CMR
INSTITUTE OF
TECHNOLOGY

USN | | | | | | | | | |

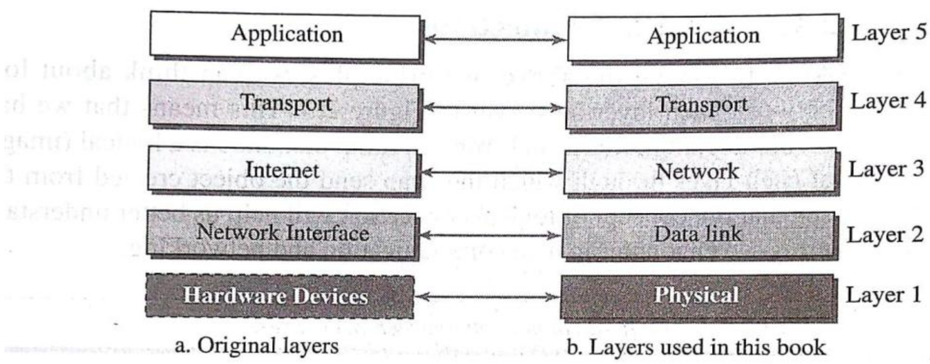
First Internal Test

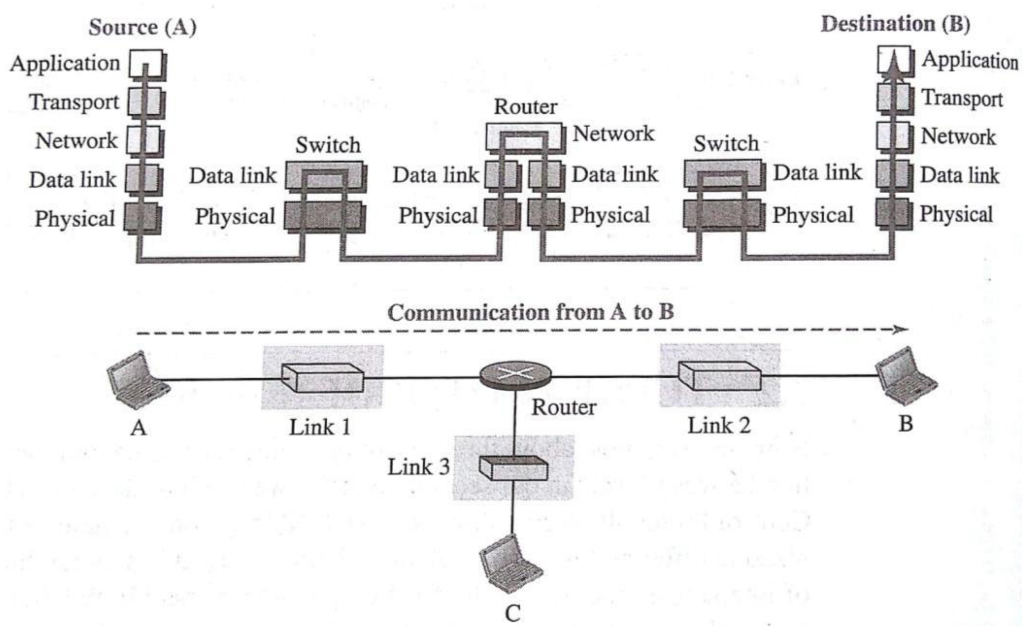
Sub:	Data Communication							Code:	17CS46
Date:	07/ 03 / 2019	Duration:	90 mins	Max Marks:	50	Sem:	IV	Branch:	**ISE**
Answer Any **FIVE FULL** Questions									

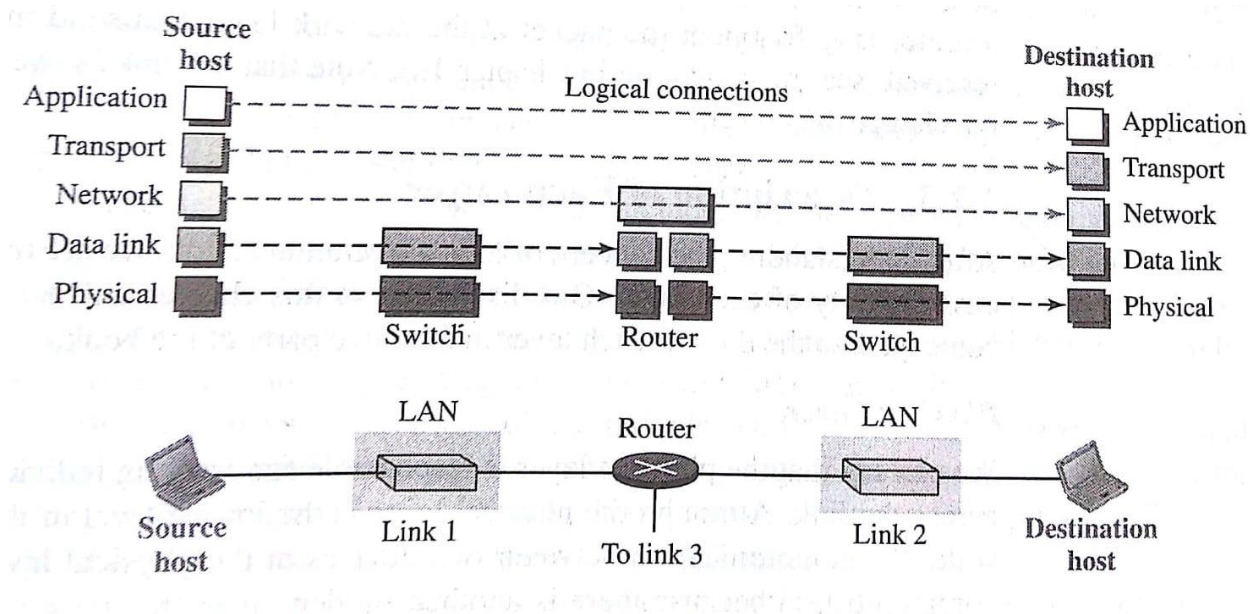
		Marks	OBE CO	OBE RBT
1	Explain the TCP/IP protocol suite with neat diagrams.	[10]	CO1	L2
2 (a)	Explain the different causes for transmission impairments during signal transmission through media.	[6]	CO2	L2
(b)	Calculate the Shannon channel capacity in the following cases: (i) BW = 20 kHz, $SNR_{dB} = 40$ (ii) BW = 200 kHz, $SNR_{dB} = 6$	[4]	CO2	L3
3 (a)	With neat diagrams, explain the 4 basic network topologies.	[8]	CO2	L2
(b)	A network with a bandwidth of 10Mbps can pass only an average of 18000 frames per minute with each frame carrying an average of 10000 bits. What is the throughput of this network?	[2]	CO2	L3
4 (a)	Define line coding. List out its characteristics.	[5]	CO2	L1
(b)	Represent the sequence 10100110 using polar and biphase schemes.	[5]	CO2	L3
5	Explain PCM and quantization process with steps and example.	[10]	CO2	L2
6 (a)	Differentiate between the 3 types of serial transmission.	[6]	CO2	L2
(b)	Define constellation diagram. Show the constellation diagrams for ASK, BPSK and QPSK signals.	[4]	CO2	L2
7	Analyze ASK, FSK and PSK mechanisms applying them over the digital data 101101.	[10]	CO2	L4
8 (a)	What is FDM? Briefly explain its multiplexing and demultiplexing process.	[6]	CO2	L2
(b)	Discuss the analog hierarchy system used in FDM.	[4]	CO2	L2

1. Explain the TCP/IP protocol suite with neat diagrams. (10 marks)

TCP/IP is a protocol suite (a set of protocols organized in different layers) used in the Internet today. It is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality. The term hierarchical means that each upper level protocol is supported by the services provided by one or more lower level protocols. The original TCP/IP protocol suite was defined as four software layers built upon the hardware. Today, however, TCP/IP is thought of as a five-layer model. Figure below shows both configurations.

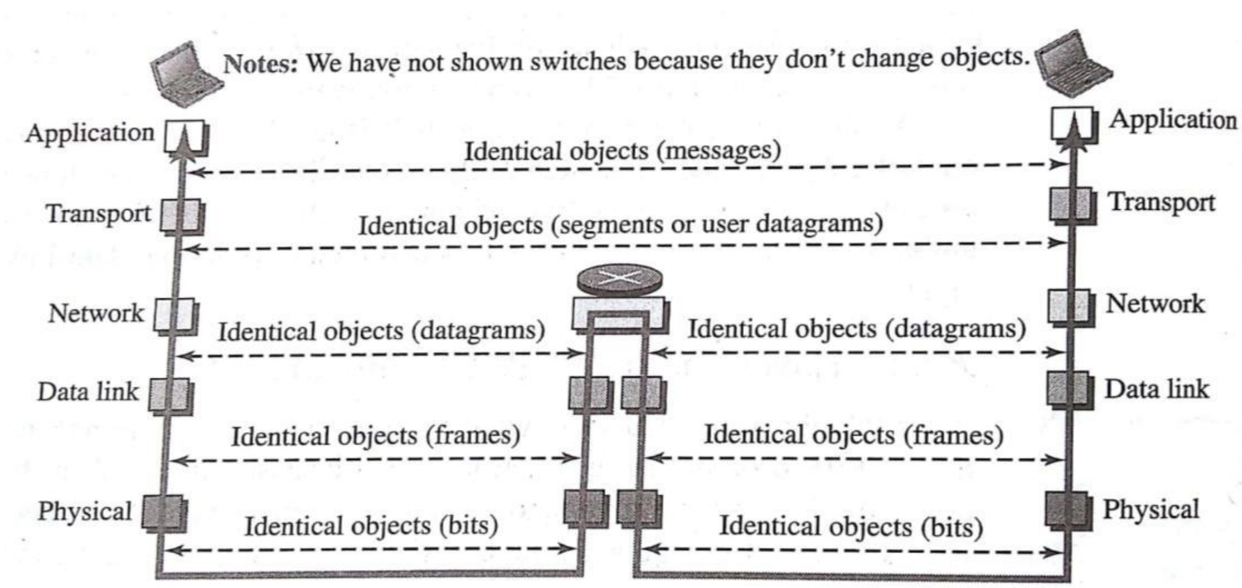






Using logical connections makes it easier for us to think about the duty of each layer. As the figure shows, the duty of the application, transport, and network layers is end-to-end. However, the duty of the data-link and physical layers is hop-to-hop, in which a hop is a host or router. In other words, the domain of duty of the top three layers is the internet, and the domain of duty of the two lower layers is the link. Another way of thinking of the logical connections is to think about the data unit created from each layer. In the top three layers, the data unit (packets) should not be changed by any router or link-layer switch. In the bottom two layers, the packet created by the host is changed only by the routers, not by the link-layer switches.

Figure below shows the second principle discussed previously for protocol layering. We show the identical objects below each layer related to each device.

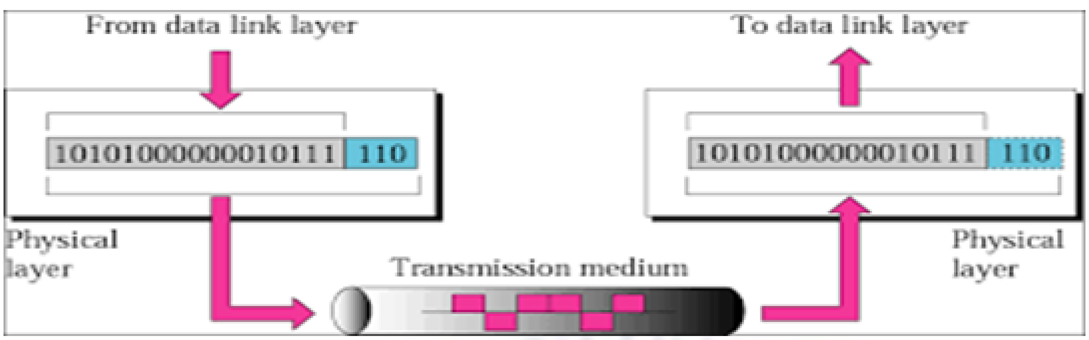


Although the logical connection at the network layer is between the two hosts, we can only say that identical objects exist between two hops in this case because a router may fragment the packet at the network layer and send more packets than received . The link between two hops does not change the object.

Physical Layer

The physical layer is responsible for carrying individual bits in a frame across the link. Although the physical layer is the lowest level in the TCP/IP protocol suite, the communication between two devices at the physical layer is still a logical communication because there is another, hidden layer, the transmission media, under the physical layer. Two devices are connected by a transmission medium (cable or air). We need to know that the transmission medium does not carry bits; it carries electrical or optical signals. So the bits received in a frame from the data-link layer are transformed and sent through the transmission media, but we can think that the logical unit between two physical layers in two devices is a bit. There are several protocols that transform a bit to a signal.

The following figure shows the position of the physical layer with respect to the transmission medium and the data link layer.



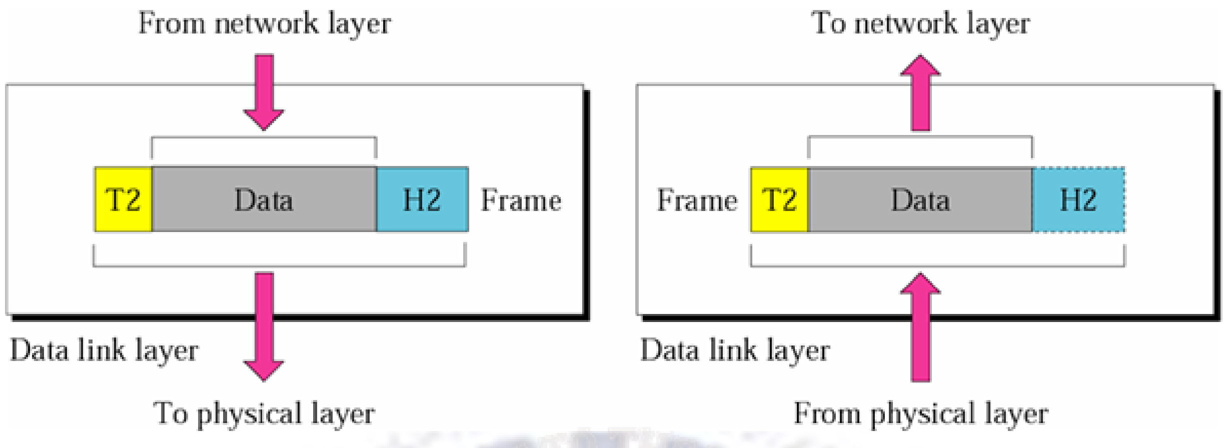
Data Link Layer

An internet is made up of several links (LANs and WANs) connected by routers. There may be several overlapping sets of links that a datagram can travel from the host to the destination. The routers are responsible for choosing the best links. However, when the next link to travel is determined by the router, the data-link layer is responsible for taking the datagram and moving it across the link. The link can be a wired LAN with a link-layer switch, a wireless LAN, a wired WAN, or a wireless WAN. We can also have different protocols used with any link type. In each case, the data-link layer is responsible for moving the packet through the link.

TCP/IP does not define any specific protocol for the data-link layer. It supports all the standard and proprietary protocols. Any protocol that can take the datagram and carry it through the link suffices for the network layer. The data-link layer takes a datagram and encapsulates it in a packet called a frame.

Each link-layer protocol may provide a different service. Some link-layer protocols provide complete error detection and correction, some provide only error correction.

The next figure shows the relationship of the data link layer to the network and physical layers.



Network Layer

The network layer is responsible for creating a connection between the source computer and the destination computer. The communication at the network layer is host-to-host. However, since there can be several routers from the source to the destination, the routers in the path are responsible for choosing the best route for each packet. We can say that the network layer is responsible for host-to-host communication and routing the packet through possible routes.

Reasons for having a separate network layer: separation of different tasks between different layers and routers do not need the application and transport layers; separating the tasks allows us to use fewer protocols on the routers.

The network layer in the Internet includes the main protocol, Internet Protocol (IP), that defines the format of the packet, called a datagram at the network layer. IP also defines the format and the structure of addresses used in this layer. IP is also responsible for routing a packet from its source to its destination, which is achieved by each router forwarding the datagram to the next router in its path.

IP is a connectionless protocol that provides no flow control, no error control, and no congestion control services. This means that if any of these services is required for an application, the application should rely only on the transport-layer protocol. The network layer also includes unicast (one-to-one) and multicast (one-to-many) routing protocols. A routing protocol does not take part in routing (it is the responsibility of IP), but it creates forwarding tables for routers to help them in the routing process.

The network layer also has some auxiliary protocols that help IP in its delivery and routing tasks. The Internet Control Message Protocol (ICMP) helps IP to report some problems when routing a packet. The Internet Group Management Protocol (IGMP) is another protocol that helps IP in multitasking. The Dynamic Host Configuration Protocol (DHCP) helps IP to get the network-layer address for a host. The Address Resolution Protocol (ARP) is a protocol that helps IP to find the link-layer address of a host or a router when its network-layer address is given.

Transport Layer

The logical connection at the transport layer is also end-to-end. The transport layer at the source host gets the message from the application layer, encapsulates it in a transport layer packet (called a segment or a user datagram in different protocols) and sends it, through the logical (imaginary) connection, to the transport layer at the destination host. In other words, the transport layer is responsible for giving services to the application layer: to get a message from an application program running on the source host and deliver it to the corresponding application program on the destination host. We have more than one protocol in the transport layer, which means that each application program can use the protocol that best matches its requirement.

The main protocol, Transmission Control Protocol (TCP), is a connection-oriented protocol that first establishes a logical connection between transport layers at two hosts before transferring data. It creates a logical pipe between two TCPs for transferring a stream of bytes. TCP provides flow control (matching the sending data rate of the source host with the receiving data rate of the destination host to prevent overwhelming the destination), error control (to guarantee that the segments arrive at the destination without error and resending the corrupted ones), and congestion control to reduce the loss of segments due to congestion in the network. The other common protocol, User Datagram Protocol (UDP), is a connectionless protocol that transmits user datagrams without first creating a logical connection. In UDP, each user datagram is an independent entity without being related to the previous or the next one (the meaning of the term connectionless). UDP is a simple protocol that does not provide flow, error, or congestion control. Its simplicity, which means small overhead, is attractive to an application program that needs to send short messages and cannot afford the retransmission of the packets involved in TCP, when a packet is corrupted or lost. A new protocol, Stream Control Transmission Protocol (SCTP) is designed to respond to new applications that are emerging in the multimedia.

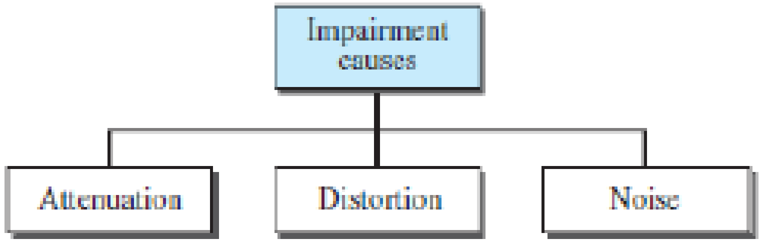
Application Layer

The logical connection between the two application layers is end to-end. The two application layers exchange messages between each other as though there were a bridge between the two layers. However, we should know that the communication is done through all the layers. Communication at the application layer is between two processes (two programs running at this layer). To communicate, a process sends a request to the other process and receives a response. Process-to-process communication is the duty of the application layer. The application layer in the Internet includes many predefined protocols, but a user can also create a pair of processes to be run at the two hosts.

The Hypertext Transfer Protocol (HTTP) is a vehicle for accessing the World Wide Web (WWW). The Simple Mail Transfer Protocol (SMTP) is the main protocol used in electronic mail (e-mail) service. The File Transfer Protocol (FTP) is used for transferring files from one host to another. The Terminal Network (TELNET) and Secure Shell (SSH) are used for accessing a site remotely. The Simple Network Management Protocol (SNMP) is used by an administrator to manage the Internet at global and local levels. The Domain Name System (DNS) is used by other protocols to find the network-layer address of a computer. The Internet Group Management Protocol (IGMP) is used to collect membership in a group.

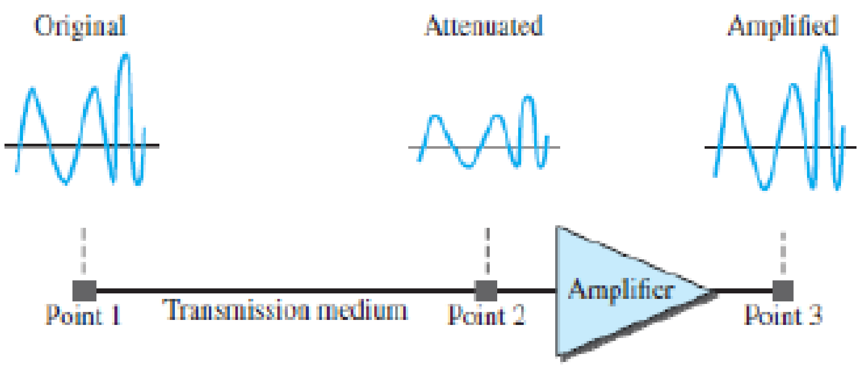
2(a). Explain the different causes for transmission impairments during signal transmission through media. (6 marks)

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



1. Attenuation

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



Decibel

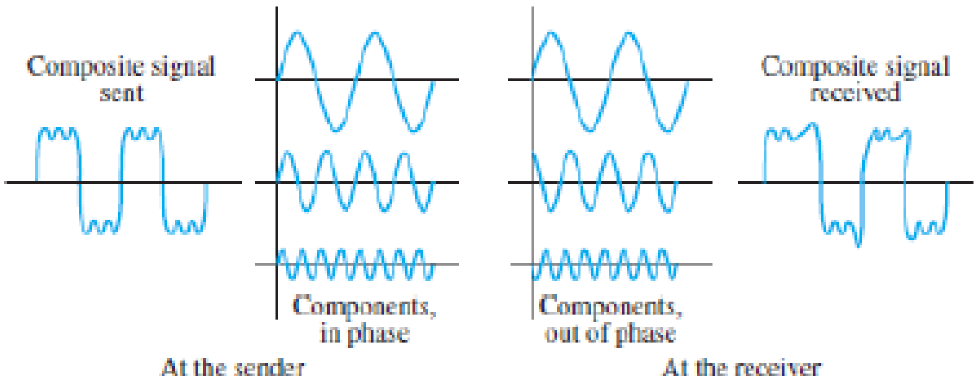
To show that a signal has lost or gained strength, engineers use the unit of the decibel. The decibel (dB) measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified.

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

Variables P1 and P2 are the powers of a signal at points 1 and 2, respectively. Note that the decibel can also be defined in terms of voltage instead of power. In this case, because power is proportional to the square of the voltage, the formula is dB = 20 $\log_{10}$ (V2/V1).

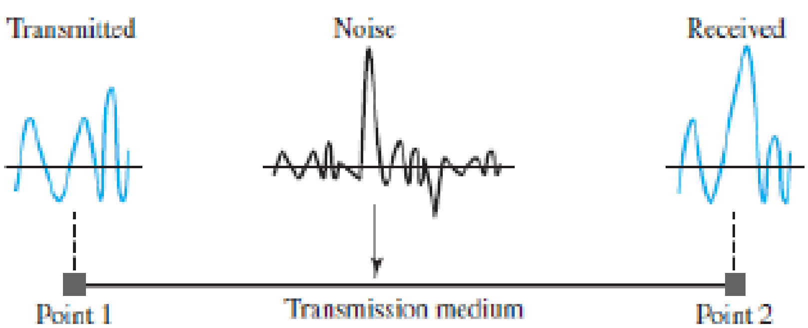
2. Distortion

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



3. Noise

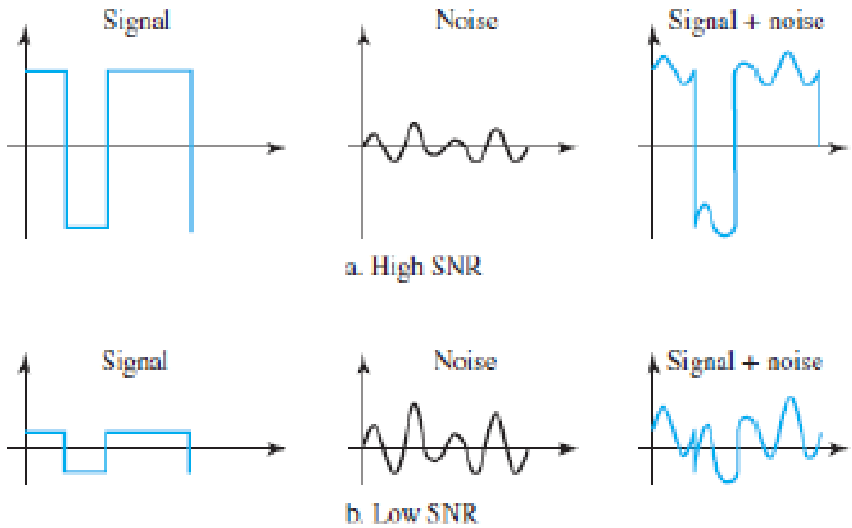
Noise is another cause of impairment. Several types of noise, such as thermal noise, induced noise, crosstalk, and impulse noise, may corrupt the signal. Thermal noise is the random motion of electrons in a wire, which creates an extra signal not originally sent by the transmitter. Induced noise comes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



Signal-to-Noise Ratio (SNR)

As we will see later, to find the theoretical bit rate limit, we need to know the ratio of the signal power to the noise power. The signal-to-noise ratio is defined as

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



SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise). A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise. Because SNR is the ratio of two powers, it is often described in decibel units, SNRdB, defined as

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

2(b). Calculate the Shannon channel capacity in the following cases:

(i) BW = 20 kHz, SNR_{dB} = 40

(ii) BW = 200 kHz, SNR_{dB} = 6

(4 marks)

2(b) (i) BW = 20 kHz ; SNR_{dB} = 40 .

Shannon channel capacity

= Bandwidth × $\log_2 (1 + SNR)$.

$SNR_{dB} = 10 \log_{10} SNR$.

ie, $\cancel{40} = \cancel{10} \log_{10} SNR \Rightarrow \log_{10} SNR = 4$

$\therefore SNR = 10^4 = \underline{10{,}000}$.

$\therefore$ Channel capacity $= 20 \times 10^3 \times \log_2 (1 + 10000)$

$= 20 \times 10^3 \times \log_2 (10001)$

$= 20 \times 10^3 \times 13.289$

$= \underline{\underline{265.78}}$ kbps .

(ii) BW = 200kHz ; SNR_{dB} = 6 .

$\therefore SNR_{dB} = 10 \log_{10} SNR$

$\Rightarrow 6 = 10 \log_{10} SNR \Rightarrow \log_{10} SNR = 0.6$

$\therefore SNR = 10^{0.6} = \underline{3.981}$

Shannon channel capacity = BW × $\log_2 (1 + SNR)$

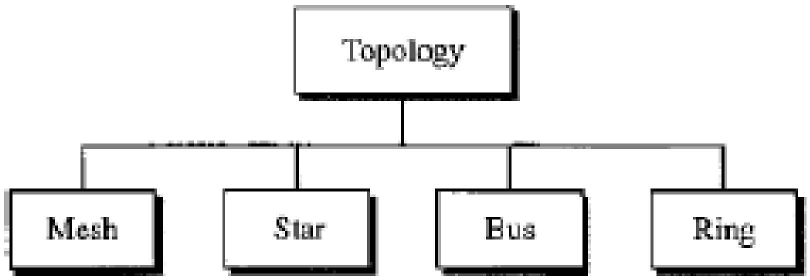
$= 200 \times 10^3 \times \log_2 (1 + 3.981)$

$= 200 \times 10^3 \times \log_2 (4.981)$

$= 200 \times 10^3 \times 2.317$

$= \underline{\underline{463.4}}$ kbps .

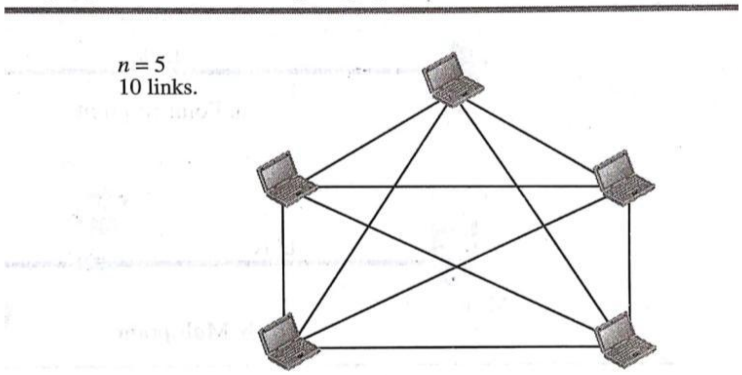
3(a). With neat diagrams, explain the 4 basic network topologies. (8 marks)



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

Mesh Topology

A fully connected mesh topology (five devices)

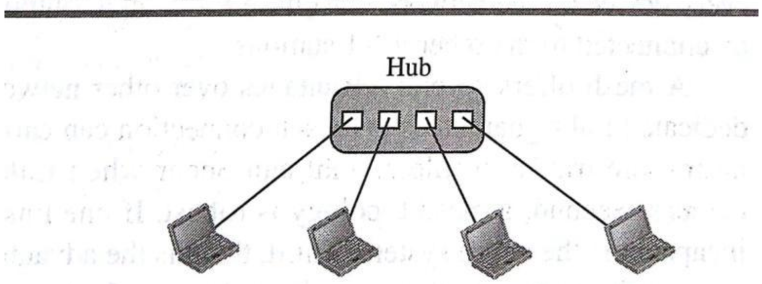


- Every device has a dedicated point-to-point link to every other device.
- The link carries traffic only between the two devices it connects.
- To find the number of physical links in a fully connected mesh network with n nodes, we first consider that each node must be connected to every other node.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

- Advantages:
  - The use of dedicated links guarantees that each connection can carry its own data load, thus eliminating the traffic problems that can occur when links must be shared by multiple devices.
  - A mesh topology is robust. If one link becomes unusable, it does not incapacitate the entire system.
  - Privacy or security - When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages.
  - Point-to-point links make fault identification and fault isolation easy. Traffic can be routed to avoid links with suspected problems. This facility enables the network manager to discover the precise location of the fault and aids in finding its cause and solution.
- Disadvantages:
  - Because every device must be connected to every other device, installation and reconnection are difficult.
  - The sheer bulk of the wiring can be greater than the available space can accommodate.
  - The hardware required to connect each link (I/O ports and cable) can be prohibitively expensive.

Star Topology

A star topology connecting four stations



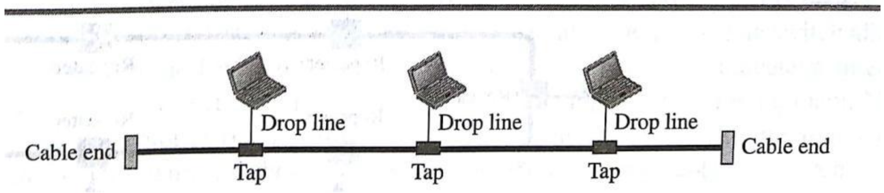
- Each device has a dedicated point-to-point link only to a central controller, usually called a hub. The devices are not directly linked to one another.
- It does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data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

- Advantages:
  - Less expensive - each device needs only one link and one I/O port to connect it to any number of others.
  - It is easy to install and reconfigure - less cabling needs to be housed, and additions, moves, and deletions involve only one connection: between that device and the hub.
  - Robustness - If one link fails, only that link is affected. All other links remain active.

- ➢ Easy fault identification and fault isolation - As long as the hub is working, it can be used to monitor link problems and bypass defective links.
- Disadvantages:
- ➢ Dependency of the whole topology on one single point, the hub - If the hub goes down, the whole system is dead.
- ➢ More cabling is required in a star than in some other topologies because although a star requires far less cable than a mesh, each node must be linked to a central hub.

Bus Topology

'e 1.6 *A bus topology connecting three stations*



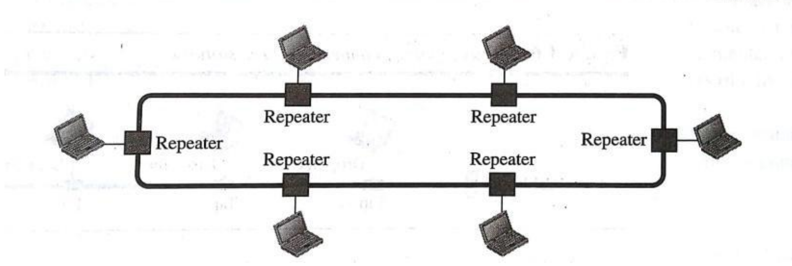
- A bus topology is multipoint. One long cable acts as a backbone to link all the devices in a network.
- Nodes are connected to the bus cable by drop lines and taps.

  A drop line is a connection running between the device and the main cable.

  A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core.
- As a signal travels along the backbone, some of its energy is transformed into heat. Therefore, it becomes weaker and weaker as it travels farther and farther. For this reason there is a limit on the number of taps a bus can support and on the distance between those taps.
- Advantages:
- ➢ Ease of installation - Backbone cable can be laid along the most efficient path, then connected to the nodes by drop lines of various lengths.
- ➢ Uses less cabling - Only the backbone cable stretches through the entire facility. Each drop line has to reach only as far as the nearest point on the backbone.
- Disadvantages:
- ➢ Difficult reconnection and fault isolation - A bus is usually designed to be optimally efficient at installation. It can therefore be difficult to add new devices.
- ➢ Signal reflection at the taps can cause degradation in quality.
- ➢ A fault or break in the bus cable stops all transmission, even between devices on the same side of the problem. The damaged area reflects signals back in the direction of origin, creating noise in both directions.

Ring Topology

Figure 1.7 *A ring topology connecting six stations*



- Each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination.
- Each device in the ring incorporates a repeater.
- When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along.
- Advantage:
  - Easy to install and reconfigure - Each device is linked to only its immediate neighbors (either physically or logically). To add or delete a device requires changing only two connections after considering media and traffic.
  - Fault isolation is simplified - Generally in a ring, a signal is circulating at all times. If one device does not receive a signal within a specified period, it can issue an alarm. The alarm alerts the network operator to the problem and its location.
- Disadvantage:
  - Unidirectional traffic – A break in the ring (such as a disabled station) can disable the entire network.

3(b). A network with a bandwidth of 10Mbps can pass only an average of 18000 frames per minute with each frame carrying an average of 10000 bits. What is the throughput of this network? (2 marks)

Bandwidth = 10 Mbps

$$\text{Throughput} = \frac{(\overset{3000}{\cancel{18000}} \times 1000\cancel{0})\ \text{bits}}{\cancel{6}\cancel{0}\ \text{sec}}$$

$$= 3000 \times 10^{3}\ \text{bps}$$

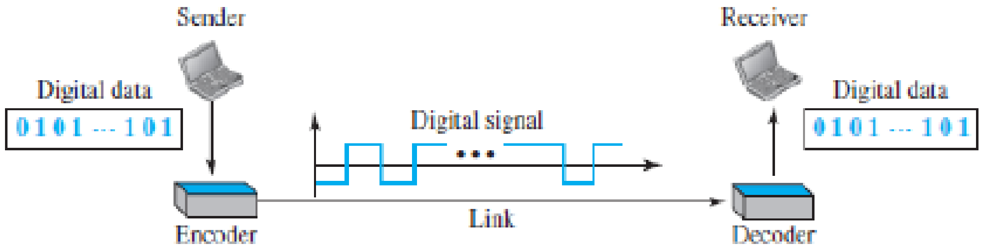
$$= 3 \times 10^{6}\ \text{bps}$$

$$= \underline{3\ \text{Mbps}}$$

4(a). Define line coding. List out its characteristics. (5 marks)

Line Coding

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

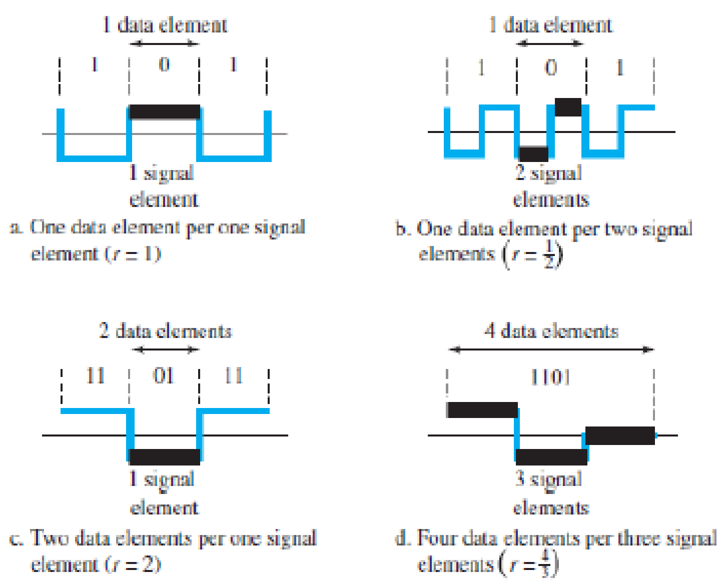


Characteristics of line coding

1. Signal Element Versus Data Element

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

- We define a ratio r which is the number of data elements carried by each signal element. Figure shows several situations with different values of r.



In part a of the figure, one data element is carried by one signal element (r = 1). In part b of the figure, we need two signal elements (two transitions) to carry each data element (r =). We will see later that the extra signal element is needed to guarantee synchronization. In part c of the figure, a signal element carries two data elements (r = 2). Finally, in part d, a group of 4 bits is being carried by a group of three signal elements (r = 4/3).

2. Data Rate Versus Signal Rate

- The data rate defines the number of data elements (bits) sent in 1s. The unit is bits per second (bps). The signal rate is the number of signal elements sent in 1s. The unit is the baud. The data rate is sometimes called the bit rate; the signal rate is sometimes called the pulse rate, the modulation rate, or the baud rate.
- One goal in data communications is to increase the data rate while decreasing the signal rate. Increasing the data rate increases the speed of transmission; decreasing the signal rate decreases the bandwidth requirement.
- Consider the relationship between data rate (N) and signal rate (S).

$$S = N/r$$

Here, r has been previously defined. This relationship, of course, depends on the value of r. It also depends on the data pattern. If we have a data pattern of all 1s or all 0s, the signal rate may be different from a data pattern of alternating 0s and 1s. To derive a formula for the relationship, we need to define three cases: the worst, best, and average. The worst case is when we need the maximum signal rate; the best case is when we need the minimum. In data communications, we are usually interested in the average case. We can formulate the relationship between data rate and signal rate as

$$S_{avg} = c \times N \times (1/r) \quad \text{baud}$$

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

3. Bandwidth

- A digital signal that carries information is nonperiodic. The bandwidth of a nonperiodic signal is continuous with an infinite range. However, most digital signals we encounter in real life have a bandwidth with finite values. In other words, the bandwidth is theoretically infinite, but many of the components have such a small amplitude that they can be ignored. The effective bandwidth is finite.
- The baud rate determines the required bandwidth for a digital signal. The bandwidth reflects the range of frequencies we need. There is a relationship between the baud rate (signal rate) and the bandwidth.
- The bandwidth (range of frequencies) is proportional to the signal rate (baud rate). The minimum bandwidth can be given as

$$B_{min} = c \times N \times (1/r)$$

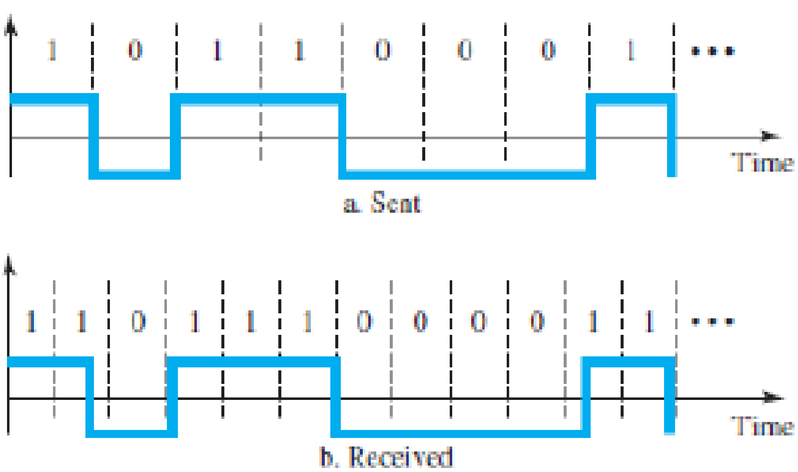
- We can solve for the maximum data rate if the bandwidth of the channel is given.

$$N_{max} = (1/c) \times B \times r$$

4. Baseline Wandering
In decoding a digital signal, the receiver calculates a running average of the received signal power. This average is called the baseline. The incoming signal power is evaluated against this baseline to determine the value of the data element. A long string of 0s or 1s can cause a drift in the baseline (baseline wandering) and make it difficult for the receiver to decode correctly. A good line coding scheme needs to prevent baseline wandering.

5. DC Components
When the voltage level in a digital signal is constant for a while, the spectrum creates very low frequencies (results of Fourier analysis). These frequencies around zero, called DC (direct-current) components, present problems for a system that cannot pass low frequencies or a system that uses electrical coupling (via a transformer). We can say that DC component means 0/1 parity that can cause base-line wondering. For example, a telephone line cannot pass frequencies below 200 Hz. Also a long-distance link may use one or more transformers to isolate different parts of the line electrically. For these systems, we need a scheme with no DC component.

6. Self-synchronization
To correctly interpret the signals received from the sender, the receiver's bit intervals must correspond exactly to the sender's bit intervals. If the receiver clock is faster or slower, the bit intervals are not matched and the receiver might misinterpret the signals. Figure 4.3 shows a situation in which the receiver has a shorter bit duration. The sender sends 10110001, while the receiver receives 110111000011.



A self-synchronizing digital signal includes timing information in the data being transmitted. This can be achieved if there are transitions in the signal that alert the receiver to the beginning, middle, or end of the pulse. If the receiver's clock is out of synchronization, these points can reset the clock.

7. Built-in Error Detection

It is desirable to have a built-in error-detecting capability in the generated code to detect some or all of the errors that occurred during transmission.

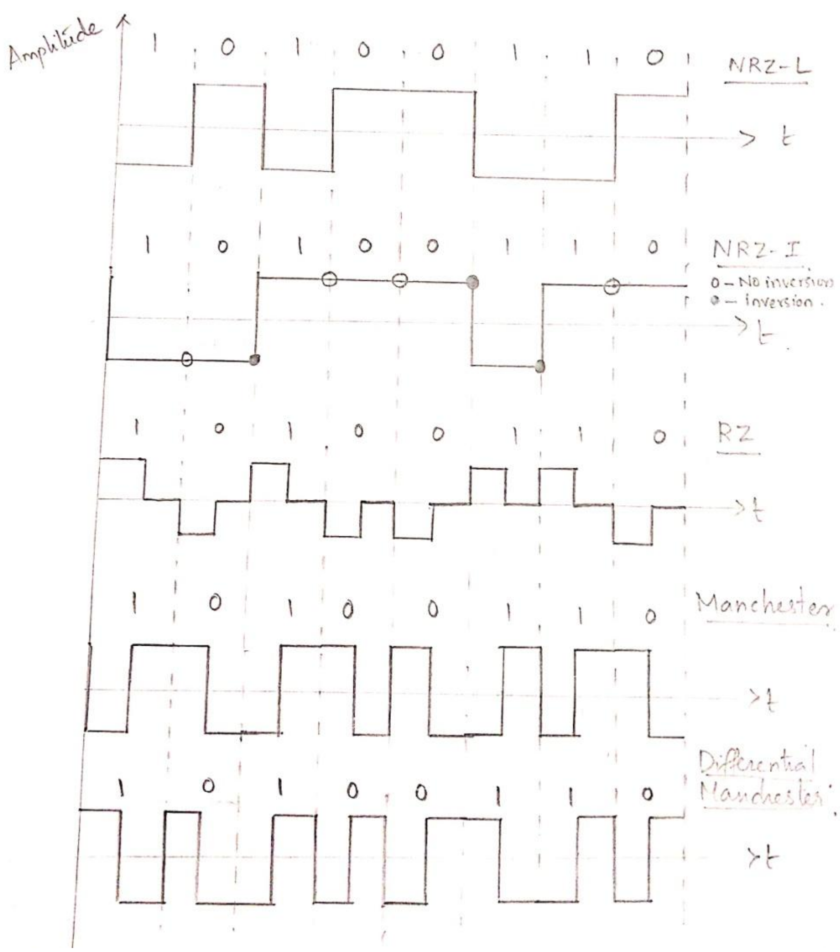
8. Immunity to Noise and Interference

Another desirable code characteristic is a code that is immune to noise and other interferences.

9. Complexity

A complex scheme is more costly to implement than a simple one. For example, a scheme that uses four signal levels is more difficult to interpret than one that uses only two levels.

4(b). Represent the sequence 10100110 using polar and biphase schemes. (5 marks)

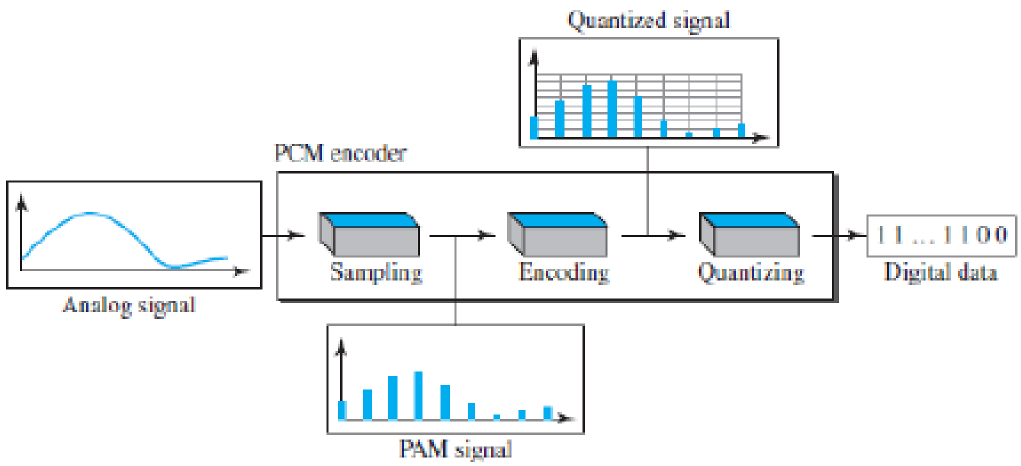


5. Explain PCM and quantization process with steps and example. (10 marks)

Sometimes, we have an analog signal such as one created by a microphone or camera. The tendency today is to change an analog signal to digital data. Two techniques - pulse code modulation and delta modulation are used for this conversion of analog signal into digital data.

Pulse Code Modulation (PCM)

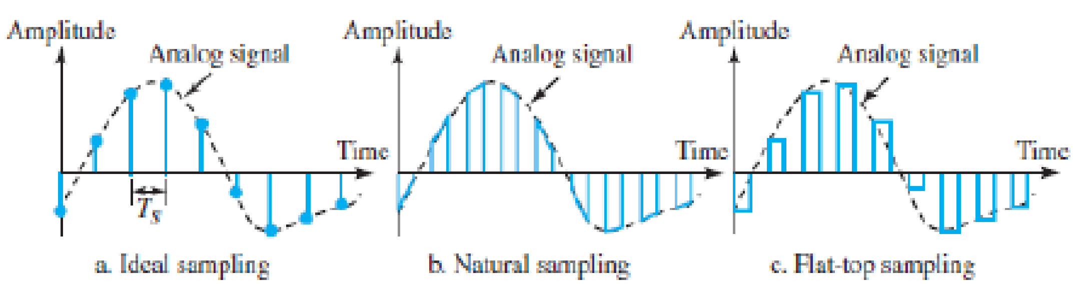
The most common technique to change an analog signal to digital data (digitization) is called pulse code modulation (PCM). A PCM encoder has three processes, as shown in the below figure.



1. The analog signal is sampled.
2. The sampled signal is quantized.
3. The quantized values are encoded as streams of bits.

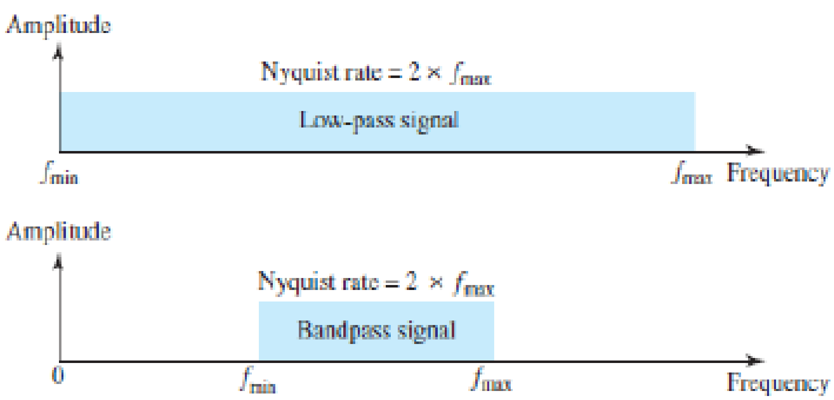
i. Sampling

The first step in PCM is sampling. The analog signal is sampled every T_s s, where T_s is the sample interval or period. The inverse of the sampling interval is called the sampling rate or sampling frequency and denoted by f_s, where $f_s = 1/T_s$. There are three sampling methods—ideal, natural, and flat-top as shown in the next figure.



In ideal sampling, pulses from the analog signal are sampled. This is an ideal sampling method and cannot be easily implemented. In natural sampling, a high-speed switch is turned on for only the small period of time when the sampling occurs. The result is a sequence of samples that retains the shape of the analog signal. The most common sampling method, called sample and hold, however, creates flat-top samples by using a circuit. The sampling process is sometimes referred to as pulse amplitude modulation (PAM). The result of sampling is still an analog signal with non-integral values.

Sampling Rate:
According to the Nyquist theorem, to reproduce the original analog signal, one necessary condition is that the sampling rate be at least twice the highest frequency in the original signal. First, we can sample a signal only if the signal is band-limited. In other words, a signal with an infinite bandwidth cannot be sampled. Second, the sampling rate must be at least 2 times the highest frequency, not the bandwidth. If the analog signal is low-pass, the bandwidth and the highest frequency are the same value. If the analog signal is band-pass, the bandwidth value is lower than the value of the maximum frequency. The next figure shows the value of the sampling rate for two types of signals.



ii. Quantization

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

a. We assume that the original analog signal has instantaneous amplitudes between V_{min} and V_{max}.

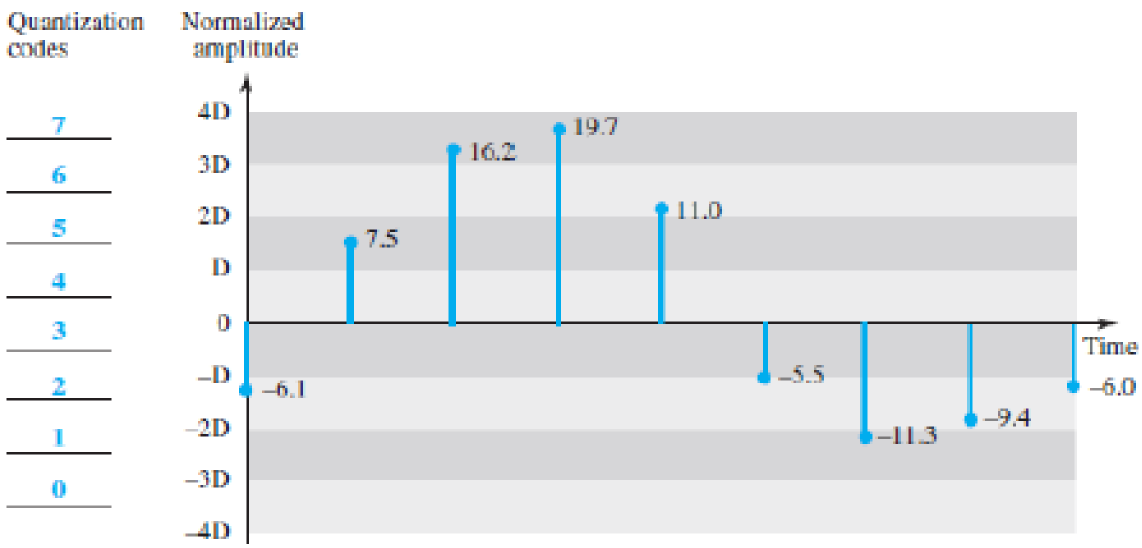
b. We divide the range into L zones, each of height Δ (delta).

$$\Delta = \frac{V_{max} - V_{min}}{L}$$

c. We assign quantized values of 0 to L − 1 to the midpoint of each zone.

d. We approximate the value of the sample amplitude to the quantized values.

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



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

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

Quantization Levels:

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

Quantization Error:

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

The quantization error changes the signal-to-noise ratio of the signal, which in turn reduces the upper limit capacity according to Shannon.

It can be proven that the contribution of the quantization error to the SNR_{dB} of the signal depends on the number of quantization levels L, or the bits per sample n_b, as shown in the following formula:

$$\mathrm{SNR_{dB}} = 6.02 n_b + 1.76 \quad \mathrm{dB}$$

Uniform Versus Nonuniform Quantization:

For many applications, the distribution of the instantaneous amplitudes in the analog signal is not uniform. Changes in amplitude often occur more frequently in the lower amplitudes than in the higher ones. For these types of applications it is better to use nonuniform zones. In other words, the height of Δ is not fixed; it is greater near the lower amplitudes and less near the higher amplitudes. Nonuniform quantization can also be achieved by using a process called companding and expanding. The signal is companded at the sender before conversion; it is expanded at the receiver after conversion. Companding means reducing the instantaneous voltage amplitude for large values; expanding is the opposite process. Companding gives greater weight to strong signals and less weight to weak ones. It has been proved that nonuniform quantization effectively reduces the SNRdB of quantization.

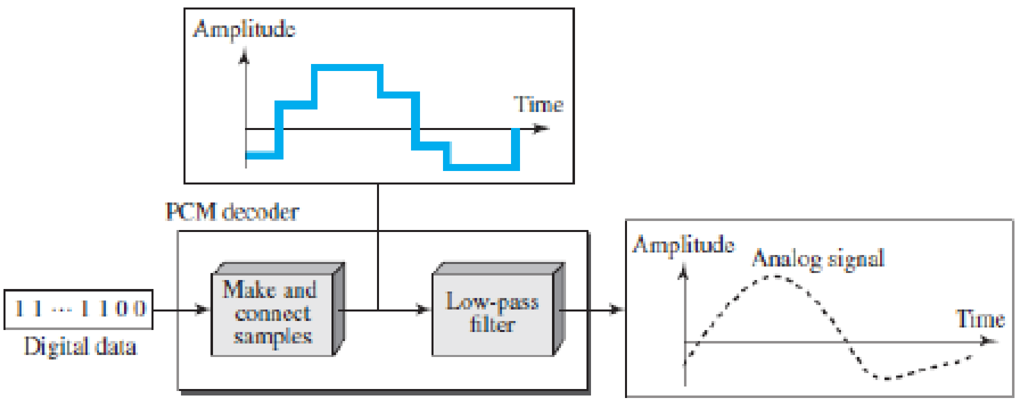
iii. Encoding

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

$$\text{Bit rate} = \text{sampling rate} \times \text{number of bits per sample} = f_s \times n_b$$

Original Signal Recovery:

The recovery of the original signal requires the PCM decoder. The decoder first uses circuitry to convert the code words into a pulse that holds the amplitude until the next pulse. After the staircase signal is completed, it is passed through a low-pass filter to smooth the staircase signal into an analog signal. The filter has the same cutoff frequency as the original signal at the sender. If the signal has been sampled at (or greater than) the Nyquist sampling rate and if there are enough quantization levels, the original signal will be recreated. The maximum and minimum values of the original signal can be achieved by using amplification. The below figure shows the simplified process.



PCM Bandwidth:

Suppose we are given the bandwidth of a low-pass analog signal. If we then digitize the signal, the new minimum bandwidth of the channel that can pass this digitized signal is obtained as follows. The minimum bandwidth of a line-encoded signal is $B_{min} = c \times N \times (1/r)$. We substitute the value of N in this formula:

$$B_{\text{min}} = c \times N \times \frac{1}{r} = c \times n_b \times f_s \times \frac{1}{r} = c \times n_b \times 2 \times B_{\text{analog}} \times \frac{1}{r}$$

When $1/r = 1$ (for a NRZ or bipolar signal) and $c = (1/2)$ (the average situation), the minimum bandwidth is

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

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

Maximum Data Rate of a Channel:

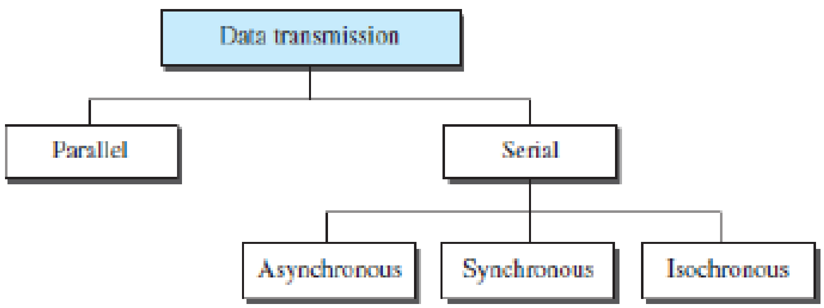
$$N_{\text{max}} = 2 \times B \times \log_2 L \quad \text{bps}$$

Minimum Required Bandwidth:

$$B_{\min} = \frac{N}{(2 \times \log_2) L} \quad \text{Hz}$$

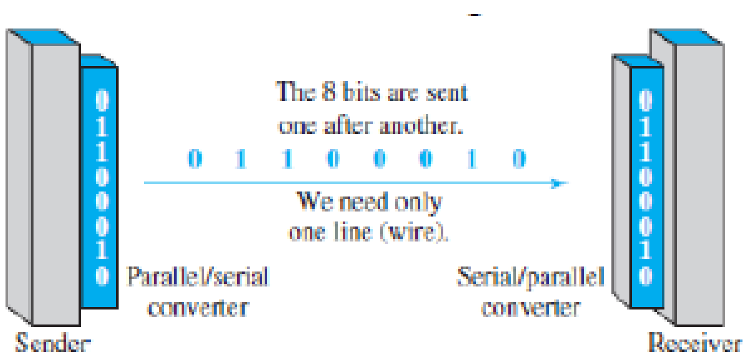
6(a). Differentiate between the 3 types of serial transmission. (6 marks)

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



Serial Transmission

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



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

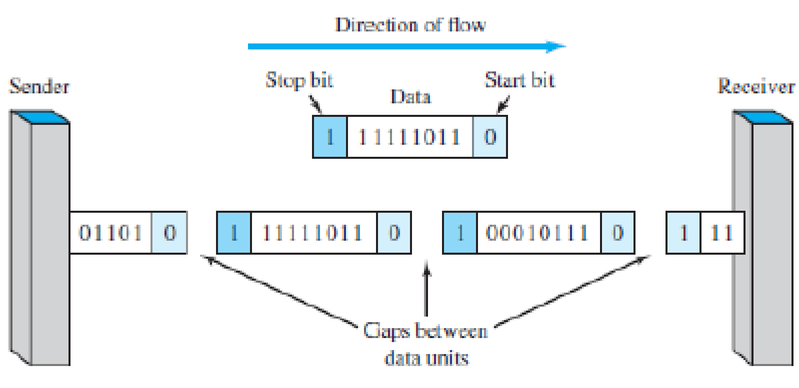
- Serial transmission occurs in one of three ways: asynchronous, synchronous, and isochronous.

i. Asynchronous Transmission

Asynchronous transmission is so named because the timing of a signal is unimportant. Instead, information is received and translated by agreed upon patterns. As long as those patterns are followed, the receiving device can retrieve the information without regard to the rhythm in which it is sent. Patterns are based on grouping the bit stream into bytes. Each group, usually 8 bits, is sent along the link as a unit. The sending system handles each group independently, relaying it to the link whenever ready, without regard to a timer.

Without synchronization, the receiver cannot use timing to predict when the next group will arrive. To alert the receiver to the arrival of a new group, therefore, an extra bit is added to the beginning of each byte. This bit, usually a 0, is called the start bit. To let the receiver know that the byte is finished, 1 or more additional bits are appended to the end of the byte. These bits, usually 1s, are called stop bits. By this method, each byte is increased in size to at least 10 bits, of which 8 bits is information and 2 bits or more are signals to the receiver. In addition, the transmission of each byte may then be followed by a gap of varying duration. This gap can be represented either by an idle channel or by a stream of additional stop bits.

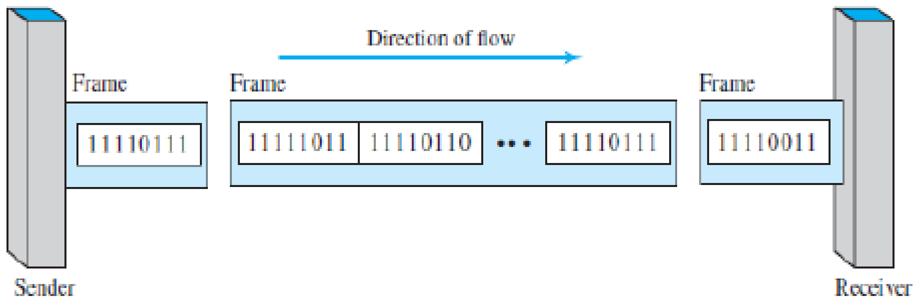
The start and stop bits and the gap alert the receiver to the beginning and end of each byte and allow it to synchronize with the data stream. This mechanism is called asynchronous because, at the byte level, the sender and receiver do not have to be synchronized. But within each byte, the receiver must still be synchronized with the incoming bit stream. That is, some synchronization is required, but only for the duration of a single byte. The receiving device resynchronizes at the onset of each new byte. When the receiver detects a start bit, it sets a timer and begins counting bits as they come in. After n bits, the receiver looks for a stop bit. As soon as it detects the stop bit, it waits until it detects the next start bit.



The above figure shows a schematic illustration of asynchronous transmission. In this example, the start bits are 0s, the stop bits are 1s, and the gap is represented by an idle line rather than by additional stop bits. The addition of stop and start bits and the insertion of gaps into the bit stream make asynchronous transmission slower than forms of transmission that can operate without the addition of control information. But it is cheap and effective, two advantages that make it an attractive choice for situations such as low-speed communication. For example, the connection of a keyboard to a computer is a natural application for asynchronous transmission. A user types only one character at a time, types extremely slowly in data processing terms, and leaves unpredictable gaps of time between characters.

ii. Synchronous Transmission

In synchronous transmission, the bit stream is combined into longer "frames," which may contain multiple bytes. Each byte, however, is introduced onto the transmission link without a gap between it and the next one. It is left to the receiver to separate the bit stream into bytes for decoding purposes. In other words, data are transmitted as an unbroken string of 1s and 0s, and the receiver separates that string into the bytes, or characters, it needs to reconstruct the information.



The above figure shows a schematic illustration of synchronous transmission. We have drawn in the divisions between bytes. In reality, those divisions do not exist; the sender puts its data onto the line as one long string. If the sender wishes to send data in separate bursts, the gaps between bursts must be filled with a special sequence of 0s and 1s that means idle. The receiver counts the bits as they arrive and groups them in 8-bit units.

Without gaps and start and stop bits, there is no built-in mechanism to help the receiving device adjust its bit synchronization midstream. Timing becomes very important, therefore, because the accuracy of the received information is completely dependent on the ability of the receiving device to keep an accurate count of the bits as they come in.

The advantage of synchronous transmission is speed. With no extra bits or gaps to introduce at the sending end and remove at the receiving end, and, by extension, with fewer bits to move across the link, synchronous transmission is faster than asynchronous transmission. For this reason, it is more useful for high-speed applications such as the transmission of data from one computer to another. Byte synchronization is accomplished in the data-link layer.

Although there is no gap between characters in synchronous serial transmission, there may be uneven gaps between frames.

iii. Isochronous

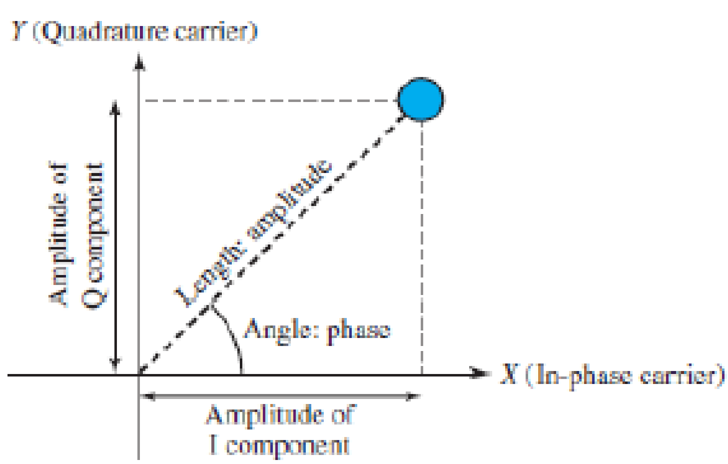
In real-time audio and video, in which uneven delays between frames are not acceptable, synchronous transmission fails. For example, TV images are broadcast at the rate of 30 images per second; they must be viewed at the same rate. If each image is sent by using one or more frames, there should be no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

6(b). Define constellation diagram. Show the constellation diagrams for ASK, BPSK and QPSK signals. (4 marks)

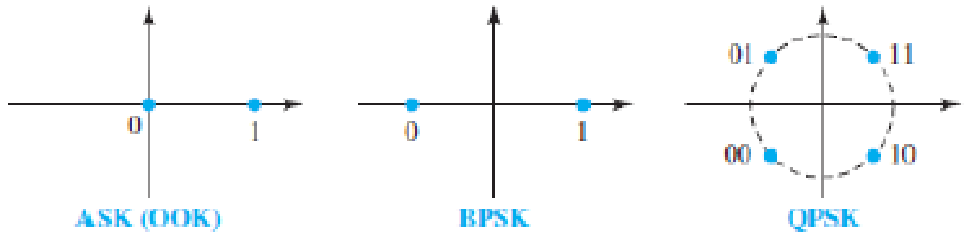
A constellation diagram can help us define the amplitude and phase of a signal element, particularly when we are using two carriers (one in-phase and one quadrature). The diagram is useful when we are dealing with multilevel ASK, PSK, or QAM. In a constellation diagram, a signal element type is represented as a dot. The bit or combination of bits it can carry is often written next to it.

The diagram has two axes. The horizontal X axis is related to the in-phase carrier; the vertical *Y* axis is related to the quadrature carrier. For each point on the diagram, four pieces of information can be deduced. The projection of the point on the X axis defines the peak amplitude of the in-phase component; the projection of the point on the *Y* axis defines the peak amplitude of the quadrature component. The length of the line (vector) that connects the point to the origin is the peak amplitude of the signal element (combination of the X and *Y* components); the angle the line makes with the X axis is the phase of the signal element. All the information we need, can easily be found on a constellation diagram. Figure below shows a constellation diagram.

Concept of a constellation diagram



Three constellation diagrams.



a. For ASK, we are using only an in-phase carrier. Therefore, the two points should be on the X axis. Binary 0 has an amplitude of 0V; binary 1 has an amplitude of 1V (for example). The points are located at the origin and at 1 unit.

b. BPSK also uses only an in-phase carrier. However, we use a polar NRZ signal for modulation. It creates two types of signal elements, one with amplitude 1 and the other with amplitude -1. This can be stated in other words: BPSK creates two different signal elements, one with amplitude 1 V and in phase and the other with amplitude 1V and 180° out of phase.

c. QPSK uses two carriers, one in-phase and the other quadrature. The point representing 11 is made of two combined signal elements, both with an amplitude of 1V. One element is represented by an in-phase carrier, the other element by a quadrature carrier. The amplitude of the final signal element sent for this 2- bit data element is $2^{1/2}$, and the phase is *45°*. The argument is similar for the other three points. All signal elements have an amplitude of $2^{1/2}$, but their phases are different *(45°*, 135°, -135°, and -45°). Of course, we could have chosen the amplitude of the carrier to be $1/(2^{1/2})$ to make the final amplitudes 1V.

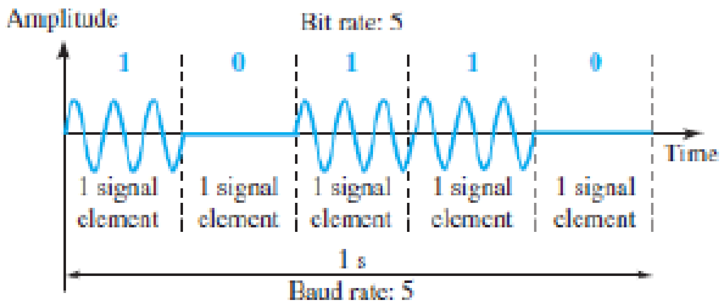
7. Analyze ASK, FSK and PSK mechanisms applying them over the digital data 101101. (10 marks)

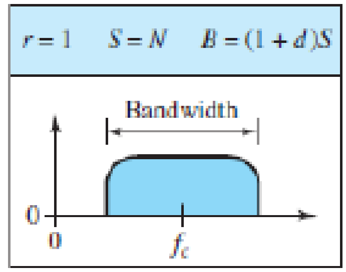
Amplitude Shift Keying

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

Binary ASK (BASK)

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





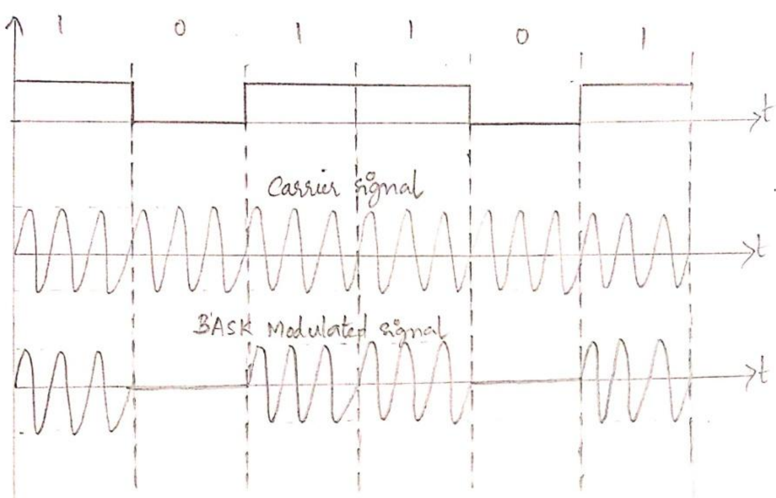
<u>Bandwidth for ASK</u>

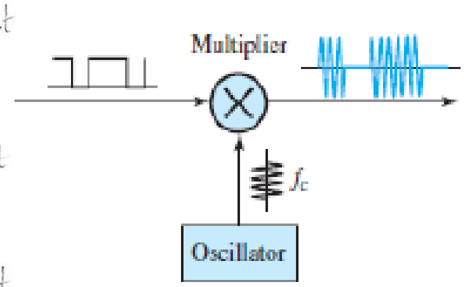
- The above figure also shows the bandwidth for ASK. Although the carrier signal is only one simple sine wave, the process of modulation produces a nonperiodic composite signal. As we expect, the bandwidth is proportional to the signal rate (baud rate). However, there is normally another factor involved, called *d*, which depends on the modulation and filtering process. The value of *d* is between 0 and 1. This means that the bandwidth can be expressed as shown, where 5 is the signal rate and the *B* is the bandwidth.

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

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

The next figure shows how we can simply implement binary ASK.





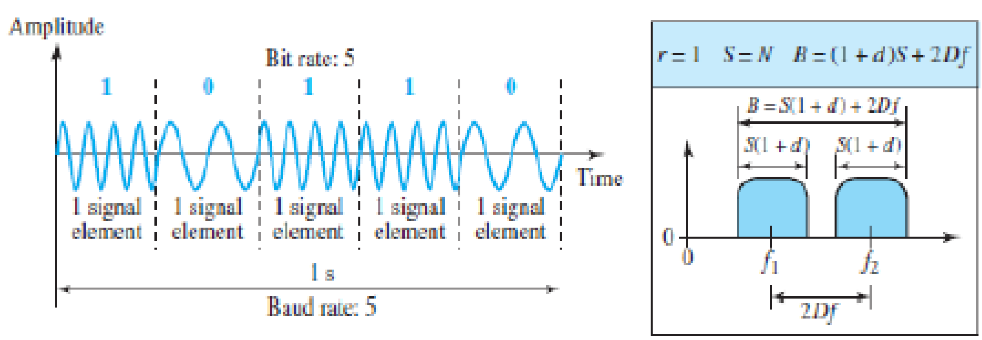
If digital data are presented as a unipolar NRZ digital signal with a high voltage of 1V and a low voltage of 0V, the implementation can achieved by multiplying the NRZ digital signal by the carrier signal coming from an oscillator. When the amplitude of the NRZ signal is 1, the amplitude of the carrier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

<u>Frequency Shift Keying</u>

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

<u>Binary FSK (BFSK)</u>

One way to think about binary FSK (or BFSK) is to consider two carrier frequencies. In Figure below, we have selected two carrier frequencies, f_1 and f_2. We use the first carrier if the data element is 0; we use the second if the data element is 1.



The middle of one bandwidth is f_1 and the middle of the other is f_2. Both f_1 and f_2 are Δf apart from the midpoint between the two bands. The difference between the two frequencies is $2\Delta f$.

<u>Bandwidth for BFSK</u>

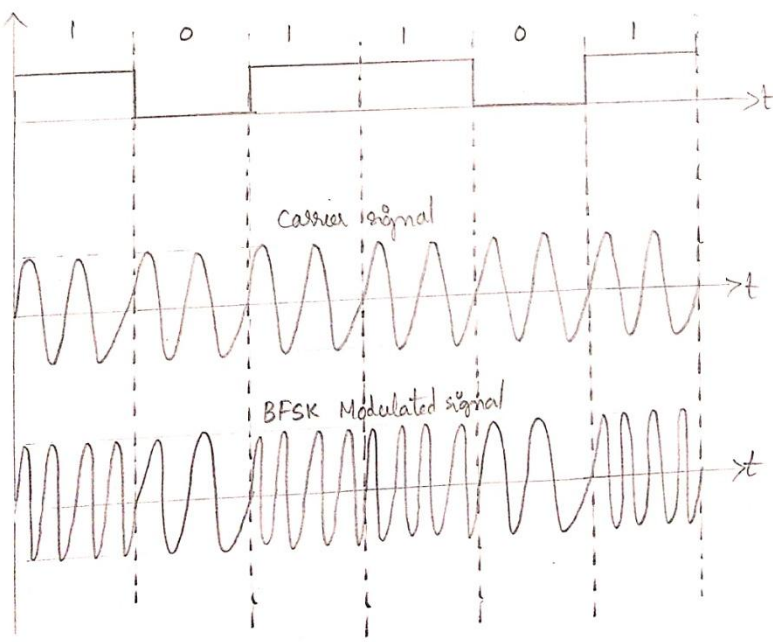
Figure also shows the bandwidth of FSK. Again the carrier signals are only simple sine waves, but the modulation creates a nonperiodic composite signal with continuous frequencies. We can think of FSK as two ASK signals, each with its own carrier frequency f_1 and f_2. If the difference between the two frequencies is $2\Delta f$, then the required bandwidth is

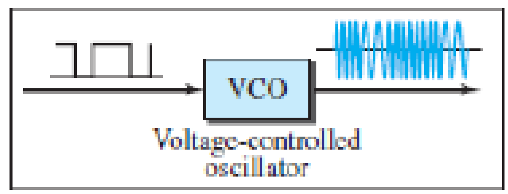
$$B = (l+d) \times S + 2\Delta f.$$

The minimum value of B should be at least S for the proper operation of modulation and demodulation.

Implementation

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



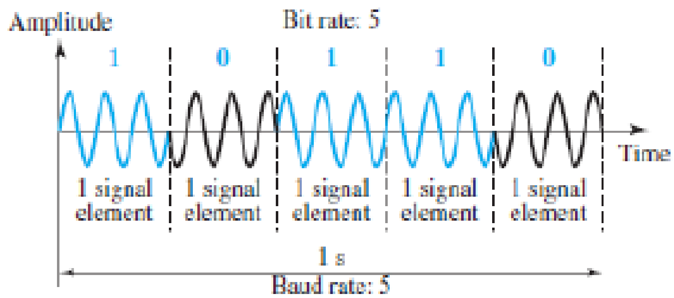


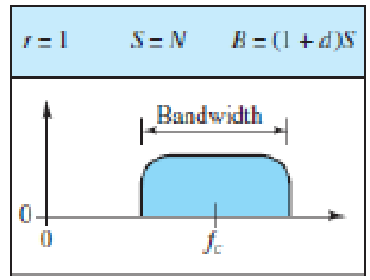
Phase Shift Keying

In phase shift keying, the phase of the carrier is varied to represent two or more different signal elements. Both peak amplitude and frequency remain constant as the phase changes. Today, PSK is more common than ASK or FSK. QAM, which combines ASK and PSK, is the dominant method of digital-to- analog modulation.

Binary PSK (BPSK)

The simplest PSK is binary PSK, in which we have only two signal elements, one with a phase of 0°, and the other with a phase of 180°. Figure below gives a conceptual view of PSK. Binary PSK is as simple as binary ASK with one big advantage-it is less susceptible to noise.





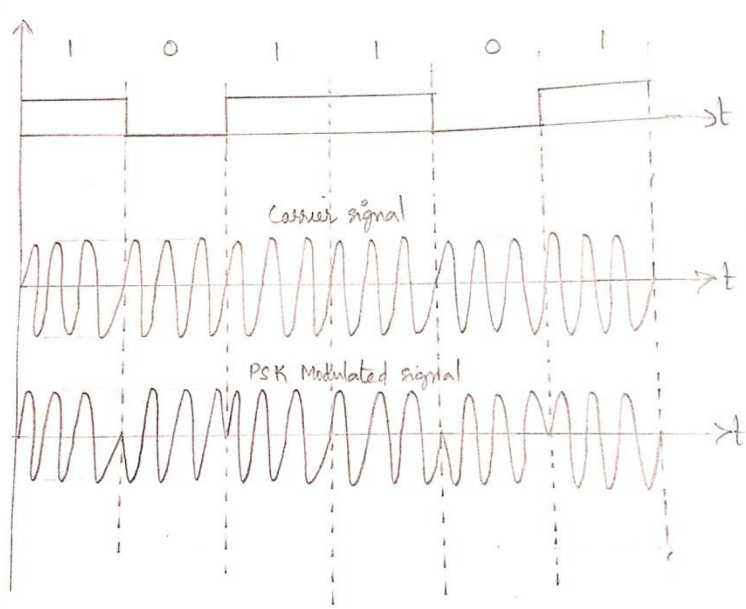
In ASK, the criterion for bit detection is the amplitude of the signal; in PSK, it is the phase. Noise can change the amplitude easier than it can change the phase. In other words, PSK is less susceptible to noise than ASK.PSK is superior to FSK because we do not need two carrier signals.

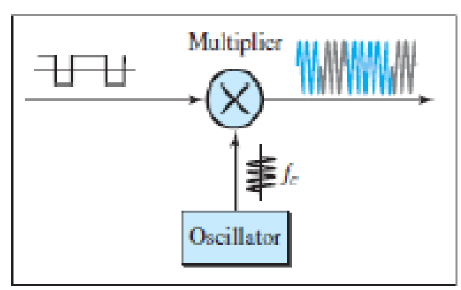
<u>Bandwidth</u>

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

<u>Implementation</u>

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

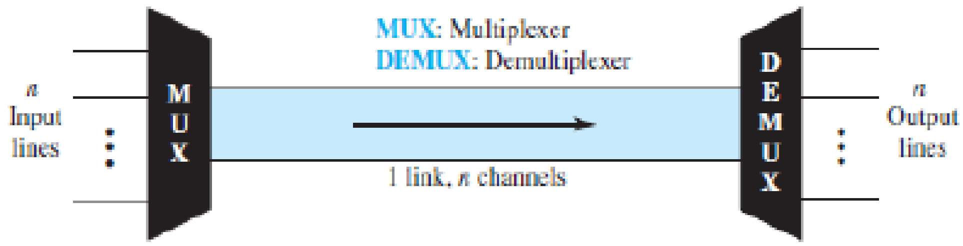




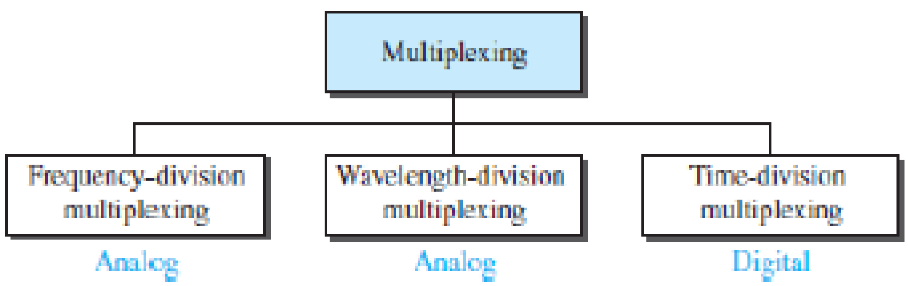
8(a). What is FDM? Briefly explain its multiplexing and demultiplexing process. (6 marks)

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

In a multiplexed system, n lines share the bandwidth of one link. Figure shows the basic format of a multiplexed system. The lines on the left direct their transmission streams to a multiplexer (MUX), which combines them into a single stream (many-to-one). At the receiving end, that stream is fed into a demultiplexer (DEMUX), which separates the stream back into its component transmissions (one-to-many) and directs them to their corresponding lines. In the figure, the word link refers to the physical path. The word channel refers to the portion of a link that carries a transmission between a given pair of lines. One link can have many (n) channels.

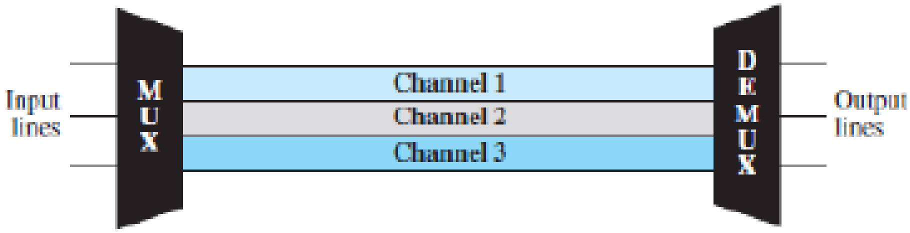


There are three basic multiplexing techniques: frequency-division multiplexing, wavelength-division multiplexing, and time-division multiplexing. The first two are techniques designed for analog signals, the third, for digital signals



Frequency-Division Multiplexing

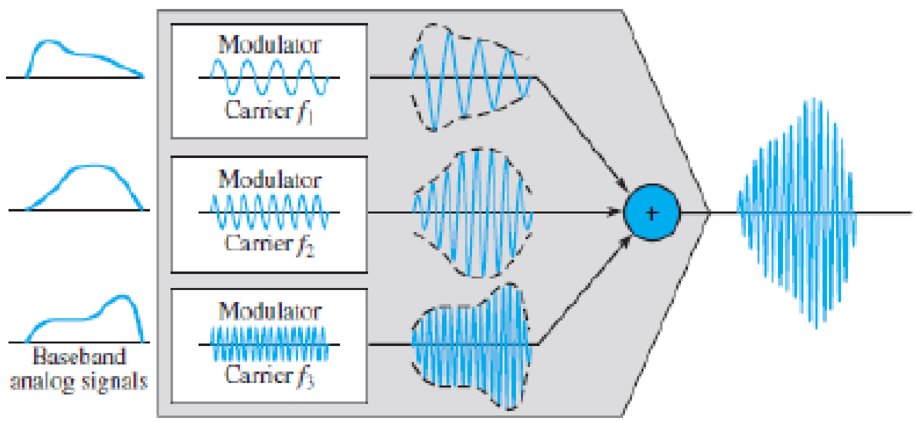
Frequency-division multiplexing (FDM) is an analog technique that can be applied when the bandwidth of a link (in hertz) is greater than the combined bandwidths of the signals to be transmitted. In FDM, signals generated by each sending device modulate different carrier frequencies. These modulated signals are then combined into a single composite signal that can be transported by the link. Carrier frequencies are separated by sufficient bandwidth to accommodate the modulated signal. These bandwidth ranges are the channels through which the various signals travel. Channels can be separated by strips of unused bandwidth-guard bands-to prevent signals from overlapping. In addition, carrier frequencies must not interfere with the original data frequencies. Figure gives a conceptual view of FDM. In this illustration, the transmission path is divided into three parts, each representing a channel that carries one transmission.



We consider FDM to be an analog multiplexing technique; however, this does not mean that FDM cannot be used to combine sources sending digital signals. A digital signal can be converted to an analog signal before FDM is used to multiplex them.

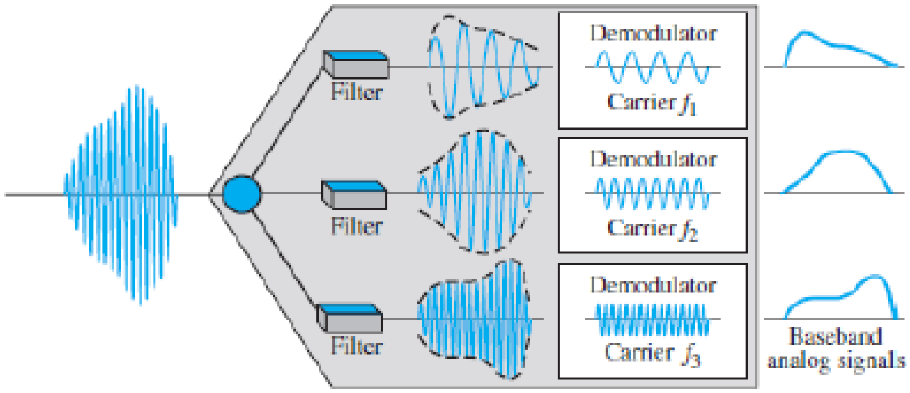
Multiplexing Process

The next figure is a conceptual illustration of the multiplexing process. Each source generates a signal of a similar frequency range. Inside the multiplexer, these similar signals modulate different carrier frequencies (f_1, f_2 and f_3). The resulting modulated signals are then combined into a single composite signal that is sent out over a media link that has enough bandwidth to accommodate it.



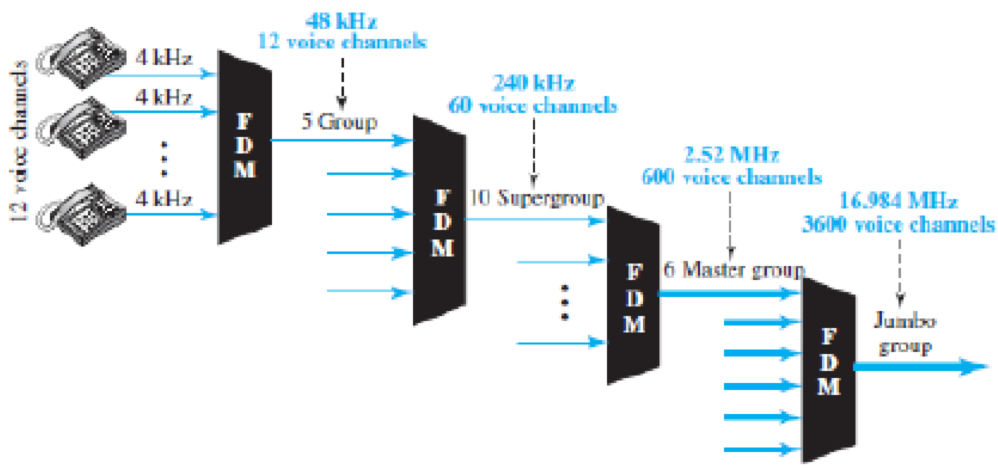
Demultiplexing Process

The demultiplexer uses a series of filters to decompose the multiplexed signal into its constituent component signals. The individual signals are then passed to a demodulator that separates them from their carriers and passes them to the output lines. The next figure is a conceptual illustration of demultiplexing process.



8(b). Discuss the analog hierarchy system used in FDM. (4 marks)

To maximize the efficiency of their infrastructure, telephone companies have traditionally multiplexed signals from lower-bandwidth lines onto higher-bandwidth lines. In this way, many switched or leased lines can be combined into fewer but bigger channels. For analog lines, FDM is used. One of these hierarchical systems used by AT&T is made up of groups, super groups, master groups, and jumbo groups



In this analog hierarchy, 12 voice channels are multiplexed onto a higher-bandwidth line to create a group. A group has 48 kHz of bandwidth and supports 12 voice channels.

At the next level, up to five groups can be multiplexed to create a composite signal called a supergroup. A supergroup has a bandwidth of 240 kHz and supports up to 60 voice channels. Supergroups can be made up of either five groups or 60 independent voice channels.

At the next level, 10 supergroups are multiplexed to create a master group. A master group must have 2.40 MHz of bandwidth, but the need for guard bands between the supergroups increases the necessary bandwidth to 2.52 MHz. Master groups support up to 600 voice channels.

Finally, six master groups can be combined into a jumbo group. A jumbo group must have 15.12 MHz (6 x 2.52 MHz) but is augmented to 16.984 MHz to allow for guard bands between the master groups.