Contents

4 Data Line Devices

Scanned by CamScanner

- Functions relating to blocking (grouping) of data characters
  - Start of text (STX)
  - End of text (ETX)
- User definable functions
  - DC1, DC2, DC3, and DC4

  DC1 and DC3 are generally used as X-ON and X-OFF for switching the transmitter.

1.2.2 Byte

Byte is a group of bits which is considered as a single unit during processing. It is usually eight bits long though its length may be different. A byte may be an element of a standard code set. But it is not necessary. It may consist of any combination of bits.

7-bit ASCII code is appended with an additional bit that makes it eight bits long. The additional bit can have significance as parity bit or it may not be of any significance than filling the vacant eighth bit position. ASCII character K can be written as an 8-bit byte 01001011.

1.3 DATA TRANSMISSION

There is always need to exchange data, commands, and other control information between a computer and its terminals or between two computers. This information, as we saw in the previous section, is in the form of bits. Data transmission refers to movement of the bits over some physical medium connecting two or more digital devices. There are two options of transmitting the bits, namely, parallel transmission, or serial transmission.

1.3.1 Parallel Transmission

In *parallel transmission*, all the bits of a byte are transmitted simultaneously on separate wires as shown in Figure 1.2. Multiple circuits interconnecting the two devices are, therefore,

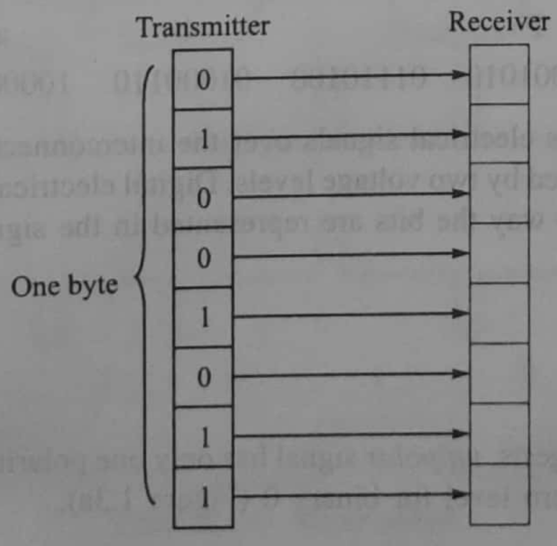

FIGURE 1.2 Parallel transmission.

required. It is practical only if the two devices, e.g. a computer and its associated printer, are close to each other.

1.3.2 Serial Transmission

In *serial transmission*, bits are transmitted serially one after the other (Figure 1.3). The Least Significant Bit (LSB) is usually transmitted first. Serial transmission requires only one circuit interconnecting the two devices. Therefore, serial transmission is suitable for interconnecting devices as a network or for transmission over long distances.

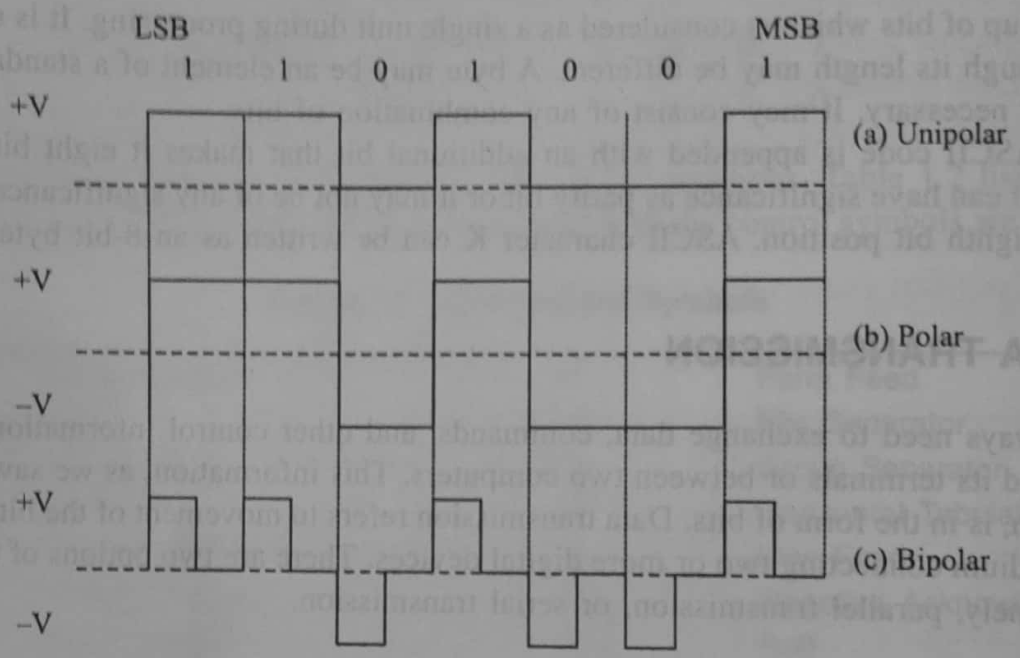

FIGURE 1.3 Serial transmission.

EXAMPLE 1.2 Write bit transmission sequence of the message given in Example 1.1.

Solution

3	P		b	a	t
11001100	00001010	01110100	01000110	10000110	00101110

Bits are transmitted as electrical signals over the interconnecting wires. The two binary states, 1 and 0, are represented by two voltage levels. Digital electrical signals can be categorized in three types based on the way the bits are represented in the signal.

- Unipolar
- Polar
- Bipolar

Unipolar. As the name suggests, *unipolar* signal has only one polarity. Usually positive polarity is used for binary 1 and zero level for binary 0 (Figure 1.3a).

Polar. A *polar* signal has two levels of opposite polarity (Figure 1.3b). Note that the signal never comes back to zero voltage level. It is always in one of the two states, +V or –V.

Bipolar. A *bipolar* signal has three level states +V, –V and zero level (Figure 1.3c). Bipolar signal is also termed as pseudo-ternary because it has three level states but uses two of these to represent the two binary digits, 0 and 1.

1.3.3 Bit Rate

Bit rate is simply the number of bits which can be transmitted in a second. If t_p is the duration of a bit, the bit rate R will be $1/t_p$. It must be noted that bit duration is not necessarily the pulse duration. In Figure 1.3a the first pulse is of two-bit duration. We will, later, come across signal formats in which the pulse duration is only half the bit duration.

1.3.4 Receiving Data Bits

The signal received at the other end of the transmitting medium is never identical to the transmitted signal as the transmission medium distorts the signal to some extent (Figure 1.4b). There may be additional pulses due to noise spikes. As a result, the receiver has to put in considerable effort in identifying the bits. The receiver must know the time instant at which it should look for a bit. Therefore, the receiver must have synchronized clock pulses which mark the location of the bits (Figure 1.4c). The received signal is sampled using the clock pulses, and depending on the polarity of a sample, the corresponding bit is identified (Figure 1.4e). Sampling serves another purpose. Unwanted noise spikes that are added during transmission are removed unless a spike occurs at the sampling instant.

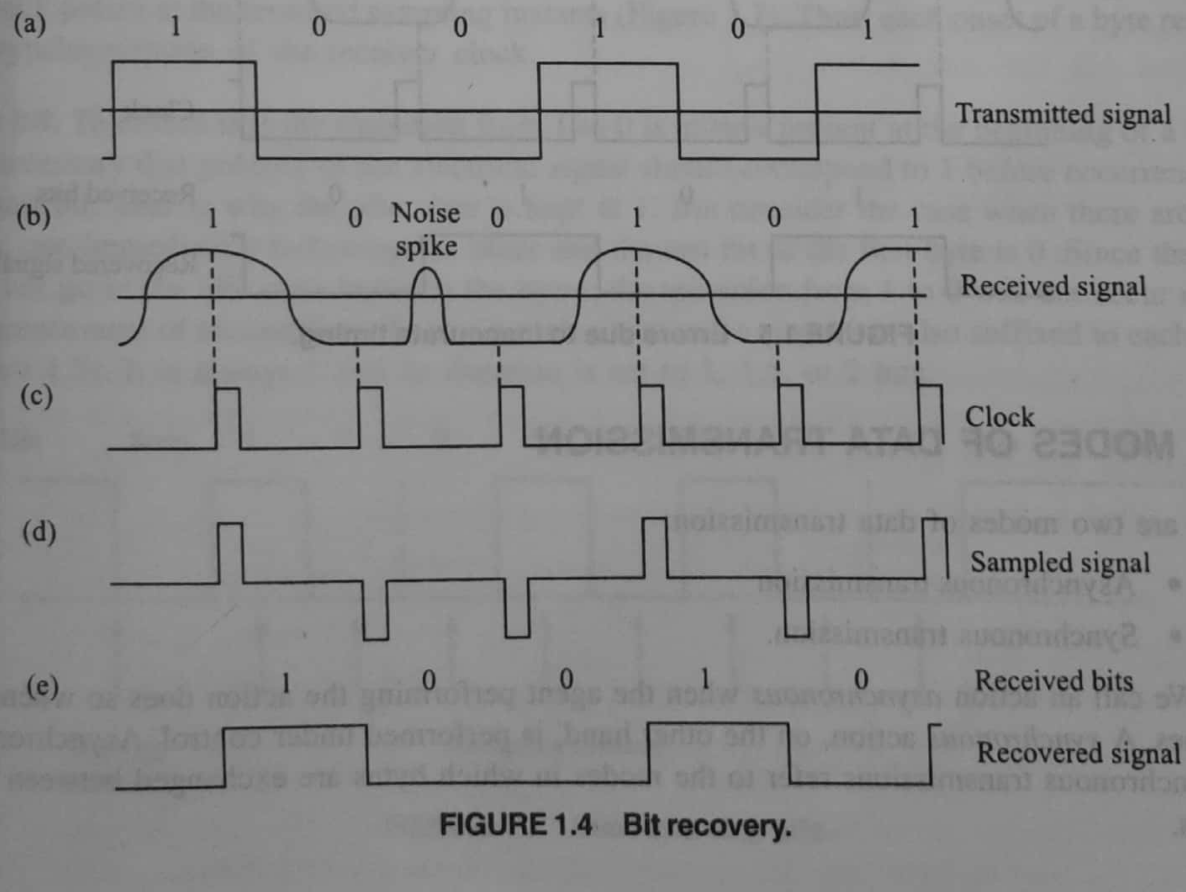

FIGURE 1.4 Bit recovery.

It is essential that the received signal is sampled at the right instants else it could be misinterpreted. Therefore, the clock frequency should be exactly the same as the transmission bit rate. Even a small difference in frequency will build up as timing error and eventually result in sampling at wrong instants. Figure 1.5 shows two situations when the receiver clock frequency is slightly faster and slightly slower than the bit rate. When clock frequency is faster, a bit may be sampled twice. A bit may be missed when the clock frequency is slower.

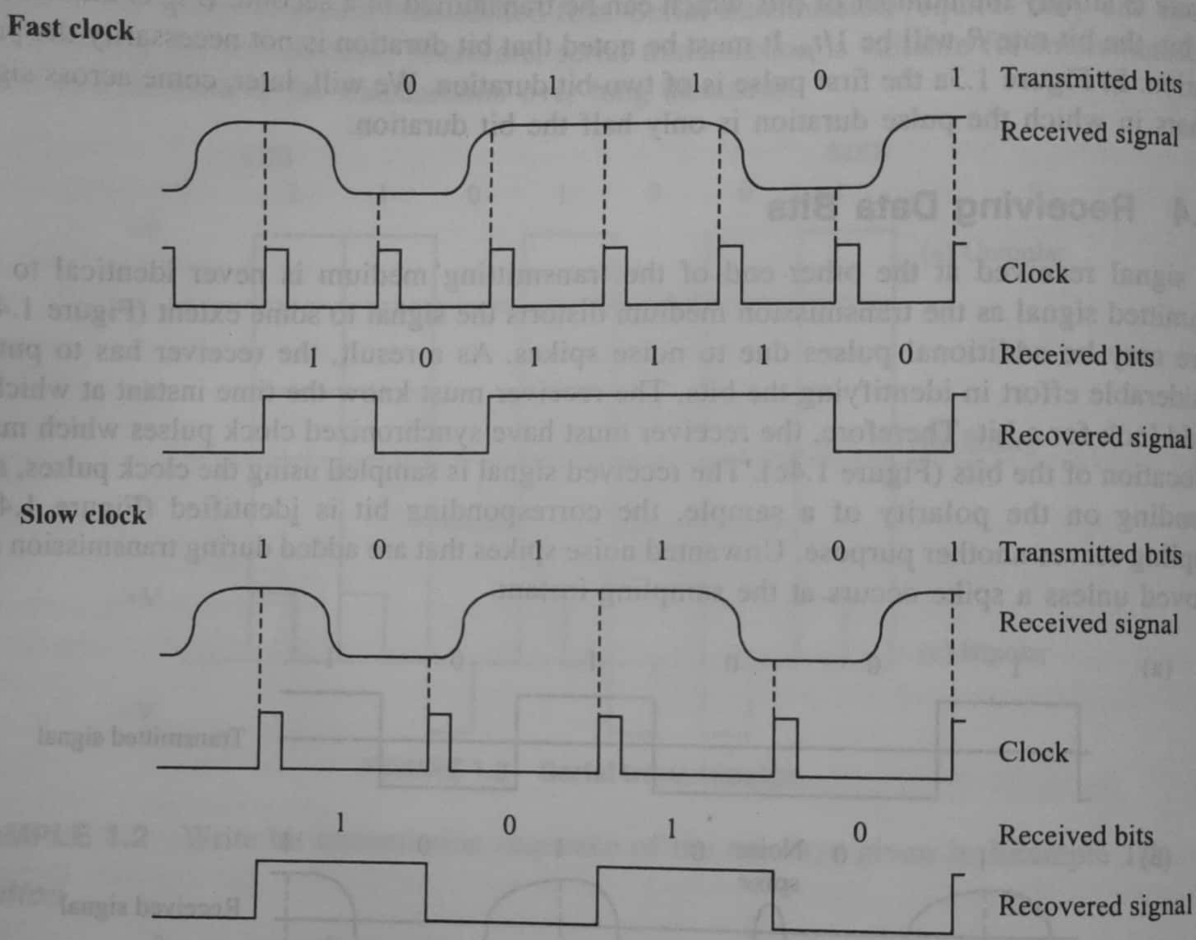

FIGURE 1.5 Errors due to inaccurate timing.

1.4 MODES OF DATA TRANSMISSION

There are two modes of data transmission:

- Asynchronous transmission
- Synchronous transmission.

We call an action *asynchronous* when the agent performing the action does so whenever it wishes. A *synchronous* action, on the other hand, is performed under control. Asynchronous and synchronous transmissions refer to the modes in which bytes are exchanged between two devices.

1.4.1 Asynchronous Transmission

Asynchronous transmission refers to the case when the sending end commences transmission of bytes at any instant of time. Only one byte is sent at a time and there is no time relation between consecutive bytes, i.e. after sending a byte, the next byte can be sent after arbitrary delay (Figure 1.6). The signal level during the idle state, when no byte is being transmitted, corresponds to 1 as per the accepted practice.

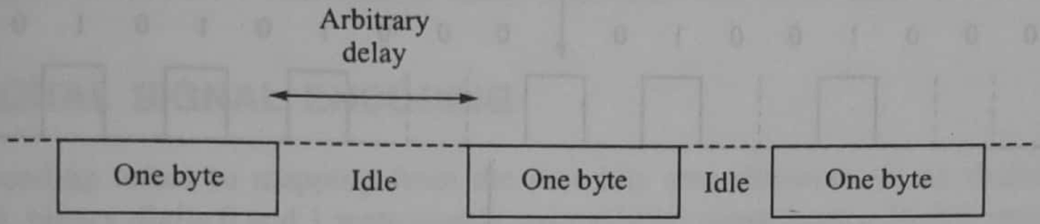

FIGURE 1.6 Asynchronous transmission.

Due to the arbitrary delay between consecutive bytes, the time occurrences of the clock pulses at the receiving end need to be synchronized repeatedly for each byte. This is achieved by providing two extra bits, a *start bit* at the beginning and a *stop bit* at the end of a byte.

Start bit. The *start bit* is always 0 and is prefixed to each byte. At the onset of transmission of a byte, it ensures that the electrical signal changes from idle state 1 to 0 and remains at 0 for one bit duration. The leading edge of the start bit is used as a time reference for generating the clock pulses at the required sampling instants (Figure 1.7). Thus, each onset of a byte results in resynchronization of the receiver clock.

Stop bit. To ensure that the transition from 1 to 0 is always present at the beginning of a byte, it is necessary that polarity of the electrical signal should correspond to 1 before occurrence of the start bit. That is why the idle state is kept at 1. But consider the case when there are two bytes, one immediately following the other and the last bit of the first byte is 0. Since the line does not go in the idle state between the bytes, the transition from 1 to 0 will not occur at the commencement of second byte. To avoid such situations, a *stop bit* is also suffixed to each byte (Figure 1.7). It is always 1 and its duration is set to 1, 1.5, or 2 bits.

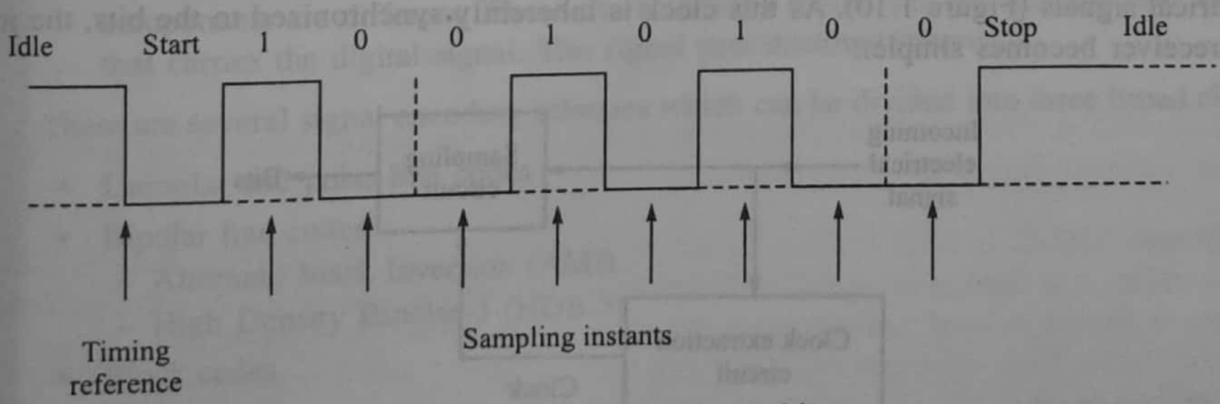

FIGURE 1.7 Start and stop bits.

EXAMPLE 1.3 Sketch the logic levels for the message 'HT' when it is transmitted in asynchronous mode with stop bit equal to one bit. Use ASCII code with parity bit 0 (Figure 1.8).

Solution

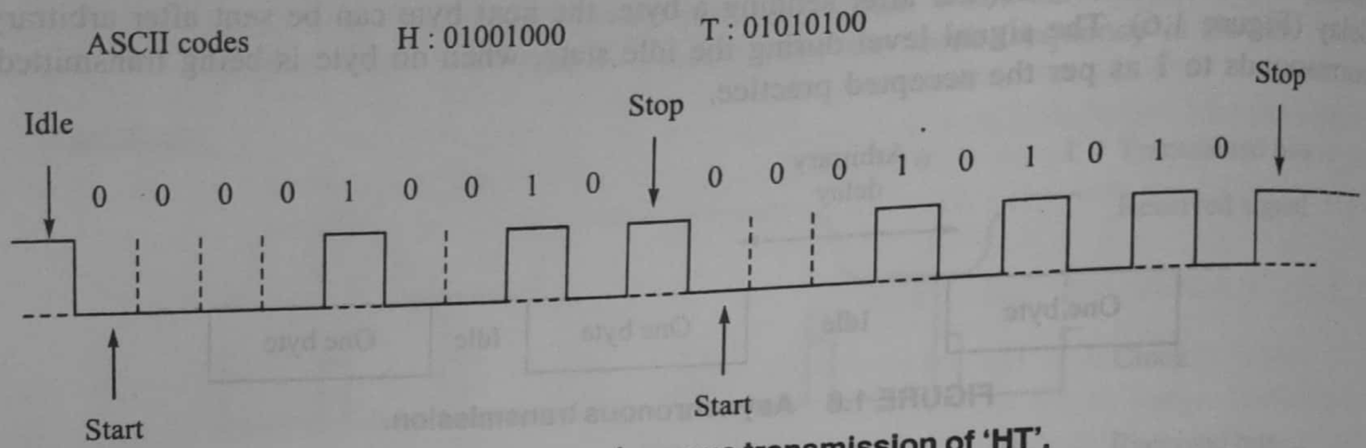

FIGURE 1.8 Asynchronous transmission of 'HT'.

1.4.2 Synchronous Transmission

A synchronous action, unlike on asynchronous action, is carried out under the control of a timing source. In synchronous transmission, all the bits are always synchronized to a reference clock irrespective of the bytes they belong to. There are no start or stop bits. Data bytes are transmitted as a block in a continuous stream of bits (Figure 1.9). Even the inter-block idle time is filled with idle characters.

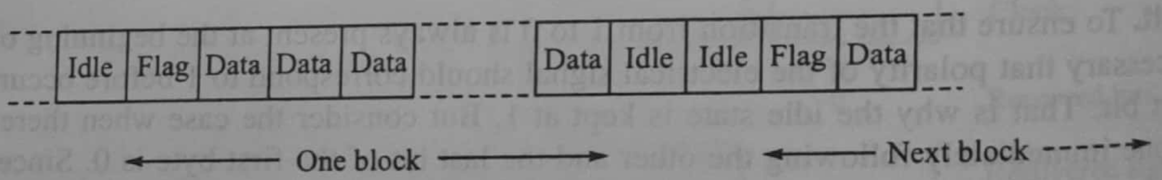

FIGURE 1.9 Synchronous transmission.

Continuous transmission of bits enables the receiver to extract the clock from the incoming electrical signals (Figure 1.10). As this clock is inherently synchronized to the bits, the job of the receiver becomes simpler.

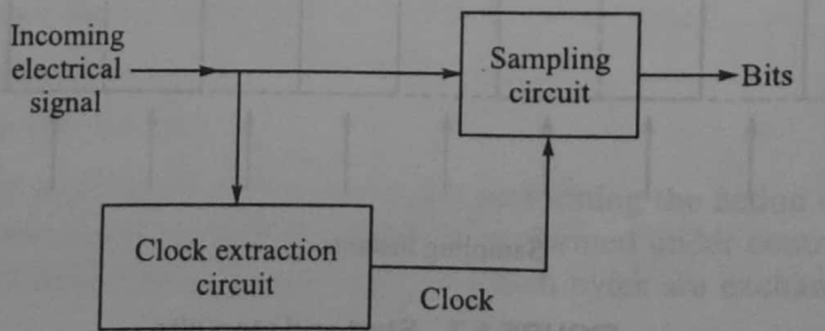

FIGURE 1.10 Bit recovery in synchronous transmission.

There is, however, still one problem. The data bytes lose their identity and therefore their boundaries need to be identified. A unique sequence of fixed number of bits, called *flag*, is prefixed to each block (Figure 1.9). The flag identifies the start of a block. Receiver first detects the flag and then identifies the boundaries of different data bytes using a counter. Just after the flag, there is first bit of the first byte.

A more common term for data block is frame. A frame contains many other fields in addition to the flag and data fields. It is also possible that the length of the data field may not necessarily be a multiple of bytes. We will discuss frame structures later in Chapter 8, Data Link Layer.

4.4 MODEM

The term 'modem' is derived from the words, MOdulator and DEModulator. A *modem* contains a modulator as well as a demodulator. The digital modulation/demodulation schemes discussed above are implemented in the modems. Most of the modems are designed for utilizing the analog voice band service offered by the telecommunication network. Therefore, the modulated carrier generated by a modem fits into the 300–3400 Hz bandwidth of the speech channel.

A typical data connection set up using modems is shown in Figure 4.16. The terminal devices which exchange digital signals are called *data terminal equipments* (DTE). Two modems are always required, one at each end. The modem at the transmitting end converts the digital signal from the DTE into an analog signal by modulating a carrier. The modem at the receiving end demodulates the carrier and hands over the demodulated digital signal to the DTE.

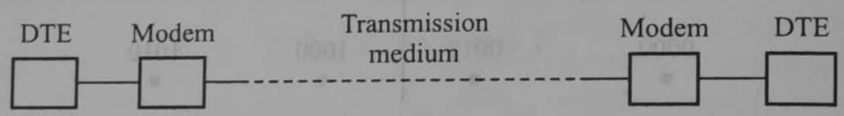

FIGURE 4.16 A data circuit implementation using modems.

The transmission medium between the two modems can be a dedicated circuit or a switched telephone circuit. In the latter case, modems are connected to the local telephone exchanges. Whenever data transmission is required, connection between the modems is established through the telephone exchanges. Modems are also required within a building to connect terminals which are located at distances usually more than 15 meters from the host.

Broadly, a modem is composed of a transmitter, a receiver, and two interfaces (Figure 4.17). The digital interface connects the modem to the DTE which generates and receives the digital signals. The line interface connects the modem to the transmission media for transmitting and receiving the modulated signals. Digital signal to be transmitted is applied to the transmitter through the digital interface. The modulated carrier that is received from the distant end at the line interface is applied to the receiver.

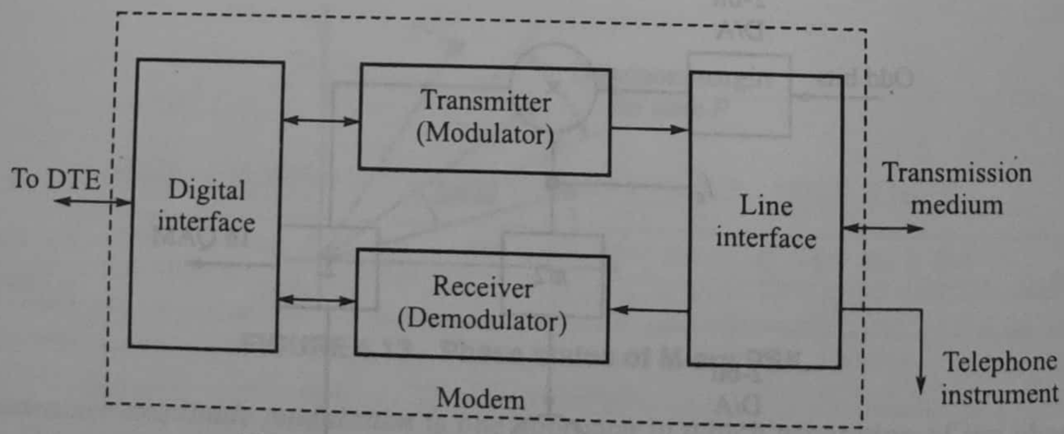

FIGURE 4.17 Building blocks of a modem.

Modems connected to telephone exchanges have additional provision for connecting a telephone instrument. The transmitter and receiver in a modem consist of several signal processing circuits which include a modulator in the transmitter and a demodulator in the receiver.

4.4.1 Types of Modems

Modems can be of several types and they can be categorized in a number of ways. Categorization is usually based on the following basic modem features:

- Directional capability—half duplex modem and full duplex modem.
- Connection to the line—2-wire modem and 4-wire modem.
- Transmission mode—asynchronous modem and synchronous modem.

Half duplex and full duplex modems. A *half duplex modem* permits transmission in one direction at a time. If a carrier is detected on the line by the modem, it gives an indication of the incoming carrier to the DTE through a control signal of its digital interface (Figure 4.18a). So long as the carrier is being received, the modem does not give clearance to the DTE to transmit.

A *full duplex modem* allows simultaneous transmission in both directions. Thus, there are two carriers on the line, one outgoing and the other incoming (Figure 4.18b).

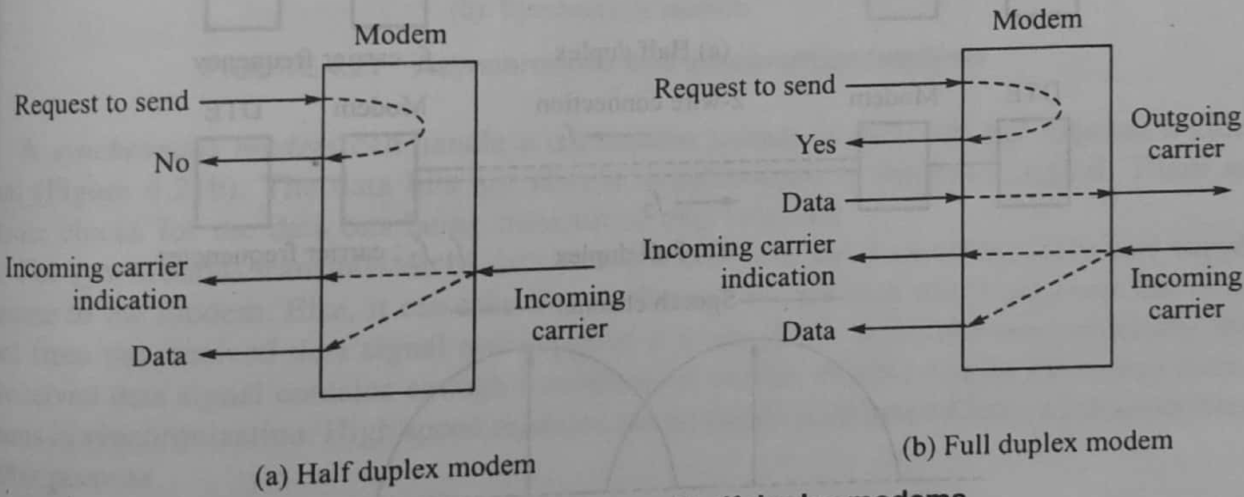

FIGURE 4.18 Full duplex and half duplex modems.

2-wire and 4-wire modems. The line interface of the modem can have a 2-wire or a 4-wire connection to transmission medium. In a 4-wire connection, one pair of wires is used for the outgoing carrier and the other pair is used for the incoming carrier (Figure 4.19). Full duplex and half duplex modes of data transmission are possible on a 4-wire connection. As the physical transmission path for each direction is separate, the same carrier frequency can be used for both the directions.

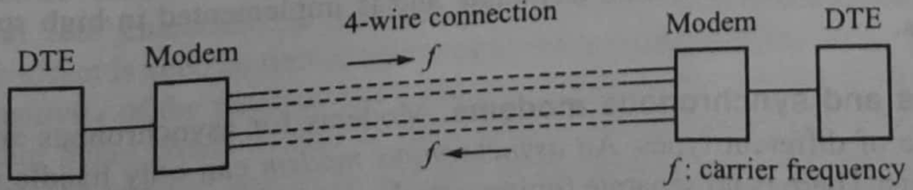

FIGURE 4.19 4-wire modems.

A leased 2-wire connection is cheaper than a 4-wire connection because only one pair of wires is extended to the subscriber's premises. The data connection established through telephone exchange is also a 2-wire connection. Modems with a 2-wire line interface are required for 2-wire connections. Such modems use the same pair of wires for outgoing and incoming carriers. Half duplex mode of transmission using the same frequency for the incoming and outgoing carriers can be easily implemented (Figure 4.20a). The transmit and receive carrier frequencies can be the same because only one of them is present on the line at a time.

For full duplex mode of operation on a 2-wire connection, it is necessary to have two transmission channels, one for the transmit direction and the other for receive direction (Figure 4.20b). This is achieved by frequency division multiplexing of two different carrier frequencies. These carriers are placed within the bandwidth of the speech channel (Figure 4.20c). A modem transmits data on one carrier and receives data from the other end on the other carrier. A hybrid is provided in the 2-wire modem to couple the line to its modulator and demodulator.

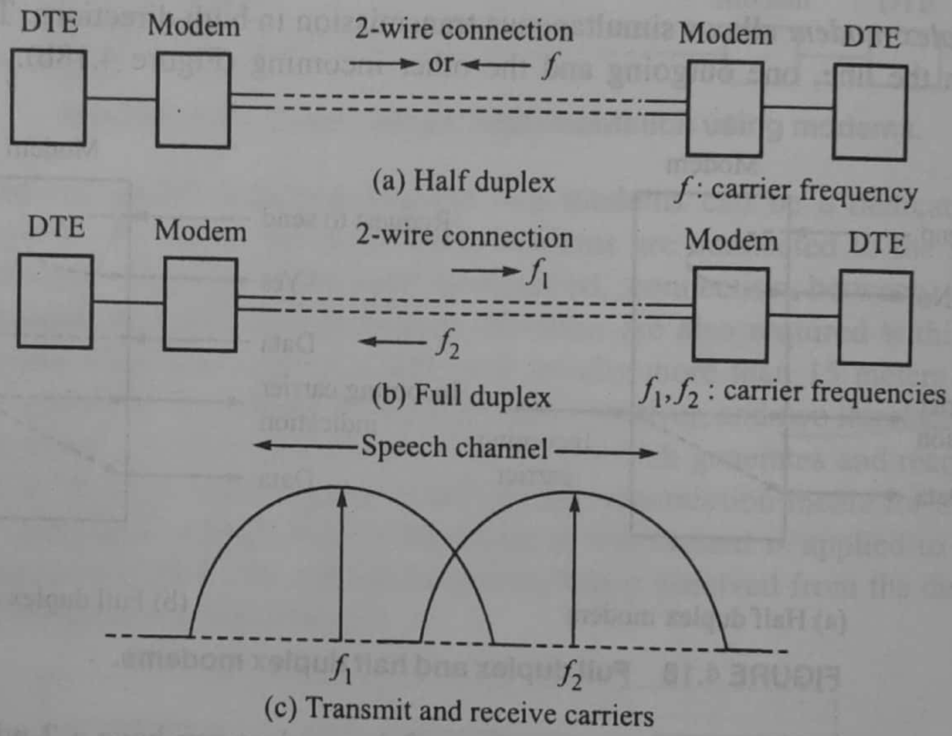

FIGURE 4.20 2-wire modems.

Note that available bandwidth for each carrier is reduced to half. Therefore, the baud rate is also reduced to half. There is a special technique which allows simultaneous transmission of incoming and outgoing carriers having the same frequency on the 2-wire transmission medium. Full bandwidth of the speech channel is made available to both the carriers simultaneously. This technique is called *echo cancellation technique* and is implemented in high speed 2-wire full duplex modems.

Asynchronous and synchronous modems. Modems for asynchronous and synchronous transmission are of different types. An *asynchronous modem* can only handle data bytes with start and stop bits. There is no separate timing signal or clock between the modem and the DTE (Figure 4.21a). The internal timing pulses are synchronized repeatedly to the leading edge of the start pulse.

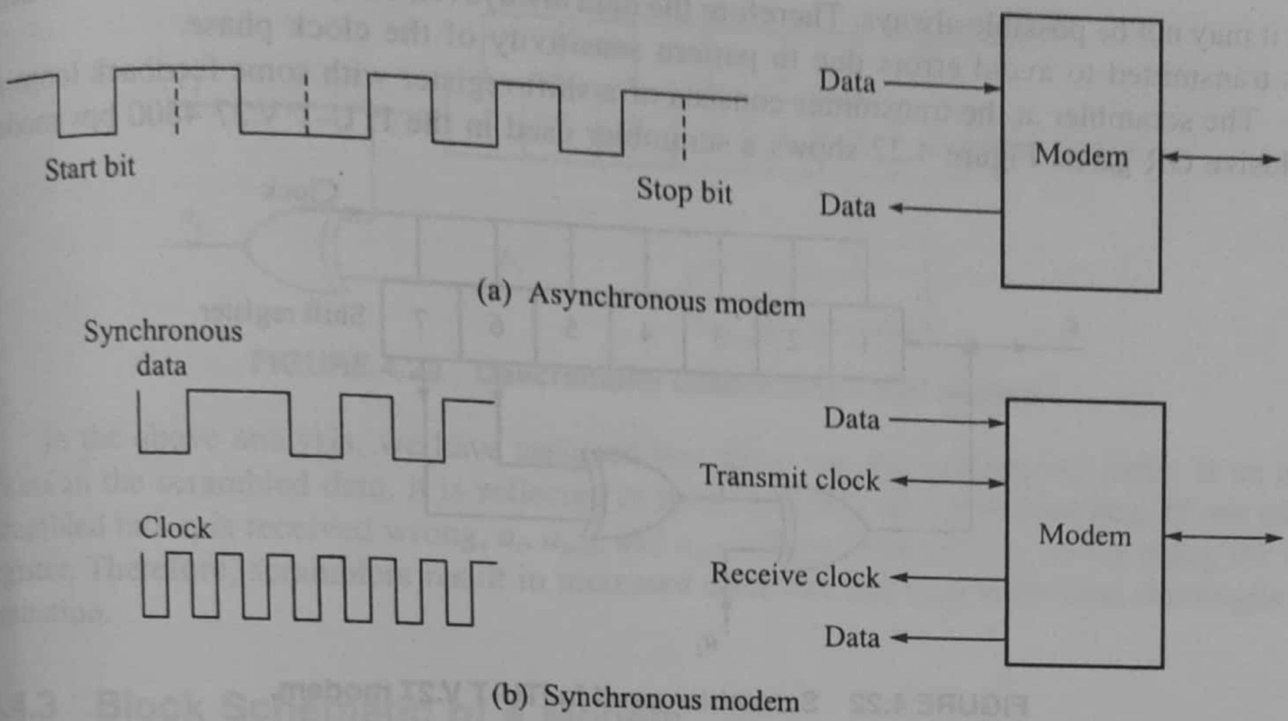

FIGURE 4.21 Asynchronous and synchronous modems.

A *synchronous modem* can handle a continuous stream of data bits but requires a clock signal (Figure 4.21b). The data bits are always synchronized to the clock signal. There are separate clocks for the data bits being transmitted and received.

For synchronous transmission of data bits, the DTE can use its internal clock and supply the same to the modem. Else, it can take the clock from the modem which recovers the clock signal from the received data signal and supplies it to the DTE. It is, however, necessary that the received data signal contains enough transitions to ensure that the timing extraction circuit remains in synchronization. High speed modems are equipped with scramblers and descramblers for this purpose.

4.4.2 Scrambler and Descrambler

As mentioned above, it is essential to have sufficient transitions in the transmitted data for clock extraction. A scrambler is provided in the transmitter to ensure this. It uses an algorithm to change the data stream received from the terminal in a controlled way so that a continuous stream of zeros or ones is avoided. The scrambled data is descrambled at the receiving end using a complementary algorithm.

There is another reason for using scramblers. It is often seen in data communications that computers transmit 'idle' characters for relatively long periods of time and then there is a sudden burst of data. The effect is seen as repeating errors at the beginning of the data. The reason for these errors is sensitivity of the receiver clock phase to certain data patterns. If the transmission line has poor group delay characteristic in some part of the spectrum and the repeated data pattern concentrates the spectral energy in that part of the spectrum, the recovered clock phase can be offset from its mean position. Drifted clock phase results in errors when the data bits are regenerated. This problem can be overcome by properly equalizing the transmission line.

But it may not be possible always. Therefore the data always randomized using scrambler before it is transmitted to avoid errors due to pattern sensitivity of the clock phase.

The scrambler at the transmitter consists of a shift register with some feedback loops and exclusive OR gates. Figure 4.22 shows a scrambler used in the ITU-T V.27 4800 bps modem.

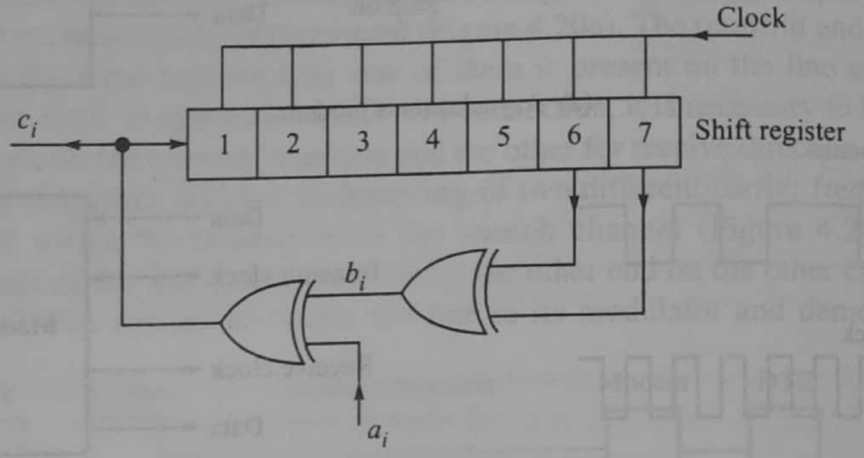

FIGURE 4.22 Scrambler used in ITU-T V.27 modem.

The output c_i (Figure 4.22) can be written as

$$c_i = a_i + b_i = a_i + c_{i-6} + c_{i-7}$$

If we represent one-bit delay using a delay operator x^{-1}, the above equation can be rewritten as follows:

$$c_i = a_i + c_i\,(x^{-6} + x^{-7})$$

or

$$c_i = a_i/(1 + x^{-6} + x^{-7})$$

Note that addition and subtraction operations are the same in modulo-2 arithmetic. Thus, a scrambler effectively divides the input data stream by polynomial $1 + x^{-6} + x^{-7}$. This polynomial is called the *generating polynomial*. By proper choice of the polynomial, it can be assured that undesirable bit sequences are avoided at the output. The generating polynomials recommended by ITU-T for scramblers are given in Table 4.5.

TABLE 4.5 ITU-T Generating Polynomials

ITU-T recommendations	Generating polynomial
V.22, V.22*bis*	$1 + x^{-14} + x^{-17}$
V.27	$1 + x^{-6} + x^{-7}$
V.26*ter*, V.29, V.32	$1 + x^{-18} + x^{-23}$, $1 + x^{-5} + x^{-23}$

To get back the data sequence at the receiving end, the scrambled data stream is multiplied by the same generating polynomial. The descrambler is shown in Figure 4.23.

In this case the scrambled bit stream c_i is the input and the descrambled output is a_i, which is obtained by multiplying c_i with the generating polynomial.

$$a_i = c_i + b_i = c_i + c_{i-6} + c_{i-7} = c_i\,(1 + x^{-6} + x^{-7})$$

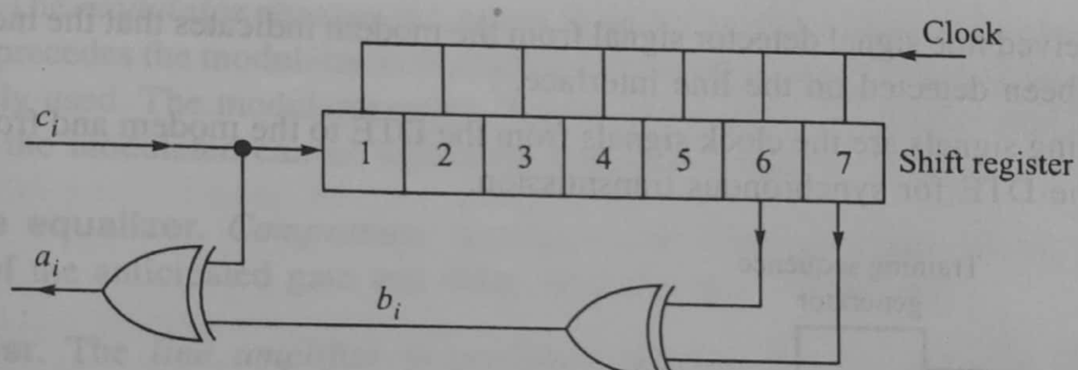

FIGURE 4.23 Descrambler used in ITU-T V.27 modem.

In the above analysis, we have assumed that there was no transmission error. If an error occurs in the scrambled data, it is reflected in three data bits after descrambling. If one of the scrambled bits c_i is received wrong, a_i, a_{i+6}, and a_{i+7} will be affected as c_i moves along the shift register. Therefore, scramblers result in increased error rate but their usefulness outweighs this limitation.

4.4.3 Block Schematic of a Modem

With this background, we can now describe the detailed block schematic of a modem. The modem design and complexity vary depending on the bit rate, type of modulation, and other basic features as discussed above. Low speed modems up to 1200 bps are asynchronous and use FSK. Medium speed modems from 2400 to 4800 bps use differential PSK. High speed modems which operate at 9600 bps and above employ QAM and are the most complex. Medium and high speed modems operate in synchronous mode of transmission.

Figure 4.24 shows important components of a typical synchronous differential PSK modem. It must, however, be borne in mind that this design gives the general functional picture of the modem. Actual implementation will vary from vendor to vendor.

11

IEEE 802.3 Ethernets

Ethernet is the most widely used local area network in the industry today. We have ethernets that operate at 10 gigabit/s bit rates. We examine the evolution of ethernet technology from its basics to the current state in this chapter. We begin with ALOHA, the contention access method that was the starting point of ethernet. We examine its throughput and the various back-off mechanisms used for improving the throughput. MAC frame formats of IEEE 802.3 Ethernet and Ethernet (DIX) are examined next. With this as the foundation, we move over to study of various types of ethernet LANs ranging from 10 Mbps to 1 Gbps. We examine the physical layer characteristics of each type of LAN in detail. 100 Mbps and higher bit rate LANs use auto-negotiation mechanism that enables the end stations to negotiate the operational parameters. We close the chapter with discussion on auto-negotiation mechanism.

11.1 CONTENTION ACCESS

In contention access methods there is no scheduled time or sequence for the stations to transmit the data frames on the medium. They contend for the use of the medium. It is, therefore, quite likely that more than one station will transmit simultaneously and the data frames will 'collide'. There are several ways of reducing the likelihood of these collisions. Carrier Sense Multiple Access/Collision Detection (CSMA/CD) is the most commonly used contention access method used in the local area networks. To understand this method, we must start from the first contention access method—Pure ALOHA, and go through its various variants.

11.1.1 Pure ALOHA

ALOHA contention access mechanism derives its name from its first implementation in ALOHA Packet Radio network that was built in 1970s to connect the various campuses of University of Hawaii. The network was based on a single radio channel for several stations to communicate with each other. There have been several variants of ALOHA subsequently. The original contention access mechanism is, therefore, referred to as Pure ALOHA. The basic scheme is as follows:

- All the stations share a common radio channel for transmitting their data frames.
- A station can transmit its data frame on the radio channel whenever it wants. There is no pre-assigned time or sequence in which the stations transmit.
- When a transmission is in progress, if another station initiates its transmission, collision (overlapping) of the two transmissions occurs. In other words, the two transmissions corrupt each other.
- A mechanism to detect collision is established (e.g. carrier detection). When a transmission is in progress, if another carrier is detected, collision is assumed to have occurred and the data frame is retransmitted.

Figure 11.1 shows transmission of data frames by four stations. Data frames A, B, D, E, and F get corrupted during transmission due to collisions with other frames. But there are instances when there is single transmission on the channel and it reaches the destination without any collision. Data frames C, G, H, and I are successfully transmitted.

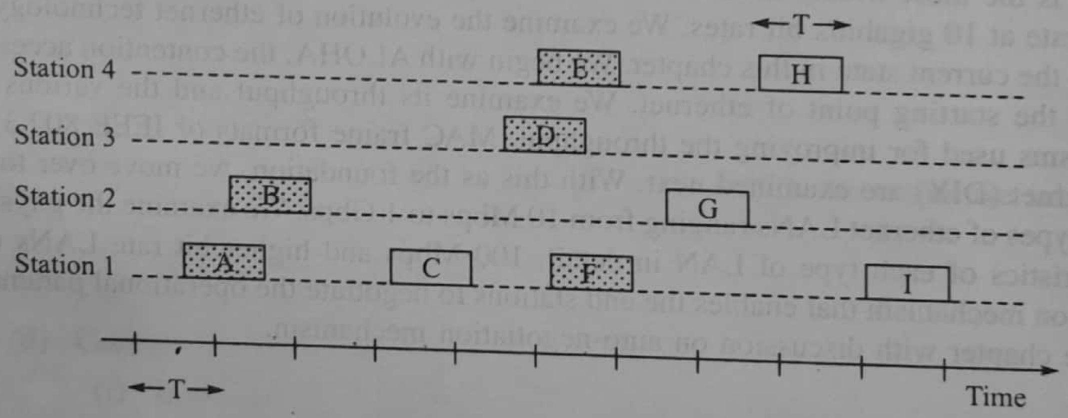

FIGURE 11.1 Frame transmission in pure ALOHA.

11.1.2 Throughput of Pure ALOHA Channel

Throughput S of a channel is defined as average number of successful transmission of the data frames on the channel per unit time. It is usually expressed as percentage of carrying capacity of the channel. To calculate the throughput of pure ALOHA channel, let us consider a simple communication model. There are N stations that send data frames to the base station on using a shared communication channel (Figure 11.2).

We assume that

- all the data frames have the same size and each frame takes time T to transmit, and
- each station generates frames independent of other stations.

Transmit time T is frame size divided by bit rate. For the sake of convenience, we take the time to transmit a frame (T) as our unit of time. Clearly, we can send on the channel at the

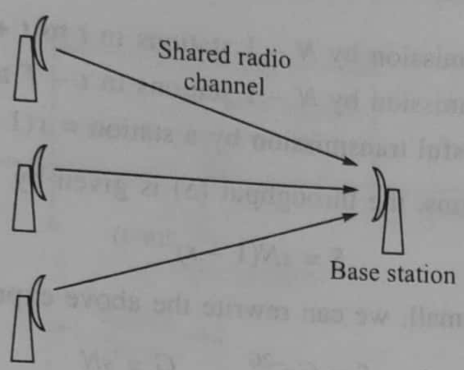

FIGURE 11.2 ALOHA network.

most one frame in time T. Therefore, the channel capacity is one frame per time unit (T). When collisions occur, some of the transmissions are lost and part of available channel time is wasted. This results in the value of throughput S always less than one. For example, throughput is equal to $4/10 = 0.4$ in Figure 11.1, as only four frames are successfully transmitted in time $10T$.

The average number of frames generated in the network in time T by the stations is called load (G). For example, in Figure 11.1, G is $9/10 = 0.9$ as nine frames are generated in time $10T$. G can have any value depending on number of stations and how frequently they generate the frames. Since there is only one channel having capacity equal to one, there are many collisions when G is more than 1. When G is low, much less than 1, there are few collisions and $S \approx G$.

Let us assume that probability of generating a data frame by a station in time T is s. Therefore, the average number of frames generated by the N stations in time T is given by $G = sN$.

The channel time is wasted whenever there is overlap of two frames. Overlap of any amount is always fatal and results in wasted channel time. Let us assume that a station generates a frame at time t. For successful transmission of the frame generated by the station at time t, it is necessary that there should not be any other frame after time $t - T$ and up to time $t + T$ (Figure 11.3).

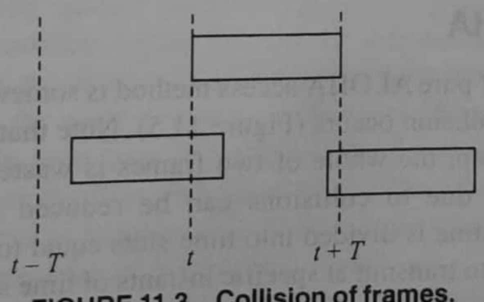

FIGURE 11.3 Collision of frames.

We can calculate the probability of a successful transmission by a station at time t as follows:

Probability of no transmission by a station in $t - T$ to $t = (1 - s)$

Probability of no transmission by $N - 1$ stations in $t - T$ to $t = (1 - s)^{N-1}$

Probability of no transmission by $N - 1$ stations in t to $t + T = (1 - s)^{N-1}$

Probability of no transmission by $N - 1$ stations in $t - T$ to $t + T = (1 - s)^{2(N-1)}$

Probability of a successful transmission by a station $= s(1 - s)^{2(N-1)}$

Since there are N stations, the throughput (S) is given by

$$S = sN(1 - s)^{2(N-1)}$$

If N is large and s is small, we can rewrite the above expression for throughput as

$$S = Ge^{-2G}, \qquad G = sN$$

The plot of throughput S with respect to load G is shown in Figure 11.4. The maximum throughput occurs at $G = 0.5$ and is equal to 0.184. This simply means that maximum throughput occurs when all the stations together generate on average one frame in time interval of $2T$. When frames are generated at this rate, only 18.4% of these frames are successfully transmitted.

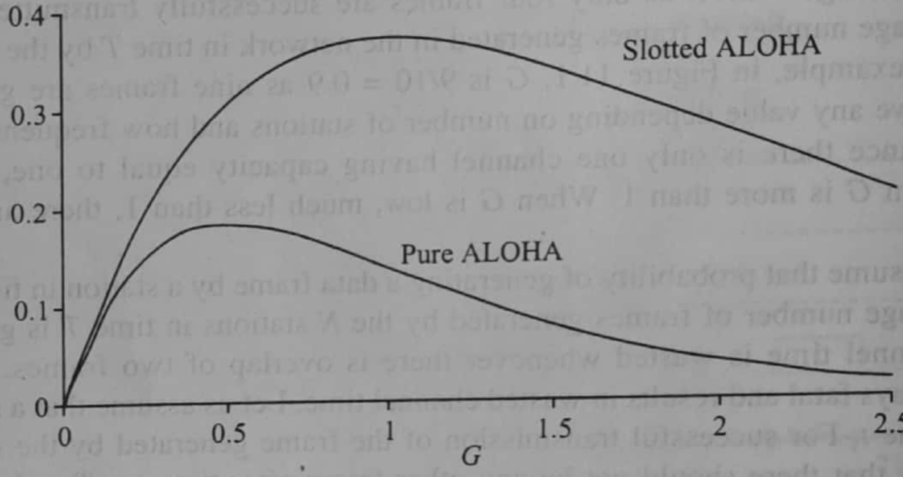

FIGURE 11.4 Thoughput of pure and slotted ALOHA.

11.1.3 Slotted ALOHA

The maximum throughput of pure ALOHA access method is somewhat modest, the reason being large wasted time when a collision occurs (Figure 11.5). Note that even though there may have been only few bits of overlap, the whole of two frames is wasted in one collision.

Wasted channel time due to collisions can be reduced if all the transmissions are synchronized. The channel time is divided into time slots equal to time to transmit a frame (T) and the stations are allowed to transmit at specific instants of time so that all transmissions arrive aligned with the time slot boundaries (Figure 11.5). Collisions will still occur but the wasted time channel time is reduced to one time slot.

In slotted ALOHA, collision can take place only if there are more than one frame in a time slot. Since the probability of having a single frame in a time slot is $s(1 - s)^{N-1}$, the throughput S of slotted ALOHA for N stations is given by

$$S = sN(1 - s)^{N-1}$$

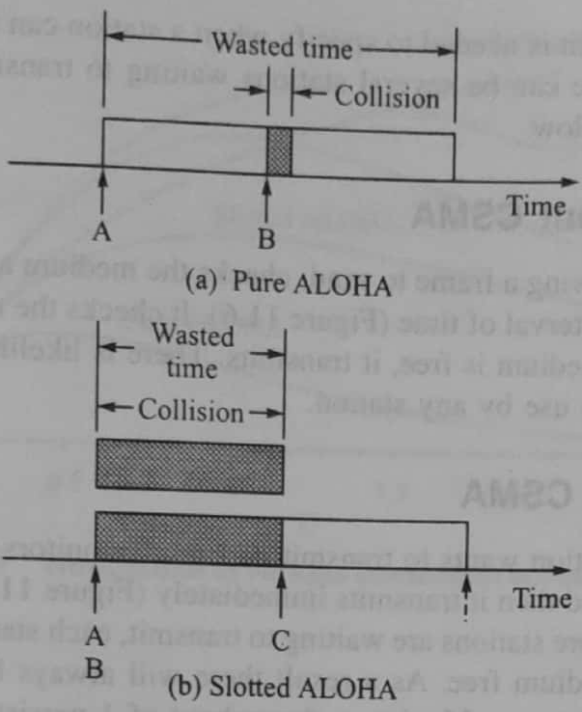

FIGURE 11.5 Wasted time due to collision.

If N is large and s is small, the throughput can be rewritten as

$$S = Ge^{-G}, \qquad G = sN$$

Plot of S with respect to G for slotted ALOHA is shown in Figure 11.4. In this case the maximum throughput occurs at $G = 1$ and its value is 0.368.

11.2 CARRIER SENSE MULTIPLE ACCESS (CSMA)

In the ALOHA channel discussed above, the possibility of collision can be reduced if some discipline is built into the totally random access mechanism. If a station checks for the presence of any other carrier on the medium before starting its own transmission, a collision can be avoided. If there is a carrier on the channel, it does not commence its transmission. Carrier Sense Multiple Access (CSMA), as the name suggests, is based on this principle.

CSMA is widely used in local area networks. In the discussion that follows, we will be using the term carrier despite the fact that most of the local area networks use baseband transmission instead of a carrier to transmit data frames. In baseband LANs, the term 'sensing carrier' connotes 'detecting presence of a baseband signal'.

When a station has a frame to send and the transmission medium is free, it starts to transmit its frame. Other stations wanting to transmit their frames soon sense presence of this frame as the signal travels along the medium, and they defer their transmissions to avoid collisions. CSMA does not avoid collisions altogether. Collision can still occur if a station commences its transmission before the first bit already transmitted by another station is heard by it. This happens because of the propagation delay. Provided propagation delay is low, CSMA is much better than ALOHA as we shall shortly see.

In CSMA, an algorithm is needed to specify when a station can transmit once the channel is found busy because there can be several stations waiting to transmit. There can be several approaches as described below.

11.2.1 Non-Persistent CSMA

In this scheme, a station having a frame to send, checks the medium and, if the medium is busy, it backs off for a random interval of time (Figure 11.6). It checks the medium again after expiry of the interval and if the medium is free, it transmits. There is likelihood of some wasted time when the channel is not in use by any station.

11.2.2 1-Persistent CSMA

In this scheme, when a station wants to transmit its frame, monitors the medium continuously until the medium is free and then it transmits immediately (Figure 11.6). The problem with this strategy is that if two or more stations are waiting to transmit, each station will transmit its frame as soon as it finds the medium free. As a result there will always be multiple frames on the medium and collision will occur. Maximum throughput of 1-persistent CSMA is 0.53.

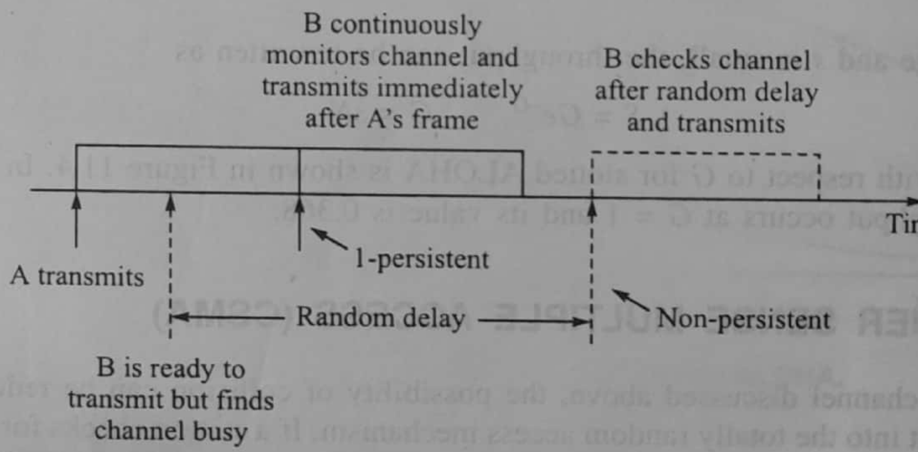

FIGURE 11.6 CSMA variants.

11.2.3 *p*-Persistent CSMA

To reduce the probability of collision in 1-persistent CSMA, not all the waiting stations are allowed to transmit immediately after the medium is free. A waiting station transmits with probability p if the medium is free. For example, if $p = 1/6$ and there are six stations waiting to transmit, on average only one of them will transmit and the rest will continue to wait. It is equivalent to throwing a dice and if a station gets six, it transmits. If two stations get six, then both will transmit and collision will take place. Likelihood of such occurrences can be reduced by reducing the transmission probability p. Optimized p-persistent CSMA can give maximum throughput of 0.8–0.9.

Figure 11.7 compares throughputs of various contention access schemes. We see that CSMA is always better than ALOHA.

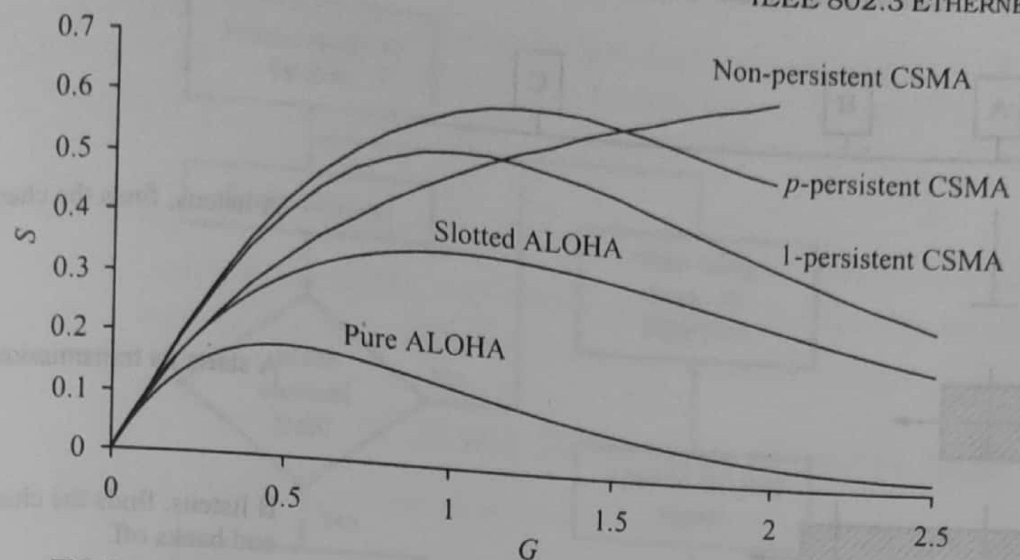

FIGURE 11.7 Throughput of various contention access methods.

11.3 CSMA/CD

The most commonly deployed multiple access technique in the local area networks is CSMA/CD, where CD stands for Collision Detection. CSMA/CD specifications were developed jointly by DEC, Intel, and Xerox (DIX) in 1980. They called this network *Ethernet.* These specifications were later adopted by IEEE as their standard IEEE 802.3 in 1985. There are some differences in IEEE 802.3 and Ethernet (DIX) specifications as we shall see shortly. These differences make them incompatible to each other.

Throughout the rest of this chapter, the term *Ethernet* refers to the network compatible to IEEE 802.3. Ethernet developed by DEC, Intel, and Xerox will be referred to as Ethernet (DIX).

11.3.1 Media Access Control in CSMA/CD

In the CSMA technique discussed above, a station continues transmission of a frame until the end of the frame even if a collision occurs. Continuing transmission of a frame that has already suffered collision results in unnecessary wastage of channel time. In CSMA/CD, transmission of a frame is abandoned as soon as collision is detected and a jam signal is appended to the frame to alert the other stations. Figure 11.8 illustrates the basic operation of the scheme. Note that for the scheme to work properly it is necessary that

- a station should still be transmitting its frame when it realizes that collision has occurred,
- the stations do not attempt to transmit again immediately after a collision has occurred. Otherwise, the same frames will collide again.

The stations are given a random back off delay for retry. If collision repeats, back off delay is progressively increased. So the network adapts itself to the traffic. In Ethernet, the random back off delay for retry after collision is doubled on each retry up to 10 retries.

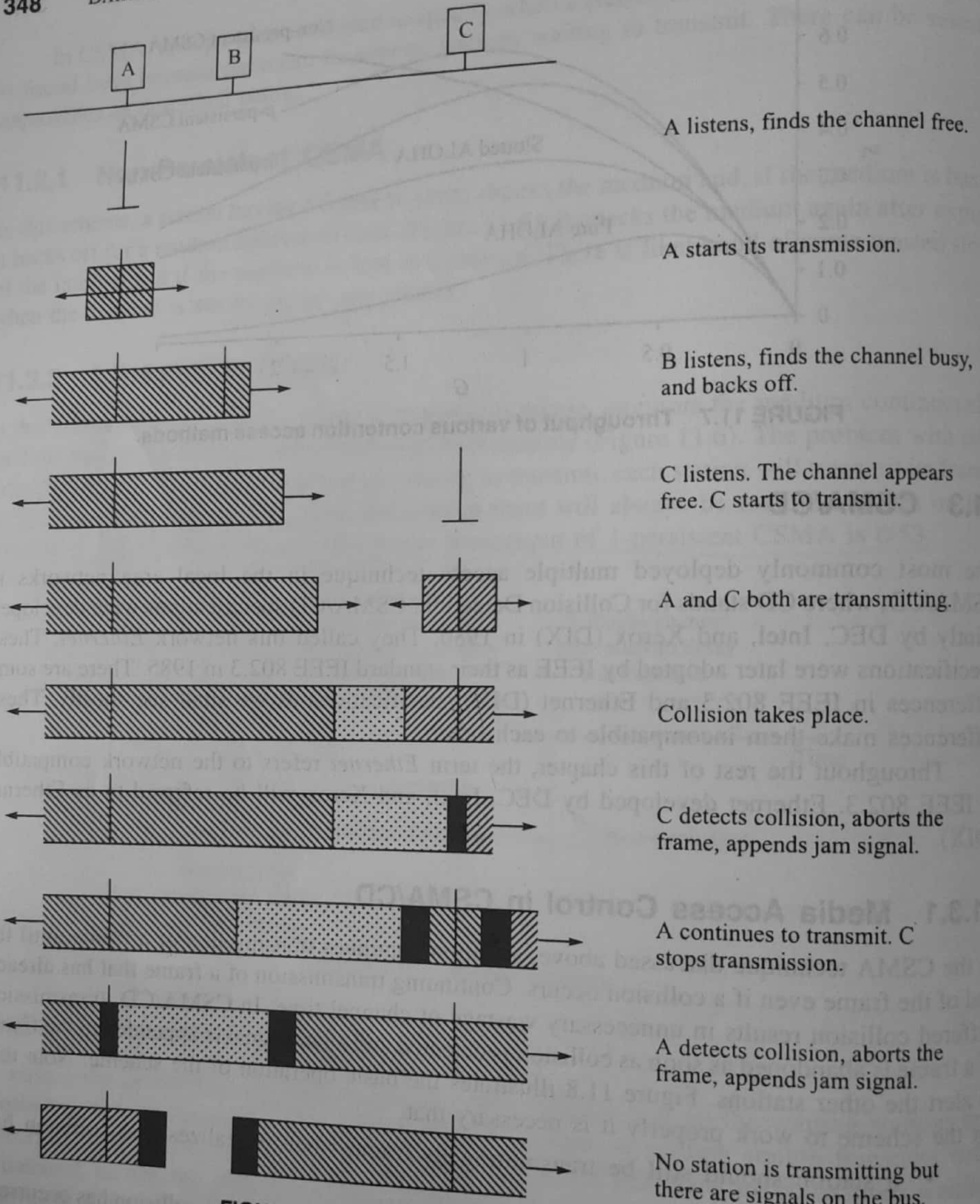

FIGURE 11.8 Collision detection in CSMA/CD.

By careful design, it is possible to achieve efficiencies of more than 90 per cent using CSMA/CD. Figure 11.9 summarizes the basic steps required for transmitting a frame.

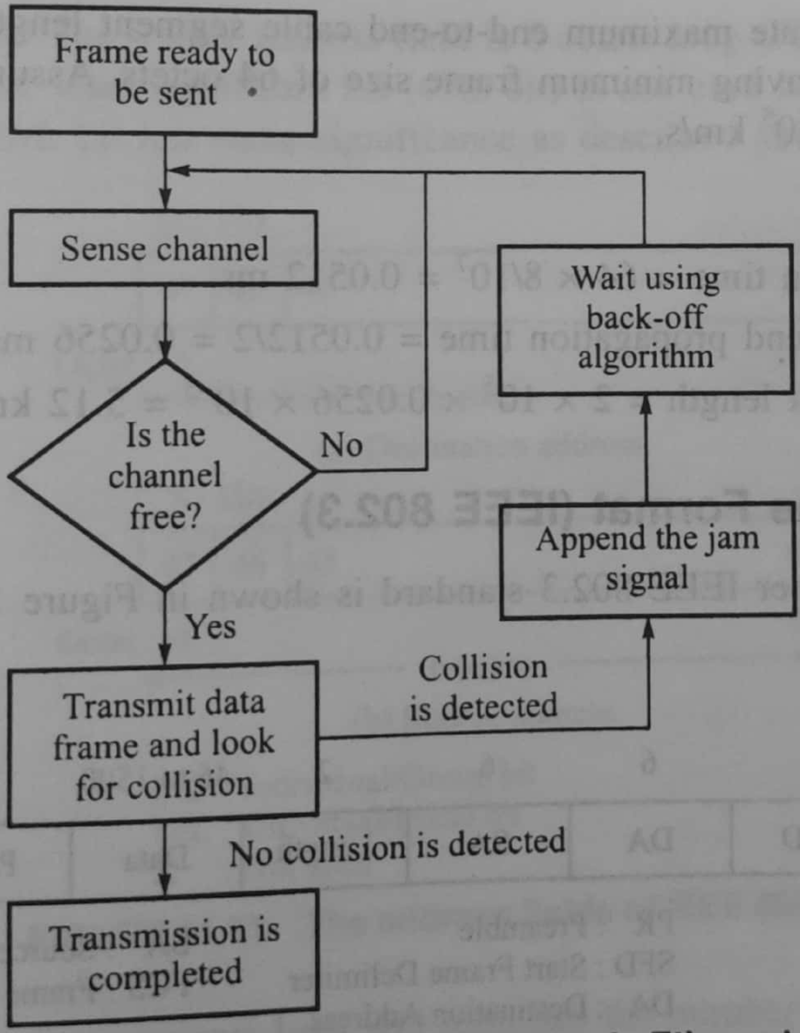

FIGURE 11.9 Frame transmission in Ethernet.

11.3.2 Maximum Cable Segment

As mentioned earlier, it is necessary that a station be transmitting if it is to detect a collision and ... This condition puts a limit on the minimum size