COMMUNICATION SYSTEMS

20A04402T

LECTURE NOTES

B.Tech – ECE – II-II Semester

Prepared by,

T. CHAKRAPANI

Associate Professor

ECE Department

Electronics and Communication Engineering

St.Johns College of Engineering & Technology

Yerrakota, Yemmiganur-518360, Kurnool(D), A.P.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR
(Established by Govt. of A.P., ACT No.30 of 2008)
ANANTHAPURAMU – 515 002 (A.P) INDIA

Electronics & Communication Engineering

Course Code	COMMUNICATION SYSTEMS		L	T	P	C
20A04402T			3	0	0	3
Pre-requisite	Signals & Systems	Semester	IV			

Course Objectives:

- To introduce various modulation and demodulation techniques of analog and digital communication systems.
- To analyze different parameters of analog and digital communication techniques.
- To Know Noise Figure in AM & FM receiver systems.
- To understand Function of various stages of AM, FM transmitters and Know Characteristics of AM &FMreceivers.
- To analyze the performance of various digital modulation techniques in the presence of AWGN.
- To evaluate the performance of each modulation scheme to know the merits and demerits interms of bandwidth and power efficiency

Course Outcomes (CO):

CO1: Recognize/List the basic terminology used in analog and digital communication techniques for transmission of information/data.
CO2: Explain/Discuss the basic operation of different analog and digital communication systems at baseband and passband level.
CO3: Compute various parameters of baseband and passband transmission schemes by applying basic engineering knowledge.
CO4: Analyze/Investigate the performance of different modulation & demodulation techniques to solve complex problems in the presence of noise.
CO5: Evaluate/Assess the performance of all analog and digital modulation techniques to know the merits and demerits of each one of them in terms of bandwidth and power efficiency.

UNIT - I	Continuous Wave Modulation	15 Hrs

Introduction: The communication Process, Communication Channels, Baseband and Passband Signals, Analog vs Digital Communications, Need for the modulation.
Amplitude Modulation(AM): AM and its modifications – DSB, SSB,VSB. Frequency Translation, Frequency Division Multiplexing (FDM).
Angle Modulation:Frequency Modulation(FM), Phase Modulation, PLL, Nonlinear Effects in FM, Superheterodyne Receivers.

UNIT - II	Noise and Pulse Modulation	12 Hrs

Introduction to Noise: Types of Noise, Receiver Model,Noise in AM, DSB, SSB, and FM Receivers, Pre-Emphasis and De-emphasis in FM.
Introduction to Pulse Modulation: The Sampling Process, PAM, TDM, Bandwidth-Noise Trade off, Quantization process, PCM, Noise considerations in PCM systems, Delta Modulation, DPCM, Coding speech at low bit rates.

UNIT - III	Baseband Pulse Transmission	10 Hrs

Introduction, Matched Filter, Properties of Matched Filter, Error rate due to noise, Inter Symbol Interference (ISI), Nyquist Criterion for distortion less baseband binary transmission, Correlative level coding, Baseband M-ary PAM transmission, QAM, MAP and ML decoding, Equalization, Eye pattern.

UNKT - IV	Digital Passband Transmission	8 Hrs

Introduction, Passband Transmission Model, Gram-Schmidt Orthogonalization Procedure, Geometric Interpretation of Signals, Response of bank of correlators in noise, Correlation receiver, Probability of Error, Detection of Signals with unknown phase.

Electronics & Communication Engineering

UNIT - V	Digital Modulation Schemes & Information Theory	12 Hrs
Coherent Digital Modulation Schemes – ASK, BPSK, BFSK, QPSK, Non-coherent BFSK, DPSK. M-ary Modulation Techniques, Power Spectra, Bandwidth Efficiency, Timing and Frequency synchronization. Information theory: Entropy, Mutual Information and Channel capacity theorem.		
Textbooks:		
1. Simon Haykin, "Communication Systems", JohnWiley& Sons, 4th Edition, 2004. 2. B. P. Lathi, Zhi Ding " Modern Digital and Analog Communication Systems", Oxford press, 2011.		
References:		
1.Sam Shanmugam, "Digital and Analog Communication Systems",JohnWiley& Sons, 1999. 2. Bernard Sklar, F. J. harris"Digial Communications: Fundamentals andApplications", Pearson Publications, 2020. 3. Taub and Schilling, " Principles of Communication Systems", Tata McGraw Hill, 2007.		

UNIT-I (a)

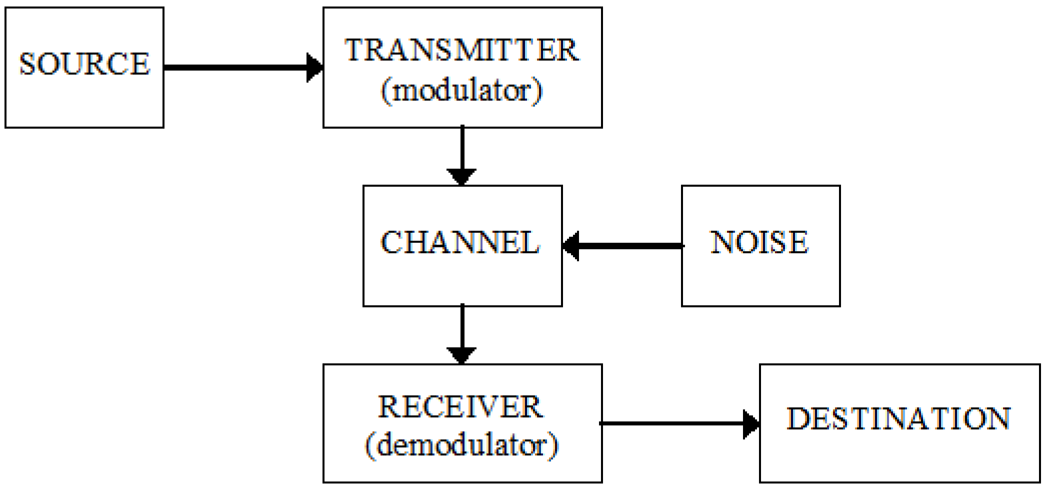
AMPLITUDE MODULATION

Introduction to Communication System

Communication is the process by which information is exchanged between individuals through a medium.

Communication can also be defined as the transfer of information from one point in spaceand time to another point.

The basic block diagram of a communication system is as follows.



- **Transmitter:** Couples the message into the channel using high frequency signals.
- **Channel:** The medium used for transmission of signals
- **Modulation:** It is the process of shifting the frequency spectrum of a signal to a frequency range in which more efficient transmission can be achieved.
- **Receiver:** Restores the signal to its original form.
- **Demodulation:** It is the process of shifting the frequency spectrum back to the original baseband frequency range and reconstructing the original form.

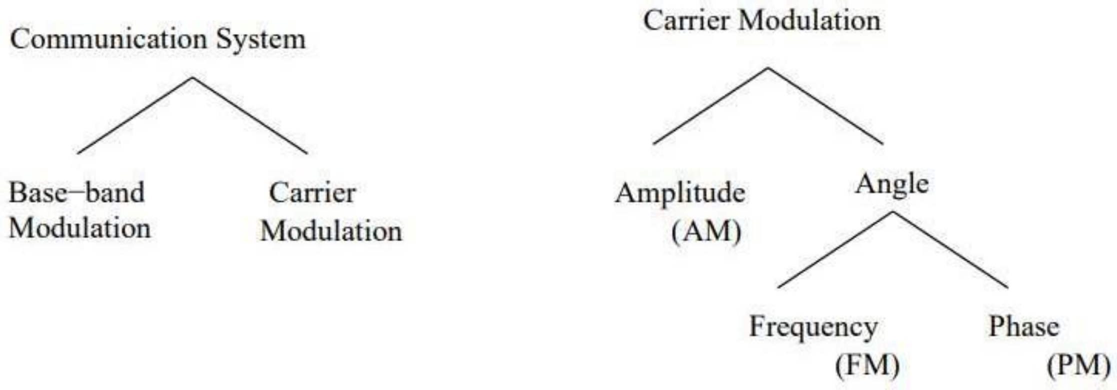
Modulation:

Modulation is a process that causes a shift in the range of frequencies in a signal.

• Signals that occupy the same range of frequencies can be separated.

• Modulation helps in noise immunity, attenuation - depends on the physical medium.

The below figure shows the different kinds of analog modulation schemes that are available



Modulation is operation performed at the transmitter to achieve efficient and reliable information transmission.

For analog modulation, it is frequency translation method caused by changing the appropriate quantity in a carrier signal.

It involves two waveforms:

- A modulating signal/baseband signal – represents the message.
- A carrier signal – depends on type of modulation.

•Once this information is received, the low frequency information must be removed from the high frequency carrier. •This process is known as "Demodulation".

Need for Modulation:

- Baseband signals are incompatible for direct transmission over the medium so, modulation is used to convey (baseband) signals from one place to another.
- Allows frequency translation:
  - Frequency Multiplexing
  - Reduce the antenna height
  - Avoids mixing of signals
  - Narrowbanding
- Efficient transmission
- Reduced noise and interference

Types of Modulation:

Three main types of modulations:

Analog Modulation

- **Amplitude modulation**
  Example: Double sideband with carrier (DSB-WC), Double- sideband suppressed carrier (DSB-SC), Single sideband suppressed carrier (SSB-SC), vestigial sideband (VSB)

- **Angle modulation (frequency modulationλ & phase modulation)**

  Example: Narrow band frequency modulation (NBFM), Widebandλ frequency modulation (WBFM), Narrowband phase modulation (NBPM), Wideband phase modulation (NBPM)

Pulse Modulation

- Carrier is a train of pulses
- Example: Pulse Amplitude Modulation (PAM), Pulse width modulation (PWM) , Pulse Position Modulation (PPM)

Digital Modulation

- Modulating signal is analog
  - Example: Pulse Code Modulation (PCM), Delta Modulationλ (DM), Adaptive Delta Modulation (ADM), Differential Pulse Code Modulation (DPCM), Adaptive Differential Pulse Code Modulation (ADPCM) etc.
- Modulating signal is digital (binary modulation)
  - Example: Amplitude shift keying (ASK), frequency Shift Keyingλ (FSK), Phase Shift Keying (PSK) etc

Amplitude Modulation (AM)

Amplitude Modulation is the process of changing the amplitude of a relatively high frequency carrier signal in accordance with the amplitude of the modulating signal (Information).

The carrier amplitude varied linearly by the modulating signal which usually consists of a range of audio frequencies. The frequency of the carrier is not affected.

- Application of AM - Radio broadcasting, TV pictures (video), facsimile transmission
- Frequency range for AM - 535 kHz – 1600 kHz
- Bandwidth - 10 kHz

Various forms of Amplitude Modulation

• Conventional Amplitude Modulation (Alternatively known as Full AM or Double Sideband Large carrier modulation (DSBLC) /Double Sideband Full Carrier (DSBFC)

• Double Sideband Suppressed carrier (DSBSC) modulation

• Single Sideband (SSB) modulation

• Vestigial Sideband (VSB) modulation

Time Domain and Frequency Domain Description

It is the process where, the amplitude of the carrier is varied proportional to that of the message signal.

Let m (t) be the base-band signal, m (t) ←→ M (ω) and c (t) be the carrier, $c(t) = A_c \cos(\omega_c t)$. fc is chosen such that fc >> W, where W is the maximum frequency component of m(t). The amplitude modulated signal is given by

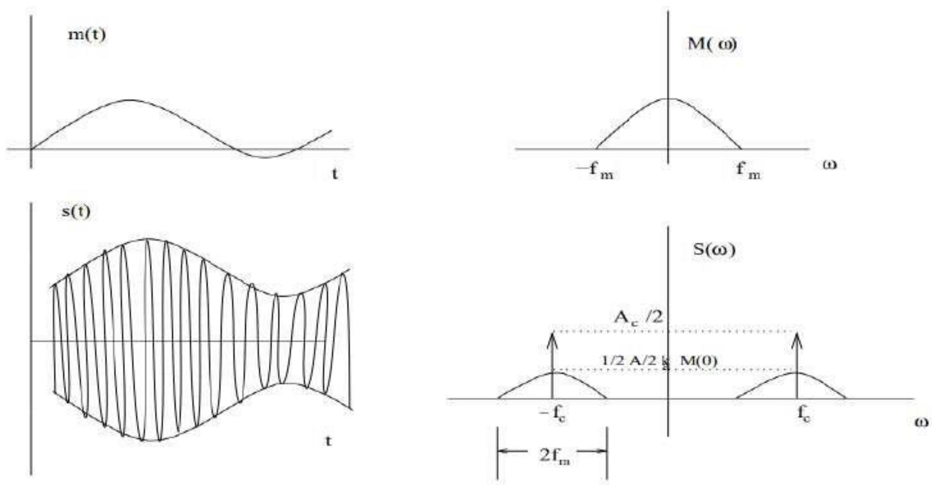
$$s(t) = Ac\,[1 + k_a m(t)]\cos(\omega ct)$$

Fourier Transform on both sides of the above equation

$$S(\omega) = \pi\, Ac/2\, (\delta(\omega - \omega c) + \delta(\omega + \omega c)) + k_a Ac/2\, (M(\omega - \omega c) + M(\omega + \omega c))$$

k_a is a constant called amplitude sensitivity.

$k_a m(t) < 1$ and it indicates percentage modulation.



Amplitude modulation in time and frequency domain

Single Tone Modulation:

Consider a modulating wave m(t) that consists of a single tone or single frequency component given by

$$m(t) = A_m \cos(2\pi f_m t) \qquad \text{......................(1)}$$

where A_m is peak amplitude of the sinusoidal modulating wave

f_m is the frequency of the sinusoidal modulating wave

Let A_c be the peak amplitude and f_c be the frequency of the high frequency carrier signal. Then the corresponding single-tone AM wave is given by

$$s(t) = A_c[1 + m\cos(2\pi f_m t)]Cos(2\pi f_c t) \qquad \text{........................(2)}$$

Let $A_{\max}$ and $A_{\min}$ denote the maximum and minimum values of the envelope of the modulated wave. Then from the above equation (2.12), we get

$$\frac{A_{\max}}{A_{\min}} = \frac{A_c(1+m)}{A_c(1-m)}$$

$$m = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}}$$

Expanding the equation (2), we get

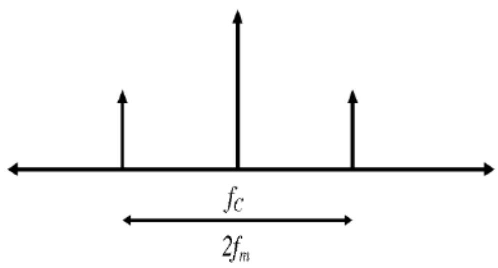
$$s(t) = A_c \cos(2\pi f_c t) + \frac{1}{2} m A_c \cos[2\pi(f_c + f_m)t] + \frac{1}{2} m A_c \cos[2\pi(f_c - f_m)t]$$

The Fourier transform of $s(t)$ is obtained as follows.

$$s(f) = \frac{1}{2} A_c[\delta(f - f_c) + \delta(f + f_c)] + \frac{1}{4} m A_c[\delta(f - f_c - f_m) + \delta(f + f_c + f_m)]$$

$$+ \frac{1}{4} m A_c[\delta(f - f_c + f_m) + \delta(f + f_c - f_m)]$$

Thus the spectrum of an AM wave, for the special case of sinusoidal modulation consists of delta functions at $\pm f_c$, $f_c \pm f_m$, and $-f_c \pm f_m$. The spectrum for positive frequencies is as shown in figure



Frequency Domain characteristics of single tone AM

Power relations in AM waves:

Consider the expression for single tone/sinusoidal AM wave

$$s(t) = A_c Cos(2\pi f_c t) + \frac{1}{2} m A_c Cos[2\pi(f_c + f_m)t] + \frac{1}{2} m A_c Cos[2\pi(f_c - f_m)t] \quad \text{..............(1)}$$

This expression contains three components. They are carrier component, upper side band and lower side band. Therefore Average power of the AM wave is sum of these three components.

Therefore the total power in the amplitude modulated wave is given by

$$Pt = \frac{V^2{}_{car}}{R} + \frac{V^2{}_{LSB}}{R} + \frac{V^2{}_{USB}}{R} \quad \text{......................(2)}$$

Where all the voltages are rms values and R is the resistance, in which the power is dissipated.

$$P_C = \frac{V^2{}_{car}}{R} = \frac{\left(A_c / \sqrt{2}\right)^2}{R} = \frac{A_c{}^2}{2R}$$

$$P_{LSB} = \frac{V_{LSB}{}^2}{R} = \left(\frac{mA_c}{2\sqrt{2}}\right)^2 \frac{1}{R} = \frac{m^2 A_c{}^2}{8R} = \frac{m^2}{4} P_c$$

$$P_{USB} = \frac{V_{USB}{}^2}{R} = \left(\frac{mA_c}{2\sqrt{2}}\right)^2 \frac{1}{R} = \frac{m^2 A_c{}^2}{8R} = \frac{m^2}{4} P_c$$

Therefore total average power is given by

$$P_t = P_c + P_{LSB} + P_{USB}$$

$$P_t = P_c + \frac{m^2}{4} P_c + \frac{m^2}{4} P_c$$

$$P_t = P_c \left(1 + \frac{m^2}{4} + \frac{m^2}{4}\right)$$

$$P_t = P_c \left(1 + \frac{m^2}{2}\right) \quad \text{............................(3)}$$

The ratio of total side band power to the total power in the modulated wave is given by

$$\frac{P_{SB}}{P_t} = \frac{P_c(m^2/2)}{P_c(1+m^2/2)}$$

$$\frac{P_{SB}}{P_t} = \frac{m^2}{2+m^2} \qquad \text{.........................(4)}$$

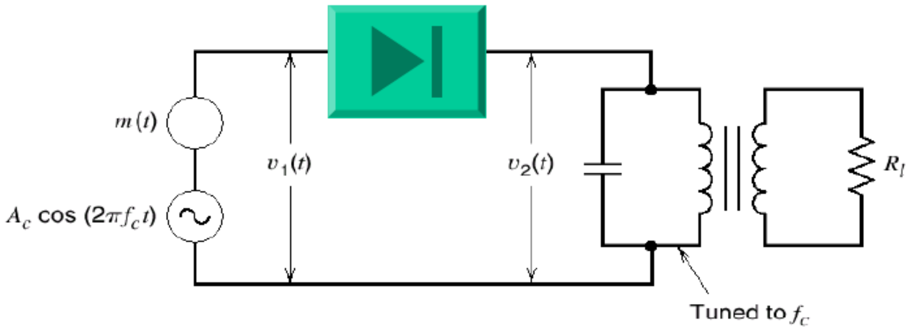
This ratio is called the efficiency of AM system

Generation of AM waves:

Two basic amplitude modulation principles are discussed. They are square law modulation and switching modulator.

Switching Modulator

Consider a semiconductor diode used as an ideal switch to which the carrier signal $c(t) = A_c \cos(2\pi f_c t)$ and information signal $m(t)$ are applied simultaneously as shown figure



Switching Modulator

The total input for the diode at any instant is given by

$$v_1 = c(t) + m(t)$$

$$v_1 = A_c \cos 2\pi f_c t + m(t)$$

When the peak amplitude of c(t) is maintained more than that of information signal, the operation is assumed to be dependent on only c(t) irrespective of m(t).

When c(t) is positive, v2=v1since the diode is forward biased. Similarly, when c(t) is negative, v2=0 since diode is reverse biased. Based upon above operation, switching response of the diode is periodic rectangular wave with an amplitude unity and is given by

$$p(t) = \frac{1}{2} + \frac{1}{\pi}\sum_{n=-\infty}^{\infty}\frac{(-1)^{n-1}}{2n-1}\cos(2\pi f_c t(2n-1))$$

$$p(t) = \frac{1}{2} + \underset{n=0,1}{\frac{2}{\pi}}\cos(2\pi f_c t) - \underset{n=-1,+2}{\frac{2}{3\pi}}\cos(6\pi f_c t) + -$$

Therefore the diode response V_o is a product of switching response *p(t)* and input v_1.

$$v_2 = v_1 * p(t)$$

$$V_2 = [A_c \cos 2\pi f_c t + m(t)]\left[\frac{1}{2} + \frac{2}{\pi}\cos 2\pi f_c t - \frac{2}{3\pi}\cos 6\pi f_c t + - + -\right]$$

Applying the Fourier Transform, we get

$$V_2(f) = \frac{A_c}{4}[\delta(f - f_c) + \delta(f + f_c)] + \frac{M(f)}{2} + \frac{A_c}{\pi}\delta(f)$$
$$+ \frac{A_c}{2\pi}[\delta(f - 2f_c) + \delta(f + 2f_c)] + \frac{1}{\pi}[M(f - f_c) + M(f + f_c)]$$
$$- \frac{A_c}{6\pi}[\delta(f - 4f_c) + \delta(f + 4f_c)] - \frac{A_c}{3\pi}[\delta(f - 2f_c) + \delta(f + 2f_c)]$$
$$- \frac{1}{3\pi}[M(f - 3f_c) + M(f + f_c)]$$

The diode output v_2 consists of

a dc component at $f = 0$.

Information signal ranging from 0 to *w* Hz and infinite number of frequency bands centered at f, $2f_c$, $3f_c$, $4f_c$, ---------

The required AM signal centred at fc can be separated using band pass filter. The lower cut off-frequency for the band pass filter should be between w and fc-w and the upper cut-off frequency between fc+w and 2fc. The filter output is given by the equation

$$S(t) = \frac{A_c}{2}\left[1 + \frac{4}{\pi}\frac{m(t)}{A_c}\right]\cos 2\pi f_c t$$

For a single tone information, let $m(t) = A_m \cos(2\pi f_m t)$

$$S(t) = \frac{A_c}{2}\left[1 + \frac{4}{\pi}\frac{A_m}{A_c}\cos 2\pi f_m t\right]\cos 2\pi f_c t$$

Therefore modulation index, $m = \frac{4}{\pi}\frac{A_m}{A_c}$

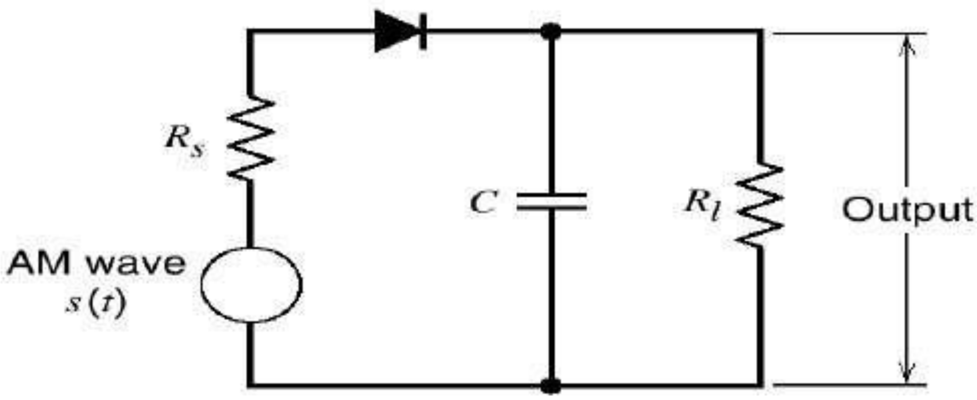
The output AM signal is free from distortions and attenuations only when $fc\text{-}w > w$ or $fc > 2w$.

Detection of AM waves

Demodulation is the process of recovering the information signal (base band) from the incoming modulated signal at the receiver. There are two methods, they are Square law Detector and Envelope Detector

Envelope Detector

It is a simple and highly effective system. This method is used in most of the commercial AM radio receivers. An envelope detector is as shown below.



Envelope Detector

During the positive half cycles of the input signals, the diode D is forward biased and the capacitor C charges up rapidly to the peak of the input signal. When the input signal falls below this value, the diode becomes reverse biased and the capacitor C discharges through the load resistor RL.

The discharge process continues until the next positive half cycle. When the input signal becomes greater than the voltage across the capacitor, the diode conducts again and the process is repeated.

The charge time constant (rf+Rs)C must be short compared with the carrier period, the capacitor charges rapidly and there by follows the applied voltage up to the positive peak when the diode is conducting.That is the charging time constant shall satisfy the condition,

$$(r_f + R_s)C << \frac{1}{f_c}$$

On the other hand, the discharging time-constant $R_L C$ must be long enough to ensure that the capacitor discharges slowly through the load resistor R_L between the positive peaks of the carrier wave, but not so long that the capacitor voltage will not discharge at the maximum rate of change of the modulating wave.

That is the discharge time constant shall satisfy the condition,

$$\frac{1}{f_c} << R_L C << \frac{1}{W}$$

Where 'W' is band width of the message signal. The result is that the capacitor voltage or detector output is nearly the same as the envelope of AM wave.

Advantages and Disadvantages of AM:

Advantages of AM:

- Generation and demodulation of AM wave are easy.
- AM systems are cost effective and easy to build.

Disadvantages:

- AM contains unwanted carrier component, hence it requires more transmission power.
- The transmission bandwidth is equal to twice the message bandwidth.

To overcome these limitations, the conventional AM system is modified at the cost of increased system complexity. Therefore, three types of modified AM systems are discussed.

DSBSC (Double Side Band Suppressed Carrier) modulation:

In DSBC modulation, the modulated wave consists of only the upper and lower side bands. Transmitted power is saved through the suppression of the carrier wave, but the channel bandwidth requirement is the same as before.

SSBSC (Single Side Band Suppressed Carrier) modulation: The SSBSC modulated wave consists of only the upper side band or lower side band. SSBSC is suited for transmission of voice signals. It is an optimum form of modulation in that it requires the minimum transmission power and minimum channel band width. Disadvantage is increased cost and complexity.

VSB (Vestigial Side Band) modulation: In VSB, one side band is completely passed and just a trace or vestige of the other side band is retained. The required channel bandwidth is therefore in excess of the message bandwidth by an amount equal to the width of the vestigial side band. This method is suitable for the transmission of wide band signals.

DSB-SC MODULATION

DSB-SC Time domain and Frequency domain Description:

DSBSC modulators make use of the multiplying action in which the modulating signal multiplies the carrier wave. In this system, the carrier component is eliminated and both upper and lower side bands are transmitted. As the carrier component is suppressed, the power required for transmission is less than that of AM.

If $m(t)$ is the message signal and $c(t) = A_c \cos(2\pi f_c t)$ is the carrier signal, then DSBSC modulated wave $s(t)$ is given by

$$s(t) = c(t)\, m(t)$$

$$s(t) = A_c \cos(2\pi f_c t) m(t)$$

Consequently, the modulated signal s(t) under goes a phase reversal , whenever the message signal m(t) crosses zero as shown below.

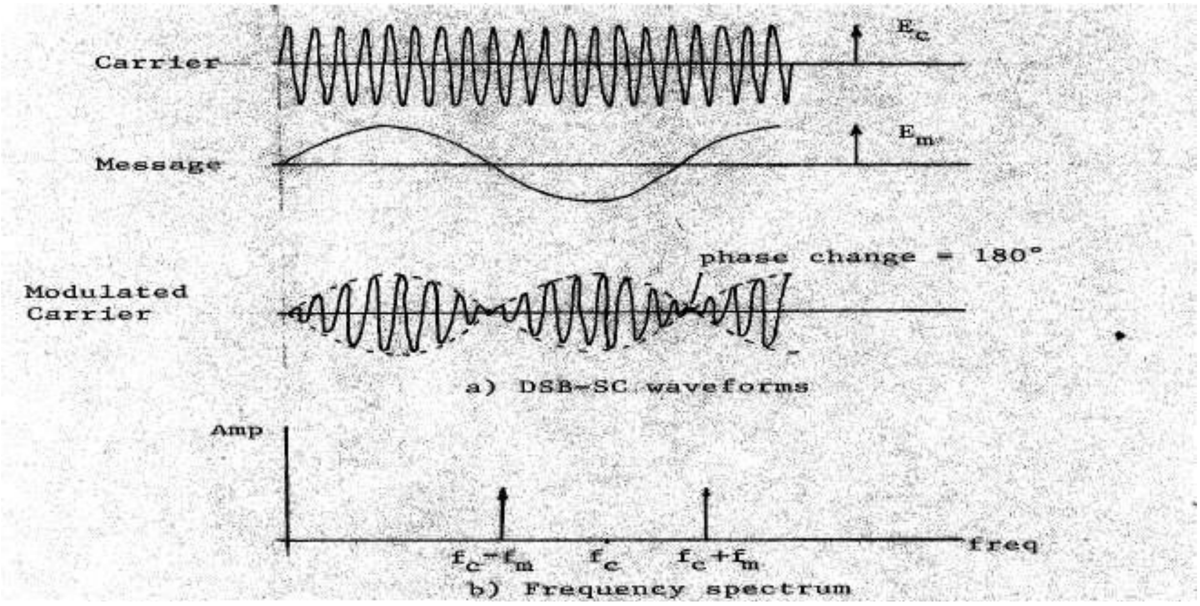


Fig.1. (a) DSB-SC waveform (b) DSB-SC Frequency Spectrum

The envelope of a DSBSC modulated signal is therefore different from the message signal and the Fourier transform of s(t) is given by

$$S(f) = \frac{A_c}{2}\left[M\left(f - f_c\right) + M\left(f + f_c\right)\right]$$

For the case when base band signal *m(t)* is limited to the interval $-W<f<W$ as shown in figure below, we find that the spectrum *S(f)* of the DSBSC wave *s(t)* is as illustrated below. Except for a change in scaling factor, the modulation process simply translates the spectrum of the base band signal by f_c. The transmission bandwidth required by DSBSC modulation is the same as that for AM.

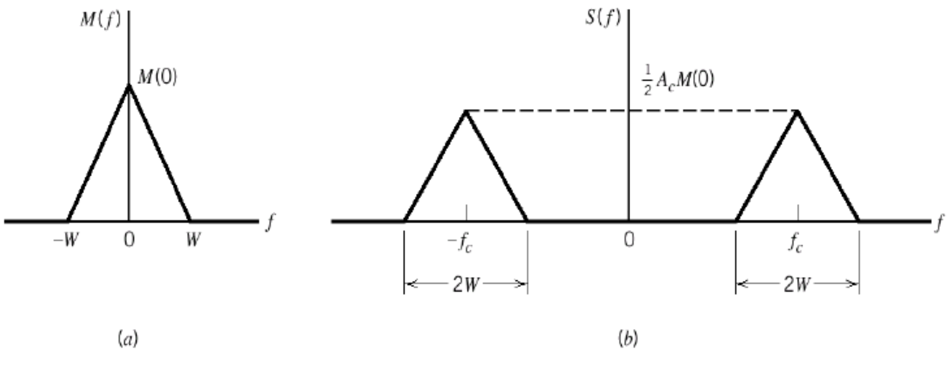


Figure: Message and the corresponding DSBSC spectrum

Generation of DSBSC Waves:

Balanced Modulator (Product Modulator)

A balanced modulator consists of two standard amplitude modulators arranged in a balanced configuration so as to suppress the carrier wave as shown in the following block diagram. It is assumed that the AM modulators are identical, except for the sign reversal of the modulating wave applied to the input of one of them. Thus, the output of the two modulators may be expressed as,

$$s_1(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$$
$$s_2(t) = A_c[1 - k_a m(t)]\cos(2\pi f_c t)$$

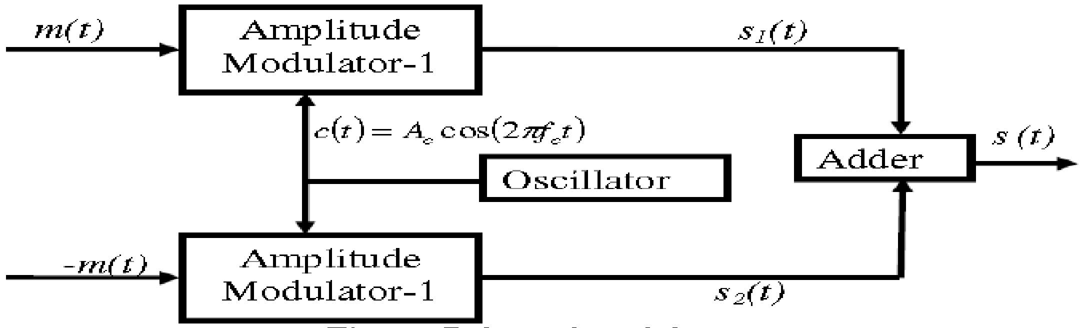


Fig 3 : Balanced modulator

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

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

Hence, except for the scaling factor 2ka, the balanced modulator output is equal to the product of the modulating wave and the carrier.

Ring Modulator

Ring modulator is the most widely used product modulator for generating DSBSC wave and is shown below.

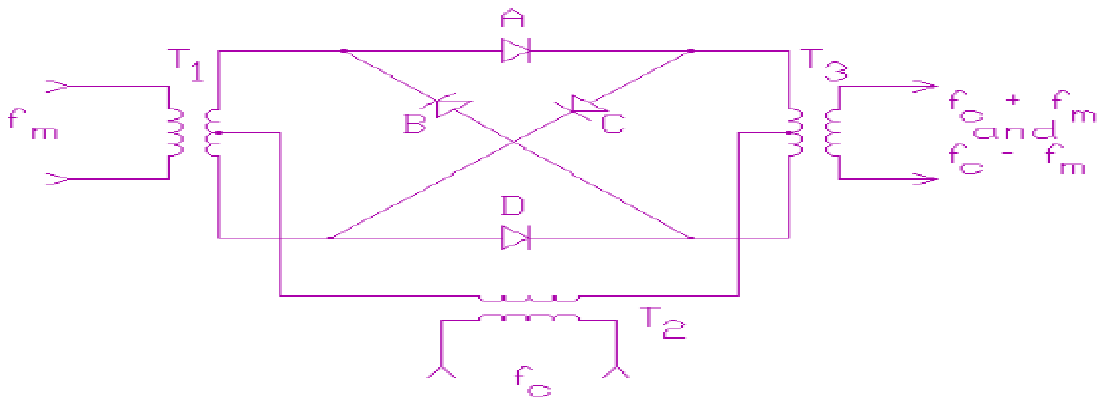


Fig.4 : Ring modulator

The four diodes form a ring in which they all point in the same direction. The diodes are controlled by square wave carrier c(t) of frequency fc, which is applied longitudinally by means of two center-tapped transformers. Assuming the diodes are ideal, when the carrier is positive, the outer diodes D1 and D2 are forward biased where as the inner diodes D3 and D4 are reverse biased, so that the modulator multiplies the base band signal m(t) by c(t). When the carrier is negative, the diodes D1 and D2 are reverse biased and D3 and D4 are forward, and the modulator multiplies the base band signal –m(t) by c(t).

Thus the ring modulator in its ideal form is a product modulator for square wave carrier and the base band signal m(t). The square wave carrier can be expanded using Fourier series as

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

Therefore the ring modulator out put is given by

$$s(t) = m(t)c(t)$$

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

From the above equation it is clear that output from the modulator consists entirely of modulation products. If the message signal m(t) is band limited to the frequency band $-w < f < w$, the output spectrum consists of side bands centred at fc.

Detection of DSB-SC waves:

Coherent Detection:

The message signal m(t) can be uniquely recovered from a DSBSC wave s(t) by first multiplying s(t) with a locally generated sinusoidal wave and then low pass filtering the product as shown.

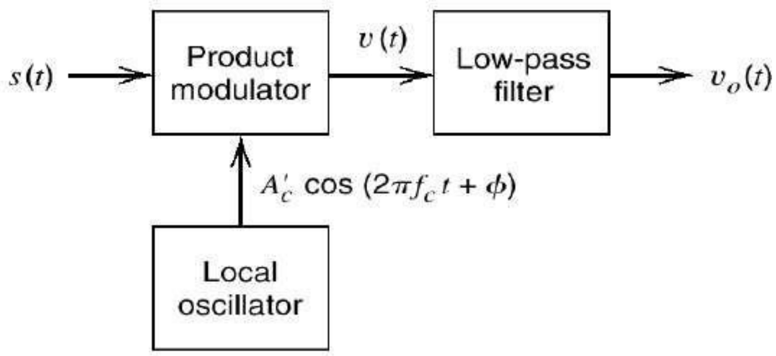


Fig.5 : Coherent detector

It is assumed that the local oscillator signal is exactly coherent or synchronized, in both frequency and phase, with the carrier wave c(t) used in the product modulator to generate s(t). This method of demodulation is known as coherent detection or synchronous detection.

Let $A_c^1 \cos(2\pi f_c t + \phi)$ be the local oscillator signal, and $s(t) = A_c \cos(2\pi f_c t)m(t)$ be the DSBSC wave. Then the product modulator output $v(t)$ is given by

$$v(t) = A_c A_c^1 \cos(2\pi f_c t)\cos(2\pi f_c t + \phi)m(t)$$

$$v(t) = \frac{A_c A_c^1}{4}\cos(4\pi f_c t + \phi)m(t) + \frac{A_c A_c^1}{2}\cos(\phi)m(t)$$

The first term in the above expression represents a DSBSC modulated signal with a carrier frequency $2f_c$, and the second term represents the scaled version of message signal. Assuming that the message signal is band limited to the interval $-w < f < w$, the spectrum of $v(t)$ is plotted as shown below.

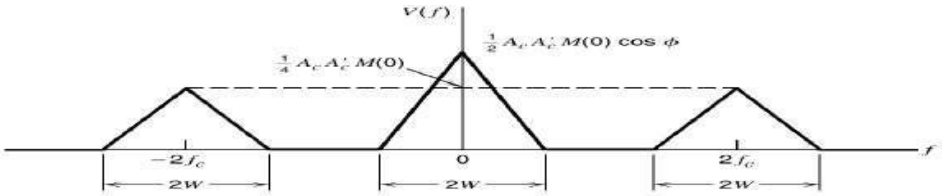


Fig.6.Spectrum of output of the product modulator

From the spectrum, it is clear that the unwanted component (first term in the expression) can be removed by the low-pass filter, provided that the cut-off frequency of the filter is greater than W but less than 2fc-W. The filter output is given by

$$v_o(t) = \frac{A_c A_c^1}{2}\cos(\phi)m(t)$$

The demodulated signal $v_o(t)$ is therefore proportional to m(t) when the phase error ϕ is constant.

Costas Receiver (Costas Loop):

Costas receiver is a synchronous receiver system, suitable for demodulating DSBSC waves. It consists of two coherent detectors supplied with the same input signal,

that is the incoming DSBSC wave $s(t) = A_c \cos(2\pi f_c t)m(t)$ but with individual local oscillator signals that are in phase quadrature with respect to each other as shown below.

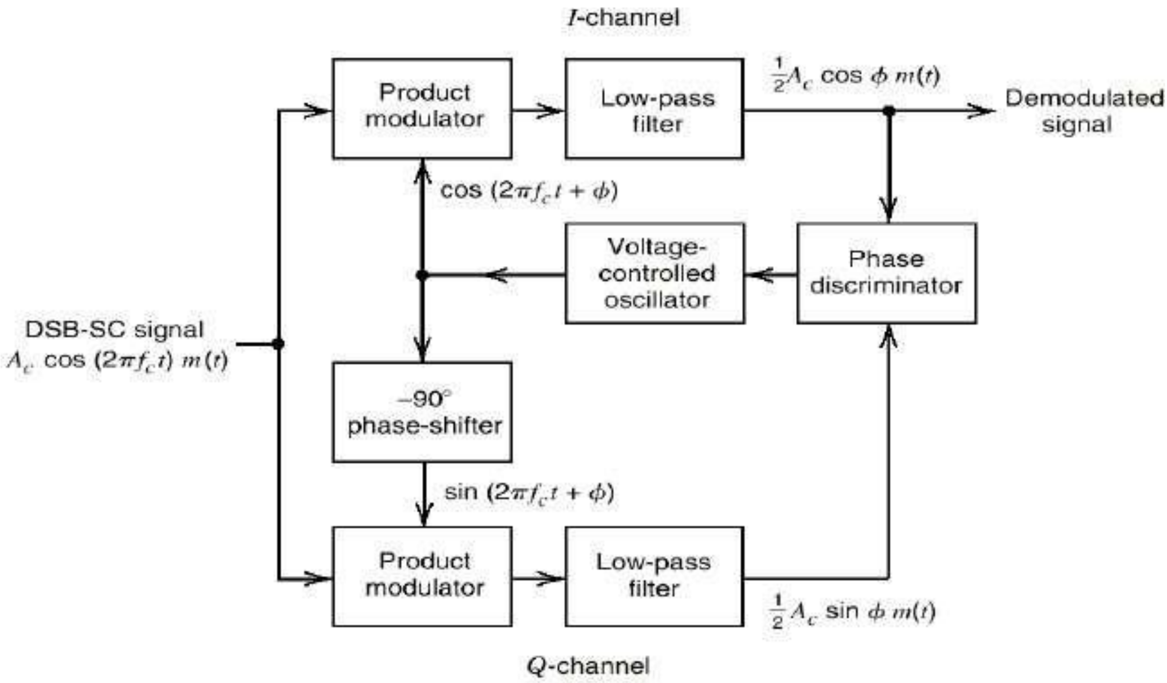


Fig.7. Costas Receiver

The frequency of the local oscillator is adjusted to be the same as the carrier frequency fc. The detector in the upper path is referred to as the in-phase coherent detector or I-channel, and that in the lower path is referred to as the quadrature-phase coherent detector or Q-channel.

These two detector are coupled together to form a negative feedback system designed in such a way as to maintain the local oscillator synchronous with the carrier wave. Suppose the local oscillator signal is of the same phase as the carrier $c(t) = A_c\cos(2\pi f_c t)$ wave used to generate the incoming DSBSC wave. Then we find that the I-channel output contains the desired demodulated signal m(t), where as the Q-channel output is zero due to quadrature null effect of the Q-channel. Suppose that the local oscillator phase drifts from its proper value by a small angle ϕ radians. The I-channel output will remain essentially unchanged, but there will be some signal appearing at the Q-channel output, which is proportional to $\sin(\phi) \approx \phi$ for small ϕ.

This Q-channel output will have same polarity as the I-channel output for one direction of local oscillator phase drift and opposite polarity for the opposite direction of local oscillator phase drift. Thus by combining the I-channel and Q-channel outputs in a phase discriminator (which consists of a multiplier followed by a LPF), a dc control signal is obtained that automatically corrects for the local phase errors in the voltage-controlled oscillator.

Introduction of SSB-SC

Standard AM and DSBSC require transmission bandwidth equal to twice the message bandwidth. In both the cases spectrum contains two side bands of width W Hz, each. But the upper and lower sides are uniquely related to each other by the virtue of their symmetry about the carrier frequency. That is, given the amplitude and phase spectra of either side band, the other can be uniquely determined. Thus if only one side band is transmitted, and if both the carrier and the other side band are suppressed at the transmitter, no information is lost. This kind of modulation is called SSBSC and spectral comparison between DSBSC and SSBSC is shown in the figures 1 and 2.

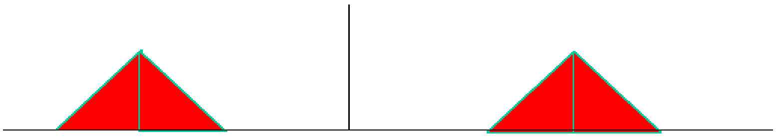


Figure.1 : Spectrum of the DSBSC wave

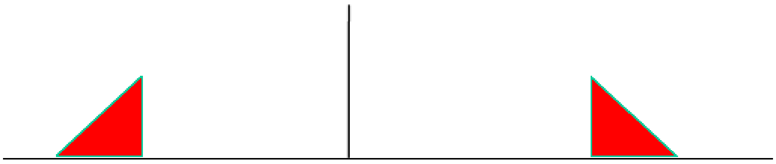


Figure .2 : Spectrum of the SSBSC wave

Frequency Domain Description

Consider a message signal $m(t)$ with a spectrum $M(f)$ band limited to the interval $-w < f < w$ as shown in figure 3 , the DSBSC wave obtained by multiplexing $m(t)$ by the carrier wave $c(t) = A_c \cos(2\pi f_c t)$ and is also shown, in figure 4 . The upper side band is represented in duplicate by the frequencies above f_c and those below $-f_c$, and when only upper

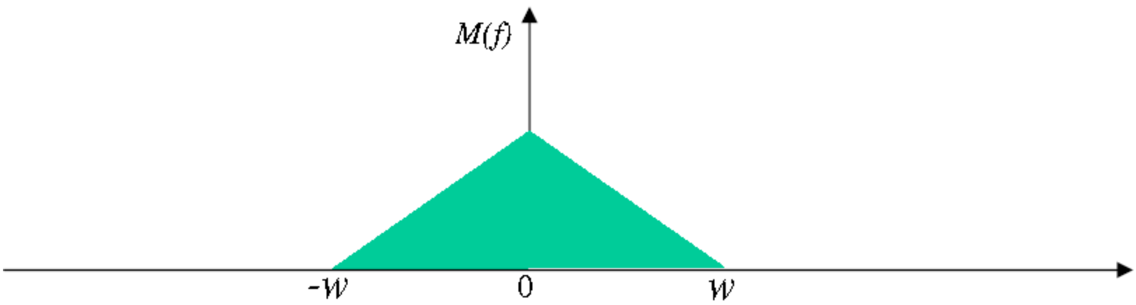


Figure 3. : Spectrum of message wave

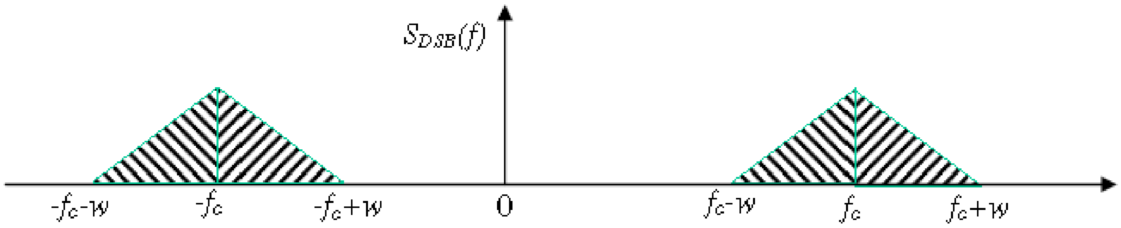


Figure .4 : Spectrum of DSBSC wave

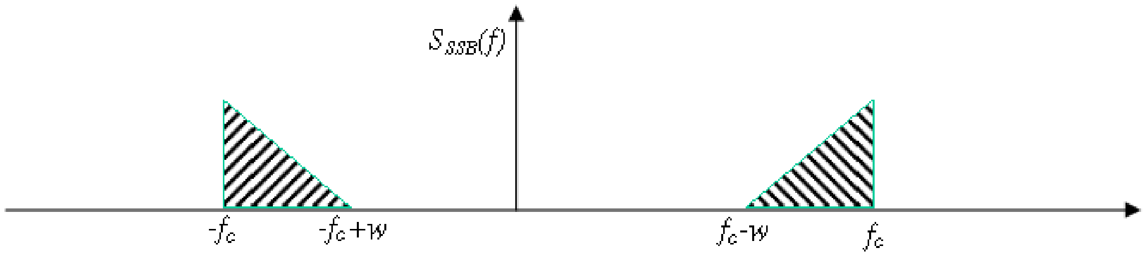


Figure.5 : Spectrum of SSBSC-LSB wave

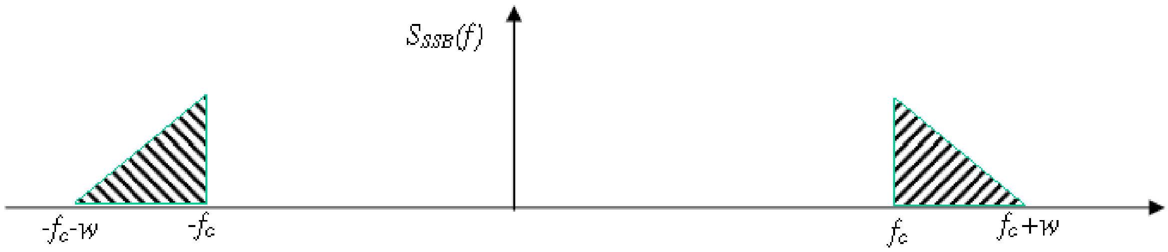


Figure .6 : Spectrum of SSBSC-USB wave

side band is transmitted; the resulting SSB modulated wave has the spectrum shown in figure 6. Similarly, the lower side band is represented in duplicate by the frequencies below fc and those above -fc and when only the lower side band is transmitted, the spectrum of the corresponding SSB modulated wave shown in figure 5.Thus the essential function of the SSB modulation is to translate the spectrum of the modulating wave, either with or without inversion, to a new location in the frequency domain. The advantage of SSB modulation is reduced bandwidth and the elimination of high power carrier wave. The main disadvantage is the cost and complexity of its implementation.

Generation of SSB wave:

Frequency discrimination method

Consider the generation of SSB modulated signal containing the upper side band only. From a practical point of view, the most severe requirement of SSB generation arises from the unwanted sideband, the nearest component of which is separated from the desired side band by twice the lowest frequency component of the message signal. It implies that, for the generation of an SSB wave to be possible, the message spectrum must have an energy gap centered at the origin as shown in figure 7. This requirement is naturally satisfied by voice signals, whose energy gap is about 600Hz wide.

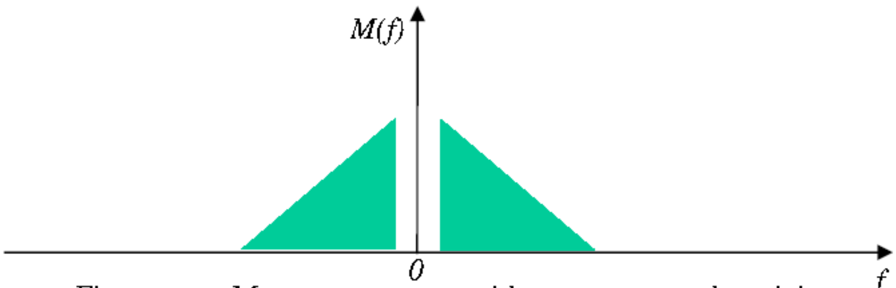


Figure .7 : Message spectrum with energy gap at the origin

The frequency discrimination or filter method of SSB generation consists of a product modulator, which produces DSBSC signal and a band-pass filter to extract the desired side band and reject the other and is shown in the figure 8.

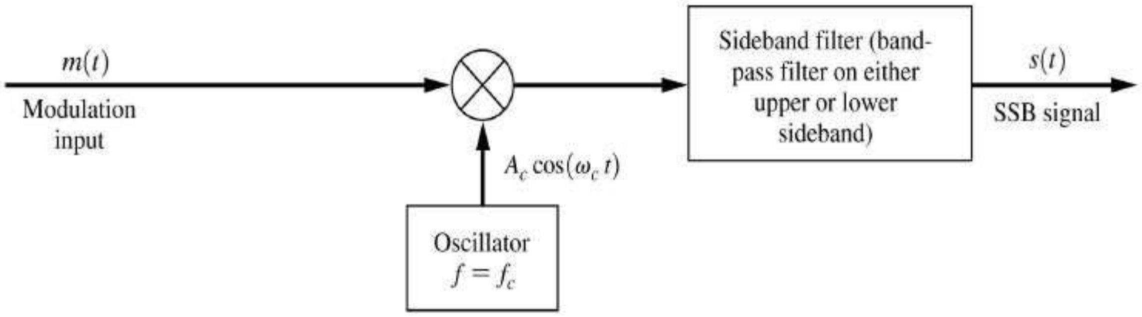


Figure .8 : Frequency discriminator to generate SSBSC wave

Application of this method requires that the message signal satisfies two conditions:
1. The message signal m(t) has no low-frequency content. Example: speech, audio, music.
2. The highest frequency component W 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.

In designing the band pass filter, the following requirements should be satisfied:
1.The pass band of the filter occupies the same frequency range as the spectrum of the desired SSB modulated wave.

2. The width of the guard band of the filter, separating the pass band from the stop band, where the unwanted sideband of the filter input lies, is twice the lowest frequency component of the message signal.

When it is necessary to generate an SSB modulated wave occupying a frequency band that is much higher than that of the message signal, it becomes very difficult to design an appropriate filter that will pass the desired side band and reject the other. In such a situation it is necessary to resort to a multiple-modulation process so as to ease the filtering requirement. This approach is illustrated in the following figure 9 involving two stages of modulation.

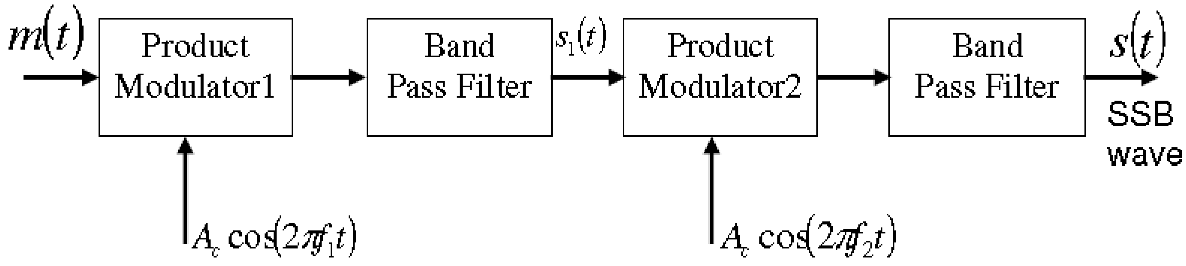


Figure .9 : Two stage frequency discriminator

The SSB modulated wave at the first filter output is used as the modulating wave for the second product modulator, which produces a DSBSC modulated wave with a spectrum that is symmetrically spaced about the second carrier frequency f2. The frequency separation between the side bands of this DSBSC modulated wave is effectively twice the first carrier frequency f1, thereby permitting the second filter to remove the unwanted side band.

Time Domain Description:

The time domain description of an SSB wave s(t) in the canonical form is given by the equation 1.

$$s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t) \quad \text{-------------------- (1)}$$

where $S_I(t)$ is the in-phase component of the SSB wave and $S_Q(t)$ is its quadrature component. The in-phase component $S_I(t)$ except for a scaling factor, may be derived from $S(t)$ by first multiplying $S(t)$ by $\cos(2\pi f_c t)$ and then passing the product through a low-pass filter. Similarly, the quadrature component $S_Q(t)$, except for a scaling factor, may be derived from $s(t)$ by first multiplying $s(t)$ by $\sin(2\pi f_c t)$ and then passing the product through an identical filter.

The Fourier transformation of $S_I(t)$ and $S_Q(t)$ are related to that of SSB wave as follows, respectively.

$$S_I(f) = \begin{cases} S(f - f_c) + S(f + f_c), -w \le f \le w \\ 0, elsewhere \end{cases} \quad \text{------------------ (2)}$$

$$S_Q(f)=\begin{cases} j[S(f-f_c)-S(f+f_c)], -w \le f \le w \\ 0, elsewhere \end{cases} \quad \text{------------------- (3)}$$

where $-w < f < w$ defines the frequency band occupied by the message signal $m(t)$.

Consider the SSB wave that is obtained by transmitting only the upper side band, shown in figure 10 . Two frequency shifted spectra $S(f-f_c)$ and $S(f+f_c)$ are shown in figure 11 and figure 12 respectively. Therefore, from equations 2 and 3 , it follows that the corresponding spectra of the in- phase component $S_I(t)$ and the quadrature component $S_Q(t)$ are as shown in figure 13 and 14 respectively.

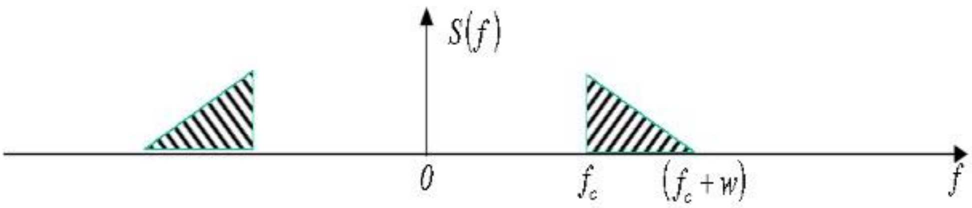


Figure 10 : Spectrum of SSBSC-USB

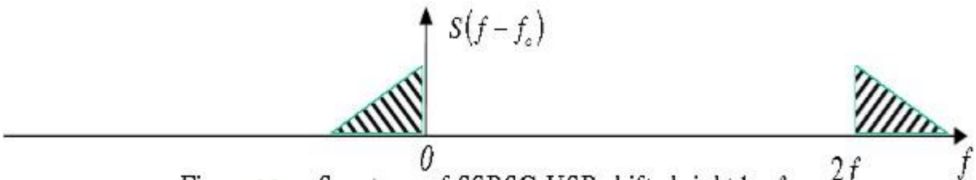


Figure 11 : Spectrum of SSBSC-USB shifted right by f_c

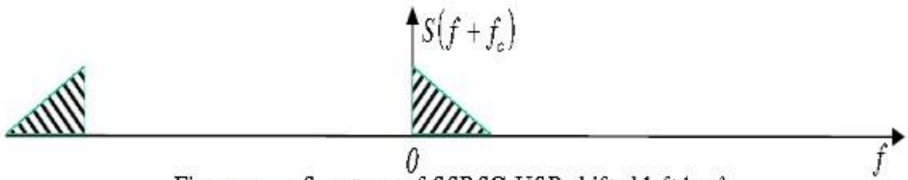


Figure 12 : Spectrum of SSBSC-USB shifted left by f_c

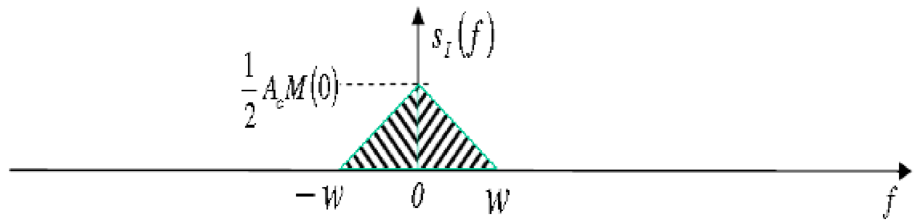


Figure 13 : Spectrum of in-phase component of SSBSC-USB

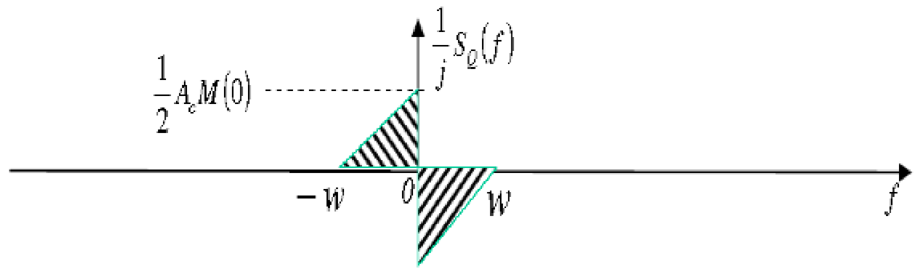


Figure 14 : Spectrum of quadrature component of SSBSC-USB

From the figure 13 , it is found that

$$S_I(f) = \frac{1}{2}A_cM(f)$$

where *M(f)* is the Fourier transform of the message signal *m(t)*. Accordingly in-phase component $S_I(t)$ is defined by equation 4

$$\boxed{s_I(t) = \frac{1}{2}A_cm(t)} \qquad \text{------------------ (4)}$$

Now on the basis of figure14 , it is found that

$$S_Q(f) = \begin{cases} \frac{-j}{2}A_cM(f), f > 0 \\ 0, f = 0 \\ \frac{j}{2}A_cM(f), f < 0 \end{cases}$$

$$S_Q(f) = \frac{-j}{2}A_c\,\mathrm{sgn}(f)M(f) \qquad \text{------------------ (5)}$$

where $\mathrm{sgn}(f)$ is the Signum function.

But from the discussions on Hilbert transforms, it is shown that

$$-j\,\mathrm{sgn}(f)M(f)=\hat{M}(f) \qquad \text{------------------ (6)}$$

where $\hat{M}(f)$ is the Fourier transform of the Hilbert transform of *m(t)*. Hence the substituting equation (6) in (5), we get

$$S_Q(f)=\frac{1}{2}A_c\hat{M}(f) \qquad \text{------------------------ (7)}$$

Therefore quadrature component $s_Q(t)$ is defined by equation 8

$$\boxed{s_Q(t)=\frac{1}{2}A_c\hat{m}(t)} \qquad \text{------------------------ (8)}$$

Therefore substituting equations (4) and (8) in equation in (1), we find that canonical representation of an SSB wave *s(t)* obtained by transmitting only the upper side band is given by the equation 9

$$\boxed{s_U(t)=\frac{1}{2}A_c m(t)\cos(2\pi f_c t)-\frac{1}{2}A_c\hat{m}(t)\sin(2\pi f_c t)} \qquad \text{------------- (9)}$$

Following the same procedure, we can find the canonical representation for an SSB wave
s(t) obtained by transmitting only the lower side band is given by

$$s_L(t)=\frac{1}{2}A_c m(t)\cos(2\pi f_c t)+\frac{1}{2}A_c\hat{m}(t)\sin(2\pi f_c t) \qquad \text{-------------- (10)}$$

Phase discrimination method for generating SSB wave:

Time domain description of SSB modulation leads to another method of SSB generation using the equations 9 or 10. The block diagram of phase discriminator is as shown in figure 15.

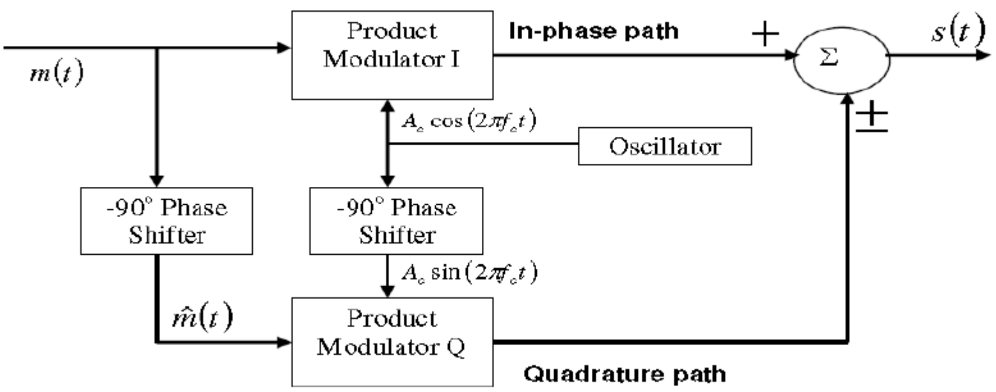


Figure 15 : Block diagram of phase discriminator

The phase discriminator consists of two product modulators I and Q, supplied with carrier waves in-phase quadrature to each other. The incoming base band signal m(t) is applied to product modulator I, producing a DSBSC modulated wave that contains reference phase sidebands symmetrically spaced about carrier frequency fc.

The Hilbert transform mˆ (t) of m (t) is applied to product modulator Q, producing a DSBSC modulated that contains side bands having identical amplitude spectra to those of modulator I, but with phase spectra such that vector addition or subtraction of the two modulator outputs results in cancellation of one set of side bands and reinforcement of the other set.

The use of a plus sign at the summing junction yields an SSB wave with only the lower side band, whereas the use of a minus sign yields an SSB wave with only the upper side band. This modulator circuit is called Hartley modulator.

Demodulation of SSB Waves:

Demodulation of SSBSC wave using coherent detection is as shown in 16 . The SSB wave *s(t)* together with a locally generated carrier $c(t) = A_c^{1}\cos(2\pi f_c t + \phi)$ is applied to a product modulator and then low-pass filtering of the modulator output yields the message signal.

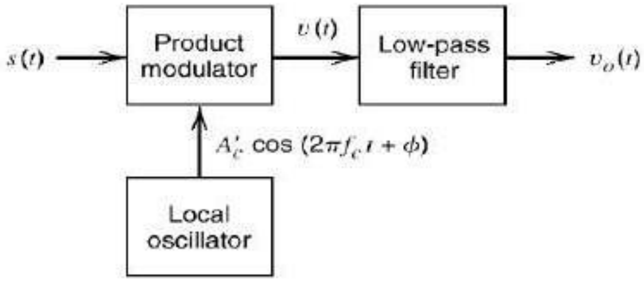


Figure 16 : Block diagram of coherent detector for SSBSC

The product modulator output *v(t)* is given by

$$v(t) = A_c^{1}\cos(2\pi f_c t + \phi)s(t) \,. \quad \text{Put } \phi = 0$$

$$v(t) = \frac{1}{2}A_c\cos(2\pi f_c t)[m(t)\cos(2\pi f_c t) \pm \hat{m}(t)\sin(2\pi f_c t)]$$

$$v(t) = \frac{1}{4}A_c m(t) + \frac{1}{4}A_c[m(t)\cos(4\pi f_c t) \pm \hat{m}(t)\sin(4\pi f_c t)] \quad \text{..............(1)}$$

The first term in the above equation 1 is desired message signal. The other term represents an SSB wave with a carrier frequency of $2f_c$ as such; it is an unwanted component, which is removed by low-pass filter.

Introduction to Vestigial Side Band Modulation

Vestigial sideband is a type of Amplitude modulation in which one side band is completely passed along with trace or tail or vestige of the other side band. VSB is a compromise between SSB and DSBSC modulation. In SSB, we send only one side band, the Bandwidth required to send SSB wave is w. SSB is not appropriate way of modulation when the message signal contains significant components at extremely low frequencies. To overcome this VSB is used.

Vestigial Side Band (VSB) modulation is another form of an amplitude-modulated signal in which a part of the unwanted sideband (called as vestige, hence the name vestigial sideband) is allowed to appear at the output of VSB transmission system.

The AM signal is passed through a sideband filter before the transmission of SSB signal. The design of sideband filter can be simplified to a greater extent if a part of the other sideband is also passed through it. However, in this process the bandwidth of VSB system is slightly increased.

Generation of VSB Modulated Signal

VSB signal is generated by first generating a DSB-SC signal and then passing it through a sideband filter which will pass the wanted sideband and a part of unwanted sideband. Thus, VSB is so called because a vestige is added to SSB spectrum.

The below figure depicts functional block diagram of generating VSB modulated signal

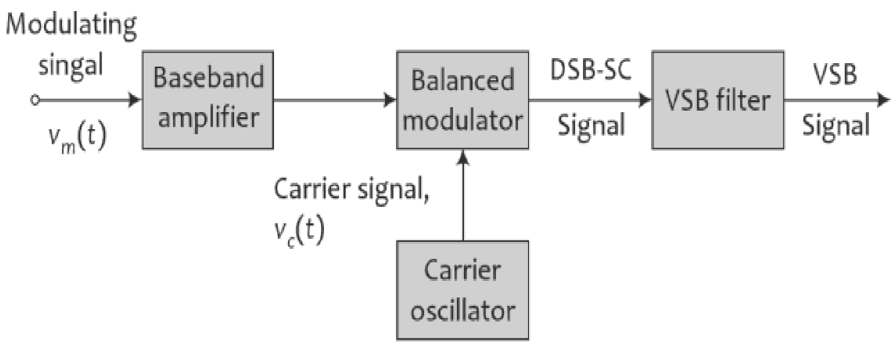


Figure: Generation of VSB Modulated Signal

A VSB-modulated signal is generated using the frequency discrimination method, in which firstly a DSB-SC modulated signal is generated and then passed through a sideband-suppression filter. This type of filter is a specially-designed bandpass filter that distinguishes VSB modulation from SSB modulation.the cutoff portion of the frequency response of this filter around the carrier frequency exhibits odd symmetry, that is, (fc-fv)≤|f|≤(fc+fv).

Accordingly the bandwidth of the VSB signal is given as

BW=(fm+fv) Hz

Where fm is the bandwidth of the modulating signal or USB, and fv is the bandwidth of vestigial sideband (VSB)

Time domain description of VSB Signal

Mathematically, the VSB modulated signal can be described in the time-domain as

$$s(t)= m(t)\ Ac\ \cos(2\pi fct) \pm m_Q(t)\ Ac\ \sin(2\pi fct)$$

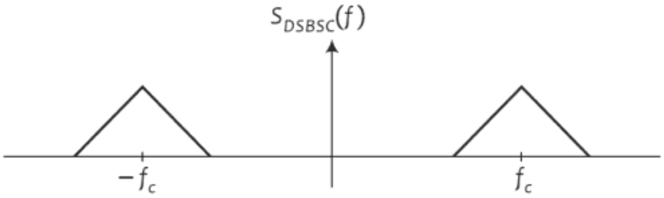
where m(t) is the modulating signal, $m_Q(t)$ is the component of m(t) obtained by passing the message signal through a vestigial filter, Ac cos(2πfct) is the carrier signal, and Ac sin(2πfct) is the 90^o phase shift version of the carrier signal.

The ± sign in the expression corresponds to the transmission of a vestige of the upper-sideband and lower-sideband respectively. The Quadrature component is required to partially reduce power in one of the sidebands of the modulated wave s(t) and retain a vestige of the other sideband as required.

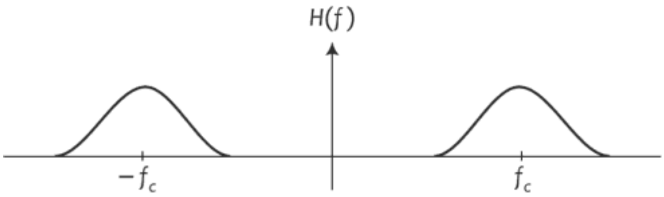
Frequency domain representation of VSB Signal

Since VSB modulated signal includes a vestige (or trace) of the second sideband, only a part of the second sideband is retained instead of completely eliminating it. Therefore, VSB signal can be generated from DSB signal followed by VSB filter which is a practical filter.

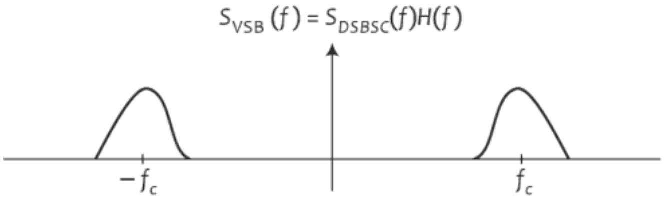
The below figure shows the DSB signal spectrum, the VSB filter characteristics, and the resulting output VSB modulated signal spectrum.



(a) DSBC Frequency-domain representation



(b) VSB Filter characteristics



(c) VSB Frequency-domain representation

Bandwidth Consideration in TV Signals

An important application of VSB modulation technique is in broadcast television. In commercial TV broadcasting system, there is a basic need to conserve bandwidth.

- The upper-sideband of the video carrier signal is transmitted upto 4MHz without any attenuation.
- The lower-sideband of the video carrier signal is transmitted without any attenuation over the range 0.75 MHz (Double side band transmission) and is entirely attenuated at 1.25MHz (single sideband transmission) and the transition is made from one o another between 0.75MHz and 1.25 MHz (thus the name vestige sideband)
- The audio signal which accompanies the video signal is transmitted by frequency modulation method using a carrier signal located 4.5 MHz above the video-carrier signal.
- The audio signal is frequency modulated on a separate carrier signal with a frequency deviation of 25 KHz. With an audio bandwidth of 10 KHz, the deviation ratio is 2.5 and an FM bandwidth of approximately 70 KHz.
- The frequency range of 100 KHz is allowed on each side of the audio-carrier signal for the audio sidebands.
- One sideband of the video-modulated signal is attenuated so that it does not interfere with the lower- sideband of the audio carrier.

Advantages of VSB Modulation

VSB transmission system has several advantages which include

- Use of simple filter design
- Less bandwidth as compared to that of DSBSC signal
- As efficient as SSB
- Possibility of transmission of low frequency components of modulating signals

Facts to Know

VSB is mainly used as a standard modulation technique for transmission of video signals in TV signals in commercial television broadcasting because the modulating video signal has large bandwidth and high speed data transmission

Envelope detection of a VSB Wave plus Carrier

To make demodulation of VSB wave possible by an envelope detector at the receiving end it is necessary to transmit a sizeable carrier together with the modulated wave. The scaled expression of VSB wave by factor k_a with the carrier component $A_c \cos(2\pi f_c t)$ can be given by

$$s(t) = A_c \cos(2\pi f_c t) + \frac{A_c}{2} k_a m(t) \cos(2\pi f_c t) - \frac{A_c}{2} k_a m_Q(t) \sin(2\pi f_c t)$$

$$= A_c \left[1 + \frac{k_a}{2} m(t)\right] \cos(2\pi f_c t) - \frac{A_c k_a}{2} m_Q(t) \sin(2\pi f_c t) \quad \text{................(1)}$$

where k_a is the modulation index; it determines the percentage modulation.

When above signal s(t) is passed through the envelope detector, the detector output is given by,

$$a(t) = A_c \left[\left(1 + \frac{k_a}{2} m(t)\right)^2 + \left(\frac{k_a}{2} m_Q(t)\right)^2\right]^{1/2}$$

$$= A_c \left[1 + \frac{k_a}{2} m(t)\right] \left(1 + \left[\frac{\frac{1}{2} k_a m_Q(t)}{1 + \frac{1}{2} k_a m(t)}\right]^2\right)^{1/2} \quad \text{................(2)}$$

The detector output is distorted by the quadrature component $m_Q(t)$ as indicated by equation (2).

Methods to reduce distortion

- Distortion can be reduced by reducing percentage modulation, k_a.
- Distortion can be reduced by reducing $m_Q(t)$ by increasing the width of the vestigial sideband.

Comparison of AM Techniques:

Sr. No.	Parameter	Standard AM	SSB	DSBSC	VSB
1	Power	High	Less	Medium	Less than DSBSC but greater than SSB
2	Bandwidth	2 f_m	f_m	2 f_m	$f_m < B_w < 2 f_m$
3	Carrier supression	No	Yes	Yes	No
4	Receiver complexity	Simple	Complex	Complex	Simple
5	Application	Radio communication	Point to point communication preferred for long distance transmission.	Point to point communication	Television broadcasting
6	Modulation type	Non linear	Linear	Linear	Linear
7	Sideband suppression	No	One sided completely	No	One sideband suppressed partly
8	Transmission efficiency	Minimum	Maximum	Moderate	Moderate

Applications of different AM systems:

- Amplitude Modulation: AM radio, Short wave radio broadcast
- DSB-SC: Data Modems, Color TV’s color signals.
- SSB: Telephone
- VSB: TV picture signals

UNIT-I (b)

ANGLE MODULATION

Introduction

There are two forms of angle modulation that may be distinguished – phase modulation and frequency modulation

Basic Definitions: Phase Modulation (PM) and Frequency Modulation (FM)

Let $\theta_i(t)$ denote the angle of modulated sinusoidal carrier, which is a function of the message. The resulting angle-modulated wave is expressed as

$$s(t) = A_c cos[\theta_i(t)] \dots\dots\dots\dots\dots (1)$$

Where A_c is the carrier amplitude. A complete oscillation occurs whenever $\theta_i(t)$ changes by 2π radians. If $\theta_i(t)$ increases monotonically with time, the average frequency in Hz, over an interval from t to $t+\Delta t$, is given by

$$f_{\Delta t}(t) = \frac{\theta_i(t+\Delta t) - \theta_i(t)}{2\pi\Delta t} \dots\dots\dots\dots\dots (2)$$

Thus the instantaneous frequency of the angle-modulated wave $s(t)$ is defined as

$$f_i(t) = \lim_{\Delta t \to 0} f_{\Delta t}(t)$$

$$f_i(t) = \lim_{\Delta t \to 0} [\frac{\theta_i(t+\Delta t) - \theta_i(t)}{2\pi\Delta t}]$$

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} \dots\dots\dots\dots\dots (3)$$

Thus, according to equation (1), the angle modulated wave s(t) is interpreted as a rotating Phasor of length Ac and angle $\theta_i(t)$. The angular velocity of such a Phasor is $d\theta_i(t)/dt$, in accordance with equ (3).In the simple case of an unmodulated carrier, the angle $\theta_i(t)$ is

$$\theta_i(t) = 2\pi f_c t + \emptyset_c$$

And the corresponding Phasor rotates with a constant angular velocity equal to $2\pi f_c$.The constant ϕ_c is the value of $\theta_i(t)$ at t=0.

There are an infinite number of ways in which the angle $\theta_i(t)$ may be varied in some manner with the baseband signal.

But the 2 commonly used methods are **Phase modulation** and **Frequency modulation.**

Phase Modulation (PM) is that form of angle modulation in which the angle $\theta_i(t)$ is varied linearly with the baseband signal m(t), as shown by

$$\boldsymbol{\theta_i(t) = 2\pi f_c t + k_p m(t)} \ldots\ldots\ldots\ldots\ldots\ldots..(\mathbf{4})$$

The term $\boldsymbol{2\pi f_c t}$ represents the angle of the unmodulated carrier, and the constant $\boldsymbol{k_p}$ represents the ***phase sensitivity*** of the modulator, expressed in radians per volt.

The phase-modulated wave *s(t)* is thus described in time domain by

$$\boldsymbol{s(t) = A_c \cos[2\pi f_c t + k_p m(t)]} \ldots\ldots\ldots\ldots\ldots(\mathbf{5})$$

Frequency Modulation (FM) is that form of angle modulation in which the instantaneous frequency $f_i(t)$ is varied linearly with the baseband signal *m(t)*, as shown by

$$\boldsymbol{f_i(t) = f_c + k_f m(t)} \ldots\ldots\ldots\ldots\ldots\ldots.(\mathbf{6})$$

The term f_c represents the frequency of the unmodulated carrier, and the constant $\boldsymbol{k_f}$ represents the ***frequency sensitivity*** of the modulator, expressed in hertz per volt.

Integrating equ.(6) with respect to time and multiplying the result by $\boldsymbol{2\pi}$, we get

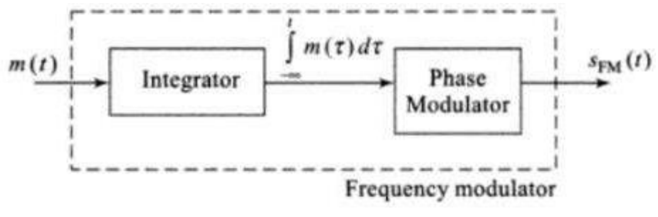
$$\boldsymbol{\theta_i(t) = 2\pi f_c t + 2\pi k_f \int_0^t m(t)\ dt} \ldots\ldots\ldots\ldots\ldots.(\mathbf{7})$$

Where, for convenience it is assumed that the angle of the unmodulated carrier wave is zero at t=0. The frequency modulated wave is therefore described in the time domain by

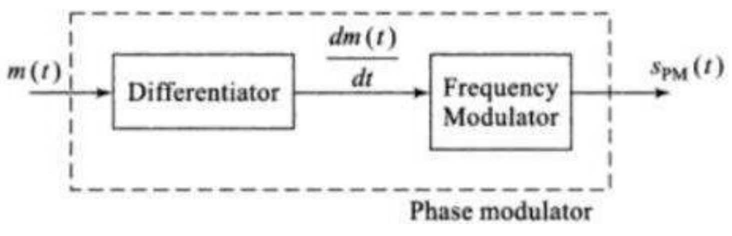
$$\boldsymbol{s(t) = A_c \cos\left[\, 2\pi f_c t + 2\pi k_f \int_0^t m(t)\ dt\right]} \ldots\ldots\ldots\ldots\ldots\ldots..(\mathbf{8})$$

Relationship between PM and FM

Comparing equ (5) with (8) reveals that an FM wave may be regarded as a PM wave in which the modulating wave is $\int_0^t m(t)dt$ in place of *m(t)*.



A PM wave can be generated by first differentiating *m(t)* and then using the result as the input to a frequency modulator.



Thus the properties of PM wave can be deduced from those of FM waves and vice versa

Single tone Frequency modulation

Consider a sinusoidal modulating wave defined by

$$\boldsymbol{m(t) = A_m \cos(2\pi f_m t) \dots \dots \dots \dots \dots (1)}$$

The instantaneous frequency of the resulting FM wave is

$$\boldsymbol{f_i(t) = f_c + k_f A_m \cos(2\pi f_m t)}$$

$$\boldsymbol{f_i(t) = f_c + \Delta f \cos(2\pi f_m t) \dots \dots \dots \dots . . (2)}$$

Where $\boldsymbol{\Delta f = k_f A_m}$ .... $\boldsymbol{(3)}$

The quantity Δf is called the ***frequency deviation***, representing the maximum departure of the instantaneous frequency of the FM wave from the carrier frequency f_c.

Fundamental characteristic of an FM wave is that the frequency deviation Δf is proportional to the amplitude of the modulating wave, and is independent of the modulation frequency.

Using equation (2), the angle $\theta_i(t)$ of the FM wave is obtained as

$$\boldsymbol{\theta_i(t) = 2\pi \int_0^t f_i(t) dt}$$

$$\boldsymbol{\theta_i(t) = 2\pi f_c t + \frac{\Delta f}{f_m} sin(2\pi f_m t) \dots \dots \dots \dots \dots \dots \dots (4)}$$

The ratio of the frequency deviation Δf *to* the modulation frequency f_m is commonly called the ***modulation index*** of the FM wave. Modulation index is denoted by $\boldsymbol{\beta}$ and is given as

$$\mathrm{Q} = \boldsymbol{\frac{\Delta f}{f_m} \dots \dots \dots \dots \dots \dots \dots . (5)}$$

And

$$\boldsymbol{\theta_i(t) = 2\pi f_c t + \mathrm{Q} sin(2\pi f_m t) \dots \dots \dots \dots \dots \dots \dots (6)}$$

In equation (6) the parameter $\boldsymbol{\beta}$ represents the phase deviation of the FM wave, that is, the maximum departure of the angle $\theta_i(t)$ from the angle $\boldsymbol{2\pi f_c t}$ of the unmodulated carrier.

The FM wave itself is given by

$$\boldsymbol{s(t) = A_c \cos[2\pi f_c t + \mathrm{Q} sin(2\pi f_m t)] \dots \dots \dots \dots \dots \dots . . (7)}$$

Depending on the value of modulation index β, we may distinguish two cases of frequency modulation. Narrow-band FM for which β is small and Wide-band FM for which β is large, both compared to one radian.

Narrow-Band Frequency modulation

Consider the Single tone FM wave

$$s(t) = A_c \cos[2\pi f_c t + \text{Q}sin(2\pi f_m t)] \; \dots\dots\dots\dots.(1)$$

Expanding this relation we get

$$s(t) = A_c \cos(2\pi f_c t) \cos[\beta \sin(2\pi f_m t)] - A_c \sin(2\pi f_c t) \sin[\beta \sin(2\pi f_m t)] \; \dots\dots\text{(2)}$$

Assuming that the modulation index β is small compared to one radian, we may use the following approximations:

$$\cos[\beta \sin(2\pi f_m t)] \simeq 1$$

and

$$\sin[\beta \sin(2\pi f_m t)] \simeq \beta \sin(2\pi f_m t)$$

Hence, Equation (2) simplifies to

$$s(t) \simeq A_c \cos(2\pi f_c t) - \beta A_c \sin(2\pi f_c t) \sin(2\pi f_m t) \; \dots\dots\dots\dots\dots\dots\text{(3)}$$

Equation (3) defines the approximate form of a narrowband FM signal produced by a sinusoidal modulating signal $A_m \cos(2\pi f_m t)$. From this representation we deduce the modulator shown in block diagram form in Figure . This modulator involves splitting the carrier wave $A_c \cos(2\pi f_c t)$ into two paths. One path is direct; the other path contains a -90 degree phase-shifting network and a product modulator, the combination of which generates a DSB-SC modulated signal. The difference between these two signals produces a narrowband FM signal, but with some distortion.

Ideally, an FM signal has a constant envelope and, for the case of a sinusoidal modulating signal of frequency f_m, the angle $\theta_i(t)$ is also sinusoidal with the same frequency.

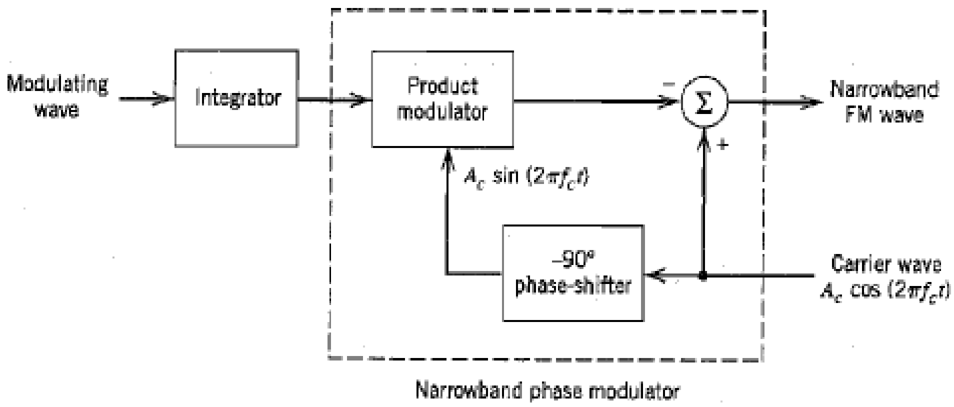


FIGURE Block diagram of a method for generating a narrowband FM signal.

But the modulated signal produced by the narrowband modulator of Figure differs from this ideal condition in two fundamental respects:

1. The envelope contains a *residual* amplitude modulation and, therefore, varies with time.
2. For a sinusoidal modulating wave, the angle $\theta_i(t)$ contains *harmonic distortion* in the form of third- and higher-order harmonics of the modulation frequency f_m.

However, by restricting the modulation index to $\beta \leq 0.3$ radians, the effects of residual AM and harmonic PM are limited to negligible levels.

Returning to Equation (3), we may expand it as follows:

$$s(t) = A_c \cos(2\pi f_c t) + \frac{1}{2}\beta A_c\{\cos[2\pi(f_c + f_m)t] - \cos[2\pi(f_c - f_m)t]\} \text{........(4)}$$

This expression is somewhat similar to the corresponding one defining an AM signal, which is as follows:

$$s_{\text{AM}}(t) = A_c \cos(2\pi f_c t) + \frac{1}{2}\mu A_c\{\cos[2\pi(f_c + f_m)t] + \cos[2\pi(f_c - f_m)t]\} \text{.........(5)}$$

where μ is the modulation factor of the AM signal. Comparing Equations (4) and (5), we see that in the case of sinusoidal modulation, the basic difference between an AM signal and a narrowband FM signal is that the algebraic sign of the lower side frequency in the narrowband FM is reversed. Thus, a narrowband FM signal requires essentially the same transmission bandwidth (i.e., $2f_m$) as the AM signal.

We may represent the narrowband FM signal with a phasor diagram as shown in Figure a , where we have used the carrier phasor as reference. We see that the resultant

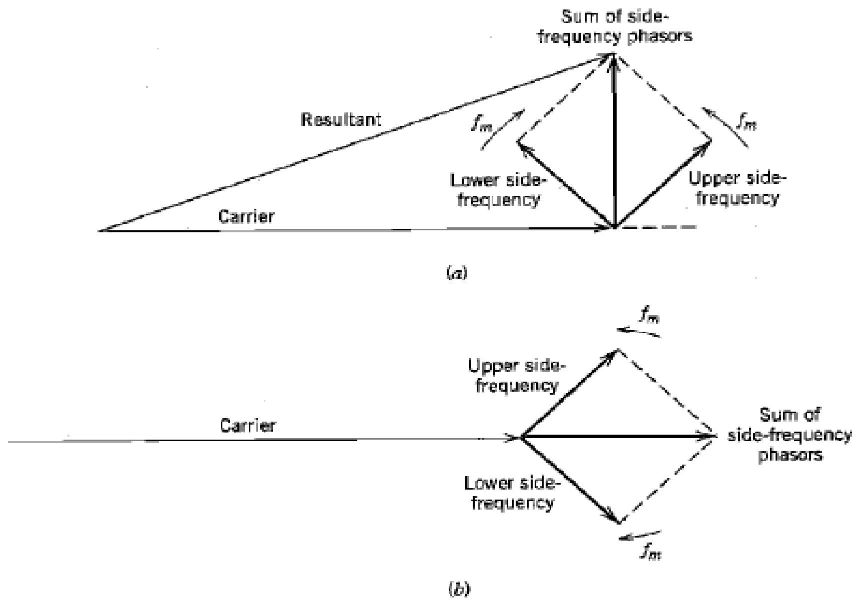


FIGURE A phasor comparison of narrowband FM and AM waves for sinusoidal modulation. (*a*) Narrowband FM wave. (*b*) AM wave.

of the two side-frequency phasors is always at right angles to the carrier phasor. The effect of this is to produce a resultant phasor representing the narrowband FM signal that is approximately of the same amplitude as the carrier phasor, but out of phase with respect to it. This phasor diagram should be contrasted with that of Figure *(b)* , representing an AM signal. In this latter case we see that the resultant phasor representing the AM signal has an amplitude that is different from that of the carrier phasor but always in phase with it.

Wide band frequency Modulation

The spectrum of the signle-tone FM wave of equation

$$\boldsymbol{s(t) = A_c \cos[2\pi f_c t + Q\sin(2\pi f_m t)]} \quad \dots\dots\dots\dots (1)$$

For an arbitrary vale of the modulation index Q is to be determined.

An FM wave produced by a sinusoidal modulating wave as in equation (1) is by itself nonperiodic, unless the carrier frequency f_c is an integral multiple of the modualtion frequency f_m. Rewriting the equation in the form

$$\begin{aligned} s(t) &= \mathrm{Re}[A_c \exp(j2\pi f_c t + j\beta \sin(2\pi f_m t))] \\ &= \mathrm{Re}[\tilde{s}(t) \exp(j2\pi f_c t)] \qquad \dots\dots\dots\dots(2) \end{aligned}$$

where $\tilde{s}(t)$ is the *complex envelope* of the FM signal $s(t)$, defined by

$$\tilde{s}(t) = A_c \exp[j\beta \sin(2\pi f_m t)] \qquad \dots\dots\dots\dots(3)$$

$\tilde{s}(t)$ is periodic function of time,with a fundamental frequency equal to the modulation frequency f_m. $\tilde{s}(t)$ in the form of complex Fourier series is as follows

$$\tilde{s}(t) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi n f_m t) \qquad \dots\dots\dots\dots(4)$$

where the complex Fourier coefficient c_n is defined by

$$\begin{aligned} c_n &= f_m \int_{-1/2f_m}^{1/2f_m} \tilde{s}(t) \exp(-j2\pi n f_m t)\, dt \\ &= f_m A_c \int_{-1/2f_m}^{1/2f_m} \exp[j\beta \sin(2\pi f_m t) - j2\pi n f_m t]\, dt \qquad \dots\dots\dots\dots(5) \end{aligned}$$

Define a new variable:

$$x = 2\pi f_m t \qquad \dots\dots\dots\dots(6)$$

Hence, we may rewrite Equation (5) in the new form

$$c_n = \frac{A_c}{2\pi} \int_{-\pi}^{\pi} \exp[j(\beta \sin x - nx)]\, dx \qquad \dots\dots\dots\dots(7)$$

The integral on the RHS of equation (7) is recognizedasthe nth order Bessel Function of the first kind and argument Q. This function is commonly denoted by the symbol J_n(Q), that is

$$J_n(\beta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp[j(\beta \sin x - nx)]\, dx \qquad \text{.........(8)}$$

Accordingly, we may reduce Equation (7) to

$$c_n = A_c J_n(\beta) \qquad \text{................(9)}$$

Substituting Equation (9) in (5), we get, in terms of the Bessel function $J_n(\beta)$, the following expansion for the complex envelope of the FM signal:

$$\tilde{s}(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \exp(j2\pi n f_m t) \qquad \text{..............(10)}$$

Next, substituting Equation (10) in (2.), we get

$$s(t) = A_c \cdot \mathrm{Re}\left[\sum_{n=-\infty}^{\infty} J_n(\beta) \exp[j2\pi(f_c + nf_m)t]\right] \qquad \text{.............(11)}$$

Interchanging the order of summation and evaluation of the real part in the right-hand side of Equation (11), we finally get

$$s(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + nf_m)t] \qquad \text{................(12)}$$

Equ. (12) is the Fourier series representation of the single-tone FM wave *s(t)* for an arbitrary value of Q.

The discrete spectrum of *s(t)* is obtained by taking the Fourier transform of both sides of equation (12); thus

$$S(f) = \frac{A_c}{2} \sum_{n=-\infty}^{\infty} J_n(\beta)[\delta(f - f_c - nf_m) + \delta(f + f_c + nf_m)] \text{.............(13)}$$

In the figure below, we have plotted the Bessel function J_n(Q) versus the modulation index Q for different positive integer value of n.

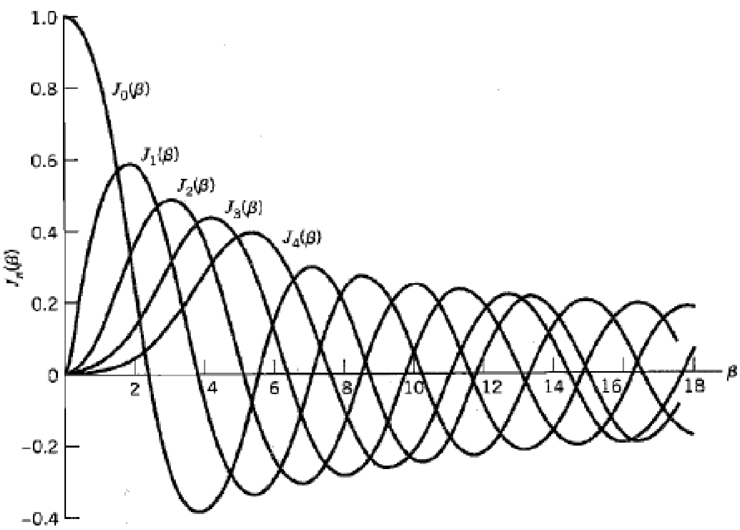


FIGURE Plots of Bessel functions of the first kind for varying order.

Properties of Bessel Function

1. $J_n(\beta) = (-1)^n J_{-n}(\beta)$ for all n, both positive and negative ....**(14)**

2. For small values of the modulation index β, we have

$$\left.\begin{aligned} J_0(\beta) &\simeq 1 \\ J_1(\beta) &\simeq \frac{\beta}{2} \\ J_n(\beta) &\simeq 0, \qquad n > 2 \end{aligned}\right\} \qquad \text{...............(15)}$$

3. $$\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1 \qquad \text{...................(16)}$$

Thus using equations (13) through (16) and the curves in the above figure, following observations are made

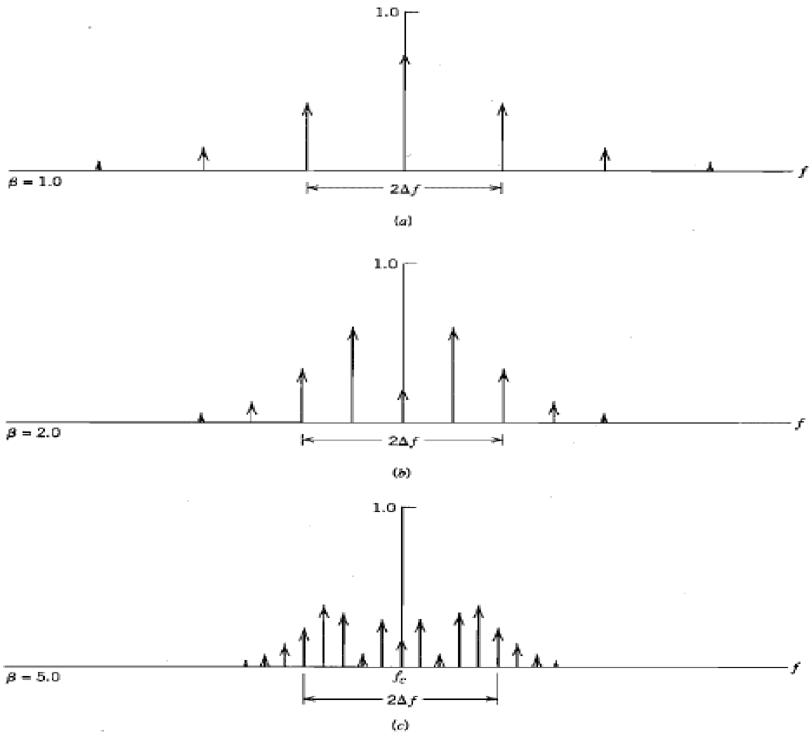
1. The spectrum of an FM signal contains a carrier component and an infinite set of side frequencies located symmetrically on either side of the carrier at frequency separations of f_m, $2f_m$, $3f_m$, $\cdots$. In this respect, the result is unlike that which prevails in an AM system, since in an AM system a sinusoidal modulating signal gives rise to only one pair of side frequencies.
2. For the special case of β small compared with unity, only the Bessel coefficients $J_0(\beta)$ and $J_1(\beta)$ have significant values, so that the FM signal is effectively composed of a carrier and a single pair of side frequencies at $f_c \pm f_m$. This situation corresponds to the special case of narrowband FM that was considered earlier.
3. The amplitude of the carrier component varies with β according to $J_0(\beta)$. That is, unlike an AM signal, the amplitude of the carrier component of an FM signal is dependent on the modulation index β. The physical explanation for this property is that the envelope of an FM signal is constant, so that the average power of such a signal developed across a 1-ohm resistor is also constant, as shown by

$$P = \frac{1}{2} A_c^2 \qquad \text{..................(17)}$$

When the carrier is modulated to generate the FM signal, the power in the side frequencies may appear only at the expense of the power originally in the carrier, thereby making the amplitude of the carrier component dependent on β. Note that the average power of an FM signal may also be determined from Equation (**12**), obtaining

$$P = \frac{1}{2} A_c^2 \sum_{n=-\infty}^{\infty} J_n^2(\beta) \qquad \text{......................(18)}$$

Spectrum Analysis of Sinusoidal FM Wave using Bessel functions



The above figure shows the Discrete amplitude spectra of an FM signal, normalized with respect to the carrier amplitude, for the case of sinusoidal modulation of varying frequency and fixed amplitude. Only the spectra for positive frequencies are shown.

Transmission Bandwidth of FM waves

In theory, an FM signal contains an infinite number of side frequencies so that the bandwidth required to transmit such a signal is similarly infinite in extent. In practice, however, we find that the FM signal is effectively limited to a finite number of significant side frequencies compatible with a specified amount of distortion. We may therefore specify an effective bandwidth required for the transmission of an FM signal. Consider first the

case of an FM signal generated by a single-tone modulating wave of frequency f_m. In such an FM signal, the side frequencies that are separated from the carrier frequency f_c by an amount greater than the frequency deviation Δf decrease rapidly toward zero, so that the bandwidth always exceeds the total frequency excursion, but nevertheless is limited. Specifically, for large values of the modulation index β, the bandwidth approaches, and is only slightly greater than, the total frequency excursion $2\Delta f$ in accordance with the situation shown in Figure 2.25. On the other hand, for small values of the modulation index β, the spectrum of the FM signal is effectively limited to the carrier frequency f_c and one pair of side frequencies at $f_c \pm f_m$, so that the bandwidth approaches $2f_m$. We may thus define an approximate rule for the transmission bandwidth of an FM signal generated by a single-tone modulating signal of frequency f_m as follows:

$$B_T \simeq 2\Delta f + 2f_m = 2\Delta f\left(1 + \frac{1}{\beta}\right)$$

This relation is known as **Carson's rule.**

Generation of FM Signal

Direct methods for FM generation

Reactance modulator:

The direct method of FM generation using the reactance modulator involves providing a voltage-variable reactance across the tank circuit of an oscillator. Though the varactor diode modulator can be called a reactance modulator, the term is generally applied to those modulators in which an active device is made to behave as a variable reactance. In Fig. we show the basic FET capacitive reactance modulator.

Under certain conditions, the impedance z, across terminals AA' is almost entirely reactive. The circuit can be made either inductive or capacitive by a simple component change and its reactance can be shown to be proportional to the transconductance of the device, which in turn can be made to depend on the variations in the gate bias. The circuit impedance is: $z = \frac{v}{i_d}$, where v is the voltage applied across AA' and i_d is the resulting drain current. For the impedance across AA' to be capacitive, the following conditions have to be met:

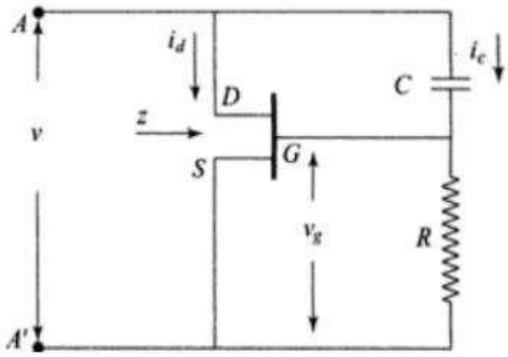


Fig. *FET capacitive reactance modulator*

- The bias current i_c must be negligible compared to the drain current i_d.
- The drain-to-gate impedance (X_c) must be greater than gate-to-source impedance (R here) by at least a factor of five.

$$v_g = i_c R = \frac{vR}{(R - jX_c)}$$

Drain current: $i_d = g_m v_g = \dfrac{g_m vR}{(R - jX_c)}$

Since $X_c \gg R$, we have:

$$z = \frac{v}{i_d} \approx -\frac{jX_c}{g_m R}$$

This is clearly a capacitive reactance with equivalent impedance:

$$X_{eq} = \frac{X_c}{g_m R} = \frac{1}{2\pi f C g_m R} = \frac{1}{2\pi f C_{eq}}$$

Hence, under these assumptions, the impedance looking into AA' is a pure reactance given by:

$$C_{eq} = g_m CR$$

Since C_{eq} depends on the transconductance g_m, it can be varied with bias voltage.

Now $X_c \gg R$. Let $X_c = nR$ at the carrier frequency. Then, $\dfrac{1}{\omega C} = nR$.

Therefore,
$$C = \frac{1}{\omega nR} = \frac{1}{2\pi fnR}$$

$$C_{eq} = g_m CR = \frac{g_m}{2\pi fn}$$

What would have happened if the positions of C and R are interchanged and if $R \gg X_c$?

$$v_g = i_c R = \left[\frac{v(-jX_c)}{(R - jX_c)}\right]$$

Drain current: $i_d = g_m v_g = \left[\frac{g_m v(-jX_c)}{(R - jX_c)}\right]$

$$z = \frac{v}{i_d} = \frac{(R - jX_c)}{(-jX_c g_m)}$$

$$= \left(\frac{1}{X_c g_m}\right)[X_c + jR] \approx \frac{jR}{(X_c g_m)}$$

Clearly, the impedance is inductive and can be written as:

$$X_{eq} = \frac{R}{X_c g_m} = \frac{(2\pi fCR)}{g_m} = 2\pi f L_{eq}, \text{ where } L_{eq} = \frac{CR}{g_m}$$

Thus, the FET reactance modulator behaves as a three-terminal reactive element (either inductive or capacitive) that may be connected across the tank circuit of the oscillator to be frequency modulated. The reactance appears between the drain and the source and its value may be controlled by a signal at the third terminal, i.e., the gate.

A disadvantage of the above methods of FM generation is that they do not provide carrier frequency stability. For attaining the required order of carrier frequency stability in the 88-108 MHz range used for FM transmission, it is necessary to use crystal oscillators. However, the high Q of crystal oscillators permits direct modulation only in some narrowband applications[2]. For wideband FM generation using crystal oscillators, the indirect method is adopted.

Indirect Method for WBFM Generation (ARMSTRONG'S Method):

In this method, an NBFM signal is generated using an integrator and a phase modulator. The NBFM signal is then converted to WBFM using a frequency multiplier. Let us first consider the principle behind a frequency multiplier.

A frequency multiplier is an amplifier whose output signal frequency is an integer multiple of the input frequency. If the input to the frequency multiplier is $A\cos\theta(t)$ then the output is $A\cos[n\theta(t)]$, where n is an integer. This can be achieved by feeding a signal frequency that is rich in harmonic distortion (e.g. from a class C amplifier) into an LC tank circuit tuned to n times the input frequency. This arrangement results in the n^{th} harmonic being the only significant output. For larger multiplication factors, a cascade of doublers and triplers are used. For example $n = 1000$ could be approximated by a cascade of ten doublers ($2^{10} = 1024$).

Effect of frequency multiplication on a NBFM signal

Consider single tone NBFM:

$s_{NBFM}(t) = A_c\cos(\omega_c' t + \beta' \sin\omega_m t)$ where $\beta' < 1$ rad and ω_c' is a stable sub-carrier frequency.

When applied to a frequency multiplier, the NBFM signal is converted to a WBFM signal given by:

$$s_{FM}(t) = A_c\cos[n(\omega_c' t + \beta' \sin\omega_m t)]$$
$$= A_c\cos[n\omega_c' t + n\beta' \sin\omega_m t]$$
$$s_{FM}(t) = A_c\cos[\omega_c t + \beta\sin\omega_m t]$$

where $\omega_c = n\omega_c'$ is the final carrier frequency and $\beta = n\beta'$ is the final β.

Note that ω_m is unaffected by frequency multiplication.

The maximum frequency deviation of the NBFM signal also gets multiplied by a factor of n since $\beta = n\beta'$ implies $(\Delta\omega/\omega_m) = n(\Delta\omega'/\omega_m)$ or $\Delta\omega = n\,\Delta\omega'$.

Detection of FM Signal

Balanced Slope Detector

Balanced slope detector. Figure **(a)** shows the schematic diagram for a *balanced slope detector*. A single-ended slope detector is a tuned-circuit frequency discriminator, and a balanced slope detector is simply two single-ended slope detectors connected in parallel and fed 180° out of phase. The phase inversion is accomplished by center tapping the tuned secondary windings of transformer T_1. In Figure **(a)**, the tuned circuits (L_a, C_a, and L_b, C_b) perform the FM-to-AM conversion, and the balanced peak detectors (D_1, C_1, R_1, and D_2, C_2, R_2) remove the information from the AM envelope. The top tuned circuit (L_a and C_a) is tuned to a frequency (f_a) that is above the IF center frequency (f_o) by approximately $1.33 \times \Delta f$ (for the FM broadcast band, this is approximately 1.33×75 kHz $= 100$ kHz). The lower tuned circuit (L_b and C_b) is tuned to a frequency (f_b) that is below the IF center frequency by an equal amount.

Circuit operation is quite simple. The output voltage from each tuned circuit is proportional to the input frequency, and each output is rectified by its respective peak detector. Therefore, the closer the input frequency is to the tank-circuit resonant frequency, the greater the tank-circuit output voltage. The IF center frequency falls exactly halfway between the resonant frequencies of the two tuned circuits. Therefore, at the IF center frequency, the output voltages from the two tuned circuits are equal in amplitude but opposite in polarity. Consequently, the rectified output voltage across $R1$ and $R2$,

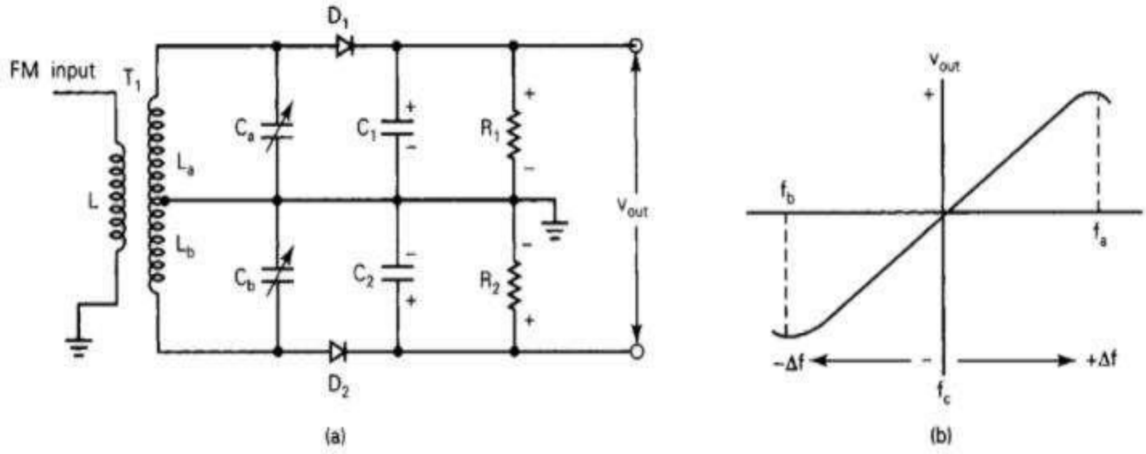


FIGURE Balanced slope detector: (a) schematic diagram; (b) voltage-versus-frequency response curve

when added, produce a differential output voltage V_{out} 5 0 V. When the IF deviates above resonance, the top tuned circuit produces a higher output voltage than the lower tank circuit, and Vout goes positive. When the IF deviates below resonance, the output voltage from the lower tank circuit is larger than the output voltage from the upper tank circuit, and V_{out} goes negative. The output-versus-frequency response curve is shown in Figure **(b).**

Although the slope detector is probably the simplest FM detector, it has several inherent disadvantages, which include poor linearity, difficulty in tuning, and lack of provisions for limiting. Because limiting is not provided, a slope detector produces an output voltage that is proportional to amplitude, as well as frequency variations in the input signal and, consequently, must be preceded by a separate limiter stage. A balanced slope detector is aligned by injecting a frequency equal to the IF center frequency and tuning C_a and C_b for 0 V at the output. Then frequencies equal to f_a and f_b are alternately injected while C_a and C_b are tuned for maximum and equal output voltages with opposite polarities.

Phase Locked Loop

Basically, the *phase-locked loop* consists of three major components: a *multiplier*, a *loop filter*, and a *voltage-controlled oscillator* (VCO) connected together in the form of a feedback system, as shown in Figure below. The VCO is a sinusoidal generator whose frequency is determined by a voltage applied to it from an external source. In effect, any frequency modulator may serve as a VCO. We assume that initially we have adjusted the VCO so that when the control voltage is zero, two conditions are satisfied:

1. The frequency of the VCO is precisely set at the unmodulated carrier frequency f_c.
2. The VCO output has a 90 degree phase-shift with respect to the unmodulated carrier wave.

Suppose then that the input signal applied to the phase-locked loop is an FM signal defined by

$$s(t) = A_c \sin[2\pi f_c t + \phi_1(t)]$$

where A_c is the carrier amplitude. With a modulating signal $m(t)$, the angle $\phi_1(t)$ is related to $m(t)$ by the integral

$$\phi_1(t) = 2\pi k_f \int_0^t m(\tau)\, d\tau$$

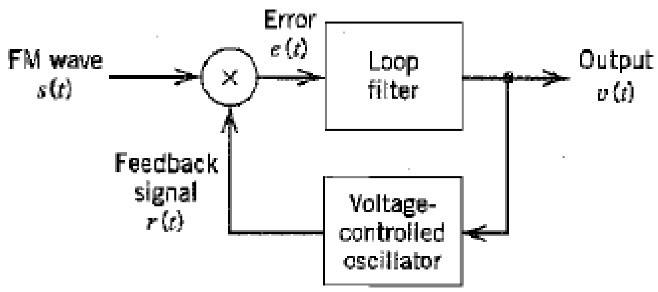


FIGURE Phase-locked loop.

where k_f is the frequency sensitivity of the frequency modulator. Let the VCO output in the phase-locked loop be defined by

$$r(t) = A_v \cos[2\pi f_c t + \phi_2(t)]$$

where A_v is the amplitude. With a control voltage $v(t)$ applied to the VCO input, the angle $\phi_2(t)$ is related to $v(t)$ by the integral

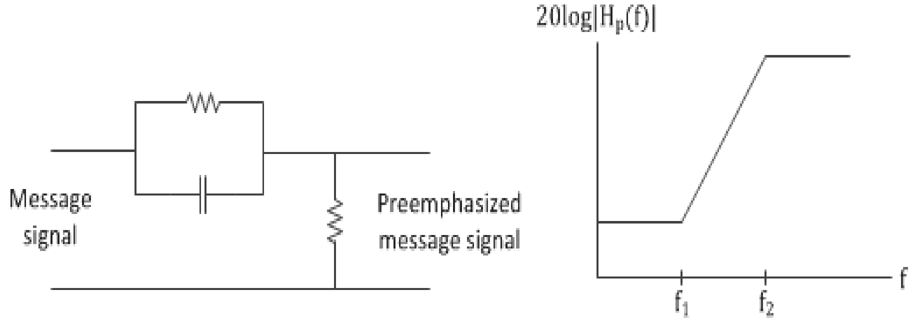
$$\phi_2(t) = 2\pi k_v \int_0^t v(\tau)\, d\tau$$

where k_v is the frequency sensitivity of the VCO, measured in Hertz per volt. The object of the phase-locked loop is to generate a VCO output $r(t)$ that has the same phase angle (except for the fixed difference of 90 degrees) as the input FM signal $s(t)$. The time-varying phase angle $\phi_1(t)$ characterizing $s(t)$ may be due to modulation by a message signal $m(t)$, in which case we wish to recover $\phi_1(t)$ and thereby produce an estimate of $m(t)$. In other applications of the phase-locked loop, the time-varying phase angle $\phi_1(t)$ of the incoming signal $s(t)$ may be an unwanted phase shift caused by fluctuations in the communication channel; in this latter case, we wish to *track* $\phi_1(t)$ so as to produce a signal with the same phase angle for the purpose of coherent detection (synchronous demodulation).

PRE-EMPHASIS AND DE-EMPHASIS NETWORKS

In FM, the noise increases linearly with frequency. By this, the higher frequency components of message signal are badly affected by the noise. To solve this problem, we can use a pre-emphasis filter of transfer function $H_p(f)$ at the transmitter to boost the higher frequency components before modulation. Similarly, at the receiver, the de-emphasis filter of transfer function $H_d(f)$can be used after demodulator to attenuate the higher frequency components thereby restoring the original message signal.

The pre-emphasis network and its frequency response are shown in Figure (a) and (b) respectively. Similarly, the counter part for de-emphasis network is shown in Figure below.



(a) (b)

Figure (a) Pre-emphasis network. (b) Frequency response of pre-emphasis network.

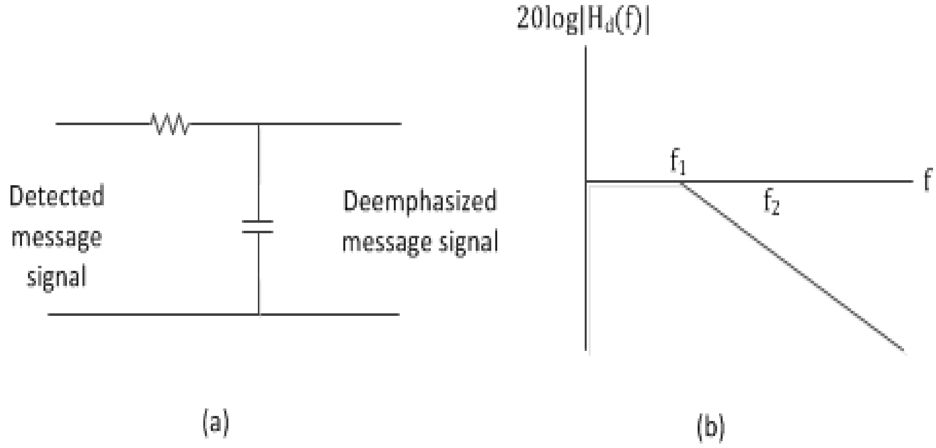


Figure (a) De-emphasis network. (b) Frequency response of De-emphasis network.

Comparison of AM and FM

S.NO	AMPLITUDE MODULATION	FREQUENCY MODULATION
1.	Band width is very small which is one of the biggest advantage	It requires much wider channel (7 to 15 times) as compared to AM.
2.	The amplitude of AM signal varies depending on modulation index.	The amplitude of FM signal is constant and independent of depth of the modulation.
3.	Area of reception is large	The area of reception is small since it is limited to line of sight.
4.	Transmitters are relatively simple & cheap.	Transmitters are complex and hence expensive.
5.	The average power in modulated wave is greater than carrier power. This added power is provided by modulating source.	The average power in frequency modulated wave is same as contained in un-modulated wave.
6.	More susceptible to noise interference and has low signal to noise ratio, it is more difficult to eliminate effects of noise.	Noise can be easily minimized amplitude variations can be eliminated by using limiter.
7.	It is not possible to operate without interference.	It is possible to operate several independent transmitters on same frequency.
8.	The maximum value of modulation index = 1, otherwise over-modulation would result in distortions.	No restriction is placed on modulation index.

UNIT - II (a)

NOISE

- Noise in communication System,
- White Noise
- Narrowband Noise –In phase and Quadrature phase components
- Noise Bandwidth
- Noise Figure
- Noise Temperature
- Noise in DSB& SSB System
- Noise in AM System
- Noise in Angle Modulation System
- Threshold effect in Angle Modulation System

Noise in communication system

A signal may be contaminated along the path by noise. Noise may be defined as any unwanted introduction of energy into the desired signal. In radio receivers, noise may produce "hiss" in the loudspeaker output. Noise is random and unpredictable.

Noise is produced both external and internal to the system. External noise includes atmospheric noise (e.g., from lightning), galactic noise (thermal radiation from cosmic bodies), and industrial noise (e.g., from motors, ignition). We can minimize or eliminate external noise by proper system design. On the other hand, internal noise is generated inside the system. It is resulted due to random motion of charged particles in resistors, conductors, and electronic devices. With proper system design, it can be minimized but never can be eliminated. This is the major constraint in the rate of telecommunications.

- Noise is unwanted signal that affects wanted signal
- Noise is random signal that exists in communication systems

Effect of noise

- Degrades system performance (Analog and digital)
- Receiver cannot distinguish signal from noise
- Efficiency of communication system reduces

Types of noise

- Thermal noise/white noise/Johnson noise or fluctuation noise
- Shot noise
- Noise temperature
- Quantization noise

Noise temperature

Equivalent noise temperature is not the physical temperature of amplifier, but a theoretical construct, that is an equivalent temperature that produces that amount of noise power

$T_e = (F - 1)$

White noise

One of the very important random processes is the *white noise* process. Noises in many practical situations are approximated by the white noise process. Most importantly, the white noise plays an important role in modelling of WSS signals.

A white noise process $\{W(t)\}$ is a random process that has constant power spectral density at all frequencies. Thus

$$S_W(\omega) = \frac{N_0}{2} \qquad -\infty < \omega < \infty$$

where N_0 is a real constant and called the *intensity* of the white noise. The corresponding autocorrelation function is given by

$$R_W(\tau) = \frac{N}{2}\delta(\tau)$$ where $\delta(\tau)$ is the Dirac delta.

The average power of white noise

$$P_{avg} = EW^2(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{N}{2}d\omega \rightarrow \infty$$

The autocorrelation function and the PSD of a white noise process is shown in Figure 1 below.

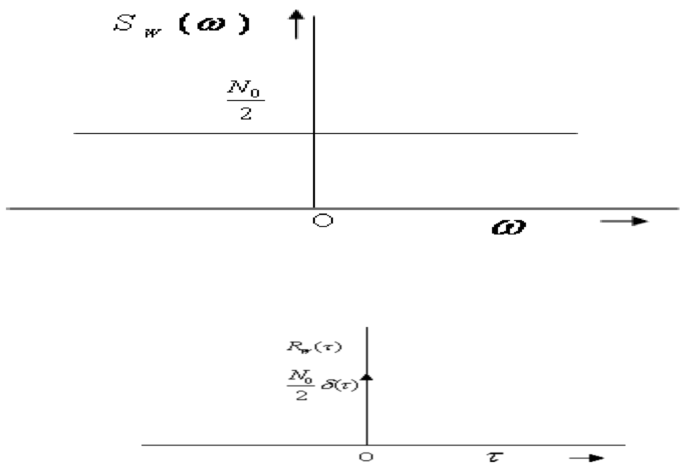


fig: auto correlation and psd of white noise

NARROWBAND NOISE (NBN)

In most communication systems, we are often dealing with band-pass filtering of signals. Wideband noise will be shaped into band limited noise. If the bandwidth of the band limited noise is relatively small compared to the carrier frequency, we refer to this as *narrowband noise*.

the narrowband noise is expressed as as

$$n(t) = x(t)\cos 2\pi f_C t - y(t)\sin 2\pi f_C t$$

where f_C is the carrier frequency within the band occupied by the noise. $x(t)$ and $y(t)$ are known as the *quadrature components* of the noise $n(t)$. The Hibert transform of *n*(*t*) is

Proof.

The Fourier transform of *n*(*t*) is

$$N(f) = \frac{1}{2}X(f-f_c) + \frac{1}{2}X(f+f_c) + \frac{1}{2}jY(f-f_c) - \frac{1}{2}jY(f+f_c)$$

Let $\hat{N}(f)$ be the Fourier transform of $\hat{n}(t)$. In the frequency domain, $\hat{N}(f) = N(f)[-j\ \mathrm{sgn}(f)]$. We simply multiply all positive frequency components of *N*(*f*) by -*j* and all negative frequency components of *N*(*f*) by *j*. Thus

$$\hat{n}(t) = H[n(t)] = x(t)\sin 2\pi f_c t + y(t)\cos 2\pi f_c t$$

$$\hat{N}(f) = -j\frac{1}{2}X(f-f_c) + j\frac{1}{2}X(f+f_c) - j\frac{1}{2}jY(f-f_c) - j\frac{1}{2}jY(f+f_c)$$

$$\hat{N}(f) = -j\frac{1}{2}X(f-f_c) + j\frac{1}{2}X(f+f_c) + \frac{1}{2}Y(f-f_c) + \frac{1}{2}Y(f+f_c)$$

and the inverse Fourier transform of $\hat{N}(f)$ is

$$\hat{n}(t) = x(t)\sin 2\pi f_c t + y(t)\cos 2\pi f_c t$$

The quadrature components x(*t*) and *y*(*t*) can now be derived from equations

$x(t) = n(t)\text{co}2\pi f_c t + n^{\wedge}(t)\sin 2\pi f_c t$ \

and

$y(t) = n(t)\cos 2\pi f_c t - n^{\wedge}(t)\sin 2\pi f_c t$

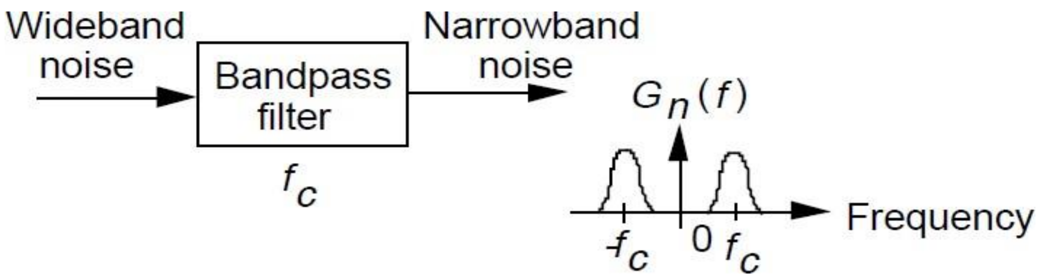


Fig: generation of narrow band noise

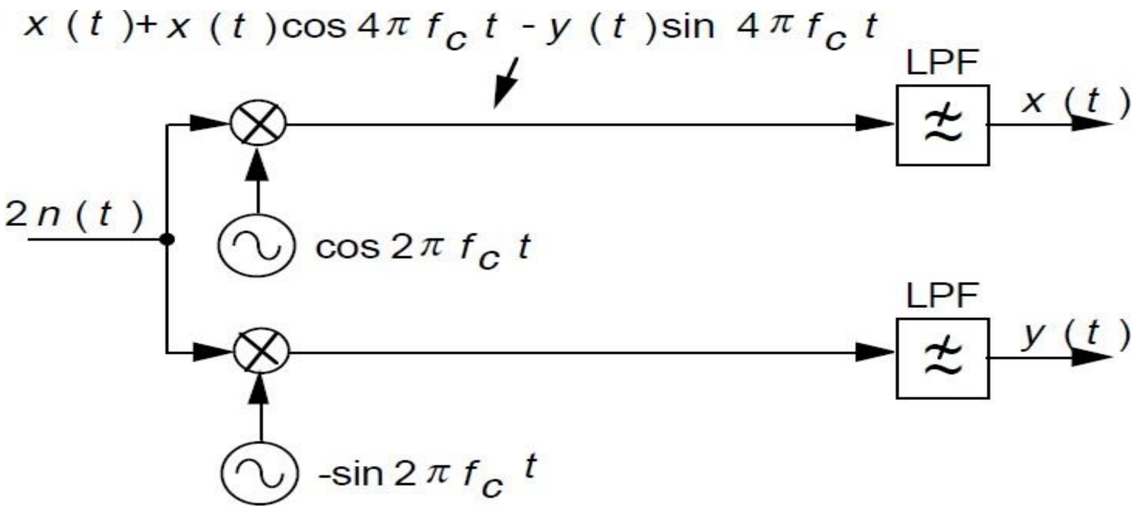


Fig: Generation of quadrature components of $n(t)$.

- Filters at the receiver have enough bandwidth to pass the desired signal but not too big to pass excess noise.
- Narrowband (NB) fc center frequency is much bigger that the bandwidth.
- Noise at the output of such filters is called narrowband noise (NBN).
- NBN has spectral concentrated about some mid-band frequency f_C
- The sample function of such NBN $n(t)$ appears as a sine wave of frequency f_C which modulates slowly in amplitude and phase

Input signal-to-noise ratio (SNR_I): is the ratio of the average power of modulated signal $s(t)$ to the average power of the filtered noise.

Output signal-to-noise ratio (SNR_O): is the ratio of the average power of demodulated message to the average power of the noise, both measured at the receiver output.

Channel signal-to-noise ratio (SNR_C): is the ratio of the average power of modulated signal $s(t)$ to the average power of the noise in the message bandwidth, both measured at the receiver input.

Noise figure

The Noise figure is the amount of noise power added by the electronic circuitry in the receiver to the thermal noise power from the input of the receiver. The thermal noise at the input to the receiver passes through to the demodulator. This noise is present in the receive channel and cannot be removed. The noise figure of circuits in the receiver such as amplifiers and mixers, adds additional noise to the receive channel. This raises the noise floor at the demodulator.

$$Noise\ Figure = \frac{Signal\ to\ noise\ ratio\ at\ input}{Signal\ to\ noise\ ratio\ at\ output}$$

Noise Bandwidth

A filter's equivalent noise bandwidth (ENBW) is defined as the bandwidth of a perfect rectangular filter that passes the same amount of power as the cumulative bandwidth of the channel selective filters in the receiver. At this point we would like to know the noise floor in our receiver, i.e. the noise power in the receiver intermediate frequency (IF) filter bandwidth that comes from kTB. Since the units of kTB are Watts/ Hz, calculate the noise floor in the channel bandwidth by multiplying the noise power in a 1 Hz bandwidth by the overall equivalent noise bandwidth in Hz.

NOISE IN DSB-SC SYSTEM:

Let the transmitted signal is

$$u(t) = A_c m(t)\cos(2\pi f_c t)$$

The received signal at the output of the receiver noise- limiting filter : Sum of this signal and filtered noise .A filtered noise process can be expressed in terms of its in-phase and quadrature components as

$$\begin{aligned} n(t) &= A(t)\cos[2\pi f_c t + \theta(t)] = A(t)\cos\theta(t)\cos(2\pi f_c t) - A(t)\sin\theta(t)\sin(2\pi f_c t) \\ &= n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \end{aligned}$$

where $n_C(t)$ is in-phase component and $n_S(t)$ is quadrature component

Received signal (Adding the filtered noise to the modulated signal)

$$\begin{aligned} r(t) &= u(t) + n(t) \\ &= A_c m(t)\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \end{aligned}$$

Demodulate the received signal by first multiplying r(t) by a locally generated sinusoid $\cos(2\pi f_c t + \phi)$, where ϕ is the phase of the sinusoid.Then passing the product signal through an ideal lowpass filter having a bandwidth W.

The multiplication of r(t) with $\cos(2\pi f_c t + \phi)$ yields

$$\begin{aligned} & r(t)\cos(2\pi f_c t + \phi) \\ &= u(t)\cos(2\pi f_c t + \phi) + n(t)\cos(2\pi f_c t + \phi) \\ &= A_c m(t)\cos(2\pi f_c t)\cos(2\pi f_c t + \phi) \\ &\quad + n_c(t)\cos(2\pi f_c t)\cos(2\pi f_c t + \phi) - n_s(t)\sin(2\pi f_c t)\cos(2\pi f_c t + \phi) \\ &= \tfrac{1}{2} A_c m(t)\cos(\phi) + \tfrac{1}{2} A_c m(t)\cos(4\pi f_c t + \phi) \\ &\quad + \tfrac{1}{2}\left[n_c(t)\cos(\phi) + n_s(t)\sin(\phi)\right] + \tfrac{1}{2}\left[n_c(t)\cos(4\pi f_c t + \phi) - n_s(t)\sin(4\pi f_c t + \phi)\right] \end{aligned}$$

The low pass filter rejects the double frequency components and passes only the low pass components.

$$y(t) = \tfrac{1}{2} A_c m(t)\cos(\phi) + \tfrac{1}{2}\left[n_c(t)\cos(\phi) + n_s(t)\sin(\phi)\right]$$

the effect of a phase difference between the received carrier and a locally generated carrier at the receiver is a drop equal to $\cos^2(\phi)$ in the received signal power.

Phase-locked loop

The effect of a phase-locked loop is to generate phase of the received carrier at the receiver. If a phase-locked loop is employed, then $\phi = 0$ and the demodulator is called a coherent or synchronous demodulator.

In our analysis in this section, we assume that we are employing a coherent demodulator. With this assumption, we assume that $\phi = 0$

$$y(t) = \tfrac{1}{2}\left[A_c m(t) + n_c(t)\right]$$

Therefore, at the receiver output, the message signal and the noise components are additive and we are able to define a meaningful SNR. The message signal power is given by

$$P_o = \frac{A_c^2}{4} P_M$$

Power P_M is the content of the messagesignal

The noise power is given by

$$P_{n_0} = \frac{1}{4} P_{n_c} = \frac{1}{4} P_n$$

The power content of n(t) can be found by noting that it is the result of passing $n_w(t)$ through a filter with bandwidth B_c.Therefore, the power spectral density of n(t) is given by

$$S_n(f) = \begin{cases} \frac{N_0}{2} & |f - f_c| < W \\ 0 & otherwise \end{cases}$$

The noise power is

$$P_n = \int_{-\infty}^{\infty} S_n(f)df = \frac{N_0}{2} \times 4W = 2WN_0$$

Now we can find the output SNR as

$$\left(\frac{S}{N}\right)_0 = \frac{P_0}{P_{n_0}} = \frac{\frac{A_c^2}{4} P_M}{\frac{1}{4} 2WN_0} = \frac{A_c^2 P_M}{2WN_0}$$

In this case, the received signal power, given by

$$P_R = A_c{}^2 P_M/2.$$

The output SNR for DSB-SC AM may be expressed as

$$\left(\frac{S}{N}\right)_{0_{DSB}} = \frac{P_R}{N_0 W}$$

which is identical to baseband SNR.

In DSB-SC AM, the output SNR is the same as the SNR for a baseband system. DSB-SC AM does not provide any SNR improvement over a simple baseband communication system.

NOISE IN SSB-SC SYSTEM:

Let SSB modulated signal is

$$u(t) = A_c m(t)\cos(2\pi f_c t) \mp A_c \hat{m}(t)\sin(2\pi f_c t)$$

Input to the demodulator

$$
\begin{aligned}
r(t) &= A_c m(t)\cos(2\pi f_c t) \mp A_c \hat{m}(t)\sin(2\pi f_c t) + n(t) \\
&= A_c m(t)\cos(2\pi f_c t) \mp A_c \hat{m}(t)\sin(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \\
&= \left[A_c m(t) + n_c(t)\right]\cos(2\pi f_c t) + \left[\mp A_c \hat{m}(t) - n_s(t)\right]\sin(2\pi f_c t)
\end{aligned}
$$

Assumption : Demodulation with an ideal phase reference.

Hence, the output of the lowpass filter is the in-phase component (with a coefficient of ½) of the precedingsignal.

Parallel to our discussion of DSB, we have

$$P_o = \frac{A_c^2}{4} P_M$$

$$P_{n_0} = \frac{1}{4} P_{n_c} = \frac{1}{4} P_n$$

$$P_n = \int_{-\infty}^{\infty} S_n(f)\,df = \frac{N_0}{2} \times 2W = WN_0$$

$$\left(\frac{S}{N}\right)_0 = \frac{P_0}{P_{n_0}} = \frac{A_c^2 P_M}{WN_0}$$

$$P_R = P_U = A_c^2 P_M$$

$$\left(\frac{S}{N}\right)_{0_{SSB}} = \frac{P_R}{N_0 W} = \left(\frac{S}{N}\right)_b$$

The signal-to-noise ratio in an SSB system is equivalent to that of a DSB system.

Noise in Conventional AM

DSB AM signal : $u(t) = A_c[1 + a m_n(t)]\cos(2\pi f_c t)$

Received signal at the input to the demodulator

$$
\begin{aligned}
r(t) &= A_c[1 + a m_n(t)]\cos(2\pi f_c t) + n(t) \\
&= A_c[1 + a m_n(t)]\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t) \\
&= \left[A_c[1 + a m_n(t)] + n_c(t)\right]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)
\end{aligned}
$$

Where a is the modulation index

$m_n(t)$ is normalized so that its minimum value is -1

If a synchronous demodulator is employed, the situation is basically similar to the DSB case, except that we have $1 + a m_n(t)$ instead of $m(t)$.

$$y(t) = \tfrac{1}{2}\left[A_c a m_n(t) + n_c(t)\right]$$

Received signal power

$$P_R = \frac{A_c^2}{2}\left[1 + a^2 P_{M_n}\right]$$

- Assumed that the message process is zero mean.

Now we can derive the output SNR as

$$\left(\frac{S}{N}\right)_{0_{AM}} = \frac{\frac{1}{4}A_c^2 a^2 P_{M_n}}{\frac{1}{4}P_{n_c}} = \frac{A_c^2 a^2 P_{M_n}}{2N_0 W} = \frac{a^2 P_{M_n}}{1 + a^2 P_{M_n}} \frac{\frac{A_c^2}{2}\left[1 + a^2 P_{M_n}\right]}{N_0 W}$$

$$= \frac{a^2 P_{M_n}}{1 + a^2 P_{M_n}} \frac{P_R}{N_0 W} = \frac{a^2 P_{M_n}}{1 + a^2 P_{M_n}}\left(\frac{S}{N}\right)_b = \eta\left(\frac{S}{N}\right)_b$$

- η denotes the modulation efficiency
- Since $a^2 P_{M_n} < 1 + a^2 P_{M_n}$, the SNR in conventional AM is always smaller than the SNR in a baseband system.

- In practical applications, the modulation index a is in the range of 0.8-0.9.
- Power content of the normalized message process depends on the message source.
- Speech signals : Large dynamic range, P_M is about 0.1.
- The overall loss in SNR, when compared to a baseband system, is a factor of 0.075 or equivalent to a loss of 11 dB.

The reason for this loss is that a large part of the transmitter power is used to send the carrier component of the modulated signal and not the desired signal. To analyze the envelope-detector performance in the presence of noise, we must use certain approximations.

This is a result of the nonlinear structure of an envelope detector, which makes an exact analysis difficult

In this case, the demodulator detects the envelope of the received signal and the noise process.

The input to the envelope detector is

$$r(t) = \left[A_c[1 + a m_n(t)] + n_c(t)\right]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

Therefore, the envelope of *r (t)* is given by

$$V_r(t) = \sqrt{\left[A_c[1 + a m_n(t)] + n_c(t)\right]^2 + n_s^2(t)}$$

Now we assume that the signal component in *r (t)* is much stronger than the noise component. Then

$$P\left(n_c(t) << A_c[1+am_n(t)]\right) \approx 1$$

Therefore, we have a high probability that

$$V_r(t) \approx A_c[1+am_n(t)] + n_c(t)$$

After removing the DC component, we obtain

$$y(t) = A_c a m_n(t) + n_c(t)$$

which is basically the same as y*(t)* for the synchronous demodulation without the ½ coefficient.

This coefficient, of course, has no effect on the final SNR. So we conclude that, under the assumption of high SNR at the receiver input, the performance of synchronous and envelope demodulators is the same.

However, if the preceding assumption is not true, that is, if we assume that, at the receiver input, the noise power is much stronger than the signal power, Then

$$\begin{aligned}
V_r(t) &= \sqrt{\left[A_c[1+am_n(t)]+n_c(t)\right]^2 + n_s^2(t)} \\
&= \sqrt{A_c^2[1+am_n(t)]^2 + n_c^2(t) + n_s^2(t) + 2A_c n_c(t)[1+am_n(t)]} \\
&\xrightarrow{a} \sqrt{\left(n_c^2(t)+n_s^2(t)\right)\left[1+\frac{2A_c n_c(t)}{n_c^2(t)+n_s^2(t)}\left(1+am_n(t)\right)\right]} \\
&\xrightarrow{b} V_n(t)\left[1+\frac{A_c n_c(t)}{V_n^2(t)}\left(1+am_n(t)\right)\right] \\
&= V_n(t) + \frac{A_c n_c(t)}{V_n(t)}\left(1+am_n(t)\right)
\end{aligned}$$

(a) : $A_c^2[1+am_n(t)]^2$ is small compared with the other components

(b) : $\sqrt{n_c^2(t)+n_s^2(t)} = V_n(t)$;the envelope of the noise process

Use the approximation

$$\sqrt{1+\varepsilon} \approx 1+\tfrac{\varepsilon}{2}, \quad \textit{for small } \varepsilon, \text{ where } \varepsilon = \frac{2A_c n_c(t)}{n_c^2(t)+n_s^2(t)}\left(1+am_n(t)\right)$$

Then $$V_r(t) = V_n(t) + \frac{A_c n_c(t)}{V_n(t)}\left(1 + a m_n(t)\right)$$

We observe that, at the demodulator output, the signal and the noise components are no longer additive. In fact, the signal component is multiplied by noise and is no longer distinguishable. In this case, no meaningful SNR can be defined. We say that this system is operating below the threshold. The subject of threshold and its effect on the performance of a communication system will be covered in more detail when we discuss the noise performance in angle modulation.

Effect of threshold in angle modulation system:

FM THRESHOLD EFFECT FM threshold is usually defined as a Carrier-to-Noise ratio at which demodulated Signal-to-Noise ratio falls 1dB below the linear relationship . This is the effect produced in an FM receiver when noise limits the desired information signal. It occurs at about 10 dB, as earlier stated in 5 the introduction, which is at a point where the FM signal-to-Noise improvement is measured. Below the FM threshold point, the noise signal (whose amplitude and phase are randomly varying) may instantaneously have amplitude greater than that of the wanted signal. When this happens, the noise will produce a sudden change in the phase of the FM demodulator output. In an audio system, this sudden phase change makes a “click”. In video applications the term “click noise” is used to describe short horizontal black and white lines that appear randomly over a picture

An important aspect of analogue FM satellite systems is FM threshold effect. In FM systems where the signal level is well above noise received carrier-to-noise ratio and demodulated signal-to-noise ratio are related by:

$$\boxed{S/N \quad = \quad 3\,\beta^2\, C/N \qquad \text{Eqn 9}}$$

where S/N = signal-to-noise ratio at output of FM demodulator
β = FM deviation ratio or modulation index = ($\Delta f/B$)
C/N = carrier-to-noise ratio at input of FM demodulator
Δf = peak deviation
B = basebandwidth of signal being modulated

The expression however does not apply when the carrier-to-noise ratio decreases below a certain point. Below this critical point the signal-to-noise ratio decreases significantly. This is known as the FM threshold effect (FM threshold is usually defined as the carrier-to-noise ratio at which the demodulated signal-to-noise ratio fall 1 dB below the linear relationship given in Eqn 9. It generally is considered to occur at about 10 dB).

Below the FM threshold point the noise signal (whose amplitude and phase are randomly varying), may instantaneously have an amplitude greater than that of the wanted signal. When this happens the noise will produce a sudden change in the phase of the FM demodulator output. In an audio system this sudden phase change makes a "click". In video applications the term "click noise" is used to describe short horizontal black and white lines that appear randomly over a picture, because satellite communications systems are power limited they usually operate with only a small design margin above the FM threshold point (perhaps a few dB). Because of this circuit designers have tried to devise techniques to delay the onset of the FM threshold effect. These devices are generally known as FM threshold extension demodulators. Techniques such as FM feedback, phase locked loops and frequency locked loops are used to achieve this effect. By such techniques the onset of FM threshold effects can be delayed till the C/N ratio is around 7 dB.

Noise in Angle Modulated Systems

Like AM, noise performance of angle modulated systems is characterized by parameter γ

$$\gamma_{FM} = \frac{3}{2}\beta^2$$

If it is compared with AM

$$\frac{\gamma_{FM}}{\gamma_{AM}} = \frac{1}{2}\left(\frac{\omega_{FM}}{\omega_{AM}}\right)^2$$

Note: if bandwidth ratio is increased by a factor 2.Then $\frac{\gamma_{FM}}{\gamma_{AM}}$ increases by a factor 4

This exchange of bandwidth and noise performance is an important feature of FM

$$\text{Figure of merit}\,(\gamma) = \frac{SNR_O}{SNR_C}$$

CW- Modulation System	SNR_O	SNR_C	Figure of merit	Figure of merit (single tone)
DSB-SC	$\frac{C^2 A_c^2 P}{2WN_0}$	$\frac{C^2 A_c^2 P}{2WN_0}$	1	1
SSB	$\frac{C^2 A_c^2 P}{4WN_0}$	$\frac{C^2 A_c^2 P}{4WN_0}$	1	1
AM	$\frac{A_c^2 k_a^2 P}{2WN_0}$	$\frac{A_c^2 (1+k_a^2 P)}{2WN_0}$	$\approx \frac{k_a^2 P}{1+k_a^2 P} < 1$	$\frac{\mu^2}{2+\mu^2}$
FM	$\frac{3A_c^2 k_f^2 P}{2N_0 W^3}$	$\frac{A_c^2}{2WN_0}$	$\frac{3k_f^2 P}{W^2}$	$\frac{3}{2}\beta^2$

P is the average power of the message signal.

C^2 is a constant that ensures that the ration is dimensionless.

W is the message bandwidth.

A_c is the amplitude of the carrier signal.

k_a is the amplitude sensitivity of the modulator.

$\mu = k_a A_m$ and A_m is the amplitude sinusoidal wave

$\beta = \frac{\Delta f}{W}$ is the modulation index.

k_f is the frequency sensitivity of the modulator.

Δf is the frequency deviation.

UNIT-II (b)

PULSE MODULATION

Introduction:

Pulse Modulation

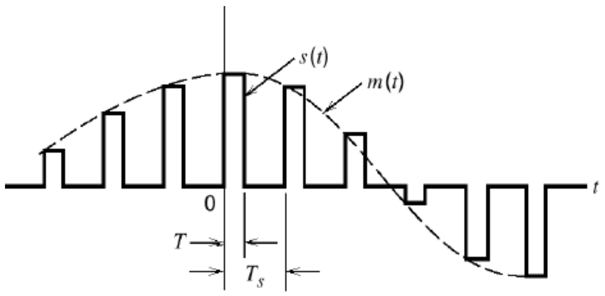
- Carrier is a train of pulses
- Example: Pulse Amplitude Modulation (PAM), Pulse width modulation (PWM) , Pulse Position Modulation (PPM)

Types of Pulse Modulation:

- The immediate result of sampling is a pulse-amplitude modulation (PAM) signal
- PAM is an analog scheme in which the amplitude of the pulse is proportional to the amplitude of the signal at the instant of sampling
- Another analog pulse-forming technique is known as **pulse-duration modulation (PDM).** This is also known as **pulse-width modulation (PWM)**
- **Pulse-position modulation** is closely related to PDM

Pulse Amplitude Modulation:

In PAM, amplitude of pulses is varied in accordance with instantaneous value of modulating signal.



PAM Generation:

The carrier is in the form of narrow pulses having frequency fc. The uniform sampling takes place in multiplier to generate PAM signal. Samples are placed Ts sec away from each other.

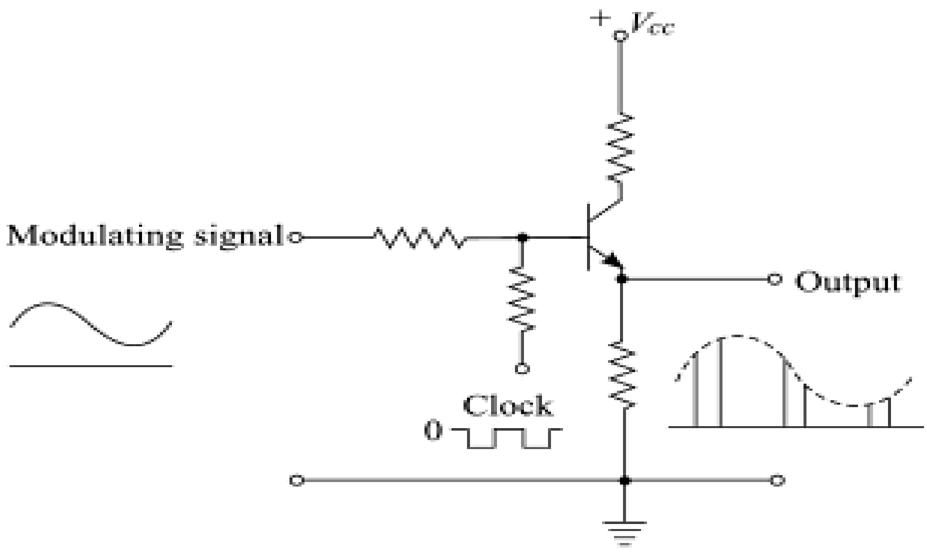


Figure PAM Modulator

- The circuit is simple emitter follower.
- In the absence of the clock signal, the output follows input.
- The modulating signal is applied as the input signal.
- Another input to the base of the transistor is the clock signal.
- The frequency of the clock signal is made equal to the desired carrier pulse train frequency.
- The amplitude of the clock signal is chosen the high level is at ground level(0v) and low level at some negative voltage sufficient to bring the transistor in cutoff region.
- When clock is high, circuit operates as emitter follower and the output follows in the input modulating signal.
- When clock signal is low, transistor is cutoff and output is zero.
- Thus the output is the desired PAM signal.

PAM Demodulator:

- The PAM demodulator circuit which is just an envelope detector followed by a second order op-amp low pass filter (to have good filtering characteristics) is as shown below

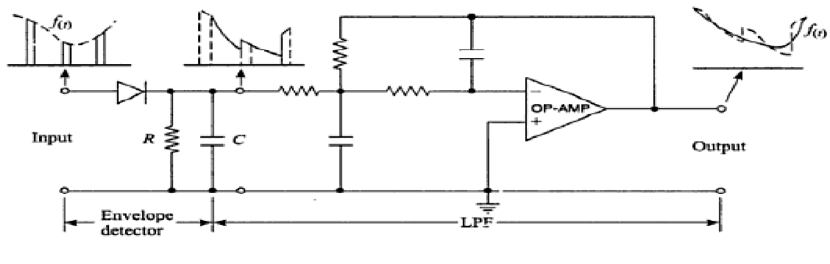
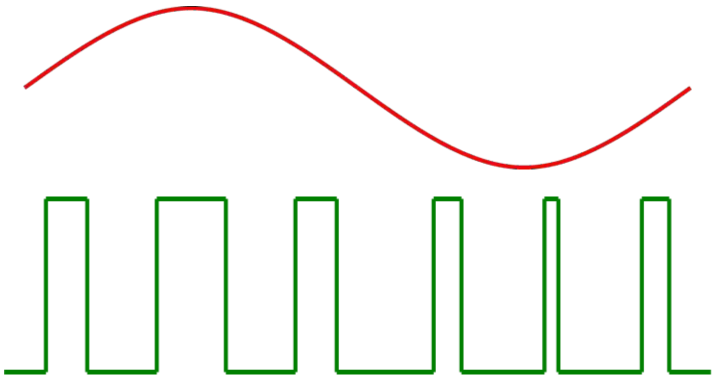


Figure PAM Demodulator

Pulse Width Modulation:

- In this type, the amplitude is maintained constant but the width of each pulse is varied in accordance with instantaneous value of the analog signal.



- In PWM information is contained in width variation. This is similar to FM.
- In pulse width modulation (PWM), the width of each pulse is made directly proportional to the amplitude of the information signal.

Pulse Position Modulation:

- In this type, the sampled waveform has fixed amplitude and width whereas the position of each pulse is varied as per instantaneous value of the analog signal.
- PPM signal is further modification of a PWM signal.

PPM & PWM Modulator:

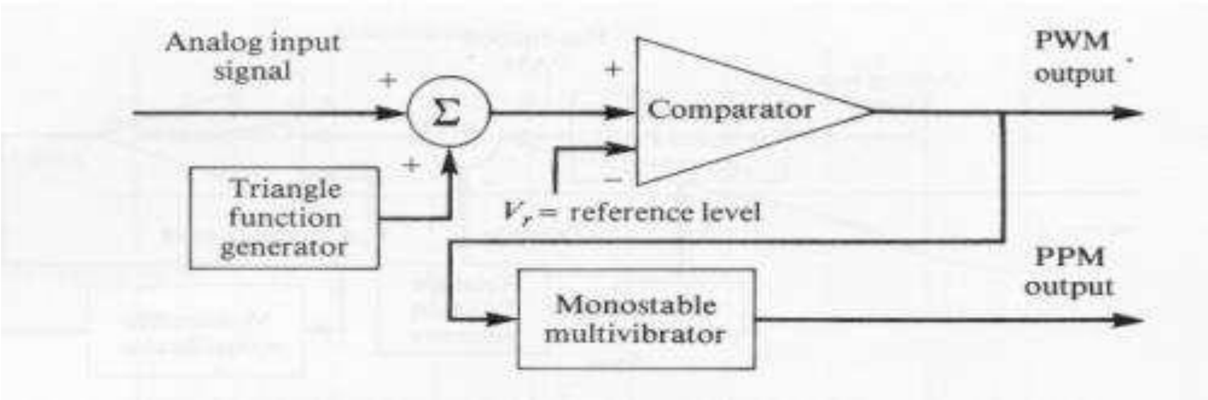


Figure PWM & PPM Modulator

- The PPM signal can be generated from PWM signal.
- The PWM pulses obtained at the comparator output are applied to a mono stable multi vibrator which is negative edge triggered.

- Hence for each trailing edge of PWM signal, the monostable output goes high. It remains high for a fixed time decided by its RC components.
- Thus as the trailing edges of the PWM signal keeps shifting in proportion with the modulating signal, the PPM pulses also keep shifting.
- Therefore all the PPM pulses have the same amplitude and width. The information is conveyed via changing position of pulses.

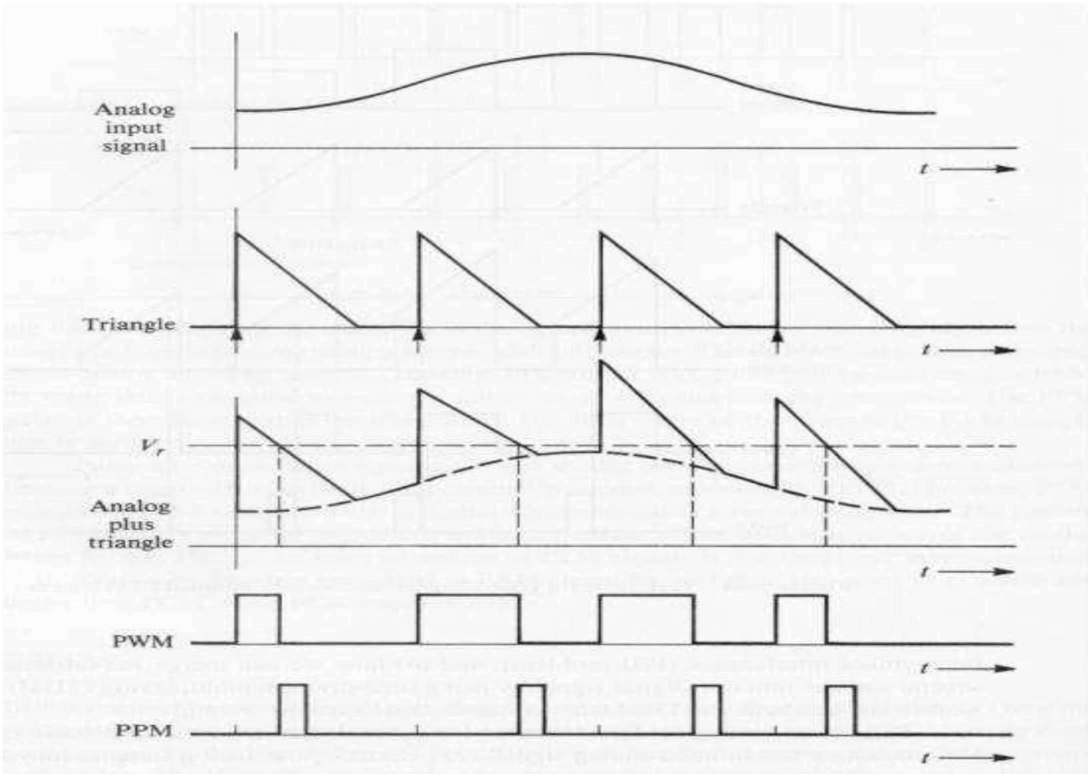


Figure PWM & PPM Modulation waveforms

PWM Demodulator:

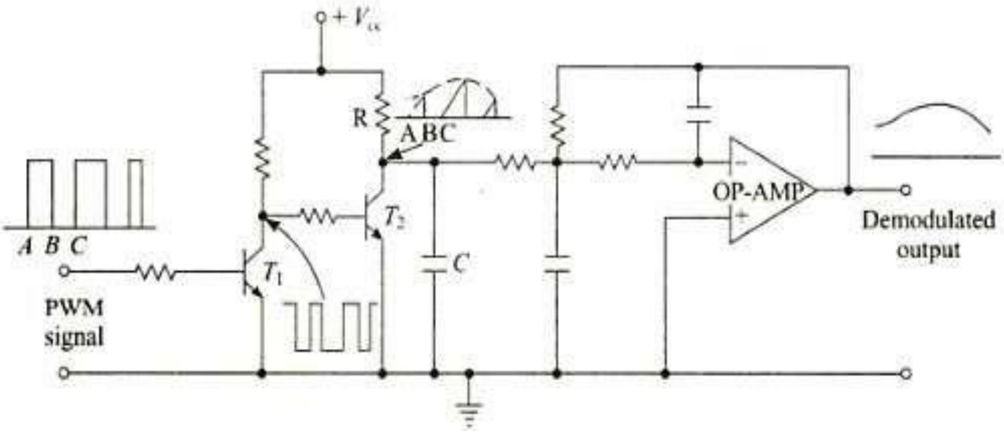


Figure PWM Demodulator

- Transistor T1 works as an inverter.
- During time interval A-B when the PWM signal is high the input to transistor T2 is low.
- Therefore, during this time interval T2 is cut-off and capacitor C is charged through an R-C combination.
- During time interval B-C when PWM signal is low, the input to transistor T2 is high, and it gets saturated.
- The capacitor C discharges rapidly through T2. The collector voltage of T2 during B-C is low.
- Thus, the waveform at the collector of T2is similar to saw-tooth waveform whose envelope is the modulating signal.
- Passing it through 2nd order op-amp Low Pass Filter, gives demodulated signal.

PPM Demodulator:

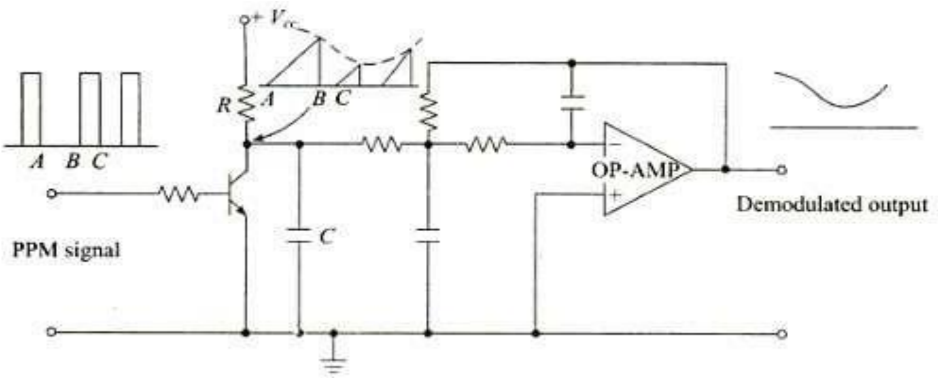


Figure PPM Demodulator

- The gaps between the pulses of a PPM signal contain the information regarding the modulating signal.
- During gap A-B between the pulses the transistor is cut-off and the capacitor C gets charged through R-C combination.
- During the pulse duration B-C the capacitor discharges through transistor and the collector voltage becomes low.
- Thus, waveform across collector is saw-tooth waveform whose envelope is the modulating signal.
- Passing it through 2nd order op-amp Low Pass Filter, gives demodulated signal.

Multiplexing

Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single common communications channel.

Multiplexing is the transmission of analog or digital information from one or more sources to one or more destination over the same transmission link.

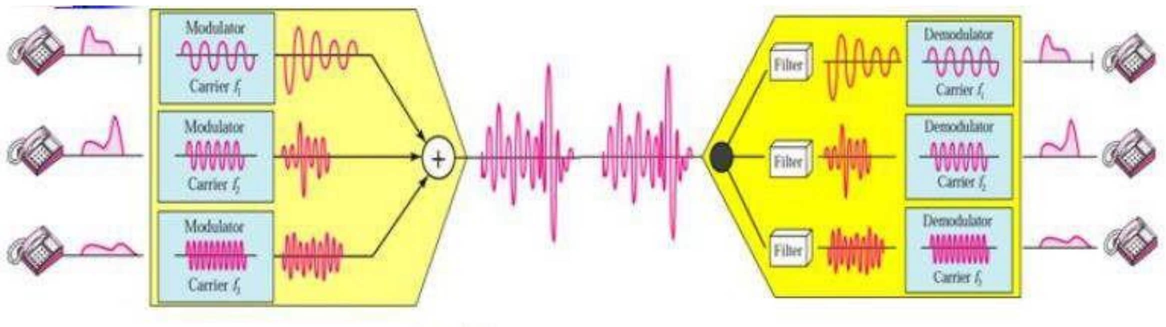
Although transmissions occur on the same transmitting medium, they do not necessarily occupy the same bandwidth or even occur at the same time.

Frequency Division Multiplexing

Frequency division multiplexing (FDM) is a technique of multiplexing which means combining more than one signal over a shared medium. In FDM, signals of different frequencies are combined for concurrent transmission.

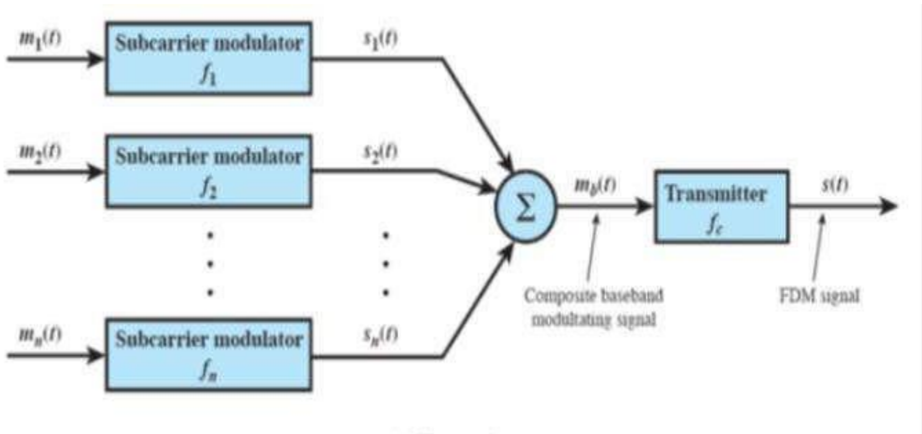
In FDM, the total bandwidth is divided to a set of frequency bands that do not overlap. Each of these bands is a carrier of a different signal that is generated and modulated by one of the sending devices. The frequency bands are separated from one another by strips of unused frequencies called the guard bands, to prevent overlapping of signals.

The modulated signals are combined together using a multiplexer (MUX) in the sending end. The combined signal is transmitted over the communication channel, thus allowing multiple independent data streams to be transmitted simultaneously. At the receiving end, the individual signals are extracted from the combined signal by the process of demultiplexing (DEMUX).



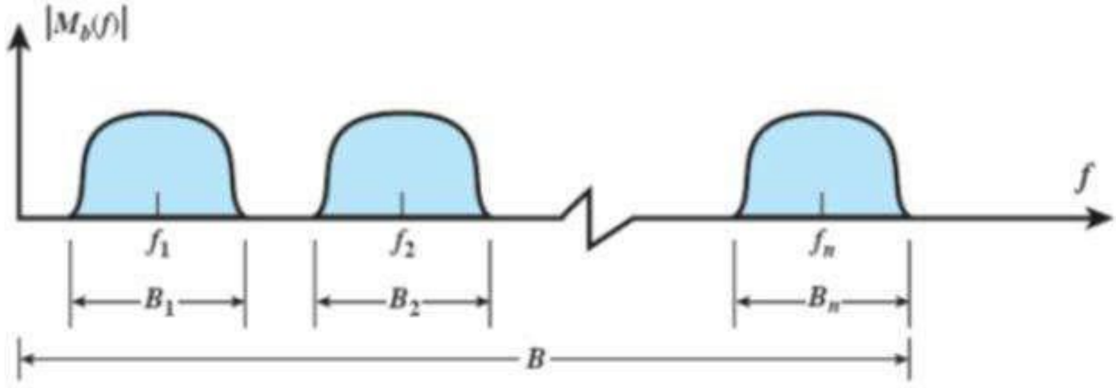
FDM system Transmitter

- Analog or digital inputs: m_i (t); i = 1,2, ... n
- Each input modulates a subcarrier of frequency fi; i=1, 2, ... n
- Signals are summed to produce a composite baseband signal denoted as $m_b(t)$
- f_i is chosen such that there is no overlap.



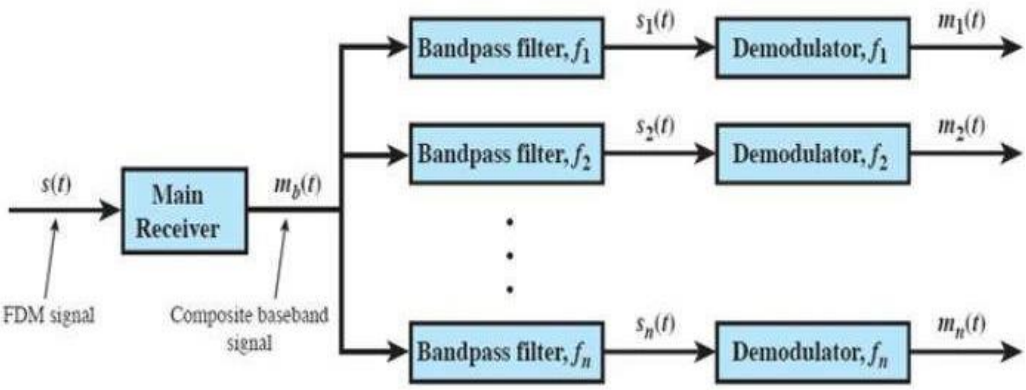
(a) Transmitter

Spectrum of composite baseband modulating signal



FDM system Receiver

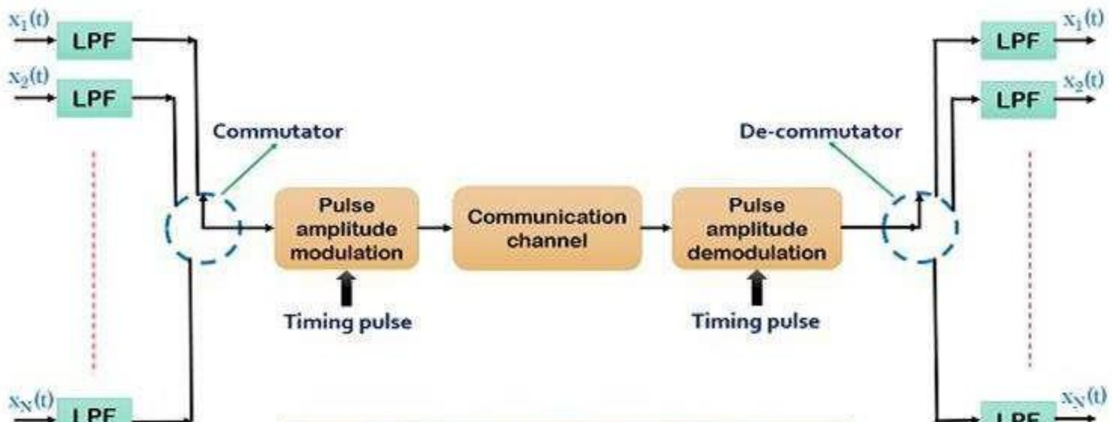
- The Composite base band signal $m_b(t)$ is passed through n band pass filters with response centred on f_i
- Each $s_i(t)$ component is demodulated to recover the original analog/digital data.



Time Division Multiplexing

TDM technique combines time-domain samples from different message signals (sampled at same rate) and transmits them together across the same channel.

The multiplexing is performed using a commutator (switch). At the receiver a decommutator (switch) is used in synchronism with the commutator to demultiplex the data.



Block diagram of PAM/TDM system

The input signals, all band limited to fm (max) by the LPFs are sequentially sampled at the transmitter by a commutator.

The Switch makes one complete revolution in Ts,(1/fs) extracting one sample from each input. Hence the output is a PAM waveform containing the individual message sampled periodically interlaced in time.

A set of pulses consisting of one sample from each input signal is called a frame.

At the receiver the de-commutator separates the samples and distributes them to a bank of LPFs, which in turn reconstruct the original messages.

Synchronizing is provided to keep the de-commutator in step with the commutator.

Elements of Digital Communication Systems

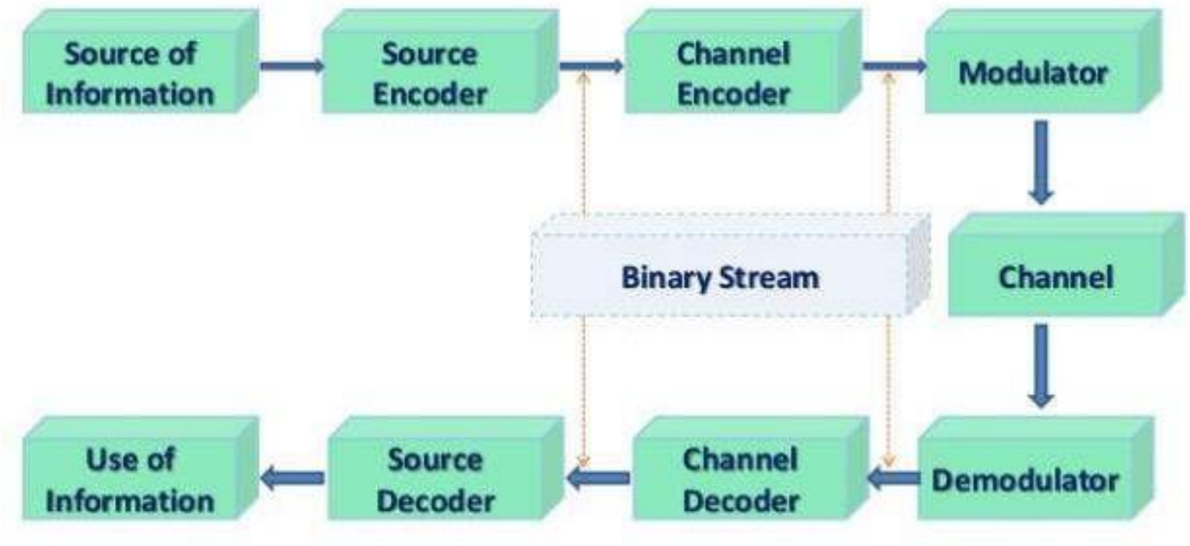


Figure Elements of Digital Communication Systems

1. **Information Source and Input Transducer:**

   The source of information can be analog or digital, e.g. analog: audio or video signal, digital: like teletype signal. In digital communication the signal produced by this source is converted into digital signal which consists of 1′s and 0′s. For this we need a source encoder.

2. **Source Encoder:**

   In digital communication we convert the signal from source into digital signal as mentioned above. The point to remember is we should like to use as few binary digits as possible to represent the signal. In such a way this efficient representation of the source output results in little or no redundancy. This sequence of binary digits is called ***information sequence.***

   Source Encoding or Data Compression: the process of efficiently converting the output of whether analog or digital source into a sequence of binary digits is known as source encoding.

3. **Channel Encoder:**

   The information sequence is passed through the channel encoder. The purpose of the channel encoder is to introduce, in controlled manner, some redundancy in the binary information sequence that can be used at the receiver to overcome the effects of noise and interference encountered in the transmission on the signal through the channel.

   For example take k bits of the information sequence and map that k bits to unique n bit sequence called code word. The amount of redundancy introduced is measured by the ratio *n/k* and the reciprocal of this ratio (k/n) is known as *rate of code or code rate.*

4. **Digital Modulator:**

   The binary sequence is passed to digital modulator which in turns convert the sequence into electric signals so that we can transmit them on channel (we will see channel later). The digital modulator maps the binary sequences into signal wave forms , for example if we represent 1 by sin x and 0 by cos x then we will transmit sin x for 1 and cos x for 0. (a case similar to BPSK)

5. **Channel:**

   The communication channel is the physical medium that is used for transmitting signals from transmitter to receiver. In wireless system, this channel consists of atmosphere , for traditional telephony, this channel is wired , there are optical channels, under water acoustic channels etc.We further discriminate this channels on the basis of their property and characteristics, like AWGN channel etc.

6. **Digital Demodulator:**

   The digital demodulator processes the channel corrupted transmitted waveform and reduces the waveform to the sequence of numbers that represents estimates of the transmitted data symbols.

7. **Channel Decoder:**

   This sequence of numbers then passed through the channel decoder which attempts to reconstruct the original information sequence from the knowledge of the code used by the channel encoder and the redundancy contained in the received data

Note: The average probability of a bit error at the output of the decoder is a measure of the performance of the demodulator – decoder combination.

8. **Source Decoder:**

   At the end, if an analog signal is desired then source decoder tries to decode the sequence from the knowledge of the encoding algorithm. And which results in the approximate replica of the input at the transmitter end.

9. **Output Transducer:**

   Finally we get the desired signal in desired format analog or digital.

Advantages of digital communication

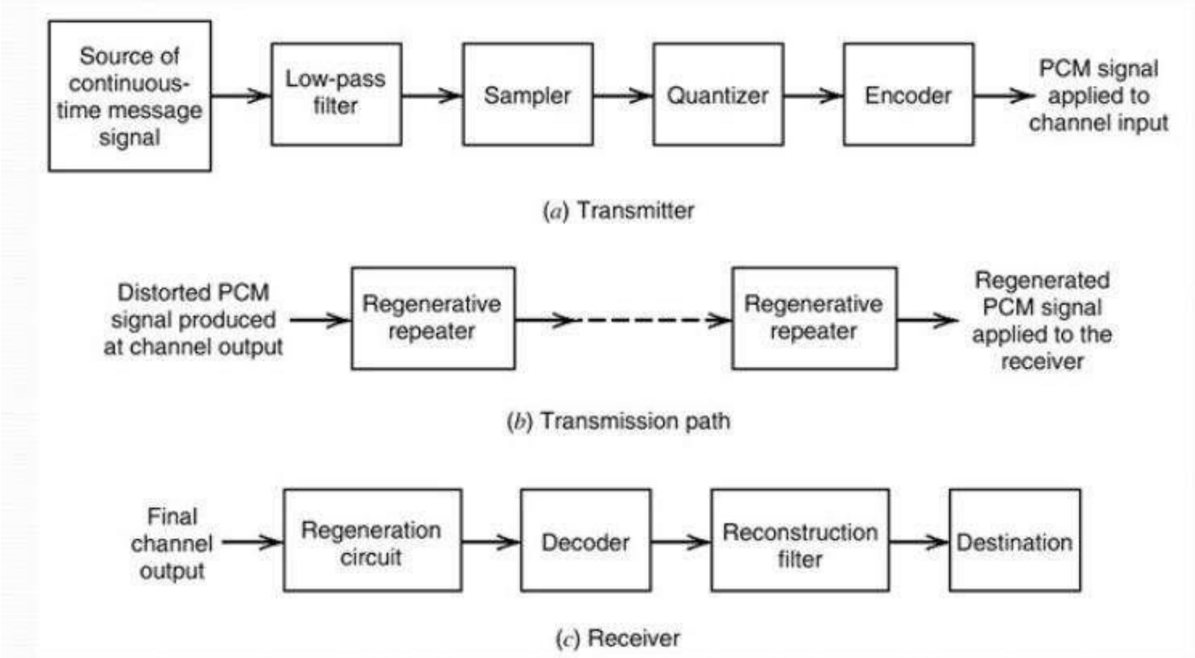
- Can **withstand channel noise and distortion** much better as long as the noise and the distortion are within limits.
- **Regenerative repeaters** prevent accumulation of noise along the path.
- Digital **hardware implementation is flexible.**
- Digital signals **can be coded** to yield extremely **low error rates**, **high fidelity** and well as **privacy**.
- Digital communication is inherently more efficient than analog in realizing the exchange of SNR for bandwidth.
- It is easier and more **efficient to multiplex** several digital signals.

- Digital signal **storage is relatively easy** and **inexpensive.**
- **Reproduction** with digital messages is extremely reliable **without deterioation.**
- The **cost** of digital hardware continues to halve every two or three years while **performance or capacity doubles** over the same time period.

Disadvantages

- TDM digital transmission is not compatible with FDM
- A Digital system requires large bandwidth.

Elements of PCM System



Sampling:

- Process of converting analog signal into discrete signal.
- Sampling is common in all pulse modulation techniques
- The signal is sampled at regular intervals such that each sample is proportional to amplitude of signal at that instant
- Analog signal is sampled every $T_s\ Secs$, called sampling interval. f_s=1/T_S is called sampling rate or sampling frequency.
- f_s=2f_m is Min. sampling rate called **Nyquist rate.** Sampled spectrum (ω) is repeating periodically without overlapping.
- Original spectrum is centered at ω=0 and having bandwidth of ω_m. Spectrum can be recovered by passing through low pass filter with cut-off ω_m.
- For f_s<2f_m sampled spectrum will overlap and cannot be recovered back. This is called **aliasing.**

Sampling methods:

- Ideal – An impulse at each sampling instant.
- Natural – A pulse of Short width with varying amplitude.
- Flat Top – Uses sample and hold, like natural but with single amplitude value.

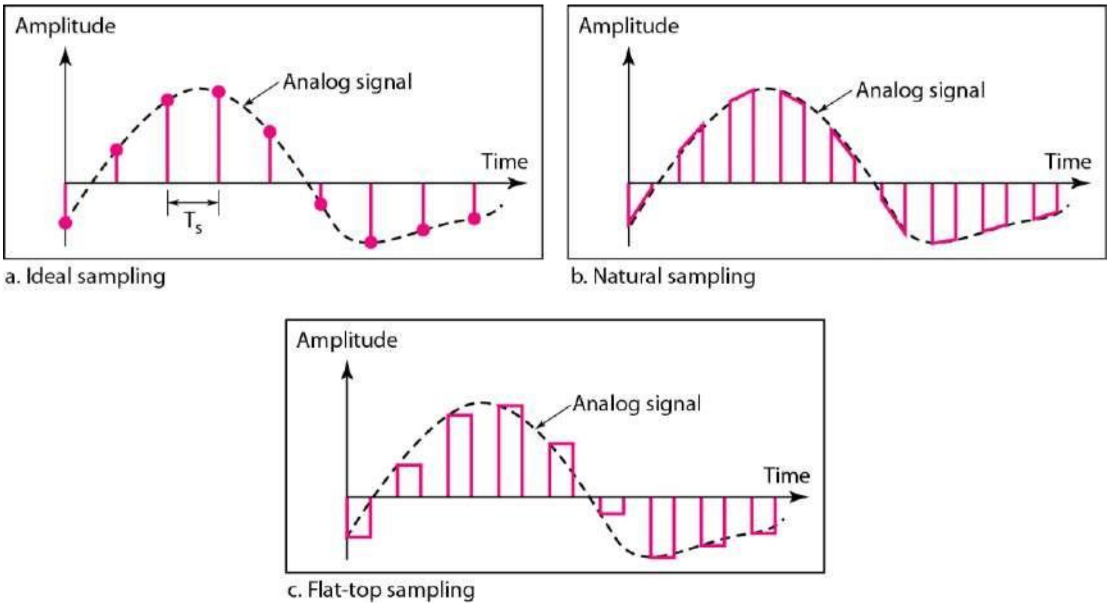


Fig. 4 Types of Sampling

Sampling of band-pass Signals:

- A band-pass signal of bandwidth $2f_m$ can be completely recovered from its samples.

  Min. sampling rate $=2\times Bandwidth$

  $$=2\times 2f_m=4f_m$$

- Range of minimum sampling frequencies is in the range of $2\times BW\ to\ 4\times BW$

Instantaneous Sampling or Impulse Sampling:

- Sampling function is train of spectrum remains constant impulses throughout frequency range. It is not practical.

Natural sampling:

- The spectrum is weighted by a **sinc** function.
- Amplitude of high frequency components reduces.

Flat top sampling:

- Here top of the samples remains constant.
- In the spectrum high frequency components are attenuated due sinc pulse roll off. This is known as **Aperture effect.**
- If pulse width increases aperture effect is more i.e. more attenuation of high frequency components.

PCM Generator

The pulse code modulator technique samples the input signal $x(t)$ at frequency $f_s \geq 2W$. This sampled 'Variable amplitude' pulse is then digitized by the analog to digital converter. The parallel bits obtained are converted to a serial bit stream. Fig. 8 shows the PCM generator.

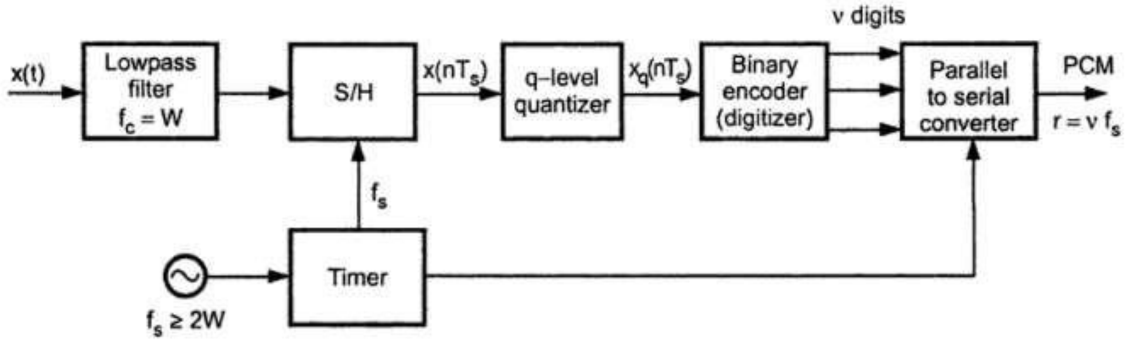


Fig. 8 PCM generator

In the PCM generator of above figure, the signal $x(t)$ is first passed through the lowpass filter of cutoff frequency 'W' Hz. This lowpass filter blocks all the frequency components above 'W' Hz. Thus $x(t)$ is bandlimited to 'W' Hz. The sample and hold circuit then samples this signal at the rate of f_s. Sampling frequency f_s is selected sufficiently above Nyquist rate to avoid aliasing i.e.,

$$f_s \geq 2W$$

In Fig. 8 output of sample and hold is called $x(nT_s)$. This $x(nT_s)$ is discrete in time and continuous in amplitude. A q-level quantizer compares input $x(nT_s)$ with its fixed digital levels. It assigns any one of the digital level to $x(nT_s)$ with its fixed digital levels. It then assigns any one of the digital level to $x(nT_s)$ which results in minimum distortion or error. This error is called *quantization error*. Thus output of quantizer is a digital level called $x_q(nT_s)$.

Now coming back to our discussion of PCM generation, the quantized signal level $x_q(nT_s)$ is given to binary encoder. This encoder converts input signal to 'v' digits binary word. Thus $x_q(nT_s)$ is converted to 'V' binary bits. The encoder is also called digitizer.

It is not possible to transmit each bit of the binary word separately on transmission line. Therefore '*v*' binary digits are converted to serial bit stream to generate single baseband signal. In a parallel to serial converter, normally a shift register does this job. The output of PCM generator is thus a single baseband signal of binary bits.

An oscillator generates the clocks for sample and hold an parallel to serial converter. In the pulse code modulation generator discussed above ; sample and hold, quantizer and encoder combinely form an analog to digital converter.

Transmission BW in PCM

Let the quantizer use 'v' number of binary digits to represent each level. Then the number of levels that can be represented by 'v' digits will be,

$$q = 2^v \quad \ldots \quad 1$$

Here 'q' represents total number of digital levels of q-level quantizer.

For example if $v = 3$ bits, then total number of levels will be,

$$q = 2^3 = 8 \text{ levels}$$

Each sample is converted to 'v' binary bits. i.e. Number of bits per sample = v

We know that, Number of samples per second = f_s

$\therefore$ Number of bits per second is given by,

$$\begin{aligned}(\text{Number of bits per second}) &= (\text{Number of bits per samples}) \\ &\quad \times (\text{Number of samples per second}) \\ &= v \text{ bits per sample} \times f_s \text{ samples per second} \quad \ldots \quad 2\end{aligned}$$

The number of bits per second is also called signaling rate of PCM and is denoted by 'r' i.e.,

$$\boxed{\text{Signaling rate in PCM : } r = v\, f_s} \quad \ldots \quad 3$$

Here $f_s \geq 2W$.

Bandwidth needed for PCM transmission will be given by half of the signaling rate i.e.,

$$\text{Transmission Bandwidth of PCM : } \begin{cases} B_T \geq \frac{1}{2} r & \ldots \quad 4 \\ B_T \geq \frac{1}{2} v\, f_s \quad \text{Since } f_s \geq 2W & \ldots \quad 5 \\ B_T \geq v\, W & \ldots \quad 6 \end{cases}$$

PCM Receiver

Fig. 9 (a) shows the block diagram of PCM receiver and Fig. 9 (b) shows the reconstructed signal. The regenerator at the start of PCM receiver reshapes the pulses and removes the noise. This signal is then converted to parallel digital words for each sample.

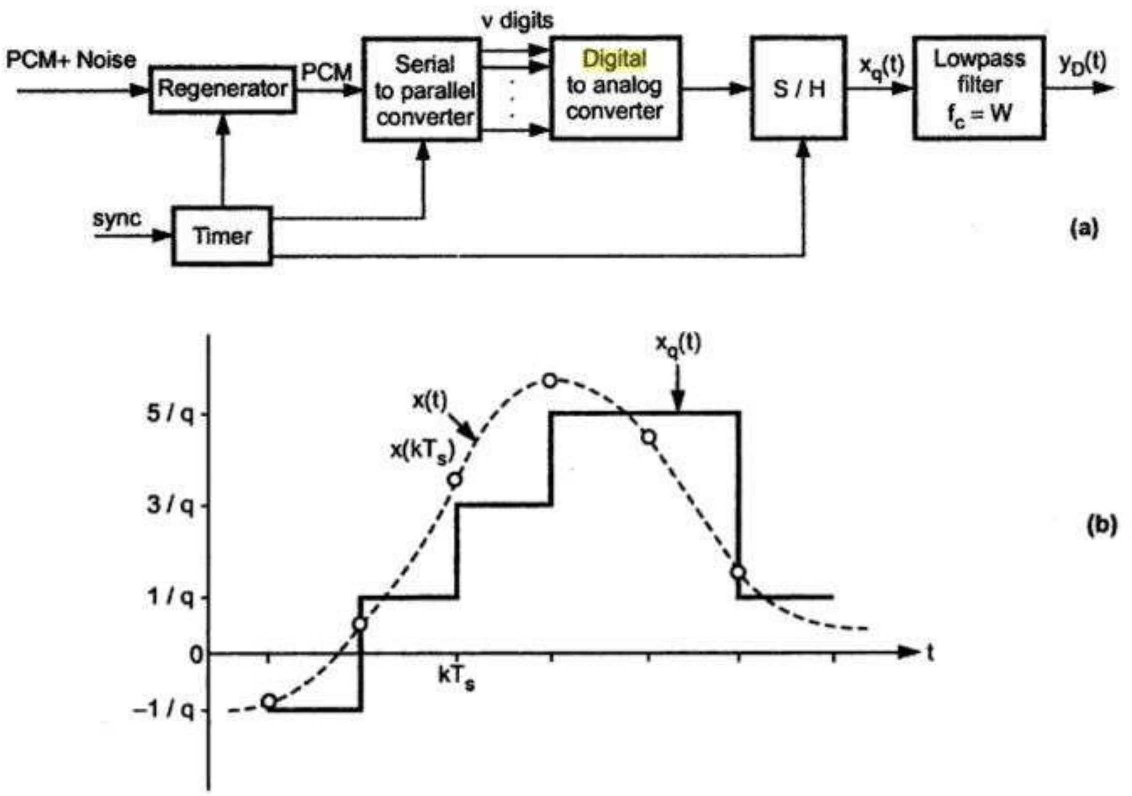


Fig. 9 (a) PCM receiver
(b) Reconstructed waveform

The digital word is converted to its analog value $x_q(t)$ along with sample and hold. This signal, at the output of S/H is passed through lowpass reconstruction filter to get $y_D(t)$. As shown in reconstructed signal of Fig. 9 (b), it is impossible to reconstruct exact original signal $x(t)$ because of permanent quantization error introduced during quantization at the transmitter. This quantization error can be reduced by increasing the binary levels. This is equivalent to increasing binary digits (bits) per sample. But increasing bits 'v' increases the signaling rate as well as transmission bandwidth as we have seen in equation 3 and equation 6. Therefore the choice of these parameters is made, such that noise due to quantization error (called as quantization noise) is in tolerable limits.

<u>Quantization</u>

- 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
- 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**.
- There are two types of Quantization
    - Uniform Quantization
    - 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.**

<u>Uniform Quantization:</u>

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

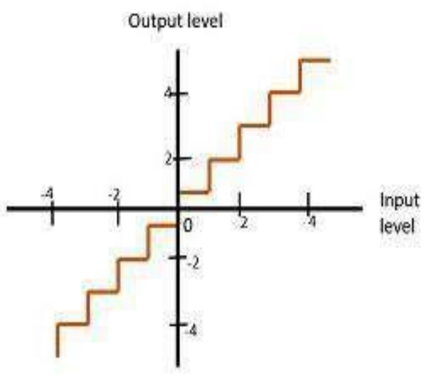


Fig 1 : Mid-Rise type Uniform Quantization

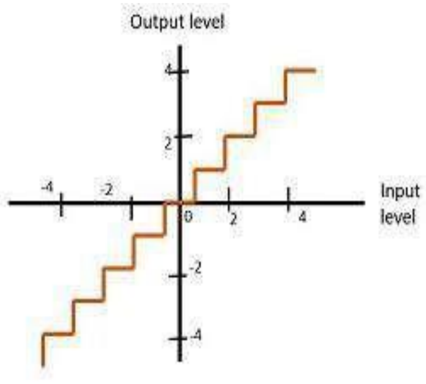


Fig 2 : Mid-Tread type Uniform Quantization

- 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 quantizer is symmetric about the origin.

Quantization Noise and Signal to Noise ratio in PCM System

Derivation of Quantization Error/Noise or Noise Power for Uniform (Linear) Quantization

Step 1 : Quantization Error

Because of quantization, inherent errors are introduced in the signal. This error is called *quantization error*. We have defined quantization error as,

$$\varepsilon = x_q(nT_s) - x(nT_s) \quad \text{.........................} (1)$$

Step 2 : Step size

Let an input $x(nT_s)$ be of continuous amplitude in the range $-x_{max}$ to $+x_{max}$.

Therefore the total amplitude range becomes,

$$\text{Total amplitude range} = x_{max} - (-x_{max}) = 2x_{max} \quad \text{.........................} (2)$$

If this amplitude range is divided into 'q' levels of quantizer, then the step size 'δ' is given as,

$$\delta = \frac{x_{max} - (-x_{max})}{q}$$

$$= \frac{2x_{max}}{q} \quad \text{..................} (3)$$

If signal $x(t)$ is normalized to minimum and maximum values equal to 1, then

$$x_{max} = 1$$
$$-x_{max} = -1 \quad \text{.........................} (4)$$

Therefore step size will be,

$$\delta = \frac{2}{q} \quad \text{(for normalized signal)} \quad \text{.........................} (5)$$

Step 3 : Pdf of Quantization error

If step size 'δ' is sufficiently small, then it is reasonable to assume that the quantization error 'ε' will be uniformly distributed random variable. The maximum quantization error is given by

$$\varepsilon_{max} = \left|\frac{\delta}{2}\right| \quad \text{.........................} (6)$$

i.e. $$-\frac{\delta}{2} \geq \varepsilon_{max} \geq \frac{\delta}{2} \quad \text{.........................} (7)$$

Thus over the interval $\left(-\frac{\delta}{2}, \frac{\delta}{2}\right)$ quantization error is uniformly distributed random variable.

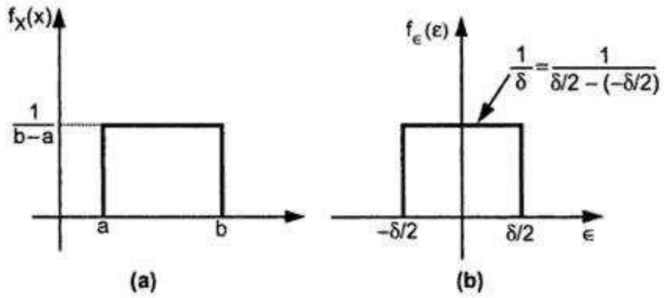


Fig. 10 (a) Uniform distribution
(b) Uniform distribution for quantization error

In above figure, a random variable is said to be uniformly distributed over an interval (a, b). Then PDF of 'X' is given by, (from equation of Uniform PDF).

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x \le a \\ \dfrac{1}{b-a} & \text{for} \quad a < x \le b \\ 0 & \text{for} \quad x > b \end{cases} \qquad \text{..........(8)}$$

Thus with the help of above equation we can define the probability density function for quantization error 'ε' as,

$$f_\varepsilon(\varepsilon) = \begin{cases} 0 & \text{for} \quad \varepsilon \le \dfrac{\delta}{2} \\ \dfrac{1}{\delta} & \text{for} \quad -\dfrac{\delta}{2} < \varepsilon \le \dfrac{\delta}{2} \\ 0 & \text{for} \quad \varepsilon > \dfrac{\delta}{2} \end{cases} \qquad \text{..........(9)}$$

Step 4 : Noise Power

quantization error 'ε' has zero average value. That is mean 'm_ε' of the quantization error is zero.

The signal to quantization noise ratio of the quantizer is defined as,

$$\frac{S}{N} = \frac{\text{Signal power (normalized)}}{\text{Noise power (normalized)}} \qquad \cdots 10$$

If type of signal at input i.e., $x(t)$ is known, then it is possible to calculate signal power.

The noise power is given as,

$$\text{Noise power} = \frac{V_{noise}^2}{R} \qquad \cdots (11)$$

Here V_{noise}^2 is the mean square value of noise voltage. Since noise is defined by random variable 'ε' and PDF $f_\varepsilon(\varepsilon)$, its mean square value is given as,

$$\text{mean square value} = E[\varepsilon^2] = \bar{\varepsilon}^2 \qquad \cdots (12)$$

The mean square value of a random variable 'X' is given as,

$$\bar{X}^2 = E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx \quad \text{By definition} \qquad \cdots (13)$$

Here
$$E[\varepsilon^2] = \int_{-\infty}^{\infty} \varepsilon^2 f_\varepsilon(\varepsilon)\,d\varepsilon \qquad \cdots (14)$$

From equation 9 we can write above equation as,

$$E[\varepsilon^2] = \int_{-\delta/2}^{\delta/2} \varepsilon^2 \times \frac{1}{\delta}\,d\varepsilon$$

$$= \frac{1}{\delta}\left[\frac{\varepsilon^3}{3}\right]_{-\delta/2}^{\delta/2} = \frac{1}{\delta}\left[\frac{(\delta/2)^3}{3} + \frac{(\delta/2)^3}{3}\right]$$

$$= \frac{1}{3\delta}\left[\frac{\delta^3}{8} + \frac{\delta^3}{8}\right] = \frac{\delta^2}{12} \qquad \cdots (15)$$

∴ From equation 1.8.25, the mean square value of noise voltage is,

$$V_{noise}^2 = \text{mean square value} = \frac{\delta^2}{12}$$

When load resistance, $R = 1$ ohm, then the noise power is normalized i.e.,

$$\text{Noise power (normalized)} = \frac{V_{noise}^2}{1} \qquad [\text{with } R = 1 \text{ in equation } 11 \quad]$$

$$= \frac{\delta^2 / 12}{1} = \frac{\delta^2}{12}$$

Thus we have,

Normalized noise power

or Quantization noise power $= \dfrac{\delta^2}{12}$ **; For linear quantization.**

or Quantization error (in terms of power) ... (16)

Derivation of Maximum Signal to Quantization Noise Ratio for Linear Quantization:

signal to quantization noise ratio is given as,

$$\frac{S}{N} = \frac{\text{Normalized signal power}}{\text{Normalized noise power}}$$

$$= \frac{\text{Normalized signal power}}{(\delta^2 / 12)} \qquad \dots (17)$$

The number of bits 'v' and quantization levels 'q' are related as,

$$q = 2^v \qquad \dots (18)$$

Putting this value in equation (3) we have,

$$\delta = \frac{2\, x_{\max}}{2^v} \qquad \dots (19)$$

Putting this value in equation 1.8.30 we get,

$$\frac{S}{N} = \frac{\text{Normalized signal power}}{\left(\frac{2\, x_{\max}}{2^v}\right)^2 + 12}$$

Let normalized signal power be denoted as 'P'.

$$\frac{S}{N} = \frac{P}{\frac{4\, x_{\max}^2}{2^{2v}} \times \frac{1}{12}} = \frac{3P}{x_{\max}^2} \cdot 2^{2v}$$

This is the required relation for maximum signal to quantization noise ratio. Thus,

$$\textbf{Maximum signal to quantization noise ratio : } \frac{S}{N} = \frac{3P}{x_{max}^2} \cdot 2^{2v} \qquad \cdots (20)$$

This equation shows that signal to noise power ratio of quantizer increases exponentially with increasing bits per sample.

If we assume that input $x(t)$ is normalized, i.e.,

$$x_{max} = 1 \qquad \cdots (21)$$

Then signal to quantization noise ratio will be,

$$\frac{S}{N} = 3 \times 2^{2v} \times P \qquad \cdots (22)$$

If the destination signal power 'P' is normalized, i.e.,

$$P \leq 1 \qquad \cdots (23)$$

Then the signal to noise ratio is given as,

$$\frac{S}{N} \leq 3 \times 2^{2v} \qquad \cdots (24)$$

Since $x_{max} = 1$ and $P \leq 1$, the signal to noise ratio given by above equation is normalized.

Expressing the signal to noise ratio in decibels,

$$\left(\frac{S}{N}\right) dB = 10 \log_{10} \left(\frac{S}{N}\right) dB \quad \text{since power ratio.}$$

$$\leq 10 \log_{10} [3 \times 2^{2v}]$$

$$\leq (4.8 + 6v)\, dB$$

Thus,

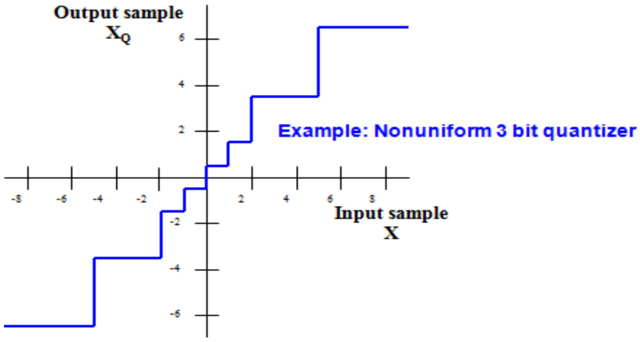
Signal to Quantization noise ratio

for normalized values of power $: \left(\frac{S}{N}\right) dB \leq (4.8 + 6v)\, dB$

'P' and amplitude of input $x(t)$ $\cdots (25)$

Non-Uniform Quantization:

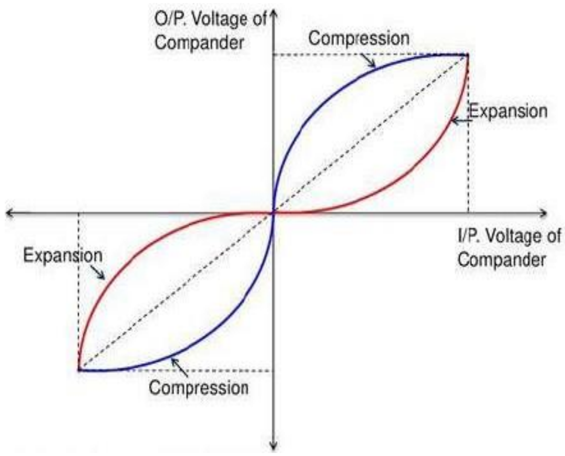
In non-uniform quantization, the step size is not fixed. It varies according to certain law or as per input signal amplitude. The following fig shows the characteristics of Non uniform quantizer.

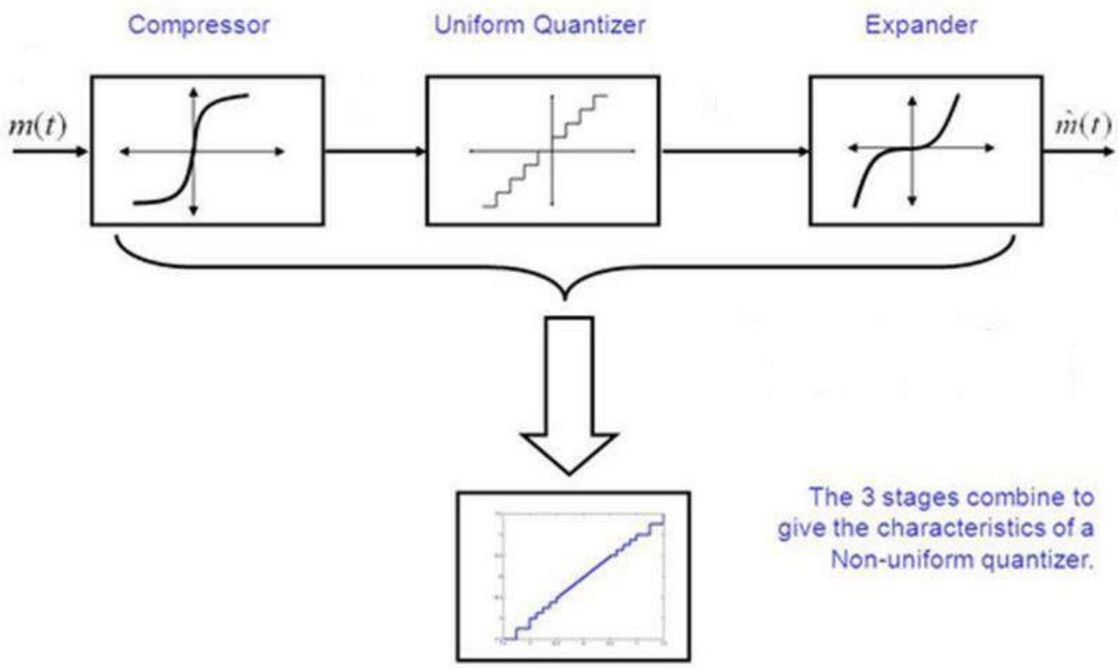


In this figure observe that step size is small at low input signal levels. Hence quantization error is also small at these inputs. Therefore signal to quantization noise power ratio is improved at low signal levels. Stepsize is higher at high input levels. Hence signal to noise power ratio remains almost same throughout the dynamic range of quantizer.

Companding PCM System

- Non-uniform quantizers are difficult to make and expensive.
- An alternative is to first pass the speech signal through nonlinearity before quantizing with a uniform quantizer.
- The nonlinearity causes the signal amplitude to be ***compressed***.
  - The input to the quantizer will have a more uniform distribution.
- At the receiver, the signal is ***expanded*** by an inverse to the nonlinearity.
- The process of compressing and expanding is called ***Companding***.





μ - Law Companding for Speech Signals

Normally for speech and music signals a μ - law compression is used. This compression is defined by the following equation,

$$Z(x) = (\mathrm{Sgn}\; x)\frac{ln(1+\mu|x|)}{ln(1+\mu)} \quad |x| \le 1 \qquad \cdots (1)$$

Below Fig shows the variation of signal to noise ratio with respect to signal level without companding and with companding.

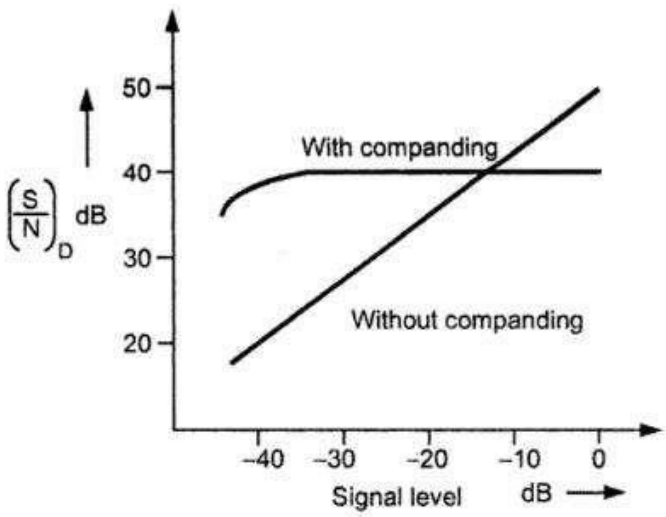


Fig. 11 PCM performance with μ - law companding

It can be observed from above figure that signal to noise ratio of PCM remains almost constant with companding.

A-Law for Companding

The A law provides piecewise compressor characteristic. It has linear segment for low level inputs and logarithmic segment for high level inputs. It is defined as,

$$Z(x) = \begin{cases} \dfrac{A|x|}{1+\ln A} & \text{for } 0 \le |x| \le \dfrac{1}{A} \\ \dfrac{1+\ln(A|x|)}{1+\ln A} & \text{for } \dfrac{1}{A} \le |x| \le 1 \end{cases} \quad \cdots (2)$$

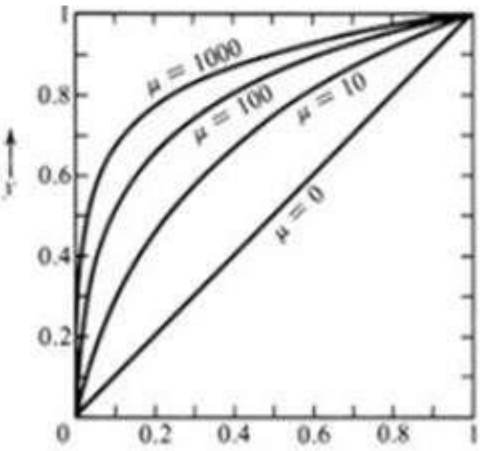
When A = 1, we get uniform quantization. The practical value for A is 87.56. Both A-law and μ-law companding is used for PCM telephone systems.

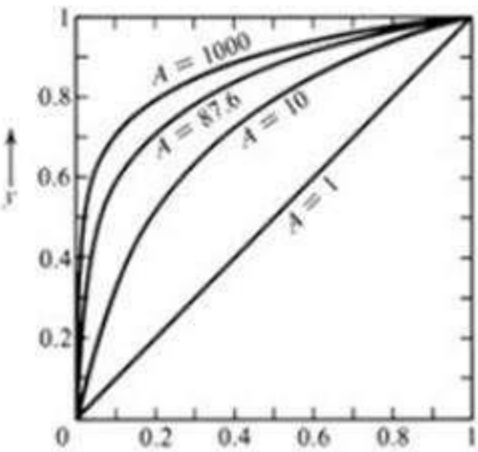
Signal to Noise Ratio of Companded PCM

The signal to noise ratio of companded PCM is given as,

$$\frac{S}{N} = \frac{3q^2}{[\ln(1+\mu)]^2} \quad \cdots (3)$$

Here $q = 2^v$ is number of quantization levels.





Differential Pulse Code Modulation (DPCM)

Redundant Information in PCM

The samples of a signal are highly corrected with each other. This is because any signal does not change fast. That is its value from present sample to next sample does not differ by large amount. The adjacent samples of the signal carry the same information with little difference. When these samples are encoded by standard PCM system, the resulting encoded signal contains redundant information.

Fig. shows a continuous time signal $x(t)$ by dotted line. This signal is sampled by flat top sampling at intervals $T_s, 2T_s, 3T_s \ldots nT_s$. The sampling frequency is selected to be higher than nyquist rate. The samples are encoded by using 3 bit (7 levels) PCM. The sample is quantized to the nearest digital level as shown by small

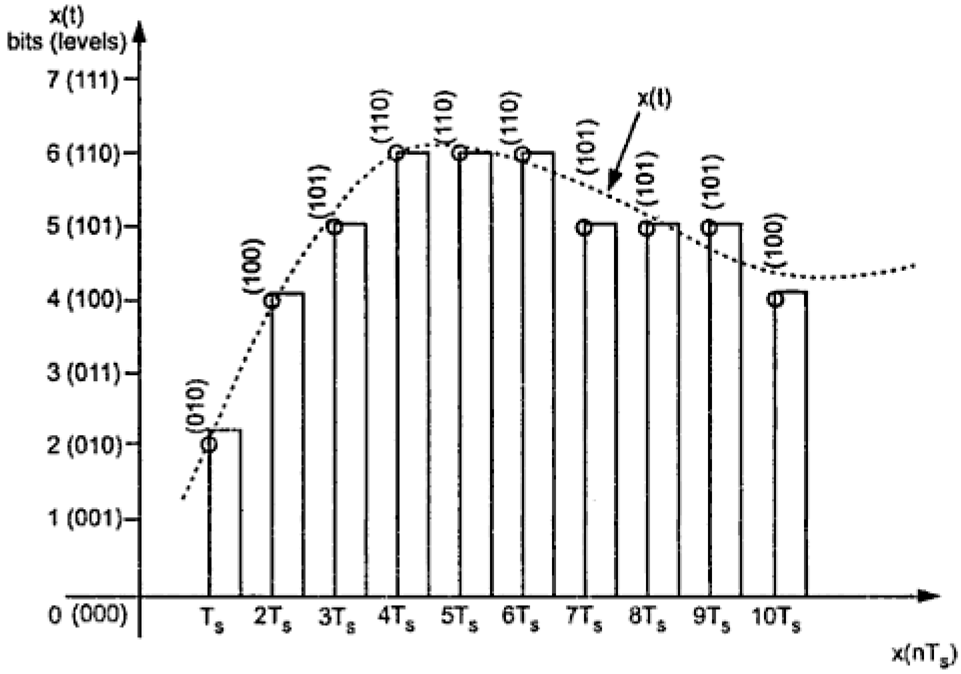


Fig. Redundant information in PCM

circles in the diagram. The encoded binary value of each sample is written on the top of the samples. We can see from Fig. that the samples taken at $4T_s, 5T_s$ and $6T_s$ are encoded to same value of (110). This information can be carried only by one sample. But three samples are carrying the same information means it is redundant. Consider another example of samples taken at $9T_s$ and $10T_s$. The difference between these samples is only due to last bit and first two bits are redundant, since they do not change.

Principle of DPCM

If this redundancy is reduced, then overall bit rate will decrease and number of bits required to transmit one sample will also be reduced. This type of digital pulse modulation scheme is called Differential Pulse Code Modulation.

DPCM Transmitter

The differential pulse code modulation works on the principle of prediction. The value of the present sample is predicted from the past samples. The prediction may not be exact but it is very close to the actual sample value. Fig. shows the transmitter of Differential Pulse Code Modulation (DPCM) system. The sampled signal is denoted by $x(nT_s)$ and the predicted signal is denoted by $\hat{x}(nT_s)$. The comparator finds out the difference between the actual sample value $x(nT_s)$ and predicted sample value $\hat{x}(nT_s)$. This is called error and it is denoted by $e(nT_s)$. It can be defined as,

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s) \qquad \text{.................}(1)$$

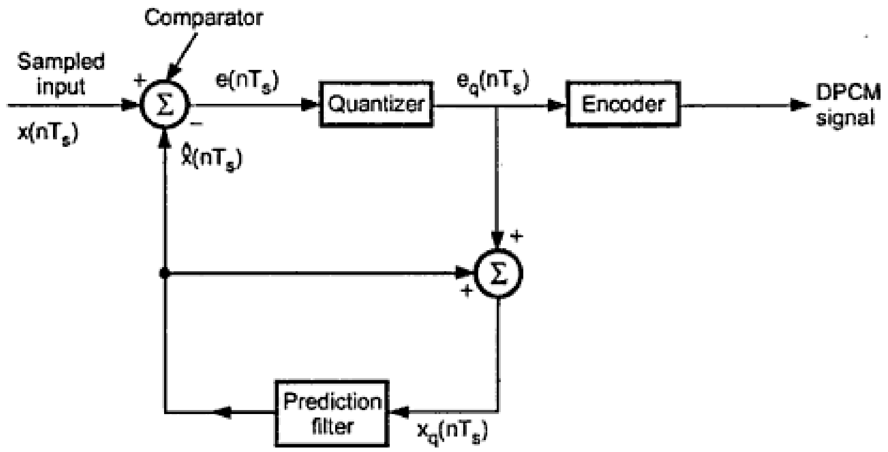


Fig. Differential pulse code modulation transmitter

Thus error is the difference between unquantized input sample $x(nT_s)$ and prediction of it $\hat{x}(nT_s)$. The predicted value is produced by using a prediction filter. The quantizer output signal $e_q(nT_s)$ and previous prediction is added and given as input to the prediction filter. This signal is called $x_q(nT_s)$. This makes the prediction more and more close to the actual sampled signal. We can see that the quantized error signal $e_q(nT_s)$ is very small and can be encoded by using small number of bits. Thus number of bits per sample are reduced in DPCM.

The quantizer output can be written as,

$$e_q(nT_s) = e(nT_s) + q(nT_s) \qquad \text{.................(2)}$$

Here $q(nT_s)$ is the quantization error. As shown in Fig. the prediction filter input $x_q(nT_s)$ is obtained by sum $\hat{x}(nT_s)$ and quantizer output i.e.,

$$x_q(nT_s) = \hat{x}(nT_s) + e_q(nT_s) \qquad \text{...................(3)}$$

Putting the value of $e_q(nT_s)$ from equation 2 in the above equation we get,

$$x_q(nT_s) = \hat{x}(nT_s) + e(nT_s) + q(nT_s) \qquad \text{................(4)}$$

Equation 1 is written as,

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s)$$

$$\therefore \quad e(nT_s) + \hat{x}(nT_s) = x(nT_s) \qquad \text{...............(5)}$$

$\therefore$ Putting the value of $e(nT_s) + \hat{x}(nT_s)$ from above equation into equation 4 we get,

$$x_q(nT_s) = x(nT_s) + q(nT_s) \qquad \text{...................(6)}$$

Thus the quantized version of the signal $x_q(nT_s)$ is the sum of original sample value and quantization error $q(nT_s)$. The quantization error can be positive or negative. Thus equation 6 does not depend on the prediction filter characteristics.

Reconstruction of DPCM Signal

Fig. shows the block diagram of DPCM receiver.

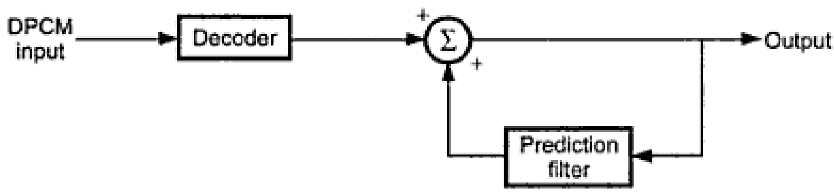


Fig. DPCM receiver

The decoder first reconstructs the quantized error signal from incoming binary signal. The prediction filter output and quantized error signals are summed up to give the quantized version of the original signal. Thus the signal at the receiver differs from actual signal by quantization error $q(nT_s)$, which is introduced permanently in the reconstructed signal.

Introduction to Delta Modulation

PCM transmits all the bits which are used to code the sample. Hence signaling rate and transmission channel bandwidth are large in PCM. To overcome this problem Delta Modulation is used.

Delta Modulation

Operating Principle of DM

Delta modulation transmits only one bit per sample. That is the present sample value is compared with the previous sample value and the indication,whether the amplitude is increased or decreased is sent. Input signal $x(t)$ is approximated to step signal by the delta modulator. This step size is fixed. The difference between the input signal $x(t)$ and staircase approximated signal confined to two levels, i.e. $+\delta$ and $-\delta$. If the difference is positive, then approximated signal is increased by one step i.e. 'δ'. If the difference is negative, then approximated signal is reduced by 'δ'. When the step is reduced, '0' is transmitted and if the step is increased, '1' is transmitted. Thus for each sample, only one binary bit is transmitted. Fig. shows the analog signal $x(t)$ and its staircase approximated signal by the delta modulator.

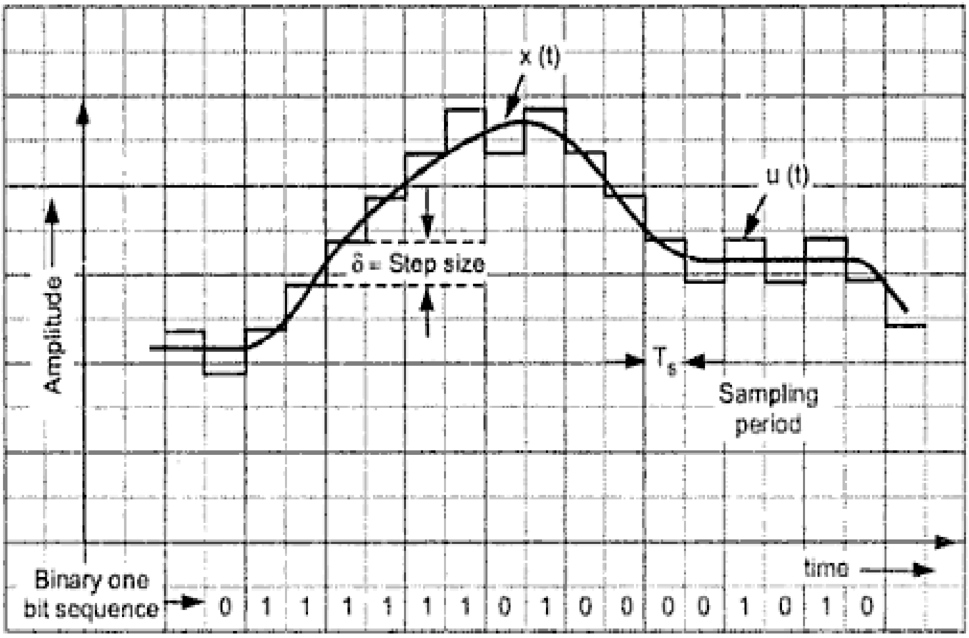


Fig. Delta modulation waveform

The principle of delta modulation can be explained by the following set of equations. The error between the sampled value of $x(t)$ and last approximated sample is given as,

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s) \quad \dots (\quad 1)$$

Here, $e(nT_s)$ = Error at present sample

$x(nT_s)$ = Sampled signal of $x(t)$

$\hat{x}(nT_s)$ = Last sample approximation of the staircase waveform.

We can call $u(nT_s)$ as the present sample approximation of staircase output.

Then, $u[(n-1)T_s] = \hat{x}(nT_s) \quad \dots (\quad 2)$

= Last sample approximation of staircase waveform.

Let the quantity $b(nT_s)$ be defined as,

$$b(nT_s) = \delta \, sgn[e(nT_s)] \quad \dots (\quad 3)$$

That is depending on the sign of error $e(nT_s)$ the sign of step size δ will be decided. In other words,

$$\begin{aligned} b(nT_s) &= +\delta \quad \text{if} \quad x(nT_s) \geq \hat{x}(nT_s) \\ &= -\delta \quad \text{if} \quad x(nT_s) < \hat{x}(nT_s) \end{aligned} \quad \dots (\quad 4)$$

If $b(nT_s) = +\delta$; binary '1' is transmitted

and if $b(nT_s) = -\delta$; binary '0' is transmitted.

T_s = Sampling interval.

DM Transmitter

Fig. (a) shows the transmitter based on equations 3 to 5.

The summer in the accumulator adds quantizer output ($\pm\delta$) with the previous sample approximation. This gives present sample approximation. i.e.,

$$\begin{aligned} u(nT_s) &= u(nT_s - T_s) + [\pm\delta] \quad \text{or} \\ &= u[(n-1)T_s] + b(nT_s) \end{aligned} \quad \dots (\quad 5)$$

The previous sample approximation $u[(n-1)T_s]$ is restored by delaying one sample period T_s. The sampled input signal $x(nT_s)$ and staircase approximated signal $\hat{x}(nT_s)$ are subtracted to get error signal $e(nT_s)$.

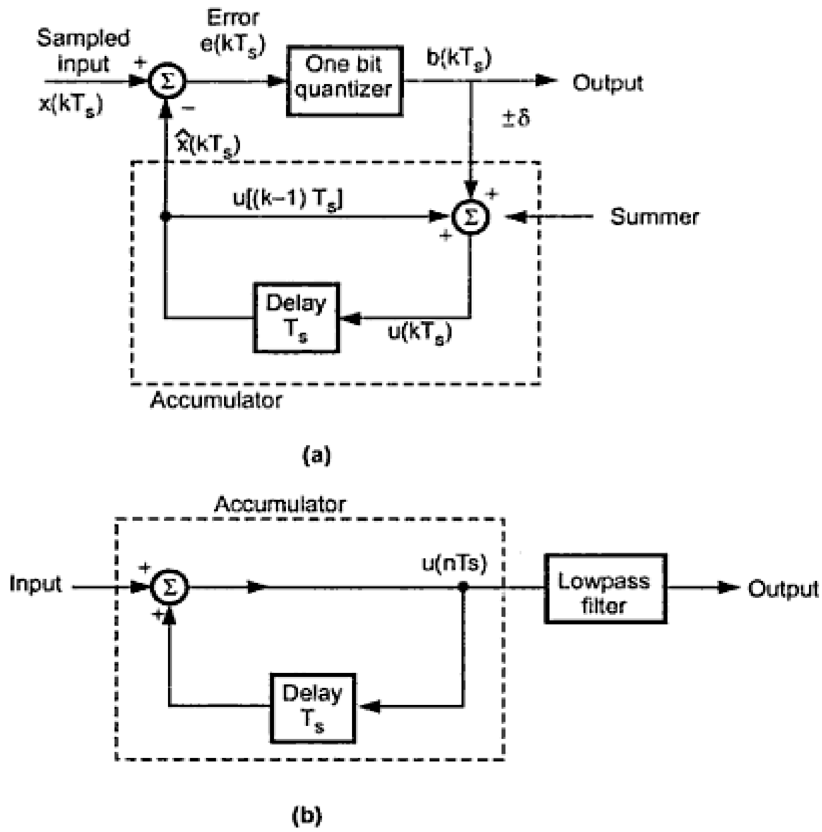


Fig. (a) Delta modulation transmitter and
(b) Delta modulation receiver

Depending on the sign of $e(nT_s)$ one bit quantizer produces an output step of $+\delta$ or $-\delta$. If the step size is $+\delta$, then binary '1' is transmitted and if it is $-\delta$, then binary '0' is transmitted.

DM Receiver

At the receiver shown in Fig. (b), the accumulator and low-pass filter are used. The accumulator generates the staircase approximated signal output and is delayed by one sampling period T_s. It is then added to the input signal. If input is binary '1' then it adds $+\delta$ step to the previous output (which is delayed). If input is binary '0' then one step 'δ' is subtracted from the delayed signal. The low-pass filter has the cutoff frequency equal to highest frequency in $x(t)$. This filter smoothen the staircase signal to reconstruct $x(t)$.

Advantages and Disadvantages of Delta Modulation

Advantages of Delta Modulation

The delta modulation has following advantages over PCM,

1. Delta modulation transmits only one bit for one sample. Thus the signaling rate and transmission channel bandwidth is quite small for delta modulation.
2. The transmitter and receiver implementation is very much simple for delta modulation. There is no analog to digital converter involved in delta modulation.

Disadvantages of Delta Modulation

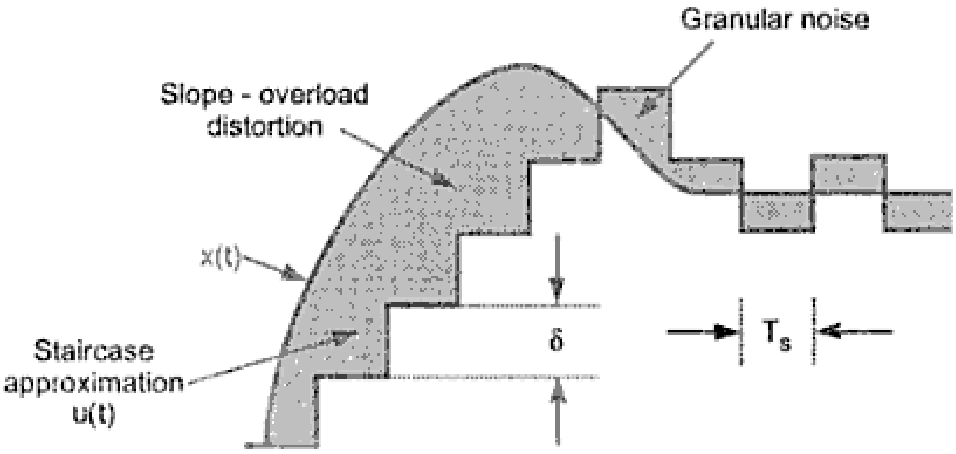


Fig. Quantization errors in delta modulation

The delta modulation has two drawbacks -

Slope Overload Distortion (Startup Error)

This distortion arises because of the large dynamic range of the input signal.

As can be seen from Fig. the rate of rise of input signal $x(t)$ is so high that the staircase signal cannot approximate it, the step size 'δ' becomes too small for staircase signal $u(t)$ to follow the steep segment of $x(t)$. Thus there is a large error between the staircase approximated signal and the original input signal $x(t)$. This error is called *slope overload distortion*. To reduce this error, the step size should be increased when slope of signal of $x(t)$ is high.

Since the step size of delta modulator remains fixed, its maximum or minimum slopes occur along straight lines. Therefore this modulator is also called Linear Delta Modulator (LDM).

Granular Noise (Hunting)

Granular noise occurs when the step size is too large compared to small variations in the input signal. That is for very small variations in the input signal, the staircase signal is changed by large amount (δ) because of large step size. Fig. shows that when the input signal is almost flat, the staircase signal $u(t)$ keeps on oscillating by $\pm\delta$ around the signal. The error between the input and approximated signal is called *granular noise.* The solution to this problem is to make step size small.

Thus large step size is required to accommodate wide dynamic range of the input signal (to reduce slope overload distortion) and small steps are required to reduce granular noise. Adaptive delta modulation is the modification to overcome these errors.

Adaptive Delta Modulation

Operating Principle

To overcome the quantization errors due to slope overload and granular noise, the step size (δ) is made adaptive to variations in the input signal $x(t)$. Particularly in the steep segment of the signal $x(t)$, the step size is increased. When the input is varying slowly, the step size is reduced. Then the method is called *Adaptive Delta Modulation (ADM).*

The adaptive delta modulators can take continuous changes in step size or discrete changes in step size.

Transmitter and Receiver

Fig. (a) shows the transmitter and (b) shows receiver of adaptive delta modulator. The logic for step size control is added in the diagram. The step size increases or decreases according to certain rule depending on one bit quantizer output.

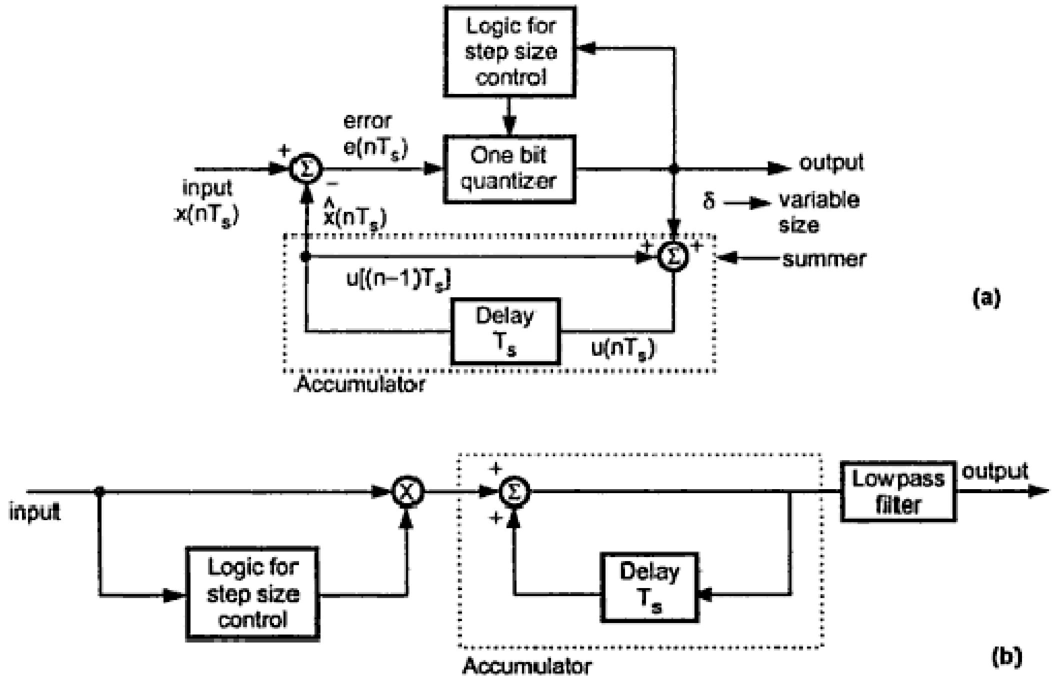


Fig. Adaptive delta modulator (a) Transmitter (b) Receiver

For example if one bit quantizer output is high (1), then step size may be doubled for next sample. If one bit quantizer output is low, then step size may be reduced by one step. Fig. shows the waveforms of adaptive delta modulator and sequence of bits transmitted.

In the receiver of adaptive delta modulator shown in Fig. (b) the first part generates the step size from each incoming bit. Exactly the same process is followed as that in transmitter. The previous input and present input decides the step size. It is then given to an accumulator which builds up staircase waveform. The low-pass filter then smoothens out the staircase waveform to reconstruct the smooth signal.

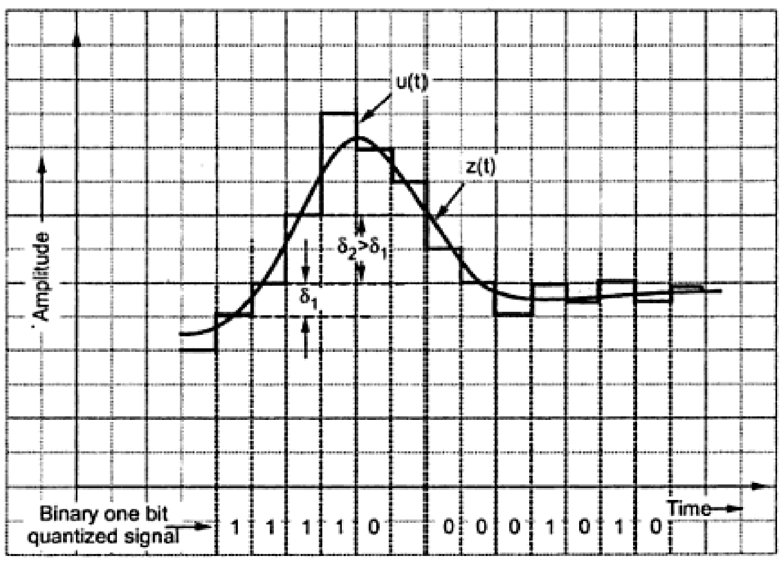


Fig. Waveforms of adaptive delta modulation

Advantages of Adaptive Delta Modulation

Adaptive delta modulation has certain advantages over delta modulation. i.e.,

1. The signal to noise ratio is better than ordinary delta modulation because of the reduction in slope overload distortion and granular noise.
2. Because of the variable step size, the dynamic range of ADM is wide.
3. Utilization of bandwidth is better than delta modulation.

Plus other advantages of delta modulation are, only one bit per sample is required and simplicity of implementation of transmitter and receiver.

Condition for Slope overload distortion occurrence

Slope overload distortion will occur if

$$A_m > \frac{\delta}{2\pi f_m T_s}$$

where T_s is the sampling period.

Let the sine wave be represented as,

$$x(t) = A_m \sin(2\pi f_m t)$$

Slope of $x(t)$ will be maximum when derivative of $x(t)$ with respect to 't' will be maximum. The maximum slope of delta modulator is given

$$\text{Max. slope} = \frac{\text{Step size}}{\text{Sampling period}}$$

$$= \frac{\delta}{T_s} \qquad \text{..............(1)}$$

Slope overload distortion will take place if slope of sine wave is greater than slope of delta modulator i.e.

$$\max\left|\frac{d}{dt}x(t)\right| > \frac{\delta}{T_s}$$

$$\max\left|\frac{d}{dt}A_m \sin(2\pi f_m t)\right| > \frac{\delta}{T_s}$$

$$\max\left|A_m\, 2\pi f_m \cos(2\pi f_m t)\right| > \frac{\delta}{T_s}$$

$$A_m\, 2\pi f_m > \frac{\delta}{T_s}$$

$$\text{or} \qquad \boxed{A_m > \frac{\delta}{2\pi f_m T_s}} \qquad \text{...............(2)}$$

Expression for Signal to Quantization Noise power ratio for Delta Modulation

To obtain signal power :

slope overload distortion will not occur if

$$A_m \le \frac{\delta}{2\pi f_m T_s}$$

Here A_m is peak amplitude of sinusoided signal

δ is the step size

f_m is the signal frequency and

T_s is the sampling period.

From above equation, the maximum signal amplitude will be,

$$A_m = \frac{\delta}{2\pi f_m T_s} \quad ....(1)$$

Signal power is given as,

$$P = \frac{V^2}{R}$$

Here V is the rms value of the signal. Here $V = \frac{A_m}{\sqrt{2}}$. Hence above equation becomes,

$$P = \left(\frac{A_m}{\sqrt{2}}\right)^2 / R$$

Normalized signal power is obtained by taking R = 1. Hence,

$$P = \frac{A_m^2}{2}$$

Putting for A_m from equation 1

$$P = \frac{\delta^2}{8\pi^2 f_m^2 T_s^2} \quad ....(2)$$

This is an expression for signal power in delta modulation.

(ii) To obtain noise power

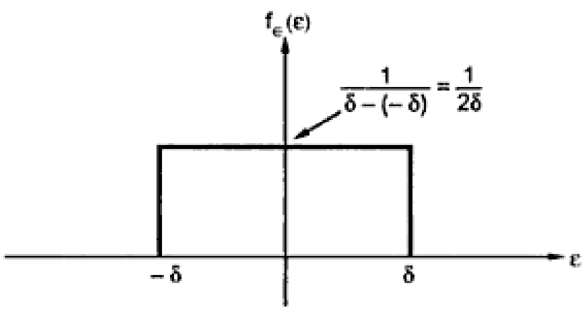


Fig. Uniform distribution of quantization error

We know that the maximum quantization error in delta modulation is equal to step size 'δ'. Let the quantization error be uniformly distributed over an interval $[-\delta, \delta]$. This is shown in Fig. From this figure the PDF of quantization error can be expressed as,

$$f_{\epsilon}(\varepsilon) = \begin{cases} 0 & \textit{for} \quad \varepsilon < \delta \\ \dfrac{1}{2\delta} & \textit{for} \quad -\delta < \varepsilon < \delta \\ 0 & \textit{for} \quad \varepsilon > \delta \end{cases} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

The noise power is given as,

$$\text{Noise power} = \frac{V_{noise}^2}{R}$$

Here V_{noise}^2 is the mean square value of noise voltage. Since noise is defined by random variable 'ε' and PDF $f_{\epsilon}(\varepsilon)$, its mean square value is given as,

$$\text{mean square value} = E[\varepsilon^2] = \overline{\varepsilon^2}$$

mean square value is given as,

$$E[\varepsilon^2] = \int_{-\infty}^{\infty} \varepsilon^2 f_{\epsilon}(\varepsilon)\, d\varepsilon$$

From equation 3

$$E[\varepsilon^2] = \int_{-\delta}^{\delta} \varepsilon^2 \cdot \frac{1}{2\delta}\, d\varepsilon$$

$$= \frac{1}{2\delta}\left[\frac{\varepsilon^3}{3}\right]_{-\delta}^{\delta}$$

$$= \frac{1}{2\delta}\left[\frac{\delta^3}{3} + \frac{\delta^3}{3}\right] = \frac{\delta^2}{3} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Hence noise power will be,

$$\text{noise power} = \left(\frac{\delta^2}{3}\right) / R$$

Normalized noise power can be obtained with R = 1. Hence,

$$\text{noise power} = \frac{\delta^2}{3} \qquad \ldots\ldots\ldots\ldots\ldots(5)$$

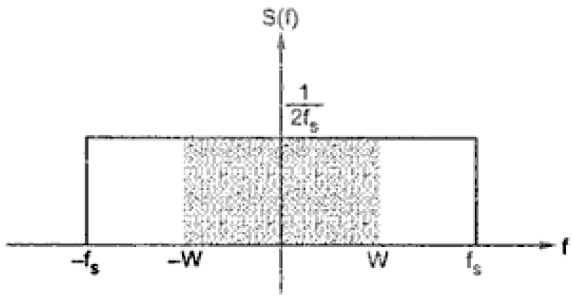


Fig. PSD of noise

This noise power is uniformly distributed over $-f_s$ to f_s range. This is illustrated in Fig. At the output of delta modulator receiver there is lowpass reconstruction filter whose cutoff frequency is 'W'. This cutoff frequency is equal to highest signal frequency. The reconstruction filter passes part of the noise power at the output as Fig. From the geometry of Fig. output noise power will be,

$$\text{Output noise power} = \frac{W}{f_s} \times \text{noise power} = \frac{W}{f_s} \times \frac{\delta^2}{3}$$

We know that $f_s = \frac{1}{T_s}$, hence above equation becomes,

$$\text{Output noise power} = \frac{W T_s \delta^2}{3} \qquad \text{.................(6)}$$

(iii) To obtain signal to noise power ratio

Signal to noise power ratio at the output of delta modulation receiver is given as,

$$\frac{S}{N} = \frac{\textit{Normalized signal power}}{\textit{Normalized noise power}}$$

From equation 2. and equation 6

$$\frac{S}{N} = \frac{\dfrac{\delta^2}{8\pi^2 f_m^2 T_s^2}}{\dfrac{W T_s \delta^2}{3}}$$

$$\boxed{\frac{S}{N} = \frac{3}{8\pi^2 W f_m^2 T_s^3}} \qquad \text{..............(7)}$$

This is an expression for signal to noise power ratio in delta modulation.

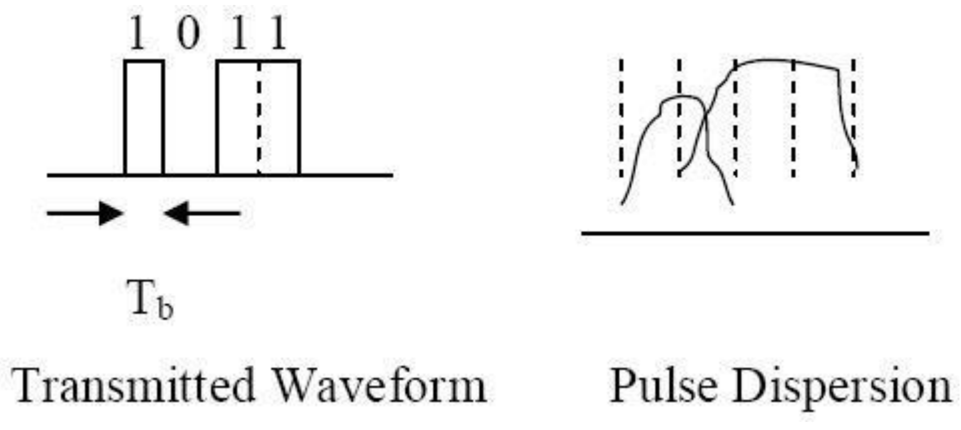
UNIT- III

BASEBAND PULSE TRANSMISSION

Inter symbol Interference:

Generally, digital data is represented by electrical pulse, communication channel is always band limited. Such a channel disperses or spreads a pulse carrying digitized samples passing through it. When the channel bandwidth is greater than bandwidth of pulse, spreading of pulse is very less. But when channel bandwidth is close to signal bandwidth, i.e. if we transmit digital data which demands more bandwidth which exceeds channel bandwidth, spreading will occur and cause signal pulses to overlap. This overlapping is called **InterSymbol Interference**. In short it is called ISI. Similar to interference caused by other sources, ISI causes degradations of signal if left uncontrolled. This problem of ISI exists strongly in Telephone channels like coaxial cables and optical fibers.

The main objective is to study the effect of ISI, when digital data is transmitted through band limited channel and solution to overcome the degradation of waveform by properly shaping pulse



(Source:Brainkart)

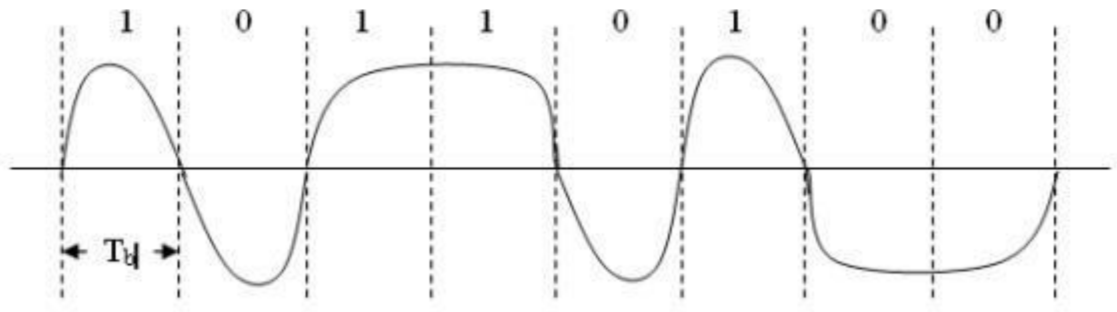
The effect of sequence of pulses transmitted through channel is shown in fig. The Spreading of pulse is greater than symbol duration, as a result adjacent pulses interfere. i.e. pulses get completely smeared, tail of smeared pulse enter into adjacent symbol intervals making it difficult to decide actual transmitted pulse. First let us have look at different formats of transmitting digital data.In base band transmission best way is to map digits or symbols into pulse waveform. This waveform is generally termed as **Line codes**.

EYE PATTERN

The quality of digital transmission systems are evaluated using the bit error rate. Degradation of quality occurs in each process modulation, transmission, and detection. The eye pattern is experimental method that contains all the information concerning the degradation of quality. Therefore, careful analysis of the eye pattern is important in analyzing the degradation mechanism.

- Eye patterns can be observed using an oscilloscope. The received wave is applied to the vertical deflection plates of an oscilloscope and the saw tooth wave at a rate equal to transmitted symbol rate is applied to the horizontal deflection plates, resulting display is eye pattern as it resembles human eye.

- The interior region of eye pattern is called eye opening



(Source:Brainkart)

We get superposition of successive symbol intervals to produce eye pattern as shown below.

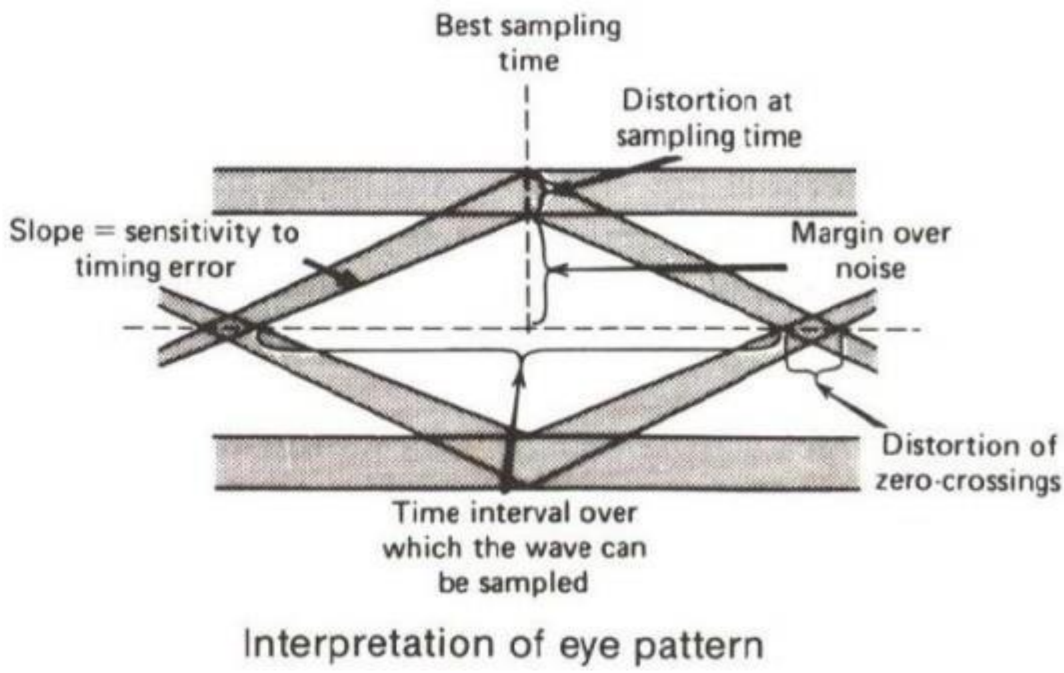


Fig 3.1Eye pattern (Source:Brainkart)

• The width of the eye opening defines the time interval over which the received wave can be sampled without error from ISI

• The optimum sampling time corresponds to the maximum eye opening

• The height of the eye opening at a specified sampling time is a measure of the margin over channel noise.

The sensitivity of the system to timing error is determined by the rate of closure of the eye as the sampling time is varied. Any non linear transmission distortion would reveal itself in an asymmetric or squinted eye. When the effected of ISI is excessive, traces from the upper portion of the eye pattern cross traces from lower portion with the result that the eye is completely closed.

Example of eye pattern:

Binary-PAM Perfect channel (no noise and no ISI)

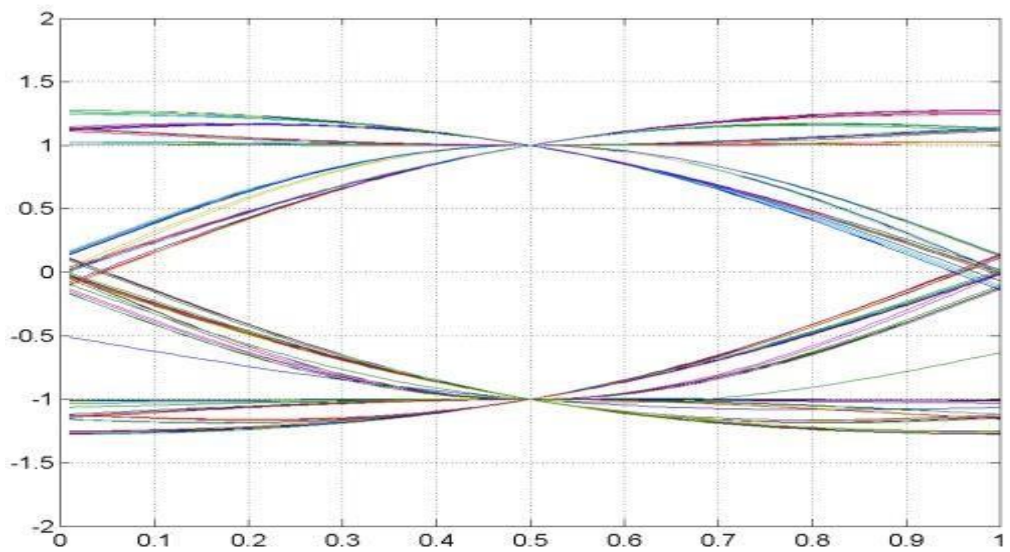


Fig 3.2 Example of eye pattern: Binary-PAM with noise no ISI (Source:Brainkart)

EQUALISING FILTER

Adaptive equalization

• An equalizer is a filter that compensates for the dispersion effects of a channel. Adaptive equalizer can adjust its coefficients continuously during the transmission of data.

Pre channel equalization

•requires feed back channel

•causes burden on transmission.

Post channel equalization

Achieved prior to data transmission by training the filter with the guidance of a training sequence transmitted through the channel so as to adjust the filter parameters to optimum values.

Adaptive equalization

It consists of tapped delay line filter with set of delay elements, set of adjustable multipliers connected to the delay line taps and a summer for adding multiplier outputs.

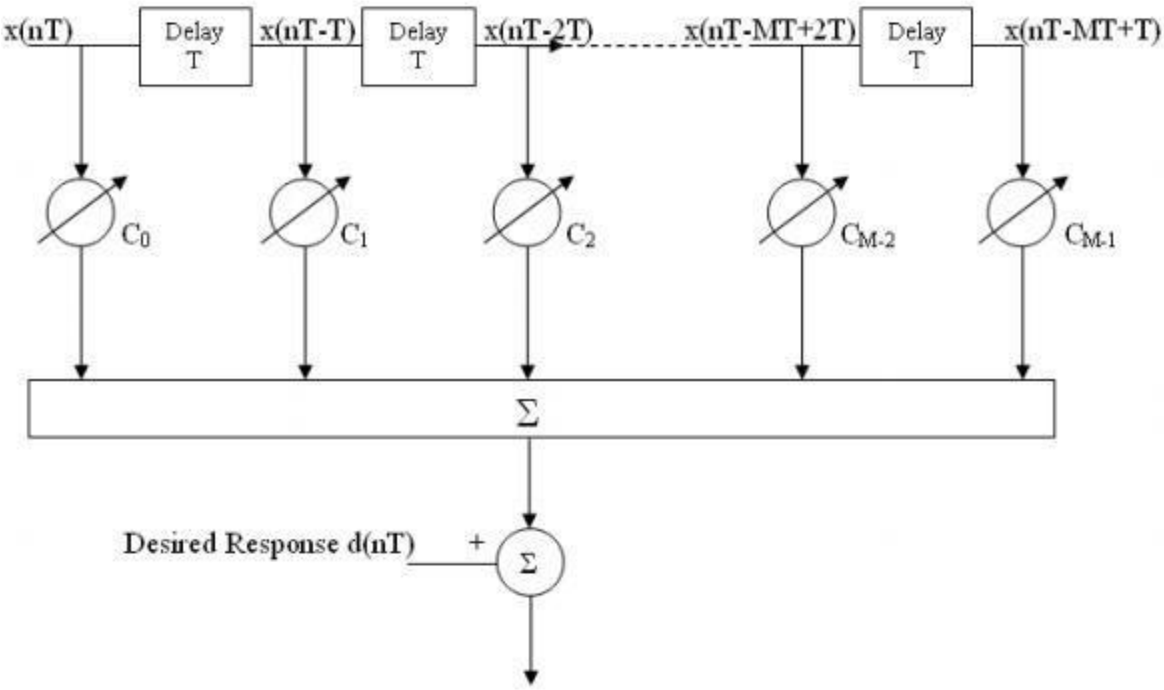


Fig 3.3 Adaptive equalization (Source:Brainkart)

The output of the Adaptive equalizer is given by

$$Y(nt)=\sum C_i\, x(nT-iT)$$

Ci is weight of the ith tap Total number of taps are M .Tap spacing is equal to symbol duration T of transmitted signal In a conventional FIR filter the tap weights are constant and particular designed response is obtained. In the adaptive equaliser the **Ci**'s are variable and are adjusted by an algorithm.

Two modes of operation

1. Training mode

2. Decision directed mode

Mechanism of adaptation

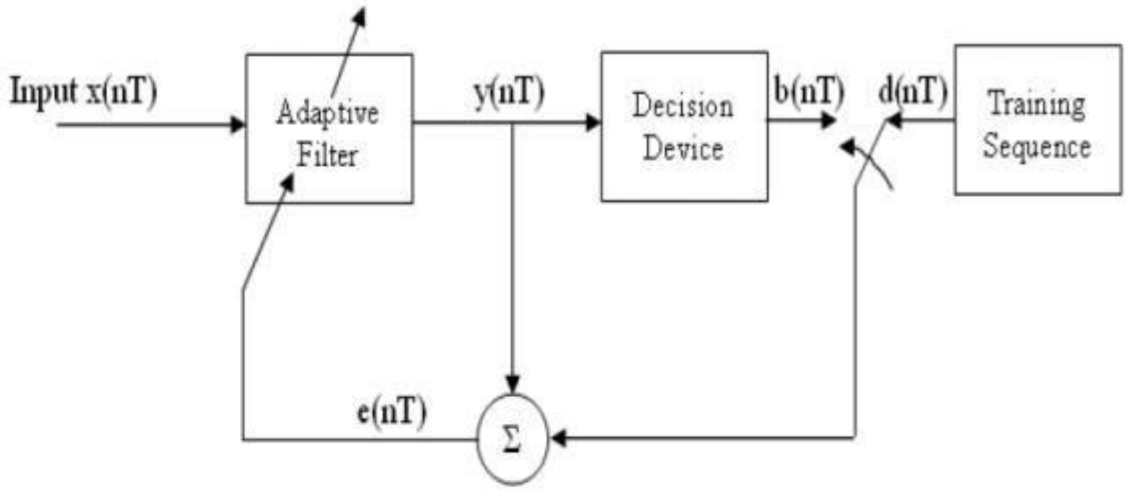


Fig 3.4 Mechanism of adaptation (Source:Brainkart)

Training mode

A known sequence d(nT) is transmitted and synchronized version of it is generated in the receiver applied to adaptive equalizer. This training sequence has maximal length PN Sequence, because it has large average power and large SNR, resulting response sequence (Impulse) is observed by measuring the filter outputs at the sampling instants. The difference between resulting response y(nT) and desired response d(nT)is error signal which is used to estimate the direction in which the coefficients of filter are to be optimized using algorithms.

Matched Filter

It is obtained by correlating a known delayed signal, or *template*, with an unknown signal to detect the presence of the template in the unknown signal. This is equivalent to convolving the unknown signal with a conjugated time-reversed version of the template. The matched filter is the optimal linear filter for maximizing the signal-to-noise ratio (SNR) in the presence of additive stochastic noise.

Matched filters are commonly used in radar, in which a known signal is sent out, and the reflected signal is examined for common elements of the out-going signal. Pulse compression is an example of matched filtering. It is so called because the impulse response is matched to input pulse signals. Two-dimensional matched filters are commonly used in image processing, e.g., to improve the SNR of X-ray observations. Matched filtering is a demodulation technique with LTI (linear time invariant) filters to maximize SNR. It was originally also known as a *North filter*.

Pulse Shaping

It is the process of changing the waveform of transmitted pulses. Its purpose is to make the transmitted signal better suited to its purpose or the communication channel, typically by limiting the effective bandwidth of the transmission. By filtering the transmitted pulses this way, the inter symbol interference caused by the channel can be kept in control. In RF communication, pulse shaping is essential for making the signal fit in its frequency band.

Typically pulse shaping occurs after line coding and modulation.

Need for pulse shaping

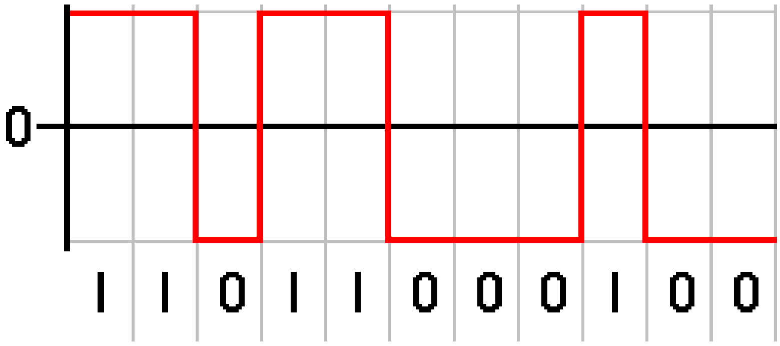
Transmitting a signal at high modulation rate through a band-limited channel can create inter symbol interference. As the modulation rate increases, the signal's bandwidth increases. When the signal's bandwidth becomes larger than the channel bandwidth, the channel starts to introduce distortion to the signal. This distortion usually manifests itself as inter symbol interference.

The signal's spectrum is determined by the modulation scheme and data rate used by the transmitter, but can be modified with a pulse shaping filter. Usually the transmitted symbols are represented as a time sequence of dirac delta pulses. This theoretical signal is then filtered with the pulse shaping filter, producing the transmitted signal.

In many base band communication systems the pulse shaping filter is implicitly a boxcar filter. Its Fourier transform is of the form *sin(x)/x*, and has significant signal power at frequencies higher than symbol rate. This is not a big problem when optical fibre or even twisted pair cable is used as the communication channel. However, in RF communications this would waste bandwidth, and only tightly specified frequency bands are used for single transmissions. In other words, the channel for the signal is band-limited. Therefore better filters have been developed, which attempt to minimize the bandwidth needed for a certain symbol rate.

An example in other areas of electronics is the generation of pulses where the rise time need to be short; one way to do this is to start with a slower-rising pulse, and decrease the rise time, for example with a step recovery diode circuit.

Pulse shaping filters



(Source:Brainkart)

A typical NRZ coded signal is implicitly filtered with a sinc filter.

Not every filter can be used as a pulse shaping filter. The filter itself must not introduce inter symbol interference — it needs to satisfy certain criteria. The Nyquist ISI criterion is a commonly used criterion for evaluation, because it relates the frequency spectrum of the transmitter signal to intersymbol interference.

Examples of pulse shaping filters that are commonly found in communication systems are:

- Sinc shaped filter
- Raised-cosine filter
- Gaussian filter

Sender side pulse shaping is often combined with a receiver side matched filter to achieve optimum tolerance for noise in the system. In this case the pulse shaping is equally distributed between the sender and receiver filters. The filters' amplitude responses are thus point wise square roots of the system filters.

Other approaches that eliminate complex pulse shaping filters have been invented. In OFDM, the carriers are modulated so slowly that each carrier is virtually unaffected by the bandwidth limitation of the channel.

Sinc filter

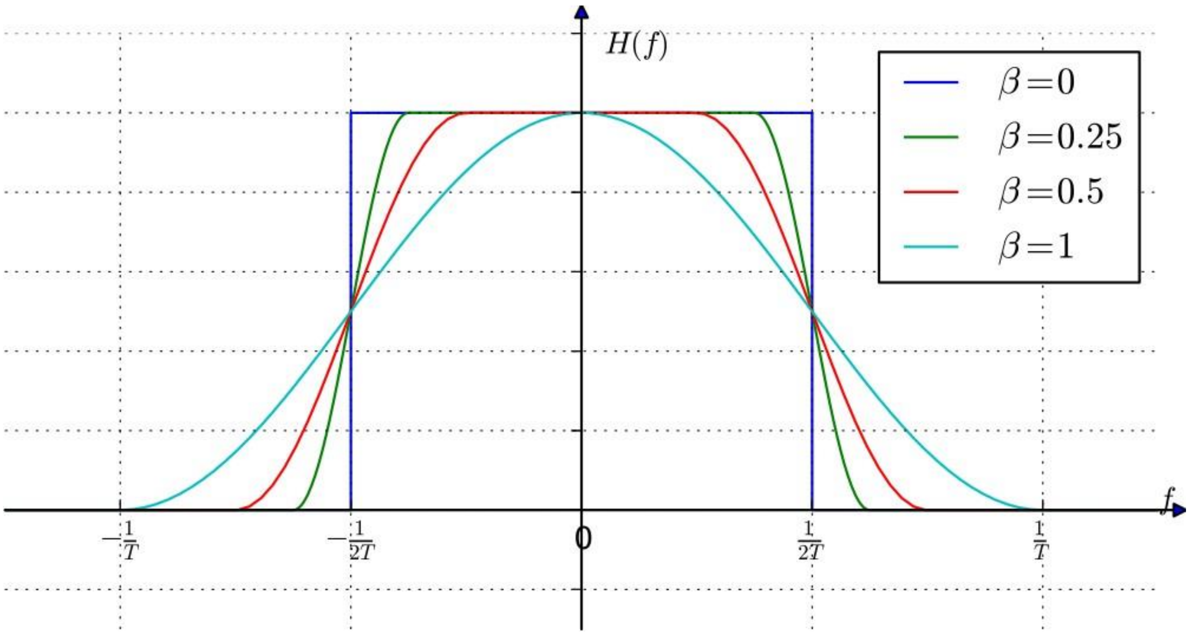


Fig 3.5 Amplitude response of raised-cosine filter with various roll-off factors (Source:Brainkart)

It is also called as Boxcar filter as its frequency domain equivalent is a rectangular shape. Theoretically the best pulse shaping filter would be the sinc filter, but it cannot be implemented precisely. It is a non-causal filter with relatively slowly decaying tails. It is also problematic from a synchronization point of view as any phase error results in steeply increasing inter symbol interference.

Raised-cosine filter

Raised-cosine is similar to sinc, with the tradeoff of smaller side lobes for a slightly larger spectral width. Raised-cosine filters are practical to implement and they are in wide use. They have a configurable excess bandwidth, so communication systems can choose a trade off between a simpler filter and spectral efficiency.

Gaussian filter

This gives an output pulse shaped like a Gaussian function.

Nyquist criterion

When the baseband filters in the communication system satisfy the Nyquist criterion, symbols can be transmitted over a channel with flat response within a limited frequency band, without ISI. Examples of such baseband filters are the raised-cosine filter, or the sinc filter as the ideal case.

UNIT – IV
DIGITAL PASSBAND TRANSMISSION

5.1.5 Passband Transmission Model

Fig. 5.1.2 shows the model of passband data transmission system

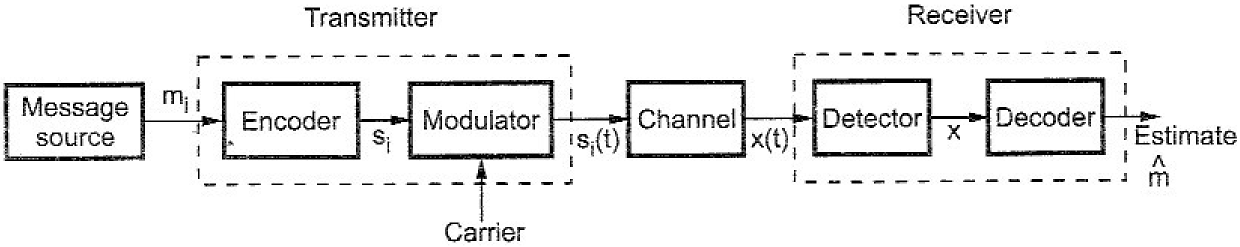


Fig. 5.1.2 Model of passband data transmission system

1. **Message source** : It emits the symbol at the rate of T seconds.
2. **Encoder** : It is signal transmission encoder. It produces the vector s_i made up of 'N' real elements. The vector s_i is unique for each set of 'M' symbols.
3. **Modulator** : It constructs the modulated carrier signal $s_i(t)$ of duration 'T' seconds for every symbol m_i. The signal $s_i(t)$ is energy signal.
4. **Channel** : The modulated signal $s_i(t)$ is transmitted over the communication channel.
   - The channel is assumed to be linear and of enough bandwidth to accommodate the signal $s_i(t)$.
   - The channel noise is white Gaussian of zero mean and psd of $\frac{N_0}{2}$.
5. **Detector** : It demodulates the received signal and obtains an estimate of the signal vector.
6. **Decoder** : The decoder obtains the estimate of symbol back from the signal vector. Here note that the detector and decoder combinely perform the reception of the transmitted signal. The effect of channel noise is minimized and correct estimate of symbol $\hat{m}$ is obtained.

5.2 Binary Phase Shift Keying (BPSK)

5.2.1 Principle of BPSK

- In binary phase shift keying (BPSK), binary symbol '1' and '0' modulate the phase of the carrier. Let the carrier be,

$$s(t) = A\cos(2\pi f_0 t) \qquad \text{... (5.2.1)}$$

'A' represents peak value of sinusoidal carrier. In the standard 1Ω load register, the power dissipated will be,

$$P = \frac{1}{2}A^2$$

$$\therefore \qquad A = \sqrt{2P} \qquad \text{... (5.2.2)}$$

- When the symbol is changed, then the phase of the carrier is changed by 180 degrees (π radians).
- Consider for example,

$$\text{Symbol '1'} \Rightarrow s_1(t) = \sqrt{2P}\cos(2\pi f_0 t) \qquad \text{... (5.2.3)}$$

if next symbol is '0' then,

$$\text{Symbol '0'} \Rightarrow s_2(t) = \sqrt{2P}\cos(2\pi f_0 t + \pi) \qquad \text{... (5.2.4)}$$

Since $\cos(\theta+\pi) = -\cos\theta$, we can write above equation as,

$$s_2(t) = -\sqrt{2P}\cos(2\pi f_0 t) \qquad \text{... (5.2.5)}$$

With the above equation we can define BPSK signal combinely as,

$$\boxed{s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t)} \qquad \text{... (5.2.6)}$$

Here $b(t)$ = $+1$ when binary '1' is to be transmitted

= -1 when binary '0' is to be transmitted

5.2.2 Graphical Representation of BPSK Signal

Fig. 5.2.1 shows binary signal and its equivalent signal $b(t)$.

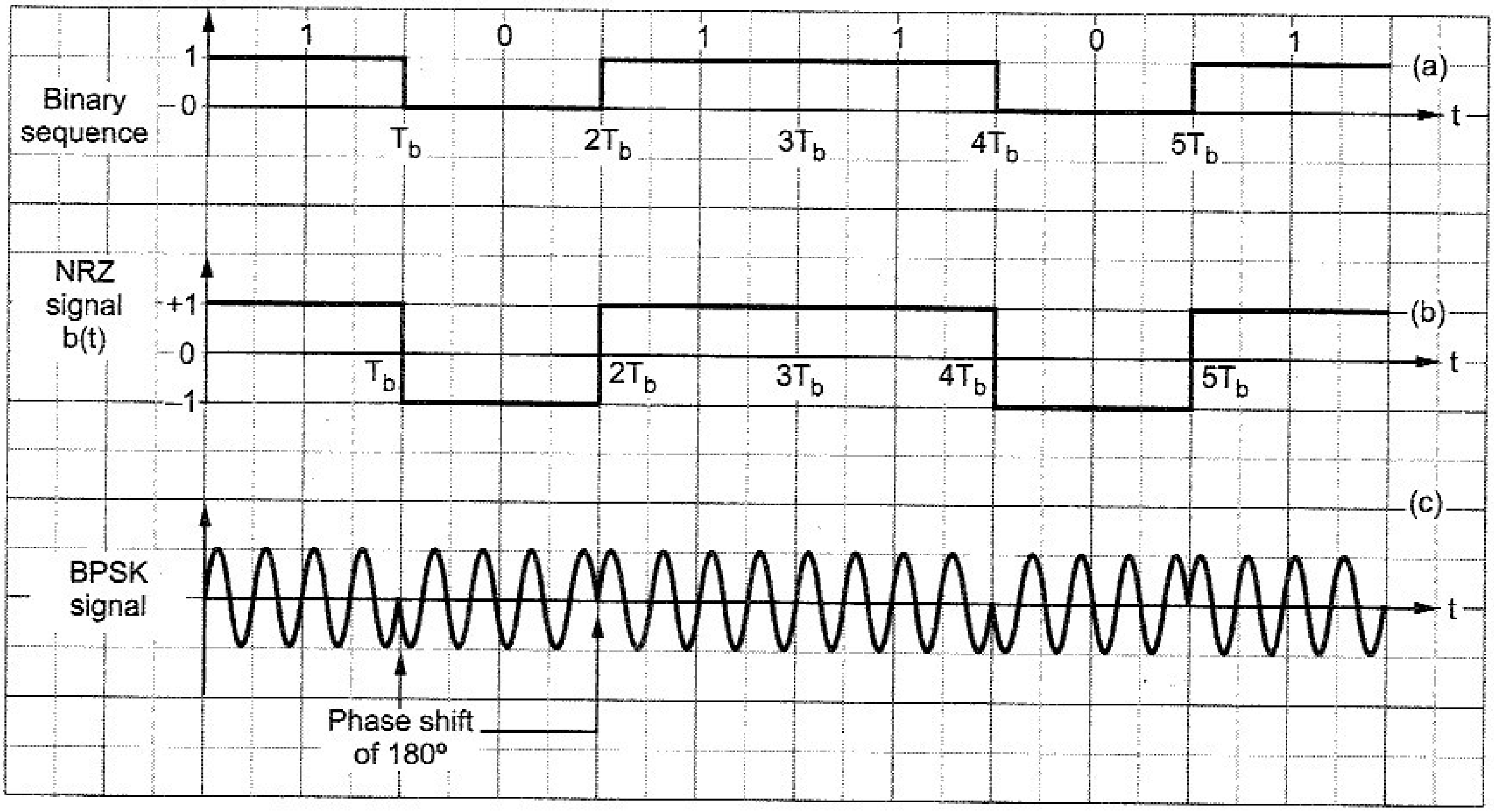


Fig. 5.2.1 (a) Binary sequence
(b) Its equivalent bipolar signal $b(t)$
(c) BPSK signal

As can be seen from Fig. 5.2.1 (b), the signal $b(t)$ is NRZ bipolar signal. This signal directly modulates carrier $\cos(2\pi f_0 t)$.

5.2.3 Generation and Reception of BPSK Signal

Nov./Dec.- 2005

5.2.3.1 Generator of BPSK Signal

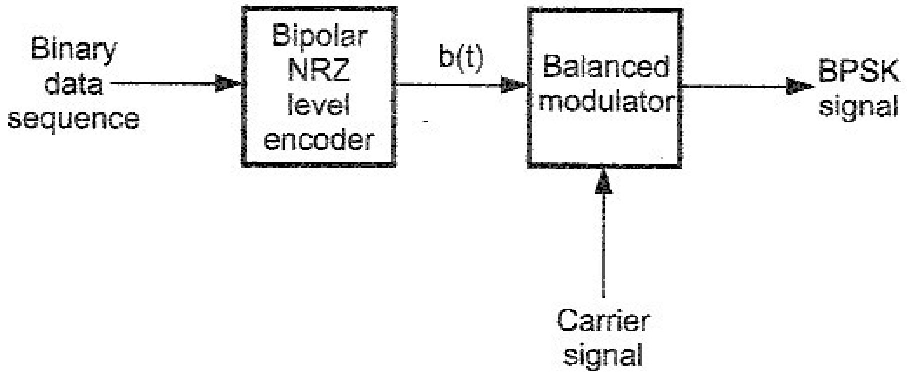


Fig. 5.2.2 BPSK generation scheme

- The BPSK signal can be generated by applying carrier signal to the balanced modulator.
- The baseband signal $b(t)$ is applied as a modulating signal to the balanced modulator. Fig. 5.2.2 shows the block diagram of BPSK signal generator.
- The NRZ level encoder converts the binary data sequence into bipolar NRZ signal.

5.2.3.2 Reception of BPSK Signal

Fig. 5.2.3 shows the block diagram of the scheme to recover baseband signal from BPSK signal. The transmitted BPSK signal is,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t)$$

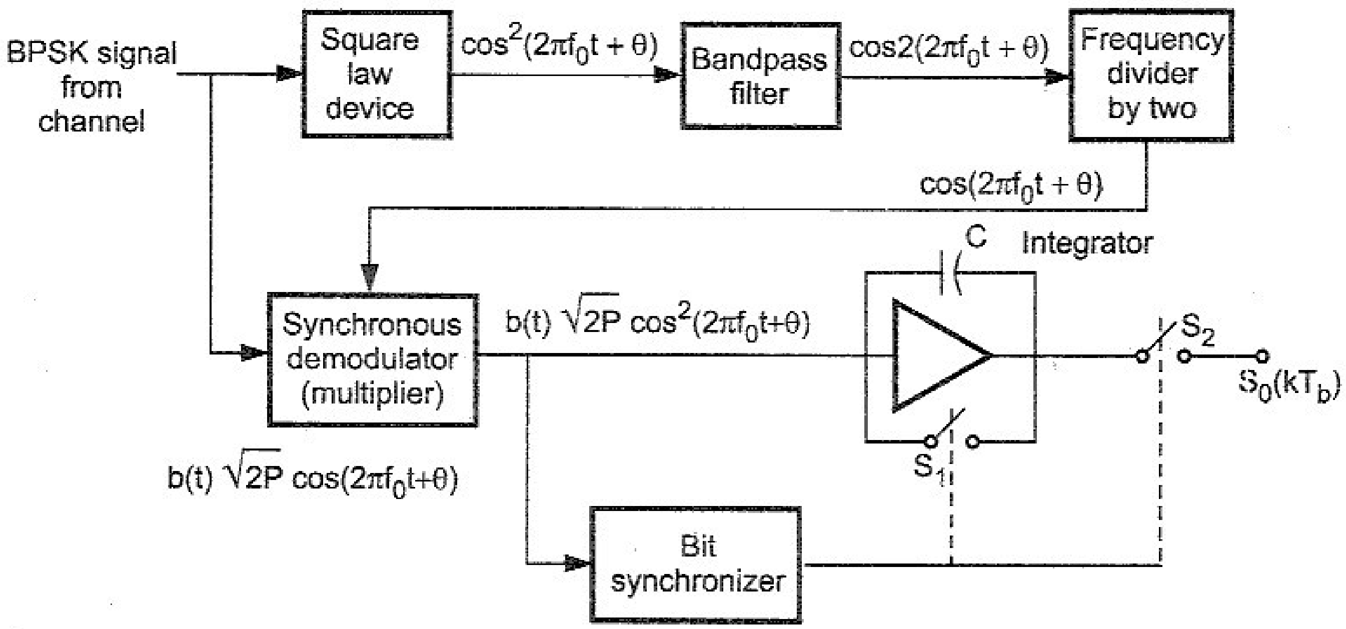


Fig. 5.2.3 Reception BPSK scheme

Operation of the receiver

1) **Phase shift in received signal** : This signal undergoes the phase change depending upon the time delay from transmitter to receiver. This phase change is normally fixed phase shift in the transmitted signal. Let the phase shift be θ. Therefore the signal at the input of the receiver is,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t + \theta) \qquad \text{... (5.2.7)}$$

2) **Square law device** : Now from this received signal, a carrier is separated since this is coherent detection. As shown in the figure, the received signal is passed through a square law device. At the output of the square law device the signal will be,

$$\cos^2(2\pi f_0 t + \theta)$$

Note here that we have neglected the amplitude, because we are only interested in the carrier of the signal.

We know that,

$$\cos^2\theta = \frac{1+\cos 2\theta}{2}$$

$$\therefore \quad \cos^2(2\pi f_0 t + \theta) = \frac{1+\cos 2(2\pi f_0 t + \theta)}{2}$$

or $$= \frac{1}{2} + \frac{1}{2}\cos 2(2\pi f_0 t + \theta)$$ Here $\frac{1}{2}$ represents a DC level.

3) **Bandpass filter** : This signal is then passed through a bandpass filter whose passband is centered around $2f_0$. Bandpass filter removes the DC level of $\frac{1}{2}$ and at its output we get,

$$\cos 2(2\pi f_0 t + \theta)$$ This signal has frequency of $2f_0$.

4) **Frequency divider** : The above signal is passed through a frequency divider by two. Therefore at the output of frequency divider we get a carrier signal whose frequency is f_0 i.e. $\cos(2\pi f_0 t + \theta)$.

5) **Synchronous demodulator** : The synchronous (coherent) demodulator multiplies the input signal and the recovered carrier. Therefore at the output of multiplier we get,

$$b(t)\sqrt{2P}\cos(2\pi f_0 t + \theta) \times \cos(2\pi f_0 t + \theta) = b(t)\sqrt{2P}\cos^2(2\pi f_0 t + \theta)$$

$$= b(t)\sqrt{2P} \times \frac{1}{2}[1 + \cos 2(2\pi f_0 t + \theta)]$$

$$= b(t)\sqrt{\frac{P}{2}}[1 + \cos 2(2\pi f_0 t + \theta)] \qquad \text{... (5.2.8)}$$

6) **Bit synchronizer and integrator :** The above signal is then applied to the bit synchronizer and integrator. The integrator integrates the signal over one bit period. The bit synchronizer takes care of starting and ending times of a bit.

- At the end of bit duration T_b, the bit synchronizer closes switch S_2 temperorily. This connects the output of an integrator to the decision device. It is equivalent to sampling the output of integrator.
- The synchronizer then opens switch S_2 and switch S_1 is closed temperorily. This resets the integrator voltage to zero. The integrator then integrates next bit.
- Let us assume that one bit period 'T_b' contains integral number of cycles of the carrier. That is the phase change occurs in the carrier only at zero crossing. This is shown in Fig. 5.2.1 (c). Thus BPSK waveform has full cycles of sinusoidal carrier.

To show that output of integrator depends upon transmitted bit

- In the k^{th} bit interval we can write output signal as,

$$s_o(kT_b) = b(kT_b)\sqrt{\frac{P}{2}} \int_{(k-1)T_b}^{kT_b} [1+\cos 2(2\pi f_0 t+\theta)]\, dt$$

from equation 5.2.8

The above equation gives the output of an interval for k^{th} bit. Therefore integration is performed from $(k-1)T_b$ to kT_b. Here T_b is the one bit period.

- We can write the above equation as,

$$s_o(kT_b) = b(kT_b)\sqrt{\frac{P}{2}} \left[\int_{(k-1)T_b}^{kT_b} 1\, dt + \int_{(k-1)T_b}^{kT_b} \cos 2(2\pi f_0 t+\theta)\, dt \right]$$

Here $\int_{(k-1)T_b}^{kT_b} \cos 2(2\pi f_0 t+\theta)\, dt = 0$, because average value of sinusoidal waveform is zero if integration is performed over full cycles. Therefore we can write above equation as,

$$s_o(kT_b) = b(kT_b)\sqrt{\frac{P}{2}} \int_{(k-1)T_b}^{kT_b} 1\, dt$$

$$= b(kT_b)\sqrt{\frac{P}{2}}\,[t]_{(k-1)T_b}^{kT_b}$$

$$= b(kT_b)\sqrt{\frac{P}{2}}\,\{kT_b-(k-1)T_b\}$$

$$\therefore \quad \boxed{s_o(kT_b) = b(kT_b)\sqrt{\frac{P}{2}}\,T_b} \quad \text{... (5.2.9)}$$

This equation shows that the output of the receiver depends on input i.e.

$$s_o(kT_b) \ \alpha \ b(kT_b)$$

Depending upon the value of $b(kT_b)$, the output $s_0(kT_b)$ is generated in the receiver.

This signal is then given to a decision device (not shown in Fig. 5.2.3), which decides whether transmitted symbol was zero or one.

5.2.4 Spectral Characteristics of BPSK Signals

Step 1 : Fourier transform of basic NRZ pulse.

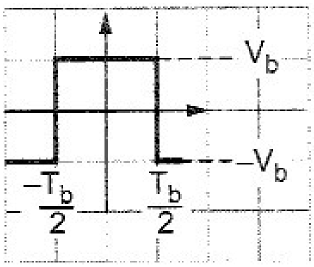


Fig. 5.2.4 NRZ pulse

We know that the waveform $b(t)$ is NRZ bipolar waveform. In this waveform there are rectangular pulses of amplitude $\pm V_b$. If we say that each pulse is $\pm\frac{T_b}{2}$ around its center as shown in Fig. 5.2.4. then it becomes easy to find fourier transform of such pulse. The fourier transform of this type of pulse is given as,

$$X(f) = V_b\,T_b\,\frac{\sin(\pi f T_b)}{(\pi f T_b)} \quad \text{By standard relations} \quad \text{... (5.2.10)}$$

Step 2 : PSD of NRZ pulse.

For large number of such positive and negative pulses the power spectral density $S(f)$ is given as

$$S(f) = \frac{\overline{|X(f)|^2}}{T_s} \quad \text{... (5.2.11)}$$

Here $\overline{X(f)}$ denotes average value of $X(f)$ due to all the pulses in $b(t)$. And T_s is symbol duration. Putting value of $X(f)$ from equation 5.2.10 in equation 5.2.11 we get,

$$S(f) = \frac{V_b^2\, T_b^2}{T_s}\left[\frac{\sin(\pi f T_b)}{\pi f T_b}\right]^2$$

Step 3 : PSD of baseband signal *b(t)*

For BPSK since only one bit is transmitted at a time, symbol and bit durations are same i.e. $T_b = T_s$. Then above equation becomes,

$$S(f) = V_b^2\, T_b\left[\frac{\sin(\pi f T_b)}{\pi f T_b}\right]^2 \qquad \text{... (5.2.12)}$$

The above equation gives the power spectral density of baseband signal $b(t)$.

Step 4 : PSD of BPSK signal.

The BPSK signal is generated by modulating a carrier by the baseband signal $b(t)$. Because of modulation of the carrier of frequency f_0, the spectral components are translated from f to $f_0 + f$ and $f_0 - f$. The magnitude of those components is divided by half.

Therefore from equation 5.2.12 we can write the power spectral density of BPSK signal as,

$$S_{BPSK}(f) = V_b^2\, T_b\left\{\frac{1}{2}\left[\frac{\sin\pi(f_0 - f)\,T_b}{\pi(f_0 - f)\,T_b}\right]^2 + \frac{1}{2}\left[\frac{\sin\pi(f_0 + f)\,T_b}{\pi(f_0 + f)\,T_b}\right]^2\right\}$$

The above equation is composed of two half magnitude spectral components of same frequency 'f' above and below f_0. Let us say that the value of $\pm V_b = \pm\sqrt{P}$. That is the NRZ signal is having amplitudes of $+\sqrt{P}$ and $-\sqrt{P}$. Then above equation becomes,

$$S_{BPSK}(f) = \frac{P\,T_b}{2}\left\{\left[\frac{\sin\pi(f - f_0)\,T_b}{\pi(f - f_0)\,T_b}\right]^2 + \frac{1}{2}\left[\frac{\sin\pi(f_0 + f)\,T_b}{\pi(f_0 + f)\,T_b}\right]^2\right\} \qquad \text{... (5.2.13)}$$

The above equation gives power spectral density of BPSK signal for modulating signal $b(t)$ having amplitudes of $\pm\sqrt{P}$. We know that modulated signal is given by equation 5.2.3 and equation 5.2.5 as,

$$s(t) = \pm\sqrt{2P}\cos(2\pi f_0 t) \qquad \text{since } A = \sqrt{2P}$$

If $b(t) = \pm\sqrt{P}$, then the carrier becomes,

$$\phi(t) = \sqrt{2}\cos(2\pi f_0 t) \qquad \text{... (5.2.14)}$$

Plot of PSD

- Equation 5.2.12 gives power spectral density of the NRZ waveform. For one rectangular pulse, the shape of $S(f)$ will be a sinc pulse as given by equation 5.2.12. Fig. 5.2.5 shows the plot of magnitude of $S(f)$.

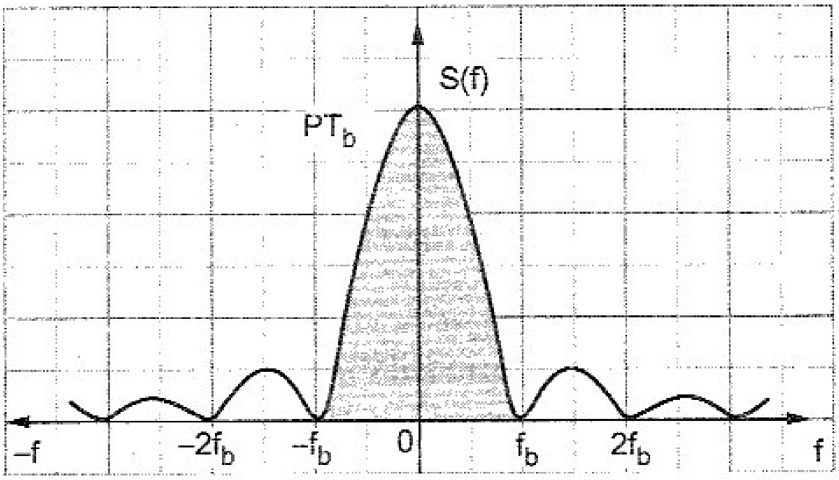


Fig. 5.2.5 Plot of power spectral density of NRZ baseband signal

Above figure shows that the main lobe ranges from $-f_b$ to $+f_b$. Here $f_b = \frac{1}{T_b}$.

Since we have taken $\pm V_b = \pm\sqrt{P}$ in equation 5.2.12, the peak value of the main lobe is PT_b.

- Now let us consider the power spectral density of BPSK signal given by equation 5.2.15. Fig. 5.2.6 shows the plot of this equation. The figure thus clearly shows that there are two lobes ; one at f_0 and other at $-f_0$. The same spectrum of Fig. 5.2.5 is placed at $+f_0$ and $-f_0$. But the amplitudes of main lobes are $\frac{PT_b}{2}$ in Fig. 5.2.6.

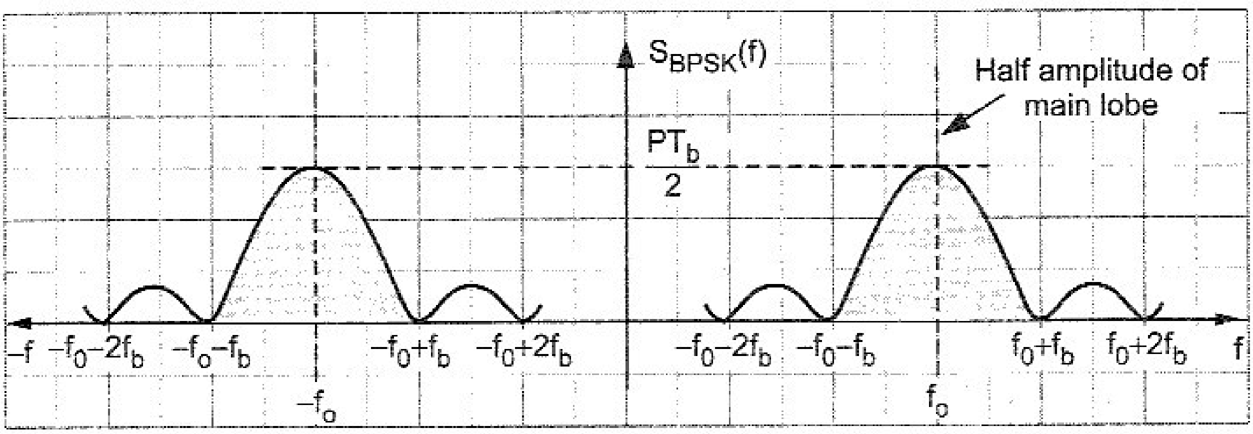


Fig. 5.2.6 Plot of power spectral density of BPSK signal

Thus they are reduced to half. The spectrums of $S(f)$ as well as $S_{BPSK}(f)$ extends over all the frequencies.

Interchannel Interference and ISI :

- Let's assume that BPSK signals are multiplexed with the help of different carrier frequencies for different baseband signals. Then at any frequency, the spectral components due to all the multiplexed channels will be present. This is because $S(f)$ as well as $S_{BPSK}(f)$ of every channel extends over all the frequency range.
- Therefore a BPSK receiver tuned to a particular carrier frequency will also receive frequency components due to other channels. This will make interference with the required channel signals and error probability will increase. This result is called *Interchannel Interference.*
- To avoid interchannel interference, the BPSK signal is passed through a filter.This filter attenuates the side lobes and passes only main lobe. Since side lobes are attenuated to high level, the interference is reduced. Because of this filtering the phase distortion takes place in the bipolar NRZ signal, i.e. $b(t)$. Therefore the individual bits (symbols) mix with adjacent bits (symbols) in the same channel. This effect is called *intersymbol interference* or ISI.
- The effect of ISI can be reduced to some extent by using equalizers at the receiver. Those equalizers have the reverse effect to that filter's adverse effects. Normally equalizers are also filter structures.

5.2.5 Geometrical Representation of BPSK Signals

We know that BPSK signal carries the information about two symbols. Those are symbol '1' and symbol '0'. We can represent BPSK signal geometrically to show those two symbols.

(i) From equation 5.2.6 we know that BPSK signal is given as,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t) \qquad \text{... (5.2.15)}$$

(ii) Let's rearrange the above equation as,

$$s(t) = b(t)\sqrt{PT_b}\cdot\sqrt{\frac{2}{T_b}}\cos(2\pi f_0 t) \qquad \text{... (5.2.16)}$$

(iii) Let $\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_0 t)$ represents an orthonormal carrier signal. Equation 5.2.14 also gives equation for carrier. It is slightly different than $\phi_1(t)$ defined here. Then we can write equation 5.2.16 as,

$$s(t) = b(t)\sqrt{PT_b}\,\phi_1(t) \qquad \text{... (5.2.17)}$$

(iv) The bit energy E_b is defined in terms of power 'P' and bit duration T_b as,

$$E_b = P T_b \quad \text{... (5.2.18)}$$

∴ Equation 5.2.17 becomes,

$$s(t) = \pm \sqrt{E_b}\, \phi_1(t) \quad \text{... (5.2.19)}$$

Here $b(t)$ is simply ± 1.

(v) Thus on the single axis of $\phi_1(t)$ there will be two points. One point will be located at $+\sqrt{E_b}$ and other point will be located at $-\sqrt{E_b}$. This is shown in Fig: 5.2.7.

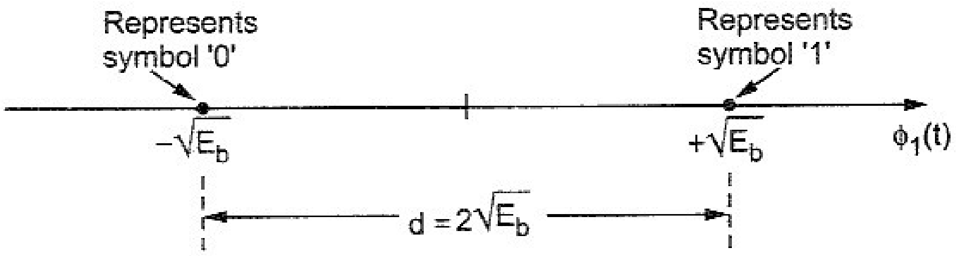


Fig. 5.2.7 Geometrical representation of BPSK signal

At the receiver the point at $+\sqrt{E_b}$ on $\phi_1(t)$ represents symbol '1' and point at $-\sqrt{E_b}$ represents symbol '0'. The separation between these two points represent the isolation in symbols '1' and '0' in BPSK signal. This separation is normally called distance 'd'. From Fig.5.2.7 it is clear that the distance between the two points is,

$$d = +\sqrt{E_b} - (-\sqrt{E_b})$$

$$\therefore \quad d = 2\sqrt{E_b} \quad \text{... (5.2.20)}$$

As this distance 'd' increases, the isolation between the symbols in BPSK signal is more. Therefore probability of error reduces.

5.2.6 Bandwidth of BPSK Signal

The spectrum of the BPSK signal is centered around the carrier frequency f_0.

If $f_b = \frac{1}{T_b}$, then for BPSK the maximum frequency in the baseband signal will be f_b see Fig. 5.2.6. In this figure the main lobe is centered around carrier frequency f_0 and extends from $f_0 - f_b$ to $f_0 + f_b$. Therefore bandwidth of BPSK signal is,

$$BW = \text{Highest frequency} - \text{Lowest frequency in the main lobe}$$

$$= f_0 + f_b - (f_0 - f_b)$$

$$\therefore \quad \boxed{BW = 2f_b} \quad \text{... (5.2.21)}$$

Thus the minimum bandwidth of BPSK signal is equal to twice of the highest frequency contained in baseband signal.

5.2.7 Drawbacks of BPSK : Ambiguity in Output Signal

Fig. 5.2.3 shows the block diagram of BPSK receiver. To regenerate the carrier in the receiver, we start by squaring $b(t)\sqrt{2P}\cos(2\pi f_0 t+\theta)$. If the received signal is $-b(t)\sqrt{2P}\cos(2\pi f_0 t+\theta)$ then the squared signal remains same as before. Therefore the recovered carrier is unchanged even if the input signal has changed its sign. Therefore it is not possible to determine whether the received signal is equal to $b(t)$ or $-b(t)$. This result in ambiguity in the output signal.

This problem can be removed if we use differential phase shift keying. But Differential Phase Shift Keying (DPSK) also has some other problems. DPSK is given in detail in the next section. Other problems of BPSK are ISI and Interchannel interference. These problems are reduced to some extent by use of filters.

➡ **Example 5.2.1 :** *Determine the minimum bandwidth for a BPSK modulator with a carrier frequency of 40 MHz and an input bit rate of 500 kbps.*

Solution : The input bit rate indicates highest frequency of the baseband signal.

Hence,

$$f_b = 500 \text{ kbps} = 500 \text{ kHz.}$$

From equation 5.2.21, the bandwidth of the BPSK system is given as,

$$BW = 2f_b = 2 \times 500 \text{ kHz} = 1 \text{ MHz}$$

Review Questions

1. *Explain BPSK system with the help of transmitter and receiver, and state its advantages/disadvantages over other system.*
2. *Derive an expression for spectrum of BPSK system and hence calculate the bandwidth required.*

In the Fig. 5.3.1 observe that the transmitted signal is given as,

$$s(t) = b(t)\sqrt{2P}\cos(2\pi f_0 t)$$
$$= \pm\sqrt{2P}\cos(2\pi f_0 t)$$

i.e.
$$s(t) = \begin{cases} \sqrt{2P}\cos(2\pi f_0 t) & \text{when} \quad b(t)=1 \\ -\sqrt{2P}\cos(2\pi f_0 t) & \text{when} \quad b(t)=0 \end{cases}$$

The above equations can also be written as,

$$s(t) = \begin{cases} \sqrt{2P}\cos(2\pi f_0 t + 0) & \text{when} \quad b(t)=1 \\ -\sqrt{2P}\cos(2\pi f_0 t + \pi) & \text{when} \quad b(t)=0 \end{cases}$$

The transmitted phase sequence is shown in Fig. 5.3.4 as per the above equation.

Review Questions

1. *With the help of block diagram, waveforms and expressions explain the operation of DPSK transmitter and receiver.*
2. *What are the advantages and disadvantages of DPSK ? What is the bandwidth requirement of DPSK ?*

5.4 Quadrature Phase Shift Keying

May/June-2006

Principle

- In communication systems we know that there are two main resources, i.e. transmission power and the channel bandwidth. The channel bandwidth depends upon the bit rate or signalling rate f_b. In digital bandpass transmission, a carrier is used for transmission. This carrier is transmitted over a channel.
- If two or more bits are combined in some symbols, then the signalling rate is reduced. Therefore the frequency of the carrier required is also reduced. This reduces the transmission channel bandwidth. Thus because of grouping of bits in symbols, the transmission channel bandwidth is reduced.
- In quadrature phase shift keying, two successive bits in the data sequence are grouped together. This reduces the bits rate of signalling rate (i.e. f_b) and hence reduces the bandwidth of the channel.
- In BPSK we know that when symbol changes the level, the phase of the carrier is changed by 180°. Since there were only two symbols in BPSK, the phase shift occurs in two levels only.
- In QPSK two successive bits are combined. This combination of two bits forms four distinct symbols. When the symbol is changed to next symbol the

phase of the carrier is changed by 45° ($\pi / 4$ radians). Table 5.4.1 shows these symbols and their phase shifts.

Sr.No.	Input successive bits		Symbol	Phase shift in carrier
$i = 1$	1(1V)	0 (−1V)	S_1	$\pi / 4$
$i = 2$	0 (−1V)	0 (−1V)	S_2	$3\pi / 4$
$i = 3$	0 (−1V)	1(1V)	S_3	$5\pi / 4$
$i = 4$	1(1V)	1(1V)	S_4	$7\pi / 4$

Table 5.4.1 Symbol and corresponding phase shifts in QPSK

Thus as shown in above table, there are 4 symbols and the phase is shifted by $\pi / 4$ for each symbol.

5.4.1 QPSK Transmitter and Receiver

5.4.1.1 Offset QPSK (OQPSK) or Staggered QPSK Transmitter

Operation and waveforms April/May-2004

Step 1 : Input Sequence Converted to NRZ type :

Fig. 5.4.1 shows the block diagram of OQPSK transmitter. The input binary sequence is first converted to a bipolar NRZ type of signal. This signal is called $b(t)$. It represents binary '1' by $+1V$ and binary '0' by $-1V$. This signal is shown in Fig.5.4.2(a).

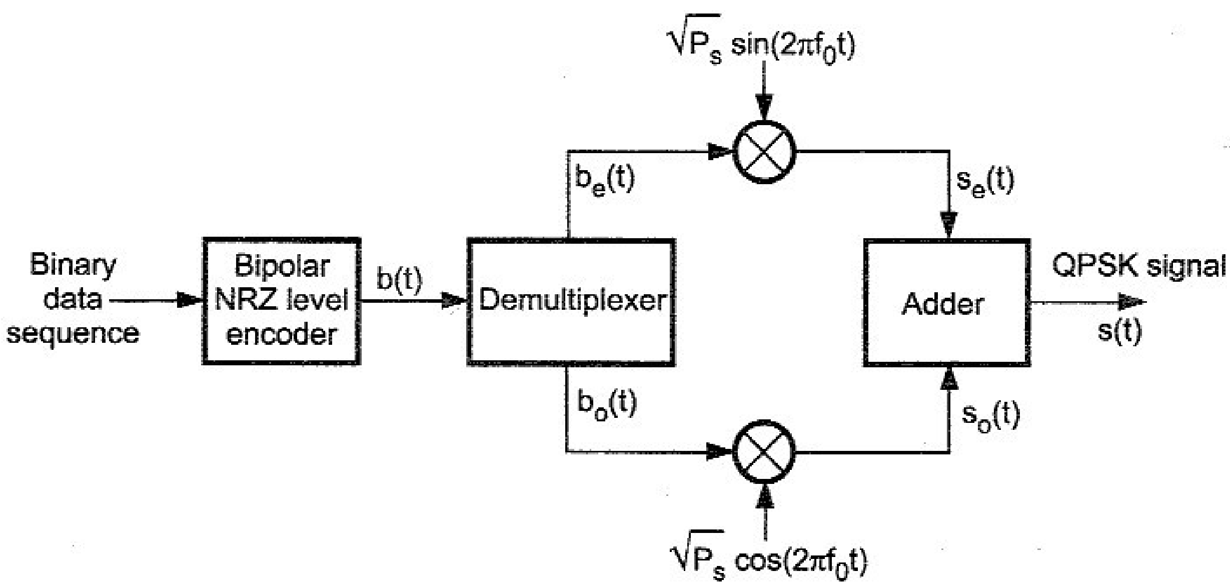


Fig. 5.4.1 An offset QPSK transmitter

Step 2 : Demultiplexing into odd and even numbered sequences

The demultiplexer divides $b(t)$ into two separate bit streams of the odd numbered and even numbered bits. $b_e(t)$ represents even numbered sequence and $b_o(t)$ represents odd numbered sequence. The symbol duration of both of these odd and even numbered sequences is $2T_b$. Thus every symbol contains two bits. Fig. 5.4.2 (b) and (c) shows the waveforms of $b_e(t)$ and $b_o(t)$.

Observe that the first even bit occurs after the first odd bit. Therefore even numbered bit sequence $b_e(t)$ starts with the delay of one bit period due to first odd bit. Thus first symbol of $b_e(t)$ is delayed by one bit period 'T_b' with respect to first symbol of $b_o(t)$. This delay of T_b is called offset. Hence the name offset QPSK is given. This shows that the change in levels of $b_e(t)$ and $b_o(t)$ cannot occur at the same time because of offset or staggering.

Step 3 : Modulation of quadrature carriers

The bit stream $b_e(t)$ modulates carrier $\sqrt{P_s}\cos(2\pi f_0 t)$ and $b_o(t)$ modulates $\sqrt{P_s}\sin(2\pi f_0 t)$. These modulators are balanced modulator. The two carriers $\sqrt{P_s}\cos(2\pi f_0 t)$ and $\sqrt{P_s}\sin(2\pi f_0 t)$ are shown in Fig. 5.4.2 (d) and (e). These carriers are also called quadrature carriers. The two modulated signals are,

$$s_e(t) = b_e(t)\sqrt{P_s}\sin(2\pi f_0 t) \qquad \text{... (5.4.1)}$$

and

$$s_o(t) = b_o(t)\sqrt{P_s}\cos(2\pi f_0 t) \qquad \text{... (5.4.2)}$$

Thus $s_e(t)$ and $s_o(t)$ are basically BPSK signals and they are similar to equation 5.2.3 and equation 5.2.5. The only difference is that $T = 2T_b$ here. The value of $b_e(t)$ and $b_o(t)$ will be $+1V$ or $-1V$. Fig. 5.4.2 (f) and (g) shows the waveforms of $s_e(t)$ and $s_o(t)$.

Step 4 : Addition of modulated carriers

The adder of Fig. 5.4.1 adds these two signals $b_e(t)$ and $b_o(t)$. The output of the adder is OQPSK signal and it is given as,

$$\begin{aligned} s(t) &= s_o(t) + s_e(t) \\ &= b_o(t)\sqrt{P_s}\cos(2\pi f_0 t) + b_e(t)\sqrt{P_s}\sin(2\pi f_0 t) \end{aligned} \qquad \text{... (5.4.3)}$$

Step 5 : QPSK signal and phase shift

Fig. 5.4.2 (h) shows the QPSK signal represented by above equation. In QPSK signal of Fig. 5.4.2 (h), if there is any phase change, it occurs at minimum duration of T_b. This is because the two signals $s_e(t)$ and $s_o(t)$ have an offset of 'T_b'. Because of this offset, the phase shift in QPSK signal is $\frac{\pi}{2}$. It is clear from the waveforms of Fig. 5.4.2 that $b_e(t)$ and $b_o(t)$ cannot change at the same time because of offset between them. Fig. 5.4.3 shows the phasor diagram of QPSK signal of equation 5.4.2.

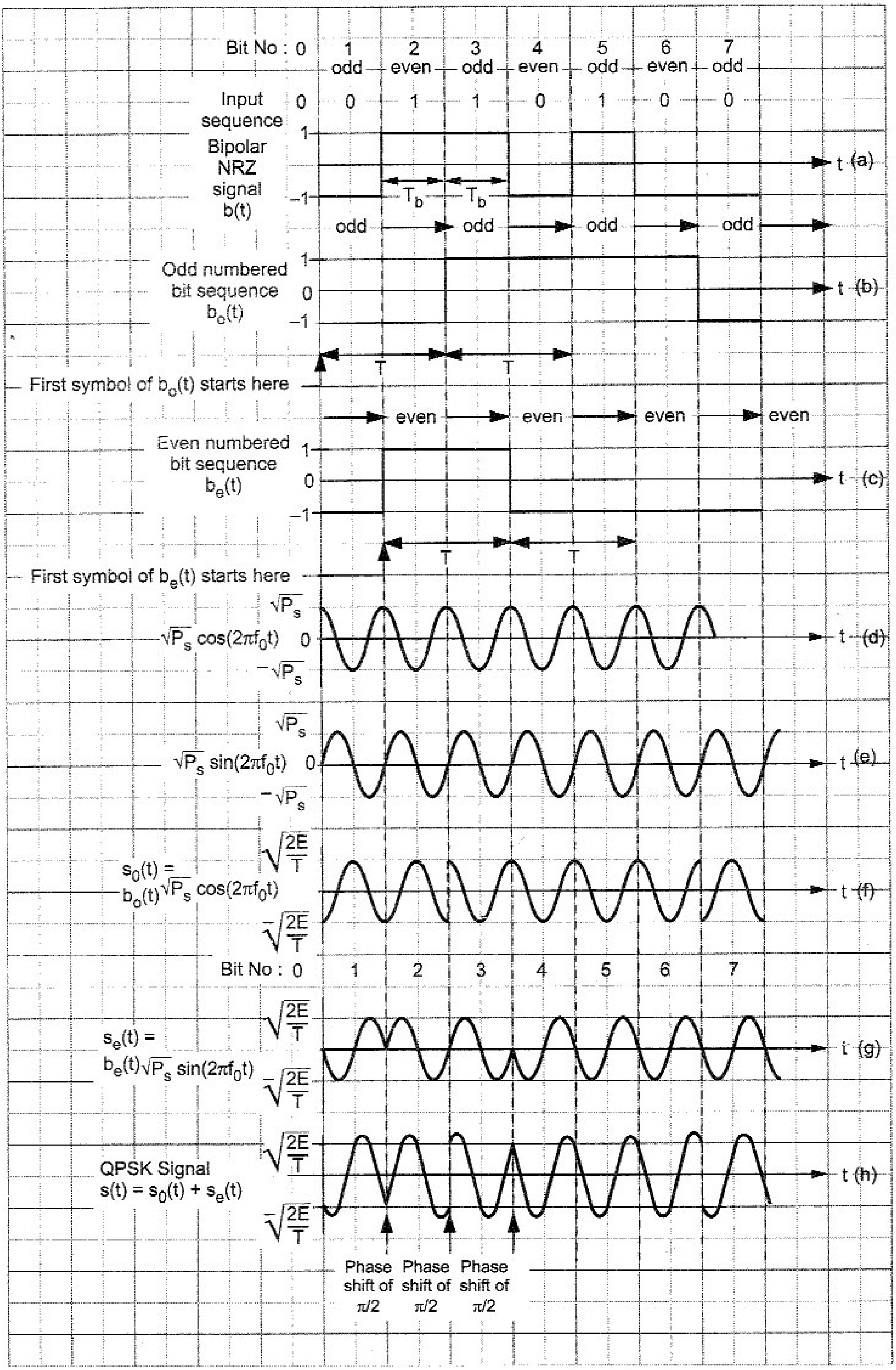


Fig. 5.4.2 QPSK waveforms (a) Input sequence and its NRZ waveform (b) Odd numbered bit sequence and its NRZ waveform (c) Even numbered bit sequence and its NRZ waveform (d) Basis function $\phi_1(t)$ (e) Basis function $\phi_2(t)$ (f) Binary PSK waveform for odd numbered channel (g) Binary PSK waveform for even numbered channel (h) Final QPSK waveform representing equation

Since $b_o(t)$ and $b_e(t)$ cannot change at the same time, the phase change in QPSK signal will be maximum $\pi / 2$. This is clear from Fig. 5.4.3.

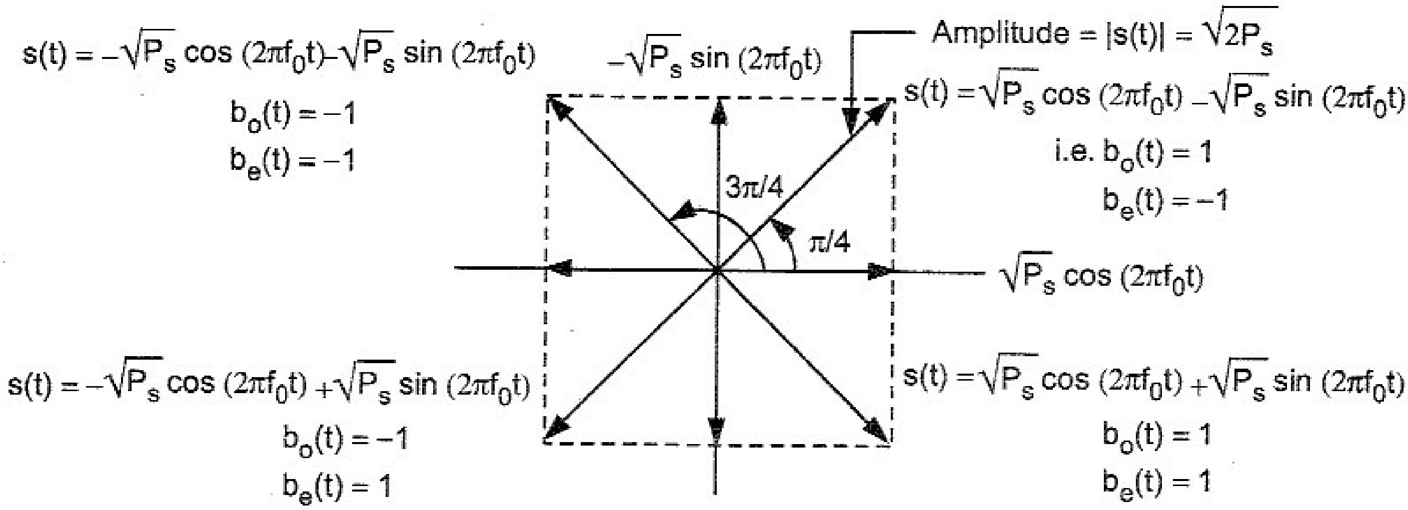


Fig. 5.4.3 Phasor diagram of QPSK signal

5.4.1.2 Non-Offset QPSK

- We known that there is an offset of 'T_b' between $b_e(t)$ and $b_o(t)$. If we delay $b_e(t)$ by 'T_b' then there will be no offset. Then the sequences $b_e(t)$ and $b_o(t)$ will change at the same time. This change will occur after minimum of '$2T_b$'.
- As a result, the signals $s_o(t)$ and $s_e(t)$ will have phase shifts at the same time. The individual phase shifts of $s_o(t)$ and $s_e(t)$ are 180°. Because of this the amplitude variations in the waveform will occur at the same time in $s_o(t)$ and $s_e(t)$. Therefore these variations will be more pronounced in non offset QPSK than OQPSK.
- Filters are used to suppress side bands in QPSK. Since phase changes by 180° in non offset QPSK, amplitude changes are more. Hence filtering affects the amplitude of non-offset QPSK. In OQPSK, the phase changes by 90°, hence amplitude changes during filtering are less.
- Since amplitude variations are more in non-offset QPSK, the signal is affected if communication takes place through repeators. These repeators highly affect the amplitude and phase of the QPSK signal.

5.4.1.3 The QPSK Receiver

April/May-2004; April/May-2005

Fig. 5.4.4 shows the QPSK receiver. This is synchronous reception. Therefore coherent carrier is to be recovered from the received signal $s(t)$.

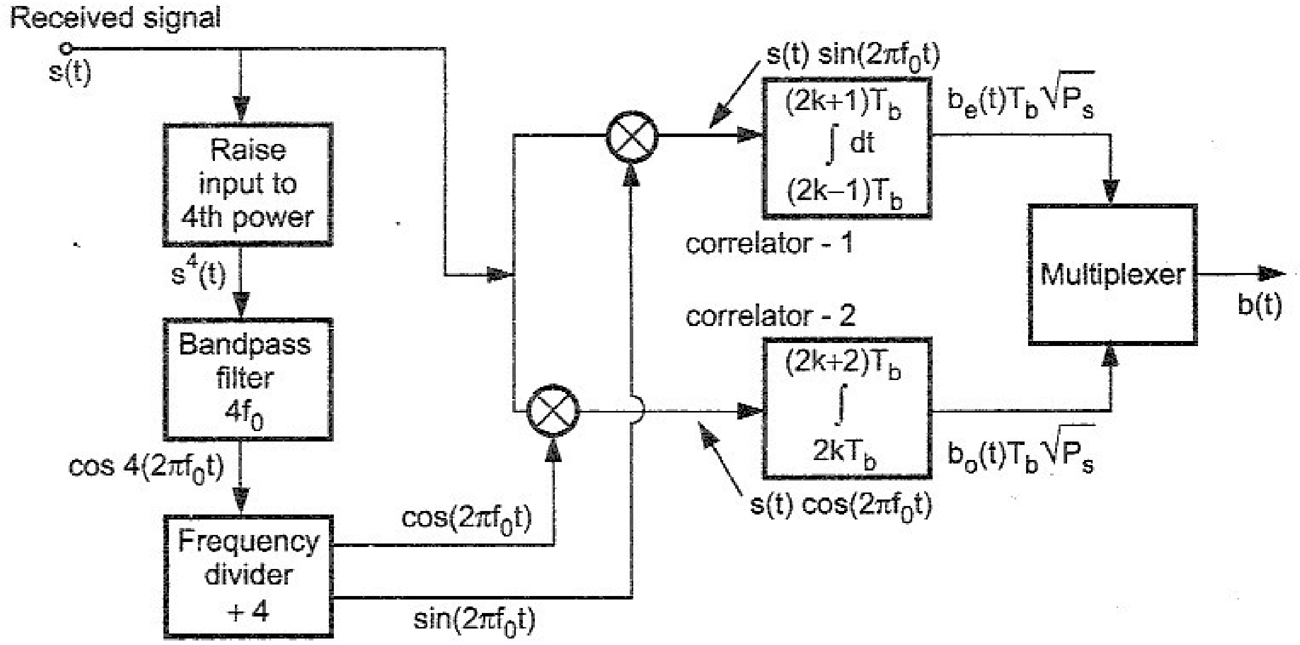


Fig. 5.4.4 QPSK receiver

Operation

Step 1 : Isolation of carrier

The received signal $s(t)$ is first raised to its 4^{th} power, i.e. $s^4(t)$. Then it is passed through a bandpass filter centered around $4f_0$. The output of the bandpass filter is a coherent carrier of frequency $4f_0$. This is divided by 4 and it gives two coherent quadrature carriers $\cos(2\pi f_0 t)$ and $\sin(2\pi f_0 (t)$.

Step 2 : Synchronous detection

These coherent carriers are applied to two synchronous demodulators. These synchronous demodulators consist of multiplier and an integrator.

Step 3 : Integration over two bits interval

The incoming signal is applied to both the multipliers. The integrator integrates the product signal over two bit interval (i.e. $T_s = 2T_b$).

Step 4 : Sampling and multiplexing odd and even bit sequences

At the end of this period, the output of integrator is sampled. The outputs of the two integrators are sampled at the offset of one bit period, T_b. Hence the output of multiplexer is the signal $b(t)$. That is, the odd and even sequences are combined by multiplexer.

To show that output of integrator depends upon respective bit sequence.

- Let's consider the product signal at the output of upper multiplier.

$$s(t) \sin(2\pi f_0 t) = b_o(t)\sqrt{P_s} \cos(2\pi f_0 t) \sin(2\pi f_0 t) + b_e(t)\sqrt{P_s} \sin^2(2\pi f_0 t) \quad \ldots (5.4.4)$$

- This signal is integrated by the upper integrator in Fig. 5.4.4.

$$\therefore \int_{(2k-1)T_b}^{(2k+1)T_b} s(t) \sin(2\pi f_0 t)\, dt = b_o(t)\sqrt{P_s} \int_{(2k-1)T_b}^{(2k+1)T_b} \cos(2\pi f_0 t) \sin(2\pi f_0 t)\, dt$$

$$+ b_e(t)\sqrt{P_s} \int_{(2k-1)T_b}^{(2k+1)T_b} \sin^2(2\pi f_0 t)\, dt$$

Since $\frac{1}{2}\sin(2x) = \sin x \cdot \cos x$

and $\sin^2(x) = \frac{1}{2}[1-\cos(2x)]$

- Using the above two trigonometric identities in the above equation,

$$\int_{(2k-1)T_b}^{(2k+1)T_b} s(t) \sin(2\pi f_0 t)\, dt = \frac{b_o(t)\sqrt{P_s}}{2} \int_{(2k-1)T_b}^{(2k+1)T_b} \sin 4\pi f_0 t\, dt + \frac{b_e(t)\sqrt{P_s}}{2} \int_{(2k-1)T_b}^{(2k+1)T_b} 1 \cdot dt$$

$$-\frac{b_e(t)\sqrt{P_s}}{2} \int_{(2k-1)T_b}^{(2k+1)T_b} \cos 4\pi f_0 t\, dt$$

- In the above equation, the first and third integration terms involves integration of sinusoidal carriers over two bit period. They have full (integral number of) cycles over two bit period and hence integration will be zero.

$$\int_{(2k-1)T_b}^{(2k+1)T_b} s(t) \sin(2\pi f_0 t)\, dt = \frac{b_e(t)\sqrt{P_s}}{2} [t]_{(2k-1)T_b}^{(2k+1)T_b}$$

$$= \frac{b_e(t)\sqrt{P_s}}{2} \times 2T_b$$

$$= b_e(t)\sqrt{P_s}\, T_b \quad \ldots (5.4.5)$$

- Thus the upper integrator responds to even sequence only. Similarly we can obtain the output of lower integrator as $b_o(t)\sqrt{P_s}\, T_b$.

Eventhough bit synchronizer is not shown in Fig. 5.4.4, it is assumed to be present with the integrator to locate starting and ending times of integration. The multiplexer is also operated by bit synchronizer. The amplitudes of signals marked in Fig. 5.4.4 are arbitrary. They can change depending upon the gains of integrator.

Ambiguity in the output :

In Fig. 5.4.4 observe that even if the received signal is negative, the recovered carrier remains unaffected because of the 4^{th} power conversion of the signal. Therefore it will not be possible to determine whether the transmitted signals were positive or negative [i.e. $+b_e(t)$ or $-b_e(t)$ and $+b_o(t)$ or $-b_o(t)$]. This is phase ambiguity in output similar to BPSK. This problem can be recovered by employing differential encoding and decoding of $b(t)$.

5.4.1.4 Carrier Synchronization in QPSK

Both the carriers are to be synchronized properly in coherent detection in QPSK. Fig. 5.4.5 shows the PLL system for carrier synchronization in QPSK.

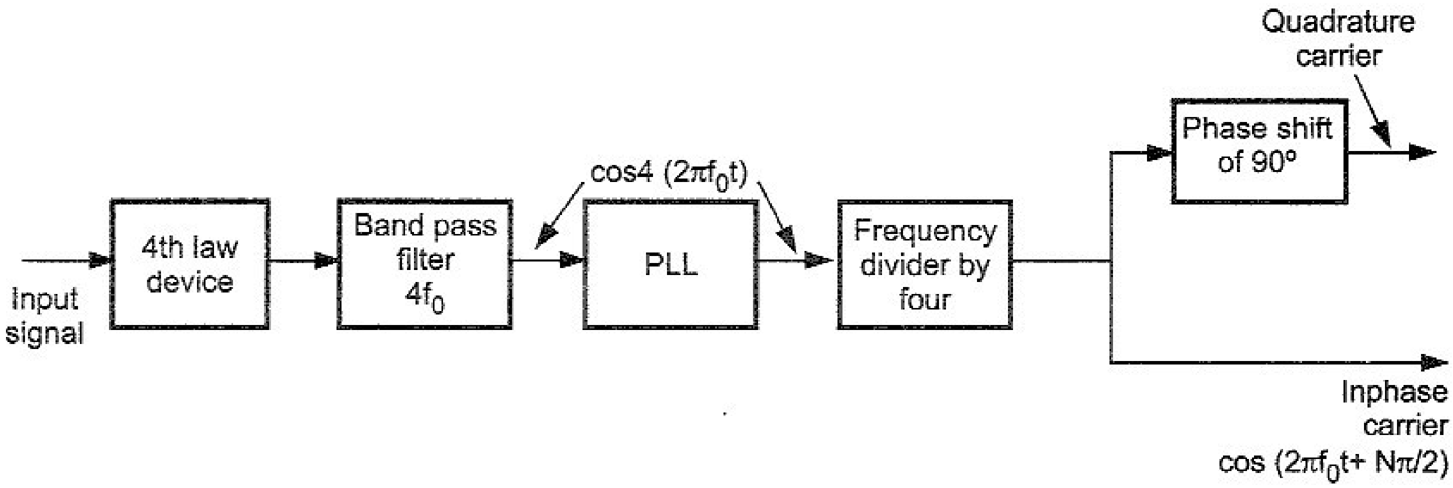


Fig. 5.4.5 PLL system for carrier synchronization

The fourth power of the input signal contains discrete frequency component at $4f_0$. We know that,

$$\cos 4(2\pi f_0 t) = \cos(8\pi f_0 t + 2\pi N)$$

Here 'N' is the number of cycles over the bit period. It is always integer value. When the frequency division by four takes place, the RHS of above equation becomes $\cos\left(2\pi f_0 t + \frac{N\pi}{2}\right)$. This shows that the output has a fixed phase error of $\frac{N\pi}{2}$. Differential encoding may be used to nullify the phase error events. The PLL remains locked with the phase of '$4f_0$' and then output of PLL is divided by 4. This gives a coherent carrier. A 90° phase shift is added to this carrier to generate a quadrature carrier.

5.4.2 Signal Space Representation of QPSK Signals

(1) Fig. 5.4.3 shows the phasor diagram of QPSK signal. Depending upon the combination of two successive bits, the phase shift occurs in carrier (see table 5.4.1). That is the QPSK signal of equation 5.4.3 can be written as,

$$s(t) = \sqrt{2P_s} \cos\left[2\pi f_0 t + (2m+1)\frac{\pi}{4}\right] \quad m = 0, 1, 2, 3 \qquad \dots (5.4.6)$$

Here, the above equation takes four values and they represent the phasors of Fig.5.4.3.

(2) The above equation can be expanded as,

$$s(t) = \sqrt{2P_s} \cos(2\pi f_0 t) \cos\left[(2m+1)\frac{\pi}{4}\right] - \sqrt{2P_s} \sin(2\pi f_0 t) \sin\left[(2m+1)\frac{\pi}{4}\right]$$

(3) Let's rearrange the above equation as,

$$s(t) = \left\{\sqrt{P_s T_s} \cos\left[(2m+1)\frac{\pi}{4}\right]\right\} \sqrt{\frac{2}{T_s}} \cos(2\pi f_0 t)$$

$$- \left\{\sqrt{P_s T_s} \sin\left[(2m+1)\frac{\pi}{4}\right]\right\} \sqrt{\frac{2}{T_s}} \sin(2\pi f_0 t) \qquad \dots (5.4.7)$$

(4) Let $\phi_1(t) = \sqrt{\frac{2}{T_s}} \cos(2\pi f_0 t)$ ... (5.4.8)

and $\phi_2(t) = \sqrt{\frac{2}{T_s}} \sin(2\pi f_0 t)$ ... (5.4.9)

The above two signals are called orthogonal signals and they are used as carriers in QPSK modulator.

(5) Let $b_o(t) = \sqrt{2} \cos\left[(2m+1)\frac{\pi}{4}\right]$ ... (5.4.10)

and $b_e(t) = -\sqrt{2} \sin\left[(2m+1)\frac{\pi}{4}\right]$ ... (5.4.11)

(6) With the use of equation 5.4.8 to equation 5.4.9 we can write equation 5.4.7 as,

$$s(t) = \sqrt{P_s T_s} \cdot \frac{1}{\sqrt{2}} b_o(t) \phi_1(t) + \sqrt{P_s T_s} \cdot \frac{1}{\sqrt{2}} b_e(t) \phi_2(t)$$

$$= \sqrt{P_s \cdot \frac{T_s}{2}} b_o(t) \phi_1(t) + \sqrt{P_s \frac{T_s}{2}} b_e(t) \phi_2(t)$$

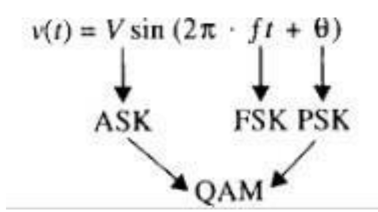
(7) T_s = symbol duration and $T_s = 2T_b$

UNIT- V (a)
DIGITAL MODULATION TECHNIQUES

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 ones.

There are many types of digital modulation techniques and we can even use a combination of these techniques as well. In this chapter, we will be discussing the most prominent digital modulation techniques.

if the information signal is digital and the amplitude (lV of the carrier is varied proportional to the information signal, a digitally modulated signal called amplitude shift keying (ASK) is produced.
If the frequency (f) is varied proportional to the information signal, frequency shift keying (FSK) is produced, and if the phase of the carrier (0) is varied proportional to the information signal,
phase shift keying (PSK) is produced. If both the amplitude and the phase are varied proportional to the information signal, quadrature amplitude modulation (QAM) results. ASK, FSK, PSK, and QAM are all forms of digital modulation:



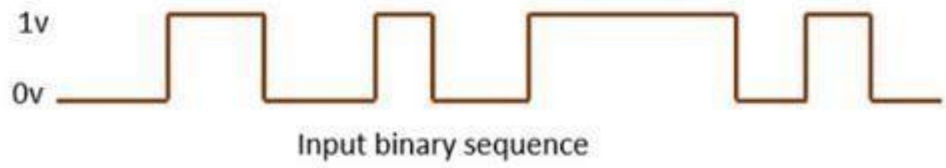
a simplified block diagram for a digital modulation system.

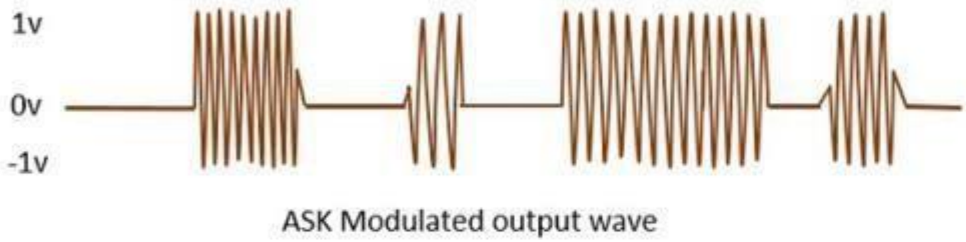
Amplitude Shift Keying

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.

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.

Following is the diagram for ASK modulated waveform along with its input.





Any modulated signal has a high frequency carrier. The binary signal when ASK is modulated, gives a zero value for LOW input and gives the carrier output for HIGH input.

Mathematically, amplitude-shift keying is

$$v_{(ask)}(t) = [1 + v_m(t)]\left[\frac{A}{2}\cos(\omega_c t)\right]$$

where vask(t) = amplitude-shift keying wave

vm(t) = digital information (modulating) signal (volts)

A/2 = unmodulated carrier amplitude (volts)

ωc= analog carrier radian frequency (radians per second, 2πfct)

In above Equation, the modulating signal [vm(t)] is a normalized binary waveform, where + 1 V = logic 1 and -1 V = logic 0. Therefore, for a logic 1 input, vm(t) = + 1 V, Equation 2.12 reduces to

$$v_{(ask)}(t) = [1 + 1]\left[\frac{A}{2}\cos(\omega_c t)\right]$$

$$= A\cos(\omega_c t)$$

Mathematically, amplitude-shift keying is (2.12) where vask(t) = amplitude-shift keying wave vm(t) = digital information (modulating) signal (volts) A/2 = unmodulated carrier amplitude (volts) ωc= analog carrier radian frequency (radians per second, 2πfct) In Equation 2.12, the modulating signal [vm(t)] is a normalized binary waveform, where + 1 V = logic 1 and -1 V = logic 0. Therefore, for a logic 1 input, vm(t) = + 1 V, Equation 2.12 reduces to and for a logic 0 input, vm(t) = -1 V,Equation reduces to

$$v_{(ask)}(t) = [1 - 1]\left[\frac{A}{2}\cos(\omega_c t)\right]$$

Thus, the modulated wave vask(t),is either A cos(ωct) or 0. Hence, the carrier is either "on "or "off," which is why amplitude-shift keying is sometimes referred to as on-off keying (OOK).

it can be seen that for every change in the input binary data stream, there is one change in the ASK waveform, and the time of one bit (tb) equals the time of one analog signaling element (t,).

B = fb/1 = fb   baud = fb/1 = fb

Example :

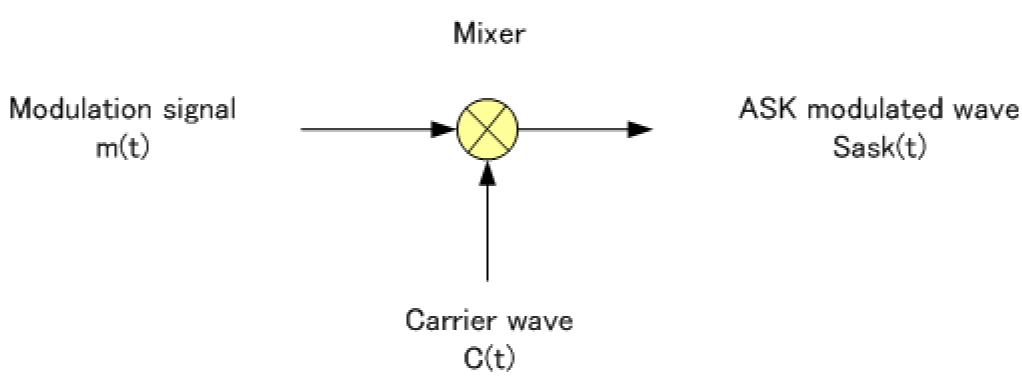
Determine the baud and minimum bandwidth necessary to pass a 10 kbps binary signal using amplitude shift keying. 10Solution For ASK, N = 1, and the baud and minimum bandwidth are determined from Equations 2.11 and 2.10, respectively:

B = 10,000 / 1 = 10,000

baud = 10, 000 /1 = 10,000

The use of amplitude-modulated analog carriers to transport digital information is a relatively low-quality, low-cost type of digital modulation and, therefore, is seldom used except for very low-speed telemetry circuits.

ASK TRANSMITTER:



The input binary sequence is applied to the product modulator. The product modulator amplitude modulates the sinusoidal carrier .it passes the carrier when input bit is '1' .it blocks the carrier when input bit is '0.'

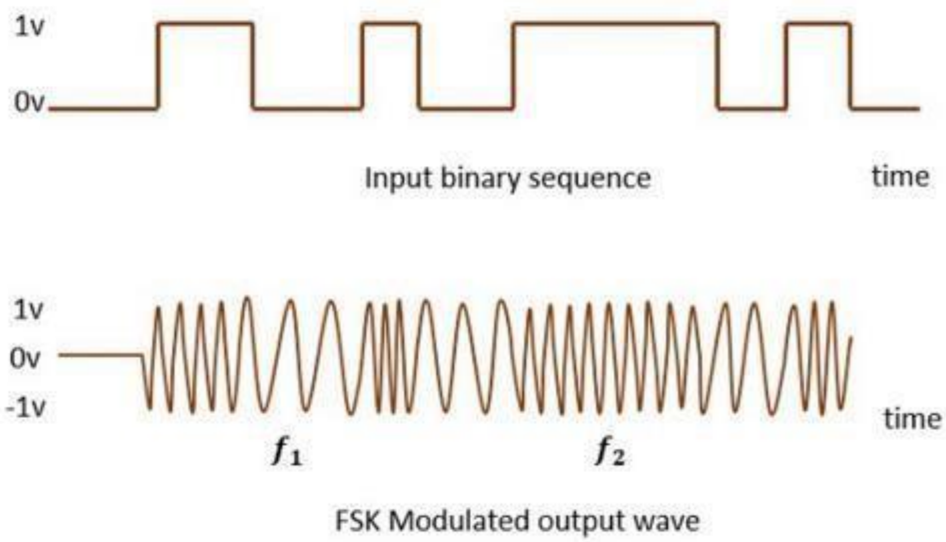
Coherent ASK DETECTOR:

FREQUENCYSHIFT KEYING

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

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

Following is the diagram for FSK modulated waveform along with its input.



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**.

FSK is a form of constant-amplitude angle modulation similar to standard frequency modulation (FM) except the modulating signal is a binary signal that varies between two discrete voltage levels rather than a continuously changing analog waveform.Consequently, FSK is sometimes called *binary FSK* (BFSK). The general expression for FSK is where $$v_{fsk}(t) = V_c \cos\{2\pi[f_c + v_m(t)\,\Delta f]t\}$$

$v_{fsk}(t)$ = binary FSK waveform

V_c = peak analog carrier amplitude (volts)

f_c = analog carrier center frequency(hertz)

f=peak change (shift)in the analog carrier frequency(hertz)

$v_m(t)$ = binary input (modulating) signal (volts)

From Equation 2.13, it can be seen that the peak shift in the carrier frequency (f) is proportional to the amplitude of the binary input signal ($v_m[t]$), and the direction of the shift is determined by the polarity.

The modulating signal is a normalized binary waveform where a logic 1 = + 1 V and a logic 0 = -1 V. Thus, for a logic l input, $v_m(t)$ = + 1, Equation 2.13 can be rewritten as

$$v_{fsk}(t) = V_c \cos[2\pi(f_c + \Delta f)t]$$

For a logic 0 input, $v_m(t)$ = -1, Equation becomes

$$v_{fsk}(t) = V_c \cos[2\pi(f_c - \Delta f)t]$$

With binary FSK, the carrier center frequency (f_c) is shifted (deviated) up and down in the frequency domain by the binary input signal as shown in Figure 2-3.

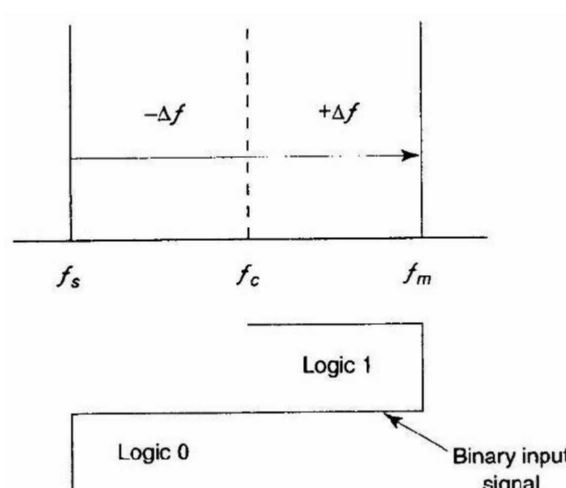


FIGURE: FSK in the frequency domain

As the binary input signal changes from a logic 0 to a logic 1 and vice versa, the output frequency shifts between two frequencies: a mark, or logic 1 frequency (f_m), and a space, or logic 0 frequency (f_s). The mark and space frequencies are separated from the carrier frequency by the peak frequency deviation (f) and from each other by 2 f.

Frequency deviation is illustrated in Figure 2-3 and expressed mathematically as

$$f = |f_m - f_s| / 2 \qquad (2.14)$$

where f = frequency deviation (hertz)

$|f_m - f_s|$ = absolute difference between the mark and space frequencies (hertz)

Figure 2-4a shows in the time domain the binary input to an FSK modulator and the corresponding FSK output.

When the binary input *(f_b)* changes from a logic 1 to a logic 0 and vice versa, the FSK output frequency shifts from a mark (f_m) to a space (f_s) frequency and vice versa.

In Figure 2-4a, the mark frequency is the higher frequency *(f_c* + f) and the space frequency is the lower frequency (f_c - f), although this relationship could be just the opposite.

Figure 2-4b shows the truth table for a binary FSK modulator. The truth table shows the input and output possibilities for a given digital modulation scheme.

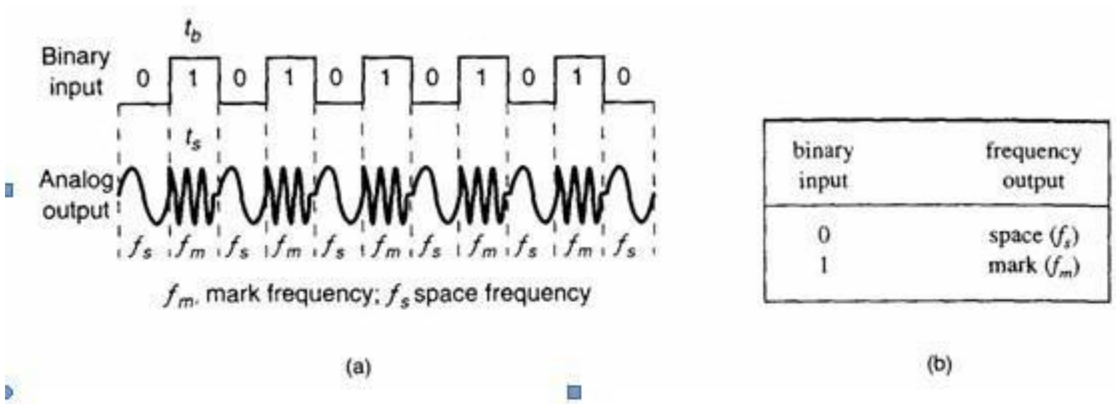


FIGURE 2-4 FSK in the time domain: (a) waveform: (b) truth table

FSK Bit Rate, Baud, and Bandwidth

In Figure 2-4a, it can be seen that the time of one bit *(t_b)* is the same as the time the FSK output is a mark of space frequency *(t_s)*. Thus, the bit time equals the time of an FSK signaling element, and the bit rate equals the baud.

The baud for binary FSK can also be determined by substituting N = 1 in Equation 2.11:

baud = f_b / 1 = f_b

The minimum bandwidth for FSK is given as

$$B= |(f_s - f_b) - (f_m - f_b)|$$

$$=|(f_s - f_m)| + 2f_b$$

and since $|(f_s - f_m)|$ equals 2 f, the minimum bandwidth can be approximated as

$$B= 2(f + f_b) \qquad (2.15)$$

where

B= minimum Nyquist bandwidth (hertz)

f= frequency deviation $|(f_m - f_s)|$ (hertz)

f_b = input bit rate (bps)

Example 2-2

Determine (a) the peak frequency deviation, (b) minimum bandwidth, and (c) baud for a binary FSK signal with a mark frequency of 49 kHz, a space frequency of 51 kHz, and an input bit rate of 2 kbps.

Solution

a. The peak frequency deviation is determined from Equation 2.14:

$$f= |149\text{kHz} - 51\text{ kHz}| / 2 =1\text{ kHz}$$

b. The minimum bandwidth is determined from Equation 2.15:

$$B = 2(100+ 2000)$$

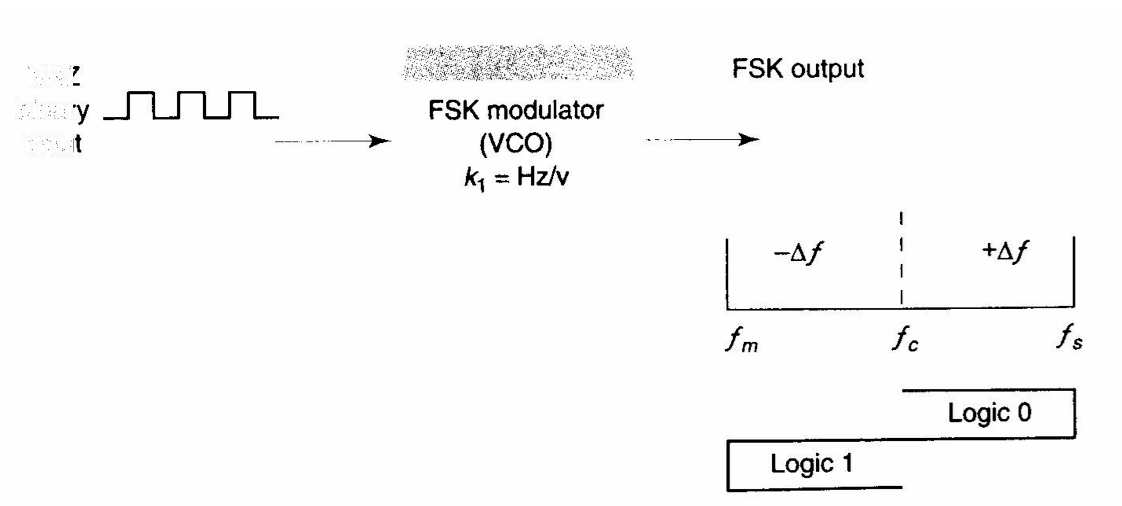
$$=6\text{ kHz}$$

c. For FSK, N = 1, and the baud is determined from Equation 2.11 as

$$\text{baud} = 2000 / 1 = 2000$$

FSK TRANSMITTER:

Figure 2-6 shows a simplified binary FSK modulator, which is very similar to a conventional FM modulator and is very often a voltage-controlled oscillator (VCO).The center frequency (f_c) is chosen such that it falls halfway between the mark and space frequencies.



A logic 1 input shifts the VCO output to the mark frequency, and a logic 0 input shifts the VCO output to the space frequency. Consequently, as the binary input signal changes back and forth between logic 1 and logic 0 conditions, the VCO output shifts or deviates back and forth between the mark and space frequencies.

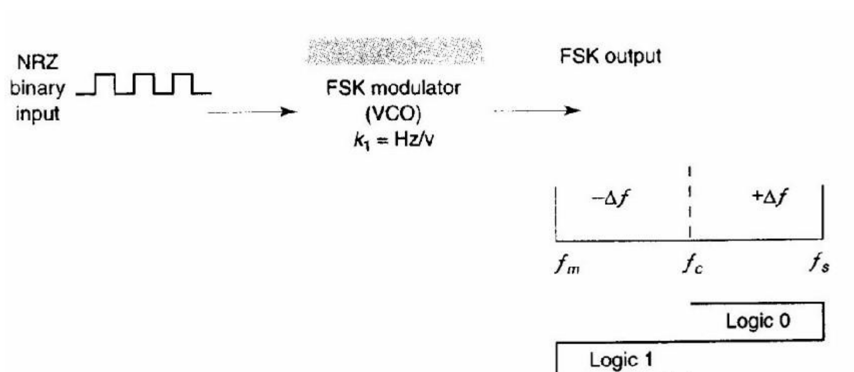


FIGURE 2-6 FSK modulator

A VCO-FSK modulator can be operated in the sweep mode where the peak frequency deviation is simply the product of the binary input voltage and the deviation sensitivity of the VCO.

With the sweep mode of modulation, the frequency deviation is expressed mathematically as

$$f = v_m(t)k_l \qquad (2\text{-}19)$$

$v_m(t)$ = peak binary modulating-signal voltage (volts)

k_l = deviation sensitivity (hertz per volt).

FSK Receiver

FSK demodulation is quite simple with a circuit such as the one shown in Figure 2-7.

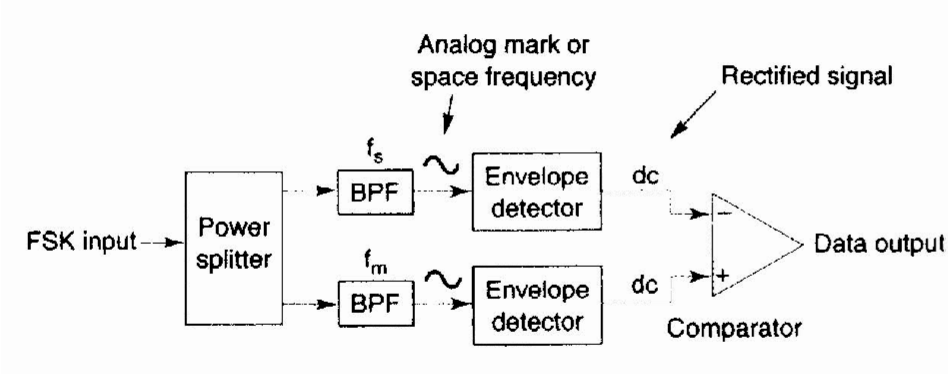


FIGURE 2-7 Noncoherent FSK demodulator

The FSK input signal is simultaneously applied to the inputs of both bandpass filters (BPFs) through a power splitter.The respective filter passes only the mark or only the space frequency on to its respective envelope detector.The envelope detectors, in turn, indicate the total power in each passband, and the comparator responds to the largest of the two powers.This type of FSK detection is referred to as noncoherent detection.

Figure 2-8 shows the block diagram for a coherent FSK receiver.The incoming FSK signal is multiplied by a recovered carrier signal that has the exact same frequency and phase as the transmitter reference.

However, the two transmitted frequencies (the mark and space frequencies) are not generally continuous; it is not practical to reproduce a local reference that is coherent with both of them. Consequently, coherent FSK detection is seldom used.

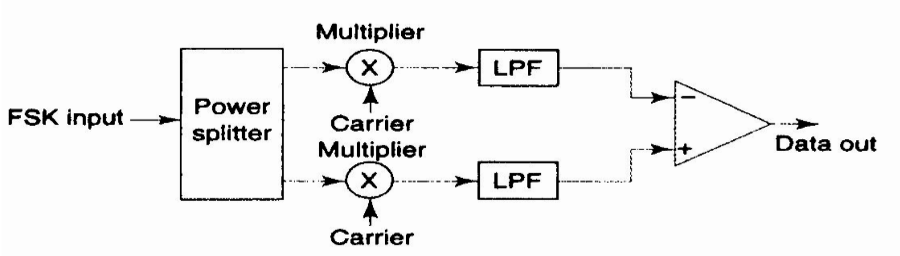
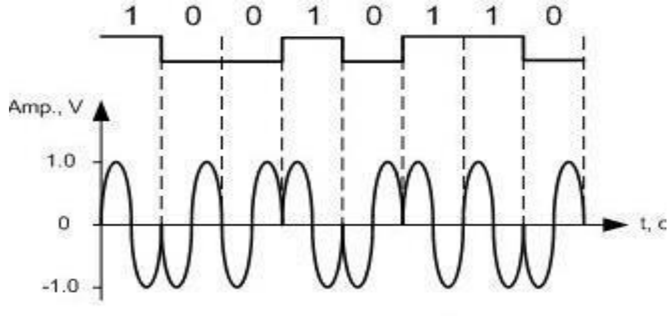


FIGURE 2-8 Coherent FSK demodulator

PHASESHIFT KEYING:

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



Phase shift keying (PSK)

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.

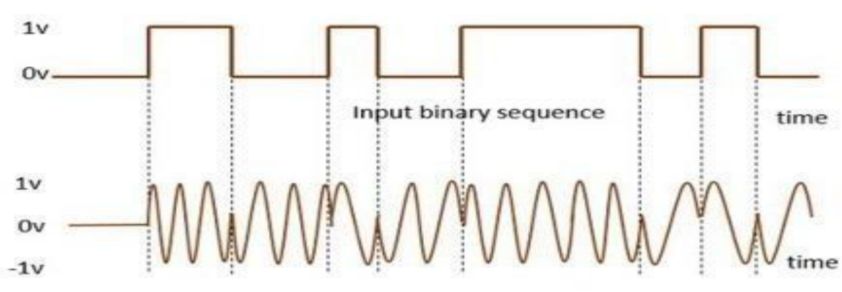
PSK is of two types, depending upon the phases the signal gets shifted. They are −

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 DSB-SC (Double Sideband Suppressed Carrier) modulation scheme, for message being the digital information.

Following is the image of BPSK Modulated output wave along with its input.



BPSK Modulated output wave

Binary Phase-Shift Keying

The simplest form of PSK is *binary phase-shift keying* (BPSK), where N = 1 and *M = 2.*Therefore, with BPSK, two phases ($2^1 = 2$) are possible for the carrier.One phase represents a logic 1, and the other phase represents a logic 0. As the input digital signal changes state (i.e., from a 1 to a 0 or from a 0 to a 1), the phase of the output carrier shifts between two angles that are separated by 180°.

Hence, other names for BPSK are *phase reversal keying* (PRK) and *biphase modulation.* BPSK is a form of square-wave modulation of a *continuous wave (CW)* signal.

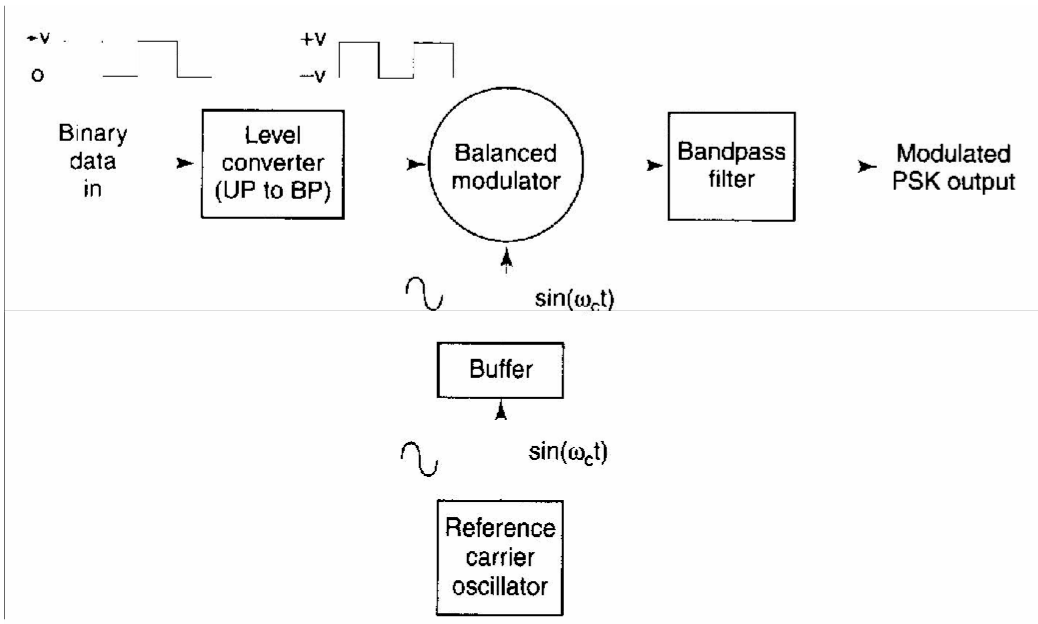


FIGURE 2-12 BPSK transmitter

BPSK TRANSMITTER:

Figure 2-12 shows a simplified block diagram of a BPSK transmitter. The balanced modulator acts as a phase reversing switch. Depending on the logic condition of the digital input, the carrier is transferred to the output either in phase or 180° out of phase with the reference carrier oscillator.

Figure 2-13 shows the schematic diagram of a balanced ring modulator. The balanced modulator has two inputs: a carrier that is in phase with the reference oscillator and the binary digital data. For the balanced modulator to operate properly, the digital input voltage must be much greater than the peak carrier voltage.

This ensures that the digital input controls the on/off state of diodes D1 to D4. If the binary input is a logic 1(positive voltage), diodes D 1 and D2 are forward biased and on, while diodes D3 and D4 are reverse biased and off (Figure 2-13b). With the polarities shown, the carrier voltage is developed across transformer T2 in phase with the carrier voltage across T

1. Consequently, the output signal is in phase with the reference oscillator.
If the binary input is a logic 0 (negative voltage), diodes Dl and D2 are reverse biased and off, while diodes D3 and D4 are forward biased and on (Figure 9-13c). As a result, the carrier voltage is developed across transformer T2 180° out of phase with the carrier voltage across T 1.

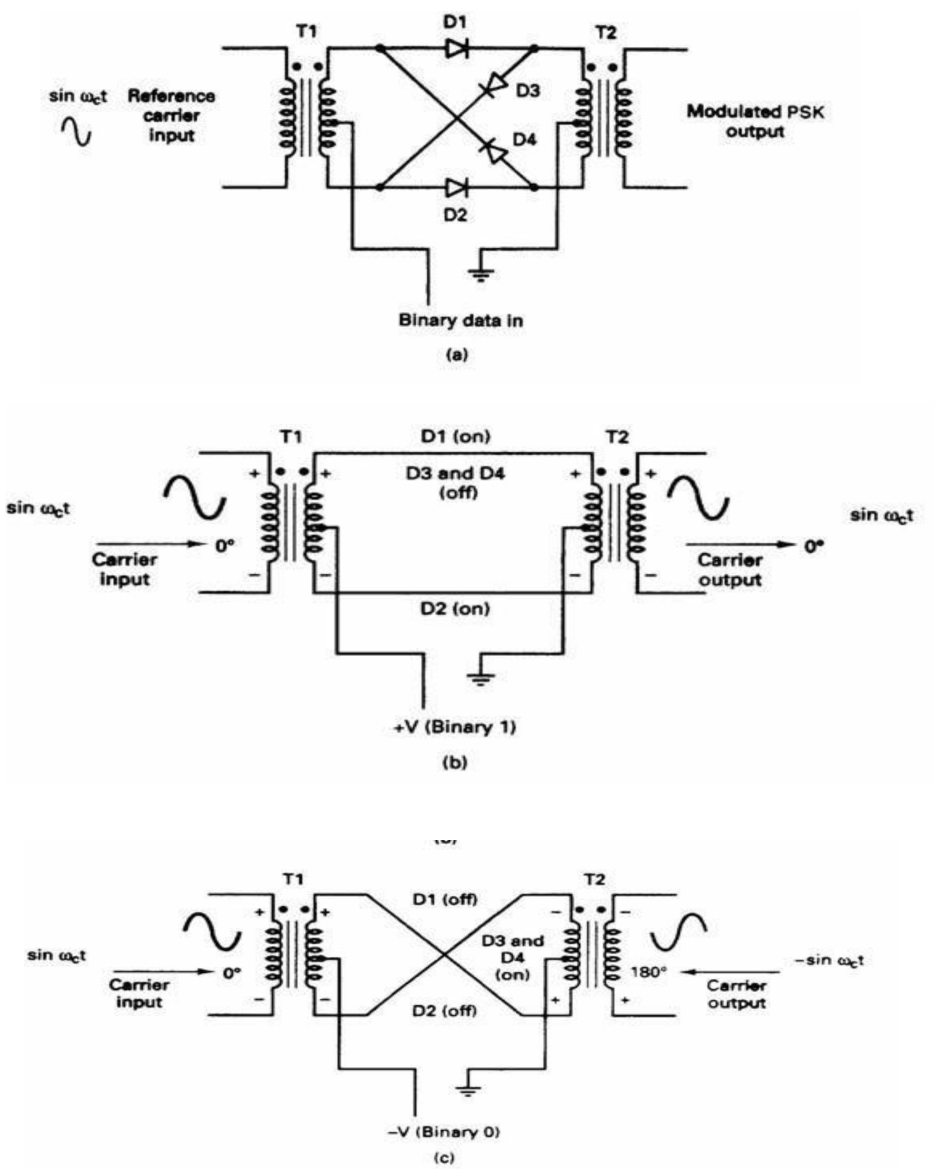


FIGURE 9-13 (a) Balanced ring modulator; (b) logic 1 input; (c) logic 0 input

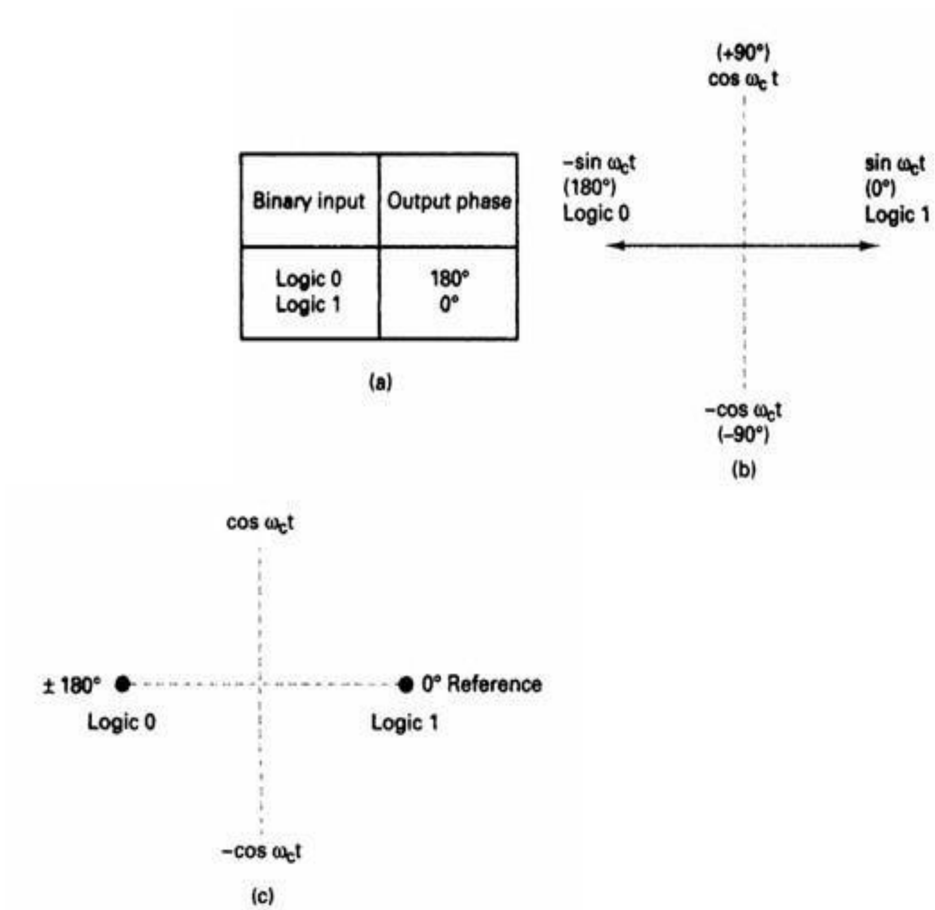


FIGURE 2-14 BPSK modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram

BANDWIDTH CONSIDERATIONS OF BPSK:

In a BPSK modulator. the carrier input signal is multiplied by the binary data.

If + 1 V is assigned to a logic 1 and -1 V is assigned to a logic 0, the input carrier (sin $\omega_c t$) is multiplied by either a + or - 1 .

The output signal is either + 1 sin $\omega_c t$ or -1 sin $\omega_c t$ the first represents a signal that is *in phase* with the reference oscillator, the latter a signal that is 180° out of phase with the reference oscillator.Each time the input logic condition changes, the output phase changes.

Mathematically, the output of a BPSK modulator is proportional to

$$\text{BPSK output} = [\sin(2\pi f_a t)] \times [\sin(2\pi f_c t)] \qquad (2.20)$$

where

f_a = maximum fundamental frequency of binary input (hertz)

f_c = reference carrier frequency (hertz)

Solving for the trig identity for the product of two sine functions,

$0.5\cos[2\pi(f_c - f_a)t] - 0.5\cos[2\pi(f_c + f_a)t]$

Thus, the minimum double-sided Nyquist bandwidth *(B) is*

$$\begin{array}{ccc} f_c + f_a & & f_c + f_a \\ -(f_c + f_a) & \text{or} & \underline{-f_c + f_a} \\ & & 2f_a \end{array}$$

and because $f_a = f_b / 2$, where f_b = input bit rate,

where *B is* the minimum double-sided Nyquist bandwidth.

Figure 2-15 shows the output phase-versus-time relationship for a BPSK waveform. Logic 1 input produces an analog output signal with a 0° phase angle, and a logic 0 input produces an analog output signal with a 180° phase angle.

As the binary input shifts between a logic 1 and a logic 0 condition and vice versa, the phase of the BPSK waveform shifts between 0° and 180°, respectively.

BPSK signaling element *(t_s)* is equal to the time of one information bit *(t_b)*, which indicates that the bit rate equals the baud.

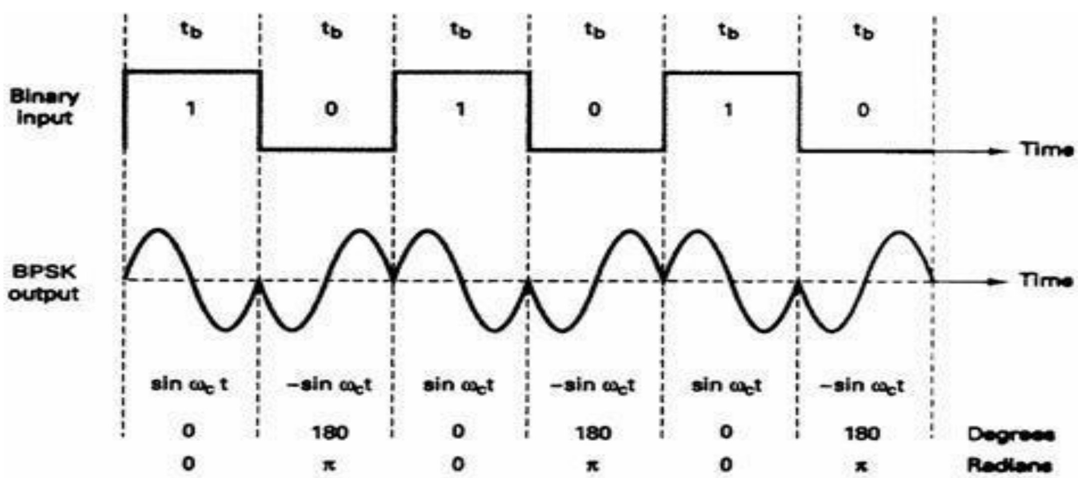


FIGURE 2-15 Output phase-versus-time relationship for a BPSK modulator

Example:

For a BPSK modulator with a carrier frequency of 70 MHz and an input bit rate of 10 Mbps, determine the maximum and minimum upper and lower side frequencies, draw the output spectrum, de-termine the minimum Nyquist bandwidth, and calculate the baud..

Solution

Substituting into Equation 2-20 yields

$$\text{output} = [\sin(2\pi f_a t)] \times [\sin(2\pi f_c t)];\ f_a = f_b / 2 = 5\text{ MHz}$$

$$= [\sin 2\pi(5\text{MHz})t)] \times [\sin 2\pi(70\text{MHz})t)]$$

$$= 0.5\cos[2\pi(70\text{MHz} - 5\text{MHz})t] - 0.5\cos[2\pi(70\text{MHz} + 5\text{MHz})t]$$

lower side frequency upper side frequency

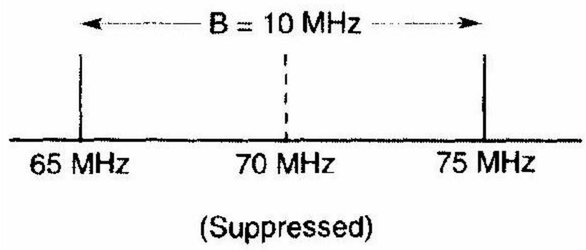
Minimum lower side frequency (LSF):

LSF=70MHz - 5MHz = 65MHz

Maximum upper side frequency (USF):

USF = 70 MHz + 5 MHz = 75 MHz

Therefore, the output spectrum for the worst-case binary input conditions *is* as follows: The minimum Nyquist bandwidth *(B) is*



B = 75 MHz - 65 MHz = 10 MHz

and the baud = f_b or 10 megabaud.

BPSK receiver:.

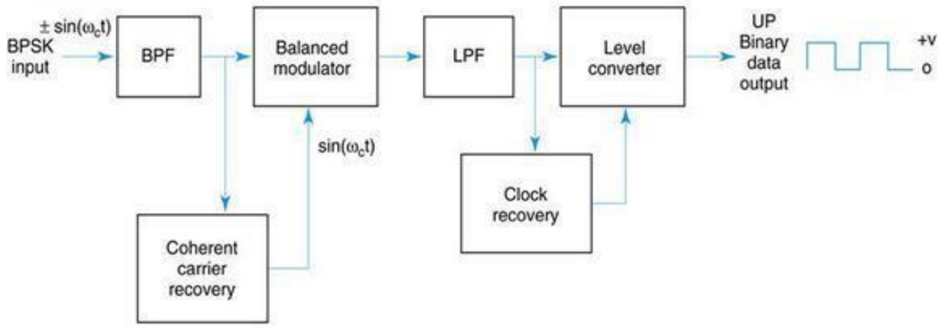
Figure 2-16 shows the block diagram of a BPSK receiver.

The input signal maybe+ sin $\omega_c t$ or - sin $\omega_c t$ .The coherent carrier recovery circuit detects and regenerates a carrier signal that is both frequency and phase coherent with the original transmit carrier.

The balanced modulator is a product detector; the output is the product d the two inputs (the BPSK signal and the recovered carrier).

The low-pass filter (LPF) operates the recovered binary data from the complex demodulated signal.

FIGURE 2-16 Block diagram of a BPSK receiver



Mathematically, the demodulation process is as follows.

For a BPSK input signal of + sin $\omega_c t$ (logic 1), the output of the balanced modulator is

$$\text{output} = (\sin \omega_c t)(\sin \omega_c t) = \sin^2\omega_c t \qquad (2.21)$$

or

$$\sin^2\omega_c t = 0.5(1 - \cos 2\omega_c t) = 0.5 - 0.5\cos 2\omega_c t$$

filtered out

leaving output = + 0.5 V = logic 1

It can be seen that the output of the balanced modulator contains a positive voltage (+[1/2]V) and a cosine wave at twice the carrier frequency (2 $\omega_c t$).

The LPF has a cutoff frequency much lower than 2 $\omega_c t$, and, thus, blocks the second harmonic of the carrier and passes only the positive constant component. A positive voltage represents a demodulated logic 1.

For a BPSK input signal of -sin $\omega_c t$ (logic 0), the output of the balanced modulator is

$$\text{output} = (-\sin \omega_c t)(\sin \omega_c t) = \sin^2\omega_c t$$

or

$$\sin^2\omega_c t = -0.5(1 - \cos 2\omega_c t) = 0.5 + 0.5\cos 2\omega_c t$$

filtered out

leaving

output = - 0.5 V = logic 0

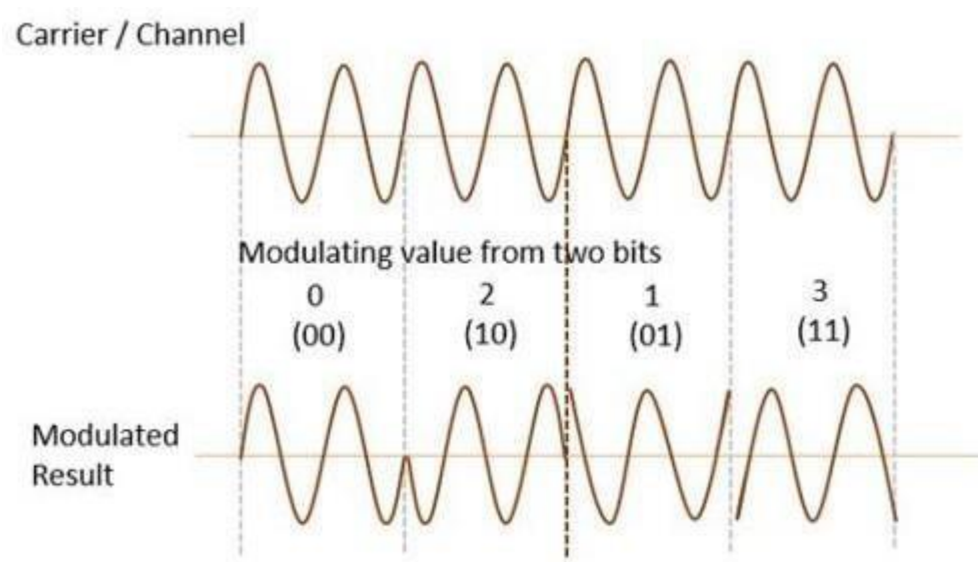
The output of the balanced modulator contains a negative voltage (-[l/2]V) and a cosine wave at twice the carrier frequency ($2\omega_c t$).

Again, the LPF blocks the second harmonic of the carrier and passes only the negative constant component. A negative voltage represents a demodulated logic 0.

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. The following figure represents the QPSK waveform for two bits input, which shows the modulated result for different instances of binary inputs.



QPSK is a variation of BPSK, and it is also a DSB-SC (Double Sideband Suppressed Carrier) 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 transmitter.

A block diagram of a QPSK modulator is shown in Figure 2-17Two bits (a dibit) are clocked into the bit splitter. After both bits have been serially inputted, they are simultaneously parallel outputted.

The I bit modulates a carrier that is in phase with the reference oscillator (hence the name "I" for "in phase" channel), and theQ bit modulate, a carrier that is 90° out of phase.

For a logic 1 = + 1 V and a logic 0= - 1 V, two phases are possible at the output of the I balanced modulator (+sin $\omega_c t$ and - sin $\omega_c t$), and two phases are possible at the output of the Q balanced modulator (+cos $\omega_c t$), and (-cos $\omega_c t$).

When the linear summer combines the two quadrature (90° out of phase) signals, there are four possible resultant phasors given by these expressions: + sin $\omega_c t$ + cos $\omega_c t$, + sin $\omega_c t$ - cos $\omega_c t$, -sin $\omega_c t$ + cos $\omega_c t$, and -sin $\omega_c t$ - cos $\omega_c t$.

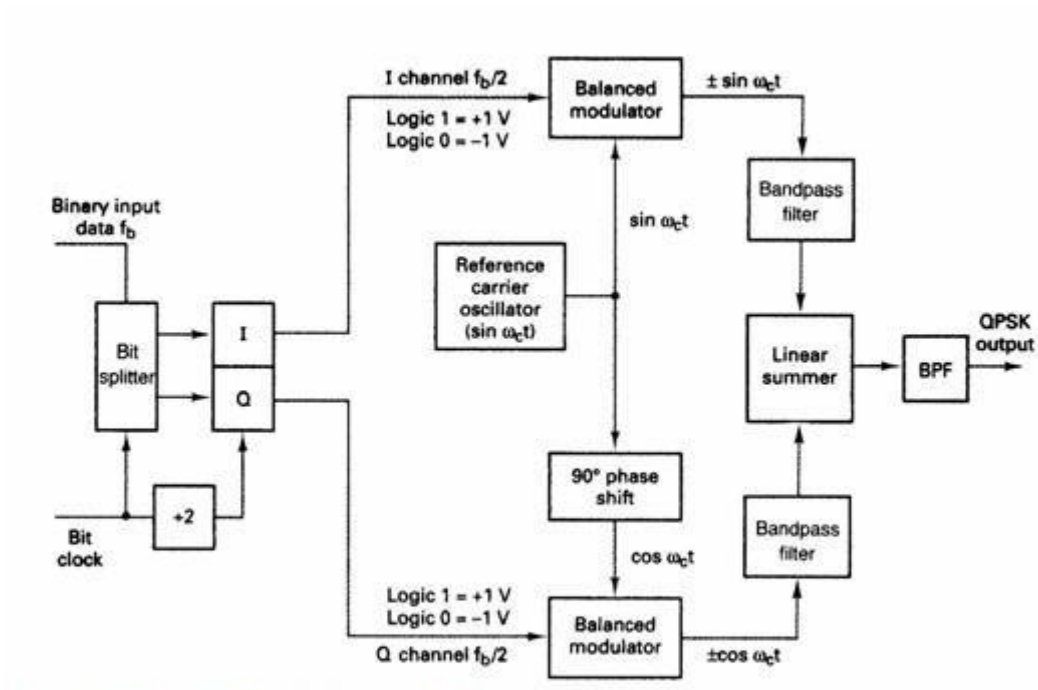


FIGURE 2-17 QPSK modulator

Example:

For the QPSK modulator shown in Figure 2-17, construct the truthtable, phasor diagram, and constellation diagram.

Solution

For a binary data input of Q = O and I= 0, the two inputs to the Ibalanced modulator are -1 and sin $\omega_c t$, and the two inputs to the Q balanced modulator are -1 and cos $\omega_c t$.

Consequently, the outputs are

I balanced modulator $=(-1)(\sin \omega_c t) = -1 \sin \omega_c t$

Q balanced modulator $=(-1)(\cos \omega_c t) = -1 \cos \omega_c t$ and the output of the linear summer is
$-1 \cos \omega_c t - 1 \sin \omega_c t = 1.414 \sin(\omega_c t - 135°)$

For the remaining dibit codes (01, 10, and 11), the procedure is the same. The results are shown in Figure 2-18a.

Binary input Q	Binary input I	QPSK output phase
0	0	−135°
0	1	−45°
1	0	+135°
1	1	+45°

(a)

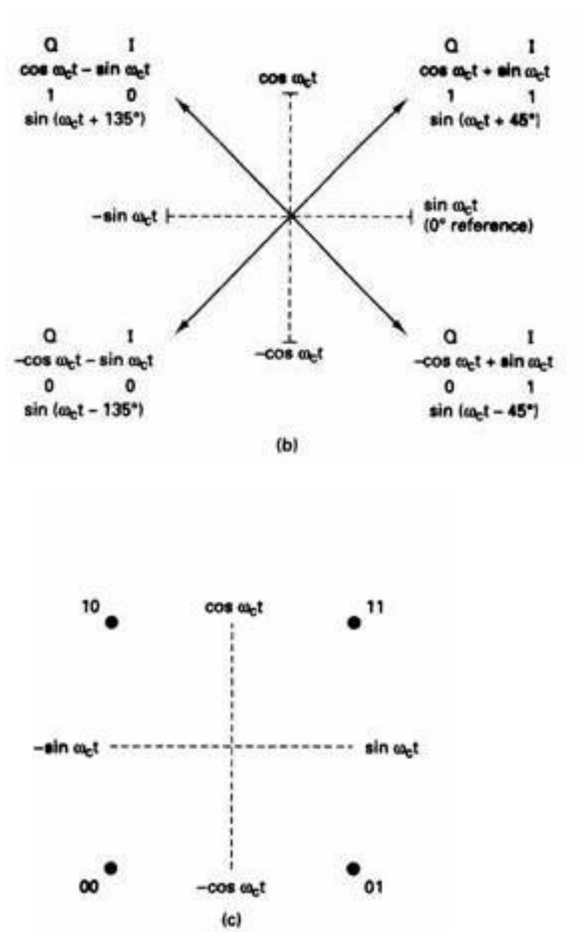


FIGURE 2-18 QPSK modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram

In Figures 2-18b and c, it can be seen that with QPSK each of the four possible output phasors has exactly the same amplitude. Therefore, the binary information must be encoded entirely in the phase of the output signal

Figure 2-18b, it can be seen that the angular separation between any two adjacent phasors in QPSK is 90°.Therefore, a QPSK signal can undergo almost a+45° or -45° shift in phase during transmission and still retain the correct encoded information when demodulated at the receiver.

Figure 2-19 shows the output phase-versus-time relationship for a QPSK modulator.

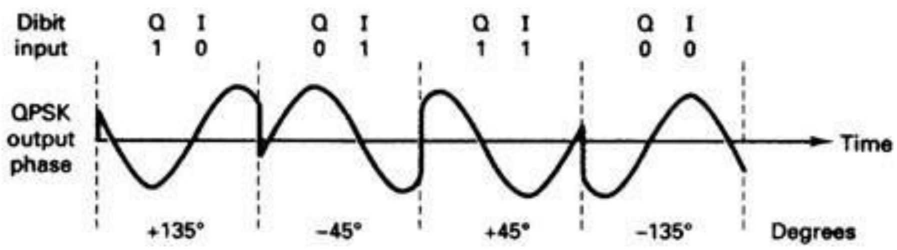


FIGURE 2-19 Output phase-versus-time relationship for a PSK modulator

Bandwidth considerations of QPSK

With QPSK, because the input data are divided into two channels, the bit rate in either the I or the Q channel is equal to one-half of the input data rate ($f_b/2$) (one-half of $f_b/2 = f_b/4$).

QPSK RECEIVER:

The block diagram of a QPSK receiver is shown in Figure 2-21

The power splitter directs the input QPSK signal to the I and Q product detectors and the carrier recovery circuit. The carrier recovery circuit reproduces the original transmit carrier oscillator signal. The recovered carrier must be frequency and phase coherent with the transmit reference carrier. The QPSK signal is demodulated in the I and Q product detectors, which generate the original I and Q data bits. The outputs of the product detectors are fed to the bit combining circuit, where they are converted from parallel I and Q data channels to a single binary output data stream. The incoming QPSK signal may be any one of the four possible output phases shown in Figure 2-18. To illustrate the demodulation process, let the incoming QPSK signal be $-\sin \omega_c t + \cos \omega_c t$. Mathematically, the demodulation process is as follows.

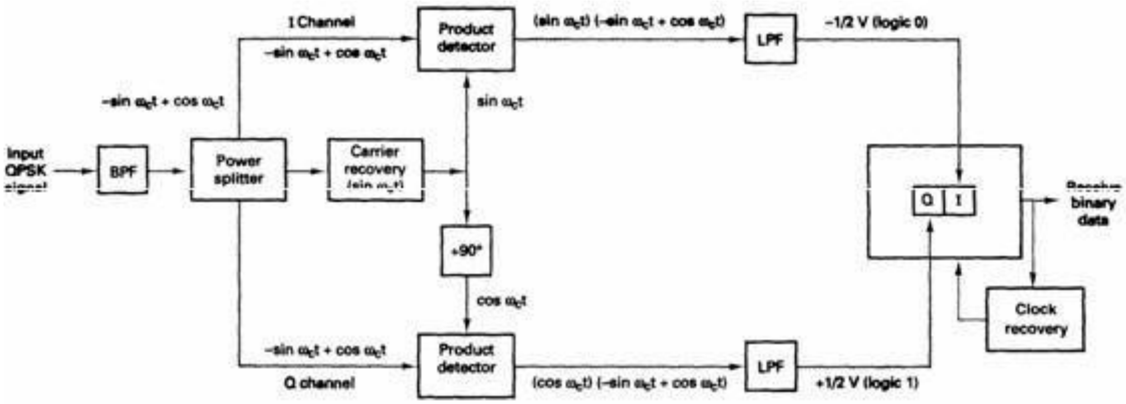


FIGURE 2-21 QPSK receiver

The receive QPSK signal (-sin ω_ct + cos ω_ct) is one of the inputs to the I product detector. The other input is the recovered carrier (sin ω_ct). The output of the I product detector is

$$I = \underbrace{(-\sin \omega_c t + \cos \omega_c t)}_{\text{QPSK input signal}} \underbrace{(\sin \omega_c t)}_{\text{carrier}}$$

$$= (-\sin \omega_c t)(\sin \omega_c t) + (\cos \omega_c t)(\sin \omega_c t)$$

$$= -\sin^2 \omega_c t + (\cos \omega_c t)(\sin \omega_c t)$$

$$= -\frac{1}{2}(1 - \cos 2\omega_c t) + \frac{1}{2}\sin(\omega_c + \omega_c)t + \frac{1}{2}\sin(\omega_c - \omega_c)t$$

$$I = -\frac{1}{2} + \underbrace{\frac{1}{2}\cos 2\omega_c t}_{\text{(filtered out)}} + \frac{1}{2}\sin 2\omega_c t + \underbrace{\frac{1}{2}\sin 0}_{\text{(equals 0)}}$$

$$= -\frac{1}{2}\text{V (logic 0)} \tag{2.23}$$

Again, the receive QPSK signal (-sin ω_ct + cos ω_ct) is one of the inputs to the Q product detector. The other input is the recovered carrier shifted 90° in phase (cos ω_ct). The output of the Q product detector is

$$Q = \underbrace{(-\sin \omega_c t + \cos \omega_c t)}_{\text{QPSK input signal}} \underbrace{(\cos \omega_c t)}_{\text{carrier}}$$

$$= \cos^2 \omega_c t - (\sin \omega_c t)(\cos \omega_c t)$$

$$= \frac{1}{2}(1 + \cos 2\omega_c t) - \frac{1}{2}\sin(\omega_c + \omega_c)t - \frac{1}{2}\sin(\omega_c - \omega_c)t$$

$$Q = \frac{1}{2} + \underbrace{\frac{1}{2}\cos 2\omega_c t}_{\text{(filtered out)}} - \frac{1}{2}\sin 2\omega_c t - \underbrace{\frac{1}{2}\sin 0}_{\text{(equals 0)}}$$

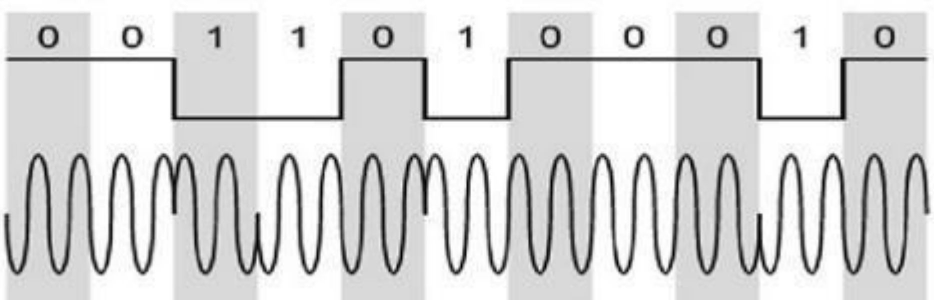
$$= \frac{1}{2}\text{V(logic 1)} \qquad (2.24)$$

The demodulated I and Q bits (0 and 1, respectively) correspond to the constellation diagram and truth table for the QPSK modulator shown in Figure 2-18.

DIFFERENTIAL PHASE SHIFT KEYING (DPSK):

In DPSK (Differential Phase Shift Keying) the phase of the modulated signal is shifted relative to the previous signal element. No reference signal is considered here. The signal phase follows the high or low state of the previous element. This DPSK technique doesn’t need a reference oscillator.

The following figure represents the model waveform of DPSK.



It is seen from the above figure that, if the data bit is LOW i.e., 0, then the phase of the signal is not reversed, but is continued as it was. If the data is HIGH i.e., 1, then the phase of the signal is reversed, as with NRZI, invert on 1 (a form of differential encoding).

If we observe the above waveform, we can say that the HIGH state represents an **M** in the modulating signal and the LOW state represents a **W** in the modulating signal.

The word binary represents two-bits. **M** simply represents a digit that corresponds to the number of conditions, levels, or combinations possible for a given number of binary variables.

This is the type of digital modulation technique used for data transmission in which instead of one-bit, two or **more bits are transmitted at a time**. As a single signal is used for multiple bit transmission, the channel bandwidth is reduced.

DBPSK TRANSMITTER.:

Figure 2-37a shows a simplified block diagram of a *differential binary phase-shift keying* (DBPSK) transmitter. An incoming information bit is XNORed with the preceding bit prior to entering the BPSK modulator (balanced modulator).

For the first data bit, there is no preceding bit with which to compare it. Therefore, an initial reference bit is assumed. Figure 2-37b shows the relationship between the input data, the XNOR output data, and the phase at the output of the balanced modulator. If the initial reference bit is assumed a logic 1, the output from the XNOR circuit is simply the complement of that shown.

In Figure 2-37b, the first data bit is XNORed with the reference bit. If they are the same, the XNOR output is a logic 1; if they are different, the XNOR output is a logic 0. The balanced modulator operates the same as a conventional BPSK modulator; a logic I produces $+\sin \omega_c t$ at the output, and A logic 0 produces $-\sin \omega_c t$ at the output.

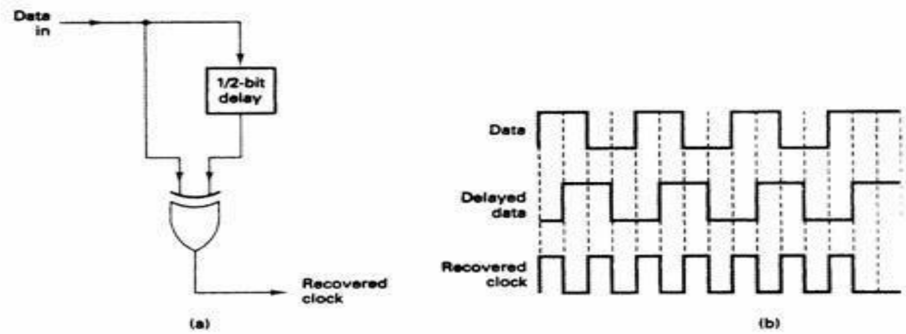


FIGURE 9-40 (a) Clock recovery circuit; (b) timing diagram

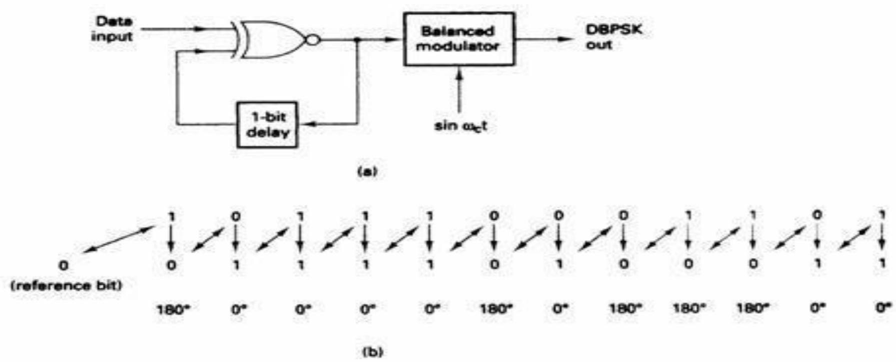


FIGURE 2-37 DBPSK modulator (a) block diagram (b) timing diagram

BPSK RECEIVER:

Figure 9-38 shows the block diagram and timing sequence for a DBPSK receiver. The received signal is delayed by one bit time, then compared with the next signaling element in the balanced modulator. If they are the same. J logic 1(+ voltage) is generated. If they are different, a logic 0 (-voltage) is generated. [f the reference phase is incorrectly assumed, only the first demodulated bit is in error. Differential encoding can be implemented with higher-than-binary digital modulation schemes, although the differential algorithms are much more complicated than for DBPS K.

The primary advantage of DBPSK is the simplicity with which it can be implemented. With DBPSK, no carrier recovery circuit is needed. A disadvantage of DBPSK is, that it requires between 1 dB and 3 dB more signal-to-noise ratio to achieve the same bit error rate as that of absolute PSK.

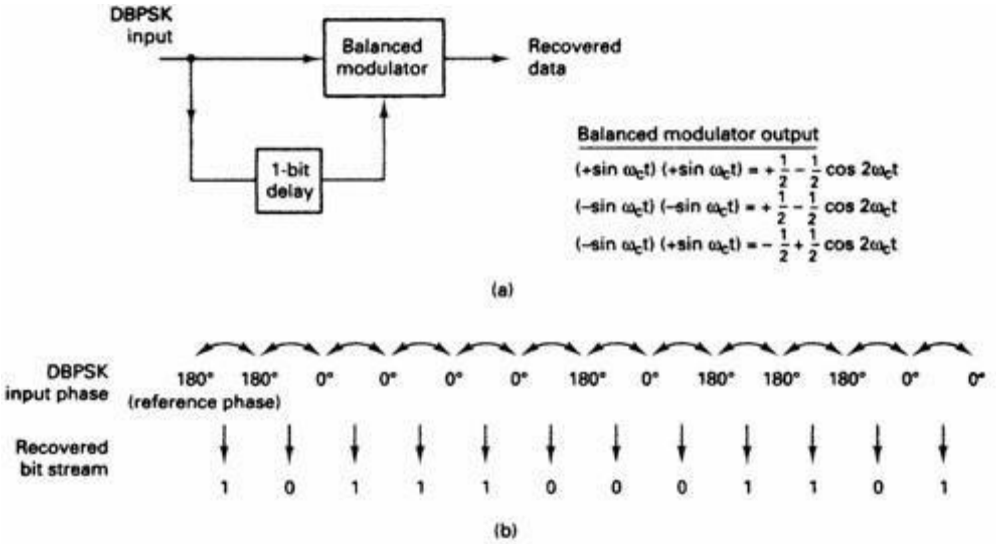


FIGURE 2-38 DBPSK demodulator: (a) block diagram; (b) timing sequence

COHERENT RECEPTION OF FSK:

The coherent demodulator for the coherent FSK signal falls in the general form of coherent demodulators described in Appendix B. The demodulator can be implemented with two correlators as shown in Figure 3.5, where the two reference signals are cos(27r f t) and cos(27r fit). They must be synchronized with the received signal. The receiver is optimum in the sense that it minimizes the error probability for equally likely binary signals. Even though the receiver is rigorously derived in Appendix B, some heuristic explanation here may help understand its operation. When s 1 (t) is transmitted, the upper correlator yields a signal 1 with a positive signal component and a noise component. However, the lower correlator output 12, due to the signals' orthogonality, has only a noise component. Thus the output of the summer is most likely above zero, and the threshold detector will most likely produce a 1. When s2(t) is transmitted, opposite things happen to the two correlators and the threshold detector will most likely produce a 0. However, due to the noise nature that its values range from -00 to m, occasionally the noise amplitude might overpower the signal amplitude, and then detection errors will happen. An alternative to Figure 3.5 is to use just one correlator with the reference signal cos (27r f t) - cos(2s f2t) (Figure 3.6). The correlator in Figure

can be replaced by a matched filter that matches cos(27r fit) - cos(27r f2t) (Figure 3.7). All implementations are equivalent in terms of error performance (see Appendix B). Assuming an AWGN channel, the received signal is

$$r(t) = s_i(t) + n(t), \qquad i = 1, 2$$

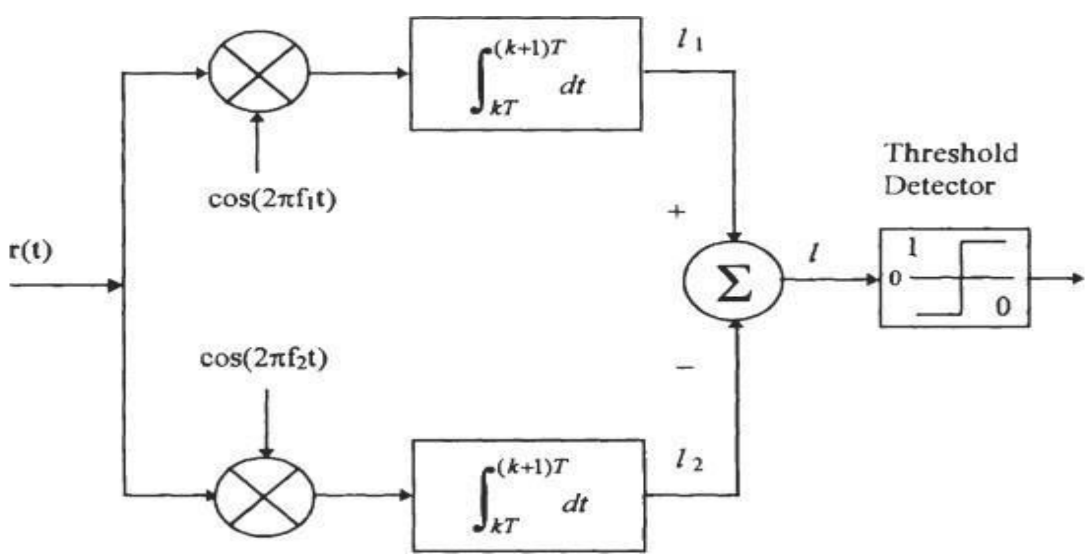
where n(t) is the additive white Gaussian noise with zero mean and a two-sided power spectral density A',/2. From (B.33) the bit error probability for any equally likely binary signals is

$$P_b = Q\left(\sqrt{\frac{E_1 + E_2 - 2\rho_{12}\sqrt{E_1 E_2}}{2N_o}}\right)$$

where No/2 is the two-sided power spectral density of the additive white Gaussian noise. For Sunde's FSK signals El = Ez = Eb, pI2 = 0 (orthogonal). thus the error probability is

$$P_b = Q\left(\sqrt{\frac{E_b}{N_o}}\right)$$

where Eb = A2T/2 is the average bit energy of the FSK signal. The above Pb is plotted in Figure 3.8 where Pb of noncoherently demodulated FSK, whose expression will be given shortly, is also plotted for comparison.



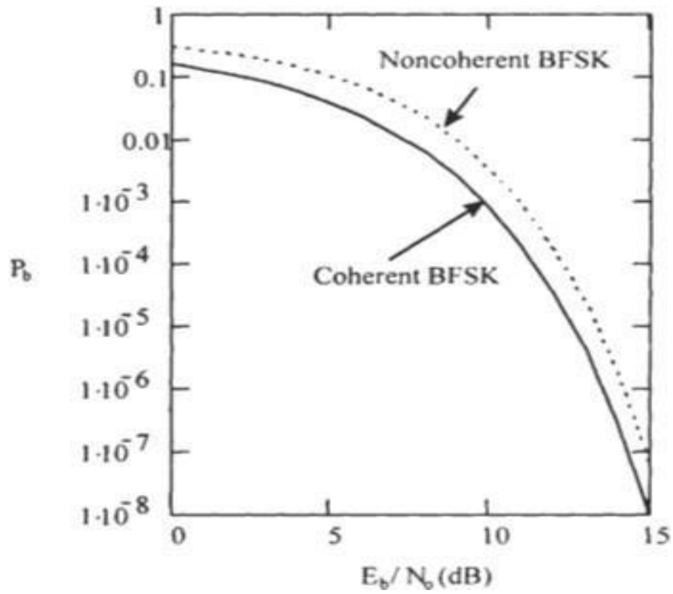


Figure: Pb of coherently and non-coherently demodulated FSK signal.

NONCOHERENT DEMODULATION AND ERROR PERFORMANCE:

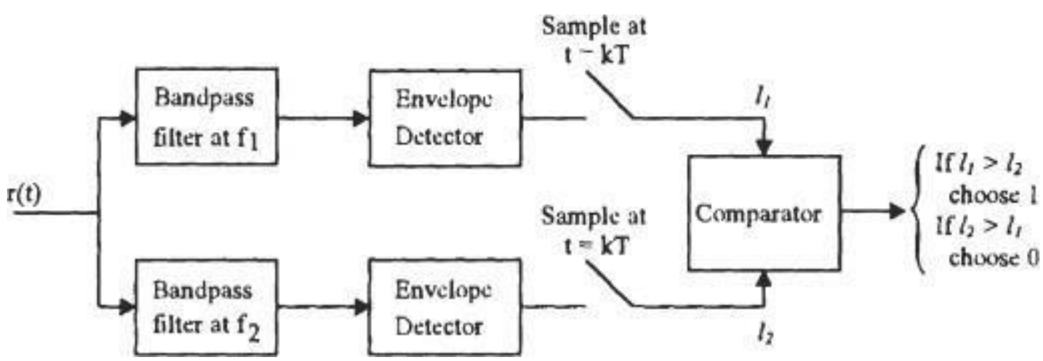
Coherently FSK signals can be noncoherently demodulated to avoid the carrier recovery. Noncoherently generated FSK can only be noncoherently demodulated. We refer to both cases as noncoherent FSK. In both cases the demodulation problem becomes a problem of detecting signals with unknown phases. In Appendix B we have shown that the optimum receiver is a quadrature receiver. It can be implemented using correlators or equivalently, matched filters. Here we assume that the binary noncoherent FSK signals are equally likely and with equal energies. Under these assumptions, the demodulator using correlators is shown in Figure 3.9. Again, like in the coherent case, the optimality of the receiver has been rigorously proved (Appendix B). However, we can easily understand its operation by some heuristic argument as follows. The received signal (ignoring noise for the moment) with an unknown phase can be written as

$$\begin{aligned} s_i(t,\theta) &= A\cos(2\pi f_i t + \theta), \qquad i = 1, 2 \\ &= A\cos\theta\cos 2\pi f_i t - A\sin\theta\sin 2\pi f_i t \end{aligned}$$

The signal consists of an in phase component A cos 8 cos 27r f t and a quadrature component A sin 8 sin 2x f,t sin 0. Thus the signal is partially correlated with cos 2s fit and partiah'y correlated with sin 27r fit. Therefore we use two correlators to collect the signal energy in these two parts. The outputs of the in phase and quadrature correlators will be cos 19 and sin 8, respectively. Depending on the value of the unknown phase 8, these two outputs could be anything in (- 5, y). Fortunately the squared sum of these two signals is not dependent on the unknown phase. That is

$$\left(\frac{AT}{2}\cos\theta\right)^2 + \left(\frac{AT}{2}\sin\theta\right)^2 = \frac{A^2T^2}{2}$$

This quantity is actually the mean value of the statistics I? when signal si (t) is transmitted and noise is taken into consideration. When si (t) is not transmitted the mean value of 1: is 0. The comparator decides which signal is sent by checking these I?. The matched filter equivalence to Figure 3.9 is shown in Figure 3.10 which has the same error performance. For implementation simplicity we can replace the matched filters by bandpass filters centered at f and fi, respectively (Figure 3.1 1). However, if the bandpass filters are not matched to the FSK signals, degradation to



various extents will result. The bit error probability can be derived using the correlator demodulator (Appendix B). Here we further assume that the FSK signals are orthogonal, then from Appendix B the error probability is

$$P_b = \frac{1}{2}e^{-E_b/2N_o}$$

Putting

$$z = \frac{1}{\sqrt{N_0}}(x_1 + \sqrt{E_b}) \tag{6.18}$$

and changing the variable of integration from x_1 to z, we may rewrite Equation (6.17) in the compact form

$$\begin{aligned} p_{10} &= \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/N_0}}^{\infty} \exp(-z^2)\, dz \\ &= \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \end{aligned} \tag{6.19}$$

where erfc(·) is the complementary error function.

Thus, averaging the conditional error probabilities p_{10} and p_{01}, we find that the *average probability of symbol error* or, equivalently, the *bit error rate for coherent binary PSK* is (assuming equiprobable symbols)

$$P_e = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.20}$$

As we increase the transmitted signal energy per bit, E_b, for a specified noise spectral density N_0, the message points corresponding to symbols 1 and 0 move further apart, and the average probability of error P_e is correspondingly reduced in accordance with Equation (6.20), which is intuitively satisfying.

UNIT – V (b)

Information Theory

Information Theory

Information theory deals with representation and the transfer of information.

There are two fundamentally different ways to transmit **messages**: via **discrete** signals and via continuous signals .....For example, the letters of the English alphabet are commonly thought of as **discrete** signals.

Information sources

Definition:

The set of source symbols is called the **source alphabet**, and the elements of the set are called the **symbols or letters.**

The number of possible answers ‘ r ’ should be linked to “information.”

“Information” should be additive in some sense.

We define the following measure of information:

$$\tilde{I}(U) \triangleq \log_b r,$$

Where ‘ r ’ is the number of all possible outcome so far an do m message U.

Using this definition we can confirm that it has the wanted property of additivity:

$$\tilde{I}(U_1, U_2, \ldots, U_n) = \log_b r^n = n \cdot \log_b r = n\tilde{I}(U).$$

The basis ‘b’ of the logarithm b is only a change of units without actually changing the amount of information it describes.

Classification of information sources

1. Discrete memory less.
2. Memory.

Discrete memory less source (DMS) can be characterized by “the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source”.

1. Information should be proportion to the uncertainty of an outcome.
2. Information contained in independent outcome should add.

Scope of Information Theory

1. Determine the irreducible limit below which a signal cannot be compressed.
2. Deduce the ultimate transmission rate for reliable communication over a noisy channel.
3. Define Channel Capacity - the intrinsic ability of a channel to convey information.

The basic setup in Information Theory has:

– a source,

– a channel and

– destination.

The output from source is conveyed through the channel and received at the destination. The source is a random variable S which takes symbols from a finite alphabet i.e.,

$$S = \{s_0, s_1, s_2, \cdot \cdot \cdot, s_{k-1}\}$$

With probabilities

$$P(S = s_k) = p_k \text{ where } k = 0, 1, 2, \cdot \cdot \cdot, k - 1$$

and

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

The following assumptions are made about the source

1. Source generates symbols that are statistically independent.
2. Source is memory less i.e., the choice of present symbol does not depend on the previous choices.

Properties of Information

1. Information conveyed by a deterministic event is nothing
2. Information is always positive.
3. Information is never lost.
4. More information is conveyed by a less probable event than a more probable event

Entropy:

The Entropy (H(s)) of a source is defined as the average information generated by a discrete memory less source.

Information content of a symbol:

Let us consider a discrete memory less source (DMS) denoted by X and having the alphabet $\{U_1, U_2, U_3, \ldots\ldots U_m\}$. The information content of the symbol xi, denoted by $I(x_i)$ is defined as

$$I(U) = \log_b \frac{1}{P(u)} = -\log_b P(U)$$

Where P (U) is the probability of occurrence of symbol U

Units of $I(x_i)$:

For two important and one unimportant special cases of b it has been agreed to use the following names for these units:

b =2(log2): bit,

b = e (ln): nat (natural logarithm),

b =10(log10): Hartley.

The conversation of these units to other units is given as

$$\log_2 a = \frac{\ln a}{\ln 2} = \frac{\log a}{\log 2}$$

Uncertainty or Entropy (i.e Average information)

Definition:

In order to get the information content of the symbol, the flow information on the symbol can fluctuate widely because of randomness involved into the section of symbols.

The uncertainty or entropy of a discrete random variable (RV) ‘U’ is defined as

$$H(U) = E[I(u)] = \sum_{i=1}^{m} P(u) I(u)$$

$$H(U) \triangleq -\sum_{u \in \text{supp}(P_U)} P_U(u) \log_b P_U(u),$$

Where PU (·) denotes the probability mass function (PMF) 2 of the RV U, and where the support of P U is defined as

$$\operatorname{supp}(P_U) \triangleq \{u \in \mathcal{U} : P_U(u) \neq 0\}.$$

We will usually neglect to mention "support" when we sum over PU (u) · $\log_b$ PU (u), i.e., we implicitly assume that we exclude all u

With zero probability PU (u) =0.

Entropy for binary source

It may be noted that for a binary source U which genets independent symbols 0 and 1 with equal probability, the source entropy H (u) is

$$H(u) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \text{ b/symbol}$$

Bounds on H (U)

If U has r possible values, then 0 ≤ H(U) ≤ log r,

Where

H(U)=0 if, and only if, PU(u)=1 for some u,

H(U)=log r if, and only if, PU(u)= 1/r ∀ u.

Hence, H(U) ≥ 0.Equalitycanonlybeachievedif −PU(u)log2 PU(u)=0

Proof. Since $0 \leq P_U(u) \leq 1$, we have

$$-P_U(u)\log_2 P_U(u) \begin{cases} = 0 & \text{if } P_U(u) = 1, \\ > 0 & \text{if } 0 < P_U(u) < 1. \end{cases}$$

For all u ∈ supp (PU), i.e., PU (u) =1forall u ∈ supp (PU).

To derive the upper bound we use at rick that is quite common in.

Formation theory: We take the deference and try to show that it must be non positive.

$$
\begin{aligned}
H(U) - \log r &= -\sum_{u \in \mathrm{supp}(P_U)} P_U(u) \log P_U(u) - \log r \\
&= -\sum_{u \in \mathrm{supp}(P_U)} P_U(u) \log P_U(u) - \sum_{u \in \mathrm{supp}(P_U)} P_U(u) \log r \\
&= -\sum_{u \in \mathrm{supp}(P_U)} P_U(u) \log \left(P_U(u) \cdot r\right) \\
&= \sum_{u \in \mathrm{supp}(P_U)} P_U(u) \log \bigg(\underbrace{\frac{1}{r \cdot P_U(u)}}_{\triangleq \xi}\bigg) \\
&\le \sum_{u \in \mathrm{supp}(P_U)} P_U(u) \left(\frac{1}{r \cdot P_U} - 1\right) \cdot \log e \\
&= \Bigg(\sum_{u \in \mathrm{supp}(P_U)} \frac{1}{r} - \underbrace{\sum_{u \in \mathrm{supp}(P_U)} P_U(u)}_{=1}\Bigg) \cdot \log e \\
&= \left(\frac{1}{r} \sum_{u \in \mathrm{supp}(P_U)} 1 - 1\right) \log e \\
&\le \left(\frac{1}{r} \sum_{u \in \mathcal{U}} 1 - 1\right) \log e \\
&= \left(\frac{1}{r} \cdot r - 1\right) \log e \\
&= (1 - 1) \log e = 0.
\end{aligned}
$$

Equality can only be achieved if

1. In the IT Inequality $\xi = 1$, i.e., if $\frac{1}{r \cdot P_U(u)} = 1 \Longrightarrow P_U(u) = \frac{1}{r}$, for all u;
2. $|\mathrm{supp}(P_U)| = r$.

Note that if Condition1 is satisfied, Condition 2 is also satisfied.

Conditional Entropy

Similar to probability of random vectors, there is nothing really new about conditional probabilities given that a particular event Y = y has occurred.

The conditional entropy or conditional uncertainty of the RV X given the event Y = y is defined as

$$H(X|Y=y) \triangleq -\sum_{x\in\text{supp}(P_{X|Y}(\cdot|y))} P_{X|Y}(x|y)\log P_{X|Y}(x|y)$$
$$= \mathsf{E}\left[-\log P_{X|Y}(X|Y) \,\middle|\, Y=y\right].$$

Note that the definition is identical to before apart from that everything is conditioned on the event Y = y

$$0 \le H(X|Y=y) \le \log r;$$
$$H(X|Y=y)=0 \quad \textit{if, and only if,} \quad P(x|y)=1 \quad \textit{for some } x;$$
$$H(X|Y=y)=\log r \quad \textit{if, and only if,} \quad P(x|y)=\frac{1}{r} \quad \forall x.$$

Note that the conditional entropy given the event Y = y is a function of y. Since Y is also a RV, we can now average over all possible events Y = y according to the probabilities of each event. This will lead to the averaged.

Mutual Information

Although conditional entropy can tell us when two variables are completely independent, it is not an adequate measure of dependence. A small value for **H(Y| X)** may implies that **X** tells us a great deal about **Y** or that **H(Y)** is small to begin with. Thus, we measure dependence using *mutual information*:

I(X,Y) =H(Y)–H(Y|X)

|

Mutual information is a measure of the reduction of randomness of a variable given knowledge of another variable. Using properties of logarithms, we can derive several equivalent definitions

$$I(X,Y)=H(X)-H(X|Y)$$

$$I(X,Y) = H(X)+H(Y)-H(X,Y) = I(Y,X)$$

In addition to the definitions above, it is useful to realize that mutual information is a particular case of the Kullback-Leibler divergence. The KL divergence is defined as:

$$D(p||q) = \int p(x)\log\frac{p(x)}{q(x)}.$$

KL divergence measures the difference between two distributions. It is sometimes called the relative entropy. It is always non-negative and zero only when **p**=**q**; however, it is not a distance because it is not symmetric.

In terms of KL divergence, mutual information is:

$$D(P(X,Y)||P(X)P(Y))) = \int P(X,Y)\log\frac{P(X,Y)}{P(X)P(Y)}$$

In other words, mutual information is a measure of the difference between the joint probability and product of the individual probabilities. These two distributions are equivalent only when **X** and **Y** are independent, and diverge as **X** and **Y** become more dependent.

Source coding

Coding theory is the study of the properties of codes and their respective fitness for specific applications. Codes are used for data compression, cryptography, error-correction, and networking. Codes are studied by various scientific disciplines—such as information theory, electrical engineering, mathematics, linguistics, and computer science—for the purpose of designing efficient and reliable data transmission methods. This typically involves the removal of redundancy and the correction or detection of errors in the transmitted data.

The aim of source coding is to take the source data and make it smaller.

All source models in information theory may be viewed as random process or random sequence models. Let us consider the example of a discrete memory less source (DMS), which is a simple random sequence model.

A DMS is a source whose output is a sequence of letters such that each letter is independently selected from a fixed alphabet consisting of letters; say a_1, a_2 ,

………a_k. The letters in the source output sequence are assumed to be random and statistically

Independent of each other. A fixed probability assignment for the occurrence of each letter is also assumed. Let us, consider a small example to appreciate the importance of probability assignment of the source letters.

Let us consider a source with four letters a_1, a_2, a_3 and a_4 with $P(a_1)=0.5$, $P(a_2)=0.25$, $P(a_3)= 0.13$, $P(a_4)=0.12$. Let us decide to go for binary coding of these four

Source letters While this can be done in multiple ways, two encoded representations are shown below:

Code Representation#1:

a_1: 00, a_2:01, a_3:10, a_4:11

Code Representation#2:

a_1: 0, a2:10, a_3:001, a_4:110

It is easy to see that in method #1 the probability assignment of a source letter has not been considered and all letters have been represented by two bits each. However in

The second method only a_1 has been encoded in one bit, a_2 in two bits and the remaining two in three bits. It is easy to see that the average number of bits to be used per source letter for the two methods is not the same. (~~a~~ for method #1=2 bits per letter and ~~a~~ for method #2 < 2 bits per letter). So, if we consider the issue of encoding a long sequence of

Letters we have to transmit less number of bits following the second method. This is an important aspect of source coding operation in general. At this point, let us note

a) We observe that assignment of small number of bits to more probable letters and assignment of larger number of bits to less probable letters (or symbols) may lead to efficient source encoding scheme.

b) However, one has to take additional care while transmitting the encoded letters. A careful inspection of the binary representation of the symbols in method #2 reveals that it may lead to confusion (at the decoder end) in deciding the end of binary representation of a letter and beginning of the subsequent letter.

So a source-encoding scheme should ensure that

1) The average number of coded bits (or letters in general) required per source letter is as small as possible and

2) The source letters can be fully retrieved from a received encoded sequence.

Shannon-Fano Code

Shannon–Fano coding, named after Claude Elwood Shannon and Robert Fano, is a technique for constructing a prefix code based on a set of symbols and their probabilities. It is suboptimal in the sense that it does not achieve the lowest possible expected codeword length like Huffman coding; however unlike Huffman coding, it does guarantee that all codeword lengths are within one bit of their theoretical ideal $I(x) = -\log P(x)$.

In Shannon–Fano coding, the symbols are arranged in order from most probable to least probable, and then divided into two sets whose total probabilities are as close as possible to being equal. All symbols then have the first digits of their codes assigned; symbols in the first set receive "0" and symbols in the second set receive "1". As long as any sets with more than one member remain, the same process is repeated on those sets, to determine successive digits of their codes. When a set has been reduced to one symbol, of course, this means the symbol's code is complete and will not form the prefix of any other symbol's code.

The algorithm works, and it produces fairly efficient variable-length encodings; when the two smaller sets produced by a partitioning are in fact of equal probability, the one bit of information used to distinguish them is used most efficiently. Unfortunately, Shannon–Fano does not always produce optimal prefix codes.

For this reason, Shannon–Fano is almost never used; Huffman coding is almost as computationally simple and produces prefix codes that always achieve the lowest expected code word length. Shannon–Fano coding is used in the IMPLODE compression method, which is part of the ZIP file format, where it is desired to apply a simple algorithm with high performance and minimum requirements for programming.

Shannon-Fano Algorithm:

A Shannon–Fano tree is built according to a specification designed to define an effective code table. The actual algorithm is simple:

For a given list of symbols, develop a corresponding list of probabilities or frequency counts so that each symbol's relative frequency of occurrence is known.

- Sort the lists of symbols according to frequency, with the most frequently occurring
  Symbols at the left and the least common at the right.
- Divide the list into two parts, with the total frequency counts of the left part being as
  Close to the total of the right as possible.
- The left part of the list is assigned the binary digit 0, and the right part is assigned the digit 1. This means that the codes for the symbols in the first part will all start with 0, and the codes in the second part will all start with 1.

Recursively apply the steps 3 and 4 to each of the two halves, subdividing groups and adding bits to the codes until each symbol has become a corresponding code leaf on the tree.

Example:

The source of information A generates the symbols {A0, A1, A2, A3 and A4} with the corresponding probabilities {0.4, 0.3, 0.15, 0.1 and 0.05}. Encoding the source symbols using binary encoder and Shannon-Fano encoder gives

Source Symbol	P_i	**Binary Code**	**Shannon-Fano**
A0	0.4	000	0
A1	0.3	001	10
A2	0.15	010	110
A3	0.1	011	1110
A4	0.05	100	1111
L_{avg}	**H = 2.0087**	**3**	**2.05**

The average length of the Shannon-Fano code is

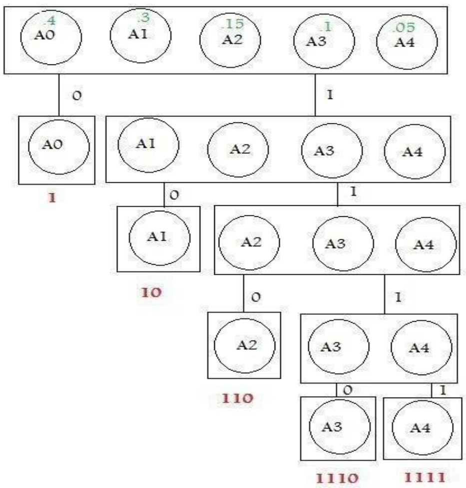
$$Lavg = \sum_{i=0}^{4} Pi\ li = 0.4 * 1 + 0.3 * 2 + 0.15 * 3 + 0.1 * 4 + 0.05 * 4 = 2.05 \text{ bit/symbol}$$

Thus the efficiency of the Shannon-Fano code is

$$\eta = \frac{H}{Lavg} = \frac{2.0087}{2.05} = 98\%$$

This example demonstrates that the efficiency of the Shannon-Fano encoder is much higher than that of the binary encoder.

Shanon-Fano code is a top-down approach. Constructing the code tree, we get



The Entropy of the source is

$$H = -\sum_{i=0}^{4} Pi\ \log_2 Pi = 2.0087 \text{ bit/symbol}$$

Since we have 5 symbols ($5<8=2^3$), we need 3 bits at least to represent each symbol in binary (fixed-length code). Hence the average length of the binary code is

$$Lavg = \sum_{i=0}^{4} Pi\ li = 3\ (0.4 + 0.3 + 0.15 + 0.1 + 0.05) = 3 \text{ bit/symbol}$$

Thus the efficiency of the binary code is

$$\eta = \frac{H}{Lavg} = \frac{2.0087}{3} = 67\%$$

Binary Huffman Coding (an optimum variable-length source coding scheme)

In Binary Huffman Coding each source letter is converted into a binary code word. It is a prefix condition code ensuring minimum average length per source letter in bits.

Let the source letters a_1, a_2, ……….a_K have probabilities $P(a_1)$, $P(a_2)$,…………. $P(a_K)$ and let us assume that $P(a_1) \geq P(a_2) \geq P(a_3) \geq \ldots \geq P(a_K)$.

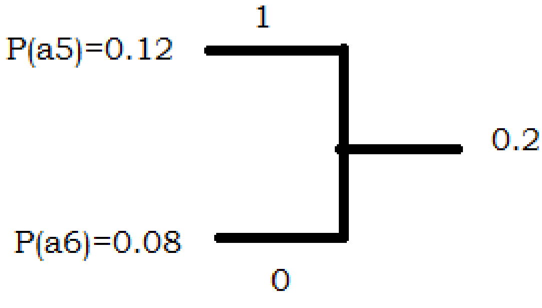
We now consider a simple example to illustrate the steps for Huffman coding.

Steps to calculate Huffman Coding

Example Let us consider a discrete memory less source with six letters having

$P(a_1)=0.3, P(a_2)=0.2, P(a_3)=0.15, P(a_4)=0.15, P(a_5)=0.12$ and $P(a_6)=0.08$.

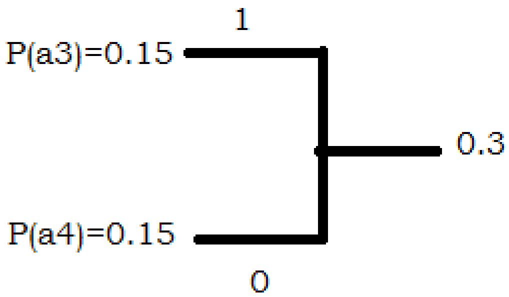
- Arrange the letters in descending order of their probability (here they are arranged).
- Consider the last two probabilities. Tie up the last two probabilities. Assign, say, 0 to the last digit of representation for the least probable letter (a_6) and 1 to the last digit of representation for the second least probable letter (a_5). That is, assign '1' to the upper arm of the tree and '0' to the lower arm.



(3) Now, add the two probabilities and imagine a new letter, say b_1, substituting for a_6 and a_5. So $P(b_1) = 0.2$. Check whether a_4 and b_1 are the least likely letters. If not, reorder the letters as per Step#1 and add the probabilities of two least likely letters. For our example, it leads to:

$$P(a_1)=0.3,\ P(a_2)=0.2,\ P(b_1)=0.2,\ P(a_3)=0.15 \text{ and } P(a_4)=0.15$$

(4) Now go to Step#2 and start with the reduced ensemble consisting of a_1 , a_2 , a_3 ,



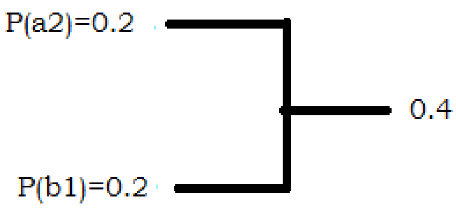
a_4 and b_1. Our example results in:

Here we imagine another letter b_1, with $P(b_2)=0.3$.

Continue till the first digits of the most reduced ensemble of two letters are assigned a ‘1’ and a ‘0’.

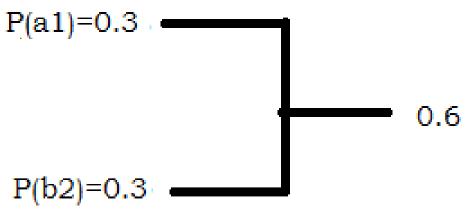
Again go back to the step (2): $P(a_1)=0.3$, $P(b_2)=0.3$, $P(a_2)=0.2$ and $P(b_1)=0.2$.

Now we consider the last two probabilities:

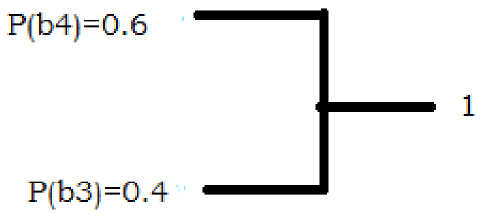


So, $P(b_3)=0.4$. Following Step#2 again, we get, $P(b_3)=0.4$, $P(a_1)=0.3$ and $P(b_2)=0.3$.

Next two probabilities lead to:



With $P(b_4) = 0.6$. Finally we get only two probabilities



6. Now, read the code tree inward, starting from the root, and construct the code words. The first digit of a codeword appears first while reading the code tree inward.

Hence, the final representation is: a_1=11, a_2=01, a_3=101, a_4=100, a_5=001, a_6=000.

A few observations on the preceding example

1. The event with maximum probability has least number of bits

2. Prefix condition is satisfied. No representation of one letter is prefix for other. Prefix condition says that representation of any letter should not be a part of any other letter.

3. Average length/letter (in bits) after coding is

$$= \sum P(a_i) n_i = 2.5 \text{ bits/letter.}$$

4. Note that the entropy of the source is: H(X)=2.465 bits/symbol. Average length per source letter after Huffman coding is a little bit more but close to the source entropy. In fact, the following celebrated theorem due to C. E. Shannon sets the limiting value of average length of code words from a DMS.

Shannon–Hartley theorem

In information theory, the Shannon–Hartley theorem tells the maximum rate at which information can be transmitted over a communications channel of a specified bandwidth in the presence of noise. It is an application of the noisy-channel coding theorem to the archetypal case of a continuous-time analog communications channel subject to Gaussian noise. The theorem establishes Shannon's channel capacity for such a communication link, a bound on the maximum amount of error-free information per time unit that can be transmitted with a specified bandwidth in the presence of the noise interference, assuming that the signal power is bounded, and that the Gaussian noise process is characterized by a known power or power spectral density.

The law is named after Claude Shannon and Ralph Hartley.

Hartley Shannon Law

The theory behind designing and analyzing channel codes is called Shannon's noisy channel coding theorem. It puts an upper limit on the amount of information you can send in a noisy channel using a perfect channel code. This is given by the following equation:

$$C = B \times \log_2(1 + SNR)$$

where C is the upper bound on the capacity of the channel (bit/s), B is the bandwidth of the channel (Hz) and SNR is the Signal-to-Noise ratio (unit less).

Bandwidth-S/N Tradeoff

The expression of the channel capacity of the Gaussian channel makes intuitive sense:

1. As the bandwidth of the channel increases, it is possible to make faster changes in the information signal, thereby increasing the information rate.

2 As S/N increases, one can increase the information rate while still preventing errors due to noise.

3. For no noise, S/N tends to infinity and an infinite information rate is possible irrespective of bandwidth.

Thus we may trade off bandwidth for SNR. For example, if S/N = 7 and B = 4kHz, then the channel capacity is C = 12 $\times 10^3$ bits/s. If the SNR increases to S/N = 15 and B is decreased to 3kHz, the channel capacity remains the same. However, as B tends to 1, the channel capacity does not become infinite since, with an increase in bandwidth, the noise power also increases. If the noise power spectral density is ɳ/2, then the total noise power is N = ɳB, so the Shannon-Hartley law becomes

$$
\begin{aligned}
C &= B \log_2 \left(1 + \frac{S}{\eta B}\right) = \frac{S}{\eta}\left(\frac{\eta B}{S}\right) \log_2 \left(1 + \frac{S}{\eta B}\right) \\
&= \frac{S}{\eta} \log_2 \left(1 + \frac{S}{\eta B}\right)^{\eta B / S}.
\end{aligned}
$$

Noting that

$$\lim_{x \to 0} (1 + x)^{1/x} = e$$

and identifying x as $x = S/\eta B$, the channel capacity as B increases without bound becomes

$$C_\infty = \lim_{B \to \infty} C = \frac{S}{\eta} \log_2 e = 1.44 \frac{S}{\eta}.$$