COS 242 - DATA COMMUICATIONS AND NETWORKING (2 Units)

COURSE CONTENTS:

Introduction, data and information, data processing system, data communication systems, signals: Introduction to analog and digital, A/D conversions, Time and frequency domain concepts, Fourier transform, Fourier series, measure of communication, channel characteristics, Nyquest and Shannon information capacity, transmission media, noise and distortions, modulation and demodulation, Multiplexing and demultiplexing, synchronous and asynchronous data transmission, Error detection and control techniques.

INTRODUCTION

OVERVIEW OF DATA AND INFORMATION

In everyday usage, the two terms: **information** and **data** are always used interchangeably. There is, however, a difference between the two in computer studies. The digits "19896206" constitute data, but they convey no information. They could be interpreted as a student registration number or as a library catalogue number.

Data, therefore is the name given to basic facts such as names and numbers. Examples are times, dates, weights, prices, costs, numbers of items sold, employee names, product names, addresses, test scores etc. Some sources of data may include:
Questionnaire survey, Face – to face interview, Published document review, Telephone interview, Experimental recordings, Group techniques such as workshops, focus/ discussion groups, Public opinion pools, Library or database.

On the other hand **information** is data which has been converted into a more useful or intelligible form. Examples are printed documents, pay slips, text, reports, etc. So a set of words would be data but text would be information e.g. Obinna, Okafor are data. "Obinna Okafor scored the highest examination mark" is information.
Different types of questions/problems require different sources of information. Some sources of information include: Observations, people, speeches, documents such as, pictures, organizations, websites, etc. Let us take a look at the differences between data and information as shown in the Table below.

Differences between data and information

Data	**Information**
Data refers to names given to basic facts such as names, numbers. Examples are dates, addresses, times, prices, costs, etc	Information is data, which has been converted into a more meaningful form.
Data is facts and statistics that are collected in raw form.	Information is processed data that make meaning to someone.
Data is normally arranged into numbers, blocks and charts.	Information is normally represented in prose form.
Data is the lowest level of knowledge.	Information is the second level of knowledge.
Data by itself is not significant without processing.	Information is by itself significant.
Observation and recording is done to obtain data.	Analysis is done to obtain information.

Note that we use the term "communication" for transfer of information and the term "transmission" for the transfer of data.

DATA PROCESSING SYSTEM

Data processing (DP) is the term given to the process of collecting data together and converting them into information. The objective of data processing, therefore, is to organize data into meaningful information.

A data processing system is made up of people, equipment and procedures that process data. There are basically five steps by which data is expressed, processed and returned to managers or other individuals as update and useful information. These are:

Step1: Preparation of Source Documents;
Step2: Input of Data;
Step3: Manipulation of Data;
Step4: Output of Information
Step5: Storage of Data.

Data Transmission and Processing Methods

In the last twenty years, there have been many developments that have increased the processing speed of computer system. The speeds of CPU and I/O devices have increased tremendously because of latest technological developments; direct access devices have also been developed and improved. In a data processing cycle, source data must be collected before process, and processed data must be distributed before it can be used. Thus, even if processing time is reduced, there can still be significant delays in the system due to distribution. Data communication systems are designed to reduce the delays in collection and distribution of data.

What is Data Communication?

Data communications means the exchange of data between two devices via some form of transmission medium such as a wire cable.

For data communications to occur, the communicating devices must be part of a communication system made up of a combination of hardware (physical equipment) and software (programs).

Characteristics of Data Communications:

The effectiveness of a data communications system depends on four fundamental characteristics: delivery, accuracy, timeliness, and jitter.

1. Delivery: The system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.

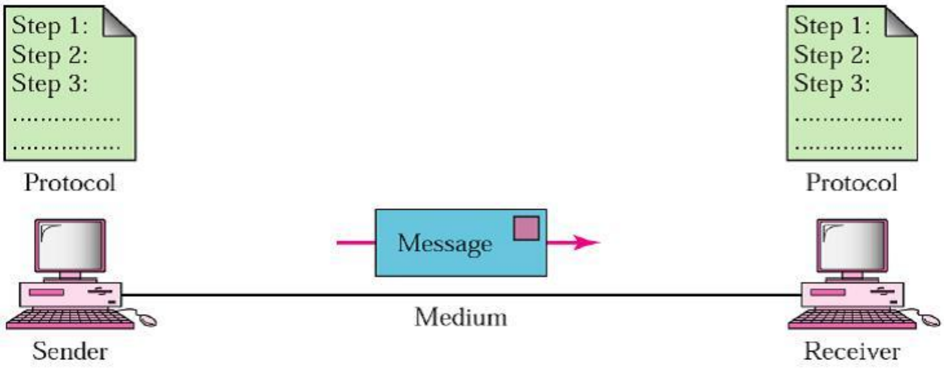
2. Accuracy: The system must deliver the data accurately. Data that have been altered in transmission and left uncorrected are unusable.

3. Timeliness: The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called real-time transmission.

4. Jitter: Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets.

Components of Data Communication

The different components of Data communication are shown in the following figure.



1. Message:

The message is the information (data) to be communicated. Popular forms of information include text, numbers, pictures, audio, and video.

2. Sender:

The sender is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.

3. Receiver:

The receiver is the device that receives the message. It can be a computer, workstation, telephone handset, television, and so on.

4. Transmission medium:

The transmission medium is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves.

5. Protocol:

A protocol is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not communicating, just as a person speaking French cannot be understood by a person who speaks only Japanese.

What is Network?

A network is a set of devices (often referred to as nodes) connected by communication links. A node can be a computer, printer, or any other device capable of sending and/or receiving data generated by other nodes on the network.

Characteristics of Network?

A network must be able to meet a certain number of criteria. The most important of these are performance, reliability, and security.

Performance:

Performance can be measured in many ways, including **transit time** and **response time**.

Transit time is the amount of time required for a message to travel from one device to another.

Response time is the elapsed time between an inquiry and a response.

The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software.

Reliability

In addition to accuracy of delivery, network reliability is measured by the frequency of failure, the time it takes a link to recover from a failure, and the network's robustness in a catastrophe.

Security

Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

TYPES OF DATA COMMUNICATION SYSTEMS

Although there are many forms of data-communication systems, they can be broken into four categories:

1) Offline;
2) Online batch-processing;
3) Online real-time and
4) Time-sharing system.

Let us examine each in turn:

1) OFFLINE SYSTEMS

Offline means that the transmission of data is not directly to or from the computer. Essentially, the system consists of terminal and communication lines: one terminal at the sending end and another at the receiving end. The terminals could be card reader at the sending end and card punch, tape drives or printers at the receiving end.

Offline system is simply a means of eliminating the delays in sending data between two geographical points.

For example, if a report is to be printed for branch office management, data can be read by tape drive in the home office and printed on a printed in the branch office management.

2) **ONLINE BATCH-PROCESSING SYSTEMS**

Online means that data is transmitted directly to the computer. Batch-processing is an important method of manual and computerized data processing characterized by the following:

a) The accumulation of transactions into batches;
b) There is some degree of delay;
c) The transactions are sorted and then processed;
d) The results of processing a particular data item are not known until the whole batch as been processed.

Examples of this system are: a punch tape is prepared from customer orders and used as input to the computer through phone lines. The computer performs editing run while writing the transactions on the magnetic tape. After the transactions from all the branches are edited and written, the tapes can be merged and processed.

Advantages of Batch Processing

1) There is no need for the user to be present since there is no interaction between him and the computer while the job is being run.
2) Preparing the work and operating the computer are done by trained people and not by users.
3) Less expensive than time sharing on large systems.

Disadvantages of Batch Processing

1) There is always a delay before work is processed and returned.
2) The user cannot take action if anything is wrong; he has to re-input the job and place it on a different batch.
3) Batching usually involves an expensive computer and a large number of staff.

3) **ONLINE REAL-TIME SYSTEMS**

A real time system is a computer system which is capable of processing data without significant delay, so that the results are available to influence the activity currently taking place.

In practical terms we could say that a real-time system is a batch processing system with a single transaction which is processed immediately on demand.

Although real-time system can be used for basic business applications such as order writing, inventory and payroll, they are more likely to be used for specialized applications as in:

a) Transport- Airline booking systems;
b) Data retrieval systems used by the police, fire services and hospitals;
c) Systems used in the control of processes such as automated process and production control in factories;

In the Airline booking system for instance, handling customer's inquiries can be a major problem for the booking department. Here, records of seat availability on all its planes will be kept by an airline on a central computer. The computer is linked through terminals to a world-wide system of agents. Each agent can gain access to the flight records and within seconds make a reservation in respect of a particular flight. This reservation is recorded immediately so the next enquiry for that flight finds that particular seat or seats reserved. Notice the computer records reflect on accurate picture of the airline's seating load at all times because there is no time lag worth mentioning. The computer would then output information for the production of the customers' ticket and flight instructions, on confirmation of the booking by the customer.

Advantages of Real-Time Systems

1.) Fast response

Disadvantages

1.) The cost of a real-time system is higher than that of batch system.
2.) A computer being used for a real-time application often cannot be used for anything else.
3) Real-time system involves hardware programming and systems design costs that have no equivalent in batch systems.

4) **TIME SHARING SYSTEMS**

A time sharing system is one in which the processor time of a central computer is divided into time slices which are shared between various interactive users appropriately. In such a system, a clock is used to divide up processor time into "time slices" and these time-slices are shared between various users in any appropriate way.

At each pulse of the clock the user program currently being executed has its execution suspended to make way for next user program. The operating system continually checks the status of each user, in strict sequence, to see whether the user requires processor time.

When it is confirmed that a user requires processor time that user's program is allowed to continue execution for the duration of the time slice.

At any given time the user may have access to the computer and because the computer switches from one user terminal to the other at a fantastic speed, each person using it may not realize that another person is using it at the same time. The number of users of any one system is limited, to ensure a high speed of response. Each has a typewriter terminal or VDU installed and linked to the computer system by telephone line. Each user has his own password and this has to be transmitted when access is required.

The usual dialog involves required for a time sharing system takes the following steps:

i) User dials the computer system telephone number.
ii) System asks for the password (or code number)
iii) User keys in his code number and is connected ("signed on") to the computer.
iv) Now by means of the keyboard he will get the system to do whatever he requires.
v) To do this the computer allocates the incoming user an area on disk, does the necessary processing in the CPU and sends back the results through telephone line to the users terminal.

vi) At the end of his “session” the user “signs off” and replaces the telephone receiver.

WAVES/SIGNALS

A wave is a disturbance, which travels through a medium and transfers energy from one point to another without causing any permanent displacement of the medium itself.

For example, waves could be observed on the surface of water if a stone is dropped into a pond of water. The water transfers energy (information) from one point to another.

TYPES OF WAVES

There are two major types of waves namely:

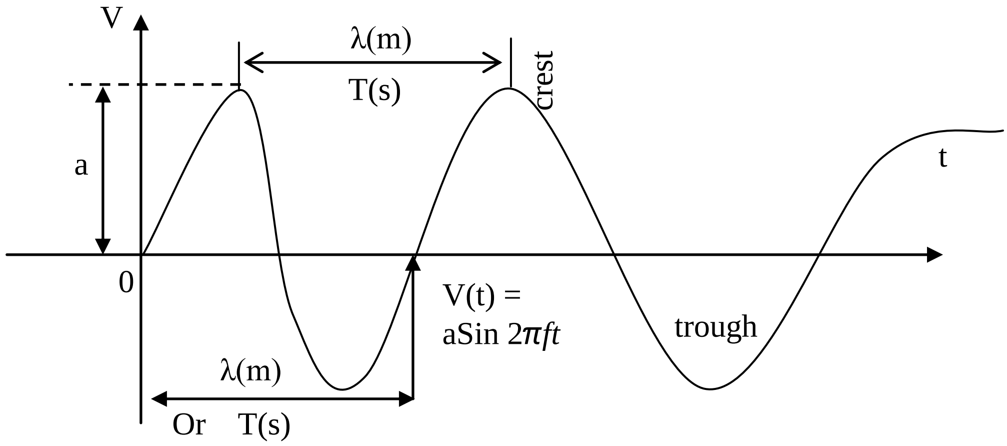
1) Mechanical and
2) Electromagnetic waves.

Mechanical waves usually require a material medium for their propagation e.g. water waves, sound waves etc.

On the other hand, Electromagnetic waves do not require a material medium for propagation. Examples are radio waves, light rays etc. These are also called microwaves.

Representation of waves

Waves whether mechanical or electromagnetic may generally be described as ripples, which alternate between positive and negative values in a sinusoidal manner as shown in figure 1 (below).



Definitions:

1) **Amplitude (a):** As the wave propagates, the particles of the medium vibrate about a mean position. The maximum displacement of particles from their mean position is called the amplitude, *a*, of the wave. It is measured in metres. **Peak amplitude** is the maximum value or strength of the signal over time.
2) **Period (T):** Is the time required for a particle to complete one cycle of the wave. Period (T) is also the amount of time it takes for one repetition; $T = \frac{1}{f}$. It is measured in seconds.
3) **Frequency (*f*)**: The number of cycles which the wave completes in one second is called its frequency. Frequency is measured in hertz (Hz). Frequency- Is the rate (in cycles per sec) at which the signal repeats. (Note that 1 kilohertz (KHz) = 10^3 Hz and 1 MHz = 10^6 Hz).
4) **Wavelength (λ):** The distance along the x -axis between successive crests or successive troughs is called the **wavelength.** It is measured in metres.
5) **Wave Speed (v):** Is the distance which the wave travels in one second.

$\therefore V = \frac{x}{t} m/s$, where

x- Is the distance traveled in metres;

t- Is the time taken to cover the distance x;

6) **Phase** ((ϕ))**:** Is a measure of the relative position in time within a **single period** of a signal = $\frac{t}{T}$.

7) **Spectrum:** Is a range of frequencies that a signal contains $S(t) = ASin(2\pi ft + \phi)$.

RELATIONSHIP BETWEEN T, f, λ AND V

By definitions, T seconds is needed to complete in 1 cycle;

$\therefore$ 1sec ” ” ” ” $\frac{1}{T}$ cycle

$$\boxed{\text{Thus: } f = \frac{1}{T}}$$

By definitions:

$$\boxed{V = \frac{\lambda}{T} = f\lambda}$$

$$\boxed{\text{Thus: V} = f\lambda}$$

<u>Solved example:</u>

A Radio station broadcasts at a frequency of 200 KHz. If the speed of the wave is $3x10^8$ m/s, calculate:

a) the period;
b) the wavelength of the wave;

<u>Solution</u>

a) Frequency f = 200 KHz = $2x10^5$ Hz

T= 1/f = 1/ ($2*10^5$) = <u>$0.5 * 10^{-5}$ sec</u>

b) Wavelength **λ** = v/f = $\frac{3x10^8}{2x10^5}$ =

λ = <u>$1.5*10^3$m</u>

Fundamentals of Data and Signals

The major function of the physical layer is to move data in the form of electromagnetic signals across a transmission medium. Whether the data may be numerical statistics from another computer, sending animated pictures from a design workstation, or causing a bell to ring at a distant control center, you are working with the transmission of data across network connections.

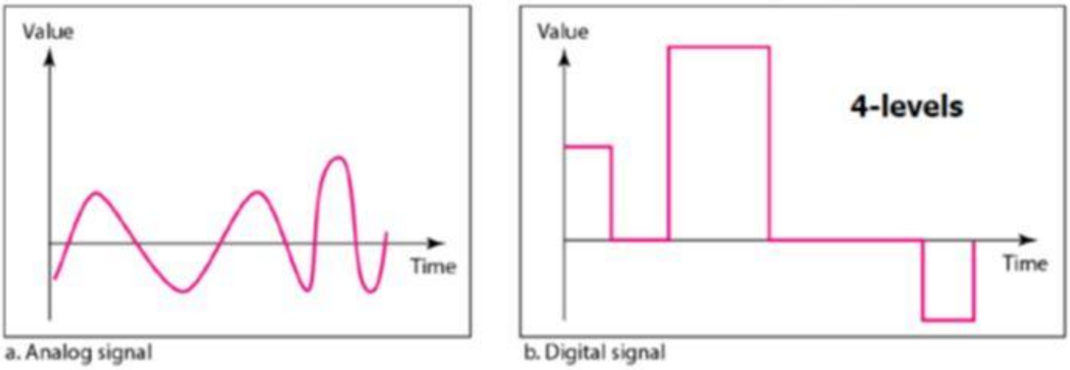
Analog and Digital Data

Data can be analog or digital. The term analog data refers to information that is continuous and take continuous values. Digital data refers to information that has discrete states and take discrete values. For example, an analog clock that has hour, minute, and second hands gives information in a continuous form, the movements of the hands are continuous. On the other hand, a digital clock that reports the hours and the minutes will change suddenly from 8:05 to 8:06.

Analog and Digital Signals:

An analog signal has infinitely many levels of intensity over a period of time. As the wave moves from value A to value B, it passes through and includes an infinite number of values along its path. A digital signal, on the other hand, can have only a limited number of defined values. Although each value can be any number, it is often as simple as 1 and 0.

The following program illustrates an analog signal and a digital signal. The curve representing the analog signal passes through an infinite number of points. The vertical lines of the digital signal, however, demonstrate the sudden jump that the signal makes from value to value.



Periodic and Non-periodic Signals:

Both analog and digital signals can be periodic or non-periodic

Periodic Signal: A periodic signal completes a pattern within a measurable time frame, called a period, and repeats that pattern over subsequent identical periods. The completion of one full pattern is called a cycle.

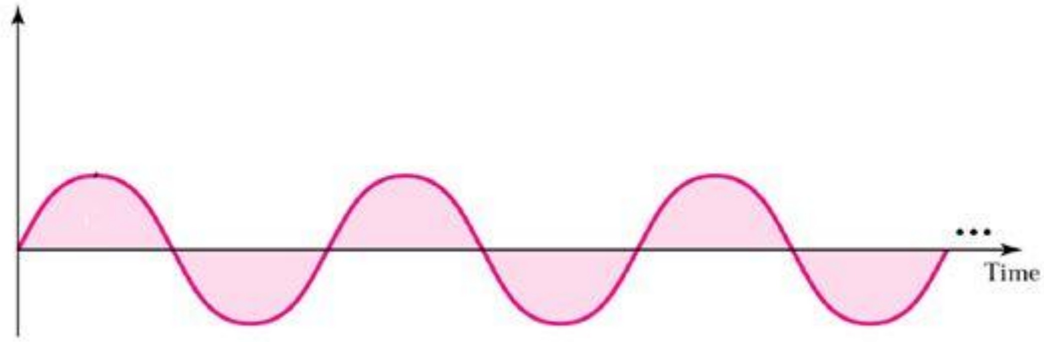
Non-periodic signal: A non-periodic signal changes without exhibiting a pattern or cycle that repeats over time.

In data communications we commonly use periodic analog signals (because they need less bandwidth) and non-periodic digital signals (because they can represent variation in data).

Periodic Analog Signal:

Periodic analog signals can be classified as simple or composite. A simple periodic analog signal, a sine wave, cannot be decomposed into simpler signals. A composite periodic analog signal is composed of multiple sine waves.

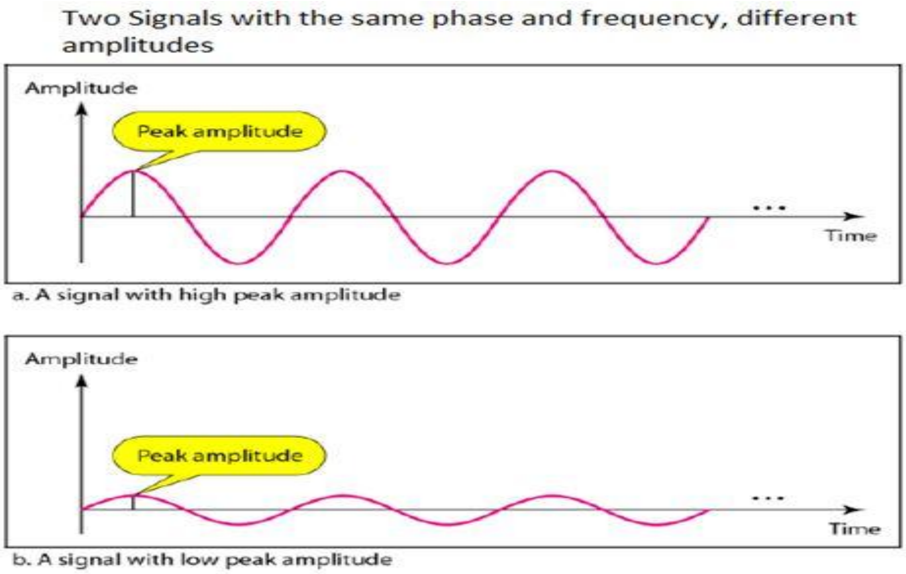
The sine wave is the most fundamental form of a periodic analog signal. When we visualize it as a simple oscillating curve, its change over the course of a cycle is smooth and consistent, a continuous, rolling flow. The following figure shows a sine wave. Each cycle consists of a single arc above the time axis followed by a single arc below it.



A sine wave can be represented by three parameters: the peak amplitude, the frequency, and the phase.

Peak Amplitude:

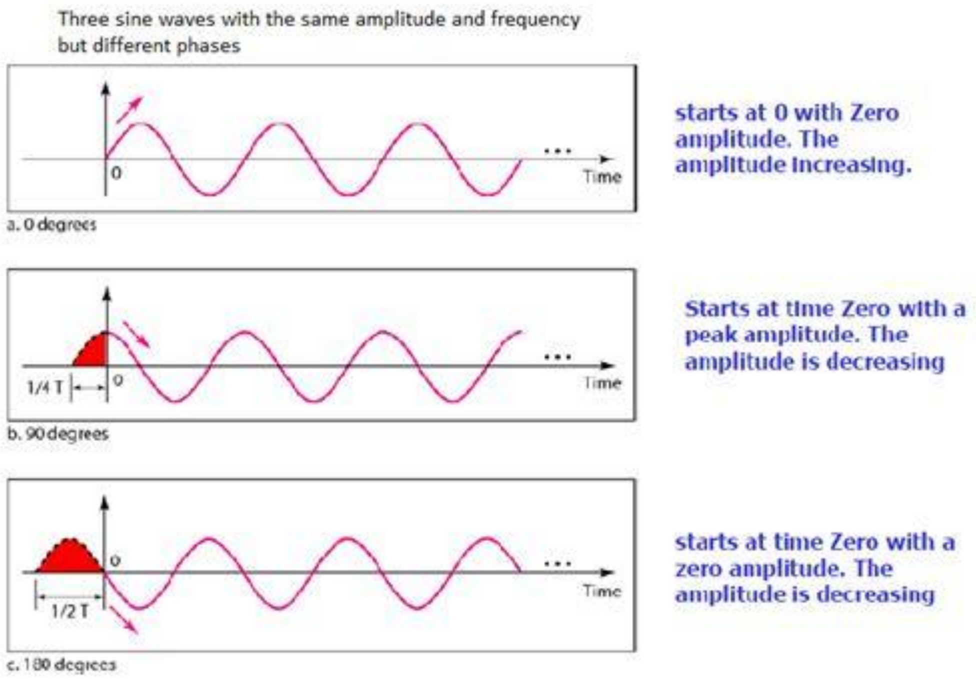
The peak amplitude of a signal is the absolute value of its highest intensity, proportional to the energy it carries. For electric signals, peak amplitude is normally measured in volts. The following Figure shows two signals and their peak amplitudes.



Period and Frequency: Period refers to the amount of time, in seconds, a signal needs to complete 1 cycle. Frequency refers to the number of periods in I s. Note that period and frequency are just one characteristic defined in two ways. Period is the inverse of frequency, and frequency is the inverse of period, as the following formulas show.

$$f= 1/T \text{ and } t= 1/F$$

Phase: The term phase describes the position of the waveform relative to time O. If we think of the wave as something that can be shifted backward or forward along the time axis, phase describes the amount of that shift. It indicates the status of the first cycle.



Wavelength:

Wavelength is another characteristic of a signal traveling through a transmission medium. Wavelength binds the period or the frequency of a simple sine wave to the propagation speed of the medium. While the frequency of a signal is independent of the medium, the wavelength depends on both the frequency and the medium. Wavelength is a property of any type of signal. In data communications, we often use wavelength to describe the transmission of light in an optical fiber. The wavelength is the distance a simple signal can travel in one period.

CHARACTERISTICS OF SIGNALS/WAVES

The characteristics of a signal may be one of a broad range of shapes, amplitudes, time durations and perhaps other physical properties, such as statistical and probabilistic.

In general, we shall examine the following characteristics, namely; periodical, symmetrical and continuity.

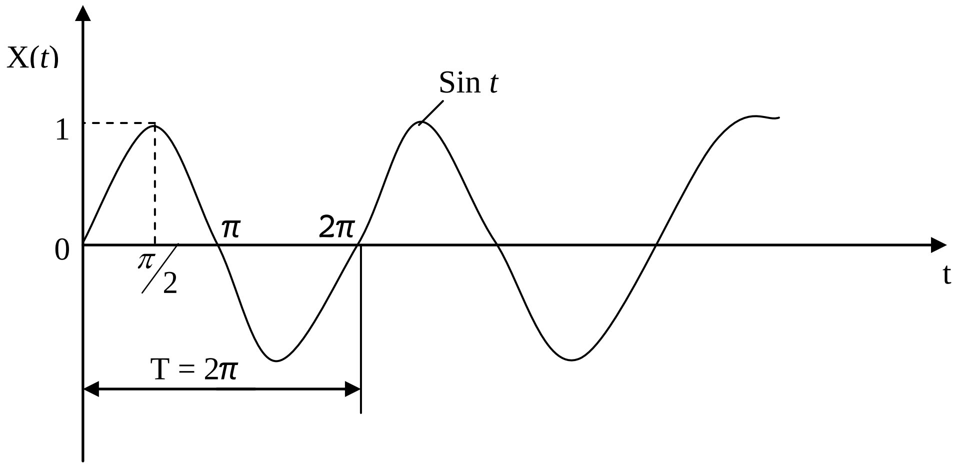
1) **Periodical**

If a signal is periodic, then it is described by the equation:

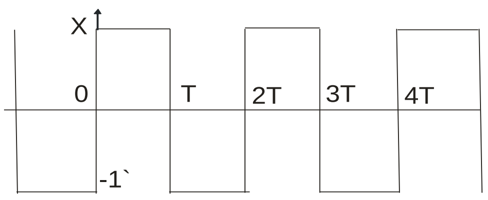
$S(t) = S(t + KT),\ K = 0, 1, 2, 3 \ldots$ (1)

Where, $-\infty < t < +\infty$ **and** T – is the period of the signal.

For instance, the sine wave sin t is periodic with period T = 2π.

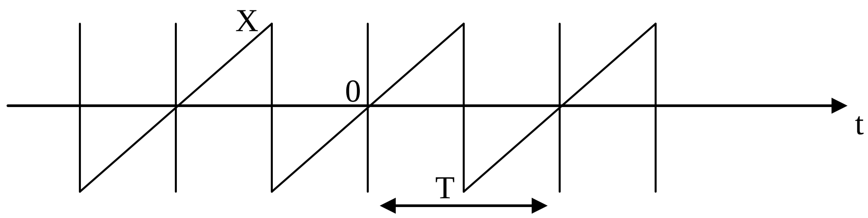


The square wave (see figure below) is another example of a periodic signal.



X(t) = $XSinwt + \frac{x}{3}Sin3wt + \frac{x}{5}\sin 5wt + \ldots$

Other signals such as the rectangular pulse, the saw-tooth wave may be considered "periodic" with an infinite period.



The saw tooth wave:

X (t) = $XSinwt - \frac{x}{2}Sin2wt + \frac{x}{3}\sin 3wt$

2) **<u>Symmetrical</u>** (or symmetrical properties)
We shall use the words **even** and **odd** to verify whether a signal is symmetrical or not.
Any even function usually obeys the relation:

S (-t) = S (t) . . . (2a)

For example:

$\cos(-\theta) = \cos\theta.$

Thus, Cosθ is an even function.

Any odd function usually obeys the relation:

S (t) = - S (t) . . . 2(b)

For example,

Sin$(-\theta) = -Sin\theta$, thus a sinθ function is an odd function.

Any signal or wave S(t) can be resolved into an even component $S_e(t)$ and an odd component S_o (t); such that

S (t) = $S_e(t) + S_o(t)$ . . . . .2(c)

Or

$S(-t) = S_e(-t) + S_o(-t) = S_e(t) - S_o(t)$ . . . .(2d)

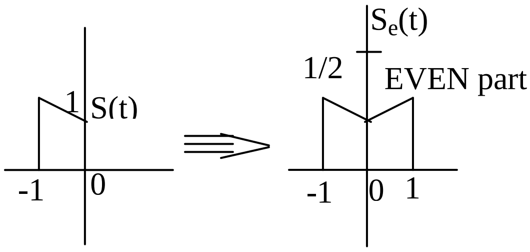
Consequently;

$$\left.\begin{aligned} S_e(t) &= 1/2\,[S(t) + S(-t)] \\ S_o(t) &= 1/2\,[S(t) + S(-t)] \end{aligned}\right\} \quad (2e)$$

Equation 2(e) is the decomposition into even and odd components.

Example: Decompose the following signals into even and odd components?

DIAGRAM

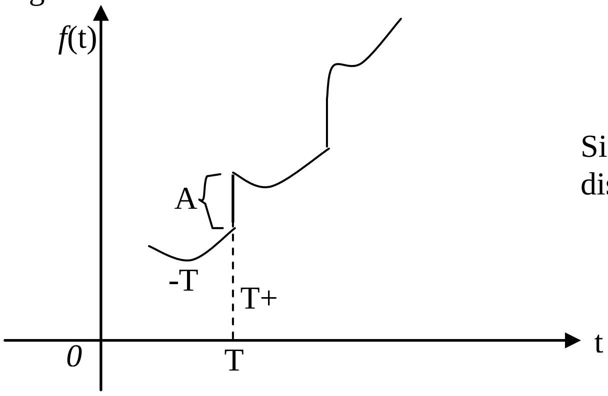


Note that the even function is symmetrical along the y-axis whereas the odd function is symmetrical along the x-axis.

A signal S(t) is continuous of $\lim_{t\to a} S(t) = S(a)$ for all a.

3) **Continuity Property**

Consider the signal shown below:



Signal with discontinuity

Observed that at t = T, the signal is discontinuous. The height of the discontinuity is

$$f(Tt) - f(T-) = A \qquad \ldots \qquad (3)$$

Where

$$\left.\begin{aligned} f(T+) &= \lim_{\alpha\to 0} f(T+\alpha) \\ \text{and} \quad f(T-) &= \lim_{\alpha\to 0} f(T-\alpha) \end{aligned}\right\} \quad \ldots \quad (4)$$

and α is a real positive quantity.

In particular, we are concerned with discontinuous in the neighborhood of t = 0 (i.e. in the neighborhood of the origin).

For equation (4) putting T = 0, the points $f(0+)$ and $f(0-)$ are:

$f(0_+) = \lim_{\alpha\to 0} : f(\alpha)$

$f(0_-) \quad \lim_{\alpha\to 0} : f(-\alpha)$

A signal f (t) is continuous if $\lim_{t\to a} f(t) = f(a)$ for all a.

FOURIER SERIES AND ANALYSIS

One of the famous characteristics of signals is periodicity. As shown above, if a signal is periodical, then it is described by the equation:

$f(t) = f(t \pm KT)$, where K = 0, 1, 2 . . . integers and T- period.

This equation is called **general periodic function**.

Following Euler's observations in the 18th century, that vibrating strings produce sinusoidal motion, on 21 December, 1807, Jean Batiste Joseph Fourier in a historic session of the French Academy in Paris, announced a thesis that opened a remarkable chapter in the history of mathematical analysis and its engineering applications.

Joseph Fourier proffered that any periodic signal or function can be represented in terms of an infinite sum of sine and cosine functions or trigonometric series that are themselves periodical. Thus we obtained:

$$f(t) = \frac{A_0}{2} + A_1 \cos wt + A_2 \cos 2wt + \ldots B_1 \sin wt + B_2 \sin 2wt + \ldots \quad (1)$$

Where $w = \frac{2\pi}{T}$ is called the radian frequency, which is $w = 2\pi f$ and is measured in radian/sec.

$nw-$with n = 2, 3, … is the nth harmonic.

The trigonometric series (equation 1) above is generally referred to as the FOURIER SERIES.
The first term in equ (1) above is a constant component or zero harmonic of a wave.
The terms with A_1 and B_1 constitute the first harmonic with w_1 radian frequency, while the terms with A_2 and B_2 constitute the second harmonics of the wave with w_2, etc.
$A_0, A_1 \ldots A_n$ and $B_1, B_2 \ldots B_n$ are constants.
In a more compact form, the Fourier series equation (1) can be expressed as follows:

$$\boxed{f(t) = \frac{A_0}{2} + \sum_{n-1}^{\infty} (A_n \cos nwt + B_n \sin nwt)} \quad \ldots \quad (2)$$

$$= \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \left(Cos nwt + \frac{B_n}{A_n} \sin nwt\right)$$

$$= \frac{A_0}{2} + \sum_{n=1}^{\infty} An(Cos nwt - \tan\phi_n Sin nwt)$$

$$= \frac{A_0}{2} + \sum_{n=1}^{\infty} \frac{A_n}{Cos\phi_n} Cos(nwt + \phi_n)$$

$$= \frac{A_0}{2} + \sum_{n=1}^{\infty} C_n Cos(nwt + \phi_n)\text{, Where}$$

$$\left.\begin{aligned} \phi_n &= \tan^{-1}\left(\frac{B_n}{A}\right) \\ C_n &= \sqrt{A_n^2 + B_n^2} \end{aligned}\right\} \quad \ldots \quad (3)$$

For a function to be Fourier series transformable, it must satisfy the DIRICHLET conditions, which ensure mathematical sufficiency, but not necessity. The Dirichlet conditions require that within a period:

i) Only a finite number of maximums and minimums can be present.
ii) The number of discontinuities must be finite.
iii) The discontinuities must be bounded. That is, the function must be absolutely integrable, which requires that

$$\int_0^T / f(t) / dt < \infty$$

The Fourier Series (equation 1) can be described completely in terms of the coefficients of its harmonic terms of A_0, A_1, A_2. . ., B_1, B . . . etc.
These coefficients constitute a frequency domain description of the signal (or wave).
These Fourier coefficients can be determined from the following equations:

$$A_0 = \frac{2}{T}\int_0^T f(t) \quad dt \qquad \ldots 4.1)$$

$$A_n = \frac{2}{T}\int_0^T f(t) \quad Cos\, nwt\, dt \qquad \ldots \qquad (4.2)$$

$$B_n = \frac{2}{T}\int_0^T f(t) \quad sin\, nwt\, dt \qquad \ldots \qquad (4.3)$$

If we integrate between the limit $(0;\pi)$ i.e.

$T = \pi$, we get;

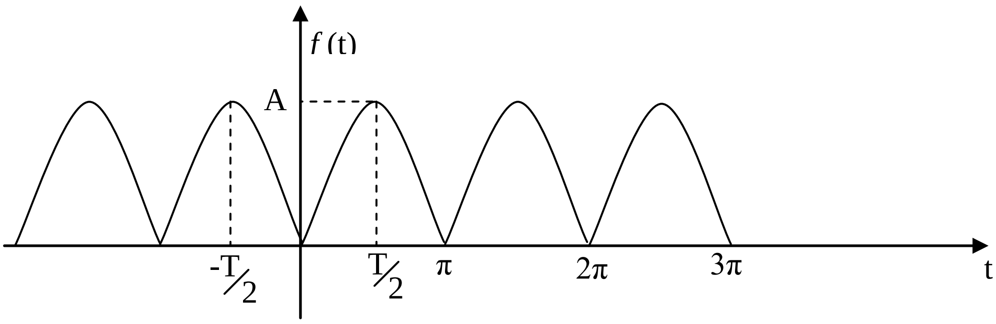
$$A_0 = \frac{2}{\pi}\int_0^\pi f(t) \quad dt \qquad \ldots \qquad (5.1)$$

$$A_n = \frac{2}{\pi}\int_0^\pi f(t) \quad \text{Cos}\, nwt\, dt \qquad \ldots \qquad (5.2)$$

$$B_n = \frac{2}{\pi}\int_0^\pi f(t) \quad \text{sin}\, nwt\, dt \qquad \ldots \qquad (5.3)$$

FOURIER ANALYSIS
Fourier analysis is concerned with determining the Fourier coefficients for a given signal and the corresponding Fourier series.
Solved Examples

1) Determine the Fourier coefficient and the Fourier series for the sine function shown below:



Solution: Here, the period is T = π, Hence; from

$$w = \frac{2\pi}{T} = \frac{2\cancel{\pi}}{\cancel{\pi}} = 2$$

$w = 2$

Hence, the signal is given as:

$$f(t) = A/\sin t/.$$

Using equation 4.3, we have;

$$B_n = \frac{2}{\pi}\int_0^{\pi} f(t)\sin 2nt\, dt = \frac{2A}{\pi}\int_0^{\pi} \sin t \sin 2nt\, dt$$

$$B_n = \frac{2A}{\pi}\int_0^{\pi} \sin 0 \sin 0\, dt = 0$$

Since Sin 0 = 0, cos 0 = 1, sin 180 = 0,

Using equation 4.1, we have;

$$A_0 = \frac{2A}{\pi}\int_0^{\pi} \sin t\, dt = \frac{2A}{\pi}\ (-1)\ [\cos t]_0^{\pi}$$

$$= \frac{2A}{\pi}(-1)[\cos \pi - \cos 0] = \frac{2A}{\pi}(-1)[-1-1]$$

$$= \frac{2A}{\pi}.(-1).(-2) = \frac{4A}{\pi}$$

Using equation 4.2, we have;

$$A_n = \frac{2}{\pi}\int_0^{\pi} f(t)\cos 2\, dt = \frac{2A}{\pi}\int_0^{\pi} \sin t \cos 2nt\, dt =$$

$$= \left(\frac{1}{1-4n^2}\right).\frac{4A}{\pi}$$

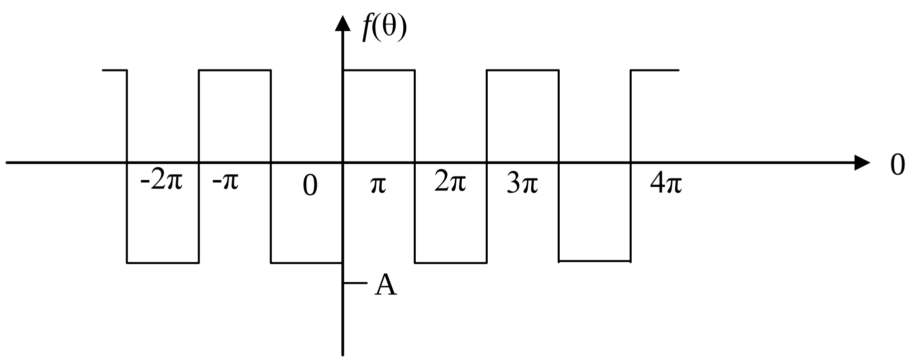
Thus, substituting the values of the coefficients obtained so far in equation (2), the Fourier series of the above sine wave is as follows:

$$f(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos nwt + B_n \sin nwt)$$

$$= \frac{A}{2} + \sum_{n=1}^{\infty} \frac{4A}{\pi(1-4n^2)}\cos 2\ nt$$

$$f(t) = \frac{A}{2} + \frac{4A}{\pi}\sum_{n=1}^{\infty}\frac{1}{1-4n^2}\cos 2\ nt$$

Example 2: Find the Fourier series for the function shown below:



<u>Solution</u>:

The function is symmetrical along the X- axis, therefore it is an odd function.

So, $A_k = 0$

The Fourier series takes the form given by:

$$f(\theta) = \sum_{k=1}^{\infty} B_k \quad \sin k\theta, \text{ where the coefficient}$$

$$B_k = \frac{2}{\pi}\int_0^{\pi} f(\theta) \; \sin K\theta \; d\theta$$

Since, $\theta = wt = 2\pi$

If $t = \pi$, therefore

$$\theta = \frac{2\pi t}{T} = \frac{2\pi\pi}{2\pi}$$

[0;π]

Using the interval of [0, π];

$f(\theta) = A$; where $0 \leq \theta \; \pi$ see (figure)

Thus, by substitution,

$$B_k = \frac{2A}{\pi}\int_0^{\pi} \sin k\theta d\theta$$

$$= \frac{-2A}{\pi K}, \left[\; CosK\theta \right]_0^{\pi} = \frac{-2A}{\pi K} \qquad (CosK\pi - 1)$$

$B_k = 0$ for k even numbers and

$B_k = \frac{4A}{\pi K}$, for k- odd numbers.

Since $f(\theta)$ is an odd function, we take the k- for odd numbers and ignore k- for even numbers.

Therefore,

$$f(\theta) = \frac{4A}{\pi}\sum_{k=1}^{\infty}\frac{SinK}{K}$$

,

$$f(\theta) = \frac{4A}{\pi} \quad (Sin\theta + \frac{1}{3}\sin 3\theta + \frac{1}{5}\sin 5\theta + \ldots)$$

Putting k =1,3,5,. . . etc, odd values.

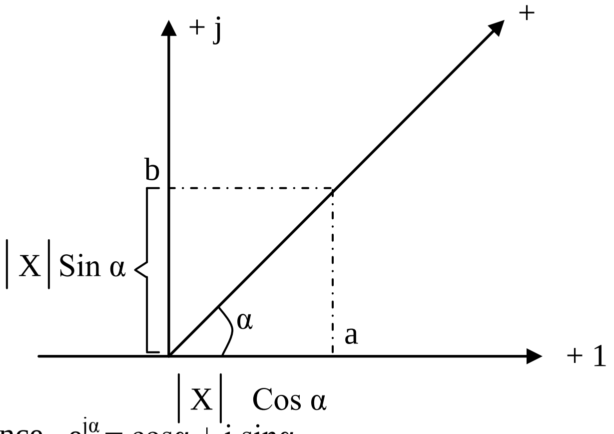
$= \frac{4A}{\pi} \sum_{n=0}^{\infty} \frac{1}{4n+1}$ Sin (2n+1)θ, which is the Fourier series.

SOME APPLICATIONS OF COMPLEX FOURIER SERIES

It has been shown that

$\boxed{x = a + jb}$ - Is a complex form of a signal.

This signal can be represented diagrammatically as follows:



Since $e^{j\alpha} = \cos\alpha + j\sin\alpha$ . .. .(1)

$$|X|\, e^{j\alpha} = \underbrace{|X| \cos\alpha}_{\alpha} + j\, \underbrace{|X| \sin\alpha}_{\beta} \; \ldots\ldots (2)$$

where $e^{j\alpha} = \sqrt{\cos\alpha^2 + \sin^2\alpha)}$ = 1 is a unit vector.

Multiplying both sides of (1) by Im, we get

$$\text{Im } e^{j\alpha} = \text{Im } \text{Cos}\alpha + j \text{ Im } \text{Sin}\alpha \qquad (3)$$

If we change, $\boxed{\alpha = wt + \varphi}$

Where φ- is the phase-shift of the signal, we get

$$\boxed{i = \text{Im}e^{j(wt+\varphi)} = \text{Im}Cos(wt+\varphi) + j\,\text{Im}\sin(wt+\varphi) \;\ldots\ldots. (4)}$$

L.H.S = Im $e^{j(wt+\varphi)}$ = Im $e^{j\varphi}.e^{jwt}$ is called the PEAK value of the current

i in complex form.. equation (5)

The value

Im = Im $e^{j\varphi}$ . . . . (6)
is called complex amplitude of current (*i*)

SOLVED EXAMPLES:

(1) If i = 2 cos (100t - 35^0). Find Im?

<u>**Solution**</u>

i = 2 cos (100t – 35^0) is of the form:

i = Im Cos (wt + φ), where Im = 2 and φ = - 35^0

Therefore, putting these values in equation (6), we have;

Im = 2 $e^{-j35°}$

(2) If the complex voltage is $U_m = 100\, e^{j60}$, Find the peak or original voltage?

Solution:

$$U_m e^{j(wt+\varphi)} = 100 e^{j(wt+60^0)}$$

$$\therefore \bigcup = 100\cos(wt + 60^0) + j100\sin(wt + 60^0)$$

(3) If the complex amplitude of a current is given by
$I_m = 3 + j4$ [A]. What is the peak value of the current (i)?

Solution:

$$Im = \sqrt{(a^2 + b^2)} = \sqrt{(9 + 16)} = 5;$$

$$\tan\varphi = \frac{b}{a} = \frac{4}{3}$$

$$\varphi = 53^0$$

$$I_m = 5e^{jwt} = 5e^{(jwt + 53)^0}$$

$$I = 5\cos(wt + 53^0)$$

(4) Given that $i = 2 \sin(100t - 30^0)$. Find I_m and rewrite in the form $I_m = a + jb$.

Solution:

$I = 2 \cos(100t - 120^0)$ (since $\sin 30^0 = -\cos 120^0$)

$$I_m = 2e^{-j120^0} = 2\cos(-120^0) + j\sin(-120) =$$

$$= 2(-\tfrac{1}{2}) - j2.\frac{\sqrt{3}}{2}$$

$I_m =$ $-(1 + j\sqrt{3})$

(5) Giving the complex amplitude $U_m = -3 + j4$. Rewrite in the form $U_m = U_m\, e^{j\varphi}$

$$\bigcup_m = \sqrt{(-3)^2 + 4^2} = 5$$

$$\text{tg}\varphi = \frac{b}{a} = \frac{-4}{3}$$

$$\varphi = 130^0$$

$$\therefore \bigcup_m = 5e^{j130^0}$$

MODEL OF A DIGITAL COMMUNICATIONS SYSTEM

Communication at its simplest level, involves the symbolic representation of thoughts, ideas quantities, and events we wish to record for later retrieval or transmit for reception at a distant point.

Operationally, this involves the transformation of one set of quantities (thoughts, ideas, etc.) into others (symbols) that are somehow more suited for transmission or recording over a degrading medium and the recovery of estimates of the original quantities at the receiving point.

The goal of communication therefore, is to achieve the maximum information throughput across channel with fixed capacity.

Figure below shows the elements of a generic digital communication system.

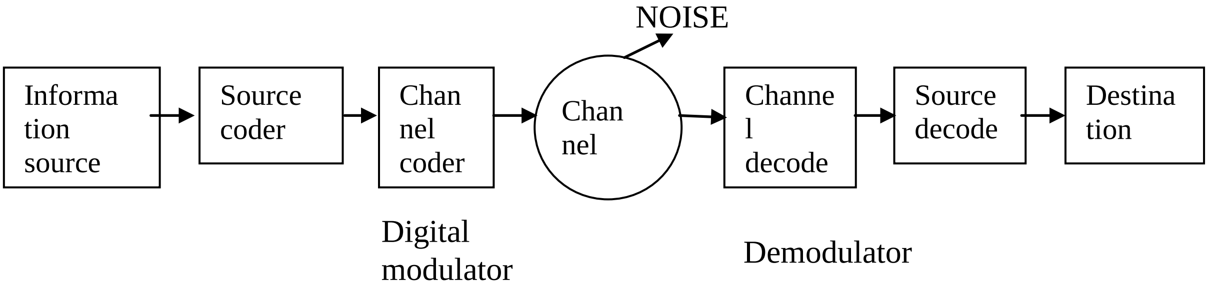


Fig ( ) The components of a generic digital communication system.

The information originates as a signal from a source, either as continuous or discrete. The process of encoding this information for transmission onto, and later retrieval from, a channel involves two conceptually distinct processes.

First, the information stream from the source must be transformed into a set of symbols- this is called **source coding**.

Source coding maps the source information into a set of symbols from a finite alphabet.

Then this information must be impressed on the physical channel properties. The overall requirement for source coding is that the process must be reversible; that is; the original information must be uniquely recoverable from its coded transcription

Encoding an information source into as few symbols as possible results in a more efficient and economical utilization of finite channel resource such as time, bandwidth, and energy.

The sequence of binary digits from the source encoder is to be transmitted through a channel to the intended receiver. For example, the real channel may be either a pair of wires, a coaxial cable, and optical fiber channel, a radio channels, a satellite channel, or some combination of these media. Such channels are basically waveform channels and, hence, they cannot be used to transmit directly the sequence of binary digits from the source. What is required is a device that converts the digital information sequence into waveforms that are compatible with the characteristics of the channel. Such a device is called a **digital modulator or channel encoder**.

In general, no real channel is ideal. There are noise disturbances and other interference that corrupt the signal transmitted through the channel.

In order to overcome such noise and interference and thus, increase the reliability of the data transmitted through the channel, it is often necessary to introduce in controlled manner some redundancy in the binary sequence from the source.

1) The redundancy introduced at the transmitter aids the receiver in decoding the desired information bearing sequence. For example, a form of encoding binary information sequence is simply to repeat each binary digit m times, where m is a positive integer.

2) Another method involves taking k information bits at a time and mapping each k-bit sequence into a unique n-bit sequence, called a **code word**. The amount of redundancy introduced by encoding

the data in this manner is measured by the $\frac{n}{k}$. In this case, the channel bandwidth must also be increased by this ratio to accommodate the added redundancy in the stream.

The reciprocal of this ratio, namely $\frac{1}{n/K} = \frac{K}{n}$ is called the rate of the code or the **code rate**.

A digital signal is a sequence of discrete, discontinuous voltage pulses. Each pulse is a signal element. Binary data are transmitted by encoding each data bit into signal elements.

<u>First</u>, the receiver must know the timing of each bit, i.e. when a bit begins and ends.
<u>Second</u>, the receiver must determine whether the signal level for each bit position is high (1) or low (0). This is done by sampling each bit position in the middle of the interval and comparing the value to a threshold.

$$D = \frac{R}{K} = \frac{R}{Log_2 M}$$

Where;

D = modulation rate or baud
R = data rat, bps
M = no of different signal element 2^K
K = no. of bits per signal element.

To elaborate on the function performed by the modulator: suppose the information is to be transmitted 1 bit at a time at some uniform rate R bits/s.

The modulator may transmit k information bits at a time by using $M = 2^K$ distinct waveforms.

$S_1(t)$, i = 1, 2, . . . M, that is, one waveform for each of the 2^K possible k-bit sequences. This is called M- ary modulation. The modulation rate is called baud. Baud refers to the rate at which the signal level is changed.

We note that a new k-bit sequence enters the modulator every k/R seconds. Hence, the amount of time available to transmit one of the M waveforms corresponding to k-bit sequence is k times the time period in a system which uses binary modulation.

At the receiving end, the digital demodulator processes the channel- corrupted transmitted waveform and reduces each waveform to a single number that represents an estimate of the transmitted data symbol (binary or M-ary).

TYPES OF COMMUNICATION SYSTEMS

We can identify three types of communications systems. These are:

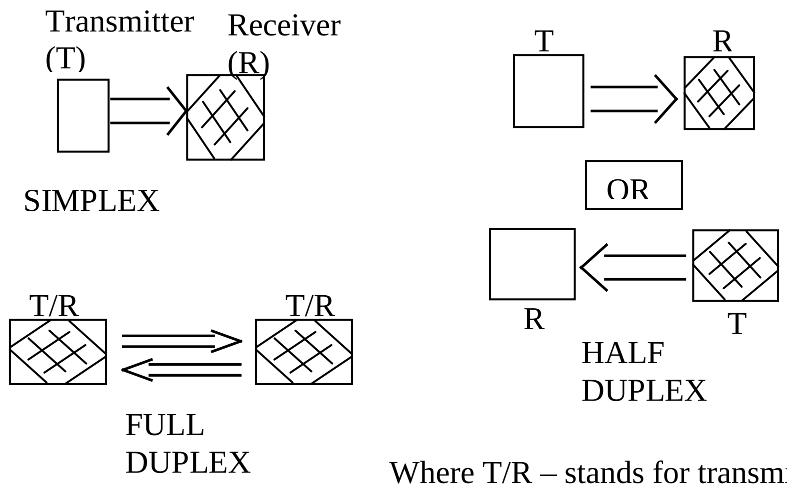
1) Simplex
2) Half Duplex
3) Full Duplex.

Simplex: This is the simplest type of communication system. It allows transmission of data in one direction only.

Half duplex: Transmission can be in both directions, but not at the same time. This means that data is transmitted only in one direction at a time.

Full duplex: In this type of transmission, data can be transmitted simultaneous in both directions.

These are illustrated in Fig ( ) below



Where T/R – stands for transmitter/Receiver

A half duplex system has a special electronic device that switches the direction of flow of data, whereas, in a full duplex there is a device that controls the data flow.

MEASURING INFORMATION

Two of the central tasks of information theory are:

i) The systematic representation of information with a suitable set of symbols and

ii) The reversible conversion from one specific representation to another.

The term information may be defined as a measure of the number of equiprobable choices between several possible alternatives. Thus, information is measured by the logarithm of the number of such alternatives.

Information implies the ability to resolve uncertainty, or a choice between several possible alternatives. The simplest uncertainty is that which is completely resolved by an answer to a YES or NO question. This corresponds to one bit of information when the anticipated answer of either YES or NO is given.

A system's capacity for storing information is fully described by a count of its distinguishable states.

Each state of a physical system is a different configuration of the system.

Some of the properties of an information capacity measure are:

1) A measure of information capacity should increase monotonically with the number of system states;
2) Information should be additive: the aggregate information capacity of two separate systems should be the sum of each system's capacity;
3) The amount of information associated with a system having only one state should be zero (i.e., $\text{Log}_2 1 = 0$).

If Ω is the total number of distinguishable states in a system, then the system's information capacity or amount of information is given by:

$$C = \log_2 \Omega \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where the base 2, denotes the two states of a binary information.

Equation (1) is called H**artley's** measure of information capacity.

The use of writing to represent information is an excellent example of a sequence or string of symbols, namely letters, numbers, and other typographical symbols, to represent information.

For example: The information capacity of a 450-page book, assuming 500 words per page with each word containing five symbols chosen at random from a 37-ary alphabet (i.e. 26 letters, 10 digits, and a blank space) is given by:

C = 450 x 500 x 5 x $\log_2 37 = 5.9\times10^6$ bits.

Here, $\log_2 37 = N$
$2^n = 37$
$N \ln 2 = l_n 37$

$$N = \frac{\ln 37}{\ln 2} = \frac{1.5682}{0.30103} = 5.227$$

Thus, C = 450 x 500 x 5 x 5.227 bits.
This is the capacity needed to store the information contained in a representative book of that size.

LOGARITHMIC MEASURE OF INFORMATION
Let X and Y be two discrete random variables with possible outcomes $X_{i,}$ (i = 1, 2, . . . n) and $Yi_{,}$ (i = 1, 2, . . .m), respectively.
Suppose we observe some outcome Y = y_j and we wish to determine quantitatively, the amount of information that the occurrence of the event Y = yj provides about the event X = xi. The problem is to select an appropriate measure for information.

Notice that when X and Y are statistically independent, the occurrence of Y =y_j provides no information about the occurrence of the event X = x_i.

On the other hand, when X and Y are fully dependent such that the occurrence of Y = y_j determines the occurrence of X = x_i, the information content is simply that provided by the event X =xi.

A suitable measure that satisfies these conditions is the logarithm of the ratio of the conditional probability:

P (X = xi/ Y = y_j) ≡ P (xi/yj) divided by the probability $P(X = x_i) \equiv P(x_1)$.

That is, the information content provided by the occurrence of the event Y = y_j about the event X = x_i, is defined as:

$$\boxed{I(x_i : y_j) = \log \frac{P(x_i / yj)}{P(x_i)}} \quad . .. .(2)$$

I($x_{i:}$ y_j) is called the mutual information between x_i and y_j.
The units of *I($x_{i:}$ y_j)* are determined by the base of the logarithm, which is usually selected as either 2 or e.
When the base is 2, the units of *I(x: y_j)* are "bits", and when the base is e, the units of *I(x_i; y_j)* are called "nats" (natural units).

When the random variables X and Y are **statistically independent,** $P(\frac{x_i}{y_j}) \equiv P(x_i)$ and hence of *I(x_i; y_j)=0*.
On the other hand, when the occurrence of the event Y = yj unique determines the occurrence of the event X = x_i, the conditional probability in the numerator of equation (2) is unity and $I(x_i, y_i) = \log \frac{1}{P(x_i)} = -\log P(x_i)$ . . . .(3)
Because (3) is just the information of the event X = x_i, it is called the "self-information" of the event *X* = x_i and is denoted as $I(x_i) = \log \frac{1}{P(x)} = -\log P(x_i)$ . . . . (4)
Note that a high-probability event conveys less information than a low-probability event
If there is only a single event x with probability P(x) = 1, then
I (x) = log 1/ P (x) = - log 1 = 0

The mutual information $I\ (x_i : y_j)$ between two discrete random variables x and y has the following properties:

i) The mutual information between x and y is symmetric; that is $I(x_i; y_j)= I(x_i; y_j)$

ii) The mutual information between x and y is always non-negative; that is, $I(x:y) \geq 0$. This property states that we cannot lose information on the average by observing the system output y.

iii) The mutual information between x and y may be expressed in terms of the entropy of y as $I(x:y) = H(y) - H(y/x)$ where H(y/x) is a conditional entropy.

Example1: Suppose we have a discrete information source that emits a binary digit either 0 or 1, with equal probability every t_s second. The information content of each output from

the source is:

$$I\ (x_i) = -\ log_2\ P(x_i). \qquad (x_{i\,=\,0,\,1,})$$

$$= -\log_2 \frac{1}{2} \qquad \text{where } P(x_i) = \frac{1}{2}$$

$$= \underline{1\ bit}$$

We define the average self-information, denoted by *H(X), as*

$$H(X) = \sum_{i=1}^{n} P(x_i)\ .\ I(x_i) = -\sum_{i=1}^{n} P(x_i)\ .\ \log P(x_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

When X represents the alphabet of possible output letters from a source, *H(x)* represents the average self information per source letter, and equation (5) is called the **entropy** of the source or **source entropy** per code word.

If the letters from the source (special case) are equally probable, $P(x_i) = 1/n$ for all i and hence $H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log(\frac{1}{n}) = \log n$.

In general, $H(x) \leq \log n$, for any given set of source letter probabilities.

The entropy H(x) is a measure of the average amount of information covered per message.

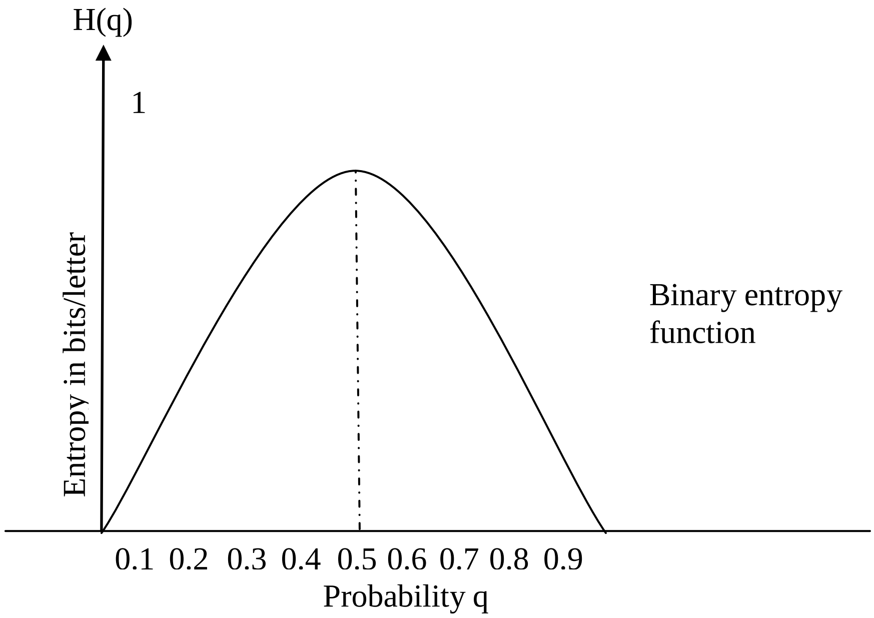
In other words, the entropy of a discrete source is a maximum when the output letters are equally probable. The average conditional self-information is called the conditional entropy and is defined as:

$$H(X/Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i; y_j) \log \frac{1}{P(x_i y_j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (6)$$

which is the uncertainty in X when Y is observed.

Example2: Consider a source that emits a sequence of statistically independent letters, where each output letter is either 0 with probability q or 1 with probability 1-q. The entropy of this source is:

$H(X) \equiv$ H (q) = -q log q - (1-q) log(1-q). . . . . (7)



CHANNEL CHARACTERISTICS

Information is communicated by transmitting over a communication channel. Being a physical system, communication channels impose certain limitations on the rate at which information can be transmitted. Quantities like the number of channel states, the state transition probabilities, bandwidth, and noise level can be related to channel physical characteristics.
A channel consisting of M input symbols $x_{1,}$ x_2, ...$x_{m\ ,}$ N output symbols y_1, y_2, . . . y_N , and with input-output transition probabilities completely specified by the first -order conditional probabilities:

{$P(y_j,/x_i)$; i = 1,2, . . . M; j = 1,2, . . .N} is called a **discrete memoryless channel**.

Let us consider the simplest model for a discrete channel called the BINARY SYMMETRIC CHANNEL (BSC.). The BSC has identical input and output symbol or alphabets, namely, the binary alphabet x: {0,1}, and the binary alphabet y: {0,1}.

The BSC is characterized by symmetrical transition probabilities for the probability that the symbol that is received is the same as the symbol that was transmitted. If the transition probability (0 → 1 or, since the channel is symmetric, 1 → 0) is P, then

prob {Y = 0/ X = 0} = 1 - p_o
prob { Y = 0/ X =1} = p_1
prob { Y = 1/ X = 0} = p_o
prob { Y = 1/ X= 1} = 1 - p_o

A convenient graphical representation of the BSC channel transition properties is shown in fig. ( ) below:

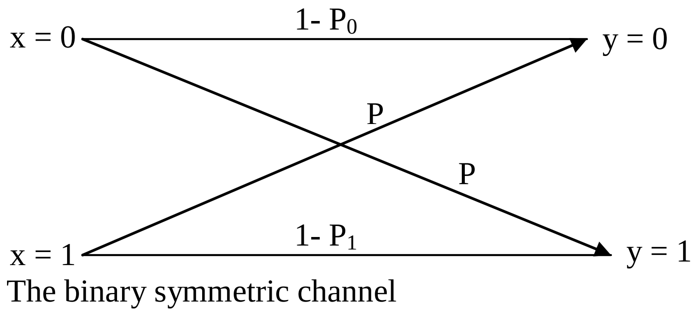
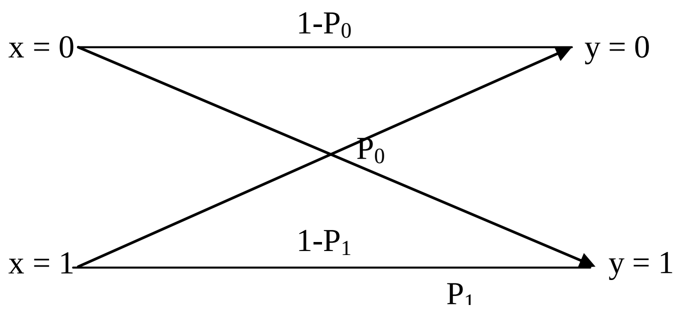


Fig ( ): The binary symmetric channel

Example: Suppose that X and Y are binary-valued {0, 1} random variables that represent the input and output of a binary-input, binary-output channel. The input symbols are equally likely and the output symbols depend on the input according to the conditional probabilities:

$P\{Y = 0/ X = 0\} = 1- P_0$
$P\{ Y = 1/ X =0\} = P_0$
$P\{ Y = 1/ X = 1\} = 1- P_1$
$P\{ Y = 0/ X = 1\} = P_1$

Let us determine the mutual information about the occurrence of the events X = 0 and X =1, given that Y = 0.
From the probabilities given previously, we obtain:
$P(Y=0) = P(Y=0/X=0)P(X=0) + P(Y=0/X=1)P(X=1) =$
(Applying the rule that $P(X=0) = P(X=1) = \frac{1}{2}$)



$$P(Y=0) = (1-P_0).\tfrac{1}{2} + P1/2 = \frac{1-P_0}{2} + \frac{P_1}{2} = \tfrac{1}{2}(1-P_0+P_1)$$

$$P(Y=1) = P(Y=1/x=0).P(x=0) + P(Y=1/x=1)P(x=1) = \tfrac{1}{2}(1-P+P_0)\,.$$

Thus, the mutual information about the occurrence of the event *x* = 0, given that y = 0 is observed:

Similarly, given that Y=0 is observed, the mutual information about the occurrence of the event x = 1

$$I(x_1;y_1) = I(0;0) = \log_2 \frac{P(y=0/x=0)}{P(y=0)} = \log_2 \frac{2(1-P_0)}{1-P_0+P_1}$$

is:

$$\boxed{I(x_2;y_1) = (1;0) = \log_2 \frac{2P_1}{(1-P_0+P_1)}} \Longrightarrow \log_2 \frac{P(Y=0/x=1)}{P(Y=0)} = \log_2 \frac{2P_1}{(1-P_0+P_1)}$$

$\boxed{\text{If } P_0 = P_1 = 0}$, the channel is called **noiseless** and $I(0;0) = \log_2 2 = 1$ bit.

Observed that the output specifies the input with certainty.

On the other hand, if $P_0 = P_1 = 1/2$, the channel is **useless** because

$$I(0;0) = \log_2 1 = 0.$$

Also by substituting P(0) = P(1) = ½ and the appropriate channel transition probabilities in the expression for mutual information I(x,y), we obtain the following expression for the capacity of the binary symmetric channel:

$$C(P) = 1 + P \log P + (1-P) \log (1-p) \quad \ldots\ldots\ldots\ldots\ldots 10$$

Or $\boxed{C(P) = 1 - h(p),}$ where

$h(p) = - p \log p - (1-p) \log (1-p)$ is the binary entropy function for the probability pair

(p,1-p). (Compare with equation 7).

Finally, the capacity of additive white-noise channel is given by:

$$C = W \log_2(1 + S/N) \text{ bit/sec.} \quad \ldots\ldots\ldots\ldots\ldots 11$$

Where,
W- Is the channel bandwidth available,
S- Signal power and
N- Noise power.

Equation (11) is the famous SHANNON'S theorem, which gives an upper bound to the information capacity of a channel, based on it bandwidth and signal-to-noise ratio.

BANDWIDTH

Let us briefly provide definitions to the following: bandwidth, bandwidth of an analog signal, bandwidth of digital signal and bandwidth of a channel.

Bandwidth can be defined as the portion of the electromagnetic spectrum occupied by the signal. It may also be defined as the frequency range over which a signal is transmitted. Bandwidth of analog and bandwidth of digital signals are calculated in different ways.

Bandwidth of Analog Signal: Analog signal bandwidth is measured in terms of its frequency (Hz). It is defined as the range of frequencies that the composite analog signal carries. It is calculated by the difference between the maximum and the minimum frequency. For example if frequency f1 =30 Hz and f2 =90 Hz, then Bandwidth (W) =f2 – f1 =90 - 30 = 60 Hz..

Bandwidth of Digital Signal: Digital Signal bandwidth is measured in terms of bit rate second (bps). It is defined as the maximum bit rate of the signal to be transmitted.

Bandwidth of a Channel (Wc): Bandwidth of signal is different from bandwidth of the medium or channel. A channel is the medium through which the signal carrying information will be passed. In terms of Analog signal, bandwidth of the channel is the range of frequencies that the channel can carry. In terms of digital signal, bandwidth of the channel is the maximum bit rate supported by the channel, i. e., the maximum amount of data that the channel can carry per second. Generally, Wc >Ws.

CHANNEL CAPACITY

The Channel capacity- Is the maximum rate at which data can be transmitted over a given communication path, or channel and under given conditions.

Data rate - is the rate at which data can be communicated in bits per second (bps)

Bandwidth – is the permissible rate of transmission expressed in cycles per second or Hertz (Hz).

Nyquist Bandwidth

In a noise free environment, the data rate equals the bandwidth of the signal.

Given a bandwidth of W, the *highest* (i.e., maximum) signal rate that can be carried is 2W. Thus by Nyquist Bandwidth:

$$C = 2W.$$

For example: if the bandwidth is 3100Hz, then the capacity C = 2W = 2 x 3100bps. (Assume binary signals of two levels).

If M possible voltage levels are used as signals, then each signal element can represent two bits.

$C = 2W \log_2 M$ – By Nyquist formula

where M is number of discrete signal or voltage levels.

SHANNON CAPACITY FORMULA

For a given level of noise, we would expect that a greater signal strength would improve the ability to receive data correctly in the presence of noise.

Signal-to-noise ratio (SNR or S/N)

SNR_{dB} = ($10\log_{10}$ signal power/noise power) dB

or ($SNR_{dB} = 10\log_{10}(SNR)$)

Shannon deduced that, the theoretical maximum channel capacity is:

$$[C = W \log_2 (1 + SNR)]$$

Where C - is the capacity of the channel in bps and;

W - is the Bandwidth.

The Shannon's formula (c) sets the upper-bound on the achievable data rate.

Note that the wider the bandwidth, the more noise is admitted to the system, thus, as W increases, SNR decreases.

Solved Example1:

Suppose that the spectrum of a channel is between 3MHz and 4 MHz and SNR_{dB} is 24 dB. Find (i) the channel capacity C, (ii) the number of discrete signal levels (M).

Solution:

(i) $W = 4\text{ MHz} – 3\text{ MHz} = 1\text{MHz} = 10^6\text{Hz}$

$SNR_{dB} = 24\text{ dB} = 10 \text{ Log}_{10} (SNR)$ $\qquad$ $24 = \log_{10} (SNR)$

$SNR = 251$ $\qquad$ $24 = \log_{10} SNR$

Using Shannon's Formula $\qquad$ $SNR = 10^{2.4}$

$$C = W \log_2 (1 + SNR) = 10^6 \times \log_2 (1+251) \cong 10^6 \times 8 = \underline{\underline{8\text{Mbps}}}$$

(ii) Using Nyquist Formula:

$C = 2 *W* \log_2 M$

$8 \times 10^6 = 2 \times (10^6) \times \log_2 M$

$4 = \log_2 M \rightarrow M = 2^4 = 16$

$M = \underline{\underline{16}}$

Example 2: Telephones with a bandwidth of roughly 3 KHz and a signal-to-noise ratio of approximately 30 dB. The maximum data rate that such a signal can support is therefore

$C = 3000 \log_2 (1+1000) = 29.9\text{Kb/S}$

Since STN ratio = $10\log_{10} \frac{\text{Wanted signal power}}{\text{Unwanted noise power}}$ $\qquad$ $= 1\phi \log_{10}(\frac{w}{unw}) = 3\phi \text{ dB}$

Here, $\frac{\text{wanted signal}}{\text{unwanted noise power}}$

$\underline{\underline{\text{Since } \log_{10} = 3}}$

Example 3: If the signal power equals the noise power in a channel of bandwidth 1Hz, what is the theoretical information rate in bits/s which can be carried through the channel?

Solution:

By Shannon' theorem:

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

If W =1, and S = N, then

$$C = \log_2 (1+ \frac{S}{N}) = \log_2 2 = 1 \text{ bit/s}$$

TRANSMISSION MEDIA (or LINES)

A transmission medium can be defined as anything that can carry information from a source to a destination. A transmission line consists of a pair of copper conductors that are separated from one another by a dielectric. We can distinguish two broad types transmission media in common use nowadays. These are wired and wireless media. The wired media may be subdivided into three types, namely:

(i) The two-wire or twin line, or twisted-pair cable
(ii) The coaxial line cable and
(iii) The optical fibre cables.

Figure ( ) below depicts the two-wire and the coaxial line

a)

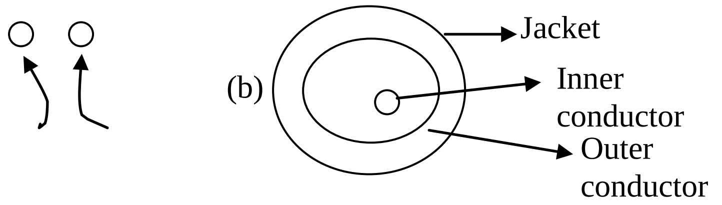
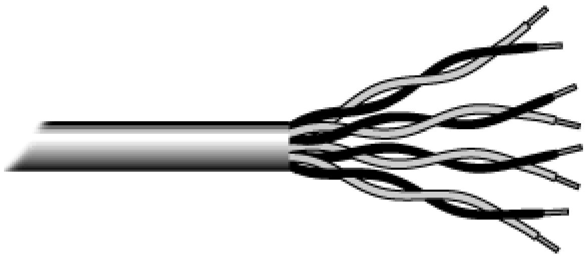


Fig (a) Twin line: two types (i) Shielded (STP) and (ii) Unshielded (UTP)
(b) Coaxial line: two types (i) Thinnet (flexible, light and is about 0.25 inches thickness) and (ii) Thicknet (does not bend, about 0.5 inches13mm in diameter).



Unshielded TP

The quality of UTP may vary from telephone-grade wire to extremely high-speed cable. The cable has four pairs of wires inside the jacket. Each pair is twisted with a different number of twists per inch to help eliminate interference from adjacent pairs and other electrical devices. The tighter the twisting, the higher the supported transmission rate and the greater the cost per foot. The EIA/TIA (Electronic Industry Association/Telecommunication Industry Association) has established standards of UTP and rated six categories of wire (additional categories are emerging).

Categories of Unshielded Twisted Pair

Category	Speed	Use

1	1 Mbps	Voice Only (Telephone Wire)
2	4 Mbps	LocalTalk & Telephone (Rarely used)
3	16 Mbps	10BaseT Ethernet
4	20 Mbps	Token Ring (Rarely used)
5	100 Mbps (2 pair) 1000 Mbps (4 pair)	100BaseT Ethernet Gigabit Ethernet
5e	1,000 Mbps	Gigabit Ethernet
6	10,000 Mbps	Gigabit Ethernet

Unshielded Twisted Pair Connector

The standard connector for unshielded twisted pair cabling is an RJ-45 connector. This is a plastic connector that looks like a large telephone-style connector (See fig. below). A slot allows the RJ-45 to be inserted only one way. RJ stands for Registered Jack, implying that the connector follows a standard borrowed from the telephone industry. This standard designates which wire goes with each pin inside the connector.

Fig.( ). RJ-45 connector

Shielded Twisted Pair (STP) Cable

Although UTP cable is the least expensive cable, it may be susceptible to radio and electrical frequency interference (it should not be too close to electric motors, fluorescent lights, etc.). If you must place cable in environments with lots of potential interference, or if you must place cable in extremely sensitive environments that may be susceptible to the electrical current in the UTP, shielded twisted pair may be the solution. Shielded cables can also help to extend the maximum distance of the cables.

Shielded twisted pair cable is available in three different configurations:

1. Each pair of wires is individually shielded with foil.
2. There is a foil or braid shield inside the jacket covering all wires (as a group).
3. There is a shield around each individual pair, as well as around the entire group of wires (referred to as double shield twisted pair).

Coaxial Cable

Coaxial cabling has a single copper conductor at its center. A plastic layer provides insulation between the center conductor and a braided metal shield (See fig. below). The metal shield helps to block any outside interference.

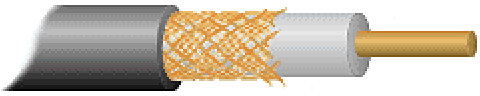


Fig. Coaxial cable

Although coaxial cabling is difficult to install, it is highly resistant to signal interference. In addition, it can support greater cable lengths between network devices than twisted pair cable. The two types of coaxial are:

(i) Thin coaxial cable is also referred to as thinnet. 10Base2 refers to the specifications for thin coaxial cable carrying Ethernet signals. The 2 refers to the approximate maximum segment length being 200 meters. In actual fact the maximum segment length is 185 meters. Thin coaxial cable has been popular.

(ii) Thick coaxial cable is also referred to as thicknet. 10Base5 refers to the specifications for thick coaxial cable carrying Ethernet signals. The 5 refers to the maximum segment length being 500 meters. Thick coaxial cable has an extra protective plastic cover that helps keep moisture away from the center conductor. This makes thick coaxial a great choice when running longer lengths in a linear bus network. One disadvantage of thick coaxial is that it does

Coaxial Cable Connectors

The most common type of connector used with coaxial cables is the Bayone-Neill-Concelman (BNC) connector (See fig. below). Different types of adapters are available for BNC connectors, including a T-connector, barrel connector, and terminator. Connectors on the cable are the weakest points in any network. To help avoid problems with your network, always use the BNC connectors that crimp, rather

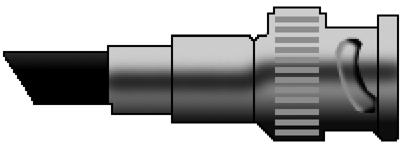


Fig. . BNC connector

Fiber Optic Cable

Fiber optic cabling consists of a center glass core surrounded by several layers of protective materials (See fig. below). It transmits light rather than electronic signals eliminating the problem of electrical interference. This makes it ideal for certain environments that contain a large amount of electrical interference. It has also made it the standard for connecting networks.

Fiber optic cable has the ability to transmit signals over much longer distances than coaxial and twisted pair. It also has the capability to carry information at vastly greater speeds. This capacity broadens communication possibilities to include services such as video conferencing and interactive services. The cost of fiber optic cabling is comparable to copper cabling.

The center core of fiber cables is made from glass or plastic fibers (see fig below). A plastic coating then cushions the fiber center, and kevlar fibers help to strengthen the cables and prevent breakage. The outer insulating jacket made of teflon or PVC.

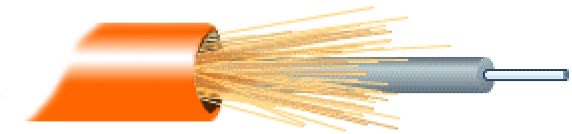


Fig.. Fiber optic cable

There are two common types of fiber cables -- single mode and multimode.

Multimode cable has a larger diameter; however, both cables provide high bandwidth at high speeds. Single mode can provide more distance, but it is more expensive.

Ethernet Cable Summary

Specification	Cable Type
10BaseT	Unshielded Twisted Pair
10Base2	Thin Coaxial
10Base5	Thick Coaxial
100BaseT	Unshielded Twisted Pair
100BaseFX	Fiber Optic
100BaseBX	Single mode Fiber
100BaseSX	Multimode Fiber
1000BaseT	Unshielded Twisted Pair
1000BaseFX	Fiber Optic

1000BaseBX	Single mode Fiber
1000BaseSX	Multimode Fiber

An optical-fibre cable consists of a cylindrical glass core that is surrounded by a glass cladding and it is able to transmit a light wave with very little loss of energy.

Advantages of optical-fibre cable over copper transmission lines are:

i) Light-weight, small-dimensioned cables;
ii) Very wide bandwidth;
iii) Freedom from electromagnetic interference;
iv) Low attenuation, i.e. low decay in signal;
v) High reliability and long life;
vi) Cheap raw materials and
vii) Scalable (negligible) crosstalk between fibres in the same line.

For these reasons, optical fibre is particularly suited to the transmission of digital signals and it is often used for the cabling of a local area network (LANS).

We can generally classify two types of parameters of a medium. These are

a) Primary parameters and
b) Secondary parameters.

Let us briefly examine them in turn.

a) Primary parameters:

The conductors that form a pair in a telephone line usually comprise of four parameters which are as follows:

i) Resistance (R)
ii) Inductance (L)
iii) Capacitance (C) and
iv) Leakance (g) or conductance of a line.

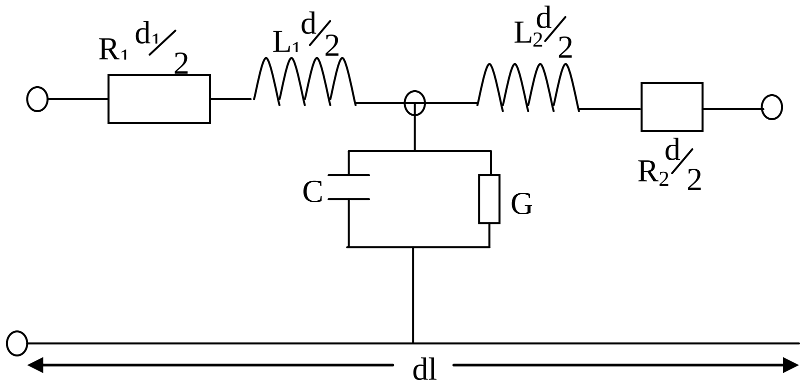
All these four parameters are uniformly distributed over the length of the line.

The resistance, R, is the loop resistance in ohms of a one-kilometre length of the line (i.e. the sum of the resistances of each conductor).

The inductance, L, is the total series inductance of both conductors, or loop inductance. It is measured in henry's per kilometer.

The capacitance, C, is the total capacitance between a one-kilometre length of the two conductors measured in microfarads per kilometer.

The leakance, G, represents the leakage of current between the two conductors. This leakage occurs partly because the insulation resistance between the conductors is not infinite, and partly because current must be supplied to supply the power losses in the dielectric as the line capacitance is charged and discharged. Figure below show a typical transmission line:



Generally, the line is considered to consist of a very large number of very short lengths dL of line connected in cascade. Each short section has a total shunt capacitance CdL and total shunt leakage GdL. The total series resistance and inductance are RdL and LdL, respectively.

b) Secondary Parameters of a line

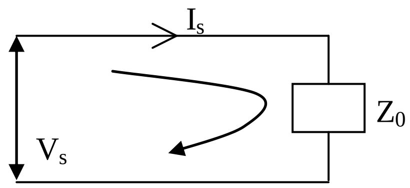
The secondary parameters of a transmission line are its:

- i) Characteristic impedance,
- ii) Attenuation coefficient
- iii) Phase-change coefficient
- iv) Velocity of propagation

i) **Characteristic impedance, Z_0** of a transmission line is the ratio between the voltage across input terminals and the current flowing into the terminals.

That is, $Z_0 = \frac{V_s}{I_s}$ (in ohms)

Or $V_s = Z_0 I_s$



ii) **Attenuation**- is the term given to the decay in the amplitude of a current, voltage or wave along a transmission line, which happens in an exponential manner. Attenuation therefore refers to the progressive reduction in propagated signal.

The percentage reduction in amplitude is exactly the same in each kilometer of the line. If for instance, the input voltage is 12V and 10% is lost in every kilometer of the line. Then the voltage that will enter the second kilometer is 10.8V, and in the third kilometer is 9.72V etc. It is measured in Decibels per kilometer.

iii) **Phase-change coefficient**

The phase different between the voltages 1 Km apart is known as the phase-change coefficient β of the line. The phase-change coefficient is measured in radians per kilometer. In each kilometer length of the line there will be the same phase shift and hence for a line that is L kilometers long the total phase difference is equal to β L rad.

iv) **Velocity of propagation**

The phase velocity v_p of a line is the velocity with which a sinusoidal wave travels along that line. The phase velocity is equal to the angular velocity (w) of the signal divided by the phase-change coefficient; $\frac{w}{\beta}(m/s)$

For a digital data waveform the ratio $\frac{w}{\beta}$ must be constant at all frequencies.

Transmission Media Problems

Round-trip delay (a) – is the time delay between the first bit of a block being transmitted by the sender and the last bit of its associated acknowledgment being received.

Round trip delay (a) = T_P/Tx

T_P – propagation delay and Tx – transmission delay

Where T_p = Physical separation (S) in meters/Velocity of propagation (V) in meters per second

and Tx = number of bits to be transmitted (N)/Link bit rate (R) in bits per second

Solved Example:

A 1000 – bit block of data is to be transmitted between two computers. Determine the ratio of the *propagation delay* to the transmission delay (a), for the following types of data link:

(i) 100m of *twisted-pair* wire and a transmission rate of 10kbps,

(ii) 10km of *coaxial cable* and a transmission rate of 1Mbps,

(iii) 50,000km of *free space* (satellite link) and a transmission rate of 10 Mbps. Assume that the velocity of propagation of an electrical signal within each type of cable is $2 \times 10^8 ms^{-1}$, and that of free space $3 \times 10^8 ms^{-1}$

Solution:

(i) $T_p = S/V = 100/2x10^8 = 5 \times 10^{-7}$ S $a = T_p/T_x = 5x10^{-7}/0.1 = 5.10^{-6}$

$T_x = N/R = 1000/10x10^3 = 0.1s$

(ii) $T_p = S/V = 10x10^3/2x10^8 = 5 \times 10^{-5}$ S $a = T_p/T_x = 5x10^{-5}/1x10^{-3} = 5x10^{-2}$

$T_x = N/R = 1000/1x10^6 = 1x10^{-3}$ S

(iii) $T_p = S/V = 5x10^7/3x10^8 = 1.67 \times 10^{-1}S$ $a=T_p/T_x= 1.67x_{10}{}^{-1}/1x10^{-4} = 1.67x10^3$

$T_x = N/R = 1000/10x10^6 = 1x10^{-4}S$

Conclusion

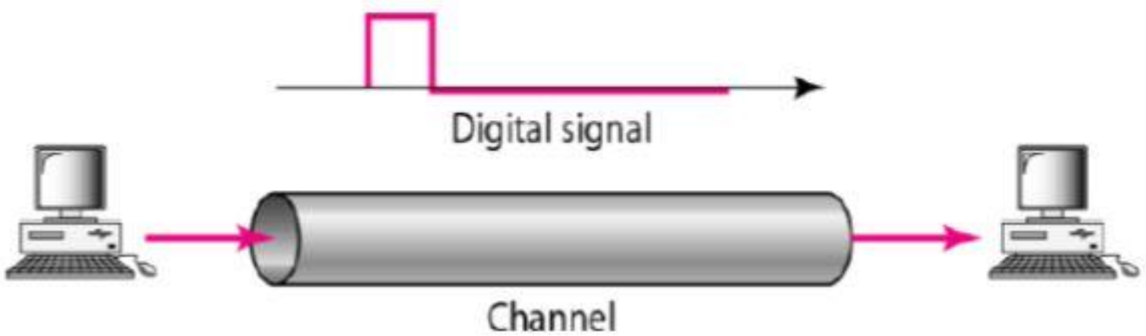
If a < 1, then the round-trip delay is determined by the transmission delay T_p.
If a = 1, then both delays have equal effect
If a > 1, then the propagation delay dominates.
Other important characteristics include distortion, bit error, noise, etc.

Different Methods for Digital Signal Transmission

A digital signal periodic or non-periodic, is a composite analog signal with frequencies between zero and infinity. We can transmit a digital signal by using one of two different approaches: baseband transmission or broadband transmission (using modulation).

1. Baseband Transmission

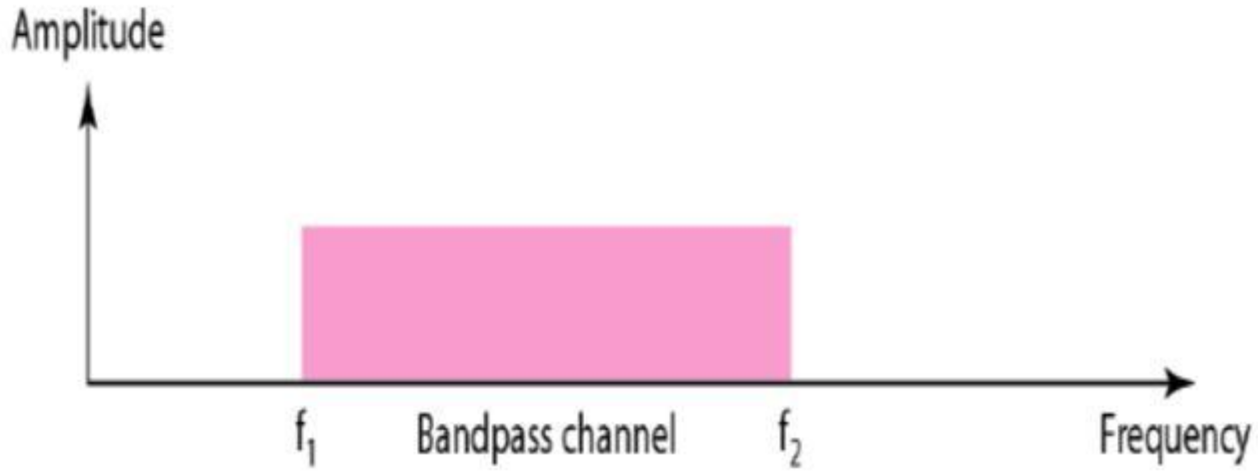
Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal. The following figure shows baseband transmission.



Baseband transmission requires a low-pass channel, a channel with a bandwidth that starts from zero. This is the case if we have a dedicated medium with a bandwidth constituting only one channel. For example, the entire bandwidth of a cable connecting two computers is one single channel. As another example, we may connect several computers to a bus, but not allow more than two stations to communicate at a time.

2. Broadband Transmission (Using Modulation)

Broadband transmission or modulation means changing the digital signal to an analog signal for transmission. Modulation allows us to use a band pass channel-a channel with a bandwidth that does not start from zero. This type of channel is more available than a low-pass channel. The following figure shows a band pass channel.



Differences between Baseband and Broadband

The Table below shows the differences between the above.

S/N	Baseband (CDMA 1.25 MHZ)	Broadband (WCDMA 5 MHZ)
1	Digital signals are used.	Analog signals are used over fibre cables, etc
2	Frequency Division Multiplexing is NOT possible.	FDM is possible.
3	It is bi-directional transmission	Transmission of data is unidirectional.
4	Signal traveling distance is short.	Signal traveling distance is long.
5	Entire bandwidth of the cable is used by a single signal.	The bandwidth is divided into a number of channels with multiple frequencies to be transmitted simultaneously.
6	Signals are of very low frequencies since Not used for long distances.	Signals are of high frequencies thus used for long distances..
7	Used in LAN implementation.	Used in WAN implementation and may travel over cables that are buried underground..

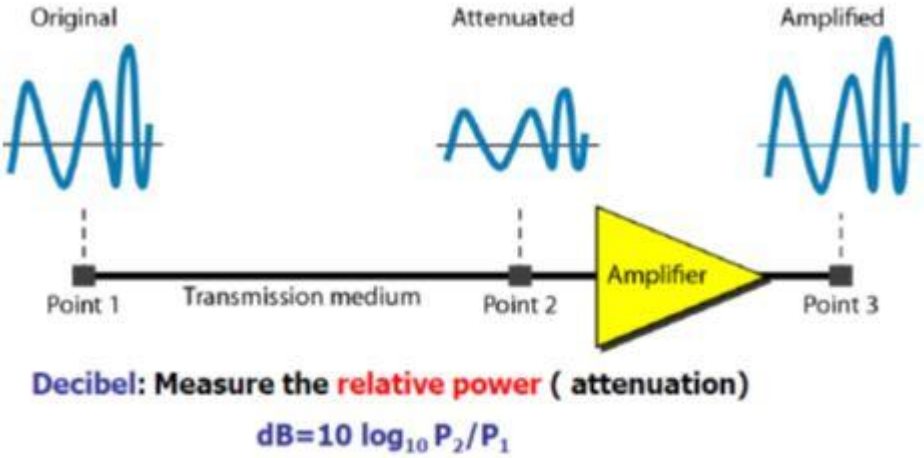
Transmission Impairment

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium. What is sent is not what is received.

The three different causes of impairment are attenuation, distortion, and noise.

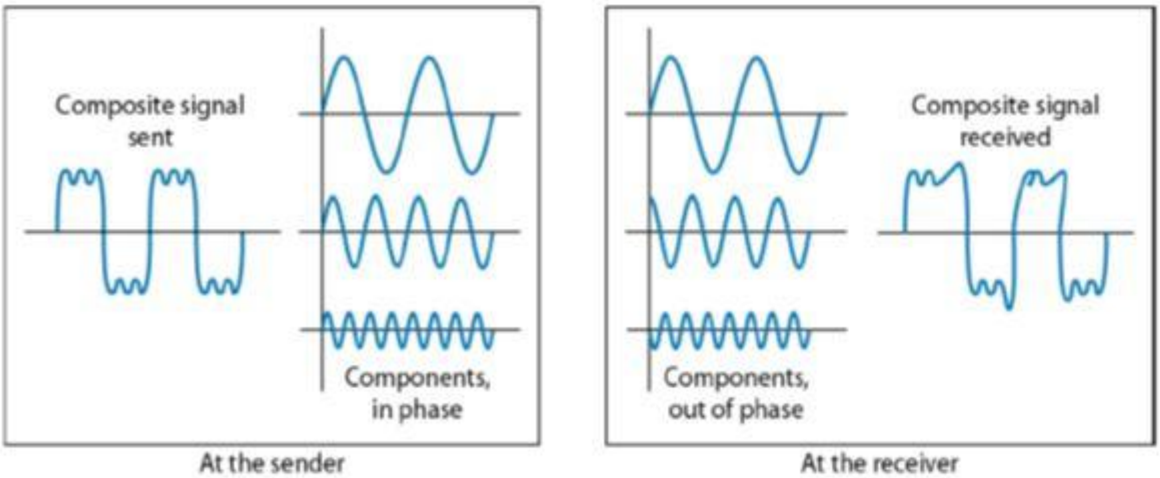
Attenuation:

Attenuation means a loss of energy. When a signal, simple or composite, travels through a medium, it loses some of its energy in overcoming the resistance of the medium. That is why a wire carrying electric signals gets warm, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To compensate for this loss, amplifiers are used to amplify the signal. The following figure shows the effect of attenuation and amplification.



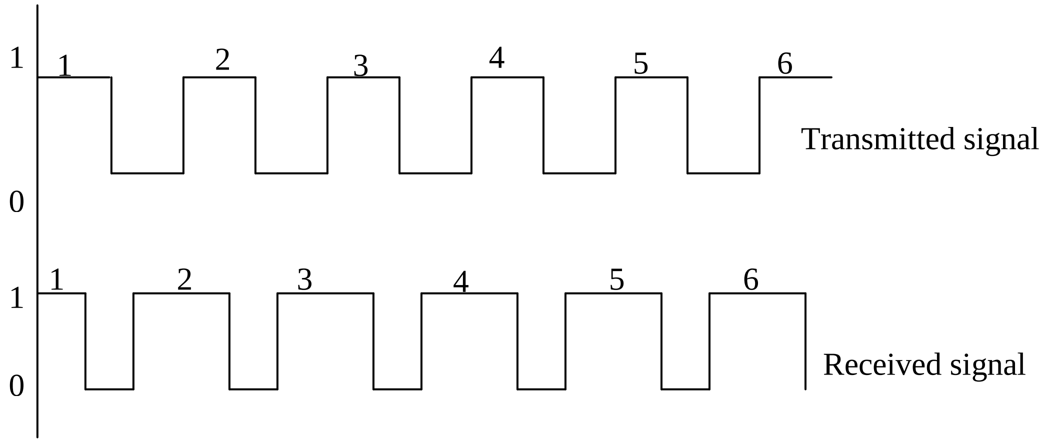
Distortion:

Distortion means that the signal changes its form or shape. Distortion can occur in a composite signal made of different frequencies. Each signal component has its own propagation speed (see the next section) through a medium and, therefore, its own delay in arriving at the final destination. Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration. In other words, signal components at the receiver have phases different from what they had at the sender. The shape of the composite signal is therefore not the same. The following figure shows the effect of distortion on a composite signal.



When a digital signal is transmitted over a telephone circuit the characteristics of that circuit will cause the received signal to be both reduced in amplitude and distorted. This is observed if the regular time interval between successive 1's and 0's of the transmitted signal is either lengthened or shortened at the receiver end. When the various frequency components making up the signal arrive at the receiver with varying delays, this is called **delay distortion.**

This can result in some of the bits incorrectly received or lost.



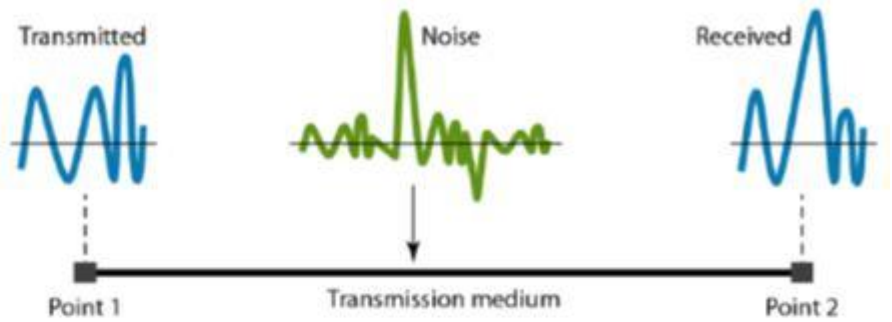
The term positive bias refers to the binary 1 pulses being lengthened and negative bias to the 0 pulses becoming longer.

$$\text{The percentage bias distortion} = \frac{T_1 - T_0}{2(T_1 + T_0)} x \frac{100}{1} \%$$

Where T_1 and T_0 are the time durations of the binary 1 and the binary 0 pulses, respectively.

Noise:

Noise is another cause of impairment. Several types of noise, such as thermal noise, induced noise, crosstalk, and impulse noise, may corrupt the signal. Thermal noise is the random motion of electrons in a wire which creates an extra signal not originally sent by the transmitter. Induced noise comes from sources such as motors and appliances.



These devices act as a sending antenna, and the transmission medium acts as the receiving antenna. Crosstalk is the effect of one wire on the other. One wire acts as a sending antenna and the other as the receiving antenna. Impulse noise is a spike (a signal with high energy in a very short time) that comes from power lines, lightning, and so on. The following figure shows the effect of noise on a signal.

Noise is also a random signal obtained as the result of measuring some physical quantity. One characteristic of physical measurements is that in addition to physical quantity of interest, other effects can influence the outcome. Noise is not the physical process itself, but rather the incomplete representation of a complex process by a signal having few degrees of freedom. Noise comes about because we operate measuring equipment in an environment that is subject to unavoidable interactions with a large number of particles in random motion.

Sources of noise

The various sources of noise that can affect a data communication circuit are:

a) Thermal agitation noise in conductors, resistors and semiconductors. Thermal noise are on the line even when no signal is being transmitted.

b) Short noise and flicker noise in semiconductors,

c) Faulty electrical connections which may cause short breaks in the transmission path,

d) Electrical and magnetic couplings to other circuits, causing cross-talk in equipment wiring and in cables.

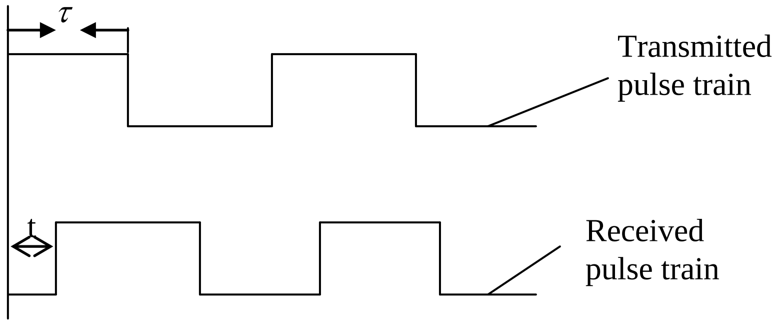
A signal –to –noise ratio (STN) may be defined as the ratio of the wanted signal to the unwanted noise power;

STN = (wanted signal power/ unwanted noise power);

STN ratio = 10 log10 [(wanted signal power)/unwanted noise power)] (in decibel).

BIT ERROR

The data transitions in the received data waveform tend to move around from their ideal positions in time. This results in an effect that is known as bit jitter. See figure ( ) below



If τ is the duration of a pulse and t is the movement of a pulse from it s idea position, then,

Bit jitter = $t_{\max} - t_{\min}$ or

Bit jitter = = $\frac{t_{\max} - t_{\min}}{\tau} x100\%$

BIT ERROR RATE (BER)

Any data circuit is always subjected to noise and interference voltages that originate from a wide variety of sources. These unwanted voltages are superimposed upon the received data voltage and usually corrupt the waveform.

At each sampling instant the receiver must determine whether the bit received at that moment is a 1 or a 0 and any waveform corruption increases the probability of this determination being incorrect and hence of an error occurring. The bit error rate (BER) is given by

$$\text{BER} = \frac{\text{Number of bits wrongly received}}{\text{Total number of bits transmitted}}$$

Example:

A message is transmitted at 2400 bits/s and it occupies a time period of 1 minute and 20 seconds. If two of the received bits are in error calculate the BER.

Solution:

At 2400 bits/s there will be no start and stop bits and so the total number of bits transmitted is 80 x 2400 =192000. Hence,

BER= 2/19200 = $10.42x10^{-6}$

MODULATION AND DEMODULATION

Signal Encoding Techniques

Time Domain Concepts

Viewed as a function of time, an electromagnetic signal can be either ANALOG or DIGITAL.

An analog signal is one in which the signal intensity varies in a smooth fashion over time. In other words, there are no breaks or discontinuities in the signal.

A digital signal is one in which the signal intensity maintains a constant level for some period of time and then changes to another constant level.

Both analog and digital information can be encoded as either analog or digital signals. The particular encoding that is chosen depends on the specific requirements to be met and the media and communications facilities available. Four combinations are available. These are:

1. **Digital data, digital signal:** This is the simplest form of digital encoding of digital data. Here, one voltage level is assigned to binary one, and another to binary zero.
2. **Digital data, Analog Signal:** In this technique, a modem is used to convert digital data to an analog signal so that it can be transmitted over an analog line. The basic techniques used are amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK). All involve altering one or more characteristics of a carrier frequency to represent binary data.
3. **Analog data, digital signal:** Analog data, such as voice and video, are often digitized to be able to use digital transmission facilities. The simplest technique is pulse code modulation (PCM), which involves sampling the analog data periodically and quantizing the samples.
4. **Analog data, analog signal:** Analog data are modulated by a carrier frequency to produce an analog signal in a different frequency band, which can be utilized on an analog transmission system. The basic techniques are amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM).

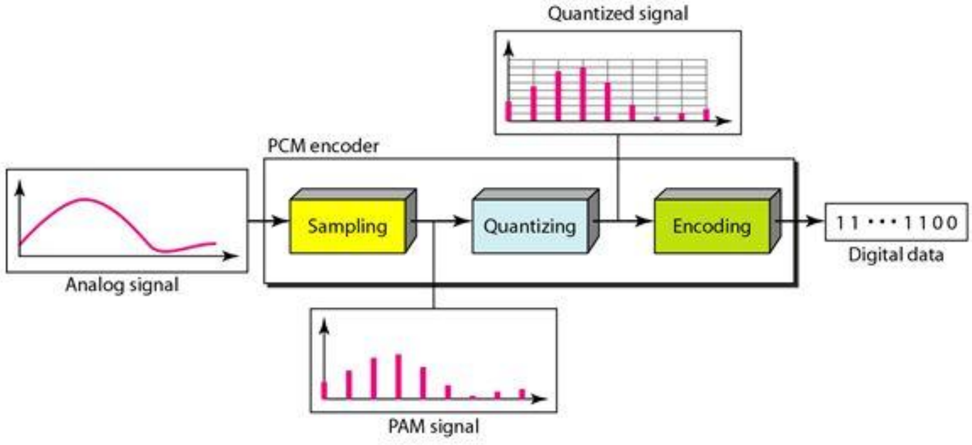
Analog to Digital Conversion Techniques

If we have an analog signal such as one created by a microphone or camera. To change an analog signal to digital data we use two techniques, **pulse code modulation** and **delta modulation.** After the digital data are created (digitization) then we convert the digital data to a digital signal.

1. Pulse Code Modulation (PCM):

Pulse Code Modulation (PCM) is the most common technique used to change an analog signal to digital data (digitization). A PCM encoder has three processes as shown in the following Figure.

(i)The analog signal is sampled.
(ii)The sampled signal is quantized.
(iii)The quantized values are encoded as streams of bits.



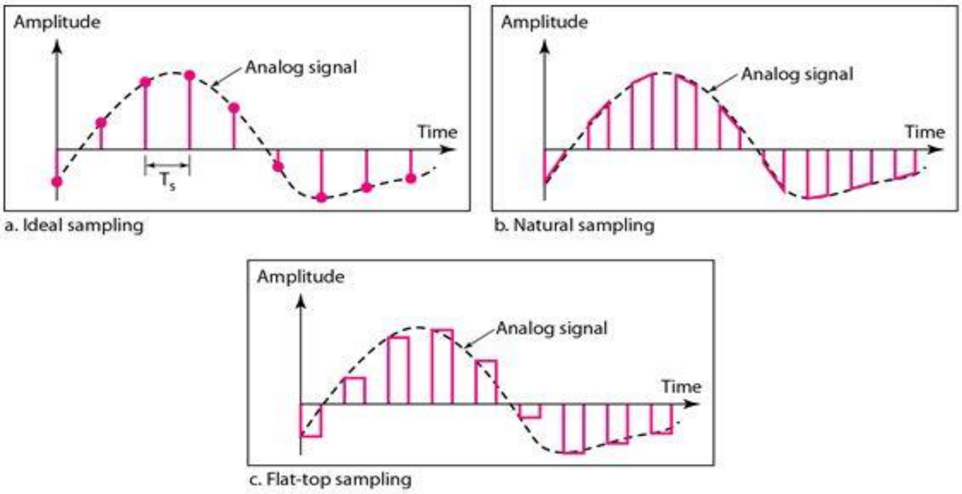
-

Sampling

The first step in PCM is sampling. The analog signal is sampled every Ts s, where Ts is the sample interval or period. The inverse of the sampling interval is called the sampling rate or sampling frequency and denoted by fs, Where fs = 1/ Ts.

There are three sampling methods-ideal, natural, and flat-top. In ideal sampling, pulses from the analog signal are sampled. This is an ideal sampling method and cannot be easily implemented. In natural sampling, a high-speed switch is turned on for only the small period of time when the sampling occurs.

The result is a sequence of samples that retains the shape of the analog signal. The most common sampling method, called sample and hold, however, creates flat-top samples by using a circuit. The sampling process is sometimes referred to as pulse amplitude modulation (PAM). The different sampling methods are as shown in the following figure.



-

Sampling Rate

One important consideration is the sampling rate or frequency. What are the restrictions on Ts? According to the Nyquist theorem, to reproduce the original analog signal, one necessary condition is that the sampling rate be at least twice the highest frequency in the original signal.
As for this Theorem, First, we can sample a signal only if the signal is band-limited i.e a signal with an infinite bandwidth cannot be sampled. Second, the sampling rate must be at least 2 times the highest frequency, not the bandwidth. If the analog signal is low-pass, the bandwidth and the highest frequency are the same value. If the analog signal is bandpass, the bandwidth value is lower than the value of the maximum frequency.

Quantization

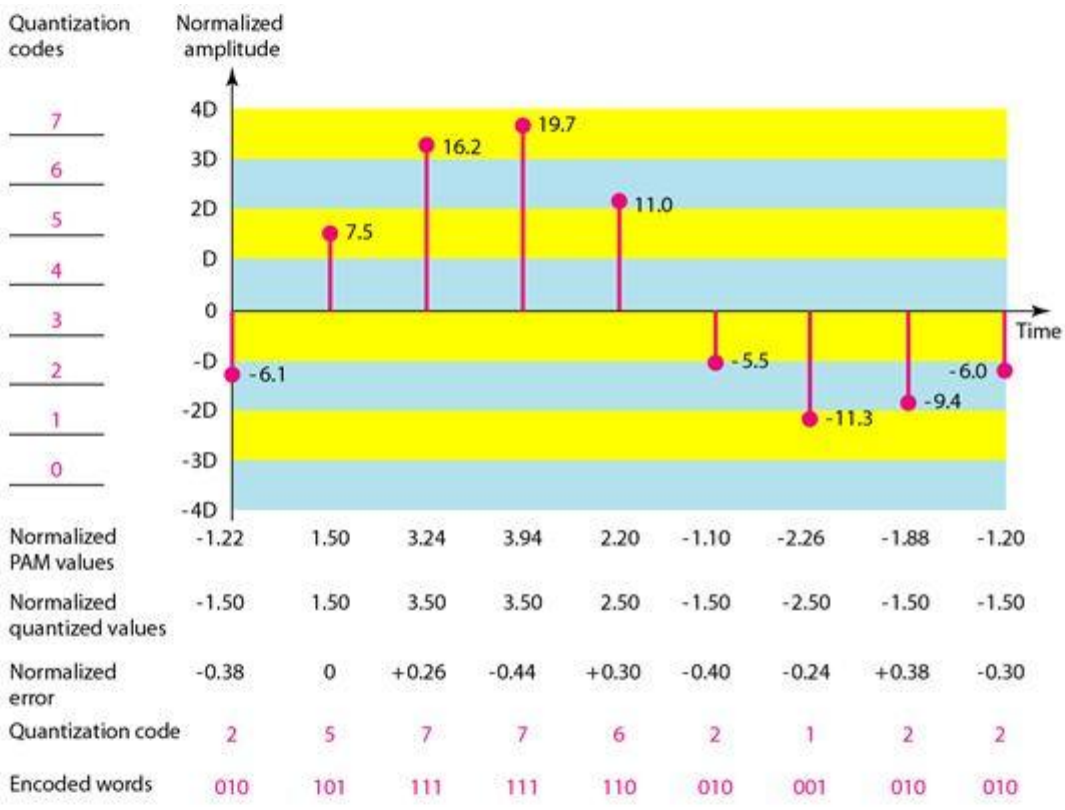
The result of sampling is a series of pulses with amplitude values between the maximum and minimum amplitudes of the signal. The set of amplitudes can be infinite with non-integral values between the two limits. These values cannot be used in the encoding process. The following are the steps in quantization:

1. We assume that the original analog signal has instantaneous amplitudes between Vmin and Vmax
2. We divide the range into L zones, each of height Δ (delta).

$$\Delta=(V_{max}-V_{min})/L$$

3. We assign quantized values of 0 to L - I to the midpoint of each zone.
4. We approximate the value of the sample amplitude to the quantized values.

As a simple example



assume that we have a sampled signal and the sample amplitudes are between -20 and +20 V.

We decide to have eight levels (L = 8). This means that Δ =5 V.

We have shown only nine samples using ideal sampling (for simplicity). The value at the top of each sample in the graph shows the actual amplitude. In the chart, the first row is the normalized value for each sample (actual amplitude/Δ).

The quantization process selects the quantization value from the middle of each zone. This means that the normalized quantized values (second row) are different from the normalized amplitudes. The difference is called the normalized error (third row). The fourth row is the quantization code for each sample based on the quantization levels at the left of the graph. The encoded words (fifth row) are the final products of the conversion.

Quantization Levels:

In the above example, we showed eight quantization levels. The choice of L, the number of levels, depends on the range of the amplitudes of the analog signal and how accurately we need to recover the signal. If the amplitude of a signal fluctuates between two values only, we need only two levels; if the signal, like voice, has many amplitude values, we need more quantization levels. In audio digitizing, L is normally chosen to be 256; in video it is normally thousands. Choosing lower values of L increases the quantization error if there is a lot of fluctuation in the signal.

Quantization Error:

One important issue is the error created in the quantization process. Quantization is an approximation process. The input values to the quantizer are the real values; the output values are the approximated values. The output values are chosen to be the middle value in the zone. If the input value is also at the middle of the zone, there is no quantization error; otherwise, there is an error. In the previous example, the normalized amplitude of the third sample is 3.24, but the normalized quantized value is 3.50. This means that there is an error of +0.26. The value of the error for any sample is less than $\Delta/2$. In other words, we have $\Delta/2 <= \text{error} <= \Delta/2$.

Uniform Versus Non uniform Quantization:

For many applications, the distribution of the instantaneous amplitudes in the analog signal is not uniform. Changes in amplitude often occur more frequently in the lower amplitudes than in the higher ones. For these types of applications it is better to use nonuniform zones. In other words, the height of Δ is not fixed; it is greater near the lower amplitudes and less near the higher amplitudes.

Nonuniform quantization can also be achieved by using a process called companding and expanding. The signal is companded at the sender before conversion; it is expanded at the receiver after conversion. Companding means reducing the instantaneous voltage amplitude for large values; expanding is the opposite process. Companding gives greater weight to strong signals and less weight to weak ones. It has been proved that nonuniform quantization effectively reduces the SNRdB of quantization.

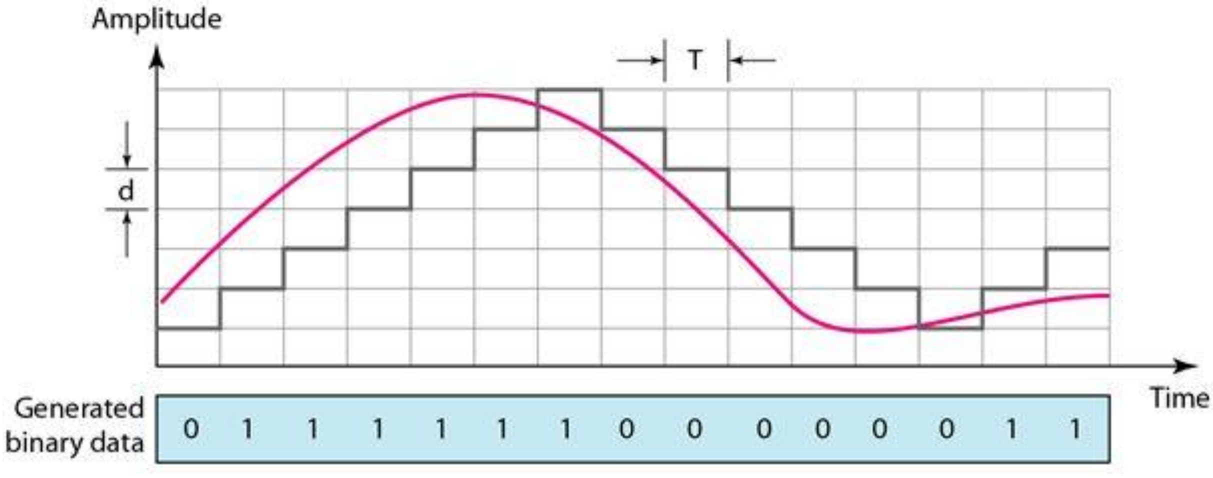
Encoding

The last step in PCM is encoding. After each sample is quantized and the number of bits per sample is decided, each sample can be changed to an nb-bit code word. In the above figure the encoded words are shown in the last row. A quantization code of 2 is encoded as 010; 5 is encoded as 101; and so on. Note that the number of bits for each sample is determined from the number of quantization levels. If the number of quantization levels is L, the number of bits is $nb = \log_2 L$. In our example L is 8 and nb is therefore 3. The bit rate can be found from the formula.

Bit-rate = Sampling rate X Number of bites per sample= fs X nb

II. Delta Modulation (DM)

PCM is a very complex technique. Number of other techniques has been developed to reduce the complexity of PCM. The simplest is delta modulation. PCM finds the value of the signal amplitude for each sample; DM finds the change from the previous sample. The following figure shows the process. Note that there are no code words here; bits are sent one after another.



Modulator

The modulator is used at the sender site to create a stream of bits from an analog signal. The process records the small positive or negative changes, called delta . If the delta is positive, the process records a 1; if it is negative, the process records a 0. However, the process needs a base against which the analog signal is compared. The modulator builds a second signal that resembles a staircase. Finding the change is then reduced to comparing the input signal with the gradually made staircase signal.

The modulator, at each sampling interval, compares the value of the analog signal with the last value of the staircase signal. If the amplitude of the analog signal is larger, the next bit in the digital data is 1; otherwise, it is O. The output of the comparator, however, also makes the staircase itself. If the next bit is I, the staircase maker moves the last point of the staircase signal up; it the next bit is 0, it moves it down. Note that we need a delay unit to hold the staircase function for a period between two comparisons.

Demodulator

The demodulator takes the digital data and, using the staircase maker and the delay unit, creates the analog signal. The created analog signal, however, needs to pass through a low-pass filter for smoothing.

Adaptive DM

A better performance can be achieved if the value of is not fixed. In adaptive delta modulation, the value of changes according to the amplitude of the analog signal.

Quantization Error

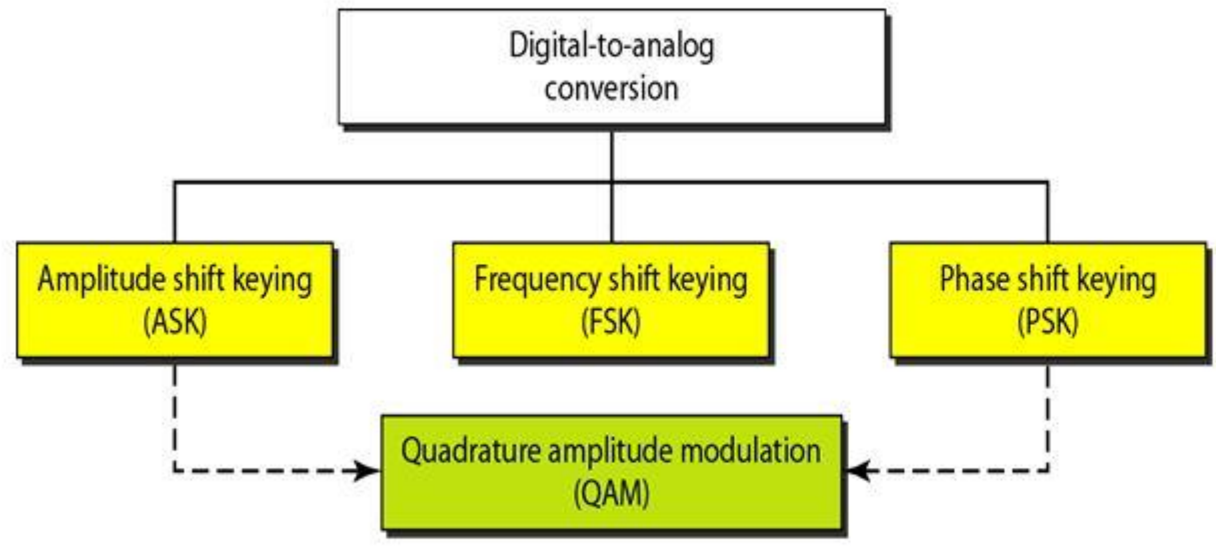
It is obvious that DM is not perfect. Quantization error is always introduced in the process. The quantization error of DM, however, is much less than that for PCM.

Digital to Analog Conversion Techniques:

Digital-to-analog conversion is the process of changing one of the characteristics of an analog signal based on the information in digital data.

A sine wave is defined by three characteristics: amplitude, frequency, and phase. When we change anyone of these characteristics, we create a different version of that wave. So, by changing one characteristic of a simple electric signal, we can use it to represent digital data.

There are three mechanisms for modulating digital data into an analog signal: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK). In addition, there is a fourth (and better) mechanism that combines changing both the amplitude and phase, called quadrature amplitude modulation (QAM).



Bandwidth

The required bandwidth for analog transmission of digital data is proportional to the signal rate except for FSK, in which the difference between the carrier signals needs to be added.

Carrier Signal

In analog transmission, the sending device produces a high-frequency signal that acts as a base for the information signal. This base signal is called the carrier signal or carrier frequency. The receiving device is tuned to the frequency of the carrier signal that it expects from the sender. Digital information then changes the carrier signal by modifying one or more of its characteristics (amplitude, frequency, or phase). This kind of modification is called modulation (shift keying).

1.Amplitude Shift-Keying

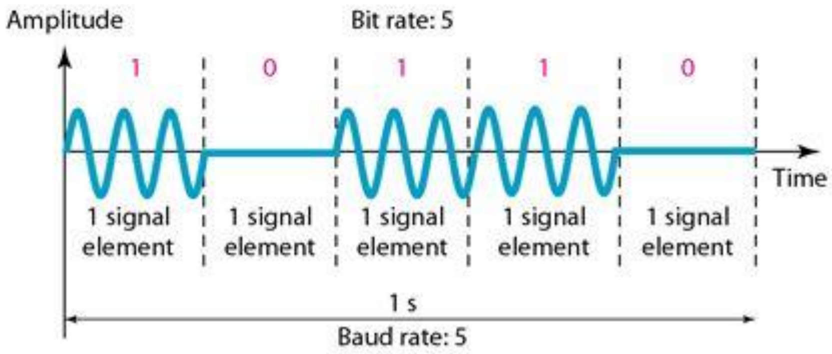
In amplitude shift keying, the amplitude of the carrier signal is varied to create signal elements. In ASK, the two binary values are represented by two different amplitudes of the carrier frequency. Usually, one of the amplitudes is zero and the other by the absence of the carrier. The resulting transmitted signal becomes as follows. Both frequency and phase remain constant while the amplitude changes.

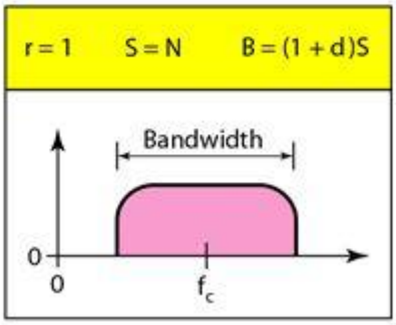
ASK: S(t)= $\begin{cases} ACos(2\pi f_0 t) & for \quad binary 1 \\ 0 & for \quad binary 0 \end{cases}$

ASK is susceptible to sudden gain changes and is a rather inefficient modulation technique.

Binary ASK (BASK)

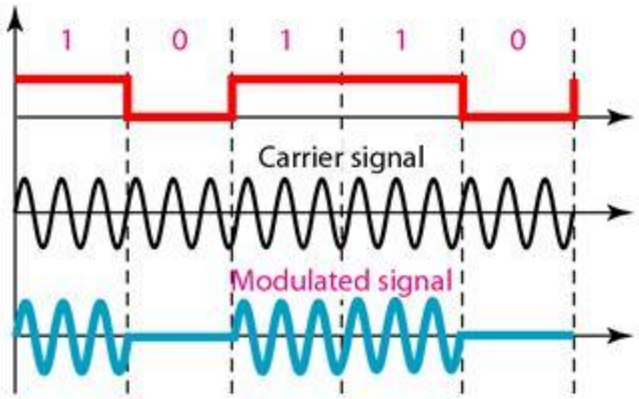
ASK is normally implemented using only two levels. This is referred to as binary amplitude shift keying or on-off keying (OOK). The peak amplitude of one signal level is 0; the other is the same as the amplitude of the carrier frequency. The following figure gives a conceptual view of binary ASKS.

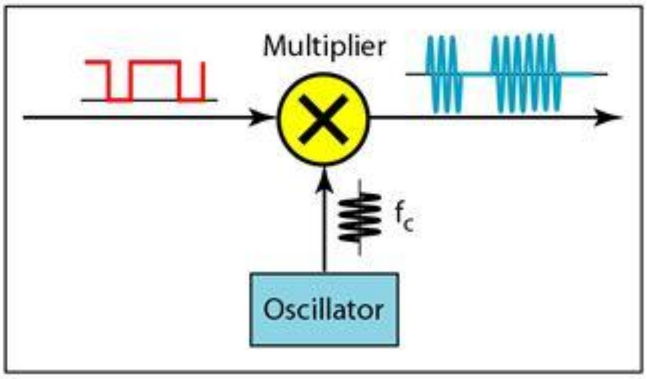




Implementation:

If digital data are presented as a unipolar NRZ digital signal with a high voltage of 1V and a low voltage of 0V, the implementation can achieved by multiplying the NRZ digital signal by the carrier signal coming from an oscillator which is represented in the following figure. When the amplitude of the NRZ signal is 1, the amplitude of the carrier frequency is held; when the amplitude of the NRZ signal is 0, the amplitude of the carrier frequency is zero.





Bandwidth for ASK:

The carrier signal is only one simple sine wave, but the process of modulation produces a non-periodic composite signal. This signal has a continuous set of frequencies. As we expect, the bandwidth is proportional to the signal rate (baud rate).

However, there is normally another factor involved, called d, which depends on the modulation and filtering process. The value of d is between 0 and 1. This means that the bandwidth can be expressed as shown, where S is the signal rate and the B is the bandwidth.

$$B = (1 + d) \times S$$

The formula shows that the required bandwidth has a minimum value of S and a maximum value of 2S. The most important point here is the location of the bandwidth. The middle of the bandwidth is where fc the carrier frequency, is located. This means if we have a bandpass channel available, we can choose our fc so that the modulated signal occupies that bandwidth. This is in fact the most important advantage of digital-to- analog conversion.

2. Frequency Shift Keying

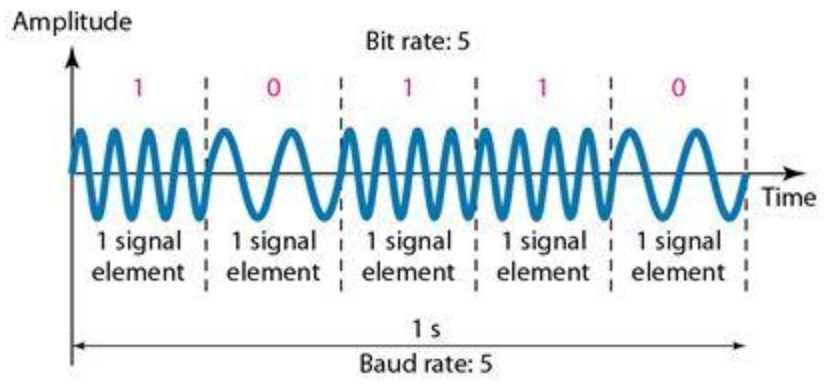
In frequency shift keying, the frequency of the carrier signal is varied to represent data. In FSK the higher frequency is used to represent binary 0, while the lower represents binary 1. The frequency of the modulated signal is constant for the duration of one signal element, but changes for the next signal element if the data element changes. Both peak amplitude and phase remain constant for all signal elements.

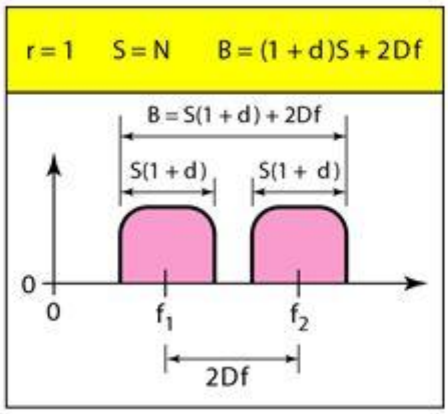
Binary FSK (BFSK)

One way to think about binary FSK (or BFSK) is to consider two carrier frequencies. In the following Figure, we have selected two carrier frequencies f1 and f2. We use the first carrier if the data element is 0; we use the second if the data element is 1. The resulting transmitted signal for one bit time is

BFSK: $S(t) = \begin{cases} ACos(2\pi f_1 t) & binary 1 \\ ACos(2\pi f_2 t) & Binary 0 \end{cases}$

where f_1 and f_2 are typically offset from the carrier frequency f_c by equal but opposite amounts.





-

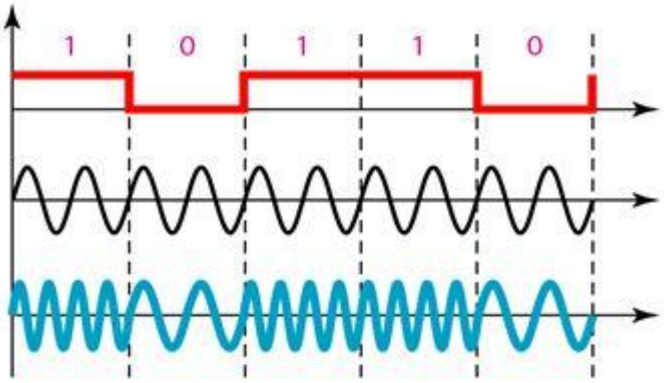
The above figure shows, the middle of one bandwidth is f1 and the middle of the other is f2. Both f1 and f2 are Δf apart from the midpoint between the two bands. The difference between the two frequencies is 2Δf.

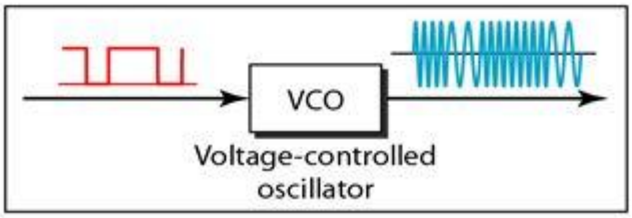
Implementation:

There are two implementations of BFSK: non-coherent and coherent. In non-coherent BFSK, there may be discontinuity in the phase when one signal element ends and the next begins. In coherent BFSK, the phase continues through the boundary of two signal elements. Non-coherent BFSK can be implemented by treating BFSK as two ASK modulations and using two carrier frequencies. Coherent BFSK can be implemented by using one voltage-controlled oscillator (VCO) that changes its frequency according to the input voltage.

The following figure shows the simplified idea behind the second implementation. The input to the oscillator is the unipolar NRZ signal. When the amplitude of NRZ is zero, the oscillator keeps its regular frequency; when the amplitude is positive, the frequency is increased.

-





Bandwidth for BFSK:

The above figure shows the bandwidth of FSK. Again the carrier signals are only simple sine waves, but the modulation creates a non-periodic composite signal with continuous frequencies. We can think of FSK as two ASK signals, each with its own carrier frequency f1 and f2. If the difference between the two frequencies is 2Δf, then the required bandwidth is

B=(l+d)XS+2Δf

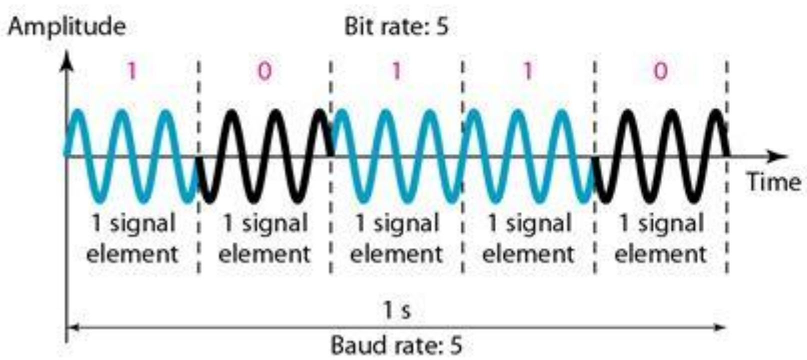
3. Phase Shift Keying:

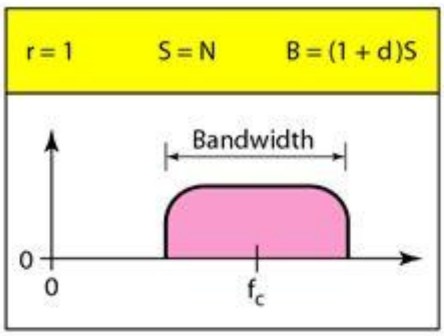
In phase shift keying, the phase of the carrier is varied to represent two or more different signal elements. Both peak amplitude and frequency remain constant as the phase changes. There are many variations of PSK. These are Two-level PSK or Binary PSK, Four-level PSK or Quadrature PSK and Multi-level PSK.

Binary PSK

The simplest PSK is binary PSK, in which we have only two signal elements, one with a phase of 0°, and the other with a phase of 180°. The following figure gives a conceptual view of PSK. Binary PSK is as simple as binary ASK with one big advantage-it is less susceptible to noise. In ASK, the criterion for bit detection is the amplitude of the signal. But in PSK, it is the phase. Noise can change the amplitude easier than it can change the phase. In other words, PSK is less susceptible to noise than ASK. PSK is superior to FSK because we do not need two carrier signals. The resulting transmitted signal for one bit time is:

BPSK: $S(t) = \begin{cases} ACos(2\pi f_c t) \\ ACos(2\pi f_c t + \pi) \end{cases} = \begin{cases} ACos(2\pi f_c t) & binary1 \\ -ACos(2\pi f_c t) & binary0 \end{cases}$



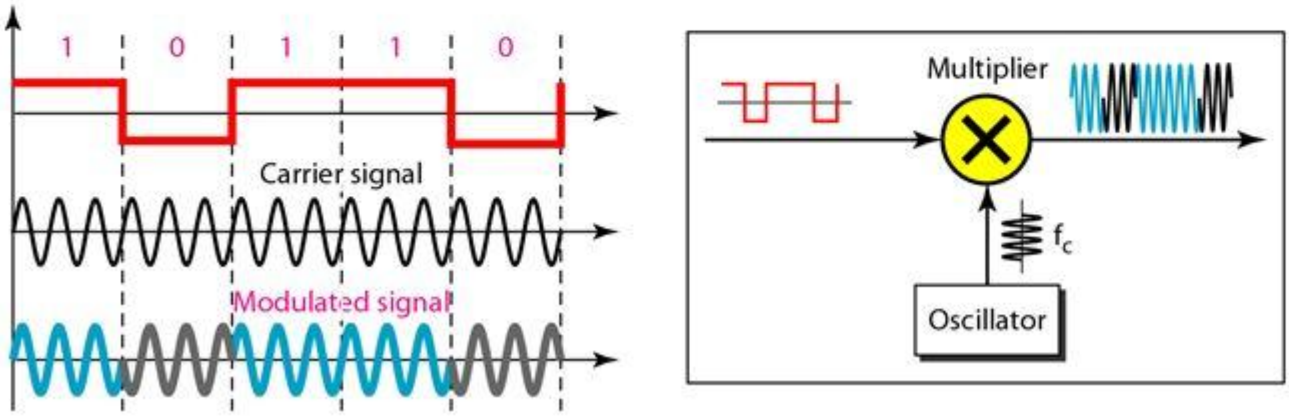


Bandwidth:

The bandwidth is the same as that for binary ASK, but less than that for BFSK. No bandwidth is wasted for separating two carrier signals.

Implementation:

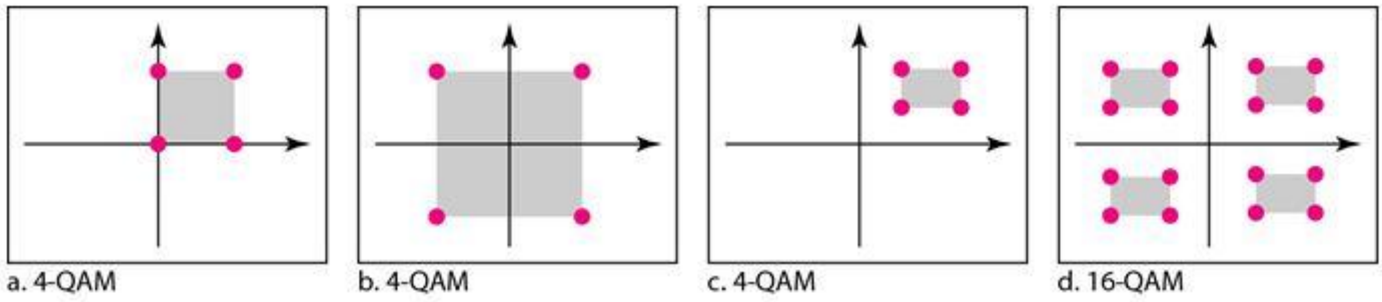
The implementation of BPSK is as simple as that for ASK. The reason is that the signal element with phase 180° can be seen as the complement of the signal element with phase 0°. This gives us a clue on how to implement BPSK. We use a polar NRZ signal instead of a unipolar NRZ signal, as shown in the following figure . The polar NRZ signal is multiplied by the carrier frequency. The 1 bit (positive voltage) is represented by a phase starting at 0° the 0 bit (negative voltage) is represented by a phase starting at 180°.



Quadrature Amplitude Modulation (QAM)

PSK is limited by the ability of the equipment to distinguish small differences in phase. This factor limits its potential bit rate. So far, we have been altering only one of the three characteristics of a sine wave at a time; but what if we alter two? Why not combine ASK and PSK? The idea of using two carriers, one in-phase and the other quadrature, with different amplitude levels for each carrier is the concept behind quadrature amplitude modulation (QAM).

The possible variations of QAM are numerous. The following figure shows some of these schemes. In the following figure Part a shows the simplest 4-QAM scheme (four different signal element types) using a unipolar NRZ signal to modulate each carrier. This is the same mechanism we used for ASK (OOK). Part b shows another 4-QAM using polar NRZ, but this is exactly the same as QPSK. Part c shows another QAM-4 in which we used a signal with two positive levels to modulate each of the two carriers. Finally, Part – d shows a 16-QAM constellation of a signal with eight levels, four positive and four negative.



Analog To Analog Conversion Techniques

Analog-to-analog conversion, or analog modulation, is the representation of analog information by an analog signal. Modulation is needed if the medium is bandpass in nature or if only a bandpass channel is available to us.

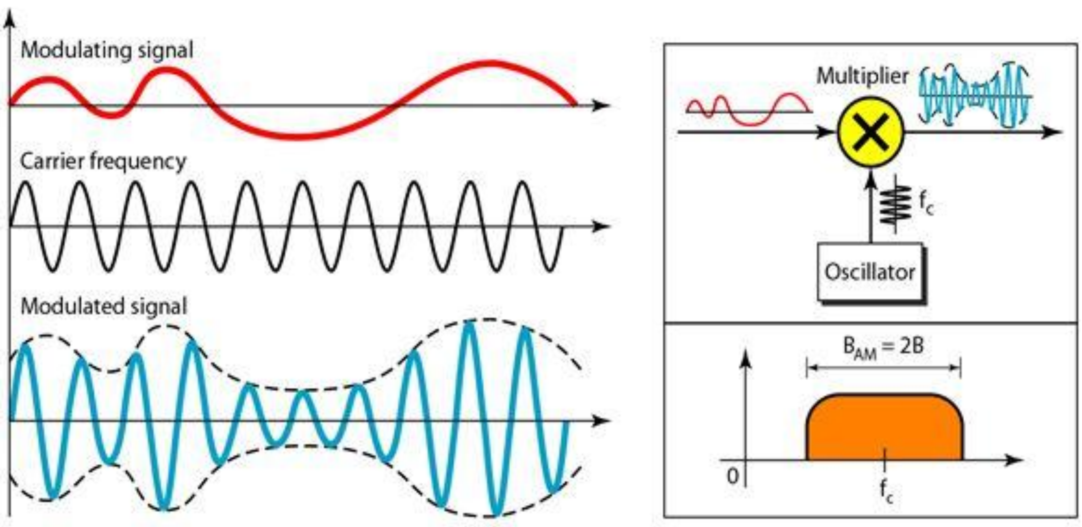
An example is radio. The government assigns a narrow bandwidth to each radio station. The analog signal produced by each station is a low-pass signal, all in the same range. To be able to listen to different stations, the low-pass signals need to be shifted, each to a different range.
Analog- to digital conversion can be accomplished in three ways, namely:

(i) Amplitude Modulation (AM)

(ii) Frequency Modulation (FM)
(iii) Phase Modulation (PM)

1. Amplitude Modulation:

In AM transmission, the carrier signal is modulated so that its amplitude varies with the changing amplitudes of the modulating signal. The frequency and phase of the carrier remain the same. Only the amplitude changes to follow variations in the information. The following figure shows how this concept works. The modulating signal is the envelope of the carrier.



AM is normally implemented by using a simple multiplier because the amplitude of the carrier signal needs to be changed according to the amplitude of the modulating signal.

AM Bandwidth:

The modulation creates a bandwidth that is twice the bandwidth of the modulating signal and covers a range centered on the carrier frequency. However, the signal components above and below the carrier frequency carry exactly the same information. For this reason, some implementations discard one-half of the signals and cut the bandwidth in half.

The total bandwidth required for AM can be determined from the bandwidth of the audio signal:

BAM =2B

Standard Bandwidth allocation for AM Radio:

The bandwidth of an audio signal (speech and music) is usually 5 kHz. Therefore, an AM radio station needs a bandwidth of 10kHz. In fact, the Federal Communications Commission (FCC) allows 10 kHz for each AM station.
AM stations are allowed carrier frequencies anywhere between 530 and 1700 kHz (1.7 MHz). However, each station's carrier frequency must be separated from those on either side of it by at least 10 kHz (one AM bandwidth) to avoid interference. If one station uses a carrier frequency of 1100 kHz, the next station's carrier frequency cannot be lower than 1110 kHz.

2. Frequency Modulation

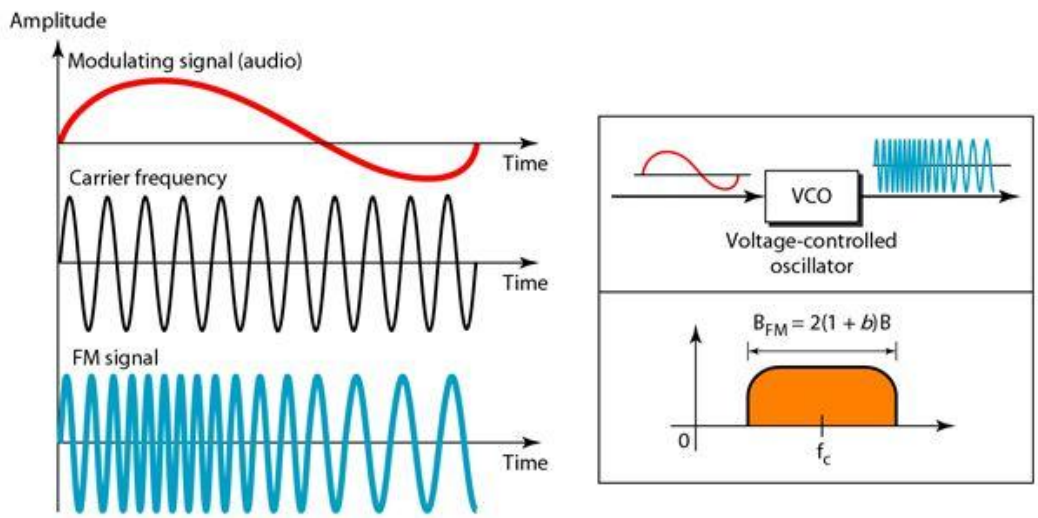
In FM transmission, the frequency of the carrier signal is modulated to follow the changing voltage level (amplitude) of the modulating signal. The peak amplitude and phase of the carrier signal remain constant, but as the amplitude of the information signal changes, the frequency of the carrier changes correspondingly.

The following figure shows the relationships of the modulating signal, the carrier signal, and the resultant FM signal. FM is normally implemented by using a voltage-controlled oscillator as with FSK. The frequency of the oscillator changes according to the input voltage which is the amplitude of the modulating signal.

FM Bandwidth

The actual bandwidth is difficult to determine exactly, but it can be shown empirically that it is several times that of the analog signal or $2(1 + \beta)B$ where β is a factor depends on modulation technique with a common value of 4.

-



Standard Bandwidth allocation for FM Radio:

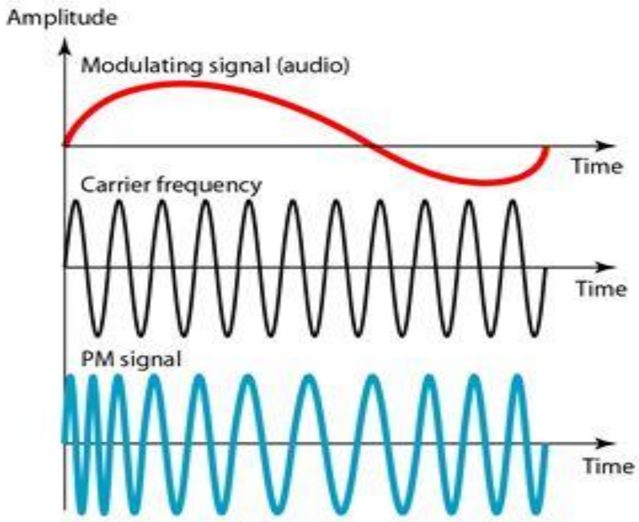
The bandwidth of an audio signal (speech and music) broadcast in stereo is almost 15 kHz. The FCC allows 200 kHz (0.2 MHz) for each station. This mean β = 4 with some extra guard band. FM stations are allowed carrier frequencies anywhere between 88 and 108 MHz. Stations must be separated by at least 200 kHz to keep their bandwidths from overlapping.

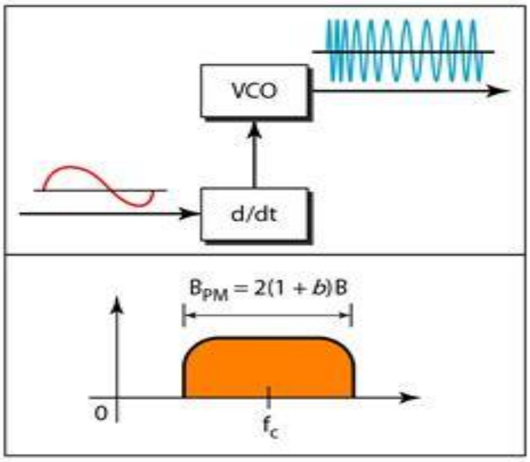
To create even more privacy, the FCC requires that in a given area, only alternate bandwidth allocations may be used. The others remain unused to prevent any possibility of two stations interfering with each other. Given 88 to 108 MHz as a range, there are 100 potential PM bandwidths in an area, of which 50 can operate at any one time.

3. Phase Modulation:

In PM transmission, the phase of the carrier signal is modulated to follow the changing voltage level (amplitude) of the modulating signal. The peak amplitude and frequency of the carrier signal remain constant, but as the amplitude of the information signal changes, the phase of the carrier changes correspondingly. It is proved mathematically that PM is the same as FM with one difference.

In FM, the instantaneous change in the carrier frequency is proportional to the amplitude of the modulating signal; in PM the instantaneous change in the carrier frequency is proportional to the derivative of the amplitude of the modulating signal. The following figure shows the relationships of the modulating signal, the carrier signal, and the resultant PM signal.





PM is normally implemented by using a voltage-controlled oscillator along with a derivative. The frequency of the oscillator changes according to the derivative of the input voltage which is the amplitude of the modulating signal.

PM Bandwidth

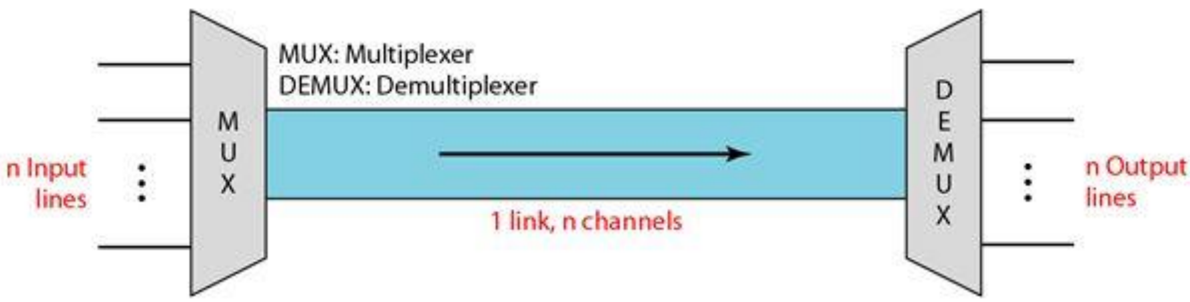
The actual bandwidth is difficult to determine exactly, but it can be shown empirically that it is several times that of the analog signal. Although, the formula shows the same bandwidth for FM and PM, the value of β is lower in the case of PM (around 1 for narrowband and 3 for wideband).

MULTIPLEXING (MUX)

Multiplexing and Types of Multiplexing

Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single data link. Whenever the bandwidth of a medium linking two devices is greater than the bandwidth needs of the devices, the link can be shared. In a multiplexed system, n lines share the bandwidth of one link.

The following figure shows the basic format of a multiplexed system. The lines on the left direct their transmission streams to a multiplexer (MUX), which combines them into a single stream (many-to-one). At the receiving end, that stream is fed into a demultiplexer (DEMUX), which separates the stream back into its component transmissions (one-to-many) and directs them to their corresponding lines.



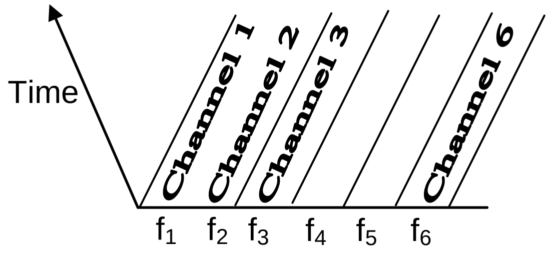
The three basic multiplexing techniques are: (i) Frequency Division Multiplexing, (ii) Wavelength Division Multiplexing and Time Division Multiplexing.

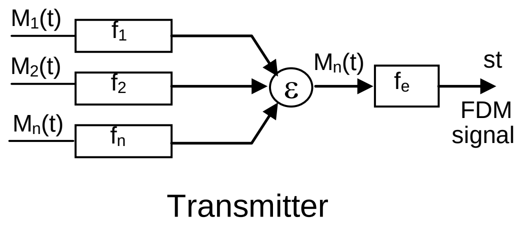
The first two are techniques designed for analog signals, while the third is for digital signal.

1. Frequency-Division Multiplexing

Frequency-division multiplexing (FDM) is an analog technique that can be applied when the bandwidth of a link (in hertz) is greater than the combined bandwidths of the signals to be transmitted.

In FDM, signals generated by each sending device modulate different carrier frequencies. These modulated signals are then combined into a single composite signal that can be transported by the link. Carrier frequencies are separated by sufficient bandwidth to accommodate the modulated signal. These bandwidth ranges are the channels through which the various signals travel.





Channels can be separated by strips of unused bandwidth-guard bands-to prevent signals from overlapping. The following figure gives a conceptual view of FDM. In this illustration, the transmission path is divided into three parts, each representing a channel that carries one transmission._

Multiplexing Process:
The following figure is a conceptual illustration of the multiplexing process. Each source generates a signal of a similar frequency range. Inside the multiplexer, these similar signals modulates different carrier frequencies (f1, f2 and f3). The resulting modulated signals are then combined into a single composite signal that is sent out over a media link that has enough bandwidth to accommodate it.

Demultiplexing Process:
The demultiplexer uses a series of filters to decompose the multiplexed signal into its constituent component signals. The individual signals are then passed to a demodulator that separates them from their carriers and passes them to the output lines.

Applications of FDM:
To maximize the efficiency of their infrastructure, telephone companies have traditionally multiplexed signals from lower-bandwidth lines onto higher-bandwidth lines.

A very common application of FDM is AM and FM radio broadcasting.

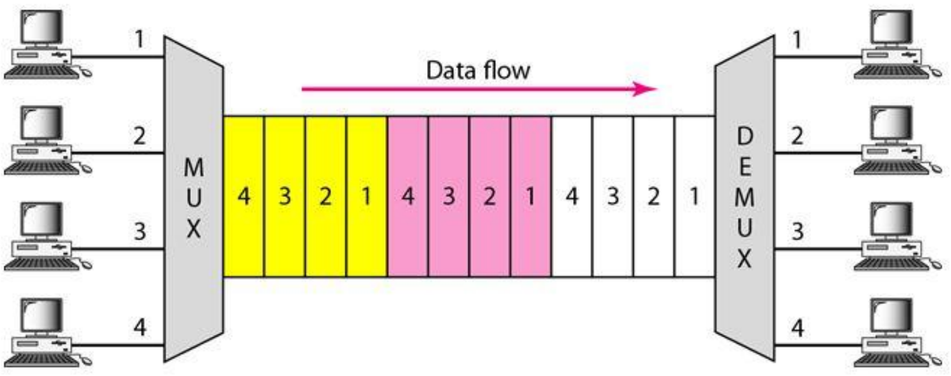
The first generation of cellular telephones (still in operation) also uses FDM.

Implementation:

FDM can be implemented very easily. In many cases, such as radio and television broadcasting, there is no need for a physical multiplexer or demultiplexer. As long as the stations agree to send their broadcasts to the air using different carrier frequencies, multiplexing is achieved. In other cases, such as the cellular telephone system, a base station needs to assign a carrier frequency to the telephone user. There is not enough bandwidth in a cell to permanently assign a bandwidth range to every telephone user. When a user hangs up, her or his bandwidth is assigned to another caller.

Time-Division Multiplexing

Time-division multiplexing (TDM) is a digital process that allows several connections to share the high bandwidth of a linle Instead of sharing a portion of the bandwidth as in FDM, time is shared. Each connection occupies a portion of time in the link.

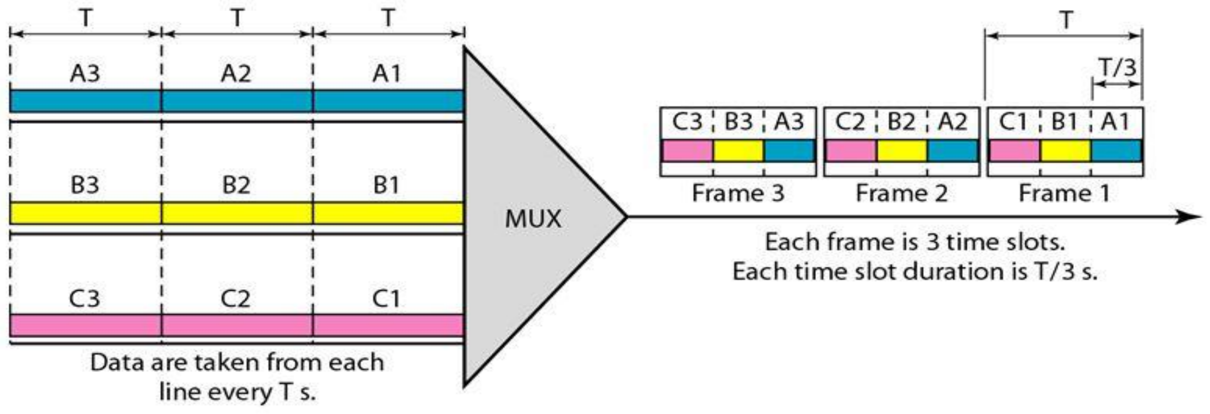


We can divide TDM into two different schemes: **synchronous and statistical**.

Synchronous Time-Division Multiplexing:

In synchronous TDM, each input connection has an allotment in the output even if it is not sending data. In synchronous TDM, the data flow of each input connection is divided into units, where each input occupies one input time slot.

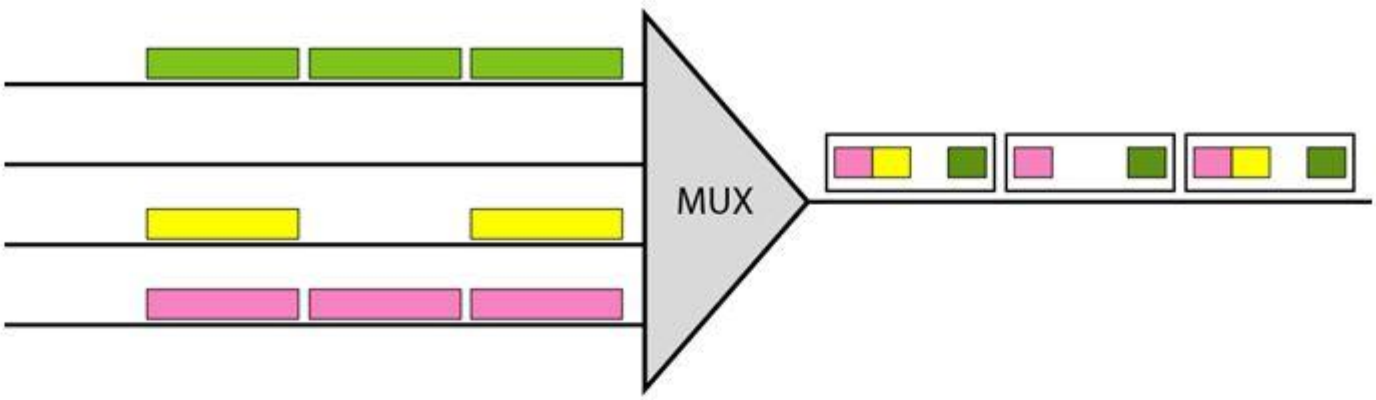
A unit can be 1 bit, one character, or one block of data. Each input unit becomes one output unit and occupies one output time slot. However, the duration of an output time slot is n times shorter than the duration of an input time slot. If an input time slot is T s, the output time slot is T/n s, where n is the number of connections. In other words, a unit in the output connection has a shorter duration; it travels faster. The following figure shows an example of synchronous TDM where n is 3.



In synchronous TDM, a round of data units from each input connection is collected into a frame. If we have n connections, a frame is divided into n time slots and one slot is allocated for each unit, one for each input line. If the duration of the input unit is T, the duration of each slot is Tin and the duration of each frame is T.
Time slots are grouped into frames. A frame consists of one complete cycle of time slots, with one slot dedicated to each sending device. In a system with n input lines, each frame has n slots, with each slot allocated to carrying data from a specific input line.

Interleaving
TDM can be visualized as two fast-rotating switches, one on the multiplexing side and the other on the demultiplexing side. The switches are synchronized and rotate at the same speed, but in opposite directions. On the multiplexing side, as the switch opens in front of a connection, that connection has the opportunity to send a unit onto the path. This process is called interleaving. On the demultiplexing side, as the switch opens in front of a connection, that connection has the opportunity to receive a unit from the path.

Empty Slots
Synchronous TDM is not as efficient as it could be. If a source does not have data to send, the corresponding slot in the output frame is empty. The following figure shows a case in which one of the input lines has no data to send and one slot in another input line has discontinuous data.



The first output frame has three slots filled, the second frame has two slots filled, and the third frame has three slots filled. No frame is full. We learn in the next section that statistical TDM can improve the efficiency by removing the empty slots from the frame.

Data Rate Management

One problem with TDM is how to handle a disparity in the input data rates. If data rates are not the same, three strategies, or a combination of them, can be used. The three different strategies are multilevel multiplexing, multiple-slot allocation, and pulse stuffing.

Multilevel Multiplexing:

Multilevel multiplexing is a technique used when the data rate of an input line is a multiple of others. For example, if we have two inputs of 20 kbps and three inputs of 40 kbps. The first two input lines can be multiplexed together to provide a data rate equal to the last three. A second level of multiplexing can create an output of 160 kbps.

Multiple-Slot Allocation:

Sometimes it is more efficient to allot more than one slot in a frame to a single input line. For example, we might have an input line that has a data rate that is a multiple of another input. The input line with a 50-kbps data rate can be given two slots in the output. We insert a serial-to-parallel converter in the line to make two inputs out of one.

Pulse Stuffing:

Sometimes the bit rates of sources are not multiple integers of each other. Therefore, neither of the above two techniques can be applied. One solution is to make the highest input data rate the dominant data rate and then add dummy bits to the input lines with lower rates. This will increase their rates. This technique is called pulse stuffing, bit padding, or bit stuffing. The input with a data rate of 46 is pulse-stuffed to increase the rate to 50 kbps. Now multiplexing can take place.

Frame Synchronizing

The implementation of TDM is not as simple as that of FDM. Synchronization between the multiplexer and demultiplexer is a major issue. If the, multiplexer and the demultiplexer are not synchronized, a bit belonging to one channel may be received by the wrong channel.

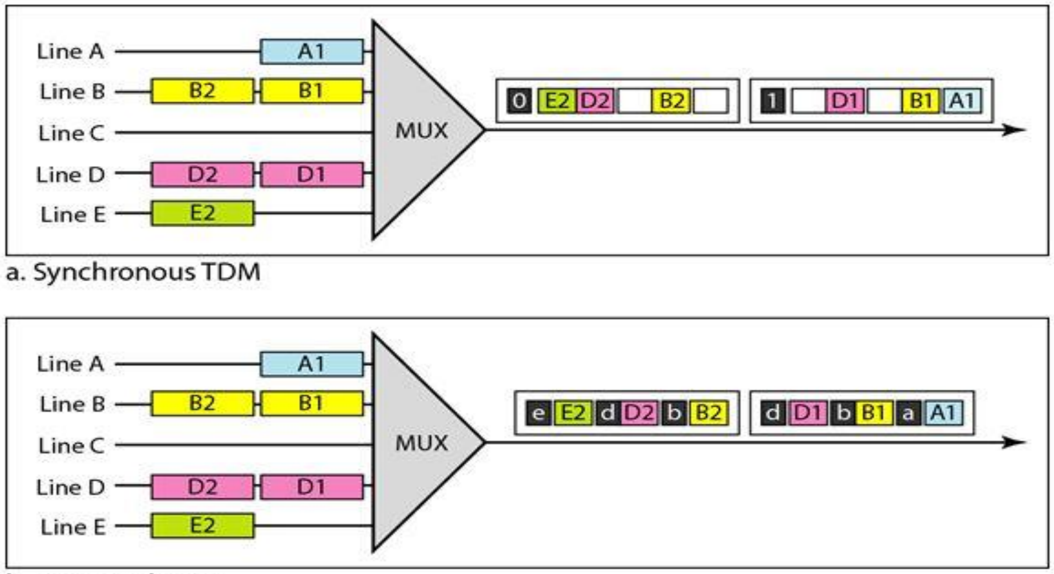
For this reason, one or more synchronization bits are usually added to the beginning of each frame. These bits, called framing bits, follow a pattern, frame to frame, that allows the demultiplexer to synchronize with the incoming stream so that it can separate the time slots accurately. In most cases, this synchronization information consists of 1 bit per frame, alternating between 0 and 1.

Statistical Time-Division Multiplexing:

In synchronous TDM, each input has a reserved slot in the output frame. This can be inefficient if some input lines have no data to send. In statistical time-division multiplexing, slots are dynamically allocated to improve bandwidth efficiency. Only when an input line has a slot's worth of data to send is it given a slot in the output frame.

In statistical multiplexing, the number of slots in each frame is less than the number of input lines. The multiplexer checks each input line in round robin fashion. It allocates a slot for an input line if the line has data to send otherwise it skips the line and checks the next line.

The following figure shows a synchronous and a statistical TDM example. In the former, some slots are empty because the corresponding line does not have data to send. In the latter, however, no slot is left empty as long as there are data to be sent by any input line.



Addressing:

The above figure also shows a major difference between slots in synchronous TDM and statistical TDM. An output slot in synchronous TDM is totally occupied by data, in statistical TDM, a slot needs to carry data as well as the address of the destination.

In synchronous TDM, there is no need for addressing. Synchronization and preassigned relationships between the inputs and outputs serve as an address. We know, for example, that input 1 always goes to input 1. If the multiplexer and the demultiplexer are synchronized, this is guaranteed. In statistical multiplexing, there is no fixed relationship between the inputs and outputs because there are no preassigned or reserved slots. We need to include the address of the receiver inside each slot to show where it is to be delivered.

The addressing in its simplest form can be n bits to define N different output lines with n =log 2 n. For example, for eight different output lines, we need a 3-bit address.

Slot Size

Since a slot carries both data and an address in statistical TDM, the ratio of the data size to address size must be reasonable to make transmission efficient. For example, it would be inefficient to send 1 bit per slot as data when the address is 3 bits. This would mean an overhead of 300 percent. In statistical TDM, a block of data is usually many bytes while the address is just a few bytes.
No Synchronization Bit
There is another difference between synchronous and statistical TDM, but this time it is at the frame level. The frames in statistical TDM need not be synchronized, so we do not need synchronization bits.

Bandwidth

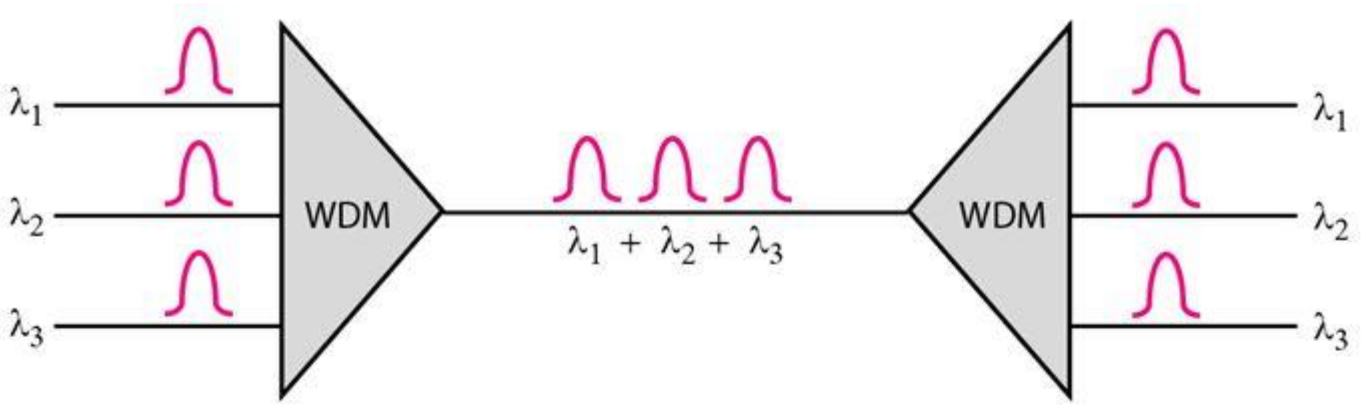
In statistical TDM, the capacity of the link is normally less than the sum of the capacities of each channel. The designers of statistical TDM define the capacity of the link based on the statistics of the load for each channel. If on average only x percent of the input slots are filled, the capacity of the link reflects this. Of course, during peak times, some slots need to wait.

Wavelength-Division Multiplexing

Wavelength-division multiplexing (WDM) is designed to use the high-data-rate capability of fiber-optic cable. The optical fiber data rate is higher than the data rate of metallic transmission cable. Using

a fiber-optic cable for one single line wastes the available bandwidth. Multiplexing allows us to combine several lines into one.
WDM is conceptually the same as FDM, except that the multiplexing and demultiplexing involve optical signals transmitted through fiber-optic channels.

The following figure gives a conceptual view of a WDM multiplexer and demultiplexer. Very narrow bands of light from different sources are combined to make a wider band of light. At the receiver, the signals are separated by the demultiplexer.



In this method, we combine multiple light sources into one single light at the multiplexer and do the reverse at the demultiplexer. The combining and splitting of light sources are easily handled by a prism. Recall from basic physics that a prism bends a beam of light based on the angle of incidence and the frequency. Using this technique, a multiplexer can be made to combine several input beams of light, each containing a narrow band of frequencies, into one output beam of a wider band of frequencies. A demultiplexer can also be made to reverse the process.

TRANSMISSION MODES

Data is transmitted between two digital devices on the network in the form of bits. Transmission mode refers to the mode used for transmitting the data. The transmission medium may be capable of sending only a single bit in unit time or multiple bits in unit time.

When a single bit is transmitted in unit time the transmission mode used is Serial Transmission and when multiple bits are sent in unit time the transmission mode used is called Parallel transmission.

Types of Transmission Modes:

There are two basic types of transmission modes (Serial and Parallel) as shown in the figure below. Serial transmission is further categorized into Synchronous and Asynchronous Serial transmission.

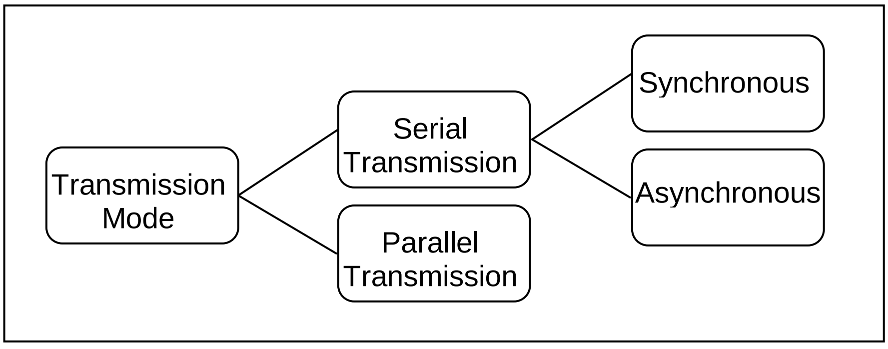


Fig.1.1 Types of Transmission Modes

Parallel Transmission

It involves simultaneous transmission of *N* bits of *N* different channels. With the parallel transmission of data all the bits making up a character are transmitted simultaneously over separate conductors. The number of conductors required for a parallel interface is known as the **bus width.** Since several conductors are necessary the parallel transmission system is only economic over fairly short distances.

Parallel Transmission increases transmission speed by a factor of *N* over serial transmission. Disadvantage of parallel transmission is the cost involved, N channels have to be used, hence, it can be used for short distance communication only.

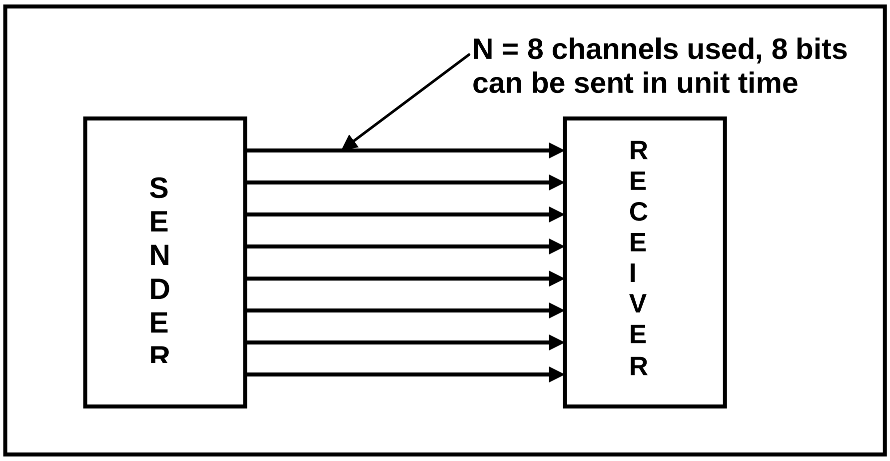


Fig.1.2 Parallel Transmission of Data over N = 8 channels

Example of Parallel Transmission is the communication between CPU and the Projector.

Serial Transmission

In Serial Transmission, as the name suggest data is transmitted serially, i.e. bit by bit, one bit at a time. Since only one bit has to be sent in unit time only a single channel is required.

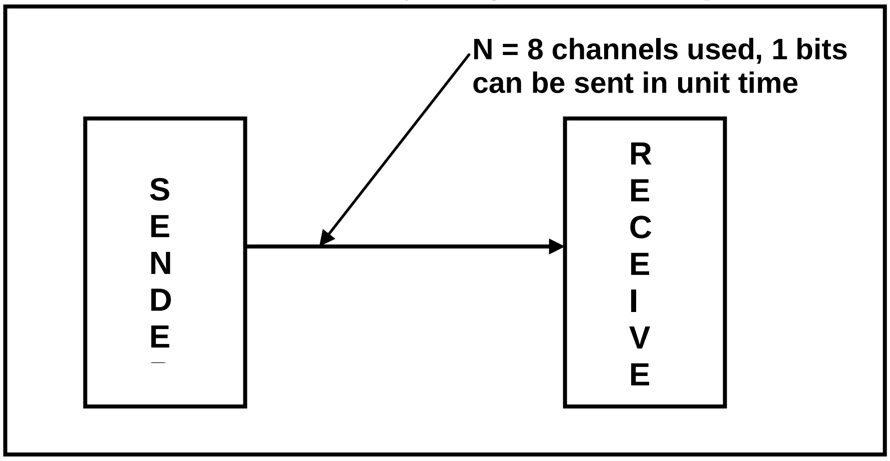


Fig.1.3 Serial Transmission of Data over N = 8 channels

Types of Serial Transmission:

Depending upon the timing of transmission of data, there are two types of serial transmission as described below:

Asynchronous Transmission

In asynchronous serial transmission, the sender and receiver are not synchronized.

The data is sent in group of 8 bits i.e. in bytes.

The sender can start data transmission at any time instant without informing the receiver.

To avoid confusing the receiver while receiving the data, “start” and “stop” bits are inserted before and after every group of 8 bits as shown below.

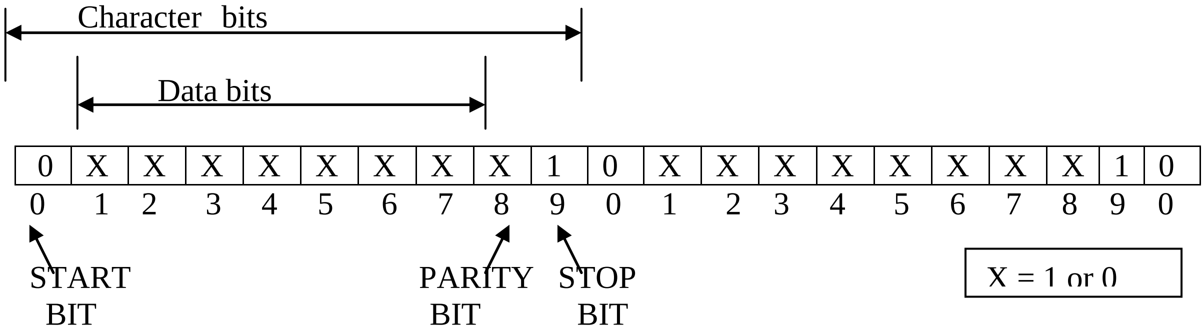


Fig.1.4 The structure of an asynchronous frame consists of four-key-bit components.

This frame has four components:

1) **A starts bit:** This component signals that a frame is starting and enables the receiving device to synchronize itself with the message.
2) **Data bits:** This component consists of a group of seven or eight bits when character data is being transmitted.
3) **A parity bit:** This component is optionally used as a crude method of detecting transmission errors. This bit can be set to either binary 1 or 0 to ensure either even for even-parity or odd for odd-parity.
4) **A stop bit or bits:** This component signals the end of the data frame.

Thus each character has a length of ten bits.

The start bit is indicated by “0” and stop bit is indicated by “1”.

The sender and receiver may not be synchronized as seen above but at the bit level they have to be synchronized i.e. the duration of one bit needs to be same for both sender and receiver for accurate data transmission. There may be gaps in between the data transmission indication that there is no data being transmitted from sender. Ex. Assume a user typing at uneven speeds, at times there is no data being transmitted from keyboard to the CPU.

Asynchronous transmission is a simple, inexpensive technology ideally suited for transmitting small frames at irregular intervals. Because start, stop and parity bits must be added to each character being transmitted, however, overhead for asynchronous transmission is high-often in the neighbourhood of nearly 20 to 30 percent. This high overhead wastes bandwidth and makes asynchronous transmission undesirable for transmitting large amounts of data.

Asynchronous transmission is frequently used for PC – to - PC and terminal-to-host communication. Data in these environments is often of the busty, character – oriented nature that is ideal for asynchronous communication. Asynchronous transmission generally requires less expensive hardwire than synchronous transmission. Following fig 1.5 is the Diagram for Asynchronous Serial Transmission.

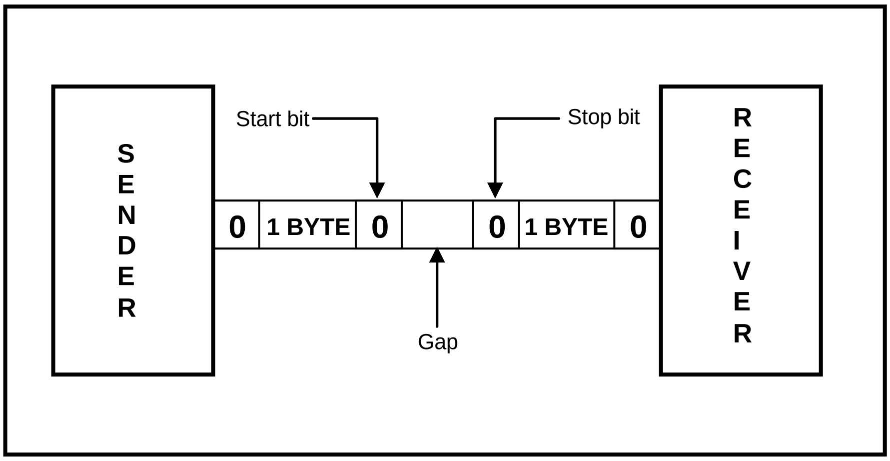


Fig:1.5 Asynchronous Serial Transmission

Advantages

1. Cheap and Effective implementation
2. Can be used for low speed communication

Disadvantages

Insertion of start bits, stop bits and gaps make asynchronous transmission slow.

Application

Keyboard

Synchronous Transmission

In Synchronous Serial Transmission, the sender and receiver are highly synchronized.
No start, stop bits are used. Instead a common master clock is used for reference.
Synchronous transmission refers to data transmission in which the time of occurrence of each signal representing a bit is related to a fixed time frame. Synchronous transmission eliminates the need for start and stop bits by synchronizing the clocks on the transmitting and receiving devices. This synchronization is accomplished in two ways:

1. By transmitting synchronization signals with data. Some data encoding techniques, by guaranteeing a signal transition with each bit transmitted, are inherently self-clocking.
2. By using a separate communication channel to carry clock signals. This technique can function with any signal encoding technique.

Figure 1.6 below, illustrates the two possible structures of messages associated with synchronous transmission.

a)

SYNCH	SYNCH	CHARACTER	….	CHARACTER	CRC	END

b)

SYNCH	SYNCH	BINARY DATA MESSAGE	CRC	END	FILL BITS	SYNCH	SYNCH	DATA	CRC	END

Fig.1.6 Structures of synchronous transmission

Both synchronous transmission methods begin with a series of SYNCH signals, which notify the receiver of the beginning of a frame.
SYNCH signals generally utilize a bit pattern that cannot appear elsewhere in messages, ensuring that the signals always are distinct and easily recognizable by the receiver.
A wide variety of data types can be transmitted. Fig. above, illustrates both character-oriented and bit-oriented data. Notice that under synchronous transmission, multiple characters or long series of bits

can be transmitter and receiver remain in synchronization for the duration of the transmission, frames maybe very long.
When frames are long, parity is no longer a suitable method for detecting errors. If errors occur, multiple bits are more likely to be affected, and parity 0 techniques are less likely to report an error. A more appropriate error-control technique for synchronous transmission is the ***Cyclic Redundancy Check*** (CRC).
In this technique, the transmitter uses an algorithm to calculate a CRC value that summarizes the entire value of the data bits. This value is then appended to the data frame,(just like a check sum value).
The receiver uses the same algorithm, recalculates the CRC, and compares the CRC, and compares the CRC in the frame almost definitely was transmitted without error.
When synchronous transmission links are idle, communicating devices generally send FILL BITS to the devices synchronized.

The sender simply send stream of data bits in group of 8 bits to the receiver without any start or stop bit. It is the responsibility of the receiver to regroup the bits into units of 8 bits once they are received. When no data is being transmitted, a sequence of 0's and 1's indicating IDLE is put on the transmission medium by the sender.

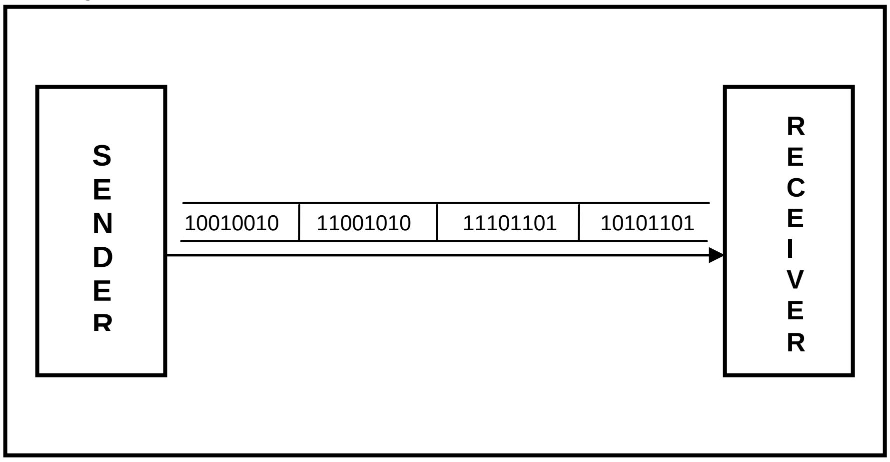


Fig:1.7 Asynchronous Serial Transmission

Advantage

1. There are no start bits, stop bits or gaps between data units. The overhead bits (SYNCH, CRC, and END) comprise a smaller portion (15%) of the overall data frame, which provides for more efficient use of available bandwidth.
2. Synchronization improves error detection and enables the devices to operate at higher speeds.
.3. Due to synchronization, there are no timing errors.

Disadvantages

Synchronous transmission requires more complex circuitry for communication, which is more expensive.
The choice between the two methods: asynchronous versus synchronous, must be based upon the required speed of response, and telephone circuit costs.

Table 1.1Comparison of serial and parallel transmission

S/N	Parameter	Parallel transmission	Serial transmission
1	Number of wire required to transmit N bits	N wire	1 wire
2	Number of bits transmitted simultaneously	N bits	1 bit

3	Speed of data transfer	False	Slow
4	Application	Higher due to more number of conductor	Low, since only one wire is used
5		Short distance communication such as computer to printer communication	Long distance computer to computer communication

Transmission Impairments & Types

Data is transmitted through transmission medium which are not perfect.
The imperfection causes signal impairment.
Due to the imperfection error is introduced in the transmitted data i.e. the original signal at the beginning of the transmission is not the same as the signal at the Receiver.
There are three causes of impairment attenuation, distortion, and noise as shown below:

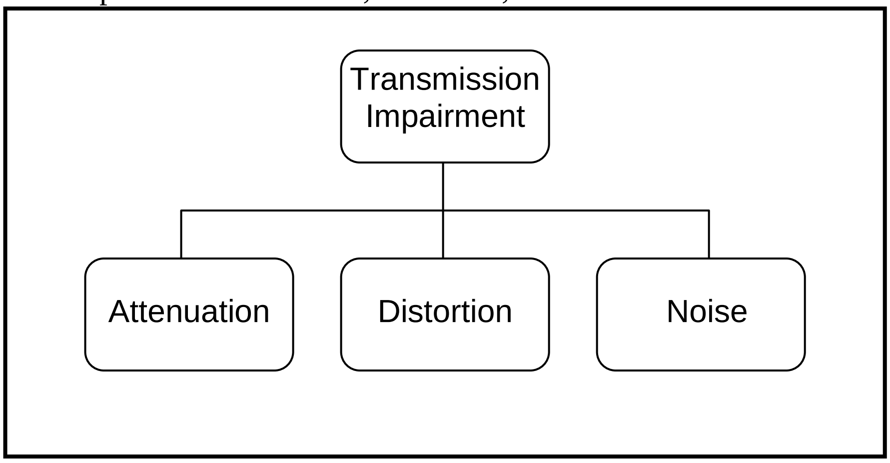


Fig:1.8 Transmission Impairment Types

Attenuation

- Attenuation results in loss of energy. When a signal travels through a medium, it loses some of its energy in overcoming the resistance of the medium.
- The electrical energy in the signal may be converted to heat.
- To compensate for this loss, amplifiers are used to amplify the signal. Figure below shows the effect of attenuation and amplification.

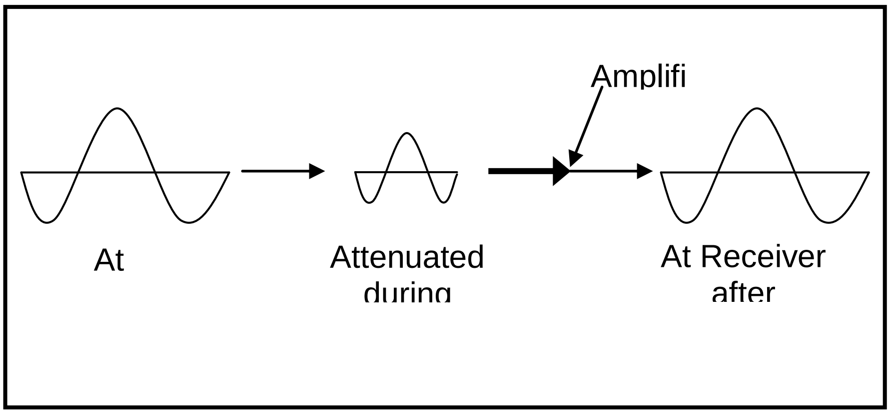


Fig.1.9 Attenuation

Distortion

Distortion changes the shape of the signal as shown below:

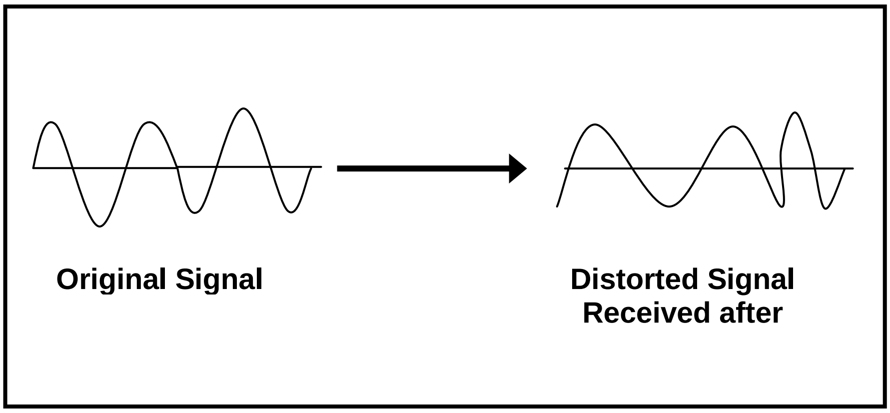


Fig.1.10 Distortion

Noise

Noise is any unwanted signal that is mixed or combined with the original signal during transmission. Due to noise the original signal is altered and signal received is not same as the one sent.

ERRORS, DETECTION & CORRECTION

INTRODUCTION

Errors in the data are basically caused due to the various impairments that occur during the process of transmission. When there is an imperfect medium or environment exists in the transmission it prone to errors in the original data.

Errors can be classified as follows: Attenuation, Noise and Distortion

TYPES OF ERRORS

If the signal comprises of binary data, there can be two types of errors which are possible during the transmission.

1. Single bit errors
2. Burst Errors

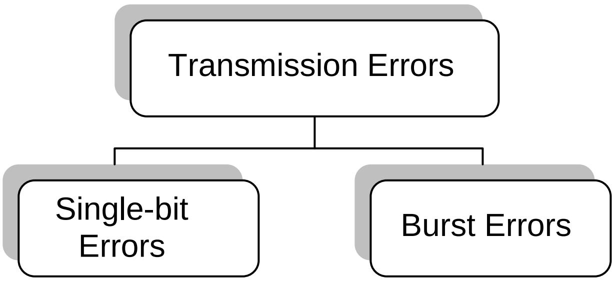


Figure2.1 Transmission errors

1. Single-bit errors:

In single-bit error, a bit value of 0 changes to bit value 1 or vice versa. Single bit errors are more likely to occur in parallel transmission. Figure below(a)

2. Burst errors:

In Burst error, multiple bits of the binary value changes. Burst error can change any two or more bits in a transmission. These bits need not be adjacent bits. Burst errors are more likely to occur in serial transmission. Figure below (b)

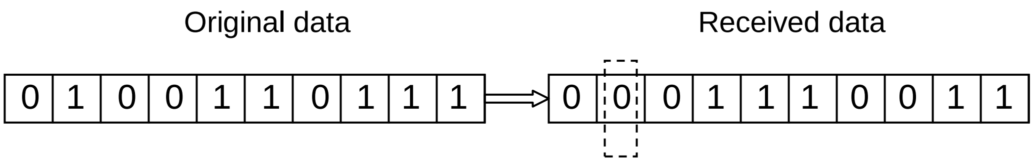
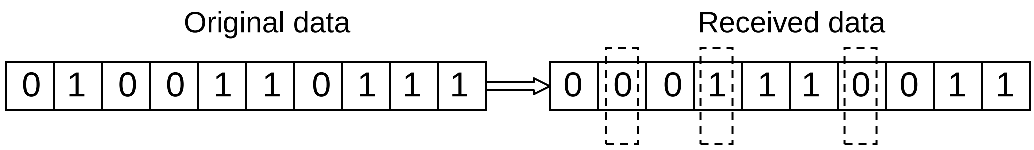


Figure (a) Single bit error



Figure(b) Burst error

REDUNDANCY

In order to detect and correct the errors in the data communication we add some extra bits to the original data. These extra bits are nothing but the redundant bits which will be removed by the receiver after receiving the data.

Their presence allows the receiver to detect or correct corrupted bits. Instead of repeating the entire data stream, a short group of bits may be attached to the entire data stream. This technique is called redundancy because the extra bits are redundant to the information, they are discarded as soon as the accuracy of the transmission has been determined.

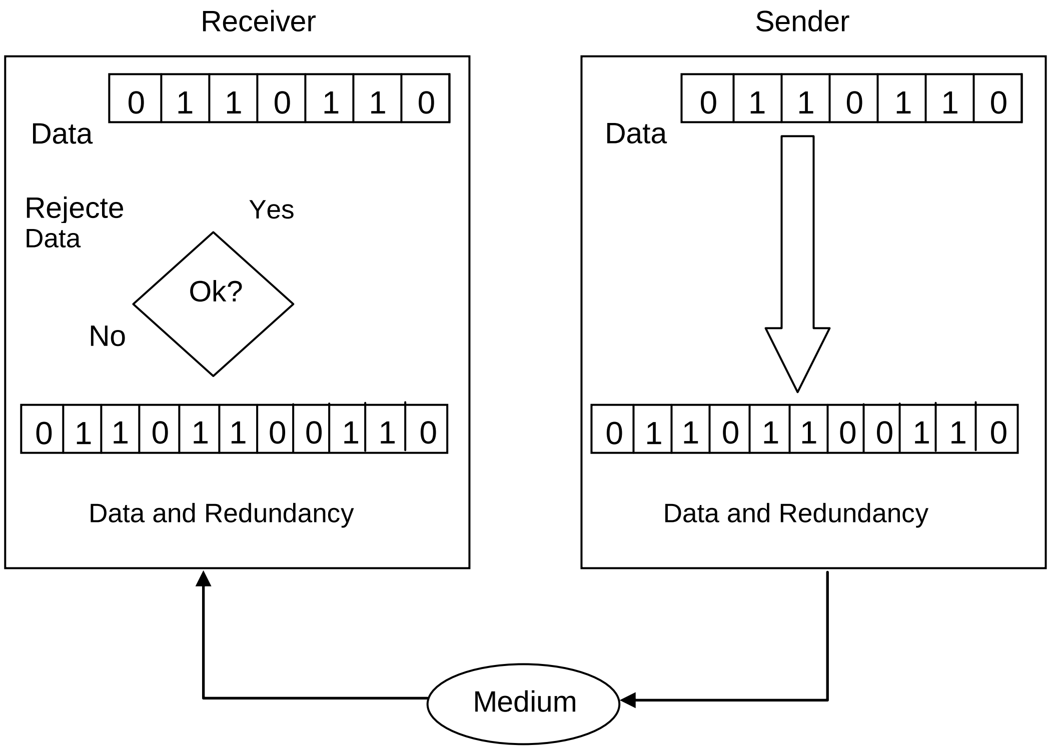


Figure 2.3

There are different techniques used for transmission error detection and correction.

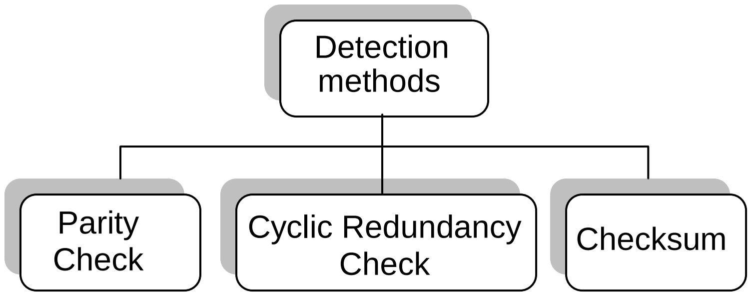
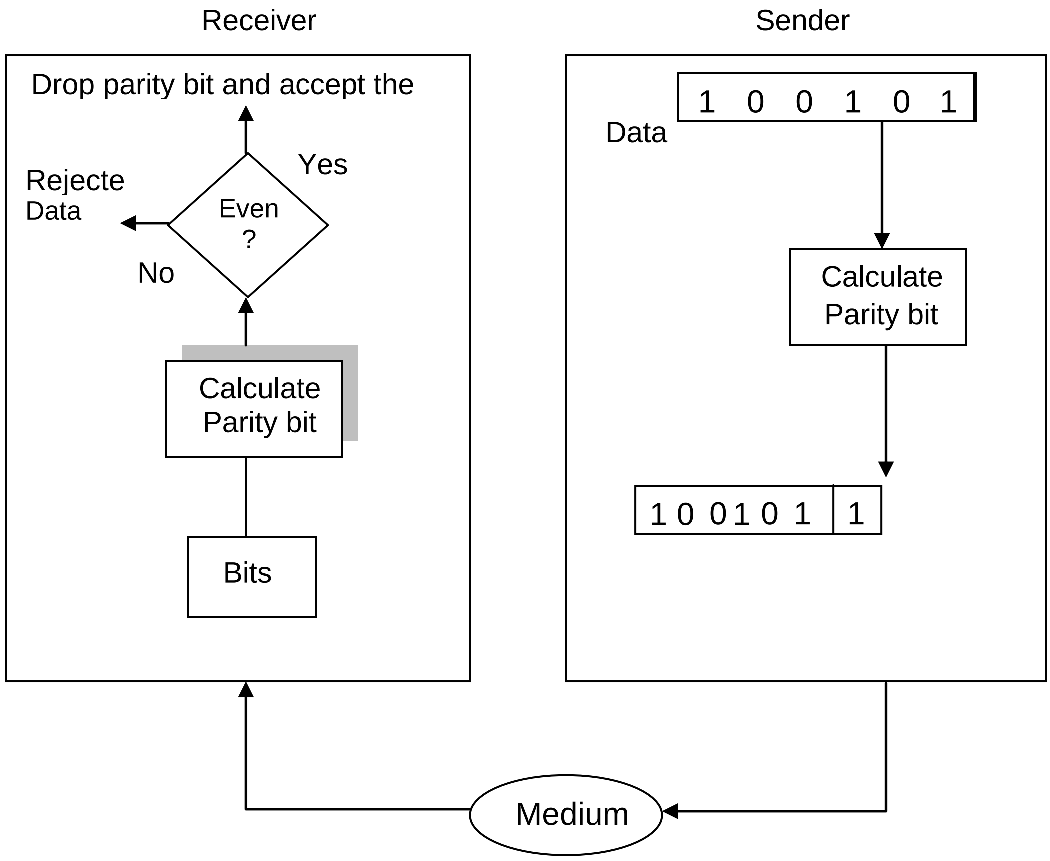


Figure 2.4

2. Parity Check

In this technique, a redundant bit called a parity bit is added to every data unit so that the total number of 1's in the unit (including the parity bit) becomes even (or odd).

Figure below shows this concept when transmitting the binary data: 100101.

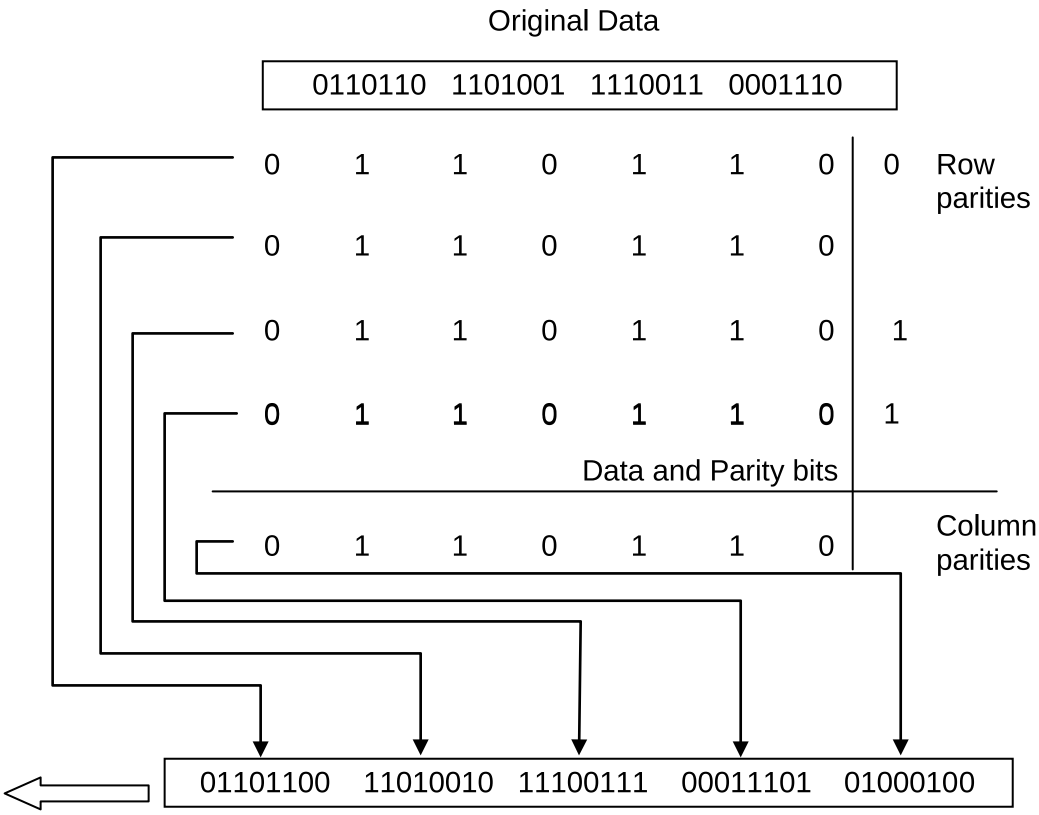


Simple parity check can detect all single-bit errors. It can also detect burst errors as long as the total number of bits changed is odd. This method cannot detect errors where the total number of bits changed is even.

Two-Dimensional Parity Check

A better approach is the two dimensional parity checks. In this method, a block of bits is organized in a table (rows and columns). First we calculate the parity bit for each data unit. Then we organize them into a table. We then calculate the parity bit for each column and create a new row of 8 bits.

Consider the following example, we have four data units to send. They are organized in the tabular form as shown below:



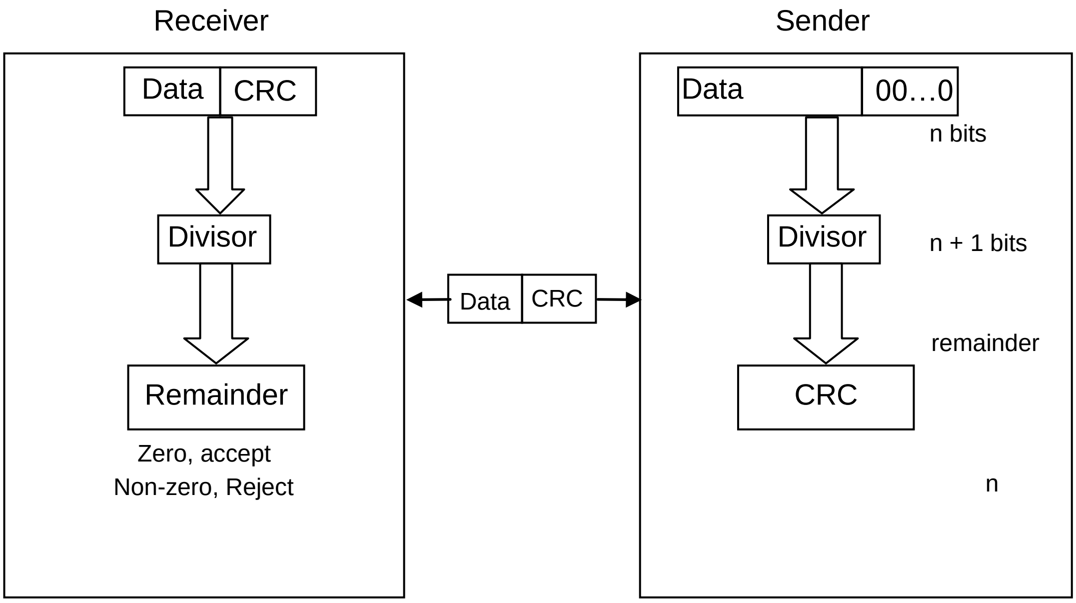
We then calculate the parity bit for each column and create a new row of 8 bits; they are the parity bits for the whole block. Note that the first parity bit in the fifth row is calculated based on all first bits; the second parity bit is calculated based on all second bits; and so on. We then attach the 8 parity bits to the original data and send them to the receiver.

Two-dimensional parity check increases the likelihood of detecting burst errors. A burst error of more than `n' bits is also detected by this method with a very high probability.

3. Cyclic Redundancy Check (CRC)

Most powerful of the redundancy checking techniques is the cyclic redundancy check (CRC). This method is based on the binary division. In CRC, the desired sequence of redundant bits are generated and is appended to the end of data unit. It is also called as CRC reminder. So that the resulting data unit becomes exactly divisible by a predetermined binary number. At its destination, the incoming data unit is divided by the same number. If at this step there is no remainder then the data unit is assumed to be correct and is therefore accepted. A remainder indicates that the data unit has been damaged in transit and therefore must be rejected. The redundancy bits used by CRC are derived by dividing the data unit by a predetermined divisor; the remainder is the CRC. To be valid, a CRC must have two qualities: it must have exactly one less bit than the divisor, and appending it to the end of the data string must make the resulting bit sequence exactly divisible by the divisor.

The following figure shows the process:



Step 1: A string of 0's is appended to the data unit. It is n bits long. The number n is 1 less if-number of bits in the predetermined divisor which is n + 1 bits.
Step 2: The newly generated data unit is divided by the divisor, using a process known as binary division. The remainder resulting from this division is the CRC.
Step 3: The CRC of n bits derived in step 2 replaces the appended 0's at the data unit. Note that the CRC may consist of all 0's.
The data unit arrives at the receiver data first, following by the CRC. The receiver treats the whole string as a unit and divides it by the same divisor that was used the CRC remainder. If the string arrives without error, the CRC checker yields a remainder of zero, the data unit passes. If the string has been changed in transit, the division yields zero remainder and the data unit does not pass.

Cyclic Redundancy Check (CRC) Technique

CRC refers to an error detection technique in which the code is the remainder resulting from dividing the bits to be checked by a predetermined binary number. In transmission the data block or message is thought of as a **stream** of serial data bits. The bits in this n-bit block are considered the coefficients of a characteristic polynomial M (x).
$M(x) = b_n X^0 + b_{n-1} X + b_{n-2} X^2 + \ldots + b_1 X^{n-1} + b_0 X^n$. Where, b_0 is the LSB and b_n is the MSB.

Example 1: Calculate the data polynomial M (x) for the 16-bit data stream 26F0H.
Solution: First representing 26F0H in binary becomes $0010\ 0110\ 1111\ 0000_2$

Now write this as M(x):
$M(x) = 0 + 0X^1 + 1X^2 + 0X^3 + 0X^4 + 1X^5 + 1X^6 + 0X^7 + 1X^8 + 1X^9 + 1X^{10} + 1X^{11} + 0X^{12} + 0X^{13} + 0X^{14} + 0X^{15}$ and eliminating the zero terms given:
$M(x) = X^2 + X^5 + X^6 + X^8 + X^9 + X^{10} + X^{11}$………………………(1)

Equation 1 is a unique polynomial representing the data in the 16-bit block. If one or more of the data bits were to be changed the polynomial would also change.
The CRC is therefore given by the formulae; $CRC = M(x) * X^n / G(x) = Q(x) + R(x)$.
Or $CRC = M(x) * 2^n / G(x)$……………………………………(2)
Where, M(x)- is a **k-bit** number;
R(x) – an n-bit number such that **k** is greater than **n; G(x) –** is an (n+1) -bit number**.**

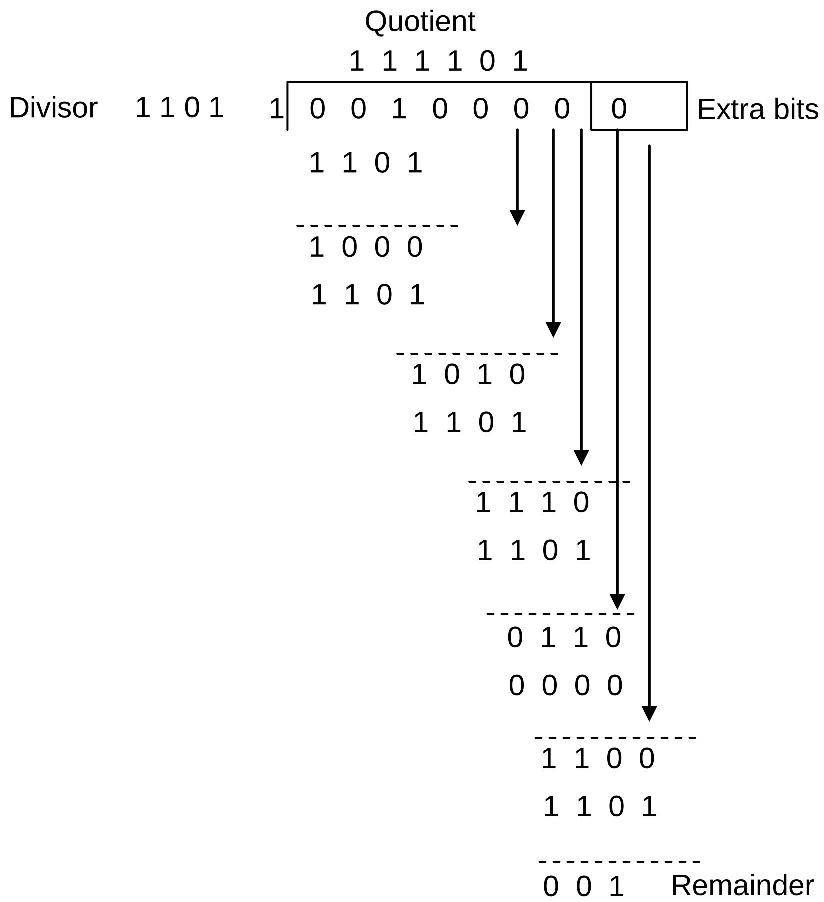
In equation 2, G(x) is called the generator polynomial. For the Bisync protocols $G(x) = X^{16} + X^{15} + X^{2} + 1$. Similarly, for SDLC (Synchronous Data link Control) protocol: $G(x) = X^{16} + X^{13} + X^{5} + 1$. When the division is performed, the result will be a quotient Q(x) and a remainder R(x). The CRC technique consists of calculating R(x) for the data stream and appending the result to the data block. With modulo-2 division of one binary number by another the rules for performing the division is as follows:

(a) If the divisor has the same number of bits as the dividend, the quotient is 1; if the divisor has fewer bits than the dividend the quotient is 0.

(b) There are no carries and 1 – 1 =0, 0 - 0 =0, 1 – 0 = 1 and 0 – 1 = 1.

Since the division is binary any remainder will always be one bit shorter than the divisor.
Following figure shows the process of generating CRC reminder:

Figure:



A CRC checker functions does exactly as the generator does. After receiving the data appended with the CRC, it does the same modulo-2 division. If the remainder is all 0's, the CRC is dropped and the data is accepted, otherwise, the received stream of bits is discarded and data is resent.
Following Figure shows the same process of division in the receiver.

Figure:

```
                    Quotient
                  1 1 1 1 0 1
Divisor  1 1 0 1 | 1 0 0 1 0 0 0 | 0  1 | CRC
                 1 1 0 1
                 -----------
                   1 0 0 0
                   1 1 0 1
                   -----------
                     1 0 1 0
                     1 1 0 1
                     -----------
                       1 1 1 0
                       1 1 0 1
                       -----------
                         0 1 1 0
                         0 0 0 0
                         -----------
                           1 1 0 1
                           1 1 0 1
                           -----------
                           0 0 0 0 Result
```

Solved Example: Given a divisor as $X^4 + X^3 + 1$ or G(x) =11001 and a data stream of M(x) =1010110101. Find the CRC?

Solution:

Performance:

CRC is a very effective error detection method. If the divisor is chosen according to the previously mentioned rules:

1. CRC can detect all burst errors that affect an odd number of bits.
2. CRC can detect all burst errors of length less than or equal to the degree of the polynomial.
3. CRC can detect, with a very probability, burst error of length greater than the degree of the polynomial.

3. Checksum

A checksum is fixed length data that is the result of performing certain operations on the data to be sent sender to the receiver. The sender runs the appropriate checksum algorithm to compute the checksum of the data, appends it as a field in the packet that contains the data to be sent, as well as various headers.

Example: Calculate the checksum byte for the following four hex data bytes: 10, 23, 45, and 04.

Solution;

The sum is calculated first as 10 + 23+45+ 04 =7C, which is 0111 1100, then invert and add 1 to the LSB (that is , forming 2"s complement code of 7C) giving 84H.

When the receiver receives the data, the receiver runs the same checksum algorithm to compute a fresh checksum. The receiver compares this freshly computed checksum with the checksum that was computed by the sender. If the two checksum matches, the receiver of the data is assured that the data has not changed during the transit.