COMPUTER NETWORKS
Module – 1 & 2

Course Outline

Over the years the subject of computer networks has grown with advancement of technology and the emergence of new technologies and new applications. In this course, this massive subject has been divided into comprehensible parts and arranged in a structured and logical manner. It is organized in the following eight modules:

- Introduction
- Data Communication Fundamentals
- Packet Transmission
- Routing Algorithms
- Internetworking
- Routing and Congestion Control
- Network Services
- Network Security

Computer networks help users on the network to share the resources and in communication. Can you imagine a world now without emails, online newspapers, blogs, chat and the other services offered by the internet?

Textbooks:

1. Data Communications and Networking, Behrouz A. Forouzan, TMH
2. Data and Computer Communications, William Stallings, PHI
3. Computer Networks, Andrew S. Tanenbaum, PHI

NETWORKS

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.

Distributed Processing

Most networks use distributed processing, in which a task is divided among multiple computers. Instead of one single large machine being responsible for all aspects of a process, separate computers (usually a personal computer or workstation) handle a subset.

Network Criteria

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. Performance is often evaluated by two networking metrics: throughput and delay. We often need more throughput and less delay. However, these two criteria are often contradictory. If we try to send more data to the network, we may increase throughput but we increase the delay because of traffic congestion in the network.

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.

Classification Based on Transmission Technology

Computer networks can be broadly categorized into two types based on transmission technologies:

Broadcast networks

Point-to-point networks

Broadcast Networks

Broadcast network have a single communication channel that is shared by all the machines on the network as shown in Fig.1.1. All the machines on the network receive short messages, called packets in certain contexts, sent by any machine. An address field within the packet specifies the intended recipient. Upon receiving a packet, machine checks the address field. If packet is intended for itself, it processes the packet; if packet is not intended for itself it is simply ignored. This system generally also allows possibility of addressing the packet to all destinations (all nodes on the network). When such a packet is transmitted and received by all the machines on the network. This mode of operation is known as Broadcast Mode. Some Broadcast systems also support transmission to a sub-set of machines, something known as Multicasting.

Point-to-Point Networks

A network based on point-to-point communication is shown in Fig. 1.2. The end devices that wish to communicate are called stations. The switching devices are called nodes. Some Nodes connect to other nodes and some to attached stations. It uses FDM or TDM for node-to-node communication. There may exist multiple paths between a source-destination pair for better network reliability. The switching nodes are not concerned with the contents of data. Their purpose is to provide a switching facility that will move data from node to node until they reach the destination. As a general rule (although there are many exceptions), smaller, geographically localized networks tend to use broadcasting, whereas larger networks normally use are point-to-point communication.

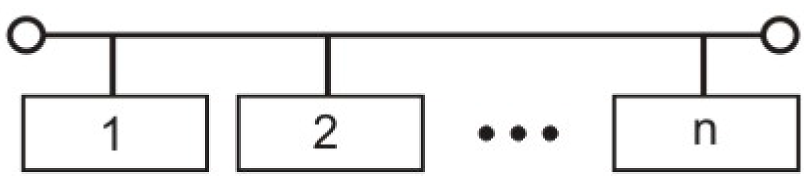


Fig.1-*Example of a broadcast network based on shared bus*

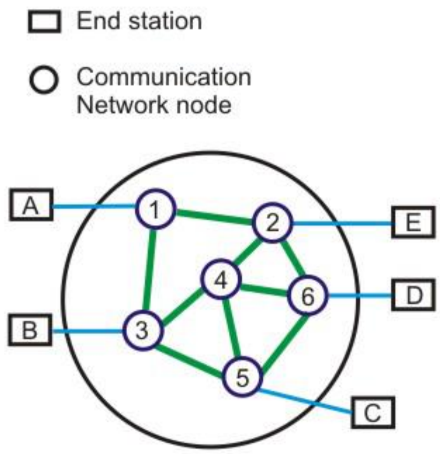


Fig. 2-*Communication network based on point-to-point communication*

Applications

In a short period of time computer networks have become an indispensable part of business, industry, entertainment as well as a common-man's life. These applications have changed tremendously from time and the motivation for building these networks are all essentially economic and technological.

Initially, computer network was developed for defense purpose, to have a secure communication network that can even withstand a nuclear attack. After a decade or so, companies, in various fields, started using computer networks for keeping track of inventories, monitor productivity, communication between their different branch offices located at different locations. For example, Railways started using computer networks by connecting their nationwide reservation counters to provide the facility of reservation and enquiry from any where across the country.

And now after almost two decades, computer networks have entered a new dimension; they are now an integral part of the society and people. In 1990s, computer network started delivering services to private individuals at home. These services and motivation for using them are quite different. Some of the services are access to remote information, person-person communication, and interactive entertainment. So, some of the applications of computer networks that we can see around us today are as follows:

Marketing and sales: Computer networks are used extensively in both marketing and sales organizations. Marketing professionals use them to collect, exchange, and analyze data related to customer needs and product development cycles. Sales application includes teleshopping, which uses order-entry computers or telephones connected to order processing network, and online-reservation services for hotels, airlines and so on.

Financial services: Today's financial services are totally depended on computer networks. Application includes credit history searches, foreign exchange and investment services, and electronic fund transfer, which allow user to transfer money without going into a bank (an automated teller machine is an example of electronic fund transfer, automatic pay-check is another).

Manufacturing: Computer networks are used in many aspects of manufacturing including manufacturing process itself. Two of them that use network to provide essential services are computer-aided design (CAD) and computer-assisted manufacturing (CAM), both of which allow multiple users to work on a project simultaneously.

Directory services: Directory services allow list of files to be stored in central location to speed worldwide search operations.

Information services: A Network information service includes bulletin boards and data banks. A World Wide Web site offering technical specification for a new product is an information service.

Electronic data interchange (EDI): EDI allows business information, including documents such as purchase orders and invoices, to be transferred without using paper.

Electronic mail: probably it's the most widely used computer network application.

Teleconferencing: Teleconferencing allows conference to occur without the participants being in the same place. Applications include simple text conferencing (where participants communicate through their normal keyboards and monitor) and video conferencing where participants can even see as well as talk to other fellow participants. Different types of equipments are used for video conferencing depending on what quality of the motion you want to capture (whether you want just to see the face of other fellow participants or do you want to see the exact facial expression).

Voice over IP: Computer networks are also used to provide voice communication. This kind of voice communication is pretty cheap as compared to the normal telephonic conversation.

Video on demand: Future services provided by the cable television networks may include video on request where a person can request for a particular movie or any clip at anytime he wish to see.

Motivation :

- Sharing of resources is more efficient
- Price/Performance
- Use each piece of equipment for what it is best at
- Centralize administration
- Computers as communication tools

Specific Instructional Objectives

At the end of this subject the students will be able to:

- Explain what is Data Communication
- Define Computer Networks
- State the evolution of Computer Networks
- Categorize different types of Computer Networks
- Specify some of the application of Computer Networks

The concept of Network is not new. In simple terms it means an interconnected set of some objects. For decades we are familiar with the Radio, Television, railway, Highway, Bank and other types of networks. In recent years, the network that is making significant impact in our day-to-day life is the Computer network.

Computer network we mean an interconnected set of autonomous computers. The term autonomous implies that the computers can function independent of others. However, these computers can exchange information with each other through the communication network system. Computer networks have emerged as a result of the convergence of two technologies of this century- Computer and Communication as shown in Fig. 3.

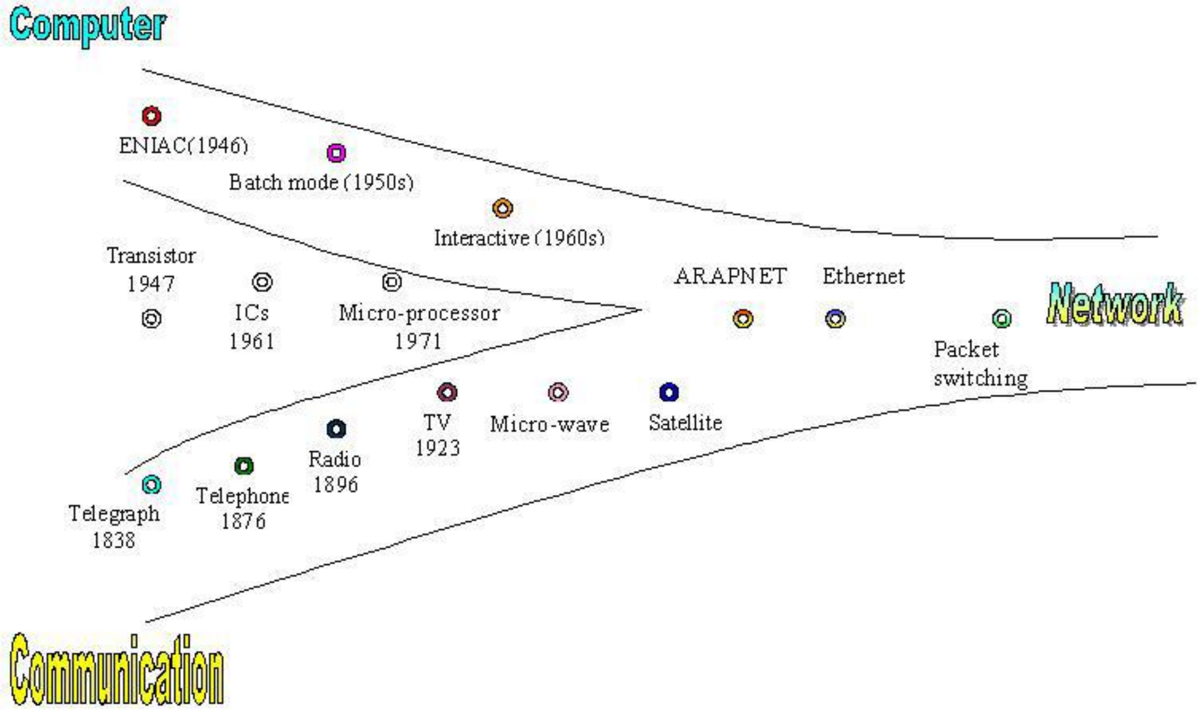


Figure 3 *Evolution of computer networks*

Historical Background

The history of electronic computers is not very old. It came into existence in the early 1950s and during the first two decades of its existence it remained as a centralized system housed in a single large room. In those days the computers were large in size and were operated by trained personnel. To the users it was a remote and mysterious object having no direct communication with the users. Jobs were submitted in the form of punched cards or paper tape and outputs were collected in the form of computer printouts. The submitted jobs were executed by the computer one after the other, which is referred to as batch mode of data processing. In this scenario, there was long delay between the submission of jobs and receipt of the results.

In the 1960s, computer systems were still centralize, but users provided with direct access through interactive terminals connected by point-to-point low-speed data links with the computer. In this situation, a large number of users, some of them located in remote locations could simultaneously access the centralized computer in time-division multiplexed mode. The users could now get immediate interactive feedback from the computer and correct errors immediately. Following the introduction of on-line terminals and time-sharing operating systems, remote terminals were used to use the central computer.

With the advancement of VLSI technology, and particularly, after the invention of microprocessors in the early 1970s, the computers became smaller in size and less expensive, but with significant increase in processing power. New breed of low-cost computers known as mini and personal computers were introduced. Instead of having a single central computer, an organization could now afford to own a number of computers located in different departments and sections.

Side-by-side, riding on the same VLSI technology the communication technology also advanced leading to the worldwide deployment of telephone network, developed primarily for voice communication. An organization having computers located geographically dispersed locations wanted to have data communications for diverse applications. Communication was required among the machines of the same kind for collaboration, for the use of common software or data or for sharing of some costly resources. This led to the development of computer networks by successful integration and cross-fertilization of communications and geographically dispersed computing facilities. One significant development was the APPANET (Advanced Research Projects Agency Network). Starting with four-node experimental network in 1969, it has subsequently grown into a network several thousand computers spanning half of the globe, from Hawaii to Sweden. Most of the present-day concepts such as packet switching evolved from the ARPANET project. The low bandwidth (3KHz on a voice grade line) telephone network was the only generally available communication system

The bandwidth was clearly a problem, and in the late 1970s and early 80s another new communication technique known as Local Area Networks (LANs) evolved, which helped computers to communicate at high speed over a small geographical area. In the later years use of optical fiber and satellite communication allowed high-speed data communications over long distances.

1.1 INTRODUCTION TO NETWORKS & DATA COMMUNICATIONS

The term telecommunication means communication at a distance. The word data refers to information presented in whatever form is agreed upon by the parties creating and using the data. Data communications are the exchange of data between two devices via some form of transmission medium such as a wire cable. For Data Communication to occur, the communicating devices must be a part of a communication system made up of a combination of hardware and software.

The effectiveness of a data communication system depends on four fundamental characteristics:

1. Delivery
2. Accuracy
3. Timeliness
4. Jitter

There are five components of data communication as shown in Fig. 1.1 below:

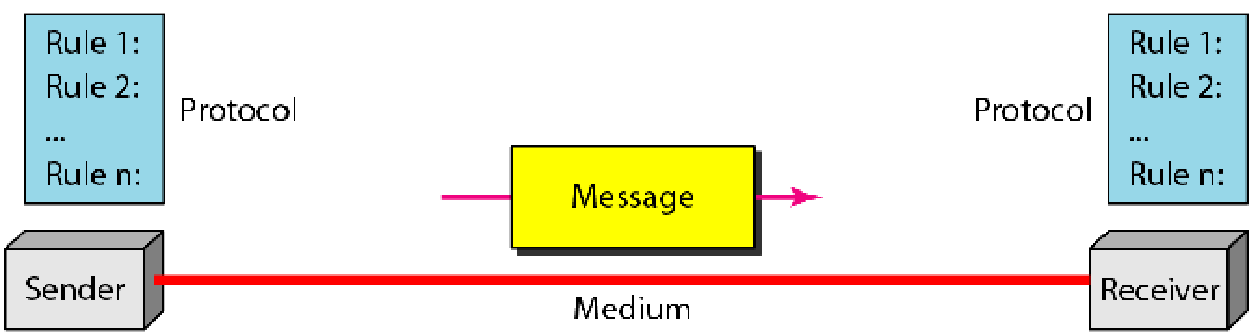


Fig 1.1 Components of Data Communication

(a) **Sender:** is the device that sends the data message.
(b) **Message**: is the information (data) to be communicated. Eg: text, numbers etc.
(c) **Transmission Medium**: is the physical path by which a message travels from sender to receiver. Eg: twisted pair cable, fiber-optic cable etc.
(d) **Receiver**: is the device that receives the message.
(e) **Protocols**: is a set of rules that govern the data communication. It represents an agreement between the communicating devices.

Moreover, Data can flow in three different ways namely Simplex, Half- Duplex and Full Duplex. In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive. In half-duplex mode, each station can both transmit and receive, but not at the same time. i.e. When one device is sending, the other can only receive, and vice versa. Whereas, in full-duplex mode (also called duplex), both stations can transmit and receive simultaneously.
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.
A network must be able to meet these three criteria's:
1. Performance: can be measured using Transit time and Response time:
(a) Transit Time: is the time required for a message to travel from one device to another.
(b) Response Time: is the elapsed time between an inquiry and a response.
2. Reliability: is measured by the frequency of failure i.e the time it takes a link to recover from a failure.
3. Security: issues include protecting data from unauthorized access and losses.
Furthermore, there are two types of connection: Point to Point and Multipoint. In Point -to-Point: Connection provides a dedicated link between two devices. Whereas, in Multi-Point: Connection is one in which more than two devices share a single link.
Network Categories: The category into which a network falls is determined by its size. Network can be categorized as: LAN, WAN, MAN, Wireless Network and Internetwork.

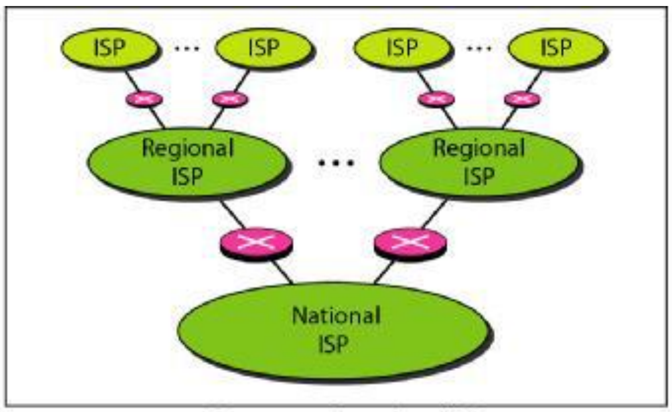
1.2 THE INTERNET

The Internet is a global, interconnected computer network in which every computer connected to it can exchange data with any other connected computer. Rather than moving through geographical space, it moves your ideas and information through cyberspace – the space of electronic movement of ideas and information.
Significance of an Internet are as follows:

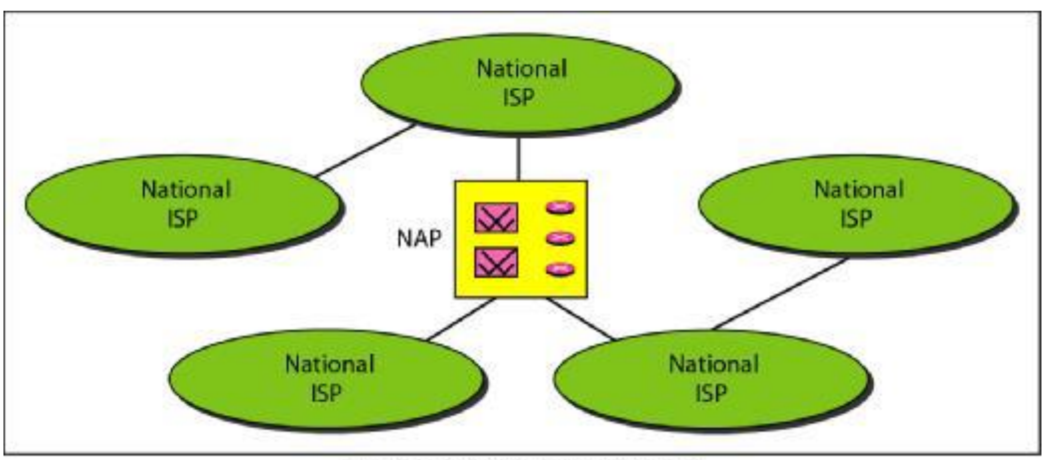
- It's the first mass medium that involves computers and uses digitized data.
- It provides the potential for media convergence, the unification of all media.
- It's transforming how we communicate, obtain information, learn, seek jobs, and maintain professional growth.
- Businesses find it an indispensable tool for their needs.

In the late 1950's the Advanced Research Projects Agency (ARPA) was founded in the United States with the primary focus of developing information technologies that could survive a nuclear attack. Scientists and military experts were especially concerned about what might happen in the event of a Soviet attack on the nation's telephone system. Just one missile, they feared, could destroy the whole network of lines and wires that made efficient long-distance communication possible. In 1962, a scientist from M.I.T. and ARPA named J.C.R. Licklider proposed a solution to this problem: a "galactic network" of computers that could talk to one another. Such a network would enable government leaders to communicate even if the Soviets destroyed the telephone system. In 1965, another M.I.T. scientist developed a way of sending information from one computer to another that he called "packet switching." Packet switching breaks data down into blocks, or packets, before sending it to its destination. That way, each packet can take its own route from place to place. In 1967 ARPA university and private sector contractors met with representatives of the Department of Defense to discuss possible protocols for sharing information via computers. Thus in 1969, ARPAnet delivered its first message: a "node-to-node" communication from one computer to another. It was a connection of computers at UCLA, Stanford, UCSB, Univ. of Utah.

The Internet has come a long way since the 1960s. The Internet today is not a simple hierarchical structure. It is made up of many wide- and local-area networks joined by connecting devices and switching stations. It is difficult to give an accurate representation of the Internet because it is continually changing-new networks are being added, existing networks are adding addresses, and networks of defunct companies are being removed. Today most end users who want Internet connection use the services of Internet service providers (ISPs). There are: International service providers, National service providers, Regional service providers and Local service providers.
The Internet today is run by private companies, not by the government.
Hierarchical organization of the Internet is shown in Fig. 1.2 below as (a) structure of a national ISP and (b) Interconnection of national ISPs.



a. Structure of a national ISP



b. Interconnection of national ISPs

Fig.-1.2 (a) structure of a national ISP and (b) Interconnection of national ISPs.

1.3 LAYERED TASKS : OSI MODEL

Why Layered architecture?

1. To make the design process easy by breaking unmanageable tasks into several smaller and manageable tasks (by divide-and-conquer approach).

2. Modularity and clear interfaces, so as to provide comparability between the different providers' components.

3. Ensure independence of layers, so that implementation of each layer can be changed or modified without affecting other layers.

4. Each layer can be analyzed and tested independently of all other layers.

We use the concept of layers in our daily life. As an example, let us consider two friends who communicate through postal mail. The process of sending a letter to a friend would be complex if there were no services available from the post office. Figure 1.3 below shows tasks involved in sending a letter:

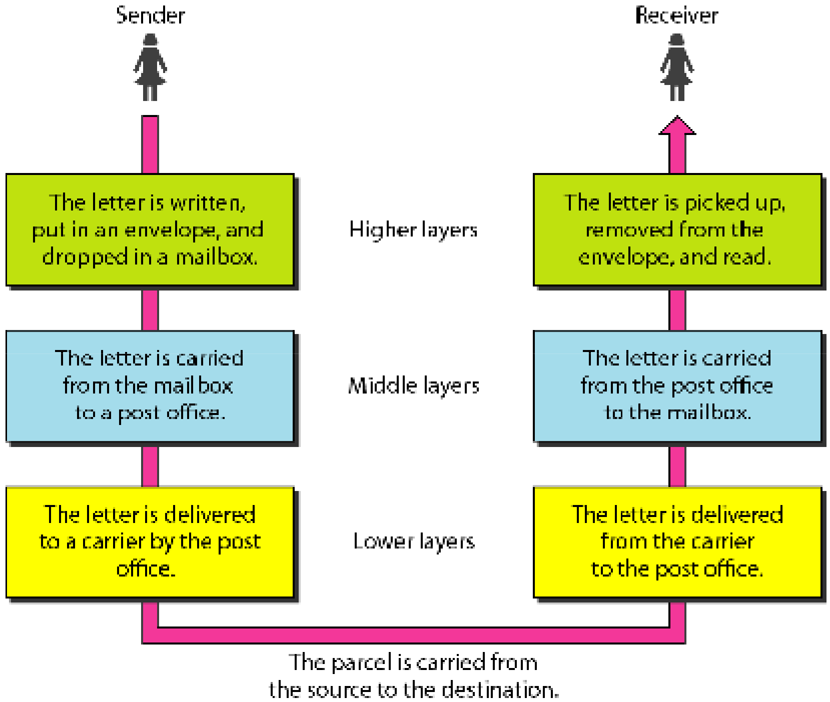


Fig 1.3 Layered Tasks

Thus from above figure it is clearly understood that layer architecture simplifies the network design. It is easy to debug network applications in a layered architecture network. There are two layered Models namely OSI Model and TCP/IP Model.

1.4 OSI MODEL: OPEN SYSTEM FOR INTERCONNECTION

International Standard Organization (ISO) established a committee in 1977 to develop architecture for computer communication. Open Systems Interconnection (OSI) reference model is the result of this effort. In 1984, the Open Systems Interconnection (OSI) reference model was approved as an international standard for communications architecture. Term “open” denotes the ability to connect any two systems which conform to the reference model and associated standards. The purpose of OSI Model is to facilitate communication between different systems without requiring changes to the logic of the underlying hardware and software. The OSI model is now considered the primary Architectural model for inter-computer communications. The OSI model describes how information or data makes its way from application programmes (such as spreadsheets) through a network medium (such as wire) to another application programme located on another network. The OSI reference model divides the problem of moving information between computers over a network medium into SEVEN smaller and more manageable problems. This separation into smaller more manageable functions is known as layering. Figure below shows interaction between layers in the OSI model:

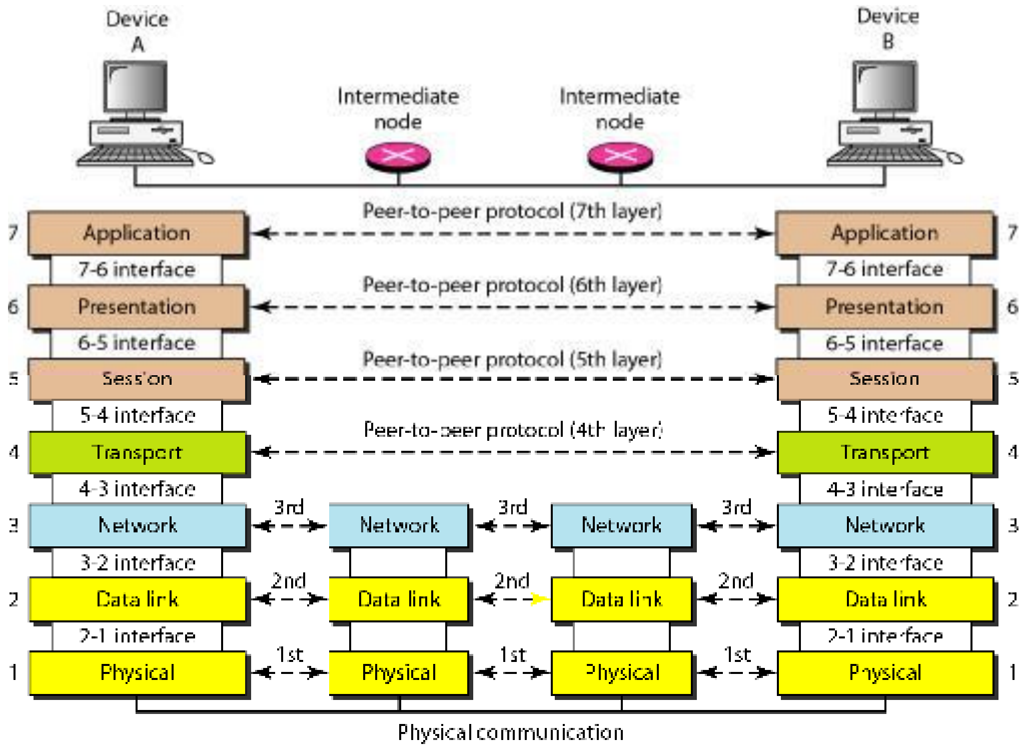


Fig 1.4 OSI Model

The process of breaking up the functions or tasks of networking into layers reduces complexity. Each layer provides a service to the layer above it in the protocol specification. Each layer communicates with the same layer's software or hardware on other computers. The lower 4 layers (transport, network, data link and physical —Layers 4, 3, 2, and 1) are concerned with the flow of data from end to end through the network. The upper four layers of the OSI model (application, presentation and session—Layers 7, 6 and 5) are orientated more toward services to the applications. Data is encapsulated with the necessary protocol information as it moves down the layers before network transit. A message begins at the top application layer and moves down the OSI layers to the bottom physical layer. As the message descends, each successive OSI model layer adds a header to it. A header is layer-specific information that basically explains what functions the layer carried out. Conversely, at the receiving end, headers are striped from the message as it travels up the corresponding layers as shown in Fig.1.5.

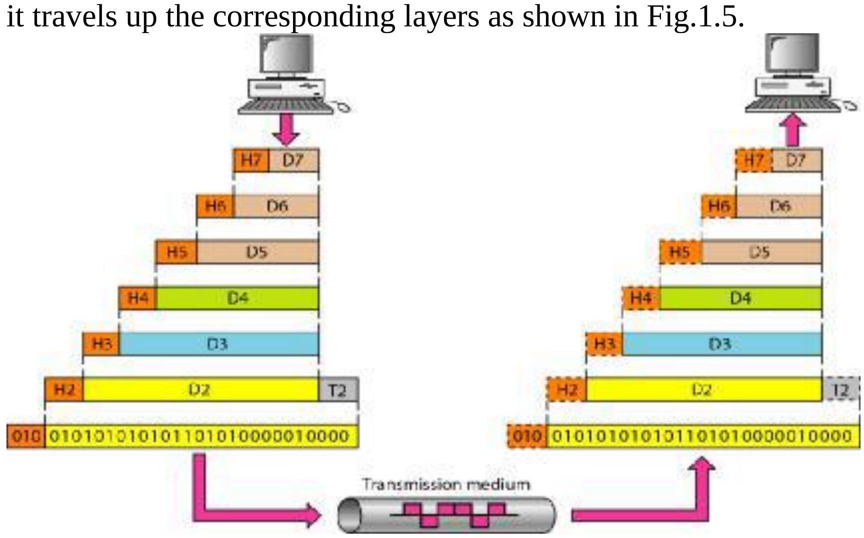


Fig 1.5 Working Principle

A) PHYSICAL LAYER

- Provides physical interface for transmission of information.
- Defines rules by which bits are passed from one system to another on a physical communication medium.
- Covers all - mechanical, electrical, functional and procedural - aspects for physical communication.
- Such characteristics as voltage levels, timing of voltage changes, physical data rates, maximum transmission distances, physical connectors, and other similar attributes are defined by physical layer specifications.
- Concerned with line configuration, physical topology and transmission mode.

B) DATA LINK LAYER

- Data link layer attempts to provide reliable communication over the physical layer interface.
- Breaks the outgoing data into frames and reassemble the received frames.
- Create and detect frame boundaries.
- Handle errors by implementing an acknowledgement and retransmission scheme.
- Implement flow control.
- Responsible for Error Control.
- Supports points-to-point as well as broadcast communication.
- Supports simplex, half-duplex or full-duplex communication.

C) NETWORK LAYER

- Implements routing of frames (packets) through the network.
- Defines the most optimum path the packet should take from the source to the destination.
- Defines logical addressing so that any endpoint can be identified.
- Handles congestion in the network.
- The network layer also defines how to fragment a packet into smaller packets to accommodate different media.

D) TRANSPORT LAYER

- Purpose of this layer is to provide a reliable mechanism for the exchange of data between two processes in different computers.
- Ensures that the data units are delivered error free.
- Ensures that data units are delivered in sequence.
- Ensures that there is no loss or duplication of data units.
- Provides connectionless or connection oriented service.
- Provides for the connection management.
- Multiplex multiple a connection over a single channel.

D) SESSION LAYER

- Session layer provides mechanism for controlling the dialogue between the two end systems.
- It defines how to start, control and end conversations (called sessions) between applications.
- This layer requests for a logical connection to be established on an end-user’s request.
- Any necessary log-on or password validation is also handled by this layer.
- Session layer is also responsible for terminating the connection.

- This layer provides services like dialogue discipline which can be full duplex or half duplex.
- Session layer can also provide check-pointing mechanism such that if a failure of some sort occurs between checkpoints, all data can be retransmitted from the last checkpoint.

F) PRESENTATION LAYER

- Presentation layer defines the format in which the data is to be exchanged between the two communicating entities.
- Also handles data compression and data encryption (cryptography).

G) APPLICATION LAYER

- Application layer interacts with application programs and is the highest level of OSI model.
- Application layer contains management functions to support distributed applications.
- Examples of application layer are applications such as file transfer, electronic mail, remote login etc.

Fig. 1.6 below shows summary of all layers of OSI Model:

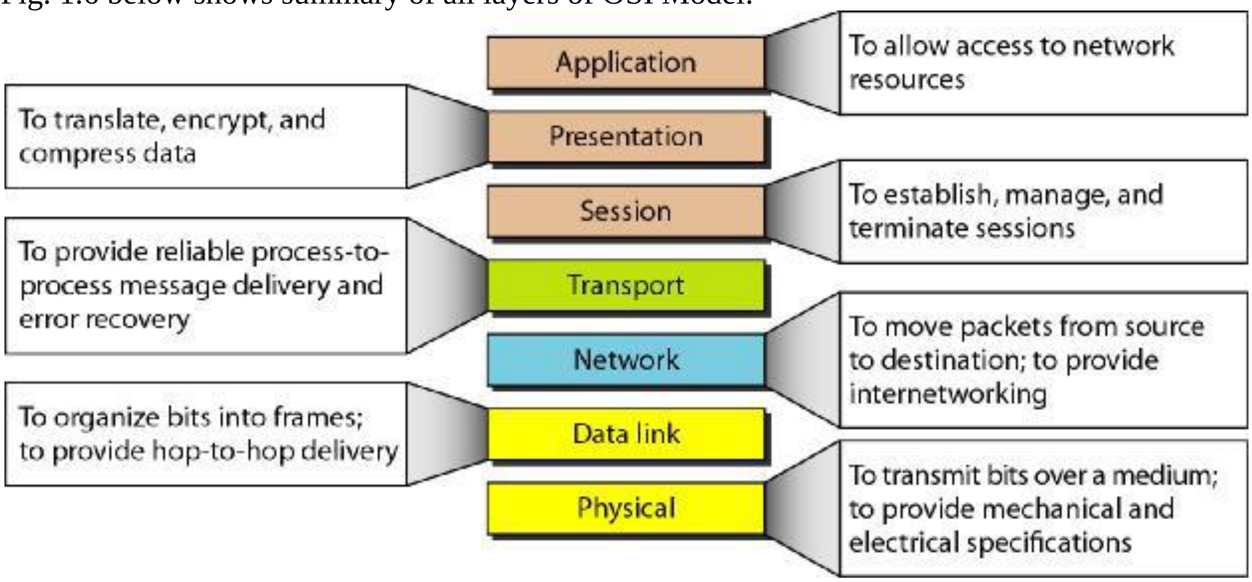


Fig 1.6 Summary of OSI Model

<u>Physical Layer</u>

Physical layer is concerned with transmitting raw bits over a communication channel. The design issues have to do with making sure that when one side sends a 1 bit, it is recieved by the other side as 1 bit and not as 0 bit. In physical layer we deal with the communication medium used for transmission.

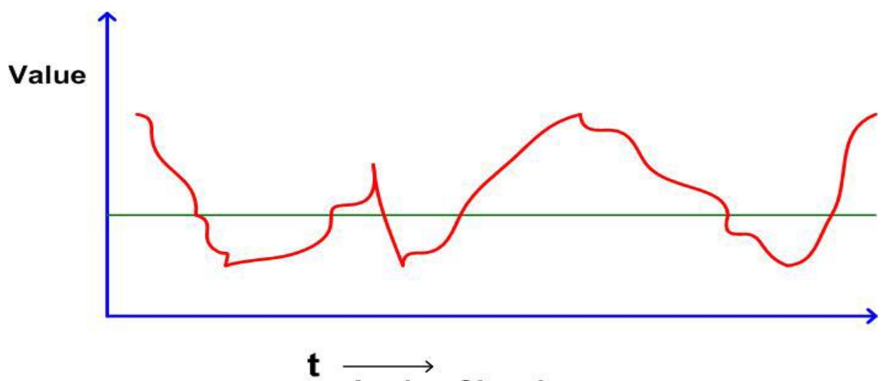
1.5 Data & Signal

Data refers to information that conveys some meaning based on some mutually agreed up rules or conventions between a sender and a receiver and today it comes in a variety of forms such as text, graphics, audio, video and animation.

Data can be of two types; analog and digital. Analog data take on continuous values on some interval. Typical examples of analog data are voice and video. The data that are collected from the real world with the help of transducers are continuous-valued or analog in nature. On the contrary, digital data take on discrete values. Text or character strings can be considered as examples of digital data. Characters are represented by suitable codes, e.g. ASCII code, where each character is represented by a 7-bit code.

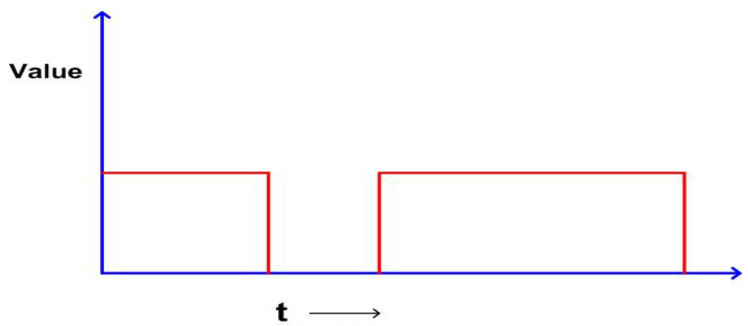
Signal

It is electrical, electronic or optical representation of data, which can be sent over a communication medium. Stated in mathematical terms, a signal is merely a function of the data. For example, a microphone converts voice data into voice signal, which can be sent over a pair of wire. Analog signals are continuous-valued; digital signals are discrete-valued. The independent variable of the signal could be time (speech, for example), space (images), or the integers (denoting the sequencing of letters and numbers in the football score). Figure shows an analog signal.



Analog Signal

Digital signal can have only a limited number of defined values, usually two values 0 and 1, as shown in Fig.



Digital Signal

Signaling: It is an act of sending signal over communication medium
Transmission: Communication of data by propagation and processing is known as transmission.

Signal Characteristics

A signal can be represented as a function of time, i.e. it varies with time. However, it can be also expressed as a function of frequency, i.e. a signal can be considered as a composition of different frequency components. Thus, a signal has both time-domain and frequency domain representation.

Time-domain concepts

A signal is continuous over a period, if
limt->a s (t) = s (a), for all a,
i.e., there is no break in the signal. A signal is discrete if it takes on only a finite number of values. A signal is periodic if and only if
s (t+T) = s (t) for - $\alpha < t < \alpha$,

where T is a constant, known as period. The period is measured in seconds.
In other words, a signal is a periodic signal if it completes a pattern within a measurable time frame. A periodic signal is characterized by the following three parameters.
***Amplitude*:** It is the value of the signal at different instants of time. It is measured in volts.
***Frequency*:** It is inverse of the time period, i.e. f = 1/T. The unit of frequency is Hertz (Hz) or cycles per second.
***Phase*:** It gives a measure of the relative position in time of two signals within a single period. It is represented by φ in degrees or radian.
A sine wave, the most fundamental periodic signal, can be completely characterized by its amplitude, frequency and phase. Examples of sine waves with different amplitude, frequency and phase are shown in below Fig. 1.7. The phase angle φ indicated in the figure is with respect to the reference waveform shown in Fig. 1.7(a).

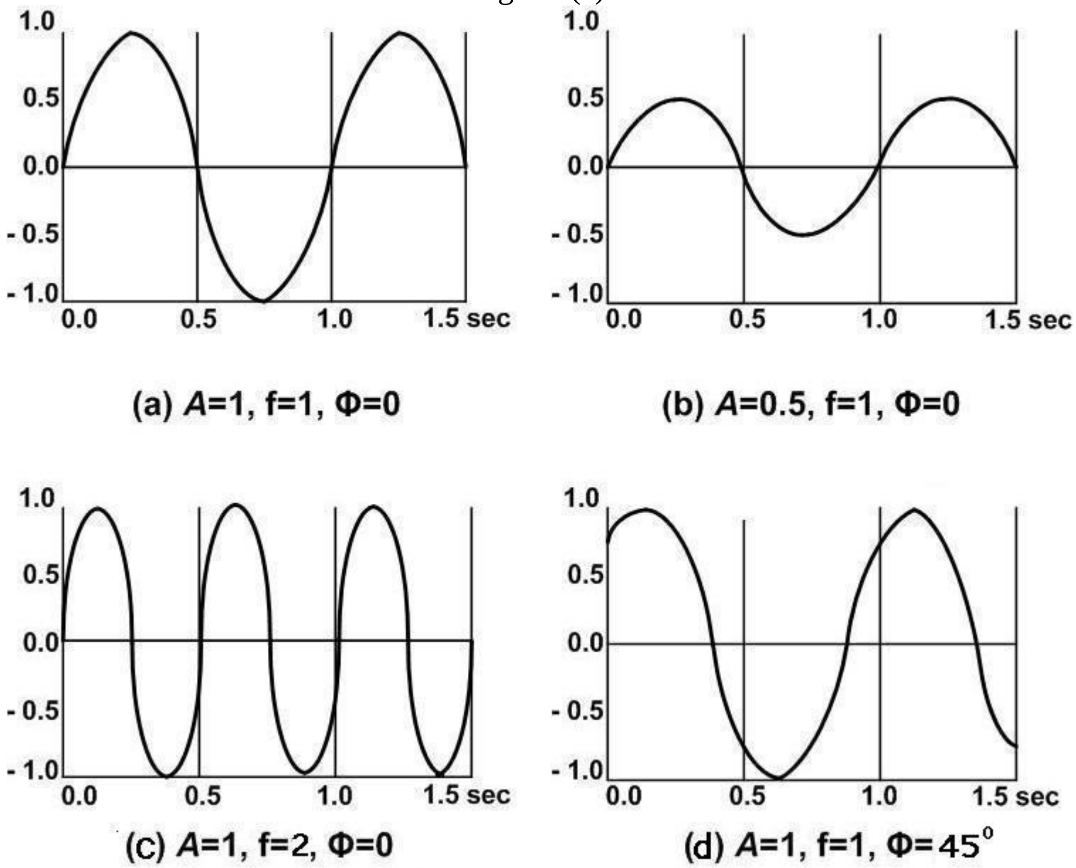


Fig.-1.7 *Examples of signals with different amplitude, frequency and phase*

1.6 Digital data to analog signals

- Digital signals have become very important in wired and wireless communications
- To send data by radio, higher frequency carrier waves are necessary
- Amplitude, frequency, and phase are all used in digital communication systems

A modem (modulator-demodulator) converts digital data to analog signal. There are 3 ways to modulate a digital signal on an analog carrier signal.

Amplitude shift keying (ASK): is a form of modulation which represents digital data as variations in the amplitude of a carrier wave. Two different amplitudes of carrier frequency represent '0' , '1'.

- The only way to achieve high data rates with a narrowband channel is to increase the number of bits/symbol
- The most reliable way to do this is with a combination of amplitude and phase modulation called *quadrature amplitude modulation* (QAM)
- A **constellation diagram** shows the possibilities for a hypothetical system with 16 amplitude/phase combinations

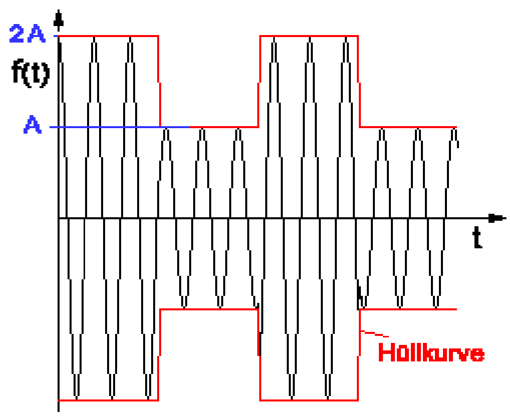


Fig.-1.7 Amplitude shift keying

Frequency shift keying (FSK): In Frequency Shift Keying, the change in frequency define different digits. Two different frequencies near carrier frequency represent '0' ,"1'.

- The simplest form of digital modulation in current use is frequency-shift keying (FSK)
- In its simplest form two frequencies are generated, one corresponding to a binary zero (space) and the other a binary one (mark)
- FSK is reliable in the presence of noise since each signal has only two possible states
- FSK is also used in high-frequency radio systems to transmit teletype information

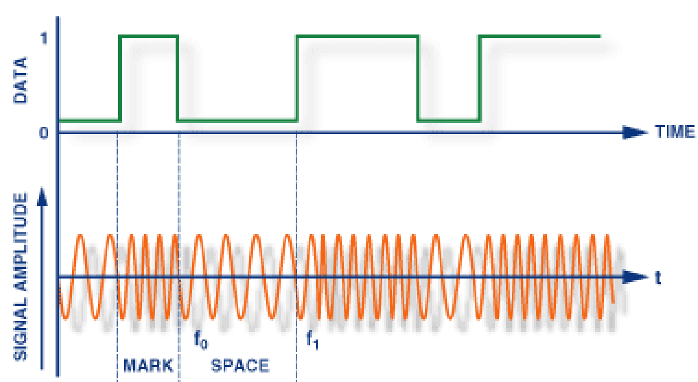


Fig.-1.8 Frequency shift keying

Phase shift keying (PSK): The phase of the carrier is discretely varied in relation either to a reference phase or to the phase of the immediately preceding signal element, in accordance with data being transmitted. Phase of carrier signal is shifted to represent '0' , '1'.

- When higher data rates are required in a band-limited channel that are capable with FSK, *phase-shift keying* is often used
- The phase of each signal is compared with that of the previous signal rather than a reference signal
- This type of PSK is called **delta phase-shift keying** (DPSK)
- Most DPSK modems use a four-phase system called **quadrature phase-shift keying** (QPSK)

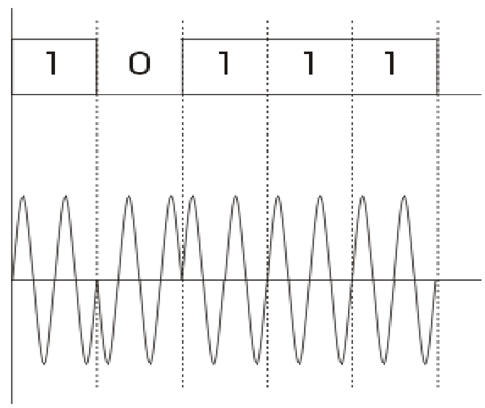


Fig.-1.9 Phase shift keying

1.7 Modem

The word "Modem" is derived from two words; "MODulator" and "DEModulator". From a data communications perspective, a modem is a device that converts the digital bit stream into analog signals that are suitable for transport over standard voice circuits.

Typical modems operate using Frequency Shift Keying (FSK), Phase Shift Keying (PSK), Amplitude Shift Keying (ASK), or a combination of basic schemes.

Telephone Modems

- Telephone modems employ all three of the modulation techniques:
  - FSK
  - PSK
  - QAM
- Because of the limitations of voice-grade telephone lines, these modems are restricted to a bandwidth of about 3 kHz
- The trend in modem design has been towards more sophisticated modulation schemes to achieve the maximum bit rate with available bandwidth

FSK Modems

The first telephone modems used FSK and this technique is still used in specialized applications such as the transmission of call-display information from the central computer to subscriber telephones

FSK Modem Operation

- The Bell 103 standard defines dial-up telephone lines and the frequencies for data communication using FSK
- This standard allows for full-duplex operation
- The modem that places the call is termed the *originating modem*. It transmits with a mark frequency of 1270 Hz and a space frequency of 1070 Hz
- The *answer modem* uses a mark frequency of 2225 Hz and a space frequency of 2025 Hz
- The Bell 103 standard is reliable but slow and is only used as a fallback system for very noisy lines

PSK Modems

- When faster data rates are needed than available with PSK, phase modulation is often used
- Most DPSK (delta phase-shift keying) systems use a four-phase system called quadrature phase-shift keying (QPSK)
- The bit rate is twice the baud rate and is referred to as a dibit system
- The Bell 212A modem is an example of this type of modulation, capable of data rates up to 1200 bits per second

QAM Modems

- Quadrature Amplitude Modulation (QAM) modems are capable of operating at several different speeds, depending upon the quality of the connection
- Modem speeds using the ITU V.34 standard are capable of 33.6 kb/s, full duplex
- V.34 modems monitor line conditions and select the appropriate speeds for a given noise level

V.90 Modems

- The V.90 standards allow data transmissions up to a theoretical limit of 56 kb/s, but because of FCC requirements, the maximum allowable is 54 kb/s
- V.90 modems appear to exceed the *Shannon Limit*
- However, the higher rate is available only in the *downstream* direction and a maximum of 33.6 kb/s is available in the *upstream* direction
- Upper limits vary greatly according to line noise, distance from a telephone substation, and availability of digital connections throughout the phone system

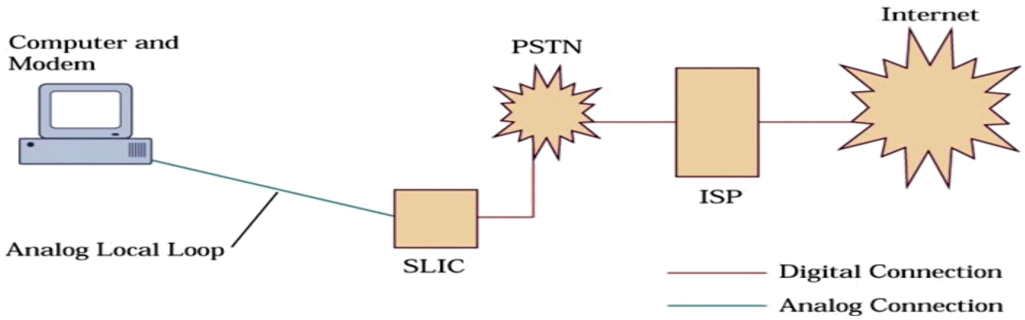


Fig.-1.10 V.90 Modem System

Fax Modems

- Whether a fax machine is stand-alone or part of a microcomputer, the modem used is specialized for facsimile functions
- There are four CCITT standards (groups) for fax transmission
- Groups one and two are now obsolete
- Group three, the most common, uses digital coding of the document followed by analog transmission using QAM at up to 14.4 kb/s
- Group four is for digital lines such as ISDN connections and can transmit up to 64 kb/s

Error Correction and Data Compression with Modems

- High-speed modems incorporate error correction and data compression
- Error correction is important at higher speeds because states in the constellation pattern are closer together
- Error-correction schemes employed are:
  - MNP 2, 3, 4, and 10
  - CCITT V.42
  - Link Access Procedure for Modems (LAPM)
  - All of these schemes use CRC to check each packet
- V.42 bis and MNP5 are the two most popular compression methods used in modems today

Modem-to-Computer Connections

- The UART is the interface between the parallel and serial transmission of data
- There are a number of standards for communication between the computer (DTE) and the modem (DCE), but the most popular is the RS-232C (EIA 232D)
- This standards defines the voltages and pin numbers for both data and control lines on a serial port

Cable Modems and Digital Subscriber Lines

- Because of the high bandwidth available on CATV systems, they're ideally suited for data transmission application
- In a CATV system, all signals go everywhere via a tree network from a single head end

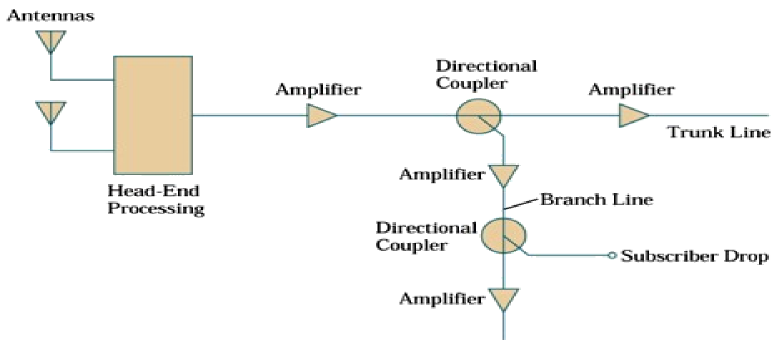


Fig.-1.11 Cable Modem

Cable Modems

- Cable modem systems dedicate one television channel for downstream use and another for upstream use
- On large cable modem systems, distribution hubs are used to add the data from several channels to the fiber-optic backbone
- The equipment that does this is called a *cable modem terminal server* (CCMTS)

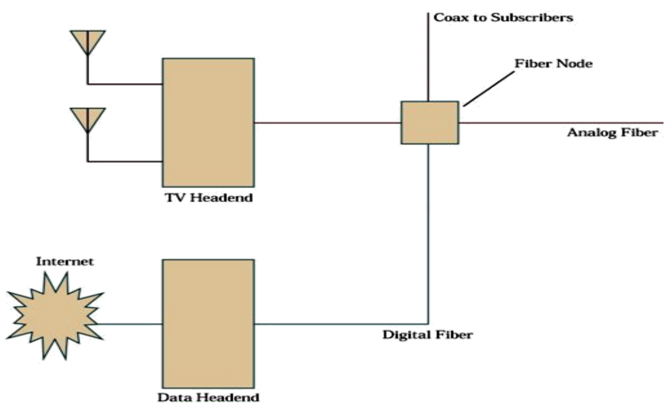


Fig.-1.12 Digital Subscriber Lines (DSL) Systems

Digital Subscriber Lines (DSL) Systems

- DSL achieves high-speed bi-directional data communications over single twisted-pair telephone lines
- There are several varieties of DSL
  - Symmetrical - equal upstream and downstream data rates
  - Asymmetrical - more common, but lower upstream than downstream speeds

Comparison of Cable Modems and ADSL

- Cable modems and ADSL use different technologies to achieve similar results
- Both are *always-on* connections
- Neither interferes with the original use of the service (telephone or CATV)
- Cable modems offer somewhat higher speeds than ADSL, but are subject to degradations in speed due to congestion from multiple users

1.8 Transmission of Digital Signal

Introduction

A computer network is used for communication of data from one station to another station in the network. We have seen that analog or digital data traverses through a communication media in the form of a signal from the source to the destination. The channel bridging the transmitter and the receiver may be a guided transmission medium such as a wire or a wave-guide or it can be an unguided atmospheric or space channel. But, irrespective of the medium, the signal traversing the channel becomes attenuated and distorted with increasing distance. Hence a process is adopted to match the properties of the transmitted signal to the channel characteristics so as to efficiently communicate over the transmission media. There are two alternatives; the data can be either converted to digital or analog signal. Both the approaches have pros and cons. What to be used depends on the situation and the available bandwidth.

Now, either form of data can be encoded into either form of signal. For digital signaling, the data source can be either analog or digital, which is encoded into digital signal, using different encoding techniques.

The basis of analog signaling is a constant frequency signal known as a *carrier signal*, which is chosen to be compatible with the transmission media being used, so that it can traverse a long distance with minimum of attenuation and distortion. Data can be transmitted using these carrier signals by a process called *modulation*, where one or more fundamental parameters of the carrier wave, i.e. amplitude, frequency and phase are being modulated by the source data. The resulting signal, called *modulated signal* traverses the media, which is *demodulated* at the receiving end and the original signal is extracted. All the four possibilities are shown in Fig. 2.4.1.

This lesson will be concerned with various techniques for conversion digital and analog data to digital signal, commonly referred to as **encoding** techniques.

Line coding characteristics

The first approach converts digital data to digital signal, known as line coding, as shown in Fig. 1.13. Important parameters those characteristics line coding techniques are mentioned below.

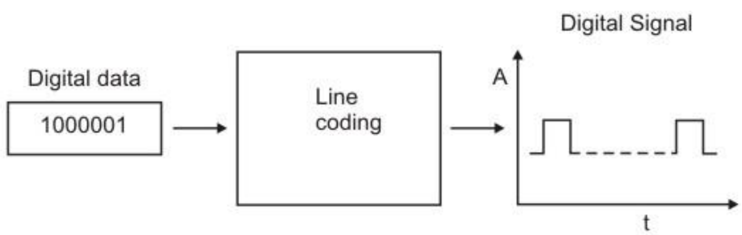
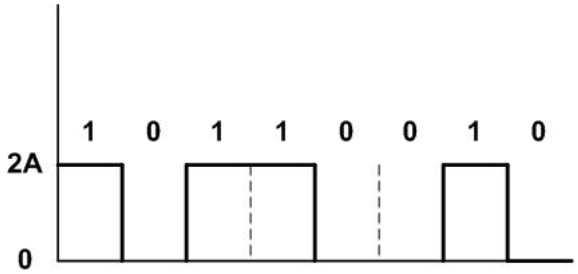


Fig. 1.13 Line coding to convert digital data to digital signal

No of signal levels: This refers to the number values allowed in a signal, known as s**ignal levels**, to represent data. Figure 1.4.(a) shows two signal levels, whereas Fig. 1.4.(b) shows three signal levels to represent binary data.

Bit rate versus Baud rate: The **bit rate** represents the number of bits sent per second, whereas the **baud rate** defines the number of signal elements per second in the signal. Depending on the encoding technique used, baud rate may be more than or less than the data rate.



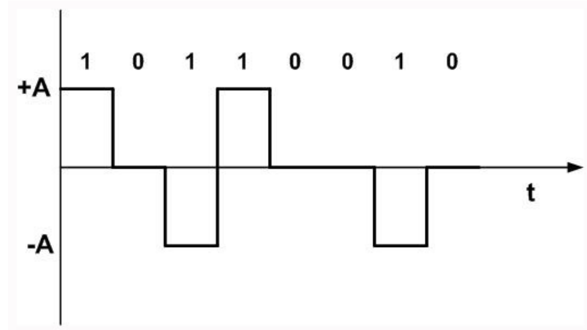


Fig. 1.14 (a) Signal with two voltage levels, (b) Signal with three voltage levels

Encoding Techniques

Line coding techniques can be broadly divided into three broad categories: Unipolar, Polar and Bipolar, as shown in Fig. 1.15.

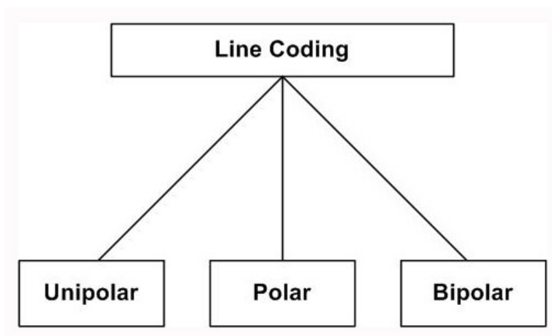


Fig. 1.15 Three basic categories of line coding techniques

Unipolar: In unipolar encoding technique, only two voltage levels are used. It uses only one polarity of voltage level as shown in Fig. 1.16. In this encoding approach, the bit rate same as data rate. Unfortunately, DC component present in the encoded signal and there is loss of synchronization for long sequences of 0's and 1's. It is simple but obsolete.

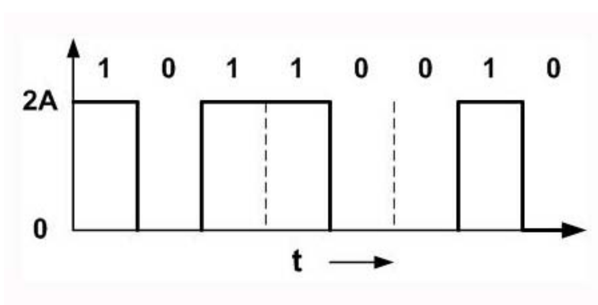


Fig. 1.16 Unipolar encoding with two voltage levels

Polar: Polar encoding technique uses two voltage levels – one positive and the other one negative. Four different encoding schemes shown in Fig. 1.17 under this category discussed below:

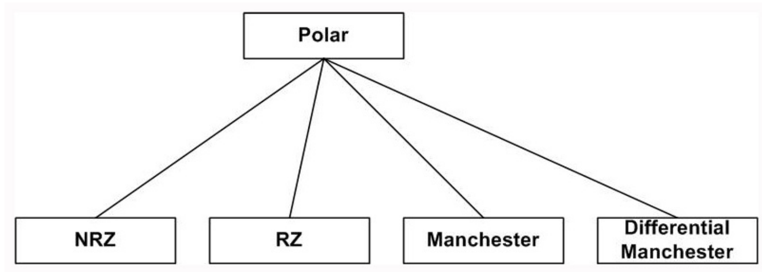
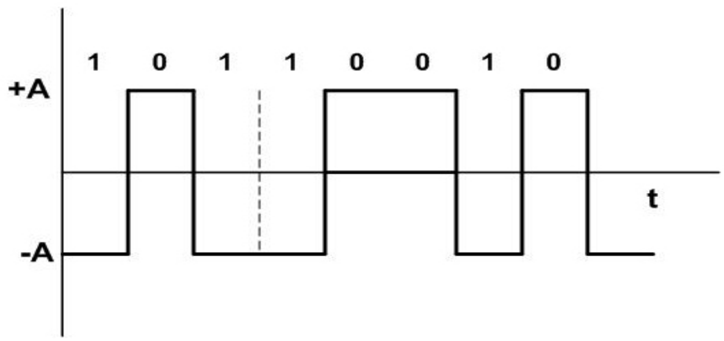
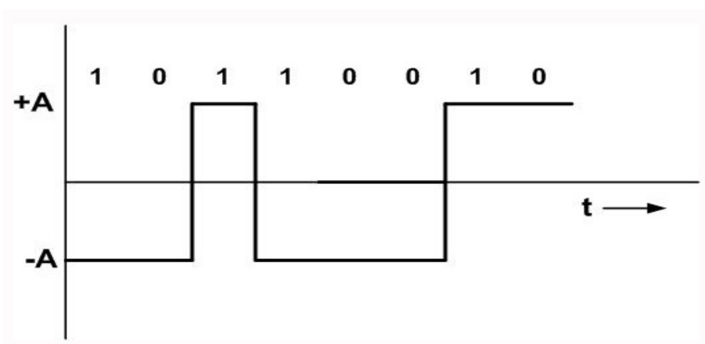


Fig. 1.17 *Encoding Schemes under polar category*

Non Return to zero (NRZ): The most common and easiest way to transmit digital signals is to use two different voltage levels for the two binary digits. Usually a negative voltage is used to represent one binary value and a positive voltage to represent the other. The data is encoded as the presence or absence of a signal transition at the beginning of the bit time. As shown in the figure below, in NRZ encoding, the signal level remains same throughout the bit-period. There are two encoding schemes in NRZ: NRZ-L and NRZ-I, as shown in Fig. 1.18.



NRZ – L 1 = low level 0 = high level



NRZ – I

- For each 1 in the bit sequence, the signal level is inverted.
- A transition from one voltage level to the other represents a 1.

Fig. 1.18 *NRZ encoding scheme*

The **advantages** of NRZ coding are:

- Detecting a transition in presence of noise is more reliable than to compare a value to a threshold.
- NRZ codes are easy to engineer and it makes efficient use of bandwidth.

The spectrum of the NRZ-L and NRZ-I signals are shown in Fig. 2.4.8. It may be noted that most of the energy is concentrated between 0 and half the bit rate. The main limitations are the presence of a dc component and the lack of synchronization capability. When there is long sequence of 0's or 1's, the receiving side will fail to regenerate the clock and synchronization between the transmitter and receiver clocks will fail.

Return to Zero RZ: To ensure synchronization, there must be a signal transition in each bit as shown in Fig. 1.19. Key characteristics of the RZ coding are:

- Three levels
- Bit rate is double than that of data rate
- No dc component
- Good synchronization
- Main limitation is the increase in bandwidth

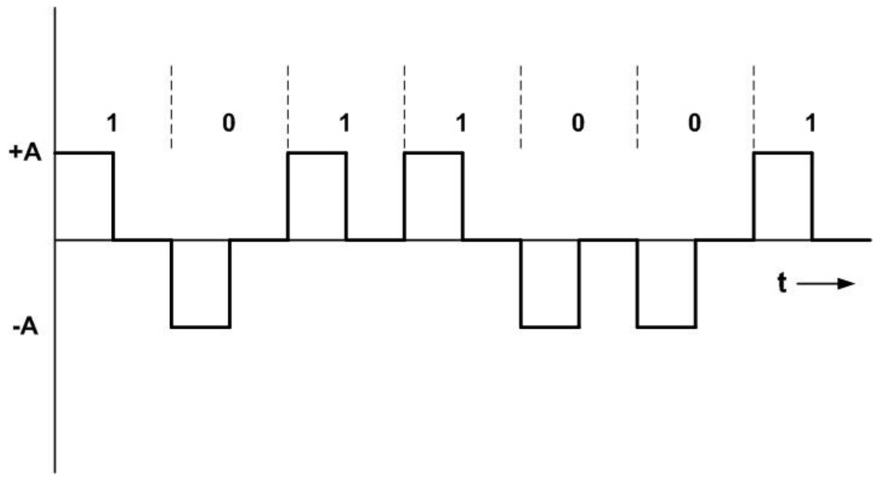


Fig. 1.19 RZ encoding technique

Biphase: To overcome the limitations of NRZ encoding, biphase encoding techniques can be adopted. Manchester and differential Manchester Coding are the two common Biphase techniques in use, as shown in Fig. 1.20. In Manchester coding the mid-bit transition serves as a clocking mechanism and also as data.

In the standard Manchester coding there is a transition at the middle of each bit period. A binary 1 corresponds to a low-to-high transition and a binary 0 to a high-to-low transition in the middle.

In Differential Manchester, inversion in the middle of each bit is used for synchronization. The encoding of a 0 is represented by the presence of a transition both at the beginning and at the middle and 1 is represented by a transition only in the middle of the bit period.

- Key characteristics are:
- Two levels
- No DC component
- Good synchronization

- Higher bandwidth due to doubling of bit rate with respect to data rate

The bandwidth required for biphase techniques are greater than that of NRZ techniques, but due to the predictable transition during each bit time, the receiver can synchronize properly on that transition. Biphase encoded signals have no DC components as shown in Fig. 2.4.11. A Manchester code is now very popular and has been specified for the IEEE 802.3 standard for base band coaxial cables and twisted pair CSMA/CD bus LANs.

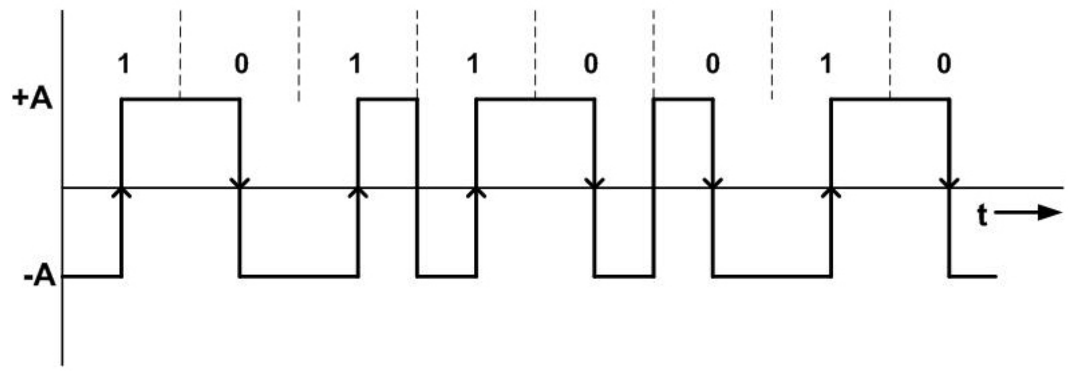


Fig. 1.14 Manchester Encoding

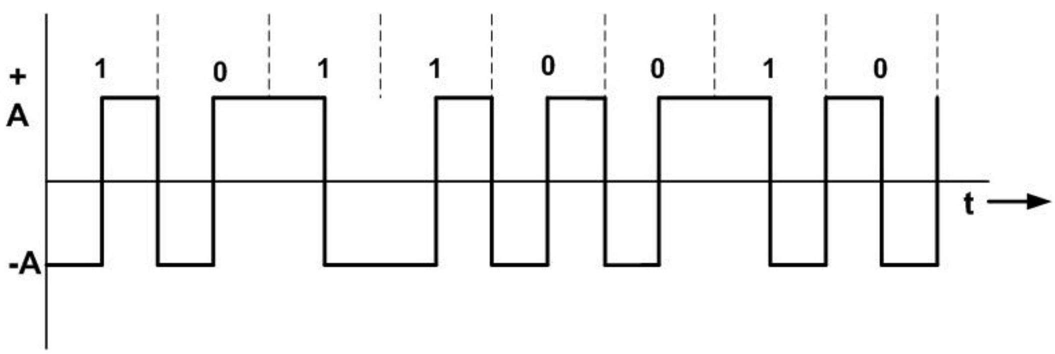


Fig. 1.20 Differential Manchester Encoding

Bipolar Encoding : Bipolar AMI uses three voltage levels. Unlike RZ, the zero level is used to represent a 0 and a binary 1's are represented by alternating positive and negative voltages, as shown in Fig 1.21.

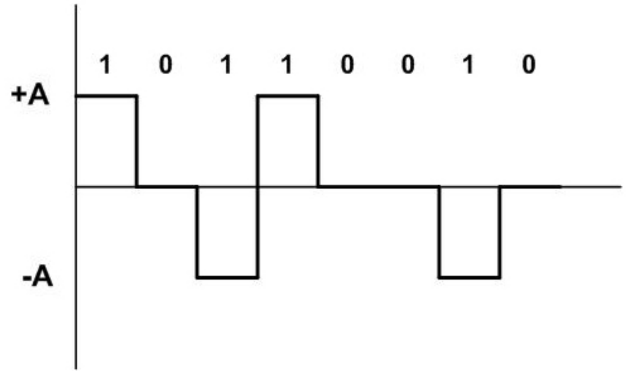


Fig. 1.21 Bipolar AMI Signal

1.9 Analog data to digital signal:

The process is called digitization. Sampling frequency must be at least twice that of highest frequency present in the the signal so that it may be fairly regenerated. Quantization - Max. and Min values of amplitude in the sample are noted. Depending on number of bits (say n) we use we divide the interval (min,max) into 2(^n) number of levels. The amplitude is then approximated to the nearest level by a 'n' bit integer. The digital signal thus consists of blocks of n bits.On reception the process is reversed to produce analog signal. But a lot of data can be lost if fewer bits are used or sampling frequency not so high.

- **Pulse Code Modulation (PCM):** Here intervals are equally spaced. 8 bit PCB uses 256 different levels of amplitude. In non-linear encoding levels may be unequally spaced.
- **Delta Modulation (DM):** Since successive samples do not differ very much we send the differences between previous and present sample. It requires fewer bits than in PCM.

Pulse Code modulation

Pulse Code Modulation involves the following three basic steps as shown in Fig. 1.22.:
Sampling – PAM
Quantization
Line coding

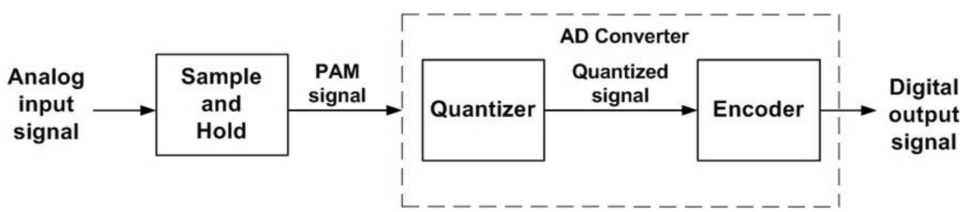


Fig. 1.22-Basic steps of pulse code modulation

Sampling: This process is based on Shannon's sampling theorem. Numbers of samples of the signal are taken at regular intervals, at a rate higher than twice the highest significant signal frequency. This basic step is known as Pulse Amplitude Modulation (PAM) as shown in Fig. 1.23. For example, during the sampling of voice data, in the frequency range 300 to 4000 Hz, 8000 samples per second are sufficient for the coding.

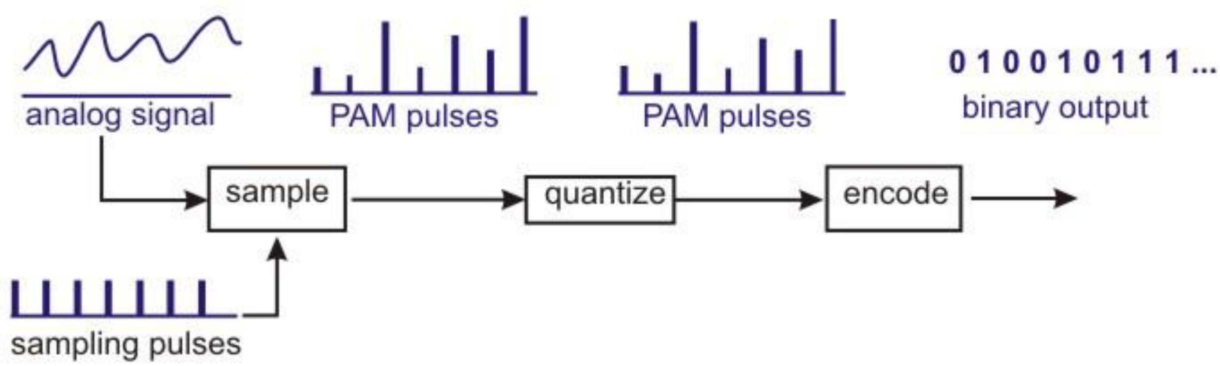


Fig. 1.23 Signal outputs after different steps of PCM

Quantization: The PAM samples are quantized and approximated to n-bit integer by using analog-to-digital converter. For example, if n = 4, then there are 16 (=24) levels available for approximating the PAM signals. This process introduces an error are known as ***quantization* error**. Quantization error depends on step size. Use of uniform step size leads to poorer S/N ratio for small amplitude signals. With the constraint of a fixed number of levels, the situation can be improved using variable step size. The effect of quantization error can be minimized by using a technique known as **companding**. In this case, instead of using uniform stage sizes, the steps are close together at low signal amplitude and further apart at high signal amplitude as shown in Fig. 1.23. It uses a compressor before encoding and expander after decoding. This helps to improve the S/N ratio of the signal.

Line coding: The digital data thus obtained can be encoded into one of digital signals discussed earlier.

At the receiving end, an Digital-to-Analog converter followed by a low-pass filter can be used to get back the analog signal as shown in Fig. 1.24.

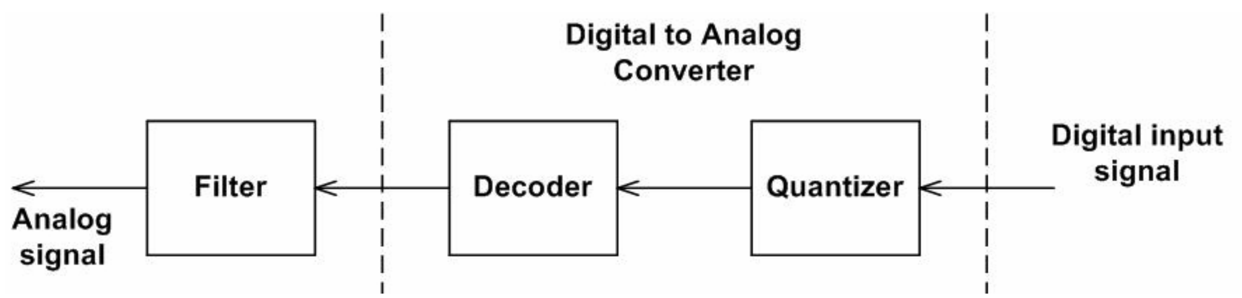


Fig. 1.24 Conversion of digital to analog signal

Limitations: The PCM signal has high bandwidth. For example, let us consider voice signal as input with bandwidth of 4 kHz. Based on Nyquist theorem, the Sampling frequency should be 8 kHz. If an 8-bit ADC is used for conversion to digital data, it generates data rate of 64 Kbps. Therefore, to send voice signal a data rate of 64 Kbps is required. To overcome this problem a technique known as **Differential PCM** (DPCM) can be used. It is based on the observation that voice signal changes slowly. So, the difference between two consecutive sample values may be sent. Since the signal changes slowly, the difference between two consecutive sample values will be small and fewer number of bits can be used with consequent reduction in data rates.

Delta Modulation (DM)

Delta Modulation is a very popular alternative of PCM with much reduced complexity. Here the analog input is approximated by a staircase function, which moves up or down by one quantization level (a constant amount) at each sampling interval. Each sample delta modulation process can be represented by a single binary digit, which makes it more efficient than the PCM technique. In this modulation technique, instead of sending the entire encoding of each and every sample, we just send the change from previous sample. If the difference between analog input and the feedback signal is positive, then encoded output is 1, otherwise it is 0. So, only one bit is to be sent per sample. Figure 1.25 shows the Delta modulation operation.

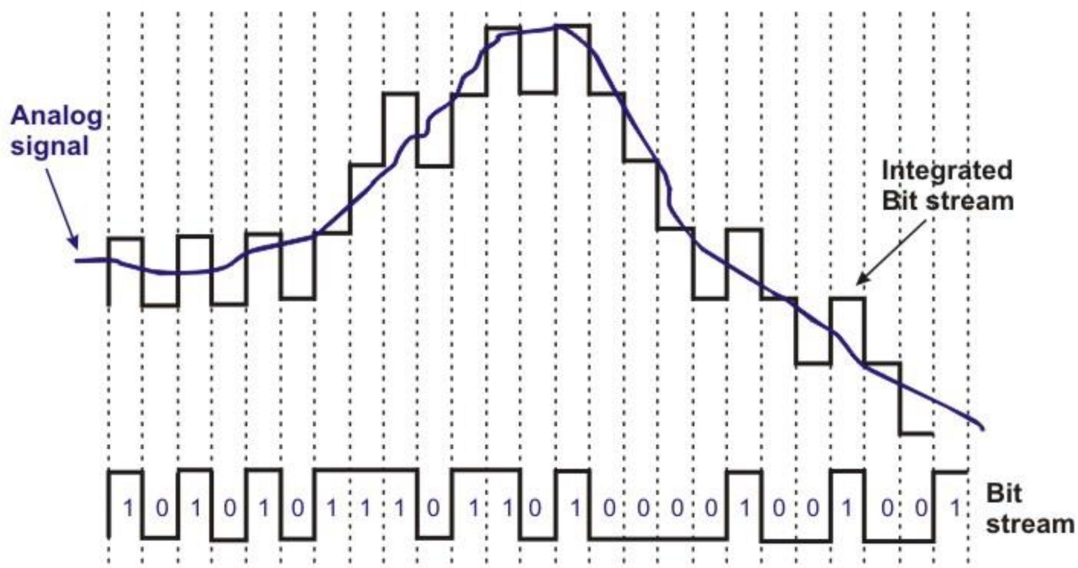


Fig. 1.25 Delta modulation

Advantages: Main advantage of Delta Modulation is its simplicity of implementation as shown in Fig. 2.4.21. Each sample is represented by a single binary digit, which makes it more efficient than the PCM technique. Two important parameters:

- The size of the step
- The sampling rate

In the transmitting end, the analog input is compared to the most recent value of the approximating staircase function at each sampling time. If the value of the sampled waveform that of the staircase functions, a 1 is generated; otherwise a 0 is generated. The output of the DM is a binary sequence that can be used to reconstruct the staircase function at the receiving end as shown in Fig. 1.26

Disadvantages: Fixed step size leads to overloading. Overloading occurs not only due to higher voltage, but due to its slope as shown in Fig. 1.25. This problem can be overcome using adaptive delta modulation. The steps sizes are small, when the signal changes are small. The steps sizes are large, when the signal changes are large

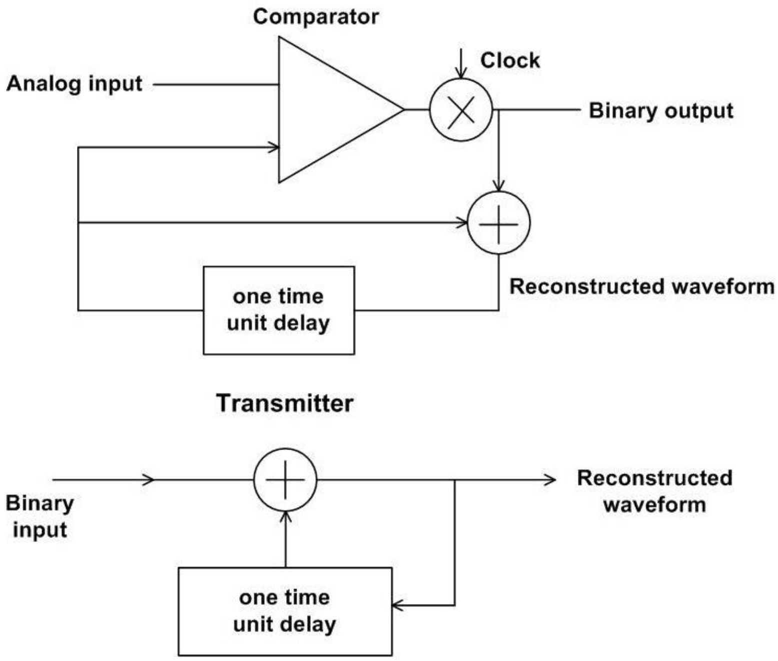


Fig. 1.26 Implementation of Delta modulation

1. Why do you need encoding of data before sending over a medium?
Ans: Suitable encoding of data is required in order to transmit signal with minimum attenuation and optimize the use of transmission media in terms of data rate and error rate.
2. What are the four possible encoding techniques? Give examples.
Ans: The four possible encoding techniques are
Digital Data to Digital Signal; Example - Transmitter
Analog Data to Digital Signal; Example - Codec (Coder-Decoder)
Digital Data to Analog Signal; Example - Modem
Analog Data to Digital Signal; Example - Telephone

3. Between RZ and NRZ encoding techniques, which requires higher bandwidth and why?

Ans: RZ encoding requires more bandwidth, as it requires two signal changes to encode one bit**.**
4.How does Manchester encoding differ from differential Manchester encoding?
Ans: In the Manchester encoding, a low-to-high transition represents a 1, and a high-to-low transition represents a 0. There is a transition at the middle of each bit period, which serves the purpose of synchronization and encoding of data.
In Differential Manchester, the encoding of a 0 is represented by the presence of a transition at the beginning of a bit period, and a 1 is represented by the absence of a transition at the beginning of a bit period. In this case, the mid bit transition is only used for synchronization.
5. How Manchester encoding helps in achieving better synchronization?

Ans: In Manchester encoding, there is a transition in the middle of each bit period and the receiver can synchronize on that transition. Hence better synchronization is achieved.
6. Why B8ZS coding is preferred over Manchester encoding for long distance communication?

Ans: The B8ZS encoding is preferred over Manchester encoding, because B8ZS encoding requires lesser bandwidth than Manchester encoding.

7. Why is it necessary to limit the band of a signal before performing sampling?

Ans: It is necessary to limit the bandwidth of a signal before sampling so that the basic requirement of sampling theorem, i.e. the sampling rate should twice or more than twice the maximum frequency component of the signal, is satisfied. This is known as Nyquist rate. If it is violated, original signal cannot be recovered from the sampled signal.

8. Distinguish between PAM and PCM signals?

Ans: In order to convert Analog data to Digital signal, initially sampling is done on the analog data by using Sample & Hold (S/H) circuit. The output of the S/H circuit is known as PAM (Pulse Amplitude Modulated) signal. The PAM signal is then converted to PCM

1.10 Digital Data Communication Techniques:

For two devices linked by a transmission medium to exchange data ,a high degree of co-operation is required. Typically data is transmitted one bit at a time. The timing (rate, duration,spacing) of these bits must be same for transmitter and receiver. There are two options for transmission of bits.

1. **Parallel** All bits of a byte are transferred simultaneously on separate parallel wires. Synchronization between multiple bits is required which becomes difficult over large distance. Gives large band width but expensive. Practical only for devices close to each other.
2. **Serial** Bits transferred serially one after other.Gives less bandwidth but cheaper. Suitable for transmission over long distances.

Transmission Techniques:

1. **Asynchronous:** Small blocks of bits(generally bytes) are sent at a time without any time relation between consecutive bytes .when no transmission occurs a default state is maintained corresponding to bit 1. Due to arbitrary delay between consecutive bytes,the time occurrences of the clock pulses at the receiving end need to be synchronized for each byte. This is achieved by providing 2 extra bits start and stop.

   Start bit: It is prefixed to each byte and equals 0. Thus it ensures a transition from 1 to 0 at onset of transmission of byte.The leading edge of start bit is used as a reference for generating clock pulses at required sampling instants. Thus each onset of a byte results in resynchronization of receiver clock.

   Stop bit: To ensure that transition from 1 to 0 is always present at beginning of a byte it is necessary that default state be 1. But there may be two bytes one immediately following the other and if last bit of first byte is 0, transition from 1 to 0 will not occur . Therefore a stop bit is suffixed to each byte equaling 1. It's duration is usually 1,1.5,2 bits.

Asynchronous transmission is simple and cheap but requires an overhead of 3 bits i.e. for 7 bit code 2 (start ,stop bits)+1 parity bit implying 30% overhead.However % can be reduced by sending larger blocks of data but then timing errors between receiver and sender can not be tolerated beyond [50/no. of bits in block] % (assuming sampling is done at middle of bit interval). It will not only result in incorrect sampling but also misaligned bit count i.e. a data bit can be mistaken for stop bit if receiver's clock is faster.

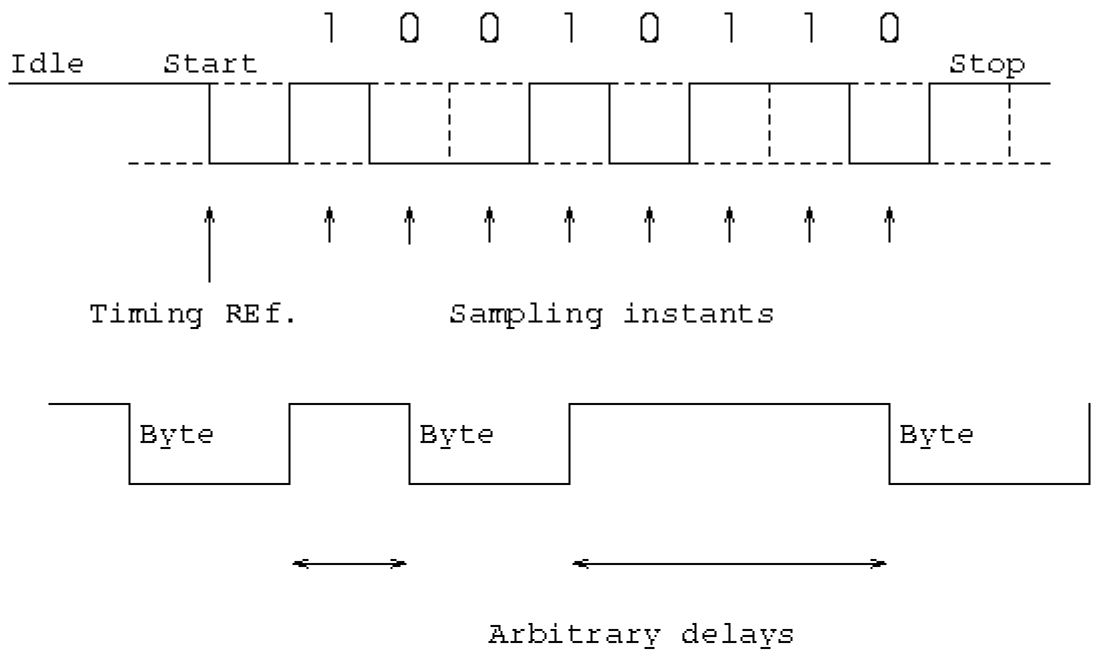


Fig. 1.27 Asynchronous transmission

2. **Synchronous** - Larger blocks of bits are successfully transmitted. Blocks of data are either treated as sequence of bits or bytes. To prevent timing drift clocks at two ends need to be synchronized. This can done in two ways:

1. Provide a separate clock line between receiver and transmitter. OR
2. Clocking information is embedded in data signal i.e. biphase coding for digital signals.

Still another level of synchronization is required so that receiver determines beginning or end of block of data. Hence each block begins with a start code and ends with a stop code.These are in general same known as flag that is unique sequence of fixed no. of bits.In addition some control characters encompass data within these flags. Data+control information is called a frame. Since any arbitrary bit pattern can be transmitted there is no assurance that bit pattern for flag will not appear inside the frame thus destroying frame level synchronization. So to avoid this we use bit stuffing

1.11 Transmission Media

Types of Medium

Transmission media can be defined as physical path between transmitter and receiver in a data transmission system. And it may be classified into two types as shown in Fig. 1.28.

Guided: Transmission capacity depends critically on the medium, the length, and whether the medium is point-to-point or multipoint (e.g. LAN). Examples are co-axial cable, twisted pair, and optical fiber.

Unguided: provides a means for transmitting electro-magnetic signals but do not guide them. Example wireless transmission.

Characteristics and quality of data transmission are determined by medium and signal characteristics. For guided media, the medium is more important in determining the limitations of transmission. While in case of unguided media, the bandwidth of the signal produced by the transmitting antenna and the size of the antenna is more important than the medium. Signals at lower frequencies are omni-directional (propagate in all directions). For higher frequencies, focusing the signals into a directional beam is possible. These properties determine what kind of media one should use in a particular application. In this lesson we shall discuss the characteristics of various transmission media, both guided and unguied. Atmosphere Communication Links.

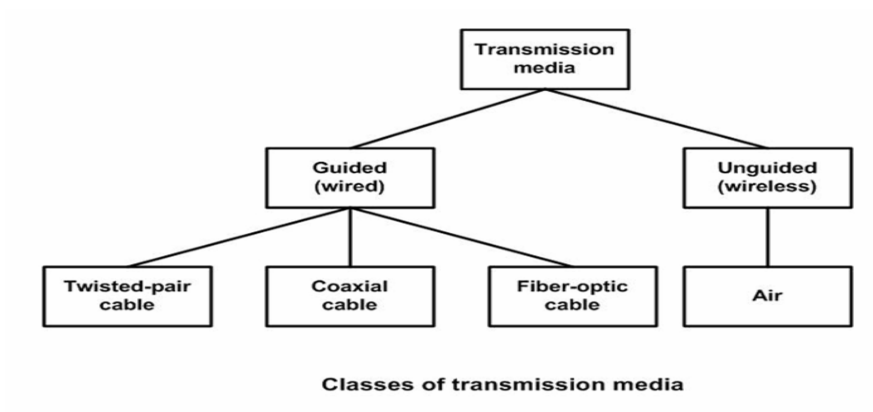


Fig. 1.28 Classification of the transmission media Classification of the transmission media

Guided transmission media

In this section we shall discuss about the most commonly used guided transmission media such as twisted-pair of cable, coaxial cable and optical fiber.

Twisted Pair

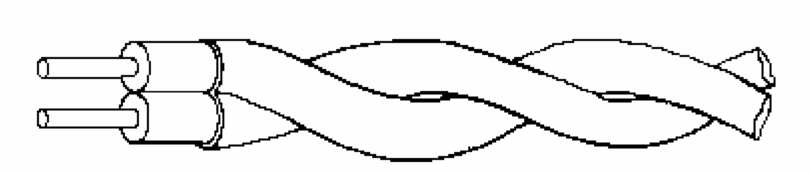


Fig. 1.29 CAT5 cable (twisted cable)

In twisted pair technology, two copper wires are strung between two points:

- The two wires are typically ``twisted" together in a helix to reduce interference between the two conductors as shown in Fig.1.29. Twisting decreases the cross-talk interference between adjacent pairs in a cable. Typically, a number of pairs are bundled together into a cable by wrapping them in a tough protective sheath.
- Can carry both analog and digital signals. Actually, they carry only analog signals. However, the ``analog" signals can very closely correspond to the square waves representing bits, so we often think of them as carrying digital data.
- Data rates of several Mbps common.
- Spans distances of several kilometers.
- Data rate determined by wire thickness and length. In addition, shielding to eliminate interference from other wires impacts signal-to-noise ratio, and ultimately, the data rate.
- Good, low-cost communication. Indeed, many sites already have twisted pair installed in offices -- existing phone lines!

Typical characteristics: Twisted-pair can be used for both analog and digital communication. The data rate that can be supported over a twisted-pair is inversely proportional to the square of the line length. Maximum transmission distance of 1 Km can be achieved for data rates up to 1 Mb/s. For analog voice signals, amplifiers are required about every 6 Km and for digital signals, repeaters are needed for about 2 Km. To reduce interference, the twisted pair can be shielded with metallic braid. This type of wire is known as Shielded Twisted-Pair (STP) and the other form is known as Unshielded Twisted-Pair (UTP).

Use: The oldest and the most popular use of twisted pair are in telephony. In LAN it is commonly used for point-to-point short distance communication (say, 100m) within a building or a room.

Advantages :

1. Cheaper and far easier to splice
2. Less susceptible to electrical interference caused by nearby equipment or wires.
3. In turn are less likely to cause interference themselves.
4. Because it is electrically "cleaner", STP wire can carry data at a faster speed.

Disadvantages :

1. STP wire is that it is physically larger and more expensive than twisted pair wire.
2. more expensive to install compare to twisted pair cable.
3. the thicker the cable, the more difficult to work with.

Coaxial Cable

Coaxial cable (or coax) carries signals of higher frequency ranges than those in twisted pair cable, in part because the two media are constructed quite differently. Instead of having two wires, coax has a central core conductor of solid or stranded wire (usually copper) enclosed in an insulating sheath, which is, in turn, encased in an outer conductor of metal foil, braid, or a combination of the two. The outer metallic wrapping serves both as a shield against noise and as the second conductor, which completes the circuit. This outer conductor is also enclosed in an insulating sheath, and the whole cable is protected by a plastic cover.

Base band Coaxial

With ``coax", the medium consists of a copper core surrounded by insulating material and a braided outer conductor as shown in Fig. 1.30. The term base band indicates digital transmission (as opposed to broadband analog).

Physical connection consists of metal pin touching the copper core. There are two common ways to connect to a coaxial cable:

- With vampire taps, a metal pin is inserted into the copper core. A special tool drills a hole into the cable, removing a small section of the insulation, and a special connector is screwed into the hole. The tap makes contact with the copper core.
- With a T-junction, the cable is cut in half, and both halves connect to the T-junction. A T-connector is analogous to the signal splitters used to hook up multiple TVs to the same cable wire.

Characteristics: Co-axial cable has superior frequency characteristics compared to twisted-pair and can be used for both analog and digital signaling. In baseband LAN, the data rates lies in the range of 1 KHz to 20 MHz over a distance in the range of 1 Km. Co-axial cables typically have a diameter of 3/8". Coaxial cables are used both for baseband and broadband communication. For broadband CATV application coaxial cable of 1/2" diameter and 75 Ω impedance is used. This cable offers bandwidths of 300 to 400 MHz facilitating high-speed data communication with low bit-error rate. In broadband signaling, signal propagates only in one direction, in contrast to propagation in both directions in baseband signaling. Broadband cabling uses either dual-cable scheme or single-cable scheme with a headend to facilitate flow of signal in one direction. Because of the shielded, concentric construction, co-axial cable is less susceptible to interference and cross talk than the twisted-pair. For long distance communication, repeaters are needed for every kilometer or so. Data rate depends on physical properties of cable, but 10 Mbps is typical.

Use: One of the most popular use of co-axial cable is in cable TV (CATV) for the distribution of TV signals. Another importance use of co-axial cable is in LAN.

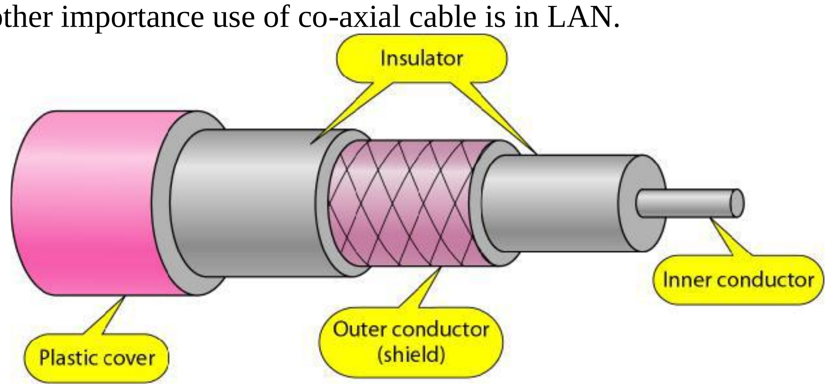


Fig. 1.30 Co-axial cable

Broadband Coaxial

The term broadband refers to analog transmission over coaxial cable. (Note, however, that the telephone folks use broadband to refer to any channel wider than 4 kHz). The technology:

- Typically bandwidth of 300 MHz, total data rate of about 150 Mbps.
- Operates at distances up to 100 km (metropolitan area!).
- Uses analog signaling.
- Technology used in cable television. Thus, it is already available at sites such as universities that may have TV classes.
- Total available spectrum typically divided into smaller channels of 6 MHz each. That is, to get more than 6MHz of bandwidth, you have to use two smaller channels and somehow combine the signals.
- Requires amplifiers to boost signal strength; because amplifiers are one way, data flowin only one direction.

Two types of systems have emerged:

1. Dual cable systems use two cables, one for transmission in each direction:
   o One cable is used for receiving data.
   o Second cable used to communicate with *headend*. When a node wishes to transmit data, it sends the data to a special node called the *headend*. The headend then resends the data on the first cable. Thus, the headend acts as a root of the tree, and all data must be sent to the root for redistribution to the other nodes.
2. *Midsplit* systems divide the raw channel into two smaller channels, with each sub channel having the same purpose as above.

Which is better, broadband or base band? There is rarely a simple answer to such questions. Base band is simple to install, interfaces are inexpensive, but doesn't have the same range. Broadband is more complicated, more expensive, and requires regular adjustment by a trained technician, but offers more services (e.g., it carries audio and video too).

Fiber Optics

In fiber optic technology, the medium consists of a hair-width strand of silicon or glass, and the signal consists of pulses of light. For instance, a pulse of light means ``1'', lack of pulse means ``0''. It has a cylindrical shape and consists of three concentric sections: the core, the cladding, and the jacket as shown in Fig. 1.31.

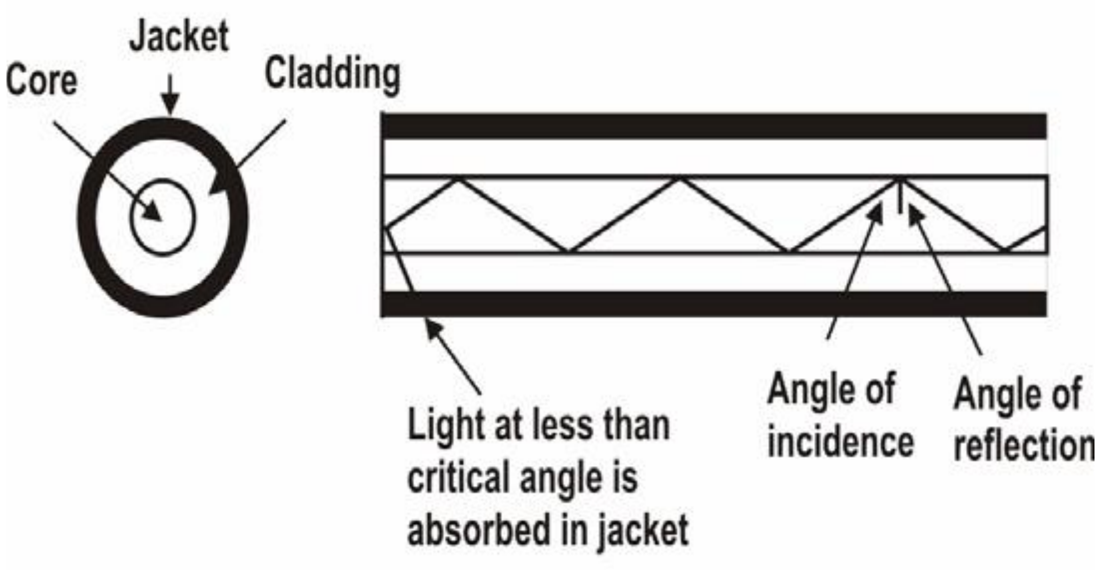


Fig. 1.31 Optical Fiber

The core, innermost section consists of a single solid dielectric cylinder of diameter d1 and of refractive index n1. The core is surrounded by a solid dielectric cladding of refractive index n2 that is less than n1. As a consequence, the light is propagated through multiple total internal reflection. The core material is usually made of ultra pure fused silica or glass and the cladding is either made of glass or plastic. The cladding is surrounded by a jacket made of plastic. The jacket is used to protect against moisture, abrasion, crushing and other environmental hazards.

Three components are required:

1. Fiber medium: Current technology carries light pulses for tremendous distances (e.g., 100s of kilometers) with virtually no signal loss.

2. Light source: typically a Light Emitting Diode (LED) or laser diode. Running current through the material generates a pulse of light.
3. A photo diode light detector, which converts light pulses into electrical signals.

Advantages:

1. Very high data rate, low error rate. 1000 Mbps (1 Gbps) over distances of kilometers common. Error rates are so low they are almost negligible.
2. Difficult to tap, which makes it hard for unauthorized taps as well. This is responsible for higher reliability of this medium.

   How difficult is it to prevent coax taps? Very difficult indeed, unless one can keep the entire cable in a locked room!

3. Much thinner (per logical phone line) than existing copper circuits. Because of its thinness, phone companies can replace thick copper wiring with fibers having much more capacity for same volume. This is important because it means that aggregate phone capacity can be upgraded without the need for finding more physical space to hire the new cables.
4. Not susceptible to electrical interference (lightning) or corrosion (rust).
5. Greater repeater distance than coax.

Disadvantages:

1. Difficult to tap. It really is point-to-point technology. In contrast, tapping into coax is trivial. No special training or expensive tools or parts are required.
2. One-way channel. Two fibers needed to get full duplex (both ways) communication.

Optical Fiber works in three different types of modes (or we can say that we have 3 types of communication using Optical fiber). Optical fibers are available in two varieties; Multi-Mode Fiber (MMF) and Single-Mode Fiber (SMF). For multi-mode fiber the core and cladding diameter lies in the range 50-200µm and 125-400µm, respectively. Whereas in single-mode fiber, the core and cladding diameters lie in the range 8-12µm and 125µm, respectively. Single-mode fibers are also known as Mono-Mode Fiber. Moreover, both single-mode and multi-mode fibers can have two types; step index and graded index. In the former case the refractive index of the core is uniform throughout and at the core cladding boundary there is an abrupt change in refractive index. In the later case, the refractive index of the core varies radially from the centre to the core-cladding boundary from n1 to n2 in a linear manner. Fig. 1.32 shows the optical fiber transmission modes.

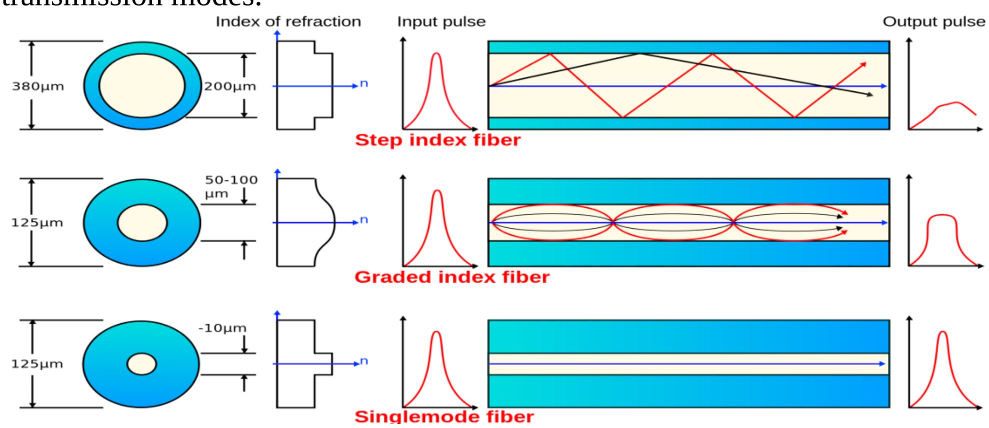


Fig. 1.32 Types transmission modes

Characteristics: Optical fiber acts as a dielectric waveguide that operates at optical frequencies (1014 to 1015 Hz). Three frequency bands centered around 850,1300 and 1500 nanometers are used for best results. When light is applied at one end of the optical fiber core, it reaches the other end by means of total internal reflection because of the choice of refractive index of core and cladding material (n1 > n2). The light source can be either light emitting diode (LED) or injection laser diode (ILD). These semiconductor devices emit a beam of light when a voltage is applied across the device. At the receiving end, a photodiode can be used to detect the signal-encoded light. Either PIN detector or APD (Avalanche photodiode) detector can be used as the light detector.

In a multi-mode fiber, the quality of signal-encoded light deteriorates more rapidly than single-mode fiber, because of interference of many light rays. As a consequence, single-mode fiber allows longer distances without repeater. For multi-mode fiber, the typical maximum length of the cable without a repeater is 2km, whereas for single-mode fiber it is 20km.

Fiber Uses: Because of greater bandwidth (2Gbps), smaller diameter, lighter weight, low attenuation, immunity to electromagnetic interference and longer repeater spacing, optical fiber cables are finding widespread use in long-distance telecommunications. Especially, the single mode fiber is suitable for this purpose. Fiber optic cables are also used in high-speed LAN applications. Multi-mode fiber is commonly used in LAN.

- Long-haul trunks-increasingly common in telephone network (Sprint ads)
- Metropolitan trunks-without repeaters (average 8 miles in length)
- Rural exchange trunks-link towns and villages
- Local loops-direct from central exchange to a subscriber (business or home)
- Local area networks-100Mbps ring networks.

Unguided Transmission

Unguided transmission is used when running a physical cable (either fiber or copper) between two end points is not possible. For example, running wires between buildings is probably not legal if the building is separated by a public street.

Infrared signals typically used for short distances (across the street or within same room),

Microwave signals commonly used for longer distances (10's of km). Sender and receiver use some sort of dish antenna.

Difficulties:

1. Weather interferes with signals. For instance, clouds, rain, lightning, etc. may adversely affect communication.
2. Radio transmissions easy to tap. A big concern for companies worried about competitors stealing plans.
3. Signals bouncing off of structures may lead to out-of-phase signals that the receiver must filter out.

Satellite Communication

Satellite communication is based on ideas similar to those used for line-of-sight. A communication satellite is essentially a big microwave repeater or relay station in the sky. Microwave signals from a ground station is picked up by a transponder, amplifies the signal and rebroadcasts it in another frequency, which can be received by ground stations at long distances as shown in Fig. 1.33.

To keep the satellite stationary with respect to the ground based stations, the satellite is placed in a geostationary orbit above the equator at an altitude of about 36,000 km. As the spacing between two satellites on the equatorial plane should not be closer than 40, there can be 360/4 =

90 communication satellites in the sky at a time. A satellite can be used for point-to-point communication between two ground-based stations or it can be used to broadcast a signal received from one station to many ground-based stations as shown in Fig. 1.33. Number of geo-synchronous satellites limited (about 90 total, to minimize interference). International agreements regulate how satellites are used, and how frequencies are allocated. Weather affects certain frequencies. Satellite transmission differs from terrestrial communication in another important way: One-way propagation delay is roughly 270 ms. In interactive terms, propagation delay alone inserts a 1 second delay between typing a character and receiving its echo.

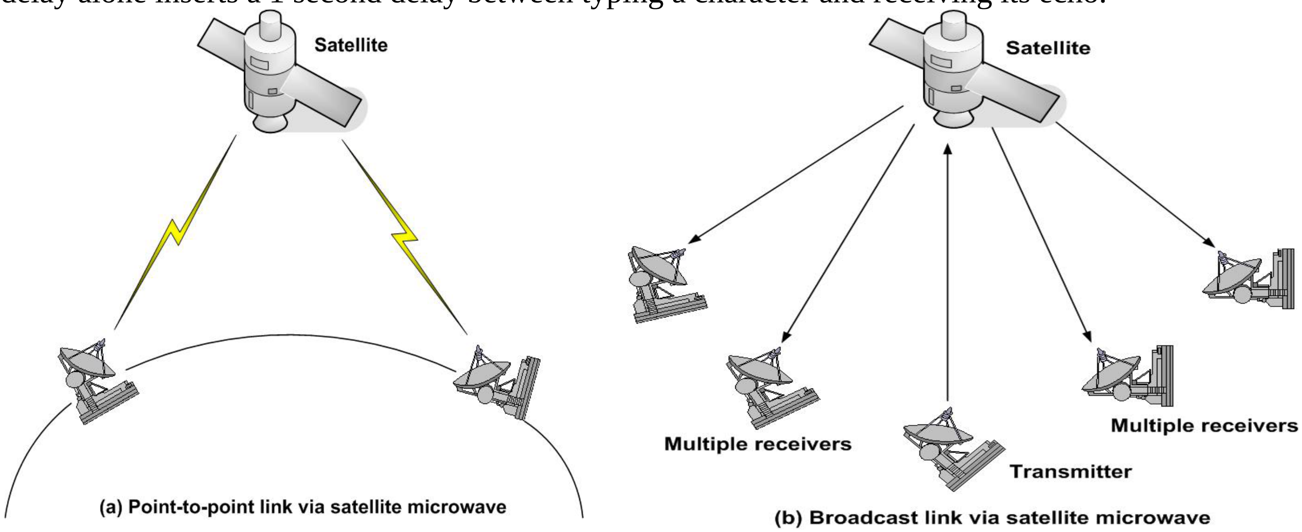


Fig. 1.33 Satellite communication

Characteristics: Optimum frequency range for satellite communication is 1 to 10 GHz. The most popular frequency band is referred to as 4/6 band, which uses 3.7 to 4.2 GHz for down link and 5.925 to 6.425 for uplink transmissions. The 500 MHz bandwidth is usually split over a dozen transponders, each with 36 MHz bandwidth. Each 36 MHz bandwidth is shared by time division multiplexing. As this preferred band is already saturated, the next highest band available is referred to as 12/14 GHz. It uses 14 to 14.5GHz for upward transmission and 11.7 to 12.2 GHz for downward transmissions. Communication satellites have several unique properties. The most important is the long communication delay for the round trip (about 270 ms) because of the long distance (about 72,000 km) the signal has to travel between two earth stations. This poses a number of problems, which are to be tackled for successful and reliable communication.

Another interesting property of satellite communication is its broadcast capability. All stations under the downward beam can receive the transmission. It may be necessary to send encrypted data to protect against piracy.

Use: Now-a-days communication satellites are not only used to handle telephone, telex and television traffic over long distances, but are used to support various internet based services such as e-mail, FTP, World Wide Web (WWW), etc. New types of services, based on communication satellites, are emerging.

Comparison/contrast with other technologies:

1. Propagation delay very high. On LANs, for example, propagation time is in nanoseconds essentially negligible.
2. One of few alternatives to phone companies for long distances.
3. Uses broadcast technology over a wide area - everyone on earth could receive a message at the same time!
4. Easy to place unauthorized taps into signal.

Satellites have recently fallen out of favor relative to fiber.

However, fiber has one big disadvantage: no one has it coming into their house or building, whereas anyone can place an antenna on a roof and lease a satellite channel.

Q.1 Why does single-mode fibres are used for large distance communications rather than multi-mode fibres?

Ans: In a multi-mode fiber, the quality of signal-encoded light deteriorates more rapidly than single-mode fiber, because of interference of many light rays. As a consequence, single-mode fiber allows longer distances without repeater. For multi-mode fiber, the typical maximum length of the cable without a repeater is 2km, whereas for single-mode fiber it is 20km.

Q.2 What is crosstalk? How is it minimized in case of twisted-pair of wire?

Ans:

(a) Crosstalk refers to the picking up of electromagnetic signals from other adjacent wires by electromagnetic induction. (b) When a pair of wires is twisted together, the electromagnetic signals generated by the two wires cancel each other as these are of opposite polarity. This helps to reduce the susceptibility of interference to the adjacent wires.

In a nework nodes are connected through links. The communication through links can be classified as

1. **Simplex :** Communication can take place only in one direction. eg. T.V broadcasting.
2. **Half-duplex :** Communication can take place in one direction at a time. Suppose node A and B are connected then half-duplex communication means that at a time data can flow from A to B or from B to A but not simultaneously. eg. two persons talking to each other such that when speaks the other listens and vice versa.
3. **Full-duplex :** Communication can take place simultaneously in both directions. eg. A discussion in a group without discipline.

Links can be further classified as

1. **Point to Point :** In this communication only two nodes are connected to each other. When a node sends a packet then it can be recieved only by the node on the other side and none else.
2. **Multipoint** : It is a kind of sharing communication, in which signal can be recieved by all nodes. This is also called broadcast.

Introduction

When a signal is transmitted over a communication channel, it is subjected to different types of impairments because of imperfect characteristics of the channel. As a consequence, the received and the transmitted signals are not the same. Outcome of the impairments are manifested in two different ways in analog and digital signals. These impairments introduce random modifications in analog signals leading to distortion. On the other hand, in case of digital signals, the impairments lead to error in the bit values. The impairment can be broadly categorised into the following three types:

- Attenuation and attenuation distortion
- Delay distortion
- Noise

In this lesson these impairments are discussed in detail and possible approaches to overcome these impairments. The concept of channel capacity for both noise-free and noisy channels have also been introduced.

Attenuation

Irrespective of whether a medium is guided or unguided, the strength of a signal falls off with distance. This is known as attenuation. In case of guided media, the attenuation is logarithmic, whereas in case of unguided media it is a more complex function of the distance and the material that constitutes the medium.

An important concept in the field of data communications is the use of on unit known as **decibel** (dB). To define it let us consider the circuit elements shown in Fig. 2.3.1. The elements can be either a transmission line, an amplifier, an attenuator, a filter, etc. In the figure, a transmission line (between points P1 and P2) is followed by an amplifier (between P2 and P3). The input signal delivers a power P1 at the input of an communication element and the output power is P2. Then the power gain G for this element in decibles is given by G = 10log2 P2/ P1. Here P2/ P1 is referred to as absolute power gain. When P2 > P1, the gain is positive, whereas if P2 < P1, then the power gain is negative and there is a power loss in the circuit element. For P2 = 5mW, P1 = 10mW, the power gain G = 10log 5/10 = 10 × -3 = -3dB is negative and it represents attenuation as a signal passes through the communication element.

Example: Let us consider a transmission line between points 1 and 2 and let the energy strength at point 2 is 1/10 of that of point 1. Then attenuation in dB is 10log10(1/10) = -10 dB. On the other hand, there is an amplifier between points 2 and 3. Let the power is 100 times at point 3 with respect to point 2. Then power gain in dB is 10log10(100/1) = 20 dB, which has a positive sign.

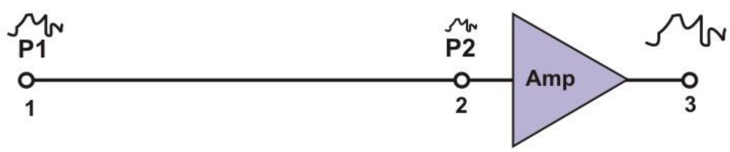


Fig. 1.34 Compensation of attenuation using an amplifier

Noise

As signal is transmitted through a channel, undesired signal in the form of noise gets mixed up with the signal, along with the distortion introduced by the transmission media. Noise can be categorised into the following four types:

- Thermal Noise
- Intermodulation Noise
- Cross talk
- Impulse Noise

The ***thermal noise*** is due to thermal agitation of electrons in a conductor. It is distributed across the entire spectrum and that is why it is also known as *white noise* (as the frequency encompass over a broad range of frequencies).

When more than one signal share a single transmission medium, ***intermodulation noise*** is generated. For example, two signals f1 and f2 will generate signals of frequencies (f1 + f2) and (f1 - f2), which may interfere with the signals of the same frequencies sent by the transmitter. Intermodulation noise is introduced due to nonlinearity present in any part of the communication system.
Cross talk is a result of bunching several conductors together in a single cable. Signal carrying wires generate electromagnetic radiation, which is induced on other conductors because of close proximity of the conductors. While using telephone, it is a common experience to hear conversation of other people in the background. This is known as *cross talk*.
Impulse noise is irregular pulses or noise spikes of short duration generated by phenomena like lightning, spark due to loose contact in electric circuits, etc. Impulse noise is a primary source of bit-errors in digital data communication. This kind of noise introduces burst errors.

Bandwidth

Bandwidth simply means how many bits can be transmitted per second in the communication channel. In technical terms it indicates the width of frequency spectrum. The range of frequencies over which most of the signal energy of a signal is contained is known as **bandwidth** or effective bandwidth of the signal. The term 'most' is somewhat arbitrary. Usually, it is defined in terms of its 3dB cut-off frequency. The frequency spectrum and spectrum of a signal is shown in Fig. 2.1.9. Here the fl and fh may be represented by 3dB below (A/√2) the maximum amplitude A.

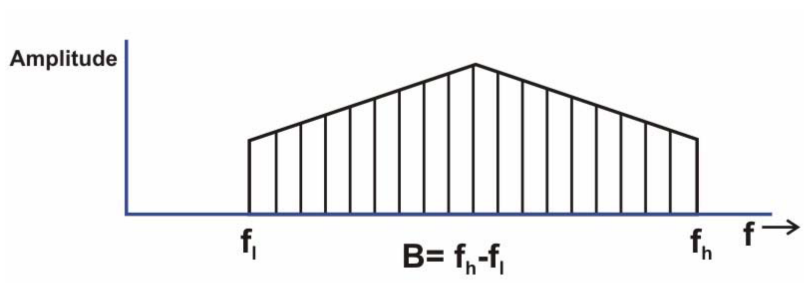


Fig. 1.35 Frequency spectrum and bandwidth of a signal

Bandwidth of a medium decides the quality of the signal at the other end. A digital signal (usually aperiodic) requires a bandwidth from 0 to infinity. So, it needs a low-pass channel characteristic as shown in Fig. 2.3.5. On the other hand, a band-pass channel characteristic is required for the transmission of analog signals, as shown in Fig. 2.3.6.

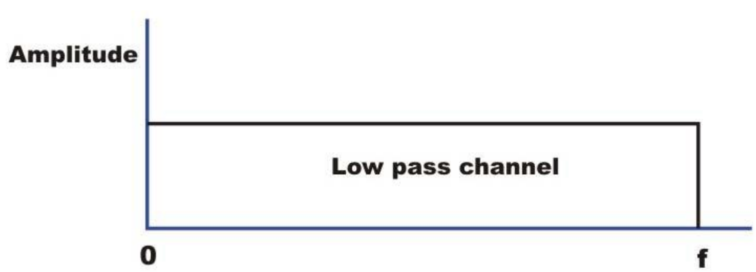


Fig. 1.36 Low-pass channel characteristic required for the transmission of digital signals

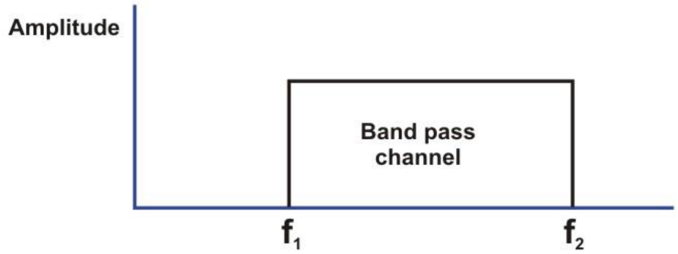


Fig. 1.37 Band-pass channel characteristic required for the transmission of analog signals

Nyquist Bit Rate

The maximum rate at which data can be correctly communicated over a channel in presence of noise and distortion is known as its channel capacity. Consider first a noise-free channel of Bandwidth B. Based on Nyquist formulation it is known that given a bandwidth B of a channel, the maximum data rate that can be carried is 2B. This limitation arises due to the effect of intersymbol interference caused by the frequency components higher than B. If the signal consists of m discrete levels, then Nyquist theorem states:

Maximum data rate $C = 2 B \log_2 m$ bits/sec,

where C is known as the channel capacity,
B is the bandwidth of the channel
and m is the number of signal levels used.

Baud Rate: The baud rate or signaling rate is defined as the number of distinct symbols transmitted per second, irrespective of the form of encoding. For baseband digital transmission m = 2. So, the maximum baud rate = 1/Element width (in Seconds) = 2B

Bit Rate: The bit rate or information rate I is the actual equivalent number of bits transmitted per second. I = Baud Rate × Bits per Baud
= Baud Rate × N = Baud Rate × $\log_2 m$

For binary encoding, the bit rate and the baud rate are the same; i.e., I = Baud Rate.

Example: Let us consider the telephone channel having bandwidth B = 4 kHz. Assuming there is no noise, determine channel capacity for the following encoding levels:
(i) 2, and (ii) 128.

Ans: (i) C = 2B = 2×4000 = 8 Kbits/s
(ii) C = 2×4000×$\log_2 128$ = 8000×7 = 56 Kbits/s

Effects of Noise

When there is noise present in the medium, the limitations of both bandwidth and noise must be considered. A noise spike may cause a given level to be interpreted as a signal of greater level, if it is in positive phase or a smaller level, if it is negative phase. Noise becomes more problematic as the number of levels increases.

Shannon Capacity (Noisy Channel)

In presence of Gaussian band-limited white noise, Shannon-Hartley theorem gives the maximum data rate capacity

$C = B \log_2 (1 + S/N)$,

where S and N are the signal and noise power, respectively, at the output of the channel. This theorem gives an upper bound of the data rate which can be reliably transmitted over a thermal-noise limited channel.

Example: Suppose we have a channel of 3000 Hz bandwidth, we need an S/N ratio (i.e. signal to noise ration, SNR) of 30 dB to have an acceptable bit-error rate. Then, the maximum data rate that we can transmit is 30,000 bps. In practice, because of the presence of different types of noises, attenuation and delay distortions, actual (practical) upper limit will be much lower.
In case of extremely noisy channel, C = 0

Between the Nyquist Bit Rate and the Shannon limit, the result providing the smallest channel capacity is the one that establishes the limit.

Example: A channel has *B* = 4 KHz. Determine the channel capacity for each of the following signal-to-noise ratios: (a) 20 dB, (b) 30 dB, (c) 40 dB.

Answer: (a) C= B log2 (1 + S/N) = 4×103×log2 (1+100) = 4×103×3.32×2.004 = 26.6 kbits/s

(b) C= B log2 (1 + S/N) = 4×103×log2 (1+1000) = 4×103×3.32×3.0 = 39.8 kbits/s

(c) C= B log2 (1 + S/N) = 4×103×log2 (1+10000) = 4×103×3.32×4.0 = 53.1 kbits/s

Example: A channel has B = 4 KHz and a signal-to-noise ratio of 30 dB. Determine maximum information rate for 4-level encoding.

Answer: For B = 4 KHz and 4-level encoding the Nyquist Bit Rate is 16 Kbps. Again for B = 4 KHz and S/N of 30 dB the Shannon capacity is 39.8 Kbps. The smallest of the two values has to be taken as the Information capacity I = 16 Kbps.

Example: A channel has B = 4 kHz and a signal-to-noise ratio of 30 dB. Determine maximum information rate for 128-level encoding.

Answer: The Nyquist Bit Rate for B = 4 kHz and M = 128 levels is 56 kbits/s. Again the Shannon capacity for B = 4 kHz and S/N of 30 dB is 39.8 Kbps. The smallest of the two values decides the channel capacity C = 39.8 kbps.

Example: The digital signal is to be designed to permit 160 kbps for a bandwidth of 20 KHz. Determine (a) number of levels and (b) S/N ratio.

(a) Apply Nyquist Bit Rate to determine number of levels.

C = 2B log2 (M),

or 160×103 = 2×20×103 log2 (M),

or M = 24, which means 4bits/baud.

(b) Apply Shannon capacity to determine the S/N ratio

C = B log2 (1+S/N),

or 160×103 = 20×103 log2 (1+S/N) ×103 log2 (M) ,

or S/N = 28 - 1,

or S/N = 255,

or S/N = 24.07 dB.

Q-1. Distinguish between attenuation distortion and delay distortion.
Ans: Attenuation distortion arises because the attenuation of the signal in the transmitting media. Attenuation distortion is predominant in case of analog signals. Delay distortion arises because different frequency components of the signal suffer different delay as the signal passes through the media. This happens because the velocity of the signal varies with frequency and it is predominant in case of digital signals.
Q-2. How the effect of delay distortion can be minimized?
Ans: Delay distortion can be minimized by using an equalizer (a kind of filter).
Q-3. What is intermodulation noise?
Ans: When a signal (having different frequency components) passes through a transmitting media, then due to non-linearity, some of the frequency components may combine to generate a different frequency component. This leads to distortion in the signal, which is known as intermodulation noise. For example, a signal may be having frequency components f1 and f2, and due to non-linearity of the media they may generate a frequency component (f1+f2). Further a frequency of (f1+f2) may be already present in the original signal. This causes intermodulation noise.
Q-4. Why does impulse noise have more effect on digital signals rather than on analog signals?
Ans: Impulse noise is random in nature and arises due to random events like lightning, electrical sparks, etc. In case of digital signal, it makes a significant effect, as '0' may become '1' and vice versa. In analog signal the effect is not that serious as some portion of the signal gets affected.
Q-5. What is crosstalk?
Ans: Crosstalk refers to the picking up of electromagnetic signals from other adjacent wires by electromagnetic induction.
Q-6. Let the energy strength at point 2 is 1/50th with respect to the point 1. Find out the attenuation in dB.
Ans: Then attenuation in dB is 10log10(1/50) = - 16.9 dB.

Multiplexing of Signals