NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

Mobile Computing-Architecture and Prototype Modem Implementation

Graduation Project
COM 400

Student: Mohammed Dauod (20001100)

Supervisor: Mr. Jamal Fathi

Nicosia 2005

ACKNOWLEDGMENT

I could not have prepared this project without the generous help of my supervisor, colleaques, friends, and family.

My deepest thanks are to my supervisor Mr. Jamal Fathi for his help and answering any question I asked him.

I would like to express my gratitude to Prof. Dr. Fakhraddin Mamedov.

Also I would like to express my gratitude to Mr. Tayseer Alshanableh and his family.

Finally, I could never have prepared this project without the encouragement and support of my mum, brothers and sister.

ABSTRACT

The mobile station (MS) consists of the mobile equipment (the terminal) and a smart card called the Subscriber Identity Module (SIM). The SIM provides personal mobility, so that the user can have access to subscribed services irrespective of a specific terminal. By inserting the SIM card into another GSM terminal, the user is able to receive calls at that terminal, make calls from that terminal, and receive other subscribed services.

The GSM technical specifications define the different entities that form the GSM network by defining their functions and interface requirements.

Each mobile uses a separate, temporary radio channel to talk to the cell site. The cell site talks to many mobiles at once, using one channel per mobile. Channels use a pair of frequencies for communication—one frequency (the forward link) for transmitting from the cell site and one frequency (the reverse link) for the cell site to receive calls from the users. Radio energy dissipates over distance, so mobiles must stay near the base station to maintain communications. The basic structure of mobile networks includes telephone systems and radio services. Where mobile radio service operates in a closed network and has no access to the telephone system, mobile telephone service allows interconnection to the telephone network.

CONTENTS

INTRODUCTION

The project of connection between computer and mobile consists of introduction , 4 chapters and conclusion.

Chapter One describes the transmission media and their impairments , guided , unguided media , fiber optic cables , optical transmitters and transmitter and receiver circuit.

Chapter Two describes all types of encoding in communication , digital-digital , analog-digital , digital-analog and analog-analog.

Chapter Three describes the transmission of digital data : interfaces and modems.

Chapter Four describes high speed wireless LAN for mobile computing - architecture and prototype modem implementation.

Conclusion presents the optained important results and contributions in the project.

1. TRANSMISSION MEDIA

1.1 Mathematical Models for Communication Channels

In the design of communication systems we find it convenient to construct mathematical models that reflect the most important characteristics of the transmission medium. Below, we provide a brief description of the channel models that are frequently used to characterize many of the physical channels that we encounter in practice.

The additive noise channel. The simplest mathematical model for a communication channel is the additive noise channel, illustrated in Figure 1.1.

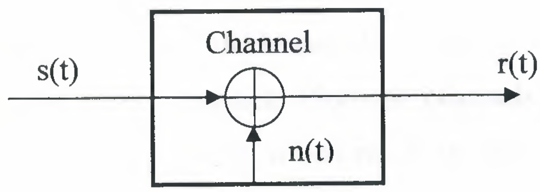


Figure 1.1 Mathematical model for communication channel.

In this model, the transmitted signal s *(t)* is corrupted by an additive random noise process *n(t)*. Physically, the additive noise process may arise from electronic components and amplifiers at the receiver of the communication system, or from interference encountered in transmission as in the case of radio signal transmission.

If the noise is introduced primarily by electronic components and amplifiers at the receiver, it may be characterized as thermal noise. This type of noise is characterized statistically as a *Gaussian noise process.* Hence, the resulting mathematical model for the channel is usually called the *additive Gaussian noise channel.* In this case the received signal is

$$r(t) = \alpha s(t) + n(t)$$

Where α represents the attenuation factor.

The linear filter channel. In some physical channels such as wire-line telephone channels, filters are used to ensure that the transmitted signals do not exceed specified bandwidth limitations and thus do not interfere with one another. Such channel (Figure 1.2) output can be characterized as

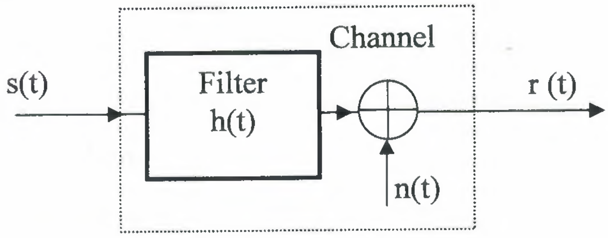


Figure 1.2 Output channel characterization

$$r(t) = s(t) * h(t) + n(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)d\tau + n(t)$$

Where h(t) is the impulse response of the linear filter and symbol * denotes convolution.

The linear time-variant filter channel. Physical channels such as underwater acoustic channels and ionosphere radio channels, which result in time-variant multi-path propagation of the transmitted signal, may be characterized mathematically as time-variant linear filters. Such system is characterized by a time-variant channel with impulse response h (τ; t) filters (Figure 1.3). For an input signal s (t), the channel output is

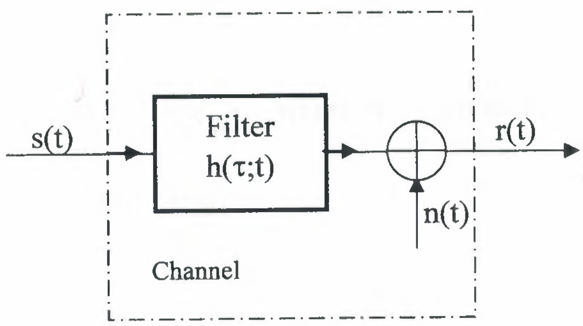


Figure 1.3 Time-variant channel with impulse response.

$$r(t) = s(t) * h(\tau ; t) + n(t)$$

The three mathematical models described above adequately characterise a majority of physical channels encountered in practice.

1.2 Transmission Impairments

The transmission medium is the physical path between transmitter and receiver. The characteristics and quality of data transmission are determined both by the nature of the signal and the nature of the medium.

With any communication system, it must be recognized that the signal that is received will differ from the signal that is transmitted due to various transmission impairments. For analog signals, these impairments introduce various random modifications that degrade the signal quality. For digital signals, bit errors are introduced: a binary 1 is transformed into a binary 0 and vice versa.

1.2.1 Attenuation

As a signal propagates along a transmission medium its amplitude decreases. This is known as signal attenuation. To compensate the attenuation, amplifiers are inserted at intervals along the cable to restore the received signal to its original level. Signal attenuation increases as a function of frequency. To overcome this problem, the amplifiers are designed to amplify different frequency by varying gains of amplifications. These devices are known as equaliser. For guided media (Twisted wires, Coaxial cables and Fiber optic cables) attenuation, is generally logarithmic and it is typically expressed as a constant number of decibels per unit distance

$$N, dB = 10\log\frac{P_2}{P_1}, \quad \text{where } N - \text{number of decibels}$$

P_1, P_2 – input and output powers.

$$N, dB = 20\log\frac{U_2}{U_1}$$

For unguided media attenuation is a more complex function of distance and the make-up of the atmosphere. An example is shown in Figure 1.4, which shows attenuation as a function of frequency for a typical wire line. In Figure 1.4, attenuation is measured relative to the attenuation at 1000 Hz. Positive values on the y-axis represent attenuation greater than that at 1000 Hz. For any other frequency f, the relative attenuation in decibels is $N_f = 10 \log_{10} P_f / P_{1000}$ .The solid line in Figure shows attenuation without equalization. The dashed line shows the effects of equalization.

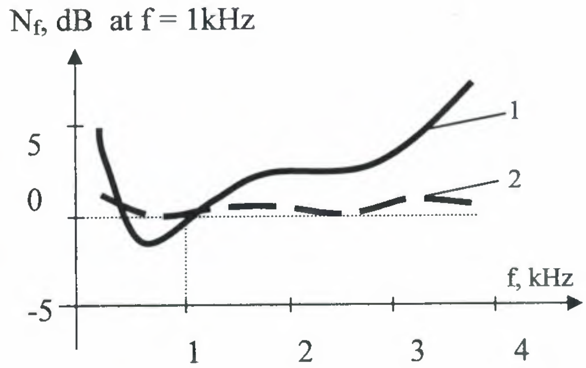


Figure 1.4. Attenuation without Equalization.

1.2.2 Delay Distortion

Delay distortion is a phenomenon peculiar to guided transmission media. The distortion is caused by the fact that the velocity of propagation of a signal through a guided medium varies with frequency. This effect is referred to as delay distortion, since the received signal is distorted due to variable delay in its components. Delay distortion is particularly critical for digital data. Consider that a sequence of bits is being transmitted, using either analog or digital signals. Because of delay distortion, some of the signal components of one bit position will spill over into other bit positions, causing inter-symbol interference, which is a major limitation to maximum bit rate over a transmission control. Equalizing techniques can also be used for delay distortion.

1.2.3 Noise

For any data transmission, the received signal will consist of the transmitted signal, modified by the various distortions imposed by the transmission system, plus additional unwanted signals that are inserted somewhere between transmission and reception. These undesired signals are referred to *Noise* and can be divided into four categories: Thermal noise, Inter-modulation noise, and Cross-talk and Impulse noise.

The Thermal noise is due to thermal agitation of electrons in a conductor. It is present in all electronic devices and transmission media and is a function of temperature. Thermal noise is uniformly distributed across the frequency spectrum and hence is often referred to as *white noise*. Thermal noise cannot be eliminated and therefore places an upper bound on communications system performance. This noise is assumed to be independent of frequency. The thermal noise in watts present in a bandwidth of W-hertz can be expressed as

$$N = kTW$$

Or, in decibel-watts:

$N = 10 \log k + 10 \log T + 10 \log W$

$N = -228.6 \text{ (dbW)} + 10 \log T + 10 \log W$

Where No - noise power density, watts/hertz;

k - Boltzmann's constant $k = 1.3803 \times 10^{-23}$ $J/^{0}K$; T - temperature, degrees Kelvin

When signals at different frequencies share the same transmission medium, the result may be *inter-modulation noise*. The effect of inter-modulation noise is to produce signals at a frequency, which is the sum or difference of the two original frequencies or multiples of those frequencies. For example, the mixing of signals at frequencies f_1 and f_2 might produce energy at the frequency $f_1 + f_2$. This derived signal could interfere with an intended signal at the frequency $f_1 + f_2$.

Inter-modulation noise is produced when there is some non-linearity in the transmitter, receiver, or interviewing transmission system.

Cross-talk has been experienced by anyone who, while using the telephone, he/she is able to hear another conversation: it is an unwanted coupling between signal paths. It can occur by electrical coupling between nearby twisted pair or rarely coaxial cable lines carrying multiple signals. Among several types of cross-talk the most limiting impairment for data communication systems is near-end cross-talk (self-cross-talk or echo), since it is caused by the strong signal output by the transmitter output being coupled with much weaker signal at the input of the local receiver circuit. Adaptive noise canceller is used to overcome this type of impairment.

4 *Impulse noise*, has short duration and have relatively high amplitude. It is generated from a variety of causes, including external electromagnetic disturbances, such as lightning, electrical impulses associated with the switching circuits used in the telephone exchange.

Impulse noise is generally only a minor annoyance for analog data. For example, voice transmission can be corrupted by short clicks and crackles with no loss of intelligibility. However, impulse noise is the primary source of error in digital data communication. For example, impulse noise of 0.01 s duration would not destroy any voice data, but would wash out about 50 bits of data is being transmitted at 4800 bps.

1.3 Channel Capacity

The rate at which data can be transmitted over a given communication channel, under given conditions, is referred to as the channel capacity.

There are four concepts here that we are trying to relate to one another.

- Data rate: This is the rate, in bits per second (bps), at which data can be transmitted.
- Bandwidth: This is the bandwidth of the transmitted signal as constrains by the transmitter and the nature of the transmission medium, expressed by Hertz.
- Noise: The average level of noise over the communications path.
- Error rate: The rate at which errors occur, where an error is the reception of a 1 when a 0 was transmitted or the reception of a 0 when a 1 was transmitted.

Communication facilities are expensive and, in general, the greater the bandwidth of a facility the greater the cost. Furthermore, all transmission channels of any practical interest are of limited bandwidth. The limitations arise from the physical properties of the transmission medium or from deliberate limitations at the transmitter on the bandwidth to prevent interference from other sources. Accordingly, we would like to make as efficient use as possible of a given bandwidth.

Let us consider the case of a channel that is noise-free. In this environment, the limitation on data rate is simply the bandwidth of the signal. A formulation of this limitation, due to Nyquist, states that if the rate of signal transmission is 2W, then a signal with frequencies no greater than W is sufficient to carry the data rate. The conserve is also true: Given a bandwidth of W, the highest signal rate that can be carried is 2W.

However, as we shall see in chapter 3, signals with more than two levels can be used; that is each signal element can represent more than one bit. For example; if M possible voltage levels are used, then each signal element can be represented by $n = \log_2 M$ numbers of bits.

With multilevel signaling, the Nyquist formulation becomes

$$C = 2\ W \log_2 M$$

Thus, for M = 8, a value used with some modems, C becomes 18600 bps.

An important parameter associated with a channel is a signal-to-noise ratio (SNR) expressed as

$$SNR = 10\log_{10} (S/N)\ dB$$

Where S/N – signal –to- noise powers ratio. Clearly a high S/N will mean a high - quality signal and a low number of required intermediate repeaters.

The signal - to noise ratio is important in the transmission of digital data because it sets the upper bound on the achievable data rate. The maximum channel capacity, in bits per second, obeys the equation attributed as the Shannon – Hartley law

$$C = W \log_2 (1+ S/N) \approx 3{,}32\ W\log_{10} (1+S/N),$$

1.4 Guided Media

The guided media includes: twisted pair, coaxial cable and fiber-optic cable (see Figure 1.5).

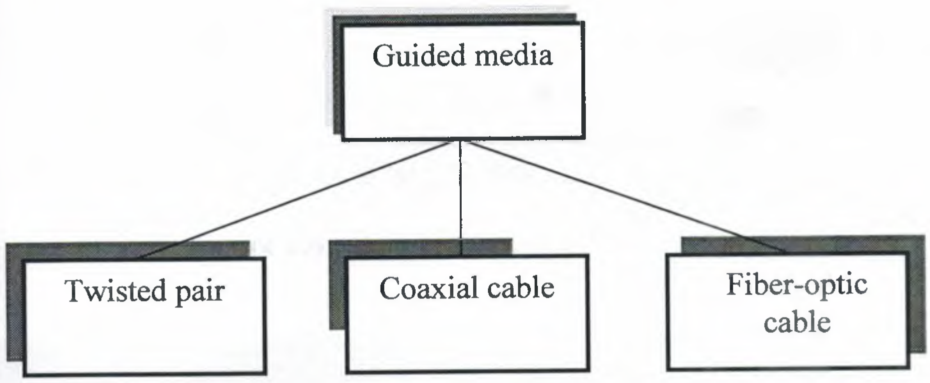


Figure 1.5 Categories of Guided Media

Table 1.1 contains the typical characteristics for guided media

Table 1.1 Typical characteristics for guided media

Medium Transmission	Total Data Rate	Bandwidth	Repeater Spacing
Twisted pair	1-100 Mbps	100Hz -5 MHz	2 - 10 km
Coaxial cable	1Mbps-1 Gbps	100 Hz – 500 MHz	1 - 10 km
Optical fiber	2 Gbps	2 GHz	10- 10 0 km

In the past two parallel flat wires were used for communications. Each wire is insulated from the other and both are open to free space. This type of line is used for connecting equipment that is up to 50 m apart using moderate rate (less than 20 kbps). The signal, typically a voltage or current level relative to some ground reference is applied to one wire while the ground reference is applied to the other. Although a two wire open line can be used to connect two computers directly, it is used mainly for connecting computers with modems. As shown in Figure 1.6 two simple wires more sensitive to noise interference.

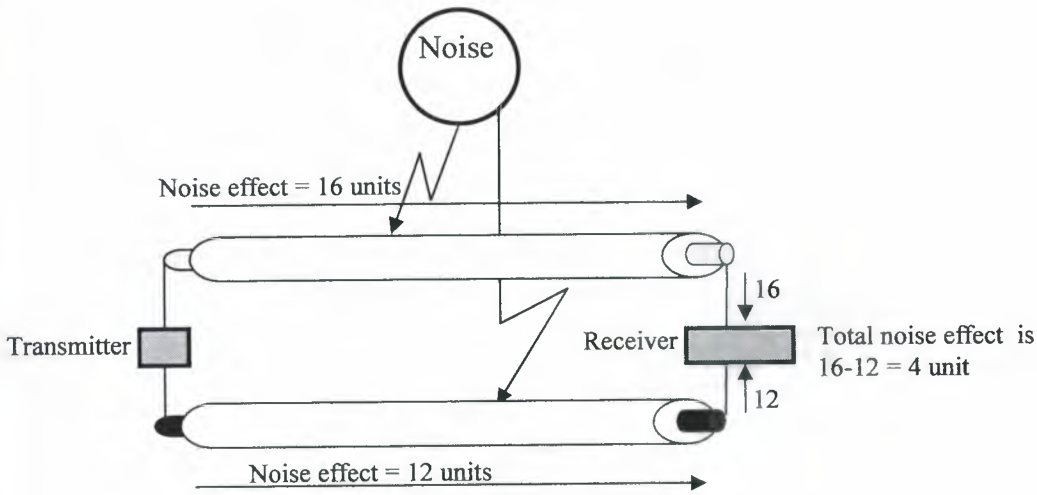


Figure 1.6 Effect of noise in parallel lines.

Figure 1.7 (a) Layouts of the coaxial cables, (b) Twisted pairs and (c) Fiber-optic cables.

1.4.1 Twisted Pair

A twisted pair consists of two insulated copper wires. Over longer distances, cables may contain hundreds of pairs. The twisting of the individual pairs minimizes electromagnetic interference between the pairs (see Figure 1.8).

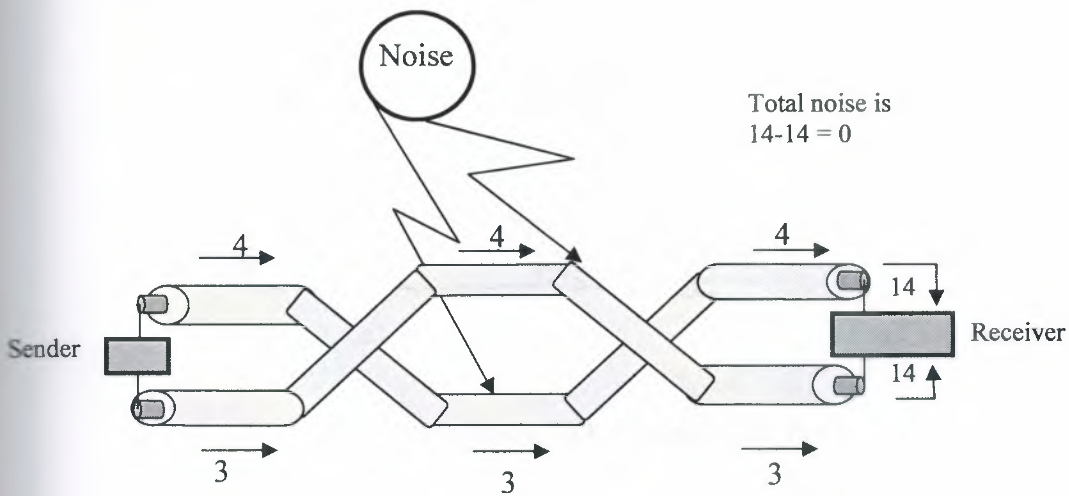


Figure 1.8 Effect of noise on twisted-pair lines

Wire pairs can be used to transmit both analog and digital signals. For analog signals, amplifiers are required about every 5 to 6 km. For digital signals, repeaters are used at every 2 or 3 km. It is the backbone of the telephone system as well as the low - cost microcomputer local network within a building. In the telephone system, individual telephone sets are connected to the local telephone exchange or "end office" by twisted - pair wire. These are referred to as "local loops". Within an office building, telephone service is often provided by means of a Private Branch Exchange (PBX). For modern digital PBX systems, data rate is about 64 kbps. Local loop connections typically require a modem, with a maximum data rate of 9600 bps. However, twisted pair is used for long - distance trucking applications and data rates of 100 Mbps or more may be achieved.

The twisted pair comes in two forms: shielded (STP) and unshielded (UTP). Figure 1.9 shows STP (a) and UTP (b, c). The metal casing prevents the penetration of electromagnetic noise and eliminates cross-talk. Materials and manufacturing requirements make STP more expensive than UTP but less susceptible to noise. UTP is cheap, flexible, and easy to use.

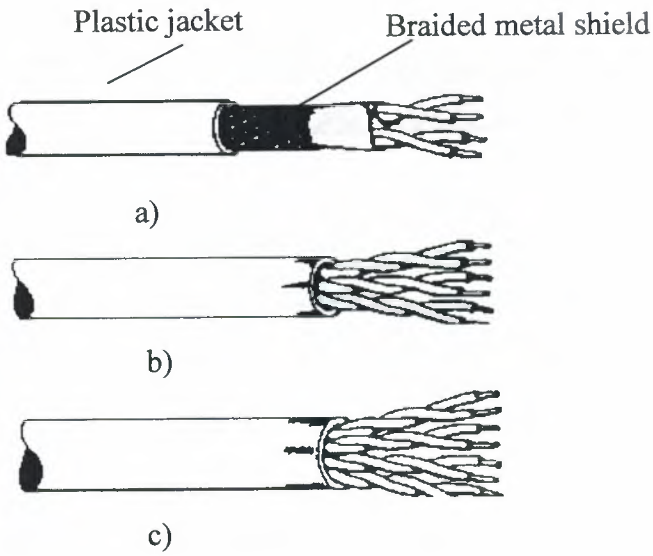


Figure 1.9 shows STP (a) and UTP (b, c).

1.4.2 Coaxial Cable

The main limiting factor of a twisted pair line are its capacity and a phenomenon known as the skin effect. As the bit rate increases, the current flowing in the wires tends to flow only on the outer surface of the wire, thus using the less available cross-section. This increases the electrical resistance of the wires for higher frequency signals, leading to the attenuation In addition, at higher frequencies, more signal power is lost as a result of radiation effect.

Coaxial cables, like twisted pairs, consist of two conductors, but are constructed differently to permit it to operate over a wider range of frequencies. Coaxial cables have been perhaps the most versatile transmission medium and is enjoying increasing utilising in a wide variety of applications. The most important of these are long-distance telephone and television transmission, television distribution, and short-range connections between devices and local area networks. In Figure 1.10 are shown the constructions of the coaxial cables. Using frequency-division multiplexing a coaxial cable can carry over 10,000 voice channels simultaneously. Coaxial cables are used to transmit both analog and digital signals.

The principal constraints on performance are attention, thermal noise, and intermodulation noise.

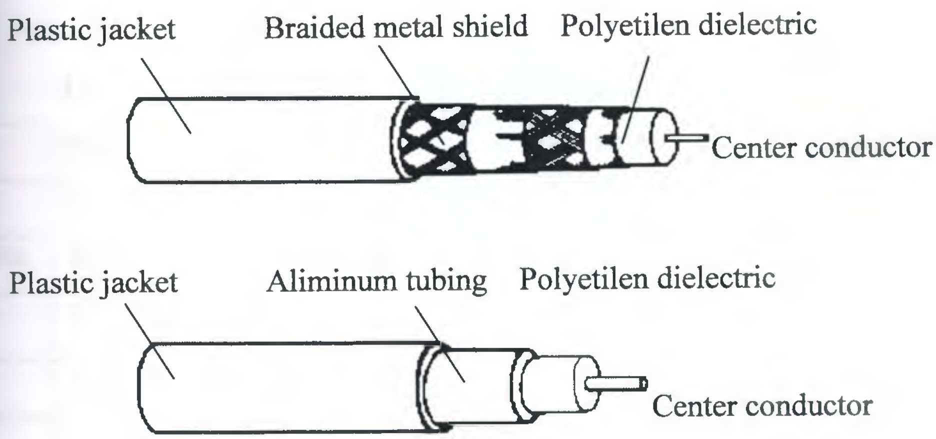


Figure 1.10 Coaxial Cable

1.5 Unguided Media

There are three basic modes of getting a radio wave from the transmitting to receiving antenna: ground wave, space wave, sky wave proportions (Figure 1.11)

The subdivision of the electromagnetic frequency range is given in the Table 1.2

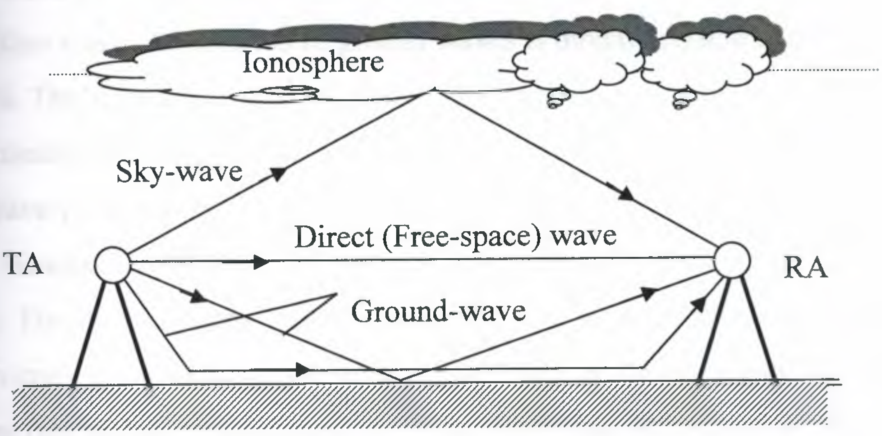


Figure 1.11 Sky Wave Proportion

Table 1.2 Frequency Range for Wireless Communication

Frequency Band	Name	Data rate	Principal applications
30 – 300 kHz	LF (Low Frequency)	0.1 – 100 bps	Navigation, Submarine
300 kHz - 3000 kHz	MF (Medium Frequency)	10 – 1000 bps	AM radio
3 – 30 MHz	HF (High Frequency)	10 – 3000 bps	Shortwave radio, CB radio
30 – 300 MHz	VHF (Very High Frequency)	To 100 kbps	VHF Television, FM radio
300 – 3000 MHz	UHF(Upper High Frequency)	To 10 Mbps	UHF Television Mobile communication Terrestrial Microwave
3 – 30 GHz	SHF (Super High Frequency)	To 100 Mbps	Satellite and Terrestrial microwaves, Radar

The frequency of the radio wave is of primary importance in considering the performance of each type of propagation.

Ground – Wave Propagation

A ground wave is a radio wave that travels along the earth's surface. It is sometimes referred to as a *surface wave.* Attenuation of ground waves is directly related to the surface impedance of the earth. This impedance is a function of conductivity and frequency. If the earth's surface is highly conductive, the absorption of wave energy, and thus its attenuation, will be reduced. Ground–wave propagation is much better over water (especially salt water) than say a very dry (poor conductivity) desert terrain. The ground losses increase rapidly with increasing frequency. For these reasons ground waves are not very effective at frequencies above 2 MHz. Ground- wave propagation is the only way to communicate into the ocean with submarines (about 100 miles distance). To minimise the attenuation of seawater, extremely low frequency (ELF) propagation is utilised. A typically used frequency is 100 Hz, the attenuation is about 0.3 dB/m.

Space – Wave (line-of-site propagation) Propagation

The two types of space waves are shown in Figure 1.12. They are the direct wave and ground reflected wave. Do not confuse these with the ground wave just discussed. The direct wave is by far the most widely used mode of antenna communications. The propagated wave is directed from transmitting to receiving antenna and does not travel along the ground. The earth's surface, therefore, does not attenuate it. The direct space wave has one severe limitation – it is basically limited to so called line–of –sight transmission distances. Thus, the antenna height and the curvature of the earth are the limiting factors. The actual radio horizon is about 1/3 times greater then the geometric line of sight due to diffraction effects and is empirically predicted by the following approximation:

$$d = \sqrt{2h_T} + \sqrt{2h_R}$$

Where d - radio horizon (mi); h_T- transmitting antenna height (ft);

h_R - receiving antenna height (ft)

Ghosting in TV reception. Any tall or massive objects obstruct space waves. This results in diffraction (and subsequent shadow zones) and reflections. Reflections pose a specific problem since, for example, reception of a TV signal may be the combined result of a direct space wave and a reflected space waves. This condition results in ghosting, which manifests itself in the form of a double – image distortion. This is due to the two signals arriving at the receiver at two different times. A possible solution to the ghosting problem is to retune the receiving antenna orientation so that the reflected wave is too weak to be displayed.

Sky Wave Propagation

The sky wave has the ability to strike the ionosphere. It can be refracted from it to the ground, strike the ground, be reflected back toward the ionosphere, and so on. A frequency occurring problem is *signal multipath.* The multipath occurs when the transmitted signal arrives at the receiver via multipath paths at different delays. Signal multipath results intersymbol interference in a digital communication system The signal components arriving via different propagation paths may add destructively, resulting in a phenomenon called *signal fading*. Sky wave propogation ceases to exist at frequencies above 30 MHz. However it is possible to have atmospheric scatter propagation at the range of 30 MHz and troposphere scattering at 40 MHz to 300MHz.

Microwaves

Two general ranges of frequencies are of interest in discussion.

1. Microwave frequencies that cover a range of about 3 to 30 GHz. At these frequencies, highly directional beams are possible, and microwave is quite suitable for point-to-point transmission.
2. Radio waves that cover a range of about 30 MHz to 1 GHz. At these frequencies, omnidirectional transmission is possible, and microwave is quite suitable for broadcasting.

We will refer to signals in the range 30 MHz to 1 GHz as radio waves.

Omnidirectional transmission is used and signals at these frequencies are suitable for broadcast applications. The most common type of microwave antenna is the parabolic "dish". A typical size is about 10 ft in diameter. The antenna is fixed rigidly and focuses a narrow beam to achieve line-of-sight transmission to the receiving antenna. Microwave antennas are usually located at substantial heights above ground level in order to extend the range between antennas and to be able to transmit over intervening obstacles.

The primary use for terrestrial microwave systems is in long-haul telecommunications service, as an alternative to coaxial cable for transmitting television and voice. Like coaxial cable, microwave can support high data rates over long distances. The microwave facility requires far fewer amplifiers or repeaters than coaxial cable for the same distance, but requires line of sight transmission.

Another increasingly common use of microwave is for short point- to point links between buildings. This can be used for closed- circuit TV or as a data link between local networks. Finally, a potential use for terrestrial microwave is to provide digital data transmission in small regions (radius < 10 km). This concept has been termed as "local data distribution" and would provide an alternative to phone lines for digital networking.

The microwave transmission covers a substantial portion of the spectrum. Common frequencies used for transmission are in the range 2 to 40 GHz. The higher the frequency used the higher the potential bandwidth, and therefore, the higher the potential data rate.

As with any transmission system, a main source of loss for microwave is attenuation.

For microwave (and radio frequency), the loss can be expressed as

$$L = 10\text{Log}\left(\frac{4\pi d}{\lambda}\right)^2 \text{ dB}; \qquad \lambda[\text{m}] = \frac{3.10^8}{f[\text{Hz}]}$$

Where d is the distance and λ is the wavelength in the same units.

This loss varies as the square of the distance. This is in contrast to twisted pair and coaxial cable where the loss varies logarithmically with distance (linear in decibels). Thus repeaters or amplifiers may be placed farther apart for microwave systems – 10 to 100 km is typical. Attenuation is increased with rainfall. Another source of impairment for microwave is interference.

Overview of Fiber Optic Cable

The fiber-optic is defined as branch of optics that deals with the transmission of light through ultra pure glass, plastic or some other form of transparent media. One of first noted experiment that demonstrated the transmission of light through a dielectric medium has been credited to John Tyndall. In 1854 John Tyndall demonstrated that light could be guided through stream of water based on the principle of total internal reflection.

In1880 Alexander Graham Bell invented the photo phone, a device that transmits voice signals over a beam of light.

Great interest in communication at optical frequencies was created in 1958 with the invention of the laser by Charles H. Townes.

In 1966 Charles K. Kao and George Hockham of Standard Telecommunications Laboratories of England performed several experiments to prove that, if glass could be made more transparent by reducing its impurities, light loss could be minimized. Their research led to a publication in which they predicted that optical fiber could be made pure enough to transmit light several kilometers. In the next two decades researchers worked intensively to reduce the attenuation to 0.16 dB/km.

In1988 the Synchronous Optical Network (SONET) was published by the American National Standards Institute (ANSI).

1995 Multimedia applications for business have become the major impetus for increased use of optical fiber within the LAN, MAN, and WAN environment.

1.6.1 Advantages and Disadvantages of the FOS

a) Advantages

The major advantages are:

Bandwidth One of the most significant advantages that fiber has over copper or other transmission media is a bandwidth. Bandwidth is directly related to the amount of information that can be transmitted per unit time. Today's advanced fiber optic systems are capable of transmitting several gigabits per second over hundreds of kilometers. Ten thousands of voice channels can now be multiplexed together and sent over a single fiber strand.

Less Lose. Currently, fiber is being manufactured to exhibit less than a few tenths of a decibel of loss per kilometer.

Less Weight and Volume. Fiber optic cables are substantially lighter in weight and occupy much less volume than copper cables with the same information capacity. For example, a 3-in. diameter telephone cable consisting of 900 twisted-pair wires can be replaced with a single fiber strand 0.005 inch in diameter (approximately the diameter of a hair strand) and retain the same information-carrying capacity. Even with a rugged protective jacket surrounding the fiber, it occupies enormously less space and weights considerably less.

Security. Since light does not radiate from a fiber optic cable, it is nearly impossible to secretly tap into it without detection. For this reason, several applications requiring communications security employ fiber-optic systems. Military information, for example, can be transmitted over fiber to prevent eavesdropping. In addition, metal detectors cannot detect fiber-optic cables unless they are manufactured with steel reinforcement for strength.

Flexibility. The surface of glass fiber is much more refined than ordinary glass. This, coupled with its small diameter, allows it to be flexible enough to wrap around a pencil. In terms of strength, a 0.005-in. strand of fiber is strong enough to cut one's finger before it breaks, if enough pressure is applied against it.

Economics. Presently, the cost of fiber is comparable to copper at approximately $0.20 to $0.50 per yard and is expected to drop as it becomes more widely used. Since transmission losses are considerably less than for coaxial cable, expensive repeaters can be spaced farther apart. Fewer repeaters mean a reduction in overall system cost and enhanced reliability.

Reliability. Once installed, a longer life span is expected with fiber over its metallic counterparts since it is more resistant to corrosion caused by environmental extremes such as temperature, corrosive gases, and liquids.

b) Disadvantages

In spite of the numerous advantages that fiber optic systems have over conventional methods of transmission, there are some disadvantages, particularly because of its newness. Many of these disadvantages are being overcome with new and competitive technology.

Interfacing costs. Electronic facilities must be converted to optics in order to interface a fiber. Often these costs are initially overlooked. Fiber-optic transmitter, receiver, couplers, and connectors, for example, must be employed as part of the communication system. Test and repair equipment is costly. If the fiber optic cable breaks, splicing can be a costly and tedious task

Strength. Fiber, by itself, has a tensile strength of approximately 1 lb, as compared to coaxial cable at 180 1b (RG59U) surrounding the fiber with stranded Kevlar and a protective PCV jacket can increase the pulling strength up to 500 1b. Installations requiring greater tensile strengths can be achieved with steel reinforcement.

Remote Powering of Devices. Occasionally it is necessary to provide electrical power to a remote device. Since this cannot be achieved through the fiber, metallic conductors are often included in the cable assembly. Several manufacturers now offer a complete line of cable types, including cables manufactured with both copper wire and fiber.

1.6.2 Theory of Light

In the seventeenth and eighteenth centuries, there were two schools of thought regarding the nature of light. Sir Isaac Newton and his followers believed that light consisted of rapidly moving particles (or corpuscles), whereas Dutch physicist Christian Huygens regarded light as being a series of waves.

The wave theory was strongly supported by an English doctor named Thomas Young. By 1905, quantum theory, introduced by Clark Maxwell, showed that when light is emitted or absorbed it is not only as a wave, but also as an electromagnetic particle called a photon. Photon is said to possess energy that is proportional to its frequency. This is known as Planck's law, which states:

$$E = h \times v$$

where E = photon's energy (J);

h = Planck's constants, 6.63×10^{-34} (J-s);

v = frequency of the photon (Hz).

Using the particle theory, Einstein and Planck were able to explain photoelectric effect: when visible light or electromagnetic radiation of a hire frequency shines on a metallic surface, electrons are emitted, which is turning an electric current.

Electromagnetic Spectrum

Fundamentally, light has been accepted as a form of electromagnetic radiation that can be categorized into a portion of the entire electromagnetic spectrum, as shown in Table 1.3. In addition, each frequency can be specified in terms of its equivalent wavelength. Frequency or wavelength are directly related to the speed of light.

$$C = f \times \lambda$$

Where c - speed of light in a vacuum or free space, 3×10^{8} (m/s);

f - frequency (Hz); λ-wavelength (m).

Table 1.3 Electromagnetic Spectrum

Range of wavelength, nm	Name of wavelength	
$10^6 - 770$	*Infrared*	*Invisible*
770 - 662	Red	*Visible*
662 - 597	Orange	
597 - 577	Yellow	
577 - 492	Green	
492 - 455	Blue	
455 - 390	Violet	
390 - 10	*Ultraviolet*	*Invisible*

The portion of the electromagnetic spectrum regarded as light has been expanded in Table 1.3 to illustrate three basic categories of light:

1. ***Infrared***: that portion of the electromagnetic spectrum having a wavelength ranging from 770 to 10^{6} nm. Fiber optic systems operate in this range.
2. ***Visible:*** that portion of the electromagnetic spectrum having a wavelength ranging from 390 to 770 nm. The human eye, responding to these wavelengths allows us to see the colours ranging from violet to red, respectively.
3. ***Ultraviolet:*** that portion of the electromagnetic spectrum ranging from 10 to 390 nm.

The light that we use for most fiber optic systems occupies a wavelength range from 800 to 900 nm. This is slightly larger than visible red light and falls within the infrared portion of the spectrum.

Snell's Law : Total Interval Reflection

For light to propagate in any medium, the medium must be transparent to some degree. The degree of transparency determines how far light will propagate. Transparent materials can be in the form of a liquid, gas, or a solid. Some examples are glass, plastic, air, and water.

One of the most fundamental principles of light is that when it strikes the interface between two transparent mediums, such as air and water, a portion of the light energy is reflected back into the first medium and a portion is transmitted into the second medium. The path in which light travels from one point to another is commonly referred to as the ray. Figure 1.14 illustrates the classic example of a ray of light incident upon the surface of water. Notice that part of the light is *reflected* off the surface of water and part of it penetrates the water. The ray penetrating to water is said to be *refracted* or bent toward the normal. The amount of refracted light is determined by the medium's index of refraction, generally denoted by the letter n. Index of refraction is the ratio of the speed of light in a vacuum - c, to the speed of light in the given medium - v. This relationship is given by the equation:

$n = c / v$. Since the speed of light is lower in mediums other than a vacuum, the index of refraction in such mediums is always greater than 1.

Example for air n = 1.003, for water n = 1.33, for fiber-optic n = 1.6.

In 1621, the Dutch mathematician Willebrard Snell established that rays of light could be traced as they propagate from one medium to another based on their indices of refraction. Snell's low is stated by the equation:

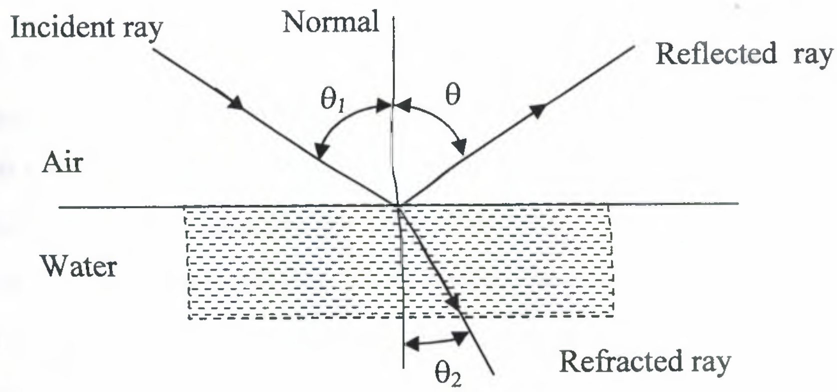


Figure 1.14 Ray Of Light Incident Upon The Surface Of Water.

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1}; \qquad n_1 \sin\theta_1 = n_2 \sin\theta_2$$

where n_1 - refractive index of material 1; θ_1 - angle of incidence; θ_2 - angle of refraction; n_2 - refractive index of material 2. When the angle of incidence, θ_1, becomes large enough to cause the sine of the refraction angle, θ_2, to exceed the value of 1, total internal reflection occurs. This angle is called the critical angle, θ_c. The critical angle, θ_c, can be derived from Snell's law as follows $\qquad n_1 \sin\theta_1 = n_2 \sin\theta_2$

$$\sin\theta_1 = n_2 \sin\theta_2 / n_1$$

When $\sin\theta_1 = \sin\theta_2$, then $\sin\theta_1 = n2 / n1$. Therefore, critical angle: $\theta_c = \sin^{-1}(n2 / n1)$

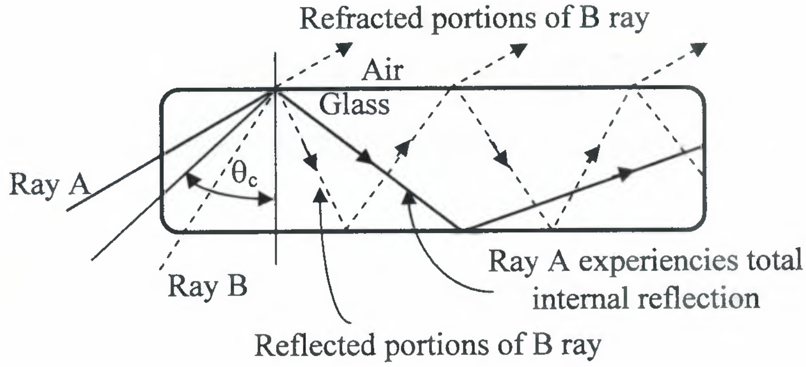


Figure 1.15 Ray A penetrates the glass-air interface at an angle exceeding the critical angle, θc.

For glass, n = 1.5; for air n = 1.0 and $\theta_c = \sin^{-1}(n2 / n1) = \sin^{-1}(1.0 / 1.5) = 41.8^0$.

By surrounding glass with material whose refraction index is less than that of the glass, total internal reflection can be achieved. This is illustrated in Figure 1.15. Ray A penetrates the glass-air interface at an angle exceeding the critical angle, θc, and therefore experiences total internal reflection. On the other hand, Ray B penetrates the glass air interface at an angle less than the critical angle. Total internal reflection does not occur. Instead, a portion of ray B escapes the glass and is refracted away from the normal as it enters the less dense medium of air. A portion is also reflected back into the glass. Ray B diminished in magnitude as it bounces back and forth between the glass-air interface. The foregoing principle is the basis for guiding light through optical fibers.

Two key elements that permit light guiding through optical fibers are its core and its cladding. The fiber's core is manufactured of ultra pure glass (silicon dioxide) or plastic. Surrounding the core is a material called cladding. A fiber cladding is also made of glass or plastic.

Its index of refraction, however, it is typically 1% less than that of its core. This permits total internal reflection of rays entering the fiber and striking the core-cladding interface above the critical angle of approximately 82-degree ($\sin^{-1}$ (1/1.01). The core of the fiber therefore guides the light and the cladding contains the light. The cladding material is much less transparent than the glass making up the core of the fiber.

This causes light rays to be absorbed if they strike the core-cladding interface at an angle less than the critical angle.

In Total internal reflection occurs as it strikes the lower index cladding material.

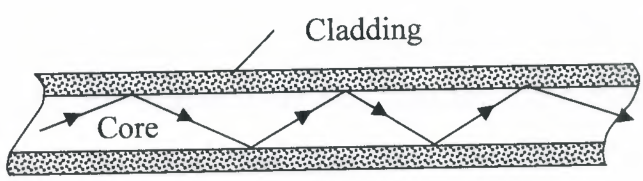


Figure 1.16 A Light Ray Is Transmitted Into The Core Of An Optical Fiber.

Block Diagram of the FOS

One of the main limitations of communication systems is their restricted information carrying capabilities. In more specific terms what this means is that the communications medium can only carry so many messages. And, as you have seen, this information-handling ability is directly proportional to the bandwidth of the communications channel. In telephone systems, the bandwidth is limited by the characteristics of the cable used to carry the signals. As the demand for telephones has increased, better cables and wiring systems have been developed. Further, multiplexing techniques have been used to transmit multiple telephone conversations over a single cable.

In radio communication systems, the information modulates a high frequency carrier. The modulation produces sidebands, and therefore, the signal occupies a narrow portion of the RF spectrum. However, the RF spectrum is finite. There is only so much space for radio signals. To increase the information capacity of a channel, the bandwidth of the channel must be increased. This reduces available spectrum space. Multiplexing techniques are used to send more signals in a given channel bandwidth, and methods have been developed to transmit more information in less bandwidth.

The information-carrying capacity of the radio signal can be increased tremendously if higher carrier frequencies are used. As the demand for increased communications capacity has gone up over the years, higher and higher RFs are being used. Today, microwaves are the preferred radio channels for this reason, but it is more complex and expensive to use these higher frequencies because of the special equipment required.

One way to expand communications capability further is to use light as the transmission medium. Instead of using an electrical signal traveling over a cable or electromagnetic waves traveling through space, the information is put on a light beam and transmitted through space or through a special cable. In the late nineteenth century, Alexander Graham Bell, the inventor of the telephone, demonstrated that information could be transmitted by light.

Light beam communication was made more practical with the invention of the laser. The laser is a special high-intensity, single frequency light source. It produces a very narrow beam of brilliant light of a specific wavelength (color). Because of its great intensity, the laser beam can penetrate atmospheric obstacles better than other types of light, thereby making light-beam communication more reliable over longer distances. The primary problem with such free-space light beam communication is that the transmitter and receiver must be perfectly aligned with one another.

Instead of using free space, some type of light carrying cable can also be used. For centuries it has been known that light is easily transmitted through various types of transparent media such as glass and water, but it wasn't until the early in 1900s that scientist were able to develop practical light carrying media. By the mid-1950s glass fibers were developed that permitted long light carrying cables to be constructed. Over the years, these glass fibers have been perfected. Further, low cost plastic fiber cable also developed. Developments in these cables permitted them to be made longer with less attenuation of the light.

Today the fiber optic cables have been highly refined. Cables many miles long can be constructed and interconnected for the purpose of transmitting information on a light beam over very long distances. Its great advantage is that light beams have an incredible information carrying capacity. Whereas hundreds of telephone conversations may be transmitted simultaneously at microwave frequencies, many thousands of signals can be carried on a light beam through a fiber optic cable. Using multiplexing techniques similar to those used in telephone and radio systems, fiber optic communications systems have an almost limitless capacity for information transfer.

The components of a typical fiber optic communications system are illustrated in Figure 1.17.

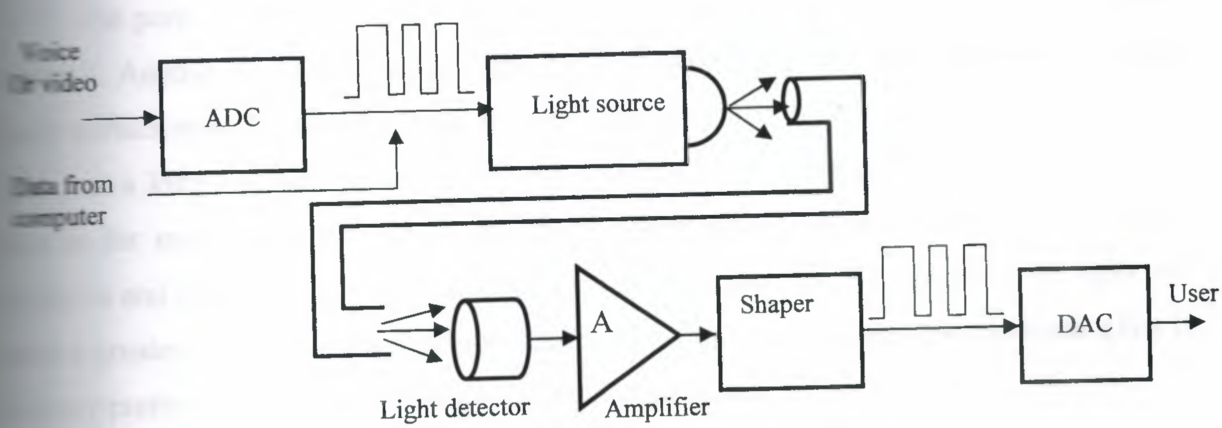


Figure 1.17 Typical Fiber Optic Communications System

The information signal to be transmitted may be voice, video, or computer data. The first step is to convert the information into a form compatible with the communications medium. This is usually done by converting continuous analog signals such as voice and video (TV) signals into a series of digital pulses. An Analog-to-Digital Converter (ADC) is used for this purpose.

Computer data is already in digital form. These digital pulses are then used to flash a powerful light source off and on very rapidly. In simple low cost systems that transmit over short distances, the light source is usually a light-emitting diode (LED). This is a semiconductor device that puts out a low intensity red light beam. Other colors are also used. Infrared beams like those used in TV remote controls are also used in transmission. Another commonly used light source is the laser emitting diode. This is also a semiconductor device that generates an extremely intense single frequency light beam.

The light beam pulses are then fed into a fiber optic cable where they are transmitted over long distances. At the receiving end, a light sensitive device known as a photocell or light detector is used to detect the light pulses. This photocell or photo detector converts the light pulses into an electrical signal. The electrical pulses are amplified and reshaped back into digital form. They are fed to a decoder, such as a Digital-to-Analog Converter (DAC), where the original voice or video is recovered for user.

1.7 Fiber Optic Cables

Just as standard electric cables come in a variety of sizes, shapes, and types, fiber optic cables are available in different configurations. The simplest cable is just a single strand of fiber, whereas complex cables are made up of multiple fibers with different layers and other elements.

The portion of a fiber optic cable (core) that carries the light is made from either glass or plastic. Another name for glass is silica. Special techniques have been developed to create nearly perfect optical glass or plastic, which is transparent to light. Such materials can carry light over a long distance. Glass has superior optical characteristics over plastic. However, glass is far more expensive and more fragile than plastic. Although the plastic is less expensive and more flexible, its attenuation of light is greater. For a given intensity, light will travel a greater distance in glass than in plastic. For very long distance transmission, glass is certainly preferred. For shorter distances, plastic is much more practical.

All fibers consist of a number of substructures including (see Figure 1.18):

- A core, which carries most of the light, surrounded by
- A cladding, which bends the light and confines it to the core, surrounded by
- A substrate layer (in some fibers) of glass which does not carry light, but adds to the diameter and strength of the fiber, covered by

- A primary buffer coating, which provides the first layer of mechanical protection, covered by
- A secondary buffer coating, which protects the relatively fragile primary coating.

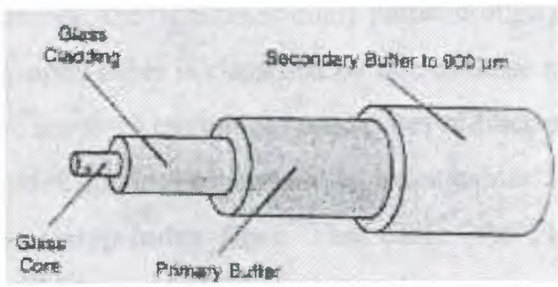


Figure 1.18 Fiber Optic Cable

The cladding is also made of glass or plastic but has a lower index of refraction. This ensures that the proper interface is achieved so that the light waves remain within the core. In addition to protecting the fiber core from nicks and scratches, the cladding adds strength. Some fiber optic cables have a glass core with a glass cladding. Others have a plastic core with a plastic cladding. Another common arrangement is a glass core with a plastic cladding. It is called plastic-clad silica (PCS) cable.

1.7.1 Basic Construction of the Fiber-Optic Cables

There are two basic ways of classifying fiber optic cables. The first way is an indication of how the index of refraction varies across the cross section of the cable. The second way of classification is by mode. Mode refers to the various paths that the light rays can take in passing through the fiber. Usually these two methods of classification are combined to define the types of cable. There are two basic ways of defining the index of refraction variation across a cable. These are step index and graded index. Step index refers to the fact that there is a sharply defined step in the index of refraction where the fiber core and the cladding interface. It means that the core has one constant index of refraction N1, while the cladding has another constant index of refraction N2.

The other type of cable has a graded index. In this type of cable, the index of refraction of the core is not constant. Instead, the index of refraction varies smoothly and continuously over the diameter of the core. As you get closer to the center of the core, the index of refraction gradually increases, reaching a peak at the center and then declining as the outer edge of the core is reached. The index of refraction of the cladding is constant.

Mode refers to the number of paths for the light rays in the cable. There are two classifications: single mode and multimode. In single mode, light follows a single path through the core. In multimode, the light takes many paths through the core.

Each type of fiber optic cable is classified by one of these methods of rating the index or mode. In practice, there are three commonly used types of fiber optic cable: multimode step index, single mode step index and multimode graded index cables.

1. **The multimode step-index fiber.** This cable (see Figure 1.19(a)) is the most common and widely used type. It is also the easiest to make and, therefore, the least expensive. It is widely used for short to medium distances at relatively low pulse frequencies.

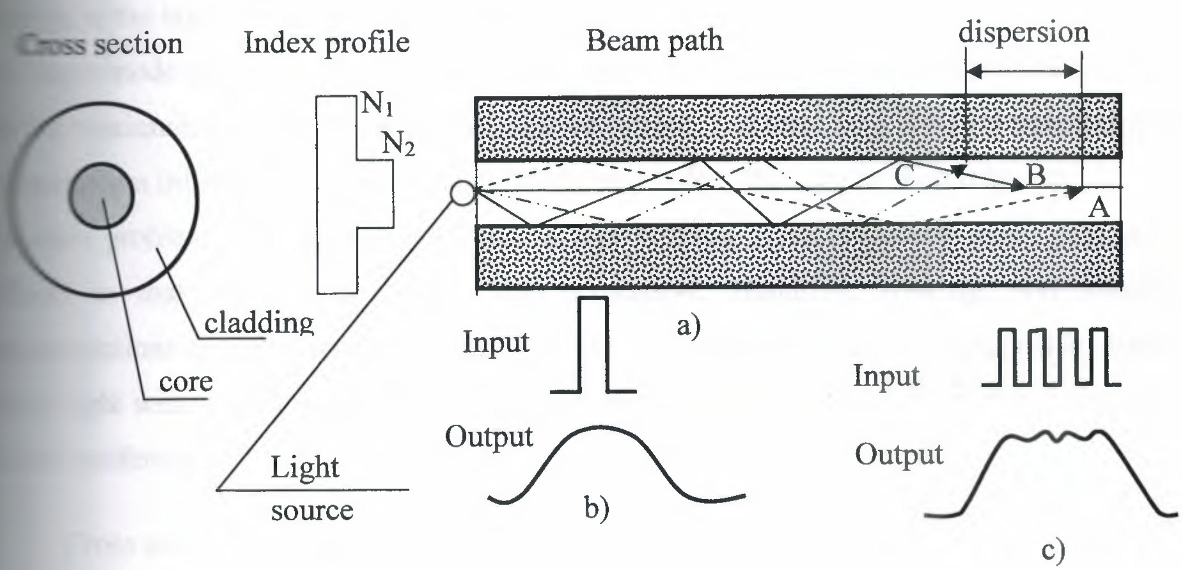


Figure 1.19 The multimode step-index fiber

The main advantage of a multimode step index fiber is the large size. Typical core diameters are in the 50-to-1000 micrometers (μm) range. Such large diameter cores are excellent at gathering light and transmitting it efficiently. This means that an inexpensive light source such as LED can be used to produce the light pulses. The light takes many hundreds of even thousands of paths through the core before exiting. Because of the different lengths of these paths, some of the light rays take longer to reach the end of the cable than others. The problem with this is that it stretches the light pulses (Figure 1.19 (b). In Figure 1.19 ray A reaches the end first, then B, and C. The result is a pulse at the other end of the cable that is lower in amplitude due to the attenuation of the light in the cable and increased in duration due to the different arrival times of the various light rays. The stretching of the pulse is referred to as modal dispersion. Because the pulse has been stretched, input pulses can not occur at a rate faster than the output pulse duration permits. Otherwise the pulses will essentially merge together as shown in Figure 1.19 (c).

At the output, one long pulse will occur and will be indistinguishable from the three separate pulses originally transmitted. This means that incorrect information will be received. The only cure for this problem is to reduce the pulse repetition rate. When this is done, proper operation occurs. But with pulses at a lower frequency, less information can be handled.

2. Single Mode Cable

In a single mode, or mono-mode, step-index fiber cable the core is so small that the total number modes or paths through the core are minimized and modal dispersion is essentially eliminated. The typical core sizes are 5 to 15 μm. The output pulse has essentially the same duration as the input pulse (see Figure 1.20).

The single mode step index fibers are by far the best since the pulse repetition rate can be high and the maximum amount of information can be carried. For very long distance transmission and maximum information content, single-mode step-index fiber cables should be used.

The main problem with this type of cable is that because of its extremely small size, it is difficult to make and is, therefore, very expensive. Handling, splicing, and making interconnections are also more difficult. Finally, for proper operation an expensive, super intense light source such as a laser must be used. For long distances, however, this is the type of cable preferred.

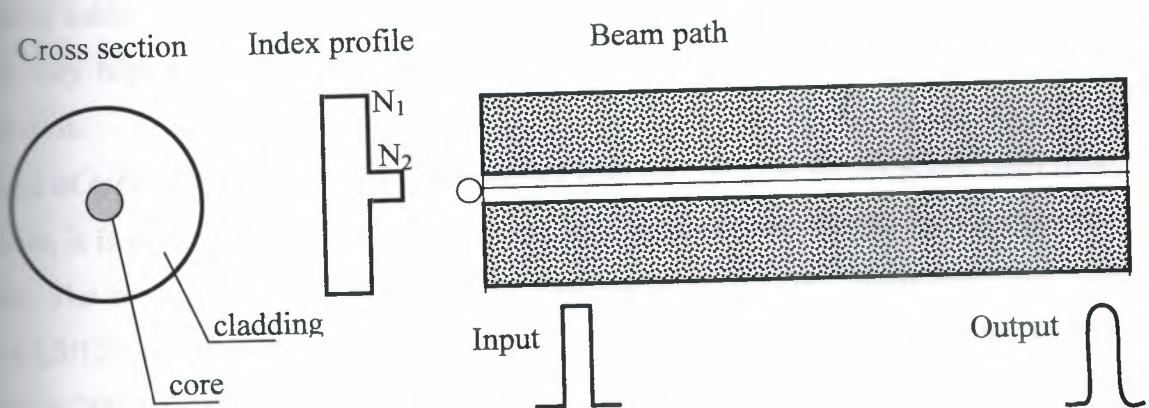


Figure 1.20 Single Mode Cable

Multimode Graded-Index Fiber Cables.

These cables have a several modes or paths of transmission through the cable, but they are much more orderly and predictable. Figure 1.8 shows the typical paths of the light beams. Because of the continuously varying index of refraction across the core, the light rays are bent smoothly and converge repeatedly at points along the cable.

The light rays near the edge of the core take a longer path but travel faster since the index of refraction is lower. All the modes or light paths tend to arrive at one point simultaneously. The result is that there is less modal dispersion.

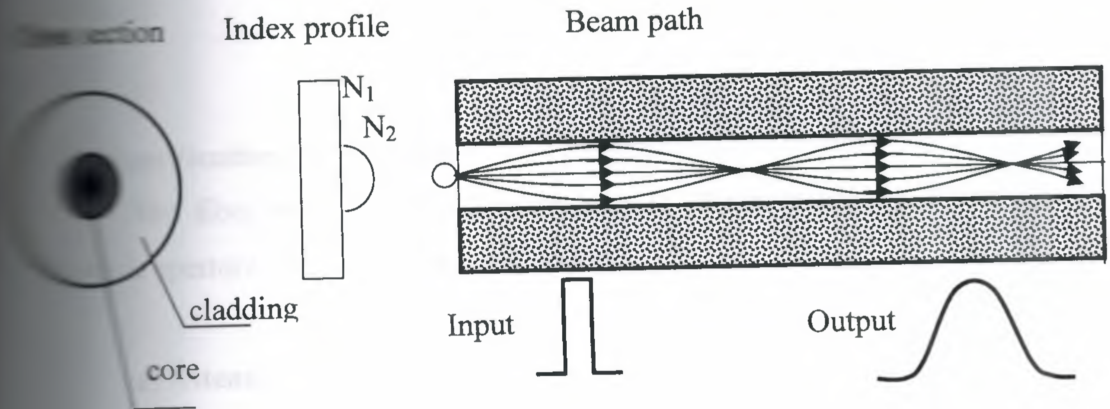


Figure 1.21 Multimode Graded-Index Fiber Cables

It is not eliminated entirely, but the output pulse is not nearly as stretched as in multimode step index cable. The output pulse is only slightly elongated. As a result, this cable can be used at very high pulse rates and, therefore, a considerable amount of information can be carried on it.

This type of cable is also much wider in diameter with core sizes in the 50 to100 (μm) range. Therefore, it is easier to splice and interconnect, and cheaper, less-intense light sources may be used. The most popular fiber-optic cables that are used in LAN: Multimode-step index cable -65.5/125; multimode-graded index cable - 50/125. The multimode-graded index cable - 100/140 or 200/300 are recommended for industrial control applications because its large size. In high data rate systems is used single mode fiber 9/125.

Typical core and cladding diameters of these cables are shown in Figure 1.22.

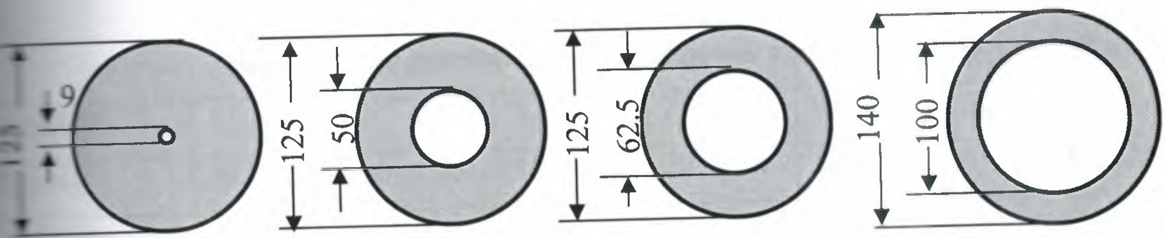


Figure 1.22 Typical core and cladding diameters of these cables

Specifications of the cables

The fiber as a transmission medium is characterized by Attenuation, A, db/km; Numeric aperture, NA and Dispersion, ns/km.

a) Attenuation

The main specification of a fiber optic cable is its attenuation.

Light power which does not reach the other end of the fiber has either left the fiber or been absorbed (converted to heat) in it. The amount of attenuation varies with the type of cable and its size. Glass has less attenuation than plastic. Wider cores have less attenuation than narrower cores. But more importantly, the attenuation is directly proportional to the length of the cable. It is obvious that the longer the distance the light has to travel the greater the loss due to absorption, scattering, and dispersion. Doubling the length of a cable doubles the attenuation, and so on.

The attenuation of a fiber optic cable is expressed in decibels per unit of length. The standard specification for fiber-optic cable is the attenuation expressed in terms of decibels per kilometers. The standard decibel formula used is

$$\text{Loss, dB} = 10 \log (P_0/P_I)$$

Where P_0 is the output power and P_I is the input power.

The table 1.4 shows the percentage of output power for various decibel loss. The attenuation range of fiber-optic cables vary over a considerable range.

Table 1.4 The percentage of output power expressed by dB

Loss (dB)	1	2	3	4	5	6	7	8	9	10	20	30
P (%)	79	63	50	40	31	25	20	14	12	10	1	0.1

The finest single mode step-index cables have an attenuation of only 1 dB/km. However, a very large core plastic fiber cables can have an attenuation of several thousands decibels per kilometer.

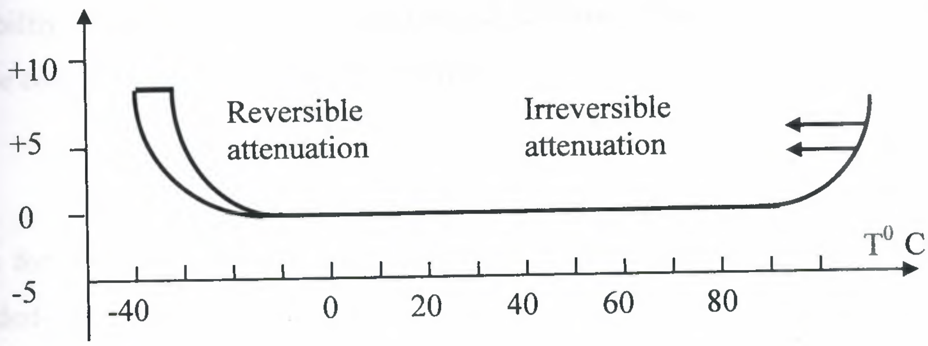


Figure 1.23 Temperature dependence of the attenuation of fiber optic cable

The following contribute to the attenuation:

Reyleigh-scattering. A mechanism called Rayleigh scattering prevents any further improvement in attenuation loss. Rayleigh scattering is caused by micro irregularities in the random molecular structure of glass. These irregularities are formed as the fiber cools from a molten state. Normally, electrons in glass molecules interact with transmitted light by absorbing and reradiating light at the same wavelength. A portion of the light, however, strikes these micro irregularities and becomes scattered in all directions of the fiber, some of which is lost in the cladding. Consequently, the intensity of the beam is diminished.

Radiation Losses.

A phenomenon called micro bending can cause radiation losses in optical fibers in excess of intrinsic losses. Micro bends are miniature bends and geometric imperfections along the axis of the fiber that occur during the manufacturing or installation of the fiber. Mechanical stresses such as pressure, tension, and twist can cause micro bending. This geometric imperfection causes light to get coupled to various unguided electromagnetic modes that radiate and escape the fiber.

Absorption. The following contribute to the absorption: Intrinsic impurities, irregularities in core diameter, IR-absorption (infrared), OH- absorption (hydroxy, humidity) and molecular agitation.

Numerical Aperture

Numerical Aperture tells how much of the light can be pass into the fiber. An important characteristic of a fiber is its *numerical aperture* (NA). NA characterizes a fiber's light-gathering capability. Mathematically, it is defined as the sine of half the angle of a fiber's *light acceptance cone*. For multimode step index fiber

$$NA = \sqrt{N^2{}_1 - N^2{}_2}$$

Typical values for NA are 0.25 to 0.4 for multimode step-index fiber and 0.2 to 0.3 for multimode graded-index fiber.

Dispersion

Dispersions classified: material dispersion, Ψ_{mat} (ns/km) and modal dispersion, Ψ_{mod} (ns/km).

Material dispersion

A light pulse is composed of light of different wavelengths depending on the spectral width of the light source. The refractive index depends weakly on the wavelength. This causes the material dispersion.

Modal dispersion. As shown above the modal dispersion due to the different arrival times of the various light rays.

The Table 1.5 shows the characteristics of the dispersions.

Table 1.5 Characteristics of the dispersions

Fiber type	Dispersion (ns / km)	
	Modal	Material
Step index	$\psi mod = t * (\Delta / 2)$	$\psi mat = 0.1\ \Delta\lambda$
Graded index	$\psi mod = t * (\Delta^2 / 2)$	
Single mode	$\psi mod = 0$	

Note: t- traveling time per km t=N/C, for N=1.5, t= 5μs/km;

$\Delta\lambda = \Delta N/N$, in practice Δ=0.01

$\Delta\lambda$ - Bandwidth of the light;

The total dispersion equals $\psi_{tot} = (\psi_{mat} + \psi_{mod})^{1/2}$

1.3 Optical Transmitters

In an optical communications system the transmitter consist of a modulator and the circuitry that generates the carrier. In this case, the carrier is a light beam that is modulated by digital pulses that turn it on and off. The basic transmitter is nothing more than a light source.

Several devices are emitters of light, both natural and artificial. Few of these devices, however, are suitable for fiber-optic transmitters. What we are interested in a light source that meets the following requirements:

- The light source must be able to turn on and off several tens of millions and even billions of time per second.
- The light source must be able to emit a wavelength that is transparent to the fiber.
- The light source must be efficient in terms of coupling light energy into the fiber.
- The power emitted must be sufficient enough to transmit through the optical fibers.
- Temperature variations should not affect the performance of the light source.
- The cost of manufacturing the light source must be relatively inexpensive.

The commonly used devices that satisfy the above requirements are: monochromatic incoherent source-the light emitting diode (LED) and monochromatic coherent source-the injection laser diode (ILD).

Light Emitting Diode

The major difference between the LED and the ILD is the manner in which light is emitted from each source. The LED is an incoherent light source that emits light in a disorderly way. A laser is a light source that emits coherent monochromatic light. Monochromatic light has a pure single frequency. Coherent refers to the fact that all the light waves emitted are in phase with one another. Coherent light waves are focused into a narrow beam, which, as a result, is extremely intense. The effect is somewhat similar to that of using highly directional antenna to focus radio waves into a narrow beam, which also increases the intensity of the signal. Figure 1.24 illustrates the differences in radiation patterns. Both devices are extremely rugged, reliable, and small in size. In terms of spectral purity, the LED's half power spectral width is approximately 50 nm, whereas the ILD's spectral width is only a few nanometers. This is shown in Figure 1.24.

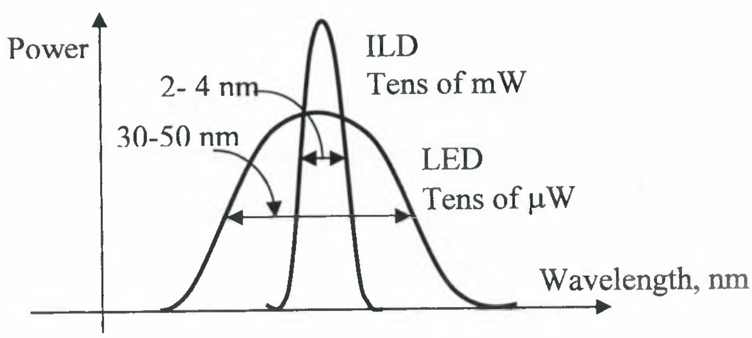


Figure 1.24 Differences In Radiation Patterns

Ideally, a single spectral line is desirable. As the spectral width of the emitter increases, attenuation and pulse dispersion increase. The spectral purity for the ILD and its ability to couple much more power into a fiber make it better suited for long-distances telecommunications links. In addition, injection laser can be turned on and off at much higher rates than an LED. The drawback, however, is its cost, which may approach several hundreds of dollars as compared to a few dollars for LED's in large quantities.

Table 1.6 lists the differences in operating characteristics between the LED and the ILD.

Table 1.6 Typical source characteristics for LED and ILD

	Output Power, μW	Peak wavelength, nm	Spectral width, nm	Rise time, ns
LED	250	820	35	12
	700	820	35	6
	1500	820	35	6
Laser	4000	820	4	1
	6000	1300	2	1

Various semiconductor materials are used to achieve this. Pure *gallium arsenide* (GaAs) emits light at a wavelength of about 900 nm. By adding a mixture of 10% aluminum (Al) to 90% GaAs, *gallium-aluminium-arsenide* (GaAlAs) is formed, which emits light at a wavelength of 820 nm. Recall that this is one of the optimum wavelengths for fiber optic transmission. By altering the amount of aluminium mixed with GaAs, wavelengths ranging from 800 to 900 nm can be obtained.

To take advantage of the reduced attenuation losses at longer wavelengths, it is necessary to include even more exotic materials. For wavelengths in the range 1000 to 1550 nm, a combination of four elements is typically used: *indium, gallium, arsenic* and *phosphorus*. These devices are commonly referred to as *quaternary* devices. Combining these four elements produces the compound *indium-gallium-arsenide-phosphide* (InGaAsP). Transfer characteristic of LED and ILD are shown Figure 1.25 (a) and (b).

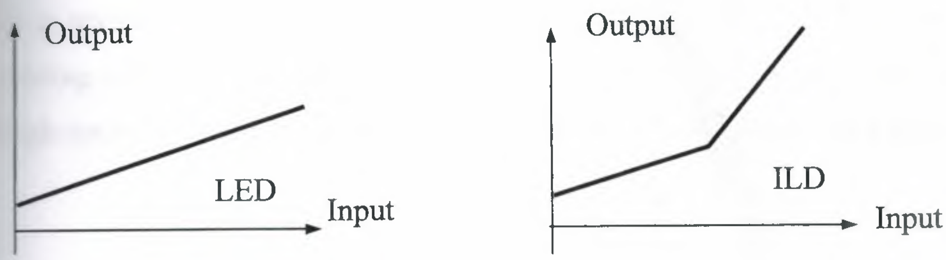


Figure 1.25 Transfer characteristic of LED and ILD

Injection Laser Diode

The term *laser* is an acronym for *light amplification by stimulated emissions of radiation.* There are many types of lasers on the market. They are constructed of gases, liquids, and solids. Laser diodes are also called injection laser diodes (ILD), because when current is passed across the PN junction, light is emitted relatively large and sophisticated device that emits a highly intense beam of visible light. Although this is in part true, the laser industry is currently devoting a great deal of effort toward the manufacture of miniature semiconductor laser diodes. Figure 1.24 illustrates the spectrum ILD. ILDs are ideally suited for use within the fiber-optic industry due to their small size, reliability, and ruggedness. Step response of ILD is shown Figure 1.26.

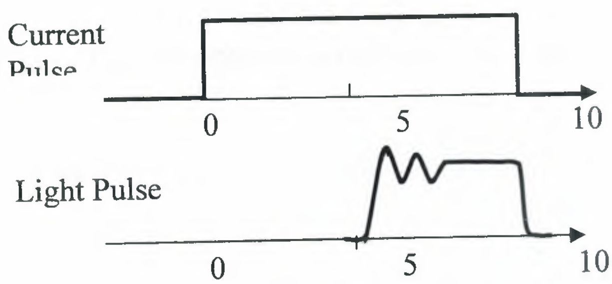


Figure 1.26 Step response of ILD

The most widely used light source in fiber optic systems is ILD. Like the LED, it is a PN junction diode usually made of GaAs. Injection laser diodes are capable of developing light power up to several watts. They are far more powerful than LEDs and, therefore, are capable of transmitting over much longer distances. Another advantage ILDs have over LEDs is theirs speed. High-speed laser diodes are capable of gigabit per second digital data rates.

Optical Transmitter Circuits

A light transmitter consists of the LED and its associated driving circuitry. An optical transmitter circuit using the LED is shown in Figure 1.27. The binary pulses are applied to a NAND gate, which, in turn, operates a transistor switch T that turns the LED off and on. A positive pulse at the NAND gate input causes the NAND output to go to zero.

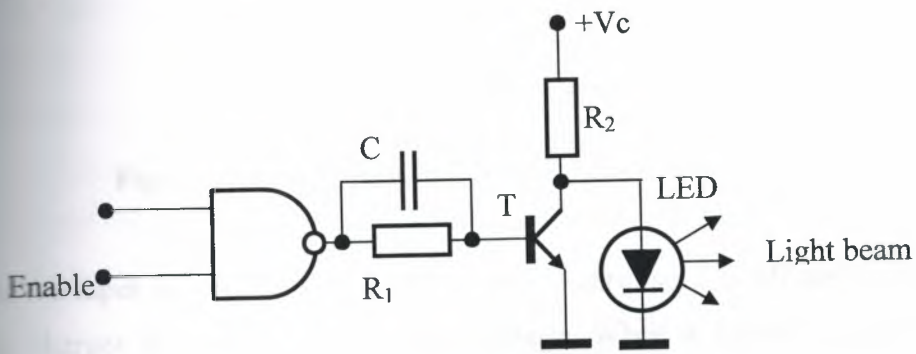


Figure 1.27 An optical transmitter circuit using the LED

This turns off T, so the LED is forward-biased through R_1 and turns on. With zero input, the NAND output is 1, so T turns on and shunts current away from the LED. Very high current pulses are used to ensure a brilliant high-transmission rates are limited. Most LED like transmitters are used for short-distance, low speed digital fiber-optic systems. With zero input, the NAND output is 1, so T turns on and shunts current away from the LED. Most LEDs are capable of generating power levels up to approximately several hundred µW. With such low intensity, LED transmitters are good for only short distances. Further the speed of the LED is limited. Turn-of and turn-on times are no faster than several nanoseconds.

A typical injection laser transmitter circuit is shown in Figure 1.28.

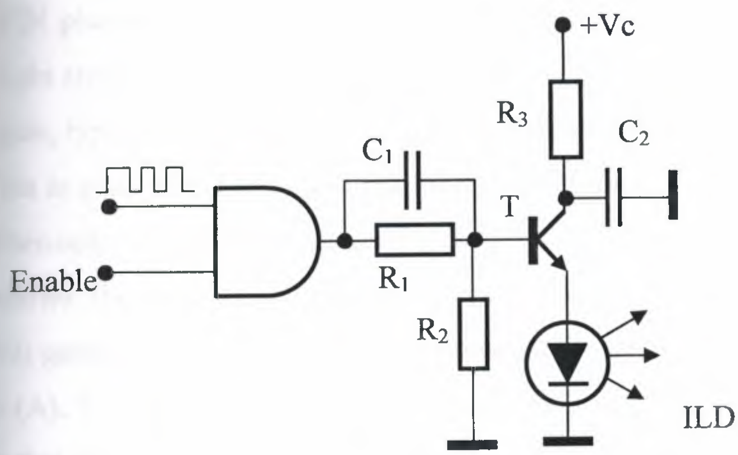


Figure 1.28 A Typical Injection Laser Transmitter Circuit

When the input is zero, the AND gate output is zero, so T is off and so is the laser. Capacitor C_2 charges through R_3 to the high voltage. When a binary 1 input occurs, T conducts connecting C_2 to the ILD. Then C_2 discharges a very high current pulse into the laser, turning it on briefly and creating the light pulse.

1.10 Optical Receiver Circuit

Two kinds of semiconductor receivers are used:

- PIN (P-intrinsic-N) diode;
- APD(Avalanche Photo-Diode).

Instead of "receiver" some times is used light detector, photo-detector or optical-to-electric converter. The receiver part of the optical communications system is relatively simple. It consists of a detector that will sense the light pulses and convert them into an electrical signal. This signal is then amplified and shaped into the original digital serial data. The most critical component, of course, is the light sensor.

The most widely used light detector is a photo diode. This is silicon PN junction diode that is sensitive to light. This diode is normally reverse-biased. Whenever light strikes the diode, this leakage current will increase significantly. It will flow through a resistor and develop a voltage drop across it. The result is an output voltage pulse.

The resulting voltage pulse is very small, so it must be amplified. This can be done by using a phototransistor. Thus the transistor amplifies the small leakage current into a larger.

The sensitivity and response time of a photo diode can be increased by adding an undraped or intrinsic (I) layer between the P and N semiconductors to form a PIN diode.

Although the PIN photodiode is extremely well suited for most fiber optic applications, its sensitively to light (responsively) is not as great as the avalanche photodiode (APD). Due to their inherent gain, typical values of responsively for APDs may range from 5 A/W to as high as 100 A/W. This is considerably higher than the PIN photodiode, which makes it extremely attractive for fiber-optic communications receivers.

Figure 1.29 shows the basic circuit used in most receivers. The current through the photodiode (PD) generated when light is sensed produced a current, which is then amplified by an amplifier (A). The pulses to ensure fast rise and fall times. The output is passed through a logic gate so that the correct binary voltage levels are produced. Most systems have a data rate of 10 to 500 Mbits/s.

The product of the bit rate and the distance usually indicates a system performance. This rating tells the fastest bit rate that can be produced over a 1-km cable. Assume a system with a 100 Mbits-km/s rating. If the distance increases, the bit rate decreases in proportion.

Another important consideration is the maximum distance between repeaters. Obviously the fewer the repeaters are better. The average distance between repeaters is now up to 100 km range.

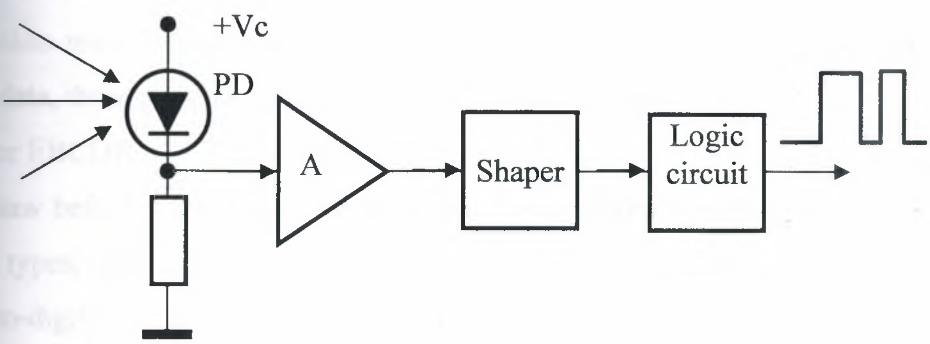


Figure 1.29 Optical Receiver Circuit

2. ENCODING

2.1 Overview

As we discussed before we must encode data into signals to send them from one place to another.

How information is encoded depends on its original format and on the format used by the communication hardware. If you want to send a love letter by smoke signal. You need to know which smoke patterns match which words in your message before you actually build your fire. Words are digital information and puffs of smoke are a digital representation of information, so defining the smoke patterns would be a form of digital-to-digital encoding. Communication technology has fundamentally the same requirements with a few additional options.

A simple signal by itself does not carry information any more than a straight line conveys words. The signal must be manipulated so that it contains identifiable changes that are recognizable to the sender and receiver as representing the information intended. First the information must be translated into agreed-upon patterns of 0s and 1s. In the case of textual data, these patterns can belong to either of two conventions:

ASCII or EBCDIC.

As we saw before, information can be of two types, digital or analog, and signals can be of two types, also digtal or analog. Therefore, four types of encoding are possible: digital-to-digital, analog-to-digital, digital-to-analog, and analog-to-analog (see Figure 2.1)

2.2 Digital-To-Digital Encoding

Digital-to-digital encoding is the representation of digital information by a digital signal. For example, when you transmit data from your computer to your printer, both the original data and the transmitted data are digital. In this type of encoding, the binary 1s and 0s generated by a computer are translated into a sequence of voltage pulses that can be propagated over a wire. Figure 2.2 shows the relationship between the digital information, the digital-to-digital encoding hardware, and the resultant digital signal.

Encoding

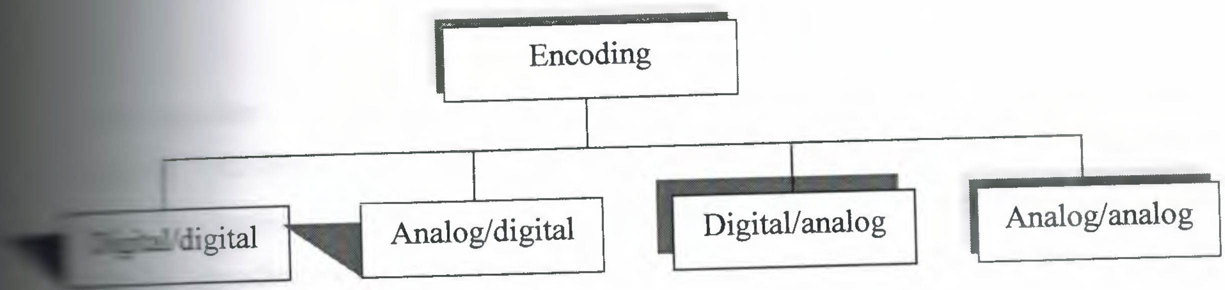


Figure 2.1 Different encoding schemes

Figure 2.2 Digital-to-digital encoding

Of the many mechanisms for digital-to-digital encoding, we will discuss only those most useful for data communication. These falls into three broad categories: unipolar, polar, and bipolar (see Figure 2.3).

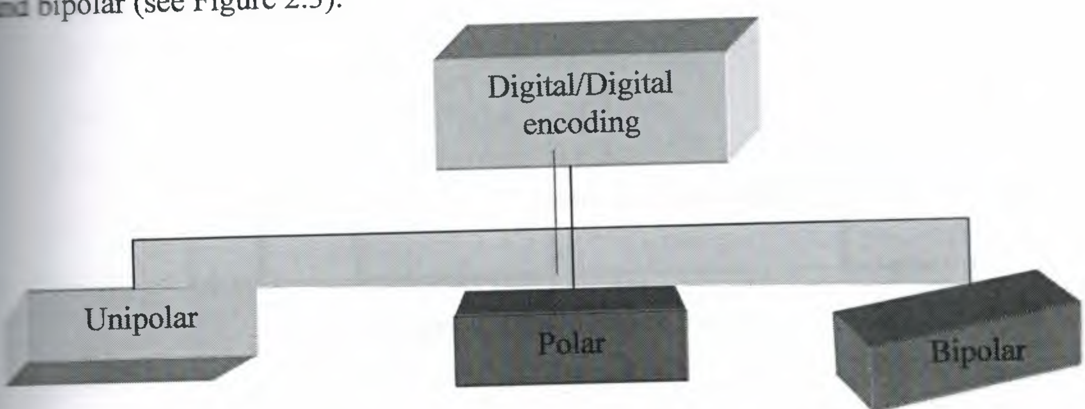


Figure 2.3 Types of digital-to-digital encoding

Unipolar encoding is simple, with only one technique in use. Polar encoding has three subcategories, NRZ, RZ and biphase, two of which have multiple variations of the own. The third option, bipolar encoding, has three variations: AMI, B8ZS, and HDB3.

Unipolar

olar encoding is very simple and very primitive. Although it is almost obsolete y, its simplicity provides an easy introduction to the concepts developed with more plex encoding systems and allows us to examine the kinds of problems that any al transmission system must overcome.

al transmission systems work by sending voltage pulses along a media link, usually re or cable. In most types of encoding, one voltage level stands for binary 0 and her level stands for binary 1. The polarity of a pulse refers to whether it is positive or tive. Unipolar encoding is so named because it uses only one polarity. Therefore, one of the two binary states is encoded. usually the 1. The other state, usually the 0, presented by zero voltage, or an idle line.

polar encoding uses only one level of value

re 2.4 shows the idea of unipolar encoding. In this example, the 1s are encoded as a tive value and the 0s are idle. In addition to being straight forward. unipolar encoding expensive to implement.

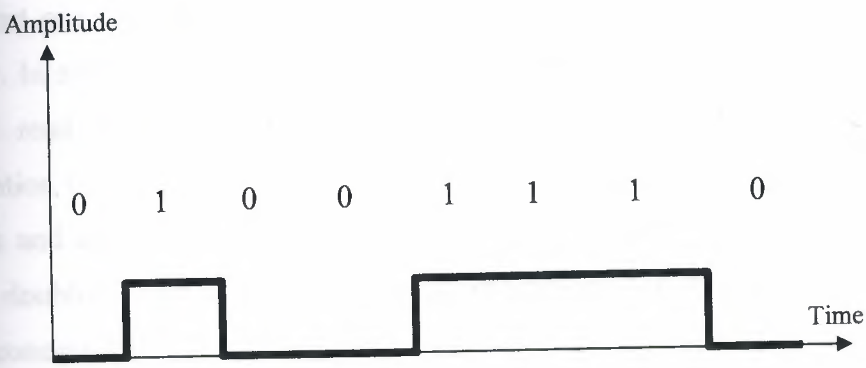


Figure 2.4 Unipolar encoding

wever, unipolar encoding has at least two problems that make it unusable: DC mponent and synchronization.

C Component

e average amplitude of a unipolar encoded signal is nonzero. This creates what is lled a direct current (DC) component (a component with zero frequency). When a signal contains a DC component, it cannot travel through media that cannot handle DC components, such as microwaves or transformers.

Synchronization

When a signal is unvarying, the receiver cannot determine the beginning and ending of each bit. Therefore, a synchronization problem in unipolar encoding can occur whenever the data stream includes a long uninterrupted series of 1s or 0s. Digital encoding schemes use changes in voltage level to indicate changes in bit type. A signal change also indicates that one bit has ended and a new bit has begun. In unipolar encoding. however, a series of one kind of bit, say seven 1s, occurs with no voltage changes, just an unbroken positive voltage that lasts seven times as long as a single 1 bit. Whenever there is no signal change to indicate the start of the next bit in a sequence, the receiver has to rely on a timer. Given an expected bit rate of 1000 bps, if the receiver detects a positive voltage lasting 0.005 seconds, it reads one 1 per 0.001 seconds, or five 1s.

Unfortunately, propagation delays can distort the timing of the signal so that, for example, five 1s can be stretched to 0006 seconds causing an extra 1 bit to be read by the receiver. That one extra bit in the data stream causes everything after it to be decoded erroneously. In addition, the receiver's clock can go out of synchronization, causing the receiver to read the bit stream erroneously. A solution developed to control the synchronization of unipolar transmission is to use a separate, parallel line that carries a clock pulse and allows the receiving device to resynchronize its timer to that of the signal. But doubling the number of lines used for transmission increases the cost and so proves uneconomical.

2.2.2 Polar

Polar encoding uses two voltage levels: one positive and one negative. By using both levels, in most polar encoding methods the average voltage level on the line is reduced and the DC component problem of unipolar encoding is alleviated In Manchester and Differential Manchester encoding, each bit consists of both positive and negative voltages, so the DC component is totally eliminated.

Polar encoding uses two levels (positive and negative) of amplitude.

Of the many existing variations of polar encoding, we will examine only the three most popular: non-return to zero (NRZ), return to zero (RZ), and biphase. NRZ encoding includes two methods: non-return to zero, level (NRZ-L), and non-return to zero-invert (NRZ-I). Biphase also refers to two methods. The first, Manchester, is the method used by Ethernet LANs. The second, Differential Manchester, is the method used by token ring LANs (see Figure 2.5).

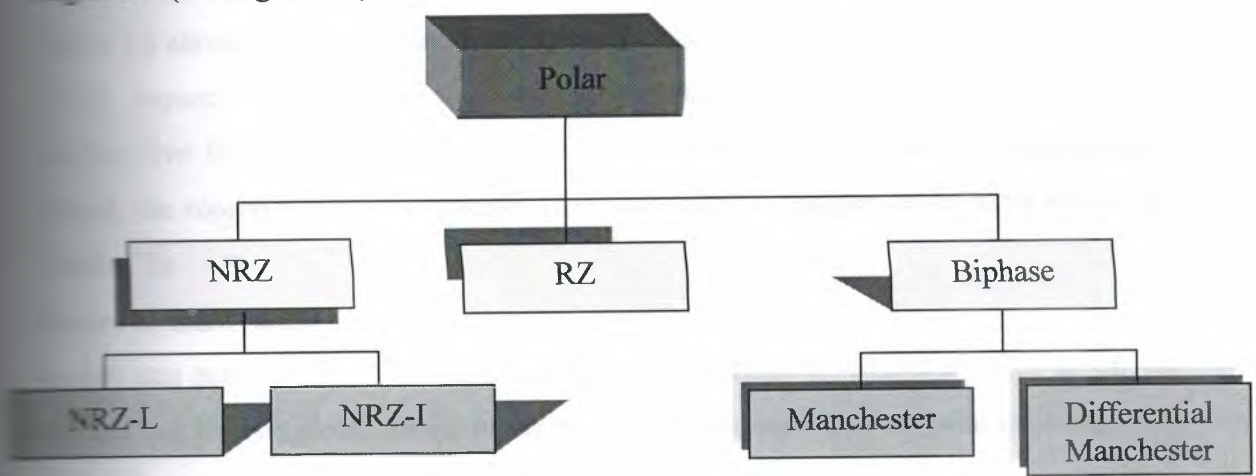


Figure 2.5 Types of Polar Encoding

Non-Return to Zero (NRZ)

In NRZ encoding, the level of the signal is always either positive or negative. Unlike in unipolar encoding, where a 0 bit is represented by an idle line, in NRZ systems if the line is idle it means no transmission is occurring at all. The two most popular methods of NRZ transmission are discussed below.

NRZ-L in NRZ-L encoding, the level of the signal depends on the type of bit it represents. A positive voltage means the bit is a 1, and a negative voltage means the bit is a 0; thus, the level of the signal is dependent upon the state of the bit. In NRZ-L the level of the signal is dependent upon the state of the bit.

NRZ-I in NRZ-I, an inversion of the voltage level represents a 1 bit. It is the transition between a positive and negative voltage, not the voltages themselves, that represents a 1 bit. A 0 bit is represented by no change. An advantage of NRZ-I over NRZ-L is that because the signal changes every time a 1 bit is encountered, it provides some synchronization.

A series of seven 1s will cause seven inversions. Each of those inversions allows the receiver to resynchronize its timer to the actual arrival of the transmission. Statistically, strings of 1s occur more frequently in transmissions than do strings of 0s. Synchronizing strings of 1s therefore goes a long way toward keeping the entire message synchronized. A string of 0s can still cause problems, but because 0s are not as likely, they are less of a threat to decoding. In NRZ-I the signal is inverted if a 1 is encountered.

Figure 2.6 shows the NRZ-L and NRZ-I representations of the same series of bits. In the NRZ-L sequence, positive and negative voltages have specific meanings: positive for 1 and negative for 0. In the NRZ-I sequence, the voltages per second are meaningless. Instead, the receiver looks for changes from one level to another as its basis for recognition of 1s.

Return to Zero (RZ)

As you can see, anytime the original data contain strings of consecutive 1s or 0s, the receiver can lose its place. As we mentioned in our discussion of unipolar encoding, one way to assure synchronization is to send a separate timing signal on a separate channel. However, this solution is both expensive and prone to errors of its own. A better solution is to somehow include synchronization in the encoded signal, something like the solution provided by NRZ-I, but one capable of handling strings of 0s as well as 1s.

To assure synchronization, there must be a signal change for each bit. The receiver can use these chances to build up, update, and synchronize its clock. As we saw above, NRZ-I accomplishes this for sequences of 1s. But to chance with every bit, we need more than just two values. One solution is return to zero (RZ) encoding, which uses three values: positive, negative, and zero. In RZ, the signal changes not between bits but during each bit. Like NRZ-L, a positive voltage means 1 and a negative voltage means 0. But, unlike NRZ-L, halfway through each bit interval, the signal returns to zero. A 1 bit is actually represented by positive-to-zero, and a 0 bit by negative-to-zero, rather than by positive and negative alone. Figure 2.7 illustrates the concept.

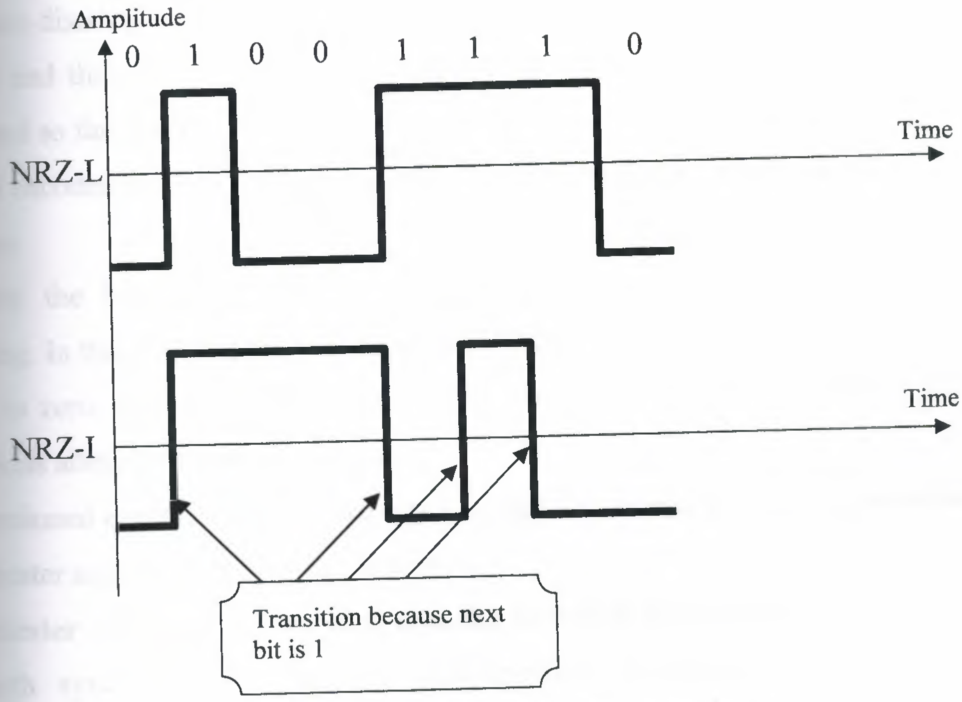


Figure 2.6 NRZ-L and NRZ-I encoding

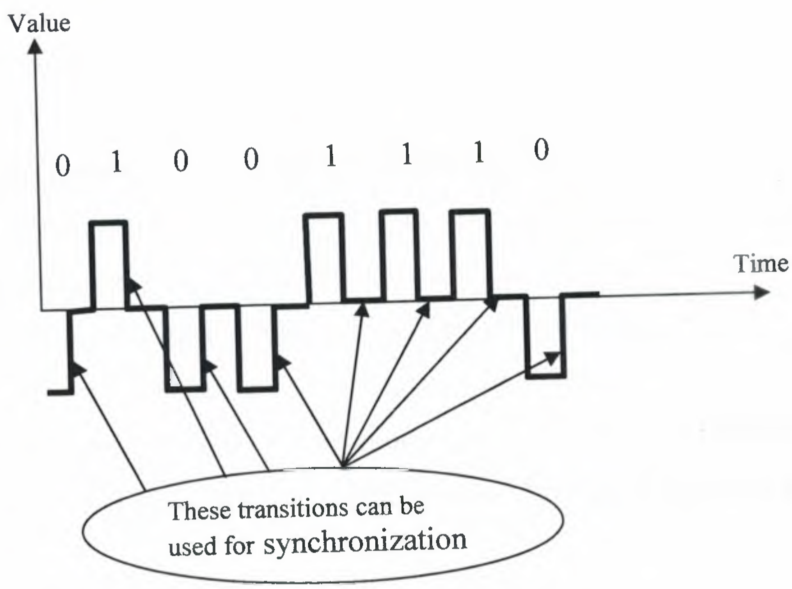


Figure 2.7 RZ encoding

The main disadvantage of RZ encoding is that it requires two signal changes to encode one bit and therefore occupies more bandwidth. But of the three alternatives we have examined so far, it is the most effective.

A good encoded digital signal must contain a provision for synchronization.

Biphase

Probably the best existing solution to the problem of synchronization is biphase encoding. In this method, the signal changes at the middle of the bit interval but does not return to zero. Instead, it continues to the opposite pole. As in RZ, these midinterval transitions allow for synchronization.

As mentioned earlier, there are two types of biphase encoding in use on networks today: Manchester and Differential-Manchester.

Manchester : Manchester encoding uses the inversion at the middle-of each bit interval for both synchronization and bit representation. A negative-to-positive transition represents binary 1 and a positive-to-negative transition represents binary 0. By using a single transition for a dual purpose, Manchester encoding achieves the same level of synchronization as RZ but with only two levels of amplitude.

In Manchester encoding the transition at the middle of the bit is used for both synchronization and bit representation.

Differential Manchester: In Differential Manchester, the inversion at the middle of the bit interval is used for synchronization, but the presence or absence of an additional transition at the beginning of the interval is used to identify the bit. A transition means binary 0 and no transition means binary 1. Differential Manchester requires two signal changes to represent binary 0 but only one to represent binary 1.

The bit representation is shown by the inversion or non-inversion at the beginning of the bit. Figure 2.8 shows the Manchester and Differential Manchester signals for the same bit pattern.

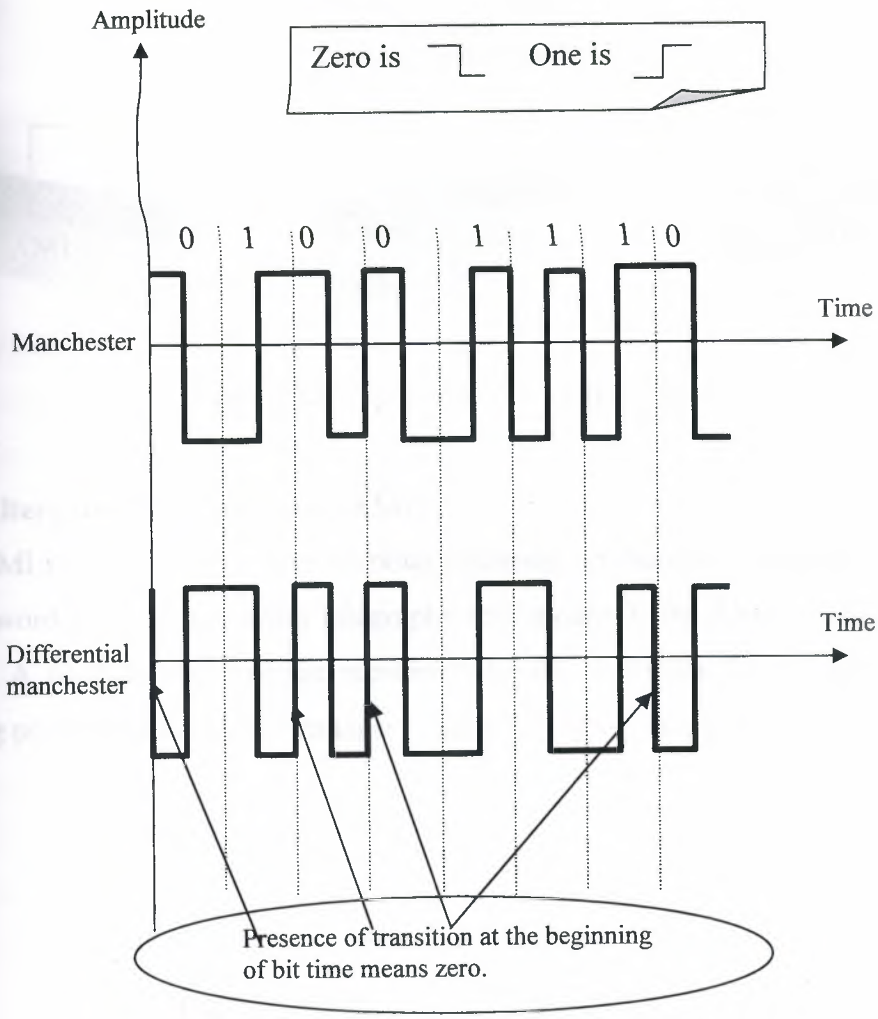


Figure 2.8 Manchester and Differential Manchester encoding

2.2.3 Bipolar

Bipolar encoding, like RZ, uses three voltage levels: positive, negative, and zero. Unlike RZ, however, the zero level in bipolar encoding is used to represent binary 0. Positive and negative voltages represent alternating 1s. If the first t bit is represented by the positive amplitude, the second will be represented by the negative amplitude, the third by the positive amplitude, and so on. This alternation occurs even when the 1 bits are not consecutive.

Three types of bipolar encoding are in popular use by the data communications industry: AMI, B8ZS, and HDB3 (see Figure 2.9).

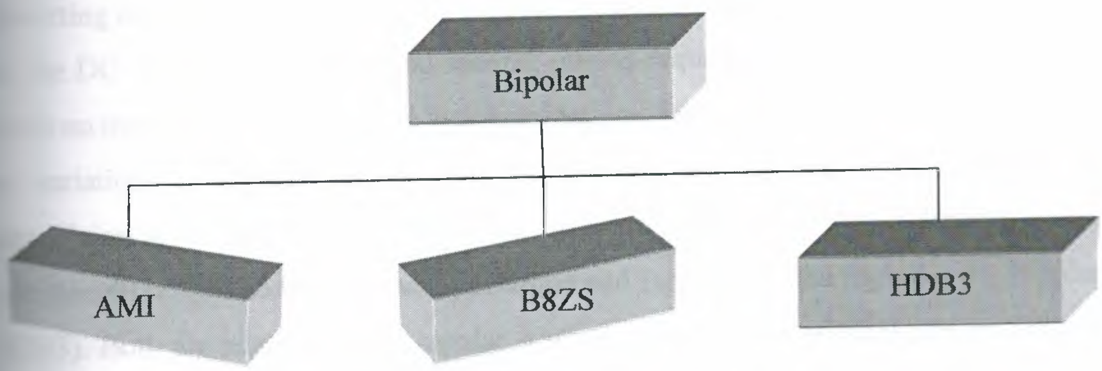


Figure 2.9 Types of Bipolar Encoding

Bipolar Alternate Mark Inversion (AMI)

Bipolar AMI is the simplest type of bipolar encoding. In the name alternate mark inversion, the word mark comes from telegraphy and means 1. So AMI means alternate 1 inversion. A neutral, zero voltage represents binary 0. Binary 1s are represented by alternating positive and negative voltages. Figure 2.10 gives an example.

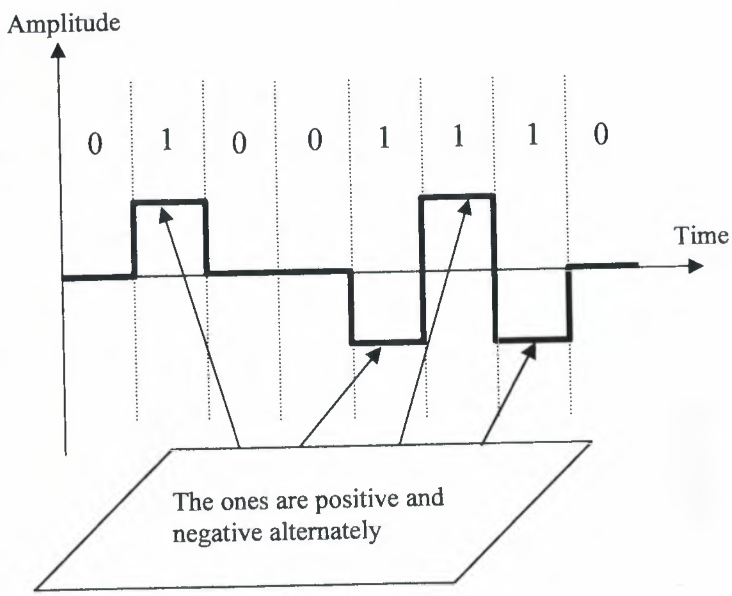


Figure 2.10 Bipolar AMI encoding

By inverting on each occurrence of a 1, bipolar AMI accomplishes two things: first, the DC component is zero, and second, a long-sequence of 1s stays synchronized. There is no mechanism to ensure the synchronization of a long string of 0s.

Two variations of bipolar AMI have been developed to solve the problem of synchronizing sequential 0s. The first, used in North America, is called bipolar 8-zero substitution (B8ZS). The second, used in Europe and Japan, is called high-density bipolar 3 (HDB3). Both are adaptations of bipolar AMI that modify the original pattern only in the case of multiple consecutive 0s.

Bipolar 8-Zero Substitution (B8ZS)

B8ZS is the convention adopted in North America to provide synchronization of long strings of 0s. In most situations, B8ZS functions identically to bipolar AMI. Bipolar AMI changes poles with every 1 it encounters. These changes provide the synchronization needed by the receiver. But the signal does not change during a string of 0s, so synchronization is often lost.

The difference between B8ZS and bipolar AMI occurs whenever eight or more consecutive 0s are encountered in the data stream. The solution provided by B8ZS is to force artificial signal changes, called violations, within the 0 string. Anytime eight 0s occur in succession, B8ZS introduces changes in the pattern based on the polarity of the previous 1 (the 1 occurring just before the 0s). See Figure 2.11

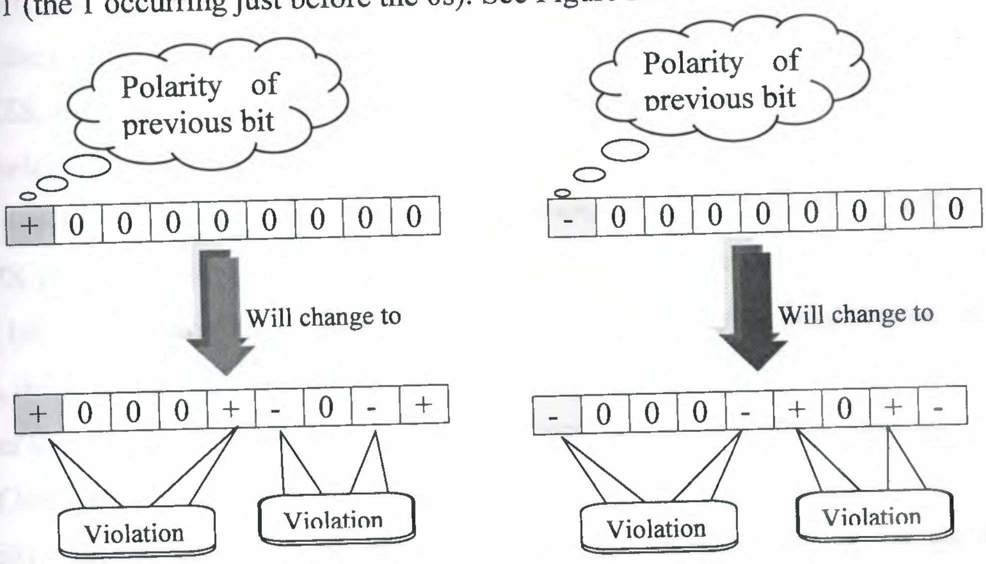


Figure 2.11 B8ZS encoding

If the previous *1 bit was positive, the eight 0s will be encoded as zero, zero, zero,* positive, negative, zero, negative, positive. Remember that the receiver is looking for alternating polarities to identify 1s.

When it finds two consecutive positive charges surrounding three 0s, it recognizes the pattern as a deliberately introduced violation and not an error. It then looks for the second pair of the expected violations. When it finds them, the receiver translates all eight bits to 0s and reverts back to normal bipolar AMI mode.

If the polarity of the previous 1 is negative, the pattern of violations is the same but with inverted polarities. Both positive and negative patterns are shown in Figure 2.11.

In B8ZS if eight 0s come one after another, we change the pattern in one of two ways based on the polarity of the previous 1.

High-Density Bipolar 3 (HDB3)

The problem of synchronizing strings of consecutive 0s is solved differently in Europe and Japan than in the United States. This convention, called HDB3, introduces changes into the bipolar AMI pattern every time four consecutive 0s are encountered instead of waiting for the eight expected by B8ZS in North America. Although the name is HDB3. the pattern changes whenever there are four 0s in succession (see Figure 2.12).

In HDB3 if four 0s come one after another, we change the pattern in one of four ways based on the polarity of the previous 1 and the number of 1s since the last substitution.

As in B8ZS, the pattern of violations in HDB3 is based on the polarity of the previous 1 bit. But unlike B8ZS, HDB3 also looks at the number of 1s that have occurred in the bit stream since the last substitution. Whenever the number of 1s since the last substitution is odd, B8ZS puts a violation in the place of the fourth consecutive 0. If the polarity of the previous bit was positive, the violation is positive. If the polarity of the previous bit was negative, the violation is negative.

Whenever the number of 1s since the last substitution is even, B8ZS puts violations in the places of both the first and the fourth consecutive 0s. If the polarity of the previous bit was positive, both violations are negative. If the polarity of the previous bit was negative, both violations are positive. All four patterns are shown in Figure 2.12.

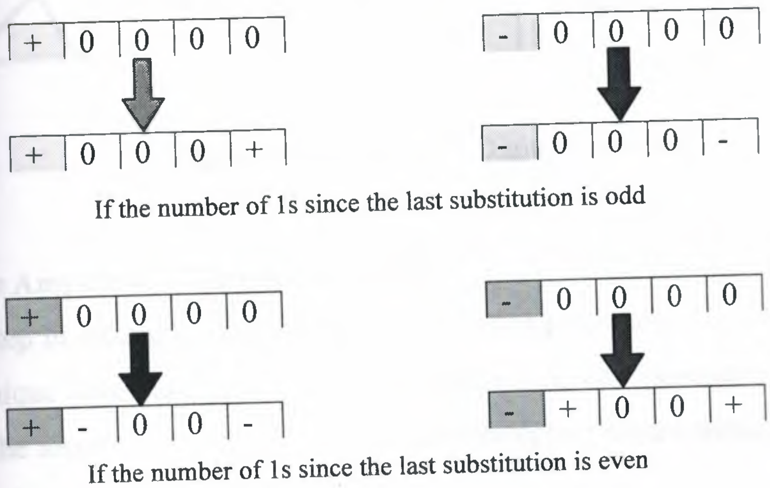


Figure 2.12 HDB3 Encoding

2.3 Analog-To-Digital Encoding

Analog-to-digital encoding is the representation of analog information by a digital signal. To record a singer's voice onto a compact disc, for example, you use digital medium to replicate analog information. To do so you need to reduce the potentially infinite number of values in an analog message so that they can be represented as a digital stream with a minimum loss of information. Several methods for analog-to-digital encoding will be discussed later in this chapter. Figure 2.13 shows the analog-to-digital encoder, called codec (coder-decoder).

In analog-to-digital encoding, we are representing the information contained continuous wave form as a series of digital pulses (1s or 0s).

So far, the encoding systems we have been examining have focused on the format of the transporting signal. The structure of the transporting signal is not the problem. Instead, the problem is how to translate information from an infinite number of values to a discrete number of values without sacrificing sense or quality.

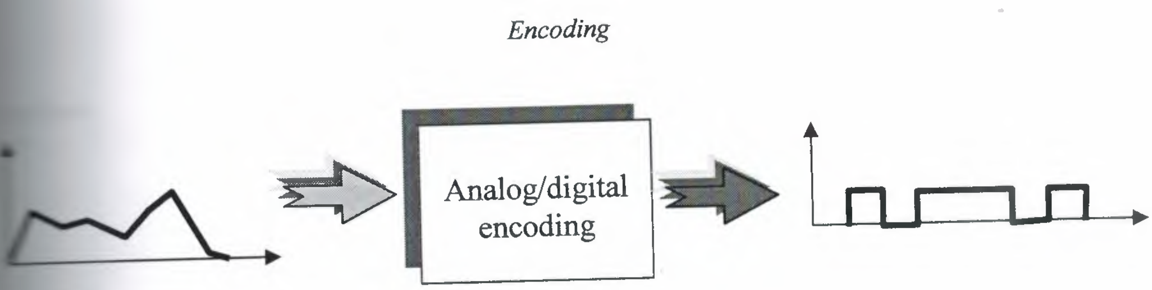


Figure 2.13 Analog-to-Digital Encoding

2.3.1 Pulse Amplitude Modulation (PAM)

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

The method of sampling used in PAM is more useful to other areas of engineering than it is to data communication. However, PAM is the foundation of an important analog-to-digital encoding method called pulse code modulation (PCM).

In PAM, the original signal is sampled at equal intervals as shown in Figure 2.14. PAM uses a technique called sample and hold. At a given moment the signal level is read, then held briefly. The sampled value occurs only instantaneously in the actual wave form, but is generalized over a still short but measurable period in the PAM result.

The reason PAM is not useful to data communications is that, although it translates the original wave form to a series of pulses, these pulses are still of any amplitude (still an analog signal, not digital). To make them digital, we must modify them by using pulse code modulation (PCM).

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

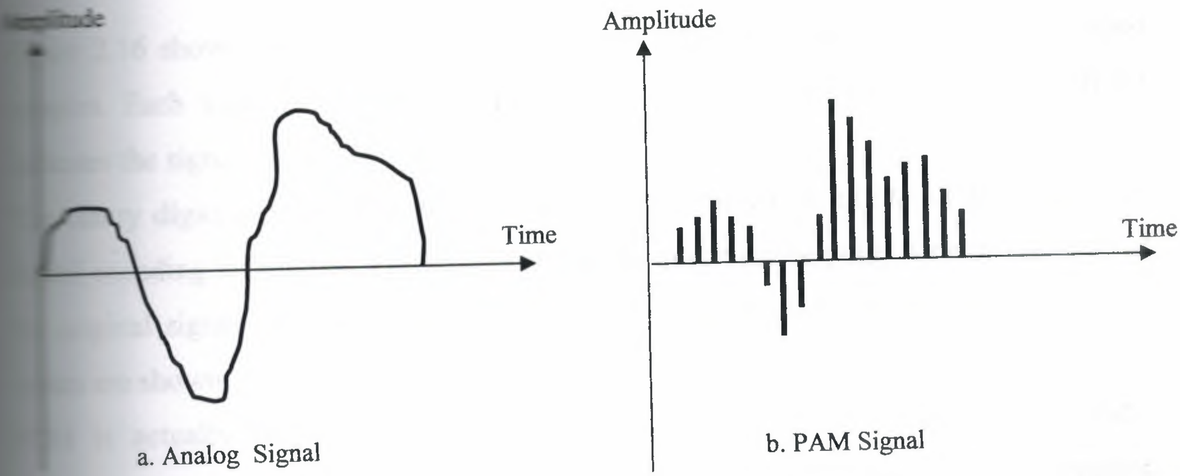


Figure 2.14 PAM

2.3.2 Pulse Code Modulation (PCM)

PCM modifies the pulses created by PAM to create a completely digital signal. To do so, PCM first quantizes the PAM pulses. Quantization is a method of assigning integral values in a specific range to sampled instances. The result of quantization is presented in Figure 2.12.

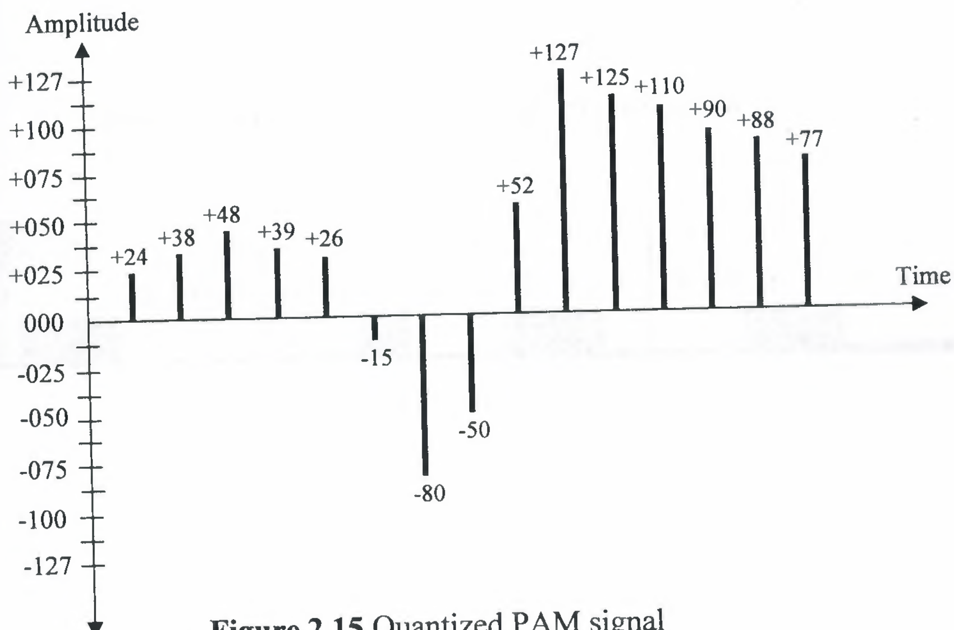


Figure 2.15 Quantized PAM signal

Figure 2.16 shows a simple method of assigning sign and magnitude values quantized samples. Each value is translated into its seven-bit binary equivalent. The eighth bit indicates the sign.

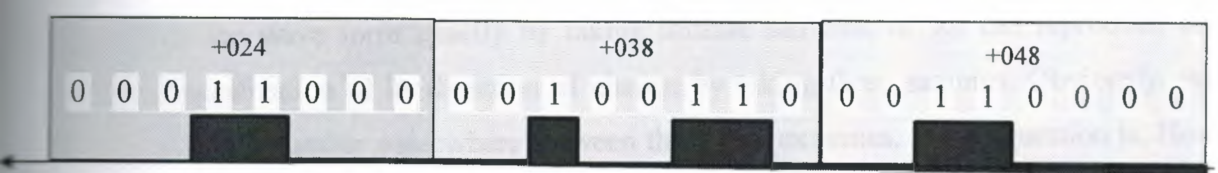
The binary digits are then transformed into a digital signal using one of the digital to-digital encoding techniques. Figure 2.17 shows the result of the pulse code modulation of the original signal encoded finally into a unipolar signal. Only the first three sampled values are shown.

PCM is actually made up of four separate processes: PAM, quantization, binary encoding, and digital-to-digital encoding. Figure 2.18 shows the entire process graphic form. PCM is the sampling method used to digitize voice in T-line transmission in the North American telecommunication system.

+024	00011000	-015	10001111	+125	01111101
+038	00100110	-080	11010000	+110	01101110
+048	00110000	-050	10110010	+090	01011010
+039	00100111	+052	00110110	+088	01011000
+026	00011010	+127	01111111	+077	01001101

Sign bit + is 0 , - is 1

Figure 2.16 Quantizing using sign and magnitude



Direction of transfer

Figure 2.17 PCM

Encoding

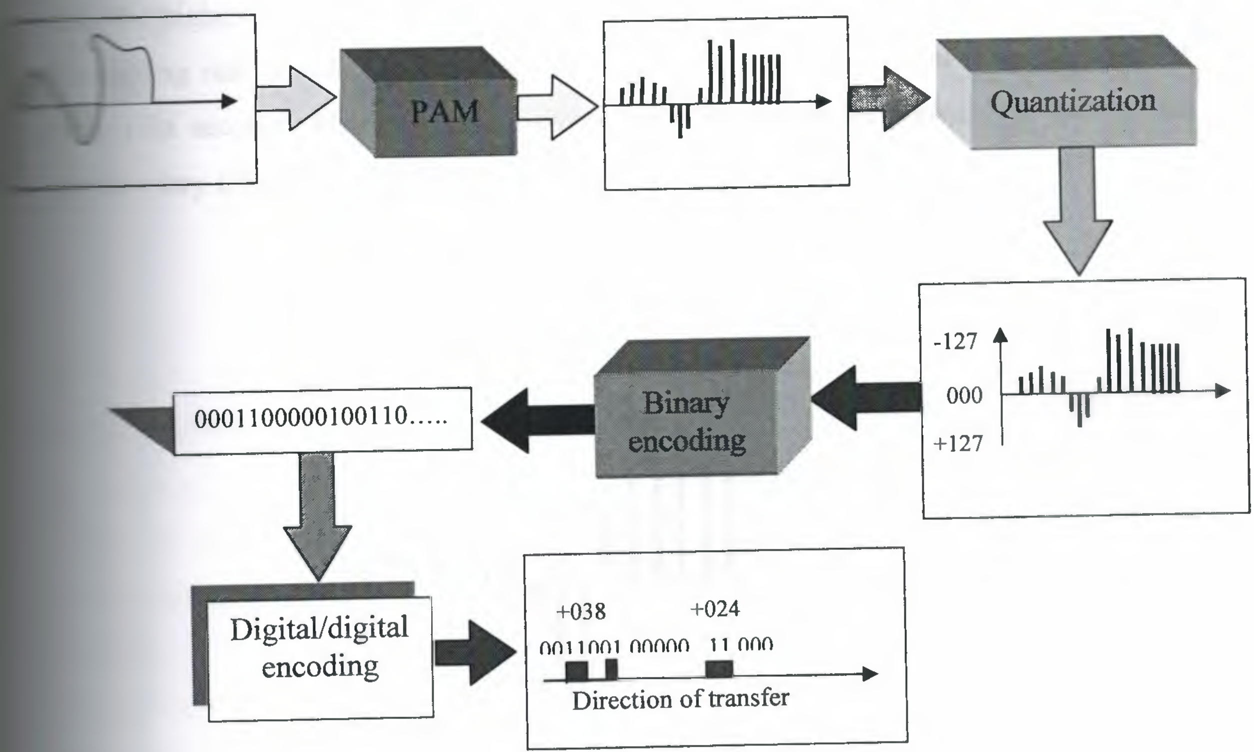


Figure 2.18 From analog signal PCM digital code

2.3.3 Sampling Rate

As you can tell from the preceding figures, the accuracy of any digital reproduction of an analog signal depends on the number of samples taken. Using PAM and PCM, we can reproduce the wave form exactly by taking infinite samples, or we can reproduce the barest generalization of its direction of change by taking three samples. Obviously, we prefer to find a number somewhere between these two extremes. So the question is, How many samples are sufficient?

Actually, it requires remarkably little information for the receiving device to reconstruct an analog signal. According to the Nyquist theorem, to ensure the accurate reproduction of an original analog signal using PAM, the sampling rate must be at least twice the highest frequency of the original signal.

So if we want to sample telephone voice information with maximum frequency 3300 Hz, we need a sampling rate of 6600 samples per second.

In actual practice, 8000 samples are taken to compensate for imperfection in later *processing. According to the Nyquist theorem, the sampling rate must be at least two times the highest frequency.*

A sampling rate of twice a frequency of x Hz means that the signal must be sampled every 1/2x seconds. Using the voice-over-phone-lines example above, that means one sample every 1/8000 second. Figure 2.19 illustrates the concept.

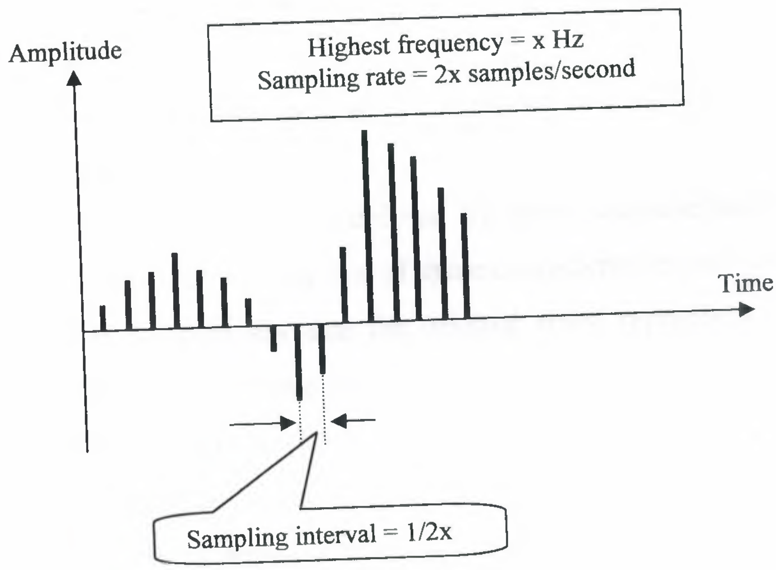


Figure 2.19 Nyquist theorem

2.4 Digital-To-Analog Encoding

Digital-to-analog encoding is the representation of digital information by an analog signal. When you transmit data from one computer to another across a public access phone line, for example, the data start out as digital, but because telephone wires carry analog signals, the data must be converted. The digital data must be encoded on an analog signal that has been manipulated to look like two distinct values that correspond to binary 1 and binary 0. Figure 2.20 shows the relationship between the digital information, the digital-to-analog encoding hardware, and the resultant analog signal.

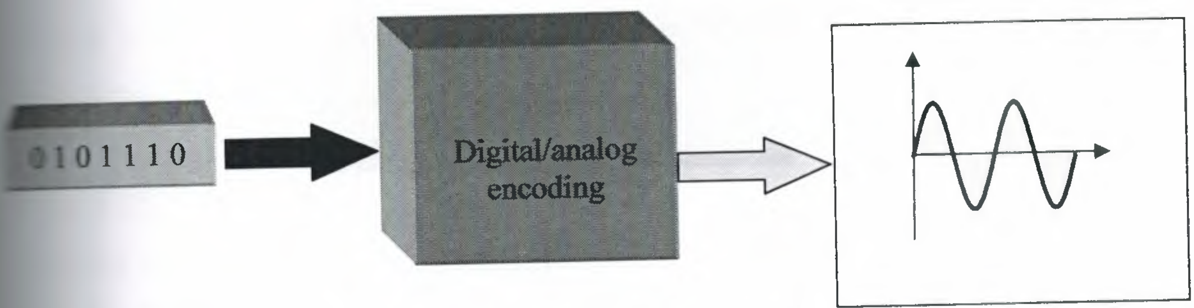


Figure 2.20 Digital-to-Analog encoding

Of the many mechanisms for digital-to-analog encoding, we will discuss only those most useful for data communications.

As discussed previously, a sine wave is defined by three characteristics: amplitude, frequency, and phase. When we vary any one of these characteristics, we create a second version of that wave. If we then say that the original wave represents binary 1. the variation can represent binary 0, or vice versa. So, by changing one aspect of a simple electrical signal back and forward, we can use it to represent digital data. Any of the three characteristics listed above can be altered in this way, giving us at least three mechanisms for encoding digital data into an analog signal: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK).

In addition, there is a fourth (and better) mechanism that combines changes in both amplitude and phase called quadrature amplitude modulation (QAM). QAM is the most efficient of these options and is the mechanism used in all modern modems (see Figure 2.21).

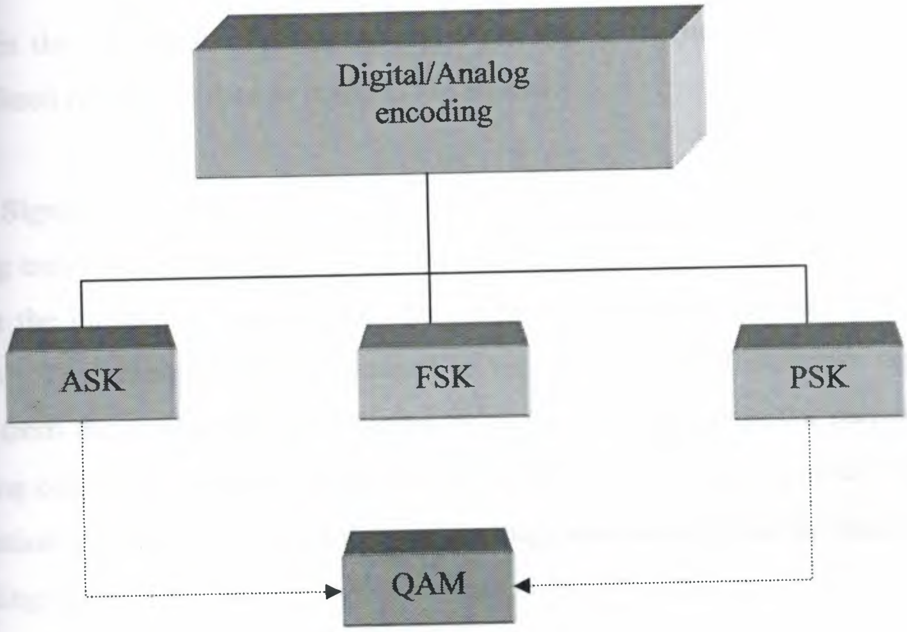


Figure 2.21 Types of digital-to-analog encoding

2.4.1 Aspects of Digital-to-Analog Encoding

Before we discuss specific methods of digital-to-analog encoding, two basic issues must be defined: bit/baud rate and carrier signal.

Bit Rate and Baud Rate

Two terms used frequently in data communication are bit rate and baud rare. Bit rate is the number of bits transmitted during one second. Baud rate refers to the number of signal Units per second that are required to represent those bits. In discussions of computer efficiency the bit rate is the more important -we want to know how long it takes to process each piece of information. In data transmission, however, we are more concerned with how efficiently we can move that data from place to place, whether in pieces or blocks. The fewer signal units required, the more efficient the system and the less bandwidth required to transmit more bits; so we are more concerned with baud rate. The baud rate determines the bandwidth required to send the signal.

Bit rate equals the baud rate times the number of bits represented by each signal unit. The baud rate equals the bit rate divided by the number of bits represented by each signal shift. Bit rate is always greater than or equal to the baud rate.

Bit rate is the number of bits per second. Baud rate is the number of signal units per second. Baud rate is less than or equal to the bit rate.

Carrier Signal

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

2.4.2 Amplitude Shift Keying (ASK)

In amplitude shift keying (ASK), the strength of the signal is varied to represent binary 1 or 0. Both frequency and phase remain constant while the amplitude changes. Which voltage represents 1 and which represents 0 is left to the system designers. A bit duration is the period of time that defines one bit. The peak amplitude of the signal during each bit duration is constant and its value depends on the bit (0 or 1). The speed of transmission using ASK is limited by the physical characteristics of the transmission medium. Figure 2.22 gives a conceptual view of ASK.

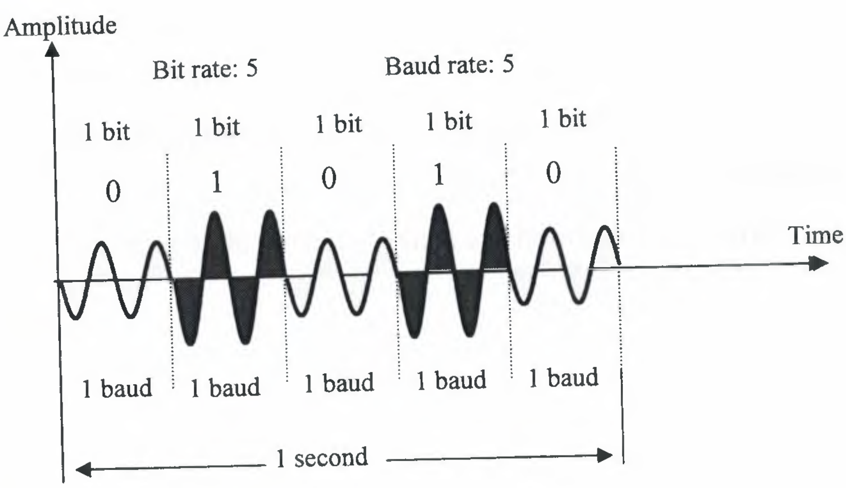


Figure 2.22 ASK Encoding

Unfortunately, ASK transmission is highly susceptible to noise interference. The term noise refers to unintentional voltages introduced onto a line by various phenomena such as heat or electromagnetic induction created by other sources. These unintentional voltages combine with the signal to change the amplitude.

Some types of noise, for example thermal noise, are constant enough not to interfere with the intelligibility of the signal. Impulse noise, however, is a sudden surge of energy that can wipe out an entire section of a transmission by inserting high-amplitude spikes where low amplitude was intended. In that case, a section of the signal that was intended to be received as one or more 0s will read as 1s. You can see how surges in voltage would be especially problematic for ASK, which relies solely on amplitude for recognition. Noise usually affects the amplitude; therefore, ASK is the encoding method most affected by noise.

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

2.4.3 Frequency Shift Keying (FSK)

In frequency shift keying (FSK), the frequency of the signal is varied to represent binary 1 or 0. The frequency of the signal during each bit duration is constant and its value depends on the bit (0 or 1): both peak amplitude and phase remain constant. Figure 2.23 gives the conceptual view of FSK.

FSK avoids most of the noise problems of ASK. Because the receiving device is looking for specific frequency changes over a given number of periods, it can ignore voltage spikes. The limiting factors of FSK are the physical capabilities of the carrier.

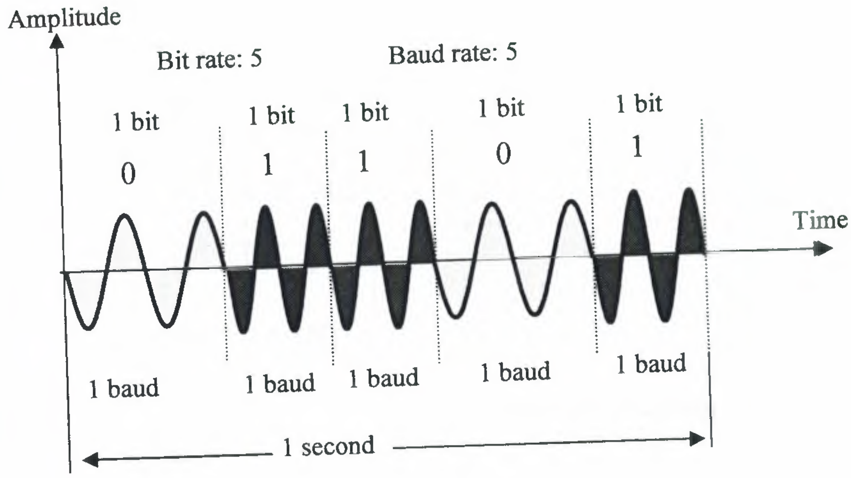


Figure 2.23 FSK encoding

2.4.4 Phase Shift Keying (PSK)

In phase shift keying (PSK), the phase is varied to represent binary 1 or 0. Both peak amplitude and frequency remain constant as the phase changes. For example, if we start with a phase of 0 degrees to represent binary 0, then we can change the phase to 180 degrees to send binary 1. The phase of the signal during each bit duration is constant and its value depends on the bit (0 or 1). Figure 2.24 gives a conceptual view of PSK.

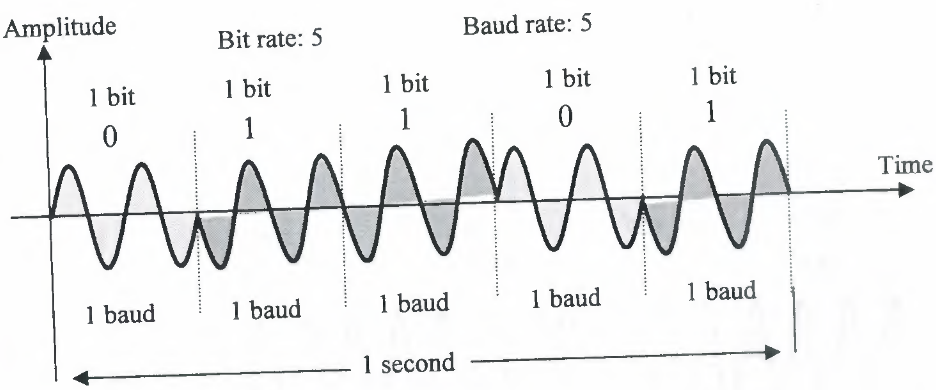
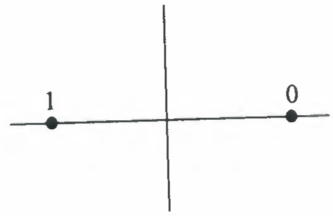


Figure 2.24 PSK

Bit	Phase
0	0
1	180



Bits

Constellation diagram

Figure 2.25 PSK constellation

The above method is often called 2-PSK, or binary PSK, because two different phases (0 and 180 degrees) are used in the encoding. Figure 2.25 makes this point clearer by showing the relationship of phase to bit value. A second diagram, called a constellation or phase-state diagram, shows the same relationship by illustrating only the phases. PSK is not susceptible to the noise degradation that affects ASK, nor to the bandwidth limitations of FSK. This means that smaller variations in the signal can be detected reliably by the receiver. Therefore, instead of utilizing only two variations of a signal, each representing one bit, we can use four variations and let each phase shift represent two bits (see Figure 2.26).

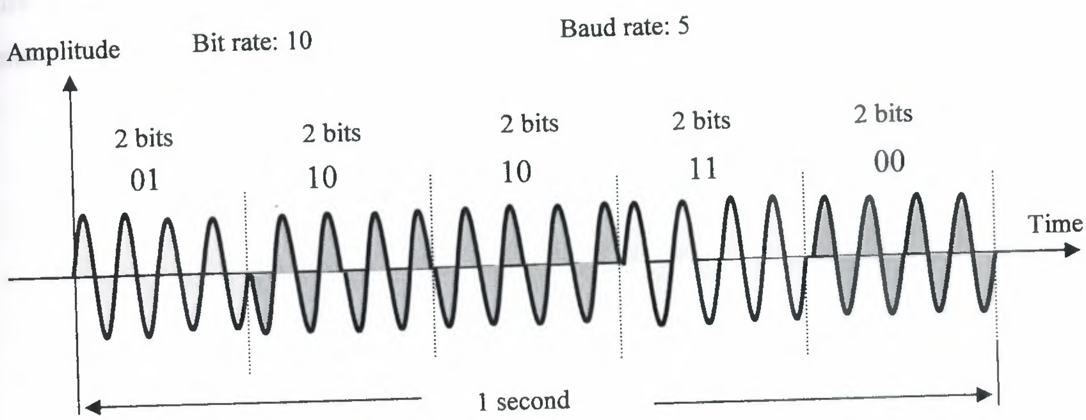
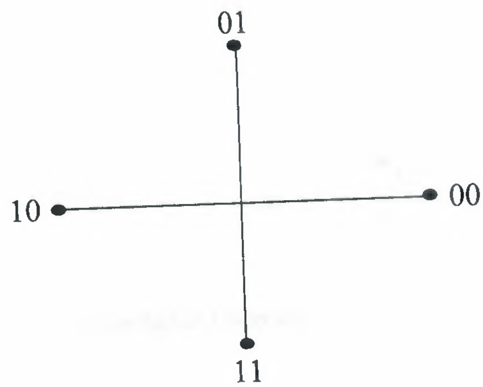


Figure 2.26 4-PSK

The constellation diagram for the signal in Figure 2.26 is given in Figure 2.27. A phase of 0 degrees now represents 00; 90 degrees represents 01; 180 degrees represents 10; and 270 degrees represents 11. This technique is called 4-PSK or Q-PSK. The pair of bits represented by each phase is. Called a Dibit. We can transmit data two times as fast using 4-PSK as we can using 2-PSK.

Dibit	Phase
00	0
01	90
10	180
11	270

Dibit
(2 bits)



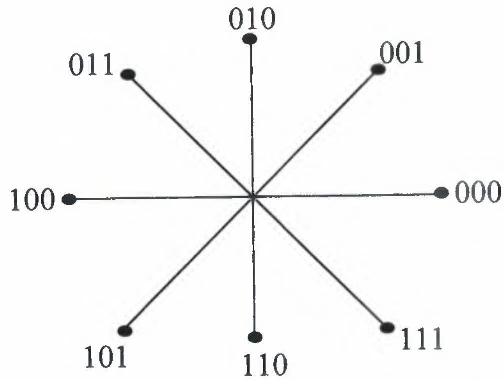
Constellation Diagram

Figure 2.27 4-PSK characteristics

We can extend this idea to 8-PSK. Instead of 90 degrees, we now vary the signal by shifts of 45 degrees. With 8 different phases, each shift can represent three bits (one tribit) at a time. (As you can see, the relationship of number of bits per shift to number of phases is a power of two. When we have four possible phases, we can send two bits at a time—2^2 equals 4. When we have eight possible phases, we can send three bits at time—2^3 equals 8). Figure 2.28 shows the relationships between the phase shifts and the tribits each one represents. 8-PSK is three times faster than 2-PSK

Tribit	Phase
000	0
001	45
010	90
011	135
100	180
101	225
110	270
111	315

Tribit
(3 bits)



Constellation Diagram

Figure 2.28 8-PSK characteristics

2.4.5 Quadrature Amplitude Modulation (QAM)

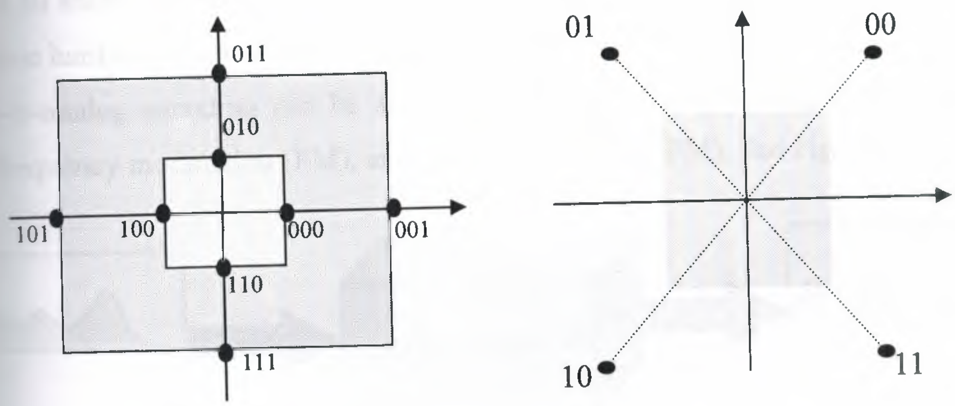
PSK is limited by the ability of the equipment to distinguish small differences in phase. This factor limits its potential bit rate.

So far, we have been altering only one of the three characteristics of a sine wave at a time to achieve our encoding, but what if we alter two? Bandwidth limitations make combinations of FSK with other changes practically useless. But why not combine ASK and PSK? Then we could have x variations in phase and y variations in amplitude, giving us x times y possible variations and the corresponding number of bits per variation. Quadrature amplitude modulation (QAM) does just that.

The term quadrature is derived from the restrictions required for minimum performance and is related to trigonometry.

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

Possible variations of QAM are numerous. Theoretically any measurable number of changes in amplitude can be combined with any measurable number of changes in phase. Figure 2.29 shows two possible configurations, 4-QAM and 8-QAM. In both cases, the number of amplitude shifts is fewer than the number of phase shifts. Because amplitude changes are susceptible to noise and require greater shift differences than do phase changes, the number of phase shifts used by a QAM system is always larger than the number of amplitude shifts. The time-domain plot corresponding to the 8-QAM signal in Figure 2.29 is shown in Figure 2.30.



8-QAM
2 amplitudes, 4 phases

4 -QAM
1 amplitude, 4 phases

Figure 2.29 4-QAM and 8-QAM constellation

Other geometric relationships besides concentric circles are also possible. The first example, three amplitudes and 12 phases, handles noise best because of a greater ratio of phase shift to amplitude. It is the ITU-T recommendation. The second example, four amplitudes an eight phases, is the OSI recommendation. If you examine the graph carefully, you will notice that although it is based on concentric circles, not every intersection of phase and amplitude is utilized. In fact, 4 times 8 should allow for 32 possible variations. But by using only half of those possibilities, the measurable differences between shifts are increased and greater signal readability is ensured. In addition, several QAM designs link specific amplitudes with specific phases. This means that even with the noise problems associated with amplitude shifting, the meaning of a shift can be recovered from phase information. In general, therefore, a second advantage of QAM encoding over ASK encoding is its lower susceptibility to noise.

2.5 Analog-To-Analog Encoding

Analog-to-analog encoding is the representation of analog information by an analog signal. Radio, that familiar utility, is an example of an analog-to-analog communication. Figure 2.30 shows the relationship between the analog information, the analog-to-analog conversion hardware, and the resultant analog signal.

Analog-to-analog encoding can be accomplished in three ways: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). See Figure 2.31.

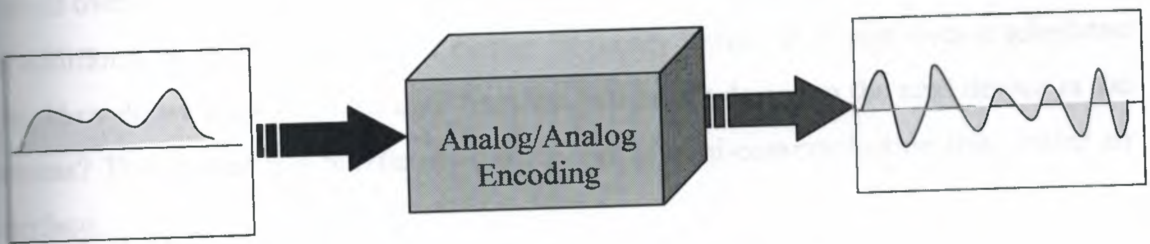


Figure 2.30 Analog-to-Analog Encoding

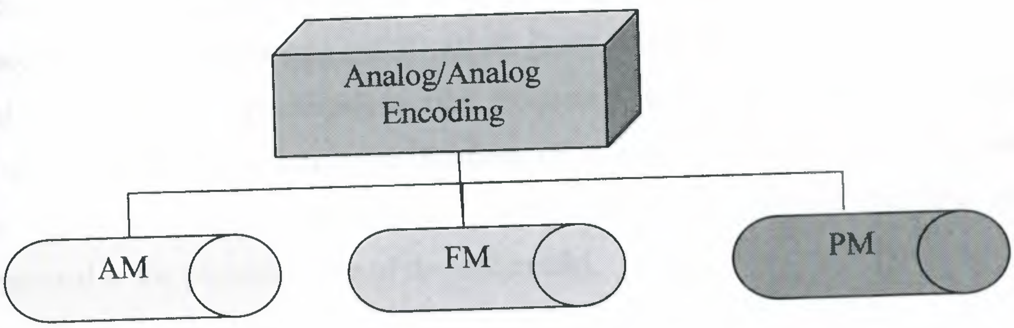


Figure 2.31 Types of analog-to-analog encoding

3. TRANSMISSION OF DIGITAL DATA: INTERFACES AND MODEMS

3.1 Overview

Once we have encoded our information into a format that can be transmitted, the next step is to investigate the transmission process itself. Information-processing equipment such as PCs generates encoded signals but ordinarily require assistance to transmit those signals over a communication link. For example, a PC generates a digital signal but needs an additional device to modulate a carrier frequency before it is sent over a telephone line. How do we relay encoded data from the generating device to the next device in the process? The answer is a bundle of wires, a sort of mini-communication link, called an interface.

Because an interface links two devices not necessarily made by the same manufacturer, its characteristics must be defined and standards must be established. Characteristics of an interface include its mechanical specifications (how many wires are used to transport the signal); its electrical specifications (the frequency, amplitude, and phase of the expected signal); and its functional specifications (if multiple wires are used, what does each one do?). These characteristics are all described by several popular standards and are incorporated in the physical layer of the OSI model.

3.2 Digital Data Transmission

Of primary concern when considering the transmission of data from one device to another is the wiring. And of primary concern when considering the wiring is the data stream. Do we send one bit at a time, or do we group bits into larger groups and, if so, how? The transmission of binary data across a link can be accomplished either in parallel mode or serial mode. In parallel mode, multiple bits are sent with each clock pulse. In serial mode, one bit is sent with each clock pulse. While there is only one way to send parallel data, there are two subclasses of serial transmission: synchronous and asynchronous (see Figure 3.1).

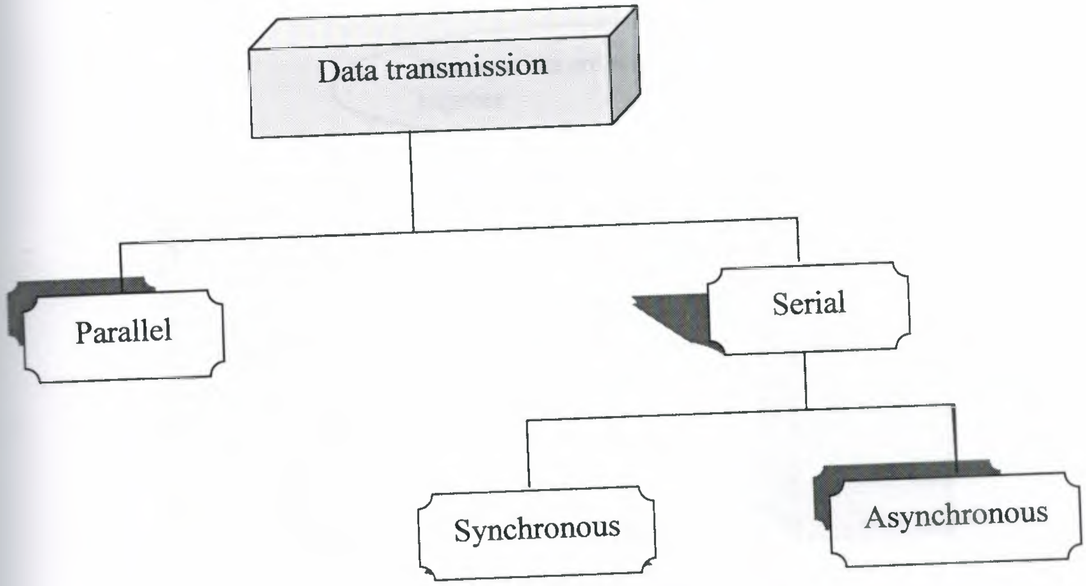


Figure 3.1 Data transmission

3.2.1 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 one. 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. That way each bit has its own wire, and all *n* bits of one group can be transmitted with each clock pulse from one device to another.

Figure 3.2 shows how parallel transmission works for $n = 8$. Typically the eight wires are bundled in a cable with a connector at each end.

The advantage of parallel transmission is speed. All else being equal, parallel transmission can increase the transfer speed by a factor of *n* over serial transmission. But there is a significant disadvantage: cost. Parallel transmission requires *n* communication lines (wires in the example) just to transmit the data stream. Because this is expensive, parallel transmission is usually limited to short distances, up to a maximum of say 25 feet.

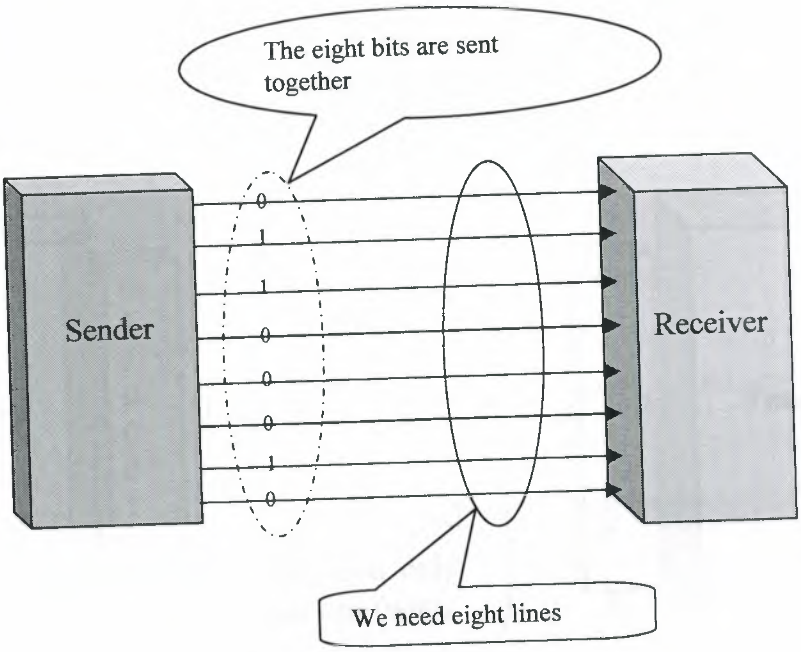


Figure 3.2 Parallel Transmission

3.2.2 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 (see Figure 3.3). The advantage of serial over parallel transmission is that with only one communication channel, serial transmission reduces the cost of transmission over parallel by roughly a factor *of n*.

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

Serial transmission occurs in one of two ways: asynchronous or synchronous.

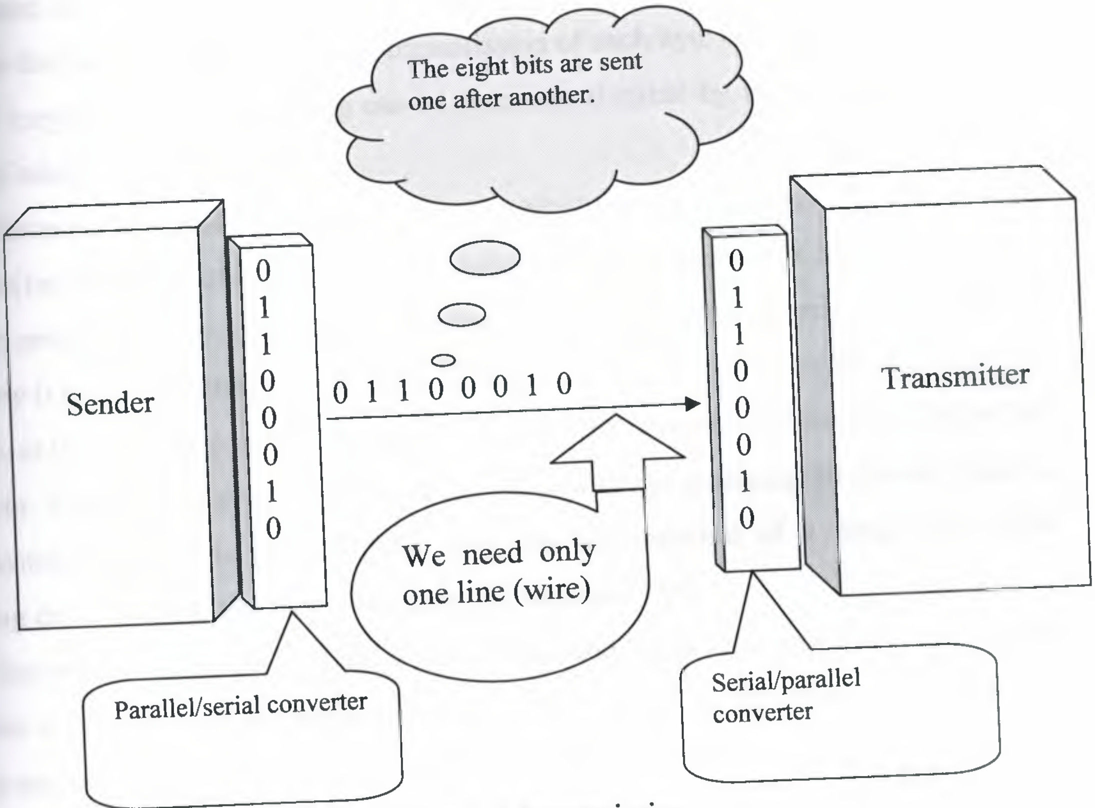


Figure 3.3 Serial Transmission

3.2.3 Asynchronous Transmission

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

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

In asynchronous transmission we send one start bit (0) at the beginning and one or more stop bits (1s) at the end of each byte. There may be a gap between each byte.

The start and stop bits and the gap alert the receiver to the beginning and end of each byte and allow it to synchronize with the data stream. This mechanism is called asynchronous because, at the byte level, sender and receiver do not have to be synchronized, But within each byte. The receiver must still be synchronized with the incoming bit stream. That is, some synchronization is required, But only for the duration of a single byte. The receiving device resynchronizes at the onset of each new byte.

When the receiver detects a start bit, it sets a timer and begins counting bits as they come in. After *n* bits the receiver looks for a stop bit. As soon as ft detects the stop bit, it ignores any received pulses until it detects the next start bit.

Asynchronous here means "asynchronous at the byte level," but the bits are still synchronized; their durations are the same.

Figure 3.4 is a schematic illustration of asynchronous transmission. In this example, the start bits are 0s, the stop bits are 1s, and the gap is represented by an idle line rather than by additional stop bits.

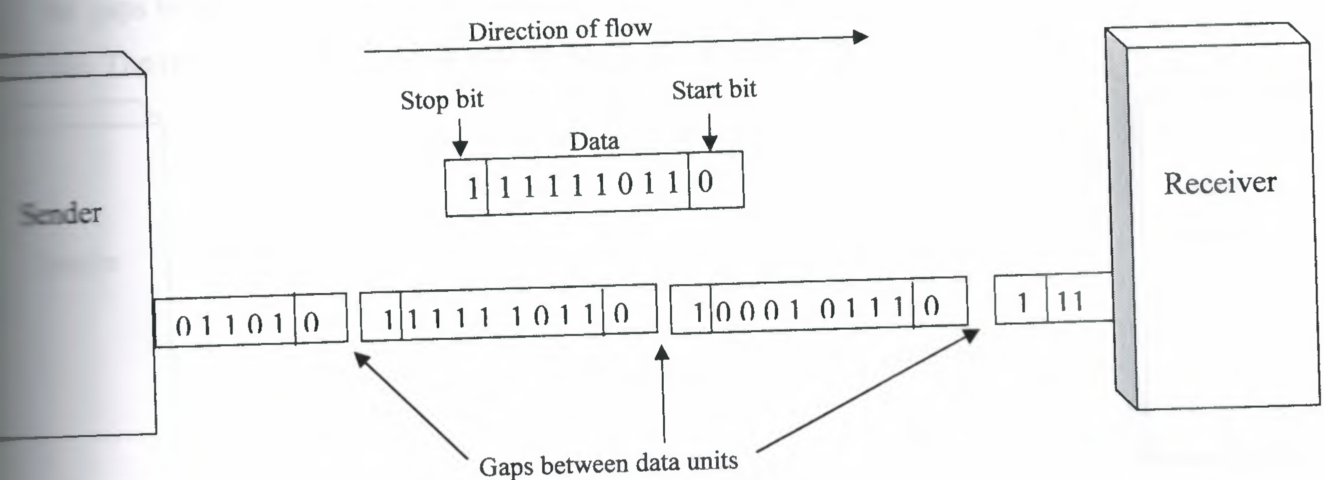


Figure 3.4 Asynchronous transmission

The addition of stop and start bits and the insertion of raps into the bit stream make asynchronous transmission slower than forms of transmission that can operate without the addition of control information. But it is cheap and effective, two advantages that make it an attractive choice for situations like low-speed communication. For example, the connection of a terminal to a computer is a natural application for asynchronous transmission. A user types only one character at a time, types extremely slowly in data processing terms, and leaves unpredictable gaps of time between each character.

3.2.4 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.

In synchronous transmission we send bits one after another without start/stop bits or gaps. It is the responsibility of the receiver to group the bits.

Figure 3.5 gives a schematic illustration of synchronous transmission. We have drawn in the divisions between bytes. In reality, those divisions do not exist; the sender puts its data onto the line as one long string. If the sender wishes to send data in separate bursts, the gaps between bursts must be filled with a special sequence of 0s and 1s that means idle. The receiver counts the bits as they arrive and groups them in eight-bit units.

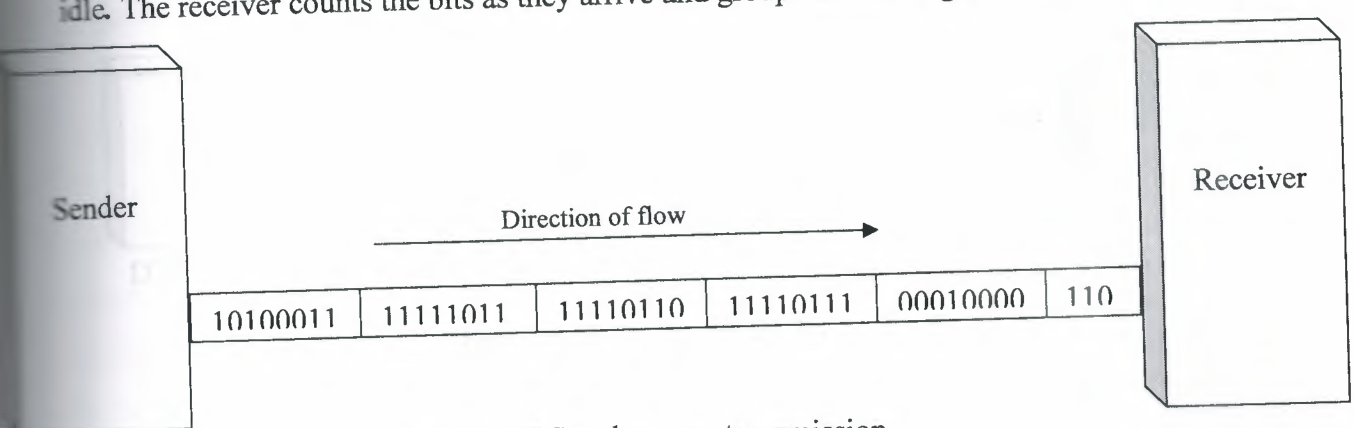


Figure 3.5 Synchronous transmission

Without gaps and start/stop bits, there is no built-in mechanism to help the receiving device adjust its bit synchronization in midstream. Timing becomes very important, therefore, because the accuracy of the received information is completely dependent on the ability of the receiving device to keep an accurate count of the bits as they come in. The advantage of synchronous transmission is speed. With no extra bits or gaps to introduce at the sending end and remove at the receiving end and, by extension, with fewer bits to move across the link, synchronous transmission is faster than asynchronous transmission. For this reason, it is more useful for high-speed applications like the transmission of data from one computer to another. Byte synchronization is accomplished in the data link layer.

3.3 DTE-DCE Interface

At this point we must clarify two terms important to computer networking: data terminal equipment (DTE) and data circuit-terminating equipment (DCE). There are usually four basic functional units involved in the communication of data: a DTE and DCE on one end and a DCE and DTE on the other end, as shown in Figure 3.6. The DTE generates the data and passes them, along with any necessary control characters, to a DCE. The DCE does the job of converting the signal to a format appropriate to the transmission medium and introducing it onto the network link. When the signal arrives at the receiving end, this process is reversed.

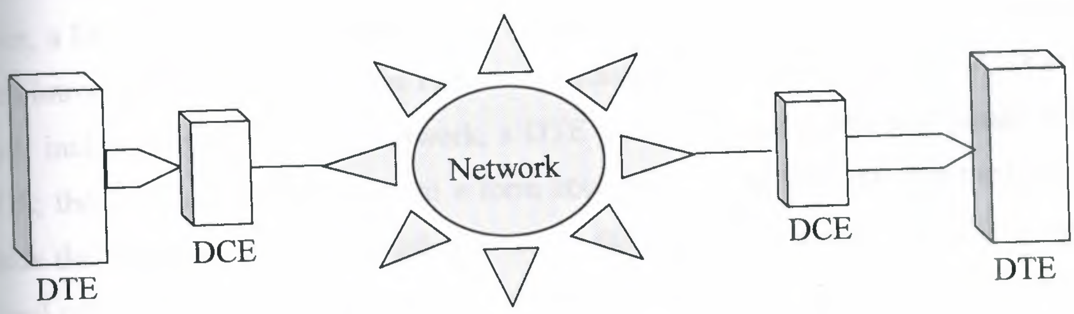


Figure 3.6 DTEs and DCEs

3.3.1 Data Terminal Equipment (DTE)

Data terminal equipment (DTE) includes any unit that functions either as a source of or as a destination for binary digital data. At the physical layer, it can be a terminal, microcomputer, computer, printer, fax machine, or any other device that generates or consumes digital data. DTEs do not often communicate directly with one another; they generate and consume information but need an intermediary to be able to communicate. Think of a DTE as operating the way your brain does when you talk. Let's say you have an idea that you want to communicate to a friend. Your brain creates the idea but cannot transmit that idea to your friend's brain by itself. Unfortunately or fortunately, we are not a species of mind readers. Instead, your brain passes the idea to your vocal chords and mouth, which convert it to sound waves that can travel through the air or over a telephone line to your friend's ear and from there to his or her brain, where it is converted back into information. In this model, your brain and your friend's brain are DTEs. Your vocal chords and mouth are your DCE. His or her ear is also a DCE. The air or telephone wire is your transmission medium.

A DTE is any device that is a source of or destination for binary digital data.

3.3.2 Data Circuit-Terminating Equipment (DCE)

Data circuit-terminating equipment (DCE) includes any functional unit that transmits or receives data in the form of an analog or digital signal through a network. At the physical layer, a DCE takes data generated by a DTE, converts them to an appropriate signal, and then introduces the signal onto the telecommunication link. Commonly used DCEs at this layer include modems. In any network, a DTE generates digital data and passes it to a DCE; the DCE converts the data to a form acceptable to the transmission medium and sends the converted signal to another DCE on the network. The second DCE takes the signal off the line, converts it to a form usable by its DTE, and delivers it. To make this communication possible, both the sending and receiving DCEs must use the same encoding method (e.g., FSK). Much the way that if you want to communicate to someone who understands only Japanese. You must speak Japanese, The two DTEs do not need to be coordinated with each other. But each of them must be coordinated with its own DCE and the DCEs must be coordinated so that data translation occurs without loss of integrity.

3.3.3 Standards

Over the years, many standards have been developed to define the connection between a DTE and a DCE (see Figure 3.7). Though their solutions differ, each standard provides a model for the mechanical, electrical, and functional characteristics of the connection. Of the organizations involved in DTE-DCE interface standards, the most active are the Electronic Industries Association (EIA) and the International Telecommunication Union—Telecommunication Standards Committee (ITU-T). The EIA standards are called, appropriately enough, EIA-232, EIA-442, EIA-449, and so on. The ITU-T standards are called the V series and the X series.

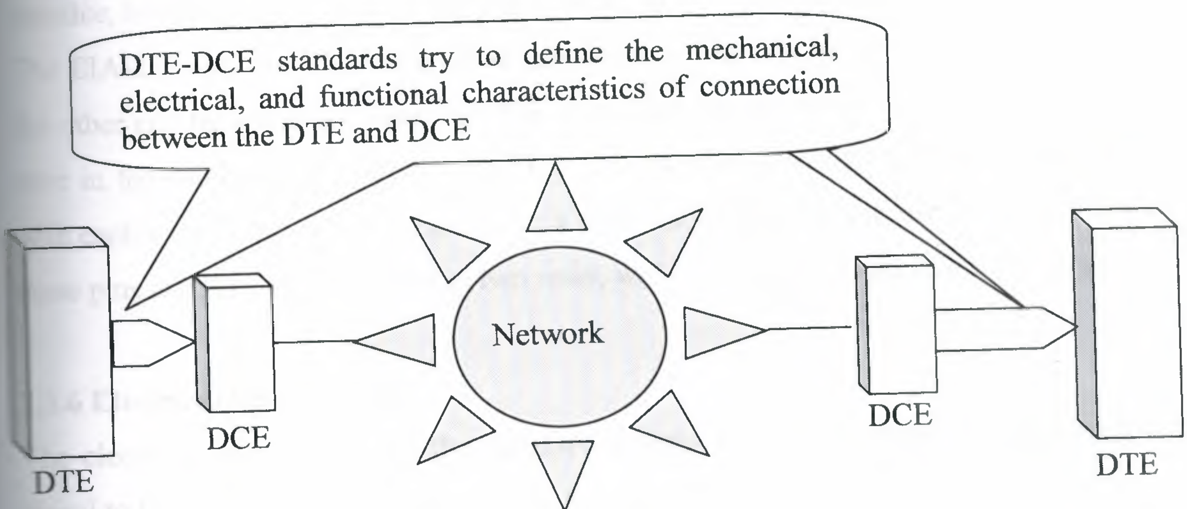


Figure 3.7 DTE-DCE interface

3.3.4 EIA-232 Interface

One important interface standard developed by the EIA is the EIA-232, which defines the mechanical, electrical, and functional characteristics of the interface between a DTE and a DCE. Originally issued in 1962 as the RS-232 standard (recommended standard), the EIA-232 has been revised several times. The most recent version. EIA-232-D, defines not only the type of connectors to be used but also the specific cable and plugs and the functionality of each pin.

EIA-232 (previously called RS-232) defines the mechanical, electrical, and functional characteristics of the interface between a DTE and a DCE.

3.3.5 Mechanical Specification

The mechanical specification of the EIA-232 standard defines the interface as a 25-wire cable with a male and a female DB-25 pin connector attached to either end. The length of the cable may not exceed 15 meters (about 50 feet).

A DB-25 connector is a plug with 25 pins or receptacles, each of which is attached to a single wire with a specific function. With this design, the EIA has created the possibility of 25 separate interactions between a DTE and a DCE. Fewer are actually used in current practice, but the standard allows for future inclusion of functionality.

The EIA-232 calls for a 25-wire cable terminated at one end by a male connector and at the other end by a female connector. The term male connector refers to a plug with each wire in the cable connecting to a pin. The term female connector refers to a receptacle with each wire in the cable connecting to a metal tube, or sheath. In the DB25 connector, these pins and tubes are arranged in two rows, with 13 on the top and 12 on the bottom.

3.3.6 Electrical Specification

The electrical specification of the standard defines the voltage levels and the type of signal to be transmitted in either direction between the DTE and the DCE.

EIA-232 states that all data must be transmitted as logical 1s and 0s (called mark and space) using non-return to zero, level (NRZ-L) encoding, with 0 defined as a positive voltage and 1 defined as a negative voltage.

Sending the Data The electrical specification for sending data is shown in Figure 3.8. Rather than defining a single range bounded by highest and lowest amplitudes, EIA-232 defines two distinct ranges, one for positive voltages and one for negative. A receiver recognizes and accepts as an intentional signal any voltage that falls within these ranges, but no voltages that fall outside the ranges. To be recognized as data, the amplitude of a signal must fall between 3 and 15 volts or between -3 and -15 volts. By allowing valid signals to fall within two 12-volt ranges, EIA-232 makes it unlikely that degradation of a signal by noise will affect its recognizability. In other words, as long as a pulse falls within one of the acceptable ranges, the precision of that pulse is unimportant.

Figure 3.8 shows a square wave degraded by noise into a curve. The amplitude of the fourth bit is lower than intended (compared to that of the second bit). and rather than staying at one single voltage, it covers a range of many voltages. If the receiver were looking only for a fixed voltage, the degradation of this pulse would have made it unrecoverable. The bit would also have been unrecoverable if the receiver were looking only for pulses that held a single voltage for their entire duration.

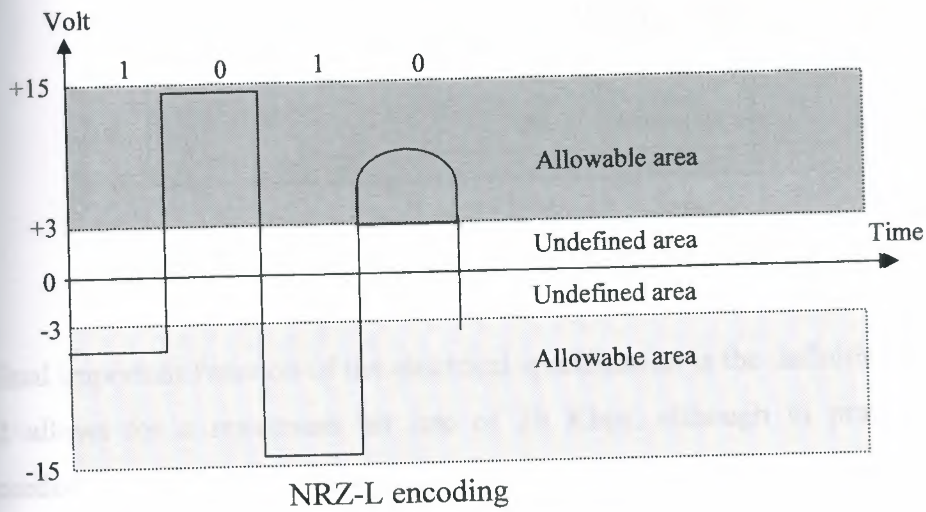


Figure 3.8 Electrical specification for sending data in EIA –232

3.3.7 Control and Timing

Only 4 wires out of the 25 available in an EIA-232 interface are used for data functions. The remaining 21 are reserved for functions like control, timing, grounding, and testing. The electrical specifications for these other wires are similar to those governing data transmission, but simpler. Any of the other functions is considered ON if it transmits a voltage of at least +3; and OFF if it transmits a voltage with a value less than -3 volts. Figure 3.9 shows one of these signals. The specification for control signals is conceptually reversed from that for data transmission. A positive voltage means ON and a negative voltage means OFF. Also note that OFF is still signified by the transmission of a specific voltage range. An absence of voltage on one of these wires while the system is running means that something is not working properly, and not that the line is turned off.

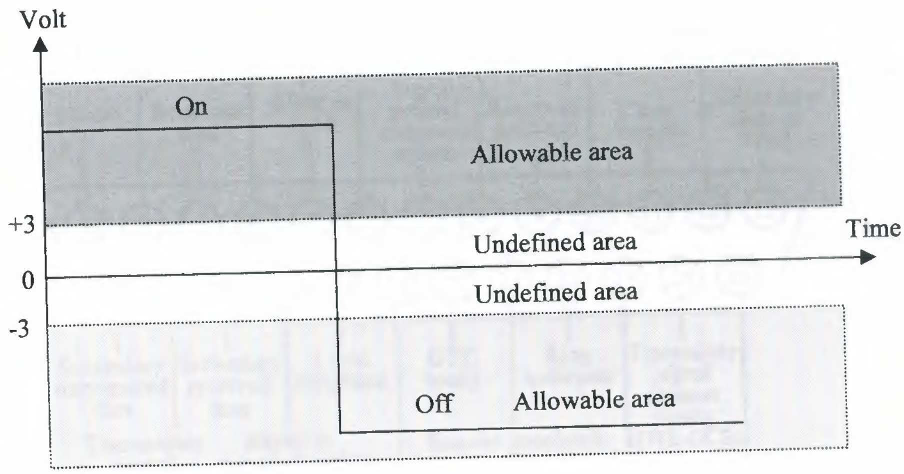


Figure 3.9 Electrical specification for control signals in EIA –232

A final important function of the electrical specification is the definition of bit rate. EIA-232 allows for a maximum bit rate of 20 Kbps, although in practice this often is exceeded.

3.3.8 Functional Specification

EIA-232 defines the functions assigned to each of the *25* pins in the DB-25 connector. Figure 3.10 shows the ordering and functionality of each pin of a male connector. Remember that a female connector will be the mirror image of the male, so that pin 1 in the plug matches tube 1 in the receptacle, and so on. Each communications function has a mirror or answering function for traffic in the opposite direction, to allow for full-duplex operation. For example, pin 2 is for transmitting data, while pin 3 is for receiving data. In this way both parties can transmit data at the same time. As you can see from Figure 3.10, not every pin is assigned a specific function. Pins 9 and 10 are reserved for future use. Pin 11 is as yet unassigned.

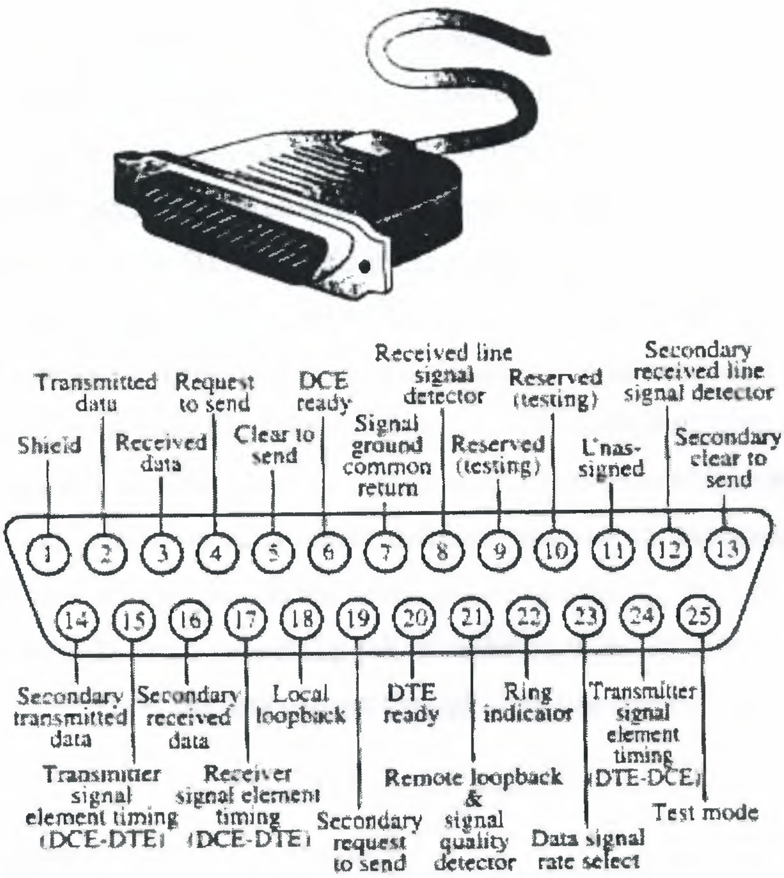


Figure 3.10 Functions of pins in EIA –232

3.3.9 Null Modem

Suppose you need to connect two DTEs in the same building, for example, two workstations, or a terminal to a workstation. Modems are not needed to connect two compatible digital devices directly; the transmission never needs to cross analog lines, such as telephone lines, and therefore does not need to be modulated. But you do need an interface to handle the exchange (readiness establishment, data transfer, receipt, etc.), just as an EIA-232 DTE-DCE cable does.

The solution, provided by the EIA standard, is called a null modem. A null modem provides the DTE-DCE/DCE-DTE interface without the DCEs. But why use a null modem? If all you need is the interface, why not just use a standard EIA-232 cable? To understand the problem, examine Figure 3.11. Part a shows a connection using a telephone network

The two DTEs are exchanging information through DCEs. Each DTE sends its data through pin 2 and the DCE receives it on pin 2; and each .DTE receives data through pin 3 that has been forwarded by the DCE using its own pin 3. As you can see, the EIA-232 cable connects DTE pin 2 to DCE pin 2 and DCE pin 3 to DTE pin 3. Traffic using pin 2 is always outgoing from the DTEs. Traffic using pin 3 is always incoming to the DTEs. A DCE recognizes the direction of a signal and passes-it along to the appropriate circuit.

Part *b* of the figure shows what happens when we use the same connections between two DTEs. Without DCEs to switch the signals to or from the appropriate pins, both DTEs are attempting to transmit over the same pin 2 wire and to receive over the same pin 3 wire. The DTEs are transmitting to each others transmit pins, not to their receive pins. The receive circuit (3) is void because it has been isolated completely from the transmission. The transmit circuit (2) therefore ends up full of collision noise and signals that can never be received by either DTE. No data can get through from one device to another.

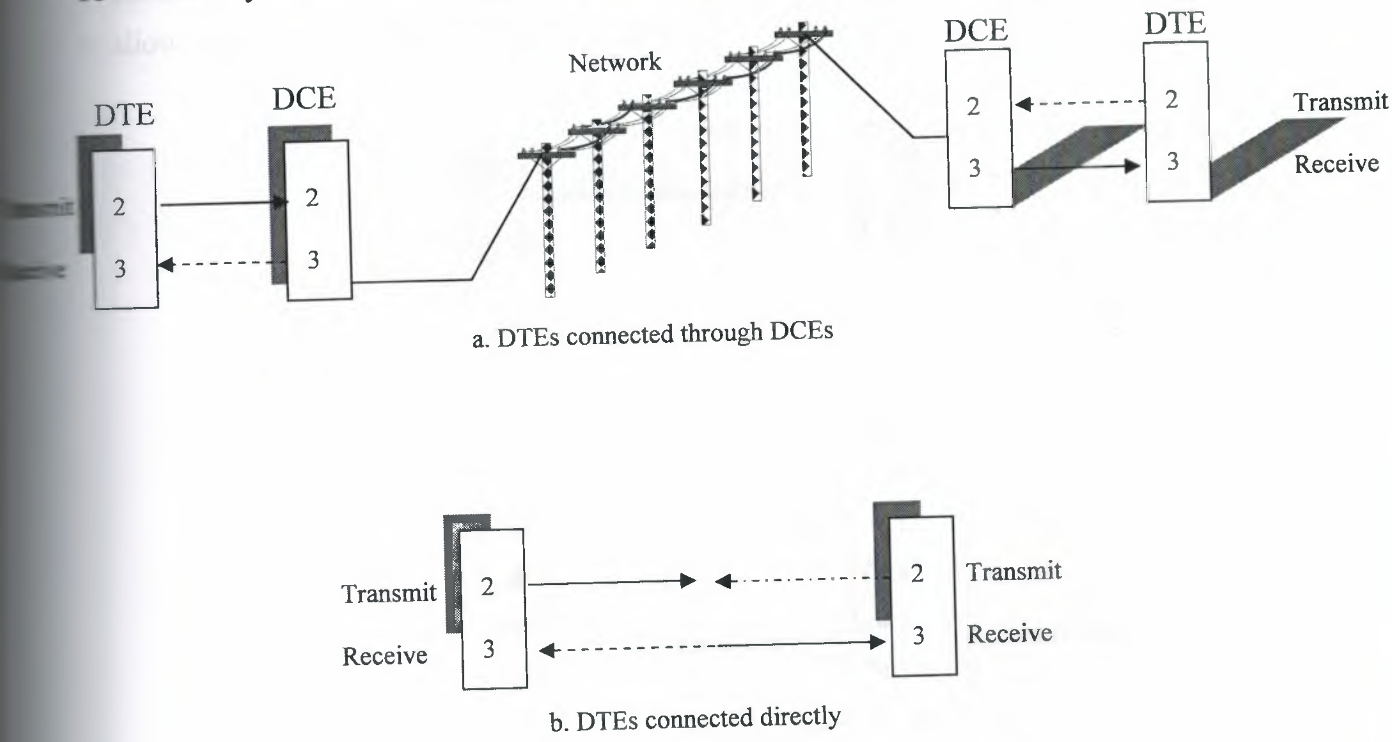


Figure 3.11 Using regular data pin connections with and without DCEs

Crossing Connections For transmission to occur, the wires must be crossed so that pin 2 of the first DTE connects to pin 3 of the second DTE; and pin 2 of the second DTE connects to pin 3 of the first. These two pins are the most important. Several other pins, however, have similar problems and their wires also need reconnection.

A null modem is an EIA-232 interface that completes the necessary circuits to fool the DTEs at either end into believing that they have DCEs and a network between them. Because its purpose is to make connections, a null modem can be either a length of cable or a device, or you can make one yourself using a standard EIA-232 cable or a breakout box that allows you to cross-connect wires in any way you desire. Of these options, the cable is the most commonly used and the most convenient (see Figure 3.12).

Other Differences Null modems have up to 25 wires, of these 25. The most important are those needed for DTE-to-DTE transmission, and are shown in Figure 3.12. Note that whereas an EIA-232 DTE-DCE interface cable has a female connector at the DTE end and a male connector at the DCE end, a null modem has female connectors at both ends to allow it to connect to the E1A-232 DTE ports, which are male.

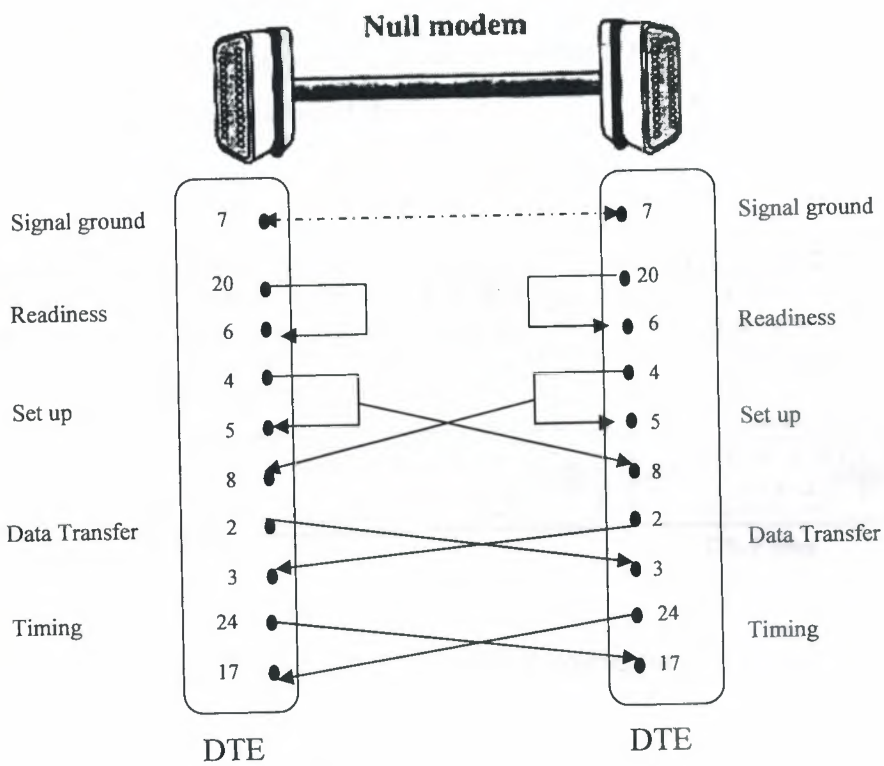


Figure 3.12 Null modem pin connections

3.4 Other Interface Standards

Both data rate and cable length (signal distance capability) are restricted by EIA-232; data rate to 20Kbps and cable length to 50 feet (15 meters). To meet the needs of users who require more speed and / or distance, the EIA and the ITU-T have introduced additional interface standards: EIA-449, EIA-530, and X.21.

3.4.1 EIA-449

The mechanical specifications of EIA-449 define a combination of two connectors: one with 37 pins (DB-37) and one with 9 pins (DB-9), for a combined 46 pins (see Figure 3.13).

The functional specifications of the EIA-449 give the DB-37 pins properties similar to those of the DB-25. The major functional difference between the 25- and 37-pin connectors is that all functions relating to the secondary channel have been removed from DB-37. Because the secondary channel is seldom used, EIA-449 separates those functions out and puts them in the second, 9-pin connector (DB-9). In this way, a second channel is available to systems that need it.

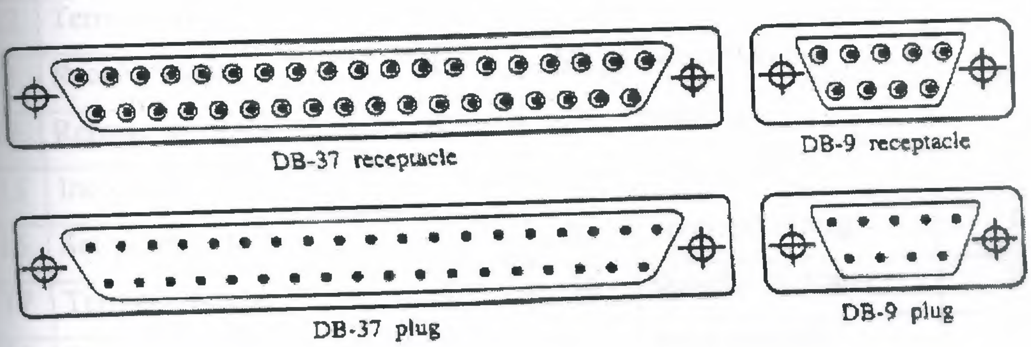


Figure 3.13 DB-37 and DB-9 connectors

3.4.2 Pin Functions

To maintain compatibility with EIA-232. EIA-449 defines two categories of pins to be used in exchanging data, control, and timing information (see Table 3.1).

Table 3.1 DB-37 pins

Pin	Function	Category	Pin	Function	Category
1	Shield		20	Receive Common	II
2	Signal rate indicator		21	Unassigned	I
3	Unassigned		22	Send data	I
4	Send data	I	23	Send timing	I
5	Send timing	I	24	Receive data	I
6	Receive data	I	25	Request to send	I
7	Request to send	I	26	Receive timing	I
8	Receive timing	I	27	Clear to send	I
9	Clear to send	I	28	Terminal in service	II
10	Local loopback	II	29	Data mode	I
11	Data mode	I	30	Terminal ready	I
12	Terminal ready	I	31	Receive ready	I
13	Receive ready	I	32	Select standby	II
14	Remote loopback	II	33	Signal quality	
15	Incoming call		34	New signal	II
16	Select frequency	II	35	Terminal timing	I
17	Terminal timing	I	36	Standby indicator	II
18	Test mode	II	37	Send common	II
19	Signal ground				

Category I Pins

Category I includes those pins whose functions are compatible with those of EIA-232 (although most have been renamed). For each Category I pin, EIA-449 defines two pins, one in the first column and one in the second column. For example, both pins 4 and 22 are called send data. These two pins have the equivalent functionality of pin 2 in EIA-232. Both pins 5 and 23 are called send timing. Arid both pins 6 and 24 are called receive data. Even more interesting, these pairs of pins are vertically adjacent to one another in the connector, with the pin from the second column occupying the position essentially below its counterpart from the first column. (Number the DB-37 connector based on the numbering of the DB-25 connector to see these relationships.) This structure is what gives EIA-449 its power. How the pins relate will become clear later in this section, when we discuss the two alternate methods of signaling defined in the electrical specifications.

Table 3.2 lists the pin functions of the DB-9 connector and shows their relationships to the EIA-232 (DB-25) equivalents.

Table 3.2 DB-9 pins

Pin	Function	EIA -232 Equivalent
1	Shield	1
2	Secondary receive ready	
3	Secondary send data	14
4	Secondary receive data	16
5	Signal ground	7
6	Receive common	12
7	Secondary request to send	19
8	Secondary clear to send	13
9	Send common	

Category II Pins

Category II pins are those that have no equivalent in EIA-232 or have been redefined The numbers and functions of these new pins are as follows:

* **Local loop-back.** Pin. 10 is used for local loop-back testing.
* **Remote loop-back.** Pin 14 is used for remote loop-back testing.
* **Select frequency.** Pin 16 is used to choose between two different frequency rates.

- **Receive common.** Pin 20 provides a common signal return line for unbalanced circuits from the DCE to the DTE.
- **Terminal in service.** Pin 28 indicates to the DCE whether or not the DTE is operational.
- **Select standby.** Pin 32 allows the DTE to request the use of standby equipment in the event of failure.
- **New signal**. Pin 34 is available for multiple point applications where a primary DTE controls several secondary DTEs. When activated, pin 34 indicates that one DTE has finished its data exchange and a new one is about to start.
- **Standby indicator**. Pin 36 provides the confirmation signal from the DCE in response to select standby (pin 32).
- **Send common**. Pin 37 provides a common signal return line for unbalanced circuits from the DTE to the DCE.

3.4.3 Electrical Specifications: RS-423 and RS-422

EIA-449 uses two other standards to define its electrical specifications: RS-423 (for unbalanced circuits) and RS-422 (for balanced circuits).

RS - 423 Unbalanced Mode

RS-423 is an unbalanced circuit specification, meaning that it defines only one line for propagating a signal. All signals in this standard use a common return (or ground) to complete the circuit. Figure 3.14 gives a conceptual view of this type of circuit as well as the specifications for the standard. In unbalanced-circuit mode, EIA-449 calls for the use of only the first pin of each pair of Category I pins and all Category II pins.

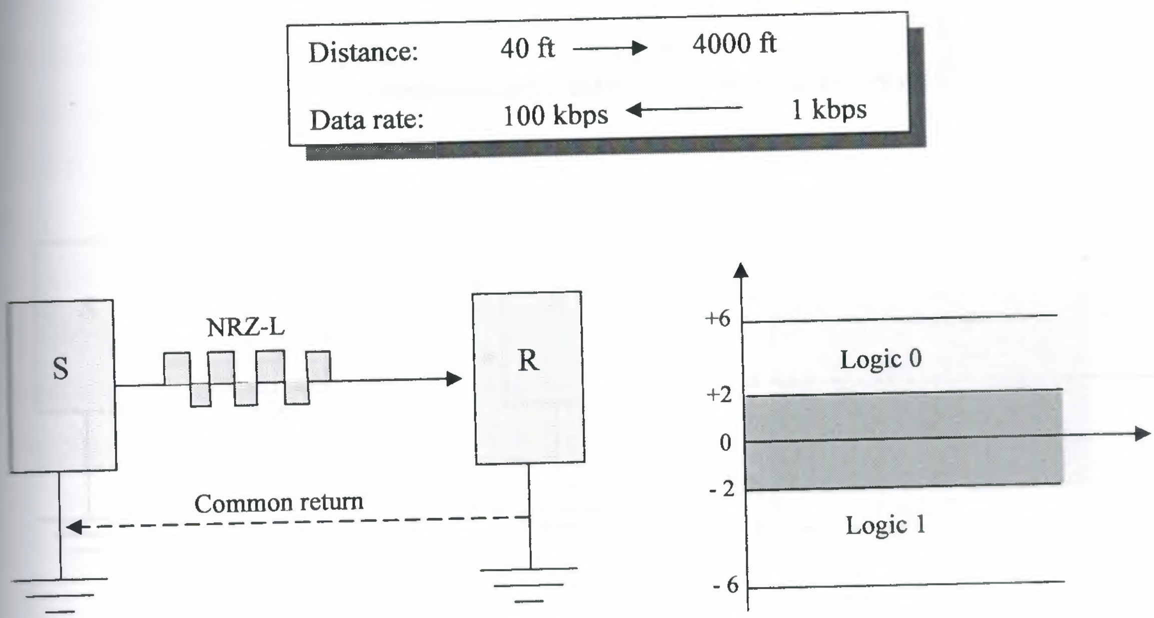


Figure 3.14 RS-423. Unbalanced Mode

RS-422: Balanced Mode

RS-422 is a balanced circuit specification, meaning that it defines two lines for the propagation of each signal. Signals again use a common return (or ground) for the return of the signal. Figure 3.15 gives a conceptual view of and the specifications for this standard. In balanced mode, EIA-449 utilizes both pins in each Category I but does not use the Category II pins. As you can see from the electrical specifications for this standard, the ratio of data rate to distance is much higher than that of the unbalanced standard or of EIA-232: 10Mbps for transmissions of 40 feet.

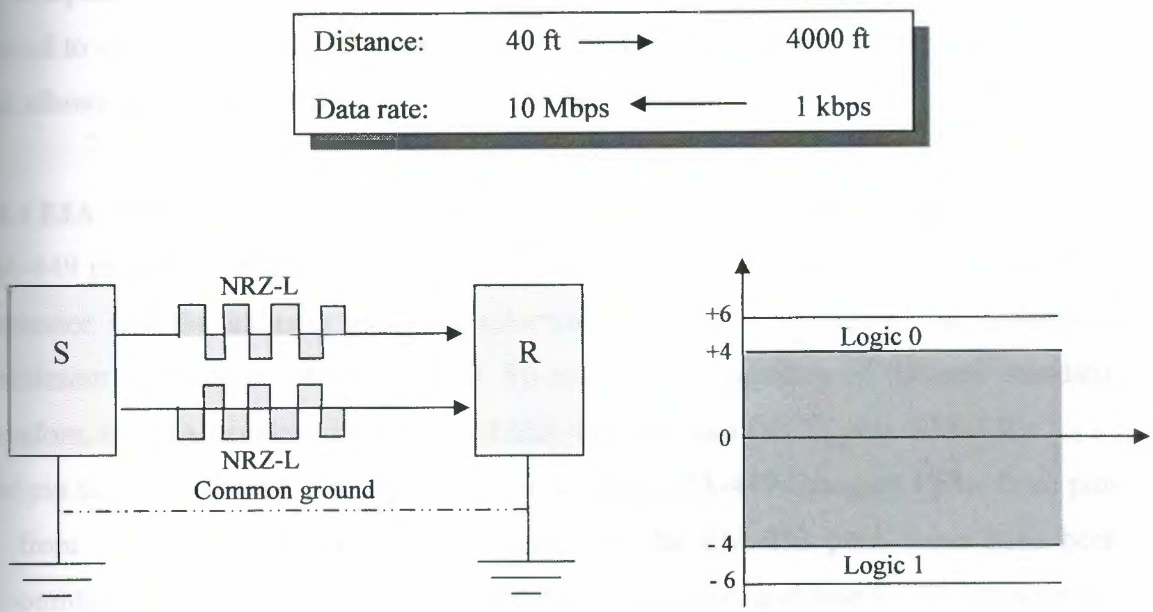


Figure 3.15 RS-422. Balanced mode

In unbalanced mode, two lines carry the same transmission. They do not, however, carry identical signals. The signal on one line is the complement of the signal on the other. When plotted, the complement looks like a mirror image of the original signal (see Figure 3.15). Instead of listening to either actual signal, the receiver detects the differences between the two. This mechanism makes a balanced circuit less susceptible to noise than an unbalanced circuit, and improves performance.

As the complementary signals arrive at the receiver, they are put through a subtracter (a differential amplifier). This mechanism subtracts the second 'signal from the first before interpretation. Because the two signals complement each other, the result of this subtraction is a doubling of the value of the first signal. For example, if at a given moment the first signal has a voltage of 5, the second signal will have a voltage of -5. The result of subtraction, therefore, is *5 - (-5),* which equals 10.

If noise is added to the transmission, it impacts both signals in the same way (positive noise affects both signals positively; negative noise affects both negatively). As a result, the noise is eliminated during the subtraction process (see Figure 3.16). For example, say that two volts of noise are introduced at the point where the first signal is at 5 volts and its complement is at -5 volts. The addition distorts the first signal to 7 volts, and the second to -3 volts. 7 - (-3) still equals 10. It is this ability to neutralize the effects of noise that allows the superior data rates of balanced transmission.

3.4.4 EIA -530

EIA-449 provides much better functionality than EIA-232. However, it requires a DB37 connector that the industry has been reluctant to embrace because of the amount of investment already put into the DB-25. To encourage acceptance of the new standard, therefore, the EIA developed a version of EIA-449 that uses DB-25 pins: EIA-530.

The pin functions of EIA-530 are essentially those of EIA-449 Category I plus three pins from Category II (the loop-back circuits). Of the EIA-232 pins, some have been omitted, including ring indicator, signal quality detector, and data signal rate selector. EIA-530 does not support a secondary circuit.

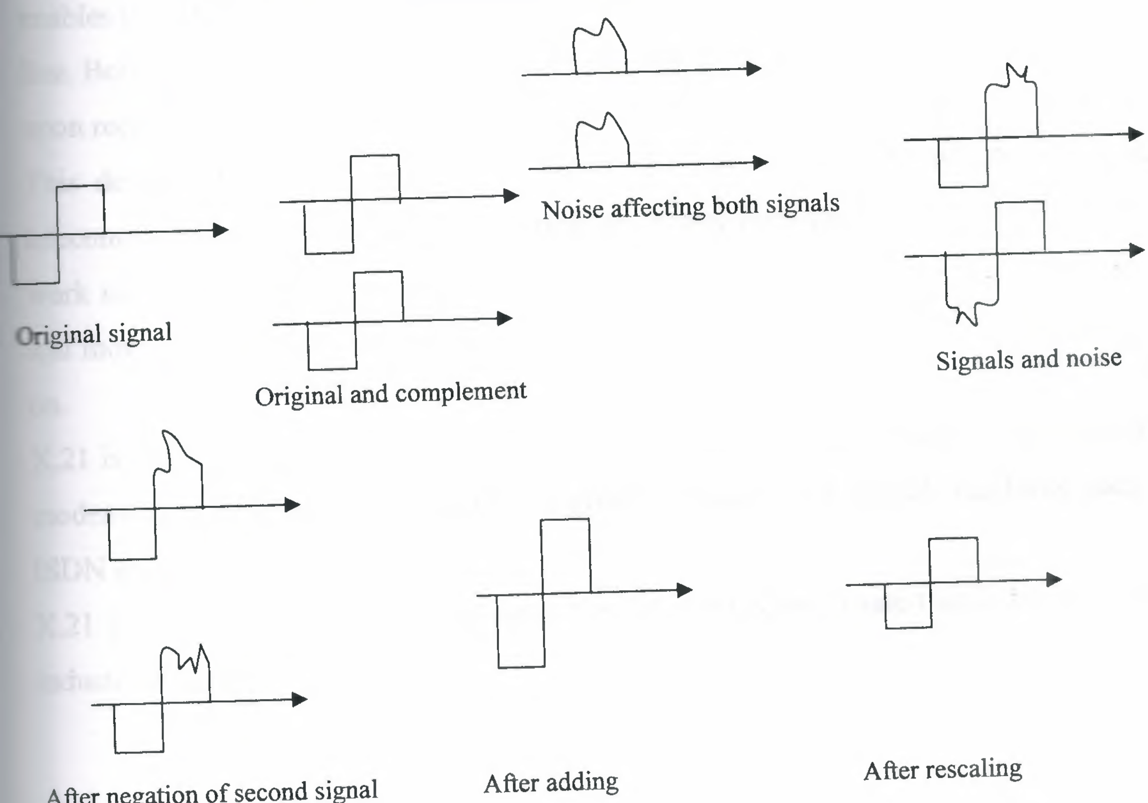


Figure 3.16 Canceling of noise using balanced mode

3.4.5 X.21

Is an interface standard designed by the ITU-T to address many of the problems existing in the EIA interfaces and, at the same time, pave the way for all-digital communication.

Using Data Circuits for Control: A large proportion of the circuits in the EIA interfaces are used for control. These circuits are necessary because the standards implement control functions as separate signals. With a separate line, control information is represented only by positive and negative voltages. But, if control signals are encoded using meaningful control characters from a system such as ASCII, they can be transmitted over data lines.

For this reason, X.21 eliminates most of the control circuits of the EIA standards and instead directs their traffic over the data circuits. To make this consolidation of functionality possible, both the DTE and the DCE must have added circuit logic that enables them to transform the control codes into bit streams that can be Sent over the data line. Both also need additional logic to discriminate between control information and data upon receipt.

This design allows X.21 not only to use fewer pins but also to be used in digital telecommunications where control information is sent from device to device over a net-work rather than just between a DTE and a DCE. As digital technology emerges. more and more control information must be handled, including dialing, redialing. hold, and so on.

X.21 is useful both as an interface to connect digital computers to analog devices such as modems and as a connector between digital computers and digital interfaces such as ISDN and X.25.

X.21 is designed to work with balanced circuits at 64 Kbps, a rate that is becoming the industry standard.

3.4.6 Pin Functions

Figure 3.17 shows the connector specified by X.21. the DB-15. As the name indicates, the DB-15 is a 15-pin connector.

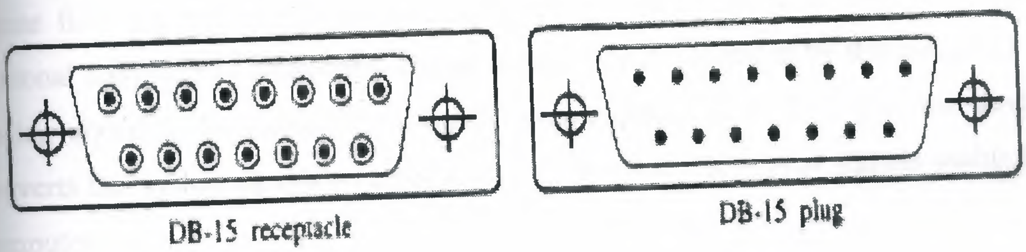


Figure 3.17 DB-15 connector

Byte timing. Another advantage offered by X.21 is that of timing lines to control byte synchronization in addition to the bit synchronization provided by the EIA standards. By adding a byte timing pulse (pins 7 and 14), X.21 improves the overall synchronization of transmissions.

Control and indication. Pins 3 and 5 of the DB-15 connector are used for the initial handshake, or agreement to begin transmitting. Pin 3 is the equivalent of request to send. Pin 5 is the equivalent of clear to send. Table 3.3 lists the functions for each pin.

Table 3.3 DB-15 pins

Pin	Function	Pin	Function
1	Shield	9	Transmit data or control
2	Transmit data or control	10	Control
3	Control	11	Receive data or control
4	Receive data or control	12	Indication
5	Indication	13	Signal element timing
6	Signal element timing	14	Byte timing
7	Byte timing	15	Reserved
8	Signal ground		

3.5 Modems

The most familiar type of DCE is a modem. Anyone who has surfed the Internet, logged on to an office computer from home, or filed a news story from a word processor over a phone line has used a modem. The external or internal modem associated with your personal computer is what converts the digital signal generated by the computer into an analog signal to be carried by a public access phone line. It is also the device that converts the analog signals received over a phone line into digital signals usable by your computer.

The term modem is a composite word that refers to the two functional entities that make up the device: a signal modulator and a signal demodulator. The relationship of the two parts is shown in Figure 3.18c.

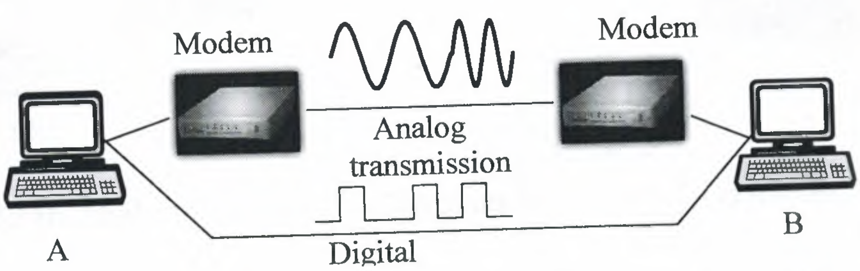


Figure 3.18a Modem connection Digital/analog connection between two computers

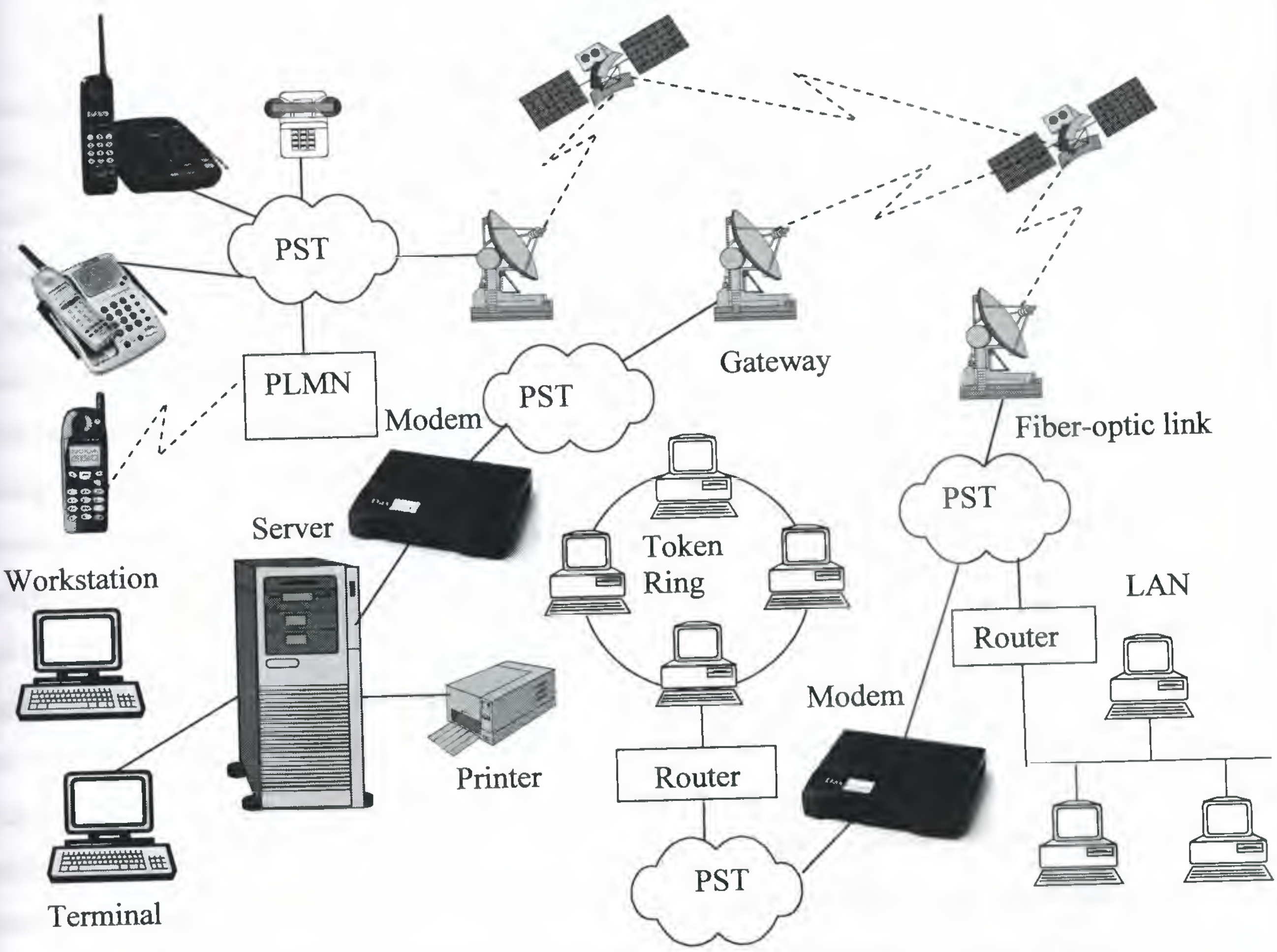


Figure 3.18b Modem Connection in Digital/analog conversion in internet working

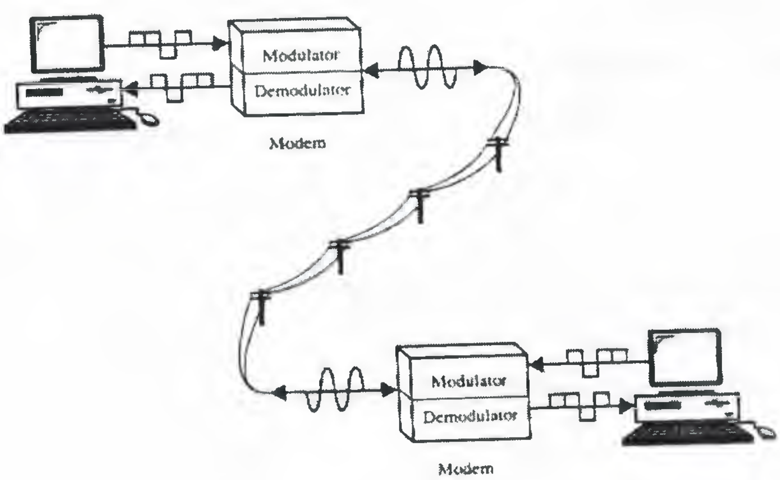


Figure 3.18c Modem concept

A modulator converts a digital signal into an analog signal. A demodulator converts an analog signal into a digital signal. While a demodulator resembles an analog to-digital encoder, it is not in fact an encoder of any kind. It does not sample a signal to create a digital facsimile; it merely reverses the process of modulation.

A modulator converts a digital signal to an analog signal. A demodulator converts an analog signal to a digital signal.

Both modulators and demodulators, however, do use the same techniques as digital-to-analog encoders: modulators to further encode a signal, and demodulators to decode it. A modulator treats a digital signal as a series of 1s and 0s, and so can transform it into a completely analog signal by using the digital-to-analog mechanisms of ASK. FSK. PSK, and QAM.

Figure 3.18 shows the relationship of modems to a communication link. The two PCs at the ends are the DTEs; the modems are the DCEs. The DTE creates a digital signal and relays it to the modem via an interface (like the EIA-232, as discussed before). The modulated signal is received by the demodulation function of the second modem. The demodulator takes the ASK, FSK, P5K, or QAM signal and decodes it into whatever format its computer can accept. It then relays the resulting digital signal to the receiving computer *via* an interface. Each DCE must be compatible with both its own DTE and with other DCEs. A modem must use the same type of encoding (such as NRZ-L), the same voltage levels to mean the same things, and the same timing conventions as its DTE. A modem must also be able to communicate with other modems.

3.5.1 Transmission Rate

You may have heard modems described as high-speed or low-speed to indicate how many bits per second a specific device is capable of transmitting or receiving. But before talking about different commercial modems and their data rates, we need to examine the limitations on the transmission rate of the medium itself.

3.5.2 Bandwidth

Here we can apply the concept of Bandwidth to physical media to see its effect on transmission. The data rate of a link depends on the type of encoding used, the duration of the signal. the size of the voltages used, and the physical properties of the transmission medium. Of these, the last imposes the greatest limitations. One way to increase the speed of data transmission is to increase the speed (frequency) of the signal carrying it. Theoretically, the faster the signal; the faster the data rate. But increasing the speed of a signal means increasing the number of changes per second (baud rate), and every line has an inherent limitation on the number of such changes it can accommodate. In other words, every line, based on its electrical qualities, can accept only a certain range of signal changes per second. If the signal is too slow, it cannot overcome the capacitance of the line. If it is too fast, it can be impeded by the inductance of the line. So we can say that every line has an upper limit and a lower limit on the frequencies of the signals it can carry. This limited range is called the bandwidth.

Every line has an upper limit and a lower limit on the frequencies of the signals it can carry. This limited range is called the bandwidth.

Traditional telephone lines can carry frequencies between 300 Hz and 3300 Hz, giving them a bandwidth of 3000 Hz. All of this range is used for transmitting voice, where a great deal of interference and distortion can be accepted without loss of intelligibility. As we have seen, however, data signals require a higher degree of accuracy to ensure integrity. For safety's sake, therefore, the edges of this range are not used for data communication. In general, we can say that the signal bandwidth must be smaller than the cable bandwidth. The effective bandwidth of a telephone line being used for data transmission is 2400 Hz, covering the range from 600 Hz to 3000 Hz. Note that today some telephone lines are capable of handling more bandwidth than traditional lines. However, modem design is still based on traditional capability (see Figure 3.19).

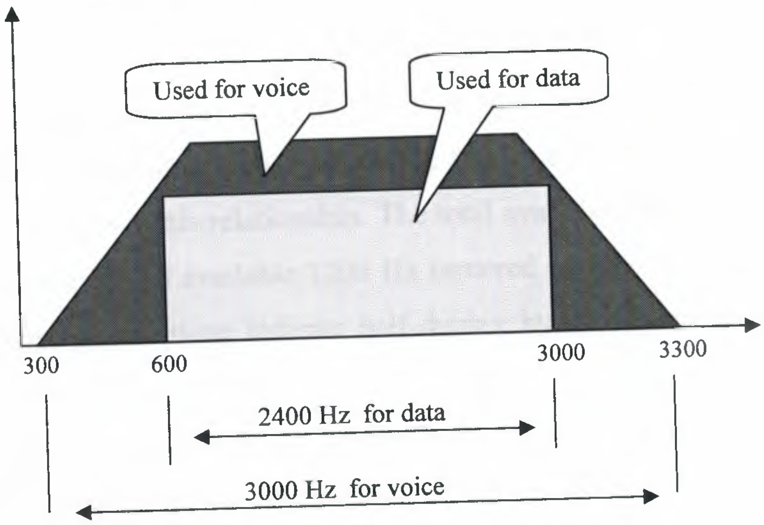


Figure 3.19 Telephone line bandwidth

3.5.3 Modem Speed

As we have seen, each type of analog encoding manipulates the signal in a different way: ASK manipulates amplitude; FSK manipulates frequency; PSK manipulates phase: and QAM manipulates both phase and amplitude.

ASK

The bandwidth required for ASK transmission is equal to the baud rate of the signal. Assuming that the entire link is being used by one signal, as it would be for simplex or half-duplex transmission, the maximum baud rate for ASK encoding is equal to the entire bandwidth of the transmission medium. Because the effective bandwidth of a telephone line is 2400 Hz, the maximum baud rate is also 2400. And, because the baud rate and bit rate are the same in ASK encoding, the maximum bit rate is also 2400 (see Figure 3.20).

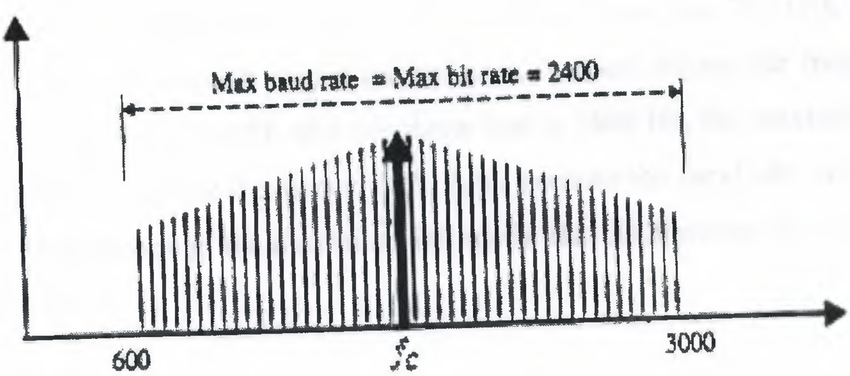


Figure 3.20 Baud rate in half-duplex ASK

For full-duplex transmission, only half of the total bandwidth can be used in either direction. Therefore, the maximum speed for ASK transmission in full-duplex mode is 1200 bps. Figure 3.21 shows this relationship. The total available bandwidth is 2400 Hz; each direction therefore has an available 1200 Hz centered on its own carrier frequency. (Note: some modem specifications indicate half-duplex by the abbreviation HDX, and full-duplex by the abbreviation FDX.).

Although ASK's bit rate equals that of more popular types of encoding, its noise problems make it impractical for use in a modem.

Although ASK has a good bit rate, it is not used today because of noise.

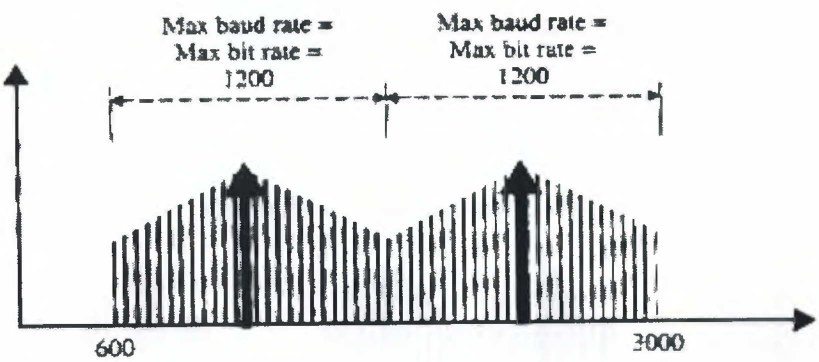


Figure 3.21 Baud rate in full-duplex ASK

FSK

The bandwidth required for FSK transmission is equal to the baud rate of the signal plus the frequency shift. Assuming that the entire link is being used by one signal, as it would for simplex or half-duplex transmission. the maximum baud rate for FSK encoding is equal to the entire bandwidth of the transmission medium minus the frequency shift. Because the effective bandwidth of a telephone line is 2400 Hz, the maximum baud rate is therefore 2400 minus the frequency shift. And, because the baud rate and bit rate are the same in FSK encoding, the maximum bit rate is also 2400 minus the frequency shift (see Figure 3.22).

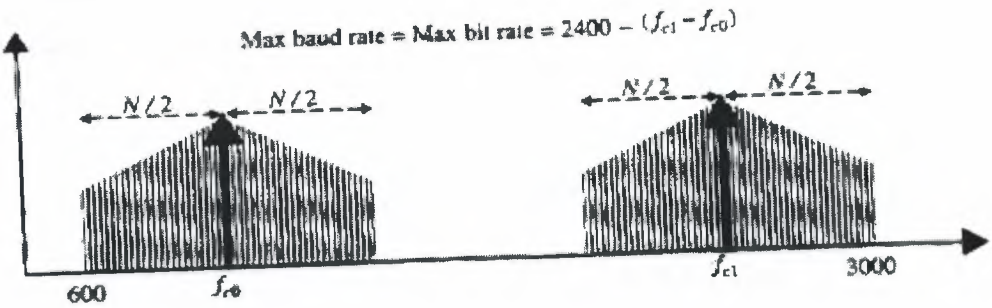


Figure 3.22 Baud rare in half-duplex FSK

For full-duplex transmission, only half of the total bandwidth of the link can be used for either direction. Therefore, the maximum theoretical rate for FSK in full-duplex mode is half of the total bandwidth minus half the frequency shift. Full-duplex FSK partitions are shown in Figure 3.23.

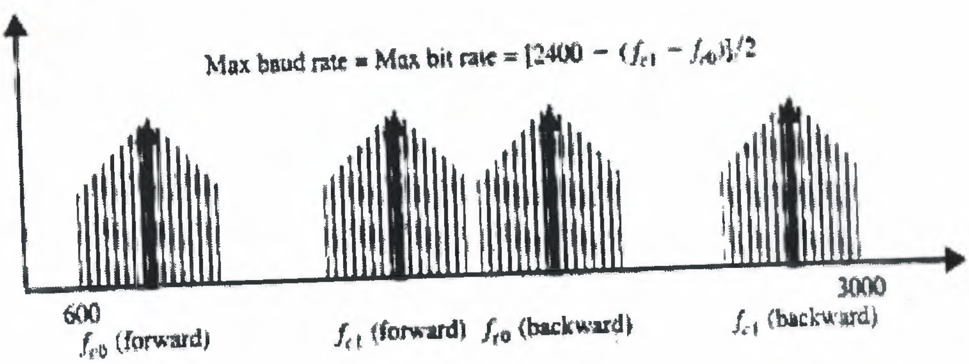


Figure 3.23 Baud rate full-duplex FSK

PSK and QAM

As you recall, the minimum bandwidth required for PSK or QAM transmission is the same as that required for ASK transmission but the bit rate can be greater depending on the number of bits that can be represented by each signal unit.

Comparison

Table 3.4 summarizes the maximum bit rate over standard twisted-wire telephone lines for each of the encoding mechanisms examined above. These figures assume a traditional two-wire line. If four-wire lines are used, the data rates for full-duplex transmission can be doubled. In that case, two wires can be used for sending and two for receiving the data, thereby doubling the available bandwidth. However, these numbers are theoretical and cannot always be achieved with available technology.

Table 3.4 Theoretical bit rates for modems

Encoding	half-Duplex	Full-Duplex
ASK, FSK, 2-PSK	2400	1200
4-PSK, 4-QAM	4800	2400
8-PSK, 8-QAM	7200	3600
16-QAM	9600	4800
32-QAM	12,000	6000
64-QAM	14,400	7200
128-QAM	16,800	8400
256-QAM	19,200	9600

4. HIGH SPEED WIRELESS LAN FOR MOBILE COMPUTING - ARCHITECTURE AND PROTOTYPE MODEM IMPLEMENTATION

4.4 Introduction

Current wireless LANs that are small enough for portable computing devices have transmission rates up to a few Mbit/s, at the lower end of that obtained in IEEE 802-compliant wired LANs. These LANs can provide a useful service when the application demands and number of users are kept low. Much higher performance, from several 10's of Mbit/s to over 100 Mbit/s, is needed to accommodate more users and multimedia traffic.

Achieving wireless access at these high rates is difficult. Wireless LAN systems face technical problems similar to those encountered in outdoor wide-area radio-based systems, including the available bandwidth and fading noise due to multipath and blockage. The goal is to transmit at maximum information rate with acceptable probability of error and minimum equipment complexity, power and cost. Competing approaches to achieve this goal use either infra-red radiation or radio waves in the microwave or millimetre-wave bands.

The success of radio-based indoor wireless LAN systems lies in access to the radio spectrum. This is fiercely competitive and alarmingly scarce in the microwave bands.

For this reason, research on higher speed wireless LANs has focussed on infra-red and millimetre-wave carriers [Fernandes et al., 1994]. With virtually unlimited and unregulated spectrum, many researchers have proposed and developed infra-red carrier techniques for wireless indoor access to local area networks.

To date the most practical infrared techniques use diffuse infrared links and have demonstrated raw link transmission rates up to 50 Mbit/s in cells a few metres in radius [Marsh & Kahn, 1995]. Nevertheless, we contend that the prospect of economically achieving reliable infrared transmission rates much above 10 Mbit/s for truly portable terminals in the next five years is poor because of the high power requirements and delay-spread of the received diffusely-reflected (multi-path) signals. Remarkable breakthroughs are needed either in the design of equalisers to accommodate the delay spread or in the development of adaptive directional links that limit the delay spread while maintaining acceptable signal-to-noise ratio.

This chapter proposes a radio solution at millimetre wavelength frequencies where there is sufficient spectrum to accommodate link speeds of hundreds-of-megabits-per-second. Using a test-bed with a burst-mode transmission capability and experimental 40 GHz radio, we have demonstrated a pico-cellular approach with a range of approximately 10 metres and link rates up to 185 Mbit/s. Additionally, we have built a prototype modem with a raw link rate of 54 Mbit/s for use in a demonstrator high-speed indoor WLAN. This chapter presents the architecture of the proposed WLAN and the design and performance of the prototype modem.

4.2 Spectrum, Propagation and Key Design Decisions

The spectral band from 54 GHz to 65 GHz is strongly absorbed in the atmosphere and is unsuitable for long distance communications. It is suitable for short distance operation on the scale of a LAN. Our propagation studies [Bird et al, 1994] and others [eg ITU-R SG3, 1996; Smulders & Wagemans, 1992; Davies et al, 1991] have shown that propagation in this band is quasi-optical with low penetration of walls and partial penetration of partitions. The signals also reflect well off many surfaces providing coverage of areas which do not have a direct line of sight to the transmitter. The unfortunate side effect of this is that signal reception occurs over multiple paths resulting in destructive interference and significant intersymbol interference.

Use of this spectral band requires a diversity scheme to overcome the destructive interference and a robust modulation and error correction coding scheme to overcome the effects of intersymbol interference. The latter becomes more difficult as the bit rate increases. Given a typical indoor environment and a user data rate above 20 Mbit/s it is difficult to achieve a data rate bandwidth efficiency of 1 bit/s/Hz1. An example of this is the HIPERLAN (High Performance Radio LAN) system [ETSI, 1995] which achieves a data rate bandwidth efficiency of significantly less than 1 bit/s/Hz. HIPERLAN uses GMSK modulation with an equaliser and BCH (31,26) FEC coding. To achieve an efficiency of 2 bit/s/Hz would require a very complex receiver.

Our WLAN uses a coded multicarrier modulation scheme to overcome multipath interference [Skellern & Percival, 1994; Lee et al., 1995; Skellern et al., 1995]. This approach, which has been widely used in other transmission applications [eg Alard & Halbert, 1987; Bingham, 1990; Chow et al., 1991; Sari et al., 1995], is less complex to implement than solutions that use a single carrier plus equaliser. The reduced complexity provides the potential for lower power consumption than solutions using a single carrier. The HIPERLAN choice of a single carrier plus equalisation is an interesting one, apparently based on a concern about distortion products in the output power amplifier arising from the non-unity peak-to-average power ratio of the multicarrier signal. However, the probability of large peak excursions in the multicarrier signal is low. Moreover, clipping of the signal does not substantially degrade the error performance.

A well-designed coded multicarrier modulation scheme, involving FFT processing, coding and interleaving, provides good protection against multipath effects. The transform processing converts a time-invariant intersymbol interference (ISI) channel to one that behaves effectively with no memory. Thus, no equaliser is needed to combat ISI in a practical system.

Figure 4.1 shows a measured 8-tone multi-carrier modulation signal suffering a frequency selective fade due to multipath interference on a 40 GHz link. The coding scheme used in our WLAN allows recovery of information lost in the faded tones.

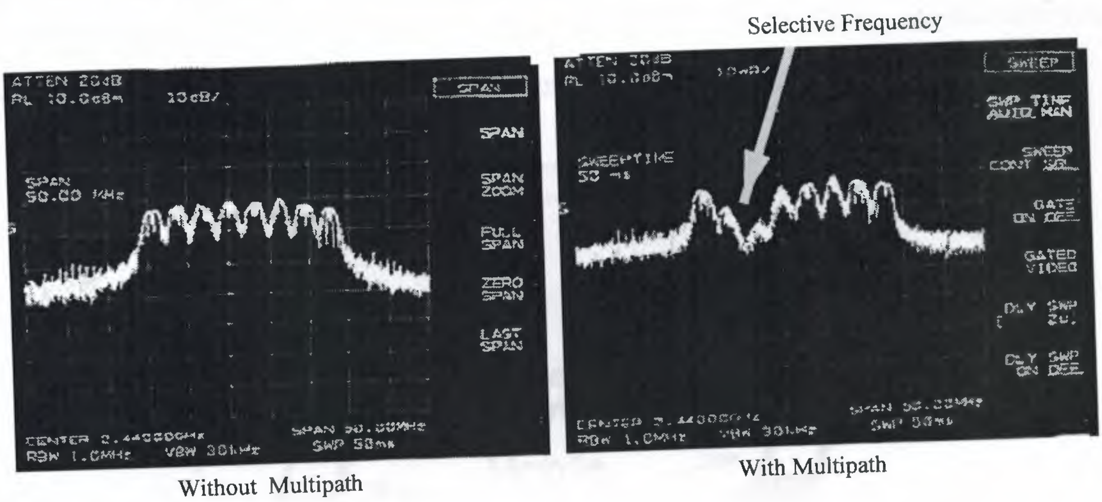


Figure 4.1 Measured multicarrier spectra at 40 GHz

Subtractive interference can cause broadband fading, which can result in an inability to obtain synchronisation or high error rates (around 10-1). This is very hard to correct with any FEC scheme and some form of diversity is required. Antenna diversity in the form of spatial diversity or beam steering is preferred over frequency hopping at the high data rates of interest in this work. Spatial diversity can be achieved by placing the antennas a significant fraction of a wavelength apart, which is quite practical at millimeter wavelengths. Beam steering is also practical at these wavelengths [Bird et al , 1994].

The USA Federal Communications Commission has notified the allocation of the 59-64 GHz band for unlicensed short range transmission [FCC, 1995]. This band is ideally suited for high speed wireless LANs. Other countries are also actively opening the bands around 60 GHz for unlicensed use [e.g. Takimoto, 1995; Meinel, 1995a].

4.3 WLAN System Architecture

The implicit low range of operation at 60 GHz is advantageous for interference minimisation and hence distributed bandwidth use. It also suggests an architecture with numerous small hub units providing access to a cabled infrastructure.

The basic entity of the WLAN is a radio cell, illustrated in Figure 4.2, containing a hub station and mobile stations. A radio cell is a volume, up to about 20m in diameter for millimetre wavelength carriers, within which it is possible to establish reliable two-way communication between the hub station and mobile stations.

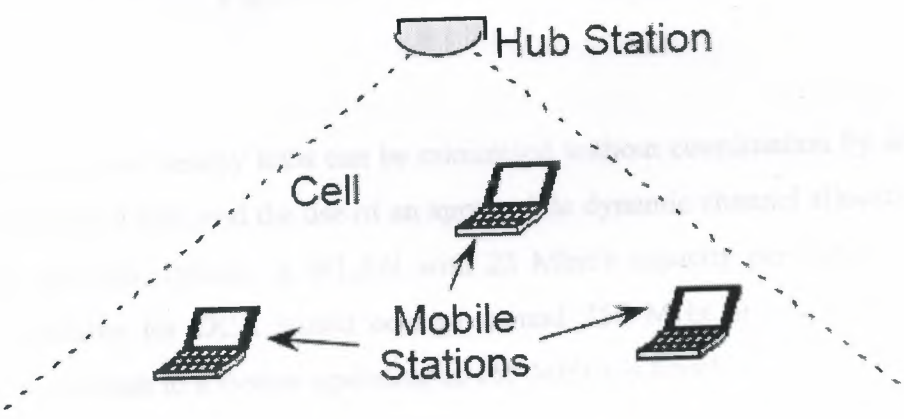


Figure 4.2 A radio Cell

A WLAN system, illustrated in Figure 4.3, consists of either a single radio cell or multiple interconnected radio cells. Interconnection is via attachment of any one or more of the hub and mobile stations in a cell to another network, which may be either wired or wireless.

The hub station has the responsibility of providing connectivity between all mobile stations and is placed to overcome the so-called 'hidden terminal problem' - not all mobile stations may be able to directly contact each other because of blockages. The hub station acts as a relay for all communications within the cell. It also fairly manages access to the wireless medium for the mobile stations within the radio cell and from the backbone.

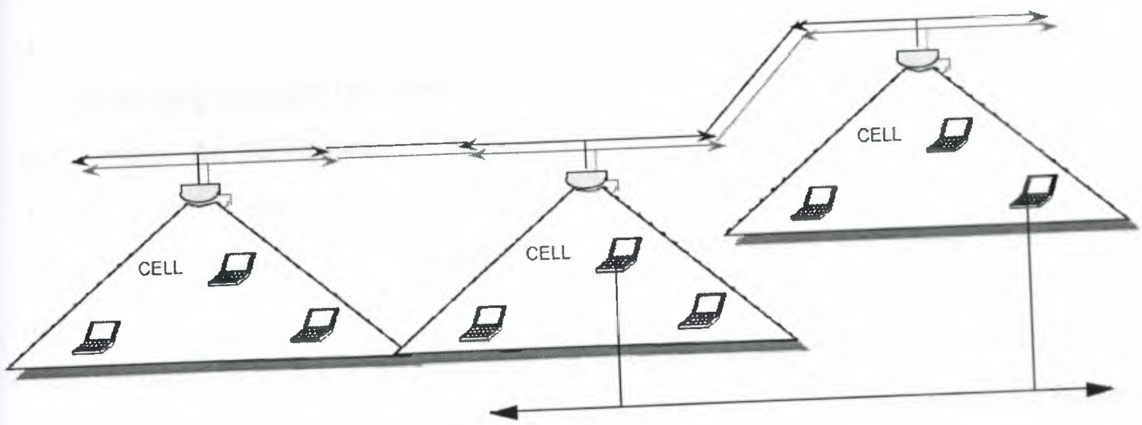


Figure 4.3 A Wireless LAN System

Interference between nearby hubs can be minimised without coordination by adoption of a spectrum channel plan and the use of an appropriate dynamic channel allocation (DCA) scheme by the hub stations. A WLAN with 25 Mbit/s capacity per hub and with ten channels available for DCA would occupy around 250 MHz or 5% of the available spectrum. Expansion to a system operating at 155 Mbit/s is feasible.

4.4 Station Implementation

A WLAN station has four principal components:

• a Physical layer (PHY) consisting of:

- a Transmission sub-layer - the 'RF section'

- a Modem sub-layer

• a Medium Access Control (MAC) sub-layer

• Station Management (SMT)

The modem sub-layer, MAC sub-layer and SMT together will be referred to as the WLAN backend. The differences between the mobile station and the hub station are their antenna systems, the MAC and SMT.

4.4.1 RF-Section

While the target carrier frequency is in the 59-64 GHz band, the prototype link uses a 40 GHz radio. An RF-section for a 61 GHz WLAN, similar to the 40 GHz transceiver frontend used in the prototype. The transceiver operates in halfduplex mode. The front-end consists of an antenna, a transmit/receive switch, a low noise amplifier, a medium power amplifier and a mixer.

For isotropic radiating antennas, there is a large variation in power received from a mobile station at a ceiling-mounted hub antenna as the mobile moves from the centre of a cell towards the edge. The WLAN hub antenna uses a shaped beam to partially compensate for this power variation. The shaping increases the gain of the antenna for shallow look angles. To completely eliminate this angle-related power variation, the combined pattern of the hub antenna and mobile antenna should be a cosec squared function of angle. However, in our system about half the potential variation is removed by the beam shaping and the remainder is handled by the AGC (automatic gain control) system. The mobile stations use a planar antenna [Bird et al, 1994].

The first mixer is a bi-directional image reject mixer that provides both up and down conversion. The first local oscillator is at 59 GHz. The MPA output power is 10 mW. The first transmit/receive switch, LNA, MPA and mixer are HEMT MMICs [Archer, 1992; Batchelor et al., 1993]. The intermediate frequency (IF) stage includes transmit/receive switches and a bandpass filter (not shown) to provide out-of-band rejection. The IF stage uses I,Q up and down conversion with a quadrature LO.

4.4.2 WLAN Station Backend

A schematic diagram of a WLAN station backend is shown in Figure 4.4.

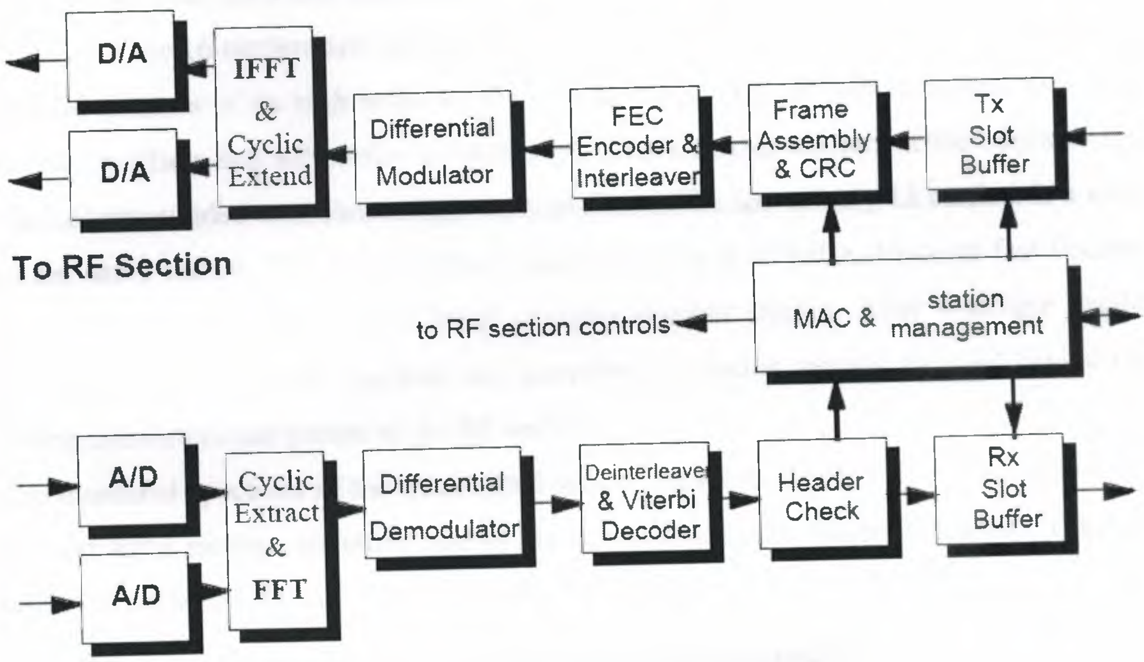


Figure 4.4 Station Backend

The backend accepts protocol data units (PDU) of length 218 bytes across the MAClayer interface connecting the WLAN system to its host computer. This PDU size conveniently holds four ATM cells plus link protocol header and is chosen recognizing the significant overhead of around 100 bits needed for synchronisation and frequency tracking, as well as some 10 msec for switching between transmit and receive. The latter involves switching of low noise, medium power amplifiers and other RF circuitry.

The data efficiency of this choice is approximately 85%. The PDU size is also short enough that Doppler effects for stations moving up to several m/s are negligible.

The MAC adds its header to the PDU, and passes the resulting slot across a UTOPIA like interface [ATM Forum, 1994] to the PHY layer where a cyclic redundancy check word is calculated and added to the end of the slot.

The data stream is encoded by a rate 1/2, memory 6, trellis coder whose output is interleaved and modulated as a block-differentially encoded QPSK signal. The DQPSK symbols are assembled into frames suitable for generating multi-carrier modulation.

Only 12 of the 16 carriers are modulated in the prototype. The carrier at zero frequency is avoided because of its high noise level resulting from receiver imperfections, including converter offsets and RF carrier breakthrough. Three further carriers at the edges of the channel are avoided to accommodate channel filters. The use of only 12 carriers is a very conservative choice. The frame is then transformed by a 45 MHz, 16-point fast Fourier transform circuit, resulting in a serial complex number stream. After 4-sample cyclic extension, these complex numbers are converted to analog signals by 8-bit digital-to-analog converters and passed to the RF section.

The measured spectrum of the transmitted signal, shown in Figure 4.5, clearly shows the 12 modulated carriers, six either side of the unmodulated zero-frequency carrier. The link bit rate is 54 Mbit/s (45 MHz * 2 bit/symbol * 12/20).

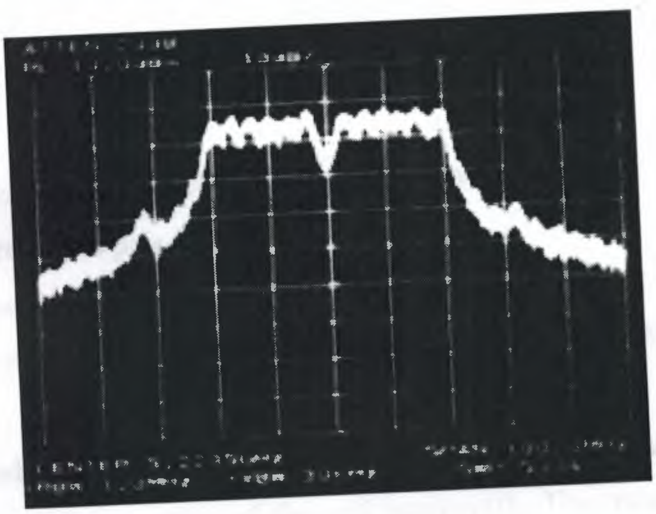


Figure 4.5 Measured spectrum of the WLAN Tx signal showing 12 modulated carriers.

On the receive side, the input is a quadrature pair of baseband signals that are digitized by 8-bit analog-to-digital converters. After cyclic extraction to complement the transmit cyclic extension process, a fast Fourier transform circuit separates the multi-carrier signals. Synchronisation circuits (not shown) obtain frequency, symbol and frame sync. Information carrying tones are fed to a soft-decision differential demodulator before deinterleaving and soft decision Viterbi decoding. The decoded bits are compiled into a slot, checked for errors and processed by the MAC and SMT entities.

The modem is implemented largely using Xilinx programmable logic parts with a custom FFT design, similar to that described by Ryan et al [1995], and a commercial Viterbi decoder. The maximum clock speed of the implementation is limited to 45 MHz by both the FFT and Viterbi decoder. Much higher speed operation can be expected of a custom chip implementation of the back-end, the design of which is entirely feasible.

The MAC builds on the half-duplex operation of the physical layer. The basic operation of the MAC protocol is as follows:

- Cell bandwidth is quantised into slots of fixed length
- Each mobile station 'owns' some fraction (one or more slots) of the total radio cell bandwidth, which it obtains upon registering with the hub station, and is guaranteed access to that bandwidth at all times if needed.
- Mobile stations may access any unused slots, which may or may not be the slots they own; the hub station has the responsibility of marking the slots to indicate their potential use.
- MAC traffic for a mobile station always appears as an exchange - a transmit-slot-tohub/ receive-slot-from-hub pair.
- Mobile stations must base their transmissions on hub timing A heavily loaded cell operates essentially in TDM mode with each mobile using its owned slots.

In a lightly loaded cell, the MAC protocol provides a service that allows any station to access a substantial proportion of the cell bandwidth. The transition between this and TDM mode is a smooth and seamless one.

SMT handles a variety of standard management function as well as the registration process and power-down modes.

4.4.3 Link Performance

The simulated BER performance of the link is shown in Figure 4.6 for two channels. The steep waterfall curve is for an AWGN channel. The second BER curve is for a representative hostile indoor channel consisting of a line-of-sight signal plus four multipath signals. The rms delay spread of the channel is 7 nsec and the maximum delay is 26 nsec.

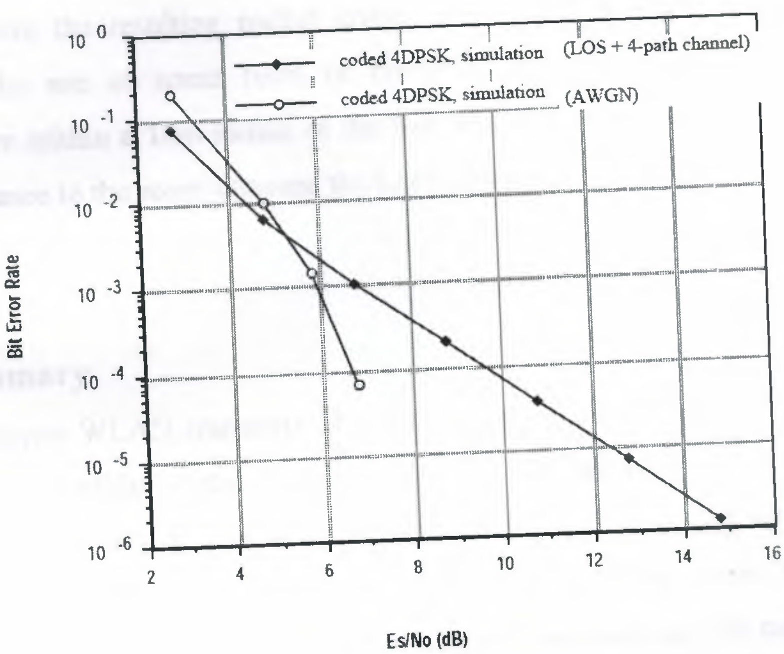


Figure 4.6 BER curves for AWGN channel and for channel with line-of-sight plus four Multipaths.

The range of channel characteristics encountered in a practical indoor environment will be great and the best indicator of overall performance is the ability to achieve an acceptable BER throughout a radio cell. A set of measurements taken for each of four beam directions at sites in an open-plan office environment (Room E6A256) at Macquarie University for a link rate of 50 Mbit/s. The room has eye-height partitions defining some 32 work corners (open cubicles), a lab measurement area and meeting area. The room has numerous metal filing cabinets, wooden bookshelves and desks, with overhead metal light fittings and several 0.5m square structural support columns. The laboratory area and left side wall contain many metal racks, cupboards and shelves. The bottom of Figure 4.8 is an exterior wall with 1.7m high windows approximately 1m from the floor occupying approximately 75% of the wall. The hub was placed at height 2.7m. Shaded clover patterns indicate beam directions for which the BER < 10-5 Four beams are shown at each measurement site. Dark shaded beams show directions in which the measured BER was less than 10-5, an error level which is considered satisfactory for computer network operation since at this level higher-layer protocols are readily able to recover from the resulting packet errors. Note that coverage even close to the hub requires the use of some form of diversity. Satisfactory performance is obtained everywhere within a 10m radius of the hub with the exception of a small region at the main entrance to the room between the Laboratory Space and Offices (at centre top of the figure).

4.5 Summary

The prototype WLAN transmits at a link speed of 54 Mbit/s to achieve reliable data transfer at 27 Mbit/s (after decoding) through the use of a robust coded multicarrier modulation scheme. Antenna diversity is needed to obtain good coverage in a typical indoor office environment. We believe that this system is the fastest WLAN yet reported and that it demonstrates the feasibility of indoor multimedia mobile communications. However, the cost of mm-wave radio systems is currently too high for consumer use. Estimates show that cost effective solutions will exist in a 5 year time frame [Meinel, 1995b; Skellern & Percival, 1994]. For shorter term solutions, lower frequencies must be considered. The prototype modem described here is also suited to lower radio frequency carriers but spectrum availability to date has concentrated our efforts on mmwave carriers. Recent regulatory changes may open the way for lower frequency multimedia WLANs.

CONCLUSION

As this brief introduction to mobile networking has shown, Mobile IP has great potential. Security needs are getting active attention and will benefit from the deployment efforts underway. Within the IETF, Mobile IP is likely to move from a proposed standard to a draft standard in the near future.

The IETF standardization process requires the working group to rigorously demonstrate interoperability among various independent implementations before the protocol can advance. FTP Software has hosted two interoperability testing sessions, and many vendors have taken advantage of the opportunity. Test results have given added confidence that the Mobile IP specification is sound, implementable, and of diverse interest throughout the Internet community. Only a few minor revisions have been needed to ensure the specification can be interpreted in only one way by the network protocol engineers and programmers who must implement it.

It is possible that the deployment pace of Mobile IP will track that of IPv6, or that the requirements for supporting mobility in IPv6 nodes will give additional impetus to the deployment of both IPv6 and mobile networking. The increased user convenience and the reduced need for application awareness of mobility can be a major driving force for adoption. Since both IPv6 and Mobile IP have little direct effect on the operating systems of mobile computers outside of the network layer of the protocol stack, application designers should find this to be an acceptable programming environment. Of course, everything depends heavily on the willingness of platform and router vendors to implement Mobile IP and/or IPv6, but indications are strong that most major vendors already have implementations either finished or underway.

REFERENCES

[1] Vineet Sachdev, System Engineer

[2] Trueposition Inc. 1111 West DeKalb Pike Wayne, PA 19087

[3] Asha Mehrotra, "GSM System Engineering"

[4] MOBILE COMMUNICATIONS SERIES, Artech House Publishers.

[5] Brian McIntosh "Telecommunications" http://telecomindustry.about.com/business/telecomindustry/library/weekly/aa1115999.ht

[6] Dick Tracy "The Applications We Promote...." http://www.comm-nav.com/commnav.htr

[7] GRAYSON WIRELESS, a division of Allen Telecom. "Geometrix Wireless Location Sensor" http://java.grayson.com/geodatasheet.htrr

[8] Louis A. Stilp "Examining the Coming Revolution in Location Services" http://www.trueposition.com

[9] Paul J, Bouchard "AccuCom Wireless Service Inc." http://www.Global-Images.com

[10] Tutorial "How GPS works?" http://www.trimble.com