is the number of bianry digits.

then, average quantizing power is,

$$N_q = \langle q_e^2 \rangle = \frac{\Delta^2}{12} = \frac{A^2}{3L^2}$$

then, $\left(\frac{S}{N_q}\right)_o = \frac{\text{Average signal power}}{\text{Average quantizing power}}$

or $$100 = \frac{30\times10^{-3}}{A^2/3L^2}$$

or $$L = \sqrt{\frac{30\times10^{-3}}{3\times100\times(3.8)^2}} = 126.67$$

or $$2^n = 128$$

Hence, $n = 7$ = number of bits required per sample.

SUMMARY

1. The use of digital communications offers several important advantages as compared to analog communications.
2. Digital communications have become the dominant form of communication technology in our society.
3. To handle the transmission of analog message signals (*i.e.*, voice and video signals) by digital means, the signal has to undergo an analog-to-digital conversion.
4. The simplest form of pulse digital modulation is called pulse code modulation (PCM).
5. In this system (PCM) the message signal is first sampled and then amplitude of each sample is rounded off to the nearest one of a finite set of allowable values known as quantization levels, so that both time and amplitude are in the discrete form. This means that in pulse code modulation both parameters *i.e.*, time and amplitude are expressed in discrete form. This process is called discretization in time and amplitude.
6. We can improve the accuracy of the quantized signal to any desired degree simply by increasing the number of levels L.
7. Pulse-code modulation is known as a digital pulse modulation techniques. The pulse-code modulation (PCM) is quite complex compared to the Analog pulse modulation techniques (*i.e.*, PAM, PWM and PPM) in the sense that the message signal is subjected to a great number of operations.
8. The essential operations in the receiver are regeneration of impaired signals, decoding and demodulation of the train of quantized samples.
9. Quantization refers to the use of a finite set of amplitude levels and the selection of a level nearest to a particular sample value of the message signal as the representation for it.
10. This operation combined with sampling, permits the use of coded pulses for representing the message signal.
11. It is the combined use of quantizing and coding that distinguishes pulse code modulation form analog modulation techniques.
12. Basically, the quantizers are of two types:
    (*i*) Uniform quantizer
    (*ii*) Non-uniform quantizer.
13. A uniform quantizer is that type of quantizer in which the 'step size' remains same through out the input range.
14. A non-uniform quantizer is that type of quantizer in which the 'step-wise' varies according to the input values.
15. Transmission bandwidth in PCM is given by

$$BW \geq \frac{1}{2}r$$

But $$r = v f_s$$

Therefore, $BW \geq \frac{1}{2} v f_s$

Again, since $f_s \geq 2 f_m$

Hence, $BW \geq v f_m$

16. Because of quantization, inherent errors are introduced in the signal. This error is called quantization error.
17. Normalized noise power
18. Thus, Signal to Quantization noise ratio for normalized values of power 'P' and amplitude of input $x(t)$, will be

$$\left(\frac{S}{N}\right) dB \leq (4.8 + 6v)\, dB$$

19. The compression of signal at transmitter and expansion at receiver is called combinely as *companding*
20. The compression and expansion is obtained by passing the signal through the amplifier havin non-linear transfer characteristic.
21. The combination of a compression and an expander is called a compander. Naturally, in an actu PCM system, the combination of compressor and uniform quantizer is located in the transmitte whereas the expander is located in the receiver.
22. In the μ-law companding, the compressor characteristic is continuous. It is described as

$$|v| = \frac{\log(1 + \mu|m|)}{\log(1 + \mu)}$$

where m and v are the normalized input and output voltages, and μ is a positive constant.

23. Another compression law that is used in practice is the so-called A-law. In the A-law compandin the compressor characteristics is piecewise, made up of a linear segment for low-level inputs a a logarithmic segment for high level inputs. It is described as

$$|v| = \begin{cases} \frac{A|m|}{1 + \log A}, & 0 \leq |m| \leq \frac{1}{A} \\ \frac{1 + \log(A|m|)}{1 + \log A}, & \frac{1}{A} \leq |m| \leq 1 \end{cases}$$

24. The signal to noise ratio of PCM remains almost constant with companding.
25. Delta modulation transmits only one bit per sample. Here, the present sample value is compa with the previous sample value and this result whether the amplitude is increased or decrease transmitted.
26. The delta modulation has certain advantages over PCM as under:
    (*i*) Since, the delta modulation transmits only one bit for one sample, therefore the signal rate and transmission channel bandwidth is quite small for delta modulation compared PCM.
    (*ii*) The transmitter and receiver implementation is very much simple for delta modolation. Th is no analog to digital converter required in delta modulation.
27. The delta modulation has two major drawbacks as under:
    (*i*) Slope overload distortion,
    (*ii*) Granular or idle noise.
28. To overcome the quantization errors due to slope overload and granular noise, the step size (made adaptive to variations in the input signal $x(t)$.
29. Adaptive delta modulation has certain advantages over delta modulation as under:
    (*i*) The signal to noise ratio becomes better than ordinary delta modulation because of the reduc in slope overload distortion and idle noise.
    (*ii*) Because of the variable step size, the dynamic range of ADM is wider than simple DM.
    (*iii*) Utilization of bandwidth is better than delta modulation.

30. If this redundancy is reduced, then overall bit rate will decrease and number of bits required to transmit one sample will also be reduced. This type of digital pulse modulation scheme is known as Differential Pulse Code Modulation.

SHORT QUESTIONS WITH ANSWERS

Q.1. What is a pulse digital modulation scheme?

Ans. It is the modulation in which the message and also the carrier are in discrete form.

These are classified as:

Pulse code modulation and Delta modulation

Q.2. What are the advantages of digital representation of analog signals?

Ans. Some of the advantages of a digital signal over analog signal are:

(i) Ruggedness to transmission noise and interference

(ii) Efficient regeneration of the coded signal along the transmission path

(iii) The possibility of a uniform format for different kinds of baseband signals.

Q.3. Draw the block diagram of PCM scheme showing the elements required for the transmission.

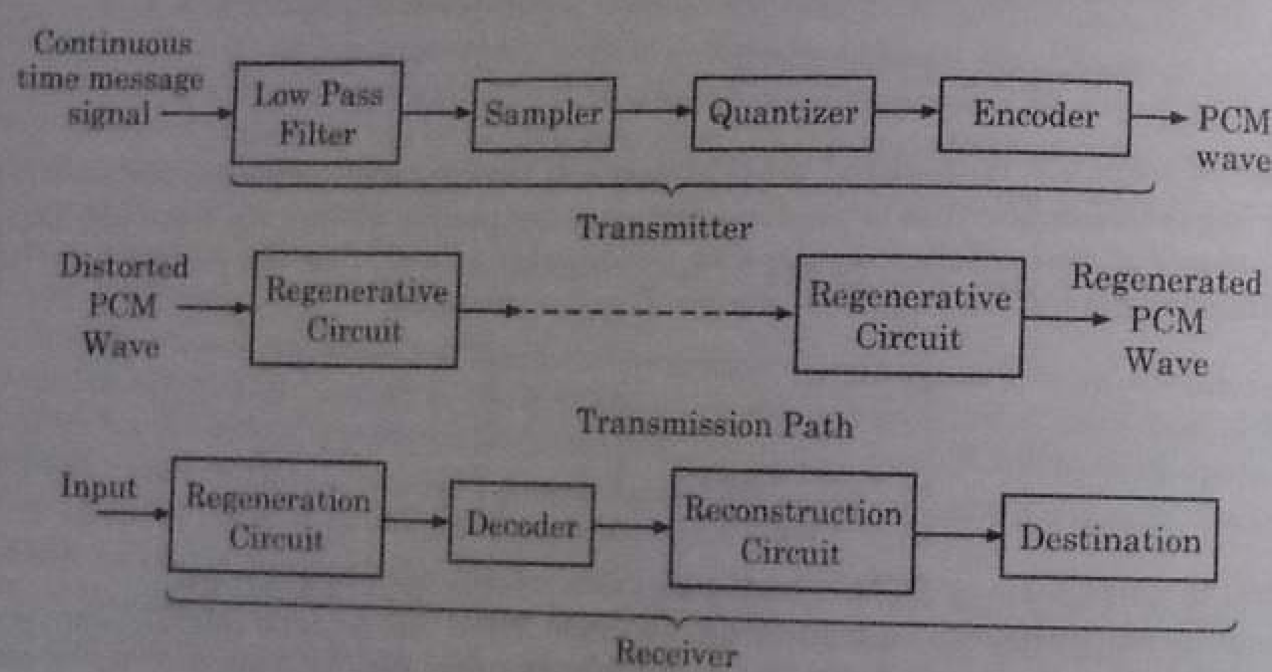


Fig. 10.29.

.4. Define Pulse Code Modulation.

s. It is the process in which the message signal is sampled and the amplitude of each sample is ...nded off to the nearest one of a finite sett of allowable values.

. Discuss Noise effect in PCM.

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

i) **Transmission Noise:** Which is introduced anywhere transmitter output and the receiver input. It is also named as channel Noise.

i) **Quantizing Noise:** This is introduced in the transmitter and is carried along to the receiver output.

Explain the importance (or) use of prediction in Differential pulse code modulation ...CM).

In standard PCM, each sample of the baseband signal is encoded independently of all others. However sometimes, in some signals, when they are samplec at the Nyquist rate (or) faster exhibit ...ficant correlation between successive samples i.e., the relative change in amplitude between the ...ssive samples is very small.

Under these circumstances, if this highly correlated signal is encoded using PCM, then the resultant signal consists of redundant Information and results in a lower bit rate. The is because the symbols that are not essential to the transmission of information are generated as a result of the encoding process.

By removing this redundancy before encoding, we obtain a more efficient coded signal.

A relatively simple solution is to encode the difference between successive samples rather than the samples themselves.

Since the difference between the samples are expected to be smaller than the actual sampled amplitudes, fewer bits are required to represent the differences.

Thus, if a sufficient part of a redundant signal is known it is possible to make the most probable estimate of the rest *i.e.*, if a small part of the sample is known, the other half can be estimated from the knowledge of previous sample.

This process is known as prediction.

Q.7. Give the block diagram of DPCM.

Ans.

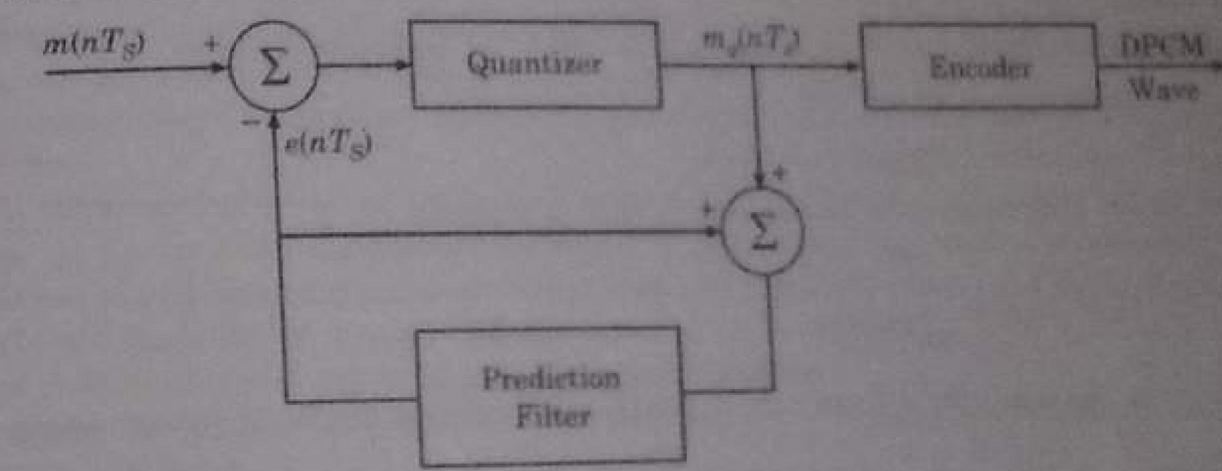


Fig. 10.30.

Q.8. What is Delta Modulation and give the comparison between DM and DPCM.

Ans. Delta Modulation is the one-bit (or) two-level version of DPCM.

They are similar except for two important differences namely, the use of a one-bit quantizer in delta modulator and the prediction filter is replaced by a single delay element.

Q.9. Discuss the Noise effects in Delta Modulation.

Ans. In Delta modulation we observe quantization noise. There are two major sources of quantizing error in DM systems. They are

(i) Slope overload distortion

(ii) Granular noise.

Q.10. Write a short notes on slope overload Distortion.

Ans. In general, the step size we choose to quantize, is fixed. So under maximum slope of the signal step size becomes small to follow the steep of the input waveform. This condition is called slope overload and the resulting quantizing error is called slop-overload distortion (noise). It is shown in figure 10.31.

Q.11. Write short notes on Granular noise.

Ans. In contrast to the slope overload distortion, the granular noise occurs when the step size Δ is too large relative to the local slope characteristics of the input waveform, thereby causing the staircase approximation to hunt around a relatively flat segment of the input waveform.

This is also known as hunting process.

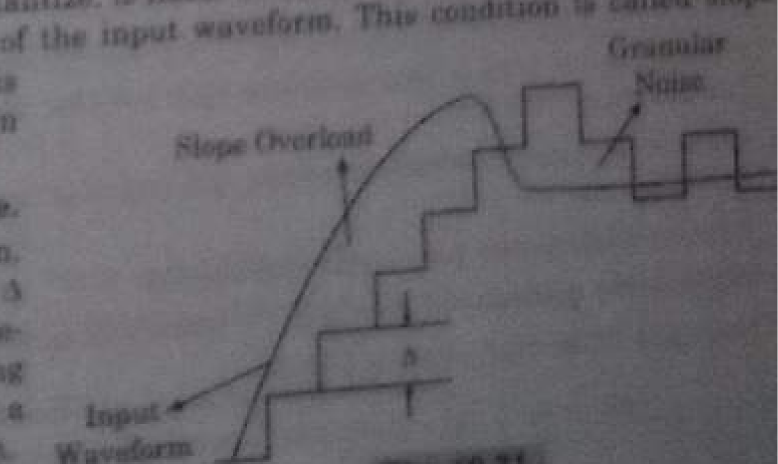


Fig. 10.31.

Q.12. Compare Delta Modulation and Pulse code Modulation schemes.

Ans. To compare the two modulation schemes, they should have some identical conditions.

Let us assume that both systems use approximately the same bandwidth for transmitting a baseband analog signal.

If f_s denotes the sampling rate of an N-bit PCM

f_s' denotes the sampling rate of N-bit DM

Then the bit rate of PCM is Nf_s and f_s' for DM.

If the signal spectrum extends up to f_m Hz, then $f_s = 2f_m$ and identical bandwidth requirements imply that

$$f'_s = 2Nf_m$$

1. SNR

If the channel signal to noise ratio is high, then the performance of PCM and DM is limited by the quantization noise.

Then the signal-to-quantizing noise power ratio for the PCM system is given by

$$(S_0/N_q)_{PCM} = Q^2 = 2^{2N} 2 \geq 2 \quad ...(i)$$

where $Q = 2^N$, the number of quantizer levels.

Similarly, for the DM system, the corresponding ratio is given by

$$(S_0/N_q)_{DM} = \frac{3}{8\pi^2}\left(\frac{f'_s}{f_m}\right)^3 \quad ...(ii)$$

From the above equations for a fixed bandwidth the performance of DM is always poorer than PCM.

For 8-bit PCM and DM, we have

$$(S_0/N_q)_{PCM} = 48 \text{ dB}$$

$$(S_0/N_q)_{DM} = 22 \text{ dB}$$

Thus the overall signal to noise ratio of a DM system is also lower than the overall signal-to-noise ratio of a PCM system using the same bandwidth.

REVIEW QUESTIONS

1. With the help of neat diagrams, explain the transmitter and receiver of pusle code modulation.
2. Explain what is uniform (linear) quantization ?
3. Explain the quantization error and derive an expression for maximum signal to noise ratio in PCM system that uses Linear quantization.
4. Derive the relations for signaling rate and transmission bandwidth in PCM system.
5. Explain Delta modulation in detail with suitable diagram. Also, explain ADM and compare its performance with DM.
6. What is the slope overload distortion and granular noise in delta modulation and how it is removed in ADM.
7. Explain Differential pulse code modulation.
8. What is the necessity of non-uniform quantization and explain companding 2.
9. Compare PCM, DM, ADM & DPCM.

NUMERICAL PROBLEMS

1. In the binary PCM system , find out the minimum number of bits reqired so that quantizing noise is less than ± k per cent of the analog level.
   [Ans. [illegible]]
2. What is the maximum power that may be transmitted without slope overload distortion?
   [Ans. [illegible]]

OBJECTIVE TYPE QUESTIONS

Fill up the Blanks

1. The essential operations in the transmitter of a PCM system are ______ and ______.
2. The quantizing and encoding are performed in the circuit which is called ______.
3. The existence of a finite number of ______ is a basic condition of PCM.
4. The conversion of an analog sample of the signal into a digital (discrete) form is called the ______ process.
5. The difference between two adjacent discrete values is called ______ (or) ______.
6. The ______ consists of difference between the input and output signals of the quantizer.
7. The use of a nonuniform quantizer is equivalent to passing the baseband signal through a ______ and then applying the compressed signal to a ______ quantizer.
8. The combination of a compressor and expander is called a ______.
9. Any plan for representing each of the discrete set of values as particular arrangement of dis... events is called a ______.
10. One of the discrete events in a code is called a ______ (or) ______.
11. A particular arrangement of symbols used in a code to represent a single value of the discre... is called a ______ (or) ______.
12. In a ______ code, each symbol may be one of the three distinct values (or) kinds.
13. The process in which the information in a binary PCM is encoded in terms of signal transiti... referred to as ______.
14. The capability of controlling the effects of distortion and noise produced by transmitting a... wave through a channel lies in reconstructing it by using ______.
15. The ______ process involves generating a pulse the amplitude of which is the linear sum of... pulses in the code word, with each pulse weighted by its place value ($2^0, 2^1, 2^2, 2^3, \ldots$) in th...
16. The basic operations performed by ______ are equalization, timing and decision making.
17. If the spacing between received pulses deviates from its assigned value, a ______ is intr... into the regenerated pulse position, thereby causing distortion.
18. ______ noise may be introduced anywhere between the transmitter, output and the receiver...
19. ______ noise may be introduced in the transmitter and is carried along to the receiver ou...
20. The average probability of error in a PCM receiver depends on the ratio of ______ to ... measured at the decoder input in the receiver.
21. The important characteristic of a PCM system is its ______ to interference.
22. The ______ is the one bit version of DPCM.
23. A Delta modulator using a fixed step size is often referred to as ______.
24. The ______ noise occurs when the step size is too large relative to the local slope charac... of the input waveform.
25. The ______ and ______ are the two noise effects in Delta modulation.
26. The method in which the step size is adapted to the level of the input signal is called ...

n terms of SNR performance, the adaptive delta modulation is ______ to linear delta modulation y ______ dB at a bit rate of ______ .

ompanding is used [ ]

) to overcome quantizing noise in PCM

) in PCM transmitters, to allow amplitude limiting in the receivers

) to protect small signals in PCM from quantizing distortion.

in PCM receivers, to overcome impulse noise.

e biggest disadvantages of PCM is [ ]

its inability to handle analog signals

the high error rate which its quantizing noise introduces

its incompatibility with TDM

the large bandwidths that are required for it.

ate which of the following pulse modulation systems is analog [ ]

PCM	(b) Differential PCM
WM	(d) Delta modulation

izing noise occurs in [ ]

me-division multiplexing	(b) FDM
CM	(d) PWM

S

ng, quantizing and encoding

to-Digital converter	3. discrete amplitude levels	4. Quantizing
n, step size	6. Quantizing error	7. Compressor, uniform
ler	9. code	10. code element, symbol
l. character	12. ternary code	13. Differential encoding
tive repeaters	15. decoding	16. Regenerative repeater
	18. Transmission noise	19. Quantizing noise
e power to the average noise power		21. Ruggedness
lulation	23. linear delta modulator	24. Granular noise
load, granular noise	26. Adaptive delta modulator	
dB, 20 kilobits/sec.	28. (a)	29. (a)
	31. (c)	

□□□

CHAPTER 11

Digital Multiplexers

Inside this Chapter

11.1. Introduction

In last two chapters, we have discussed the pulse analog modulation and pulse digital modulation methods. In this chapter, we shall discuss an important process in communication system known as multiplexing. In the articles to follow, we shall discuss what is multiplexing and what are the types of multiplexing and then proceed for a detailed aspects of multiplexing.

11.2. Multiplexing

Multiplexing may be defined as a technique which allows many users to share a common communication channel simultaneously. There are two major types of multiplexing techniques. They are as under:

(i) Frequency division multiplexing (FDM),

(ii) Time division multiplexing (TDM).

11.2.1. Frequency Division Multiplexing (FDM)

This technique permits a fixed frequency band to every user in the complete channel bandwidth. Such frequency slot is allotted continuously to that user. As an example consider that the channel bandwidth is 1 MHz. Let there be ten users, each requiring upto 100 kHz bandwidth. Then the complete channel bandwidth of 1 MHz can be divided into ten frequency bands, i.e. each of 100 kHz and every user can be allotted one independent frequency band. This technique is known as **Frequency Division Multiplexing** (FDM). It is mainly used for modulated signal. This is due to the fact that a modulated signal can be placed in any frequency band by just changing the carrier frequency. However, at the receiver, these frequency multiplexed signals can be separated by the use of tuned circuits (*i.e.*, bandpass filters) of their respective frequency band. And for every band, there are independent tuned circuits and demodulators.

11.2.2. Time Division Multiplexing (TDM)

As discussed earlier, in PAM, PPM and PDM the pulse is present for a short duration and for most of the time between the two pulses, no signal is present. This free space between the pulses can be occupied by pulses from other channels. This is known as Time Division Multiplexing (TDM). Thus, time division multiplexing (TDM) makes maximum utilization of the transmission channel.

Hence, we can say that in FDM, all the signals are transmitted simultaneously over the same communication medium, and the signals occupy frequency slots. However, in TDM, the signals to be multiplexed are transmitted sequentially one after the other. Each signal occupies a short time slot as shown in figure 11.1. Thus, the signals are isolated from each other in the time domain, but all of them occupy the same slot in the frequency spectrum. Therefore, in TDM the complete bandwidth of the communication channel is available to each signal being transmitted.

Figure 11.1 shows the concept of TDM.

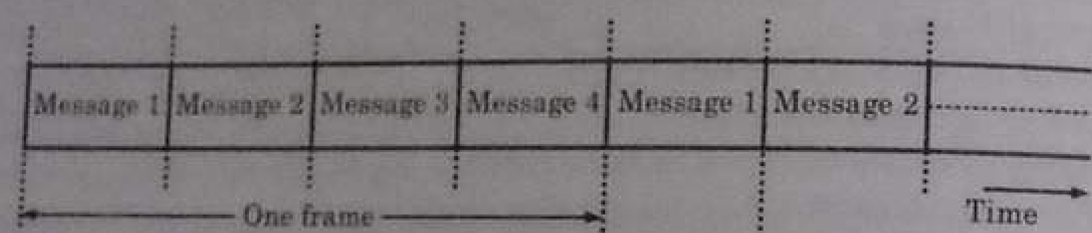


Fig. 11.1. *Illustration of TDM concept*

At this stage, it may be noted that in context of TDM, we define one important term *i.e.*, frame. One frame corresponds to the time period required to transmit all the signals once on the transmission channel. This has been shown in figure 11.1. Here, we have total four message signals to be transmitted. Hence, one frame will correspond to the time period required to transmit all the four signals once on the channel.

The TDM system can be used to multiplex analog or digital signals, however it is more suitable for the digital signal multiplexing.

11.3. A PAM/TDM System

Now, let us discuss a PAM/TDM system. Infact, this system combines the concepts of PAM and TDM both as shown in figure 11.2.

Working Principle

Here, the multiplexer is a single pole rotating switch or commutator. This switch can be a mechanical switch or an electronic switch and it rotates at f_s rotations per second. As the switch arm rotates, it is going to make contact with the position 1, 2, 3 or N for a short time. There are N analog signals, to be multiplexed, which are connected to these contracts. Hence, the switch arm will connect these N input signals one by one to the communication channel.

The waveform of a TDM signal which is being transmitted has been shwon in figure 11.3. It shows that the rotary switch samples each message during each of its rotations.

Since, each rotation corresponds to one frame, therefore, one frame is completed in T_s seconds where $T_s = 1/f_s$.

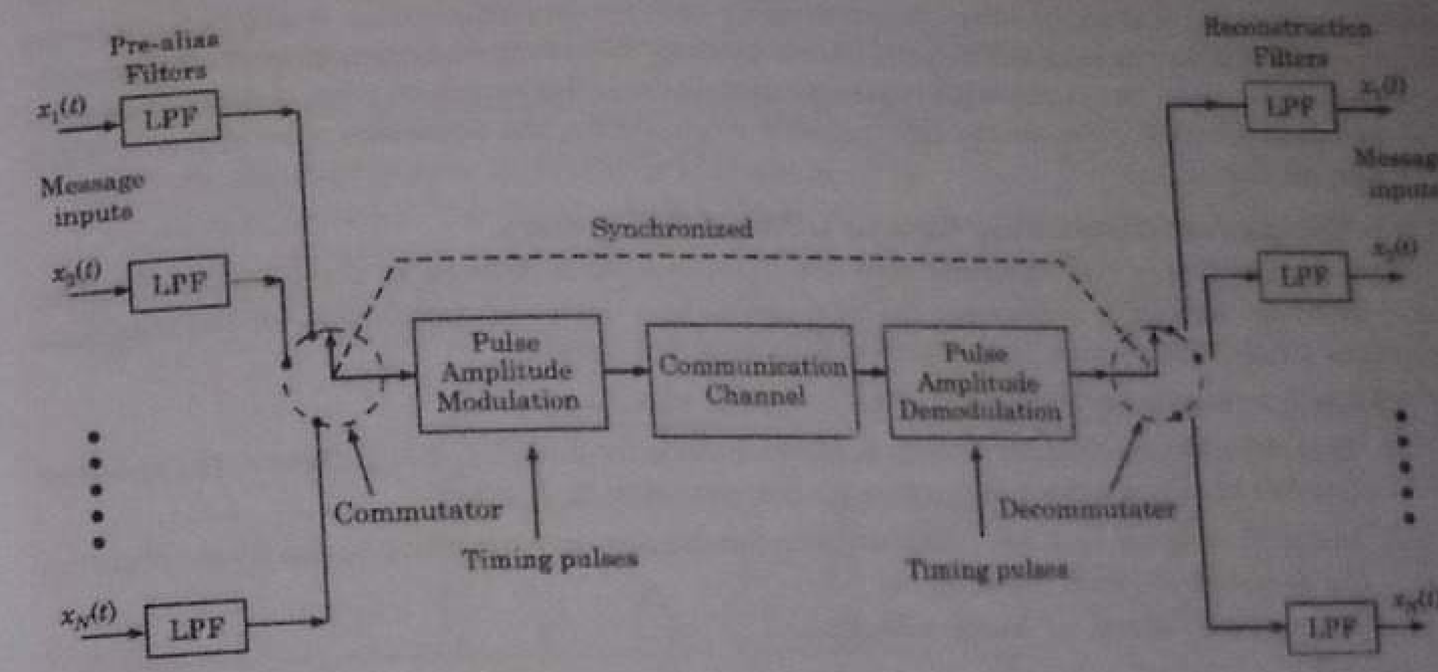


Fig. 11.2. *Block diagram of a PAM/TDM system*

Hence, the function of the commutator is two fold as under:

(*i*) To take narrow sample of each input message at a rate f_s which is higher than $2f_m$.

(*ii*) To sequentially interleave the N samples inside the interval $T_s = 1/f_s$.

Now, the multiplexed signal at the output of the commutator is applied to a pulse amplitude modulator. It converts the PAM pulses into a form suitable for transmission over the communication channel. The input message signals are passed through low pass filters before applying them to the commutator. These filters are actually the antialiasing filters which avoid the aliasing. The cutoff frequency of each low pass filter (LPF) is f_m Hz.

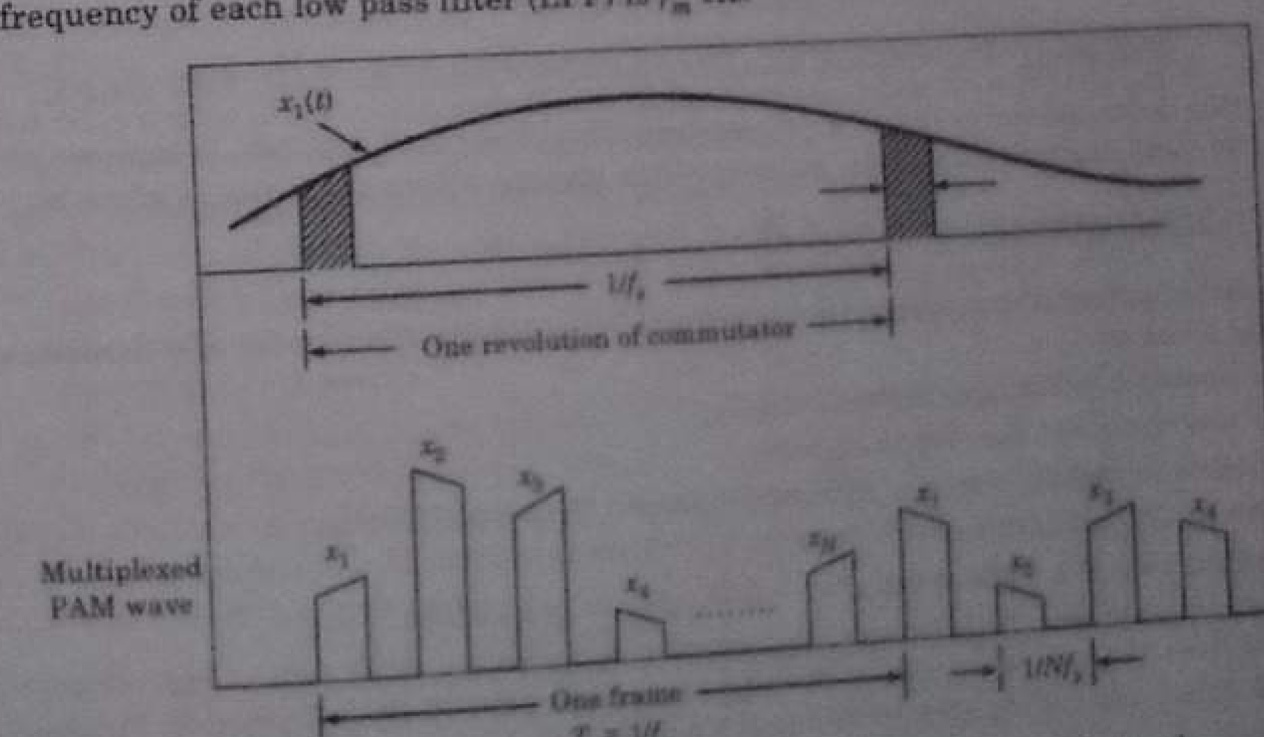


Fig. 11.3. *(a) Sampling of the first input (b) Multiplexed PAM signal transmitted on the transmission channel*

At the receiving end of PAM/TDM system, the received signal is applied to a pulse amplit... demodulator which performs the reverse operation of pulse amplitude modulator. At the recei...

there is one more rotating switch or decommutator used for demultiplexing. It will be interes... to know that this switch must rotate at the same speed as that of the commutator at the transm... and its position must be synchronized with commutator in order to ensure proper demultiple... The low pass filters (LPFs) on the receiver side are used for the reconstruction of the origi... message signals.

11.3.1. Evaluation of Signaling Rate in a PAM/TDM System

As a matter of fact, the signaling rate of a TDM system is defined as the number of pul... transmitted per second. It is represented by r. Let us now derive an expression for the signali... rate of the PAM/TDM system in the form of following few points:

(i) Let f_m = maximum frequency of all the input signals x_1 to x_N.

(ii) Therefore, as per Nyquist criterion, the sampling frequency $f_s \geq 2f_m$. Hence, the speed ... rotation of commutators is f_s rotations per second with $f_s \geq 2 f_m$.

(iii) As shown in figure 11.4, one revolution of commutators corresponding to one frame contai... one sample from each input signal.

Hence, 1 revolution = 1 frame = N pulses ...(11...

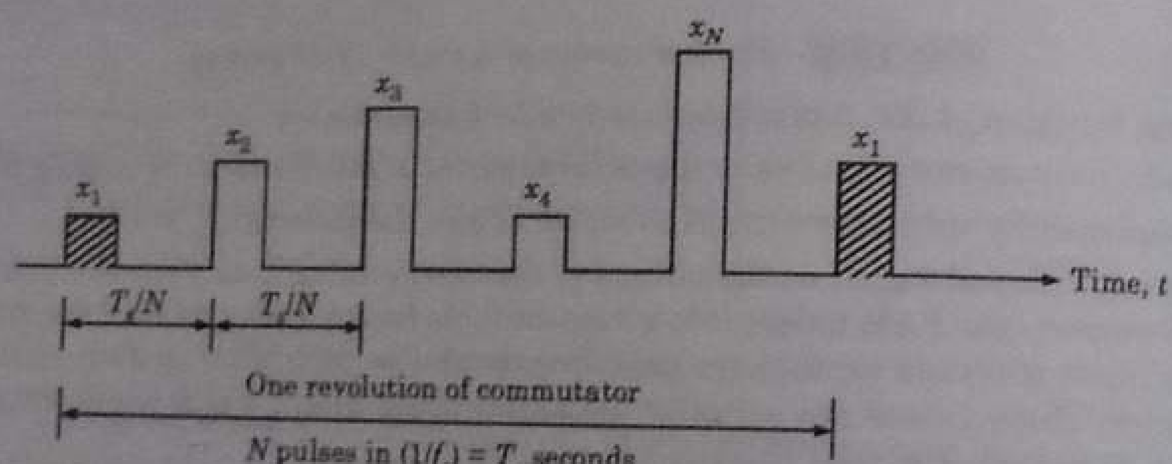


Fig. 11.4. *Evaluation of number of pulses per second for PAM/TDM system.*

(iv) One frame period is $(1/f_s)$ i.e., T_s seconds. Therefore, in T_s seconds, N number of pulses are transmitted. Hence, the pulse to pulse spacing within the frame is given by,

$$\text{Pulse to pulse spacing} = \frac{T_s}{N} = \frac{1}{Nf_s} \quad ...(11...$$

(v) As the period of one pulse (ON + OFF) is $(1/Nf_s)$ seconds, the number of pulses per second is given by,

Number of pulses per second = Nf_s

This is nothing but the signaling rate.

Therefore, signaling rate of a TDM system = $r = Nf_s$ pulses/second.

But as $f_s \geq 2f_m$, therefore,

Signaling rate of a TDM system = $r \geq 2Nf_m$ pulses per second.

It may be noted that TDM system is supposed to have its signaling rate as high as possible. It is evident from the expression above that the signaling rate can be increased by increasing the sampling rate f_s and/or the number of input signals N.

11.3.2. Transmission Bandwidth of a PAM/TDM Channel

The minimum transmission bandwidth of a PAM-TDM channel is given by

$$BW = \frac{1}{2}\text{(signaling rate)}$$

Therefore, minimum transmission bandwidth $BW \geq \frac{1}{2} \times 2N f_m$

Hence, minimum transmission bandwidth $BW = Nf_m$

11.3.3. Synchronization in PAM/TDM System

As a matter of fact, the multiplexed PAM signals can be received properly if and only if th... transmitter and receiver commutators are synchronized to each other in terms of the speed an... the position. In order to ensure synchronization, a marker pulse is introduced at the end of ea... frame in the transmitted signals as shown in figure 11.5.

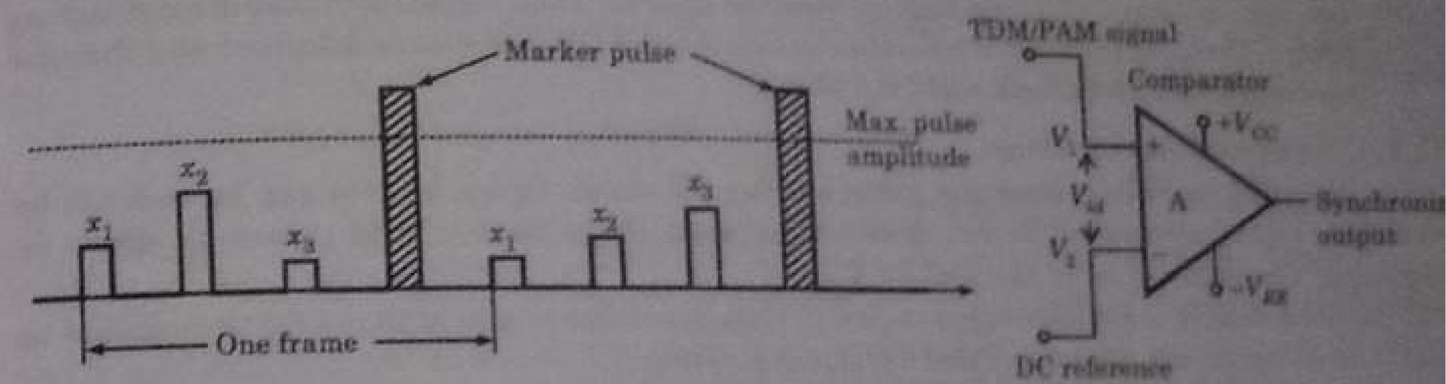


Fig. 11.5. *Illustration of frame synchronization and detection.*

The amplitude of this pulse is kept higher than the maximum permissible amplitude of multiplexed channels. At the receiver end, the received signal is compared with a DC refere... level. The comparator responds to only the marker pulse to produce output. Thus, the ma... pulse is separated from the remaining multiplexed channels. Due to the introduction of synchroni... pulse, only three signals instead of four can now be transmitted.

11.3.4. Concept of Crosstalk in a PAM/TDM System

Crosstalk basically means interference between the adjacent TDM channels. Infact, it ... unwanted coupling of information from one channel to the other. The guard time 'T_g' is the spacing introduced between the adjacent TDM channels.

Let us note few important points about crosstalk as under:

(i) The communication channel over which the TDM signal is travelling should ideally ... an infinite bandwidth in order to avoid the signal distortion. However, in practic... the communication channels have a finite bandwidth. Such channels are known a... bandlimited channels.

(ii) Whenever, a signal is passed over such bandlimited channel, the shape of the signa... change as shown in figure 11.6 (a).

(iii) Whenever a PAM/TDM signal is transmitted over a bandlimited channel, the s... corresponding to $x_1(t)$ will get mixed with $x_2(t)$ as shown in figure 11.6 (b) and this ov... will result into crosstalk.

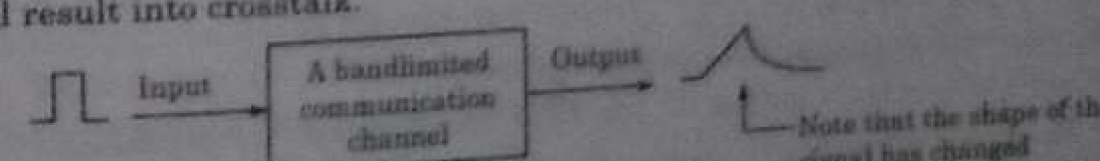


Fig. 11.6. *(a) Transmission of signal over a bandlimited channel*

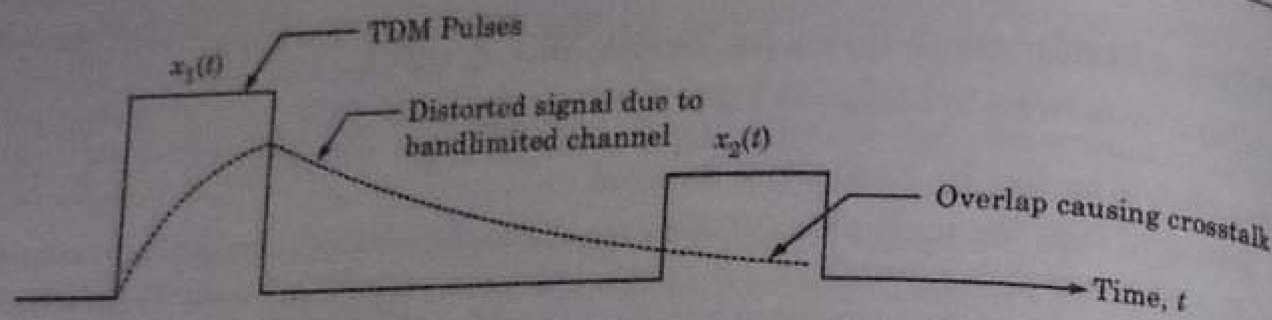


Fig. 11.6. *(b) Illustration of crosstalk in TDM*

(*iv*) One more reason for the crosstalk between the adjacent TDM signals is the use of bandlimiting filters. Because of these filters, the shapes of the TDM pulses are distorted and they get overlapped and crosstalk will take place.

11.3.5. Concept of Guard Time

The crosstalk resulting from the pulse overlap shown in figure 11.6(*b*) can be reduced by introducing guard time of sufficient duration between the adjacent TDM pulses as shown in figure 11.7. The guard time is denoted by T_g.

If the crosstalk is to be kept below – 30 dB, then the value of guard time (T_g) that should be introduced between the adjacent TDM signals is given by,

Guard time $$T_g > \frac{0.55}{BW} \quad ...(11.3)$$

where BW = Bandwidth of the channel.

The equation (11.3) shows that the guard time required to avoid the crosstalk will decrease with increase in the channel bandwidth. As T_g reduces, we can increase the signaling rate of the PAM/TDM system.

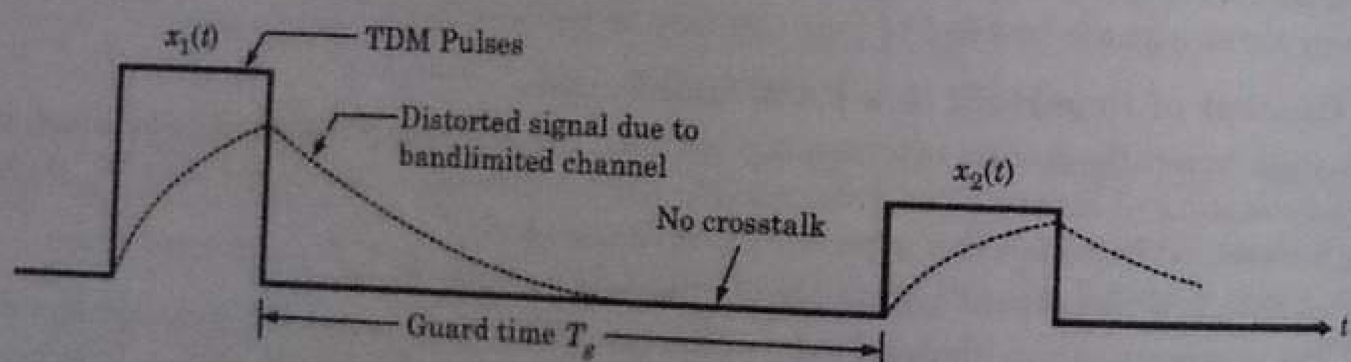


Fig. 11.7. *Elimination of crosstalk due to guard time.*

11.3.6. Advantages of TDM

(*i*) Full available channel bandwidth can be utilized for each channel.

(*ii*) Intermodulation distortion is absent.

(*iii*) TDM circuitry is not very complex.

(*iv*) The problem of crosstalk is not severe.

11.3.7. Disadvantages of TDM

(*i*) Synchronization is essential for proper operation.

(*ii*) Due to slow narrowband fading, all the TDM channels may get wiped out.

Example 11.1. To analog signals $x_1(t)$ and $x_2(t)$ are to be transmitted over a common channel by means of time division multiplexing (TDM). The highest frequency of $x_1(t)$ is 4 kHz and that of $x_2(t)$ is 4.5 kHz. What will be the minimum value of permissible sampling rate?

Solution: The highest frequency component of the composite signal consisting of $x_1(t)$ and $x_2(t)$ is 4.5 kHz. Therefore, the minimum value of permissible sampling rate will be,

$$f_{s(min)} = 2 \times 4.5 \text{ kHz} = 9 \text{ kHz} \qquad \textbf{Ans.}$$

Example 11.2. A signal $x_1(t)$ is bandlimited to 3 kHz. There are three more signals $x_2(t)$, $x_3(t)$ and $x_4(t)$ which are bandlimited to 1 kHz each. These signals are to be transmitted by a TDM system.

(*i*) Design a TDM scheme where each signal is sampled at its Nyquist rate.

(*ii*) What must be the speed of the commutator?

(*iii*) Calculate the minimum transmission bandwidth of the channel.

Solution: (*i*) Table 11.1 shows different message signals with corresponding Nyquist rates.

Table 11.1

S.No.	Message signal	Bandwidth	Nyquist rate
1.	$x_1(t)$	3 kHz	6 kHz
2.	$x_2(t)$	1 kHz	2 kHz
3.	$x_3(t)$	1 kHz	2 kHz
4.	$x_4(t)$	1 kHz	2 kHz

If the sampling commutator rotates at the rate of 2000 rotations per second then the signals $x_2(t)$, $x_3(t)$ and $x_4(t)$ will be sampled at their Nyquist rate. But, we have to sample $x_1(t)$ also at its Nyquist rate which is three times higher than that of the other three. In order to achieve this, we should sample $x_1(t)$ three times in one rotation of the commutator. Therefore, the commutator must have atleast 6 poles connected to the signals as shown in figure 11.8.

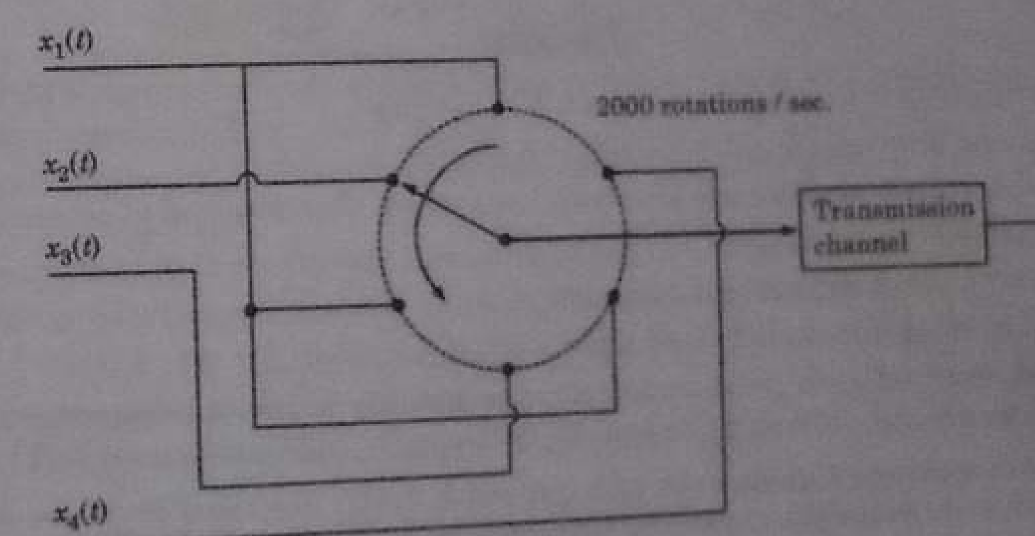


Fig. 11.8.

(*ii*) The speed of rotation of the commutator is 2000 rotations/sec.

(*iii*) Number of samples produced per second is calculated as under:

$x_1(t)$ produces 3 × 2000 = 6000 samples/sec.

$x_2(t)$, $x_3(t)$ and $x_4(t)$ produce 2000 samples/sec. each.

Therefore, number of samples per second = 6000 + (3 × 2000) = 12000 samples/sec.

∴ Signaling rate = 12000 samples/sec.

(*iv*) The minimum channel bandwidth will be

$$BW = \frac{1}{2} \text{ (signaling rate)} = 12000/2$$

or $$BW = 6000 \text{ Hz} \qquad \textbf{Ans.}$$

Example 11.3. Twenty-four voice signals are sampled uniformly and then time division multiplexed. The sampling operation uses flat top samples with 1 μs duration. The multiplexing operation includes provision for synchronization by adding an extra pulse of appropriate

amplitude and 1 μs duration. The highest frequency component of each voice signal is 3.4 kHz.

(i) Assuming a sampling rate of 8 kHz, calculate the spacing between successive pulses of the multiplexed signal.

(ii) Repeat (i) assuming the use of Nyquist rate sampling.

Solution: (i) Given that

Sampling rate = 8 kHz = 8000 samples/sec.

There are 24 voice signals + 1 synchronizing pulse.

Pulse width of each voice channel and synchronizing pulse is 1 μs.

Now, time taken by the commutator for 1 rotation = $\frac{1}{8000}$ = 125 μ sec.

Number of pulses produced in 1 rotation = 24 + 1 = 25

Therefore, the leading edges of the pulses are at $\frac{125}{25}$ = 5 μ seconds distance as shown in figure 11.9.

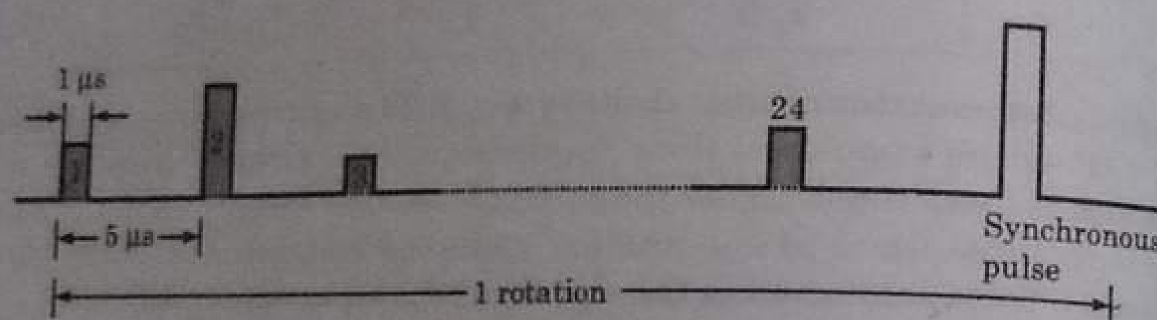


Fig. 11.9.

Hence, spacing between successive pulses = 5 – 1 = 4 μs

(ii) Nyquist rate of sampling = 2 × 3.4 kHz = 6.8 kHz.

This means that 6800 samples are produced per second. One rotation of commutator takes

1/6800 = 147 μs time

Therefore, 147 μ sec corresponds to 25 pulses

Therefore, 1 pulse corresponds to 5.88 μ sec.

As the pulse width of each pulse is 1 μ sec, the spacing between adjacent pulses will be 4.88 μ sec and if we assume τ = 0 then the spacing between the adjacent pulses will be 5.88 μ sec.

Example 11.4. Six message signals each of bandwidth 5 kHz are time division multiplexed and transmitted. Determine the signaling rate and the minimum channel bandwidth of the PAM/TDM channel.

Solution: The number of channels $N = 6$

Bandwidth of each channel, $f_m = 5$ kHz

∴ Minimum sampling rate = 2 × 5 kHz = 10 kHz

Signaling rate = Number of bits per second

= 6 × 10 kHz = 60 K bits/sec. **Ans.**

Minimum, channel bandwidth to avoid cross talk in PAM/TDM is,

$BW = Nf_m$

= 6 × 5 kHz = 30 kHz **Ans.**

Example 11.5. Sketch a channel interleaving scheme for the time division multiplexing the following PAM signals: Five 4 kHz telephone channels and one 20 kHz music channel. Find the pulse repetition rate of the multiplexed signal and estimate the minimum system bandwidth required.

Solution: Let us note the following points:

(i) Each telephone channel of bandwidth 4 kHz must be sampled at Nyquist rate i.e., 2 × 4 kHz = 8 kHz using a TDM commutator.

(ii) The 20 kHz music channel must be sampled at 40 kHz (Nyquist rate) hence a separate sampler is required.

(iii) The sampled signals are applied to two 4-bit A-D converters to obtain the equivalent digital signals.

(iv) These signals are finally multiplexed using a multiplexer as shown in figure 11.10.

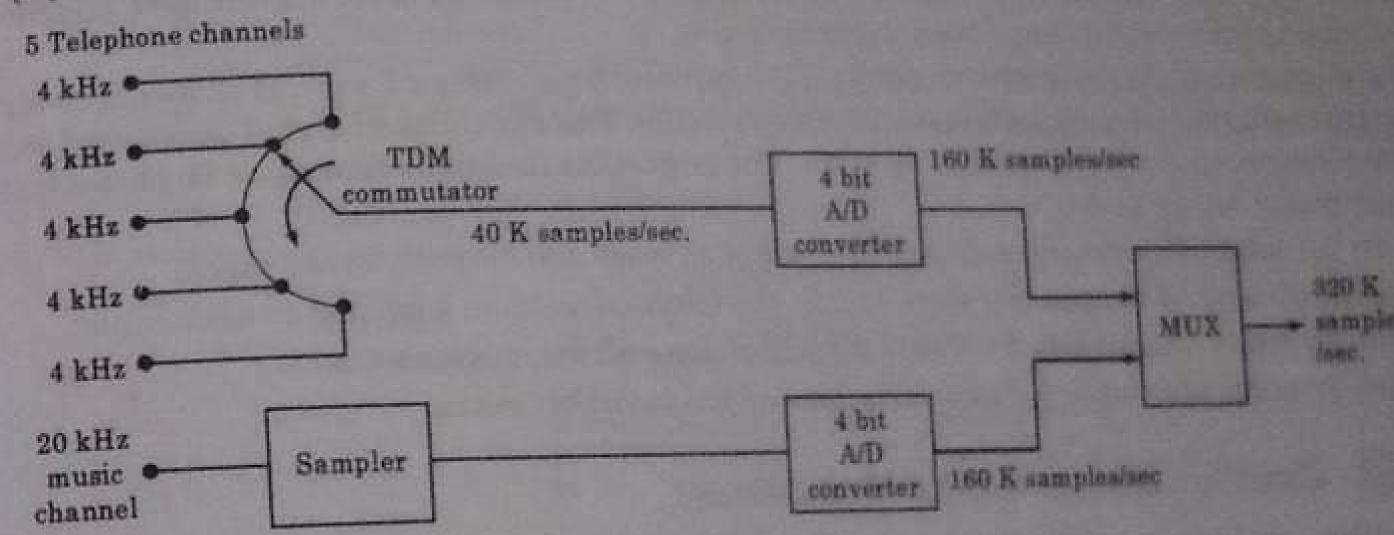


Fig. 11.10. *PAM/TDM system for example 11.5.*

(v) The TDM commutator output has a pulse repletion rate of 40 k samples/sec as there are 5 channel and sampling rate is 8 kHz.

(vi) Similarly, the output of the separate sampler has a pulse repetition rate of 40 K samples/sec.

(vii) The outputs of A-D converters has pulse repetition rates of 40 × 4 = 160 K samples/sec.

(viii) Therefore, pulse repetition rate at the output of a multiplexer is

160 + 160 = 320 K samples/sec.

Hence, pulse repetition rate of the system = 320 kHz

Therefore, bandwidth required = bit rate = 320 kHz. **Ans.**

11.4. Introduction to Digital Multiplexers

In Article 11.3, we have discussed the idea of TDM. In this section, let us consider the multiplexing of digital signals at different bit rates. The digital multiplexing will enable us to combine many digital signals such as computer outputs, digital voice, digitized facsimile and TV signals.

Figure 11.11 illustrates the concept of digital multiplexing and demultiplexing.

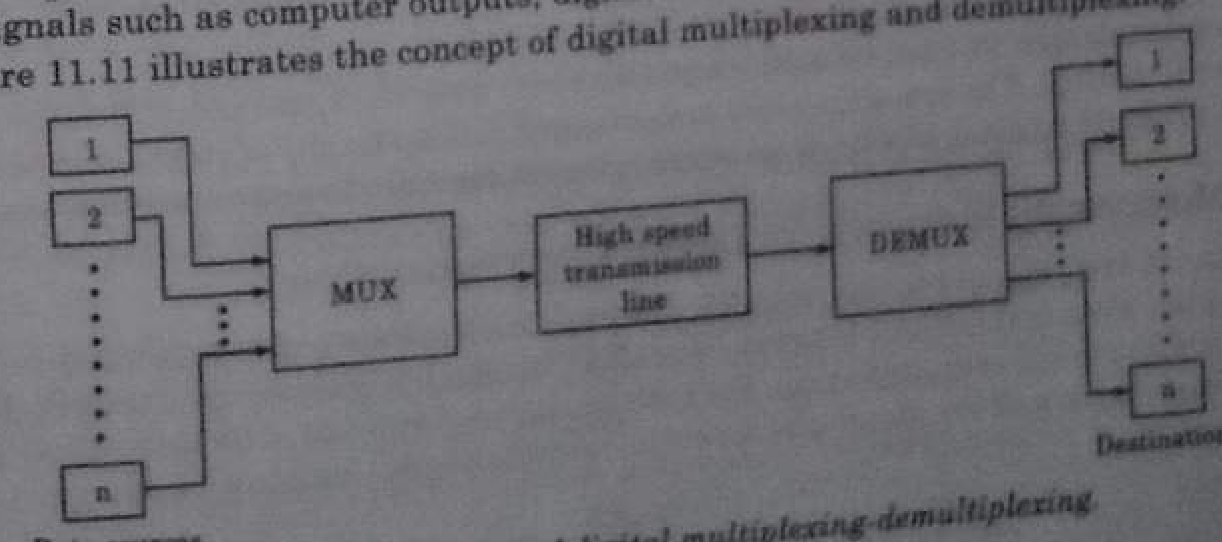


Fig. 11.11. *Concept of digital multiplexing-demultiplexing*

The digital data can be multiplexed by using a bit-by-bit interleaving procedure. This is achieved by using a selector switch which sequentially selects a bit from each input and places over the high speed transmission line. At the receiving end, the bits received on the common line are separated out and delivered to their respective destinations.

11.4.1. Principle of Digital Multiplexing

Digital multiplexing is based on the principle of "interleaving symbols" from two or more digital signals. This is similar to TDM but it does not require the periodic sampling and waveform preservation. The signals which are to be multiplexed may come from digital data sources or from analog sources that have been digitally coded.

A digital multiplexer is used to merge the input bits from different sources to form one signal from transmission via a digital communication system. The multiplexed signal consists of source digits interleaved bit by bit or word by word. The important functions that must be performed by a multiplexer are as under:

(i) To establish a frame. A frame consists of at least one bit from every input.
(ii) A number of unique bits slots within the frame should be assigned to each input.
(iii) To insert control bits for frame identification and synchronization.
(iv) To make allowance for any variations of the input bit rates.

11.5. Classification of Digital Multiplexers

The various digital sources that are to be multiplexed will have different bit rates. In practice the bit rate variation poses the most serious design problem and leads to the three categories of multiplexers, as under:

(i) Synchronous multiplexers
(ii) Asynchronous multiplexers
(iii) Quasi-synchronous multiplexers.

(i) Synchronous Multiplexers

In the synchronous digital multiplexers, a master clock governs all the sources. Therefore all the sources will operate at the same bit rate. As the bit rate variations are completely eliminated the synchronous multiplexing systems attain a very high throughput efficiency. But, they need elaborate provisions to distribute the master clock signal, to all the sources.

(ii) Asynchronous Multiplexers

The asynchronous multiplexers are used for the digital data sources which operate in the start/stop mode. These sources produce data in the form of "bursts" of characters with a variable spacing between the bursts. The techniques such as "buffering" and "character interleaving" makes it possible to merge these sources into a synchronous multiplexed bit stream.

(iii) Quasi-synchronous Multiplexer

These multiplexers are used when the input bit rates have the same nominal value but vary within specific bounds. These multiplexers are arranged in a hierarchy of increasing bit rates to constitute the basic building blocks of an interconnected digital telecommunication system.

11.6. Multiplexing Hierarchy for Digital Communication

As a matter of fact, there are two slightly different multiplexing patterns used for digital communication namely the AT&T hiearchy and CCIT hierarchy as shown in Table 11.2.

There are four levels of multiplexing. For both the types of systems, a 64 kb/s voice PCM unit is used as basic input and the structural layout is same as shown in figure 11.12.

Table 11.2. Multiplexing Hierarchies

S.No.	Level	AT & T		CCIT	
		Numer of inputs	Output rate Mb/s	Number of inputs	Output rate Mb/s
1.	First level	24	1.544	30	2.048
2.	Second level	4	6.312	4	8.448
3.	Third level	7	44.736	4	34.368
4.	Fourth level	6	274.176	4	139.264

Figure 11.12 shows the multiplexing hierarchy for digital communication.

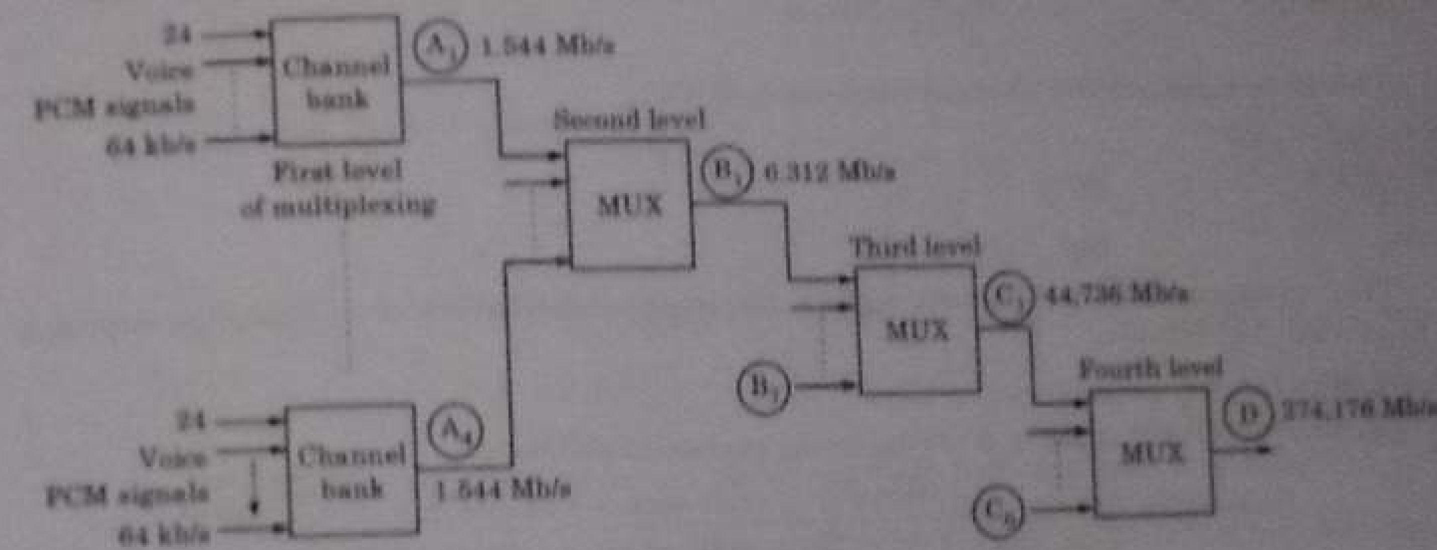


Fig. 11.12. *Multiplexing hierarchy for digital communication.*

At the first level of multiplexing, a number of 64 kb/s PCM voice channels are multiplexed together. The number of such voice inputs is 24 for each channel bank (for AT and T system). The bit rate at point A_1 in figure 11.12 is 1.544 Mb/s. Four such channel banks produce outputs A_1, A_2, A_3 and A_4. These outputs are further multiplexed by using a second level multiplexer. The bit rate at the second level multiplexer output *i.e.*, at point B_1 is 6.312 Mb/s. Seven such outputs i.e., B_1, B_2, ..., B_7 are multiplexed further by using a third level multiplexer. The bit rate at the output of the third level multiplexer i.e., at point C_1 is 44,736 Mb/s. Six such outputs *i.e.*, C_1, C_2, ..., C_6 are then multiplexed in the fourth level multiplexer. The bir rate at the output of the fourth level multiplexer i.e., at point D is 274,176 Mb/s.

11.6.1. Calculation of Bit Rate at Each Level

For an *AT&T* system, the bit rates at various levels of multiplexing are as under:

(i) Bit rate at the first level of multiplexing = 64 kb/sec × 24 Channels ...(11.6)
= 1.536 Mb/sec ...(11.7)

Bit rate assigned at the output of first level = 1.544 Mb/sec

(ii) Bit rate at the output of second level = 1.544 Mb/sec × 4 Inputs ...(11.8)
= 6.176 Mb/sec ...(11.9)

Bit rate assigned at the output of second level = 6.312 Mb/sec

(iii) Bit rate at the output of third level = 6.312 Mb/sec × 7 Inputs ...(11.10)
= 44.184 Mb/sec ...(11.11)

Bit rate assigned at the output of third level = 44.736 Mb/sec

(iv) Bit rate at the output of fourth level = 44.736 Mb/sec × 6 Inputs
= 268.416 Mb/sec ...(11.12)

Bit rate assigned at the output of fourth level = 274.176 Mb/sec ...(11.13)

(v) Total number of voice PCM channels = 24 × 4 × 7 × 6 = 4032 ...(11.14)

Note: *At this stage, it may be noted that at every level of multiplexing, the output bit rate is lower than the bit rate assigned at that level. This higher value allows us to add control bits and additional bits called "stuff bits" required to be added to yield a steady output rate.*

11.6.2. Bit Rate and Transmission Channel Bandwidth

The bit rate at the output of fourth level of *AT&T* system is given by,

Bit rate, $r = 274.176$ Mb/sec.

Therefore transmission channel bandwidth, $BW \geq \frac{r}{2}$

or $$BW \geq \frac{274.176}{2}$$

or $$BW \geq 137 \text{ Mb/sec.}$$

Thus, to transmit 4032 PCM voice signals, the required bandwidth of multiplexing system is greater than 137 Mb/sec.

11.6.3. Bandwidth Efficiency

The bandwidth efficiency of a multiplexing system is defined as,

$$\text{Bandwidth Efficiency} = \frac{\text{Number of voice signals} \times \text{Bandwidth of each signal}}{\text{Bandwidth of the multiplexing system}}$$

or $$\text{Bandwidth Efficiency} = \frac{4032 \times 4 \text{ kHz}}{137 \text{ MHz}} = 11.77\%$$

Thus bandwidth efficiency of a digital multiplexing system is only 11.77% which is extremely poor as compared to a much higher bandwidth efficiency of an analog multiplexing system (typically above 85%).

However this disadvantage of poor bandwidth efficiency is outweighted by the other advantages of digital transmission.

11.6.4. Advantages of Digital Multiplexing

(i) Hardware cost reduction due to the use of digital ICs.

(ii) Powder cost reduction due to use of regenerative repeaters.

(iii) More flexibility as compared to the analog multiplexers. This has been illustrated in figure 11.13 which shows a digital multiplexer which can multiplex voice signals, digital data, TV and videophone together.

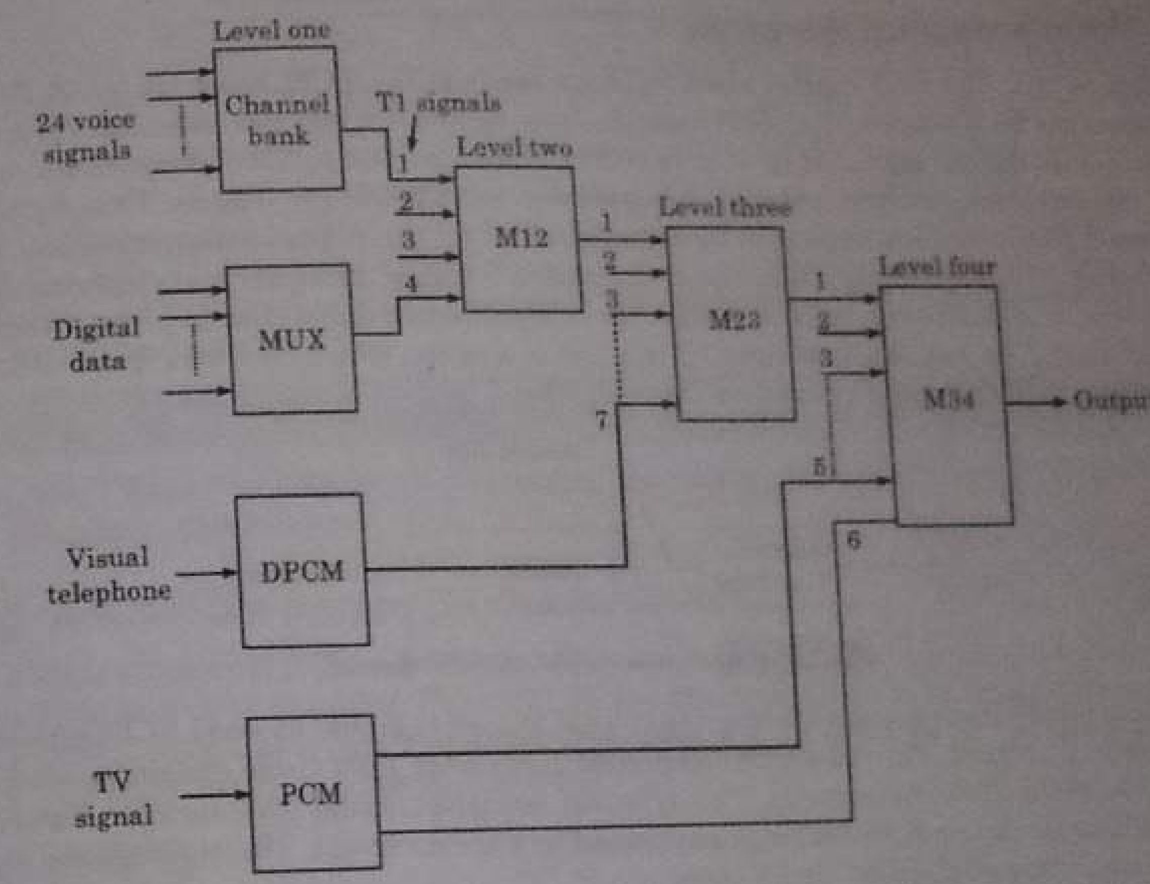


Fig. 11.13. *A digital multiplexer used to multiplex different types of signals.*

Example 11.6. For CCIT hierarchy, assume that the first level multiplexer is a synchronous voice PCM bank with 30 input signals. The output bit rate of this multiplexer is 2.048 Mb/sec. If bit robbing has not been implemented, calculate the number of framing plus signaling bits per frame.

Solution: One frame of transmitted signal is as shown in figure 11.14.

It consists of signal bits for all the 30 input signals plus one frame bit. Let us assume that there are "x" bits per frame.

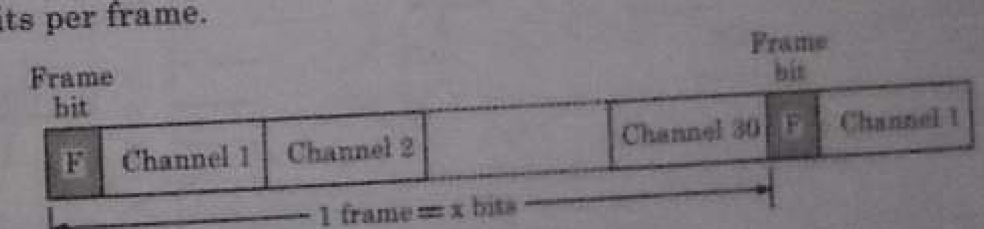


Fig. 11.14. *Frame format for a CCIT digital multiplexer.*

(i) Channels 1 to 30 each are 64 kb/s PCM encoded voice channels. Therefore, the minimum sampling rate required to multiplex them is, ...(i)

$$f_{s(min)} = 64 \times 10^3 \times 2 = 128 \text{ kHz}$$

(ii) Number of frames transmitted per second = $f_{s(min)} = 128 \times 10^3$

(iii) Number of bits per frame = x

(iv) Bit rate = Number of frames/sec × Number of bits/frame = $128 \times 10^3 x$

But it is given that bit rate = 2.048 Mb/sec

Hence, $128 \times 10^3 x = 2.048 \times 10^6$

or $x = 16$

Thus, number of bits per frame = 16. **Ans.**

11.7. North American Hierarchy

As a matter of fact, the first digital signal in true sense is the PCM voice signal. A PCM voice signal represents 64 k bits/sec *i.e.*, 8000 samples per second × 8 bits per samples. Infact, such a signal is called as digital signal at level zero (DS0). Note that due to 8000 samples/sec, sampling rate, the time duration between adjacent samples will be (1/8000) *i.e.*, 125 μs. This digital signal at level zero (DS0) is the fundamental building block of all the digital communication systems. But, practically, the DS0 signal is never transmitted because most of the telephone lines are analog. Hence, in a telephone central office, the subscribers analog line is passed through an antialiasing filter. The bandlimited signal is applied to a codec, which converts it into DS0 signal. The generation of DS0 signal is shown in figure 11.15.

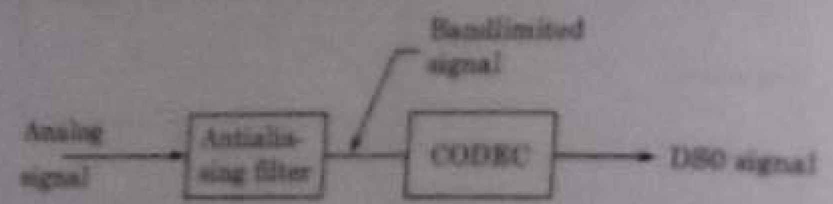


Fig. 11.15. *Generation of DS0 signal.*

Now, 24 such DS0 lines are multiplexed into a *DS*1 (digital signal at level 1). Commonly this signal is called as *T*1 signal. The telephone companies implement TDM (time division multiplexing) through the hierarchy of digital signals. This is called as digital signal (*DS*) service. Figure 11.16 shows the DS hieararchy and the bit rates supported by various levels. The multiplexed signal is converted into a frame at the *DS*1 or *T*1 level.

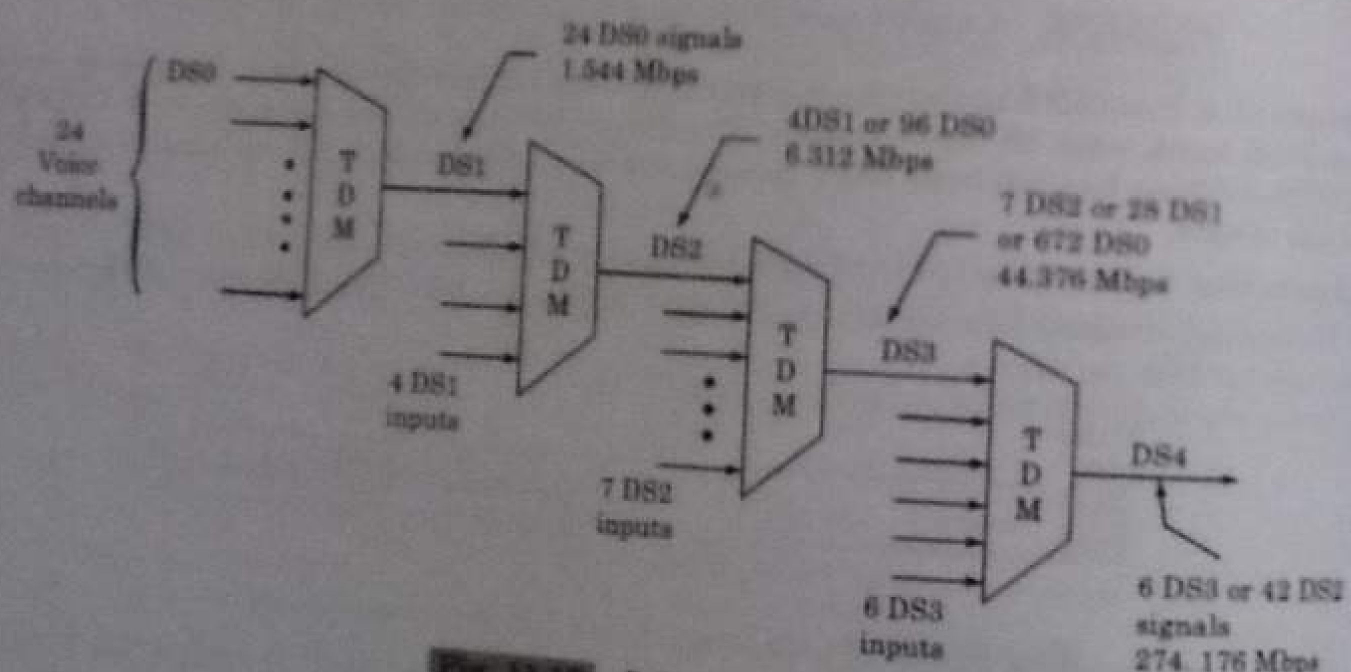


Fig. 11.16. *D.S. hierarchy.*

11.8. T Lines

*DS*0, *DS*1, *DS*2, ... etc. are the names of the services. The telephone companies use the *T* lines ($T-0$, $T-2$) ... etc.) to implement these services.

The *T* lines have capacities which precisely match with the bit rates of the corresponding services as shown in Table 11.3.

Table 11.3. **Relation between DS and T lines**

Service	*Line*	*Rate (Mbps)*	*Number of voice channels*
DS-1	*T*-1	1.544	24
DS-2	*T*-2	6.312	96
DS-3	*T*-3	44.736	672
DS-4	*T*-4	274.176	4032

Thus, *T*-1 line implements *DS*-1 service, *T*-2 implements DS-2 service and so on. *DS*0 is defined as the basic service.

Note: *T lines are digital lines which are designed to conveying digital data, audio or video.*

But, the *T* lines can also be used or analog communication. For example *T*1 line can be used for the telephone applications.

11.9. A PCM-TDM System (T1 Carrier System)

When a large number of PCM signals are to be transmitted over a common channel, multiplexing of these PCM signals is required. Figure 11.17 shows the basic time division multiplexing scheme, called as the ***T*1-digital system** or **T1 carrier system**. This system is used to convey multiple signals over telephone lines using wideband coaxial cable.

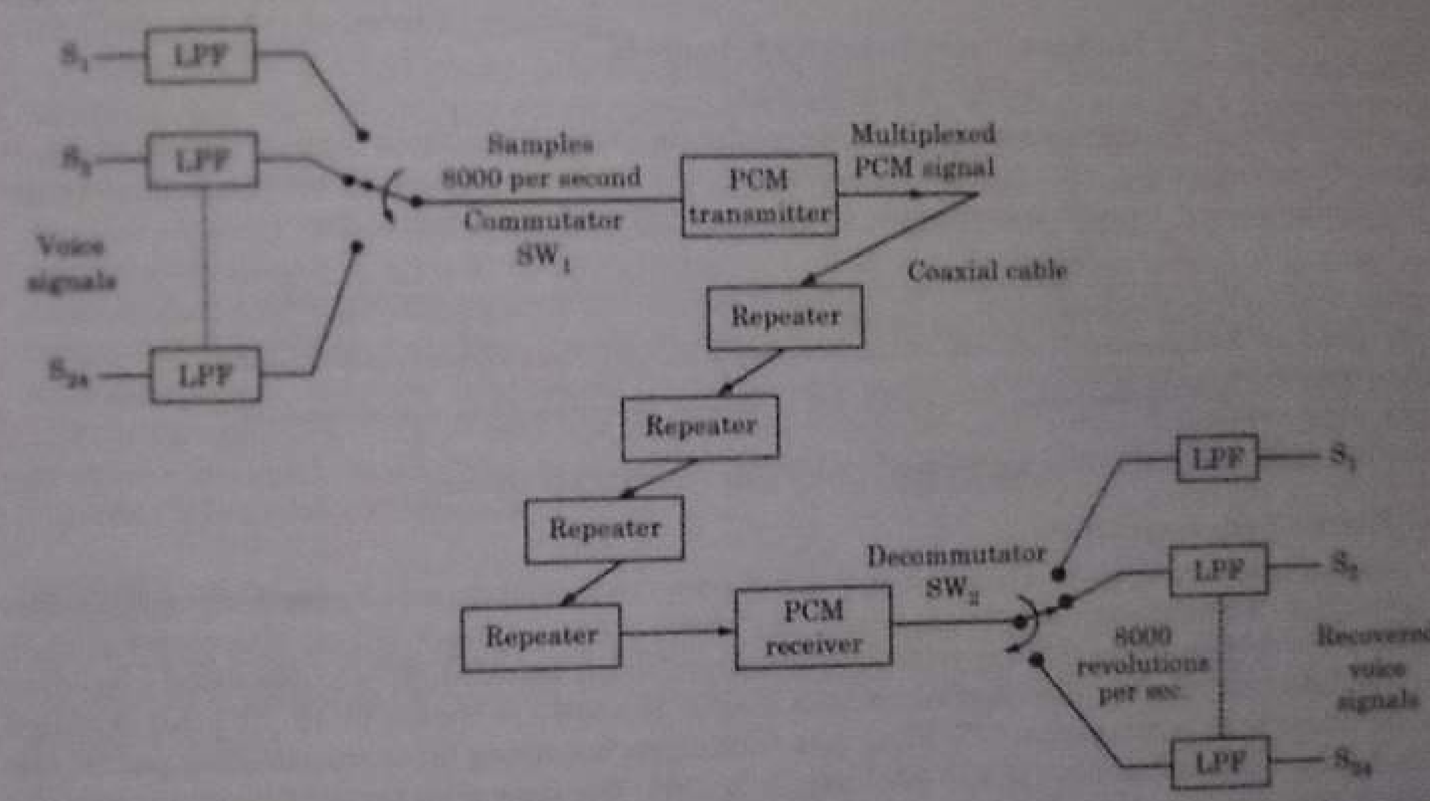


Fig. 11.17. *Block diagram of a basic PCM-TDM system or T1 carrier system.*

Working Operation of the T1 Carrier System

The working operation of the PCM-TDM system shown in figure 11.17 can be explained in the form of few points as under:

(*i*) This system has been designed to accommodate 24 voice channels marked S_1 to S_{24}. Each signal is bandlimited to 3.3 kHz, and the sampling is done at a standard rate of 8 kHz. This is higher than the Nyquist rate. The sampling is done by the commutator switch SW_1.

(ii) These voice signals are selected one by one and connected to a PCM transmitter by the commutator switch SW_1.

(iii) Each sampled signal is then applied to the PCM transmitter which converts it into a digital signal by the process of A to D conversion and companding, as explained earlier.

(iv) The resulting digital waveform is transmitted over a co-axial cable.

(v) Periodically, after every 6000 ft, the PCM-TDM signal is regenerated by amplifiers called "Repeaters". They eliminate the distortion introduced by the channel and remove the superimposed noise and regenerate a clean PCM-TDM signal at their output. This ensures that the received signal is free from the distortions and noise.

(vi) At the destination, the signal is companded, decoded and demultiplexed, using a PCM receiver. The PCM receiver output is connected to different low pass filters via the decommutator switch SW_2.

(vii) Synchronization between the transmitter and receiver commutators SW_1 and SW_2 is essential in order to ensure proper communication.

Now, let us discuss few important terms related to a T1 carrier system as under:

(i) Bits/Frame

The commutators sweep continuously from S_1 to S_{24} and back to S_1 at the rate of 8000 revolutions per second. This will generate 8000 samples per second of each signal (S_1 to S_{24}). Each sample is then encoded (converted) into an eight bit digital word.

Thus, the digital signal generated during one complete sweep (revolution) of the commutator is given by

$$1 \text{ Frame} \equiv 1 \text{ revolution} = 24 \text{ channels}$$
$$= 24 \times 8 \text{ bits} = 192$$

One frame of $T1$ carrier system is shown in figure 11.18. Each voice signal from S_1 to S_{24} is encoded into eight bits. One frame corresponds to the time corresponding transmission of each signal once. Hence 1-frame corresponds to one-revolution of the commutator.

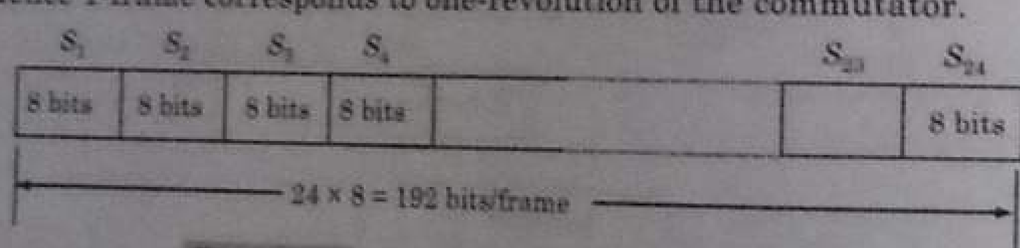


Fig. 11.18. *One frame and bits per frame.*

(ii) Frame Synchronization

As discussed earlier, the synchronization between the transmitter and receiver commutators is essential. Without such synchronization, the receiver cannot know which bits correspond to which of the original signals. To provide such synchronization, an extra bit is transmitted preceding the 192 bits carrying the information in each frame, as shown in figure 11.19. This bit is called as the frame synchronizing bit "F". Thus, one frame synchronizing bit is transmitted per frame. This makes the total number of bits per frame to be 193. The time slots for the 24 signals and the extra frame synchronizing bit is as shown in figure 11.19.

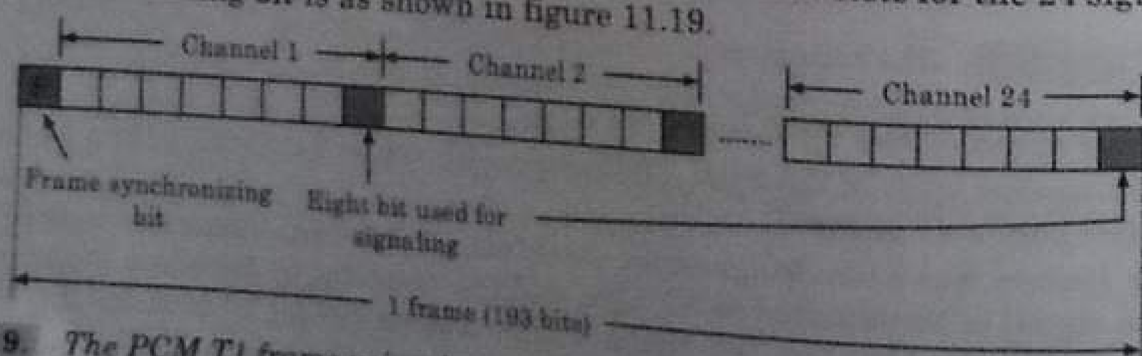


Fig. 11.19. *The PCM T1 frame using frame synchronization and channel associated signaling.*

Further, twelve successive F slots are used to transmit a 12 bit code. The code is 1101 1100 1000. This code is transmitted repeatedly once every 12 frames and it is used at the receiver to obtain synchronization.

(iii) Bit Rate

Bit rate means number of bits transmitted by a system per second.

In the $T1$ carrier system, each signal is sampled 8000 times per second, therefore,

1 frame (1 revolution of commutator) $\equiv 1/8000 = 125\ \mu$ sec.

But, 1 frame consists of 193 bits.

Hence, 193 bits are transmitted in 125 μ sec.

$$\text{Also, number of bits in 1 sec.} = \frac{193}{125 \times 10^{-6}} = 1.544 \times 10^6$$

So, bit rate of $T1$ carrier system = 1.544 Mbits/sec.

(iv) Bandwidth of $T1$ Carrier System

$$\text{Minimum bandwidth } BW = \frac{1}{2} \text{ (bit rate)} = \frac{1}{2} \times (1.544 \times 10^6) = 772 \text{ kHz}$$

Duration of each bit can be found as follows:

Since $\quad 193 \equiv 125\ \mu s$

therefore, $\quad 1 \text{ bit} \equiv (125/193)\ \mu s = 0.6476\ \mu s$

(v) Channel Associated Signaling

When the PCM-TDM system is being used for the telephony, it is expected to transmit certain signaling and supervisory signals alongwith the speech information. The signaling information consists of the signals such as a call is being initiated or a call is being terminated, or the address of calling party etc. In analog system such a signaling information is transmitted over a separate channel other than the voice channel. But in the $T1$ carrier system which is a digital system, a separate channel is not used. In this system, the signaling information is sent using the same data bit slots which are used to send the voice information. The technique used is "bit slot sharing". In the "bit slot sharing" method, for the first five frames, all the 24 channels are encoded into an 8 bit digital code. However in the sixth frame, all the channels are coded into a 7 bit code and the LSB (least significant bit) of each channel is used to transmit the signaling information. This is as shown in figure 11.19. This is called as "**channel associated signaling**". This pattern is repeated after every six frames.

11.10. E Lines

As a matter of fact, Europeans use a version of T lines called E lines. The T lines and E lines are conceptually identical but their capacities and number of voice channels which they can carry will be different.

Table 11.4 shows the E lines and their capacities.

Table 11.4.

S.No.	*Line*	*Rate (Mbps)*	*Number of voice*
1.	$E-1$	2.048	30
2.	$E-2$	8.448	120
3.	$E-3$	34.368	480
4.	$E-4$	139.264	1920

SUMMARY

1. Multiplexing may be defined as a technique which allows many users to share a common communication channel simultaneously. There are two major types of multiplexing techniques. They are as under:
   (i) Frequency division multiplexing (FDM).
   (ii) Time division multiplexing (TDM)
2. The TDM system can be used to multiplex analog or digital signals, however it is more suitable for the digital signal multiplexing.
3. The signaling rate of a TDM system is defined as the number of pulses transmitted per second. It is represented by "r".
4. TDM system is supposed to have its signaling rate as high as possible. It is evident from the expression above that the signaling rate can be increased by increasing the sampling rate f_s and/or the number of input signals N.
5. The minimum transmission bandwidth of a PAM-TDM channel is given by

$$BW = \frac{1}{2}\text{(signaling rate)}$$

6. Therefore, minimum transmission bandwidth $BW \geq \frac{1}{2} \times 2N f_m$
   Hence, minimum transmission bandwidth $BW = Nf_m$.
7. The multiplexed PAM signals can be received properly if and only if the transmitter and receiver commutators are synchronized to each other in terms of the speed and the position. In order to ensure synchronization, a marker pulse is introduced at the end of each frame in the transmitted signals.
8. Crosstalk basically means interference between the adjacent TDM channels it is the unwanted coupling of information from one channel to the other. The guard time T_g is the time spacing introduced between the adjacent TDM channels.
9. The communication channel over which the TDM signal is travelling should ideally have an infinite bandwidth in order to avoid the signal distortion. However, in practice, all the communication channels have a finite bandwidth. Such channels are known as the bandlimited channels.
10. One more cause for the crosstalk between the adjacent TDM signals is the use of bandlimiting filters. Because of these filters, the shapes of the TDM pulses are distorted and they get overlapped and crosstalk will take place.
11. The crosstalk resulting from the pulse overlap can be reduced by introducing guard time of sufficient duration between the adjacent TDM pulses. The guard time is denoted by T_g.
12. Advantages of TDM
    (i) Full available channel bandwidth can be utilized for each channel.
    (ii) Intermodulation distortion is absent.
    (iii) TDM circuitry is not very complex.
    (iv) The problem of crosstalk is not severe.
13. Disadvantages of TDM
    (i) Synchronization is essential for proper operation.
    (ii) Due to slow narrowband fading, all the TDM channels may get wiped out.
14. The digital multiplexing will enable us to combine many digital signals such as computer outputs, digital voice, digitized facsimile and TV signals.
15. The digital data can be multiplexed by using a bit-by-bit interleaving procedure. This can be achieved by using a selector switch which sequentially selects a bit from each input and places it over the high speed transmission line. At the receiving end the bits received on the common line are separated out and delivered to their respective destinations.

16. Digital multiplexing is based on the principle of "interleaving symbols" from two or more dig signals. This is similar to TDM but it does not require the periodic sampling and waveform preservat The signals which are to be multiplexed may come from digital data sources or from analog sou that have been digitally coded.
17. A digital multiplexer is used to merge the input bits from different sources to form one si from transmission via a digital communication system. The multiplexed signal consists of se digits interleaved bit by bit or word by word.
18. The various digital sources that are to be multiplexed will have different bit rates. In practice bit rate variation poses the most serious design problem and leads to the three categori multiplexers, as under:
    (i) Synchronous multiplexers
    (ii) Asynchronous multiplexers
    (iii) Quasi-synchronous multiplexers.
19. The bandwidth efficiency of a multiplexing system is defined as,

$$\text{Bandwidth Efficiency} = \frac{\text{Number of voice signals} \times \text{Bandwidth of each signal}}{\text{Bandwidth of the mutiplexing system}}$$

$$\text{or Bandwidth Efficiency} = \frac{4032 \times 4\text{ kHz}}{137\text{ MHz}} = 11.77\%$$

   Thus bandwidth efficiency of a digital multiplexing system is only 11.77% which is ext poor as compared to a much higher bandwidth efficiency of an analog multiplexing system (t above 85%).
20. Advantages of Digital Multiplexing
    (i) Hardware cost reduction due to the use of digital ICs.
    (ii) Powder cost reduction due to use of regenerative repeaters.
    (iii) More flexibility as compared to the analog multiplexers.
21. DS0, DS1, DS2, ... etc. are the names of the services. The telephone companies use the T – 0, T – 2) ... etc.) to implement these services.
    The T lines have capacities which precisely match with the bit rates of the corresponding
22. When a large number of PCM signals are to be transmitted over a common channel, mul of these PCM signals is required.
23. When the PCM-TDM system is being used for the telephony, it is expected to transmi signaling and supervisory signals alongwith the speech information. The signaling inf consists of the signals such as a call is being initiated or a call is being terminated, or th of calling party etc.

SHORT QUESTIONS WITH ANSWERS

Q.1. What do you mean by Multiplexing?

Ans. Multiplexing may be defined as a technique which allows many users to share a communication channel simultaneously. There are two major types of multiplexing te They are as under:

(i) Frequency division multiplexing (FDM).
(ii) Time division multiplexing (TDM).

Q.2. Explain Frequency Division Multiplexing (FDM).

Ans. This technique permits a fixed frequency band to every user in the complete channel b Such frequency slot is allotted continuously to that user. As an example consider that th bandwidth is 1 MHz. Let there be ten users, each requiring upto 100 kHz bandwidth complete channel bandwidth of 1 MHz can be divided into ten frequency bands, i.e. each of [illegible] kHz and every user can be allotted one independent frequency band. This technique is known as **Frequency Division Multiplexing** (FDM).

Q.3. What is Transmission Bandwidth of a PAM/TDM Channel?

Ans. The minimum transmission bandwidth of a PAM-TDM channel is given by

$$BW = \frac{1}{2}\,(\text{signaling rate})$$

Therefore, minimum transmission bandwidth $BW \geq \frac{1}{2} \times 2N f_m$

Hence, minimum transmission bandwidth $BW = N f_m$

Q.4. What is Crosstalk in PAM/TDM System?

Ans. Crosstalk basically means interference between the adjacent TDM channels. It is [illegible] the unwanted coupling of information from one channel to the other. The guard time T_g is the time spacing introduced between the adjacent TDM channels.

Q.5. Write the Advantages of TDM.

Ans. (i) Full available channel bandwidth can be utilized for each channel.

(ii) Intermodulation distortion is absent.

(iii) TDM circuitry is not very complex.

(iv) The problem of crosstalk is not severe.

Q.6. Write the Disadvantages of TDM.

Ans. (i) Synchronization is essential for proper operation.

(ii) Due to slow narrowband fading, all the TDM channels may get wiped out.

Q.7. Explain the Principle of Digital Multiplexing.

Ans. Digital multiplexing is based on the principle of "interleaving symbols" from two or more digital signals. This is similar to TDM but it does not require the periodic sampling and waveform preservation. The signals which are to be multiplexed may come from digital data sources or from analog sources that have been digitally coded.

A digital multiplexer is used to merge the input bits from different sources to form one signal for transmission via a digital communication system. The multiplexed signal consists of source digits interleaved bit by bit or word by word. The important functions that must be performed by a multiplexer are as under:

(i) To establish a frame. A frame consists of at least one bit from every input.

(ii) A number of unique bits slots within the frame should be assigned to each input.

(iii) To insert control bits for frame identification and synchronization.

(iv) To make allowance for any variations of the input bit rates.

Q.8. How can you classify Digital Multiplexers?

Ans. The various digital sources that are to be multiplexed will have different bit rates. In practice, the bit rate variation poses the most serious design problem and leads to the three categories of multiplexers, as under:

(i) Synchronous multiplexers

(ii) Asynchronous multiplexers

(iii) Quasi-synchronous multiplexers.

Q.9. Write the Advantages of Digital Multiplexing.

Ans. (i) Hardware cost reduction due to the use of digital ICs.

(ii) Powder cost reduction due to use of regenerative repeaters.

(iii) More flexibility as compared to the analog [illegible]

Q.10. What is Channel Associated Signaling? Explain.

Ans. When the PCM-TDM system is being used for the telephony, it is expected to transmit certain signaling and supervisory signals alongwith the speech information. The signaling information consists of the signals such as a call is being initiated or a call is being terminated, or the address of calling party etc. In analog system such a signaling information is transmitted over a separate channel other than the voice channel. But in the T1 carrier system which is a digital system, a separate channel is not used. In this system, the signaling information is sent using the same data bit slots which are used to send the voice information. The technique used is "bit slot sharing". In the "bit slot sharing" method, for the first five frames, all the 24 channels are encoded into an 8 bit digital code. However in the sixth frame, all the channels are coded into a 7 bit code and the LSB (least significant bit) of each channel is used to transmit the signaling information. This is called as "**channel associated signaling**". This pattern is repeated after every six frames.

REVIEW QUESTIONS

1. What do you mean by multiplexing? Explain TDM and FDM.
2. Explain a PAM/TDM system in detail with a block diagram and write briefly about the following:
   (i) frame
   (ii) signaling rate
   (iii) transmission BW
   (iv) synchronization
   (v) crosstalk
   (vi) guard time
3. Explain the principle of digital multiplexing. What are the types of digital multiplexers?
4. Explain multiplexing hierarchy for digital communication.

□□□

Hence, in view of above discussion, the received signal $x(t)$ is expressed as

$$x(t) = s_i(t) + w(t), \begin{cases} 0 \le t \le T \\ i = 1, 2, \ldots M \end{cases}$$

Q.2. Explain the concept of Optimum Receiver.

Ans. the receiver observes the received signal $x(t)$ for a duration of T seconds and makes a be estimate of the transmitted signal $s_i(t)$ or equivalently the estimate of symbols m_i. But, due to the presence of channel noise, this decision-making process is statistical in nature. As a result of this he receiver is likely to make occasional errors.

Therefore, the requirement is to design the receiver so as to minimize the average probability f symbol error.

This average probability of symbol error may be defined as

$$P_e = \sum_{i=1}^{M} p_i \, P(\hat{m} \neq m_i \mid m_i)$$

where m_i = transmitted symbol.

$\hat{m}$ = estimate produced by the receiver, and

$P(\hat{m} \neq m_i \mid m_i)$ = the conditional error probability given that the ith symbol was sent.

3. What is Geometric Representation of Signals? Explain.

ns. In geometric representation of signals, we represent any set of M energy signals $\{s_i(t)\}$ as line mbinations of N orthonormal basis functions, where $N \le M$.

This means that given a set of real-valued signals $s_1(t)$, $s_2(t)$, ..., $s_M(t)$, each of duration T onds, we may write $s_i(t)$ as under:

$$s_i(t) = \sum_{j=1}^{N} s_{ij} \phi_j(t), \begin{cases} 0 \le t \le T \\ i = 1, 2, \ldots M \end{cases}$$

ere, the coefficients of the expansion can be defined as

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) \, dt, \begin{cases} i = 1, 2, \ldots M \\ j = 1, 2, \ldots N \end{cases}$$

Now, the real-valued basis functions $\phi_1(t)$, $\phi_2(t)$, ..., $\phi_N(t)$ are orthonormal. Here, the wo honormal' implies that

$$\int_0^T \phi_i(t) \phi_j(t) \, dt = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

re δ_{ij} is the Kronecker delta.

REVIEW QUESTIONS

Draw the block diagram of a most basic form of digital communication system and write the xpression for probable that symbol m_i is emitted by an information source.

xplain the concept of AWGN channel.

hat is the concept of an optimum receiver? Explain.

xplain the geometric representation of signals.

hat is Gram-Schmidt orthogonalization procedure? Explain.

□□□

CHAPTER 14

Digital Modulation Techniques

Inside this Chapter

14.1. Introduction

As discussed earlier, Modulation is defined as the process by which some characteristics of a carrier is varied in accordance with a modulating signal. In digital communications, the modulating signal consists of binary data or an M-ary encoded version of it. This data is used to modulate a carrier wave (usually sinusoidal) with fixed frequency. In fact, the input data may represent the digital computer outputs or PCM waves generated by digitizing voice or video signals. The channel may be a telephone channel, microwave radio link, satellite channel or an optical fiber. In digital communication, the modulation process involves switching or keying the amplitude, frequency or phase of the carrier in accordance with the input data.

Thus, there are three basic modulation techniques for the transmission of digital data. They are known as amplitude-shift keying (ASK), frequency shift keying (FSK) and phase-shift keying (PSK) which can be viewed as special cases of amplitude modulation frequency modulation and phase modulation respectively.

The present chapter is devoted to detailed discussion of digital modulation techniques, their noise performance, spectral properties, their merits and limitations, applications and other related aspects.

14.2. Digital Modulation Formats

When we have to transmit a digital signal over a long distance, we need continuous-wave (CW) modulation. For this purpose, the transmission medium can be in form of radio, cable or other type of channel. Also, a carrier signal having some frequency f_c is used for modulation. Then the modulating digital signal modulates some prameter like frequency, phase or amplitude of the carrier. Due to this process, there is some deviation in carrier frequency f_c. This deviation is known as the bandwidth of the channel. This means that the channel has to transmit some range or band of frequencies. Such type of transmission is known as **bandpass transmission** and the communication channel is known as **bandpass channel.**

Here, the word bandpass is used since the range of frequencies does not start from zero Hz to f_m Hz. In fact, the range of frequencies from zero Hz to f_m Hz is known as **low-pass signal** and such channel is known as **low-pass channel.**

Now, when it is required to transmit digital signals on a bandpass channel, the amplitude, frequency or phase of the sinusoidal carrier is varied in accordance with the incoming digital data. Since the digital data is in discrete steps, the modulation of the bandpass sinusoidal carrier is also done in discrete steps. Due to this reason, this type of modulation (*i.e.*, Digital modulation) is also known as switching or signaling. Now, if an amplitude of the carrier is switched depending on the input digital signal, then it is called Amplitude shift keying (ASK).

This process is quite similar to analog amplitude modulation. If the frequency of the sinusoidal carrier is switched depending upon the input digital signal, then it is known as the frequency shift keying (FSK). This is very much similar to the analog frequency modulation. If the phase of the carrier is switched depending upon the input digital signal, then it is called phase shift keying (PSK). This is similar to phase modulation. Since the phase and frequency modulation has constant amplitude envelope, therefore FSK and PSK also has a constant amplitude envelope. Because of constant amplitude of FSK and PSK, the effect of non-linearities, noise interference is minimum on signal detection. However, these effects are more pronounced on ASK. Therefore, FSK and PSK are preferred over ASK.

Figure 14.1 shows the waveforms for amplitude-shift keying, phase-shift keying and frequency shift keying. In these waveforms, a single feature of the carrier (*i.e.*, amplitude, phase or frequency) undergoes modulation.

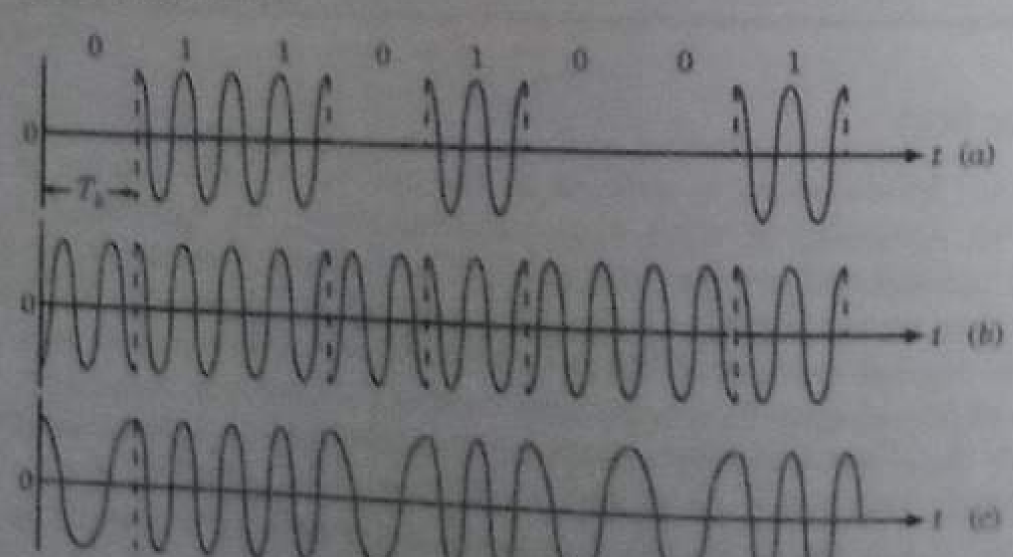


Fig. 14.1. *The three basic forms of signaling binary information. (a) Amplitude-shift keying (b) Phase-shift keying. (c) Frequency shift keying with continuous phase*

In digital modulations, instead of transmitting one bit at a time, we transmit two or more bits simultaneously. This is known as M-ary transmission. This type of transmission results in reduced channel bandwidth. However, sometimes, we use two quadrature carriers for modulation. This process is known as **Quadrature modulation.**

Thus, we see that there are a number of modulation schemes available to the designer of a digital communication system required for data transmission over a bandpass channel.

Every scheme offers system trade-offs of its own. However, the final choice made by the designer is determined by the way in which the available primary communication resources such as transmitted power and channel bandwidth are best exploited. In particular, the choice is made in favour of a scheme which possesses as many of the following design characteristics as possible:

(*i*) Maximum data rate.
(*ii*) Minimum probability of symbol error.
(*iii*) Minimum transmitted power.
(*iv*) Maximum channel bandwidth.
(*v*) Maximum resistance to interfering signals.
(*vi*) Minimum circuit complexity.

14.3. Types of Digital Modulation Techniques

(U.P. Tech., Sem. Examination, 2003-2004)

Basically, digital modulation techniques may be classified into coherent or non-coherent techniques, depending on whether the receiver is equipped with a phase-recovery circuit or not. The phase-recovery circuit ensures that the oscillator supplying the locally generated carrier wave receiver is synchronized* to the oscillator supplying the carrier wave used to originally modulate the incoming data stream in the transmitter.

(*i*) Coherent Digital Modulation Techniques

Coherent digital modulation techniques are those techniques which employ coherent detection. In coherent detection, the local carrier generated at the receiver is phase locked with the carrier at the transmitter. Thus, the detection is done by correlating received noisy signal and locally generated carrier. The coherent detection is a synchronous detection.

(*ii*) Non-coherent Digital Modulation Techniques

Non-coherent digital modulation techniques are those techniques in which the detection process does not need receiver carrier to be phase locked with transmitter carrier.

The advantage of such type of system is that the system becomes simple. But the drawback of such a system is that the error probability increases.

In fact, the different digital modulation techniques are used for various specific application areas.

14.4. Coherent Binary Modulation Techniques

As mentioned earlier, the binary (*i.e.*, digital) modulation has three basic forms amplitude shift keying (ASK), phase-shift keying (PSK) and frequency-shift keying (FSK). In this section let us discues different coherent binary modulation techniques.

14.5. Coherent Binary Amplitude Shift Keying or On-Off Keying

Amplitude shift keying (ASK) or ON-OFF keying (OOK) is the simplest digital modulation technique. In this method, there is only one unit energy carrier and it is switched on or off depending upon the input binary sequence. The ASK waveform may be represented as,

$$s(t) = \sqrt{2P_s}\cos(2\pi f_c t) \quad \text{(To transmit '1')} \qquad ...(14.1)$$

* In both frequency and phase.

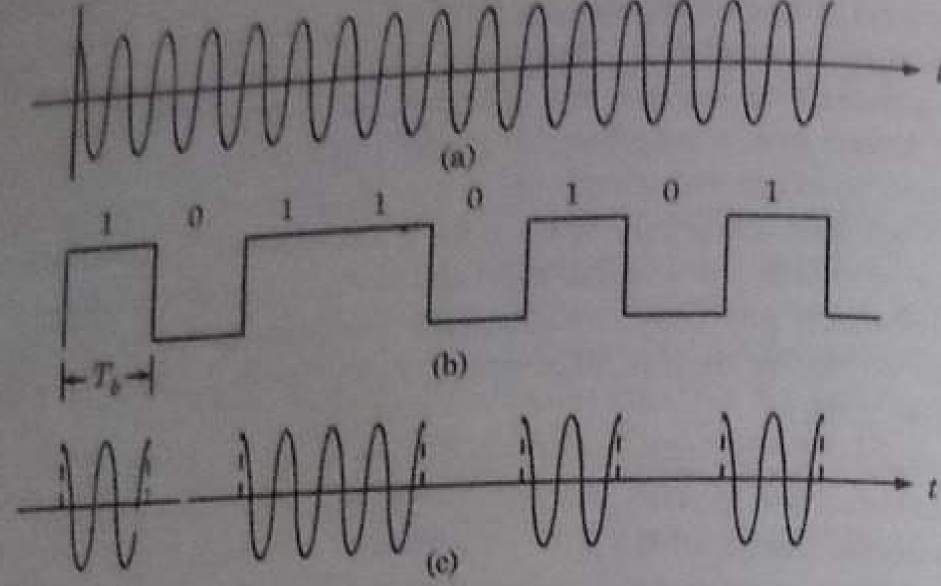


Fig. 14.2. *Amplitude-shift keying waveforms, (a) Unmodulated carrier, (b) Unipolar bit sequence, (c) ASK waveform.*

To transmit symbol '0', the signal $s(t) = 0$ *i.e.*, no signal is transmitted. Signal $s(t)$ contai some complete cycles of carrier frequency 'f_c'.

Hence, the ASK waveform looks like an ON-OFF of the signal. Therefore, it is als known as the ON-OFF keying (OOK). Figure 14.2 shows the ASK waveform.

14.5.1. Signal Space Diagram of ASK

The ASK waveform of equation (14.1) for symbol '1' can be represented as,

$$s(t) = \sqrt{P_s T_b} \cdot \sqrt{2/T_b} \cos(2\pi f_c t)$$

or

$$s(t) = \sqrt{P_s T_b}\, \phi_1(t) \quad ...(14.$$

This means that there is only one carrier function $\phi_1(t)$. The signal space diagram wi have two points on $\phi_1(t)$. One will be at zero and other will be at $\sqrt{P_s T_b}$. Figure 14.3 show this aspect.

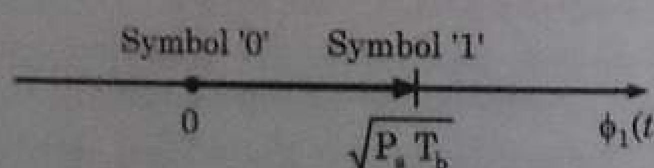


Fig. 14.3. *Signal space diagram of ASK.*

Thus, the distance between the two signal points is,

$$d = \sqrt{P_s T_b} = \sqrt{E_b} \quad ...(14.$$

14.5.2. Generation of ASK Signal

ASK signal may be generated by simply applying the incoming binary data (represent in unipolar form) and the sinusoidal carrier to the two inputs of a product modulator (i.e balanced modulator). The resulting output will be the ASK waveform. This is shown in figure 14.4. Modulation causes a shift of the baseband signal spectrum. The ASK signal which is basically the product of the binary sequence and the carrier signal, has a power spectral density (PSD) same as that of the baseband on-off signal but shifted in the frequency domain by $\pm f_c$. This is shown in figure 14.5. It may be noted that two impulses occur at $\pm f_c$. The spectrum of the ASK signal shows that it has an infinite bandwidth. However for practical purpose, the bandwidth is often defined as the bandwidth of an ideal bandpass filter centered at f_c whose output contains about 95% of the total average power content of the ASK signal. It may be proved that according to this criterion the bandwidth of the ASK signal

is approximately $3/T_b$ Hz. The bandwidth of the ASK signal can however, be reduced by usi smoothed versions of the pulse waveform instead of rectangular pulse waveforms.

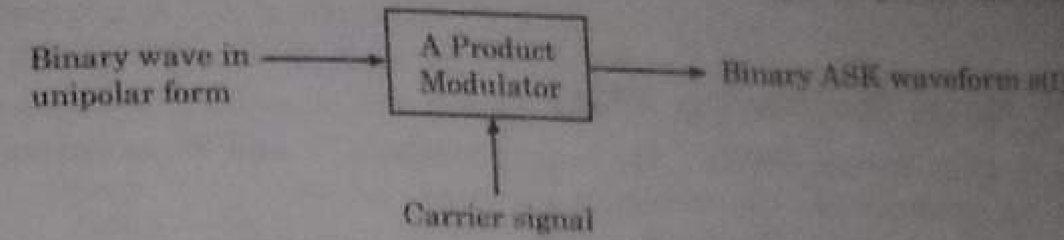


Fig. 14.4. *Generation of binary ASK waveform.*

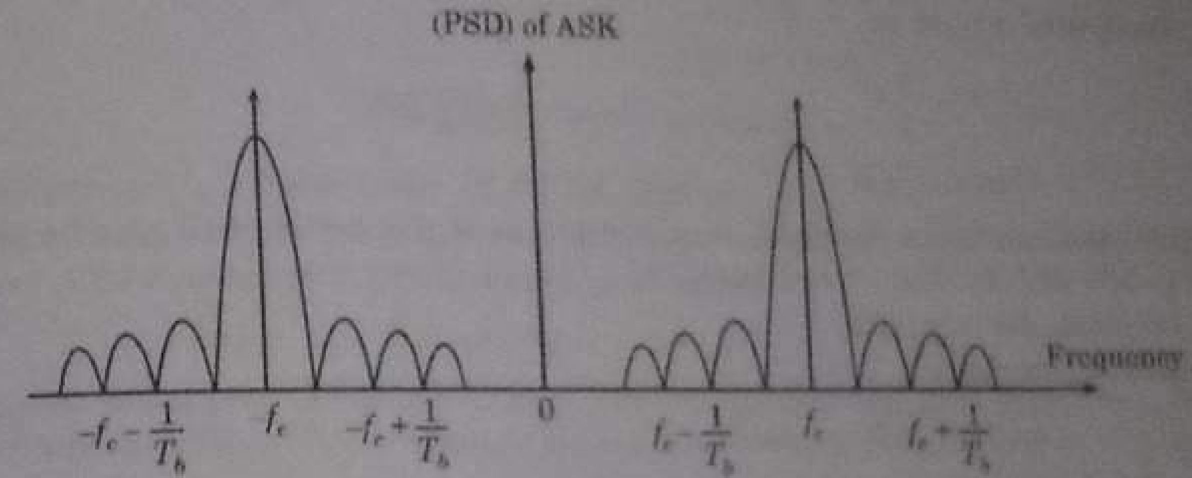


Fig. 14.5. *Power spectral density of ASK signal.*

14.5.3. Coherent Demodulation of Binary ASK

The demodulation of binary ASK waveform can be achieved with the help of *coherent de* as shown in figure 14.6. It consists of a product modulator which is followed by an inte and a decision-making device. The incoming ASK signal is applied to one input of the p modulator. The other input of the product modulator is supplied with a sinusoidal c which is generated with the help of a local oscillator. The output of the product mod goes to input of the integrator. The integrator operates on the output of the multipl successive bit intervals and essentially performs a low-pass filtering action. The out the integrator goes to the input of a decision-making device.

Now, the decision-making device compares the output of the integrator with a threshold. It makes a decision in favour of symbol 1 when the threshold is exceeded favour of symbol 0 otherwise. The *coherent detection* makes the use of linear operat this method we have assumed that the local carrier is in perfect synchronisation wi carriers used in the transmitter. This means that the frequency and phase of the generated carrier is same as those of the carriers used in the transmitter.

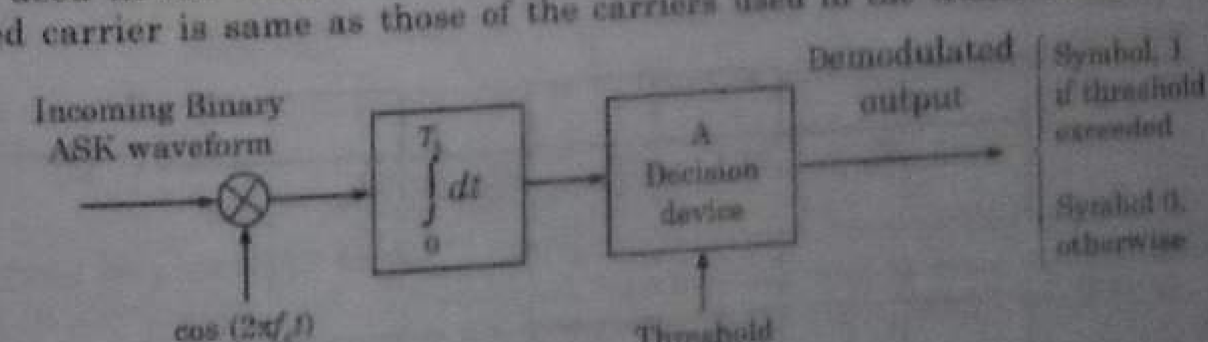


Fig. 14.6. *Coherent detection of binary ASK signals.*

The following two forms of synchronisation are required for the operation of co (or synchronous detector):

(i) *Phase synchronisation which ensures that carrier wave generated locally in the is locked in phase with respect to one that is employed in the transmitter.*

(ii) Timing synchronization which enable proper timing of the decision making opera- the receiver with respect to switching instants (switching between 1 and 0) in the origi- binary data.

14.6. Binary Phase Shift Keying (BPSK)

In a binary phase shift keying (BPSK), the binary symbols '1' and '0' modulate the phase of the carrier. Let us assume that the carrier is given as,

$$s(t) = A \cos (2\pi f_c t)$$

Here 'A' represents peak value of sinusoidal carrier. For the standard 1 Ω load resistor, the power dissipated would be,

$$P = \frac{1}{2}A^2$$

or

$$A = \sqrt{2P}$$

Now, when the symbol is changed, then the phase of the carrier will also be changed by an amount of 180 degrees (i.e., π radians).

Let us consider, for example,

For symbol '1', we have

$$s_1(t) = \sqrt{2P} \cos(2\pi f_c t)$$

If next symbol is '0', then we have

For symbol '0', we have

$$s_2(t) = \sqrt{2P} \cos (2\pi f_c t + \pi)$$

Now, because cos (θ + π) = – cos θ, therefore, the last equation can be written as

$$s_2(t) = -\sqrt{2P} \cos (2\pi f_c t)$$

With the above equation, we can define BPSK signal combinely as,

$$s(t) = b(t) \sqrt{2P} \cos (2\pi f_c t)$$

where $b(t)$ = + 1 when binary '1' is to be transmitted

– 1 when binary '0' is to be transmitted

Figure 14.7 illustrates binary signal and its equivalent signal $b(t)$.

Note: *It may be observed from figure 14.7(b) that the signal b(t) is a NRZ bipolar signal. In fact, this signal directly modulates the carrier signal cos (2π f_c t).*

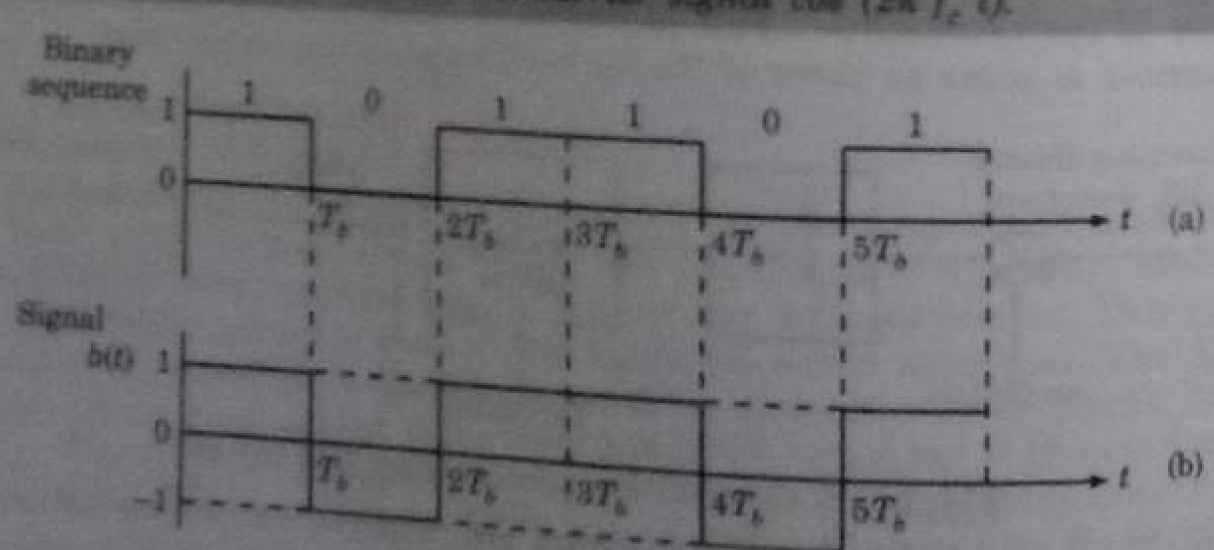


Fig. 14.7. *(a) Binary sequence, (b) The corresponding bipolar signal b(t).*

14.6.1. Generation of BPSK Signal

The BPSK signal may be generated by applying carrier signal to a balanced modulator. Here, the bipolar signal $b(t)$ is applied as a modulating signal to the balanced modulator. Figure 14.8 shows the block diagram of a BPSK signal generator.

A NRZ level encoder converts the binary data sequence into bipolar NRZ signal.

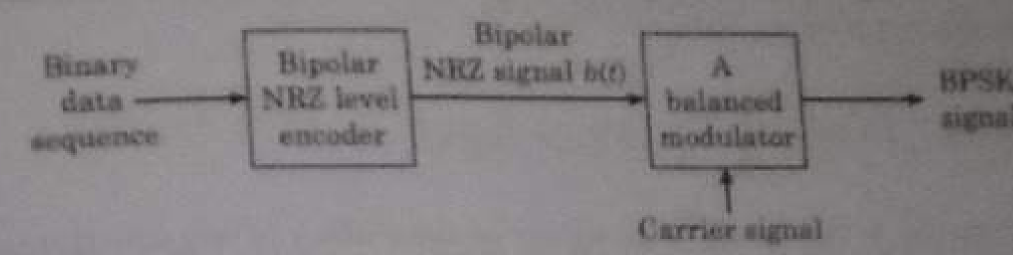


Fig. 14.8. *Generation of BPSK.*

14.6.2. Reception (i.e. Detection) of BPSK Signal

Figure 14.9 shows the block diagram of the scheme to recover baseband signal from BPSK signal. The transmitted BPSK signal is given as

$$s(t) = b(t) \sqrt{2P} \cos (2\pi f_c t)$$

This signal undergoes the phase change depending upon the time delay from transmitter end to receiver end. This phase change is, usually, a fixed phase shift in the transmitted signal.

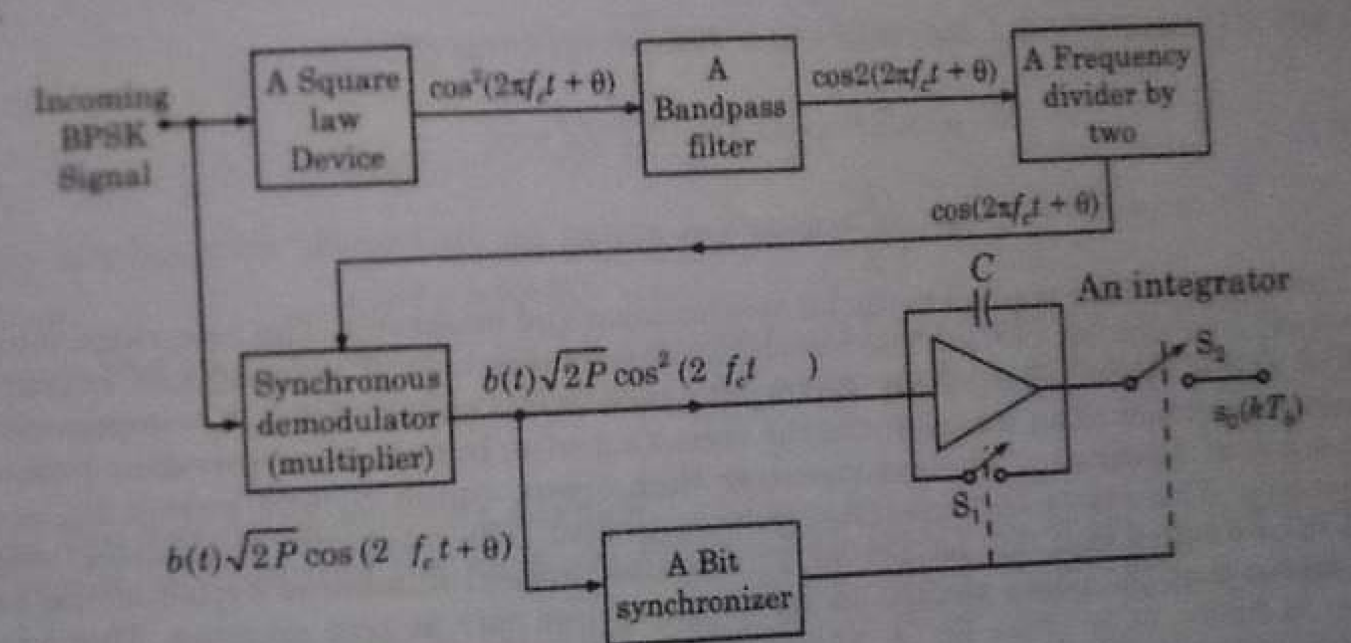


Fig. 14.9. *Reception of baseband signal in BPSK signal.*

Let us consider that this phase shift is θ. Because of this, the signal at the input of the receiver can be written as ...(14.10)

$$s(t) = b(t) \sqrt{2P} \cos (2\pi f_c t + \theta)$$

Now, from this received signal, a carrier is separated because this is coherent detection. As shown in the figure 14.9, the received signal is allowed to pass through a square law device. At the output of the square law device, we get a signal which is given as

$$\cos^2 (2\pi f_c t + \theta)$$

Here, it may be noted that we have neglected the amplitude, since we are only interested in the carrier of the signal

Again, we know that

$$\cos^2\theta = \frac{1+\cos 2\theta}{2}$$

Therefore, we have

$$\cos^2(2\pi f_c t + \theta) = \frac{1+\cos 2(2\pi f_c t+\theta)}{2}$$

$$\cos^2(2\pi f_c t + \theta) = \frac{1}{2}+\frac{1}{2}\cos 2(2\pi f_c t+\theta)$$

Here, $\frac{1}{2}$ represents a DC level. This signal is then allowed to pass through a band... filter (BPF) whose passband is centred around $2f_c$. Bandpass filter removes the DC level... and at the output, we obtain,

$$\cos 2(2\pi f_c t + \theta)$$

This signal is having frequency equal to $2f_c$. Hence, it is passed through a frequ... divider by two. Thus, at the output of frequency divider, we get a carrier signal wh... frequency is f_c i.e., $\cos(2\pi f_c t + \theta)$.

The synchronous (*i.e.*, coherent) demodulator multiplies the input signal and the recov... carrier. Hence, at the output of multiplier, we get

$$b(t)\sqrt{2P}\cos(2\pi f_c t+\theta)\times\cos(2\pi f_c t+\theta) = b(t)\sqrt{2P}\cos^2(2\pi f_c t+\theta)$$

$$= b(t)\sqrt{2P}\times\frac{1}{2}[1+\cos 2(2\pi f_c t+\theta)]$$

$$= b(t)\sqrt{\frac{P}{2}}[1+\cos 2(2\pi f_c t+\theta)] \quad ...(14.1...$$

This signal is then applied to the bit synchronizer and integrator. The integrator integra... the signal over one bit period. The bit synchronized takes care of starting and ending ti... of a bit. At the end of bit duration T_b, the bit synchronizer closes switch S_2 temporarily. Th... connects the output of an integrator to the decision device. In fact, it is equivalent to sampl... the output of integrator. The synchronizer then opens switch S_2 and switch S_1 is clos... temperorily. This resets the integrator voltage to zero. The integrator then integrates ne... bit. Let us assume that one bit period 'T_b' contains integral number of cycles of the carr... This means that the phase change occurs in the carrier only at zero crossing. This has be... shown in figure 14.10. This BPSK waveform has full cycles of sinusoidal carrier.

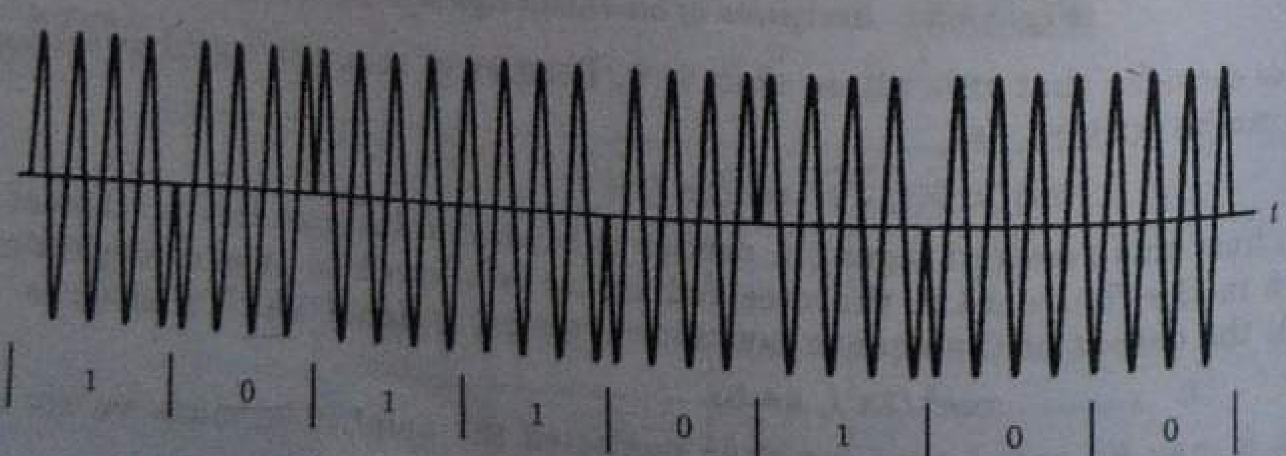


Fig. 14.10. *The BPSK waveform.*

Also, in the k^{th} bit interval, we can write output signal as under:

$$s_0(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}\int_{(k-1)T_b}^{kT_b}[1+\cos(2\pi f_c t+\theta)]\,dt$$

This equation gives the output of an interval for k^{th} bit. Hence, integration is performed from $(k-1)T_b$ to kT_b. Here, T_b is the one bit period. We can write the above equation as under :

$$s_0(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}\left[\int_{(k-1)T_b}^{kT_b}1dt + \int_{(k-1)T_b}^{kT_b}\cos(2\pi f_c t+\theta)dt\right]$$

where $\int_{(k-1)T_b}^{kT_b}\cos 2(2\pi f_c t+\theta)\,dt = 0$, since average value of sinusoidal waveform is zero if integration is done over full cycles. Hence, we can write above equation as,

$$s_0(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}\int_{(k-1)T_b}^{kT_b}1dt$$

or

$$s_0(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}[t]_{(k-1)T_b}^{kT_b}$$

or

$$s_0(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}\{kT_b-(k-1)T_b\}$$

or

$$s_0(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}T_b \quad ...(14.12)$$

The last eqaution shows that the output of the receiver depends on input.

Thus, $\quad s_0(kT_b) \propto b(kT_b)$

Depending upon the value of $b(kT_b)$, the output $s_0(kT_b)$ is generated in receiver.

This signal is then applied to a decision device which decides whether transmitted symbol was zero or one.

14.6.3. The Spectrum of BPSK Signals

Type of we know that the waveform $b(t)$ is a NRZ binary waveform. In this waveform, there are rectangular pulses of amplitude $\pm V_b$. If we assume that each pulse is $\pm\frac{T_b}{2}$ around its centre, then it becomes easy to find Furrier transform of such pulse. The Fourier transform of this type of pulse is given as,

$$X(f) = V_b T_b \frac{\sin(\pi f T_b)}{(\pi f T_b)}\,* \quad ...(14.13)$$

For a large number of such positive and negative pulses, the power spectral density $S(f)$ is expressed as

$$S(f) = \frac{\overline{|X(f)|^2}}{T_s} \quad ...(14.14)$$

Here, $\overline{X(f)}$ denotes average value of $X(f)$ due to all the pulses in $b(t)$. And T_s is symbol duration. Substituting value of $X(f)$ from equation (13.13) in equation (14.14), we get

* Making use of standard relation.

$$S(f) = \frac{V_b^2 T_b^2}{T_s}\left[\frac{\sin(\pi f T_b)}{\pi f T_b}\right]^2$$

For BPSK, because only one bit is transmitted at a time, therefore, symbol and ... durations are same i.e., $T_b = T_s$. Then the last equation becomes,

$$S(f) = V_b^2 T_b\left[\frac{\sin(\pi f T_b)}{\pi f T_b}\right]^2 \quad ...(14...$$

This equation gives the power spectral density (psd) of baseband signal $b(t)$. The BPSK signal is generated by modulating a carrier by the baseband signal $b(t)$. Due to modulation of the carrier of frequency f_c, the spectral components are translated from f to $f_c + f$ and $f_c -$ ... The magnitude of these components is divided by half.

Therefore, from equation (14.15) we can write the power spectral density of BPSK signal as under:

$$S_{BPSK}(f) = V_b^2 T_b\left\{\frac{1}{2}\left[\frac{\sin\pi(f_c - f)T_b}{\pi(f_c - f)T_b}\right] + \frac{1}{2}\left[\frac{\sin\pi(f_c + f)T_b}{\pi(f_c + f)T_b}\right]^2\right\}$$

It may be noted that this equation consists of two half magnitude spectral components of same frequency 'f' above and below f_c. Let us assume that the value of $\pm V_b = \pm\sqrt{P}$. This means that the NRZ signal is having amplitudes of $+\sqrt{P}$ and $-\sqrt{P}$. Then the last equation becomes,

$$S_{BPSK}(f) = \frac{PT_b}{2}\left\{\left[\frac{\sin\pi(f - f_c)T_b}{\pi(f - f_c)T_b}\right]^2 + \frac{1}{2}\left[\frac{\sin\pi(f_c + f)T_b}{\pi(f_c + f)T_b}\right]^2\right\} \quad ...(14.16)$$

This equation gives power spectral density (psd) of BPSK signal for modulating signal $b(t)$ having amplitudes equal to $\pm\sqrt{P}$.

Further, we know that the modulated signal is given as

$$s(t) = \pm\sqrt{2P}\cos(2\pi f_c t) \quad [\because A = \sqrt{2P}]$$

If $b(t) = \pm\sqrt{P}$, then the carrier becomes,

$$\phi(t) = \sqrt{2}\cos(2\pi f_c t) \quad ...(14.17)$$

Equation (14.15) describes power spectral density (psd) of the NRZ waveform. For one rectangular pulse, the shape of $S(f)$ will be a sinc pulse as shown in figure 14.11.

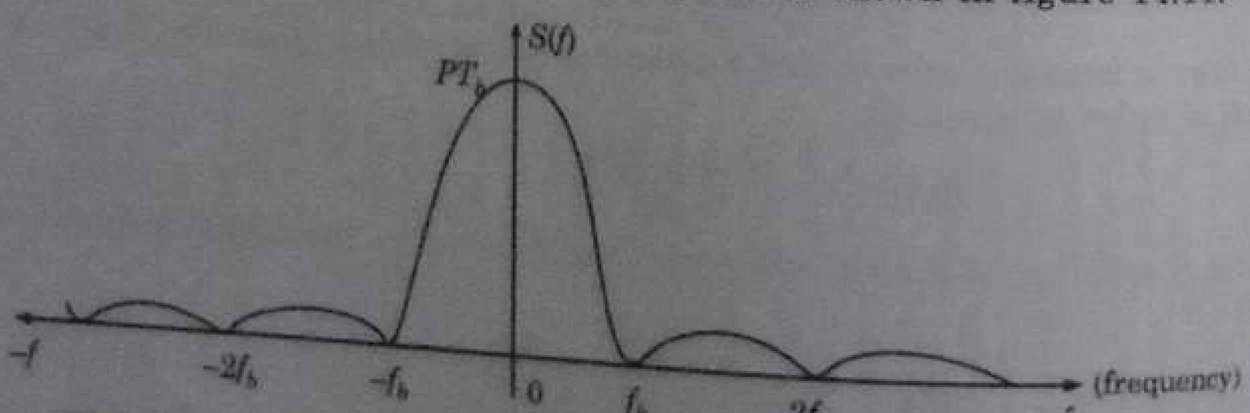


Fig. 14.11. *Plot of power spectral density (psd) of NRZ baseband signal.*

It may be observed from this figure that the main lobe ranges from $-f_b$ to $+f_b$.* Because we have taken $\pm V_b = \pm\sqrt{P}$ in equation (14.15), therefore, the peak value of the main lobe is PT_b. Now let us consider the power spectral density (psd) of BPSK signal expressed by equation (14.16).

Figure 14.12 shows the plot of this equation. This figure, thus, clearly shows that there are two lobes, one at f_c and other at $-f_c$. The same spectrum of figure 14.11 has been placed at $+f_c$ and $-f_c$. However, the amplitudes of main lobes are $\frac{PT_b}{2}$ in figure 14.12.

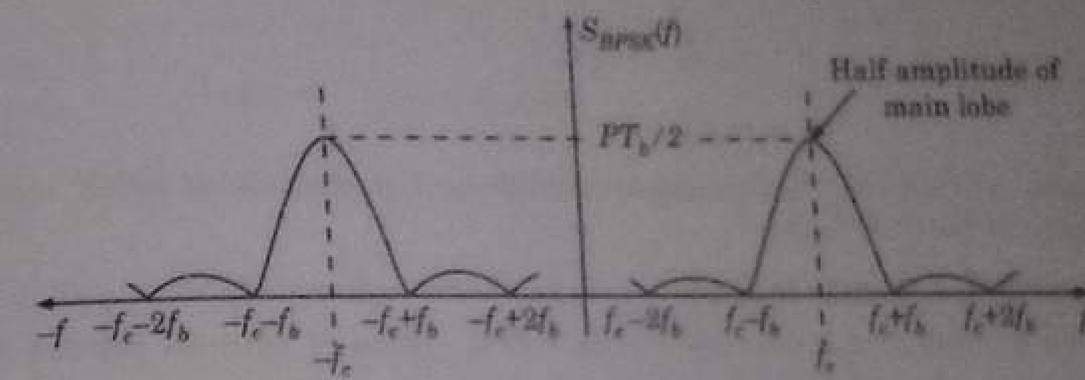


Fig. 14.12. *Plot of power spectral density of BPSK signal.*

Hence, they are reduced to half. The spectrum of $S(f)$ as well as $S_{BPSK}(f)$ extends overall the frequencies.

14.6.4. A Geometrical Representation for BPSK Signals

We know that BPSK signal carries the information about two symbols. These symbols are symbol '1' and symbol '0'. We can represent BPSK signal geometrically to show those two symbols. From equation (14.9), we know that BPSK signal is expressed as,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_c t) \quad ...(14.18)$$

Let us rearrange the last equation as,

$$s(t) = b(t)\sqrt{PT_b}\sqrt{\frac{2}{T_b}}\cos(2\pi f_c t) \quad ...(14.19)$$

Now, let $\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t)$ represents an orthonormal carrier signal. Equation (14.17) also gives equation for carrier. It is slightly different than $\phi_1(t)$ defined here. Then, we may write equation (14.19) as,

$$s(t) = b(t)\sqrt{PT_b}\,\phi_1(t) \quad ...(14.20)$$

The bit energy E_b is defined in terms of power 'P' and bit duration T_b as,

$$E_b = PT_b \quad ...(14.21)$$

Therefore, equation (14.20) becomes,

$$s(t) = \pm\sqrt{E_b}\,\phi_1(t) \quad ...(14.22)$$

Here, $b(t)$ is simply ± 1.

Thus, on the single axis of $\phi_1(t)$, there will be two points. One point will be located at $+\sqrt{E_b}$ and other point will be located at $-\sqrt{E_b}$. This has been shown in figure 14.13.

* Here, $f_b = \frac{1}{T_b}$.

Fig. 14.13. *Geometrical representation of BPSK signal.*

At the receiver end, the point at $+\sqrt{E_b}$ on $\phi_1(t)$ represents symbol '1' and point at $-\sqrt{E_b}$ represents symbol '0'. The separation between these two points represents the isolation between symbols '1' and '0' in BPSK signal. This separation is generally called distance 'd'. From figure 14.13, it is obvious that the distance between the two points is,

$$d = +\sqrt{E_b} - \left(-\sqrt{E_b}\right)$$

or

$$d = 2\sqrt{E_b} \qquad \text{...(14.2[illegible])}$$

As this distance 'd' increases, the isolation between the symbols in BPSK signal is more. Thus, probability of error reduces.

14.6.5. Bandwidth for BPSK Signal

As discussed earlier, the spectrum of the BPSK signal is centred around the carrier frequency f_c.

If $f_b = \frac{1}{T_b}$, then for BPSK, the maximum frequency in the baseband signal will be f_b as shown in figure 13.12. In this figure, the main lobe is centred around carrier frequency f_c and extends from $f_c - f_b$ to $f_c + f_b$.

Therefore Bandwidth of BPSK signal will be,

BW = Highest frequency – Lowest frequency in the main lobe

$$BW = f_c + f_b - (f_c - f_b)$$

or

$$BW = 2f_b \qquad \text{...(14.24)}$$

Hence, the minimum bandwidth of BPSK signal is equal to twice of the highest frequency contained in baseband signal.

14.6.6. Error Probability of BPSK Signal Employing Coherent Reception or Matched Filter

The BPSK receiver shown in figure 14.9 is equivalent to the coherent receiver. It is nothing but coherent receiver. Let us say that the signal $x(t)$ represents $s_1(t)$ and $s_2(t)$. Then the BPSK receiver of figure 14.9 can be approximated to coherent as shown in figure 14.14.

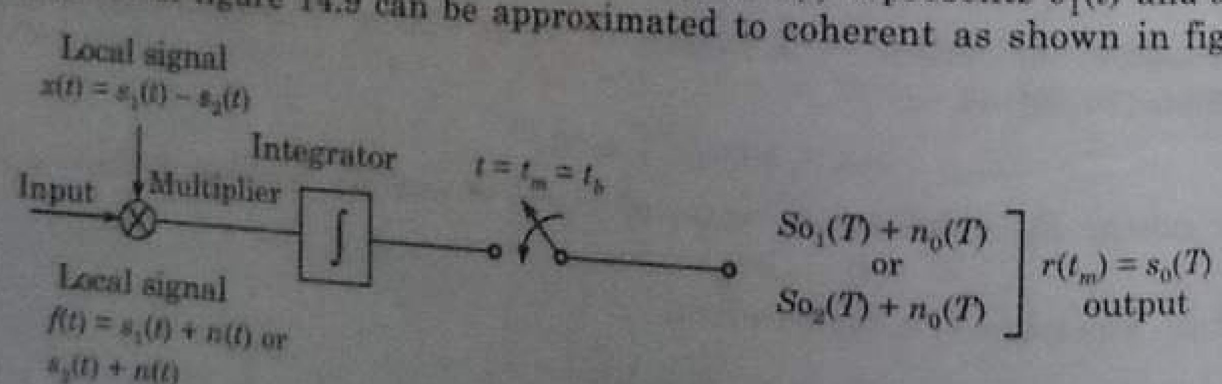


Fig. 14.14. *Coherent reception system*

Because, the correlator of figure 14.14 is equivalent to matched filter, we can apply the analysis of matched filter detection. We know that signal to noise ratio ρ_{max} at the output of receiver is given by

$$\rho_{max} = \frac{2}{N_0}\int_{-\infty}^{\infty} x^2(t)\,dt \qquad \text{...(14.25)}$$

Here, $x(t) = s_1(t) - s_2(t)$ i.e., input difference signal, then we have

$$\rho_{max} = \frac{2}{N_0}\int_{-\infty}^{\infty} [s_1(t) - s_2(t)]^2\,dt$$

or

$$\rho_{max} = \frac{2}{N_0}\left[\int_{-\infty}^{\infty} s_1^2(t)\,dt + \int_{-\infty}^{\infty} s_2^2(t)\,dt - 2\int_{-\infty}^{\infty} s_1(t)s_2(t)\,dt\right]$$

Since $s_1(t) = -s_2(t)$, we have

$$\int_{-\infty}^{\infty} s_1(t)\,s_2(t)\,dt = -E_b$$

and $\int_{-\infty}^{\infty} s_1^2(t)\,dt = \int_{-\infty}^{\infty} s_2^2(t)\,dt = E_b$

Hence,

$$\rho_{max} = \frac{2}{N_0}[E_b + E_b + 2E_b] \qquad \text{by putting values}$$

or

$$\rho_{max} = \frac{8E_b}{N_0}$$

For the correlator, we can write

$$\left[\frac{s_{01}(t) - s_{02}(t)}{\sigma}\right]^2 = \frac{8E_b}{N_0}$$

or

$$\frac{s_{01}(t) - s_{02}(t)}{\sigma} = 2\sqrt{2}\sqrt{\frac{E_b}{N_0}} \qquad \text{...(14.26)}$$

The probability of error is given as,

$$P(e) = \frac{1}{2}\,erfc\left[\frac{s_{01}(T) - s_{02}(T)}{2\sqrt{2}\,\sigma}\right]$$

From equation (14.26), we can write above equation as under:

$$P(e) = \frac{1}{2}\,erfc\left[\frac{1}{2\sqrt{2}}\cdot 2\sqrt{2}\sqrt{\frac{E_b}{N_0}}\right]$$

or

$$P(e) = \frac{1}{2}\,erfc\sqrt{\frac{E_b}{N_0}} \qquad \text{...(14.27)}$$

This is the required expression for error probability of BPSK reception using coherent/matched filter detection.

14.6.7. Drawbacks of BPSK

Figure 14.9 shows the block diagram of BPSK receiver. To regenerate the carrier in the receiver, we start by squaring $b(t)\sqrt{2P}\cos(2\pi f_c t + \theta)$. If the received signal is $-b(t)\sqrt{2P}\cos(2\pi f_c t + \theta)$, then the squared signal remains same as before. Hence, the recovered carrier is unchanged even if the input signal has changed its sign. Therefore, it is not possible to determine whether the received signal is equal to $b(t)$ or $-b(t)$. Infact, this results in ambiguity in the output signal.

This problem can be removed if we use differential phase shift keying (DPSK). However differential phase shift keying (DPSK) also has some other problems. DPSK will be discussed in detail later on in this chapter. Other problems of BPSK are ISI and Interchannel interference. However, these problems can be reduced to some extent by making use of filters.

14.7. Coherent Binary Frequency Shift Keying (BFSK)

In binary frequency shift keying (BFSK), the frequency of the carrier is shifted according to the binary symbol. However, the phase of the carrier is unaffected. This means that we have two different frequency signals according to binary symbols. Let there be a frequency shift by Ω. Then we can write following equations.

If $\quad b(t) = \text{'1'}, \quad$ then $s_H(t) = \sqrt{2P_s}\cos(2\pi f_c + \Omega)\,t$ ...(14.28)

If $\quad b(t) = \text{'0'}, \quad$ then $s_L(t) = \sqrt{2P_s}\cos(2\pi f_c - \Omega)\,t$ ...(14.29)

Hence, there is increase or decrease in frequency by Ω. Let us use the following conversion table to combine above two FSK equations:

Table 14.1. Conversion table for BPSK representation

$b(t)$ Input	$d(t)$	$P_H(t)$	$P_L(t)$
1	+ 1V	+ 1V	0V
0	– 1V	0V	+ 1V

The equations (14.28) and (14.29) combinely may be written as

$$s(t) = \sqrt{2P_s}\cos[(2\pi f_c + d(t)\Omega)\,t] \quad ...(14.30)$$

Hence, if symbol '1' is to be transmitted, the carrier frequency will be $f_c + \left(\frac{\Omega}{2\pi}\right)$. If symbol '0' is to be transmitted, then the carrier frequency will be $f_c - \left(\frac{\Omega}{2\pi}\right)$.

Therefore, we have

Thus, $\quad f_H = f_c + \frac{\Omega}{2\pi} \quad$ for symbol '1' ...(14.31)

$\quad f_L = f_c - \frac{\Omega}{2\pi} \quad$ for symbol '0' ...(14.32)

14.7.1. Generation of BFSK

It may be observed from Table 14.1 that $P_H(t)$ is same as $b(t)$ and also $P_L(t)$ is inverted version of $b(t)$. The block diagram for BFSK generation is shown in figure 14.15.

We know that input sequence $b(t)$ is same as $P_H(t)$. An inverter is added after $b(t)$ to get $P_L(t)$. The level shifter $P_H(t)$ and $P_L(t)$ are unipolar signals. The level shifter converts the '+1' level to $\sqrt{P_sT_b}$. Zero level is unaffected. Thus, the output of the level shifters will be either $\sqrt{P_sT_b}$ (if '+1') or zero (if input is zero). Further, there are product modulators after level shifter. The two carrier signals $\phi_1(t)$ and $\phi_2(t)$ are used. $\phi_1(t)$ and $\phi_2(t)$ are orthogonal to each other. In one bit period of input signal (*i.e.*, T_b), $\phi_1(t)$ or $\phi_2(t)$ have integral number of cycles. Thus, the modulated signal is having continuous phase. Figure 14.16 shows such type of BFSK signal. The adder then adds the two signals.

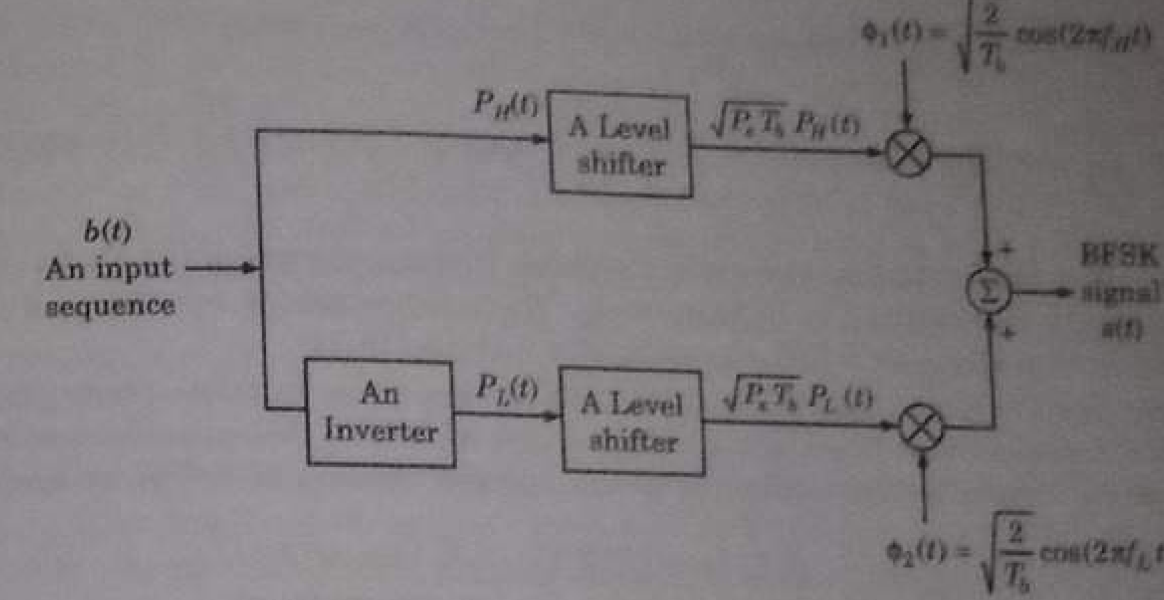


Fig. 14.15. *Block diagram for BFSK generation.*

Note: *Here it may be noted that outputs from both the multipliers are not possible at a time. This is because $P_H(t)$ and $P_L(t)$ are complementary to each other. Therefore, if $P_H(t) = 1$, then output will be only due to upper modulator and lower modulator output will be zero [since $P_L(t) = 0$].*

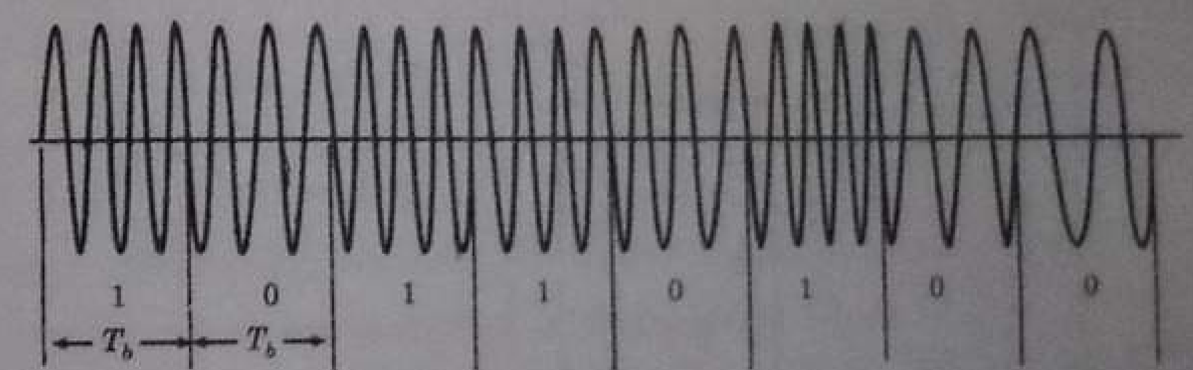


Fig. 14.16. *The BFSK signal.*

14.7.2. The Spectrum of BFSK Signal

In figure 14.15, the BFSK signal $s(t)$ may be written as,

$$s(t) = \sqrt{2P_s}\,P_H(t)\cos(2\pi f_H t) + \sqrt{2P_s}\,P_L(t)\cos(2\pi f_L t) \quad ...($$

This is the expression for BFSK signal. Let us compare this equation with BPSK eq... which is written below:

$$S_{BPSK}(t) = b(t)\sqrt{2P}\cos(2\pi f_c t) \quad ...($$

It may be noted that this equation is identical to BFSK equation. In BPSK equatio... $P_H(t)$ or $P_L(t)$ are un... is a bipolar signal where as in BFSK, the similar coefficients $P_H(t)$ or $P_L(t)$ are un... Hence, let us convert these coefficients in bipolar form as under:

$$P_H(t) = \frac{1}{2} + \frac{1}{2}P'_H(t) \quad ...$$

and

$$P_L(t) = \frac{1}{2} + \frac{1}{2}P'_L(t) \quad ...$$

where $P'_H(t)$ and $P'_L(t)$ will be bipolar (*i.e.*, + 1 or – 1).

Substituting these values in equation (14.33), we obtain

$$s(t) = \sqrt{2P_s}\left[\frac{1}{2}+\frac{1}{2}P_H'(t)\right]\cos(2\pi f_H t)+\sqrt{2P_s}\left[\frac{1}{2}+\frac{1}{2}P_L'(t)\right]\cos(2\pi f_L t)$$

or $$s(t) = \sqrt{\frac{P_s}{2}}\cos(2\pi f_H t)+\sqrt{\frac{P_s}{2}}\cos(2\pi f_L t)+\sqrt{\frac{P_s}{2}}P_H'(t)\cos(2\pi f_H t)+\sqrt{\frac{P_s}{2}}P_L'(t)\cos(2\pi f_L t)$$ ...(14.37)

In this equation, the first term represents the single frequency impulse situated at frequency f_H. The second term represents the impulse at f_L. These are constant amplitude pulses. The last two terms are identical to BPSK equation of equation (14.34).

Here $P_H'(t)$ and $P_L'(t)$ are equivalent to $b(t)$. Therefore, these last two terms in equation (14.37) produce the spectrum which are similar to that of BPSK. One spectrum is located at f_H and other at f_L. Hence, we can write the power spectral density of BFSK as under:

$$S(f) = \sqrt{\frac{P_s}{2}}\left[\delta(f-f_H)+\delta(f+f_L)+\frac{P_sT_b}{2}\left\{\frac{\sin(\pi f_H T_b)}{\pi f_H T_b}\right\}^2+\frac{P_sT_b}{2}\left\{\frac{\sin(\pi f_L T_b)}{\pi f_L T_b}\right\}^2\right]$$ ...(14.38)

Figure 14.17 illustrates the plot of power spectral density of BFSK signal expressed by equation (14.38).

Also, f_H and f_L are selected such that,

$$f_H - f_L = 2f_b$$ ...(14.39)

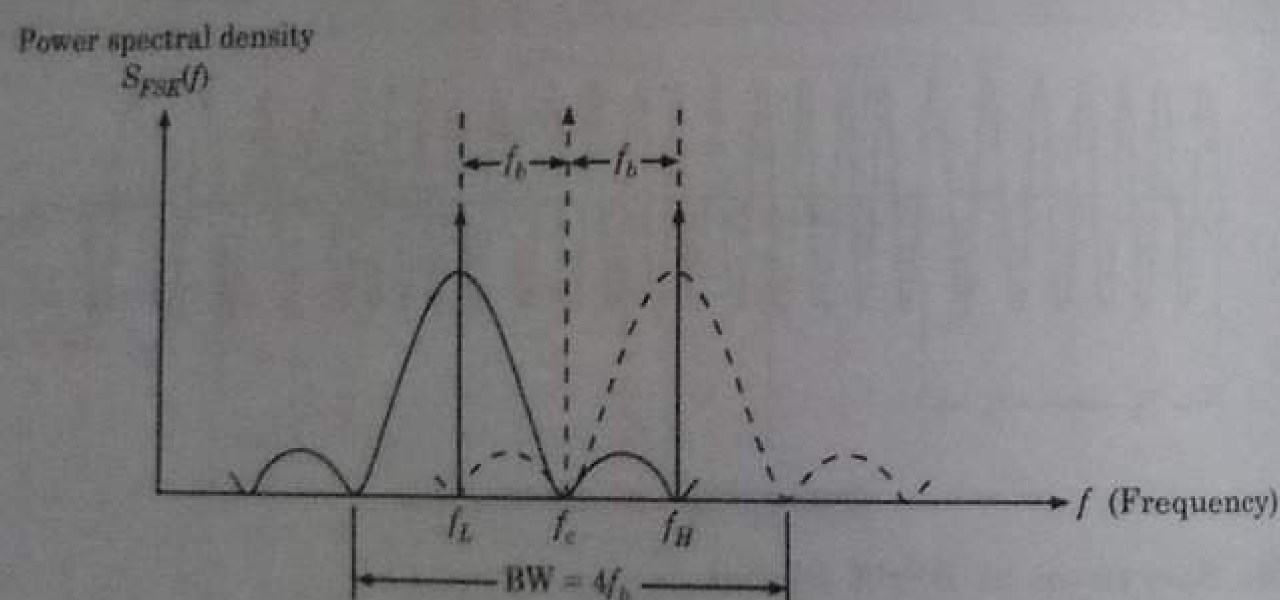


Fig. 14.17. *Illustration of Power spectral density (psd) of a BFSK signal.*

With such types of selection, it is obvious from the spectrums in the above figure that the two frequencies f_H and f_L may be identified properly. The interference between the spectrums is not much with the above assumption.

14.7.3. Bandwidth of BFSK Signal

From figure 14.17, it is obvious that the width of one lobe is $2f_b$. The two main lobes due to f_H and f_L are placed such that the total width due to both main lobes is $4f_b$.

Therefore, we have

Bandwidth of BFSK = $2f_b + 2f_b$

or $$BW = 4f_b$$ ...(14.40)

Now, if we compare this bandwidth with that of BPSK, we note that,

$$BW\ (BFSK) = 2 \times BW(BPSK)$$ ...(14.41)

14.7.4. Detection of BFSK

Figure 14.18 shows the block diagram of a scheme for demodulation of BFSK wave using coherent detection technique. The detector consists of two correlators that are individually tuned to two different carrier frequencies to represent symbols '1' and '0'. A correlator consists of a multiplier followed by an integrator. Then, the received binary FSK signal is applied to the multipliers of both the correlators. To the other input of the multipliers, carriers with frequency f_{c1} and f_{c2} are applied as shown in figure 14.18. The multiplied output of each multiplier is subsequently passed through integrators generating output l_1 and l_2 in the two paths. The output of the two integrators are then fed to the decision making device. The decision making device is essentially a comparator which compares the output l_1 (in the upper path) and output l_2 (in the lower path). If the output l_1 produced in the upper path (associated with frequency f_{c1}) is greater than the output l_2 produced in the lower path (associated with frequency f_{c2}), the detector makes a decision in favour of symbol 1. If the output l_1 is less than l_2, then the decision making device decides in favour of symbol 0 (say). This type of digital communication receivers are also called *correlation receivers*. As discussed earlier, the detector based upon coherent detection requires phase and timing synchronisation.

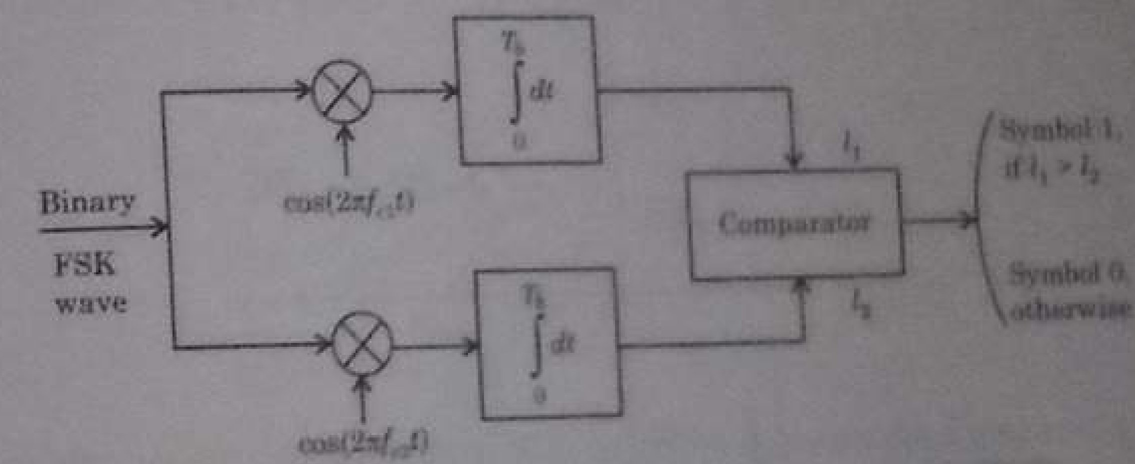


Fig. 14.18. *Block diagram of BFSK receiver (detection of BFSK).*

14.7.5. Geometrical Representation of Orthogonal BFSK

As a matter of fact, orthogonal carriers are used for M-ary PSK and QASK. The different signal points are represented geometrically in ϕ_1 ϕ_2-plane. For geometrical representation of BFSK signals, such orthogonal carriers are required. From figure 14.15, we know that two carriers $\phi_1(t)$ and $\phi_2(t)$ of two different frequencies f_H and f_L are used for modulation. To make $\phi_1(t)$ and $\phi_2(t)$ orthogonal, the frequencies f_H and f_L must be some integer multiple of band frequency 'f_b'.

Thus, $$f_H = m\, f_b$$ ...(14.42)

and $$f_L = n\, f_b$$ ...(14.43)

Here: $f_b = \frac{1}{T_b}$, then the carriers would be

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi m\, f_b\, t)$$ ...(14.44)

and $$\phi_2(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi\, m f_b\, t)$$ ...(14.45)

The carriers $\phi_1(t)$ and $\phi_2(t)$ are orthogonal over the period T_b. We can write equation (14.28) and equation (14.29) as,

$$s_H(t) = \sqrt{P_sT_b}\sqrt{\frac{2}{T_b}}\cos(2\pi f_H t)$$

$$s_L(t) = \sqrt{P_s T_b}\sqrt{\frac{2}{T_b}}\cos(2\pi f_L t)$$

Here

$$f_H = f_c + \frac{\Omega}{2\pi}$$

and

$$f_L = f_c - \frac{\Omega}{2\pi}$$

Using the relations in equations (14.42) to (14.45), we can write above equations as,

$$s_H(t) = \sqrt{P_s T_b}\cdot\phi_1(t) \quad ...(14.46)$$

and

$$s_L(t) = \sqrt{P_s T_b}\cdot\phi_2(t) \quad ...(14.47)$$

Thus, based on the above two equations, we can draw the signal space diagram as shown in figure 14.19.

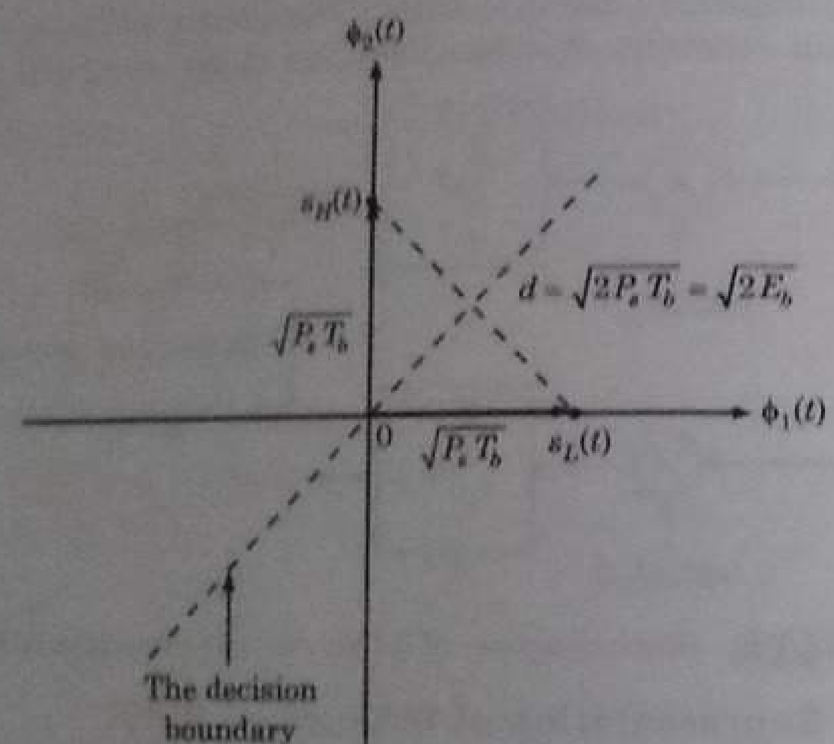


Fig. 14.19. *Illustration of signal space representation of orthogonal BFSK signal.*

14.7.5.1. Distance Between Signal Points

Note that there are two signal points in the signal space. The distance between these two points may be evaluated as under:

$$d^2 = \left(\sqrt{P_s T_b}\right)^2 + \left(\sqrt{P_s T_b}\right)^2$$

or

$$d^2 = 2P_s T_b \quad \text{or} \quad d = \sqrt{2P_s T_b} \quad ...(14.48)$$

Since $P_s T_b = E_b$, we can write above relation (*i.e.*, equation (14.48)) as under:

$$d = \sqrt{2E_b} \quad ...(14.49)$$

As compared to the distance of BPSK, we may observe that this distance is smaller than BPSK.

14.7.6. Geometrical Representation of Non-Orthogonal BFSK Signals

As a matter of fact, whenever the carriers $\phi_1(t)$ and $\phi_2(t)$ are non-orthogonal, then the signal point $S_H(t)$ or $S_L(t)$ would not lie exactly on the axes $\phi_1(t)$ and $\phi_2(t)$. Such a representation has been shown in figure 14.20.

The distance 'd' for non-orthogonal signal shown in figure 14.20 may be given approximately as,

$$d^2 = 2E_b\left[1 - \frac{\sin 2\pi(f_H - f_L)T_b}{2\pi(f_H - f_L)T_b}\right] \quad ...(14.50)$$

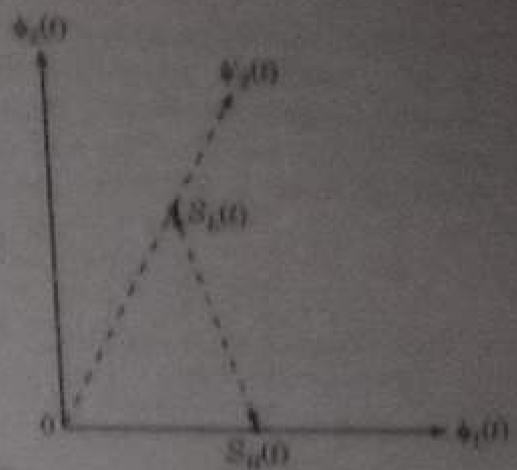


Fig. 14.20. *Geometrical representation of non-orthogonal BFSK signals.*

14.7.7. Probability of Error, P_e of BFSK Signals

The generated equation derived using union bound approximation gives probability of error as under:

$$P(e) = \frac{1}{2}\sum_{\substack{k=1\\k\neq i}}^{M} erfc\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \quad \text{for all } i \quad ...(14.50a)$$

Here, we have only two points in BFSK. Let $S_H(t)$ be $S_1(t)$ and $S_L(t)$ be $S_2(t)$. Then, distance between these two points can be written as,

$$d = d_{12} = \sqrt{2E_b}$$

Since there are only two points (*i.e.*, $M = 2$) take $i = 2$ and $k = 1$, then equation (14.50a) becomes,

$$P(e) = \frac{1}{2}erfc\left(\frac{d_{12}}{2\sqrt{N_0}}\right)$$

or

$$P(e) = \frac{1}{2}erfc\left(\frac{\sqrt{2E_b}}{2\sqrt{N_0}}\right)$$

or

$$P(e) = \frac{1}{2}erfc\sqrt{\frac{E_b}{2N_0}}$$

This equation shows that error probability of BFSK is higher compared to that of BPSK.

14.7.8. Advantages and Disadvantages of BFSK Signals

Even though the generation of BFSK is easier, it has many disadvantages compared to BPSK signal. Firstly, its bandwidth is greater than $4/f_b$, which is almost double the bandwidth of BPSK. Also, the distance between the signal points is less in case of BFSK. Therefore, the error rate of BFSK is more compared to BPSK.

14.8. Non-Coherent Binary Modulation Techniques

As discussed earlier, coherent detection exploits knowledge of the carrier wave's phase reference and thus providing the optimum error performance attainable with a digital modulation format of interest. However, when it is impractical to have knowledge of the carrier phase at the receiver, we make use of **non-coherent detection**. Thus, in this section, we shall study non-coherent binary modulation techniques *i.e.*, we shall study non-coherent detection of ASK and FSK. In the case of phase-shift keying (PSK), we cannot have "non-coherent PSK" since non-coherent means doing without phase information. However, there is a 'pseudo PSK' technique known as differential phase-shift keying (DPSK) which can be viewed as the non-coherent form of PSK.

14.9. Non-Coherent Binary Amplitude Shift Keying (ASK)

In the binary ASK case, the transmitted signal is defined as

$$s(t) = \sqrt{2P_s}\cos(2\pi f_c t)$$

Binary ASK signal can also be demodulated non-coherently using envelope detector. This greatly simplifies the design consideration required in synchronous detection. Non-coherent detection schemes do not require a phase-coherent local oscillator. This method involves some form of rectification and low pass filtering at the receiver. The block diagram of a non-coherent receiver for ASK signal has been shown in figure 14.21.

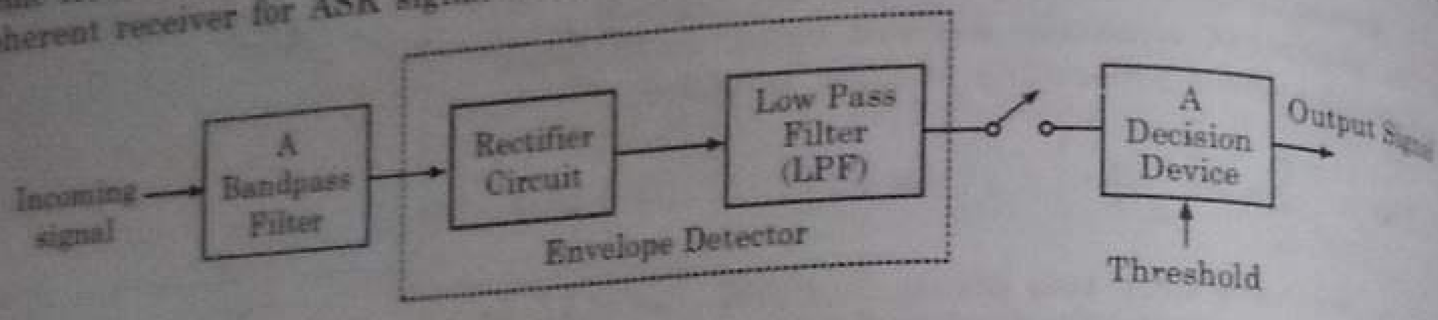


Fig. 14.21. *Non-coherent ASK detector.*

14.10. Non-Coherent Detection of FSK

Binary FSK waves may be demodulated non-coherently using envelope detector. The received FSK signal is applied to a bank of two bandpass filters, one tuned to frequency f_{c1} and the other tuned to f_{c2}. Each filter is followed by an envelope detector. The resulting outputs of the two envelope detectors are sampled and then compared with each other. The arrangement for non-coherent detection of FSK signal has been shown in figure 14.22.

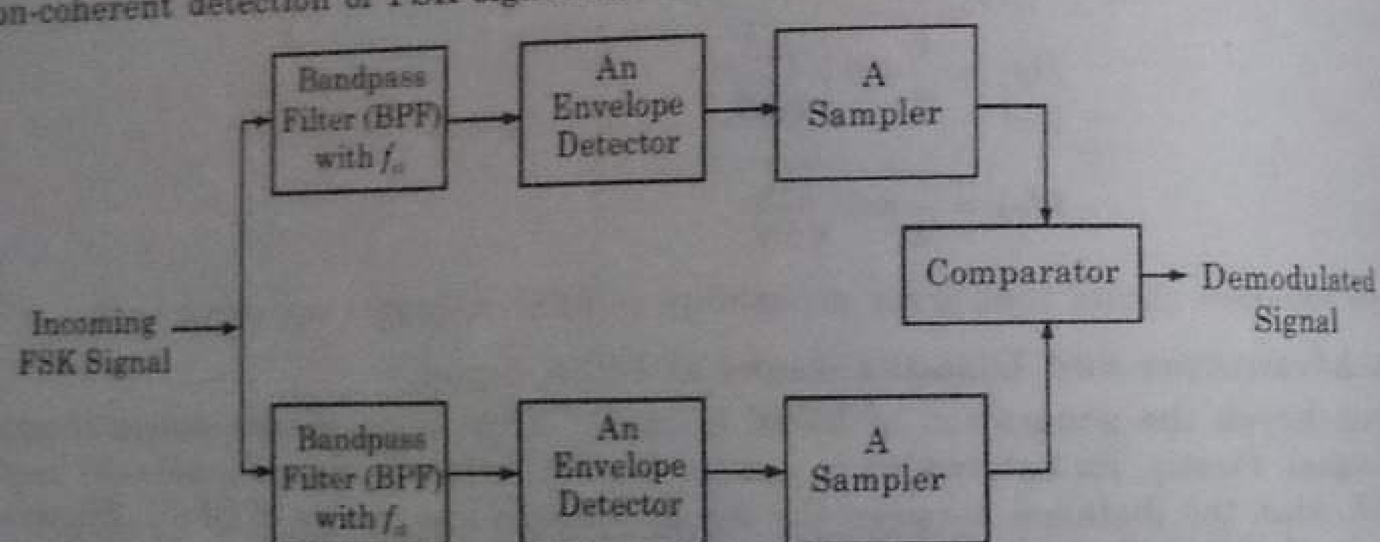


Fig. 14.22. *Non-coherent detection of FSK binary signals.*

A decision is made in favour of symbol '1' if the envelope detector output derived from the filter tuned to frequency f_{c1} is larger than that derived from the second filter. Otherwise a decision is made in favour of the symbol 0.

14.11. Differential Phase Shift Keying (DPSK)

(U.P. Tech., Semester, Examination, 2003-2004)

We can view differential phase-shift keying as the non-coherent version of the PSK. Differential phase shift keying (DPSK) is differentially coherent modulation method. DPSK does not need a synchronous (coherent) carrier at the demodulator. The input sequence of binary bits is modified such that the next bit depends upon the previous bit. Therefore, in the receiver the previous received bits are used to detect the present bit.

14.11.1. Generation of DPSK

Thus, in order to eliminate the need for phase synchronisation of coherent receiver with PSK, a differential encoding system can be used with PSK. The digital information content of the binary data is encoded in terms of signal transitions. As an example, the symbol 0 may

be used to represent transition in a given binary sequence (with respect to the previous encoded bit) and symbol '1' to indicate no transition. This new signaling technique which combines differential encoding with phase-shift keying (PSK) is known as *differential phase-shift keying (DPSK).*

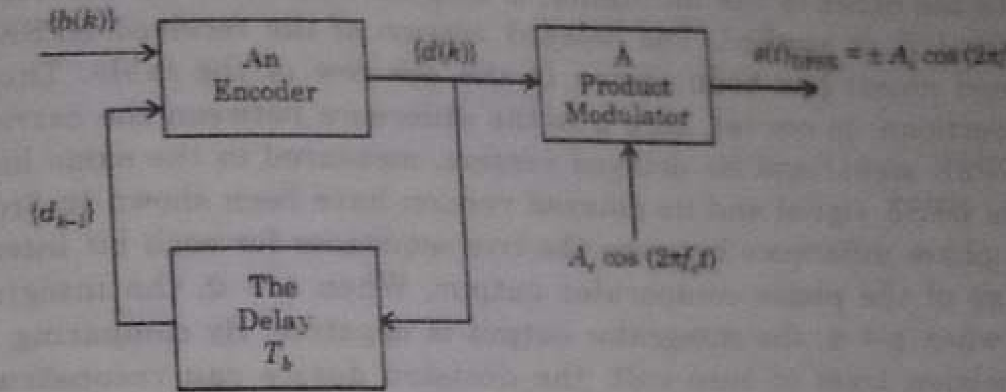


Fig. 14.23. *Illustration of the scheme to generate DPSK signals.*

A schematic arrangement for generating DPSK signal has been shown in figure 14.23. The data stream $b(t)$ is applied to the input of the encoder. The output of the encoder is applied to one input of the product modulator. To the other input of this product modulator, a sinusoidal carrier of fixed amplitude and frequency is applied. The relationship between the binary sequence and its differentially encoded version is illustrated in Table 14.2 for a assumed data sequence 0 0 1 0 0 1 0 0 1 1. In this illustration it has been assumed that the encoding has been done in such a way that *transition* in the given bianry sequence with respect to the previous encoded bit is represented by a symbol 0 and no *transition* by symbol '1'. It may be noted that an extra bit (symbol 1) has been arbitrarily added as an initial bit. This is essential to determine the encoded sequence. The phase of the generated DPSK signal has been shown in the third row of Table 14.2.

Table 14.2. **Differentially encoded sequences with phase.**

Binary data $\{b(k)\}$		0	0	1	0	0	1	0	0	1	1
Differentially encoded data $\{d(k)\}$	1*	0	1	1	0	1	1	0	1	1	1
Phase of DPSK	0	π	0	0	π	0	0	π	0	0	0
Shifted differentially encoded data $\{d_{k-1}\}$		1	0	1	1	0	1	1	0	1	1
Phase of shifted DPSK		0	π	0	0	π	0	0	π	0	0
Phase comparison output		−	−	+	−	−	+	−	−	+	+
Detected binary sequence		0	0	1	0	0	1	0	0	1	1

* Arbitrary starting reference bit.

14.11.2. Detection of DPSK

For detection of the differentially encoded PSK (*i.e.*, DPSK), we can use the receiver arrangement as shown in figure 14.24. The received DPSK signal is applied to one input of the multiplier. To the other of the multiplier, a delayed version of the received DPSK signal by the time interval T_b is applied. The delayed version of the received DPSK signal (in absence of channel noise) has been shown in the 4th row of the table. The output of difference is proportional to cos (φ), here φ is the difference between the carrier phase of the received DPSK signal and its delayed version, measured in the same bit interval. The phase angle of the DPSK signal and its delayed version have been shown in 3rd and 5th rows respectively. The phase difference between the two sequences for each bit interval is used to determine the sign of the phase comparator output. When $\phi = 0$, the integrator output is positive whereas when $\phi = \pi$, the integrator output is negative. By comparing the integrator output with a decision level of zero volt, the decision device can reconstruct the binary sequence by assigning a symbol '0' for negative output and a symbol '1' for positive output. The reconstructed binary data is shown in the last row of the table. It is thus seen that in absence of noise, the receiver can reconstruct the transmitted bianry data exactly. *DPSK may be viewed as a non-coherent version of PSK. It may also be noted that the reconstruction is invariant with the choice of the initial bit in the encoded data. This has been illustrated in the example 14.1 given below.*

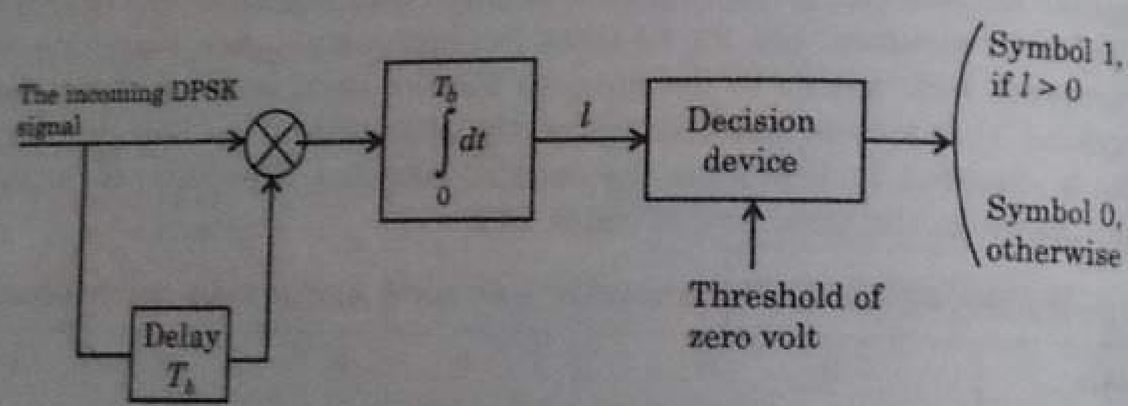


Fig. 14.24. *Receiver for the detection of DPSK signals.*

Example 14.1. A binary data strem 0 0 1 0 0 1 0 0 1 1 needs to be transmitted using DPSK technique. Prove that the reconstruction of the DPSK signal by the technique discussed in the previous article is independent of the choice of the extra bit.

Solution: In the last article, we have observed that DPSK signal can be detected accurately (in the absence of channel noise) without having a local oscillator for generation of synchronous carrier. The initial bit in the differentially encoded data was assumed to be '1'. In this example we use the initial bit to be symbol '0' and verify that the reconstruction is invariant with the choice of the initial bit. The results obtained for this case are given in Table 14.3. It can be easily verified that the extra chosen bit 0 changes the phase of the DPSK sequence but the detected sequence remains invariant.

Table 14.3. Differentially encoded sequences with phase.

Binary data $\{b(k)\}$		0	0	1	0	0	1	0	0	1	1
Differentially encoded data $\{d(k)\}$	0*	1	0	0	1	0	0	1	0	0	0
Phase of DPSK	π	0	π	π	0	π	π	0	π	π	π
Shifted differentially encoded data $\{d_{k-1}\}$		0	1	0	0	1	0	0	1	0	0
Phase of shifted DPSK		π	0	π	π	0	π	π	0	π	π
Phase comparison output		–	–	+	–	–	+	–	–	+	+
Detected binary sequence		0	0	1	0	0	1	0	0	1	1

* Starting reference bit.

14.11.3. Evaluation of Bandwidth of DPSK Signal

As discussed earlier that one previous bit is used to decide the phase shift of next bit. Thus, change in $b(t)$ occurs only if input bit is at level '1'. No change happens if input bit is at level '0'. Because, one previous bit is always used to define the phase shift in next bit, therefore, the symbol can be said to have two bits. Hence, one symbol duration (T) is equivalent to two bits duration ($2T_b$) *i.e.*,

Symbol duration $\quad T = 2T_b \quad$ ...(14.51)

Bandwidth is expressed as

$$BW = \frac{2}{T} = \frac{1}{T_b} = f_b \quad \text{...(14.52)}$$

Hence, the minimum bandwidth in DPSK is equal to f_b, *i.e.*, maximum baseband signal frequency.

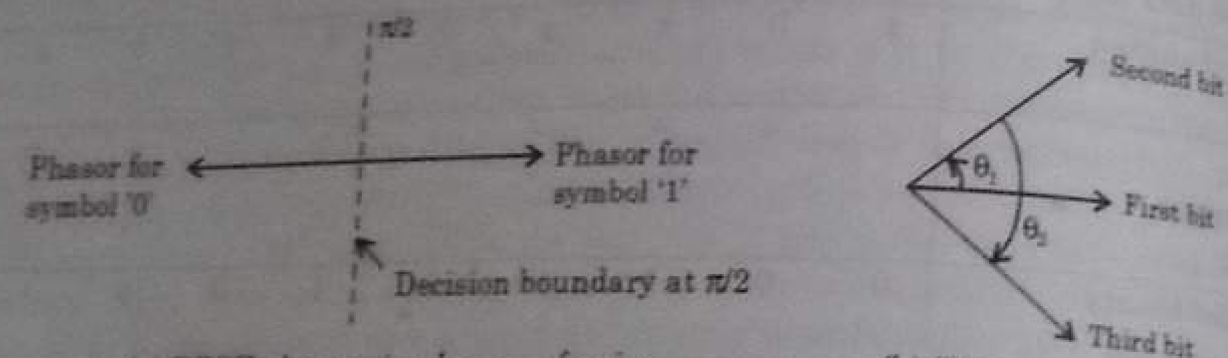
14.11.4. Error Probability of DPSK

Figure 14.25(a) shows the phasor diagram of DPSK signal when no noise is present. This means that in the absence of noise and transmission delay, the phase shift of the DPSK signal is either '0' or 'π'. Therefore, a decision boundary is drawn at $\frac{\pi}{2}$ as shown in figure 14.25(a). Therefore, we consider that the transmitted symbol is '1', if the phase difference between two consecutive bits differs by less than $\frac{\pi}{2}$. If the pahse difference between two consecutive bits differs by more than $\frac{\pi}{2}$, then decision is taken in favour of zero.

Figure 14.25(b) shows three consecutive bits. The first bit signal contains no noise, hence its phasor is along the horizontal line. Therefore, the symbol transmitted in first bit is assumed to be '1'. Because of noise, there is a phase difference of 'θ_1' between first and second bit. Since $\theta_1 < \frac{\pi}{2}$, second bit is also taken as symbol '1'. The phase difference between second and third bit is θ_2. From figure, it is clear that $\theta_2 > \frac{\pi}{2}$, hence third bit is ta... symbol zero. Since the phase differences are not exact between two successive bits, ... error is introduced in the decision. Therefore, DPSK system is called 'sub-optimum' ... If the synchronization is used to make phase diferences exact, then it becomes PSK ...

Because of the sub-optimum nature of DPSK, the error probability is higher than ... BPSK. The average probability of error of non-coherent receiver is given as,

$$P(e) = \frac{1}{2} e^{-E/2N_0}$$



(a) DPSK phasors in absence of noise (b) Three consecutive bits

Fig. 14.25.

Here, E is the energy per symbol and $\frac{N_0}{2}$ is spectral density of white Gaussian noise. ... know that symbol duration $T = 2T_b$. Hence, energy of the symbol will be

$$E = 2E_b$$

Hence, expression for $P(e)$ becomes,

$$P(e) = \frac{1}{2} e^{-E_b/N_0}$$

This is called average probability of error or bit error rate (BER) of DPSK system.

14.11.5. Advantages and Disadvantages of DPSK

We have observed in above discussion that DPSK has some advantages over BPSK; however, at the same time, it has some drawbacks.

(i) Advantages

(i) DPSK does not need carrier at the receiver end. This means that the complicated circuitry for generation of local carrier is not required.

(ii) The bandwidth requirement of DPSK is reduced as compared to that of BPSK.

(ii) Disadvantages

(i) The probability or error (*i.e.*, bit error rate) of DPSK is higher than that of BPSK.

(ii) Because DPSK uses two successive bits for its reception, error in the first bit creates error in the second bit. Therefore, error propagation in DPSK is more. On the other hand, in BPSK single bit can go in error since detection of each bit is independent.

(iii) Noise interference in DPSK is more.

***Note:** In DPSK, previous bit is used to detect next bit. Hence, if error is present in previous bit, detection of next bit can also be wrong. Hence, error is created in next bit also. Therefore, there is tendency of appearing errors in pairs in DPSK.*

14.12. Quadrature Phase Shift Keying (QPSK)

As a matter of fact, in communication systems, we have two main resources. These are the transmission power and the channel bandwidth. The channel bandwidth depends upon the bit rate or signaling rate f_b. In digital bandpass transmission, we use a carrier for transmission. This carrier is transmitted over a channel. If two or more bits are combined in some symbols, then the signaling rate will be reduced. Thus, the frequency of the carrier needed is also reduced. This reduces the transmission channel bandwidth. Hence, because of grouping of bits in symbols, the transmission channel bandwidth can be reduced. In quadrature phase shift keying (QPSK), two successive bits in the data sequence are grouped together. This reduces the bits rate or signaling rate (*i.e.*, f_b) and thus reduces the bandwidth of the channel.

In case of BPSK, we know that when symbol changes the level, the phase of the carrier is changed by 180°. Because, there were only two symbols in BPSK, the phase shift occurs in two levels only. However, in QPSK, two successive bits are combined. Infact, this combination of two bits forms four distinct symbols. When the symbol is changed to next symbol, then the phase of the carrier is changed by 45° (π/4 radians). Table 14.4 shows these symbols and their phase shifts.

Table 14.4. Symbol and corresponding phase shifts in QPSK

S. No.	*Input successive bits*		*Symbol*	*Phase shift in carrier*
$i = 1$	1(1 V)	0 (–1 V)	S_1	π/4
$i = 2$	0(–1 V)	0(–1 V)	S_2	3π/4
$i = 3$	0(–1 V)	1(1 V)	S_3	5π/4
$i = 4$	1(1 V)	1(1 V)	S_4	7π/4

Hence as shown in Table 14.4, there are four symbols and the phase is shifted by π/4 for each symbol.

14.13. Generation of QPSK

Figure 14.26 shows the block diagram of QPSK transmitter. Here, the input binary sequence is first converted to a bipolar NRZ type of signal. This signal is denoted by $b(t)$. It represents binary '1' by + 1 V and binary '0' by – 1 V. This signal has been shown in figure 14.27(*a*). The demultiplexer divides $b(t)$ into two separate bit streams of the odd numbered and even numbered bits. Here, $b_e(t)$ represents even numbered sequence and $b_0(t)$ represents odd numbered sequence. The symbol duration of both of these odd and even numbered sequences is $2T_b$. Hence, each symbol consists of two bits. Figure 14.27(*b*) and (*c*) illustrate the waveform of $b_e(t)$ and $b_0(t)$.

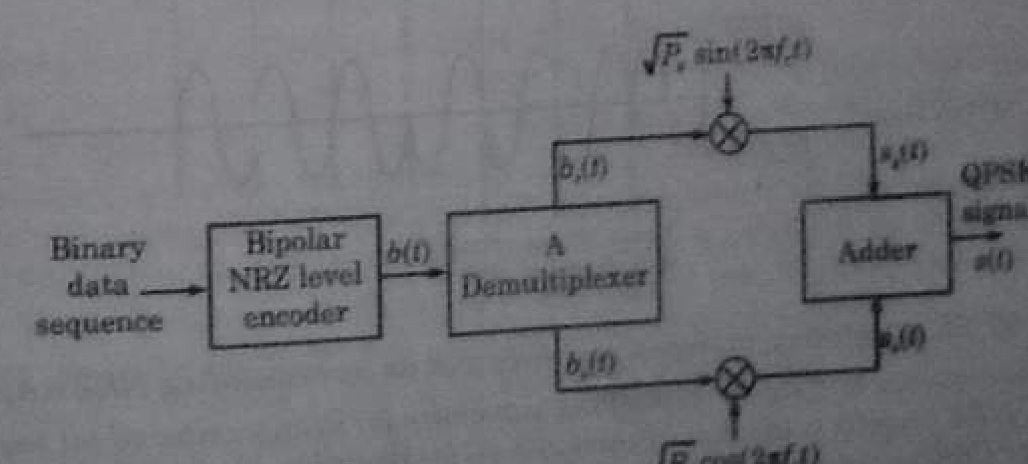


Fig. 14.26. *Generation of QPSK*

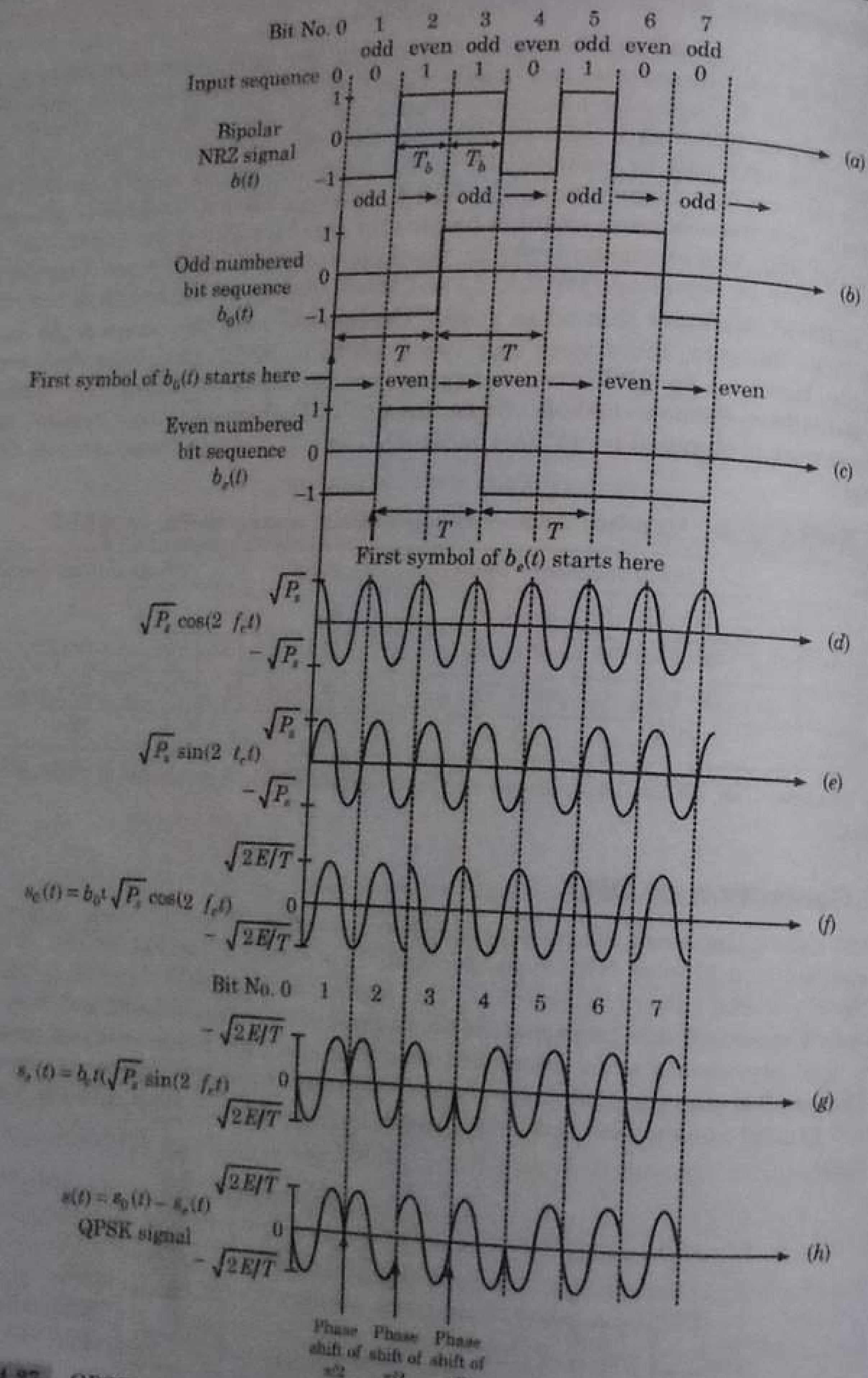


Fig. 14.27. *QPSK waveforms. (a) Input sequence and its corresponding NRZ waveform. (b) Odd numbered bit sequence and its corresponding waveform (c) Even numbered bit sequence and its NRZ waveform (d) Basis function $\phi_1(t)$ (e) Basis function $\phi_2(t)$ (f) Binary PSK waveform for odd numbered channel (g) Binary PSK waveform for even numbered channel (h) Final QPSK waveform.*

It may be observed that the first even bit occurs after the first odd bit. Hence, even numbered bit sequence $b_e(t)$ starts with the delay of one bit period due to first odd bit. Thus, first symbol of $b_e(t)$ is delayed by one bit period 'T_b' with respect to first symbol of $b_0(t)$. This delay of T_b is known as offset. This shows that the change in levels of $b_e(t)$ and $b_0(t)$ cannot occur at the same time due to offset or staggering.

Also, the bit steam $b_e(t)$ modulates carrier $\sqrt{P_s}\cos(2\pi f_ct)$ and $b_0(t)$ modulates $\sqrt{P_s}\sin(2\pi f_ct)$. These modulators are the balanced modulators. The two carriers $\sqrt{P_s}\cos(2\pi f_ct)$ and $\sqrt{P_s}\sin(2\pi f_ct)$ have been shown in figure 14.27(*d*) and (*e*). There carriers are also known as **quadrature carriers.**

The two modulated sigals can be written as,

$$s_e(t) = b_e(t)\sqrt{P_s}\sin(2\pi f_c t) \qquad ...(14.53)$$

and

$$s_0(t) = b_0(t)\sqrt{P_s}\cos(2\pi f_c t) \qquad ...(14.54)$$

Hence, $s_e(t)$ and $s_0(t)$ are basically BPSK signals. The only difference is that $T = 2T_b$ here. The value of $b_e(t)$ and $b_0(t)$ would be + 1*V* or – 1*V*. Figure 14.27 (*f*) and (*g*) shows the waveforms of $s_e(t)$ and $s_0(t)$. The adder in figure 14.26 adds these two signals $b_e(t)$ and $b_0(t)$.

The output of the adder is QPSK signal and it is given by,

$$s(t) = s_0(t) + s_e(t)$$

or

$$s(t) = b_0(t)\sqrt{P_s}\cos(2\pi f_ct) + b_e(t)\sqrt{P_s}\sin(2\pi f_ct) \qquad ...(14.55)$$

Figure 14.27(*h*) shows the QPSK signal represented by equation (14.55). In QPSK signal in figure 14.27(*h*), if there is any phase change, it occurs at minimum duration of T_b. This is because the two signals $s_e(t)$ and $s_0(t)$ have an offset of 'T_b'. Due to this offset, the phase shift in QPSK signal is $\frac{\pi}{2}$.

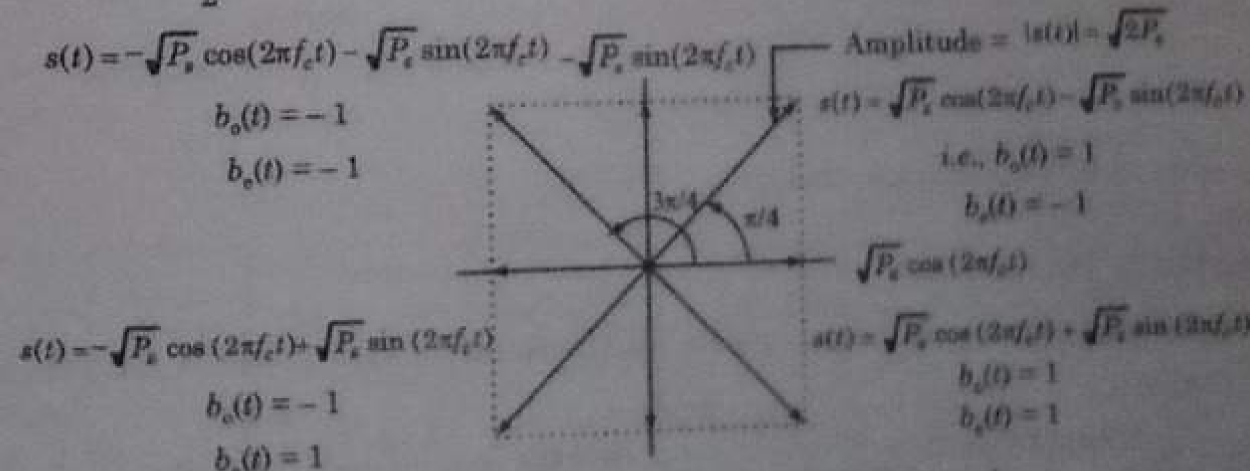


Fig. 14.28. *The Phasor diagram of QPSK signal*

14.13.1. Reception of QPSK (*i.e.* Detection of QPSK)

Figure 14.29 shows the QPSK receiver. This is synchronous reception. Hence, the coheren carrier is to be recovered from the received signal $s(t)$. The received signal $s(t)$ is first raise to its 4th power, *i.e.*, $s^4(t)$. After that, it is allowed to pass through a bandpass filter (BPF which is centred around $4f_c$. The output of the bandpass filter is a coherent carrier of frequenc $4f_c$. This is divided by 4 and it provides two coherent quadrature carriers, *i.e.*, $\cos(2\pi f_ct)$ an $\sin(2\pi f_ct)$. These coherent carriers are applied to two synchronous demodulators. Thes synchronous demodulators consist of multiplier and an integrator.

The incoming signal is applied to both the multipliers. Here, the integrator integrate the product signal over two bit interval (*i.e.*, $T_s = 2T_b$). At the end of this period, the output o integrator is sampled. The outputs of the two integrators are sampled at the offset of one b period, T_b. Hence, the output of the multiplexer is the signal $b(t)$. This means that the od and even sequences are combined by multiplexer.

Now, let us consider the product signal at the output of upper multiplier, *i.e.*,

$$s(t)\sin(2\pi f_c t) = b_0(t)\sqrt{P_s}\cos(2\pi f_c t)\sin(2\pi f_c t) + b_e(t)\sqrt{P_s}\sin^2(2\pi f_c t) \quad ...(14.56)$$

This signal is integrated by the upper integrator in figure 14.29.

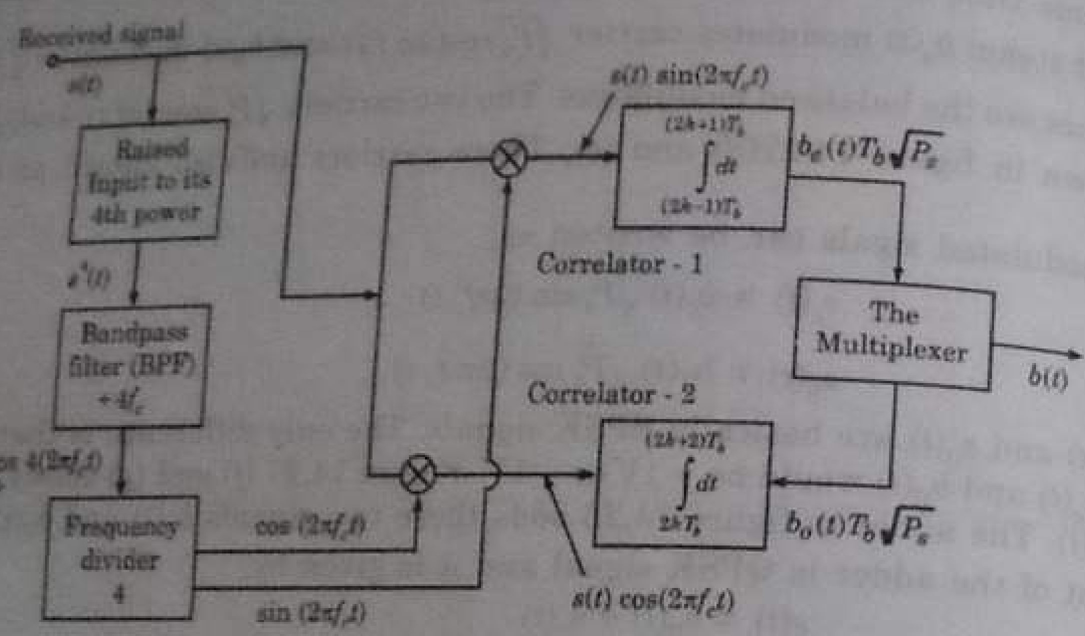


Fig. 14.29. *Reception of QPSK.*

Therefore, we have

$$\int_{(2k-1)T_b}^{(2k+1)T_b} s(t)\sin(2\pi f_c t)\,dt = b_0(t)\sqrt{P_s}\int_{(2k-1)T_b}^{(2k+1)T_b}\cos(2\pi f_c t)\sin(2\pi f_c t)\,dt$$

$$+\, b_e(t)\sqrt{P_s}\int_{(2k-1)T_b}^{(2k+1)T_b}\sin^2(2\pi f_c t)\,dt \quad ...(14.57)$$

Now, since $\frac{1}{2}\sin(2x) = \sin x \cdot \cos x$

and $\sin^2(x) = \frac{1}{2}[1-\cos(2x)]$

Therefore, using these two trigonometric identities in equation (14.57), we get

$$\int_{(2k-1)T_b}^{(2k+1)T_b} s(t)\sin(2\pi f_c t)\,dt = \frac{b_0(t)\sqrt{P_s}}{2}\int_{(2k-1)T_b}^{(2k+1)T_b}\sin 4\pi f_c t\,dt + \frac{b_e(t)\sqrt{P_s}}{2}\int_{(2k-1)T_b}^{(2k+1)T_b} 1\cdot dt$$

$$-\frac{b_e(t)\sqrt{P_s}}{2}\int_{(2k-1)T_b}^{(2k+1)T_b}\cos 4\pi f_c t\,dt$$

In this equation, the first and third integration terms involve integration of sinusoidal riers over two bit period. They have full (integral number of) cycles over two bit periods d thus integration will be zero, *i.e.*,

$$\int_{(2k-1)T_b}^{(2k+1)T_b} s(t)\sin(2\pi f_c t)\,dt = \frac{b_e(t)\sqrt{P_s}}{2}[t]_{(2k-1)T_b}^{(2k+1)T_b}$$

$$= \frac{b_e(t)\sqrt{P_s}}{2}\times 2T_b = b_e(t)\sqrt{P_s}\,T_b \quad ...(14.58)$$

Hence, the upper integrator responds to even sequence only. Similarly, we can obtain the output of lower integrator as $b_0(t)\sqrt{P_s}T_b$.

Note: *Even though bit synchronizer has not been shown in figure 14.29, it is assumed to be present with the integrator to locate starting and ending times of integration. The multiplexer is also operated by bit synchronizer. The amplitudes of voltage marked in figure 14.29 are arbitrary. They can change depending upon the gains of the integrator.*

14.13.2. Concept of Carrier Synchronization in QPSK

Both the carriers are to be synchronized properly in coherent detection in QPSK. Figure 14.30 shows the PLL system for carrier synchronization in QPSK.

The fourth power of the input signal consists of discrete frequency component at $4f_c$. We know that,

$$\cos 4(2\pi f_c t) = \cos(8\pi f_c t + 2\pi N)$$

where '*N*' is the number of cycles over the bit period. It is always an integer value. When the frequency division by four takes place, the RHS of this equation becomes $\cos\left(2\pi f_c t + \frac{N\pi}{2}\right)$.

This indicates that the output has a fixed phase error of $\frac{N\pi}{2}$. Differential encoding can be used to nullify the phase error events. The PLL remains locked with the phase of '$4f_c$' and then output of PLL is divided by 4. This provides a coherent carrier. A 90° phase shift is added to this carrier to produce a quadrature carrier.

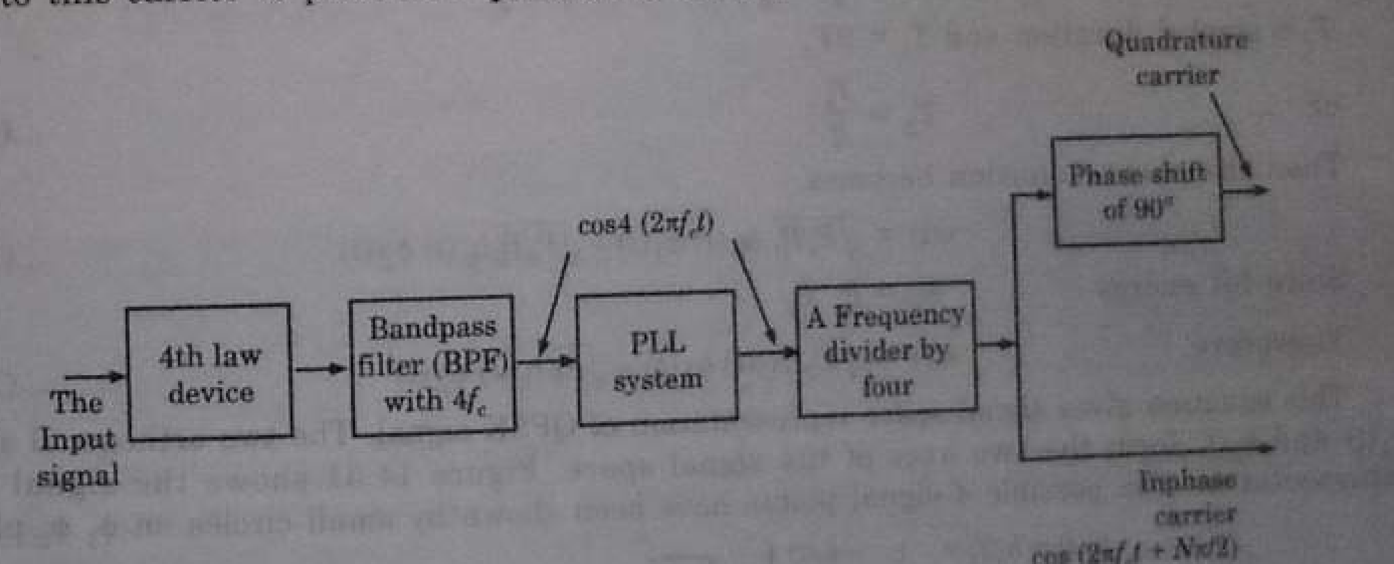


Fig. 14.30. *PLL system used for carrier synchronization in QPSK.*

14.13.3. Signal Space Representation in QPSK Signals

Figure 14.31 shows the phasor diagram of QPSK signal. Depending upon the combination of two successive bits, the phase shift occurs in carrier. This means that the QPSK signal in equation (14.55) can be written as,

$$s(t) = \sqrt{2P_s}\cos\left[2\pi f_c t + (2m+1)\frac{\pi}{4}\right] \qquad m = 0, 1, 2, 3 \quad ...(14.60)$$

Here, this equation takes four values and they represent the phasors of figure 14.31. This equation can be expanded as under:

$$s(t) = \sqrt{2P_s}\cos(2\pi f_c t)\cos\left[(2m+1)\frac{\pi}{4}\right] - \sqrt{2P_s}\sin(2\pi f_c t)\sin\left[(2m+1)\frac{\pi}{4}\right]$$

Let us rearrange the above equation as under:

$$s(t) = \left\{\sqrt{P_s T_s}\cos\left[(2m+1)\frac{\pi}{4}\right]\right\}\sqrt{\frac{2}{T_s}}\cos(2\pi f_c t)$$

$$-\left\{\sqrt{P_s T_s}\sin\left[(2m+1)\frac{\pi}{4}\right]\right\}\sqrt{\frac{2}{T_s}}\sin(2\pi f_c t) \quad ...(14$$

Again, let $\phi_1(t) = \sqrt{\frac{2}{T_s}}\cos(2\pi f_c t)$ ...(14

and $\phi_2(t) = \sqrt{\frac{2}{T_s}}\sin(2\pi f_c t)$ ...(14

These two signals are known as orthogonal signals and they are used as carriers in QPSK modulator.

Let $b_0(t) = \sqrt{2}\cos\left[(2m+1)\frac{\pi}{4}\right]$ ...(14

and $b_e(t) = -\sqrt{2}\sin\left[(2m+1)\frac{\pi}{4}\right]$ ...(14

With the use of equations (14.61) to (14.64) we can write equation (14.60) as under

$$s(t) = \sqrt{P_s T_s}\cdot\frac{1}{\sqrt{2}}b_0(t)\,\phi_1(t) + \sqrt{P_s T_s}\cdot\frac{1}{\sqrt{2}}b_e(t)\,\phi_2(t)$$

or

$$s(t) = \sqrt{P_s\cdot\frac{T_s}{2}}\,b_0(t)\,\phi_1(t) + \sqrt{P_s\frac{T_s}{2}}\,b_e(t)\,\phi_2(t)$$

T_s = symbol duration and $T_s = 2T_b$

or $T_b = \frac{T_s}{2}$ ...(14.6

Then the above equation becomes,

$$s(t) = \sqrt{P_s T_b}\,b_0(t)\,\phi_1(t) + \sqrt{P_s T_b}\,b_e(t)\,\phi_2(t) \quad ...(14.67$$

Since bit energy $E_b = P_s T_b$

Therefore, $s(t) = \sqrt{E_b}\,b_0(t)\,\phi_1(t) + \sqrt{E_b}\,b_e(t)\,\phi_2(t)$ ...(14.68

This equation gives signal space representation of QPSK signal. The two orthogonal signals $\phi_1(t)$ and $\phi_2(t)$ form the two axes of the signal space. Figure 14.31 shows the signal space representation. The possible 4 signal points have been shown by small circles on ϕ_1 ϕ_2-plane.

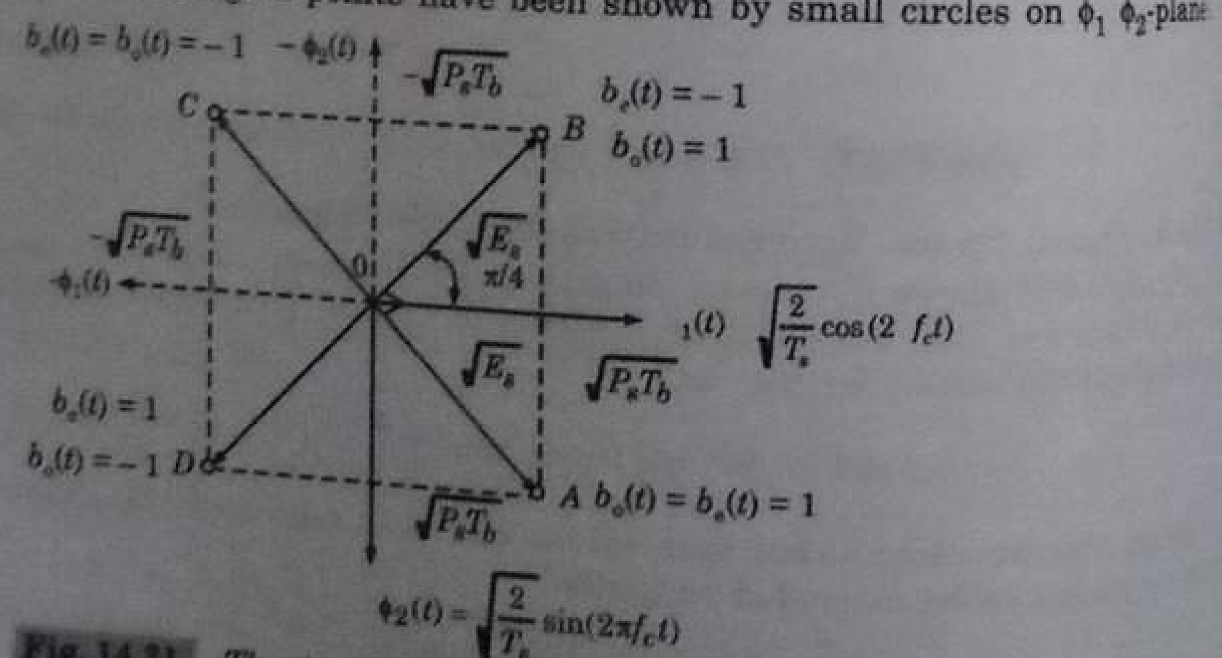


Fig. 14.31. *The signal space representation for QPSK signals.*

From each signal point, we obtain two bits. For example, from point 'A', we obtain two bits as (1, 1) and from 'B' we obtain bits as (– 1, 1). The distance of any signal point from origin 'O', given as,

$$\text{Distance 'OB'} = \sqrt{P_sT_b + P_sT_b} = \sqrt{2P_sT_b}$$

$$= \sqrt{P_sT_s} \qquad [\because 2T_b = T_s] \quad ...(14.69)$$

or

$$\text{'OB'} = \sqrt{E_s} \qquad [\because P_s T_s = E_s] \quad ...(14.70)$$

Hence, the length of each signal point from origin is $\sqrt{E_s}$. We know that $b_e(t)$ and $b_o(t)$ represent two successive bits. There is an offset of 'T_b' between $b_e(t)$ and $b_o(t)$. Therefore, $b_e(t)$ and $b_o(t)$ both cannot change their levels simultaneously. Hence, either $b_e(t)$ or $b_o(t)$ can change at a time.

Let us say that $b_e(t) = b_o(t) = 1$ is representing signal point 'A' in figure 14.31. In the next bit interval, if $b_o(t) = -1$, then signal point will be 'D'. Otherwise, if $b_e(t)$ changes its level [i.e., $b_e(t) = -1$], then next signal point will be 'B'. Hence, from signal point 'A', then next signal points will be either 'D' or 'B'.

14.13.3.1. Distance Between Signal Points

As a matter of fact, the ability to determine a bit without error is measured by the distance between two nearest possible signal points in the signal space. Such points differed in signal bit. For example, signal points 'A' and 'B' are two nearest points since they differ by a signal bit $b_e(t)$. As 'A' and 'B' become closer to each other, the possibility of error increases. Therefore, this distance must be as large as possible. This distance is denoted by 'd'. In figure 14.31, the distance between signal points 'A' and 'B' can be given by,

$$d^2 = \left(\sqrt{E_s}\right)^2 + \left(\sqrt{E_s}\right)^2 = \sqrt{2E_s} \quad ...(14.71)$$

or

$$d = 2\sqrt{P_sT_b} = 2\sqrt{E_b} \quad ...(14.72)$$

Note: *If we compare this distance with the distance of BPSK signals, then this shows that the distance for QPSK is the same as that for BPSK. Because, this distance represents noise immunity of the system, it shows that noise immunities of BPSK and QPSK are same.*

14.13.4. Spectrum of QPSK Signal

The input sequence $b(t)$ is of bit duration T_b. Also, it is a NRZ bipolar waveform. Recall, the power spectral density of such waveform can be given as,

$$S(f) = V_b^2\,T_b\left[\frac{\sin(\pi f T_b)}{\pi f\,T_b}\right]^2$$

Also, $V_b = \sqrt{P_s}$, then this equation becomes,

$$S(f) = P_sT_b\left[\frac{\sin(\pi f\,T_b)}{\pi\,f\,T_b}\right]^2 \quad ...(14.73)$$

This equation gives power spectral density (psd) of signal $b(t)$. This signal is divided into $b_e(t)$ and $b_0(t)$ each of bit period $2T_b$. If we consider that symbols 1 and 0 are equally likely, then we can write power spectral densities (psds) of $b_e(t)$ and $b_0(t)$ as,

$$S_e(f) = P_sT_s\left[\frac{\sin(\pi fT_s)}{\pi fT_s}\right]^2 \quad ...(14.74)$$

$$S_Q(f) = P_s T_s \left[\frac{\sin(\pi f T_s)}{\pi f T_s} \right]^2 \quad ...(14.7\ldots)$$

and

In these two equations, we have just replaced T_b by T_s and T_s is the period of bit in $b_I(t)$ and $b_Q(t)$. Because, inphase and quadrature components $[b_e(t)$ and $b_Q(t)]$ are statistically independent, the baseband power spectral density of QPSK signal equals the sum of the individual power spectral densities of $b_e(t)$ and $b_Q(t)$ *i.e.*,

$$S_B(f) = S_e(f) + S_Q(f)$$

or

$$S_B(f) = 2P_s T_s \left[\frac{\sin(\pi f T_s)}{\pi f T_s} \right]^2 \quad ...(14.7\ldots)$$

This equation gives baseband power spectral density of QPSK signal. Upon modulation of carrier of frequency f_c, the spectral density given by above equation is shifted at $\pm f_c$. The plots of power spectral density of QPSK will be similar to that BPSK.

14.13.5. Bandwidth of QPSK Signal

We have observed that the bandwidth of BPSK signal is equal of $2f_b$. Here, $T_b = \frac{1}{f_b}$ is the one bit period. In QPSK, the two waveforms $b_e(t)$ and $b_Q(t)$ form the baseband signals. One bit period for both of these signals is equal to $2T_b$. Therefore, bandwidth of QPSK signal will be

$$BW = 2 \times \frac{1}{2T_b} = f_b \quad ...(14.77)$$

Hence, the bandwidth of QPSK signal is half of the bandwidth of BPSK signal. Earlier, we have observed that noise immunity of QPSK and BPSK is same. This shows that inspite of the reduction in bandwidth in QPSK, the noise immunity remains same as compared to BPSK. *BW* of QPSK can also be obtained by plotting equation (14.71) as shown in figure 14.32 below.

BW = Highest frequency – Lowest frequency in main lobe

$$BW = \frac{1}{T_s} - \left(-\frac{1}{T_s} \right) \text{ since carrier frequency } f_c \text{ cancels our}$$

$$BW = \frac{2}{T_s}$$

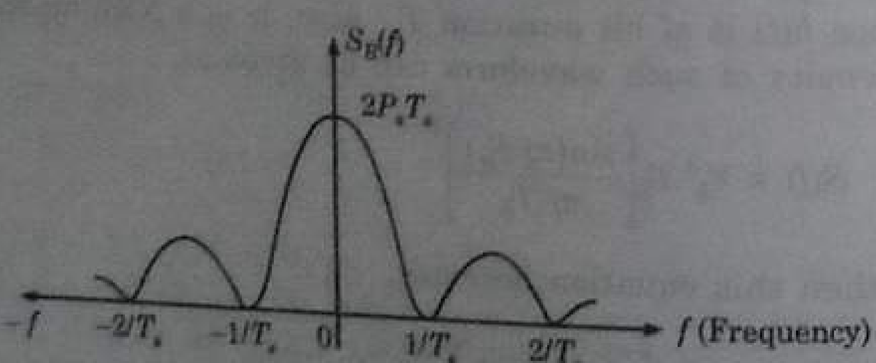


Fig. 14.32. *Plot of power spectral density (psd) of QPSK signal.*

We know that $T_s = 2T_b$

or $BW = \frac{2}{2T_b} = \frac{1}{T_b}$

or $BW = f_b \quad ...(14.78)$

14.13.6. Probability of Error of QPSK System

Observe figure 14.31 carefully. Between signal phasors 'OA' and 'OB', the axis $\phi_1(t)$ can be called decision boundary. Let us say signal vector 'OB' is present, but because of imperfect phase synchronization, it is detected as 'OA'. For this to happen, the phase shift must be at least $\frac{\pi}{4}$. The same thing can happen in case of other phasors also.

There are two correlators in the QPSK receiver. One correlator is used to detect even bits and other detects odd bits. Thus, any correlator can make a mistake if phase shift of $\frac{\pi}{4}$ occurs in the corresponding carrier. Therefore, the probability that correlator 1 or correlator 2 will make a mistake is given as,

$$P'_1(e) = P'_2(e) = \frac{1}{2} \, erfc \sqrt{\frac{E \cos^2 \theta}{N_0}}$$

This equation has been written because each correlator is independent PSK receiver. 'E_b' is replaced by 'E' in above equation, since energy of one bit in $b_e(t)$ or $b_Q(t)$ is E (also written as E_s). Putting phase shift $\theta = \frac{\pi}{4}$ in above equation we obtain,

$$P'_1(e) = P'_2(e) = \frac{1}{2} \, erfc \sqrt{\frac{E \cos^2 \frac{\pi}{4}}{N_0}}$$

or

$$P'_1(e) = \frac{1}{2} \, erfc \sqrt{\frac{E}{2N_0}}$$

We can verify the above relation for BPSK as under:

We know that $E = E_s = P_s T_s$

$T_s = 2T_b$

and $E = P_s 2T_b = 2P_s T_b$

Therefore, $E_b = P_s T_b$

$E = 2E_b$

By using the above relation, we obtain error probability of BPSK *i.e.*,

$$P'_1(e) = P'_2(e) = \frac{1}{2} \, erfc \sqrt{\frac{2E_b}{2N_0}}$$

or

$$P'_1(e) = \frac{1}{2} \, erfc \sqrt{\frac{E_b}{N_0}}$$

The above equation gives bit error probability of QPSK. Thus, bit error probability of QPSK and BPSK is same.

The probability $P(e)$ that the QPSK receiver will correctly detect the transmitted signal is equal to product of probabilities that both correlator 1 and correlator 2 will receive their bits correctly. The probability of correct reception of correlators 1 and 2 is,

$$P'_1(e) = 1 - P_1(e) \quad \text{and} \quad P_2(e) = 1 - P'_2(e)$$

$$\therefore \quad P(e) = P'_1(e) \times P'_2(e)$$

$$= [1 - P'_1(e)] \times [1 - P'_2(e)]$$

$$= 1 - P'_1(e) - P'_2(e) + P'_1(e) \times P'_2(e)$$

$P'_1(e) = P'_2(e)$ we can write above equation as,

$\therefore$ $P(e) = 1 - 2P'(e) + P'^2(e)$

$P'_1(e) = P'_2(e) = P'^2(e).$

Here,

The term $P'^2(e)$ will be very very small and can be neglected. Hence,

$$P(e) = 1 - 2P'(e)$$

$\therefore$ Probability of error of QPSK system is,

$$P(e) = 1 - P'(e)$$

or $P(e) = 1 - 1 + 2P'(e)$

or $P(e) = 2P'(e)$

Putting value of $P'(e) = \frac{1}{2} erfc\sqrt{\frac{E}{2N_o}}$, we get

$$P(e) = erfc\sqrt{\frac{E}{2N_o}} = erfc\sqrt{\frac{E_b}{N_o}} \quad \text{(since } E = 2E_b\text{)}$$

The individual probabilities $P'_1(e)$ and $P'_2(e)$ correlators are some times called as bit error probabilities or Bit Error Rate (BER). Thus bit error rate of QPSK is given as,

$$BER = \frac{1}{2} erfc\sqrt{\frac{E_b}{N_o}}$$

14.13.7. Advantages of QPSK

QPSK has some certain advantages as compared to BPSK and DPSK as under:

(i) For the same bit error rate, the bandwidth required by QPSK is reduced to half as compared to BPSK.

(ii) Because of reduced bandwidth, the information transmission rate of QPSK is higher.

14.14. Minimum Shift Keying (MSK)

(U.P. Tech., Sem., Examination 2003-2004)

We have discussed QPSK technique in last article. The bandwidth requirement of QPSK is high. Filters or other methods can overcome these problems, but they have other side effects. For example, filters alter the amplitude of the waveform.

MSK overcomes these problems. In MSK, the output waveform is continuous in phase hence there are no abrupt changes in amplitude. The sidelobes of MSK are very small hence bandpass filtering is not required to avoid interchannel interference. Figure 14.33 shows the waveform of MSK. The binary bit sequence is shown at the top. Figure 14.33(*a*) shows the corresponding NRZ waveform $b(t)$. From $b(t)$, two waveforms are generated for odd and even bits. $b_o(t)$ represents odd bits and $b_e(t)$ represents even bits. Figure 14.33(*b*) and (*c*) shows the waveform of $b_o(t)$ and $b_e(t)$. As shown in those waveforms b_1, b_3, b_5 etc. are represented by odd waveform *i.e.*, $b_o(t)$.

The duration of each bit in $b_o(t)$ or $b_e(t)$ is $2T_b$, whereas it is T_b in $b(t)$ *i.e.*,

$$T_s = 2T_b \qquad \text{...(14.79)}$$

The waveforms $b_o(t)$ and $b_e(t)$ have an offset of T_b. This offset is essential in MSK. Two waveforms $\sin 2\pi(t/4T_b)$ and $\cos 2\pi(t/4T_b)$ are generated as shown in figure 14.33(*d*). The waveform of $\sin 2\pi(t/4T_b)$ passes through zero at the end of symbol time in $b_o(t)$. Hence, one symbol duration of $b_o(t)$ consists of complete half cycle of $\cos 2\pi(t/4T_b)$. This means that similarly one symbol duration of $b_e(t)$ contains half cycle of $\sin 2\pi(t/T_b)$. Thus there is a phase shift of 'T_b' in sine and cosine waveforms. $b_e(t)$ is multiplied by $\sin 2\pi(t/4T_b)$ and $b_o(t)$ is multiplied by $\cos 2\pi(t/4T_b)$. These product waveforms are shown in figure 14.33 (*e*) and (*f*). The transmitted MSK signal is represented as under:

$$s(t) = \sqrt{2P_s}\,[b_e(t)\sin(2\pi t/4T_b)]\cos(2\pi f_c t) + \sqrt{2P_s}\,[b_o(t)\cos(2\pi t/4T_b)]\sin(2\pi f_c t) \qquad \text{...(14.80)}$$

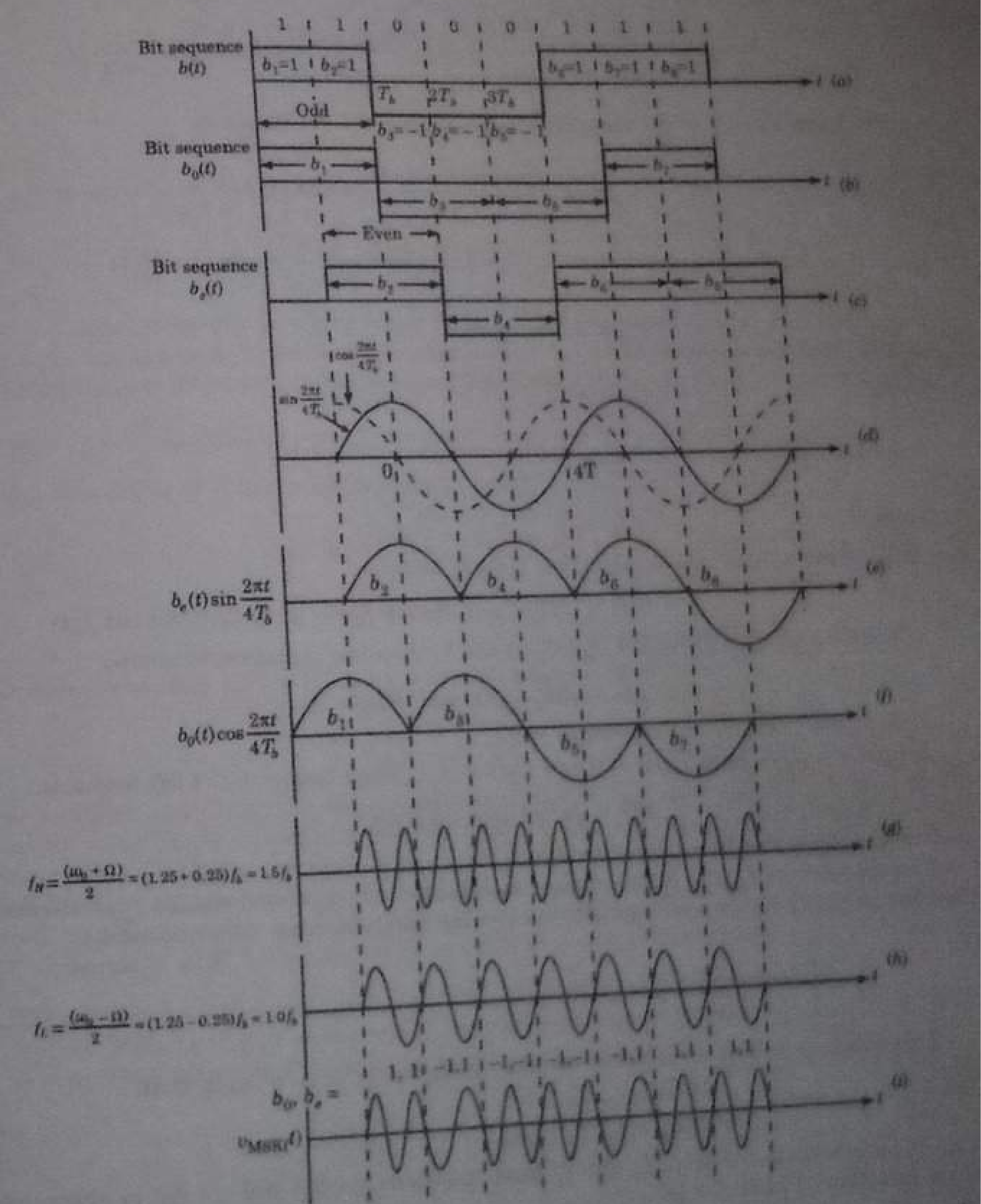


Fig. 14.33. *(a) Bipolar NRZ waveform representing bit sequence (b) Odd bit sequence waveforms b_o (c) Even bit sequence waveform $b_e(t)$ (d) Waveforms of frequency $f_b/4$ used for smoothing of $b_e(t)$ and b_o (e) Modulating waveform of even sequence (f) Modulating waveform of odd sequence (g) Waveform of frequency f_H (h) Waveform of frequency f_L (i) MSK waveform.*

This means that the product ... $(2\pi t/4T_b)$ modulat...
quadrature carriers of frequency f_c. We can write last equation as,

$$s(t) = \sqrt{2P_s}\left[\frac{b_o(t)+b_e(t)}{2}\right]\sin 2\pi\left(f_c+\frac{1}{4T_b}\right)t$$

$$+\sqrt{2P_s}\left[\frac{b_0(t)-b_e(t)}{2}\right]\sin 2\pi\left(f_c-\frac{1}{4T_b}\right)t \quad ...(14$$

We know that $f_b = \frac{1}{T_b}$, then the last equation (14.81) becomes,

$$s(t) = \sqrt{2P_s}\left[\frac{b_0(t)+b_e(t)}{2}\right]\sin 2\pi\left(f_c+\frac{f_b}{4}\right)t$$

$$+\sqrt{2P_s}\left[\frac{b_0(t)-b_e(t)}{2}\right]\sin 2\pi\left(f_c-\frac{f_b}{4}\right)t \quad ...(14$$

Let $$C_H(t) = \frac{b_0(t)+b_e(t)}{2}$$

and $$C_L(t) = \frac{b_0(t)+b_e(t)}{2}$$

and let $$f_H = f_c+\frac{f_b}{4} \quad ...(14$$

and $$f_L = f_c-\frac{f_b}{4} \quad ...(14$$

with these substitutions, equation (14.82) becomes,

$$s(t) = \sqrt{2P_s}\,C_H(t)\sin(2\pi f_H t)+\sqrt{2P_s}\,C_L(t)\sin(2\pi f_L t) \quad ...(14$$

If $b_0(t) = b_e(t)$ then $C_L(t) = 0$ and $C_H(t) = \pm 1$, then last equation becomes,

$$s(t) = \sqrt{2P_s}\,C_H(t)\sin(2\pi f_H t) \quad ...(14.8$$

Hence, the transmitted frequency is f_H.

If $b_0(t) = -b_e(t)$, then $C_H(t) = 0$ and $C_L(t) = \pm 1$. Then equation (14.86) becomes,

$$s(t) = \sqrt{2P_s}\,C_H(t)\sin(2\pi f_L t) \quad ...(14.8$$

Hence, the transmitted frequency is f_L.

The frequencies f_H and f_L are chosen such that $\cos(2\pi f_H t)$ and $\sin(2\pi f_L t)$ are orthogonal over the interval T_b. For orthogonality following relation must be satisfied *i.e.*,

$$\int_0^{T_b}\sin(2\pi f_H t)\sin(2\pi f_L t)\,dt = 0 \quad ...(14.88)$$

This relation will be satisfied if we have integers '*m*' and '*n*' such that,

$$2\pi(f_H - f_L)\,T_b = n\pi \quad ...(14.89)$$

and

$$2\pi(f_H + f_L)\,T_b = m\pi \quad ...(14.90)$$

Let us substitue values of f_H and f_L from equations (14.81) and (14.82) in above relations. From equation (14.87), we get

$$2\pi\left(f_c+\frac{f_b}{4}-f_c+\frac{f_b}{4}\right)T_b = n\pi$$

or

$$f_b T_b = n$$

or

$$f_b \times \frac{1}{f_b} = n \quad \Rightarrow \quad n = 1 \quad ...(14$$

Similarly from equation (14.90), we get

$$2\pi\left(f_c+\frac{f_b}{4}+f_c-\frac{f_b}{4}\right)T_b = m\pi$$

or

$$4f_c T_b = m$$

or

$$4f_c \times \frac{1}{f_b} = m \quad \Rightarrow \quad f_c = \frac{m}{4}f_b \quad ...(1$$

with $n = 1$ in equation (14.89), we get

$$2\pi(f_H - f_L)\,T_b = 1\times\pi$$

or

$$(f_H - f_L) = \frac{1}{2T_b} = \frac{f_b}{2} \quad ...(1$$

Here $n = 1$ means the difference between f_H and f_L is minimum and at the same (MSK) they are orthogonal. Therefore, this technique is called **minimum shift k** **(MSK)**. This minimum difference is given by equation (14.93) above. From equation (1

we know that $f_c = \frac{m}{4}f_b$. This shows that carrier frequency 'f_c' is integer multiple

Substituting, this value of f_c in equation (14.83), we get

$$f_H = f_c+\frac{f_b}{4} = m\frac{f_b}{4}+\frac{f_b}{4}$$

or

$$f_H = (m+1)\frac{f_b}{4}$$

Similarly, substituting $f_b = m\frac{f_b}{4}$ in equation (14.84), we get

$$f_L = (m-1)\frac{f_b}{4}$$

Figure 14.32(*g*) and (*h*) shows the waveforms of $\sin(2\pi f_H t)$ and $\sin(2\pi f_L t)$. F waveforms $m = 5$. Using equations (14.94) and (14.90), f_H and f_L are calculat $m = 5$. Figure 14.33(*i*) shows the final MSK waveform. From equation (14.86), we kn if $b_0(t) = b_e(t)$, then transmitted waveform is of frequency f_H. And if $b_0(t) = -b_e(t)$ transmitted waveform is given by equation (14.82), which has frequency of f_L. Th that MSK is basically FSK with reduced bandwidth and continuous phase.

14.14.1. Signal Space Representation of MSK and Distance between the Signal Points (*i.e.*, Geometrical Representation of MSK)

Let us rearrange equation (14.80) as follows

$$s(t) = C_H(t)\sqrt{P_sT_s}\cdot\sqrt{\frac{2}{T_s}}\sin(2\pi f_H t) + C_L(t)\sqrt{P_sT_s}\cdot\sqrt{\frac{2}{T_s}}\sin(2\pi f_L t)$$

Here let, $$\phi_H(t) = \sqrt{2/T_s}\sin(2\pi f_H t)$$

$$\phi_L(t) = \sqrt{2/T_s}\sin(2\pi f_L t)$$

The carriers $\phi_H(t)$ and $\phi_L(t)$ are in quadrature. They are in quadrature beca frequencies are in quadrature. In QPSK the carriers are in quadrature because shift. Depending on the values of $C_H(t)$ and $C_L(t)$, there will be four signal point plane. This has been illustrated in figure 14.34. The distance of each signal point from origin is $\sqrt{P_s T_s}$.

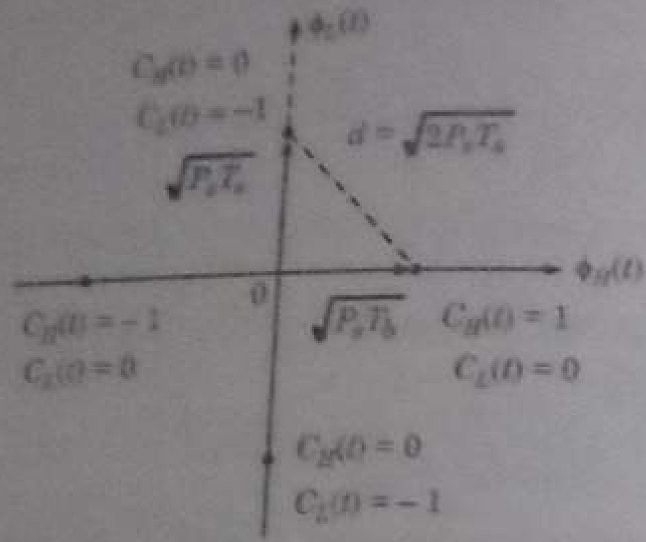


Fig. 14.34. *Geometrical (Signal Space) representation of MSK signals.*

Distance Between Signal Points

Since the points are symmetric, the distance between any two nearest points is [illegible] i.e.,

$$d^2 = \left(\sqrt{P_s T_s}\right)^2 + \left(\sqrt{P_s T_s}\right)^2$$

or $$d = \sqrt{2P_s T_s}$$ ...(14.[illegible])

or $$d = \sqrt{2E_s} \quad (\text{since } P_s\,T_s = E_s)$$ ...(14.[illegible])

or $$d = \sqrt{4E_b} \quad (\text{since } E_s = 2E_b) = 2\sqrt{E_b}\,)$$ ...(14.[illegible])

These relations give distance between signal points in MSK. This distance is same as QPSK.

14.14.2. Power Spectral Density (psd) and Bandwidth of MSK

Let us consider the baseband signal of equation (14.85). The waveform which modulates $\sin(2\pi f_c t)$ is,

$$p(t) = \sqrt{2P_s}\,[b_0(t)\cos(2\pi t/4T_b)]$$ ...(14.[illegible])

$$= \sqrt{2P_s}\, b_0(t)\cos(\pi f_b t/2)$$ ...(14.[illegible])

The power spectral density (psd) of above waveform is expressed as,

$$S_p(f) = \frac{32E_b}{\pi^2}\left[\frac{\cos(2\pi f T_b)}{1-(4fT_b)^2}\right]^2$$ ...(14.[illegible])

when this signal modulates the carrier 'f_c' then the total power spectral density (psd) of baseband signal is divided by '4' and is placed at $\pm f_c$, i.e.,

$$S(f) = \frac{8E_b}{\pi^2}\left\{\frac{\cos 2\pi(f-f_c)T_b}{1-[4(f-f_c)T_b]^2}\right\}^2 + \frac{8E_b}{\pi^2}\left\{\frac{\cos 2\pi(f+f_c)T_b}{1-[4(f+f_c)T_b]^2}\right\}^2$$ ...(14.[illegible])

The above equation gives power spectral density (psd) of MSK signal. Figure 14.34 shows the normalized spectral densities of MSK and QPSK. Normalization means maximum amplitudes of signals are scaled with respect to '1'.

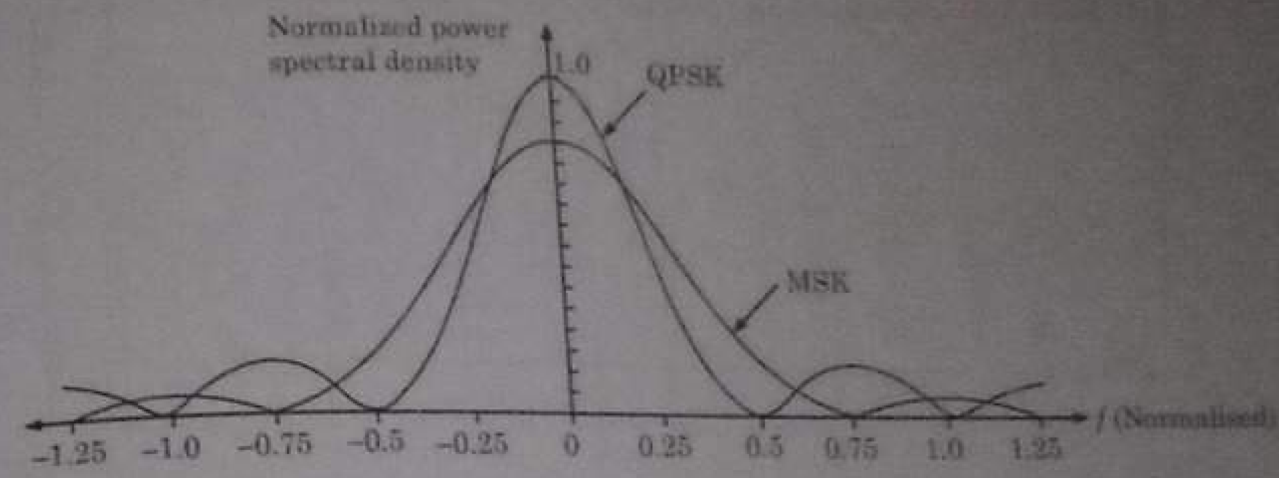


Fig. 14.35. *Power spectral densities (psd) of MSK and QPSK.*

The above plots show that the main lobe in MSK is wider than QPSK. The sides lobes in MSK are very small compared to QPSK.

Bandwidth Calculation of MSK

From figure 14.35, we observe that the width of main lobe in MSK is ± 0.75 i.e.,

$$fT_b = \pm\, 0.75$$

or $$f = \pm\, 0.75\, f_b$$

Hence, bandwidth will be equal to width of the main lobe i.e.,

$$BW = 0.75\, f_b - (-\,0.75\, f_b)$$
$$= 1.5\, f_b$$ ...(14.107)

Thus, the BW of MSK is higher than that of QPSK.

14.14.3. Generation of MSK

Figure 14.36 shows the block diagram of MSK transmitter. The two sinusoidal signals $\sin(2\pi f_c t)$ and $\cos(2\pi t/4T_b)$ are mixed (i.e., multiplied). The bandpass filters then pass only sum and difference components $f_c + \frac{f_b}{4}$ and $f_c - \frac{f_b}{4}$. The outputs of bandpass filters (BPFs) a[illegible]

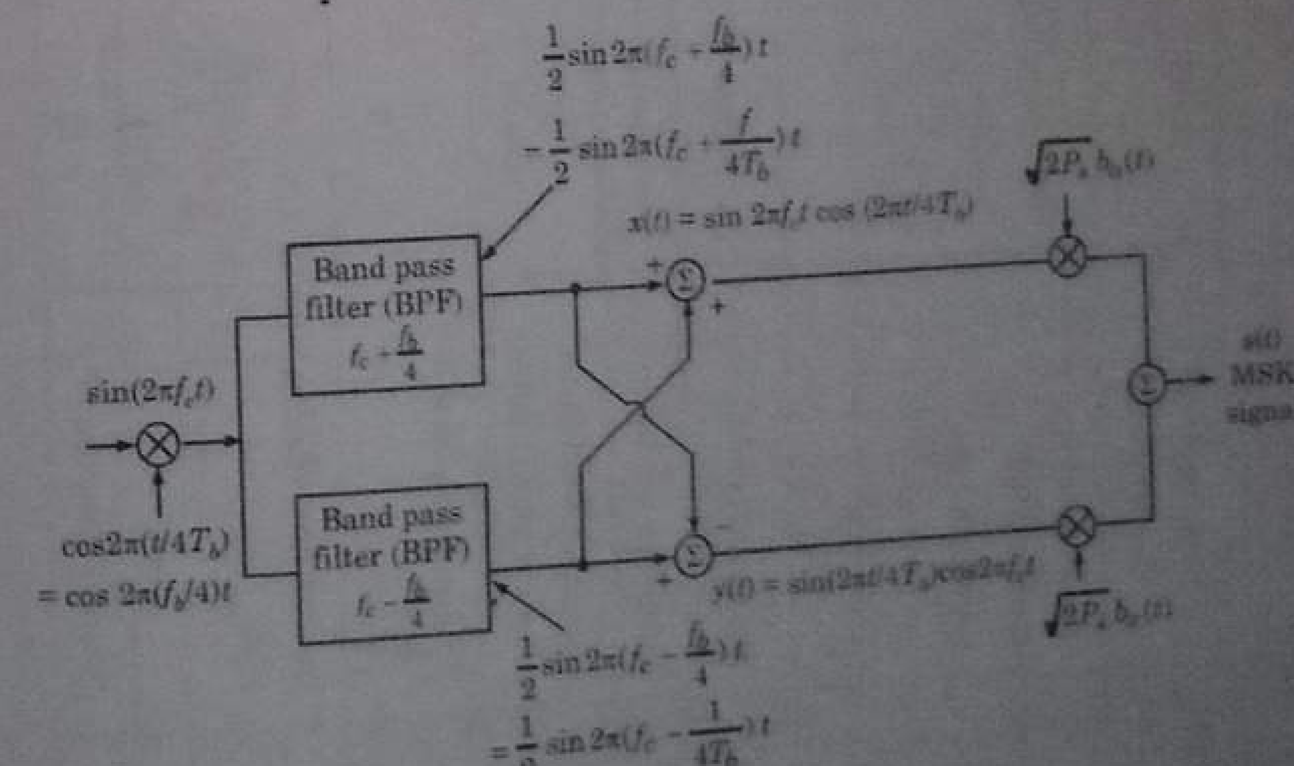


Fig. 14.36. *MSK transmitter block diagram.*

Table 14.5.

Sr. No.	*Parameter of Comparision*	*BPSK*	*DPSK*	*QPSK*	*M-ary PSK*
1	Equation of the transmitted signal $s(t)$	$s(t) = b(t)\sqrt{2P_s}$ $\cos(2\pi f_C t)$	$s(t) = b(t)\sqrt{2P_s}$ $\cos(2\pi f_C t)$ $b(t)$ differentially coded	$s(t) = \sqrt{2P_s}\cos$ $[2\pi f_C t + (2m+1)\frac{\pi}{4}]$ $m = 0, 1, 2, 3, \ldots$	$s(t) = \sqrt{2P_s}$ $\cos(2\pi f_C t + \phi_m)$ $\phi_m = (2m+1)\frac{p}{M}$ $m = 0, 1, 2, \ldots M-1$
2	Bits per symbol	One	one	Two	N
3	Detection method	Coherent	Non coherent	coherent	coherent
4	Minimum Euclidean distance signal points	$2\sqrt{E_b}$	–	$2\sqrt{E_b}$	$2\sqrt{E_s}\sin\frac{\pi}{M}$
5	Minimum Bandwidth (BW)	$2f_b$	f_b	f_b	$\frac{2f_b}{N}$
6	Probability of error $P(e)$	$\frac{1}{2}\,erfc\sqrt{\frac{E_b}{No}}$	$\frac{1}{2}e^{-E_b/No}$	$\text{erfc}\sqrt{\frac{E_b}{N}}$	$\leq \text{erfc}\left(\sqrt{\frac{E_s}{No}}\sin\frac{\pi}{M}\right)$
7	Symbol duration (T_s)	T_b	$2T_b$	$2T_b$	NT_b

(Contd.)

QASK	*BFSK*	*M-ary FSK*	*MSK*	*ASK*
$s(t) = k_1\sqrt{0.2P_s}$ $\cos 2\pi f_C t) + k_2\sqrt{0.2P_s}$ $\sin(2\pi f_C t)$ $k_1, k_2 = \pm 1$ or ± 3 for $M = 16$	$s(t) = \sqrt{2P_s}\cos$ $[(2\pi f_C + d(t)\,\Omega)f]$	$s(t) = \sqrt{2P_s}$ $\cos(2\pi f_1 t)$ $i = 1, 2, \ldots M$	$s(t) = b_0(t)\sqrt{2P_s}\sin 2\pi$ $\left[f_C + b_e(t)\,b_o(t)\frac{f_b}{4}\right]t$ $b_e(t), b_0(t)$ = odd/even sequence	$s(t) = 2\sqrt{2P_s}$ $\cos(2\pi f_C t)$ for symbol '1' = 0 for symbol '0'
N	one	N	Two	one
coherent	non coherent	non coherent	coherent	coherent
$\sqrt{0.4E_s}$ for $M = 16$	$\sqrt{2E_b}$	$\sqrt{2NE_b}$	$2\sqrt{E_b}$	$\sqrt{E_b}$
$\frac{2f_b}{N}$	$4f_b$	$\frac{2^{N+1}}{N}f_b$	$1.5f_b$	–
$\leq 2\,\text{erfc}\sqrt{\frac{0.4E_b}{No}}$ for $M = 16$	$\frac{1}{2}\text{erfc}\sqrt{\frac{E_b}{2No}}$	$\leq \frac{2^N - 1}{2}\text{erfc}\sqrt{\frac{NE_b}{2N_0}}$	$\text{erfc}\sqrt{\frac{E_b}{N_o}}$	$\frac{1}{2}\text{erfc}\sqrt{\frac{E_b}{4N_o}}$
NT_b	T_b	NT_b	$2T_b$	T_b

then added and subtracted such that two signals s(t) and y(t) are generated. Signal s(t) is multiplied by $\sqrt{2P_s}\, b_e(t)$ and y(t) is multiplied by $\sqrt{2P_s}\, b_o(t)$. The outputs of the multipliers are then added to give final MSK signal. Thus the block diagram of figure 14.36 is the direct implementation of equation (14.80).

14.14.4. Reception of MSK (i.e. Detection of MSK)

Figure 14.37 shows the block diagram of MSK receiver. MSK uses synchronous detection. The signals x(t) and y(t) are multiplied with the received MSK signal. Here x(t) and y(t) have same values as shown in transmitter block diagram of figure 14.37. The outputs of the multipliers are $b_e(t)$ and $b_o(t)$. The integrators integrate over the period of $2T_b$. For the upper correlator, the sampling switch samples output of integrator at $t = (2k + 1)T_b$. Then the decision device decides whether $b_e(t)$ is + 1 or − 1. Similarly, lower correlator output is $b_o(t)$. The outputs of two decision devices are staggered by T_b. The switch S_3 operates at $t = kT_b$ and multiplexes the two correlator outputs.

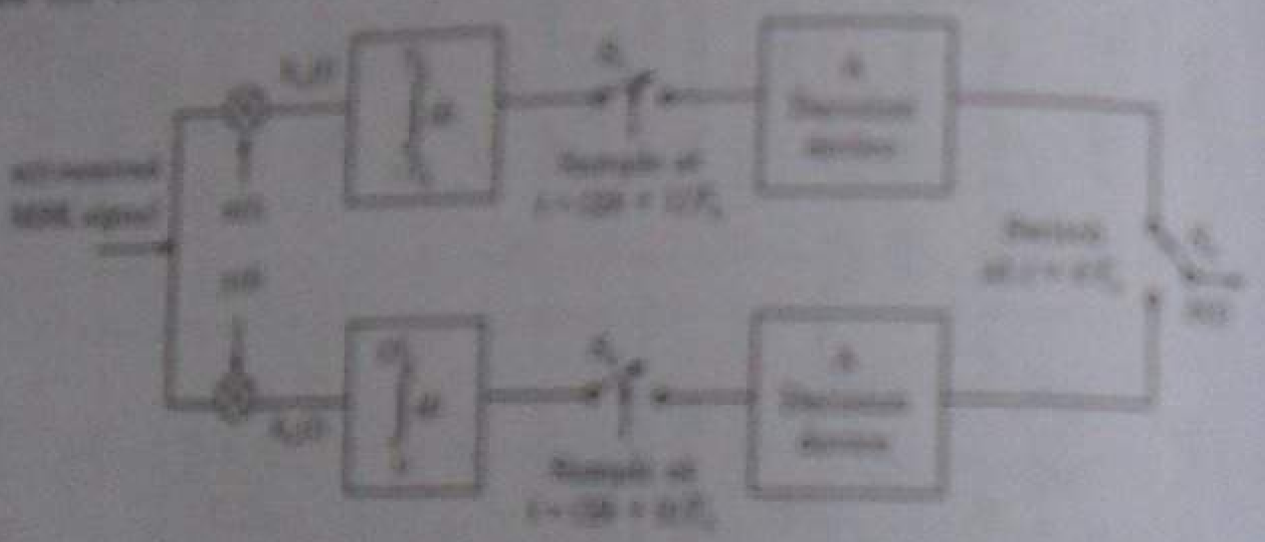


Fig. 14.37. MSK receiver block diagram.

14.14.5. Advantages and Disadvantages of MSK as Compared to QPSK

From the discussion of MSK, we can now compare the advantages of MSK over QPSK.

Advantages:

1. The MSK baseband waveforms are smoother compared to QPSK.
2. MSK signal have continuous phase in all the cases, whereas QPSK has abrupt phase shift of $\frac{\pi}{2}$ or π.
3. MSK waveform does not have amplitude variations, whereas QPSK signals have abrupt amplitude variations.
4. The main lobe of MSK is wider than that of QPSK. Main lobe of MSK contains around 99% of signal energy whereas QPSK main lobe contains around 90% signal energy.
5. Side lobes of MSK are smaller compared to that of QPSK. Hence, interchannel interference because of side lobes is significantly large in QPSK.
6. To avoid interchannel interference due to sidelobes, QPSK needs bandpass filtering whereas it is not required in MSK.
7. Bandpass filtering changes the amplitude waveform of QPSK because of abrupt changes in phase. This problem does not exist in MSK.

The distance between signal points is same in QPSK as well as in MSK. Hence, the probability of error is also same. However, there are some drawback of MSK.

(ii) Drawbacks

1. The bandwidth requirement of MSK is 1.5 f_b, whereas it is f_b in QPSK. Actually, this cannot be said serious drawback of MSK. Because power to bandwidth ratio of MSK is more. In fact, 99% of signal power can be transmitted within the bandwidth of 1.2 f_b in MSK. While QPSK needs around 8 f_b to transmit the same power.
2. The generation and detection of MSK is slightly complex. Because of incorrect synchronization, phase jitter can be present in MSK. This degrades the performance of MSK.

14.15 Comparison of Digital Modulation Techniques

Table 14.6 shows the comparison of various digital modulation techniques. They are compared on the basis of various parameters like bits transmitted per symbol, detection method, Euclidean distance, bandwidth, error probability, symbol duration etc. Various other important parameters like bandwidth efficiency, spectrum of transmitted signal etc., are not compared. QPSK, ASK have amplitude variations hence noise interference is more in these techniques. Normally PSK and FSK methods have less noise interference. M-ary techniques are more complex compared to binary techniques.

SUMMARY

1. Modulation is defined as the process by which some characteristics of a carrier is varied in accordance with a modulating signal.
2. In digital communications, the modulating signal consists of binary data or an M-ary encoded version of it.
3. The channel may be a telephone channel, microwave radio link, satellite channel or an optical fi... In digital communication, the modulation process involves switching or keying the amplitude, frequency or phase of the carrier in accordance with the input data.
4. There are three basic modulation techniques for the transmission of digital data. They are known as amplitude shift keying (ASK), frequency shift keying (FSK) and phase-shift keying (PSK) w... can be viewed as special cases of amplitude modulation frequency modulation and phase modulation respectively.
5. When we have to transmit a digital signal over a long distance, we need continuous-wave ... modulation. For this purpose, the transmission medium can be in form of radio, cable or other ... of channel. Also, a carrier signal having some frequency f_c is used for modulation. Then the modul... digital signal modulates some parameter like frequency, phase or amplitude of the carrier.
6. There is some deviation in carrier frequency f_c. This deviation is known as the bandwidth of ... channel. This means that the channel has to transmit some range or band of frequencies. Such ... of transmission is known as bandpass transmission and the communication channel is known ... bandpass channel.
7. When it is required to transmit digital signals on a bandpass channel, the amplitude, frequency ... phase of the sinusoidal carrier is varied in accordance with the incoming digital data. Since ... digital data is in discrete steps, the modulation of the bandpass sinusoidal carrier is also d... discrete steps. Due to this reason, this type of modulation (i.e., Digital modulation) is also k... as switching or signaling.
8. Because of constant amplitude of FSK and PSK, the effect of non-linearities, noise interference ... minimum on signal detection. However, these effects are more pronounced on ASK. Therefore ... and PSK are preferred over ASK.
9. In digital modulations, instead of transmitting one bit at a time, we transmit two or more ... simultaneously. This is known as M-ary transmission. This type of transmission results in r...

CHAPTER 15

Information Theory

Inside this Chapter

15.1. Introduction

As discussed earlier in chapter 1, the purpose of a communication system is to carry informatio- bearing baseband signals from one place to another place over a communication channel. In the last few chapters, we have discussed a number of modulation schemes to accomplish this purpose. But what is the meaning of the word "Information". To answer this question we need to discuss information theory. Infact, information theory is a branch of probability theory which may be applied to the study of the communication systems. This broad mathematical discipline has made fundamental contributions, not only to communications, but also computer science, statistical physics and probability and statistics.

Further, in the context of communications, information theory deals with mathematical modelling and analysis of a communication system rather than with physical sources and physical channels. As a matter of fact, Information theory was invented by communication scientists while they were studying the statistical structure of electronic communication equipments. When the communique is readily measurable, such as an electric current, the study of the communication system is relatively easy. But, when the communique is information, the study becomes rather difficult. How to define the measure for an amount of information? And also having described a suitable measure, how can it be applied to improve the communication of information? Information theory provides answers to all these questions. Thus, this chapter is devoted to a detailed discussion of information theory.

15.2. What is Information?

Before discussing the quantitative measure of information, let us review a basic concept about the amount of information in a message. Few messages produced by an information source contain more information than others. This may be best understood with the help of following example:

Consider you are planning a tour a city located in such an area where rain fall is very rare. To know about the weather forecast you will call the weather bureau and may receive one of the following information:

(i) It would be hot and sunny,

(ii) There would scattered rain,

(iii) There would be a cyclone with thunderstorm.

It may be observed that the amount of information received is clearly different for the three messages. The first message, just for instance, contains very little information because the weather in a desert city in summer is expected to be hot and sunny for maximum time. The second message forecasting a scattered rain contains some more information because it is not an event that occurs often. The forecast of a cyclonic storm contains even more information compared to the second message. This is because the third forecast is a rearest event in the city. Hence, on an conceptual basis the amount of information received from the knowledge of occurrence of a event may be related to the likelihood or probability of occurrence of that event. The message associated with the least likelihood event thus consists of maximum information. The above amount of information in a message depends only upon the uncertainty of the underlying event rather than its acutal content. Now, let us discuss few important concepts related to Information theory in the sections to follow.

15.3. Information Sources

An information source may be viewed as an object which produces an event, the outcome of which is selected at random according to a probability distribution. A practical source in a communication system is a device which produces messages, and it can be either analog or discrete. In this chapter, we deal mainly with the discrete sources since analog sources can be transformed to discrete sources through the use of sampling and quantization techniques, described in chapter 10. As a matter of fact, a discrete information source is a source which has only a finite set of symbols as possible outputs. The set of source symbols is called the **source alphabet**, and the elements of the set are called **symbols** or **letters.**

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

A discrete memoryless source (DMS) can be characterized by the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source.

15.4. Information Content of a Discrete Memoryless Source (DMS)

The amount of information contained in an event is closely related to its uncertainty. Mes... containing knowledge of high probability of occurrence convey relatively little information. ... note that if an event is certain (that is, the event occurs with probability 1), it conveys ... information. Thus, a mathematical measure of information should be a function of the probabi... of the outcome and should satisfy the following axioms:

(i) Information should be proportional to the uncertainty of an outcome.

(ii) Information contained in independent outcomes should add.

15.5. Information Content of a Symbol (i.e., Logarithmic Measure of Information)

Let us consider a discrete memoryless source (DMS) denoted by X and having alphabet $\{x_1, x_2, ... x_m\}$. The *information content* of a symbol x_i, denoted by $I(x_i)$ is defined by

$$I(x_i) = \log_b \frac{1}{P(x_i)} = -\log_b P(x_i) \quad ...(15.$$

where $P(x_i)$ is the probability of occurrence of symbol x_i.

Note that $I(x_i)$ satisfies the following properties:

$$I(x_i) = 0 \text{ for } P(x_i) = 1 \quad ...(15.$$

$$I(x_i) \geq 0 \quad ...(15.$$

$$I(x_i) > I(x_j) \text{ if } P(x_i) < P(x_j) \quad ...(15.$$

$$I(x_i x_j) = I(x_i) + I(x_j) \text{ if } x_i \text{ and } x_j \text{ are indepedent} \quad ...(15.$$

The unit of $I(x_i)$ is the bit (binary unit) if $b = 2$, Hartely or decit if $b = 10$, and nat (natural unit) if $b = e$. It is standard to use $b = 2$. Here the unit bit (abbreviated "b") is a measure of information content and is not to be confused with the term 'bit' meaning "binary digit". The conversion of these units to other units can be achieved by the following relationships.

$$\log_2 a = \frac{\ln a}{\ln 2} = \frac{\log a}{\log 2} \quad ...(15.$$

According to our concept, the information content or amount of information of a symbol x_i denoted by $I(x_i)$, must be inversely related to $P(x_i)$. Also from our intuitive concept $I(x_i)$ must staisfy the following requirements:

(i) $I(x_i)$ must approach 0 as $P(x_i)$ approaches infinity. For example, consider the message 'Sun will rise in the east'. This message does not contain any information since sun will rise in the east with probability 1.

(ii) The information content $I(x_i)$ must be a non-negative quantity since each message contains some information. In the worst case $I(x_i)$ can be equal to zero.

(iii) The information content of a message having higher probability of occurrence is less than the information content of a message having lower probability.

Now let us discuss few numerical examples to illustrate all the above concepts.

Example 15.1. A source produces one of four possible symbols during each interval having probabilities $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, $P(x_3) = P(x_4) = \frac{1}{8}$. Obtain the information content of each of these symbols.

Solution: We know that the information content of each symbol is given as

$$I(x_i) = \log_2 \frac{1}{P(x_i)}$$

Thus, we can write

$$I(x_1) = \log_2 \frac{1}{\frac{1}{2}} = \log_2 (2) = 1 \text{ bit}$$

$$I(x_2) = \log_2 \frac{1}{\frac{1}{4}} = \log_2 2^2 = 2 \text{ bits}$$

$$I(x_3) = \log_2 \frac{1}{\frac{1}{8}} = \log_2 2^3 = 3 \text{ bits}$$

$$I(x_4) = \log_2 \frac{1}{\frac{1}{8}} = 3 \text{ bits} \quad \text{Ans.}$$

Example 15.2. Calculate the amount of information if it is given that $P(x_i) = \frac{1}{4}$.

Solution: We know that amount of information is given as,

$$I(x_i) = \log_2 \frac{1}{P(x_i)} = \frac{\log_{10} \frac{1}{P(x_i)}}{\log_{10} 2}$$

Substituting given value of $P(x_i)$ in above equation, we get

$$I(x_i) = \frac{\log_{10} 4}{\log_{10} 2} = 2 \text{ bits} \quad \text{Ans.}$$

Example 15.3. Calculate the amount of information if binary digits (binits) occur with equal likelihood in a binary PCM system.

Solution: We know that in binary PCM, there are only two binary levels i.e., 1 or 0. Since, they occur with equal likelihood, their probabilities of occurrence would be.

$$P(x_1) \text{ ('0' level)} = P(x_2) \text{ ('1' level)} = \frac{1}{2}$$

Therefore, the amount of information content will be given as

$$I(x_1) = \log_2 \frac{1}{P(x_1)}$$

and $$I(x_2) = \log_2 \frac{1}{P(x_2)}$$

or $$I(x_1) = \log_2 2$$

and $$I(x_2) = \log_2 2$$

or $$I(x_1) = I(x_2) = \frac{\log_{10} 2}{\log_{10} 2} = 1 \text{ bit} \quad \text{Ans.}$$

Hence, the correct identification of binary digit (binit) in binary PCM carries 1 bit of information. **Ans.**

Example 15.4. In a binary PCM if '0' occur with probability $\frac{1}{4}$ and '1' occur with probability equal to $\frac{3}{4}$, then calculate the amount of information carried by each binit.

Solution: Here, given that the binit '0' has $P(x_1) = \frac{1}{4}$

$$[P(X,Y)] = \begin{bmatrix} \alpha & 0 \\ 0 & 1-\alpha \end{bmatrix} \begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}$$

or

$$[P(X,Y)] = \begin{bmatrix} \alpha(1-p) & \alpha p & 0 \\ 0 & (1-\alpha)p & (1-\alpha)(1-p) \end{bmatrix}$$

or

$$[P(X,Y)] = \begin{bmatrix} P(x_1,y_1) & P(x_1,y_2) & P(x_1,y_3) \\ P(x_2,y_1) & P(x_2,y_2) & P(x_2,y_3) \end{bmatrix}$$

In addition, from equations (15.24) and (15.26), we can calculate

$$H(Y) = -\sum_{j=1}^{3} P(y_j)\log_2 P(y_j)$$

$$= -\alpha(1-p)\log_2 \alpha(1-p) - p\log_2 p - (1-\alpha)(1-p)\log_2 [(1-\alpha)(1-p)]$$

$$= (1-p)[-\alpha \log_2 \alpha - (1-\alpha)\log_2 (1-\alpha)] - p\log_2 p - (1-p)\log_2 (1-p)$$

Also, we have $H(Y|X) = -\sum_{j=1}^{3}\sum_{i=1}^{2} P(x_i,y_j)\log_2 P(y_j|x_i)$

$$= -\alpha(1-p)\log_2 (1-p) - \alpha p\log_2 p - (1-\alpha)\,p\log_2 p - (1-\alpha)(1-p)\log_2 (1-p)$$

$$= -p\log_2 p - (1-p)\log_2 (1-p)$$

Thus, by equations (15.33) and (15.57), we have

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= (1-p)[-\alpha\log_2 \alpha - (1-\alpha)\log_2 (1-\alpha)] = (1-p)H(X)$$

And by equations (15.35) and (15.58), we have

$$C_s = \max_{\{P(x_i)\}} I(X;Y) = \max_{\{P(x_i)\}} (1-p)H(X) = (1-p)\max_{\{P(x_i)\}} H(X) = 1-p \quad \text{Ans.}$$

15.13. Entropy Relations for a Continuous Channel

In a continuous channel, an information source produces a continuous signal $x(t)$. The set of possible signals is considered as an ensemble of waveforms generated by some ergodic random process. It is further assumed that $x(t)$ has a finite bandwidth so that $x(t)$ is completely characterized by its periodic sample values. Hence, at any sampling instant, the collection of possible sample values constitutes a continuous random variable X described by it probability density function $f_X(x)$.

The average amount of information per sample value of $x(t)$ (i.e., entropy of a continuous source) is measured by

$$H(X) = -\int_{-\infty}^{\infty} f_X(x)\log_2 f_X(x)\,dx \text{ bits/sample} \quad \text{...(15.45)}$$

The entropy $H(X)$ defined by equation (15.45) is known as the *differential entropy* of X.

Also, the average mutual information in a continuous channel is defined (by analogy with the discrete case) as

$$I(X;Y) = H(X) - H(X|Y)$$

or

$$I(X;Y) = H(Y) - H(Y|X)$$

where $$H(Y) = -\int_{-\infty}^{\infty} f_Y(y)\log_2 f_Y(y)\,dy \quad \text{...(15.46)}$$

$$H(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_X(x|y)\,dxdy \quad \text{...(15.47)}$$

$$H(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_Y(y|x)\,dxdy \quad \text{...(15.48)}$$

15.14. Capacity of an Additive White Gaussian Noise (AWGN) Channel: Shannon-Hartley Law

In an additive white Gaussian noise (AWGN) channel, the channel output Y is given by

$$Y = X + n \quad \text{...(15.48)}$$

where X is the channel input and n is an additive bandlimited white Gaussian noise with zero mean and variance σ^2.

The capacity C_s of an AWGN channel is given by

$$C_s = \max_{\{f_X(x)\}} I(X;Y)$$

or

$$C_s = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \text{ b/sample} \quad \text{...(15.49)}$$

where S/N is the signal-to-noise ratio at the channel output. If the channel bandwidth B Hz is fixed then the output $y(t)$ is also a bandlimited signal completely characterized by its periodic sample values taken at the Nyquist rate $2B$ samples/s.

Then the capacity C(b/s) of the AWGN channel is given by

$$C = 2B \times C_s = B\log_2\left(1+\frac{S}{N}\right) \text{ b/s} \quad \text{...(15.5}$$

Equation (15.50) is known as the **Shannon-Hartley law.**

The Shannon-Hartley law underscores the fundamental role of bandwidth and signal-to-noise ratio in communication. It also shows that we can exchange increased bandwidth for decreased signal power for a system with given capacity C.

15.15. Channel Capacity

We know that the bandwidth and the noise power place a restriction upon the rate of information that can be transmitted by a channel. It may be shown that in a channel which is disturbed by a white Gaussian noise, one can transmit information at a rate of C bits per second, where C is the the channel capacity and is expressed as

$$C = B\log_2\left(1+\frac{S}{N}\right) \quad \text{...(15.}$$

In this expression,

B = channel bandwidth in Hz

S = Signal power

N = Noise power

It may be noted that the expression (equation 15.50) for channel capacity is valid for white Ga... noise. However, for other types of noise, the expression is modified.

Proof: Let us present a proof of channel capacity formula based upon the assumption that if a sig... is mixed with noise, the signal amplitude can be recognized only within the root main square no... voltage. In other words, we can say that the uncertainty in recognizing the exact signal amplitude... equal to the root mean square noise voltage.

Again, let us assume that the average signal power and the noise power are S watts and... watts respectively. This means that the root mean square value of the received signal is $\sqrt{S+...}$ volts and the root mean square value of the noise voltage is $\sqrt{N}$ volts.

Now, we have to distinguish the received signal of the amplitude $\sqrt{S+N}$ volts in the presen... of the noise amplitude $\sqrt{N}$ volts.

As a matter of fact, the input signal variation of less than $\sqrt{N}$ volts will not be distinguishab... at the receiver end.

Therefore, the number of the distinct levels that can be distinguished without error can b... expressed as

$$M = \frac{\sqrt{S+N}}{\sqrt{N}} = \sqrt{1+\frac{S}{N}} \quad ...(15.5...)$$

Thus, equation (15.51) expresses the maximum value of M.

Now, the maximum amount of information carried by each pulse having $\sqrt{1+\frac{S}{N}}$ distinc... levels is given by

$$I = \log_2 \sqrt{1+\frac{S}{N}} = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \text{ bits} \quad ...(15.5...)$$

Now, after establishing expression in equation (8.15), we can determine the channel capacity... In fact, the channel capacity is the maximum amount of information that can be transmitted pe... second by a channel.

If a channel can transmit a maximum of K pulses per second, then, the channel capacity C is... given by

$$C = \frac{K}{2}\log_2\left(1+\frac{S}{N}\right) \text{ bits per second} \quad ...(15.5...)$$

Recall that for bandwidth requirements of PAM signals, it has been shown that a system of... bandwidth nf_m Hz can transmit $2n\,f_m$ independent pulses per second. Further, under these conditions... the received signal will yield the correct values of the amplitudes of the pulses but will not reproduce... the details of the pulse shapes.

Now, since, we are interested only in the pulse amplitudes and not their shapes, it is conclude... that a system with bandwidth B Hz can transmit a maximum of $2B$ pulses per second. Further... since, each pulse can carry a maximum information of $\frac{1}{2}\log_2\left(1+\frac{S}{N}\right)$ bits, it follows that a system... of bandwidth B can transmit the information at a following maximum rate:

$$C = B\log_2\left(1+\frac{S}{N}\right) \text{ bits per second} \quad ...(15.54)$$

Therefore, the channel capacity C is limited by the bandwidth of the channel (or system) and the noise signal.

For a noiseless channel, $N = 0$ and the channel capacity will be infinite. However, practically, N is always finite and therefore, the channel capacity is finite.*

The expression in equation (15.54) is also known as the **Hartley-Shannon law** and is treated as the central theorem of information theory.

From Hartley-Shannon law, it is obvious that the bandwidth and the signal power can be exchanged for one another. To transmit the information at a given rate, we may reduce the signal power transmitted provided that the bandwidth is increased correspondingly. In a similar manner, the bandwidth may be reduced if we have to increase the signal power.

As a matter of fact, the process of modulation is actually a means of effecting this exchange between the bandwidth and the signal-to-noise ratio.**

Note: *It may be noted that the channel capacity represents the maximum amount of information that can be transmitted by a channel per second. To achieve this rate of transmission, the information has to be processed properly or coded in the most efficient manner.*

15.16. Transmission of Continuous Signals

Now, let us further illustrate the Hartley-Shannon law for the exchange of bandwidth and signal-to-noise ratio by a continuous signal which is bandlimited to f_m Hz.

According to the sampling theorem, the information of a continuous-time signal, which i... bandlimited to f_m Hz, is completely specified by $2f_m$ samples per second.

Hence, to transmit the information of such a signal, it is necessary to transmit only the... discrete samples.

Theoretical, Aspect

Now, one important question arises. How much information does each sample contain...

It depends upon how many discrete levels or values the samples may assume. In fact, th... samples can assume any value and hence to transmit such samples, we require pulses capabl... assuming infinite levels. Clearly, the information carried by each sample is infinite bits.

Therefore, the information contained in a continuous bandlimited signal is infinite.

* This is true even if the bandwidth B is infinite. The noise signal is a white noise with a unif... power density spectrum over the entire frequency range. Therefore, as the bandwidth B is increa... N also increases and hence the channel capacity remains finite even if $B = \infty$.

If $\eta/2$ is the power density, then we have

$$N = \eta B, \text{ and}$$

$$C = B\log_2\left(1+\frac{S}{\eta B}\right)$$

and $$\lim_{B\to\infty} C = \frac{S}{\eta}\frac{\eta B}{S}\log_2\left(1+\frac{S}{\eta B}\right)$$

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

$$\lim_{x\to 0}\frac{1}{x}\log_2(1+x) = \log_2 e = 1.44$$

Therefore, we have

$$\lim_{B\to\infty} C = \frac{S}{\eta}\log_2 e = 1.44\frac{S}{\eta}$$

** The improvement in the signal-to-noise ratio in wideband FM and PCM can be properly under...

In the presence of noise*, the channel capacity is finite. Therefore, it is impossible to tra[nsmit] complete information in a bandlimited signal by a physical channel in the presence of noise**.

In the absence of noise, $N = 0$, the channel capacity is infinite and hence any desired sig[nal] can be transmitted.

It is quite obvious that it is impossible to transmit the complete information contained i[n] continuous signal unless the transmitted signal power is made infinite.

Because of presence of noise, there is always a certain amount of uncertainty in the rece[ived] signal. In fact, the transmission of complete information in a signal will mean a zero amount [of] uncertainty. Actually, the amount of uncertainty can be made arbitrarily small by increasing t[he] channel capacity***. However, it can never be made zero.

15.17. Uncertainty in the Transmission Process

As a matter of fact, the uncertainty is introduced in the process of transmission. Therefore, altho[ugh] it is possible to transmit the complete information in a continuous signal at the transmitter e[nd,] it is impossible to recover this infinite amount of information at the receiver end. The amount [of] information that can be recovered per second at the receiver is C bits per second where C is t[he] channel capacity.

In place of transmitting all of the information at the transmitter, we can approximate t[he] signal so that its information contents are reduced to C bits per second and transmit this approxima[ted] signal which has a finite information content.

Now, it will be possible to recover all of the information that has been transmitted. In fa[ct,] this happens in pulse code modulation (PCM) system.

Now, the question arises: How can we approximate a signal so that the approximated sign[al] has a finite information content per second?

In fact, this can be done by a process known as **quantization.**

For illustration, let us consider the continuous signal bandlimited to f_m Hz as shown [in] figure 15.14.

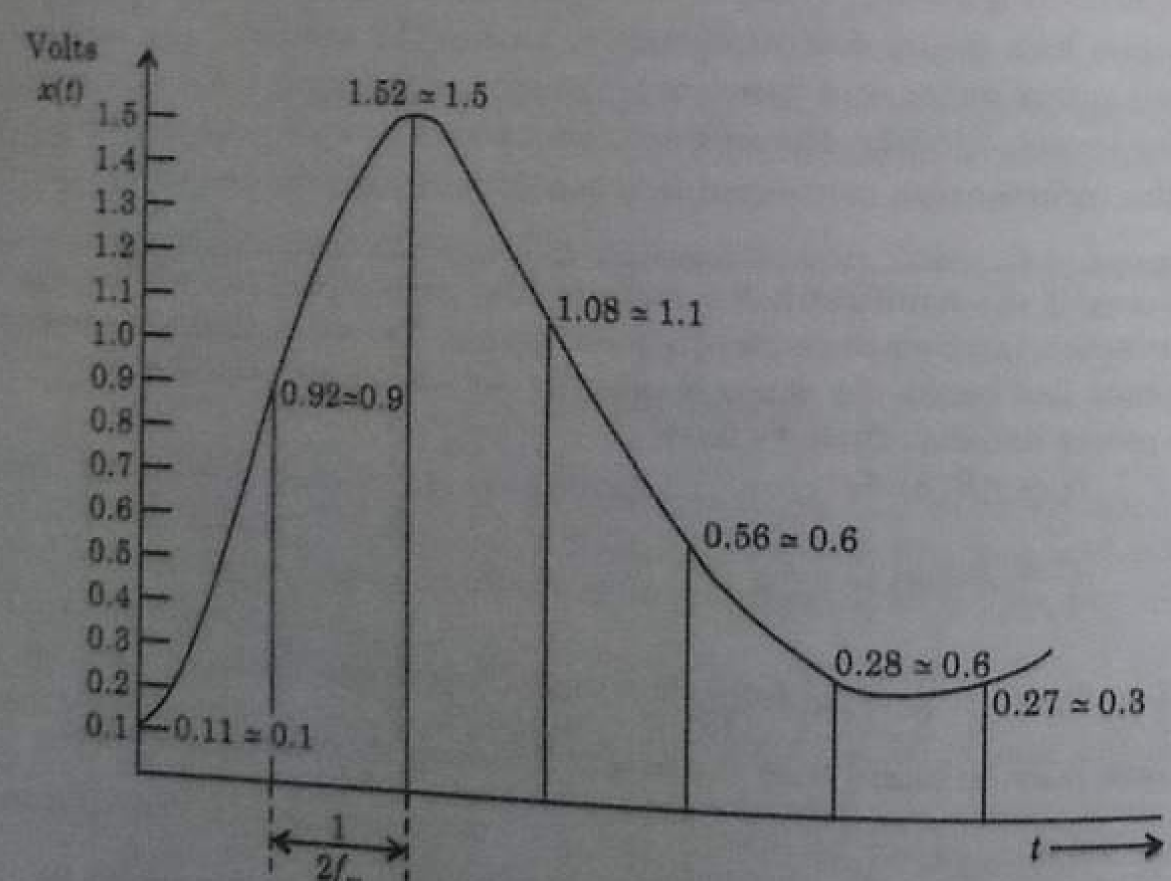


Fig. 15.14.

* That finite value of N.
** Existing over the same band.
*** By increasing the bandwidth and/or increa...

Now, to transmit the information in this signal, we require to transmit only $2f_m$ samples per second. Figure 15.14 also shows samples.

The samples can take any value, and to transmit them directly, we require pulses which can assume an infinite number of levels.

Therefore, instead of transmitting the exact values of these pulses, we round off the amplitudes to the nearest one of the finite number of permitted values.

In this example, all of the pulses are approximated to the nearest tenth of a volt.

It may be noted from the figure that each of the pulses transmitted assumes any one of the 16 levels, and thus, carries an amount of information of $\log_2 16 = 4$ bits. Also, since, there are $2f_m$ samples per second, the total information content of the approximated signal is $8 f_m$ bits per second.

It the channel capacity is greater than or equal to $8 f_m$ bits per second, all of the information that has been transmitted will be recovered completely without any uncertainty. This means that the received signal will be an exact replica of the approximated signal that was transmitted.

It can be shown that if the channel capacity is $8 f_m$ bits per second, the process of transmission does not introduce an additional degree of uncertainty.

Now, let us consider that we are using a channel of bandwidth f_m Hz to transmit these samples, then, since the channel capacity required will be $8 f_m$ bits per second, the required signal-to-noise power ratio will be given by

$$8 f_m = f_m \log_2\left(\frac{S+N}{N}\right)$$

Therefore, $\quad \dfrac{S+N}{N} = 256$

We have already discussed that the number of levels that can be distinguished at the receiver is $\sqrt{\dfrac{S+N}{N}}$.

It is obvious that in this case, the receiver can distinguish the 16 states without error. Thus, although the process of transmission introduces some noise in the desired signal, the levels are far enough apart to be distinguishable at the receiver.

In other words, we can say that a channel of the capacity of $8f_m$ bits per second can transmit information of $8 f_m$ bits per second virtually without error.

15.18. Exchange of Bandwidth for Signal-to-Noise Ratio

As a matter of fact, a given signal can be transmitted with a given amount of uncertainty by a channel of finite capacity. We have already discussed that a given channel capacity may be obtained by any number of combinations of bandwidth and signal power. Actually, it is possible to exchange one for the other. Now, let us discuss how such an exchange can be affected.

Again, let us consider the transmission of signal $x(t)$ shown in figure 15.14, we have already observed that if an uncertainty of 0.1 volt is tolerated, the information content of the signal is given by $8 f_m$ bits per second. Now, let us show that this information can be transmitted by various combinations of bandwidths and signal power.

One important and possible way of transmission is to send $2 f_m$ samples per second directly. Each of the samples can assume any of the 16 states. In this case, we must have a signal-to-noise ratio which allows us to distinguish 16 states.

Clearly, $\quad \dfrac{\sqrt{S+N}}{\sqrt{N}} = 16.$

Moreover, in order to transmit $2f_m$ pulses per second, we require a channel of bandwidth f_m Hz.

Therefore, the required channel capacity C is expressed as

$$C = f_m \log_2 \frac{S+N}{N}$$

or

$$C = f_m \log_2 (16)^2 = 8 f_m \text{ bits per second.}$$

Hence, the channel capacity C is exactly equal to the amount of information per second in the signal $x(t)$.

In other method of transmission, we may transmit the samples in figure 8.1 by quaternary pulse (pulses that can assume four states). Thus, it is obvious that we require a group of two quaternary pulses to transmit each sample that can assume 16 states.

In this case, the signal-to-noise ratio required at the receiver to distinguish pulses that assume four distinct states is $\frac{\sqrt{S+N}}{\sqrt{N}} = 4$.

Hence, in this mode of transmission, the required signal power is reduced. But, now, we have to transmit twice as many pulses per second *i.e.*, $4 f_m$ pulses per second.

Hence, the required bandwidth is $2 f_m$ Hz.

In this case, the channel capacity C will be

$$C = 2f_m \log_2 \frac{S+N}{N}$$

or

$$C = 2f_m \log_2 (4)^2 = 8f_m \text{ bits per second.}$$

Example 15.32. Find the differential entropy $H(X)$ of the uniformly distributed random variable X with the following probability density function (pdf):

$$f_X(x) = \begin{cases} \frac{1}{a} & \text{for} \quad 0 \le x \le a \\ 0 & \text{elsewhere} \end{cases}$$

for (*i*) $a = 1$, (*ii*) $a = 2$, and (*iii*) $a = \frac{1}{2}$.

Solution: We know that the differential entropy of X is given by

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x)\, dx \quad \text{bits/sample}$$

Making use of given pdf, we have

$$H(X) = -\int_0^a \frac{1}{a} \log_2 \frac{1}{a} dx = \log_2 a$$

Now, we have

(*i*) $a = 1, H(X) = \log_2 1 = 0$

(*ii*) $a = 2, H(X) = \log_2 2 = 1$

(*iii*) $a = \frac{1}{2}, H(X) = \log_2 \frac{1}{2} = -\log_2 2 = -1$

It may be noted that the differential entropy $H(X)$ is not an absolute measure of information.

Example 15.33. The differential entropy of a random variable X is defined by following equation:

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x)\, dx$$

Find the probability density function $f_X(x)$ for which $H(X)$ is maximum.

Solution: We know that $f_X(x)$ must satisfy the following two conditions:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \quad ...(i)$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx = \sigma^2 \quad ...(ii)$$

where μ is the mean of X and σ^2 it its variance. Since the problem is the maximization of $H(X)$ under constraints of equations (*i*) and (*ii*), therefore, we use the method of *Lagrange multipliers* as under:

First, we form the function:

$$G[f_X(x), \lambda_1, \lambda_2] = H(X) + \lambda_1 \left[\int_{-\infty}^{\infty} f_X(x) dx - 1\right] + \lambda_2 \left[\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx - \sigma^2\right]$$

$$= \int_{-\infty}^{\infty} [-f_X(x) \log_2 f_X(x) + \lambda_1 f_X(x) + \lambda_2 (x-\mu)^2 f_X(x)]\, dx - \lambda_1 - \lambda_2 \sigma^2$$

where the parameters λ_1 and λ_2 are the Lagrange multipliers. Then the maximization of $H(X)$ requires that

$$\frac{\partial G}{\partial f_X(x)} = -\log_2 f_X(x) - \log_2 e + \lambda_1 + \lambda_2 (x-\mu)^2 = 0$$

Thus,

$$\log_2 f_X(x) - \log_2 e + \lambda_1 + \lambda_2 (x-\mu)^2$$

or

$$\ln f_X(x) = -1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e} (x-\mu)^2$$

Hence, we obtain

$$f_X(x) = \exp\left[-1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e} (x-\mu)^2\right]$$

In view of the constraints of equations (*i*) and (*ii*), it is required that $\lambda_2 < 0$. Let

$$\exp\left(-1 + \frac{\lambda_1}{\log_2 e}\right) = a \quad \text{and} \quad \left(\frac{\lambda_2}{\log_2 e}\right) = -b^2$$

Then, equation (*v*) can be rewritten as

$$f_X(x) = a e^{-b^2 (x-\mu)^2}$$

Substituting equation (*vi*) into equations (*i*) and (*ii*), we get

$$a \int_{-\infty}^{\infty} e^{-b^2 (x-\mu)^2} dx = a \frac{\sqrt{\pi}}{b} = 1$$

$$a \int_{-\infty}^{\infty} (x-\mu)^2 e^{-b^2 (x-\mu)^2} dx = a \frac{\sqrt{\pi}}{2b^3} = \sigma^2$$

Solving equations (*vii*) and (*viii*) for a and b^2, we get

$$a = \frac{1}{\sqrt{2\pi}\sigma} \text{ and } b^2 = \frac{1}{2\sigma^2}$$

Substituting these values in equation (*vi*), we observe that the desired $f_X(x)$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$ **Hence proved.**

which is the probability density function of Gaussian random variable X of mean μ and variance σ^2.

Example 15.34. Show that the channel capacity of an ideal AWGN channel with infinite bandwidth is given by

$$C_\infty = \frac{1}{\ln 2} \cdot \frac{S}{\eta} \cong 1.44 \frac{S}{\eta} \text{ b/s} \quad \text{...(i)}$$

where S is the average signal power and $\eta/2$ is the power spectral density (psd) of white gaussian noise. *(U.P.S.C. I.E.S. Examination-1999)*

Solution: We know that the noise power N is given by $N = \eta B$.

Also, according to Shannon-Hartley law, the channel capacity is given by

$$C = B\log_2\left(1 + \frac{S}{N}\right) b/s$$

In this expression, substituting $N = \eta B$, we get

$$C = B\log_2\left(1 + \frac{S}{\eta B}\right)$$

Let $S/(\eta B) = \lambda$. Then, we write

$$C = \frac{S}{\eta\lambda}\log_2(1+\lambda) = \frac{1}{\ln 2}\frac{S}{\eta}\frac{\ln(1+\lambda)}{\lambda} \quad \text{...(ii)}$$

Now,

$$C_\infty = \lim_{B\to\infty} B\log_2\left(1 + \frac{S}{\eta B}\right)$$

$$C_\infty = \frac{1}{\ln 2} \cdot \frac{S}{\eta} \lim_{\lambda\to 0} \frac{\ln(1+\lambda)}{\lambda}$$

Since $\lim_{\lambda\to 0} [\ln(1+\lambda)]/\lambda = 1$, we get

$$C_\infty = \frac{1}{\ln 2} \cdot \frac{S}{\eta} \cong 1.44 \frac{S}{\eta} \text{ b/s}$$ **Hence proved.**

Note: *It may be noted that equation (i) can be used to estimate upper limits on the performance of any practical communication system whose transmission channel can be approximated by the AWGN channel.*

Example 15.35. Given an AWGN channel with 4 kHz bandwidth and the noise power spectral density $\eta/2 = 10^{-12}$ W/Hz. The signal power required at the receiver is 0.1 mW. Calculate the capacity of this channel.

Solution: Given that

$$B = 4000 \text{ Hz}$$

$$S = 0.1(10^{-3}) W$$

$$N = \eta B = 2(10^{-12})(4000)$$

$$N = 8(10^{-9}) \text{ W}$$

Thus,

$$\frac{S}{N} = \frac{0.1(10^{-3})}{8(10^{-9})} = 1.25(10^4)$$

And, by equation (14.50), we have

$$C = B\log_2\left(1 + \frac{S}{N}\right) = 4000 \log_2 [1 + 1.25(10^4)] = 54.44(10^3) \text{ b/s Ans.}$$

Example 15.36. An Analog signal having 4 kHz bandwidth is sampled at 1.25 times the Nyqu rate, and each sample is quantized into one of equally likely levels. Assume that the success samples are statistically independent.

(*i*) What is the information rate of this source?

(*ii*) Can the output of this source be transmitted without error over an AWGN chan with a bandwidth of 10 kHz and an S/N ratio of 20 dB?

(*iii*) Find the S/N ratio required for error-free transmission for part (*i*).

(*iv*) Find the bandwidth required for an AWGN channel for error-free transmission of output of this source if the S/N ratio is 20 dB? *(U.P.S.C. I.E.S. Examination, 1*

Solution: (*i*) Here, $f_m = 4(10^3)$Hz

We know that Nyquist rate f_s is given by

$$f_s = 2 f_m$$

$$\text{Nyquist rate} = 2 f_m = 8(10^3) \text{ samples/s}$$

Also, we have

$$r = 8(10^3)(1.25) = 10^4 \text{ samples/s}$$

Further, we know that entropy is expressed as

$$H(X) = -\sum_{i=1}^{m} P(x_i)\log_2 P(x_i) \text{ bits/symbol}$$

Here $P(x_i) = \frac{1}{256}$

Hence, $H(X) = \log_2 256 = 8$ bits/sample

The information rate R of the source is given by

$$R = rH(X) = 10^4(8) \text{ b/s} = 80 \text{ kb/s} \quad \textbf{Ans.}$$

(*ii*) Again, we know that channel capacity is given by

$$C = B\log_2\left(1 + \frac{S}{N}\right) \text{ bits/s}$$

Hence,

$$C = B\log_2\left(1 + \frac{S}{N}\right) = 10^4\log_2(1 + 10^2) = 66.6(10^3) \text{ b/s}$$

Here, since $R > C$, *error*-free transmission is not possible.

(*iii*) The required S/N ratio can be found by

$$C = 10^4\log_2\left(1 + \frac{S}{N}\right) \geq 8(10^4)$$

or $$\log_2\left(1 + \frac{S}{N}\right) \geq 8$$

or $$1 + \frac{S}{N} \geq 2^8 = 256$$

or $$\frac{S}{N} \geq 255 \; (= 24.1 \; dB) \quad \textbf{Ans.}$$

Thus, the required S/N ratio must be greater than or equal to 24.1 dB for error-free tra

(ii) The required bandwidth B can be found by

$$C = B\log_2(1+100) \geq 8(10^4)$$

or

$$B \geq \frac{8(10^4)}{\log_2(1+100)} \geq 1.2(10^4)\text{ Hz} \geq 12\text{ kHz}$$

and the required bandwidth of the channel must be greater than or equal to 12 kHz. Ans.

15.19. The Source Coding

A conversion of the output of a DMS into a sequence of binary symbols (*i.e.*, binary code word) is called **source coding.** The device that performs this conversion is called the **source encoder**, shown in figure 15.15.

An objective of source coding is to minimize the average bit rate required for representation of the source by reducing the redundancy of the information source.

15.19.1. The Code Length and Code Efficiency

Let X be a DMS with finite entropy $H(X)$ and an alphabet $\{x_1, \ldots, x_m\}$ with corresponding probabilities of occurrence $P(x_i)(i = 1,\ldots,m)$.

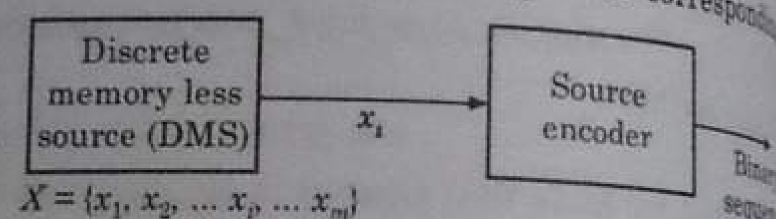


Fig. 15.15. *Block diagram for source Coding*

Let the binary codeword assigned to symbol x_i by the encoder have length n_i, measured in bits. The length of a codeword is the number of binary digits in the codeword. The average codeword length L, per source symbol is given by

$$L = \sum_{i=1}^{m} P(x_i)n_i \qquad \ldots(15.5\text{[illegible]})$$

The parameter L represents the average number of bits per source symbol used in the source coding process.

Also, the *code efficiency* η is defined as

$$\eta = \frac{L_{min}}{L} \qquad \ldots(15.5\text{[illegible]})$$

where L_{min} is the minimum possible value of L. When η approaches unity, the code is said to be *efficient.*

The code *redundancy* γ is defined as

$$\gamma = 1 - \eta \qquad \ldots(15.53)$$

15.19.2. The Source Coding Theorem

The source coding theorem states that for a DMS X, with entropy $H(X)$, the average codeword length L per symbol is bounded as

$$L \geq H(X) \qquad \ldots(15.54)$$

and further, L can be made as close to $H(X)$ as desired for some suitably chosen code.

Thus, with $L_{min} = H(X)$, the code efficiency can be rewritten as

$$\eta = \frac{H(X)}{L} \qquad \ldots(15.55)$$

15.19.3. Classification of Codes

Classification of codes is best illustrated by an example. Let us consider Table 15.2 where a source of size 4 has encoded in binary codes with symbol 0 and 1.

Table 15.2. Binary Codes

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

15.19.3.1. Fixed-Length Codes

A *fixed-length* code is one whose codeword length is fixed. Code 1 and code 2 of Table 15.2 are fixed-length with length 2.

15.19.3.2. Variable-Length Codes

A *variable-length* code is one whose codeword length is not fixed. All codes of Table 15.2 except codes 1 and 2 are variable-length codes.

15.19.3.3. Distinct codes

A code is *distinct* if each codeword is distinguishable from other codewords. All codes of Table (15.2) except code 1 are distinct codes – notice the codes for x_1 and x_3.

15.19.3.4. Prefix-Free Codes

A code in which no codeword can be formed by adding code symbols to another codeword is called a *prefix-free code.* Thus, in a prefix-free code, no codeword is a *prefix* of another. Codes 2 and 6 of Table 15.2 are prefix-free codes.

15.19.3.5. Uniquely Decodable Codes

A distinct code *uniquely decodable* if the original source sequence can be reconstructed perfectly from the encoded binary sequence. Note that code 3 of Table 15.2 is not a uniquely decodable code. For, example, the binary sequence 1001 may correspond to the source sequences $x_2x_3x_2$ or x_2x_1... A sufficient condition to ensure that a code is uniquely decodable is that no code word is a prefix of another. Thus, the prefix-free codes 2, 4, and 6 are uniquely decodable codes. Note that the prefix-free condition is not a necessary condition for unique decodability. For example, code 5 of Table 15.2 does not satisfy the prefix-free condition, and yet it is uniquely decodable since the bit 0 indicates the beginning of each codeword of the code.

15.19.3.6. Instantaneous Codes

A uniquely decodable code is called an *instantaneous code* if the end of any codeword is recognizable without examining subsequent code symbols. The instantaneous codes have the property previously mentioned that no codeword is a prefix of another codeword. For this reason, prefix-free codes are sometimes known as instantaneous codes.

15.19.3.7. Optimal Codes

A code is said to be *optimal* if it is instantaneous and has minimum average L for a given source with a given probability assuagement for the sourcesymbols.

15.19.4. The Kraft Inequality

Let X be a DMS with alphabet $\{x_i\}$ $(i = 1, 2, \ldots m)$. Assume that the length of the assigned binary codeword corresponding to x_i is n_i.

A necessary and sufficient condition for the existence of an instantaneous binary code

$$K = \sum_{i=1}^{m} 2^{-n_i} \leq 1$$

which is known as the **Kraft inequality.**

Note: It may be noted that Kraft inequality assures us of the existence of an instantaneously decodable code with codeword lengths that satisfy the inequality. But it does not show us how to obtain these codewords, nor does it say any code satisfies the inequality is automatically uniquely decodable.

Example 15.37. Consider a DMS X with two symbols x_1 and x_2 and $P(x_1) = 0.9$, $P(x_2) = 0.1$. Symbols x_1 and x_2 are encoded as follows (Table 15.3)

Table 15.3.

x_i	$P(x_i)$	Code
x_1	0.9	0
x_2	0.1	1

Find the efficiency η and the redundancy γ of this code.

Solution: We know that the average code length L per symbol is

$$L = \sum_{i=1}^{2} P(x_i)n_i = (0.9)(1) + (0.1)(1) = 1 \text{ b}$$

Using equation (15.7), we have

$$H(X) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i)$$

$$= -0.9 \log_2 0.9 - 0.1 \log_2 0.1 = 0.469 \text{ b/symbol}$$

Also, the code efficiency η is

$$\eta = \frac{H(X)}{L} = 0.469 = 46.9\%$$

And, the code redundancy γ is given by

$$\gamma = 1 - \eta = 0.531 = 53.1\% \text{ Ans.}$$

Example 15.38. The second-order extension of a DMS X, denoted by X^2, is formed by taking the source symbols two at a time. The coding of this extension has been shown in Table 15.4. Find the efficiency η and the redundancy γ of this extension code.

Solution:

Table 15.4.

a_i	$P(a_i)$	Code
$a_1 = x_1x_1$	0.81	0
$a_2 = x_1x_2$	0.09	10
$a_3 = x_2x_1$	0.09	110
$a_4 = x_2x_2$	0.01	111

We have $$L = \sum_{i=1}^{4} P(a_i)n_i = 0.81(1) + 0.09(2) + 0.09(3) + 0.01(3) = 1.29 \text{ b/symbol}$$

The entropy of the second-order extension of X, $H(X^2)$, is given by

$$H(X^2) = -\sum_{i=1}^{4} P(a_i)\log_2 P(a_i)$$

$$= -0.81 \log_2 0.81 - 0.09 \log_2 0.09 - 0.09 \log_2 0.09 - 0.01 \log_2 0.01$$

or $$H(X^2) = 0.938 \text{ b/symbol}$$

Therefore, the code efficiency η is

$$\eta = \frac{H(X^2)}{L} = \frac{0.938}{1.29} = 0.727 = 72.7\%$$

Also, the code redundancy γ will be

$$\gamma = 1 - \eta = 0.273 = 27.3\% \text{ Ans.}$$

Note that $H(X^2) = 2H(X)$.

Example 15.39. Consider a DMS X with symbols x_i, $i = 1, 2, 3, 4$. Table 15.5 lists four possible binary codes.

Table 15.5

x_i	Code A	Code B	Code C	Code D
x_1	00	0	0	0
x_2	01	10	11	100
x_3	10	11	100	110
x_4	11	110	110	111

(i) Show that all the codes except code B satisfy the Kraft inequality.

(ii) Show that codes A and D are uniquely decodable but codes B and C are not uniquely decodable.

Solution: We have

(i) For code A, $$n_1 = n_2 = n_3 = n_4 = 2$$

therefore, $$K = \sum_{i=1}^{4} 2^{-n_i} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$$

For code B, $$n_1 = 1, n_2 = n_3 = 2, n_4 = 3$$

therefore, $$K = \sum_{i=1}^{4} 2^{-n_i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} = 1\frac{1}{8} > 1$$

For code C, $$n_1 = 1, n_2 = 2, n_3 = n_4 = 3$$

therefore, $$K = \sum_{i=1}^{4} 2^{-n_i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = 1$$

For code D, $$n_1 = 1, n_2 = n_3 = n_4 = 3$$

therefore, $$K = \sum_{i=1}^{4} 2^{-n_i} = \frac{1}{2} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{7}{8} < 1$$

Thus, all codes except code B satisfy the Kraft inequality.

(ii) Codes A and D are prefix-free codes. They are, therefore, uniquely decodable. Code B does not satisfy the Kraft inequality, and it is not uniquely decodable. Although code C does satisfy the Kraft inequality, but it is not uniquely decodable. This can be seen by the following example: Given the binary sequence 0110110. This sequence may correspond to the source sequences $x_1x_2x_1x_4$ or $x_1x_4x_4$.

Example 15.40. Verify the following expression:

$$L \geq H(X)$$

where L is the average codeword length per symbol and $H(X)$ is the source entropy.

Solution: We know that

$$\sum_{i=1}^{m} P_i \log_2 \frac{Q_i}{P_i} \leq 0$$

where the equality holds only if $Q_i = P_i$.

Let
$$Q_i = \frac{2^{-n_i}}{K}$$

where
$$K = \sum_{i=1}^{m} 2^{-n_i}$$

Now, we have

$$\sum_{i=1}^{m} Q_i = \frac{1}{K}\sum_{i=1}^{m} 2^{-n_i} = 1$$

and
$$\sum_{i=1}^{m} P_i \log_2 \frac{2^{-n_i}}{KP_i} = \sum_{i=1}^{m} P_i\left(\log_2 \frac{1}{P_i} - n_i - \log_2 K\right)$$

$$= -\sum_{i=1}^{m} P_i \log_2 P_i - \sum_{i=1}^{m} P_i n_i - (\log_2 K)\sum_{i=1}^{m} P_i$$

$$= H(X) - L - \log_2 K \leq 0$$

From the Kraft inequlity (14.56), we have

$$\log_2 K \leq 0$$

Thus, $\quad H(X) - L \leq \log_2 K \leq 0$

or $\quad L \geq H(X)$

The equality holds when $K = 1$ and $P_i = Q_i$. **Hence proved.**

Example 15.41. Let X be a DMS with symbols x_i and corresponding probabilities $P(x_i)$ = [illegible] **$i = 1,2, \ldots m$. Show that for the optimum source encoding, we require that**

$$K = \sum_{i=1}^{m} 2^{-n_i} = 1$$

and
$$n_i = \log_2 \frac{1}{P_i} = I_i$$

where n_i is the length of the codeword corresponding to x_i and I_i is the information conte[nt] **of x_i.**

Solution: From the result of last problem, the optimum source encoding with $L = H(X)$ requi[res] $K = 1$ and $P_i = Q_i$. Thus equations using (ii) and (i) of last problem, we have

$$K = \sum_{i=1}^{m} 2^{-n_i} = 1 \quad \ldots(i)$$

and
$$P_i = Q_i = 2^{-n_i} \quad \ldots(ii)$$

Hence
$$n_i = -\log_2 P_i = \log_2 \frac{1}{P_i} = I_i$$

Note: *Note that equation (ii) implies the following commonsense principle.*

Symbols which occur with high probability should be assigned shorter codewords than symbols that occur with low probability. **Hence proved.**

Example 15.42. Consider a DMS X with symbols x_i and corresponding probabilities $P(x_i) = P_i$, $i = 1, 2, \ldots m$. Let n_i be the length of the codeword of x_i such that

$$\log_2 \frac{1}{P_i} \leq n_i \leq \log_2 \frac{1}{P_i} + 1 \quad \ldots(i)$$

Show that this relationship satisfies the Kraft inequality (15.56), and find the bound on K in equation (15.56).

Solution: Equation (*i*) can be rewritten as

$$-\log_2 P_i \leq n_i \leq -\log_2 P_i + 1 \quad \ldots(ii)$$

or $\quad \log_2 P_i \geq -n_i \geq \log_2 P_i - 1$

Then, $\quad 2\log_2 P_i \geq 2^{-n_i} \geq 2^{\log_2 P_i}\, 2^{-1}$

or
$$P_i \geq 2^{-n_i} \geq \frac{1}{2} P_i \quad \ldots(iii)$$

Thus,
$$\sum_{i=1}^{m} P_i \geq \sum_{i=1}^{m} 2^{-n_i} \geq \frac{1}{2}\sum_{i=1}^{m} P_i \quad \ldots(iv)$$

or
$$1 \geq \sum_{i=1}^{m} 2^{-n_i} \geq \frac{1}{2} \quad \ldots(v)$$

which indicates that the Kraft inequality (15.56) is satisfied, and the bound on K will be

$$\frac{1}{2} \leq K \leq 1 \quad \ldots(vi)$$

Hence proved.

Example 15.43. Consider a DMS X with symbols x_i and corresponding probabilities $P(x_i) = P_i$, i = 1, 2, ..., m. Show that a code constructed in agreement with equation (i) in last problem will satisfy the following relation:

$$H(X) \leq L \leq H(X) + 1 \quad \ldots(i)$$

where $H(X)$ is the source entropy and L is the average codeword length.

Solution: Multiplying equation (*ii*) in last problem by P_i and summing over i yields

$$-\sum_{i=1}^{m} P_i \log_2 P_i \leq \sum_{i=1}^{m} n_i P_i \leq \sum_{i=1}^{m} P_i(-\log_2 P_i + 1) \quad \ldots(ii)$$

Now,
$$\sum_{i=1}^{m} P_i(-\log_2 P_i + 1) = -\sum_{i=1}^{m} P_i \log_2 P_i + \sum_{i=1}^{m} P_i = H(X) + 1$$

Thus, equation (*ii*) reduces to

$$H(X) \leq L \leq H(X) + 1$$

Hence proved.

15.20. Entropy Coding

The design of a variable-length code such that its average codeword length approaches the entropy of DMS is often referred to as **entropy coding**. In this section, we present two examples of entropy coding.

15.30.1. Shannon-Fano Coding

An efficient code can be obtained by the following simple procedure, known as Shannon-Fano algorithm:

1. List the source symbols in order of decreasing probability.
2. Partition the set into two sets that are as close to equiprobables as possible, and assign 0 to the upper set 1 to the lower set.
3. Continue this process, each time partitioning the sets with as nearly equal probabilities as possible until further partitioning is not possible.

An example of Shannon-Fano encoding is shown in Table 15.6. Note in Shannon-Fano encoding the ambiguity may arise in the choice of approximately equiprobable sets.

Table 15.6. Shannon-Fano Encoding

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Step 4	Code
x_1	0.30	0	0			00
x_2	0.25	0	1			01
x_3	0.20	1	0			10
x_4	0.12	1	1	0		110
x_5	0.08	1	1	1	0	1110
x_6	0.05	1	1	1	1	1111

$$H(X) = 2.36 \text{ b/symbol}$$
$$L = 2.38 \text{ b/symbol}$$
$$\eta = H(X)/L = 0.99$$

(U.P. Tech., Sem. Examination

15.30.2. The Huffman Encoding

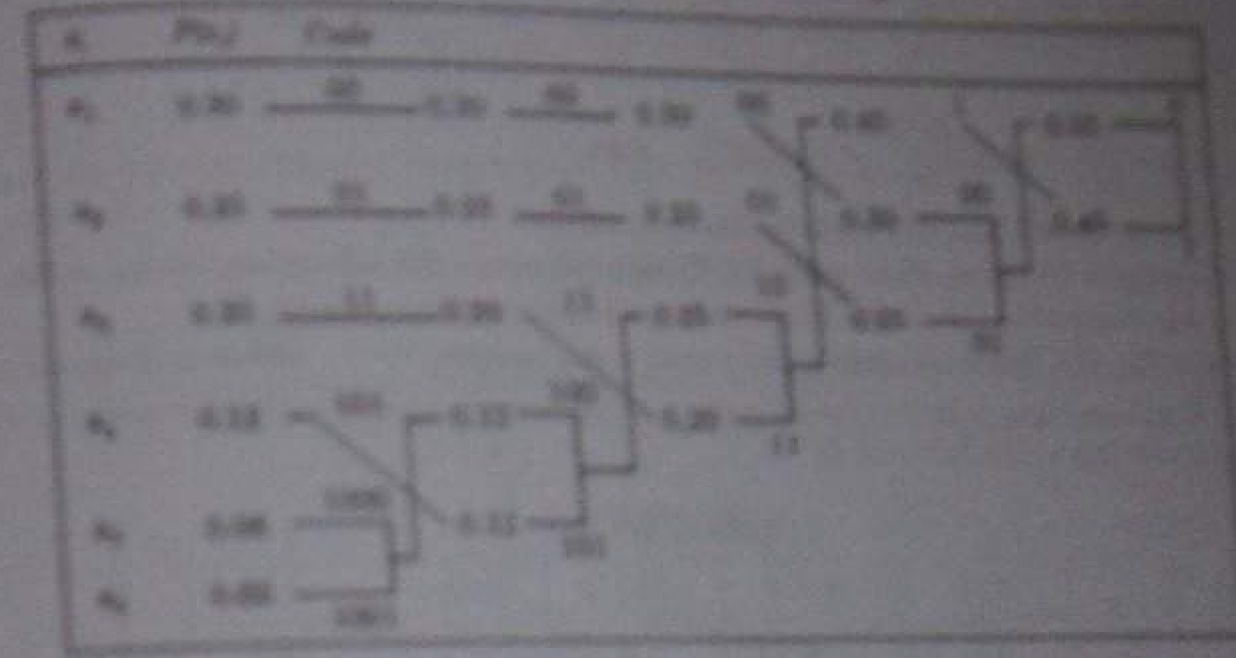
In general, Huffman encoding results in an optimum code. Thus, it is the code for ... highest efficiency. The Huffman encoding procedure is as follows:

1. List the source symbols in order of decreasing probability.
2. Combine the probabilities of the two symbols having the lowest probabilities and ... the resultant probabilities, this step is called reduction 1. The same procedure is ... until there are two ordered probabilities remaining.
3. Start encoding with the last reduction, which consist of exactly two ordered probabilities. Assign 0 as the first digit in the codewords for all the source symbols associated with the first probability; assign 1 to the second probability.
4. Now go back and assign 0 and 1 to the second digit for the two probabilities that were combined in the previous reduction step, retaining all assignments made in Step 3.
5. Keep regressing this way until the first column is reached.

An example of Huffman encoding is shown in Table 15.7

$$H(X) = 2.36 \text{ b/symbol}$$
$$L = 2.38 \text{ b/symbol}$$
$$\eta = 0.99$$

Table 15.7. Huffman Encoding



Example 15.44. A DMS X has four symbols x_1, x_2, x_3, and x_4 with $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, and $P(x_3) = P(x_4) = \frac{1}{8}$. Construct a Shannon-Fano code for X; show that this code has the optimum property that $n_i = I(x_i)$ and that the code efficiency is 100 percent.

Solution: The Shannon-Fano code is constructed as follows (see Table 15.8).

Table 15.8

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Code
x_1	1/2	0			0
x_2	1/4	1	0		10
x_3	1/8	1	1	0	110
x_4	1/8	1	1	1	111

$$I(x_1) = -\log_2 \frac{1}{2} = 1 = n_1$$

$$I(x_2) = -\log_2 \frac{1}{4} = 2 = n_2$$

$$I(x_3) = -\log_2 \frac{1}{8} = 3 = n_3$$

$$I(x_4) = -\log_2 \frac{1}{8} = 3 = n_4$$

We know that,

$$H(X) = \sum_{i=1}^{4} P(x_i) I(x_i)$$

$$= \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) = 1.75$$

$$L = \sum_{i=1}^{4} P(x_i)n_i = \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) = 1.75$$

Also, $$\eta = \frac{H(X)}{L} = 1 = 100\%$$ **Ans.**

Example [illegible]. A DMS X has five equally likely symbols.

(i) Construct a Shannon-Fano code for X, and calculate the efficiency of the code.

(ii) Construct another Shannon-Fano code and compare the results.

(iii) Repeat for the Huffman code and compare the results. [illegible]

Solution: A Shannon-Fano code [by choosing two approximately equiprobable (0.4 versus 0.6) sets] is constructed as follows (see Table 15.9)

Table 15.9

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Code
x_1	0.2	0	0		00
x_2	0.2	0	1		01
x_3	0.2	1	0		10
x_4	0.2	1	1	0	110
x_5	0.2	1	1	1	111

$$H(X) = -\sum_{i=1}^{5} P(x_i)\log_2 P(x_i) = 5(-0.2\log_2 0.2) = 2.32$$

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+2+2+3+3) = 2.4$$

The efficiency η is $$\eta = \frac{H(X)}{L} = \frac{2.32}{2.4} = 0.967 = 96.7\%$$ **Ans.**

(ii) Another Shannon-Fano code [by choosing another two approximately equiprobable (0.6 versus 0.4) sets] is constructed as follows (see Table 15.10)

Table 15.10

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Code
x_1	0.2	0	0		00
x_2	0.2	0	1	0	010
x_3	0.2	0	1	1	011
x_4	0.2	1	0		10
x_5	0.2	1	1		11

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+3+3+2+2) = 2.4$$

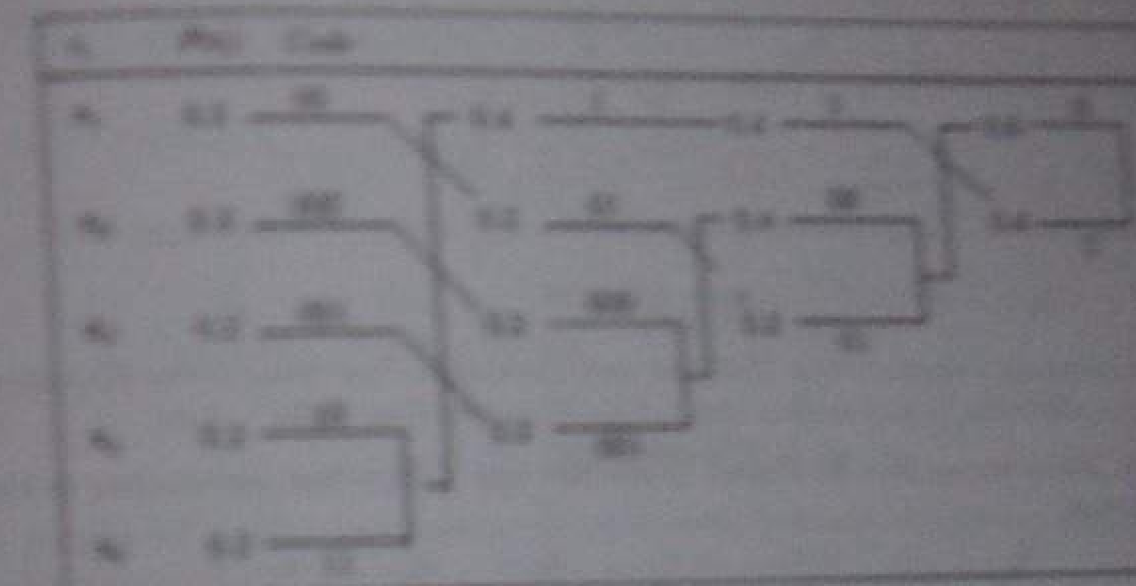
Since the average codeword length is the same as that for the code of (i), the efficiency [illegible]

(iii) The Huffman code is constructed as follows (see Table 15.11)

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+3+3+2+2) = 2.4$$

Since the average code word length is the same as that for the Shannon-Fano code, the efficiency is also the same.

Table 15.11



Example [illegible]. A DMS X has five symbols x_1, x_2, x_3, x_4 and x_5 with $P(x_1) = 0.4$, $P(x_2) = 0.19$, $P(x_3) = 0.16$, $P(x_4) = 0.15$, and $P(x_5) = 0.1$.

(i) Construct a Shannon-Fano code for X, and calculate the efficiency of the code.

(ii) Repeat for the Huffman code and compare the results.

Solution: The Shannon-Fano code is constructed as follows (see Table 15.12)

Table 15.12

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Code
x_1	0.4	0	0		00
x_2	0.19	0	1		01
x_3	0.16	1	0		10
x_4	0.15	1	1	0	110
x_5	0.1	1	1	1	111

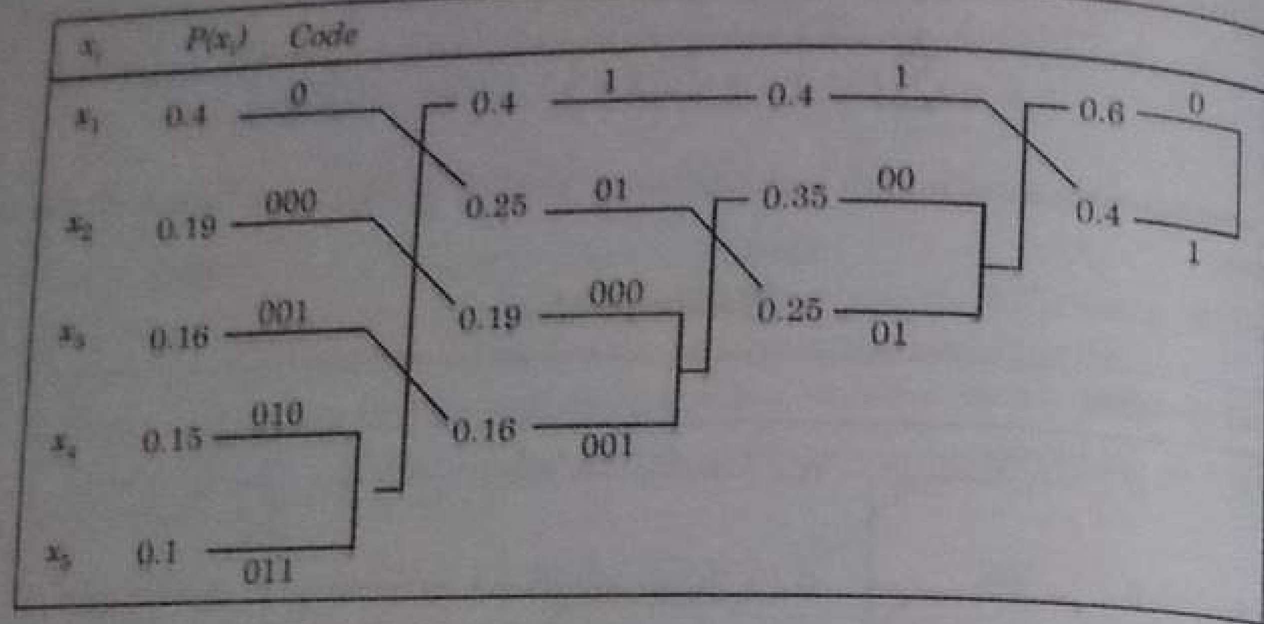
$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.4(2) + 0.19(2) + 0.16(2) + 0.15(3) + 0.1(3) = 2.25$$

Also, $$\eta = \frac{H(X)}{L} = \frac{2.15}{2.25} = 0.956 = 95.6\%$$ **Ans.**

(ii) The Huffman code is constructed as follows (see Table 15.13)

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.4(1) + (0.19 + 0.16 + 0.15 + 0.1)(3) = 2.2$$ **Ans.**

Table 15.18.



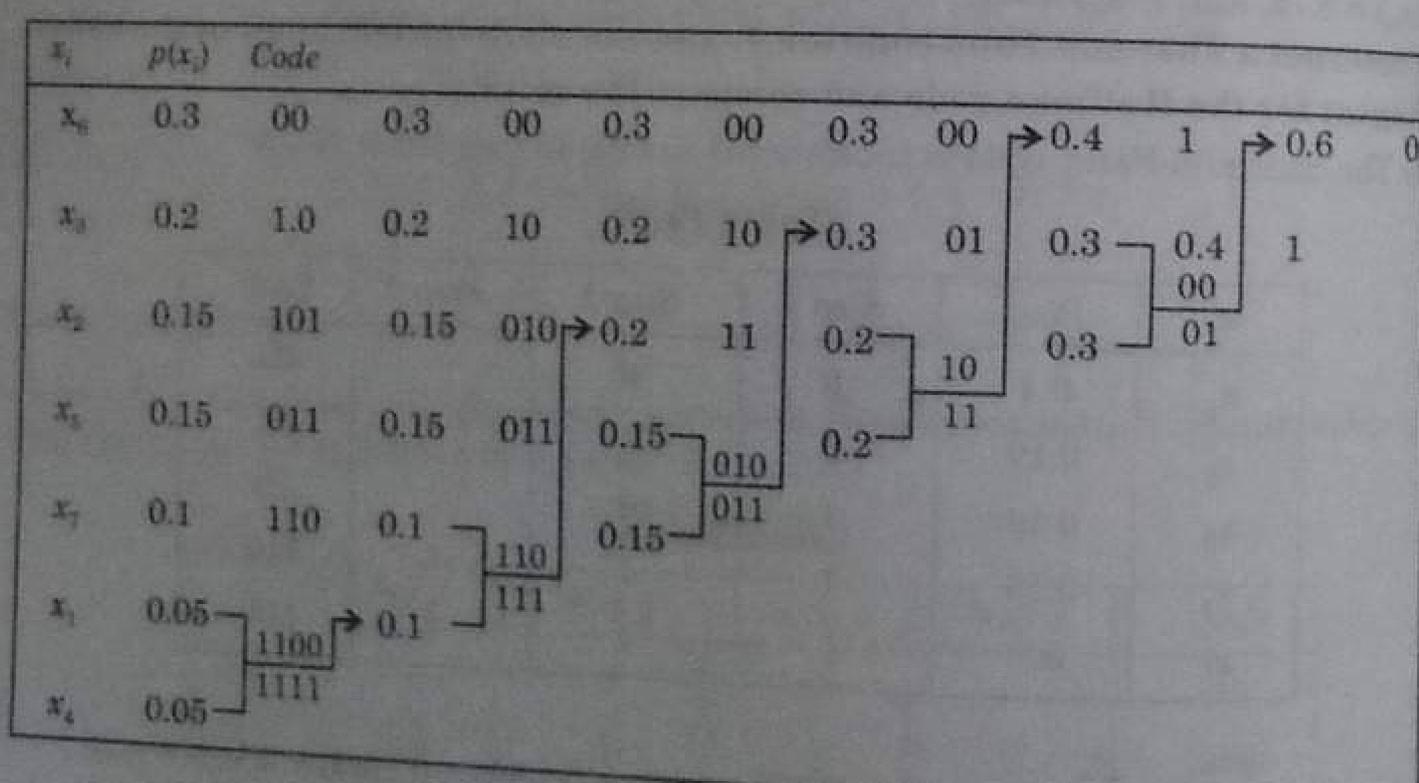
$$\eta = \frac{H(X)}{L} = \frac{2.15}{2.2} = 0.977 = 915.7\%$$

The average codeword length of the Huffman code is shorter than that of the Shannon-Fano code, and thus the efficiency is higher than that of the Shannon-Fano code. **Ans.**

Example 15.47. Determine the Huffman code for the following messages with their probabilities given

x_1	x_2	x_3	x_4	x_5	x_6	x_7
0.05	0.15	0.2	0.05	0.15	0.3	0.1

Solution: Arranging and groping of messages is done as shown below:



Message	*Prob.*	*Code*	*No. of bits in code*
x_1	0.05	1110	4
x_2	0.15	010	3
x_3	0.2	10	2
x_4	0.05	1111	4
x_5	0.15	011	3
x_6	0.3	00	2
x_7	0.1	110	3

∴ Average length L is given by

$$L = \sum_{i=1}^{7} n_i P(x_i)$$

or

$$L = 4(0.05 + 0.05) + 3(0.15 + 0.15 + 0.1) + 2(0.2 + 0.3)$$

$$L = 2.6 \text{ bits}$$

Entropy $H(X)$ is given by

$$H(X) = \sum_{i=1}^{7} P(x_i) \log_2 \frac{1}{P(x_i)}$$

$$= 0.3 \log_2\left(\frac{1}{0.3}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.3\log\left(\frac{1}{0.15}\right) + 0.1\log_2\left(\frac{1}{0.3}\right) + 0.1\log_2\left(\frac{1}{0.0}\right.$$

$$= 2.57 \text{ bits}$$

$$\eta = \frac{H(X)}{L \log_2 M} = \frac{2.57}{2.6 \log_2 2} = \frac{2.57}{2.6} = 98.85\%$$

SUMMARY

1. The purpose of a communication system is to carry information-bearing baseband one place to another place over a communication channel.
2. Information theory is a branch of probability theory which may be applied to the communication systems.
3. In the context of communications, information theory deals with mathematical analysis of a communication system rather than with physical sources and physica
4. Information theory was invented by communication scientists while they were studying structure of electronic communication equipments. When the communique is readily such as an electric current, the study of the communication system is relatively ea
5. An information source may be an object produces an event, the outcome of which random according to a probability distribution. A practical source in a communicatio device which produces messages, and it can be either analog or discrete.
6. A discrete information source is a source which has only a finite set of symbols as pos The set of source symbols is called the source alphabet, and the elements of the symbols or letters.
7. Information sources can be classified as having memory or being memoryless. A source is one for which a current symbol depends on the previous symbols. A memoryless so which each symbol produced is independent of the previous symbols.
8. A discrete memoryless source (DMS) can be characterized by the list of the symbols, t assignment to these symbols, and the specification of the rate of generating these s source.
9. The amount of information contained in an event is closely related to its uncertai containing knowledge of high probability of occurrence convey relatively little infor
10. In a practical communication system, we usually transmit long sequences of sym information source. Thus, we are more interested in the average information that a so than the information content of a single symbol.
11. For quantitative representation of average information per symbol we make the followin
    (i) The source is stationary so that the probabilities may remain constant with ti
    (ii) The successive symbols are statistically independent and come form the sourc rate of r symbols per second.

used in the receiver to despread the receive signal which is synchronized to the PN sequence used to spread the transmitted signal in the transmitter.

A solution to synchronisation problem consists of two parts. They are:

(i) Acquisition

(ii) Tracking.

Acquisition: In acquisition (or) coarse synchronisation the two PN code are aligned to within a fraction of a chip in a short a time as possible.

Tracking: Once the incoming PN code has been acquired tracking (or) fine synchronisation takes place.

PN acquisition proceeds in two-steps:

First, the received signal is multiplied by a locally generated PN code to produce a measure of correlation between in and the PN code used in transmitter.

Secondly, An appropriate decision rule and strategy is used to process the measure of correlation so obtained to determine whether the two codes are in synchronism

As for tracking, it is accomplished using phase techniques very similar to those used for the local generation of coherent carrier references.

REVIEW QUESTIONS

1. A true random waveform has no DC term. Why is there a DC term in the power density of the PN code?
2. (*a*) What are the sequences generated by the polynomials $x^5 + x^4 + x^3 + x^2 + 1$ and $x^5 + x^4 + x^2 + x + 1$? Assume an initial conditions of (11111) in both cases.
   (*b*) Compute the plot the autocorrelation function for each sequence in part (*a*).
   (*c*) Compute and plot the cross-correlation function for the two sequences.
   (*d*) Draw the circuit diagram for generating each of the sequences.
3. A PN sequences is $12^{15} - 1$ in length. How many runs of four 1s would be expected?

□□□

CHAPTER 18

Cellular and Mobile Communications

Inside this Chapter

18.1. Introduction

Although the concept of cellular communication was developed in 1947 in AT&T Bell laboratorie
USA, but the first tests were conducted to explore the possibility in commercial application
After that it then took another eight years when the Federal Communication Commission
USA, set aside new radio frequency for "Land Mobile communication". In this particular ye
(i.e., 1970), AT&T proposed to establish the very first high capacity cellular telephone system.
was then called as the advanced mobile phone service or AMPS.

Thus, in nineteen six*ies, more than half of the world cellular subscribers were using
AMPS system developed by AT&T. Now, the cellular carriers and equipment manufactures h
upgraded their system from analog to digital technology.

The idea behind this is to offer a higher quantity, higher capacity and more feature rich service for cellular users. Firstly, CTIA (Cellular Telecommunication Industry Association) adopted TDMA as its digital transmission standard in 1990 and then continued to support the system. The TDMA claims three times the transmission capacity of analog systems used previously. However, as a matter of fact, TDMA's strongest competitors is CDMA (*i.e.*, Code Division Multiple Access), a superior system developed by QUALCOMM., Inc. In fact, CDMA is based on spread spectrum technique which was originally developed for the military to scatter signals across a wide frequency band and hence making it difficult to intercept or jam it. In addition to its superior qualities, the most important feature of CDMA is that it offers at least 10 times the communication capacity of the present analog communication system.

One more digital transmission scheme, broadband CDMA, is being promoted by Intel corporation, Digital communication corporation which claims that B-CDMA, provides additional capacity to the network and also improves additional capacity to the network and improved voice quality. However, new equipment manufacturers today support all the major techniques, *i.e.*, AMPS, Narrow band AMPS, TDMA, CDMA and the GSM–the European digital standard. India selected GSM, though costly and yet to be field proven on a under scale. This digital system was proposed and developed by conference of European posts and Telecommunication (CEPT), with strong backing from European commission. The technical work in devising a common system has been co-ordinated by the CEPT's Groups Special Mobile (GSM) committee. Afterwards system has been renamed as Global system for Mobile communications. Now before discussing all this in detail, let us see the chart which illustrates the frequency allocated for different services. It has been illustrated in figure 18.1.

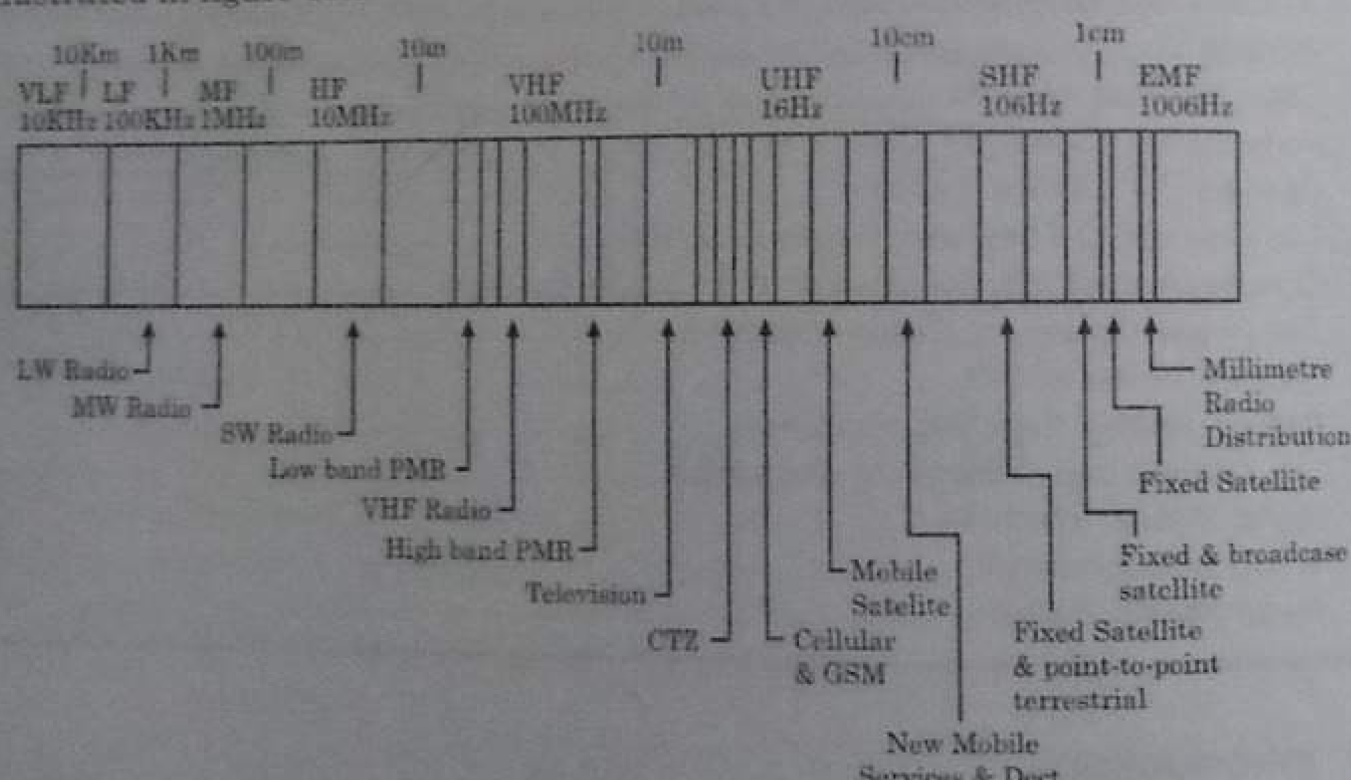


Fig. 18.1. *Chart showing the frequencies allocated for different services.*

18.2. Main Methods of Radio Transmission

Although there are several methods of radio transmission but mainly there are three methods of radio transmission listed as under:

(*i*) FDMA/SCPC (Frequency Division Multiple Access/ Single channel per carrier). Here each channel uses a separate carrier.

(*ii*) Wideband TDMA (Time Division Multiple Access). Here a single carrier is modulated to cover the whole band.

(*iii*) Narrowband TDMA (or FDM frequency division multiplex-TDMA). Here a number of carriers operate at different f...

Speech is coded using a regular pulse excitation long term prediction linear predictive coder into a data stream of 13 kbit/s, although there is an option to halve the coding rate to double the spectrum efficiency and system capacity.

In terms of technology GSM is a most demanding system with the full range of digital techniques, *viz.*, equalisation, frequency hopping, sophisticated speech coding, error correction coding, echo cancellation block interleaving and advanced modulation provided to maximise the performance. The degree of processing is such that the battery current drain of the integrated circuits in the mobile is comparable with the current required to provide the RF power for the transmitter.

18.3. GSM Standards for Cellular Telephony

The GSM air interface provides the physical link between the mobile and the network. Some of the important characteristic of the air interface are given in Table 18.1. GSM is a digital system employing time division multiple access (TDMA) technique and operates at 900 MHz.

Table 18.1.

1.	Frequency band mobile-base	890-915 MHz
2.	Frequency band base-mobile	935-960 MHz
3.	124 radio carriers spaced by	200 MHz
4.	TDMA structure with 8 time slots per radio carrier	
5.	Gaussian minimum shift keying (GMSK) modulation with	BT = 0.3
6.	Slow frequency hopping at 217 hops per second	
7.	Block and combustion channel coding with interleaving	
8.	Down link and up line control	
9.	Discontinuous transmission and reception.	

The CEPT has made available two frequency bands in the GSM system: (*i*) 890 MHz for the mobile to base station (up, link), and (*ii*) 985 MHz to 960 MHz for the base station to mobile (down link).

These 25 MHz bands are divided into 124 pairs of carriers spaced by 200 MHz. Each of the carriers is divided into 8 TDMA time slots of 0.577 m sec length, such that the frame length is 4.615 m sec.

The recurrence of each time slot makes up one physical channel, such that each carrier can support eight physical channels, both in up link and down link directions. Figure 18.2 show features of GSM standard.

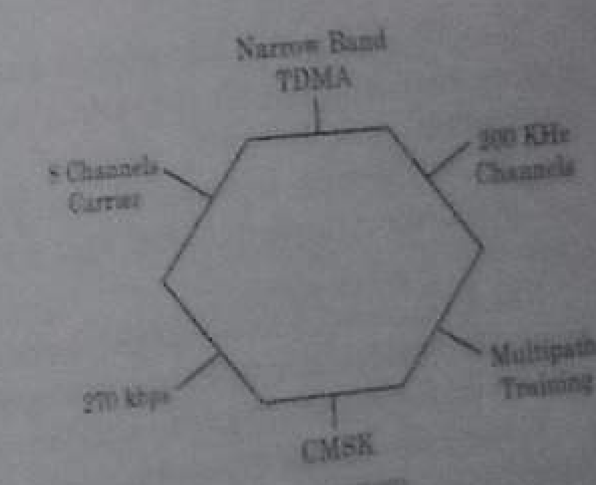


Fig. 18.2.

18.4. Architecture of GSM

Although the basic architecture of different cellular standards are the same, their individual components and configuration may differ drastically. Basic components of GSM include, base transceiver station (BTS), base station controller (BSC), Mobile switching control (MSC) and a variety of registers and network management systems shown in figure 18.3.

The mobile station comprises a mobile equipment and a subscriber identity mobile (SIM) for security and authentication of subscriber. The BTS and BSC together constitute the base station sub-system (BSS) and perform all the functions related to the radio channel for speech, data signalling and frequency hopping control and power level control.

The MSC, VLR and HLR are concerned with mobility management functions. These include authentication and registration of a mobile customers, location updating, call setup and release.

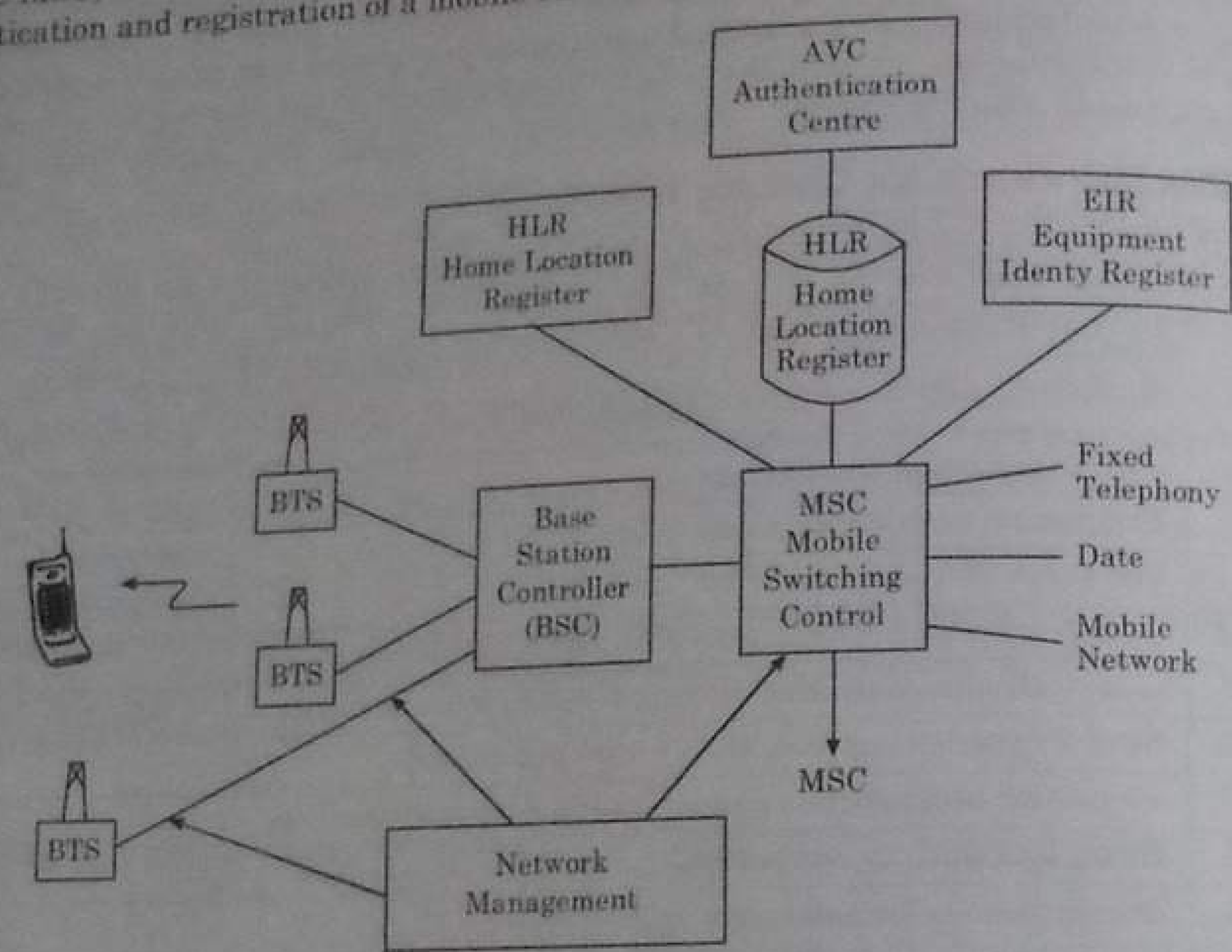


Fig. 18.3.

The HLR is the master subscriber data base and carrier information about individual subscriber numbers. Subscription levels, call restrictions, supplementary services and the most recent location of subscriber.

The VLR acts as a temporary subscriber data base for all subscribers and contains similar information as that in MLR VLR deviates a need of the MCS to access the HLR for energy transaction.

The authentication centre (AUC) works closely with the HLR and provides information to authenticated all cells in order to guard against fraud. The equipment identity register (EIR) is used for equipment security and validation of different types of mobile equipment.

Network management is used to monitor and control the major elements of the GSM Network. In particular it monitor and reports faults and performance data besides helping in reconfiguration of the network.

GSM also defines several interfaces which include the radio interface, the interface between MSC and BSC, interface the external data device and signalling interface that allows roaming between different GSM Network.

Features of GSM

The primary objective of GSM is to provide a full roaming mobile telephony service. Three broad categories of services provided by GSM are

(*i*) Teleservices,

(*ii*) Bearer services, and

(*iii*) Supplementary services.

(i) Teleservice

Teleservice are the services which are provided on a user terminal basis. Paramount teleservices include voice communication and facsimile transmission.

(ii) Bearer Services

For bearer services, the terminal equipment is provided by the user, the responsibility of the network service provide ending at the point connection. Data rates between 300 and 9600 bps fall into this category.

Supplementary services will be developed along the lines of ISDN services but will very from country to country. GSM uses the international standards organisation (ISO) and open systems interconnection (OSI) model. This model envisages structuring data communication on network in general.

OSI consists of seven layers. Figure 18.4 shows an OSI model for the mobile parts depicting the first three layers.

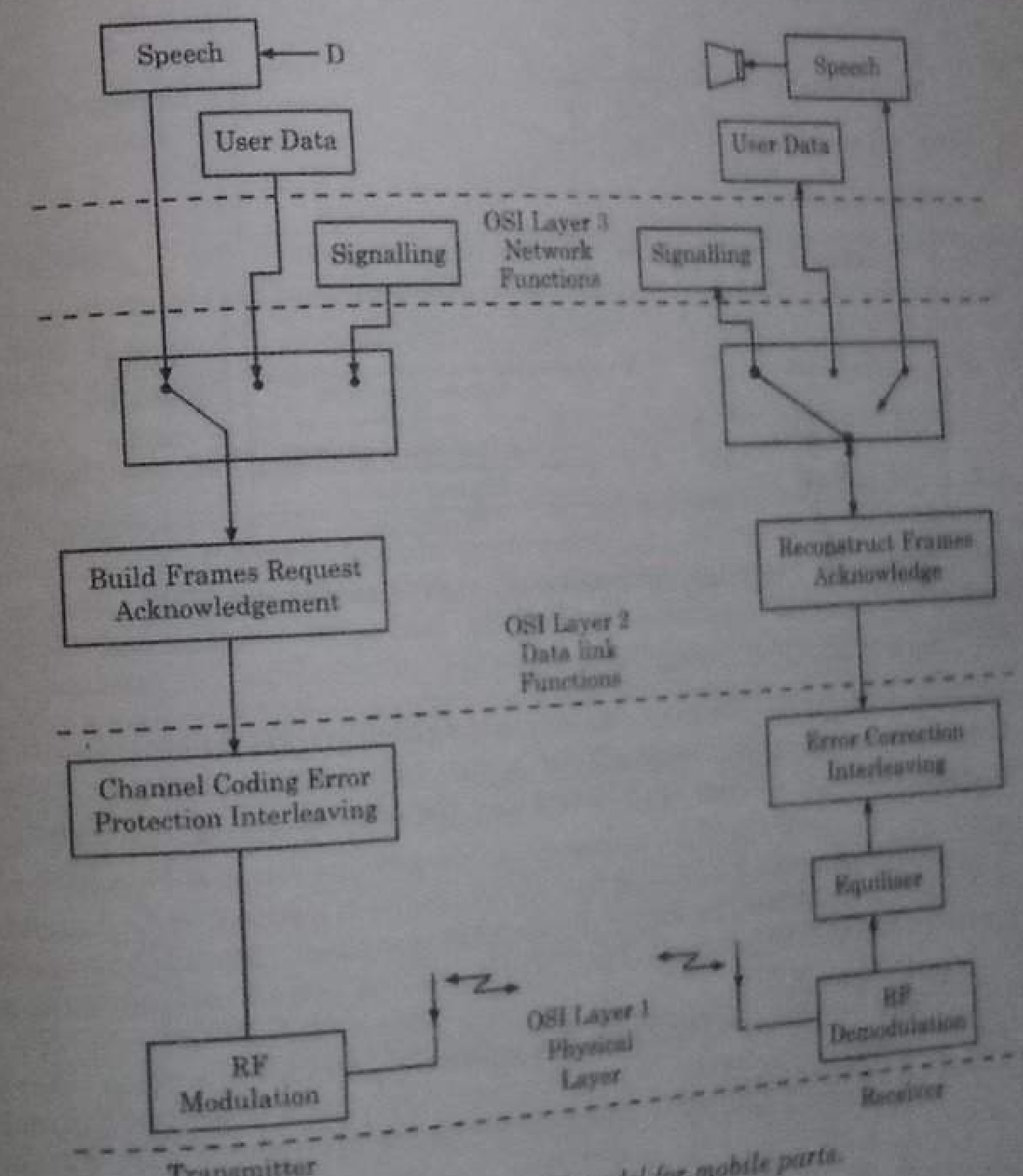


Fig. 18.4. *An OSI model for mobile parts.*

(iii) Supplementary Services

In the lowest layer 1, the physical characteristics of the transmission or radio path medi... are special in reference to GSM radio link. These include frequencies, modulation, and elem... of error protection coding.

Layer 2, the data link layer consists of element responsible for safe communication of messages or frames between radio stations.

Layer 3, the network layer, is responsible for managing calling and related activity of the radio network. Figure 18.5 shows the graphical representation of OSI model and GSM.

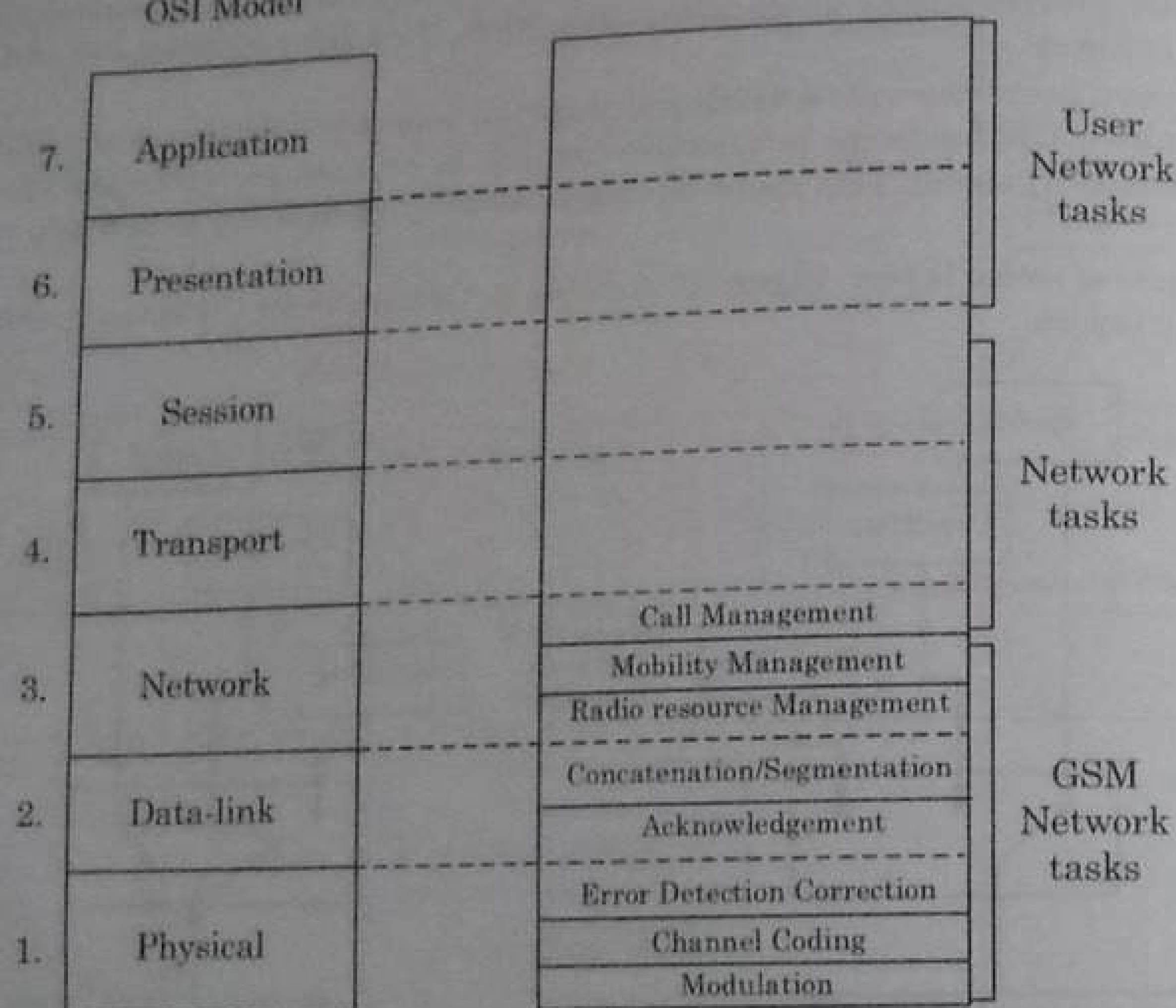
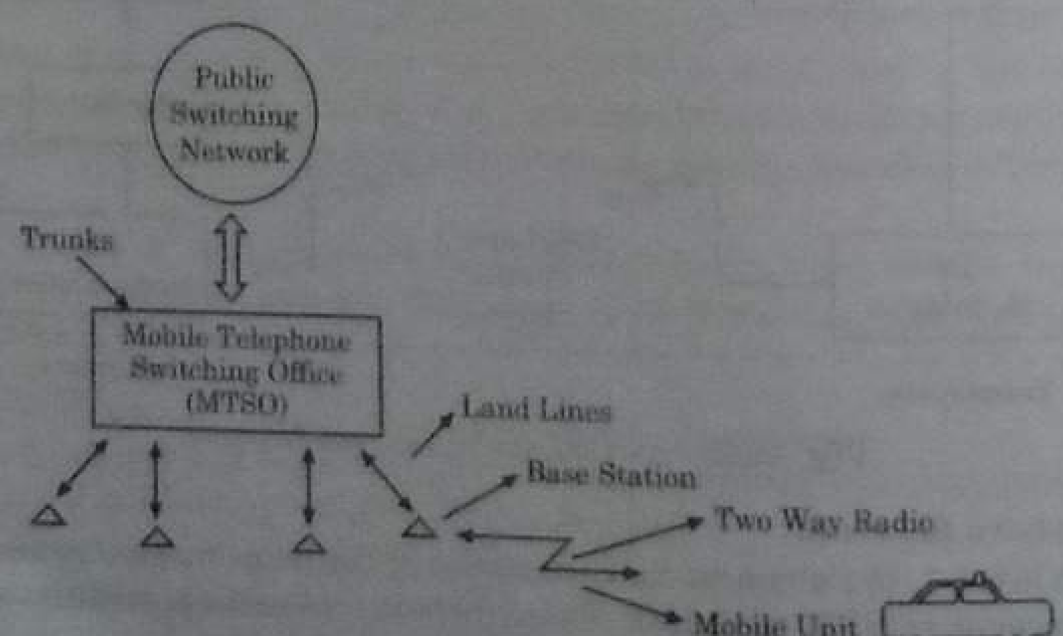


Fig. 18.5. *The graphical representation of OSI model and GSM*

18.5. The Cellular Mobile Radio Systems

Cellular mobile radio systems have come in a big way in the global market for providing telephone services to people on the move. The basic cellular mobile radio system shown in figure 18.6. These can be adopted for fixed cellular applications also, the principle being the same.



In cellular radio system, the geographical area under consideration is divided into a number of cells as shown in figure 18.7. These cells are usually hexagonal in shape and are organised into clusters with most cellular systems using seven cells cluster. The radio channels are allocated across the seven channels. The clusters are then repeated over and over again to cover the entire geographical area served by the system. Since cells using the same channels are separated from each other, and also since the transmitted signal power is low, interference is less likely. The shape of the cells need not always be hexagonal but depends on the terrain the radio propagation.

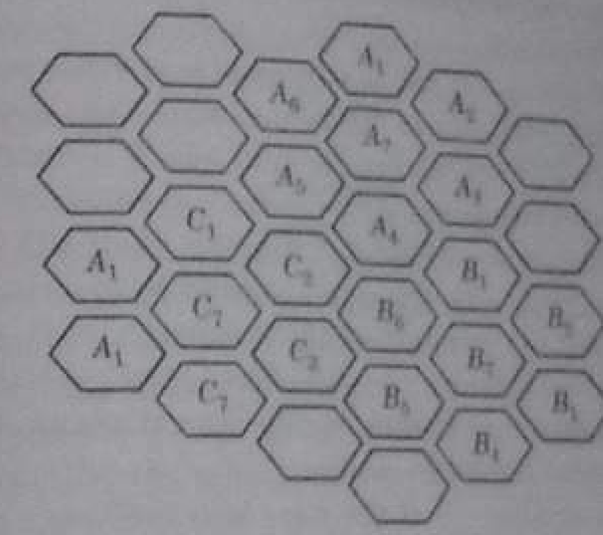


Fig. 18.7. *Cells in a cellular system permit frequency reuse without interference.*

Cellular systems provide an excellent breakthrough as they permit 'frequency reuse' in non-adjacent cells. That is, the same frequency channels can be used in different cells which are separated by a sufficient distance, without interference.

18.6. Structure of Cellophone

The compact cellphone is a very complex piece of digital engineering which has been made possible by advanced ICs and advancements in telecom technology. As a matter of fact the handset can be though to consist of two units–the mobile terminal (or the phone itself) and the Subscriber Identity Module (SIM). The SIM is a credit card-sized plastic module which fits into your mobile phone. This SIM is 'smart card', and consists of all the subscriber related information, like your cellular identification number and other preferences. It can also be used for storing messages and phone numbers.

It also enables charges to be automatically billed to the card-holder, regardless of who owns the phone. This means that if your handset is out of order, you can always use any other handset with your SIM card without affecting the other person's billing.

Functionally, the handset itself may be divided into three main parts. Terminal adaptations, radio modem and the radio frequency (RF) unit. Terminal adaptations comprise the human user interface elements (like the mic, speaker, display and keyboard), and interfaces to other equipment such as a PC or a PCMCIA data card.

The radio modem is a digital processing unit which handles the conversion of data or speech signals into digital form. It interfaces between the terminal adaptations and the RF (radio frequency) unit. This RF unit handles the actual wireless communication by receiving and transmitting radio signals to the cellular network.

18.7. Working

When you dial a number on the keypad of the phone, the handset transmits the digits, with the help of a built-in radio transceiver, to a nearby radio base station (also called cell). A group of transceiver stations are controlled by a base station controller, which in turn is connected to a

Mobile Switching Centre (MSC). The MSC, in turn, is linked to other cellular and fixed line networks.

All the switching functions within a GSM network can be handled by the MSC, which is the intelligence of the network and performs all the functions like call routing, cell control, switching plus all accounting and charging activities.

Once a cell is forwarded to the MSC, it determines how to route the call and set up the required link to enable the conversation. If the call is destined for a fixed (or normal telephone) the MSC sends it to DOT's public telephone exchange, over a leased line, which then switches the call to the desired telephone set.

However, it may be noted that if the call is destined for another mobile phone, things become more complicated. Firstly, the MSC has to figure out where the desired mobile phone is, and then forward the call to the radio base station (or cell) which is nearest to it.

But how does the MSC figure out where a particular cell phone is? It is assisted by cellphone in this task. When the handset is powered on, it initializes itself and scan the control channels. These control channels are special radio frequencies used by the cell transmitter to send and receive control data. Based upon the strength of the received signal, the handset assigns itself to a specific cell. In this process, it informs the cell of its location so that it may be passed. The handset keeps monitoring the data that is sent on the control channel till its own ID is paged- and then puts itself into the receiving mode.

However, if the handset is mobile, *i.e.* if the user is travelling by a car while calling, the cellular system also needs to keep track of the phone and the call is progressed, so that it can automatically switch the call to another cell as the caller moves from one area to another. This process of switching calls between cells is user transparent, and is known as **cell handoff.**

Although as a user you never need to be actually aware of how the base station or the network and switching subsystems work, it helps to be knowledgeable about how they function so that you can take advantage of new services whenever they are offered by the service provider.

18.8. Other Services of GSM

One advantage of the GSM system is that it enables service providers to offer a number of new and useful value-added services. These include: Call forwarding (default, busy, no-answer, unconditional), call holding and retrieving, call acceptance, preferred language, priority access, remote facilities configuration, automatic reverse charging, selective call acceptance, etc.

Currently, all cellular operations provide the Calling Line Identification service which displays the incoming caller's number. This lets you decide if you want to take the call on your cellular, or call, back on your regular phone. Voice mail retrieval is another facility offered by some operators which allows users to retrieve voice message stored in the network. These messages are typically left by parties calling the user while the phone was in use, did not answer or had phone switched off.

The Short Message Service (SMS) is a service, currently available in Europe, which lets users send and receiver messages of up to 160 characters on their cell phones. The messages may be read on the display of the phone or a PC. SMS is a useful way to transmit data because the message can be sent or received even while a voice call is in progress.

Facsimile transmission (fax) has already entered the Indian market in a big way. Tough not many Indian companies are producing fax machines, they are keenly marketing the foreign products.

The latest buzz phrase that is luring the private and public telecom industries is the radio paging system. With their ability to alert the paged person by a beep, a buzz, a voice, a vibration or a text message, paging systems are half way to the kind of totally mobile telecom facilities that cellular or car telephones allow.

Radio paging systems include input devices such as telephones and computer keyboards through which the message is transmitted with a special code number of the person to be paged. This massage goes to the switching network station where it is incoded into a signalling format required by the pagers. Then the signal is transmitted using VHF/UHF radio frequencies by digital transmitters. Various kinds of pagers which can receive tone and voice, and allow numeric or alphanumeric display, being the message to the receiver.

18.9. Performance Criteria for Cellular Phones

There are three categories for specifying performance criteria of Mobile phones as under:

1. Voice quality,
2. Service quality,
3. Special features.

18.9.1. Voice Quality

Voice quality is very difficult to judge without subjective test from user's opinions. In this technical area engineers cannot decide how to build and systems without knowing the voice quality that twill satisfy the users. For as given chimerical communication system, the voice quality will be based upon the following criterion: a set value at which by per cent of customers rate the system voice quality (from transmitter to receiver) as good or excellent the top circuit merits (CM) of the five listed below.

Circuit Merits (CM)	*Score*	*The Quality scale*
CM5	5	Excellent (speech perfectly understandable
CM4	4	Good (speech easily understable, some noise)
CM3	3	Fair (speech understandable only with considerable effort frequentry repetitions needed)
CM2	2	Poor (speech understandable only with considerable effort frequently repetitions needed)
CM1	1	Unsatisfactory (speech not understandable)

As the per centage of customers choosing CM 4 + CM 5 increases, the cost of building the system rises.

The average of the CM scores obtained from all the listeners is known as **mean opinion score (MOS)**. Usually the toll-quality voice is around MOS ≥ 4.

18.9.2. Service Quality

Three items are required for service quality:

1. ***Coverage:*** The system must serve an area as large as possible. With ratio coverage, however because of irregular terrain configurations, it is usually not practical to cover 100 per cent of the area for two reasons:

(i) The transmitted power would have to be very high to illuminate weak spots with sufficient reception, a significant cost factor.

(ii) The higher the transmitted power, the harder it becomes to control interference.

Therefore, systems usually try to cover 90 per cent of an area in flat terrain and 75 percent are in hilly terrain. The combined voice quality and coverage criteria in AMPS cellular system state that 75 per cent of users rate the voice quality between good an excellent in 90 per cent of the served area, which is generally flat terrain. The voice quality and coverage criteria would be adjusted as per decided various terrain conditions. In hilly terrain, 90 per cent of users must rate voice quality good or excellent in 75 per cent of the served area. A system operator can lower the per centage values stated above for a low-performance an blow-cost system.

2. *Required grade of service:* For a normal start-up system the grade of service is specified for a blocking probability of 02 for initiating calls at the bust hour. This is average value. However, the blocking probability at each cell site will be different. At the busy hour, near freeway, automobile traffic is usually heavy, so the blocking probability at certain cell sites may be higher than 2 per cent, especially when car accidents occur. The decrease the blocking probability requires a good system plan and a sufficient number of radio channels.

3. *Number of dropped calls:* During Q calls in an hour, if a call is dropped and $Q - 1$ calls are completed, then the call drop rate is $1/Q$. This drop rate must be kept low. A high drop rate could be caused by either coverage problems or hand off problems related to inadequate channel availability. How to estimate the number of dropped calls will be described in Chapter 9.

18.9.3. Special Features

A system would like to provide as many special features as possible, such as call forwarding, call waiting, voice stored (VSR) box, automatic roaming, or navigation services. However, sometimes the customers may not be willing to pay extra charges the these special services.

18.10. Operation of Cellular Systems

In this article, let us describe the operation of the cellular mobile system from a customer's perception without touching on the design parameters. The operation can be divided into four parts and a hand off procedure.

Mobile unit initialization: When a user sitting in a car activates the receiver of the mobile unit, the receiver scans 21 set-up channels which are designated among the 416 channels. It then selects the strongest and locks on for a certain time. Since each site is assigned a different set-up channel, locking onto the strongest set-up channel usually means selecting, the nearest cell site. This self location scheme is used in the idle stage and is user-independent. It has a great advantage because it eliminates the load on the transmission at the cell site for locating the mobile unit. The disadvantage of the self-location scheme is that no location information of idle mobile units appears at each cell site. Therefore, when the call initiates from the land line to a mobile unit, the paging process is longer. Since a large per centage of calls originates at the mobile unit, the use of self-location schemes is justified. After 60's the self-location procedure is repeated. In the future, when land line originated calls increase, feature called "registration" can be used.

Mobile originated call: The user places the called number into an originating register in the mobile unit, checks to see that the number is correct, and pushes the "send" button. A request for service is sent on a selected set-up channel obtained from self-locations scheme. The cell site receives it, and in directional cell sites, selects the best directive antenna for the voice channel to use. At the same time the cell site sends requisite to the mobile telephone switching office (MTSO) via a high-speed data link. The MTSO selects an appropriate voice channel for the call, and the cell site acts on it through the best directive antenna to link the mobile unit. The MTSO also connects the wire-line party through the telephone company zone office.

Network originated call: A landline party dials a mobile unit number. The telephone company zero office recognize that the number is mobile and forward the cell to the MTSO. The MTSO sends a paging message to certain cell sites based on the mobile unit number and the search algorithm. Each cell site transmits the page on its own set-up channel, locks onto it, and responds to the cell site. The mobile unit also follows the instruction to tune to an assigned voice channel and initiate user alert.

Call termination: When the mobile user turns off the transmitter, a particular signal (signalling tone) transmits to the cell site, and both sides free the voice channel. The mobile unit resumes monitoring pages through the strongest set-up channel.

Handoff procedure: During the call, two parties are on a voice channel. When the mobile unit moves out of the coverage area of a particular cell site, the reception becomes weak. The present cell site requests a handoff. The system switches the call to a new frequency channel in a new cell site without either interrupting the call or alerting the near. The call continues as long as the user is talking. The user does not notice the handoff occurrences. *Handoff* was first used by the AMPS system, them renamed *handover* by the European systems because the different meanings in English and American English.

18.11. The Concept of Frequency Reuse Channels

A radio channel consists of a pair of frequencies, one for each direction of transmission which may be used for full-duplex operation. A particular radio channel, say F_1, used in one geographic zone to call a call, say C_1, with a coverage radius R can be used in another cell with the same coverage radius at a distance D away.

Frequency reuse in the core concept of the cellular mobile radio system. In this frequency reuse system, users in different geographic locations (different cells may simultaneously use the same.

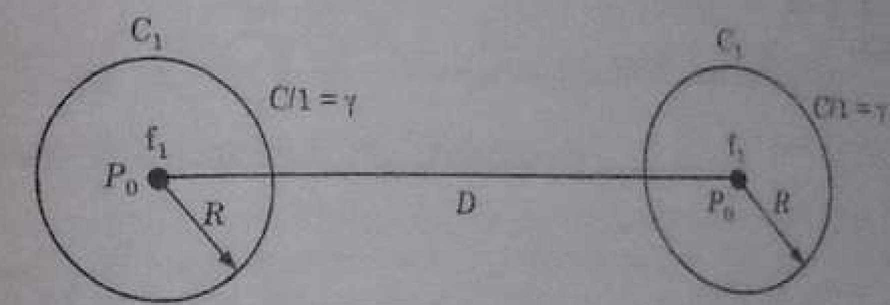


Fig. 18.8. *The ratio of D/R.*

Frequency channel (figure 18.8). The frequency reuse system can drastically increase the spectrum efficiency, but if the system is not properly designed, serious interference may occur. Interference due to the common use of the same channel is called ***cochannel interference*** and is our major concern in the concept of frequency reuse.

18.11.1. Frequency Reuse Schemes

The frequency reuse concept may be utilized in the time domain and the space domain. Frequency reuse in the time domain results in the occupation of the same frequency in different time slots. It is known as ***time-division multiplexing (TDM)***. Frequency reuse in the space domain can be divided into two categories.

(*i*) Same frequency assigned in two different geographic area, such as AM or FM radio stations using the same frequency in different cities.

(*ii*) Same frequency repeatedly used in a same generally area in one system–the scheme is used in cellular systems. There are many cochannel cells in the system. The total frequency spectrum allocation is divided into K frequency reuse patterns, as illustrated in Fig. 2.3 for $k = 4, 7, 12$ and 19.

18.11.2. The Frequency Reuse Distance

The minimum distance which allows the same frequency to be reused will depend upon several factors, such as the number of cochannel cells in the vicinity of the centre cell, the type of geographic terrain contour, the antenna height, and the transmitted power at each cell site.

The frequency reuse distance D may be found from the expression

$$D = \sqrt{3K}\,R$$

where K is frequency reuse pattern shown in figure 18.9, then

$$D = \begin{cases} 3.46R & K = 4 \\ 4.6R & K = 7 \\ 6R & K = 12 \\ 7.55R & K = 19 \end{cases}$$

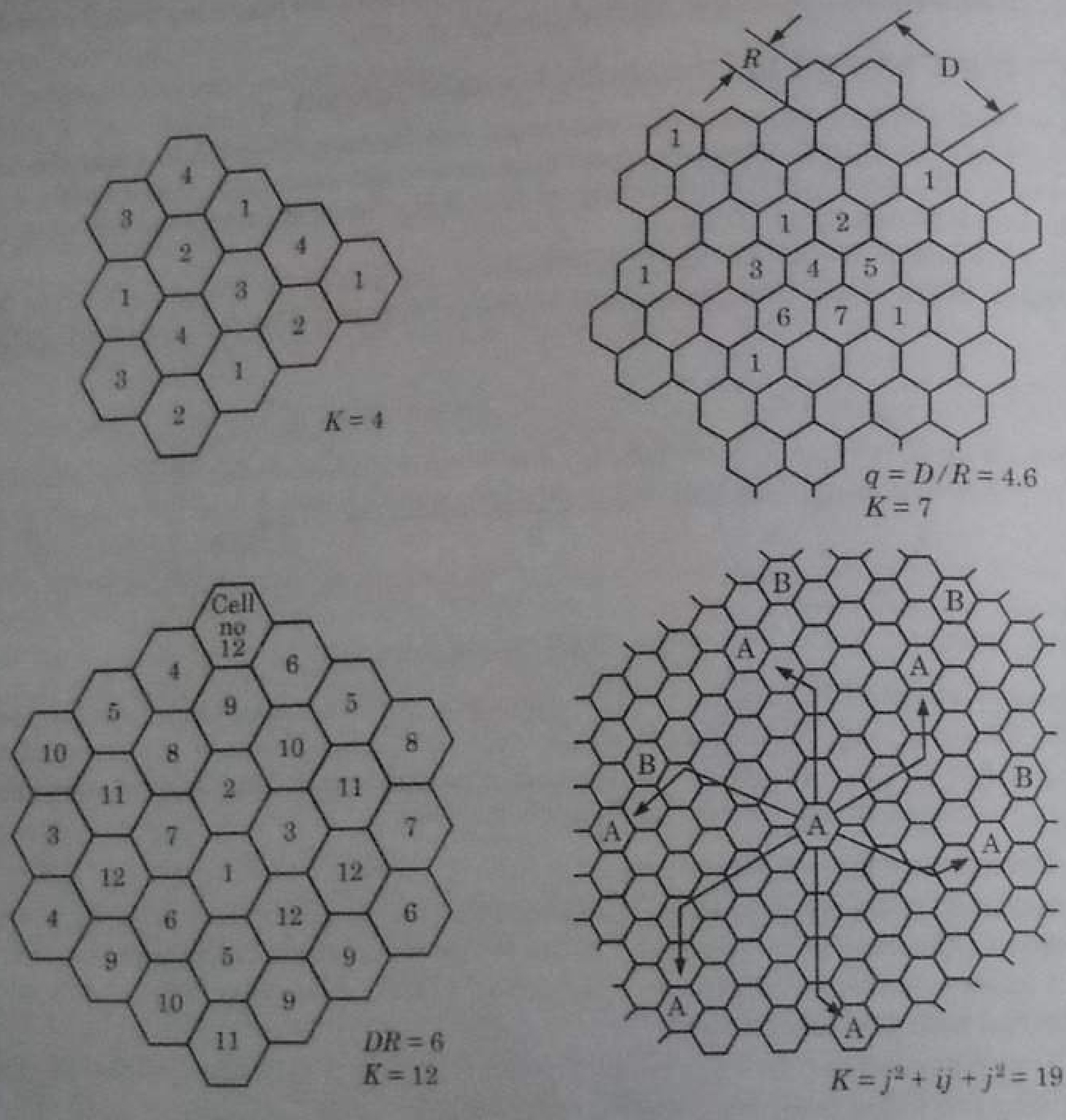


Fig. 18.9. *N-Cell reuse pattern.*

18.12. Consideration of the Components in a Cellular Systems

The elements of cellular mobile radio system design have been mentioned in the previous sections. Here we must also consider the components of cellular systems, like mobile radios, antennas, cell-site controller, and MTSO. Infact, they will affect our system design if we do not choose the right one. The general view of the cellular system is shown in figure 18.10. Even though the EIA (Electronic Industries Association) and the FCC have specified standards for radio equipment at the cell sites and the mobile sites, we still need to be concerned about that equipment. The issues affecting choice of antennas, switching equipment, and data links are briefly described here

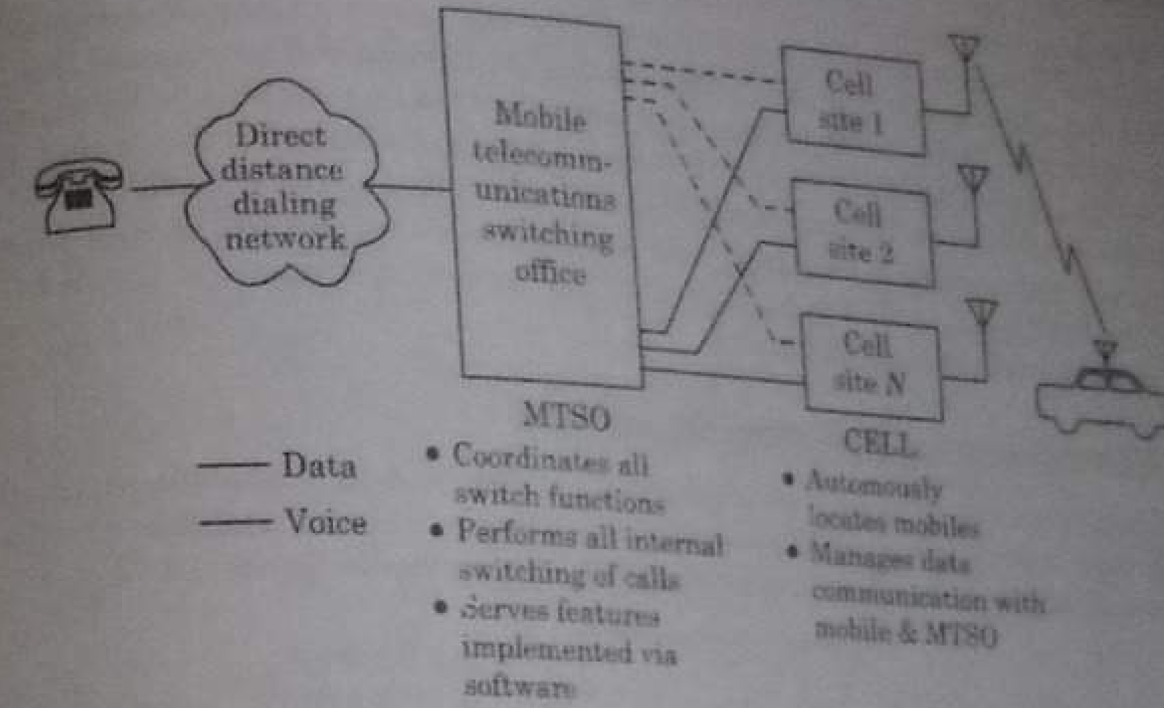


Fig. 18.10. *A general view of cellular telecommunications systems.*

18.13. The Power Control for Cellular Systems

The power level can be controlled only by the mobile transmitted switching office (MTSO), not by the mobile units, and there can be only limited power control by the cell sites as a result of system limitations.

The reasons are as under:

The mobile transmitted power level assignment must be controlled by the MTSO or the cell site, not the mobile unit. Or, alternatively, the mobile unit can lower the power level but cannot arbitrarily increase it. This is because the MTSO is capable of monitoring the performance of the whole system and can increase or decrease the transmitted power level of those mobile units to render optimum performance. The MTSO will also optimize performance for any particular mobile unit unless a special arrangement is made.

18.14. Function of the MTSO

The MTSO controls the transmitted power levels at both the cell sites and the mobile units. Th advantages the having the MSTO control the power levels are as under:

(i) Control of the mobile transmitted power levels. When the mobile unit is approaching th cell site, the mobile unit power level must be reduced for the following factors:

(a) Reducing the chance of generating intermodulation products from a saturated receivin amplifier.

(b) Lowering the power level is equivalent to reducing the chance of interfering with oth channel cell sites.

(c) Reducing the near-end-far and interference ratio.

(ii) Control of the cell-site transmitted power level. When the signal received from the mob unit at the cell site is quite strong, then MTSO must reduce the transmitted power level that particular radio at the cell site and also at the same time, lower the transmitted pow level at the mobile unit. The advantages are as under:

(a) For a particular radio channel, the cell size decreases significantly, the cochannel re distance increase, and the cochannel interference is further reduced. In other wo cell size and cochannel interferences inversely proportional to cochannel reuse distan

(b) The adjacent channel interference in the system is also reduced. However, in most cellular systems, it is not possible to reduce only one or a few channel power levels at the cell site due to of the design limitation of the combiner. The channel isolation in the combiner is 18 dB. If the transmitted power level of one channel is lower, the channel having high transmitted design an unequal-power combiner for the system operator so that che power level of each channel can be controlled at the cell site.

3. The power transmitted from a small cell is always reduced, and so is that from a mobile unit. The MTSO can facilitate adjustment of the transmitted power of the mobile units as soon as they enter the cell boundary.

18.15. Cellular Analog Switching Equipment

Most analog switching equipment consists of processors, memory, switching network, trunk circuitry and miscellaneous service circuitry. The control is usually centralized, and there is always some degree of redundancy. A common control system has been shown in figure 18.11.

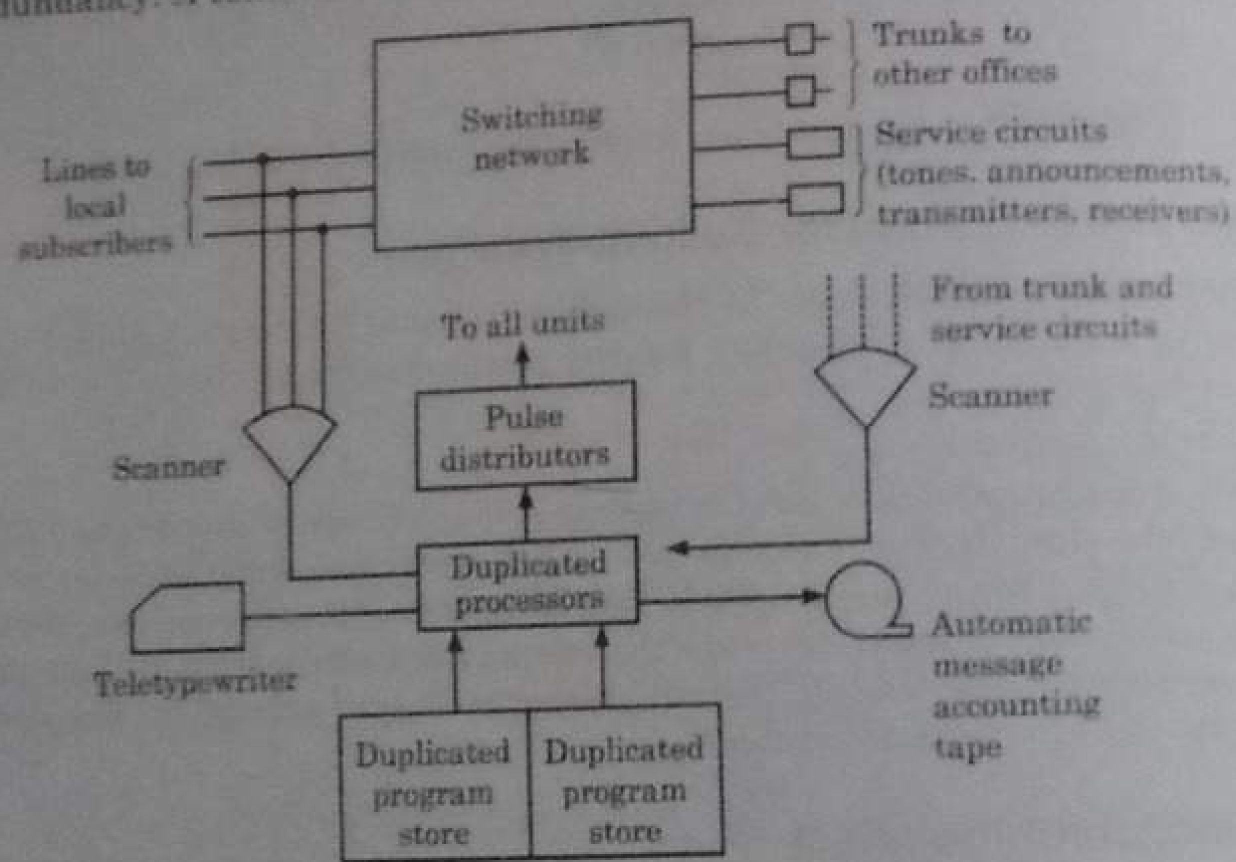


Fig. 18.11. *A typical analog switching system.*

The programs are stored in the memory which provides the logic for controlling telephone calls. The processor and the memory for programming and calls are duplicated. The switching network provides a means for interconnecting the local lines and trunks. The scanners are read under the control of the central processor. The changes in every connection at the line side and at the trunk side are also controlled by the processor. The central processor sends the order to all the units (switching network, trunk, service circuits) through pulse distributions. The automatic message accounting (AMA) tapes are used for recording the call usage. Three programmes are stored in most switching equipment (1) call processing (set, up, hand off, or disconnect a call), (2) hardware maintenance (diagnose failed or suspected failed units), and (3) administration (collect customer recorder, truck records, building data, and traffic count).

18.15.1. Modified Analog Switching Equipment

The local line side has to change to the trunk side as illustrated in figure 18.12, since the mobile unit does not have a fixed frequency channel. Hence, the mobile unit itself acts as a trunk line. In addition, the processors have to be modified to handle cellular call processing, the locating algorithm, the handoff algorithm, the special disconnect algorithm, billing (air time and wire line), and diagnosis (radio, switching, and other hardware failure).

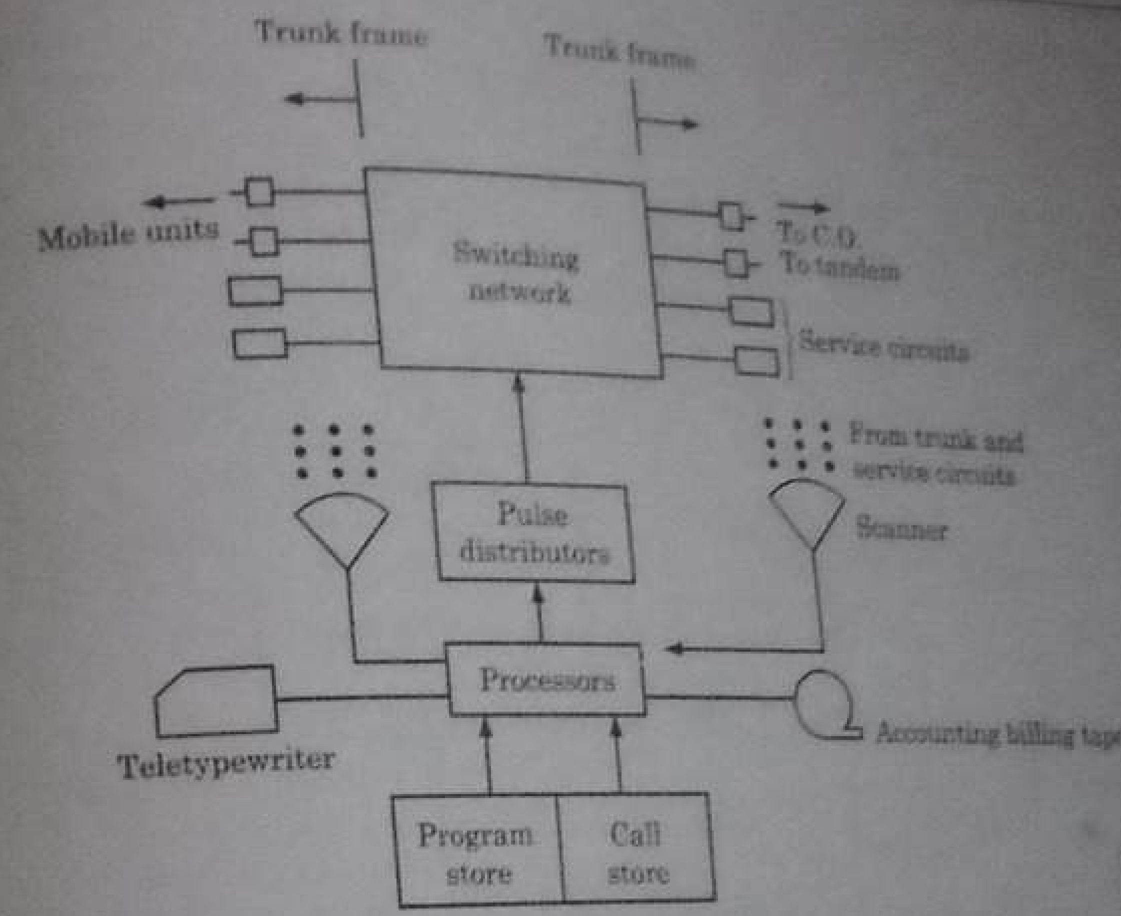


Fig. 18.12. *Illustration of a modified analog switching system for cellular mobile systems.*

18.15.2. Cell-Site Controllers and Hardware

Mobile telephone switching office (MTSO) system manufacturers designed their own cell-site controllers and transceivers (radios). Cell-site equipment is shown in figure 18... The cell site can be rendered "smarter", that is, programmed to handle many semiautonomous functions under the direction of the MTSO. Cell-site equipment consists of two basic frames under:

1. **Data frame—consists of controller and both data and locating radios**
   (a) Providing RF radiation, reception, and distribution.
   (b) Providing data communication with MTSO and with the mobile units.
   (c) Locating mobile units.
   (d) Data communication over voice channels.
2. **Maintenance and test frame**
   (a) Testing each transmitting channel for:
      (i) Incident and reflected power to and from the antenna.
      (ii) Transmitter frequency and its deviation.
      (iii) Modulation quality.
   (b) Testing each receiving channel for:
      (i) Sensitivity.
      (ii) Audio quality.

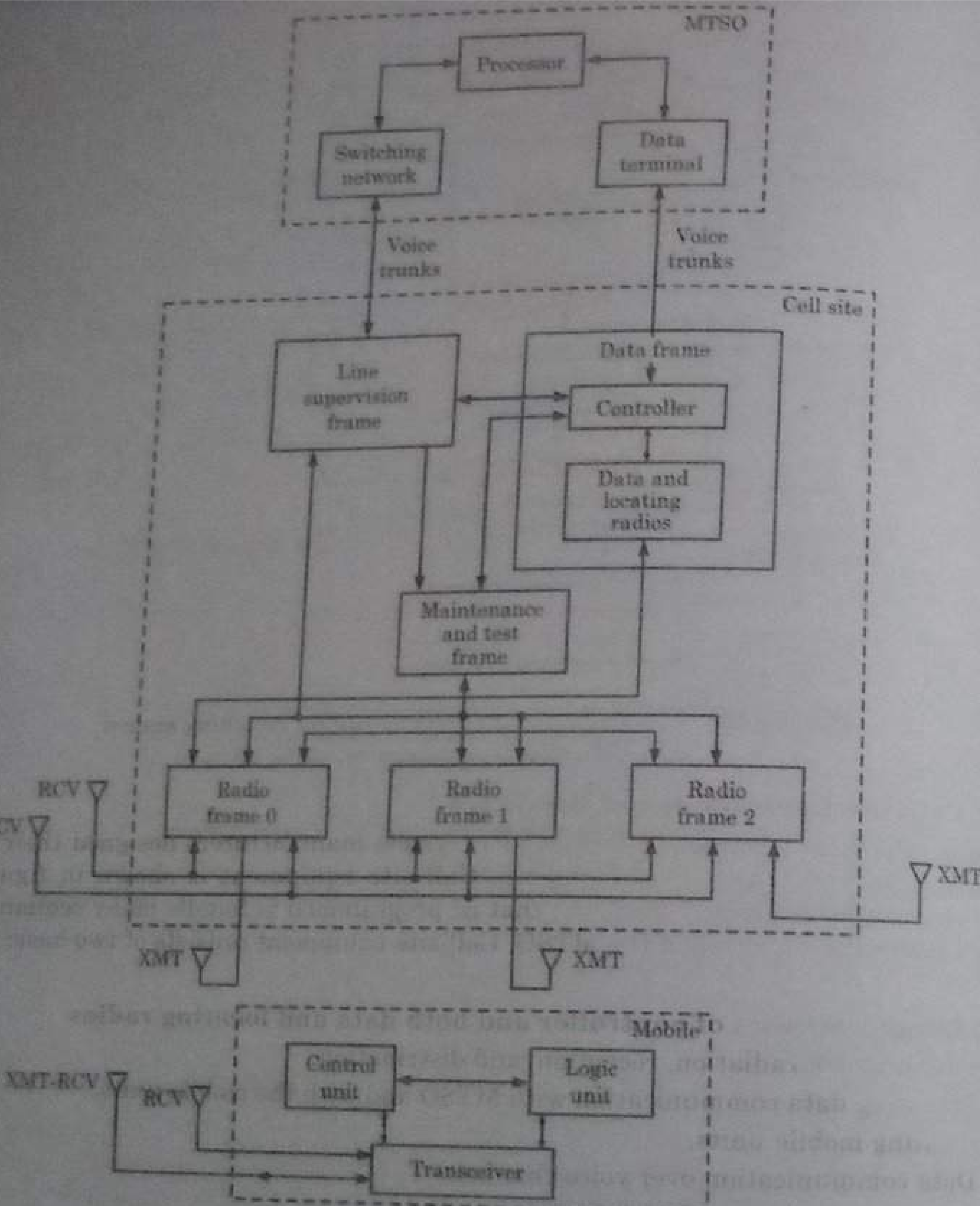


Fig. 18.13. *Illustration of Cellular mobile major systems.*

6. The Cellular Digital Switching Equipment

1. General Concept

e digital switching, which is usually the message switch, handles the digitized message alog switch, which is the circuit switch, must hold a call throughout the entire duration of . The digital (message) switch can send and message or transmit the voice in digital from: it can break a message into small pieces and send it at a fast rate. Thus, the digital switch rnate between ON the OFF modes periodically during a call. During the OFF mode, the an handle other calls. Hence, the call-processing efficiency of digital switching is higher at of analog switching.

The further digital switch may be a switching packed which would send digital information in a non-periodic fashion on request. There are several other advantages to digital switching besides its greater efficiency. Digital switches are always quite small, consume less power, require less human effort to operate, and are easier to maintain. Digital switching is also flexible and can grow medullary. Digital switching equipment can be either centralized or decentralized. A centralized system of a digital system has an architecture similar to that of an analog system: Motorola's EMX2500, Ericession's AXE-10, and Northern Telecom incarnation's DMS-MTXM are large centralized digital system. A decentralized system is described here.

A decentralized system is slightly different from a remote-control switching system. In the remote-control switching system, a main switch is used to control a remote secondary switch. In a decentralized system, all the switches are treated equally, i.e., there is no main switch.

18.16.2. Elements of Switching

One decentralized switching system can be introduced here for illustration. It is the American Telephone and Telegraph (AT&T) Autoplex 1000, which consists of an executive cellular processor (ECP), digital cellular switching (DCS), an interprocess message switch (IMS), RPC (ring peripheral control), and nodes.

1. ECP transports messages from the one processor to another.
2. IMS attached to token ring (IMS uses a token ring technology provides interfaces between ECP, DCS, cell sites, and other network. The RPC attached to the ring permits direct communication among all the elements through the ring.
3. DCS, which are digital cellular switches, function is modules to allow the systems to grow. Of these coders, LPC is attractive because of its performance and degree of complexity.

18.17. Digital Mobile Telephony

Because voice communication is the key service in cellular mobile systems, when we think of the digital systems, we must think of a digital voice.

In present-day mobile cellular systems, transmission of a digital voice in a multipath fading environment is a challenging job. The major considerations in implementing digital voice in cellular mobile systems are discussed below, along with a tentatively recommended transmission rate for the cellular mobile system.

Digital Voice in the Mobile Radio Environment

1. The criterion the judge, a good digital voice through a wire line is employed in three existing digital voice schemes:
   (*a*) In a continuously variable step delta (CVSD) modulation scheme, the present transmission rate is 16 kbps. This is not toll-quality voice transmission and is commonly used by the military.
   (*b*) In a LPC scheme, the present transmission rate of .4 kbps provides a synthetic quality voice, but a rate of .8 kbps using vector quantisation may proved a communications quality voice. A rate of 16 kbps can provide a toll-quality voice.
   (*c*) In a pulse code modulation (PCM) scheme, the present transmission rates of 32 kbps and 64 kbps is used commercially. Of the three schemes, LPC seems most attractive because if its low transmission rate. However, LPC is more vulnerable in terms of distortion to the mobile fading environment.
2. Digital voice has to be processed in real time, which imposes constraints on the digital processing time. This adversely affects LPC but not CVSD.
3. When sending a digital stream (voice) through a radio channel in a fading environment, in general an LPC scheme needs more code protection than CVSD scheme does because LPC is not implemented in a continuous waveform in either the frequency domain or the time domain while CVSD is implemented in a continuous waveform in the time domain.

4. Because the mobile unit is moving, sometimes rapidly, sometimes slowly, insertion of extra synchronization bits is needed in the normal digital stream.

onsiderations for a digital voice transmission in cellular mobile systems

The following factors are significant which are to be considered:

Digital Transmission Rate

(a) Present cellular signalling rate: The present signalling format is designed on the sumption that the mobile unit moves at an average of 30 mi/h and that the transmission rate 10 kbps. The 21 synchronization bits (10 synchronization bits and 11 frame bits) occur in front every code word of 48 bits ensure that the bits are not falling out of synchrony before the ynchronization takes place.

(b) Considerations of LPC scheme: If a rate of 4.8 kbps using LPC for a communications ality voice is accepted its rate is almost half of the present transmission rate, and at this nsmission rate a 48-bit word would be acceptable in a fading environment. The resynchronization eme for a mobile receiver should take place in front of every code work of 48 bits (21 synchronization s) + (a code word of 48 bits = 69 bits). The number of synchronization bits is almost half the nber of bits in a code word. Therefore, the transmission rate would be approximately (4.8 × = 7.2 kbps.

(c) Redundancy of transmission: The protection of synchronization in a mobile radio ronment is not sufficient. If the digital stream were to occur in a signal fade, partial or whole words would be lost. In order to prevent fading, redundancy of transmission is often used would take a minimum redundancy scheme; for example, we would transmit the same message hree times and take a "2-out -of-3 majority vote" on each bit to minimize the fading impairment e message bits. For LPC of 4.8 kbps, an RF transmission rate of (4.8 kbps × 1.5)3 = 21.6 kbps eded. It is reasonable for a 30-kHz channel to carry a transmission rate of 21.6 kbps over a r fading medium. When an RF transmission rate of 21.6 kbps over a severe fading medium n an RF transmission rate is given, the channel bandwidth can be narrower with a trade-off smitted powers.

d) Modulation, diversity coding, ARQ, and scrambling: Diversity and modulation can in reducing the RF transmission rate for the digital voice. However ARQ schemes, fancy g schemes, and complicated scrambling schemes cannot be implemented for voice transmission is because the digital voice must be processed in real time, and these three schemes usual ire be processed in real time, and these three schemes usually require a fair amount of time rocessing. These schemes can be used for data transmission.

. Word Error Rate: In the multipath fading environment, the bit error rate P_e is not the concept for voice-quality measurement; the word error rate P_w is also important and varies vehicle speed. However, information on the word error rate for transmission of digital voice a mobile radio environment only appears in two extreme. Assume that we know the required d P_w. We can covert P_e and P_w to a required carrier-to-noise ratio C/N. If a two-branch sity scheme has been used, the bit error rate of 10^{-3} in a relatively slow fading case requires / level of approximately 15 dB. The C/N level, a word error rate of a 4-bit word is about In general, if the work error rate is the same as or lower than the bit error rate for a given he C/N level is acceptable. In our case, P_w and P_e are the same at $C/N = 15$; therefore, the 15 dB is justified.

Relationship between C/N and E_b/N_0: The relationship between the carrier-to-noise /N, the energy-per-bit-noise-per-hertz ratio E_b/N_0, the transmission rate R, and the bandwidth be expressed as

$$\frac{C}{N} = \frac{E_b}{N_0}\frac{R}{B}$$

When the number of levels C/N increases, the bandwidth decreases. Keeping E_b/N_0 constant, we see that when the bandwidth decreases, the required carrier-to-noise ratio C/N increases. Previously we calculated that $C/N = 15$ dB words for a two-level (binary) system. If the number of levels increases, the C/N will be higher than 15 dB.

Example 18.1. Let $E_b/N_0 = 15$ dB for a two-level system and R_0 and B_0 be the transmission rate and transmission bandwidth, respectively, of the two-level system. Now if we reduce the bandwidth $B_1 = 0.5\ B_0$, then

$$\left(\frac{C}{N}\right) = (31.6)\frac{R_0}{0.5B_0} = 2\left(\frac{C}{N}\right)_0 = \left(\frac{C}{N}\right) + 3dB$$

This means that the power increases by 3 dB. If the transmitted power was 50 W, now it is 100 W.

18.18. MTSO Interconnection

18.18.1. Connection to Wire-Line Network

The MTSO operates on a truck-to-truck basis. The MTSO interconnection arrangement is similar to a private-branch exchange (PBX) or a class 5 central office (a tandem connection) see figure 18.14. The MTSO has three types of interconnection links.

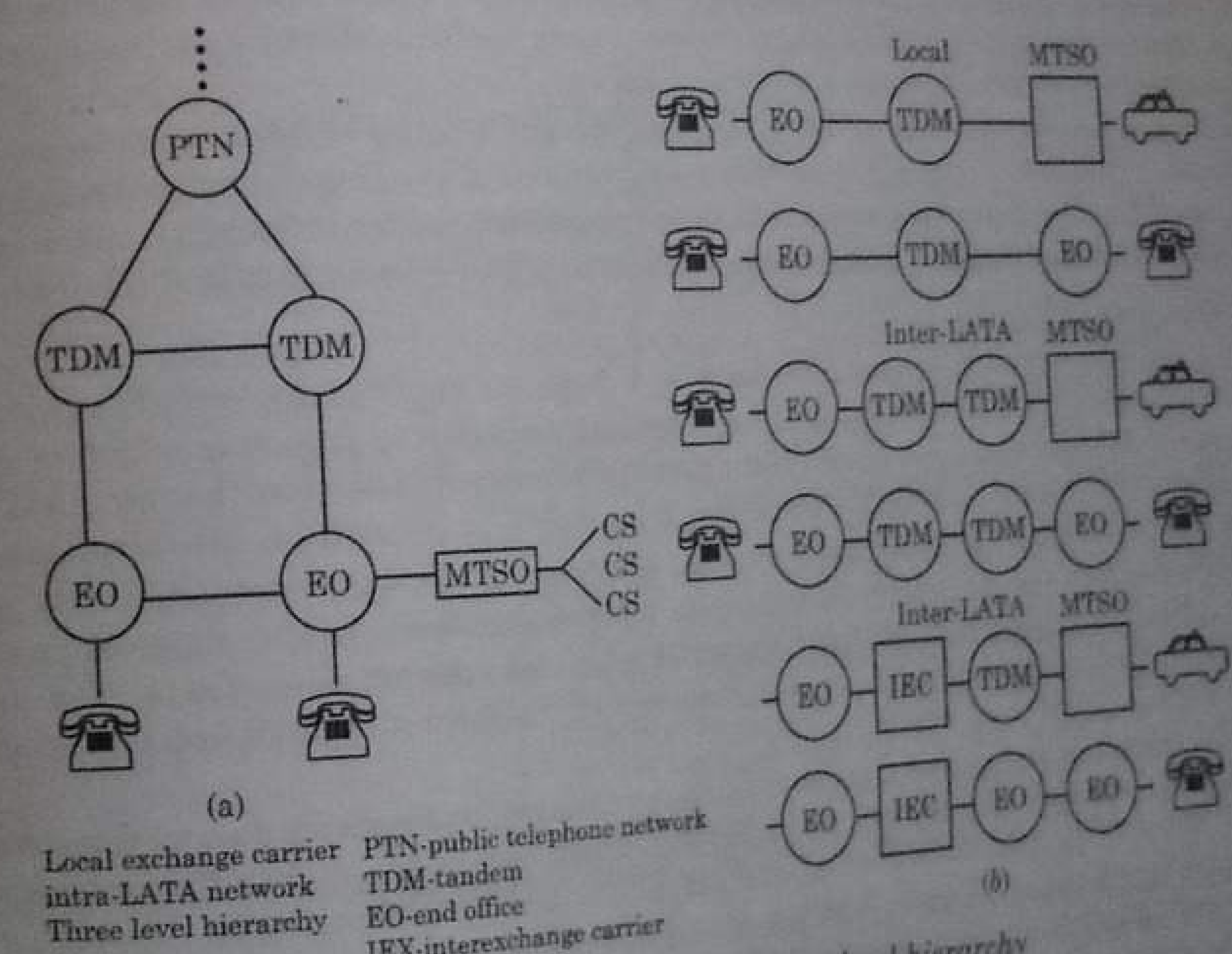


Fig. 18.14. *Illustration of Three-level hierarchy (a) Interconnection of MTSO; (b) three types of call.*

Type 1—interconnects a MTSO to a local-exchange carrier (LEC) end office.

Type 2A—interconnects an LEC tandem office.

Type 2B—interconnects to an LEC end office in conjunction with type 2A on a high-usage alternate-routing basis.

The three-level hierarchy of a public telephones network is shown in figure 18.15. With this diagram, we can illustrate the three types of calls: (1) a local call, (2) an intra-LATA (local access and transport area) call, and (3) an inter-LATA call.

18.18.5. Connection to a Cell Site

Two types of facility are used as under: [illegible]

1. Cell site trunks provide a voice communication path. Each trunk is physically connected to a cell site voice radio. The number of trunks is decided on the basis of the traffic and the desired blocking probability (grade of service).
2. The cell site acts as a traffic concentrator for the MTSO. For instance, we may design an average busy hour radio channel occupancy of [illegible] for high-traffic cells.

Both T1 carrier cables and microwave links are used. The duplication is needed for reliability.

18.19. Radio Paging Systems

As a matter of fact, two persons are required for any form of [illegible] communication; however, for good communication everyone needs to be contactable. Mobile radio is for communication, with people on the move. Paging is a simple land mobile radio system to solve the all-times, all-places, all-people communication problems.

With very economic use of the radio spectrum, paging systems can keep thousands of people in touch, with a small, light-weight, unobtrusive, body-worn equipment. Present day paging developments are available with several innovative features and powerful message capabilities and are supported by nation-wide and international transmitting and control facilities. Even in India, a sophisticated modern paging system with many useful features is being planned for Mumbai and other metropolitan too.

This article reviews the basic paging concepts, on site paging history, wide area paging developments, operation of a public paging system, constituents of a paging control and transmitting system, anatomy of miniaturised pager unit, innovative features/facilities of an alive modern system, user's multiple benefits, numerous beneficiaries, global growth rate and future trends.

18.19.1. Concept of Paging

Paging literally means to summon by sending a page (i.e., attendant) to call someone by name. As a matter of fact, Radio paging is fundamentally a one-way personal selective calling system and relies on the availability of wire line telephone network to support the communication process. The user is always in contact even though he mobile or away from his telephone set.

When the call is sent out on a paging system, the pager (radio bell) carried by the person required immediately sounds a beep or gives a distinctive alert tone. In the simplest case, the person receiving the tone-alert signal has to contact back to take the actual message. Now pagers are available with additional facility of displaying messages in the form of numeric and alphanumeric form or conveying direct voice messages.

In earlier days, it was common practice to use public address systems for paging the required people instantly, whether in a big office or a factory. That was a simple beginning of paging. Later, such direct audio systems were replaced and improved by radio methods.

Radio paging emerged as a useful method of communication via radio in the mid-1950s. The problem of calling doctors/nursing staff silently without distressing patients was addressed by the radio technology of the day. In 1956, the first paging system used low frequencies like 30-50 kHz and its range was restricted to below 500 metres due to the use of inductive loop techniques. Inductive loop systems have the advantage of excellent spectrum conservation properties. However, the cost of installation, extension and coverage predictions makes such a system less attractive in relation to its radio alternatives.

During 1960s and 1970s on-site paging expanded rapidly with frequency allocation of 27-41 MHz and 470 MHz bands. Both amplitude and frequency modulations were used. The signalling systems were based on sequential tones in the audio frequency band of 300 MHz to 3 kHz. Paging service was initially limited to tone alert operation.

By the early 1970s, digital signalling methods became popular and paging became real [illegible]. In the early 1980s digital code formats were improved to permit paging of numeric, alphanumeric display and voice messages. Now there are many different proprietary digital signalling systems commercially available giving rise to the problem of non-compatibility between different manufacturers.

On-site paging provides service within a private or professional organisation. With a typical range of less than 1 km, on-site paging grew as an extension of the internal telephone, public address and intercom system business. However, application for the facilities offered in the on-site situations were soon found for much wider area systems requiring longer ranges. Fireman's call-out and medical answering services are best examples. The early system required limited receiver capability but excellent sensitivity to achieve wider range. Because of low capacity, they tend to be simple and less costly. The early wide area paging systems, unlike on-site systems, were seen as an extension of mobile radio business.

The significant milestones in the growth of the wide area paging are shown in figure 18.15. Bell Boy was first introduced in North America in 1963, followed by Eurocall in Europe in 1967 and Pocket Bell in Japan in 1968. The early systems used sequential tone signalling. They can be generally classified as 2-tone and 5-tone signalling methods. The EIA/Motorola 5-tone system [illegible].

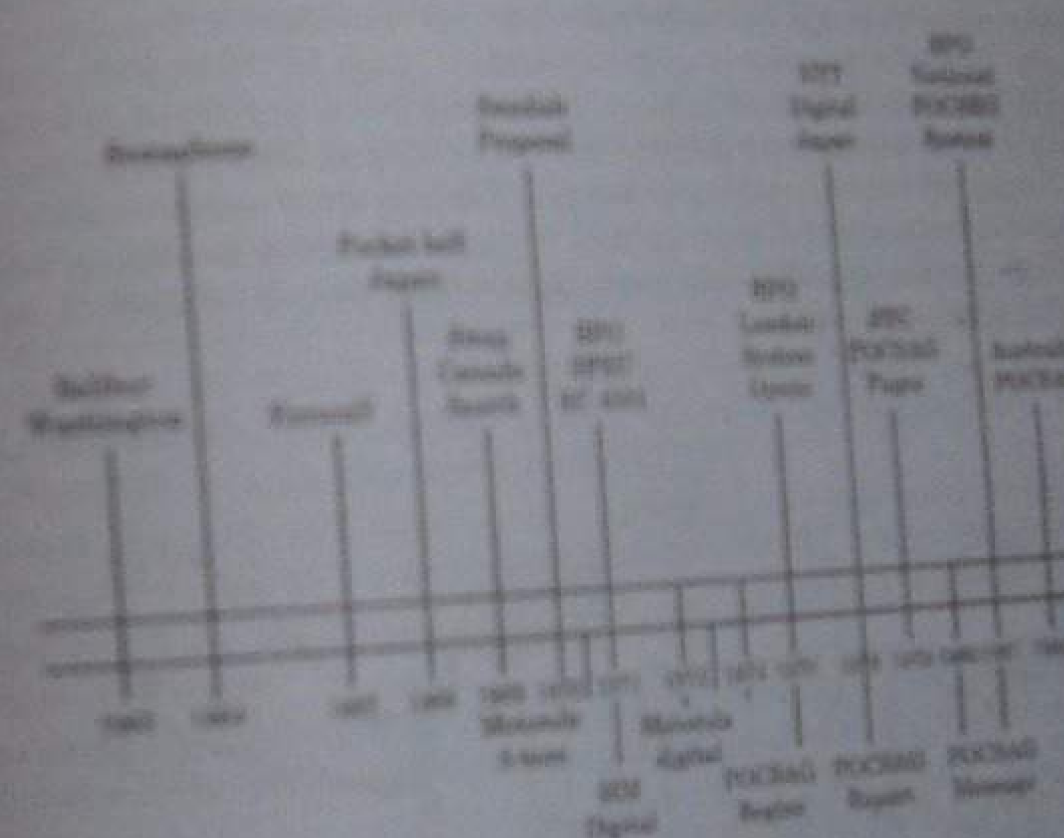


Fig. 18.15. *Illustration of Paging milestones*

The later developments were in digital signalling. The first truly binary digital paging system was introduced in 1970 in Canada and subsequently aroused great interest in such method of signalling. During the mid-1970s digital paging systems were developed by some of the major companies like the Swedish PTT and NTT, Japan.

However, during this period, one of the most adventurous single channel national paging system in the word was designed and planned in the UK by BPO (British Post Office). System wide area paging (SWAP) was initially proposed in which numerous proprietary digital paging codes could be accommodated by computer control. Such a system was commissioned in [illegible] in 1976. However, the BPO realised that it was necessary to have a transmission system using both simultaneous and sequential techniques to optimise traffic handling for a nation-wide network, while it was also important to standardise a paging code and a signalling format for channel utilisation and occupancy.

Thus, in December 1975, the Post Office Code Standardisation Advisory Group (POCSAG) was established. A code structure and format originally proposed by Philips Research Laboratories was finally accepted with some modifications. POCSAG paging code, published in 1978, satisfied ideal code properties for large national use, with an address capacity of eight million and pager capacity of two million. The credit for evolving an international paging code goes to British Post Officer.

In addition to the POCSAG (a British code in real sense), three other notable and rival codes which are in wide use are as under:

(i) GSC (Golay sequential code) is essentially of Americal origin. It is supported by the largest manufactures with a full range of products and is in volume service, notable in the USA.

(ii) NTT code is in high volume service in Japan and also sold by NEC in other countries.

(iii) MBS code is in use exclusively in Sweden.

18.19.2. The Components of a Paging System

The block diagram in figure 18.16 illustrates how a typical wide area public paging system works. The radio paging system basically consists of paging control terminal (PCT), radio base stations (subscriber units). PCT is usually installed in the centralised exchange premises for public paging. Radio transmitters are installed at suitable sites to take the advantage of antenna height and better radio coverage. Normally more than one transmitter is used to cover a wide area in a big city. Typical high power transmitter outputs range for 50 to 1000 watts for public wide area paging, while 5 watts is the maximum transmitter power permitted in India for private paging. Frequency bands available for paging cover both VHF and UHF ranges (30, 70, 150, 400 and 900 MHz bands). RF channel spacing is just kHz like any other single voice channel radio system. Paging broadcasts are received selectivey by the subscribers pssessing pager units. A typical PCT mainly contains of central processor, input-output processor (coders, storage units, queuing units etc.) and peripherals.

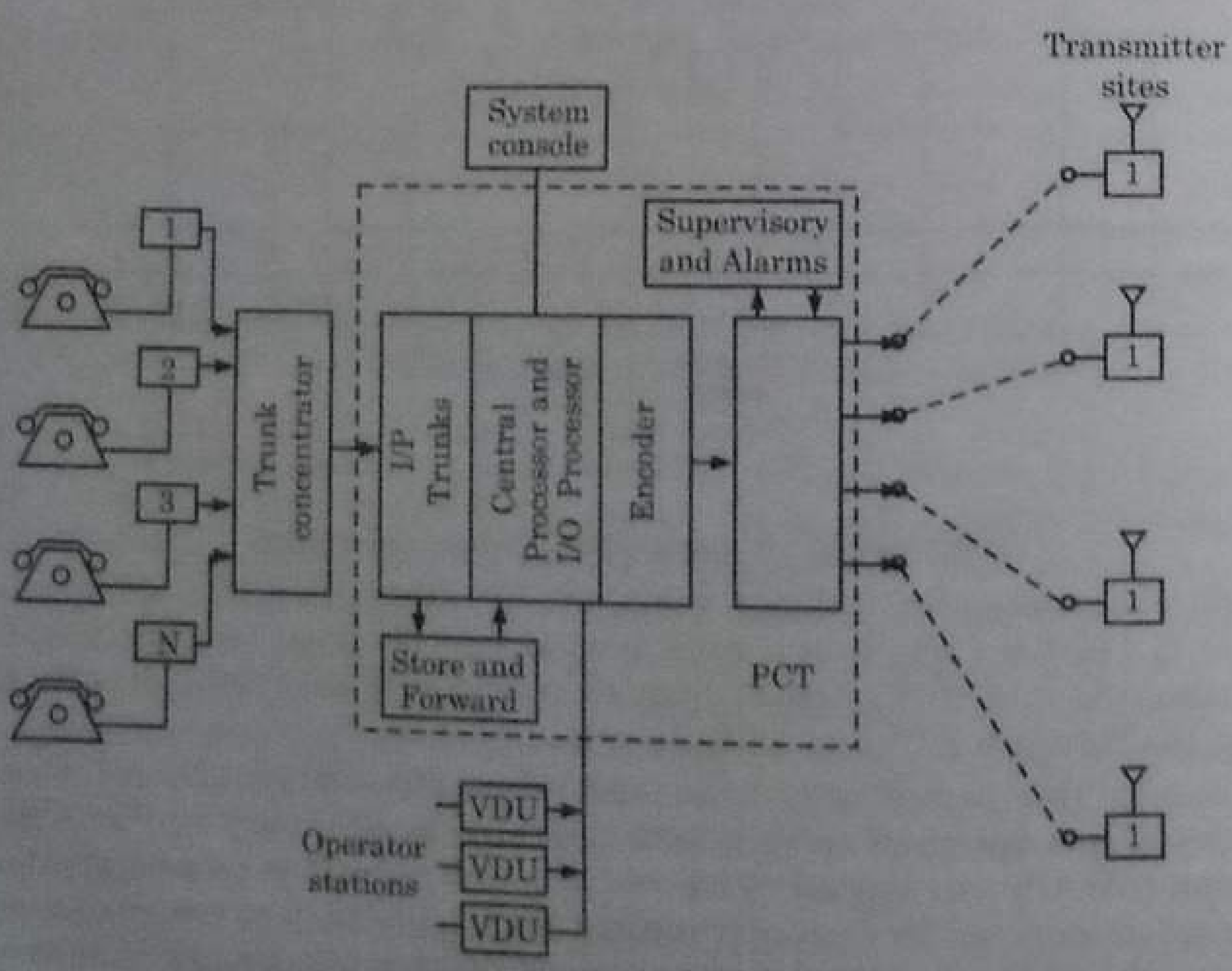


Fig. 18.16. *Illustration of wide area public system.*

Since PCT is the heat of the paging system, generally it is configured with full dynamic redundancy with auto change over facility to increase system reliability. Dynamic redundancy includes redundancy by of central processor and input-output processor with associated RAM and disc drives. For smaller and less important places redundant protection can be reduced from 1 + 1 mode (full redundancy on unit by unit basis) to 1 + 1 mode (one unit is redundant for N working units) or non-redundant configuration may resorted to for simplicity and economy.

PCT also provides centralised remote supervision, fault diagnosis, online testing, remote control, traffic analysis, billing etc. It generally has facilities for digital paging (numeric and alphanumeric messages). Some PCTs have facility for voice messages too. PCT is capable of being operated manually (through an operator control) or automatically when interfaced with the existing PSTN (public switching telephone network).

18.19.3. Operational Features of a Paging System

When a subscriber dials a pager number form any telephone with an intention of sending a message, PCT first receives call from the PSTN through various exchanges directly or through a trunk concentrator. PCT feeds back either a suitable tone or a voice announcement to the calling subscriber to indicate that the call is being processed. After decoding and recognising the pager number as that of a valid paging subscriber, the system returns a go-ahead signal to the calling subscriber to enter the actual message. Store and forward facility will record messages when the system is busy and then transmits as soon as it becomes free. Completion of message transmission is also conveyed to the caller.

The encoded data is distributed through a zone controller and is carried to multiple transmitter sites on VF (voice frequency) channel either by wire-line or radio. The electrical path lengths of these links should be properly equalised for compensating data delays. After modulation (generally FM is used), the transmitters broadcast the RF signals. In multiple transmitter zones, a single radio frequency channel is preferred so as to avoid multichannel recoveries. The transmitters can operate either sequentially or simultaneously. To avoid null zones between individual transmitter service areas, very high stability of RF carrier must be maintained.

PCT has facilities for group calling, priority paging, secure paging, repeat paging, greeting messages, canned messages, message retrieval etc.

Group call. In a group call, a page (same message) is sent simultaneously to multiple subscribers. Thus it is possible to call at a time several subscribers having common interest.

Priority page. By assigning priority status to any subscriber, his page will be transmitted before nonpriority pages.

Secure page. Secure paging is applicable to operator control systems. Each operator is assigned an identification code and password to gain access to the system, to protect unauthorised access to the system.

Repeat page. In repeat paging, repetition of the same message is done more than once for special subscribers requiring extra reliability to ensure that the message has successfully been received. For example, if a paging subscriber is inside a basement, due to penetration losses the pager would be out-of-range. As a first step, the pager provides an indication/warning. All messages transimitted during such times may be lost and repeat paging at programmed intervals will increase the probability of reception.

Greetings. With greeting facility, the caller can be greeted by prerecorded voice greetings. These greetings can be programmed by the subscriber.

Canned messages. Canned or predefined messages can be sent by dialling unique number for each message, on the lines of telegraphic greetings.

Message retrieval. Messages can be retrieved from the paging system by disclosing a password for identification, to verify messages or find out missing messages, if any.