University of Anbar
College of Engineering
Department of Electrical Engineering

LECTURE NOTES FOR

Analog and Digital Communications

FOR THE THIRD YEAR STUDENTS

DR. MOHAMMED ALI ALMAHAMDY
mohammed.almahamdy@uoanbar.edu.iq

REFERENCES

- *Introduction to Communication Systems*. Ferrel G. Stremler.
- Instructor's Lectures.

BIBLIOGRAPHY:

(1) *Modern Digital and Analog Communication Systems*. B. P. Lathi.

(2) *Communication Systems, an Introduction to Signals and Noise in Electrical Communications*. Bruce Carlson.

(3) *Communication Systems Engineering*. John G. Proakis and Masoud Salehi.

(4) *Digital and Analog Communication Systems*. Leon W. Couch.

Course Timeline

First Semester

Week	
1	Introduction to Communication System, Channel; Signal Classification and Characteristics
2	Introduction to Waves Propagation, Multipath
3	Introduction to Fourier Transform, Filters & Bandwidth
4	Analog Communications, Amplitude Modulation Systems, AM-DSB-SC (Mod/Demod)
5	AM-DSB-LC (Modulation-Demodulation)
6	AM-SSB & AM-VSB (Modulation-Demodulation)
7	FDM, Frequency Conversion, Super-Heterodyne Receiver
8	FM: Introduction, NBFM, WBFM
9	Spectrum Plotting Using Bessel Function, Power in FM
10	FM Generation: Direct (VCO) and Indirect Method (Armstrong)
11	FM Detection: Discriminator, Zero Crossing Detector, PLL
12	Introduction, Noise Sources, Mathematical Representation of Noise, Noise Figure
13	Thermal Noise, White & Filtered Noise, Equivalent Temperature, Noise in Multistage System
14	Noise & SNR in: AM (DSB / SSB / normal AM), FM, Noise Reduction in FM Using Emphasis

Second Semester

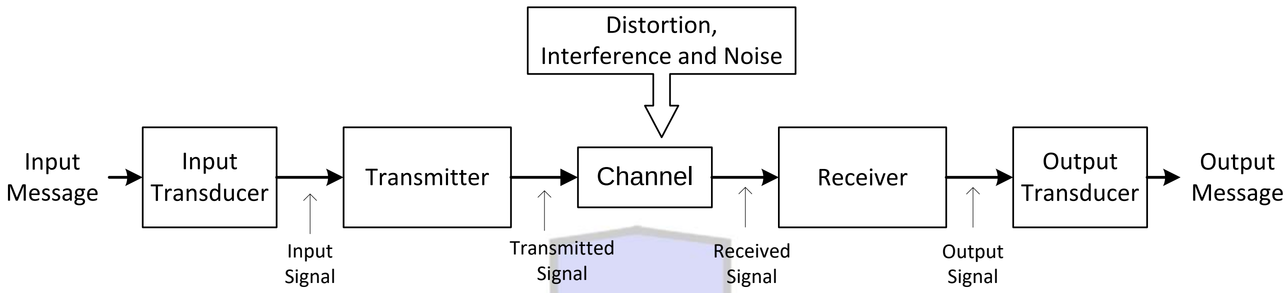
Week	
1	The Sampling Theory (Ideal & Natural), Nyquist Sampling Rate & Aliasing in Reconstruction
2	PCM
3	Non-Uniform Quantization & Companding, SNR in PCM
4	Delta PCM, Deferential PCM, DM Systems, Noise in DM, Adaptive DM
5	PAM-TDM, Crosstalk and Guard Times, PCM-TDM, ISI & Eye Diagram
6	Channel Capacity, Multi-Level Baseband Signaling (M-ary)
7	Probability of Error at Reception
8	Examples & Problems
9	Introduction Digital Modulation, BASK (Mod/DeMod/P_E)
10	BFSK (Mod/DeMod/P_E)
11	BPSK (Mod/DeMod/P_E)
12	Comparison Through: Average Power, Peak Power, Spectral Efficiency, Their Systems
13	M-ary Modulation, QPSK (Mod/DeMod/P_E)
14	QAM (Mod/DeMod/P_E)

TABLE OF CONTENTS

Section *Page*

Part 1 INTRODUCTION TO COMMUNICATION SYSTEMS

A typical Communication system can be modeled as:



The *source* originates the message. If this message is non-electrical (such as human voice or a television picture), it must be converted by an *input transducer* into electrical waveforms referred as the baseband signal.

For efficient transmission, the *transmitter* processes and amplifies the input signal to produce a transmitted signal suited to the characteristics of the transmission channel. Signal processing involves: Modulation and Coding.

The *channel* is the medium that bridges the distance from source to destination, such as: wire, coaxial cable, optical fiber, radio link… etc.

Generally, the signal passes a channel suffers from the following main problems:

(1) Attenuation: the signal power gradually decreases along the distance.

(2) Distortion: the channel changes the shape of the signal.

(3) Noise: the signal is corrupted by random and unpredictable electrical signals. These unwanted signals are produced by natural processes in both internal and external to the system.

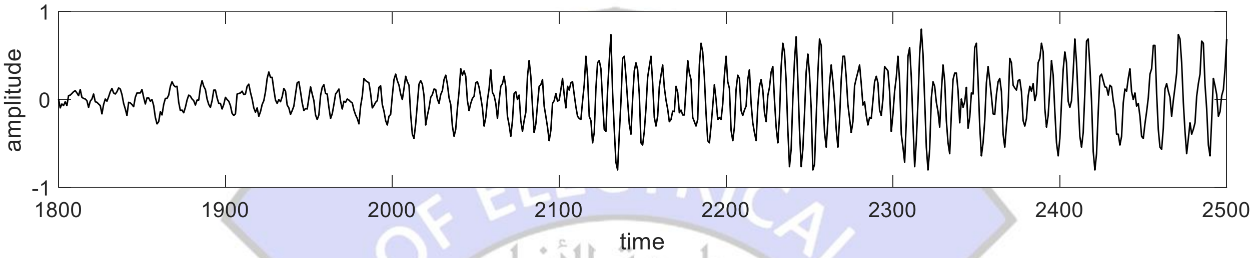
(4) Interference: is the contamination by extraneous signals, such as: other transmitters, power lines and machinery, switching circuits… etc.

The *receiver* compensates the received signal for the channel effects, and reprocesses the received signal by undoing the signal modification made at the transmitter through the demodulation and the decoding. The receiver output is fed to the *output transducer* to convert the electrical signals to its original form: the message.

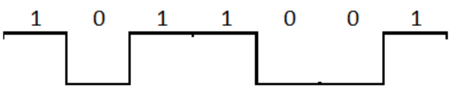
The main objective of a "brilliant" communication engineer is to design and implement a maximum efficiency system which can send and receiving data with minimum required resources (power, frequency band, time, and cost) beside many other issues that must be considered (such as data security, reliability, immunity to noise, better control, weight, size, etc.).

1.1 BASIC DEFINITIONS:

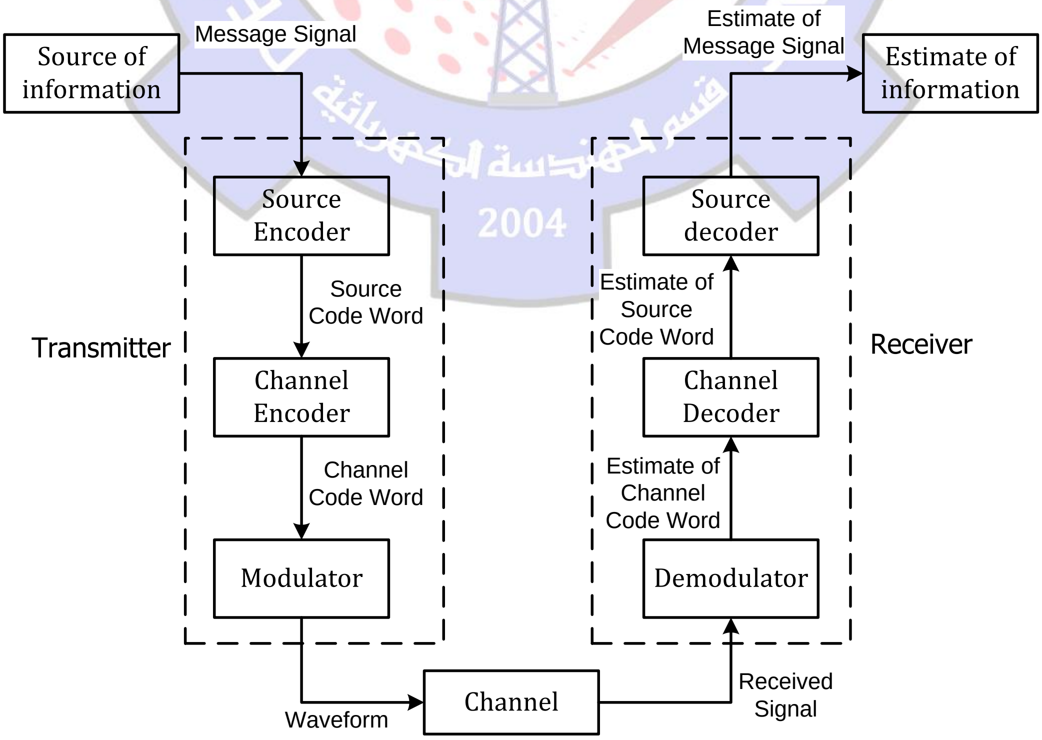
An *Analogue* signal is defined as a physical time varying quantity and is usually smooth and continuous, e.g. acoustic pressure variation when speaking. The performance of an analogue communications system is often specified in terms of its fidelity or quality.



A *digital* signal on the other hand is made up of discrete symbols selected from a finite set, e.g. letters from the alphabet or binary data. The performance of a digital system is specified in terms of accuracy of transmission e.g. Bit Error Rate (BER) and Symbol Error Rate (SER).



The main elements of a digital system are shown in below diagram:



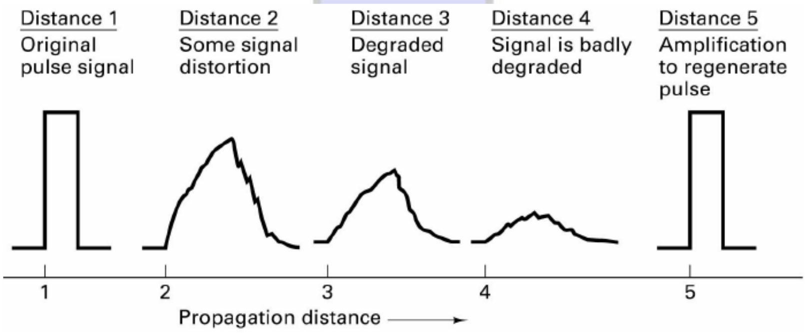
The *source encoder* removes redundant information from the message signal to facilitate the transmission. The opposite is true for the *channel encoder*. It adds redundant bits to the transmission to provide the capability of the error correction and control at the receiver. Finally, the modulator represents each symbol of the channel code word by an analog symbol. The sequence of analog symbols is called a waveform, which is suitable for transmission over the physical channel.

At the receiver, the channel output (the received signal) is processed in the order reverse to that of the transmitter, thereby reconstructing a recognizable version of the original message signal to be finally delivered to the user.

From this description, it is apparent that the design of a digital communication system is rather complex, but nowadays electronics are inexpensive, due to the ever-increasing availability of VLSI circuits in the form of silicon chips. Besides being easy to build, digital communications offer greater tolerance of physical effects (e.g. temperature variations, aging, mechanical vibrations) than its analog counterparts.

Advantages of digital communication systems:

(1) Relatively inexpensive digital circuits may be used.
(2) Privacy is preserved by using data encryption.
(3) Voice, video, and data sources may be merged and transmitted over a common digital system.
(4) Errors can often be corrected using coding.
(5) It is easy to regenerate the transmitted signal to be able to extend the receiver distance. For example: A regenerative repeater:



Digital communication also has disadvantages:

- Generally, more bandwidth is required than that for analog systems.
- Synchronization is required.

Despite the trend toward the ever-increasing use of digital communications, analog communications are still in use; e.g., in radio. It is important to understand the fundamentals of analog modulation techniques, thereby getting better insights into why digital communications is more preferred nowadays.

1.2 Synchronous/Asynchronous Communications

Synchronous (coherent) transmission: A synchronous system is one in which the transmitter and receiver are operating continuously at the same number of symbols per second in the desired phase relationship.

Asynchronous (non-coherent) transmission: in an asynchronous system, no rigid timing constraint is applied between the transmitter and the receiver.

Advantages of synchronous data communications

- Superior noise immunity due to matched filtering; that is. the symbol or bit is averaged over its entire duration giving optimum noise and interference rejection and maximizing signal power.
- Can accommodate higher data rates than asynchronous systems.

Disadvantages of synchronous data communications

- Requires finite time for synchronization to occur.
- Is more expensive and complex than asynchronous operation.

1.3 Gain/Loss in Decibel (dB)

The unit of signal strength based on ratio of power is:

$$\text{dB} = 10\log_{10}\left(\frac{\text{Measured Power}}{\text{Reference Power}}\right) = 10\log_{10}\left(\frac{P_o}{P_i}\right)$$

So, if an amplifier makes an output signal 100 times stronger than its input, then it has gain of:

$$\text{Gain}_{\text{dB}} = 10\log_{10}\left(\frac{100P_i}{P_i}\right) = 20\text{dB}$$

Also, if a signal reduced to 5% of original strength when passing through a cable, then this cable has dB loss (negative gain) of:

$$\text{Loss}_{\text{dB}} = 10\log_{10}\left(\frac{0.05}{1}\right) = -13\text{dB}$$

If signal intensity measured in Volts or Amperes, then:

$$\text{dB} = 10\log_{10}\left(\frac{V_o^2/R}{V_i^2/R}\right) = 20\log_{10}\left(\frac{V_o}{V_i}\right) \quad , \quad \text{dB} = 10\log_{10}\left(\frac{I_o^2 R}{I_i^2 R}\right) = 20\log_{10}\left(\frac{I_o}{I_i}\right)$$

H.W. Find P_o/P_i & V_o/V_i ratios for -3dB system gain?

H.W. what is the dBm?

Signals strength is multiplied through cascaded blocks; and added using dB of these strengths!

1.4 Communication Modes

In any communication link connecting two devices, data can be sent in one of three communication modes. These are:

A *simplex*: the communication flow can only occur in one direction. e.g. broadcast radio.

A *half-duplex*: communication in both directions, but not at the same time. e.g. 'walkie-talkies'.

A *full-duplex:* system can support simultaneous two-way communication. e.g. telephone.

1.5 Transmission Characteristics

1.5.1 Signaling Rate

The signaling rate of a communication link is a measure of how many times the physical signal changes per second and is expressed as the *baud rate*. An oscilloscope trace of the data transfer would show pulses at the baud rate.

1.5.2 Data Rate

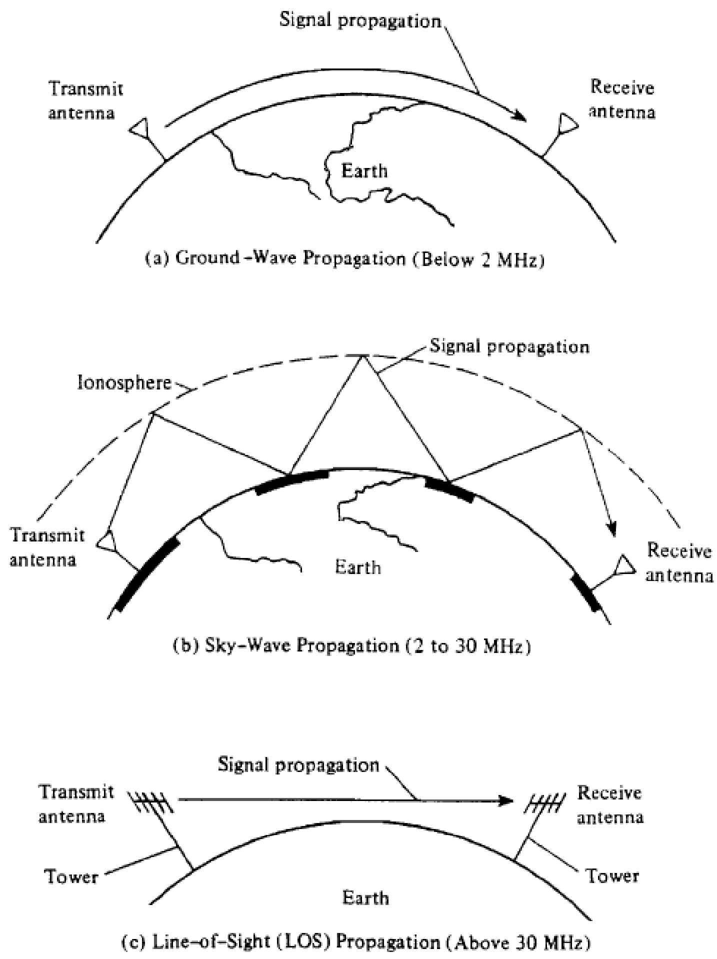
The data rate or bit rate is expressed in bits per second (bps). This represents the actual number of data bits transferred per second.

1.5.3 Signal to Noise Ratio

The signal to noise (S/N) ratio of a communications link is an important limiting factor of the quality and rate of the communication. Noise sources may be external or internal, as discussed in Part 4 in this document.

1.6 WAVE PROPAGATION

Generally, there are three types of traveling waves in wireless channels: the ground waves, the sky waves, and the Line-of-Sight (LoS).



Propagation Types of Radio Frequencies

The ground waves communication is efficient using LF band. While the sky waves communication is working only in the HF band. Higher frequency (VHF or higher) are not reflected by ionosphere and they pass directly to space, so they are used in the satellite communications.

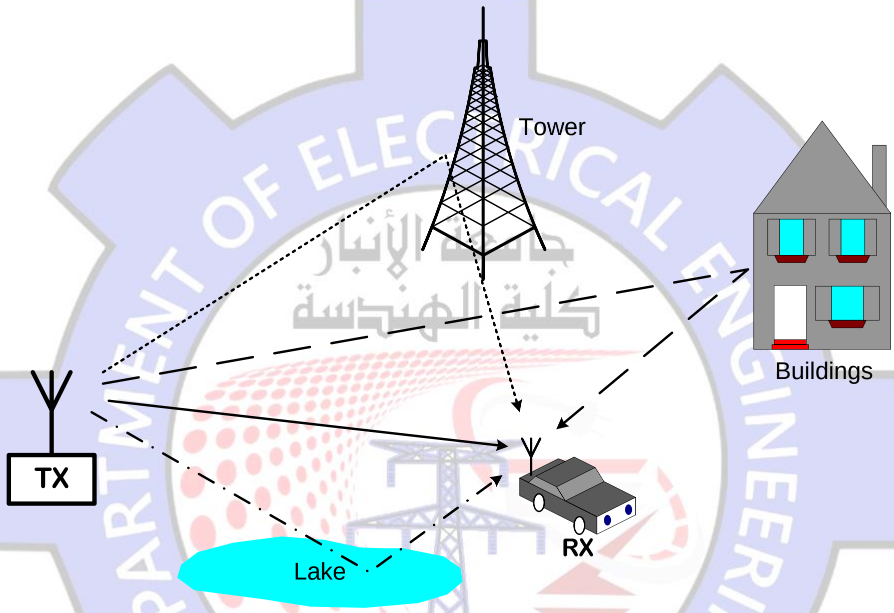
The ionosphere is a region of ionized particles extends from 100km to 700km above the earth surface. The ionization is caused by the action of the sun's radiation on the upper atmosphere of the earth. The free electrons act as reflectors for HF radio waves, and their density varies with: height, period of the day and year. During day time, the reflection capability is strong, but it decreases during night. Most AM signals are transmitted over the HF band 3MHz → 30MHz. The transmitted HF signal reaches the receiver via the sky waves that have been reflected by the ionosphere.

While, the FM signals are transmitted over the VHF band which pass through the atmosphere layers (not reflected), hence the FM broadcasting coverage is limited. High frequencies, therefore, are transmitted through the LoS communication.

1.7 Multipath Effect

When two or more radio waves arrive at the receiver point along different paths, their phases (and sometimes frequencies) are not the same, results in cancellation of one signal by the others, which results in a loss of the signal strength. This phenomenon is called '*fading*'.

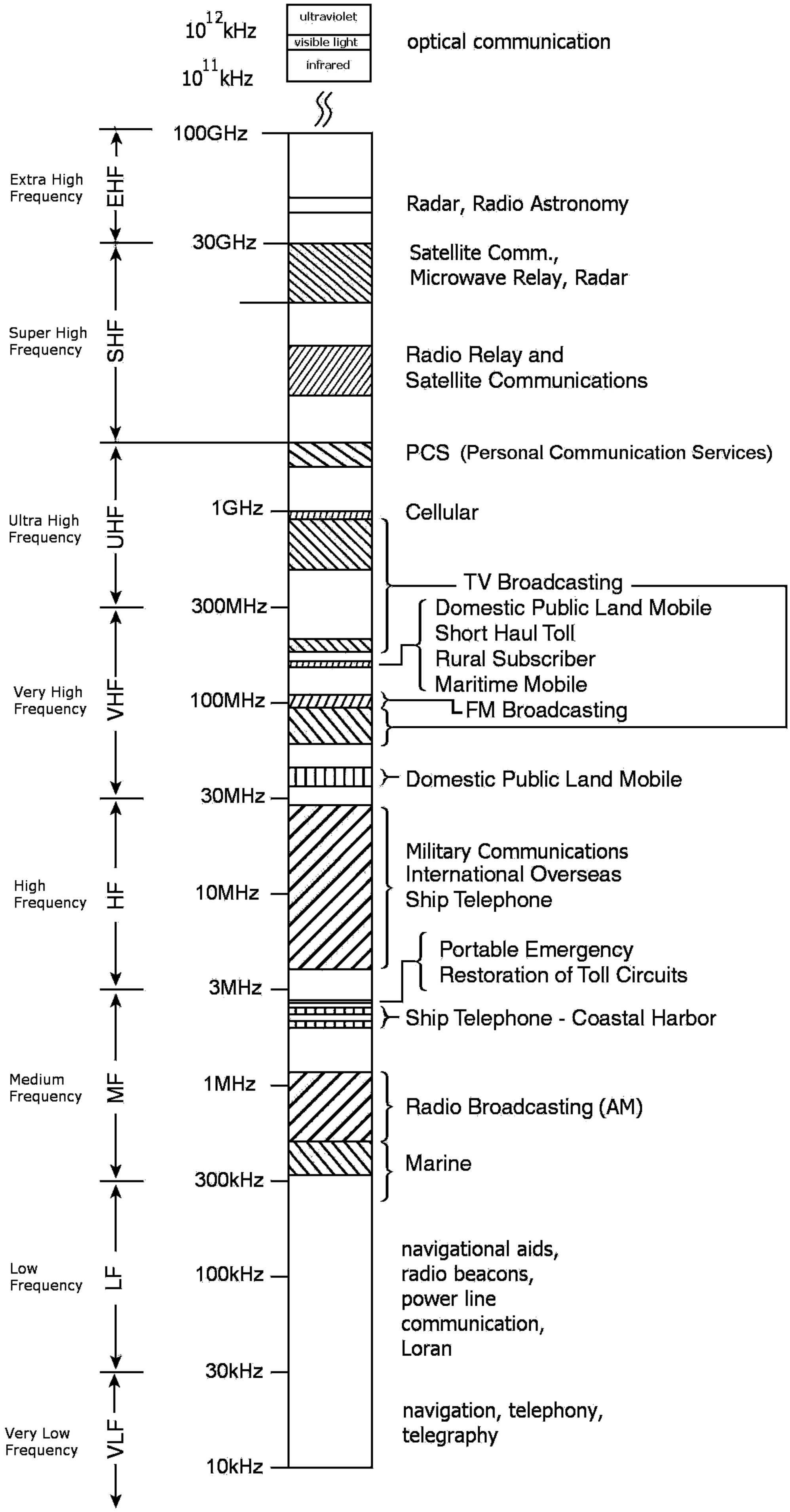
The fading effect is usually divided into two types, namely *large-scale fading*, mainly due to path loss as a function of distance and shadowing by large objects such as mountains and tall buildings, and *small-scale fading* due to the constructive and destructive combination of randomly scattered, reflected, diffracted, and delayed multiple path signals.



1.8 Government Regulations and Standards

In communications, perhaps more than any other field, the need for standards to ensure correct interoperation of equipment is paramount. Communication systems are always interworking with other devices, possibly located on the other side of the world.

The drawing up of standards falls to a small number of national and international bodies, with, for example, ITU (International Telecommunications Union) being responsible for the drafting standards of most of the new wireless and wired communications.


ultraviolet
visible light
infrared
10^{12}kHz
10^{11}kHz
optical communication
100GHz
Extra High Frequency
EHF
Radar, Radio Astronomy
30GHz
Satellite Comm.,
Microwave Relay, Radar
Super High Frequency
SHF
Radio Relay and
Satellite Communications
PCS (Personal Communication Services)
1GHz
Cellular
Ultra High Frequency
UHF
TV Broadcasting
300MHz
Domestic Public Land Mobile
Short Haul Toll
Rural Subscriber
Maritime Mobile
Very High Frequency
VHF
100MHz
FM Broadcasting
Domestic Public Land Mobile
30MHz
Military Communications
International Overseas
Ship Telephone
High Frequency
HF
10MHz
Portable Emergency
Restoration of Toll Circuits
3MHz
Ship Telephone - Coastal Harbor
Medium Frequency
MF
1MHz
Radio Broadcasting (AM)
Marine
300kHz
Low Frequency
LF
100kHz
navigational aids,
radio beacons,
power line
communication,
Loran
30kHz
Very Low Frequency
VLF
navigation, telephony,
telegraphy
10kHz

Part 2 SIGNALS & SPECTRA

2.1 BASIC DEFINITIONS

2.1.1 CLASSIFICATION OF SIGNALS

- Deterministic Signal vs. Random Signal.
- Periodic Signal vs. Non-periodic Signal.
- Analog Signal vs. Discrete Signal.
- Energy Signal vs. Power Signal: where the mean power and the total energy are:

$$P = x_{rms}^2 = \frac{1}{T}\int_0^T x^2(t)\cdot dt \quad , \quad E = \int_{-\infty}^{\infty} x^2(t)\cdot dt$$

- Power signal: If $x(t)$ is periodic, it has finite mean power but infinite energy.
- Energy signal: If $x(t)$ is non-periodic, it has finite energy and zero mean power.

2.1.2 CORRELATION

The correlation function between $x_1(t)$ and $x_2(t)$ is:

For finite energy signals:

$$R(\alpha) = \int_{-\infty}^{\infty} x_1(t)x_2(t+\alpha)dt$$

For finite power signals:

$$R(\alpha) = \frac{1}{T}\int_{-T/2}^{T/2} x_1(t)x_2(t+\alpha)dt$$

2.1.3 CONVOLUTION

The convolution between $x_1(t)$ and $x_2(t)$ is:

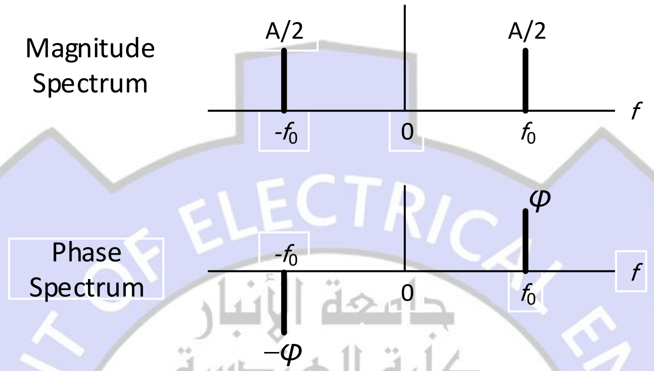
$$c_{12}(\alpha) = x_1 \star x_2 = \frac{1}{T}\int_{-\infty}^{\infty} x_1(t)\cdot x_2(-t+\alpha)\cdot dt$$

2.2 FOURIER TRANSFORM

Although an electrical signal physically exists in the time domain, we can also represent it in the *frequency domain*. We view the signal as it consists of sinusoidal components at various frequencies. This frequency-domain description is called: *spectrum*.

To understand the frequency spectrum, let's consider the signal $x(t) = A\cos(\omega_0 t + \varphi)$ where: A: the peak value or the amplitude, ω_0: The radian frequency and φ: The phase angle.

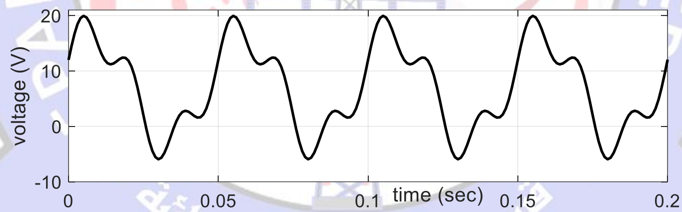
The spectrum of $x(t)$ will be look like:



Next, consider the periodic function:

$$x(t) = 7 + 10\cos(40\pi t - 60°) + 4\sin(120\pi t)$$

which is plotted in the time domain as:



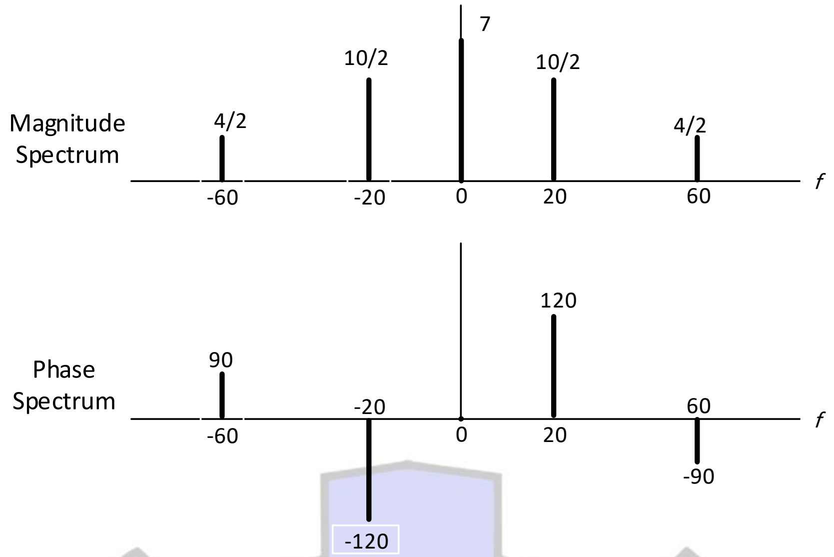
To get the frequency spectrum of $y(t)$, first, we convert $y(t)$ to cosines using:

$\sin\omega t = \cos(\omega t - 90°)$ and $-\cos\omega t = \cos(\omega t \pm 180°)$.

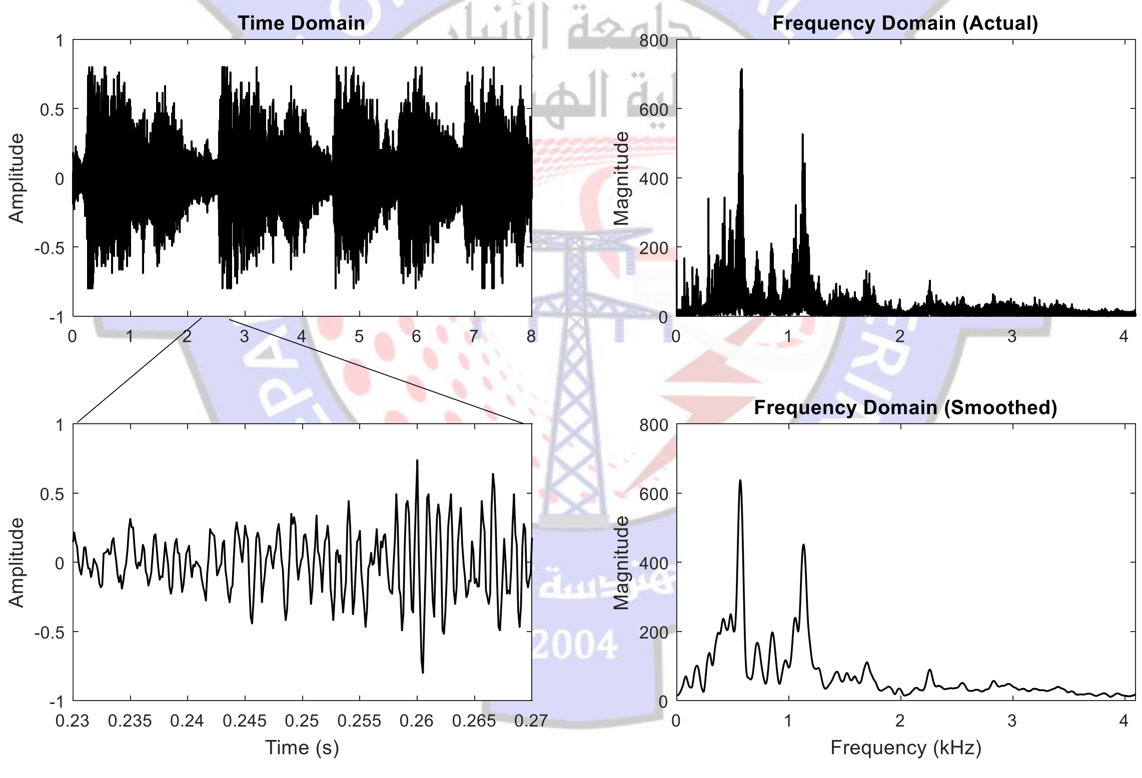
This yields to:

$$y(t) = 7\cos(2\pi 0t) + 10\cos(2\pi 20t + 120°) + 4\cos(2\pi 60t - 90°)$$

Whose spectrum will be:



Practical signals, such as speech, consist of large number of frequency components, and they may look like:



The basic equations relating the time domain version $x(t)$ and the frequency domain version $X(f)$ are known as the Fourier Transform, which are:

$$X(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$$

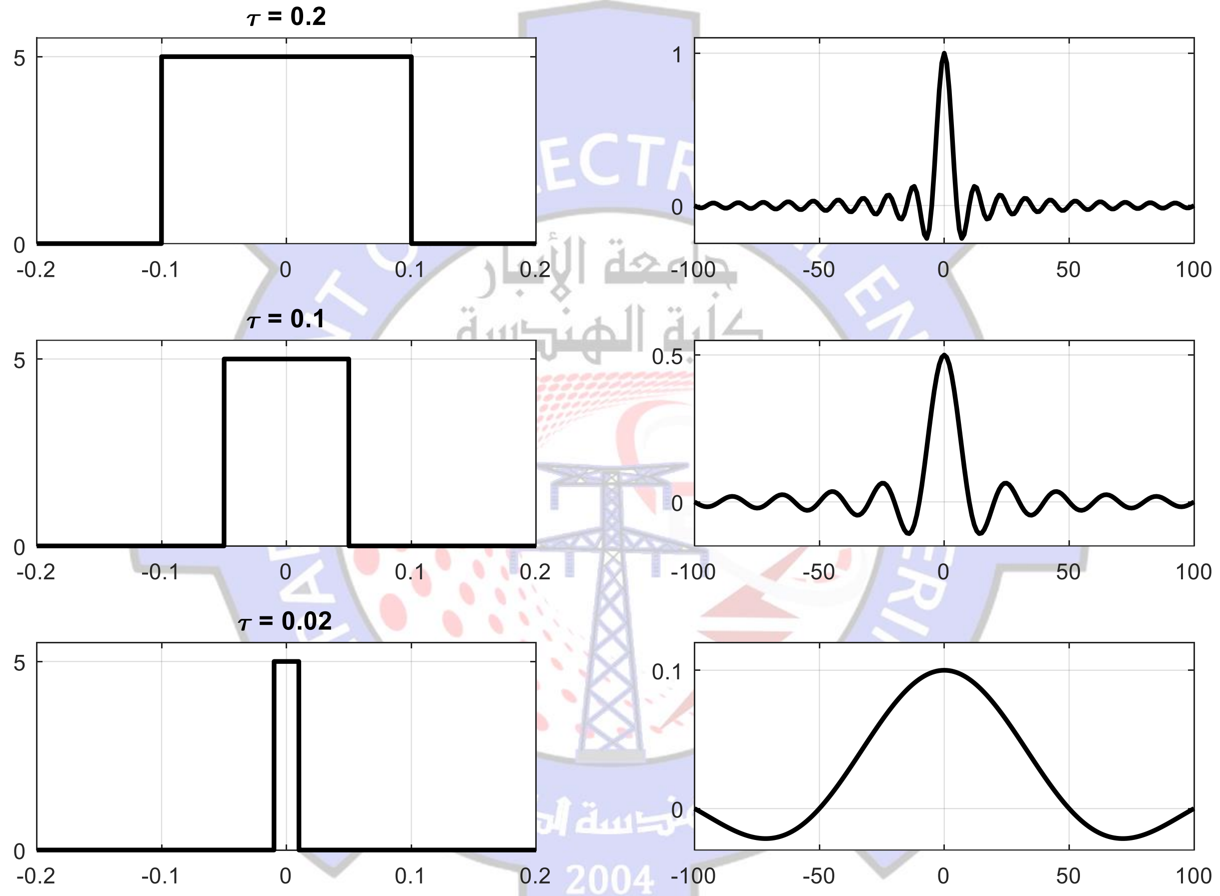
$$x(t) = \mathcal{F}^{-1}\{X(f)\} = \int_{-\infty}^{\infty} X(f)e^{j\omega t}df$$

These general equations are valid for non-periodic signals whose frequency domain are continuous. The spectrum of periodic functions is discrete and the Fourier Transform becomes:

$$X_n = \mathcal{F}\{x(t)\} = \frac{1}{T_0}\int_{T_0} x(t)e^{-jn\omega_0 t}dt$$

$$x(t) = \mathcal{F}^{-1}\{X_n\} = \sum_{n=-\infty}^{\infty} X_n e^{jn\omega_0 t}$$

For example, when $A = 5\text{V}$, and $\tau = (0.2, 0.1, 0.02)$ seconds, we get the first nulls respectively located at $(5, 10, 50)\text{Hz}$.



Main properties of the Fourier Transform: (Str. Fig: 3.3, Tables: 3.1 and 3.2)

Signal (in TD)	Spectrum (in FD)
$x(t)$	$X(f)$
$a \cdot x(t)$	$a \cdot X(f)$
$x_1(t) + x_2(t)$	$X_1(f) + X_2(f)$
$x_1(t) \times x_2(t)$	$X_1(f) \star X_2(f)$
$x_1(t) \star x_2(t)$	$X_1(f) \times X_2(f)$

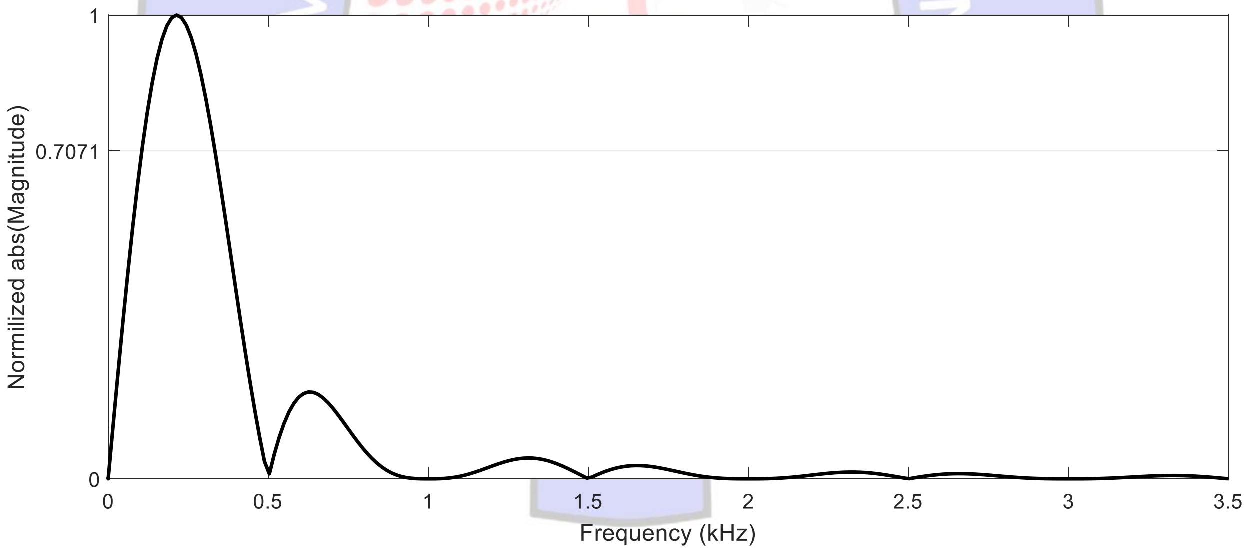
2.3 BANDWIDTH OF A SYSTEM

The set of all frequencies that are present in the signal is the frequency *content*, and if the frequency content consists of all frequencies below some given B, then the signal is said to be *bandlimited to* B. Some bandlimited signals are:

- Telephone quality speech: maximum frequency ~4kHz.
- Audible music: maximum frequency ~20kHz.

As we have seen, many signals have a bandwidth that is theoretically infinite. Several definitions of the bandwidth are commonly used.

- *Absolute bandwidth*: it contains all the frequency componets of the signal, i.e. the spectrum is zero outside it.
- *3dB bandwidth (or the half-power bandwidth)*: it contains the frequency components whose values are at least $1/\sqrt{2}$ times the maximum component.
- *Null-to-null bandwidth (or zero-crossing bandwidth)*.
- *Power banwidth*: it contains the frequency components that sum ~99% of the total power.

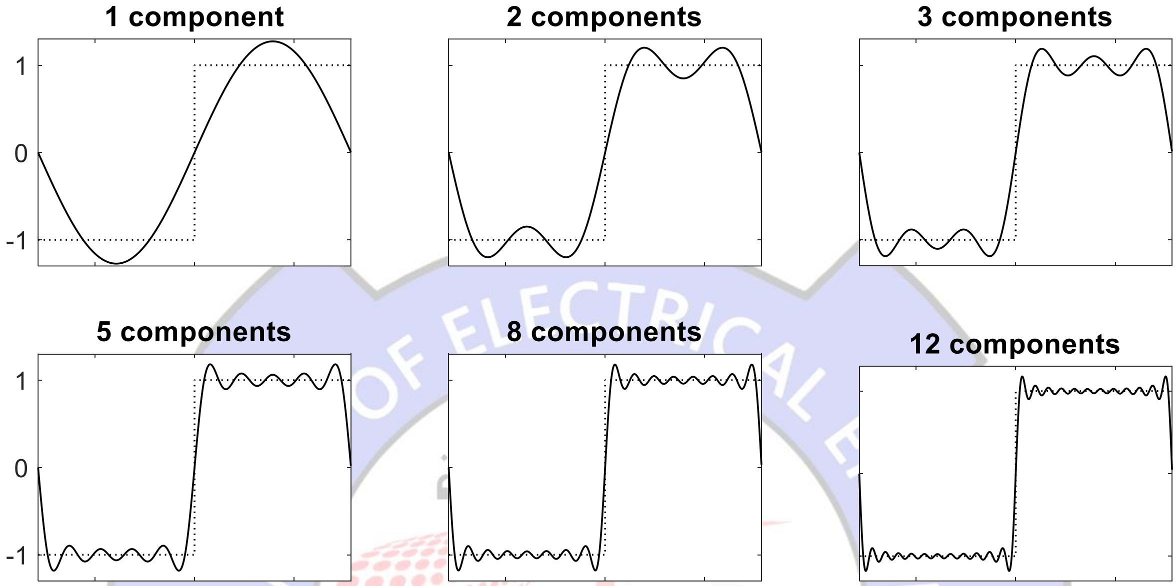


Limiting the frequency response of a system and channel removes some of the frequency components of these signals and causes the time-domain representation to be distorted.

Although real communication systems have a solid theoretical base, they also involve many practical considerations. There is always a trade-off between fidelity to the original signal (that is, the absence of any distortion of its waveform) and such factors as bandwidth and cost.

Increasing the bandwidth often increases the cost of a communication system. Not only is the hardware likely to be more expensive, but bandwidth itself may be in short supply.

In communication over cables, the total bandwidth of a given cable is fixed by the technology employed. The more bandwidth used by each signal, the fewer signals can be carried by the cable.



$$x(t) = \frac{4}{\pi}\left\{\sin t + \frac{1}{3}\sin 3t + \frac{1}{5}\sin 5t + \cdots + \frac{1}{2k-1}\sin(2k-1)t\right\}$$

Generally, the bandwidth of a system (B) is defined as: the interval of positive frequencies over which the magnitude $|V_O(f)/V_I(f)|$ remains within -3dB (that is $\frac{1}{\sqrt{2}}$ in voltage or ½ in power).

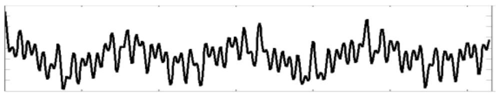
Note: We consider the required B for transmitting a baseband signal = Symbol Rate ÷ 2

2.4 Filters

A filter is a device that passes electric signals at certain frequency ranges while preventing the passage of others. For illustration, let the input signal is:

$$x(t) = 2 + 3\cos(2\pi 800t) + 4\cos(2\pi 1200t + 45°) + 5\cos(2\pi 6000t) + 6\cos(2\pi 9000t) + 7\cos(2\pi 11500t) + 8\cos(2\pi 13000t) + 9\cos(2\pi 20000t)$$

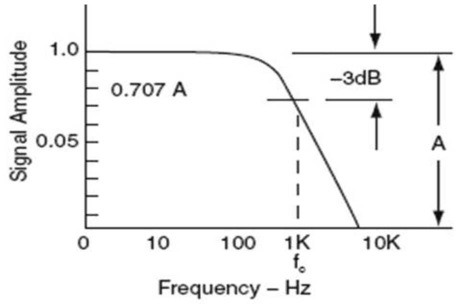
This is plotted in time domain (T.D.) as:

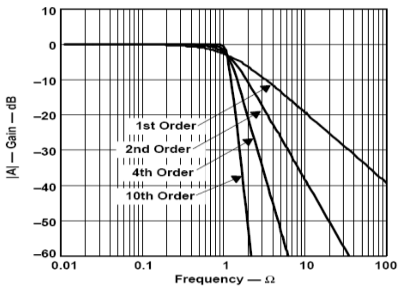


Plot it in frequency domain (F.D.)

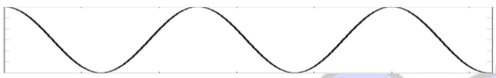
2.4.1 LOW PASS FILTER (LPF)

$B = 0 \rightarrow f_k$





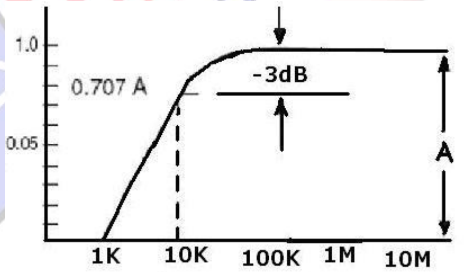
$y(t) = 2 + 3\cos(2\pi 800t)$, This is plotted in T.D. as:



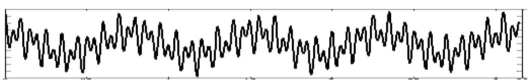
Plot it in frequency domain (F.D.)!

2.4.2 HIGH PASS FILTER (HPF)

$B = f_k \rightarrow \infty$



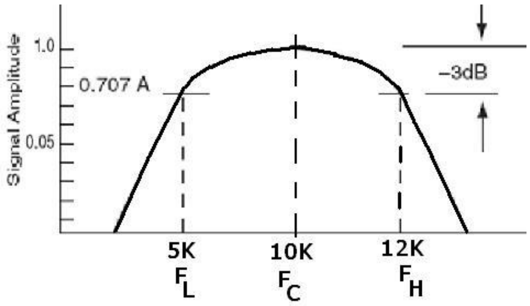
$y(t) = 7\cos(2\pi 11500t) + 8\cos(2\pi 13000t) + 9\cos(2\pi 20000t)$, plotted in T.D. as:

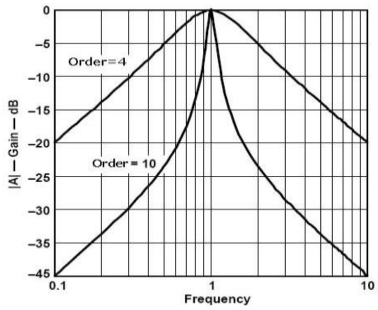


Plot it in frequency domain (F.D.)!

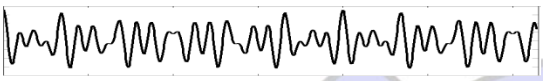
2.4.3 Band Pass Filter (BPF)

$B = f_L \to f_H$, centered at f_C





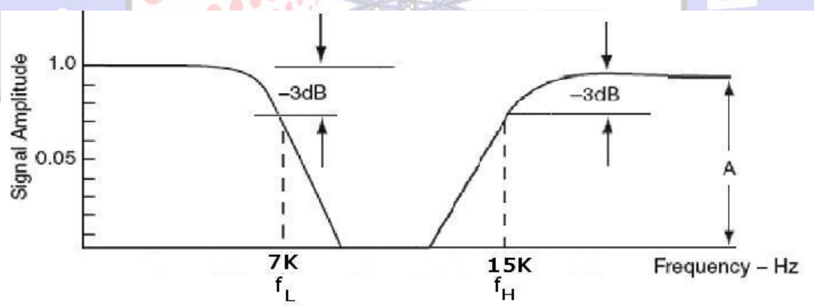
$y(t) = 5\cos(2\pi 6000t) + 6\cos(2\pi 9000t) + 7\cos(2\pi 11500t)$



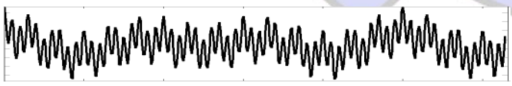
Plot it in frequency domain (F.D.)!

2.4.4 Band Reject Filter (BRF)

$B = (0 \to f_L) + (f_H \to \infty)$



$y(t) = 2 + 3\cos(2\pi 800t) + 4\cos(2\pi 1200t + 45°) + 5\cos(2\pi 6000t) + 9\cos(2\pi 20000t)$



2.4.5 All Pass Filter (APF)

H.W. Find the output signal of the mentioned filters if the input signal is:

$$x(t) = -2.3 + 7\cos(2\pi 50t) + \cos(2\pi 950t) + 3\cos(2\pi 2340t + 45°) + 8\cos(2\pi 6720t) + 4.4\cos(2\pi 8800t) + 5.7\cos(2\pi 11000t) + 6\cos(2\pi 1400t) + 9\cos(2\pi 10^6 t)$$

Part 3 ANALOG MODULATION

3.1 INTRODUCTION

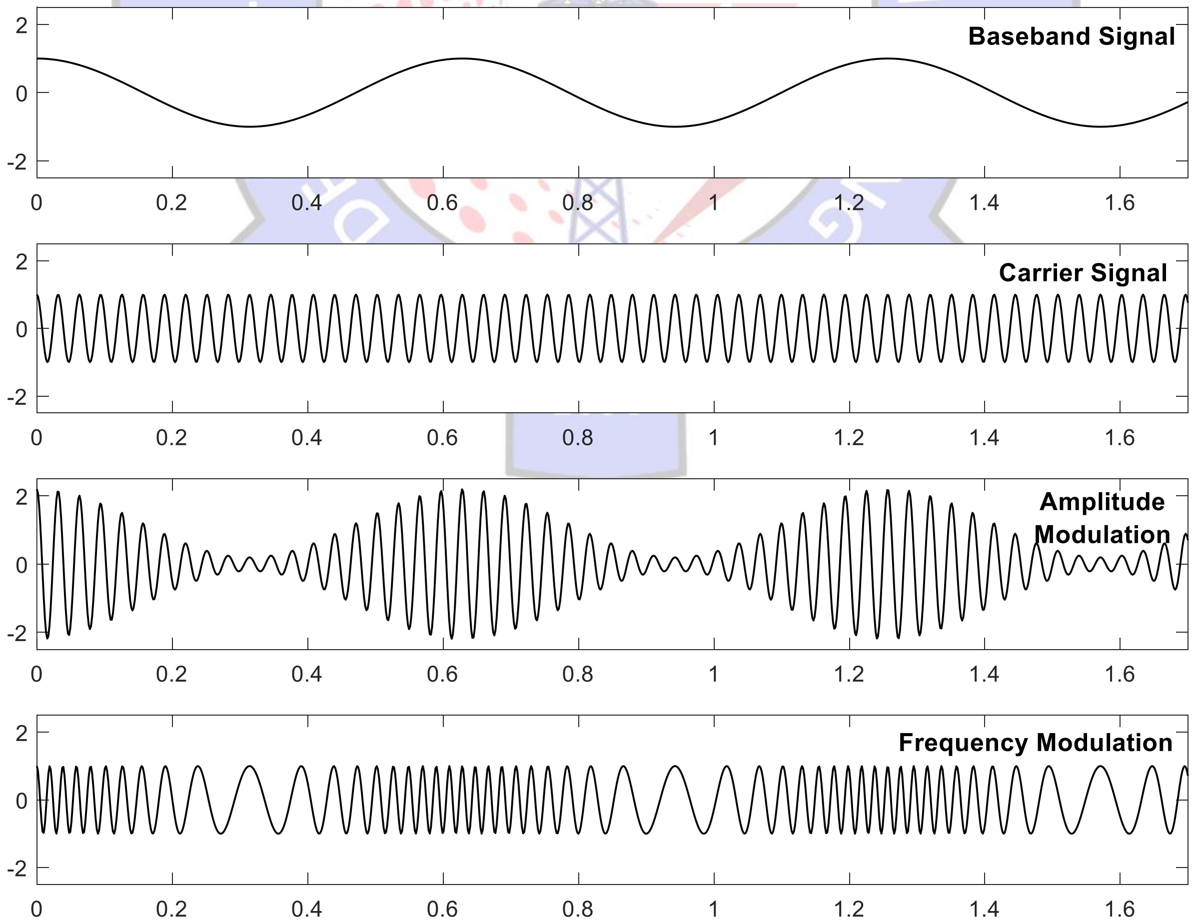
Baseband signals produced by various sources are not always suitable for direct transmission over a given channel. These signals are further modified to facilitate transmission using 'modulation', which is simply: is a process that causes a shift in the range of frequencies of a signal.

In this part, we study the principles of continuous-wave (CW) modulation. This analog form of modulation uses a sinusoidal carrier whose amplitude or angle is varied in accordance with a message signal.

A carrier is a sinusoid of high frequency and one of its parameters (amplitude, frequency or phase) is varied in proportion to the Baseband signal. Accordingly, we have Amplitude Modulation (AM), Frequency Modulation (FM) And Phase Modulation (PM).

$$y(t) = a(t)\cos\{\theta(t) \cdot t + \varphi(t)\}$$

Changing only [$a(t)$ for AM] or [$\theta(t)$ for FM] or [$\varphi(t)$ for PM]



At the receiver, the modulated signal must pass through the reverse process called Demodulation, to reconstruct the Baseband signal.

FM and PM are very close relatives (in fact you can't have one without the other). Hence, we will consider AM and FM only. Both are used in ordinary radio broadcasts. Commercial radio stations are licensed to use carrier frequencies between about 500kHz to 1600kHz using AM, and frequencies between 88MHz and 108MHz using FM.

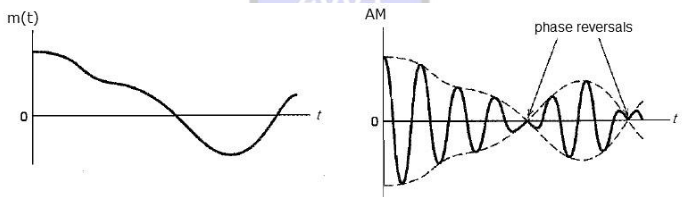
Among the most important reasons for using modulation:

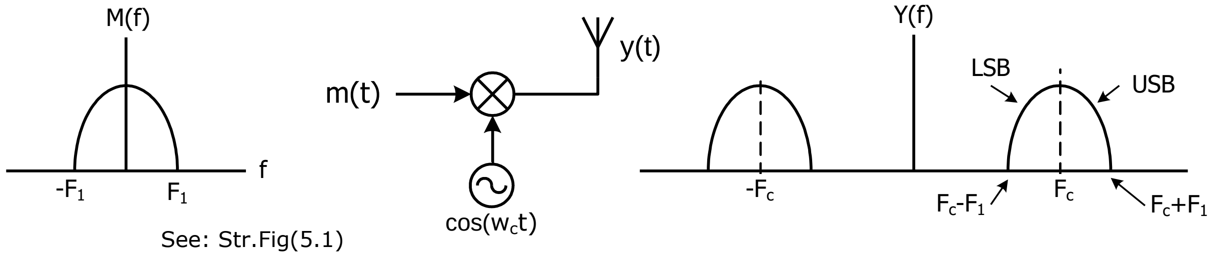
- Ease of Radiation: For efficient radiation of electromagnetic energy, the antenna length must be ≥ λ/10. For many baseband signals, the wavelength is too large for reasonable antenna dimension. For example, the main frequency band of speech = (100→3000)Hz or λ=(100→3000)km i.e. the antenna length ≈ 300km !! ; But if we modulate such audio signal using a 1MHz carrier, the required length for the antenna becomes ≈ 30m which is reasonable size.
- Simultaneous Transmission of Several Signals: Most Baseband signals occupy the same frequency band, so they cannot be transmitted over the same channel at the same time. So, via modulation, it is possible to send several signals over the same channel at the same time.
- Overcome the Channel Problems.

3.2 Amplitude Modulation

3.2.1 DSB-SC (Double Side Band – Suppressed Carrier)

Let $m(t)$ = Baseband signal, the transmitted modulated signal is $y(t) = m(t)\cos(\omega_c t)$





See: Str.Fig(5.1)

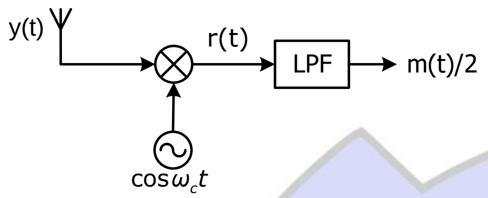
AM shifts the frequency spectrum of a signal from the zero to $\pm f_c$ without changing its shape.

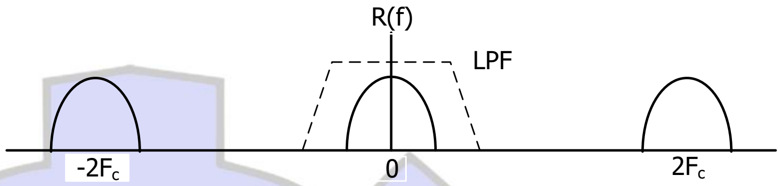
Here, the bandwidth of the AM signal is double of $m(t)$, so it is called DSB:

Positive frequency of $m(t)$ → Upper Side Band (USB)

Negative frequency of $m(t)$ → Lower Side Band (LSB)

Demodulation is restoring the shifted spectrum of $m(t)$ in $y(t)$ back to its original position:



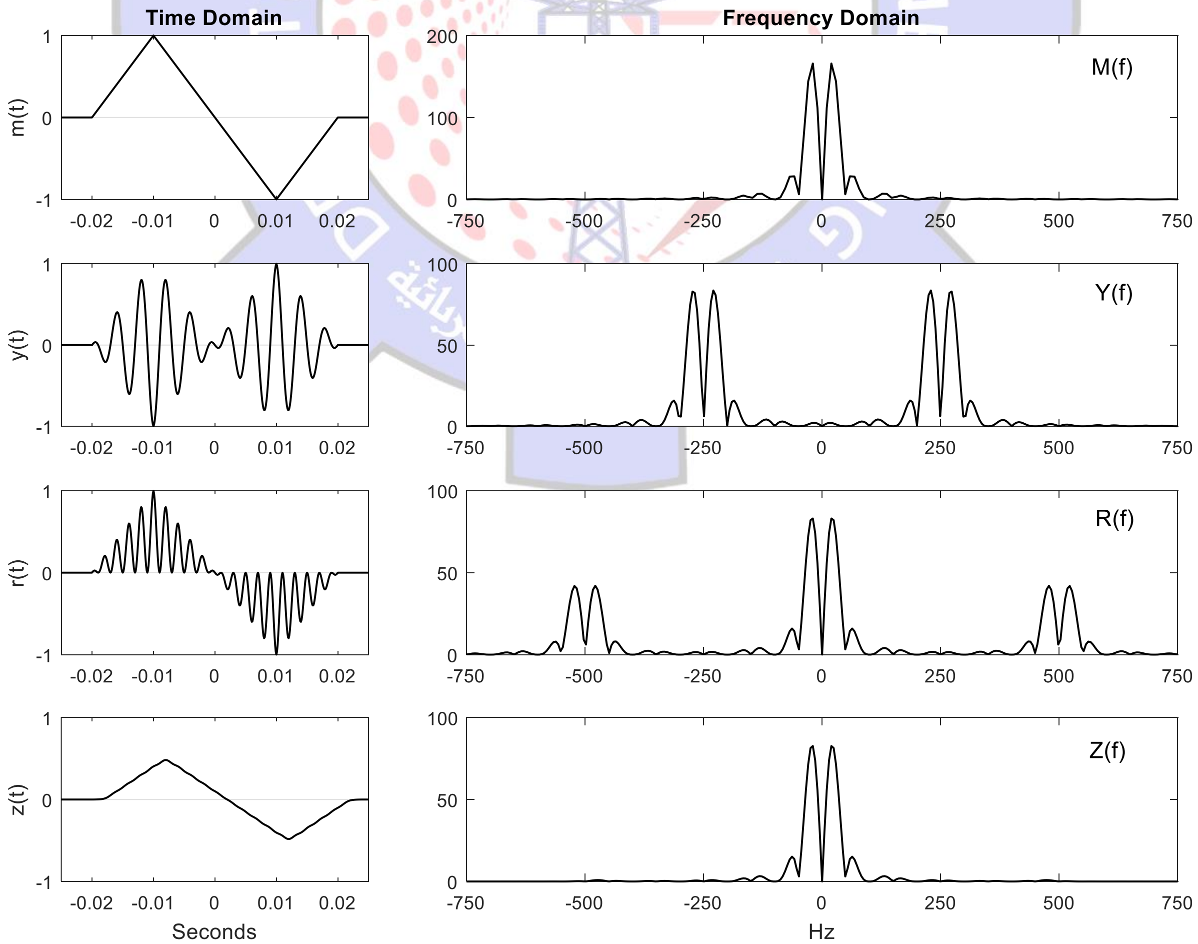


$$r(t) = y(t) \times \cos(\omega_c t) = m(t)\cos^2(\omega_c t)$$

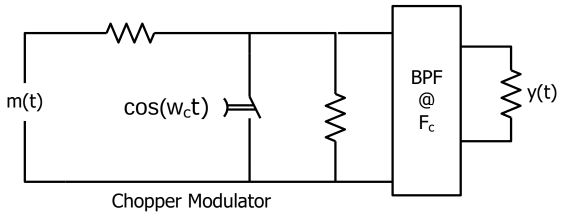
$$= \frac{1}{2}m(t) + \frac{1}{2}m(t)\cos(2\omega_c t)$$

The last term is eliminated by LPF. So, the output becomes:

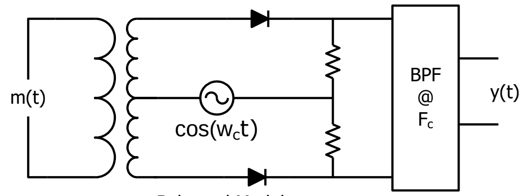
$$z(t) = \frac{1}{2}m(t)$$



Generation of DSB-SC



Chopper Modulator



Balanced Modulator

Demodulation of DSB-SC

Due to several reasons, the locally generated carrier signal at the reception end has some differences in frequency and/or phase. If the signal $y(t) = m(t)\cos\omega_c t$ is received, and the receiver carrier is $\cos[(\omega_c + \Delta)t + \theta]$, then

$$r(t) = y(t)\cos[(\omega_c + \Delta)t + \theta]$$

$$= m(t)\cos(\omega_c t)\cos[(\omega_c + \Delta)t + \theta]$$

$$= \frac{1}{2}m(t)\cos(\Delta t + \theta) + \frac{1}{2}m(t)\cos[(2\omega_c + \Delta)t + \theta]$$

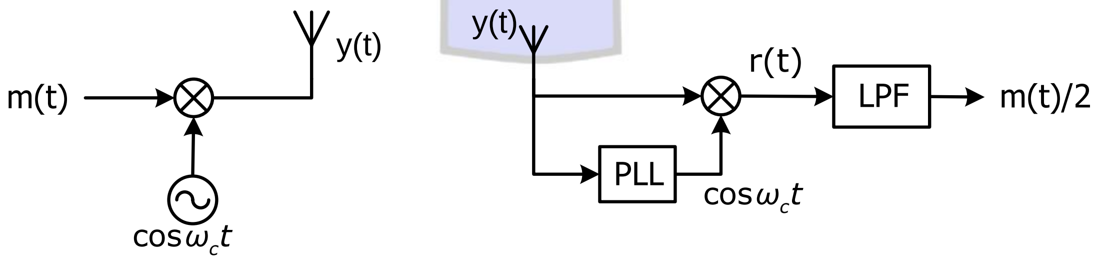
The second term will be removed by LPF, yields $m(t)/2$ multiplied by a factor ≤ 1, as:

$$z(t) = \frac{1}{2}m(t)\cos(\Delta t + \theta)$$

The value of this undesirable scale is governed by Δ and θ. Or:

- $\cos(\Delta t + \theta) = 1$ when Δ & $\theta = 0$ (Synchronous or Coherent Reception)
- $\cos(\Delta t + \theta) < 1$ when Δ & $\theta \neq 0$

So, it is important to perform the synchronous detection of DSB to maximize the output. This can be done through a PLL: when the modulator and the demodulator are remotely located, all synchronous receivers must involve a PLL to re-generate a fresh and in-phase carrier.



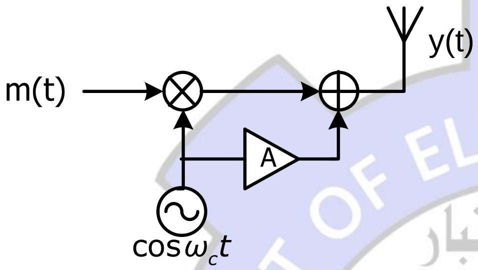
3.2.2 DSB-LC (DOUBLE SIDE BAND – LARGE CARRIER)

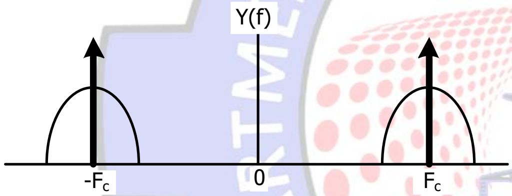
Suppressed carrier signals require complicated circuits at receivers for synchronization. To use inexpensive receivers, the carrier information is included as a part of the waveform being transmitted in the same spectral width. This is achieved by putting a large carrier term at f_c . This technique is used in the commercial broadcasting stations.

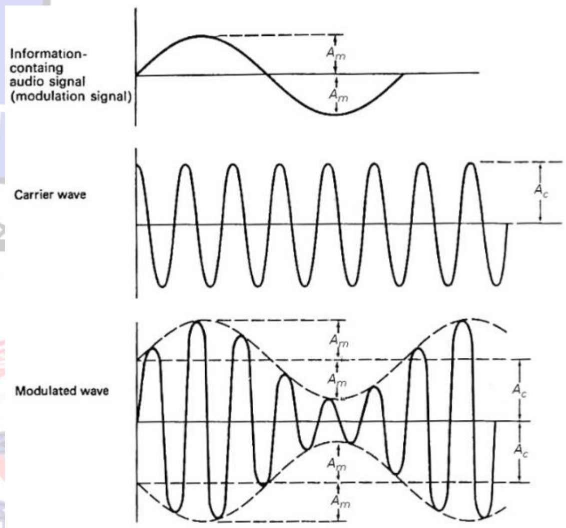
$$y(t) = A_c \cos(\omega_c t) + m(t)\cos(\omega_c t)$$

$$= \{A_c + m(t)\}\cos(\omega_c t)$$

Where A_c = the peak value of the un-modulated carrier







If we set $m(t) = A_m \cos \omega_m t$, where: A_m as the *minimum* value of $m(t)$, f_m is the maximum frequency component of the baseband signal, then:

$$y(t) = A_c\left\{1 + \frac{A_m}{A_c}\cos(\omega_m t)\right\}\cos(\omega_c t) = A_c\{1 + \mu\cos(\omega_m t)\}\cos(\omega_c t)$$

Where μ is the modulation index:

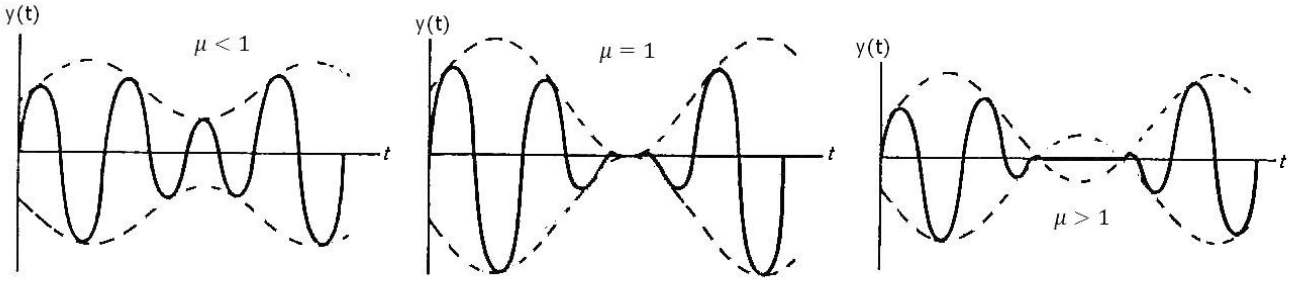
$$\mu = \frac{|\text{min. value of modulating signal}|}{\text{peak value of un-modulated carrier}} = \frac{|A_m|}{A_c}$$

To make proper modulation, the value of carrier amplitude A_c must be set as:

$$A_c - |A_m| \geq 0 \;\rightarrow\; A_c \geq |A_m|$$

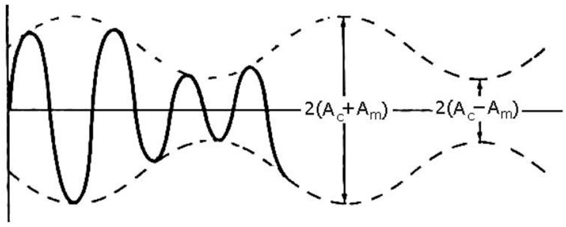
But generally, according to the values of A_c and $|A_m|$:

- If $\mu < 1$, under-modulation, $y(t)$ trace out $m(t)$, hence correctly demodulation.
- If $\mu = 1$, optimum modulation/demodulation. *Why?*
- If $\mu > 1$, over-modulation, distortion in demodulation.



Note that:

$$\mu = \frac{\text{max p-p} - \text{min p-p}}{\text{max p-p} + \text{min p-p}} = \frac{2(A_c + A_m) - 2(A_c - A_m)}{2(A_c + A_m) + 2(A_c - A_m)} = \frac{A_m}{A_c}$$

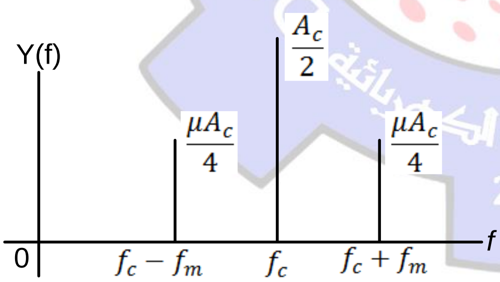


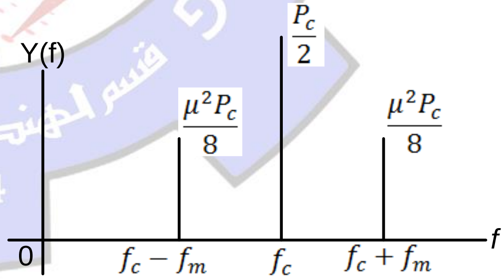
SPECTRUM OF DSB-LC

For illustration, we'll reconsider $m(t) = A_m \cos(\omega_m t)$, hence the voltage or current equation will be:

$$y(t) = A_c\{1 + \mu \cos(\omega_m t)\} \cos(\omega_c t)$$

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





POWER IN DSB-LC

The carrier does not contain any information about $m(t)$, but it is the price to make cheep receivers available.

If $m(t) = A_c \cos(\omega_c t) + A_c \mu \times m(t) \cos(\omega_c t)$, using: R=1Ω, $m(t) = \cos(\omega_m t)$, the mean power (mean square value) is:

$$P_{DSBLC} = \overline{y^2(t)} = A_c^2\,\overline{\cos^2(\omega_c t)} + \frac{\mu^2 A_c^2}{4}\overline{\cos^2[2\pi(f_c - f_m)t]} + \frac{\mu^2 A_c^2}{4}\overline{\cos^2[2\pi(f_c + f_m)t]}$$

$$= \frac{A_c^2}{2} + \frac{A_c^2\mu^2}{8} + \frac{A_c^2\mu^2}{8} \equiv P_c + P_{USB} + P_{LSB}$$

$$= \frac{A_c^2}{2} + \frac{A_c^2\mu^2}{4} = P_c + P_{SB} = P_c + \frac{\mu^2}{2}P_c$$

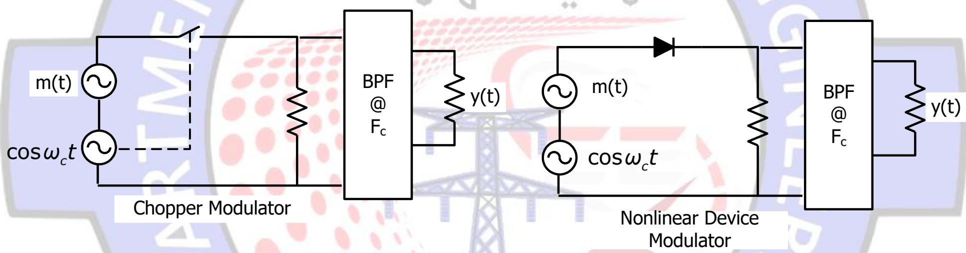
$$\therefore P_T = P_c\left(1 + \frac{\mu^2}{2}\right)$$

Now, let the information power to the total power ratio is Transmission Efficiency:

$$\rho = \frac{P_{SB}}{P_T} = \frac{P_{DSBSC}}{P_{DSBLC}} = \frac{\mu^2}{2 + \mu^2}$$

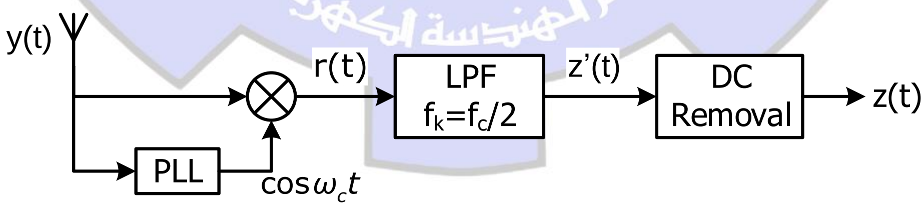
Since the best demodulation occurs at $\mu \le 1$, so best ρ for DSB-LC system is 33%, i.e. 67% of the total power is spent in the carrier and represents a wasted power.

Generation of DSB-LC

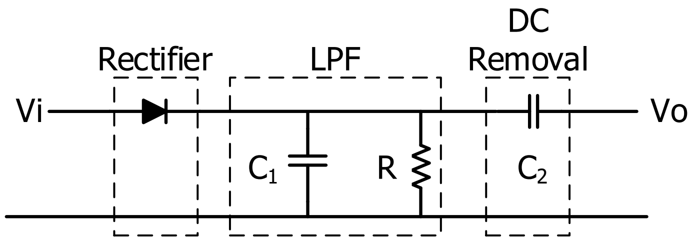


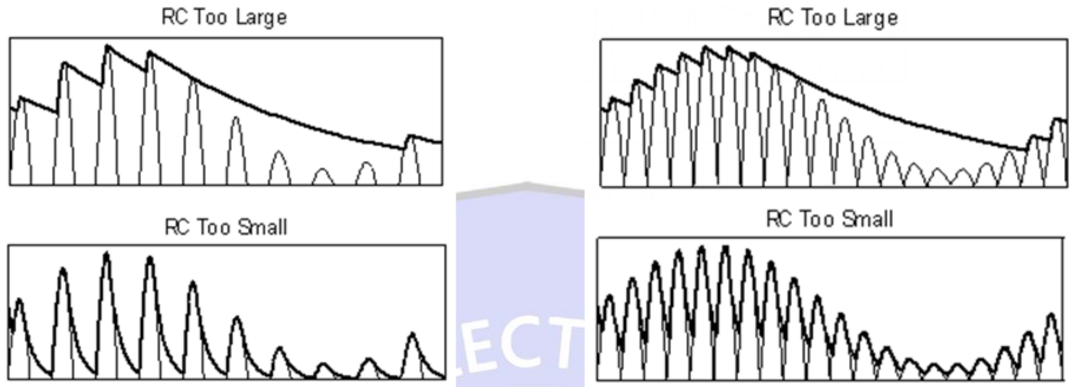
Demodulation of DSB-LC

(1) Synchronous (coherent) detector: (via PLL)

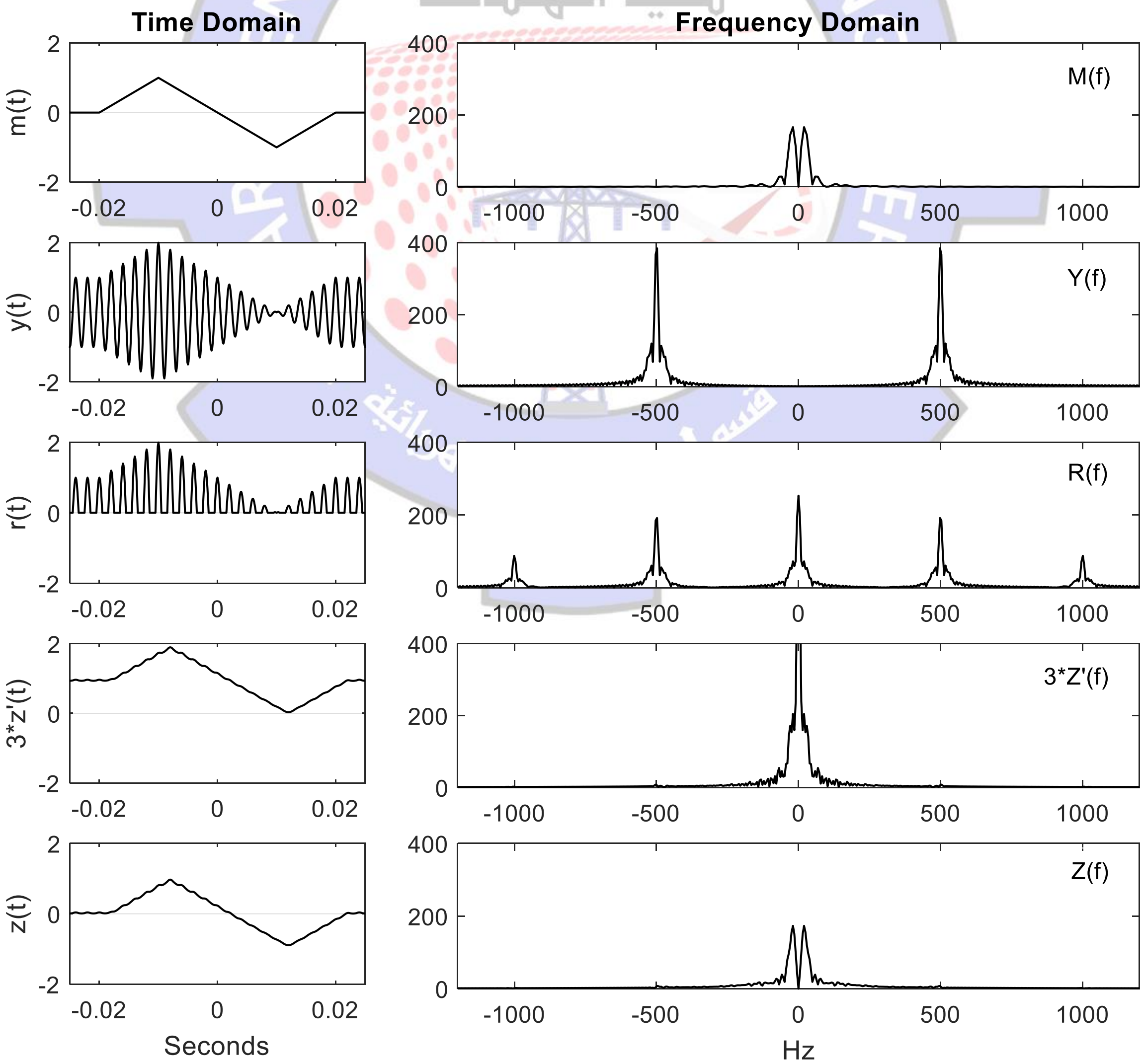


(2) Asynchronous: (using an envelope detector), as $m(t)$ is available in the envelope of DSB-LC signal, it is possible to use simple envelop detector circuit as a demodulator. Why we use asynchronous detection?



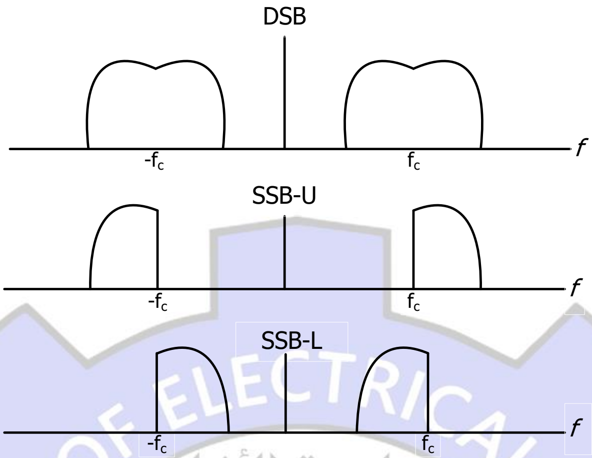


The following is an example that illustrates the DSB-LC.



3.2.3 SSB (Single Side Band)

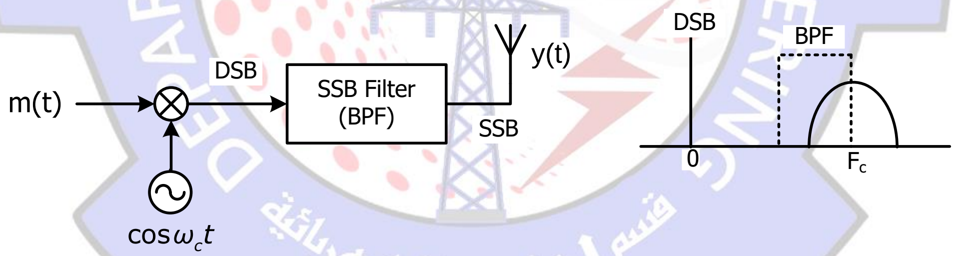
DSB method results in doubling of $m(t)$ bandwidth, which is disadvantage when channel is crowded or expensive. So, we can use only one sideband to be transmitted as it contains all the information about $m(t)$.



Generation of SSB

(1) Filtering method:

SSB can be generated by filtering the DSB signal. In practice, this operation is not easy because it is difficult to meet such filter requirements.



(2) Phase shift method:

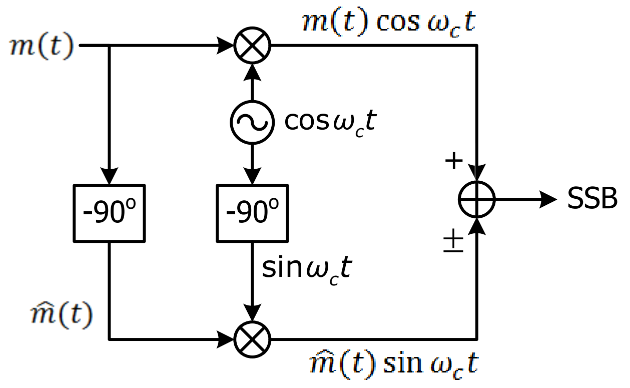
Since the ordinary AM signal is $y(t) = m(t)\cos(\omega_c t)$, letting $m(t) = \cos(\omega_m t)$, so:

$y(t) = \cos(\omega_m t) \times \cos(\omega_c t) = \frac{1}{2}\cos(\omega_c + \omega_m)t + \frac{1}{2}\cos(\omega_c - \omega_m)t =$DSB

Yields:

$\cos(\omega_c + \omega_m)t = \cos(\omega_c t)\cos(\omega_m t) - \sin(\omega_c t)\sin(\omega_m t)\ \equiv$ Upper SSB

$\cos(\omega_c - \omega_m)t = \cos(\omega_c t)\cos(\omega_m t) + \sin(\omega_c t)\sin(\omega_m t)\ \equiv$ Lower SSB



But $\cos\theta = \sin(\theta \pm 90°)$, hence:

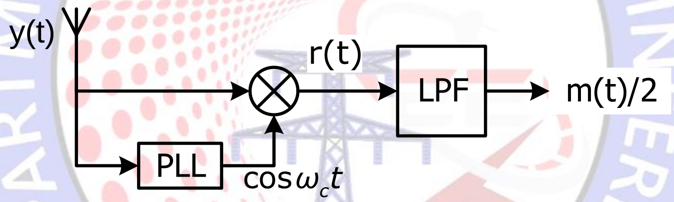
Upper SSB is $y(t) = m(t)\cos(\omega_c t) - \hat{m}(t)\sin(\omega_c t)$

Lower SSB is $y(t) = m(t)\cos(\omega_c t) + \hat{m}(t)\sin(\omega_c t)$

Where $\hat{m}(t)$ is shifting the phase of $m(t)$ by 90°

The main problem of SSB systems is the practical realization of the 90° phase shifter.

Demodulation of SSB



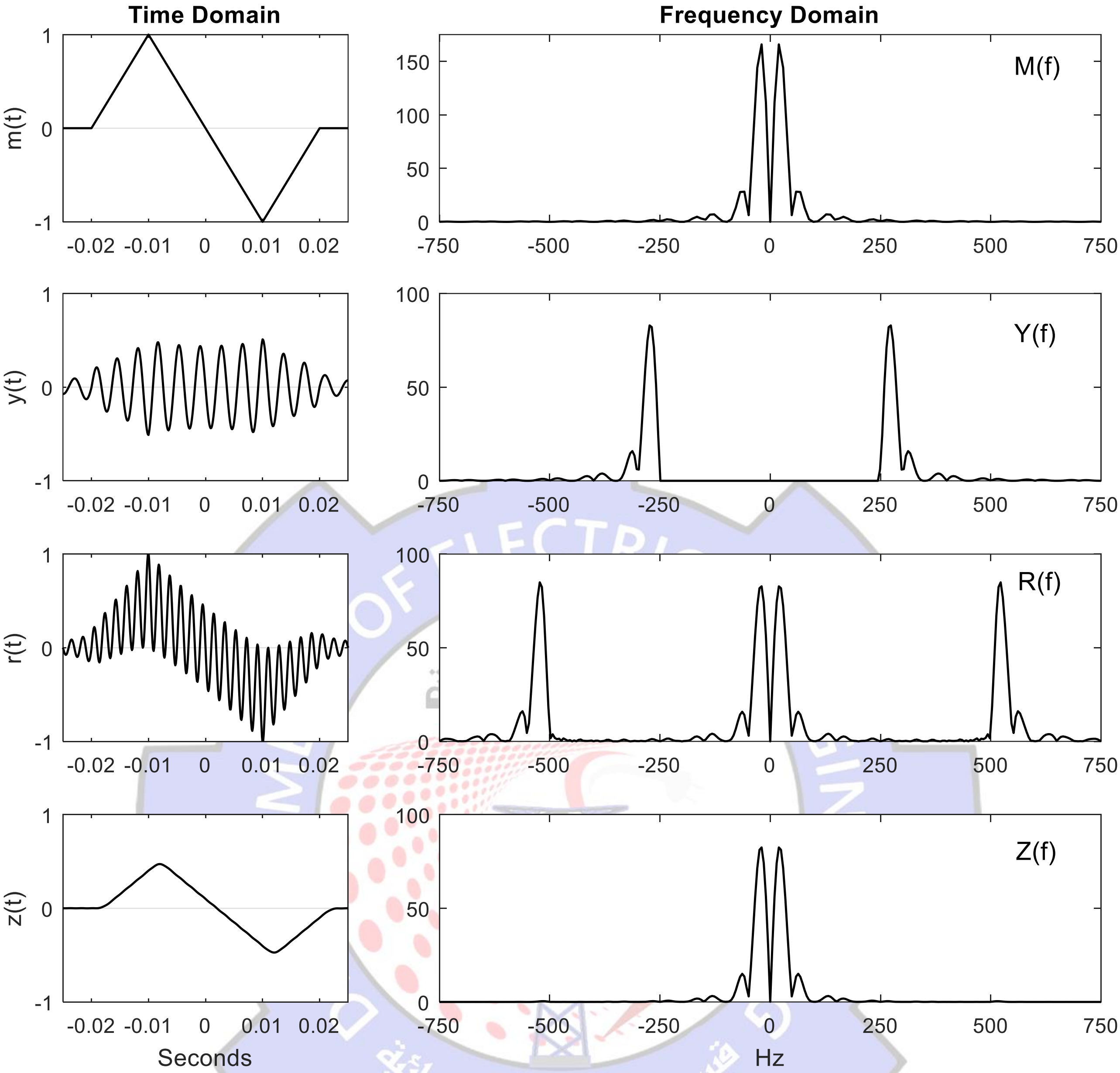
$y(t) = m(t)\cos(\omega_c t) \pm \hat{m}(t)\sin(\omega_c t)$

$R(t) = y(t)\cos(\omega_c t)$

$$= m(t)\cos^2(\omega_c t) \pm \hat{m}(t)\sin(\omega_c t)\cos(\omega_c t)$$

$$= \frac{1}{2}m(t) + \frac{1}{2}m(t)\cos(2\omega_c t) \pm \frac{1}{2}\hat{m}(t)\sin(2\omega_c t)$$

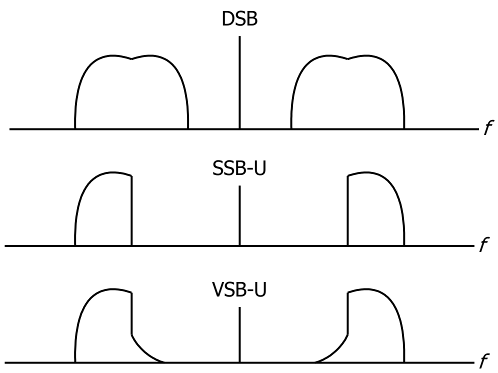
The high frequency parts are removed by LPF, yields: $m(t)/2$.



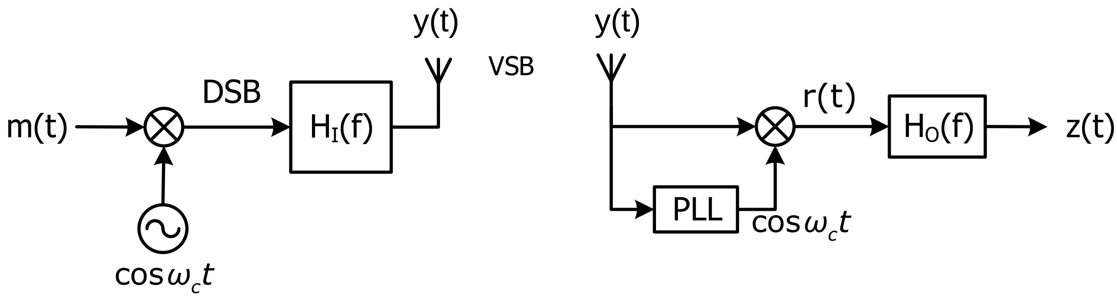
3.2.4 VSB (Vestigial Side Band)

Because of double BW usage in DSB method, and the selective-filtering and phase shifter limitations in SSB method, VSB is a compromise between DSB and SSB.

Instead of rejecting one sideband completely as in SSB, a gradual cutoff of one sideband is accepted: (VSB BW=one sideband +25% of the other sideband)



GENERATION/DETECTION OF VSB



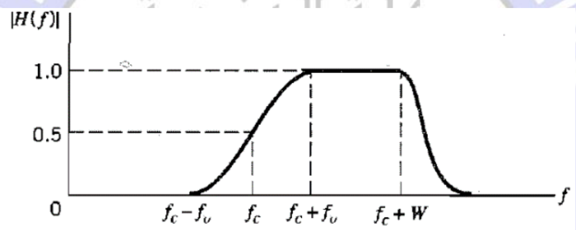
$Y_{VSB}(f) = \{M(f+f_c) + M(f-f_c)\} \times H_I(f)$

$r(t) = y_{VSB}(t)\cos\omega_c t \Leftrightarrow R(f) = Y_{VSB}(f+f_c) + Y_{VSB}(f-f_c)$

$$Z(f) = M(f)\{H_I(f+f_c) + H_I(f-f_c)\} \times H_O(f)$$

And to obtain $z(t) = m(t)$,

$$H_O(f) = \frac{1}{H_I(f+f_c) + H_I(f-f_c)}$$



The VSB technique is used in the analog TV broadcasting system for video signals. The following figure is a typical analog TV signal (Audio Video), *why?*

3.2.5 AM SUMMARY

(1) Full amplitude modulation (DSB-LC), in which the upper and lower sidebands are transmitted in full, accompanied by the carrier wave. Accordingly, demodulation of an AM signal is done rather simply in the receiver by using an envelope detector, for example. It is for this reason we find that full AM is commonly used in commercial AM radio broadcasting, which involves a single powerful transmitter and numerous receivers that are relatively inexpensive to build.

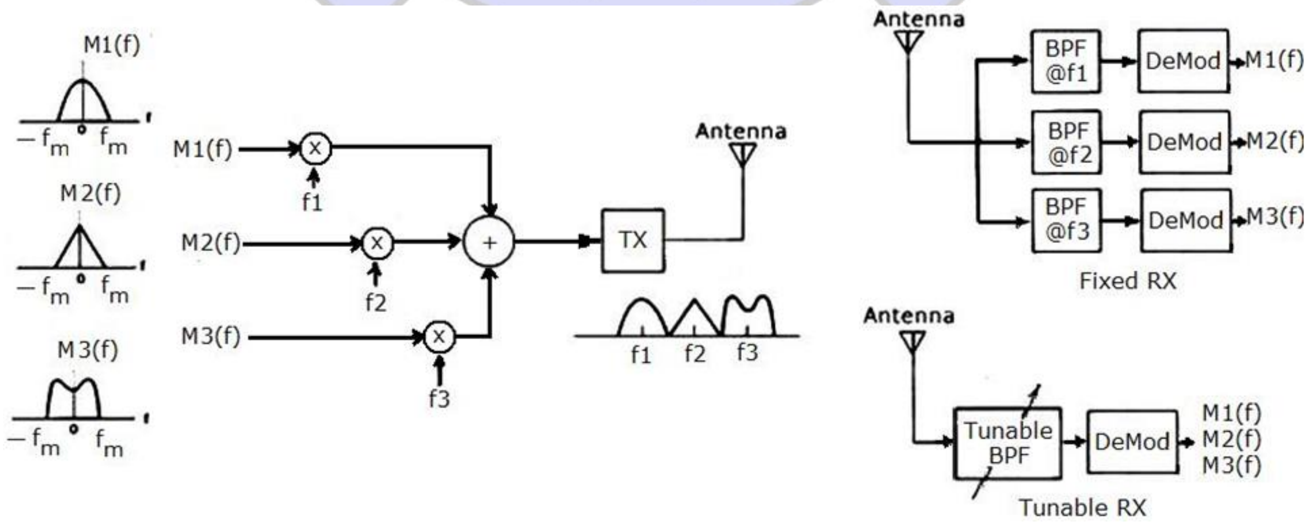
(2) DSB-SC, in which only the upper and lower sidebands are transmitted. The suppression of the carrier wave means that DSB-SC modulation requires much less power than full AM to transmit the same message signal; this advantage of, however, attained at the expense of increased receiver complexity. DSB-SC is therefore well suited for point-to-point communication involving one transmitter and one receiver.

(3) SSB, in which only the upper sideband or lower sideband is transmitted. This is the optimum form of CW modulation if is required the minimum transmitted power and the minimum channel bandwidth for conveying a message signal from one point to another. However, its use is limited to message signals with an energy gap centered on zero frequency beside its complex systems.

(4) VSB, in which almost all of one sideband and a vestige of the other sideband are transmitted. It requires a channel bandwidth that is between that required for SSB and DSB-SC systems, and the saving in bandwidth can be significant if modulating signals with large bandwidths are being handled, as in the case of television signals and high-speed data.

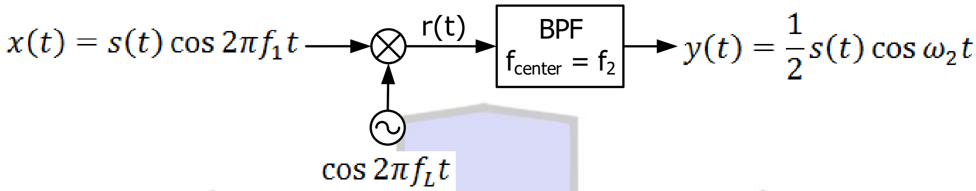
3.3 Frequency Division Multiplexing (FDM)

It is possible to send several signals simultaneously by choosing different carrier frequency for each signal. These carriers must be chosen so that the signals spectra do not overlapping. This technique is used basically in the commercial radio and TV stations. In practical AM radio broad-casting stations, they are allocated 10kHz per station i.e. 5kHz per sideband. *Is it sufficient?*



3.4 FREQUENCY CONVERSION

Which sometimes referred to as frequency translating, changing, mixing, or heterodyning. Suppose that we have a modulated wave $s(t)$ whose spectrum is centered on a carrier frequency f_1 and the requirement is to translate it upward/downward in frequency such that its carrier frequency is changed from f_1 to a new value f_2. This requirement may be accomplished using the mixer shown in Figure below.



Where the local oscillator frequency $f_L = f_1 \pm f_2$

If $x(t) = s(t)\cos(\omega_1 t)$ then

$$r(t) = x(t)\cos(\omega_L t)$$

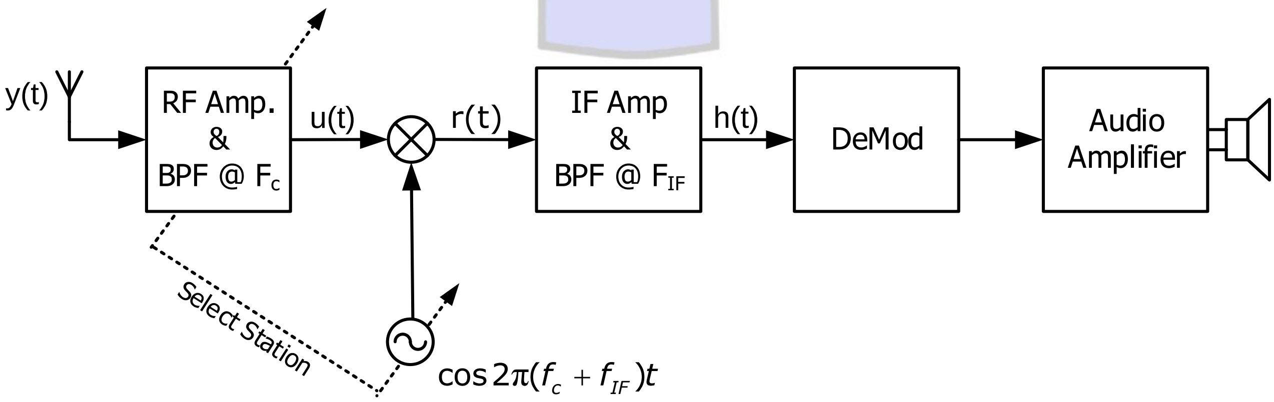
$$= s(t) \times \cos(\omega_1 t) \times \cos[(\omega_1 + \omega_2)t]$$

$$= \tfrac{1}{2}s(t)\cos(\omega_2 t) + \tfrac{1}{2}s(t)\cos[(2\omega_1 + \omega_2)t]$$, the last term removed by the BPF Amplifier.

$$\therefore y(t) = \frac{1}{2}s(t)\cos(\omega_2 t)$$

3.5 SUPER-HETERODYNE RECEIVER

For commercial radio receivers, it is difficult to design a good demodulator for wide range of frequencies. In this type, the carrier of the incoming signal is translated to a fixed frequency value (called intermediate frequency $f_{IF} = 455\text{kHz}$), and finally demodulated by well designed system.



If $y(t) = m(t)\cos\omega_c t$ then for the moment, let $u(t) = y(t)$, the frequency convertor results:

$$h(t) = \frac{1}{2}m(t)\cos(\omega_{IF}t)$$

Which represent the main message signal $m(t)$ re-modulated by new carrier of the value f_{IF} .

IMAGE STATION PROBLEM

Now, suppose there is another station with a carrier $f_C + 2f_{IF}$ is available at the receiving input, so:

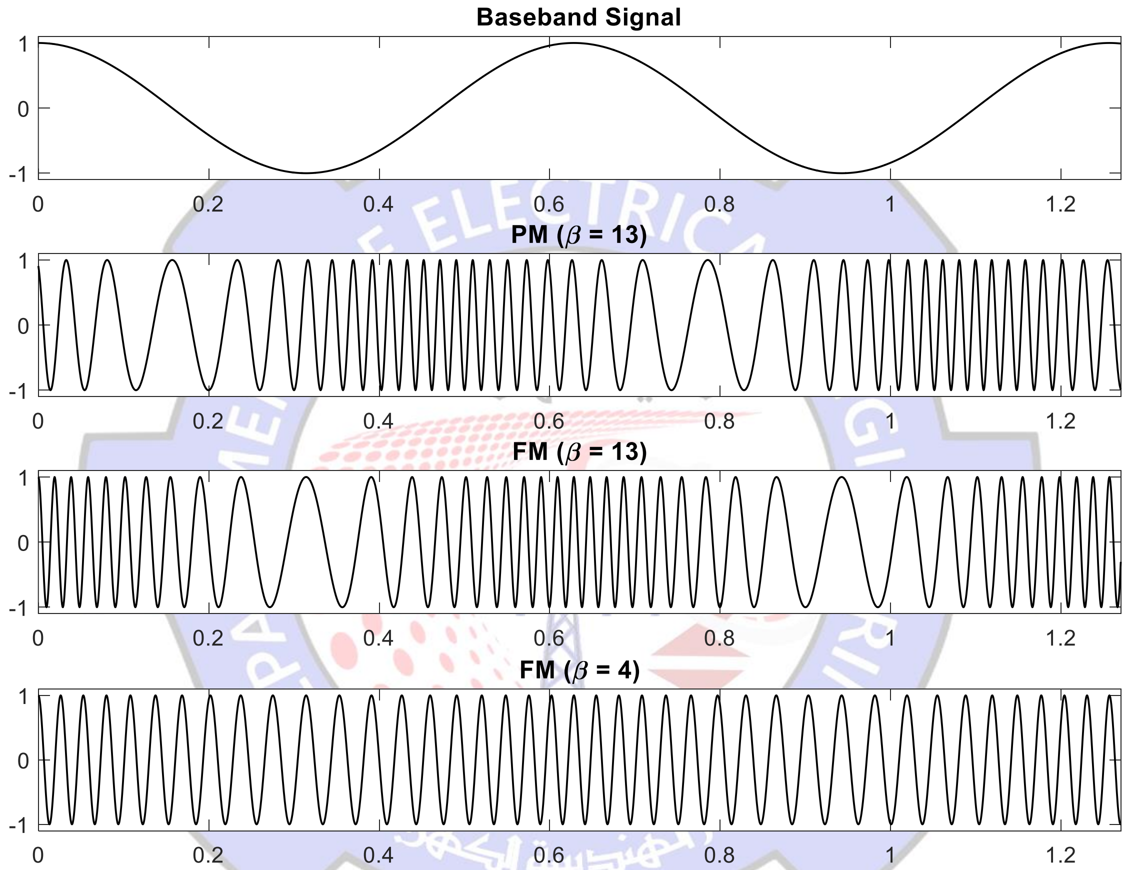
$$u(t) = y_1(t) + y_2(t)$$
$$= m_1(t)\cos(\omega_c t) + m_2(t)\cos[(\omega_c + 2\omega_{IF})t]$$
$$r(t) = \frac{1}{2}m_1(t)\{\cos(\omega_{IF}t) + \cos[(2\omega_c + \omega_{IF})t]\} + \frac{1}{2}m_2(t)\{\cos(\omega_{IF}t) + \cos[(2\omega_c + 3\omega_{IF})t]\}$$
$$\therefore h(t) = \frac{1}{2}[m_1(t) + m_2(t)]\cos(\omega_{IF}t)$$

Which means both m_1 and m_2 will pass to the demodulator (overlapping). So, in the super-heterodyne receivers, for each desired station frequency f_c there is an *image station* at frequency $= f_C + 2f_{IF}$ which called the *image frequency*. The RF-Amplifier stage of the super-heterodyne receiver prevents an image station from passing.

3.6 FREQUENCY MODULATION

We saw in AM that the information contents of $m(t)$ is transmitted through changing the amplitude of the carrier in proportion to the amplitude of $m(t)$. In the angle modulation, the information content is transmitted through changing the frequency of the carrier instead. The instantaneous frequency of the carrier is determined by the value of the baseband signal.

In this modulation type, the carrier frequency varies in a proportion to the amplitude of the message signal $m(t)$.



3.6.1 DEFINITIONS

Let us consider the carrier function as: $y(t) = A_c \cos \theta(t)$, where $\theta(t)$ is the angle.

Since the instantaneous frequency $\omega_i(t)$ is the slop of $\theta(t)$, or $\omega_i(t) = \frac{d\theta(t)}{dt}$, then

$$\theta(t) = \int_0^t \omega_i(\varepsilon) d\varepsilon$$

In the case of the Phase Modulation (PM), the angle varies linearly with $m(t)$ as:

$\theta(t) = \omega_c t + k_P m(t)$

where: k_P = modulation sensitivity = constant, and ω_c = the center frequency at $m(t) = 0$.

The resulting PM wave is:

$$y_{PM}(t) = A_c \cos[\omega_c t + k_P m(t)]$$

The instantaneous frequency in this case is:

$$\omega_i(t) = \frac{d\theta(t)}{dt} = \omega_c + k_P m'(t)$$

In words: the instantaneous frequency varies linearly with the derivative of the modulating signal.

In the case of the Frequency Modulation (FM), the instantaneous frequency $\omega_i(t)$ is varied linearly with the modulating signal, or:

$$\omega_i(t) = \omega_c + k_F m(t)$$

where: k_F = modulation sensitivity = constant. The angle of the carrier becomes:

$$\theta(t) = \int_0^t [\omega_c + k_F m(\varepsilon)]d\varepsilon + \theta_0$$

If we assume $\theta_0 = 0$, the FM signal becomes:

$$y_{FM}(t) = A_c \cos\left(\omega_c t + k_F \int_0^t m(\varepsilon)d\varepsilon\right)$$

In words: the instantaneous frequency varies linearly with the integral of the modulating signal.

FM and PM are closely related to each other; if we know the properties of the one, we can determine those of the other. For this reason, the material on angle modulation hereafter is devoted to FM.

To simplify the analysis of this equation, we consider $m(t) = A_m \cos(\omega_m t)$. So,

$$\omega_i(t) = \omega_c + k_F A_m \cos \omega_m t = \omega_c + \Delta_f \cos \omega_m t$$

Where $\Delta_f = A_m k_F$ =is the peak frequency deviation or the maximum frequency shift away from f_c (Note: $\Delta_f \ll f_C$). The instantaneous angle of the carrier is:

$$\theta(t) = \int_0^t \omega_i(\varepsilon)d\varepsilon = \omega_c t + \frac{\Delta_f}{\omega_m} \sin \omega_m t\,, \text{ at } \theta_0 = 0$$

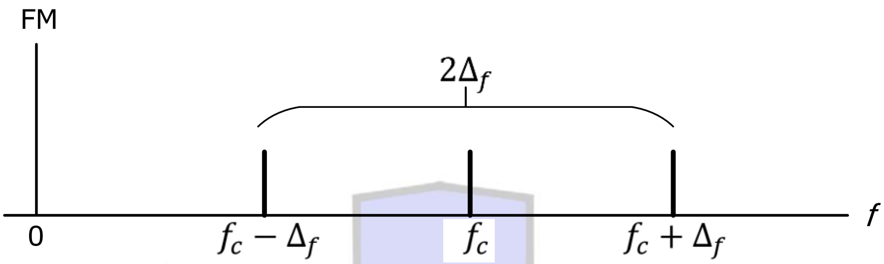
$$\Rightarrow \theta(t) = \omega_c t + \beta \sin \omega_m t$$

Where $\beta = \frac{\Delta_f}{f_m}$ = modulation index of the FM signal

Hence, the formula of the FM waveform will be:

$$y_{FM}(t) = A_c \cos[\omega_c t + \beta \sin(\omega_m t)]$$

Since the peak deviation is given by Δ_f, the carrier swing $= 2\Delta_f$.



Note that:

- There is no 'over-modulation' situation with FM signal.
- β can take on any value from 0 to infinity. Its range is not limited as it is for AM.
- As β is increased, the signal becomes more resistant to interfering noise however occupies more bandwidth.

3.6.2 Spectrum of FM Signals

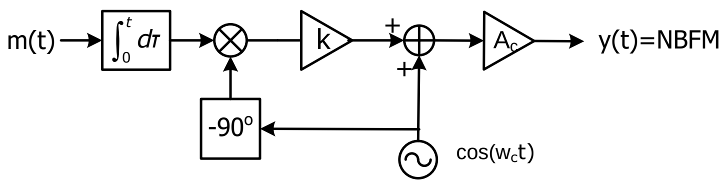
The spectrum of FM signals is rather 'messy' as it has may different sidebands spaced at multiples of f_m from the carrier. As a result, the bandwidth needed to accommodate an FM signal is greater than that for an AM signal having the same modulating frequency, (Except for NBFM).

According to the value of β, we have the following FM formats:

Narrow Band FM

In the FM systems where β is small (< 0.2), several approximations lead to the following method of FM generation:

$$y(t) = A_c \left[\cos(\omega_c t) + k_F \sin(\omega_c t) \times \int_0^t m(\varepsilon) d\varepsilon \right]$$



We can easily notice in the Bessel chart that the spectrum and BW occupied by this NBFM waveform is $2f_m$ (just like the AM-DSB).

WIDE BAND FM

To determine the spectrum of FM signal with values of $\beta > 0.2$ (which is typical), we may simplify matters by using the complex representation of band-pass signals when $m(t) = A_m \cos \omega_m t$:

$$y(t) = A_c \cos[\omega_c t + \beta \sin \omega_m t]$$
$$= \text{Re}[A_c \exp(j\omega_c t + j\beta \sin(\omega_m t))]$$
$$= \text{Re}[\tilde{y}(t) \exp(j\omega_c t)]$$

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

$$\tilde{y}(t) = \text{Re}[A_c \exp(j\beta \sin(\omega_m t))]$$

Thus, unlike the original FM signal $y(t)$, the complex envelope $\tilde{y}(t)$ is a periodic function of time with a fundamental frequency equal to the modulation frequency f_m. We may therefore expand $\tilde{y}(t)$ in the form of a complex Fourier series as follows:

$$\tilde{y}(t) = \sum_{n=-\infty}^{\infty} C_n \exp(j\omega_m n t)$$

where the complex Fourier coefficient C_n is defined by

$$C_n = f_m \int_{-\pi/\omega_m}^{\pi/\omega_m} \tilde{y}(t) \exp(-j\omega_m n t)\, dt$$

$$= f_m A_c \int_{-\pi/\omega_m}^{\pi/\omega_m} \exp(j\beta \sin(\omega_m t) - j\omega_m n t)\, dt$$

Define a new variable $x = \omega_m t$,

$$C_n = \frac{A_c}{2\pi} \int_{-\pi}^{\pi} \exp(j\beta \sin(x) - jnx)\, dx$$

Here the integral on the right-hand side, is recognized as the n^{th} order Bessel function of the first kind and argument β. This function is commonly denoted by the symbol $J_n(\beta)$, as shown by:

$$J_n(\beta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta\sin(x)-nx)}dx$$

Accordingly, we may reduce C_n Equation to $C_n = A_c J_n(\beta)$.

We substitute to get the complex envelope of the FM signal in terms of the $J_n(\beta)$

$$\tilde{y}(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta)\exp(j\omega_m nt)$$

And

$$y(t) = A_c \operatorname{Re}\left[\sum_{n=-\infty}^{\infty} J_n(\beta)\exp[j2\pi(f_c+nf_m)t]\right]$$

$$\therefore y(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta)\cos[2\pi(f_c+nf_m)t] \qquad \text{Eq.1}$$

$$= A_c \begin{Bmatrix} J_0(\beta)\cos[2\pi f_c t] \\ +J_1(\beta)\cos[2\pi(f_c+f_m)t] \quad +J_1(\beta)\cos[2\pi(f_c-f_m)t] \\ +J_2(\beta)\cos[2\pi(f_c+2f_m)t] + J_2(\beta)\cos[2\pi(f_c-2f_m)t] \\ +J_3(\beta)\cos[2\pi(f_c+3f_m)t] + J_3(\beta)\cos[2\pi(f_c-3f_m)t] \\ \ldots \end{Bmatrix}$$

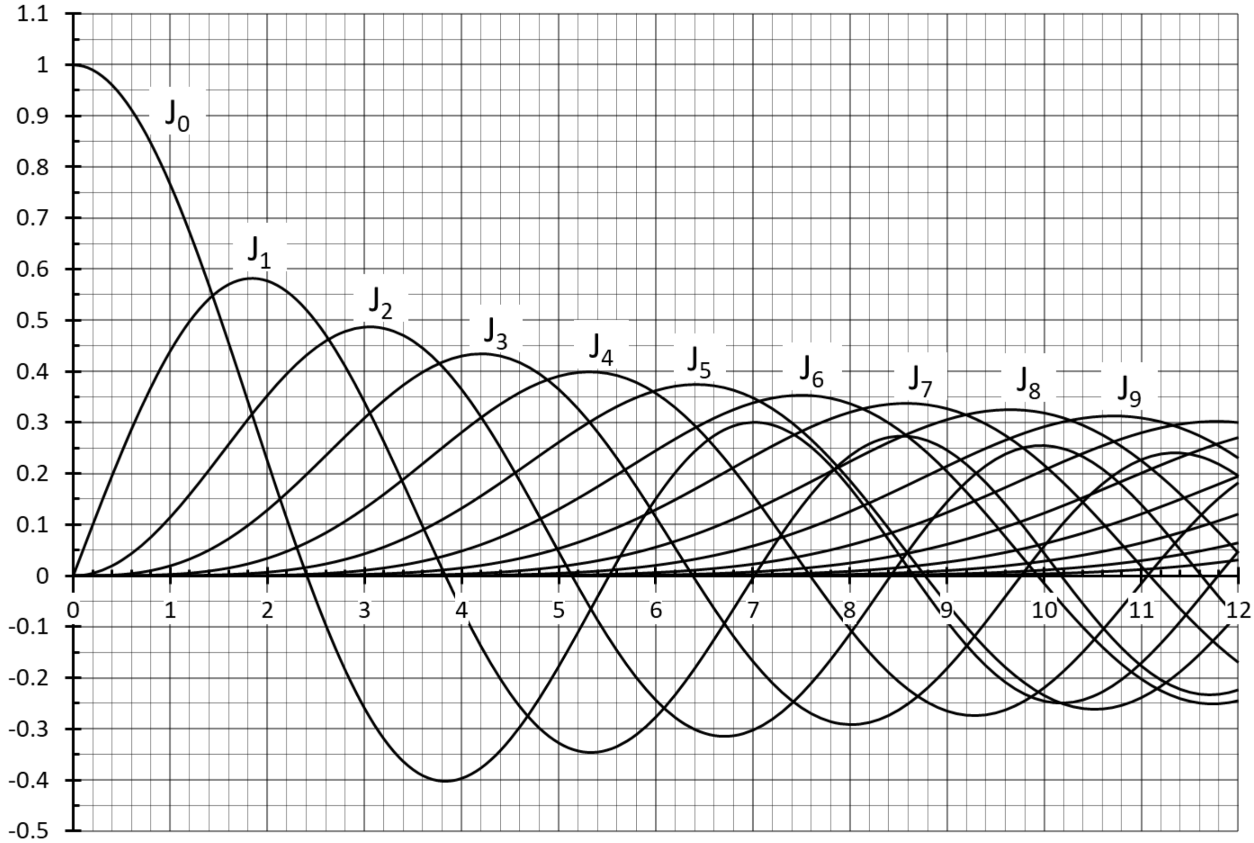
$$= A_c \begin{Bmatrix} J_0(\beta)\cos[2\pi f_c t] \\ +2J_1(\beta)\cos[2\pi(f_c+f_m)t] \\ +2J_2(\beta)\cos[2\pi(f_c+2f_m)t] \\ +2J_3(\beta)\cos[2\pi(f_c+3f_m)t] \\ \ldots \end{Bmatrix}$$

The discrete spectrum of $y(t)$ is obtained by taking the Fourier transforms of both sides, yields:

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

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

In the figure below, we plot the Bessel function $J_n(\beta)$ versus the modulation index β for different positive integer values of n, (see the properties of Bessel function: Str. Page 291)



Thus, using Equations 2 & 3 and the curves of the above Figure, we may observe:

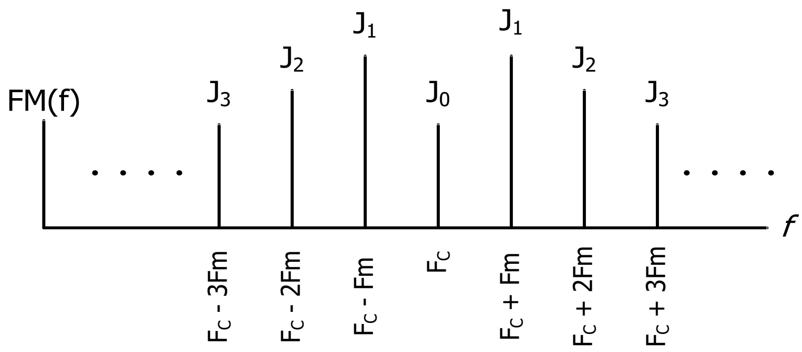
(1) The spectrum of an FM signal contains a carrier component and an infinite set of side frequencies located symmetrically on either side of the carrier at frequency separations of $f_m, 2f_m, 3f_m$ …

(2) For the special case of $\beta < 0.2$, 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 $f_c \pm f_m$. (NBFM)

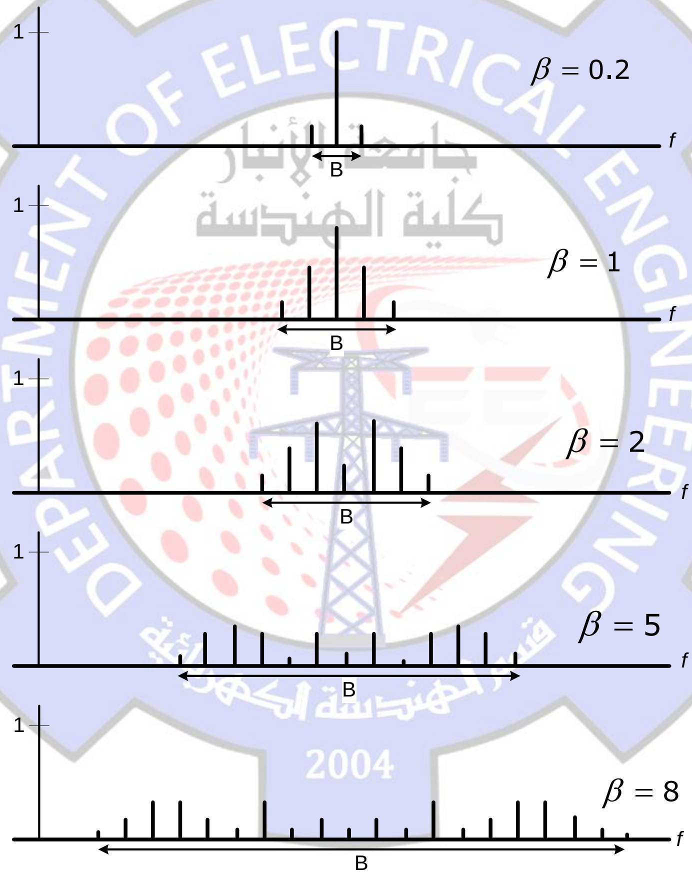
(3) The amplitude of the carrier component varies with β according to $J_0(\beta)$. That is, unlike an AM signal, the amplitude of the carrier component of an FM signal is dependent on β. As the envelope of an FM signal is constant, so that the average power of such a signal is:

$$P = \overline{y^2(t)} = \frac{A_c^2}{2R} \qquad , \qquad P_{Total} = \frac{A_c^2}{2R} \sum_{n=-\infty}^{\infty} J_n^2(\beta)$$

To sketch the spectrum of an FM signal at a certain β, we must get the values of $J_n(\beta)$ from a plot or a table of Bessel function (see **Error! Reference source not found.**). So, the plot may l ook like the following depiction, which is plotted using Eq.2 (Note: we assume $A_c = 2\text{V}$)



It is evident that, the frequency modulation of $m(t) = \cos\omega_m t$ has infinite number of sidebands. However, the magnitude of the spectral components of the higher order sidebands becomes negligible



3.6.3 BANDWIDTH OF AN FM SIGNAL

The bandwidth of an FM signal could be calculated as:

$$B = 2nf_m$$

where n = maximum number of *significant* sideband,

f_m = maximum frequency component of the baseband signal.

If we do not have the Bessel plot or table, we can *approximate* B according to the value of β.

- large β (≥ 12): $n \approx \beta$ so $B = 2nf_m \approx 2\beta f_m = 2\Delta_f$
- very small values of β (≤ 0.2), only J_0 and J_1 are significant, so: $B \approx 2f_m$ (NBFM)
- And generally, B can be approximated for intermediate β values as:
  $B \approx 2(\Delta_f + f_m) = 2f_m(1+\beta)$.

3.6.4 Power in FM

For sinusoidal input signal, the FM formula is $y(t) = A_c \cos(\omega_c t + \beta \sin(\omega_m t))$

Mean square power of each sideband of an FM signal: $P_n = \dfrac{A_c^2}{2R} J_n^2(\beta)$

Mean square power of an **unmodulated** FM carrier is: $P_c = \dfrac{A_c^2}{2R} = P_{Total}$

Mean square power of total FM signal power is that delivered to the load:

$$P_L = \sum_n P_n = P_c \left\{ J_0^2 + 2(J_1^2 + J_2^2 + J_3^2 + J_4^2 + \cdots) \right\}$$

when $n \to \infty$ (i.e. All sidebands are included), $P_L \approx P_c \equiv P_{Total}$

3.6.5 Generation of WBFM

Direct Method

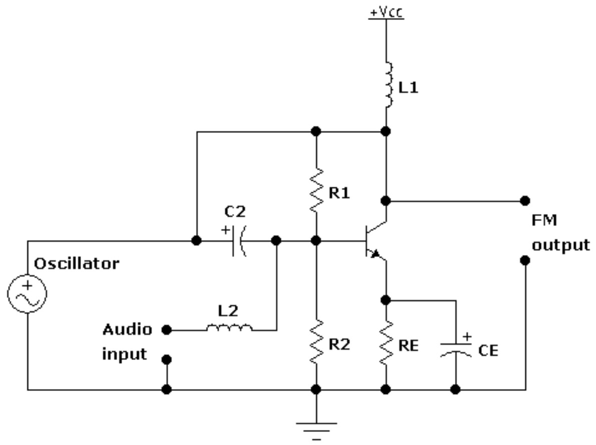
Using the Voltage–Controlled Oscillator (VCO), at which the output signal frequency varies linearly with the input control voltage. The principle of such circuit is charging L and/or C in a tuned oscillator as:

$$f_o = \frac{1}{2\pi\sqrt{LC}}$$

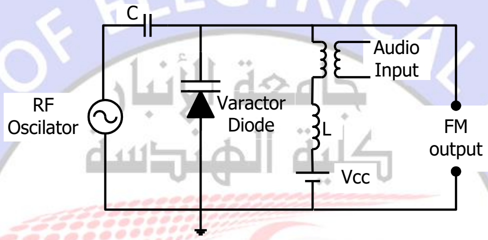
The following are examples of a VCO:

(1) *The Transistor Reactance Modulator*: Here, the h_{fe} of the transistor is caused to vary by changing the operating point of the transistor according to the varying in audio signal. The equivalent C is changed as:

$$C_{eq} = \frac{h_{fe} R_2 C_2}{h_{ie} + R_2}$$

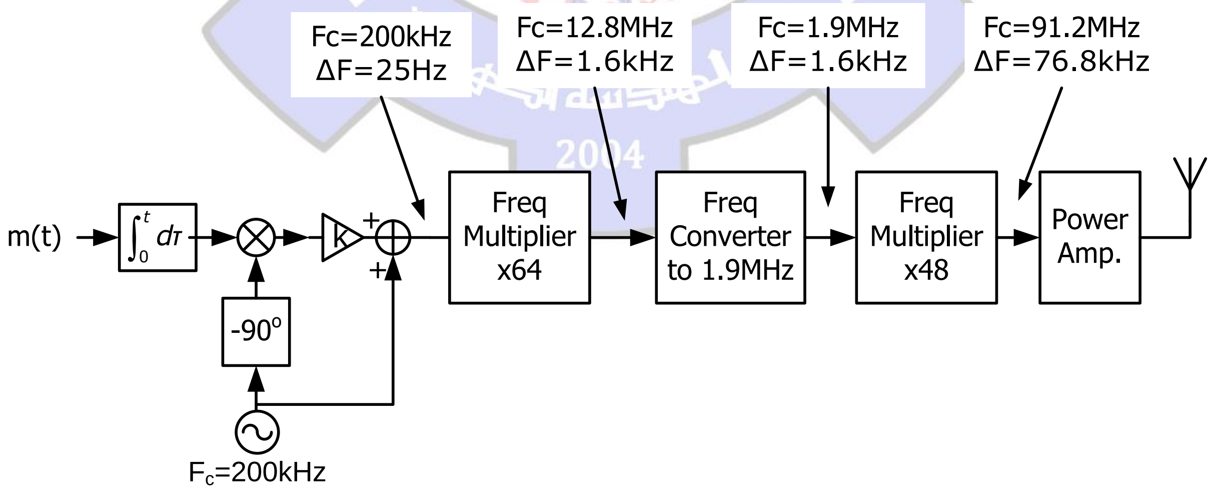


(2) *The varactor diode modulator*: The varactor diode capacitance varies with its bias (V_{cc} and the audio input signal).



Indirect Method

Also called Armstrong's method. First, it generates NBFM, then it changes the signal to WBFM via frequency multipliers and converters.



3.6.6 DEMODULATION OF FM SIGNALS

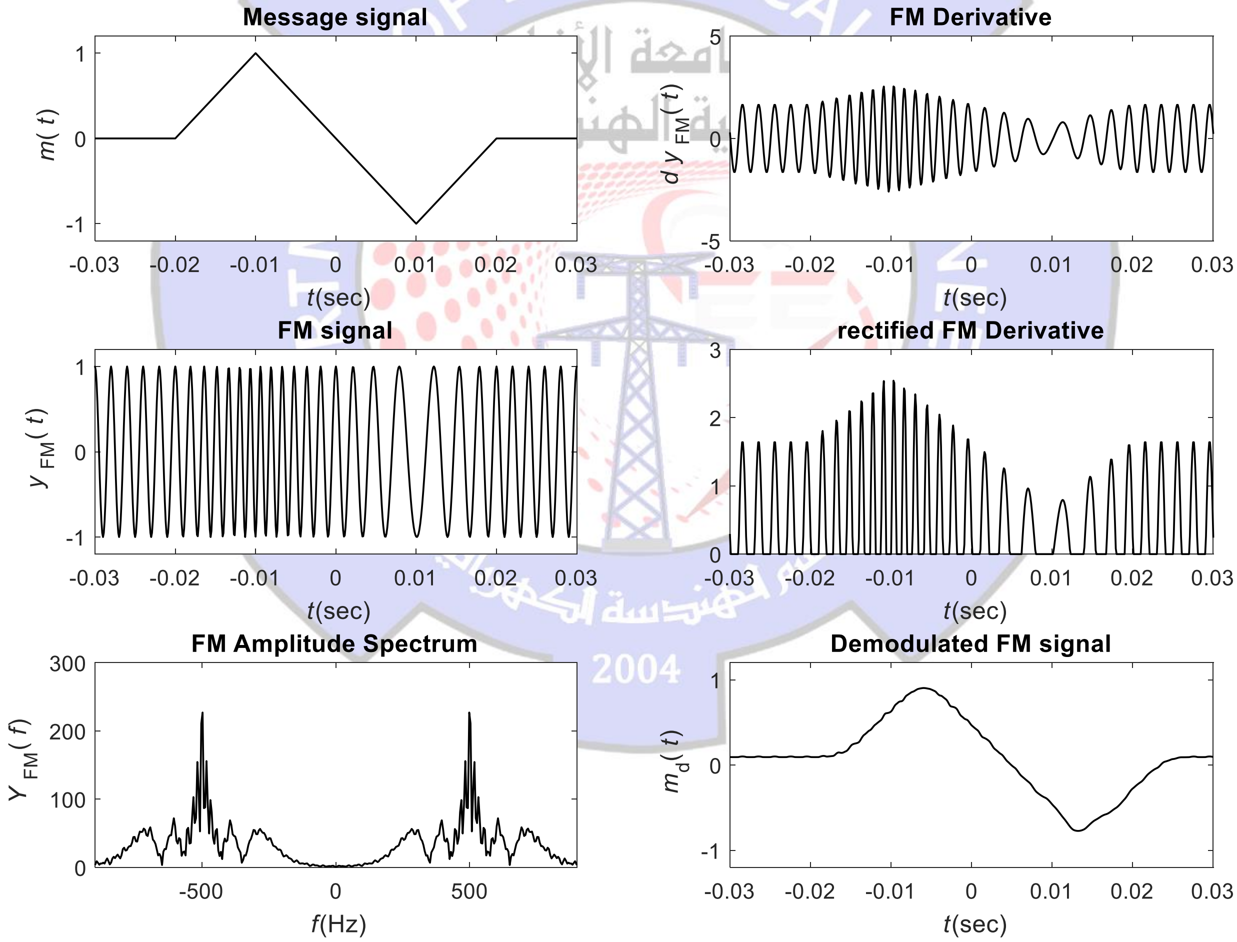
(1) *Differentiator*: the equation of the FM (below) can be demodulated using a differentiating circuit.

$$y_{FM}(t) = A_c \cos\left(\omega_c t + k_F \int_0^t m(\varepsilon)d\varepsilon\right)$$

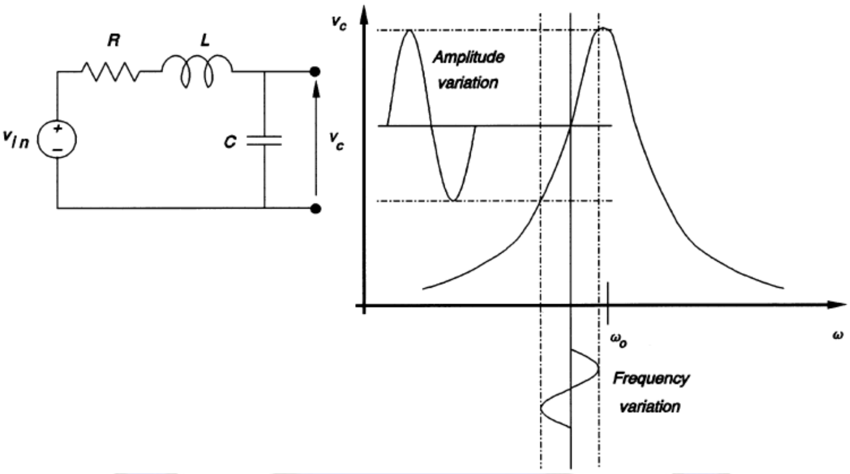
$$\frac{d}{dt}\big(y_{FM}(t)\big) = \frac{d}{dt}\left\{A_c \cos\left(\omega_c t + k_F \int_0^t m(\varepsilon)d\varepsilon\right)\right\}$$

$$= A_c[\omega_c t + k_F m(t)] \sin\left(\omega_c t + k_F \int_0^t m(\varepsilon)d\varepsilon - \pi\right)$$

The following example illustrates the decoding procedure.

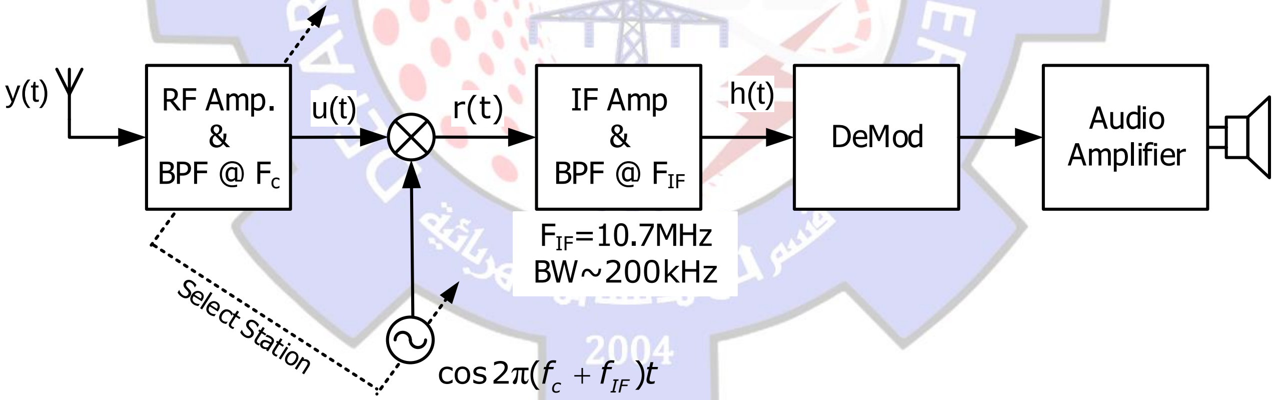


(2) *Discriminator*: The purpose of the discriminator is to convert the variation of frequency to a variation of amplitude (inverse of VCO). The simplest circuit of Discriminator is a tuned RLC circuit that has a rapid change of amplitude with frequency on both sides of the resonance frequency, especially when the Q factor of the circuit is high.



(3) *Zero Crossing Detectors*: because the information is contained in the zero crossings of the FM waveform, it is possible to clip (limit) the amplitude of the FM waveform, which results in a square wave. Counting the zero crossings in each time interval is an indication of the amplitude of the Baseband signal.

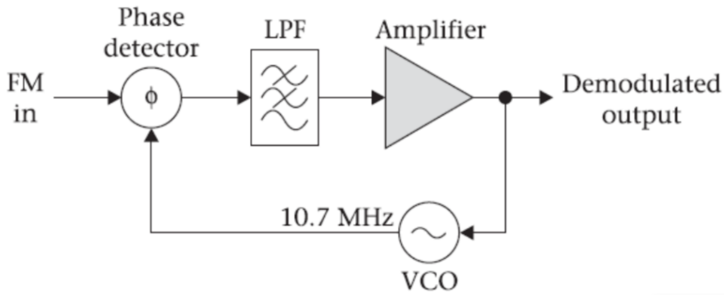
(4) *Super–heterodyne*: is the typical receiver of the commercial FM broadcast stations. This system is like the AM Super–heterodyne receiver but with $f_{IF} = 10.7$MHz.



The frequency band of FM radio broadcasting is 88MHz to 108MHz, with $\beta \geq 5, \Delta_f \leq$ 75kHz. While $\Delta_f \leq 25$kHz for the sound portion of TV broadcasting. Practically, $\Delta_f =$ 15kHz for both FM radio and TV.

Each commercial FM radio broadcast station is allocated a 150kHz channel plus 50kHz guard band. The bandwidth given in FM radio stations is very large compared to AM radio stations; therefore, the sound resolution is significantly enhanced.

(5) *PLL*: A typical modern FM receiver uses a PLL subsystem as the detector. The PLL is insensitive to amplitude variations and can perform the Frequency-to-Voltage conversion; it can therefore be used as an FM detector.



The PLL detector works in the following manner: The free VCO runs at the intermediate frequency 10.7MHz. The incoming signal, if unmodulated, locks up with the VCO signal, causing there to be no signal output from the LPF stage. Now let us assume that the incoming signal has been modulated by a single audio tone. The phase detector will output an error voltage to the VCO to drive the VCO into lock-up with the incoming signal. Because the incoming signal frequency is deviating both above and below the 10.7 MHz rest frequency at a certain number of cycles per second, the VCO will do the same, following the input-signal frequency variations. The error voltage from the LPF, which drives the VCO, will be identical to the original modulating signal, and hence is taken as the demodulated output.

3.6.7 SUMMARY

- Generally, AM techniques are simpler than those for FM in terms of the required electronic circuits and the system structure (Tx and Rx). Hence AM systems are cheaper. However, one significant drawback of AM systems is that they tend to be rather sensitive to impulsive interference which can be caused by say: lightening or car ignition noise, since the information is contained in the instantaneous amplitude of the signal.
- Unlike AM, FM is a nonlinear modulation process. Accordingly, spectral analysis of FM is more difficult than for AM.
- For a certain f_m, the bandwidth of FM is controlled by β.

Part 4 NOISE

4.1 INTRODUCTION

Noise in communication systems is caused by unwanted signals. Basically, the sources of these signals are:

(1) External Noise:

- Atmospheric Noise: It is caused by naturally occurring disturbances in the earth's atmosphere such as: lightening discharges, thunderstorms and other natural electric disturbances. At 30MHz and above atmospheric noise is less severe.
- Industrial Noise: this includes sources like ignitions of cars and aircrafts, electric motors, switching equipment, leakage from high voltage lines, etc.
- Extra-terrestrial Noise:
  - Solar Noise: it originates from the sun which radiates a broad spectrum of frequencies, including those which are used for broadcasting.
  - Cosmic Noise: Distant stars also radiate noise in much the same way as the sun. Noise also comes from distant galaxies in much the same way as they come from the Milky Way.

(2) Internal Noise: This type is generated by any of the active or passive devices found in the receiver. It is proportional to the bandwidth over which it is measured.

By careful engineering, the effects of many unwanted signals could be reduced or eliminated, however, other signals could not be removed. One unavoidable cause of electrical noise is the thermal effects of electrons motion in conducting media, wiring, resistors…etc.

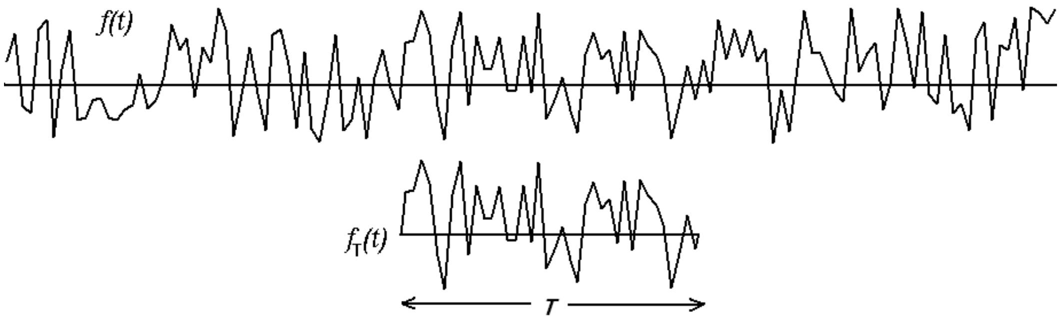
4.2 DEFINITIONS

4.2.1 POWER SPECTRAL DENSITY

Because we deal with random signals, we need to review the PSD. The Power Spectral density (PDF) describes the distribution of power versus the frequency for energy signals.

$$G(f) = \lim_{T\to\infty} \frac{|F_T(f)|^2}{T}$$

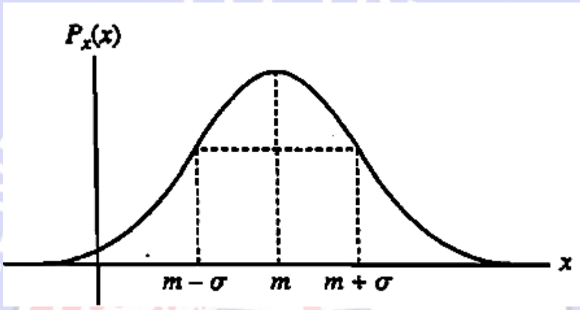
where $F_T(f)$: is the Fourier transform of the *truncated* segment from the random signal.



4.2.2 GAUSSIAN PDF

A Gaussian random variable x is continuous with mean m, variance σ^2, and PDF:

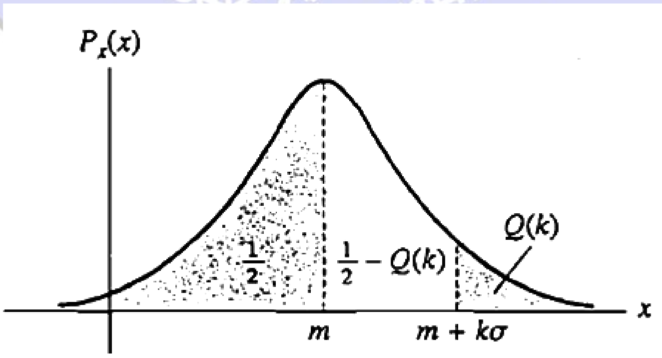
$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2} \qquad -\infty < x < \infty$$



The curve of this function has even symmetry about the peak, i.e. at $x = m$. For some integer k, we find the probability of the event $x > m + k\sigma$ using:

$$Q(k) \cong \frac{1}{\sqrt{2\pi}} \int_k^\infty e^{-\lambda^2/2} d\lambda \qquad \text{at } \lambda = (x-m)/\sigma$$

$Q(k)$ represents the area under the Gaussian tail, as illustrated by Figure below:



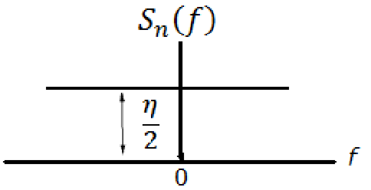
Since this integral cannot be evaluated in closed form, numerical methods are used to generate extensive tables of the normalized integral.

4.3 WHITE NOISE AND FILTERED NOISE

Most noise sources in the electrical systems are Gaussian. They also have a flat spectral density over a wide frequency range. This means the noise is almost the same at all frequencies in equal proportion. Therefore, it is called "white noise" by analogy to white light. White noise is a convenient model (and often an accurate one) in communications. The assumption of a Gaussian process allows us to invoke all the noise properties, however, some applications (beyond our scope) may need a more advanced model for the noise.

We'll write the spectral density of the white noise in general as:

$$S_n(f) = \frac{\eta}{2} \quad \text{for all } f \text{ (two-sided PSD)}$$



Where: η (in Watt per Hertz) is the power spectral density of the white noise. The PSD $S_n(f)$ is η when it is measured for the positive frequency only, and hence it is referred as one-sided PSD. For the two-sided PSD, $S_n(f) = \frac{\eta}{2}$ when it is used for all the frequencies. The value of η depends upon two factors: the type of noise, and the type of spectral density. If the source is a thermal resistor; then the mean square voltage and mean square current densities are

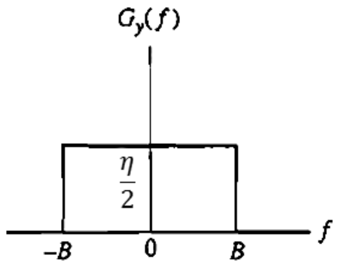
$$\eta_v = 4kTR \quad \text{Volts} \quad , \quad \eta_i = 4kT/R \quad \text{Ampers}$$

We can find noise power P_n as:

$$P_n = \int_{-\infty}^{\infty} \frac{\eta}{2} df \to \infty$$

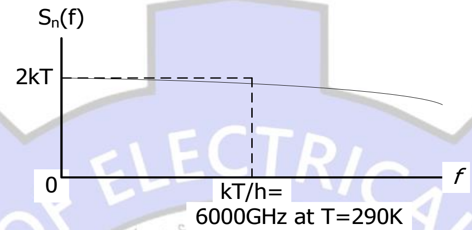
Normally, the communication systems are band limited to B Hz

$$P_n = N = \int_{-B}^{B} \frac{\eta}{2} df = \eta B \text{ Watt}$$



4.4 THERMAL NOISE

Among the mentioned noise sources, we can focus on the thermal noise as it is one of the most commonly considered sources. Thermal noise is an important noise source of white noise. Thermal Noise is produced because of the thermally excited random motion of free electrons in a conducting medium, such as resistor. The path of each electron in motion is randomly oriented due to collisions. The net effect of the motion of the electrons is an electric current in the resistor which is random with a mean of zero. From thermodynamic and quantum mechanical considerations, the PSD of thermal noise is:



$$S_n(f) = \frac{2h|f|}{e^{\frac{h|f|}{kT}} - 1} \approx 2kT \text{ Watt/Hz} \quad \text{for} \quad |f| \ll \frac{kT}{h}$$

T = temperature of conducting medium in Kelvin

k = Boltzmann's constant = 1.38×10^{-23} Joule/Kelvin

h = Planck's constant = 6.625×10^{-34} Joule·sec

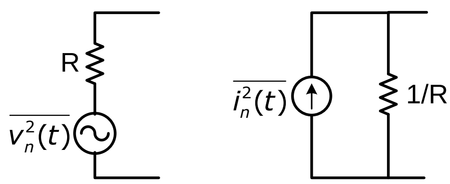
$$P_n = \int_{-B}^{B} S_n(f)df = 4kTB \text{ Watt}$$

So, the mean-square voltage and current generated by a resistor R within the band width B is:

$$\overline{v_n^2(t)} = RP_n = 4kTRB \quad \text{Volt}^2$$

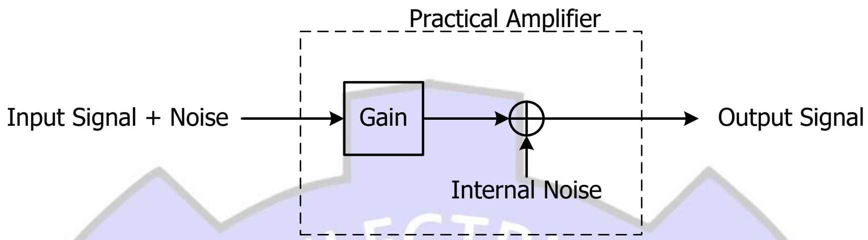
$$\overline{i_n^2(t)} = \frac{P_n}{R} = \frac{4kTB}{R} \quad \text{Ampere}^2$$

For band-limited thermal noise, the models for voltage and current equivalent circuits are shown below, assuming R is noise-free and bandwidth B.



4.5 AMPLIFIER NOISE

Since the received signal is very weak, it is crucial to have an amplifier at an early stage of the reception. Because the received signal has already been corrupted by the noise from the external noise, it is important to make the internal noise within the receiver itself as small as possible. Since we cannot practically implement a *noiseless* amplifier, we may consider *theoretically* an amplifier like:



Let the received signal S_r comprise S_i as the message without noise, and N_i as the additive noise. So, if the input to the amplifier of the receiver is $S_r = S_i + N_i$, the output will be:

$$S_o = S_r G + N_{int} = GS_i + (GN_i + N_{int}) = GS_i + N_o$$

Where: G is the gain of the ideal amplifier, N_{int} is the internally generated noise. It is obvious that the final noise component represents the addition of the internal and the amplified value of the external noise. To discuss the noise within the receiver, we need to use describe the following measures and terms.

4.6 NOISE COMPUTATIONS

4.6.1 SNR

The Signal-to-Noise Ratio (SNR) is an important measure for the performance of a system. It reflects the amount of noise that has been induced in the system in proportion to the desired signal.

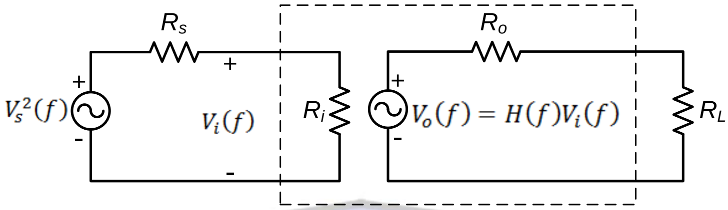
$$\frac{S}{N} = \frac{\text{Signal Power}}{\text{Noise Power}} = \frac{\overline{s^2(t)}}{\overline{n^2(t)}} \quad \text{or} \quad \left(\frac{S}{N}\right)_{dB} = 10\log\left(\frac{S}{N}\right)$$

4.6.2 AVAILABLE POWER AND T_E

The figure below depicts the circuit model of a noiseless amplifier inserted between a source V_s and a load R_L. This model has an input resistance R_i, an output resistance R_o, and a voltage transfer function $H(f)$. The source generates a mean square voltage density $V_s^2(f)$

representing noise or an information signal or both. The available power density from the source is $\eta_s(f) = V_s^2(f)/4R$. The available power density at the output of the amplifier is:

$$\eta_o(f) = \frac{V_o^2(f)}{4R_o} = \frac{|H(f)|^2 V_i^2(f)}{4R_o} = \frac{|H(f)|^2}{4R_o}\left(\frac{R_i}{R_s + R_i}\right)^2 V_s^2(f)$$

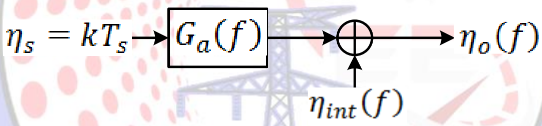


We define $G_a(f)$ as the gain of the amplifier's available power:

$$G_a(f) \approx \frac{\eta_o(f)}{\eta_s(f)} = \frac{V_o^2(f) R_s}{V_s^2(f) R_o} = \left(\frac{|H(f)| R_i}{R_s + R_i}\right)^2 \frac{R_s}{R_o}$$

Assuming that the source generates thermal white noise at the temperature T_s, then $\eta_s = kT_s$. The available noise power at the output of a noiseless amplifier is:

$$\eta_o(f) = G_a(f)\eta_s(f) = G_a(f)kT_s$$



Since the internal noise is independent of the source noise, we write:

$$\eta_o(f) = G_a(f)kT_s + \eta_{int}(f)$$

Integration then yields the total available output noise power:

$$N_o = \int_0^\infty \eta_o(f)df = kT_s \int_0^\infty G_a(f)df + \int_0^\infty \eta_{int}(f)df$$

Most amplifiers in a communication system have a frequency-selective response, with maximum power gain G and noise equivalent bandwidth B_N, so:

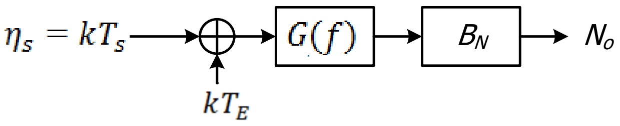
$$GB_N = \int_0^\infty G_a(f)df$$

We define the effective noise temperature of the amplifier to be:

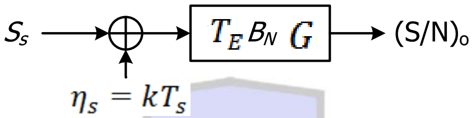
$$T_E = \frac{1}{GkB_N}\int_0^\infty \eta_{int}(f)df$$

Hence, the total output noise power becomes:

$$N_o = Gk(T_s + T_E)B_N$$



Now let the figure below represent a noisy amplifier with signal plus white noise at the input.



So, the available signal power at the output will be $S_o = GS_s$. Thus

$$\left(\frac{S}{N}\right)_o = \frac{GS_s}{N_o} = \frac{S_s}{k(T_s + T_E)B_N}$$

Since $\left(\frac{S}{N}\right)_s = \frac{S_s}{kT_sB_N}$, then:

$$\left(\frac{S}{N}\right)_o = \frac{1}{1 + T_E/T_s}\left(\frac{S}{N}\right)_s$$

we see that in general $\mathrm{SNR_o} < \mathrm{SNR_s}$, however we may get $\mathrm{SNR_o} \approx \mathrm{SNR_s}$ when $T_E \ll T_s$. At carrier frequencies below ~30MHz, the internal noise has little effect, and the amplifier appears to be noiseless. At higher frequencies, T_E becomes significant which often affect the design of the receivers and the repeaters.Some very low-noise amplifiers have $T_E \approx 10\mathrm{K} \rightarrow 30\mathrm{K}$ while it is $T_E \approx 1000\mathrm{K}$ for some systems. Knowing T_E is not necessarily the ambient temperature of the amplifier, sometimes cryogenic cooling is employed to lower the noise temperature. In this class, we set $T_E = 290\mathrm{K}$.

4.6.3 Noise Figure

Another measure of amplifier noise is the noise figure F (also called 'figure-of-merit'). It is expressed as $F = \mathrm{SNR}_I \div \mathrm{SNR}_o$ (normally in dB). Now, for the case where $T_s = T_0 = 290\mathrm{K}$,

$$F = 1 + \frac{T_E}{T_0} \quad \text{or} \quad F_{dB} = 10\log_{10}(F)$$

A very noisy amplifier has $T_E \gg T_0 \rightarrow F \gg 1$

At room temperature, $F = 2$

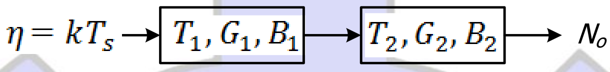
A low-noise amplifier has $T_E < T_0 \rightarrow 1 < F < 2$: it is the case receivers usually should work at.

HW: What is the value of F of a perfect system?

4.6.4 NOISE IN MULTI-STAGE SYSTEMS

Receivers comprise consecutive different stages to process the incoming signal. Since each stage generates internal noise differently, they must be carefully designed to perform the optimum reception.

To develop expressions for the overall performance of the system in terms of the parameters of the individual stages, let's consider the following cascade of two noisy two-ports sub-systems. In this figure, the subscripts identify: the effective noise temperature, the maximum power gain, and the noise bandwidth per stage.



The overall power gain then equals the product, i.e. $G = G_1 G_2$

The total output noise power consists of three terms:

(1) Source noise amplified by both stages;
(2) Internal noise from the first stage, amplified by the second stage;
(3) Internal noise from the second stage.

Assuming $B_2 \leq B_1$ and $B_N \approx B_2$, thus:

$$N_o = GkT_sB_N + G_2(G_1kT_1B_N) + G_2kT_2B_N = Gk\left(T_s + T_1 + \frac{T_2}{G_1}\right)B_N$$

The overall effective noise temperature and noise figure are:

$$T_E = T_1 + \frac{T_2}{G_1} \quad \text{and} \quad F = 1 + \frac{T_1}{T_0} + \frac{T_2}{G_1T_0} = F_1 + \frac{F_2 - 1}{G_1}$$

The foregoing analysis readily generalizes to:

$$T_E = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1G_2} + \cdots \quad \text{and} \quad F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1G_2} + \cdots$$

Both expressions bring out the fact that:

The first stage plays a critical role and it must be given careful attention in system design.

See Str. Example 4.7.5

4.6.5 TIME REPRESENTATION OF BANDPASS NOISE

We can approximate the noise with a phasor representation as: the portion of the noise which is the in-phase component $n_c(t)$, and the quadrature component $n_s(t)$. So:

$n(t) = n_c(t)\cos(\omega_0 t) - n_s(t)\sin(\omega_0 t)$

Where f_0 = center frequency. Note that: $\overline{n^2(t)} = \overline{n_c^2(t)} = \overline{n_s^2(t)}$

4.7 Noise in AM Systems

4.7.1 Synchronous DSB-SC

We assume synchronous demodulation. So, if the input is $y(t) = m(t)\cos(\omega_c t)$, then:

$S_I = \overline{y^2(t)} = \frac{1}{2}\overline{m^2(t)}$, and as the output is $\frac{1}{2}m(t)$ then: $S_O = \overline{\left[\frac{1}{2}m(t)\right]^2} = \frac{1}{4}\overline{m^2(t)} = \frac{1}{2}S_I$.

If the input noise is $n_I(t)$ then $N_I = \overline{n_i^2(t)}$, The multiplier output

$r(t) = n_I(t)\cos(\omega_c t) = [n_c(t)\cos(\omega_c t) - n_s(t)\sin(\omega_c t)]\cos(\omega_c t)$

$$= \frac{1}{2}n_c(t) + \frac{1}{2}n_c(t)\cos(2\omega_c t) - \frac{1}{2}n_s(t)\sin(2\omega_c t)$$

The output from the LPF will be: $n_o(t) = \frac{1}{2}n_c(t)$, so:

$$N_O = \overline{n_O^2(t)} = \frac{1}{4}\overline{n_c^2(t)} = \frac{1}{4}\overline{n_i^2(t)} = \frac{1}{4}N_I$$

As a result:

$$\therefore \frac{S_O}{N_O} = 2\frac{S_I}{N_I}$$

This means that the detector improves the SNR in DSB-SC system by factor of two. This improvement is achieved because the coherent (synchronous) detector rejects $n_s(t)$ and halves $n_c(t)$.

4.7.2 Synchronous DSB-LC

The use of synchronous detector for DSB-LC provides the same advantages mentioned above. The same computations can be used here as well by using $m(t)$ as $\left(A + m(t)\right)$ instead. So:

$S_I = \frac{1}{2}A^2 + \frac{1}{2}\overline{m^2(t)}$ assuming $\overline{m^2(t)} = 0$.

Hence $$\frac{S_O}{N_O} = \frac{2\overline{m^2(t)}}{A^2 + \overline{m^2(t)}} \cdot \frac{S_I}{N_I}$$

Note: for $m(t) = \cos(\omega_m t)$, then:

$$\frac{S_O}{N_O} = 2\rho \cdot \frac{S_I}{N_I}$$

4.7.3 ENVELOPE DETECTOR DSB-LC

The input signal to the envelope detector is

$\Phi(t) = y(t) + n_I(t) = [A + m(t)]\cos(\omega_c t) + n_c(t)\cos(\omega_c t) - n_s(t)\sin(\omega_c t)$

For high input SNR, it is just like synchronous. i.e.:

$$\frac{S_O}{N_O} = \frac{2\overline{m^2(t)}}{A^2 + \overline{m^2(t)}} \cdot \frac{S_I}{N_I}$$

For input SNR< 10dB

$$\frac{S_O}{N_O} = \frac{S_I}{\eta f_m}$$

In fact, for input SNR< 1 (0dB), the noise takes over the control of the envelop.

4.7.4 SSB-SC

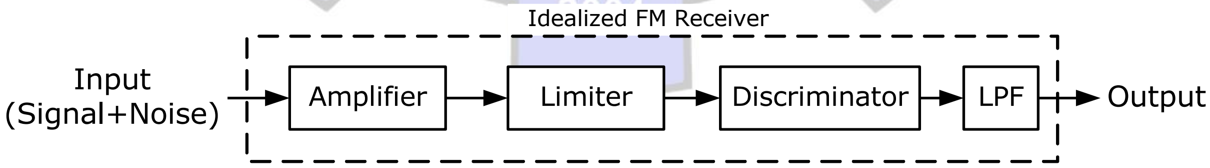
Since the SSB-SC signal is: $y(t) = m(t)\cos(\omega_c t) \pm \hat{m}(t)\sin(\omega_c t)$, we have:

$S_I = \overline{y^2(t)} = \frac{1}{2}\overline{m^2(t)} \pm \frac{1}{2}\overline{\hat{m}^2(t)}$

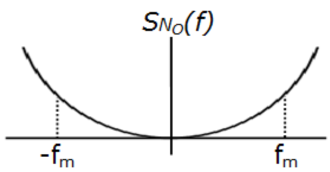
But $\hat{m}(t)$ is only 90° shifting of $m(t)$, so $S_I = \overline{m^2(t)}$. And: $S_O = \overline{\left[\frac{1}{2}m(t)\right]^2} = \frac{1}{4}\overline{m^2(t)} = \frac{1}{4}S_I$. This results in:

$$\therefore \frac{S_O}{N_O} = \frac{S_I}{N_I}$$

4.8 NOISE IN FM SYSTEMS



In FM signals, $s_I(t) = A\cos\theta$, where $\theta = \omega_c t + k\int_0^t m(\tau)d\tau$, and $s_o(t) = \frac{d\theta}{dt} - \omega_c = km(t) \Leftrightarrow$ $S_o = \overline{s_o^2(t)} = k^2\overline{m^2(t)}$. The PSD of output noise of FM demodulator is $S_{N_o} = \frac{\eta}{A^2}f^2$.



Since it has a parabolic spectrum, therefore, the effect of noise in FM for higher frequency components is much higher than the effect of noise on lower frequency.

Assuming LPF cutoff frequency (or B) $= f_m$, the output noise power from the LPF is:

$$N_o = \int_{-f_m}^{f_m} S_{N_o}(f)df = \int_{-f_m}^{f_m} \frac{\eta}{A^2} f^2 df = \frac{\eta f^3}{A^2}$$

The output signal power $S_o = k^2 P_M$, where P_M is the average power of $m(t)$. So:

$$\frac{S_o}{N_O} = \frac{3k^2A^2P_M}{2\eta f_m^3}$$

But $\beta = \dfrac{k \max|m(t)|}{f_m}$ and $S_I = \dfrac{A^2}{2}$ and $N_I = \eta f_m$.

This yields:

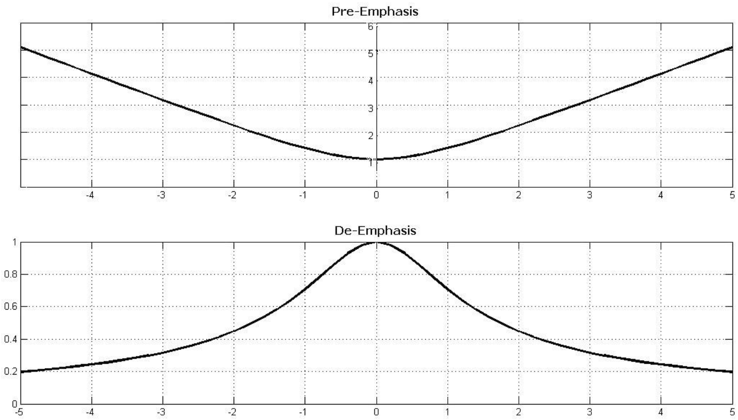
$$\frac{S_o}{N_o} = \frac{3\beta^2 P_M}{[\max|m(t)|]^2}$$

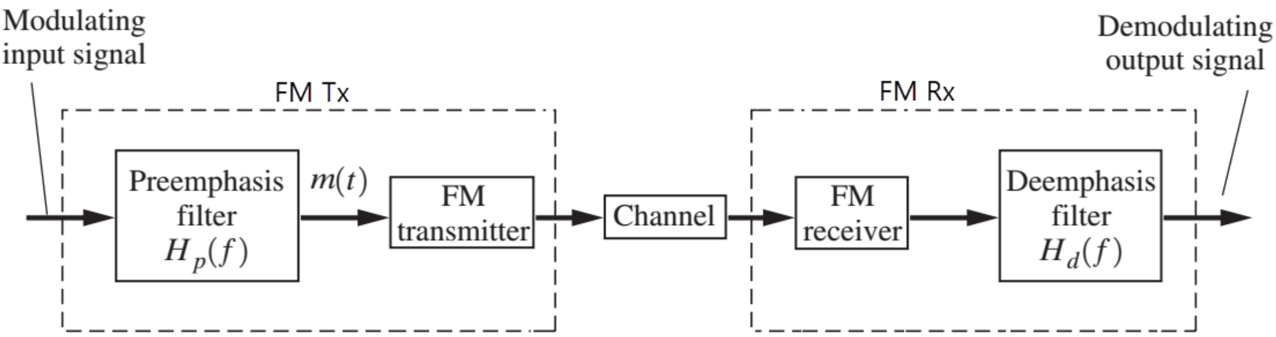
and in case of $m(t) = \cos(\omega_m t)$ we get:

$$\frac{S_O}{N_O} = \frac{3\beta^2}{2}\frac{S_I}{N_I}$$

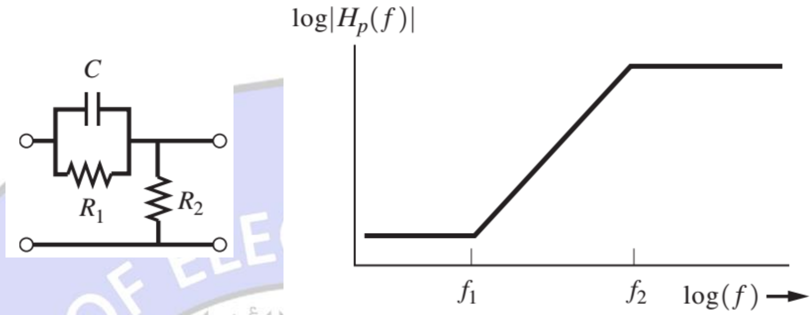
SNR Improvement in FM using Pre-emphasis/De-emphasis

The audio signals have most of the energy at the lower frequency (300 → 3300)Hz. And as in the output of the FM demodulator the noise PSD rises parabolically with the frequency, that means: the noise PSD is largest in the frequency range where the signal PSD is smallest. To remedy this situation, we emphasize the high-frequency components in the input signal at the transmitter. At the output of the FM demodulator in the receiver the inverse operation is performed. The signal spectrum is restored to its original shape but the noise which was added after the pre-emphasis is now reduced.

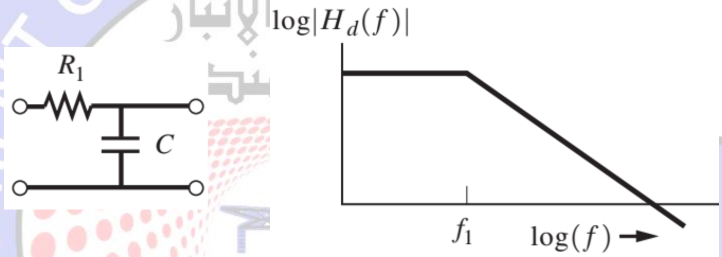




Preemphasis Filter:



Deemphasis Filter:

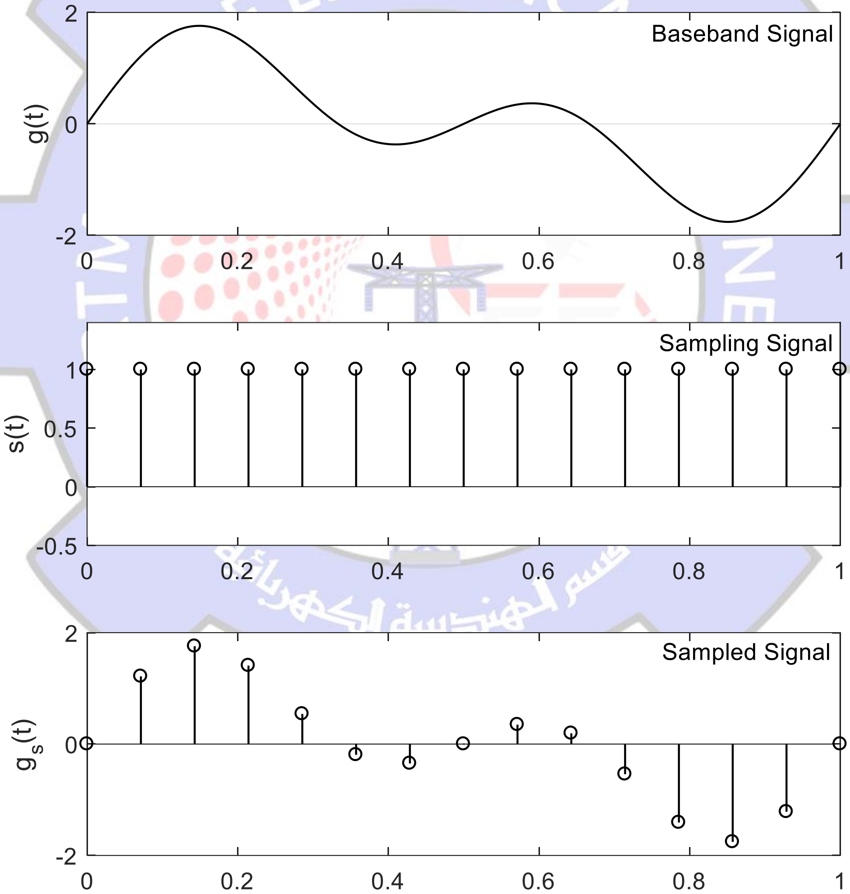


Part 5 SAMPLING & PULSE MODULATION

5.1 INTRODUCTION

Experimental data and mathematical functions are frequently displayed as continuous curves, even though a finite number of discrete points were used to construct the graph. If these points or samples have sufficiently close spacing, we obtain a smooth curve drawn through them. Therefore, it can be said that: a continuous curve is adequately described by its sample points alone.

Sampling, therefore, makes it is possible to transmit messages in the form of pulse modulation rather than as continuous signals. Usually, the pulses are quite short compared to the time between them.



Pulse modulation offers two main advantages over continuous wave modulation:

(1) The transmitted power can be concentrated into short bursts instead of being generated continuously.

(2) The time interval between pulses can be filled with sample values from other signals. (TDM).

However, the main disadvantage is the requirement for large transmission bandwidth compared to the original message bandwidth.

5.2 SAMPLING THEOREM

5.2.1 IDEAL SAMPLING

Ideal sampling♣ is done by multiplying the baseband signal $g(t)$ by the impulse signal $s(t)$.

$$g_s(t) = g(t) \cdot s(t)$$

$$s(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_s)$$

The Fourier transform of $s(t)$ is:

$$S(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} \delta(f - kf_s) \quad \text{where } f_s = \frac{1}{T_s}$$

Let's now evaluate $G_s(f)$, the Fourier transform of the output signal $g_s(t)$. So

$$G_s(f) = \mathcal{F}\{g_s(t)\} = \mathcal{F}\{g(t) \cdot s(t)\}$$

Multiplication in the time domain is convolution in the frequency domain:

$$G_s(f) = G(f) \star S(f) = G(f) \star \frac{1}{T_s} \sum_{k=-\infty}^{\infty} \delta(f - kf_s)$$

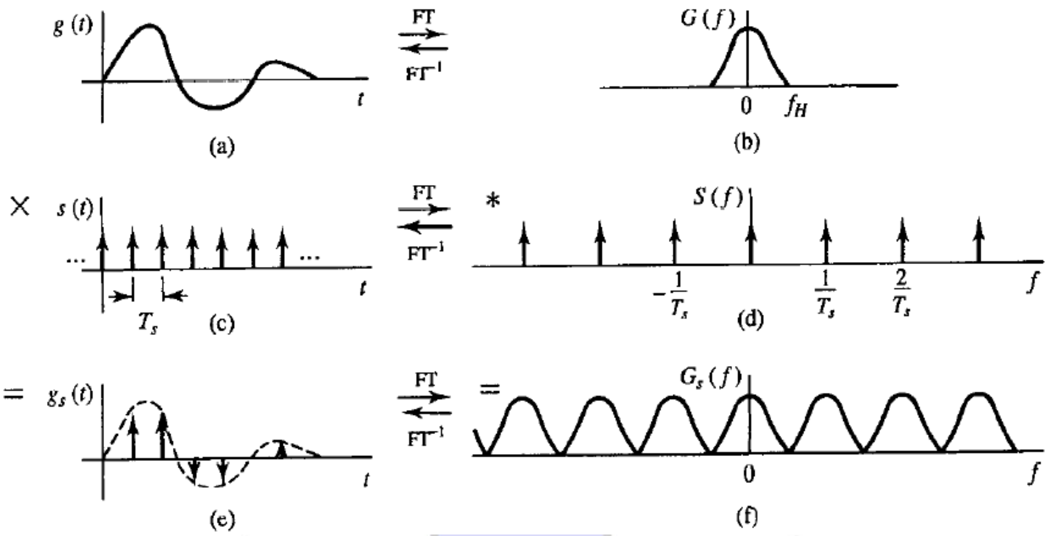
Where $\star$ denotes convolution. Next, we apply simple properties of convolution to move $G(f)$ inside the sum; that is, $G_s(f)$ becomes:

$$G_s(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} G(f) \star \delta(f - kf_s)$$

by applying the shifting property of the delta function:

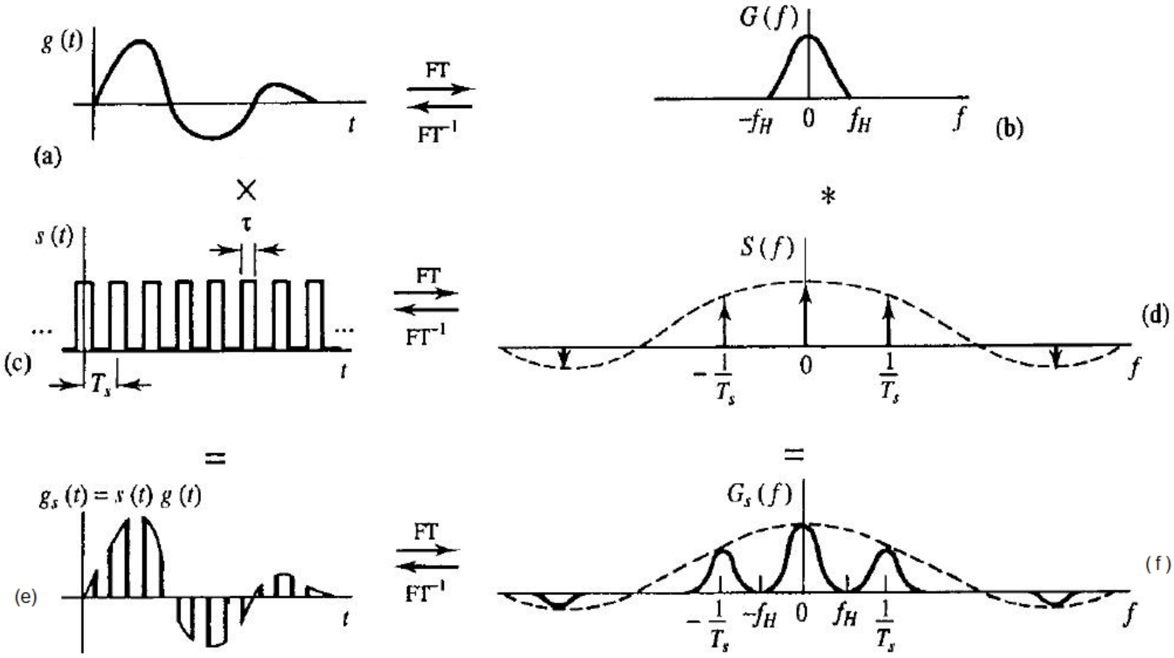
$$G_s(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} G(f - kf_s) = \frac{1}{T_s} \left\{ \begin{matrix} G(f) + G(f + f_s) + G(f + 2f_s) + G(f + 3f_s) + \cdots \\ +G(f - f_s) + G(f - 2f_s) + G(f - 3f_s) + \cdots \end{matrix} \right\}$$

♣ It is impossible to physically carry out ideal sampling, but it is used to understand the sampling theorem.



5.2.2 NATURAL SAMPLING

The practical method of sampling is a naturally sampled signal which is produced by multiplying the baseband information signal $g(t)$ by the periodic pulse train $s(t)$.



Here, we also have: $G_s(f) = \mathcal{F}\{g_s(t)\} = \mathcal{F}\{g(t) \cdot s(t)\}$

And because $s(t)$ is a periodic signal:

$$s(t) = \sum_{k=-\infty}^{\infty} C_k e^{j2\pi k f_s t}$$

Where C_k is the Fourier series coefficient:

$$C_k = \frac{1}{T_s}\text{sinc}\left(\frac{nT}{T_s}\right)$$

So:

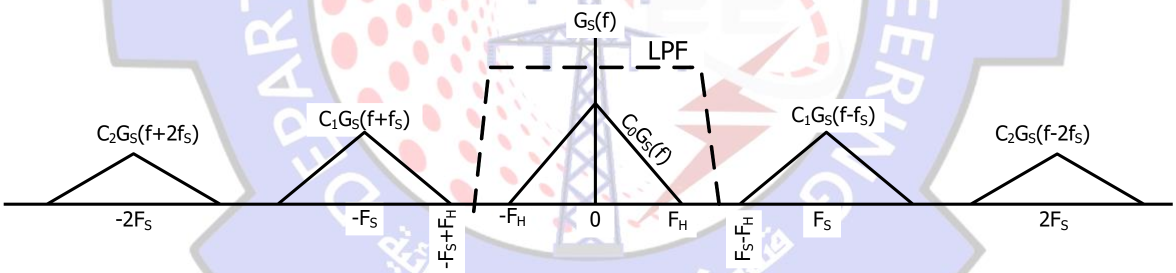
$$G_s(f) = \mathcal{F}\left\{g(t)\sum_{k=-\infty}^{\infty} C_k e^{j2\pi k f_s t}\right\} \equiv \sum_{k=-\infty}^{\infty} C_k \mathcal{F}\{g(t)e^{j2\pi k f_s t}\}$$

$$\Rightarrow G_s(f) = \sum_{k=-\infty}^{\infty} C_k G(f - kf_s)$$

This means: $G_s(f)$ consists of many copies of $G(f)$ added together, where the k^{th} copy is shifted by kf_s and multiplied by C_k.

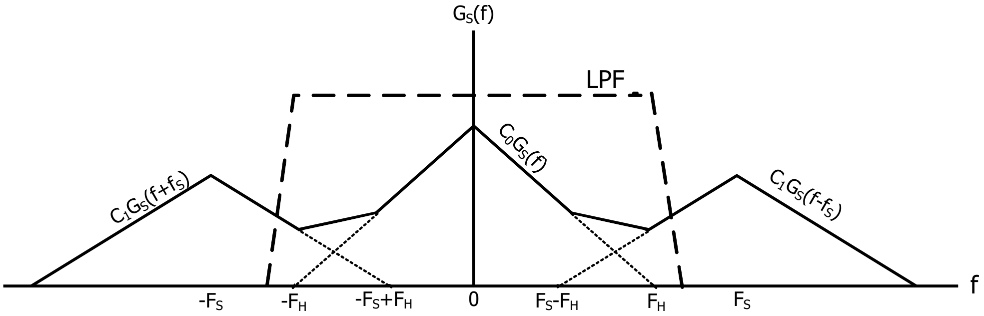
5.2.3 RECONSTRUCTION

To perform successful sampling, we keep $f_s - f_H \geq f_H$, i.e. $f_s \geq 2f_H$. This ensures the signal $G(f)$ always contained perfectly in $G_s(f)$. As a result, $G(f)$ can be recovered exactly by simply passing $G_s(f)$ through a LPF that passes only $C_0G(f)$. Finally, we introduce a gain of $1/C_0$ in the LPF to restore the original signal.



5.2.4 ALIASING

When $f_s < 2f_H$, the copies overlap, and hence, the perfect reconstruction becomes impossible. If this Nyquist criterion is not considered, the folded back portion overlaps the original spectrum. This results a new shape of the reconstructed spectrum filtered by LPF. And this, undoubtedly, gives a signal different from $g(t)$. This problem is called "aliasing".

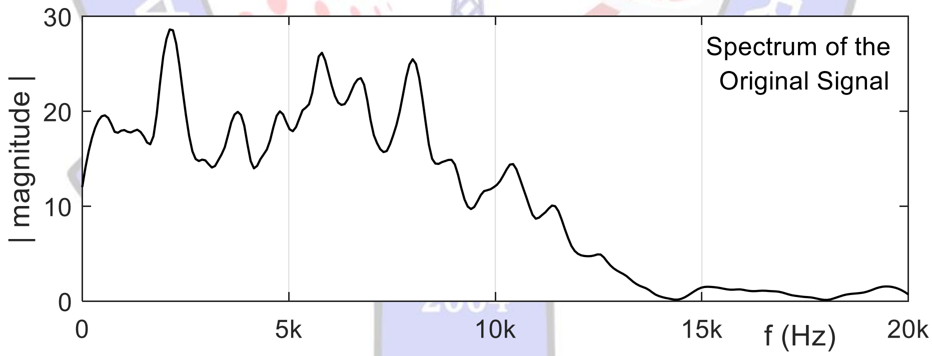


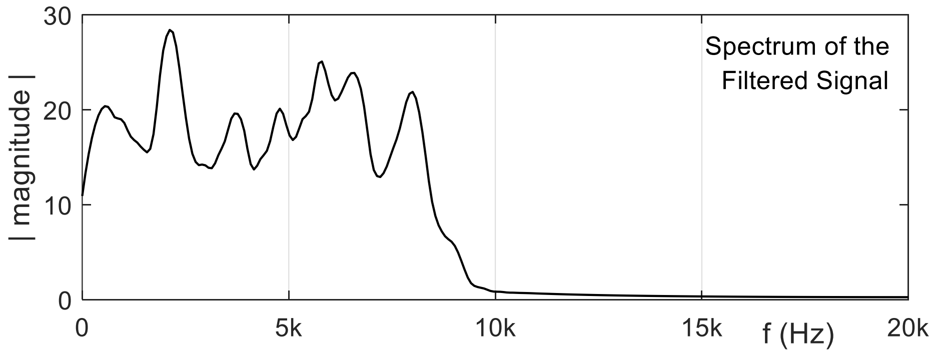
Two conditions are necessary to avoid the aliasing:

(1) The input signal must be limited according the Nyquist condition, i.e. $f_H \leq \frac{f_s}{2}$.

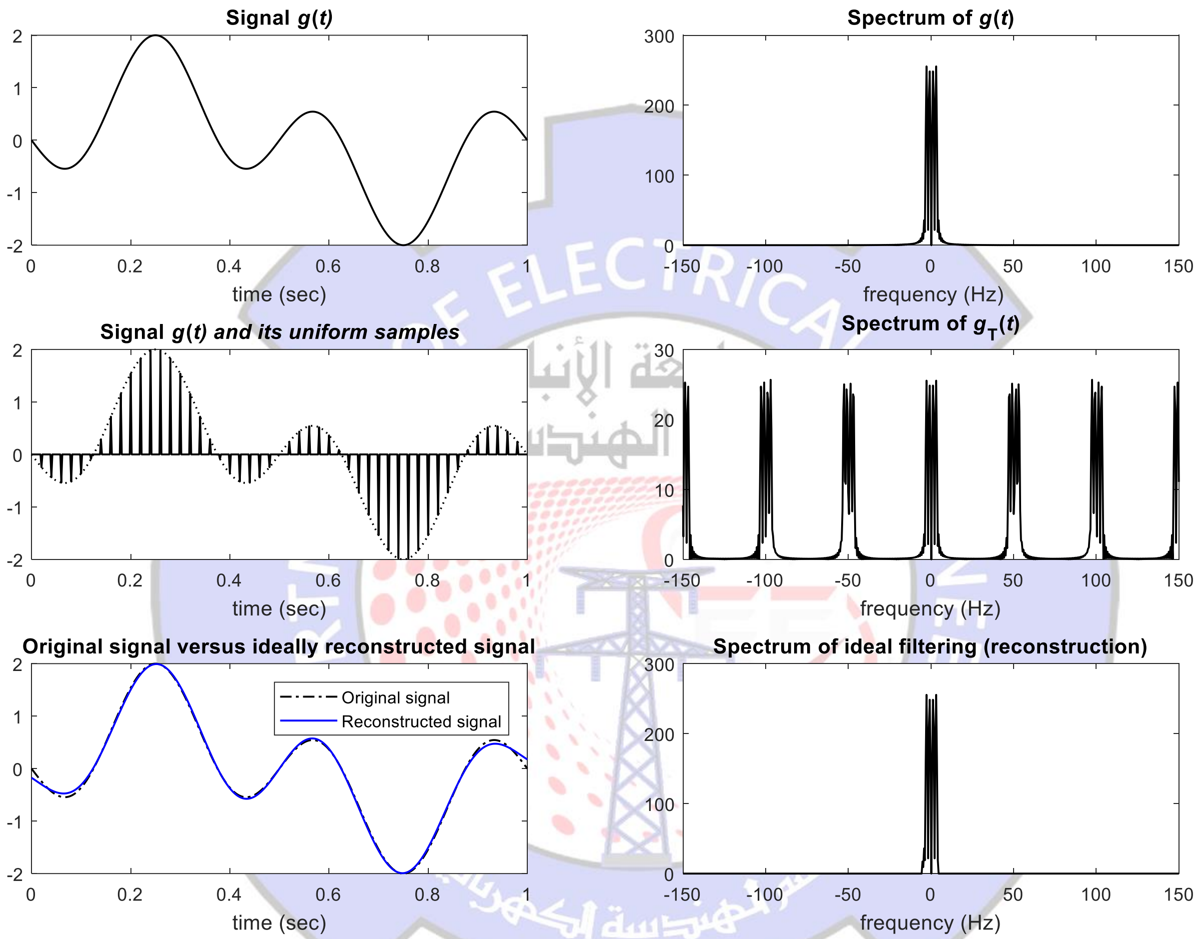
(2) The sampling frequency must sufficiently greater than the maximum frequency component of the signal, i.e. $f_s \geq 2f_H$. This condition is called "Nyquist Rate".

The illustration below shows a baseband signal whose spectrum extends up to 20kHz. If we want to perform Nyquist sampling at the rate 20kHz, we must use an anti-aliasing filter whose cutoff frequency equals $\frac{f_s}{2} \leq 10$kHz.



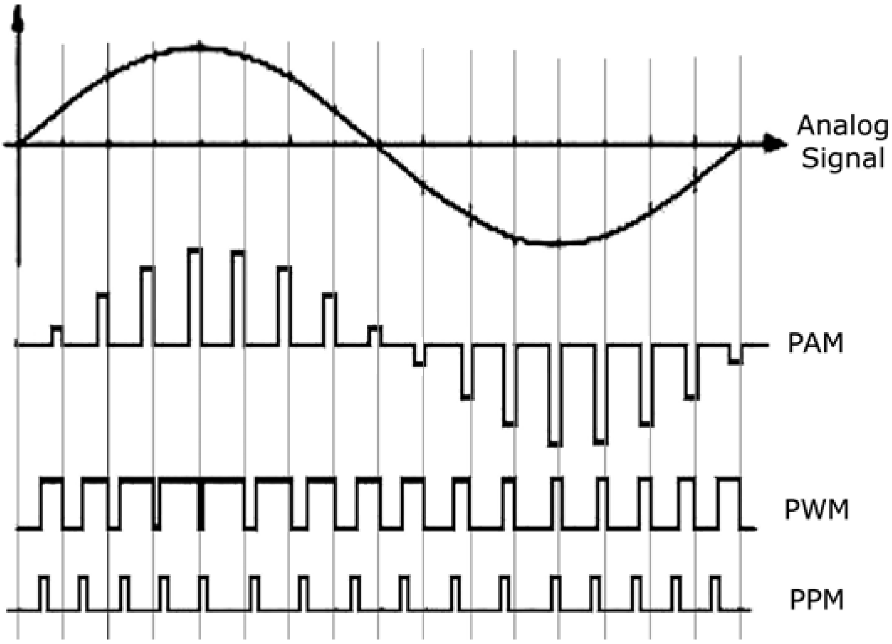


Another example from practical voice transmission systems. Although the average voice spectrum extends well up to 20kHz, most of the energy is concentrated in the range 100Hz→1kHz. For intelligibility, the bandwidth 3kHz is sufficient. A voice wave is pre-filtered by a 3.3kHz LPF and then sampled at 8kHz. These are the standard values for telephone sampling.



5.3 Pulse Modulation

Pulse modulation describes the process whereby the amplitude, the width or the position of individual pulses in a periodic pulse train are varied (i.e. modulated) in sympathy with the amplitude of a baseband information signal $m(t)$.



The pulses of PAM have: amplitude=*variable*, width=*fixed*, position=*fixed*.

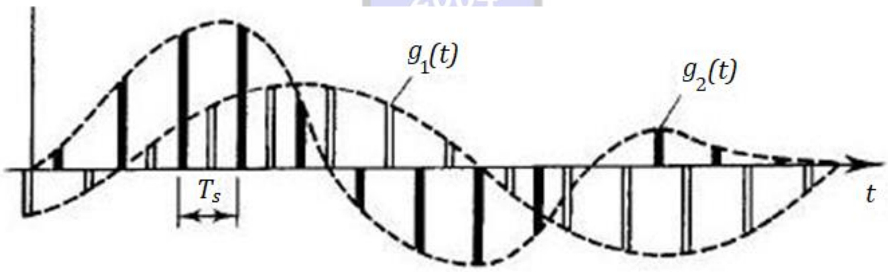
The pulses of PWM have: amplitude= *fixed*, width= *variable*, position=*fixed*.

The pulses of PPM have: amplitude= *fixed*, width=*fixed*, position= *variable*.

What modulation scheme could be used in a simple TV remote controller?

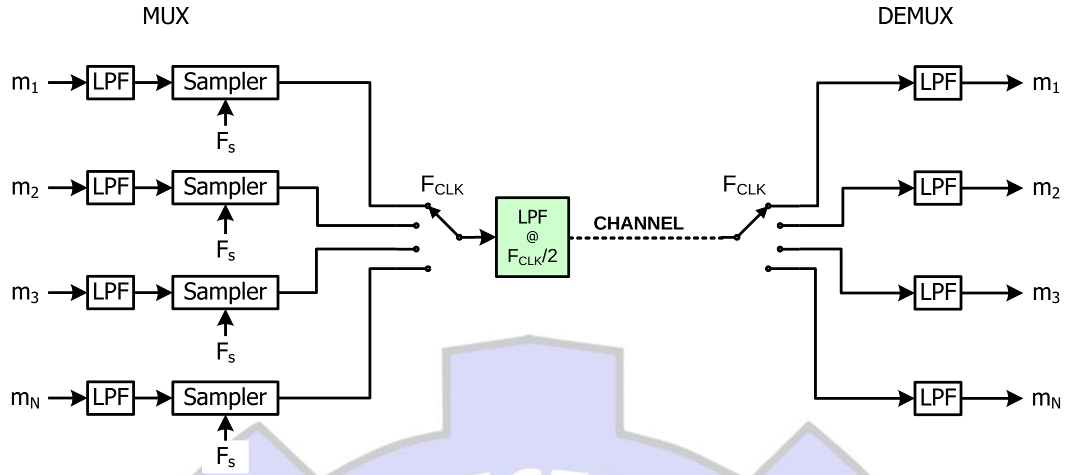
5.4 TIME DIVISION MULTIPLEXING (TDM)

The use of short pulse width in PAM, leaves sufficient spaces between samples for insertion of pulses other sampled signals. The method of combining several sampled signals in a defined sequence is called TDM.

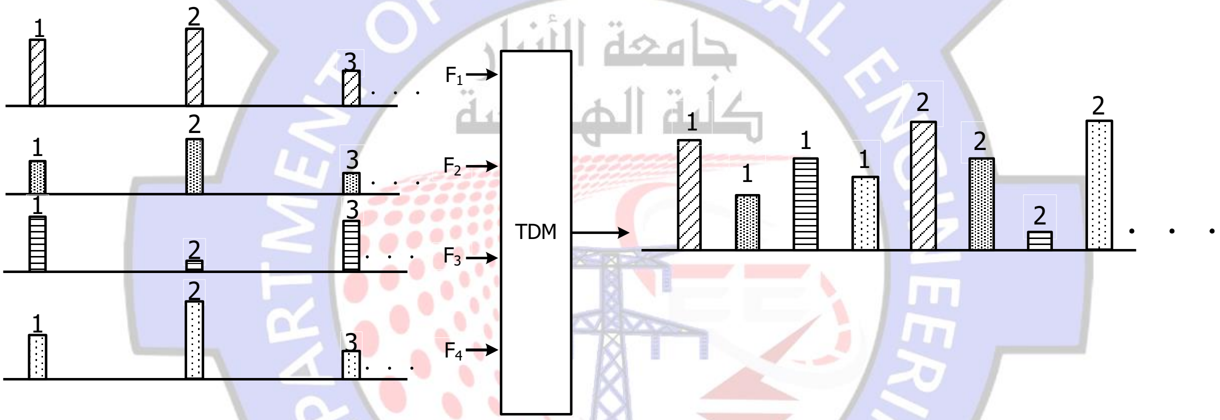


The block diagram below demonstrates the PAM-TDM principle. In this system, 4 different signals are sampled at the same rate (f_s samples per seconds). The clock frequency f_{CLK} must be fast enough to send the all the samples without missing a sample from a signal (in this case

$f_{CLK} = 4f_s$. The input signals are pre-filtered to prevent the aliasing. Complete synchronization between MUX and DEMUX is critical for correct reception.

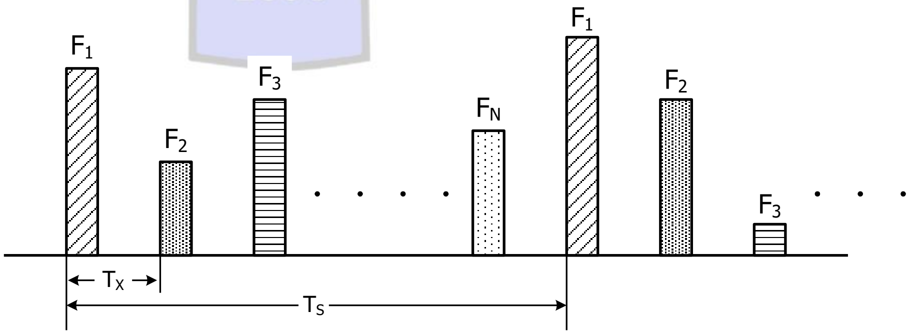


The depiction below illustrates the scenario of multiplexing.



In general, we need to understand the exact design for a TDM system. In the diagram below, let T_x be the time spacing between adjacent samples in a TDM signal. If all input signals are sampled equally at $f_S = 1/T_s$ then $T_x = T_s/N$, where N = number of multiplexed sampled signals. The output TDM stream frequency= clock frequency is

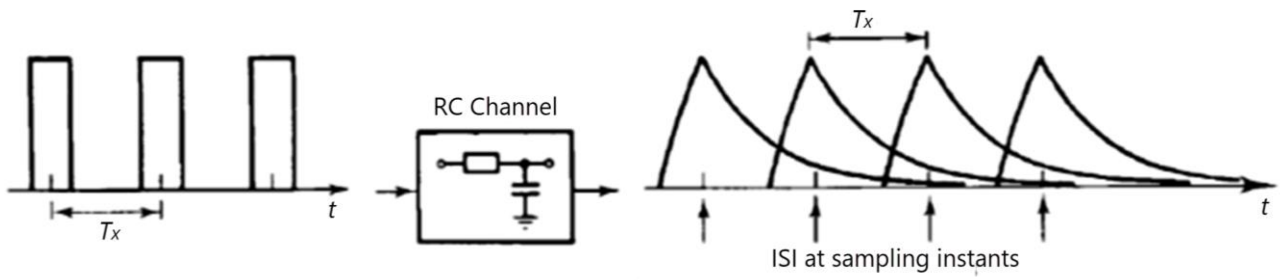
$$f_{CLK} = \frac{1}{T_x} = \frac{N}{T_s} = Nf_s$$



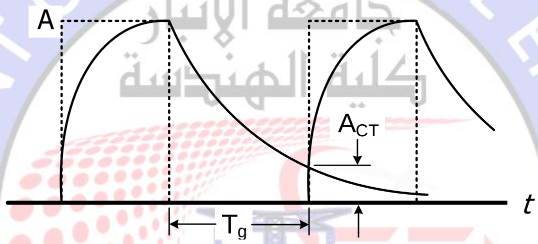
What is the TDMA?

Crosstalk and Guard Times

In addition to ensuring the synchronization, TDM system must avoid inter-channel crosstalk. A TDM signal suffers from crosstalk if the transmission channel results in pulses whose tails overlap into the next time slots of the frame.



If we assume the transmission channel acts like a 1st order LPF, the response to a rectangular pulse is an exponential decay. To reduce the cross talk, the transmitted pulses must be separated by T_g. The guard time T_g is the minimum pulse spacing so that the tail of the pulse decays to a value less than A_{CT} by the time next pulse arrives.



5.5 PULSE CODE MODULATION (PCM)

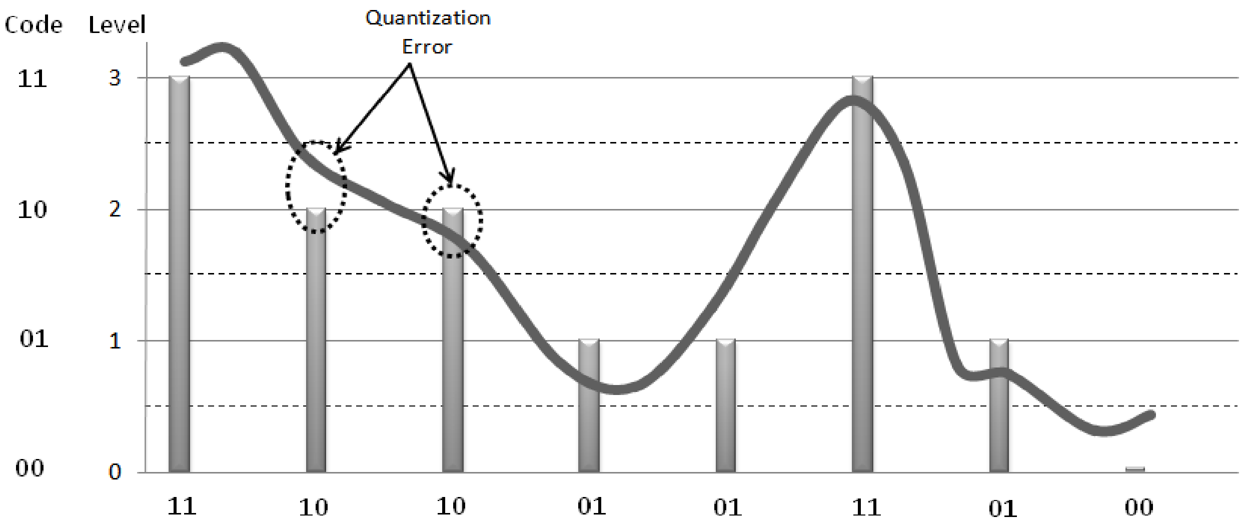
After sampling an analog signal, it is possible to code the samples using discrete symbols, such as binary. The communication through binary alphabets is more efficient than the analog transmission. If the value of a sample is sent using only two possible elements, the reception becomes more reliable. It is easier for the receiver to discriminate the reception into two possible signals than estimating the value of a continuous signal. This significantly helps in reducing the effects of distortion and of additive noise.

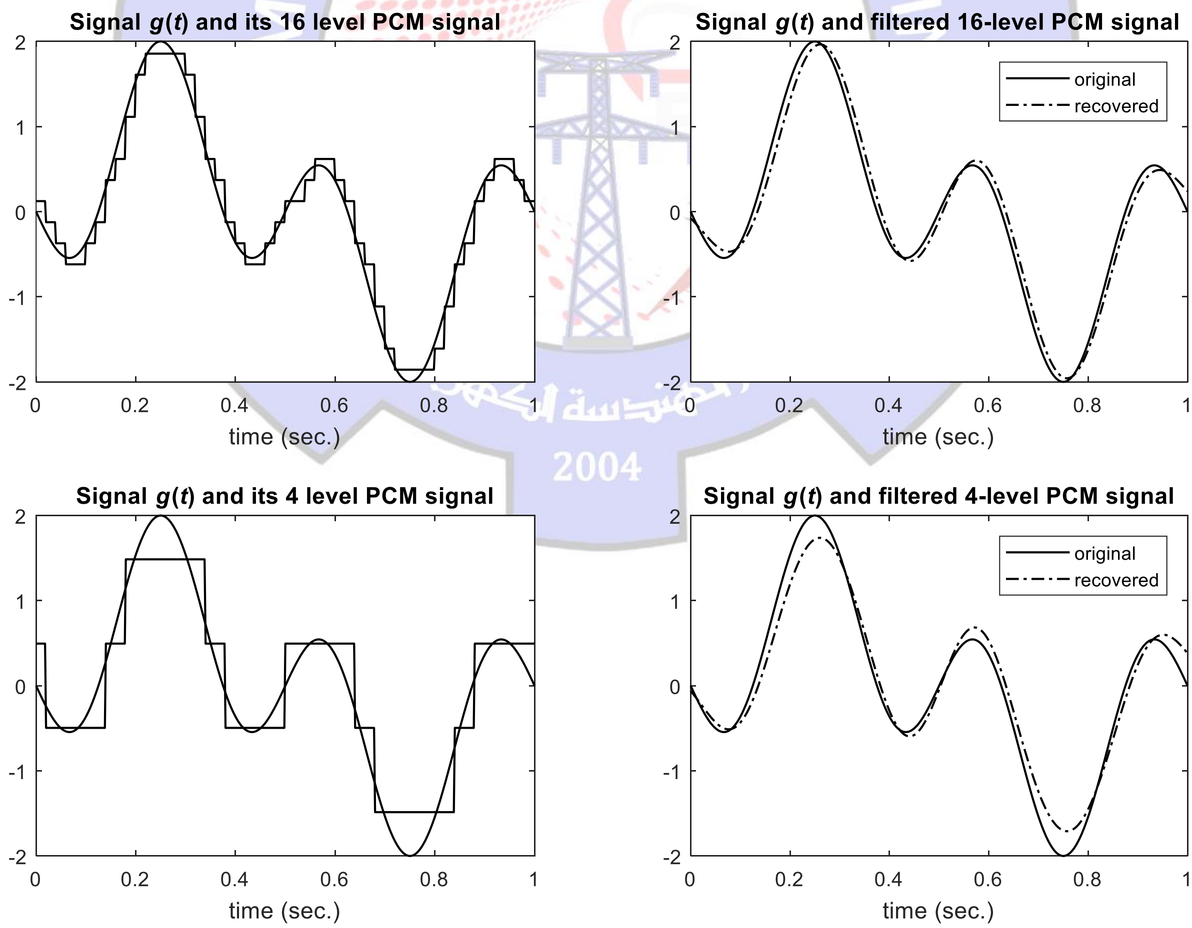
In PCM, the sample values are rounded to the values of certain levels. The rounding off operation is known as "quantization". Then each sample is coded into a binary number which is equivalent to the index of the quantization level that is closest to the sample value.

5.5.1 QUANTIZATION

When an information signal is pulse amplitude modulated, it becomes discrete in time only. It remains analogue in amplitudes since all the values within the specified range are allowed. PAM signal is said to be quantized when each pulse of the PAM signal is adjusted in amplitude to coincide with the nearest level within a finite set.

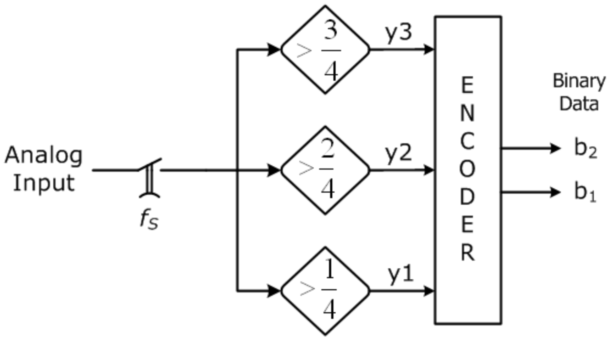
It is clear from the figure below, quantization error (noise) can be reduced by increasing the number of quantization levels (L), i.e. decreasing the intervals (q) between the levels.





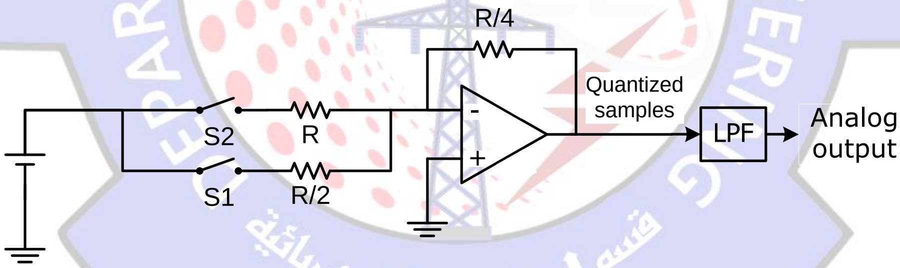
5.5.2 ENCODING

The quantized samples are now to be coded with l bits per sample ($l = \log_2 L$). One of the most popular quantizer/encoder circuits is the parallel quantizer, which requires $L - 1$ comparators. The following is a 2-bit PCM coder.

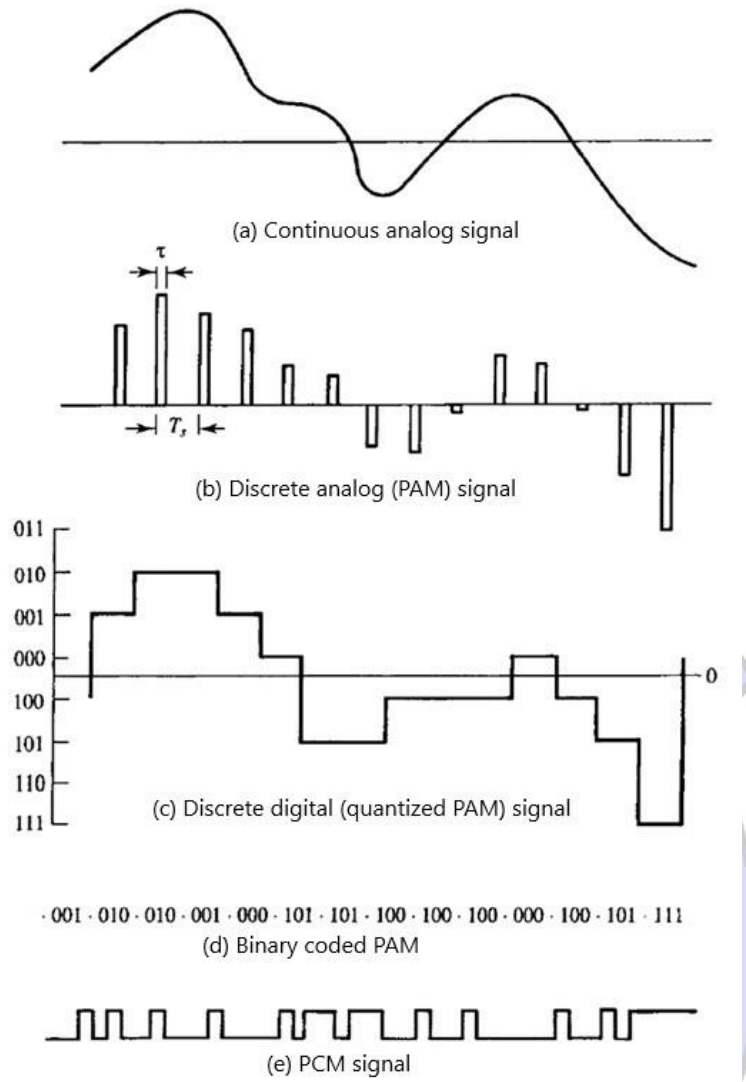


5.5.3 DECODING

The transmitter sends binary codes to the receiver via a channel. The receiver must decode the bit sequence back to a time function. This is done by associating each group of bits with the corresponding quantization level; thereby reconstructing the quantized waveform by LPF. In this simple 2-bit PCM decoder, S1 & S2 are received binary, S1=MSB and S2=LSB.

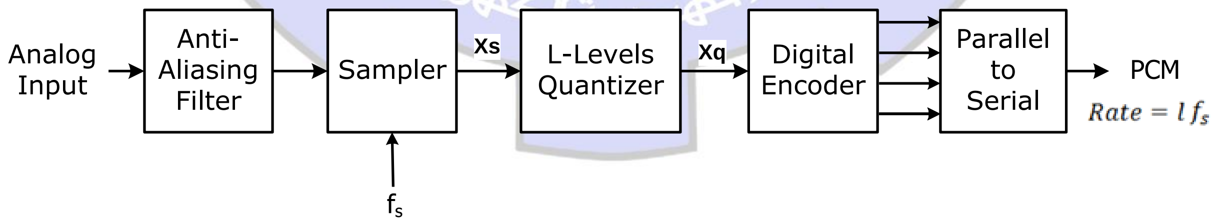


The PCM technique is considered as an Analog to Digital Converter (ADC) at the transmitter and Digital to Analog Converter (DAC) at the receiver. The following figure illustrates the ADC

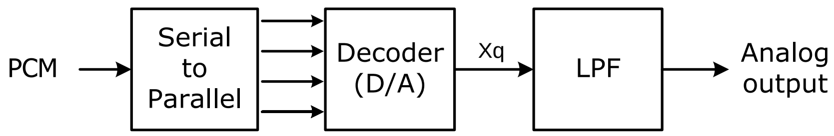


The complete block diagram of PCM system is:

Modulator:



Demodulator:



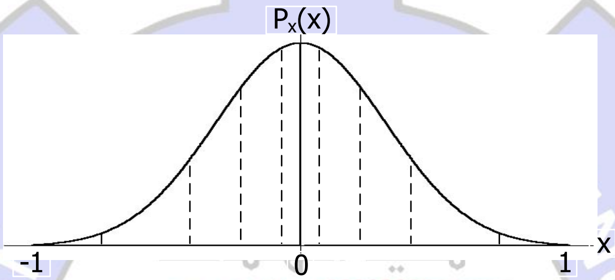
5.5.4 NON-UNIFORM QUANTIZATION

UNEQUAL STEP-SIZE

Uniform quantization assumes that the information signal has uniform PDF, i.e. all quantization levels are used equally. For most signals, it is not the case.

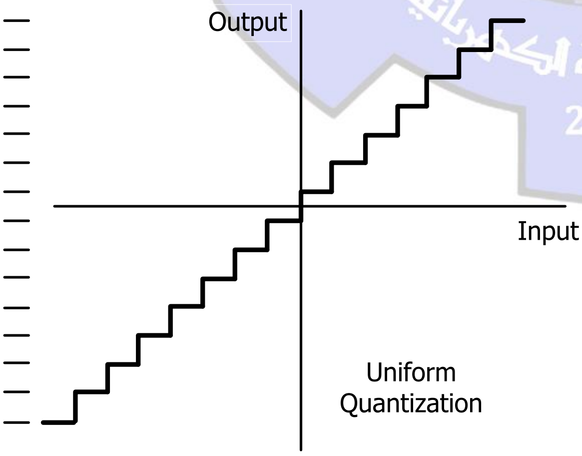
If the PDF of the information signal is not uniform (nevertheless known and constant with time), then we can optimize the locations of the quantization levels to obtain minimum quantization noise introduced.

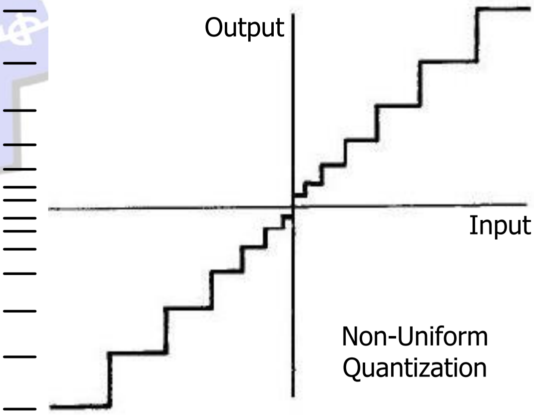
As an illustration, let the normalized signal $x(t)$ has the typical probability function $P_x(x)$:



The shape of $P_x(x)$ means $|x(t)| \ll 1$ most of the time. Therefore, we can use non-uniform quantization as indicated by the dashed lines. The quantization lines are located here close to each other near $x = 0$. They are sparse for large values of $|x(t)|$, as large $|x(t)|$ occurs infrequently.

The depictions below show the uniform and the non-uniform distribution of the quantization levels.

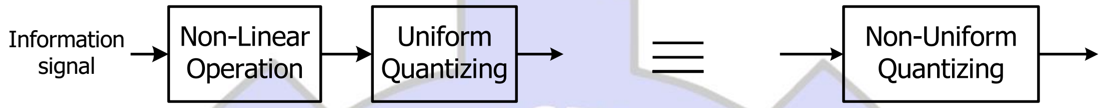




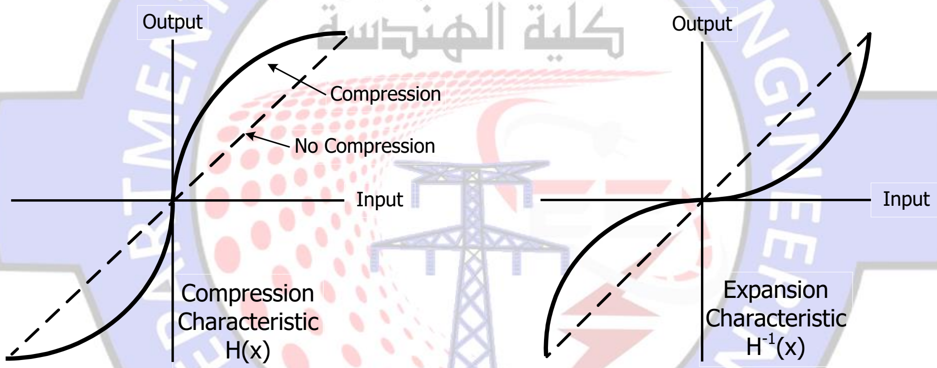
Practically, such optimization is a difficult procedure because it requires prior knowledge of the signal PDF. Problems arise if the information signal has an unknown PDF or if its PDF changes with time. However, they are similar in some signals. For example, in the case of voice signals, the PDF shape of different speakers is usually similar, but the gross level can vary widely between speakers, e.g. man is shouting and woman softly spoken. Therefore, the approach taken in practice is to use uniform quantizing after non-linear compression (the *companding*).

COMPANDING (*COMPRESSING-EXPANDING*)

This is the process of *compressing* the information signal prior to linear quantization at transmission. The compression is achieved via a non-linear amplitude characteristic circuit.



The receiver *expands* the reconstructed signal with the inverse characteristic to restore the original waveform.



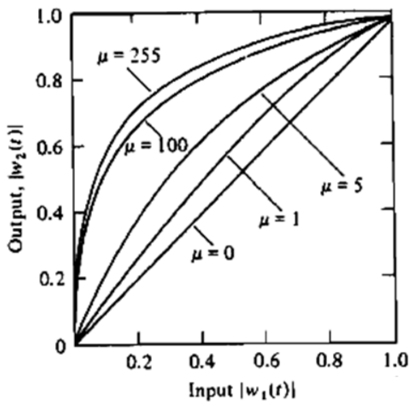
So, PCM with companding system will be:

Companding Methods

There are two types of companders that are practically and widely used for speech coding:

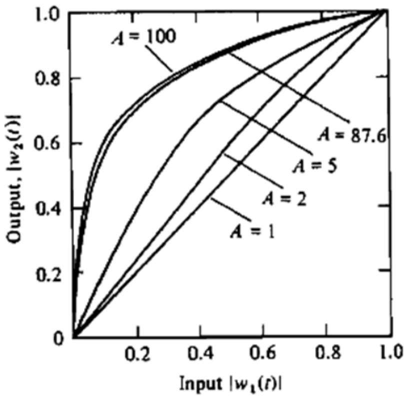
(1) The μ-law compander (used in the US, Canada and Japan). The parameter μ controls the amount of compression and expansion. The standard compressor uses $\mu = 255$ followed by a uniform quantizer with 128 levels (7 bits per sample).

$$H(x) = \frac{\log(1+\mu|x|)}{\log(1+\mu)}$$



(2) The A-law compander (used in most countries). A is chosen to be 87.56.

$$H(x) = \begin{cases} \dfrac{A|x|}{1+\ln A} & \text{for } 0 \le |x| \le \dfrac{1}{A} \\[2ex] \dfrac{1+\ln(A|x|)}{1+\ln A} & \text{for } \dfrac{1}{A} \le |x| \le 1 \end{cases}$$



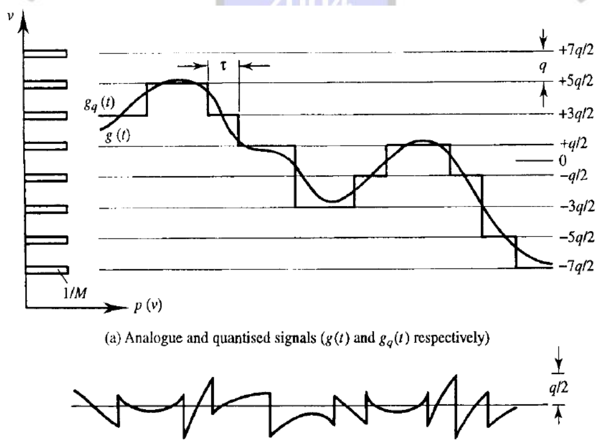
5.5.5 SIGNAL TO QUANTIZATION NOISE RATIO

It is important to consider the quantization noise in the overall system quality. To calculate SN_qR of a uniformly quantized signal, it is suitable to make the following assumptions:

(1) Linear quantization (i.e. equal increments between quantization levels).

(2) Zero mean signal (i.e. symmetrical PDF around the 0 Volt).

(3) Uniform signal PDF (i.e. all signal levels equally likely).



(a) Analogue and quantised signals ($g(t)$ and $g_q(t)$ respectively)

(b) Quantised minus analogue signal, $\varepsilon_q(t) = g_q(t) - g(t)$

Let: L be the number of the levels of the quantizer, l be the number of bits per a PCM word ($L = 2^l$), and V_p be the peak design level of the quantizer. The quantization interval q becomes:

$$q = \frac{\text{p.p. voltage}}{\text{No. of levels}} = \frac{2V_p}{L}$$

The PDF of the allowed levels is given by:

$$p(v) = \sum_{\substack{k=-L \\ k=\text{odd}}}^{L} \frac{1}{L}\delta\left(v - \frac{qk}{2}\right)$$

The mean square signal after quantization is:

$$\overline{v^2} = \int_{-\infty}^{\infty} v^2 p(v) dv = \frac{2}{L}\left[\int_0^{\infty} v^2\delta\left(v - \frac{q}{2}\right) dv + \int_0^{\infty} v^2\delta\left(v - \frac{3q}{2}\right) dv + \cdots\right]$$

$$= \frac{2}{L}\left(\frac{q}{2}\right)^2 [1^2 + 3^2 + 5^2 + \cdots + (L-1)^2] = \frac{2}{L}\left(\frac{q}{2}\right)^2 \left[\frac{L(L-1)(L+1)}{6}\right]$$

$$\therefore \overline{v^2} = \frac{q^2}{12}(L^2 - 1)$$

Denoting the quantization error (i.e. the difference between the unquantized and quantized signals) as ε_q, then the PDF of ε_q is uniform:

$$p(\varepsilon_q) = \begin{cases} \frac{1}{q} & -\frac{q}{2} \le \varepsilon_q < \frac{q}{2} \\ 0 & \text{elsewhere} \end{cases}$$

The mean square quantization error (noise) is:

$$\overline{\varepsilon_q^2} = \int_{-q/2}^{q/2} \varepsilon_q^2\, p(\varepsilon_q)\, d\varepsilon_q = \frac{q^2}{12}$$

Therefore, the average $\mathrm{SN_qR}$ will be:

$$\mathrm{SN_qR} = \overline{v^2}/\overline{\varepsilon_q^2} = L^2 - 1$$

Since the peak signal level is $\frac{qL}{2}$ Volts then the peak $\mathrm{SN_qR}$ will be:

$$\mathrm{SN_qR} = \frac{(Lq/2)^2}{\overline{\varepsilon_q^2}} = 3L^2$$

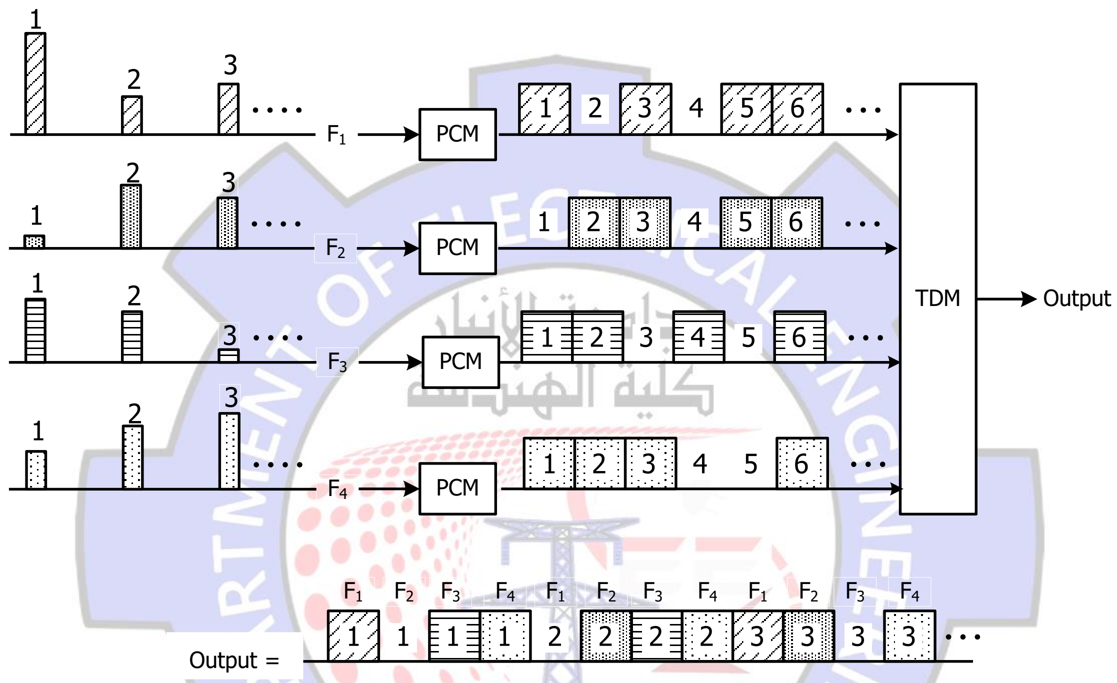
We may express it in decibels as

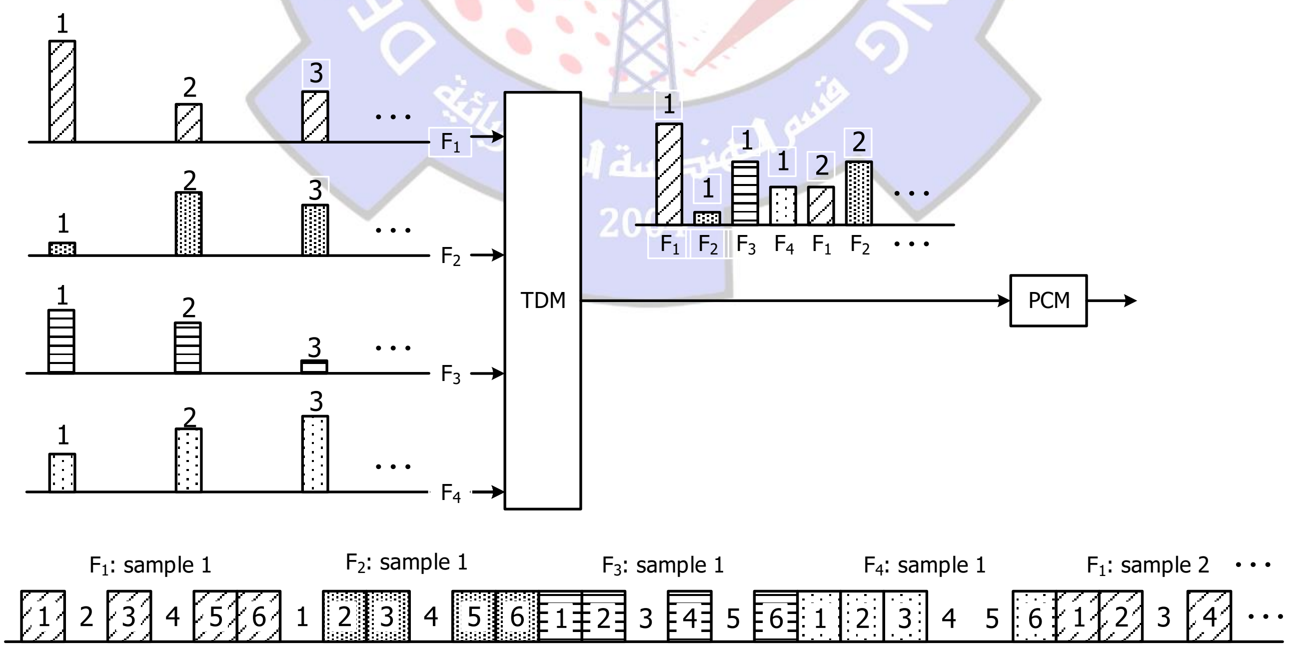
$$\mathrm{SN_qR} = 6.02l + \alpha$$

Where $\alpha = 4.77$ for the peak SN_qR, and $\alpha = 0$ for the average SN_qR. This equation is called the *6dB rule*, and it points out that: an additional 6dB improvement in the SN_qR is obtained for each bit added to the PCM word.

5.5.6 PCM MULTIPLEXING

The output PCM signal rate $= Nlf_s$ where: N = number of multiplexed signals, l = number of bits per sample and f_s = sampling frequency.





What is the difference between *Information Rate* and *Baud Rate*?

5.6 PCM Quality Versus Required Rate

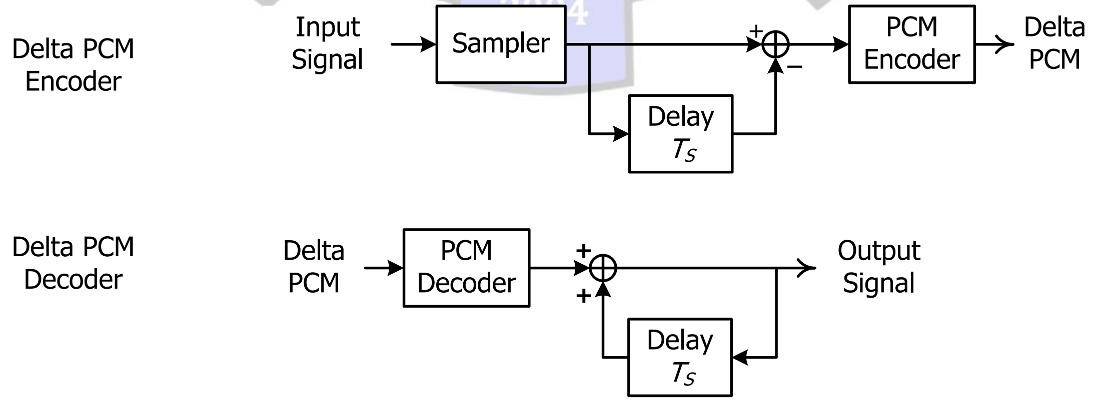
In addition to the channel effects, PCM performance depends primarily on the quantization noise. To make the reconstructed signal like the original baseband signal we must reduce the quantizing noise by increasing L. Now, we learned that increasing the number of quantization levels requires more bits per sample to be transmitted. Large l is not perfect for crowded channels and due to some channels limitations. So, beside the non-uniform quantization, several techniques are used to reduce the quantization noise at the same number of the L.

5.7 Bandwidth Reduction Techniques

The channel bandwidth is limited and it is a valuable resource. A frequent objective of the communications engineer is to transmit the maximum information rate via the minimum possible bandwidth. This is especially true for radio communications in which radio spectrum is a scarce, and therefore valuable, resource. The following systems are used to maintain the same coding fidelity using fewer bits per sample.

5.7.1 Delta PCM

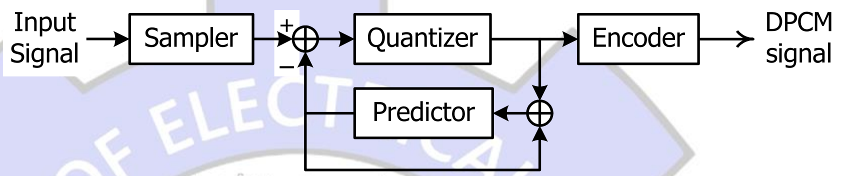
Because the samples of most of the baseband signals are highly correlated, it is possible to transmit the information about the changes between samples instead of sending the sample values themselves. A simple way for such systems is the Delta PCM. This method transmits the difference between adjacent samples through code words. This difference is significantly less than the actual sample values, hence it is coded using fewer binary symbols per word than the conventional PCM. However, Delta PCM systems cannot accommodate rapidly varying transient signals.



5.7.2 DEFERENTIAL PCM

Since neighbor samples within many information signals are highly correlated, Deferential PCM (DPCM) uses an algorithm to predict future values. Such algorithms monitor the trend of the baseband samples and use some models to predict the value of the incoming samples. Then DPCM waits until the actual value becomes available for examination and transmits the correction to the already predicted value. The correction signal represents the information signal's unpredictable part. By this means, DPCM reduces the redundancy in signal and allows the information to be transmitted using fewer symbols, less spectrum, and shorter time.

DPCM Encoder

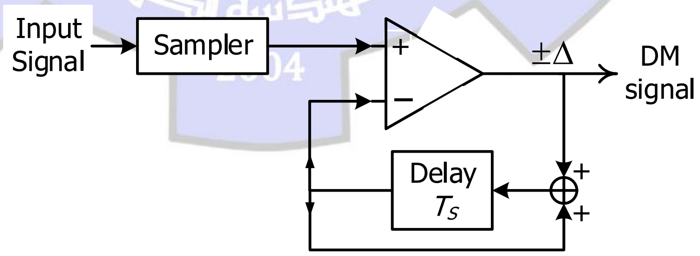


DPCM Decoder

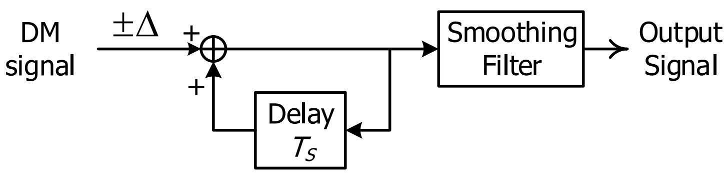
5.8 DELTA MODULATION (DM)

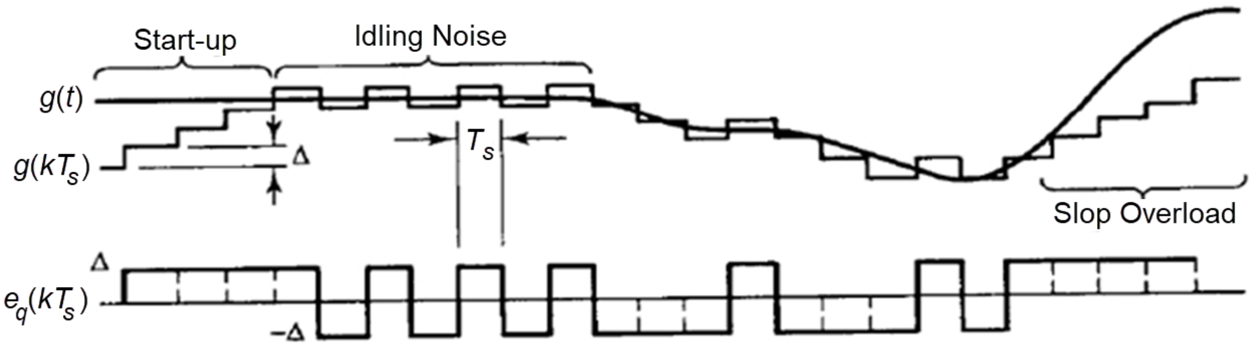
If the quantizer of the DPCM system is restricted to one bit (i.e. the two levels only: $\pm\Delta$) and the predictor to one sample delay, then the resulting scheme is called DM. The information signal is represented by a stepped waveform. The resolution of this waveform depends on Δ & T_s values.

DM Encoder



DM Decoder

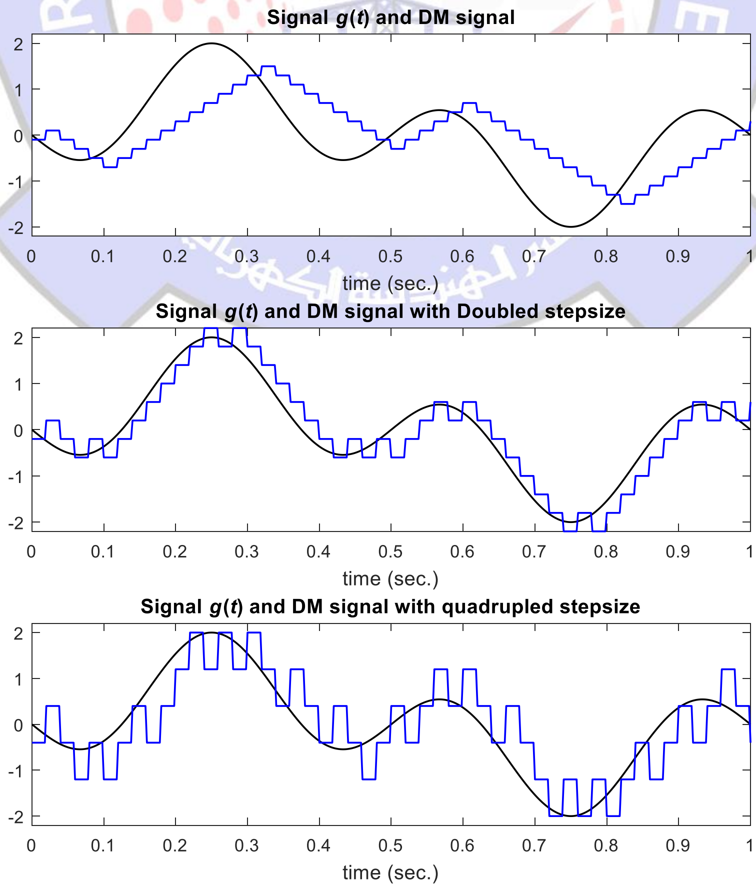




We observe from the figure above that:

- The system requires notable time for the start-up to catch the required level of the input.
- If the input remains constant, the reconstructed DM waveform exhibits ripple known as *idling noise*, which almost filtered out at the receiver.
- The rate-of-rise overload problem occurs when the input changes too rapidly for the stepped waveform to follow.

To resolve these problems, we are clearly in a dilemma about how to choose the best values for Δ and T_s. A small step size (Δ) is desirable for accuracy, but the clock rate ($f_s = 1/T_s$) should be fast to avoid *slop-overload*, which is not recommended. Inversely, as f_s get smaller, many small details of the information signal will be lost. This dilemma is depicted in the figures below:



However, we can calculate the optimum values for these parameters that overcome the slop-overload problem. An estimate of the rate-of-rise condition for DM may be obtained quit easily for sinusoidal modulation. Let the input be $g(t) = b\cos(\omega_m t)$, so that:

$$\max\left[\frac{dg(t)}{dt}\right] = 2\pi b f_m = \text{ maximum slop of this signal.}$$

Since the maximum rate-of-rise $= \frac{\Delta}{T_s} = \Delta f_s$, then

$$\Delta f_s \geq 2\pi b f_m \quad \Leftrightarrow f_s \geq \frac{2\pi b f_m}{\Delta} \quad \Leftrightarrow \quad b \leq \frac{\Delta f_s}{2\pi f_m}$$

The above f_s condition may be applied to band-limited signals by letting f_m be the highest frequency component, and $b = \max|g(t)|$.

Since the quantization noise:

$$\overline{e^2} = \int_{-\Delta}^{\Delta} \frac{e^2}{2\Delta} de = \frac{\Delta^2}{3}$$

And by filtering this noise to a bandwidth B, we get:

$$N_q = \overline{n_q^2} = \frac{\Delta^2 B}{3 f_s}$$

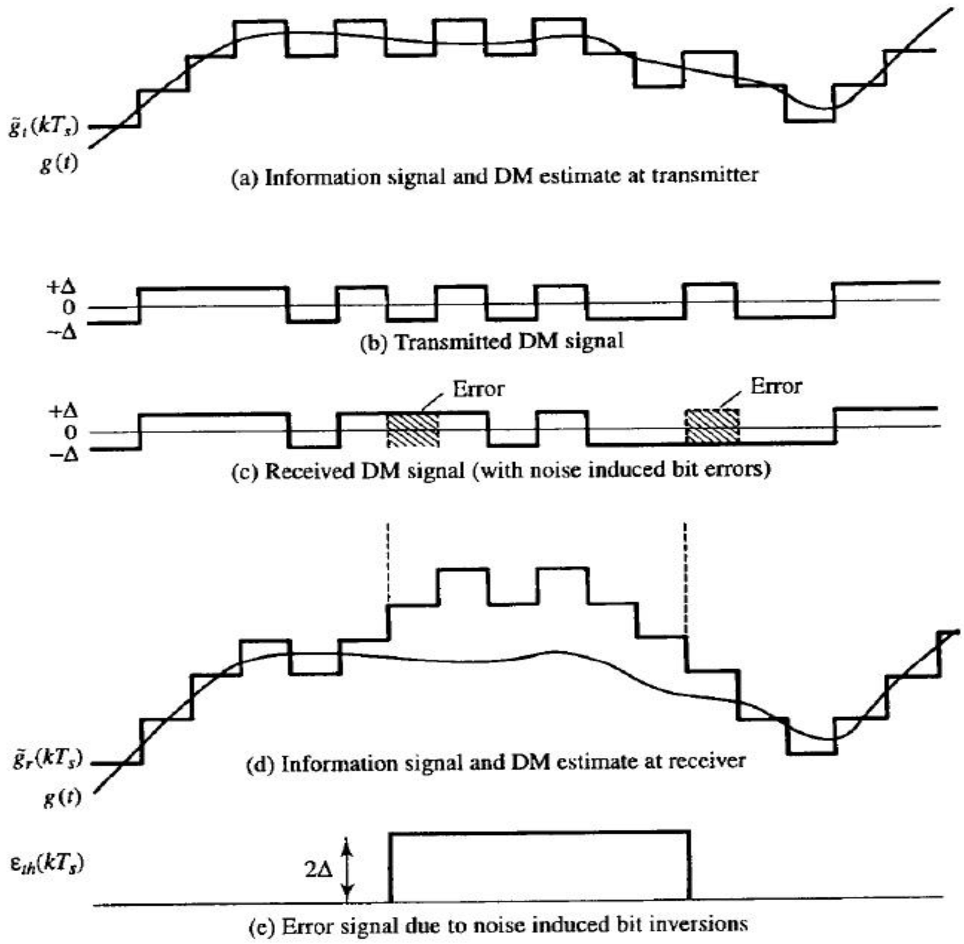
The mean-square value of the information signal is:

$$S = \overline{g^2(t)} = \left(\frac{V_p}{\sqrt{2}}\right)^2 = \frac{b^2}{2} = \frac{1}{8}\left(\frac{\Delta f_s}{\pi f_m}\right)^2$$

$$\therefore \text{SN}_\text{q}\text{R} = \frac{3 f_s^3}{8\pi^2 f_m^2 B}$$

Errors in DM

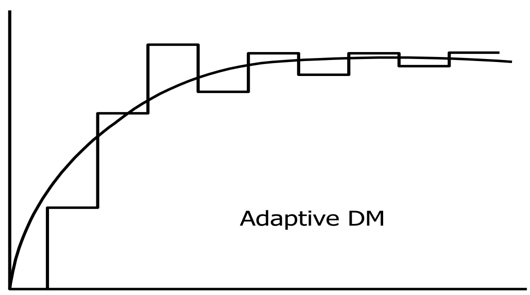
If the SNR is not sufficiently high, then the DM receiver will occasionally interpret a received symbol in error (i.e. $+\Delta$ instead of $-\Delta$ or the converse), and this is equivalent to the addition of an error of 2Δ to the accumulated signal at the DM receiver, as shown below. This situation continues until another error occurs which either cancels the first error or double it!



Adaptive Delta Modulation

We have seen that: a large step size causes unacceptable quantization noise, and a small step size results in sample-overload distortion. This means that a good choice for Δ is a "medium" value, but in some cases, the performance of the best "medium" values is not satisfactory. An approach that works well in these cases is to change the step size according to changes in the input: if the input tends to change rapidly, the step size is chosen to be large (and vice versa). So, the output can follow the input quickly without distortion.

DM is primarily used for telemetry systems and speech transmission in telephone. It has been found that: PCM is preferable for high quality speech transmission, whereas DM is easier to implement and yields transmission of acceptable quality.



5.9 CHANNEL CAPACITY

For any communication system, it is required to send data as fast as possible. But in the presence of noise and distortion along the channel, it is very hard to avoid errors at the reception. The Hartly-Shannon theorem of channel capacity states that: the maximum rate of information transmission $R_{\max}$ over a channel of the bandwidth B and the received signal to ratio SNR is given by:

$$R_{\max} = C = B\log_2\left(1+\frac{S}{N}\right) \quad \text{bps}$$

Where C = channel capacity = the maximum rate at which information can be transmitted across that channel without error; it is measured in bits per second (bps).

let R be the operating information rate.

- If $R \leq C$, it is possible to receive data with small probabilities of error, even with noise.
- If $R > C$, errors can not be avoided regardless of the coding technique used.

If we would like to increase $R_{\max}$ in the above equation, we can increase B and/or the SNR.

- Larger SNR implies working at, for example, higher Tx powers or shorter distances. In some cases, this is not that easy to achieve. But in general, as the SNR increased R would also increase without errors for a given B channel. However, you must note that: when $N \to 0$ then $SNR \to \infty$ and hence $R_{\max} \to \infty$ regardless of B (is it possible?).
- As for increasing B, it requires changing the medium or buying a license for extra bandwidth. In general, as B increased, it can follow faster changes in the information signal, thereby increasing R. Nevertheless, when $B \to \infty$, C does not approach ∞. The noise is assumed to be white: the wider the bandwidth, the more the noise admitted to the system. This means, as B increases, SNR decreases at the same S.

Theoretical Capacity

Suppose that the noise is white with PSD $\eta/2$ (W/Hz), and assume the received signal power is fixed at a value S (W), the channel capacity would be:

$$R_{\max} = C = B\log_2\left(1+\frac{S}{\eta B}\right)$$

thus, when $B \to \infty$ we get:

$$C = \lim_{B\to\infty}\left\{B\log_2\left(1+\frac{S}{\eta B}\right)\right\} = \lim_{B\to\infty}\left\{\frac{S}{\eta}\log_2\left(1+\frac{S}{\eta B}\right)^{\frac{\eta B}{S}}\right\} \approx \frac{S}{\eta}\log_2 e$$

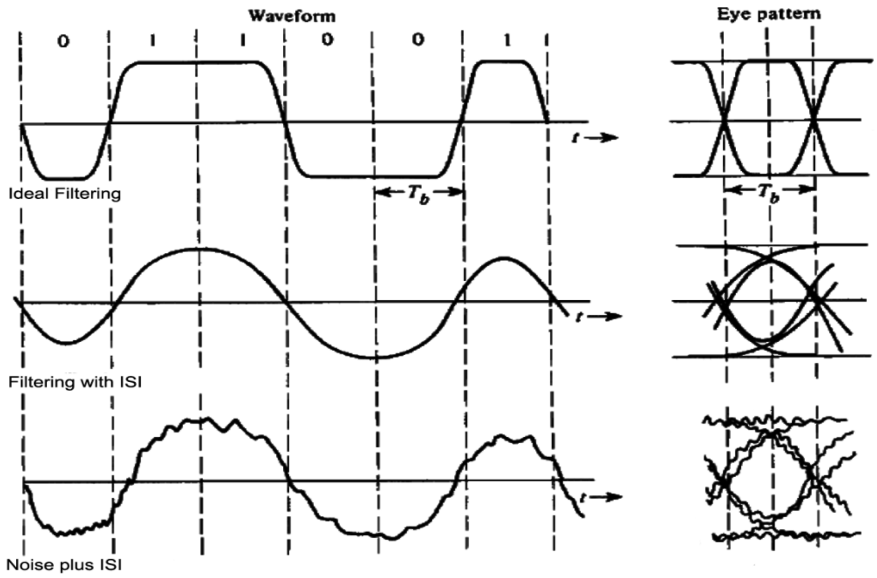
$$\therefore C = 1.44\frac{S}{\eta}$$

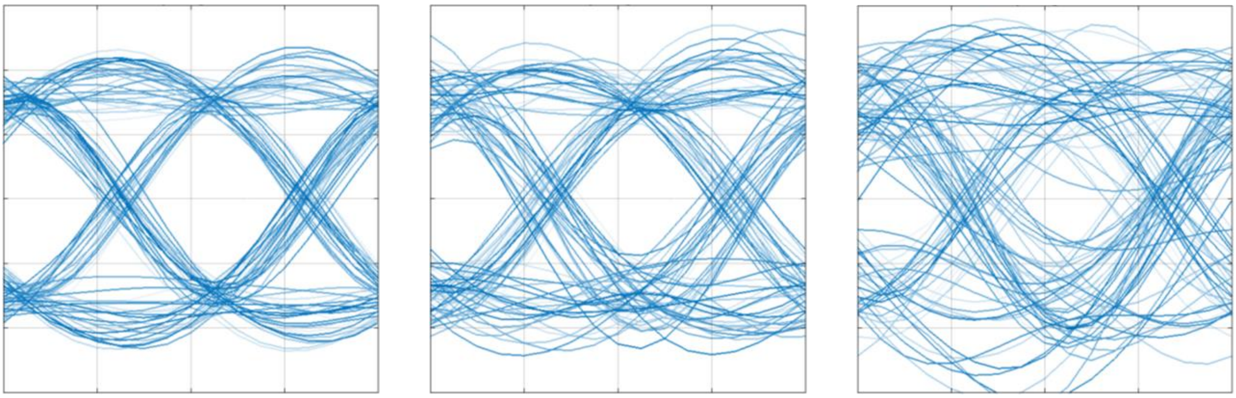
This gives the maximum possible channel capacity as a function of the received signal power and the noise PSD. In actual systems design, the channel capacity might be compared to this value to decide whether a further increase in B is worthwhile.

5.10 INTER-SYMBOL INTERFERENCE (ISI)

Previously, we discussed the crosstalk problem (distortion caused by time dispersion). This results in spreading of time signals and overlapping among adjacent bit waveforms. This overlap is also known as ISI. It is caused not only by channel distortion but also by multi-path effects.

To illustrate the ISI problem, we should first introduce the *eye pattern*. *(in class)*





SNR=20dB SNR=15dB SNR=10dB

To reduce the ISI, we must:

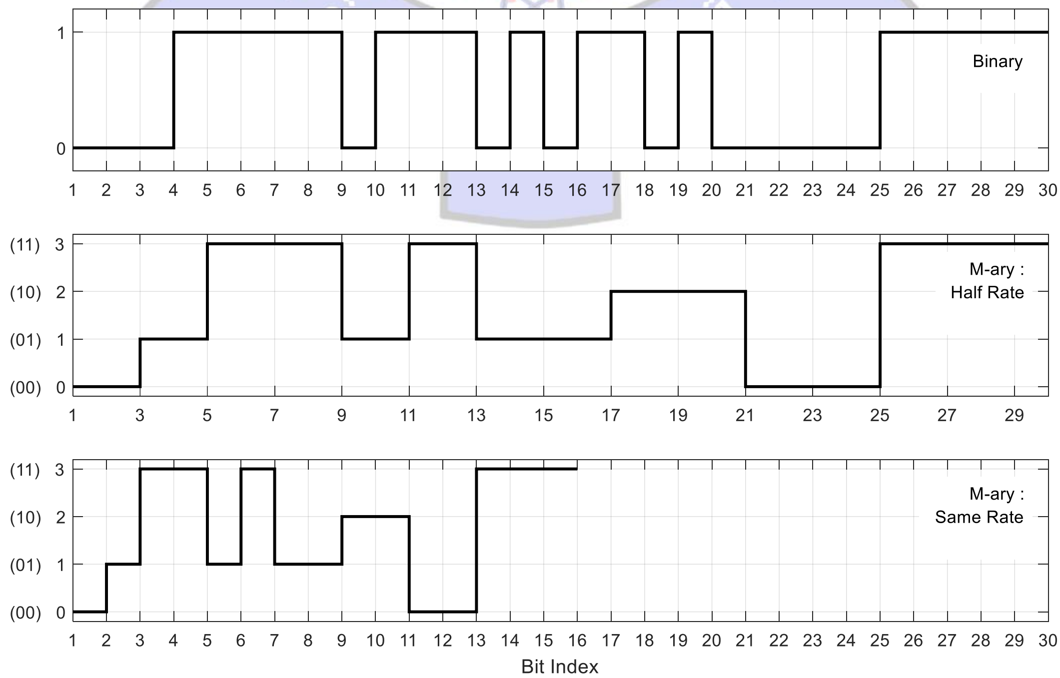
(a) Consider the Shannon criterion for the maximum rate.

(b) Frequency limitation of the transmitted signal to fit B.

(c) Reduce the effects of the multi-path problem *(How?)*.

5.11 MULTI-LEVEL BASEBAND SIGNALING (M-ARY)

Binary shift keying means sending a single bit over the symbol interval T_b (or at the bit rate $1/T_b$ bits/sec). To increase the data transfer rate, it is possible to combine several bits in one symbol. In this technique, we send one symbol per m data bits. So, we the transmitted symbols range between M levels (know that $M = 2^m$). The plots below illustrate the M-ary transmission.



In this example, we set 2 bits per symbol (hence $2^2 = 4$ levels). The information rate is still unchanged, but the symbol rate is halved (as in the second plot). On other words, we can send twice the information rate at the same symbol rate (as in the third plot). Practically, it is difficult to consider the multi-level baseband signaling through noisy and distorting channels. the receiver now must distinguish the incoming symbols according to their levels. Thus the probability of error increases as M becomes larger.

Advantage:

- A higher information transfer rate is possible for a given symbol rate and a channel bandwidth.

Disadvantages:

- M-ary baseband signaling results in reduced noise/interference immunity when it is compared to the binary signaling.
- It involves more complex symbol recovery processing in the receiver.
- It imposes a greater requirement for linearity and/or reduced distortion in the Tx/Rx hardware and in the channel.

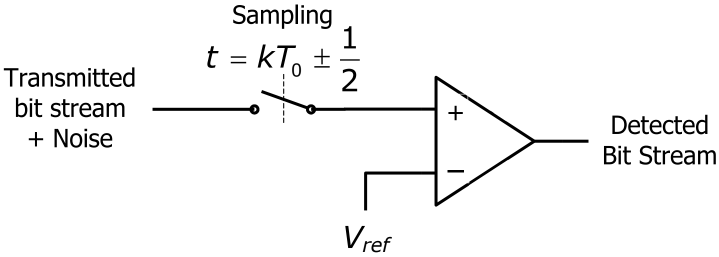
5.12 Probability of Error at Reception

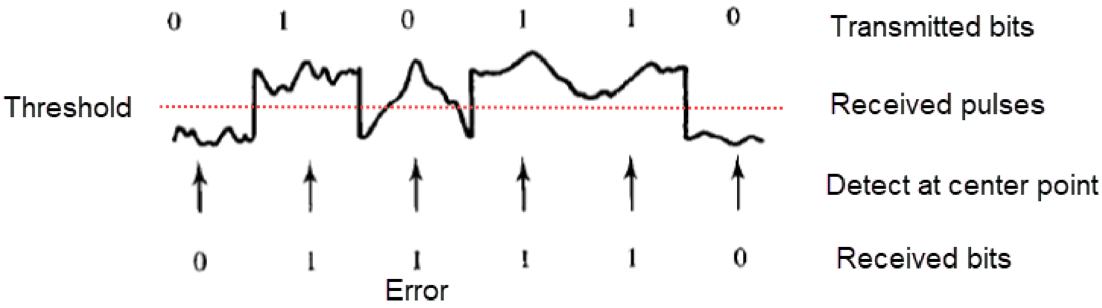
The detection of digital signals involves two processes:

(1) Reduction of each received voltage pulse (i.e. symbol) to a single numerical value, (just like quantization).

(2) Comparison of this value with a reference voltage to determine which symbol was transmitted.

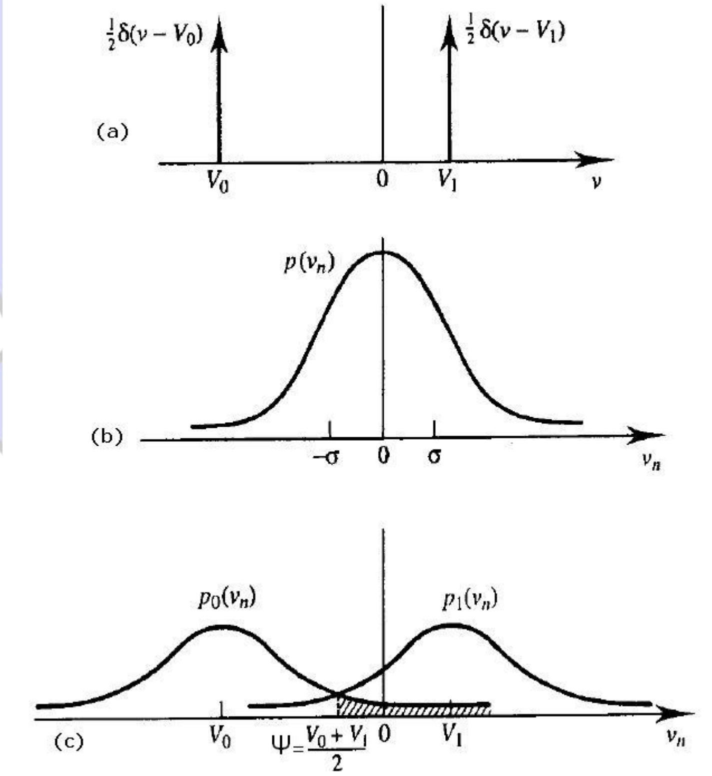
The unipolar binary symbols (0 and 1) is represented by two voltage levels (e.g. 0V and 3V). Intuition tells us that a sensible strategy would be to set the reference, V_{ref}, mid-way between the two voltage levels (i.e. at 1.5V).





In general, for equiprobable symbols the decision level is set to $\psi = \frac{V_0+V_1}{2}$. For unequal transmission of the symbols, we need to determine the optimum threshold using more advanced mathematics.

Since the noise with a Gaussian PDF♣ is common and analytically tractable, the bit error rate (BER) of a communication system is often modeled assuming Gaussian noise alone.



(a) The PDF of a binary information signal which can employ voltage levels V_0 and V_1 only. (b) The PDF of a zero mean Gaussian noise process, $v_n(t)$, with RMS value σ Volts. (c) The PDF of the sum of the signal and the noise.

♣ See Str. Section 8.6.4.

Let the probability of sending symbol 0 (the voltage level V_0) is:

$$p_0(v_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_n - V_0)^2}{2\sigma^2}}$$

And the PDF of sending symbol 1 (the voltage level V_1) is:

$$p_1(v_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_n - V_1)^2}{2\sigma^2}}$$

Now, if the symbol 0 is transmitted, let P_{e1} be the probability of the received signal plus noise that is above the threshold at the decision instant (i.e. seen as voltage level V_1), [shaded area under the curve $p_0(v_n)$ in Figure (c)]. The equation of the probability of error is:

$$P_{e1} = \int_{\psi}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_n - V_0)^2}{2\sigma^2}} dv_n$$

Also, if the digital symbol 1 is transmitted, let P_{e0} be the probability that the received signal plus noise is below the threshold at the decision instant (i.e. seen as voltage level V_0). So:

$$P_{e0} = \int_{-\infty}^{\psi} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_n - V_1)^2}{2\sigma^2}} dv_n$$

It is clear from the symmetry of this problem that P_{e0} is identical to P_{e1}, and for equiprobable symbols $p_0(v_n) = p_1(v_n) = \frac{1}{2}$, the net probability of error P_E will be:

$$P_E = p_0 P_{e0} + p_1 P_{e1} = \frac{1}{2}(P_{e0} + P_{e1}) = \int_{\psi}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_n - V_0)^2}{2\sigma^2}} dv_n$$

using $x = \dfrac{v_n - V_0}{\sigma}$ $\Rightarrow P_E = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{\frac{\psi}{\sigma}}^{\infty} e^{-x^2} dx$

This integral cannot be evaluated analytically but it can be recast as a complementary error function, which is defined by:

$$\text{Erfc}(z) \approx \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-x^2} dx$$

Thus $$P_E = \text{Erfc}\left(\frac{\psi}{\sigma}\right)$$

The advantage of using Erfc (also called Q-function) in the expression for P_E is that this function has been extensively tabulated. Sometimes is tabulated as erf(z) or erfc(z)♣.

For **unipolar** binary $\left(V_0 = 0,\ \ V_1 = A\ \ \rightarrow \psi = \frac{A}{2}\right)$, the average signal power is: $S = \frac{(0^2+A^2)}{2}$

And the average noise power is $N = \sigma^2$, so that: $P_E = \text{Erfc}\left\{\sqrt{\frac{S}{2N}}\right\}$

For **polar** binary $\left(V_0 = \frac{-A}{2},\ \ V_1 = \frac{A}{2} \rightarrow \psi = 0\right)$, the average signal power is: $S = \frac{\left[\left(\frac{-A}{2}\right)^2+\left(\frac{A}{2}\right)^2\right]}{2}$

And the average noise power is $N = \sigma^2$, we get: $P_E = \text{Erfc}\left\{\sqrt{\frac{S}{N}}\right\}$

Therefore, the average transmitted power for the unipolar signal must be twice that of the polar binary signal to achieve the same probability of error. For equiprobable signals, the polar binary signaling also has an advantage in the optimum decision threshold. It is simply set at the zero Volt, whereas the receiver for the ON-OFF binary signaling, the threshold must be adjusted to half the amplitude of the received signal.

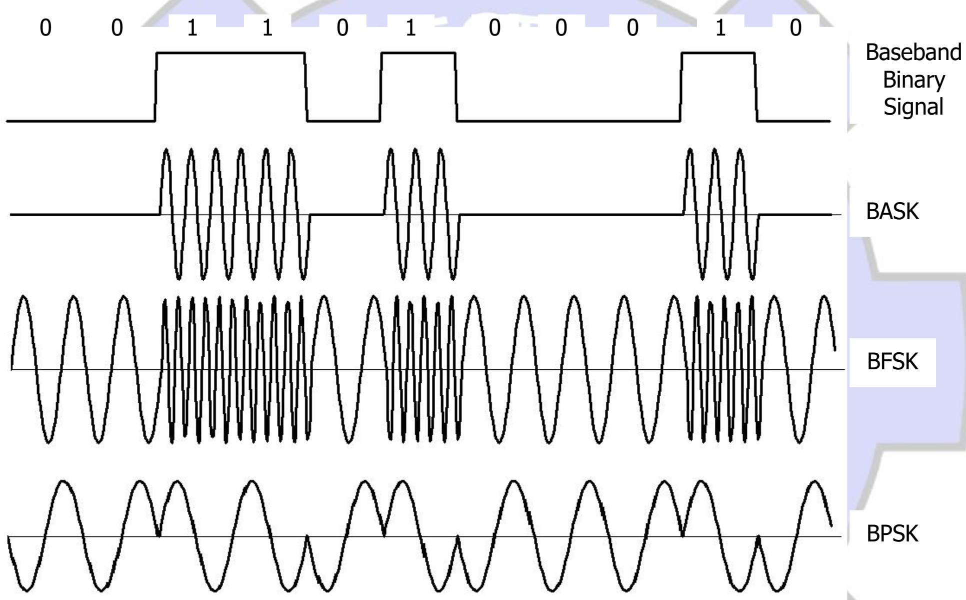
So, which is the better: the polar or the unipolar signaling in terms of the probability of error? why?

♣ See **Error! Reference source not found.** in this document and Str. Appendix G.

Part 6 DIGITAL MODULATION

6.1 BINARY DIGITAL MODULATION

Since pulse-modulated signals consist of "low" frequencies, they cannot be efficiently transmitted through a channel with band-pass characteristics. Hence, for communication systems employing band-pass channels, it becomes advantageous to modulate a carrier signal with the digital data stream prior to transmission. Three basic forms of digital modulation corresponding to AM, FM & PM are known as Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), and Phase Shift Keying (PSK).



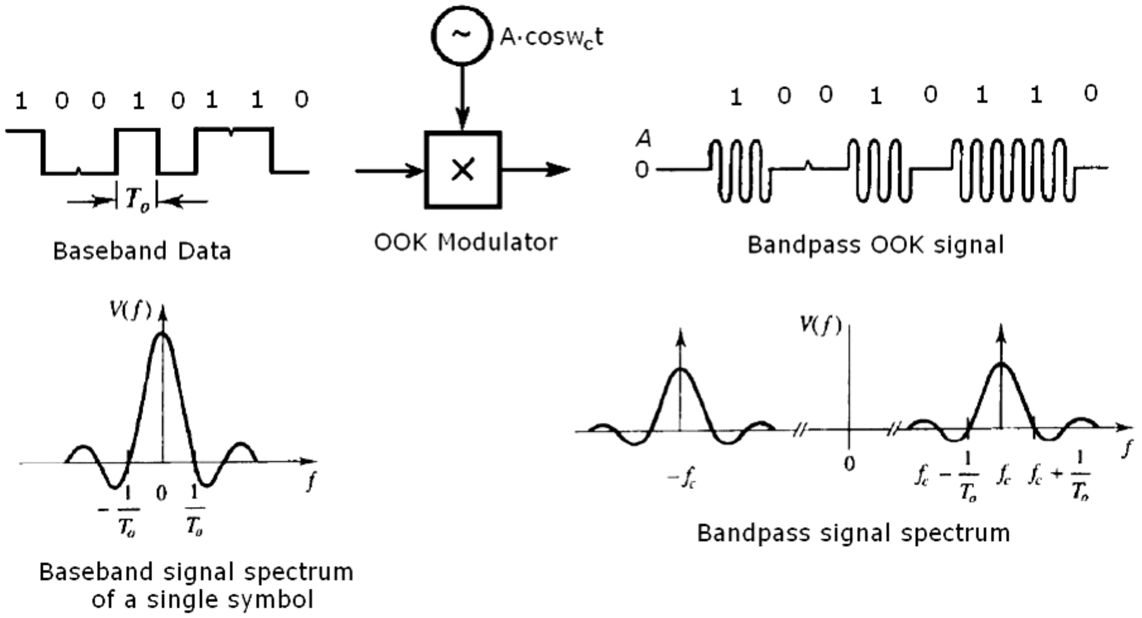
6.2 BINARY AMPLITUDE SHIFT KEYING (BASK)

In BASK, the amplitude of a high-frequency carrier is switched between two values, ON-OFF Keying (OOK):

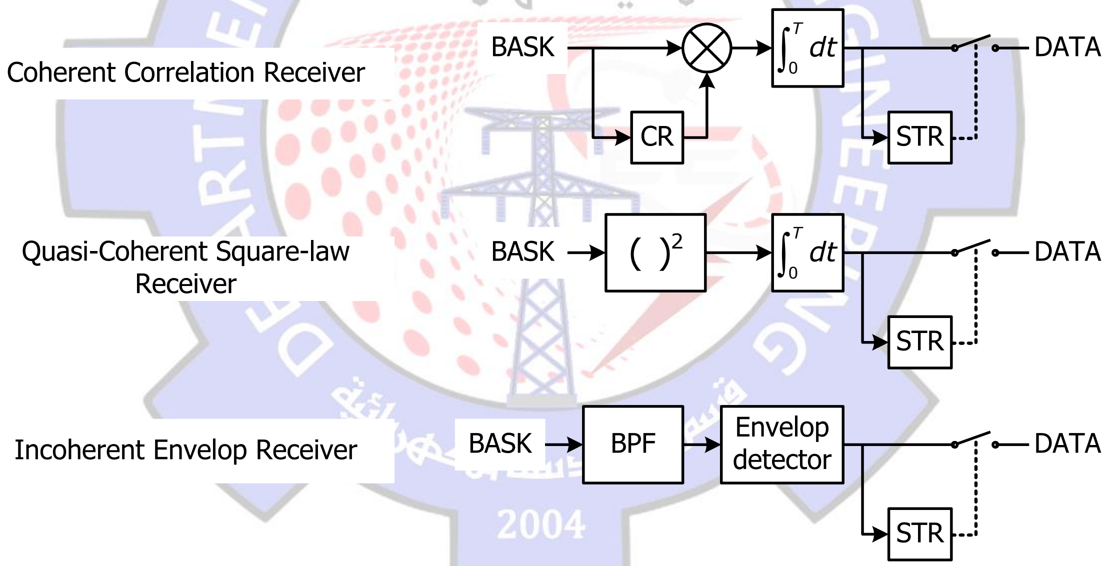
$$x(t) = \begin{cases} A\cos(\omega_c t) & \text{for logic 1} \\ 0 & \text{for logic 0} \end{cases}$$

The bandwidth of BASK signal = ?

6.2.1 GENERATION



6.2.2 DEMODULATION



6.2.3 PROBABILITY OF ERROR IN BASK

The receiver must decide based on two possibilities:

$$y(t) = \begin{cases} A\cos(\omega_c t) + n_0(t) & \text{for logic 1} \\ n_0(t) & \text{for logic 0} \end{cases}$$

Where $n_0(t)$ is the input noise to the decision maker.

As mentioned earlier, for $p_0(v_n) = p_1(v_n) = \frac{1}{2}$, the optimum decision threshold is set at $E_1/2$, (E_1= received signal energy for logic bit 1, when $E_0 = 0$). So, with the Gaussian distributed noise, the probability of error is:

$$P_E = \text{Erfc}\left(\sqrt{\frac{E_1}{2\eta}}\right)$$

since $S = \dfrac{E_{AV}}{T_0} = E_{AV}R$ and $N = \eta B \Rightarrow \dfrac{E_{AV}}{\eta} = T_0 B \dfrac{S}{N}$

Where: S = average received signal power,

E_{AV} = average received signal energy $= \dfrac{(E_0 + E_1)}{2} = \dfrac{E_1}{2}$

R = binary data rate $= \dfrac{1}{T_0}$

T_0 = Binary bit interval.

B = Decision maker bandwidth.

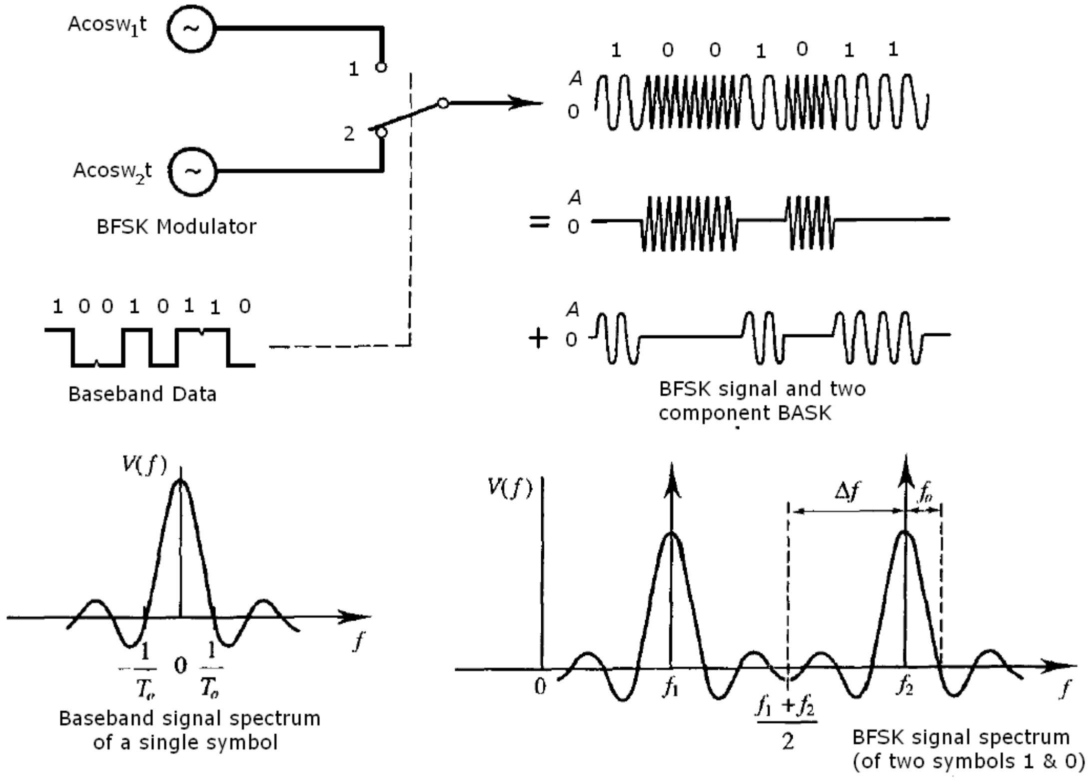
$P_E = \text{Erfc}\left(\sqrt{\dfrac{S}{N}}\right)$ Coherent Detection at $T_0 B = 1$

$P_E \approx \dfrac{1}{2}\exp\left(\dfrac{-E_{AV}}{2\eta}\right)$ Incoherent Detection

6.3 Binary Frequency Shift Keying (BFSK)

In BFSK, the instantaneous frequency of the carrier is switched between two values in response to the binary code. We can consider the BFSK waveform as a composition of two BASK waveforms of different carrier frequencies.

$$x(t) = \begin{cases} A\cos(\omega_1 t) & \text{for logic 1} \\ A\cos(\omega_2 t) & \text{for logic 0} \end{cases}$$



6.3.1 GENERATION

As shown above, a BFSK carrier frequency is:

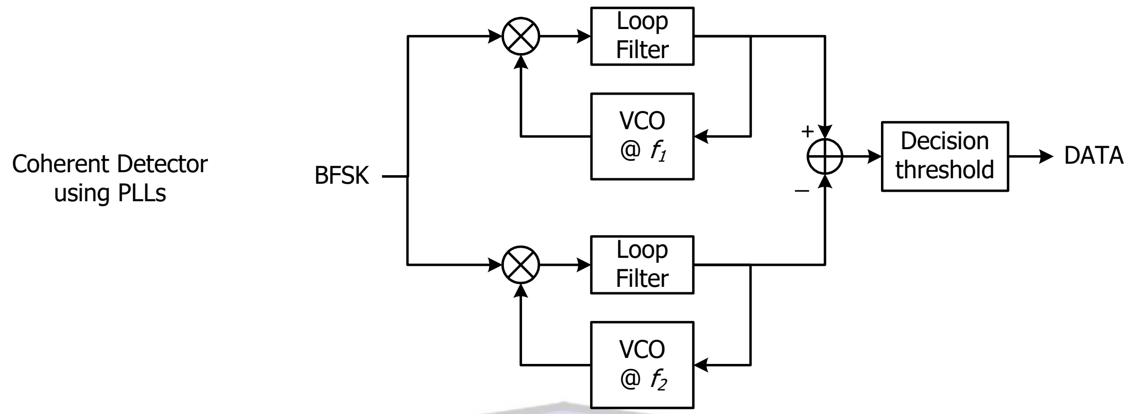
$$f_c = \frac{f_1 + f_2}{2}$$

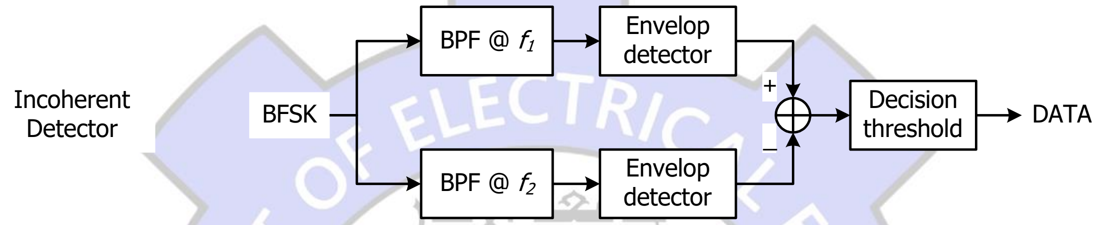
And the BFSK frequency deviation is:

$$\Delta_f = \frac{f_2 - f_1}{2}$$

To define the bandwidth of a BFSK signal, we'll consider the bandwidth ends at the first zero crossing point in the BFSK spectrum. So, $B = 2\Delta_f + 2R$, where $R = 1/T_0$ = the baud rate of the baseband data stream, which is assumed here to be the nominal bandwidth of the binary.

6.3.2 DEMODULATION





6.3.3 PROBABILITY OF ERROR IN BFSK

As $E_{AV} = \dfrac{(E_0 + E_1)}{2} = \dfrac{2E_1}{2} = E_1 = E_0$

$$P_E = \text{Erfc}\left(\sqrt{\frac{E_1}{\eta}}\right) = \text{Erfc}\left(\sqrt{\frac{E_{AV}}{\eta}}\right)$$

$$P_E = \text{Erfc}\left(\sqrt{\frac{S}{N}}\right) \qquad \text{Coherent Detection at } T_0B = 1$$

$$P_E \approx \frac{1}{2}\exp\left(\frac{-E_{AV}}{2\eta}\right) \qquad \text{Incoherent Detection}$$

It seems just like the BASK results, is there any difference?

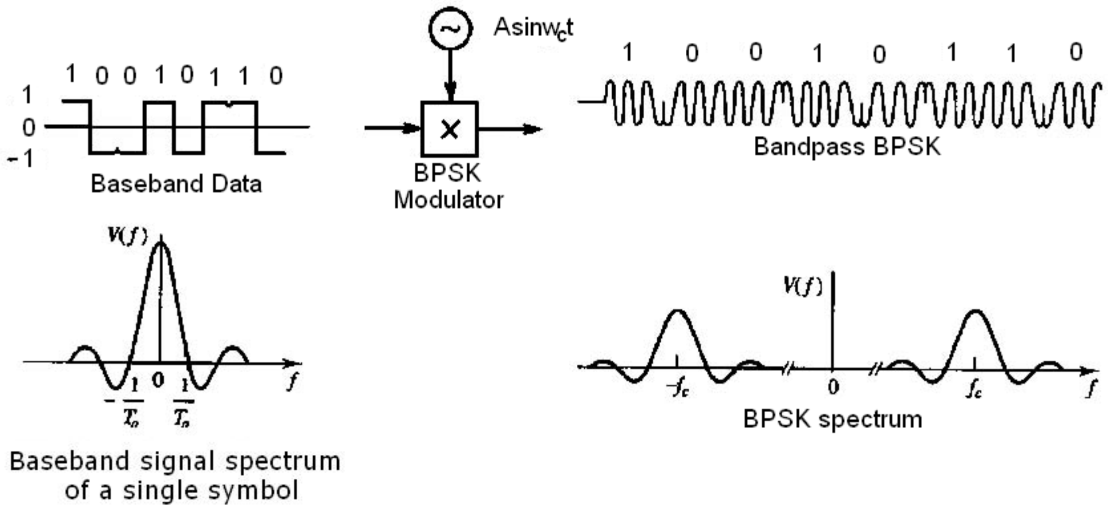
6.4 BINARY PHASE SHIFT KEYING (BPSK)

BPSK converts the baseband binary to passband by changing the carrier's phase in sympathy with the baseband digital data.

$$x(t) = \begin{cases} A\sin(\omega_c t) & \text{for logic 1} \\ A\sin(\omega_c t + \phi) & \text{for logic 0} \end{cases}$$

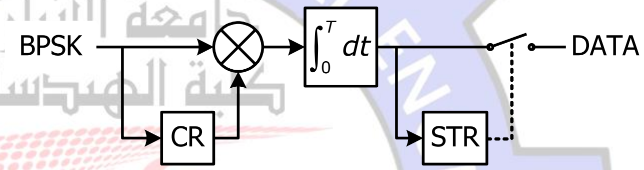
In BPSK, $\phi = 180°$.

6.4.1 GENERATION



6.4.2 DEMODULATION

Coherent Correlation Receiver



6.4.3 PROBABILITY OF ERROR IN BPSK

As $E_{AV} = E_1 = E_0$, then the Coherent Detection at $T_0B = 1$

$$P_E = \text{Erfc}\left(\sqrt{\frac{2E_1}{\eta}}\right) = \text{Erfc}\left(\sqrt{\frac{2E_{AV}}{\eta}}\right) = \text{Erfc}\left(\sqrt{\frac{2S}{N}}\right)$$

It gives the impression that the BPSK is the best!

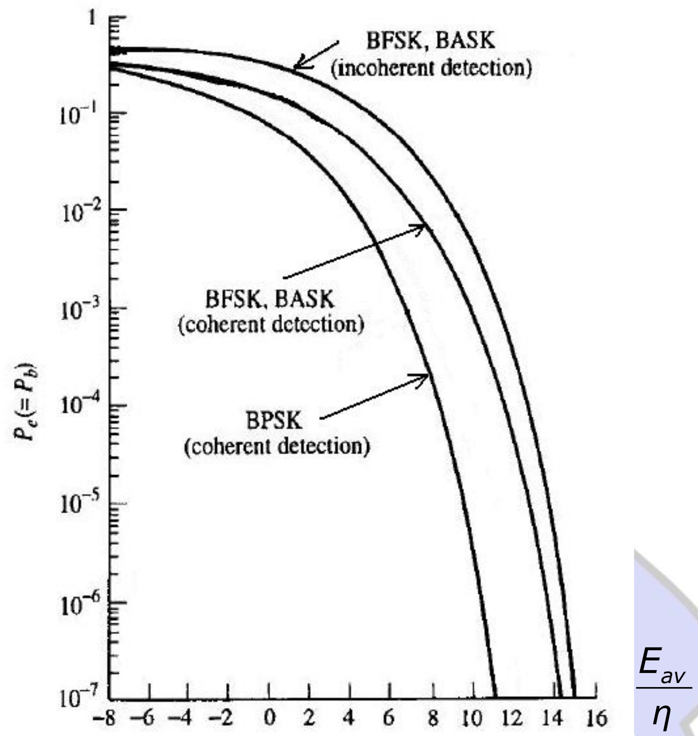
Where is the incoherent detection?

6.5 COMPARISON OF BINARY KEYING TECHNIQUES

6.5.1 THROUGH THE AVERAGE POWER

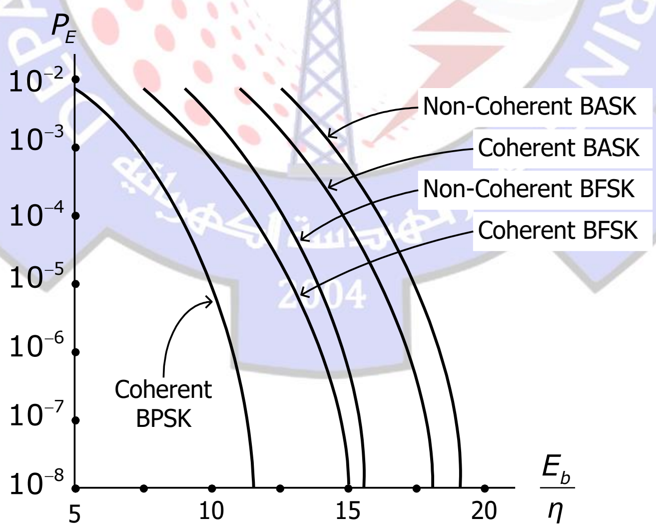
BASK & BFSK have equivalent E_{AV}, hence the same P_E. BPSK has better performance (less P_E).

As shown:



6.5.2 THROUGH THE PEAK POWER

To get the same P_E, BFSK requires double the power of BPSK but also requires half the power of BASK, as all the energy of BASK transmission is squeezed into only one type of data. As shown:



6.5.3 THROUGH THE SPECTRAL EFFICIENCY

If we define Γ as the spectral efficiency of data transmission as:

$$\Gamma = \frac{R_b}{B}\log_2 M \qquad \text{bits/sec/Hz}$$

M = number of levels $= 2^m$ (for binary $M = 2, m = 1$)

R_b = binary data rate (in bps) $= f_0 = \frac{1}{T_0}$

B = the required channel bandwidth of the transmitted signal.

Then the comparison will be:

	Baseband	BASK	BFSK	BPSK
Data Rate (bps)	f_0	f_0	f_0	f_0
Nominal Bandwidth	$\frac{f_0}{2}$	f_0	$2(\Delta_f + f_0)$	f_0
Nominal spectral efficiency (Γ)	2	1	$\frac{f_0}{2(\Delta_f + f_0)}$	1

6.5.4 Through Systems

ASK:

- The transmitters of ASK are very easy to build.
- They have the advantage of transmitting no power at 0. Such systems find some applications in short-range miniature telemetry systems.
- The receivers for non-coherent ASK are easy to build. The difference in performance between coherent & non-coherent detection is slight compared to the increase in complexity required. Therefore, coherent detection of ASK is generally not common.
- The decision threshold in the receiver must be adjusted with changes in the levels of the received signal. Therefore, it requires an Automatic Gain Control (AGC).

FSK:

- In contrast to ASK, the FSK systems operate symmetrically about a zero-decision threshold level, regardless of carrier signal strength.
- Transmitters of FSK is slightly more complex than those for ASK.
- Non-coherent FSK receiver is relatively an easy and a popular choice for low to medium data transmission rates.
- FSK requires more bandwidth than ASK & PSK.

PSK:

- The performance of PSK systems are superior to both ASK & FSK systems.
- PSK systems require less transmitted power for a given P_E.
- Synchronous detection is required, and carrier recovery systems are more difficult (therefore more expensive) to build.

The decision of which scheme of the digital modulation to be selected depends on the trade-off between (performance, cost, bandwidth…etc.). In addition (propagation, distortion, fading, non-white noise (e.g. interference), non-Gaussian noise…etc.) may affect the choice.

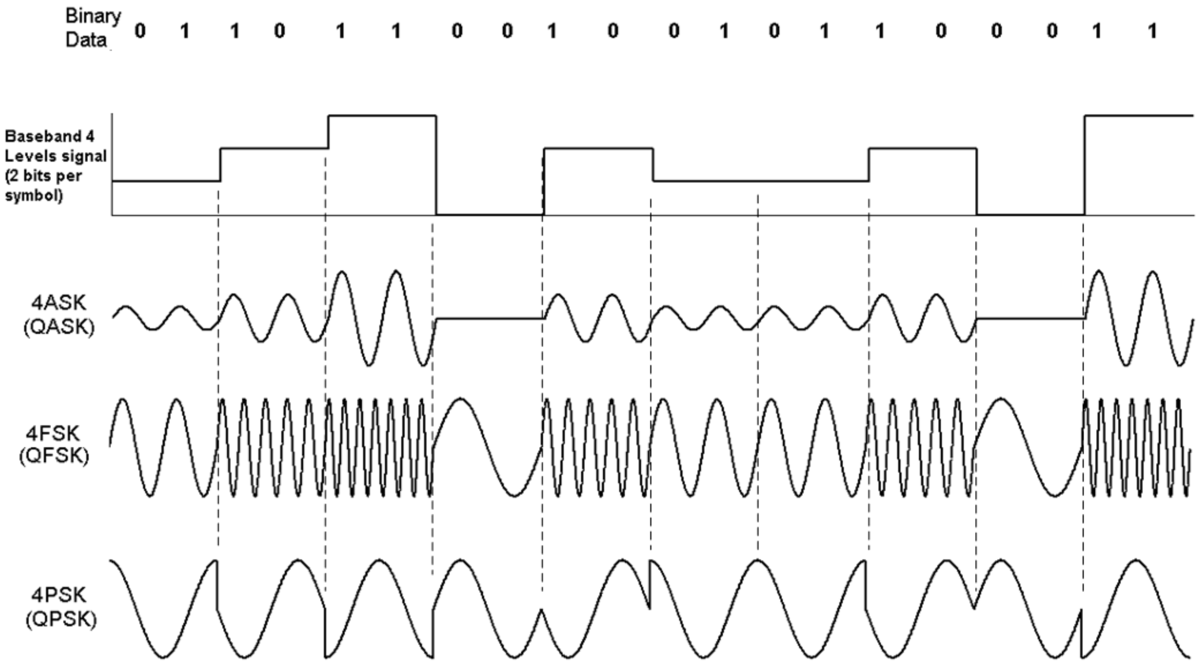
6.6 MODULATION TECHNIQUES WITH INCREASED SPECTRAL EFFICIENCY

In addition to the reliability and the low P_E, the efficient usage of the bandwidth is an important aspect in the design of digital communication systems.

As we know, the best design is to transmit the maximum information rate through the minimum possible bandwidth. We defined earlier the spectral efficiency Γ. To make Γ as large as possible, we can:

- Decrease B by filtering the transmitted signal prior to the transmission. *How far is this method useful?*
- Increase the number of transmitted bits per second (R), but we must consider the Shannon-Hartley limit of channel capacity. So, to avoid ISI, $T_0 B \geq 0.5$ for baseband signals, and $T_0 B \geq 1$ for Bandpass signals.
- The final option: increase M (i.e. increase the number of bits per a transmitted symbol). Practical systems currently exist with M of 4, 8, 16, … 1024.

The multi-level per symbol (*M-symbol*) signaling for the studied digital modulations will referred as: MASK, MFSK and MPSK.



In practice, only MPSK and QAM are used, because:

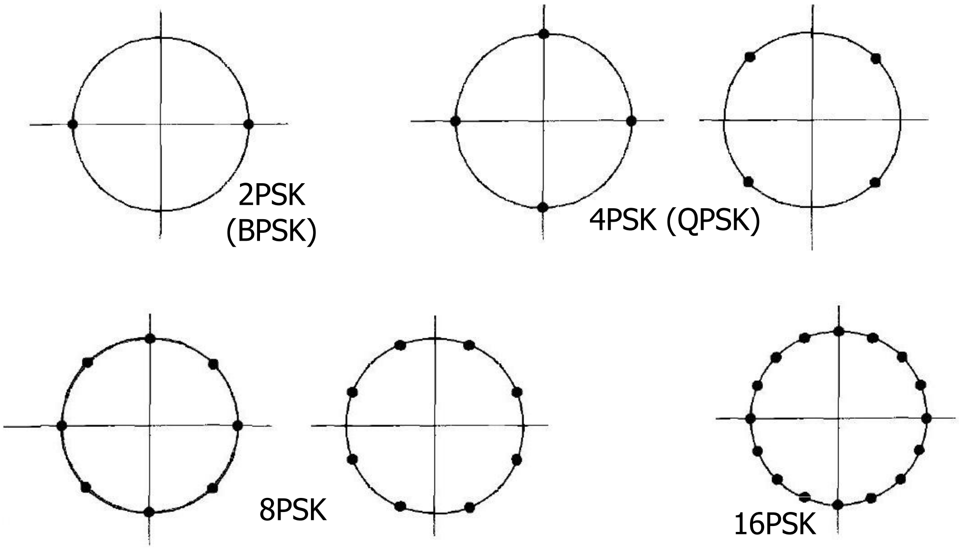
- the increase of M in MASK results in level ambiguity at detection. The channel effects like the noise and distortion make the process of the detection very hard. Nevertheless, increasing M, undoubtedly, increases P_E. But it is still acceptable for small M through good channels.
- as we studied before, the increase of M in MFSK requires larger bandwidth to accommodate the modulated signal, which decreases Γ.

6.6.1 M-Symbol Phase Shift Keying (MPSK)

MPSK implies the extension of the number of the allowed phasor states from 2 to 4, 8, 16……2^m. i.e. while the carrier amplitude being constant, the phase is changing according to the input taking one of the designed phase state.

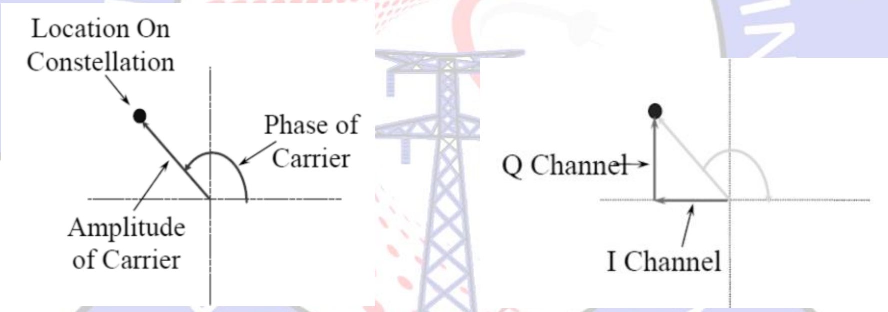
$$\text{MPSK} = \cos(\omega_c t + \theta_k) \quad \text{where } k = 0,1,2,\dots M-1 \quad , \quad \theta_k = \frac{2\pi k}{M} \quad \text{or} \quad \theta_k = \frac{(2k+1)\pi}{M}$$

Some examples on MPSK is shown in the following figure:



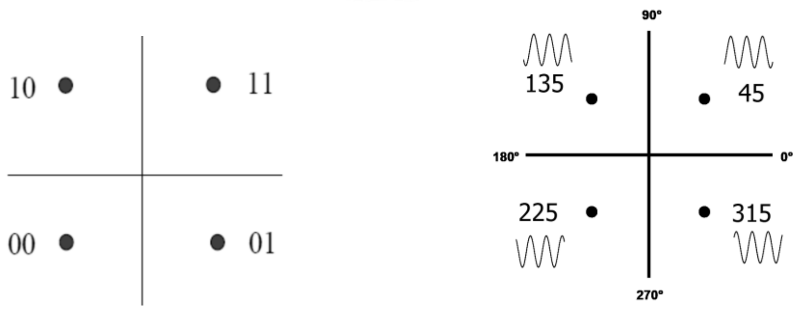
Quadrature Modulation

The phase and amplitude of the carrier at any given time determine the location on the Constellation. The amplitudes of the I and Q channels are derived from the rectangular coordinates of the carrier's amplitude and phase.



Quadri-phase-Shift Keying (QPSK)

Here, we study the coherent QPSK as an example of MPSK. In QPSK, as with BPSK, information carried by the transmitted signal is contained in the phase. The phase of the carrier takes on one of four equally spaced values, as:



For this set of values, we may define the transmitted signal as:

$$s(t) = \sqrt{2}\cos\left[2\pi f_c t + \frac{(2i-1)\pi}{4}\right] \qquad \text{where } i = 1, 2, 3, 4$$

$$= \sqrt{2}\cos\left[\frac{(2i-1)\pi}{4}\right]\cos\omega_c t - \sqrt{2}\sin\left[\frac{(2i-1)\pi}{4}\right]\sin\omega_c t$$

Based on this representation, we can make the following observations

- There are two orthogonal basis functions, defined by a pair of quadrature carriers:

$$\phi_x = \cos\omega_c t \quad and \quad \phi_y = \sin\omega_c t$$

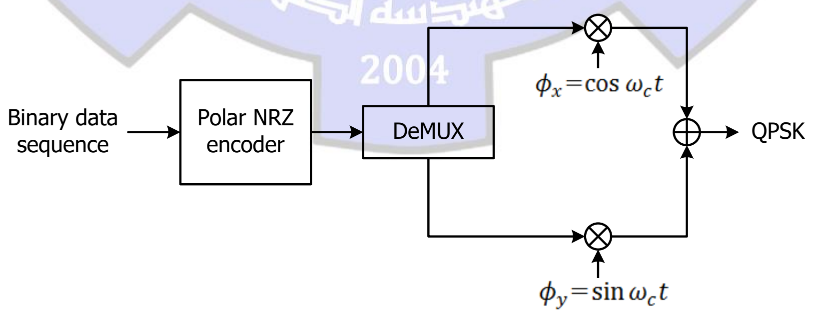
- There are four message points, and the associated signal vectors are defined by:

$$s = \begin{bmatrix} s_x \\ s_y \end{bmatrix} = \begin{bmatrix} \sqrt{2}\cos\left[\frac{(2i-1)\pi}{4}\right] \\ -\sqrt{2}\sin\left[\frac{(2i-1)\pi}{4}\right] \end{bmatrix} \text{where } i = 1, 2, 3, 4$$

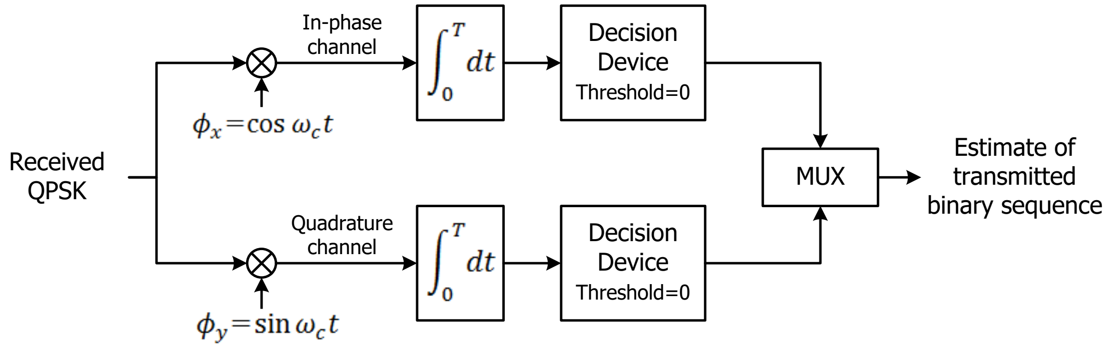
The elements of the signal vectors have their values summarized in the table below:

i	Gray-encoded input Di-bits	Phase of QPSK signal (rad)	Coordinates of Message points	
			s_x	s_y
1	10	$\pi/4$	$+1$	$+1$
2	00	$3\pi/4$	-1	$+1$
3	01	$5\pi/4$	-1	-1
4	11	$7\pi/4$	$+1$	-1

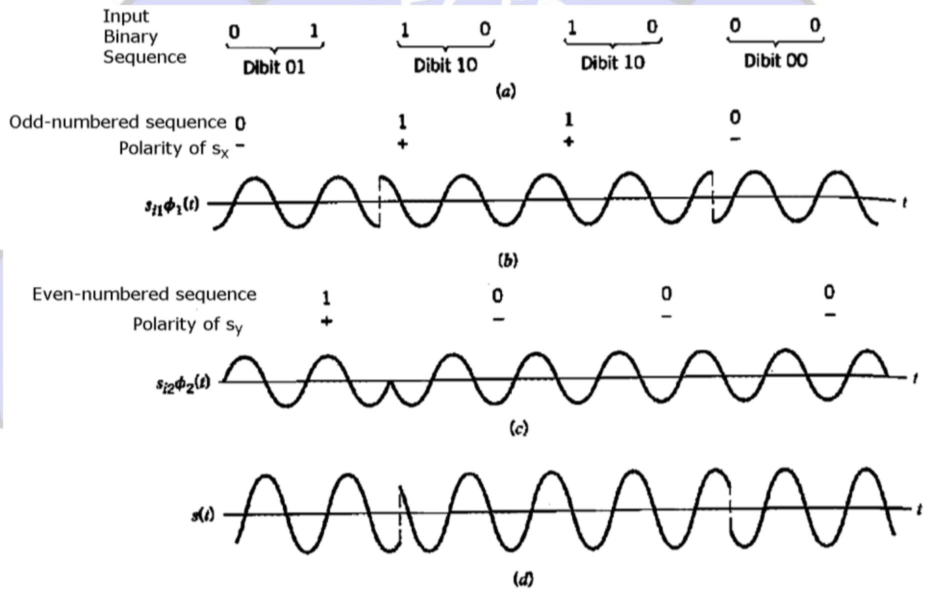
Modulator *



Demodulator *



The figure below illustrates the sequences and waveforms involved in the generation of a QPSK signal



Probability of Error in MPSK

As Γ increases with *M*, the probability of error would also be increased. So, the *symbol* probability of error is:

$$P_E \approx 2\,\text{Erfc}\left(\sqrt{\frac{2E_{AV}}{\eta}}\sin\frac{\pi}{M}\right) \quad \& \quad P_E \approx 2\,\text{Erfc}\left(\sqrt{T_0 B\frac{2S}{N}}\sin\frac{\pi}{M}\right)$$

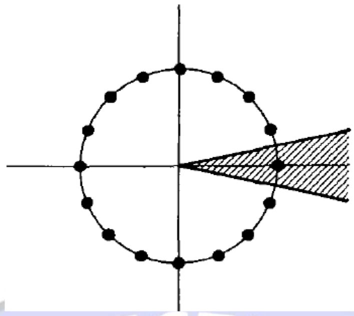
Note: these formulas are for $M \geq 4$ only, and if we use M=2, we'll get twice the correct result of BPSK.

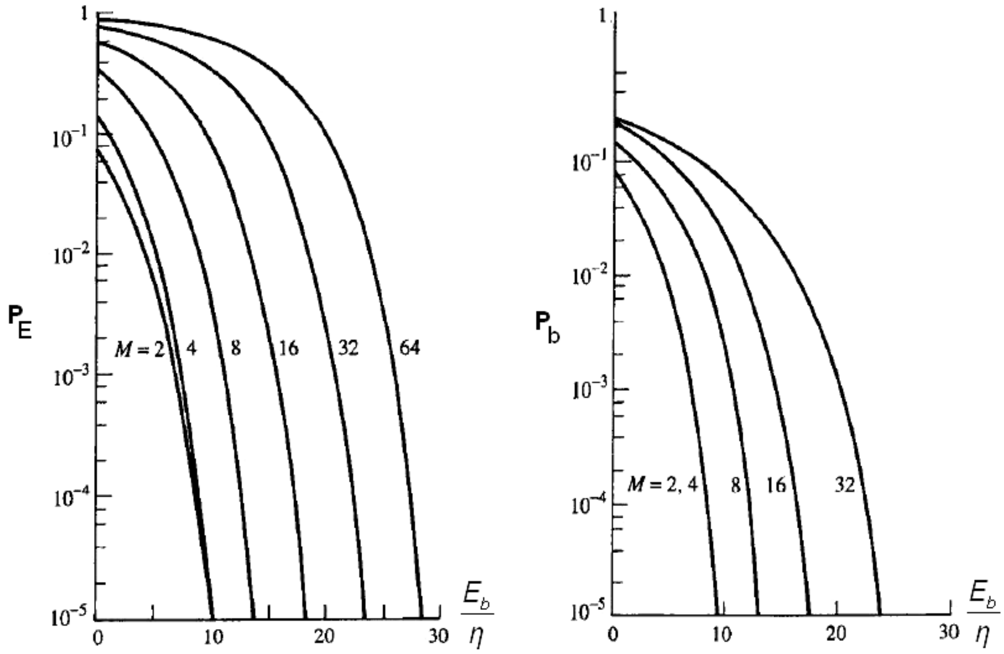
Now, if we use gray code to map binary symbols to phasor state, the probability of *bit* error is:

$$P_b = \frac{P_E}{\log_2 M}$$

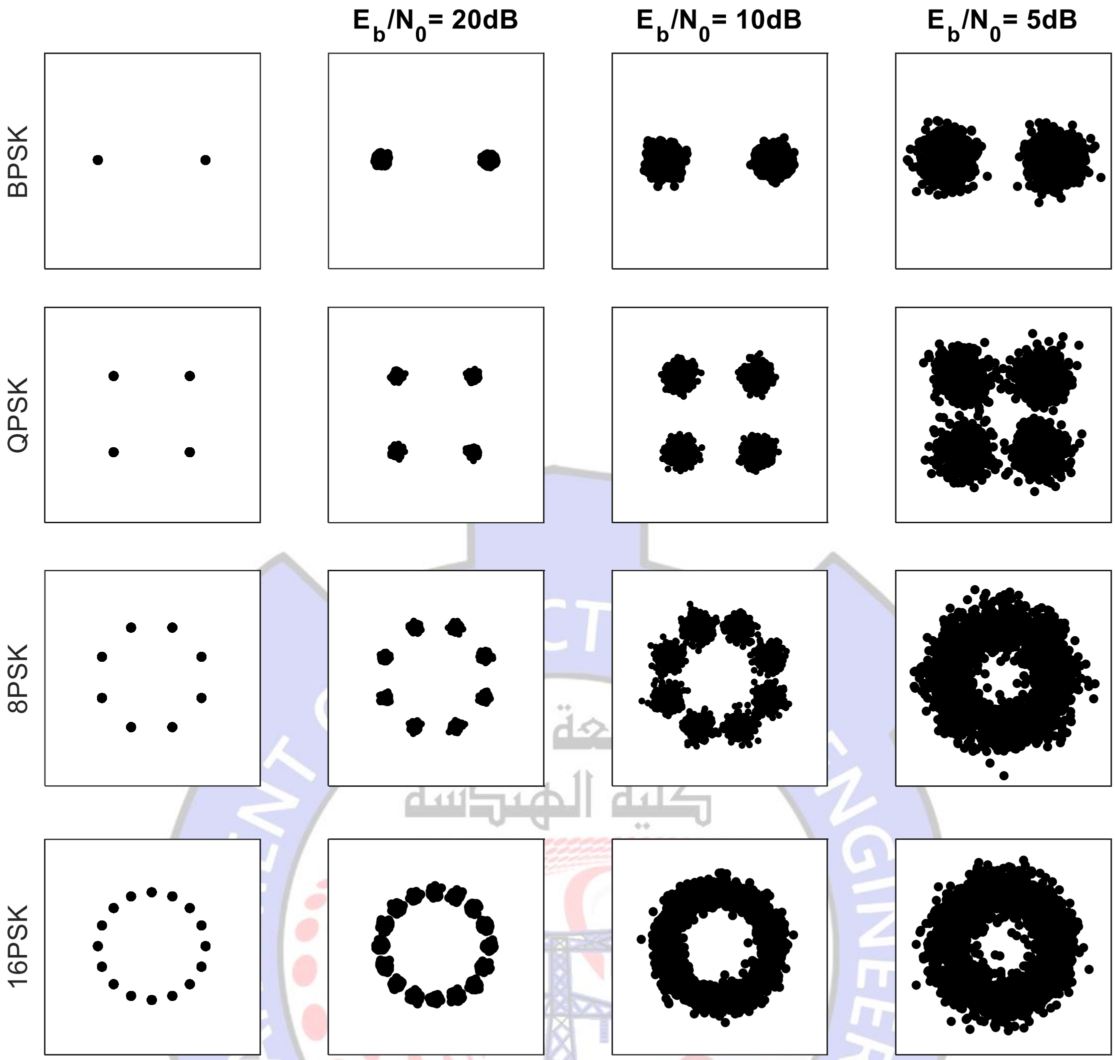
To compare the performance of MPSK through M, we should express P_E in terms of E_b, where:

$$E_b = \frac{E_{AV}}{\log_2 M}$$





For 32PSK, 64PSK and higher M values, it will become so difficult (or sometimes impossible) to distinguish between close phases.



Bandwidth of MPSK Signals

The power spectra of MPSK signals possess a main lobe bounded by well-defined spectral nulls (i.e., frequencies at which the power spectral density is zero). Accordingly, the spectral width of the main lobe provides a simple and popular measure for the bandwidth (null-to-null bandwidth).

So, the channel bandwidth required to pass MPSK signals (more precisely, the main spectral lobe of M-ary signals) is given by:

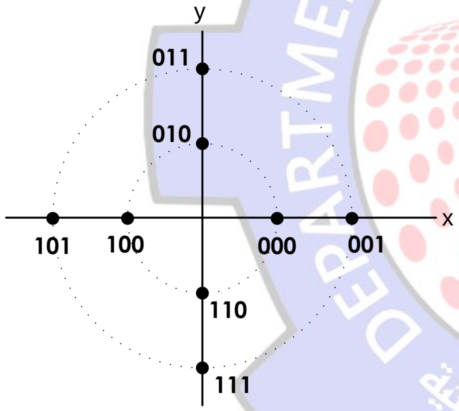
$$B = R_s = \frac{R_b}{\log_2 M}$$

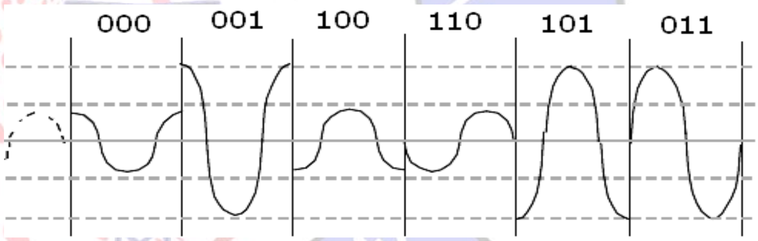
Where R_s is the symbol rate (bauds per second), R_b is the binary data rate (bits per second).

6.6.2 HYBRID AMPLITUDE/PHASE MODULATION (QAM)

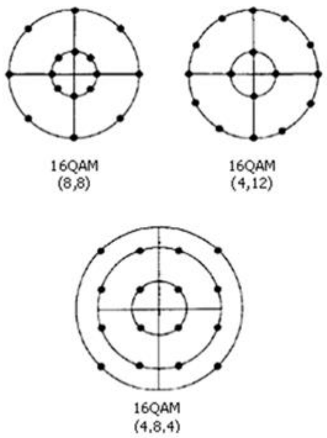
Also called Quadrature Amplitude Modulation. Here, it is possible to introduce amplitude as well as phase modulation to give an improved distribution of signal state in the phasor diagram. The following illustrates the 8QAM:

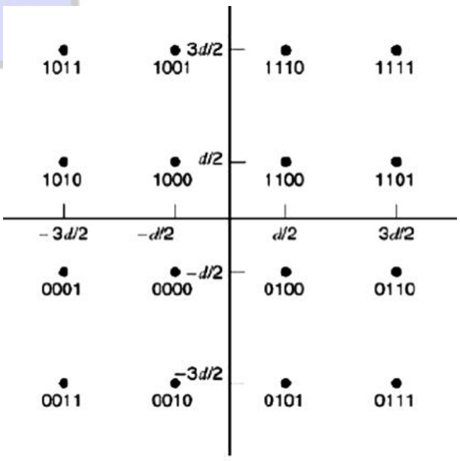
Bit Combination	Phase Shift (Deg.)	Amplitude
000	0	Low
001	0	High
010	90	Low
011	90	High
100	180	Low
101	180	High
110	270	Low
111	270	High

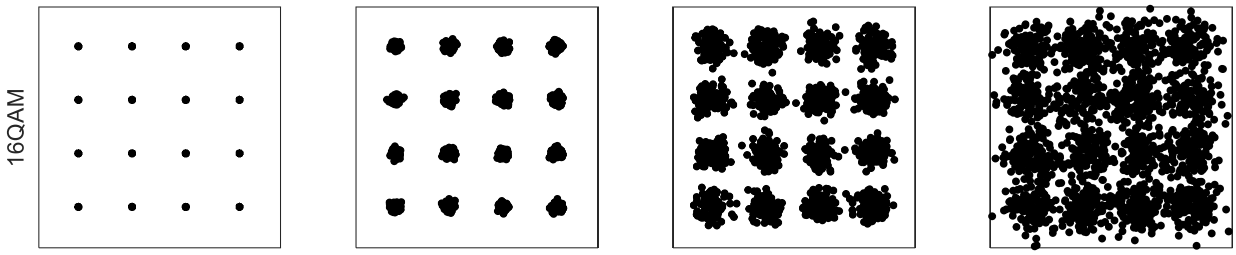




The following figures illustrate the re-distribution of 16PSK to create 16QAM:



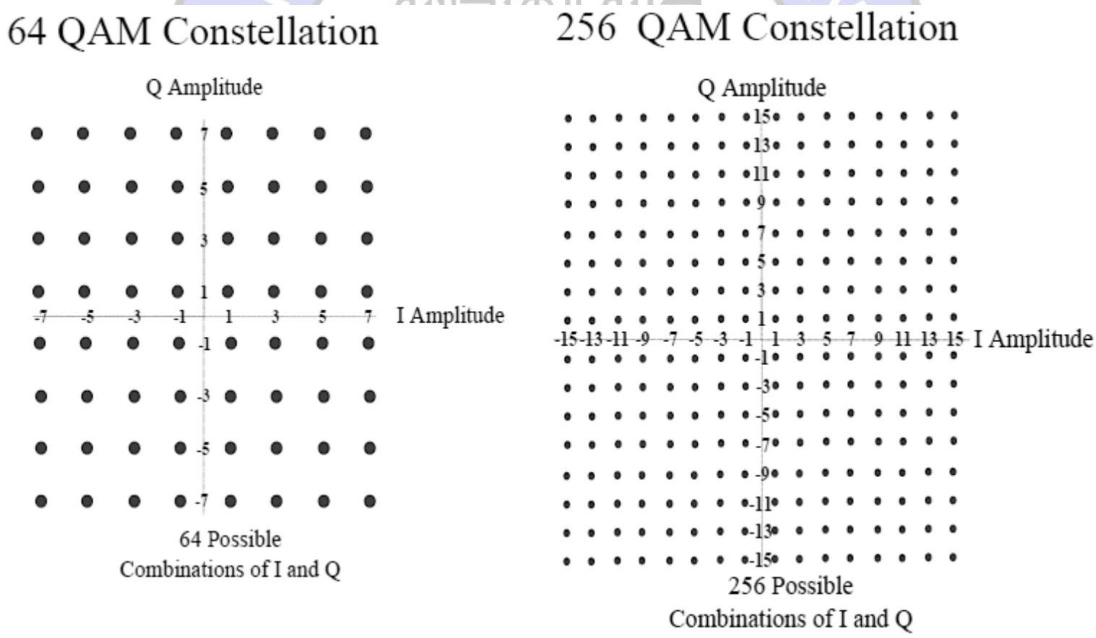




64 and 256 QAM Constellations

By adding more levels to the I and Q channels, higher data rates can be carried.

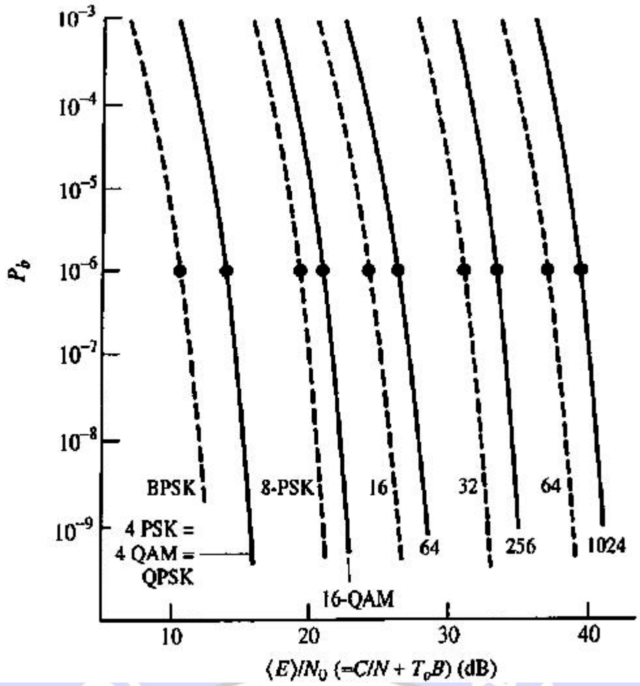
- The higher the number of levels, the more effect there will be from noise or interference.
- 64 QAM uses 8 levels in the I direction and 8 levels in the Q direction for a total of 8 squared or 64 symbols.
- 256 QAM uses 16 levels in the I direction and 16 levels in the Q direction for a total of 16 squared or 256 symbols.



Probability of Error in MQSK*

$$P_E \approx 4\left(1-\frac{1}{\sqrt{M}}\right)\text{Erfc}\left(\sqrt{\frac{3E_{AV}}{(M-1)\eta}}\right)$$

This figure shows clearly that the MQAM is much better than the MPSK for the same M.


P_b
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
BPSK
8-PSK
16
32
64
4 PSK =
4 QAM =
QPSK
64
256
1024
16-QAM
10
20
30
40
$\langle E \rangle / N_0$ ($=C/N + T_pB$) (dB)