Nagaland Board of School Education
Upper Bayavü, Kohima - 797001
Website: *www.nbsenl.edu.in*, Email: *nagaboard@gmail.com*

NOTIFICATION NO.132/2024
Dated Kohima, the 16th December, 2024

NO.NBE-14/Ad-Tb(12)/2024-25 :: It is hereby notified for information of all Higher Secondary Schools having Science stream that the NBSE have included the topic **"Transistor"** and the Unit of **"Communication system"** in the **Class XII Physics syllabus**. But now, in the revised NCERT Textbook of Physics Class XII, it is found that this topic/unit is not included in the textbook. In this regard, the NBSE is sharing the soft copy of this topic/unit contents to students and subject teacher as a supplement to the textbook.

Therefore, all the Principals are requested to share this soft copy to the Physics teacher and the students of Class XII Science stream positively till the revision of syllabus/textbook is taken up by the Board.

Sd/-Asano Sekhose
Chairperson

NO.NBE-14/Ad-Tb(12)/2024-25/3567 Dated Kohima, the 16th December, 2024

Copy to:

1. The Commissioner & Secretary to the Government of Nagaland, School Education & SCERT, Kohima for information.
2. The State Mission Director, Samagra Shiksha, Nagaland, Kohima.
3. The Principal Director, School Education, Nagaland, Kohima.
4. All the Principals of Higher Secondary Schools having Science Stream for information and necessary action.

16/12/24
(Iluheing Nsarange)
Jt. Secretary (Academic)

SEMICONDUCTOR

ELECTRONICS: MATERIALS, DEVICES AND SIMPLE CIRCUITS

Junction Transistor

The credit of inventing the transistor in the year 1947 goes to J. Bardeen and W.H. Brattain of Bell Telephone Laboratories, U.S.A. That transistor was a point-contact transistor. The first junction transistor consisting of two back-to-back p-n junctions was invented by William Schockley in 1951.

As long as only the junction transistor was known, it was known simply as transistor. But over the years new types of transistors were invented and to differentiate it from the new ones it is now called the Bipolar Junction Transistor (BJT). Even now, often the word transistor is used to mean BJT when there is no confusion. Since our study is limited to only BJT, we shall use the word transistor for BJT without any ambiguity.

Transistor: structure and action

A transistor has three doped regions forming two p-n junctions between them. Obviously, there are two types of transistors, as shown in Fig. 14.27.

(i) **n-p-n transistor**: Here two segments of n-type semiconductor (emitter and collector) are separated by a segment of p-type semiconductor (base).

(ii) **p-n-p transistor**: Here two segments of p-type semiconductor (termed as emitter and collector) are separated by a segment of n-type semiconductor (termed as base).

The schematic representations of an n-p-n and a p-n-p configuration are shown in Fig. 14.27(a). All the three segments of a transistor have different thickness and their doping levels are also different. In the schematic symbols used for representing p-n-p and n-p-n transistors [Fig. 14.27(b)] the arrowhead shows the direction of conventional current in the transistor. A brief description of the three segments of a transistor is given below:

• *Emitter*: This is the segment on one side of the transistor shown in Fig. 14.27(a). It is of moderate size and heavily doped. It supplies a large number of majority carriers for the current flow through the transistor.

• *Base*: This is the central segment. It is very thin and lightly doped.

• *Collector*: This segment collects a major portion of the majority carriers supplied by the emitter. The collector side is moderately doped and larger in size as compared to the emitter.

We have seen earlier in the case of a p-n junction, that there is a formation of depletion region across the junction. In case of a transistor depletion regions are formed at the emitter base-junction and the base-collector junction. For understanding the action of a transistor, we have to consider the nature of depletion regions formed at these junctions. The charge carriers move across different regions of the transistor when proper voltages are applied across its terminals.

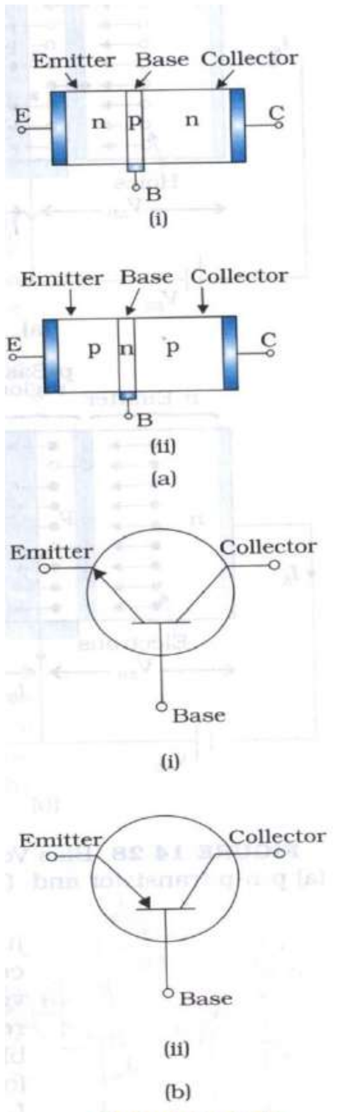

FIGURE 14.27 (a) Schematic representations of a n-p-n transistor and p-n-p transistor, and (b) Symbols for n-p-n and p-n-p transistors.

The biasing of the transistor is done differently for different uses. The transistor can be used in two distinct ways. Basically, it was invented to function as an amplifier, a device which produces an enlarged copy of a signal. But later its use as a switch acquired equal importance. We shall study both these functions and the ways the transistor is biased to achieve these mutually exclusive functions.

First we shall see what gives the transistor its amplifying capabilities. The transistor works as an amplifier, with its emitter -base junction forward biased and the base-collector junction reverse biased. This situation is shown in Fig. 14.28, where V_{CC} and V_{EE} are used for creating the respective biasing. When the transistor is biased in this way it is said to be in active state. We represent the voltage between emitter and base as V_{EB} and that between the collector and the base as V_{CB}. In Fig. 14.28, base is a common terminal for the two power supplies whose other terminals are connected to emitter and collector, respectively. So the two power supplies are represented as V_{EE}, and V_{CC}, respectively. In circuits, where emitter is the common terminal, the power supply between the base and the emitter is represented as V_{BB} and that between collector and emitter as V_{CC}.

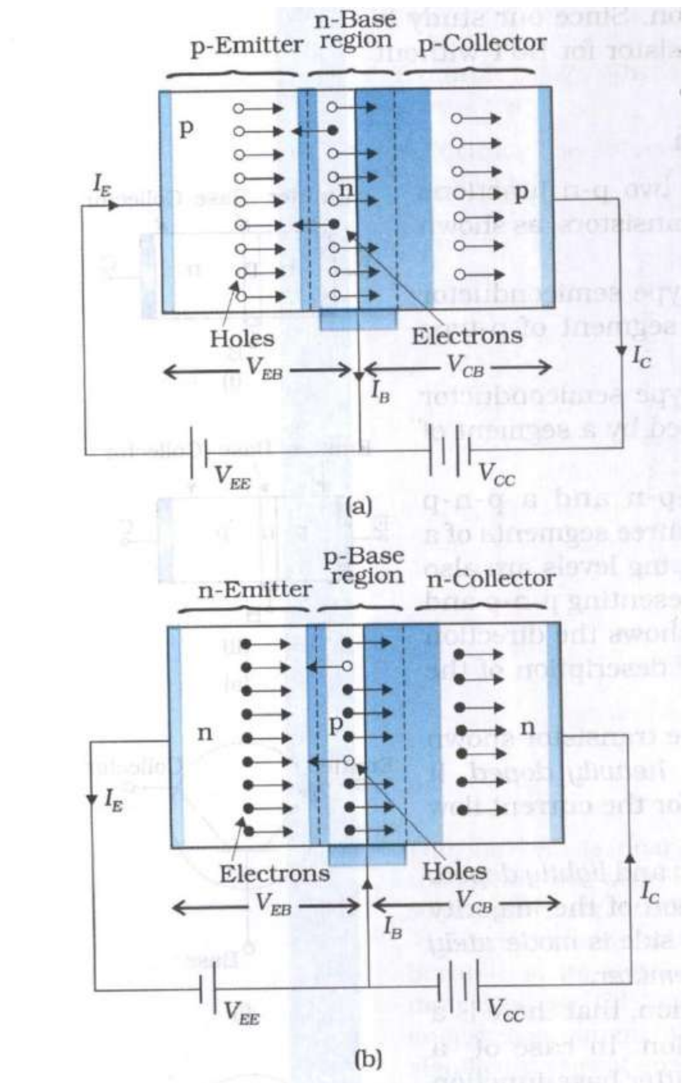

FIGURE 14.28 Bias Voltage applied on: (a) p-n-p transistor and (b) n-p-n transistor.

Let us see now the paths of current carriers in the transistor with emitter-base junction forward biased and base-collector junction reverse biased. The heavily doped emitter has a high concentration of majority carriers, which will be holes in a p-n-p transistor and electrons in an n-p-n transistor. These majority carriers enter the base region in large numbers. The base is thin and lightly doped. So the majority carriers there would be few. In a p-n-p transistor the majority carriers in the base are electrons since base is of n-type semiconductor. The large number of holes entering the base from the emitter swamps the small number of electrons there. As the base collector-junction is reverse biased, these holes, which appear as minority carriers at the junction, can easily cross the junction and enter the collector. The holes in the base could move either towards the base terminal to combine with the electrons entering from outside or cross the junction to enter into the collector and reach the collector terminal. The base is made thin so that most of the holes find themselves near the reverse-biased base-collector junction and so cross the junction instead of moving to the base terminal.

It is interesting to note that due to forward bias a large current enters the emitter-base junction, but most of it is diverted to adjacent reverse-biased base-collector junction and the current coming out of the base becomes a very small fraction of the current that entered the junction. If we represent the hole current and the electron current crossing the forward biased junction by I_h and I_e respectively then the total current in a forward biased diode is the sum $I_h + I_e$. We see that the emitter current $I_E = I_h + I_e$ but the base current $I_B << I_h + I_e$, because a major part of I_E goes to collector instead of coming out of the base terminal. The base current is thus a small fraction of the emitter current.

The current entering into the emitter from outside is equal to the emitter current I_E. Similarly the current emerging from the base terminal is I_B and that from collector terminal is I_C. It is obvious from the above description and also from a straight forward application of Kirchhoff's law to Fig. 14.28(a) that the emitter current is the sum of collector current and base current:

$$I_E = I_C + I_B$$

We also see that $I_C \approx I_E$

Our description of the direction of motion of the holes is identical with the direction of the conventional current. But the direction of motion of electrons is just opposite to that of the current. Thus in a p-n-p transistor the current enters from emitter into base whereas in a n-p-n transistor it enters from the base into the emitter. The arrowhead in the emitter shows the direction of the conventional current.

The description about the paths followed by the majority and minority carriers in a n-p-n is exactly the same as that for the p-n-p transistor. But the current paths are exactly opposite, as shown in Fig. 14.28. In Fig. 14.28(b) the electrons are the majority carriers supplied by the n-type emitter region. They cross the thin p-base region and are able to reach the collector to give the collector current, I_C. From the above description we can conclude that in the active state of the transistor the emitter-base junction acts as a low resistance while the base collector acts as a high resistance.

Basic transistor circuit configuration and transistor characteristics

In a transistor, only three terminals are available, viz., Emitter (E), Base (B) and Collector (C). Therefore, in a circuit the input/output connections have to be such that one of these (E, B or C) is common to both the input and the output. Accordingly, the transistor can be connected in either of the following three configurations:

Common Emitter (CE), Common Base (CB), Common Collector (CC)

The transistor is most widely used in the CE configuration and we shall restrict our discussion to only this configuration. Since more commonly used transistors are n-p-n Si transistors, we shall confine our discussion to such transistors only. With p-n-p transistors the polarities of the external power supplies are to be inverted.

Common emitter transistor characteristics

When a transistor is used in CE configuration, the input is between the base and the emitter and the output is between the collector and the emitter. The variation of the base current I_B with the base-emitter voltage V_{BE} is called the input characteristic. Similarly, the variation of the collector current I_C with the collector-emitter voltage V_{CE} is called the output characteristic. You will see that the output characteristics are controlled by the input characteristics. This implies that the collector current changes with the base current.

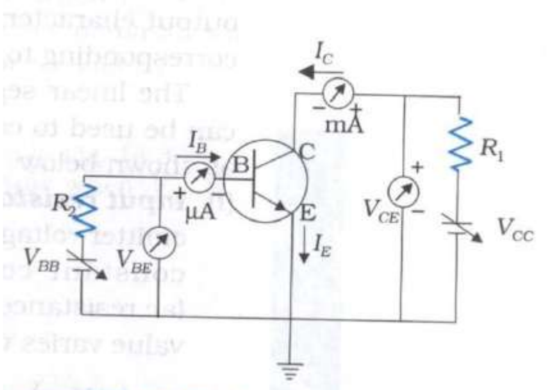

FIGURE 14.29 Circuit arrangement for studying the input and output characteristics of n-p-n transistor in CE configuration.

The input and the output characteristics of an n-p-n transistors can be studied by using the circuit shown in Fig. 14.29.

To study the input characteristics of the transistor in CE configuration, a curve is plotted between the base current I_B against the base-emitter voltage V_{BE}. The collector-emitter voltage V_{CE} is kept fixed while studying the dependence of I_B on

V_{BE}. We are interested to obtain the input characteristic when the transistor is in active state. So the collector -emitter voltage V_{CE} is kept large enough to make the base collector junction reverse biased. Since $V_{CE} = V_{CB} + V_{BE}$ and for Si transistor V_{BE} is 0.6 to 0.7 V, V_{CE} must be sufficiently larger than 0.7 V. Since the transistor is operated as an amplifier over large range of V_{CE}, the reverse bias across the base collector junction is high most of the time. Therefore, the input characteristics may be obtained for V_{CE} somewhere in the range of 3 V to 20 V. Since the increase in V_{CE} appears as increase in V_{CB}, its effect on I_B is negligible. As a consequence, input characteristics for various values of V_{CE} will give almost identical curves. Hence, it is enough to determine only one input characteristics. The input characteristics of a transistor are as shown in Fig. 14.30(a).

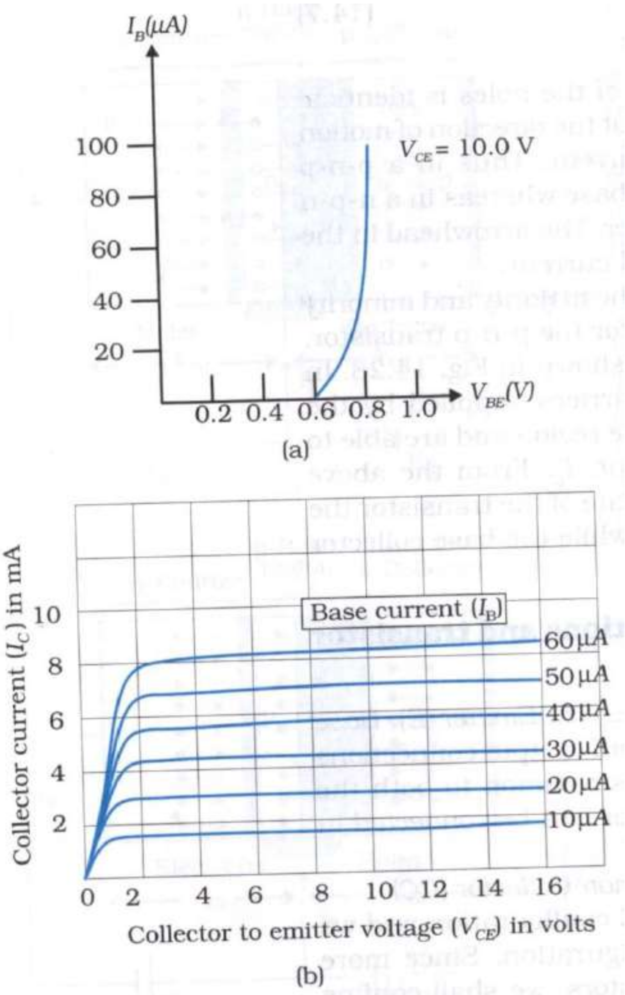

FIGURE 14.30 (a) Typical input characteristics, and (b) Typical output characteristics.

The output characteristic is obtained by observing the variation of I_C as V_{CE} is varied keeping I_B constant. It is obvious that if V_{BE} is increased by a small amount, both hole current from the emitter region and the electron current from the base region will increase. As a consequence both I_B and I_C will increase proportionately. This shows that when I_B increases I_C also increases. The plot of I_C versus V_{CE} for different fixed values of I_B gives one output characteristic. So there will be different output characteristics corresponding to different values of I_B as shown in Fig. 14.30(b).

The linear segments of both the input and output characteristics can be used to calculate some important ac parameters of transistors as shown below.

(i) **Input resistance (r_i):** This is defined as the ratio of change in base-emitter voltage (ΔV_{BE}) to the resulting change in base current (ΔI_B) at constant collector-emitter voltage (V_{CE}). This is dynamic (ac resistance) and as can be seen from the input characteristic, its value varies with the operating current in the transistor:

$$r_i = \left(\frac{\Delta V_{BE}}{\Delta I_B}\right)_{V_{CE}} \qquad (14.8)$$

The value of r_i can be anything from a few hundreds to a few thousand ohms.

(ii) **Output resistance (r_o):** This is defined as the ratio of change in collector-emitter voltage (ΔV_{CE}) to the change in collector current (ΔI_C) at a constant base current I_B.

$$r_o = \left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B} \qquad (14.9)$$

The output characteristics show that initially for very small values of V_{CE}, I_C increases almost linearly. This happens because the base-collector junction is not reverse biased and the transistor is not in active state. In fact, the transistor is in the saturation state and the current is controlled by the supply voltage V_{CC} (=V_{CE}) in this part of the characteristic. When V_{CE} is more than that required to reverse bias the base-collector junction, I_C increases very little with V_{CE}. The reciprocal of the slope of the linear part of the output characteristic gives the values of r_o. The output resistance of the transistor is mainly controlled by the bias of the base-collector junction. The high magnitude of the output resistance (of the order of 100 kΩ) is due to the reverse-biased state of this diode. This also explains why the resistance at the initial part of the characteristic, when the transistor is in saturation state, is very low.

(iii) **Current amplification factor (β):** This is defined as the ratio of the change in collector current to the change in base current at a constant collector-emitter voltage (V_{CE}) when the transistor is in active state.

$$\beta_{ac} = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}} \qquad (14.10)$$

This is also known as small *signal current gain* and its value is very large.

If we simply find the ratio of I_C and I_B we get what is called β_{dc} of the transistor. Hence,

$$\beta_{dc} = \frac{\Delta I_C}{\Delta I_B} \qquad (14.11)$$

Since I_C increases with I_B almost linearly and I_C = 0 when I_B = 0, the values of both β_{dc} and β_{ac} are nearly equal. So, for most calculations β_{dc} can be used. Both β_{ac} and β_{dc} vary with V_{CE} and I_B (or I_C) slightly.

Example From the output characteristics shown in Fig. 14.30(b), calculate the values of β_{ac} and β_{dc} of the transistor when V_{CE} is 10 V and I_C = 4.0 mA.

Solution

$$\beta_{ac} = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}},\ \beta_{dc} = \frac{I_C}{I_B}$$

For determining β_{ac} and β_{dc} at the stated values of V_{CE} and I_C one can proceed as follows. Consider any two characteristics for two values of I_B which lie above and below the given value of I_C . Here I_C = 4.0 mA. (Choose characteristics for I_B= 30 and 20 µA.) At V_{CE} = 10 V we read the two values of I_C from the graph.
Then ΔI_B = (30 – 20) µA = 10 µA, ΔI_C = (4.5 – 3.0) mA = 1.5 mA
Therefore, β_{ac} = 1.5 mA/ 10 µA = 150
For determining β_{dc}, either estimate the value of I_B corresponding to I_C = 4.0 mA at V_{CE} = 10 V or calculate the two values of $\beta_{dc}$for the two characteristics chosen and find their mean.
Therefore, for I_C= 4.5 mA and I_B = 30 µA,
β_{dc} = 4.5 mA/ 30 µA = 150
and for I_C = 3.0 mA and I_B = 20 µA
β_{dc}=3.0 mA / 20 µA = 150
Hence, β_{dc} = (150 + 150) /2 = 150

Transistor as an amplifier (CE-Configuration)

To operate the transistor as an amplifier it is necessary to fix its operating point somewhere in the middle of its active region. If we fix the value of V_{BB} corresponding to a point in the middle of the linear part of the transfer curve then the dc base current I_B would be constant and corresponding collector current I_C will also be constant. The dc voltage $V_{CE} = V_{CC} - I_C R_C$ would also remain constant. The operating values of V_{CE} and I_B determine the operating point, of the amplifier.

If a small sinusoidal voltage with amplitude v_s is superposed on the dc base bias by connecting the source of that signal in series with the V_{BB} supply, then the base current will have sinusoidal variations superimposed on the value of I_B. As a consequence the collector current also will have sinusoidal variations superimposed on the value of I_C, producing in turn corresponding change in the value of V_o. We can measure the ac variations across the input and output terminals by blocking the dc voltages by large capacitors.

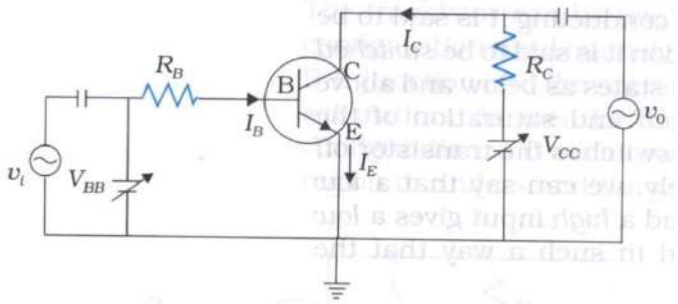

FIGURE 14.32 A simple circuit of a CE-transistor amplifier.

In the description of the amplifier given above we have not considered any ac signal. In general, amplifiers are used to amplify alternating signals. Now let us superimpose an ac input signal v_i (to be amplified) on the bias V_{BB} (dc) as shown in Fig. 14.32. The output is taken between the collector and the ground.

The working of an amplifier can be easily understood, if we first assume that v_i = 0. Then applying Kirchhoff's law to the output loop, we get

$$V_{CC} = V_{CE} + I_C R_L \quad (14.15)$$

Likewise, the input loop gives

$$V_{BB} = V_{BE} + I_B R_B \quad (14.16)$$

When v_i is not zero, we get

$$V_{BE} + v_i = V_{BE} + I_B R_B + \Delta I_B (R_B + r_i)$$

The change in V_{BE} can be related to the input resistance r_i [see Eq. (14.8)] and the change in I_B. Hence

$$v_i = \Delta I_B (R_B + r_i)$$
$$= r\, \Delta I_B$$

The change in I_B causes a change in I_C. We define a parameter β_{ac}, which is similar to the β_{dc} defined in Eq. (14.11), as

$$\beta_{ac} = \frac{\Delta I_C}{\Delta I_B} = \frac{i_c}{i_b} \quad (14.17)$$

which is also known as the *ac current gain* A_i. Usually β_{ac} is close to β_{dc} in the linear region of the output characteristics.

The change in I_C due to a change in I_B causes a change in V_{CE} and the voltage drop across the resistor R_L because V_{CC} is fixed.

These changes can be given by Eq. (14.15) as

$$\Delta V_{CC} = \Delta V_{CE} + R_L \Delta I_C = 0$$

or $\Delta V_{CE} = -R_L \Delta I_C$

The change in V_{CE} is the output voltage v_o. From Eq. (14.10), we get

$$v_o = \Delta V_{CE} = -\beta_{ac} R_L \Delta I_B$$

The voltage gain of the amplifier is

$$A_v = \frac{v_o}{v_i} = \frac{\Delta V_{CE}}{r \Delta I_B}$$
$$= -\frac{\beta_{ac} R_L}{r} \quad (14.18)$$

The negative sign represents that output voltage is opposite with phase with the input voltage.

From the discussion of the transistor characteristics you have seen that there is a current gain β_{ac} in the CE configuration. Here we have also seen the voltage gain A_v. Therefore the power gain A_p can be expressed as the product of the current gain and voltage gain. Mathematically

$$A_p = \beta_{ac} \times A_v \quad (14.19)$$

Since β_{ac} and A_v are greater than 1, we get ac power gain. However it should be realised that transistor is not a power generating device. The energy for the higher ac power at the output is supplied by the battery.

Example For a CE transistor amplifier, the audio signal voltage across the collector resistance of 2.0 kΩ is 2.0 V. Suppose the current amplification factor of the transistor is 100, what should be the value of R_B in series with V_{BB} supply of 2.0 V if the dc base current has to be 10 times the signal current? Also calculate the dc drop across the collector resistance. (Refer to Fig. 14.33).

Solution

The output ac voltage is 2.0 V. So, the ac collector current i_C= 2.0/2000 = 1.0 mA. The signal current through the base is, therefore given by $i_B = i_C/\beta$ = 1.0 mA/100 = 0.010 mA. The dc base current has to be 10× 0.010 = 0.10 mA.

From Eq.14.16, $R_B = (V_{BB} - V_{BE})/I_B$. Assuming V_{BE}= 0.6 V, R_B = (2.0 – 0.6)/0.10 = 14 kΩ.

The dc collector current I_C = 100×0.10 = 10 mA.

Feedback amplifier and transistor oscillator

In an amplifier, we have seen that a sinusoidal input is given which appears as an amplified signal in the output. This means that an external input is necessary to sustain ac signal in the output for an amplifier. In an oscillator, we get ac output without any external input signal. In other words, the output in an oscillator is self-sustained. To attain this, an amplifier is taken. A portion of the output power is returned back (feedback) to the input in phase with the starting power (this process is termed positive feedback) as shown in Fig. 14.33(a). The feedback can be achieved by inductive coupling (through mutual inductance) or LC or RC networks. Different types of oscillators essentially use different methods of coupling the output to the input (feedback network), apart from the resonant circuit for obtaining oscillation at a particular frequency. For understanding the oscillator action, we consider the circuit shown in Fig. 14.33(b) in which the feedback is accomplished by inductive coupling from one coil winding (T_1) to another coil winding (T_2). Note that the coils T_2 and T_1 are wound on the same core and hence are inductively coupled through their mutual inductance. As in an amplifier, the base-emitter junction is forward biased while the base-collector junction is reverse biased. Detailed biasing circuits actually used have been omitted for simplicity.

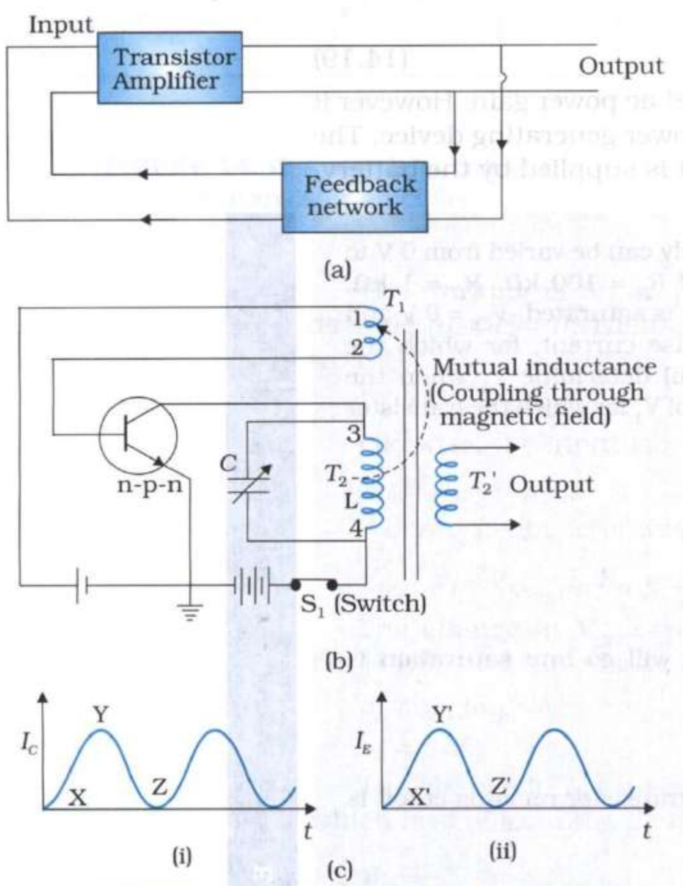

FIGURE 14.33 (a) Principle of a transistor amplifier with positive feedback working as an oscillator and (b) Tuned collector oscillator, (c) Rise and fall (or built up) of current I_c and I_e due to the inductive coupling.

500

Let us try to understand how oscillations are built. Suppose switch S_1 is put on to apply proper bias for the first time. Obviously, a surge of collector current flows in the transistor. This current flows through the coil T_2 where terminals are numbered 3 and 4 [Fig. 14.33(b)]. This current does not reach full amplitude instantaneously but increases from X to Y, as shown in Fig. [14.33(c)(i)]. The inductive coupling between coil T_2 and coil T_1 now causes a current to flow in the emitter circuit (note that this actually is the ‘feedback’ from input to output). As a result of this positive feedback, this current (in T_1 ; emitter current) also increases from X´ to Y´ [Fig. 14.33(c)(ii)]. The current in T_2 (collector current) connected in the collector circuit acquires the value Y when the transistor becomes saturated. This means that maximum collector current is flowing and can increase no further. Since there is no further change in collector current, the magnetic field around T_2 ceases to grow. As soon as the field becomes static, there will be no further feedback from T_2 to T_1 . Without continued feedback, the emitter current begins to fall. Consequently, collector current decreases from Y towards Z [Fig. 14.33(c)(i)]. However, a decrease of collector current causes the magnetic field to decay around the coil T_2 . Thus, T_1 is now seeing a decaying field in T_2 (opposite from what it saw when the field was growing at the initial start operation). This causes a further decrease in the emitter current till it reaches Z′ when the transistor is cut-off. This means that both I_E and I_C cease to flow. Therefore, the transistor has reverted back to its original state (when the power was first switched on). The whole process now repeats itself. That is, the transistor is driven to saturation, then to cut-off, and then back to saturation. The time for change from saturation to cut-off and back is determined by the constants of the tank circuit or tuned circuit (inductance L of coil T_2 and C connected in parallel to it). The resonance frequency (ν) of this tuned circuit determines the frequency at which the oscillator will oscillate.

$$v = \frac{1}{2\pi\sqrt{LC}} \qquad (14.20)$$

In the circuit of Fig. 14.33(b), the tank or tuned circuit is connected in the collector side. Hence, it is known as tuned collector oscillator. If the tuned circuit is on the base side, it will be known as tuned base oscillator. There are many other types of tank circuits (say RC) or feedback circuits giving different types of oscillators like Colpitt’s oscillator, Hartley oscillator, RC-oscillator.

POINTS TO PONDER

1. In transistors, the base region is both narrow and lightly doped, otherwise the electrons or holes coming from the input side (say, emitter in CE-configuration) will not be able to reach the collector.
2. We have described an oscillator as a positive feedback amplifier. For stable oscillations, the voltage feedback (V_{fb}) from the output voltage (V_o) should be such that after amplification (A) it should again become V_o. If a fraction β′ is feedback, then V_{fb}= V_o. β′ and after amplification its value A(v_o. β′) should be equal to V_o. This means that the criteria for stable oscillations to be sustained is Aβ′ = 1. This is known as Barkhausen’s Criteria.
3. In an oscillator, the feedback is in the same phase (positive feedback). If the feedback voltage is in opposite phase (negative feedback), the gain is less than 1 and it can never work as oscillator. It will be an amplifier with reduced gain. However, the negative feedback also reduces noise and distortion in an amplifier which is an advantageous feature.

EXERCISES

1. For transistor action, which of the following statements are correct:

   (a) Base, emitter and collector regions should have similar size and doping concentrations.
   (b) The base region must be very thin and lightly doped.
   (c) The emitter junction is forward biased and collector junction is reverse biased.
   (d) Both the emitter junction as well as the collector junction are forward biased.

2. For a transistor amplifier, the voltage gain

   (a) remains constant for all frequencies.
   (b) is high at high and low frequencies and constant in the middle frequency range.
   (c) is low at high and low frequencies and constant at mid frequencies.
   (d) None of the above.

3. For a CE-transistor amplifier, the audio signal voltage across the collected resistance of 2 kΩ is 2 V. Suppose the current amplification factor of the transistor is 100, find the input signal voltage and base current, if the base resistance is 1 kΩ.
4. Two amplifiers are connected one after the other in series (cascaded). The first amplifier has a voltage gain of 10 and the second has a voltage gain of 20. If the input signal is 0.01 volt, calculate the output ac signal.
5. Two amplifiers are connected one after the other in series (cascaded). The first amplifier has a voltage gain of 10 and the second has a voltage gain of 20. If the input signal is 0.01 volt, calculate the output ac signal.

COMMUNICATION SYSTEMS

INTRODUCTION

Communication is the act of transmission of information. Every living creature in the world experiences the need to impart or receive information almost continuously with others in the surrounding world. For communication to be successful, it is essential that the sender and the receiver understand a common language. Man has constantly made endeavors to improve the quality of communication with other human beings. Languages and methods used in communication have kept evolving from prehistoric to modern times, to meet the growing demands in terms of speed and complexity of information. It would be worthwhile to look at the major milestones in events that promoted developments in communications, as presented in Table 15.1

Modern communication has its roots in the 19^{th} and 20^{th} century in the work of scientists like J.C. Bose, F.B. Morse, G. Marconi and Alexander Graham Bell. The pace of development seems to have increased dramatically after the first half of the 20^{th} century. We can hope to see many more accomplishments in the coming decades. The aim of this chapter is to introduce the concepts of communication, namely the mode of communication, the need for modulation, production and detection of amplitude modulation.

ELEMENTS OF A COMMUNICATION SYSTEM

Communication pervades all stages of life of all living creatures. Irrespective of its nature, every communication system has three essential elements - transmitter, medium/channel and receiver. The block diagram shown in fig. 15.1 depicts the general form of a communication system.

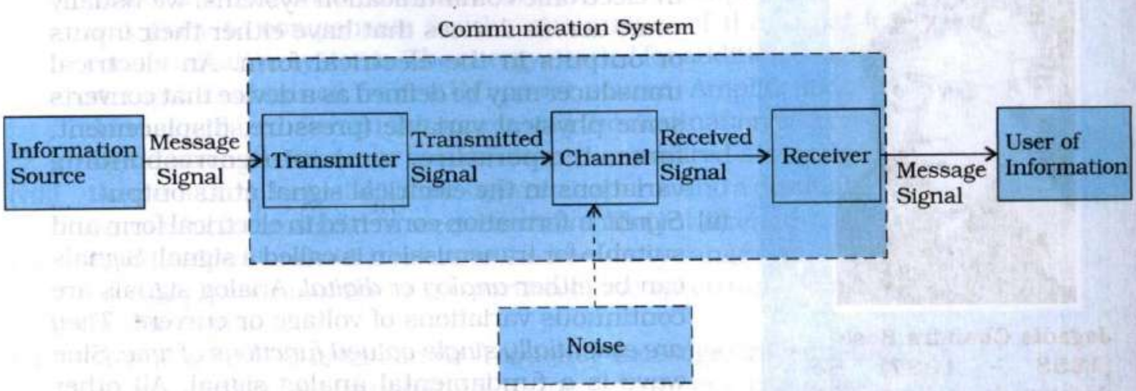

FIGURE 15.1 Block diagram of a generalised communication system.

In communication system, the transmitter is located at one place, the receiver is located at some other place (far or near) separate from the transmitter and the channel is the physical medium that connects them. Depending upon the type of communication system, a channel may be in the form of wires or cables connecting the transmitter and the receiver or it may be wireless. The purpose of the transmitter is to convert the message signal produced by the source of information into a form suitable for transmission through the channel. If the output of the information source is a non-electrical signal like a voice signal, a transducer converts it to electrical form before giving it as an input to the transmitter. When a transmitted signal propagates along the channel it may get distorted due to channel imperfection. Moreover, noise adds to the transmitted signal and the receiver receives a corrupted version of the transmitted signal. The receiver has the task of operating on the received signal. It reconstructs a recognizable form of the original message signal for delivering it to the user of information.

There are two basic modes of communication: *point-to-point* and *broadcast*.

In point-to-point communication mode, a communication takes place over a link between a single transmitter and a receiver. Telephony is an example of such a mode of communication. In contrast, in the broadcast mode, there are a large number of receivers corresponding to a single transmitter. Radio and television are examples of broadcast mode of communication.

Table 15.1 Some major milestones in the history of communication

Year	Event	Remarks
Around 1565 A.D.	The reporting of the delivery of a child by queen using drum beats from a distant place to King Akbar.	It is believed that minister Birbal experimented with the arrangement to decide the number of drummers posted between the place where the queen stayed and the place where the king stayed.
1835	Invention of telegraph by Samuel F.B. Morse and Sir Charles Wheatstone	It resulted in tremendous growth of messages through post offices and reduced physical travel of messengers considerably.
1876	Telephone invented by Alexander Graham Bell and Antonio Meucci	Perhaps the most widely used means of communication in the history of mankind.
1895	Jagadis Chandra Bose and Guglielmo Marconi demonstrated wireless telegraphy.	It meant a giant leap – from an era of communication using wires to communicating without using wires. (wireless)
1936	Television broadcast(John Logi Baird)	First television broadcast by BBC
1955	First radio FAX transmitted across continent.(Alexander Bain)	The idea of FAX transmission was patented by Alexander Bain in 1843.
1968	ARPANET- the first internet came into existence(J.C.R. Licklider)	ARPANET was a project undertaken by the U.S. defence department. It allowed file transfer from one computer to another connected to the network.
1975	Fiber optics developed at Bell Laboratories	Fiber optical systems are superior and more economical compared to traditional communication systems.
1989-91	Tim Berners-Lee invented the **World Wide Web**.	WWW may be regarded as the mammoth encyclopedia of knowledge accessible to everyone round the clock throughout the year.

BASIC TERMINOLOGY USED IN ELECTRONIC COMMUNICATION SYSTEMS

By now, we have become familiar with some terms like information source, transmitter, receiver, channel, noise, etc. It would be easy to understand the principles underlying any communication, if we get ourselves acquainted with the following basic terminology.

i. *Transducer:* Any device that converts one form of energy into another can be termed as a transducer. In electronic communication systems, we usually come across devices that have either their inputs or outputs in the electrical form. An electrical transducer may be defined as a device that converts some physical variable (pressure, displacement, force, temperature, etc.) into corresponding variations in the electrical signal at its output.

ii. *Signal*: Information converted in electrical form and suitable for transmission is called a signal. Signals can be either analog or digital. Analog signals are continuous variations of voltage or current. They are essentially single-valued functions of time. Sine wave is a fundamental analog signal. All other analog signals can be fully understood in terms of their sine wave components. Sound and picture signal in TV are analog in nature. Digital signals are those which can take only discrete stepwise values. Binary system that is extensively used in digital electronics employs just two levels of a signal. '0' corresponds to a low level and '1' corresponds to a high level of voltage/current. There are several coding schemes useful for digital communication. They employ suitable combinations of number systems such as the binary coded decimal (BCD)*. American Standard Code for Information Interchange (ASCII)** is a universally popular digital code to represent numbers, letters and certain characters. (Nowadays, optical signal are also in use.)

iii. *Noise*: Noise refers to the unwanted signals that tend to disturb the transmission and processing of message signals in a communication system. The source generating the noise may be located inside or outside the system.

iv. *Transmitter*: A transmitter processes the incoming message signal so as to make it suitable for transmission through a channel and subsequent reception.

v. *Receiver*: A receiver extracts the desired message signals from the received signals at the channel output.

vi. *Attenuation*: The loss of strength of a signal while propagating through a medium is known as attenuation.

vii. *Amplification*: It is the process of increasing the amplitude (and consequently the strength) of a signal using an electronic circuit called the amplifier (reference Chapter 14). Amplification is necessary to compensate for the attenuation of the signal in communication systems. The energy needed for additional signal strength is obtained from a DC power source. Amplification is done at a place between the source and the destination wherever signal strength becomes weaker than the required strength.

viii. *Range*: It is the largest distance between a source and a destination up to which the signal is received with sufficient strength.

ix. *Bandwidth*: Bandwidth refers to the frequency range over which an equipment operates or the portion of the spectrum occupied by the signal.

x. *Modulation*: The original low frequency message/information signal cannot be transmitted to long distances because of reasons given in Section 15.7. Therefore, at the transmitter, information contained in the low frequency message signal is superimposed on a high frequency wave, which acts as a carrier of the information. This process is known as modulation. As will be explained later, there are several types of modulation, abbreviated as AM, FM and PM.

** In BCD, a digit is usually represented by four binary (0 or 1) bits. For example the numbers 0, 1, 2, 3, 4 in the decimal system are written as 0000, 0001, 0010, 0011 and 0100. 1000 would represent eight.*

*** It is a character encoding in terms of numbers based on English alphabet since the computer can only understand numbers.*

xi. *Demodulation*: The process of retrieval of information from the carrier wave at the receiver is termed demodulation. This is the reverse process of modulation.

xii. *Repeater*: A repeater is a combination of a receiver and a transmitter. A repeater, picks up the signal from the transmitter, amplifies and retransmits it to the receiver sometimes with a change in carrier frequency. Repeaters are used to extend the range of a communication system as shown in Fig. 15.2. A communication satellite is essentially a repeater station in space.

FIGURE 15.2 Use of repeater station to increase the range of communication.

BANDWIDTH OF SIGNALS

In a communication system, the message signal can be voice, music, picture or computer data. Each of these signals has different ranges of frequencies. The type of communication system needed for a given signal depends on the band of frequencies which is considered essential for the communication process.

For speech signals, frequency range 300 Hz to 3100 Hz is considered adequate. Therefore speech signal requires a bandwidth of 2800 Hz (3100 Hz – 300 Hz) for commercial telephonic communication. To transmit music, an approximate bandwidth of 20 kHz is required because of the high frequencies produced by the musical instruments. The audible range of frequencies extends from 20 Hz to 20 kHz.

Video signals for transmission of pictures require about 4.2 MHz of bandwidth. A TV signal contains both voice and picture and is usually allocated 6 MHz of bandwidth for transmission.

In the preceeding paragraph, we have considered only analog signals. Digital signals are in the form of rectangular waves as shown in Fig. 15.3. One can show that this rectangular wave can be decomposed into a superposition of sinusoidal waves of frequencies v_0, $2v_0$, $3v_0$, $4v_0$ ... nv_0 where n is an integer extending to infinity and $v_0 = 1/T_0$. The fundamental (v_0), fundamental (v_0) + second harmonic (v_0), and fundamental (v_0) + second harmonic ($2v_0$) + third harmonic ($3v_0$), are shown in the same figure to illustrate this fact. It is clear that to reproduce the rectangular wave shape exactly we need to superimpose all the harmonics v_0, $2v_0$, $3v_0$, $4v_0$ ..., which implies an infinite bandwidth. However, for practical purposes, the contribution from higher harmonics can be neglected, thus limiting the bandwidth. As a result, received waves are a distorted version of the transmitted one. If the bandwidth is large enough to accommodate a few harmonics, the information is not lost and the rectangular signal is more or less recovered. This is so because the higher the harmonic, less is its contribution to the wave form.

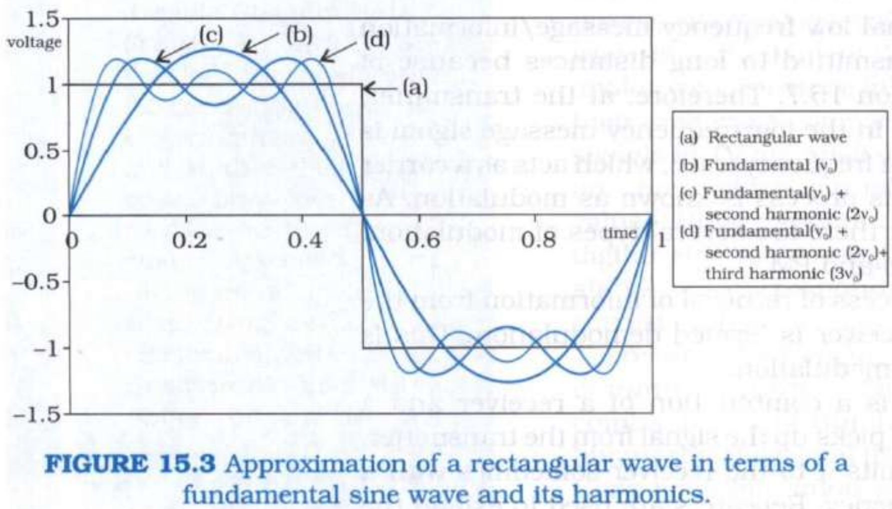

FIGURE 15.3 Approximation of a rectangular wave in terms of a fundamental sine wave and its harmonics.

BANDWIDTH OF TRANSMISSION MEDIUM

Similar to message signals, different types of transmission media offer different bandwidths. The commonly used transmission media are wire, free space and fiber optic cable. Coaxial cable is a widely used wire medium, which offers a bandwidth of approximately 750 MHz. Such cables are normally operated below 18 GHz. Communication through free space using radio waves takes place over a very wide range of frequencies: from a few hundreds of kHz to a few GHz. This range of frequencies is further subdivided and allocated for various services as indicated in Table 15.2. Optical communication using fibers is performed in the frequency range of 1 THz to 1000 THz (microwaves to ultraviolet). An optical fiber can offer a transmission bandwidth in excess of 100 GHz.

Spectrum allocations are arrived at by an international agreement. The International Telecommunication Union (ITU) administers the present system of frequency allocations.

TABLE 15.2 SOME IMPORTANT WIRELESS COMMUNICATION FREQUENCY BANDS

Service	Frequency bands	Comments
Standard AM broadcast	540-1600 kHz	
FM broadcast	88-108 MHz	
Television	54-72 MHz 76-88 MHz 174-216 MHz 420-890 MHz	VHF (very high frequencies) TV UHF (ultra high frequencies) TV
Cellular Mobile Radio	896-901 MHz 840-935 MHz	Mobile to base station Base station to mobile
Satellite Communication	5.925-6.425 GHz 3.7-4.2 GHz	Uplink Downlink

PROPAGATION OF ELECTROMAGNETIC WAVES

In communication using radio waves, an antenna at the transmitter radiates the Electromagnetic waves (em waves), which travel through the space and reach the receiving antenna at the other end. As the em wave travels away from the transmitter, the strength of the wave keeps on decreasing. Several factors influence the propagation of em waves and the path they follow. At this point, it is also important to understand the composition of the earth's atmosphere as it plays a vital role in the propagation of em waves. A brief discussion on some useful layers of the atmosphere is given in Table 15.3.

TABLE 15.3 DIFFERENT LAYERS OF ATMOSPHERE AND THEIR INTERACTION WITH THE PROPAGATING ELECTROMAGNETIC WAVES

Name of the stratum (layer)		Approximate height over earth's surface	Exists during	Frequencies most affected
Troposphere		10 km	Day and night	VHF (up to several GHz)
D (part of stratosphere)	PARTS OF IONOSPHERE	65-75 km	Day only	Reflects LF, absorbs MF and HF to some degree
E (part of Stratosphere)		100 km	Day only	Helps surface waves, reflects HF
F_1 (Part of Mesosphere)		170-190 km	Daytime, merges with F_2 at night	Partially absorbs HF waves yet allowing them to reach F_2
F_2 (Thermosphere)		300 km at night, 250-400 km during daytime	Day and night	Efficiently reflects HF waves, particularly at night

Ground Wave

To radiate signals with high efficiency, the antennas should have a size comparable to the wavelength λ of the signal (at least ~ λ/4). At longer wavelengths (i.e., at lower frequencies), the antennas have large physical size and they are located on or very near to the ground. In standard AM broadcast, ground based vertical towers are generally used as transmitting antennas. For such antennas, ground has a strong influence on the propagation of the signal. The mode of propagation is called surface wave propagation and the wave glides over the surface of the earth. A wave induces current in the ground over which it passes and it is attenuated as a result of absorption of energy by the earth. The attenuation of surface waves increases very rapidly with increase in frequency. The maximum range of coverage depends on the transmitted power and frequency (less than a few MHz).

Sky Waves

In the frequency range from a few MHz up to 30 to 40 MHz, long distance communication can be achieved by ionospheric reflection of radio waves back towards the earth. This mode of propagation is called sky wave propagation and is used by short wave broadcast services. The ionosphere is so called because of the presence of a large number of ions or charged particles. It extends from a height of ~ 65 Km to about 400 Km above the earth's surface. Ionisation occurs due to the absorption of the ultraviolet and other high-energy radiation coming from the sun by air molecules. The ionosphere is further subdivided into several layers, the details of which are given in Table 15.3. The degree of ionisation varies with the height. The density of atmosphere decreases with height. At great heights the solar radiation is intense but there are few molecules to be ionised. Close to the earth, even though the molecular concentration is very high, the radiation intensity is low so that the ionisation is again low. However, at some intermediate heights, there occurs a peak of ionisation density. The ionospheric layer acts as a reflector for a certain range of frequencies (3 to 30 MHz). Electromagnetic waves of frequencies higher than 30 MHz penetrate the ionosphere and escape. These phenomena are shown in the Fig. 15.4. The phenomenon of bending of em waves so that they are diverted towards the earth is similar to total internal reflection in optics*.

** Compare this with the phenomenon of mirage.*

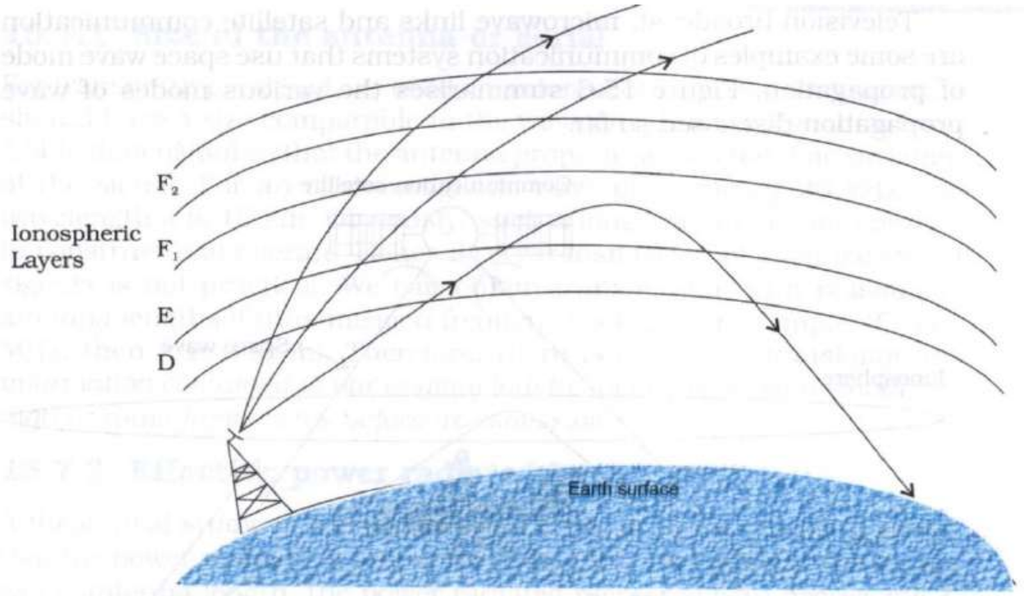

FIGURE 15.4 Sky wave propagation. The layer nomenclature is given in Table 15.3.

Space Wave

Another mode of radio wave propagation is by space waves. A space wave travels in a straight line from transmitting antenna to the receiving antenna. Space waves are used for line-of-sight (LOS) communication as well as satellite communication. At frequencies above 40 MHz, communication is essentially limited to line-of-sight paths. At these frequencies, the antennas are relatively smaller and can be placed at heights of many wavelengths above the ground. Because of line-of-sight nature of propagation, direct waves get blocked at some point by the curvature of the earth as illustrated in Fig. 15.5. If the signal is to be received beyond the horizon then the receiving antenna must be high enough to intercept the line-of-sight waves.

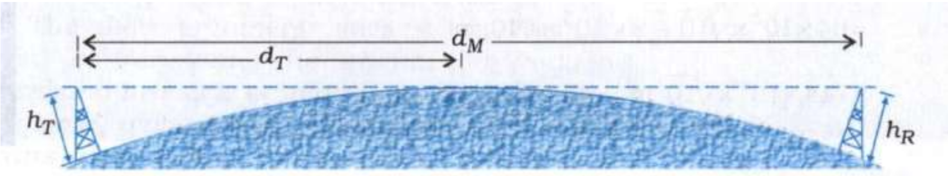

FIGURE 15.5 Line of sight communication by space waves.

If the transmitting antenna is at a height h_T, then you can show that the distance to the horizon d_T is given as $d_T = \sqrt{2Rh_T}$, where R is the radius of the earth (approximately 6400 km). d_T is also called the radio horizon of the transmitting antenna. With reference to Fig. 15.5 the maximum line-of-sight distance d_M between the two antennas having heights h_T and h_R above the earth is given by

$$d_M = \sqrt{2Rh_T} + \sqrt{2Rh_R} \qquad (15.1)$$

where h_R is the height of receiving antenna.

Television broadcast, microwave links and satellite communication are some examples of communication systems that use space wave mode of propagation. Figure 15.6 summarises the various modes of wave propagation discussed so far.

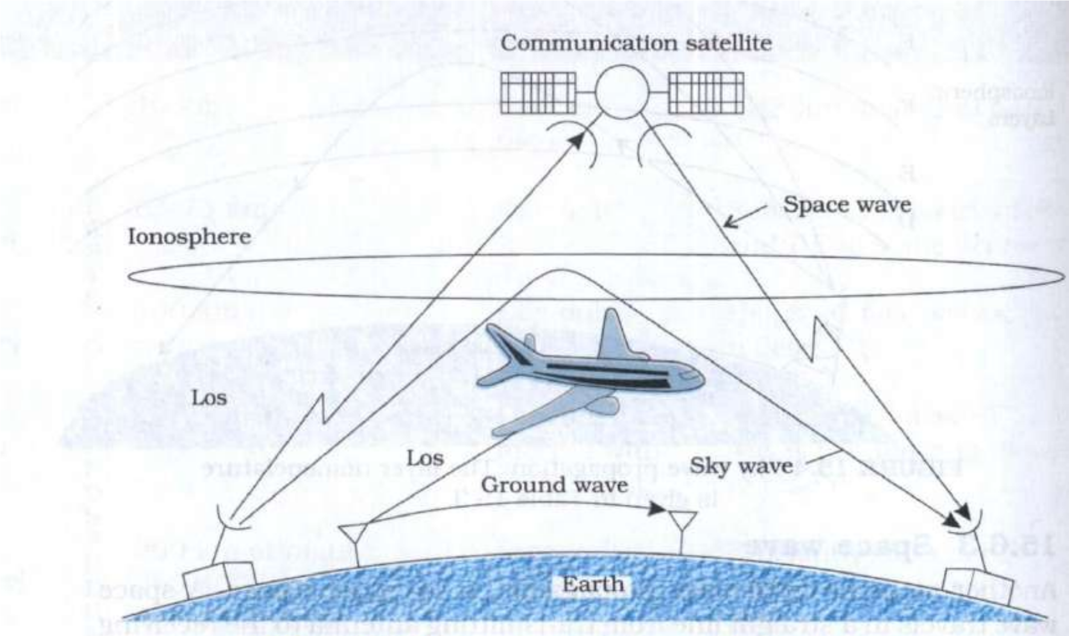

FIGURE 15.6 Various propagation modes for em waves.

MODULATION AND ITS NECESSITY

As already mentioned, the purpose of a communication system is to transmit information or message signals. Message signals are also called baseband signals, which essentially designate the band of frequencies representing the original signal, as delivered by the source of information. No signal, in general, is a single frequency sinusoid, but it spreads over a range of frequencies called the signal bandwidth. Suppose we wish to transmit an electronic signal in the audio frequency (AF) range (baseband signal frequency less than 20 kHz) over a long distance directly. Let us find what factors prevent us from doing so and how we overcome these factors.

Size of the antenna or aerial

For transmitting a signal, we need an antenna or an aerial. This antenna should have a size comparable to the wavelength of the signal (at least λ/4 in dimension) so that the antenna properly senses the time variation of the signal. For an electromagnetic wave of frequency 20 kHz, the wavelength λ is 15 km. Obviously, such a long antenna is not possible to construct and operate. Hence direct transmission of such baseband signals is not practical. We can obtain transmission with reasonable antenna lengths if transmission frequency is high (for example, if ν is 1 MHz, then λ is 300 m). Therefore, there is a need of *translating the information contained in our original low frequency baseband signal into high or radio frequencies before transmission.*

Effective power radiated by an antenna

A theoretical study of radiation from a linear antenna (length *l*) shows that the power radiated is proportional to $(l/\lambda)^2$. This implies that for the same antenna length, the power radiated increases with decreasing λ, i.e., increasing frequency. Hence, the effective power radiated by a long wavelength baseband signal would be small. For a good transmission, we need high powers and hence this also points out to the need of using high frequency transmission.

Mixing up of signals from different transmitters

Another important argument against transmitting baseband signals directly is more practical in nature. Suppose many people are talking at the same time or many transmitters are transmitting baseband information signals simultaneously. All these signals will get mixed up and there is no simple way to distinguish between them. This points out towards a possible solution by using communication at high frequencies and allotting a band of frequencies to each message signal for its transmission.

The above arguments suggest that there is a *need for translating the original low frequency baseband message or information signal into high frequency wave before transmission such that the translated signal continues to possess the information contained in the original signal.* In doing so, we take the help of a high frequency signal, known as the carrier wave, and a process known as modulation which attaches information to it. The carrier wave may be continuous (sinusoidal) or in the form of pulses as shown in Fig. 15.7.

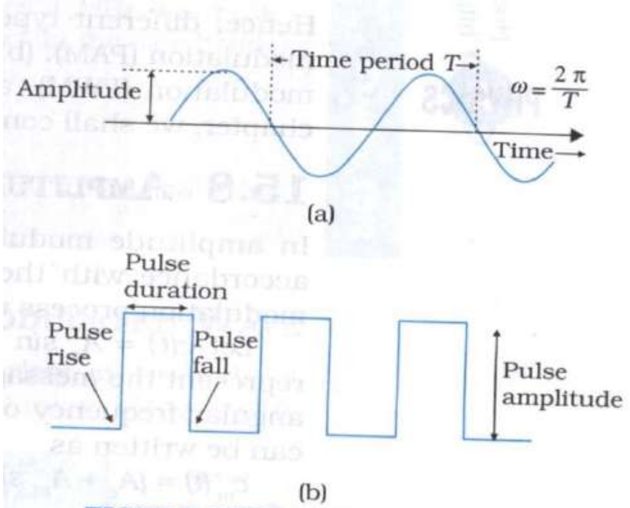

FIGURE 15.7 (a) Sinusoidal, and (b) pulse shaped signals.

A sinusoidal carrier wave can be represented as

$$c(t) = A_c \sin(\omega_c t + \phi) \quad (15.2)$$

where c(t) is the signal strength (voltage or current), A_c is the amplitude, ω_c ($= 2\pi v_c$) is the angular frequency and ϕ is the initial phase of the carrier wave. During the process of modulation, any of the three parameters, viz A_c, ω_c and ϕ, of the carrier wave can be controlled by the message or information signal. This results in three types of modulation: (i) Amplitude modulation (AM), (ii) Frequency modulation (FM) and (iii) Phase modulation (PM), as shown in Fig. 15.8.

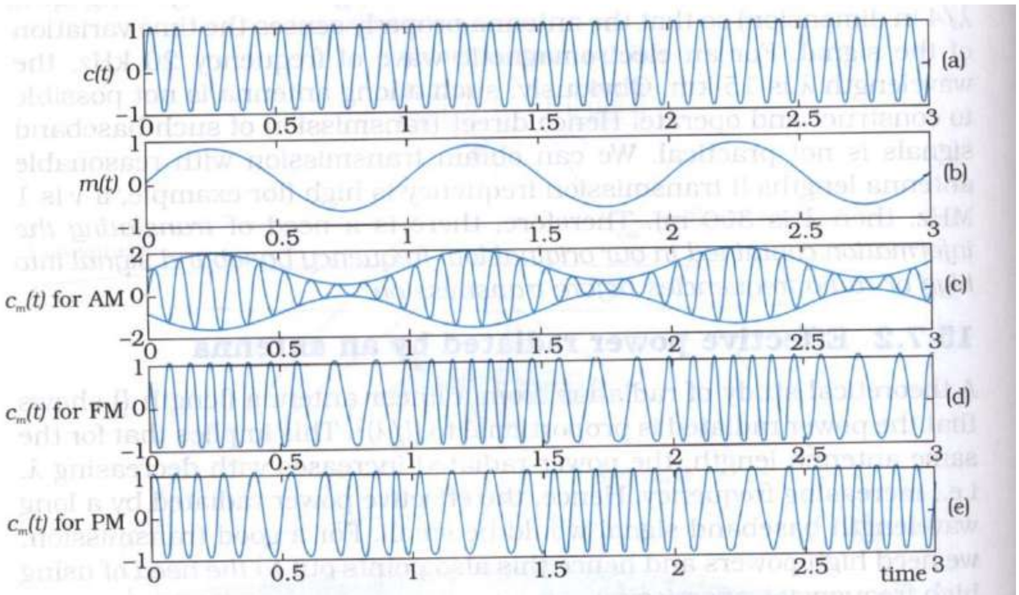

FIGURE 15.8 Modulation of a carrier wave: (a) a sinusoidal carrier wave; (b) a modulating signal; (c) amplitude modulation; (d) frequency modulation; and (e) phase modulation.

Similarly, the significant characteristics of a pulse are: pulse amplitude, pulse duration or pulse Width, and pulse position (denoting the time of rise or fall of the pulse amplitude) as shown in Fig. 15.7(b). Hence, different types of pulse modulation are: (a) pulse amplitude modulation (PAM), (b) pulse duration modulation (PDM) or pulse width modulation (PWM), and (c) pulse position modulation (PPM). In this chapter, we shall confine to amplitude modulation only.

AMPLITUDE MODULATION

In amplitude modulation the amplitude of the carrier is varied in accordance with the information signal. Here we explain amplitude modulation process using a sinusoidal signal as the modulating signal.

Let $c(t) = A_c \sin \omega_c t$ represent carrier wave and $m(t) = A_m \sin \omega_m t$ represent the message or the modulating signal where $\omega_m = 2\pi f_m$is the angular frequency of the message signal. The modulated signal $c_m(t)$ can be written as

$$c_m(t) = (A_c + A_m \sin \omega_m t) \sin \omega_c t$$

$$= A_c(1 + \frac{A_m}{A_c} \sin \omega_m t) \sin \omega_c t \quad (15.3)$$

Note that the modulated signal now contains the message signal. This can also be seen from Fig. 15.8(c). From Eq. (15.3), we can write,

$$c_m(t) = A_c \sin \omega_c t + \mu A_c \sin \omega_m t \sin \omega_c t \quad (15.4)$$

Here $\mu = A_m/A_c$ is the modulation index; in practice, μ is kept ≤1 to avoid distortion.

Using the trigonometric relation sinA sinB = ½ (cos(A – B) – cos (A + B)), we can write $c_m(t)$ of Eq. (15.4) as

$$c_m(t) = A_c \sin \omega_c t + \frac{\mu A_c}{2} \cos(\omega_c - \omega_m)t - \frac{\mu A_c}{2} \cos(\omega_c + \omega_m)t \quad (15.5)$$

Here $\omega_c - \omega_m$ and $\omega_c + \omega_m$are respectively called the lower side and upper side frequencies. The modulated signal now consists of the carrier wave of frequency ω_c plus two sinusoidal waves each with a frequency slightly different from, known as side bands. The frequency spectrum of the amplitude modulated signal is shown in Fig. 15.9.

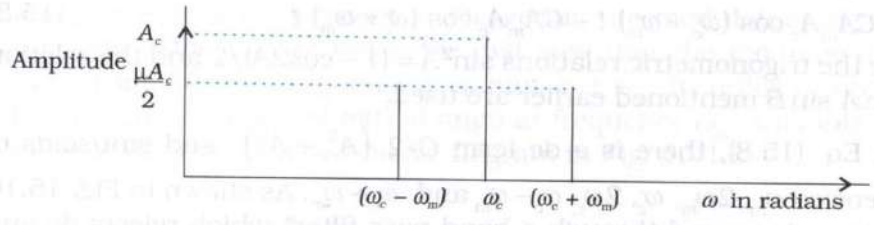

FIGURE 15.9 A plot of amplitude versus ω for an amplitude modulated signal.

As long as the broadcast frequencies (carrier waves) are sufficiently spaced out so that sidebands do not overlap, different stations can operate without interfering with each other.

ADDITIONAL INFORMATION

The Internet

It is a system with billions of users worldwide. It permits communication and sharing of all types of information between any two or more computers connected through a large and complex network. It was started in 1960's and opened for public use in 1990's. With the passage of time it has witnessed tremendous growth and it is still expanding its reach. Its applications include

(i) *E mail* – It permits exchange of text/graphic material using email software. We can write a letter and send it to the recipient through ISP's (Internet Service Providers) who work like the dispatching and receiving post offices.

(ii) *File transfer* – A FTP (File Transfer Programmes) allows transfer of files/software from one computer to another connected to the Internet.

(iii) *World Wide Web (WWW)* – Computers that store specific information for sharing with others provide websites either directly or through web service providers. Government departments, companies, NGO's (Non-Government Organisations) and individuals can post information about their activities for restricted or free use on their websites. This information becomes accessible to the users. Several search engines like Google, Yahoo! etc. help us in finding information by listing the related websites. Hypertext is a powerful feature of the web that automatically links relevant information from one page on the web to another using HTML (hypertext markup language).

(iv) *E-commerce* – Use of the Internet to promote business using electronic means such as using credit cards is called E-commerce. Customers view images and receive all the information about various products or services of companies through their websites. They can do on-line shopping from home/office. Goods are dispatched or services are provided by the company through mail/courier.

(v) *Chat* – Real time conversation among people with common interests through typed messages is called chat. Everyone belonging to the chat group gets the message instantaneously and can respond rapidly.

Facsimile (FAX)

It scans the contents of a document (as an image, not text) to create electronic signals. These signals are then sent to the destination (another FAX machine) in an orderly manner using telephone lines. At the destination, the signals are reconverted into a replica of the original document. Note that FAX provides image of a static document unlike the image provided by television of objects that might be dynamic.

Mobile telephony

The concept of mobile telephony was developed first in 1970's and it was fully implemented in the following decade. The central concept of this system is to divide the service area into a suitable number of cells centred on an office called MTSO (Mobile Telephone Switching Office). Each cell contains a low-power transmitter called a base station and caters to a large number of mobile receivers (popularly called cell phones). Each cell could have a service area of a few square kilometers or even less depending upon the number of customers. When a mobile receiver crosses the coverage area of one base station, it is necessary for the mobile user to be transferred to another base station. This procedure is called handover or handoff. This process is carried out very rapidly, to the extent that the consumer does not even notice it. Mobile telephones operate typically in the UHF range of frequencies (about 800-950 MHz).

SUMMARY

1. Electronic communication refers to the faithful transfer of information or message (available in the form of electrical voltage and current) from one point to another point.

2. Transmitter, transmission channel and receiver are three basic units of a communication system.

3. Two important forms of communication system are: Analog and Digital. The information to be transmitted is generally in continuous waveform for the former while for the latter it has only discrete or quantised levels.

4. Every message signal occupies a range of frequencies. The bandwidth of a message signal refers to the band of frequencies, which are necessary for satisfactory transmission of the information contained in the signal. Similarly, any practical communication system permits transmission of a range of frequencies only, which is referred to as the bandwidth of the system.

5. Low frequencies cannot be transmitted to long distances. Therefore, they are superimposed on a high frequency carrier signal by a process known as modulation.

6. In modulation, some characteristic of the carrier signal like amplitude, frequency or phase varies in accordance with the modulating or message signal. Correspondingly, they are called Amplitude Modulated (AM), Frequency Modulated (FM) or Phase Modulated (PM) waves.

7. Pulse modulation could be classified as: Pulse Amplitude Modulation (PAM), Pulse Duration Modulation (PDM) or Pulse Width Modulation (PWM) and Pulse Position Modulation (PPM).

8. For transmission over long distances, signals are radiated into space using devices called antennas. The radiated signals propagate as electromagnetic waves and the mode of propagation is influenced by the presence of the earth and its atmosphere. Near the surface of the earth, electromagnetic waves propagate as surface waves. Surface wave propagation is useful up to a few MHz frequencies.

9. Long distance communication between two points on the earth is achieved through reflection of electromagnetic waves by ionosphere. Such waves are called sky waves. Sky wave propagation takes place up to frequency of about 30 MHz. Above this frequency, electromagnetic waves essentially propagate as space waves. Space waves are used for line-of-sight communication and satellite communication.

10. If an antenna radiates electromagnetic waves from a height h_T , then the range d_T is given by $\sqrt{2Rh_T}$ where R is the radius of the earth.

11. Amplitude modulated signal contains frequencies $(\omega_c - \omega_m)$, ω_c and $(\omega_c + \omega_m)$.

12. Amplitude modulated waves can be produced by application of the message signal and the carrier wave to a non-linear device, followed by a band pass filter.

13. AM detection, which is the process of recovering the modulating signal from an AM waveform, is carried out using a rectifier and an envelope detector.

POINTS TO PONDER

1. In the process of transmission of message/ information signal, noise gets added to the signal anywhere between the information source and the receiving end. Can you think of some sources of noise?

2. In the process of modulation, new frequencies called sidebands are generated on either side (higher and lower than the carrier frequency) of the carrier by an amount equal to the highest modulating frequency. Is it possible to retrieve the message by transmitting (a) only the side bands, (b) only one side band?

3. In amplitude modulation, modulation index $\mu \leq 1$ is used. What will happen if $\mu > 1$?

EXERCISES

15.1 Which of the following frequencies will be suitable for beyond-the horizon communication using sky waves?

(a) 10 kHz (b) 10 MHz (c) 1 GHz (d) 1000 GHz

15.2 Frequencies in the UHF range normally propagate by means of:

(a) Ground waves.

(b) Sky waves.

(c) Surface waves.

(d) Space waves.

15.3 Digital signals

(i) do not provide a continuous set of values,

(ii) represent values as discrete steps,

(iii) can utilize binary system, and

(iv) can utilize decimal as well as binary systems.

Which of the above statements are true?

(a) (i) and (ii) only

(b) (ii) and (iii) only

(c) (i), (ii) and (iii) but not (iv)

(d) All of (i), (ii), (iii) and (iv).

15.4 Is it necessary for a transmitting antenna to be at the same height as that of the receiving antenna for line-of-sight communication? A TV transmitting antenna is 81m tall. How much service area can it cover if the receiving antenna is at the ground level?

15.5 A carrier wave of peak voltage 12V is used to transmit a message signal. What should be the peak voltage of the modulating signal in order to have a modulation index of 75%?

15.6 Due to economic reasons, only the upper sideband of an AM wave is transmitted, but at the receiving station, there is a facility for generating the carrier. Show that if a device is available which can multiply two signals, then it is possible to recover the modulating signal at the receiver station.

15.7 For an amplitude modulated wave, the maximum amplitude is found to be 10V while the minimum amplitude is found to be 2V. Determine the modulation index, μ. What would be the value of μ if the minimum amplitude is zero volt?

15.8 A modulating signal is a square wave, as shown in Fig. 15.14.

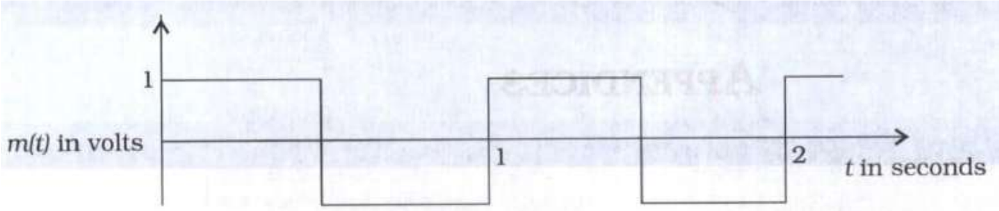

FIGURE 15.14

The carrier wave is given by $c(t) = 2\sin(8\pi t)$ volts.

(i) Sketch the amplitude modulated waveform

(ii) What is the modulation index?

15.9 Give three reasons why modulation of a message signal is necessary for long distance transmission.

15.10 Show graphically an audio signal, a carrier wave and an amplitude modulated wave.

15.11 A message signal of frequency 20 kHz and peak voltage of 20 volts is used to modulate a carrier signal of frequency 2 MHz and peak voltage of 40 volts. Determine (i)modulation index, (ii)the side bands produced. Draw the corresponding frequency spectrum of amplitude modulated signal.

15.12 The carrier wave is given by

$C(t) = 2\sin 8\pi t$ volts.

The modulating signal is a square wave as shown in Fig. 15.15. Calculate modulation index.

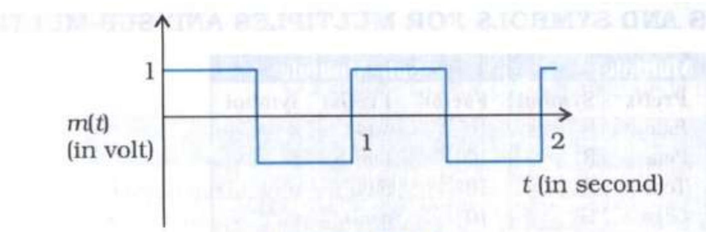

FIGURE 15.15

15.13 Figure 15.16 shows a block diagram of a detector for amplitude modulated signal. Identify the boxes X and Y and write their functions.

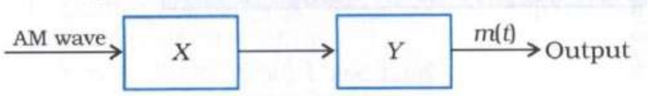

FIGURE 15.16