**INTERNET HISTORY**

Now that we have given an overview of the Internet, let us give a brief history of the internet. This brief history makes it clear how the Internet has evolved from a private network to a global one in less than 40 years.

**Early History**

There were some communication networks, such as telegraph and telephone networks, before 1960. These networks were suitable for constant-rate communication at that time, which

means that after a connection was made between two users, the encoded message (telegraphy) or voice (telephony) could be exchanged.

**ARPANET**

In the mid-1960s, mainframe computers in research organizations were stand-alone

devices. Computers from different manufacturers were unable to communicate with one another.

The Advanced Research Projects Agency (ARPA) in the Department of Defense (DOD) was

interested in finding a way to connect computers so that the researchers they funded could share their findings, thereby reducing costs and eliminating duplication of effort. In 1967, at an Association for Computing Machinery (ACM) meeting, ARPA presented its ideas for the Advanced Research Projects Agency Network (ARPANET), a small network of connected computers. The idea was that each host computer (not necessarily from the same manufacturer)

would be attached to a specialized computer, called an interface message processor (IMP). The

IMPs, in turn, would be connected to each other. Each IMP had to be able to communicate with

other IMPs as well as with its own attached host.

Birth of the Internet

In 1972, Vint Cerf and Bob Kahn, both of whom were part of the core ARPANET group,

collaborated on what they called the Internet ting Project. TCPI/P Cerf and Kahn's landmark

1973 paper outlined the protocols to achieve end-to-end delivery of data. This was a new version of NCP. This paper on transmission control protocol (TCP) included concepts such as

encapsulation, the datagram, and the functions of a gateway. Transmission Control Protocol

(TCP) and Internet Protocol (IP). IP would handle datagram routing while TCP would be

responsible for higher level functions such as segmentation, reassembly, and error detection. The new combination became known as TCPIIP.

MILNET

In 1983, ARPANET split into two networks: Military Network (MILNET) for military users and ARPANET for non military users.

CSNET

Another milestone in Internet history was the creation of CSNET in 1981. Computer Science Network (CSNET) was a network sponsored by the National Science Foundation (NSF).

NSFNET

With the success of CSNET, the NSF in 1986 sponsored the National Science Foundation Network (NSFNET), a backbone that connected five supercomputer centers located throughout the United States.

ANSNET

In 1991, the U.S. government decided that NSFNET was not capable of supporting the rapidly increasing Internet traffic. Three companies, IBM, Merit, and Verizon, filled the void by forming a nonprofit organization called Advanced Network & Services (ANS) to build a new,

high-speed Internet backbone called Advanced Network Services Network (ANSNET).

Internet Today

Today, we witness a rapid growth both in the infrastructure and new applications. The Internet today is a set of pier networks that provide services to the whole world. What has made the internet so popular is the invention of new applications.

World Wide Web

The 1990s saw the explosion of Internet applications due to the emergence of the World Wide Web (WWW). The Web was invented at CERN by Tim Berners-Lee. This invention has added the commercial applications to the Internet.

Multimedia

Recent developments in the multimedia applications such as voice over IP (telephony), video over IP (Skype), view sharing (YouTube), and television over IP (PPLive) has increased the number of users and the amount of time each user spends on the network.

Peer-to-Peer Applications

Peer-to-peer networking is also a new area of communication with a lot of potential.

STANDARDS AND ADMINISTRATION

In the discussion of the Internet and its protocol, we often see a reference to a standard or an administration entity. In this section, we introduce these standards and administration entities for those readers that are not familiar with them; the section can be skipped if the reader is familiar with them.

INTERNET STANDARDS

An Internet standard is a thoroughly tested specification that is useful to and adhered to by those who work with the Internet. It is a formalized regulation that must be followed. There is

a strict procedure by which a specification attains Internet standard status. A specification begins as an Internet draft. An Internet draft is a working document (a work in progress) with no official status and a six-month lifetime. Upon recommendation from the Internet authorities, a draft may be published as a Request for Comment (RFC). Each RFC is edited, assigned a number, and made available to all interested parties. RFCs go through maturity levels and are categorized according to their requirement level.

Maturity Levels

An RFC, during its lifetime, falls into one of six maturity levels: proposed standard, draft standard, Internet standard, historic, experimental, and informational. Proposed Standard. A proposed standard is a specification that is stable, well understood, and of sufficient interest to the Internet community. At this level, the specification is usually tested and implemented by

several different groups.

Draft Standard. A proposed standard is elevated to draft standard status after at least two successful independent and interoperable implementations. Barring difficulties, a draft standard, with modifications if specific problems are encountered, normally becomes an Internet standard.

Internet Standard. A draft standard reaches Internet standard status after demonstrations of successful implementation.

Historic The historic RFCs are significant from a historical perspective. They either have been superseded by later specifications or have never passed the necessary maturity levels to become an Internet standard. Experimental An RFC classified as experimental describes work related to an experimental situation that does not affect the operation of the Internet. Such an RFC should not be implemented in any functional Internet service.

Informational An RFC classified as informational contains general, historical, or tutorial information related to the Internet. It is usually written by someone in a non-Internet organization, such as a vendor.

Requirement Levels

RFCs are classified into five requirement levels: required, recommended, elective, limited use,

and not recommended. Required An RFC is labeled required if it must be implemented by all Internets systems to achieve minimum conformance. For example, IF and ICMP are required protocols. Recommended An RFC labeled recommended is not required for minimum conformance; it is recommended because of its usefulness. For example, FTP and TELNET are recommended protocols.

Elective An RFC labeled elective is not required and not recommended. However, a system can use it for its own benefit. Limited Use An RFC labeled limited use should be used only in limited situations. Most of the experimental RFCs fall under this category.

Not Recommended An RFC labeled not recommended is inappropriate for general use. Normally a historic (deprecated) RFC may fall under this category.

INTERNET ADMINISTRATION

The Internet, with its roots primarily in the research domain, has evolved and gained a

broader user base with significant commercial activity. Various groups that coordinate Internet

issues have guided this growth and development. Appendix G gives the addresses, e-rnail

addresses, and telephone numbers for some of these groups. Shows the general organization of

Internet administration. E-rnail addresses and telephone numbers for some of these groups.

Below figure shows the general organization of Internet administration. Isoc

The Internet Society (ISOC) is an international, nonprofit organization formed in 1992 to provide support for the Internet standards process. ISOC accomplishes this through maintaining and supporting other Internet administrative bodies such as lAB, IETF,IRTF, and IANA (see the following sections). ISOC also promotes research and other scholarly activities relating to the Internet.

lAB

The Internet Architecture Board (lAB) is the technical advisor to the ISOC. The main purposes of the lAB are to oversee the continuing development of the TCP/IP Protocol Suite and to serve in a technical advisory capacity to research members of the Internet community. lAB

accomplishes this through its two primary components, the Internet Engineering Task Force

(IETF) and the Internet Research Task Force (IRTF). Another responsibility of the lAB is the

editorial management of the RFCs, described earlier. lAB is also the external liaison between the Internet and other standards organizations and forums.

JETF

The Internet Engineering Task Force (IETF) is a forum of working groups managed by the Internet Engineering Steering Group (IESG). IETF is responsible for identifying operational

problems and proposing solutions to these problems. IETF also develops and reviews

specifications intended as Internet standards. The working groups are collected into areas, and each area concentrates on a specific topic. Currently nine areas have been defined. The areas include applications, protocols, routing, network management next generation (lPng), and

security.

JRTF

The Internet Research Task Force (IRTF) is a forum of working groups managed by the Internet Research Steering Group (IRSG). IRTF focuses on long-term research topics related to Internet protocols, applications, architecture, and technology.

**COMPARISION OF OSI AND TCP/IP REFERENCE MODEL**

When we compare the two models, we find that two layers, session and presentation, are

missing from the TCP/IP protocol suite. These two layers were not added to the TCP/IP protocol suite after the publication of the OSI model. The application layer in the suite is usually considered to be the combination of three layers in the OSI model.

Two reasons were mentioned for this decision. First, TCP/IP has more than one transport-layer protocol. Some of the functionalities of the session layer are available in some of

the transport-layer protocols. Second, the application layer is not only one piece of software.

Many Applications can be developed at this layer. If some of the functionalities mentioned in the session and presentation layers are needed for a particular application, they can be included in the development of that piece of software.

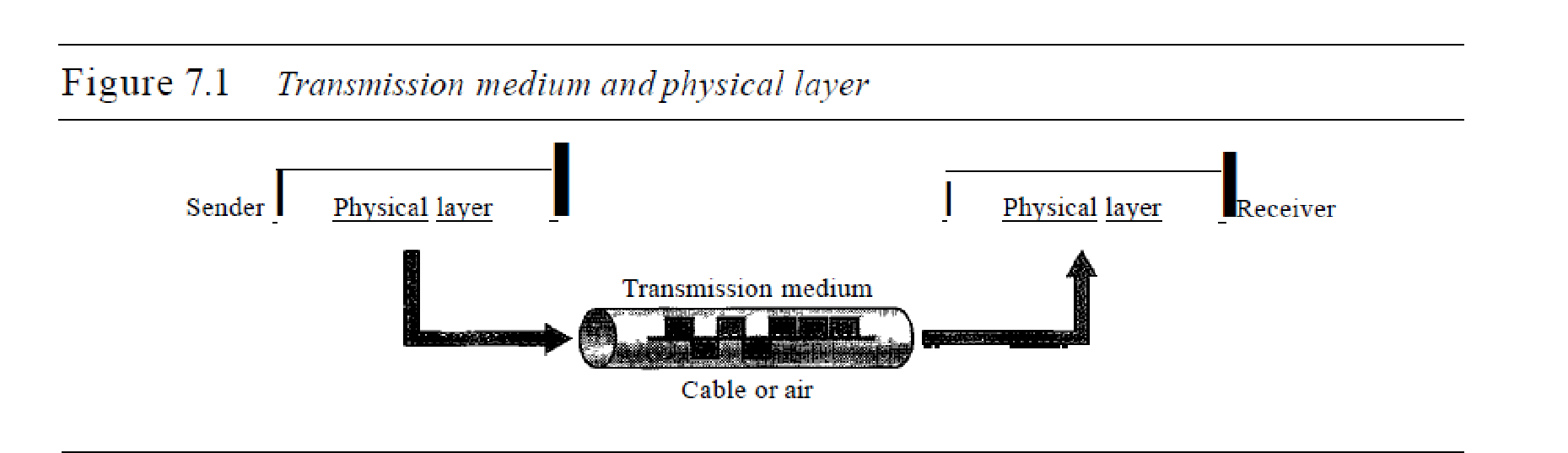
Lack of OSI Model's Success

The OSI model appeared after the TCP/IP protocol suite. Most experts were at first excited and thought that the TCP/IP protocol would be fully replaced by the OSI model. This did not happen for several reasons, but we describe only three, which are agreed upon by all experts in the field. First, OSI was completed when TCP/IP was fully in place and a lot of time and money had been spent on the suite; changing it would cost a lot. Second, some layers in the OSI model were never fully defined. For example, although the services provided by the presentation and the session layers were listed in the document, actual protocols for these two layers were not fully defined, nor were they fully described, and the corresponding software was not fully developed. Third, when OSI was implemented by an organization in a different application, it did not show a high enough level of performance to entice the Internet authority to switch from the TCP/IP protocol suite to the OSI model.

PHYSICAL LAYER

One of the major functions of the physical layer is to move data in the form of electromagnetic signals across a transmission medium. Whether you are collecting numerical statistics from another computer, sending animated pictures from a design workstation, or causing a bell to ring at a distant control center, you are working with the transmission of data across network connections. Generally, the data usable to a person or application are not in a form that can be transmitted over a network. For example, a photograph must first be changed to a form that transmission media can accept. Transmission media work by conducting energy along a physical path. For transmission, data needs to be changed to signals.

**Dr.Renukadevi M.N:** **Physical Layer**: Introduction to Guided transmission media and wireless transmission media.



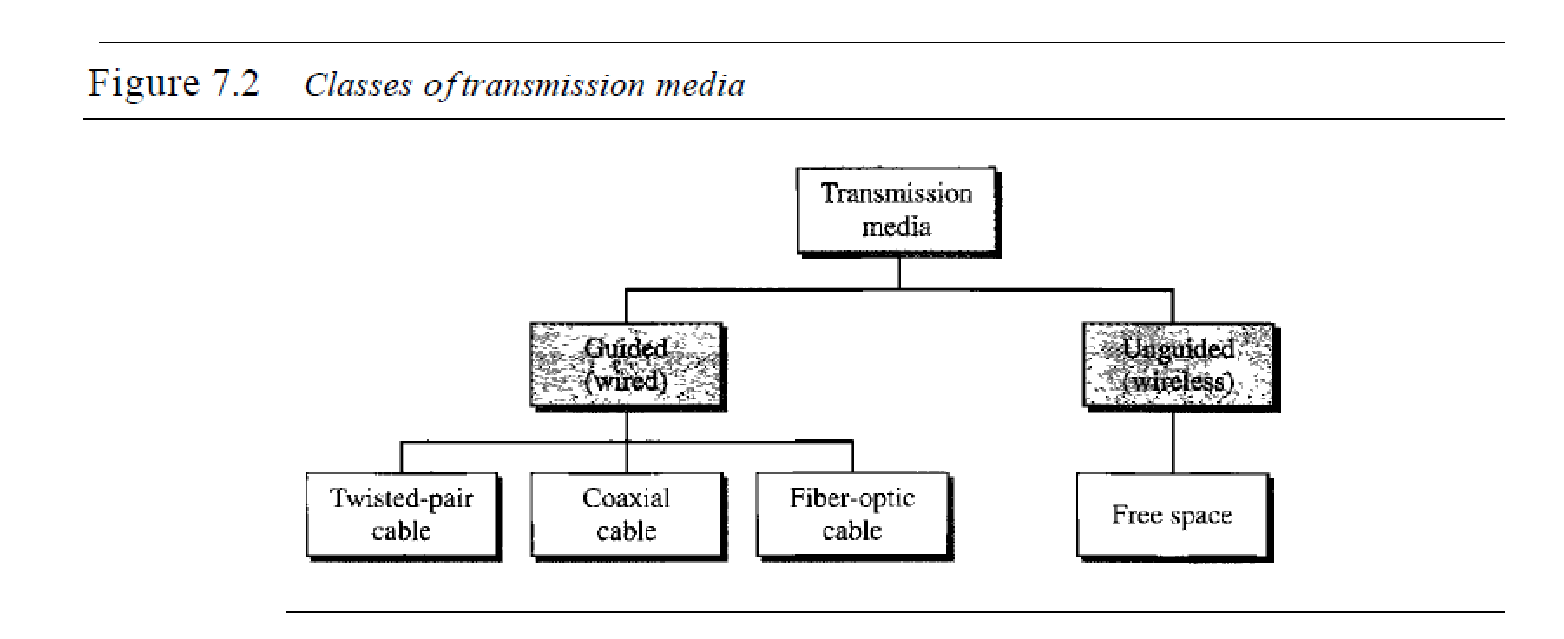
A transmission **medium** can be broadly defined as anything that can carry information

from a source to a destination

In telecommunications, transmission media can be divided into two broad categories:

guided and unguided. Guided media include twisted-pair cable, coaxial cable, and

fiber-optic cable. Unguided medium is free space. Figure 7.2 shows this taxonomy.



**GUIDED MEDIA**

Guided media, which are those that provide a conduit from one device to another,

include twisted-pair cable, coaxial cable, and fiber-optic cable. A signal traveling

along any of these media is directed and contained by the physical limits of the

medium. Twisted-pair and coaxial cable use metallic (copper) conductors that accept

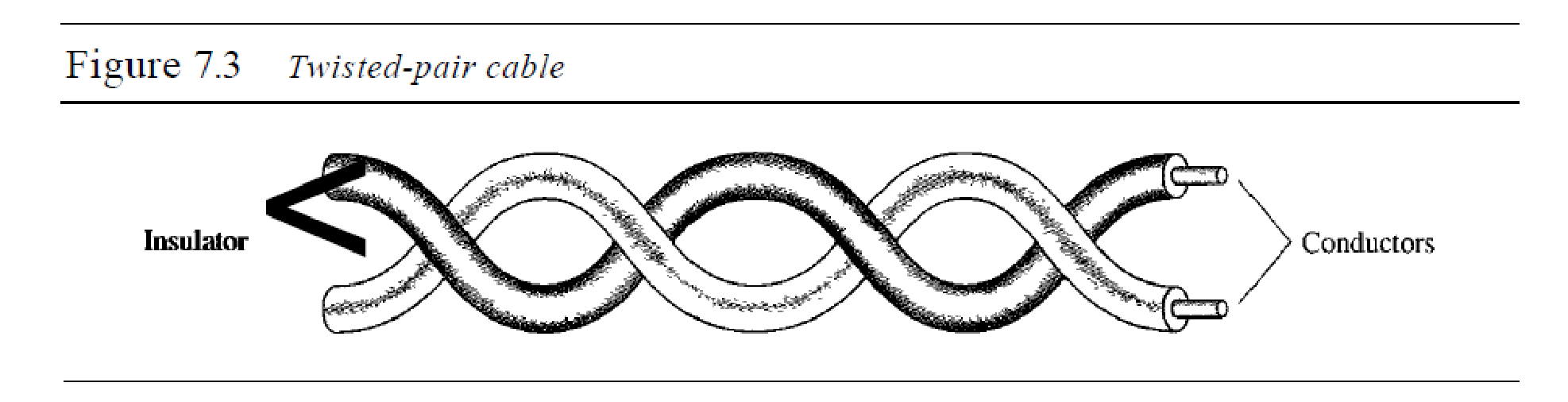
and transport signals in the form of electric current. Optical fiber is a cable that accepts

and transports signals in the form of light.

Twisted-Pair Cable

A twisted pair consists of two conductors (normally copper), each with its own plastic

insulation, twisted together, as shown in Figure 7.3



One of the wires is used to carry signals to the receiver, and the other is used only

as a ground reference. The receiver uses the difference between the two.

In addition to the signal sent by the sender on one of the wires, interference (noise)

and crosstalk may affect both wires and create unwanted signals.

If the two wires are parallel, the effect of these unwanted signals is not the same in

both wires because they are at different locations relative to the noise or crosstalk sources

(e,g., one is closer and the other is farther). This results in a difference at the receiver. By

twist,ing the pairs, a balance is maintained. For example, suppose in one twist, one wire

is closer to the noise source and the other is farther; in the next twist, the reverse is true.

Twisting makes it probable that both wires are equally affected by external influences

(noise or crosstalk). This means that the receiver, which calculates the difference between

the two, receives no unwanted signals. The unwanted signals are mostly canceled out.

From the above discussion, it is clear that the number of twists per unit of length

(e.g., inch) has some effect on the quality of the cable.

*Unshielded Versus Shielded Twisted-Pair Cable*

The most common twisted-pair cable used in communications is referred to as

unshielded twisted-pair (UTP). IBM has also produced a version of twisted-pair cable

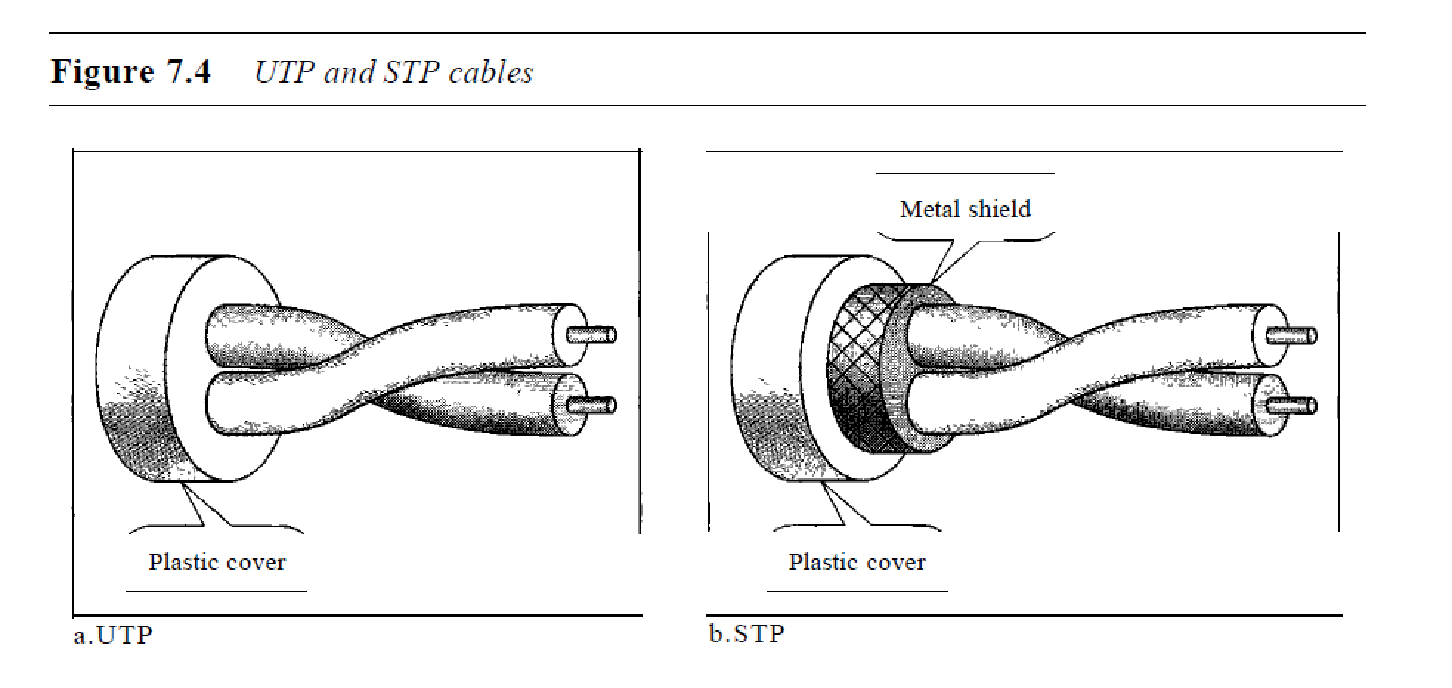
for its use called shielded twisted-pair (STP). STP cable has a metal foil or braidedmesh

covering that encases each pair of insulated conductors. Although metal casing

improves the quality of cable by preventing the penetration of noise or crosstalk, it is

bulkier and more expensive. Figure 7.4 shows the difference between UTP and STP.

Our discussion focuses primarily on UTP because STP is seldom used outside of IBM.



*Categories*

The Electronic Industries Association (EIA) has developed standards to classify

unshielded twisted-pair cable into seven categories. Categories are determined by cable

quality, with 1 as the lowest and 7 as the highest. Each EIA category is suitable for

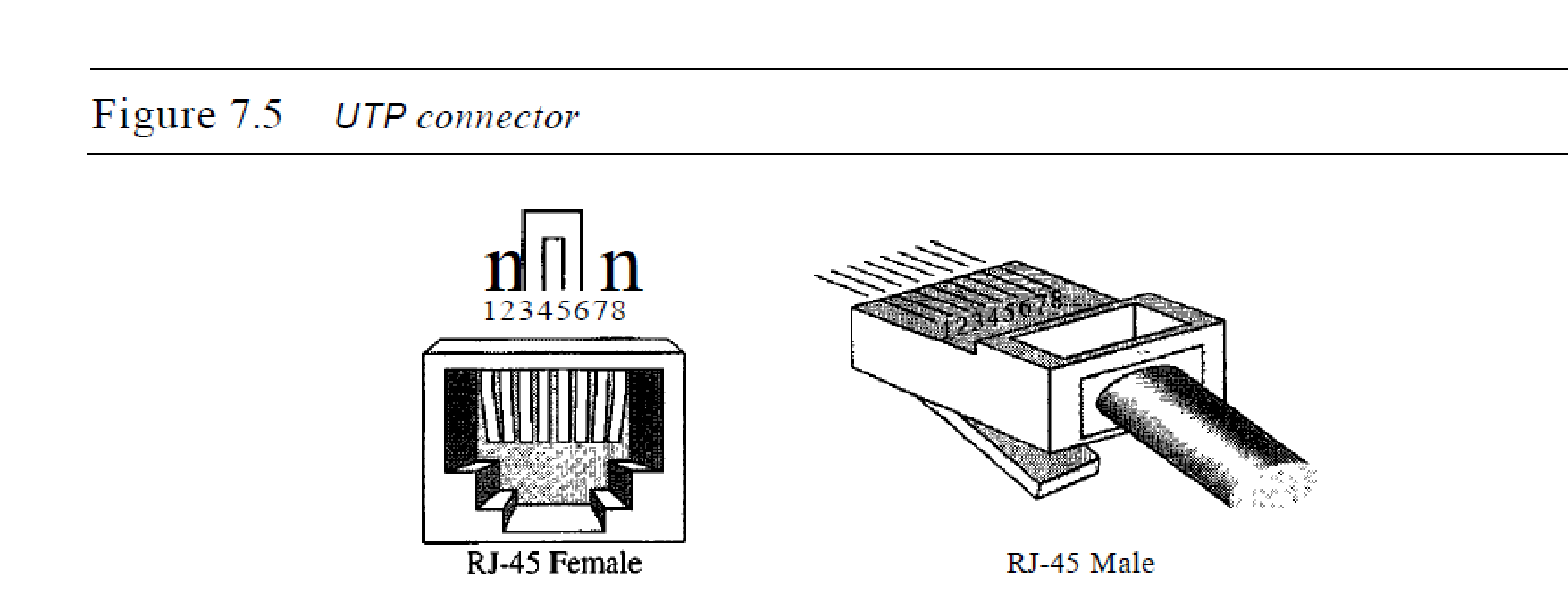
specific uses. Table 7. I shows these categories.

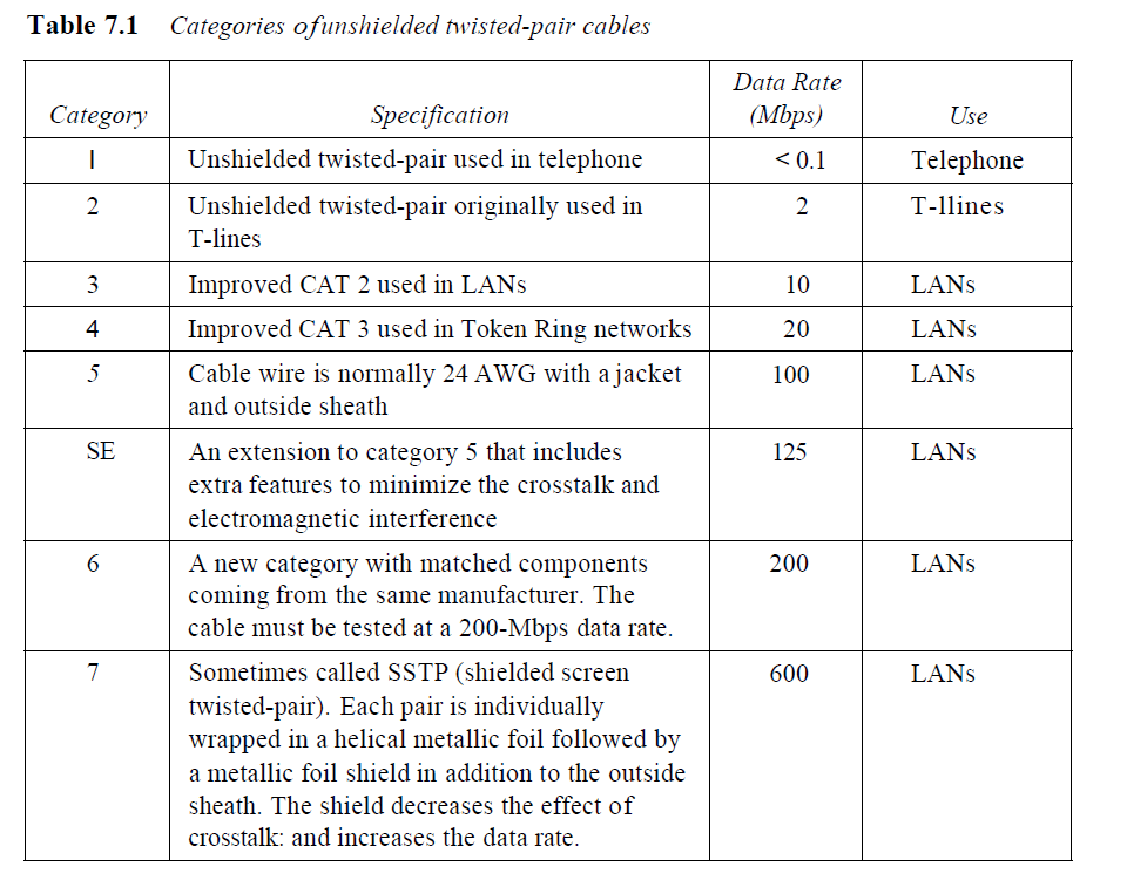
*Connectors*

The most common UTP connector is RJ45 (RJ stands for registered jack), as shown

in Figure 7.5. The RJ45 is a keyed connector, meaning the connector can be inserted in

only one way.





*Performance*

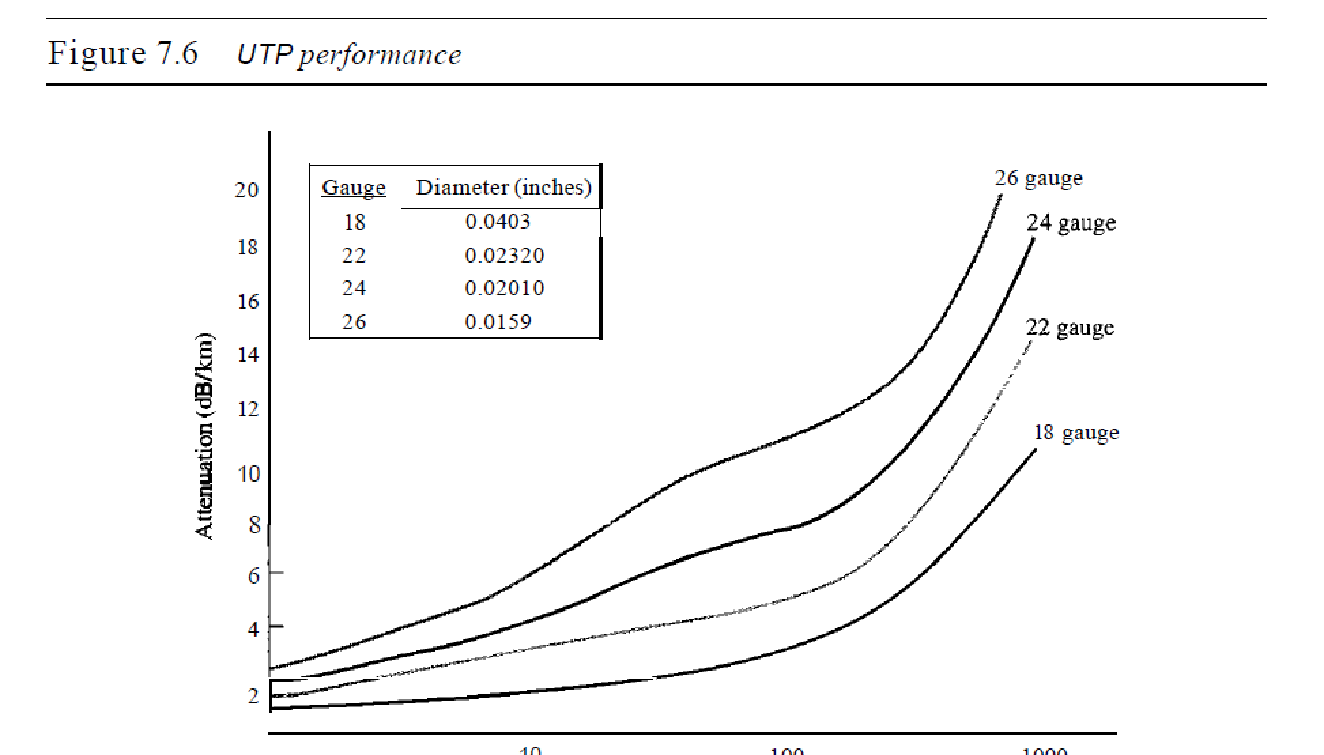
One way to measure the performance of twisted-pair cable is to compare attenuation

versus frequency and distance. A twisted-pair cable can pass a wide range of frequencies.

However, Figure 7.6 shows that with increasing frequency, the attenuation, measured in

decibels per kilometer (dB/km), sharply increases with frequencies above 100 kHz. Note

that *gauge* is a measure of the thickness of the wire.



*Applications*

Twisted-pair cables are used in telephone lines to provide voice and data channels. The

local loop-the line that connects subscribers to the central telephone office---commonly

consists of unshielded twisted-pair cables. We discuss telephone networks in Chapter 9.

The DSL lines that are used by the telephone companies to provide high-data-rate

connections also use the high-bandwidth capability of unshielded twisted-pair cables.

We discuss DSL technology in Chapter 9.

Local-area networks, such as lOBase-T and lOOBase-T, also use twisted-pair cables.

We discuss these networks in Chapter 13.

**Coaxial Cable**

Coaxial cable (or *coax)* carries signals of higher frequency ranges than those in twistedpair

cable, in part because the two media are constructed quite differently. Instead of having two wires, coax has a central core conductor of solid or stranded wire (usually

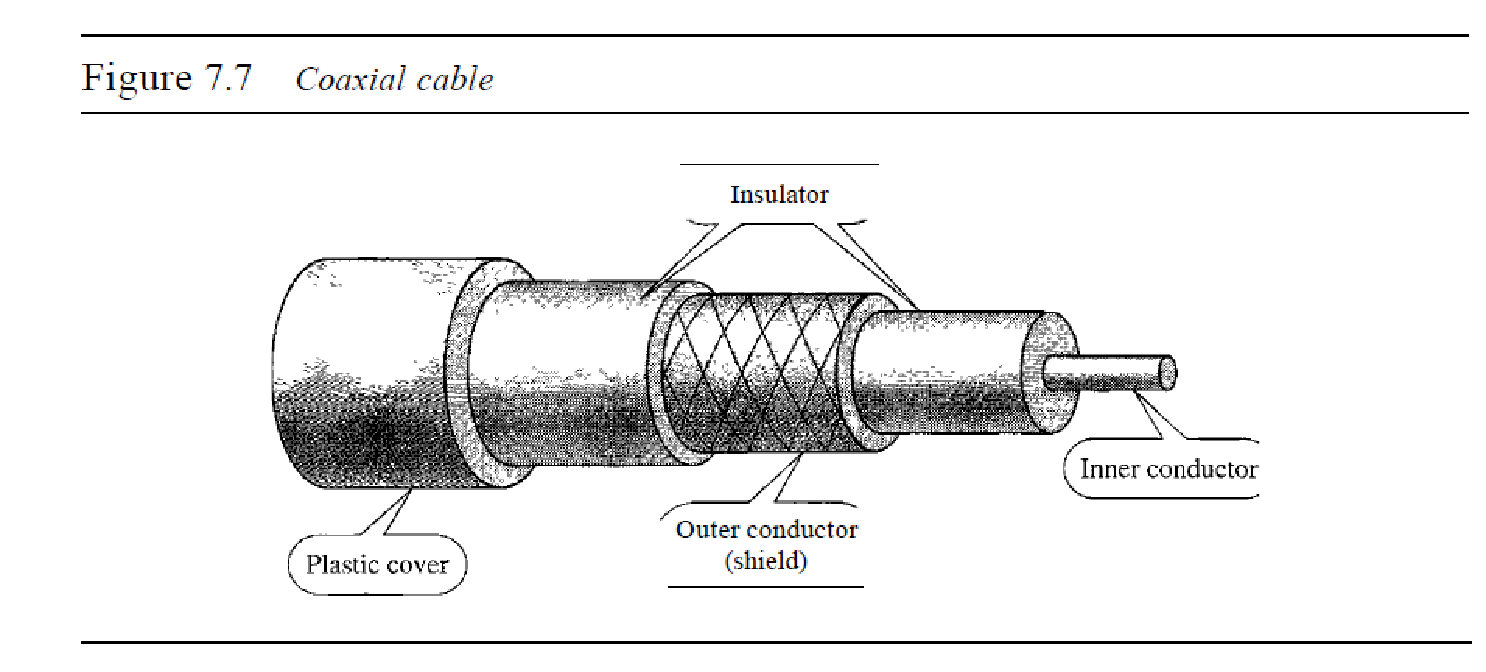
copper) enclosed in an insulating sheath, which is, in turn, encased in an outer conductor

of metal foil, braid, or a combination of the two. The outer metallic wrapping serves

both as a shield against noise and as the second conductor, which completes the circuit.

This outer conductor is also enclosed in an insulating sheath, and the whole cable is

protected by a plastic cover (see Figure 7.7).



*Coaxial Cable Standards*

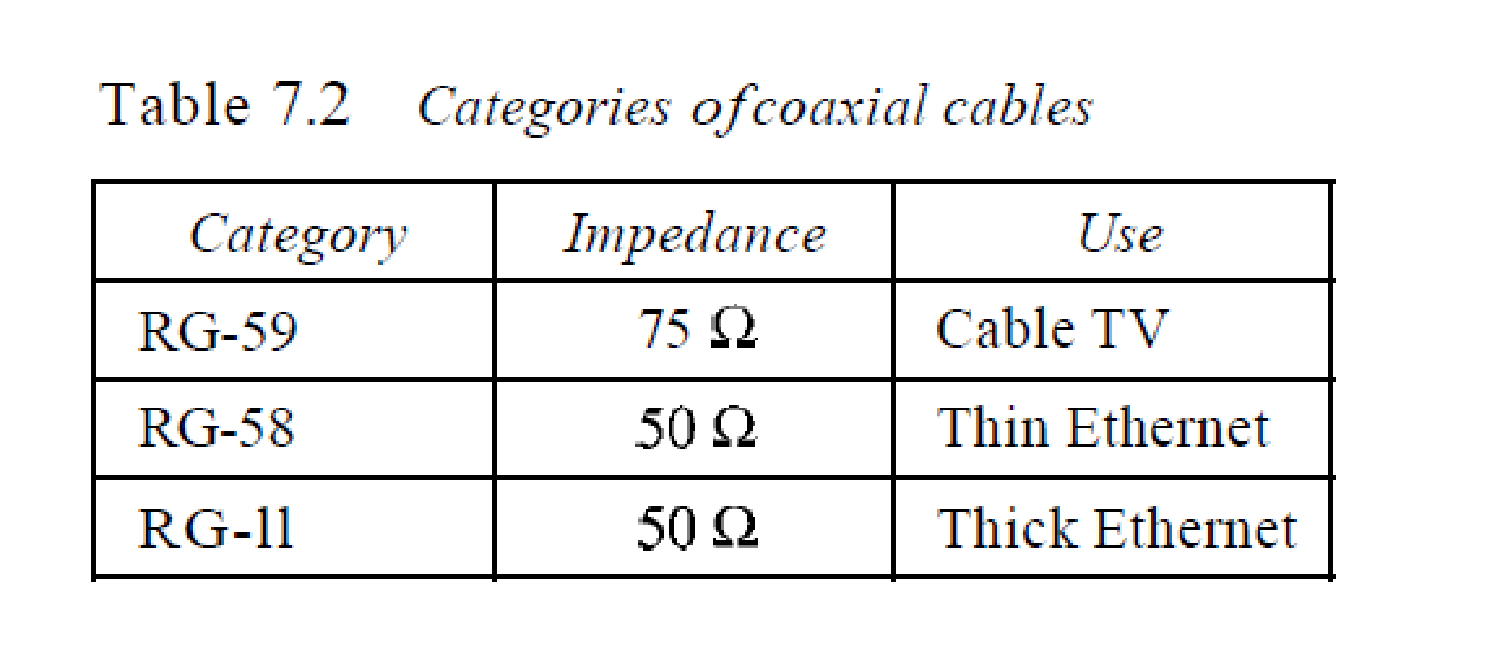
Coaxial cables are categorized by their radio government (RG) ratings. Each RG number

denotes a unique set of physical specifications, including the wire gauge of the

inner conductor, the thickness and type of the inner insulator, the construction of the

shield, and the size and type of the outer casing. Each cable defined by an RG rating is

adapted for a specialized function, as shown in Table 7.2.



*Coaxial Cable Connectors*

To connect coaxial cable to devices, we need coaxial connectors. The most common

type of connector used today is the Bayone-Neill-Concelman (BNe), connector.

Figure 7.8 shows three popular types of these connectors: the BNC connector, the

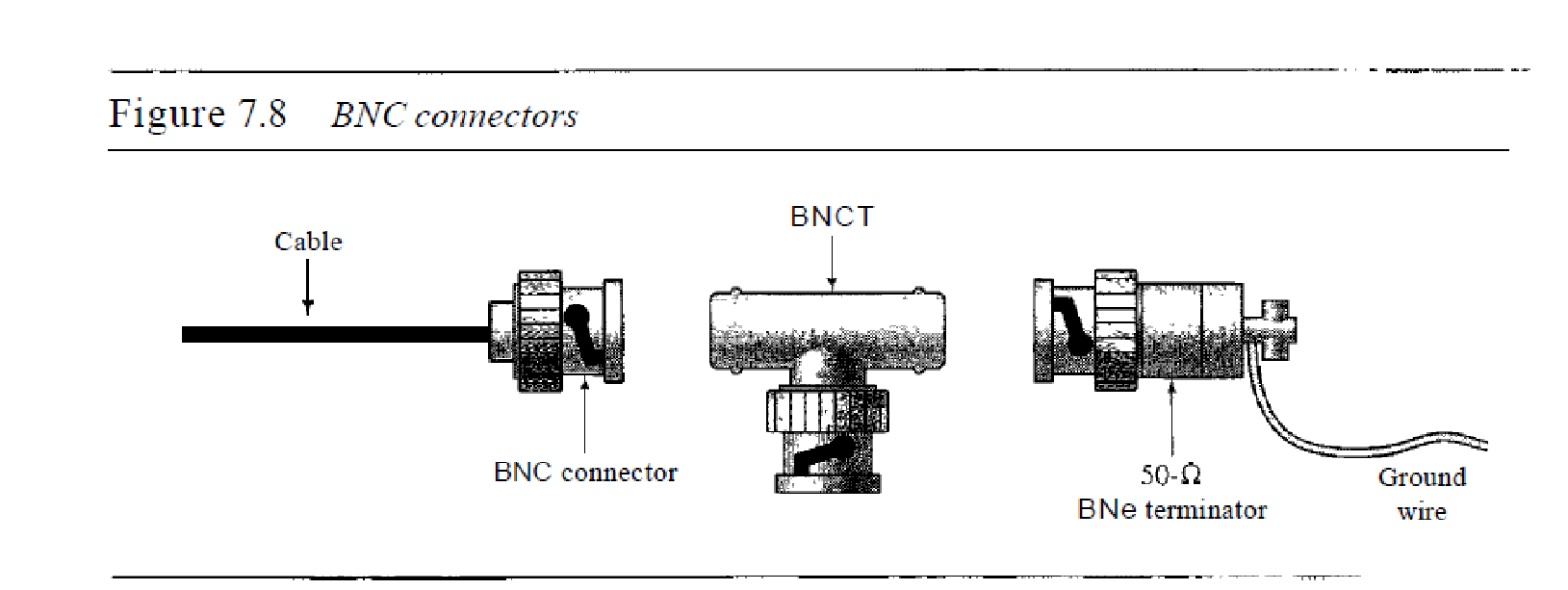
BNC T connector, and the BNC terminator.

The BNC connector is used to connect the end of the cable to a device, such as a

TV set. The BNC T connector is used in Ethernet networks (see Chapter 13) to branch

out to a connection to a computer or other device. The BNC terminator is used at the

end of the cable to prevent the reflection of the signal.



*Performance*

As we did with twisted-pair cables, we can measure the performance of a coaxial cable.

We notice in Figure 7.9 that the attenuation is much higher in coaxial cables than in

twisted-pair cable. In other words, although coaxial cable has a much higher bandwidth,

the signal weakens rapidly and requires the frequent use of repeaters

*Applications*

Coaxial cable was widely used in analog telephone networks where a single coaxial network

could carry 10,000 voice signals. Later it was used in digital telephone networks

where a single coaxial cable could carry digital data up to 600 Mbps. However, coaxial

cable in telephone networks has largely been replaced today with fiber-optic cable.

Cable TV networks (see Chapter 9) also use coaxial cables. In the traditional cable

TV network, the entire network used coaxial cable. Later, however, cable TV providersreplaced most of the media with fiber-optic cable; hybrid networks use coaxial cable

only at the network boundaries, near the consumer premises. Cable TV uses RG-59

coaxial cable.

Another common application of coaxial cable is in traditional Ethernet LANs (see

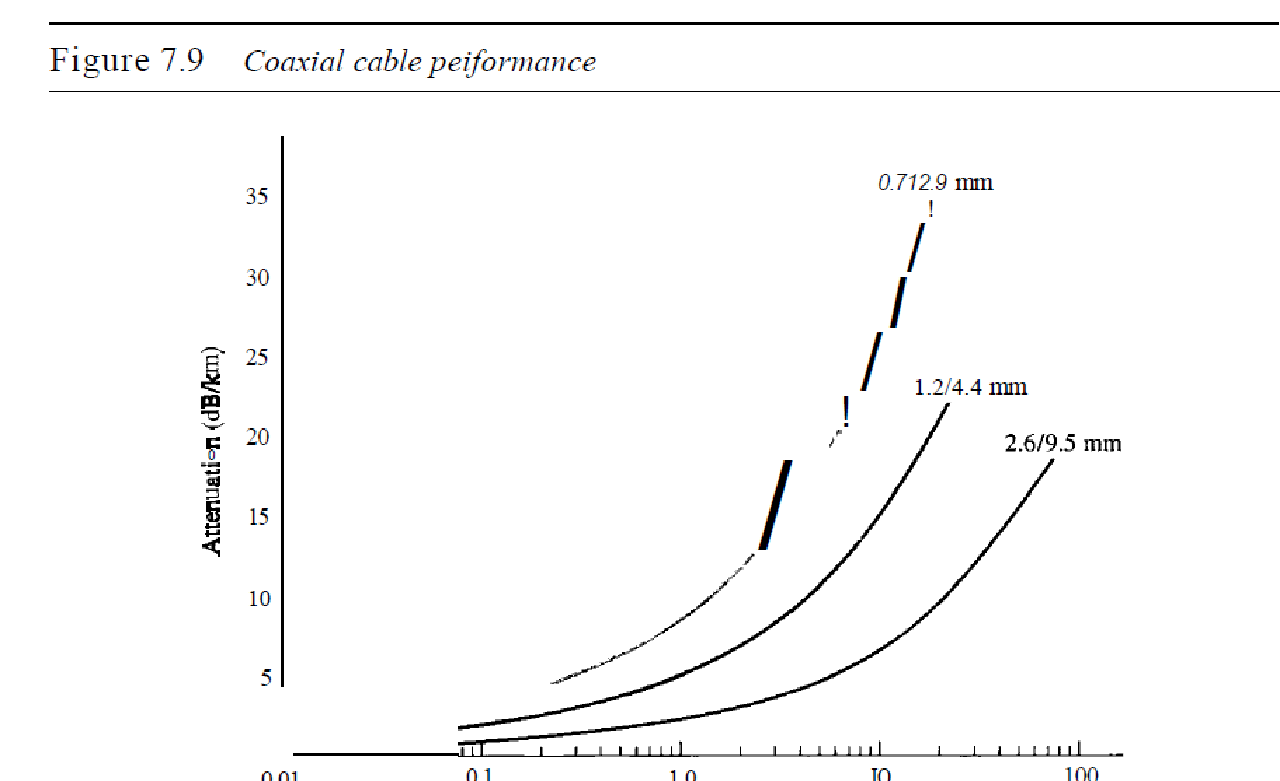
Chapter 13). Because of its high bandwidth, and consequently high data rate, coaxial

cable was chosen for digital transmission in early Ethernet LANs. The 10Base-2, or Thin

Ethernet, uses RG-58 coaxial cable with BNe connectors to transmit data at 10 Mbps

with a range of 185 m. The lOBase5, or Thick Ethernet, uses RG-11 (thick coaxial cable)

to transmit 10 Mbps with a range of 5000 m. Thick Ethernet has specialized connectors.



Fiber-Optic Cable

A fiber-optic cable is made of glass or plastic and transmits signals in the form of light.

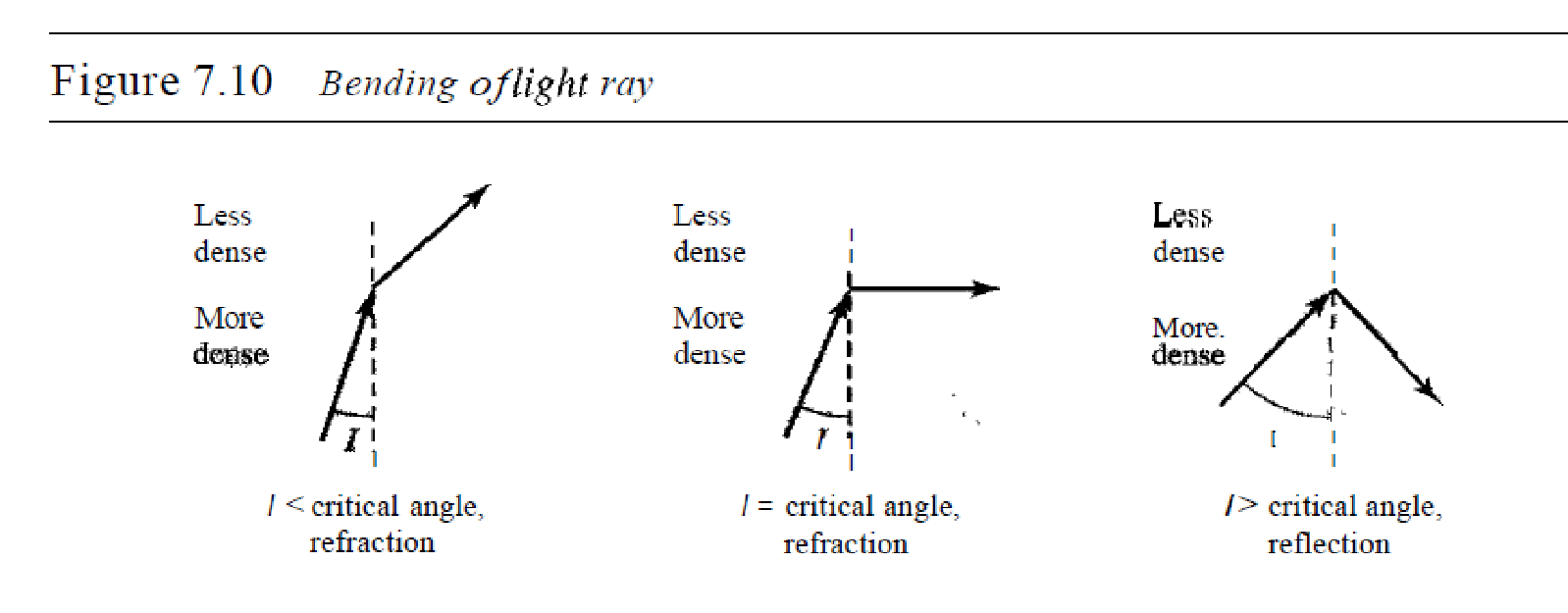
To understand optical fiber, we first need to explore several aspects of the nature of light.

Light travels in a straight line as long as it is moving through a single uniform substance.

If a ray of light traveling through one substance suddenly enters another substance

(of a different density), the ray changes direction. Figure 7.10 shows how a ray of light

changes direction when going from a more dense to a less dense substance.



As the figure shows, if the angle of incidence *I* (the arIgle the ray makes with the

line perpendicular to the interface between the two substances) is less than the critical

angle, the ray refracts and moves closer to the surface. If the angle of incidence is

equal to the critical angle, the light bends along the interface. If the angle is greater than

the critical angle, the ray reflects (makes a turn) and travels again in the denser substance.

Note that the critical angle is a property of the substance, and its value differs

from one substance to another.

Optical fibers use reflection to guide light through a channel. A glass or plastic core

is surrounded by a cladding of less dense glass or plastic. The difference in density of the

two materials must be such that a beam of light moving through the core is reflected off

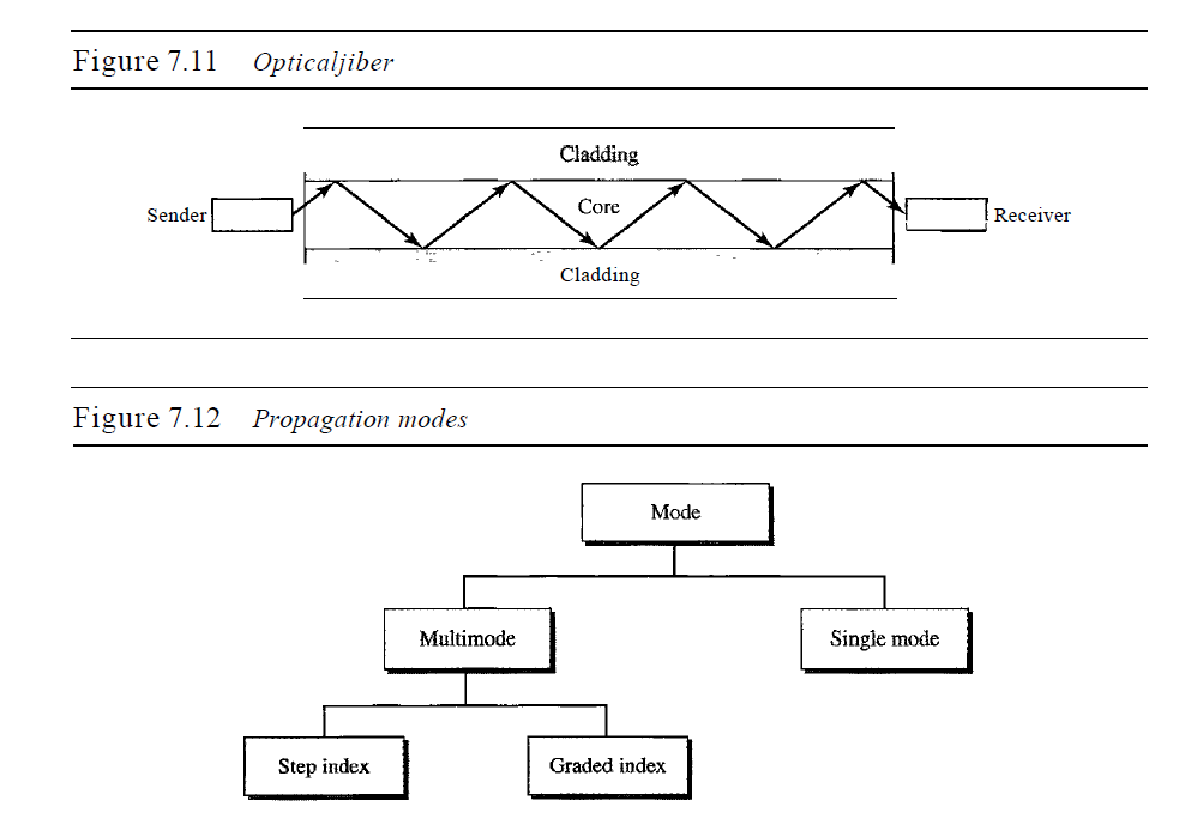
the cladding instead of being refracted into it. See Figure 7.11.

*Propagation Modes*

Current technology supports two modes (multimode and single mode) for propagating light

along optical channels, each requiring fiber with different physical characteristics. Multimode

can be implemented in two forms: step-index or graded-index (see Figure 7.12).



Multimode Multimode is so named because multiple beams from a light source

move through the core in different paths.

In multimode step-index fiber, the density of the core remains constant from the

center to the edges. A beam of light moves through this constant density in a straight

line until it reaches the interface of the core and the cladding. At the interface, there is

an abrupt change due to a lower density; this alters the angle of the beam's motion. The

term *step index* refers to the suddenness of this change, which contributes to the distortion

of the signal as it passes through the fiber.

A second type of fiber, called multimode graded-index fiber, decreases this distortion

of the signal through the cable. The word *index* here refers to the index of refraction.

As we saw above, the index of refraction is related to density. A graded-index fiber,

therefore, is one with varying densities. Density is highest at the center of the core and

decreases gradually to its lowest at the edge. Figure 7.13 shows the impact of this variable

density on the propagation of light beams.

Single-Mode Single-mode uses step-index fiber and a highly focused source of light

that limits beams to a small range of angles, all close to the horizontal. The singlemode

fiber itself is manufactured with a much smaller diameter than that of multimode

fiber, and with substantiallY lower density (index of refraction). The decrease in density

results in a critical angle that is close enough to 90° to make the propagation of beams

almost horizontal. In this case, propagation of different beams is almost identical, and

delays are negligible. All the beams arrive at the destination "together" and can be

recombined with little distortion to the signal (see Figure 7.13).

*Fiber Sizes*

Optical fibers are defined by the ratio of the diameter of their core to the diameter of

their cladding, both expressed in micrometers. The common sizes are shown in Table 7.3.

Note that the last size listed is for single-mode only.

*Advantages and Disadvantages of Optical Fiber*

Advantages Fiber-optic cable has several advantages over metallic cable (twistedpair

or coaxial).

D Higher bandwidth. Fiber-optic cable can support dramatically higher bandwidths

(and hence data rates) than either twisted-pair or coaxial cable. Currently, data rates

and bandwidth utilization over fiber-optic cable are limited not by the medium but

by the signal generation and reception technology available.

D Less signal attenuation. Fiber-optic transmission distance is significantly greater

than that of other guided media. A signal can run for 50 km without requiring

regeneration. We need repeaters every 5 km for coaxial or twisted-pair cable.

D Immunity to electromagnetic interference. Electromagnetic noise cannot affect

fiber-optic cables.

o Resistance to corrosive materials. Glass is more resistant to corrosive materials

than copper.

o Light weight. Fiber-optic cables are much lighter than copper cables.

o Greater immunity to tapping. Fiber-optic cables are more immune to tapping than

copper cables. Copper cables create antenna effects that can easily be tapped.

Disadvantages There are some disadvantages in the use of optical fiber.

o Installation and maintenance. Fiber-optic cable is a relatively new technology. Its

installation and maintenance require expertise that is not yet available everywhere.

o Unidirectional light propagation. Propagation of light is unidirectional. If we

need bidirectional communication, two fibers are needed.

o Cost. The cable and the interfaces are relatively more expensive than those of other

guided media. If the demand for bandwidth is not high, often the use of optical fiber

cannot be justified.

UNGUIDED MEDIA: WIRELESS

Unguided media transport electromagnetic waves without using a physical conductor.

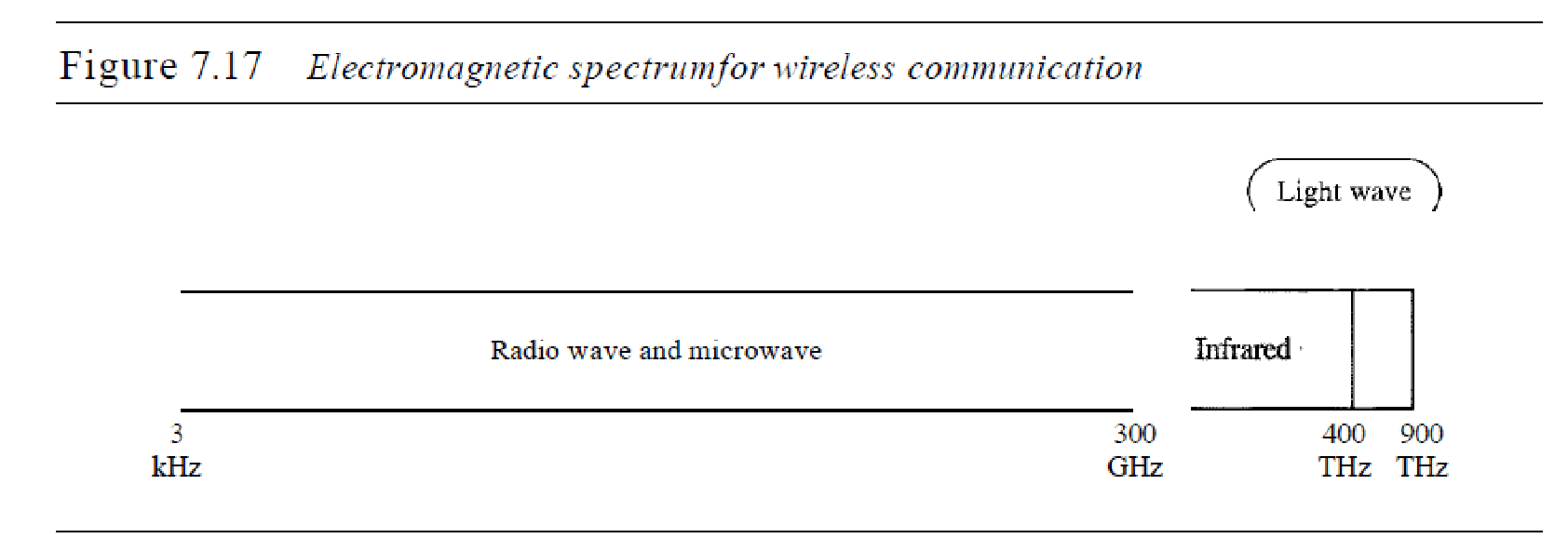
This type of communication is often referred to as wireless communication. Signals

are normally broadcast through free space and thus are available to anyone who has a

device capable of receiving them.

Figure 7.17 shows the part of the electromagnetic spectrum, ranging from 3 kHz to

900 THz, used for wireless communication.



Unguided signals can travel from the source to destination in several ways: ground

propagation, sky propagation, and line-of-sight propagation, as shown in Figure 7.18.

In ground propagation, radio waves travel through the lowest portion of the

atmosphere, hugging the earth. These low-frequency signals emanate in all directions

from the transmitting antenna and follow the curvature of the planet. Distance depends

on the amount of power in the signal: The greater the power, the greater the distance. In

sky propagation, higher-frequency radio waves radiate upward into the ionosphere

(the layer of atmosphere where particles exist as ions) where they are reflected back to

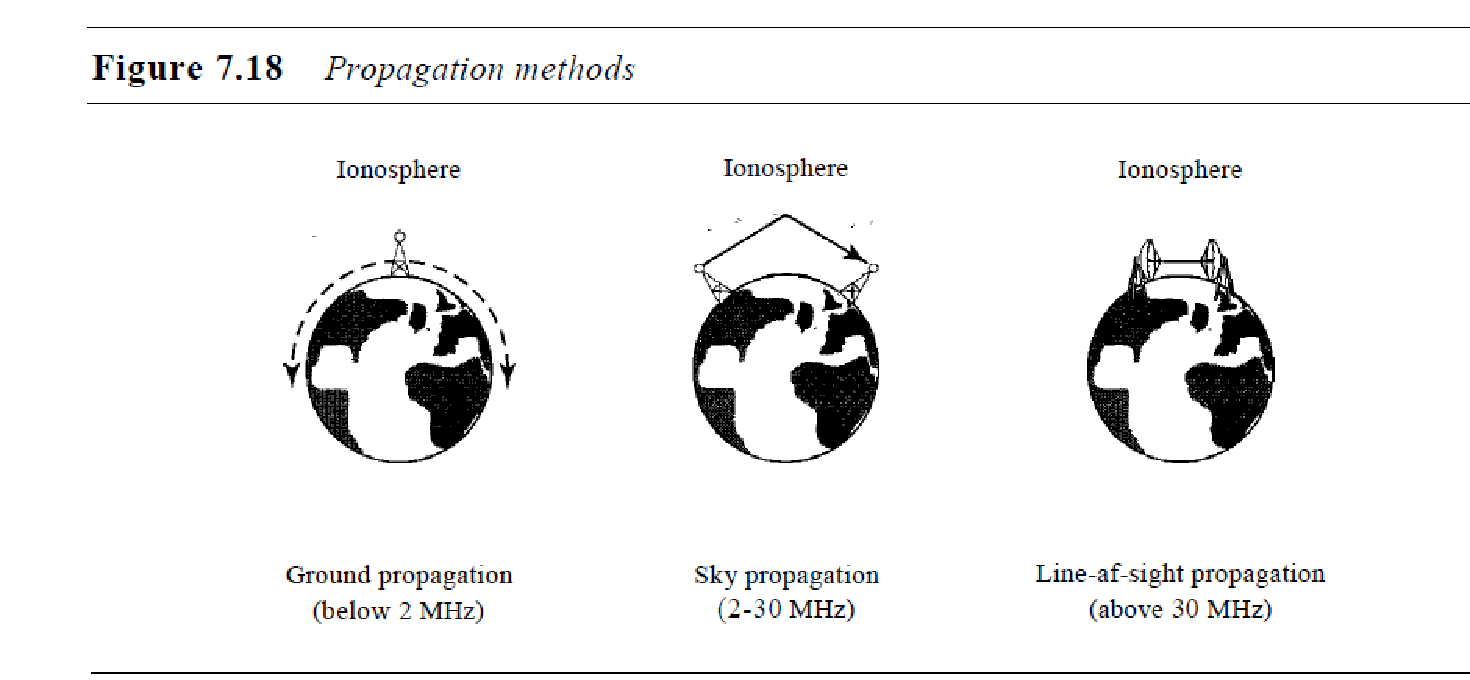
earth. This type of transmission allows for greater distances with lower output power.

In line-or-sight propagation, very high-frequency signals are transmitted in straight

lines directly from antenna to antenna. Antennas must be directional, facing each other, and either tall enough or close enough together not to be affected by the curvature of

the earth. Line-of-sight propagation is tricky because radio transmissions cannot be

completely focused.

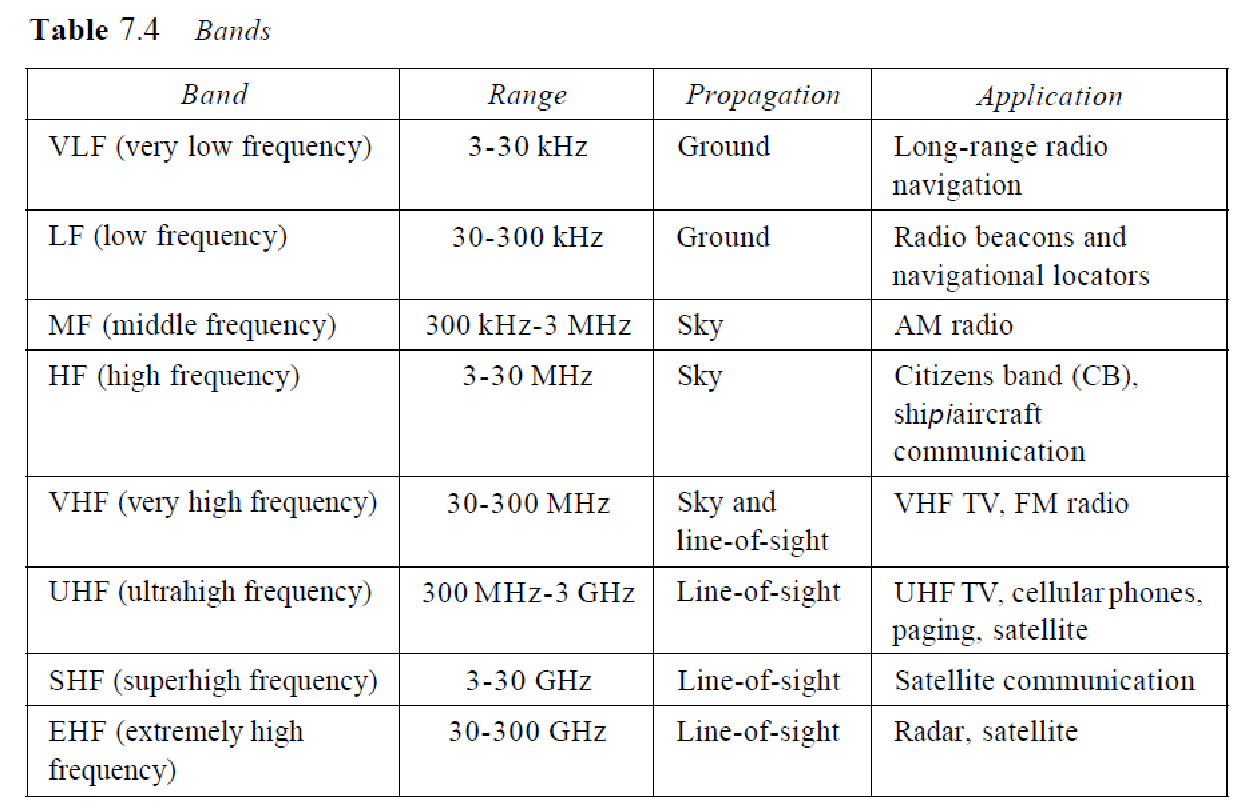


The section of the electromagnetic spectrum defined as radio waves and microwaves

is divided into eight ranges, called *bands,* each regulated by government authorities.

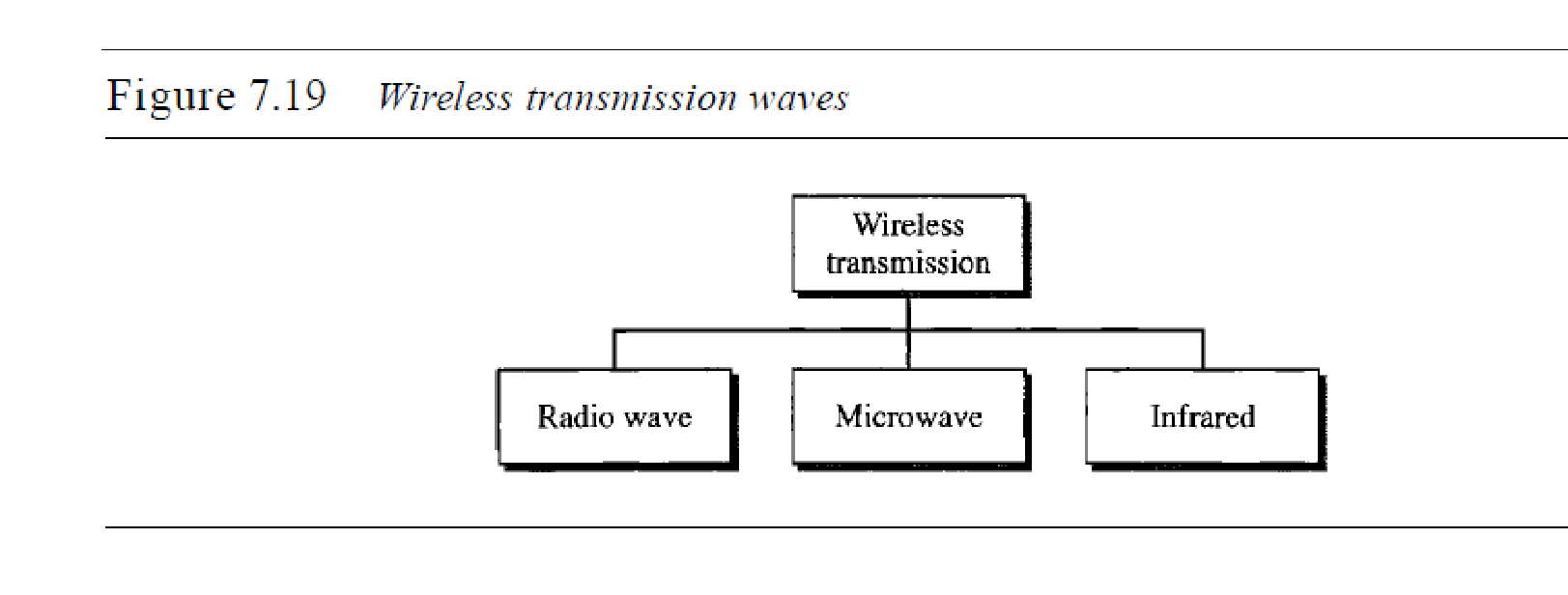
These bands are rated from *very low frequency* (VLF) to *extremely highfrequency* (EHF).

Table 7.4 lists these bands, their ranges, propagation methods, and some applications



We can divide wireless transmission into three broad groups: radio waves, microwaves,

and infrared waves. See Figure 7.19.



Radio Waves

Although there is no clear-cut demarcation between radio waves and microwaves, electromagnetic

waves ranging in frequencies between 3 kHz and 1 GHz are normally called

radio waves; waves ranging in frequencies between 1 and 300 GHz are called microwaves.

However, the behavior of the waves, rather than the frequencies, is a better

criterion for classification.

Radio waves, for the most part, are omnidirectional. When an antenna transmits

radio waves, they are propagated in all directions. This means that the sending and

receiving antennas do not have to be aligned. A sending antenna sends waves that can

be received by any receiving antenna. The omnidirectional property has a disadvantage,

too. The radio waves transmitted by one antenna are susceptible to interference by

another antenna that may send signals using the same frequency or band.

Radio waves, particularly those waves that propagate in the sky mode, can travel

long distances. This makes radio waves a good candidate for long-distance broadcasting

such as AM radio.

Radio waves, particularly those of low and medium frequencies, can penetrate walls.

This characteristic can be both an advantage and a disadvantage. It is an advantage

because, for example, an AM radio can receive signals inside a building. It is a disadvantage

because we cannot isolate a communication to just inside or outside a building. The

radio wave band is relatively narrow, just under 1 GHz, compared to the microwave

band. When this band is divided into subbands, the subbands are also narrow, leading to a

low data rate for digital communications.

Almost the entire band is regulated by authorities (e.g., the FCC in the United

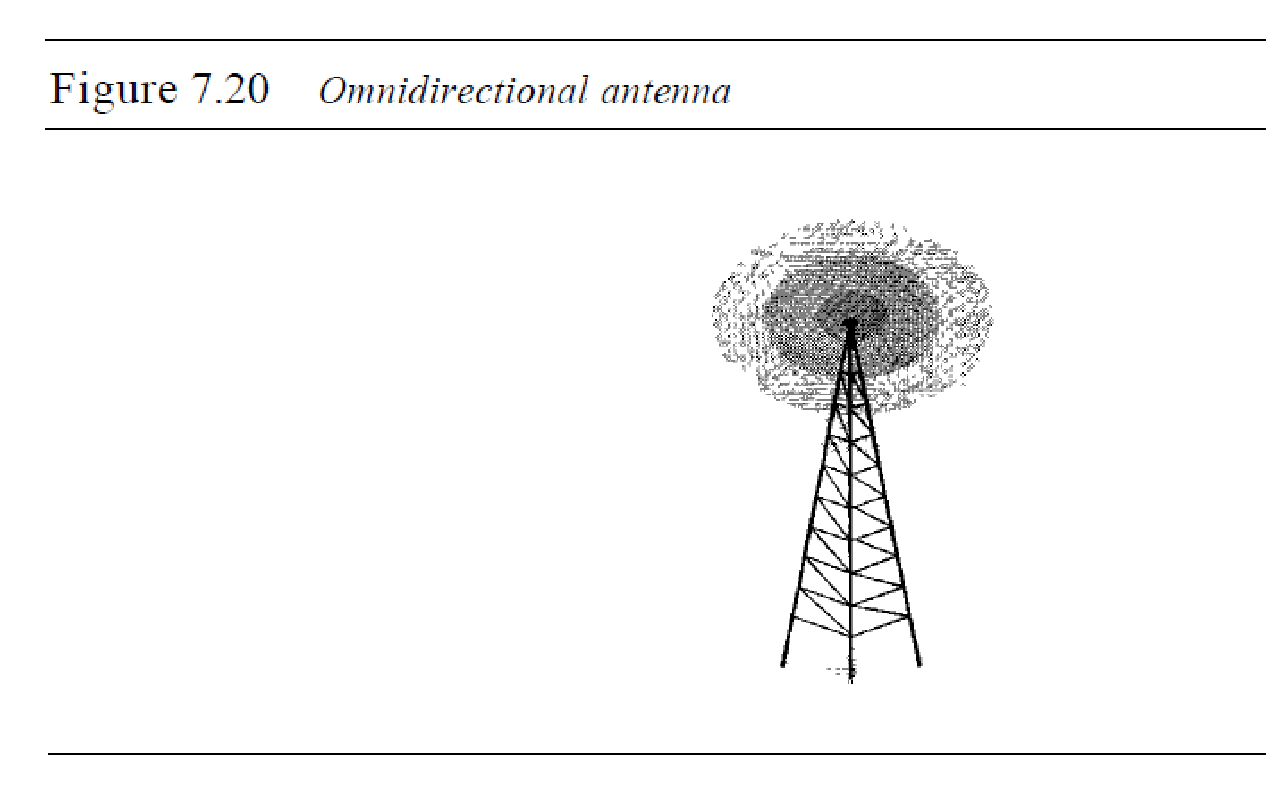
States). Using any part of the band requires permission from the authorities.

*Omnidirectional Antenna*

Radio waves use omnidirectional antennas that send out signals in all directions.

Based on the wavelength, strength, and the purpose of transmission, we can have several

types of antennas. Figure 7.20 shows an omnidirectional antenna.



*Applications*

The omnidirectional characteristics of radio waves make them useful for multicasting,

in which there is one sender but many receivers. AM and FM radio, television, maritime

radio, cordless phones, and paging are examples of multicasting.

Microwaves

Electromagnetic waves having frequencies between I and 300 GHz are called microwaves.

Microwaves are unidirectional. When an antenna transmits microwave waves, they

can be narrowly focused. This means that the sending and receiving antennas need to

be aligned. The unidirectional property has an obvious advantage. A pair of antennas

can be aligned without interfering with another pair of aligned antennas. The following

describes some characteristics of microwave propagation:

o Microwave propagation is line-of-sight. Since the towers with the mounted antennas

need to be in direct sight of each other, towers that are far apart need to be very tall.

The curvature of the earth as well as other blocking obstacles do not allow two short

towers to communicate by using microwaves. Repeaters are often needed for longdistance

communication.

o Very high-frequency microwaves cannot penetrate walls. This characteristic can be

a disadvantage if receivers are inside buildings.

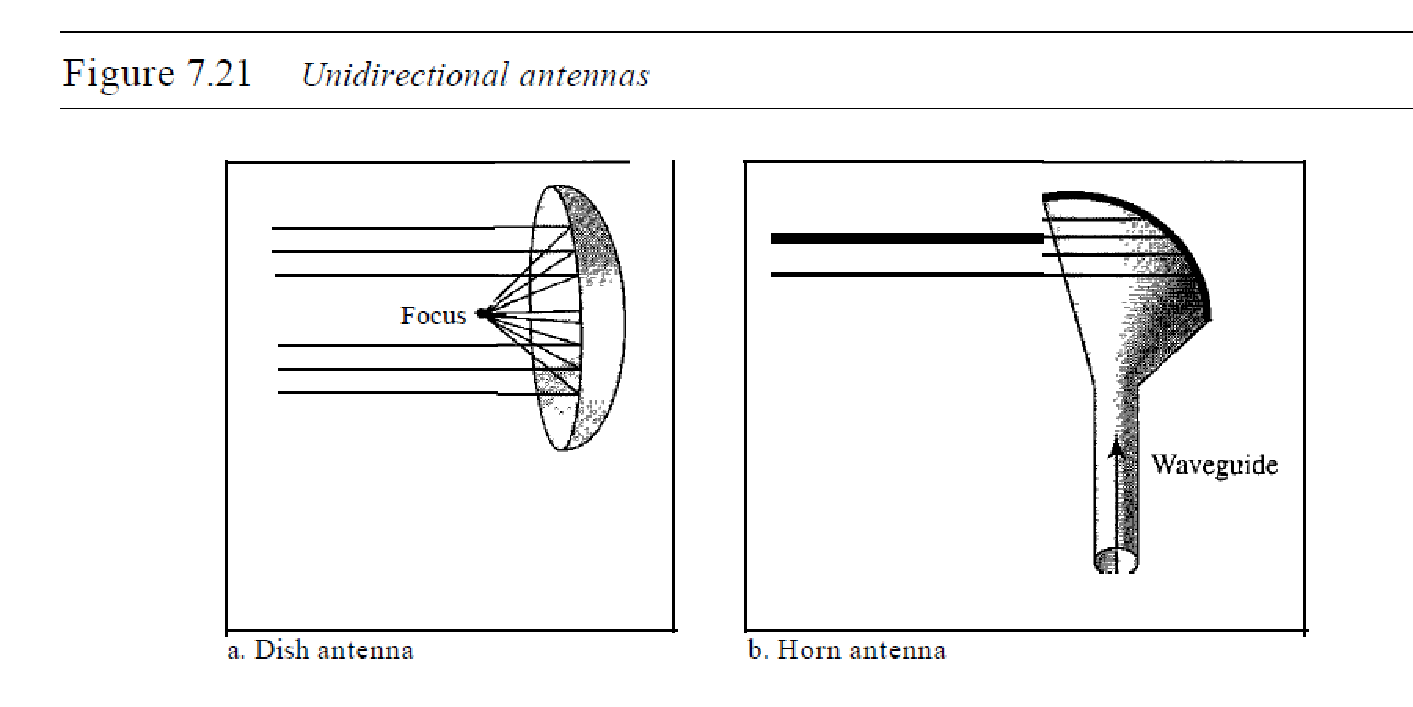
o The microwave band is relatively wide, almost 299 GHz. Therefore wider subbands

can be assigned, and a high data rate is possible

o Use of certain portions of the band requires permission from authorities.

*Unidirectional Antenna*

Microwaves need unidirectional antennas that send out signals in one direction. Two

types of antennas are used for microwave communications: the parabolic dish and the hom (see Figure 7.21). 

A parabolic dish antenna is based on the geometry of a parabola: Every line

parallel to the line of symmetry (line of sight) reflects off the curve at angles such that

all the lines intersect in a common point called the focus. The parabolic dish works as afunnel, catching a wide range of waves and directing them to a common point. In

this way, more of the signal is recovered than would be possible with a single-point

receiver.

Outgoing transmissions are broadcast through a horn aimed at the dish. The microwaves

hit the dish and are deflected outward in a reversal of the receipt path.

A horn antenna looks like a gigantic scoop. Outgoing transmissions are broadcast

up a stem (resembling a handle) and deflected outward in a series of narrow parallel

beams by the curved head. Received transmissions are collected by the scooped shape of

the horn, in a manner similar to the parabolic dish, and are deflected down into the stem.

*Applications*

Microwaves, due to their unidirectional properties, are very useful when unicast

(one-to-one) communication is needed between the sender and the receiver. They are

used in cellular phones (Chapter 16), satellite networks (Chapter 16), and wireless LANs

(Chapter 14).

*Microwaves are used for unicast communication such as cellular telephones,*

*satellite networks, and wireless LANs.*

Infrared

Infrared waves, with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm

to 770 nm), can be used for short-range communication. Infrared waves, having high

frequencies, cannot penetrate walls. This advantageous characteristic prevents interference

between one system and another; a short-range communication system in one room

cannot be affected by another system in the next room. When we use our infrared remote

control, we do not interfere with the use of the remote by our neighbors. However, this

same characteristic makes infrared signals useless for long-range communication. In

addition, we cannot use infrared waves outside a building because the sun's rays contain

infrared waves that can interfere with the communication.

*Applications*

The infrared band, almost 400 THz, has an excellent potential for data transmission.

Such a wide bandwidth can be used to transmit digital data with a very high data rate.

The *Infrared Data Association* (IrDA), an association for sponsoring the use of infrared

waves, has established standards for using these signals for communication between

devices such as keyboards, mice, PCs, and printers. For example, some manufacturers

provide a special port called the IrDA port that allows a wireless keyboard to communicate

with a PC. The standard originally defined a data rate of 75 kbps for a distance

up to 8 m. The recent standard defines a data rate of 4 Mbps.

Infrared signals defined by IrDA transmit through line of sight; the IrDA port on

the keyboard needs to point to the PC for transmission to occur.

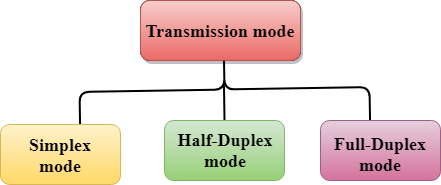
Infrared signals can be used for short-range communication

in a closed area using line-of-sight propagation.

**Transmission mode:**

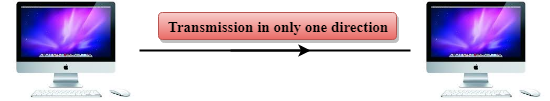
* The way in which data is transmitted from one device to another device is known as **transmission mode**.
* The transmission mode is also known as the communication mode.
* Each communication channel has a direction associated with it, and transmission media provide the direction. Therefore, the transmission mode is also known as a directional mode.
* The transmission mode is defined in the physical layer.

The Transmission mode is divided into three categories:



* Simplex mode
* Half-duplex mode
* Full-duplex mode

Simplex mode



* In Simplex mode, the communication is unidirectional, i.e., the data flow in one direction.
* A device can only send the data but cannot receive it or it can receive the data but cannot send the data.
* This transmission mode is not very popular as mainly communications require the two-way exchange of data. The simplex mode is used in the business field as in sales that do not require any corresponding reply.
* The radio station is a simplex channel as it transmits the signal to the listeners but never allows them to transmit back.
* Keyboard and Monitor are the examples of the simplex mode as a keyboard can only accept the data from the user and monitor can only be used to display the data on the screen.
* The main advantage of the simplex mode is that the full capacity of the communication channel can be utilized during transmission.

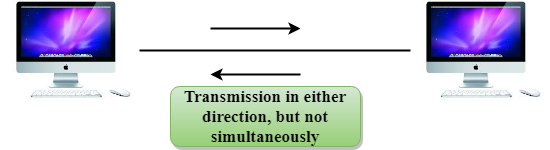
Advantage of Simplex mode:

* In simplex mode, the station can utilize the entire bandwidth of the communication channel, so that more data can be transmitted at a time.

Disadvantage of Simplex mode:

* Communication is unidirectional, so it has no inter-communication between devices.

Half-Duplex mode



* In a Half-duplex channel, direction can be reversed, i.e., the station can transmit and receive the data as well.
* Messages flow in both the directions, but not at the same time.
* The entire bandwidth of the communication channel is utilized in one direction at a time.
* In half-duplex mode, it is possible to perform the error detection, and if any error occurs, then the receiver requests the sender to retransmit the data.
* A **Walkie-talkie** is an example of the Half-duplex mode. In Walkie-talkie, one party speaks, and another party listens. After a pause, the other speaks and first party listens. Speaking simultaneously will create the distorted sound which cannot be understood.

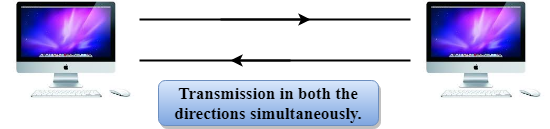
Advantage of Half-duplex mode:

* In half-duplex mode, both the devices can send and receive the data and also can utilize the entire bandwidth of the communication channel during the transmission of data.

Disadvantage of Half-Duplex mode:

* In half-duplex mode, when one device is sending the data, then another has to wait, this causes the delay in sending the data at the right time.

Full-duplex mode



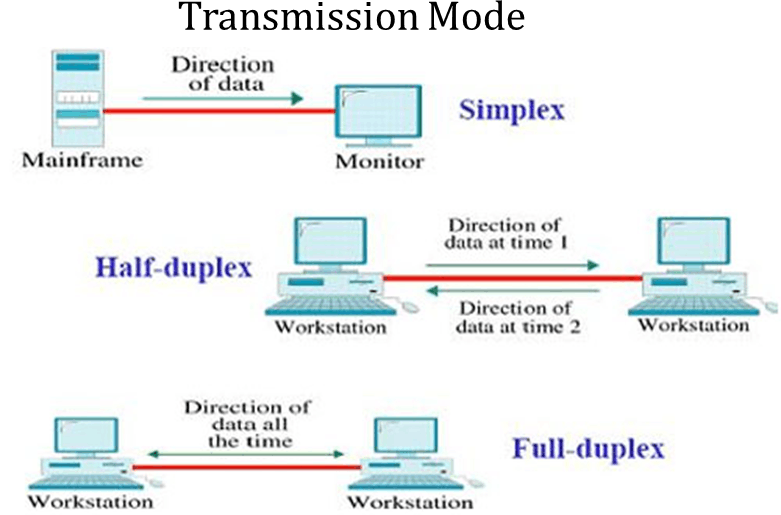
* In Full duplex mode, the communication is bi-directional, i.e., the data flow in both the directions.
* Both the stations can send and receive the message simultaneously.
* Full-duplex mode has two simplex channels. One channel has traffic moving in one direction, and another channel has traffic flowing in the opposite direction.
* The Full-duplex mode is the fastest mode of communication between devices.
* The most common example of the full-duplex mode is a telephone network. When two people are communicating with each other by a telephone line, both can talk and listen at the same time.

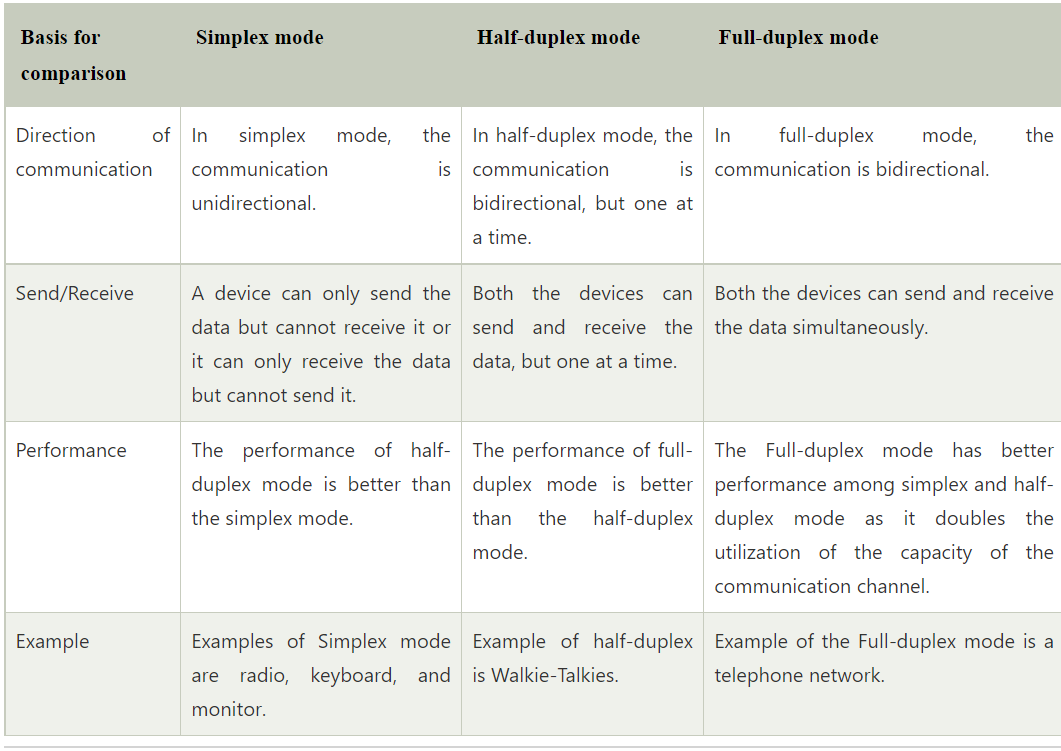
Advantage of Full-duplex mode:

* Both the stations can send and receive the data at the same time.

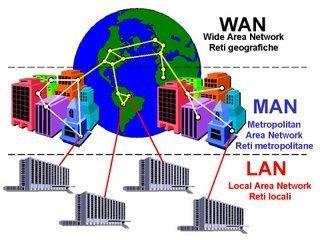
Disadvantage of Full-duplex mode:

* If there is no dedicated path exists between the devices, then the capacity of the communication channel is divided into two parts.



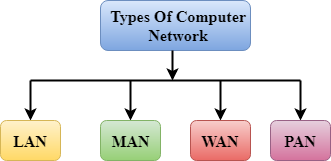


**Classification of Networks:**



A computer network is a group of computers linked to each other that enables the computer to communicate with another computer and share their resources, data, and applications.

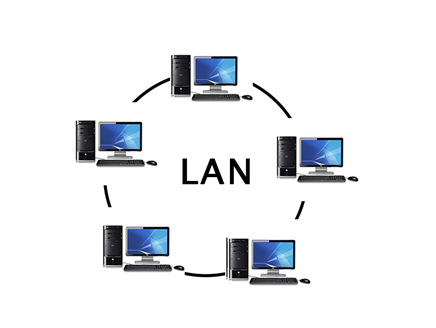
A computer network can be categorized by their size. A **computer network** is mainly of **four types**:



* LAN (Local Area Network)
* PAN (Personal Area Network)
* MAN (Metropolitan Area Network)
* WAN (Wide Area Network)

LAN (Local Area Network)

* Local Area Network is a group of computers connected to each other in a small area such as building, office.
* LAN is used for connecting two or more personal computers through a communication medium such as twisted pair, coaxial cable, etc.
* It is less costly as it is built with inexpensive hardware such as hubs, network adapters, and ethernet cables.
* The data is transferred at an extremely faster rate in Local Area Network.
* Local Area Network provides higher security.

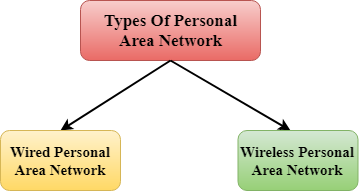


PAN (Personal Area Network)

* Personal Area Network is a network arranged within an individual person, typically within a range of 10 meters.
* Personal Area Network is used for connecting the computer devices of personal use is known as Personal Area Network.
* **Thomas Zimmerman** was the first research scientist to bring the idea of the Personal Area Network.
* Personal Area Network covers an area of **30 feet**.
* Personal computer devices that are used to develop the personal area network are the laptop, mobile phones, media player and play stations.



**There are two types of Personal Area Network:**



* Wired Personal Area Network
* Wireless Personal Area Network

**Wireless Personal Area Network:** Wireless Personal Area Network is developed by simply using wireless technologies such as WiFi, Bluetooth. It is a low range network.

Skip Ad

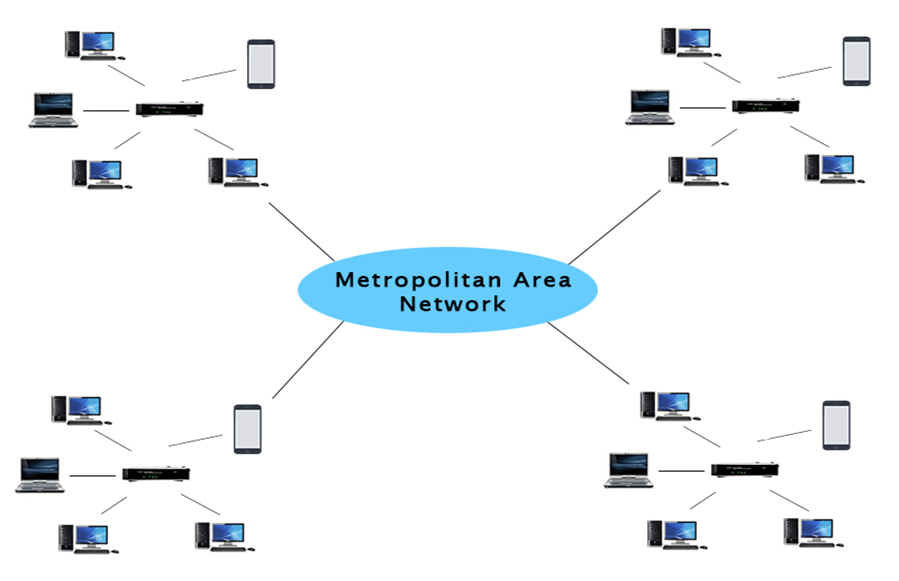
**Wired Personal Area Network:** Wired Personal Area Network is created by using the USB.

Examples Of Personal Area Network:

* **Body Area Network:** Body Area Network is a network that moves with a person. **For example**, a mobile network moves with a person. Suppose a person establishes a network connection and then creates a connection with another device to share the information.
* **Offline Network:** An offline network can be created inside the home, so it is also known as a **home network**. A home network is designed to integrate the devices such as printers, computer, television but they are not connected to the internet.
* **Small Home Office:** It is used to connect a variety of devices to the internet and to a corporate network using a VPN

MAN (Metropolitan Area Network)

* A metropolitan area network is a network that covers a larger geographic area by interconnecting a different LAN to form a larger network.
* Government agencies use MAN to connect to the citizens and private industries.
* In MAN, various LANs are connected to each other through a telephone exchange line.
* The most widely used protocols in MAN are RS-232, Frame Relay, ATM, ISDN, OC-3, ADSL, etc.
* It has a higher range than Local Area Network(LAN).

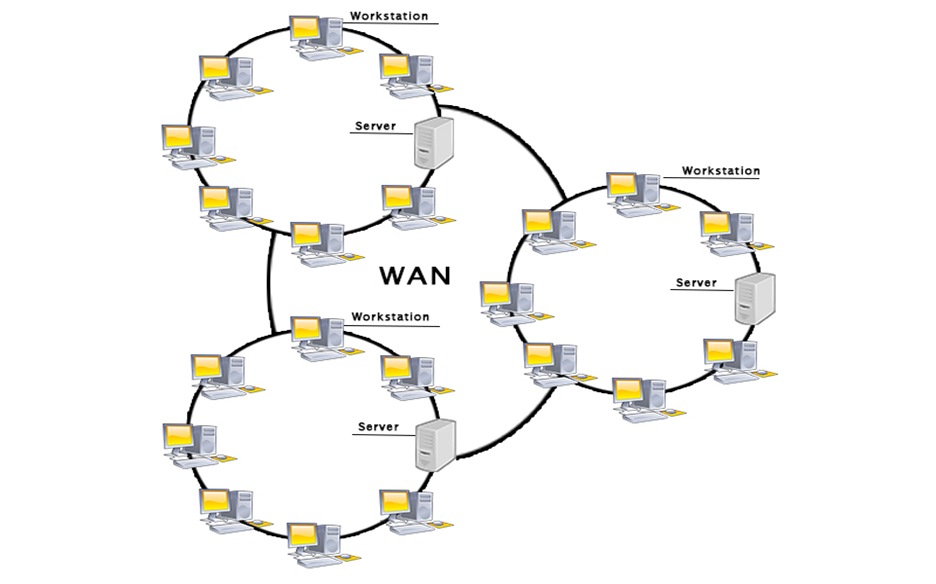


Uses Of Metropolitan Area Network:

* MAN is used in communication between the banks in a city.
* It can be used in an Airline Reservation.
* It can be used in a college within a city.
* It can also be used for communication in the military.

WAN (Wide Area Network)

* A Wide Area Network is a network that extends over a large geographical area such as states or countries.
* A Wide Area Network is quite bigger network than the LAN.
* A Wide Area Network is not limited to a single location, but it spans over a large geographical area through a telephone line, fibre optic cable or satellite links.
* The internet is one of the biggest WAN in the world.
* A Wide Area Network is widely used in the field of Business, government, and education.



Examples Of Wide Area Network:

* **Mobile Broadband:** A 4G network is widely used across a region or country.
* **Last mile:** A telecom company is used to provide the internet services to the customers in hundreds of cities by connecting their home with fiber.
* **Private network:** A bank provides a private network that connects the 44 offices. This network is made by using the telephone leased line provided by the telecom company.

Advantages Of Wide Area Network:

Following are the advantages of the Wide Area Network:

* **Geographical area:** A Wide Area Network provides a large geographical area. Suppose if the branch of our office is in a different city then we can connect with them through WAN. The internet provides a leased line through which we can connect with another branch.
* **Centralized data:** In case of WAN network, data is centralized. Therefore, we do not need to buy the emails, files or back up servers.
* **Get updated files:** Software companies work on the live server. Therefore, the programmers get the updated files within seconds.
* **Exchange messages:** In a WAN network, messages are transmitted fast. The web application like Facebook, Whatsapp, Skype allows you to communicate with friends.
* **Sharing of software and resources:** In WAN network, we can share the software and other resources like a hard drive, RAM.
* **Global business:** We can do the business over the internet globally.
* **High bandwidth:** If we use the leased lines for our company then this gives the high bandwidth. The high bandwidth increases the data transfer rate which in turn increases the productivity of our company.

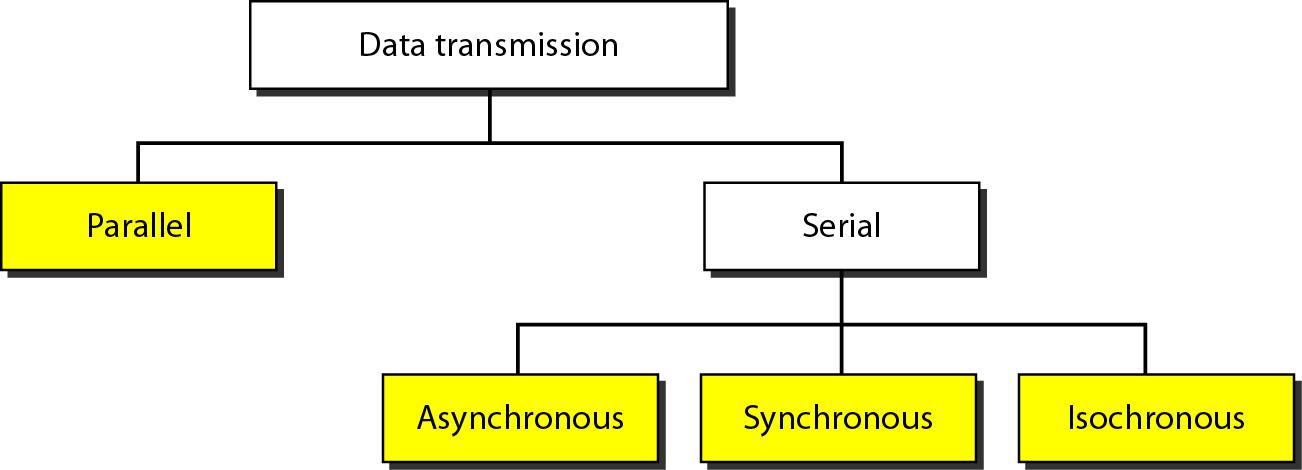
Disadvantages of Wide Area Network:

The following are the disadvantages of the Wide Area Network:

* **Security issue:** A WAN network has more security issues as compared to LAN and MAN network as all the technologies are combined together that creates the security problem.
* **Needs Firewall & antivirus software:** The data is transferred on the internet which can be changed or hacked by the hackers, so the firewall needs to be used. Some people can inject the virus in our system so antivirus is needed to protect from such a virus.
* **High Setup cost:** An installation cost of the WAN network is high as it involves the purchasing of routers, switches.
* **Troubleshooting problems:** It covers a large area so fixing the problem is difficult.

**Parallel Transmission and Serial Transmission:**

* The transmission of binary data across a link can be accomplished in either parallel or serial mode.
* Parallel Transmission-multiple bits are sent with each clock tick.
* Serial Transmission-1 bit is sent with each clock tick



**Parallel transmission:**

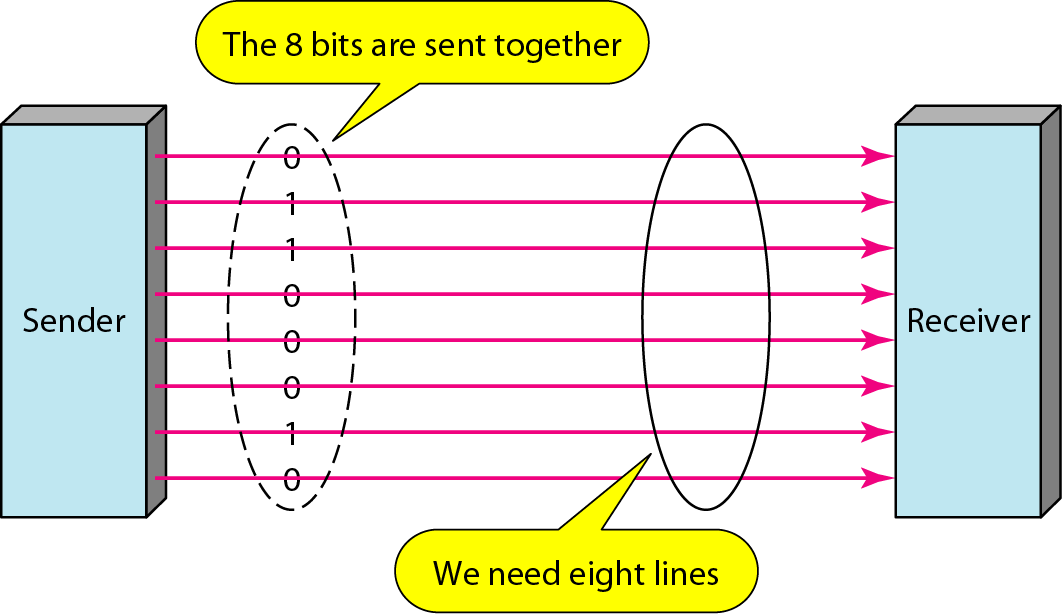
Binary data, consisting of 1s and 0s, may be organized into groups of n bits each.

● Computers produce and consume data in groups of bits much as we conceive of and use spoken language in the form of words rather than letters.

● By grouping, we can send data n bits at a time instead of 1. This is called parallel transmission.

● The mechanism for parallel transmission is a conceptually simple one: Use ‘n’ wires to send ‘n’ bits at one time.

● The advantage of parallel transmission is speed.

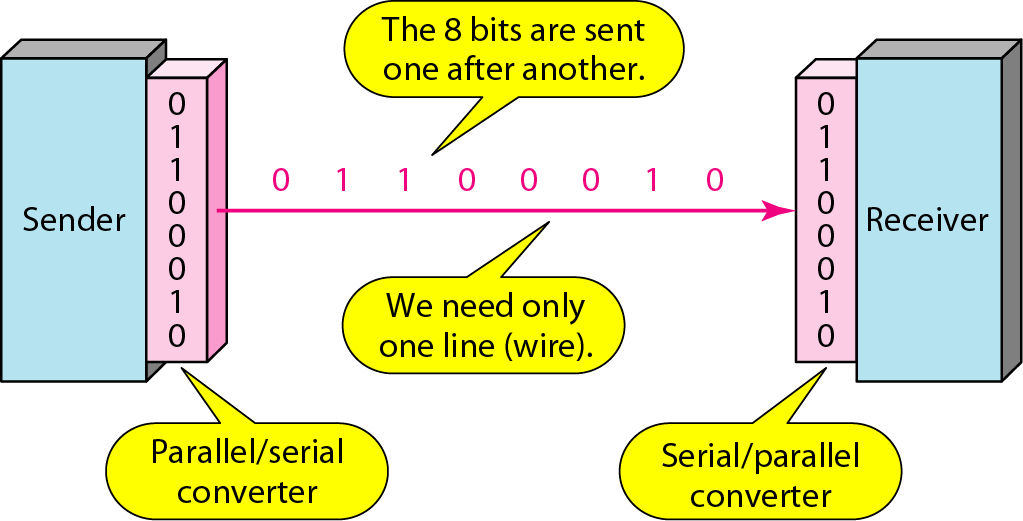


**Serial transmission:**

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

● Reduces the cost of transmission over parallel by roughly a factor of n.

● Since communication within devices is parallel, conversion devices are required at the interface between the sender and the line (parallel-to-serial) and between the line and the receiver (serial-to-parallel).



Serial transmission occurs in one of three ways: 1. Asynchronous. 2. Synchronous. 3. Isochronous.

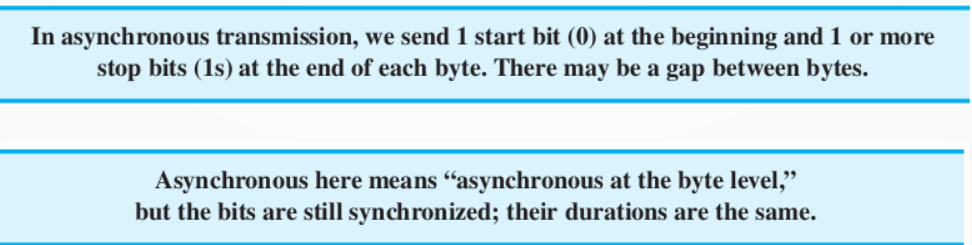
**Asynchronous Transmission**:

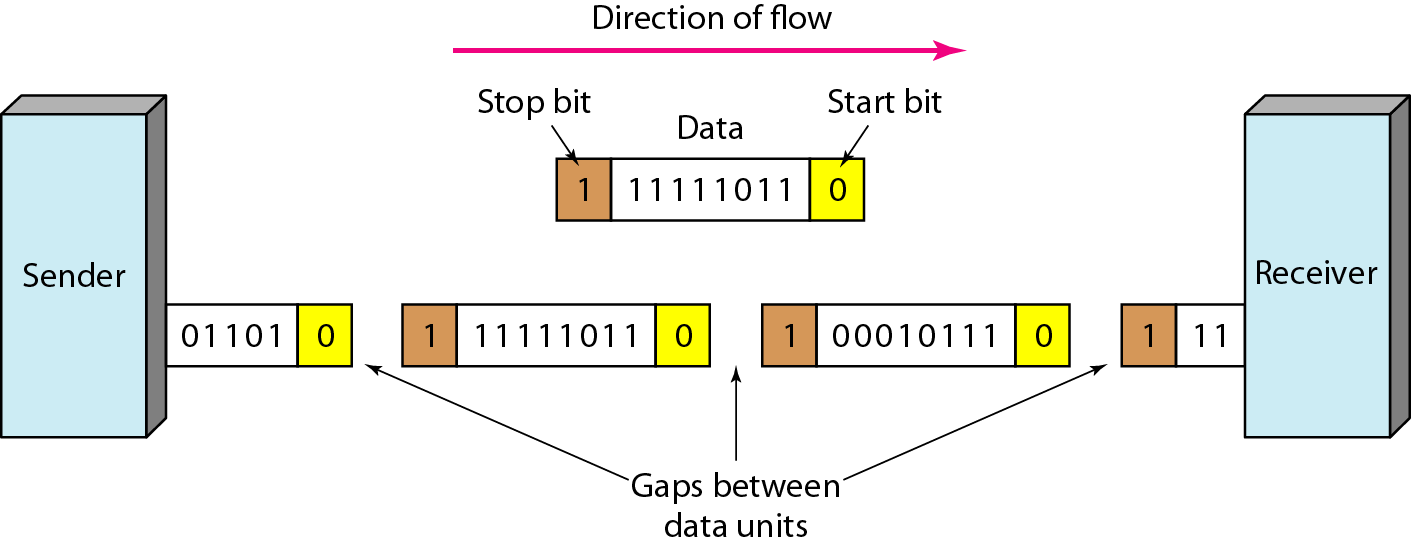
● Asynchronous transmission is so named because the timing of a signal is unimportant.

● Information is received and translated by agreed upon patterns.

● Without synchronization, the receiver cannot use timing to predict when the next group will arrive.

● To alert the receiver to the arrival of a new group, therefore, an extra bit is added to the beginning of each byte





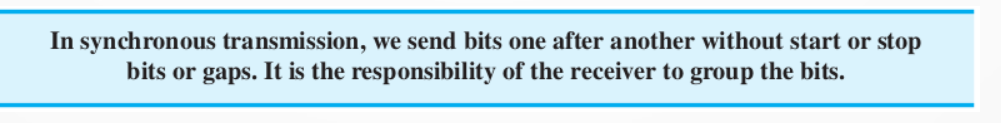
**Synchronous Transmission:**

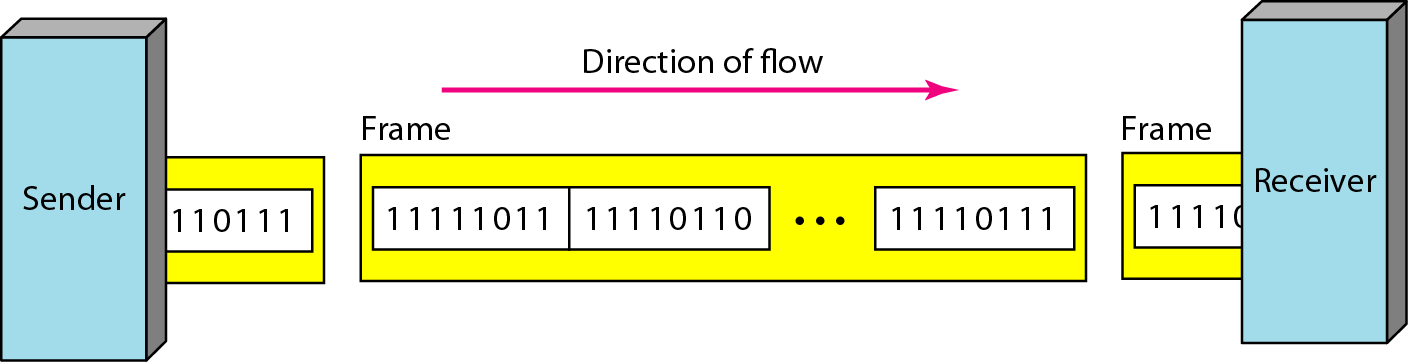
● In synchronous transmission, the bit stream is combined into longer “frames,” which may contain multiple bytes.

● Each byte, however, is introduced onto the transmission link without a gap between it and the next one.

● It is left to the receiver to separate the bit stream into bytes for decoding purposes.

● In other words, data are transmitted as an unbroken string of 1s and 0s, and the receiver separates that string into the bytes, or characters, it needs to reconstruct the information





**Isochronous Transmission:**

● In real-time audio and video, in which uneven delays between frames are not acceptable, synchronous transmission fails. – For example, TV images are broadcast at the rate of 30 images per second; they must be viewed at the same rate. – If each image is sent by using one or more frames, there should be no delays between frames.

● For this type of application, synchronization between characters is not enough; the entire stream of bits must be synchronized.

● The isochronous transmission guarantees that the data arrive at a fixed rate

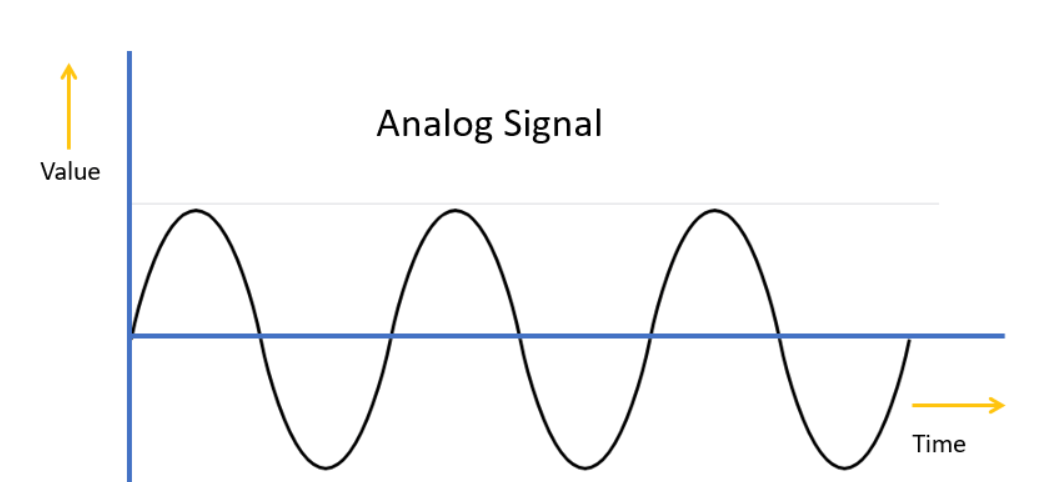
**Analog and Digital Signal:**

## Signal:

A signal is an electromagnetic or electrical current that is used for carrying data from one system or network to another. The signal is a function that conveys information about a phenomenon.

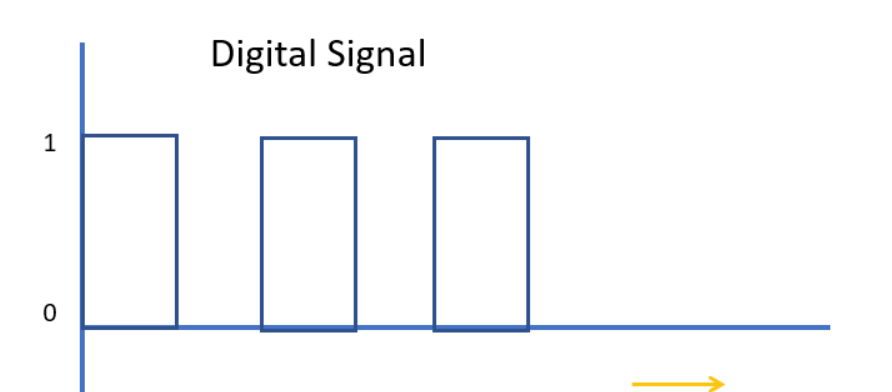
In electronics and telecommunications, it refers to any time-varying voltage that is an electromagnetic wave which carries information. A signal can also be defined as an observable change in quality such as quantity. There are two main types of signals: Analog signal and Digital signal.

## Analog Signal:



Analog signal is a continuous signal in which one time-varying quantity represents another time-based variable. These kind of signals works with physical values and natural phenomena such as earthquake, frequency, volcano, speed of wind, weight, lighting, etc.

## Digital Signal:



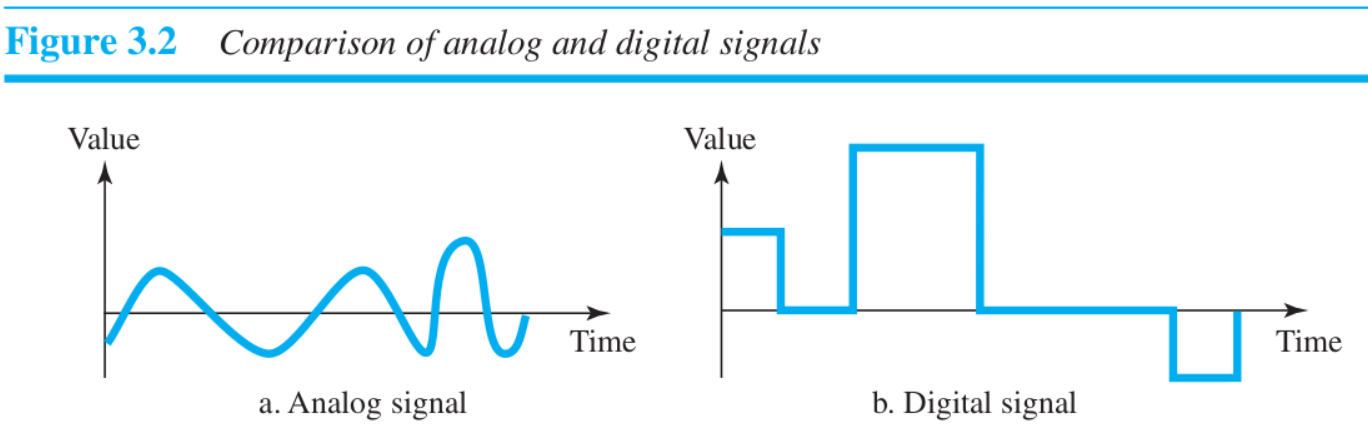
Time

A digital signal is a signal that is used to represent data as a sequence of separate values at any point in time. It can only take on one of a fixed number of values. This type of signal represents a real number within a constant range of values. Now, let’s learn some key difference between Digital and Analog signals.

Signals can be either analog or digital.

● An analog signal has infinitely many levels of intensity over a period of time. – As the wave moves from value A to value B, it passes through and includes an infinite number of values along its path.

● A digital signal, on the other hand, can have only a limited number of defined values. Although each value can be any number, it is often as simple as 1 and 0.



## KEY DIFFERENCES:

* An analog signal is a continuous signal whereas Digital signals are time separated signals.
* Analog signal is denoted by sine waves while it is denoted by square waves
* Analog signal uses a continuous range of values that help you to represent information on the other hand digital signal uses discrete 0 and 1 to represent information.
* Comparing Digital vs Analog signals, the analog signal bandwidth is low while the bandwidth of the digital signal is high.
* Analog instruments give considerable observational errors whereas Digital instruments never cause any kind of observational errors.
* Analog hardware never offers flexible implementation, but Digital hardware offers flexibility in implementation.
* Comparing Analog vs Digital signal, Analog signals are suited for audio and video transmission while Digital signals are suited for Computing and digital electronics.

## Characteristics of Analog Signal:

Here, are essential characteristics of Analog Signal

* These types of electronic signals are time-varying
* Minimum and maximum values which is either positive or negative.
* It can be either periodic or non-periodic.
* Analog Signal works on continuous data.
* The accuracy of the analog signal is not high when compared to the digital signal.
* It helps you to measure natural or physical values.
* Analog signal output form is like Curve, Line, or Graph, so it may not be meaningful to all.

## Characteristics of Digital Signals:

Here, are essential characteristics of Digital signals

* Digital signal are continuous signals
* This type of electronic l signals can be processed and transmitted better compared to analog signal.
* Digital signals are versatile, so it is widely used.

**PERIODIC & APERIODIC SIGNALS**

Data transmitted over the network can be Analog or Digital. Both Analog and Digital signals can take one of two forms:

* Periodic
* Non-Periodic (Aperiodic)

A **periodic** signal is a signal that repeats the sequence of values exactly after a fixed length of time, known as the period. It completes the pattern within a measurable time frame. The completion of one full pattern is called a cycle.

**Ex:** When a flight is detected by the radar and until the radar exists, the radar signal zone is an example of a periodic real-time task.

A **Non-periodic (Aperiodic)** changes without any pattern or cycle that repeats over time.

**Ex:** Typing on the Keyboard.

Both analog and digital signals can be periodic and aperiodic. But in data communications, **periodic analog signals** and **aperiodic digital signals** are commonly used.

**Periodic Analog Signals:**

Periodic Analog Signals can be of two types:

* **Simple** : A simple periodic analog signal is a Sine wave
* **Composite:**  A composite periodic analog signal is composed of multiple sine waves.

**Sine Wave:**

The sine wave is the most fundamental form of a periodic signal, represented in Fig.1.

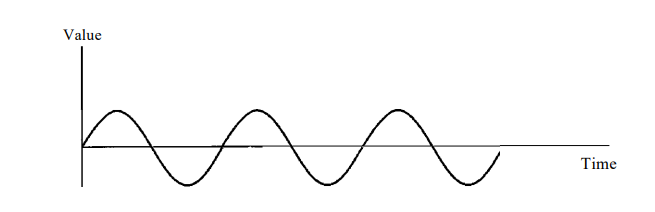


Fig.1. Sine Wave

To describe and understand the sine wave completely, it can be done using three parameters. they are:

* Peak amplitude
* Frequency
* Phase

**Peak Amplitude:**  The peak amplitude of a signal is the value of its highest intensity. It is measured in volts.

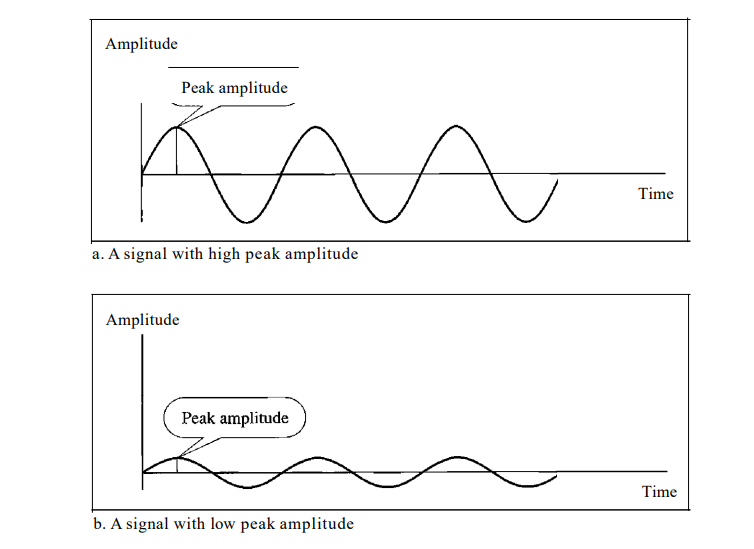


Fig.2. Two different signals with different Peak Amplitudes

**Ex:** The power in the house can be represented by sine wave with a peak amplitude of 155 to 170 volts.

**Period and Frequency:** Period refere to the amount of time needed for a signal to complete one cycle. It is measured in seconds.

Frequency is the number of periods in one second. It is measured or expressed in Hertz (Hz). Period is the inverse of frequency and frequency is the inverse of the period. The formulae is given as follows:

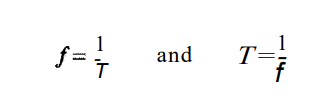


Fig.3. Formulae for Frequency and Period.

**Example:**

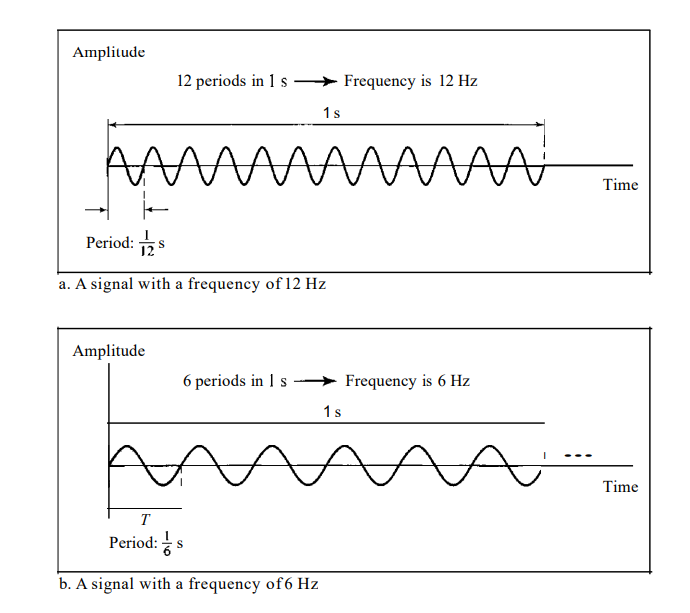


Fig.4. Example

**Phase:**  The term Phase describes the position of waveform relative to time 0. It is measured in radians.

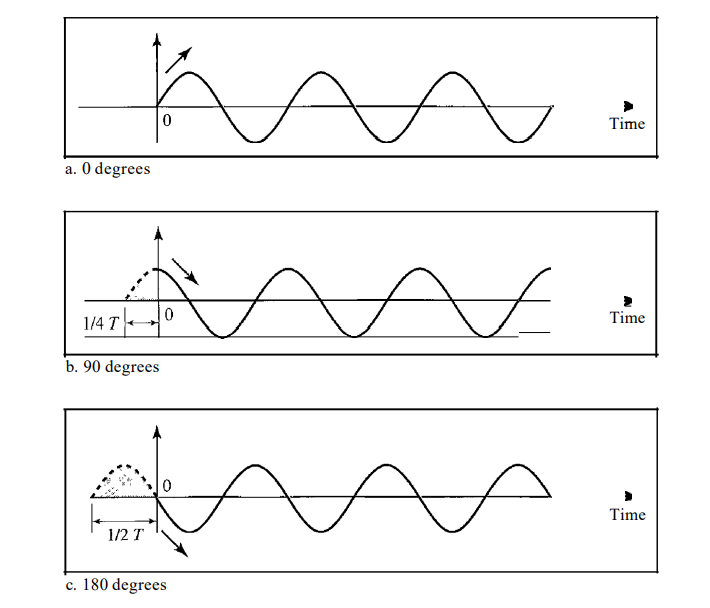


Fig.5. Three sine waves with different Phases.

From Fig.5.

* A sine wave with a phase of 0° starts at time 0 with a zero amplitude. The amplitude is increasing.
* A sine wave with a phase of 90° starts at time 0 with a peak amplitude. The amplitude is decreasing.
* A sine wave with a phase of 180° starts at time 0 with a zero amplitude. The amplitude is decreasing.

Another way to look at the phase is in terms of shift or offset. We can say that

* A sine wave with a phase of 0° is not shifted.
* A sine wave with a phase of 90° is shifted to the left by ¼ cycle.
* A sine wave with a phase of 180° is shifted to the left by ½ cycle.

**Encoding** is the process of converting the data or a given sequence of characters, symbols, alphabets etc., into a specified format, for the secured transmission of data. **Decoding** is the reverse process of encoding which is to extract the information from the converted format.

Data Encoding

Encoding is the process of using various patterns of voltage or current levels to represent **1s** and **0s** of the digital signals on the transmission link.

The common types of line encoding are Unipolar, Polar, Bipolar, and Manchester.

Encoding Techniques

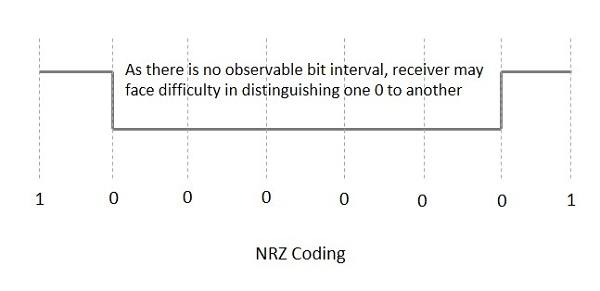
The data encoding technique is divided into the following types, depending upon the type of data conversion.

* **Analog data to Analog signals** − The modulation techniques such as Amplitude Modulation, Frequency Modulation and Phase Modulation of analog signals, fall under this category.
* **Analog data to Digital signals** − This process can be termed as digitization, which is done by Pulse Code Modulation PCMPCM. Hence, it is nothing but digital modulation. As we have already discussed, sampling and quantization are the important factors in this. Delta Modulation gives a better output than PCM.
* **Digital data to Analog signals** − The modulation techniques such as Amplitude Shift Keying ASKASK, Frequency Shift Keying FSKFSK, Phase Shift Keying PSKPSK, etc., fall under this category. These will be discussed in subsequent chapters.
* **Digital data to Digital signals** − These are in this section. There are several ways to map digital data to digital signals. Some of them are −

**Non Return to Zero NRZNRZ**

NRZ Codes has **1** for High voltage level and **0** for Low voltage level. The main behavior of NRZ codes is that the voltage level remains constant during bit interval. The end or start of a bit will not be indicated and it will maintain the same voltage state, if the value of the previous bit and the value of the present bit are same.

The following figure explains the concept of NRZ coding.



If the above example is considered, as there is a long sequence of constant voltage level and the clock synchronization may be lost due to the absence of bit interval, it becomes difficult for the receiver to differentiate between 0 and 1.

There are two variations in NRZ namely −

**NRZ - L NRZ–LEVELNRZ–LEVEL**

There is a change in the polarity of the signal, only when the incoming signal changes from 1 to 0 or from 0 to 1. It is the same as NRZ, however, the first bit of the input signal should have a change of polarity.

**NRZ - I NRZ–INVERTEDNRZ–INVERTED**

If a **1** occurs at the incoming signal, then there occurs a transition at the beginning of the bit interval. For a **0** at the incoming signal, there is no transition at the beginning of the bit interval.

NRZ codes has a **disadvantage** that the synchronization of the transmitter clock with the receiver clock gets completely disturbed, when there is a string of **1s** and **0s**. Hence, a separate clock line needs to be provided.

**Bi-phase Encoding**

The signal level is checked twice for every bit time, both initially and in the middle. Hence, the clock rate is double the data transfer rate and thus the modulation rate is also doubled. The clock is taken from the signal itself. The bandwidth required for this coding is greater.

There are two types of Bi-phase Encoding.

* Bi-phase Manchester
* Differential Manchester

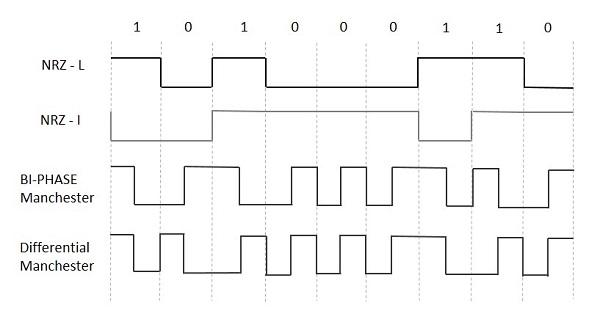
**Bi-phase Manchester**

In this type of coding, the transition is done at the middle of the bit-interval. The transition for the resultant pulse is from High to Low in the middle of the interval, for the input bit 1. While the transition is from Low to High for the input bit **0**.

**Differential Manchester**

In this type of coding, there always occurs a transition in the middle of the bit interval. If there occurs a transition at the beginning of the bit interval, then the input bit is **0**. If no transition occurs at the beginning of the bit interval, then the input bit is **1**.

The following figure illustrates the waveforms of NRZ-L, NRZ-I, Bi-phase Manchester and Differential Manchester coding for different digital inputs.



**Block Coding**

Among the types of block coding, the famous ones are 4B/5B encoding and 8B/6T encoding. The number of bits are processed in different manners, in both of these processes.

**4B/5B Encoding**

In Manchester encoding, to send the data, the clocks with double speed is required rather than NRZ coding. Here, as the name implies, 4 bits of code is mapped with 5 bits, with a minimum number of **1** bits in the group.

The clock synchronization problem in NRZ-I encoding is avoided by assigning an equivalent word of 5 bits in the place of each block of 4 consecutive bits. These 5-bit words are predetermined in a dictionary.

The basic idea of selecting a 5-bit code is that, it should have **one leading 0** and it should have **no more than two trailing 0s**. Hence, these words are chosen such that two transactions take place per block of bits.

**8B/6T Encoding**

We have used two voltage levels to send a single bit over a single signal. But if we use more than 3 voltage levels, we can send more bits per signal.

For example, if 6 voltage levels are used to represent 8 bits on a single signal, then such encoding is termed as 8B/6T encoding. Hence in this method, we have as many as 729 3636 combinations for signal and 256 2828 combinations for bits.

These are the techniques mostly used for converting digital data into digital signals by compressing or coding them for reliable transmission of data.

**RS-232C PROTOCOL**

**DATA LINK LAYER DESIGN ISSUES**

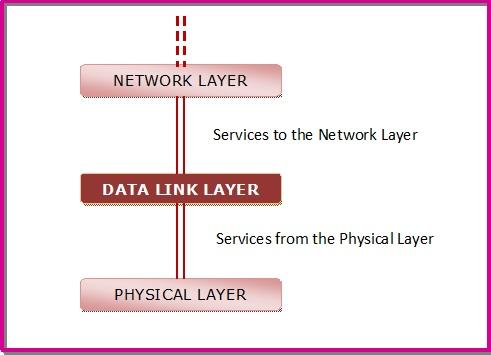
The data link layer in the OSI (Open System Interconnections) Model, is in between the physical layer and the network layer. This layer converts the raw transmission facility provided by the physical layer to a reliable and error-free link.

The main functions and the design issues of this layer are

* Providing services to the network layer
* Framing
* Error Control
* Flow Control

**Services to the Network Layer**

In the OSI Model, each layer uses the services of the layer below it and provides services to the layer above it. The data link layer uses the services offered by the physical layer.The primary function of this layer is to provide a well defined service interface to network layer above it.



The types of services provided can be of three types −

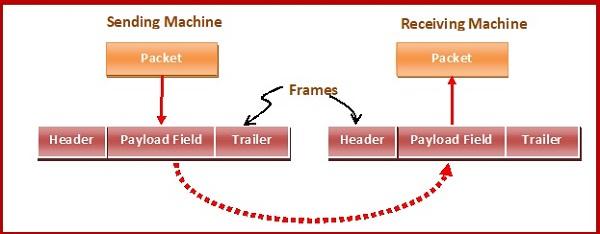
* Unacknowledged connectionless service
* Acknowledged connectionless service
* Acknowledged connection - oriented service

**Framing**

The data link layer encapsulates each data packet from the network layer into frames that are then transmitted.

A frame has three parts, namely −

* Frame Header
* Payload field that contains the data packet from network layer
* Trailer



**Error Control**

The data link layer ensures error free link for data transmission. The issues it caters to with respect to error control are −

* Dealing with transmission errors
* Sending acknowledgement frames in reliable connections
* Retransmitting lost frames
* Identifying duplicate frames and deleting them
* Controlling access to shared channels in case of broadcasting

**Flow Control**

The data link layer regulates flow control so that a fast sender does not drown a slow receiver. When the sender sends frames at very high speeds, a slow receiver may not be able to handle it. There will be frame losses even if the transmission is error-free. The two common approaches for flow control are −

* Feedback based flow control
* Rate based flow control

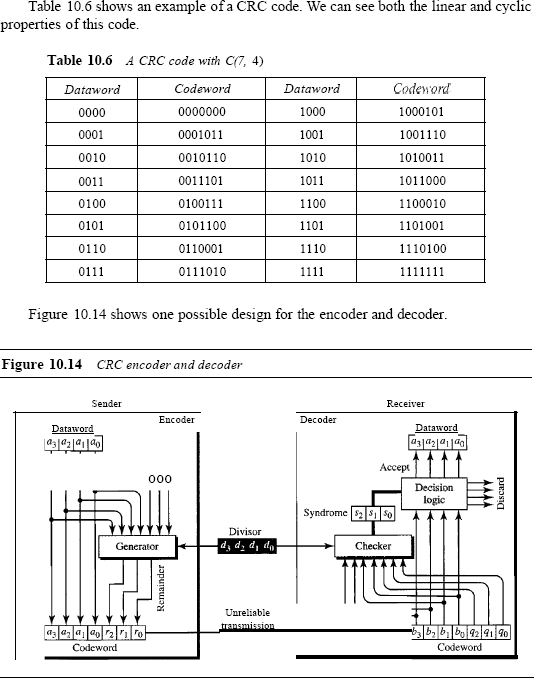
**Module 1**

**Topics: CRC codes, Elementary Data Link Layer Protocols, Stop and Wait, Sliding Window, go-back-N protocols**

**CYCLIC CODES**

Cyclic codes are special linear block codes with one extra property. In a cyclic code, if a codeword is cyclically shifted (rotated), the result is another codeword. For example, if 1011000 is a codeword and we cyclically left-shift, then 0110001 is also a codeword

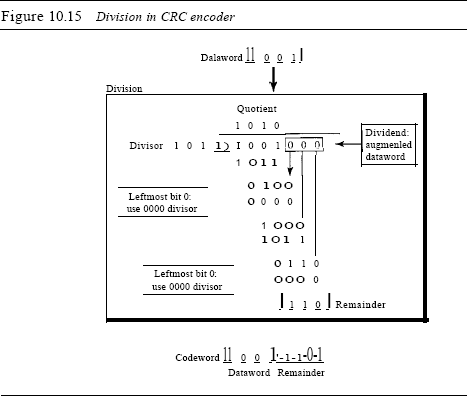
We can create cyclic codes to correct errors. Here, we discuss a category of cyclic codes called the cyclic redundancy check (CRC) that is used in networks such as LANs and WANs.



In the encoder, the dataword has *k* bits (4 here); the codeword has *n* bits. The size of the dataword is augmented by adding *n* - *k* (3 here) Os to the right-hand side of the word. The n-bit result is fed into the generator. The generator uses a divisor of size *n* - *k* + I (4 here), predefined and agreed upon. The generator divides the augmented dataword by the divisor (modulo-2 division). The quotient of the division is discarded; the remainder *(r2 rl r0)* is appended to the dataword to create the codeword. The decoder receives the possibly corrupted codeword. A copy of all *n* bits is fed to the checker which is a replica of the generator. The remainder produced by the checker is a syndrome of *n* - *k* (3 here) bits, which is fed to the decision logic analyzer. The analyzer has a simple function. If the syndrome bits are all as, the 4 leftmost bits of the codeword are accepted as the dataword (interpreted as no error); otherwise, the 4 bits are discarded (error).

. The encoder takes the dataword and augments it with *n* -

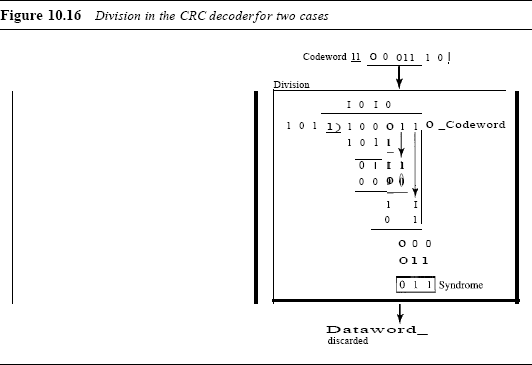
*k* number of as. It then divides the augmented dataword by the divisor, as shown in Figure 10.15.



The process of modulo-2 binary division is the same as the familiar division process we use for decimal numbers. However, as mentioned at the beginning of the chapter, in this case addition and subtraction are the same. We use the XOR operation to do both. As in decimal division, the process is done step by step. In each step, a copy of the divisor is XORed with the 4 bits of the dividend. The result of the XOR operation (remainder) is 3 bits (in this case), which is used for the next step after 1 extra bit is pulled down to make it 4 bits long. There is one important point we need to remember in this type of division. If the leftmost bit of the dividend (or the part used in each step) is 0, the step cannot use the regular divisor; we need to use an all-0s divisor. When there are no bits left to pull down, we have a result. The 3-bit remainder forms the check bits *(r2' rl'* and *ro).* They are appended to the dataword to create the codeword.

*Decoder*

The codeword can change during transmission. The decoder does the same division process as the encoder. The remainder of the division is the syndrome. If the syndrome is all 0s, there is no error; the dataword is separated from the received codeword and accepted. Otherwise, everything is discarded. Figure 10.16 shows two cases: The left hand figure shows the value of syndrome when no error has occurred; the syndrome is 000. The right-hand part of the figure shows the case in which there is one single error. The syndrome is not all 0s (it is 01l).



Advantages of Cyclic Codes

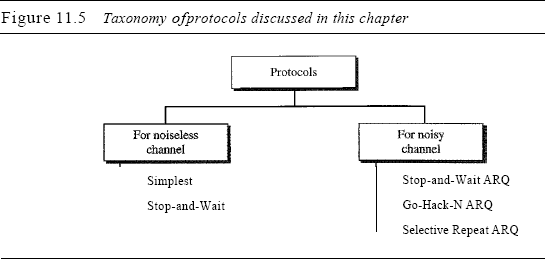
Cyclic codes have a very good performance in detecting single-bit errors, double errors, an odd number of errors, and burst errors. They can easily be implemented in hardware and software. They are especially fast when implemented in hardware. This has made cyclic codes a good candidate for many networks.

**Elementary Data Link Layer Protocols**

FLOW AND ERROR CONTROL:

Flow control coordinates the amount of data that can be sent before receiving an acknowledgment and is one of the most important duties of the data link lay

Error control is both error detection and error correction. It allows the receiver to inform the sender of any frames lost or damaged in transmission and coordinates the retransmission of those frames by the sender



**Stop-and-Wait Protocol**

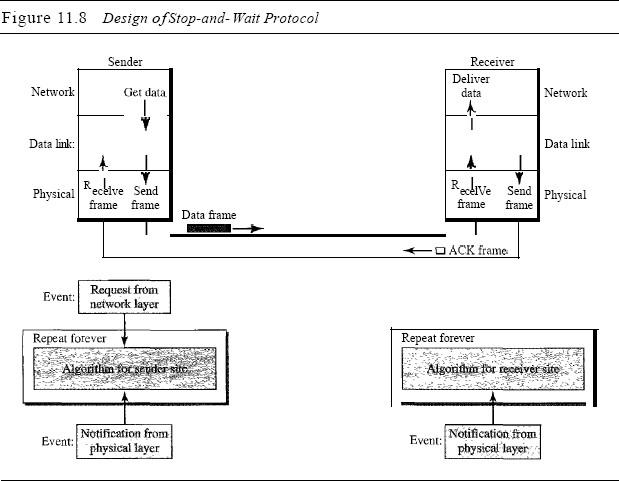
If data frames arrive at the receiver site faster than they can be processed, the frames must be stored until their use. Normally, the receiver does not have enough storage space, especially if it is receiving data from many sources. This may result in either the discarding of frames or denial of service. To prevent the receiver from becoming overwhelmed with frames. There must be feedback from the receiver to the sender. The protocol we discuss now is called the Stop-and- Wait Protocol because the sender sends one frame, stops until it receives confirmation from the receiver (okay to go ahead), and then sends the next frame.

*Design*

Figure 11.8 illustrates the mechanism. Comparing this figure with Figure 11.6, can see the traffic on the forward channel (from sender to receiver) and the reverse channel. At any time, there is either one data frame on the forward channel or one ACK frame on the reverse channel. We therefore need a half-duplex link.

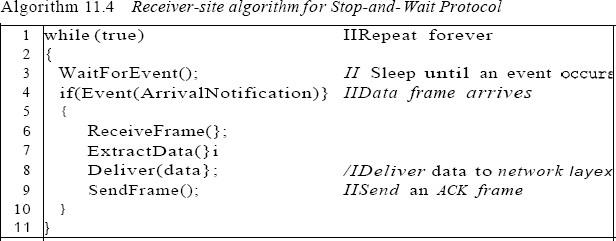
*Algorithms*

Algorithm 11.3 is for the sender site.



Analysis Here two events can occur: a request from the network layer or an arrival notification from the physical layer. The responses to these events must alternate. In other words, after a frame is sent, the algorithm must ignore another network layer request until that frame is acknowledged. We know that two arrival events cannot happen one after another because the channel is error-free and does not duplicate the frames. The requests from the network layer, however, may happen one after another without an arrival event in between. To prevent the immediate sending of the data frame. there are several methods used a simple *canSend* variable that can either be true or false. When a frame is sent, the variable is set to false to indicate that a new network request cannot be sent until *can Send* is true. When an ACK is received, can Send is set to true to allow the sending of the next frame.

Algorithm 11.4 shows the procedure at the receiver site.



* 1. NOISY CHANNELS

### Stop-and-Wait Automatic Repeat Request

Our first protocol, called the Stop-and-Wait Automatic Repeat Request (Stop-and Wait ARQ), adds a simple error control mechanism to the Stop-and-Wait Protocol. Let us see how this protocol detects and corrects errors. To detect and correct corrupted frames, we need to add redundancy bits to our data frame. When the frame arrives at the receiver site, it is checked and if it is corrupted, it is silently discarded. The detection of errors in this protocol is manifested by the silence of the receiver.

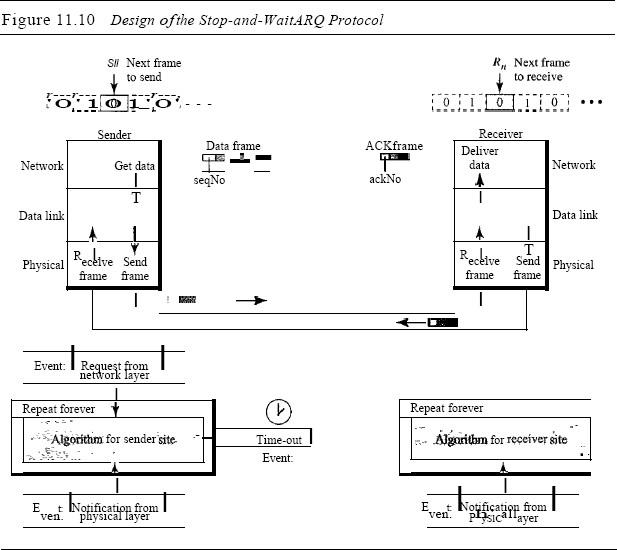
Lost frames are more difficult to handle than corrupted ones. In our previous protocols, there was no way to identify a frame. The received frame could be the correct one, or a duplicate, or a frame out of order. The solution is to number the frames. When the receiver receives a data frame that is out of order, this means that frames were either lost or duplicated.

The completed lost frames need to be resent in this protocol. If the receiver does not respond when there is an error, how can the sender know which frame to resend? To remedy this problem, the sender keeps a copy of the sent frame. At the same time, it starts a timer. If the timer expires and there is no ACK for the sent frame, the frame is resent, the copy is held, and the timer is restarted. Since the protocol uses the stop-and-wait mechanism, there is only one specific frame that needs an ACK even though several copies of the same frame can be in the network.

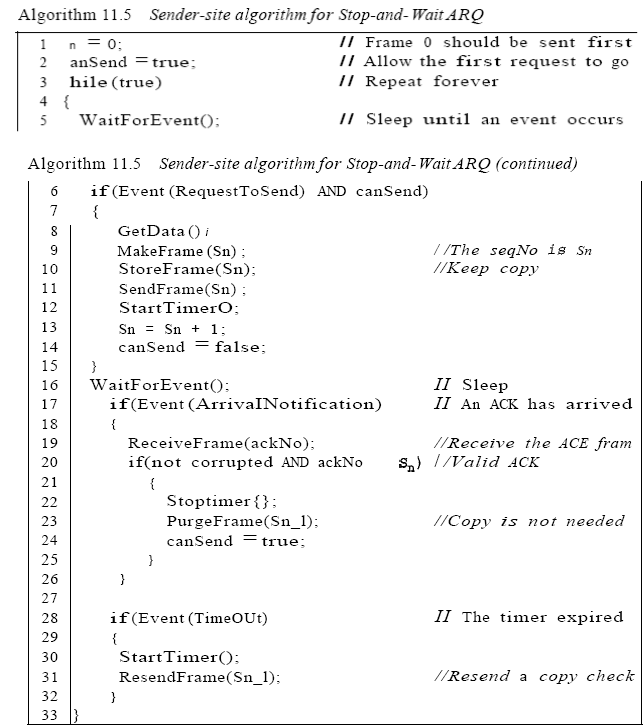
Since an ACK frame can also be corrupted and lost, it too needs redundancy bits and a sequence number. The ACK frame for this protocol has a sequence number field. In this protocol, the sender simply discards a corrupted ACK frame or ignores an out-of-order one.

*Design*

Figure 11.10 shows the design of the Stop-and-WaitARQ Protocol. The sending device keeps a copy of the last frame transmitted until it receives an acknowledgment for that frame. A data frames uses a seqNo (sequence number); an ACK frame uses an ackNo (acknowledgment number). The sender has a control variable, which we call *Sn* (sender, next frame to send), that holds the sequence number for the next frame to be sent (0 or 1).



The receiver has a control variable, which we call *Rn* (receiver, next frame expected), that holds the number of the next frame expected. When a frame is sent, the value of *Sn* is incremented (modulo-2), which means if it is 0, it becomes 1 and vice versa. When a frame is received, the value of *Rn* is incremented (modulo-2), which means if it is 0, it becomes 1 and vice versa. Three events can happen at the sender site; one event can happen at the receiver site. Variable *Sn* points to the slot that matches the sequence number of the frame that has been sent, but not acknowledged; *Rn* points to the slot that matches the sequence number of the expected frame.

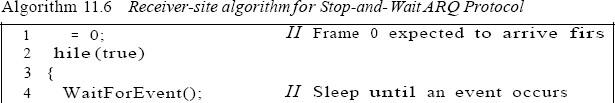


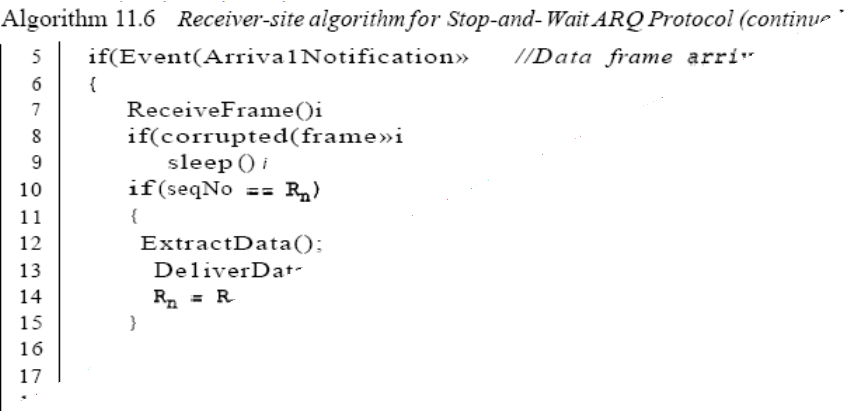
Analysis We first notice the presence of *Sn'* the sequence number of the next frame to be sent. This variable is initialized once (line 1), but it is incremented every time a frame is sent (line 13) in preparation for the next frame. However, since this is modulo-2 arithmetic, the sequence numbers are

0, 1,0, 1, and so on. Note that the processes in the first event (SendFrame, StoreFrame, and Purge Frame) use an *Sn* defining the frame sent out. We need at least one buffer to hold this frame until we are sure that it is received safe and sound. Line 10 shows that before the frame is sent, it is stored.

The copy is used for resending a corrupt or lost frame. We are still using the canSend variable to prevent the network layer from making a request before the previous frame is received safe and sound. If the frame is not corrupted and the ackNo of theACK frame matches the sequence number of the next frame to send, we stop the timer and purge the copy of the data frame we saved. Otherwise, we just ignore this event and wait for the next event to happen. After each frame is sent, a timer is started.

When the timer expires (line 28), the frame is resent and the timer is restarted.

Algorithm 11.6 shows the procedure at the receiver site.



Analysis This is noticeably different from Algorithm 11.4. First, all arrived data frames that are corrupted are ignored. If the SeqNo of the frame is the one that is expected *(Rn ),* the frame is accepted, the data are delivered to the network layer, and the value of *Rn* is incremented. However, there is one subtle point here. Even if the sequence number of the data frame does not match the next frame expected, an ACK is sent to the sender. This ACK, however, just reconfirms the previous ACK instead of confirming the frame received. This is done because the receiver assumes that the previous ACK might have been lost; the receiver is sending a duplicate frame. The resent ACK may solve the problem before the time-out does it.

*Efficiency*

The Stop-and-Wait ARQ discussed in the previous section is very inefficient if our channel

is *thick* and *long.* By *thick,* we mean that our channel has a large bandwidth; by *long,* mean the round-trip delay is long. The product of these two is called the bandwidth delay product, as we discussed in Chapter 3. We can think of the channel as a pipe. The bandwidth-delay product then is the volume of the pipe in bits. The pipe is always there. If we do not use it, we are inefficient. The bandwidth-delay product is a measure of the number of bits we can send out of our system while waiting for news from the receiver.

*Pipelining*

In networking and in other areas, a task is often begun before the previous task has ended.

This is known as pipelining. There is no pipelining in Stop-and-Wait ARQ because need to wait for a frame to reach the destination and be acknowledged before the next frame can be sent. However, pipelining does apply to our next two protocols because several frames can be sent before we receive news about the previous frames. Pipelining improves the efficiency of the transmission if the number of bits in transition is large with respect to the bandwidth-delay product.

Go-Back-N Automatic Repeat Request

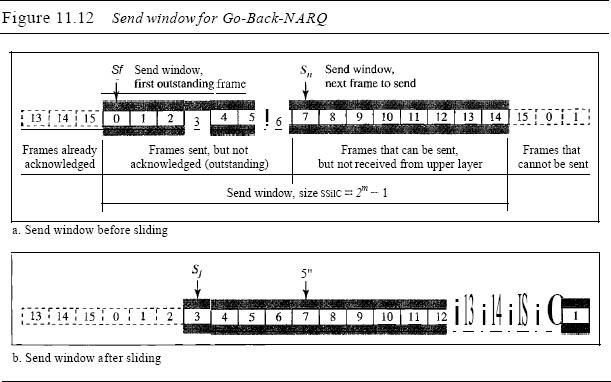
To improve the efficiency of transmission (filling the pipe), multiple frames must be in transition while waiting for acknowledgment. In other words, we need to let more than one frame be outstanding to keep the channel busy while the sender is waiting for acknowledgment. In this section, we discuss one protocol that can achieve this goal; in the next section, we discuss a second. The first is called Go-Back-N Automatic Repeat Request (the rationale for the name will become clear later). In this protocol we can send several frames before receiving acknowledgments; we keep a copy of these frames until the acknowledgments arrive.

*Sequence Numbers*

Frames from a sending station are numbered sequentially. However, because we need to include the sequence number of each frame in the header, we need to set a limit. If the header of the frame allows *m* bits for the sequence number, the sequence numbers range from 0 to *2m* - 1. For example, if *m* is 4, the only sequence numbers are 0 through 15 inclusive. However, we can repeat the sequence. So the sequence numbers are 0, 1,2,3,4,5,6, 7,8,9, 10, 11, 12, 13, 14, 15,0, 1,2,3,4,5,6,7,8,9,10, 11, ... In other words, the sequence numbers are modulo-2m

*Sliding Window*

In this protocol (and the next), the sliding window is an abstract concept that defines the range of sequence numbers that is the concern of the sender and receiver. In other words, the sender and receiver need to deal with only part of the possible sequence numbers. The range which is the concern of the sender is called the send sliding window; the range that is the concern of the receiver is called the receive sliding window. We discuss both here. The send window is an imaginary box covering the sequence numbers of the data frames which can be in transit. In each window position, some of these sequence numbers define the frames that have been sent; others define those that can be sent. The maximum size of the window is *2m* - 1 for reasons that we discuss later. In this chapter, we let the size be fixed and set to the maximum value, but we will see in future chapters that some protocols may have a variable window size. Figure 11.12 shows a sliding window of size 15 *(m* =4). The window at any time divides the possible sequence numbers into four regions. The first region, from the far left to the left wall of the window, defines the sequence



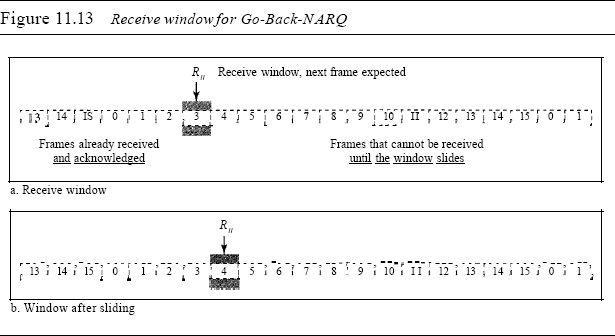
numbers belonging to frames that are already acknowledged. The sender does not worry about these frames and keeps no copies of them. The second region, colored in Figure 11.12a, defines the range of sequence numbers belonging to the frames that are sent and have an unknown status. The sender needs to wait to find out if these frames have been received or were lost. We call these outstanding frames. The third range, white in the figure, defines the range of sequence numbers for frames that can be sent; however, the corresponding data packets have not yet been received from the network layer. Finally, the fourth region defines sequence numbers that cannot be used until the window slides, as we see next.

The window itself is an abstraction; three variables define its size and location at any time. We call these variables *Sf(send* window, the first outstanding frame), *Sn* (send window, the next frame to be sent), and Ssize (send window, size). The variable *Sf* defines the sequence number of the first (oldest) outstanding frame. The variable *Sn* holds the sequence number that will be assigned to the next frame to be sent. Finally, the variable Ssize defines the size of the window, which is fixed in our protocol.

Figure 11.12b shows how a send window can slide one or more slots to the right when an acknowledgment arrives from the other end. As we will see shortly, theacknowledgments

in this protocol are cumulative, meaning that more than one frame can be acknowledged by an ACK frame. In Figure 11.12b, frames 0, I, and 2 are acknowledged,

so the window has slid to the right three slots. Note that the value of*Sf* is 3 because frame 3 is now the first outstanding frame. The receive window makes sure that the correct data frames are received and that the correct acknowledgments are sent. The size of the receive window is always 1. The receiver is always looking for the arrival of a specific frame. Any frame arriving out of order is discarded and needs to be resent. Figure 11.13 shows the receive window.



Note that we need only one variable *Rn* (receive window, next frame expected) to define this abstraction. The sequence numbers to the left of the window belong to the frames already received and acknowledged; the sequence numbers to the right of this window define the frames that cannot be received. Any received frame with a sequence number in these two regions is discarded. Only a frame with a sequence number matching the value of *Rn* is accepted and acknowledged. The receive window also slides, but only one slot at a time. When a correct frame is received (and a frame is received only one at a time), the window slides.

*Timers*

Although there can be a timer for each frame that is sent, in our protocol we use only one. The reason is that the timer for the first outstanding frame always expires first; we send all outstanding frames when this timer expires.

*Acknowledgment*

The receiver sends a positive acknowledgment if a frame has arrived safe and sound and in order. If a frame is damaged or is received out of order, the receiver is silent and

will discard all subsequent frames until it receives the one it is expecting. The silence of

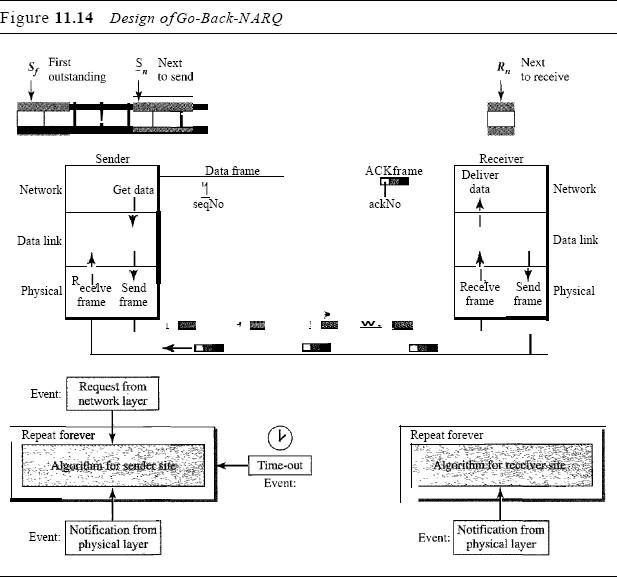
the receiver causes the timer of the unacknowledged frame at the sender site to expire. This, in turn, causes the sender to go back and resend all frames, beginning with the one with the expired timer. The receiver does not have to acknowledge each frame received. It can send one cumulative acknowledgment for several frames.

*Resending a Frame*

When the timer expires, the sender resends all outstanding frames. For example, suppose the sender has already sent frame 6, but the timer for frame 3 expires. This means that frame 3 has not been acknowledged; the sender goes back and sends frames 3, 4,5, and 6\ again. That is why the protocol is called *Go-Back-N* ARQ.

*Design*

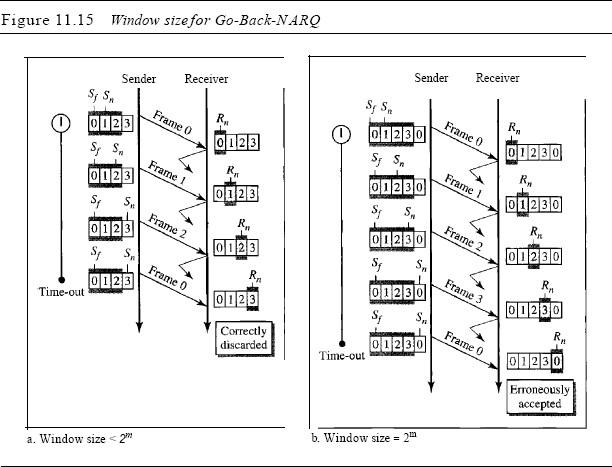
Figure 11.14 shows the design for this protocol. As we can see, multiple frames can be in transit in the forward direction, and multiple acknowledgments in the reverse direction. The idea is similar to Stop-and-Wait ARQ; the difference is that the send



Window allows us to have as many frames in transition as there are slots in the send window.

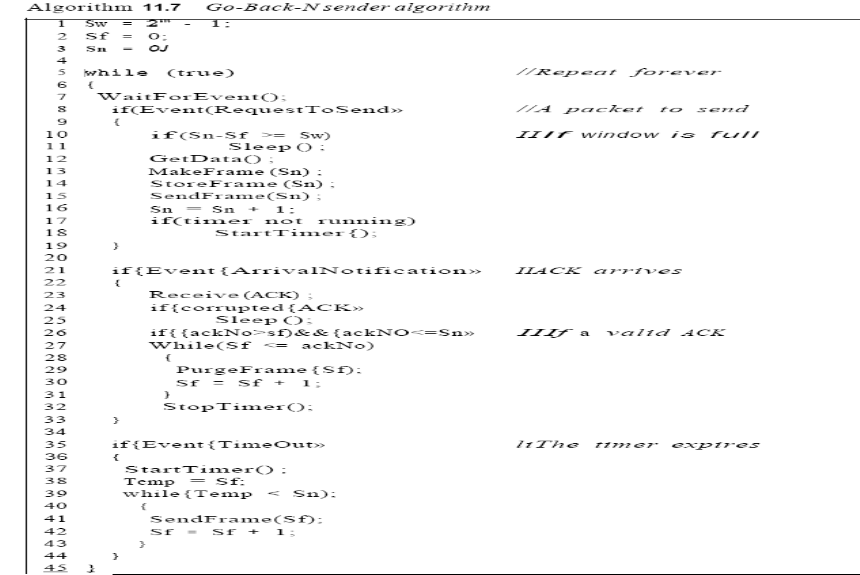
*Send Window Size*

We can now show why the size of the send window must be less than *2m.* As an example, we choose *m* =2, which means the size of the window can be *2m* - 1, or 3. Figure 11.15 compares a window size of 3 against a window size of 4. If the size of the window is (less than 22) and all three acknowledgments are lost, the frame timer expires and all three frames are resent. The receiver =is now expecting frame 3, not frame 0, so the duplicate frame is correctly discarded. On the other hand, if the size of the window is 4 (equal to 22) and all acknowledgments are lost, the sender will send a duplicate of frame 0. However, this time the window of the receiver expects to receive frame 0, so it accepts frame 0, not as a duplicate, but as the first frame in the next cycle. This is an error.



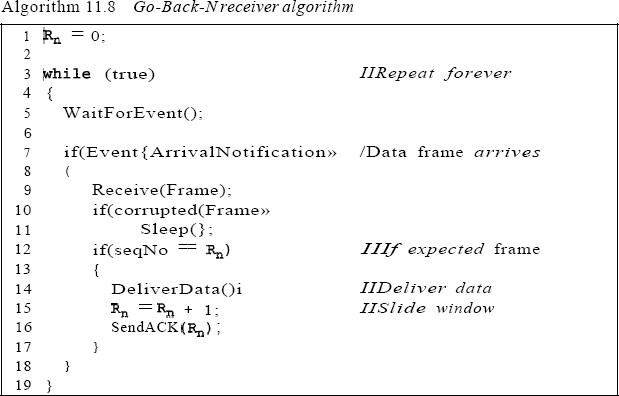
*Algorithms*

Algorithm 11.7 shows the procedure for the sender in this protocol



Analysis This algorithm first initializes three variables. Unlike Stop-and-Wait ARQ, this protocol allows several requests from the network layer without the need for other events to occur; we just need to be sure that the window is not full (line 12). In our approach, if the window is full, the request is just ignored and the network layer needs to try again. Some implementations use other methods such as enabling or disabling the network layer. The handling of the arrival event is more complex than in the previous protocol. If we receive a corrupted ACK, we ignore it. If the ade Na belongs to one of the outstanding frames, we use a loop to purge the buffers and move the left wall to the right. The time-out event is also more complex.

Algorithm 11.8 is the procedure at the receiver site.

 Analysis This algorithm is simple. We ignore a corrupt or out-of-order frame. If a frame

arrives with an expected sequence number, we deliver the data, update the value of*Rn,* and send an ACK with the ackNa showing the next frame expected.

*Go-Back-N ARQ Versus Stop-and- Wait ARQ*

You may find that there is a similarity between *Go-Back-N*ARQ and Stop-and-Wait ARQ. We can say that the Stop-and-WaitARQ Protocol is actually a *Go-Back-N*ARQ in which there are only two sequence numbers and the send window size is 1. In other words, *m* = 1, *2m* - 1 = 1. In *Go-Back-N*ARQ, we said that the addition is modulo-2m; in Stop-and-WaitARQ it is 2, which is the same as *2m* when *m* = 1.