ONLINE COURSE WARE (OCW)

for

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

Paper Name: Communication Engineering
Paper Code: CS (EC) 504C
Contacts: 3L

Credits: 3 **Total 32 Hours**

Prepared By,
Mr. Surajit Basak (GNIT Coordinator)
Mr. Kaushik Sarkar (NIT Member)
Dr. Indranath Sarkar (JISCE Member)

Prerequisites:

- Knowledge in different types of signals
- Exponential Fourier series
- Fourier transform and its properties
- Energy and power signal
- Probability and statistics

COURSE OBJECTIVES:

To present the fundamentals of analog and modern digital communication system design. Students should evaluate the performance of analog and digital signalling schemes on realistic communication channels. Emphasis is placed on physical layer digital communications and coding techniques, including waveform analysis, transmitter design and receiver design. The student will learn about theoretical bounds on the rates of digital data transportation systems.

COURSE OUTCOME:

Sem. No.	Course Title (Code)	CO Codes	Course Outcomes
			On completion of the course students will be able to
5th	Communication Engineering CS(EC)504C	CO. CS(EC)504C .1	Apply the fundamental concepts of engineering principles in design issues in various communication systems.
		CO. CS(EC)504C .2	Apply the basic concepts for analyzing the modulation techniques including amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM) that are widely used in analogue communication systems in the time and frequency domains.
		CO. CS(EC)504C .3	Demonstrate the concepts of sampling, Pulse Modulation techniques and their comparison.
		CO. CS(EC)504C .4	Design Matched filter, demonstrate the effects of Inter Symbol Interference (ISI) and compare Eye pattern analysis.
		CO. CS(EC)504C .5	Illustrate various types of coherent and non-coherent digital modulation techniques, analyse immunity parameters and calculate their error probabilities.

		CO. CS(EC)504C .6	Inspect recent trend and performance issues for different digital modulation techniques.

CO-PO MAPPING:

Sem. No.	Course Title (Code)	CO Codes	Program Outcomes (POs)											
			PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
5th	Communication Engineering CS(EC)504C	CO. CS(EC)504C .1	H	H	H		L	L			M			H
		CO. CS(EC)504C .2	H	H		H	H		M			L	M	H
		CO. CS(EC)504C .3	H	H	H	H	M	M			L			H
		CO. CS(EC)504C .4	H	H	H	M	H		M			M	M	H
		CO. CS(EC)504C .5	H	H		H	H	M						H
		CO. CS(EC)504C .6	H	H	H			M	M	H	L	M		H

Module-1
Elements of Communication system, Analog Modulation & Demodulation, Noise, SNR

Introduction to Analog Communication:

Communication is simply the process of conveying message at a distance or communication is the basic process of exchanging information. Typical examples of communication system are line telephony, radio telephony, radar communication, television broadcasting, radio aids to navigation, radio aids to aircraft landing etc. The communication based on analog signals and analog values is known as Analog Communication.

Elements of a communication system:

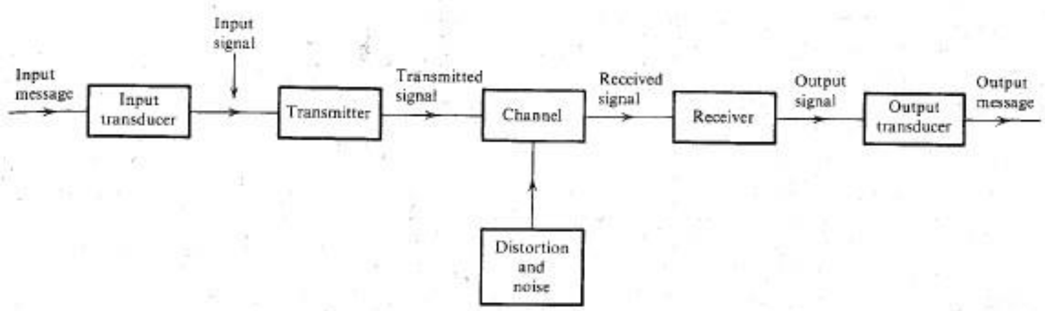


Fig. 1.1 Block diagram of a communication system

Concept of modulation:

For a signal to be transmitted to a distance, without the effect of any external interference or noise addition and without getting faded away, it has to undergo a process called as **Modulation**. It improves the strength of the signal without disturbing the parameters of the original signal.

A message carrying a signal has to get transmitted over a distance and for it to establish a reliable communication; it needs to take the help of a high frequency signal which should not affect the original characteristics of the message signal.

The characteristics of the message signal, if changed, the message contained in it also alters. Hence, it is a must to take care of the message signal. A high frequency signal can travel up to a longer distance, without getting affected by external disturbances. We take the help of such high frequency signal which is called as a **carrier signal** to transmit our message signal. Such a process is simply called as Modulation.

Modulation is the process of changing the parameters of the carrier signal, in accordance with the instantaneous values of the modulating signal.

Need for Modulation:

Baseband signals are incompatible for direct transmission. For such a signal, to travel longer distances, its strength has to be increased by modulating with a high frequency carrier wave, which doesn't affect the parameters of the modulating signal.

Advantages of Modulation

The antenna used for transmission, had to be very large, if modulation was not introduced. The range of communication gets limited as the wave cannot travel a distance without getting distorted.

Following are some of the advantages for implementing modulation in the communication systems.

- Reduction of antenna size
- No signal mixing
- Increased communication range
- Multiplexing of signals
- Possibility of bandwidth adjustments
- Improved reception quality

Signals in the Modulation Process:

Following are the three types of signals in the modulation process.

Message or Modulating Signal

The signal which contains a message to be transmitted is called as a **message signal**. It is a baseband signal, which has to undergo the process of modulation, to get transmitted. Hence, it is also called as the **modulating signal**.

Carrier Signal

The high frequency signal, which has a certain amplitude, frequency and phase but contains no information, is called as a **carrier signal**. It is an empty signal and is used to carry the signal to the receiver after modulation.

Modulated Signal

The resultant signal after the process of modulation is called as a **modulated signal**. This signal is a combination of modulating signal and carrier signal.

Types of Modulation:

There are many types of modulations. Depending upon the modulation techniques used, they are classified as shown in the following figure.

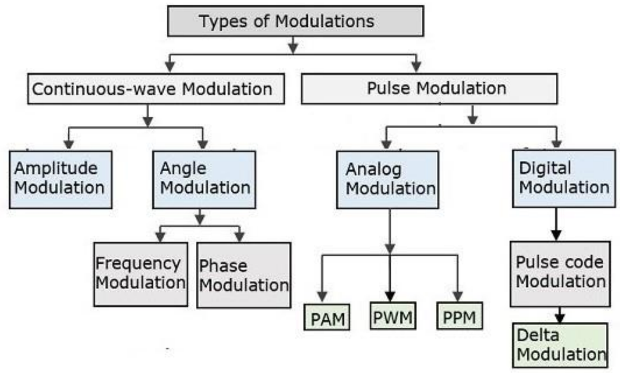


Fig. 1.2 Classification of Modulation

The types of modulations are broadly classified into continuous-wave modulation and pulse modulation.

Continuous-wave Modulation

In continuous-wave modulation, a high frequency sine wave is used as a carrier wave. This is further divided into amplitude and angle modulation.

- If the amplitude of the high frequency carrier wave is varied in accordance with the instantaneous amplitude of the modulating signal, then such a technique is called as **Amplitude Modulation**.

- If the angle of the carrier wave is varied, in accordance with the instantaneous value of the modulating signal, then such a technique is called as **Angle Modulation**. Angle modulation is further divided into frequency modulation and phase modulation.

    - If the frequency of the carrier wave is varied, in accordance with the instantaneous value of the modulating signal, then such a technique is called as **Frequency Modulation**.

    - If the phase of the high frequency carrier wave is varied in accordance with the instantaneous value of the modulating signal, then such a technique is called as **Phase Modulation**.

Pulse Modulation

- In Pulse modulation, a periodic sequence of rectangular pulses is used as a carrier wave. This is further divided into analog and digital modulation.
- In analog modulation technique, if the amplitude or duration or position of a pulse is varied in accordance with the instantaneous values of the baseband modulating signal, then such a technique is called as Pulse Amplitude Modulation (PAM) or Pulse Duration/Width Modulation (PDM/PWM), or Pulse Position Modulation (PPM).

In digital modulation, the modulation technique used is Pulse Code Modulation (PCM) where the analog signal is converted into digital form of 1s and 0s. As the resultant is a coded pulse train, this is called as PCM. This is further developed as Delta Modulation (DM).

Continuous Wave Linear Modulation:

A continuous-wave goes on continuously without any intervals and it is the baseband message signal, which contains the information. This wave has to be modulated.

According to the standard definition, “The amplitude of the carrier signal varies in accordance with the instantaneous amplitude of the modulating signal.” Which means, the amplitude of the carrier signal containing no information varies as per the amplitude of the signal containing information, at each instant. This can be well explained by the following figures.

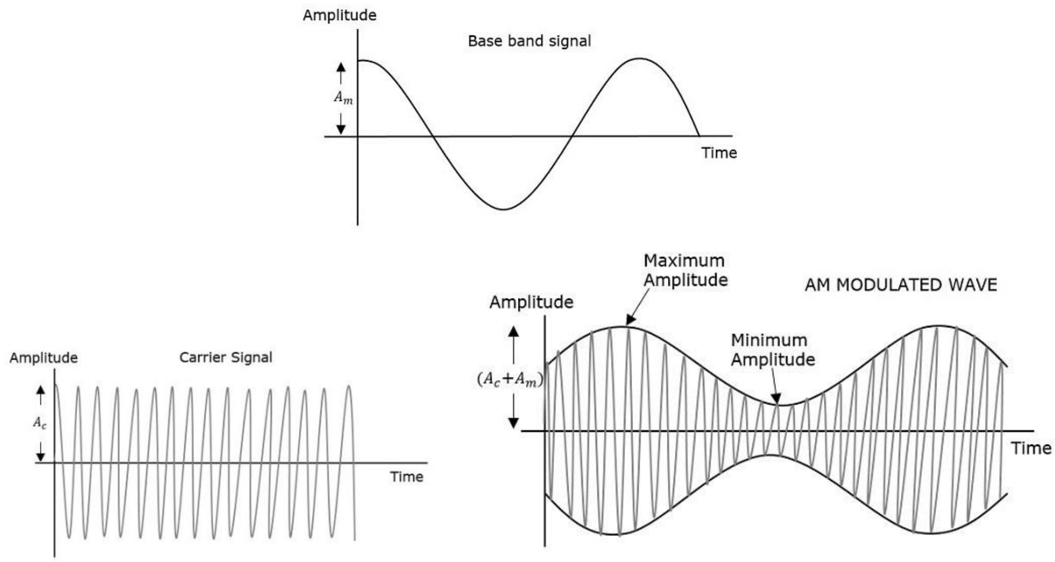


Fig. 1.3 Amplitude modulated waveform

The first figure shows the modulating wave, which is the message signal. The next one is the carrier wave, which is a high frequency signal and contains no information. While, the last one is the resultant modulated wave.

It can be observed that the positive and negative peaks of the carrier wave are interconnected with an imaginary line. This line helps recreating the exact shape of the modulating signal. This imaginary line on the carrier wave is called as **Envelope**. It is the same as that of the message signal.

Time-domain Representation of AM:

Let the modulating signal be,

$$m(t) = A_m \cos(2\pi f_m t)$$

and the carrier signal be,

$$c(t) = A_c \cos(2\pi f_c t)$$

Where, A_m and A_c are the amplitude of the modulating signal and the carrier signal respectively.

f_m and f_c are the frequency of the modulating signal and the carrier signal respectively.

Then, the equation of Amplitude Modulated wave will be

$$s(t) = [A_c + A_m \cos(2\pi f_m t)] \cos(2\pi f_c t) \tag{1.5}$$

Modulation Index

A carrier wave, after being modulated, if the modulated level is calculated, then such an attempt is called as **Modulation Index** or **Modulation Depth**. It states the level of modulation that a carrier wave undergoes.

Rearrange the (1) as below.

$$s(t) = A_c\left[1 + \left(\frac{A_m}{A_c}\right)\cos(2\pi f_m t)\right]\cos(2\pi f_c t)$$

$$s(t) = A_c[1 + \mu\cos(2\pi f_m t)]\cos(2\pi f_c t) \tag{1.6}$$

Where, μ is Modulation index and it is equal to the ratio of A_m and A_c. Mathematically, we can write it as

$$\mu = \frac{A_m}{A_c} \tag{1.7}$$

Hence, we can calculate the value of modulation index by using the above formula, when the amplitudes of the message and carrier signals are known.

Now, let us derive one more formula for Modulation index by considering (1). We can use this formula for calculating modulation index value, when the maximum and minimum amplitudes of the modulated wave are known.

Let A_{max} and A_{min} be the maximum and minimum amplitudes of the modulated wave.

We will get the maximum amplitude of the modulated wave, when $\cos(2\pi f_m t)$ is 1.

$$A_{max} = A_c + A_m \tag{1.8}$$

We will get the minimum amplitude of the modulated wave, when $\cos(2\pi f_m t)$ is -1.

$$A_{min} = A_c - A_m \tag{1.9}$$

Add (1.8) and (1.9)

$$A_{max} + A_{min} = A_c + A_m + A_c - A_m = 2\,A_c$$

$$A_c = \frac{A_{max} + A_{min}}{2} \tag{1.10}$$

Subtract (1.9) from (1.8)

$$A_{max} - A_{min} = A_c + A_m - A_c + A_m = 2\,A_m$$

$$A_m = \frac{A_{max} - A_{min}}{2} \tag{1.11}$$

The ratio of (1.11) and (1.10) will be as follows

$$\mu = \frac{A_m}{A_c} = \frac{(A_{max} - A_{min})}{(A_{max} + A_{min})} \tag{1.12}$$

Therefore, (1.7) and (1.12) are the two formulas for Modulation index. The modulation index or modulation depth is often denoted in percentage called as Percentage of Modulation. We will get the **percentage of modulation**, just by multiplying the modulation index value with 100.

For a perfect modulation, the value of modulation index should be 1, which implies the percentage of modulation should be 100%.

For instance, if this value is less than 1, i.e., the modulation index is 0.5, and then the modulated output would look like the following figure. It is called as **Under-modulation**. Such a wave is called as an **under-modulated wave**.

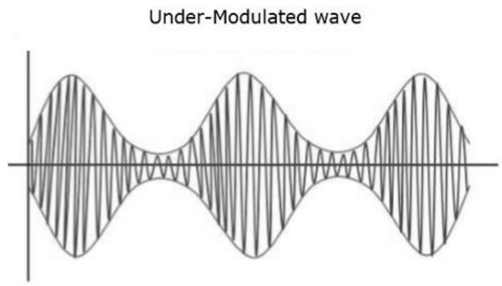


Fig. 1.4 Under-modulated AM waveform

If the value of the modulation index is greater than 1, i.e., 1.5 or so, then the wave will be an **over-modulated wave**. It would look like the following figure.

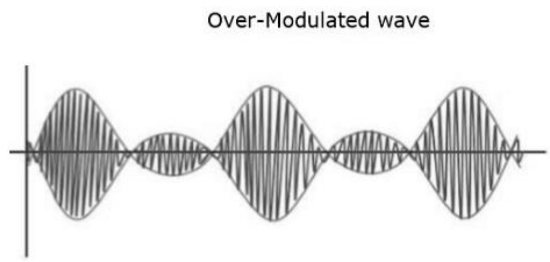


Fig. 1.5 Over-modulated AM waveform

As the value of the modulation index increases, the carrier experiences a 180° phase reversal, which causes additional sidebands and hence, the wave gets distorted. Such an over-modulated wave causes interference, which cannot be eliminated.

Bandwidth of AM Wave:

Bandwidth (BW) is the difference between the highest and lowest frequencies of the signal. Mathematically, we can write it as

$$BW = f_{max} - f_{min} \tag{1.13}$$

Consider the following equation of amplitude modulated wave.

$$s(t) = A_c[1 + \mu \cos(2\pi f_m t)] \cos(2\pi f_c t) \tag{1.14}$$

$$s(t) = A_c \cos(2\pi f_c t) + A_c \mu \cos(2\pi f_c t) \cos(2\pi f_m t) \tag{1.15}$$

$$s(t) = A_c \cos(2\pi f_c t) + \frac{A_c \mu}{2} \cos[2\pi(f_c + f_m)t] + \frac{A_c \mu}{2} \cos[2\pi(f_c - f_m)t] \tag{1.16}$$

Hence, the amplitude modulated wave has three frequencies. Those are carrier frequency f_c, upper sideband frequency $f_c + f_m$ and lower sideband frequency $f_c - f_m$.

Here, $f_{max} = f_c + f_m$ and $f_{min} = f_c - f_m$

Substitute, f_{max} and f_{min} values in bandwidth formula.

$$BW = (f_c + f_m) - (f_c - f_m) = 2f_m \tag{1.17}$$

Thus, it can be said that the bandwidth required for amplitude modulated wave is twice the frequency of the modulating signal.

Power Calculations of AM Wave:

Consider the following equation of amplitude modulated wave.

$$s(t) = A_c \cos(2\pi f_c t) + \frac{A_c\mu}{2}\cos[2\pi(f_c + f_m)t] + \frac{A_c\mu}{2}\cos[2\pi(f_c - f_m)t] \tag{1.18}$$

Power of AM wave is equal to the sum of powers of carrier, upper sideband, and lower sideband frequency components.

$$P_t = P_c + P_{USB} + P_{LSB} \tag{1.19}$$

We know that the standard formula for power of cos signal is

$$P = \frac{v_{rms}^2}{R} = \frac{(v_m/\sqrt{2})^2}{2} \tag{1.20}$$

Where,

v_{rms} is the rms value of cos signal.

v_m is the peak value of cos signal.

First, let us find the powers of the carrier, the upper and lower sideband one by one.

Carrier power, $P_c = \frac{(A_c/\sqrt{2})^2}{R} = \frac{A_c^2}{2R}$

$$\text{Upper sideband power, } P_{USB} = \frac{(A_c\mu/2\sqrt{2})^2}{R} = \frac{A_c^2\mu^2}{8R} \tag{1.21}$$

Similarly, we will get the lower sideband power same as that of the upper side band power.

$$P_{LSB} = \frac{(A_c\mu/2\sqrt{2})^2}{R} = \frac{A_c^2\mu^2}{8R} \tag{1.22}$$

Now, let us add these three powers in order to get the power of AM wave.

$$P_t = \frac{A_c^2}{2R} + \frac{A_c^2\mu^2}{8R} + \frac{A_c^2\mu^2}{8R} = \frac{A_c^2}{2R}\left[1 + 2 \times \frac{\mu^2}{4}\right] = P_c\left[1 + \frac{\mu^2}{2}\right] \tag{1.23}$$

We can use the above formula to calculate the power of AM wave, when the carrier power and the modulation index are known.

If the modulation index μ=1μ=1 then the power of AM wave is equal to 1.5 times the carrier power. So, the power required for transmitting an AM wave is 1.5 times the carrier power for a perfect modulation.

In the process of Amplitude Modulation, the modulated wave consists of the carrier wave and two sidebands. The modulated wave has the information only in the sidebands. **Sideband** is nothing but a band of frequencies, containing power, which are the lower and higher frequencies of the carrier frequency.

The transmission of a signal, which contains a carrier along with two sidebands, can be termed as **Double Sideband Full Carrier** system or simply **DSBFC**. It is plotted as shown in the following Fig. 1.6.

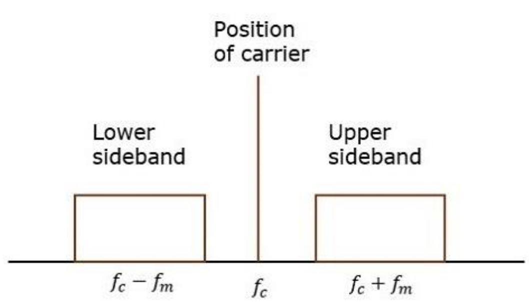


Fig. 1.6 Spectrum of DSB-FC

However, such a transmission is inefficient. Because, two-thirds of the power is being wasted in the carrier, which carries no information.

If this carrier is suppressed and the saved power is distributed to the two sidebands, then such a process is called as **Double Sideband Suppressed Carrier** system or simply **DSBSC**. It is plotted as shown in the following Fig. 1.7.

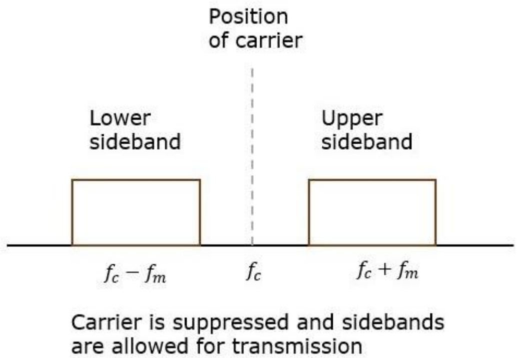


Fig. 1.7 Spectrum of DSB-SC signal

Mathematical Expressions

Let us consider the same mathematical expressions for modulating and carrier signals as we have considered in the earlier chapters.

i.e. Modulating signal,

$$m(t) = A_m \cos(2\pi f_m t) \tag{1.24}$$

Carrier signal

$$c(t) = A_c \cos(2\pi f_c t) \tag{1.25}$$

Mathematically, we can represent the **equation of DSBSC wave** as the product of modulating and carrier signals.

$$s(t) = m(t) \times c(t) \tag{1.26}$$

$$s(t) = A_m A_c \cos(2\pi f_m t)\cos(2\pi f_c t) \tag{1.27}$$

Bandwidth of DSBSC Wave

We know the formula for bandwidth (BW) is

$$BW = f_{max} - f_{min} \tag{1.28}$$

Consider the equation of DSBSC modulated wave.

$$s(t) = A_m A_c \cos(2\pi f_m t)\cos(2\pi f_c t) \tag{1.29}$$

$$s(t) = \frac{A_m A_c}{2}\cos[2\pi(f_c + f_m)t] + \frac{A_m A_c}{2}\cos[2\pi(f_c - f_m)t] \tag{1.30}$$

The DSBSC modulated wave has only two frequencies. So, the maximum and minimum frequencies are $f_c + f_m$ and $f_c - f_m$ respectively.

i.e., $f_{max} = f_c + f_m$ and $f_{min} = f_c - f_m$

Substitute, f_{max} and f_{min} values in the bandwidth formula.

$$BW = (f_c + f_m) - (f_c - f_m) = 2f_m \tag{1.31}$$

Thus, the bandwidth of DSBSC wave is same as that of AM wave and it is equal to twice the frequency of the modulating signal.

Power Calculations of DSBSC Wave:

Consider the following equation of DSBSC modulated wave.

$$s(t) = \frac{A_m A_c}{2}\cos[2\pi(f_c + f_m)t] + \frac{A_m A_c}{2}\cos[2\pi(f_c - f_m)t] \tag{1.32}$$

Power of DSBSC wave is equal to the sum of powers of upper sideband and lower sideband frequency components.

$$P_t = P_{USB} + P_{LSB} \tag{1.33}$$

We know the standard formula for power of cos signal is

$$P = \frac{v_{rms}^2}{R} = \frac{(v_m/\sqrt{2})^2}{R} \tag{1.34}$$

First, let us find the powers of upper sideband and lower sideband one by one.

Upper sideband power

$$P_{USB} = \frac{(A_m A_c/2\sqrt{2})^2}{R} = \frac{{A_m}^2 {A_c}^2}{8R} \tag{1.35}$$

Similarly, we will get the lower sideband power same as that of upper sideband power.

$$P_{LSB} = P_{USB} = \frac{{A_m}^2 {A_c}^2}{8R} \tag{1.36}$$

Now, let us add these two sideband powers in order to get the power of DSBSC wave.

$$P_t = \frac{{A_m}^2 {A_c}^2}{8R} + \frac{{A_m}^2 {A_c}^2}{8R} = \frac{{A_m}^2 {A_c}^2}{4R} \tag{1.37}$$

Therefore, the power required for transmitting DSBSC wave is equal to the power of both the sidebands.

Single Sideband Suppressed Carrier (SSB-SC):

The DSBSC modulated signal has two sidebands. Since, the two sidebands carry the same information, there is no need to transmit both sidebands. We can eliminate one sideband.

The process of suppressing one of the sidebands along with the carrier and transmitting a single sideband is called as **Single Sideband Suppressed Carrier** system or simply **SSBSC**. It is plotted as shown in the following Fig. 1.8.

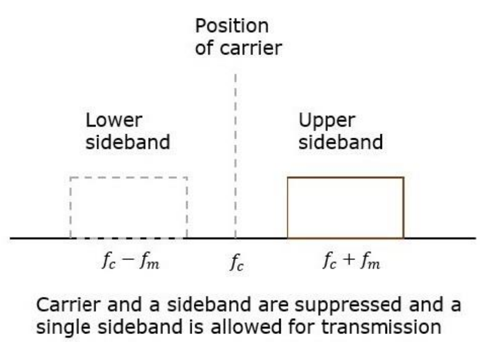


Fig. 1.8 Spectrum of SSB-SC signal

In the above figure, the carrier and the lower sideband are suppressed. Hence, the upper sideband is used for transmission. Similarly, we can suppress the carrier and the upper sideband while transmitting the lower sideband.

This SSBSC system, which transmits a single sideband, has high power, as the power allotted for both the carrier and the other sideband is utilized in transmitting this Single Sideband.

Mathematical Expressions

Let us consider the same mathematical expressions for the modulating and the carrier signals as we have considered in the earlier chapters.

i.e., Modulating signal

$$m(t) = A_m \cos(2\pi f_m t) \tag{1.38}$$

and the carrier signal,

$$c(t) = A_c \cos(2\pi f_c t) \tag{1.39}$$

Mathematically, we can represent the equation of SSBSC wave as

$$s(t) = \frac{A_m A_c}{2} \cos[2\pi(f_c + f_m)t] \qquad \text{for the upper sideband} \tag{1.40}$$

Or

$$s(t) = \frac{A_m A_c}{2} \cos[2\pi(f_c - f_m)t] \qquad \text{for the lower sideband} \tag{1.41}$$

Bandwidth of SSBSC Wave:

We know that the DSBSC modulated wave contains two sidebands and its bandwidth is $2f_m$. Since the SSBSC modulated wave contains only one sideband, its bandwidth is half of the bandwidth of DSBSC modulated wave.

i.e., ***Bandwidth of SSBSC modulated wave*** $\frac{2f_m}{2} = f_m$

Therefore, the bandwidth of SSBSC modulated wave is f_m and it is equal to the frequency of the modulating signal.

Power Calculations of SSBSC Wave:

Consider the following equation of SSBSC modulated wave.

$$s(t) = \frac{A_m A_c}{2} \cos[2\pi(f_c + f_m)t] \qquad \text{for the upper sideband} \tag{1.42}$$

Or

$$s(t) = \frac{A_m A_c}{2} \cos[2\pi(f_c - f_m)t] \qquad \text{for the lower sideband} \tag{1.43}$$

Power of SSBSC wave is equal to the power of any one sideband frequency components.

$$P_t = P_{USB} = P_{LSB} \tag{1.44}$$

We know that the standard formula for power of cos signal is

$$P = \frac{v_{rms}^2}{R} = \frac{\left(v_m/\sqrt{2}\right)^2}{R} \tag{1.45}$$

In this case, the power of the upper sideband is

$$P_{USB} = \frac{\left(A_m A_c/2\sqrt{2}\right)^2}{R} = \frac{{A_m}^2 {A_c}^2}{8R} \tag{1.46}$$

Similarly, we will get the lower sideband power same as that of the upper side band power.

$$P_{LSB} = \frac{{A_m}^2 {A_c}^2}{8R} \quad (1.47)$$

Therefore, the power of SSBSC wave is

$$P_t = P_{USB} = P_{LSB} = \frac{{A_m}^2 {A_c}^2}{8R} \quad (1.48)$$

Advantages:

- Bandwidth or spectrum space occupied is lesser than AM and DSBSC waves.
- Transmission of more number of signals is allowed.
- Power is saved.
- High power signal can be transmitted.
- Less amount of noise is present.
- Signal fading is less likely to occur.

Disadvantages:

- The generation and detection of SSBSC wave is a complex process.
- The quality of the signal gets affected unless the SSB transmitter and receiver have excellent frequency stability.

Applications:

- For power saving requirements and low bandwidth requirements.
- In land, air, and maritime mobile communications.
- In point-to-point communications.
- In radio communications.
- In television, telemetry, and radar communications.
- In military communications, such as amateur radio, etc.

Vestigial Side Band (VSB):

SSBSC modulated signal has only one sideband frequency. Theoretically, we can get one sideband frequency component completely by using an ideal band pass filter. However, practically we may not get the entire sideband frequency component. Due to this, some information gets lost.

To avoid this loss, a technique is chosen, which is a compromise between DSBSC and SSBSC. This technique is known as Vestigial Side Band Suppressed Carrier (VSBSC) technique. The word "vestige" means "a part" from which, the name is derived.

VSBSC Modulation is the process, where a part of the signal called as vestige is modulated along with one sideband. The frequency spectrum of VSBSC wave is shown in the following figure.

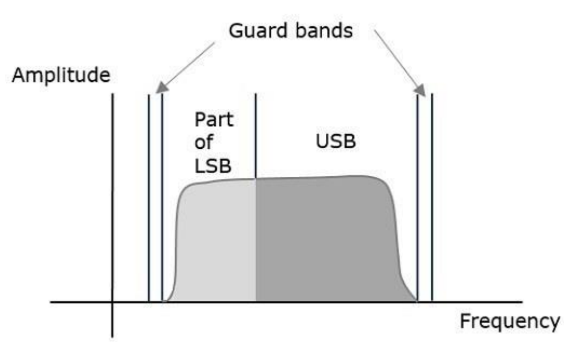


Fig. 1.9 Spectrum of VSB signal

Along with the upper sideband, a part of the lower sideband is also being transmitted in this technique. Similarly, we can transmit the lower sideband along with a part of the upper sideband. A guard band of very small width is laid on either side of VSB in order to avoid the interferences. VSB modulation is mostly used in television transmissions.

Bandwidth of VSBSC Modulation:
We know that the bandwidth of SSBSC modulated wave is f_m. Since the VSBSC modulated wave contains the frequency components of one side band along with the vestige of other sideband, the bandwidth of it will be the sum of the bandwidth of SSBSC modulated wave and vestige frequency f_v.

i.e., *Bandwidth of VSBSC Modulated Wave =* $f_m + f_v$

Advantages:
Following are the advantages of VSBSC modulation.

- Highly efficient.
- Reduction in bandwidth when compared to AM and DSBSC waves.
- Filter design is easy, since high accuracy is not needed.
- The transmission of low frequency components is possible, without any difficulty.
- Possesses good phase characteristics.

Disadvantages:
Following are the disadvantages of VSBSC modulation.

- Bandwidth is more when compared to SSBSC wave.
- Demodulation is complex.

Applications:
The most prominent and standard application of VSBSC is for the transmission of television signals. Also, this is the most convenient and efficient technique when bandwidth usage is considered.

Now, let us discuss about the modulator which generates VSBSC wave and the demodulator which demodulates VSBSC wave one by one.

Generation of amplitude modulated signal

Switching modulator
Efficient high level modulators are arranged so that undesired modulation products never fully developed & need not be filtered out. This can be accomplished with the help of switching modulator. The circuit diagram of switching modulator is shown below.

The carrier wave $e_c(t)$ is applied to the diode has been assumed to be of large amplitude so as to swing right across the characteristic curve of the diode. The diode is assumed to be ideal in the sense that it offers zero resistance in the forward direction ($e_c(t)>0$) & infinite resistance in the reverse direction ($e_c(t)<0$). Thus we approximate the transfer characteristics of the diode load resistor combination by a piece-wise linear characteristics as shown in the figure.
The input voltage $V_{in}(t)$ can be written as:

$$\begin{aligned} V_{in}(t) &= e_c(t) + e_m(t) \\ &= A_C \cos(2\pi f c t) + e_m(t) \end{aligned} \tag{2.1}$$

Where, mod $e_m(t) << A_C$

The resulting load voltage Vout(t) is

$$V_{out}(t) \approx V_{in}(t) e_c(t) > 0$$
$$\approx 0 \qquad e_c(t) < 0$$

This means that the load voltage Vout(t) varies periodically between the values V_{in} (t) & zero at a rate equal to carrier frequency f_c. Thus assuming the modulating wave to be weak campared to the carrier wave, we have effectively replaced the non-linear behavior of the diode by an approximately equivalent piecewise linear time varying operation & can be expressed as

$$V_{out}(t) \cong [A_C \cos(2\pi fct) + e_m(t)] g_p(t) \qquad \ldots(2.2)$$

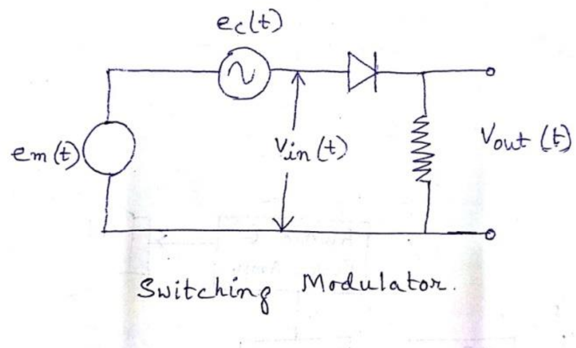


Figure 1.10: Switching Modulator

Where, g_p (t) is an periodic pulse train of duty cycle equal to one-half & period $T_0 = \frac{1}{fc}$

$$g_p(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos[2\pi f_c t(2n-1)]$$
$$\ldots(2.3)$$

Substituting equation (3)in equation (2) we find that the output voltage consists of the sum of two components:

i)The component

$$\frac{Ac}{2}\left[1 + \frac{4}{\pi Ac} e_m(t)\right] \cos(2\pi fct)$$

Which is the desired AM wave with amplitude sensitivity $k_a = 4/\pi Ac$

ii) An unwanted component, the spectrum of which contains delta functions at 0, $\pm$2fc, $\pm$ 4fc& so on and which occupies frequency intervals of width 2W centered at 0, $\pm$3fc, $\pm$ 5fc& so on, where W is the bandwidth of message signal. The unwanted terms can be removed from the load voltage V_{out}(t) by means of band pass filter with center frequency fc & bandwidth 2W.

Square Law Modulator

A square law modulator requires a means to add up the carrier & modulating signal, a non-linear element (like diode, transistor) and a band-pass filter to get the desired modulated wave. A schematic diagram is shown below:

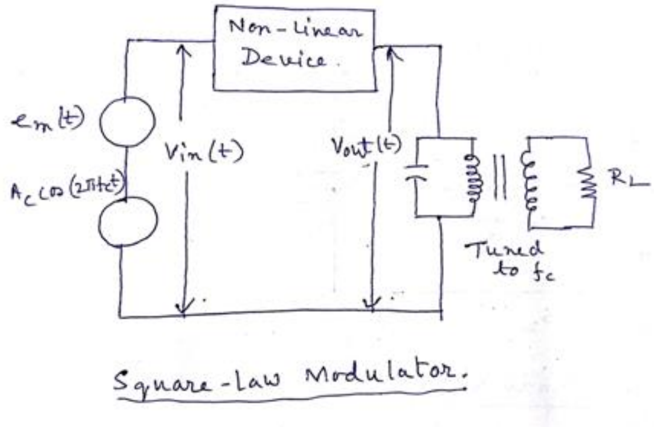


Figure 1.11: Square law modulator

The mathematical expression of square law modulator is expressed as:

$$v_{out}(\mathrm{t}) = a_1 v_{in}(\mathrm{t}) + a_2 v_{in}(\mathrm{t})^2 \qquad \ldots(2.4)$$

Where, a_1& a_2 are constants, vin(t)&Vout(t) are input & output voltages respectively.

The input voltage vin(t) consists of carrier & message signal, which is given by

$$v_{in}(\mathrm{t}) = A_c \cos(2\pi\mathrm{fct}) + e_m(\mathrm{t}) \qquad \ldots(2.5)$$

Substituting equation (2.5) in equation (2.4) we get,

$$v_{out}(\mathrm{t}) = a_1\mathrm{Ac}\ (1 + \frac{2a_2}{\mathrm{a1}}\mathrm{em(t)}\cos(2\pi\mathrm{fct}) + a_1 e_m(\mathrm{t}) + a_2 e_m(\mathrm{t})^2 + a_2\mathrm{Ac}^2\cos^2(2\pi\mathrm{fct}) \quad \ldots(2.6)$$

The 1st term in equation (2.6) is the AM wave with amplitude sensitivity,

$$\mathrm{k_a} = \frac{2\mathrm{a}2}{\mathrm{a}1}$$

The remaining three terms are unwanted & can be removed by Band-pass filter.

Balanced Modulator

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. The balanced modulator circuit is shown in Fig. 1.12

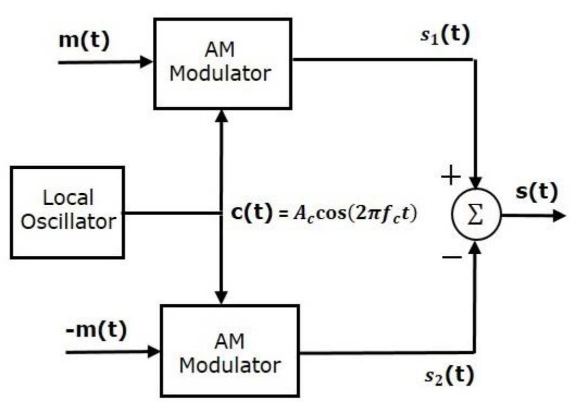


Figure 1.12: Balanced Modulator

One input to each modulator is from an oscillator which generates sinusoidal carrier. Other input is the message signal m(t). Note that the baseband signal applied to one of the modulators has a sign reversal. The output of the two AM modulators can be expressed as follows:

$$s_1(t) = A_C(1 + K_a m(t))\cos(2\pi f_c t)$$

And $$s_2(t) = A_C(1 - K_a m(t))\cos(2\pi f_c t)$$

Subtracting $s_2(t)$ from $s_1(t)$, we get

$$s(t) = s_1(t) - s_2(t) = 2A_c K_a m(t)\cos(2\pi f_c t)$$

Thus, except for a scaling factor $2K_a$, the balanced modulator output is equal to the product of the modulating wave & the carrier wave, which is nothing but the DSB-SC signal.

Generation of SSB

In DSB-SC it is observed that there is symmetry in the band structure. So, even if one half is transmitted, the other half can be recovered at the received. By doing so, the bandwidth and power of transmission is reduced by half. Depending on which half of DSB-SC signal is transmitted, there are two types of SSB modulation 1. Lower Side Band (LSB) Modulation 2. Upper Side Band (USB) Modulation

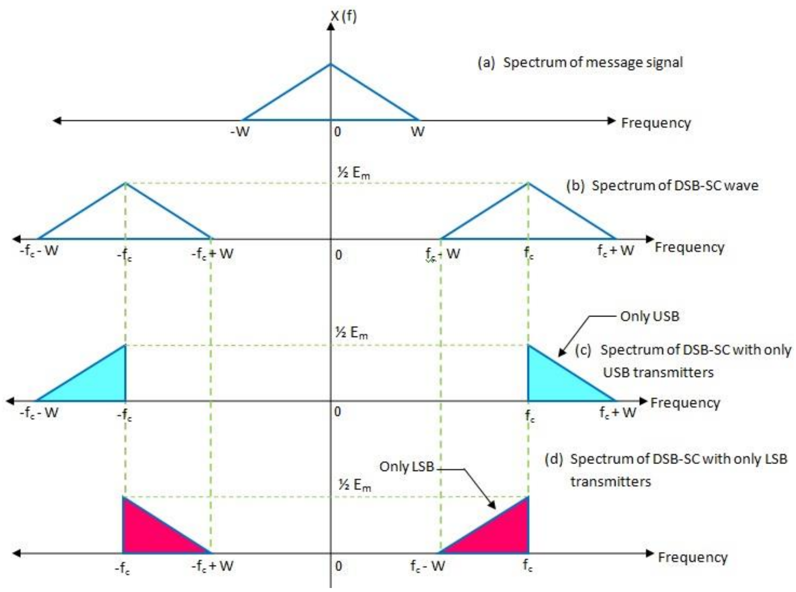


Figure 1.13: Generation of SSB signal

Filter method

An SSB signal can be derived by using of a suitable band pass filter - commonly called, an SSB sideband filter. This, the filter method, is probably the most common method of SSB generation. But these filters are generally only available at 'standard' frequencies (for example 455 kHz, 10.7 MHz) and SSB generation by the filter method at other frequencies can be expensive.

In this method a DSB-SC signal is generated simply by using a simple product modulator or balanced modulator. This DSB-SC contains two symmetrical sidebands. One of the sidebands is filtered out by BPF (Band pass filter). This is shown in the figure 1.14. It is important here to notice that to design of BPF is quite critical.

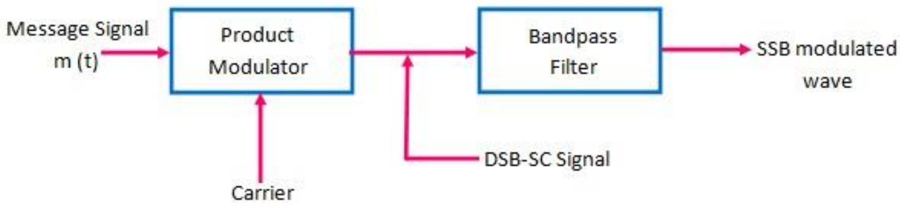


Figure 1.14: Filter method to generate SSB signal

The highest frequency component of the message signal m(t) is much less than the carrier frequency fc. Then, under these conditions, the desired side band will appear in a non-overlapping interval in the spectrum in such a way that it may be selected by an appropriate filter.

Phase shift method

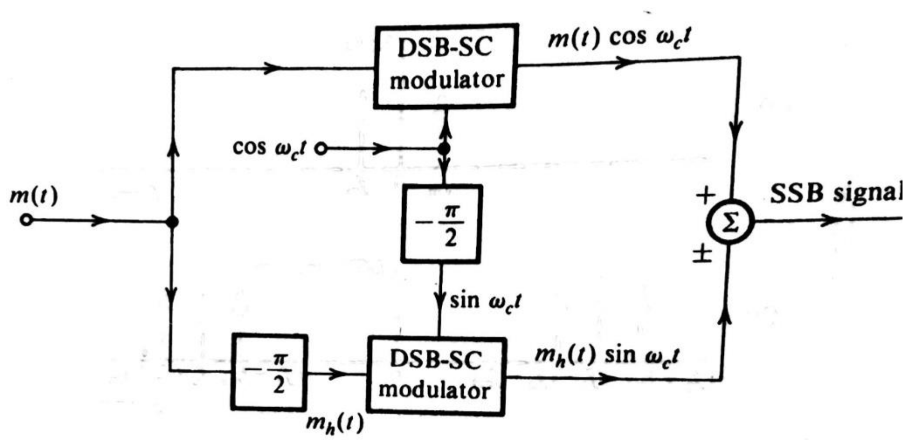


Figure 1.15: Phase shift method

The phase shift method uses the SSB signal as described by the following equation.

$$x(t)=m(t)\cos\omega ct \mp mh(t)\sin\omega ct$$

Figure 2.6 shows the implementation. The box marked π/2 is a phase shifter, which delays the phase of every positive spectral component by π/2. Hence, it is a nothing but a Hilbert transform. When the message m(t) has a dc null and very little low frequency content, the practical approximation of this ideal phase shifter has almost no real effect and does not affect the accuracy of SSB modulation.

Demodulation

The process of detection provides a means of recovering the modulating Signal from modulating signal. Demodulation is the reverse process of modulation. The detector circuit is employed to separate the carrier wave and eliminate the side bands. Since the envelope of an AM wave has the same shape as the message, independent of the carrier frequency and phase, demodulation can be accomplished by extracting envelope.

Demodulation of AM signals by envelope detector

The output of the envelop detector follows the envelope of the modulated signal. The circuit is shown in figure 2.8(b).On positive half-cycle of RF input signal the diode is forward biased and the capacitor charges up rapidly to the peak value of RF input signal. When RF input falls below the output voltage then the diode becomes reverse-biased and the capacitor discharges slowly through the load resistor.

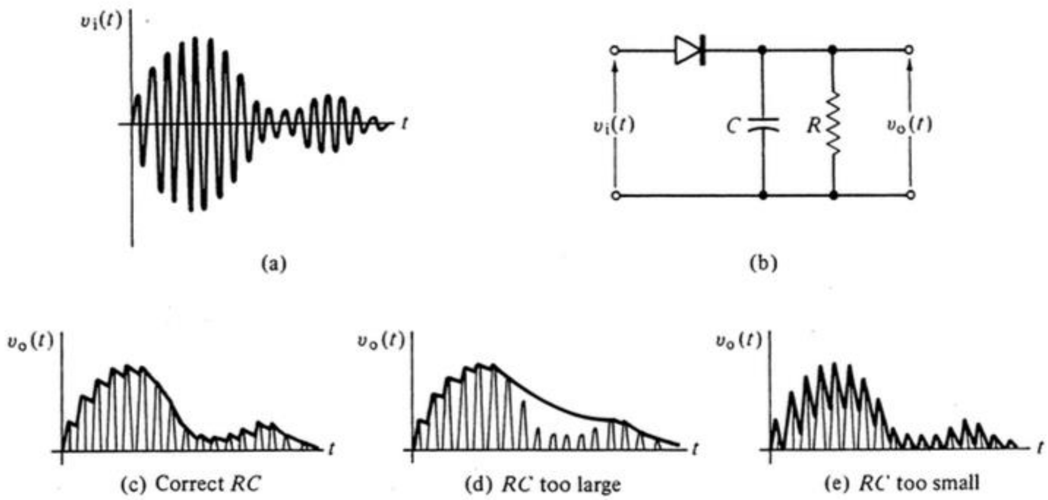


Figure 1.16: Envelope detector and the detected signals under different conditions

The charging and discharging depends on the RC time constant of the circuit. During each positive cycle, the capacitor charges up to the peak voltage of the input signal and decays slowly until the next positive cycle as shown in figure 1.16(c). The output voltage thus follows the envelope of the input AM signal as shown in figure 1.16(a). If RC value is too large, then the capacitor discharge very slow and as a result it can't track the actual envelope as shown in figure 1.16(d). On the other hand, very small RC may cause an unnecessary ripple during the tracking of the envelope as shown in figure 1.16(e) as the capacitor discharges too fast compared to the frequency of the input signal. So the selection of RC is an important factor for correct recovery of the message. Hence

$$\boldsymbol{T_c \ll \tau \ll T_m}$$

Where Tc is time period of carrier signal, T_m is the time period of modulating signal and τ is the time constant RC. If this condition prevails then output demodulated signal is the good approximation of the input message signal.

Synchronous detection

The coherent detection of the DSB-SC signal is shown in figure 2.10 .The DSB-SC wave s(t) is applied to a product modulator in which it is multiplied with the locally generated carrier cos ($2\pi f_c t$). We assume that this locally generated carrier is exactly coherent or synchronized in both frequency and phase with the original carrier wave c(t) used to generate the DSB-SC wave. This method of detection is therefore called as coherent detection or synchronous detection. The output of the product modulator is applied to the low pass filter (LPF) which eliminates all the unwanted frequency components and produces the message signal.

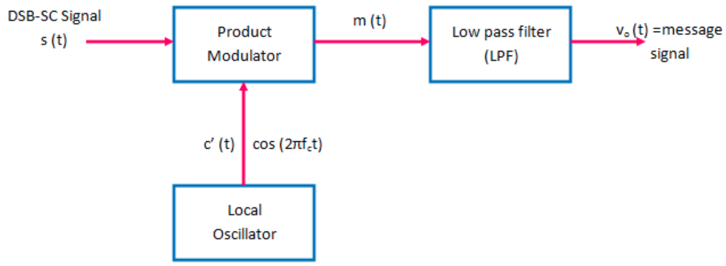


Figure 1.17: Synchronous detection of DSB-SC modulated wave

Let the output of the local oscillator be given by

$$c^{'}(t) = \cos(2\pi f_c t + \varphi)$$

The output of the product modulator is

$$m(t) = \ s(t)c^{'}(t)$$

where $s(t) = x(t)E_c \cos(2\pi f_c t)$
So the modulator output is given by

$$m(t) = \ x(t)E_c \cos(2\pi f_c t) \cos(2\pi f_c t + \varphi)$$

Hence

$$\cos(2\pi f_c t + \varphi) \cos(2\pi f_c t) = \frac{1}{2}[\cos(\ 2\pi f_c t + \varphi + 2\pi f_c t) + cos\ \varphi]$$

$$= \frac{1}{2}[\cos(4\pi f_c t + \varphi) + cos\ \varphi$$

Thus

$$m(t) = \frac{1}{2} x(t)E_c cos\varphi + \ \frac{1}{2}\ x(t)E_c \cos(4\pi f_c t + \varphi)$$

The signal m(t) is then passed through a LPF to suppress the high frequency term and only to get the first term of the above equation.

Therefore, the filter output is

$$v_o(t) = \frac{1}{2} E_c\ x(t) cos\ \varphi$$

Thus the output voltage of the synchronous demodulator is proportional to the message signal x(t) if the phase error $cos\ \varphi$ is constant.

Superheterodyne Receiver:
The radio receiver used in broadcast AM and FM system, is called the Superheterodyne receiver. In 1918, Edwin Armstrong invented the idea of a Heterodyne receiver. Heterodyning is the translation of a signal from a higher Radio Frequency (RF) carrier signal to a lower Intermediate Frequency (IF).
The block diagram is shown in figure 1.18. The RF section consists of tunable filter and amplifier that picks up the desired station. The frequency converter translates the carrier frequency ω_c to a fixed frequency ω_{IF}. For that the there is a local oscillator having frequency f_{LO} and they are related as $f_{LO}=f_c+f_{IF}$, where f_{IF}=455 KHz.
The entire selectivity is realized in IF section. The purpose of RF section is to suppression of image frequency.

The output of a mixer is the difference between the incoming (f_c) and the local oscillator frequency (f_{LO}) , so, $f_{IF}=f_{LO}-f_c$.

For example, if f_c=1000KHz, then $f_{LO}=f_c + f_{IF}$ =1000+455=1455 KHz. At the same time, another carrier frequency, say, $f1_c$=1455+455=1910 KHz. So along with the 1455 KHz, 1910 KHz station also be picked up by the RF section. The station 1910 KHz is known as the image of the station 1000 KHz. So the AM stations those are $2f_{IF}$=910 KHz apart are called the image stations (Figure 2.12). The RF filter has poor selectivity against a station separated by 10 KHz but has good selectivity for stations separated by 910 KHz.

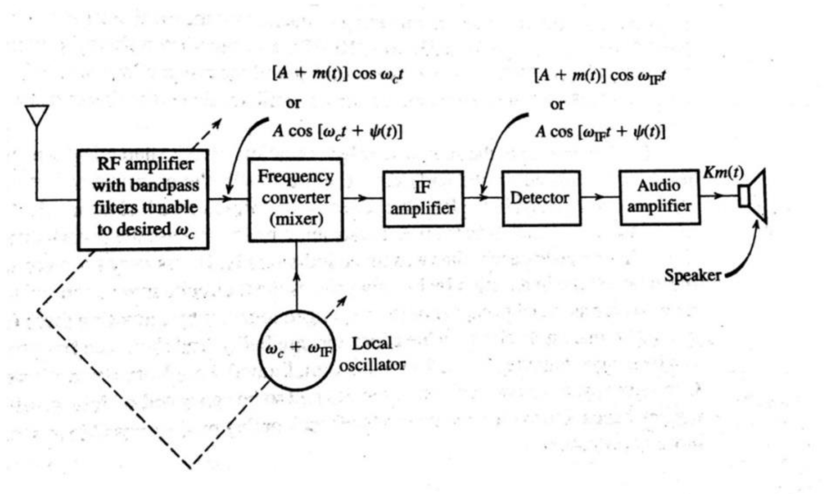


Figure 1.18: Superheterodyne Receiver

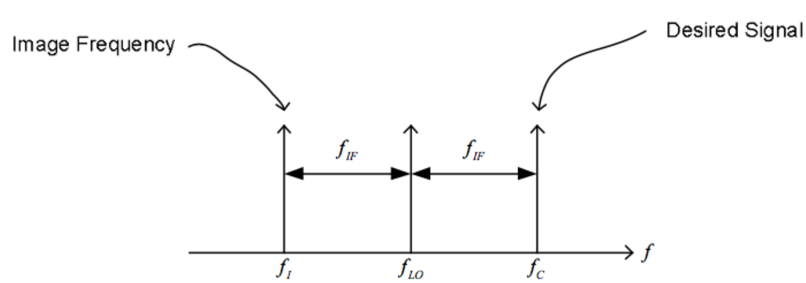


Figure 1.19: Image frequency

In the superheterodyne receiver, the received RF signal voltage is combined with the local oscillator voltage & is converted into a signal of lower fixed frequency. This frequency is called intermediate frequency(IF). The intermediate frequency signal contains the same modulation as the original carrier & can be subsequently amplified & demodulated to reproduce the original information. The additional units needed in a superheterodyne receiver include local oscillator, mixer & an IF amplifier.

A constant frequency difference is maintained between the local oscillator & the RF stage, through capacitance tuning in which capacitance are ganged together.

The IF Amplifier contains a no. of transformers each consisting of a pair of mutually coupled tuned circuits. The IF amplifier provides most of the gain & bandwidth requirements of the receiver. In other words, it determines sensitivity & selectivity of the receiver.

The IF frequency is represented by

$$f_{IF} = f_{LO} - f_c$$

where, f_{LO} is the frequency of local oscillator & f_c is the carrier received carrier frequency from RF stage.
The main purpose of RF amplifier stage is to select the desired frequency & reject other interfering frequencies such as image frequency.
The intermediate frequency is so called because it has a value intermediate between the received carrier frequency & the audio frequency.
The output of the IF amplifier is demodulated by using AM detector. The intelligent signal from the detector output is finally given audio amplification to drive the speaker.

Angle Modulation

Angle Modulation:

Let us consider a sinusoidal signal, $A_c \cos(2\pi f_c t + \varphi_0)$, where A_c is the (constant) amplitude, f_c is the (constant) frequency in Hz and φ_0 is the initial phase angle. Let the sinusoid be written as $A_c \cos[\theta(t)]$ where $\theta(t) = 2\pi f_c t + \varphi_0$. The case that we shall now examine where A_c is a constant but $\theta(t)$, instead of being equal to $2\pi f_c t + \varphi_0$, is a function of m(t) . This is known as the angle modulated signal. Two important cases of angle modulation are Frequency Modulation (FM) and Phase modulation (PM).

Features of Angle Modulation:

- They can provide much better protection to the message against the channel noise as compared to the linear (amplitude) modulation schemes.
- Due to constant amplitude nature, they can withstand nonlinear distortion and amplitude fading.

Limitations of Angle Modulation:

- Increased transmission bandwidth.
  To define Phase Modulation and Frequency Modulation let us consider a signal $s(t) = A_c \cos[\theta_i(t)]$ where $\theta_i(t)$, the instantaneous angle quantity, is a function of m(t) . We define the instantaneous frequency of the angle modulated wave s(t) , as $f_i = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt}$-----(3.1)
  (The subscript i in $\theta_i(t)$ or f_i is indicative of our interest in the instantaneous behavior of these quantities). If $\theta_i(t) = 2\pi f_c t + \varphi_0$ then $f_i(t)$ reduces to the constant f_c, which is in perfect agreement with our established notion of frequency of a sinusoid.

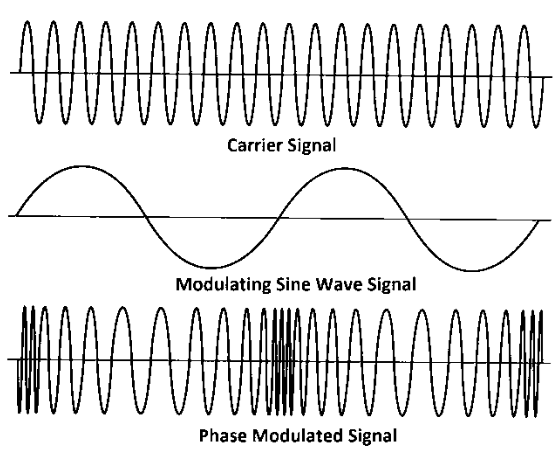


*Figure 1.20**Error! Use the Home tab to apply 0 to the text that you want to appear here.**: Time domain PM signal*

Phase modulation

For PM, $\theta_i(t)$ is given by $\theta_i(t)= 2\pi f_c t + k_p m(t)$--------(3.2)

Where, $2\pi f_c t$ is the angle of the un-modulated carrier and the constant k_p is the phase sensitivity of the modulator with the units, radians per volt. (For convenience, the initial phase angle of the un-modulated carrier is assumed to be zero). Using Eq. 3.2, the phase modulated wave s (t) can be written as $[s(t)]_{PM} = A_c \cos [2\pi f_c t + k_p m(t)]$ --------(3.3)

From Eq. 3.2 and 3.3, it is evident that for PM, the phase deviation of s(t) from that of the un-modulated carrier phase is a linear function of the base-band message signal, m(t).

The instantaneous frequency of a phase modulated signal depends on $\frac{dm(t)}{dt} = m'(t)$ because $\frac{1}{2\pi}\frac{d\theta i(t)}{dt} = f_c + \frac{k_p}{2\pi} m'(t)$

Frequency Modulation

Let us now consider the case where $f_i(t)$ is a function of m(t) ; that is,

$f_i(t) = f_c + k_f m(t)$ (3.4)

or, $\theta_i(t) = 2\pi \int_{-\infty}^{t} f_i(\tau)d\tau$ ----------(3.5)

or, $\theta_i(t) = 2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau$ ----------(3.6) k_f is a constant, which we will identify shortly. A frequency modulated signal s(t) is described in the time domain by $s(t)|_{FM} = A_c \cos(2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau)$ ----------(3.7),k_f is termed as the frequency sensitivity of the modulator with the units Hz/volt. From Eq. 3.4 we infer that for an FM signal, the instantaneous frequency deviation of s(t) from the (un-modulated) carrier frequency f_c is a linear function of m(t).

From these illustrations, we observe the following:

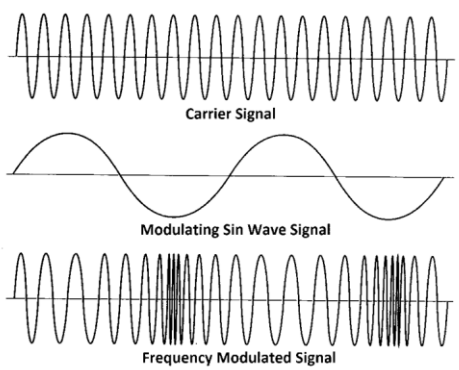


Figure 1.21: Time domain FM signal

i) Unlike AM, the zero crossings of PM and FM waves are not uniform (zero crossings refer to the time instants at which a waveform changes from negative to positive and vice versa).

ii) Unlike AM, the envelope of PM or FM wave is a constant.

Spectral representation of FM and PM:

Consider a single tone modulating sinusoid m (t) = $A_m \cos \omega_m t$

$$s(t)|_{FM} = A_c \cos(2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} A_m \cos \omega_m \tau d\tau)$$

$$s(t)|_{FM} = A_c \cos(2\pi f_c t + \frac{2\pi A_m k_f \sin \omega_m t}{\omega_m}) \text{--------(3.8)}$$

$$m_f = \frac{2\pi A_m k_f}{\omega_m} = \frac{k_f A_m}{f_m} = \frac{\Delta f|_{max}}{f_m} = \text{modulation index of FM}$$

Equ. 8, gives, $s(t)|_{FM} = A_c \cos(2\pi f_c t + m_f \sin \omega_m t)$

Carson's rule

Bandwidth of FM = $2(\Delta f|_{max} + f_m) = 2\left(k_f A_m + f_m\right) = 2\left(m_f + 1\right) f_m$--------(3.9)

- The bandwidth of a frequency modulated signal varies with both deviation and modulating frequency.
- Increasing modulating frequency reduces modulation index - it reduces the number of sidebands with significant amplitude and hence the bandwidth.
- Increasing modulating frequency increases the frequency separation between sidebands.
- The frequency modulation bandwidth increases with modulation frequency but it is not directly proportional to it.

NBFM

From equ. 8 we get $s(t)|_{FM} = A_c \cos(2\pi f_c t + m_f \sin \omega_m t)$

$$s(t)|_{FM} = A_c \cos(2\pi f_c t) \cos\left(m_f \sin \omega_m t\right) - A_c \sin(2\pi f_c t) \sin(m_f \sin \omega_m t)$$

In NBFM m_f<1. Thus $\cos\left(m_f \sin \omega_m t\right) = 1$ and $\sin\left(m_f \sin \omega_m t\right) = m_f \sin \omega_m t$

Then we get, $s(t)|_{FM} = A_c \cos(2\pi f_c t) - m_f A_c \sin(\omega_m t) \sin(2\pi f_c t)$

$= A_c \cos(2\pi f_c t) - \frac{A_c m_f}{2} \cos(\omega_c - \omega_m)t + \frac{A_c m_f}{2} \cos(\omega_c + \omega_m)t$----------(3.10)

WBFM

From equ. 8 we get $s(t)|_{FM} = A_c \cos(2\pi f_c t + m_f \sin \omega_m \tau)$

In WBFM m_f>1

Therefore, $s(t)|_{FM} = A_c \cos(2\pi f_c t) \cos\left(m_f \sin \omega_m t\right) - A_c \sin(2\pi f_c t) \sin(m_f \sin \omega_m t)$----(3.11)

$$\cos\left(m_f \sin \omega_m t\right) = J_0\left(m_f\right) + 2[J_2\left(m_f\right) \cos(2\omega_m t) + J_4\left(m_f\right) \cos(4\omega_m t) + \cdots]$$

And $\sin\left(m_f \sin \omega_m t\right) = 2[J_1\left(m_f\right) \sin(\omega_m t) + J_3\left(m_f\right) \sin(3\omega_m t) + J_5\left(m_f\right) \sin(5\omega_m t) + \cdots]$

J_n is a ***Bessel function*** of first kind. $J_n(m_f)$ is negligible if $n > m_f + 1$.

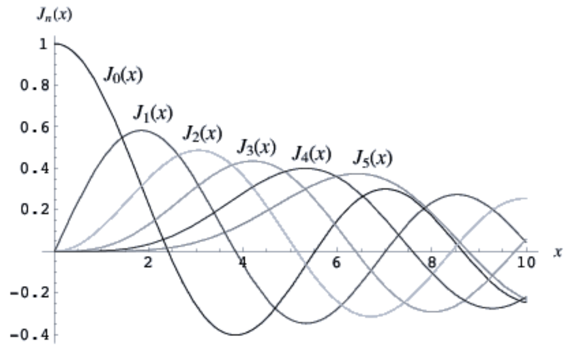


Figure 1.22: Bessel function

From equation (11) we can further get,

$$s(t)|_{FM} = A_c\cos(2\pi f_c t)\,[J_0(m_f) + 2[J_2(m_f)\cos(2\omega_m t) + J_4(m_f)\cos(4\omega_m t) + \cdots]] \\ - A_c\sin(2\pi f_c t)[2[J_1(m_f)\sin(\omega_m t) + J_3(m_f)\sin(3\omega_m t) \\ + J_5(m_f)\sin(5\omega_m t) + \cdots]]$$

$$s(t)|_{FM} = A_c[J_0(m_f)\cos(2\pi f_c t) + 2[J_2(m_f)\cos(2\pi f_c t)\cos(2\omega_m t) \\ + J_4(m_f)\cos(2\pi f_c t)\cos(4\omega_m t) + \cdots] - 2[J_1(m_f)\sin(2\pi f_c t)\sin(\omega_m t) \\ + J_3(m_f)\sin(2\pi f_c t)\sin(3\omega_m t) + J_5(m_f)\sin(2\pi f_c t)\sin(5\omega_m t) + \cdots]]$$

Since $2\pi f_c = \omega_c$,

$$s(t)|_{FM} = A_c[J_0(m_f)\cos(\omega_c t) + J_2(m_f)\,[\cos(\omega_c + 2\omega_m)t + \cos(\omega_c - 2\omega_m)t] \\ + J_4(m_f)\,[\cos(\omega_c + 4\omega_m)t + \cos(\omega_c - 4\omega_m)t] + \cdots \\ - J_1(m_f)[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] - J_3(m_f)[\cos(\omega_c - 3\omega_m)t \\ - \cos(\omega_c + 3\omega_m)t] + \cdots]$$

$$s(t)|_{FM} = A_c[J_0(m_f)\cos(\omega_c t) + J_2(m_f)\cos(\omega_c + 2\omega_m)t + J_2(m_f)\cos(\omega_c - 2\omega_m)t \\ + J_4(m_f)\cos(\omega_c + 4\omega_m)t + J_4(m_f)\cos(\omega_c - 4\omega_m)t + \cdots \\ - J_1(m_f)\cos(\omega_c - \omega_m)t + J_1(m_f)\cos(\omega_c + \omega_m)t - J_3(m_f)\cos(\omega_c - 3\omega_m)t \\ + J_3(m_f)\cos(\omega_c + 3\omega_m)t - \cdots]$$

Above expression of WBFM clearly shows that the un-modulated carrier signal is associated with $J_0(m_f)$ only. The message frequency ω_m is associated to $J_1(m_f)$ closest sideband components to the carrier frequencies. Remaining sideband components are distributed away from the carrier frequency and theoretically infinite in number.

Power Calculation:

From the above equations average power in WBFM can be calculated as given below, $P_{av} = \frac{[A_c J_0(m_f)]^2}{2} + 2\left[\frac{[A_c J_1(m_f)]^2}{2} + \frac{[A_c J_2(m_f)]^2}{2} + \cdots\right] = P_c[J_0(m_f)^2 + \sum_{n=0}^{\infty} J_n(m_f)^2]$

Where $P_c = \frac{A_c{}^2}{2}$

FM: Narrowband vs Wideband

- m_f<1 for Narrowband FM (NBFM) and m_f>1 for Wideband FM (WBFM)
- Maximumdeviation5 kHz in NBFM but 75 kHz in WBFM
- Range of modulating frequency : 30 Hz to 3 kHz in NBFM but 30 Hz to 15 kHz in WBFM
- Bandwidth is small in NBFM and same as of AM but in WBFM theoretically infinite.
- NBFM finds applications in mobile communication like police wireless, ambulance etc whereas WBFM is used in entertainment broadcasting.

Generation of Narrowband Angle Modulation

1. Armstrong's method

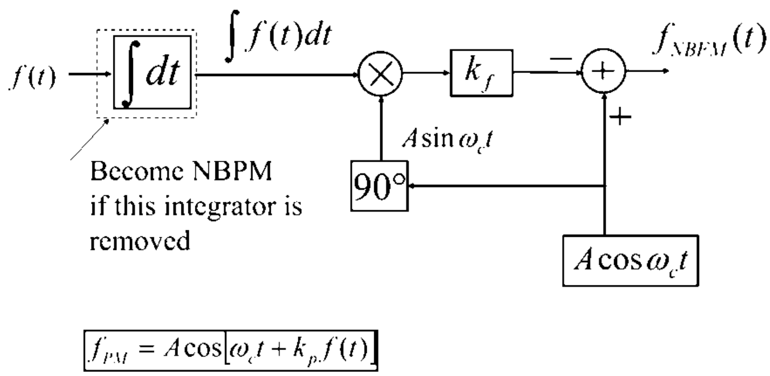


Figure 1.23: Armstrong's method of Narrow Band Angle Modulation

Generation of Wide-band angle modulation

1. Direct Method:

WBFM/WBPM is generated by passing the message or differentiated message through VCO respectively. This method is called as the Direct Method because a wide band FM/PM wave is generated by the device directly from message/differentiated message input. In this method, Voltage Controlled Oscillator (VCO) is used to generate WBFM. VCO produces an output signal, whose frequency is proportional to the input signal voltage. This is similar to the definition of FM wave. The block diagram of the generation of WBFM wave is shown in the following figure.

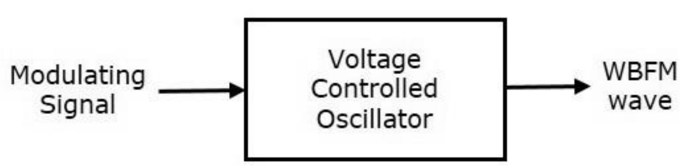


Figure 1.24: Voltage Controlled Oscillator

a. Angle Modulator VCO (Hartley Oscillator):

Hartley oscillator is a voltage controlled oscillator. Under the direct method of FM generation, the instantaneous frequency of the carrier wave is directly varied in accordance with instantaneous values of the message signal by means of voltage controlled oscillator. The frequency determining network in the oscillator is chosen with high quality factor (Q-factor) and the oscillator is controlled by the incremental variation of the reactive components in the tank circuit of the oscillator. A Hartley Oscillator can be used for this purpose.

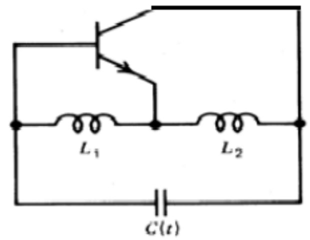


Figure 1.25: Hartley oscillator

The portion of the tank circuit in the oscillator is shown in Fig. 3.8. The capacitive component of the tank circuit consists of a fixed capacitor (with capacitance C_0) shunted by a voltage-variable capacitor (with frequency deviation of ΔC F/volt). The resulting capacitance is represented by C(t) in the figure. The voltage variable capacitor is commonly called as varactor / varicap. This is one whose capacitance depends on the voltage applied (say $\mathrm{Cos}(\omega_m t)$) across its electrodes. The varactor diode in the reverse bias condition can be used as a voltage variable capacitor. The larger the voltage applied across the diode, the smaller the transition capacitance of the diode. Thus changes the frequency of oscillation.

The frequency of oscillation of the Hartley oscillator is given by:

$$f_i(t) = \frac{1}{2\pi\sqrt{(L_1 + L_2)c(t)}}$$

Where the L_1 and L_2 are the inductances in the tank circuit and the total capacitance, c(t) is the fixed capacitor and voltage variable capacitor and given by:

$C(t) = C_0 + \Delta C \mathrm{Cos}(\omega_m t)$

$$f_i(t) = f_0\left[1 + \frac{\Delta c}{c_0}\cos(2\pi f_m t)\right]^{-\frac{1}{2}}$$

where $f_0 = \frac{1}{2\pi\sqrt{(L_1 + L_2)c_0}}$, since $\omega_m = 2\pi f_m$

$$\therefore f_i(t) = f_0\left[1 + \frac{\Delta c}{c_0}\cos(2\pi f_m t)\right]^{-\frac{1}{2}}$$

$$\cong f_0\left[1 - \frac{\Delta c}{2c_0}\cos(2\pi f_m t)\right]$$

Thus the instantaneous frequency $f_i(t)$ is defined as:

$$\therefore f_i(t) \cong f_0 + \Delta f \cos(2\pi f_m t)$$

The term, Δf represents the frequency deviation and the relation with Δc is given by:

$$\left(\frac{\Delta c}{2c_0} = -\frac{\Delta f}{f_0}\right)$$

Thus the output of the oscillator will be an FM wave.

But the direct method of generation has the disadvantage that the carrier frequency will not be stable as it is not generated from a highly stable oscillator. Generally, in FM transmitter the frequency stability of the modulator is achieved by the use of an auxiliary stabilization circuit as shown in the Fig. 1.26.

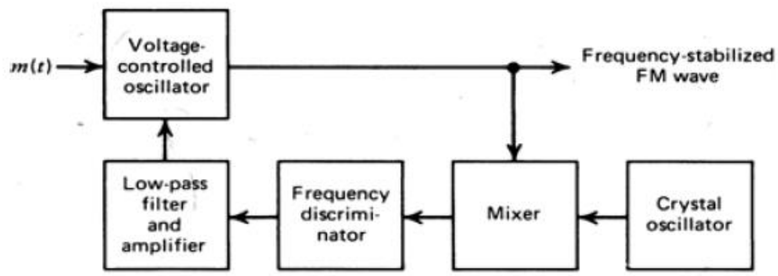


Figure 1.26: Frequency stabilized FM Modulator

The output of the FM generator is applied to a mixer together with the output of crystal controlled oscillator and the difference frequency signal is obtained. The mixer output is applied to a frequency discriminator, which gives an output voltage proportional to the instantaneous frequency of the FM wave applied to its input. The discriminator is filtered by a low pass filter and then amplified to provide a dc voltage. This dc voltage is applied to a voltage controlled oscillator (VCO) to modify the frequency of the oscillator of the FM generator.
The deviations in the transmitter carrier frequency from its assigned value will cause a change in the dc voltage in a way such that it restores the carrier frequency to its required value.

2. Indirect Method:
In indirect method message is passed through a narrowband frequency modulator followed by a frequency multiplier (Fig. 3.10).The job of the frequency multiplier is to increase the operating frequency to a desired value. A frequency multiplier consists of a nonlinear device followed by a band pass filter. If NBFM wave whose modulation index m_f is less than 1 is applied as the input of frequency multiplier, then the frequency multiplier produces an output signal, whose modulation index is ‘n’ times m_f and the frequency also ‘n’ times the frequency of WBFM wave.
Sometimes, we may require multiple stages of frequency multiplier and mixers in order to increase the frequency deviation and hence modulation index of FM wave.

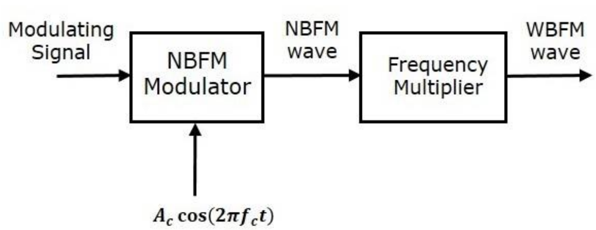


Figure 1.27: Indirect method of FM generation

Advantages of FM over AM:

1. Less radiated power.
2. Low distortion due to improved signal to noise ratio (about 25dB) w.r.t. to manmade interference.
3. Smaller geographical interference between neighboring stations.
4. Well defined service areas for given transmitter power.

Disadvantages of FM over AM:

1. Much more Bandwidth.
2. More complicated receiver and transmitter.

FM Demodulation:

Frequency demodulation is the process that enables us to recover the original modulating signal from a frequency modulated signal. Frequency Demodulator produces an output signal with amplitude directly proportional to the instantaneous frequency of FM wave.

Frequency demodulators are broadly classified into two categories:

(i) Direct method – examples: Phase Discriminators, Ratio Detector

(ii) Indirect method – example: phase locked loop.

The direct methods use the direct application of the definition of instantaneous frequency. The indirect method depends on the use of feed back to track variations in the instantaneous frequency of the input signal.

Slope Circuit:

This is a circuit in which the output voltage is proportional to the input frequency. An example is a differentiator. The output of the differentiator, x(t) = ds(t)/dt and the transfer function H(f)=j2πf (Fig. 3.11)

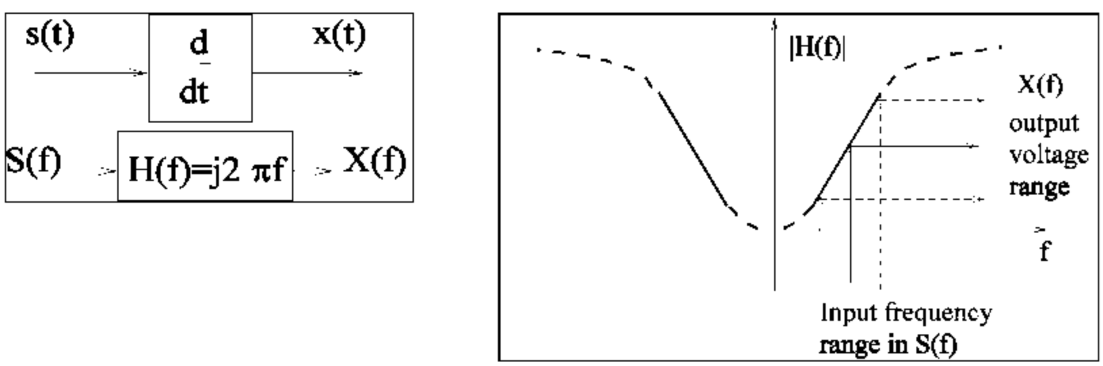
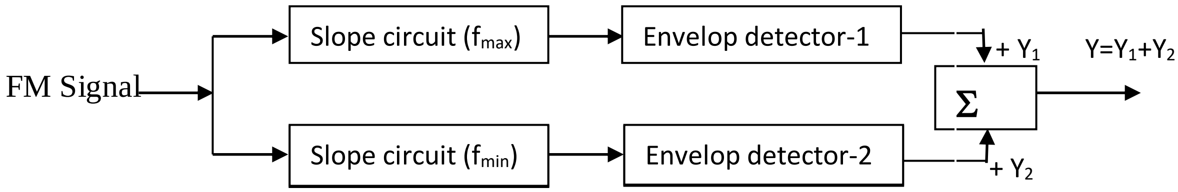


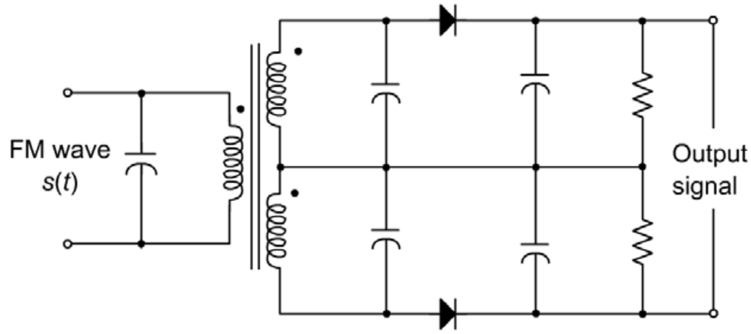
Figure 1.28: Differentiator as slope detector

Slope detector:

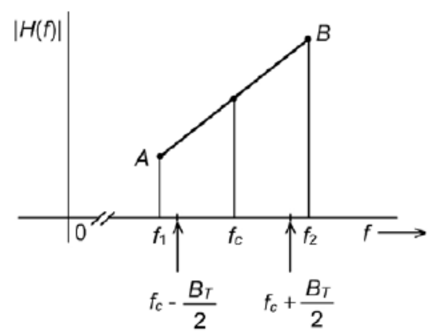
A slope detector circuit consists of two units: a slope circuit and an envelope detector. The slope circuit converts the frequency variations in the FM signal into a voltage signal, which resembles an AM signal. The envelope detector circuit obtains the output carrier signal of which magnitude is proportional to the message signal. The output of the slope circuit followed by envelope detector circuit is thus proportional to the message signal, m(t).A simple slope detector and envelope detector circuit arrangement is shown if Fig. 1.29.



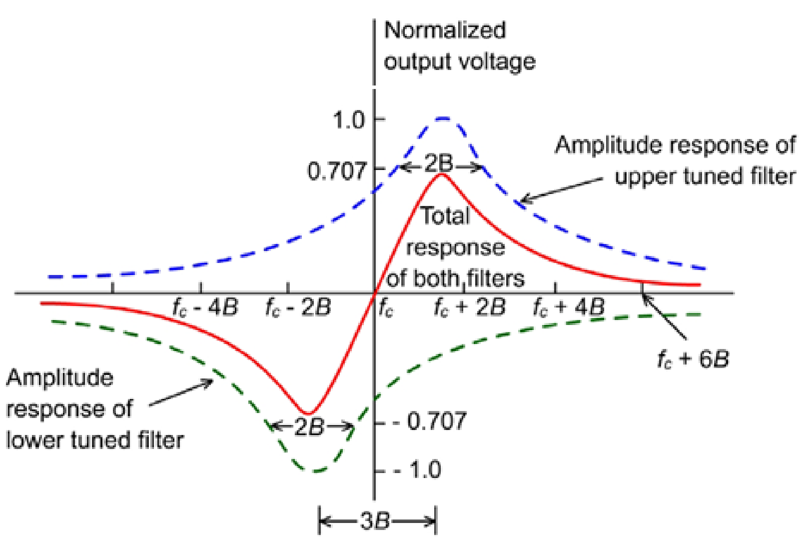
(a)



(a)



(b) $B_T=2(\Delta f_{max}+f_m)$



(b)

Figure 1.29: (a) direct method FM demodulation, (b) Transfer function of slope detector

Slope circuits are tuned to $f_{max}=f_c+\Delta f_{max}$ and $f_{min}=f_c-\Delta f_{max}$ respectively to extract the positive and negative excursion of message through envelope detector.
Finally the adder circuit at the end receives signal in either of its input at a time. Hence 'Y' will produce the total message.

Phase shift discriminator:
This method of FM demodulation involves converting frequency variations into phase variations and detecting the phase changes. In other words, this method makes use of linear phase networks instead of the linear amplitude characteristic of the circuits used in the previous method. Under this category, we have the Foster-Seely discriminator (and its variant the ratio detector).

1. Foster-Seely discriminator:
Fig. 3.13 illustrates the circuit diagram of this discriminator where all the resonant circuits involved are tuned to the same frequency. Note the similarity between this circuit and the circuit of Fig. 3.12a. Major differences are a by-pass capacitor C between the primary and secondary, an additional inductance L and only a single tuned circuit on the secondary (L2 ||C2).

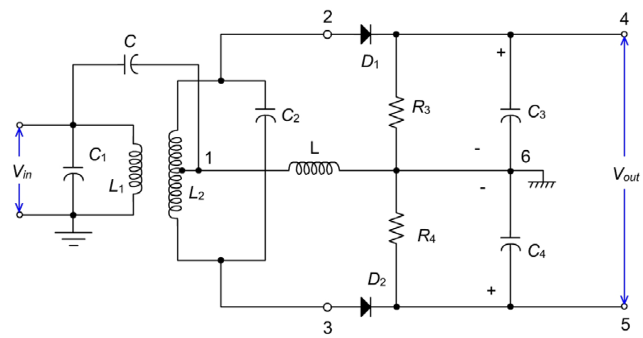


Figure 1.30: Circuit schematic of Foster-Seely discriminator

In the frequency band of operation, C, C3 and C4 are essentially short circuits, which implies that the entire input voltage Vin would appear across L. The mutually coupled double tuned circuit has high primary and secondary Q and low mutual inductance. When evaluating the primary current, we may neglect the primary resistance and any impedance coupled from the secondary. Hence, the primary current

$$I_p = \frac{V_{in}}{j\omega L_1}....(1)$$

(Note that all the voltages and currents are phasor quantities)
The voltage induced in series with the secondary as a result of the primary current is given by

$$V_s = \pm j\,\omega M I_p....(2)$$

Where, the sign depends on the direction of the winding. Taking the negative sign and using Eq. (1) in Eq. (2), we have

$$V_S = -\left(\frac{M}{L_1}\right)V_{in}$$

Assuming the diode circuit will draw very little current, we calculate the current in the secondary because of V_s as,

$$I_S = \frac{V_S}{R_2 + j(X_{L2} + X_{C2})}$$

Where, R_2 is the resistance associated with the inductance L_2 , $X_{L2} = \omega_{L2}$ and$X_{C2} = 1/\omega_{C2}$. Hence, the voltage across the terminals 2, 3 is given by

$V_{23}=I_S(-jX_{C2})=\frac{V_S(-jX_{C2})}{R_2+j(X_{L2-}X_{C2})} = \frac{jMV_{in}X_{C2}}{L_1(R_2+jX_2)}$where, $X_2=(X_{L2}-X_{C2})$

The voltage applied to diode D_1, is

$$V_{62}=V_L+V_{23}/2 = V_{in}+ V_{23}/2 \ldots.(3)$$

Similarly, the voltage applied to diode $D_2$is

$$V_{63} = V_{in} - V_{23}/2\ldots.(4)$$

The final output voltage V_{54} is, $V_{54} = V_{64} - V_{65}$ which is proportional to$\{|V_{62}|- |V_{63}|\}$. We will consider three different cases: $f = f_c$, $f > f_c$ and $f < f_c$.

i) When the input frequency $f = f_c$ we have

$$V_{23} = \frac{jMV_{in}X_{C2}}{L_1R_2} = j\left(\frac{MX_{C2}}{L_1R_2}\right)V_{in}\ldots.(5a)$$

That is, the secondary voltage V_{23} leads the primary voltage by 90°.

Thus $V_{23}/2$ will lead V_{in} by 90° and $V_{23}/2$ will lag V_{in} by 90°. Let us construct a phasor diagram by taking V_{in} as reference. This is shown in Fig. 3.14(a). As the magnitude of the voltage vectors applied to the diodes $D_1$and D_2 , V_{62} and V_{63} respectively are equal, the voltages V_{64} and V_{65} are equal and hence the final output V_{54} is zero.

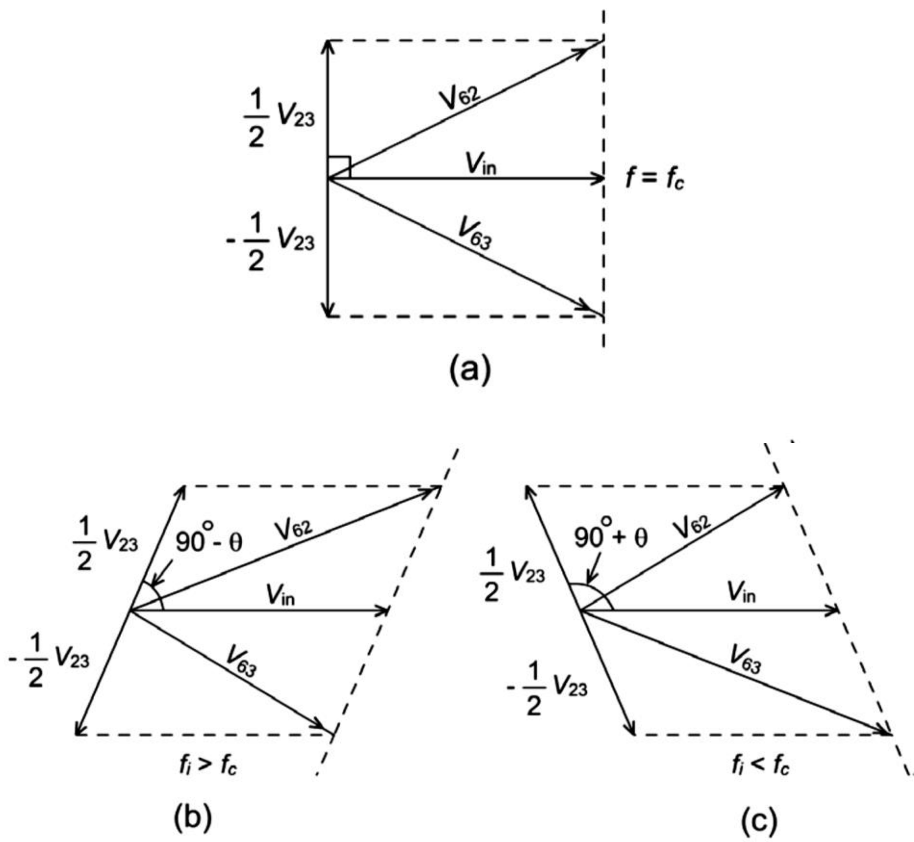


Figure 1.31: Phasor diagram illustrative of the operation of Foster-Seely discriminator

ii) when the input frequency exceeds f_c , $X_2=(X_{L2}-X_{C2})$ is positive. Let $R_2+jX_2=|Z_2|\exp(j\theta)$

then, $V_{23} = \frac{jMV_{in}X_{C2}}{L_1(R_2+jX_2)} = \left(\frac{MV_{in}X_{C2}}{L_1|Z_2|}\right)e^{j(90°-\theta)}\ldots.(5b)$

That is, V_{23} leads in V_{in} by less than 90° and $- V_{23}$ lags in V_{in} by more than 90°. This is shown in Fig. 3.14 (b). As the magnitude of the vector V_{62} is greater than that of V_{63}, $V_{64}> V_{65}$ which implies the final output $V_{54} = V_{64} - V_{65}$ is positive.

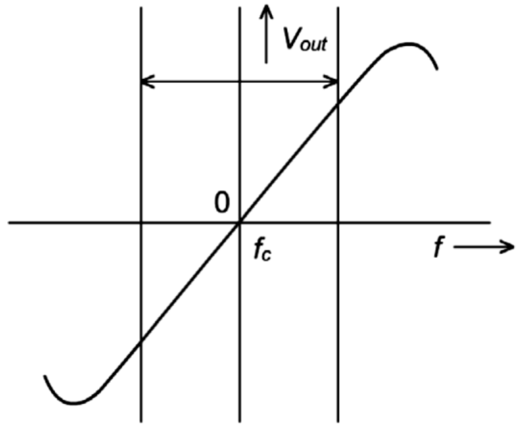


Figure 1.32: Response curve of the Foster-Seely

iii) Similarly, based on the phasor diagram of Fig. 1.31(c), we can easily argue that the final output would be negative when $f < f_c$. The actual value of the final output depends on how far away the input frequency is from f_c. Fig. 3.15 give the plot of the frequency response of the Foster-Seely discriminator, which is usually termed as the S-curve of the discriminator.

Useful range of the discriminator (frequency range of linear response, shown in Fig. 1.32) normally lies between the 3 dB points of the tuned circuit which forms part of the discriminator circuit. Foster-Seely discriminator responds also to input amplitude variations. Let the input to the discriminator be $f_i(t) = f_c$. Then, the voltages across R_3 and $R_4$are equal and let this value be 3V. Now, let $f_i(t)$ be such that voltage across R_3increases while that across R_4 decreases. Let the voltage increase on R_3 be 2Volts. We can take the voltage decrease on R_4 also as 2V. In other words, for the given frequency deviation, say Δf_1, we have the voltage at point 4 equal to 5volts whereas the voltage at point 5 equal to 1 V.

$V_{out} = (5 - 1) = 4$. Let V_{64} denote voltage across R_3 and V_{65}, the voltage across R_4. Then $\frac{V_{64}}{V_{65}} = 5$

Let the input signal strength be increased such that, when $f_i(t) = f_c, V_{64} = V_{65} = 6$. Now let $f_i(t)$ change such that we have the deviation Δf_1 as in the previous case. Then V_{64} will become 10 Volts whereas V_{65} becomes 2V, with their difference being equal to 8V. Though the ratio $\frac{V_{64}}{V_{65}}$ remains at 5, V_{out} changes from the previous value. That is, the circuit responds not only to frequency changes but also to changes in the incoming carrier strength. Hence, Foster-Seely discriminator has to be the preceded by a BPL.

2. **Ratio Detector:**

By making a few changes in the Foster-Seely discriminator, it is possible to have a demodulator circuit which has built in capability to handle the amplitude changes of the input FM signal, thereby obviating the need for an amplitude limiter. The resulting circuit is called the ratio detector which has been shown in Fig. 1.33.Comparing the ratio detector circuit with that of the Foster-Seely discriminator, we find the following differences: direction of D_2 is reversed, a parallel RC combination consisting of $(R_5 + R_6)$ and C_5 has been added and the output V_{out} is taken across a different pair of points. We shall now briefly explain the operation of the circuit.

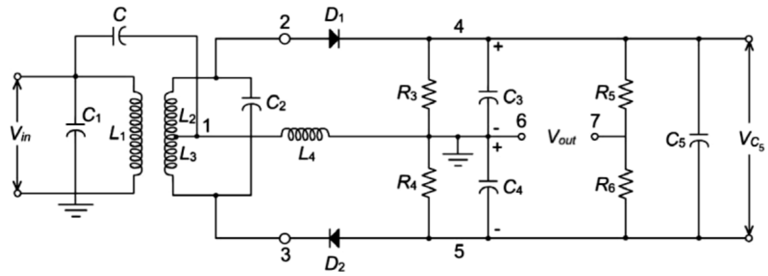


Figure 1.33: Circuit schematic of a ratio detector Reexamining

Fig. 1.33 and the corresponding phasor diagrams, we find that by and large, the sum $V_{62} + V_{63}$ remains constant. Hence, any variation in the magnitude of this sum voltage can be considered to be spurious. Suppression of these spurious variations would result in a detector that is unaffected by input voltage fluctuations which implies that circuit does not require a separate limiter stage. How the sum voltage is kept constant would be explained a little later. With the diode D_2 being reversed, we find that the voltages V_{64} and $V_{65}$are series aiding rather than series opposing and as such, the voltage V_{54} represents the sum voltage. Taking $R_5 = R_6$, we find

$V_{out} = V_{64} + V_{47} = V_{64} - V_{74}$

$V_{64} - V_{54}/2 = V_{64} - (V_{56}+V_{64})/2 = k(|V_{62}|-|V_{63}|)$.

Usually, $C_3 = C_4$ and $R_3 = R_4$. Hence at resonance, $V_{64} = V_{56}$ which implies that V_{out} is zero. Above resonance, as $V_{64} > V_{56}$, the output is positive whereas below resonance $V_{56} > V_{64}$, and the output is negative. In the circuit Fig. 1.33, C_5 is a capacitor of a rather large value. For example, C_5 is of the order of 5 μF whereas C_3 and C_4 are of the order 300pF. If V_{in} is constant,C_5 charges to the full potential existing between the points5 and 4, which, as indicated earlier is essentially a constant. If V_{in} tries to increase, C_5 will tend to oppose any rise in V_{out}. This is because as the input voltage tries to rise, extra diode current flows trying to charge C_5. But V_{54}remains constant at first because C_5 is a fairly large capacitance and it is not possible for the voltage across it to change instantaneously. The situation now is that the current in the diodes' load has risen but the voltage across the load has not changed. This being so, the secondary of the ratio detector transformer is more heavily damped, Q falls and so does the gain of the amplifier driving the detector. This nearly counteracts the rise in the input voltage. Similarly, when V_{in} increases, the damping is reduced. The gain of the driving amplifier increases, thereby counteracting the fall in the input voltage. Thus the ratio detector provides variable damping.

3. Phase Locked Loop:

PLL is a versatile building block of the present day communication systems. Besides FM demodulation, it has a large number of other applications such as carrier tracking, timing recovery, frequency synthesis etc. The basic aim of a PLL is to lock (or synchronize) the instantaneous angle of a VCO output to the instantaneous angle of a signal that is given as input to the PLL. In the case of demodulation of FM, the input signal to PLL is the received FM signal. In its simplest form, PLL consists of a phase detector and a VCO connected as shown in Fig. 1.34 (a).

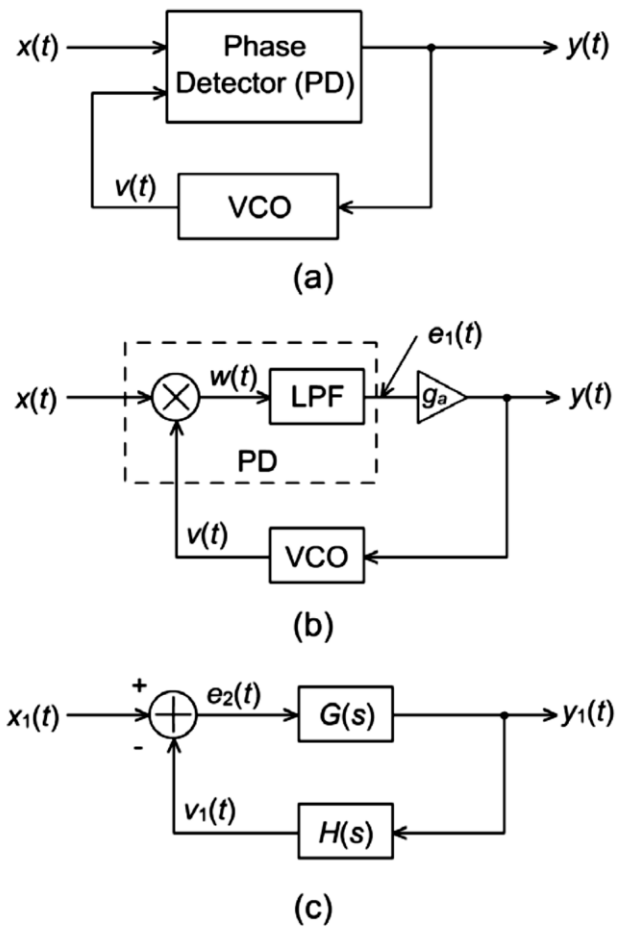


Fig. 1.34: Phase lock loop (a) Basic configuration
(b) PD of (a) in functional form
(c) PLL as a negative feedback loop

PD makes the comparison of the instantaneous phase of x(t) and v(t),and is designed such that θv(t), the instantaneous phase of v(t) locks on toθx(t) , the instantaneous phase of x(t) , if necessary with some fixed phase difference. (This will become evident later.)A number of circuits are available which have been used as phase detectors. In the context of FM demodulation, the most common PD is that of an analog multiplier followed by a LPF (Fig. 3.17(b)). The scheme resembles closely that of a negative feedback amplifier configuration, shown in Fig. 3.17(c). In this figure, s is the variable of the Laplace transform, G(s) is system function in the forward path whereas H(s) is the network in the feedback path. A properly designed negative feedback system ensures that the error quantity e2 (t) is fairly close to zero so that $v_1(t) = x_1(t)$. This is ensured by providing sufficiently high loop gain. Similarly, by making the amplifier gain g_a sufficiently large, it is possible to make $\theta_v(t)$ follow the changes in $\theta_x(t)$.

Let $x(t) = \cos[\omega_c t + \phi(t)]$ and let the VCO output be, $v(t) = \cos[\omega_c t + \psi(t)]$. Then, from Fig. 3.17(b), $w(t) = x(t)v(t) = \cos[\omega_c t + \phi(t)]\cos[\omega_c t + \psi(t)]$

$$w(t) = x(t)v(t) = \frac{1}{2}\{\cos[2\omega_c t + \phi(t) + \psi(t)] + \cos[\phi(t) - \psi(t)]\}$$

Only the term $\frac{1}{2}\{\cos[\phi(t) - \psi(t)]\}$ will appear at the output of the LPF;

That is $e_1(t) = \frac{1}{2}\{\cos[\phi(t) - \psi(t)]\}$….(1)

(We are assuming that the LPF has unit gain)As the phase detector, we want e(t) to be zero where $\phi(t) = \psi(t)$; but from Eq. 1, $e_1(t)$ is maximum when $\phi(t) = \psi(t)$. This is not the characteristic of a proper phase detector. This anomaly can be corrected, if the loop provides a $\pi/2$ phase shift so that the output of the VCO is $\sin[\omega_c t + \psi(t)]$. That is, the loop locks in phase quadrature. Here after, we shall assume this $\pi/2$ phase shift in the VCO output.

Now let us look at the demodulation of FM. Let

$x(t) = s(t) = \cos[\omega_c t + \phi(t)]$

where $\phi(t) = 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau$

The VCO is designed such that when y(t), the control voltage is zero, its frequency is f_c. (As mentioned earlier, in a superhet, the demodulator follows the IF stage and f_c is actually f_{IF}, which is a known quantity). This is called the free running frequency of the VCO. Hence, the VCO output can be written as

$v(t) = A_v \sin[\omega_c t + \psi(t)]$….(2a)

$\psi(t)=\int 2\pi k_v y(\tau)d\tau$….(2b)

and k_v is the voltage sensitivity of the VCO, in units of Hz/volt.

Or $\psi'(t) = 2\pi k_v y(t)$ ….(2c)

$= K_1 y(t)$, where $K_1 = 2\pi k_v$

$e_1(t)$ of Fig. 3.17(b) is,$e_1(t) = \frac{A_c A_v}{2}\sin[\phi(t) - \psi(t)] * h(t)$….(3a)

where h(t) is the impulse response of the LPF,

and $y(t) = g_a\, e_1(t)$….(3b)

Let $\theta_e(t) = \phi(t) - \psi(t)$ ….(4)

Then, $y(t) = \frac{A_c A_v g_a}{2}\sin[\theta_e(t)] * h(t)$

Or, $y(t) = K_2 \sin[\theta_e(t)] * h(t)$, where $K_2 = \frac{A_c A_v g_a}{2}$….(5)

Using Eq. 4, 5and 2(c), we can draw following block diagram (Fig. 3.18) for the PLL.

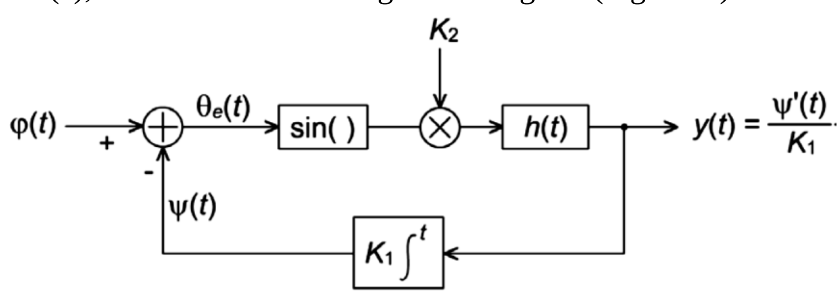


Figure 1.35: Equivalent circuit of a PLL

Fig. 1.35 brings out much more clearly the negative feedback nature of the PLL, where in the quantities involved are instantaneous phase deviations of the input and the VCO output.

Let the loop be in lock so that $\theta_e(t) << 1$ for all t. Then $\sin\theta_e(t) \cong \theta_e(t)$, $\psi(t) \cong \phi(t)$ and $\psi'(t) \cong \phi'(t)$.

As $\phi(t) = 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau$

we have$\phi'(t) = 2\pi k_f m(t)$

$\phi'(t) = 2\pi k_f\, m(t)$, $y(t) = \frac{\psi'(t)}{K_1} = \frac{2\pi k_f m(t)}{K_1}$

That is we get, $y(t) \propto m(t)$.

Pulse Modulation

SAMPLING THEOREM:

If a band limited signal g(t) contains no frequency components for $|f| < W$, then it is completely described by instantaneous values $g(kT_s)$ uniformly spaced in time with period $Ts \leq 1/2W$ (Fig. 1.36). If the sampling rate, f_s is at least equal to the Nyquist rate (i.e. $f_s \geq 2W$), then signal g(t) can be exactly reconstructed from the knowledge of its samples using a low pass reconstruction filter.

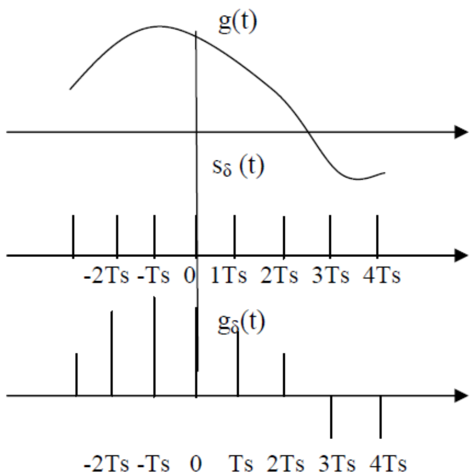


Figure 1.36: band limited signal g(t), pulse train $s_\delta(t)$ and the sampled signal $g_\delta(t)$

A signal is said to be band limited if it has frequency not more than a mentioned frequency. For example 4 khz signal is said to be band limited if it contains highest frequency component not more than 4 kHz.

Natural Sampling:

In this method of sampling, an electronic switch (Fig. 1.37) is used to periodically shift between the two contacts at a rate of $f_s = (1/T_s)$ Hz, staying on the input contact for small amount of time and on the grounded contact for the remaining of each sampling period (Fig. 1.38).The output $x_s(t)$ of the sampler consists of segments of x(t) and hence $x_s(t)$ can be considered as the product of x(t) and sampling function s(t). Hence circuit for natural sampling resembles a multiplier. $x_s(t) = x(t) . s(t)$

The sampling function s(t) is a periodic pulse train with period T_s.

$$S(t) = \begin{cases} 1 & -\tau/2 < t < \tau/2 \\ 0 & \tau/2 < |t| < Ts/2 \end{cases} \quad \text{-------} \ (1)$$

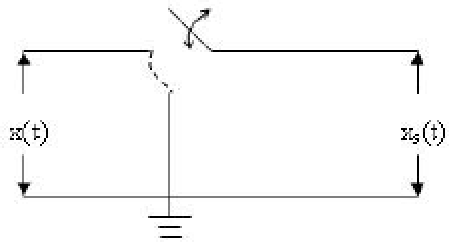


*Figure1.37***Error! Use the Home tab to apply 0 to the text that you want to appear here.***: Natural sampling using switch circuit*

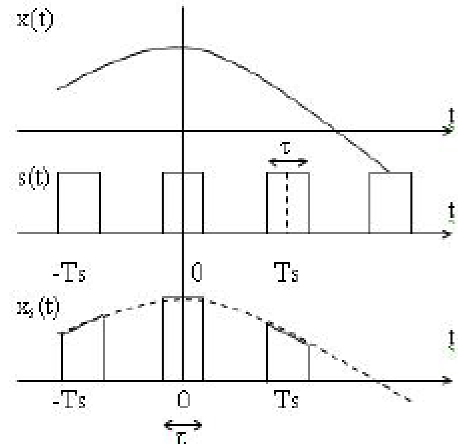


Figure 1.38*: Natural sampling waveform*

Flattop sampling:

In this method, the sampled waveform produced by practical sampling devices, the pulse p(t) is a flat – topped pulse of duration $\tau \leq T_s$,

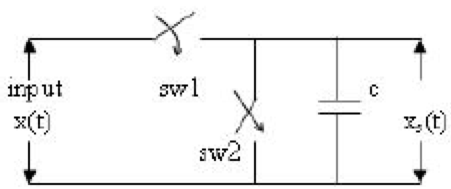


Figure 1.39: Flat top sampling circuit

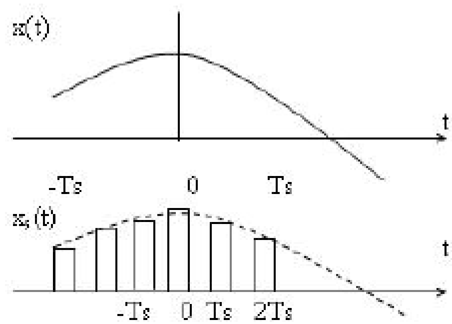


Figure 1.40: Flat top sampling waveform

Aliasing:

Aliasing effect occurs in sampled signal when the signal before sampling is not band limited. In this case message before sampling contain frequency component higher than the frequency determining the nyquist rate. Thus higher frequency components takes on the identity of lower frequency component called aliasing (Fig. 1.41).

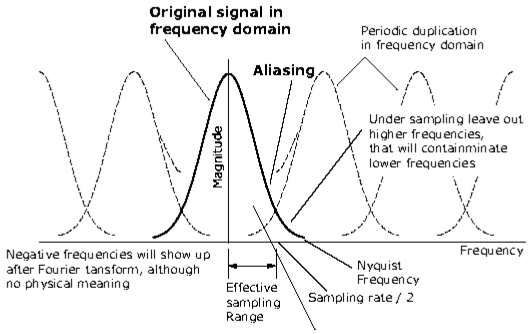


Figure 1.41: Aliasing effect

Pulse Modulation:

1. **Pulse Amplitude Modulation**

In Pulse Amplitude Modulation (PAM) technique, the amplitude of the pulse carrier varies, which is proportional to the instantaneous amplitude of the message signal.
The pulse amplitude modulated signal will follow the amplitude of the original signal, as the signal traces out the path of the whole wave. In natural PAM, a signal sampled at Nyquist rate can be reconstructed, by passing it through an efficient Low Pass Filter (LPF) with exact cutoff frequency. The following figures explain the Pulse Amplitude Modulation.

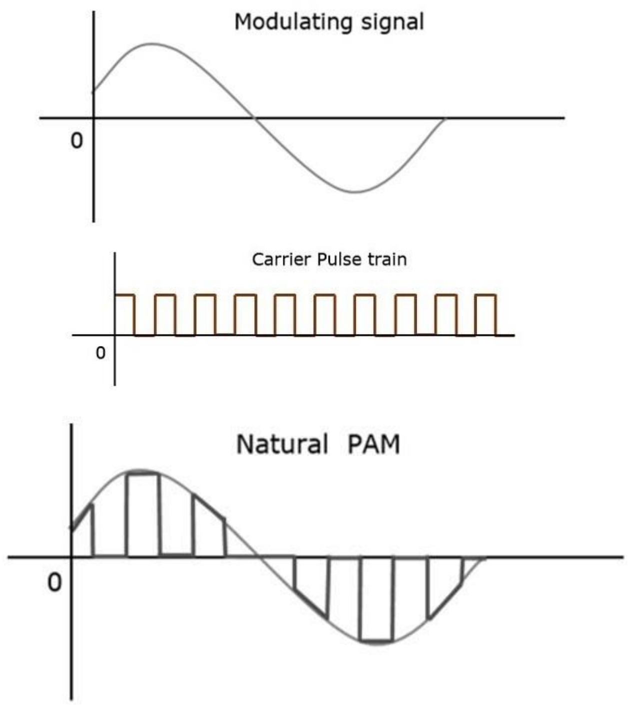


Figure 1.42: Natural sampled PAM waveform

Though the PAM signal is passed through a LPF, it cannot recover the signal without distortion. Hence, to avoid this noise, use flat-top sampling. The flat-top PAM signal is shown in the following figure.

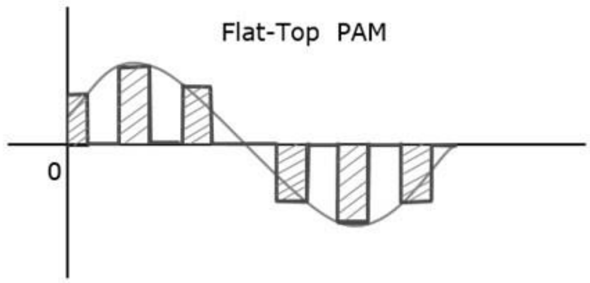


Figure 1.43: Flat-top sampled PAM waveform

Flat-top sampling is the process in which, the sampled signal can be represented in pulses for which the amplitude of the signal cannot be changed with respect to the analog signal, to be sampled. The tops of amplitude remain flat. This process simplifies the circuit design.

2. Pulse Width Modulation

In Pulse Width Modulation (PWM) or Pulse Duration Modulation (PDM) or Pulse Time Modulation (PTM) technique, the width or the duration or the time of the pulse carrier varies, which is proportional to the instantaneous amplitude of the message signal.

The width of the pulse varies in this method, but the amplitude of the signal remains constant. Amplitude limiters are used to make the amplitude of the signal constant. These circuits clip off the amplitude to a desired level, and hence the noise is limited.

The following figure explains the types of Pulse Width Modulations.

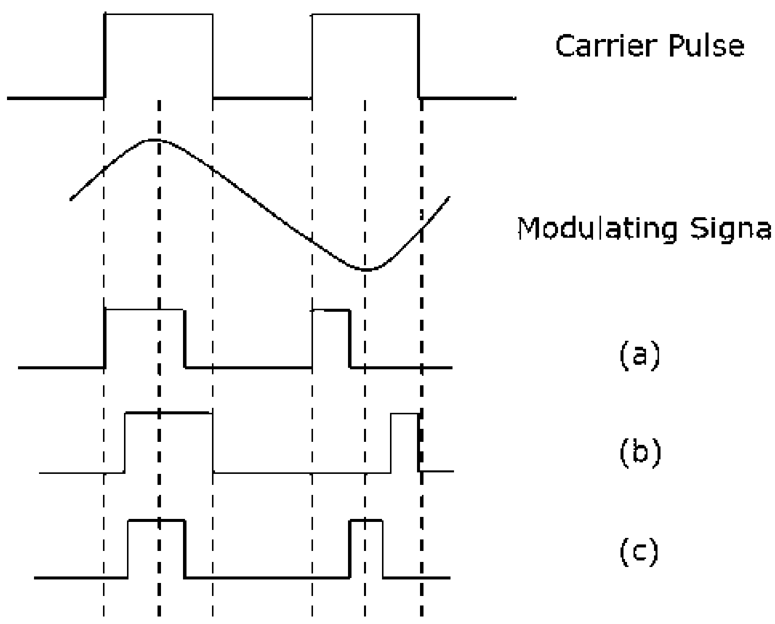


Figure 1.44: PWM waveform

There are three types of PWM.

The leading edge of the pulse being constant, the trailing edge varies according to the message signal. The waveform for this type of PWM is denoted as (a) in the above figure.

The trailing edge of the pulse being constant, the leading edge varies according to the message signal. The waveform for this type of PWM is denoted as (b) in the above figure.

The center of the pulse being constant, the leading edge and the trailing edge varies according to the message signal. The waveform for this type of PWM is denoted as (c) shown in the above figure.

3. Pulse Position Modulation

Pulse Position Modulation (PPM) is an analog modulation scheme in which, the amplitude and the width of the pulses are kept constant, while the position of each pulse, with reference to the position of a reference pulse varies according to the instantaneous sampled value of the message signal.

The transmitter has to send synchronizing pulses (or simply sync pulses) to keep the transmitter and the receiver in sync. These sync pulses help to maintain the position of the pulses. The following figures explain the Pulse Position Modulation.

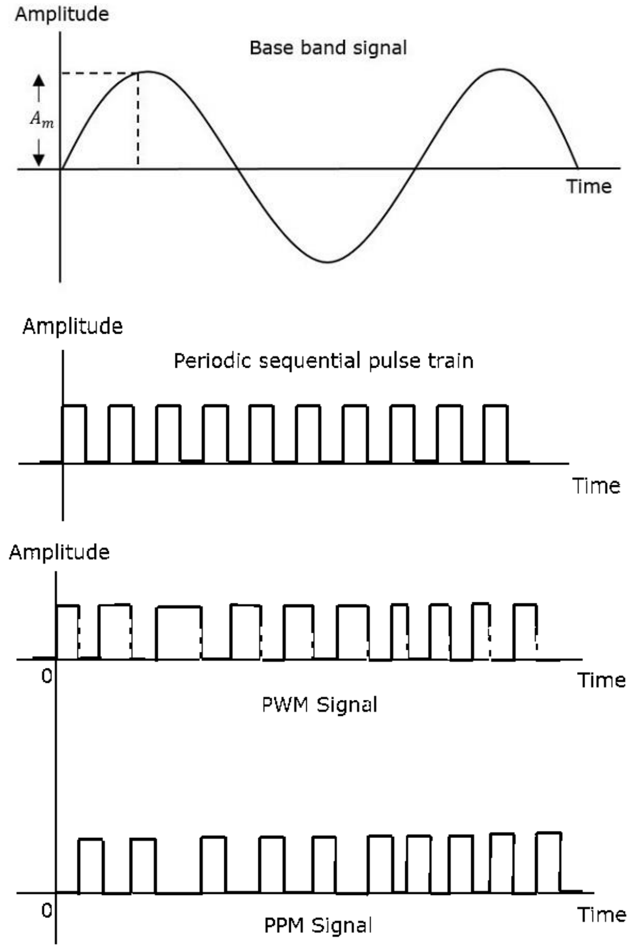


Figure 1.45: PPM waveform

Pulse position modulation is done in accordance with the pulse width modulated signal. Each trailing edge of the pulse width modulated signal becomes the starting point for pulses in PPM signal. Hence, the position of these pulses is proportional to the width of the PWM pulses.

Advantage:
As the amplitude and the width are constant, the power handled is also constant.
Disadvantage:
The synchronization between the transmitter and the receiver is a must.

Comparison between PAM, PWM, and PPM:

The following table presents the comparison between three modulation techniques.

PAM	PWM	PPM

Amplitude is varied	Width is varied	Position is varied
Bandwidth depends on the width of the pulse	Bandwidth depends on the rise time of the pulse	Bandwidth depends on the rise time of the pulse
Instantaneous transmitter power varies with the amplitude of the pulses	Instantaneous transmitter power varies with the amplitude and the width of the pulses	Instantaneous transmitter power remains constant with the width of the pulses
System complexity is high	System complexity is low	System complexity is low
Noise interference is high	Noise interference is low	Noise interference is low
It is similar to amplitude modulation	It is similar to frequency modulation	It is similar to phase modulation

Problem 1:

A modulating signal $m(t) = 10\cos(2\pi \times 10^3 t)$ is amplitude modulated with a carrier signal $c(t) = 50\cos(2\pi \times 10^5 t)$. Find the modulation index, the carrier power, and the power required for transmitting AM wave.

Solution

Given, the equation of modulating signal as

$$m(t) = 10\cos(2\pi \times 10^3 t)$$

We know the standard equation of modulating signal as

$$m(t) = A_m \cos(2\pi f_m t)$$

By comparing the above two equations, we will get Amplitude of modulating signal as A_m=10 volts

and Frequency of modulating signal as, $f_m = 10^3\ Hz$ = 1 KHz

Given, the equation of carrier signal is

$$c(t) = 50\cos(2\pi \times 10^5 t)$$

The standard equation of carrier signal is

$$c(t) = A_c \cos(2\pi f_c t)$$

By comparing these two equations, we will get

Amplitude of carrier signal as A_c=50volts
and Frequency of carrier signal as $f_c = 10^5\ Hz = 100\ KHz$
We know the formula for modulation index as

$$\mu = \frac{A_m}{A_c}$$

Substitute, A_m and A_c values in the above formula.

$$\mu = \frac{10}{50} = 0.2$$

Therefore, the value of **modulation index is 0.2** and percentage of modulation is 20%.

The formula for Carrier power, P_c is

$$P_c = \frac{A_c^2}{2R}$$

Assume $R = 1$ and substitute A_c value in the above formula.

$$P_c = \frac{50^2}{2} = 1250\,W$$

Therefore, the **Carrier power**, $P_c$is **1250 watts**.
We know the formula for **power** required for **transmitting AM** wave is

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

Substitute P_c and μ values in the above formula

$$P_t = 1250\left(1 + \frac{0.2^2}{2}\right) = 1275\,W$$

Therefore, the **power required for transmitting AM** wave is **1275 watts**.

Problem 2:

The equation of amplitude modulated wave is given by $s(t) = 20[1 + 0.8\cos(2\pi \times 10^3 t)]\cos(4\pi \times 10^5 t)$. Find the carrier power, the total sideband power, and the band width of AM wave.

Solution

Given, the equation of Amplitude modulated wave is

$$s(t) = 20[1 + 0.8\cos(2\pi \times 10^3 t)]\cos(4\pi \times 10^5 t)$$

Re-write the above equation as

$$s(t) = 20[1 + 0.8\cos(2\pi \times 10^3 t)]\cos(2\pi \times 2 \times 10^5 t)$$

We know the equation of Amplitude modulated wave is

$$s(t) = A_c[1 + \mu\cos(2\pi f_m t)]\cos(2\pi f_c t)$$

By comparing the above two equations, we will get

Amplitude of carrier signal as $A_c = 20$ volts
Modulation index as $\mu = 0.8$
Frequency of modulating signal as $f_m = 10^3\,Hz = 1\,KHz$
Frequency of carrier signal as $f_c = 2 \times 10^5 Hz = 200\,KHz$
The formula for Carrier power, P_c is

$$P_c = \frac{A_c^2}{2R}$$

Assume $R = 1\Omega$ and substitute $A_c$value in the above formula.

$$P_c = \frac{20^2}{2} = 200\,W$$

Therefore, the **Carrier power** P_c is **200watts**.
We know the formula for total side band power is

$$P_{SB} = \frac{P_c \mu^2}{2}$$

Substitute P_c and μ values in the above formula

$$P_{SB} = \frac{200 \times (0.8)^2}{2} = 64\ W$$

Therefore, the **total side band power** is **64 watts.**

We know the formula for bandwidth of AM wave is

$$BW = 2f_m$$

Substitute f_m value in the above formula.

$$BW = 2(1\ K) = 2\ KHz$$

Therefore, the **bandwidth** of AM wave is **2 KHz.**

Problem 3:

A single-tone FM is represented by the voltage equation as:

$$v(t) = 12\cos(6 \times 10^8 t + 5\sin 1250t)$$

Determine the following:

(a) Carrier frequency, (ii) modulating frequency, (iii) the modulation index, (iv) maximum deviation, (v) what power will this FM wave dissipate in 10Ω resistors.

Solution:

We know that the standard expression for a single-tone FM wave is given as

$$v(t) = A\cos\left(\omega_c t + m_f \sin\omega_m t\right)$$

The given expression is

$$v(t) = 12\cos(6 \times 10^8 t + 5\sin 1250t)$$

By comparing the above two equations, we get

(i) Carrier frequency

$$\omega_c = 6 \times 10^8 \text{ rad/sec}$$

$$f_c = \frac{6\times10^8}{2\pi} = 95.5 \text{ MHz}$$

(ii) Modulating frequency

$$\omega_m = 1250 \text{ rad/sec}$$

$$f_m = \frac{1250}{2\pi} = 199 \text{ Hz}$$

(iii) $m_f = 5$

(iv) Maximum frequency deviation is given as

$$m_f = \frac{\Delta f}{f_m}$$

$$\Delta f = m_f f_m = 5 \times 199 = 995 \text{ Hz}$$

(v) The power dissipated is

$$P = \frac{v_{rms}^2}{R} = \frac{\left(12/\sqrt{2}\right)^2}{10} = 7.2 \text{ watts}$$

Problem 4:

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, $\Delta f = 75$ kHz and $f_m = 8$ 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
$BW = 2(\Delta f + f_m) = 2(150 + 8) = 316$ kHz

MCQ Questions:

1. If the radiated power of AM transmitter is 10 kW, the power in the carrier for modulation index of 0.6 is nearly

a) 8.24 kW b) 8.47 kW c) 9.26 kW d) 9.6 kW

2. For video transmission of television, which of the following is used?

a) AM b) DSB-SC c) VSB d) SSB-SC

3. A 10 kW transmitter is modulated to 80%. The average sideband power will be

a) 1.8 kW b) 8 kW
c) 3.2 kW d) 4.6 kW

4. The length of the antenna to transmit a signal must be at least

a) 1/3 wavelength b) 2/3 wavelength
c) 1/4 wavelength d) 3/4 wavelength

5. Two sinusoidal signals are simultaneously modulating a carrier, the modulation indices being 0.3 and 0.4. The overall modulation index is

a) 0.5 b) 0.1 c) 0.7 d) 0.12

6. AM demodulation techniques are

a. Square law demodulator b. Envelope detector
c. PLL detector d. Both a and b are correct

7. In synchronous detection of AM signal
a. Carrier is locally generated
b. Passed through a low pass filter
c. The original signal is recovered
d. All of the above

8. Advantages of using an RF amplifier are:
a. Better selectivity
b. Better sensitivity
c. Improved signal to noise ratio
d. All of the above

9. Intermediate frequency (IF) should be carefully chosen as
a. High IF results in poor selectivity
b. High IF results in problems in tracking of signals

c. Image frequency rejection becomes poor at low IF
d. All of the above

10. The standard value for Intermediate frequency (IF) in AM receivers is
a. 455 KHz
b. 580 KHz
c. 10.7 MHz
d. 50 MHz

11. Function of frequency mixer in super heterodyne receiver is
a. Amplification
b. Filtering
c. Multiplication of incoming signal and the locally generated carrier
d. None of the above

12. One of the following cannot be used to remove the unwanted sideband in SSB. This is the
a. filter system
b. phase-shift method
c. third method
d. balanced modulator

13. In Amplitude Demodulation, the condition which the load resistor R must satisfy to discharge capacitor C slowly between the positive peaks of the carrier wave so that the capacitor voltage will not discharge at the maximum rate of change of the modulating wave (W is message bandwidth and ω is carrier frequency, in rad/sec) is
a. RC < 1/W
b. RC > 1/W
c. RC < 1/ω
d. RC > 1/ω

14. If a receiver has poor capacity of blocking adjacent channel interference then the receiver has
a. Poor selectivity
b. Poor Signal to noise ratio
c. Poor sensitivity
d. None of the above

15. Demodulation is:
a. Detection
b. Recovering information from modulated signal
c. Both a and b
d. None of the above

Sample Questions:

1. What are the basic components of a communication system? Why modulation is necessary for communication?
2. Calculate the percent power saving for the DSB-SC and SSB-SC signal if the AM wave is modulated to a depth of 100%
3. A modulated signal is given by, $s(t) = m_1(t)\cos(2\pi f_c t) + m_2(t)\sin(2\pi f_c t)$

Where the baseband signal $m_1(t)$ and $m_2(t)$ have bandwidths of 10 kHz and 15 kHz respectively. What is the bandwidth of the modulated signal?
4. Define Modulation index for AM and calculate it in terms V_{max} and V_{min} of modulated wave. The carrier $V_c(t) = 15x10^3 Cos\ (600\pi\ x\ 10^3 t)$ is modulated by the baseband signal $m_1(t) = 30\ Cos\ (6\pi\ x\ 10^3 t)$. Calculate the transmission efficiency if AM is used.
5. Prove that the efficiency of a single tone AM is 33.3% for perfect modulation. What will be the efficiency if the value of modulation index is 0.5?
6. A carrier signal $A_c Cos\omega_c t$ is amplitude modulated by a message signal $A_m Cos\omega_m t$, where $A_m < A_c$. (i) Write down the expression for the modulated signal. (ii) Write down the

expression for carrier component and side band component. (iii) Draw the phasor diagram of the modulated signal.
7. What do you mean by modulation? How to generate AM wave using square-law modulator?
8. Briefly explain the generation of DSB-SC using balanced modulator.
9. What is heterodyning? Draw the block diagram of superheterodyne receiver & explain the operation.
10. Determine the image frequency for a standard broadcast band AM receiver using a 455 KHz IF & tuned to a station at 640 KHz.
11. What are the generating methods for SSB-SC signals?
12. Draw the block diagram of the filter method for generating SSB signal for an audio baseband signal.
13. Explain the phase shift method for SSB generation.
14. Explain with suitable block diagram the generation of FM signal using Armstrong method.
15. Discuss about the roles of pre-emphasis and de-emphasis circuit in FM broadcasting.
16. A frequency modulated signal is represented as follows:

$$e_{FM} = 10\sin(16\pi \times 10^6 t + 20\sin 2\pi \times 10^3 t) \text{ volts.}$$

Determine:
(i) carrier frequency, (ii) modulating frequency, (iii) modulation index, (iv) frequency deviation, (v) power dissipated in 10 Ω resistor.
17. Explain with block diagram the indirect method of WBFM generation.
18. Explain the direct method of generation of FM signals using a varactor diode.
19. Explain FM demodulation scheme using PLL.

Module-2
Digital Transmission

Digital communication is the process of devices communicating information digitally. This tutorial helps the readers to get a good idea on how the signals are digitized and why digitization is needed. By the completion of this tutorial, the reader will be able to understand the conceptual details involved in digital communication.

Advantages of Digital Communication

As the signals are digitized, there are many advantages of digital communication over analog communication, such as −

- The effect of distortion, noise, and interference is much less in digital signals as they are less affected.
- Digital circuits are more reliable.
- Digital circuits are easy to design and cheaper than analog circuits.
- The hardware implementation in digital circuits, is more flexible than analog.
- The occurrence of cross-talk is very rare in digital communication.
- The signal is un-altered as the pulse needs a high disturbance to alter its properties, which is very difficult.
- Signal processing functions such as encryption and compression are employed in digital circuits to maintain the secrecy of the information.
- The probability of error occurrence is reduced by employing error detecting and error correcting codes.
- Spread spectrum technique is used to avoid signal jamming.

- Combining digital signals using Time Division Multiplexing (TDM) is easier than combining analog signals using Frequency Division Multiplexing (FDM).
- The configuring process of digital signals is easier than analog signals.
- Digital signals can be saved and retrieved more conveniently than analog signals.
- Many of the digital circuits have almost common encoding techniques and hence similar devices can be used for a number of purposes.
- The capacity of the channel is effectively utilized by digital signals.

Elements of Digital Communication

The elements which form a digital communication system is represented by the following block diagram for the ease of understanding.

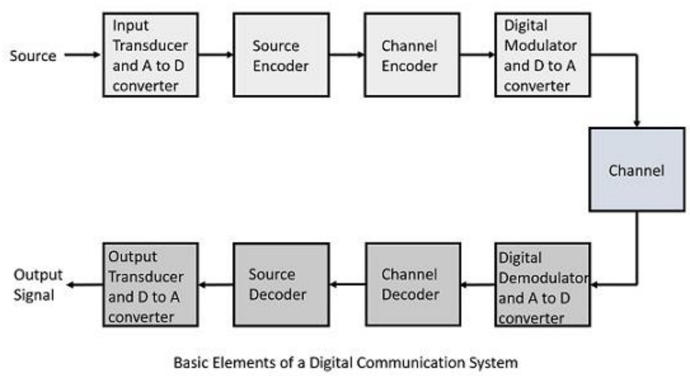


Figure 2.1: Generic block diagram of a digital communication system

Following are the sections of the digital communication system.

Source

The source can be an analog signal. Example: A Sound signal

Input Transducer

This is a transducer which takes a physical input and converts it to an electrical signal (Example: microphone). This block also consists of an analog to digital converter where a digital signal is needed for further processes.

A digital signal is generally represented by a binary sequence.

Source Encoder

The source encoder compresses the data into minimum number of bits. This process helps in effective utilization of the bandwidth. It removes the redundant bits (unnecessary excess bits, i.e., zeroes).

Channel Encoder

The channel encoder, does the coding for error correction. During the transmission of the signal, due to the noise in the channel, the signal may get altered and hence to avoid this, the channel encoder adds some redundant bits to the transmitted data. These are the error correcting bits.

Digital Modulator

The signal to be transmitted is modulated here by a carrier. The signal is also converted to analog from the digital sequence, in order to make it travel through the channel or medium.

Channel

The channel or a medium, allows the analog signal to transmit from the transmitter end to the receiver end.

Digital Demodulator
This is the first step at the receiver end. The received signal is demodulated as well as converted again from analog to digital. The signal gets reconstructed here.
Channel Decoder
The channel decoder, after detecting the sequence, does some error corrections. The distortions which might occur during the transmission are corrected by adding some redundant bits. This addition of bits helps in the complete recovery of the original signal.
Source Decoder
The resultant signal is once again digitized by sampling and quantizing so that the pure digital output is obtained without the loss of information. The source decoder recreates the source output.
Output Transducer
This is the last block which converts the signal into the original physical form, which was at the input of the transmitter. It converts the electrical signal into physical output (Example: loud speaker).
Output Signal
This is the output which is produced after the whole process. Example − the sound signal received.

Modulation is the process of varying one or more parameters of a carrier signal in accordance with the instantaneous values of the message signal.
The message signal is the signal which is being transmitted for communication and the carrier signal is a high frequency signal which has no data, but is used for long distance transmission.
There are many modulation techniques, which are classified according to the type of modulation employed. Of them all, the digital modulation technique used is **Pulse Code Modulation (PCM)**.
A signal is pulse code modulated to convert its analog information into a binary sequence, i.e., **1s** and **0s**. The output of a PCM will resemble a binary sequence. The following figure shows an example of PCM output with respect to instantaneous values of a given sine wave.

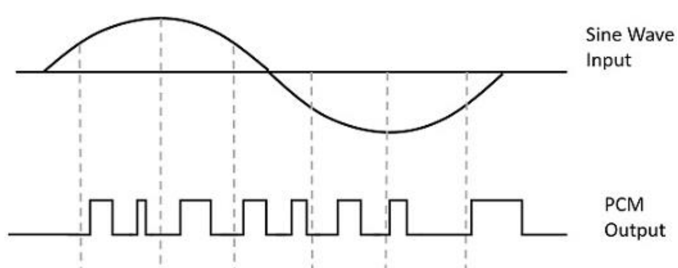


Figure 2.2: Example of a PCM waveform

Instead of a pulse train, PCM produces a series of numbers or digits, and hence this process is called as **digital**. Each one of these digits, though in binary code, represents the approximate amplitude of the signal sample at that instant.
In Pulse Code Modulation, the message signal is represented by a sequence of coded pulses. This message signal is achieved by representing the signal in discrete form in both time and amplitude.

Basic Elements of PCM

The transmitter section of a Pulse Code Modulator circuit consists of **Sampling, Quantizing** and **Encoding**, which are performed in the analog-to-digital converter section. The low pass filter prior to sampling prevents aliasing of the message signal.

The basic operations in the receiver section are **regeneration of impaired signals, decoding,** and **reconstruction** of the quantized pulse train. Following is the block diagram of PCM which represents the basic elements of both the transmitter and the receiver sections.

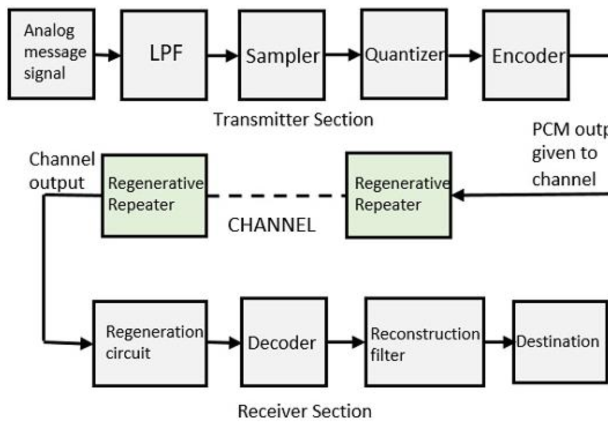


Figure 2.3: Block diagram of a PCM system

Low Pass Filter

This filter eliminates the high frequency components present in the input analog signal which is greater than the highest frequency of the message signal, to avoid aliasing of the message signal.

Sampler

This is the technique which helps to collect the sample data at instantaneous values of message signal, so as to reconstruct the original signal. The sampling rate must be greater than twice the highest frequency component **W** of the message signal, in accordance with the sampling theorem.

Quantizer

Quantizing is a process of reducing the excessive bits and confining the data. The sampled output when given to Quantizer, reduces the redundant bits and compresses the value.

Encoder

The digitization of analog signal is done by the encoder. It designates each quantized level by a binary code. The sampling done here is the sample-and-hold process. These three sections (LPF, Sampler, and Quantizer) will act as an analog to digital converter. Encoding minimizes the bandwidth used.

Regenerative Repeater

This section increases the signal strength. The output of the channel also has one regenerative repeater circuit, to compensate the signal loss and reconstruct the signal, and also to increase its strength.

Decoder

The decoder circuit decodes the pulse coded waveform to reproduce the original signal. This circuit acts as the demodulator.

Reconstruction Filter

After the digital-to-analog conversion is done by the regenerative circuit and the decoder, a low-pass filter is employed, called as the reconstruction filter to get back the original signal.
Hence, the Pulse Code Modulator circuit digitizes the given analog signal, codes it and samples it, and then transmits it in an analog form. This whole process is repeated in a reverse pattern to obtain the original signal.
Sampling is defined as, "The process of measuring the instantaneous values of continuous-time signal in a discrete form."
Sample is a piece of data taken from the whole data which is continuous in the time domain.
When a source generates an analog signal and if that has to be digitized, having **1s** and **0s** i.e., High or Low, the signal has to be discretized in time. This discretization of analog signal is called as Sampling.
The following figure indicates a continuous-time signal **x (t)** and a sampled signal **x_s (t)**. When **x (t)** is multiplied by a periodic impulse train, the sampled signal **x_s (t)** is obtained.

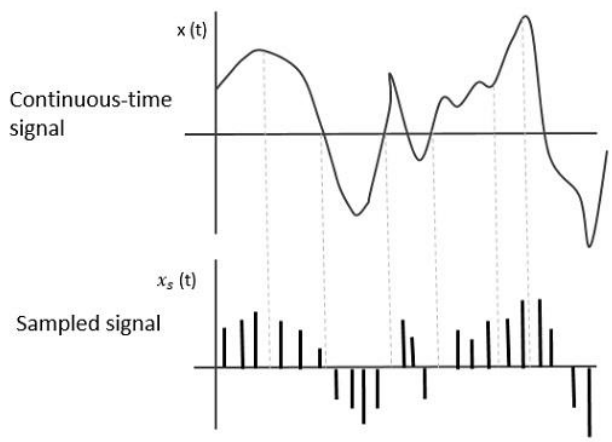


Figure 2.4: Sampled waveform

Sampling Rate
To discretize the signals, the gap between the samples should be fixed. That gap can be termed as a **sampling period T_s**.
SamplingFrequency$= \frac{1}{T_s} = f_s$
Where,
- T_s is the sampling time
- f_s is the sampling frequency or the sampling rate

Sampling frequency is the reciprocal of the sampling period. This sampling frequency, can be simply called as **Sampling rate**. The sampling rate denotes the number of samples taken per second, or for a finite set of values.
For an analog signal to be reconstructed from the digitized signal, the sampling rate should be highly considered. The rate of sampling should be such that the data in the message signal should neither be lost nor it should get over-lapped. Hence, a rate was fixed for this, called as Nyquist rate.
Nyquist Rate
Suppose that a signal is band-limited with no frequency components higher than **W** Hertz. That means, **W** is the highest frequency. For such a signal, for effective reproduction of the original signal, the sampling rate should be twice the highest frequency.

Which means,

$$f_s = 2W$$

Where,

- f_s is the sampling rate
- **W** is the highest frequency

This rate of sampling is called as **Nyquist rate**.

A theorem called, Sampling Theorem, was stated on the theory of this Nyquist rate.

Sampling Theorem

The sampling theorem, which is also called as **Nyquist theorem**, delivers the theory of sufficient sample rate in terms of bandwidth for the class of functions that are bandlimited.

The sampling theorem states that, "a signal can be exactly reproduced if it is sampled at the rate $\mathbf{f_s}$ which is greater than twice the maximum frequency **W**."

To understand this sampling theorem, let us consider a band-limited signal, i.e., a signal whose value is **non-zero** between some **–W** and **W** Hertz.

Such a signal is represented as x(f)=0 for |f|>W

For the continuous-time signal **x (t)**, the band-limited signal in frequency domain, can be represented as shown in the following figure.

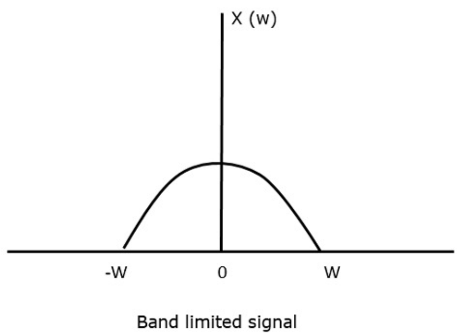


Figure 2.5: Message waveform

We need a sampling frequency, a frequency at which there should be no loss of information, even after sampling. For this, we have the Nyquist rate that the sampling frequency should be two times the maximum frequency. It is the critical rate of sampling.

If the signal **x(t)** is sampled above the Nyquist rate, the original signal can be recovered, and if it is sampled below the Nyquist rate, the signal cannot be recovered.

The following figure explains a signal, if sampled at a higher rate than **2w** in the frequency domain.

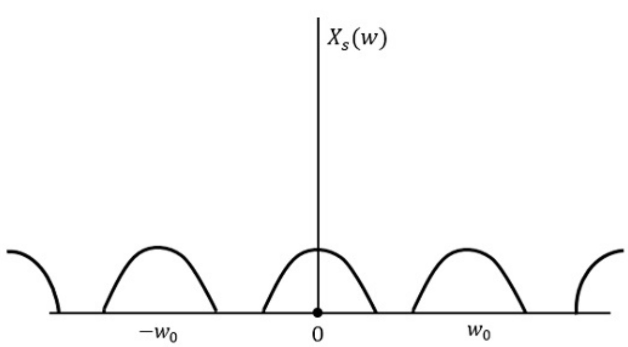


Figure 2.6: Fourier transform of message waveform

The above figure shows the Fourier transform of a signal $x_s(t)$. Here, the information is reproduced without any loss. There is no mixing up and hence recovery is possible.
The Fourier Transform of the signal $x_s(t)$ is

$$X_s(w) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(w - nw_0)$$

Where T_s = **Sampling Period** and $w_0 = \frac{2\pi}{T_s}$

Let us see what happens if the sampling rate is equal to twice the highest frequency (**2W**)

That means, $f_s = 2W$

Where,

- f_s is the sampling frequency
- **W** is the highest frequency

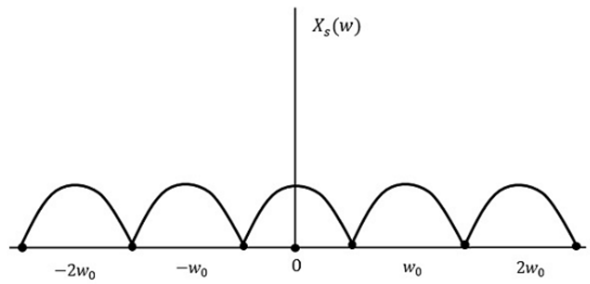


Figure 2.7: Fourier transform of message waveform when sampling rate=2W

The result will be as shown in the above figure. The information is replaced without any loss. Hence, this is also a good sampling rate.

Now, let us look at the condition,

fs<2W

The resultant pattern will look like the following figure.

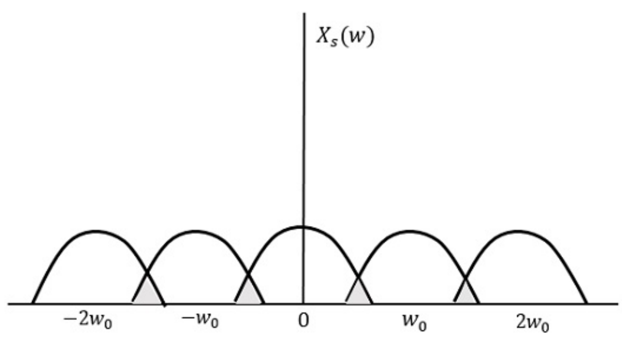


Figure 2.8: Aliasing effect

We can observe from the above pattern that the over-lapping of information is done, which leads to mixing up and loss of information. This unwanted phenomenon of over-lapping is called as **Aliasing**.

The digitization of analog signals involves the rounding off of the values which are approximately equal to the analog values. The method of sampling chooses a few points on the analog signal and then these points are joined to round off the value to a near stabilized value. Such a process is called as **Quantization**.

Quantizing an Analog Signal

The analog-to-digital converters perform this type of function to create a series of digital values out of the given analog signal. The following figure represents an analog signal. This signal to get converted into digital, has to undergo sampling and quantizing.

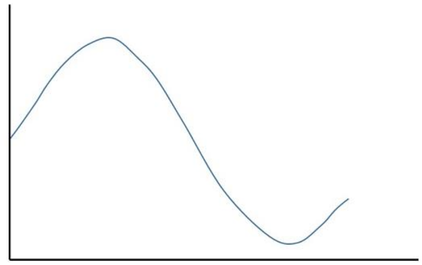


Figure 2.9(a): Analog message signal

The quantizing of an analog signal is done by discretizing the signal with a number of quantization levels. **Quantization** is representing the sampled values of the amplitude by a finite set of levels, which means converting a continuous-amplitude sample into a discrete-time signal.
The following figure shows how an analog signal gets quantized. The blue line represents analog signal while the brown one represents the quantized signal.

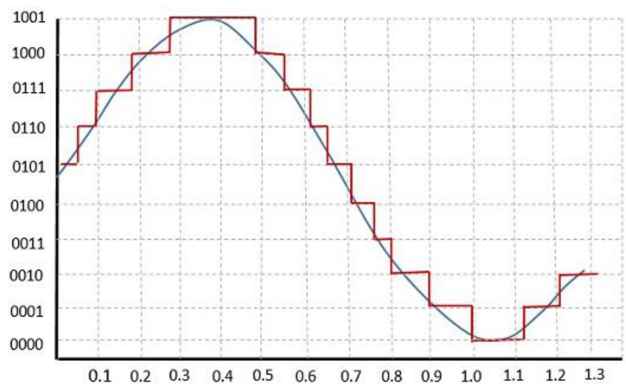


Figure 2.9(b): Quantization process

Both sampling and quantization result in the loss of information. The quality of a Quantizer output depends upon the number of quantization levels used. The discrete amplitudes of the quantized output are called as **representation levels** or **reconstruction levels**. The spacing between the two adjacent representation levels is called a **quantum** or **step-size**.
The following figure shows the resultant quantized signal which is the digital form for the given analog signal.

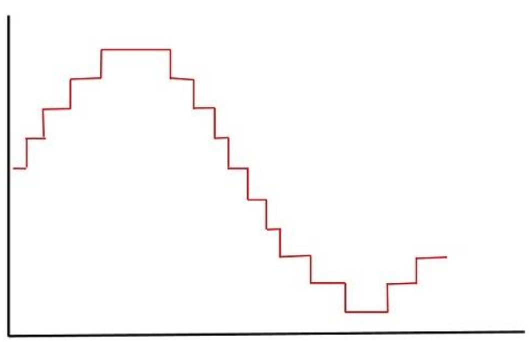


Figure 2.9(c): Quantized signal

This is also called as **Stair-case** waveform, in accordance with its shape.

Types of Quantization

There are two types of Quantization - Uniform Quantization and Non-uniform Quantization.

The type of quantization in which the quantization levels are uniformly spaced is termed as a **Uniform Quantization**. The type of quantization in which the quantization levels are unequal and mostly the relation between them is logarithmic, is termed as a **Non-uniform Quantization**.

There are two types of uniform quantization. They are Mid-Rise type and Mid-Tread type. The following figures represent the two types of uniform quantization.

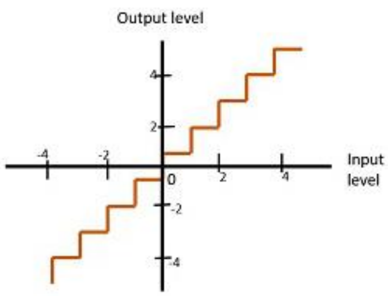


Fig 1 : Mid-Rise type Uniform Quantization

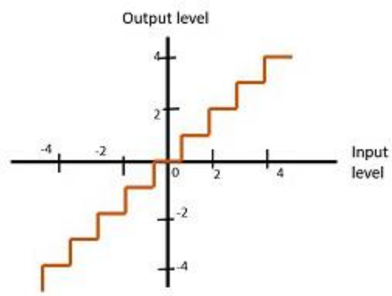


Fig 2 : Mid-Tread type Uniform Quantization

Figure 1 shows the mid-rise type and figure 2 shows the mid-tread type of uniform quantization.

- The **Mid-Rise** type is so called because the origin lies in the middle of a raising part of the stair-case like graph. The quantization levels in this type are even in number.
- The **Mid-tread** type is so called because the origin lies in the middle of a tread of the stair-case like graph. The quantization levels in this type are odd in number.
- Both the mid-rise and mid-tread type of uniform quantizers are symmetric about the origin.

Quantization Error

For any system, during its functioning, there is always a difference in the values of its input and output. The processing of the system results in an error, which is the difference of those values.

The difference between an input value and its quantized value is called a **Quantization Error**. A **Quantizer** is a logarithmic function that performs Quantization (rounding off the value). An analog-to-digital converter (**ADC**) works as a quantizer.

The following figure illustrates an example for a quantization error, indicating the difference between the original signal and the quantized signal.

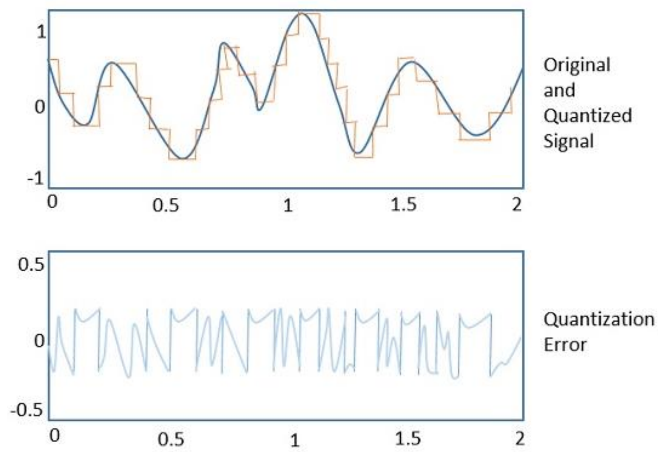


Figure 2.10: Example of quantization error

Quantization Noise

It is a type of quantization error, which usually occurs in analog audio signal, while quantizing it to digital. For example, in music, the signals keep changing continuously, where regularity is not found in errors. Such errors create a wideband noise called as **Quantization Noise**.

Companding in PCM

The word **Companding** is a combination of Compressing and Expanding, which means that it does both. This is a non-linear technique used in PCM which compresses the data at the transmitter and expands the same data at the receiver. The effects of noise and crosstalk are reduced by using this technique.

There are two types of Companding techniques. They are −

A-law Companding Technique

- Uniform quantization is achieved at **A = 1**, where the characteristic curve is linear and no compression is done.
- A-law has mid-rise at the origin. Hence, it contains a non-zero value.
- A-law companding is used for PCM telephone systems.

µ-law Companding Technique

- Uniform quantization is achieved at **µ = 0**, where the characteristic curve is linear and no compression is done.
- µ-law has mid-tread at the origin. Hence, it contains a zero value.
- µ-law companding is used for speech and music signals.

µ-law is used in North America and Japan.

For the samples that are highly correlated, when encoded by PCM technique, leave redundant information behind. To process this redundant information and to have a better output, it is a wise decision to take a predicted sampled value, assumed from its previous output and summarize them with the quantized values. Such a process is called as **Differential PCM (DPCM)** technique.

DPCM Transmitter

The DPCM Transmitter consists of Quantizer and Predictor with two summer circuits. Following is the block diagram of DPCM transmitter.

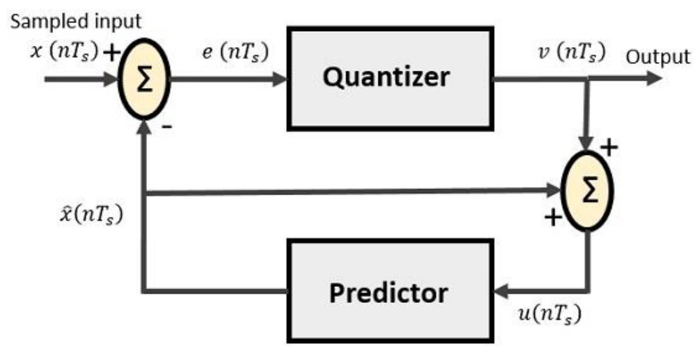


Figure 2.11(a): Block diagram of a DPCM transmitter

DPCM Receiver

The block diagram of DPCM Receiver consists of a decoder, a predictor, and a summer circuit. Following is the diagram of DPCM Receiver.

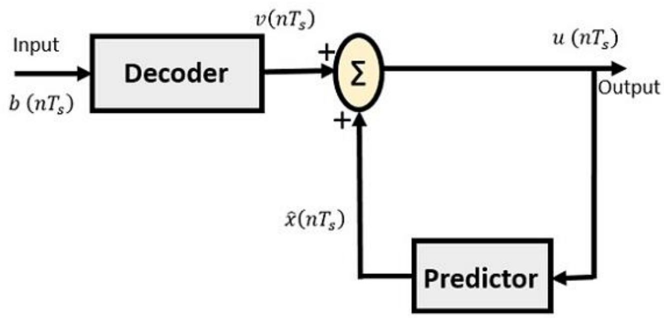


Figure 2.11(b): Block diagram of a DPCM receiver

Delta Modulation

The type of modulation, where the sampling rate is much higher and in which the step size after quantization is of a smaller value Δ, such a modulation is termed as delta modulation.

Delta Modulator

The Delta Modulator comprises of a 1-bit quantizer and a delay circuit along with two summer circuits. Following is the block diagram of a delta modulator.

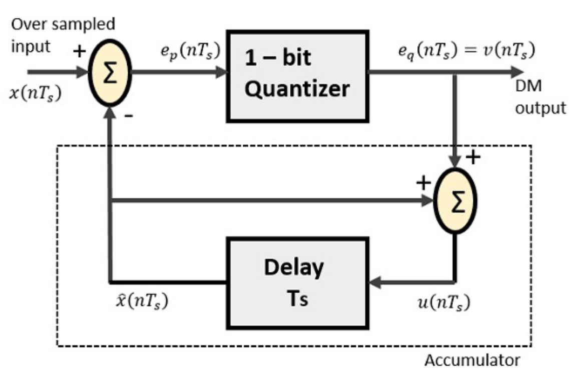


Figure 2.12(a): Block diagram of a DM transmitter

From the above diagram, we have the notations as −

- x(nTs) = over sampled input
- ep(nTs) = summer output and quantizer input
- eq(nTs) = quantizer output = v(nTs)

- $x\hat{}(nTs)$ = output of delay circuit
- $u(nTs)$ = input of delay circuit

Using these notations, now we shall try to figure out the process of delta modulation.

$ep(nTs)=x(nTs)-x\hat{}(nTs)$

$=x(nTs)-u([n-1]Ts)$

$=x(nTs)-[x\hat{}[[n-1]Ts]+v[[n-1]Ts]]$

Further,

$v(nTs)=eq(nTs)=S.sig.[ep(nTs)]$

$u(nTs)=x\hat{}(nTs)+eq(nTs)$

Where,

- $x\hat{}(nTs)$ = the previous value of the delay circuit
- $eq(nTs)$ = quantizer output = $v(nTs)$

Hence,

$u(nTs)=u([n-1]Ts)+v(nTs)$

Which means,

The present input of the delay unit

= (The previous output of the delay unit) + (the present quantizer output)

A Stair-case approximated waveform will be the output of the delta modulator with the step-size as delta (Δ). The output quality of the waveform is moderate.

Delta Demodulator

The delta demodulator comprises of a low pass filter, a summer, and a delay circuit. The predictor circuit is eliminated here and hence no assumed input is given to the demodulator.

Following is the diagram for delta demodulator.

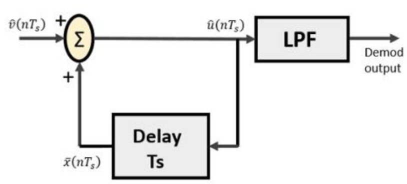


Figure 2.12(b): Block diagram of a DM receiver

However, there exists some noise in DM.

- Slope Over load distortion (when **Δ** is small)
- Granular noise (when **Δ** is large)

Adaptive Delta Modulation (ADM)

In digital modulation, we have come across certain problem of determining the step-size, which influences the quality of the output wave.

A larger step-size is needed in the steep slope of modulating signal and a smaller step-size is needed where the message has a small slope. The minute details get missed in the process. So, it would be better if we can control the adjustment of step-size, according to our requirement in order to obtain the sampling in a desired fashion. This is the concept of **Adaptive Delta Modulation**.

Following is the block diagram of Adaptive delta modulator.

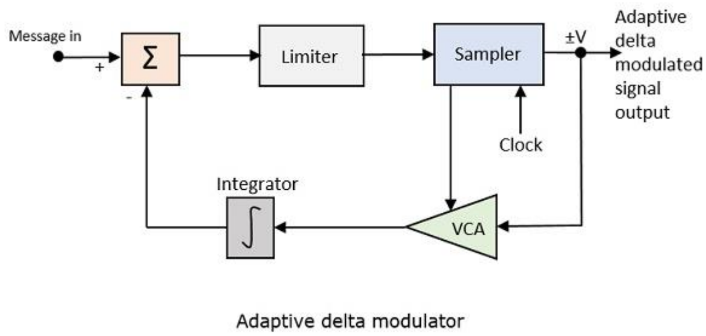


Figure 2.13: Block diagram of ADM transmitter

The gain of the voltage controlled amplifier is adjusted by the output signal from the sampler. The amplifier gain determines the step-size and both are proportional.

ADM quantizes the difference between the value of the current sample and the predicted value of the next sample. It uses a variable step height to predict the next values, for the faithful reproduction of the fast varying values.

Multiplexing

Multiplexing is the process of combining multiple signals into one signal, over a shared medium. These signals, if analog in nature, the process is called as analog multiplexing. If digital signals are multiplexed, it is called as digital multiplexing.

Multiplexing was first developed in telephony. A number of signals were combined to send through a single cable. The process of multiplexing divides a communication channel into several number of logical channels, allotting each one for a different message signal or a data stream to be transferred. The device that does multiplexing, can be called as a MUX. The reverse process, i.e., extracting the number of channels from one, which is done at the receiver is called as de-multiplexing. The device which does de-multiplexing is called as DEMUX.

The following figures represent MUX and DEMUX. Their primary use is in the field of communications.

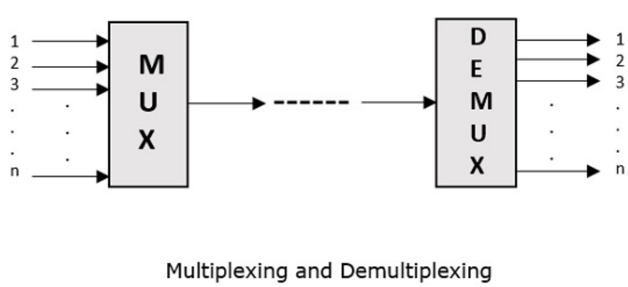
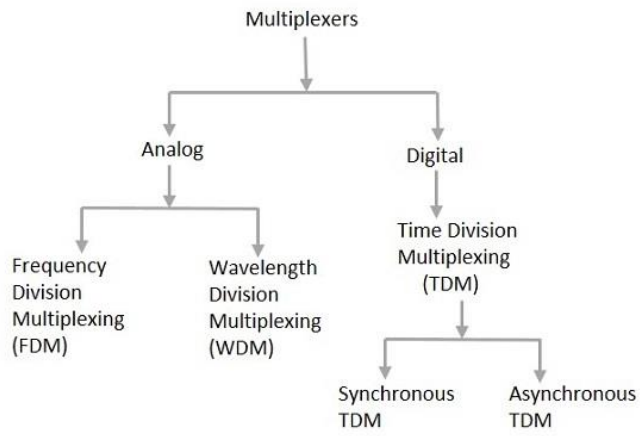


Figure 2.14: Multiplexing and demultiplexing



Analog Multiplexing

The analog multiplexing techniques involve signals which are analog in nature. The analog signals are multiplexed according to their frequency (FDM) or wavelength (WDM).

Frequency Division Multiplexing (FDM)

In analog multiplexing, the most used technique is Frequency Division Multiplexing (FDM). This technique uses various frequencies to combine streams of data, for sending them on a communication medium, as a single signal.

Example − A traditional television transmitter, which sends a number of channels through a single cable, uses FDM.

Wavelength Division Multiplexing (WDM)

Wavelength Division multiplexing is an analog technique, in which many data streams of different wavelengths are transmitted in the light spectrum. If the wavelength increases, the frequency of the signal decreases. A prism which can turn different wavelengths into a single line, can be used at the output of MUX and input of DEMUX.

Example − Optical fiber communications use WDM technique to merge different wavelengths into a single light for communication.

Digital Multiplexing

The term digital represents the discrete bits of information. Hence, the available data is in the form of frames or packets, which are discrete.

Time Division Multiplexing (TDM)

In TDM, the time frame is divided into slots. This technique is used to transmit a signal over a single communication channel, by allotting one slot for each message.

Of all the types of TDM, the main ones are Synchronous and Asynchronous TDM.

Synchronous TDM

In Synchronous TDM, the input is connected to a frame. If there are ‘n’ number of connections, then the frame is divided into ‘n’ time slots. One slot is allocated for each input line.

In this technique, the sampling rate is common to all signals and hence the same clock input is given. The MUX allocates the same slot to each device at all times.

Asynchronous TDM

In Asynchronous TDM, the sampling rate is different for each of the signals and a common clock is not required. If the allotted device, for a time-slot, transmits nothing and sits idle, then that slot is allotted to another device, unlike synchronous. This type of TDM is used in Asynchronous transfer mode networks.

Match Filter

Receiver Model:

In our simple model, the signal $s_i(t)$, denoting one the two possible received signals $s_0(t)$ and $s_1(t)$ is processed through a filter and then sampled at time T_0. The received signal is corrupted by noise which also passes through the filter and corrupts the sample value which is thus $\mathbb{Y} = \hat{s}_i(T_0) + N(T_0)$.

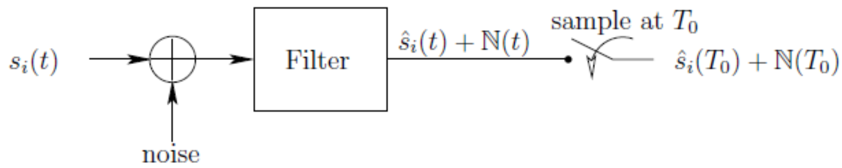


Figure 2.15: Block diagram of a baseband receiver

The pdf of $N(T_0)$ does not depend on whether $s_0(t)$ or $s_1(t)$ is received, that is, the noise is assumed to be independent of the received signal. A good model for $N(T_0)$ is a zero-mean Gaussian random variable with variance σ^2. The *min-max* decision rule is the same as the *maximum-likelihood* decision rule is the same as the *minimum-error-probability* decision rule (since we are assuming equally likely signals) and consists of deciding that a 0 or a 1 was transmitted according as $|\mathbb{Y} - \hat{s}_0(T_0)|$ is smaller than or larger than $|\mathbb{Y} - \hat{s}_1(T_0)|$. Equivalently, assuming that $\hat{s}_0(T_0) > \hat{s}_1(T_0)$, the receiver compares the sample value $\mathbb{Y}$ to a threshold $\theta = \frac{(\hat{s}_0(T_0)+\hat{s}_1(T_0))}{2}$ and decides that a 0 or a 1 was transmitted according as $\mathbb{Y}$ is larger than or smaller than θ. The error probability P_e achieved by this decision rule is $Q\left(\frac{\hat{s}_0(T_0)+\hat{s}_1(T_0)}{2\sigma}\right) = Q(SNR)$ where SNR (signal-to-noise ratio) is the value of the argument of the $Q(.)$ function. Note that P_e decreases rapidly as SNR increases. It is also worth noting that there are many different definitions for SNR that are in common use, and thus, care must be taken in comparing systems from different designers since they be defining SNR differently. However, regardless of the exact definition of SNR, P_e decreases as SNR increases in a properly designed system, that is, everyone is in agreement that increasing SNR is a good thing to do.

The noise in the system is almost always referred to as channel noise though most of it actually arises in the front end of the receiver. The random motion of electrons in the electrical conductors comprising the front end of the receiver creates small time-varying voltages - referred to as *thermal noise* voltages - that are on the order of a few micro-volts or so. Based on experimental evidence, the thermal noise is modelled as a stationary Gaussian random process. It suffices to note that if we sample the noise at the receiver input at any time instant, then an reasonable model for this noise sample is a zero-mean Gaussian random variable whose variance is the same regardless of the choice of time instant. The reader may then wonder why it is necessary to use a filter as shown in the above figure of the receiver model. Why not just sample at the receiver input itself and make a decision based on that sample value? After all, the sample will be $s_i(T_0)$(instead of $\hat{s}_0(T_0)$) plus a Gaussian noise variable. Unfortunately, in many instances, the noise voltages can be of the same order of magnitude as (or even considerably larger than) the voltages created by the received signals, especially when the transmitter is far away, or is restricted in transmitter power. Thus, the error probability can be unacceptably large if we were to make a decision based on a sample taken at the receiver input. But, can filtering ameliorate this situation? In almost all instances, it can. Informally speaking, the noise at the receiver input is broadband noise in comparison to which the received signals are narrowband signals.1 Thus, even a simple band-pass filter which passes the signals $s_0(t)$ and $s_1(t)$ (and the in-band noise) unchanged while eliminating the out of-band noise will reduce the noise variance considerably, and thereby reduce the error probability achieved. The use of a filter can be beneficial even when the received signals are strong enough that sampling at the receiver input gives acceptably small error probability. In such a case, the filter can be used not as a device for improving the error probability from acceptably small to fantastically small, but rather as a cost-effective way of achieving the same acceptably small error probability while increasing the spatial distance over which the communication system can operate, or reducing the transmitter power which in turn can have additional side benefits such as a reduction in the size and weight of the transmitter, an increase in battery life, etc. Finally, for those still unconvinced of the utility of filtering before sampling, consider that it is universal practice to amplify the receiver input before any sampling is done, and for reasons of power efficiency and ease of implementation, amplifiers also act as band-pass filters. The reason for amplification is that it is much easier to design a circuit that triggers when its input exceeds 0.25V (where the threshold may vary $\pm$0:01V (say) due to manufacturing process variations or circuit component tolerances) than it is to design a circuit that triggers at a threshold of 0:25 $\pm$0:01 μV. Note that since P_e is determined by the ratio of signal level to noise level, and the amplifier increases the signal level and the noise level by the same factor, amplification by itself does not change the error probability−but amplification does make implementation of the sampler, A/D converter, threshold device etc. a lot easier. In summary, digital communication receivers amplify and filter the received signals (plus noise) before sampling and making a decision as to which bit was transmitted. Since the analysis of error probability is unaffected by the amplifier gain, we do not include amplifier gain explicitly in our analysis, though we do incorporate the filtering in the amplifier(s)[2] into the filter shown in the figure above. One other consequence of amplification is important, and simplifies our analysis. Remember that thermal noise is present in all the electrical conductors in the amplifier/filter combination and not just in those in the front end of the receiver. However, because of the amplification that the thermal noise present in the front end of the receiver undergoes as it passes through the amplifier/filter combination, this amplified noise from the front end is the predominant noise present in the output of the amplifier/filter combination, completely swamping out the few micro-volts contribution to the noise from the electrical conductors in the output impedance of the amplifier/filter. In other words, for all practical purposes, the noise $N(t)$ shown in the figure above can be taken to be the result of filtering the noise process in the front end of the receiver. Notice that such an assumption would not be valid in the absence of amplification since the thermal noise generated at the filter output would be comparable in magnitude to that passing through the filter and appearing at its output. Finally, following convention, we blame it all on the channel and say that the noise $N(t)$ at the filter output is the result of filtering the Gaussian channel noise process. The channel itself is called a Gaussian noise channel.

Having justified the need for the filter shown in our model, let us consider what kind of filter we should use. For a Gaussian noise channel, the smallest error probability that can possibly be achieved with given received signals $s_0(t)$ and $s_1(t)$ is achievable by using a suitable linear filter, that is, a linear time-invariant system. Call this smallest achievable error probability P_e^*. The *optimum linear filter* that achieves P_e^* is called a matched filter for signals $s_0(t)$ and $s_1(t)$. As the name implies, different signal sets have different matched filters (and achieve different minimum error probabilities). For given signals $s_0(t)$ and $s_1(t)$, the use of a linear filter not matched to them will result in a receiver with $P_e > P_e^*$. Notice also that the claim does not mean that a receiver with a nonlinear filter cannot achieve error probability P_e^*: it might well do so, but so can the linear matched filter receiver achieve error probability P_e^* . What the claim does mean is that no receiver, whether using a linear filter or a nonlinear filter, can achieve an error probability smaller than P_e^* . That is the least error probability that we can achieve with signals $s_0(t)$ and $s_1(t)$, and we can achieve it with a linear (matched filter) receiver. Looking at receivers with nonlinear filters in the hope of getting error probability smaller than P_e^* is futile.

Restricting ourselves to linear filters, what can be said about the noise process at the filter output? Since the channel noise process is a zero-mean stationary Gaussian random process that passes through the filter, the noise process at the filter output is also a zero-mean stationary Gaussian random process. Thus, regardless of the choice of sampling instant T_0, $N(T_0)$is a zero-mean Gaussian random variable with fixed variance σ^2.

Line Coding

A **line code** is the code used for data transmission of a digital signal over a transmission line. This process of coding is chosen so as to avoid overlap and distortion of signal such as inter-symbol interference.

Properties of Line Coding

Following are the properties of line coding −

- As the coding is done to make more bits transmit on a single signal, the bandwidth used is much reduced.
- For a given bandwidth, the power is efficiently used.
- The probability of error is much reduced.
- Error detection is done and the bipolar too has a correction capability.
- Power density is much favorable.
- The timing content is adequate.
- Long strings of **1s** and **0s** is avoided to maintain transparency.

Types of Line Coding

There are 3 types of Line Coding

- Unipolar
- Polar
- Bi-polar

Unipolar Signaling

Unipolar signaling is also called as **On-Off Keying** or simply **OOK**.

The presence of pulse represents a **1** and the absence of pulse represents a **0**.

There are two variations in Unipolar signaling −

- Non Return to Zero (NRZ)
- Return to Zero (RZ)

Unipolar Non-Return to Zero (NRZ)

In this type of unipolar signaling, a High in data is represented by a positive pulse called as **Mark**, which has a duration $\mathbf{T_0}$ equal to the symbol bit duration. A Low in data input has no pulse.

The following figure clearly depicts this.

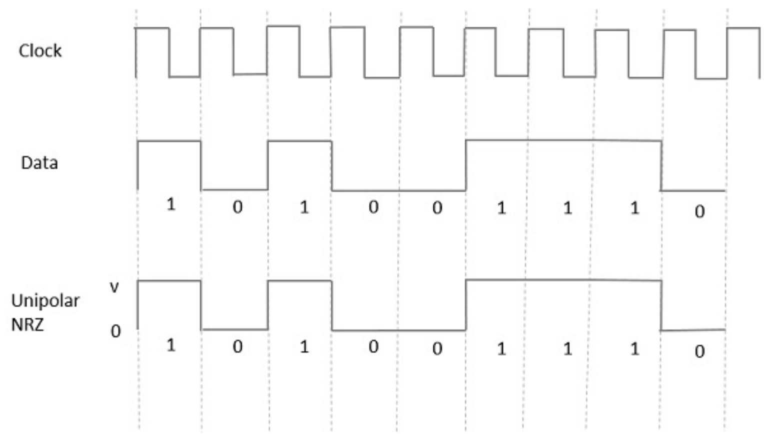


Figure 2.16: Unipolar-NRZ format

Advantages

The advantages of Unipolar NRZ are −

- It is simple.
- A lesser bandwidth is required.

Disadvantages

The disadvantages of Unipolar NRZ are −

- No error correction done.
- Presence of low frequency components may cause the signal droop.
- No clock is present.
- Loss of synchronization is likely to occur (especially for long strings of **1s** and **0s**).

Unipolar Return to Zero (RZ)

In this type of unipolar signalling, a High in data, though represented by a **Mark pulse**, its duration $\mathbf{T_0}$ is less than the symbol bit duration. Half of the bit duration remains high but it immediately returns to zero and shows the absence of pulse during the remaining half of the bit duration.

It is clearly understood with the help of the following figure.

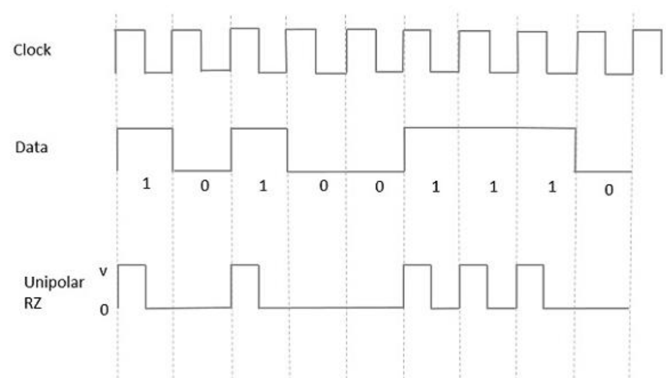


Figure 2.17: Unipolar-RZ format

Advantages

The advantages of Unipolar RZ are −

- It is simple.
- The spectral line present at the symbol rate can be used as a clock.

Disadvantages

The disadvantages of Unipolar RZ are −

- No error correction.
- Occupies twice the bandwidth as unipolar NRZ.
- The signal droop is caused at the places where signal is non-zero at 0 Hz.

Polar Signalling

There are two methods of Polar Signalling. They are −

- Polar NRZ
- Polar RZ

Polar NRZ

In this type of Polar signalling, a High in data is represented by a positive pulse, while a Low in data is represented by a negative pulse. The following figure depicts this well.

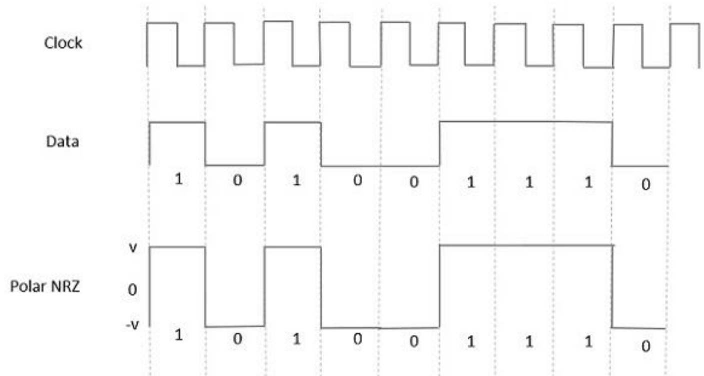


Figure 2.18: Polar-NRZ format

Advantages

The advantages of Polar NRZ are −

- It is simple.
- No low-frequency components are present.

Disadvantages

The disadvantages of Polar NRZ are −

- No error correction.
- No clock is present.
- The signal droop is caused at the places where the signal is non-zero at **0 Hz**.

Polar RZ

In this type of Polar signaling, a High in data, though represented by a **Mark pulse**, its duration $\mathbf{T_0}$ is less than the symbol bit duration. Half of the bit duration remains high but it immediately returns to zero and shows the absence of pulse during the remaining half of the bit duration.

However, for a Low input, a negative pulse represents the data, and the zero level remains same for the other half of the bit duration. The following figure depicts this clearly.

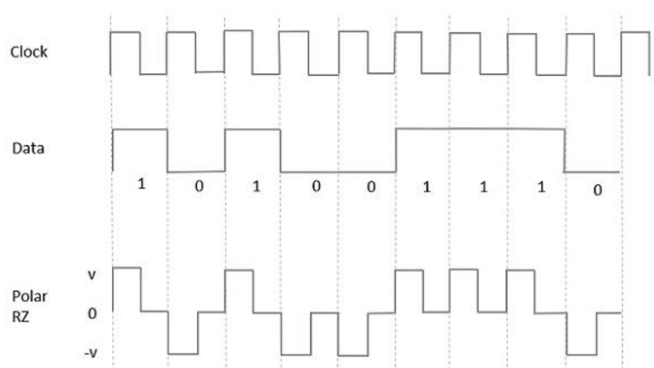


Figure 2.19: Polar-RZ format

Advantages

The advantages of Polar RZ are −

- It is simple.

- No low-frequency components are present.

Disadvantages

The disadvantages of Polar RZ are −

- No error correction.
- No clock is present.
- Occupies twice the bandwidth of Polar NRZ.
- The signal droop is caused at places where the signal is non-zero at **0 Hz**.

Bipolar Signalling

This is an encoding technique which has three voltage levels namely **+, -** and **0**. Such a signal is called as **duo-binary signal**.

An example of this type is **Alternate Mark Inversion (AMI)**. For a **1**, the voltage level gets a transition from + to – or from – to +, having alternate **1s**to be of equal polarity. A **0** will have a zero voltage level.

Even in this method, we have two types.

- Bipolar NRZ
- Bipolar RZ

From the models so far discussed, we have learnt the difference between NRZ and RZ. It just goes in the same way here too. The following figure clearly depicts this.

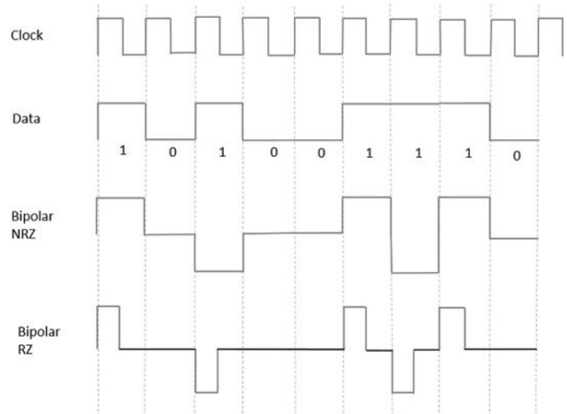


Figure 2.20: Bipolar signaling format

The above figure has both the Bipolar NRZ and RZ waveforms. The pulse duration and symbol bit duration are equal in NRZ type, while the pulse duration is half of the symbol bit duration in RZ type.

Advantages

Following are the advantages −

- It is simple.
- No low-frequency components are present.
- Occupies low bandwidth than unipolar and polar NRZ schemes.
- This technique is suitable for transmission over AC coupled lines, as signal drooping doesn't occur here.
- A single error detection capability is present in this.

Disadvantages

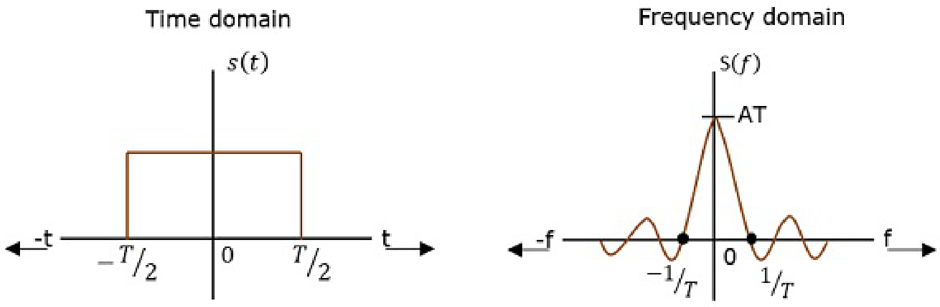
Following are the disadvantages −

- No clock is present.
- Long strings of data causes loss of synchronization.

Power Spectral Density

The function which describes how the power of a signal got distributed at various frequencies, in the frequency domain is called as **Power Spectral Density (PSD)**.

PSD is the Fourier Transform of Auto-Correlation (Similarity between observations). It is in the form of a rectangular pulse.



$$s(t) = A \quad |t| < T/2$$
$$= 0 \quad |t| > T/2$$

$$s(f) \;=\; AT\frac{\sin(\pi T f)}{\pi T f}$$

PSD Derivation

According to the Einstein-Wiener-Khintchine theorem, if the auto correlation function or power spectral density of a random process is known, the other can be found exactly.

Hence, to derive the power spectral density, we shall use the time auto-correlation $\left(R_x(\tau)\right)$ of a power signal $x(t)$ as shown below.

$$R_x(\tau) = \lim_{T_p \to \infty} \frac{1}{T_p} \int_{-\frac{T_p}{2}}^{\frac{T_p}{2}} x(t)\, x(t+\tau) dt$$

Since $x(t)$ consists of impulses, $R_x(\tau)$ can be written as

$$R_x(\tau) = \frac{1}{T} \sum_{n=-\infty}^{\infty} R_n\, \delta(\tau - nT)$$

Where $R_n = \lim_{N \to \infty} \frac{1}{N} \sum_k a_k a_{k+n}$

Getting to know that $R_n = R_{-n}$ for real signals, we have

$$S_x(w) = \frac{1}{T}\left(R_0 + 2\sum_{n=1}^{\infty} R_n \cos nwT\right)$$

Since the pulse filter has the spectrum of (w)↔f(t), we have

$$s_y(w) = |F(w)|^2 S_x(w)$$

$$= \frac{|F(w)|^2}{T}\left(\sum_{n=-\infty}^{\infty} R_n\, e^{-jnwT_b}\right)$$

$$= \frac{|F(w)|^2}{T}\left(R_0 + 2\sum_{n=1}^{\infty} R_n \cos nwT\right)$$

Hence, we get the equation for Power Spectral Density. Using this, we can find the PSD of various line codes.

Bi-phase Encoding

The signal level is checked twice for every bit time, both initially and in the middle. Hence, the clock rate is double the data transfer rate and thus the modulation rate is also doubled. The clock is taken from the signal itself. The bandwidth required for this coding is greater.

There are two types of Bi-phase Encoding.

- Bi-phase Manchester
- Differential Manchester

Bi-phase Manchester

In this type of coding, the transition is done at the middle of the bit-interval. The transition for the resultant pulse is from High to Low in the middle of the interval, for the input bit 1. While the transition is from Low to High for the input bit **0**.

Differential Manchester

In this type of coding, there always occurs a transition in the middle of the bit interval. If there occurs a transition at the beginning of the bit interval, then the input bit is **0**. If no transition occurs at the beginning of the bit interval, then the input bit is **1**.

The following figure illustrates the waveforms of NRZ-L, NRZ-I, Bi-phase Manchester and Differential Manchester coding for different digital inputs.

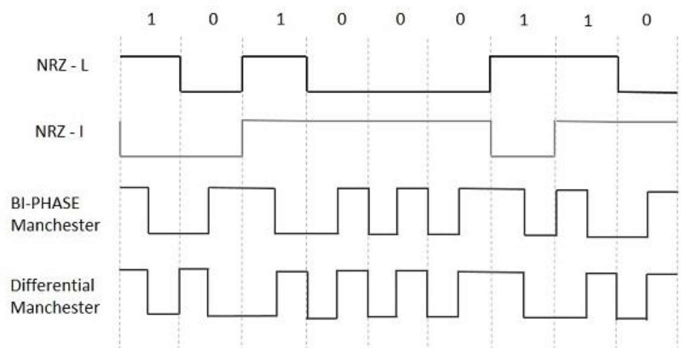


Figure 2.21: Manchester signaling format

After going through different types of coding techniques, we have an idea on how the data is prone to distortion and how the measures are taken to prevent it from getting affected so as to establish a reliable communication.

There is another important distortion which is most likely to occur, called as **Inter-Symbol Interference (ISI)**.

Inter Symbol Interference

This is a form of distortion of a signal, in which one or more symbols interfere with subsequent signals, causing noise or delivering a poor output.

Causes of ISI

The main causes of ISI are −

- Multi-path Propagation
- Non-linear frequency in channels

The ISI is unwanted and should be completely eliminated to get a clean output. The causes of ISI should also be resolved in order to lessen its effect.

To view ISI in a mathematical form present in the receiver output, we can consider the receiver output.

The receiving filter output $y(t)$ is sampled at time $t_i = iT_b$ (with **i** taking on integer values), yielding –

$$y(t_i) = \mu \sum_{k=-\infty}^{\infty} a_k\, p(iT_b - kT_b) = \mu a_i + \mu \sum_{\substack{k=-\infty \\ k\neq i}}^{\infty} a_k p(iT_b - kT_b)$$

In the above equation, the first term μa_i is produced by the $\mathbf{i^{th}}$ transmitted bit.

The second term represents the residual effect of all other transmitted bits on the decoding of the i^{th} bit. This residual effect is called as **Inter Symbol Interference**.

In the absence of ISI, the output will be −

$$y(t_i) = \mu a_i$$

This equation shows that the i^{th} bit transmitted is correctly reproduced. However, the presence of ISI introduces bit errors and distortions in the output.

While designing the transmitter or a receiver, it is important that you minimize the effects of ISI, so as to receive the output with the least possible error rate.

Correlative Coding

So far, we've discussed that ISI is an unwanted phenomenon and degrades the signal. But the same ISI if used in a controlled manner, is possible to achieve a bit rate of **2W** bits per second in a channel of bandwidth **W** Hertz. Such a scheme is called as **Correlative Coding** or **Partial response signaling schemes**.

Since the amount of ISI is known, it is easy to design the receiver according to the requirement so as to avoid the effect of ISI on the signal. The basic idea of correlative coding is achieved by considering an example of **Duo-binary Signalling**.

Duo-binary Signalling

The name duo-binary means doubling the binary system's transmission capability. To understand this, let us consider a binary input sequence $\{a_k\}$ consisting of uncorrelated binary digits each having a duration T_a seconds. In this, the signal **1** is represented by a **+1** volt and the symbol **0** by a **-1** volt.

Therefore, the duo-binary coder output c_k is given as the sum of present binary digit a_k and the previous value a_{k-1} as shown in the following equation.

$$c_k = a_k + a_{k-1}$$

The above equation states that the input sequence of uncorrelated binary sequence $\{a_k\}$ is changed into a sequence of correlated three level pulses $\{c_k\}$. This correlation between the pulses may be understood as introducing ISI in the transmitted signal in an artificial manner.

Eye Pattern

An effective way to study the effects of ISI is the **Eye Pattern**. The name Eye Pattern was given from its resemblance to the human eye for binary waves. The interior region of the eye pattern is called the **eye opening**. The following figure shows the image of an eye-pattern.

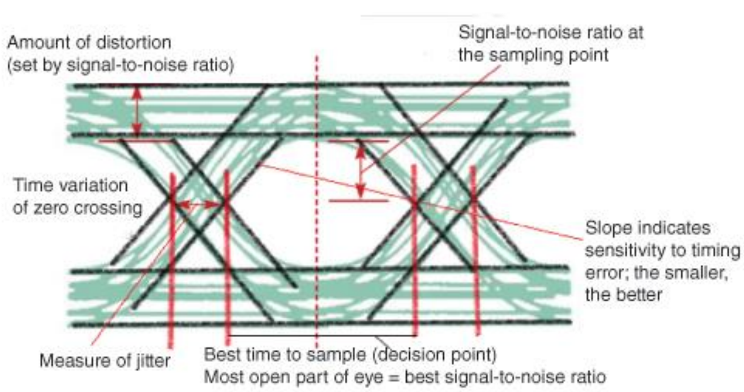


Figure 2.22: Eye pattern

Jitter is the short-term variation of the instant of digital signal, from its ideal position, which may lead to data errors.

When the effect of ISI increases, traces from the upper portion to the lower portion of the eye opening increases and the eye gets completely closed, if ISI is very high.
An eye pattern provides the following information about a particular system.

- Actual eye patterns are used to estimate the bit error rate and the signal-to-noise ratio.
- The width of the eye opening defines the time interval over which the received wave can be sampled without error from ISI.
- The instant of time when the eye opening is wide, will be the preferred time for sampling.
- The rate of the closure of the eye, according to the sampling time, determines how sensitive the system is to the timing error.
- The height of the eye opening, at a specified sampling time, defines the margin over noise.

Hence, the interpretation of eye pattern is an important consideration.

Equalization

For reliable communication to be established, we need to have a quality output. The transmission losses of the channel and other factors affecting the quality of the signal, have to be treated. The most occurring loss, as we have discussed, is the ISI.
To make the signal free from ISI, and to ensure a maximum signal to noise ratio, we need to implement a method called **Equalization**. The following figure shows an equalizer in the receiver portion of the communication system.

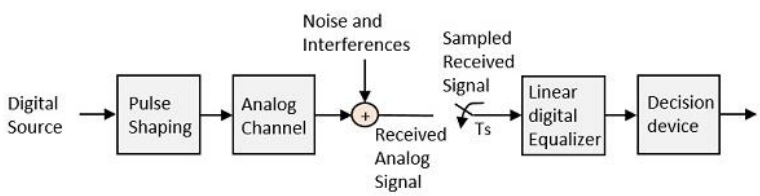


Figure 2.23: Block diagram of an equalizer

The noise and interferences which are denoted in the figure, are likely to occur, during transmission. The regenerative repeater has an equalizer circuit, which compensates the transmission losses by shaping the circuit. The Equalizer is feasible to get implemented.

Multiple choice questions:

(1) The number of bits required to represent a 256 quantization level in PCM is
a) 7 b) 8 c) 5 d) 6

(2) Eye pattern is used to study
a) Bit error rate b) Error vector magnitude
c) Quantization noise d) Inter Symbol Interference

(3) Alternate Mark Inversion (AMI) signaling is known as
a) Bipolar signaling b) Polar signaling
c) Manchester signaling d) Unipolar signaling

(4) The use of non-uniform quantization leads to
a) Reduction of transmission bandwidth b) Increase in maximum SNR
c) Increase in SNR for low bend signal d) Simplification of quantization process

(5) The spectral density of white noise is
a) Exponential b) Uniform c) Poisson d) Gaussian

(6) A rectangular pulse of duration T is applied to matched filter. The output of the filter is a
a) Rectangular pulse of duration T b) Rectangular pulse of duration 2T
c) Triangular pulse d) Sine function

(7) The main advantage of PCM system is
a) possibility of TDM b) less channel bandwidth

c) less transmission power d) better noise performance

Sample Questions:

1. What are the advantages of digital communication system over analog communication System?
2. State and explain sampling theorem for band-limited signals.
3. What are quantization and quantization error?
4. Derive the relation for signaling rate and transmission bandwidth in PCM system and also derive signal to noise ratio.
5. What is companding? Why is it needed? Explain A-Law and μ-Law companding.
6. Explain the delta modulation with transmitter and receiver block diagram.
7. What is the slope overload distortion and granular noise in delta-modulation?
8. What are the advantages of adaptive-delta modulation over ordinary delta-modulation?
9. What do you mean by DPCM?
10. A Television signal having a bandwidth of 10.2 MHz is transmitted using binary PCM system. Given that the number of quantization levels is 512. Determine: (i) Code word length, (ii) Transmission bandwidth, (iii) Final bit rate and (iv) Output signal to quantization noise ratio.
11. Why do we need to use the discrete PAM formats? Write the properties of Line Coding.
12. What is the difference between source coding and line coding?
13. Given the data stream 1110010100. Sketch the transmitted sequence of rectangular pulses for each of the following line codes: a) Unipolar NRZ, b) Unipolar RZ, c) Polar RZ, d) Polar NRZ e) Bipolar NRZ and f) Manchester
14. Determine the bandwidth of bipolar format from the PSD. What are the advantages and disadvantages of bipolar signaling format? -Explain.
15. What do you mean by match filter?
16. Prove that the SNR at the output of a matched filter is $8E_s/\eta$. Where Es is the signal energy and $\eta/2 = G_n(f)$, for white gaussian noise. And hence deduce the transfer function of a matched filter.
17. What is intersymbol interference(ISI)?
18. What is Nyquist criterion for zero ISI?
19. What are the limitations of Nyquist pulse? How is it solved using Raised Cosine pulse?
20. What is Eye pattern? How is it generated in CRO? What information we get from it?

Module-3

Digital Modulation

Digital Modulation provides more information capacity, high data security, quicker system availability with great quality communication. Hence, digital modulation techniques have a greater demand, for their capacity to convey larger amounts of data than analog modulation techniques.

There are many types of digital modulation techniques and also their combinations, depending upon the need. Of them all, we will discuss the prominent ones.

ASK – Amplitude Shift Keying

The amplitude of the resultant output depends upon the input data whether it should be a zero level or a variation of positive and negative, depending upon the carrier frequency.

FSK – Frequency Shift Keying

The frequency of the output signal will be either high or low, depending upon the input data applied.

PSK – Phase Shift Keying

The phase of the output signal gets shifted depending upon the input. These are mainly of two types, namely Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK), according to the number of phase shifts. The other one is Differential Phase Shift Keying (DPSK) which changes the phase according to the previous value.

Amplitude Shift Keying (ASK) is a type of Amplitude Modulation which represents the binary data in the form of variations in the amplitude of a signal.
Any modulated signal has a high frequency carrier. The binary signal when ASK modulated, gives a zero value for Low input while it gives the carrier output for High input.
The following figure represents ASK modulated waveform along with its input.

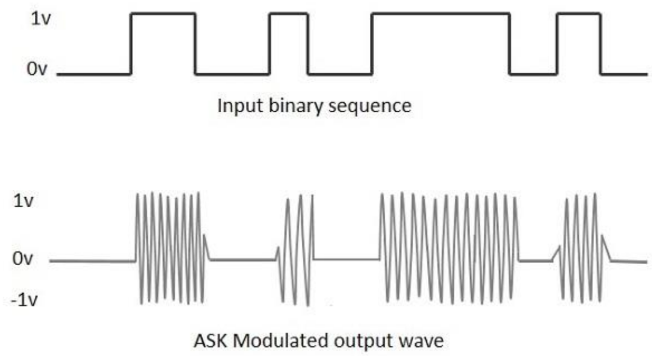


Figure 3.1: ASK modulated waveform

ASK Modulator

The ASK modulator block diagram comprises of the carrier signal generator, the binary sequence from the message signal and the band-limited filter. Following is the block diagram of the ASK Modulator.

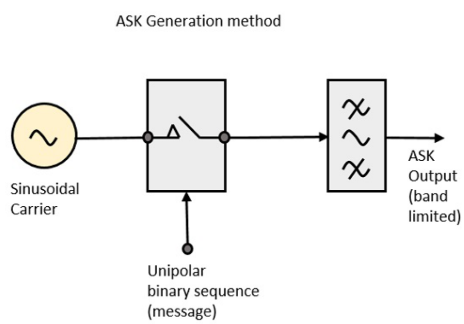


Figure 3.2(a): ASK transmitter

ASK Demodulator

Synchronous ASK detector consists of a Square law detector, low pass filter, a comparator, and a voltage limiter. Following is the block diagram for the same.

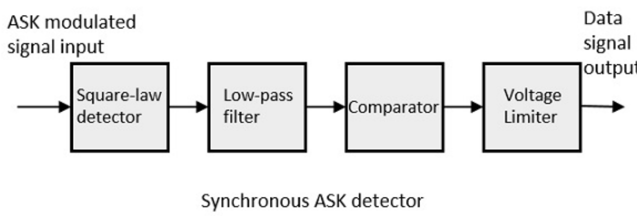


Figure 3.2(b): ASK receiver

Frequency Shift Keying (FSK) is the digital modulation technique in which the frequency of the carrier signal varies according to the digital signal changes. FSK is a scheme of frequency modulation.

The output of a FSK modulated wave is high in frequency for a binary High input and is low in frequency for a binary Low input. The binary 1s and 0s are called Mark and Space frequencies.

The following image is the diagrammatic representation of FSK modulated waveform along with its input.

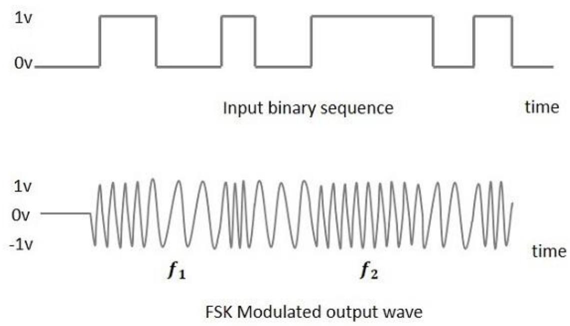


Figure 3.3: FSK modulated waveform

FSK Modulator

The FSK modulator block diagram comprises of two oscillators with a clock and the input binary sequence. Following is its block diagram.

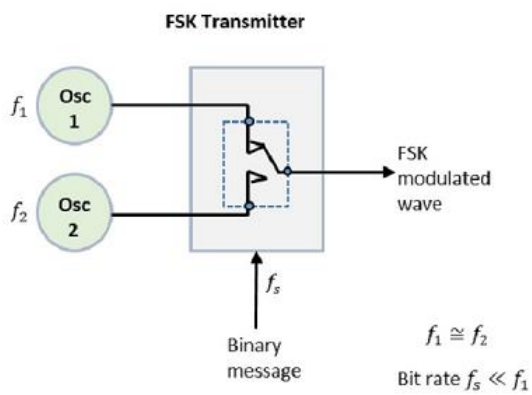


Figure 3.4(a): FSK transmitter

FSK Demodulator

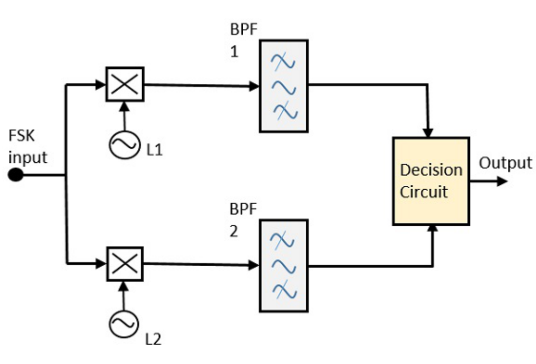


Figure 3.4(b): FSK receiver

Phase Shift Keying (PSK) is the digital modulation technique in which the phase of the carrier signal is changed by varying the sine and cosine inputs at a particular time. PSK technique is widely used for wireless LANs, bio-metric, contactless operations, along with RFID and Bluetooth communications.

PSK is of two types, depending upon the phases the signal gets shifted. They are −

Binary Phase Shift Keying (BPSK)

This is also called as 2-phase PSK or Phase Reversal Keying. In this technique, the sine wave carrier takes two phase reversals such as 0° and 180°.

BPSK is basically a Double Side Band Suppressed Carrier (DSBSC) modulation scheme, for message being the digital information.

Quadrature Phase Shift Keying (QPSK)

This is the phase shift keying technique, in which the sine wave carrier takes four phase reversals such as 0°, 90°, 180°, and 270°.

If this kind of techniques are further extended, PSK can be done by eight or sixteen values also, depending upon the requirement.

BPSK Modulator

The block diagram of Binary Phase Shift Keying consists of the balance modulator which has the carrier sine wave as one input and the binary sequence as the other input. Following is the diagrammatic representation.

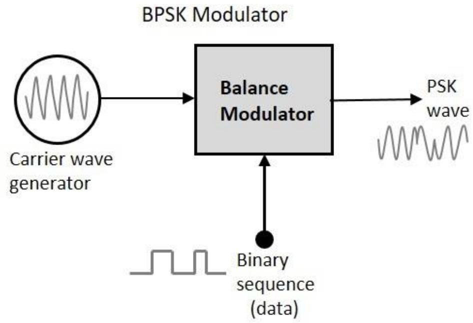


Figure 3.5(a): BPSK transmitter

The modulation of BPSK is done using a balance modulator, which multiplies the two signals applied at the input. For a zero binary input, the phase will be 0° and for a high input, the phase reversal is of 180°.

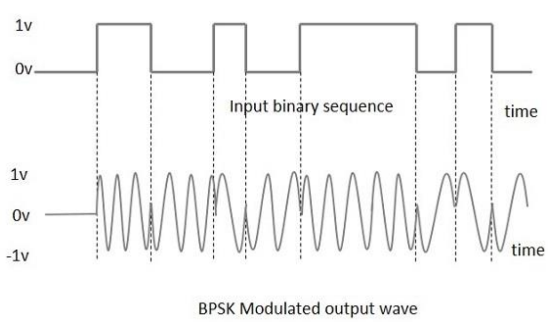


Figure 3.5(b): BPSK modulated waveform

BPSK Demodulator

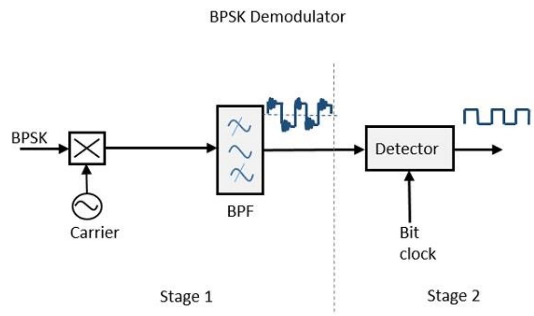


Figure 3.5(c): BPSK detection

The **Quadrature Phase Shift Keying (QPSK)** is a variation of BPSK, and it is also a Double Side Band Suppressed Carrier (DSBSC) modulation scheme, which sends two bits of digital information at a time, called as **bigits**.

Instead of the conversion of digital bits into a series of digital stream, it converts them into bit pairs. This decreases the data bit rate to half, which allows space for the other users.

QPSK Modulator

The QPSK Modulator uses a bit-splitter, two multipliers with local oscillator, a 2-bit serial to parallel converter, and a summer circuit. Following is the block diagram for the same.

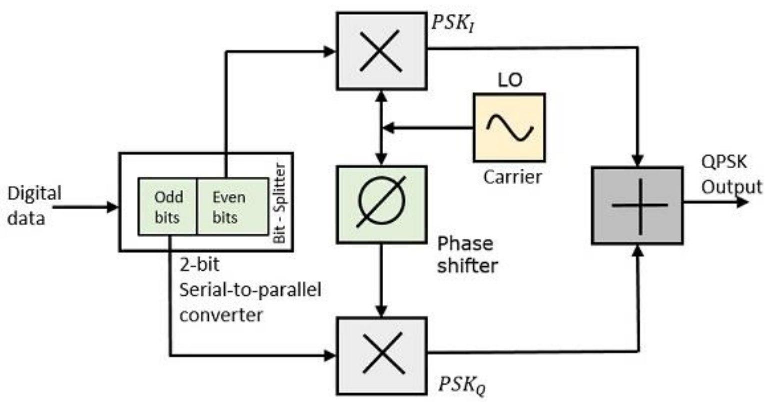


Figure 3.6(a): QPSK generation

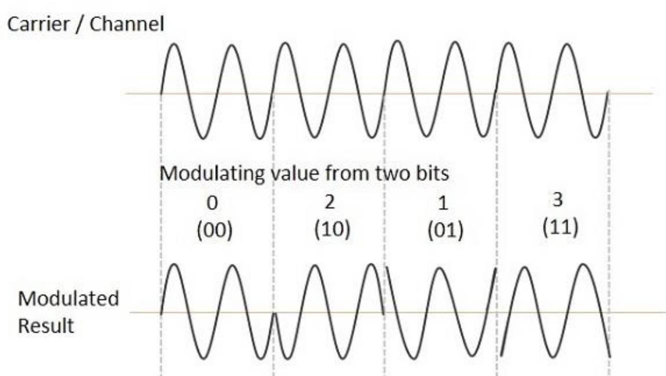


Figure 3.6(b): QPSK waveform

QPSK Demodulator

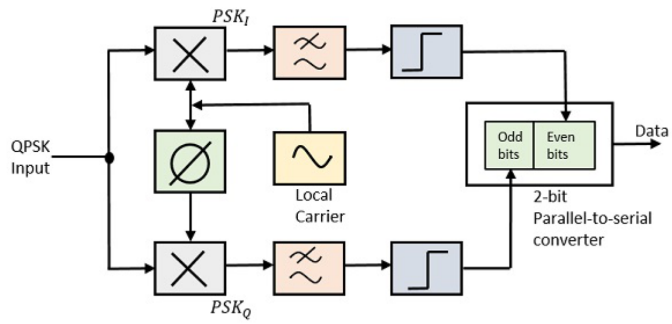


Figure 3.6(c): QPSK detection

Spread Spectrum Modulation

A collective class of signalling techniques are employed before transmitting a signal to provide a secure communication, known as the Spread Spectrum Modulation. The main advantage of spread spectrum communication technique is to prevent “interference” whether it is intentional or unintentional.

The signals modulated with these techniques are hard to interfere and cannot be jammed. An intruder with no official access is never allowed to crack them. Hence, these techniques are used for military purposes. These spread spectrum signals transmit at low power density and has a wide spread of signals.

Pseudo-Noise Sequence

A coded sequence of 1s and 0s with certain auto-correlation properties, called as Pseudo-Noise coding sequence is used in spread spectrum techniques. It is a maximum-length sequence, which is a type of cyclic code.

Narrow-band and Spread-spectrum Signals

Both the Narrow band and Spread spectrum signals can be understood easily by observing their frequency spectrum as shown in the following figures.

The **Narrow-band signals** have the signal strength concentrated as shown in the following frequency spectrum figure.

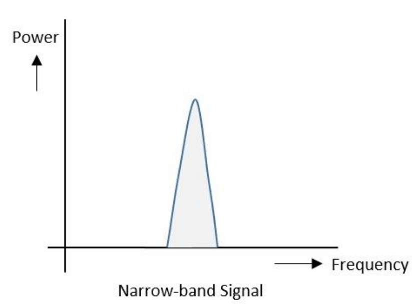


Figure 3.7(a): Narrowband signal

Following are some of its features −

- Band of signals occupy a narrow range of frequencies.
- Power density is high.
- Spread of energy is low and concentrated.

Though the features are good, these signals are prone to interference.

The **spread spectrum signals** have the signal strength distributed as shown in the following frequency spectrum figure.

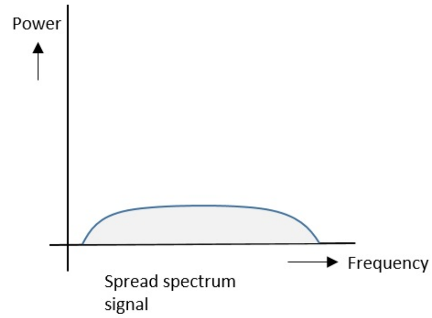


Figure 3.7(b): Spread spectrum signal

Following are some of its features −

- Band of signals occupy a wide range of frequencies.
- Power density is very low.
- Energy is wide spread.

With these features, the spread spectrum signals are highly resistant to interference or jamming. Since multiple users can share the same spread spectrum bandwidth without interfering with one another, these can be called as **multiple access** techniques.

Sample Questions:

1) What are the differences between coherent and non-coherent binary modulation techniques?
2) Describe ASK demodulation through coherent detection and find the bandwidth.
3) Describe PSK demodulation through coherent detection.
4) What are the disadvantages of BPSK and how is it improved?
5) Derive the expression for the spectrum of BPSK and sketch the same. Also find the bandwidth.
6) Draw the block diagram for generation and detection of the BFSK signal and explain clearly its operation.
7) Explain the principle of operation of QPSK transmitter and receiver with suitable block diagram. Sketch its signal space diagram.
8) Compare the bandwidths of QPSK and BPSK.
9) Explain the concept of spread spectrum.
10) What is a PN sequence? What are the properties of a PN sequence?
11) Why PN sequence is called pseudo noise?

Module-4

Information Theory & Coding

Information is the source of a communication system, whether it is analog or digital. **Information theory** is a mathematical approach to the study of coding of information along with the quantification, storage, and communication of information.

Conditions of Occurrence of Events

If we consider an event, there are three conditions of occurrence.

- If the event has not occurred, there is a condition of **uncertainty**.
- If the event has just occurred, there is a condition of **surprise**.
- If the event has occurred, a time back, there is a condition of having some **information**.

These three events occur at different times. The difference in these conditions help us gain knowledge on the probabilities of the occurrence of events.

Entropy:

When we observe the possibilities of the occurrence of an event, how surprising or uncertain it would be, it means that we are trying to have an idea on the average content of the information from the source of the event.

Entropy can be defined as a measure of the average information content per source symbol. **Claude Shannon**, the "father of the Information Theory", provided a formula for it as −

$$H = -\sum_i p_i \log_b p_i$$

Where $\mathbf{p_i}$ is the probability of the occurrence of character number **i** from a given stream of characters and **b** is the base of the algorithm used. Hence, this is also called as **Shannon's Entropy**.

The amount of uncertainty remaining about the channel input after observing the channel output, is called as **Conditional Entropy**. It is denoted by H(x|y)

Mutual Information

Let us consider a channel whose output is **Y** and input is **X**

Let the entropy for prior uncertainty be **X = H(x)**

(This is assumed before the input is applied)

To know about the uncertainty of the output, after the input is applied, let us consider Conditional Entropy, given that $\mathbf{Y = y_k}$

$$H(x|y_k) = \sum_{j=0}^{j-1} p(x_j|y_k) \log_2 \left[\frac{1}{(x_j|y_k)}\right]$$

This is a random variable for $H(X|y = y_0)$ … … … … … … … … … … . $H(X|y = y_k)$ with probabilities $p(y_0)$ … … … … $p(y_{k-1})$ respectively.

The mean value of $H(X|y = y_k)$for output alphabet **y** is −

$$H(X|Y) = \sum_{k=0}^{k-1} H(X|y = y_k)\, p(y_k)$$

$$= \sum_{k=0}^{k-1} \sum_{j=0}^{j-1} p(x_j, y_k) \log_2 \left[\frac{1}{(x_j|y_k)}\right]$$

Now, considering both the uncertainty conditions (before and after applying the inputs), we come to know that the difference, i.e. H(x)−H(x|y)H(x)−H(x|y) must represent the uncertainty about the channel input that is resolved by observing the channel output.

This is called as the **Mutual Information** of the channel.

Denoting the Mutual Information as $I(x;y)$, we can write the whole thing in an equation, as follows

$$I(x;y) = H(x) - H(x|y)$$

Hence, this is the equational representation of **Mutual Information**.

Channel Capacity

We have so far discussed mutual information. The maximum average mutual information, in an instant of a signaling interval, when transmitted by a discrete memory-less channel, the probabilities of the rate of maximum reliable transmission of data, can be understood as the **channel capacity**.

It is denoted by C and is measured in **bits per channel** use.

Source Coding Theorem

The Code produced by a discrete memory-less source, has to be efficiently represented, which is an important problem in communications. For this to happen, there are code words, which represent these source codes.

For example, in telegraphy, we use Morse code, in which the alphabets are denoted by **Marks** and **Spaces**. If the letter **E** is considered, which is mostly used, it is denoted by **"."** Whereas the letter **Q** which is rarely used, is denoted by **"--.-"**

Let us take a look at the block diagram.

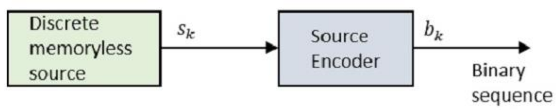


Figure 4.1: Block diagram of source coding

Where **S$_k$** is the output of the discrete memory-less source and **b$_k$** is the output of the source encoder which is represented by **0s** and **1s**.

The encoded sequence is such that it is conveniently decoded at the receiver.

Let us assume that the source has an alphabet with **k** different symbols and that the **kth** symbol **S$_k$** occurs with the probability **P$_k$**, where **k = 0, 1…k-1**.

Let the binary code word assigned to symbol **S$_k$**, by the encoder having length **l$_k$**, measured in bits.

Hence, we define the average code word length L of the source encoder as

$$\bar{L} = \sum_{k=0}^{k-1} p_k l_k$$

$\bar{L}$ represents the average number of bits per source symbol

If $L_{min} = minimum\ possible\ value\ of\ \bar{L}$

Then **coding efficiency** can be defined as

$$\eta = \frac{L_{min}}{\bar{L}}$$

With $\bar{L} \geq L_{min}$ we will have $\eta \leq 1$

However, the source encoder is considered efficient when $\eta = 1$

For this, the value L_{min} has to be determined.

Let us refer to the definition, *"Given a discrete memory-less source of entropy $H(\delta)$, the average code-word length **L** for any source encoding is bounded as $\bar{L} \geq H(\delta)$."*

Hence with $L_{min} = H(\delta)$, the efficiency of the source encoder in terms of Entropy $H(\delta)$ may be written as

$$\eta = \frac{H(\delta)}{\bar{L}}$$

This source coding theorem is called as **noiseless coding theorem** as it establishes an error-free encoding. It is also called as **Shannon's first theorem**.

Error Control Coding

Noise or Error is the main problem in the signal, which disturbs the reliability of the communication system. **Error control coding** is the coding procedure done to control the occurrences of errors. These techniques help in Error Detection and Error Correction.

There are many different error correcting codes depending upon the mathematical principles applied to them. But, historically, these codes have been classified into **Linear block codes** and **Convolution codes**.

Linear Block Codes

In the linear block codes, the parity bits and message bits have a linear combination, which means that the resultant code word is the linear combination of any two code words.

Let us consider some blocks of data, which contains **k** bits in each block. These bits are mapped with the blocks which has **n** bits in each block. Here **n** is greater than **k**. The transmitter adds redundant bits which are **(n-k)** bits. The ratio **k/n** is the **code rate**. It is denoted by **r** and the value of **r** is **r < 1**.

The **(n-k)** bits added here, are **parity bits**. Parity bits help in error detection and error correction, and also in locating the data. In the data being transmitted, the left most bits of the code word correspond to the message bits, and the right most bits of the code word correspond to the parity bits.

Systematic Code

Any linear block code can be a systematic code, until it is altered. Hence, an unaltered block code is called as a **systematic code**.

Following is the representation of the **structure of code word**, according to their allocation.

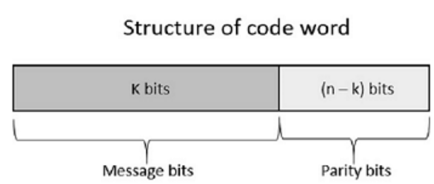


Figure 4.2: Generation of systematic code

If the message is not altered, then it is called as systematic code. It means, the encryption of the data should not change the data.

Convolution Codes

So far, in the linear codes, we have discussed that systematic unaltered code is preferred. Here, the data of total **n** bits if transmitted, **k** bits are message bits and **(n-k)** bits are parity bits. In the process of encoding, the parity bits are subtracted from the whole data and the message bits are encoded. Now, the parity bits are again added and the whole data is again encoded. The following figure quotes an example for blocks of data and stream of data, used for transmission of information.

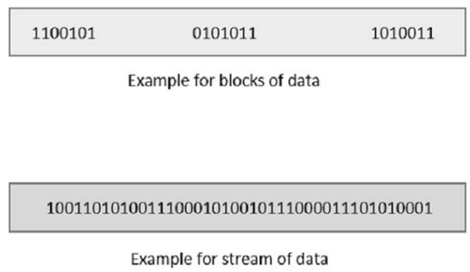


Figure 4.3: Generation of convolution code

The whole process, stated above is tedious which has drawbacks. The allotment of buffer is a main problem here, when the system is busy.

This drawback is cleared in convolution codes. Where the whole stream of data is assigned symbols and then transmitted. As the data is a stream of bits, there is no need of buffer for storage.

Hamming Codes

The linearity property of the code word is that the sum of two code words is also a code word. Hamming codes are the type of **linear error correcting** codes, which can detect up to two bit errors or they can correct one bit errors without the detection of uncorrected errors.

While using the hamming codes, extra parity bits are used to identify a single bit error. To get from one-bit pattern to the other, few bits are to be changed in the data. Such number of bits can be termed as **Hamming distance**. If the parity has a distance of 2, one-bit flip can be detected. But this can't be corrected. Also, any two bit flips cannot be detected.

However, Hamming code is a better procedure than the previously discussed ones in error detection and correction.

BCH Codes

BCH codes are named after the inventors **B**ose, **C**haudari and **H**ocquenghem. During the BCH code design, there is control on the number of symbols to be corrected and hence multiple bit correction is possible. BCH codes is a powerful technique in error correcting codes.

For any positive integers $\mathbf{m \geq 3}$ and $\mathbf{t < 2^{m-1}}$ there exists a BCH binary code. Following are the parameters of such code.

Block length $\mathbf{n = 2^m - 1}$

Number of parity-check digits $\mathbf{n - k \leq mt}$

Minimum distance $\mathbf{d_{min} \geq 2t + 1}$

This code can be called as **t-error-correcting BCH code**.

Cyclic Codes

The cyclic property of code words is that any cyclic-shift of a code word is also a code word. Cyclic codes follow this cyclic property.

For a linear code **C**, if every code word i.e., **C = (C1, C2, ... Cn)** from C has a cyclic right shift of components, it becomes a code word. This shift of right is equal to **n-1** cyclic left shifts. Hence, it is invariant under any shift. So, the linear code **C**, as it is invariant under any shift, can be called as a **Cyclic code**.

Cyclic codes are used for error correction. They are mainly used to correct double errors and burst errors.

Hence, these are a few error correcting codes, which are to be detected at the receiver. These codes prevent the errors from getting introduced and disturb the communication. They also prevent the signal from getting tapped by unwanted receivers.

Multiple Choice Questions:

1) Which coding terminology deals with the inverse operation of assigned words of second language corresponding to the words in the first language?

a. Enciphering **b.** Deciphering

c. Codeword **d.** Codebook

2) Huffman coding technique is adopted for constructing the source code with ________ redundancy.

a. Maximum **b.** Constant

c. Minimum **d.** Unpredictable

3) Which type of channel does not represent any correlation between input and output symbols?

a. Noiseless channel b. Lossless channel

c. Useless channel d. Deterministic channel

4) In digital communication system, smaller the code rate, _________are the redundant bits.

a. less b. more c. equal d. unpredictable

5) A zero source generates two messages with prob. 0.8 and 0.2. These are coded as 1 and 0.2. The code efficiency is

a. 0.2 b. 0.5 c. 0.7 d. 1.0

6) The capacity of a communication channel with a bandwidth of 4 KHz and 15 SNR is approx

a. 20 kbps b. 16 kbps c. 10 kbps d. 8 kbps

7) A source generates 4 messages. The entropy of the source will be maximum when

a. all probabilities equal.

b. One of the probabilities equal 1 and 2 others are 0

c. the probabilities are ½, ¼ and 1/2.

d. the two probabilities are ½ each and other zero.

Sample Questions:

1) Explain the concept of information.
2) What is entropy?
3) What is information rate?
4) What are discrete memory-less channels (DMC)?
5) What is source coding?
6) What are the types of error control methods?
7) Write the advantages and disadvantages of Cyclic Codes.

References:

1. Taub and Schilling , "Principles of Communication Systems", 2nd ed., Mc-Graw Hill

2. B.P.Lathi -Communication Systems- BS Publications

3. V Chandra Sekar – Analog Communication- Oxford University Press

4. Carlson—Communication System,4/e , Mc-Graw Hill

5. Proakis & Salehi Fundamentals of Communication Systems- Pearson

6. Singh & Sapre—Communication Systems: 2/e, TMH

7. P K Ghosh- Principles of Electrical Communications- University Press

8. L.W.Couch Ii, “Digital and Analog Communication Systems”, 2/e, Macmillan Publishing

9. Blake, Electronic Communication Systems- Cengage Learning

10. S Sharma, Analog Communication Systems- Katson Books