DATA COMMUNICATION AND COMPUTER NETWORK

user

Now a days the growth of data communication technology has become very fast in development of various application areas. This subject will expose the learner to have an idea about the architecture computer network and different protocols to be followed to communicate. Further they will have an idea about different mode of communication.

HP

[Type the company address]

[Type the phone number]

[Type the fax number]

[Pick the date]

Unit-1. Network& Protocol

Data communications

When we communicate, we are sharing information. This sharing can be local or remote. Between individuals, local communication usually occurs face to face, while remote communication takes place over distance.

The term telecommunication, which includes telephony, telegraphy, and television, means communication at a distance (tele is Greek for "far").

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

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

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

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

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

4. Jitter. Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, let us assume that video packets are sent every 3D ms. If some of the packets arrive with 3D-ms delay and others with 4D-ms delay, an uneven quality in the video is the result

Components

A data communications system has five components

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

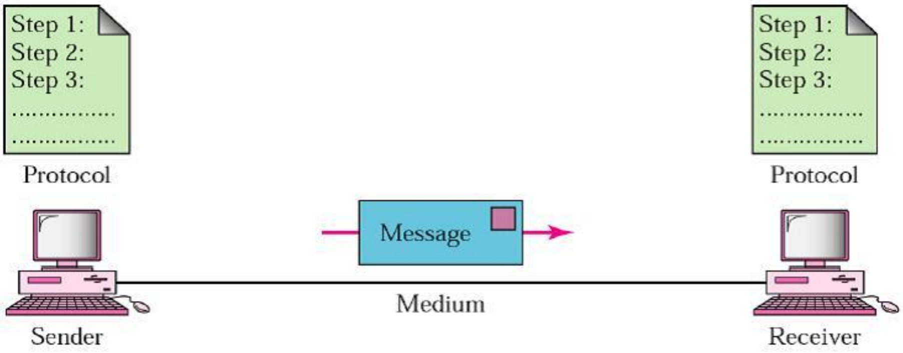
2. Sender. The sender is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.

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

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

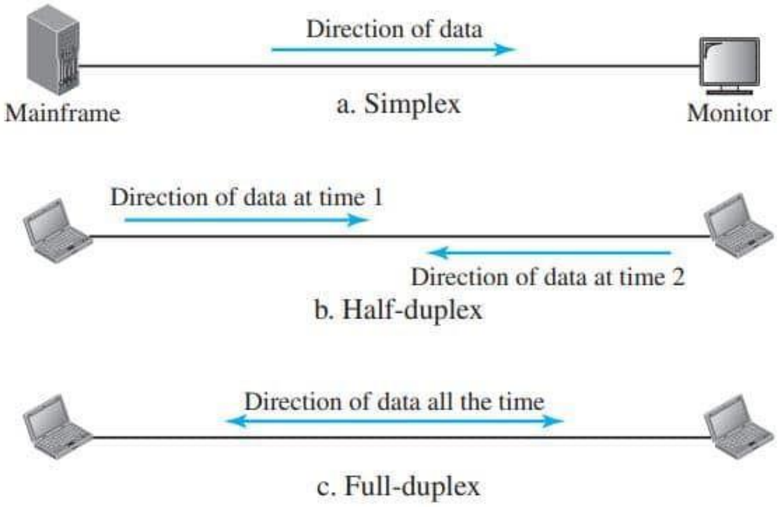
5. Protocol. A protocol is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not communicating, just as a person speaking French cannot be understood by a person who speaks only Japanese.

(Data Representation Information today comes in different forms such as text, numbers, images, audio, and video)



Data Flow

Communication between two devices can be simplex, half-duplex, or full-duplex


Direction of data
Mainframe
a. Simplex
Monitor
Direction of data at time 1
Direction of data at time 2
b. Half-duplex
Direction of data all the time
c. Full-duplex

Simplex

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 . Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output. The simplex mode can use the entire capacity of the channel to send data in one direction.

Half-Duplex

In half-duplex mode, each station can both transmit and receive, but not at the same time. : When one device is sending, the other can only receive, and vice versa.

The half-duplex mode is like a one-lane road with traffic allowed in both directions. When cars are traveling in one direction, cars going the other way must wait. In a half-duplex transmission, the entire capacity of a channel is taken over by whichever of the two devices is transmitting at the time. Walkie-talkies and CB (citizens band) radios are both half-duplex systems.

Full-Duplex

In full-duplex (also called duplex), both stations can transmit and receive simultaneously . The full-duplex mode is like a two way street with traffic flowing in both directions at the same time.

In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other direction.

This sharing can occur in two ways: Either the link must contain two physically separate transmission paths, one for sending and the other for receiving; or the capacity of the channel is divided between signals traveling in both directions.

One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time.

NETWORK

A network is a set of devices (often referred to as nodes) connected by communication links.

A node can be a computer, printer, or any other device capable of sending and/or receiving data generated by other nodes on the network.

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.

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.

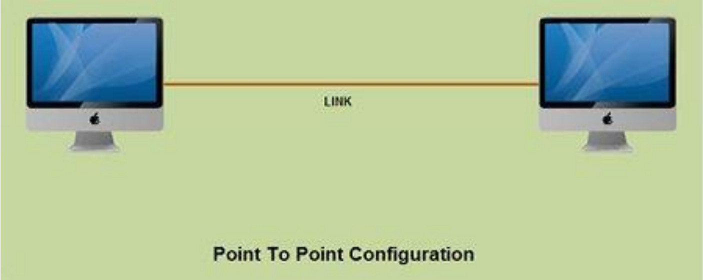
Physical Structures

Type of Connection

There are two possible types of connections: point-to-point and multipoint.

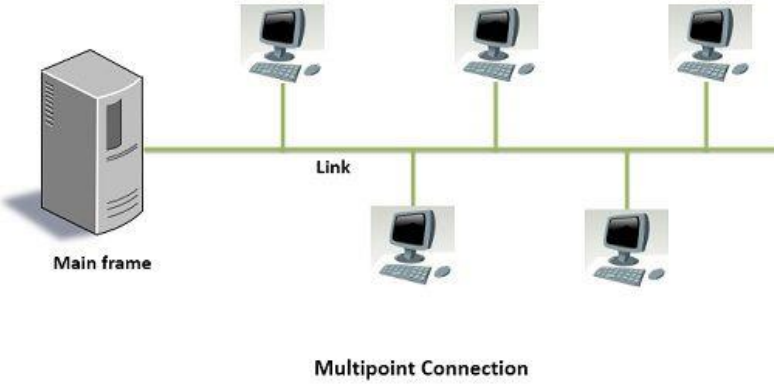
Point-to-Point

- A point-to-point connection provides a dedicated link between two devices.
- The entire capacity of the link is reserved for transmission between those two devices.
- Most point-to-point connections use an actual length of wire or cable to connect the two ends, but other options, such as microwave or satellite links, are also possible



Multipoint

- A multipoint (also called multidrop) connection is one in which more than two specific devices share a single link .
- In a multipoint environment, the capacity of the channel is shared, either spatially or temporally.
- If several devices can use the link simultaneously, it is a spatially shared connection. If users must take turns, it is a timeshared connection.



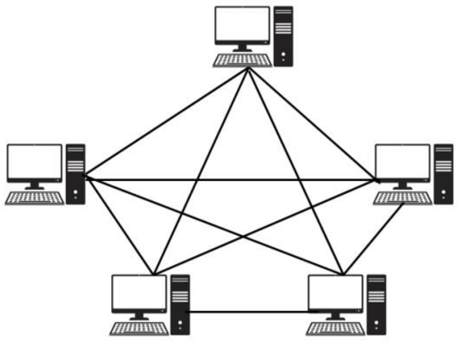
Physical Topology

- The term physical topology refers to the way in which a network is laid out physically.:
- Two or more devices connect to a link; two or more links form a topology.
- The topology of a network is the geometric representation of the relationship of all the links and linking devices (usually called nodes) to one another.
- There are four basic topologies possible: **mesh, star, bus, and ring**

Mesh

- In a mesh topology, every device has a dedicated point-to-point link to every other device.
- The term dedicated means that the link carries traffic only between the two devices it connects.
- To find the number of physical links in a fully connected mesh network with n nodes, we first consider that each node must be connected to every other node. Node 1 must be connected to n - I nodes, node 2 must be connected to n - 1 nodes, and finally node n must be connected to n - 1 nodes. We need n(n - 1) physical links.
- However, if each physical link allows communication in both directions (duplex mode), we can divide the number of links by 2. In other words, we can say that in a mesh topology, we need n(n -1) /2 duplex-mode links.

Mesh Topology



Advantages

- The use of dedicated links guarantees that each connection can carry its own data load
- Mesh topology is robust means if one link becomes unusable, it does not incapacitate the entire system.
- Privacy or security

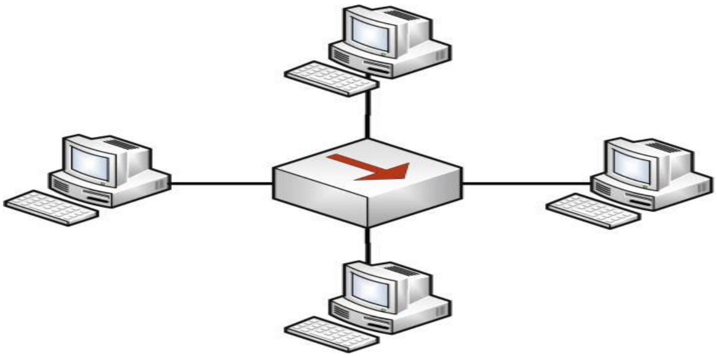
Disadvantages

- The main disadvantages of a mesh are related to the amount of cabling and the number of I/O ports required.

One practical example of a mesh topology is the connection of telephone regional offices in which each regional office needs to be connected to every other regional office.

Star Topology

- In a star topology, each device has a dedicated point-to-point link only to a central controller, usually called a hub.
- The devices are not directly linked to one another.
- Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data to the controller, which then relays the data to the other connected device.



Advantages

- A star topology is less expensive than a mesh topology.
- In a star, each device needs only one link and one I/O port to connect it to any number of others which makes it easy to install and reconfigure.
- Far less cabling needs to be housed, and additions, moves, and deletions involve only one connection: between that device and the hub.
- Star topology is also robust.

The star topology is used in local-area networks (LANs)

Bus Topology

A bus topology is multipoint. One long cable acts as a backbone to link all the devices in a network.

- Nodes are connected to the bus cable by drop lines and taps.
- A drop line is a connection running between the device and the main cable.
- A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core.
- As a signal travels along the backbone, some of its energy is transformed into heat. Therefore, it becomes weaker and weaker as it travels farther and farther. For this reason there is a limit on the number of taps a bus can support and on the distance between those taps.

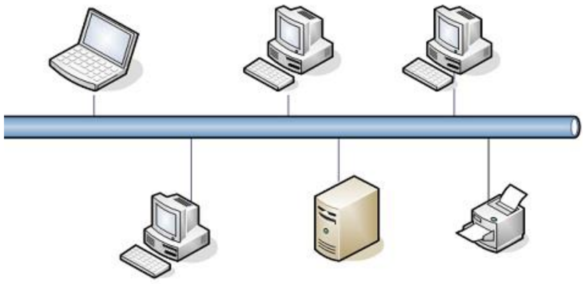
Advantages of BUS Topology

- Ease of installation.
- It uses less cabling than mesh or star topologies.

Disadvantages of BUS Topology

- Disadvantages include difficult reconnection and fault isolation. A bus is usually designed to be optimally efficient at installation. It can therefore be difficult to add new devices. Signal reflection at the taps can cause degradation in quality. This degradation can be controlled by limiting the number and spacing of devices connected to a given length of cable. Adding new devices may therefore require modification or replacement of the backbone.
- A fault or break in the bus cable stops all transmission

Bus topology was the one of the first topologies used in the design of early localarea networks



Ring Topology

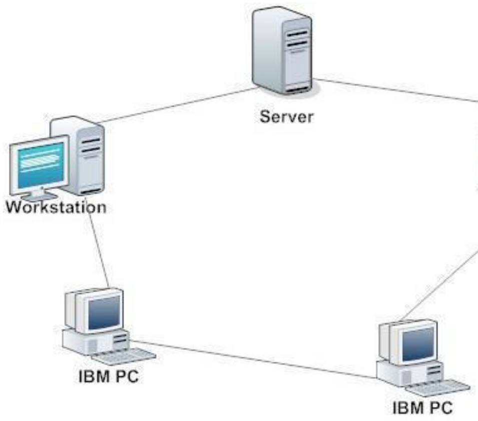
- In a ring topology, each device has a dedicated point-to-point connection with only the two devices on either side of it.
- A signal is passed along the ring in one direction, from device to device, until it reaches its destination.
- Each device in the ring incorporates a repeater.
- When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along

Advantages of Ring Topology

- A ring is relatively easy to install and reconfigure.
- Each device is linked to only its immediate neighbors (either physically or logically).
- To add or delete a device requires changing only two connections.
- Fault isolation is simplified. Generally in a ring, a signal is circulating at all times. If one device does not receive a signal within a specified period, it can issue an alarm. The alarm alerts the network operator to the problem and its location.

Disadvantages of Ring Topology

However, unidirectional traffic can be a disadvantage. In a simple ring, a break in the ring (such as a disabled station) can disable the entire network. This weakness can be solved by using a dual ring or a switch capable of closing off the break.

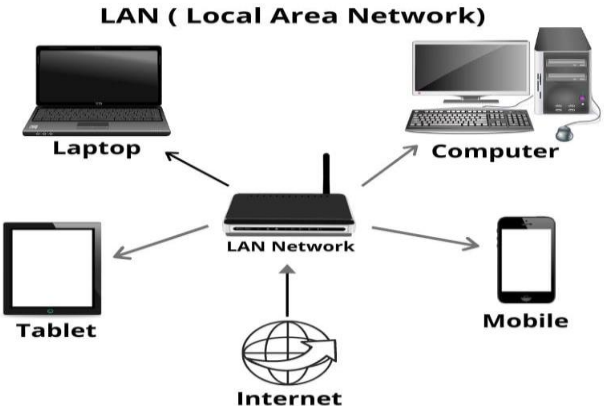


Categories of Networks

1. Local Area Network
2. Wide Area Network
3. Metropolitan Area Networks

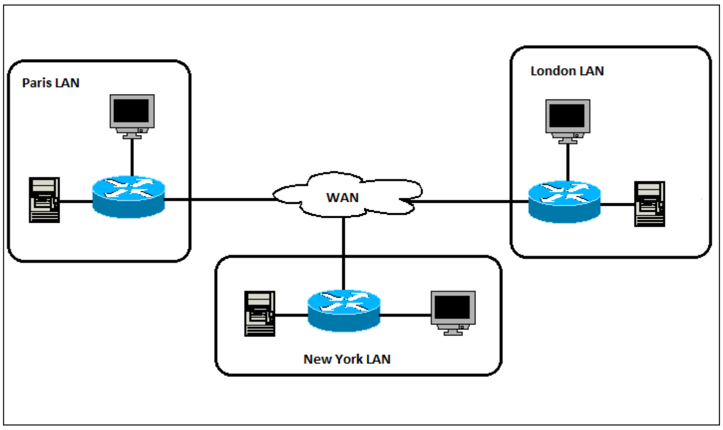
Local Area Network

- A local area network (LAN) is usually privately owned and links the devices in a single office, building, or campus.
- Currently, LAN size is limited to a few kilometers.
- LANs are designed to allow resources to be shared between personal computers or workstations.
- The resources to be shared can include hardware (e.g., a printer), software (e.g., an application program), or data.
- In addition to size, LANs are distinguished from other types of networks by their transmission media and topology.
- In general, a given LAN will use only one type of transmission medium.
- The most common LAN topologies are bus, ring, and star.
- Early LANs had data rates in the 4 to 16 megabits per second (Mbps) range. Today, however, speeds are normally 100 or 1000 Mbps.



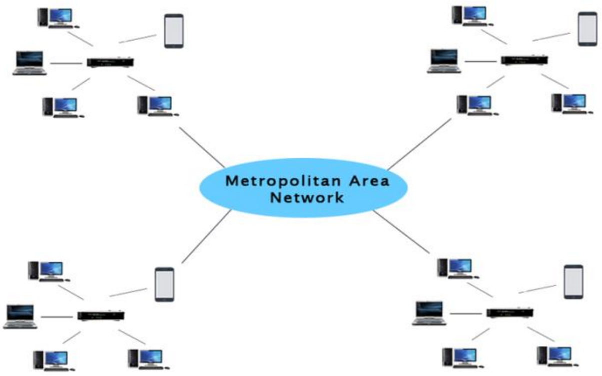
Wide Area Network

- A wide area network (WAN) provides long-distance transmission of data, image, audio, and video information over large geographic areas that may comprise a country, a continent, or even the whole world



Metropolitan Area Networks

- A metropolitan area network (MAN) is a network with a size between a LAN and a WAN. It normally covers the area inside a town or a city. It is designed for customers who need a high-speed connectivity
- A good example of a MAN is the part of the telephone company network that can provide a high-speed DSL line to the customer. Another example is the cable TV network that originally was designed for cable TV, but today can also be used for high-speed data connection to the Internet.



THE INTERNET

- An internet (note the lowercase letter i) is two or more networks that can communicate with each other.
- The Internet is a communication system that has brought a wealth of information to our fingertips and organized it for our use.
- The Internet is a structured, organized system.
- The most notable internet is called the Internet - a collaboration of more than hundreds of thousands of interconnected networks.
- Private individuals as well as various organizations such as government agencies, schools, research facilities, corporations, and libraries in more than 100 countries use the Internet.
- The Internet Today
- 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 (lSPs).
- There are international service providers, national service providers, regional service providers, and local service providers.

PROTOCOLS AND STANDARDS

- A protocol is a set of rules that govern data communications.
- A protocol defines what is communicated, how it is communicated, and when it is communicated.
- The key elements of a protocol are syntax, semantics, and timing.

Syntax

- The term syntax refers to the structure or format of the data, meaning the order in which they are presented. For example, a simple protocol might expect the first 8 bits of data to be the address of the sender, the second 8 bits to be the address of the receiver, and the rest of the stream to be the message itself.

Semantics

- The word semantics refers to the meaning of each section of bits. How is a particular pattern to be interpreted, and what action is to be taken based on that interpretation? For example, does an address identify the route to be taken or the final destination of the message?

Timing

- The term timing refers to two characteristics: when data should be sent and how fast they can be sent. For example, if a sender produces data at 100 Mbps but the receiver can process data at only 1 Mbps, the transmission will overload the receiver and some data will be lost.

STANDARDS

- Standards provide guidelines to manufacturers, vendors, government agencies, and other service providers to ensure the kind of interconnectivity necessary in today's marketplace and in international communications. Data communication standards fall into two categories: de facto (meaning "by fact" or "by convention") and de jure (meaning "by law" or "by regulation").

De facto.

- Standards that have not been approved by an organized body but have been adopted as standards through widespread use are de facto standards.

- De facto standards are often established originally by manufacturers who seek to define the functionality of a new product or technology.

De jure

- Those standards that have been legislated by an officially recognized body are de jure standards.
- Standards Creation Committees
- While many organizations are dedicated to the establishment of standards, data telecommunications in North America rely primarily on those published by the following:

International Organization for Standardization (ISO).

- The ISO is a multinational body whose membership is drawn mainly from the standards creation committees of various governments throughout the world. The ISO is active in developing cooperation in the realms of scientific, technological, and economic activity.
- International Telecommunication Union-Telecommunication Standards Sector (ITU-T).
- By the early 1970s, a number of countries were defining national standards for telecommunications, but there was still little international compatibility. The United Nations responded by forming, as part of its International Telecommunication Union (ITU), a committee, the Consultative Committee for International Telegraphy and Telephony (CCITT). This committee was devoted to the research and establishment of standards for telecommunications in general and for phone and data systems in particular. On March 1, 1993, the name of this committee was changed to the International Telecommunication

UnionTelecommunication Standards Sector (ITU-T).

American National Standards Institute (ANSI).

- Despite its name, the American National Standards Institute is a completely private, nonprofit corporation not affiliated with the U.S. federal government. However, all ANSI activities are undertaken with the welfare of the United States and its citizens occupying primary importance.

Institute of Electrical and Electronics Engineers (IEEE).

- The Institute of Electrical and Electronics Engineers is the largest professional engineering society in the world. International in scope, it aims to advance theory, creativity, and product quality in the fields of electrical engineering, electronics, and radio as well as in all related branches of engineering. As one of its goals, the IEEE oversees the development and adoption of international standards for computing and communications.

Electronic Industries Association (EIA).

- Aligned with ANSI, the Electronic Industries Association is a nonprofit organization devoted to the promotion of electronics manufacturing concerns. Its activities include public awareness education and lobbying efforts in addition to standards development. In the field of information technology, the EIA has made significant contributions by defining physical connection interfaces and electronic signaling specifications for data communication.

SUMMARY

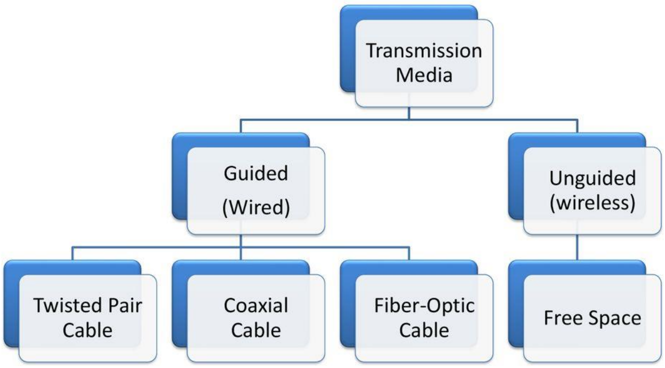
- Data communications are the transfer of data from one device to another via some form of transmission medium.
- A data communications system must transmit data to the correct destination in an accurate and timely manner.
- The five components that make up a data communications system are the message, sender, receiver, medium, and protocol.
- Text, numbers, images, audio, and video are different forms of information.
- Data flow between two devices can occur in one of three ways: simplex, half-duplex, or full-duplex.
- A network is a set of communication devices connected by media links.
- In a point-to-point connection, two and only two devices are connected by a dedicated link.
- In a multipoint connection, three or more devices share a link.
- Topology refers to the physical or logical arrangement of a network.
- Devices may be arranged in a mesh, star, bus, or ring topology.
- A network can be categorized as a local area network or a wide area network.
- A LAN is a data communication system within a building, plant, or campus, or between nearby buildings.
- A WAN is a data communication system spanning states, countries, or the whole world.
- An internet is a network of networks.
- The Internet is a collection of many separate networks.
- There are local, regional, national, and international Internet service providers.
- A protocol is a set of rules that govern data communication; the key elements of a protocol are syntax, semantics, and timing.
- Standards are necessary to ensure that products from different manufacturers can work together as expected.
- The ISO, ITD-T, ANSI, IEEE, and EIA are some of the organizations involved in standards creation.

Unit-2. Data Transmission & Media

TRANSMISSION MEDIUM

- A transmission medium can be broadly defined as anything that can carry information from a source to a destination.
- The transmission medium is usually free space, metallic cable, or fiber-optic cable.
- In telecommunications, transmission media can be divided into two broad categories: **guided and unguided.**
- **Guided media** include twisted-pair cable, coaxial cable, and fiber-optic cable.
- **Unguided medium** is free space.

Classification of Transmission Media



GUIDED MEDIA

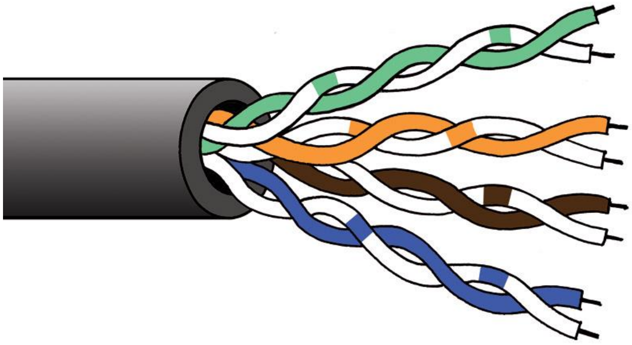
- Guided media, which are those that provide a conduit from one device to another, include twisted-pair cable, coaxial cable, and fiber-optic cable.
- A signal traveling along any of these media is directed and contained by the physical limits of the medium.
- Twisted-pair and coaxial cable use metallic (copper) conductors that accept and transport signals in the form of electric current.
- Optical fiber is a cable that accepts and transports signals in the form of light.

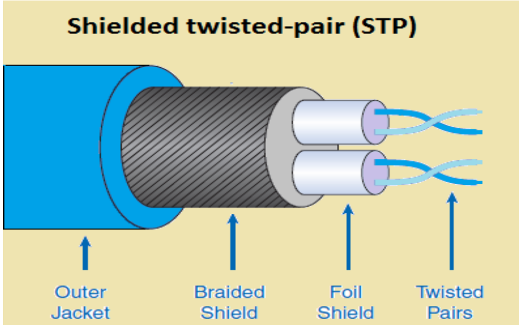
1. Twisted-Pair Cable

- A twisted pair consists of two conductors (normally copper), each with its own plastic insulation, twisted together.
- One of the wires is used to carry signals to the receiver, and the other is used only as a ground reference. The receiver uses the difference between the two.
- Twisting makes it probable that both wires are equally affected by external influences (noise or crosstalk). This means that the receiver, which calculates the difference between the two, receives no unwanted signals. The unwanted signals are mostly cancelled out.

Types of Twisted pair cable

1. Unshielded Twisted-Pair Cable. (UTP)
2. Shielded Twisted-Pair Cable. (STP)

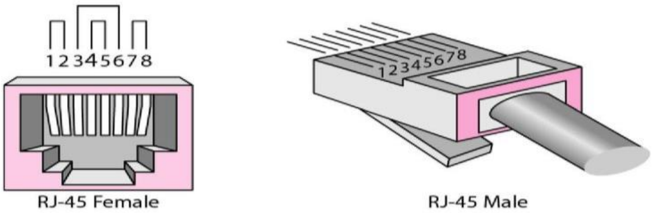




- The most common twisted-pair cable used in communications is referred to as unshielded twisted-pair (UTP)
- STP cable has a metal foil or braided mesh covering that encases each pair of insulated conductors. Although metal casing improves the quality of cable by preventing the penetration of noise or crosstalk, it is bulkier and more expensive.

Connectors

The most common UTP connector is RJ45 (RJ stands for registered jack). The RJ45 is a keyed connector, meaning the connector can be inserted in only one way.



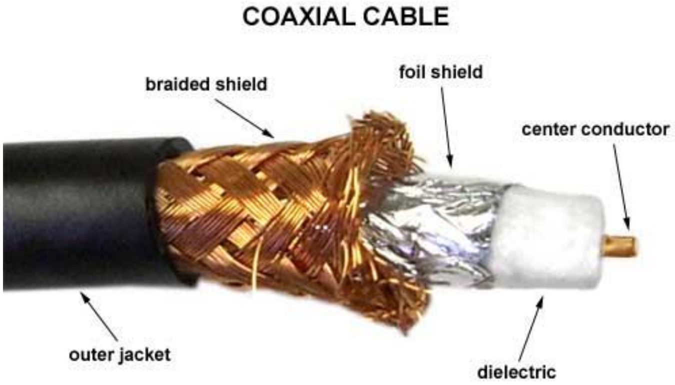
Applications

Twisted-pair cables are used in telephone lines to provide voice and data channels.

2. **<u>Coaxial Cable</u>**

- Coaxial cable (or coax) carries signals of higher frequency ranges than those in twistedpair 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.



Coaxial Cable Connectors

To connect coaxial cable to devices, we need coaxial connectors. The most common type of connector used today is the Bayone-Neill-Concelman (BNe), connector.

There are three popular types of these connectors: the BNC connector, the BNC T connector, and the BNC terminator.

Applications

Cable TV networks use coaxial cables.

Coaxial cable was widely used in analog telephone networks where a single coaxial network could carry 10,000 voice signals.

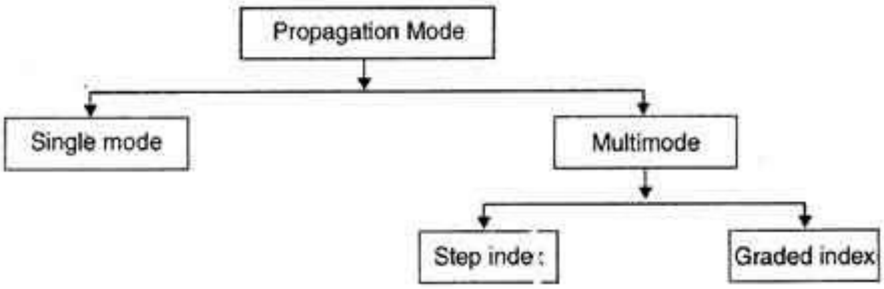
Another common application of coaxial cable is in traditional Ethernet LANs.

3. **Fiber-Optic Cable**
   - A fiber-optic cable is made of glass or plastic and transmits signals in the form of light.

- Light travels in a straight line as long as it is moving through a single uniform substance. If a ray oflight traveling through one substance suddenly enters another substance (of a different density), the ray changes direction.
- As the figure shows, if the angle of incidence I (the arIgle the ray makes with the line perpendicular to the interface between the two substances) is less than the critical angle, the ray refracts and moves closer to the surface.
- If the angle of incidence is equal to the critical angle, the light bends along the interface.
- If the angle is greater than the critical angle, the ray reflects (makes a turn) and travels again in the denser substance. Note that the critical angle is a property of the substance, and its value differs from one substance to another.
- Optical fibers use reflection to guide light through a channel. A glass or plastic core is surrounded by a cladding of less dense glass or plastic. The difference in density of the two materials must be such that a beam of light moving through the core is reflected off the cladding instead of being refracted into it.

Propagation Modes

Current technology supports two modes (multimode and single mode) for propagating light along optical channels, each requiring fiber with different physical characteristics. Multimode can be implemented in two forms: step-index or graded-index.



Multimode

- Multimode is so named because multiple beams from a light source move through the core in different paths.
- How these beams move within the cable depends on the structure of the core.

Multimode step-index fiber

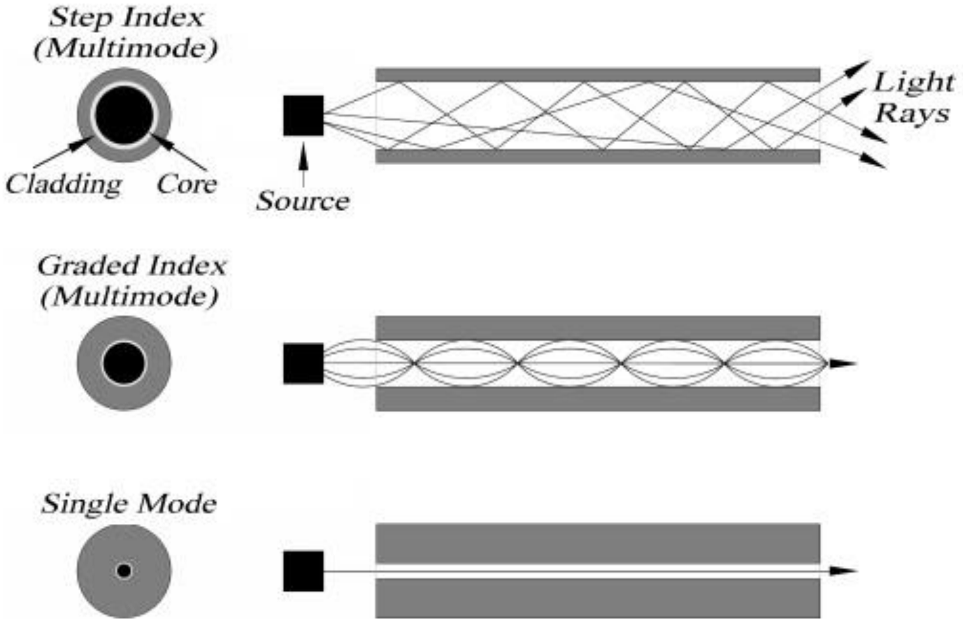
- **In multimode step-index fiber** the density of the core remains constant from the center to the edges.
- A beam of light moves through this constant density in a straight line until it reaches the interface of the core and the cladding.
- At the interface, there is an abrupt change due to a lower density; this alters the angle of the beam's motion.
- The term step index refers to the suddenness of this change, which contributes to the distortion of the signal as it passes through the fiber.

Multimode graded-index fiber

- This fiber decreases the distortion of the signal through the cable.
- The word index here refers to the index of refraction.
- A graded-index fiber, therefore, is one with varying densities.
- Density is highest at the center of the core and decreases gradually to its lowest at the edge.

Single-Mode

- Single-mode uses step-index fiber and a highly focused source of light that limits beams to a small range of angles, all close to the horizontal.
- The single mode fiber itself is manufactured with a much smaller diameter than that of multimode fiber, and with substantially lower density (index of refraction).
- The decrease in density results in a critical angle that is close enough to 90° to make the propagation of beams almost horizontal.
- In this case, propagation of different beams is almost identical, and delays are negligible. All the beams arrive at the destination "together" and can be recombined with little distortion to the signal.



Fiber Sizes:

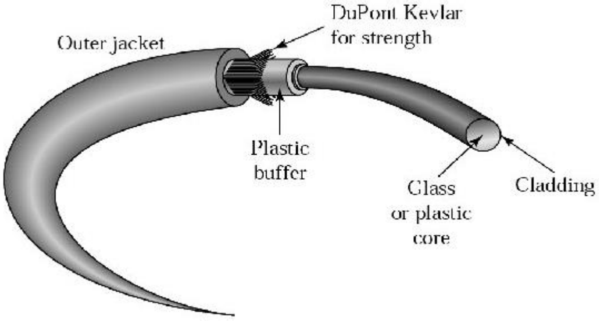
Optical fibers are defined by the ratio of the diameter of their core to the diameter of their cladding, both expressed in micrometers.

Cable Composition:

The figure below shows the composition of a typical fiber-optic cable.

- The outer jacket is made of either PVC or Teflon.

- Inside the jacket are Kevlar strands to strengthen the cable.
- Kevlar is a strong material used in the fabrication of bulletproof vests.
- Below the Kevlar is another plastic coating to cushion the fiber.
- The fiber is at the center of the cable, and it consists of cladding and core.

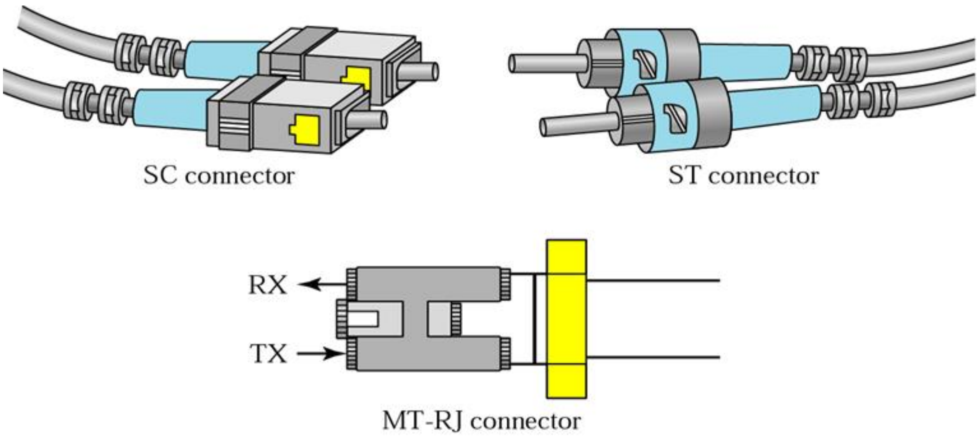


Fiber-Optic Cable Connectors:

There are three types ofconnectors for fiber-optic cables

1. SC Connector
2. ST Connector
3. MT-RJ

- The subscriber channel (SC) connector is used for cable TV.
- It uses a push/pull locking system.
- The straight-tip (ST) connector is used for connecting cable to networking devices.
- It uses a bayonet locking system and is more reliable than SC.
- MT-RJ is a connector that is the same size as RJ45.



Applications

- Fiber-optic cable is often found in backbone networks because its wide bandwidth is cost-effective.
- Some cable TV companies use a combination of optical fiber and coaxial cable, thus creating a hybrid network.
- Local-area networks such as 100Base-FX network (Fast Ethernet) and 1000Base-X also use fiber-optic cable.

Advantages and Disadvantages of Optical Fiber

Advantages:

- **Higher bandwidth**: Fiber-optic cable can support dramatically higher bandwidths (and hence data rates) than either twisted-pair or coaxial cable.
- **Less signal attenuation**: Fiber-optic transmission distance is significantly greater than that of other guided media. A signal can run for 50 km without requiring regeneration. We need repeaters every 5 km for coaxial or twisted-pair cable.
- **Immunity to electromagnetic interference**: Electromagnetic noise cannot affect fiber-optic cables.
- **Resistance to corrosive materials**: Glass is more resistant to corrosive materials than copper.
- **Light weight**: Fiber-optic cables are much lighter than copper cables.
- **Greater immunity to tapping:** Fiber-optic cables are more immune to tapping than copper cables.

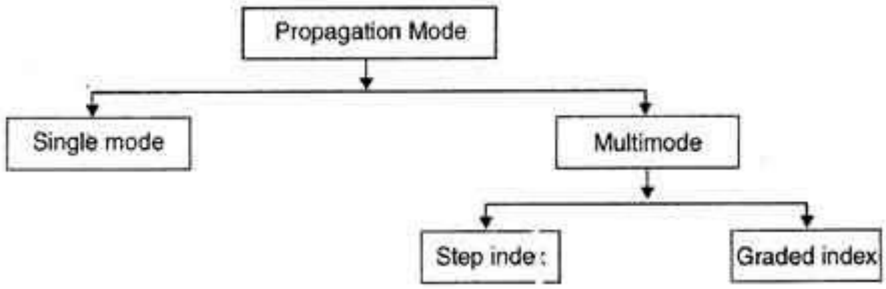
Disadvantages:

There are some disadvantages in the use of optical fiber.

- **Installation and maintenance:** Fiber-optic cable is a relatively new technology. Its installation and maintenance require expertise that is not yet available everywhere.
- **Unidirectional light propagation**: Propagation of light is unidirectional. If we need bidirectional communication, two fibers are needed.
- **Cost**: The cable and the interfaces are relatively more expensive than those of other guided media.

Propagation Modes

Current technology supports two modes (multimode and single mode) for propagating light along optical channels, each requiring fiber with different physical characteristics. Multimode can be implemented in two forms: step-index or graded-index.



Multimode

- Multimode is so named because multiple beams from a light source move through the core in different paths.
- How these beams move within the cable depends on the structure of the core.

Multimode step-index fiber

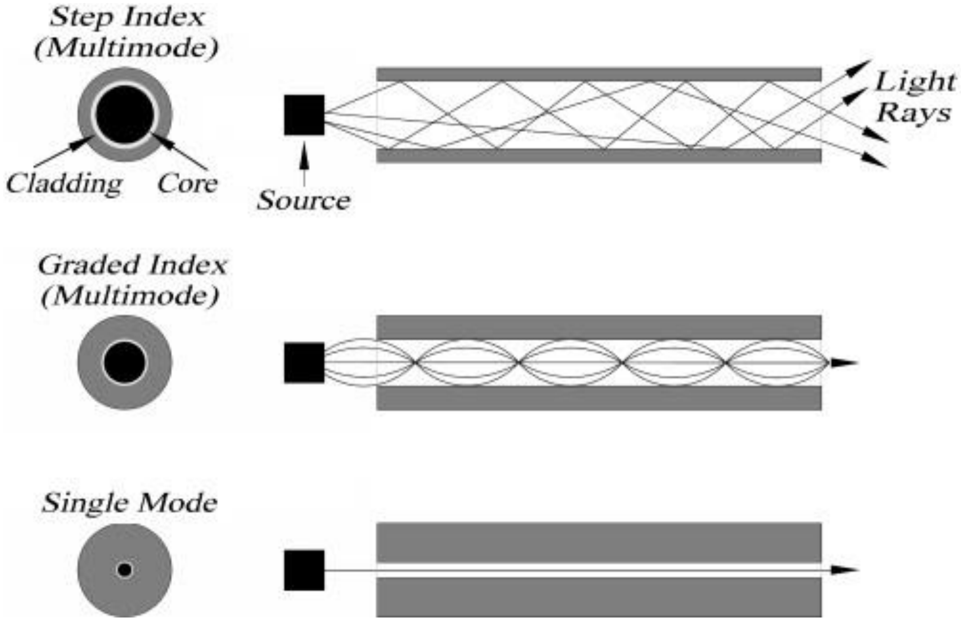
- **In multimode step-index fiber** the density of the core remains constant from the center to the edges.
- A beam of light moves through this constant density in a straight line until it reaches the interface of the core and the cladding.
- At the interface, there is an abrupt change due to a lower density; this alters the angle of the beam's motion.
- The term step index refers to the suddenness of this change, which contributes to the distortion of the signal as it passes through the fiber.

Multimode graded-index fiber

- This fiber decreases the distortion of the signal through the cable.
- The word index here refers to the index of refraction.
- A graded-index fiber, therefore, is one with varying densities.
- Density is highest at the center of the core and decreases gradually to its lowest at the edge.

Single-Mode

- Single-mode uses step-index fiber and a highly focused source of light that limits beams to a small range of angles, all close to the horizontal.
- The single mode fiber itself is manufactured with a much smaller diameter than that of multimode fiber, and with substantially lower density (index of refraction).
- The decrease in density results in a critical angle that is close enough to 90° to make the propagation of beams almost horizontal.
- In this case, propagation of different beams is almost identical, and delays are negligible. All the beams arrive at the destination "together" and can be recombined with little distortion to the signal.



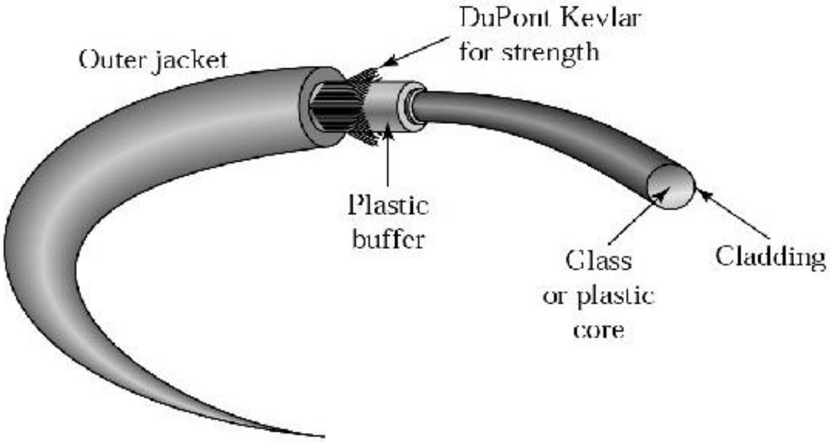
Fiber Sizes:

Optical fibers are defined by the ratio of the diameter of their core to the diameter of their cladding, both expressed in micrometers.

Cable Composition:

The figure below shows the composition of a typical fiber-optic cable.

- The outer jacket is made of either PVC or Teflon.
- Inside the jacket are Kevlar strands to strengthen the cable.
- Kevlar is a strong material used in the fabrication of bulletproof vests.
- Below the Kevlar is another plastic coating to cushion the fiber.
- The fiber is at the center of the cable, and it consists of cladding and core.

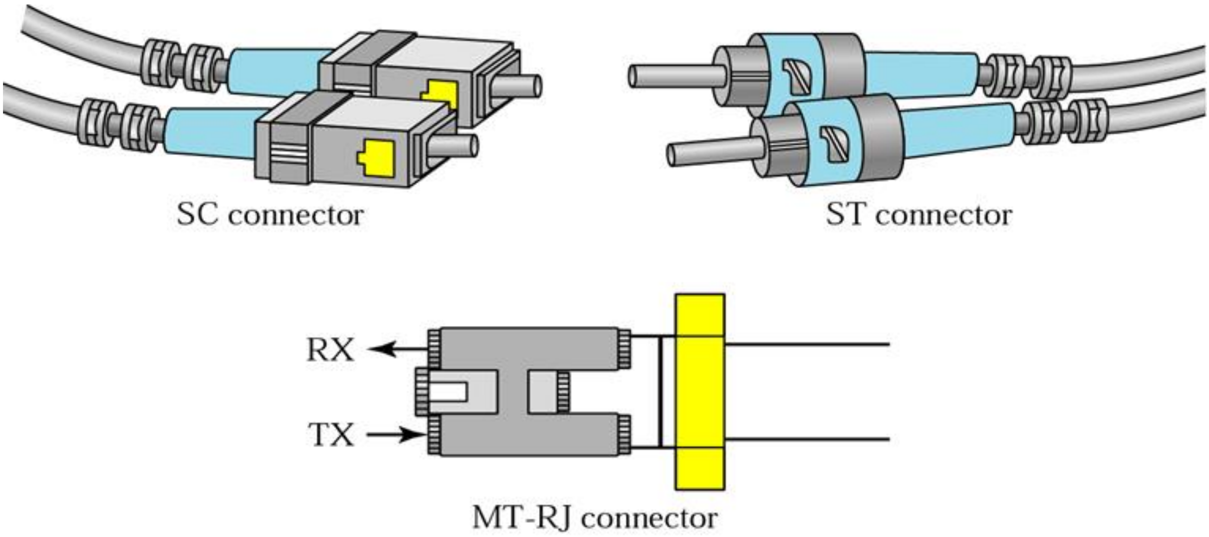


Fiber-Optic Cable Connectors:

There are three types ofconnectors for fiber-optic cables

4. SC Connector
5. ST Connector
6. MT-RJ

- The subscriber channel (SC) connector is used for cable TV.
- It uses a push/pull locking system.
- The straight-tip (ST) connector is used for connecting cable to networking devices.
- It uses a bayonet locking system and is more reliable than SC.
- MT-RJ is a connector that is the same size as RJ45.



Applications

- Fiber-optic cable is often found in backbone networks because its wide bandwidth is cost-effective.
- Some cable TV companies use a combination of optical fiber and coaxial cable, thus creating a hybrid network.
- Local-area networks such as 100Base-FX network (Fast Ethernet) and 1000Base-X also use fiber-optic cable.

Advantages and Disadvantages of Optical Fiber

Advantages:

- **Higher bandwidth**: Fiber-optic cable can support dramatically higher bandwidths (and hence data rates) than either twisted-pair or coaxial cable.
- **Less signal attenuation**: Fiber-optic transmission distance is significantly greater than that of other guided media. A signal can run for 50 km without requiring regeneration. We need repeaters every 5 km for coaxial or twisted-pair cable.
- **Immunity to electromagnetic interference**: Electromagnetic noise cannot affect fiber-optic cables.
- **Resistance to corrosive materials**: Glass is more resistant to corrosive materials than copper.
- **Light weight**: Fiber-optic cables are much lighter than copper cables.
- **Greater immunity to tapping:** Fiber-optic cables are more immune to tapping than copper cables.

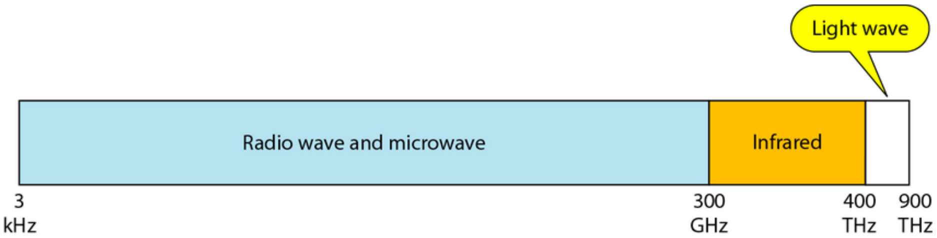
Disadvantages:

There are some disadvantages in the use of optical fiber.

- **Installation and maintenance:** Fiber-optic cable is a relatively new technology. Its installation and maintenance require expertise that is not yet available everywhere.
- **Unidirectional light propagation**: Propagation of light is unidirectional. If we need bidirectional communication, two fibers are needed.
- **Cost**: The cable and the interfaces are relatively more expensive than those of other guided media.

UNGUIDED MEDIA: WIRELESS

- Unguided media transport electromagnetic waves without using a physical conductor.
- This type of communication is often referred to as **wireless communication**.
- Signals are normally broadcast through free space and thus are available to anyone who has a device capable of receiving them.



Electromagnetic spectrum for wireless communication

Unguided signals can travel from the source to destination in several ways:

1. Ground propagation
2. Sky propagation
3. Line-of-sight propagation

Ground propagation

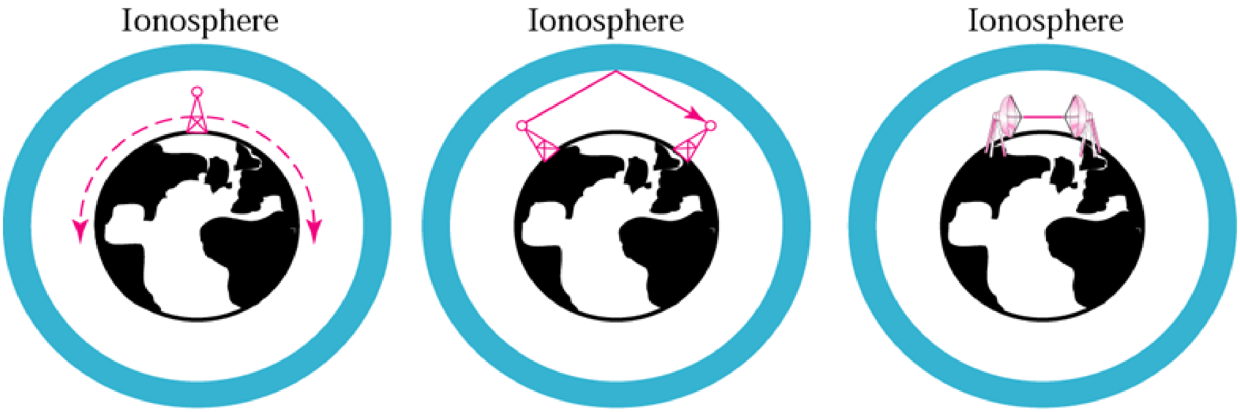
- In ground propagation, radio waves travel through the lowest portion of the atmosphere, hugging the earth.
- These low-frequency signals emanate in all directions from the transmitting antenna and follow the curvature of the planet.
- Distance depends on the amount ofpower in the signal: The greater the power, the greater the distance.

Sky propagation

- In sky propagation, higher-frequency radio waves radiate upward into the ionosphere (the layer of atmosphere where particles exist as ions) where they are reflected back to earth.
- This type of transmission allows for greater distances with lower output power.

Line-of-sight propagation

- In line-or-sight propagation, very high-frequency signals are transmitted in straight lines directly from antenna to antenna. Antennas must be directional, facing each other, and either tall enough or close enough together not to be affected by the curvature of the earth.



Ground propagation (below 2 MHz)

Sky propagation (2 - 30 MHz)

Line-of-sight propagation (above 30 MHz)

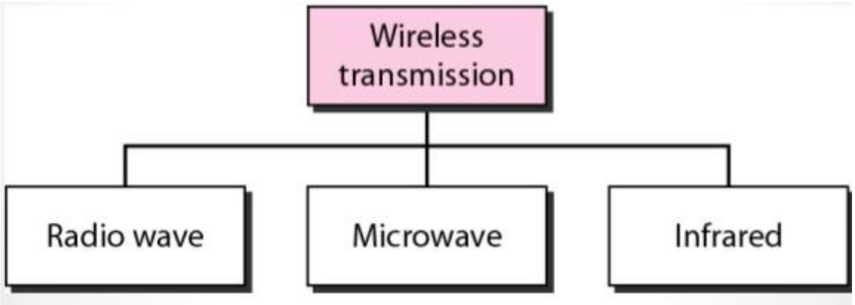
The section of the electromagnetic spectrum defined as radio waves and microwaves is divided into eight ranges, called bands, each regulated by government authorities.

These bands are rated from very low frequency (VLF) to extremely high frequency (EHF). Table below shows these bands, their ranges, propagation methods, and some applications.

Sl No	Band	Range	Propagation	Application
1	VLF (very low frequency)	3-30 kHz	Ground	Long-range radio navigation
2	LF (low frequency)	30-300 kHz	Ground	Radio beacons and navigational locators
3	MF (middle frequency)	300 kHz-3 MHz	Sky	AM radio
4	HF (high frequency)	3-30 MHz	Sky	Citizens band (CB), ship/aircraft communication
5	VHF (very high frequency)	30-300 MHz	Sky and Line-of-Sight	VHF TV, FM radio line-of-sight
6	UHF (ultrahigh frequency)	300 MHz-3 GHz	Line-of-sight	UHF TV, cellular phones, paging, satellite
7	SHF (super high frequency)	3-30 GHz	Line-of-sight	Satellite communication
8	EHF (extremely high frequency)	30-300 GHz	Line-of-sight	Radar, satellite

We can divide wireless transmission into three broad groups:

1. Radio waves
2. Microwaves
3. Infrared waves



Radio Waves

Although there is no clear-cut demarcation between radio waves and microwaves,

- Electromagnetic waves ranging in frequencies between **3 kHz and 1 GHz** are normally called **radio waves**;
- Waves ranging in frequencies between **1 and 300 GHz** are called **microwaves**.
- Radio waves, for the most part, are omni directional i.e.when an antenna transmits radio waves, they are propagated in all directions. This means that the sending and receiving antennas do not have to be aligned. A sending antenna sends waves that can be received by any receiving antenna. The omni directional property has a disadvantage, too. The radio waves transmitted by one antenna are susceptible to interference by another antenna that may send signals using the same frequency or band.
- Radio waves, particularly those waves that propagate in the sky mode, can travel long distances. This makes radio waves a good candidate for long-distance broadcasting such as AM radio.

Microwaves

Electromagnetic waves having frequencies between I and 300 GHz are called microwaves.

- Microwaves are unidirectional.
- When an antenna transmits microwave waves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. The unidirectional property has an obvious advantage. A pair of antennas can be aligned without interfering with another pair of aligned antennas.
- Microwave propagation is line-of-sight. Since the towers with the mounted antennas need to be in direct sight of each other, towers that are far apart need to be very tall. The curvature of the earth as well as other blocking obstacles do not allow two

short towers to communicate by using microwaves. Repeaters are often needed for long distance communication.

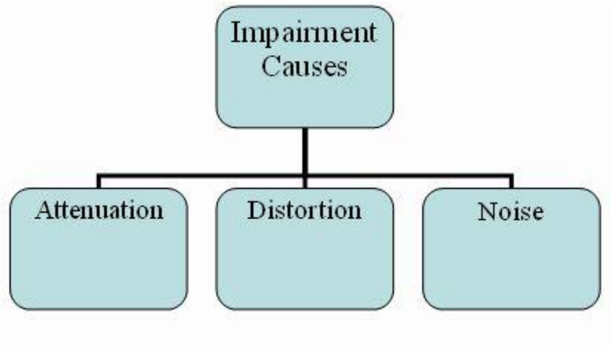
- Very high-frequency microwaves cannot penetrate walls. This characteristic can be a disadvantage ifreceivers are inside buildings.
- The microwave band is relatively wide, almost 299 GHz. Therefore wider subbands can be assigned, and a high data rate is possible

Infrared

- Infrared waves, with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm to 770 nm), can be used for short-range communication.
- Infrared waves, having high frequencies, cannot penetrate walls. This advantageous characteristic prevents interference between one system and another; a short-range communication system in one room cannot be affected by another system in the next room. When we use our infrared remote control, we do not interfere with the use of the remote by our neighbors. However, this same characteristic makes infrared signals useless for long-range communication.
- In addition, we cannot use infrared waves outside a building because the sun's rays contain infrared waves that can interfere with the communication.

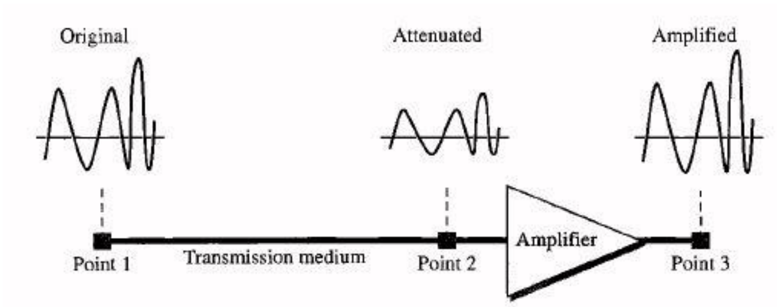
TRANSMISSION IMPAIRMENT

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



Attenuation

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



Decibel

To show that a signal has lost or gained strength, engineers use the unit of the decibel. **The decibel (dB) measures the relative strengths of two signals or one signal at two different points.** This decibel is negative if a signal is attenuated and positive if a signal is amplified.

$$\mathbf{dB=10\log_{10}\frac{P_2}{P_1}}$$

Where P_1 and P_2 are the powers of a signal at points 1 and 2, respectively.

We also define the decibel in terms of voltage instead of power. In this case, because power is proportional to the square of the voltage, the formula is

$$\mathbf{dB=20\log_{10}\frac{V_2}{V_1}}$$

Discussion :

1.Suppose a signal travels through a transmission medium and its power is reduced to one-half. This means that $P_2 = \frac{1}{2} P_1$

In this case, the attenuation (loss ofpower) can be calculated as

Attenuation $\mathbf{=10\log_{10}\frac{P_2}{P_1}}$

$$\mathbf{= 10\log_{10}\frac{0.5P_1}{P_1}}$$

$$\mathbf{=10\log_{10}0.5}$$

=-3 dB

A loss of3 dB (-3 dB) is equivalent to losing one-half the power.

2. A signal travels through an amplifier, and its power is increased 10 times. Find the amplification (gain of power)

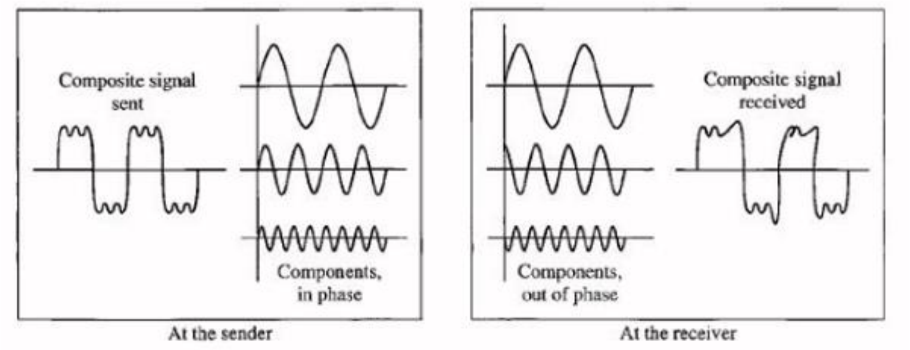
P_2=10 P_1

dB=10 $\log_{10} \frac{10 \text{ P1}}{P_1}$

=10

Distortion

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

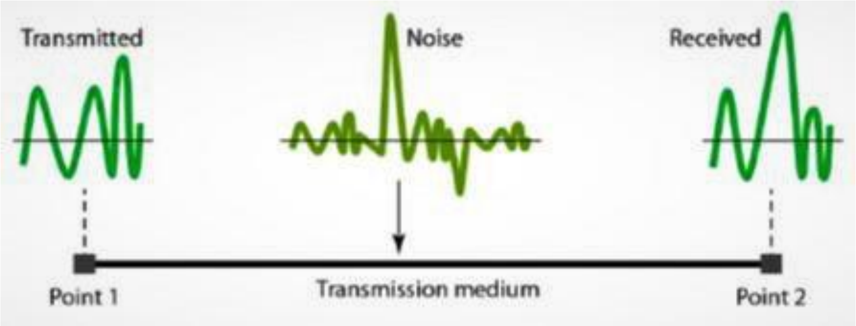


Distortion

Noise

- Noise is another cause of impairment.
- Several types of noise, such as thermal noise, induced noise, crosstalk, and impulse noise, may corrupt the signal.
- **Thermal noise is the random motion of electrons in a wire which creates an extra signal not originally sent by the transmitter**.
- **Induced noise comes from sources such as motors and appliances**. These devices act as a sending antenna, and the transmission medium acts as the receiving antenna.
- Crosstalk is the effect of one wire on the other. One wire acts as a sending antenna and the other as the receiving antenna.
- Impulse noise is a spike (a signal with high energy in a very short time) that comes from power lines, lightning, and so on.

The figure below shows the effect of noise on a signal



Signal-to-Noise Ratio (SNR)

It is useful in finding the theoretical bit rate limit.

The signal-to-noise ratio is defined as

$$\text{SNR}=\frac{\textbf{average signal power}}{\textbf{average noise power}}$$

- We need to consider the average signal power and the average noise power because these may change with time.
- SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise).
- A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise.
- Because SNR is the ratio of two powers, it is often described in decibel units.

$$\text{SNR}_{\text{dB}}=10\ \log_{10} SNR$$

Discussion: The power of a signal is 10 mW and the power of the noise is 1 μW; what are the values of SNR and SNR_{dB}?

The values of SNR and SNR_{dB} can be calculated as follows:

$$\text{SNR} =\frac{\textbf{average signal power}}{\textbf{average noise power}}=\frac{10^{-2}}{10^{-6}}=10^{4}=10000$$

$$\text{SNR}_{\text{dB}} = 10\ \log_{10} 10^{4}= 40$$

DATA RATE LIMITS

A very important consideration in data communications is how fast we can send data, in bits per second over a channel. Data rate depends on three factors:

1. The bandwidth available
2. The level of the signals we use
3. The quality of the channel (the level of noise)

Two theoretical formulas were developed to calculate the data rate:

- One by Nyquist for a noiseless channel.
- Other by Shannon for a noisy channel.

Noiseless Channel: Nyquist Bit Rate

For a noiseless channel, the Nyquist bit rate formula defines the theoretical maximum bit rate

Bit Rate= 2 x bandwidth x $\log_2 L$

Where bandwidth is the bandwidth of the channel
L is the number of signal levels used to represent data
Bit Rate is the bit rate in bits per second.

Discussion:

Consider a noiseless channel with a bandwidth of 3000 Hz transmitting a signal with two signal levels.

The maximum bit rate can be calculated as

Bit Rate= 2 x bandwidth x $\log_2 L$
=2 x 3000 x $\log_2 2$
=6000 bps

Consider the same noiseless channel transmitting a signal with four signal levels (for each level, we send 2 bits).

The maximum bit rate can be calculated as

Bit Rate= 2 x bandwidth x $\log_2 4$
=2 x 3000 x $\log_2 4$
=12000 bps

Noisy Channel: Shannon Capacity

- In reality, we cannot have a noiseless channel; the channel is always noisy.
- In 1944, Claude Shannon introduced a formula, called the Shannon capacity, to determine the theoretical highest data rate for a noisy channel:

Capacity =bandwidth X $\log_2$ (1 +SNR)

Where bandwidth is the bandwidth of the channel
SNR is the signal-to noise ratio
Capacity is the capacity of the channel in bits per second

Discussion:

Consider an extremely noisy channel in which the value of the signal-to-noise ratio is almost zero. In other words, the noise is so strong that the signal is faint. For this channel the capacity C is calculated as

$$C=B \log_2 (1 + SNR)=B \log_2 (1 + 0) =B \log_2 (1)=B \times 0=0$$

This means that the capacity of this channel is zero regardless of the bandwidth. In other words, we cannot receive any data through this channel.

A telephone line normally has a bandwidth of 3000 Hz (300 to 3300 Hz) assigned for data communications. The signal-to-noise ratio is usually 3162. For this channel the capacity is calculated as

$C = B \log_2 (1 + SNR)$
$= 3000 \log_2 (l + 3162)$
$= 3000 \log_2 3163$
$= 3000 \times 11.62 = 34{,}860$ bps

The signal-to-noise ratio is often given in decibels. Assume that SNR(dB) = 36 and the channel bandwidth is 2 MHz.

$SNR(dB) = 10 \log_{10}(SNR)$
$36 = 10 \log_{10}(SNR)$
$SNR = 10^{3.6} = 3981$

The theoretical channel capacity can now be calculated as

$C = B \log_2 (1 + SNR)$
$C = 2 \times 10^6 (1 + 3981)$
$C = 24$ Mbps

PERFORMANCE

One important issue in networking is the performance of the network-how good is it?

- **Bandwidth**
  One characteristic that measures network performance is bandwidth. However, the term can be used in two different contexts with two different measuring values: **bandwidth in hertz and bandwidth in bits per second**
  Bandwidth in Hertz Bandwidth in hertz is the range of frequencies contained in a composite signal or the range offrequencies a channel can pass
  Bandwidth in Bits per Seconds The term bandwidth can also refer to the number of bits per second that a channel, a link, or even a network can transmit
- **Throughput**
  The throughput is a measure of how fast we can actually send data through a network. Although, at first glance, bandwidth in bits per second and throughput seem the same, they are different. A link may have a bandwidth of B bps, but we can only send T bps through this link with T always less than B. In other words, the bandwidth is a potential measurement of a link; the throughput is an actual measurement of how fast we can send data.
- **Latency (Delay)**
  The latency or delay defines how long it takes for an entire message to completely arrive at the destination from the time the first bit is sent out from the source. We can say that latency is made of four components: propagation time, transmission time, queuing time and processing delay.
  Latency=propagation time+ transmission time +queuing time + processing delay

  Propagation Time
  Propagation time measures the time required for a bit to travel from the source to the destination. The propagation time is calculated by dividing the distance by the propagation speed.

Propagation time = Distance / Propagation speed

Transmission Time

In data communications we don't send just 1 bit, we send a message. The first bit may take a time equal to the propagation time to reach its destination; the last bit also may take the same amount of time. However, there is a time between the first bit leaving the sender and the last bit arriving at the receiver. The first bit leaves earlier and arrives earlier; the last bit leaves later and arrives later. The time required for transmission of a message depends on the size of the message and the bandwidth of the channel.

Transmission time=Message size Bandwidth

Queuing Time

The third component in latency is the queuing time, the time needed for each intermediate or end device to hold the message before it can be processed. The queuing time is not a fixed factor; it changes with the load imposed on the network. When there is heavy traffic on the network, the queuing time increases. An intermediate device, such as a router, queues the arrived messages and processes them one by one. If there are many messages, each message will have to wait.

- **Jitter**

Another performance issue that is related to delay is jitter. We can roughly say that jitter is a problem if different packets of data encounter different delays and the application using the data at the receiver site is time-sensitive (audio and video data, for example). If the delay for the first packet is 20 ms, for the second is 45 ms, and for the third is 40 ms, then the real-time application that uses the packets endures jitter.

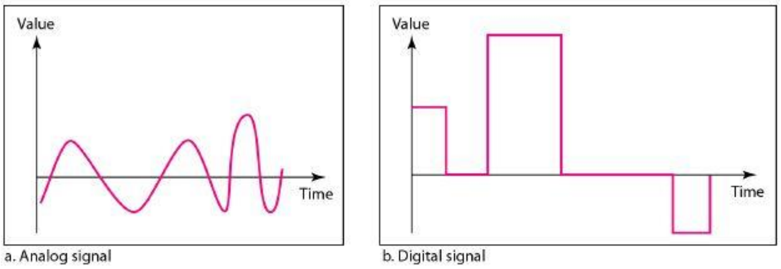
Unit-3. Data Encoding

Analog and Digital Data

- Data can be analog or digital.
- The term **analog data** refers to information that is continuous
- **Digital data** refers to information that has discrete states.
- For example, an analog clock that has hour, minute, and second hands gives information in a continuous form; the movements of the hands are continuous. On the other hand, a digital clock that reports the hours and the minutes will change suddenly from 8:05 to 8:06.
- Analog data, such as the sounds made by a human voice, take on continuous values. When someone speaks, an analog wave is created in the air. This can be captured by a microphone and converted to an analog signal or sampled and converted to a digital signal.

Analog and Digital Signals

- Like the data we represent, signals can be either **analog or digital**.
- An **analog signal** has infinitely many levels of intensity over a period of time.
- A digital signal, on the other hand, can have only a limited number of defined values.



a. Analog signal b. Digital signal

Periodic and Non periodic Signals

- Both analog and digital signals can take one of two forms: **periodic or non periodic**
- A periodic signal completes a pattern within a measurable time frame, called a period, and repeats that pattern over subsequent identical periods. The completion of one full pattern is called a cycle.
- A nonperiodic signal changes without exhibiting a pattern or cycle that repeats over time.
- Both analog and digital signals can be periodic or nonperiodic. In data communications, we commonly use periodic analog signals and nonperiodic digital signals.

Periodic Analog Signals

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

Sine Wave

- The sine wave is the most fundamental form of a periodic analog signal. When we visualize it as a simple oscillating curve, its change over the course of a cycle is smooth and consistent, a continuous, rolling flow.
- A sine wave can be represented by three parameters: **the peak amplitude, the frequency, and the phase**. These three parameters fully describe a sine wave

Peak Amplitude

The peak amplitude of a signal is the absolute value of its highest intensity, proportional to the energy it carries. For electric signals, peak amplitude is normally measured in volts.

Period and Frequency

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

$$f=1/T$$

and

$$T=1/f$$

- Period is formally expressed in seconds. Frequency is formally expressed in Hertz (Hz), which is cycle per second

Phase

- The term phase describes the position of the waveform relative to time 0. If we think of the wave as something that can be shifted backward or forward along the time axis, phase describes the amount of that shift. It indicates the status of the first cycle.
- Phase is measured in degrees or radians .

Wavelength

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

Discussion: Find the wavelength of red light in air whose frequency is **4×10^{14} Hz**

$$\lambda=c/f$$

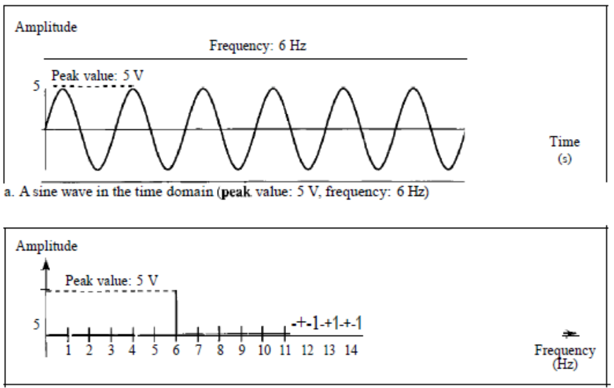
$$=3 \times 10^{8} / 4 \times 10^{14}$$

$$=0.75 \times 10^{-6}=0.75\ \mu m$$

Time and Frequency Domains

- A sine wave is comprehensively defined by its amplitude, frequency, and phase.
- We have been showing a sine wave by using what is called a time-domain plot.
- The time-domain plot shows changes in signal amplitude with respect to time (it is an amplitude-versus-time plot). Phase is not explicitly shown on a time-domain plot.

- To show the relationship between amplitude and frequency, we can use what is called a frequency-domain plot. A frequency-domain plot is concerned with only the peak value and the frequency. Changes of amplitude during one period are not shown.



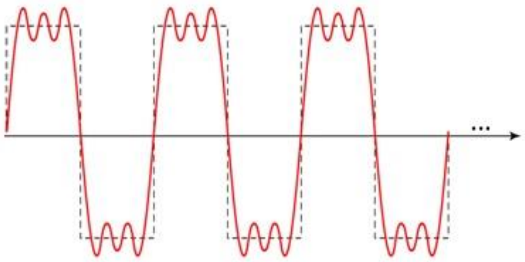
a. A sine wave in the time domain (peak value: 5 V, frequency: 6 Hz)

b. The same sine wave in the frequency domain (peak value: 5 V, frequency: 6 Hz)

- It is obvious that the frequency domain is easy to plot and conveys the information that one can find in a time domain plot.
- The advantage of the frequency domain is that we can immediately see the values of the frequency and peak amplitude.
- A complete sine wave is represented by one spike.
- The position of the spike shows the frequency; its height shows the peak amplitude.

Composite Signals

- A composite signal is made of many simple sine waves.
- Any composite signal is actually a combination of simple sine waves with different frequencies, amplitudes, and phases.
- A composite signal can be periodic or nonperiodic.
- A periodic composite signal can be decomposed into a series of simple sine waves with discrete frequencies, frequencies that have integer values (1, 2, 3, and so on).
- A nonperiodic composite signal can be decomposed into a combination of an infinite number of simple sine waves with continuous frequencies, frequencies that have real values.



Composite Periodic Signal

Bandwidth

- The range of frequencies contained in a composite signal is its bandwidth.
- The bandwidth is normally a difference between two numbers. For example, if a composite signal contains frequencies between 1000 and 5000, its bandwidth is 5000 - 1000, or 4000.

Discussion: If a periodic signal is decomposed into five sine waves with frequencies of 100, 300, 500, 700, and 900 Hz, what is its bandwidth?

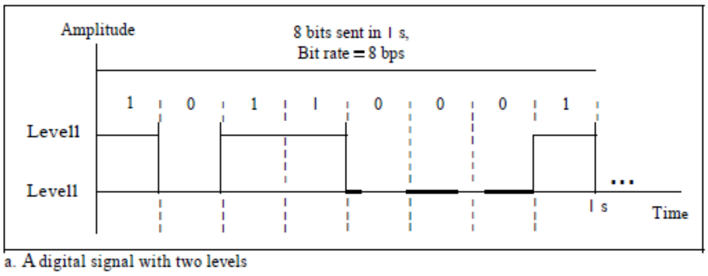
A periodic signal has a bandwidth of 20 Hz. The highest frequency is 60 Hz. What is the lowest frequency?

$100ms=100 \times 10^{-3} = 10^{-1}=10^{-1} \times 10^{6} \times 10^{-6} =10^{5}$ μsec

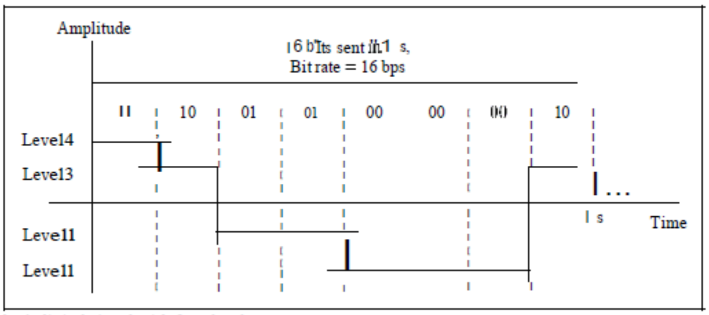
DIGITAL SIGNALS

- In addition to being represented by an analog signal, information can also be represented by a digital signal.
- For example, a 1 can be encoded as a positive voltage and a 0 as zero voltage. A digital signal can have more than two levels. In this case, we can send more than 1 bit for each level.

Two digital signals: one with two signal levels and the other with four signal levels



a. A digital signal with two levels



b. A digital signal with four levels

-

We send 1 bit per level in part a of the figure and 2 bits per level in part b of the figure. In general, if a signal has L levels, each level needs $\log_2 L$ bits.

Discussion: A digital signal has eight levels. How many bits are needed per level?

Number of bits per level $=\log_2 8 =3$

Each signal level is represented by 3 bits.

Bit Rate

- Most digital signals are non periodic, and thus period and frequency are not appropriate characteristics.
- Another term-bit rate (instead of frequency)-is used to describe digital signals.
- The bit rate is the number of bits sent in 1s, expressed in bits per second (bps).

- The above figure shows the bit rate for two signals. One is 8bps and other one is 16bps.

Bit Length

- The wavelength for an analog signal is the distance one cycle occupies on the transmission medium.
- We can define something similar for a digital signal: the bit length.
- **The bit length is the distance one bit occupies on the transmission medium.**

Bit length=propagation speed x bit duration

Digital Signal as a Composite Analog Signal

- Based on Fourier analysis, a digital signal is a composite analog signal.
- The bandwidth is infinite.
- A digital signal, in the time domain, comprises connected vertical and horizontal line segments.
- A vertical line in the time domain means a frequency of infinity (sudden change in time);
- A horizontal line in the time domain means a frequency of zero (no change in time). Going from a frequency of zero to a frequency of infinity (and vice versa) implies all frequencies in between are part of the domain.

Transmission of Digital Signals

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

Baseband Transmission

- Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal.
- Baseband transmission requires that we have a low-pass channel, a channel with a bandwidth that starts from zero.
- Baseband transmission of a digital signal that preserves the shape of the digital signal is possible only if we have a low-pass channel with an infinite or very wide bandwidth.
- In baseband transmission, the required bandwidth is proportional to the bit rate; if we need to send bits faster, we need more bandwidth.

Broadband Transmission (Using Modulation)

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

DIGITAL-TO-DIGITAL CONVERSION

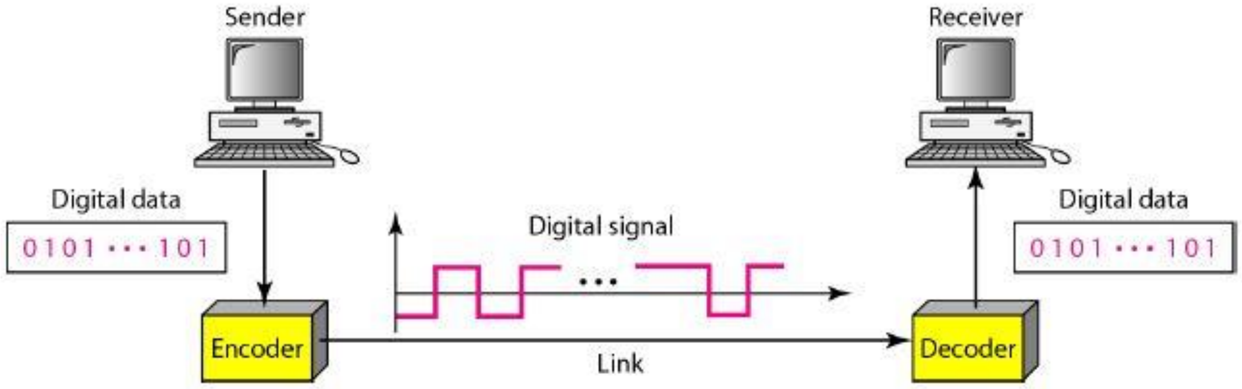
We discussed data and signals. We know that data can be either **digital or analog**. We also know that signals that represent data can also be digital or analog. In this chapter, we see how we can represent digital data by using digital signals. The conversion involves three techniques:

- Line coding
- Block coding
- Scrambling

Line Coding

- Line coding is the process of converting digital data to digital signals.
- We assume that data, in the form of text, numbers, graphical images, audio, or video, are stored in computer memory as sequences of bits .
- Line coding converts a sequence of bits to a digital signal.
- At the sender, digital data are encoded into a digital signal; at the receiver, the digital data are recreated by decoding the digital signal.

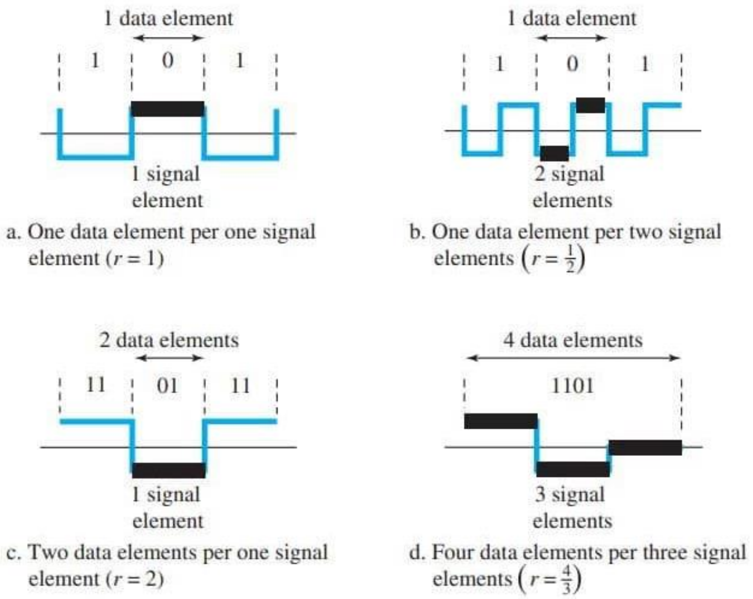
Figure below shows the process.



Characteristics of Line Coding

1. Signal Element Versus Data Element
   - A data element is the smallest entity that can represent a piece of information: this is the bit.
   - A signal element carries data elements.
   - A signal element is the shortest unit (time wise) of a digital signal.
   - In other words, data elements are what we need to send; signal elements are what we can send.
   - Data elements are being carried; signal elements are the carriers.

We define a ratio r which is the number of data elements carried by each signal element



a. One data element per one signal element ($r = 1$)

b. One data element per two signal elements $\left(r = \frac{1}{2}\right)$

c. Two data elements per one signal element ($r = 2$)

d. Four data elements per three signal elements $\left(r = \frac{4}{3}\right)$

2. Data Rate Versus Signal Rate
   - The data rate defines the number of data elements (bits) sent in 1s.
   - The unit is bits per second (bps).
   - The signal rate is the number of signal elements sent in 1s.
   - The unit is the baud.
   - The data rate is sometimes called the bit rate
   - The signal rate is sometimes called the pulse rate, the modulation rate, or the baud rate.
   - We can formulate the relationship between data rate and signal rate as

$$S = c \times N \times 1/r \qquad \text{baud}$$

Where N is the data rate (bps); c is the case factor, which varies for each case; S is the number of signal elements; and r is the previously defined factor.

Discussion: A signal is carrying data in which one data element is encoded as one signal element (r = 1). If the bit rate is 100 kbps, what is the average value of the baud rate if c is between 0 and 1?

We assume that the average value of c is 1/2. The baud rate is then

$$S = 1/2 \times 100 \times 1 = 50 \text{ kbaud}$$

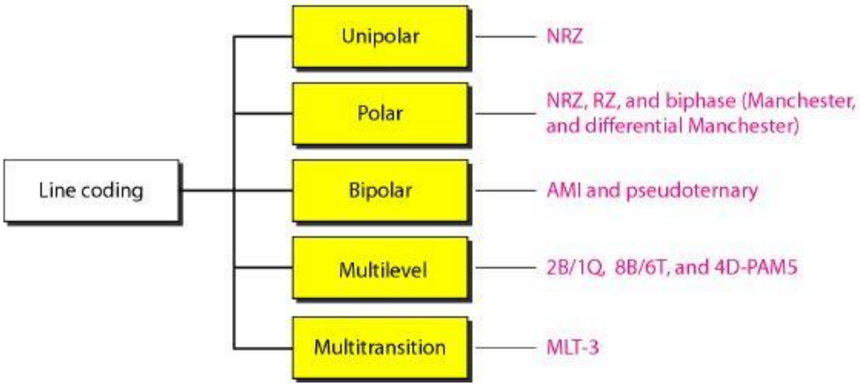
3. Baseline Wandering
   - In decoding a digital signal, the receiver calculates a running average of the received signal power.
   - This average is called the baseline.
   - The incoming signal power is evaluated against this baseline to determine the value of the data element. A long string of 0s or 1s can cause a drift in the baseline

(baseline wandering) and make it difficult for the receiver to decode correctly. A good line coding scheme needs to prevent baseline wandering.

4. DC Components
    - When the voltage level in a digital signal is constant for a while, the spectrum creates very low frequencies (results of Fourier analysis).
    - These frequencies around zero, called DC (direct-current) components, present problems for a system that cannot pass low frequencies or a system that uses electrical coupling (via a transformer). For example, a telephone line cannot pass frequencies below 200 Hz. Also a long-distance link may use one or more transformers to isolate different parts of the line electrically. For these systems, we need a scheme with no DC component.
5. Self-synchronization
    - To correctly interpret the signals received from the sender, the receiver's bit intervals must correspond exactly to the sender's bit intervals.
    - If the receiver clock is faster or slower, the bit intervals are not matched and the receiver might misinterpret the signals.
6. Built-in Error Detection
    - It is desirable to have a built-in error-detecting capability in the generated code to detect some of or all the errors that occurred during transmission.
7. Immunity to Noise and Interference
    - Another desirable code characteristic is a code that is immune to noise and other interferences.
8. Complexity
    - A complex scheme is more costly to implement than a simple one. For example, a scheme that uses four signal levels is more difficult to interpret than one that uses only two levels.

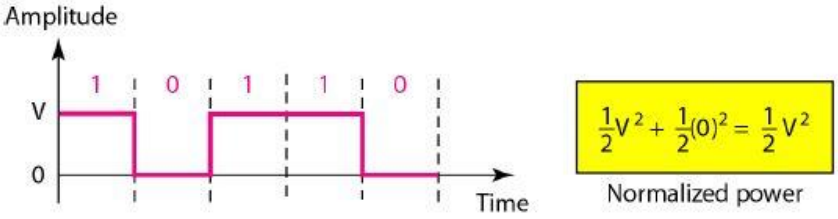
Line Coding Schemes

We can roughly divide line coding schemes into five broad categories.


Line coding
Unipolar
NRZ
Polar
NRZ, RZ, and biphase (Manchester, and differential Manchester)
Bipolar
AMI and pseudoternary
Multilevel
2B/1Q, 8B/6T, and 4D-PAM5
Multitransition
MLT-3

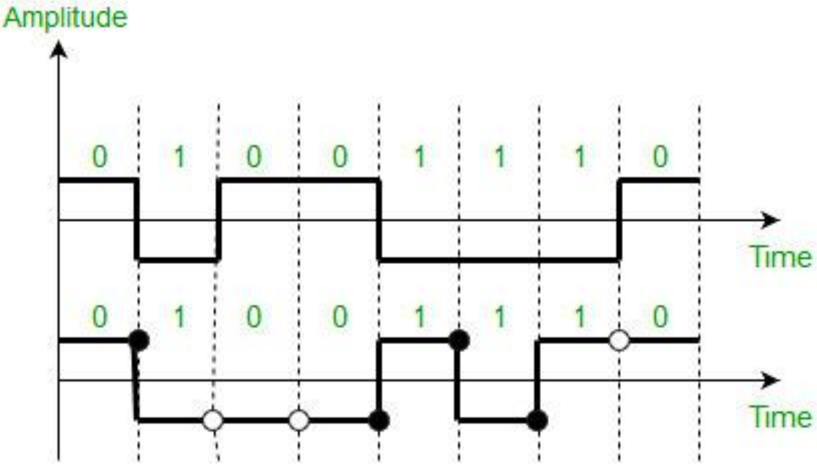
1. **<u>Unipolar Scheme</u>**
    - In a unipolar scheme, all the signal levels are on one side of the time axis, either above or below.

- <u>NRZ (Non-Return-to-Zero)</u> Traditionally, a unipolar scheme was designed as a non-return-to-zero (NRZ) scheme in which the positive voltage defines bit 1 and the zero voltage defines bit 0. It is called NRZ because the signal does not return to zero at the middle of the bit. Figure below show a unipolar NRZ scheme.



2. <u>Polar Schemes</u>

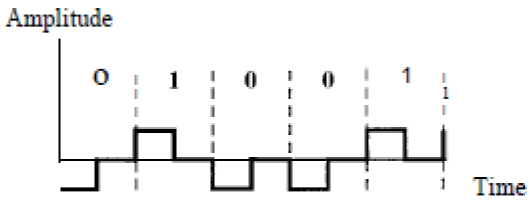
- In polar schemes, the voltages are on the both sides of the time axis. For example, the voltage level for 0 can be positive and the voltage level for 1 can be negative.
- <u>Non-Return-to-Zero (NRZ)</u> In polar NRZ encoding, we use two levels of voltage amplitude. We can have two versions of polar NRZ: NRZ-L and NRZ-I.
- In the first variation, NRZ-L (NRZ-Level), the level of the voltage determines the value of the bit. In the second variation, NRZ-I (NRZ-Invert), the change or lack of change in the level of the voltage determines the value of the bit. If there is no change, the bit is 0; if there is a change, the bit is 1.

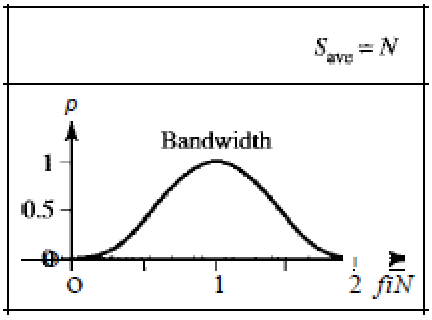


Return to Zero (RZ)

- The main problem with NRZ encoding occurs when the sender and receiver clocks are not synchronized.
- The receiver does not know when one bit has ended and the next bit is starting.
- One solution is the **return-to-zero (RZ) scheme**, which uses three values: **positive, negative, and zero**.
- In RZ, the signal changes not between bits but during the bit.
- In the below figure the signal goes to 0 in the middle of each bit. It remains there until the beginning of the next bit.

- The main disadvantage of RZ encoding is that it requires two signal changes to encode a bit and therefore occupies greater bandwidth.
- There is no DC component problem.
- Another problem is the complexity: RZ uses three levels of voltage, which is more complex to create and discern.
- As a result of all these deficiencies, the scheme is not used today.
- Instead, it has been replaced by the better-performing **Manchester and differential Manchester schemes**



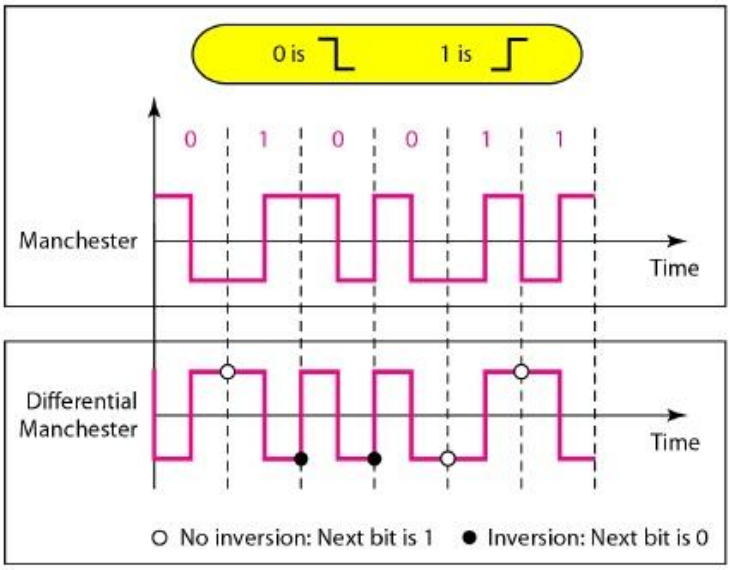


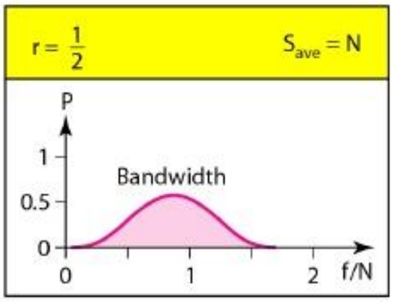
Biphase

Manchester and Differential Manchester

- The idea of RZ (transition at the middle of the bit) and the idea of NRZ-L are combined into the Manchester scheme.
- In Manchester encoding, the duration of the bit is divided into two halves.
- The voltage remains at one level during the first half and moves to the other level in the second half.
- The transition at the middle of the bit provides synchronization.
- Differential Manchester, on the other hand, combines the ideas of RZ and NRZ-I. There is always a transition at the middle of the bit, but the bit values are determined at the beginning of the bit.
- If the next bit is 0, there is a transition; if the next bit is 1, there is none.

Figure below shows both Manchester and differential Manchester encoding

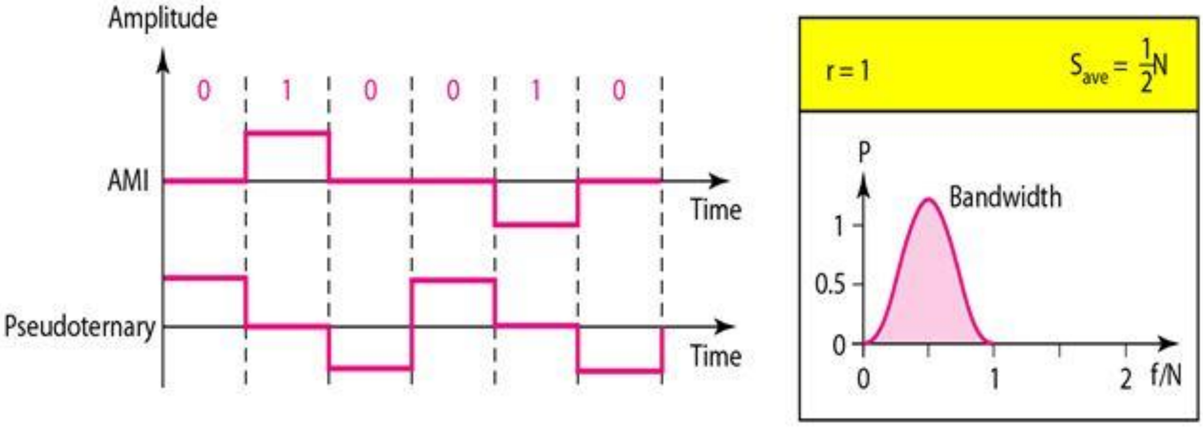




- The Manchester scheme overcomes several problems associated with NRZ-L, and differential Manchester overcomes several problems associated with NRZ-I.
- First, there is no baseline wandering.
- There is no DC component because each bit has a positive and negative voltage contribution.
- The only drawback is the signal rate.
- The signal rate for Manchester and differential Manchester is double that for NRZ. The reason is that there is always one transition at the middle of the bit and maybe one transition at the end of each bit.

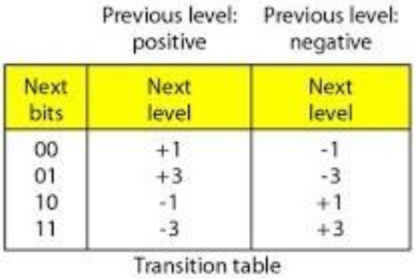
Bipolar Schemes

- In bipolar encoding (sometimes called multilevel binary), there are three voltage levels: positive, negative, and zero.
- The voltage level for one data element is at zero, while the voltage level for the other element alternates between positive and negative.
- There are two variations of bipolar encoding: **AMI and pseudoternary**.
- A common bipolar encoding scheme is called bipolar **alternate mark inversion** (AMI). In the term alternate mark inversion, the word mark comes from telegraphy and means 1. So AMI means alternate 1 inversion.
- A neutral zero voltage represents binary 0. Binary 1s are represented by alternating positive and negative voltages.
- A variation of AMI encoding is called pseudoternary in which the 1 bit is encoded as a zero voltage and the 0 bit is encoded as alternating positive and negative voltages.



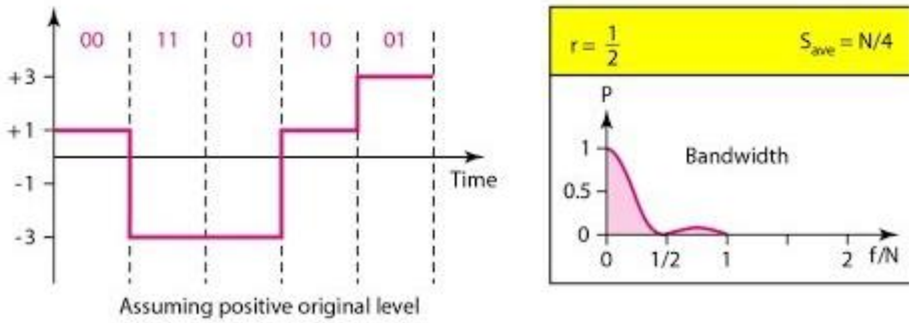
- The bipolar scheme was developed as an alternative to NRZ.
- The bipolar scheme has the same signal rate as NRZ, but there is no DC component.
- The NRZ scheme has most of its energy concentrated near zero frequency, which makes it unsuitable for transmission over channels with poor performance around this frequency.
- The concentration of the energy in bipolar encoding is around frequency N/2.
- Multilevel Scheme
- The desire to increase the data speed or decrease the required bandwidth has resulted in the creation of many schemes.
- The goal is to increase the number of bits per baud by encoding a pattern of m data elements into a pattern of n signal elements.
- We only have two types of data elements (0s and 1s), which means that a group of m data elements can produce a combination of 2m data patterns.
- We can have different types of signal elements by allowing different signal levels. If we have L different levels, then we can produce Ln combinations of signal patterns.
- If 2m = Ln, then each data pattern is encoded into one signal pattern.
- If 2m< Ln, data patterns occupy only a subset of signal patterns. The subset can be carefully designed to prevent baseline wandering, to provide synchronization, and to detect errors that occurred during data transmission.
- Data encoding is not possible if 2m > Ln because some of the data patterns cannot be encoded.
- The code designers have classified these types of coding as mBnL, where m is the length of the binary pattern, B means binary data, n is the length of the signal pattern, and L is the number of levels in the signaling. A letter is often used in place of L: B(binary)for L=2, T (ternary) for L =3, and Q (quaternary) for L =4.
-
- 2B1Q

- The first mBnL scheme is two binary, one quaternary (2B1Q), uses data patterns of size 2 and encodes the 2-bit patterns as one signal element belonging to a four-level signal. In this type of encoding m =2, n =1, and L =4 (quatenary).
- Figure below shows an example of a 2B1Q signal.
- The average signal rate of 2B1Q is S =N/4. This means that using 2B1Q, we can send data 2 times faster than by using NRZ-L.
- However, 2B 1Q uses four different signal levels, which means the receiver has to discern four different thresholds.
- The reduced bandwidth comes with a price. There are no redundant signal patterns in this scheme because 22 =41
-

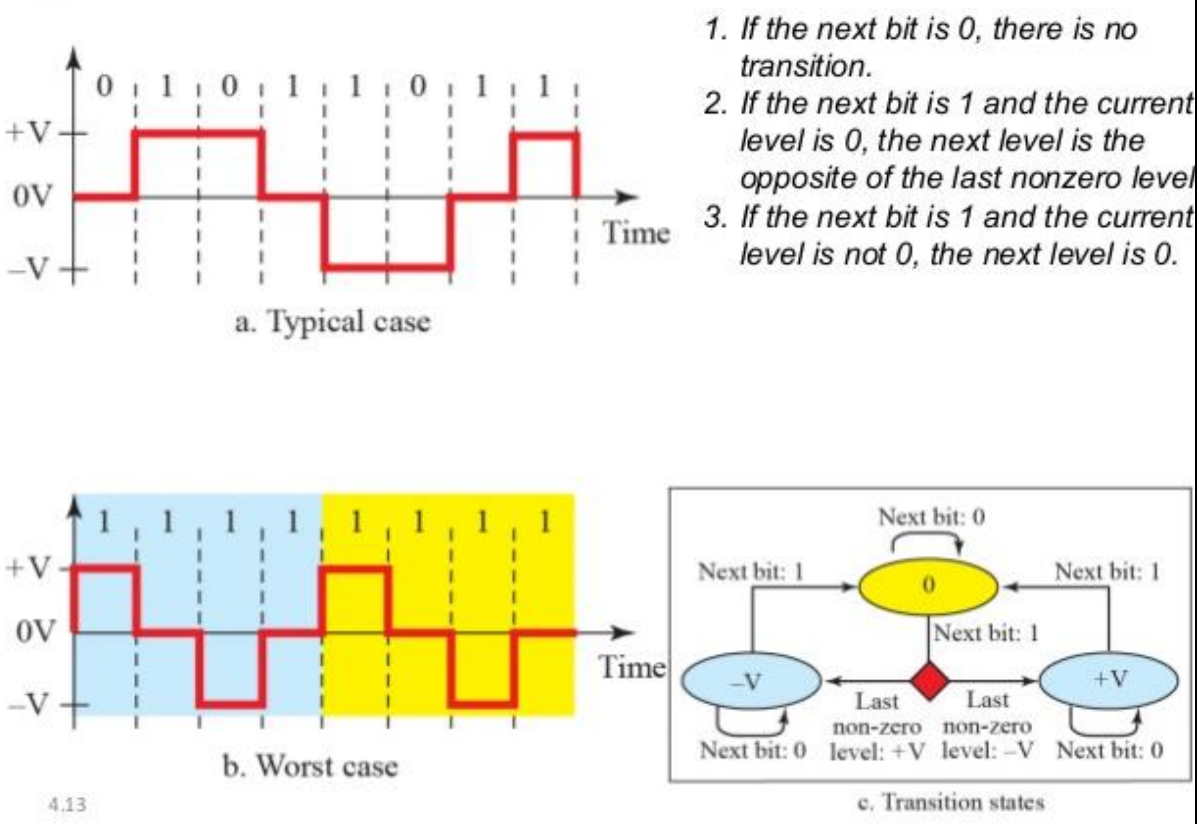


	Previous level: positive	Previous level: negative
Next bits	Next level	Next level
00	+1	-1
01	+3	-3
10	-1	+1
11	-3	+3

Transition table



-
-
- Multiline Transmission:
- MLT-3
- NRZ-I and differential Manchester are classified as differential encoding but use two transition rules to encode binary data (no inversion, inversion).
- If we have a signal with more than two levels, we can design a differential encoding scheme with more than two transition rules. MLT-3 is one of them.
- The multiline transmission, three level (MLT-3) scheme uses three levels (+V, 0, and - V) and three transition rules to move between the levels.
- If the next bit is 0, there is no transition.
- If the next bit is 1 and the current level is not 0, the next level is 0.
- If the next bit is 1 and the current level is 0, the next level is the opposite of the last nonzero level.



1. *If the next bit is 0, there is no transition.*
2. *If the next bit is 1 and the current level is 0, the next level is the opposite of the last nonzero level*
3. *If the next bit is 1 and the current level is not 0, the next level is 0.*

a. Typical case

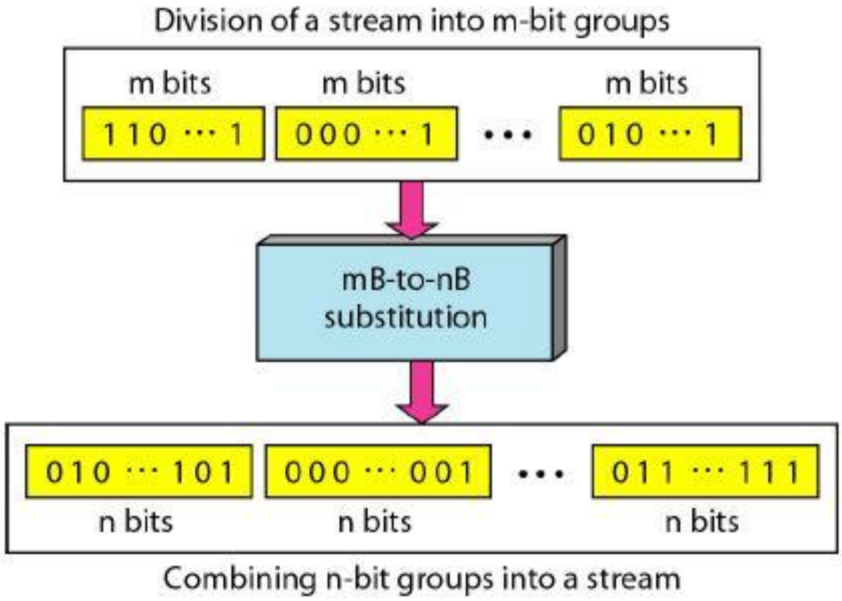
b. Worst case

4.13

c. Transition states

Block Coding

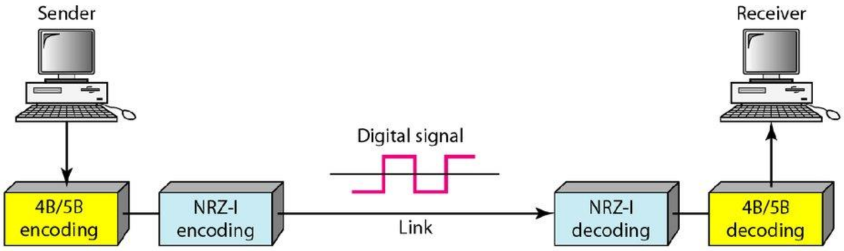
- We need redundancy to ensure synchronization and to provide some kind of inherent error detecting.
- Block coding can give us this redundancy and improve the performance of line coding.
- In general, block coding changes a block of m bits into a block of n bits, where n is larger than m.
- Block coding is referred to as an mB/nB encoding technique.
- The slash in block encoding (for example, 4B/5B) distinguishes block encoding from multilevel encoding (for example, 8B6T), which is written without a slash.
- Block coding normally involves three steps: division, substitution, and combination.
- In the division step, a sequence of bits is divided into groups of m bits.
- For example, in 4B/5B encoding, the original bit sequence is divided into 4-bit groups. The heart of block coding is the substitution step.
- In this step, we substitute an m-bit group for an n-bit group.
- For example, in 4B/5B encoding we substitute a 4-bit code for a 5-bit group.
- Finally, the n-bit groups are combined together to form a stream. The new stream has more bits than the original bits. Figure below shows the procedure.



4B/5B

- The four binary/five binary (4B/5B) coding scheme was designed to be used in combination with NRZ-I.
- NRZ-I has a good signal rate, one-half that of the biphase, but it has a synchronization problem. A long sequence of 1s can make the receiver clock lose synchronization.
- One solution is to change the bit stream, prior to encoding with NRZ-I, so that it does not have a long stream of 0s.
- The 4B/5B scheme achieves this goal. The block-coded stream does not have more than three consecutive 0s, as we will see later.
- At the receiver, the NRZ-I encoded digital signal is first decoded into a stream of bits and then decoded to remove the redundancy. Figure below shows the idea.

Using block coding 4B/5B with NRZ-I line coding scheme



ANALOG-TO-DIGITAL CONVERSION

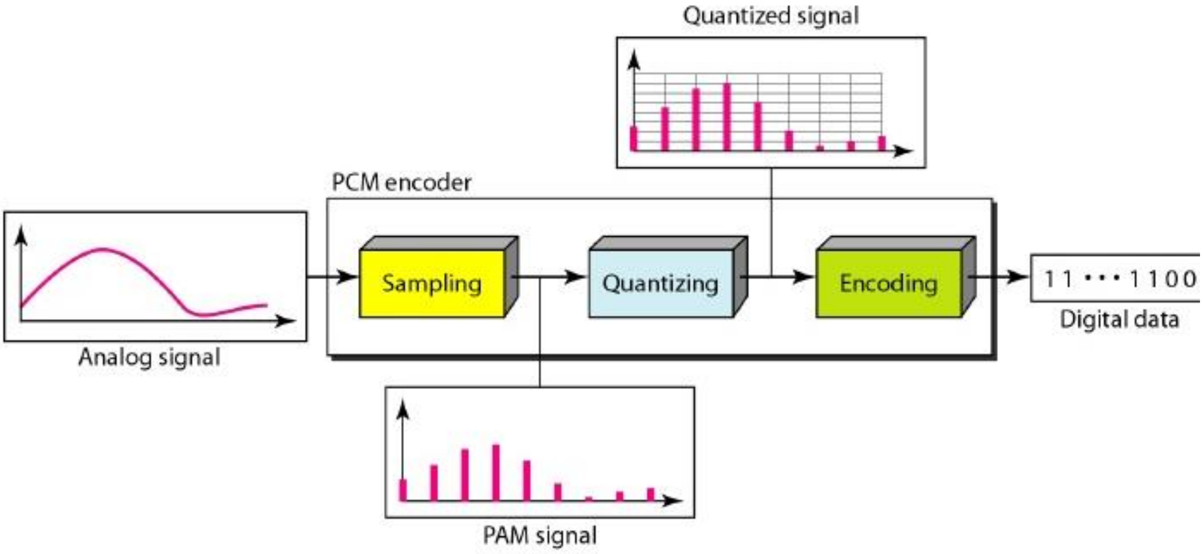
Pulse Code Modulation (PCM)

The most common technique to change an analog signal to digital data (digitization) is called pulse code modulation (PCM). A PCM encoder has three processes, as shown in figure below

1. The analog signal is sampled.

2. The sampled signal is quantized.

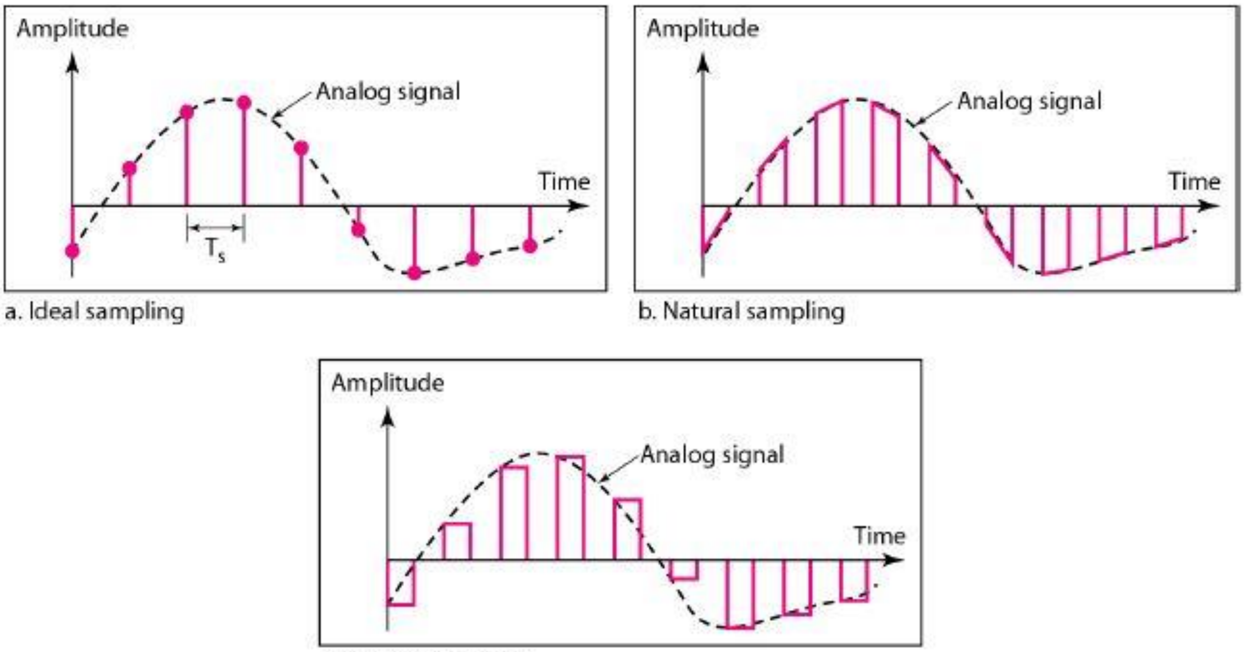
3. The quantized values are encoded as streams of bits.



Sampling

- The first step in PCM is sampling.
- The analog signal is sampled every T_s sec, where T_s is the sample interval or period.
- The inverse of the sampling interval is called the sampling rate or sampling frequency and denoted by f_s, where $f_s = 1/ T_s$.
- There are three sampling methods-**ideal, natural, and flat-top**-as shown in figure.
- In **ideal sampling**, pulses from the analog signal are sampled. This is an ideal sampling method and cannot be easily implemented.
- In **natural sampling**, a high-speed switch is turned on for only the small period of time when the sampling occurs. The result is a sequence of samples that retains the shape of the analog signal.
- The most common sampling method, called sample and hold, however, creates flat-top samples by using a circuit.

The sampling process is sometimes referred to as **pulse amplitude modulation** (PAM).



a. Ideal sampling

b. Natural sampling

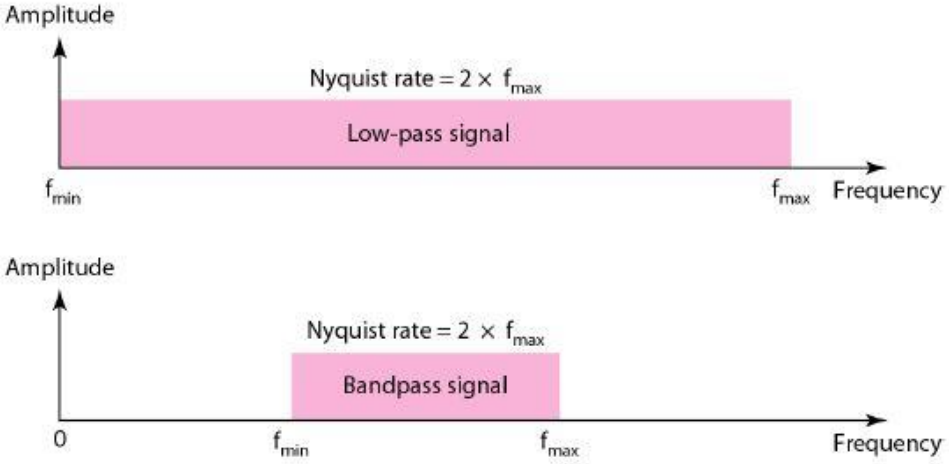
c. Flat-top sampling

Sampling Rate

One important consideration is the sampling rate or frequency.

According to the Nyquist theorem, **to reproduce the original analog signal, one necessary condition is that the sampling rate be at least twice the highest frequency in the original signal.**

- First, we sample a signal only if the signal is band-limited. In other words, a signal with an infinite bandwidth cannot be sampled.
- Second, the sampling rate must be at least 2 times the highest frequency, not the bandwidth. If the analog signal is low-pass, the bandwidth and the highest frequency are the same value. If the analog signal is bandpass, the bandwidth value is lower than the value of the maximum frequency. Figure below shows the value of the sampling rate for two types of signals.



Quantization

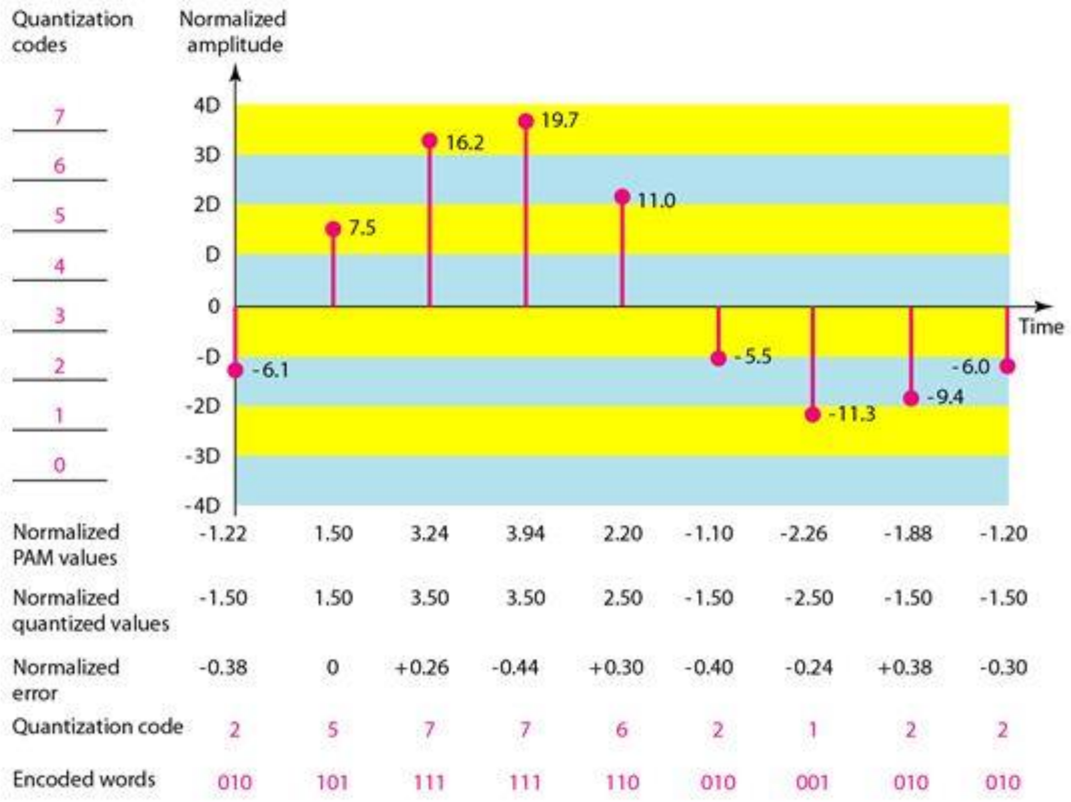
- The result of sampling is a series of pulses with amplitude values between the maximum and minimum amplitudes of the signal.
- The set of amplitudes can be infinite with nonintegral values between the two limits.
- These values cannot be used in the encoding process. The following are the steps in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

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

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

3. We assign quantized values of 0 to L - 1 to the midpoint of each zone.
4. We approximate the value of the sample amplitude to the quantized values.

As a simple example, assume that we have a sampled signal and the sample amplitudes are between -20 and +20 V. We decide to have eight levels (L = 8). This means that Δ =5 V. Figure below shows this example.



Normalized PAM values	-1.22	1.50	3.24	3.94	2.20	-1.10	-2.26	-1.88	-1.20
Normalized quantized values	-1.50	1.50	3.50	3.50	2.50	-1.50	-2.50	-1.50	-1.50
Normalized error	-0.38	0	+0.26	-0.44	+0.30	-0.40	-0.24	+0.38	-0.30
Quantization code	2	5	7	7	6	2	1	2	2
Encoded words	010	101	111	111	110	010	001	010	010

- We have shown only nine samples using ideal sampling (for simplicity).
- The value at the top of each sample in the graph shows the actual amplitude.
- In the chart, the first row is the normalized value for each sample (actual amplitude/Δ).
- The quantization process selects the quantization value from the middle of each zone. This means that the normalized quantized values (second row) are different from the normalized amplitudes.
- The difference is called the normalized error (third row).

- The fourth row is the quantization code for each sample based on the quantization levels at the left of the graph.
- The encoded words (fifth row) are the final products of the conversion.

Quantization Levels

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

Quantization Error

- One important issue is the error created in the quantization process.
- Quantization is an approximation process.
- The input values to the quantizer are the real values; the output values are the approximated values.
- The output values are chosen to be the middle value in the zone.
- If the input value is also at the middle of the zone, there is no quantization error; otherwise, there is an error.
- In the previous example, the normalized amplitude of the third sample is 3.24, but the normalized quantized value is 3.50. This means that there is an error of +0.26.
- The value of the error for any sample is less than $\Delta/2$. In other words, we have $-\Delta/2 \leq$ error $\leq \Delta/2$.
- The quantization error changes the signal-to-noise ratio of the signal, which in turn reduces the upper limit capacity according to Shannon.

Encoding

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

The bit rate can be found from the formula

Bit rate= Sampling rate x number of bits per sample=f_s x n_b

Discussion: We want to digitize the human voice. What is the bit rate, assuming 8 bits per sample?

The human voice normally contains frequencies from 0 to 4000 Hz. So the sampling rate and bit rate are calculated as follows:

Sampling rate= 4000 x 2 = 8000 samples/sec

Bitrate = 8000 x 8 bits/sec=64,000 bps =64 kbps

Original Signal Recovery

The recovery of the original signal requires the PCM decoder.

The decoder first uses circuitry to convert the code words into a pulse that holds the amplitude until the next pulse.

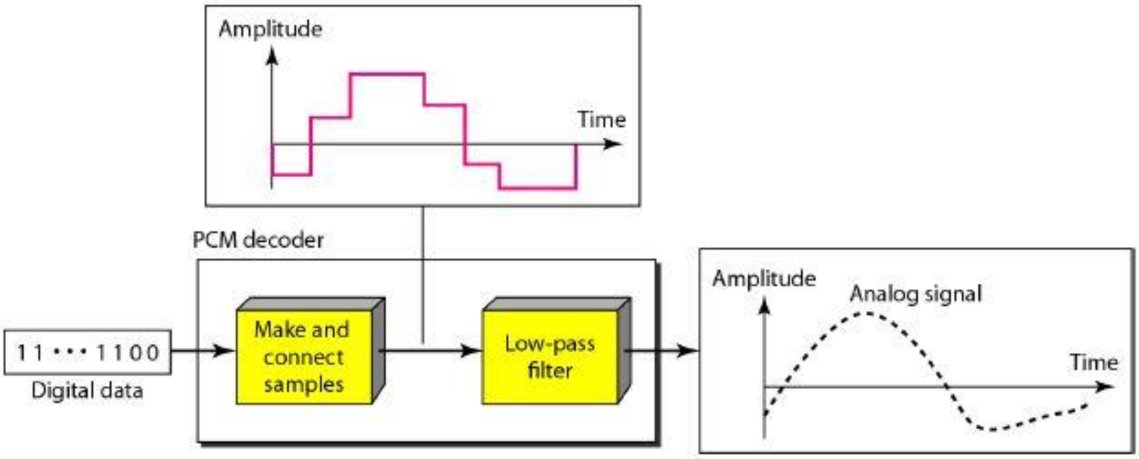
After the staircase signal is completed, it is passed through a low-pass filter to smooth the staircase signal into an analog signal.

The filter has the same cutoff frequency as the original signal at the sender.

If the signal has been sampled at (or greater than) the Nyquist sampling rate and if there are enough quantization levels, the original signal will be recreated.

The maximum and minimum values of the original signal can be achieved by using amplification.

Figure below shows the simplified process



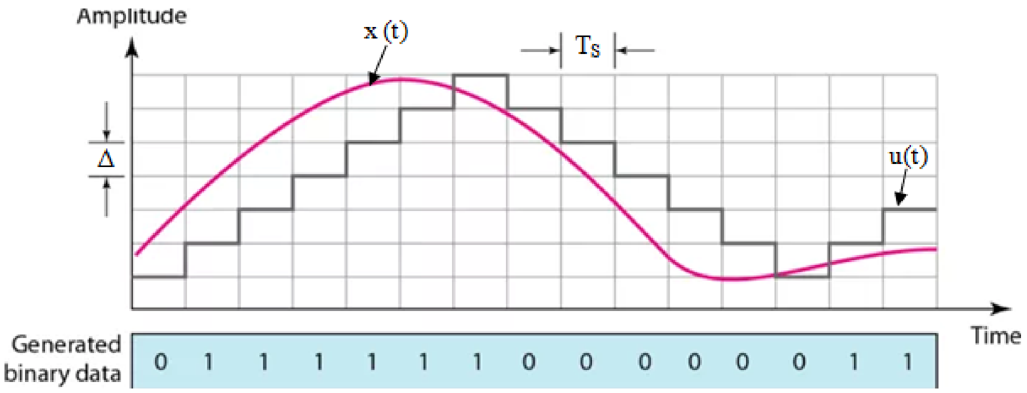
PCM Bandwidth

$B_{min}= n_b \times B_{analog}$

This means the minimum bandwidth of the digital signal is n_b times greater than the bandwidth of the analog signal. This is the price we pay for digitization.

Delta Modulation (DM)

- PCM is a very complex technique.
- Other techniques have been developed to reduce the complexity of PCM.
- The simplest is delta modulation.
- PCM finds the value of the signal amplitude for each sample
- DM finds the change from the previous sample.
- There are no code words here; bits are sent one after another.

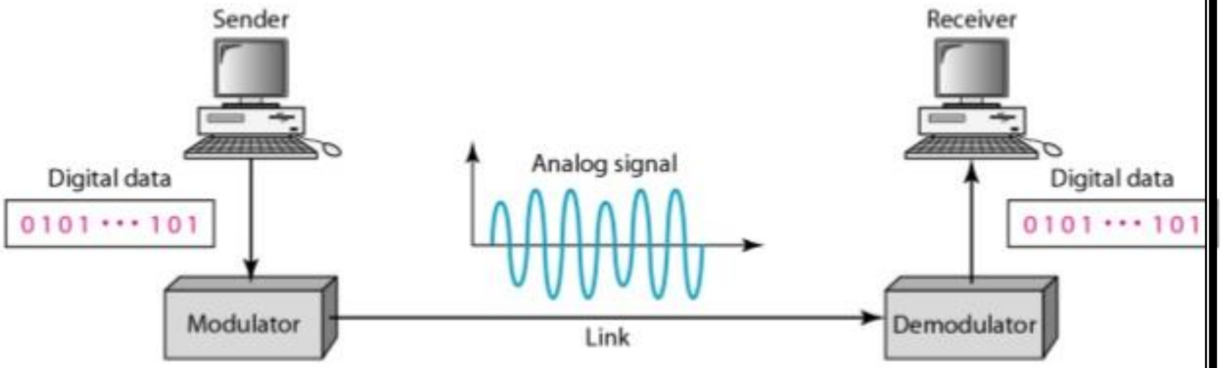


<u>Modulator</u>

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

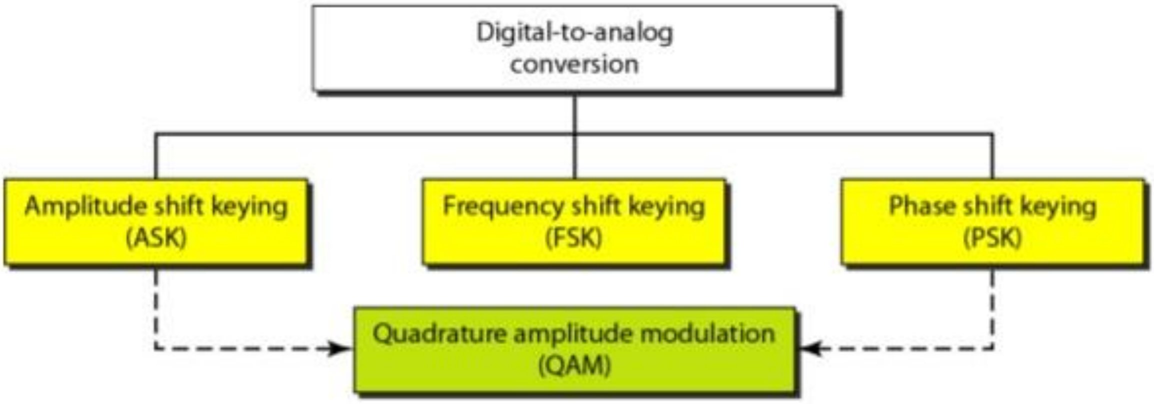
DIGITAL-TO-ANALOG CONVERSION

- Digital-to-analog conversion is the process of changing one of the characteristics of an analog signal based on the information in digital data.
- Figure below shows the relationship between the digital information, the digital-to-analog modulating process, and the resultant analog signal.



- A sine wave is defined by three characteristics: **amplitude, frequency, and phase**.
- When we vary anyone of these characteristics, we create a different version of that wave.
- So, by changing one characteristic of a simple electric signal, we can use it to represent digital data.
- Any of the three characteristics can be altered in this way, giving us at least three mechanisms for modulating digital data into an analog signal:

  Amplitude Shift Keying (ASK),
  Frequency Shift Keying (FSK),
  Phase Shift Keying (PSK).
- In addition, there is a fourth (and better) mechanism that combines changing both the amplitude and phase, called quadrature amplitude modulation (QAM).
- QAM is the most efficient of these options and is the mechanism commonly used today.


Digital-to-analog conversion
Amplitude shift keying (ASK)
Frequency shift keying (FSK)
Phase shift keying (PSK)
Quadrature amplitude modulation (QAM)

Aspects of Digital-to-Analog Conversion

Before we discuss specific methods of digital-to-analog modulation, two basic issues must be reviewed: bit and baud rates and the carrier signal.

Data Element Versus Signal Element

- We defined a data element as the smallest piece of information to be exchanged, the bit.
- We also defined a signal element as the smallest unit of a signal that is constant. Here the nature of the signal element is little bit different in analog transmission.

Data Rate Versus Signal Rate

We can define the data rate (bit rate) and the signal rate (baud rate) as we did for digital transmission. The relationship between them is

$S=N \times 1/r$ baud

where N is the data rate (bps) and r is the number of data elements carried in one signal element. The value of r in analog transmission is $r = \log_2 L$, where L is the type of signal element, not the level.

Discussion

An analog signal carries 4 bits per signal element. If 1000 signal elements are sent per second, find the bit rate.

In this case, r = 4, S = 1000, and N is unknown. We can find the value of N from

$N=S \times r=1000 \times 4=4000$ bps

An analog signal has a bit rate of 8000 bps and a baud rate of 1000 baud. How many data elements are carried by each signal element? How many signal elements do we need?

In this example, S =1000, N =8000, and L & r are unknown. We find first the value of r and then L.

S=N x 1/r baud

r=N/S

r=8000/1000=8

$r=\log_2(L)=8$

$L=2^r=2^8=256$

Carrier Signal

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

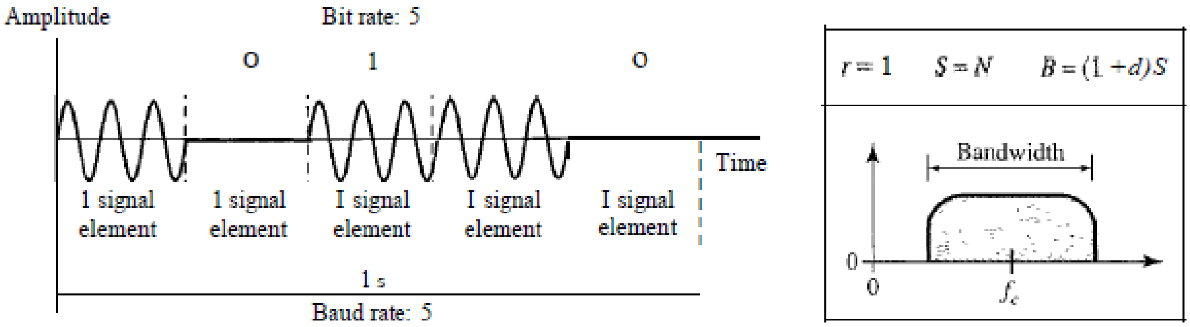
Amplitude Shift Keying

In amplitude shift keying, the amplitude of the carrier signal is varied to create signal elements. Both frequency and phase remain constant while the amplitude changes.

Binary ASK (BASK)

- Although we can have several levels (kinds) of signal elements, each with a different amplitude, ASK is normally implemented using only two levels.
- This is referred to as binary amplitude shift keying or on-off keying (OOK).
- The peak amplitude of one signal level is 0; the other is the same as the amplitude of the carrier frequency.

Figure below gives a conceptual view of binary ASK.



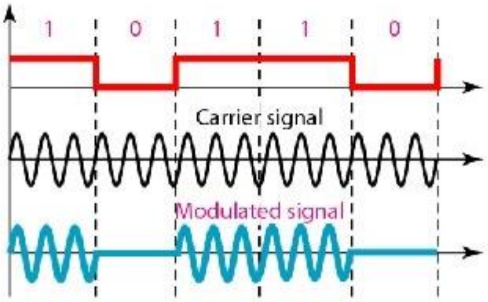
Bandwidth for ASK

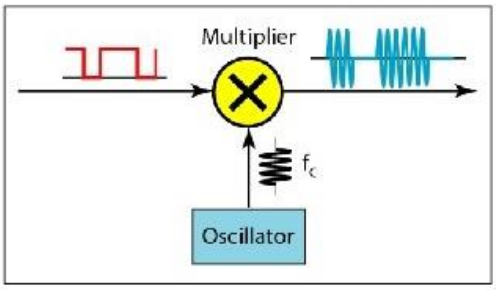
- Although the carrier signal is only one simple sine wave, the process of modulation produces a nonperiodic composite signal.
- As we expect, the bandwidth is proportional to the signal rate (baud rate).
- However, there is normally another factor involved, called d, which depends on the modulation and filtering process.
- The value of d is between 0 and 1. This means that the bandwidth can be expressed as shown, where S is the signal rate and the B is the bandwidth.

B =(1 +d) x S

Implementation

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





Discussion: We have an available bandwidth of 100 kHz which spans from 200 to 300 kHz. What are the carrier frequency and the bit rate if we modulate our data by using ASK with d =1?

The middle of the bandwidth is located at 250 kHz. This means that our carrier frequency can be at fc=250 kHz. We can use the formula for bandwidth to find the bit rate (with d =1 and r =1).

B =(1 +d) x S

B=2 x N x 1/r

B=2 x N x 1=100

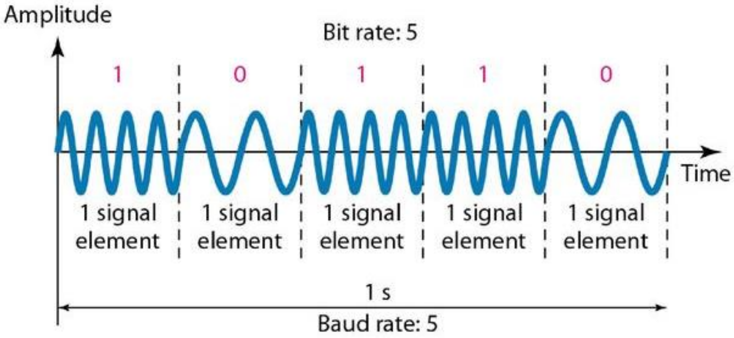
N=50kbps

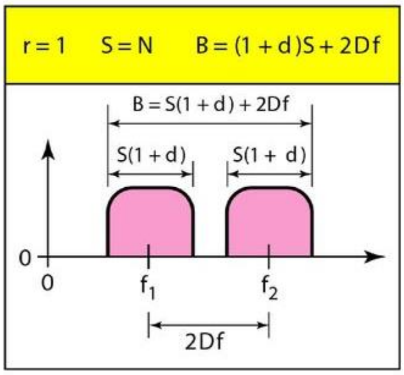
Frequency Shift Keying

- In frequency shift keying, the frequency of the carrier signal is varied to represent data.
- The frequency of the modulated signal is constant for the duration of one signal element, but changes for the next signal element if the data element changes.
- Both peak amplitude and phase remain constant for all signal elements.

Binary FSK (BFSK)

- One way to think about binary FSK (or BFSK) is to consider two carrier frequencies.
- In the below figure , we have selected two carrier frequencies, f1 and f2.
- We use the first carrier if the data element is 0; we use the second if the data element is 1.





As shown in the figure, the middle of one bandwidth is f_1 and the middle of the other is f_2. Both f_1 and f_2 are D_f apart from the midpoint between the two bands. The difference between the two frequencies is 2 D_f

Bandwidth for BFSK

- Again the carrier signals are only simple sine waves, but the modulation creates a nonperiodic composite signal with continuous frequencies.
- We can think of FSK as two ASK signals, each with its own carrier frequency f_1 or f_2. If the difference between the two frequencies is 2 D_f , then the required bandwidth is

$$\mathbf{B=(1+d) \times S + 2\,D_f}$$

Discussion:

We have an available bandwidth of 100 kHz which spans from 200 to 300 kHz. What should be the carrier frequency and the bit rate if we modulate our data by using FSK with d =1?

This problem is similar to the above problem, but we are modulating here by using FSK. The midpoint of the band is at 250 kHz. We choose 2 D_f to be 50 kHz; this means

$$B = (1 + d) \times S + 2D_f = 100$$

$$2 \times S = 50$$

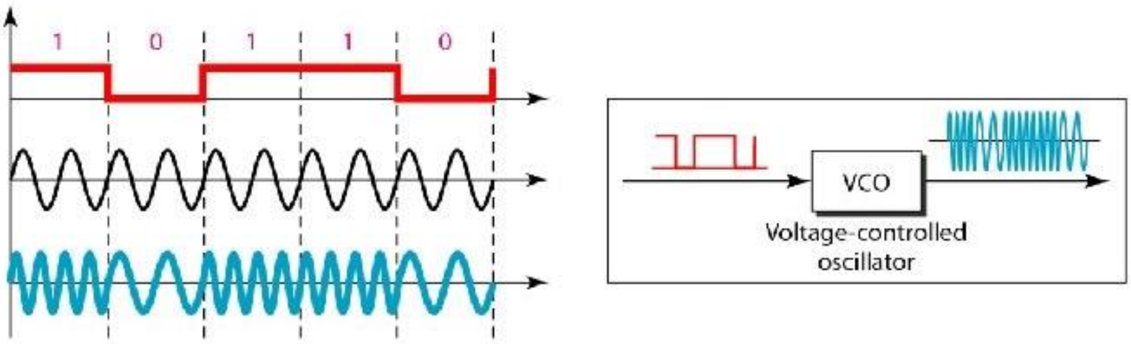
$$S = 25 \text{ kbaud}$$

$$N = 25 \text{ kbps}$$

Compared to the previous problem, we can see the bit rate for ASK is 50 kbps while the bit rate for FSK is 25 kbps.

Implementation

- There are two implementations of BFSK: noncoherent and coherent.
- In noncoherent BFSK, there may be discontinuity in the phase when one signal element ends and the next begins.
- In coherent BFSK, the phase continues through the boundary of two signal elements.
- Noncoherent BFSK can be implemented by treating BFSK as two ASK modulations and using two carrier frequencies.
- Coherent BFSK can be implemented by using one voltage-controlled oscillator (VCO) that changes its frequency according to the input voltage.
- Figure below shows the simplified idea behind the second implementation.
- The input to the oscillator is the unipolar NRZ signal. When the amplitude of NRZ is zero, the oscillator keeps its regular frequency; when the amplitude is positive, the frequency is increased.



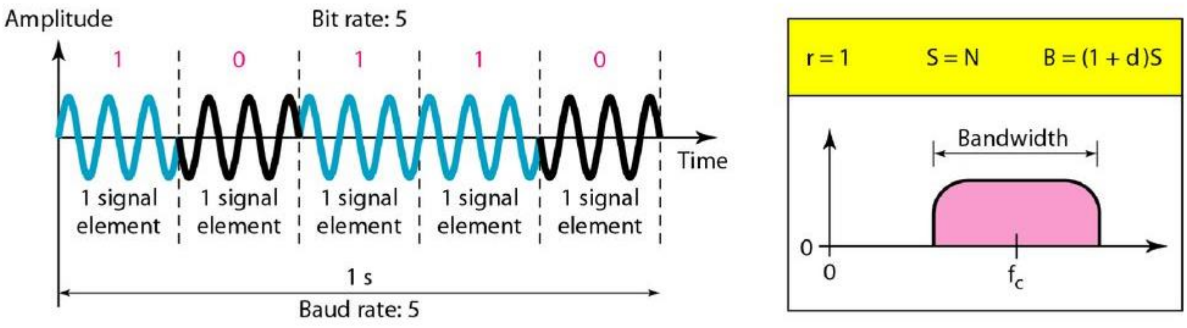
Phase Shift Keying

- In phase shift keying, the phase of the carrier is varied to represent two or more different signal elements.

- Both peak amplitude and frequency remain constant as the phase changes.
- Today, PSK is more common than ASK or FSK..

Binary PSK (BPSK)

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



Bandwidth

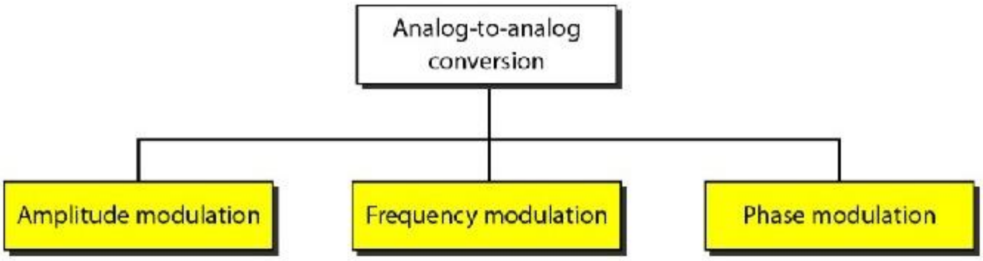
The bandwidth is the same as that for binary ASK, but less than that for BFSK. No bandwidth is wasted for separating two carrier signals.

Implementation

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

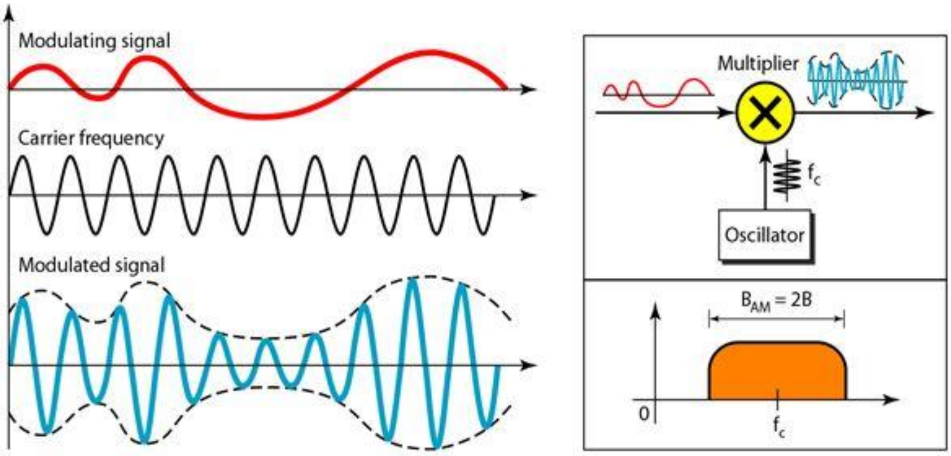
ANALOG-TO-ANALOG CONVERSION

- Analog-to-analog conversion, or analog modulation, is the representation of analog information by an analog signal.
- Analog-to-analog conversion can be accomplished in three ways: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM).
- FM and PM are usually categorized together.



Amplitude Modulation

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



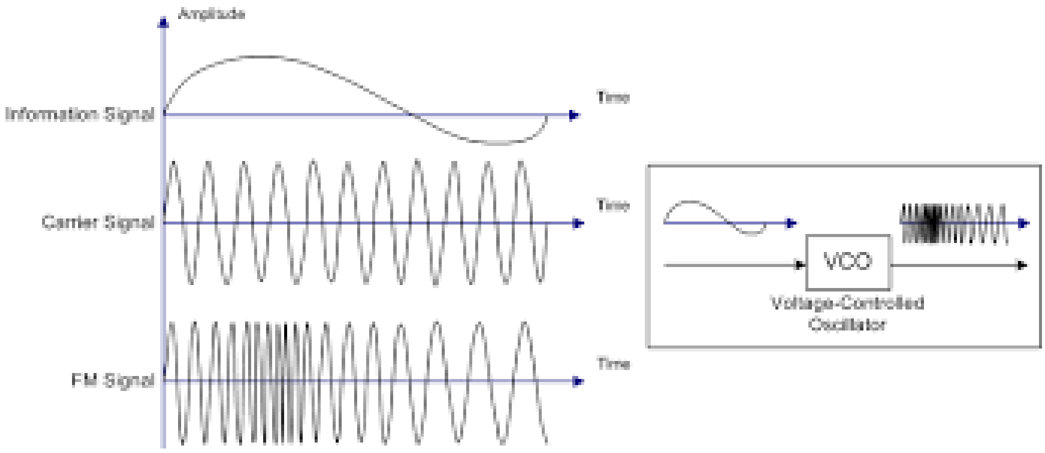
- AM is normally implemented by using a simple multiplier because the amplitude of the carrier signal needs to be changed according to the amplitude of the modulating signal.

AM Bandwidth

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

Frequency Modulation

- In FM transmission, the frequency of the carrier signal is modulated to follow the changing voltage level (amplitude) of the modulating signal.
- The peak amplitude and phase of the carrier signal remain constant, but as the amplitude of the information signal changes, the frequency of the carrier changes correspondingly.
- Figure below shows the relationships of the modulating signal, the carrier signal, and the resultant FM signal.
- FM is normally implemented by using a voltage-controlled oscillator as with FSK. The frequency of the oscillator changes according to the input voltage which is the amplitude of the modulating signal.



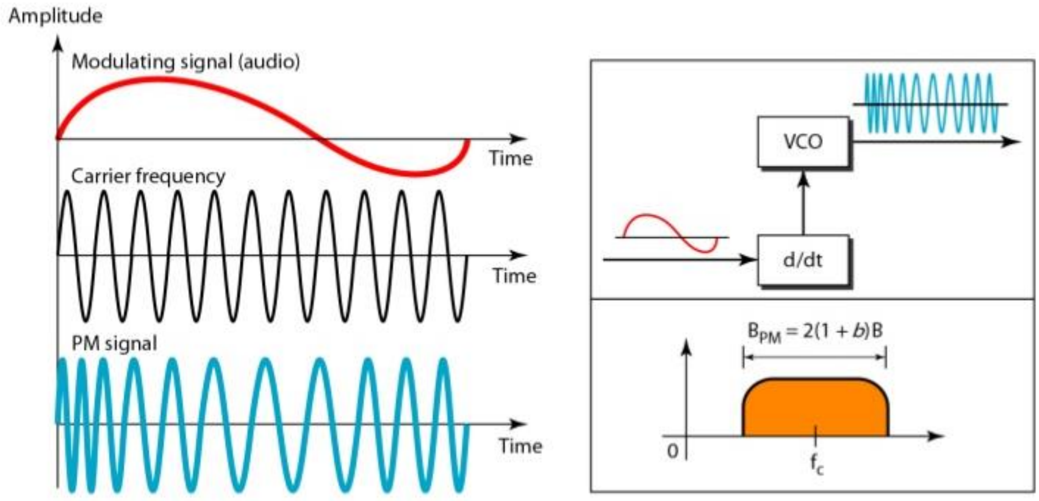
FM Bandwidth

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

Phase Moduiation

- In PM transmission, the phase of the carrier signal is modulated to follow the changing voltage level (amplitude) of the modulating signal.

- The peak amplitude and frequency of the carrier signal remain constant, but as the amplitude of the information signal changes, the phase of the carrier changes correspondingly.
- It can proved mathematically that PM is the same as FM with one difference. In FM, the instantaneous change in the carrier frequency is proportional to the amplitude of the modulating signal
- In PM the instantaneous change in the carrier frequency is proportional to the derivative of the amplitude of the modulating signal.
- Figure below shows the relationships of the modulating signal, the carrier signal, and the resultant PM signal.

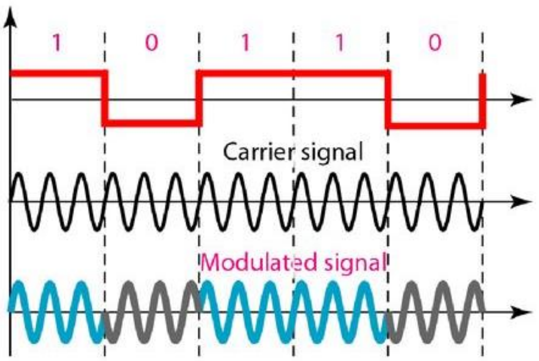


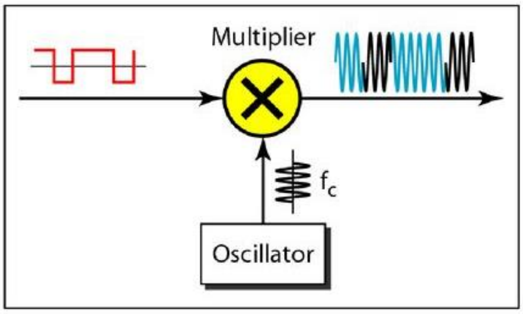
- PM is normally implemented by using a voltage-controlled oscillator along with a derivative.
- The frequency of the oscillator changes according to the derivative of the input voltage which is the amplitude of the modulating signal.

PM Bandwidth

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

$$BW=2(1 + \beta)B.$$

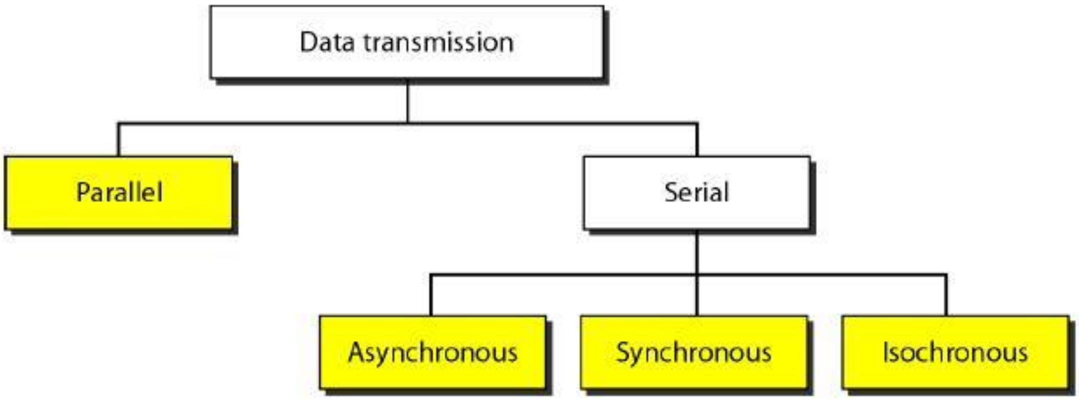

1
0
1
1
0
Carrier signal
Modulated signal


Multiplier
f_c
Oscillator

Unit-4. Data Communication & Data link control

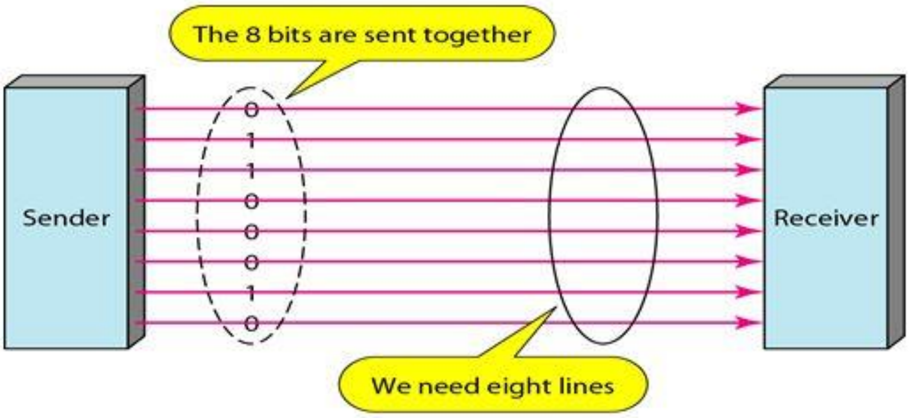
TRANSMISSION MODES

- The transmission of binary data across a link can be accomplished in either parallel or serial mode.
- In parallel mode, multiple bits are sent with each clock tick.
- In serial mode, 1 bit is sent with each clock tick.
- While there is only one way to send parallel data, there are three subclasses of serial transmission: **asynchronous, synchronous, and isochronous**



Parallel Transmission

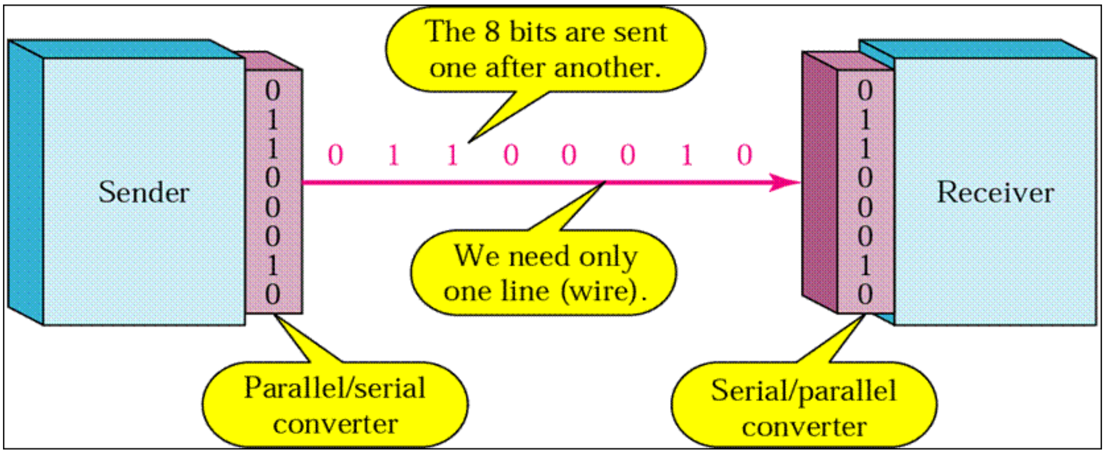
- Binary data, consisting of 1s and 0s, may be organized into groups of n bits each.
- By grouping, we can send data n bits at a time instead of 1. This is called parallel transmission.
- The mechanism for parallel transmission is a conceptually simple one: Use n wires to send n bits at one time. That way each bit has its own wire, and all n bits of one group can be transmitted with each clock tick from one device to another.
- Figure below shows how parallel transmission works for n =8. Typically, the eight wires are bundled in a cable with a connector at each end.
- The advantage of parallel transmission is **speed**. All else being equal, parallel transmission can increase the transfer speed by a factor of n over serial transmission.



- But there is a significant disadvantage: **cost.**
- Parallel transmission requires n communication lines (wires in the example) just to transmit the data stream. Because this is expensive, parallel transmission is usually limited to short distances.

Serial Transmission

- In serial transmission one bit follows another, so we need only one communication channel rather than n to transmit data between two communicating devices

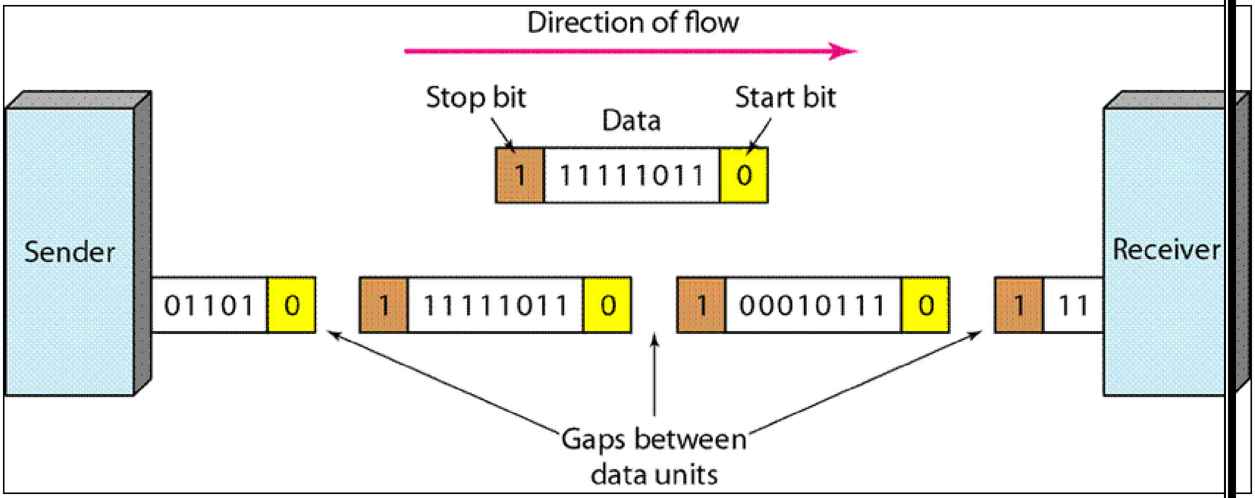

The 8 bits are sent one after another.
Sender
0 1 1 0 0 0 1 0
0 1 1 0 0 0 1 0
0 1 1 0 0 0 1 0
Receiver
We need only one line (wire).
Parallel/serial converter
Serial/parallel converter

- The advantage of serial over parallel transmission is that with only one communication channel, serial transmission reduces the cost of transmission over parallel by roughly a factor of n.
- Since communication within devices is parallel, conversion devices are required at the interface between the sender and the line (parallel-to-serial) and between the line and the receiver (serial-to-parallel). Serial transmission occurs in one of three ways: **asynchronous, synchronous, and isochronous.**

Asynchronous Transmission

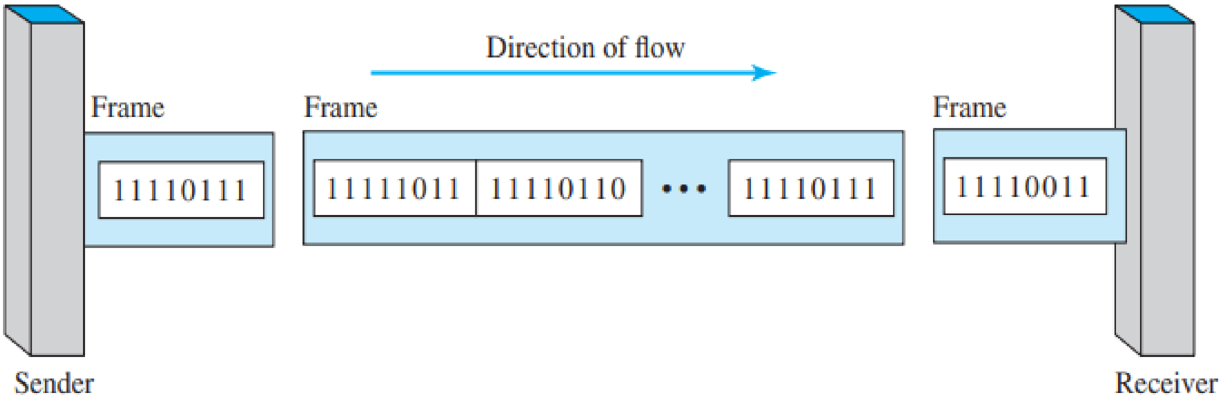
- Asynchronous transmission is so named because the timing of a signal is unimportant.
- Instead, information is received and translated by agreed upon patterns.
- As long as those patterns are followed, the receiving device can retrieve the information without regard to the rhythm in which it is sent.
- Patterns are based on grouping the bit stream into bytes.
- Each group, usually 8 bits, is sent along the link as a unit.
- The sending system handles each group independently, relaying it to the link whenever ready, without regard to a timer.
- Without synchronization, the receiver cannot use timing to predict when the next group will arrive.
- To alert the receiver to the arrival of a new group, therefore, an extra bit is added to the beginning of each byte. This bit, usually a 0, is called the start bit. To let the receiver know that the byte is finished, 1 or more additional bits are appended to the end of the byte. These bits, usually 1s, are called stop bits.

- By this method, each byte is increased in size to at least 10 bits, of which 8 bits is information and 2 bits or more are signals to the receiver.
- In addition, the transmission of each byte may then be followed by a gap of varying duration. This gap can be represented either by an idle channel or by a stream of additional stop bits.
- Figure below is a schematic illustration of asynchronous transmission.
- In this example, the start bits are 0s, the stop bits are 1s, and the gap is represented by an idle line rather than by additional stop bits.
- The addition of stop and start bits and the insertion of gaps into the bit stream make asynchronous transmission slower than forms of transmission that can operate without the addition of control information.
- But it is cheap and effective, two advantages that make it an attractive choice for situations such as low-speed communication. For example, the connection of a keyboard to a computer is a natural application for asynchronous transmission. A user types only one character at a time, types extremely slowly in data processing terms, and leaves unpredictable gaps of time between each character.



Synchronous Transmission

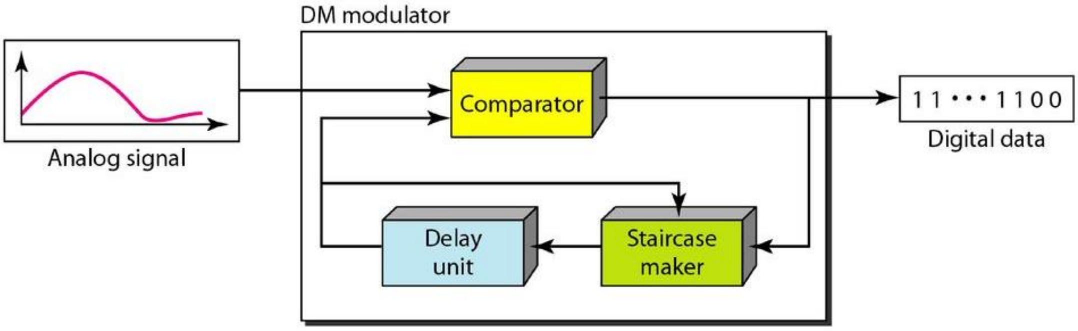
- In synchronous transmission, the bit stream is combined into longer "frames," which may contain multiple bytes.
- Each byte, however, is introduced onto the transmission link without a gap between it and the next one. It is left to the receiver to separate the bit stream into bytes for decoding purposes.
- In other words, data are transmitted as an unbroken string of 1s and 0s, and the receiver separates that string into the bytes, or characters, it needs to reconstruct the information.
- Figure below gives a schematic illustration of synchronous transmission.
- We have drawn in the divisions between bytes. In reality, those divisions do not exist; the sender puts its data onto the line as one long string.
- If the sender wishes to send data in separate bursts, the gaps between bursts must be filled with a special sequence of 0s and 1s that means idle. The receiver counts the bits as they arrive and groups them in 8-bit units.



- Without gaps and start and stop bits, there is no built-in mechanism to help the receiving device adjust its bit synchronization midstream.
- Timing becomes very important, therefore, because the accuracy of the received information is completely dependent on the ability of the receiving device to keep an accurate count of the bits as they come in.
- The advantage of synchronous transmission is speed.
- With no extra bits or gaps to introduce at the sending end and remove at the receiving end, and, by extension, with fewer bits to move across the link, synchronous transmission is faster than asynchronous transmission.
- For this reason, it is more useful for high-speed applications such as the transmission of data from one computer to another. Byte synchronization is accomplished in the data link layer. We need to emphasize one point here. Although there is no gap between characters in synchronous serial transmission, there may be uneven gaps between frames.

Isochronous

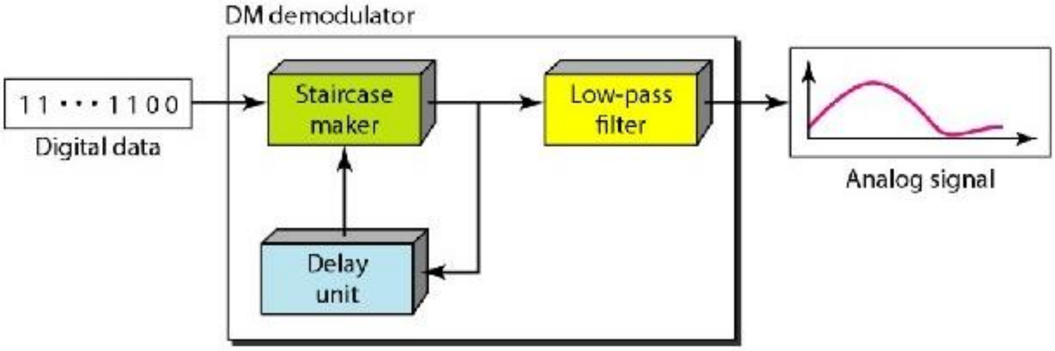
- In real-time audio and video, in which uneven delays between frames are not acceptable, synchronous transmission fails.
- For example, TV images are broadcast at the rate of 30 images per second; they must be viewed at the same rate.
- If each image is sent by using one or more frames, there should be no delays between frames.
- For this type of application, synchronization between characters is not enough; the entire stream of bits must be synchronized.
- The isochronous transmission guarantees that the data arrive at a fixed rate.



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

Demodulator

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

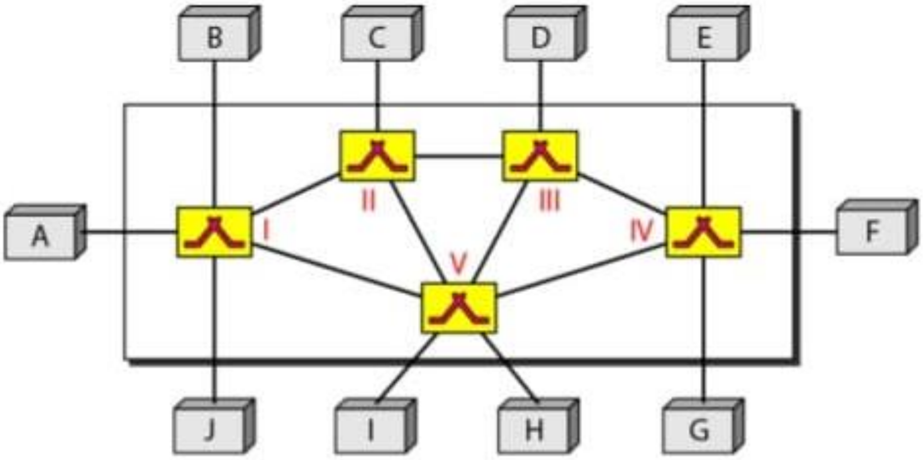

DM demodulator
11 ··· 1100
Digital data
Staircase maker
Low-pass filter
Delay unit
Analog signal

Unit-5. Switching & Routing

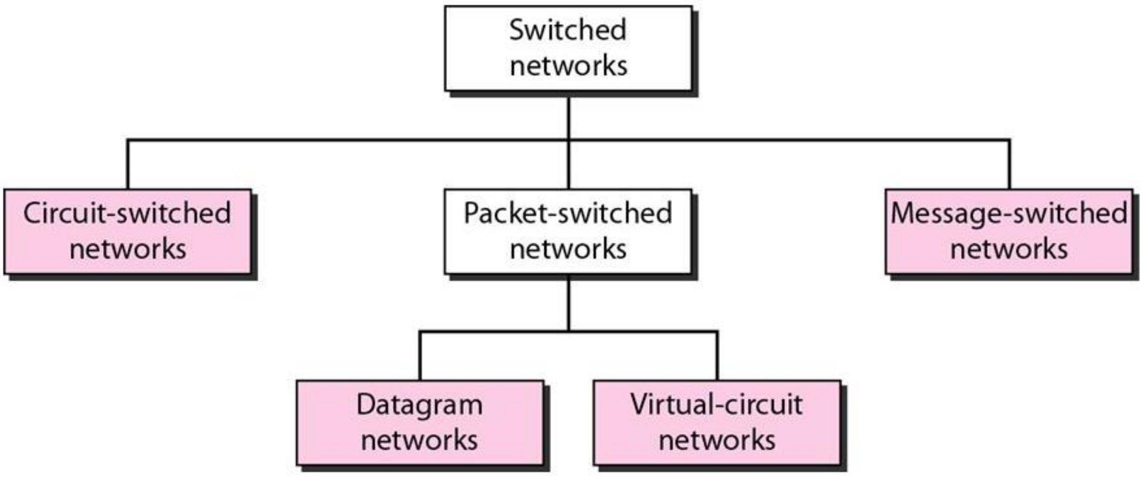
Switching

A network is a set of connected devices.

- Whenever we have multiple devices, we have the problem of how to connect them to make one-to-one communication possible.
- One solution is to make a point-to-point connection between each pair of devices (a mesh topology) or between a central device and every other device (a star topology).
- These methods, however, are impractical and wasteful when applied to very large networks. The number and length of the links require too much infrastructure to be cost-efficient, and the majority of those links would be idle most of the time.
- Other topologies employing multipoint connections, such as a bus, are ruled out because the distances between devices and the total number of devices increase beyond the capacities of the media and equipment.
- A better solution is switching.
- A switched network consists of a series of interlinked nodes, called switches.
- **Switches are devices capable of creating temporary connections between two or more devices linked to the switch.** In a switched network, some of these nodes are connected to the end systems (computers or telephones, for example). Others are used only for routing. Figure below shows a switched network.

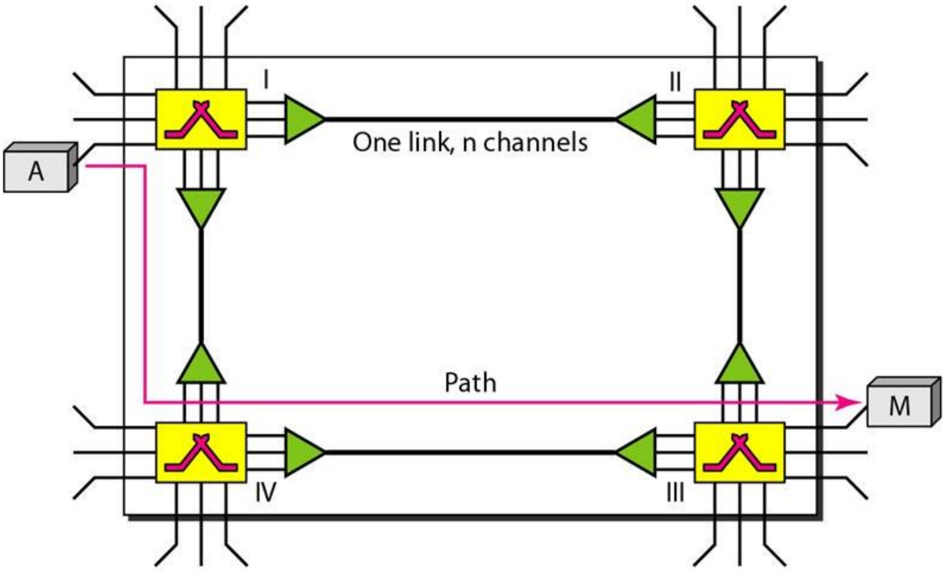


The end systems (communicating devices) are labeled A, B, C, D, and so on, and the switches are labeled I, II, III, IV, and V. Each switch is connected to multiple links.



CIRCUIT-SWITCHED NETWORKS

- A circuit-switched network consists of a set of switches connected by physical links.
- A connection between two stations is a dedicated path made of one or more links. However, each connection uses only one dedicated channel on each link.
- Each link is normally divided into n channels by using FDM or TDM .
- Figure below shows a trivial circuit-switched network with four switches and four links. Each link is divided into n (n is 3 in the figure) channels by using FDM or TDM.



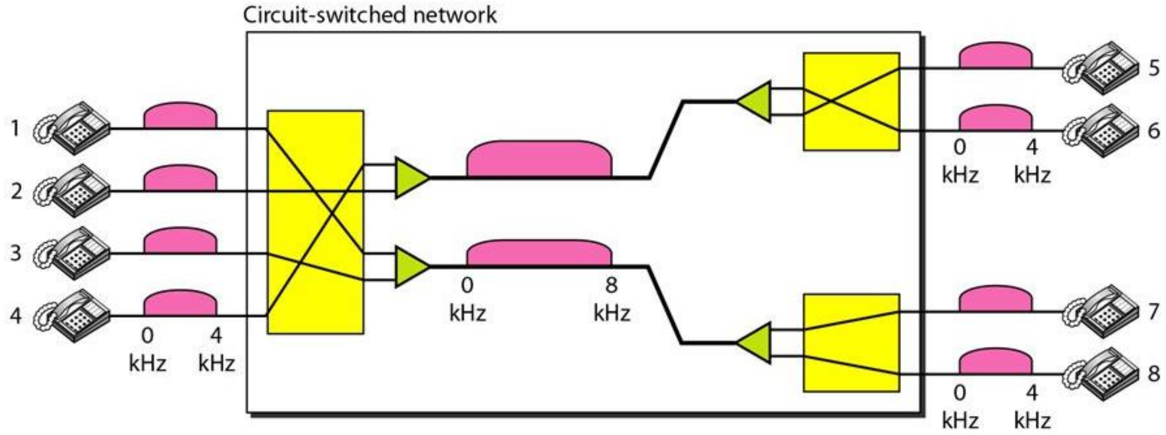
- The multiplexing symbols is shown to emphasize the division of the link into channels even though multiplexing can be implicitly included in the switch fabric.
- The end systems, such as computers or telephones, are directly connected to a switch.
- Only two end systems are shown for simplicity.

- When end system A needs to communicate with end system M, system A needs to request a connection to M that must be accepted by all switches as well as by M itself. This is called the setup phase; a circuit (channel) is reserved on each link, and the combination of circuits or channels defines the dedicated path. After the dedicated path made of connected circuits (channels) is established, data transfer can take place. After all data have been transferred, the circuits are tom down.

We need to emphasize several points here:

- Circuit switching takes place at the physical layer.
- Before starting communication, the stations must make a reservation for the resources to be used during the communication.
- These resources, such as channels (bandwidth in FDM and time slots in TDM), switch buffers, switch processing time, and switch input/output ports, must remain dedicated during the entire duration of data transfer until the teardown phase.
- Data transferred between the two stations are not packetized (physical layer transfer of the signal).
- The data are a continuous flow sent by the source station and received by the destination station, although there may be periods of silence.
- There is no addressing involved during data transfer.
- The switches route the data based on their occupied band (FDM) or time slot (TDM). Of course, there is end-to end addressing used during the setup phase, as we will see shortly.

As a trivial example, let us use a circuit-switched network to connect eight telephones in a small area. Communication is through 4-kHz voice channels. We assume that each link uses FDM to connect a maximum of two voice channels. The bandwidth of each link is then 8 kHz. Figure 8.4 shows the situation. Telephone 1 is connected to telephone 7; 2 to 5; 3 to 8; and 4 to 6. Ofcourse the situation may change when new connections are made. The switch controls the connections.



Three Phases

The actual communication in a circuit-switched network requires three phases:

- **connection setup**
- **data transfer**
- **connection teardown**

Setup Phase

- Before the two parties (or multiple parties in a conference call) can communicate, a dedicated circuit (combination of channels in links) needs to be established.
- The end systems are normally connected through dedicated lines to the switches, so connection setup means creating dedicated channels between the switches.
- For example, in the above figure, when system A needs to connect to system M, it sends a setup request that includes the address of system M, to switch I. Switch I finds a channel between itself and switch IV that can be dedicated for this purpose. Switch I then sends the request to switch IV, which finds a dedicated channel between itself and switch III. Switch III informs system M of system A's intention at this time. In the next step to making a connection, an acknowledgment from system M needs to be sent in the opposite direction to system A. Only after system A receives this acknowledgment is the connection established.
- Note that end-to-end addressing is required for creating a connection between the two end systems. These can be, for example, the addresses of the computers assigned by the administrator in a TDM network, or telephone numbers in an FDM network.

Data Transfer Phase

After the establishment of the dedicated circuit (channels), the two parties can transfer data.

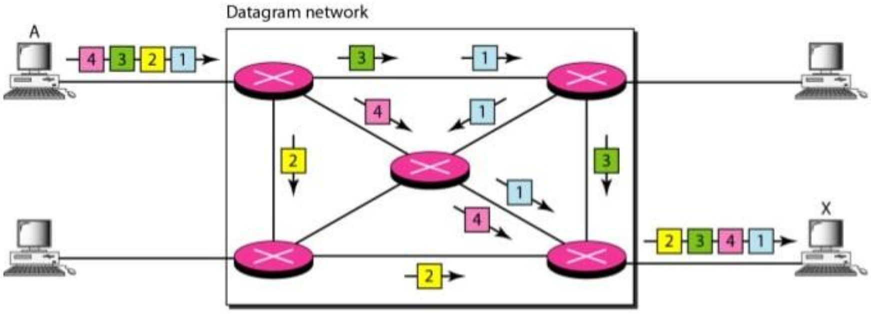
Teardown Phase

When one of the parties needs to disconnect, a signal is sent to each switch to release the resources.

DATAGRAM NETWORKS

- In data communications, we need to send messages from one end system to another.
- If the message is going to pass through a packet-switched network, it needs to be divided into packets of fixed or variable size.
- The size of the packet is determined by the network and the governing protocol.
- In packet switching, there is no resource allocation for a packet.
- This means that there is no reserved bandwidth on the links, and there is no scheduled processing time for each packet.
- Resources are allocated on demand. The allocation is done on a firstcome, first-served basis.

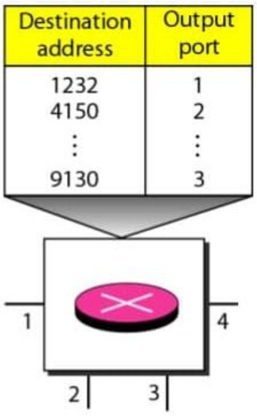
- When a switch receives a packet, no matter what is the source or destination, the packet must wait if there are other packets being processed. As with other systems in our daily life, this lack of reservation may create delay.
- In a datagram network, each packet is treated independently of all others.
- Even if a packet is part of a multi packet transmission, the network treats it as though it existed alone.
- Packets in this approach are referred to as datagrams.
- Datagram switching is normally done at the network layer.
- Figure below shows how the datagram approach is used to deliver four packets from station A to station X.
- The switches in a datagram network are traditionally referred to as routers.
- That is why we use a different symbol for the switches in the figure.



- In this example, all four packets (or datagrams) belong to the same message, but may travel different paths to reach their destination. This is so because the links may be involved in carrying packets from other sources and do not have the necessary bandwidth available to carry all the packets from A to X. This approach can cause the datagrams of a transmission to arrive at their destination out of order with different delays between the packets. Packets may also be lost or dropped because of a lack of resources. In most protocols, it is the responsibility of an upper-layer protocol to reorder the datagrams or ask for lost datagrams before passing them on to the application.
- The datagram networks are sometimes referred to as connectionless networks. The term connectionless here means that the switch (packet switch) does not keep information about the connection state. There are no setup or teardown phases. Each packet is treated the same by a switch regardless of its source or destination.

Routing Table

If there are no setup or teardown phases, how are the packets routed to their destinations in a datagram network? In this type of network, each switch (or packet switch) has a routing table which is based on the destination address. The routing tables are dynamic and are updated periodically. The destination addresses and the corresponding forwarding output ports are recorded in the tables. This is different from the table of a circuit switched network in which each entry is created when the setup phase is completed and deleted when the teardown phase is over. Figure below shows the routing table for a switch.



Destination address	Output port
1232	1
4150	2
⋮	⋮
9130	3

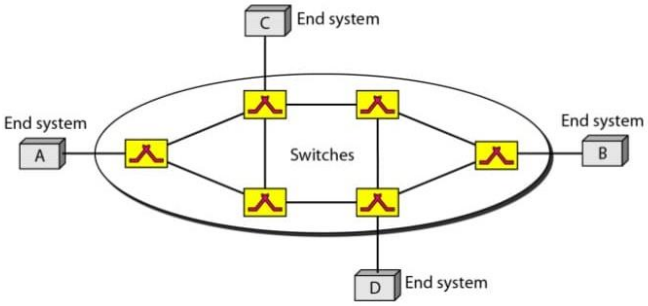
Destination Address

Every packet in a datagram network carries a header that contains, among other information, the destination address of the packet. When the switch receives the packet, this destination address is examined; the routing table is consulted to find the corresponding port through which the packet should be forwarded. This address, unlike the address in a virtual-circuit-switched network, remains the same during the entire journey of the packet.

VIRTUAL-CIRCUIT NETWORKS

A virtual-circuit network is a cross between a circuit-switched network and a datagram network. It has some characteristics of both.
As in a circuit-switched network, there are setup and teardown phases in addition to the data transfer phase.
Resources can be allocated during the setup phase, as in a circuit-switched network, or on demand, as in a datagram network.
As in a datagram network, data are packetized and each packet carries an address in the header. However, the address in the header has local jurisdiction, not end-to-end jurisdiction.
As in a circuit-switched network, all packets follow the same path established during the connection.
A virtual-circuit network is normally implemented in the data link layer, while a circuit-switched network is implemented in the physical layer and a datagram network in the network layer.

Figure below is an example of a virtual-circuit network. The network has switches that allow traffic from sources to destinations. A source or destination can be a computer, packet switch, bridge, or any other device that connects other networks.



Addressing

In a virtual-circuit network, two types of addressing are involved: global and local (virtual-circuit identifier). Global Addressing

A source or a destination needs to have a global address-an address that can be unique in the scope of the network or internationally if the network is part of an international network. A global address in virtual-circuit networks is used only to create a virtual-circuit identifier.

Virtual-Circuit Identifier

The identifier that is actually used for data transfer is called the virtual-circuit identifier (VCl). A VCI, unlike a global address, is a small number that has only switch scope; it is used by a frame between two switches. When a frame arrives at a switch, it has a VCI; when it leaves, it has a different VCl.

X.25 An ITU-T standard that defines the interface between a data terminal device and a packet-switching network

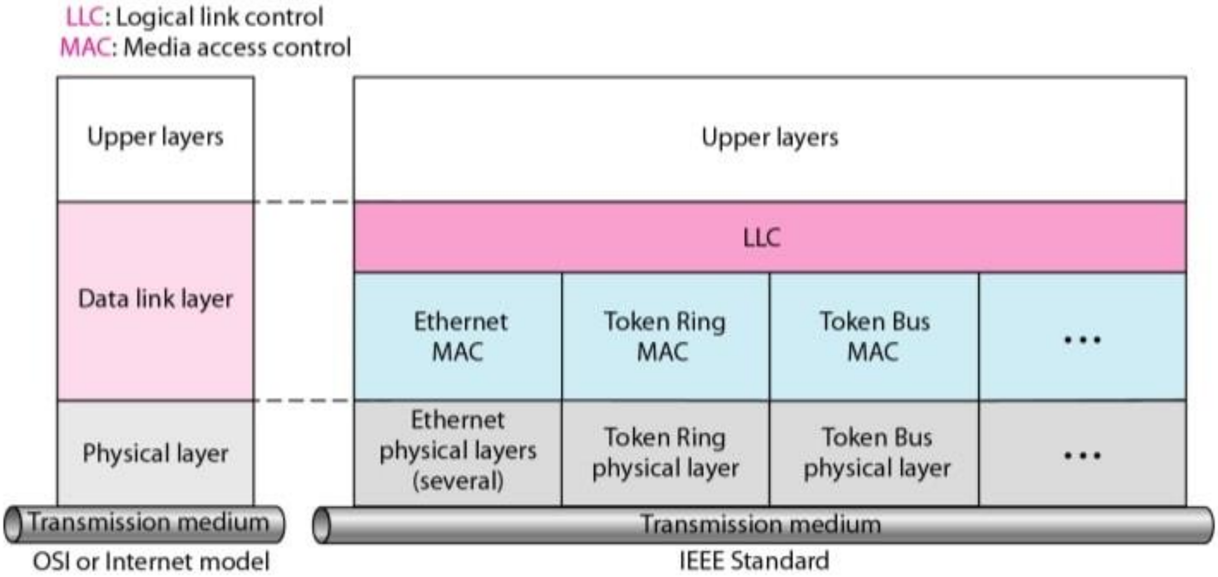
Unit-6. LAN Technology

Wired LANs: Ethernet

- Local area network (LAN) is a computer network that is designed for a limited geographic area such as a building or a campus.
- Although a LAN can be used as an isolated network to connect computers in an organization for the sole purpose of sharing resources, most LANs today are also linked to a wide area network (WAN) or the Internet.

IEEE STANDARDS

- In 1985, the Computer Society of the IEEE started a project, called Project 802, to set standards to enable intercommunication among equipment from a variety of manufacturers.
- Project 802 does not seek to replace any part of the OSI or the Internet model.
- Instead, it is a way of specifying functions of the physical layer and the data link layer of major LAN protocols.
- The standard was adopted by the American National Standards Institute (ANSI).
- In 1987, the International Organization for Standardization (ISO) also approved it as an international standard under the designation ISO 8802.
- The relationship of the 802 Standard to the traditional OSI model is shown in below figure .
- The IEEE has subdivided the data link layer into two sublayers: logical link control (LLC) and media access control (MAC).
- IEEE has also created several physical layer standards for different LAN protocols.



Logical Link Control (LLC)

- We know that data link control handles framing, flow control, and error control.
- In IEEE Project 802, flow control, error control, and part of the framing duties are collected into one sublayer called the logical link control.
- Framing is handled in both the LLC sublayer and the MAC sublayer.
- The LLC provides one single data link control protocol for all IEEE LANs.
- In this way, the LLC is different from the media access control sublayer, which provides different protocols for different LANs.
- A single LLC protocol can provide interconnectivity between different LANs because it makes the **MAC sublayer transparent.**

Need for LLC

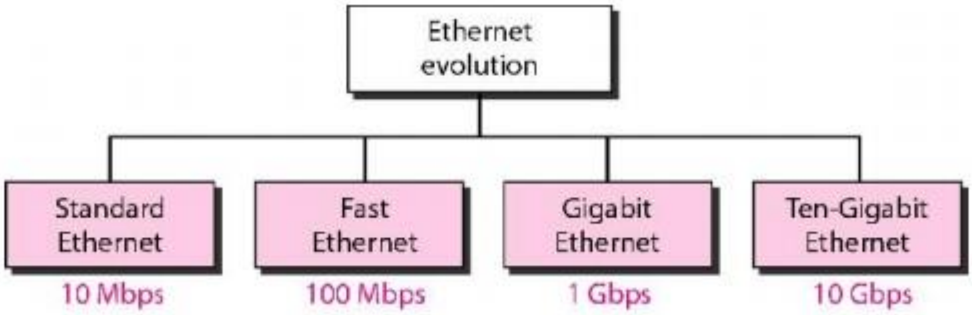
- The purpose of the LLC is to provide flow and error control for the upper-layer protocols that actually demand these services.

Media Access Control (MAC)

- IEEE Project 802 has created a sublayer called media access control that defines the specific access method for each LAN. For example, it defines CSMA/CD as the media access method for Ethernet LANs and the token passing method for Token Ring and Token Bus LANs.
- As we know, part of the framing function is also handled by the MAC layer.
- In contrast to the LLC sublayer, the MAC sublayer contains a number of distinct modules; each defines the access method and the framing format specific to the corresponding LAN protocol

STANDARD ETHERNET

- The original Ethernet was created in 1976 at Xerox's Palo Alto Research Center (PARC).
- Since then, it has gone through four generations: Standard Ethernet (10 Mbps), Fast Ethernet (100 Mbps), Gigabit Ethernet (1 Gbps), and Ten-Gigabit Ethernet (10 Gbps).



Frame Format

- The Ethernet frame contains seven fields: preamble, SFD, DA, SA, length or type of protocol data unit (PDU), upper-layer data, and the CRe. Ethernet does not provide any mechanism for acknowledging received frames, making it what is known as an unreliable medium. Acknowledgments must be implemented at the higher layers. The format ofthe MAC frame is shown in Figure 13.4.

Preamble. The first field of the 802.3 frame contains 7 bytes (56 bits) of alternating Os and Is that alerts the receiving system to the coming frame and enables it to synchronize its input timing. The pattern provides only an alert and a timing pulse. The 56-bit pattern allows the stations to miss some bits at the beginning of the frame. The preamble is actually added at the physical layer and is not (formally) part of the frame.

Start frame delimiter (SFD). The second field (l byte: 10101011) signals the beginning of the frame. The SFD warns the station or stations that this is the last chance for synchronization. The last 2 bits is 11 and alerts the receiver that the next field is the destination address.

Destination address (DA). The DA field is 6 bytes and contains the physical address ofthe destination **station or stations to receive the packet. We will discuss addressing shortly.**

Source address (SA). The SA field is also 6 bytes and contains the physical address ofthe sender ofthe packet. We will discuss addressing shortly.

Length or type. This field is defined as a type field or length field. The original Ethernet used this field as the type field to define the upper-layer protocol using the MAC frame. The IEEE standard used it as the length field to define the number of bytes in the data field. Both uses are common today.

Data. This field carries data encapsulated from the upper-layer protocols. It is a minimum of 46 and a maximum of 1500 bytes, as we will see later.

CRC. The last field contains error detection information, in this case a CRC-32.

Addressing

- Each station on an Ethernet network (such as a PC, workstation, or printer) has its own network interface card (NIC).
- The NIC fits inside the station and provides the station with a 6-byte physical address.
- As shown below , the Ethernet address is 6 bytes (48 bits), normally written in hexadecimal notation, with a colon between the bytes.

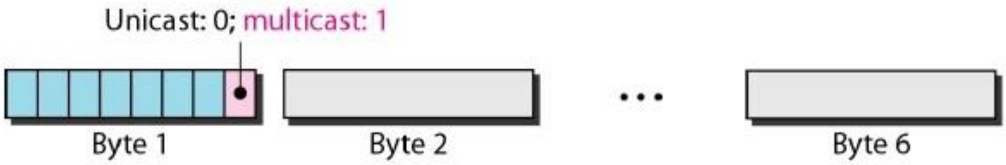
e.g. **06:01:02:01:2C:4B**

6 bytes =12 hex digits =48 bits

Unicast, Multicast, and Broadcast Addresses

- A source address is always a unicast address-the frame comes from only one station.
- The destination address, however, can be unicast, multicast, or broadcast.

- Figure below shows how to distinguish a unicast address from a multicast address.
- If the least significant bit of the first byte in a destination address is 0, the address is unicast; otherwise, it is multicast.



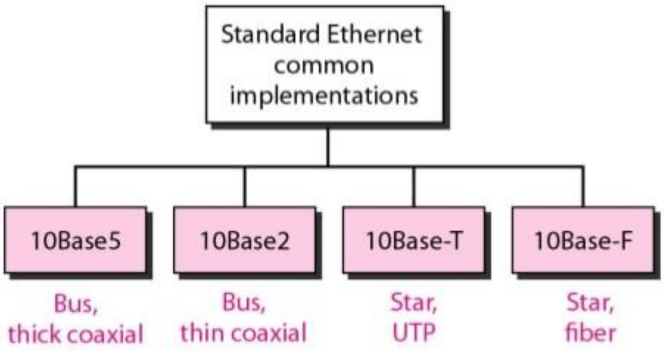
- **A unicast** destination address defines only one recipient; the relationship between the sender and the receiver is one-to-one.
- **A multicast** destination address defines a group of addresses; the relationship between the sender and the receivers is one-to-many.
- The **broadcast** address is a special case of the multicast address; the recipients are all the stations on the LAN.

Define the type of the following destination addresses:

a. 4A:30:10:21:10:1A
b. 47:20:1B:2E:08:EE
c. FF:FF:FF:FF:FF:FF
To find the type of the address, we need to look at the second hexadecimal digit from the left.

If it is even, the address is unicast.
If it is odd, the address is multicast.
If all digits are F's, the address is broadcast.
Therefore, we have the following:
a. This is a unicast address because A in binary is 1010 (even).
b. This is a multicast address because 7 in binary is 0111 (odd).
c. This is a broadcast address because all digits are F's.

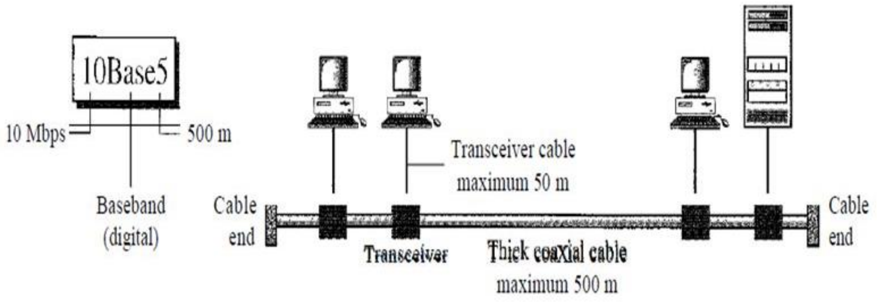
Physical Layer
The Standard Ethernet defines several physical layer implementations; four of the most common, are shown in below figure



.

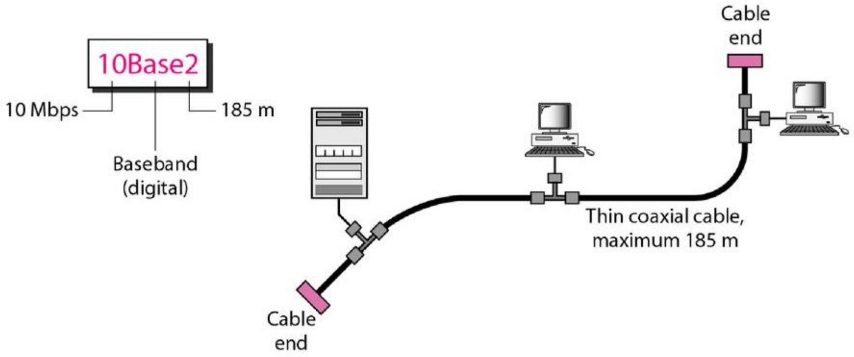
10Base5: Thick Ethernet

- The first implementation is called 10Base5, thick Ethernet, or Thicknet.
- The nickname derives from the size of the cable, which is roughly the size of a garden hose and too stiff to bend with your hands.
- 10Base5 was the first Ethernet specification to use a bus topology with an external transceiver (transmitter/receiver) connected via a tap to a thick coaxial cable.
- Figure below shows a schematic diagram of a 10Base5 implementation.



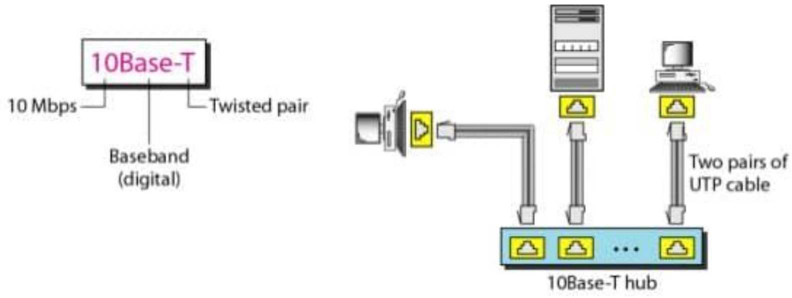
10Base2: Thin Ethernet

- The second implementation is called 10Base2, thin Ethernet, or Cheapernet.
- 10Base2 also uses a bus topology, but the cable is much thinner and more flexible.
- The cable can be bent to pass very close to the stations. In this case, the transceiver is normally part of the network interface card (NIC), which is installed inside the station.
- Figure below shows the schematic diagram of a 10Base2 implementation.



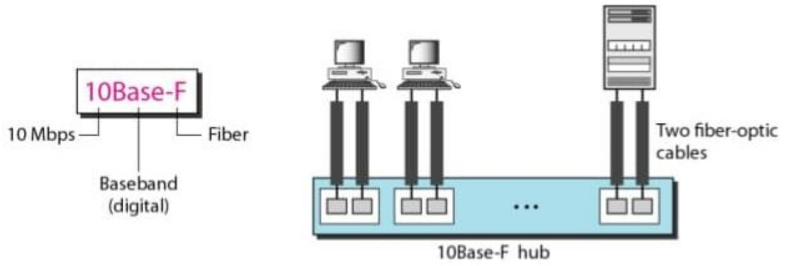
10Base-T: Twisted-Pair Ethernet

- The third implementation is called 10Base-T or twisted-pair Ethernet.
- 10Base-T uses a physical star topology.
- The stations are connected to a hub via two pairs of twisted cable, as shown in the figure.
- The two pairs of twisted cable create two paths (one for sending and one for receiving) between the station and the hub. Any collision here happens in the hub.
- Compared to 10Base5 or 10Base2, we can see that the hub actually replaces the coaxial cable as far as a collision is concerned. The maximum length of the twisted cable here is defined as 100 m, to minimize the effect of attenuation in the twisted cable.

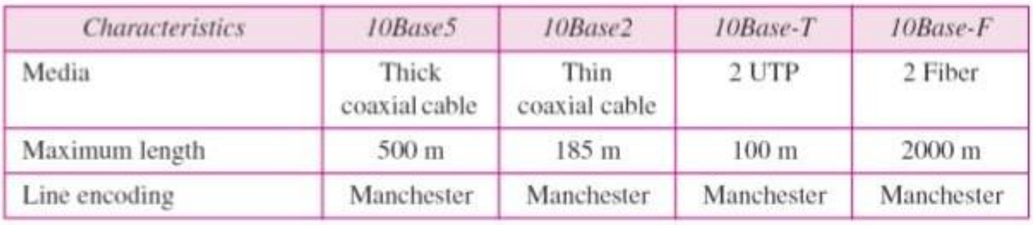


10Base-F: Fiber Ethernet

- Although there are several types of optical fiber 10-Mbps Ethernet, the most common is called 10Base-F.
- 10Base-F uses a star topology to connect stations to a hub.
- The stations are connected to the hub using two fiber-optic cables, as shown in figure.



Comparison between the Ethernet standards



Characteristics	*10Base5*	*10Base2*	*10Base-T*	*10Base-F*
Media	Thick coaxial cable	Thin coaxial cable	2 UTP	2 Fiber
Maximum length	500 m	185 m	100 m	2000 m
Line encoding	Manchester	Manchester	Manchester	Manchester

FAST ETHERNET

- Fast Ethernet was designed to compete with LAN protocols such as FDDI or Fiber Channel (or Fibre Channel, as it is sometimes spelled).
- IEEE created Fast Ethernet under the name 802.3u.
- Fast Ethernet is backward-compatible with Standard Ethernet, but it can transmit data 10 times faster at a rate of 100 Mbps.

The goals of Fast Ethernet can be summarized as follows:

1. Upgrade the data rate to 100 Mbps.
2. Make it compatible with Standard Ethernet.
3. Keep the same 48-bit address.
4. Keep the same frame format.
5. Keep the same minimum and maximum frame lengths.

Autonegotiation

A new feature added to Fast Ethernet is called autonegotiation. It allows a station or a hub a range of capabilities. Autonegotiation allows two devices to negotiate the mode or data rate of operation. It was designed particularly for the following purposes:

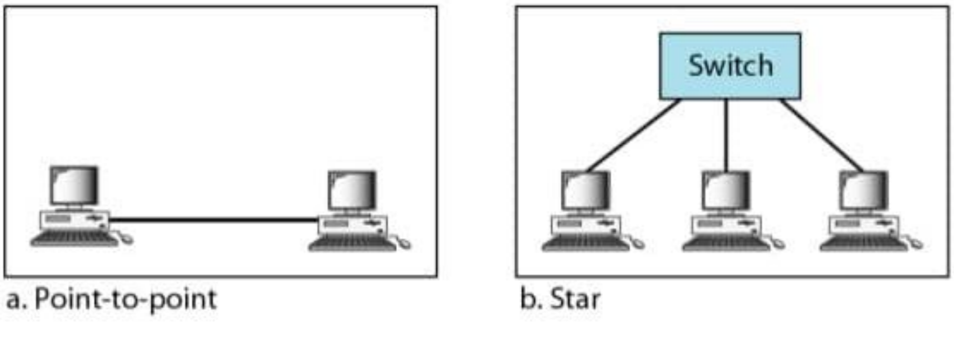
- To allow incompatible devices to connect to one another. For example, a device with a maximum capacity of 10 Mbps can communicate with a device with a 100 Mbps capacity (but can work at a lower rate).
- To allow one device to have multiple capabilities.
- To allow a station to check a hub's capabilities.

Physical Layer

The physical layer in Fast Ethernet is more complicated than the one in Standard Ethernet.

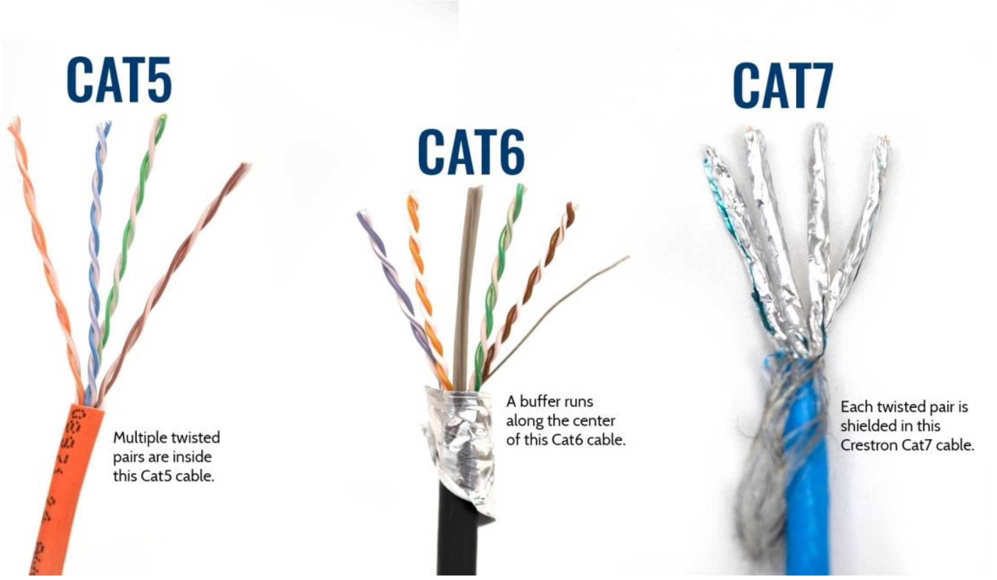
Topology

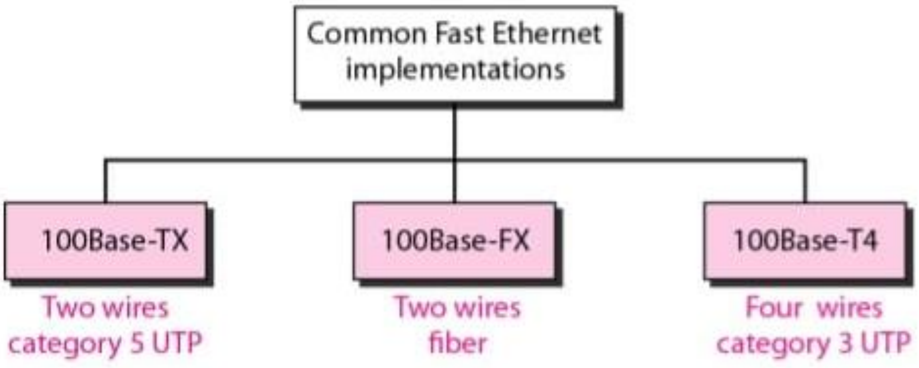
- Fast Ethernet is designed to connect two or more stations together.
- If there are only two stations, they can be connected point-to-point. Three or more stations need to be connected in a star topology with a hub or a switch at the center, as shown in the below figure.



Implementation

- Fast Ethernet implementation at the physical layer can be categorized as either two-wire or four-wire.
- The two-wire implementation can be either category 5 UTP (100 Base-TX) or fiber-optic cable (100Base-FX).
- The four-wire implementation is designed only for category 3 UTP (l00Base-T4).


CAT5
CAT6
CAT7
Multiple twisted pairs are inside this Cat5 cable.
A buffer runs along the center of this Cat6 cable.
Each twisted pair is shielded in this Crestron Cat7 cable.



100Base-TX

- It uses two pairs of twisted-pair cable (either category 5 UTP or STP).
- For this implementation, the MLT-3 scheme was selected since it has good bandwidth performance .
- However, since MLT-3 is not a self-synchronous line coding scheme, 4B/5B block coding is used to provide bit synchronization by preventing the occurrence of a long sequence of 0s and Is .
- This creates a data rate of 125 Mbps, which is fed into MLT-3 for encoding.

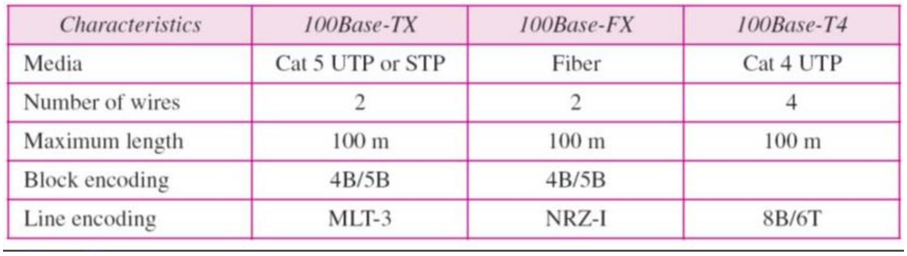
100Base-FX

- It uses two pairs of fiber-optic cables.
- Optical fiber can easily handle high bandwidth requirements by using simple encoding schemes.
- The designers of 100Base-FX selected the NRZ-I encoding scheme for this implementation.
- However, NRZ-I has a bit synchronization problem for long sequences of 0s (or 1s, based on the encoding).
- To overcome this problem, the designers used 4B/5B block encoding as we described for 100Base-TX.
- The block encoding increases the bit rate from 100 to 125 Mbps, which can easily be handled by fiber-optic cable.

100Base-T4,

- It was designed to use category 3 or higher UTP.
- The implementation uses four pairs of UTP for transmitting 100 Mbps.
- Encoding/decoding in 100Base-T4 is more complicated.
- As this implementation uses category 3 UTP, each twisted-pair cannot easily handle more than 25 Mbaud.
- In this design, one pair switches between sending and receiving.
- Three pairs of UTP category 3, however, can handle only 75 Mbaud (25 Mbaud) each.
- We need to use an encoding scheme that converts 100 Mbps to a 75 Mbaud signal.
- In 8B/6T, eight data elements are encoded as six signal elements.
- This means that 100 Mbps uses only (6/8) x 100 Mbps, or 75 Mbaud.

Comparison table



Characteristics	*100Base-TX*	*100Base-FX*	*100Base-T4*
Media	Cat 5 UTP or STP	Fiber	Cat 4 UTP
Number of wires	2	2	4
Maximum length	100 m	100 m	100 m
Block encoding	4B/5B	4B/5B	
Line encoding	MLT-3	NRZ-I	8B/6T

Discussion

Q.1.How is the preamble field different from the SFD field?

Q.2. What is the purpose of an NIC?

Q.3. What is the difference between a unicast, multicast, and broadcast address?

Q.4.What are the common Standard Ethernet implementations?

Q.5.What is the hexadecimal equivalent of the following Ethernet address?

01011010 00010001 01010101 00011000 10101010 00001111

GIGABIT ETHERNET

- The need for an even higher data rate resulted in the design of the Gigabit Ethernet protocol (1000 Mbps).
- The IEEE committee calls the Standard 802.3z.

The goals of the Gigabit Ethernet design can be summarized as follows:

1. Upgrade the data rate to 1 Gbps.
2. Make it compatible with Standard or Fast Ethernet.
3. Use the same 48-bit address.
4. Use the same frame format.
5. Keep the same minimum and maximum frame lengths.
6. To support autonegotiation as in Fast Ethernet.

MAC Sublayer

A main consideration in the evolution of Ethernet was to keep the MAC sublayer untouched. However, to achieve a data rate 1 Gbps, this was no longer possible. Gigabit Ethernet has two distinctive approaches for medium access: half-duplex and full-duplex. Almost all implementations of Gigabit Ethernet follow the full-duplex approach.

Full-Duplex Mode

- In full-duplex mode, there is a central switch connected to all computers or other switches.
- In this mode, each switch has buffers for each input port in which data are stored until they are transmitted.
- There is no collision in this mode, as we discussed before.
- This means that CSMA/CD is not used.
- Lack of collision implies that the maximum length of the cable is determined by the signal attenuation in the cable, not by the collision detection process.

Half-Duplex Mode

- Gigabit Ethernet can also be used in half-duplex mode, although it is rare.
- In this case, a switch can be replaced by a hub, which acts as the common cable in which a collision might occur.
- The half-duplex approach uses CSMA/CD.
- However, as we saw before, the maximum length of the network in this approach is totally dependent on the minimum frame size.
- Three methods have been defined: traditional, carrier extension, and frame bursting.

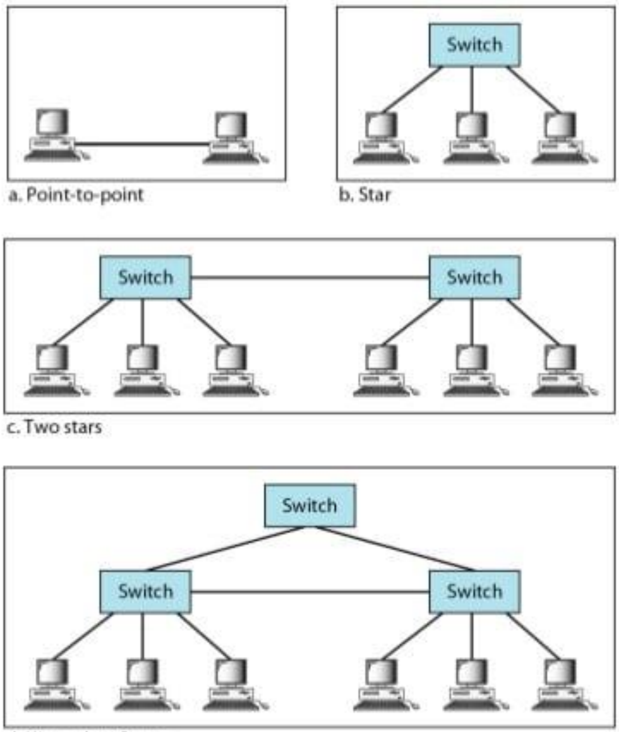
Physical Layer

The physical layer in Gigabit Ethernet is more complicated than that in Standard or Fast Ethernet. We briefly discuss some features of this layer.

Topology

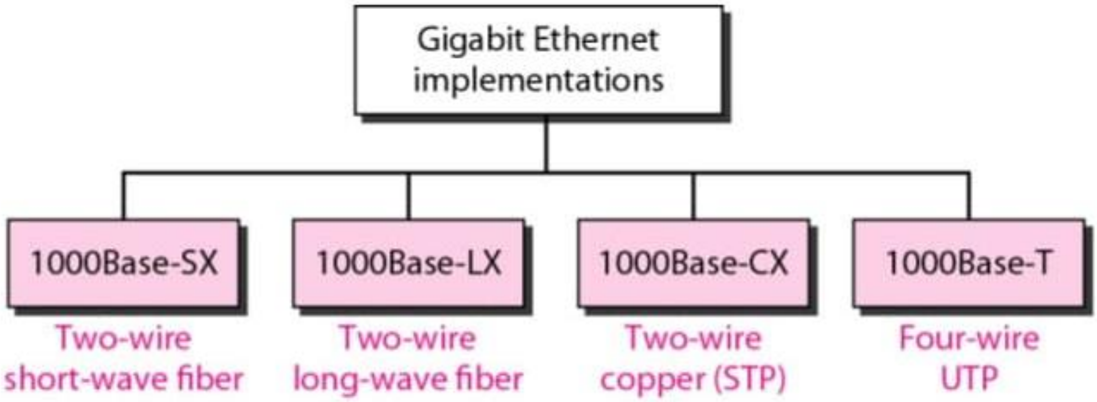
- Gigabit Ethernet is designed to connect two or more stations.
- If there are only two stations, they can be connected point-to-point.

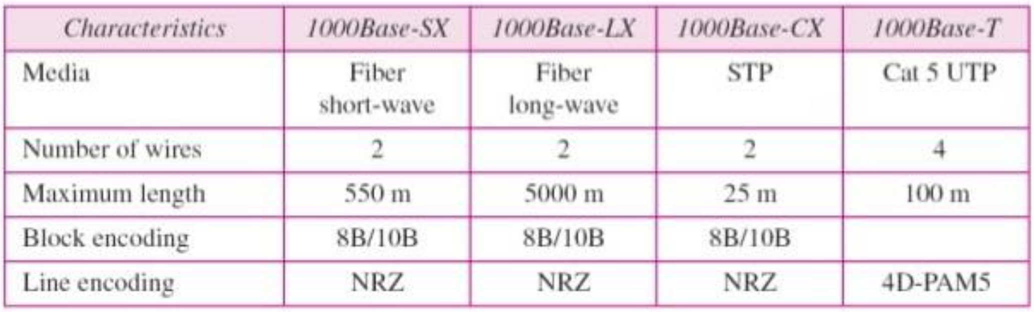
- Three or more stations need to be connected in a star topology with a hub or a switch at the center.
- Another possible configuration is to connect several star topologies or let a star topology be part of another .



Implementation

- Gigabit Ethernet can be categorized as either a two-wire or a four-wire implementation.
- The two-wire implementations use fiber-optic cable (1000Base-SX, short-wave, or 1000Base-LX, long-wave), or STP (1000Base-CX).
- The four-wire version uses category 5 twisted-pair cable (1000Base-T).
- In other words, we have four implementations, as shown in the below figure





Characteristics	*1000Base-SX*	*1000Base-LX*	*1000Base-CX*	*1000Base-T*
Media	Fiber short-wave	Fiber long-wave	STP	Cat 5 UTP
Number of wires	2	2	2	4
Maximum length	550 m	5000 m	25 m	100 m
Block encoding	8B/10B	8B/10B	8B/10B	
Line encoding	NRZ	NRZ	NRZ	4D-PAM5

Unit-7. TCP/IP

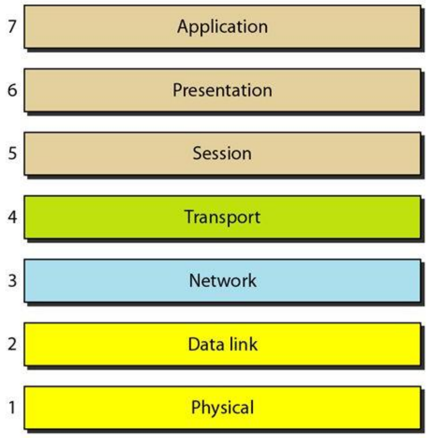
Network Models

A network is a combination of hardware and software that sends data from one location to another. The hardware consists of the physical equipment that carries signals from one point of the network to another. The software consists of instruction sets that make possible the services that we expect from a network.

The layered model that dominated data communications and networking literature before 1990 was the Open Systems Interconnection (OSI) model. Everyone believed that the OSI model would become the ultimate standard for data communications, but this did not happen. The TCP/IP protocol suite became the dominant commercial architecture because it was used and tested extensively in the Internet; the OSI model was never fully implemented.

THE OSI MODEL

- Established in 1947, the International Standards Organization (ISO) is a multinational body dedicated to worldwide agreement on international standards.
- An ISO standard that covers all aspects of network communications is the Open Systems Interconnection model.
- It was first introduced in the late 1970s.
- An open system is a set of protocols that allows any two different systems to communicate regardless of their underlying architecture.
- The purpose of the OSI model is to show how to facilitate communication between different systems without requiring changes to the logic of the underlying hardware and software.
- The OSI model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust, and interoperable.
- The OSI model is a layered framework for the design of network systems that allows communication between all types of computer systems. It consists of seven separate but related layers, each of which defines a part of the process of moving information across a network .
- An understanding of the fundamentals of the OSI model provides a solid basis for exploring data communications.



Layered Architecture

The OSI model is composed of seven ordered layers: physical (layer 1), data link (layer 2), network (layer 3), transport (layer 4), session (layer 5), presentation (layer 6), and application (layer 7).

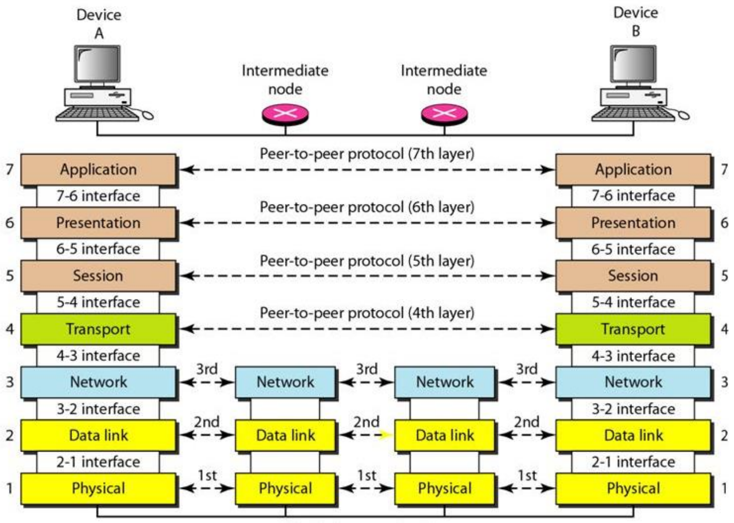
Peer-to-Peer Processes

At the physical layer, communication is direct

In the figure below, device A sends a stream of bits to device B (through intermediate nodes).

At the higher layers, however, communication must move down through the layers on device A, over to device B, and then back up through the layers.

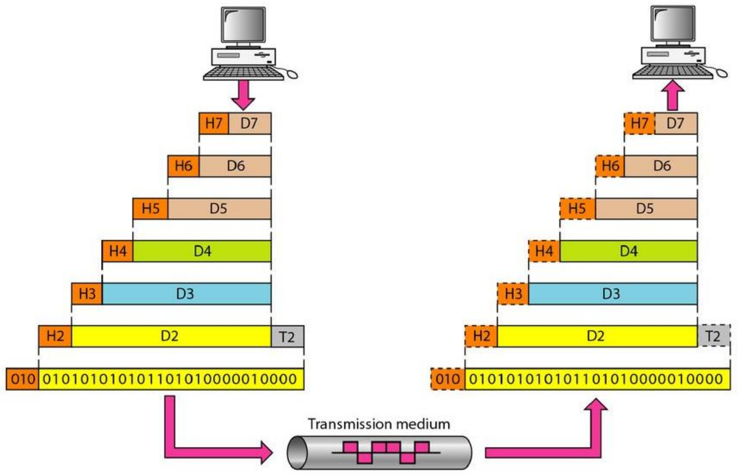
Each layer in the sending device adds its own information to the message it receives from the layer just above it and passes the whole package to the layer just below it.



Organization of the Layers

- The seven layers can be thought of as belonging to three subgroups.
- Layers I, 2, and 3-physical, data link, and network-are the network support layers; they deal with the physical aspects of moving data from one device to another (such as electrical specifications, physical connections, physical addressing, and transport timing and reliability).
- Layers 5, 6, and 7-session, presentation, and application-can be thought of as the user support layers; they allow interoperability among unrelated software systems.
- Layer 4, the transport layer, links the two subgroups and ensures that what the lower layers have transmitted is in a form that the upper layers can use.
- The upper OSI layers are almost always implemented in software; lower layers are a combination of hardware and software, except for the physical layer, which is mostly hardware.

Figure below gives an overall view of the OSI layers, D7 means the data unit at layer 7, D6 means the data unit at layer 6, and so on. The process starts at layer 7 (the application layer), then moves from layer to layer in descending, sequential order. At each layer, a header, or possibly a trailer, can be added to the data unit. Commonly, the trailer is added only at layer 2. When the formatted data unit passes through the physical layer (layer 1), it is changed into an electromagnetic signal and transported along a physical link.



Encapsulation

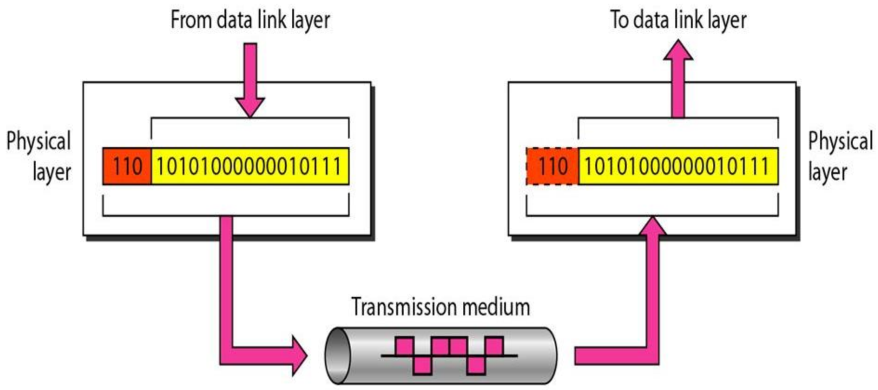
Another aspect of data communications in the OSI model is encapsulation.

A packet (header and data) at level 7 is encapsulated in a packet at level 6. The whole packet at level 6 is encapsulated in a packet at level 5, and so on. In other words, the data portion of a packet at level N - 1 carries the whole packet (data and header and maybe trailer) from level N. The concept is called encapsulation; level N - 1 is not aware of which part of the encapsulated packet is data and which part is the header or trailer. For level N – 1 , the whole packet coming from level N is treated as one integral unit.

LAYERS IN THE OSI MODEL

Physical Layer

- The physical layer coordinates the functions required to carry a bit stream over a physical medium.
- It deals with the mechanical and electrical specifications of the interface and transmission medium.
- It also defines the procedures and functions that physical devices and interfaces have to perform for transmission to Occur.
- Figure below shows the position of the physical layer with respect to the transmission medium and the data link layer.



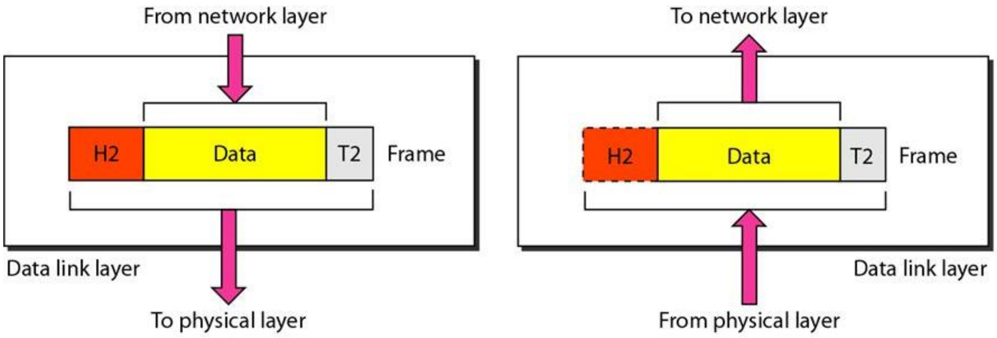
The physical layer is also concerned with the following:

- **Physical characteristics of interfaces and medium**. The physical layer defines the characteristics of the interface between the devices and the transmission medium. It also defines the type of transmission medium.
- **Representation of bits**. The physical layer data consists of a stream of bits (sequence of 0s or 1s) with no interpretation. To be transmitted, bits must be encoded into signals--electrical or optical. The physical layer defines the type of encoding (how 0s and Is are changed to signals).
- **Data rate**. The transmission rate-the number of bits sent each second-is also defined by the physical layer. In other words, the physical layer defines the duration of a bit, which is how long it lasts.
- **Synchronization of bits**. The sender and receiver not only must use the same bit rate but also must be synchronized at the bit level. In other words, the sender and the receiver clocks must be synchronized.
- **Line configuration**. The physical layer is concerned with the connection of devices to the media. In a point-to-point configuration, two devices are connected through a dedicated link. In a multipoint configuration, a link is shared among several devices.
- **Physical topology**. The physical topology defines how devices are connected to make a network. Devices can be connected by using a mesh topology (every device is connected to every other device), a star topology (devices are connected through a central device), a ring topology (each device is connected to the next, forming a ring), a bus topology (every device is on a common link), or a hybrid topology (this is a combination of two or more topologies).
- **Transmission mode**. The physical layer also defines the direction of transmission between two devices: simplex, half-duplex, or full-duplex. In simplex mode, only one device can send; the other can only receive. The simplex mode is a one-way communication. In the half-duplex mode, two devices can send and receive, but not at the same time. In a full-duplex (or simply duplex) mode, two devices can send and receive at the same time.

Data Link Layer

- The data link layer transforms the physical layer, a raw transmission facility, to a reliable link.
- It makes the physical layer appear error-free to the upper layer (network layer).

- Figure below shows the relationship of the data link layer to the network and physical layers.

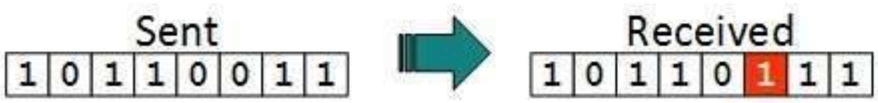


Other responsibilities of the data link layer include the following:

- **Framing**. The data link layer divides the stream of bits received from the network layer into manageable data units called frames.
- **Physical addressing**. If frames are to be distributed to different systems on the network, the data link layer adds a header to the frame to define the sender and/or receiver of the frame. If the frame is intended for a system outside the sender's network, the receiver address is the address of the device that connects the network to the next one.
- **Flow control**. If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.
- **Error control**. The data link layer adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged or lost frames. It also uses a mechanism to recognize duplicate frames. Error control is normally achieved through a trailer added to the end of the frame.
- **Access control**. When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

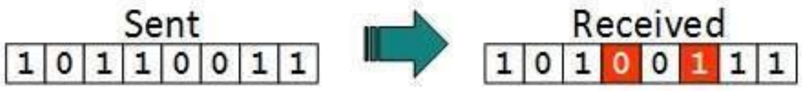
There may be three types of errors:

- **Single bit error**



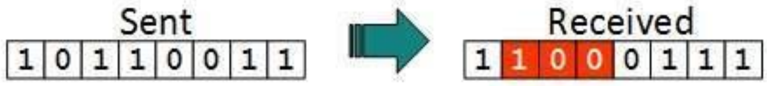
In a frame, there is only one bit, anywhere though, which is corrupt.

- **Multiple bits error**



Frame is received with more than one bits in corrupted state.

- **Burst error**



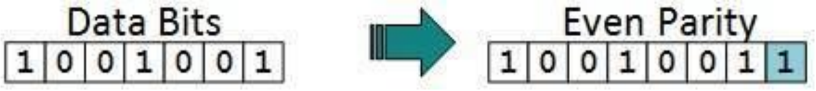
Frame contains more than1 consecutive bits corrupted.

Error control mechanism may involve two possible ways:

- Error detection
- Error correction

Parity Check

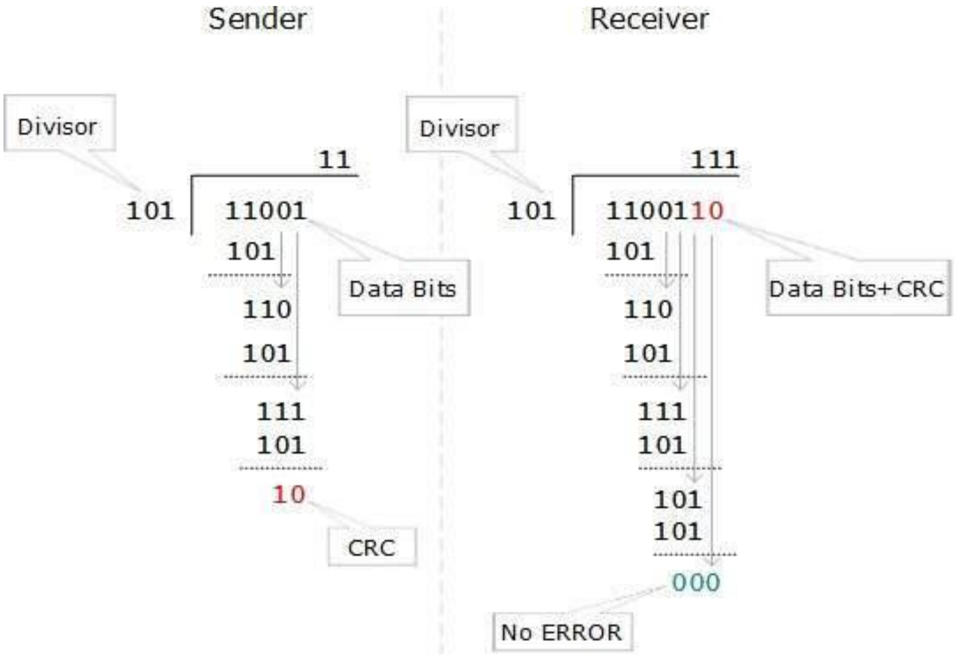
- One extra bit is sent along with the original bits to make number of 1s either even in case of even parity, or odd in case of odd parity.
- The sender while creating a frame counts the number of 1s in it. For example, if even parity is used and number of 1s is even then one bit with value 0 is added. This way number of 1s remains even.If the number of 1s is odd, to make it even a bit with value 1 is added.



- The receiver simply counts the number of 1s in a frame. If the count of 1s is even and even parity is used, the frame is considered to be not-corrupted and is accepted. If the count of 1s is odd and odd parity is used, the frame is still not corrupted.
- If a single bit flips in transit, the receiver can detect it by counting the number of 1s. But when more than one bits are erroneous, then it is very hard for the receiver to detect the error.

Cyclic Redundancy Check (CRC)

- CRC is a different approach to detect if the received frame contains valid data. This technique involves binary division of the data bits being sent. The divisor is generated using polynomials. The sender performs a division operation on the bits being sent and calculates the remainder. Before sending the actual bits, the sender adds the remainder at the end of the actual bits. Actual data bits plus the remainder is called a codeword. The sender transmits data bits as codewords.



-
- At the other end, the receiver performs division operation on codewords using the same CRC divisor. If the remainder contains all zeros the data bits are accepted, otherwise it is considered as there some data corruption occurred in transit.

Error Correction

In the digital world, error correction can be done in two ways:

- **Backward Error Correction** When the receiver detects an error in the data received, it requests back the sender to retransmit the data unit.
- **Forward Error Correction** When the receiver detects some error in the data received, it executes error-correcting code, which helps it to auto-recover and to correct some kinds of errors.

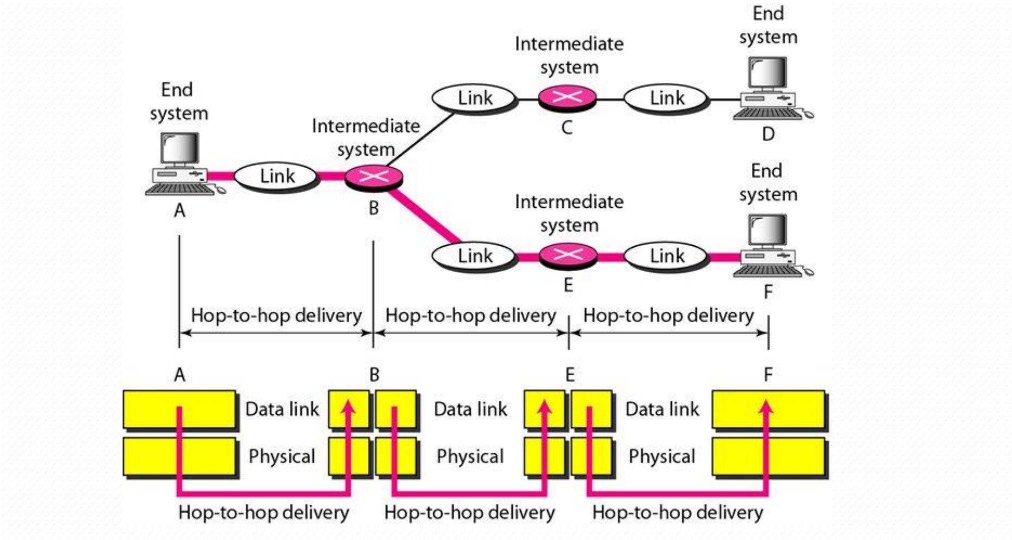
The first one, Backward Error Correction, is simple and can only be efficiently used where retransmitting is not expensive. For example, fiber optics. But in case of wireless transmission retransmitting may cost too much. In the latter case, Forward Error Correction is used.

To correct the error in data frame, the receiver must know exactly which bit in the frame is corrupted. To locate the bit in error, redundant bits are used as parity bits for error detection. For example, we take ASCII words (7 bits data), then there could be 8 kind of information we need: first seven bits to tell us which bit is error and one more bit to tell that there is no error.

For m data bits, r redundant bits are used. r bits can provide 2^r combinations of information. In m+r bit codeword, there is possibility that the r bits themselves may get corrupted. So the number of r bits used must inform about m+r bit locations plus no-error information, i.e. m+r+1.

$$2^r >= m+r+1$$

Figure below illustrates hop-to-hop (node-to-node) delivery by the data link layer

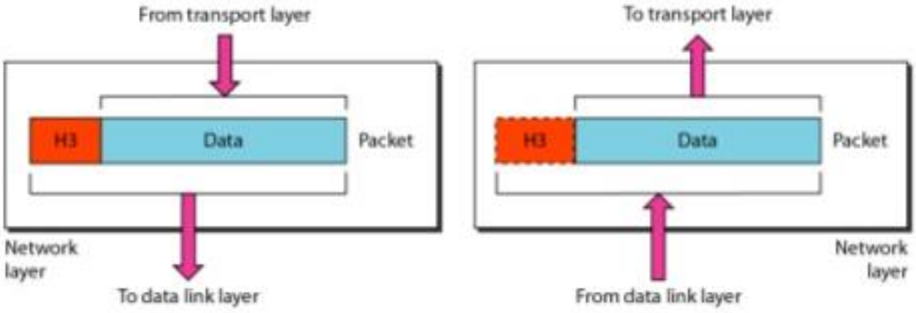


As the figure shows, communication at the data link layer occurs between two adjacent nodes.

- To send data from A to F, three partial deliveries are made. First, the data link layer at A sends a frame to the data link layer at B (a router).
- Second, the data link layer at B sends a new frame to the data link layer at E.
- Finally, the data link layer at E sends a new frame to the data link layer at F.
- The frames that are exchanged between the three nodes have different values in the headers.
- The frame from A to B has B as the destination address and A as the source address.
- The frame from B to E has E as the destination address and B as the source address.
- The frame from E to F has F as the destination address and E as the source address.
- The values of the trailers can also be different if error checking includes the header of the frame.

Network Layer

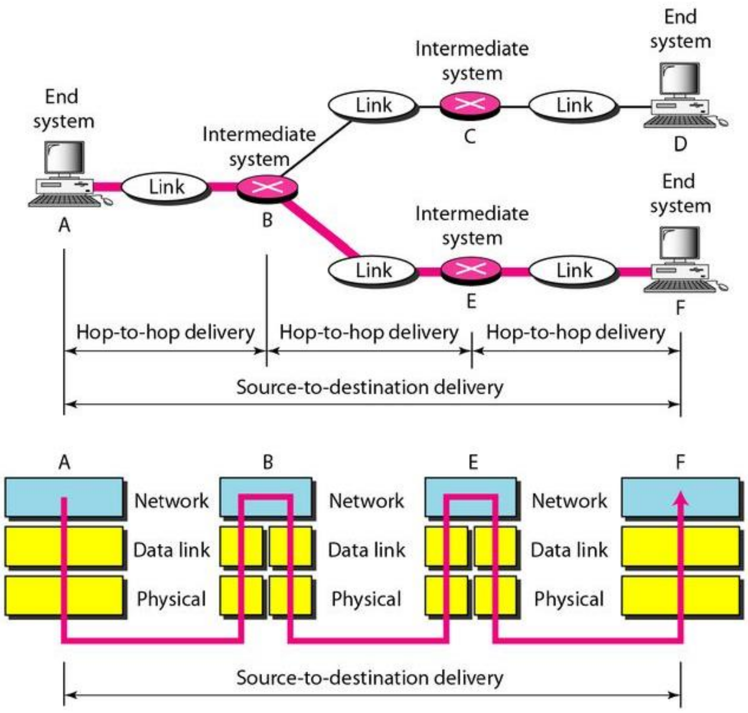
- The network layer is responsible for the **source-to-destination** delivery of a packet, possibly across multiple networks (links).
- Whereas the data link layer oversees the delivery of the packet between two systems on the same network (links), the network layer ensures that each packet gets from its point of origin to its final destination.
- If two systems are connected to the same link, there is usually no need for a network layer. However, if the two systems are attached to different networks (links) with connecting devices between the networks (links), there is often a need for the network layer to accomplish source-to-destination delivery.
- Figure below shows the relationship of the network layer to the data link and transport layers.



Other responsibilities of the network layer include the following:

Logical addressing. The physical addressing implemented by the data link layer handles the addressing problem locally. If a packet passes the network boundary, we need another addressing system to help distinguish the source and destination systems. The network layer adds a header to the packet coming from the upper layer that, among other things, includes the logical addresses of the sender and receiver.

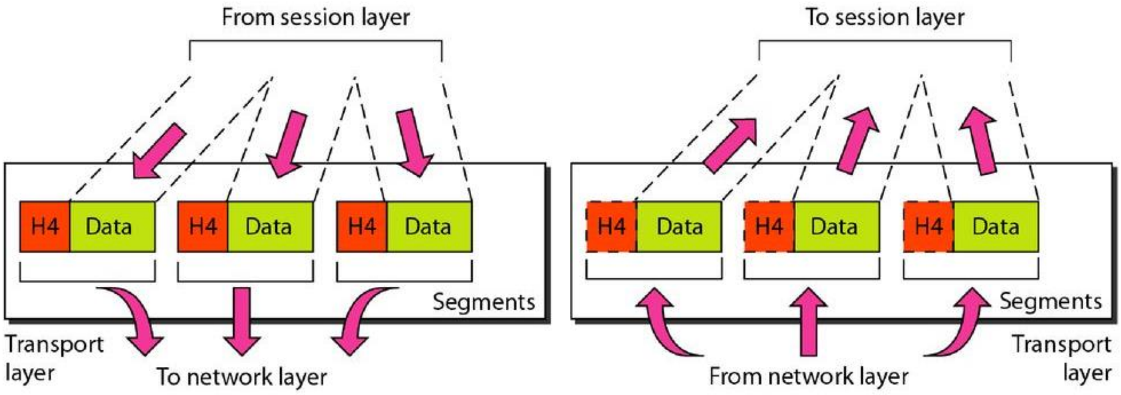
Routing. When independent networks or links are connected to create internetworks (network of networks) or a large network, the connecting devices (called routers or switches) route or switch the packets to their final destination. One of the functions of the network layer is to provide this mechanism



- The network layer at A sends the packet to the network layer at B.
- When the packet arrives at router B, the router makes a decision based on the final destination (F) of the packet.
- Router B uses its routing table to find that the next hop is router E.
- The network layer at B, therefore, sends the packet to the network layer at E.
- The network layer at E, in turn, sends the packet to the network layer at F.

Transport Layer

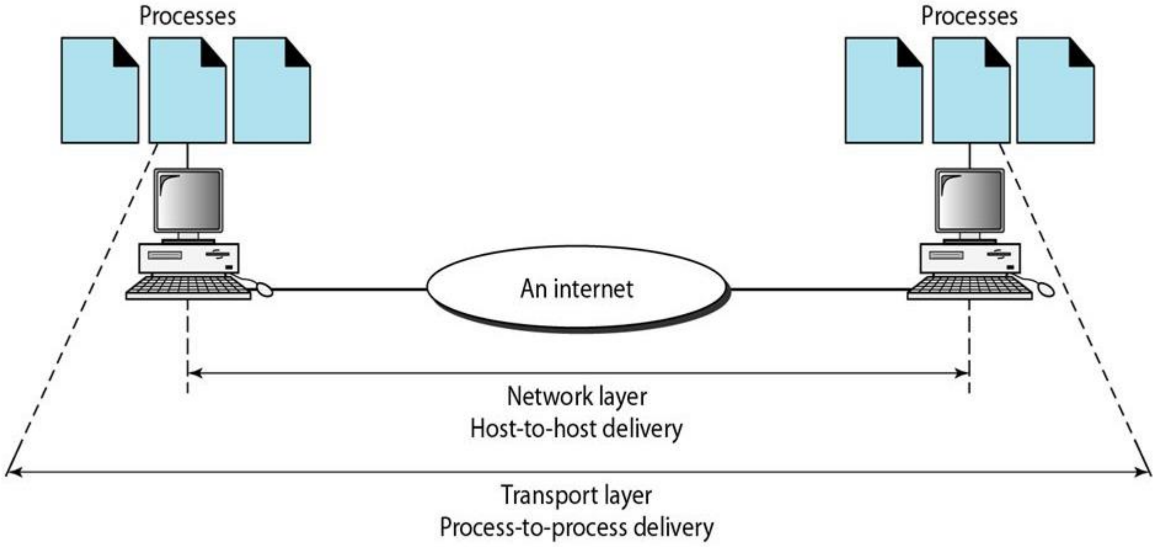
- The transport layer is responsible for process-to-process delivery of the entire message.
- A process is an application program running on a host.
- Whereas the network layer oversees source-to-destination delivery of individual packets, it does not recognize any relationship between those packets. It treats each one independently, as though each piece belonged to a separate message, whether or not it does.
- The transport layer, on the other hand, ensures that the whole message arrives intact and in order, overseeing both error control and flow control at the source-to-destination level.
- Figure below shows the relationship of the transport layer to the network and session layers.



Other responsibilities of the transport layer include the following:

- **Service-point addressing**. Computers often run several programs at the same time. For this reason, source-to-destination delivery means delivery not only from one computer to the next but also from a specific process (running program) on one computer to a specific process (running program) on the other. The transport layer header must therefore include a type of address called a service-point address (or port address). The network layer gets each packet to the correct computer; the transport layer gets the entire message to the correct process on that computer.
- **Segmentation and reassembly**. A message is divided into transmittable segments, with each segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly upon arriving at the destination and to identify and replace packets that were lost in transmission.

- **Connection control**. The transport layer can be either connectionless or connection oriented. A connectionless transport layer treats each segment as an independent packet and delivers it to the transport layer at the destination machine. A connection oriented transport layer makes a connection with the transport layer at the destination machine first before delivering the packets. After all the data are transferred, the connection is terminated.
- **Flow control**. Like the data link layer, the transport layer is responsible for flow control. However, flow control at this layer is performed end to end rather than across a single link.
- **Error control**. Like the data link layer, the transport layer is responsible for error control. However, error control at this layer is performed process-to process rather than across a single link. The sending transport layer makes sure that the entire message arrives at the receiving transport layer without error (damage, loss, or duplication). Error correction is usually achieved through retransmission.



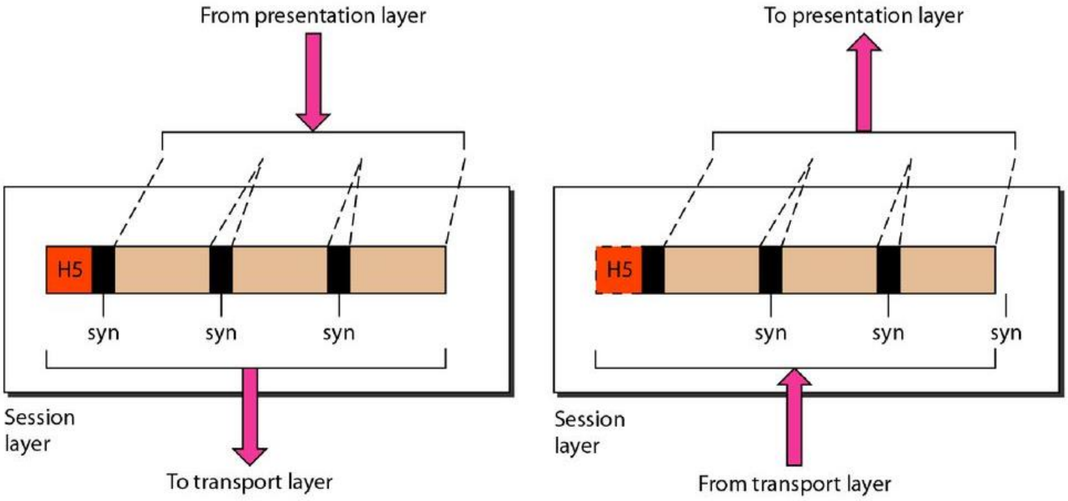
Session Layer

- The services provided by the first three layers (physical, data link, and network) are not sufficient for some processes. The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems.

Specific responsibilities of the session layer include the following:

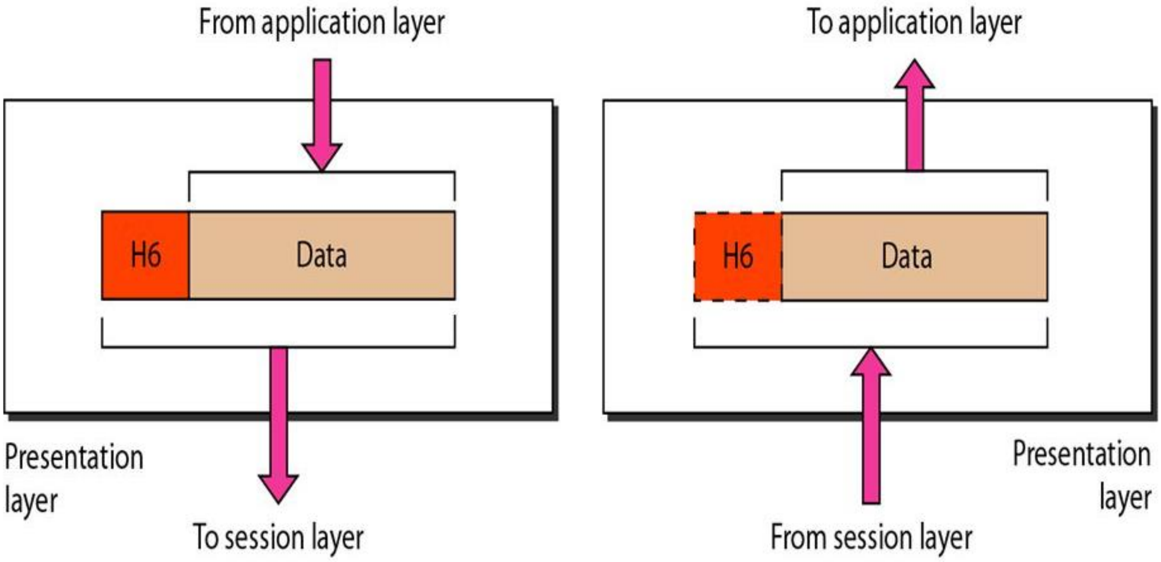
- **Dialog control**. The session layer allows two systems to enter into a dialog. It allows the communication between two processes to take place in either half duplex (one way at a time) or full-duplex (two ways at a time) mode.
- **Synchronization**. The session layer allows a process to add checkpoints, or synchronization points, to a stream of data. For example, if a system is sending a file of2000 pages, it is advisable to insert checkpoints after every 100 pages to ensure that each 100-page unit is received and acknowledged independently. In this case, if a crash happens during the transmission of page 523, the only pages that need to be resent after system recovery are pages 501 to 523. Pages previous to 501 need not be resent.

- Figure below illustrates the relationship of the session layer to the transport and presentation layers.



Presentation Layer

- The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems.
- Figure below shows the relationship between the presentation layer and the application and session layers.



Specific responsibilities of the presentation layer include the following:

Translation.

- The processes (running programs) in two systems are usually exchanging information in the form of character strings, numbers, and so on.
- The information must be changed to bit streams before being transmitted.
- Because different computers use different encoding systems, the presentation layer is responsible for interoperability between these different encoding methods.

- The presentation layer at the sender changes the information from its sender-dependent format into a common format.
- The presentation layer at the receiving machine changes the common format into its receiver-dependent format.

Encryption.

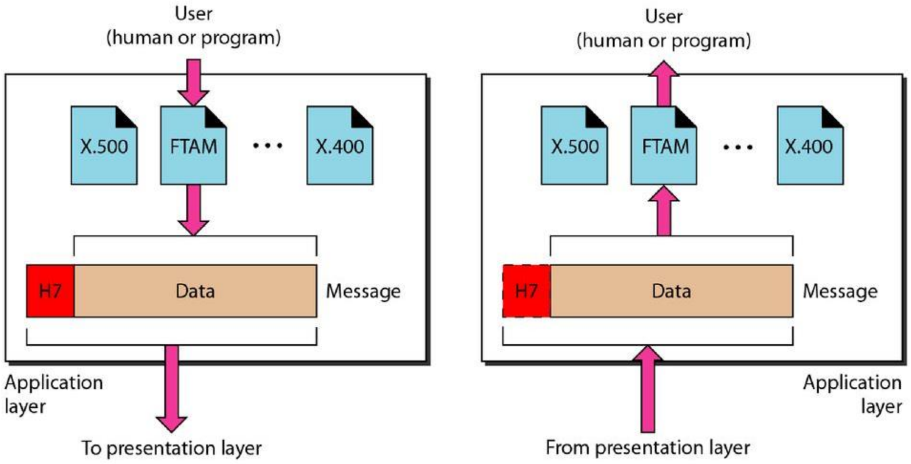
- To carry sensitive information, a system must be able to ensure privacy. Encryption means that the sender transforms the original information to another form and sends the resulting message out over the network. Decryption reverses the original process to transform the message back to its original form.

Compression.

- Data compression reduces the number of bits contained in the information.
- Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

Application Layer

- The application layer enables the user, whether human or software, to access the network.
- It provides user interfaces and support for services such as electronic mail, remote file access and transfer, shared database management, and other types of distributed information services.
- Figure below shows the relationship of the application layer to the user and the presentation layer. Of the many application services available, the figure shows only three: X400 (message-handling services), X.500 (directory services), and file transfer, access, and management (FTAM). The user in this example employs X400 to send an e-mail message.



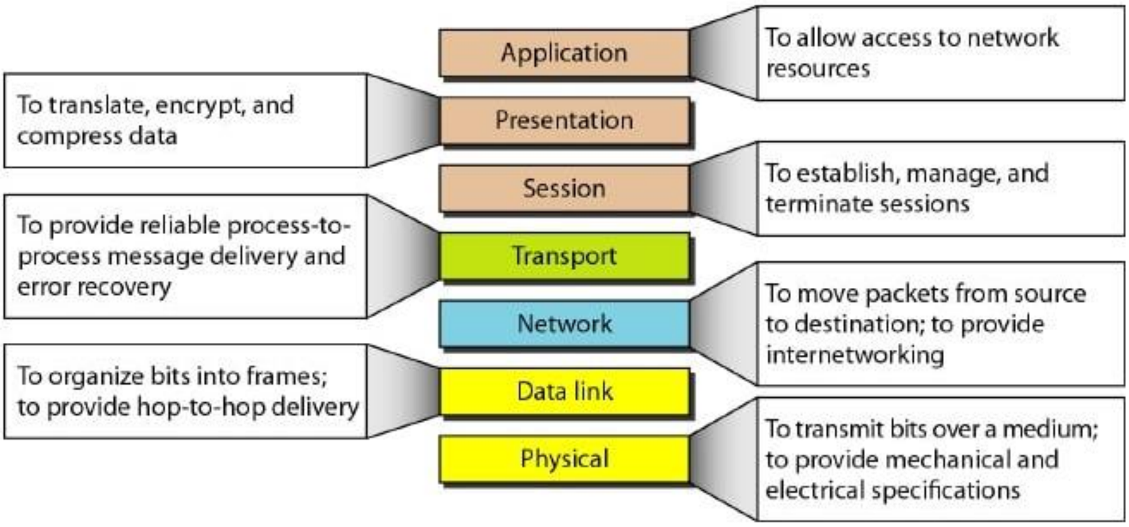
Specific services provided by the application layer include the following:

1. **Network virtual terminal**.
- A network virtual terminal is a software version of a physical terminal, and it allows a user to log on to a remote host.
- To do so, the application creates a software emulation of a terminal at the remote host.
- The user's computer talks to the software terminal which, in turn, talks to the host, and vice versa. The remote host believes it is communicating with one of its own terminals and allows the user to log on.

2. **File transfer, access, and management.**
- This application allows a user to access files in a remote host (to make changes or read data), to retrieve files from a remote computer for use in the local computer, and to manage or control files in a remote computer locally.

3. **Mail services.**
- This application provides the basis for e-mail forwarding and storage.

4. **Directory services**.
- This application provides distributed database sources and access for global information about various objects and services.


Application
To allow access to network resources
To translate, encrypt, and compress data
Presentation
Session
To establish, manage, and terminate sessions
To provide reliable process-to-process message delivery and error recovery
Transport
Network
To move packets from source to destination; to provide internetworking
To organize bits into frames; to provide hop-to-hop delivery
Data link
Physical
To transmit bits over a medium; to provide mechanical and electrical specifications

TCP/IP PROTOCOL SUITE

- The TCP/IP protocol suite was developed prior to the OSI model.
- Therefore, the layers in the TCP/IP protocol suite do not exactly match those in the OSI model.
- The original TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application.
- However, when TCP/IP is compared to OSI, we can say that the host-to-network layer is equivalent to the combination of the physical and data link layers. The internet layer is equivalent to the network layer, and the application layer is roughly doing the job of the

session, presentation, and application layers with the transport layer in TCP/IP taking care of part of the duties of the session layer.

- So here, we assume that the TCP/IP protocol suite is made of five layers: **physical, data link, network, transport, and application.**
- The first four layers provide physical standards, network interfaces, internetworking, and transport functions that correspond to the first four layers of the OSI model. The three topmost layers in the OSI model, however, are represented in TCP/IP by a single layer called the application layer

Flow control and Error control:

Data-link layer is responsible for implementation of flow control and error control mechanism.

Flow Control:

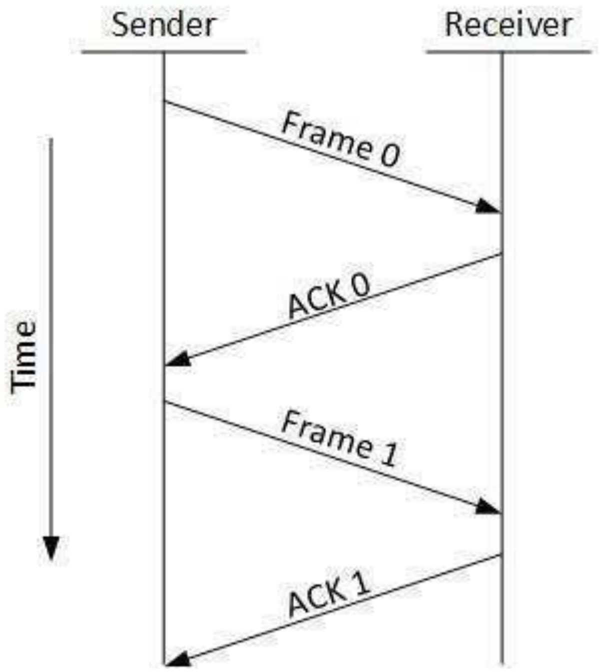
When a data frame is sent from one host to another over a single medium, it is required that the sender and receiver should work at the same speed. That is, sender sends at a speed on which the receiver can process and accept the data. If sender is sending too fast the receiver may be overloaded, and unable to process and receive then the data may lose.

There are two types of mechanism used to control the flow

i- Stop and wait ii- Sliding window

Stop and Wait

This flow control mechanism forces the sender after transmitting a data frame to stop and wait until the acknowledgement of the data-frame sent is received.



Sliding Window

In this flow control mechanism, both sender and receiver agree on the number of data-frames after which the acknowledgement should be sent. Unlike stop and wait flow control mechanism wastes resources, this protocol tries to make use of underlying resources as much as possible.

Error Control

→When data-frame is transmitted, there is a probability that data-frame may be lost in the transit or it is received corrupted.

→In both cases, the receiver does not receive the correct data-frame and sender does not know anything about any loss.

→In such case, both sender and receiver use some protocols which helps them to detect transit errors such as loss of data-frame.

→So that, either the sender retransmits the data-frame or the receiver may request to resend the previous data-frame.

Requirements for error control mechanism:

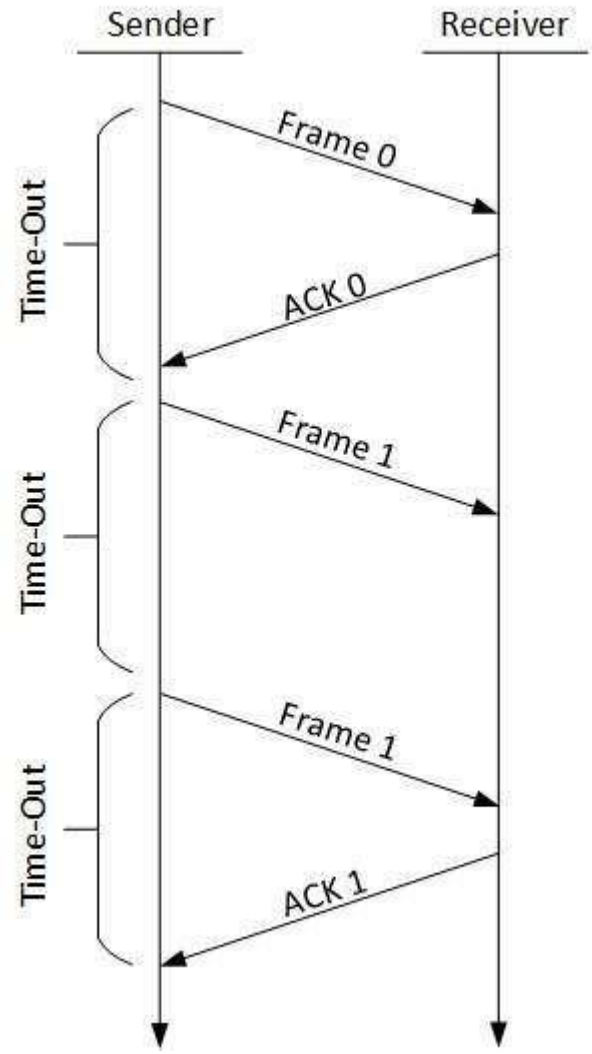
- Error detection - The sender and receiver, either both or any, must ascertain that there is some error in the transit.
- Positive ACK - When the receiver receives a correct frame, it should acknowledge it.
- Negative ACK - When the receiver receives a damaged frame or a duplicate frame, it sends a NACK back to the sender and the sender must retransmit the correct frame.
- Retransmission: The sender maintains a clock and sets a timeout period. If an acknowledgement of a data-frame previously transmitted does not arrive before the timeout the sender retransmits the frame, thinking that the frame or it's acknowledgement is lost in transit.

Data link layer uses three types of techniques available to control the errors by Automatic Repeat Requests (ARQ):

i. Stop-and-wait ARQ **ii.**
Go-Back-N ARQ

iii. Selective Repeat ARQ

Stop-and-wait ARQ

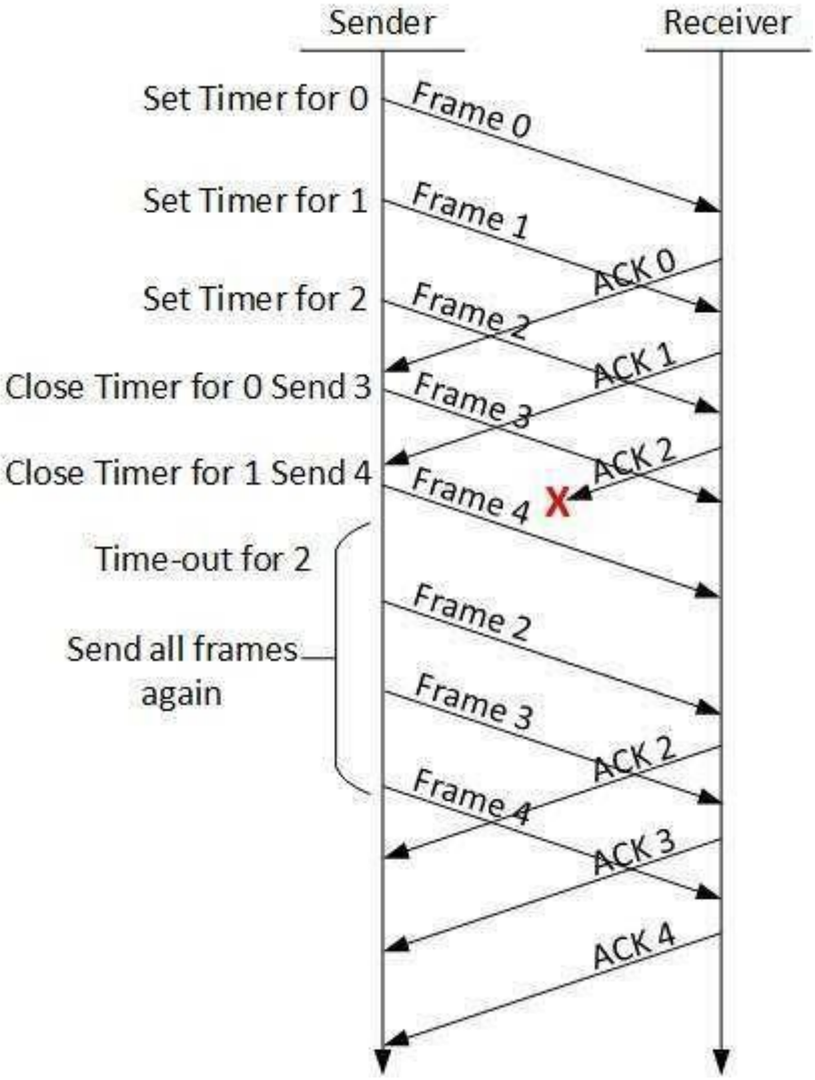


The following transition may occur in Stop-and-Wait ARQ:

- The sender maintains a timeout counter.
- When a frame is sent, the sender starts the timeout counter.
- If acknowledgement of frame comes in time, the sender transmits the next frame in queue. o If acknowledgement does not come in time, the sender assumes that either the frame or its acknowledgement is lost in transit.
  Sender retransmits the frame and starts the timeout counter. o If a negative acknowledgement is received, the sender retransmits the frame.

- **Go-Back-N ARQ**

  Stop and wait ARQ mechanism does not utilize the resources at their best. When the acknowledgement is received, the sender sits idle and does nothing. In Go-Back-N ARQ method, both sender and receiver maintain a window.

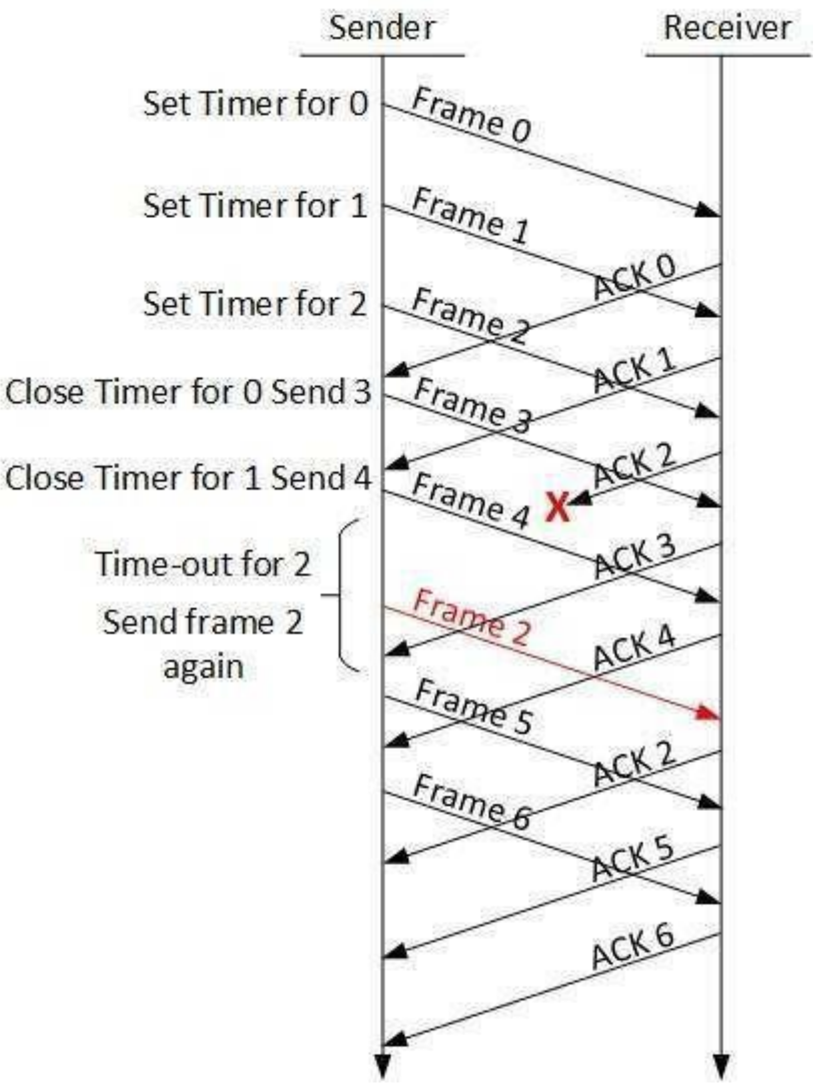


The sending-window size enables the sender to send multiple frames without receiving the acknowledgement of the previous ones. The receiving-window enables the receiver to receive multiple frames and acknowledge them. The receiver keeps track of incoming frame's sequence number.

When the sender sends all the frames in window, it checks up to what sequence number it has received positive acknowledgement. If all frames are positively acknowledged, the sender sends next set of frames. If sender finds that it has received NACK or has not receive any ACK for a particular frame, it retransmits all the frames after which it does not receive any positive ACK.

- **Selective Repeat ARQ**

  In Go-back-N ARQ, it is assumed that the receiver does not have any buffer space for its window size and has to process each frame as it comes. This enforces the sender to retransmit all the frames which are not acknowledged.



In Selective-Repeat ARQ, the receiver while keeping track of sequence numbers, buffers the frames in memory and sends NACK for only frame which is missing or damaged.

The sender in this case, sends only packet for which NACK is received.

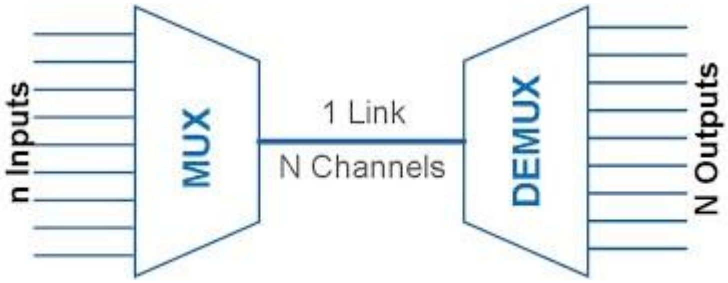
Multiplexing:

Multiplexing which means multiple sources but one link.

→It is a technique by which different analog and digital streams of transmission can be simultaneously processed over a shared link.

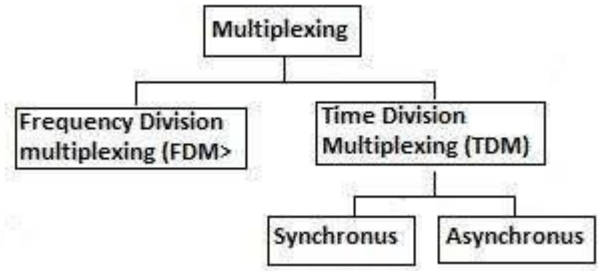
→Multiplexing divides the high capacity medium into low capacity logical medium which is then shared by different streams.

→When multiple senders try to send over a single medium, a device called Multiplexer divides the physical channel and allocates one to each. →At the receiver side a De-multiplexer receives data from a single medium, identifies each, and sends to different receivers.



Types of Multiplexing:
There are two basic techniques:

1. FDM (Frequency Division Multiplexing)
2. TDM (Time-Division Multiplexing) i- Synchronous TDM ii- Asynchronous TDM


Multiplexing
Frequency Division multiplexing (FDM>
Time Division Multiplexing (TDM)
Synchronus
Asynchronus

1. Frequency Division Multiplexing (FDM) –

→In this a number of signals are transmitted at the same time, and each source transfers its signals in the allotted frequency range.

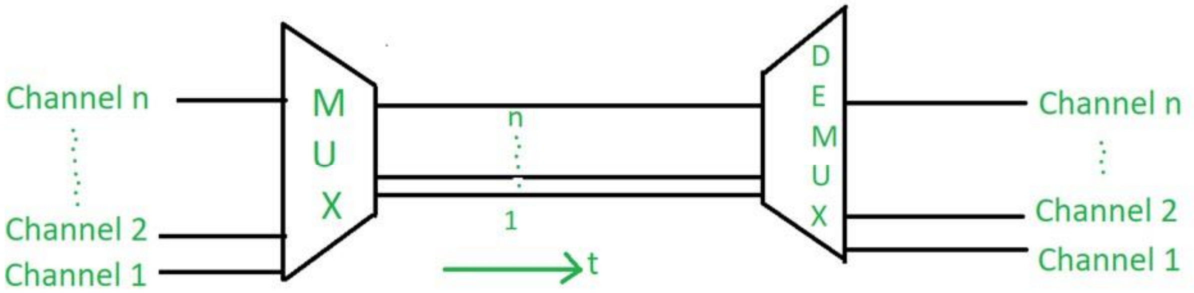
→There is a suitable frequency gap between the 2 adjacent signals to avoid over-lapping.

→Since the signals are transmitted in allotted time so this decreases the probability of collision.

→The frequency spectrum is divided into several logical channels, in which every user feels that they posses a particular bandwidth.

→A number of signals are sent simultaneously on the same time allocating separate frequency band or channel to each signal.

→Therefore to avoid interference between two successive channels **Guard bands** are used.



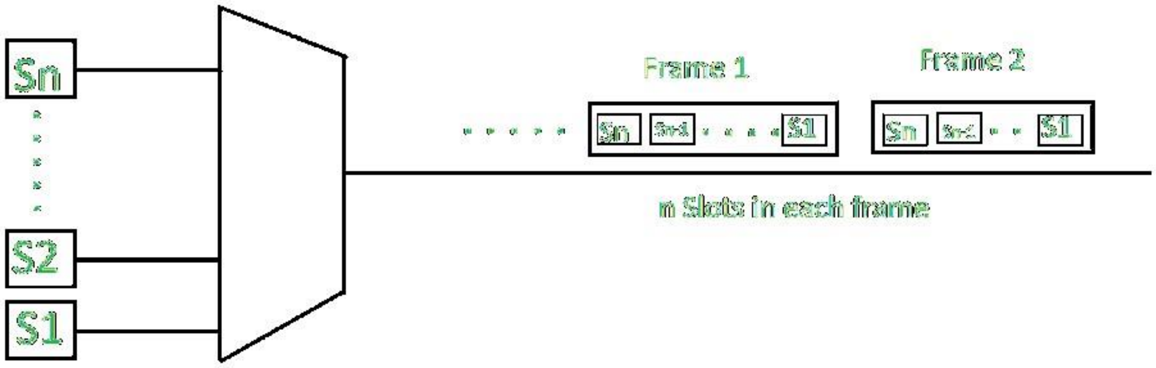
Advantages of FDM:

1. A large number of signals (channels) can be transmitted simultaneously.
2. FDM does not need synchronization between its transmitter and receiver for proper operation.
3. Demodulation of FDM is easy.
4. Due to slow narrow band fading only a single channel gets affected.

2. Time Division Multiplexing (TDM) –

→TDM Occurs when data transmission rate of media is greater than that of the source, and each signal is allotted a definite amount of time.

→In frequency division multiplexing all the signals operate at the same time with different frequencies, but in time division multiplexing all the signals operate with same frequency at different times.



It is of two types:

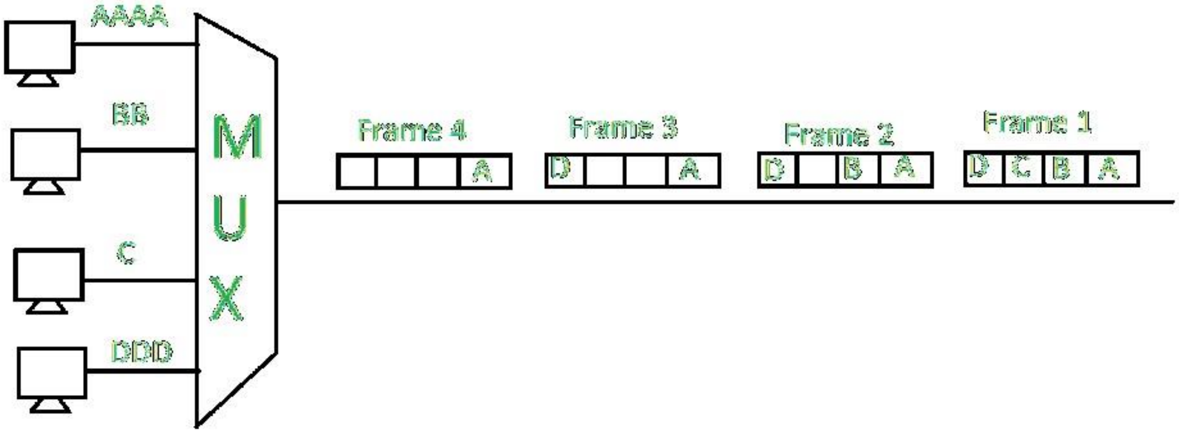
1. Synchronous TDM –

→The time slots are pre-assigned and fixed.

→This slot is even given if the source is not ready with data at this time.

→In this case the slot is transmitted empty.

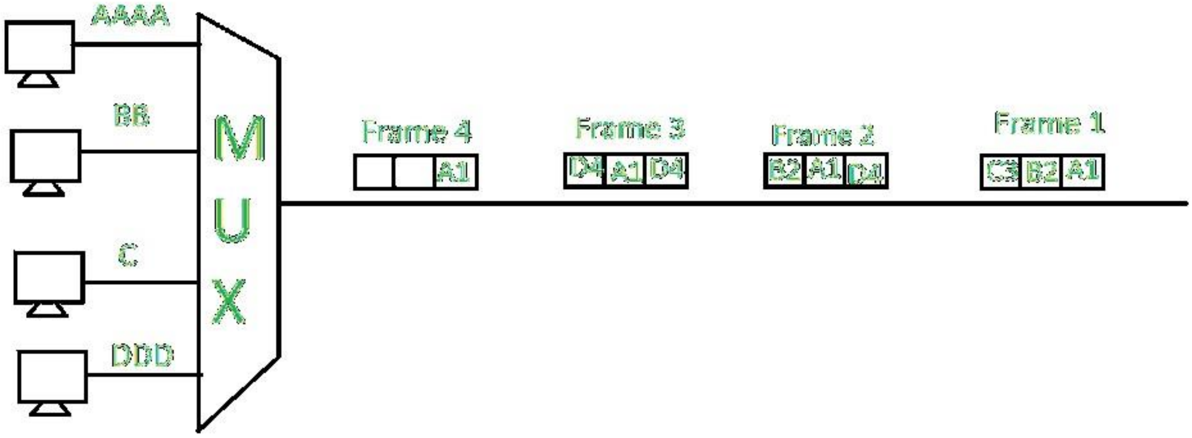
→It is used for multiplexing digitized voice stream.



2. Asynchronous (or statistical) TDM –

→The slots are allocated dynamically depending on the speed of source or their ready state.

→It dynamically allocates the time slots according to different input channel's needs, thus saving the channel capacity.



Module-5
Switching and Routing:

What is switching

It is a mechanism which exchanges the information between different networks or different computers.

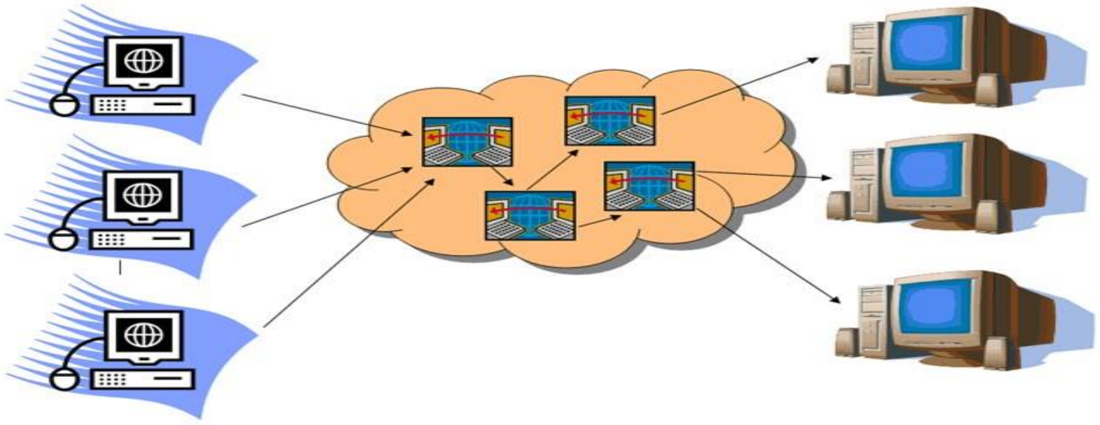
Different types of data use different types of switching techniques.

There are three type of switching techniques available: **Circuit Switching, Packet Switching and Message Switching**.

Among three Circuit and Packet Switching are the most used techniques.

Circuit Switching

- Circuit switching is a switching method where an end-to-end path is created between two stations within a network before starting the data transfer.
- It has three phases: **Circuit establishment, Transferring the data and circuit disconnect**.
- Circuit switching method has a fixed data rate and both the subscribers need to operate at this fixed rate.
- Circuit switching is the simplest method of data communication where dedicated physical connections are established between two individual senders and receiver.
- To create these dedicated connections, a set of switches are connected by physical links.



- In the above picture three computers on the left side are connected with three desktops PCs on the right side with physical links, depending on the four circuit switches.
- If the circuit switching is not used, they need to be connected with point-topoint connections, where many numbers of dedicated lines are required, which will not only increase the connection cost but also increase the complexity of the system.
- In this case routing decision is made when the routing path is being established in the network.
- After the dedicated routing path is established the data continuously transmitted to the receiver destination.
- The connection is maintained until the end of the data transfer.

Three Phases in Circuit switching

<u>Setup phase</u>

- A dedicated routing or connection path is established between the sender and the receiver.
- At this period End to End addressing, like source address, destination address is must create a connection between two physical devices.
- The circuit switching happens in the physical layers. <u>Data transfer</u>
- Data transfer happens only after the setup phase is completed and a dedicated path is established.
- No addressing method is involved in this phase.
- The switches use time slot (TDM) or the occupied band (FDM) to route the data from the sender to the receiver.
- The data sending in this method is continuous and there may be periods of silence in data transmitting.
- All internal connections are made in duplex form.

<u>Circuit disconnect phase</u>

- In this phase any one from the sender or receiver needs to disconnect the path, a disconnect signal is sent to all involved switches to release the resource and break the connection.
- This phase also called as Teardown phase in circuit switching method.

Advantages of Circuit Switching

1. The data rate is fixed and dedicated because the connection is established using dedicated physical connection or circuits.
2. As there are dedicated transmission routing paths involved, it is a good choice for continuous transmission over a long duration.
3. The data transmission delay is negligible. No waiting time is involved in switches. So, the data gets transmitted without any prior delay in the transmission.

Disadvantages of Circuit Switching

1. Whether the communication channel is free or busy, the dedicated channel could not be used for other data transmission.
2. It requires more bandwidth, and continuous transmission offers wastage of bandwidth when there is a silence period.
3. It is highly inefficient when utilizing the system resource.
4. It takes huge time during the establishment of physical links between senders and receivers.

Packet Switching

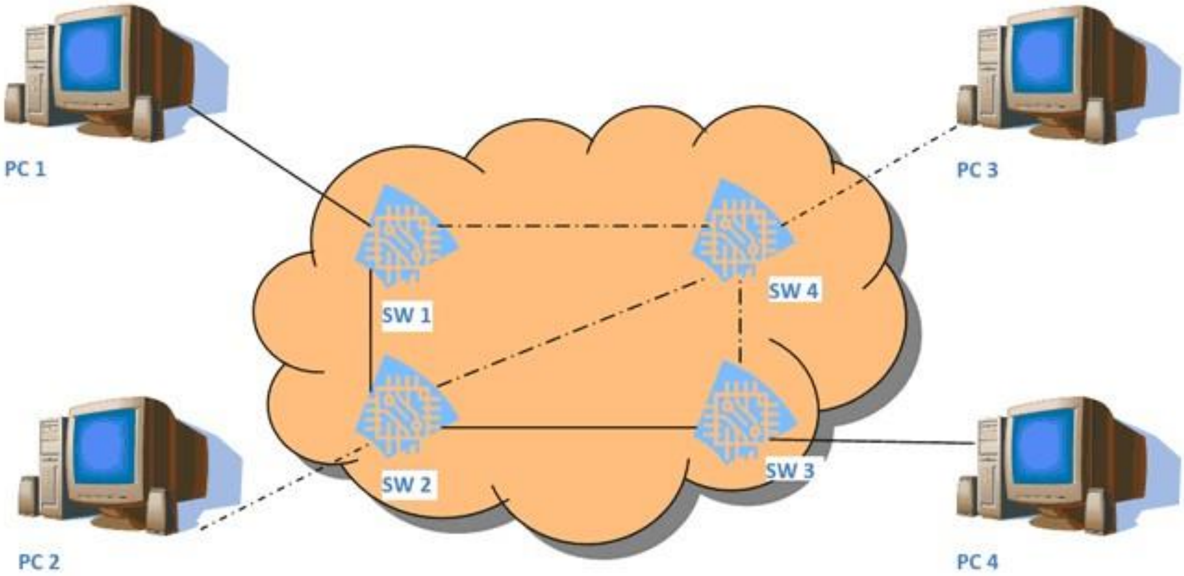
- In this method the data is divided into small pieces of variable lengths and then transmitted to the network line.
- Each individual pieces of data are called as packets.
- After receiving those individual data or packets, all are reassembled at the destination and thus making a complete file.
- Due to this method, the data gets transferred fast and in an efficient manner.
- In this method, no pre-setup or resource reservation is required like circuit switching method.
- This method use Store and Forward techniques.
- So each hop will store the packet first and then forward the packets to the next host destination.
- Each packet contains control information, source address and destination address. Due to this packets can use any route or paths in an existing network.
- If there is congestion at some path, packets are allowed to choose different path possible over existing network.

There are two approaches in the packet switching network

1. Virtual circuit based packet switching
2. Datagram approach

Virtual circuit(VC) based packet switching

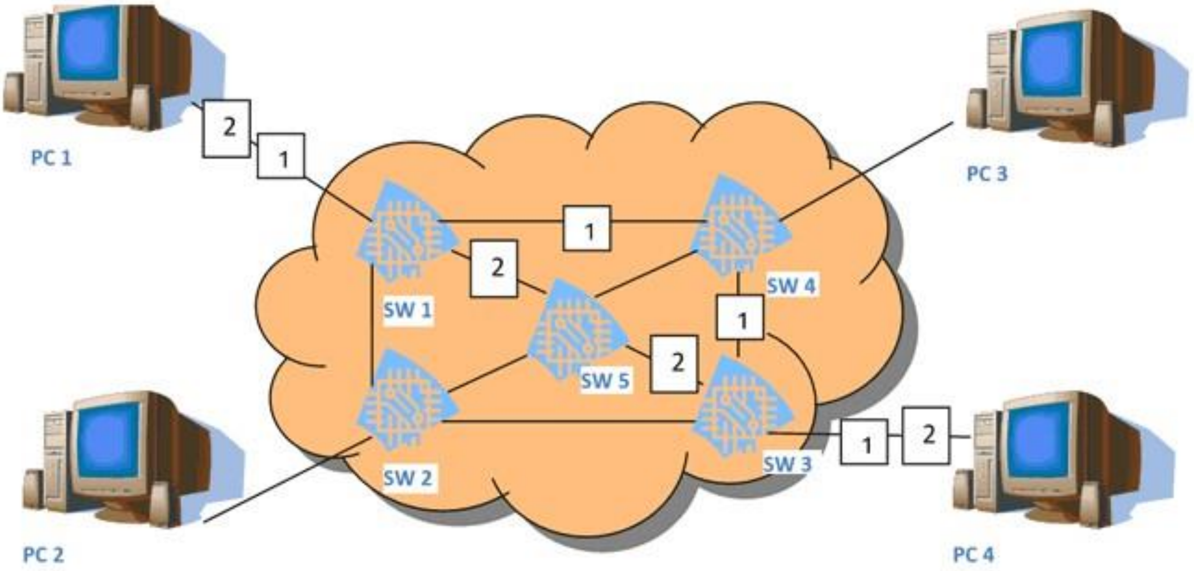
- Virtual circuit based package switching is a mode of packet switching where a logical path or virtual circuit connection is done between sender and receiver.
- VC stands for Virtual Circuit.
- In this mode of packet switching operation, a predefined route is created and all packets will follow the predefined paths.
- All routers or switches which are involved in the logical connection are provided a unique Virtual Circuit ID to uniquely identify the virtual connections.
- It also has the same three-phase protocol used in circuit switching, Setup Phase, Data Transfer Phase and Tear down Phase.



- In the above image, 4 PCs are connected with a 4 switch network and the data flow will be packet switching in Virtual circuit mode.
- As we can see switches are connected with each other and share the communications path with each other.
- Now in the virtual circuit, a predefined route needs to be established. If we want to transfer data from PC1 to the PC 4 the path will be directed from the SW1 to SW2 to SW3 and then finally at PC4.
- This route is predefined and All SW1, SW2, SW3 are provided with a unique ID to identify the data paths, so the data is bound by the paths and could not choose another route.

Datagram approach:

- Datagram switching is completely different from VC based packet switching technology.
- In Datagram switching, the path is dependent on the data.
- The Packets have all the necessary information like Source address, destination address and Port identity etc.
- So in connectionless datagram-based packet switching mode, each packet is treated independently.
- They can choose different routes and the routing decisions are made dynamically when data are transmitting inside the network.
- So, at the destination, the packets can be received out of order or in any sequence, there is no predefined route and the guaranteed packet delivery is not possible. To secure guaranteed packet receiving, additional end system protocols need to be configured.
- In this mode of packet switching, there is no setup, transmit and teardown phase is involved.



 Again in the above image, 4 computers are connected and we transferring data from PC1 to PC4.

- The data contains two packets labeled as 1 and 2. As we can see, in Datagram mode, packet 1 chose to follow the SW1- SW4-SW3 path whereas Packet 2 chose the route path of SW1- SW5- SW3 and finally reached PC4.
- The packets can choose different path depending on the delay time and congestion on other paths in Datagram packet switching network.

Advantages of Packet Switching

Packet switching has following advantages over the circuit switching as given below:

1. Efficient in terms of Bandwidth.
2. Transmission delay is minimum
3. Missing packets can be detected by the destination.
4. Cost-effective implementation.
5. Reliable when busy path or links breakdown is detected in the network. Packets can be transmitted by other links or can use a different path.

Difference between packet switching and circuit switching

Circuit – Switching	Packet – Switching
It is a connection oriented network switching technique.	It is a connectionless network switching technique.
A dedicated path has to be established between the source and the destination before transfer of data commences. Once, the data is transmitted, the path is relinquished.	There is no need to establish a dedicated path from the source to the destination.

It is inflexible in nature since data packets are routed along the same dedicated path.	Each packet is routed separately. Consequently, it is flexible in nature where the different data packets follow different paths.
It was initially designed for voice transfer.	It was initially designed for data transfer.
The entire message is received in the order sent by the source.	The individual packets of the message are received out of order and so need to be reassembled at the destination.
It is implemented at Physical Layer.	It is implemented at Network Layer.
It is not a store and forward transmission.	It is store and forward transmission.
Data is processed and transmitted at the source only.	Data is processed and transmitted, not only at the source but at each switching station.

X.25

X.25 was a standard suite of protocols used for packet-switched communications over a wide area network — a WAN. A protocol is an agreedupon set of procedures and rules. Two devices that follow the same protocols can understand each other and exchange data.

X.25 was developed in the 1970s to carry voice over analog telephone lines — dial-up networks — and is one of the oldest packet-switched services.

Presently, it is used for networks for ATMs and credit card verification.

X.25 Structure:

Each X.25 packet contained up to 128 bytes of data.

The X.25 network handled packet assembly at the source device, the delivery, and the reassembly at the destination.

X.25 has three protocol layers

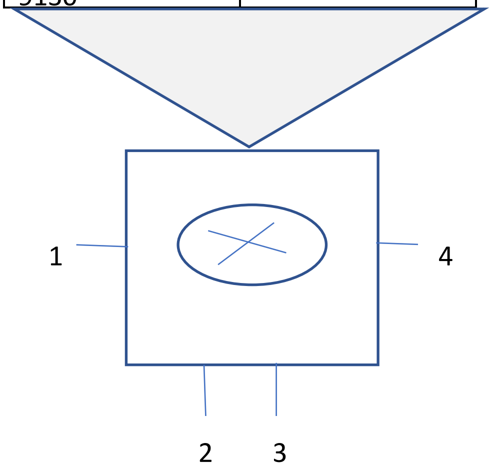
- **Physical Layer:** It lays out the physical, electrical and functional characteristics that interface between the computer terminal and the link to the packet switched node. X.21 physical implementer is commonly used for the linking.

- **Data Link Layer:** It comprises the link access procedures for exchanging data over the link. Here, control information for transmission over the link is attached to the packets from the packet layer to form the LAPB frame (Link Access Procedure Balanced). This service ensures a bit-oriented, error-free, and ordered delivery of frames.
- **Packet Layer:** This layer defines the format of data packets and the procedures for control and transmission of the data packets. It provides external virtual circuit service. Virtual circuits may be of two types: virtual call and permanent virtual circuit. The virtual call is established dynamically when needed through call set up procedure, and the circuit is relinquished through call clearing procedure. Permanent virtual circuit, on the other hand, is fixed and network assigned. Equipment used

- **X.21** implementer
- **DTE**: Data Terminal Equipment
- **DCTE**: Data Circuit Terminating Equipment **Routing in Packet switching:**

- The routing technique is used in packet switched datagram network.
- As there is no setup or teardown phases, how are the packets routed to their destinations in a datagram network?
- For that reason, each switch has a routing table which is based on destination address.
- The routing tables are dynamic and it gets updated periodically.
- The destination addresses and the corresponding output port are updated in the routing table.
- Every packet in the datagram network contains a header which consists of destination address and port number.
- When switch receives a packet, the destination address of the packet is examined.
- Then from the routing table the corresponding port number is being retrieved and packet forwarded.

Destination address	Output port
1232	1
4150	2
.	.
.	.
.	.
9130	3



Network layer Performance

The performance of a network can be measured in terms of delay, throughput and packet loss.

And the congestion control is used to improve the performance.

Network delay:

The amount of time it takes for a packet to go from point A to point B. If Point A is the source and point B is the destination, then the delay is called an end to end delay.

The types of delays encountered in a packet-switched network are:

- Propagation delay
- Transmission delay
- Queuing delay
- Processing delay

Throughput

The number of data is transmitted during a specified time period over a network is called throughput.

Packet Loss:

It is defined as the number of packets lost during transmission.

The causes of packet loss include inadequate signal strength at the destination, natural or human-made interference, excessive system noise, hardware failure, software corruption or overburdened network nodes.

Congestion:

Congestion in a network may occurs due to the load on the network. If the number of packets sent to the network is greater than the capacity of the network which a network can handle, then congestion occurs and it slows down the response time.

Effect of congestion:

- As delay increases, performance decreases.

- If delay increases, retransmission occurs, making situation worse.

Congestion control in packet switching network:
Congestion control refers to the techniques used to control or prevent congestion. Congestion control techniques can be broadly classified into two categories:
i-Open Loop Congestion Control ii-
Closed Loop Congestion Control Open
Loop Congestion Control:

Open loop congestion control policies are applied to prevent congestion before it happens. The congestion control is handled either by the source or the destination.

List of policies used to prevent congestion by Open Loop Congestion Control:

1. Retransmission Policy:

It is the policy in which retransmission of the packets are taken care. If the sender feels that a sent packet is lost or corrupted, the packet needs to be retransmitted. This retransmission may increase the congestion in the network.

To prevent congestion, retransmission policy and timers must be designed to prevent congestion and optimize efficiency.

2. Window Policy:

The type of window at the sender side may also affect the congestion. Several packets in the Go-back-n window are resent, although some packets may be received successfully at the receiver side. This duplication may increase the congestion in the network.

Therefore, Selective repeat window should be adopted as it sends the specific packet that may have been lost.

3. Discarding Policy:

A good discarding policy adopted by the routers is that the routers may prevent congestion and at the same time partially discards the corrupted or less sensitive package and also able to maintain the quality of a message.
In case of audio file transmission, routers can discard less sensitive packets to prevent congestion and also maintain the quality of the audio file.

4. Acknowledgment Policy:

Since acknowledgement are also the part of the load in network, the acknowledgment policy imposed by the receiver may also affect congestion. Several approaches can be used to prevent congestion related to acknowledgment.

The receiver should send acknowledgement for N packets rather than sending acknowledgement for a single packet. The receiver should send a acknowledgment only if it has to sent a packet or a timer expires.

5. Admission Policy:

In admission policy a mechanism should be used to prevent congestion. Switches in a flow should first check the resource requirement of a network flow before transmitting it further. If there is a chance of a congestion or there is a congestion in the network, router should deny establishing a virtual network connection to prevent further congestion.

Closed Loop Congestion Control

Closed loop congestion control technique is used to treat or alleviate congestion after it happens.

Techniques are used by Closed Loop Congestion Control are:

Backpressure:

Backpressure is a technique in which a congested node stop receiving packet from upstream node.

Choke Packet:

Choke packet technique is applicable to both virtual networks as well as datagram subnets. A choke packet is a packet sent by a node to the source to inform it of congestion.

Module-6
LAN Technology

What is LAN protocol architecture?

The LAN architecture consists of three layers: Physical, MAC (Medium Access Control) and LLC (Logical Link Control).

LLC provides connection management, if needed.

MAC is a protocol for accessing high speed physical links and for transferring data frames from one station to another.

LLC:

The logical link control (LLC) is the upper sublayer of the data link layer of the OSI model

It acts act an interface between the network layer and the medium access control (MAC) sublayer of the data link layer. **Functions of LLC Sublayer**

- The primary function of LLC is to multiplex protocols over the MAC layer while transmitting and likewise to de-multiplex the protocols while receiving.
- LLC provides hop-to-hop flow and error control.
- It allows multipoint communication over computer network.
- Frame Sequence Numbers are assigned by LLC.
- In case of acknowledged services, it tracks acknowledgements **Medium Access Control/ Media Access Control:**

MAC is the lower sub layer of the data link layer.

The MAC sublayer acts as an interface between the logical link control (LLC) Ethernet sublayer and physical layer.

A MAC address is a 12-digit hexadecimal number (48 bits in long). MAC addresses are usually written in one of these formats:

- 00:0A:89:5B:F0:11
- 00-0A-89-5B-F0-11 first half of a MAC address contains the ID number of the adapter manufacturer. These IDs are regulated by an Internet standards body. The second half of a MAC address represents the serial number assigned to the adapter by the manufacturer. **Ethernet (CSMA/CD):**

Carrier Sense Multiple Access with Collision Detection (CSMA/CD) is a network protocol for carrier transmission that operates in the Medium Access Control (MAC) layer.
It senses whether the shared channel for transmission is busy or not, and wait for transmissions until the channel is free.

The collision detection technology detects collisions by sensing transmissions from other stations.

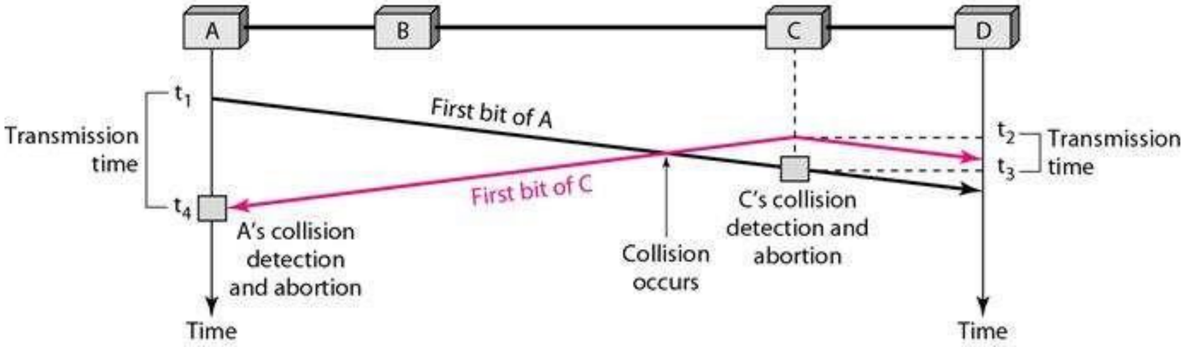
On detection of a collision, the station stops transmitting, sends a jam signal, and then waits for a random time interval before retransmission.

Algorithms

The algorithm of CSMA/CD is:

- When a frame is ready, the transmitting station checks whether the channel is idle or busy.
- If the channel is busy, the station waits until the channel becomes idle.
- If the channel is idle, the station starts transmitting and continually monitors the channel to detect collision.
- If a collision is detected, the station starts the collision resolution algorithm.
- The station resets the retransmission counters and completes frame transmission

Ex: let us look at the first bits transmitted by the two stations A and C in the figure given below are involved in the collision.



- At time t 1, station A has executed its persistence procedure and starts sending the bits of its frame.

- At time t2, station C has not yet sensed the first bit sent by A. Station C executes its persistence procedure and starts sending the bits in its frame, which propagate both to the left and to the right.

- The collision occurs sometime after time t2' Station C detects a collision at time t3 when it receives the first bit of A's frame. Station C immediately (or after a short time, but we assume immediately) aborts transmission.
- Station A detects collision at time t4 when it receives the first bit of C's frame; it also immediately aborts transmission.

- Looking at the figure, we see that A transmits for the duration t4 – t1. C transmits for the duration t3 - t2. The protocol to work, the length of any frame divided by the bit rate in this protocol must be more than either of these durations. At time t4, the transmission of A’s frame, though incomplete, is aborted. At time t3, the transmission of C's frame, though incomplete, is aborted.

Fibre Channel

Fibre Channel is a high-speed networking technology primarily used for transmitting data among data centres, computer servers, switches and storage at data rates of up to 128 Gbps. It was developed to overcome the shortcomings of the Small Computer System Interface (SCSI) and High-Performance Parallel Interface (HIPPI) by filling the need for a reliable and scalable high-throughput and low-latency protocol and interface.

→Fibre Channel is especially suited for connecting servers to shared storage devices and interconnecting storage controllers and drives.

Network Connecting devices:

Hub

Hub is a centralized device that connects multiple devices in a single LAN network. When Hub receives the data signals from a connected device on any of its port, except that port, it forwards those signals to all other connected devices from the remaining ports.

→It operates only in the physical layer.
There are two types of the Hub.
Passive Hub: - It forwards data signals in the same format in which it receives them. It does not change the data signal in any manner.

Active Hub: - It also works same as the passive Hub works. But before forwarding the data signals, it amplifies them. Due to this added feature, the active Hub is also known as the repeater.

Bridge
Bridge is used to divide a large network into smaller segments.

Basic functions of the Bridge are the following: -
Breaking a large network into smaller segments.

- Connecting different media types. Such as connects UTP with the fiber optic.
- Connecting different network architectures. Such as connects Ethernet with the Token ring.

Switch
Switch is a network device that connects other devices to Ethernet networks through twisted pair cables. It uses packet switching technique to receive, store and forward data packets on the network. The switch maintains a list of network addresses of all the devices connected to it.
→On receiving a packet, it checks the destination address and transmits the packet to the correct port. Before forwarding, the packets are checked for collision and other network errors. The data is transmitted in full duplex mode →It operates both in physical and datalink layers.

Router
Router is a device that forwards data packets along networks. A router is connected to at least two networks, commonly two LANs or WANs or a LAN and its ISP's network. Routers are located at gateways, the places where two or more networks connect.
→Routers use headers and forwarding tables to determine the best path for forwarding the packets, and they use protocols such as ICMP to communicate with each other and configure the best route between any two hosts.
→The router is three layer device. It operates in the physical, datalink and network layers.

Wireless LAN Technology:

Wireless LANs are those Local Area Networks that use high frequency radio waves instead of cables for connecting the devices in LAN. Users connected by WLANs can move around within the area of network coverage. Most WLANs are based upon the standard IEEE 802.11 or WiFi.

IEEE 802.11(Wireless Ethernet) Architecture

The components of an IEEE 802.11 architecture are as follows

1) Stations (STA): Stations comprise all devices and equipments that are connected to the wireless LAN. A station can be of two types:
   1. Wireless Access Point (WAP): WAPs or simply access points (AP) are generally wireless routers that form the base stations or access.
   2. Client: Clients are workstations, computers, laptops, printers, smartphones, etc.

Each station has a wireless network interface controller.

2) Basic Service Set (BSS): A basic service set is a group of stations communicating at physical layer level. BSS can be of two categories depending upon mode of operation:

   1. Infrastructure BSS: Here, the devices communicate with other devices through access points.
   2. Independent BSS: Here, the devices communicate in peer-to-peer basis in an ad hoc manner.

3) Extended Service Set (ESS): It is a set of all connected BSS.
4) Distribution System (DS): It connects access points in ESS.

Advantages of WLANs

1. They provide clutter free homes, offices and other networked places.
2. The LANs are scalable in nature, i.e. devices may be added or removed from the network at a greater ease than wired LANs.
3. The system is portable within the network coverage and access to the network is not bounded by the length of the cables.
4. Installation and setup is much easier than wired counterparts.
5. The equipment and setup costs are reduced.

Disadvantages of WLANs

1. Since radio waves are used for communications, the signals are noisier with more interference from nearby systems.
2. Greater care is needed for encrypting information. Also, they are more prone to errors. So, they require greater bandwidth than the wired LANs.
3. WLANs are slower than wired LANs.

Module-7

TCP/IP Internet

Protocol:

IP is the main protocol of network layer.

Internet protocol version 4(IPv4) is responsible for packetizing, forwarding and delivery of a packet at the network layer.

IPv4 is an unreliable datagram protocol- a best effort delivery service. The term best-effort means the IPv4 packets can be corrupted, be lost, arrive out of order, be delayed and may create congestion for the network.

For reliable service the IPv4 must be paired with the TCP.

IPv4 is also a connectionless protocol that uses the datagram approach. In this approach each datagram is handled independently and each datagram can follow a different route to reach the destination.

That means the datagram send by the source to destination may arrive out of order.

IP datagram format:

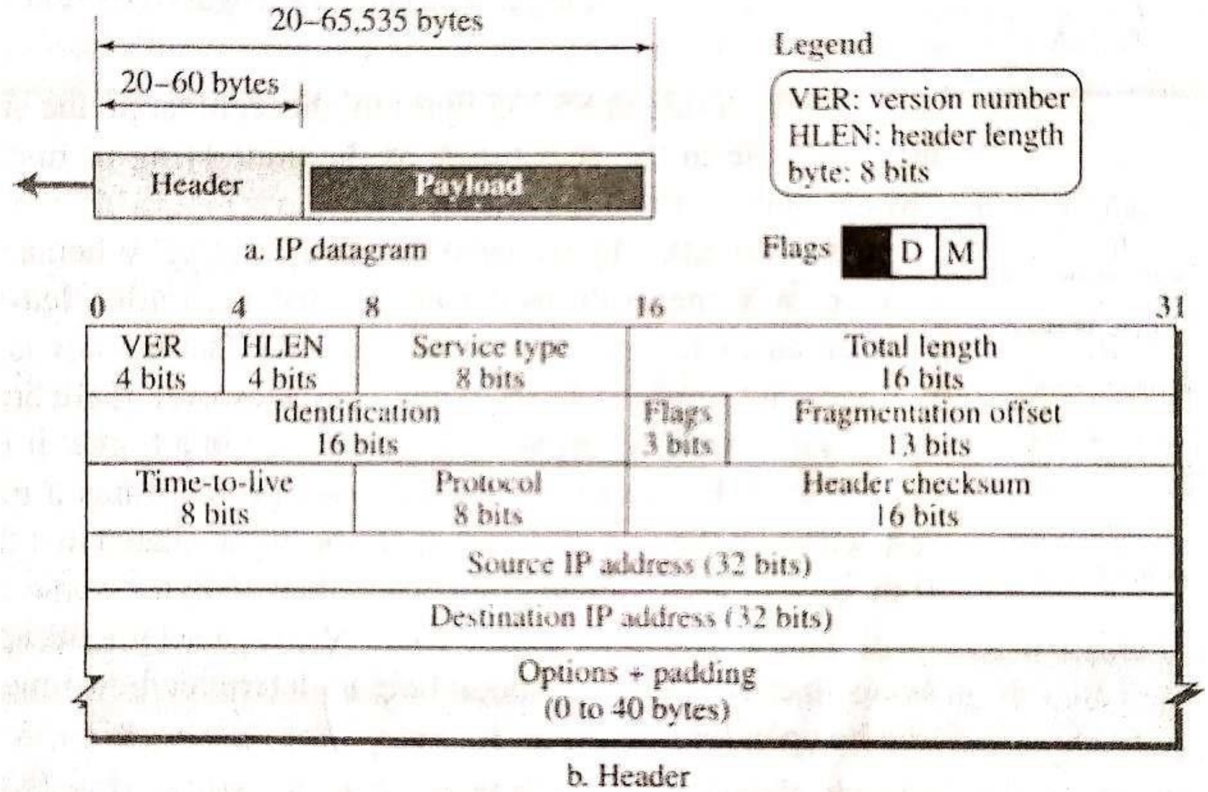
Packets used by the IP are also called datagrams.

A datagram is a variable-length packet consisting of two parts i.e header and payload(data).

IP Datagram Header:

The size of the header is 20 to 60 bytes and contains information on routing and delivery.

Due to the presence of options, the size of the datagram header can be of variable length (20 bytes to 60 bytes).



a. IP datagram

0	4	8	16	31
VER 4 bits	HLEN 4 bits	Service type 8 bits	Total length 16 bits	
Identification 16 bits			Flags 3 bits	Fragmentation offset 13 bits
Time-to-live 8 bits		Protocol 8 bits	Header checksum 16 bits	
Source IP address (32 bits)				
Destination IP address (32 bits)				
Options + padding (0 to 40 bytes)				

b. Header

IP header includes many relevant information including Version Number, which, in this context, is 4. Other details are as follows − □ **Version:** Version no. of Internet Protocol used (e.g. IPv4).

- **HLEN:** IP header length (4 bits), which is the number of 32-bit words in the header. The minimum value for this field is 5 and the maximum is 15.
- **Type of service:** Low Delay, High Throughput, Reliability (8 bits)

- **Total Length:** This 16-bit field defines the total length i.e header plus data of the IP datagram in bytes. A 16-bit number can define a total length up to 65,535 bytes of data. This field helps the receiving device to know when the packet has completely arrived.

  Length of data=total length-(HLEN) *4

- **Identification:** If IP packet is fragmented during the transmission, all the fragments contain same identification number. to identify original IP packet, they belong to.
- **Flags:** As required by the network resources, if IP Packet is too large to handle, these ‘flags’ tells if they can be fragmented or not. In this 3-bit flag, the MSB is always set to ‘0’.
- **Fragment Offset:** This offset tells the exact position of the fragment in the original IP Packet.
- **Time to Live:** This TTL field is used to control the maximum number of routers visited by the datagram. To avoid looping in the network, every packet is sent with some TTL value set, which tells the network how many routers (hops) this packet can visit. This value is approximately two times the number of routers between any two host. At each hop(router), its value is decremented by one and when the value reaches zero, the packet is discarded.
- **Protocol:** Tells the Network layer at the destination host, to which Protocol this packet belongs to, i.e. the next level Protocol. For example protocol number of ICMP is 1, TCP is 6 and UDP is 17.
- **Header Checksum:** This field is used to keep checksum value of entire header which is then used to check if the packet is received error-free.
- **Source Address:** 32-bit address of the Sender (or source) of the packet.
- **Destination Address:** 32-bit address of the Receiver (or destination) of the packet.
- **Options:** This is optional field, which is used if the value of IHL is greater than 5. These options may contain values for options such as Security, Record Route, Time Stamp, etc.

ICMP:

ICMP stands for Internet Control Message Protocol.

The Ipv4 has no error reporting and error correcting mechanism.
It depends on Internet Control Message Protocol (ICMP) to provide an error control. It is used for reporting errors and management queries. ICMP messages are divided into two categories:

Error Reporting messages and query messages

The error reporting messages report the problems that a router or a host may encounter when it processes an IP packet.

The query messages helps a host or network manager to get specific information from router.

ICMP message has an 8-byte header and a variable size data section.

IGMP:

IGMP stands for Internet Group Message Protocol.

It's used to establish host memberships in particular multicast groups on a single network.

IPv4 Addresses

The identifier used in the IP layer of the TCP/IP protocol suite to identify the connection of each device to the internet is called the Internet Address or IP address.

Definition:

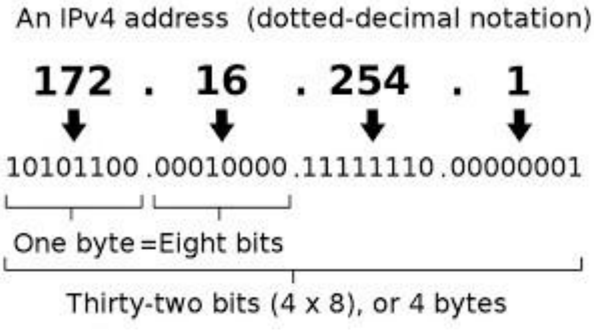
An IPv4 address is a 32-bit address that uniquely and universally identifies the connection of a host or a router to the Internet.

The IP address is the address of the connection not the host or router, because if the device is moved to another network the IP address may be changed.

Notation of IPv4 Address:

In binary notation an IPv4 address is displayed as 32 bits and each octet is separated by one or more spaces or dot as shown below.

For easier to read it is written in decimal form with a decimal point(dot) separating the number which is called as dotted decimal notation is given below.


An IPv4 address (dotted-decimal notation)
172 . 16 . 254 . 1
10101100 .00010000 .11111110 .00000001
One byte =Eight bits
Thirty-two bits (4 x 8), or 4 bytes

This IPv4 Address is divided into two parts.
The first part of the address is called prefix which is used to define the network id and second part of the address is called suffix which is used to define host id.
The length is n bits and suffix length is (32-n) bits.

Classful Addressing

The 32-bit IP address is divided into five sub-classes. These are:

- Class A
- Class B
- Class C
- Class D

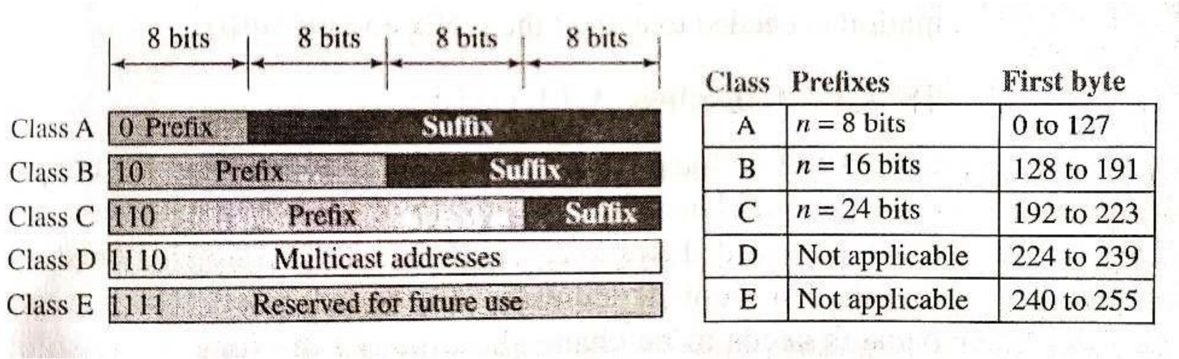
- Class E

Each of these classes has a valid range of IP addresses. Classes D and E are reserved for multicast and experimental purposes respectively. The order of bits in the first octet determine the classes of IP address.

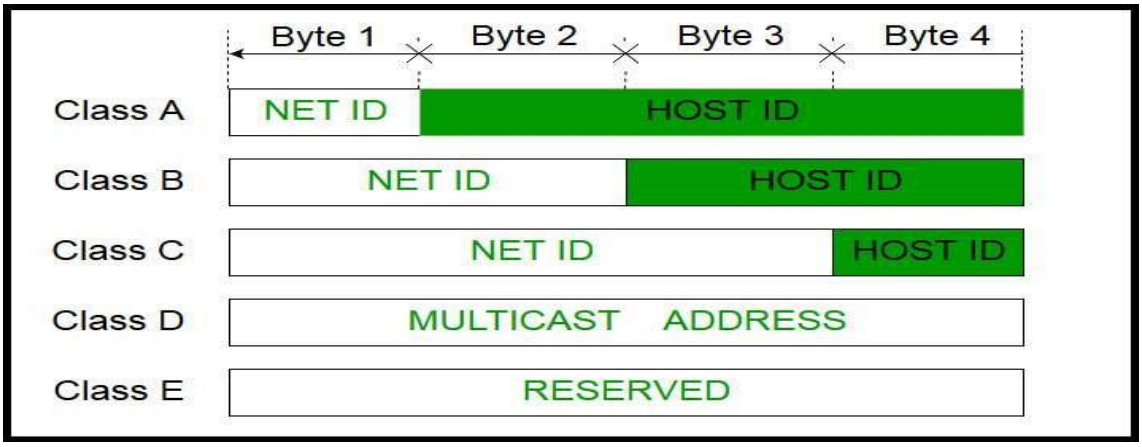
IPv4 address is divided into two parts:

- Network ID
- Host ID

The class of IP address is used to determine the bits used for network ID and host ID and the number of total networks and hosts possible in that particular class. Each ISP or network administrator assigns IP address to each device that is connected to its network.



Class	Prefixes	First byte
A	n = 8 bits	0 to 127
B	n = 16 bits	128 to 191
C	n = 24 bits	192 to 223
D	Not applicable	224 to 239
E	Not applicable	240 to 255



Note: IP addresses are globally managed by Internet Assigned Numbers Authority (IANA) and regional Internet registries (RIR).

Note: While finding the total number of host IP addresses, 2 IP addresses are not counted and are therefore, decreased from the total count because the first IP address of any network is the network number and whereas the last IP address is reserved for broadcast IP.

Class A:

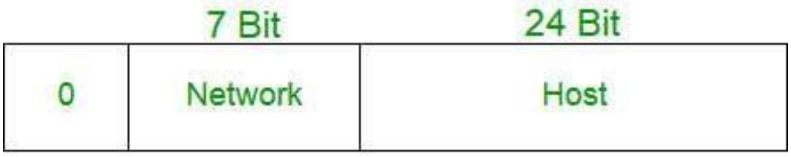
IP address belonging to class A are assigned to the networks that contain a large number of hosts.

- The network ID is 8 bits long.
- The host ID is 24 bits long.

The higher order bit of the first octet in class A is always set to 0. The remaining 7 bits in first octet are used to determine network ID. The 24 bits of host ID are used to determine the host in any network. The default subnet mask for class A is 255.x.x.x. Therefore, class A has a total of:

- 2^7-2= 126 network ID (**Here 2 address is subtracted because 0.0.0.0 and 127.x.y.z are special address**.)
- 2^24 – 2 = 16,777,214 host ID

IP addresses belonging to class A ranges from 1.x.x.x – 126.x.x.x.



Class A

Class B:

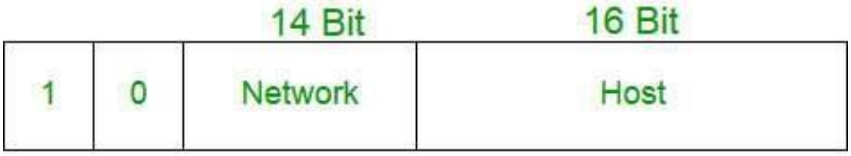
IP address belonging to class B are assigned to the networks that ranges from medium-sized to large-sized networks.

- The network ID is 16 bits long.
- The host ID is 16 bits long.

The higher order bits of the first octet of IP addresses of class B are always set to 10. The remaining 14 bits are used to determine network ID. The 16 bits of host ID is used to determine the host in any network. The default sub-net mask for class B is 255.255.x.x. Class B has a total of:

- 2^14 = 16384 network address • 2^16 – 2 = 65534 host address

IP addresses belonging to class B ranges from 128.0.x.x – 191.255.x.x.



Class B

Class C:

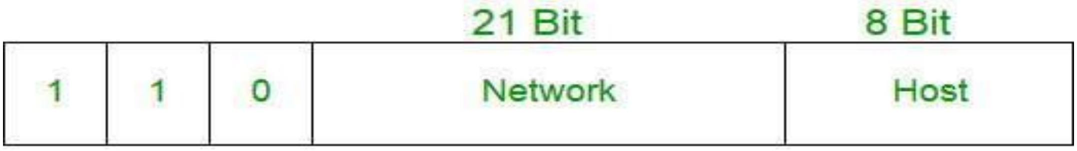
IP address belonging to class C are assigned to small-sized networks.

- The network ID is 24 bits long.
- The host ID is 8 bits long.

The higher order bits of the first octet of IP addresses of class C are always set to 110. The remaining 21 bits are used to determine network ID. The 8 bits of host ID is used to determine the host in any network. The default sub-net mask for class C is 255.255.255.x. Class C has a total of:

- 2^21 = 2097152 network address
- 2^8 – 2 = 254 host address

IP addresses belonging to class C ranges from 192.0.0.x – 223.255.255.x.

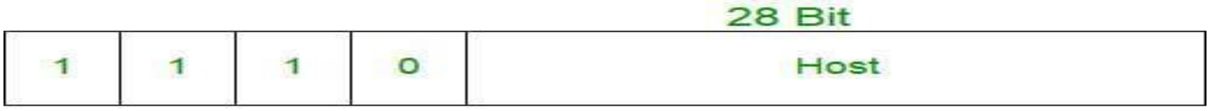


Class C

Class D:

IP address belonging to class D are reserved for multi-casting. The higher order bits of the first octet of IP addresses belonging to class D are always set to 1110. The remaining bits are for the address that interested hosts recognize.

Class D does not posses any sub-net mask. IP addresses belonging to class D ranges from 224.0.0.0 – 239.255.255.255.

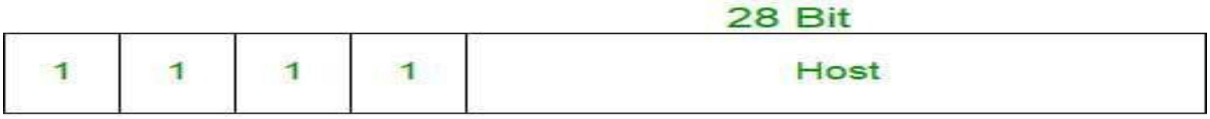


				28 Bit
1	1	1	0	Host

Class D

Class E:

IP addresses belonging to class E are reserved for experimental and research purposes. IP addresses of class E ranges from 240.0.0.0 – 255.255.255.254. This class doesn’t have any sub-net mask. The higher order bits of first octet of class E are always set to 1111.



				28 Bit
1	1	1	1	Host

Class E

Model Questions (2 Marks, 5 Marks and 10 marks)

Unit-1

2 Marks:

1. Define Computer Network?
2. Define Internet?
3. What is ARPANET?
4. What do mean by network topology?
5. What is a network? And what are the benefits of the networks?
6. What do mean by half-duplex?
7. Define multiplexing and demultiplexing?
8. What are the different transmission media?

5 Marks:

9. Explain the types of transmission modes.
10. What is network topology? Explain the different network topologies.
11. What are the different types of networks? Explain in detail.
12. What are the applications of Computer Networks?

10 Marks:

13. Explain the OSI reference model with neat diagram.
14. 5.Explain the TCP/IP reference model with neat diagram
15. What is OSI Model? Explain the functions and protocols and services of each layer?
16. Briefly explain different types of transmission Medias in computer networks?

Unit-2

2 Marks:

1. Define Analog and Digital signal?
2. Define frequency?
3. What is Composite signal?
4. What is baud rate?
5. Define bandwidth?
6. What is Attenuation?
7. Define Jitter?
8. Define Line coding?
9. What is data rate and signal rate?
10. Define Block coding?

5 Marks:

11. Describe Manchester and Differential Manchester scheme?

10 Marks:

12. Briefly explain various Line coding schemes?

Unit-3

2 Marks:

1. What do mean by MODEM?
2. What is Sampling?

5 Marks:

3. Explain AM, FM and PM
4. Describe the Pulse Code Modulation (PCM) technique with neat diagram?
5. Explain Delta Modulation (DM) technique with neat diagram?
6. Describe Frequency shift keying?

10 Marks:

7. Briefly explain the different data transmission modes (Parallel and Serial)?
8. Explain ASK, FSK, and PSK with neat diagram?

Unit-4

2 Marks:

1. Define Multiplexing ?
2. Define checksum?
3. What is an error?
4. Define framing?
5. List out the Protocols in Datalink layer.
6. What is Piggybacking
7. What are the flow characteristics related to QoS?
8. Define flow control?
9. What is FDM?
10. What do mean by redundancy?
11. Define Parity Check?
12. Define CRC?
13. What is Hamming code?
14. What do mean by error control?

5 Marks:

15. What is an error ? Explain the types of errors ?
16. Explain the Point to Point Protocol in detail.
17. Explain the various error detection methods?

18. Briefly explain the CRC error detection scheme ?
19. Explain about Sliding Window Protocols?

10 Marks:

20. Define Multiplexing. Explain the various types of multiplexing?

Unit-5

2 Marks:

1. What is Switching?
2. Define congestion?
3. What do mean by congestion control?
4. What is datagram?
5. What is packet fragmentation?

5 Marks:

6. Define Switching and Briefly explain the Datagram Approach in Packet Switching Method with neat diagram?
7. Differentiate packet switching and circuit switching?

10 Marks:

8. Write short notes on circuit switching, packet switching and message switching.

Unit-6

2 Marks:

1. What are the two sublayers of Data link layer?
2. What do mean by wireless LAN?
3. Define WWW ?
4. What is throughput?
5. What is pure ALOHA ?
6. What is Media Access Control?
7. What is Router?
8. What is CSMA/CD?
9. What do mean by carrier sense multiple access?

5 Marks:

10. What is transmission media? Explain different forms of transmission media?
11. Explain about the different types of connecting devices in computer networks?

10 Marks:

12. Briefly describe the types of transmission media?
13. Briefly explain the CSMA/CD method with neat diagram?
14. Define computer networks? Discuss various types of networks topologies in computer network with its advantages and disadvantages.
15. Define the following terms
    - i- Hub
    - ii- Switch
    - iii- Router
    - iv- Bridge
    - v- Gateway

Unit-7

2 Marks:

1. What is POP?
2. Define HTTP?
3. What is DNS?
4. Define ICMP?
5. What are the types of classful addressing?
6. What do mean by ARP?
7. Define unicast and Multicast?
8. Define Ip address?
9. Define TCP?
10. Define UDP?

5 Marks:

11. Briefly Explain TCP protocol?
12. Write short notes on
    - i- TCP
    - ii- UDP
13. Briefly describe the classful addressing?

10 Marks:

14. Briefly explain IP datagram header format with neat diagram?