CSC 422 (DATA COMMUNICATION AND INFORMATION THEORY)

COMPILED BY

ABIKOYE, OLUWAKEMI CHRISTIANA(Ph.D)

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF COMMUNICATION AND INFORMATION SCIENCES

UNIVERSITY OF ILORIN

2019/2020 SESSION

1. DATA COMMUNICATION AND INFORMATION THEORY

1.1 A Mathematical Theory of Communication

"A Mathematical Theory of Communication" is an influential 1948 article by mathematician Claude E. Shannon. It was renamed "*The* Mathematical Theory of Communication" in the book, a small but significant title change after realizing the generality of this work.

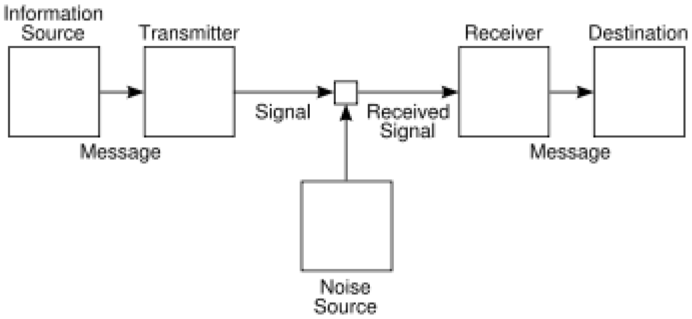


Figure 1.1: Shannon’s General Communications system

Shannon's diagram of a general communications system, which shows the process that produces a message.

The article was the founding work of the field of information theory. It was later published in 1949 as a book titled *The Mathematical Theory of Communication* (ISBN 0-252-72546-8), which was published as a paperback in 1963 (ISBN 0-252-72548-4). The book contains an additional article by Warren Weaver, providing an overview of the theory for a more general audience. Shannon's article laid out the basic elements of communication:

- An information source that produces a message
- A transmitter that operates on the message to create a signal which can be sent through a channel
- A channel, which is the medium over which the signal, carrying the information that composes the message, is sent
- A receiver, which transforms the signal back into the message intended for delivery
- A destination, which can be a person or a machine, for whom or which the message is intended

It also developed the concepts of information entropy and redundancy, and introduced the term bit as a unit of information.

1.2 Introduction to Data communication

Data: Data refers to the information or message, which is present in the form that is agreed upon by user and creator of data (mostly Digital data)

Data Communication: is exchange of data between two devices via some form of transmission medium.

Message or Signal: is electrical or electromagnetic wave sent through medium from one point to another, which contains encoded message. A messages can be in the form of sound, text, numbers, pictures, video or combinations of these.

Sender: A sender is device which sends the message, example : computer, workstation, video camera, telephone etc.

Medium: It is physical path over which data travels from a sender to receiver.

Receiver: A receiver is a device which receives the message, example : computer, TV receiver, workstation, telephone receiver, radio receiver etc.

Protocol: A protocol is defined as the set of rules which governs data communication. The connection of two devices takes places via the communication medium but the actual communication between them will take place with take place with the help of a protocol.

1.3 Elements of information theory

Information theory can be defined as the mathematical study of the coding of information in the form of sequence of symbols, impulses etc and of how rapidly such information can be transmitted. For example, through computer circuits or telecommunications channels.

Information theory is a branch of applied mathematics, electrical engineering, and computer science involving the quantification, storage, and communication of information. Information theory was originally developed by Claude E. Shannon to find fundamental limits on signal processing and communication operations such as data compression. Since its inception in a landmark 1948 paper by Shannon entitled "A Mathematical Theory of Communication".

Information theory studies the transmission, processing, utilization, and extraction of information. Abstractly, information can be thought of as the resolution of uncertainty. In the case of communication of information over a noisy channel, this abstract concept was made concrete in 1948 by Claude Shannon in his paper "A Mathematical Theory of Communication", in which "information" is thought of as a set of possible messages, where the goal is to send these messages over a noisy channel, and then to have the receiver reconstruct the message with low probability of error, in spite of the channel noise

A key measure in information theory is "entropy". Entropy quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process. For example, identifying the outcome of a fair coin flip (with two equally likely outcomes) provides less information (lower entropy) than specifying the outcome from a roll of a die (with six equally likely outcomes). Some other important measures in information theory are mutual information, channel capacity, error exponents and relative entropy.

Information theory often concerns itself with measures of information of the distributions associated with random variables. Important quantities of information are entropy, a measure of information in a single random variable, and mutual information, a measure of information in common between two random variables. The former quantity is a property of the probability distribution of a random variable and gives a limit on the rate at which data generated by independent samples with the given distribution can be reliably compressed. The latter is a property of the joint distribution of two random variables, and is the maximum rate of reliable communication across a noisy channel in the limit of long block lengths, when the channel statistics are determined by the joint distribution.

1.3.1 Entropy

If is the set of all messages $\{x_1, \dots, x_n\}$ that X could be, and $p(x)$ is the probability of some $x \in \mathbb{X}$, then the entropy, H, of X is defined:

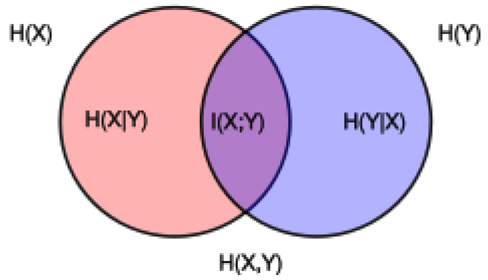
$$H(X) = \mathbb{E}_X[I(x)] = -\sum_{x \in \mathbb{X}} p(x) \log p(x).$$

(Here, $I(x)$ is the **self-information**, which is the entropy contribution of an individual message, and $_X$ is the expected value.) A property of entropy is that it is maximized when all the messages in the message space are equiprobable *p*(*x*) = 1/*n*; i.e., most unpredictable, in which case *H*(*X*) = log *n*.

1.3.2 Measures of Information theory

Mutual Information

Mutual information (MI) of two random variables is a measure of the mutual dependence between the two variables. More specifically, it quantifies the "amount of information" (in units such as bits) obtained about one random variable, through the other random variable. The concept of mutual information is intricately linked to that of entropy of a random variable, a fundamental notion in information theory, that defines the "amount of information" held in a random variable



Venn diagram for various information measures associated with correlated variables X and Y. The area contained by both circles is the joint entropy H(X,Y). The circle on the left (red and violet) is the individual entropy H(X), with the red being the conditional entropy H(X|Y). The circle on the right (blue and violet) is H(Y), with the blue being H(Y|X). The violet is the mutual information I(X;Y).

Channel Capacity

Channel capacity is the tight upper bound on the rate at which information can be reliably transmitted over a communication channel.

By the noisy-channel coding theorem, the channel capacity of a given channel is the limiting information rate (in units of information per unit time) that can be achieved with arbitrarily small error probability.

Error Exponent

In information theory, the **error exponent** of a channel code or source code over the block length of the code is the logarithm of the error probability. For example, if the probability of error of a decoder drops as $e^{-n\alpha}$, where n is the block length, the error exponent is α

Relative entropy

Relative entropy is a measure of the difference between two probability distributions *P* and *Q*. It is not symmetric in *P* and *Q*. In applications, *P* typically represents the "true" distribution of data, observations, or a precisely calculated theoretical distribution, while *Q* typically represents a theory, model, description, or approximation of *P*.

2. SIGNALS

Signal is an electrical transmission of alternating current (AC) on network cabling that is generated by a networking component such as a network interface card (NIC). An electromagnetic signal is transmitted through air, vacuum to satellite or antenna to mobile. Signals can be either analog or digital.

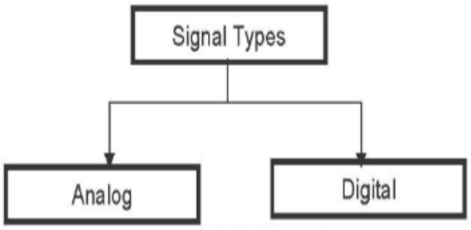


Figure 2.1: Signal types

2.1 Analog Signal

An analog signal is a continuous wave form that changes smoothly over time. An analog signal can take on any value in a specified range of values. As the wave moves from value A to B, it passes through and includes an infinite number of values along its path. A simple example is alternating current (AC), which continually varies between about +110 volts and -110 volts in a sine wave fashion 50 times per second. A more complex example of an analog signal is the time-varying electrical voltage generated when a person speaks into a dynamic microphone or telephone.

Analog signals such as telephone speech contain a wealth of detail, but are not readily accessible to computers unless they are converted to digital form using a device such as an analog-to-digital converter (ADC). Analog signals are usually specified as a continuously varying voltage over time and can be displayed on a device known as an oscilloscope. **Amplitude** is absolute value of signal at an instance. The maximum voltage displacement of a periodic (repeating) analog signal is called its amplitude, and the shortest distance between crests of a periodic analog wave is called its **wavelength**.

An example of analog data is the human voice. When somebody speaks, a continuous wave is created in the air. Analog data --voice, video --continuously varying patterns of different intensity (amplitude). Analog signal can be classified as simple or composite. A simple analog signal, a sine wave, cannot be decomposed into simpler signals. A composite analog signal is composed of multiple sine waves. Three characteristics namely – amplitude, frequency and phase fully describes a sine wave.

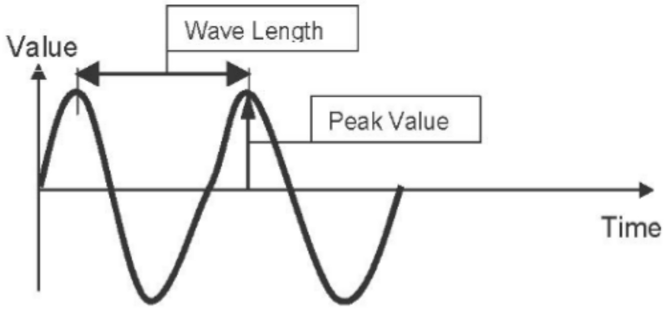


Figure 2:2: Analog Signal

2.1.1 Characteristics of a sine wave

Peak Amplitude

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

Period and Frequency

Period refers to the amount of time, in second, a signal needs to complete one cycle. Frequency refers to the number of periods in one second.

Period is expressed in seconds and frequency is expressed in hertz (Hz)

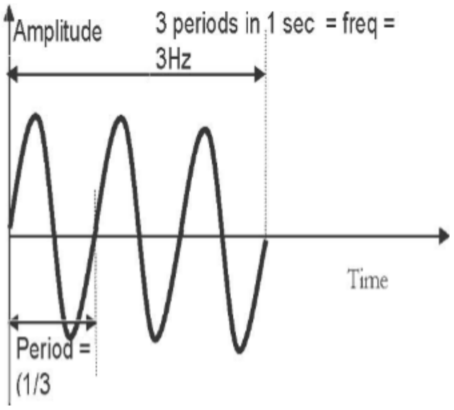


Figure 2.3: Period and frequency

Phase

Phase describes the position of the waveform relative to the time zero. Phase is measured in degrees or radians. The Phase is denoted by symbol [Φ]. A periodic signal is represented by an equation.

$$x(t)=x(t)+t0 \text{ OR } x(t)=A\sin(2\pi ft + \Phi]$$

Table 1.1: Units of Period and Frequency

Unit time	Equivalent	Unit frequency	Equivalent
Seconds (s)	1 s	Hertz (Hz)	1 Hz
Milliseconds (ms)	10_{-3} s	Kilohertz (KHz)	10_{3} Hz = 1 KHz
Microseconds (μs)	10_{-6} s	Megahertz (MHz)	10_{6} Hz = 1 MHz
Nano second	10_{-9} s	Gegahertz (GHz)	10_{9} Hz = 1 GHz

2.2 Digital Signal

Digital signal is the transmission of signals that vary discretely with time between two values of some physical quantity, one value representing the binary number 0 and the other representing 1. Digital signals use discrete values for the transmission of binary information over a communication medium such as a network cable or a telecommunications link. On a serial transmission line, a digital signal is transmitted 1-bit at a time.

A digital signal is discrete. It has only a limited number of definite discreet values, as 1 and 0. An example of digital data is data stored in memory of a computer in the form of 0's and 1's. For example, a 1 can be encoded as a positive voltage and a 0 as zero voltage. Digital data --text, digitized images --takes discrete values, usually binary (0,1). Example of digitized text is the ASCII code. 8 -bits so 255 patterns including -upper and lower case
characters, integers 0-9, special characters and some "control" characters are used in communication. Bit interval and baud rate are used to describe digital signals.

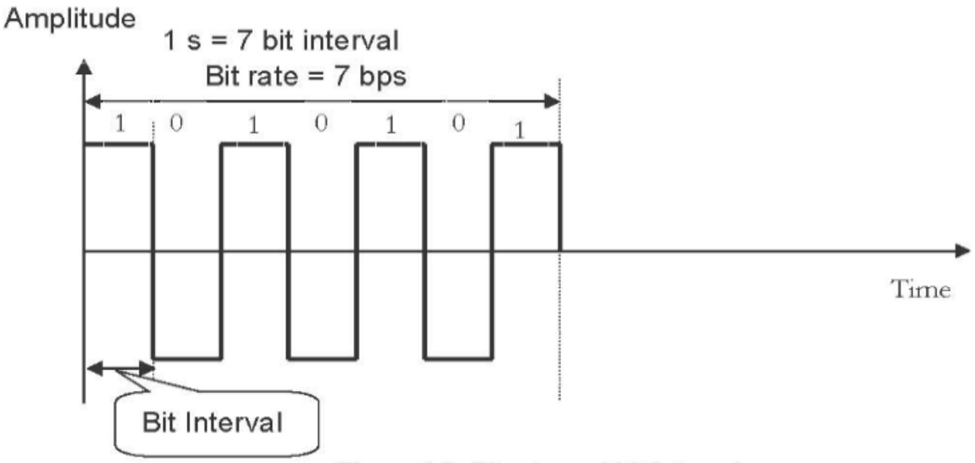


Figure 2.4:   Bit rate and bit interval

Bit Interval

The bit interval is the time required to send one single bit. The bit rate is the number of bit intervals per second. This means that the bit rate is number of bits sent in one second, usually expressed in bits per second (bps).

Baud rate

Baud rate refers to the number of signal units per second that are required to

represent those bits. Baud rate is less than or equal to the bit rate.
The difference between baud rate and bit rate occurs as they define different but related information. Thus Baud rate is effective measure of information transmitted and bit rate is measure of the data transmitted (which might include error correcting codes, frame, frame-packet numbers etc.].

Example 1

A signal carries three bits in each signal element. If 1200 signal elements are sent per second, find the baud rate and the bit rate.

Solution

Baud rate = Number of signal elements = 1200 bps

Bit rate = baud rate × Number of bits per signal element

= 1200 × 3

= 3600 bps

Example 2

The bit rate of a signal is 2000. If each signal element carries five bits, what is the baud rate?

Solution

Baud rate = Bit rate / Number of bits per signal element

= 2000 / 5

= 400 bps

3. FOURIER ANALYSIS

Fourier analysis is a method of defining periodic waveform s in terms of trigonometric functions. The method gets its name from a French mathematician and physicist named Jean Baptiste Joseph, Baron de Fourier, who lived during the 18th and 19th centuries. Fourier analysis is used in electronics, acoustics, and communications.

Many waveforms consist of energy at a fundamental frequency and also at harmonic frequencies (multiples of the fundamental). The relative proportions of energy in the fundamental and the harmonics determines the shape of the wave.

The wave function (usually amplitude , frequency, or phase versus time) can be expressed as of a sum of sine and cosine function s called a Fourier series , uniquely defined by constants known as Fourier coefficient s. If these coefficients are represented by a , a_1 , a_2 , a_3 , ..., a_n , ... and b_1 , b_2 , b_3 , ..., b_n , ..., then the Fourier series $F(x)$, where x is an independent variable (usually time), has the following form:

$F(x) = a/2 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ldots$
$+ a_n \cos nx + b_n \sin nx + \ldots$

In Fourier analysis, the objective is to calculate coefficients a , a_1 , a_2 , a_3 , ..., a_n and b_1 , b_2 , b_3 , ..., b_n up to the largest possible value of n . The greater the value of n (that is, the more terms in the series whose coefficients can be determined), the more accurate is the Fourier-series representation of the waveform.

4. DATA TRANSMISSION

4.1 Analog VERSUS Digital transmission

ANALOG TRANSMISSION --a means of transmitting ONLY analog signals.

- Data can be analog or digital; signal is always analog.
- Propagation can be over guided [wired, coaxial, optical fiber, cable] or unguided medium (space, atmosphere).
- Analog signal will become weaker in signal strength (attenuate) over distance and will be impaired by noise.
- An AMPLIFIER will boost the energy of the signal but also the noise. Noise is a undesirable random electrical transmission on network cabling that is generated by networking components such as network interface cards (NICs) or induced in cabling by proximity to electrical equipment that generates electromagnetic interference (EMI).
- No coding is possible and thus no self-error correction is possible.

DIGITAL TRANSMISSION --a means of transmitting both digital and analog signals. Usually assume that the signal is carrying digital (or digitized) data.

- Digital transmission can propagate to a limited distance before attenuation distorts the signal and compromises the data integrity.

- A REPEATER retrieves the (digital) signal; recovers the (digital) data, e.g., a pattern of 1's and 0's; and retransmits a new signal. Digital transmission is the preferred method for several reasons :
- Equipment used for digital transmission is cheaper as compared to analog transmission.
- Use of repeaters, which recover the data and retransmit, are preferred over amplifiers, which boost both signal and noise.
- Errors are not cumulative and so it is possible to transmit over longer distances, using lower quality guided medium with better data integrity.
- Multiplexing -transmission links have high bandwidth and must propagate multiple signals simultaneously to utilize the bandwidth. In digital transmission, time- division multiplexing is used. Signals share the same medium over different time slots. This is easier than analog transmission where the analog signals occupy different frequency spectrum (frequency-division).
- Encryption of signal is possible for security and privacy.
- Coding is possible and self-error correction is possible.

4.2 Data rate in digital communication (Measure of Communication)

How fast data can be sent, in bits per second, through a channel depends on three factors.

1 The bandwidth available.

2 The levels of signals.

3 The quality of the channel.

Two theoretical formulas were developed to calculate the data rate : one by Nyquist for a noiseless channel, another by Shannon for a noisy channel.

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

$$Bitrate = 2 \times Bandwidth \ \times \log_2 L$$

Where, Bandwidth is the bandwidth of the channel

L is the number of signal levels used to represent data, and Bitrate is the bit rate in bits per second.

Example 3

Consider a noiseless channel with a bandwidth of 2000 Hz transmitting a signal with two signal levels. Calculate the bit rate.

Solution

Bitrate= 2 × 2000 × $\log_2 2$ = 4000 bps

Example 4

Consider the same noiseless channel, transmitting a signal with four signal levels.

Solution

Bitrate = 2 × 2000 × $\log_2 4$ = 8000 bps.

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

$$Capacity = Bandwidth \times \log_2(1 + SNR)$$

Where, Bandwidth is the bandwidth of the channel

SNR is the signal-to-noise ratio, and Capacity is the capacity of the channel in bits per second.

The signal-to-noise ratio is the statistical ration of the power of the signal to the power of the noise.

Example 5

Calculate the channel capacity of telephone line using Shannon formula.

Solution

A telephone line has a bandwidth of 3000 Hz (300 Hz to 3300 Hz). The signal-to-nose ration is usually 3162. For this channel capacity is :

$$\begin{aligned}\text{Capacity} &= \text{Bandwidth} \times \log_2(1 + \text{SNR}) \\ &= 3000 \times \log_2(1 + 3162) \\ &= 3000 \times \log_2(3163) \\ &= 3000 \times 11.62 \\ &= 34{,}860 \text{ bps}\end{aligned}$$

That is, the highest bit rate for a telephone line is 34.860 Kbps. If data want to be sent faster than this, the bandwidth of the line can be increased or signal-to-¬noise ratio improved.

4.3 Transmission Mode

Transmission of data across link can be accomplished in either parallel or serial. Data transfer from one device to another device is always either by parallel transmission mode or serial transmission mode. In parallel mode, multiple bits are sent with each clock tick. In serial mode, one bit is sent with each clock tick.

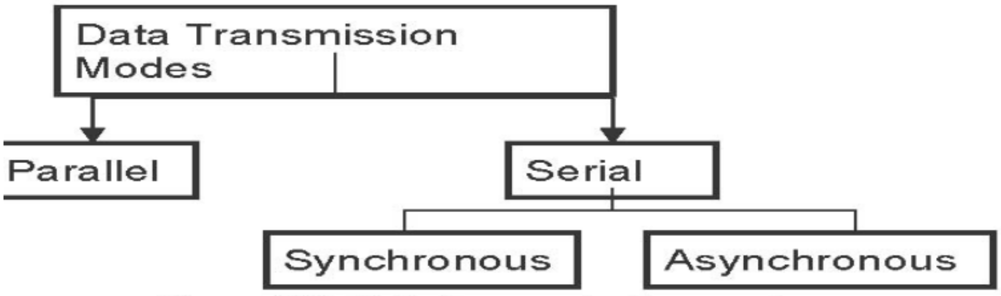


Figure 4.1: Data Transmission modes

4.3.1 Parallel Transmission

A form of signal transmission that sends information 8 or more bits at a time over a cable. Parallel interfaces are used mainly to connect printers, hard drives, and other peripherals to computers. The mechanism for parallel transmission is a conceptually simple one. Use n wires to send n bits at a time. For example, to send 8 bits at a time use 8 data wires. Typically, the eight wires are bundled in a cable with connector at each end. That way each bit has its own wire, and all n bits of one group can be transmitted with each clock tick from one device to another device.

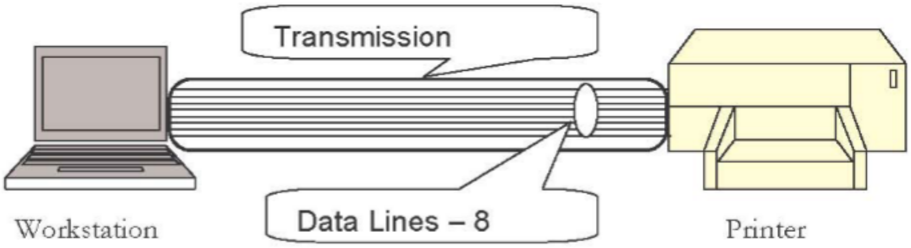


Figure 4.2: Parallel Transmission

4.3.2 Serial Transmission

A form of signal transmission that sends information one bit at a time over a single data channel or data link. Serial interfaces are generally used to connect data communications equipment (DCE) such as modems to data terminal equipment (DTE) such as computers and terminals and for connecting a DCE to a DTE. RS-232 is the most commonly used serial interface in ordinary network communication, which supports transmission over a range of 0 to 20 Kbps at distances of up to 50 feet (15.24 meters).

In serial transmission one bit follows another, so we need only one communication channel or wire rather than n to transmit data between two communicating devices. Serial transmission is possible in one of two ways: synchronous and asynchronous.

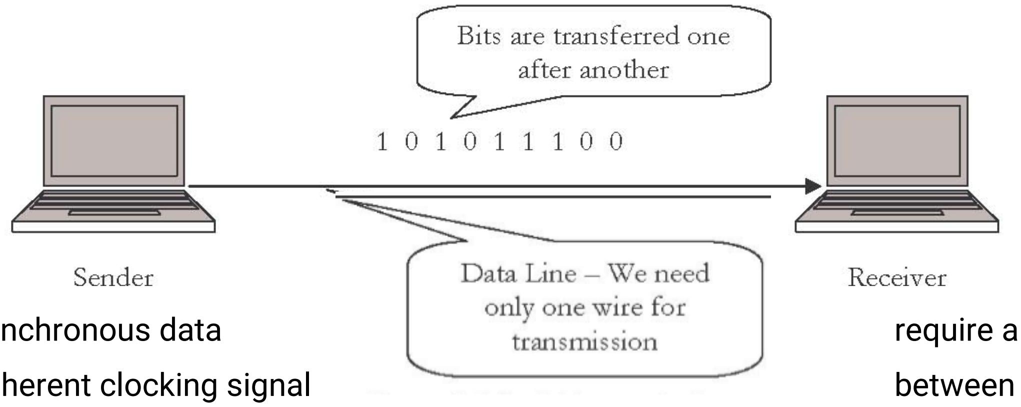
4.3.2.1 Serial Data Formats

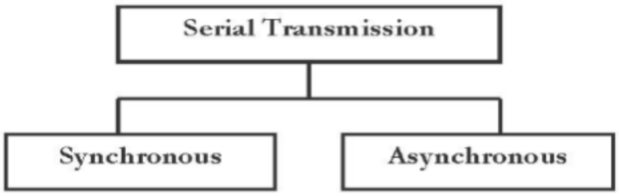
Whether data are sent as bits or symbols, it is transmitted serially in one of two forms, synchronous or asynchronous. Synchronous means serial data that requires a synchronizing clock signal between sender and receiver. And, Asynchronous means serial data that does not require a synchronizing clock or signal between sender and receiver.

Figure 4.3: Serial Transmission

Figure 4.4: Types of Serial transmission

Synchronous Data


Bits are transferred one after another
1 0 1 0 1 1 1 0 0
Sender
Receiver
Data Line – We need only one wire for transmission



Synchronous data require a coherent clocking signal between transmitter and receiver, called a data clock, to synchronize the interpretation of the data sent and received. The data clock is extracted from the serial data stream at the receiver by special circuits called clock recovery circuits.

Once the clock is recovered at the receiving end, bit and character synchronization can be established. Bit synchronization requires that the high and low condition of the binary data sent matches that received and is not in an inverted state.

Character synchronization implies that the beginning and end of a character word is established so that these characters can be decoded and defined. Overall the clock recovered from the message stream itself maintains synchronization. Figure 4.5(a) shows how a synchronous binary transmission would send the ASCII character E (hex 45 or 1000101). The least significant bit (LSB) is transmitted first, followed by the remaining bits of the character. There are no additional bits added to the transmission.

With synchronous transmission, a block of bits is transmitted in a steady stream without start and stop codes. The block may be many bits in length. To prevent timing drift between transmitter and receiver, their clocks must somehow be synchronized. One possibility is to provide a separate clock line between transmitter and receiver. One side (transmitter or receiver) pulses the line regularly with one short pulse per bit-time. The other side uses these regular pulses as a clock. This technique works well over short distances, but over longer distances the clock pulses are subject to the same impairments as the data signal, and timing errors can occur. The other alternative is to embed the clocking information in the data signal; for digital signals, this can be accomplished with Manchester or Differential Manchester encoding. For analog signals, a number of techniques can be used; for example, the carrier frequency itself can be used to synchronize the receiver based on the phase of the carrier.

With synchronous transmission, there is another level of synchronization required to allow the receiver to determine the beginning and end of a block of data; to achieve this, each block begins with a preamble bit pattern and generally ends with a postamble bit pattern.

Figure 4.6 shows, in general terms. a typical frame format for synchronous transmission. Typically, the frame starts with a preamble called a flag, which is eight bit-long. The same flag is used as a postamble. The receiver looks for the occurrence of the flag pattern to signal the start of a frame. This is followed by some number of control fields, then a data field (variable length for most protocols), more control fields, and finally the flag is repeated.

For sizable blocks of data, synchronous transmission is far more efficient than asynchronous. Asynchronous transmission requires 20 percent or more overheads. The control information, preamble, and postamble in synchronous transmission are typically less than 100 bits. For example, one of the more common schemes, HDLC, contains 48 bits of control, preamble, and postamble. Thus, for a 1000-character block of data, each frame consists of 48 bits of overhead and 1000 X 8 = 8,000 bits of data, for a percentage overhead of only 0.6%.

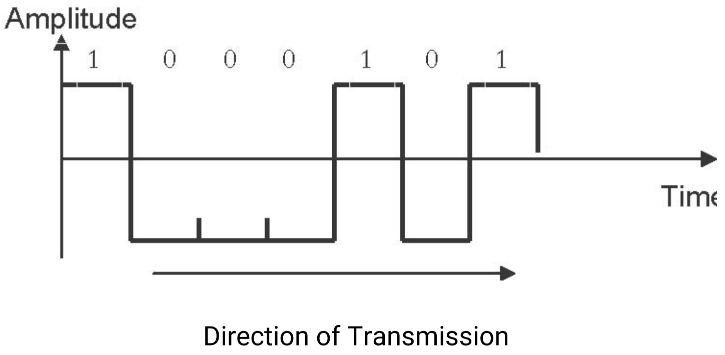


Figure 4.5(a): Synchronous E (hex 45 or 1000101).

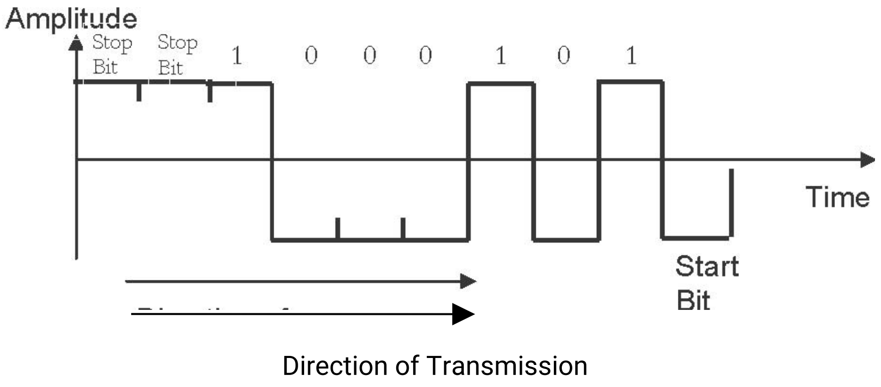


Figure 4.5(6): Asynchronous E (hex 45 or 1000101).

Flag Field	Control field	Data field (payload)	Control field	Flag Field

Figure 4.6: Frame format for synchronous transmission

Asynchronous Data

Asynchronous data formats incorporate the use of framing bits to establish the beginning (start bit) and ending (stop bit) of a data character word as shown in Figure 4.5 (b). A clocking signal is not recovered from the data stream, although the internal clocks of the transmitter and receiver must be the same frequency for data to be correctly received. To understand the format of an asynchronous character, it is first necessary to be aware of the state of the transmission line when it is idle and no data is being sent. The idle condition results from the transmission line being held at a logic 1, high state, or mark condition. The receiver responds to a change in the state of the line as an indication that data has been sent to it. This change of state is indicated by the line going low or logic 0, caused by the transmission of a start bit at the beginning of the character transmission as shown in Figure 4.5 (b). Data bits representing the code of the character being sent follow next ending with one or two stop bits. The stop bits actually specify the minimum time the line must return to logic 1 condition before the receiver can detect the next start bit of the next character.

Asynchronous transmission is simple and cheap but requires an overhead of two to three bits per character. For example, for an 8-bit code, using a 1-bit-long stop bit, two out of every ten bits convey no information but are there merely for synchronization; thus the overhead is 20%. Of course, sending larger blocks of bits between the start and stop bits could reduce the percentage overhead.

Table 4.1: Comparison of serial and parallel transmission

S/No	Parameter	Parallel transmission	Serial transmission
1	Number of wire required to transmit N bits	N wire	1 wire
2	Number of bits transmitted	N bits	1 bit

	simultaneously		
3	Speed of data transfer	Fast	Slow
4	Cost	Higher due to more number of conductor	Low, since only one wire is used
5	Application	Short distance communication such as computer to printer communication	Long distance computer to computer communication

4.3.3 Transmission Efficiency

Notice that the synchronous data uses just the seven bits required for the E character’s code while the asynchronous stream needs 10 bits (one start, seven data, and two stop bits). The synchronous stream is more efficient than the asynchronous because it does not require the overhead (framing) bits that the asynchronous stream needs. Efficiency is a mark of performance and is calculated as a ratio of data or information bits sent to total bits sent as shown in Equation 4.1.

Efficiency = (data bits/total bits) * 100………………. Equation 4.1

A more efficient stream of data takes less time to be transmitted simply because there are less bits to be sent. However, the overall efficiency of a transmission relies on more than the efficiency of individual characters within a message. For asynchronous data, the entire message will retain a 70% efficiency because no additional bits or overhead are required to send the data. Bit and character synchronization are built into the framing bits.

Synchronous data, on the other hand, requires a preamble message, which is a set pattern of binary ones and zeros used to facilitate clock recovery, so the data to be bit and character synchronized before data can be correctly received. This adds additional bits to be sent and reduces the overall efficiency of the transmission. Despite this added burden, synchronous transmissions remain more efficient than asynchronous ones.

4.4 Transmission Impairment

Transmission is the act of propagation through the medium and receiving and processing of the signal. Transmission media are not perfect. The imperfections cause impairment in the signal sent through the medium. This means that the signal at the beginning and end of the medium are not the same. What is sent is not what is received.

Signal that is received will be impaired or distorted during transmission. For analog signals, signal quality is reduced. For digital signals, errors are introduced

,1 recognized

as 0 and vice versa. Transmission medium is imperfect and these impairments affect the capacity of the channel. Three types of impairments can occur : attenuation, distortion, and noise.

4.4.1 Types of Impairments

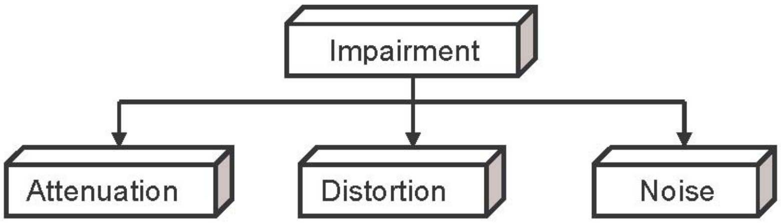


Figure 4.7: Types of Impairment

i) ATTENUATION

Attenuation means loss of energy. Signal strength reduces over time. When a signal travels through a medium, it losses some of energy so that it can overcome the resistance of the medium.

Attenuation is the weakening in strength, of a signal as it passes through the medium. As the signal travels through the transmission medium, some of its power is absorbed, the signal gets weaker, and the receiving equipment has less and less chance of correctly interpreting the data. Thus a wire carrying electrical signal gets warm, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To compensate for this loss, amplifiers are used to amplify the signal.

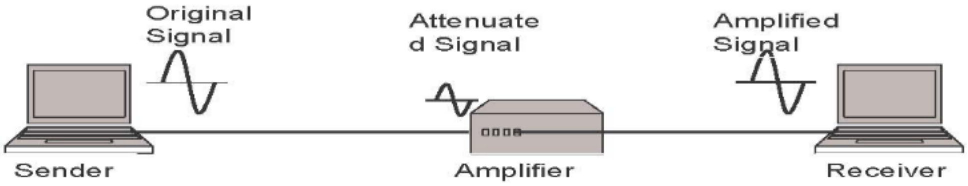


Figure 4.8: Attenuation

The loss of signal strength with long distances when signals travel along cabling. Attenuation values for actual cables are measured in units of decibels (dB) – a standard measurement value used in communication for expressing the ratio of two values of voltage, power, or some other signal-related quantity. For example, a drop of 3 dB corresponds to a decrease in signal strength of 50 percent or 2:1, while a drop of 6 dB corresponds to a decrease of 75 percent or 4:1. Attenuation values for cabling media are expressed in units of decibels per 1000 feet, which express the amount of attenuation in decibels for a standard 1000 -foot length of cabling composed of that media.

Loss of signal strength is expressed in dB decibel. It is a ratio between final and initial power, using logarithms. Loss will be negative dB and gain will be positive dB.

loss in dB = 10 * log{base-10} þ (Pfinal / Pinitial).

Attenuation increases at higher frequencies. Copper cabling has much greater attenuation than fiber-optic cabling; therefore, copper is suitable only for relatively short cable runs. Typical attenuation values for copper category 5 cabling vary with frequency and are shown in the table that follows. Attenuation for lower-grade cable is slightly higher.

Signal Frequency	Attenuation
4 MHz	13 dB/1000 feet
10 MHz	20 dB/1000 feet
20 MHz	28 dB/1000 feet
100 MHz	67 dB/1000 feet

Table 4.2: Attenuation Values for Copper Category 5 Cabling

Attenuation is caused by signal absorption, connector loss, and coupling loss. To minimize attenuation, use high-grade cabling such as enhanced category 5 cabling. Also try to minimize the number of connector devices or couplers, ensuring that these are high-grade components as well. When a signal attenuates a large amount, the receiving device might not be able to detect it or might misinterpret it, therefore causing errors.

ii) DISTORTION

Distortion means that the signal changes its form or shape. The distortion of electrical signals occurs as they pass through metallic conductors. Attenuation Distortion occurs because high frequencies lose power more rapidly than low frequencies during transmission. Thus the received signal is distorted by unequal loss of its component

frequencies. Signals that start at the source as clean, rectangular pulses may be received as rounded pulses with ringing at the rising and falling edges. These effects are properties of transmission through metallic conductors, and become more pronounced as the conductor length increases. To compensate for distortion, signal power must be increased or the transmission rate decreased.

Delay Distortion

It occurs when the method of transmission involves transmission at different frequencies.

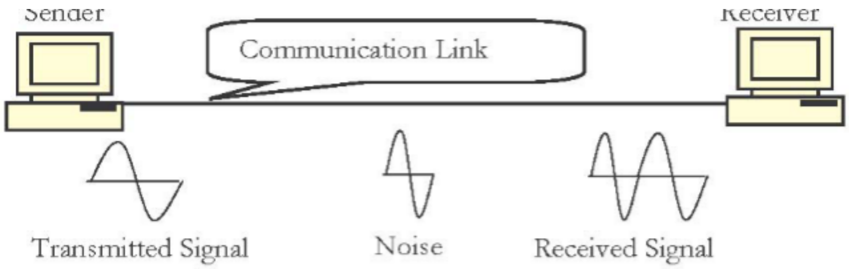
The bits transmitted at one frequency may travel slightly faster than the bits transmitted at another frequency. Delay distortion occurs in guided medium. All frequency components of the signal may not "travel" at the same speed --can cause distortion --e.g., for two consecutive bits, the portion of the signal carrying one bit may overlap with the portion of the signal carrying the neighboring bit. The various frequency components in digital signal arrive at the receiver with varying delays, resulting in delay distortion.

As bit rate increases, some of the frequency components associated with each bit transition are delayed and start to interfere with frequency components associated with a later bit, causing inter-symbol interference, which is a major limitation o maximum bit rate.

iii) NOISE

Noise refers to unintentional signal (voltages) introduced in a line by various phenomenons such as heat or electromagnetic induction created by other sources.

Noise is an undesirable random electrical transmission on network cabling that is generated by networking components such as network interface cards (NICs) or induced in cabling by proximity to electrical equipment that generates electromagnetic interference (EMI). Noise is generated by all electrical and electronic devices, including motors, fluorescent lamps, power lines, and office equipment, and it can interfere with the transmission of signals on a network. The better the signal-to-noise ratio of an electrical transmission system, the greater the throughput of information on the system.



Sender

Receiver

The binary data being transmitted will be altered by noise and result in incorrect data received. Noise and momentary electrical disturbances may cause data to be changed as it passes through a communications channel.

The noisy signals, which cause the data lost or corruption, are classified into different types such as : white noise or thermal noise, induced noise, interference, crosstalk, impulse noise and human errors may corrupt the signal.

White noise is present in all electronic devices and cannot be eliminated by any circuits. It increases with temperature, but it is independent of frequency. That means the white noise covers the whole frequency spectrum and will be picked up by both low and high frequency devices. As bandwidth increases, (thermal) white noise power increases. White noise is also called as thermal noise or additive noise. The amount of noise is directly proportional to the temperature of the medium. White noise usually is not a problem unless it becomes so strong that it obliterates the transmission. Thermal noise (or additive noise) is the random motion of electrons in a wire that created an extra signal not originally sent by the transmitter.

Figure 4.9: Noise

Thermal noise is also called as additive noise. Additive noise is generated internally by components such as resistors and solid-state devices used to implement the communication system. Thermal noise is the most common impairment in a wireless communication system. There are three general sources:

1) The noise that enters the antenna with the signal, aptly called antenna noise,

2) the noise generated due to ohmic absorption in the various passive hardware components, and

3) noise produced in amplifiers through thermal action within semiconductors.

Induced noise comes from sources such as motors and appliances with coils. These devices act as a sending antenna and the transmission medium acts as a receiving antenna.

What can we do to minimize the white noise?

The medium should be kept as cool as possible

Impulse noise is a spike (a signal with high energy in a very short period of time) that comes from power lines, lightning, and so on. Impulse Noise consisting of random occurrences of energy spikes having random amplitude and spectral content. Impulse noise in a data channel can be a definitive cause of data transmission errors.

Interference is caused by picking up the unwanted electromagnetic signals nearby such as crosstalk due to adjacent cables transmitting electronic signals or lightning causing power surge.

Crosstalk is the undesired effect of one circuit (or channel) on another circuit (or channel). It occurs when one line picks up some of the signal traveling down another line. Crosstalk effect can be experienced during telephone conversations when one can hear other conversations in the background. Crosstalk is a form of interference in which signals in one cable induce electromagnetic interference (EMI) in an adjacent cable. The twisting in twisted-pair cabling reduces the amount of crosstalk that occurs, and crosstalk can be further reduced by shielding cables or physically separating them. Crosstalk is a feature of copper cables only—fiber-optic cables do not experience crosstalk

The ability of a cable to reject crosstalk in Ethernet networks is usually measured using a scale called near-end crosstalk (NEXT). NEXT is expressed in decibels (dB), and the higher the NEXT rating of a cable, the greater its ability to reject crosstalk. A more complex scale called Power Sum NEXT (PS NEXT) is used to quantify crosstalk in high-speed Asynchronous Transfer Mode (ATM) and Gigabit Ethernet networks.

Human Error

Noise sometimes is caused by human being such as plugging or unplugging the signal cables, or power on/off the related communications equipment.

The effects of noise may be minimized by increasing the power in the transmitted signal. However, equipment and other practical constraints limit the power level in the transmitted signal. Another basic limitation is the available channel bandwidth. A bandwidth constraint is usually due to the physical limitations of the medium and the electronic components used to implement the transmitter and the receiver.

These two limitations result in constraining the amount of data that can be transmitted reliably over any communications channel. Shannon's basic results relate the channel capacity to the available transmitted power and channel bandwidth.

Signal to noise ratio to quantify noise

Signal-to-noise ratio (S/N) is a parameter used to quantify how much noise there is in a signal. A high SNR means a high power signal relative to noise level, resulting in a good-quality signal. SNR is represented in decibel (db).

$$S/N = 10 \text{ Log10 } (S/N)$$

Where S = average signal power

N = noise power

Bit Error Rate

The BER (Bit Error Rate) is the probability of a signal bit being corrupted in a define time interval. BER of 10^{-5} means on average 1 bit in 10^{-5} will be corrupted.

Note that, a BER of 10^{-5} over voice-graded line is typical and BER of less than 10^{-6} over digital communication is common.

A Bit Error Rate (BER) is a significant measure of system performance in terms of noise. A BER of 10^{-6}, for example, means that one bit of every million may be destroyed during transmission.

Several factors affect the BER:

- Bandwidth
- S/N (Signal-to-noise ratio)
- Transmission medium
- Transmission distance
- Environment
- Performance of transmitter and receiver

4.5 Communication Channel

A communications channel is a pathway over which information can be conveyed. i.e a channel may be defined as a path between a transmitter and receiver . This path may be logical or physical in nature. It may also be hard wired or wireless. It may be defined by a physical wire that connects communicating devices, or by a radio, laser, or other radiated energy source that has no obvious physical presence.

The communication channel provides the connection between the transmitter and the receiver. The physical channel may be a pair of wires that carry the electrical signal, or an optical fiber that carries the information on a modulated light beam, or an underwater ocean channel in which the information is transmitted acoustically, or free space over which the information-bearing signal is radiated by use of an antenna. Other media that can be characterized as communication channels are data storage media, such as magnetic tape, magnetic disks, and optical disks.

Information sent through a communications channel has a source from which the information originates, and a destination to which the information is delivered. Although information originates from a single source, there may be more than one destination, depending upon how many receive stations are linked to the channel and how much energy the transmitted signal possesses. In a digital communications channel, the information is represented by individual data bits, which may be encapsulated into multibit message units. A byte, which consists of eight bits, is an example of a message unit that may be conveyed through a digital communications channel. A collection of bytes may itself be grouped into a frame or other higher-level message unit. Such multiple levels of encapsulation facilitate the handling of messages in a complex data communications network

In some cases, the information may not be reproduced or the information may not reach the receiver at all. Such phenomena can be understood from the following channel characteristics issues :

4.5.1 Channel characteristics

1. Channel Noise

It is a slight background interference present on the channel or unwanted electrical or electromagnetic energy that carries no data or information on but interfaces with the information or data. Hence, noise degrades the quality of information and data by affecting files and communicating of all types including text, programmers, images audio and telemetry. Information and data may be treated as signals in either electrical form. This may be considered as the main source of transmission errors.

The noise may e classified as external or internal noise based upon the sources. External noise is generally picked up from electrical appliances in the vicinity, from electrical transformers, the atmosphere, on even from outer space. Normally this noise does not seriously hamper the performance. However there are a number of electrical appliances or heavy current machines in use, external noise can affect communications. It also impacts communication during severe thunderstorms.

The external noise is generated in inverse proportion to the frequency and in direct proportion to the wavelength and therefore has a remarkable impact on wireless systems than on hard-wired systems. The noise generated because of electricity or atmosphere disturbances is 300khz that is lower than the high frequency range of 300MHz and therefore may have more interference with the signal or information.

Noise generated inside channels or receivers is known as internal noise. Internal noise is less dependent on frequency but has a significant effect at higher frequencies because less dependent on frequencies but has a significant effect at a higher frequencies because external noise has less effect at these frequencies. Minimizing the signal bandwidth may contain noise but this will limit the maximum speed of the data that can be delivered.

2. Channel Bandwidth

Channel bandwidth may be defined as the size of the range of frequencies that can be transmitted through a channel. In order words, it is the volume of information per unit time that a computer, person or transmission medium can handle. It is measured in Hertz Bandwidth is expressed as data speed in bits per second in digital systems and as the difference between highest frequency to lowest frequency in analog system. Bandwidth determines how fast data flows on a given transmission path.

3. Channel Capacity

It is the amount of information per unit time handled by either a link or a node

(system element). The messages transmitted may be either similar or different. It is usually measured in bits per second.

4 **Transmission Time**

This is the time required transmitting a message through the channel. It is the size of message in bits divided by the data rate in bits per second of the channel over which the transmission takes place. It is also given as the packet length divided by the channel capacity.

5. **Propagation Time**

This is the amount of time needed for information to propagate from source to destination through the channel. It is the distance divided by the signal propagation speed. Channel latency depends on media characteristics, signal propagation speed, and transmission distance.

4.6 Channel Capacity

Channel can be defined as a single path provided by a transmission medium via either (a) physical separation, such as by multipair cable or (b) electrical separation, such as by frequency-or time-division multiplexing.

Channel capacity Definition: The maximum bit rate that can be handled by a channel. Channel capacity is also defined as maximum number of television channels that a cable system can carry simultaneously.

Signal-to-Noise Ratio (S/N Ratio) is a very important parameter in assessing the channel capacity or throughput of a data channel. From Shannon's Law, the maximum data rate (bit rate), which a channel can possibly support, is given by the product of the line bandwidth and the signal-to-noise ratio of the channel. Channel capacity, shown often as "C" in communication formulas, is the amount of discrete information bits that a defined area or segment in a communications medium can hold. Thus, a telephone wire may be considered a channel in this sense.

Shannon‘s Law

The maximum data rate of a noisy channel whose bandwidth W in Hz, and whose signal-to-noise ratio is S/N, is given by

$$C= W \text{ Log2}(1 + S/N)$$

Where W = Bandwidth in Hz

S = Average signal power in watts

N = Random noise power in watts

C = Maximum data rate possible

Example

Calculate maximum data rate for telephone line, which having 30 dB signal-to-noise ratio.

Solution

Bandwidth (W) of telephone line = 3300 – 300 Hz = 3000 Hz.
S/N= 39 dB = 1000
C = 3000 × Log2 (1 + 1000)

C = 3000 × Log2 (1001) C = 29,897 bps

4.8

5. MODULATION/DEMODULATION

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

The process of changing some characteristic (e.g. amplitude, frequency or phase)

≈ 30 Kbps

of a carrier wave in accordance with the intensity of the signal is known as **modulation**.

5.1 Types of Modulation

Signal modulation can be divided into two broad categories: Analog and Digital modulation. The aim of **digital modulation** is to transfer a digital bit stream over an analog band pass channel

The aim of **analog modulation** is to transfer an analog baseband (or low pass) signal, for example an audio signal or *TV* signal, over an analog band pass channel, for example a limited radio frequency band or a cable *TV* network channel.

5.2 Analog Modulation methods

Analog Modulation can be accomplished in three ways:

Amplitude, Frequency, and Phase modulation

1. Amplitude Modulation (AM)

In AM transmission, the carrier signal is modulated so that its amplitude varies with the changing amplitudes of the modulating signal. The frequency and phase of the carrier remain the same; only the amplitude changes to follow variations in the information. The modulating signal becomes the envelope of the carrier.

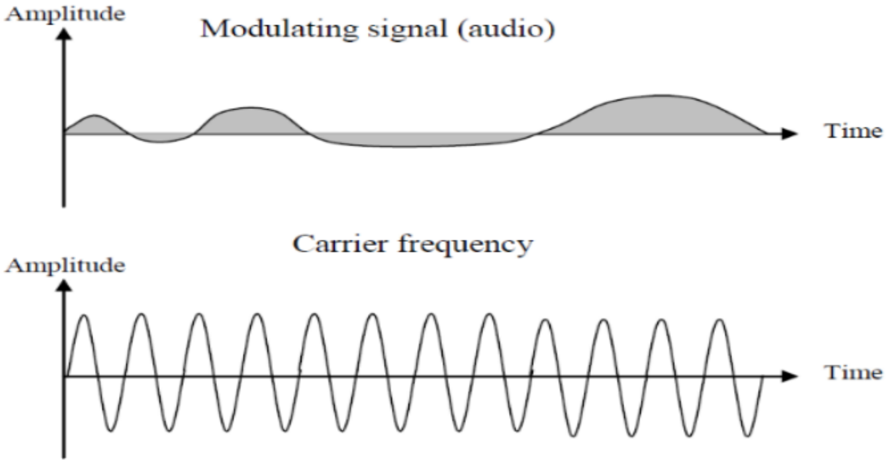


Figure 5.1: Modulating signal and Carrier frequency

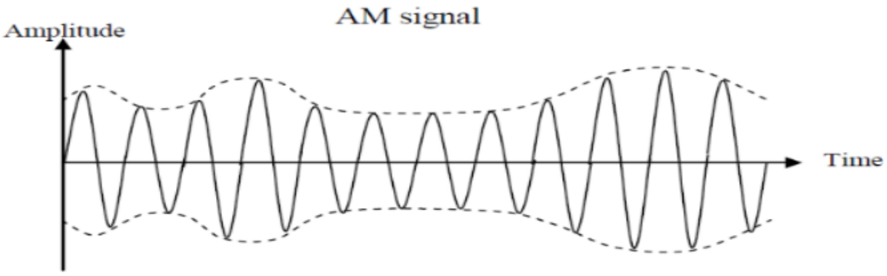


Figure 5.2: AM Signal

AM stations are allowed carrier frequencies anywhere between 530 and 1700 kHz (1.7 MHz). However, each station's carrier frequency must be separated from those on either side by at least 10 kHz (one AM bandwidth) to avoid interference.

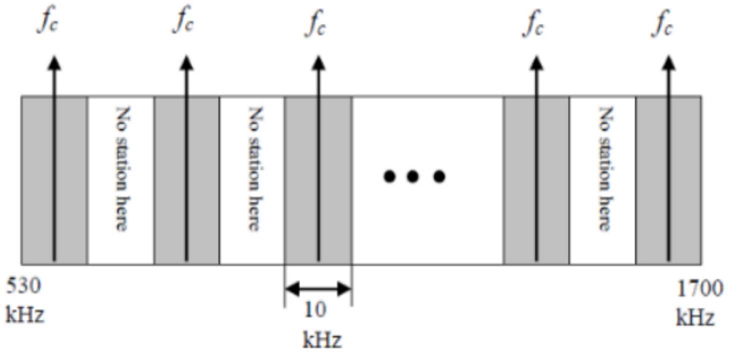


Figure 5.3: Station's carrier frequency

The following points are worth noting in amplitude modulation:

(*i*) The amplitude of the carrier wave changes according to the intensity of the signal.

(*ii*) The amplitude variations of the carrier wave is at the signal frequency.

(*iii*) The frequency of the amplitude modulated wave remains the same *i.e*. carrier frequency *fc*.

2. Frequency Modulation (FM)

In FM transmission, the frequency of the carrier signal is modulated to follow the changing voltage level (amplitude) of the modulating signal. The peak amplitude and phase of the carrier signal remain constant, but as the amplitude of the information signal changes, the frequency of the carrier changes correspondingly. When the frequency of carrier wave is changed in accordance with the intensity of the signal, it is called frequency modulation (FM).

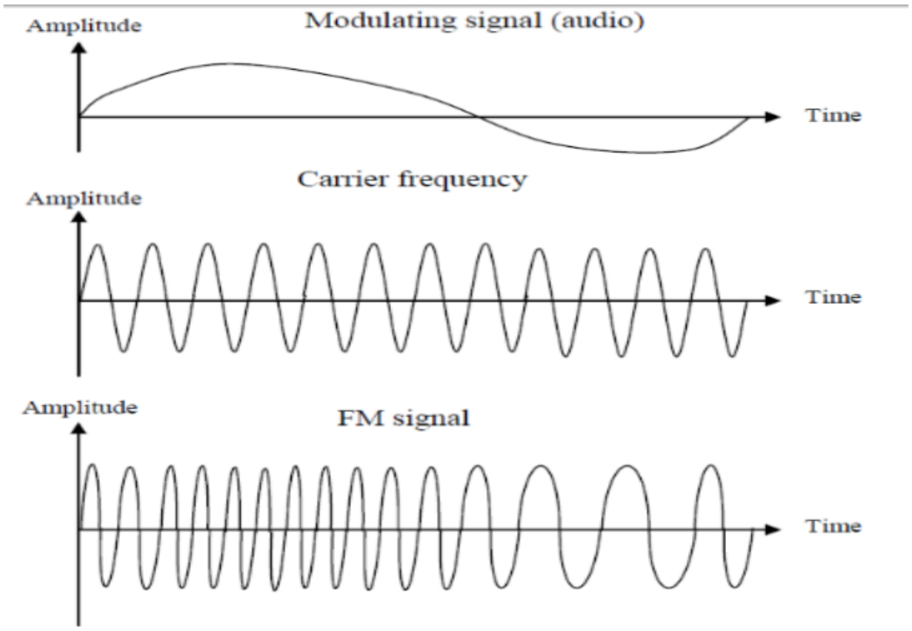
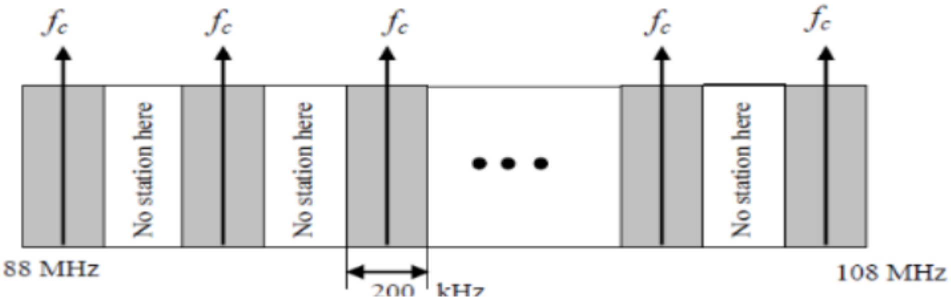


Figure 5.4: Modulating signal, Carrier frequency and FM signal

FM stations are allowed carrier frequencies anywhere between 88 and 108 MHz. However, stations must be separated from by at least 200 kHz to avoid overlapping.



Advantages: The following are the advantages of FM over AM:

(i) It gives noiseless reception. As discussed before, noise is a form of amplitude variations and a FM receiver will reject such signals.

(ii) The operating range is quite large.

(iii) It gives high-fidelity reception.

(iv) The efficiency of transmission is very high.

The comparison of FM and AM is given in the table 4.3 below.

S. N	FM	AM

1.	The amplitude of carrier remains constant with modulation	The amplitude of carrier changes with modulation.
2.	The carrier frequency changes with modulation.	The carrier frequency remains constant with modulation.
3.	The carrier frequency changes according to the strength of the modulating signal.	The carrier amplitude changes according to the strength of the modulating signal.
4.	The value of modulation index (*mf*) can be more than 1.	The value of modulation factor (*m*) cannot be more than 1 for distortionless AM signal.

3. Phase Modulation (PM)

Due to simpler hardware requirements, PM is used in some systems as an alternative to FM. In PM transmission, the phase of the carrier signal is modulated to follow the changing voltage level of the modulating signal. The peak amplitude and frequency of the carrier signal remain constant, but as the amplitude of the information signal changes, the phase of carrier changes correspondingly. The analysis and final result (modulating signal) are similar to those of FM.

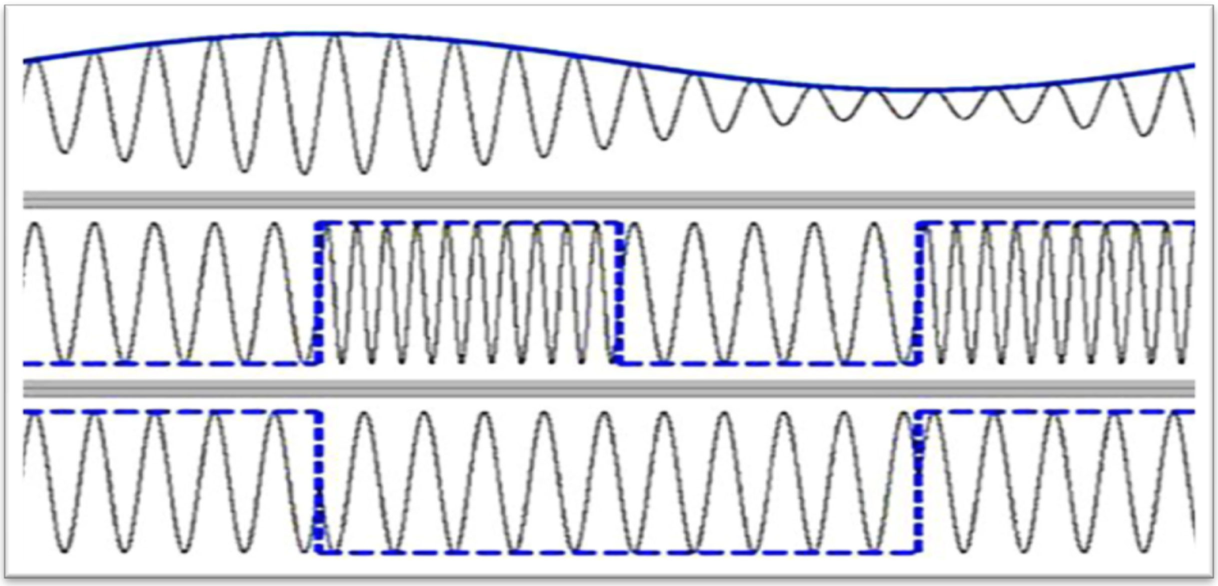


Figure 5.5: Comparison of AM, FM & PM

5.3 Digital Modulation methods

Modulation of binary data or digital-to-analog modulation is the process of changing one of the characteristics of an analog signal based on the information in a digital signal (0s and 1s). When we transmit data from one computer to another across a public access phone line, for example, the original data are digital, but because telephone wires carry analog signal, the data must be converted. The digital data must be modulated on an analog signal that has been manipulated to look like two distinct values corresponding to binary 1 and binary 0.

Two terms used frequently in data communication are bit rate and baud rate. Bit rate is the number of bits transmitted during 1s. Baud rate refers to the number of signal units per second that are required to represent those bits. A signal unit is composed of one or more bits. Bit rate is the number of bits per second. Baud rate is the number of signal units per second. Baud rate is always less than or equal to the bit rate.

An analogy can clarify the concept of baud and bits. In transportation, a baud is analogous to a car, and a bit is analogous to passenger. A car can carry one or more passengers. If 2000 cars go from one location to another, carrying only one passenger, then 2000 passengers are transported. However, if each car carries two passengers, then 4000 passengers are transported. Note that number of cars, not the number of passengers, determines the traffic and therefore, the need for wider highways. Similarly, the number of bauds determines the required bandwidth, not the number of bits.

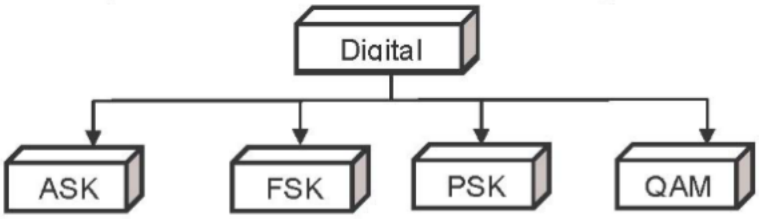


Figure 5.6: Digital Modulation Methods

1. Amplitude Shift Keying (ASK)

In ASK the strength of the carrier signal is varied to represent binary 1 or 0. In ASK both the frequency and phase remain constant while the amplitude changes. Which voltage represents 1 and which represents 0 is left to the system designers. Unfortunately, ASK transmission is highly susceptible to noise interference. The term noise refers to unintentional voltage introduced onto a line by various phenomena such as heat or electromagnetic indication created by other sources. These unintentional voltages combine with the signal to change the amplitude. A 0 can be changed to 1, and a 1 to 0.

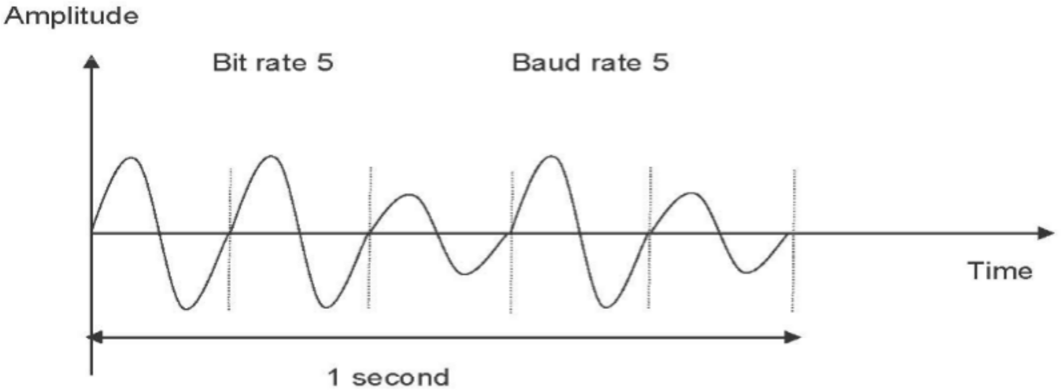
Advantages

Scheme is simple, so it is easy to implement transmitter and receiver with several components.

Low bandwidth requirements

Disadvantage

ASK is heavily affected by noise and interference and can be easily demodulated.



A popular ASK technique is called on-off keying (OOK). In OOK one of the bit values is represented by no voltage. The advantage is reduction in the amount of energy required to transmit information. Unfortunately, ASK transmission is highly susceptible to noise interference.

Figure 5.7: Amplitude Shift Keying

2. Frequency Shift Keying

In FSK the frequency of the carrier signal is varied to represent binary 1 or 0. The frequency of the signal during each bit duration is constant.

In FSK, value of frequency depends on the bit 1 or 0. In FSK both the amplitude and phase remains constant.

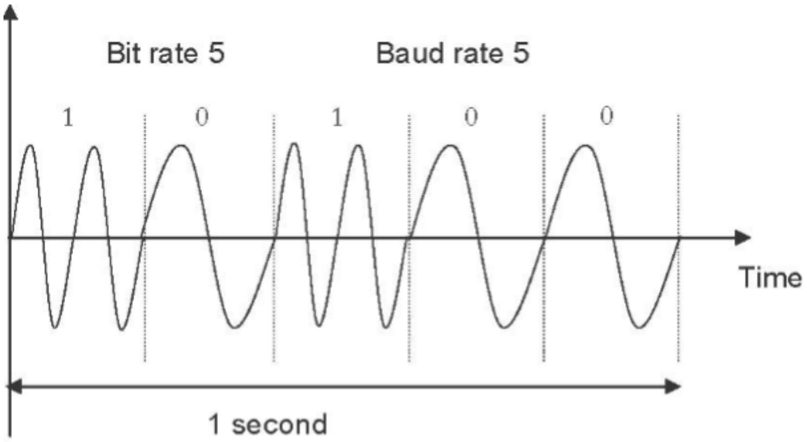


Figure 5.8: Frequency Shift Keying

In FSK, two fixed amplitude carrier signal are used, one for a binary 0 and the other for a binary 1. The different between the two carriers is known as the frequency shift. FSK avoids most of the noise problem of ASK. Because the receiving device is looking for specific frequency changes over a given number of periods, it can ignore voltage spikes. The limiting factors of FSK are the physical capabilities of the carrier.

Advantages

FSK is insensitive to channel fluctuations and not easily effected by noise.

Resilient to signal strength variations

Does not require linear amplifiers in the transmitter

Disadvantage

FSK is a low performance type of digital modulation.

3. Phase Shift Keying (PSK)

In PSK the phase of the carrier is varied to represent binary 1 or 0. For example, if we start with phase of 0 degrees to represent binary 0, then we can change the phase to 180 degrees to send binary 1. The above method is often called 2-PSK, or binary PSK, because two different phases (0 and 180 degrees) are used.

In PSK, both the amplitude and frequency remains constant as the phase changes. PSK is not susceptible to the noise degradation that mostly affects ASK, nor to the bandwidth limitations of FSK.

Advantages

PSK, phase shift keying enables data to be carried on a radio communications signal in a more efficient manner than Frequency Shift

Disadvantage: Implementation is complex and expensive.

Figure 5.9: Phase Shift Keying

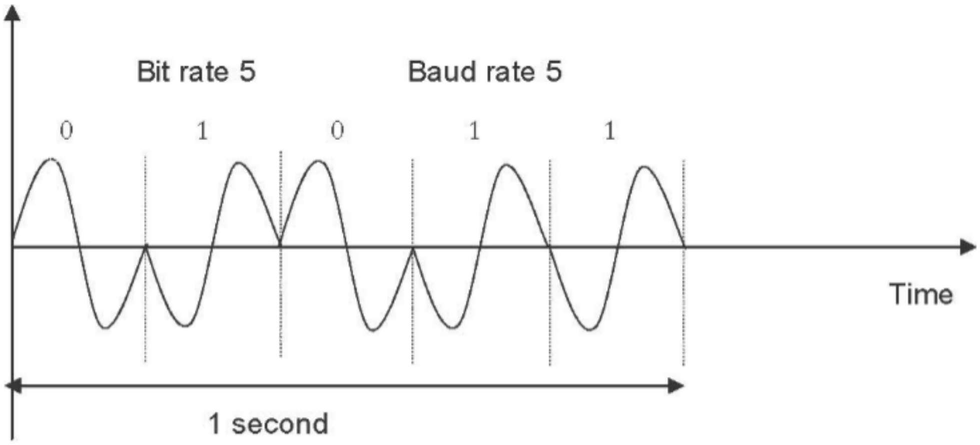
4. QAM (Quadrature Amplitude Modulation)

So far, we have been altering only one of the three characteristics (amplitude, frequency, and phase) of a sine wave at a time, but what if we alter two ?

FSK having bandwidth limitation so combining it with other is practically useless. But why not combine ASK and PSK ? Then we have x variation in amplitude and y variation in phase, giving us x times y possible variations.

QAM means combining ASK and PSK in such a way that we have maximum contrast between each bit, dibit, tribit, and so on.

Possible variations of QAM are numerous. Theoretically, any measurable number of changes in amplitude can be combined with any measurable number of changes in phase. For example : 4-QAM or 8-QAM. In 4-QAM, two-amplitude change and 2 phase shift as shown in figure. In 8-QAM, 2-amplitude change and 4 phase shift. In 8-QAM number of amplitude shifts is less than number of phase shifts. Because amplitude changes are susceptible to noise and require greater shift differences than do phase changes, the number of phase shifts used by a QAM system always larger than the number of amplitude shifts.



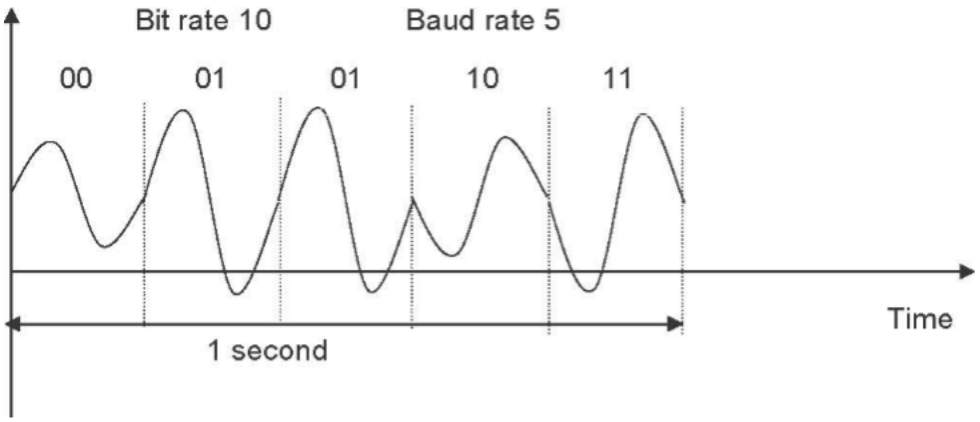


Figure 5.10: 4-QAM (2 amplitudes, 2 phases)

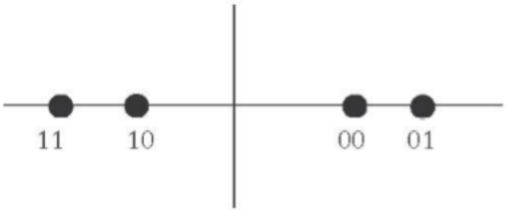


Figure 5.11: Constellation Diagram of 4-QAM

5.4 MODEM

The devices (computers) that generate the digital data (DTE) usually generate a sequence of digital pulses which are not suitable for transmission on a medium. Typically there is another device -a modem (DCE) which prepares this signal for the transmission medium.

Modem stands for modulator/demodulator. A modulator converts a digital signal into an analog signal using Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), Phase Shift Keying (PSK), or Quadrature Amplitude Modulation (QAM) appropriate for telephone lines. A demodulator converts an analog signal into a digital signal.

The two PCs at the end are the DTEs; the modems are the DCEs. The DTE creates a digital signal and relay it to the modem via an interface (like the EIA 232). The modulated signal is received by demodulation function of second modem. It decodes it and then relays the resulting digital signal to the receiving computer via an interface. Generally, modem is any type of data communications equipment (DCE) that enables digital data transmission over the analog Public Switched Telephone Network (PSTN). The term "modem" (which actually stands for "modulator/demodulator") is usually reserved for analog modems, which interface, through a serial transmission connection such as the RS-232 interface, with data terminal equipment (DTE) such as computers. The modem converts the digital signal coming from the computer into an analog signal that can be carried over a Plain Old Telephone Service (POTS) line. The term "digital modem" is sometimes used for ISDN terminal adapters.

Modems were developed in the 1960s by Bell Labs, which developed a series of standards called the Bell Standards. These standards defined modem technologies of up to a 9600-bps transmission speed. But after the breakup of Bell Telephone, the task of developing modem standards was taken over by the International Telegraph and Telephone Consultative Committee (CCITT), which is now called the International Telecommunication Union (ITU). According to ITU specifications, modem standards are classified by a series of specifications known as the V series. The International Telecommunication Union (ITU), which defines standards of up to V.90 (which supports 56-Kbps downloads and 33.6-Kbps uploads).

Modems generally have two interfaces:

- An RS-232 serial transmission interface for connecting to the DTE, usually the computer
- An RJ-11 telephone interface for connecting to the 4-wire PSTN telephone outlet in the local loop connection

Physical types of modem types include the following :

- Internal modems, which are installed as interface cards inside the computer and might use some of the machine's CPU processing power for functions such as encoding and data compression.

- External modems, which are generally more expensive and connect to the serial port on the computer using a DB9 or DB25 connector. External modems are useful when several users need to share a modem.
- PCMCIA modems, which are credit-card-sized modems for laptop computers used by mobile workers.
- Voice/data/fax modems, which can be used for file transfer, sending and receiving faxes, and voice mail using associated software.

Logical types of modem

- Asynchronous and synchronous

Low speed modems are designed to operate asynchronously. Each data frame conforms an asynchronous transmission mechanism. High-speed modems as well as leased-lines modems use synchronous transmission. The two modems use a common time base and operate continuously at substantially that same frequency and phase relationship by circuit that monitor the connection.

- Half duplex and Full duplex

A half-duplex modem must alternately send and receives signals. Half-duplex allows more of the channel bandwidth to be put to use but slows data communications. A full-duplex modem can simultaneously handle two signals using two carriers to transmit and receive data. Each carrier uses half of the bandwidth available to it and its modulation.

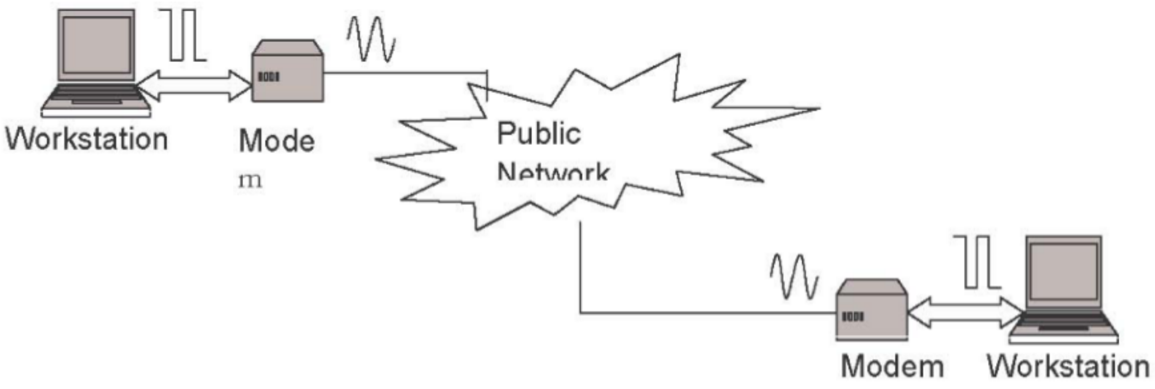


Figure 5.12: Modem

Digital Modem

Any type of modem used for synchronous transmission of data over circuit-switched digital lines. One example of a digital modem is an ISDN terminal adapter. Digital modems are not used for changing analog signals into digital signals because they operate on end-to-end digital services. Instead, they use advanced digital modulation techniques for changing data frames from a network into a format suitable for transmission over a digital line such as an Integrated Services Digital Network (ISDN) line. They are basically data framing devices, rather than signal modulators.

Analog Modem

A modem used for asynchronous transmission of data over Plain Old Telephone Service (POTS) lines. Analog modems are still a popular component for remote communication between users and remote networks. The word “modem” stands for “modulator/demodulator,” which refers to the fact that modems convert digital transmission signals to analog signals and vice versa. For example, in transmission, an analog modem converts the digital signals it receives from the local computer into audible analog signals that can be carried as electrical impulses over POTS to a destination computer or network. To transmit data over a telephone channel, the modem modulates the incoming digital signal to a frequency within the carrying range of analog phone lines (between 300 Hz and 3.3 kHz). To accomplish this, multiplexing of the digital signal from the computer with a carrier signal is performed. The resulting modulated signal is transmitted into the local loop and transmitted to the remote station where a similar modem demodulates it into a digital signal suitable for the remote computer.

5.3 Analog to Digital Conversion using modulation

There are 3 modulation techniques: 1. PAM 2. PCM 3. PWM

1. PAM (Pulse Amplitude Modulation)

The first step in A/D encoding is called pulse amplitude modulation (PAM). This technique takes analog information, samples it, and generates a series of pulses based on the results of sampling. The term sampling means measuring the amplitude of the signal at equal time intervals. In PAM, the original signal is sampled at equal intervals.

PAM has some applications, but it is not used by itself in data communications. However, it is the first step in another very popular encoding method called pulse code modulation (PCM).

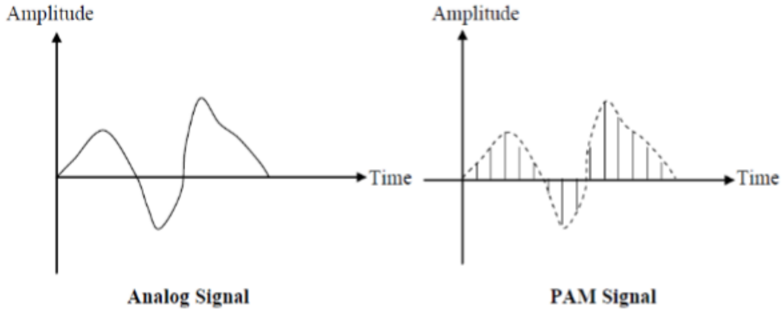


Figure 5.13: PAM

2. PCM (Pulse Code Modulation)

PCM is a general scheme for transmitting analog data in a digital and binary way, independent of the complexity of the analog waveform. With PCM all forms of analog data like video, voice, music and telemetry can be transferred.

3. PWM (Pulse Width Modulation)

Pulse Width Modulation refers to a method of carrying information on a train of pulses, the information being encoded in the width of the pulses. In applications to motion control, it is not exactly information we are encoding, but a method of controlling power in motors without (significant) loss.

There are several schemes to accomplish this technique. One is to switch voltage on and off, and let the current recirculate through diodes when the transistors have switched off.

In battery systems PWM is the most effective way to achieve a constant voltage for battery charging by switching the system controller's power devices on and off.

6. MULTIPLEXING

Multiplexing (or *muxing*) is a way of sending multiple signals or streams of information over a communications link at the same time in the form of a single, complex signal; the receiver recovers the separate signals, a process called *demultiplexing* (or *demuxing*).

In telecommunications and computer networks, multiplexing (sometimes contracted to muxing) is a method by which multiple analog message signals or digital data streams are combined into one signal over a shared medium. The aim is to share an expensive resource. For example, in telecommunications, several telephone calls may be carried using one wire. Multiplexing originated in telegraphy in the 1870s, and is now widely applied in communications. In telephony, George Owen Squier is credited with the development of telephone carrier multiplexing in 1910.

The multiplexed signal is transmitted over a communication channel, such as a cable. The multiplexing divides the capacity of the communication channel into several logical channels, one for each message signal or data stream to be transferred. A reverse process, known as demultiplexing, extracts the original channels on the receiver end. A device that performs the multiplexing is called a multiplexer (MUX), and a device that performs the reverse process is called a demultiplexer (DEMUX or DMX).

Networks use multiplexing for two reasons:

- To make it possible for any network device to talk to any other network device without having to dedicate a connection for each pair. This requires shared media;
- To make a scarce or expensive resource stretch further -- e.g., to send many signals down each cable or fiber strand running between major metropolitan areas, or across one satellite uplink.

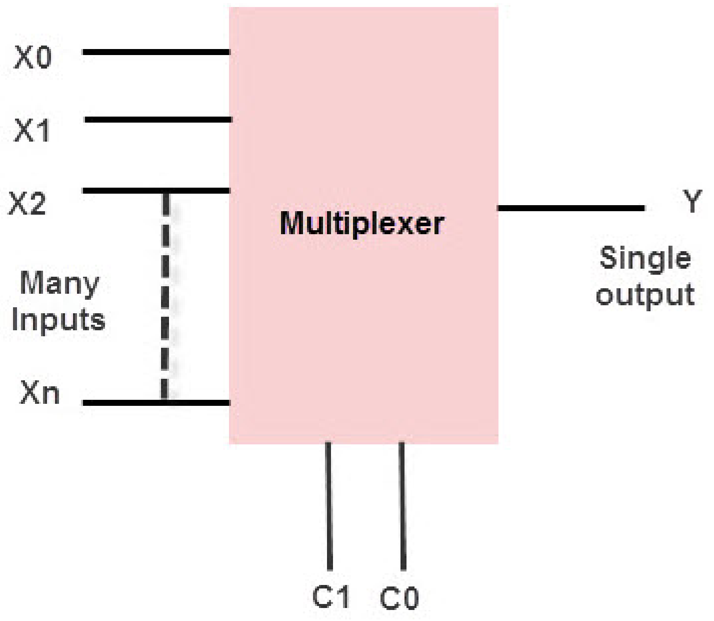


Figure 6.1: Multiplexing

6.1 Types of Multiplexing

1. Frequency Division Multiplexing (FDM)

Frequency Division Multiplexing is a technique which uses various frequencies to combine many streams of data for sending signals over a medium for communication purpose. It carries frequency to each data stream and later combines various modulated frequencies to transmission. When the carrier is frequency, FDM is used. FDM is an analog technology. FDM divides the spectrum or carrier bandwidth in logical channels and allocates one user to each channel. Each user can use the channel frequency independently and has exclusive access of it. All channels are divided in such a way that they do not overlap with each other. Channels are separated by guard bands. Guard band is a frequency which is not used by either channel.

One of the most common applications for FDM is traditional radio and television broadcasting from terrestrial, mobile or satellite stations, or cable television. Only one cable reaches a customer's residential area, but the service provider can send multiple television channels or signals simultaneously over that cable to all subscribers without interference. Receivers must tune to the appropriate frequency (channel) to access the desired signal

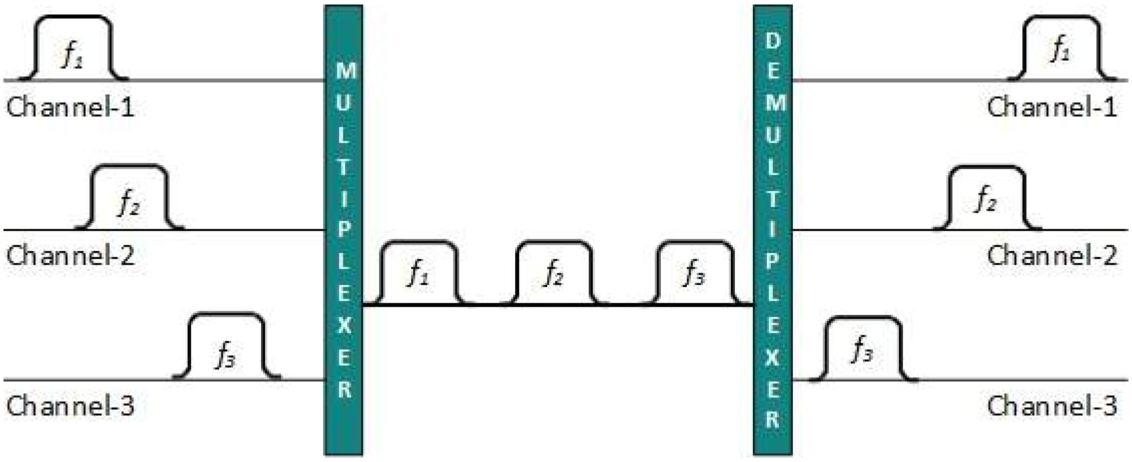


Figure 6.2: FCM

2. Time Division Multiplexing (TDM)

TDM is applied primarily on digital signals but can be applied on analog signals as well. In TDM the shared channel is divided among its user by means of time slot. Each user can transmit data within the provided time slot only. Digital signals are divided in frames, equivalent to time slot i.e. frame of an optimal size which can be transmitted in given time slot.

TDM works in synchronized mode. Both ends, i.e. Multiplexer and De-multiplexer are timely synchronized and both switch to next channel simultaneously.

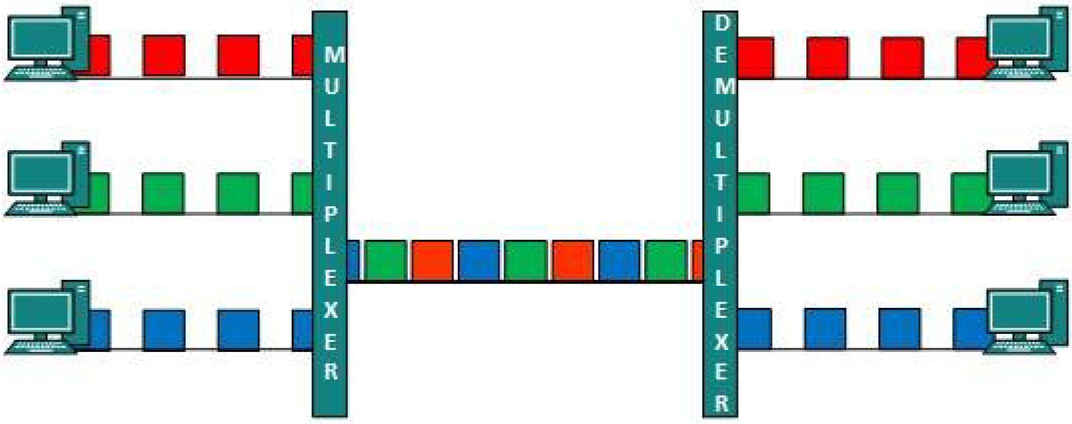


Figure 6.3: TCM

When channel A transmits its frame at one end, the De-multiplexer provides media to channel A on the other end. As soon as the channel A's time slot expires, this side switches to channel B. On the other end, the De-multiplexer works in a synchronized manner and provides media to channel B. Signals from different channels travel the path in interleaved manner.

A drawback to standard TDM is that each sending device has a reserved time slot in each cycle, regardless of whether it is ready to transmit. This can result in empty slots and underutilization of the multiplexed communication channel.

Statistical TDM (STDM) represents an improvement over standard TDM. In STDM, if a sender is not ready to transmit in a cycle, the next sender that is ready can transmit. This reduces the number of wasted slots and increases the utilization of the communication channel. STDM data blocks are known as packets and must contain header information to identify the receiving destination.

Applications that use TDM include long-distance telephone service over a T-1 wire line and the Global System for Mobile Communications (GSM) standard for cellular phones. STDM is used in packet-switching networks for LAN and Internet communications.

3. Wavelength Division Multiplexing (WDM)

Light has different wavelength (colors). It modulates many data streams on light spectrum. This multiplexing is used in optical fiber. In fiber optic mode, multiple optical carrier signals are multiplexed into an optical fiber by using different wavelengths. This is an analog multiplexing technique and is done conceptually in the same manner as FDM but uses light as signals.

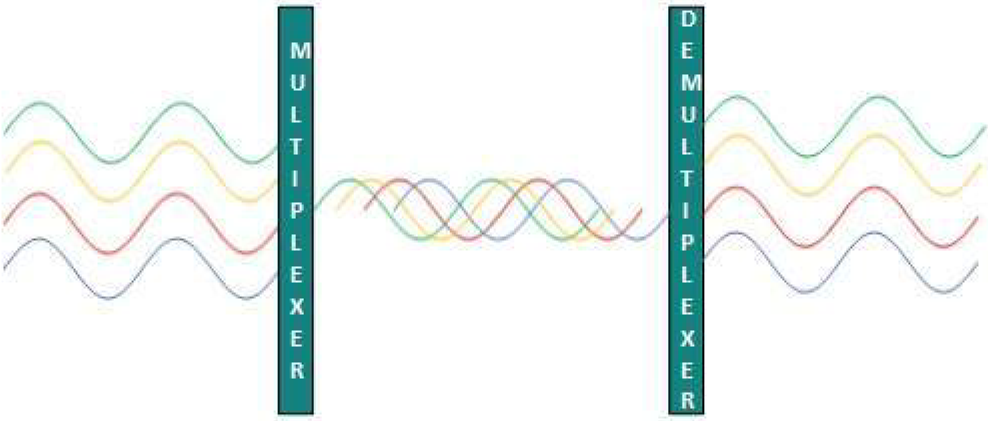


Figure 6.4: WDM

It is FDM optical equivalent. Various signals in WDM are optical signal that will be light and were transmitted through optical fiber.WDM similar to FDM as it mixes many signals of different frequencies into single signal and transfer on one link. Wavelength of wave is reciprocal to its frequency, if wavelength increase then frequency decreases. Several light waves from many sources are united to get light signal which will be transmitted across channel to receiver.

WDM used in Synchronous Optical Network (SONET). It utilizes various optical fiber lines that are multiplexed and demultiplexed. Further, on each wavelength time division multiplexing can be incorporated to accommodate more data signals.

4. Dense Wavelength Division Multiplexer (DWDM)

In Dense Wavelength Division Multiplexing, an optical technology used to expand bandwidth onto fiber optic. Bit rate and protocol are independent and these are the main advantage of DWDM. Dense Wavelength Division Multiplexing (DWDM) operated by combining different signals simultaneously at different wavelengths. On fiber is changed to multiple fibers. By increasing the carrier capacity of fiber from 2.5 Gb/s to 20 Gb/s, an eight OC 48 signals can be multiplexed into single fiber. Single fibers are able to transfer data at a speed upto 400 GB/s due to DWDM.DWDM transfers data or information in IP, SONET, ATM and Ethernet It also carries different type of traffic at a range of speeds on an optical channel.

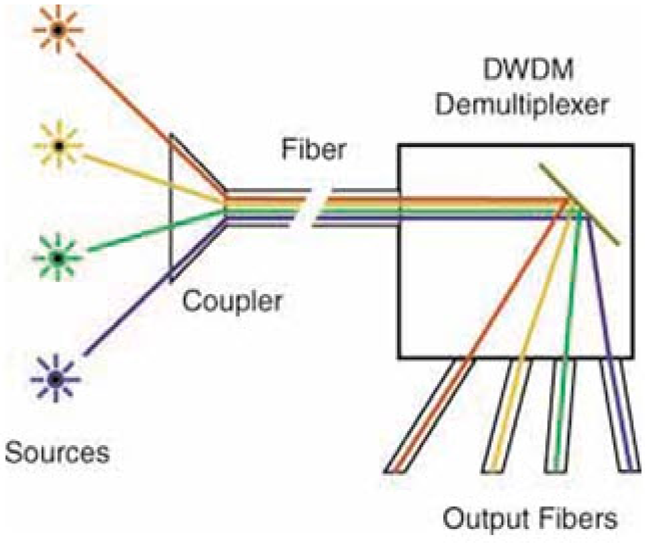


Figure 6.5: DWDM

5. Code Division Multiplexing (CDM)

Code division multiplexing (CDM), Code division multiple access (CDMA) or spread spectrum is a class of techniques where several channels simultaneously share the same frequency spectrum, and this spectral bandwidth is much higher than the bit rate or symbol rate. One form is frequency hopping, another is direct sequence spread spectrum. Multiple data signals can be transmitted over a single frequency by using Code Division Multiplexing. FDM divides the frequency in smaller channels but CDM allows its users to full bandwidth and transmit signals all the time using a unique code. CDM uses orthogonal codes to spread signals.

Each channel transmits its bits as a coded channel-specific sequence of pulses called chips. Each station is assigned with this unique code. Signals travel with these codes independently, inside the whole bandwidth. The receiver knows in advance the chip code signal it has to receive. Number of chips per bit, or chips per symbol, is the spreading factor. This coded transmission typically is accomplished by transmitting a unique time-dependent series of short pulses, which are placed within chip times within the larger bit time. All channels, each with a different code, can be transmitted on.

Advantages over conventional techniques are that variable bandwidth is possible (just as in statistical multiplexing) and it is more secure. CDM is widely used in digital television and radio broadcasting and in 3G mobile cellular networks. Where CDM allows multiple signals from multiple sources, it is called Code-Division Multiple Access (CDMA). A significant application of CDMA is the Global Positioning System (GPS).
the same fiber or radio channel or other medium, and asynchronously demultiplexed.

6. Polarization-division multiplexing

Polarization-division multiplexing uses the polarization of electromagnetic radiation to separate orthogonal channels. It is in practical use in both radio and optical communications, particularly in 100 Gbit/s per channel fiber optic transmission systems.

7. Orbital angular momentum multiplexing

Orbital angular momentum multiplexing is a relatively new and experimental technique for multiplexing multiple channels of signals carried using electromagnetic radiation over a single path. It can potentially be used in addition to other physical multiplexing methods to greatly expand the transmission capacity of such systems.

8. Space-division multiplexing

In wired communication, space-division multiplexing is the use of separate point-to-point electrical conductors for each transmitted channel. Examples include an analogue stereo audio cable, with one pair of wires for the left channel and another for the right channel, and a multi-pair telephone cable, a switched star network such as a telephone access network, a switched Ethernet network, and a mesh network.

In wireless communication, space-division multiplexing is achieved with multiple antenna elements forming a phased array antenna. Examples are multiple-input and multiple-output (MIMO), single-input and multiple-output (SIMO) and multiple-input and single-output (MISO) multiplexing.

6.2 Difference between Mux and Demux

- A Multiplexer is a device used to communicate by means of a medium with combination of multiple signals.
- A DE multiplexer is a process of separating multiplexed signals from transmission line.
- Both Mux and DMux are mixed into single device which has the capability to process outgoing and incoming signals
- An electronic multiplexer is a multiple-input, single-output switch.
- A DE multiplexer as a single-input, multiple-output switch
- MUX allows many signals to share one device.
- Example: one A/D converter or one communication line

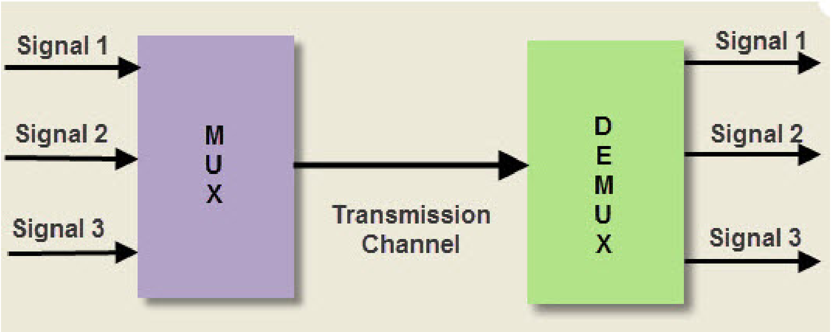


Figure 6.6: Difference between MUX and Demux

6.3 Applications of Multiplexers

A Multiplexer is used in numerous applications like, where multiple data can be transmitted using a single line.

Communication System – A Multiplexer is used in communication systems, which has a transmission system and also a communication network. A Multiplexer is used to increase the efficiency of the communication system by allowing the transmission of data such as audio & video data from different channels via cables and single lines.

Computer Memory – A Multiplexer is used in computer memory to keep up a vast amount of memory in the computers, and also to decrease the number of copper lines necessary to connect the memory to other parts of the computer.

Telephone Network – A multiplexer is used in telephone networks to integrate the multiple audio signals on a single line of transmission.

Transmission from the Computer System of a Satellite:

A Multiplexer is used to transmit the data signals from the computer system of a satellite to the ground system by using a GSM communication.

7. ERROR DETECTION AND CORRECTION

7.1 Introduction

Errors are all around us. We hate them, but we make them all the time, we expect them, and we try to learn to live with them. We say “to err is human," but it is not only humans who err. Tools, instruments, and machines of all kinds also misbehave or err, bite back, break down from time to time. One area where errors are particularly critical is data processing and communications. One bad bit in a computer program can completely corrupt the program. Similarly, the smallest error in a data file can change the meaning of the data in a crucial way. Fortunately, it is possible to detect, and often correct errors in data and data communication.

Every time information is transmitted, on any channel, it may get corrupted by noise. In fact, even when information is stored in a storage device, errors may suddenly occur, because no piece of hardware is absolutely reliable. This also applies to non-computer information. Printed information may fade with time and may deteriorate from high use. Speech sent through the air may deteriorate due to noise, wind, and variations in temperature. Speech, in fact, is a good starting point for understanding the principles of error-detecting and error-correcting codes. Imagine a noisy cocktail party where everybody talks simultaneously, on top of blaring music. We know that even in such a situation it is possible to carry on a conversation, except that more attention than usual is needed.

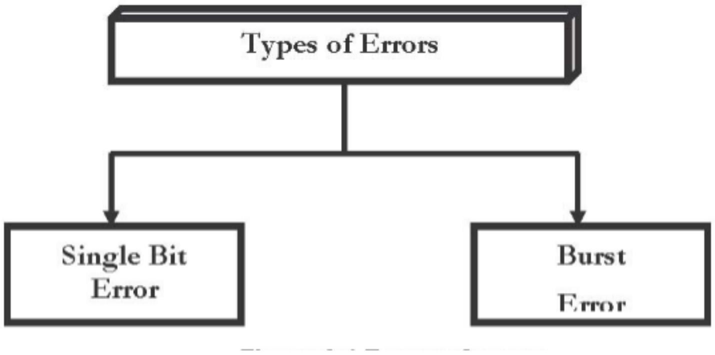
Data processing and transmission systems use a variety of techniques to detect and correct errors that occur, usually for any of the following reasons:

Electrostatic interference from nearby machines or circuits Attenuation of the signal caused by a resistance to current in a cable Distortion due to inductance and capacitance Loss in transmission due to leakages Impulses from static in the atmosphere

Most of the LAN technologies and optical cable networks reduce errors considerably. Wireless networks and WAN links can have high error rates. The occurrence of a data bit error in a serial stream of digital data is an infrequent occurrence. Even less frequent is the experience of numerous errors within the transmission of a single message. If a number of errors occur then it is presumed that either a significant interference occurred affecting the transmission line or that there is a major failure in the communications path. Largely because of the extremely low bit-error rates in data transmissions, most error detection methods and algorithms are designed to address the detection or correction of a single bit error.

7.2 Bit Errors

Bit errors are errors that corrupt bits during transmission, turning a 1 into a 0, or 0 into a 1.These errors are caused by power surges and other interference. Packet errors occur when packets are lost or corrupted. Packet loss can occur during times of network congestion when buffers become full and network devices start discarding packets. Errors and packet loss also occur during network link failures. There are two types of errors namely, single bit error and burst errors



7.3 Types of Errors

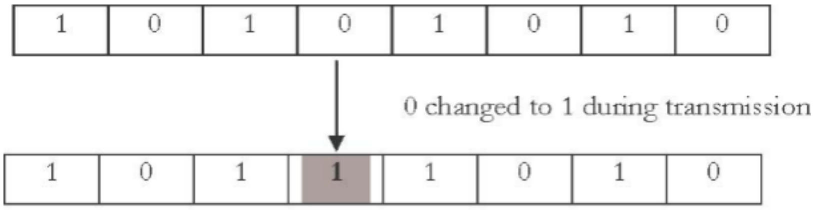
Figure 7.1: Types of Errors

Single-Bit Error

The term single-bit error means that 1 bit of a message is changed from 0 to 1 or from 1 to 0 during transmission.

Burst Error

The term burst error means that 2 or more bits of messages changed from 0 to 1 or from 1 to 0 during transmission. Burst error does not mean that the error occurred in consecutive bits. It might be possible some bits in between not been corrupted. The length of burst is measured from the first changed bit to the last changed/corrupted bit.



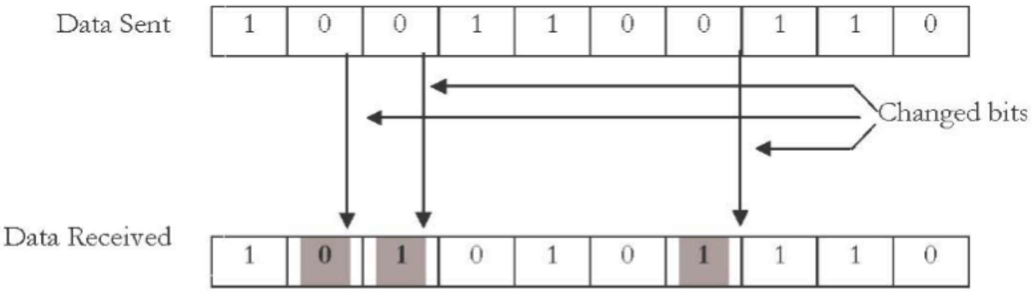


Figure 7.2: Single Bit-error

Figure 7.3: Burst Error of length 6

Length of Burst Error

In practice, bits are sent on a wire as voltages. A binary 0 may, e.g., be represented by any voltage in the range 3-25 volts. A binary 1 may similarly be represented by the voltage range of — 25v to — 3v. Such voltages tend to drop over long lines and have to be amplified periodically. In the telephone network there is an amplifier (a repeater) every 20 miles or so. It looks at every bit received, decides if it is a 0 or a 1 by measuring the voltage, and sends it to the next repeater as a clean, fresh pulse. If the voltage has deteriorated enough in passage, the repeater may make a wrong decision when sensing it, which introduces an error into the transmission. At present, typical transmission lines have error rates of about one in a billion but, under extreme conditions—such as in a lightning storm, or when the electric power suddenly fluctuates—the error rate may suddenly increase, creating a burst of errors.

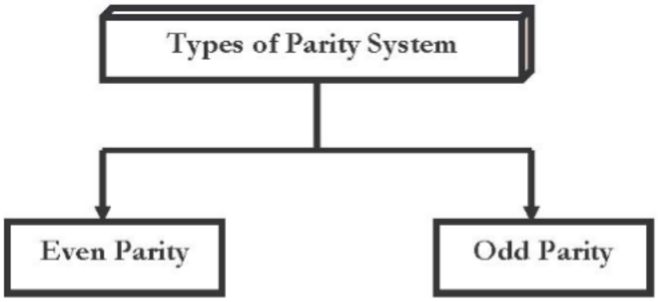
7.4 Data Error Detection Methods

7.4.1 Asynchronous Data Error Detection Methods

Probably the most common and oldest method of error detection is the use of parity. While parity is used in both asynchronous and synchronous data streams, it finds greater use in low-speed asynchronous transmission applications; however, its use is not exclusive to this.

7.4.1.1 Parity Error Detection

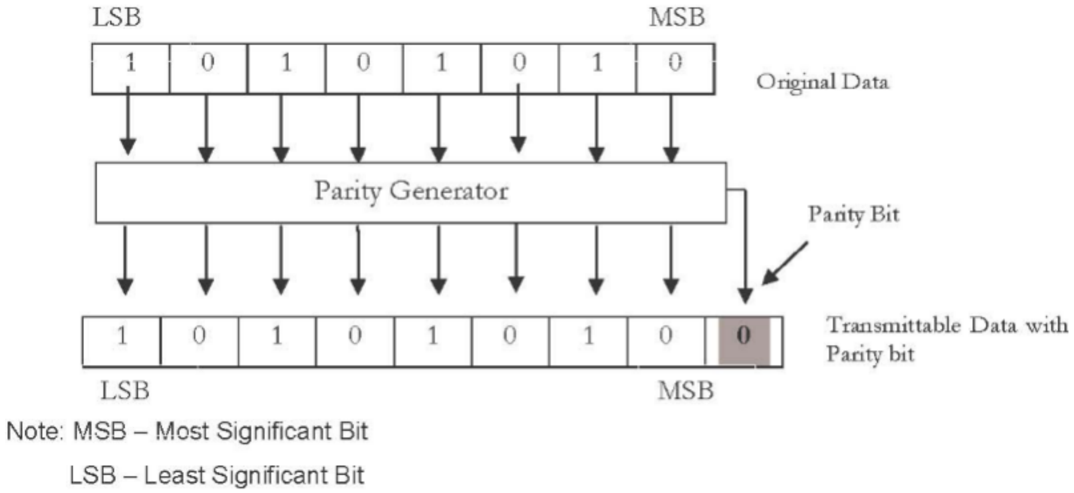
Parity works by adding an additional bit to each character (word) transmitted. The state of this bit is determined by a combination of factors, the first of which is the type of parity system employed. The second factor is the number of logic 1 bits in the data character The two types are even and odd parity.



In an even parity system, the parity bit is set to a low state if the number of logic 1s in the data word is even. If the count is odd, then the parity bit is set high. For an odd parity system, the state of the parity bit is reversed. For an odd count, the bit is set low, and for an even count, it is set high.

Figure 7.4: Types of Parity System

To detect data errors, each character word that is sent has a parity bit computed for it and appended after the last bit of each character is sent as illustrated in Figure 7.5. At the receiving site, parity bits are recalculated for each received character. The parity bits sent with each character are compared to the parity bits the receiver computes. If their states do not match, then an error has occurred. If the states do match, then the character may be error-free.


LSB
MSB
1 0 1 0 1 0 1 0
Original Data
Parity Generator
Parity Bit
1 0 1 0 1 0 1 0 0
Transmittable Data with Parity bit
LSB
MSB
Note: MSB – Most Significant Bit
LSB – Least Significant Bit

LSB

Least Significant Bit

Example 1

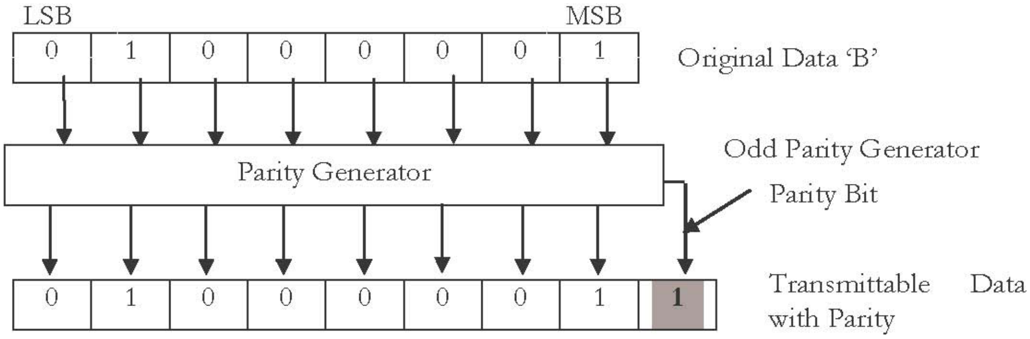
What is the state of the parity bit for both an odd and an even parity system for the extended ASCII character A?

Figure 7.5: Appending parity bit

Solution

The extended ASCII character ‘A’ has a bit pattern of 0100 0001 (41 H). The number of logic 1s in that pattern is two, which is an even count. For an even parity system, the parity bit would be set low so that the total number of 1’s in transmittable data unit including parity bit will become even and for an odd parity system, it would be set high so that the total number of 1’s in transmittable data unit including parity bit will become odd.

B 8	B 7	B 6	B 5	B 4	B 3	B 2	B 1



LSB MSB

Figure 7.6(a): Even Parity for ASCII character A

LSB MSB

Figure 7.6(b): Odd Parity for ASCII character A

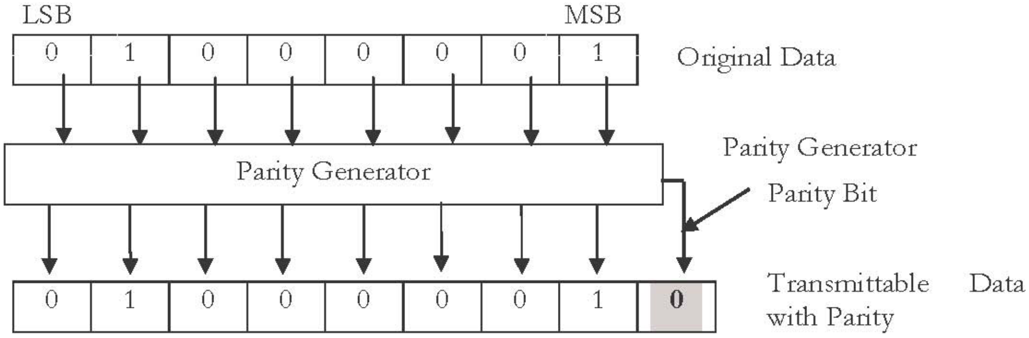
Example 2

The ASCII character A (0100 0001 = 41h) is transmitted with an even-parity bit appended to it. Illustrate how the receiver would detect an error.

Solution

As shown in Figure 7.6(a), the state of the even-parity bit for the ASCII A is low, so the complete data stream for the character sent, starting with the least significant bit (LSB) is: 010000010. Notice there is now nine bits -eight bits for the extended ASCII character A and one for the parity bit. The breakdown of the data stream is:

0	1	0	0	0	0	0	1	0



LSB MSB Parity

Suppose that the LSB becomes corrupted during transmission. The receiver receives the character as: 110000010. When the receiver computes a parity bit for the character data, it results in a high state of the parity. This is compared with the transmitted parity, which is a low state. Since they do not agree, the receiver determines that an error has occurred. Note that the receiver cannot determine which bit is bad, only that one of them is wrong.

A match between transmitted parity and receiver-calculated parity does not guarantee that the data has not been corrupted. Indeed, if an even number of errors occurs in a single character, then the parity for the corrupted data will be the same state as the good data. This does not present a major problem, since the occurrence of two errors in an eight-bit character is excessive and usually indicates a major problem in the system. Such a problem would cause errors to occur in other characters and one of them would eventually be detected.

7.4.2 Synchronous Data Error Methods

Synchronous data are transmitted at higher data rates in as an efficient manner as possible. Start-and-stop framing and parity bits are omitted from the data stream to reduce overhead. Overhead is defined as any bits sent that do not contain actual data information. This includes framing bits, preambles, error-detection characters, or bits, etc. It should be noted that in some synchronous data systems, parity is occasionally employed for error detection. Most high-speed synchronous transmissions, however, do not follow this practice. The reason is that most errors in high-speed transmissions occur in bursts, which could render parity-error detection less effective. These error bursts result from some external interference or other effect on the line that causes several bits to be corrupted at once. Single-bit errors occur less frequently. The duration of noise is normally longer than the duration of one bit, which means that when noise affects data, it affects set of bits. The number of affected bits depends on the data rate and duration of noise. For example, if we are sending data at 1 Mbps, a noise of 1/100 second can affect 10,000 bits. Because of this, and the desire to reduce the overhead in synchronous transmissions, error-detection methods have evolved to detect single and multiple errors within a data stream.

Synchronous error detection works by creating an additional character to be sent with the data stream. At the receive site, the process is duplicated and the two error detection characters are compared similarly to comparing two parity bits. If the characters match then the data received has no errors. If they do not match, an error has occurred and the message has to be retransmitted. Note that one major difference between using error-detection characters versus single-parity bits, is that if the transmitted and received characters match, then the data is good. Using parity, matching parity bits does not guarantee that the character received was good. The computation of error characters is carried on quickly to support the higher data rates of transmission.

7.4.2.1 Cyclic Redundancy Check (CRC)

One of the most frequently used error-detection methods for synchronous data transmissions is cyclic redundancy check (CRC) developed by IBM. This method uses a pseudo-binary-division process to create the error or CRC character, which is appended to the end of the message. The hardware circuitry that generates the CRC character at the transmitter is duplicated at the receiver. This circuitry is incorporated into the transmit¬ and-receive shift registers that send and receive the actual message.

Unlike the parity check which is based on addition, CRC is based on binary division. In CRC, instead of adding bits to achieve a desired parity, a sequence of redundant bits, called the CRC or the CRC remainder, is appended to the end of the message so that resulting data unit becomes exactly divisible by a second, predetermined binary number. At its destination, the incoming data unit is divided by the same number. If at this step there is no remainder, the data unit or message is assumed to be intact and is therefore accepted. A remainder indicates that the data unit has been damaged during transmission and therefore must be rejected.

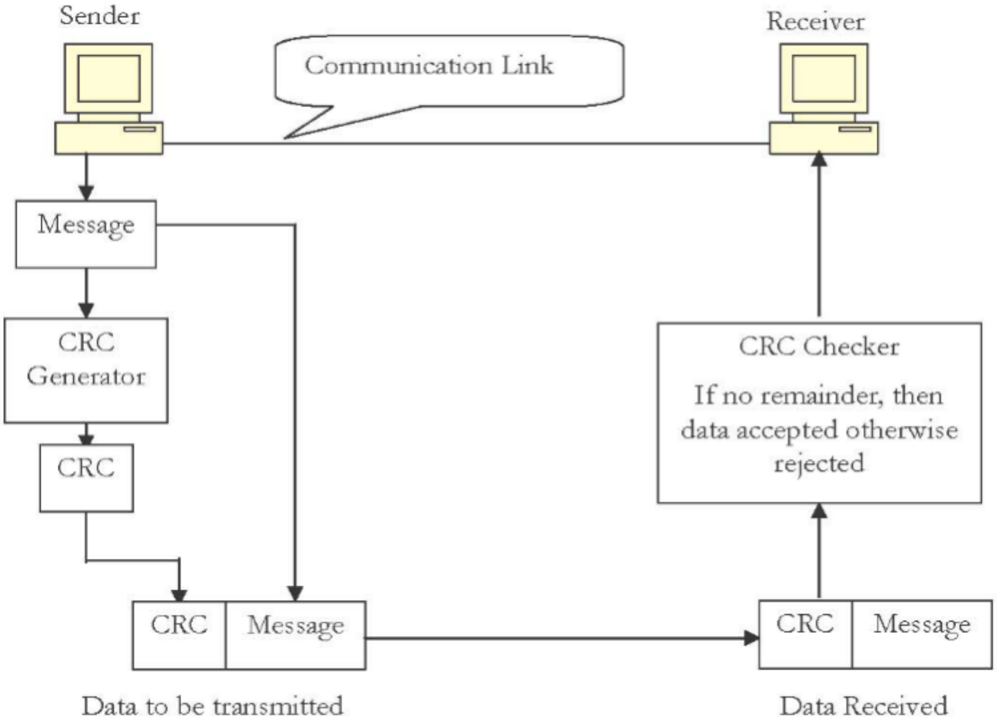


Figure 7.7: CRC generator and checker

Example 3

Compute the CRC-4 character for the following message using a “divisor” constant of 10011 on data unit 1100 0110 1011 01

Solution

Notice that the “divisor” is 5-bits, one more than the number indicated by the CRC type

(CRC-4). We start the process by adding four zeros to the data stream and removing the spaces we have been using for convenience. Next, set up the problem to appear as a division problem: Start the “division” process by exclusive OR the “divisor” with the first five bits of the message:

Now bring “down” one bit so that the result of the exclusive OR process is filled out to the “divisor” size and repeat the process: Continue with the process until all of the bits in the message plus the added four zeros are used

```
1 1 0 0 0 1 1 0 1 0 1 1 0 1 0 0 0 0
1 0 0 1 1
-----------------------------------------------
0 1 0 1 1 1 1 0 1 0 1 1 0 1 0 0 0 0
  1 0 0 1 1
--------------------------------------------
0 0 0 1 0 0 1 0 1 0 1 1 0 1 0 0 0 0
      1 0 0 1 1
-------------------------------------------
      0 0 0 0 1 1 0 1 1 0 1 0 0 0 0
              1 0 0 1 1
--------------------------------------------
              0 1 0 0 0 0 1 0 0 0 0
                1 0 0 1 1
------------------------------------------
             0 0 0 0 1 1 1 0 0 0 0
                      1 0 0 1 1
------------------------------------------
            0 0 0 0 0 1 1 1 1 0 0
                      1 0 0 1 1
-------------------------------------------
                      0 1 1 0 1 0
                        1 0 0 1 1
------------------------------------------
                      0 0 1 0 0 1
```

The CRC is 1001

The CRC character is appended onto the end of the message and transmitted. At the receiver, the process is repeated, except that there are no zeros added to the message. Instead, the CRC character fills up those positions. If the result of the process at the receiver produces zero then no errors occurred. If any bit or combinations of bits are wrong, then the receiver will yield a non-zero result

Example 4
Demonstrate how a receiver detects a good message and a message with several errors in it.

Solution
Repeat steps of EX 3 but this time use CRC character in place of extra zeros:

The changes in the bits brought down are highlighted. Notice how they produce different results from EXAMPLE 3. This eventually results in a CRC of 0000 if everything is correct. It means receiver will accept data unit if CRC checker generates CRC Character Zero at receiver side.

```
1 1 0 0 0 1 1 0 1 0 1 1 0 1 1 0 0 1
1 0 0 1 1
--------------------------------------------------
0 1 0 1 1 1 1 0 1 0 1 1 0 1 1 0 0 1
  1 0 0 1 1
-----------------------------------------------
0 0 0 1 0 0 1 0 1 0 1 1 0 1 1 0 0 1
      1 0 0 1 1
----------------------------------------------
      0 0 0 0 1 1 0 1 1 0 1 1 0 0 1
              1 0 0 1 1
-----------------------------------------------
              0 1 0 0 0 0 1 1 0 0 1
                1 0 0 1 1
---------------------------------------------
             0 0 0 0 1 1 1 1 0 0 1
                     1 0 0 1 1
---------------------------------------------
            0 0 0 0 0 1 1 0 1 0 1
                      1 0 0 1 1
----------------------------------------------
                     0 1 0 0 1 1
                       1 0 0 1 1
---------------------------------------------
                     0 0 0 0 0 0
```

The Message does not have any error

NOTE

Given the small size (CRC-4) of this example, there could easily be error combinations that would produce a zero CRC result. This is the reason that most CRC systems today use either CRC-16, CRC-32 or CRC-64.

```
1
1
0          1
1          -
0
1
0
0
1
1
```

7.4.2.2 CheckSum Error Detection

Another method of error detection uses a process known as checksum to generate an error-detection character. The character results from summing all the bytes of a message together, discarding and carry-over for the addition. Again, the process is repeated at the receiver and the two checksums are compared. A match between receiver checksum and transmitted checksum indicates good data. A mismatch indicates an error has occurred.

This method, like CRC, is capable of detecting single or multiple errors in the message. The major advantage of checksum is that it is simple to implement in either hardware or software. The drawback to checksum is that, unless you use a fairly large checksum (16¬or 32-bit instead of 8-bit), there are several data-bit patterns that could produce the same checksum result, thereby decreasing its effectiveness. It is possible that if enough errors occur in a message that a checksum could be produced that would be the same as a good message. This is why both checksum and CRC error-detection methods do not catch 100% of the errors that could occur, they both come pretty close.

Example 5

What is the checksum value for the extended ASCII message “Help”?

01001000 H e l p Checksum
01100101
01101100
01110000
00110001

Solution

The checksum value is found by adding up the bytes representing the Help Characters:

Example 6

What is the checksum value for the extended ASCII message “Hello”?

Solution

The checksum value is found by adding up the bytes representing the Hello Characters:

01001000 H 01100101 e 01101100 l 01101100 l 01101111 0
00110010 Checksum

7.5 Error Correction

It is important to understand the meaning of the word error in data storage and transmission. When an n-bit data or message is sent and received, the receiver always receives n bits, but some of them may be bad. A bad bit does not disappear, nor does it change into something other than a bit. A bad bit simply changes its value, either from 0 to 1 or from 1 to 0. This makes it relatively easy to correct the bit. The code should tell the receiver which bits are bad, and the receiver can then easily correct the bits by inverting them.

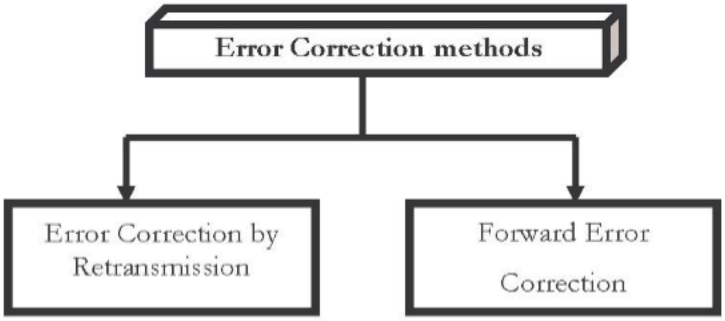


Figure 7.8: Error Correction Methods

Error correction can be handled by two ways: error correction by retransmission and forward error correction.

7.5.1 Error Correction by Retransmission

In error correction by retransmission, when an error is discovered, the receiver can have sender retransmits the entire data unit. In this mechanism, it allows receiver to inform the sender of the damaged or corrupted data units during transmission and coordinates the retransmission of those data units by sender.

However, if propagation delays, due to distance, are large, the technique may become as inefficient as to be useless.

7.5.2 Forward Error Correction

Error correction by retransmission is an acceptable method of handling data errors in LAN-based networks because retransmissions of most messages result in a short delay and a little extra use of bandwidth resources. Now, imagine a satellite orbiting around Jupiter or Saturn, transmitting critical visual data as binary stream information. The time it takes for those transmissions to reach Earth is measured in hours. During this time, the satellite has adjusted its orbit and is soaring across new territory and sending additional data. Correcting errors in these messages cannot be done by retransmission. A request for that retransmission takes as long to get to the satellite as the original message took to get to Earth. Then consider the time it would take to retransmit the message. What

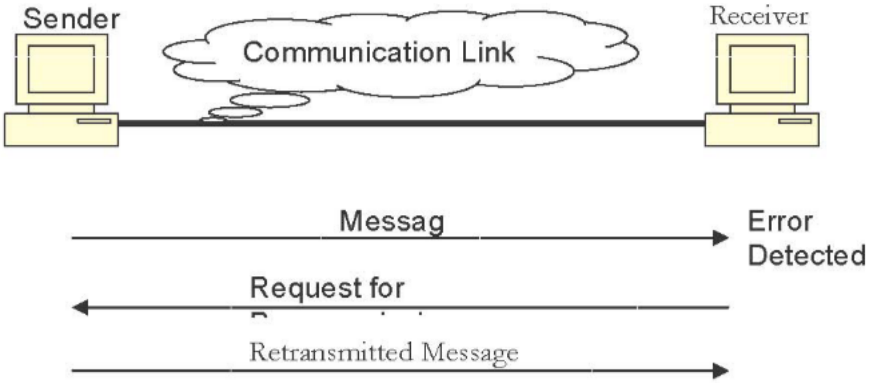


Figure 7.9: Error recovery by detection and retransmission

would the satellite do with new data, reach it while it tries to handle the retransmitting of old data? The memory needed to hold the old data in case it would need to be resent is astronomical to say the least. Instead, a forward error-correcting method such as the Hamming code is used so that errors can be corrected as they are detected.

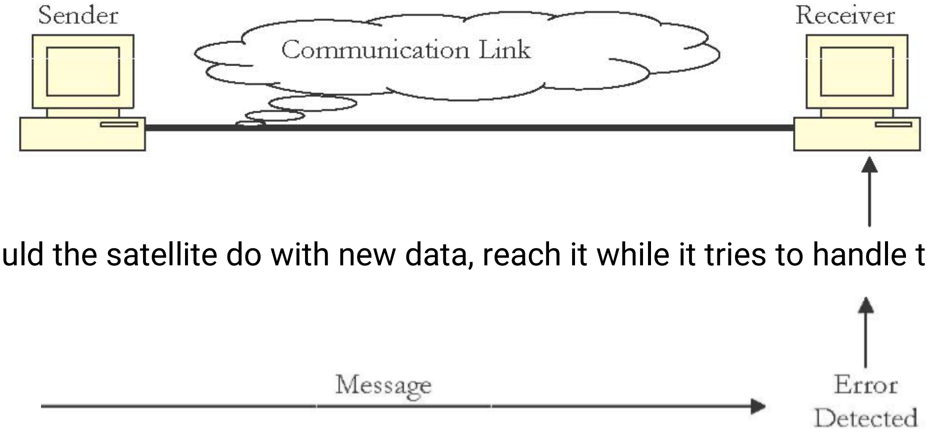


Figure 7.10: Forward Error Correction

7.5.2.1 Error Correction using Vertical Redundancy Check (VRC)/ Longitudinal Redundancy Check (LRC) or Two-¬dimensional Parity Check

Parity check is primarily used for detecting errors in a serial data character. A bad parity match indicates a logic error has occurred in one of the character’s data bits. The use of parity called a vertical redundancy check (VRC) can be extended to allow single-bit error correction to take place in a received data stream. By having the ability to correct an error, a receiver would not require a message to be retransmitted, but could do the correction itself. The trade-off in using an error-correction scheme is that an additional character has to be sent with the message and additional software and/or hardware must be used to create and interpret that character. For asynchronous data transmission, that character is known as the longitudinal redundancy check (LRC) character.

Using a VRC/LRC or 2-dimensional parity check system, the message is sent with each character containing the regular even-parity bit known as the VRC bit. As with error-detection schemes, any mismatch between transmitted and received VRCs indicates that the character contains a bad data bit. In order to correct the bad bit, what is left to be done is to determine which of the character's bits the bad one is. This is where LRC comes in. It is used to create a cross-matrix type of configuration where the VRC bit denotes the row (character) and the LRC, the column (bit position) of the message's bad bit. At the sending site, each of the data bits of each character is exclusive ORed with the bits of all the other data bits.

This error-correction method and others, which are similar, are known as forward error correction (FEC) because errors are corrected as the message is received. There is no requirement to retransmit the message as long as the errors remain infrequent. If more than one error occurs in a message, then more than one LRC and one VRC bit will be bad and there is no way to determine which LRC bit goes with which VRC character. In this case, the excessive number of bit errors is indicative of a severe condition. Once the cause of the problem is resolved, the message will have to be retransmitted fully.

7.5.2.2 Hamming Code

For synchronous data streams, an error-correcting process called Hamming code is commonly used. This method is fairly complex from the standpoint of creating and interpreting the error bits. It is implemented in software algorithms and relies on a lot of preliminary conditions agreed upon by the sender and receiver. Error bits, called Hamming bits, are inserted into the message at random locations. It is believed that the randomness of their locations reduces the statistical odds that these Hamming bits themselves would be in error. This is based on a mathematical assumption that because there are so many more messages bits compared to Hamming bits, that there is a greater chance for a message bit to be in error than for a Hamming bit to be wrong.

Hamming code differs from other error detection and correction codes such as CRC or Checksum. In the other error detection and correction codes such as checksum or CRC, redundant bit or error control bits are appended at the end of data unit but in case of Hamming code, error control codes are randomly inserted into data unit. But each and every bit in the message, including the Hamming bits, has the same chance of being corrupted as any other bit. Be that as it may, Hamming bits are inserted into the data stream randomly. The only crucial point in the selection of their locations is that both the sender and receiver are aware of where they actually are.

The first step in the process is to determine how many Hamming bits (H) are to be inserted between the message (M) bits. Then their actual placement is selected. The number of bits in the message (M) are counted and used to solve the following equation to determine the number of Hamming (H) bits:

$$2^{H} \geq M + H - 1$$

Number of Data Bits (M)	Number of Redundancy Bits (H)
1	2
2	3
3	3
4	3
5	4
6	4
7	4
8	4

Table 7.1: Relationship between data and redundancy bits.

Once the number of Hamming bits is determined, the actual placement of the bits into the message is performed. It is important to note that despite the random nature of the Hamming bit placements, the exact same placements must be known and used by both the transmitter and the receiver. This is necessary so that the receiver can remove the Hamming bits from the message sent by the transmitter and compare them with a similar set of bits generated at the receiver.

Example 7

How many Hamming bits are required when using the Hamming code with the extended ASCII synchronous message “Help!” ?

Error Method	Data Type	Detection / correction	Overhead
Parity	Asynchronous	Detection only	One Bit Added per Character
LRC/VRC	Asynchronous	Detection and Correction	One Bit per Character Plus LRC Character
Checksum	Asynchronous/ Synchronous	Detection only	Checksum Character at End of Message
CRC	Synchronous	Detection only	CRC Bytes at End of Message
Hamming Code	Synchronous	Detection and Correction	Hamming Bits Inserted into Data Stream

Solution

The total number of bits in the message is:

$$M = 8\text{-bits/character} - 5 \text{ characters} = 40 \text{ bits}$$

This number is used in Equation ($2^H \geq M + H - 1$) to determine the number of Hamming bits:

$$2^H \geq 40 + H + 1$$

The closest value to try is 6 bits for H, since $2^6 = 64$, which is greater than $40 + 6 + 1 = 47$. This satisfies the equation. Number of Redundancy bits or Hamming bits required are six.

The Hamming code can be applied to data units of any length, which uses the relationship between data and redundancy bits discussed above. Once the Hamming bits are inserted into their positions within the message, their states (high or low) need to be determined. Starting with the least significant bit (LSB) as bit 1, the binary equivalent of each message-bit position with a high (1) state is exclusive ORed with every other bit position containing a 1. The result of the exclusive-OR process is the states of the Hamming bits.

Table 7.2: Error Methods Summary

8. BUS STRUCTURE AND LOOP SYSTEM

8.1 Bus Structure

If different functional units of a computer are connected in a organized way then only the computer will carry out the task. There are different ways in connecting. When a word of data is transferred between units, all its bits are transferred in parallel, so this requires many number of wires. A collection of lines that connects several devices is called a bus.

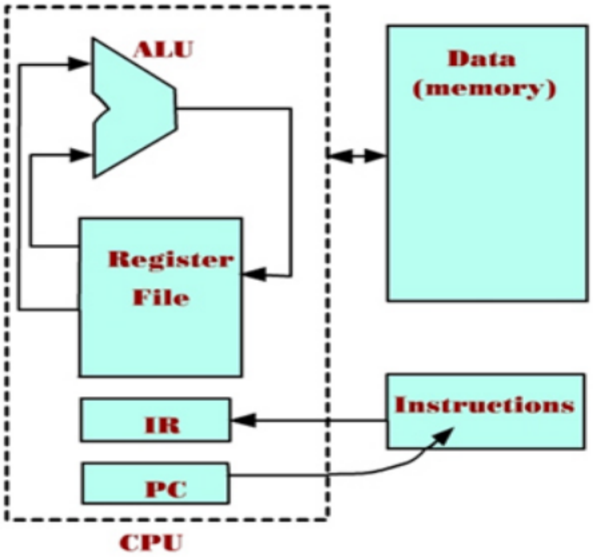


Figure 8.1: Load and Store Machine Architecture

i.e a group of lines that serves as a connecting path for several devices is called .In addition to the lines that carry the data, the bus must have lines for address and control purposes.

In computer architecture, a **bus** is a subsystem that transfers data between computer components inside a computer or between computers. A computer bus structure is provided which permits replacement of removable modules during operation of a computer wherein means are provided to precharge signal output lines to within a predetermined range prior to the usage of the signal output lines to carry signals, and further, wherein means are provided to minimize arcing to pins designed to carry the power and signals of a connector. In a specific embodiment, pin length, i.e., separation between male and female components of the connector, are subdivided into long pin length and short pin length. Ground connections and power connections for each voltage level are assigned to the long pin lengths. Signal connections and a second power connection for each voltage level is assigned to the short pin lengths.

The precharge/prebias circuit comprises a resistor divider coupled between a power source and ground with a high impedance tap coupled to a designated signal pin, across which is coupled a charging capacitor or equivalent representing the capacitance of the signal line. Bias is applied to the precharge/prebias circuit for a sufficient length of time to precharge the signal line to a desired neutral signal level between expected high and low signal values prior to connection of the short pin to its mate.

Early computer buses were literally parallel electrical buses with multiple connections, but the term is now used for any physical arrangement that provides the same logical functionality as a parallel electrical bus. Modern computer buses can use both parallel and bit-serial connections, and can be wired in either a multidrop (electrical parallel) or daisy chain topology, or connected by switched hubs, as in the case of USB.

8.1.1 Description of BUS

At one time, "bus" meant an electrically parallel system, with electrical conductors similar or identical to the pins on the CPU. This is no longer the case, and modern systems are blurring the lines between buses and networks. Buses can be parallel buses, which carry data words in parallel on multiple wires, or serial buses, which carry data in bit-serial form. The addition of extra power and control connections, differential drivers, and data connections in each direction usually means that most serial buses have more conductors than the minimum of one used in the 1-Wire and UNI/O serial buses. As data rates increase, the problems of timing skew, power consumption, electromagnetic interference and crosstalk across parallel buses become more and more difficult to circumvent. One partial solution to this problem has been to double pump the bus. Often, a serial bus can actually be operated at higher overall data rates than a parallel bus, despite having fewer electrical connections, because a serial bus inherently has no timing skew or crosstalk. Universal Serial Bus (USB), FireWire, and Serial ATA are examples of this. Multidrop connections do not work well for fast serial buses, so most modern serial buses use daisy-chain or hub designs.

Most computers have both internal and external buses. An *internal bus* connects all the internal components of a computer to the motherboard (and thus, the CPU and internal memory). These types of buses are also referred to as a local bus, because they are intended to connect to local devices, not to those in other machines or external to the computer. An *external bus* connects external peripherals to the motherboard.

Network connections such as Ethernet are not generally regarded as buses, although the difference is largely conceptual rather than practical. The arrival of technologies such as InfiniBand and HyperTransport is further blurring the boundaries between networks and buses. Even the lines between internal and external are sometimes fuzzy, Inter-Integrated Circuit I²C can be used as both an internal bus, or an external bus (where it is known as ACCESS.bus), and InfiniBand is intended to replace both internal buses like Peripheral component interconnect (PCI) as well as external ones like Fibre Channel. In the typical desktop application, USB serves as a peripheral bus, but it also sees some use as a networking utility and for connectivity between different computers, again blurring the conceptual distinction.

System bus- This consists of data bus, address bus and control bus
Data bus- A bus which carries data to and from memory/IO is called as data bus
Address bus- This is used to carry the address of data in the memory and its width is equal to the number of bits in the MAR of the memory.
For example . If comp. memory of 64K has 32 bit words then the computer will have a

data bus of 32 bits wide and the address bus of 16 bits wide

Control Bus- carries the control signals between the various units of the computer. Ex: Memory Read/write, I/O Read/write

8.1.2 Two types of Bus organizations

Single Bus organization

Two bus Organization

8.1.2.1 Single Bus Architecture

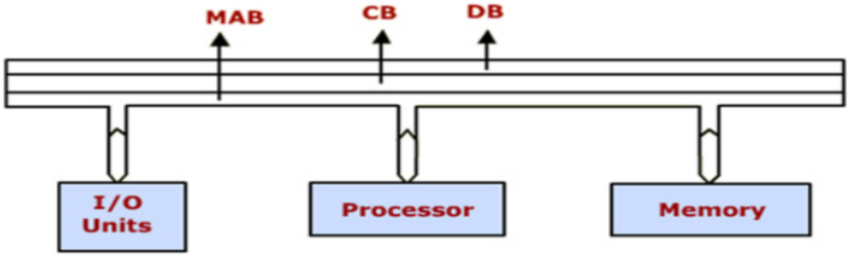


Figure 8.2: Single-Bus Organization

- Three units share the single bus. At any given point of time, information can be transferred between any two units
- Here I/O units use the same memory address space (Memory mapped I/O)
- So no special instructions are required to address the I/O, it can be accessed like a memory location
- Since all the devices do not operate at the same speed, it is necessary to smooth out the differences in timings among all the devices A common approach used is to include buffer registers with the devices to hold the information during transfers

Example: Communication between the processor and printer

8.1.2.2 Two Bus Architecture

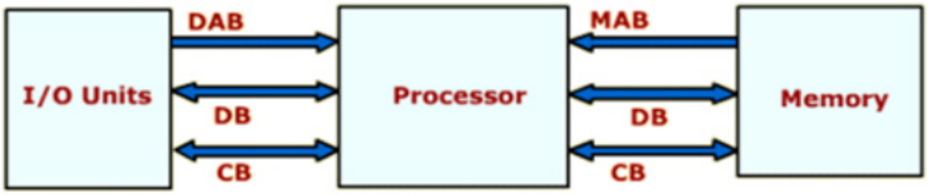


Figure 8.3: Two Bus Organization

- Various units are connected through two independent buses
- I/O units are connected to the processor though an I/O bus and Memory is connected to the processor through the memory bus
- I/O bus consists of address, data and control bus Memory bus also consists of address, data and control bus In this type of arrangements processor completely supervises the transfer of information to and from I/O units. All the information is first taken to processor and from there to the memory . Such kind of transfers are called as program controlled transfer

Alternative Two Bus Architecture

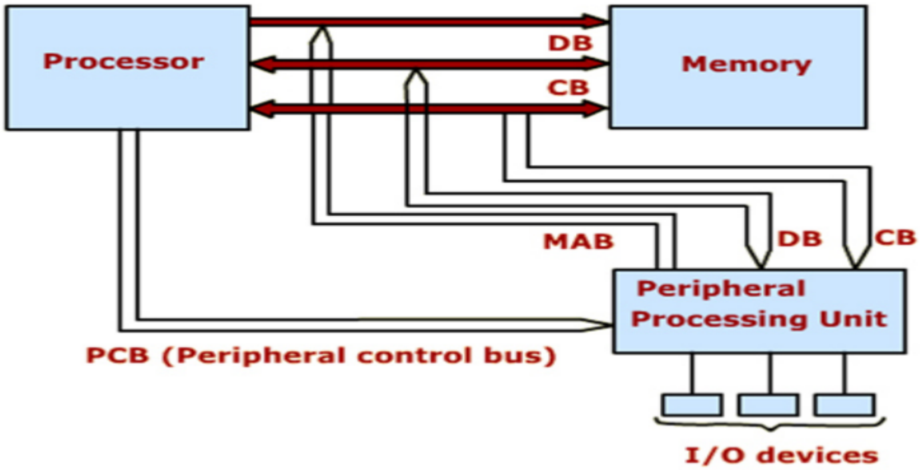


Figure 8.4: Alternative Two-Bus Organization

In this I/O units are directly connected to the memory and not to the processor

The I/O units are connected to special interface logic known as Direct Memory Access (DMA) or an I/O channel. This is also called as Peripheral Processor Unit (PPU)

In this the data from the I/O device is directly sent to memory bypassing the processor.

8.1.3 Bus standards

- ISA (Industry standard Architecture)

  -Developed by IBM

  -Speed is 8 MHz

  -16 bit Interface

- EISA (Extended Industry Standard Architecture)
- VESA (Video Electronics Industry Standard Architecture)
- PCI (Peripheral Component Interconnect)

  -Developed by Intel

  -Speed is 33MHz,Also available in 66MHz speed

  -64 bit interface

8.1.4 Computer Mother Board

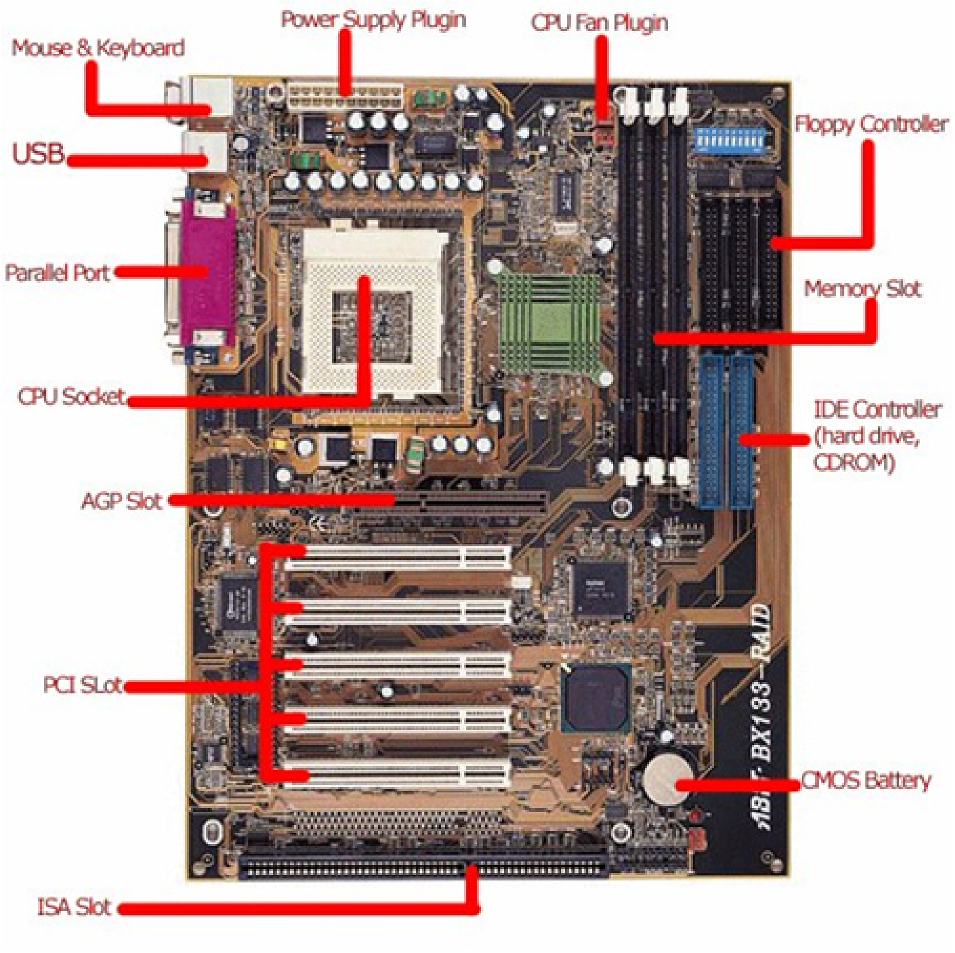


Figure 8.5: Computer Mother Board

8.2 Control Loop System

A **control system** is a device, or set of devices, that manages, commands, directs or regulates the behaviour of other devices or systems to achieve desire results. In other words **control system** is a system, which controls other system. Industrial control systems are used in industrial production for controlling equipment or machines.

There are two common classes of control systems, open loop control systems and closed loop control systems. In open loop control systems output is generated based on inputs. In closed loop control systems current output is taken into consideration and corrections are made based on feedback. A closed loop system is also called a feedback control system.

8.2.1 Feature of Control System

The main feature of control system is, there should be a clear mathematical relation between input and output of the system. When the relation between input and output of the system can be represented by a linear proportionality, the system is called linear control system. Again when the relation between input and output cannot be represented by single linear proportionality, rather the input and output are related by some non-linear relation, the system is referred as non-linear control system.

8.2.2 Requirement of Good Control System

Accuracy: Accuracy is the measurement tolerance of the instrument and defines the limits of the errors made when the instrument is used in normal operating conditions. Accuracy can be improved by using feedback elements. To increase accuracy of any control system error detector should be present in control system.

Sensitivity: The parameters of control system are always changing with change in surrounding conditions, internal disturbance or any other parameters. This change can be expressed in terms of sensitivity. Any control system should be insensitive to such parameters but sensitive to input signals only.

Noise: An undesired input signal is known as noise. A good control system should be able to reduce the noise effect for better performance.

Stability: It is an important characteristic of control system. For the bounded input signal, the output must be bounded and if input is zero then output must be zero then such a control system is said to be stable system.

Bandwidth: An operating frequency range decides the bandwidth of control system. Bandwidth should be large as possible for frequency response of good control system.

Speed: It is the time taken by control system to achieve its stable output. A good control system possesses high speed. The transient period for such system is very small.

Oscillation: A small numbers of oscillation or constant oscillation of output tend to system to be stable

A. Open loop system

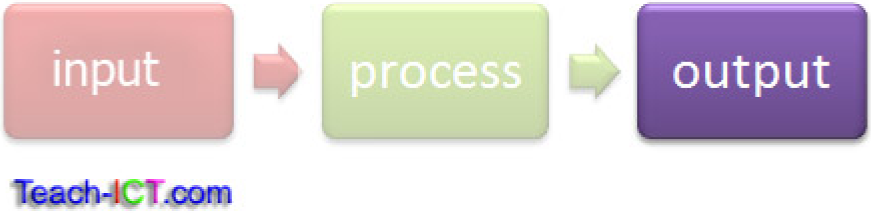


Figure 8.6: Open Loop System

The function of any electronic system is to automatically regulate the output and keep it within the systems desired input value or “set point”. If the systems input changes for whatever reason, the output of the system must respond accordingly and change itself to reflect the new input value.

Likewise, if something happens to disturb the systems output without any change to the input value, the output must respond by returning back to its previous set value. In the past, electrical control systems were basically manual or what is called an Open-loop System with very few automatic control or feedback features built in to regulate the process variable so as to maintain the desired output level or value.A control system in which the control action is totally independent of output of the system is called open loop control system. Open Loop System, only looks at its input signal in order to decide what to do.It takes no account at all of what is happening to its output.

For example, an electric clothes dryer. Depending upon the amount of clothes or how wet they are, a user or operator would set a timer (controller) to say 30 minutes and at the end of the 30 minutes the drier will automatically stop and turn-off even if the clothes are still wet or damp.

In this case, the control action is the manual operator assessing the wetness of the clothes and setting the process (the drier) accordingly.So in this example, the clothes dryer would be an open-loop system as it does not monitor or measure the condition of the output signal, which is the dryness of the clothes. Then the accuracy of the drying process, or success of drying the clothes will depend on the experience of the user (operator).

However, the user may adjust or fine tune the drying process of the system at any time by increasing or decreasing the timing controllers drying time, if they think that the original drying process will not be met. For example, increasing the timing controller to 40 minutes to extend the drying process. Consider the following open-loop block diagram.

Open-loop Drying System

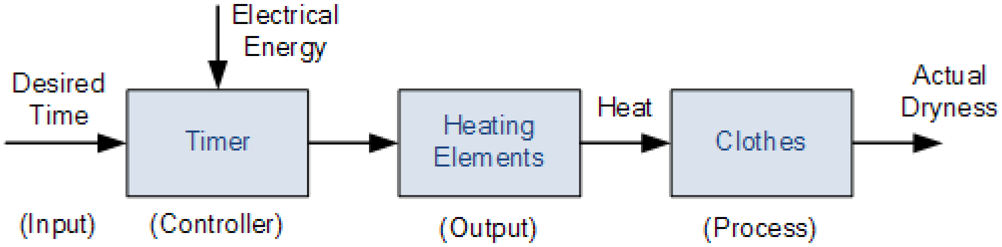


Figure 8.7: Open-Loop Drying System

Then an **Open-loop system**, also referred to as non-feedback system, is a type of continuous control system in which the output has no influence or effect on the control action of the input signal. In other words, in an open-loop control system the output is neither measured nor “fed back” for comparison with the input. Therefore, an open-loop system is expected to faithfully follow its input command or set point regardless of the final result. Also, an open-loop system has no knowledge of the output condition so cannot self-correct any errors it could make when the preset value drifts, even if this results in large deviations from the preset value.

Another disadvantage of open-loop systems is that they are poorly equipped to handle disturbances or changes in the conditions which may reduce its ability to complete the desired task. For example, the dryer door opens and heat is lost. The timing controller continues regardless for the full 30 minutes but the clothes are not heated or dried at the end of the drying process. This is because there is no information fed back to maintain a constant temperature.

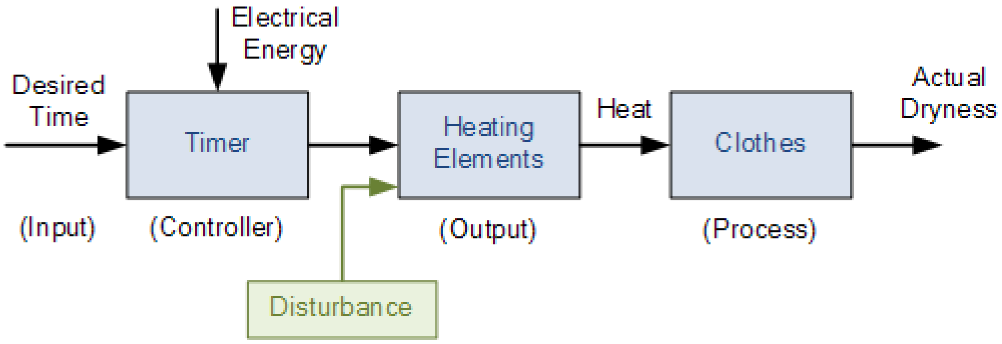
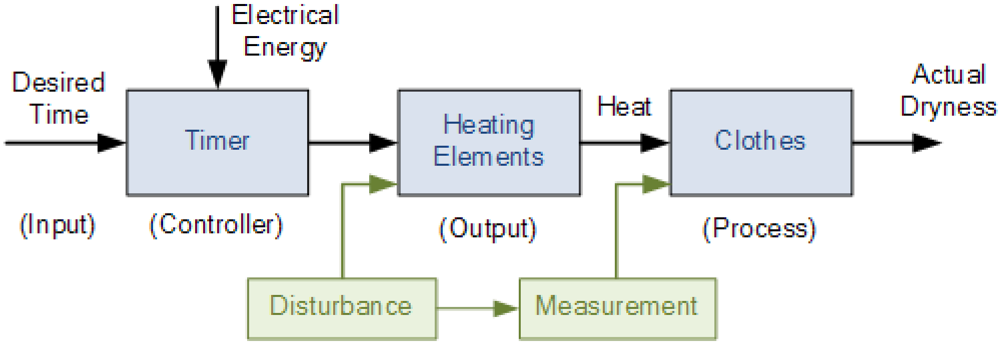


Figure 8.8: Open-Loop Drying System(with disturbance)

Then we can see that open-loop system errors can disturb the drying process and therefore requires extra supervisory attention of a user (operator). The problem with this anticipatory control approach is that the user would need to look at the process temperature frequently and take any corrective control action whenever the drying process deviated from its desired value of drying the clothes. This type of manual open-loop control which reacts before an error actually occurs is called **Feed forward Control.**

The objective of feed forward control, also known as predictive control, is to measure or predict any potential open-loop disturbances and compensate for them manually before the controlled variable deviates too far from the original set point. So for our simple example above, if the dryers door was open it would be detected and closed allowing the drying process to continue.



If applied correctly, the deviation from wet clothes to dry clothes at the end of the 30 minutes would be minimal if the user responded to the error situation (door open) very quickly. However, this feed forward approach may not be completely accurate if the system changes, for example the drop in drying temperature was not noticed during the 30 minute process.

Then we can define the main characteristics of an “Open-loop system” as being:

- There is no comparison between actual and desired values.
- An open-loop system has no self-regulation or control action over the output value.
- Each input setting determines a fixed operating position for the controller.
- Changes or disturbances in external conditions does not result in a direct output change.
  (unless the controller setting is altered manually)

You set the microwave oven to run for two minutes. After cooking
for two minutes, the control system turns the microwave off. It has no idea whether your food is still frozen, burnt or cooked perfectly.

Practical Examples of Open Loop System

1. Electric Hand Drier – Hot air (output) comes out as long as you keep your hand under the machine, irrespective of how much your hand is dried.
2. Automatic Washing Machine – This machine runs according to the pre-set time irrespective of washing is completed or not.
3. Bread Toaster - This machine runs as per adjusted time irrespective of toasting is completed or not.
4. Automatic Tea/Coffee Maker – These machines also function for pre adjusted time only.
5. Timer Based Clothes Drier – This machine dries wet clothes for pre – adjusted time, it does not matter how much the clothes are dried.
6. Light Switch – lamps glow whenever light switch is on irrespective of light is required or not.
7. Volume on Stereo System – Volume is adjusted manually irrespective of output volume level.

Advantages of Open Loop Control System

1. Simple in construction and design

2. Economical.
3. Easy to maintain.
4. Generally stable.
5. Convenient to use as output is difficult to measure.

Disadvantages of Open Loop Control System

1. They are inaccurate.
2. They are unreliable.
3. Any change in output cannot be corrected automatically

B. Closed loop system

We have seen that systems in which the output quantity has no effect upon the input to the control process are called open-loop control systems, and that open-loop systems are just that, open ended non-feedback systems. But the goal of any electrical or electronic control system is to measure, monitor, and control a process.

One way in which we can accurately Control the Processis by monitoring its output and "feeding" some of it back to compare the actual output with the desired output so as to reduce the error and if disturbed, bring the output of the system back to the original or desired response. The measure of the output is called the "feedback signal" and the type of control system which uses feedback signals to control itself is called a **Close-loop System**. A **Closed-loop Control System**, also known as a *feedback control system* is a control system which uses the concept of an open loop system as its forward path but has one or more feedback loops (hence its name) or paths between its output and its input. The reference to "feedback", simply means that some portion of the output is returned "back" to the input to form part of the systems excitation.

Closed-loop systems are designed to automatically achieve and maintain the desired output condition by comparing it with the actual condition. It does this by generating an error signal which is the difference between the output and the reference input. In other words, a "closed-loop system" is a fully automatic control system in which its control action being dependent on the output in some way.

So for example, consider our electric clothes dryer from the previous open-loop. Suppose we used a sensor or transducer (input device) to continually monitor the temperature or dryness of the clothes and feed a signal relating to the dryness back to the controller as shown below.

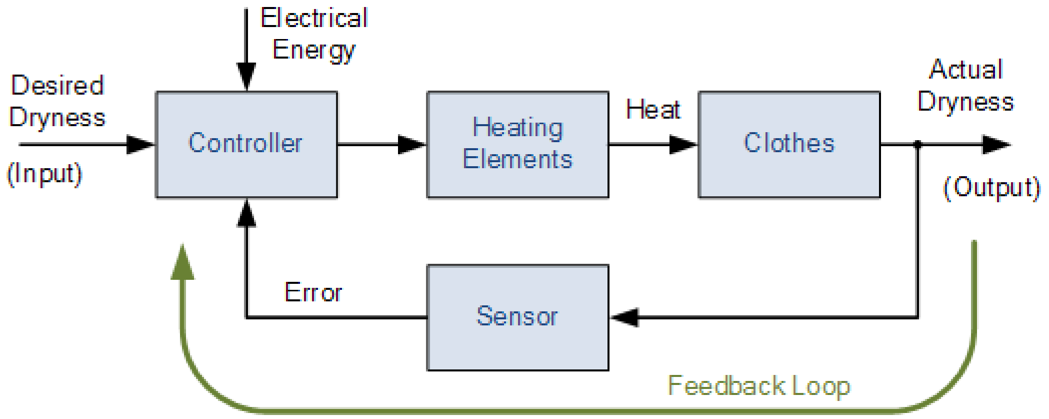


Figure 8.9: Closed-Loop System

This sensor would monitor the actual dryness of the clothes and compare it with (or subtract it from) the input reference. The error signal (error = required dryness – actual dryness) is amplified by the controller, and the controller output makes the necessary correction to the heating system to reduce any error. For example if the clothes are too wet the controller may increase the temperature or drying time. Likewise, if the clothes are nearly dry it may reduce the temperature or stop the process so as not to overheat or burn the clothes, etc.

Then the closed-loop configuration is characterised by the feedback signal, derived from the sensor in our clothes drying system. The magnitude and polarity of the resulting error signal, would be directly related to the difference between the required dryness and actual dryness of the clothes.Also, because a closed-loop system has some knowledge of the output condition, (via the sensor) it is better equipped to handle any system disturbances or changes in the conditions which may reduce its ability to complete the desired task.

For example, as before, the dryer door opens and heat is lost. This time the deviation in temperature is detected by the feedback sensor and the controller self-corrects the error to maintain a constant temperature within the limits of the preset value. Or possibly stops the process and activates an alarm to inform the operator.As we can see, in a closed-loop control system the error signal, which is the difference between the input signal and the feedback signal (which may be the output signal itself o function of the output signal), is fed to the controller so as to reduce the systems error and bring the output of the system back to a desired value. In our case the dryness of the clothes. Clearly, when the error is zero the clothes are dry.

The term Closed-loop control always implies the use of a feedback control action in order to reduce any errors within the system, and its “feedback” which distinguishes the main differences between an open-loop and a closed-loop system.The accuracy of the output thus depends on the feedback path, which in general can be made very accurate and within electronic control systems and circuits, feedback control is more commonly used than open-loop or feed forward control.

Closed-loop systems have many advantages over open-loop systems. The primary advantage of a closed-loop feedback control system is its ability to reduce a system’s sensitivity to external disturbances, for example opening of the dryer door, giving the system a more robust control as any changes in the feedback signal will result in compensation by the controller.

Then we can define the main characteristics of Closed-loop Control as being:

- To reduce errors by automatically adjusting the systems input.
- To improve stability of an unstable system.
- To increase or reduce the systems sensitivity.
- To enhance robustness against external disturbances to the process.
- To produce a reliable and repeatable performance.

Whilst a good closed-loop system can have many advantages over an open-loop control system, its main disadvantage is that in order to provide the required amount of control, a closed-loop system must be more complex by having one or more feedback paths. Also, if the gain of the controller is too sensitive to changes in its input commands or signals it can become unstable and start to oscillate as the controller tries to over-correct itself, and eventually something would break. So we need to “tell” the system how we want it to behave within some pre-defined limits.

Practical Examples of Closed Loop Control System

1. Automatic Electric Iron – Heating elements are controlled by output temperature of the iron.
2. Servo Voltage Stabilizer – Voltage controller operates depending upon output voltage of the system.
3. Water Level Controller– Input water is controlled by water level of the reservoir.
4. Missile Launched & Auto Tracked by Radar – The direction of missile is controlled by comparing the target and position of the missile.
5. An Air Conditioner – An air conditioner functions depending upon the temperature of the room.
6. Cooling System in Car – It operates depending upon the temperature which it controls.

Advantages of Closed Loop Control System

1. Closed loop control systems are more accurate even in the presence of non-linearity.
2. Highly accurate as any error arising is corrected due to presence of feedback signal.
3. Bandwidth range is large.
4. Facilitates automation.
5. The sensitivity of system may be made small to make system more stable.
6. This system is less affected by noise

Disadvantages of Closed Loop Control System

1. They are costlier.
2. They are complicated to design.
3. Required more maintenance.
4. Feedback leads to oscillatory response.

5. Overall gain is reduced due to presence of feedback.
6. Stability is the major problem and more care is needed to design a stable closed loop system.

8.3 Comparison of Closed Loop and Open Loop Control System

Sr. No.	Open loop control system	Closed loop control system
1	The feedback element is absent.	The feedback element is always present.
2	An error detector is not present.	An error detector is always present.
3	It is stable one.	It may become unstable.
4	Easy to construct.	Complicated construction.
5	It is an economical.	It is costly.
6	Having small bandwidth.	Having large bandwidth.
7	It is inaccurate.	It is accurate.
8	Less maintenance.	More maintenance.
9	It is unreliable.	It is reliable.
10	Examples: Hand drier, tea maker	Examples: Servo voltage stabilizer, perspiration

9. COMMUNICATION PROTOCOL

Communication protocols are formal descriptions of digital message formats and rules. They are required to exchange messages in or between computing systems and are required in telecommunications. Communications protocols cover authentication, error detection and correction, and signaling. They can also describe the syntax, semantics, and synchronization of analog and digital communications. Communications protocols are implemented in hardware and software. There are thousands of communications protocols that are used everywhere in analog and digital communications. Computer networks cannot exist without them.

Communications devices have to agree on many physical aspects of the data to be exchanged before successful transmission can take place. Rules defining transmissions are called protocols.

Protocol

- It is a set of rules governing the format and meaning of frames, packets, or message s that are
exchanged by peer entities within a layer.
- Protocol are used for communications between entities in a systems.
- Entities use protocols in order to implement their service definitions.

The key elements of a protocol are:
Syntax : Include Time data formats and signal levels
Semantics: Includes control information and error handling

There are many properties of a transmission that a protocol can define. Common ones include: packet size, transmission speed, error correction types, handshaking and synchronization techniques, address mapping, acknowledgement processes, flow control, packet sequence controls, routing, address formatting

Protocols are generally described using a layered architecture known as the Open System Interconnection (OSI) reference model, which abstracts the details of the protocol and allows a simple description of the service provided by the protocol to the protocol layer above and the service required by protocol layer from the layer below.

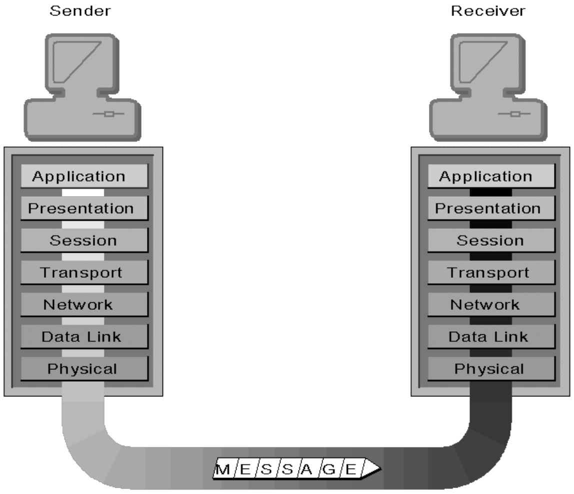
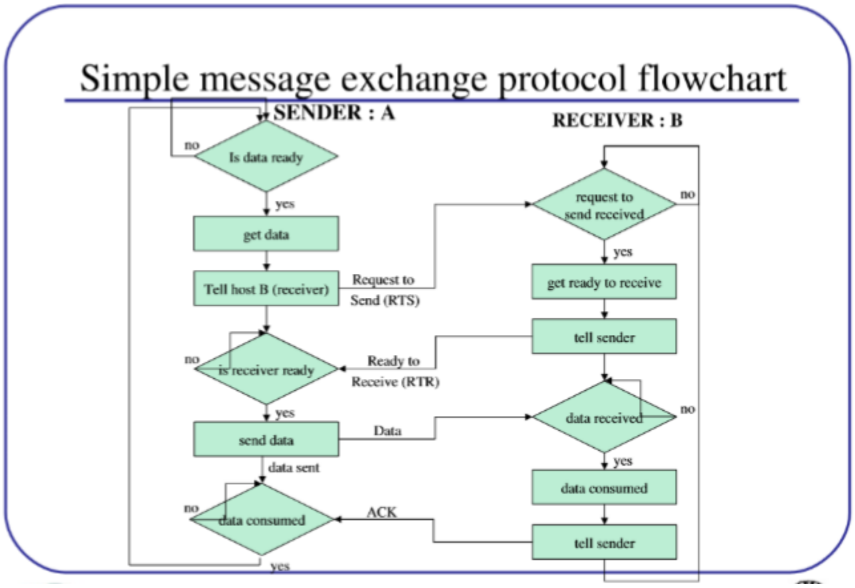


Figure 9.1: OSI reference Model

Popular protocols include: File Transfer Protocol (FTP), TCP/IP, User Datagram Protocol (UDP), Hypertext Transfer Protocol (HTTP), Post Office Protocol (POP3), Internet Message Access Protocol (IMAP), Simple Mail Transfer Protocol (SMTP).


Simple message exchange protocol flowchart
SENDER : A
RECEIVER : B
no
Is data ready
yes
get data
Tell host B (receiver)
Request to
Send (RTS)
no
is receiver ready
Ready to
Receive (RTR)
yes
send data
Data
data sent
no
data consumed
ACK
yes
request to
send received
no
yes
get ready to receive
tell sender
data received
no
yes
data consumed
tell sender