UNIT 1 ELEMENTS OF COMMUNICATION

Structure

1.1 INTRODUCTION

We always communicate with others, may be by directly talking, if the person is nearby or by using modern technological tools, if the person is stationed far away. Communicating with each other is a characteristic of most of the living creatures in the world.

The word *communication* implies that there is some *information*, which one entity wants to pass on to the other. In the animal kingdom, the communication is used mainly for two purposes:

The first is for *defence* or *maintenance*, following the law of survival of the fittest. This information transfer can be done in different ways, e.g. by mechanical means like biting of ants or dogs, or via audio signals like chirping of sparrows or the cry of a child etc. The second purpose is to *communicate* between members of same society, for joy and maintenance of the society, e.g. mechanical signals in the form of dancing of bees to indicate direction of flowers or sound signals as used by all the birds and mammals or chemical signals used by the ants for indicating track of the food.

You must have noticed that these communications are based on mainly four domains in Physics, viz. mechanical, chemical, optical and electrical. The other domains like nuclear and magnetic are not used significantly.

Human beings possess two very strong means of communication. Our mouth (with tongue and throat) has a special structure to produce about 10 vowels and 35 to 40 consonants very distinctly; and thus we, particularly Indians, have come out with phonetic languages. The sounds produced by mouth are heard (sensed) by our ears.

We also use light signals, as we have a very good optical sensor in the form of eyes, but for this, we need some external source of light like the sun, lamps etc. Though we make use of purely mechanical signals for very small distances like tapping on shoulder (direct contact), more often we use opto-mechanical signals like nodding or hand-gestures for calling or by showing a picture. Sometimes direct optical signals are used; e.g. light signals for traffic control. Thus, in general, we use sound (audio frequencies: 20 Hz to 20 kHz) and light (visible: red to violet) as the main modes for non-contact communication.

In audio communication, the information in the form of sound signal reaches our ears as a series of sounds one after the other. This is a *serial* communication. Our eyes can process a picture or view, which is in 2 or 3 dimensions. That is, it accepts the information in a *parallel* form. These are the two major modes of information transport in communication.

Sometimes we use touch and smell as a way of communication. However, in this course, we will be concentrating on the electronic communication; and so far audio and visual are the basic communications carried out using electronic means.

In this course, we mainly concentrate on the electronic modes of communications. This includes telephone, radio, television, cellular mobiles and communication via computers. Though these are different types of communication tools, the basic building blocks of any communication system are broadly the same. There is always a source of signal. This signal is first converted into electrical form and then processed through electronic circuit, in order to make it suitable for transmission. This processed signal is then transmitted through a medium of transmission like free space, metal wire, optical fibre etc. Once the signal reaches the destination, the receiver has a circuit that converts the received signal back into the form which is understood by human perception.

The audio and visual communications use serial and parallel modes respectively. You will learn about these modes in Sec. 1.2. The physics involved in the treatment of communication signals is described in Sec. 1.3. The electronic systems used in transmission, receiving and modulation of signals are discussed in Sec. 1.4. The communication takes place through various media like metal wire, optical fibre, free space, dielectric medium etc. You will learn about the characteristics of transmission media in Sec. 1.5.

Objectives

After studying this unit, you should be able to:

- describe the various forms of communications in nature;
- explain the basics of human communication;
- distinguish and compare audio (serial) and visual (parallel) signals;
- draw the block diagram of a typical communication system and describe the significance of each block;
- explain the modern communication terminology; and
- list the various media used in communication.

1.2 SERIAL VERSUS PARALLEL SIGNALS AND SCANNING

You know that we mainly use two modes of communication, viz. audio and visual. There is one basic difference between these two signals. The sound signal is emitted by mouth serially, i.e. one tone followed by other tone and received serially by ears. Our two ears receive it simultaneously and give us a *stereophonic* effect, i.e. the sense of distance and direction of the source. This happens due to slight phase difference in the signals received by two ears caused by the time difference between their receipts. As against this, the visual signal like a picture has many points, which give out light simultaneously. In electronics terminology, these points are called picture cells or **pixels**. The intensity variation at these points (for any colour) is the signal. The optical signal emitted by many points of the picture or scene is received by the lens in our eye *simultaneously* or *parallely*. The visual signal seen by our eye at any instant may be considered to be a two-dimensional (2-D) source.

In the case of electronic communication, radio and television are the most popular communication media. As you will understand in the further units of this course, the radio wave transmitting antenna (AE) is a point source, and sends the signal (radio waves) serially, which are received by the receiving antenna serially. This signal is further sent along the connecting wires and the following circuit serially. This is also true for the telephony lines (transmission lines) and the optical fibres.

For television also, the signals to be transmitted by an antenna have to be in a serial form. Hence, for TV, we need to convert the picture signal, which is parallel in nature, into a serial form and then it can go for transmission via a transmitter antenna. This is done by a *scanning* or *rastering* process as shown in Fig. 1.1. A picture frame to be transmitted is segmented in different horizontal lines. Then each line is scanned pixel by pixel from one end to the other. Then the scanner is moved to the next line as shown by the retrace line in the figure. This way the picture is converted to a serial form which contains information about light intensity at every pixel. This information is transmitted serially. At the receiver end, in our TV sets, this signal is received and shown on the screen using the same format that was used to initially scan the picture. Hence, we see the exact replica of the original picture.

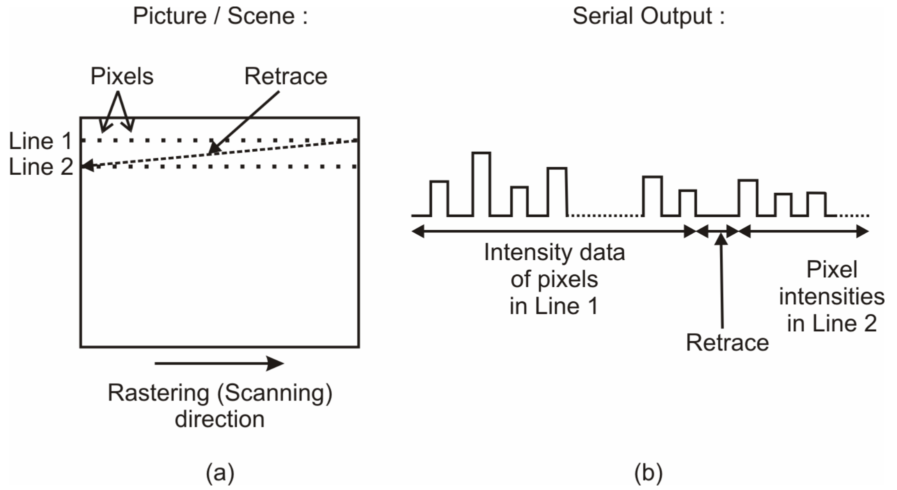


Fig. 1.1: Rastering process of parallel to serial signal conversion

When you read this page you are scanning it from left to right and up to down. In TV, similar but sufficiently fast scanning is used so that our eye sees the picture as a whole and not as scanned one. This is possible due to persistence of vision. Note that the response time of our eyes is in milliseconds ($1/16^{th}$ of a second) and scanning rate is kept much faster than this.

The scanning process is also used in facsimile (Fax) transmission for pictures and documents, but this scanning process is slower than in TV.

SAQ 1

Spend 2 Min.

What difficulties would arise if we wish to send pictures in parallel form over a long distance?

When you read a page of any book, you can read it easily, since it is narrow enough to fit in your field of vision. But if you are seeing a picture, which is much broader than your field of vision, you look at it by moving your neck. This type of scanning is used in case of the ground survey done by the satellites. Here, the pictures of earth surface are taken one after the other as the satellite moves over the earth. Then the entire scene is reconstructed by joining the sequential pictures.

After understanding the basic requirements of communication, we will now discuss the concepts in Physics that govern the communication technology in the modern era.

1.3 BLOCK DIAGRAM OF COMMUNICATION SYSTEMS

For the scientific understanding of any system, it is essential to model the system into its basic elements. For example, when we analyse the electrical system, we first represent it into basic elements *R*, *C*, *L*, sources (voltage/current), transformer, ammeters, voltmeters and switch. What, then, are the blocks of a communication system? In Fig. 1.2 we show the basic building blocks of a typical communication system.

You are using the communication in everyday life and you can convince yourself that, if seen closely, any communication system consists of (a) a signal source (like picture or mouth), a sensor transducer (like camera or microphone) and a transmitter; (b) an intervening medium; and (c) signal receiver (sensor or sink) and an actuator transducer (like picture tube or loud speaker).

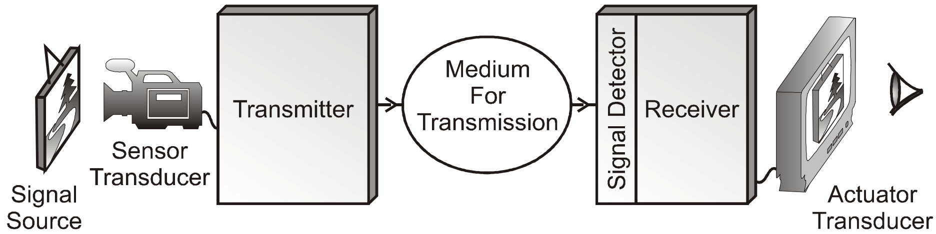


Fig. 1.2: Basic blocks of a typical communication system

Now we shall discuss these blocks of the communication system in details.

1.3.1 Signal Sources

To give a model, i.e. an approximately valid theoretical equation which can be experimentally verified, is the basis of Physics. Law of gravity, with point mass approximation is the best example. The equivalent circuit (black box) is other example in electricity or electronics.

The signal sources commonly used are either audio or visual sources. Audio sources are mainly mouth, drum, clapping, musical instruments etc. The visual sources are the emitted or reflected light, e.g. traffic signal or 7 segment display, picture, photograph, written script, television screen etc.

These sources can be parameterised by a mathematical model signal, which is the simple harmonic motion, extensively used in physics for sound and optics. It can be expressed as:

$$a = \hat{a}\sin(\omega t + \varphi) = \hat{a}\ \sin\theta \tag{1.1}$$

Here *a* is the *instantaneous* signal amplitude, which is a function of following parameters:

(i) **Maximum amplitude ($\hat{a}$)** of the signal, in proper units. For sound it is a maximum amplitude of displacement of vibrating air molecules. In case of optics it is the electric field (**E**) which is related to voltage (*V*). The square of amplitude gives the energy or energy density or intensity or power (V^2 or $\mathbf{E}\times\mathbf{H}$). For sound, it is customary to measure the power in decibels (dB). For a given system, the range of amplitudes called *dynamic range* will be fixed.
The actual amplitude *a*, of course, varies at every instant and its variation with time is the *real information*. If this amplitude were to be constant (like a blue sky or a constant whistle before the radio programme starts) then it is a smallest *bit* of information, in which nobody will be much interested.

The word *bit* is used here in a simple way, as a word, but as you will see later in Unit 2, in modern terminology, it is used as the smallest unit of information, viz. binary digit (bit).

(ii) The next parameter of signal is the **frequency**, $f = \omega/2\pi$ where ω is the angular frequency in rad s^{-1}. It may be varying from instant to instant and often many frequencies may be present simultaneously at any *one instant*. The range of frequencies used, for any system is also limited. For example, for sound signal, the frequencies used are in the range 20Hz to 20kHz, but for normal telephony (handling speech signals) the range is only up to 4kHz whereas for police wireless telephony only 1kHz frequency range is good enough. This is called **band-limited communication**. Of course, in this case, the corresponding voice (sound signal) is somewhat distorted, and is of inferior quality.

For visual (optical) signal, however, the situation is quite different. In total electromagnetic (e.m.) frequency range, the visible range is a small part which has frequency of $\sim 10^{13}$ Hz. The e.m. spectrum relevant for communications is shown in Fig. 1.3. In the figure you will observe that the measurements are expressed in different units, viz. wavelength (λ) and frequency (ν). Sometimes the spectrum is also expressed in terms of wave number ($1/\lambda$) or energy ($\hbar\nu$) but this is applicable to frequencies normally not used in communications. Usually, lower frequencies upto RF are expressed in frequency terms; microwave to ultraviolet are referred in terms of wavelengths or wave numbers while deep *UV* to γ- rays are referred in terms of their energies. These different representations are partly due to different applications of these radiations and partly due to convenience.

You should note that *theoretically* the frequency of only single sine wave can be talked about. But in electronics we may talk about frequencies of square waves, which contain many (harmonic) frequencies.

The parameter of interest to human eye or video signal is not the optical frequency (10^{13} Hz related to colours) but the **intensity variation** of different colour shades. The term **optical frequency spectrum** is often used to indicate *the visible frequencies present in the source or emitted by the source*. These are of the order of 10^{13} Hz (see Fig. 1.3) and are perceived by us as colour. However, this colour definition is only approximate.

(iii) The last parameter in the model signal in Eq. (1.1) is the **instantaneous phase** ϕ. Note that concept of phase needs a reference zero phase which can be at a time $t = 0$. But to define $t = 0$ is not easy. Practically, the phase of a *sine wave* of a certain frequency can be compared with that of the other sine wave with the *same frequency*.

(iv) There is one more hidden parameter for an electromagnetic signal. You should remember that with this signal there is an associated electric field **E**. Since this is a vector, it is accompanied by a direction in the space, i.e. it has got a **polarisation**. The nature of waves depends on the oscillation of the field at the source and hence on the motion of the charged particles generating the field. If the charge is oscillating along a straight line (a = $\hat{a}$ sin ωt), it generates linear polarisation. If it is moving along a circle due to the combination of two harmonic sinusoidal oscillations at right angles to each other and out of phase by 90° (i.e., $x = \hat{x}$ sin ωt and $y = \hat{y}$ cos ωt, with $|\hat{x}| = |\hat{y}|$), then it generates circular polarisation. If $|\hat{x}|$ and $|\hat{y}|$ are not equal, then it is elliptical polarisation. If the motion of charges is random, then it will generate unpolarised light or randomly polarised light. This may be generated by many electrons moving randomly at the source. Normally we will be assuming linearly polarised electric waves for simplicity.

SAQ 2

Spend 2 Min.

Eq. (1.1) represents an oscillatory nature of the source. How would you represent a propagating wave?

Though our model may appear simple, the actual situation can be very complex because of $\hat{a}$ and ω variations. Here we have discussed only two types of signals (audio, video) so far, which are directly used in communication, but there are many signals and waves present in other *physics domains* e.g. (a) atmospheric *temperature* variations are important for meteorologists; (b) seismic waves indicating earthquake, sea waves, ripples are *mechanical* signals; (c) *magnetic* field variation can indicate the presence of iron ores, as observed by the satellites; and (d) smell variation signal is *chemical* in nature.

Frequency Bands	Sub-Bands	Frequency (ν)	Wavelength(λ)	Application
Optical	Visible Light			
		300 THz	0.001 mm	Fibre optic communication
	Infrared	30 THz	0.01 mm	Remote Controls
		3 THz	0.1 mm	
		300 GHz	1 mm	
Microwaves	Extremely High Frequency (EHF)			
		30 GHz	10 mm	Post & Telegraph TV earth stations Satellite Communication Cellular Mobile Cordless Telephone Commercial TV Radio Radar
	Super High Frequency (SHF)			
		3 GHz	100 mm	
	Ultra High Frequency (UHF)			
		300 MHz	1 m	
	Very High Frequency (VHF)	100 MHz	3 m	FM Radio
Radio Frequency (RF)		30 MHz	10 m	
	High Frequency (HF)	10 MHz	30 m	Short Wave (SW) Radio
		3 MHz	100 m	
	Medium Frequency (MF)			
		300 kHz	1 km	Medium Wave (MW) Radio
	Low Frequency (LF)			
Audio Frequency (AF)		30 kHz	10 km	Marine Navigation
	Very Low Frequency (VLF)			Ultra Sound SONAR
		3 kHz	100 km	
	Ultra Low Frequency (ULF)			
		300 Hz	1000 km	
		50 Hz	6000 km	Mains Electricity
		30 Hz	10000 km	
	Extremely Low Frequency (ELF)			
		3 Hz	100000 km	Earth's Electric & Magnetic Fields
		0		Battery Supply DC

kHz = 10^3 Hz; MHz = 10^6 Hz; GHz = 10^9 Hz; THz = 10^{12} Hz.

Fig. 1.3: The Electromagnetic Spectrum for Communication Applications

1.3.2 Receiver

Another important block of the communication system, shown in Fig. 1.2, is the receiver. It comprises the signal detector and relevant electronics to process the received signal and convert it into usable form. It is also called the **sink** of the information. In practice this received information may be further used for some successive actions. In modern systems, the output of the receiver may be given to some actuator (for control system) or display (like TV-CRT) or stored in electronic or magnetic memory for future use.

The main specifications or the parameters of the receiving system are (i) **sensitivity** to input signal; (ii) **amplitude (dynamic) range** of the input signal, which can be received and converted to output; (iii) **linearity** between the input and output signal; (iv) **frequency response** to reproduce input signals faithfully, i.e. **fidelity**.

Another important parameter of the communication system is the **signal to noise ratio** (SNR or S/N). If the signal is to be used gainfully, the system should not introduce any internal noise to mask the signal. Further, if there is any other stray (external) noise entering the communication link, along with the information signal, then we should use some signal processing technique like filtering to improve the SNR. There are various methods to improve the SNR, which will be discussed in the next unit. Of course, the simplest thumb rule is that the signal should be detectable above the noise level at any instant. To achieve this, you have to either increase the signal intensity or reduce the noise level so that the signal amplitude is above the lower threshold level of the receiver.

We have already discussed that the communication links, particularly the sink part, should be linear in amplitude response, i.e. the relation between the input and output amplitude should be linear as shown in the middle portion of the curve in Fig. 1.4a. It should also have good fidelity, i.e. the output amplitude should not be affected by the frequency of input signal at least in the frequency range of the operation (source spectrum). This is shown in Fig. 1.4b.

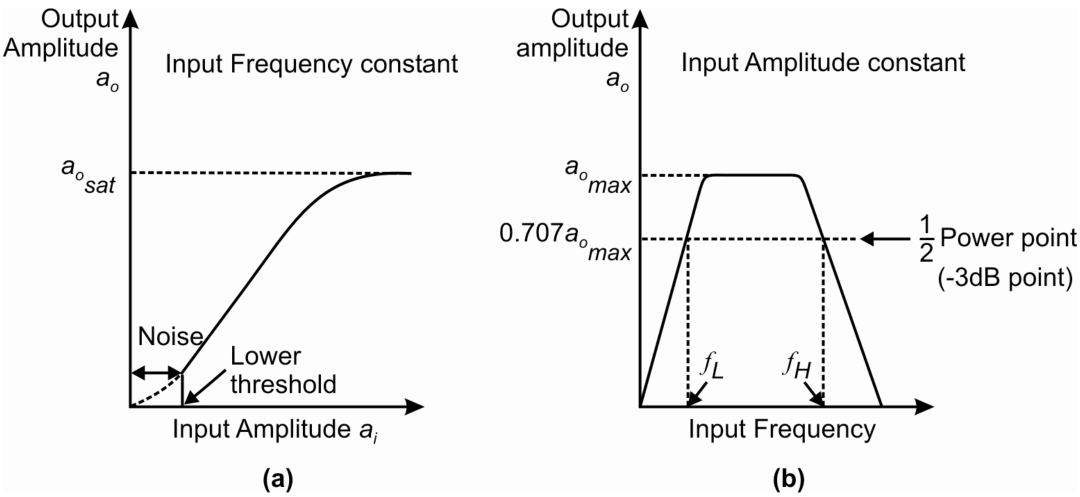


Fig. 1.4: a) Linear amplitude response of the receiver at constant frequency; and b) flat frequency response of receiver at constant input amplitude

The term *sensitivity* is used to mean either

(i) $\frac{\text{output}}{\text{input}}$

or

(ii) lowest limit of sensing, i.e. threshold of detection governed by the noise.

The amplitude response curve decides the **amplitude range** or **dynamic range** of the system. It is often governed by the amplifier saturation at the upper limit and by noise at the lower limit. The lower limit is called the **threshold input** or **sensitivity**. The frequency response decides the **frequency range** or **bandwidth** over which there is good communication, within tolerable noise limits. When the input signal crosses these ranges, we get amplitude and frequency distortions, which can be related to internal noise generation.

Note that these amplitude and frequency responses are similar to that of an amplifier; but this is not an amplifier system since the output power may not be more than the input power. In case of amplifier, the ac output power at the output is more than the input, at the expense of dc power.

Harmonics and Intermodulation terms: The real question is, what happens when the source signal crosses the above two limits (shown in Fig.1.4a & b) of the receiver. The following two things happen in this case:

(a) In Fig. 1.4a, we observe saturation of output amplitude, after certain input level. This results into a waveform distortion as seen in Fig. 1.5.

(b) Fig. 1.4b, shows an amplitude reduction if the signal frequencies are outside the receiver bandwidth. Here, we lose some information contained in the frequencies falling outside the range. This is observed particularly, if the incoming signal intensity is too low.

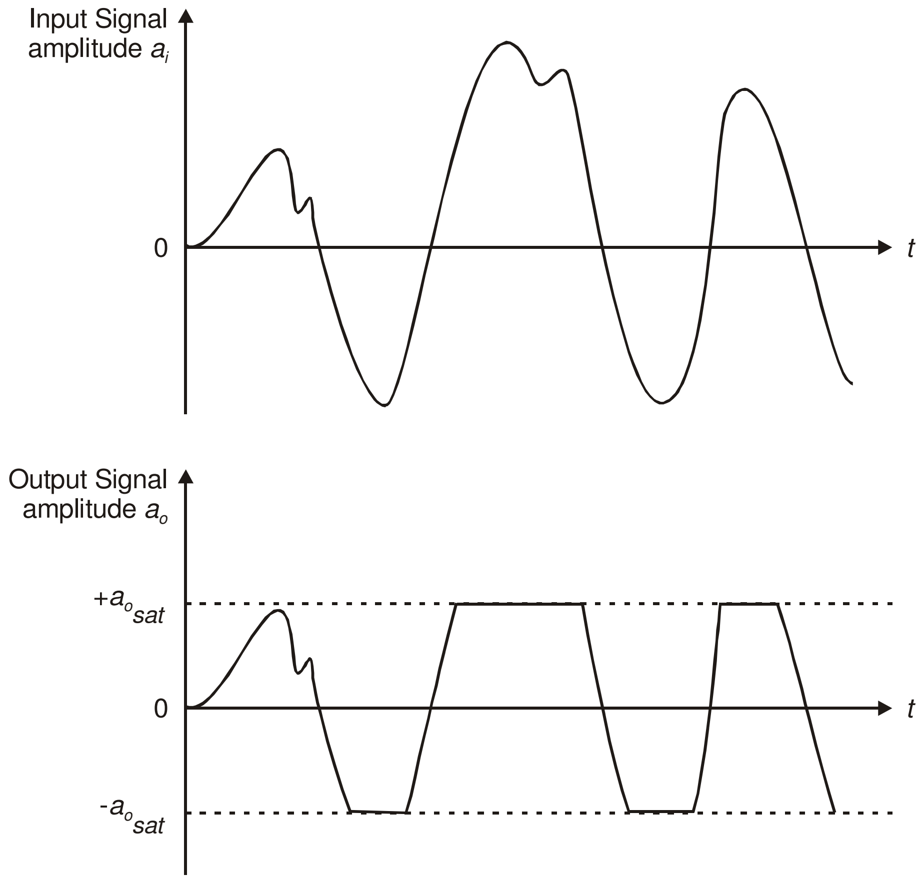


Fig. 1.5: Input signal with peak amplitude $\hat{a}_i$ results into output clipped at $a_{o_{sat}}$

Distortion means non-linear relation between input and output. This can be expressed as power series

$v_o = c_0 + c_1 v_i + c_2 v_i^2 + \dots$

When $v_i = \hat{v}_i \sin \omega t$, this results into a Fourier series.

Using Fourier theorem we can show that, even if the input signal is sinusoidal with only one term in Fourier series (corresponding to input frequency ω), the distorted output waveform will give a series

$$a_{out} = a_0 + a_1 \sin \omega t + a_2 \sin 2\omega t + \dots \qquad (1.2)$$

with some dc amplitude (a_0) and many higher **harmonics** (like $a_2 \sin 2\omega t$). If the input contains more than one frequency, say ω_1 and ω_2 , then the output of such non-linear system will contain not only harmonics $n\omega_1$ and $m\omega_2$, but also **intermodulation terms** i.e. ($n\omega_1 \pm m\omega_2$). You know that, if it were a good amplifier, it would contain only ω_1 and ω_2 frequencies at output and no beat frequencies (ω_1-ω_2). For generation of new frequencies, some **non-linear** element is essential which gives distortion.

You may now like to solve one SAQ.

SAQ 3

Spend 2 Min.

Name any two devices, which can be used for new frequency generation.

1.3.3 Medium for Communication

The difference between a propagating wave in Eq. (1.3) and an oscillation signal in Eq. (1.1) is that there is no velocity parameter in Eq. (1.1)

The third important block in the communication system is the *medium* between the transmitter and receiver. When a person taps on back of other person, by direct contact, the medium is not needed. But to get a signal at some distance (r), there are only two possibilities:

(i) Use of a wave

$$a = \hat{a} \sin \omega \left(t - \frac{r}{u} \right) \qquad (1.3)$$

where u is the velocity of the wave. (You must have recognised this expression which you had deduced in SAQ 2.)

or

(ii) Use of a field like magnetic or gravitational field of the earth, for remote sensing by a satellite.

For general communication purposes, these fields are not used and we will restrict our discussion to the waves only.

The basic need of communication system is that the signal needs to be transmitted without distortion through the propagation medium. Hence it should be *linear*, so that only the superposition principle applies. If not, we may get harmonics and intermodulation frequencies added to the signal. For both sound and optical (e.m.) waves, under normal intensity conditions, air is a linear medium. EM waves can travel through free space as well, and it is a linear medium for e.m. waves. But some media may show non-linearity, particularly at higher intensities, because of distortions at molecular/atomic levels. (e.g. distortions caused in some ionospheric regions by high intensity laser beams).

The second and more troublesome problem in wave propagation through medium is the *dispersion*. This is caused by the refractive index of the medium, which varies with the frequency of the wave. The change in refractive index causes variation in the velocity of the wave in the medium. As the signal contains many frequencies, some frequencies, i.e. some parts of the signal, are received earlier than the other parts and this gives distortion of the signal. As you will learn in later units, dispersion plays an important role in the performance of optical fibre communication system.

There are many other properties of wave propagation in medium which may affect the transmission of signal through the medium. They are (i) refraction; (ii) diffraction; (iii) reflection from ground or other objects and related interference; (iv) scattering and absorption losses; and (v) polarisation changes on path the of e.m. waves. These may give rise to distortions and noises.

After discussing the general aspects of a communication system, let us now elaborate about some special communication related circuits.

1.4 MODERN COMMUNICATION SYSTEMS

In the last sections we discussed various Physics principles involved in communications, viz. different domains of signals and source/sink/medium parameters along with the non-linearity effects. Now we will discuss about the modern communication systems.

1.4.1 Needs of Communication System and Related Parameters

The modern world is fast and vast. It expects the signal transfer to be quick (higher velocity) and to a longer distance with lesser path-losses. Further, the information content packed in a wave must be more and noise free. Let us now focus on different propagation-related parameters.

a. **Sound versus e.m. waves**: Use of sound signals for direct propagation is not economic as compared to optical (e.m.) signals because:

(i) The **velocity** of sound is very low (330 m s^{-1}) as compared to that of e.m. waves (3×10^8 m s^{-1}).

(ii) Sound waves require a **medium** for propagation, so we cannot use them much beyond the earth's surface (≤ 6-8 km) and hence not in the free space for satellites.

(iii)The **frequency band available** with sound waves in communication is mostly limited to ≤ 50MHz, because the atomic vibrations are much sluggish compared to e.m. field vibrations. The higher frequencies of e.m. waves and corresponding larger available bandwidths allow packing of more information in the e.m. waves. From this packing point of view, optical range is most preferred.

(iv) Higher velocity also means that for the same frequency (*f*), we have a longer wavelength (λ=v/*f*). For example, at 30 kHz, for sound, the wavelength λ is 1 cm and for e.m. wave it is 10 km. The longer e.m. wavelengths do not see small objects like building etc. on path, hence they can propagate to longer distances with lesser attenuation and without much diffraction or scattering, than the sound wave of same frequency. However, as can be seen from Fig. 1.3, at microwave frequencies (GHz), λ is few cm while at optical frequencies it is in microns.

b. **Long versus short e.m. waves:** Continuing the above discussion, if we have to make a choice between sound and e.m. waves, on all the four accounts, e.m. waves is a better choice for electronic communications. Amongst e.m. waves, the longer waves in the frequency range of 500 kHz to 30 GHz are the preferred ones, because they satisfy all the above conditions of high velocity, free space transmission, comparatively high band capacity as well as longer wave lengths. The optical waves, because of their shorter wavelengths, are stopped or scattered readily by even small obstacles like dust particles in the path of transmission.

You must have noticed that the wavelength is a crucial factor in determining the performance of the waves used. The smaller wavelengths provide large band capacity, however they pose a problem in transmission over the obstacles. The problem due to smaller wavelengths can be overcome by any of the following solutions:

(i) **Repeated reflection** is used in optical fibres and in Shortwave Radio band transmission.

(ii) Since the intensity of signal will fall as it travels longer distances, instead of just a reflector, a **repeater** can be used, which receives, amplifies and retransmits the signal in forward direction. You must have seen the repeater towers installed at regular intervals for use by the Post and Telegraph department and railway microwaves links. A geo-stationary communication satellite also uses a repeater which receives signals sent from the earth, amplifies them to adequate level (compensating for losses occurring in propagation) and transmits them back to the earth to be received by the ground receivers.

(iii) One more possibility is to use **a guided e.m. wave** instead of free space wave, i.e. a waveguide or optical fibre guide or transmission line.

(iv) Generally speaking, *no parallel communication system is easily possible for long distances*. For example a camera can capture the whole image simultaneously, but the same picture is converted to a scanned image and then transmitted serially, pixel by pixel.

Till now our discussion was related only to the basic, i.e. base-band signal and its *direct transmission* by sound wave and visible light (e.m.) waves. But it is a common practice to convert this signal into a suitable form and then carry out the transmission.

The simplest solution for the problems related with sound waves is *to convert sound signal to electrical signal* (voltage $\boldsymbol{V}$ or field $\mathbf{E}$) using proper transducer, i.e. a microphone and use these Audio Frequency (AF) electrical signals for transmission. We can transmit it, on a pair of conducting wires called **transmission lines** working as the medium of transmission. The wire/line communication is used in telephony and Local Area Network (LAN) of computers. (These days even these small distances are covered by wireless communication). At the receiver end, we have to convert these electrical signals back to sound by a transducer-actuator, i.e. a loud speaker.

The second possible solution to transmit the output of the microphone after necessary amplification as an *e.m. radiation* (field $\mathbf{E}$) is by using **an antenna**. This can decrease the cost of wires and increase the distance of transmission. Further, it can be received by many receivers simultaneously.

Advantage of transmission line communication is that the current in a circuit remains practically constant, unlike the electric field in radiated wave.

The first method is called *one to one* communication and the second is *one to many*. This latter solution uses the free space as medium, with no cost. However it is rarely used in practice directly for basic signals because we need a very big antenna. The reason is that, for efficient transmission, the size of antenna should be ~λ/4. Hence, for 30 kHz AF signal, we need an antenna of 2.5 km size.

Please note that the first solution can also be *many to many* or *any to any* users, by proper intermediate switching facility. This is done in telephony for providing access between any two subscribers, as you will learn in Unit 7.

You may like to solve one SAQ now.

SAQ 4

Spend 2 Min.

Give any two examples where e.m. radiation is *not* used for signal communication.

---

c. Modulation: The problem posed by a large antenna size can be easily overcome by reducing the wavelength of the transmitted wave. This is done by the modulation technique. Here, one more signal frequency, called *carrier frequency* (ω_c) is used, which is much higher than the basic signal frequency (ω_m) and ω_m is used to change or *modulate* the carrier wave. From Eq. (1.1), we can write these two signals as:

Carrier signal:

$$v_c = \hat{v}_c \sin(\omega_c t + \phi_c) \qquad (1.4a)$$

and modulating base-band signal:

$$v_m = \hat{v}_m \sin(\omega_m t) \qquad (1.4b)$$

Here we neglect the phase ϕ_m for convenience.

Note that v_m is the *instantaneous* signal voltage value at time *t*.

Modulation means changing the $\hat{v}_c$ or ω_c or ϕ_c (and sometimes the polarisation), according to the *instantaneous* value of the intelligent signal, i.e. according to v_m. Thus the modulated carrier signal will be, either

$$v_c^{mod} = (\hat{v}_c + m_a v_m)\,[\sin(\omega_c t + \phi_c)] \qquad (1.5a)$$

or $$v_c^{mod} = \hat{v}_c \sin\left[(\omega_c + m_f v_m)\,t + \phi_c\right] \qquad (1.5b)$$

or $$v_c^{mod} = \hat{v}_c \sin\left[\omega_c t + (\phi_c + m_p v_m)\right] \qquad (1.5c)$$

These equations represent amplitude (AM), frequency (FM) and phase (PM) modulations respectively. *m* is often called **modulation index**. Obviously m_a is a dimensionless fraction, but m_f has dimensions of Hz V^{-1} and m_p has dimensions of rad V^{-1}.

In general, we can have any complex rule (code) of modulation. Then, we have to use corresponding demodulation/detection at the receiver. Such complexity is normally introduced for military secrecy, but for commercial applications like radio, TV, telephony, only the predefined methods with international conventions are used.

Modulation can be used for both the modes of transmission, viz. (i) transmission line communication, and (ii) wireless communication using antenna.

There are two main classes of modulations (i) Analog (including AM, FM, PM) and (ii) Digital or pulse. You will study the details of these modulations and their practical variants in Unit 5 and 6. Presently, we will mention about two basic physics principles involved in modulation.

(i) **Sampling**: As you know, sampling is the basis of statistics. The basic principle behind sampling is: *you should take sufficiently large number of samples at sufficiently small intervals with respect to period of the phenomenon you want to study*. For example, if you want seasonal variation of rain, you take monthly samples, but for variation of solar radiation you have to take hourly samples. Similarly in modulation (Eq. 1.5), we sample the basic information signal (ω_m). For example, if we consider amplitude modulation, we can write Eq. (1.5a) as:

$$v_c^{mod} = \left[\hat{v}_c + m_a v_m(t)\right]\sin\omega_c t = \left[\hat{v}_c^{mod}(t)\right]\sin\omega_c t \qquad (1.6)$$

i.e. the peak amplitude of modulated carrier changes with time at the frequency of ω_m, as shown in Fig. 1.6. For transmitting ω_m signal, we take its samples by high frequency ω_c carrier, which gives samples of $v_m(t)$ at each peak of ω_c, indicated by crosses in Fig. 1.6.

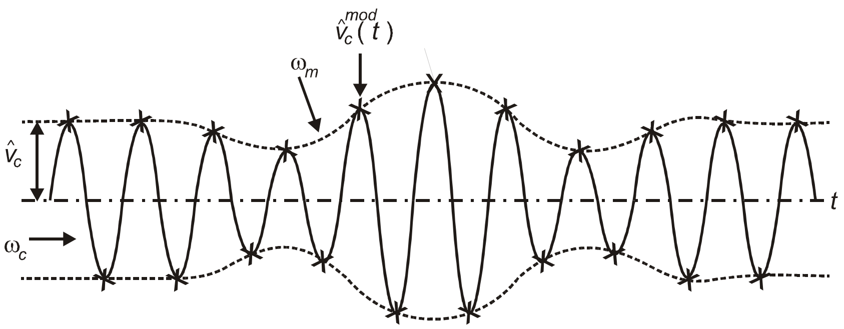


Fig 1.6: Sampling of analog signal v_m at the peaks of ω_c in amplitude modulation

The similar principle can be used to discretise an analog signal. A process of sampling is used for this purpose as shown in Fig. 1.7.

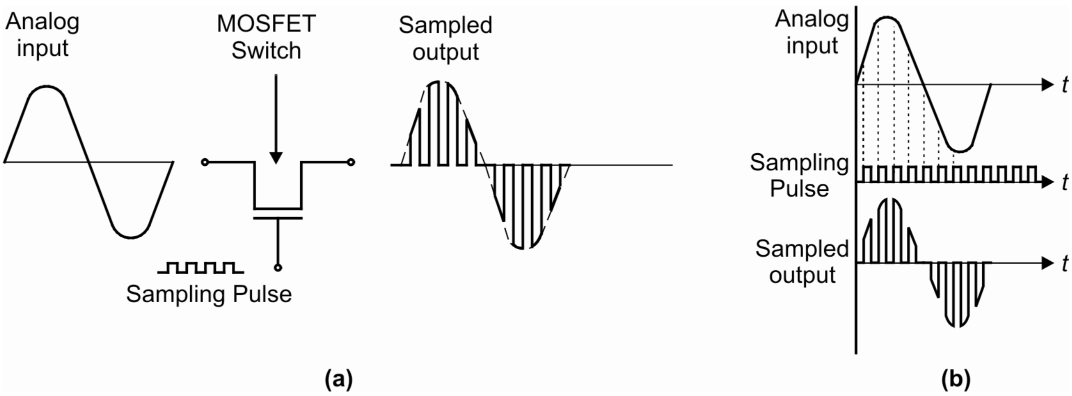
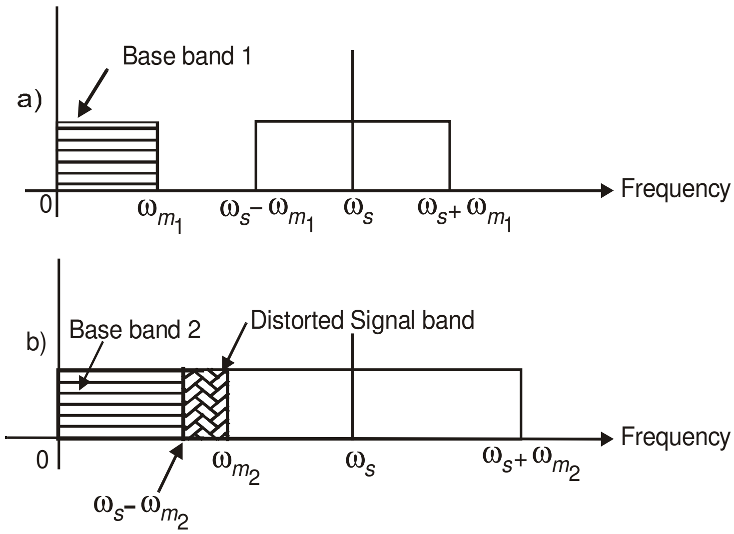


Fig. 1.7: (a) Circuit for discretising analog signal; and (b) Discrete form of analog signal

If you observe the sampled output shown in Fig. 1.7b, it appears similar as the amplitude modulation, where the modulating signal is sampled at the peaks of the sampling signal. When the basic (base-band) signal with frequency ω_m is sampled using pulse carrier or sampling carrier with frequency ω_s, it is called *pulse amplitude modulation*. In this case, the sampling frequency ω_s should be sufficiently high with respect to ω_m. It is essential to have $\omega_s \geq 2\omega_m$. The reason for this constraint will be clear from Fig. 1.8.



Fig, 1.8: Frequency spectrum of pulse amplitude modulation a) $\omega_s > 2\omega_m$; and b) $\omega_s < 2\omega_m$

We can rewrite Eq.(1.6) as

$$v_s^{mod} = \left[\hat{v}_s + m_a v_m(t)\right] \sin \omega_s t \quad (1.7)$$

where $v_m(t) = \hat{v}_m \sin \omega_m t$ is the modulating signal and $v_s(t) = \hat{v}_s \sin \omega_s t$ is the sampling (or carrier) signal.

Hence, we write

$$v_s^{mod} = \left[\hat{v}_s + m_a \hat{v}_m \sin \omega_m t\right] \sin \omega_s t. \quad (1.8)$$

We use the identity
$2 \sin A \sin B = \cos (A-B) - \cos (A+B)$

In this expression, the second term can be written as

$$m_a \hat{v}_m \sin \omega_m t \sin \omega_s t = K\left[\cos(\omega_s - \omega_m)t - \cos(\omega_s + \omega_m)t\right] \quad (1.9)$$

where $k = \dfrac{\hat{v}_m m_a}{2}$

From Eq. (1.9), we observe that the sampled output will have three frequency components viz. ω_s, $(\omega_s - \omega_m)$ and $(\omega_s + \omega_m)$. Here, $(\omega_s - \omega_m)$ is the lower side-band frequency, where as $(\omega_s + \omega_m)$ is the upper side-band frequency. The last component is the up-converted frequency where the modulated frequency is higher than ω_s by ω_m.

In Fig. 1.8a, frequency is $\omega_s \geq 2\omega_{m_1}$, hence the modulated frequency band (from $\omega_s - \omega_{m_1}$ to $\omega_s + \omega_{m_1}$) is well separated from ω_{m_1} and we can use a filter to remove the base band frequency ω_{m_1}. Now, in Fig. 1.8b, the modulating base band frequency is such that $\omega_s < 2\,\omega_{m_2}$. In this case, the modulated frequency band and the base band frequency overlap; and when we try to use a filter to take away ω_{m_2}, we get distorted signal due to overlapping bands shown as double shaded frequency band. In such cases, we have to restrict or band-limit the base band signal in the distortion free range.

From this discussion, we can conclude that *the sampling frequency should be at least double the signal frequency for distortion free signal processing*. This is known as the **sampling theorem**.

The sampling frequency is not directly used for transmission purposes. It is actually an intermediate frequency used inside a circuit, for frequency multiplexing or line-transmission, or for coding the signal. For the actual transmission by radiation, this (sampled) signal is further modulated on a very high frequency carrier wave.

(ii) **Need of a non-linear device**: Basic mathematical physics says that *only infinite train of sine wave* (represented by Eq. (1.1)) *can have a single frequency*. Any deviation from this assumption will mean that the signal contains more frequencies. These can be obtained, mathematically, by Fourier analysis expressed in Eq. (1.2). But to get the new frequencies practically, a non-linear electronic device is essential. e.g. diode detector/rectifier (to get dc from ac) or amplitude saturation shown in Fig. 1.5. The non-linear device will generate many harmonics and inter-modulation frequencies. To select only the required frequency at output, a filter is essential.

Multiplexing is facility for any user to contact any other user. Switching is the tool used for realising multiplexing.

d. Multiplexing: You know that there are two main modes of transmission viz. line communication and wireless communication. As seen above, the line communication is basically for *one to one* communication but with proper switching it can be used for *any of many* to *any of many* communication. This is achieved by *multiplexing*.

However, space communication is basically for *one* to *many*. This radiative system can also be used for *any one of many* to *many*. This is done in radio **broadcasting** by different stations like Mumbai, Kolkata, Chennai and Delhi. Each station is assigned a different carrier frequency for broadcasting. This is called **Frequency Division Multiplexing (FDM)**. In short, the same medium can be used for many signals (customers) by multiplexing methods.

1.4.2 Typical Communication Circuits

You are already conversant with the basic electronic circuits like amplifiers, oscillators, power supply etc. In communication system, apart from these circuits, some special circuits like modulators are used. A typical communication system has the following main building blocks: (i) sensor transducer for converting source signal into electrical signal (like microphone), (ii) input voltage amplifier (buffer), (iii) modulation circuit, (iiia) carrier frequency generator, (iv) final high power amplifier, (v) transmitter antenna, (vi) medium for transmission, (vii) the receiver antenna, (viii) low noise amplifier (LNA) for received signal, (ix) demodulation (detector) circuit, (x) output base band amplifier (with sufficient power) and (xi) output actuator transducer (like loud speaker). These blocks are shown in Fig. 1.9. You will be learning the details of these blocks in the following units.

You must have noticed one important common speciality of the special communication blocks i.e. (a) modulator (frequency up-converter), and (b) demodulator/detector (frequency down-converter) uses a single physics principle: *new frequencies cannot be generated unless some non-linearity is used.* The easiest way to achieve this goal is to generate the new frequency by selecting the Q point of the device transistor, such that it gives distorted output of the input signal. This can

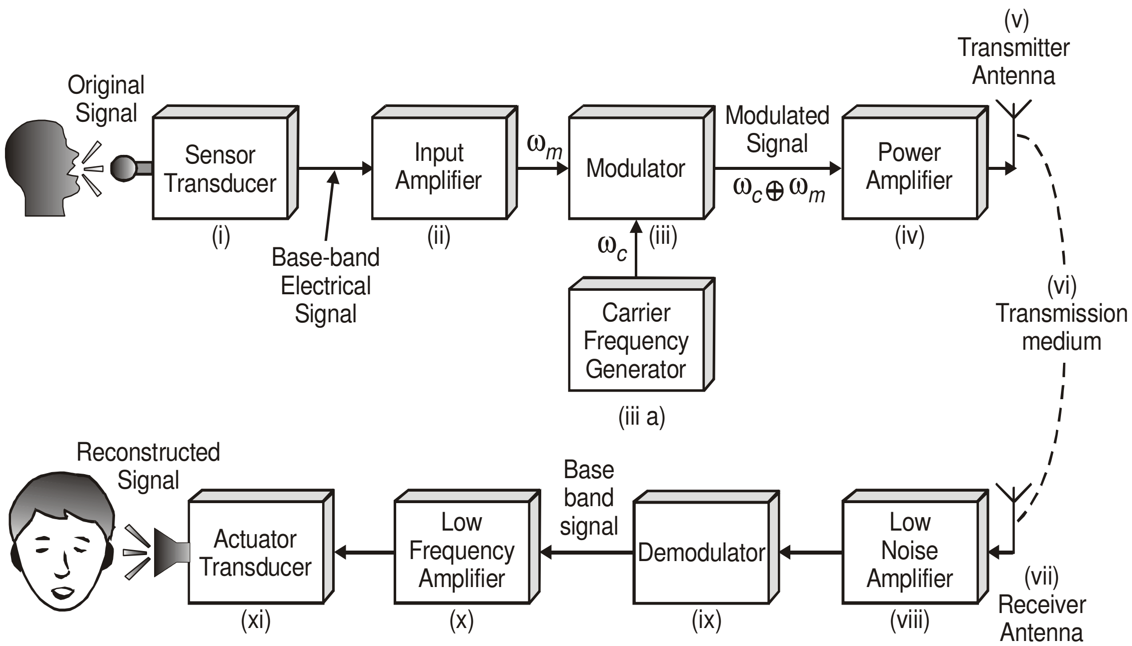


Fig. 1.9: Detailed block diagram of a typical communication system

be done by using bias nearer the cut off or saturation. Thus class B or C amplifier circuits are used for the purpose. Class C is used for generating the carrier frequency ω_c. In modulator, very often, Q point is selected nearer to the cut off region of the device. Note that the Q point is selected such that it distorts ω_c, but *not* ω_m. In other words, we need to use proper (electrical) boundary conditions, for using the non-linear nature of the device such that we do not distort the original signal ω_m. In demodulation, generally, diode rectifiers are used.

Till now, we have discussed only the *electronic* blocks, which are non-linear. But there is one more special communication block, which is electrical and linear (and reciprocal), viz. the **Antenna (AE)** or aerial for transmitting and receiving the electromagnetic (RF and microwave) waves. The normal transmission lines do not radiate much, because the two wires carry opposite current and being near to each other, nullify the radiation. But if transmission line of $\lambda/4$ length is opened out (to 180°) to get a $\lambda/2$ *dipole antenna*, as shown in Fig. 1.10, it starts giving out radiation and the corresponding receiver antenna starts reciprocally, catching the radiation. We can have many types of wire geometries corresponding to different types of antennae but a basic radiator is a small element of wire with length δl ($<<\lambda$) carrying constant RF current along δl. This is called an **elementary electrical dipole (EED)** or a **Hertzian dipole**. An actual antenna gives a *directional radiation pattern,* which is basically just the *interference pattern* in space, summed over all such elemental δl. The wire antennae are *primary* radiators, but we can have *secondary radiators*, like parabolic reflectors. You will learn further details about antennae in Unit 3.

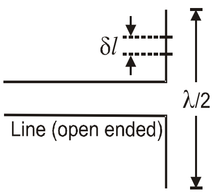


Fig. 1.10: λ/2 Antenna

One of the important parameters associated with the radiating sources is the wavefront. Imagine a simple case of charge moving sinusoidally and harmonically along a straight line. Then we get a cylindrical wavefront going round it, at right angles to the axis of motion of the charge (i.e. a.c. current) as shown in Fig.1.11. This is the model situation assumed in electrical field. Any wire antenna system formed out of small *elementary electrical dipoles* (EED) results into wavefront that gives rise to its directional radiation pattern.

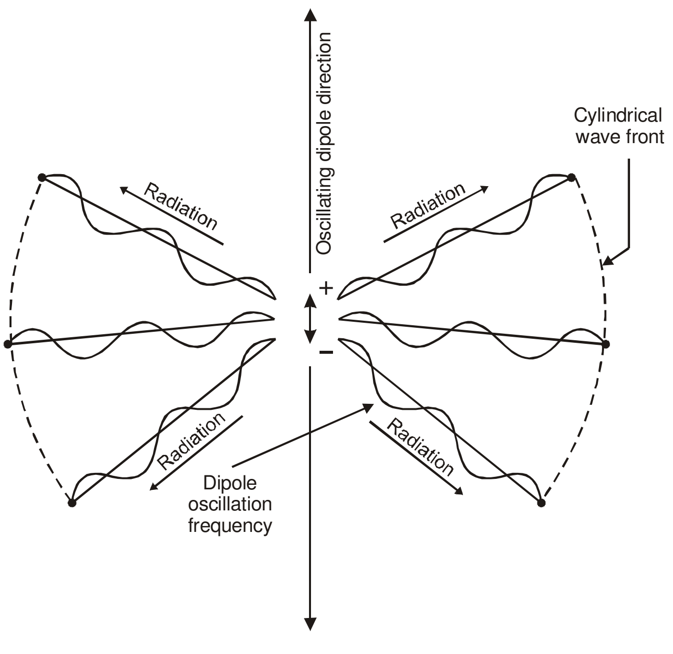


Fig. 1.11: Cylindrical wavefront due to harmonic movement of charge

Spend 1 *Min.*

SAQ 5

What will be the shape of the wavefront if the radiator is in the form of a point source?

In optics, the convention is slightly different. As an approximation, we consider the source to be a point source. When such a point source has many electrons in it, moving randomly in all directions, it results into an average spherical wavefront. This is an extensively used assumption in ray optics or lens/mirror theory but theoretically, a point source cannot exist in electromagnetics.

When we consider a very small area-element of spherical or cylindrical wavefront at a distance sufficiently away from the source, we can approximate it as a *plane wavefront*, i.e. a plane with in-phase field at every point on it. These wavefronts are used to calculate the directional radiation patterns of the antenna systems.

1.4.3 Modern Variants of Modulation

So far, we have discussed the direct modulation of continuously varying audio and video signals. These signals were of analog form. However, by *sampling* and *quantisation* of these signals, we can convert them into digital signals. Because of various advantages like noise immunity and circuit simplicity, these digital signals find wide application in communication. The merger of computers with the electronic communication technology has given rise to the vast field of **Information Technology** (IT).

In the last section, you studied the sampling of the signal by sinusoidal waveform for modulation. However, pulses can also be used for sampling. The pulse sampling is more suited for present day digital circuitry. Digital techniques have gone far ahead now with different digital modulation techniques like pulse code modulation (PCM) using binary code; phase shift keying (PSK); Frequency Shift Keying (FSK) etc. You will learn the details of digital modulation techniques in Unit 6.

One important point to be noted in the present age of integrated circuits (ICs) is that they need to be fabricated in bulk quantity so that they become commercially viable and cheaper. You already know that AND, OR and NOT are the three basic gates. From these gates we can generate NAND and NOR gates. Either NAND or NOR can be used as the building blocks for making different digital circuits. With some basic manipulation rules and hardware IC fabrication processes, we can make any complex circuits ranging from Medium Scale Integration (MSI) to very large scale integration (VLSI) with device density of the order of thousands of transistors, in a single chip of mm^2 size. This is not so easy with analog circuits, but is very easy with digital circuits because of limited number of basic gates. This makes digital communication a promising field. Still some analog blocks like phase locked loop (PLL) as detector and voltage controlled oscillator (VCO) as local oscillator are used in communication systems.

Two important parameters, not to be missed, in digital systems are (i) sampling frequency and (ii) maximum number of levels (and corresponding number of binary digits, i.e. bits required) for any sample. Any signal is a time variation of amplitude, which is expected to be continuous. But the user sensors (eye, ears) have got finite response time which limits the capacity for resolution. For example, a human ear cannot separate out sound, if its variations come faster than 30 μs (30 kHz). So we should sample the signal at a rate sufficiently faster than this response time (minimum twice i.e. at least at every 15 μs).

The next step is the digitisation of the dynamic amplitude range, which is invariably done in binary code with base 2 as digits, instead of base 10 in decimals system. Thus the total range of signal amplitude is to be divided in maximum 2 or 4 or 8 or…..2^n levels. This amplitude range can be then represented by 1 or 2 or 3 or……n binary digits (called bits). Conversion of analog voltage into digital form is achieved by an *analog to digital converter* (ADC). In digital communication (like satellite communication), systems with specific number of bits are used (often 16-bits or 32-bits).

You will learn the details of digital signal and its processing in the next unit.

Spend 3 Min.

SAQ 6

In your opinion, can we reconstruct exactly the original analog signal from its digitised form?

After studying about the essentials of transmitter and receiver, let us now discuss briefly the media used for signal transmission.

1.5 COMMUNICATION MEDIA

In a communication system, we need a medium between transmitter and receiver to communicate the information signal. Ideally, this medium has to be linear and reciprocal. Let us now discuss the basic physics behind wave propagation.

Disturbance at a point disturbs other nearby points, which in turn disturb the further points. This is the Huygen's principle in optics. It states that, each point on the wavefront of a wave acts as a *secondary source*. In case of sound, it is the displacement of particles causing pressure disturbance. In case of e.m. waves, it is a changing electric field producing a changing electric current which, in turn, produces magnetic field and hence electric field at the next point. The current can be a motion of actual charged particle (in wire like medium) or can be the Maxwell's displacement current in free space. This displacement current gives the free space e.m. wave for line of sight propagation in microwave repeaters.

The e.m. wave can be initially provided by an EED (discussed in the last section). In an EED, the current can flow along any bend on a conducting wire and the associated e.m. field wave travels along the surface of the wire. The same is true of a metal sheet or a sea surface or the ground. Hence we get e.m. **surface** or **ground waves**, which travel even along the curved surface of the earth. But in this case, the scattered radiation losses limit the distance (d) up to which the waves are strong enough to be used for communication. The ratio of the distance and this wavelength of the wave (d/λ) gives the attenuation factor, which determines the performance of the wave transmission. A larger wavelength (λ) results into smaller attenuation factor i.e. the ground waves are more useful at lower frequencies. The mid-frequency (MF) band of radio (~300 kHz to 3 MHz) uses this mode of transmission. It is interesting to note that, some times the surface waves at 30 kHz, produced by lightening discharges, easily propagate even across the continents on the earth.

Another way of transmitting waves is to conduct it in a medium with confined boundary as shown in Fig.1.12. In optical fibres, repeated total internal reflection is used to guide the optical wave even along the bend fibres as shown in Fig. 1.12a. At microwave frequencies hollow metallic conductors called **waveguides** are used (see Fig. 1.12b).

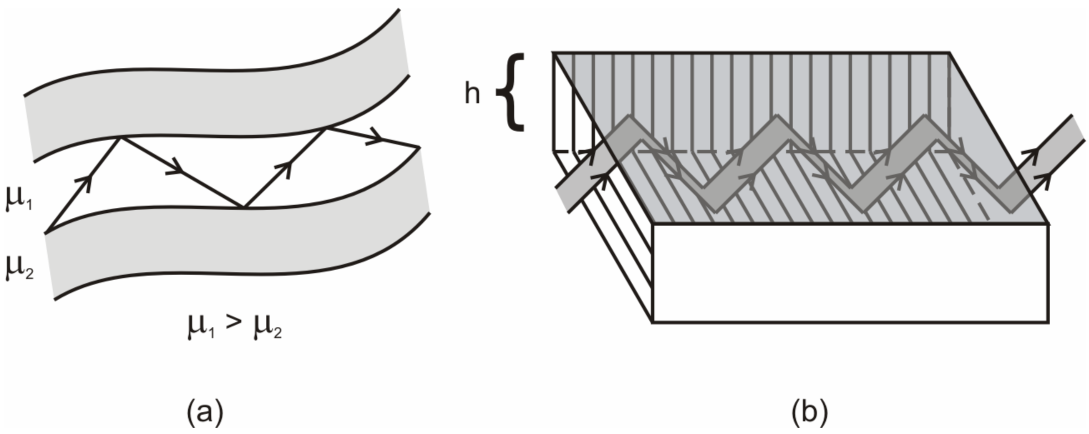


Fig. 1.12: a) Wave propagation in optical fibre; and b) Waveguide for microwaves

All the above discussion is a physical picture but basic physics of the e.m. wave propagation is given by the Maxwell's second order differential equations (in **E** and **H**). Its solution with different boundary conditions gives different modes of propagating wave equation. The simplest solution is the plane wave solution used extensively in the geometric optics.

Due to the higher attenuation factor, at higher frequencies (above 1 MHz) the surface wave cannot propagate over a long distance. But it is possible to get propagation through air between the transmitter and receiver antennae fixed on elevated towers shown in Fig. 1.13(i). This mode of transmission is called **line-of-sight** communication. It is also possible to transmit the signal by reflecting it from the ground. This is shown by (ii) in Fig. 1.13. You should not confuse this reflected wave with ground wave discussed above. The line-of-sight and ground reflected waves are called **space waves**. Microwave signals are sent in this mode.

In space wave transmission there is a problem caused due to ground reflected wave, which gives interference and distorts the signal at the receiver antenna. Also earth's curved surface limits the distance of line-of-sight communication.

There is one more type of radio wave propagation, called **ionospheric** or **skywave** propagation. This is shown in (iii) of Fig.1.13. In the atmosphere, at the height of 60 km to 300 km, the solar (UV) radiation ionises the air molecules. Due to reducing density of air with increasing height, there is a formation of different ionised layers, as

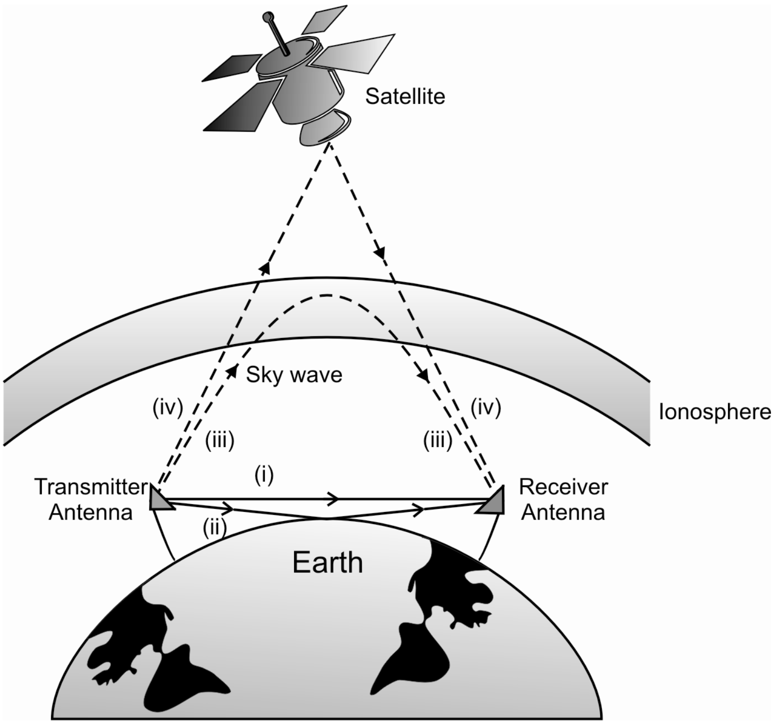


Fig. 1.13: Free space and surface wave transmission

shown in Fig. 1.14. During the daytime there are four ionospheric layers D, E, F_1 and F_2 at 60,100,200 and 300 km respectively. At night, since there is no sunlight to cause ionisation, there are only two layers, E and F_2. When the e.m. wave passes through the ionospheric plasma, it sets the electrons in the layer into vibration and in turn they produce e.m. radiation. The resultant field produces refraction and at certain frequency called *critical frequency of layer* it gives rise to total internal reflection. Since these layers have different spectro-photochemical reactivities, they have different critical frequencies. The critical frequency range is between 1.5 to 30 MHz. At night, the radio wave transmission can occur with lesser scattering loss in lower layers and can cover longer distance.

Fig. 1.14: Ionospheric layers

At higher frequencies than the critical frequency, the radiation skips out of the layer, as shown by (iv) in Fig. 1.13. These types of waves can be used for reaching the transmitter-receivers (transponders) in the satellites, which are placed in the geo-stationary orbit at about 36,000 km height. Typical frequencies used for satellite communication are between 1 GHz and 10 GHz. Typically the signal of 6 GHz is up-linked or sent in form of highly directional narrow beam aimed at the satellite. The receiver on the satellite receives it, amplifies and transmits it back at 4 GHz frequency. This is called down-linking. The difference in the up- and down-link frequencies is kept to avoid any interference between the received and transmitted signals. You will be learning the details of satellite communication in Unit 11.

Thus, we have the following possible propagation modes and media:

- Surface (ground) wave (MF band);
- Free space - line of sight at UHF/VHF/Microwave for radio and TV transmission over short distance and microwave for satellite communications;
- Ionospheric reflected (SW band);
- Two wire transmission lines and co-axial cable (AF to microwaves);
- Waveguides (microwave transmission over short distance); and
- Optical fibre guides (IR and visible range).

Let us now summarise the points you learnt in this Unit.

1.6 SUMMARY

- Communication is used by biological entities and it can be in anyone of the many physics domains, viz. mechanical, light, electrical, chemical etc.
- Human communication is mainly in two forms, viz. audio and visual.
- The audio signal is often a point source and emits the sound waves serially one followed by the other.
- The visual picture signal source is a 2-D with an array-matrix of point sources (called pixels) emitted simultaneously. However, for transmission, the signal needs to be in a serial form, hence for the TV, scanning process is needed.
- Mathematical model of signal can be represented as periodic wave $a = \hat{a} \sin\ (\omega t + \phi)$ where $\hat{a}$ is the maximum signal amplitude, ω is the frequency of signal and ϕ is the phase factor.
- The signal can be analog or digital in nature.
- Sampling and digitisation can convert analog signal into digital form.
- Any communication system consists of three parts: (i) source or transmitter; (ii) medium of transmission; and (iii) sink or receiver.
- In order to make signal transmission of high quality, cost effective, practical, efficient and with low noise, modulation techniques are used.
- The special electronic communication circuit blocks, like modulator and demodulator (detector), use non-linear devices in order to generate required new frequencies.
- The media used for transmission of signal should be linear in nature.
- Antenna has directional pattern, which is nothing but an interference pattern.
- Most frequently used communication media are free space, metal conductors and optical fibres.

1.7 TERMINAL QUESTIONS

Spend 15 *Minutes*

1. What are the frequency ranges included in following bands?

   (i) RF, (ii) Microwave, (iii) Infrared and (iv) Visible.

2. What are the frequency ranges included in following sub-bands?
   (i) MF (ii) HF, (iii) VHF, (iv) UHF (v) SHF and (vi) EHF
3. List three parameters of a high frequency carrier that may be varied by a low-frequency intelligence signal.
4. What are two basic limitations that restrict the ideal performance of a communication circuit?
5. Calculate the minimum sampling frequency required to sample the audio range signal.

1.8 SOLUTIONS AND ANSWERS

Self Assessment Questions

1. If a picture is to be transmitted in parallel from, each pixel should have a dedicated channel of communication (like wire or optical fibre). Typically a TV signal is sent in 700×525 format. Hence we would have to run more than three hundred thousand wires or fibres over hundreds of kilometre length. This is not feasible.
2. $a = \hat{a} \sin \omega \left(t - \frac{r}{u} \right)$, where u is the velocity of the wave.
3. Diode, bipolar junction transistor, field effect transistors or vacuum tube devices, are non-linear devices, which can be used for new frequency generation in communication circuits when biased properly.
4. Microwave oven, microwave plasma for coating the materials, room illumination with natural or artificial light, and use of infrared radiation for heating are examples of the use of e.m. radiation for applications not directly related to communications.
5. Spherical wavefront.
6. When the signal is converted from analog to digital, an error is introduced due to discrete voltage steps of digital form. When converted back from digital to analog form, the output is in the form of voltage steps, corresponding to the digital value. This kind of quantisation of signal introduces error and the original continuous signal is not obtained back exactly. By increasing the number of digital steps, this error can be minimized.

Terminal Questions

1. (i) Radio Frequency (RF) 300 kHz–300 MHz
   (ii) Microwaves 3 GHz–300 GHz
   (iii) Infrared (IR) 300 GHz–400 THz
   (iv) Visible 400 THz–750 THz
2. (i) Medium Frequency (MF) 300 kHz–3 MHz
   (ii) High Frequency (HF) 3 MHz–30 MHz
   (iii) Very High Frequency (VHF) 30 MHz–300 MHz
   (iv) Ultra High Frequency (UHF) 300 MHz–3 GHz
   (v) Super High Frequency (SHF) 3 GHz–30 GHz
   (vi) Extremely High Frequency (EHF) 30 GHz–300 GHz
3. The amplitude, frequency or phase of the high frequency carrier can be modified by the low frequency intelligence signal. This is *modulation*.
4. The amplitude and frequency response of the circuit.

5. For audio range, the maximum base band frequency is 20kHz. Hence the minimum sampling frequency should be $2 \times 20\text{kHz} = 40$ kHz.

Reference Material:

1. *Communication Systems* by Haykin, Simon; (III Edition) (John Weiley & Sons)
2. *Modern Digital and Analog Communication Systems* by Lathi, B.P.; (III Edition) (Oxford University Press)
3. *Electronic Communication Systems* by Kennedy, George; (III Edition) (Tata McGraw-Hill)