Communication Engineering

Table of Contents

Reference book: Behrouz A. Forouzan: Data Communications and Networking (4th & 5th Edition)

Chapter 1: Introduction

1.1 DATA COMMUNICATIONS

Q. Define data communication.

Data communication is the process of exchanging data between two or more devices through a transmission medium, such as wire cables or wireless signals.

Q. Briefly describe the fundamental characteristics of data communication system.

The fundamental characteristics of a data communication system are (or, The effectiveness of a data communications system depends on four fundamental characteristics: delivery, accuracy, timeliness, and jitter):

1. **Delivery:** The system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.
2. **Accuracy:** The system must deliver the data accurately. Data that have been altered in transmission and left uncorrected are unusable.
3. **Timeliness:** The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called real-time transmission.
4. **Jitter:** Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, if video packets arrive with inconsistent delays, it causes noticeable distortions in playback.

1.1.1 Components

Q. Describe a simple data communication system with diagram.

A data communications system has five components.

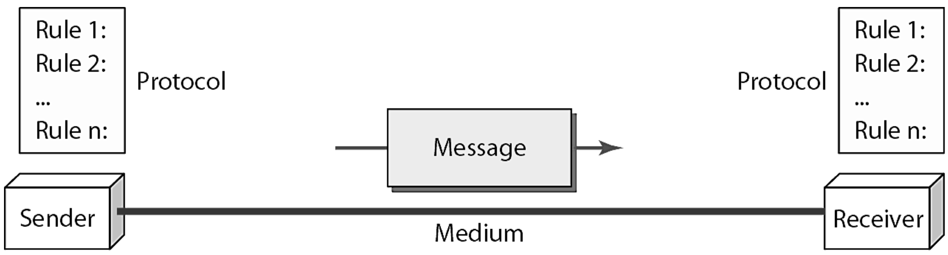


Figure 1.1 Five components of data communication

1. **Message:** The message is the information (data) to be communicated. Popular forms of information include text, pictures, audio, and video.
2. **Sender:** The sender is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.
3. **Receiver:** The receiver is the device that receives the message. It can be a computer, workstation, telephone handset, television, and so on.
4. **Transmission medium:** The transmission medium is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves.
5. **Protocol:** A protocol is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not communicating, just as a person speaking French cannot be understood by a person who speaks only Japanese.

1.1.3 Data Flow / Data Transmission Modes

Communication between two devices can be simplex, half-duplex, or full-duplex as shown in Figure 1.2.

or, Data transmission modes describe how data flows between devices in a communication system. There are three primary types of data transmission modes: Simplex, Half-Duplex, and Full-Duplex.

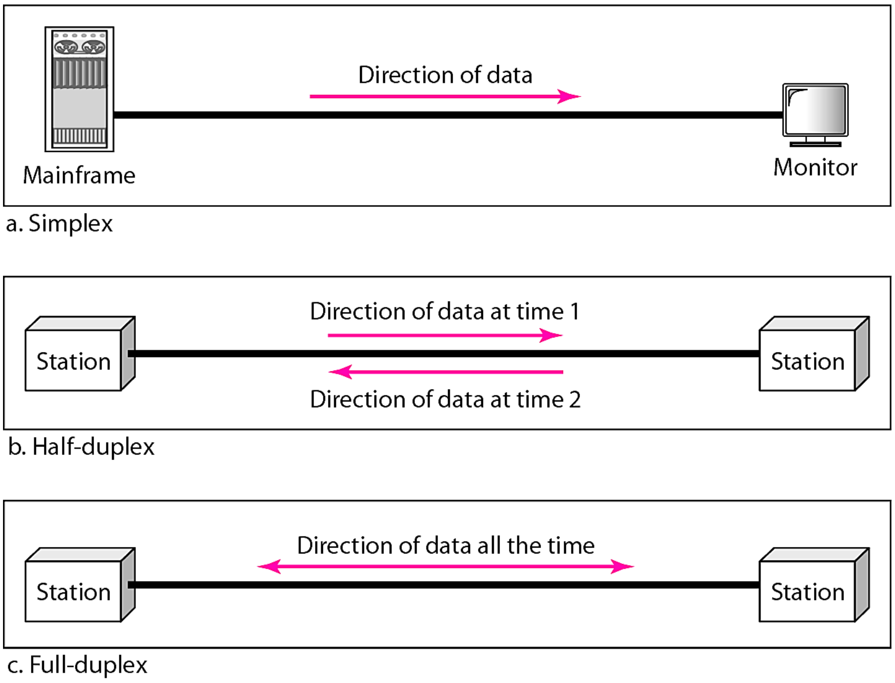


Figure 1.2 Data flow (simplex, half-duplex, and full-duplex)

Simplex

In simplex mode, the communication is unidirectional, meaning data flows only in one direction, similar to a one-way street. Only one of the two devices on a link can transmit; the other can only receive (see Figure 1.2a).

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

Half-Duplex

In half-duplex mode, the communication is bidirectional but not simultaneous. This means that data can flow in both directions, but only one direction at a time. When one device is sending, the other can only receive, and vice versa (see Figure 1.2b). Push-to-talk (PTT) communication systems, such as walkie-talkies and CB (citizens band) radios are typical examples of half-duplex communication.

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.

Full-Duplex

In full-duplex mode (also called duplex), data transmission is bidirectional and simultaneous. This means that data can flow in both directions at the same time, i.e., both devices can transmit and receive at the same time. (see Figure 1.2c).

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

The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel, however, must be divided between the two directions.

Topic Related Questions:

1. Explain different types of data transmission modes with proper diagrams and examples.
2. What do you mean by i) Simplex ii) Half duplex and (ii) Full duplex transmission?

1.2 NETWORKS

1.2.1 Network Criteria

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

Performance

Performance can be measured in many ways, including transit time and response time. Transit time is the amount of time required for a message to travel from one device to another. Response time is the elapsed time between an inquiry and a response. The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software.

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

Reliability

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

Security

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

1.2.2 Physical Structures

Physical Topology

The topology refers to the physical layout or arrangement of the network components, such as computers, servers, routers, switches, and links. 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 (see Figure 1.4).

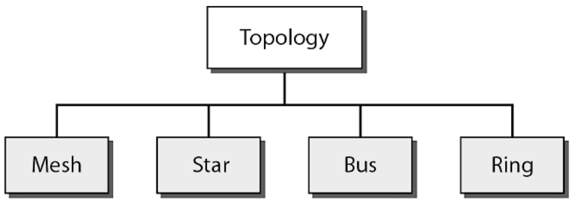


Figure1.4 Categories of topology

Mesh Topology

In a mesh topology, every device has a dedicated point-to-point link to every other device. The term dedicated means that the link carries traffic only between the two devices it connects. To establish such a topology, we need $n(n-1)/2$ duplex-mode links, where n is the number of devices. Additionally, each device needs $n-1$ I/O ports (see Figure 1.5) to connect with the other $n-1$ devices.

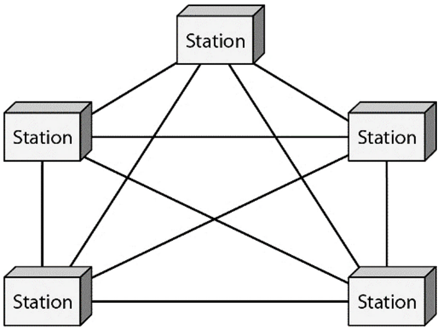


Figure 1.5 A fully connected mesh topology (five devices)

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

Advantages

A mesh offers several advantages over other network topologies.

1. 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.
2. A mesh topology is robust. It remains operational even if a single link fails, as alternative paths exist.
3. The use of dedicated links provides a high level of privacy and security for data transmission.
4. Point-to-Point links make fault identification and fault isolation easy.

Disadvantages

1. The main disadvantages of a mesh are related to the amount of cabling and the number of I/O ports required.
2. Since every device must be connected to every other device, installation and reconnection are difficult.
3. The need for a large number of I/O ports and cables increases hardware costs, especially in larger networks.

Star Topology

In a star topology, each device has a dedicated point-to-point link only to a central controller, usually called a hub. The devices are not directly connected to one another. Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data to the controller, which then forwards the data to the other connected device (see Figure 1.6).

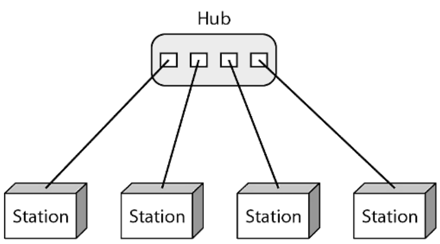


Figure 1.6 A star topology connecting four stations

The star topology is used in local-area networks (LANs). High-speed LANs often use a star topology with a central hub.

Advantages

1. A star topology is less expensive than a mesh topology due to its reduced cabling.
2. In a star, each device needs only one link and one I/O port to connect it to any number of others. This factor makes it easy to install and reconfigure.
3. A star topology is robust. If one link fails, only that link is affected. All other links remain active. This factor lends itself to easy fault identification and fault isolation.

Disadvantages

- One big disadvantage of a star topology is the dependency of the whole topology on one single point, the hub. If the hub goes down, the whole system is dead.
- Although a star requires far less cable than a mesh, each node must be linked to a central hub. For this reason, often more cabling is required in a star than in some other topologies (such as ring or bus).

Bus Topology

In a bus topology, all devices are connected to a single shared communication line, known as a bus. Devices are connected to the bus cable via drop lines and taps. Data is transmitted along the bus, and each device receives and processes the data intended for it.

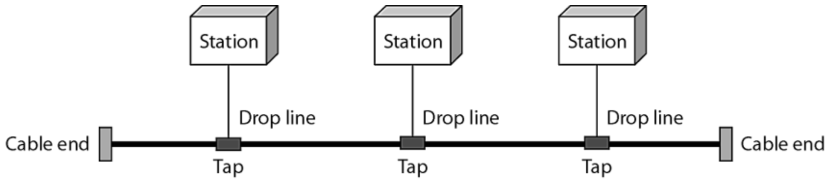


Figure 1.7 A bus topology connecting three stations

Bus topology is famously used for the Local Area Network (LANs).

Advantages

- A bus topology uses less cabling than mesh or star topologies.
- It operates efficiently when there is a small network.
- Very cost-effective as compared to other network topology, i.e., mesh and star.

Disadvantages

- Bus topology is not great for large networks.
- If the main cable is damaged, the whole network fails or splits into two.
- Identification of problems becomes difficult if the whole network goes down.

Ring Topology

In a ring topology, each device is connected in a closed loop, with a dedicated point-to-point connection to the two devices adjacent to it. Data travels in a single direction along the ring, moving from one device to the next until it reaches its destination. Each device in the ring incorporates a repeater to regenerate and forward data bits (see Figure 1.8).

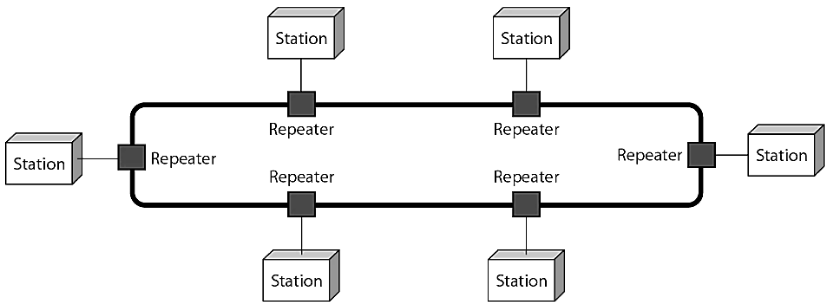


Figure 1.8 A ring topology connecting six stations

Advantages

- A ring is relatively easy to install and reconfigure.
- Speed to transfer the data is very high in this type of topology.

Disadvantages

- Due to the unidirectional traffic, a data packet (token) must have to pass through all the nodes.
- Failure of a single device or connection can disable the entire network.

A network can be hybrid. For example, we can have a main star topology with each branch connecting several stations in a bus topology as shown in Figure 1.9.

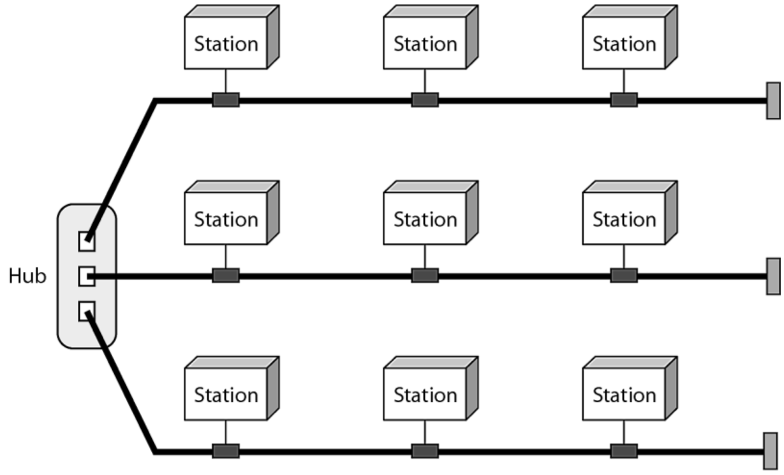


Figure 1.9 A hybrid topology: a star backbone with three bus networks

Topic Related Questions:

1. Classify network topologies. Draw block diagram, and write the advantages and disadvantages of each type.
2. Draw the diagram of four basic network topologies and write the advantages of each type.

1.4 PROTOCOLS AND STANDARDS

Q. Define protocol and protocol layering.

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

- **Syntax:** The term syntax refers to the structure or format of the data, meaning the order in which they are presented.
- **Semantics:** The word semantics refers to the meaning of each section of bits. How is a particular pattern to be interpreted, and what action is to be taken based on that interpretation?
- **Timing:** The term timing refers to two characteristics: when data should be sent and how fast they can be sent.

Protocol layering: Protocol layering is a design principle used in network communication systems where complex communication processes are divided into hierarchical layers, each with specific functions and responsibilities. Each layer serves the layer above it and is served by the layer below it.

Chapter 2: Network Models

2.2 THE OSI MODEL

Established in 1947, the International Standards Organization (ISO) is a multinational body dedicated to worldwide agreement on international standards. An ISO standard that covers all aspects of network communications is the Open Systems Interconnection (OSI) model. The OSI model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust, and interoperable.

The OSI model is a layered framework for the design of network systems that allows communication between all types of computer systems. It consists of seven separate but related layers, each of which defines a part of the process of moving information across a network (see Figure 2.2).

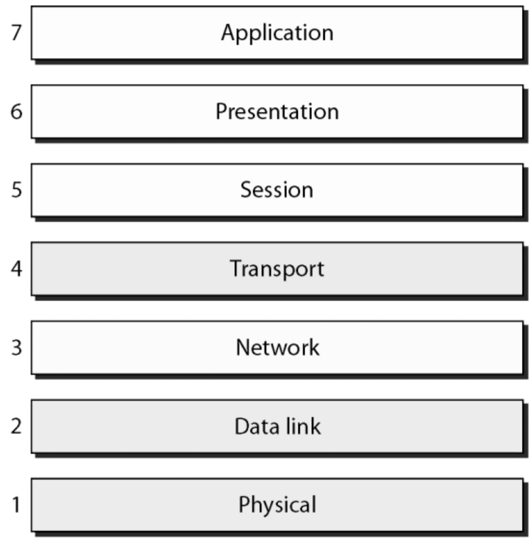


Figure 2.2 Seven layers of the OSI model

The OSI model is composed of seven ordered layers: physical (layer 1), data link (layer2), network (layer 3), transport (layer 4), session (layer 5), presentation (layer 6), and application (layer 7). Here are the functionalities of each layer of the OSI model:

1. **Physical Layer (Layer 1):** The main functionality of the physical layer is to transmit the individual bits from one node to another node. This layer defines the hardware, cabling wiring, power output, pulse rate etc.
2. **Data Link Layer (Layer 2):** This layer is responsible for data framing, error detection, and flow control.
3. **Network Layer (Layer 3):** The Network Layer handles routing and forwarding of data packets between different networks. This layer is responsible for logical addressing, such as IP addresses, and determines the best path for data to reach its destination.
4. **Transport Layer (Layer 4):** The transport layer is responsible for process-to-process delivery of the entire message. It is responsible for segmenting data into smaller chunks, managing flow control, and providing error recovery through retransmission.
5. **Session Layer (Layer 5):** The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems.
6. **Presentation Layer (Layer 6):** The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems. This layer is responsible for translation, compression, and encryption.
7. **Application Layer (Layer 7):** The application layer is responsible for providing services to the user.

Topic Related Questions:

1. Write a short note on ISO reference model.

2.4 TCP/IP PROTOCOL SUITE

TCP/IP stands for Transmission Control Protocol/Internet Protocol and is a suite of communication protocols used to interconnect network devices on the internet. 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 one or more lower-level protocols. The original TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application. Today, however, TCP/IP is thought of as a five layer model: physical, data link, network, transport, and application. Below describes each layer of TCP/IP model.

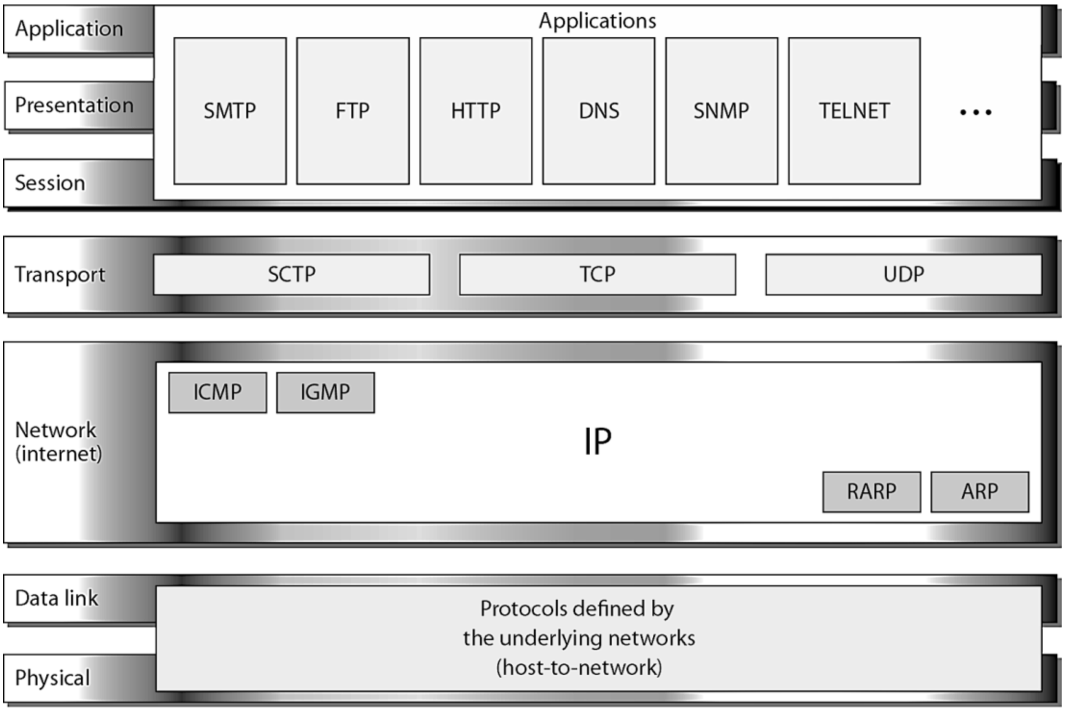


Figure 2.16 TCP/IP and OSI model

Physical Layer

The physical layer is the lowest layer in the TCP/IP protocol suite. It is responsible for carrying individual bits in a frame across the link and requesting connections.

Data Link Layer

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. This layer is also responsible for the reliable and error-free transfer of data frames between two directly connected devices over a physical link. The data link layer takes a datagram and encapsulates it in a packet called a frame. TCP/IP does not define any specific protocol for the data link layer. It supports all the standard and proprietary protocols.

Network Layer

The network link layer, also called the internet layer, is responsible for creating a connection between the source computer and the destination computer. This layer is also responsible for specifying the path that the data packets will use for transmission. The communication at the network layer is host-to-host. The main protocol in this layer is Internet Protocol (IP) and it is supported by the protocols ICMP, IGMP, RARP, and ARP.

Transport Layer

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. Protocols used in this layer are TCP and UDP. The main protocol, Transmission Control Protocol (TCP), is a connection-oriented protocol that handles communications between hosts and provides flow control, multiplexing and reliability. User Datagram Protocol (UDP) is a connectionless protocol that transmits user datagrams without first creating a logical connection.

Application Layer

The application layer is the topmost layer in the TCP/IP model. Communication at the application layer is between two processes. 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. This layer uses various protocols to transfer the data between applications. Some standard protocols used in this layer are: TELNET, DNS, HTTP, FTP, SMTP, SNMP, etc.

Q. Discuss in short about the services of data link layer.

The Data Link Layer is responsible for reliable data transfer between nodes on a network. It provides the following key services:

1. **Framing:** The data link layer divides the stream of bits received from the network layer into manageable data units called frames.
2. **Error Detection and Correction:** The layer uses mechanisms like checksums, cyclic redundancy checks (CRC), and acknowledgment/retransmission techniques to detect and correct errors in data transmission.
3. **Flow Control:** Manages the rate of data transmission between sender and receiver to prevent congestion.
4. **Physical Addressing:** Adds a header containing the MAC (Media Access Control) address of the sender and receiver to ensure proper delivery of frames within the local network.
5. **Access Control:** In shared network environments, the data link layer determines how devices access the shared medium to avoid collisions.

Q. Discuss in short about the services of Network layer.

The Network Layer is responsible for routing data packets from the source to the destination across different networks. It provides the following key services:

1. **Logical Addressing:** Assigns and uses IP addresses to uniquely identify devices on the network.
2. **Routing:** Determines the best path for data to travel from the source to the destination across multiple networks.
3. **Packet Forwarding:** Moves packets from the source toward the destination, either directly or via intermediate routers.
4. **Fragmentation and Reassembly:** Breaks down large packets into smaller fragments to match the Maximum Transmission Unit (MTU) of the network and reassembles them at the destination.
5. **Error Handling and Diagnostics:** Provides tools for error reporting (e.g., using ICMP messages for destination unreachable, time exceeded, etc.) and diagnostics (e.g., ping and traceroute utilities).

Q. Write the services of application layer.

The Application layer in the TCP/IP model provides a variety of services to users. Here are some of the key services:

1. **File transfer, access, and management:** This service enables users to access files on remote systems, make modifications, read data, retrieve files to a local machine, and manage files stored on remote systems. Example Protocols: File Transfer Protocol (FTP), Trivial File Transfer Protocol (TFTP).
2. **Mail services:** This service facilitates the forwarding and storage of emails. It provides the foundation for email communication across devices. Example Protocols: Simple Mail Transfer Protocol (SMTP), Post Office Protocol (POP), Internet Message Access Protocol (IMAP).

3. **Directory Services:** This service provides distributed databases that store and provide access to global information about various objects and services on the network. Example Protocols: Lightweight Directory Access Protocol (LDAP), Domain Name System (DNS).
4. **Web Browsing and Hypertext Access:** This service enables web browsing and communication between web browsers and web servers. Example Protocols: Hypertext Transfer Protocol (HTTP) and its secure variant (HTTPS).

Q. Compare TCP/IP and OSI model.

Comparison between the TCP/IP and OSI model:

TCP/IP Model	OSI Model
TCP/IP stands for Transmission Control Protocol/Internet Protocol.	OSI stands for Open Systems Interconnection.
The TCP/IP model has five layers: physical, data link, network, transport, and application.	The OSI model consists of seven layers: Application, Presentation, Session, Transport, Network, Data Link, and Physical.
It is protocol dependent.	It is protocol independent.
In this model, the session and presentation layer are not different layers. Both layers are included in the application layer.	In this model, the session and presentation layers are separated, i.e., both the layers are different.
In TCP/IP model, the transport layer does not guarantees delivery of packets.	In OSI model, the transport layer guarantees the delivery of packets.
It is more reliable than OSI Model.	It is less reliable than TCP/IP Model.
This model is highly used.	The usage of this model is very low.
Used in modern networks, especially the Internet.	Used primarily as a reference model for designing networks.

2.5 ADDRESSING

Four levels of addresses are used in an internet employing the TCP/IP protocols: physical (link) addresses, logical (IP) addresses, port addresses, and specific addresses (see Figure 2.17).

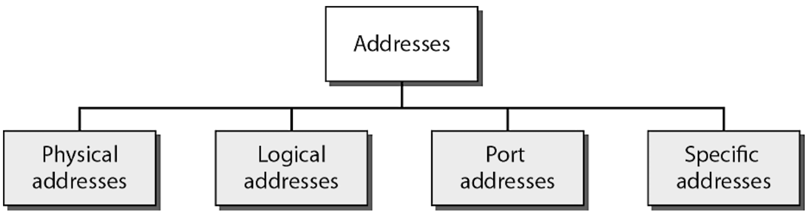


Figure 2.17 Addresses in TCP/IP

Q. Define i) Physical address ii) Logical address and iii) Port address.

Physical Address: The physical address, also known as the link address, is the unique hardware address of a device on a local network. It operates at the Data Link Layer and is used for communication within a local network segment.

Logical Address: A logical address is an identifier used to uniquely identify a device on a network and allow communication across different networks. It operates at the Network Layer.

Port Address: A port address identifies a specific application or service on a device. It operates at the Transport Layer and allows multiple applications to communicate simultaneously on the same device.

Chapter 3: Data and Signals

Q. Define bit rate, baud rate and bandwidth.

Bit Rate: Bit rate refers to the number of bits transmitted/sent per second in a communication channel. It is measured in bits per second (bps). Bit rate determines the speed of data transfer. The bit rate is sometimes called the data rate.

Baud Rate: Baud rate is the number of signal units/elements transmitted per second in a communication channel. Each signal unit may carry one or more bits, depending on the modulation technique. Baud rate is measured in baud. The baud rate is sometimes called the pulse rate, the modulation rate, or the signal rate.

$$Bit\ Rate = Baud\ Rate \times Number\ of\ Bits\ per\ Signal\ Unit$$

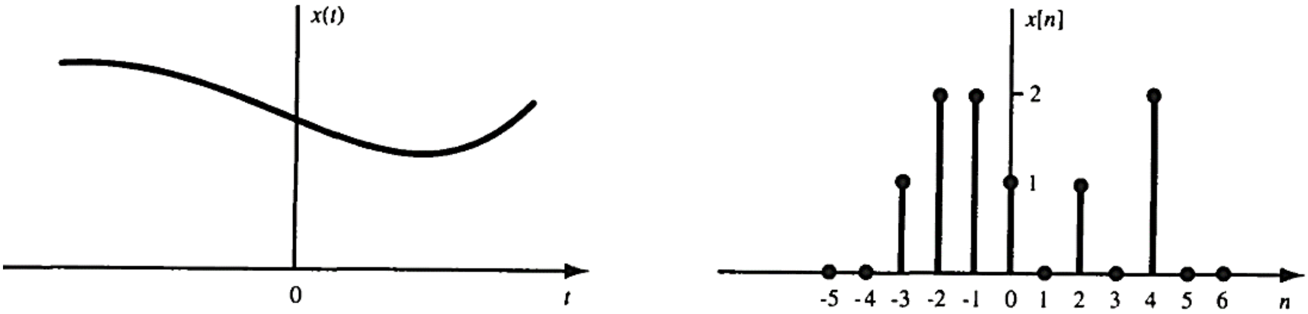
Bandwidth: Bandwidth is the range of frequencies that a communication channel can support for data transmission. It is calculated as the difference between the highest and the lowest frequencies of a signal that a communication channel can transmit without significant loss. It is measured in Hertz (Hz). Wider bandwidth allows more data to be transmitted simultaneously.

Q. Define continuous time signal and discrete time signal.

A **continuous-time signal** is a signal that is defined for every instant of time. It is represented as a continuous function of time $x(t)$, where t is a continuous variable. For example, a sin wave, $x(t) = A\sin(2\pi ft)$, is a continuous-time signal.

A **discrete-time signal** is a signal that is defined only at discrete points in time. In other words, if t is a discrete variable, that is, $x(t)$ is defined at discrete times, then $x(t)$ is a discrete-time signal. Since a discrete-time signal is defined at discrete times, a discrete-time signal is often represented as a sequence of numbers, denoted by $\{x_n\}$ or $x[n]$, where n = integer.

Illustrations of a continuous-time signal $x(t)$ and of a discrete-time signal $x[n]$ are shown in Fig. 3-1.



(a) (b)

Fig. 3-1 Graphical representation of (a) continuous-time and (b) discrete-time signals.

Q. What is your concept about deterministic signal and random signal?

Deterministic signals are those signals whose values are completely specified for any given time. Thus, a deterministic signal can be completely described mathematically by a known function of time t. For example, a sin wave, $x(t) = A\sin(2\pi ft)$.

Random signals (or stochastic signals) are those signals that take random values at any given time. These signals can be described statistically but not exactly. For example, white noise: a signal with equal intensity at different frequencies and random amplitudes.

Q. Explain how we obtain discrete time signal from continuous time signal.

We obtain a discrete-time signal from a continuous-time signal $x(t)$ through a process called sampling. Sampling is the process of measuring the amplitude of a continuous-time signal at regular time intervals. The rate at which the continuous signal is sampled is called the sampling rate (or sampling frequency) and is denoted by f_s. It is measured in samples per second (Hz). The time interval between successive samples is the sampling interval, T_s, which is the reciprocal of the sampling rate.

$$T_s = \frac{1}{f_s}$$

The resulting discrete-time signal is denoted as $x[n]$, where:

$$x[n] = x(nT_s), \qquad n = 0, 1, 2, \ldots$$

To ensure that the discrete-time signal $x[n]$ fully represents the original continuous-time signal $x(t)$ without loss of information, the sampling rate f_s must satisfy the Nyquist criterion:

$$f_s \geq 2f_{max}$$

where, f_{max} is the highest frequency present in the signal. If this condition is not met, aliasing occurs, which introduces distortion into the discrete-time signal.

Q. Write the advantages of discrete signal in communication.

Discrete signals, also known as digital signals, have several advantages in communication systems. Here are some key benefits:

1. **Noise Resistance:** Digital signals are more resistant to noise and interference compared to analog signals. This ensures better signal integrity during transmission.
2. **Error Detection and Correction:** Discrete signals can incorporate error detection and correction mechanisms (e.g., parity bits, checksums) to ensure reliable communication, even in noisy environments.
3. **Compatibility with Digital Systems:** Discrete signals are inherently compatible with digital devices, such as computers, microprocessors, and digital communication systems, making them ideal for modern technologies.
4. **Compression:** Discrete signals can be compressed to save bandwidth or storage space, enabling efficient utilization of resources.
5. **Efficiency in Processing:** Digital signals can be easily processed, stored, and manipulated using digital circuits and computer systems. Operations such as compression, encryption, and modulation are more efficiently performed on digital data.

3.3 DIGITAL SIGNALS

3.3.4 Transmission of Digital Signals

We can transmit a digital signal by using one of two different approaches: baseband transmission or broadband transmission (using modulation).

Q. What are baseband and broadband transmissions?

Baseband Transmission: Baseband transmission refers to sending signals over a channel without changing their original form. In baseband transmission, the entire bandwidth of the communication channel is used to transmit a single signal. This type of transmission is typically used for short-distance communication within a single network. Ethernet is a common example of baseband transmission.

Broadband Transmission: Broadband transmission refers to sending signals by modulating them onto a carrier signal, which allows multiple signals to share the same medium simultaneously. This technique is commonly used in wireless and cable communications, such as Cable TV, DSL internet, and satellite communications.

3.4 TRANSMISSION IMPAIRMENT

Q. What is transmission impairment? Classify and briefly describe various transmission impairments.

Transmission impairment refers to any kind of degradation that occurs during the transmission of data signal over a communication channel, resulting in errors or loss of information.

Transmission impairments can be broadly classified into three main categories: attenuation, distortion, and noise.

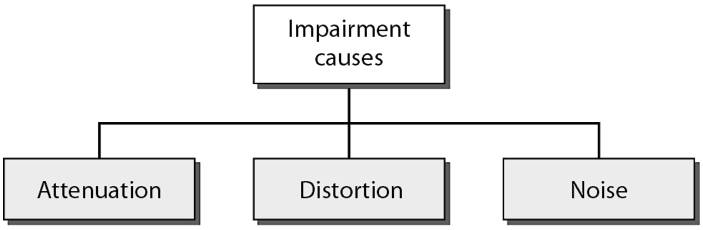


Figure 3.25 Causes of impairment / Impairment types

Attenuation

Attenuation refers to the loss of signal strength/energy as it propagates through a medium. This loss occurs due to the resistance of the medium, leading to a reduction in signal amplitude. To compensate for this loss, amplifiers are used to amplify the signal. Figure 3.27 shows the effect of attenuation and amplification.

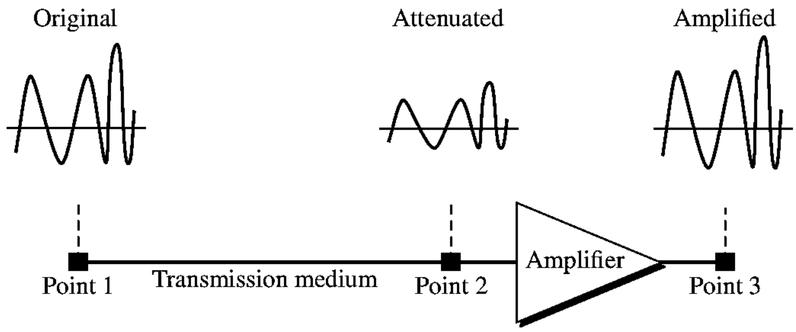


Figure 3.26 Attenuation

Distortion

Distortion means that the signal changes its form or shape during transmission. It can occur in a composite signal composed of multiple frequencies. Each signal component has its own propagation speed through the medium, resulting in varying delays upon arrival at the destination. Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration. Consequently, the composite signal at the receiver is not the same as the one sent by the sender. Figure 3.29 illustrates the effect of distortion on a composite signal.

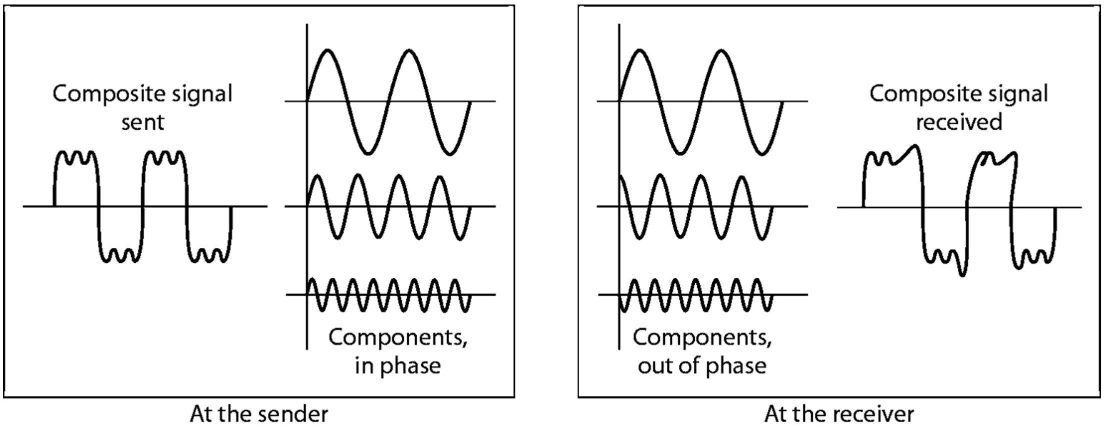


Figure 3.28 Distortion

Noise

Noise is another cause of signal impairment. It refers to unwanted electrical or electromagnetic signals that interfere with the transmitted data. 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 external 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 3.30 shows the effect of noise on a signal.

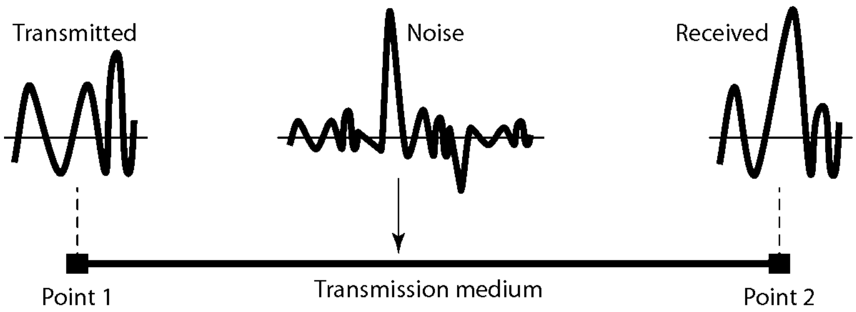


Figure 3.29 Noise

Example 3.26: Suppose a signal travels through a transmission medium and its power is reduced to one-half. This means that $P_2 = \frac{1}{2} P_1$. In this case, the attenuation (loss of power) can be calculated as

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{0.5P_1}{P_1} = 10 \log_{10} 0.5 = 10(-0.3) = -3 \text{ dB}$$

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

Q. What is meant by signal to noise ratio? Discuss the importance of SNR in radio receivers.

Signal-to-Noise Ratio (SNR)

The signal-to-noise ratio (SNR) is a measure of the strength of the desired signal relative to the background noise in a communication system. It is mathematically defined as:

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

We need to consider the average signal power and the average noise power because these may change with time. A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise.

Since SNR is the ratio of two powers, it is often described in decibel (dB) units, and defined as

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

Importance of SNR in radio receivers

The Signal-to-Noise Ratio (SNR) is crucial in radio receivers as it determines the quality and clarity of the received signal. A higher SNR means the signal is much stronger than the noise, ensuring clear audio or data reception. This reduces error rates, extends the receiver's effective range, and enhances its ability to manage interference. High SNR is essential for reliable communication, especially in environments with significant noise or interference, such as urban areas or during long-distance transmissions.

3.5 DATA RATE LIMITS

3.5.1 Noiseless Channel: Nyquist Bit Rate

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

$$\text{BitRate} = 2 \times \text{bandwidth} \times \log_2 L$$

In this formula, $\textbf{bandwidth}$ is the bandwidth of the channel, L is the number of signal levels used to represent data, and $\textbf{BitRate}$ is the bit rate in bits per second.

Example 3.35: Consider a noiseless channel with a bandwidth of 3000 Hz transmitting a signal with four signal levels (for each level, we send 2 bits). The maximum bit rate can be calculated as

$$\text{BitRate} = 2 \times 3000 \times \log_2 4 = 12{,}000 \text{ bps}$$

Q. State and explain Sampling theorem.

The Sampling Theorem states that: For a continuous-time signal to be perfectly reconstructed from its discrete samples, it must be sampled at a rate that is at least twice the highest frequency present in the analog signal. This minimum sampling rate is known as the Nyquist rate.

Let f_s be the sampling rate, which is the number of samples per second, and f_m be the highest frequency in the signal. The theorem states that $f_s \geq 2f_m$. This means the sampling rate must be at least twice the highest frequency to capture all its details without loss of information. If f_s is less than $2f_m$, then aliasing occurs. Aliasing is a distortion or artifact caused when the sampling rate is insufficient, causing high-frequency components to be misrepresented in the sampled signal.

If the signal is sampled at or above the Nyquist rate, it can be perfectly reconstructed from its samples. After sampling, a low-pass filter can be applied to eliminate any frequencies above half of the sampling rate (the Nyquist frequency). This helps prevent aliasing when reconstructing the signal. The sampled signal can be reconstructed using an interpolation technique (e.g., linear, spline, or polynomial interpolation) to approximate the original continuous signal.

Q. Explain Nyquist Criterion for signaling. When is it considered?

The Nyquist Criterion is derived from the Nyquist Theorem, which states that for a band-limited signal (i.e., a signal whose frequency content is limited to a certain bandwidth) to be transmitted without distortion over a noiseless channel, it must be sampled at a rate that is at least twice the bandwidth of the signal. This rate is known as the Nyquist rate.

When Considered:

The Nyquist Criterion is particularly relevant when designing digital communication systems, especially when dealing with high-speed data transmission over limited bandwidth channels. It is considered when the communication system needs to transmit data efficiently without errors or distortion.

3.6 PERFORMANCE

Q. What is meant by the effective bandwidth of a signal?

The effective bandwidth of a signal refers to the range of frequencies that contains most of the signal's energy and is essential for accurate transmission and reconstruction. It is the portion of the frequency spectrum where the signal's power is concentrated. In practical terms, the effective bandwidth determines the minimum channel capacity required to transmit the signal without significant loss or distortion.

Q. How we measure the efficiency of digital transmission?

To measure the efficiency of digital transmission, several key parameters and metrics are used:

1. **Bandwidth:** This measures the range of frequencies used for transmission. A lower bandwidth requirement indicates higher efficiency.
2. **Bit Error Rate (BER):** This measures the number of bits received incorrectly compared to the total number of bits transmitted. A lower BER indicates higher efficiency.
3. **Throughput:** The throughput is a measure of how fast we can actually send data through a network. Higher throughput indicates better efficiency.
4. **Power Efficiency:** This measures the amount of power required to transmit a given amount of data. Lower power consumption indicates higher efficiency.
5. **Latency (Delay):** Latency or delay refers to the time it takes for an entire message to reach its destination, starting from the moment the first bit is transmitted from the source. Lower latency denotes greater efficiency.

Q. What are the conditions for distortion less line?

For a distortionless transmission line, the following conditions must be met:

1. **Attenuation Constant (α) must be independent of frequency:** This means that all frequency components of the signal are attenuated equally, preserving the signal's shape.
2. **Phase Velocity (v) must be independent of frequency:** This ensures that all frequency components of the signal travel at the same speed, preventing dispersion and maintaining the signal's shape.

Q. What is the thermal noise level of a channel with bandwidth of 10 kHz carrying 100W of power operating at 500°C?

The thermal noise power of a channel can be calculated using the formula:

$$N = kTB$$

Where:

k is the Boltzmann constant (1.38×10^{-23} J/K),
T is the absolute temperature in Kelvin,
B is the bandwidth in Hertz.

Given,

Temperature, $T = 500℃ = 500 + 273.15 = 773.15$ K
Bandwidth, $B = 10$ kHz $= 10000$ Hz

Now, calculate:

$$N = kTB = (1.38 \times 10^{-23}) \times 773.15 \times 10000 = 1.067 \times 10^{-16} \text{ W}$$

Therefore, the thermal noise level of the channel is approximately 1.067×10^{-16} W.

Chapter 4: Digital Transmission

4.1 DIGITAL-TO-DIGITAL CONVERSION

4.1.1 Line Coding

Q. What do you mean by line coding?

Line coding is the process of converting digital data to digital signals. It involves converting a sequence of bits into a digital signal. At the sender, digital data are encoded into a digital signal; at the receiver's end, the digital signal are decoded to reconstruct the original digital data.

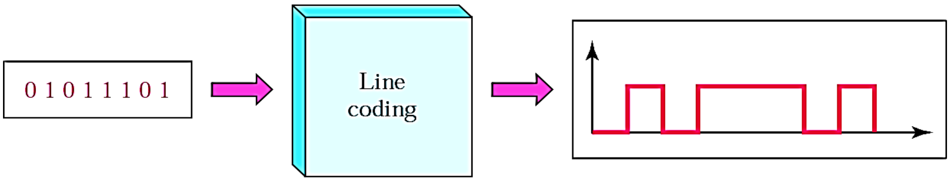


Figure 4.1 Line coding

Q. Why data encoding is necessary?

Data encoding is necessary to ensure efficient and reliable communication over a transmission medium. It helps in:

1. **Signal Compatibility:** Converts data into a form suitable for the physical characteristics of the transmission medium.
2. **Error Detection and Correction:** Embeds mechanisms to identify and correct transmission errors.
3. **Synchronization:** Encoding techniques can help synchronize the sender and receiver, ensuring that both devices are aligned in time.
4. **Bandwidth Utilization:** Optimizes the use of available bandwidth for efficient data transfer.

Q. Define a DC component and write its effect on digital transmission.

When the voltage level in a digital signal is constant for a while, the spectrum creates very low frequencies (results of Fourier analysis). These low frequencies (around zero), referred to as DC (direct-current) components.

Effects on Digital Transmission:

A DC component can cause several issues in digital transmission:

1. **DC Coupling Problems:** Many transmission media, such as telephone lines and transformers, are designed to block DC components. A DC component in the signal can interfere with the transmission process, leading to signal distortion or loss.
2. **Base-Line Wander:** DC component can cause baseline wandering, where the reference voltage drifts, leading to difficulty in accurate signal interpretation.
3. **Synchronization Issues:** Long sequences of constant voltage levels (e.g., continuous 1s or 0s) make it difficult for the receiver to maintain synchronization with the sender.
4. **Increased Power Requirements:** The presence of a DC component requires additional power for transmission, reducing energy efficiency.

4.1.2 Line Coding Schemes

We can roughly divide line coding schemes into five broad categories, as shown in Figure 4.4. There are several schemes in each category.

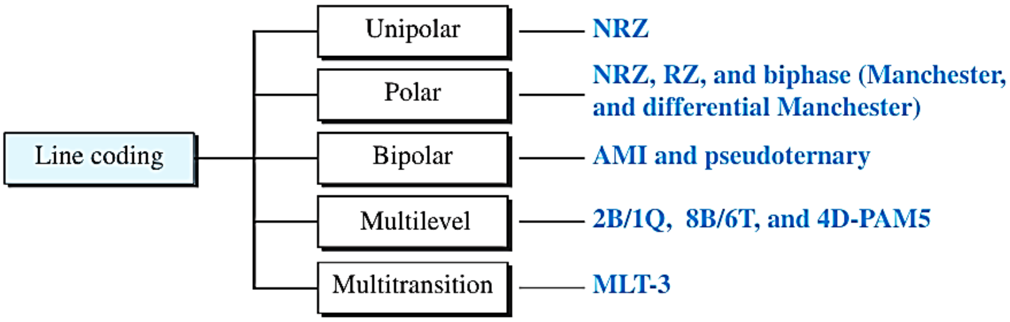
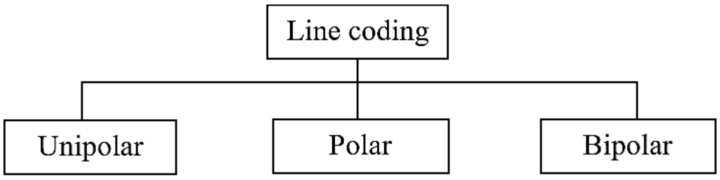


Figure 4.4 Line coding schemes

or, Line coding schemes can be broadly classified into three main categories:



Unipolar Scheme

In a unipolar line coding scheme, all signal levels are on one side of the time axis, either above or below. In other words, this scheme uses only one voltage level, meaning it uses only one polarity of voltage, either positive or negative.

NRZ (Non-Return-to-Zero)

Traditionally, a unipolar scheme is implemented/designed as a non-return-to-zero (NRZ) scheme, where a positive voltage represents bit 1, and zero voltage represents bit 0. It is called NRZ because the signal does not return to zero at the middle of the bit. Figure 4.5 shows a unipolar NRZ scheme.

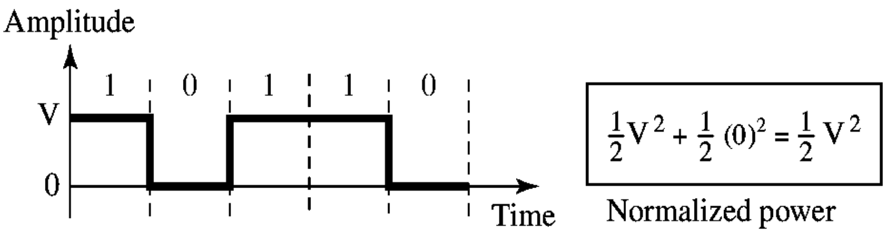


Figure 4.5 Unipolar NRZ scheme

Compared to its polar counterpart, this scheme is very costly. The normalized power (the power needed to send 1 bit per unit line resistance) is double that for polar NRZ. For this reason, this scheme is normally not used in data communications today.

Polar Schemes

In polar schemes, the signal levels are on both sides of the time axis. In other words, this scheme uses two voltage levels, i.e., positive and negative voltages. For example, the voltage level for 0 can be positive and the voltage level for 1 can be negative. The major types of polar schemes are: NRZ, RZ, Manchester, and Differential Manchester.

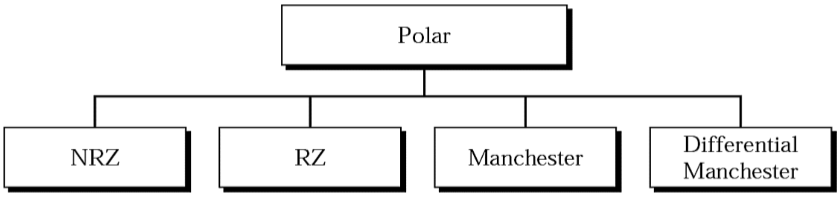


Figure: Types of polar encoding

Non-Return-to-Zero (NRZ)

The NRZ line coding technique uses two distinct voltage levels to represent binary data. In this technique, the signal remains constant during the entire bit interval without returning to zero. There are two variations of NRZ: NRZ-L (NRZ-Level) and NRZ-I (NRZ-Invert), as shown in Figure 4.6.

In **NRZ-L** the level of the voltage determines the value of the bit. A positive voltage represents one binary value (e.g., 0), while a negative voltage represents the other binary value (e.g., 1).

In **NRZ-I** the inversion or the lack of inversion determines the value of the bit. A binary 1 is represented by a change in the signal level (inversion of polarity), whereas a binary 0 is represented by no change in the signal level (the current level remains the same).

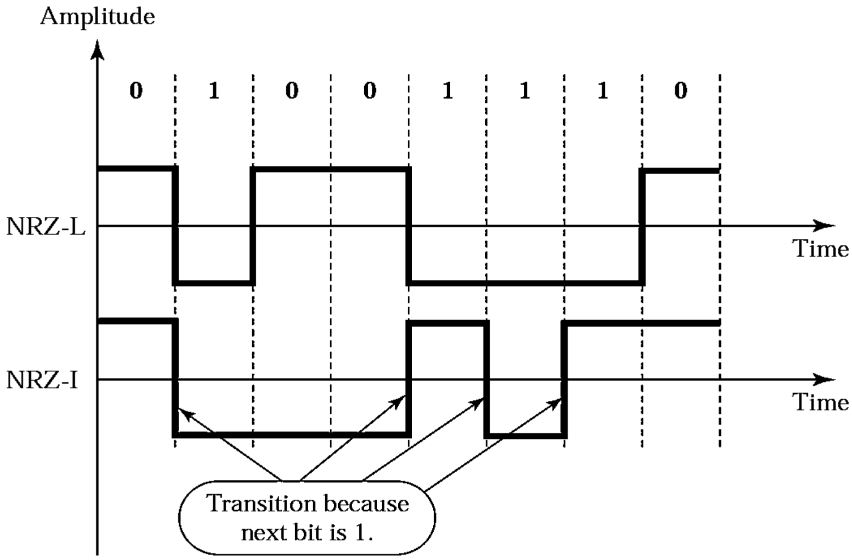


Figure 4.6 Polar NRZ-L and NRZ-I schemes

Let us compare NRZ-L and NRZ-I based on the defined criteria:

1. **Baseline Wandering:** Although baseline wandering is a problem for both variations, it is twice as severe in NRZ-L. If there is a long sequence of 0s or 1s in NRZ-L, the average signal power becomes skewed. The receiver might have difficulty discerning the bit value. In NRZ-I this problem occurs only for a long sequence of 0s. If somehow we can eliminate the long sequence of 0s, we can avoid baseline wandering.
2. **Synchronization:** The synchronization problem (sender and receiver clocks are not synchronized) also exists in both schemes. Again, this problem is more serious in NRZ-L than in NRZ-I. While a long sequence of 0s can cause a problem in both schemes, a long sequence of 1s affects only NRZ-L.
3. **Polarity Changes:** Another problem with NRZ-L occurs when there is a sudden change of polarity in the system. For example, if twisted-pair cable is the medium, a change in the polarity of the wire results in all 0s interpreted as 1s and all 1s interpreted as 0s. NRZ-I does not have this problem.
4. **Signal Rate:** Both schemes have an average signal rate of $N/2$ Bd.

Return-to-Zero (RZ)

The main problem with NRZ encoding occurs when the sender and receiver clocks are not synchronized. The receiver does not know when one bit has ended and the next bit is starting. One solution is the return-to-zero (RZ) scheme, which uses three signal levels: positive, negative, and zero.

In the Polar RZ scheme, the signal does not remain constant throughout the entire bit period. Instead, it returns to zero in the middle of each bit. As shown in Figure 4.7, the signal goes to zero at the midpoint of each bit and remains/stays there until the start/beginning of the next bit.

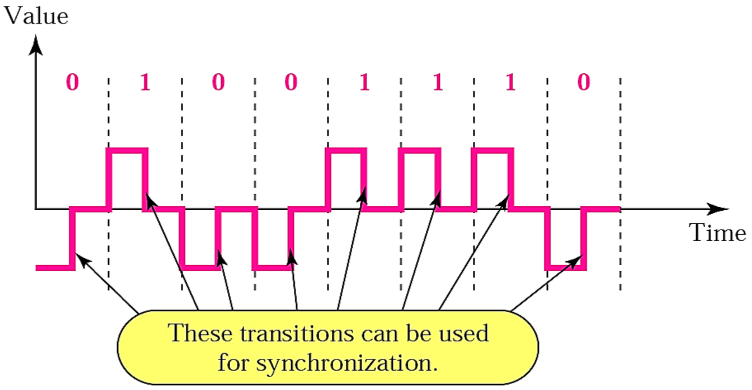


Figure 4.7 Polar RZ encoding scheme

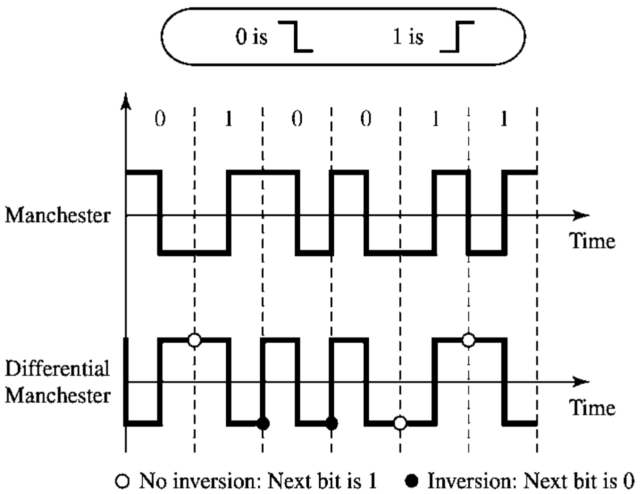
The main disadvantage of RZ encoding is that it requires two signal changes to encode a bit and therefore occupies greater bandwidth. Additionally, a sudden change of polarity resulting in all 0s interpreted as 1s and all 1s interpreted as 0s, but there is no DC component problem. Another problem is the complexity: RZ uses three levels of voltage, which makes it more complex to generate and decode accurately. Due to these limitations, the RZ scheme is not used today. Instead, it has been replaced by the better-performing Manchester and differential Manchester schemes.

Biphase: Manchester and Differential Manchester

Manchester encoding scheme combines the ideas/concepts of RZ (transition at the middle of the bit) and NRZ-L. In Manchester encoding, the duration of the bit is divided into two halves. The voltage remains at one level during the first half and moves to the other level in the second half. The transition at the middle of the bit provides synchronization.

Differential Manchester, on the other hand, combines the ideas of RZ (transition at the middle of the bit) and NRZ-I (bit value determined by transitions). There is always a transition at the middle of the bit to provide synchronization. The bit value is determined by whether there is an inversion (transition) or no inversion (no transition) at the beginning of the bit interval. If the next bit is 0, there is an inversion at the start of the bit. If the next bit is 1, there is no inversion at the start of the bit.

Figure 4.8 shows both Manchester and differential Manchester encoding schemes.



or,

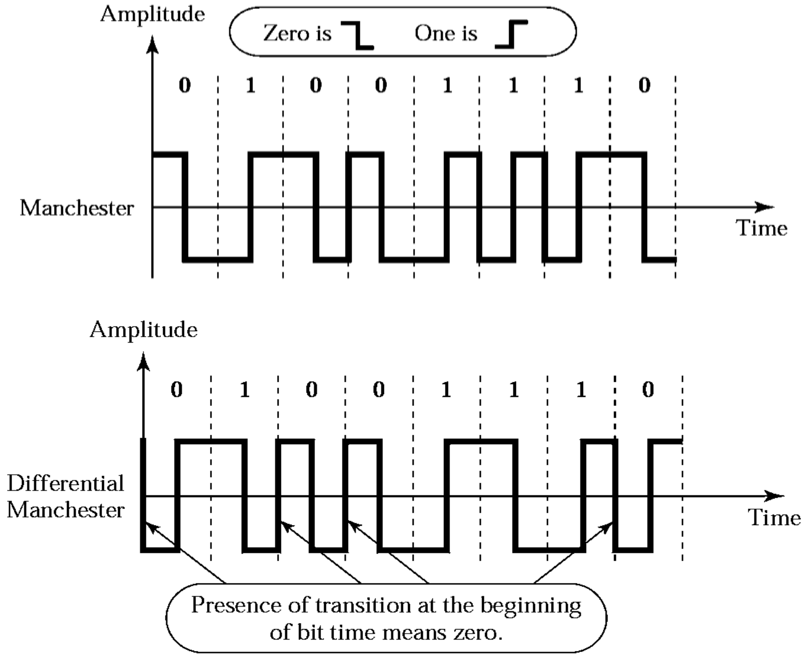


Figure 4.8 Polar biphase: Manchester and differential Manchester schemes

The Manchester scheme overcomes several problems associated with NRZ-L, and differential Manchester overcomes several problems associated with NRZ-I. First, there is no baseline wandering. Second, there is no DC component because each bit has a positive and negative voltage contribution.

The main drawback of Manchester and Differential Manchester encoding is the higher signal rate. The signal rate for these schemes is twice/double that of NRZ because there is always one transition in the middle of each bit and maybe one transition at the end of each bit. As a result, the minimum bandwidth of Manchester and differential Manchester is 2 times that of NRZ.

Note that Manchester and differential Manchester schemes are also called **biphase** schemes.

Bipolar Schemes

In bipolar encoding (sometimes called multilevel binary), there are three voltage levels: positive, negative, and zero. The voltage level for one data element is at zero, while the voltage level for the other element alternates between positive and negative.

AMI and Pseudoternary

Figure 4.9 shows two variations of bipolar encoding: AMI and pseudoternary. A common bipolar encoding scheme is called bipolar alternate mark inversion (AMI). In the term alternate mark inversion, the word mark comes from telegraphy and means 1. So AMI means alternate 1 inversion. A neutral zero voltage represents binary 0. Binary 1s are represented by alternating positive and negative voltages. A variation of AMI encoding is called pseudoternary in which the 1 bit is encoded as a zero voltage and the 0 bit is encoded as alternating positive and negative voltages.

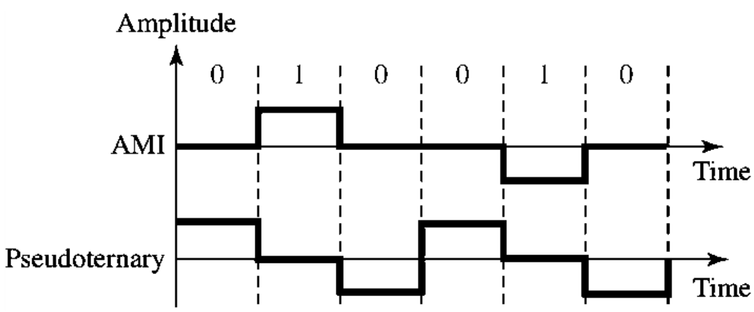


Figure 4.9 Bipolar schemes: AMI and pseudoternary

Tutorial:

- https://www.youtube.com/watch?v=2OJXnk1QjJk&list=PLncy2sD7w4YoYG5EFgry-AcdDAy2pFvMJ
- https://www.youtube.com/watch?v=ts_iiEoLV7I&list=PLncy2sD7w4YoYG5EFgry-AcdDAy2pFvMJ&index=2
- https://www.youtube.com/watch?v=sGDynGq6Vy0&list=PLncy2sD7w4YoYG5EFgry-AcdDAy2pFvMJ&index=3
- https://www.youtube.com/watch?v=myI7t0tt9ak&list=PLncy2sD7w4YoYG5EFgry-AcdDAy2pFvMJ&index=4

Topic Related Questions:

1. List three techniques of digital-to-digital conversion.
2. Classify line coding schemes and describe NRZ-L line coding technique.
3. Classify line coding scheme and describe NRZ-I line coding technique.
4. Encode the following data stream 0001100111 using NRZ-I encoding technique.
5. For the bit stream 10110011, sketch the waveforms for Non-return to Zero Invert (NRZ-I) and Manchester encoding format.
6. For the bit stream 01011001, sketch the waveforms for Non-return to Zero Inverted (NRZ-I) and Manchester encoding format.
7. Discuss in brief the Manchester and Differential Manchester bi-phase encoding in use on networks.
8. For the bit stream 01011101, sketch the waveform for Manchester encoding format.
9. Describe differential Manchester encoding technique.
10. Write the advantages of bi-phase coding technique.

4.1.3 Block Coding

Block Coding is a technique used in digital communication to introduce redundancy for synchronization and error detection, thereby improving the performance of line coding. In general, block coding changes/transforms a block of m bits into a block of n bits, where n is larger than m. This method is commonly referred to as the mB/nB encoding technique.

Block coding normally involves three steps: division, substitution, and combination. In the division step, a sequence of bits is divided into groups of m bits. For example, in 4B/5B encoding, the original bit sequence is divided into 4-bit groups. The heart of block coding is the substitution step. In this step, we substitute an m-bit group with an n-bit group. For example, in 4B/5B encoding we substitute a 4-bit group with a 5-bit group. Finally, the n-bit groups are combined to form a stream. The new stream has more bits than the original bits. Figure 4.14 shows the procedure.

Note: The slash in block encoding (for example, 4B/5B) distinguishes block encoding from multilevel encoding (for example, 8B6T), which is written without a slash.

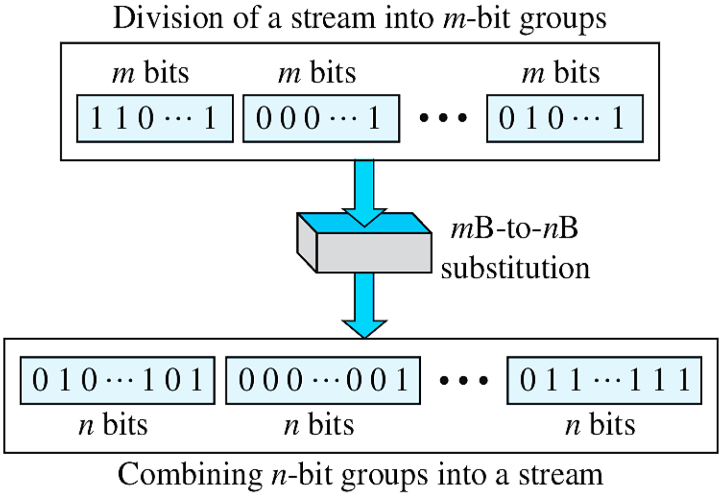
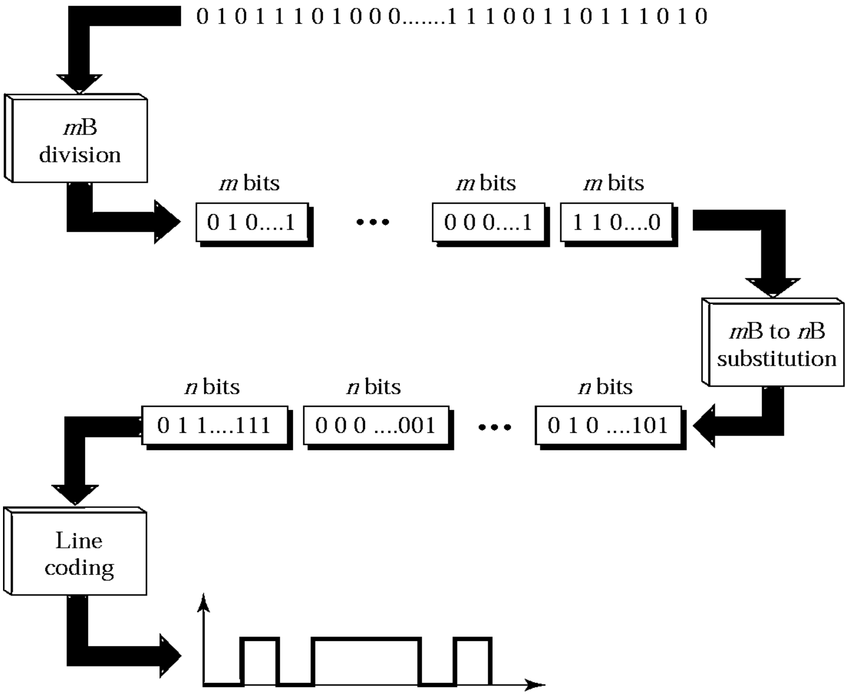


Figure 4.14 Block coding concept

or,



Purpose of Block Coding

1. **Error Detection and Correction:** By adding redundant bits to the original data, block coding helps detect and correct errors that may occur during data transmission.
2. **Synchronization:** Block codes help maintain synchronization between the sender and receiver by including specific patterns within the data stream.
3. **DC Component Elimination:** Reduces or eliminates DC components, ensuring compatibility with systems that cannot handle low-frequency signals.
4. **Improved Signal Performance:** Provides better performance by encoding data into patterns that meet the requirements of the transmission channel, such as balancing the number of 0s and 1s.

Topic Related Questions:

1. What is block coding and give its purpose.

4.2 ANALOG-TO-DIGITAL CONVERSION

Analog-to-digital conversion is the process of transforming continuous-time analog signals into discrete digital data for processing, storage, or transmission in digital systems. In this section we describe two techniques for Analog-to-Digital Conversion: pulse code modulation and delta modulation.

4.2.1 Pulse Code Modulation (PCM)

Q. Describe PCM with block diagram.

Pulse Code Modulation (PCM) is a technique that converts an analog signal to digital data (digitization). This process involves three main steps: sampling, quantization, and encoding, as shown in Figure 4.21.

Sampling

The first step in PCM is sampling. In this step, the analog signal is sampled at regular intervals to capture its amplitude at discrete moments in time. 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.

The sampling rate is chosen according to the Nyquist Theorem, which states that the sampling rate must be at least twice the highest frequency of the analog signal to avoid aliasing.

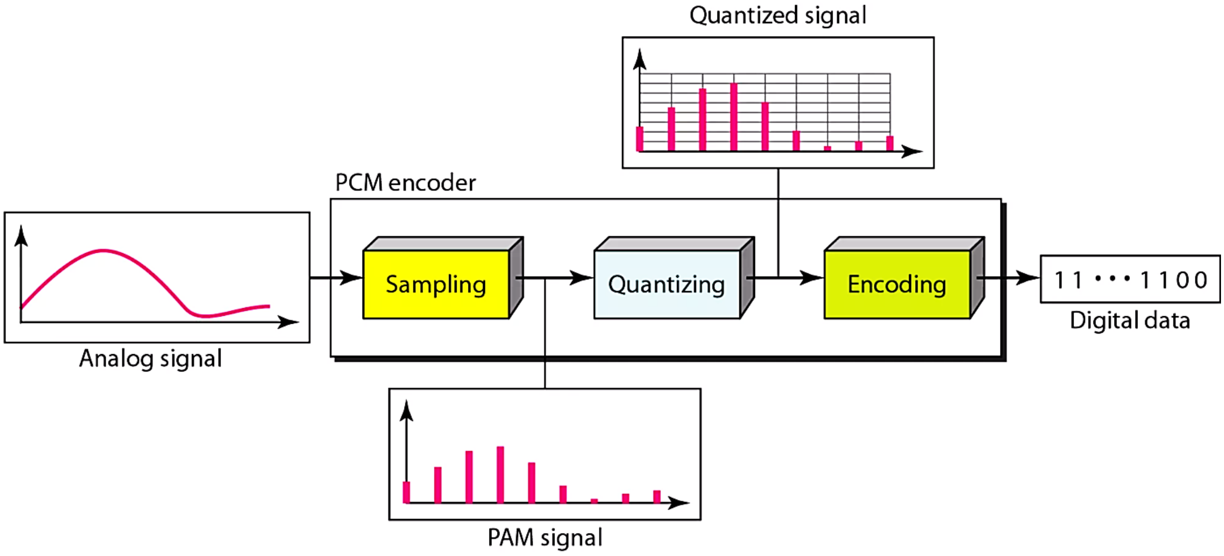


Figure 4.21 Components of PCM encoder

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-integer values between the two limits. These values cannot be used in the encoding process. In quantization, each sampled amplitude is mapped to the nearest value within a finite set of levels (quantization levels). This process introduces quantization error, which is the difference between the actual sample value and the quantized value.

Quantization can be either uniform quantization (equal step sizes) or non-uniform quantization (step sizes vary).

Encoding

The last step in PCM is encoding. In this step, the quantized values are encoded into binary form to represent the digital signal. Each sample is converted into an n_b-bit code word.

The steps involved in quantization are as follows:

1. **Determine the range:** The analog signal's amplitude is assumed to lie between a minimum value (V_{min}) and a maximum value (V_{max}).
2. **Divide the range:** The range is divided into L zones, each of height Δ (delta).
$$\Delta = \frac{V_{max} - V_{min}}{L}$$
3. **Assign quantized values:** Assign quantized values of 0 to L − 1 to the midpoint of each zone.
4. **Approximate amplitudes:** The sample amplitudes are approximated to the nearest quantized value.

Q. Write the steps that involved in conversion of analog signal to digital signal.

The process of converting an analog signal to a digital signal involves following steps:

1. **Sampling:** The analog signal is sampled at regular intervals.
2. **Quantization:** Each sampled amplitude is mapped to the nearest value within a finite set of levels.
3. **Encoding:** The quantized values are then encoded into binary form.
4. **Digital Representation:** The encoded binary data is used to represent the digital signal.

4.2.2 Delta Modulation (DM)

Q. Explain Delta modulation? Describe its advantages.

Delta Modulation (DM) is a simple and efficient method of converting an analog signal into a digital signal. Unlike Pulse Code Modulation (PCM), which encodes the absolute amplitude of each sample, DM encodes the change (or delta) in the signal amplitude between successive samples. If the current sample is higher than the previous sample, a 1 is generated. If the current sample is lower than the previous sample, a 0 is generated. Figure 4.28 shows the process.

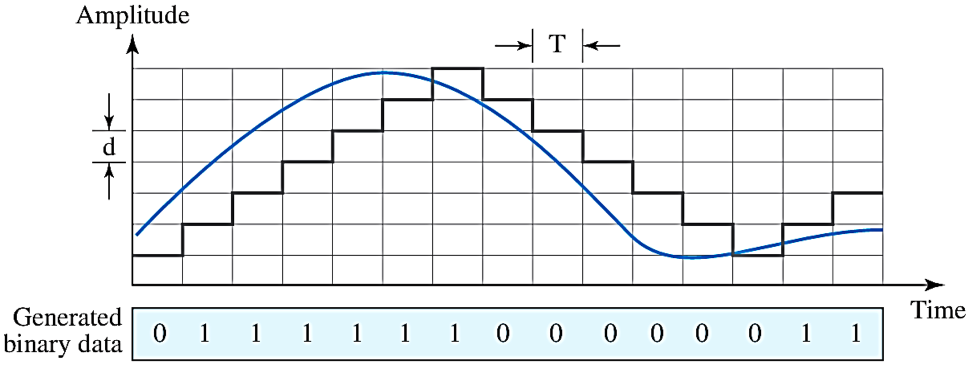


Figure 4.28 The process of delta modulation

This technique is particularly useful when the analog signal changes slowly over time. Note that there are no code words here; bits are sent one after another.

Advantages of Delta Modulation:

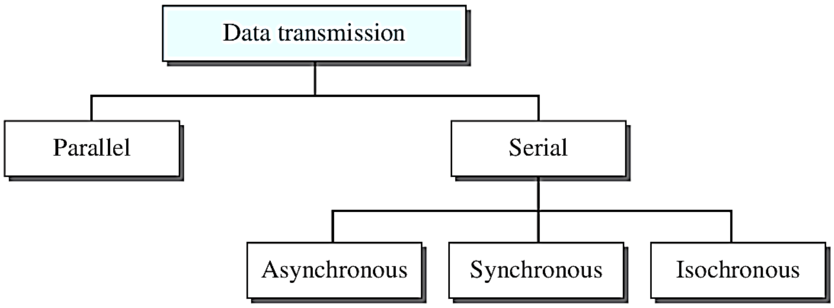
1. **Simplicity:** Delta Modulation is simpler to implement than PCM because it transmits only one bit per sample.
2. **Lower Bandwidth Requirement:** Since it transmits only 1 bit per sample, DM requires less bandwidth compared to PCM.
3. **Reduced Quantization Error:** Delta Modulation continuously tracks the signal, which can reduce the quantization error compared to PCM, especially for slowly varying signals.

Q. Compare and contrast PCM and DM.

Pulse Code Modulation (PCM)	Delta Modulation (DM)
PCM encodes the absolute amplitude of the signal at each sample.	DM encodes the change (difference) in amplitude between successive samples.
Quantizes and encodes each sample into a multi-bit binary code.	Encodes the difference as a single bit (increase or decrease).
Requires multiple bits per sample, depending on the quantization level (e.g., 8 bits, 16 bits).	Requires only 1 bit per sample.
Higher complexity due to multi-bit quantization and encoding.	Simpler implementation since it uses 1-bit quantization.
PCM requires higher bandwidth due to multi-bit transmission.	DM requires lower bandwidth as it transmits only 1 bit per sample.
PCM has good signal to noise ratio.	While DM has poor signal to noise ratio.
Reconstructs the signal more accurately with appropriate quantization levels.	May struggle with accurate reconstruction for rapidly changing signals.
More expensive due to higher processing requirements.	Less expensive.
Suitable for signals requiring high fidelity (e.g., audio, video).	Suitable for slowly varying signals (e.g., speech).

4.3 TRANSMISSION MODES

The transmission of binary data across a link can be accomplished in either parallel or serial mode. In parallel mode, multiple bits are sent with each clock tick. In serial mode, 1 bit is sent with each clock tick. While there is only one way to send parallel data, there are three subclasses of serial transmission: asynchronous, synchronous, and isochronous (see Figure 4.31).



or,

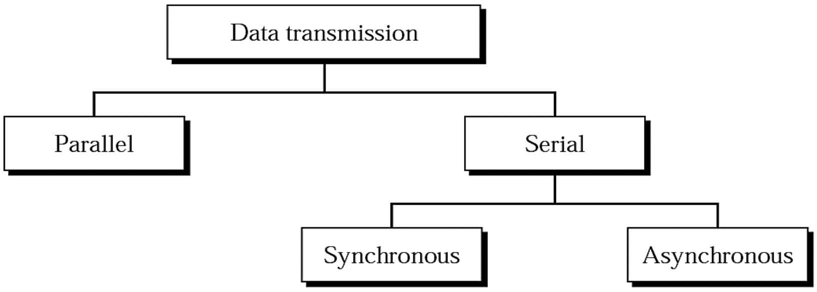


Figure 4.31 Data transmission and modes

Q. Write the advantages and disadvantages of parallel transmission over serial transmission.

Parallel transmission and serial transmission are two methods used to transmit data between devices. Each has its own strengths and weaknesses. Here are the advantages and disadvantages of parallel transmission over serial transmission:

Advantages of Parallel Transmission over Serial Transmission

1. **Higher Speed:** In parallel transmission, multiple bits are sent simultaneously, typically 8, 16, or 32 bits at a time, through separate channels. This making it faster than serial transmission, where bits are sent one after the other.
2. **Low Latency:** Parallel transmission reduces the time it takes to transmit a group of bits since they are sent all at once.
3. **More Suitable for Short Distances:** It performs well over short distances, such as inside a computer (e.g., data buses between CPU and RAM).

Disadvantages of Parallel Transmission over Serial Transmission

1. **Higher Cost:** Requires multiple wires or channels for data transfer, increasing the cost of cables and connectors.
2. **Signal Degradation:** Over long distances, signal degradation (crosstalk and electromagnetic interference) becomes significant due to closely packed wires. In this reason, parallel transmission is not suitable for long-distance communication.
3. **Complexity in Large-scale Systems:** Managing multiple data channels simultaneously increases design complexity for hardware systems.

Q. What are the differences between Synchronous and Asynchronous communication/Transmission?

Synchronous Communication	Asynchronous Communication
In Synchronous communication, data is sent in form of blocks or frames.	In Asynchronous communication, data is sent in form of bytes or characters.
Requires a shared clock between sender and receiver for synchronization.	No shared clock; synchronization is achieved using start and stop bits.
Faster, as data is sent in a frames without start/stop bits.	Slower, as extra bits (start and stop bits) are added to each data unit.
More efficient for large amounts of data due to continuous transmission.	Less efficient because of the additional overhead of start and stop bits.
Requires more complex hardware for synchronization and timing.	Simpler hardware, as no clock synchronization is needed.
Higher cost due to synchronization and clock management.	Lower cost as it does not require a shared clock.
Best for high-speed, large-volume data transfer.	Suitable for low-speed, intermittent communication.

Figures for understanding transmission modes

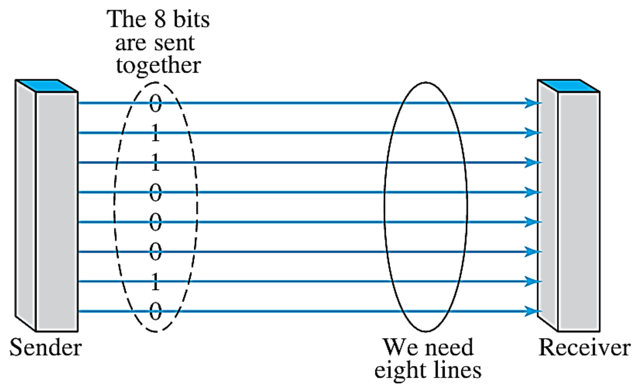


Figure 4.32 Parallel transmission

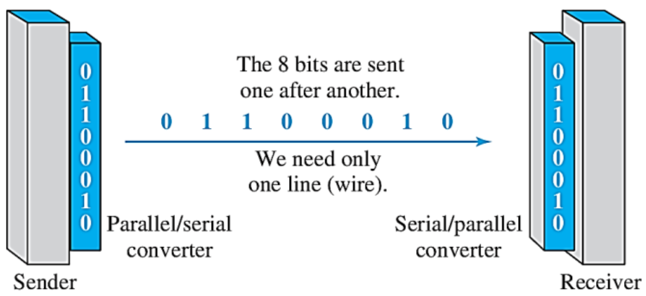


Figure 4.33 Serial transmission

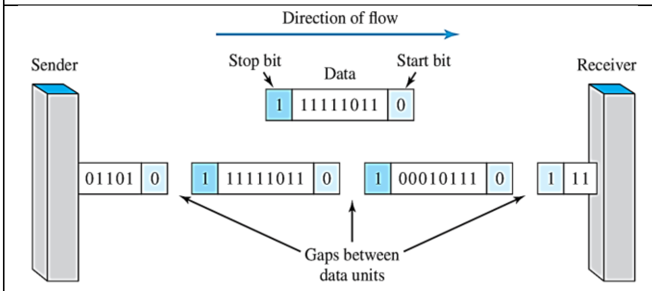


Figure 4.34 Asynchronous transmission

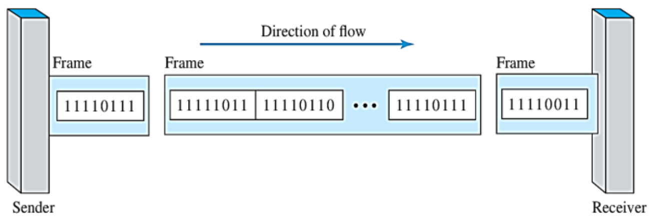


Figure 4.35 Synchronous transmission

Chapter 5: Analog Transmission

5.1 DIGITAL-TO-ANALOG CONVERSION/MODULATION

Q. What is digital to analog modulation?

Digital-to-analog modulation (also known as Digital Modulation) is the process of converting digital data into an analog signal. This technique modulates certain properties of a carrier signal (such as amplitude, frequency, or phase) to represent the digital data.

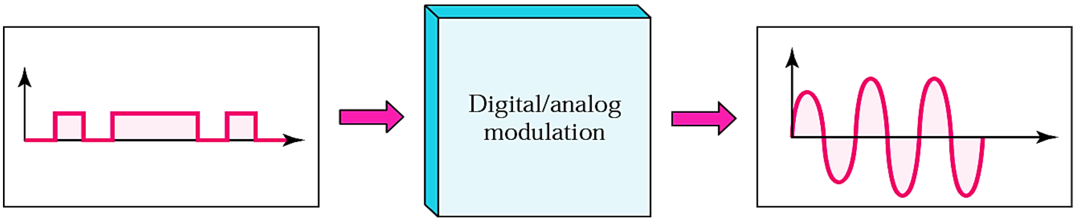


Figure 5.1 Digital-to-analog modulation

Types of digital-to-analog conversion

A sine wave is defined by three characteristics: amplitude, frequency, and phase. When we vary any one of these characteristics, we create a different version of that wave. So, by changing one characteristic of a simple electric signal, we can use it to represent digital data. Any of the three characteristics can be altered in this way, giving us at least three mechanisms for modulating digital data into an analog signal: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK). In addition, there is a fourth (and better) mechanism that combines changing both the amplitude and phase, called quadrature amplitude modulation (QAM). QAM is the most efficient of these options and is the mechanism commonly used today (see Figure 5.2).

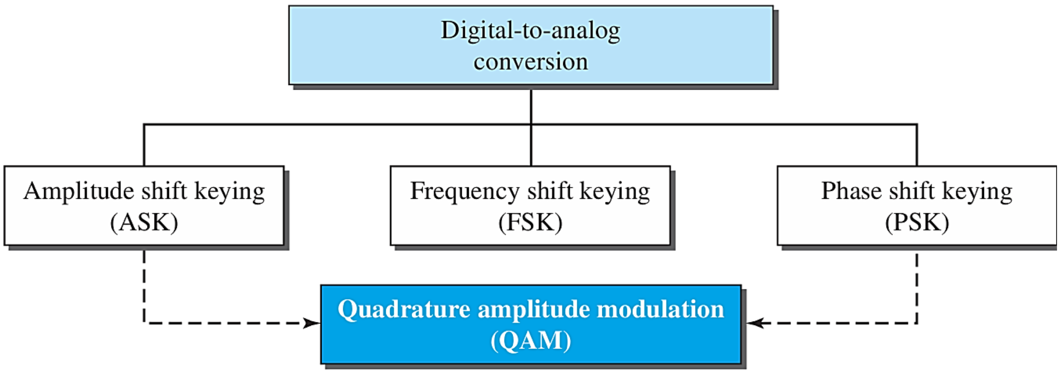


Figure 5.2 Types of digital-to-analog conversion

5.1.1 Aspects of Digital-to-Analog Conversion

Data Rate (Bit Rate) Versus Signal Rate (Baud Rate)

Bit rate is the number of bits per second. Baud rate is the number of signal elements per second. The relationship between them is

$$S = N \times \frac{1}{r} \text{ baud}$$

where N is the data rate (bps) and r is the number of data elements carried in one signal element. The value of r in analog transmission is $r = \log_2 L$, where L is the number of different signal elements. In the analog transmission of digital data, the baud rate is less than or equal to the bit rate.

Example 5.1: An analog signal carries 4 bits per signal element. If 1000 signal elements are sent per second, find the bit rate.

Solution

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

$$S = N \times \frac{1}{r}$$

$$\text{or}, N = S \times r = 1000 \times 4 = 4000 \text{ bps}$$

Example 5.2: An analog signal has a bit rate of 8000 bps and a baud rate of 1000 baud. How many data elements are carried by each signal element? How many signal elements do we need?

Solution

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

$$S = N \times \frac{1}{r}$$

$$\text{or}, r = \frac{N}{S} = \frac{8000}{1000} = 8 \text{ bits/baud}$$

and

$$r = \log_2 L$$

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

Carrier Signal

Q. Define carrier signal and its role in analog transmission.

Carrier Signal: In analog transmission, the sending device generates a high-frequency signal that serves as a base for the information signal. This base signal is called the carrier signal or carrier frequency. The carrier signal is modulated by the information signal to encode and transmit data effectively over the communication medium.

Role of Carrier Signal in Analog Transmission:

1. **Efficient Transmission:** Carrier signals operate at high frequencies, which makes them suitable for long-distance transmission without significant attenuation.
2. **Multiplexing:** The carrier enables the simultaneous transmission of multiple signals by assigning different carrier frequencies to each signal (frequency division multiplexing).
3. **Adaptation to Medium:** The carrier signal helps match the characteristics of the communication medium (e.g., coaxial cables, optical fibers, or radio waves), facilitating efficient propagation.
4. **Modulation Platform:** The carrier provides a platform for modulation, where the information is encoded by varying one or more of its properties (amplitude, frequency, or phase).

Digital information changes the carrier signal by modifying one or more of its characteristics (amplitude, frequency, or phase). This kind of modification is called modulation (shift keying).

Q. What do you mean by modulation?

Modulation is the process of varying one or more properties of a carrier signal according to the data signal. The carrier signal is typically a high-frequency signal that can efficiently travel over long distances, while the data signal (which could be analog or digital) contains the actual information.

Q. Define low level modulation and high level modulation.

Low-Level Modulation: Low-level modulation is a type of modulation where the modulating signal (information signal) is combined with the carrier signal before amplification. In this case, the power of the carrier signal is relatively low during modulation, and the resulting modulated signal is amplified later for transmission.

High-Level Modulation: High-level modulation is a type of modulation where the modulating signal is combined with a high-power carrier signal. The carrier signal is amplified before modulation occurs.

5.1.2 Amplitude Shift Keying

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

Binary ASK (BASK)

Although we can have several levels (kinds) of signal elements, each with a different amplitude, ASK is normally implemented using only two levels. This is referred to as binary amplitude shift keying or on-off keying (OOK). The peak amplitude of one signal level is 0; the other is the same as the amplitude of the carrier frequency. Figure 5.3 gives a conceptual view of binary ASK.

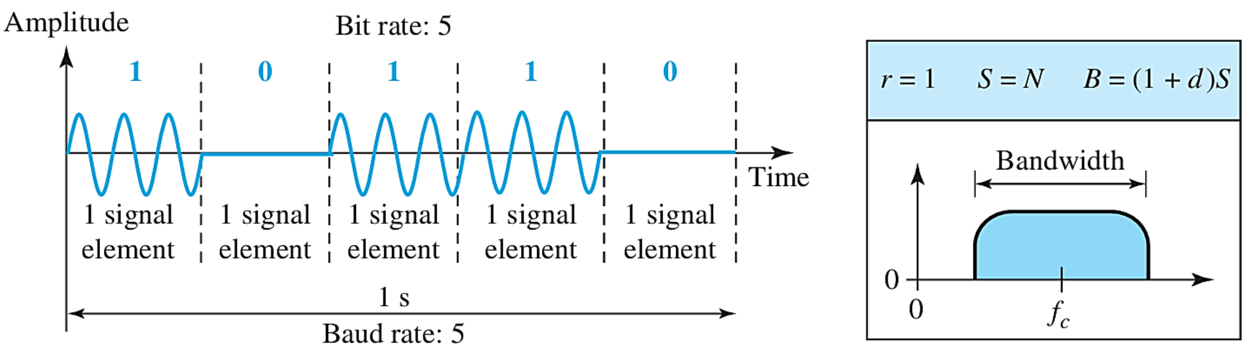


Figure 5.3 Binary amplitude shift keying

Figure 5.3 also shows the bandwidth for ASK. The value of d is between 0 and 1. This means that the bandwidth can be expressed as shown, where S is the signal rate and the B is the bandwidth.

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

Implementation

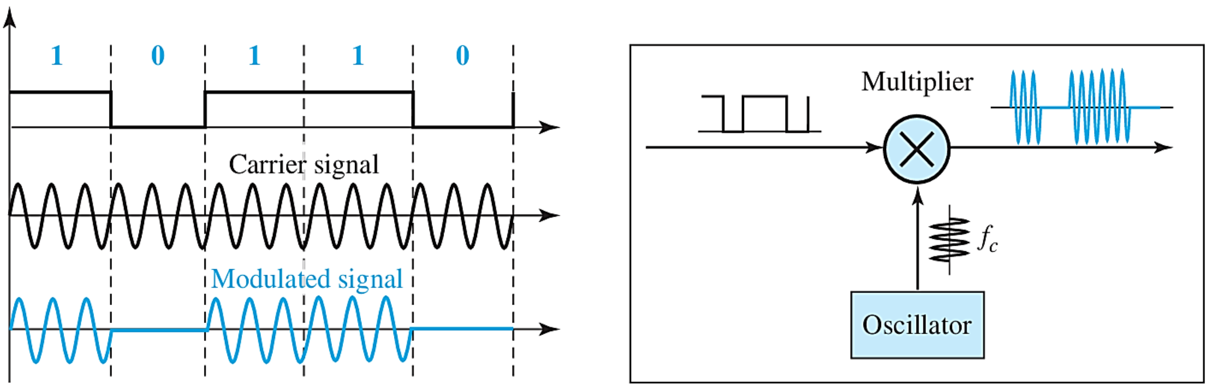


Figure 5.4 Implementation of binary ASK

If digital data are presented as a unipolar NRZ digital signal with a high voltage of 1 V and a low voltage of 0 V, the implementation can 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 signal/frequency is held; when the amplitude of the NRZ signal is 0, the amplitude of the carrier signal is zero.

5.1.3 Frequency Shift Keying

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

Binary FSK (BFSK)

BFSK uses two carrier frequencies, f_1 and f_2, to represent binary data: one for a binary 0 and another for a binary 1. However, note that this is an unrealistic example used only for demonstration purposes. Normally the carrier frequencies are very high, and the difference between them is very small.

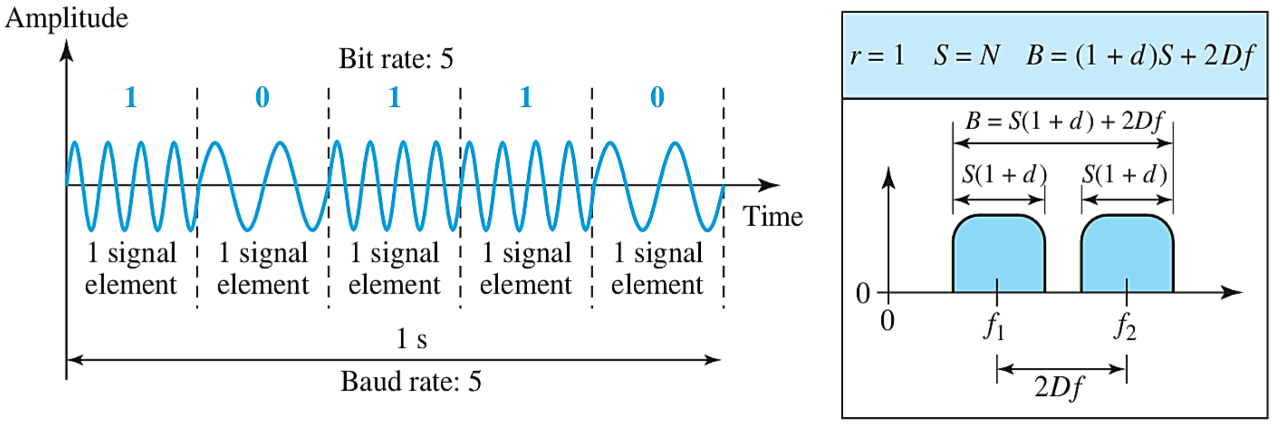


Figure 5.6 Binary frequency shift keying

Figure 5.6 also shows the bandwidth of FSK. If the difference between the two frequencies is $2\Delta_f$, then the required bandwidth is

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

Implementation

Figure 5.7 shows the implementation of BFSK. 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.

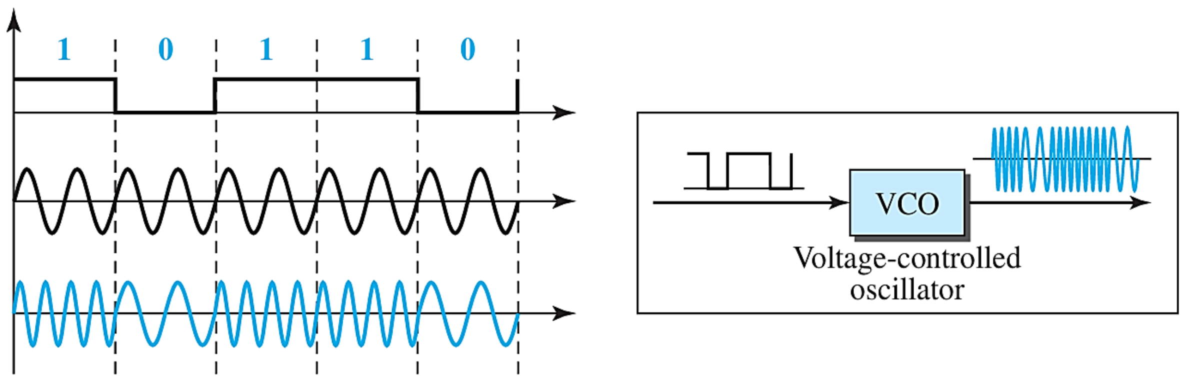


Figure 5.7 Implementation of BFSK

Multilevel FSK (MFSK)

In MFSK, we can use more than two frequencies to represent binary data. For example, we can use four different frequencies f_1, f_2, f_3, and f_4 to send 2 bits at a time. To send 3 bits at a time, we can use eight frequencies. And so on.

Topic Related Questions:

1. Briefly describe Frequency Shift Keying (FSK) modulation technique.

Q. Find the minimum bandwidth for an FSK signal transmitting at 2000 bps. Transmission is in half duplex mode and the carriers are separated by 3000 Hz.

For FSK,

$$\begin{aligned} \text{BW} &= \text{baud rate} + f_{c1} - f_{c0} \\ &= 2000 + 3000 \\ &= 5000 \text{ Hz} \end{aligned}$$

Hence, the minimum bandwidth required is 5000 Hz (5 kHz).

Q. Find the maximum bit rates for an FSK signal if the bandwidth of the medium is 12,000 Hz and the difference between the two carriers is 2000 Hz. Transmission is in full-duplex mode.

Because the transmission is full duplex, only 6000 Hz is allocated for each direction.

For FSK,

$$\text{BW} = \text{baud rate} + f_{c1} - f_{c0}$$

$$\text{or, baud rate} = \text{BW} - (f_{c1} - f_{c0}) = 6000 - 2000 = 4000$$

But because the baud rate is the same as the bit rate, the bit rate is 4000 bps.

5.1.4 Phase Shift Keying

In Phase Shift Keying (PSK), the phase of the carrier signal is varied to represent data. Both amplitude and frequency remain constant as the phase changes. Today, PSK is more common than ASK or FSK.

Binary PSK (BPSK)

The simplest form of PSK is Binary PSK (BPSK), where only two signal elements are used: one with a phase of 0° and the other with a phase of 180°. Figure 5.9 illustrates the basic concept of BPSK. This technique is as simple as Binary ASK with one big advantage: it is less affected by noise.

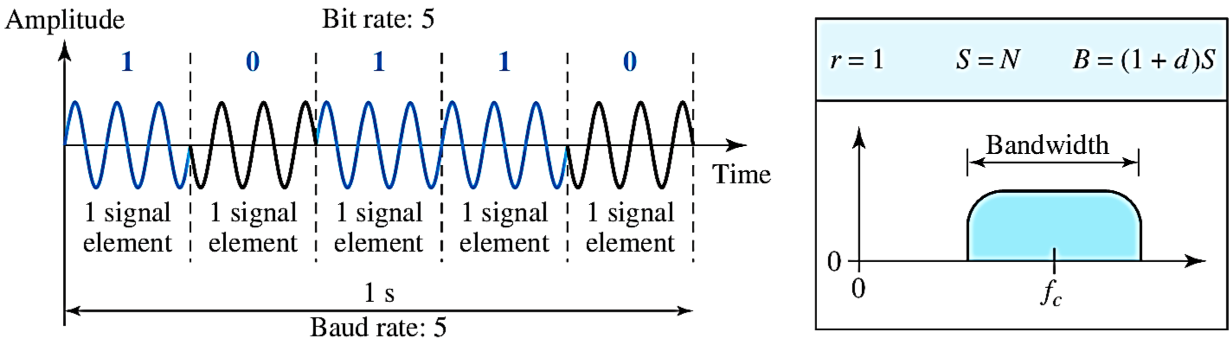


Figure 5.9 Binary phase shift keying

Figure 5.9 also shows the bandwidth for BPSK. The bandwidth is the same as that for binary ASK, but less than that for BFSK. No bandwidth is wasted for separating two carrier signals.

Implementation

The implementation of BPSK is as simple as that for ASK. The reason is that the signal element with phase 180° can be seen as the complement of the signal element with phase 0°. This gives us a clue on how to implement BPSK.

To implement BPSK, we follow a similar approach to ASK but use a polar NRZ signal instead of a unipolar NRZ signal, as illustrated in Figure 5.10. In this technique, 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°.

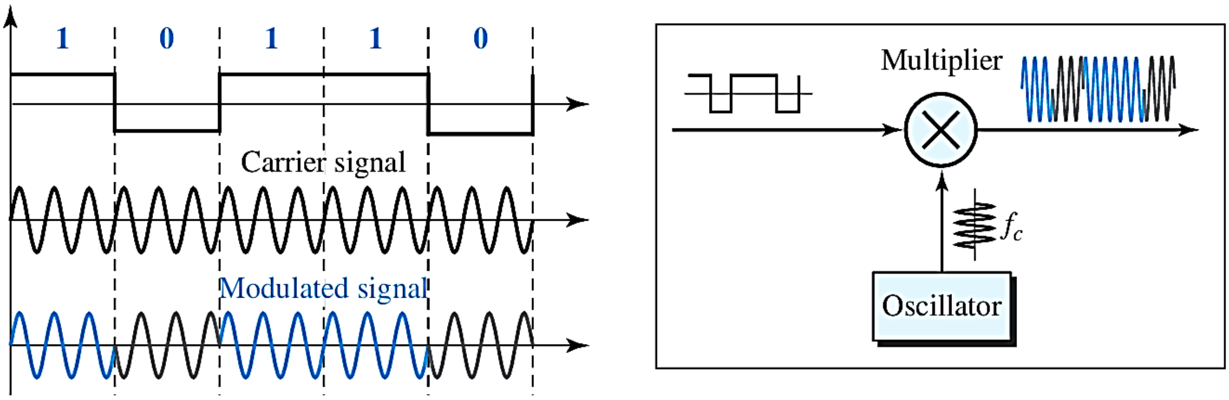


Figure 5.10 Implementation of BASK

Quadrature PSK (QPSK)

The implementation of Quadrature Phase Shift Keying (QPSK) builds upon the foundation of Binary Phase Shift Keying (BPSK), extending it to encode 2 bits per signal element. This enhancement decreases the baud rate and eventually the required bandwidth. QPSK achieves this by using two separate BPSK modulations: one in-phase and the other quadrature (90° out-of-phase). The incoming bits are first passed through a serial-to-parallel conversion that sends one bit to one modulator and the next bit to the other modulator. If the duration of each bit in the incoming signal is T, the duration of each bit sent to the corresponding BPSK signal is $2T$. This means that the bit to each BPSK signal has one-half the frequency of the original signal. Figure 5.11 shows the idea.

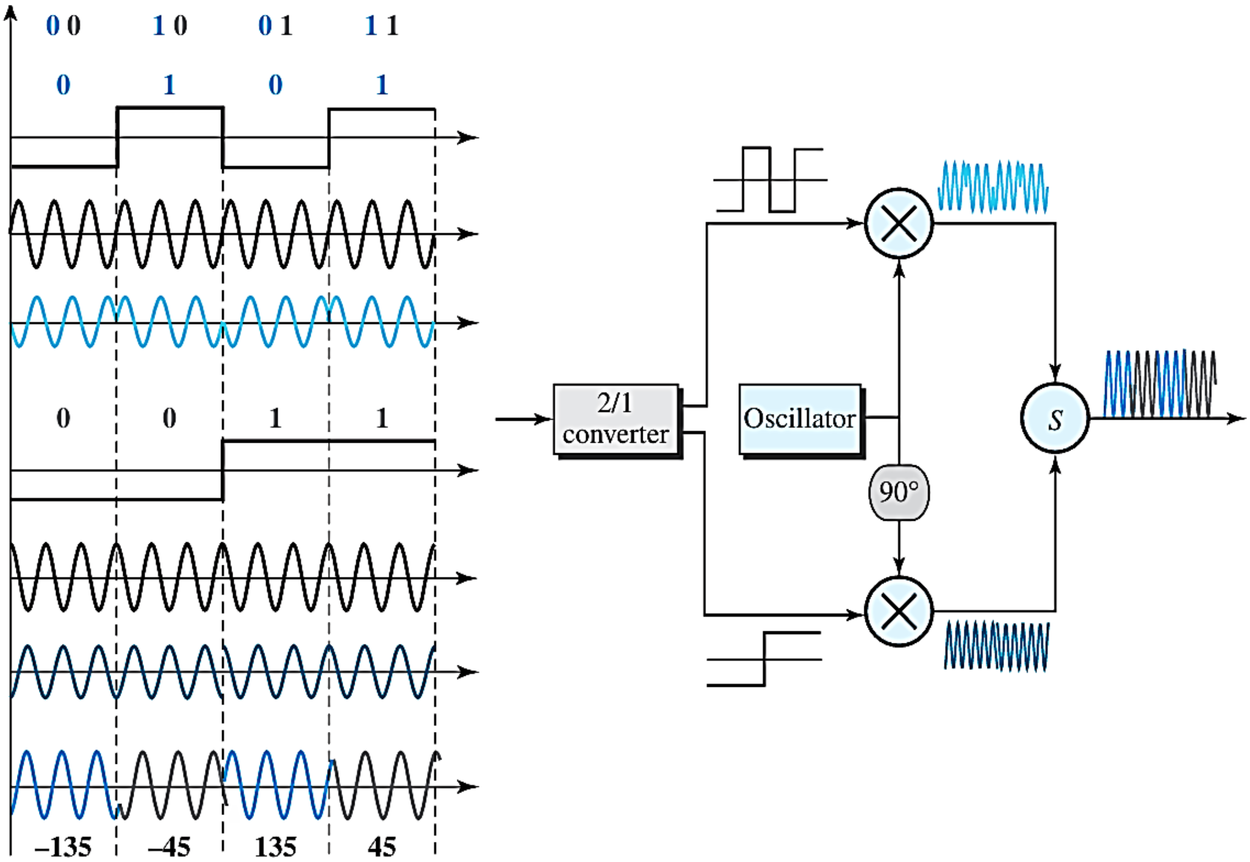


Figure 5.11 QPSK and its implementation

The two composite signals created by each multiplier are sine waves with the same frequency, but different phases. These signals are combined to produce the final QPSK signal. The resulting waveform has one of four possible phases: 45°, −45°, 135°, and −135°. Since there are four distinct signal elements ($L = 4$), each signal element represents 2 bits ($r = 2$).

Topic Related Questions:

1. Discuss the implementation of BPSK and QPSK.

Q. What are the major factors that make PSK superior to ASK and FSK?

The major factors that make Phase Shift Keying (PSK) superior to Amplitude Shift Keying (ASK) and Frequency Shift Keying (FSK) are:

1. **Noise Resistance:** In ASK, bit detection depends on the amplitude of the signal; in PSK, it is the phase. Noise can change the amplitude more easily than it can change the phase. In other words, PSK is less susceptible to noise than ASK.
2. **Fewer Carrier Signals Needed:** Unlike FSK, which uses multiple carrier frequencies to represent data, PSK only requires one carrier signal. This simplification reduces hardware complexity, power consumption, and the overall cost of the modulation process.
3. **Higher Efficiency:** PSK can transmit more bits per baud compared to ASK and FSK. This means that for a given bandwidth, PSK can transmit more data per unit time.

Q. How does the phase of the carrier vary for the message m(n) = {1, 0, 1, 1, 0, 1}?

In Binary Phase Shift Keying (BPSK), the phase of the carrier signal is shifted to represent binary data. Typically, a phase shift of 180° (or π radians) represents a binary 0, and a phase shift of 0° (or 0 radians) represents a binary 1.

For the given message m(n) = {1, 0, 1, 1, 0, 1}, the phase variations of the carrier signal would be:

m(1) = 1: 0° phase shift

m(2) = 0: 180° phase shift (no change)

m(3) = 1: 0° phase shift

m(4) = 1: 0° phase shift

m(5) = 0: 180° phase shift (no change)

m(6) = 1: 0° phase shift

Therefore, the phase of the carrier signal varies as: 0°, 180°, 0°, 0°, 180°, 0°.

5.1.5 Quadrature Amplitude Modulation

Q. Mention the advantages of QAM over PSK.

Quadrature Amplitude Modulation (QAM) offers several advantages over Phase Shift Keying (PSK):

1. **Higher Data Rates:** QAM modulates both the amplitude and phase of the carrier signal, allowing it to transmit more bits per symbol than PSK.
2. **Improved Bandwidth Efficiency:** Since QAM uses both amplitude and phase to convey data, it achieves higher bandwidth efficiency. This means more data can be transmitted over the same bandwidth.
3. **Flexibility:** QAM offers a wide range of variations, such as 16-QAM, 64-QAM, and 256-QAM, allowing for different levels of data rate and robustness depending on the channel conditions and system requirements.

5.2 ANALOG-TO-ANALOG CONVERSION/MODULATION

Analog-to-analog conversion, or Analog modulation, is the process of modifying one or more properties of a carrier signal, such as its amplitude, frequency, or phase, in proportion to an analog information signal. Modulation is needed if the transmission medium is bandpass in nature or if only a bandpass channel is available. A common example is radio broadcasting. The government assigns a narrow bandwidth to each radio station. The analog signal produced by each station is a low-pass signal, all in the same range. To be able to listen to different stations, these low-pass signals must be shifted to distinct frequency ranges through modulation.

Analog-to-analog conversion can be accomplished in three ways: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM) (See Figure 5.15).

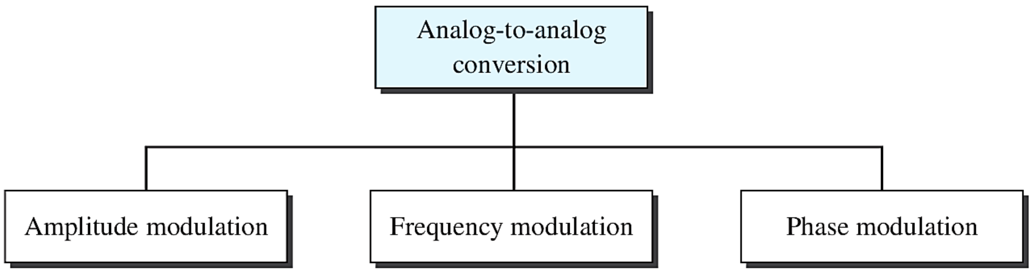


Figure 5.15 Types of analog-to-analog modulation

5.2.1 Amplitude Modulation (AM)

Amplitude Modulation (AM) is a type of analog modulation technique where the amplitude of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal. The frequency and phase of the carrier signal remain constant. Figure 5.16 shows how this concept works. The modulating signal is the envelope of the carrier.

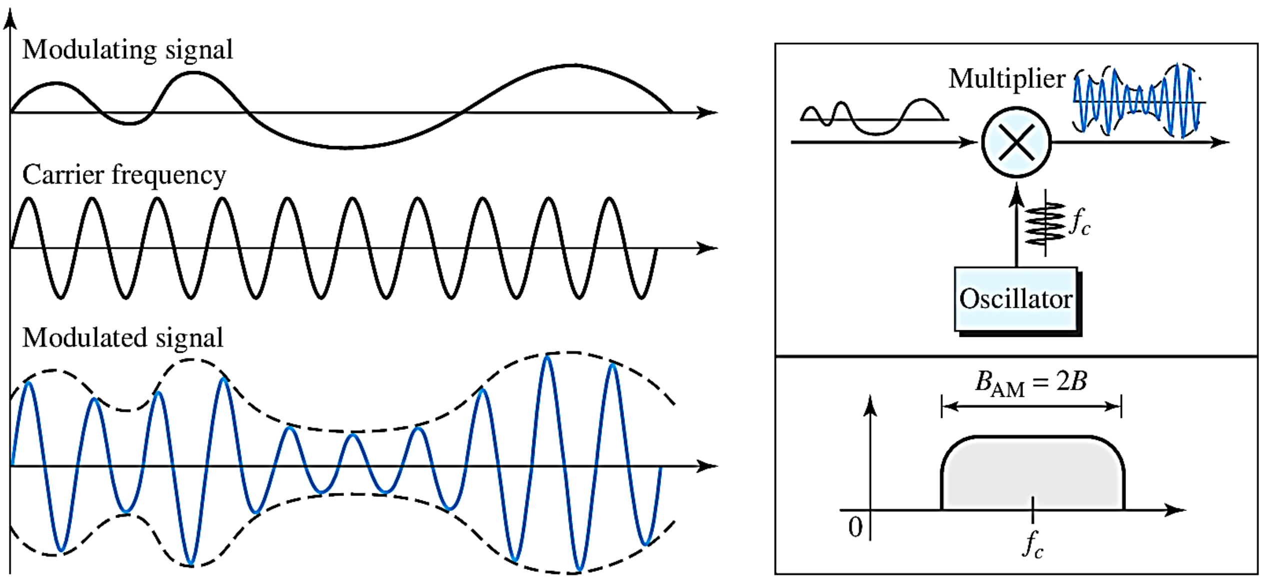


Figure 5.16 Amplitude modulation

As Figure 5.16 shows, AM is typically implemented using a simple multiplier, which adjusts the carrier signal's amplitude in accordance with the amplitude of the modulating signal.

Figure 5.16 also shows the bandwidth of an AM signal. The modulation creates a band width that is twice the bandwidth of the modulating signal and covers a range centered on the carrier frequency.

Topic Related Questions:

1. Define Amplitude Modulation and draw the frequency spectrum of Amplitude Modulated signal.

5.2.2 Frequency Modulation (FM)

Frequency Modulation (FM) is an analog modulation technique where the frequency of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal. The amplitude of the carrier signal remains constant, while its frequency changes to represent the information being transmitted. Figure 5.18 shows the relationships of the modulating signal, the carrier signal, and the resultant FM signal.

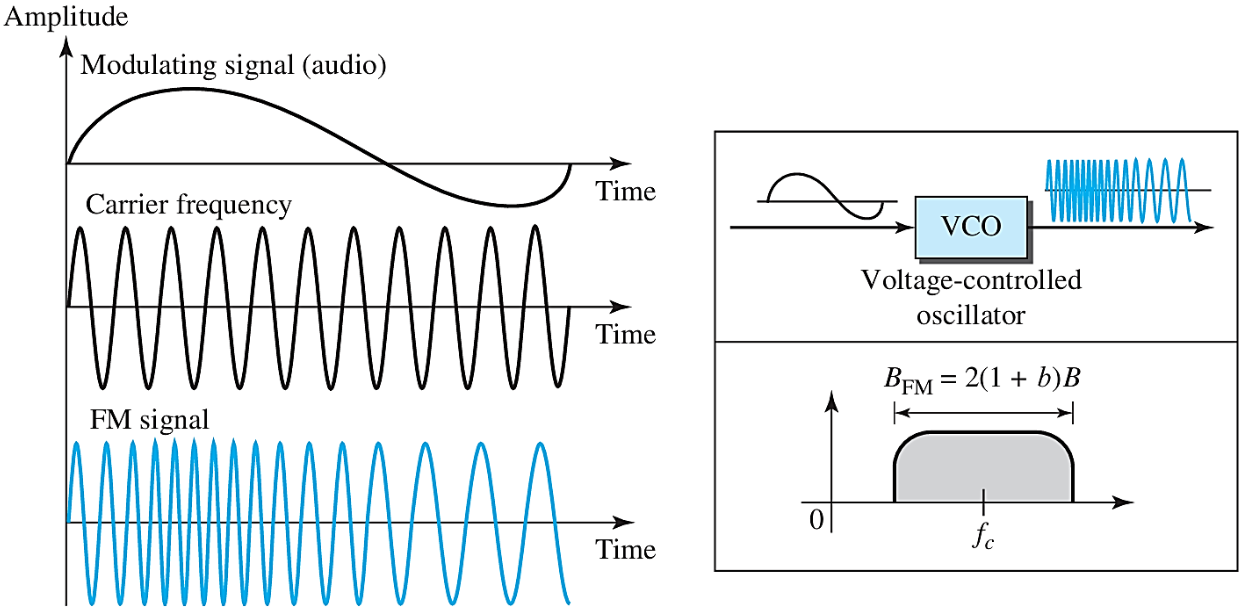


Figure 5.18 Frequency modulation

As Figure 5.18 shows, FM is normally implemented by using a voltage-controlled oscillator as with FSK. The frequency of the oscillator changes according to the input voltage which is the amplitude of the modulating signal.

Q. The carrier frequency of an FM broadcast transmission is 100 MHz and maximum frequency deviation is 75 KHz. Find the Bandwidth of the signal when the highest audio frequency modulating the carrier is 15 KHz.

Carson's Rule states that the approximate bandwidth (BW) of an FM signal is given by:

$$BW = 2(\Delta f + f_m)$$

where,

Δf is the maximum frequency deviation,

f_m is the maximum modulating (audio) frequency.

Given Information:

Carrier Frequency (f_c) = 100 MHz
Maximum Frequency Deviation (Δf) = 75 KHz
Highest Audio Frequency (f_m) = 15 KHz

Using Carson's Rule:

$$BW = 2(\Delta f + f_m) = 2(75 \text{ kHz} + 15 \text{ kHz}) = 180 \text{ kHz}$$

Therefore, the bandwidth of the FM signal is 180 kHz.

5.2.3 Phase Modulation (PM)

Phase Modulation (PM) is a type of analog modulation technique where the phase of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal. In PM, the frequency and amplitude of the carrier signal remain constant, while its phase is altered to encode the information. Figure 5.20 shows the relationships of the modulating signal, the carrier signal, and the resultant PM signal.

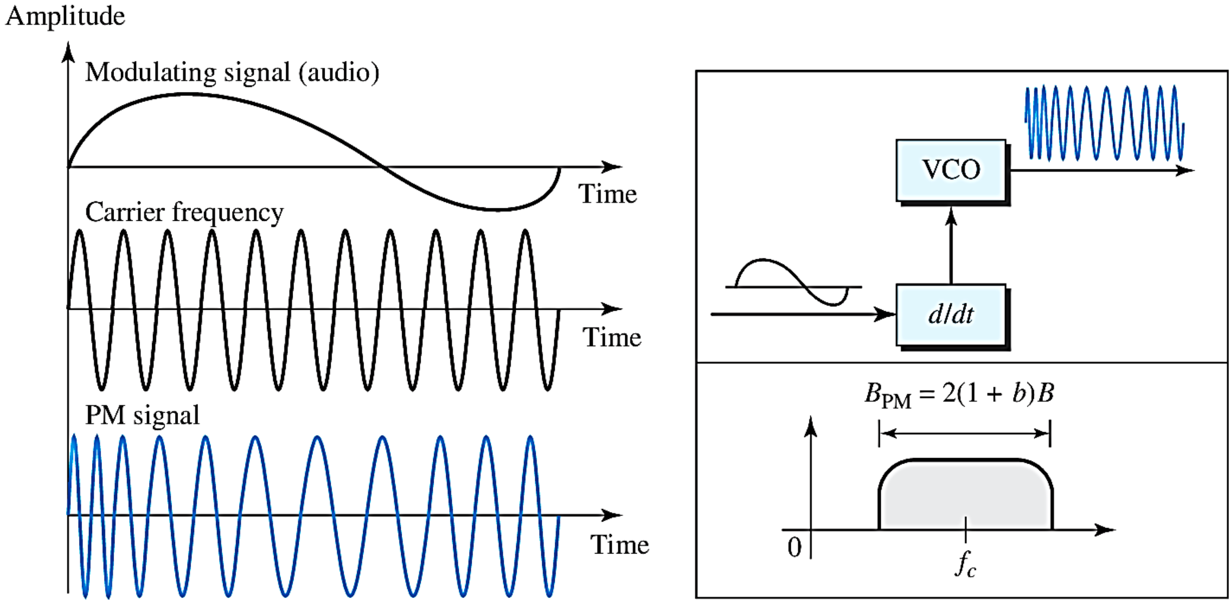


Figure 5.20 Phase modulation

As Figure 5.20 shows, PM is normally implemented by using a voltage-controlled oscillator along with a derivative. The frequency of the oscillator changes according to the derivative of the input voltage, which is the amplitude of the modulating signal.

Figure 5.20 also shows the bandwidth of a PM signal. The total bandwidth required for PM can be determined from the bandwidth and maximum amplitude of the modulating signal: $B_{PM} = 2(1 + \beta)B$ (Similar to FM as phase changes also affect frequency).

Q. Which of the three analog-to-analog conversion techniques (AM, FM, or PM) is the most susceptible to noise? Defend your answer.

Among the three analog-to-analog conversion techniques—Amplitude Modulation (AM), Frequency Modulation (FM), and Phase Modulation (PM)—AM is the most susceptible to noise.

In Am modulation, the amplitude of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal. Most types of noise, such as thermal or atmospheric noise, primarily affect the amplitude of a signal. Any unwanted variation in amplitude directly distorts the information carried by the AM signal, making it highly vulnerable to noise.

In contrast, Frequency Modulation (FM) encodes information in the frequency of the carrier signal. Since frequency is unaffected by amplitude-based noise, FM is significantly more resistant to noise than AM. Similarly, Phase Modulation (PM) encodes information in the phase of the carrier signal. Like FM, PM is less affected by amplitude noise, as phase variations are relatively immune to such disturbances.

Q. Given comparison of AM, FM & PM with suitable example.

Amplitude Modulation (AM)	Frequency Modulation (FM)	Phase Modulation (PM)
Amplitude of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal.	Frequency of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal.	Phase of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal.
Frequency and phase remain constant.	Amplitude and phase remain constant.	Amplitude and frequency remain constant.
Lowest noise immunity.	High noise immunity.	Moderate noise immunity.
The total bandwidth required for AM: $\boldsymbol{B_{AM} = 2B}$	The total bandwidth required for FM: $\boldsymbol{B_{FM}} = 2(1 + \beta)B$	The total bandwidth required for PM: $\boldsymbol{B_{PM}} = 2(1 + \beta)B$
AM requires a narrow bandwidth compared to FM and PM.	Requires a wider bandwidth than AM.	Bandwidth is similar to FM.
Example: AM Radio Broadcasting.	Example: FM Radio Broadcasting.	Example: Digital communication systems like Wi-Fi and Bluetooth.

Chapter 10: Error Detection and Correction

10.1 INTRODUCTION

10.1.1 Types of Errors

Errors during data transmission occur when bits are altered due to interference, noise, or other impairments. These errors can be categorized into two main types: single-bit error and burst error, based on the number and pattern of affected bits.

Single-Bit Error

A single-bit error occurs when only one bit of a data unit (e.g., byte, character, or packet) is changed during transmission. This means that a bit flips from 0 to 1 or 1 to 0 during transmission.

Figure 10.1 shows the effect of a single-bit error on a data unit. In this case, 00000010 was sent, but 00001010 was received.

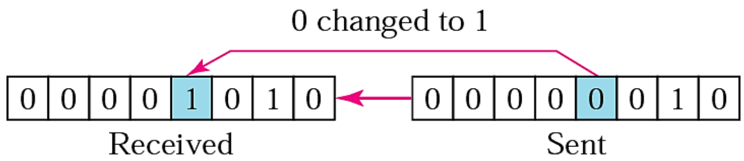


Figure 10.1 Single-bit error

Single-bit errors are rare/uncommon in serial data transmission because the duration of noise is usually longer than the time required to send one bit.

Burst Error

A burst error occurs when two or more bits in the data unit are altered during transmission.

Figure 10.2 shows the effect of a burst error on a data unit. In this case, 0100010001000011 was sent, but 0101110101100011 was received. Note that a burst error does not necessarily mean that the errors occur in consecutive bits. The length of the burst is measured from the first corrupted bit to the last corrupted bit. It may include bits that remain unaffected in between.

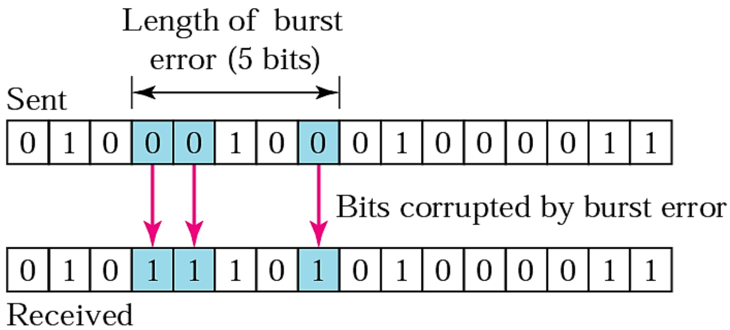


Figure 10.2 Burst error of length 5

A burst error is more likely to occur than a single-bit error. The duration of noise is normally longer than the duration of 1 bit, which means that when noise affects data, it affects a set of bits.

Topic Related Questions:

1. Briefly describe various types of error that occur during transmission.
2. Describe single bit error and burst error with appropriate example.

10.1.2 Redundancy

Q. Describe the concept of redundancy in error detection.

Error detection uses the concept of redundancy, which means adding extra bits to the data for detecting errors at the destination. These redundant bits are added by the sender and removed by the receiver. Their presence allows the receiver to detect or correct corrupted bits that may occur during transmission.

When a sender transmits a message, it includes redundant bits generated by a specific algorithm. Upon receiving the message, the receiver applies the same algorithm to verify the integrity of the data. If the computed result matches the received redundant data, the message is assumed to be error-free. Otherwise, errors are detected, prompting retransmission or other corrective actions. Figure 10.3 illustrates this process.

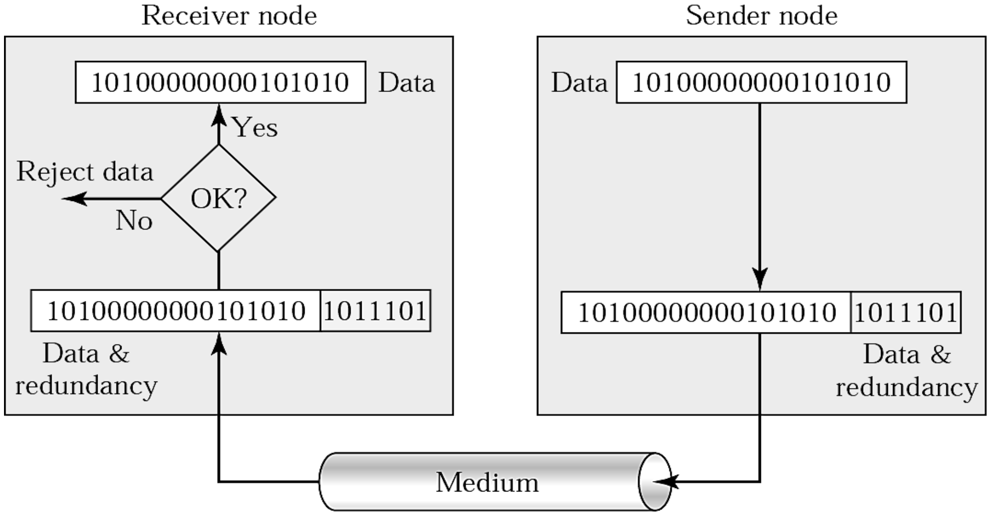


Figure 10.3 Redundancy

10.4 CYCLIC CODES

Modulo 2 Arithmetic:
The process of modulo-2 binary division is the same as the familiar division process we use for decimal numbers. Just that instead of subtraction, we use XOR here.

- In each step, a copy of the divisor (or data) is XORed with the k bits of the dividend (or key).
- The result of the XOR operation (remainder) is $(n-1)$ bits, which is used for the next step after 1 extra bit is pulled down to make it n bits long.
- When there are no bits left to pull down, we have a result. The $(n-1)$-bit remainder which is appended at the sender side.

Polynomial division: Polynomial division in the context of CRC (Cyclic Redundancy Check) is a method of dividing binary polynomials using modulo-2 arithmetic.

10.4.1 Cyclic Redundancy Check

The Cyclic Redundancy Check (CRC) is a widely used error-detection technique that relies on polynomial division. It involves the following steps:

1. **Polynomial Division:** A predetermined polynomial, also known as the generator polynomial, is chosen. This polynomial is used for division. At the sender's end, the data is treated as a binary number and divided by the predetermined generator polynomial.
2. **Appending the CRC Code:** The remainder from this division, which represents the CRC code, is appended to the data and sent to the receiver.
3. **Error Detection:** At the receiver's end, the received data (including the CRC code) is again divided by the same generator polynomial. If the remainder of this division is zero, the data is assumed to be error-free. Otherwise, it indicates the presence of errors.

CRC generator/encoder

1. The CRC generator takes the original data bits (message) and appends $k - 1$ zeros to the end, where k is the number of bits of the divisor (or key).
2. The extended data (message + appended zeros) is divided by the divisor polynomial (or key) using binary (modulo-2) division.
3. The remainder from the division is the CRC code. This code is appended to the original data and sent to the receiver.

Example 10.1:

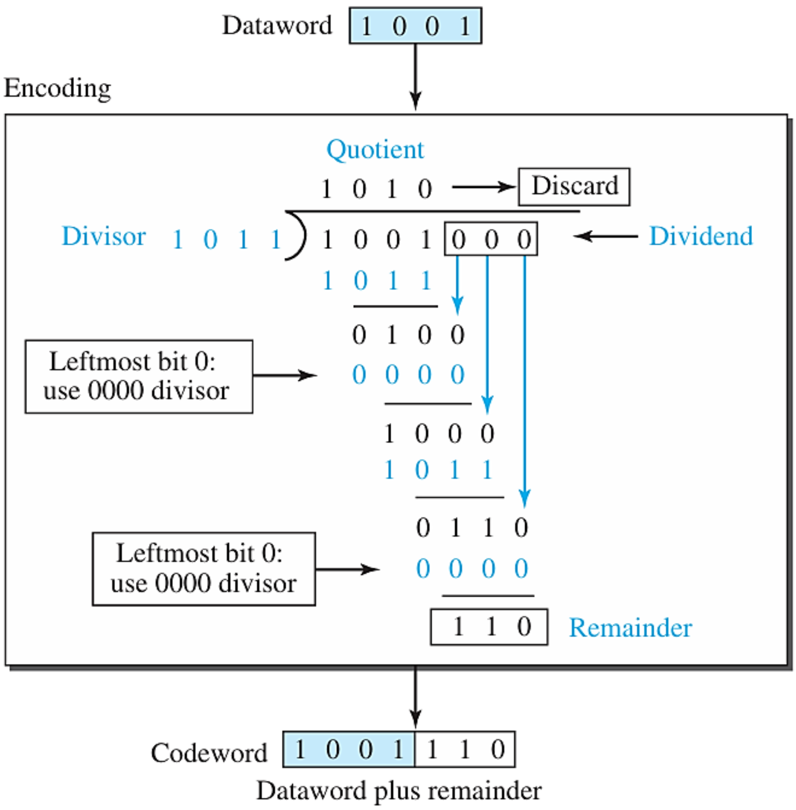


Figure 10.6 Division in CRC encoder

Example 10.2:

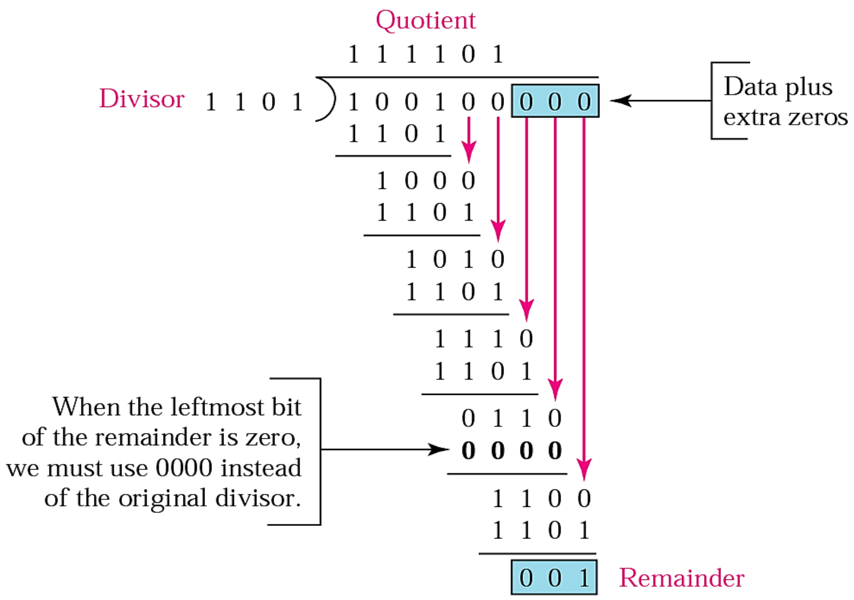


Figure: Binary division in a CRC generator

Tutorial: https://www.youtube.com/watch?v=A9g6rTMblz4

CRC checker/decoder

1. The receiver obtains the transmitted frame, which consists of the original data bits followed by the CRC code (redundant bits).
2. The receiver divides the entire frame (data + CRC) by the same divisor (polynomial) used by the sender during CRC generation.
3. If the remainder is zero, the data is considered error-free; otherwise, an error is detected.

Example 10.3: Checker for example 10.1

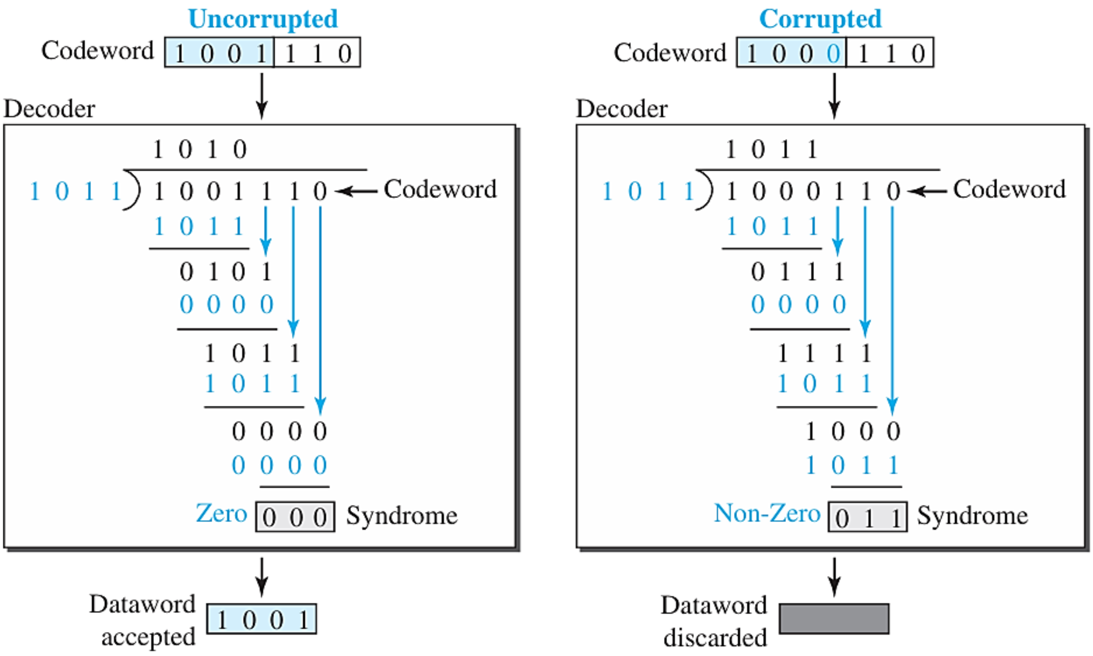


Figure 10.7 Division in the CRC decoder for two cases

Example 10.4: Checker for example 10.2

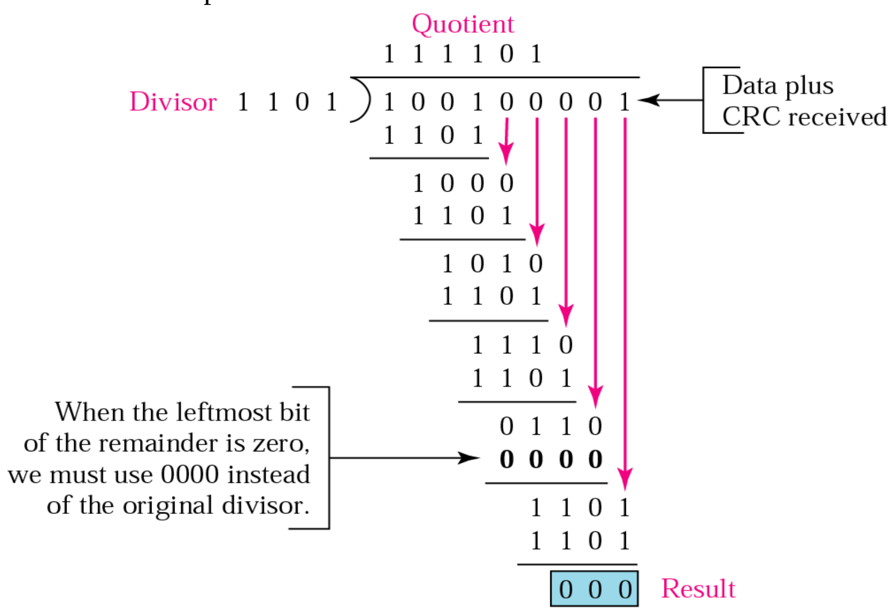


Figure: Binary division in CRC checker

Tutorial: https://www.youtube.com/watch?v=wQGwfBS3gpk

Q. Given a remainder of 111, a data unit of 10110011 and a divisor of 1001. Is there any error in data unit?

Append remainder to data unit: 10110011111. Show division steps and calculate remainder. Remainder is 000, which indicates no error.

10.5 CHECKSUM

Checksum is a simple error-detection technique used in data communication to verify the integrity of transmitted data. In this method, the sender divides the data into equal segments of n bits. The checksum generator then creates an extra n bit unit called the checksum, which is sent with the message. At the receiver's end, a checksum checker recalculates the checksum based on the received data and transmitted checksum. If the new checksum is all zeros, the data is considered error-free and accepted; otherwise, the data is rejected/discarded (Figure 10.12).

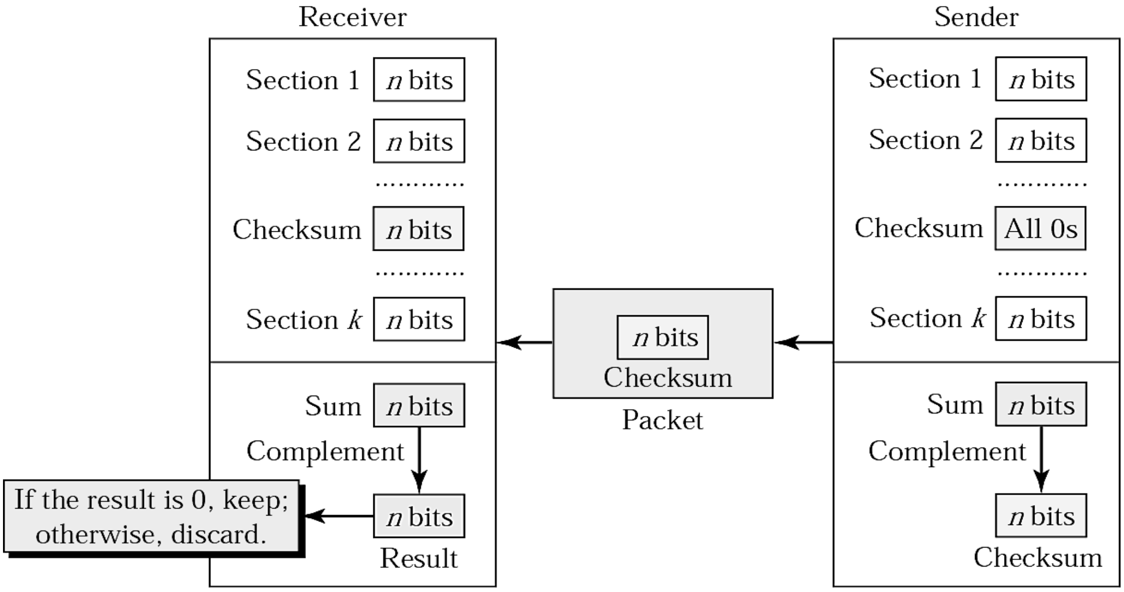


Figure 10.12 Checksum

Checksum Generator

The checksum generator is a tool or algorithm used at the sender's side to create a checksum value for error detection. It follows these steps:

1. The data unit is divided into k sections, each of n bits.
2. The value of the checksum is initially set to zero.
3. All the sections are added together using binary addition. If the sum produces a carry, it is wrapped around and added to the result.
4. The final sum is complemented (inverted) to generate the checksum.

The checksum is appended to the data and sent to the receiver.

Example 10.5: Suppose the following block of 16 bits is to be sent using a checksum of 8 bits.

10101001 00111001

The 16-bit block is divided into two 8-bit sections:

Section 1: 10101001

Section 2: 00111001

The two sections are added using one's complement addition

	10101001
	00111001
Sum	11100010
Checksum	00011101

The final transmitted/sent pattern includes the original data and the checksum:

10101001 00111001 00011101

Checksum Checker

The checksum checker is a tool or algorithm used at the receiver's side to verify the integrity of transmitted data. It follows these steps:

1. The receiver divides the received data, including the checksum, into equal-sized sections of n bits.
2. All sections are added together using binary addition. If the sum generates a carry, it is wrapped around and added to the result.
3. The final sum is complemented (inverted).
4. If the result is all zeros, the data is considered error-free and accepted; otherwise, errors are detected, and the data is rejected or requested for retransmission.

Example 10.6: Suppose the receiver receives the pattern sent in Example 10.5 without any errors.

10101001 00111001 00011101

When the receiver adds the three sections, it will get all 1s, which, after complementing, is all 0s and shows that there is no error.

	10101001
	00111001
	00011101
Sum	11111111
Complement	00000000

Example 10.7: Now suppose there is a burst error of length 5 that affects 4 bits.

10101111 11111001 00011101

When the receiver adds the three sections, it gets

	10101111
	11111001
	00011101
Partial Sum	1 11000101
Carry	1
Sum	11000110
Complement	00111001

Since the complement is not all zeros, the pattern is corrupted.

Tutorial: https://www.youtube.com/watch?v=AtVWnyDDaDI

Topic Related Questions:

1. Briefly describe checksum error detection method.
2. Briefly describe the checksum generator.
3. Suppose the receiver of checksum method receives the patterns as follows: 10101111, 11111001, and 00011101. Check the data whether it is erroneous or not.

Chapter 11: Data Link Control (DLC)

Q. What is data link? List some of the requirements needed in data link control for effective data communication.

Data link: A data link refers to the communication connection between two directly connected nodes in a network. It is responsible for transferring data across a physical link while ensuring reliable and error-free communication.

Some of the key requirements needed in data link control for effective data communication are:

1. **Framing:** Dividing the stream of bits into manageable data units called frames to facilitate synchronization and proper interpretation by the receiver.
2. **Flow Control:** Managing the rate of data transmission to ensure that the sender does not overwhelm the receiver.
3. **Error Detection and Correction:** Implementing mechanisms like checksums, CRC, and ARQ to detect and correct errors in transmitted data.
4. **Addressing:** Identifying the sender and receiver of data for proper delivery in multi-node networks.
5. **Link Management:** Establishing, maintaining, and terminating communication links between devices.
6. **Access Control:** Regulating access to the shared communication medium in multi-access environments to avoid collisions.
7. **Reliability:** Ensuring data is delivered without loss, duplication, or corruption.
8. **Efficiency:** Minimizing overhead and maximizing data transfer speed.

These requirements ensure that communication between devices is smooth, accurate, and efficient.

11.2 Flow and Error Control

Q. Why flow control is necessary?

Flow control is necessary in data communication to ensure efficient and reliable transmission of data between sender and receiver. Without flow control, a faster sender might overwhelm a slower receiver, leading to buffer overflow. This can result in data loss as the receiver cannot process the incoming data quickly enough. Additionally, flow control prevents network congestion, ensuring optimal performance and minimizing packet loss.

Note: Each receiving device has a block of memory, called a buffer, reserved for storing incoming data until they are processed.

11.4 NOISELESS CHANNELS

11.4.2 Stop-and-Wait Protocol

Q. Explain the stop and wait flow control methods.

The Stop-and-Wait flow control method is a simple mechanism used to manage data transmission over a network. In this method, the sender sends a single frame of data and then waits for an acknowledgment from the receiver before sending the next frame. This ensures that data is only sent when it can be confirmed received correctly, preventing data loss due to packet collisions or other issues. Figure 11.8 illustrates the mechanism.

This method is straightforward but can be inefficient, especially over long distances or high latency networks, as it results in idle time waiting for acknowledgments.

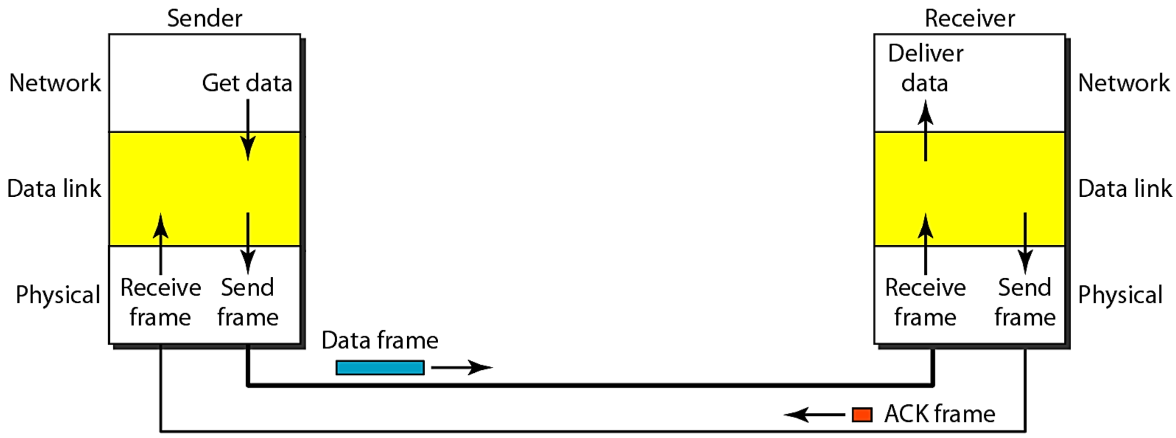


Figure 11.8 Design of Stop-and-Wait Protocol

11.5 NOISY CHANNELS

11.5.1 Stop-and-Wait Automatic Repeat Request

Q. Describe how to handle Stop-and-Wait automatic repeat request flow and error control mechanism with lost of frame.

The Stop-and-Wait ARQ (Automatic Repeat Request) is a flow control and error control mechanism that ensures reliable data transmission between a sender and a receiver. In this approach, the sender transmits a single frame and then waits for an acknowledgment (ACK) from the receiver before sending the next frame. If the receiver successfully receives the frame, it sends back an acknowledgment. If the sender does not receive an acknowledgment within a specified timeout period, it assumes the frame or the ACK was lost and retransmits the same frame.

For handling lost frames, the sender keeps a copy of the last transmitted frame and starts a timer after sending the frame. If the timer expires before an acknowledgment (ACK) is received, the sender assumes that the frame has been lost or the acknowledgment has not arrived. In this case, the sender retransmits the same frame, restarts the timer, and waits for an acknowledgment again. This cycle continues until the sender receives the acknowledgment, ensuring that each frame is correctly received by the receiver. This mechanism helps maintain data integrity by repeatedly sending the frame until it is accurately received, although it can lead to inefficiencies due to the waiting periods involved.

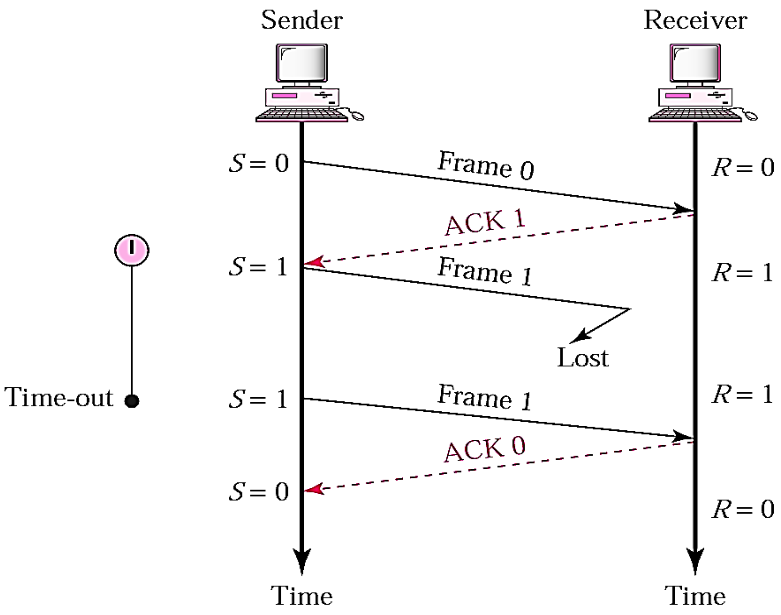


Figure: Stop-and-Wait ARQ, lost frame

Q. Describe how to handle Stop-and-Wait automatic repeat request flow and error control mechanism with lost of acknowledgement.

In the Stop-and-Wait ARQ protocol, the sender transmits a single frame and waits for an acknowledgment (ACK) before sending the next frame. If the sender does not receive an acknowledgment within a specified timeout period, it assumes the frame or the ACK was lost and retransmits the same frame. In the case of a lost acknowledgment (ACK), specific steps are taken to ensure the sender can identify the issue and retransmit the correct frame.

When the sender transmits a frame to the receiver, it starts a timer and waits for an acknowledgment. If the receiver successfully receives the frame and detects no errors, it sends an acknowledgment back to the sender. However, in cases where the acknowledgment is lost in transit or delayed, the sender's timer will eventually expire because it does not receive the expected ACK within a specified time. Upon the timer's expiration, the sender assumes that the frame or acknowledgment has been lost and retransmits the same frame.

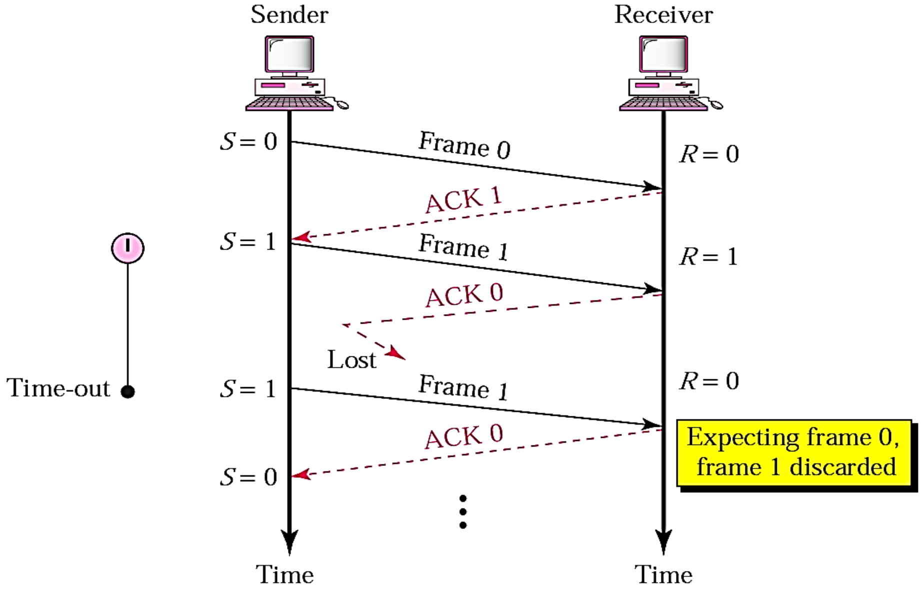


Figure: Stop-and-Wait ARQ, lost ACK frame

11.6 HDLC

High-level Data Link Control (HDLC) is a bit-oriented protocol for communication over point-to-point and multipoint links. It implements the Stop-and-Wait protocol.

Q. Draw and explain the HDLC frame format.

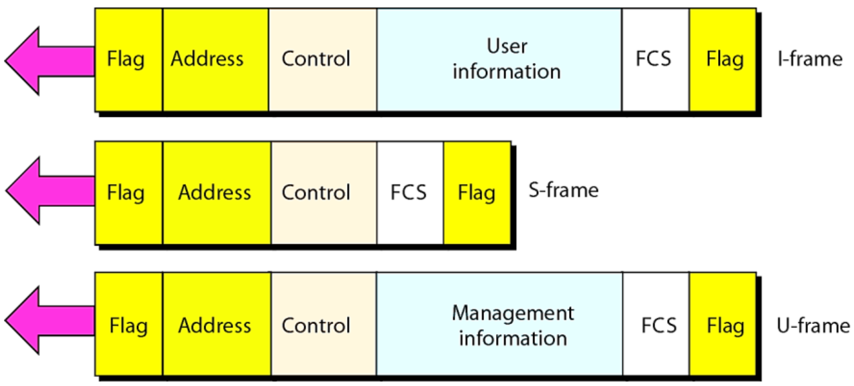


Figure 11.16 HDLC frames

Each frame in HDLC may contain up to six fields, as shown in Figure 11.16: a beginning flag field, an address field, a control field, an information field, a frame check sequence (FCS) field, and an ending flag field. In multiple-frame transmissions, the ending flag of one frame can serve as the beginning flag of the next frame.

Here's an overview of the HDLC frame fields:

- **Flag field:** This field contains synchronization pattern 01111110, which identifies both the beginning and the end of a frame.
- **Address field:** This field contains the address of the secondary station. If a primary station created the frame, it contains a to address. If a secondary station creates the frame, it contains a $from$ address. The address field can be one byte or several bytes long, depending on the needs of the network.
- **Control field:** The control field is one or two bytes used for flow and error control. It determines the type of frame and defines its functionality.
- **Information field:** The information field contains the user's data from the network layer or management information. Its length can vary from one network to another.
- **FCS field:** The frame check sequence (FCS) is the HDLC error detection field. It can contain either a 2- or 4-byte CRC.

Q. What is the difference between polling and selection? (Chapter 12)

Polling is a method where a primary device repeatedly checks (or "polls") multiple secondary devices to see if they are ready to communicate or have data to send. The primary device queries all secondary devices sequentially to determine which one is ready. It is used when the primary device is unsure which secondary device has data to send (e.g., in shared communication systems). However, this method may introduce delays since it involves querying multiple devices even if they are not ready.

On the other hand, selection is a method where the primary device directly addresses a specific secondary device, instructing it to send or receive data. This approach is used when the primary device wants to send data to a particular secondary device. Selection is generally more efficient than polling.