UNIT - I

Introduction to Data Communications:

In Data Communications, *data* generally are defined as information that is stored in digital form. *Data communications* is the process of transferring digital information between two or more points. *Information* is defined as the knowledge or intelligence. Data communications can be summarized as the transmission, reception, and processing of digital information. 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: delivery, accuracy, timeliness, and jitter.

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.
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.

Standards Organizations for Data Communications

An association of organizations, governments, manufacturers and users form the standards organizations and are responsible for developing, coordinating and maintaining the standards. The intent is that all data communications equipment manufacturers and users comply with these standards. The primary standards organizations for data communication are:

1. International Standard Organization (ISO)

ISO is the international organization for standardization on a wide range of subjects. It is comprised mainly of members from the standards committee of various governments throughout the world. It is even responsible for developing models which provides high level of system compatibility, quality enhancement, improved productivity and reduced costs. The ISO is also responsible for endorsing and coordinating the work of the other standards organizations.

2. International Telecommunications Union-Telecommunication Sector (ITU-T)

ITU-T is one of the four permanent parts of the International Telecommunications Union based in Geneva, Switzerland. It has developed three sets of specifications: the *V series* for modem interfacing and data transmission over telephone lines, the *X series* for data transmission over public digital networks, email and directory services; the *I and Q series* for Integrated Services Digital Network (ISDN) and its extension Broadband ISDN. ITU-T membership consists of government authorities and representatives from many countries and it is the present standards organization for the United Nations.

3. **Institute of Electrical and Electronics Engineers (IEEE)**

IEEE is an international professional organization founded in United States and is compromised of electronics, computer and communications engineers. It is currently the world's largest professional society with over 200,000 members. It develops communication and information processing standards with the underlying goal of advancing theory, creativity, and product quality in any field related to electrical engineering.

4. **American National Standards Institute (ANSI)**

ANSI is the official standards agency for the United States and is the U.S voting representative for the ISO. ANSI is a completely private, non-profit organization comprised of equipment manufacturers and users of data processing equipment and services. ANSI membership is comprised of people form professional societies, industry associations, governmental and regulatory bodies, and consumer goods.

5. **Electronics Industry Association (EIA)**

EIA is a non-profit U.S. trade association that establishes and recommends industrial standards. EIA activities include standards development, increasing public awareness, and lobbying and it is responsible for developing the RS (recommended standard) series of standards for data and communications.

6. **Telecommunications Industry Association (TIA)**

TIA is the leading trade association in the communications and information technology industry. It facilitates business development opportunities through market development, trade promotion, trade shows, and standards development. It represents manufacturers of communications and information technology products and also facilitates the convergence of new communications networks.

7. **Internet Architecture Board (IAB)**

IAB earlier known as Internet Activities Board is a committee created by ARPA (Advanced Research Projects Agency) so as to analyze the activities of ARPANET whose purpose is to accelerate the advancement of technologies useful for U.S military. IAB is a technical advisory group of the Internet Society and its responsibilities are:

I. Oversees the architecture protocols and procedures used by the Internet.
II. Manages the processes used to create Internet Standards and also serves as an appeal board for complaints regarding improper execution of standardization process.
III. Responsible for administration of the various Internet assigned numbers
IV. Acts as a representative for Internet Society interest in liaison relationships with other organizations.
V. Acts as a source of advice and guidance to the board of trustees and officers of Internet Society concerning various aspects of internet and its technologies.

8. **Internet Engineering Task Force (IETF)**

   The IETF is a large international community of network designers, operators, vendors and researchers concerned with the evolution of the Internet architecture and smooth operation of the Internet.

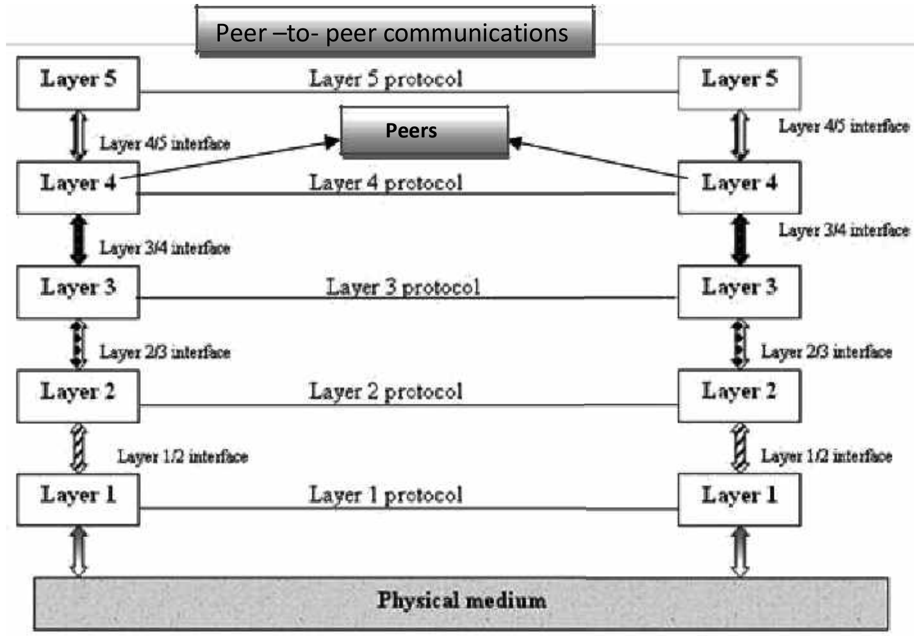
9. **Internet Research Task Force (IRTF)**

   The IRTF promotes research of importance to the evolution of the future Internet by creating focused, long-term and small research groups working on topics related to Internet protocols, applications, architecture and technology.

Layered Network Architecture

To reduce the design complexity, most of the networks are organized as a series of **layers** or **levels**, each one build upon one below it. The basic idea of a layered architecture is *to divide the design into small pieces*. Each layer adds to the services provided by the lower layers in such a manner that the highest layer is provided a full set of services to manage communications and run the applications. The benefits of the layered models are modularity and clear interfaces, i.e. open architecture and comparability between the different providers' components. A basic principle is to ensure independence of layers by defining services provided by each layer to the next higher layer without defining how the services are to be performed. This permits changes in a layer without affecting other layers. The basic elements of a layered model are services, protocols and interfaces. A ***service*** is a set of actions that a layer offers to another (higher) layer. ***Protocol*** is a set of rules that a layer uses to exchange information with a peer entity. These rules concern both the contents and the order of the messages used. Between the layers service interfaces are defined. The messages from one layer to another are sent through those interfaces.

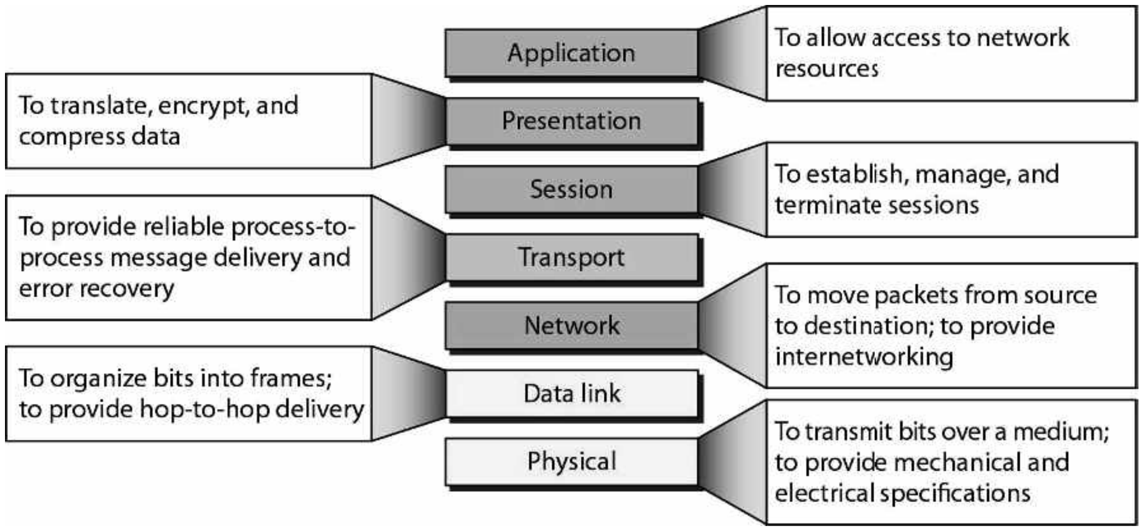
In a *n-layer* architecture, layer n on one machine carries on conversation with the layer n on other machine. The rules and conventions used in this conversation are collectively known as the *layer-n protocol*. Basically, a protocol is an agreement between the communicating parties on how communication is to proceed. Five-layer architecture is shown below**;** the entities comprising the corresponding layers on different machines are called ***peers***. In other words, it is the peers that communicate using protocols. In reality, no data is transferred from layer n on one machine to layer n of another machine. Instead, each layer passes data and control information to the layer immediately below it, until the lowest layer is reached. Below layer-1 is the physical layer through which actual communication occurs.



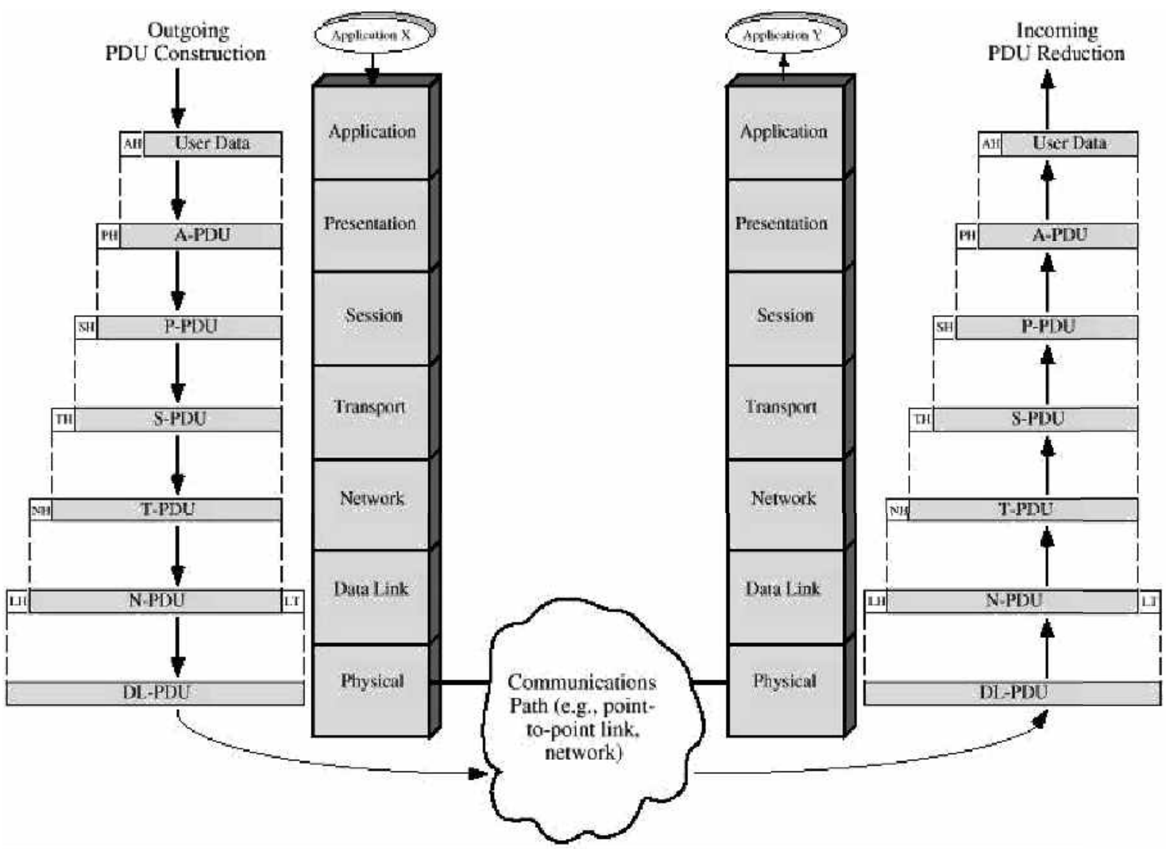
With layered architectures, communications between two corresponding layers requires a unit of data called a ***protocol data unit (PDU)***. A PDU can be a header added at the beginning of a message or a trailer appended to the end of a message. Data flows downward through the layers in the source system and upwards at the destination address. As data passes from one layer into another, headers and trailers are added and removed from the PDU. This process of adding or removing PDU information is called ***encapsulation/decapsulation.*** Between each pair of adjacent layers there is an ***interface***. The *interface* defines which primitives operations and services the lower layer offers to the upper layer adjacent to it. A set of layers and protocols is known as **network architecture**. A list of protocols used by a certain system, one protocol per layer, is called **protocol stack**.

Open Systems Interconnection (OSI)

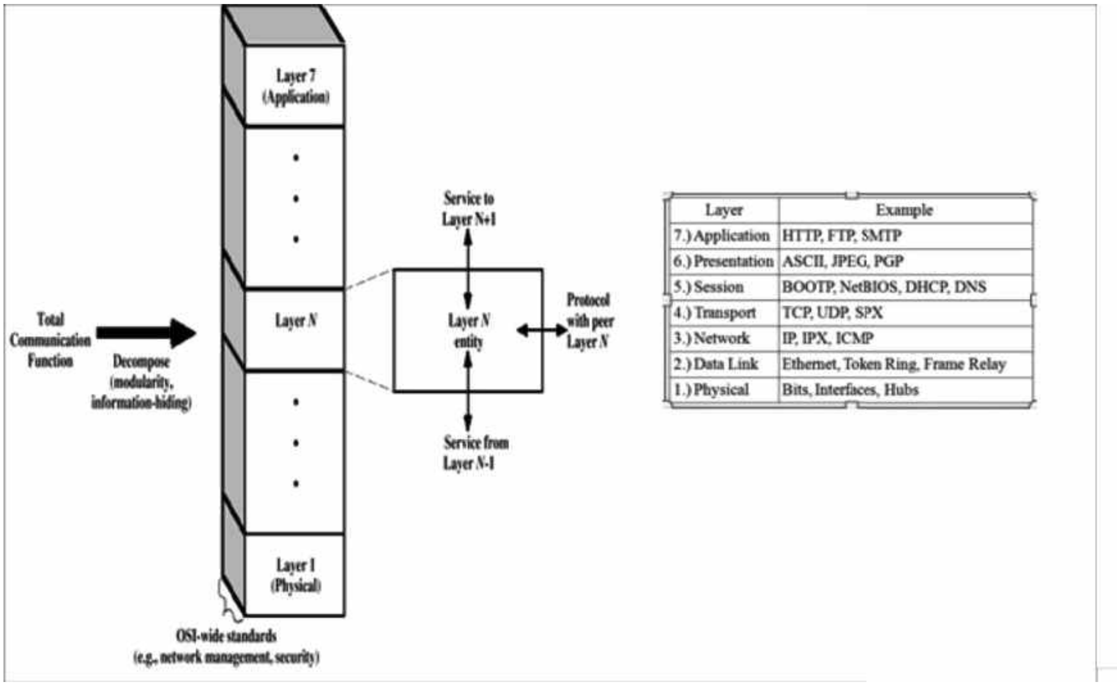
International standard organization (ISO) established a committee in 1977 to develop architecture for computer communication and the OSI model is the result of this effort. In 1984, the Open Systems Interconnection (OSI) reference model was approved as an international standard for communications architecture. The term "*open*" denotes the ability to connect any two systems which conform to the reference model and associated standards. The OSI model describes how information or data makes its way from application programmes (such as spreadsheets) through a network medium (such as wire) to another application programme located on another network. The OSI reference model divides the problem of moving information between computers over a network medium into **SEVEN** smaller and more manageable problems. The seven layers are:



The lower 4 layers (transport, network, data link and physical —Layers 4, 3, 2, and 1) are concerned with the flow of data from end to end through the network. The upper four layers of the OSI model (application, presentation and session—Layers 7, 6 and 5) are orientated more toward services to the applications. Data is Encapsulated with the necessary protocol information as it moves down the layers before network transit.

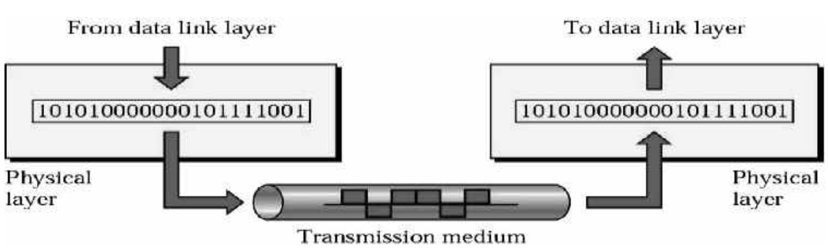


As with any layered architecture, overhead information is added to a PDU in the form of headers and trailers. Each layer provides a service to the layer above it in the protocol specification. Each layer communicates with the same layer's software or hardware on other computers.



Physical Layer {***the physical layer is responsible for transmitting individual bits from one node to the next}***

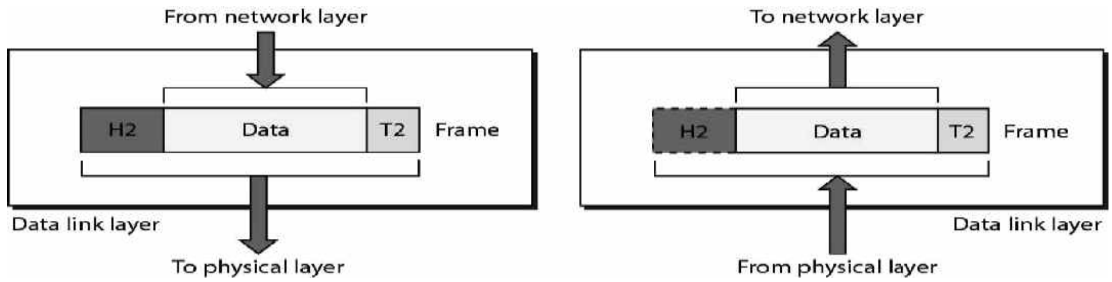
The physical layer is the lowest layer of the OSI hierarchy and coordinates the functions required to transmit a bit stream over a physical medium. It also defines the procedures and functions that physical devices and interfaces have to perform for transmission occur. The physical layer specifies the type of transmission medium and the transmission mode (simplex, half duplex or full duplex) and the physical, electrical, functional and procedural standards for accessing data communication networks.



Transmission media defined by the physical layer include metallic cable, optical fiber cable or wireless radio-wave propagation. The physical layer also includes the *carrier system* used to propagate the data signals between points in the network. The carrier systems are simply communication systems that carry data through a system using either metallic or optical fiber cables or wireless arrangements such as microwave, satellites and cellular radio systems.

Data-link Layer {***the data link layer is responsible for transmitting frames from one node to the next}***

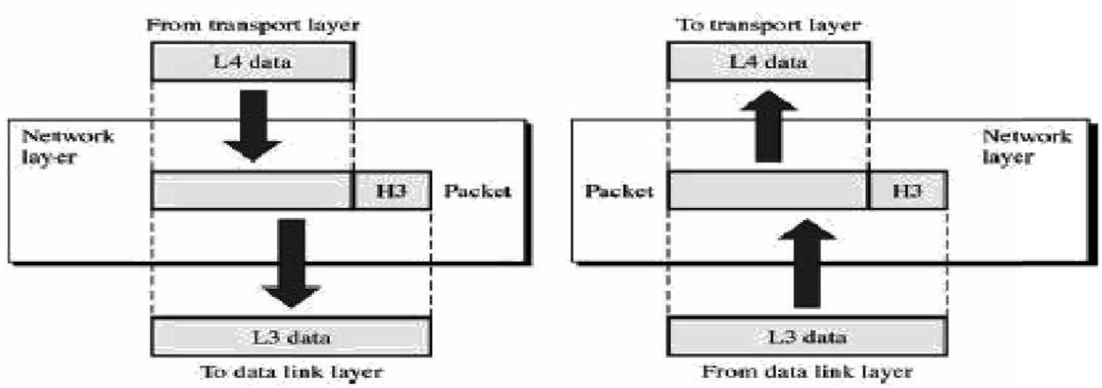
The data link layer transforms the physical layer, a raw transmission facility, to a reliable link and is responsible for node-to-node delivery. It makes the physical layer appear error free to the upper layer (network layer).



The data link layer packages data from the physical layer into groups called blocks, frames or packets. 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 physical address of the sender (source address) and/or receiver (destination address) of the frame. The data-link layer provides flow-control, access-control, and error-control.

Network Layer **{is responsible for the delivery of individual packets from the source host to the destination host}**

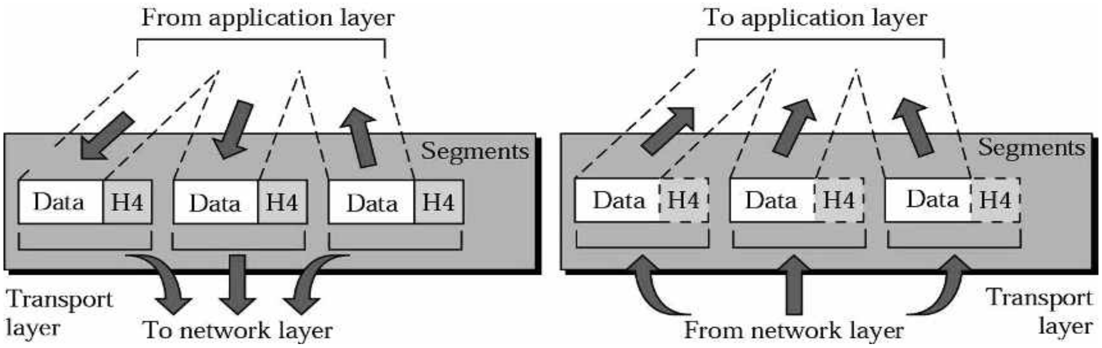
The network layer provides details that enable data to be routed between devices in an environment using multiple networks, subnetworks or both. This is responsible for addressing messages and data so they are sent to the correct destination, and for translating logical addresses and names (like a machine name FLAME) into physical addresses. This layer is also responsible for finding a path through the network to the destination computer.



The network layer provides the upper layers of the hierarchy with independence from the data transmission and switching technologies used to interconnect systems. Networking components that operate at the network layer include routers and their software.

Transport Layer {***is responsible for delivery of a message from one process to another}***

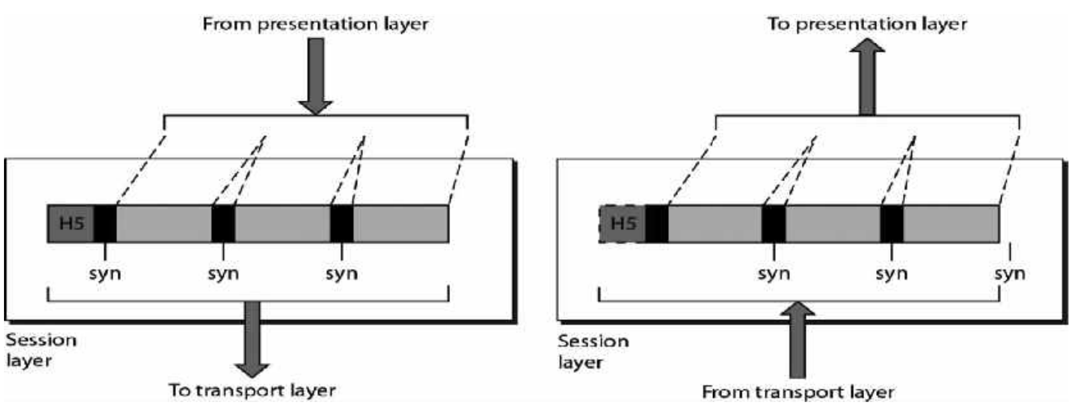
The transport layer controls and ensures the end-to-end integrity of the data message propagated through the network between two devices, providing the reliable, transparent transfer of data between two endpoints.



Transport layer responsibilites includes message routing, segmenting, error recovery and two types of basic services to an upper-layer protocol: connection oriented and connectionless. The transport layer is the highest layer in the OSI hierarchy in terms of communicatons and may provide data tracking, connection flow control, sequencing of data, error checking, and application addressing and identification.

Session Layer {**responsible for dialog control and synchronization**}

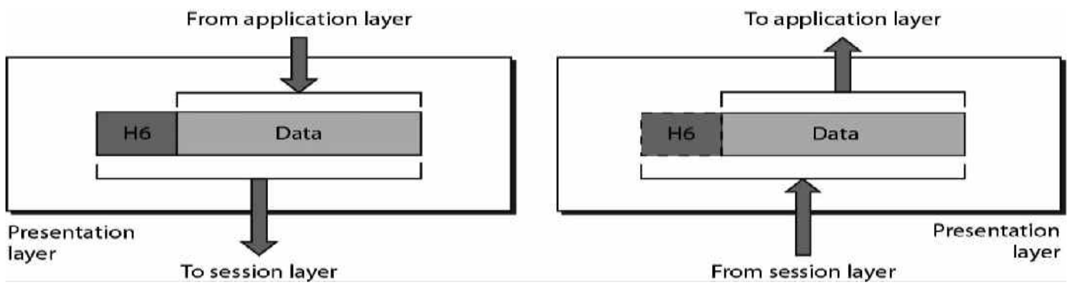
Session layer, some times called the dialog controller provides mechanism for controlling the dialogue between the two end systems. It defines how to start, control and end conversations (called sessions) between applications.



Session layer protocols provide the logical connection entities at the application layer. These applications include file transfer protocols and sending email. Session responsibilities include network log-on and log-off procedures and user authentication. Session layer characteristics include virtual connections between applications, entities, synchronization of data flow for recovery purposes, creation of dialogue units and activity units, connection parameter negotiation, and partitioning services into functional groups.

Presentation Layer {**responsible for translation, compression, and encryption**}

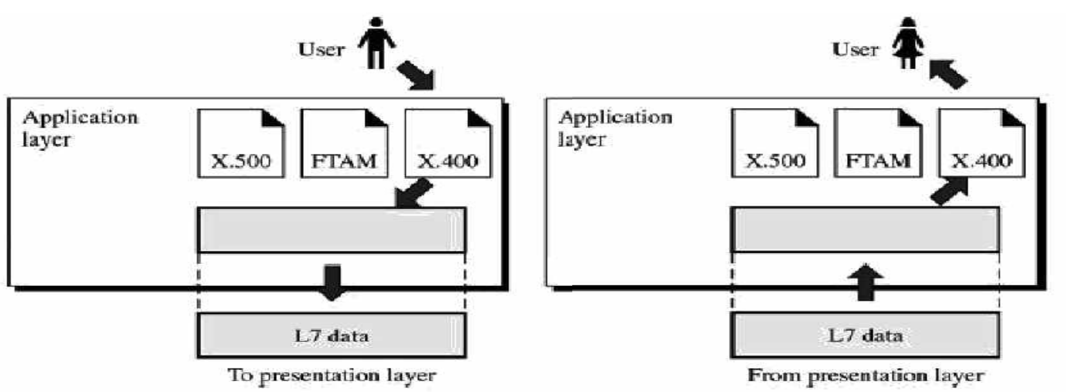
The presentation layer provides independence to the application processes by addressing any code or syntax conversion necessary to present the data to the network in a common communications format. It specifies how end-user applications should format the data.



The presentation layer translated between different data formats and protocols. Presentation functions include data file formatting, encoding, encryption and decryption of data messages, dialogue procedures, data compression algorithms, synchronization, interruption, and termination.

Application Layer {**responsible for providing services to the user**}

The application layer is the highest layer in the hierarchy and is analogous to the general manager of the network by providing access to the OSI environment. The applications layer provides distributed information services and controls the sequence of activities within and application and also the sequence of events between the computer application and the user of another application.



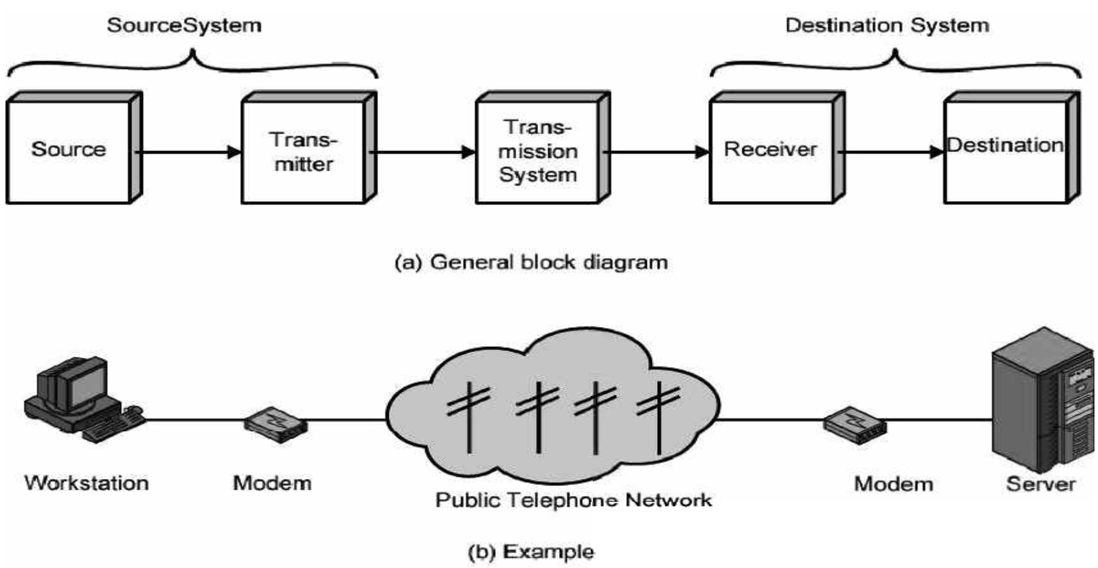
The application layer communicates directly with the user's application program. User application processes require application layer service elements to access the networking environment. The service elements are of two types: CASEs *(common application service elements)* satisfying particular needs of application processes like association control, concurrence and recovery. The second type is SASE (*specific application service elements)* which include TCP/IP stack, FTP, SNMP, Telnet and SMTP.

Data Communication Circuits

The underlying purpose of a digital communications circuit is to provide a transmission path between locations and to transfer digital information from one station (node, where computers or other digital equipment are located) to another using electronic circuits. Data communications circuits utilize electronic communications equipment and facilities to interconnect digital computer equipment. Communication facilities are physical means of interconnecting stations and are provided to data communications users through public telephone networks (PTN), public data networks (PDN), and a multitude of private data communications systems.

The following figure shows a simple two-station data communications circuit. The main components are:

Source: - This device generates the data to be transmitted; examples are mainframe computer, personal computer, workstation etc. The source equipment provides a means for humans to enter data into system.



Transmitter: - A transmitter transforms and encodes the information in such a way as to produce electromagnetic signals that can be transmitted across some sort of transmission system. For example, a modem takes a digital bit stream from an attached device such as a personal computer and transforms that bit stream into an analog signal that can be handled by the telephone network.

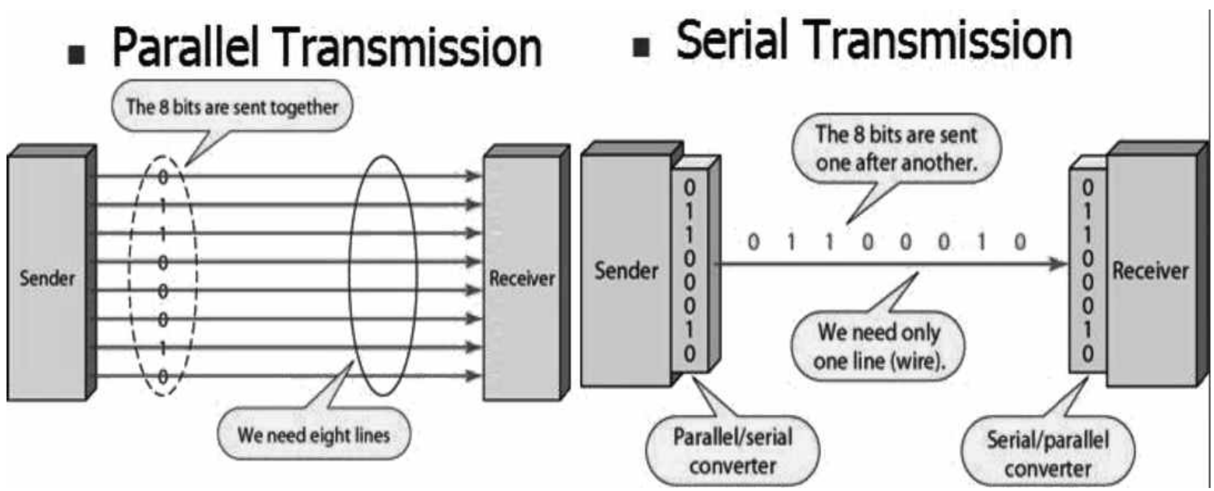
Transmission medium: - The transmission medium carries the encoded signals from the transmitter to the receiver. Different types of transmission media include free-space radio transmission (i.e. all forms of wireless transmission) and physical facilities such as metallic and optical fiber cables.

Receiver: - The receiver accepts the signal from the transmission medium and converts it into a form that can be handled by the destination device. For example, a modem will

accept an analog signal coming from a network or transmission line and convert it into a digital bit stream.
Destination: - Takes the incoming data from the receiver and can be any kind of digital equipment like the source.

Serial and Parallel Data Transmission

There are two methods of transmitting digital data namely ***parallel and serial*** transmissions. In parallel data transmission, all bits of the binary data are transmitted simultaneously. For example, to transmit an 8-bit binary number in parallel from one unit to another, eight transmission lines are required. Each bit requires its own separate data path. All bits of a word are transmitted at the same time. This method of transmission can move a significant amount of data in a given period of time. Its disadvantage is the large number of interconnecting cables between the two units. For large binary words, cabling becomes complex and expensive. This is particularly true if the distance between the two units is great. Long multiwire cables are not only expensive, but also require special interfacing to minimize noise and distortion problems. Serial data transmission is the process of transmitting binary words a bit at a time. Since the bits time-share the transmission medium, only one interconnecting lead is required.



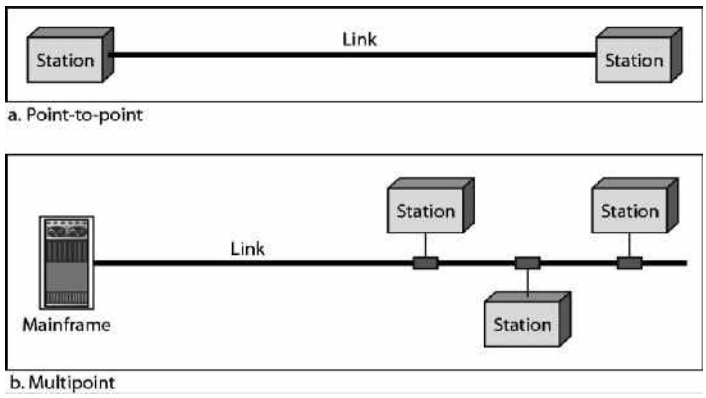
While serial data transmission is much simpler and less expensive because of the use of a single interconnecting line, it is a very slow method of data transmission. Serial data transmission is useful in systems where high speed is not a requirement. Parallel communication is used for short-distance data communications and within a computer, and serial transmission is used for long-distance data communications.

Data Communication Circuit Arrangements

A data communications circuit can be described in terms of circuit configuration and transmission mode.

Circuit Configurations

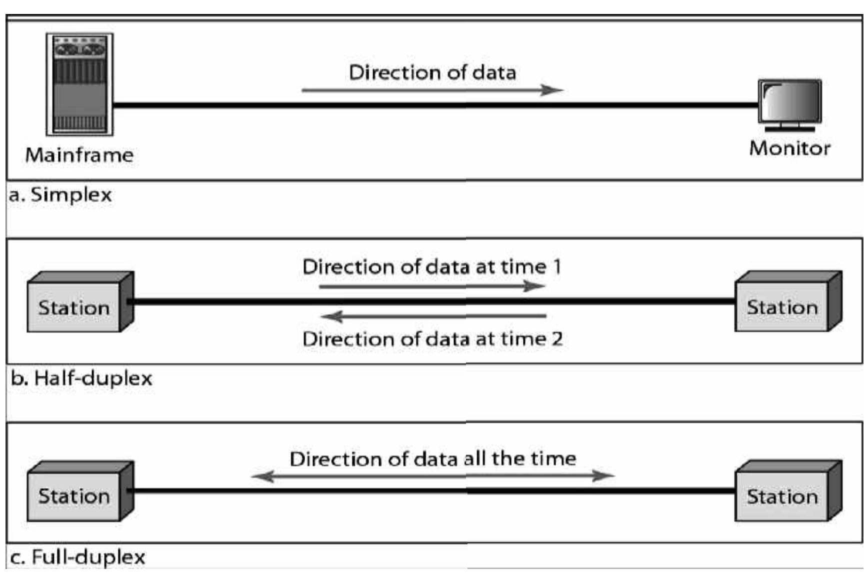
Data communications networks can be generally categorized as either two point or multipoint. A ***two-point*** configuration involves only two locations or stations, whereas a ***multipoint*** configuration involves three or more stations.



A two-point circuit involves the transfer of digital information between a mainframe computer and a personal computer, two mainframe computers or two data communications networks. A multi-point network is generally used to interconnect a single mainframe computer (host) to many personal computers or to interconnect many personal computers and capacity of the channel is either *Spatially shared*: Devices can use the link simultaneously or *Timeshare*: Users take turns

Transmission Modes

There are four modes of transmission for data communications circuits:



In ***simplex mode(SX)***, 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. Commercial radio broadcasting is an example. Simplex lines are also called receive-only, transmit-only or one-way-only lines.

In ***half-duplex(HDX)*** 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 used in cases where there is no need for communication in both directions at the same time; the entire capacity of the channel can be utilized for each direction. Citizens band (CB) radio is an example where push to talk (PTT) is to be pressed or depressed while sending and transmitting.

In ***full-duplex mode(FDX)*** (called duplex), both stations can transmit and receive simultaneously. One common example of full-duplex communication is the telephone network. The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel must be divided between the two directions.

In ***full/full duplex (F/FDX)*** mode, transmission is possible in both directions at the same time but not between the same two stations (i.e. station 1 transmitting to station 2, while receiving from station 3). F/FDX is possible only on multipoint circuits. Postal system can be given as a person can be sending a letter to one address and receive a letter from another address at the same time.

Data Communications Networks

Any group of computers connected together can be called a *data communications network,* and the process of sharing resources between computers over a data communications network is called *networking*. The most important considerations of a data communications network are *performance, transmission rate, reliability and security.*

Network Components, Functions, and Features

The major components of a network are end stations, applications and a network that will support traffic between the end stations. Computer networks all share common devices, functions, and features, including servers, clients, transmission media, shared data, shared printers and other peripherals, hardware and software resources, network interface card (NIC), local operating system (LOS) and the network operating system (NOS).

Servers: Servers are computers that hold shared files, programs and the network operating system. Servers provide access to network resources to all the users of the network and different kinds of servers are present. Examples include file servers, print servers, mail servers, communication servers etc.

Clients: Clients are computers that access and use the network and shared network resources. Client computers are basically the customers (users) of the network, as they request and receive service from the servers.

Shared Data: Shared data are data that file servers provide to clients, such as data files, printer access programs, and e-mail.

Shared Printers and other peripherals: these are hardware resources provided to the users of the network by servers. Resources provided include data files, printers, software, or any other items used by the clients on the network.

Network interface card: Every computer in the network has a special expansion card called network interface card (NIS), which prepares and sends data, receives data, and controls data flow between the computer and the network. While transmitting, NIC passes frames of data on to the physical layer and on the receiver side, the NIC processes bits received from the physical layer and processes the message based on its contents.

Local operating system: A local operating system allows personal computers to access files, print to a local printer, and have and use one or more disk and CD drives that are located on the computer. Examples are MS-DOS, PC-DOS, UNIX, Macintosh, OS/2, Windows 95, 98, XP and Linux.

Network operating system: the NOS is a program that runs on computers and servers that allows the computers to communicate over a network. The NOS provides services to clients such as log-in features, password authentication, printer access, network administration functions and data file sharing.

Network Models

Computer networks can be represented with two basic network models: peer-to-peer client/server and dedicated client/server. The client/server method specifies the way in which two computers can communicate with software over a network.

Peer-to-peer client/server network: Here, all the computers share their resources, such as hard drives, printers and so on with all the other computers on the network. Individual resources like disk drives, CD-ROM drives, and even printers are transformed into shared, collective resources that are accessible from every PC. Unlike client-server networks, where network information is stored on a centralized file server PC and made available to tens, hundreds, or thousands client PCs, the information stored across peer-to-peer networks is uniquely decentralized. Because peer-to-peer PCs have their own hard disk drives that are accessible by all computers, each PC acts as both a client (information requestor) and a server (information provider). The peer-to-peer network is an appropriate choice when there are fewer than 10 users on the network, security is not an issue and all the users are located in the same general area.

The advantages of peer-to-peer over client-server NOSs include:

- No need for a network administrator
- Network is fast/inexpensive to setup & maintain
- Each PC can make backup copies of its data to other PCs for security.
- Easiest type of network to build, peer-to-peer is perfect for both home and office use.

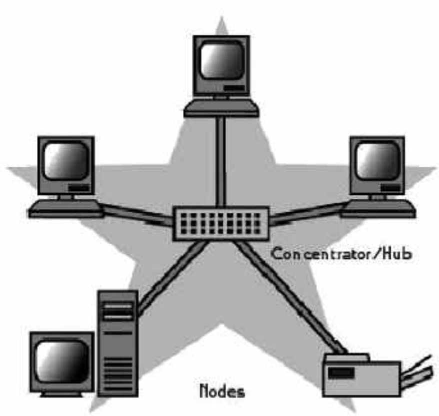
Dedicated client/server network: Here, one computer is designated as server and the rest of the computers are clients. Dedicated Server Architecture can improve the efficiency of client server systems by using one server for each application that exists within an organization. The designated servers store all the networks shared files and applications programs and function only as servers and are not used as a client or workstation. Client computers can access the servers and have shared files transferred to them over the transmission medium. In some client/server networks, client computers submit jobs to one of the servers and once they process the jobs, the results are sent back to the client computer.

In general, the dedicated client/server model is preferable to the peer-to-peer client/server model for general purpose data networks.

Network Topologies

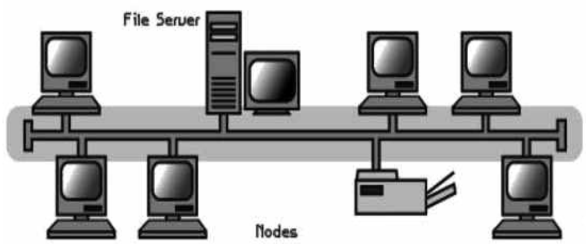
In computer networking, *topology* refers to the layout of connected devices, i.e. how the computers, cables, and other components within a data communications network are interconnected, both physically and logically. The physical topology describes how the network is actually laid out, and the logical topology describes how the data actually flow through the network. Two most basic topologies are point-to-point and multipoint. A point-to-point topology usually connects two mainframe computers for high-speed digital information. A multipoint topology connects three or more stations through a single transmission medium and some examples are *star, bus, ring, mesh and hybrid.*

Star topology: A star topology is designed with each node (file server, workstations, and peripherals) connected directly to a central network hub, switch, or concentrator. Data on a star network passes through the hub, switch, or concentrator before continuing to its destination. The hub, switch, or concentrator manages and controls all functions of the network. It also acts as a repeater for the data flow.



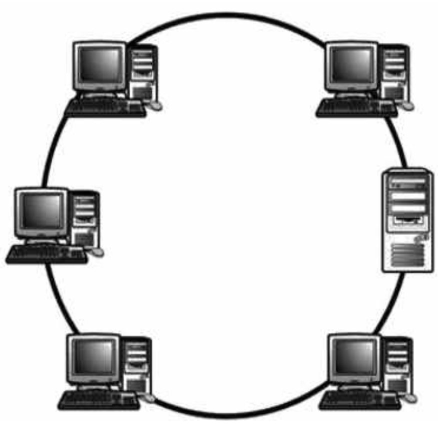
Advantages	Disadvantages
Easily expanded without disruption to the network	Requires more cable
Cable failure affects only a single user	A central connecting device allows for a single point of failure
Easy to troubleshoot and isolate problems	More difficult to implement

Bus topology: Bus networks use a common backbone to connect all devices. A single cable, (the backbone) functions as a shared communication medium that devices attach or tap into with an interface connector. A device wanting to communicate with another device on the network sends a broadcast message onto the wire that all other devices see, but only the intended recipient actually accepts and processes the message. The bus topology is the simplest and most common method of interconnecting computers. The two ends of the transmission line never touch to form a complete loop. A bus topology is also known as multidrop or linear bus or a horizontal bus.



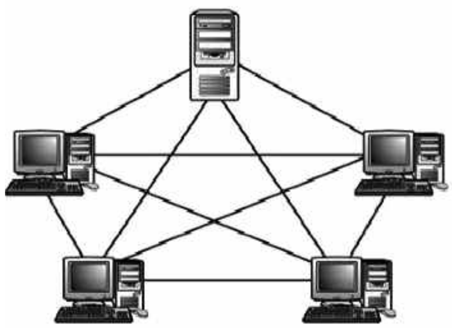
Advantages	Disadvantages
Cheap and easy to implement	Network disruption when computers are added or removed
Require less cable	A break in the cable will prevent all systems from accessing the network.
Does not use any specialized network equipment.	Difficult to troubleshoot.

Ring topology: In a ring network (sometimes called a loop), every device has exactly two neighbours for communication purposes. All messages travel through a ring in the same direction (either "clockwise" or "counter clockwise"). All the stations are interconnected in tandem (series) to form a closed loop or circle. Transmissions are unidirectional and must propagate through all the stations in the loop. Each computer acts like a repeater and the ring topology is similar to bus or star topologies.



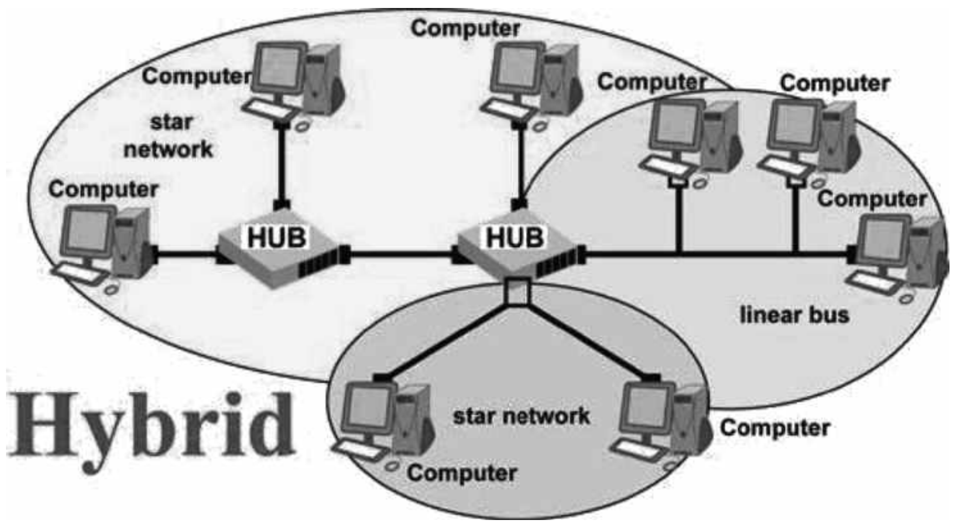
Advantages	Disadvantages
Cable faults are easily located, making troubleshooting easier	Expansion to the network can cause network disruption
Ring networks are moderately easy to install	A single break in the cable can disrupt the entire network.

Mesh topology: The *mesh* topology incorporates a unique network design in which each computer on the network connects to every other, creating a point-to-point connection between every device on the network. Unlike each of the previous topologies, messages sent on a mesh network can take any of several possible paths from source to destination. A mesh network in which every device connects to every other is called a full mesh. A disadvantage is that, a mesh network with *n* nodes must have *n(n-1)/2* links and each node must have *n-1* I/O ports (links).



Advantages	Disadvantages
Provides redundant paths between devices	Requires more cable than the other LAN topologies
The network can be expanded without disruption to current uses	Complicated implementation

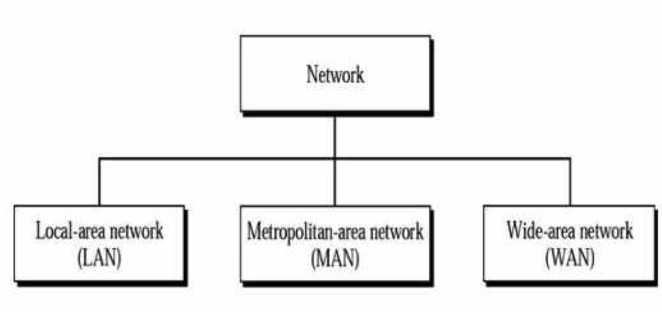
Hybrid topology: This topology (sometimes called mixed topology) is simply combining two or more of the traditional topologies to form a larger, more complex topology. Main aim is being able to share the advantages of different topologies.

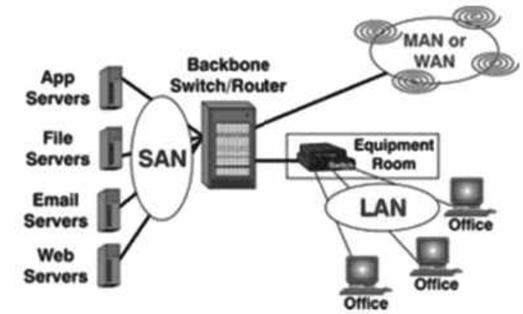


Network Classifications

One way to categorize the different types of computer network designs is by their scope or scale. Common examples of area network types are:

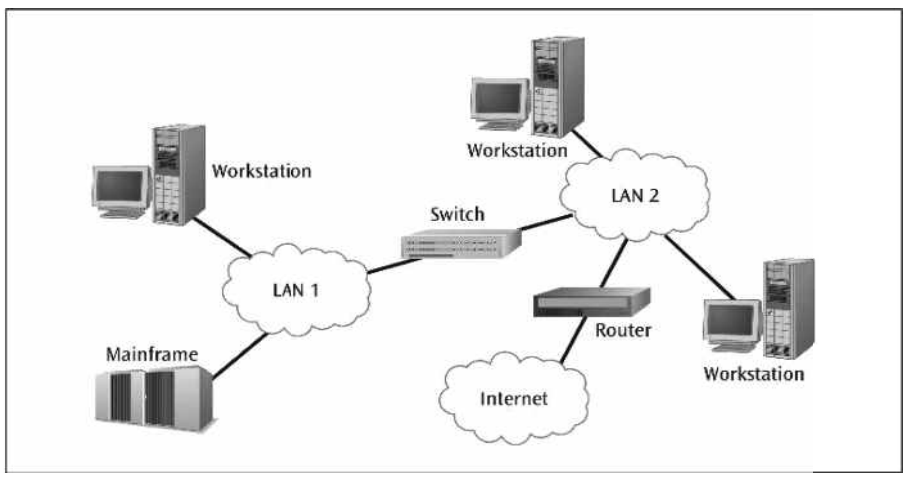
- LAN - Local Area Network
- WLAN - Wireless Local Area Network
- WAN - Wide Area Network
- MAN - Metropolitan Area Network
- SAN - Storage Area Network, System Area Network, Server Area Network, or sometimes Small Area Network
- CAN - Campus Area Network, Controller Area Network, or sometimes Cluster Area Network
- PAN - Personal Area Network
- DAN - Desk Area Network





<u>Local area network</u>: A local area network (LAN) is a network that connects computers and devices in a limited geographical area such as home, school, computer laboratory, office building, or closely positioned group of buildings. LANs use a network operating system to provide two-way communications at bit rates in the range of 10 Mbps to 100 Mbps. In addition to operating in a limited space, LANs are also typically owned, controlled, and managed by a single person or organization. They also tend to use certain connectivity technologies, primarily Ethernet and Token Ring.

Figure 7-1
A local area network interconnecting another local area network, the Internet, and a mainframe computer



<u>Advantages of LAN:</u>

- ✓ Share resources efficiently
- ✓ Individual workstation might survive network failure if it doesn't rely upon others
- ✓ Component evolution independent of system evolution
- ✓ Support heterogeneous hardware/software
- ✓ Access to other LANs and WANs
- ✓ High transfer rates with low error rates

~~Metropolitan area network~~: A MAN is optimized for a larger geographical area than a LAN, ranging from several blocks of buildings to entire cities. Its geographic scope falls between a WAN and LAN. A MAN might be a single network like the cable television network or it usually interconnects a number of local area networks (LANs) using a high-capacity backbone technology, such as fiber-optical links, and provides up-link services to wide area networks and the Internet. MANs typically operate at speeds of 1.5 Mbps to 10 Mbps and range from five miles to a few hundred miles in length. Examples of MANs are FDDI (fiber distributed data interface) and ATM (asynchronous transfer mode).

~~**Wide area network**~~: Wide area networks are the oldest type of data communications network that provide relatively slow-speed, long-distance transmission of data, voice and video information over relatively large and widely dispersed geographical areas, such as country or entire continent. WANs interconnect routers in different locations. A WAN differs from a LAN in several important ways. Most WANs (like the Internet) are not owned by any one organization but rather exist under collective or distributed ownership and management. WANs tend to use technology like ATM, Frame Relay and X.25 for connectivity over the longer distances.

~~**Global area network**~~: A GAN provides connections between countries around the entire globe. Internet is a good example and is essentially a network comprised of other networks that interconnect virtually every country in the world. GANs operate from 1.5 Mbps to 100 Gbps and cover thousands of miles.

Campus Area Network: - a network spanning multiple LANs but smaller than a MAN, such as on a university or local business campus.

Storage Area Network: - connects servers to data storage devices through a technology like Fibre Channel.

System Area Network: - Links high-performance computers with high-speed connections in a cluster configuration. Also known as Cluster Area Network.

Building backbone: - It is a network connection that normally carries traffic between departmental LANs within a single company. It consists of a switch or router to provide connectivity to other networks such as campus backbones, enterprise backbones, MANs, WANs etc

Camus backbone: - **It** is a network connection used to carry traffic to and from LANs located in various buildings on campus. It normally uses optical fiber cables for the transmission media between buildings and operates at relatively high transmission rates.

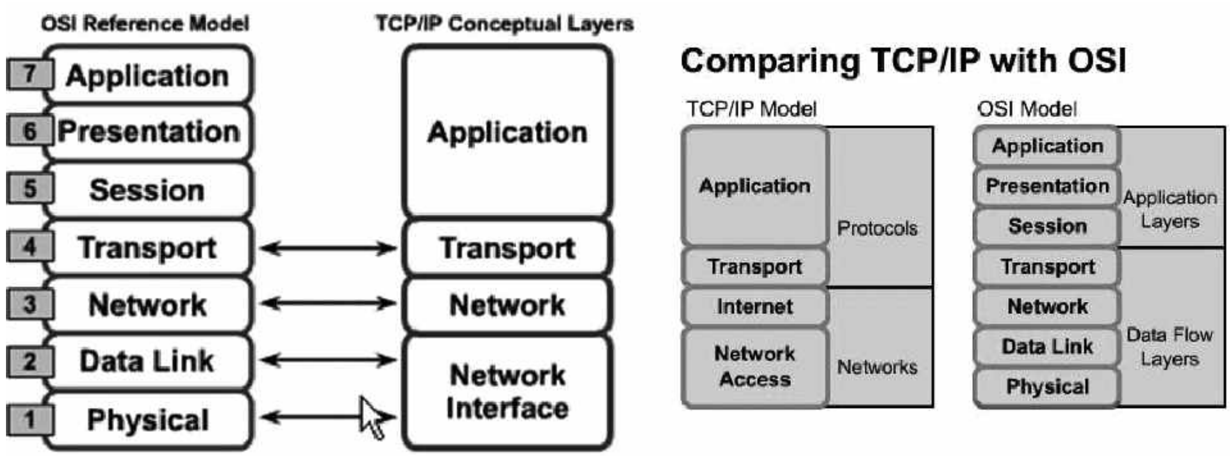
Enterprise networks: - It includes some or all of the above networks and components connected in a cohesive and manageable fashion.

Alternate Protocol Suites

The protocols other than OSI that are in wide spread used are TCP?IP and the Cisco three-layer hierarchical model.

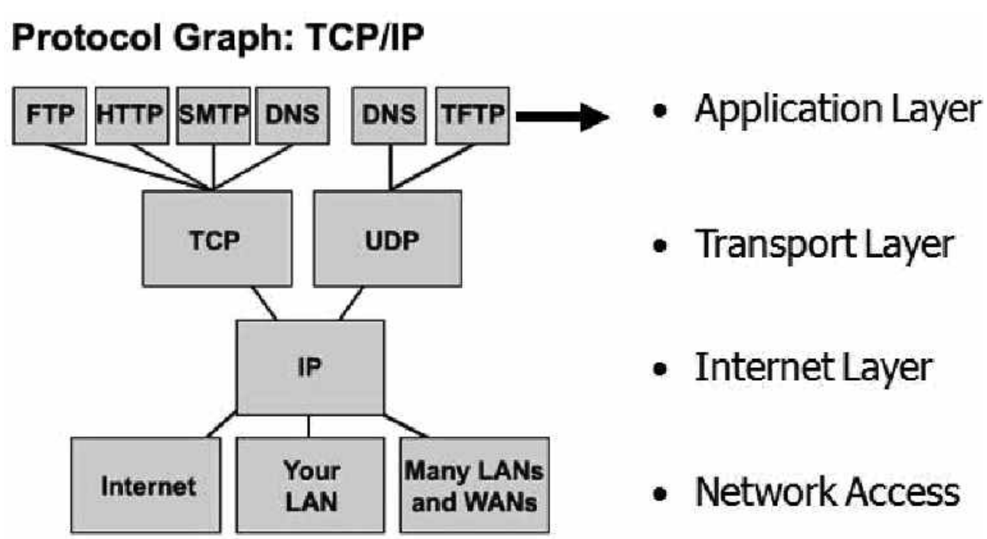
TCP/IP Protocol Suite

The U.S. Department of Defense (*DoD*) created the TCP/IP reference model because it wanted a network that could survive any conditions, even a nuclear war. Transmission Control Protocol/Internet Protocol (TCP/IP) {commonly known as internet suite} model is a set of communication protocols that allow communication across multiple diverse networks. TCP/IP is a hierarchical protocol comprised of either three or four layers. The three-layer version of TCP/IP contains the network, transport and application layers. Four layer version specifies the host to network layer.



The designers of TCP/IP felt that the higher level protocols should include the *session* and *presentation* layer details. They simply created an **application** layer that handles high-level protocols, issues of representation, encoding, and dialog control. The TCP/IP combines all application-related issues into one layer, and assures this data is properly packaged for the next layer.

The TCP/IP **transport layer** deals with the quality-of-service issues of reliability, flow control, and error correction. One of its protocols, the transmission control protocol (TCP), provides excellent and flexible ways to create reliable, well-flowing, low-error network communications. TCP is a *connection-oriented protocol*. The other protocol is User Datagram Protocol (UDP) which is a connection less protocol.



Common TCP/IP Protocols

The purpose of the ***Internet layer*** is to send source packets from any network on the internetwork and have them arrive at the destination independent of the path and networks they took to get there. The specific protocol that governs this layer is called the **Internet protocol** (IP). *Best path determination* and *packet switching* occur at this layer.

The **network access layer** also called the *host-to-network* layer is concerned with all of the issues of physically delivering data packets using frames or cells.

Differences between OSI and TCP/IP

- TCP/IP combines the presentation and session layer issues into its application layer
- TCP/IP combines the OSI data link and physical layers into one layer
- TCP/IP appears simpler because it has fewer layers
- TCP/IP protocols are the standards around which the Internet developed, so the TCP/IP model gains credibility just because of its protocols. In contrast, typically networks aren't built on the OSI protocol, even though the OSI model is used as a guide.

Cisco Three Layer Model

Cisco has defined a hierarchical model known as the hierarchical internetworking model. This model simplifies the task of building a reliable, scalable, and less expensive hierarchical internetwork because rather than focusing on packet construction; it focuses on the three functional areas, or layers, of your network.

Core layer: This layer is considered the backbone of the network and includes the high-end switches and high-speed cables such as fiber cables. This layer of the network does not route traffic at the LAN. In addition, no packet manipulation is done by devices in this layer. Rather, this layer is concerned with speed and ensures reliable delivery of packets.

Distribution layer: This layer includes LAN-based routers and layer 3 switches. This layer ensures that packets are properly routed between subnets and VLANs in your enterprise. This layer is also called the Workgroup layer. It also provides policy-based network connectivity, including:

- Packet filtering (firewalling): Processes packets and regulates the transmission of packets based on its source and destination information to create network borders.
- QoS: The router or layer 3 switches can read packets and prioritize delivery, based on policies set.
- Access Layer Aggregation Point: The layer serves the aggregation point for the desktop layer switches.
- Control Broadcast and Multicast: The layer serves as the boundary for broadcast and multicast domains.
- Application Gateways: The layer allows you to create protocol gateways to and from different network architectures.
- The distribution layer also performs queuing and provides packet manipulation of the network traffic.

Access layer: This layer includes hubs and switches. This layer is also called the desktop layer because it focuses on connecting client nodes, such as workstations to the network. This layer ensures that packets are delivered to end user computers. At the access layer, you can:

- Enable MAC address filtering: It is possible to program a switch to allow only certain systems to access the connected LANs.
- Create separate collision domains: A switch can create separate collision domains for each connected node to improve performance.
- Share bandwidth: You can allow the same network connection to handle all data.
- Handle switch bandwidth: You can move data from one network to another to perform load balancing.

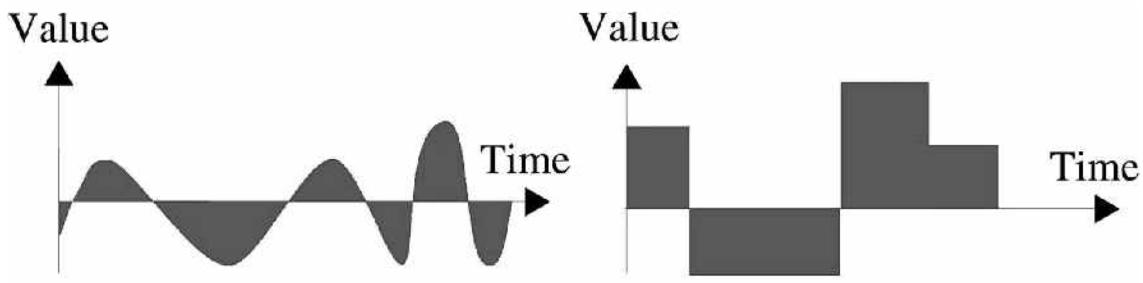
The benefits of the Cisco hierarchical model includes:

- High Performance: You can design high performance networks, where only certain layers are susceptible to congestion.
- Efficient management & troubleshooting: Allows you to efficiently organize network management and isolate causes of network trouble.
- Policy creation: You can easily create policies and specify filters and rules.
- Scalability: You can grow the network easily by dividing your network into functional areas.
- Behavior prediction: When planning or managing a network, the model allows you determine what will happen to the network when new stresses are placed on it.

UNIT-II

Signals, Noise, Modulation and Demodulation

Computers transmit data using digital signals, sequences of specified voltage levels. Computers sometimes communicate over telephone lines using analog signals, which are formed by continuously varying voltage levels. Electrical signals can be in analog or digital form. With analog signals, the amplitude changes continuously with respect to time with no breaks or discontinuities. A sine wave is the most basic analog signal.



a. Analog signal b. Digital signal

Digital signals are described as discrete; their amplitude maintains a constant level for a prescribed period of time and then it changes to another level. If only two levels are possible, it is called a binary signal. All binary signals are digital, but all digital signals are not necessarily binary. Converting information signals to a different form is called ***modulation*** and the reverse process is called ***demodulation***. The modulating signal is the information and the signal being modulated is the ***carrier***.

Two basic types of electronic communications systems are analog and digital. An analog digital communications system is a communications system in which energy is transmitted and received in analog form and are also propagated through the system in analog form. Digital communications covers a broad range of communications techniques including digital transmission and digital modulation.

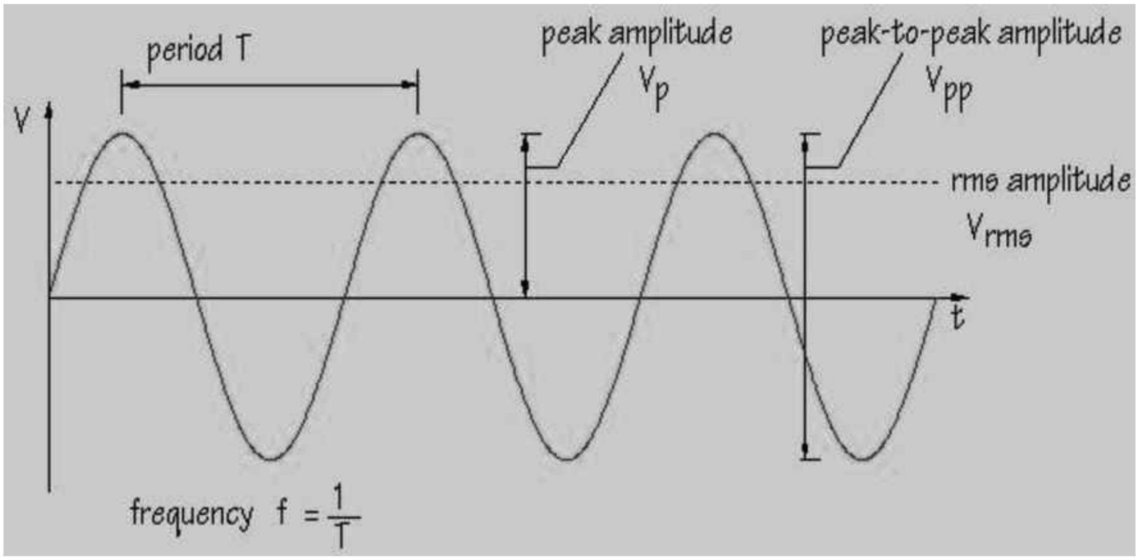
Signal Analysis

Mathematical signal analysis is used to analyze and predict the performance of the circuit on the basis of the voltage distribution and frequency composition of the information signal.

Amplitude, Frequency and Phase

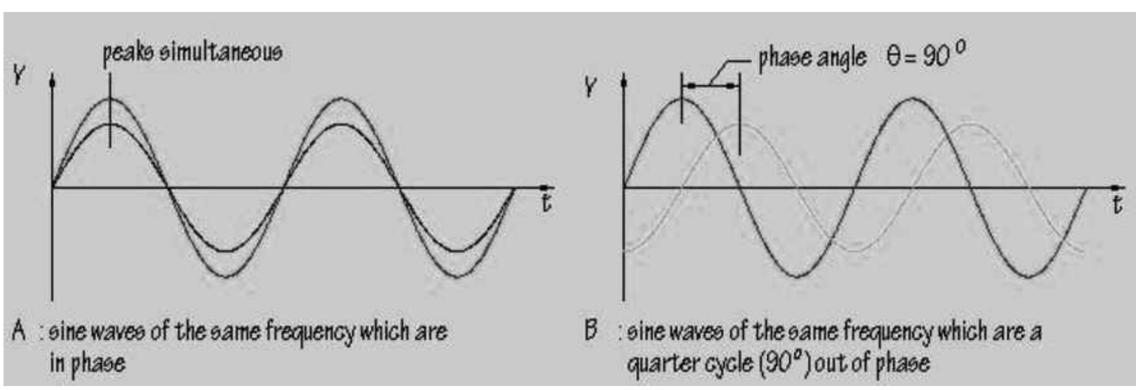
A ***cycle*** is one complete variation in the signal, and the ***period*** is the time the waveform takes to complete on cycle. One cycle constitutes 360 degrees (or 2π radians). Sine waves can be described in terms of three parameters: ***amplitude, frequency and phase.***

Amplitude (A): It is analogous to magnitude or displacement. The amplitude of a signal is the magnitude of the signal at any point on the waveform. The amplitude of electrical signal is generally measured in voltage. The maximum voltage of a signal in respect to its average value is called its peak amplitude or peak voltage.

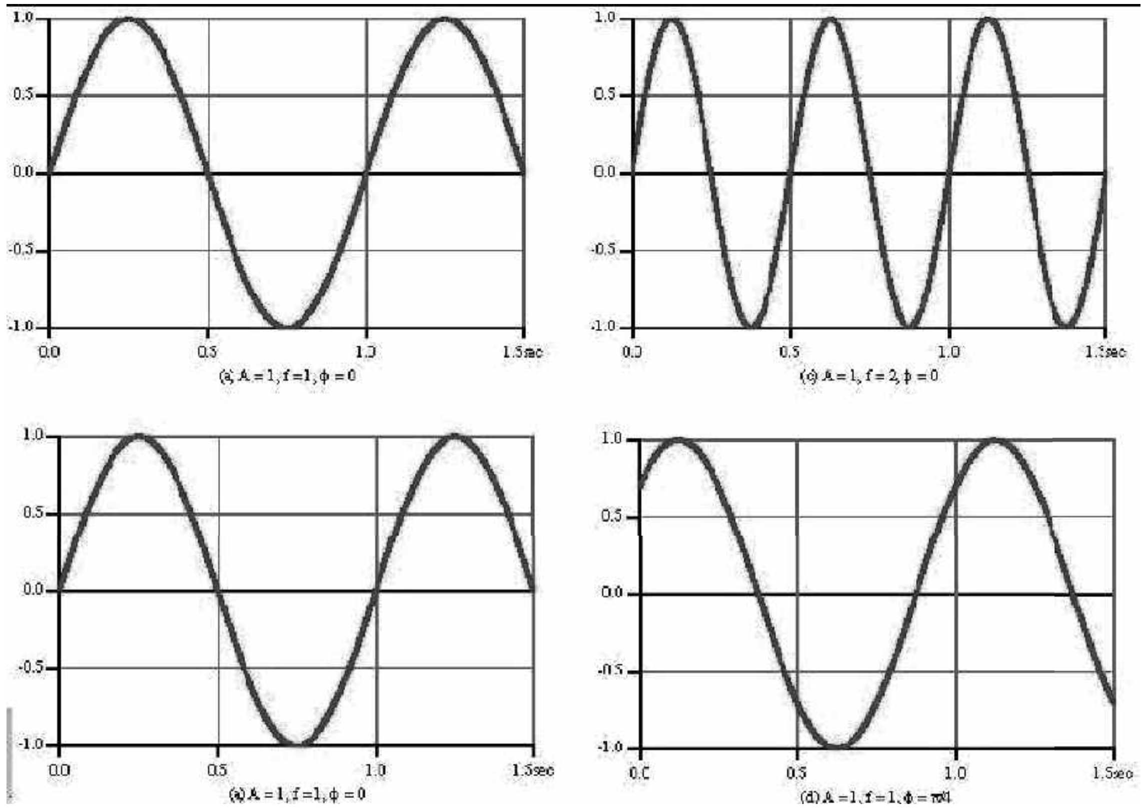


Frequency (f): The time of one cycle of a waveform is its period, which is measured in seconds. Frequency is the number of cycles completed per second. The measurement unit for frequency is the **hertz**, **Hz**. 1 Hz = 1 cycle per second. The frequency of the signal can be calculated from**T=1/*f***

Phase (Ø): The phase of the signal is measured in degrees or radians with respect to a reference point. A phase shift of 180 degrees corresponds to a shift of half a cycle.



A phase shift of 360 degrees corresponds to a shift of one complete cycle. If two sine waves have the same frequency and occur at the same time, they are said to be **in phase**, or they are said to out of phase. The difference in phase can be measured in degrees, and is called the **phase angle**, θ



Varying Sine wave with respect to frequency and phase

Periodic Signals

A signal is periodic if it completes a pattern within a measurable time and is characterized by amplitude, frequency and phase. Mathematically, a single frequency voltage waveform is

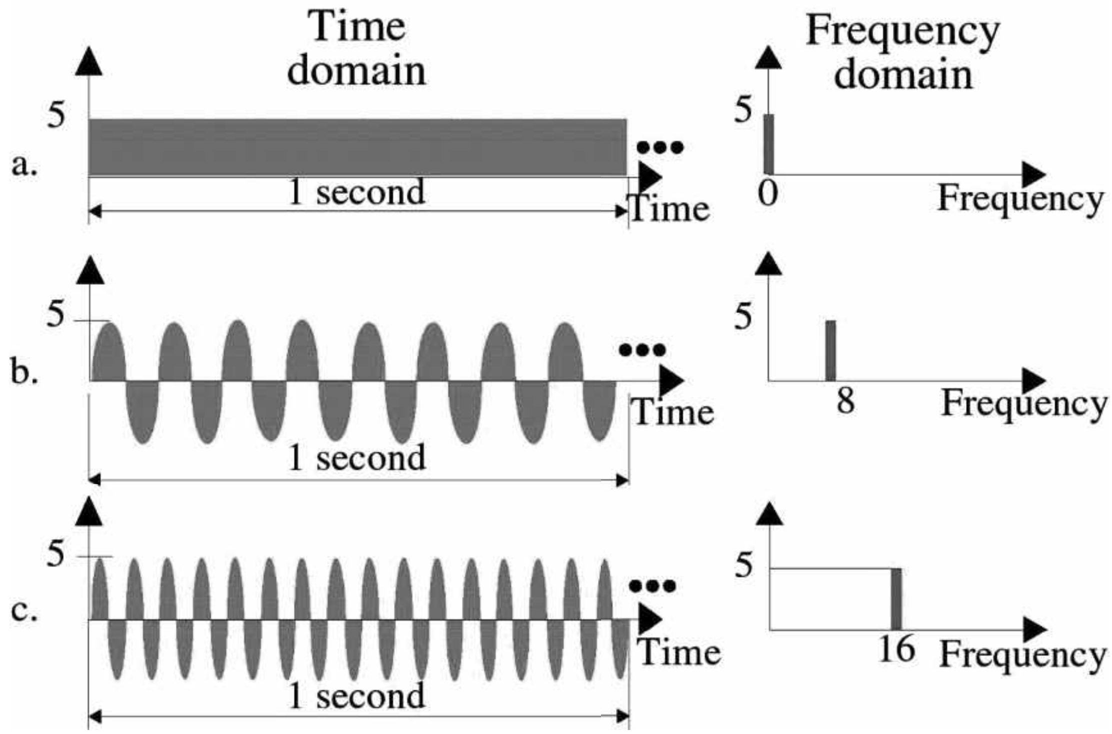
$$v(t) = V \sin(2\pi f t + \theta)$$, where

- $v(t)$ is time-varying voltage sine wave
- V is peak amplitude in volts
- f is frequency in hertz
- t is time in seconds
- θ is phase in degrees or radians

It is called a periodic wave because, it repeats at a uniform rate. A series of sine, cosine or square waves constitute an example of periodic waves, which can be analyzed in either the time domain or the frequency domain.

Time domain: Time domain is a term used to describe the analysis of mathematical functions, or physical signals, with respect to time. In the time domain, the signal or function's value is known for all real numbers, for the case of continuous time, or at various separate instants in the case of discrete time. An oscilloscope is a time-domain tool commonly used to visualize real-world signals in the time domain. A time domain graph shows how a signal changes over time.

Frequency Domain: frequency domain is a term used to describe the analysis of mathematical functions or signals with respect to frequency, rather than time. A spectrum analyser is a frequency-domain instrument which displays amplitude-versus frequency plot (called a frequency spectrum). The horizontal axis represents frequency and the vertical axis amplitude showing a vertical deflection for each frequency present in the waveform, which is proportional to the amplitude of the frequency it represents.



Examples for time domain and frequency domain

Complex Signals

Any repetitive waveform that is comprised of more than one harmonically related sine or cosine wave is called a nonsinusoidal, complex wave. Fourier series is used to analyze the complex periodic waves.

Fourier series: The Fourier series is used in signal analysis to represent the sinusoidal components of nonsinusoidal periodic waveforms. A **Fourier series** decomposes a periodic function or periodic signal into a sum of simple oscillating functions, namely sines and cosines. It can be expressed as:

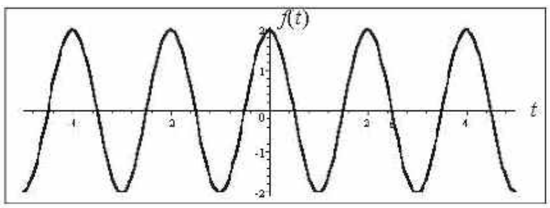
$$\mathbf{f(t) = A_0 + A_1 \cos\alpha + A_2 \cos 2\alpha + A_3 \cos 3\alpha + \ldots\ldots A_n \cos n\alpha}$$
$$\mathbf{+ A_0 + B_1 \sin\beta + B_2 \sin 2\beta + B_3 \sin 3\beta + \ldots\ldots\ldots B_n \sin n\beta}$$

Where $\alpha=\beta$

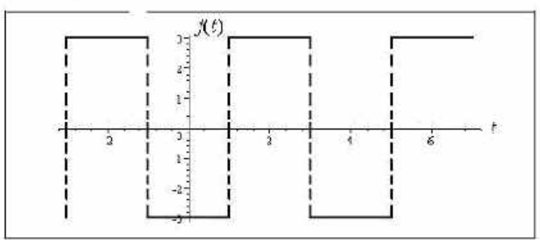
Any periodic waveform is comprised of an average dc component and a series of harmonically related sines or cosine waves. A harmonic is an integral multiple of the fundamental frequency. Fundamental frequency is the first harmonic and equal to the frequency (repetition rate) of the waveform. Second multiple is called second harmonic, third multiple is called third harmonic and so forth.

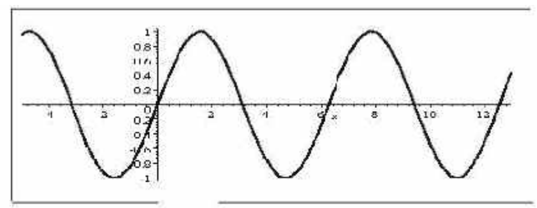
Wave symmetry: It describes the symmetry of a waveform in the time domain, i.e., its relative position with respect to the horizontal (time) and vertical (amplitude) axes.

Even symmetry: If a periodic voltage waveform is symmetric about the vertical axis, it is said to have axes, or mirror, symmetry and is called an even function. For all even functions, the β coefficients are zero. Even function satisfy the condition *f(t) = f(-t)*

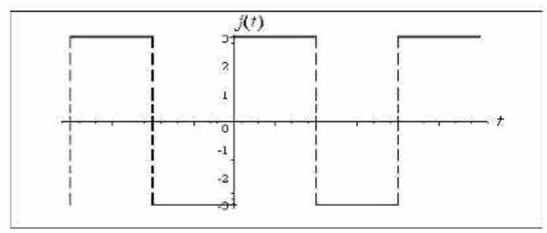


Examples of Even waves: sine wave and square wave

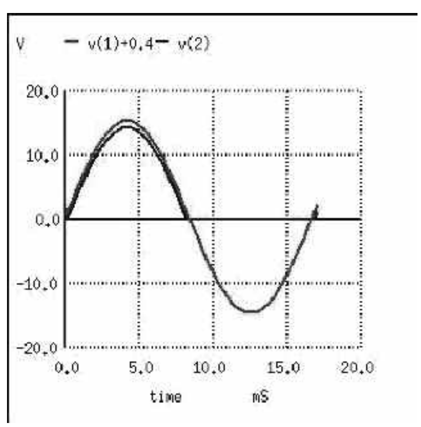




Examples of Odd waves: sine wave and square wave



Odd symmetry: If a periodic voltage waveform is symmetric about a line midway between the vertical axis and the negative horizontal axis and passing through the coordinate origin, it is said to have to point or skew, symmetry and is called an odd function. For all odd functions, the α coefficients are zero. Odd function satisfies *f(t) = -f(-t)*



Half-wave symmetry: If a periodic voltage waveform is such that the waveform for the first half cycle repeats itself except with the opposite sign for the second half cycle, it is called to have half-wave symmetry. Half-wave symmetry implies that the second half of the wave is exactly opposite to the first half. A function with half-wave symmetry does not have to be even or odd, as this property requires only that the shifted signal is opposite. Half-wave functions satisfy the condition $f(t) = -f(T+t)/2$

Frequency Spectrum and Bandwidth

The frequency spectrum of a waveform consists of all the frequencies contained in the waveform and their respective amplitudes plotted in the frequency domain.

Bandwidth of an information signal is simply the difference between the highest and lowest frequencies contained in the information and the bandwidth of a communication channel is the difference between the highest and lowest frequencies that the channel will allow to pass through it.

Electrical Noise and Signal-To-Noise Ratio

Noise is any disturbance or distortion that comes in the process of communication. Electrical noise is defined as any undesirable electrical energy that falls within the passband of the signal. A noise signal consists of a mixture of frequencies with random amplitudes. Noise can originate in various ways. The most prevalent and most interfering to data communication signals are *man-made noise, thermal noise, correlated noise, and impulse noise.*

Man-made noise: It is the kind of noise produced by mankind. The main sources are spark-producing mechanisms like commutators in electric motors, automobile ignition systems, ac power-generating and switching equipment, and fluorescent lights. It is impulsive in nature and contains a wide range of frequencies propagated in the free space like the radio waves. Man-made noise is most intense in more densely populated areas and sometimes is referred to as *industrial noise.*

Thermal noise: This is the noise generated by thermal agitation of electrons in a conductor. It is also referred to as white noise because of its uniform distribution across the entire electromagnetic frequency spectrum. Noise power density is the thermal noise power present in a 1-Hz bandwidth and is given by $N_o = KT$.

Thermal noise is independent of frequency and thus thermal noise present in any bandwidth is

$$N = KTB$$

where **N** is thermal noise power in watts, **K** is Boltzmann's constant in joules per Kelvin, ***T*** is the conductor temperature in kelvin ($0K = -273^{o}C$), and **B** is the bandwidth in hertz. Noise power is often measured in dBm. From the equation above, noise power in a resistor at room temperature, in dBm, is: $N_{dBm} = -174\ dBm + 10 \log B$

Correlated noise: this noise is correlated to the signal and cannot be present in a circuit unless there is a signal. Correlated noise is produced by nonlinear amplification and includes harmonic distortion and intermodulation distortion. Harmonic distortion occurs when unwanted harmonics of a signal are produced through nonlinear amplification and is also called amplitude distortion. Intermodulation distortion is the generation of unwanted sum and difference frequencies produced when two or more signals are amplified in a nonlinear device.

Impulse noise: This noise is characterized by high-amplitude peaks of short duration in the total noise spectrum. It consists of sudden bursts of irregularly shaped pulses that generally last between a few microseconds and several milliseconds, depending on their amplitude and origin. In case of voice communications, impulse noise is very annoying as it generates a sharp popping or crackling sound where as it is devastating in data circuits.

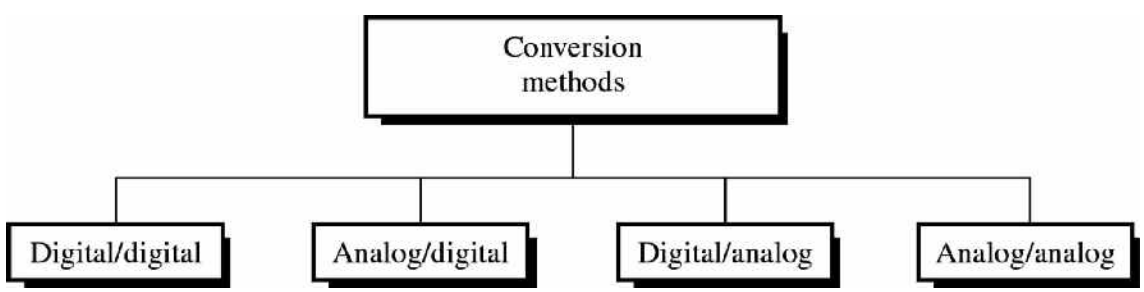
Signal-to-noise power ratio: **Signal-to-noise ratio** (often abbreviated **SNR** or **S/N**) is defined as the ratio of signal power to the noise power corrupting the signal. A ratio higher than 1:1 indicates more signal than noise. Signal-to-noise ratio is defined as the power ratio between signal (meaningful information) and the background noise (unwanted signal)

$$\mathrm{SNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}},$$

where *P* is average power in watts. The ratio often expressed in decibels as ***S/N** $_{(dBm)}$ **= 10 log(P_S/P_N)***

Analog Modulation Systems

A sine wave has three main components: amplitude, frequency and phase and can be expressed as $v(t) = V \sin(2\pi ft + \theta)$. If the information signal is analog and the amplitude (V) of the carrier is varied proportional to the informational signal, *amplitude modulation* (AM) is produced. If the frequency (f) is varied proportional to the information signal, *frequency modulation* (FM) is produced and if the phase (θ) is varied proportional to the information signal, *phase modulation* (PM) is produced. Frequency and phase modulation are similar and often combined and are simply called *angle modulation*.

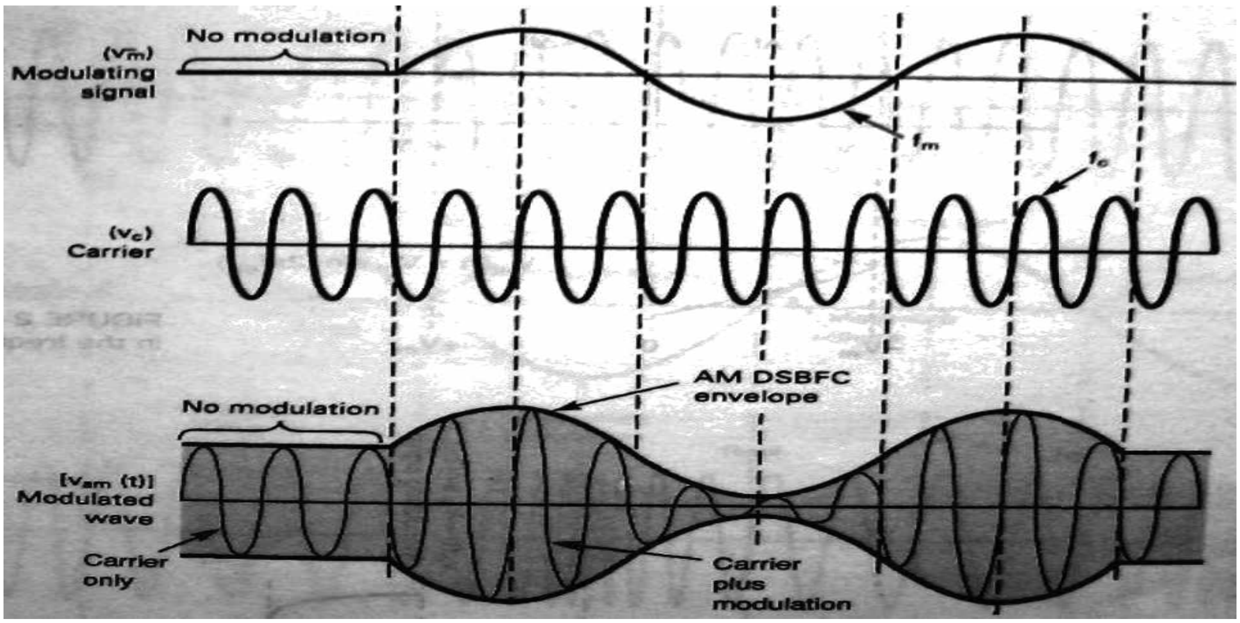


The process of impressing relatively low-frequency information signals onto a high-frequency carrier signal is called *modulation* and the reverse process is called *demodulation*.

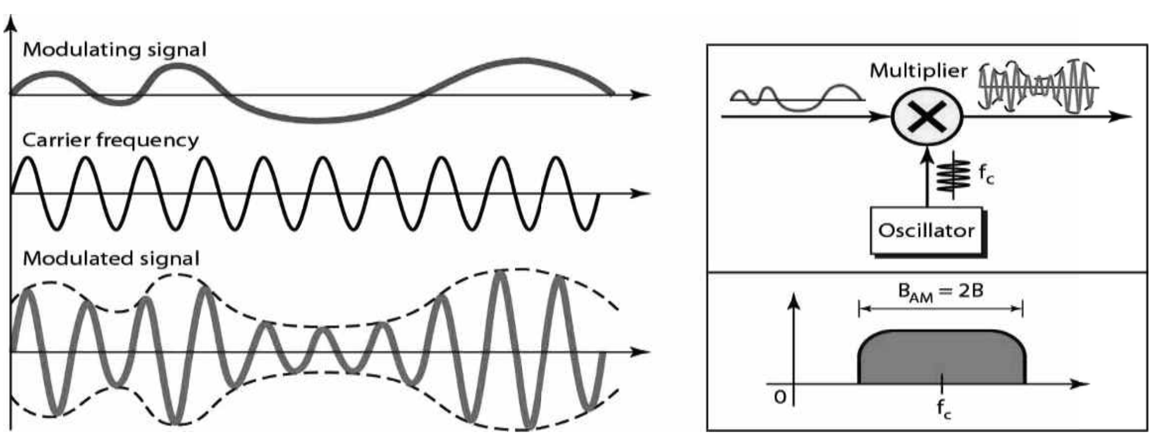
Analog modulation is used for the transmission of conventional analog signals, such as voice, music, and video and not particularly useful for data communication systems.

Amplitude Modulation

Amplitude modulation is the process of changing the amplitude of a relatively high frequency carrier signal in proportion to the instantaneous value of the modulating signal (information). AM modulators are two-input devices, one of them is a single, relatively high frequency carrier signal of constant amplitude and the second is the relatively low-frequency information signal. The following figure shows generation of AM waveform when a single-frequency modulating signal acts on a high frequency carrier signal.



AM generation



Advantages of AM are simple to implement, needs a circuit with very few components and inexpensive. The disadvantages include inefficient power usage and use of bandwidth and also prone to noise. The total bandwidth required for AM can be determined from the bandwidth of the audio signal: $B_{AM} = 2B$

Angle Modulation

Angle modulation results whenever the phase angle of a sinusoidal signal is varied with respect to time and includes both FM and PM. Whenever the frequency of a carrier signal is varied, the phase is also varied and vice versa. If the frequency of the carrier is varied directly in accordance with the information signal, FM results, whereas if the phase is varied directly, PM results.

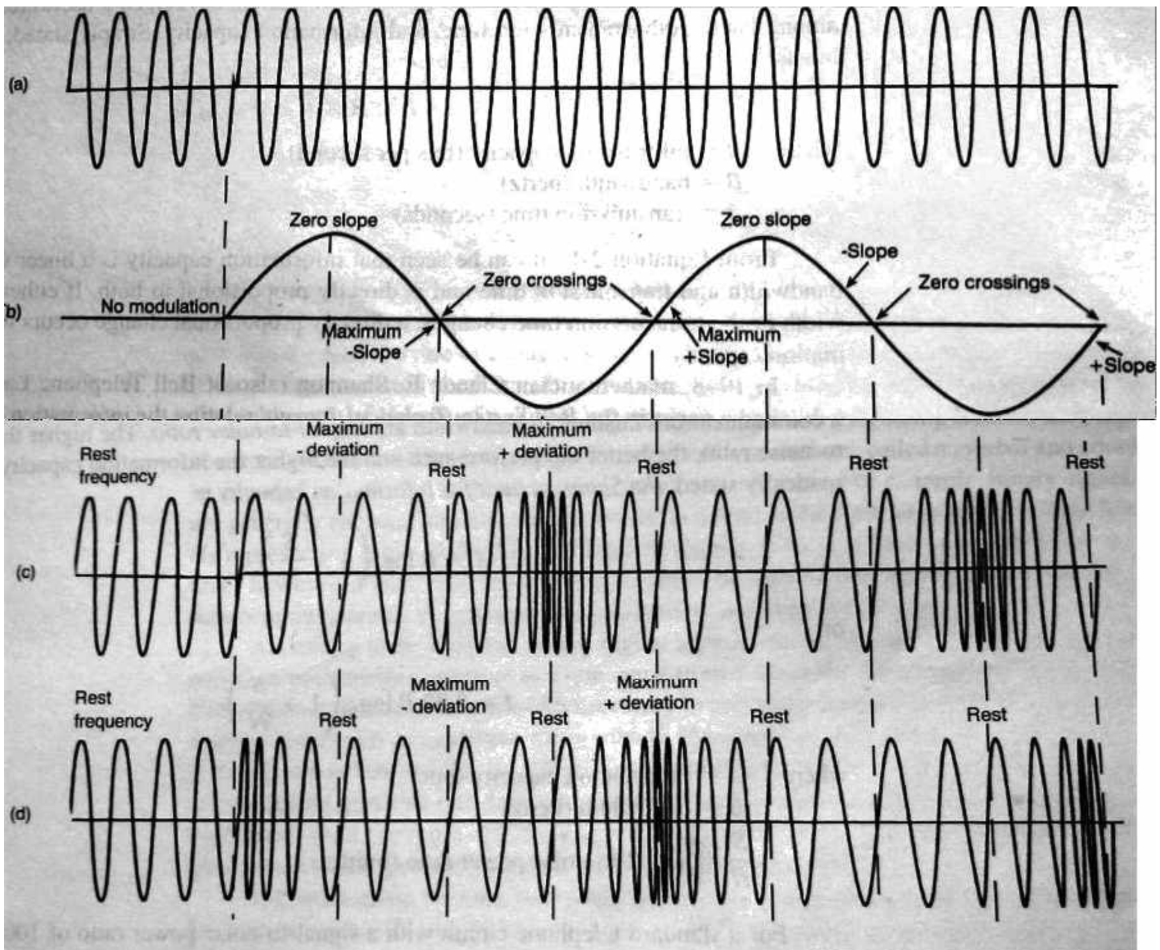
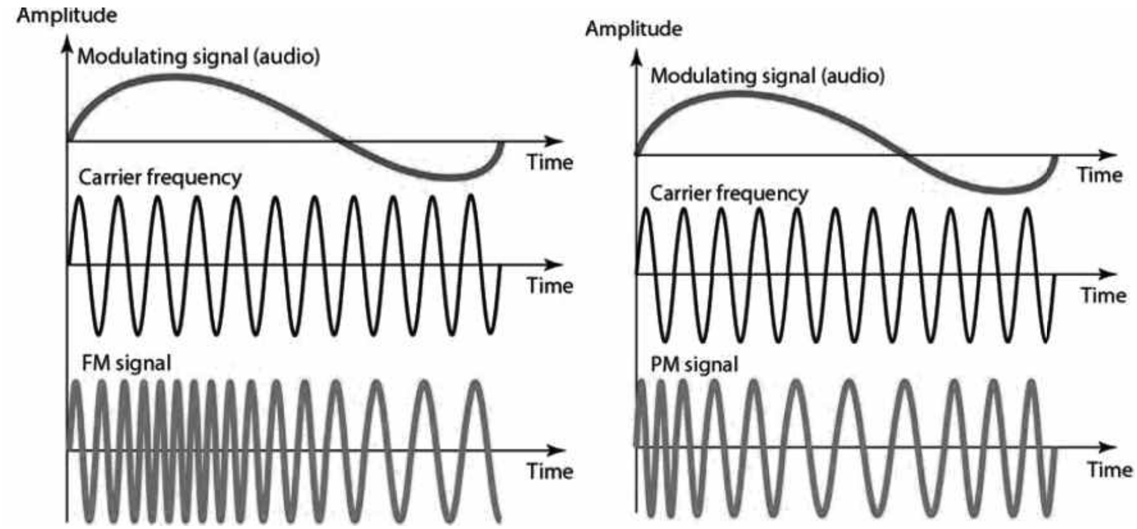


FIGURE 2-16 Phase and frequency modulation of a sine-wave carrier by a sine-wave signal: (a) unmodulated carrier; (b) modulating signal; (c) frequency-modulated wave; (d) phase-modulated wave

The above figure shows the FM and PM of a sinusoidal carrier by a single-frequency modulating signal. Both FM and PM waveforms are identical except for their time relationship (phase). With FM, the maximum frequency deviation occurs during the maximum positive and negative peaks of the modulating signal. With FM, the maximum frequency deviation occurs during the zero crossings in the modulating signal.



Comparison of FM and PM

An important feature of FM and PM is that they can provide much better protection to the message against channel noise when compared to AM. Also because of their constant amplitude nature, they can withstand nonlinear distortion and amplitude fading.

Information Capacity, Bits, Bit Rate, Baud, and M-ARY Encoding

Information capacity is a measure of how much information can be propagated through a communication system and a function of bandwidth and transmission time. It represents the number of independent symbols that can be carried through a system in a given unit of time. The most basic digital symbol used to represent information is the ***binary digit, or bit***. ***Bit rate*** is simply the number of bits transmitted during 1 second and is expressed as *bits per second (bps).*

R.Hartley developed a useful relationship among bandwidth, transmission time and information capacity called **Hartley's law** given by:

$$I \propto B \times t$$

Where*, I is the information capacity in bps, B is bandwidth in hertz and t is transmission time in sec's*

Relation between information capacity of a communication channel to a bandwidth and signal-to-noise ratio is given by Claude E. Shannon. The higher the signal-to-noise ratio, the better the performance and also information capacity is higher. The **Shannon limit of information capacity** is

$$I = B \log_2 (1 + S/N) \quad \text{or} \quad I = 3.32\, B \log_{10} (1 + S/N)$$

Where I is information capacity in bps, B is bandwidth in hertz and S/N is signal to noise ratio.

M-ary Encoding

M-ary is a term derived from the word binary. M simply represents a digit that corresponds to the number of conditions, levels, or combinations possible for a given number of binary variables. For example, a digital signal with four possible conditions is an M-ary system where M= 4 and if there are eight possible conditions, then M= 8. The number of bits necessary to produce a given number of conditions is expressed mathematically as:

$\mathbf{N = \log_2 M}$ or it can be written as $\mathbf{M = 2^N}$**,** where N is no of bits necessary and M is number of conditions, levels or combinations possible with N bits. From the equation, it can be said that if there is one bit, only 2^1 or two conditions are possible. For two bits 2^2 or four conditions are possible.

Baud and Minimum Bandwidth

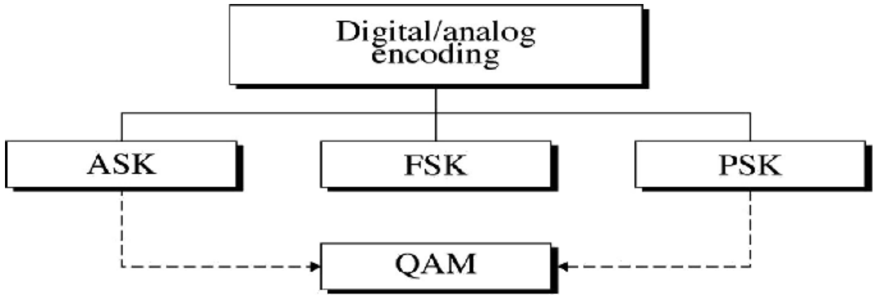
Baud, like bit rate is a rate of change. Baud refers to the rate of change of the signal on the transmission medium after encoding and modulation have occurred. Baud is the reciprocal of the time of one output signalling element, and a signalling element may represent several information bits. Baud is also transmitted one at a time and a baud may represent more than one information bit. So, the baud of the data communications system may be considerably less than the bit rate.

According to H.Nyquist, binary digital signals can be propagated through an ideal noiseless medium at a rate equal to twice the bandwidth of the medium. The minimum theoretical bandwidth necessary to propagate a signal is called the minimum Nyquist bandwidth or sometimes the Nyquist bandwidth. Using multilevel signalling, the Nyquist formulation for channel capacity is $f_b = B \log_2 M$ where, f_b is channel capacity in bps, B is minimum Nyquist bandwidth in hertz and M is no of discrete signal or voltage levels. If N is substituted, we get

$\mathbf{B = baud = f_b/N}$**,** where N is number of bits encoded into each signalling element.

Digital Modulation

Digital modulation is the transmission of digitally modulated analog signals between two or more points in a communications system. Analog and Digital modulation systems use analog carriers to transport information through the system, but digital modulation uses digital modulating (information) signal. Analog systems use analog signal only. In, $v(t) = V \sin(2\pi ft + \theta)$, if the information signal is digital and amplitude (V) of the carrier is varied proportional to the information signal, a digitally modulated signal called amplitude-shift keying (ASK) is produced. If the frequency (f) is varied proportional to the information signal, frequency-shift keying (FSK) is produced and if the phase is varied proportional to the information signal, phase-shift keying (PSK) is produced. If both amplitude and phase are varied proportional to the information signal, quadrature amplitude modulation (QAM) results.



Digital modulation is ideally suited to a multitude of communications applications including both cable and wireless systems. Applications include relatively low-speed voice-band data communications systems, high-speed data transmission systems, digital satellite communication systems and personal communication systems (PCS).

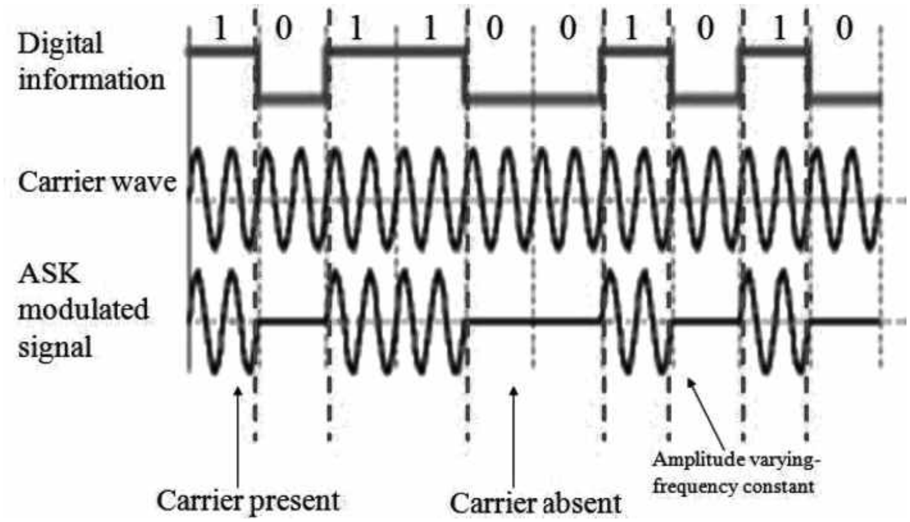
Modulation format	Application
MSK, GMSK	GSM, CDPD
BPSK	Deep space telemetry, cable modems
QPSK, $\pi/4$ DQPSK	Satellite, CDMA, NADC, TETRA, PHS, PDC, LMDS, DVB-S, cable (return path), cable modems, TFTS
OQPSK	CDMA, satellite
FSK, GFSK	DECT, paging, RAM mobile data, AMPS, CT2, ERMES, land mobile, public safety
8PSK	Satellite, aircraft, telemetry pilots for monitoring broadband video systems
16 QAM	Microwave digital radio, modems, DVB-C, DVB-T
32 QAM	Terrestrial microwave, DVB-T
64 QAM	DVB-C, modems, broadband set top boxes, MMDS
256 QAM	Modems, DVB-C (Europe), Digital Video (US)

Amplitude-Shift Keying

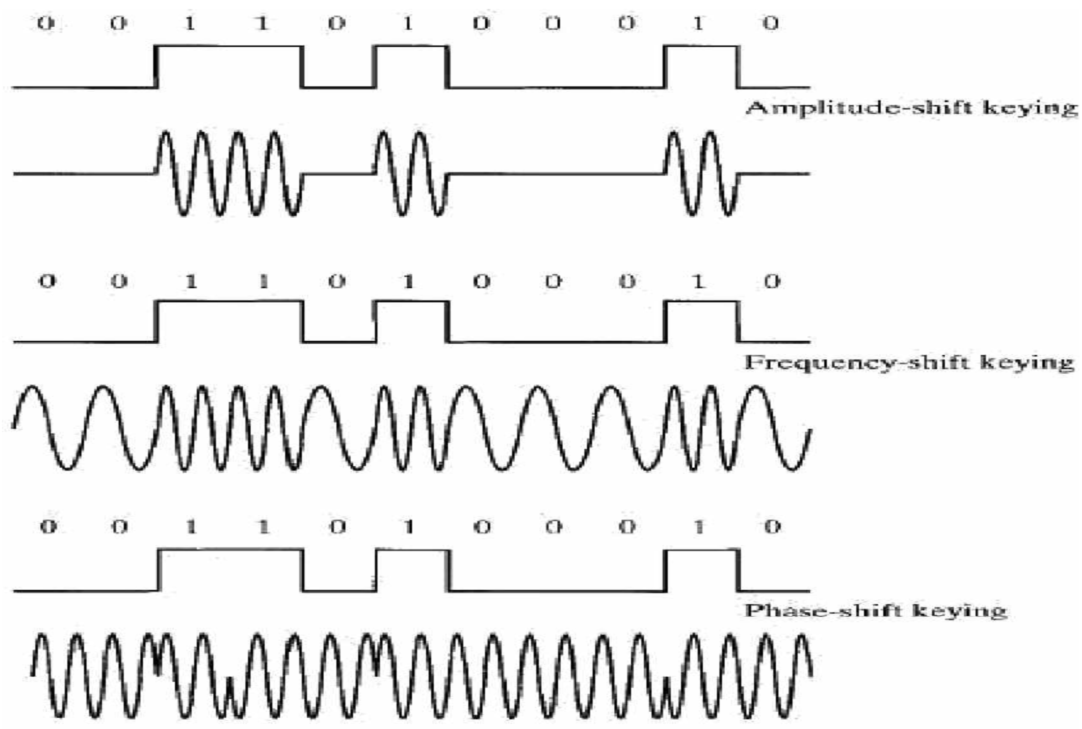
It is the simplest digital modulation technique where a binary information signal directly modulates the amplitude of an analog carrier. Only two output amplitudes are possible and ASK is sometimes called as digital amplitude modulation (DAM). Amplitude shift keying is given in mathematical terms as:

$$v_{ask}(t) = [\,1 + v_m(t)\,]\left[\,A/2 \cos(\omega_c t)\right]$$

Where $v_{ask}(t)$ is amplitude-shift keying wave, $v_m(t)$ is digital modulation (modulating) signal in volts, $A/2$ is unmodulated carrier amplitude in volts and ω_c is analog carrier radian frequency in radians per second.



In the above equation, for the modulating signal $v_m(t)$, logic 1 is represented by +1V and logic 0 is represented by -1V. So the modulated wave $v_{ask}(t)$ is either $A\cos(\omega_c t)$ or 0 i.e., the carrier is either on or off. ASK is sometimes referred as on-off keying (OOK). The rate of change of the ASK waveform (baud) is the same as the rate of change of the binary input making bit rate equal to baud. With ASK, the bit rate is also equal to the minimum Nyquist bandwidth.



Modulation of analog signals for digital data.

Frequency Shift Keying

FSK is another simple, low-performance type of digital modulation. It is similar to FM, except the modulating signal is a binary signal varying between two discrete voltage levels. FSK is sometimes called as *binary* FSK (BFSK). FSK is generally expressed as

$$v_{fsk}(t) = V_c \cos\{ 2\pi[f_c + v_m(t)\Delta f]t \}$$

Where $v_{fsk}(t)$ is binary FSK waveform, V_c is peak analog carrier amplitude in volts, f_c is analog carrier center frequency in hertz, f is peak change or shift in the analog carrier frequency and $v_m(t)$ is binary input(modulating) signal in volts. For logic 1, $v_m(t) = +1$ and for logic 0, $v_m(t) = -1$ reducing the equation to $v_{fsk}(t) = V_c \cos\{ 2\pi[f_c + f]t \}$ and $v_{fsk}(t) = V_c \cos\{ 2\pi[f_c - f]t \}$

As the binary signal changes from a logic 0 to a logic 1 and vice versa, the output frequency shifts between two frequencies: a mark, or logic 1 frequency (f_m) and a space or logic 0 frequency (f_s). The mark and space frequencies are separated from the carrier frequency by the peak frequency deviation (f) and from each other by 2 f.

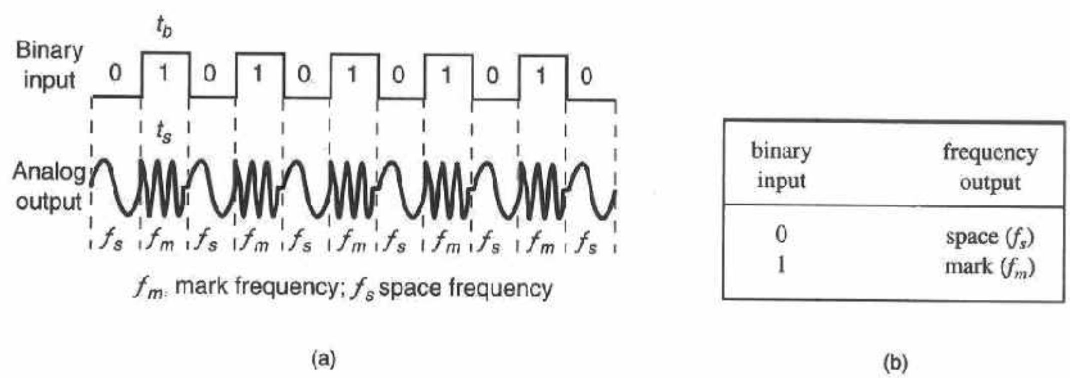


FIGURE 9-4 FSK in the time domain: (a) waveform; (b) truth table

With FSK, ***frequency deviation*** is defined as the difference between either the mark or space frequency and the center frequency or half the difference between the mark and space frequencies. Frequency deviation can be expressed as $f = |f_m - f_s| / 2$

The **baud** for BFSK is determined by placing N = 1, i.e., baud = $f_b/1 = f_b$

The **minimum bandwidth** for FSK is determined from;

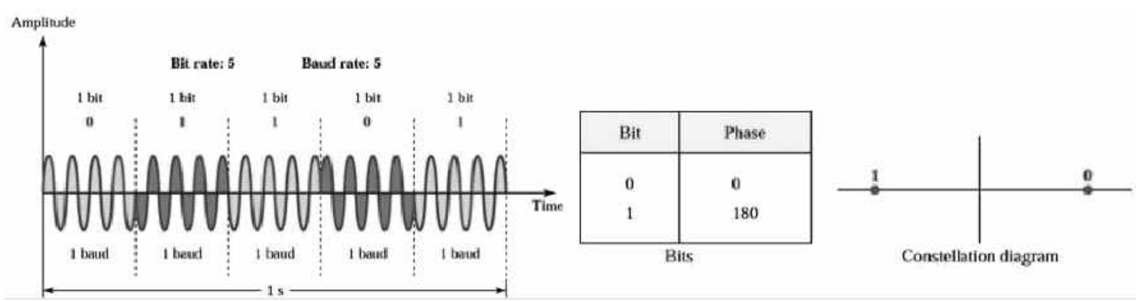
B = $|(f_s - f_b) - (f_m - f_b)| = |f_s - f_m| + 2f_b$. But $|f_s - f_m| = 2f$,

Therefore, **B = $2(f + f_b)$**, where B is minimum Nyquist bandwidth in hertz and f **is** frequency deviation and f_b is input bit rate.

Phase-Shift Keying

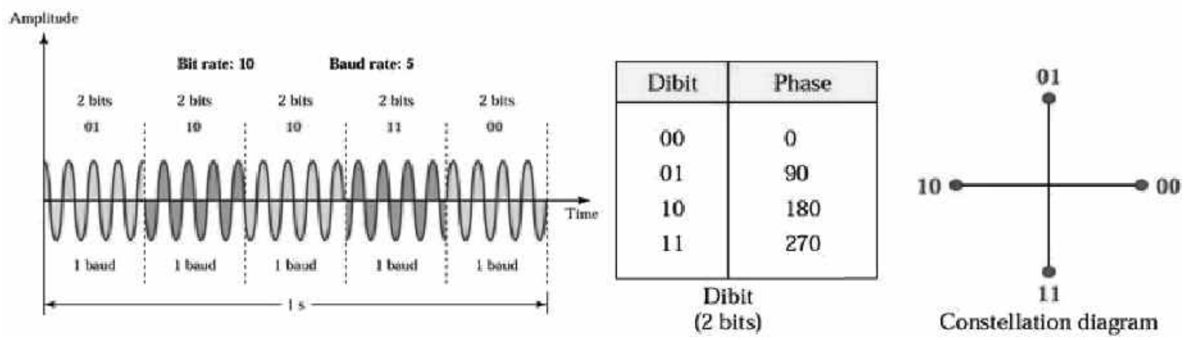
Phase-shift keying (PSK) is a digital modulation scheme that conveys data by changing, or modulating, the phase of a reference signal (the carrier wave). PSK uses a finite number of phases; each assigned a unique pattern of binary digits. Usually, each phase encodes an equal number of bits. PSK is not susceptible to the noise degradation that affects ASK or to the bandwidth limitations of FSK.

Binary phase-shift keying: The simplest PSK technique is called binary phase-shift keying (BPSK), where N = 1 and M = 2. Therefore, with BPSK two phases are possible for the carrier. It uses two opposite signal phases (0 and 180 degrees). The digital signal is broken up timewise into individual bits (binary digits). The state of each bit is determined according to the state of the preceding bit. If the phase of the wave does not change, then the signal state stays the same (0 or 1). If the phase of the wave changes by 180 degrees -- that is, if the phase reverses -- then the signal state changes (from 0 to 1 or from 1 to 0). Because there are two possible wave phases, BPSK is sometimes called ***biphase modulation or phase-reversal keying (PRK).***



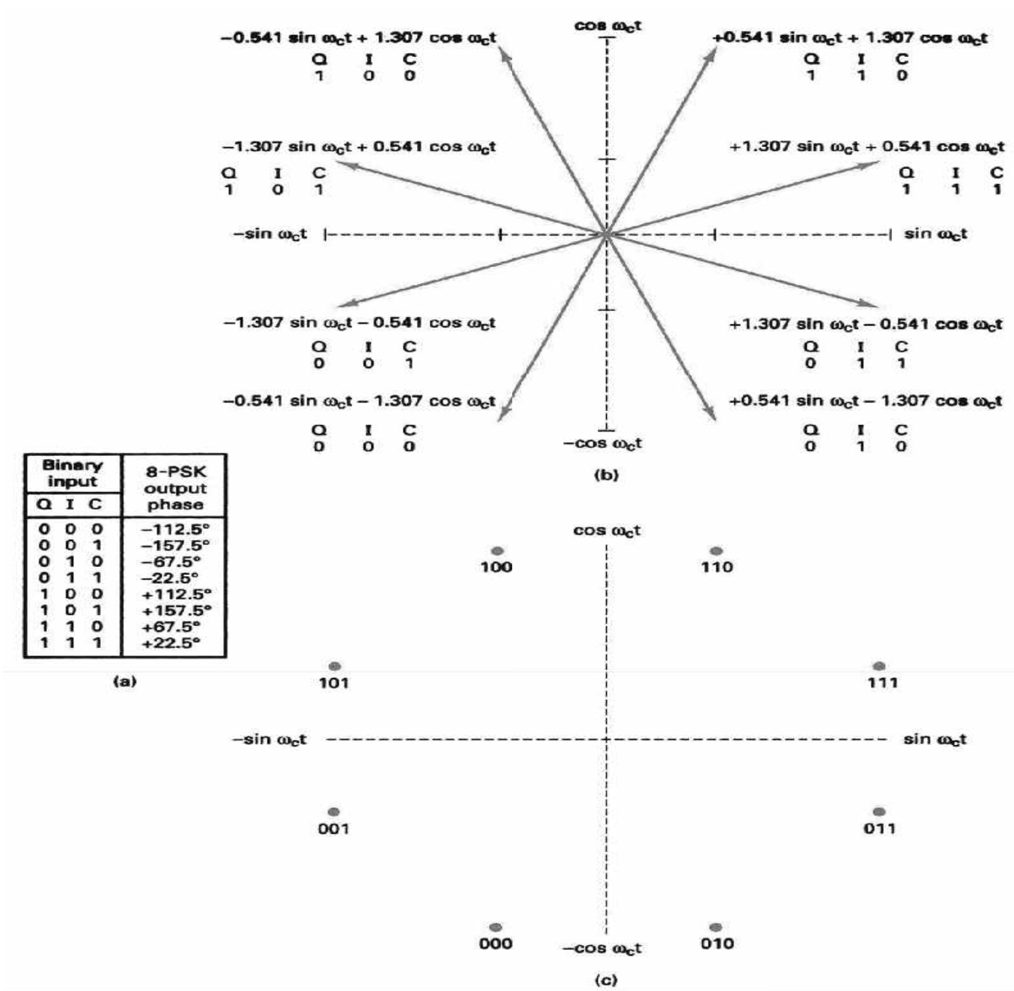
More sophisticated forms of PSK exist. In M-ary or multiple phase-shift keying (MPSK), there are more than two phases, usually four (0, +90, -90, and 180 degrees) or eight (0, +45, -45, +90, -90, +135, -135, and 180 degrees). If there are four phases (m = 4), the MPSK mode is called ***quadrature phase-shift keying*** or quaternary phase-shift keying (QPSK), and each phase shift represents two signal elements. If there are eight phases (m = 8), the MPSK mode is known as ***octal phase-shift keying (OPSK)***, and each phase shift represents three signal elements. In MPSK, data can be transmitted at a faster rate, relative to the number of phase changes per unit time, than is the case in BPSK.

QPSK is an M-ary encoding scheme where N = 2 and M = 4, which has four output phases are possible for a single carrier frequency needing four different input conditions. With two bits, there are four possible conditions: 00, 01, 10, and 11. With QPSK, the binary input data are combined into groups of two bits called ***dibits***.

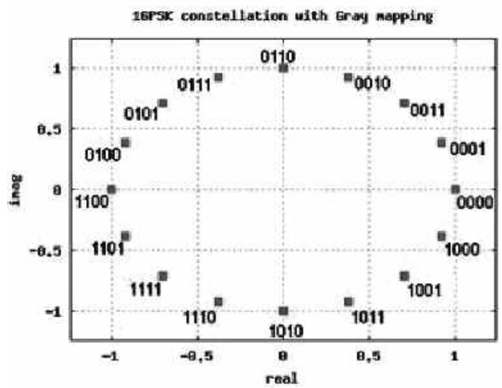


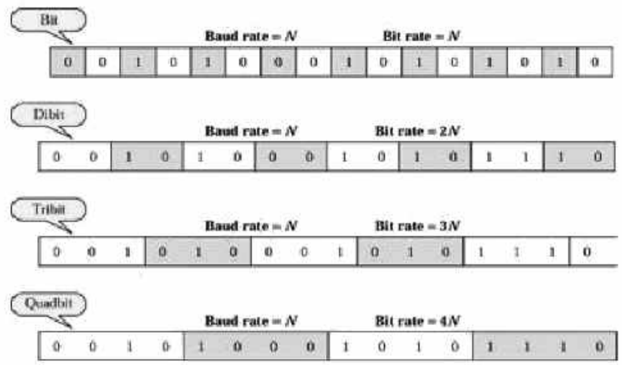
The above figure shows the output phase-versus-time relationship, truth table, and constellation diagram for QPSK. A phase of 0^0 now represents 00; 90^0 represents 01; 180^0 represents 10; and 270^0 represents 11. Data can be transmitted twice as efficiently using 4-PSK than 2-PSK.

With 8-PSK, three bits are encoded forming ***tribits*** and producing eight different output phases. With 8-PSK, N = 3, M = 8, and the minimum bandwidth and baud equal one third the bit rate (f_b /3). 8-PSK is 3 times as efficient as 2-PSK.



With 16-PSK, four bits called ***quadbits*** are combined, producing 16 different outputs phases. With 16-PSK, N = 4, M = 16, and the minimum bandwidth and baud equal one-fourth the bit rate (f_b /4).



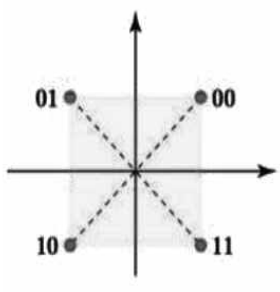


Modulation	Bit Rate	Encoding Scheme	Bandwidth Efficiency	Outputs Possible	Minimum Bandwidth	Baud
ASK	N	Single bit	1	2	f^{a}_{b}	f_b
FSK	N	Single bit	1	2	$>f_b$	f_b
BPSK	N	Single bit	1	2	f_b	f_b
QPSK	2N	Dibits	2	4	f_b /2	f_b /2
8-PSK	3N	Tribits	3	8	f_b /3	f_b /3
16-PSK	4N	Quadibits	4	16	f_b /4	f_b /4

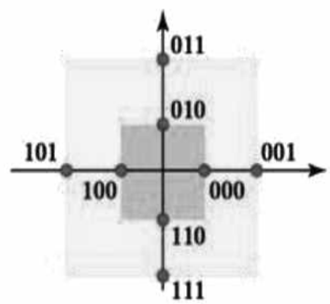
Quadrature Amplitude Modulation (QAM)

PSK is limited by the ability of the equipment to distinguish small differences in phase. Bandwidth limitations make combinations of FSK with other changes practically useless. Quadrature amplitude modulation is a combination of ASK and PSK so that a maximum contrast between each signal unit (bit, dibit, tribit, and so on) is achieved. QAM is used extensively as a modulation scheme for digital telecommunication systems. The primary advantage of QAM over PSK is immunity to transmission impairments, especially phase impairments that are inherent in all communication systems.

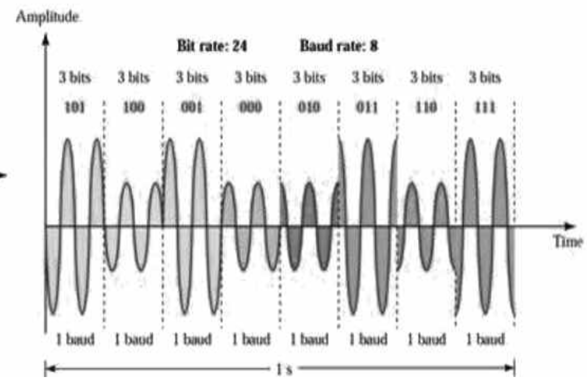
In 4-QAM and 8-QAM, number of amplitude shifts is fewer than the number of phase shifts. Because amplitude changes are susceptible to noise and require greater shift differences than do phase changes, the number of phase shifts used by a QAM system is always larger than the number of amplitude shifts.



4-QAM
1 amplitude, 4 phases



8-QAM
2 amplitudes, 4 phases



With 16-QAM, there are 12 phases and three amplitudes that are combined to produce 16 different output conditions. With QAM, there are always more phases possible than amplitude.

Bandwidth Efficiency

Bandwidth efficiency is often used to compare the performance of one digital modulation technique to another. It is the ration of transmission bit rate to the minimum bandwidth required for a particular modulation scheme. Mathematically represented as:

Bη = transmission bit rate (bps) / minimum bandwidth (Hz)

Modulation	Encoding Scheme	Outputs Possible	Minimum Bandwidth	Baud	Bη
ASK	Single bit	2	f_b	f_b	1
FSK	Single bit	2	f_b	f_b	1
BPSK	Single bit	2	f_b	f_b	1
QPSK	Dibits	4	$f_b/2$	$f_b/2$	2
8-PSK	Tribits	8	$f_b/3$	$f_b/3$	3
8-QAM	Tribits	8	$f_b/3$	$f_b/3$	3
16-PSK	Quadbits	16	$f_b/4$	$f_b/4$	4
16-QAM	Quadbits	16	$f_b/4$	$f_b/4$	4
32-PSK	Five bits	32	$f_b/5$	$f_b/5$	5
64-QAM	Six bits	64	$f_b/6$	$f_b/6$	6

Note: f_b indicates a magnitude equal to the input bit rate.

ASK, FSK, PSK, and QAM Summary

Trellis Code Modulation

Data Transmission Rates in excess of 56 kbps can be achieved over standard telephone circuits using an encoding scheme called trellis code modulation(TCM) developed by

Dr. Ungerboeck. It combines encoding and modulation to reduce the probability of error, thus improving the bit error performance and it uses conventional (tree) codes.

Trellis coding defines the manner in which signal-state transitions are allowed to occur, and transitions that do not follow this pattern are interpreted as transmission errors. TCM can improve error performance by restricting the manner in which signals are allowed to transition. TCM improves on standard QAM by increasing the distance between symbols on the constellation *(called Euclidean distance)*.

Appendix (some additional figures)

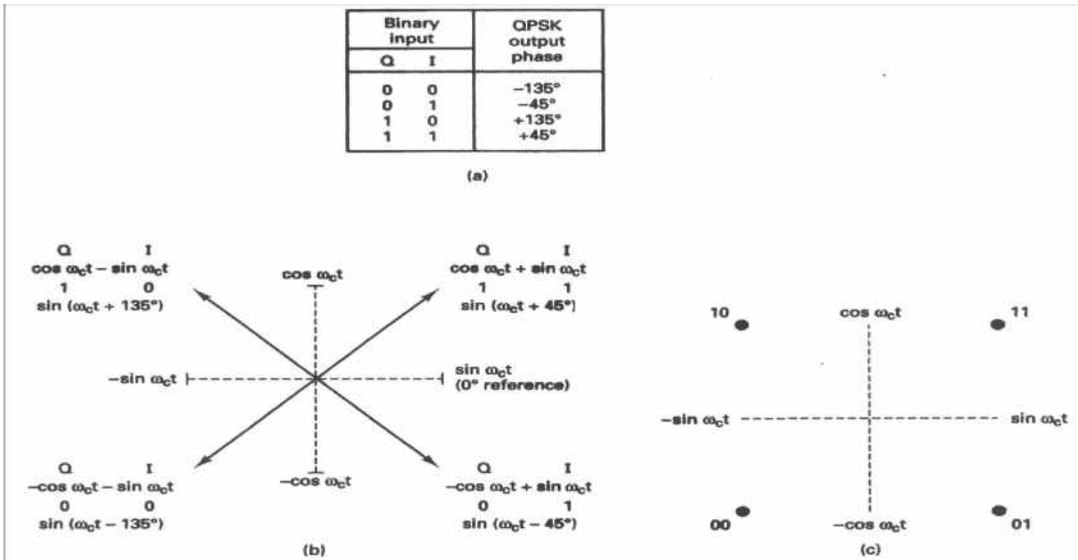


FIGURE 9-18 QPSK modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram

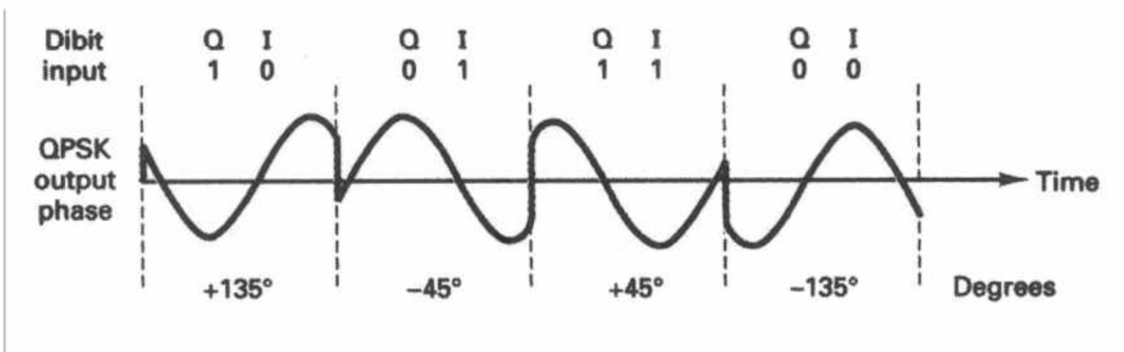
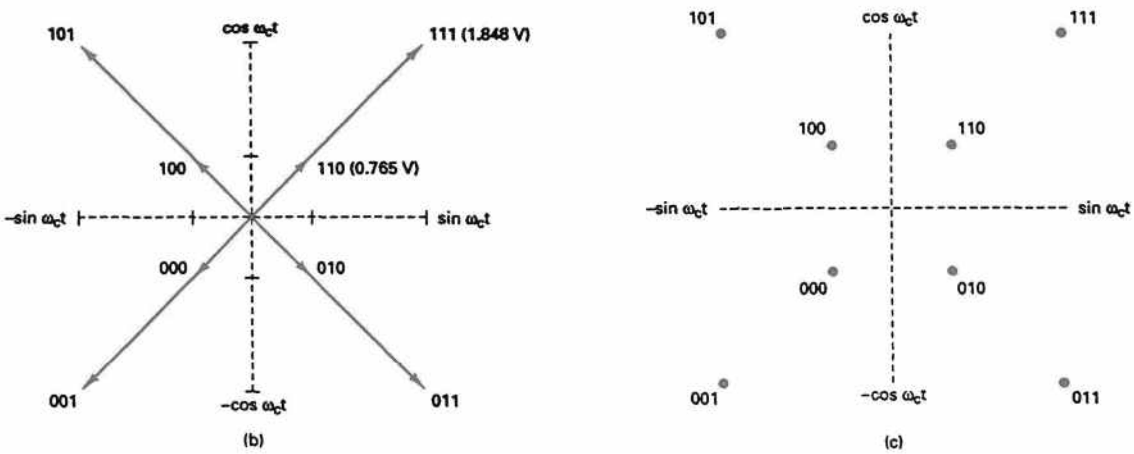


FIGURE 9-19 Output phase-versus-time relationship for a QPSK modulator

Binary input			8-QAM output	
Q	I	C	Amplitude	Phase
0	0	0	0.765 V	−135°
0	0	1	1.848 V	−135°
0	1	0	0.765 V	−45°
0	1	1	1.848 V	−45°
1	0	0	0.765 V	+135°
1	0	1	1.848 V	+135°
1	1	0	0.765 V	+45°
1	1	1	1.848 V	+45°

(a)



8-QAM modulator: (a) truth table; (b) phasor diagram; (c) constellation diagram

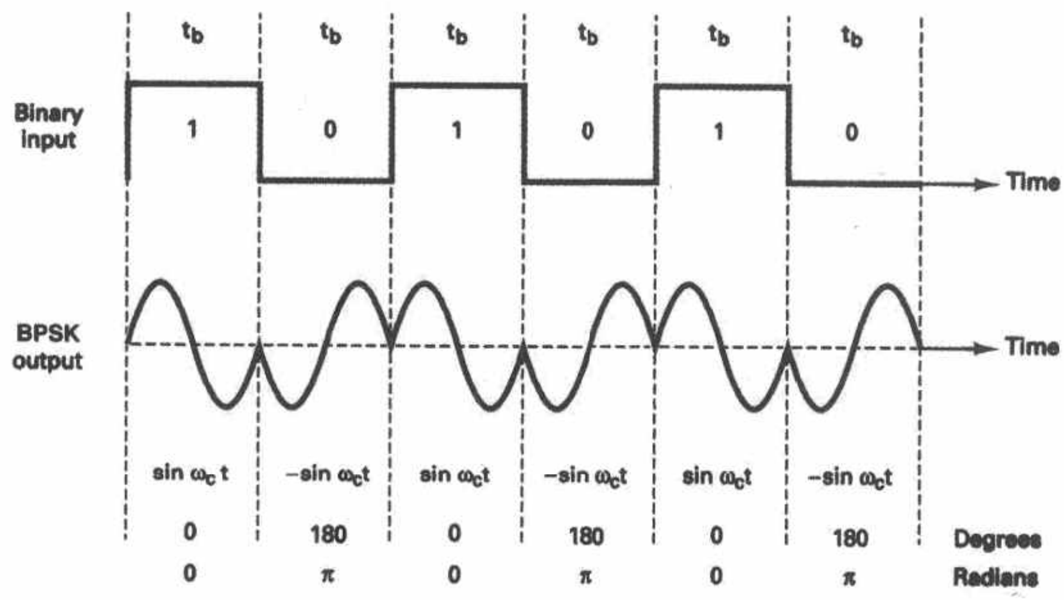


FIGURE 9-15 Output phase-versus-time relationship for a BPSK modulator

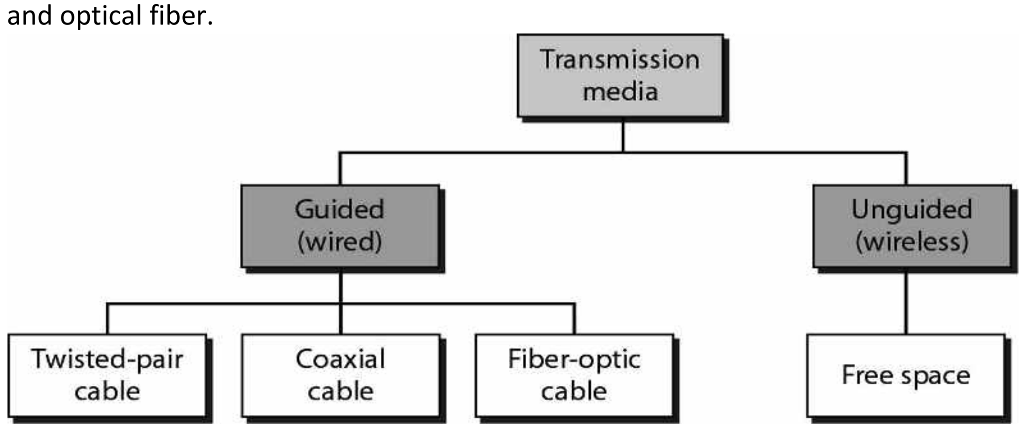
UNIT - III

METALLIC CABLE TRANSMISSION MEDIA :
Metallic Transmission Lines, Transverse Electromagnetic Waves, Characteristics of Electromagnetic Waves, Transmission Line Classifications, Metallic Transmission Line Types, Metallic Transmission Line Equivalent Circuit, Wave Propagation on Metallic Transmission Lines, Metallic Transmission Line Losses.

Introduction

The ***transmission medium*** is the physical path between transmitter and receiver in a data transmission system. It is included in the physical layer of the OSI protocol hierarchy. The transmission medium is usually free space, metallic cable, or fiber-optic cable. The information is usually a signal that is the result of a conversion of data from another form.

Transmission media can be generally categorized as either *unguided or guided.* Guided Transmission Media uses a "cabling" system (or some sort of conductor) that guides the data signals along a specific path. The data signals are bound by the "cabling" system. Guided Media is also known as Bound Media. The conductor directs the signal propagating down it. Only devices physically connected to the medium can receive signals propagating down a guided transmission medium. Examples of guided transmission media are copper wire and optical fiber.



Unguided Transmission Media consists of a means for the data signals to travel but nothing to guide them along a specific path. The data signals are not bound to a cabling media and as such are often called Unbound Media. Unguided transmission media are wireless systems. Signals propagating down an unguided transmission medium are available to anyone who has a device capable of receiving them.

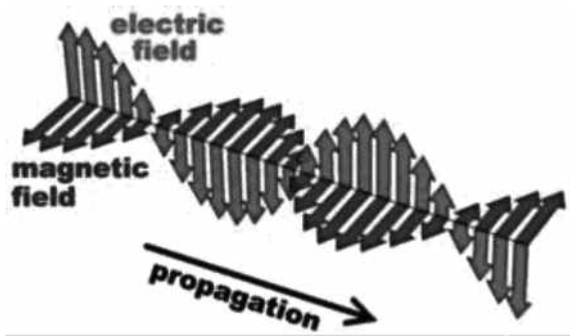
A physical facility is one that occupies space and has weight as opposed to wireless media such as earth's atmosphere or a vacuum and includes metallic cables and optical cables. Metallic transmission lines includes open-wire, twin-lead, and twisted-pair copper wire as well as coaxial cable, and optical fibers include plastic- and glass-core fibers encapsulated in various kinds of cladding materials.

Metallic Transmission Lines

A transmission line is a metallic conductor system used to transfer electrical energy from one point to another using electrical current flow. It is two or more electrical conductors separated by a nonconductive insulator (dielectric). It can be of varied lengths varying from few inches to several thousand miles. It can be used to propagate dc or low-frequency ac and also very high frequencies such as microwave radio-frequency signals.

Transverse Electromagnetic Waves

The two basic kinds of waves are longitudinal and transverse. With longitudinal waves, the displacement is in the direction of propagation. A surface wave or sound waves can be said as examples of longitudinal waves. With transverse waves, the direction or displacement is perpendicular to the direction of propagation. Electromagnetic waves are transverse waves.



Propagation of electrical power along a transmission line occurs in the form of transverse electromagnetic (TEM) waves. TEM wave propagates primarily in the non-conductor that separates the two conductors of the transmission line. The electric field (E) and magnetic field (H) are perpendicular to each other at all points. This is referred to as space or quadrature. Electromagnetic waves that travel along a transmission line from the source to the load are called *incident waves* and those that travel from the load back towards the source are called *reflected waves*.

Characteristics of Electromagnetic waves

The three main characteristics are wave velocity, frequency and wavelength.

Wave velocity: Waves travel at different speeds depending on the type of wave and the characteristics of the propagation medium. Sound travels at 1100 feet/second in normal atmosphere where electromagnetic waves travel much faster. In free space i.e. in vacuum, TEM waves travel at the speed of the light, **c** (approximately at 186,000 miles/sec) and slightly slower in air and considerably slower along a transmission line.

Frequency and Wavelength: The oscillations of an electromagnetic wave are periodic and repetitive. The rate at which the periodic wave repeats is its frequency. The distance of one cycle occurring in space is called the wavelength and is given by

Distance = velocity × time

If the time for one cycle is substituted above, we get the length of one cycle which is called wavelength and is given by

λ **= velocity × period =** $\boldsymbol{v \times T}$, where λ is wavelength, v is velocity and T is period because $T = 1/f$, we can write $\boldsymbol{\lambda = v/f}$

As for free space propagation, v = c; the length of one cycle is $\lambda = c/f = 3\times10^{8}\ \text{m/s} / f\ \text{cycles/s}$

Transmission line classifications

Balanced Transmission Line

In two wire balanced lines, both conductors carry current. One conductor carries the signal and the other conductor in the return path. This type of transmission is called *differential or balanced signal transmission*. Both conductors in a balanced line carry signal currents, which are equal in magnitude with respect to electrical ground but travel in opposite directions.

Currents that flow in opposite directions in a balanced wire pair are called *metallic circuit currents* and currents that flow in same direction are called *longitudinal currents*. The chief advantage of the balanced line format is good rejection of external noise. Common forms of balanced line are twin-lead, used for radio frequency signals and twisted pair, used for lower frequencies.

Unbalanced Transmission Line

With an unbalanced transmission line, one wire is at ground potential, whereas the other wire is at signal potential. This type of transmission line is called *single-ended or unbalanced* signal transmission. The ground wire may also be the reference for other signal-carrying wires and must go anywhere any of the signal wires go.

Unbalanced transmission lines have the advantage of requiring only one wire for each signal and only one ground line is required no matter how many signals are grouped into one conductor. Balanced transmission lines can be connected to unbalanced transmission lines and vice versa with special transformers called *balums.*

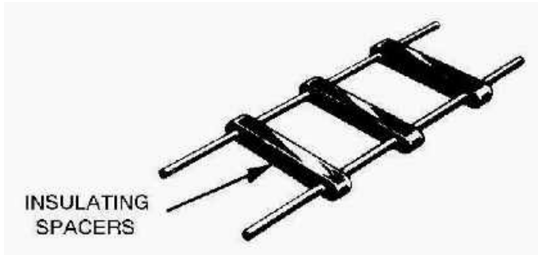
Metallic Transmission Line Types

All data communications systems and computer networks are interconnected to some degree with cables, which form the most important part of the transmission medium transporting signals between computers.

Parallel-Conductor Transmission Lines

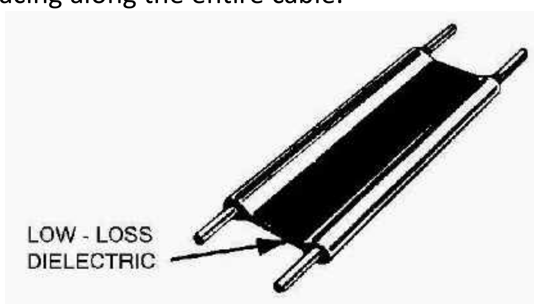
Parallel-wire transmission lines are comprised of two or more metallic conductors separated by a nonconductive insulating material called a dielectric. Common dielectric materials include air, rubber, polyethylene, paper, mica, glass and Teflon. The most common parallel-conductor transmission lines are open-wire, twin lead and twisted pair, including unshielded twisted pair (UTP) and shielded twisted pair (STP).

Open-Wire Transmission Lines: These are two-wire parallel conductors, closely spaced and separated by air. Non conductive spacers are placed at periodic intervals not only for support but also to keep the distance between the conductors constant. TEM wave propagates in the air between the conductors, which acts as dielectric. The main advantage is its simple construction.



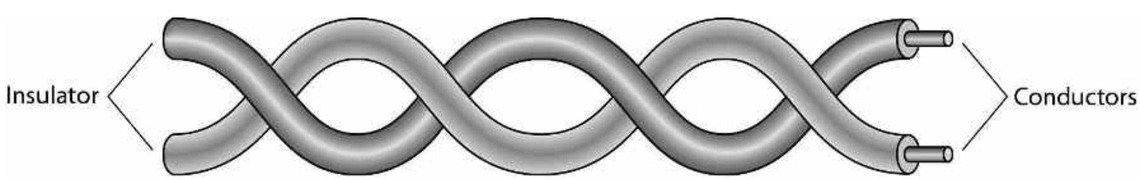
Since no shielding is present, the radiation losses are high and cable is susceptible to picking up signals through mutual induction, which produces crosstalk. The primary usage is in standard voice-grade telephone applications.

Twin lead: Twin-lead is essentially the same as open-wire transmission line except that the spacers between the two conductors are replaced with a continuous solid dielectric ensuring the uniform spacing along the entire cable.



It is mainly used to connect televisions to rooftop antennas. Common dielectric materials used with twin-lead cable are Teflon and polyethylene.

Twisted-pair transmission lines: A twisted-pair (TP) transmission line is formed by twisting two insulated conductors around each other. Usually, a number of pairs of these wires are put together into a cable. The cable may contain more than a hundred pairs of wires for long-distance communications. Twisted-pair wires are the most common media in a telephone network. These wires support both analog and digital signals and can transmit the signal at a speed of 10 Mbps over a short distance. The twisting of wires with different twisting lengths reduces the effect of cross talk and low-frequency interference.

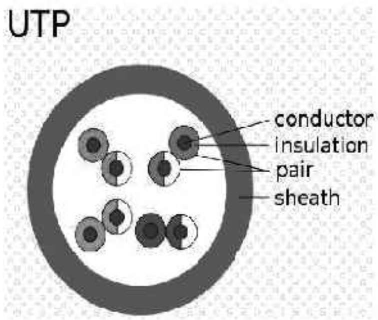


Twisted-pair cable

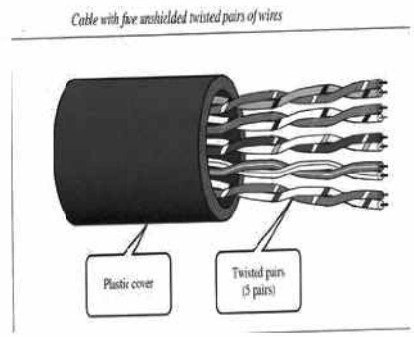
Twisted-pair transmission lines are also the transmission medium of choice for most local area networks because twisted-pair cable is simple to install and relatively independent when compared to coaxial and optical fiber cables.

The two basic types of twisted-pair transmission lines specified are unshielded twisted pair (UTP) and shielded twisted pair (STP).

Unshielded twisted-pair: An UTP cable consists of two copper wires where each wire is seperately encapsulated in PVC (polyvinyl chloride) insulation. Bandwidth can be improved by controlling the number of twists per foot and also the manner in which multiple pairs are twisted around each other. The minimum number of twists for UTP cable is two per foot.



Unshielded twisted pair cable with different twist rates



UTPs are cheaper, more flexible, and easier to install. They provide enough support for telephone systems and are not covered by metal insulation. They offer acceptable performance for a long-distance signal transmission, but as they are uninsulated, they are affected by cross talk, atmospheric conditions, electromagnetic interference, and adjacent twisted pairs, as well as by any noise generated nearby. The majority of the telephone twisted pairs are unshielded and can transmit signals at a speed of 10 Mbps.

The Electronic Industries Association (EIA) has developed standard to grade UTP cable by quality; Category 1 as the lowest quality and category 6 as the highest quality.

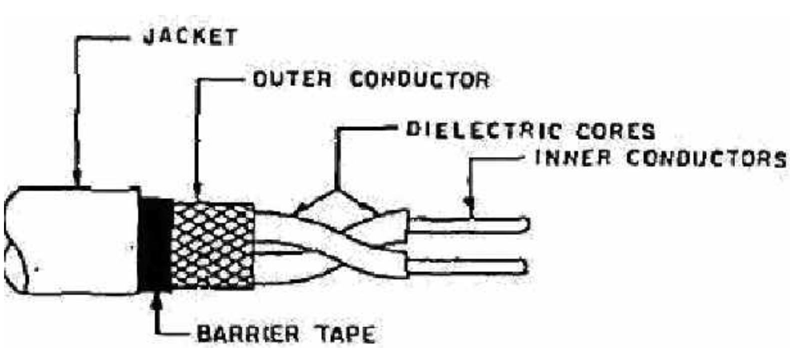
1. Category 1: The basic twisted-pair cabling used in telephone systems. This level of quality is fine for voice but inadequate for data transmission.
2. Category 2: This category is suitable for voice and data transmission of up to 2Mbps.
3. Category 3: This category is suitable for data transmission of up to 10 Mbps. It is now the standard cable for most telephone systems. At least three twist per feet
4. Category 4: This category is suitable for data transmission of up to 20 Mbps.
5. Category 5: This category is suitable for data transmission of up to 100 Mbps.
6. Category 6: CAT- 6 is recently proposed cable type comprised of four pairs of wire capable of operating at transmission rates of up to 400Mbps.

Advantages of UTP are its easy to terminate, installation costs are less and more lines can be run through the same wiring ducts. Disadvantages of UTP are its a bit noisy and prone to interference.

Category	*Specification*	*Data Rate (Mbps)*	*Use*
1	Unshielded twisted-pair used in telephone	< 0.1	Telephone
2	Unshielded twisted-pair originally used in T-lines	2	T-1 lines
3	Improved CAT 2 used in LANs	10	LANs
4	Improved CAT 3 used in Token Ring networks	20	LANs
5	Cable wire is normally 24 AWG with a jacket and outside sheath	100	LANs
5E	An extension to category 5 that includes extra features to minimize the crosstalk and electromagnetic interference	125	LANs
6	A new category with matched components coming from the same manufacturer. The cable must be tested at a 200-Mbps data rate.	200	LANs
7	Sometimes called SSTP (shielded screen twisted-pair). Each pair is individually wrapped in a helical metallic foil followed by a metallic foil shield in addition to the outside sheath. The shield decreases the effect of crosstalk and increases the data rate.	600	LANs

Categories of unshielded twisted-pair cables

Shielded Twisted Pair (STP) Cable: STP cable is a parallel two-wire transmission line consisting of two copper conductors separated by a solid dielectric material. The wires and dielectric are enclosed in a conductive-metal sleeve called a foil. If the sleeve is woven into a mesh, it's called braid.

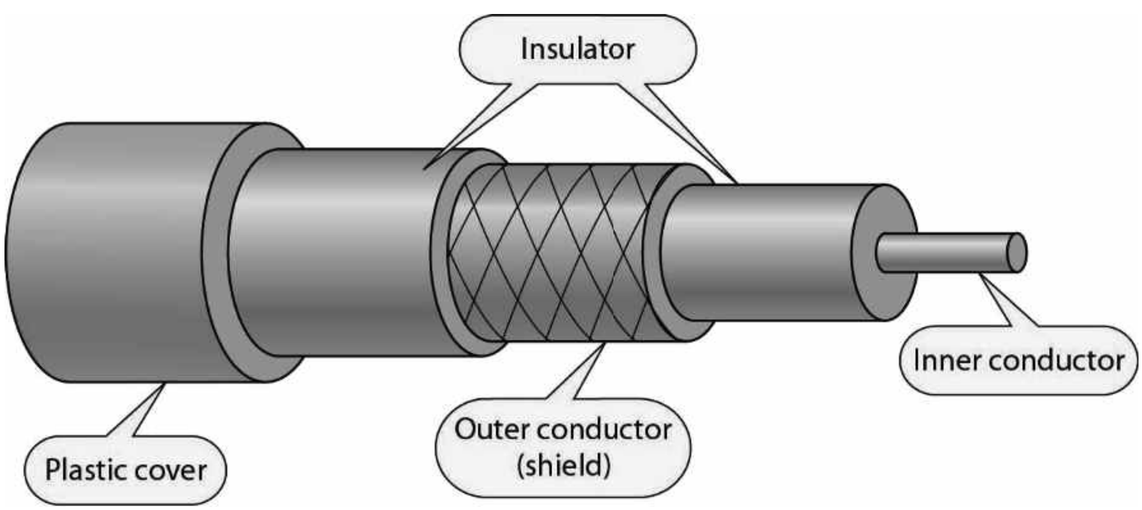


The metal casing prevents the penetration of electromagnetic noise. Materials and manufacturing requirements make STP more expensive than UTP but less susceptible to noise.

Plenum Cable: Plenum cables are the electrical or telecommunication cables (or wires) which are installed in environmental air spaces in the interior of many commercial and residential buildings. It is common practice to route communication cables and the like for computers, data devices, and alarm systems through plenums in building constructions. If a fire occurs in a building which includes plenums or risers, the non-fire retardant plenum construction would enable the fire to spread very rapidly throughout the entire building. Typically plenum data cables have two or more pairs of insulated conductors in a common jacket. The insulation can be made of several types of flame retardant insulation. A plenum is defined as a compartment or chamber to which one or more air ducts are connected and which forms part of the air distribution system of the structure. Plenum cables have a plurality of twisted pair conductors surrounded by a jacket. The twisted pairs generally all have the same twist or substantially the same twist. A typical and widely used flame retardant insulation for conductors in data plenum cables is fluorinated ethylene-propylene. Category 5 plenum cable made of jacketed twisted pairs of insulated conductors has to satisfy a number of electrical requirements set by the EIA/TIA specification 568A.

Coaxial (Concentric) Transmission Lines

Because of the advent of modern UTP and STP twisted pair cables, coaxial cable is seen very less in computer networks, but still has very high importance in analog systems, such as cable television distribution networks. The basic coaxial cable consists of a center conductor surrounded by a dielectric material (insulation), then a concentric (uniform distance from the center) shielding, and finally a rubber environmental protection outer jacket. A coaxial cable with one layer of foil insulation and one layer of braided shielding is referred to as *dual shielded* and if two layers of foil insulation and two layers of braided metal shielding are present, it's called *quad shielding.*



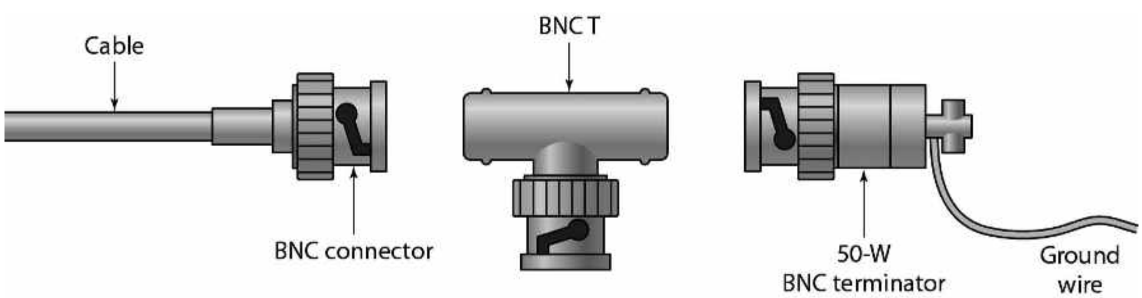
Two basic types of coaxial cables are present: *rigid air filled and solid flexible.* Rigid air-filled cables are relatively expensive and are tough to maintain. Coaxial cables are capable of operating at higher bit rates than their parallel-wire counterparts, very secure than twisted-pair cable, can be used over long distances, immune to external radiation and radiate little themselves. Disadvantages of coaxial transmission lines are their poor cost-to-performance ratio, low reliability, and high maintenance.

The RG numbering system used with coaxial cables refers to cables approved by U.S. Department of Defense (DoD).

Category	Impedance	Use
RG-59	75 Ω	Cable TV
RG-58	50 Ω	Thin Ethernet
RG-11	50 Ω	Thick Ethernet

Categories of coaxial cables

To connect coaxial cable to devices, it is necessary to use coaxial connectors. The most common type of connector is the Bayone-Neill-Concelman, or BNC, connectors. BNC connectors are sometimes referred to as bayonet mount, as they can be easily twisted on or off.



There are three types: the BNC connector, the BNC T connector, the BNC terminator. Applications include cable TV networks, and some traditional Ethernet LANs like 10Base-2, or 10-Base5.

Metallic Transmission Line Equivalent Circuit

The characteristics of a transmission line are determined by its electrical properties like wire conductivity, insulator dielectric constant and its physical properties like wire diameter and conductor spacing. These properties in turn determine the primary electric constants: *series resistance (R), series inductance (L), shunt capacitance (C), and shunt conductance (G).* Resistance and inductance occur along the line, whereas capacitance and conductance occur between the conductors.

Characteristic Impedance

For maximum power transfer from the source to load, a transmission line must be terminated in a purely resistive load equal to the characteristic impedance of the transmission line. Transmission line stores energy in its distributed inductance and capacitance.

Using Ohm's law, the characteristic impedance is simply the ratio of the source voltage (E_O) to the line current (I_O), given by

$Z_0 = E_0 / I_0$, where Z_O is characteristic impedance in ohms, E_O is source voltage in volts and I_O is transmission line current in amps.

- Characteristic impedance of a two wire parallel transmission line with an air dielectric can be determined from its physical dimensions $Z_0 = 276 \log D/r$ where D is distance between the centres of the two conductors and R is radius of the conductors.
- Characteristic impedance of a coaxial cable can also be determined from its physical dimensions: $Z_o = \frac{138}{\sqrt{\varepsilon_r}} (\log D/d)$ where, D is inside diameter of the conductor and ε_r is relative dielectric constant of the insulating material.

Wave Propagation on Metallic Transmission Lines

EM waves travel at the speed of light through vacuum and nearly the same through air, but they travel considerably slowly in metallic transmission lines, where the conductor is generally copper and the dielectric materials vary with cable type.

Velocity Factor and Dielectric Constant

Velocity factor is defined as the ratio of the actual velocity of propagation of an electromagnetic wave through a given medium to the velocity of propagation through a vacuum. Mathematically, given as:

$V_f = V_p / c$, where V_f is velocity factor, V_p is actual velocity of propagation and c is velocity of propagation through a vacuum (3×10^8 m/s).

Dielectric constant is simply the relative permittivity of a material. The dielectric constant depends on the type of insulating material used. The velocity at which an EM wave propagates along a transmission line varies with the inductance and capacitance of the cable. Time can be given as: $T = \sqrt{LC}$. Inductance, capacitance ad velocity of propagation can be given by the formula, **velocity × Time = Distance**

Therefore, $V_p = \text{Distance}/\text{Time} = D/T$ which can be written as $V_p = D/\sqrt{LC}$

If the distance is normalized to 1 meter, the velocity of propagation for a lossless transmission line is $V_p = 1/\sqrt{LC}$

Metallic Transmission Line Losses

Signal power is lost in a transmission line through different ways: *conductor loss, radiation loss, dielectric heating loss, coupling loss and corona.* All these losses are lumped together and are specified as attenuation loss in decibels per unit length.

Conductor Losses: As electrical current flows through a metallic transmission line, there is an inherent and unavoidable power loss because of the finite resistance present in the line. This loss is termed as conductor loss or conductor heating loss and is simply I^2r power loss.

Radiation Losses: Radiation and Induction losses are similar in that both are caused by the fields surrounding the conductors. Induction losses occur when the electromagnetic field about a conductor cuts through any nearby metallic object and a current is induced in that object. Radiation losses are reduced by properly shielding the cable. Therefore, STP and coaxial cables have less radiation than UTP, twin lead and openwire.

Coupling Losses: Coupling loss occurs whenever a connection is made to or from a transmission line or when two sections of transmission line are connected together. Discontinuities are the locations where dissimilar materials meet and they tend to heat up, radiate energy, and dissipate power.

Corona: Corona is a luminous discharge that occurs between the two conductors of a transmission line, when the difference of potential between them exceeds the breakdown voltage of the dielectric insulator. When corona occurs, the transmission line is destroyed.

OPTICAL FIBER TRANSMISSION MEDIA :

Advantages of Optical Fiber Cables, Disadvantages of Optical Fiber Cables, Electromagnetic spectrum, Optical Fiber Communications System Block Diagram, Optical Fiber construction, The Physics of Light, Velocity of Propagation, Propagation of Light Through an Optical fiber Cable, Optical Fiber Modes and Classifications, Optical Fiber Comparison, Losses in Optical Fiber Cables, Light sources, Light Detectors, Lasers.

An optical communications system is one that uses light as the carrier of information. They use glass or plastic fiber cables to contain the light waves and guide them in a manner similar to the way EM waves are guided through a metallic transmission media.

Advantages of Optical Fiber Cables

- *Wider bandwidth and greater information capacity*: The light wave occupies the frequency range between 2×10^{12} Hz to 37×10^{12} Hz. This makes the information carrying capability of fiber optic cables is much higher.
- *Immunity to crosstalk*: Since fiber optic cables use glass and plastic fibers, which are non-conductors of electrical current, no magnetic field is present. No magnetic induction means no crosstalk.
- *Immunity to static interference*: As optical fiber cables are non-conductors, they are immune to electromagnetic interference (EMI) caused by lightning, electric motors, relays, fluorescent lights and other electrical noise sources.
- *Environmental immunity*: Optical fibers are more immune to environmental extremes. They can operate over large temperature variations and are also not affected by corrosive liquids and gases.
- *Safety and convenience*: As only glass and plastic fibers are present, no electrical currents or voltages are associated with them. Also they can be used around any volatile liquids and gasses without worrying about their causing explosions or fires.
- *Lower transmission loss*: Fiber optic cables offers less signal attenuation over long distances. Typically, it is less than 1 dB/km
- *Security*: Optical fibers are more secure as they are almost impossible to tap into because they do not radiate signals. No ground loops exist between optical fibers hence they are more secure.
- *Durability and reliability*: Optical cables last longer and are more reliable than metallic facilities because fiber cables have a higher tolerance to changes in environmental conditions and are immune to corrosive materials.
- *Economics*: Cost of optical fiber cables is same as metallic cables. Fiber cables have less loss and require fewer repeaters, which in turn needs lower installation and overall system costs.

Disadvantages of Optical Fiber Cables

- *Interfacing costs*: As optical cables need to be connected standard electronic facilities requiring expensive interfaces
- *Strength*: Optical cables have lower tensile strength than coaxial cable. They need an extra coating of Kevlar and also a protective jacket of PVC. Glass fiber is also fragile making them less attractive in case of hardware portability is required

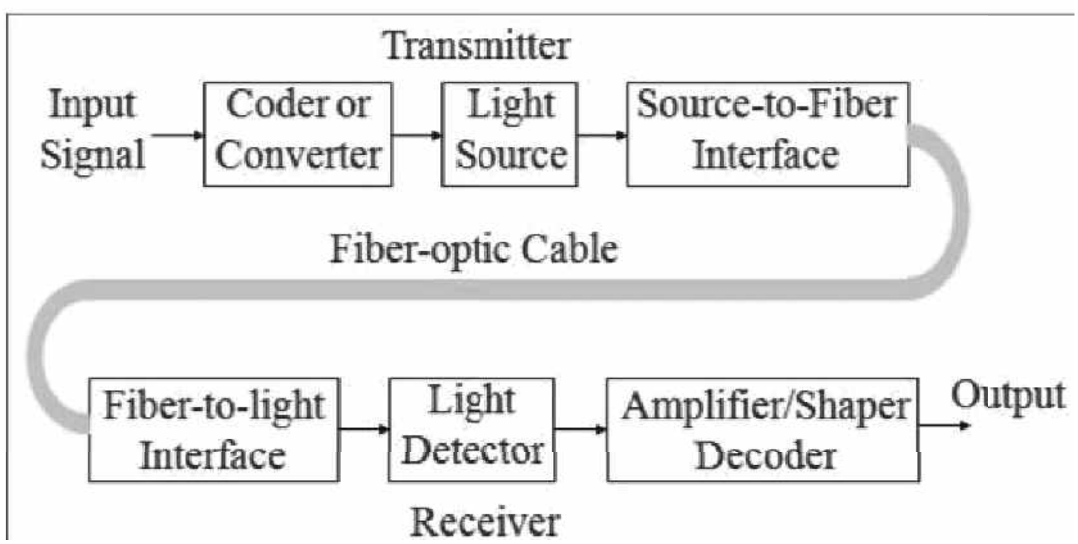
- *Remote electrical power:* Occasionally, electrical power needs to be provided to remote interfaces, which cannot be accomplished using optical cables
- *Losses through bending*: Bending the cable causes irregularities in the cable dimensions, resulting in loss of signal power. Also, optical cables are prone to manufacture defects causing an excessive loss of signal power.
- *Specialized tools, equipment and training*: Special tools are required to splice and repair cables and special test equipment are needed to make routine measurements. Technicians working on optical cables need special skills and training.

Electromagnetic Spectrum

The **electromagnetic spectrum** is the range of all possible frequencies of electromagnetic radiation. The "electromagnetic spectrum" of an object is the characteristic distribution of electromagnetic radiation emitted or absorbed by that particular object. The frequency spectrum extends from the subsonic frequencies (a few hertz) to cosmic rays (10^{23} Hz). The light frequency spectrum can be divided into three general bands.

1. *Infrared:* The band of frequencies that is too high to be seen by the human eye with wavelengths ranging between 770nm and 10^{6} nm. Optical fibers generally operate in infrared band.
2. *Visible:* The band of light frequencies to which the human eye will respond with wave lengths ranging between 390nm and 770nm. This band is visible to human eye.
3. *Ultraviolet*: The band of light frequencies, that are too low to be seen by the human eye with wave lengths ranging between 10nm and 390nm.

Optical Fiber Communications System Block Diagram



The three primary building blocks are transmitter, receiver and the optical fiber cable. The transmitter is comprised of a voltage-to-current converter, a light source, and source-to-fiber interface. The fiber guide is the transmission medium, which is either an ultrapure glass or a plastic cable. The receiver includes a fiber-to-interface, a photodetector, and a current-to-voltage converter.

Optical Fiber Construction

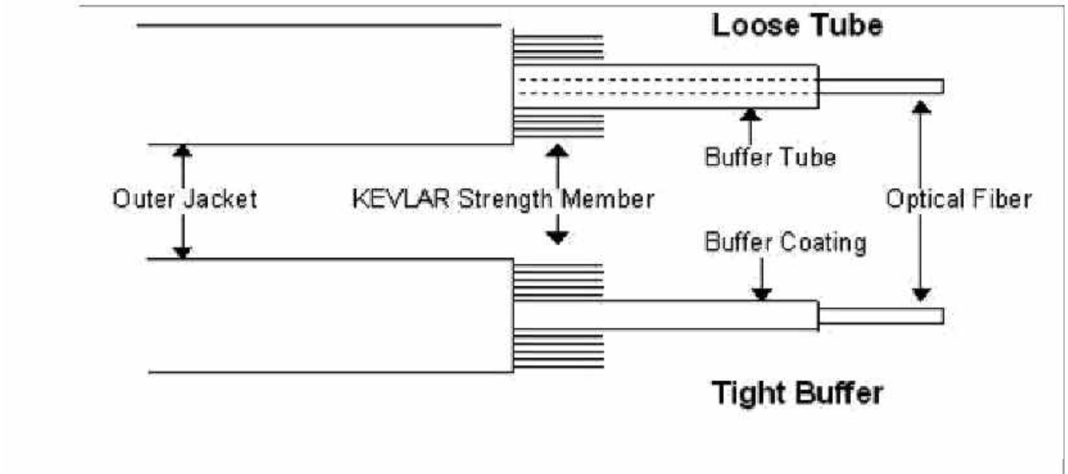
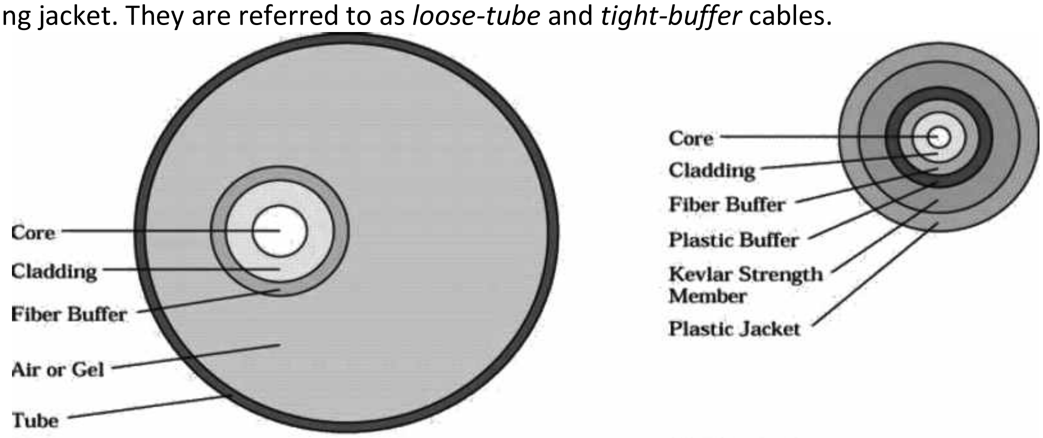


Figure 2, Basic Fiber Optic Cable Construction

There are two basic types of fiber-optic cable. The difference is whether the fiber is free to move inside a tube with a diameter much larger than the fiber or is inside a relatively tight-fitting jacket. They are referred to as *loose-tube* and *tight-buffer* cables.



(a) Loose-tube construction (b) Tight-buffer construction

Both methods of construction have advantages

- Loose-tube cables—all the stress of cable pulling is taken up by the cable's strength members and the fiber is free to expand and contract with temperature
- Tight-buffer cables are cheaper and generally easier to use

Physics of Light

Albert Einstein and Max Planck showed that when light is emitted or absorbed, it behaves like an electromagnetic wave and also like a particle called a photon, which possesses energy proportional to its frequency. This is known as Planck's Law. It states that "when visible light or high-frequency electromagnetic radiation illuminates a metallic surface, electrons are emitted". It is expressed mathematically as:

$$E_p = hf,$$

where E_p is energy of the photons in joules, h is Planck's constant and f is frequency of light

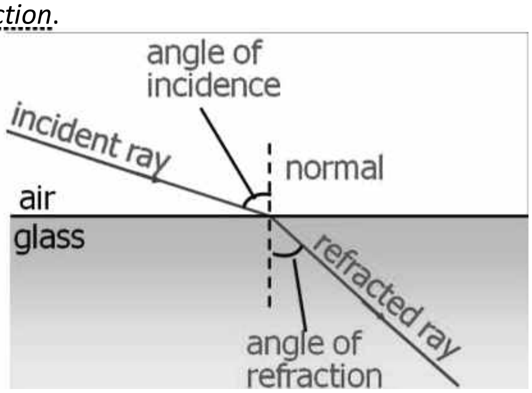
The process of decaying from one energy level to another energy level is called *spontaneous decay or spontaneous emission.* The process of moving from one energy level to another is called *absorption*.

Optical power measures the rate at which electromagnetic waves transfer light energy. It is described as the flow of light energy past a given point in a specified time. Expressed mathematically as:

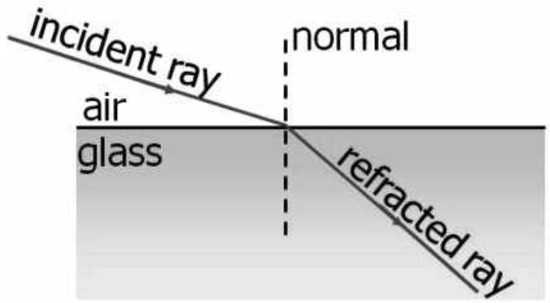
P = d(energy)/ d(time) = *dQ/ dt* , where P is optical power in watts and dQ is instantaneous charge in joules

Velocity of Propagation

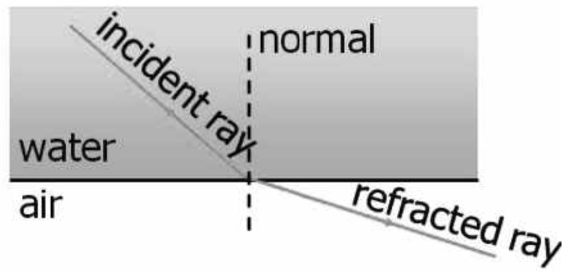
Refraction: Refraction is the bending of light when the light passes from one medium to another. The angle between the light ray and the normal as it leaves a medium is called the *angle of incidence*. The angle between the light ray and the normal as it enters a medium is called the *angle of refraction*.



When an electromagnetic wave is reduced as it passes from one medium to another medium of denser material, the light ray changes direction or refracts (bends) toward the normal. When an electromagnetic wave passes from a more dense material into a less dense material, the light ray is refracted away from the normal. The normal is simply an imaginary line drawn perpendicular to the interface of the two materials at the point of incidence



From Air to Glass: Light is bent **towards** the normal



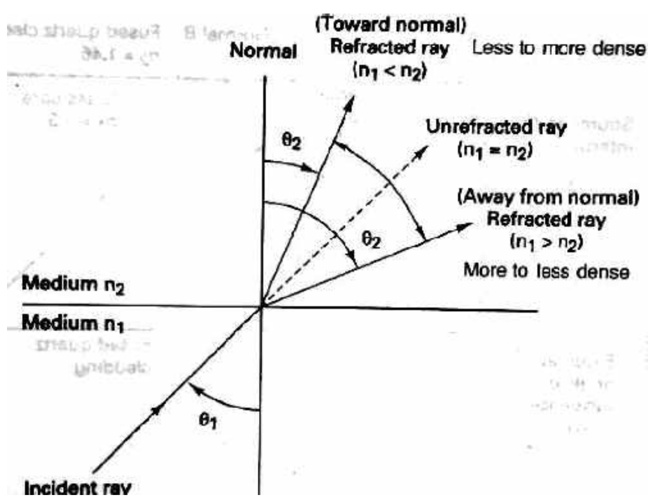
From Water to Air: Light is bent **away** from the normal

Refractive Index: Refractive index is simply the ratio of the velocity of propagation of light ray in free space to the velocity of propagation of a light ray in a given material. Given by,

$n = c/v$, where n is refractive index and c is speed of light (m/sec) and v is speed of light in a given material (m/sec). Typical indexes of refraction of some materials are given below:

Material	Refractive index
Glass	1.5 – 1.7
Water	1.33
Air	1.0001
Diamond	2.0 - 2.42
Vacuum	1.0

Snell's Law: This relationship between the angles of incidence and refraction and the indices of refraction of the two medium is known as ***Snell's Law***. Snell's law applies to the refraction of light in any situation, regardless of what the two media are.



Refractive model for Snell's Law

Snell's Law is stated mathematically as:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

Where, n_1 is refractive index of material 1, n_2 is refractive index of material2, θ_1 is angle of incidence and θ_2 is angle of refraction.

Critical Angle: The angle of incidence is called the critical angle (θ_c), which is defined as the minimum angle of incidence at which a light ray may strike the interface of two media and result in an angle of refraction of 90 degrees or greater. Light ray has to travel from medium of higher refractive index to that of lower refractive index. Expressed as:

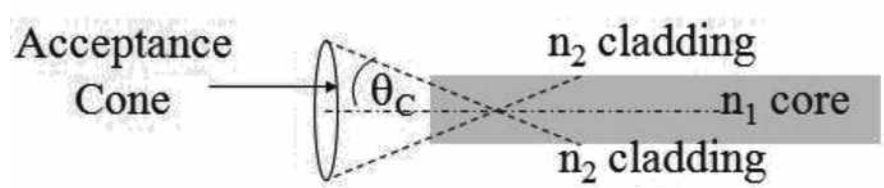
$$\theta_c = \sin^{-1} n_2/n_1$$

Acceptance angle, acceptance cone and numerical aperture: For a ray of light to propagate down the cable, it must strike the internal core/cladding interface at an angle that is greater than the critical angle.

$$\theta_{in(max)} = \sin^{-1} \sqrt{(n_1^2 - n_2^2)}$$

Where, $\theta_{in(max)}$ is acceptance angle or acceptance cone half angle. It defines the maximum angle in which external light rays may strike the air/glass interface and still propagate down the fiber.

Rotating the acceptance angle around the fiber axis, a cone pattern is obtained, called as acceptance cone of the fiber input. The cone of acceptance is the angle within which the light is accepted into the core and is able to travel along the fiber. Launching light wave will be easier for large acceptance cone.



Numerical Aperture (NA) is used to describe the light-gathering or light-collecting ability of an optical fiber. Larger the magnitude of NA, greater the amount of external light the fiber will accept. Described as, $NA = \sin \theta_{in}$ and $NA = \sqrt{(n_1^2 - n_2^2)}$. Therefore, it can be written:

$$\theta_{in} = \sin^{-1} NA$$

Propagation of Light through an Optical Fiber Cable

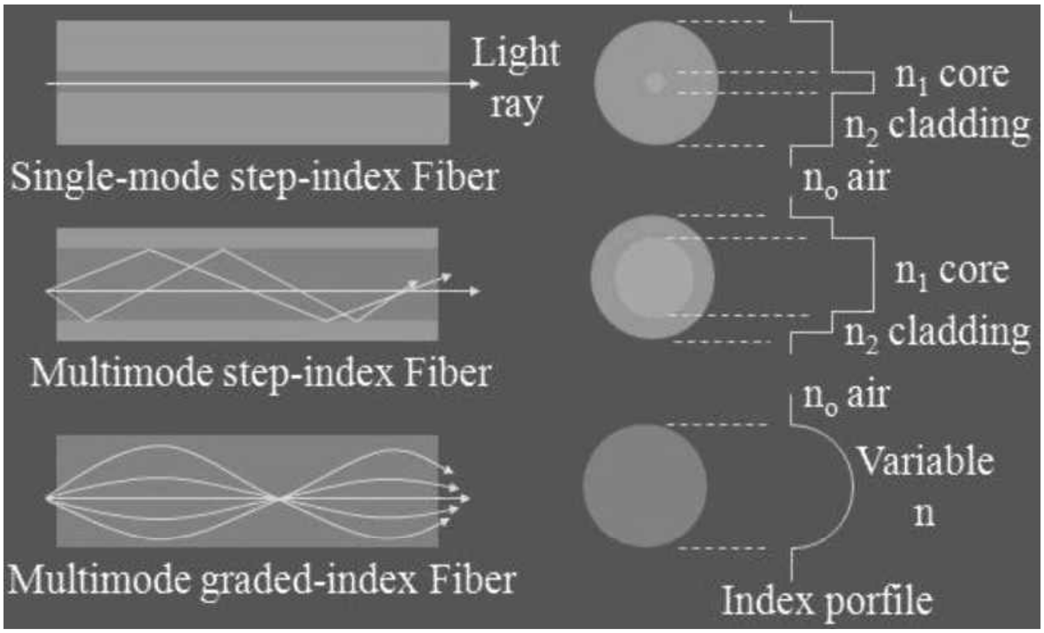
Light can be propagated using either refraction or reflection and the way light propagates depends on the mode of propagation and the index profile of the fiber.

Modes of propagation

Mode simply means path. If there is only one path for light rays to take down a cable, it is called single mode and if there is more than one path, it is called multimode. In single mode, the light travels directly down the center of the cable, whereas for multimode, light rays propagate down the cable in a zigzagging fashion following several paths. The number of modes possible for a given cable can be given by:

$$N \approx \left[\frac{\pi d}{\lambda} \sqrt{(n_1^2 - n_2^2)}\right]$$

Where N is number of modes, d is core diameter and λ is wave length and n1 is refractive index of core and n2 is refractive index of cladding.



Index Profile

Index profile of an optical fiber is a graphical representation of the magnitude of the refractive index across the fiber. The above figure shows the index profiles of three types of fibers. Two basic types of index profiles are present. A *step-index* fiber has a central core with a uniform refractive index. A *graded-index* fiber has no cladding and the refractive index of the core is nonuniform. It is highest at the center of the core and decreases gradually with distance towards the outer edge.

Optical Fiber modes and Classifications

Three practical types of optical fiber configurations: single-mode step index, multimode step index and multimode graded index.

Single-Mode Step-Index Optical Fiber: The fiber has a central core that is sufficiently small that there is essentially only one path for light ray through the cable. In most cases, the outside cladding is air making this fiber to have a wide external acceptance angle making it relatively easy to couple to a light source. But, this type of fiber is very weak and difficult to splice or terminate. A more practical approach will be single mode step-index fiber that has a cladding other than air. This would be physically stronger than air-clad fiber but critical angle will be higher resulting in a small acceptance angle. This makes it difficult to couple light into the fiber from a light source.

Advantages:

- Minimum dispersion: all rays take same path, same time to travel down the cable. A pulse can be reproduced at the receiver very accurately.
- Less attenuation can run over longer distance without repeaters.
- Larger bandwidth and higher information rate

- Difficult to couple light in and out of the tiny core
- Highly directive light source (laser) is required.
- Interfacing modules are more expensive

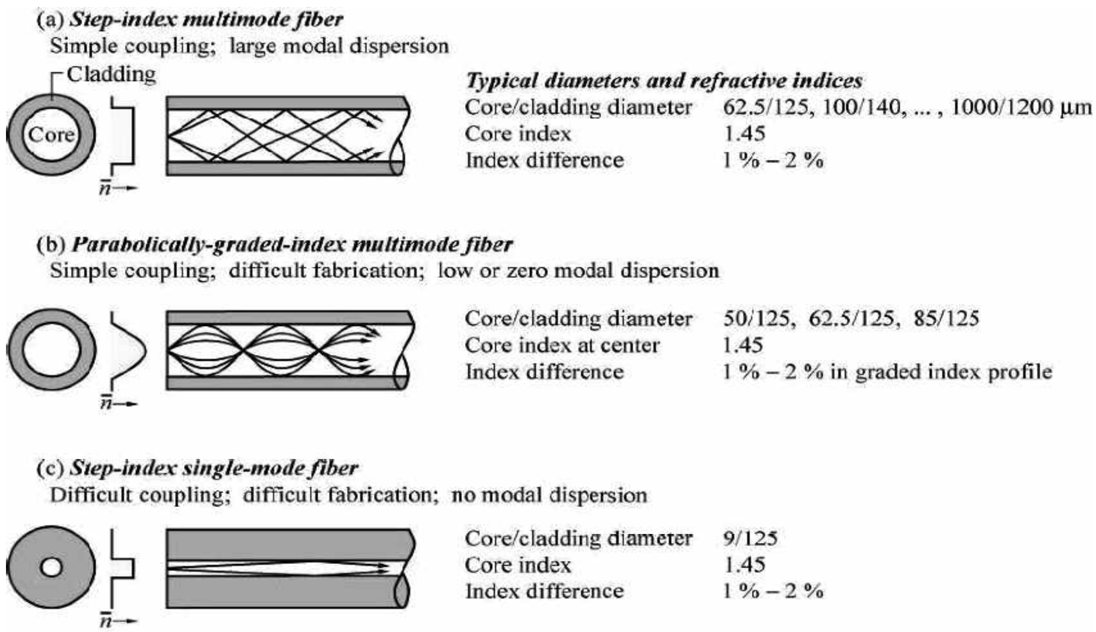


Fig. 22.1. (a) Step-index multimode fibers allow for the propagation of several optical modes. (b) Parabolically graded-index multimode fibers allow for the propagation of several modes with similar propagation constant. Graded-index multimode fibers have a lower modal dispersion than step-index multimode fibers. (c) Step-index single-mode fibers have a small core diameter and no modal dispersion.

E. F. Schubert
Light-Emitting Diodes (Cambridge Univ. Press)
www.LightEmittingDiodes.org

Multimode Step-Index Optical Fiber: These are similar to single mode step-index fibers except that the center core is much larger with the multimode configuration. This type has a large light-to-fiber aperture and therefore allows more external light to enter the cable. Light rays travel down the cable in a zigzag fashion continuously reflecting off the interface boundary. Light rays travel in many paths as it propagates down the fiber. So, all light rays do not follow the same path and do not take same amount of time to travel the length of the cable.

Advantages

- These are relatively expensive and simple to manufacture
- It is easier to couple light into and out of multimode step-index fiber as they have a relatively large source-to-fiber aperture.

Disadvantages

- As light rays travel in different paths, large difference in propagation times results. So, the rays travelling down have a tendency to spread out. Consequently the pulse of light propagating down is more distorted than other types of fibers.
- Less bandwidths and lower rate of information transfer rates when compared to other types.

Multimode Graded-Index Optical Fiber: These fibers are characterized by a central core with a nonuniform refractive index. Cables density is maximum at centre and decreases gradually towards the edge. Light ray is propagated through refraction. As the light propagates across the core toward the center it intersects a less dense to more dense medium. Consequently, light rays constantly being refracted resulting in continuous bending of light rays. The light rays take approximately the same amount of time to travel the length of the fiber. This cable is mostly used for long distance communication.

Losses in Optical Fiber Cables

Power loss in optical fiber cables is often called attenuation and results in reduction of power of light wave as it travels down the cable. Generally, total power loss is expressed as:

$\mathbf{A_{(dB)} = 10 \log (P_{out} / P_{in})}$ where $A_{(dB)}$ is total reduction in power level, attenuation and P_{out} is cable output power and P_{in} is cable input power. Multimode fibers tend to have more attenuation than single-mode cables because of increased scattering of light wave.

Transmission losses in optical fibers result in reduction in light power, thus reducing the system bandwidth, information transmission rate, efficiency, and overall system capacity. The predominant losses are:

Absorption Losses: It is analogous to power dissipation in copper cables as impurities in the fiber absorb the light and convert it to heat. Three main factors contribute to absorption losses.

- Ultraviolet absorption:- Caused by valence electrons in the silica material from which fibers are manufactured.
- Infrared absorption: - Result of photons of light that are absorbed by the atoms of the glass core molecules.
- Ion resonance absorption: - Caused by OH- ion in the material. Iron, copper and chromium molecules also cause ion absorption

Material or Rayleigh Scattering Losses: Rayleigh scattering of light is due to small localized changes in the refractive index of the core and cladding material. Two main causes for this:

- The first is due to slight fluctuation in mixing of ingredients. The random changes because of this are impossible to eliminate completely.
- The other cause is slight change in density as the silica cools and solidifies. When light ray strikes such zones, it gets scattered in all directions. The amount of scatter

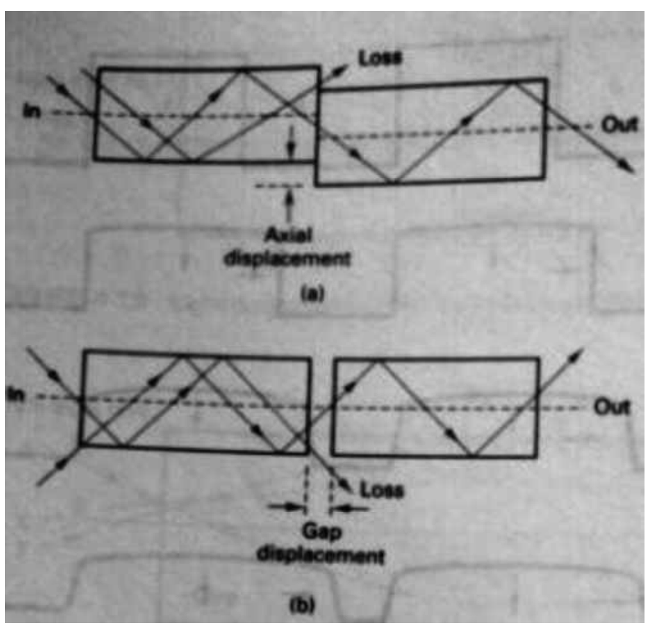
depends on the size of the discontinuity compared with the wavelength of the light. So the shortest wavelength suffers most scattering.

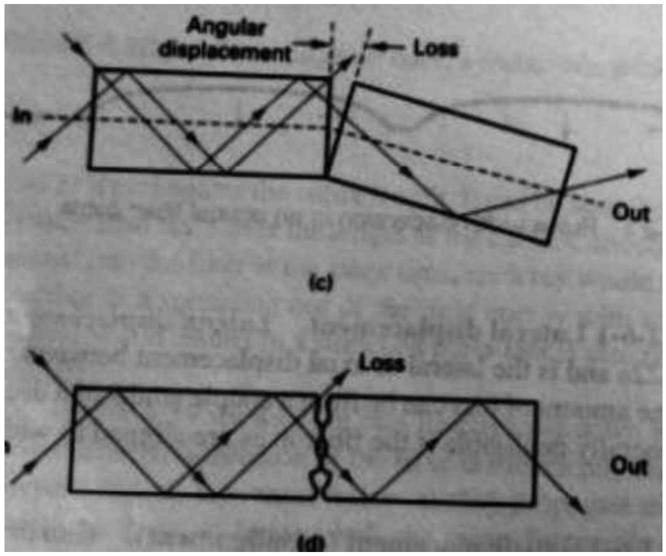
Chromatic Distortion or Wavelength Dispersion: Light rays that are simultaneously emitted from an LED and propagated down an optical fiber do not arrive at the far end of the fiber at the same time, which results in an impairment called chromatic distortion. It occurs in only in fibers with a single mode of transmission and can be eliminated using monochromatic light sources like injection laser diode (ILD).

Radiation Losses: These are caused predominantly by small bends wand kinks in the fiber. The two types of bends are: microbends and constant-radius bends. Microbending occurs as result of differences in the thermal contraction rates between core and cladding material and results in a material bend along the axis of the fiber and represents a discontinuity where Rayleigh scattering occurs. Constant-radius bends are caused by excessive pressure and tension and generally occur when fibers are bent during installation.

Modal dispersion: Modal dispersion (called pulse spreading) is caused by the difference in the propagation times of light rays that take different paths down a fiber and occurs only in multimode fibers. It can be reduced considerably by using graded index fibers and almost entirely eliminated using single-mode step-index fibers. If three rays of light are emitted into the fiber at the same time, each ray would reach the far end at a different time resulting in a spreading out of light energy with respect to time. This is called modal dispersion.

Coupling Losses: -These losses are caused by imperfect physical connections. These occur at three types of junctions: light source-to-fiber connections, fiber-to-fiber connections, and fiber-to-photodetector connections. They are caused by one of the following alignment problems:





- Lateral displacement: It is the lateral or axial displacement between two pieces of adjoining fiber cables.
- Gap displacement (misalignment): When splices are made in optical fibers, the fibers should actually touch. The farther apart the fibers are, the greater the loss of light.
- Angular displacement: It is sometimes called angular displacement and if it is less than 2 degrees, the loss will typically be less than 0.5 dB.
- Imperfect surface finish: The ends of two adjoining fibers should be highly polished and fit together squarely. If the fiber ends are less than 3 degrees off from perpendicular, the losses will typically be less than 0.5 dB.

Light Sources

Light sources are used in fiber optic communication to generate light pulses at wavelengths efficiently propagated by the optical fiber. They also should produce sufficient power to allow the light to propagate through the fiber without causing distortion in the cable or receiver. Two types of practical light sources used to generate light for optical fiber communications systems: light-emitting diodes (LED's) and injection laser diodes (ILD's).

Light Emitting Diodes: A LED is a p-n junction diode, usually made from a semiconductor material such as aluminum-gallium-arsenide (AlGaAs) or gallium-arsenide-phosphide (GaAsP). LED's emit light by spontaneous emission-light is emitted as a result of the recombination of electrons and holes. LEDs can provide light output when forward biased. The LED has a low output power, slower switching speed and greater spectral width, hence more dispersion. These deficiencies make it not useful for high speed and long distance communication. The output of LED is non-coherent and coupling efficiency is very low.

Injection Laser Diode: ILD's are similar to LED's and they act similarly below a certain threshold current. Above the threshold current, and ILD oscillates and lasing occurs. As current passes through a forward biased p-n junction diode, light is emitted by spontaneous emission at a frequency determined by the energy gap of the semiconductor material. The radiant output power of ILD is more directive than LED. After lasing occurs, the optical power increases dramatically, with small increases in drive current. Advantages

1. ILD's emit coherent (orderly) light compared to incoherent (disorderly) light emitted by LED. So ILD have a more direct radian pattern, making it easier to couple light emitted by the ILD into an optical fiber cable. Coupling losses are reduced and also small fibers can be used.
2. The radiant output power of ILD is greater than that for an LED. Typically the output power for an ILD is 5 mW and only 0.5mW for LED. This allows ILD's to provide a higher drive power and can be used for operation over longer distances.
3. ILD's can be used at higher bit rates than LED's
4. ILD's generate monochromatic light, which reduces chromatic or wavelength dispersion

Disadvantages

1. ILD's are typically 10 times more expensive than LED's
2. As ILD's operate at higher powers, they have a short lifetime
3. ILD's are more temperature dependent than LED's

Light Detectors

Two devices are commonly used to detect light energy in optical fiber communications receivers: PIN (p-type-intrinsic –n-type) diodes and APD (avalanche photodiodes). PIN diodes are the most common device used and operate just the opposite of an LED. APD's are more sensitive than pin diodes and require less additional amplification. The disadvantages of APD's are relatively long transmit times and additional internally generated noise due to avalanche multiplication factor.

Characteristics of light detectors

1. Responsivity: A measure of the conversion efficiency of a photodetector. It is the ratio of the output current of a photodiode to the input optical power and has the unit of amperes/watt.
2. Dark Current: The leakage current that flows through a photodiode with no light input.
3. Transit time: The time it takes a light-induced carrier to travel across the depletion region of a semiconductor. Determines the maximum bit-rate possible
4. Spectral Response: The range of wavelength values that a given photodiode will respond to.
5. Light Sensitivity: The minimum optical power a light detector can receive and still produce a usable electrical output signal.

Lasers

Laser stands for light amplification stimulated by the emission of radiation. It deals with the concentration of light into a very small, powerful beam. There are four types of lasers:

1. *Gas lasers*: Gas lasers use a mixture of helium and neon enclosed in a glass tube. A flow of coherent light waves is emitted when an electric current is discharged into the gas. The continuous light-wave output in monochromatic (one color)

2. *Liquid lasers*: They use organic dyes enclosed in a glass tube for an active medium. A powerful pulse of light excites the organic dye.
3. *Solid lasers*: They use a solid, cylindrical crystal such as ruby, for the active medium. Both ends of ruby are polished and parallel and the ruby is excited by a tungsten lamp tied to an ac power supply. It produces a continuous wave.
4. *Semiconductor lasers*: They are made from semiconductor p-n junctions and are commonly called injection laser diodes. The excitation mechanism is a dc power supply that controls the amount of current to the active medium. The output light is easily modulated making it very useful in many electronic communication systems.

Laser Characteristics

All types of lasers use

1. an active material to convert energy into laser light.
2. a pumping source to provide power or energy
3. optics to direct the beam through the active material to be amplified
4. optics to direct the beam into a narrow powerful cone of divergence
5. a feedback mechanism to provide continuous operation
6. an output coupler to transmit power out of the laser

UNIT-IV

DIGITAL TRANSMISSION

Pulse Modulation, Pulse code Modulation, Dynamic Range, Signal Voltage –to-Quantization Noise Voltage Ration, Linear Versus Nonlinear PCM Codes, Companding, PCM Line Speed, Delta Modulation PCM and Differential PCM

Digital transmission is the transmittal of digital signals between two or more points in a communications system. The signals can be binary or any other form of discrete-level digital pulses. With digital transmission systems, a physical facility, such as pair of wires, a coaxial cable or an optical fiber cable is required to interconnect the various points within the system.

Digital transmission has several advantages over analog transmission:

- Important advantage is the noise immunity as digital signals are inherently less susceptible than analog signals to interference caused by noise.
- Digital signals are better suited than analog signals for processing and combining using a technique called multiplexing.
- Digital transmission systems are more resistant to analog systems to additive noise because they use signal regeneration rather than signal amplification.
- Digital signals are simpler to measure and evaluate than analog signals.

Disadvantages:

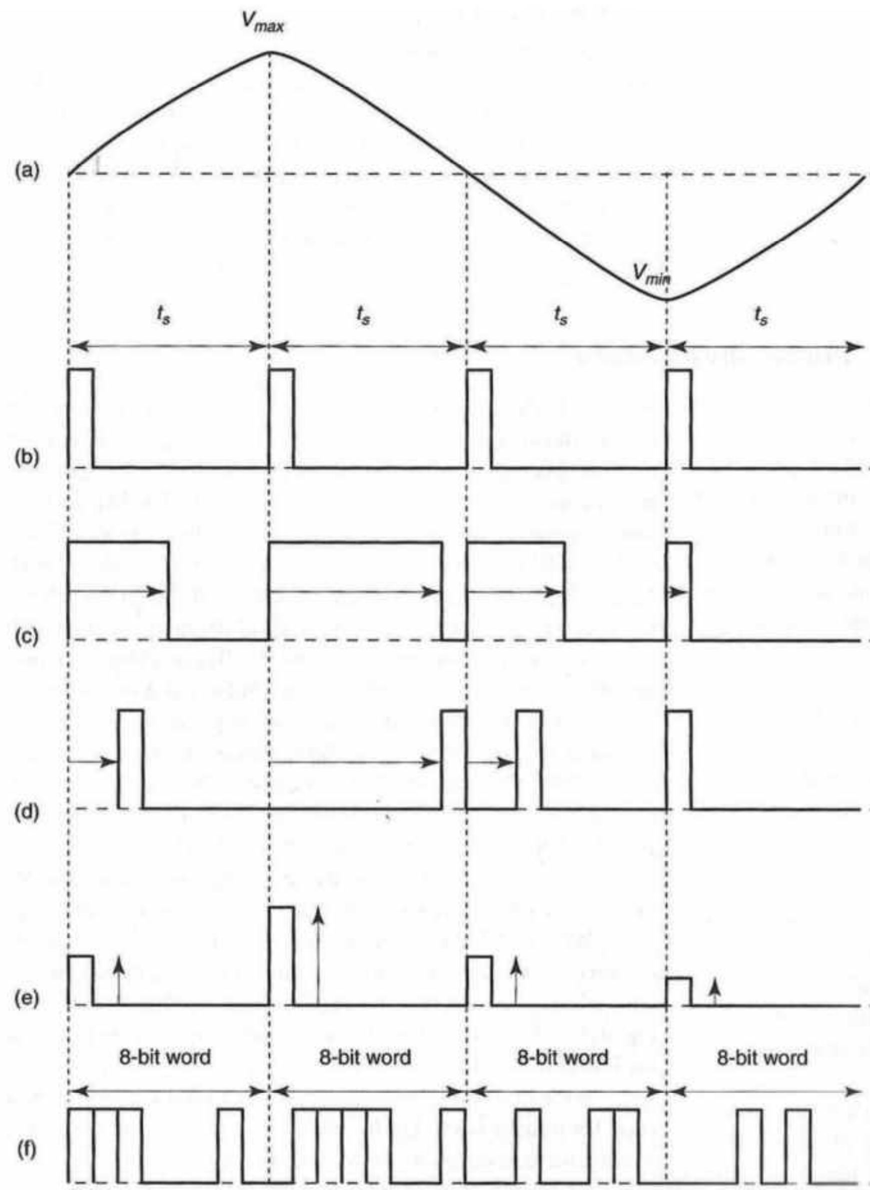
- Transmitting digitally encoded analog signals requires more bandwidth than simply transmitting the original analog signal, which makes it expensive.
- Also conversion of analog signals into digital pulses prior and after transmission requires additional encoding and decoding circuitry.
- Precise time synchronization between the clocks in transmitters and receivers is required in digital transmission
- Digital transmission systems are incompatible with older analog transmission systems.

Pulse Modulation

The process of sampling analog information signals and then converting those samples into discrete pulses and then transporting the pulses from a source to a destination over a physical transmission medium is called ***Pulse Modulation***. Pulse modulation involves communication using a train of recurring pulses. There are four different types of pulse modulation techniques.

Pulse Width Modulation (PWM): The width of a constant amplitude pulse is varied proportional to the amplitude of the analog signal at the time the signal is sampled. It is sometimes called as pulse duration modulation (PDM) or pulse length modulation (PLM). It is very popular in digital circuits because of its easy generation and its applications include voltage regulators and class-D audio amplifiers.

Pulse Position Modulation (PPM): The position of a constant-width pulse within a prescribed time slot is varied according to the amplitude of the sample of the analog signal. It is commonly used in communications over optic fibers as multipath fading is minimal. It is also used in communications for RC aircraft/cars etc as demodulation is easy allowing a low-cost receiver.



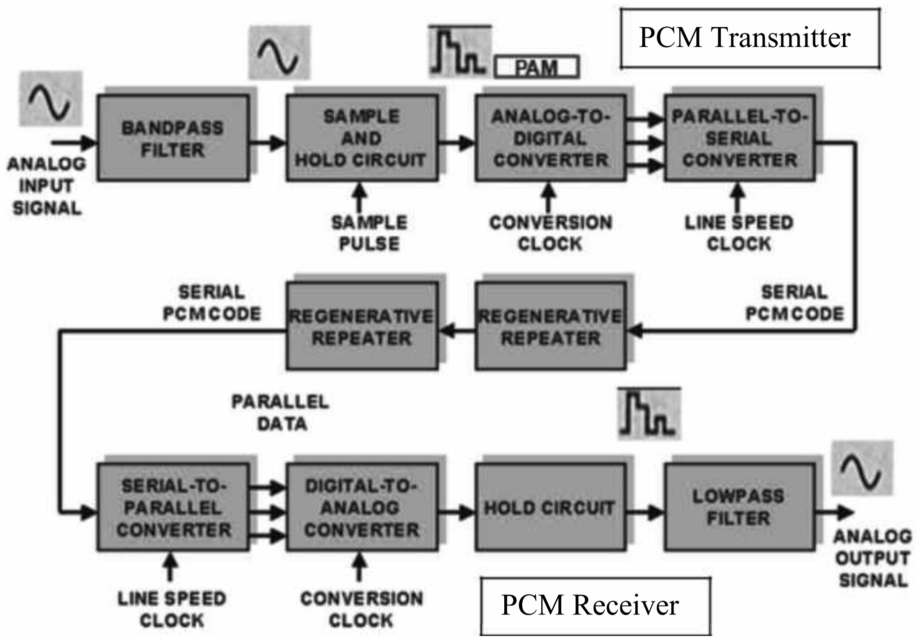
Pulse modulation: (a) analog signal; (b) sample pulse; (c) PWM; (d) PPM; (e) PAM; (f) PCM

Pulse Amplitude Modulation (PAM): The amplitude of a constant width, constant-width, and constant-position is varied according to the amplitude of the sample of the analog signal. It resembles the original analog signal more than the wave forms for PWM or PPM. Telephone modems faster than 300 bits/sec and Ethernet use PAM.

Pulse Code Modulation (PCM): The analog signal is sampled and then converted to a serial n-bit binary code for transmission. Each code has the same number of bits and requires the same length of time for transmission. Applications include digital audio in computers and CDs.

Pulse Code Modulation

PCM invented by Alex H. Reeves in 1937 is the preferred method of communications within the public switched telephone network because with PCM, it is easy to combine digitized voice and digital data into a single, high-speed digital signal and propagate it over either metallic or optical fiber cables. With PCM, the pulses are of fixed length and fixed amplitude. PCM is a binary system where a pulse of lack of pulse within a prescribed time slot represents either logic 1 or logic 0 conditions.



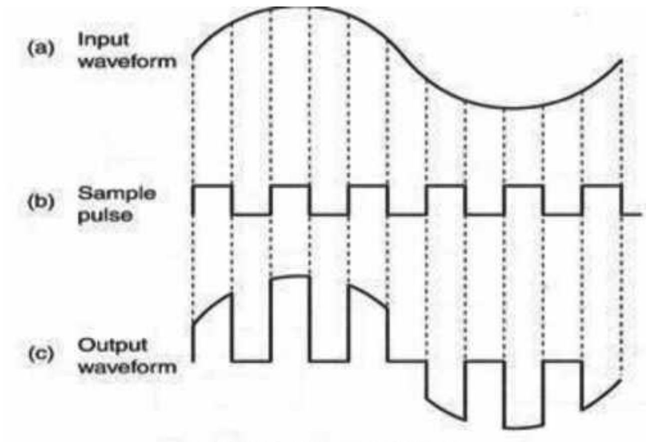
The above figure shows a simplified block diagram of a single-channel, simplex PCM system. The *bandpass filter* limits the frequency of the analog input signal to the standard voice-band frequency range of 300 Hz to 3000 Hz. The *sample and hold circuit* periodically samples the analog input signal and converts those samples to a multiple PAM signal. The *analog-to-digital converter (ADC)* converts the PAM samples to parallel PCM codes, which are converted to serial binary data in the *parallel-to-serial converter* and then outputted into the *transmission line* as serial digital pulses. *Repeaters* are placed at prescribed distances to regenerate the digital pulses.

At receiver side, the *serial-to-parallel converter* converts serial pulses received from the transmission line to parallel PCM codes. The *digital-to-analog converter (DAC)* converts the parallel PCM codes to multilevel PAM signals. The *hold circuit* is basically a low pass filter that converts the PAM signals back to its original analog form. An integrated circuit that performs the PCM encoding and decoding functions is called a codec (coder/decoder).

PCM Sampling

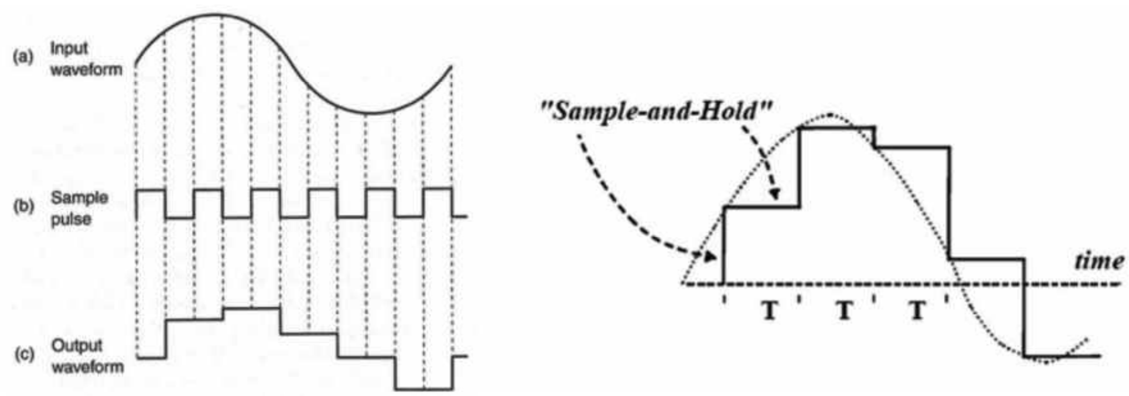
Sampling circuit in a PCM transmitter periodically samples the continually changing analog input voltage and converts those samples to a series of constant amplitude pulses, which can be more easily converted to binary PCM code. Two basic techniques to perform sampling exist.

Natural Sampling: In natural sampling the pulse amplitude takes the shape of the analogue waveform for the period of the sampling pulse. The frequency spectrum of the sampled output is different from that of an ideal sample.



Natural sampling

Flat-top sampling: It is accomplished in a sample-and-hold circuit. Its purpose is to periodically sample the continually changing analog input voltage and covert those samples to a series of constant voltage PAM voltage levels. With flat-top sampling, the input voltage is sampled with a narrow pulse and then relatively held constant until next sample is taken.



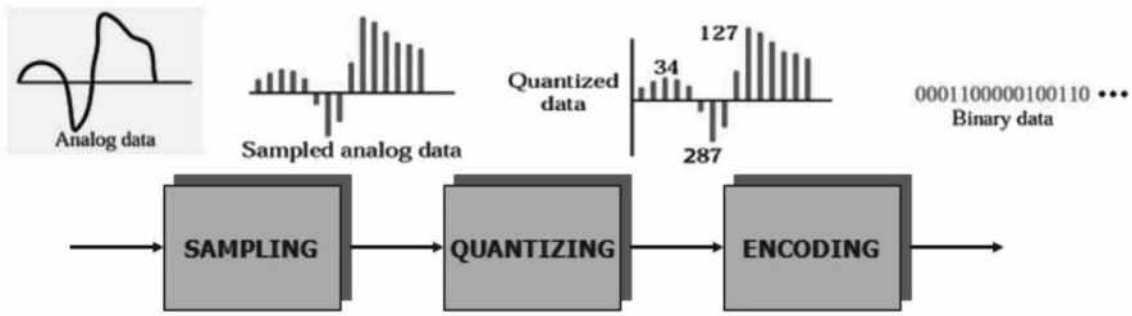
Flat-top sampling

Flat-top sampling introduces less aperture distortion than natural sampling and can operate with a slower analog-to-digital converter.

Sampling Rate

Nyquist sampling theorem states that for a sample to be reproduced accurately, minimum sampling rate, f_s must be twice the higher input frequency, f_a. Mathematically, the minimum Nyquist sampling rate, fs is $\boldsymbol{f_s \geq 2 f_a}$, where f_s is minimum Nyquist sample rate in hertz and f_a is maximum analog input frequency in hertz.

If $\boldsymbol{f_s}$ is less than twice $\boldsymbol{f_a}$, i.e. $\boldsymbol{f_s < 2 f_a}$, **aliasing or *foldover* distortion** occurs. This can be overcome by using anti-aliasing filter before sampling to suppress the component before sampling.



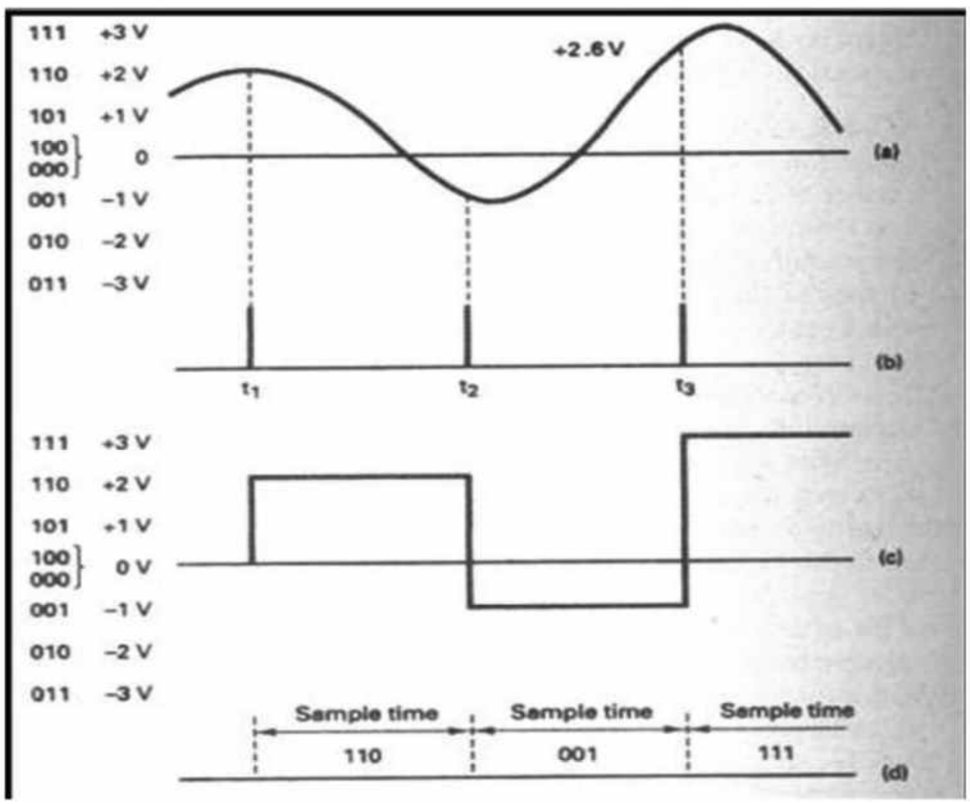
Quantization and the Folded Binary Code

Quantization is the process of converting a continuous range of values into a finite range of discreet values. This is a function of analog-to-digital converters, which create a series of digital values to represent the original analog signal. Quantization is required to convert the analog signal to a PCM code with a limited number of combinations. Taking an example, a sine wave with peak amplitude of 5v, varying between +5V and -5V passing through every amplitude between them. A PCM code could have only eight bits, which equates to only 2^8 or 256 combinations and to be converted, the sine wave values have to be rounded off.

Sign	Magnitude	Decimal Value	Quantization Range
1	11	+3	+2.5 - +3.5
1	10	+2	+1.5 - +2.5
1	01	+1	+0.5 - +1.5
1	00	+0	+0 - +0.5
0	00	-0	-0 - -0.5
0	01	-1	-0.5 - -1.5
0	10	-2	-1.5 - -2.5
0	11	-3	-2.5 - -3.5

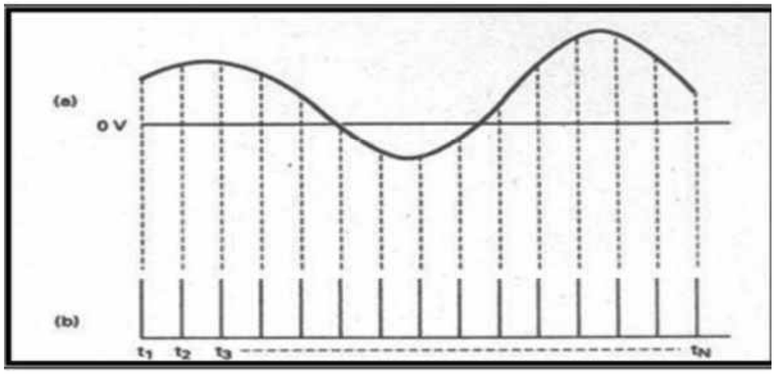
The above table shows the three bit PCM code, which is a three-bit sign magnitude code with eight possible combinations (four +ve and four –ve). The left most bit is the sign bit (1 = + and 0 = -), and the remaining two right most bits represent magnitude. This type of code is called *folded binary code* because the codes on the bottom half of the table are an exact mirror image of the codes on the top except for the sign bit. The magnitude difference between adjacent steps is called the *quantization interval or quantum* (1V for above table). For the above code, the maximum signal magnitude that can be encoded is +3V (111) or -3V (011) and the minimum is +1V (101) or -1V (001). If the magnitude of the sample exceeds the highest quantization interval, *overload distortion (peak limiting)* occurs.

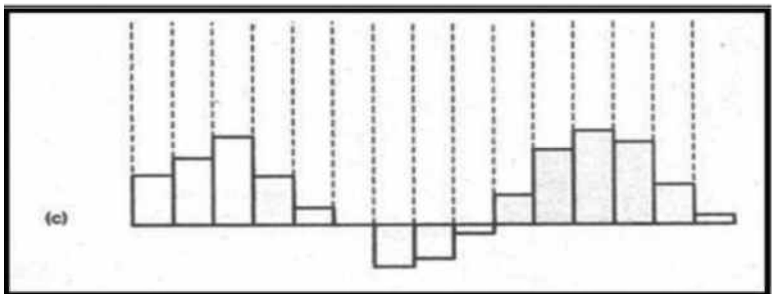
Assigning PCM codes to absolute magnitudes is called quantizing. The magnitude of a quantum is also called the *resolution*. It is equal to the voltage of the least significant bit (V_{lsb}) of the PCM code. The smaller the magnitude of a quantum, the better the resolution and the more accurately the quantized signal will resemble the original analog signal.



(a) Analog input signal; (b) sample pulse; (c) PAM; (d) PCM code

Each sample voltage is rounded off to the closest available quantization level and then converted to its corresponding PCM code. The PAM signal in the transmitter is essentially the same PAM signal produced in the receiver. So, any round-off errors in the transmitted signal are reproduced when the code is converted back to the analog by the DAC in the receiver. This error is called the *quantization error (Qe),* which is also quantization noise (*Qn*). The quantized signal shown above roughly resembles the original input signal as with three-bit PCM code, poor resolution results and only three samples are taken from analog signal.





As shown above, the quality of PAM signal can be improved by using a PCM code with more bits, reducing the magnitude of quantum and improving the resolution. Also, the sampling the analog signals at a faster rate increases the quality and the PAM signal resembles the analog signal closely. Quantization error is given by

$$Q_e = \frac{V_{MIN}}{2} = \frac{\text{Resolution}}{2}$$

Dynamic Range

It is the ratio of the largest possible magnitude to the smallest possible magnitude (other than 0V) that can be decoded by the digital-to-analog converter in the receiver. Mathematically,

$DR = \frac{V_{max}}{V_{min}}$, where, DR is dynamic range and V_{min} is the quantum value (resolution) and V_{max} is the maximum voltage magnitude that can be discerned by the receivers DACs. Dynamic range can be expressed in decibels as

$$DR_{(dB)} = 20 \log \frac{V_{max}}{V_{min}}$$

Or $DR_{(dB)} = 20 \log (2^n - 1)$, where n is the number of PCM bits.

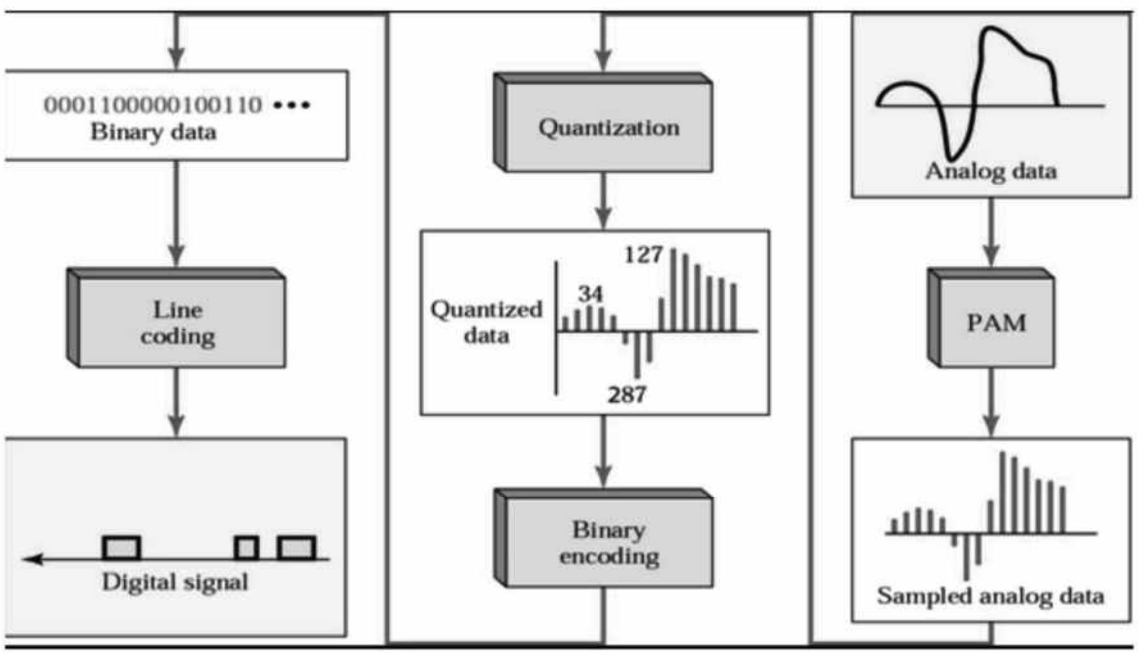
Signal Voltage –To-Quantization Noise Voltage Ratio

The maximum quantization noise is half the resolution. Therefore, the worst possible signal voltage-to-quantization noise voltage ratio (SQR) occurs when the input signal is at its minimum amplitude. Mathematically, the worst-case voltage SQR is 2.

For linear PCM codes, the signal power-to-quantizing noise power ratio is determined by the formula:

$$SQR_{(dB)} = 10 \log \frac{v^2/R}{(q^2/12)/R}$$

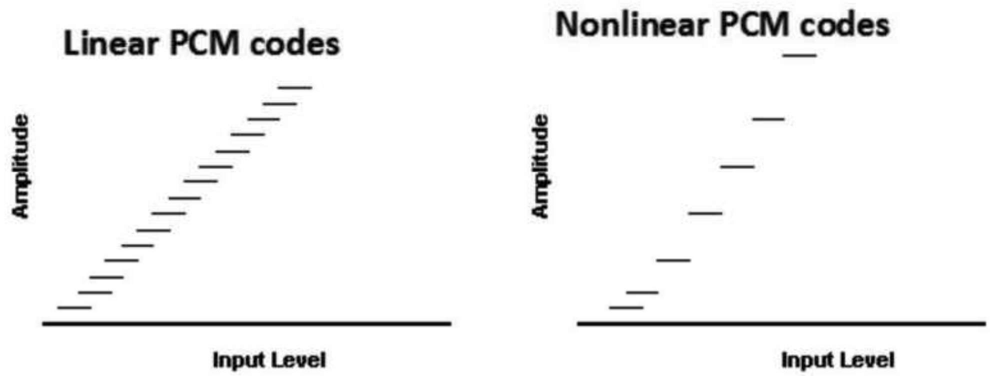
Where, R is resistance in ohms, v is rms signal voltage in volts, q is quantization interval in volts, v^2/R is average signal power in watts and $(q^2/12)/R$ is average quantization noise power in watts.



From analog signal to PCM Digital Codes

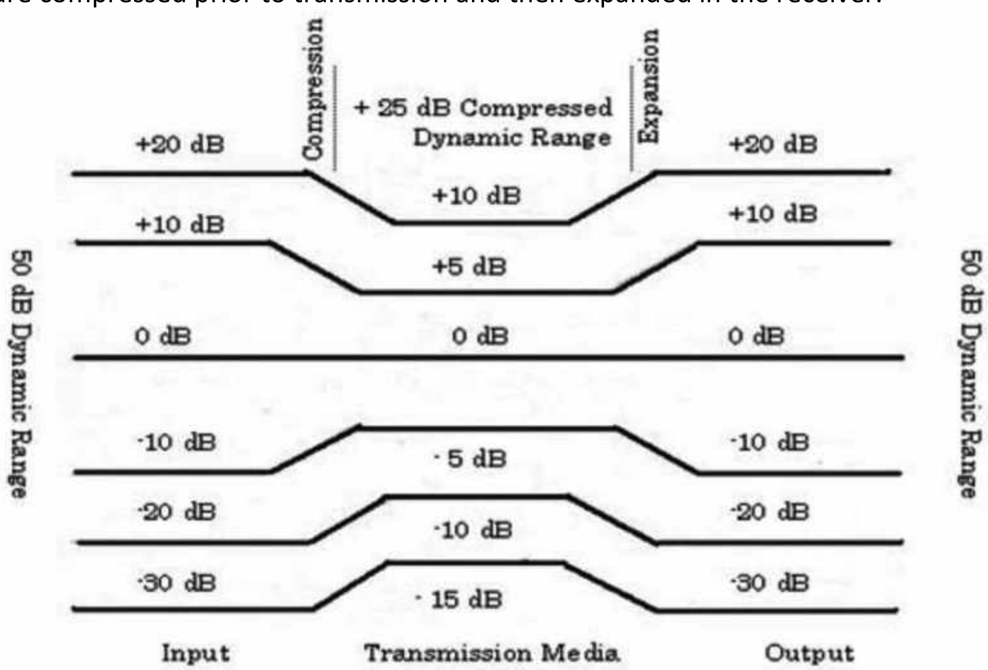
Linear versus Nonlinear PCM codes

- Linear codes – magnitude change between any two successive steps in uniform
  - Resolution/accuracy is the same for lower and higher amplitude signal
  - SQR for low amplitude signal is less than the SQR for higher amplitude signal
- Nonlinear – step size increases with the amplitude of the input signal
  - More codes at the bottom
  - Distance between successive codes is greater for higher amplitude signals
  - Vmax/Vmin is increased



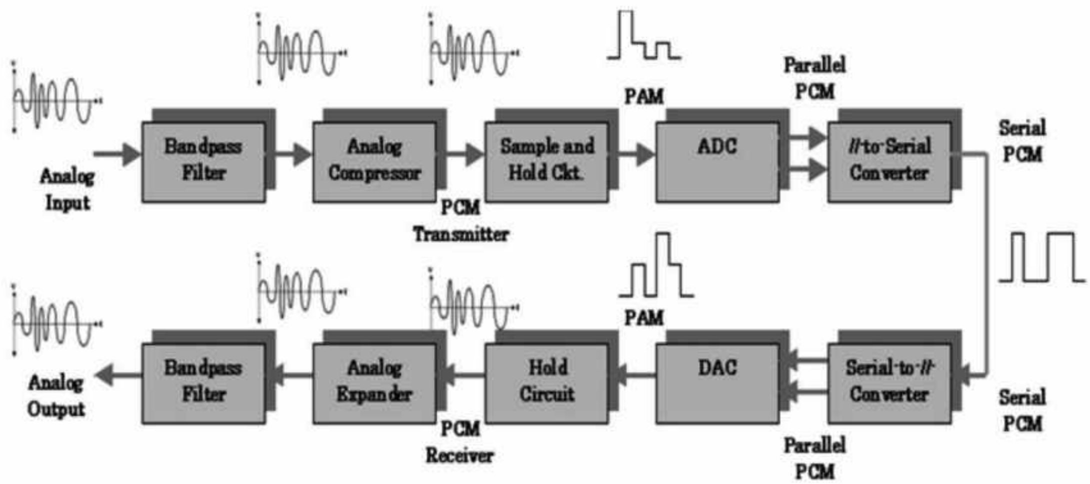
Companding

Companding is the process of compressing and expanding and is a means of increasing the dynamic range of a communications system. Higher-amplitude analog signals are compressed prior to transmission and then expanded in the receiver.



An analog input signal with a dynamic range of 50dB is compressed to 25dB prior to transmission and then expanded back at the receiver. With PCM, companding may be accomplished using analog or digital techniques.

Analog Companding

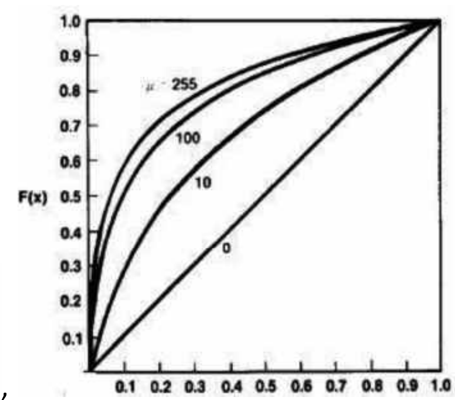


PCM system with analog companding

In the transmitter, the dynamic range of the analog signal is compressed, sampled, and then converted to a linear PCM code. In the receiver, the PCM code is converted to a PAM signal, filtered and then expanded back to its original dynamic range. Two methods of analog companding exist (also called log-PCM codes).

μ **– Law companding:** u-law is a companding scheme used in telephone network to get more dynamics to the 8 bit samples that is available with linear coding. Compression characteristics for **μ – Law** is

$$V_{out} = \frac{V_{max} \ln(1 + \mu\{V_{in} / V_{max}\})}{\ln(1 + \mu)}$$,



μ *– Law Compression characteristics*

Where,

V_{max} = maximum uncompressed analog input amplitude (volts)

V_{in} = amplitude of the input signal at particular instant of time (volts)

μ = parameter used to define the amount of compression

(unitless) V_{out} = compressed output amplitude (volts)

The graph shows the compression curves for several values of μ. The higher the μ, the more compression and also for μ = 0, the graph is linear.

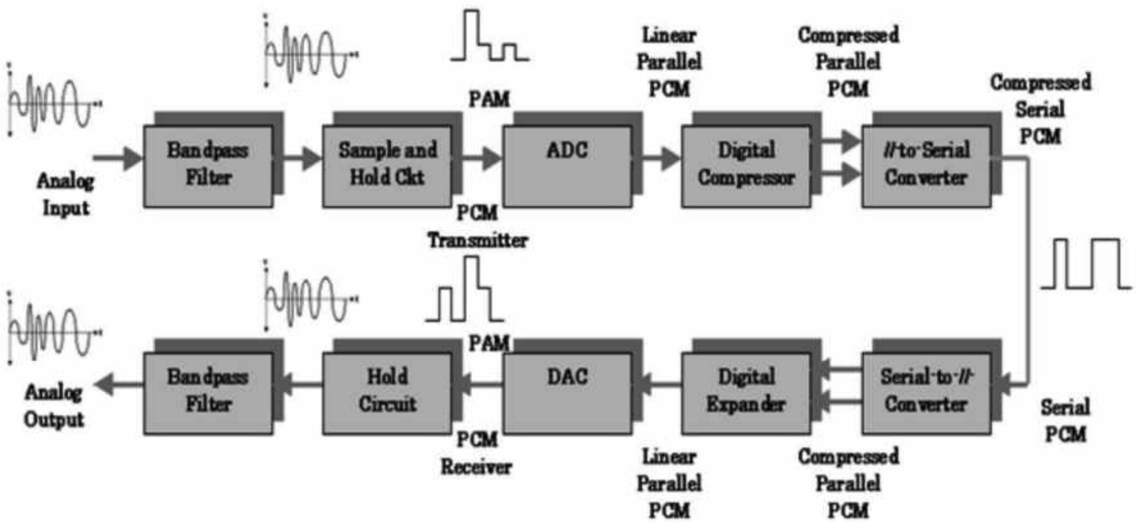
A – Law: In Europe, the ITU-T has established A-law companding to be used to approximate true logarithmic companding. Compression characteristic for A- law companding is

$$V_{out} = V_{max} \frac{AV_{in} / V_{max}}{1 + \ln A} \qquad 0 \le \frac{V_{in}}{V_{max}} \le \frac{1}{A}$$

$$V_{out} = V_{max} \frac{1 + \ln(AV_{in} / V_{max})}{1 + \ln A} \qquad \frac{1}{A} \le \frac{V_{in}}{V_{max}} \le 1$$

Digital Companding

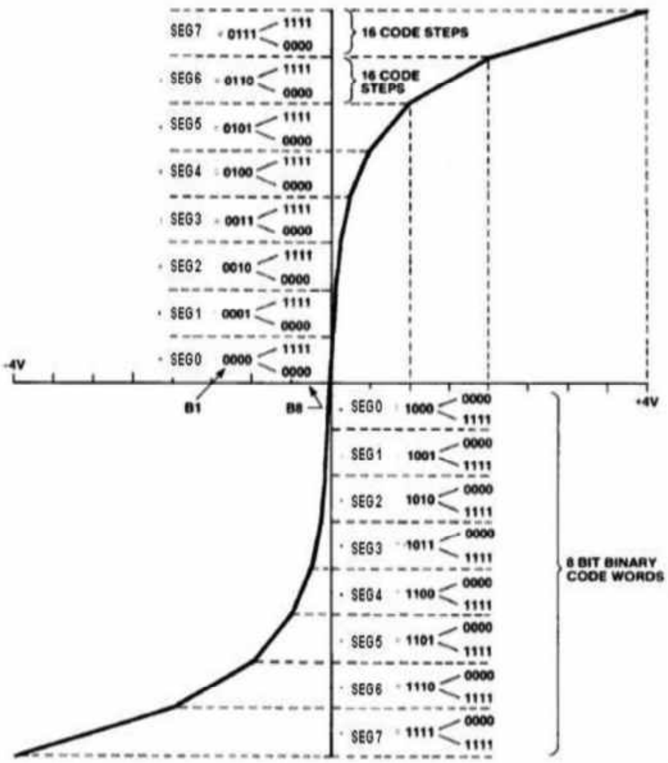
Digital companding involves compression in the transmitter after the input sample has been converted to a linear PCM code and then expansion in the receiver prior to PCM decoding.

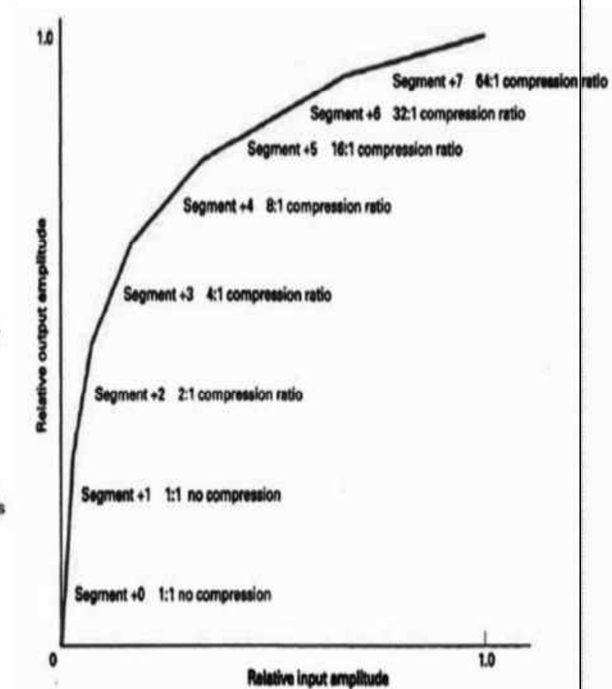


Recent digitally compressed PCM systems use a 12-bit linear PCM code and an eight-bit compressed PCM code. The compression and expansion curves closely resemble analog µ-law curves with a µ = 255.

The following figure shows 12- to eight-bit digital compression curve for positive values only. For negative values, it's identical but just inverse. Though 16 segments are present, this scheme is called 13-segment compression because the curve for segments +0, +1, -0 and -1 is a straight line.

The digital companding algorithm for a 12-bit linear to eight-bit compressed code is quite simple and the compressed code consists of a sign bit, a three-bit segment identifier, and a five-bit magnitude code which specifies the quantization interval within the specified segment.





13 Segment Scale **μ- 255 compression characteristics**

SIGN BIT	3-BIT SEGMENT IDENTIFIER	4-BIT QUANTIZATION INTERVAL			
1 = +	000 TO 111	A	B	C	D
0 = -		0000 TO 1111			

Eight bit μ- 255 compressed code format

SEGMENT	12-BIT LINEAR CODE	8-BIT COMPRESSED CODE
0	s0000000ABCD	s000ABCD
1	s0000001ABCD	s001ABCD
2	s000001ABCDX	s010ABCD
3	s00001ABCDXX	s011ABCD
4	s0001ABCDXXX	s100ABCD
5	s001ABCDXXXX	s101ABCD
6	s01ABCDXXXXX	s110ABCD
7	s1ABCDXXXXXX	s111ABCD

8-BIT COMPRESSED CODE	12-BIT RECOVERED CODE	SEGMENT
s000ABCD	s0000000ABCD	0
s001ABCD	s0000001ABCD	1
s010ABCD	s000001ABCD1	2
s011ABCD	s00001ABCD10	3
s100ABCD	s0001ABCD100	4
s101ABCD	s001ABCD1000	5
s110ABCD	s01ABCD10000	6
s111ABCD	s1ABCD100000	7

μ- 255 encoding and decoding table

In the encoding table shown above, the bit positions designated with an X are truncated during compression and thereafter lost. Bits designated by A, B,C, D along with the sign bit are transmitted as is. The analog signal is sampled and converted to a linear 12-bit sign-magnitude code. The sign bit is transferred directly to an eight-bit compressed code. The segment number in the eight-bit code is determined by counting the number of leading 0's in the 11-bit magnitude portion of the linear code beginning with the most-significant bit and then subtracting the number of leading 0s from 7, which is the segment number. The segment number is converted to a three-bit binary number and inserted into the eight-bit compressed code as the segment identifier. The four magnitude bits (A, B, C, D) represent the quantization interval and are submitted into the least-significant four bits of the eight-bit compressed code.

Segment	12-Bit linear code		12-Bit expanded code	Subsegment
7	s11111111111 ... s11111000000	64 : 1	s11111100000	15
7	s11110111111 ... s11110000000	64 : 1	s11110100000	14
7	s11101111111 ... s11101000000	64 : 1	s11101100000	13
7	s11100111111 ... s11100000000	64 : 1	s11100100000	12
7	s11011111111 ... s11011000000	64 : 1	s11011100000	11
7	s11010111111 ... s11010000000	64 : 1	s11010100000	10
7	s11001111111 ... s11001000000	64 : 1	s11001100000	9
7	s11000111111 ... s11000000000	64 : 1	s11000100000	8
7	s10111111111 ... s10111000000	64 : 1	s10111100000	7
7	s10110111111 ... s10110000000	64 : 1	s10110100000	6
7	s10101111111 ... s10101000000	64 : 1	s10101100000	5
7	s10100111111 ... s10100000000	64 : 1	s10100100000	4
7	s10011111111 ... s10011000000	64 : 1	s10011100000	3
7	s10010111111 ... s10010000000	64 : 1	s10010100000	2
7	s10001111111 ... s10001000000	64 : 1	s10001100000	1
7	s10000111111 ... s10000000000	64 : 1	s10000100000	0
	s1ABCD------			

Segments 2 through 7 are subdivided into smaller subsegments. Each segment consists of 16 subsegments corresponding to the 16 conditions possible for bits A, B,C and D (0000 – 1111). In segment 2, there are two codes per subsegment and in segment 3, there are four. The number of codes per subsegment doubles with each subsequent segment. So, in segment 7, each subsegment has 64 codes. In the decoder, the most significant of the truncated bits is reinserted as logic 1. The remaining truncated bits are reinserted as 0s. This minimizes the magnitude of error introduced by compression and expansion process.

Digital Compression Error

The magnitude of the compression error is not the same for all samples. However, the maximum percentage is the same in each segment (other than segments 0 and 1, where there is no compression error), which is calculated using:

$$\% \text{ error} = \frac{\text{12-bit encoded voltage - 12-bit decoded voltage}}{\text{12-bit decoded voltage}} \text{ X } 100$$

Every function performed by a PCM encoder and decoder is now accomplished with a single integrated-circuit chip called *codec.* Some of the most recent developed codecs are called combo chips, as they include an antialiasing (band-pass) filter, a sample-and-hold circuit, and an ADC in transaction and a DAC, a hold circuit, and a band pass filter in the receive section.

PCM Line Speed

Line speed is the data rate at which serial PCM bits are clocked out of the PCM encoder onto the transmission line. Line speed is dependent on the sample rate and the number of bits in the compressed PCM code. Mathematically, it is:

$$\text{line speed} = \frac{\text{samples}}{\text{second}} \times \frac{\text{bits}}{\text{sample}}$$

, where

Line speed is the transmission rate in bits per second, samples/second is sample rate (f_s) and bits/sample is number of bits in the compressed PCM code.

Delta Modulation PCM and Differential PCM

Delta modulation PCM uses a single-bit PCM code to achieve digital transmission of analog signals. Here, only a single bit is transmitted, which simply indicates whether the present sample is larger or smaller in magnitude than the previous sample. If the current sample is larger in magnitude than a previous sample, a logic 1 is transmitted and if its smaller, logic 0 is transmitted.

Differential pulse code modulation (DPCM) takes advantage of the sample-to-sample redundancies in typical speech waveforms. With DPCM, a binary code proportional to the difference in the amplitude of two successive samples is transmitted rather than a binary code of an actual sample. As the range of sample differences is less than the range of individual sample amplitudes, fewer bits are required for DPCM than for conventional PCM.

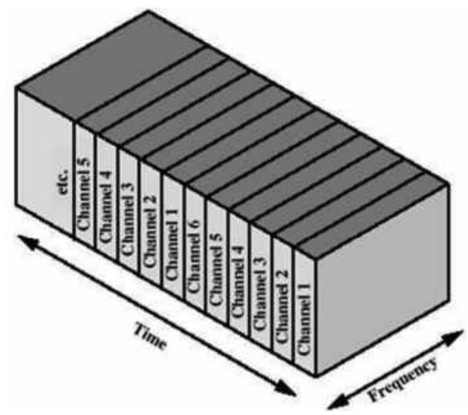
MULTIPLEXING AND T CARRIERS

Time- Division Multiplexing, T1 Digital Carrier System, North American Digital Multiplexing Hierarchy, Digital Line Encoding, T Carrier systems, European Time- Division Multiplexing, Statistical Time – Division Multiplexing, Frame Synchronization, Frequency-Division Multiplexing, Wavelength- Division Multiplexing, Synchronous Optical Network

Multiplexing is the transmission of information from more than one source to more than one destination over the same transmission medium. Multiplexing is accomplished in various domains such as space, time, phase, frequency and wavelength.

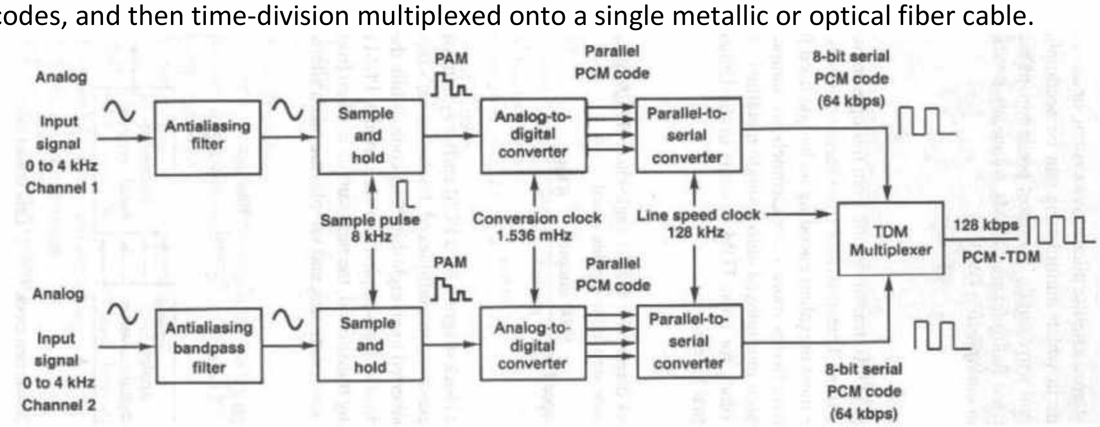
Time Division Multiplexing

With TDM system, transmission from multiple sources occurs on the same transmission medium but not at the same time. Transmission from various sources is interleaved in time domain. The two basic forms of TDM are: Synchronous TDM (STDM) and Asynchronous (or) Statistical TDM (STATDM)

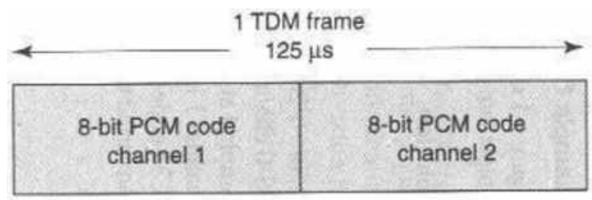


In synchronous TDM, time slot 'x' is assigned to user m alone and cannot be used by any other user or other device. T-1 and ISDN telephone lines are common examples of synchronous time division multiplexing. Asynchronous TDM networks assign time slots only when they are to be used and delete them when they are idle. STATDM is used in high density and high traffic applications.

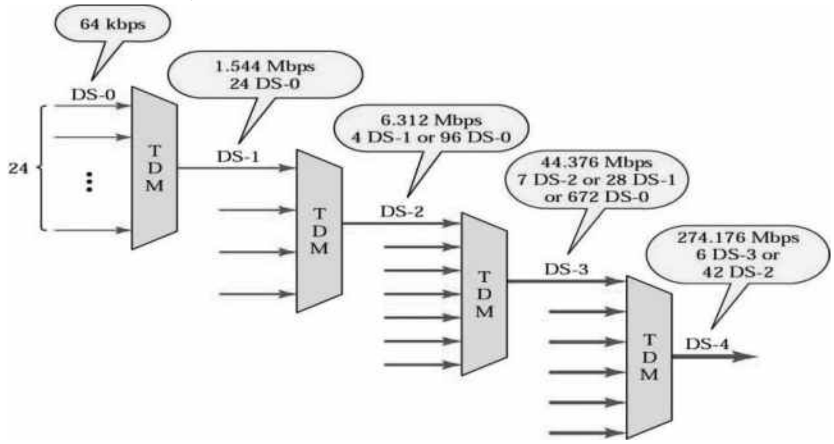
With PCM-TDM system, two or more voice channels are sampled, converted to PCM codes, and then time-division multiplexed onto a single metallic or optical fiber cable.



The above figure shows a block diagram for a PCM carrier system comprised of two DS-0 channels that have been time-division multiplexed. Each channel's input is alternately sampled at an 8-kHz rte and converted to an eight-bit PCM code. While the PCM code for channel-1 is being transmitted, channel-2 is sampled and converted to PCM code. When its turn for channel-2's PCM code to be transmitted, the next sample is taken from channel-1 and converted to PCM code. This is a continuous process.



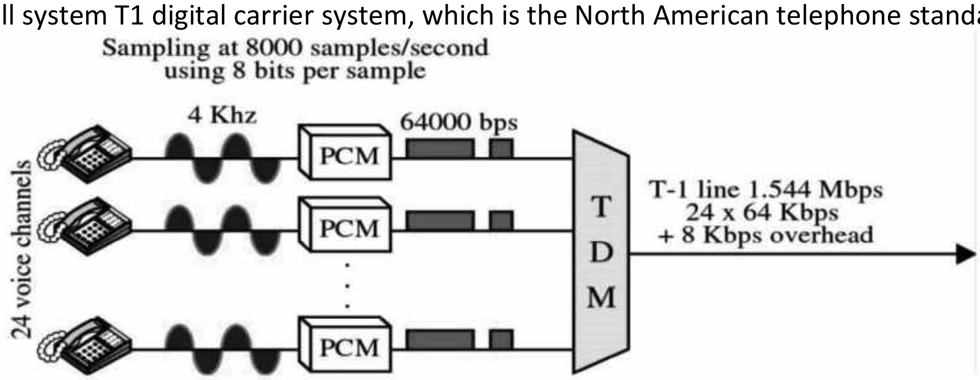
The multiplexer is simply an electronically controlled digital switch with two inputs and one output. One eight-bit PCM code from each channel is called a *<u>TDM frame</u>* and the time it takes to transmit one TDM frame is called *<u>frame time</u>* and it is equal to reciprocal of sample rate. The above figure shows the frame allocation for a two channel PCM system. The PCM code for each channel occupies a fixed time slot within the total TDM frame.

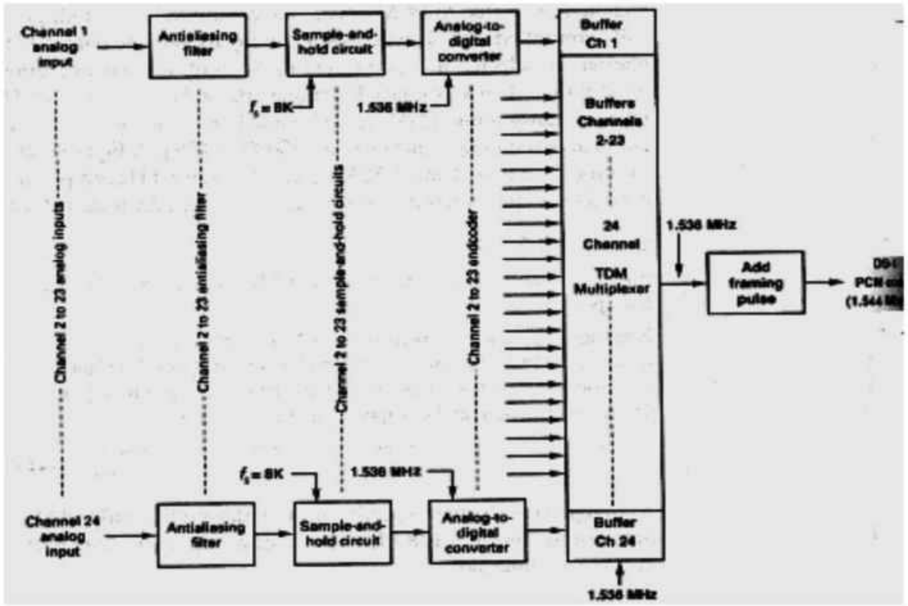


DS-hierarchy

T1 Digital Carrier System

A digital carrier system is a communications system that uses digital pulse rather than analog signals to encode information. The following figure shows a block diagram of for the Bell system T1 digital carrier system, which is the North American telephone standard.



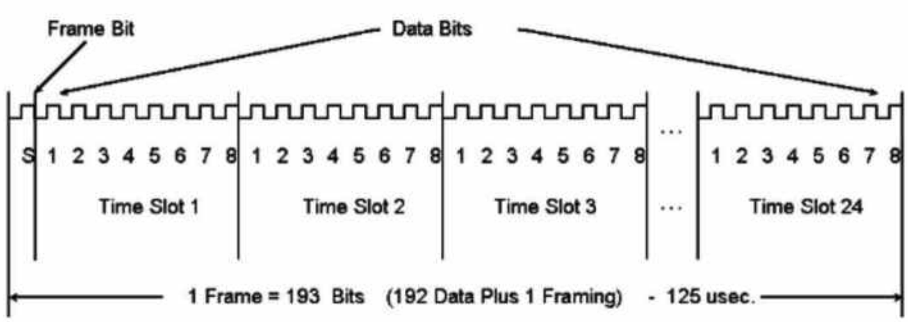


A T1 carrier system time division multiplexes PCM encoded samples from 24 voice band channels for transmission over a single metallic wire pair or a fiber optic cable. The multiplexer has 24 independent inputs and one time-division multiplexed output. The 24 PCM output signals are sequentially selected and connected through the multiplexer to the transmission line. To become a T1 carrier, the system has to be line encoded and placed on special conditioned cables called T1 lines.

A transmitting portion of a Channel Bank digitally encodes the 24 analog channels, adds signalling information into each channel, and multiplexes the digital stream onto the transmission medium. The receiving portion reverses the process. Each of the 24 channels contains an eight-bit PCM code and is sampled 8000 times a second. Each channel is sampled at the same rate, but may not be at the same time. The line speed is calculated as:

$$\frac{24\ \text{channels}}{\text{frame}} \times \frac{8\ \text{bits}}{\text{channel}} = 192\ \text{bits/frame} \Rightarrow \frac{192\ \text{bits/frame}}{\text{frame}} \times \frac{8000\ \text{frames}}{\text{second}} = 1.536\ \text{Mbps}$$

Later, an additional bit called the framing bit is added to each frame. The framing bit occurs once per frame and is recovered at the receiver and its main purpose is to maintain frame and sample synchronization between TDM transmitter and receiver.



As a result of this extra bit, each frame now contains 193 bits and the line speed for a T1 digital carrier system is 1.544 Mbps. { 193 bits × 8000 frames = 1.544 Mbps}

D-Type Channel Banks

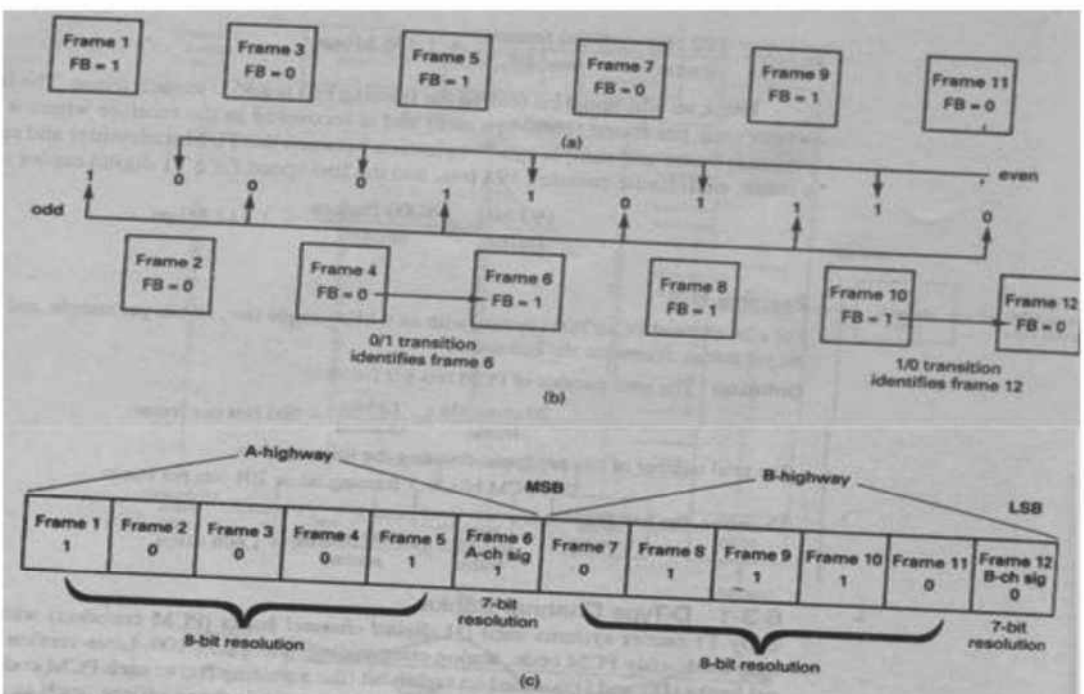
D type Channel Bank refers to the terms used in T1 technology. Channel Bank defines the type of formatting that is required for transmission on T1 trunk. The purpose of a Channel Bank in the telephone company is to form the foundation of multiplexing and demultiplexing the 24 voice channels (DS0). D type Channel Bank is one of the types of Channel Bank which is used for digital signals. There are five kinds of Channel Banks that are used in the System: D1, D2, D3, D4, and DCT (Digital Carrier Trunk).

Earlier T1 carrier systems used D1 digital channel banks (PCM encoders) with a seven-bit magnitude-only PCM code, analog companding and a $\mu = 100$. Modern versions use digital companded, eight-bit sign magnitude-compressed PCM codes with a $\mu = 255$.

Superframe TDM Format

The 8-kbps signalling rate used with the early digital channel banks was excessive for signaling on standard telephone voice circuits. Therefore, with modern channel banks, a signaling bit is substituted only into the least-significant bit (LSB) of every sixth frame. Hence, five of every six frames have eight-bit resolution, while one in every six frames (the signaling frame) has only seven-bit resolution.

Because only every sixth frame includes a signaling bit, it is necessary that all the frames be numbered so that the receiver knows when to extract the signaling bit. Also, because signaling is accomplished with a two-bit binary word, it is necessary to identify the most- and least-significant bits (MSB and LSB, respectively) of the signaling word. A Superframe format was devised as shown below.



Framing bit sequence for the T1 superframe format using D2 and D3 channel banks: (a) frame synchronizing bits (odd-numbered frames); (b) signaling frame alignment bits (even-numbered frames); (c) composite frame alignment

Within each super- frame are 12 consecutively numbered frames (1 to12). The signaling bits are substituted in frames 6 and 12, the MSB into frame 6, and the LSB into frame 12. Frames1to 6 are called the A highway, with frame 6 designated the A channel signaling frame. Frames 7 to 12 are called the B high way, with frame 12 designated the B channel signaling frame. Therefore, in addition to identifying the signaling frames, the sixth and twelfth frames must also be positively identified. To identify frames 6 and 12, a different framing bit sequence is used for the odd- and even-numbered frames. The odd frames (frames 1, 3, 5, 7, 9, and 11) have an alternating 1/0) pattern, and the even frames (frames 2, 4, 6, 8, 10, and 12) have a 00 1110repetitive pattern. As a result, the combined framing bit pattern is 1000 11011100. The odd numbered frames are used for frame and sample Synchronization and the even-numbered frames are used to identify the A and B channel signaling frames (frames 6 and 12). Frame 6 is identified by a 0/1 transition in the framing bit between frames 4 and 6. Frame 12 is identified by a 1/0 transition in the framing bit between frames *10* and 12.

D4 channel banks time-division multiplex 48 voice-band telephone channels and operate at a transmission rate of 3.152 Mbps, which is slightly more than twice the line speed for 24-channel D1, D2, or D3 channel banks because with D4 channel banks, rather than transmitting a single framing bit with each frame, a 10-bit frame synchronization pattern is used.

Line speed is calculated as: total no of bits is 8 bits/channel × 48 channels = 384 bits/frame
An additional 10 bits are added for frame: so 394 bits/frame. Therefore, line speed of DS-1C system is 394×8000 = **3.152 Mbps**

Extended Superframe Format

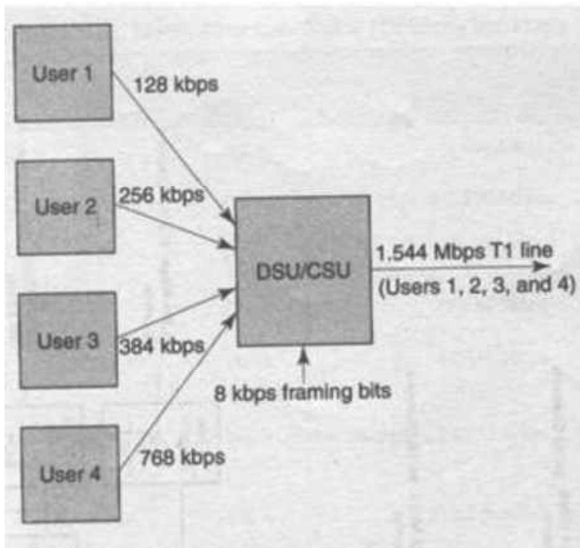
In telecommunication, an **Extended Super Frame** (**ESF**) is a T1 framing standard, sometimes called D5 framing because it was first used in the D5 Channel Bank, invented in the 1980s. It requires less frequent synchronization than the earlier superframe or D-4 format, and provides on-line, real-time testing of circuit capability and operating condition.

In ESF, a superframe is 24 frames long, and the 193rd bit of each frame is used as framing bit. Only 6 of the 24 framing bits are used for frame synchronization. Frame synchronization bits occur in frames 4, 8, 12, 16, 20 and 24 and have a bit sequence of 001011. Six additional framing bits in frames 1, 5, 9, 13, 17, and 21 are used for an error-detection code called CRC-6 (cyclic redundancy checking). The 12 remaining framing bits provide for a management channel called the facilities data link (FDL). FDL bits occur in frames 2, 3, 6, 7, 10, 11, 14, 15, 18, 19, 22, and 23.

The extended superframe format supports a four-bit signaling word with signaling bits provided in the second least-significant bit of each channel during every sixth frame. The signaling bit in frame 6 is called the A bit, in frame 12 is called the B bit, in frame 18 is C bit and in frame 24 is called D bit. These signaling streams are sometimes called the A, B, C and D signaling channels (or signaling highways).

Fractional T Carrier Service

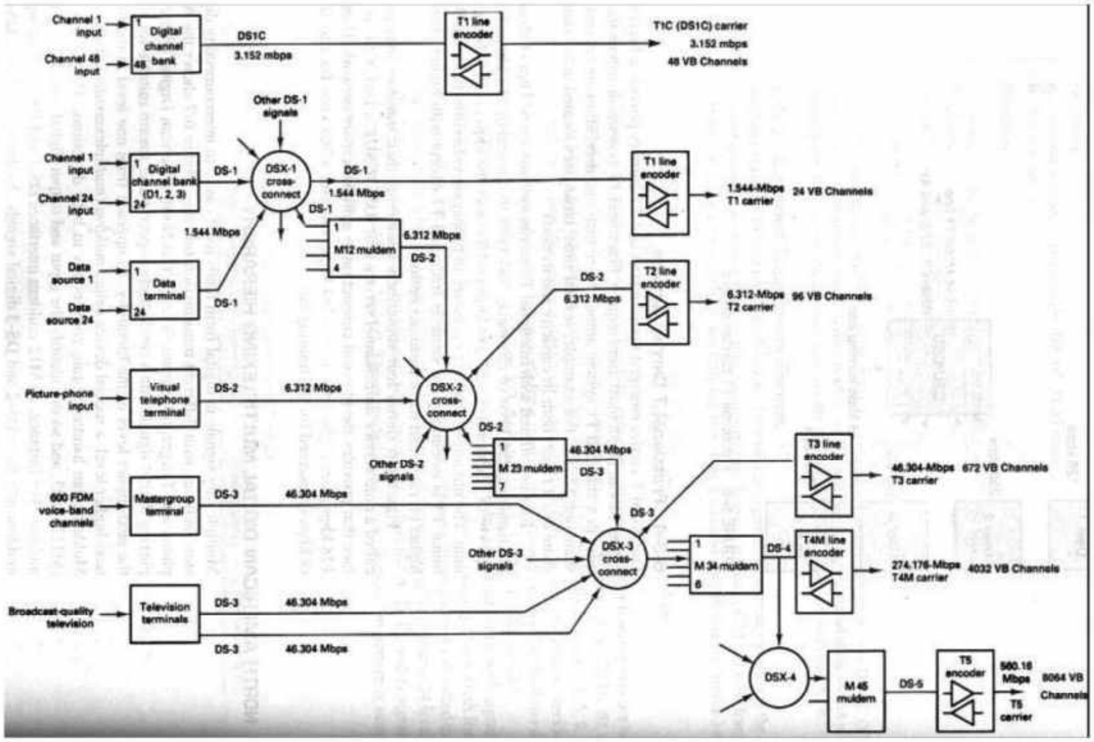
Fractional T carrier emerged because standard T1 earners provide a higher capacity (i.e., higher bit rate) than most users require. Fractional T1 systems distribute the channels (i.e., bits) in a standard T1 system among more than one user, allowing several subscribers to share one T1line. Bit rates offered with fractional T1 carrier systems are 64kbps (1 channel), 128 kbps (2 channels), *256* kbps (4 channels), 384 kbps (6 channels), 512 kbps (8 channels), and 768 kbps (12 channels), with 384 kbps (1/4 T1) and 768 kbps (1/2 T1) being the most common. The minimum data rate necessary to propagate video information is 384 kbps.



Fractional T1 carrier service

The above figure shows four subscribers combining their transmissions in a special unit called a *data service unit/channel service unit* (DSU/CSU). A DSU/CSU is a digital interface that provides the physical connection to a digital carrier network. User 1 is allocated 128 kbps, user 2 - 256 kbps, user 3 - 384 kbps, and user 4 - 768 kbps for a total of 1.536 kbps (8 kbps is reserved for the framing bit).

North American Digital Multiplexing hierarchy



The above figure shows the American Telephone and Telegraph Company's (AT&T's) North American Digital Hierarchy for multiplexing digital signals into a single higher-speed pulse stream suitable for transmission on the next higher level of the hierarchy. A special device called ***muldem*** (multiplexers/demultiplexer) is used to upgrade from one level in the hierarchy to the next-higher level. They handle bit-rate conversions in both directions and are designated as M12, M23 etc. which identifies the respective input and output digital signals. As shown, an M12 muldem interfaces DS-1 and DS-2 digital signals. Also DS-1 signals can be further multiplexed or line encoded and placed on specially conditioned cables called ***T1 lines***.

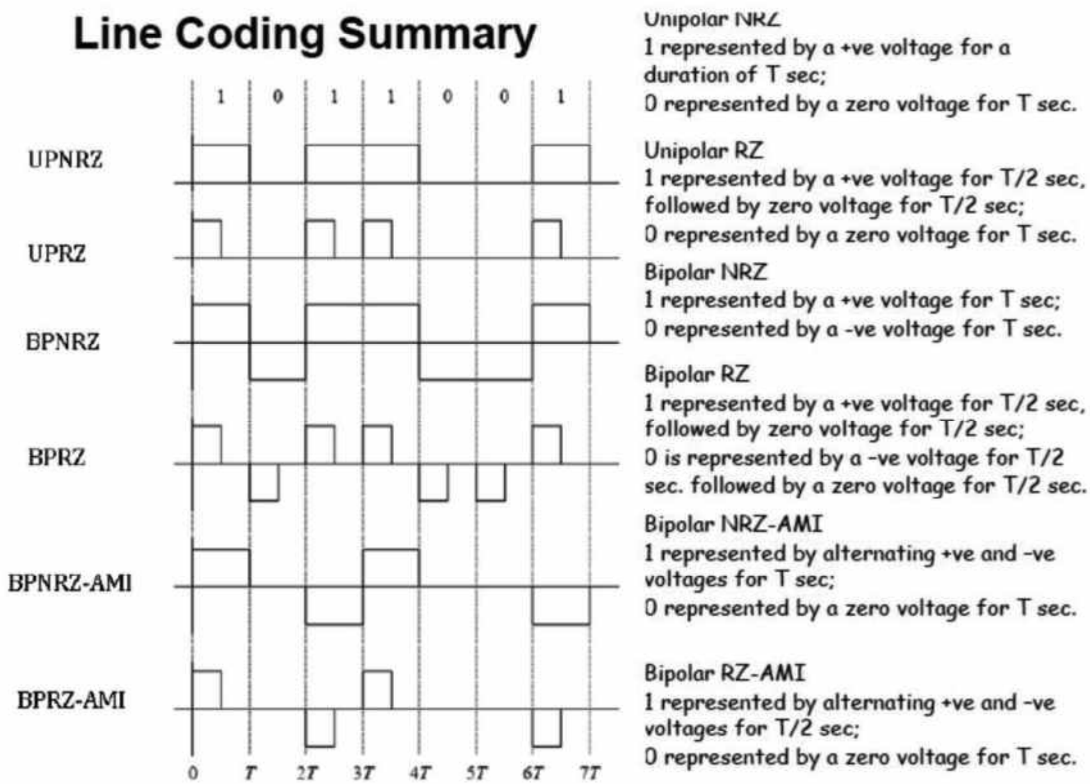
Digital signals are routed at central locations called ***digital cross-connects (DSX)***, which are convenient for making patchable interconnections and routine maintenance and troubleshooting. Each digital signal (i.e. DS-1, DS-2, etc) has its own digital switch (DSX-1, DSX-2…).

Digital Line Encoding

It involves converting standard logic levels to a form more suitable for telephone line transmission. Six factors must be considered

Transmission voltages and DC component: Transmission voltages or levels can be categorized as being either *unipolar (UP) or bipolar (BP)*. Unipolar transmission involves the transmission of only a single nonzero voltage level (either +ve or –ve for logic 1 and a 0 V for logic 0). In bipolar transmission, two nonzero voltages are involved (+ve voltage for logic 1 and equal –ve voltage for logic 0).

Duty cycle: The duty cycle of a binary pulse can be used to categorize the type of transmission. If the binary pulse is maintained for the entire bit time, this is called *nonreturn to zero (NRZ)*. If the active time of the binary pulse is less than 100% of the bit time, it's called *return to zero (RZ)*. Unipolar and Bipolar transmission voltages can be combined with either RZ or NRZ in several ways to achieve a particular line-encoding scheme. *Alternate mark inversion (AMI)* scheme involves two nonzero voltage levels (-V and +V) but both polarities represent logic 1s and 0V represents logic 0



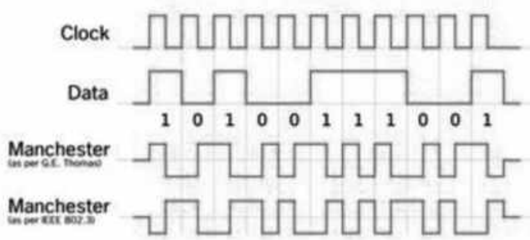
With NRZ encoding, a long string of either logic 1 or logic 0's produces a condition in which a receiver may lose its amplitude reference for discrimination between received 1's and 0's. This condition is called *dc wandering*.

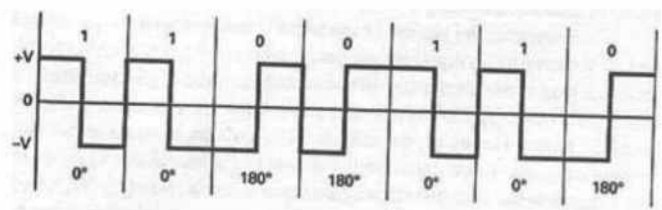
Bandwidth Requirements: The minimum bandwidth required to propagate a line-encoded digital signal is determined by the highest fundamental frequency, which is in turn determined by the worst-case (fastest transition) binary sequence. For, UPNRZ and BPNRZ the worst-case is alternating 1/0 sequence making the highest fundamental frequency one-half of the bit rate $(f_b/2)$. With BPRZ, it occurs for successive logic 1's and 0's making the minimum bandwidth equal to bitrate f_b. With BPRZ-AMI, the worst-case condition is two or more consecutive logic 1's, and minimum bandwidth is one-half of bitrate $(f_b/2)$.

Clock and framing bit recovery: To maintain clock and framing bit synchronization, there must be sufficient transitions in the data waveform. Among all, BPRZ is the best encoding scheme for clock recovery as a transition occurs in each position regardless of whether the bit is a 1 or 0.

Error detection: With UPNRZ, BPNRZ, UPRZ, and BPRZ encoding, there is no way to determine if the received data have errors. However, with BPRZ-AMJ encoding, an error in any bit will cause a bipolar violation (BPV—the reception of two or more consecutive logic is with the same polarity). Therefore, BPRZ-AMI has a built-in error-detection mechanism. T carriers use BPRZ-AMI, with +3V and —3 V representing logic 1 and 0 V representing a logic 0.

Digital Biphase: *Digital biphase* (sometimes called the *Manchester code* or diphase) is a popular type of line encoding that produces a strong timing component for clock recovery and does not dc wandering. Biphase is a form of BPRZ encoding that uses one cycle of a square wave at 0^0 phase to represent a logic 1 and one cycle of a square wave at 180^0 phase to represent a logic 0.

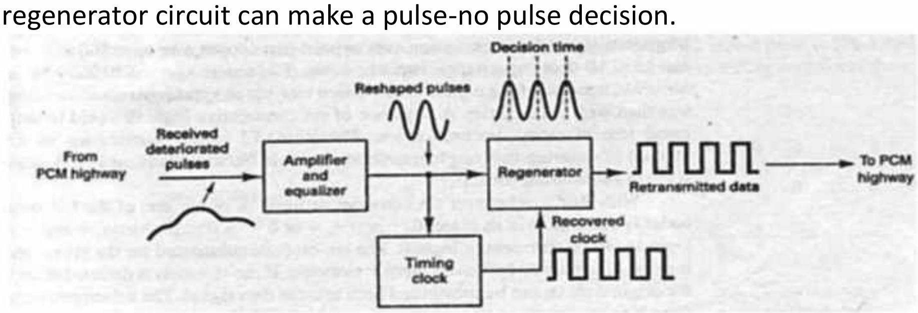




Manchester codes always have a transition at the middle of each bit period, and depending on the state of the signal, may have a transition at the beginning of the period as well. In addition, assuming equal probability of 1s and 0s, the average dc voltage is 0 V. and there is no dc wandering. A disadvantage of biphase is that it contains no means of error detection.

T Carrier Systems

T carriers are used for the transmission of PCM-encoded time-division multiplexed digital signals. Digital signals deteriorate as they propagate along a cable and regenerative repeaters are placed at periodic intervals. It has three functional blocks: an *amplifier/equalizer,* a *timing clock recovery circuit,* and the *regenerator* itself. The amplifier/equalizer filters and shapes the incoming digital signal and raises its power level so that the regenerator circuit can make a pulse-no pulse decision.

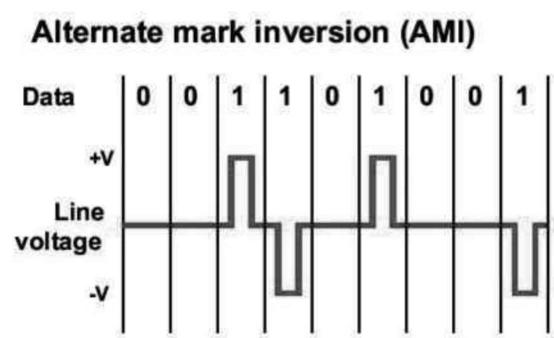


The timing clock recovery circuit reproduces the clocking information from the received data and provides the proper timing information to the regenerator so that samples can be made at the optimum time, minimizing the chance of an error occurring. A regenerative repeater is simply a threshold detector that compares the sampled voltage received to a reference level and determines whether the bit is logic 1 or logic 0. Spacing of repeaters is designed to maintain an adequate signal-to-noise ratio for error-free performance.

T1 Carrier System

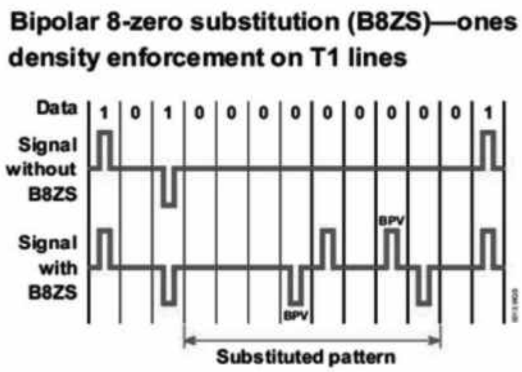
T1 carrier systems were designed to combine PCM and TDM techniques for the transmission of 24 64-kbps channels with each channel capable of carrying digitally encoded voiceband telephone signals or data. The transmission bit rate (line speed) for a T1 carrier is 1.544 Mbps. Using TDM, T1 divides this bandwidth into 24 individual DS-0 channels, sampling each channel 8000 times per second. Thus 8 × 8000 samples per second give each of the 24 DS-0 channels a data rate of 64kbps. All 24 DS-0 channels combined has a data rate of 1.544 Mbps; this digital signal level is called DS-1. Therefore T1 lines are sometimes referred to as DS-1 lines.

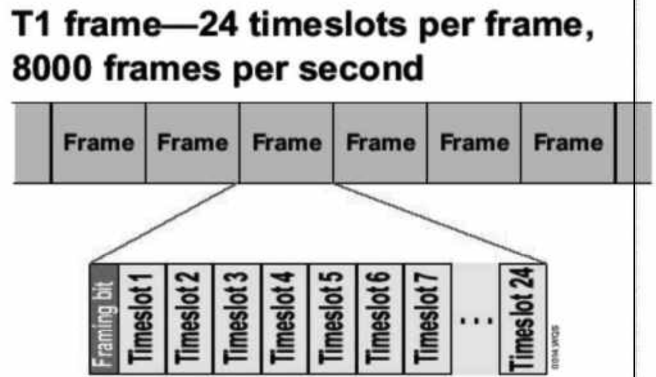
Alternate mark inversion (AMI) is the type of line coding used for T1 lines. Electrically, the signal transmitted on a T1 line is a bipolar, return-to-zero (RZ) signal. This simply means that each logical 1 bit is transmitted as a positive or a negative pulse, after which the line voltage always returns to zero. A logical 0 bit is transmitted as a zero voltage on the line. This format is known as AMI because each logical 1 bit (pulse or mark) is of opposite polarity from the previous one.

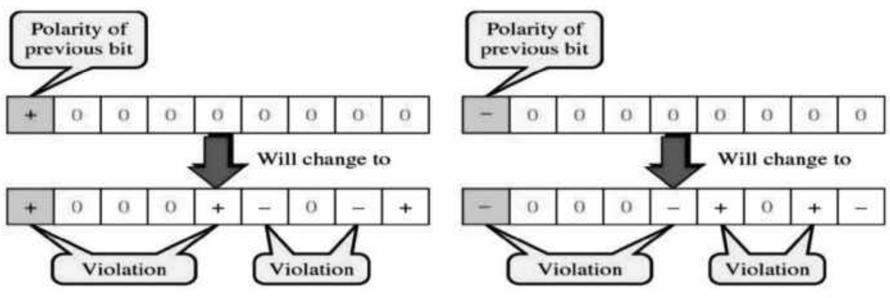


One additional benefit of the AMI bipolar format is that it permits detection of line errors. If a line problem causes a pulse to be deleted or an unintended pulse to be transmitted, two consecutive pulses with the same polarity on the line will result, called a bipolar violation (BPV).

With modern T1 carriers, a technique called *binary eight zero substitution* (B8ZS) is used to ensure that sufficient transitions occur in the data to maintain clock synchronization. Here, whenever eight consecutive 0s are encountered, one of two special patterns is substituted for the eight 0s, either **+-0-+000** or **-+0+-000**. The + and – represent bipolar logic 1 conditions and a 0 indicates a logic 0 condition.







(a) (b)

The eight-bit pattern substituted for the eight consecutive 0s is the one that purposely induces bipolar violations in the fourth and seventh bit positions. This code is then interpreted at the remote end of the connection. A full 1.544 Mbps T1 line contains 24 fractional T1 lines (abbreviated as FT1), each with a bandwidth of 64 kbps.

Some of the limitations of T1 services are: They are very expensive, installation cost is also very high and some case improper utilization of bandwidth.

T2 Carrier System

T2 carriers time-division multiplex 96 64-kbps voice or data channels into a single 6.312 Mbps data signal for transmission over twisted-pair copper wire upto 500 miles over a special LOCAP (low capacitance) metallic cable. Higher transmission rates make clock synchronization even more critical. So, an alternative method called *binary six zero substitution* (B6ZS) is used to ensure that ample transitions occur in the data.

B6ZS Example:

Original data:	0	-	0	0	0	0	0	0	+	0	-	+
After substitution:	0	-	0	-	+	0	+	-	0	+	0	-	+
Original data:	0	+	0	0	0	0	0	0	-	0	+	-
After substitution:	0	+	0	+	-	0	-	+	0	-	0	+	-

Whenever six consecutive logic 0s occur, either **0-+0+-** or **0+-0-+** is substituted, and this code is selected to create a bipolar violations in the second and fourth bits of the substituted patterns.

T3 Carrier System

T3 carriers time-division multiplex 672 64-kbps voice or data channels for transmission over a single 3A-RDS coaxial cable. The transmission bit rate is 44.736 Mbps and coding technique used with T3 carriers is *binary three zero substitution* (B3ZS).

T4M Carrier System

T4M carriers time division multiplex 4032 64-kbps voice or data channels for transmitting over a single T4M coaxial cable upto 500 miles. The transmission rate is very high (274.16 kbps) making substituting patterns impractical. They transmit scrambled unipolar NRZ digital signals.

T5 Carrier System

T5 carriers time-division multiplex 8064 64-kbps voice or data channels and transmits them at 560.16 Mbps over a single coaxial cable.

European Time-Division Multiplexing

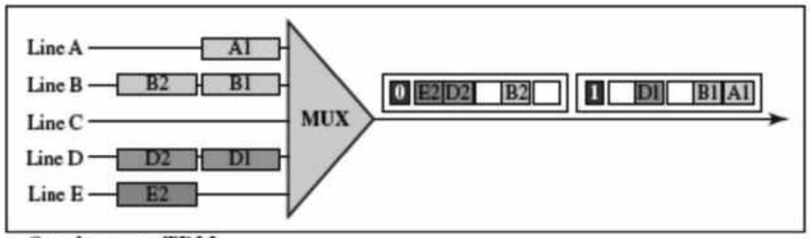
In Europe, a different version of T carrier lines is used called E lines. With the basic E1 system, a 125µs frame is divided into 32 equal time slots. Time slot 0 is used for a frame alignment pattern and for an alarm channel. Time slot 17 is used for a common signalling channel (CSC). The signalling for all 30 voice-band channels is accomplished on the common signalling channel. Consequently, 30 voice-band channels are time-division multiplexed into each E1 frame. Every slot has eight bits. So the number of bits per frame is given as:

$$\frac{\text{8 bits}}{\text{time slot}} \times \frac{\text{32 time slots}}{\text{frame}} = \text{256 bits/frame}$$

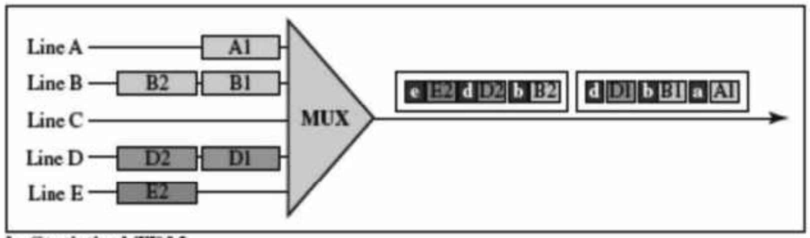
And the line speed can be given as **256 bits/frame × 8000 frames/second = 2.408 Mbps**

Statistical Time Division Multiplexing

Statistical time division multiplexing, is one method for transmitting several types of data simultaneously across a single transmission cable or line. STA-TDM is often used for managing data being transmitted via a local area network (LAN) or a wide area network (WAN). A statistical TDM multiplexer exploits the natural breaks in transmissions by dynamically allocating time slots on a demand basis. As like synchronous TDM system, a statistical mux has a finite number of low-speed data input lines with one high-speed multiplexed data output line, and each input line has its own digital encoder and buffer. With the statistical mux, there are *n* input lines and *k* time slots available (k >n). The multiplexer scans the input buffers, collecting data until a frame is filled, at which time the frame is transmitted. On the receiving end, the demultiplexer removes the data from the time slots and distributes to their appropriate output buffers. Statistical multiplexers require low data rate than synchronous multiplexers. Also, they can support more users operating at the same transmission rate.



a. Synchronous TDM



b. Statistical TDM

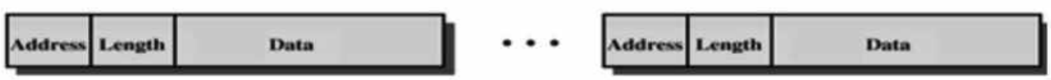
With Statistical multiplexing, control bits must be included in the frame. The following figure shows the overall frame format for a statistical TDM multiplexer.

(a) Overall frame

The frame includes a beginning flag and ending flag to indicate the start and end of frame, an address field that indicates the transmitting device, a control field, a statistical TDM subframe, and a frame check sequence field (FCS), which provides error detection.

(b) Subframe with one source per frame

The above figure shows the frame when only one data source is transmitting. The transmitting devise is identified in the address field. The data field is variable and this scheme works well in times of light loads, but inefficient for heavy loads.



(c) Subframe with multiple sources per frame

The above figure shows a way to improve efficiency by allowing more than one data source to be included within a single frame.

Frame Synchronization

With TDM systems, it is important not only that a frame has to be identified, but also individual timeslots within the frame be identified. There are several methods used to establish frame synchronization, including added digit, robbed digit, added channel, statistical and unique coding. Considerable amount of overhead is added to transmission to achieve frame synchronization.

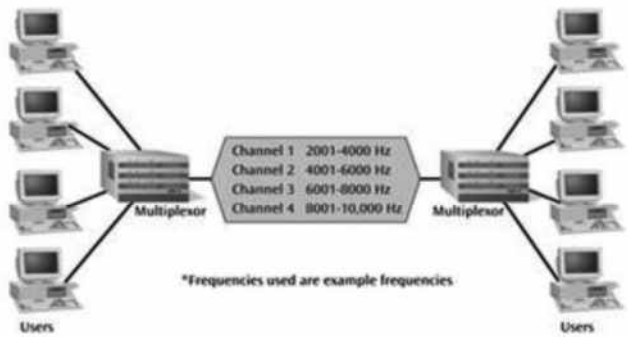
1. ***Added-Digit Framing***: - T1 carriers using D1, D2 or D3 channel banks use added-digit framing. A special framing digit (framing pulse) is added to each frame. The maximum average synchronization time is given by

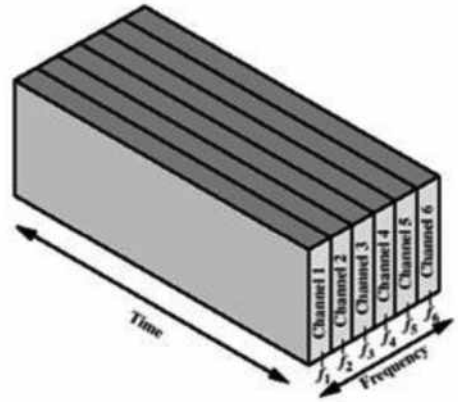
   Synchronization time = 2NT = $2N^2 t_b$

   Where N is number of bits per frame and T is frame period of Nt_b and t_b is bit time.
2. ***Robbed-Digit Framing***: - Added-digit framing is inefficient when a short frame is used in the case of single-channel PCM systems. As an alternative, the least significant bit of every n^{th} frame is replaced with a framing bit. This process is called robbed-digit framing and it does not interrupt transmission, but instead periodically replaces information bits with forced data errors to maintain frame synchronization.
3. ***Added-Channel Framing***: - It is essentially same as added-digit framing except that digits are added in groups or words instead of as individual bits. The average number of bits to acquire frame synchronization using added-channel framing is $\mathbf{N^2/2(2^K - 1)}$, where N is number of bits per frame and K is number of bits in the synchronizing word.
4. ***Statistical Framing***: - Here no robbing or adding digits is done. As a signal that has a centrally peaked amplitude distribution generates a high probability of logic 1 in the second digit, the second digit of a given channel can be used for the framing bit.
5. ***Unique-Line Code Framing***: - Some property of the framing bit is different from the data bits. The framing bit is either made higher or lower in amplitude or with a different time duration. The advantage is that synchronization is immediate and automatic. The disadvantage is additional processing requirements necessary to generate and recognize the unique bit.

Frequency Division Multiplexing

Assignment of non-overlapping frequency ranges to each “user” or signal on a medium, such that all signals are transmitted at the same time, each using different frequencies. FDM is used for combining many relative narrowband sources into a single wideband channel, such as in public telephone systems. Essentially FDM is taking a given bandwidth and subdividing it into narrow segments with each segment carrying different information. FDM is an analog multiplexing scheme.





In the above figure-b, five signal sources are fed into a multiplexer that modulates each signal onto a different frequency (f_1, f_2, f_3, f_4 , f_5). To prevent interference, the channels are separated by guard bands, which are unused portions of the spectrum. With FDM, each user has its own modulating circuitry, a transmitter, a receiver and a demodulator. The channel is common to all users. Since each transmitter is using a carrier of a different frequency, there is no interference unless the sidebands or carriers are incorrectly assigned and therefore overlap. AM, FM and cable TV broadcasting are most common examples of FDM where each station uses a different frequency band.

Advantages of FDM:

1. In FDM system, users can be added to the system by simply adding another pair of transmitter modulator and receiver demodulators.
2. FDM system support full duplex information flow which is required by most of the applications
3. Noise problem for analog communication has less effect.

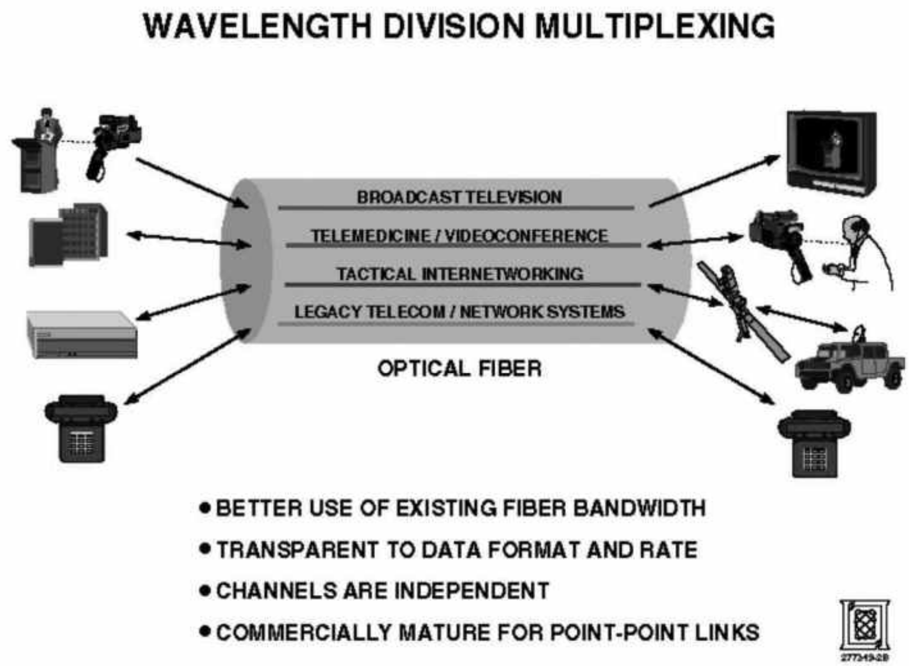
Disadvantages of FDM:

1. The initial cost is high, which includes the cable between the two ends and associated connectors for the cable.
2. One users problem can sometimes affect others
3. Each user requires a precise carrier frequency.

Wavelength-Division Multiplexing

WDM involves transmission of multiple digital signals using several wavelengths without their interfering with one another. This technology enables many optical signals to be transmitted simultaneously by a single fiber cable. It is also referred to as wave-division multiplexing.

WDM is accomplished by modulating injection laser diodes, which are transmitting highly concentrated light waves at different wavelengths (i.e. at different optical frequencies). Therefore WDM is coupling light at two or more discrete wavelengths into and out of an optical fiber. Each wavelength is capable of carrying vast amounts of information in either analog or digital form, and the information can already be time- or frequency-division multiplexed.



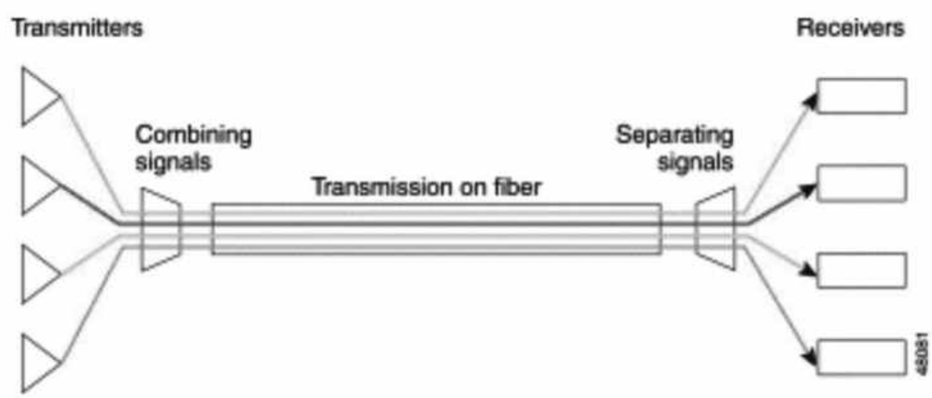
Advantages of WDM:

1. Enhanced capacity as full-duplex transmission is also possible with a single fiber.
2. WDM is inherently easier to reconfigure (i.e. adding of removing channels)
3. Usage of optical components makes it simpler, more reliable and often less

costly Disadvantages of WDM:

1. Signals cannot be placed so close in the wavelength spectrum that they interfere with each other.
2. The overall signal strength should be approximately the same for each wavelength which may not be possible.
3. Light waves carrying WDM are limited to a two-point circuit or a combination of many two-point circuits that can go only where the cable goes.

Dense wavelength division multiplexing (DWDM) is a fiber-optic transmission technique that employs light wavelengths to transmit data parallel-by-bit or serial-by-character.



Advantages:

- Protocol & Bit Rate independence
- Increased overall capacity at much lower cost
  - Current fiber plant investment can be optimized by a factor of at least 32
- Transparency
  - Physical layer architecture → supports both TDM and data formats such as ATM, Gigabit Ethernet, etc.
- Scalability
  - Utilize abundance of dark fibers in metropolitan areas and enterprise networks

Disadvantages:

- Dispersion
  - Chromatic dispersion
  - Polarization mode dispersion
- Attenuation
  - Intrinsic: Scattering, Absorption, etc.
  - Extrinsic: Manufacturing Stress, Environment, etc.
- Four wave mixing
  - Non-linear nature of refractive index of optical fiber
  - Limits channel capacity of the DWDM System

Advantages and disadvantages of multiplexing techniques

Multiplexing Technique	Advantages	Disadvantages
Frequency Division Multiplexing	Simple Popular with radio, TV, cable TV Relatively inexpensive All the receivers, such as cellular telephones, do not need to be at the same location	Analog signals only Limited by frequency ranges
Synchronous Time Division Multiplexing	Digital signals Relatively simple Commonly used with T-1 and ISDN	Wastes bandwidth
Statistical Time Division Multiplexing	More efficient use of bandwidth Packets can be various sizes Frame can contain control and error information	More complex than synchronous time division multiplexing
Dense Wavelength Division Multiplexing	Very high capacities over fiber Scalable Signals can have varying speeds	Cost Complexity
Code Division Multiplexing	Large capacities Scalable	Complexity

Synchronous Optical Network (SONET)

The synchronous optical network is a multiplexing system similar to conventional time-division multiplexing except SONET was developed to be used with optical fibers. SONET is the name for a standard family of interfaces for high speed optical links. These start at 51.84 Mbps, which is referred to as synchronous transport level 1 (STS-1). It is comprised of 28 DS-1 signals. Each DS-1 signal is equivalent to a single 24-channel T1 digital carrier system. With STS-1, it is possible to extract or add individual DS-1 signals with completely disassembling the entire frame. OC-48 is the second level of SONET multiplexing. It has a transmission bit rate of 2.48Gbps.

SONET Applications:

1. High speed backbone networks
2. Basic architecture for B-ISDN
3. Basic architecture for ATM
4. High speed optical networks for data communications.

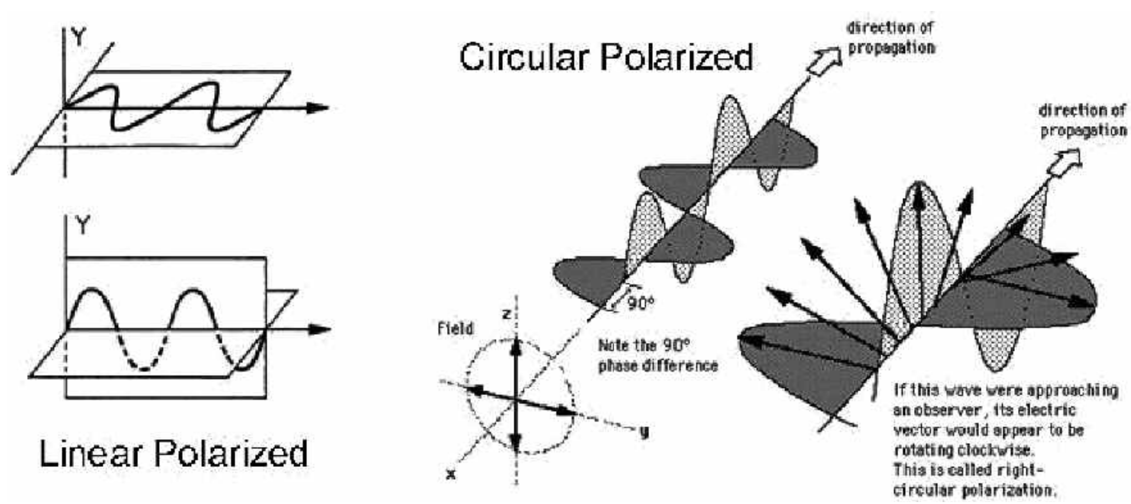
UNIT - V

WIRELESS COMMUNICATIONS SYSTEMS

Electromagnetic Polarization, Rays and Wavefronts, Electromagnetic Radiation, Spherical Wavefront and the Inverse Square Law, wave Attenuation and Absorption, Optical Properties of Radio Waves, Terrestrial Propagation of Electromagnetic Waves, Skip Distance, Free-Space Path Loss, Microwave Communications Systems, Satellite Communications Systems

With ***wireless communication systems***, electromagnetic signals are emitted from an antenna, propagate through the earth's atmosphere (air) or free space (a vacuum), and are then received (captured) by another antenna. Sometimes, it is impractical to interconnect two pieces of equipment physically. So, free space or earth's atmosphere is often used as the transmission medium. Free space propagation of electromagnetic waves is often called radio-frequency (RF) propagation or simply radio propagation. Wireless communications include terrestrial and satellite microwave radio systems, broadcast radio systems, two-way mobile radio and cellular telephone.

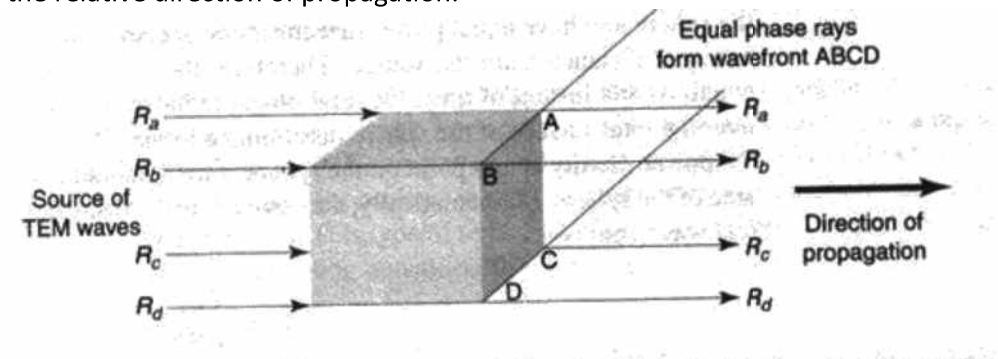
Electromagnetic Polarization



Electromagnetic waves are comprised of an electric and a magnetic field at 90 degrees to each other. The *polarization* of a plane electromagnetic wave is simply the orientation of the electric field vector in respect to earth's surface. If the polarization remains constant, it is described as *linear polarization.* Horizontal and vertical polarizations are two forms of linear polarization. A wave is *horizontally polarized* if the electric field propagates parallel to the earth's surface, and the wave is *vertically polarized* if the electric field propagates perpendicular to the earth's surface. The wave is described as having ***circular polarization*** if the polarization vector rotates 360 degrees, as the wave moves one wavelength through space and the field strength is equal at all angles of polarization. When the field strength varies with changes in polarization, this is described as ***elliptical polarization.*** A rotating wave can turn in either direction. If the vector rotates in a clockwise direction, it is ***right handed,*** and if the vector rotates in a counter-clockwise direction, it is considered ***left handed.***

Rays and Wavefronts

Rays and wavefronts are used for analysing electromagnetic waves. A ray is a line drawn along the direction of propagation of an electromagnetic wave. Rays are used to show the relative direction of propagation.



Plane wave comprised of rays R_a, R_b, R_c, and R_d forming wavefront ABCD

A wavefront shows a surface of constant phase of electromagnetic waves. A wavefront is formed when points of equal phase on rays propagating from the same source are joined together. The above figure shows a wavefront with a surface that is perpendicular to the direction of propagation (rectangle ABCD). When a surface is plane, its wavefront is perpendicular to the direction of propagation. A point source is a single location from which rays propagate equally in all directions (i.e. isotropic source). The wavefront generated from a point source is simply a sphere with radius R and its center located at the point of origin of the waves.

Electromagnetic Radiation

The flow of electromagnetic waves (energy) in the direction of propagation is called electromagnetic radiation. The rate at which energy passes through a given surface area in free space is called power density, usually given in watts per square meter. Mathematically, Power density $\mathcal{P} = \mathcal{E}\mathcal{H}$, where P is power density (watt/m^2), E represents rms electric field intensity (volts/meter) and H represents rms magnetic field intensity (ampere turns/meter).

Spherical Wavefront and Inverse Square Law

A <u>*spherical wavefront*</u> is obtained by an isotopic radiator. All points at distance R(radius) from the source lie on the surface of the sphere and have equal power densities. At an instance of time, the total power radiated P_{rad} is uniformly distributed over the total surface of the sphere. Therefore, the power density at any point on the sphere is the total radiated power divided by the total area of the sphere and can be given as,

$$\mathcal{P} = P_{rad} / (4\pi R^2)$$

The power density becomes smaller as the distance from isotropic source increases. The total radiated power is same. But as the area of the sphere increases in direct proportion to the square of distance from source, the power density is inversely proportional to the square of the distance from the source. This relationship is called *inverse square law.*

Wave Attenuation and Absorption

When waves propagate through free space, they spread out, resulting in reduction of power density. This is called *attenuation loss* and it occurs in free space as well as earth's atmosphere. Earth's atmosphere contains different particles which absorb electromagnetic energy, causing reduction in power, called as *absorption loss*. The reduction in power density with increase in distance is equivalent to a power loss and is called wave attenuation. Because it's due to spherical spreading of wave in space, it is sometimes called space attenuation. Mathematically, wave attenuation is

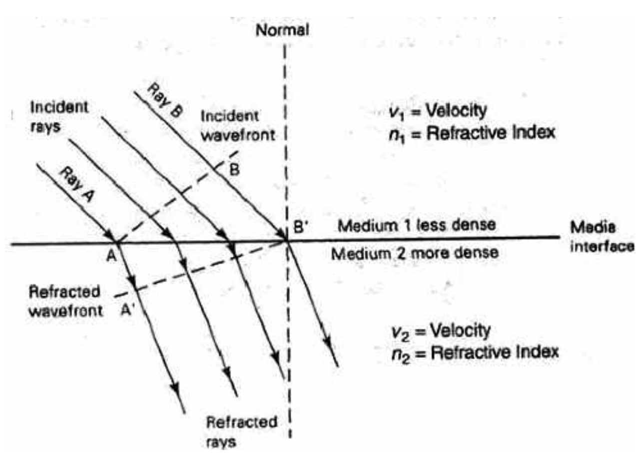
$\gamma_A = 10 \log (P_1/P_2)$, where γ_A represent wave attenuation in dB, P_1 is power density at point 1 and P_2 is power density at point 2.

Earth's atmosphere is not a vacuum and it consists of atoms, molecules of various substances such as gases, liquids and solids, which are quite capable of absorbing EM waves. As the wave propagates, energy is transferred from the wave to the atoms and molecules and this transfer is known as *wave absorption* and is analogous to I2R power loss. Once absorbed, energy is lost forever and causes reduction in the power density.

Optical Properties of Radio Waves

The free space behaviour of propagation is altered by optical effects such as refraction, reflection, diffraction and interference.

Refraction: Electromagnetic *refraction* is the change in direction of an electromagnetic wave as it passes obliquely from one medium to another medium with a different density (refractive index).



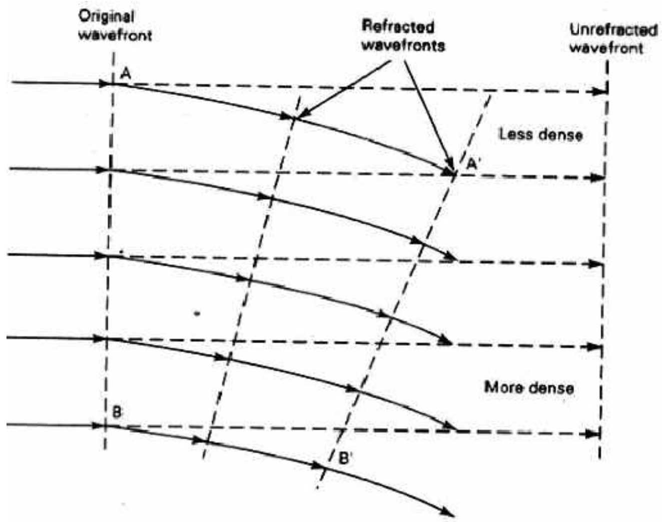
Refraction at a plane boundary between two media

Whenever a ray passes from a less dense to a more dense medium, it is effectively bent toward the normal (imaginary line drawn perpendicular to the interface at the point of incidence). Conversely, whenever a ray passes from a more dense to a less dense medium, it is effectively bent away from the normal. The *angle of incidence* is the angle formed between the incident wave and the normal, and the *angle of refraction* is the angle formed between the refracted wave and the normal. Snell's law states that,

$$sin\theta_1\left(\frac{n_1}{n_2}\right) = sin\theta_2$$

, where θ_1 and θ_2 are angles of incidence and refraction and n_1 and n_2 are refractive indexes of material1 and material2.

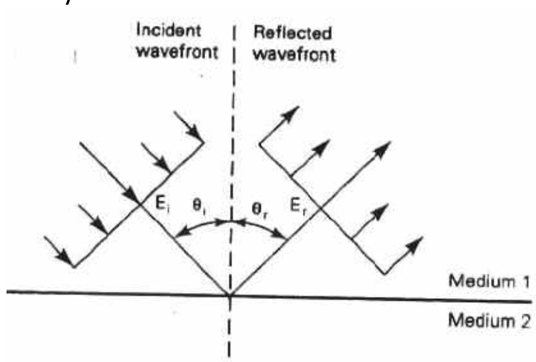
Refraction also occurs when a wavefront propagates in a medium that has a density gradient that is perpendicular to the direction of propagation. The following figure shows wavefront refraction in earth's atmosphere (which has gradient refractive index).



Wavefront refraction in a gradient medium

The medium is more dense near the bottom and less dense near the top (upper atmosphere). Therefore, rays travelling in the upper layers of the atmosphere travel faster than rays travelling near earth's surface and, consequently, the wavefront tilts downward. The tilting occurs in a gradual fashion as the wave progresses.

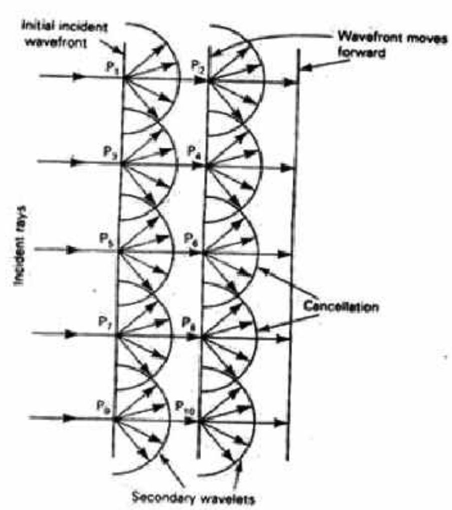
Reflection: Electromagnetic wave *reflection* occurs when an incident wave strikes a boundary of two media and some or all of the incident power does not enter the second material (i.e., they are reflected). The following figure shows electromagnetic wave reflection at a plane boundary between two media.

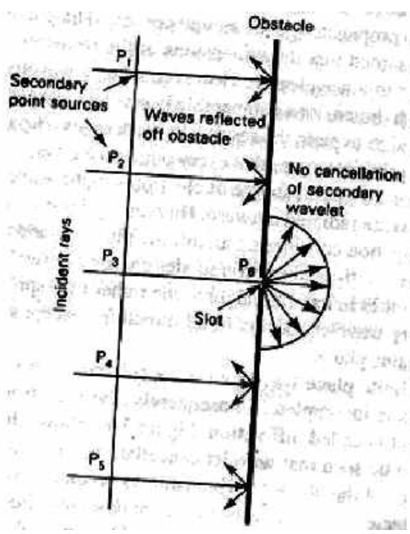


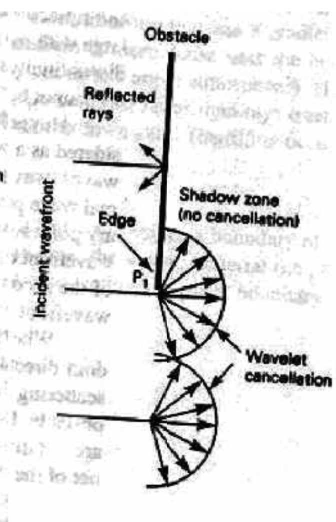
Because all the reflected waves remain in medium1, angle of reflection equals the angle of incidence ($\theta_i = \theta_r$). The ratio of reflected to incident power is $\boldsymbol{\Gamma}$, expressed as $\Gamma = \mathbf{P}_r/\mathbf{P}_i$ where $\boldsymbol{\Gamma}$ is reflection coefficient and Pr and Pi are reflected and incident power.

For perfect conductors, $\boldsymbol{\Gamma} = 1$ and all incident power is reflected. Reflection also occurs when the reflective surface is irregular. When an incident wavefront strikes an irregular surface, it is randomly scattered in many directions. Such a condition is called *diffuse* ref*lection,* whereas reflection from a perfectly smooth surface is called *specular* (mirror like) *reflection.*

Diffraction: *Diffraction* is defined as the modulation or redistribution of energy within a wavefront when a density it passes near the edge of an opaque object. Diffraction is the phenomenon that allows light or radio waves to propagate (peek) around corners. Huygen's principle states that every point on a given spherical wavefront can be considered as a secondary point source of electromagnetic waves from which other secondary waves (wavelets) are radiated outward. Huygen's principle is illustrated below.



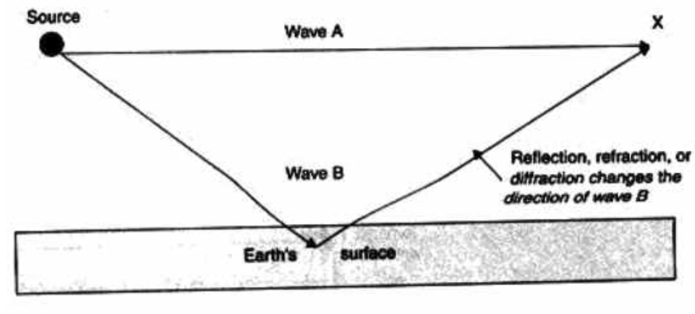




The first figure shows normal wave propagation considering an infinite plane. Each secondary point source *(P1, P2* and so on) radiates energy outward in all directions. But, the wavefront continues in its original direction rather than spreading out because cancellation of the secondary wavelets occurs in all directions except straight forward. Therefore, the wavefront remains plane. When a finite plane wavefront is considered, as in second figure, cancellation in random directions is incomplete. So, the wavefront spreads out or scatters. This scattering effect is called diffraction.

The third figure shows diffraction around the edge of an obstacle. It can be seen that wavelet cancellation occurs only partially. Diffraction occurs around the edge of the obstacle, which allows secondary waves to "sneak" around the corner of the obstacle into what is called the *shadow zone.*

Interference: Radio wave *interference* occurs when two or more electromagnetic waves combine in such a way that system performance is degraded. Interference, on the other hand, is subject to the principle of *linear superposition* of electromagnetic waves and occurs whenever two or more waves simultaneously occupy the same point in space.

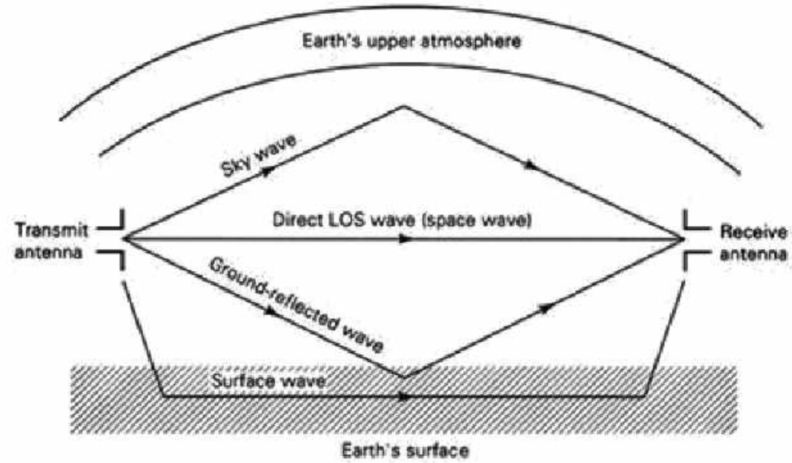


Electromagnetic wave interference

In the above figure, it can be seen that, at point X the two waves occupy the same area in space. However, wave B has travelled a different path than wave A and, therefore, their relative phase angles may be different. If the difference in distance travelled is an odd-integral multiple of one-half wavelength, reinforcement takes place. If the difference is an even-integral multiple of one- half wavelength, total cancellation occurs.

Terrestrial Propagation of Electromagnetic Waves

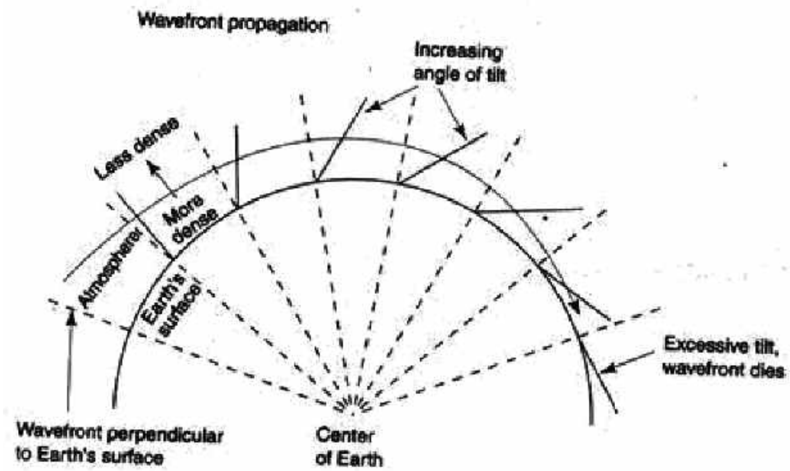
Electromagnetic waves travelling within earth's atmosphere are called terrestrial waves and communications between two or more points on earth is called terrestrial radio communications. There are three modes of propagating EM wave within earth's atmosphere: ground wave propagation, space wave propagation and sky wave propagation.



Normal modes of wave propagation

Ground Wave Propagation: Ground waves are the electromagnetic waves that travel along the surface of earth and are also called as surface waves. Ground waves must be vertically polarized and the changing electric field induces voltages in earth's surface, which cause currents to flow that are very similar to those in a transmission line. Ground waves are attenuated as they propagate because of the presence of resistance and dielectric losses in the earth's surface. Ground waves propagate best over a surface that is a good conductor, such as salt water and poorly over dry desert areas. Also losses in ground waves increase rapidly with frequency, ground wave propagation is limited to frequencies below 2 MHz

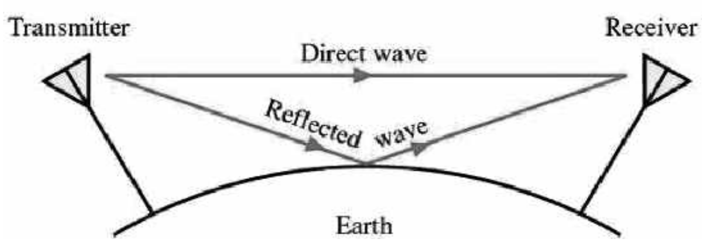
The following figure shows ground wave propagation. Because of earth's gradient density, ground wave propagates around the earth, remaining close to its surface.



Surface (ground) wave propagation

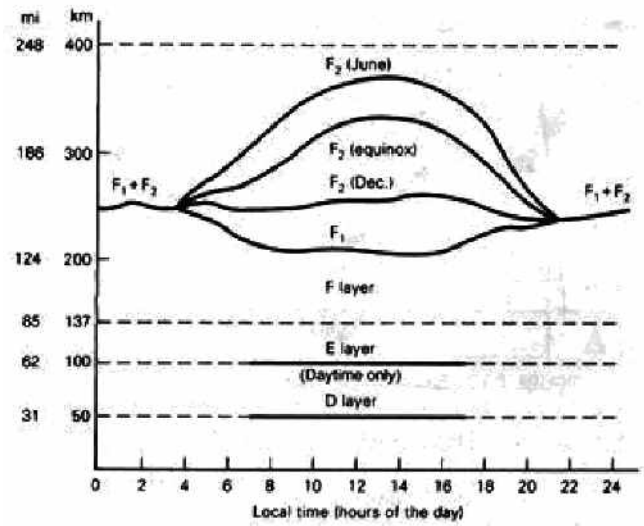
The frequency and terrain over which the ground wave propagates has to be selected carefully to ensure that the wavefront does not tilt excessively and simply turn over, lie flat on the ground and cease to propagate. Ground wave communication is commonly used for ship-to-ship and ship-to-shore communications, for radio navigation and for maritime mobile communications.

Space Wave Propagation: It includes radiated energy that travels in the lower few miles of earth's atmosphere. Space wave include both direct and ground reflected waves.



Direct waves travels essentially in a straight line between transmit and receive antennas. And this propagation with direct waves is commonly called *line-of-sight* (LOS) *transmission.* Direct space wave propagation is limited by the curvature of the earth. *Ground-reflected waves* are waves reflected by earth's surface as they propagate between transmit and receive antennas. The field intensity at the receive antenna depends on the distance between the two antennas (attenuation and absorption) and whether the direct and ground-reflected waves are in phase (interference). The curvature of earth presents a horizon to space wave propagation commonly called the *radio horizon.* Because the conditions in earth's lower atmosphere are subject to change, the degree of refraction can vary with time. A special condition called *duct propagation* occurs when the density of the lower atmosphere is such that electromagnetic waves can propagate within the duct for great distances, causing them to propagate around earth following its natural curvature.

Sky Wave Propagation: Electromagnetic waves that are directed above the horizon level are called *sky waves.* Sky waves are radiated toward the sky, where they are either reflected or refracted back to earth by the *ionosphere.* Because of this, sky wave propagation is sometimes called *ionospheric propagation.* The ionosphere is the upper portion of earth's atmosphere and is located approximately 50km to 400km (31 mi to 248 mi) above earth's surface. Because of the ionosphere's non uniform composition and its temperature and density variations, it is *stratified.* Essentially, three layers make up the ionosphere (the D, E, and F layers) and are shown below:

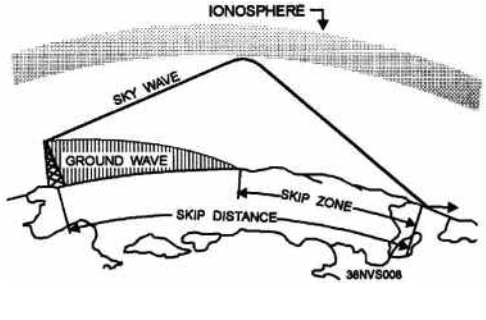


Ionospheric Layers

All three layers of the ionosphere vary in location and in *ionization density* with the time of day. The ionosphere is most dense during times of maximum sunlight. Because the density and location of the ionosphere vary over time, the effects it has on electromagnetic radio wave propagation also vary.

Skip Distance

The *skip distance* is the distance from the transmitter to the point where the sky wave first returns to the earth. The skip distance depends on the wave's frequency and angle of incidence, and the degree of ionization.



The SKIP ZONE is a zone of silence between the point where the ground wave becomes too weak for reception and the point where the sky wave is first returned to Earth. The size of the skip zone depends on the extent of the ground wave coverage and the skip distance. When the ground wave coverage is great enough or the skip distance is short enough that no zone of silence occurs, there is no skip zone.

Free-Space Path Loss

In telecommunication, **free-space path loss** (**FSPL**) is the loss in signal strength of an electromagnetic wave that would result from a line-of-sight path through free space, with no obstacles nearby to cause reflection or diffraction. With free-space path loss, no electromagnetic energy is actually lost—it merely spreads out as it propagates away from the source resulting in a lower power density. It's also referred as spreading loss, which occurs simply because of the inverse square law. Spreading loss is a function of distance from the source and the wavelength (frequency) of the electromagnetic wave. Mathematically, free-space path loss is proportional to the square of the distance between the transmitter and receiver, and also proportional to the square of the frequency of the radio signal.

$$\begin{aligned}\mathrm{FSPL} &= \left(\frac{4\pi d}{\lambda}\right)^2 \\ &= \left(\frac{4\pi d f}{c}\right)^2\end{aligned}$$

Where:
λ is the signal wavelength (in metres),
f is the signal frequency (in hertz),
d is the distance from the transmitter (in metres),
c is the speed of light in a vacuum, 2.99792458×10^8 metres per second

For typical radio applications, it is common to find f measured in units of MHz and d in km, in which case the FSPL equation becomes

$$\mathrm{FSPL(dB)} = 20\log_{10}(d) + 20\log_{10}(f) + 32.45$$

Microwave Communication Systems

Microwaves are generally described as electromagnetic waves with frequencies that range from approximately 500 MHz to 300 GHz. Because of their high frequencies, microwaves have relatively short wavelengths. Microwave systems are used for carrying long-distance voice telephone service, metropolitan area networks, wide area networks and the Internet. There are different types of microwave systems operating over distances that vary from 15 miles to 4000 miles in length. Intrastate or feeder service microwave systems are generally classified as short haul because they are used to carry information for relatively short distances, such as between cities within the same state. Long-haul microwave systems are those used to carry information for relatively long-distances, such as interstate and backbone route applications. Microwave radio system capacities range from less than 12 voice grade telephone circuits to more than 22,000 voice and data channels.

Advantages of Microwave Radio Communication:

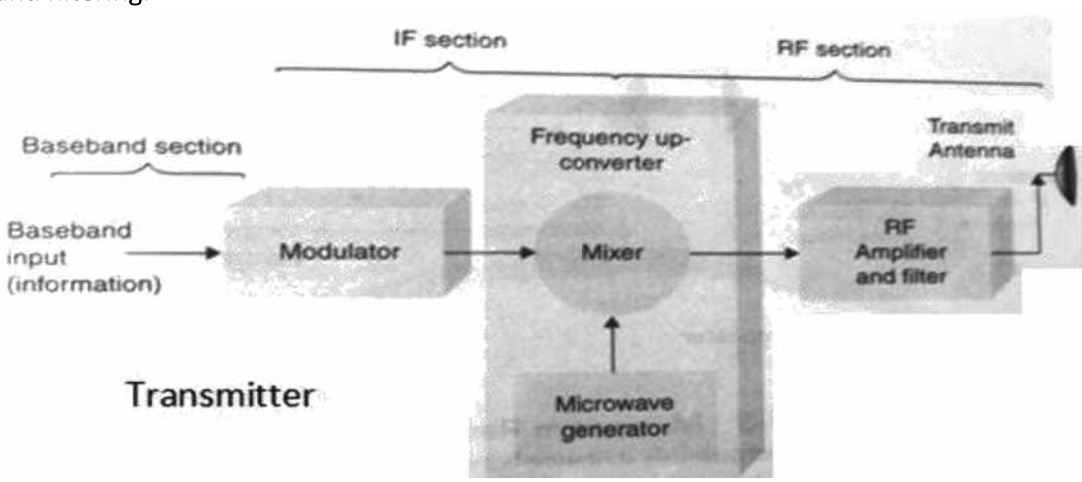
1. Radio systems do not require a right-of-way acquisition between stations.
2. Each station requires the purchase or lease of only a small area of land.
3. Because of their high operating frequencies, microwave radio systems can carry large quantities of information.
4. High frequencies mean short wavelengths, which require relatively small antennas
5. Radio signals are more easily propagated around physical obstacles, such as water and high mountains.
6. Microwave Systems require fewer repeaters for amplification.
7. Distances between switching centers are less.
8. Underground facilities are minimized.
9. Minimum delay times are introduced.
10. Minimal crosstalk exists between voice channels.

Disadvantages of Microwave radio systems:

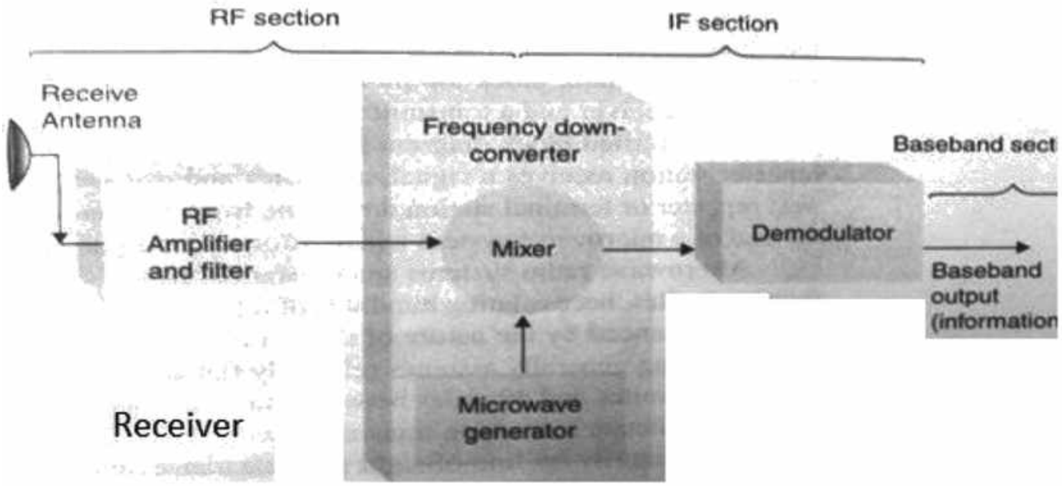
1. The electronic circuits used with microwave frequencies are more difficult to analyze.
2. Conventional components, such as resistors, inductors, and capacitors, are more difficult to manufacture and implement at microwave frequencies.
3. Microwave components are more expensive.
4. Transistor transit time is a problem with microwave devices.
5. Signal amplification is more difficult with microwave frequencies.

Microwave Radio Link

The following figure shows a simplex microwave radio link. The transmitter includes a modulator, mixer, and microwave generator and several stages of amplification and filtering.



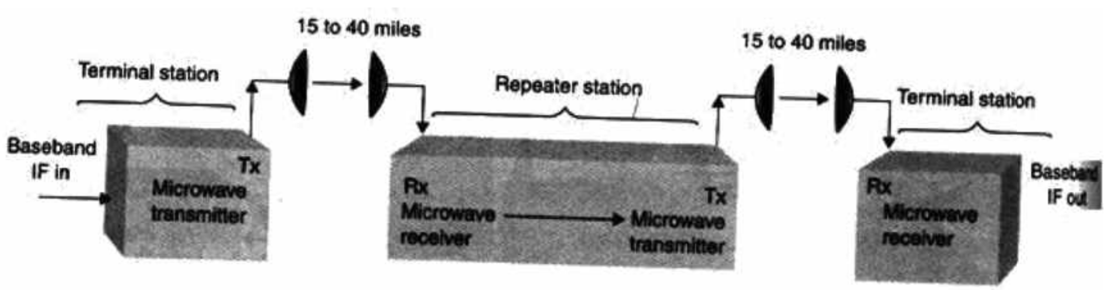
The modulator may perform frequency modulation or some form of digital modulation such as PSK or QAM. The output of modulator is an intermediate frequency (IF) carrier that has been modulated or encoded by the baseband input signal. The baseband signal is simply the information. The mixer and microwave generator (oscillator) combine to perform frequency up-conversion through nonlinear mixing. The up-converter is to translate IF frequencies to RF microwave frequencies.



The receiver consists of a radio-frequency (RF) amplifier, a frequency down-converter and a demodulator. The RF amplifier and filter increase the received signal level so that the down-converter can convert the RF signals to IF signals. The demodulator can be for FM, PSK or QAM. The output of demodulator is the original baseband (information) signals.

Microwave Radio Repeaters

With systems longer than 40 miles or when geographical obstructions block the transmission path, repeaters are needed. A microwave repeater is a receiver and transmitter placed back to back in the system.



The repeater station receives a signal, amplifies and reshapes it, and then retransmits it to the next repeater or terminal station down line from it. A terminal station is simply a station at the end of a microwave system where information signals originate and terminate.

Satellite Communication Systems

A satellite is a celestial body that orbits around a planet. In other terms, a satellite is a space vehicle launched by humans that orbits earth or another celestial body. Communication satellites are manmade satellites that orbit earth, providing a multitude of communications services to a wide variety of consumers, including military, governmental, private and commercial subscribers. The main purpose of communications satellite is to relay signals between two or more earth stations. A satellite repeater is called a transponder, and a satellite may have many transponders. Transmissions to and from satellites are categorized as either bus or payload. The bus includes control mechanisms that support the payload operation. The payload is the actual user information. Satellites utilize many of the same frequency bands as terrestrial microwave radio systems.

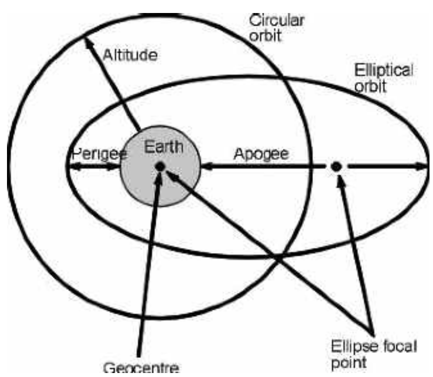
Satellite Elevation Categories

Satellites are generally classified as having a low earth orbit (LEO), medium earth orbit (MEO), or geosynchronous earth orbit (GEO).

- LEO satellites operate in the 1.0 GHz to 2.5 GHz frequency range. Main advantage is that the path loss between earth stations and space vehicles is much lower thereby resulting in lower transmit powers, smaller antennas and less weight. Example is Motorola's satellite-based mobile telephone system, Iridium.
- MEO satellites operate in the 1.2 GHz to 1.67 GHz frequency band and orbit between 6000 miles and 12,000 miles above earth. Example is DOD's satellite based global positioning system, NAVSTAR.
- Geosynchronous or geostationary satellites operate primarily in the 2 GHz to 18 GHz frequency spectrum with orbits 22,300 miles above the earth's surface. They orbit in a circular pattern with an angular velocity equal to that of earth and have an orbital time of approx 24 hours (i.e. same as earth).

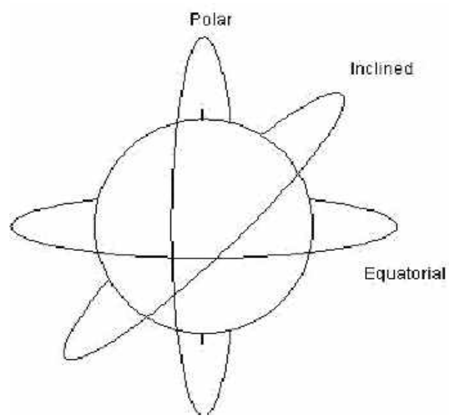
Satellite Orbits and Orbital Patterns

Satellites are classified as either synchronous or nonsynchronous. Synchronous satellites orbit earth above the equator with the same angular velocity as earth and therefore appear to be stationary and remain in the same location with respect to a given point on earth. Nonsynchronous satellites rotate around earth in circular or elliptical pattern as shown below.

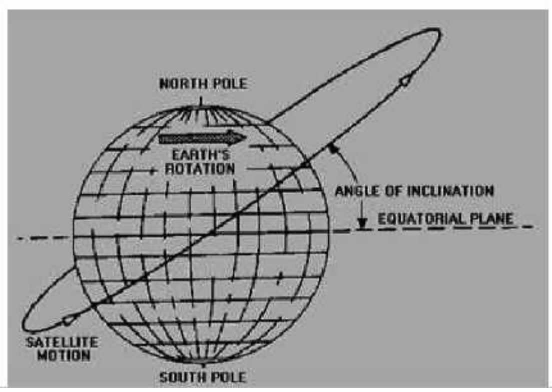


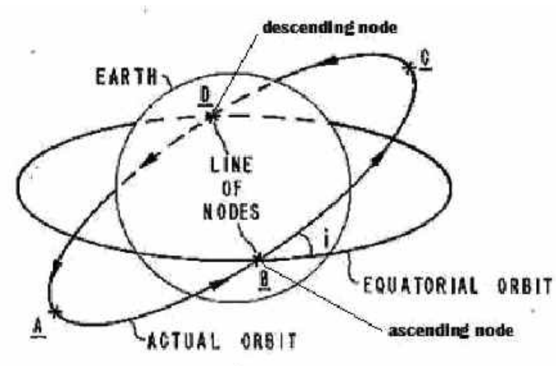
In circular orbit, the speed or rotation is constant. With elliptical orbits, the velocity of a satellite is greatest when satellite is closest to earth. The point in an elliptical orbit farthest from earth is called the ***apogee***, and the point on the orbit closest to earth is called ***perigee.*** If satellite rotation is in the same direction as earth's rotation with angular velocity greater than that of earth, the orbit is called **prograde or posigrade orbit**. If it's in opposite direction with angular velocity less than that of earth, then it's called a **retrograde orbit**. Nonsynchronous satellites revolve around in a prograde orbit, resulting in change of position continuously in respect to a fixed position on earth. So expensive and complicated tracking equipment is needed to locate and lock the antennas onto the satellite track.

Out of infinite number of orbital paths possible, only three are used for communication satellites: inclined, equatorial, or polar. When satellites orbit the Earth, either in a circular or elliptical orbit, the satellite orbit forms a plane that passes through the centre of gravity called **geocentre** of the Earth.



Inclined orbits are virtually all orbits except those that travel directly above the equator or directly above the North and South Poles.





The angle of inclination is the angle between the earth's equatorial plane and the orbital plane of a satellite measured counter clockwise at the point in the orbit where it crosses the equatorial plane from south to north and this point is called **ascending node**. If it's passing from north to south, it is called **descending node**. Angles of inclination vary between 0 degrees and 90 degrees. The line joining both these nodes through the center of earth is called **line of nodes**.

An *equatorial orbit* is when the satellite rotates in an orbit directly above the equator, usually in a circular path. With an equatorial orbit, the angle of inclination is 0 degrees. All geosynchronous satellites are in equatorial orbits. A *polar orbit* is when the satellite rotates in a path that takes it over the North and South Poles in an orbital pattern that is perpendicular to the equatorial plane. The angle of inclination of a satellite in a polar orbit is nearly 90 degrees. 100% of earth's surface can be covered with a single satellite in a polar orbit. Satellites in polar orbits rotate around earth in a longitudinal orbit while earth is rotating on its axis in a latitudinal rotation.

Geosynchronous Satellites

Also referred to as *geostationary,* it refers to the movement of communications satellites where the satellite circles the globe over the equator, in a movement that is synchronized with the earth's rotation. Because of this synchronization, the satellite appears to be stationary, and they also offer continuous operation in the area of visibility. Geosynchronous orbits are circular. There is only one geosynchronous earth orbit, which is occupied by a large number of satellites.

Geosynchronous orbit requirements: The most important requirement is that the orbit must have a 0-degree angle of elevation. They also must orbit in the same direction as earth's rotation with the same angular velocity. Using Kepler's third law, it can be shown that geosynchronous satellites must revolve around earth in a circular pattern 42,164 km from the center of the earth. The circumference of a geosynchronous satellite orbit is
C = 2 π (42,164km) = 264,790 km, and the velocity (v) is ***v* = 264,790 km/24hr = 6840 mph**

Clarke orbit: Synonymous with geostationary orbit. It is so-named because noted author Arthur C. Clarke was the first person to realize that this orbit would be useful for communication satellites. The Clarke orbit meets the concise setoff specifications for geosynchronous satellite orbits: (1) be located directly above the equator, (2) travel in the same direction as earth's rotation with a velocity of 6840 mph, (3) have an altitude of 22,300 miles above earth and (4) complete one revolution in 24 hours

Geosynchronous satellite advantages and disadvantages:
Some of the advantages are,

1. Geosynchronous satellites remain almost stationary in respect to a given earth station; therefore, expensive tracking equipment is not required at the earth stations.
2. Geosynchronous satellites are available to all earth stations within their *shadow* 100% of the time. The shadow of a satellite includes all the earth stations that have a line-of-sight path to the satellite.
3. Switching from one geosynchronous satellite to another as they orbit overhead is not necessary. Consequently, there are no transmission breaks due to switching times

Disadvantages are;

1. An obvious disadvantage of geosynchronous satellites is they require sophisticated and heavy propulsion devices on board to keep them in a fixed orbit.
2. High-altitude geosynchronous satellites introduce much longer propagation delays. The roundtrip propagation delay between two earth stations through a geosynchronous satellite is typically between 500 ms and 600 ms.
3. Geosynchronous satellites require higher transmit power levels and more sensitive receivers because of the longer distances and greater path losses.
4. High precision spacemanship is required to place a geosynchronous satellite into orbit and to keep it there.

Satellite Look Angles

Two angles have to be determined to ensure the earth station antenna is pointed directly at the satellite: the **azimuth and the elevation angle**. Both of them together are referred to as **look angles**. With geosynchronous satellites, the look angles of earth station antennas need to be adjusted only once, as the satellite will remain in a given position permanently except for minor variations. The pint on the surface of earth directly below the satellite is used to identify its location is called the **subsatellite point** (SSP) and for geosynchronous satellites, SSP must fall on the equator.

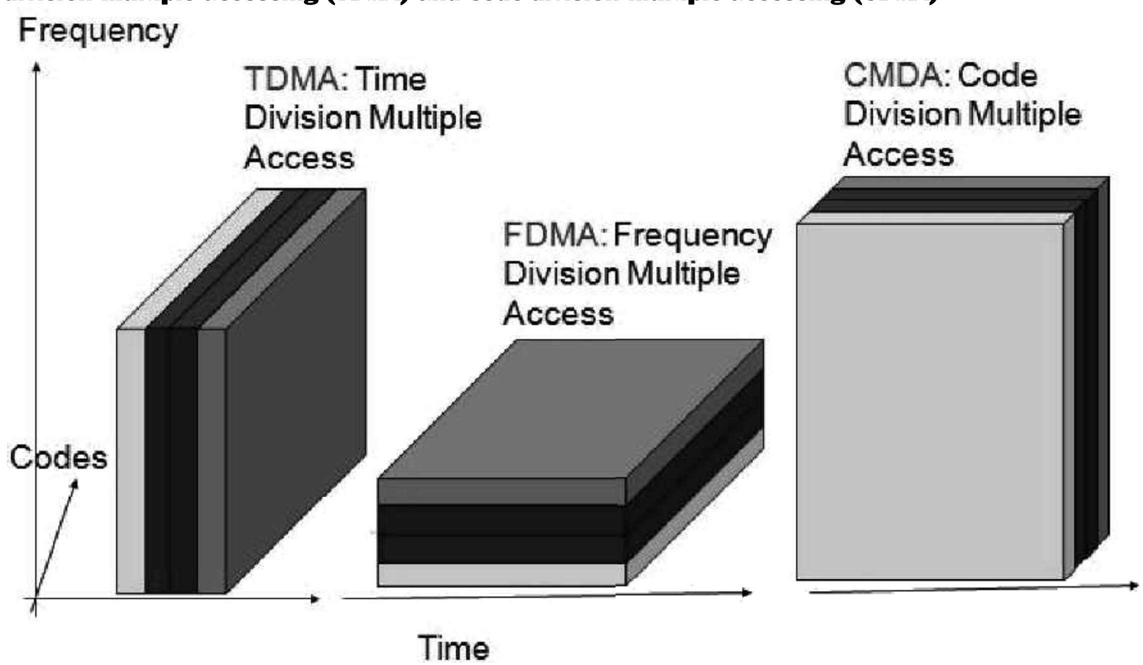
Satellite Antenna Radiation Patterns: Footprints

The geographical representation of the area on earth illuminated by the radiation from a satellite's antenna is called a *footprint* or sometimes a *footprint map.* In essence, a footprint of a satellite is the area on earth's surface that the satellite can receive from or transmit to. The shape of a satellite's footprint depends on the satellite's orbital path, height, and the type of antenna used. The higher the satellite, the more of the earth's surface it can cover.

The radiation pattern from a satellite's antenna is sometimes called a beam. The smallest and most directive beam is called a spot beam, followed by zonal beams, hemispherical beams, and earth (global) beams.

Satellite Multiple-Accessing Arrangements

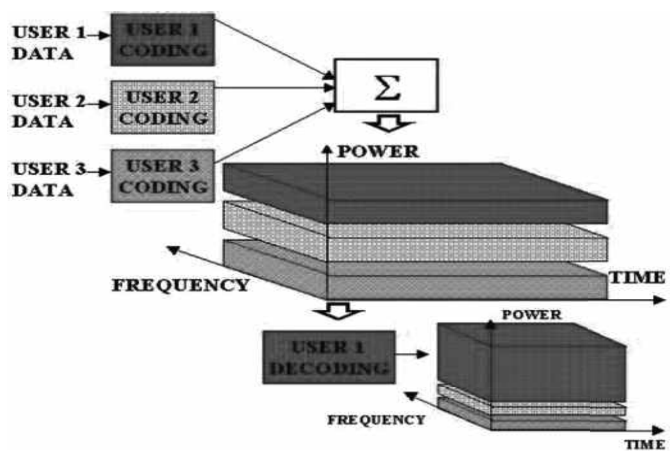
Satellite multiple accessing implies that more than one user has access to one or more transponders within a satellite's bandwidth allocation. The three most commonly used multiple accessing arrangements are **frequency division multiple accessing (FDMA), time-division multiple accessing (TDMA) and code-division multiple accessing (CDMA)**.



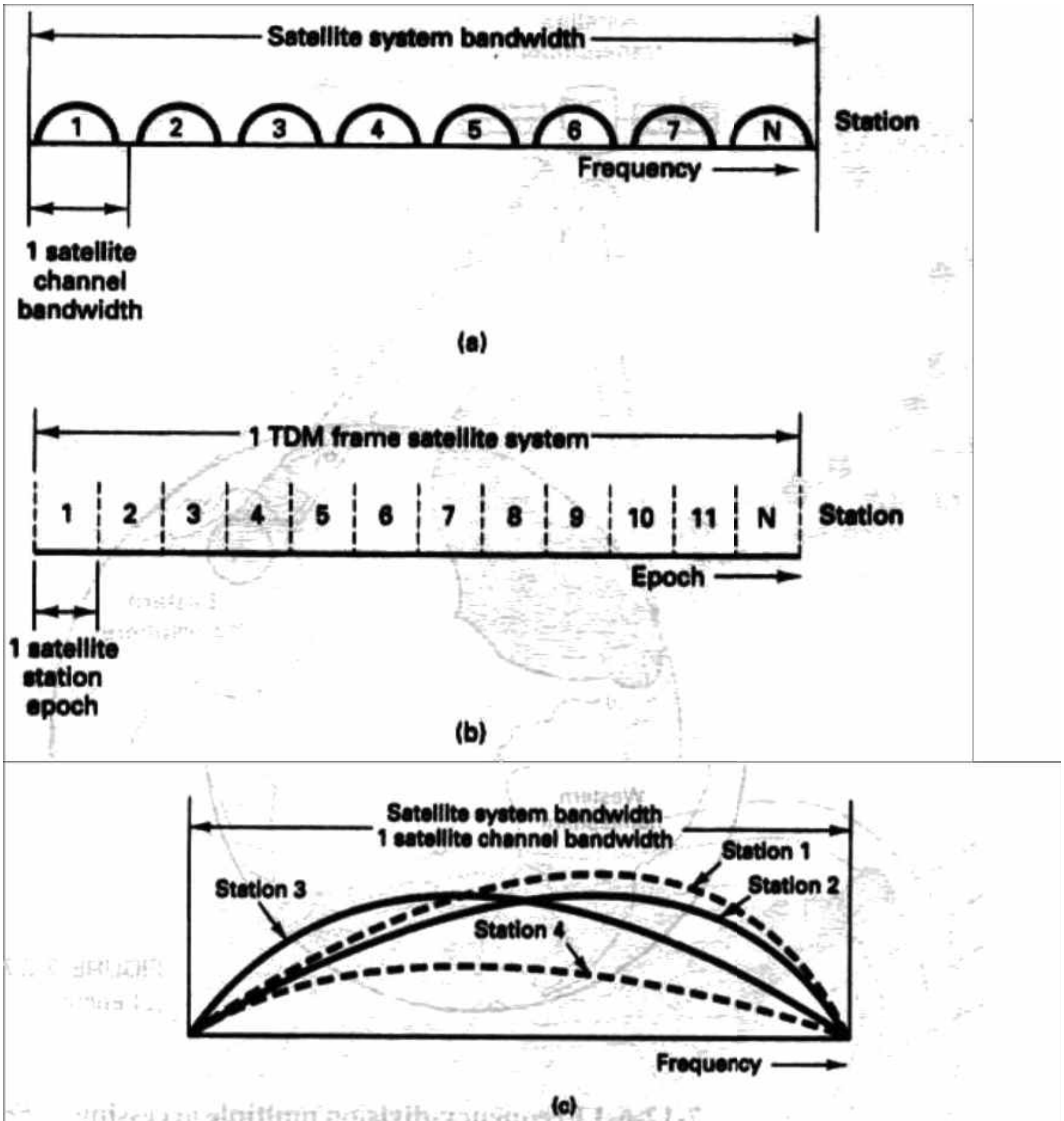
Frequency division multiple accessing: FDMA is a method of multiple accessing where a given RF bandwidth is divided into smaller frequency bands called subdivisions. FDMA transmissions are separated in the frequency domain and must share the total available transponder bandwidth as well as total transponder power. A control mechanism is used to ensure that two or more earth stations do not transmit in the same subdivision at the same time. Essentially, the control mechanism designates a receive station for each of the subdivisions. Thus, with FDMA, transmission can occur from more than one station at the same time, but the transmitting stations must share the allocated power, and no two stations can utilize the same bandwidth.

Time-division multiple accessing: TDMA is the predominant multiple-accessing method used today. TDMA is a method of time-division multiplexing digitally modulated carriers between participating earth stations within a satellite network using a common satellite transponder. With TDMA, each earth stat ion transmits a short burst of information during a specific time slot within a TDMA frame. The bursts must be synchronized so that each station's burst arrives at the satellite at a different time, thus avoiding a collision with another station's carrier. TDMA transmissions are separated in the tune domain, and with TDMA, the entire transponder bandwidth and power are used for each transmission but for only a prescribed interval of time. Thus, with TDMA, transmission cannot occur from more than one station at the same time. However, the transmitting station can use all the allocated power and the entire bandwidth during its assigned time slot.

Code-division multiple accessing: CDMA is based on the use of modulation technique known as **spread spectrum.** Users are separated both by frequency and time.



Because there are no limitations on bandwidth, CDMA is sometimes referred to as *spread-spectrum multiple accessing* (SSMA). With CDMA, all earth stations transmit within the same frequency band and, for all practical purposes, have no limitations on when they may transmit or on which carrier frequency. Thus, with CDMA, the entire satellite transponder bandwidth is used by all stations on a continuous basis. Signal separation is accomplished with envelope encryption/decryption techniques



Multiple-accessing arrangements: (a) FDMA; (b) TDMA; (c) CDMA

Comparison: In both FDMA and TDMA, only one subscriber at a time is assigned to a channel. No other conversion can access this channel until the subscriber's call is finished or until that original call to be handed off to a different channel by the system. Voice data tends to be burst in nature. So much of the time, no data is being sent over the channel. This inefficiency tends to limit the capacity of the system. The above drawbacks are overcome in this third technique in which the users are spread across both frequency and time in the same channel. This is a hybrid combination of FDMA and TDMA. For example, *frequency hopping* may be employed to ensure during each successive time slot, the frequency bands assigned to the users are recorded in random manner. An important advantage of CDMA over FDMA and TDMA is that it can provide for secure communication.

Questions

1. What is a radio wave? What are the optical properties of radio waves? Explain all the details of how they relate to radio wave propagation?

2. What is meant by a free space path loss of an electromagnetic wave? Give the mathematical equation in decibel form. Determine, in dB, the free space path loss for a frequency of 6 GHz travelling a distance of 50 km.

3. What are the three modes of terrestrial propagation of electromagnetic waves? Explain.

4. What is a satellite multiple accessing arrangement? List and describe, in detail with neat diagrams, the three forms of satellite multiple accessing arrangements.

5. Explain the term skip distance, satellite footprint and give the advantages of geosynchronous satellites

6. List the advantages and disadvantages of microwave communications over cable transmission facilities.

7. Compare FDMA, TDMA and CDMA

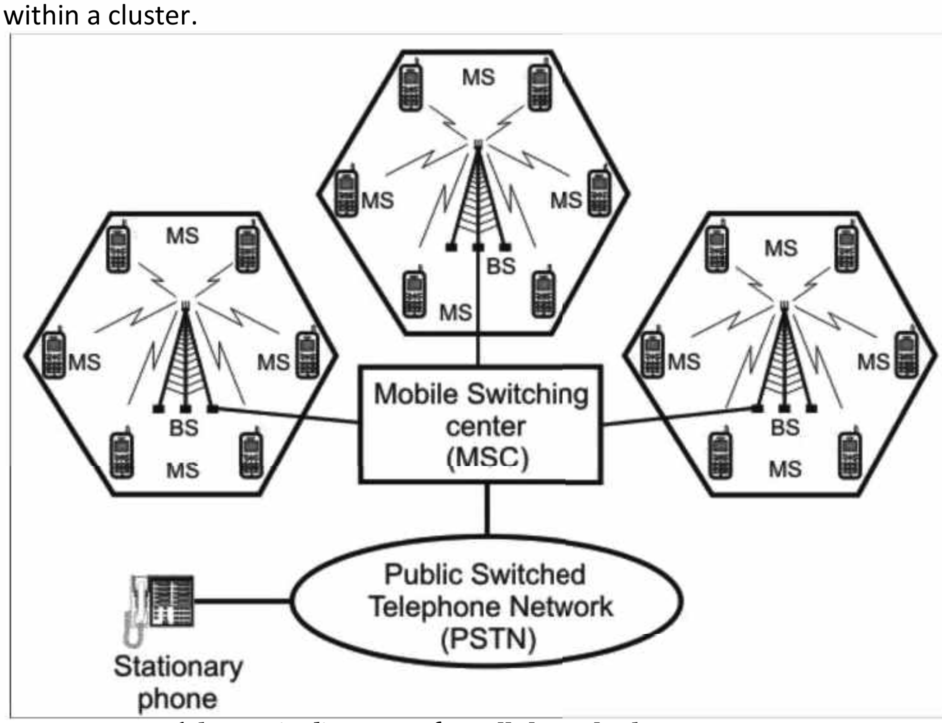
CELLULAR TELEPHONE SYSTEMS

First- Generation Analog Cellular Telephone, Personal Communications system, Second-Generation Cellular Telephone Systems, N-AMPS, Digital Cellular Telephone, Interim Standard, North American Cellular and PCS Summary, Global system for Mobile Communications, Personal Communications Satellite System

Introduction

Cellular system was developed to provide mobile *telephony*: *telephone* access "anytime, anywhere." Cellular telephony is a system-level concept, which replaces a single high power transmitter with a large number of low-power transmitters for communication between any two devices over a large geographic area. Primary goal of the cellular telephone network is to provide wireless communication between two moving devices, called *mobile stations* or between one mobile unit and a stationary unit, commonly referred to as *land-line* unit. To accommodate a large number of users over a large geographic area, the cellular telephone system uses a large number of low-power wireless transmitters to create *cells*. Variable power levels allow cells to be sized according to subscriber density and demand within a particular region.

A **cell** is a basic geographic unit of a cellular system. The term cellular comes from the honeycomb shape of the areas into which a coverage region is divided. Cells are base stations transmitting over small geographic areas that are represented as hexagons. As mobile users travel from cell to cell, their conversations are handed off between *cells*. Channels (frequencies) used in one cell can be reused in another cell some distance away, which allows communication by a large number stations using a limited number of radio frequencies. To summarize, the basic concept of reuse allows a fixed number of channels to serve an arbitrarily large number of users. A cluster is a group of cells and no channels are reused within a cluster.

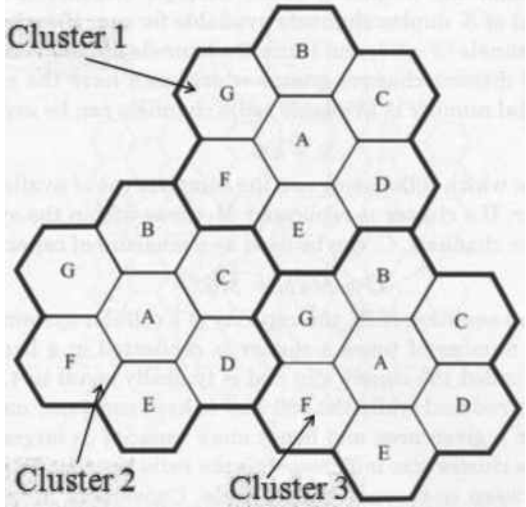


Schematic diagram of a cellular telephone system

As shown above, a cellular system comprises of the following basic components:

- **Mobile Stations (MS):** Mobile handsets, which is used by an user to communicate with another user
- **Cell**: Each cellular service area is divided into small regions called cell (5 to 20 Km)
- **Base Stations (BS):** Each cell contains an antenna, which is controlled by a small office.
- **Mobile Switching Center (MSC)**: Each base station is controlled by a switching office, called mobile switching center

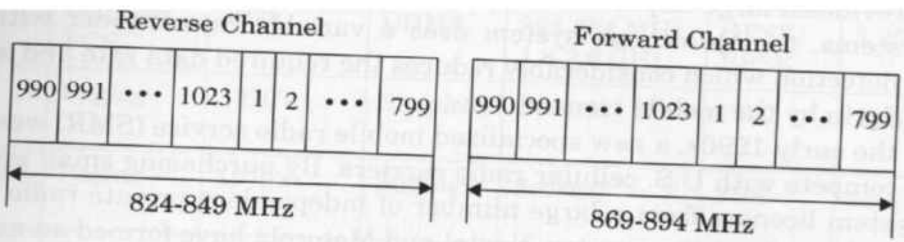
Frequency reuse is the process in which the same set of frequencies (channels) can be allocated to more than one cell, provided the cells are separated by sufficient distance. The figure shows a geographic cellular radio coverage area containing three groups of cells called clusters. Each cluster has seven cells in it, and all cells are assigned the same number of full-duplex cellular telephone channels. Cells with the same letter use the same set of channel frequencies. A, B, C, D, E, F and G denote the seven sets of frequencies.



Handoff: At any instant, each mobile station is logically in a cell and under the control of the cell's base station. When a mobile station moves out of a cell, the base station notices the MS's signal fading away and requests all the neighbouring BSs to report the strength they are receiving. The BS then transfers ownership to the cell getting the strongest signal and the MSC changes the channel carrying the call. The process is called *handoff*. There are two types of handoff; Hard Handoff and Soft Handoff. In a *hard handoff*, which was used in the early systems, a MS communicates with one BS. As a MS moves from cell A to cell B, the communication between the MS and base station of cell A is first broken before communication is started between the MS and the base station of B. As a consequence, the transition is not smooth. Hard handoff is often called as *break before-make*. Hard handoffs are intended to be instantaneous in order to minimize the disruption to the call. A hard handoff is perceived by network engineers as an event during the call. For smooth transition from one cell (say A) to another (say B), an MS continues to talk to both A and B. As the MS moves from cell A to cell B, at some point the communication is broken with the old base station of cell A. This is known as *soft handoff* (also called as make before break). A soft handoff may involve using connections to more than two cells, e.g. connections to three, four or more cells can be maintained by one phone at the same time. Softer handoffs are possible when the cells involved in the handoff have a single cell site.

First Generation Analog Cellular Telephone

AMPS (Advanced Mobile Telephone System) was invented at Bell Labs and initially deployed in the U.S. in the early 1980's. The frequencies allocated to AMPS by the Federal Communications Commission (FCC) range between 824 to 849 MHz in reverse channels (mobile to base) and 869 to 894 MHz in forward channels (base to mobile). Simultaneous transmission in both directions in a transmission mode is called full duplex (FDX) or simply duplexing. Frequency-division duplexing (FDD) is used with AMPS and occurs when two distinct frequency bands are provided to each user. A special device called duplexer is used in each mobile unit and base station to allow simultaneous transmission and reception on duplex channels. Transmissions from base stations to mobile units are called forward links, whereas transmissions from mobile units to base stations are called reverse links. In 1989, the FCC added an additional 10-MHz frequency spectrum to the original 40-MHz band, which increased the simplex channels to a total of 832 (416 full duplex).



	Channel Number	Center Frequency (MHz)
Reverse Channel	$1 \le N \le 799$	$0.030N + 825.0$
	$990 \le N \le 1023$	$0.030(N - 1023) + 825.0$
Forward Channel	$1 \le N \le 799$	$0.030N + 870.0$
	$990 \le N \le 1023$	$0.030(N - 1023) + 870.0$

(Channels 800 - 989 are unused)

Complete AMPS Frequency spectrum

The 832 channels are divided into four categories:

1. Control (base to mobile) to manage the system.
2. Paging (base to mobile) to alert mobile users to calls for them.
3. Access (bidirectional) for call setup and channel assignment.
4. Data (bidirectional) for voice, fax, or data.

Each physical channel is 30 kHz wide and is dedicated to a single mobile station for the duration of the call while the mobile is in the current cell. Each call uses a dedicated forward channel paired with a dedicated reverse channel at a 45 MHz offset. Some of the channel pairs (21 of them) are used for control purposes in the AMPS environment. Analog *frequency modulation (FM)* with 8 kHz deviation is used in the *traffic channels*, which convey voice conversations. Binary *frequency shift keying (FSK)* at 10 kbps-a digital modulation technique-is used in the *control channels* used for signalling.

AMPS Identification Codes: The AMPS system uses several identification codes for each mobile unit. The **mobile identification number** (MIN) is a 34-bit binary code, which is the programmed handset phone number used to call the subscriber. This programmed identifier is associated with the subscriber and is stored in erasable non-volatile memory in the handset. The second identifier is the **electronic serial number (*ESN*)**, which is a manufactured characteristic of the mobile unit. This identifier is permanent and associated with the physical equipment. It is 32 bits in length, with the first 8 bits identifying the manufacturer. The third identification code is four bit **station class mark (SCM)**, which indicates whether the terminal has access to all 832 channels or not. The SCM also specifies the maximum radiated power for the unit. The **system identifier (SID)** is a 15-bit binary code issued by the FCC to an operating company when it issues a license to provide AMPS cellular service to an area. Local operating companies assign a tow-bit digital **color code (DCC)** and a **supervisory audio tone (SAT)** to each of their base stations to help the mobile units distinguish one base station from a neighbouring base station.

AMPS Control Channels: Control channels are used in cellular telephone systems to enable mobile units to communicate with the cellular network through base stations and are used for call origination, call termination, and to obtain system information. With AMPS system, voice channels are analog FM, while control channels are digital and employ FSK. Base stations broadcast on the forward control channel (FCC) and listen on the reverse control channel (RCC). All AMPS base stations continuously transmit FSK data on the FCC so that idle cellular telephones can maintain a lock on the strongest FCC regardless of their location. A subscriber's unit must be locked on an FCC before it can originate or receive calls.

Personal Communications System

The FCC defines PCS mobile telephone as "a family of mobile or portable radio communications services, which provides services to individuals and business and is integrated with a variety of competing networks". PCS is North American implementation of the European GSM standard. Differences between PCS systems and standard cellular telephone systems generally include but are certainly not limited to the following:

(1) smaller cell size, (2) all digital and (3) additional features. Cellular systems generally classified as PCS include IS-136 TDMA, GSM and IS-95 CDMA.

The fundamental concept of PCS is to assign each mobile unit a PTN that is stored in a database on the SS7 common signalling network. The database keeps track of where mobile units are. When a call is placed for a mobile unit, the SS7 artificial intelligence network determines where the call should be directed. The PCS network is similar to

D-AMPS system in that the MTSO stores three essential-databases: home location register, visitor location register, and equipment identification registry.

The HLR is a database that stores information about the user, including home subscription information and also the supplementary services like call waiting, call hold, call forwarding etc subscribed by the user. The VLR stores information about subscribers in a particular MTSO serving area, such as whether the unit is on or off and whether any of the supplementary services are activated or deactivated. The EIR stores information pertaining to the identification and type of equipment that exists in the mobile unit. The EIR also helps the network identify stolen or fraudulent mobile units.

Some of the services offered by PCS systems are:

- Available mode: It allows all calls to pass through the network to the subscriber except for a minimal number of telephone numbers that can be blocked.
- Screen mode: It is PCS equivalent to caller ID. The calling party's name appears on the mobile units display allowing users to screen calls. Unanswered calls are automatically forwarded to a forwarding destination specified by the subscriber.
- Private mode: Here, all calls except those specified by the subscriber are automatically forwarded to a forwarding destination without ringing the subscriber's handset.
- Unavailable mode: no calls are allowed to pass through to the subscriber. So, all calls are automatically forwarded to a forwarding destination.

The primary disadvantage of PCS is network cost. Employing small cells requires using more base stations, which equates to more transceivers, antennas, and trunk circuits. PCS networks rely extensively on the SS7 signalling network for interconnecting to other telephone networks and databases.

N-AMPS

Narrowband Advanced Mobile Phone Service (NAMPS) is an improved version of AMPS systems. NAMPS is a cellular call-handling system that uses digital signalling techniques to split the existing 30 kHz wideband voice channels into three 10 kHz narrowband voice channels. Each 10-KHz subchannel is capable of handling its own calls. The result is three times more voice channel capacity than the traditional AMPS system provides.

With narrow bandwidths, voice channels are more vulnerable to interference than standard AMPS channels and would require a higher frequency reuse factor. This is compensated for, with the addition of an interference avoidance scheme called **Mobile Reported Interference (MRI)**, which uses voice companding to provide synthetic voice channel quieting. NAMPS cellular phones are manufactured for dual mode operation, and they are compatible with traditional AMPS systems. N-AMPS systems use standard AMPS control channels for call setup and termination. N-AMPS mobile units are capable of utilizing **four types of handoffs**: *wide channel to wide channel (30 kHz to 30 kHz), wide channel to narrow channel (30 kHz to 10 kHz), narrow channel to narrow channel (10 kHz to 10 kHz) and narrow channel to wide channel (10 kHz to 30 kHz).* To conclude, with N-AMPS, user capacity can be expanded by subdividing existing channels (band splitting), partitioning cells into smaller subcells (cell splitting), and modifying antenna radiation patterns (sectoring).

Digital Cellular Telephone

With the rapidly expanding customer base while working with unchanged allocated frequency spectrum, it was a growing problem for the cellular companies. Digital cellular telephone systems have several inherent advantages over analog cellular telephone systems, including better utilization of bandwidth, more privacy and incorporation of error detection and correction. Consequently, the ***United States Digital Cellular (USDC)*** system was designed and developed with the intent of supporting a higher user density within a fixed-bandwidth frequency spectrum. Cellular telephone systems that use digital modulation, such as USDC, are called digital cellular. USDC cellular systems comply with IS-54, which specifies dual-mode operation and backward compatibility with standard AMPS and because of this reason, they are also known as ***Digital AMPS (D-AMPS or DAMPS).*** The USDC system has an additional frequency band in the 1.9 GHz that is not compatible with AMPS frequency allocation.

Time-Division Multiple Accessing

USDC uses time-division multiple accessing (TDMA) as well as FDMA. However, TDMA allows more than one mobile unit to use a channel at the same time by further dividing transmissions within each cellular channel into time slots, one for each mobile unit using that channel. Unlike AMPS FDMA systems, with USDC TDMA systems, mobile-unit subscribers can only hold a channel while they are actually talking on it. During pauses or other normal breaks in a conversation, users must relinquish their channel so that other mobile units can use it. This time sharing technique significantly increases the capacity of a system, allowing more mobile-unit subscribers to use a system at virtually the same time within a geographical area.

A USDC TDMA transmission frame consist of six equal-duration time slots enabling each 30-kHz AMPS channel to support three full-rate or six half-rate users. Hence USDC offers as much as six times the channel capacity as AMPS. The advantages of digital TDMA multiple-accessing systems over analog AMPS FDMA systems are given below:

1. Time domain multiple accessing allows for a threefold to sixfold increase in the number of mobile subscribers using a single cellular channel.

2. Digital signals are much easier to process than analog signals as most of the modern modulation techniques are developed to be used in a digital environment.
3. Digital signals (bits) can be easily encrypted and decrypted, safeguarding against eavesdropping.
4. The entire telephone system is compatible with other digital formats, such as those used in computers and computer networks.
5. Digital systems inherently provide a quieter (less noisy) environment than their analog counterparts.

EIA/TIA Interim Standard 54

In 1990, the Electronics Industries Association and Telecommunications Industry Association (EIA/TIA) standardized the dual-mode USDC/AMPS system as Interim Standard 54 (IS-54), Cellular Dual Mode Subscriber Equipment. Using IS-54, a cellular telephone carrier could convert any or all of its existing analog channels to digital. To achieve dual-mode operation, IS-54 provides digital control channels and both analog and digital voice channels. Dual-mode mobile units can operate in either the digital or the analog mode for voice and access the system with the standard AMPS digital control channel. IS-54 specifies a 48.6 kbps rate per 30-kHz voice channel divided among three simultaneous users. Each user is allocated 13 kbps, and the remaining 9.6 kbps is used for timing and control overhead.

USDC Control Channels and IS-136.2

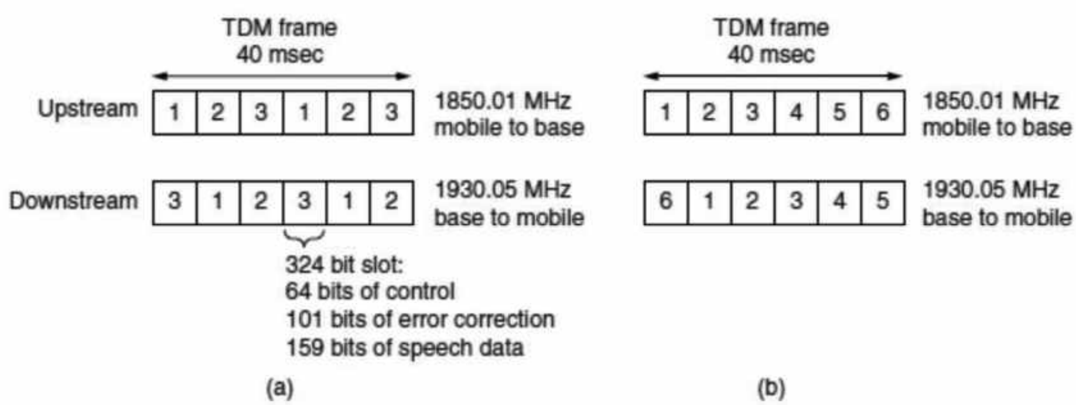
The IS-54 USDC standard specifies the same 42 primary control channels as AMPS and 42 additional control channels called secondary control channels. So USDC offers twice as many control channels as AMPS and is therefore capable of providing twice the capacity of control traffic within a given market area. To maintain compatibility with existing AMPS cellular telephone systems, the primary forward and reverse control channels in USDC cellular systems use the same signalling techniques and modulation scheme (FSK) as AMPS. However, a new standard IS-136.2 replaces FSK with π/4 DQPSK modulation for the 42 dedicated USDC secondary control channels, allowing digital mobile units to operate entirely in the digital domain. The IS-136.2 standard is called North American- Time Division Multiple Accessing (NA-TDMA). IS 136 was developed to provide a host of new features and services. An additional “sleep mode” which conserves power is also provided.

The IS-54 standard specifies three types of channels: analog control channels, analog voice channels, and a 10-kbps binary FSK digital control channel (DCCH). The IS-136 standard provides the above three channels and an additional one: a digital control channel with a signalling rate of 48.6 kbps on USDC-only control channels. The new digital control channel includes several logical channels with different functions, including the random access channel (RACH), the SMS point-to-point, paging, and access response channel (SPACH); the broadcast control channel (BCCH) and the shared channel feedback (SCF) channel.

- RACH: It is a unidirectional channel used by mobile units to request access to the cellular telephone system.
- SPACH: It is used to transmit information from base to specific mobile station and information transmitted on SPACH channel includes three separate logical subchannels: SMS point-to-point messages, paging messages, and access response messages.
- BCCH: It is an acronym referring to the F-BCCH, E-BCCH and S-BCCH. The fast broadcast channel (F-BCCH) broadcasts digital control channel (DCCH) structure parameters. Mobile units use F-BCCH information when initially accessing the system to determine the beginning and ending of each logical channel in the DCCH frame. The extended broadcast control channel (E-BCCH) carries information about neighbouring analog and TDMA cells and optional messages, such as emergency information, time and date messaging etc. The SMS broadcast channel (S-BCCH) is a logical channel used for sending short messages to individual mobile units.
- SCF: It is used to support random access channel operation by providing information about which time slots the mobile unit can use for access attempts and also if a mobile unit's previous RACH transmission was successfully received.

USDC Digital Voice Channel

Like AMPS, each USDC voice channel is assigned a 30-kHz bandwidth on both the forward and the reverse link. With USDC, each channel can support as many as three full-rate mobile users simultaneously by using digital modulation and a TDMA format called North American Digital Cellular (NADC). Each radio-frequency voice channel in the total AMPS FDMA frequency band consists of one 40-ms TDMA frame comprised of six time slots containing 324 bits each. The average cost per subscriber per base station equipment is lower with TDMA since each base station transceiver can be shared by upto six users at a time.



(a) A D-AMPS channel with three users. (b) A D-AMPS channel with six users.

E-TDMA: General Motors Corporation implemented a TDMA scheme called E-TDMA {Extended or Enhanced TDMA}, which incorporates six half-rate users transmitting at half the bit rate of standard USDC TDMA systems. E-TDMA systems also incorporate digital speech interpolation (DSI) to dynamically assign more than one user to a time slot, deleting silence on the calls. Consequently E-TDMA can handle approximately 12 times the user traffic as standard AMPS systems and four times that of systems complying with IS-54.

Each time slot in every USDC voice-channel frame contains four data channels-three for control and one for digitized voice and user data. The full-duplex digital traffic channel (DTC) carries digitized voice information and consists of a reverse digital traffic channel (RDTC) and a forward digital traffic channel (FDTC) that carry digitized speech information or user data. The three supervisory channels are given below:

- *Coded digital verification color code (CDVCC)*: Its purpose is to provide co-channel identification similar to the SAT signal transmitted in the AMPS system. It is a 12 bit message transmitted in every time slot.
- *Slow associated control channel (SACCH)*: It is a signalling channel for transmission of control and supervision messages between the digital mobile unit and the base station while the mobile unit is involved with a call. It is also used by the mobile unit to report signal strength measurements of neighbouring base stations, so when needed the base station can initiate a mobile-assisted handoff (MAHO).
- *Fast associated control channel (FACCH)*: It is a second signalling channel for transmission of control and specialized supervision and traffic messages between the base station and the mobile units. It is a blank-and-burst type of transmission than when transmitted replaces digitized speech information with control and supervision messages within a subscriber's time slot.

USDC Digital modulation scheme

To achieve a transmission bit rate of 48.6 kbps in a 30-kHz AMPS voice channel, a bandwidth (spectral) efficiency of 1.62 bps/Hz is required, binary FSK is incapable. USDC voice and control channels use a symmetrical differential, phase-shift keying technique known as π/4 DQPSK or π/4 differential quadriphase shift keying, which offers several advantages such as improved co-channel rejection and bandwidth efficiency. In π/4 DQPSK modulator, data bits are split into two parallel channels that produce a specific phase shift in the analog carrier, and since there are four possible bit pairs, there are four possible phase shifts using a quadrature I/Q modulator and the four phase changes are π/4, - π/4, 3 π/4 and -3 π/4, which define eight possible carrier phases. Using pulse shaping with π/4 DQPSK allows for the simultaneous transmission of three separate 48.6-kbps speech signals in a 30-kHz bandwidth.

Interim Standard 95

Interim Standard 95 (IS-95) is the first CDMA-based digital cellular standard by Qualcomm. The brand name for IS-95 is **cdmaOne**. IS-95 is also known as TIA-EIA-95. CDMA allows users to differentiate from one another by a unique code rather than a frequency or time assignment and hence has several advantages over TDMA and FDMA cellular systems such as increased capacity, improved performance and reliability. IS-95 is designed to be compatible with existing analog systems (AMPS).

CDMA

IS-95 specifies a direct-sequence, spread spectrum CDMA system and does not follow the channelization principles of traditional cellular radio communications systems. Rather than dividing the allocated frequency spectrum into narrow bandwidth channels, one for each user, information is transmitted (spread) over a very wide frequency spectrum with as many as 20 mobile subscriber units using the same carrier frequency within the same frequency band. IS-95 is not asymmetrical as it specifies a different modulation and spreading technique for the forward (digital QPSK) and reverse (digital OQPSK) channels. On the forward channel, the base station simultaneously transmits user data from all current mobile units in that cell by using different spreading sequences (codes) for each user's transmissions. A pilot code is transmitted with the user data at a higher power level, thus allowing all mobile units to use coherent detection. On the reverse link, all mobile units respond in an asynchronous manner (i.e. no time or duration limitations) with a constant signal level controlled by the base station. The speech coder used with IS-95 is the Qualcomm 9600-bps Code-Excited Linear Predictive (QCELP) coder. The vocoder converts an 8-kbps compressed data stream to a 9.6 kbps data stream.

Advantages of CDMA:

- Frequency diversity – frequency-dependent transmission impairments have less effect on signal
- Multipath resistance – chipping codes used for CDMA exhibit low cross correlation and low autocorrelation
- Privacy – privacy is inherent since spread spectrum is obtained by use of noise-like signals
- Graceful degradation – system only gradually degrades as more users access the system

Limitations of CDMA

- Self-jamming – arriving transmissions from multiple users not aligned on chip boundaries unless users are perfectly synchronized
- Near-far problem – signals closer to the receiver are received with less attenuation than signals farther away

- Soft handoff – requires that the mobile acquires the new cell before it relinquishes the old; this is more complex than hard handoff used in FDMA and TDMA schemes

CDMA frequency and channel allocations

Each IS-95 channel is allocated a 1.25-MHz frequency spectrum for each one-way CDMA communications channel. A single CDMA radio channel takes up the same bandwidth as approximately 42 30-kHz AMPS voice channels. But because of the frequency reuse advantage of CDMA, CDMA offers approximately a 10-to-1 channel advantage over standard analog AMPS and a 3-to-1 advantage over USDC digital AMPS. Each CDMA channel is 1.23 MHz wide with a 1.25 MHz frequency separation between adjacent carriers, producing a 200-kHz guard band between CDMA channels. There are as many as nine CDMA carriers available for A and B band operator in the AMPS frequency spectrum.

For the forward (uplink) channel, subscriber data are encoded using convolutional coding with rate ½, interleaved and spread by one of 64 orthogonal Walsh codes. For the downlink channels, a different spreading strategy is used as each mobile unit's received signal takes a different transmission path and therefore, arrives at the base station at a different time. A convolutional coding rate of 1/3 is used and long sequences are used to separate the signals from different users on the reverse link (CDMA).

Each mobile unit in a given cell is assigned a unique spreading sequence that ensures near perfect separation among the signals from different subscriber units and allows transmission differentiation between users. All signals in a particular cell are scrambled using a pseudorandom sequence of length 2^{15} chips. This reduces radio frequency interference between mobiles in neighbouring cells that may be using the same spreading sequence and provides the desired wideband spectral characteristics. Two commonly used techniques for spreading the spectrum are frequency hopping and direct sequencing.

Frequency-hopping spread spectrum: FH – CDMA is a kind of spread spectrum technology that enables many users to share the same channel by employing a unique hopping pattern to distinguish different users' transmission. The type of spread spectrum in which the carrier hops randomly from one frequency to another is called FH spread spectrum. A common modulation format for FH system is that of M-ary frequency shift keying (MFSK).the combination is referred to as FH/MFSK.

A major advantage of frequency hopping is that it can be implemented over a much larger frequency band than it is possible to implement DS- spreading, and the band can be non-contiguous. Another major advantage is that frequency hopping provides resistance to multiple – access interference while not requiring power control to prevent near – far problems. The sequence in which the frequencies are selected must be known by both the transmitter and the receiver prior to the beginning of the transmission. Each transmitter in the system has a different hopping sequence to prevent one subscriber from interfering with transmissions from other subscribers using the same radio channel frequency.

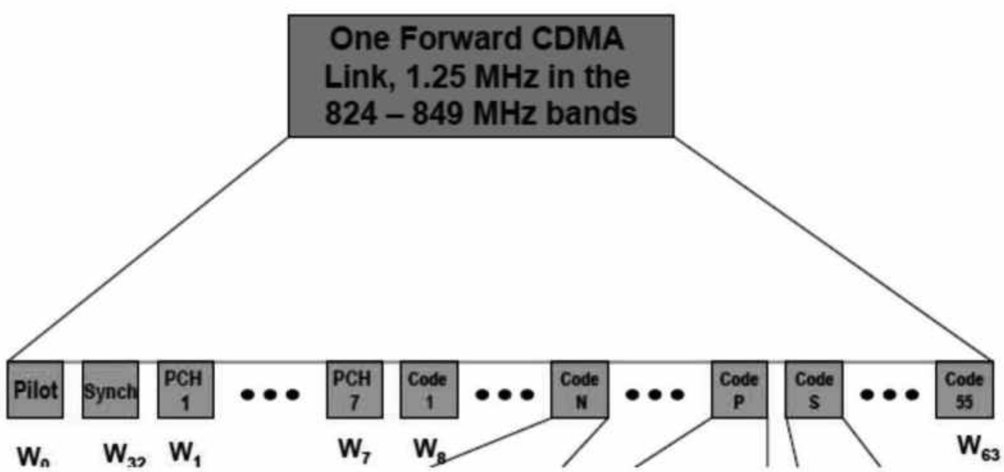
Direct-sequence spread spectrum: Here, a high-bit-rate pseudorandom code is added to a low-bit-rate information signal to generate a high-bit-rate pseudorandom signal closely resembling to noise that contains both the original data signal and the pseudorandom code. The code has to known both to the transmitter and the intended receiver. The receiver upon detection of direct-sequence transmission, simply subtracts the pseudorandom signal from the composite receive signal to extract the information data.

Adding a high bit-rate pseudorandom signal to the voice information makes the signal more dominant and less susceptible to interference, allowing lower power transmission and hence, a lower number of transmitters and less expensive receivers.

CDMA Traffic Channels

CDMA traffic channels consist of a downlink (base station to mobile unit) channel and an uplink (mobile station to base station) channel. The downlink traffic channel consists up to 64 channels, including a broadcast channel used for control and traffic channels used to carry subscriber information.

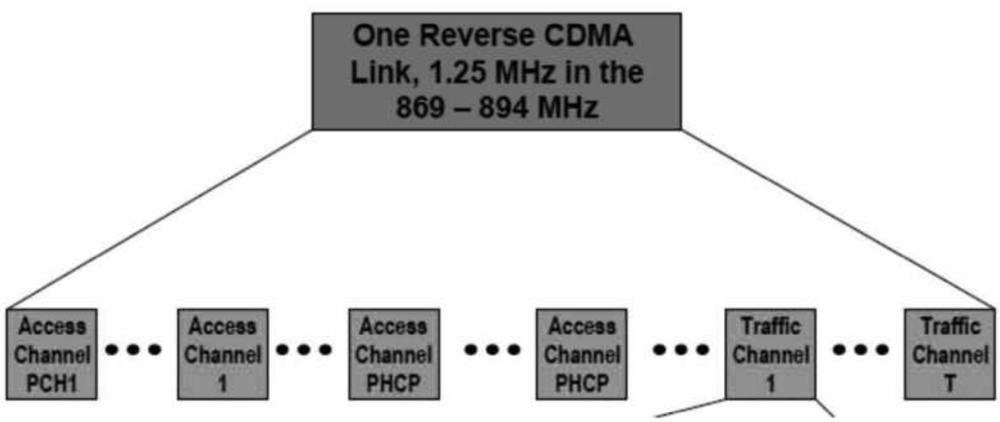
The forward link uses the same frequency spectrum as AMPS (824-849 MHz). Four types of logical channel are present i.e. a pilot, a synchronization, 7 paging, and upto 63 traffic channels. All these channels share the same 1.25-MHz CDMA frequency assignment. The traffic channels are identified by a distinct user-specific long-code sequence, and each access channel is identified by a distinct access channel long-code sequence.



The pilot channel is included in every cell with the purpose of providing a signal for the receiver to use to acquire timing and provide a phase reference for coherent demodulation. It is also used by mobile units to compare signal strengths between base stations to determine when a handoff should be initiated. The pilot contains no information but it is the strongest signal on the forward link, containing at least 20% of the total power on the forward link. The synchronization channel uses a Walsh W32 code and same pseudorandom sequence and phase offset as the pilot channel, allowing it to be demodulated by any receiver that can acquire the pilot signal. The synchronization channel broadcasts synchronization messages to mobile units and operates at 1200 bps. Once the mobile is synchronized with the base station the sync channel is ignored.

The paging channels are used to transmit overhead information (i.e. commands and pages) to the mobile. When a call is being set up the commands and traffic channel assignment are sent on the paging channel. Once a traffic channel is established the paging channel is ignored by the mobile. Paging channels are optional and can range in number between zero and seven. A single 9600-bps pilot channel can typically support about 180 pages per second for a total capacity of 1260 pages per second. Data on the downlink traffic channel are grouped into 20-ms frames. The data are first convolutionally coded and then formatted and interleaved to compensate for differences in the actual data rates. The resulting signal is spread with Walsh code with a long pseudorandom sequence at a rate of 1.2288 Mchips/s.

The uplink radio channel transmitter consists of access channels and upto 62 uplink traffic channels. The access channel is used by the mobile when not assigned to a traffic channel. The access channels is used by the mobile to register with the network, originate calls, respond to pages and commands from the base station, and transmit overhead messages to the base station. Typical access channel messages include acknowledgements and sequence number, mobile identification parameter messages and authentication parameters. The access channel is a random access channel with each channel subscriber uniquely identified by their pseudorandom codes.



The uplink traffic channel operates at a variable data rate mode, and the access channels operate at a fixed 4800-bps rate. The reverse traffic channel is used when there is a call. Subscriber data on the uplink radio channel transmitter are also grouped into 20-ms frames, convolutionally encoded, block interleaved, modulated by a 64-ary orthogonal modulation and spread prior to transmission.

CDMA radiated Power

IS-95 specifies complex procedures for regulating the power transmitted by each mobile unit. The goal is to make all reverse-direction signals within a single CDMA channel arrive at the base station with approximately the same signal strength (±1 dB), which is essential for CDMA operation. As signal paths change continuously with moving units, the mobile units perform power adjustments as many as 800 times per second under the control of the base station. Base stations instruct the mobile units to increase or decrease their transmitted power in 1-dB increments.

When a mobile unit is first turned on, it measures the power of the signal received from the base station. The mobile unit assumes that the signal loss is the same in each direction and adjusts its transmit power on the basis of the power level of the signal it receives from the base station. This process is called open-loop power setting. Mobile units use the following formula to compute their transmit power:

$$\mathbf{P_t\ dBm = -76\ dB - P_r}$$

where Pt is transmit power in dBm and Pr is received power in dBm.

With CDMA, rather than limit the maximum transmit power, the minimum and maximum effective isotropic radiated power (EIRP) is specified. The maximum radiated power of base stations is limited to 100 W per 1.23 MHz CDMA channel.

Global System for Mobile Communications

Throughout the evolution of cellular telecommunications, various systems have been developed without the benefit of standardized specifications. This presented many problems directly related to compatibility. GSM standard is intended to address these problems. GSM was the world’s first totally digital cellular telephone system designed to use the services of SS7 signalling and an all-digital data network called integrated services digital network (ISDN) to provide a wide range of network services. GSM is now the world’s most popular standard for new cellular telephone and personal communications equipment.

Advantages of GSM

- Communication: mobile, wireless communication, support for voice and data services
- Total mobility: international access, chip-card enables use of access points of different providers.
- Worldwide connectivity: one number, the network handles every location.
- High capacity: better frequency efficiency, smaller cells, more customers per cell.
- High transmission quality: high audio quality and reliability for wireless, uninterrupted phone calls at higher speeds (e.g., from cars, trains).

GSM Services

GSM telephone services are broadly classified into three categories: bearer services, teleservices, and supplementary services. Teleservices are mainly voice services that provide subscribers with the complete capability to communicate with other subscribers. Data services provide the capacity necessary to transmit appropriate data signals between two access points creating an interface to the network. Some of the subscriber services are given below:

- dualtone multifrequency (DTMF): DTMF is a tone signalling scheme used for various control purposes via the telephone network, such as remote control of an answering machine.
- facsimile group III: GSM supports CCITT group 3 facsimile. This enables a GSM connected fax to communicate with any analog fax in the network.
- short message service: A message consisting of 160 alphanumeric characters can be sent to or from a mobile station. If the mobile station is off or not in the coverage area, the message is stored and then offered back ensuring that the message will be received.
- cell broadcast: a message of maximum of 93 characters can be broadcast to all mobile subscribers in a given geographic area. Typical applications include traffic congestion warnings and reports on accidents.
- voice mail: This service is actually an answering machine within the network controlled by the subscriber. Calls can be forwarded to the subscriber’s voice mail box, which can be checked later by the subscriber via a personal code.
- fax mail: with this service, the subscriber can receive fax at any fax machine.

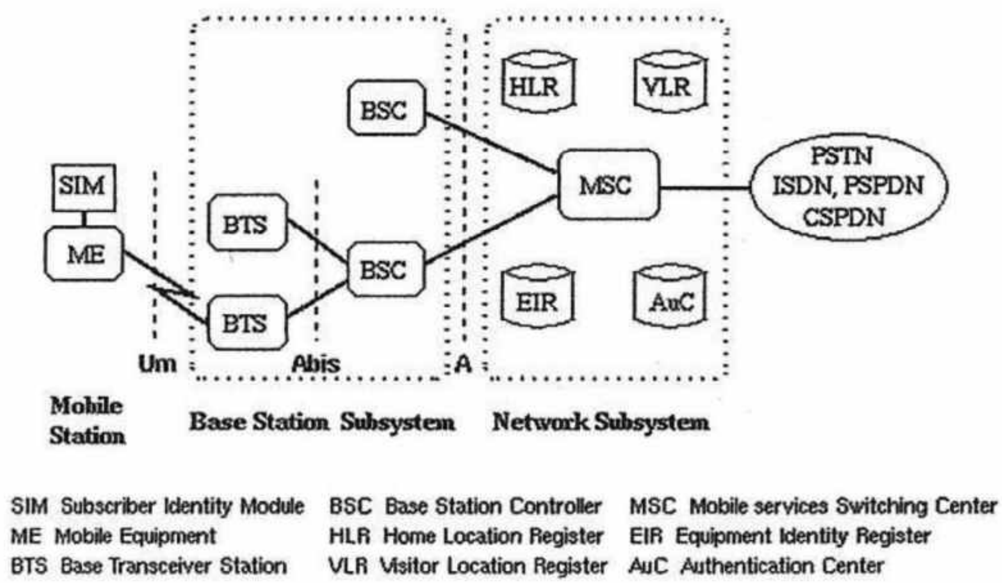
GSM supports a set of supplementary services that can complement and support both telephony and data services. These are defined by GSM and are termed as revenue generating services. Some of them are listed below:

- Call forwarding: It gives the subscriber the ability to forward incoming calls to another number if the called unit is not reachable, not answering, or busy.
- barring of outgoing calls: this service makes it possible for a subscriber to prevent all outgoing calls
- barring of incoming calls: It allows the subscriber to prevent incoming calls either completely or if in roaming
- advise of charge: The AoC service provides the mobile subscriber with an estimate of the call charges.
- call hold: This service enables the subscriber to interrupt an ongoing call and then subsequently re-establish the call.
- call waiting: It allows the mobile subscriber to be notified of an incoming call during a conversation. The subscriber then can answer, reject or ignore the incoming call.

- multiparty service: It enables a mobile subscriber to establish a multiparty conversation i.e. a simultaneous communication between three and six users.
- closed user groups: CUG's are generally comparable to a PBX. They are a group of subscribers who are capable of only calling themselves and certain numbers.
- calling line identification presentation/restriction: these services supply the called party with the integrated services digital network (ISDN) number of the calling party.

GSM architecture

The GSM network is divided into three major systems: the Network Switching Subsystem (NSS), the Base Station Subsystem (BSS) and the Operation and Support System (OSS). The basic GSM elements are shown below:



Network Switching Subsystem: The NSS is responsible for performing call processing and subscriber related functions. The switching system includes the following functional units:

- home location register (HLR): It is a database used for storage and management of subscriptions. HLR stores permanent data about subscribers, including a subscribers service profile, location information and activity status. When an individual buys a subscription from the PCS provider, he or she is registered in the HLR of that operator.
- Visitor location register (VLR): It is a database that contains temporary information about subscribers that is needed by the MSC in order to service visiting subscribers. VLR is always integrated with the MSC. When a MS roams into a new MSC area, the VLR connected to that MSC will request data about the mobile station from the HLR. Later if the mobile station needs to make a call, VLR will be having all the information needed for call setup.

- Authentication center (AUC): A unit called the AUC provides authentication and encryption parameters that verify the users identity and ensure the confidentiality of each call.
- Equipment identity register (EIR): It is a database that contains information about the identity of mobile equipment that prevents calls from stolen, unauthorized or defective mobile stations.
- Mobile switching center (MSC): The MSC performs the telephony switching functions of the system. It controls calls to and from other telephone and data systems.

Base Station Subsystem (BSS): All radio related functions are performed in the BSS, which is also known as radio subsystem. It provides and manages radio-frequency transmission paths between mobile units and MSC. It consists of many base station controllers (BSC) and base transceiver stations (BTS).

- Base station controllers (BSC): The BSC provides all the control functions and physical links between the MSC and BTS. It is a high capacity switch that provides functions such as handover, cell configuration data, and control of radio frequency (RF) power levels in BTS. A number of BSC's are served by and MSC.
- Base transceiver station (BTS): The BTS handles the radio interface to the mobile station. The BTS is the radio equipment (transceivers and antennas) needed to service each cell in the network. A group of BTS's are controlled by an BSC.

Operation and Support system: The operations and maintenance center (OMC) is connected to all equipment in the switching system and to the BSC. Implementation of OMC is called operation and support system (OSS). The OSS is the functional entity from which the network operator monitors and controls the system. The purpose of OSS is to offer the customer cost-effective support for centralized, regional and local operational and maintenance activities that are required for a GSM network. OSS provides a network overview and allows engineers to monitor, diagnose and troubleshoot every aspect of the GSM network.

The mobile station (MS) consists of the mobile equipment (the terminal) and a smart card called the Subscriber Identity Module (SIM). The SIM provides personal mobility, so that the user can have access to subscribed services irrespective of a specific terminal. By inserting the SIM card into another GSM terminal, the user is able to receive calls at that terminal, make calls from that terminal, and receive other subscribed services.

The mobile equipment is uniquely identified by the International Mobile Equipment Identity (IMEI). The SIM card contains the International Mobile Subscriber Identity (IMSI) used to identify the subscriber to the system, a secret key for authentication, and other information. The IMEI and the IMSI are independent, thereby allowing personal mobility. The SIM card may be protected against unauthorized use by a password or personal identity number.

GSM Radio Subsystem

GSM uses two 25-MHz frequency bands that have been set aside for system use in all member companies. The 890 MHz to 915 MHz band is used for mobile unit-to-base station transmissions (reverse link transmissions), and the 935-MHz to 960-MHz frequency band is used for bas station-to-mobile unit transmission (forward link transmission). GSM uses frequency-division duplexing and a combination of TDMA and FDMA techniques to provide base stations simultaneous access to multiple mobile units. The available forward and reverse frequency bands are subdivided into 200-kHz wide voice channels called absolute radio-frequency channel numbers (ARFCN). The ARFCN number designates a forward reverse channel pair with 45-MHz separation between them. Each voice channel is shared among as many as eight mobile units using TDMA.

Each of the ARFCN channel subscribers occupies a unique time slot within the TDMA frame. Radio transmissions in both directions is at a 270.833-kbps rate using binary Gaussian minimum shift keying (GMSK) modulation with an effective channel transmission rate of 33.833 kbps per user.

Personal Communications Satellite System

Personal communications satellite services, however, use low earth orbit (LEO) and medium earth orbit (MEO) satellites that communicate directly with small, low power mobile telephone units. The intention of PCSS mobile telephone is to provide the same features and services offered by traditional, terrestrial cellular telephone providers. PCSS telephones will be able to make or receive calls at anytime, anywhere in the world. The Personal Communication Satellite System (PCSS) is the mother of the Iridium satellite system.

The Iridium System is a satellite-based, wireless personal communications network to permit a wide range of mobile telephone services including voice, data, networking, facsimile, and paging. The iridium uses GSM-based telephony architecture to provide a digitally switched telephone network and global roaming feature is designed in to the system. Each subscriber is assigned a personal phone number and will receive only one bill, no matter in what country or area they use the telephone.

ADVANTAGES

- Less reliance on wire-line networks
- Continuous talk time
- Fewer outages
- Don't need to be in the in the same footprint as the gateway

DISADVANTAGES

- High risk associated with designing, building, and launching satellites.
- High cost for the terrestrial-based networking and interface infrastructure.
- low power, dual mode transceivers are more cumbersome and expensive

APPLICATIONS

- Fixed cellular telephone service
- Complementary and back up telephone service in fields of:
  - Manufacturing
  - Military
  - Government
  - Transportation

Comparison between iridium and traditional satellite systems: -

- Iridium is the first mobile satellite to incorporate sophisticated, onboard digital processing on each satellite.
- Entire global coverage by a single wireless network system.
- Only provider of truly global voice and data solutions.
- With this system the subscriber will never listen a message called "OUT OF COVERAGE AREA". This list provides just a few of absolutely inexhaustible list of comparisons.

Appendix diagrams

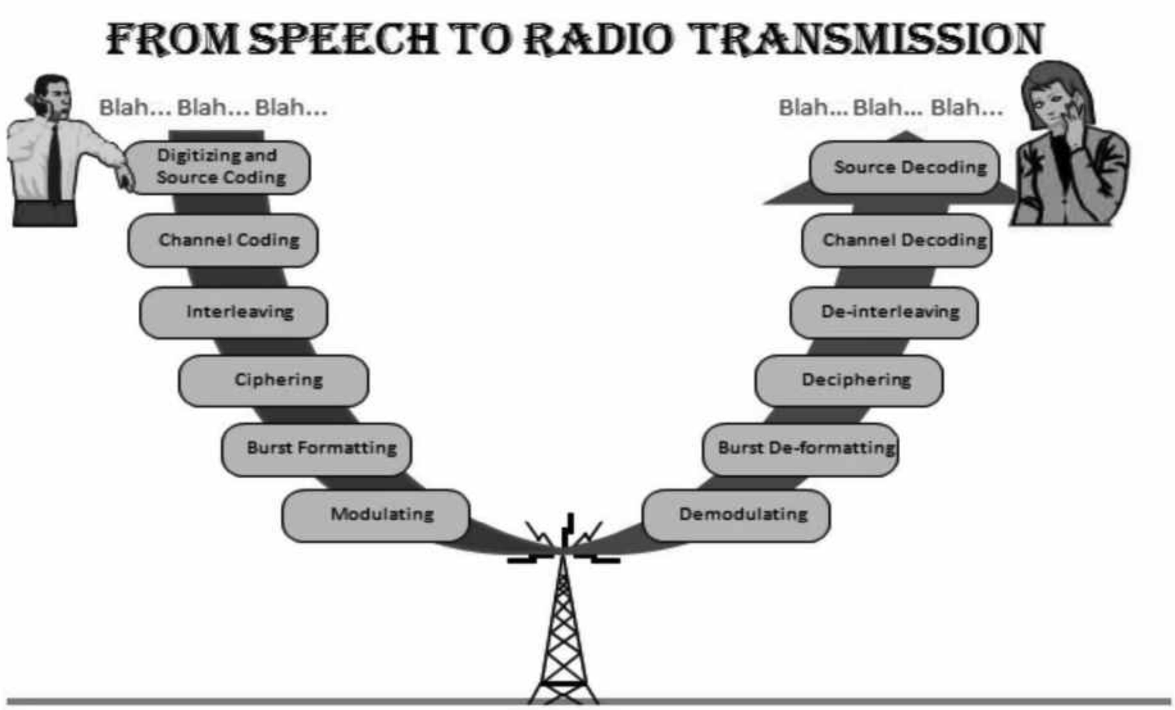
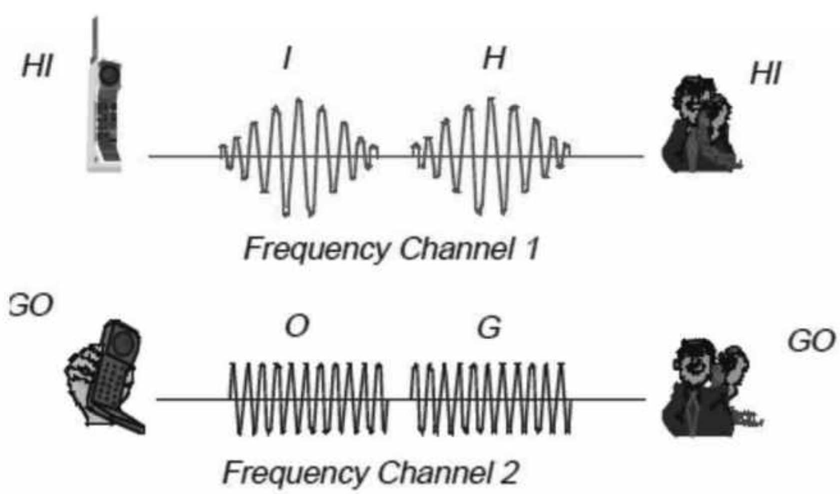


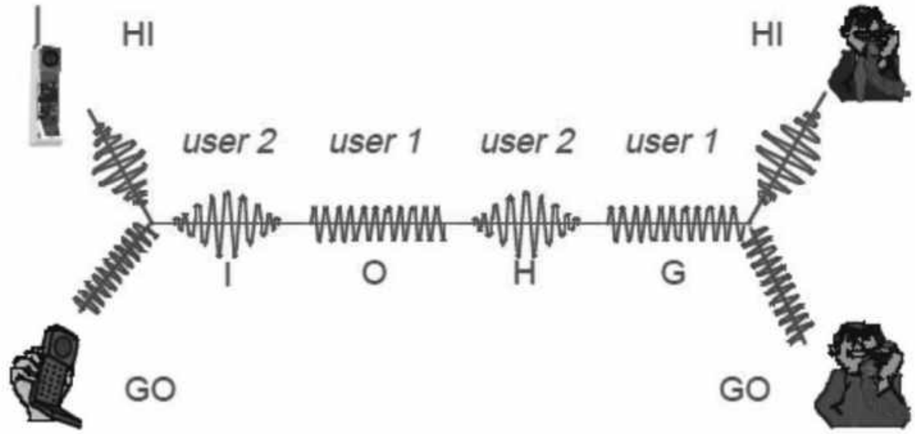
Table 1.1 Major Mobile Radio Standards in North America

Standard	Type	Year of Introduction	Multiple Access	Frequency Band	Modulation	Channel Bandwidth
AMPS	Cellular	1983	FDMA	824-894 MHz	FM	30 kHz
NAMPS	Cellular	1992	FDMA	824-894 MHz	FM	10 kHz
USDC	Cellular	1991	TDMA	824-894 MHz	$\pi/4$-DQPSK	30 kHz
CDPD	Cellular	1993	FH/Packet	824-894 MHz	GMSK	30 kHz
IS-95	Cellular/PCS	1993	CDMA	824-894 MHz 1.8-2.0 GHz	QPSK/BPSK	1.25 MHz
GSC	Paging	1970's	Simplex	Several	FSK	12.5 kHz
POCSAG	Paging	1970's	Simplex	Several	FSK	12.5 kHz
FLEX	Paging	1993	Simplex	Several	4-FSK	15 kHz
DCS-1900 (GSM)	PCS	1994	TDMA	1.85-1.99 GHz	GMSK	200 kHz
PACS	Cordless/PCS	1994	TDMA/FDMA	1.85-1.99 GHz	$\pi/4$-DQPSK	300 kHz
MIRS	SMR/PCS	1994	TDMA	Several	16-QAM	25 kHz

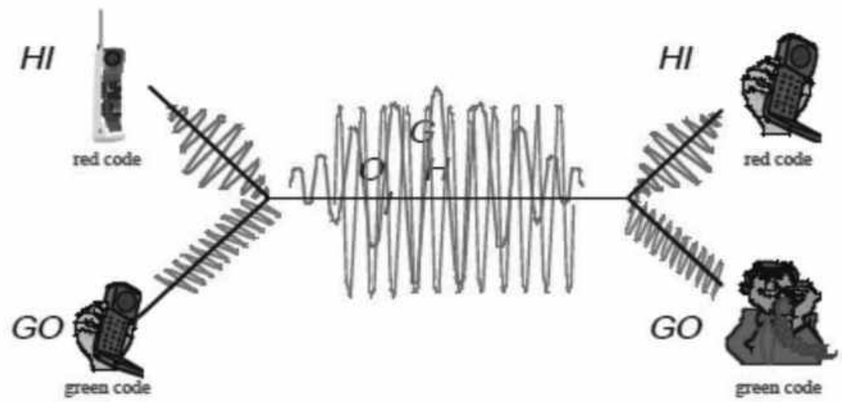
Multiple Accessing Schemes: HOW IT WORKS



FDMA


HI
HI
user 2
user 1
user 2
user 1
I
O
H
G
GO
GO

TDMA


HI
red code
HI
red code
G
O
H
GO
green code
GO
green code

CDMA

Questions

1. (a) What is a digital Cellular System? List the advantages of a digital cellular system
   (b) Explain the classifications of CDMA radiated power. Determine the transmit power for a CDMA mobile unit that is receiving a signal from the base station at -100dB

2. (a) What is GSM cellular telephone system? Describe the services offered by GSM
   (b) What is meant by false handoff? What are the four types of handoff's possible with N-AMPS? Compare macro cellular system and digital cellular system

3. (a) What is N-AMPS cellular telephone system? Explain the operation of N-AMPS cellular telephone system
   (b) List the basic parameters of GSM and briefly describe the GSM radio system?

4. (a) What are the three primary subsystems of GSM? Describe in detail, the GSM system architecture
   (b) Explain with diagrams the CDMA traffic channels

5. (a) Explain the TDMA scheme used with USDC and its advantages
   (b) what is an interference avoidance scheme

6.(a) Outline the advantages and disadvantages of PCSS over terrestrial cellular telephone systems?
   (b) Briefly describe the E-TDMA scheme?