

CHAPTER 2

ANTENNA FUNDAMENTALS

In this chapter, the basic concept of an antenna is provided and its working is explained. Next, some critical performance parameters of antennas are discussed. Finally, some common types of antennas are introduced.

2.1 Introduction

Antennas are metallic structures designed for radiating and receiving electromagnetic energy. An antenna acts as a transitional structure between the guiding device (e.g. waveguide, transmission line) and the free space. The official IEEE definition of an antenna as given by Stutzman and Thiele [4] follows the concept: “That part of a transmitting or receiving system that is designed to radiate or receive electromagnetic waves”.

2.2 How an Antenna radiates

In order to know how an antenna radiates, let us first consider how radiation occurs. A conducting wire radiates mainly because of time-varying current or an acceleration (or deceleration) of charge. If there is no motion of charges in a wire, no radiation takes place, since no flow of current occurs. Radiation will not occur even if charges are moving with uniform velocity along a straight wire. However, charges moving with uniform velocity along a curved or bent wire will produce radiation. If the charge is oscillating with time, then radiation occurs even along a straight wire as explained by Balanis [5].

The radiation from an antenna can be explained with the help of Figure 2.1 which shows a voltage source connected to a two conductor transmission line. When a sinusoidal voltage is applied across the transmission line, an electric field is created which is sinusoidal in nature and this results in the creation of electric lines of force which are tangential to the electric field. The magnitude of the electric field is indicated by the bunching of the electric lines of force. The free electrons on the conductors are forcibly displaced by the electric lines of force and the movement of these charges causes the flow of current which in turn leads to the creation of a magnetic field.

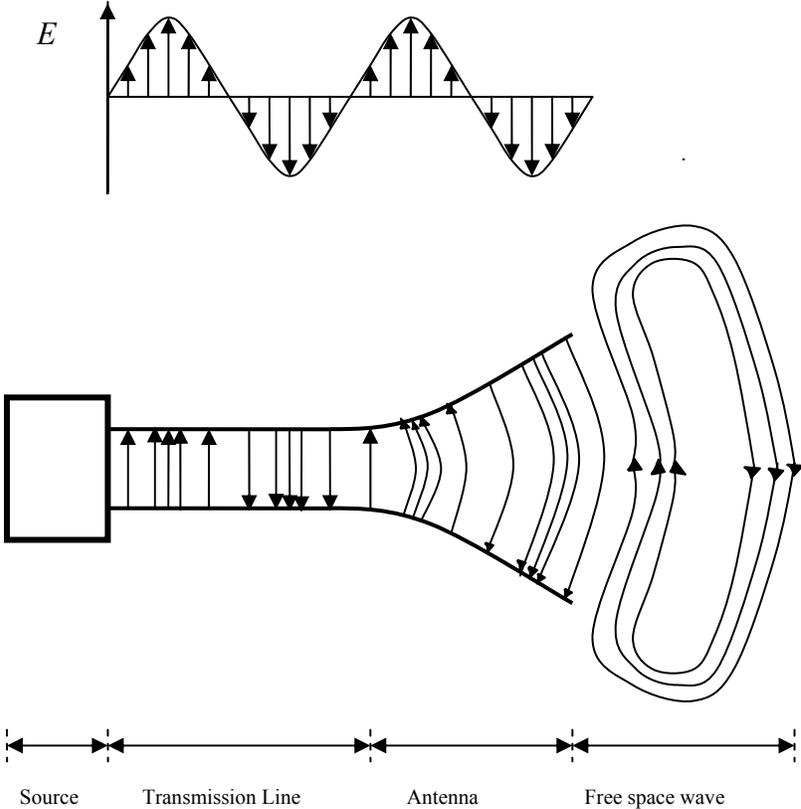


Figure 2.1 Radiation from an antenna

Due to the time varying electric and magnetic fields, electromagnetic waves are created and these travel between the conductors. As these waves approach open space, free space waves are formed by connecting the open ends of the electric lines. Since the sinusoidal source continuously creates the electric disturbance, electromagnetic waves are created continuously

and these travel through the transmission line, through the antenna and are radiated into the free space. Inside the transmission line and the antenna, the electromagnetic waves are sustained due to the charges, but as soon as they enter the free space, they form closed loops and are radiated [5].

2.3 Near and Far Field Regions

The field patterns, associated with an antenna, change with distance and are associated with two types of energy: - radiating energy and reactive energy. Hence, the space surrounding an antenna can be divided into three regions.

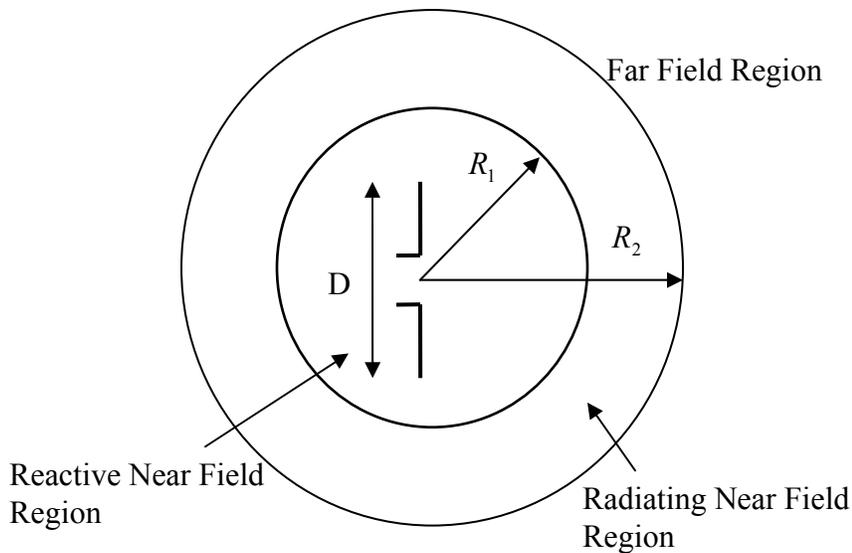


Figure 2.2 Field regions around an antenna

The three regions shown in Figure 2.2 are:

- Reactive near-field region: In this region, the reactive field dominates. The reactive energy oscillates towards and away from the antenna, thus appearing as reactance. In this region, energy is only stored and no energy is dissipated. The outermost boundary for this region is at a distance $R_1 = 0.62\sqrt{D^3/\lambda}$ where R_1 is the distance from the antenna surface, D is the largest dimension of the antenna and λ is the wavelength.

- Radiating near-field region (also called Fresnel region): This is the region which lies between the reactive near-field region and the far field region. Reactive fields are smaller in this field as compared to the reactive near-field region and the radiation fields dominate. In this region, the angular field distribution is a function of the distance from the antenna. The outermost boundary for this region is at a distance $R_2 = 2D^2 / \lambda$ where R_2 is the distance from the antenna surface.
- Far-field region (also called Fraunhofer region): The region beyond $R_2 = 2D^2 / \lambda$ is the far field region. In this region, the reactive fields are absent and only the radiation fields exist. The angular field distribution is not dependent on the distance from the antenna in this region and the power density varies as the inverse square of the radial distance in this region.

2.4 Far field radiation from wires

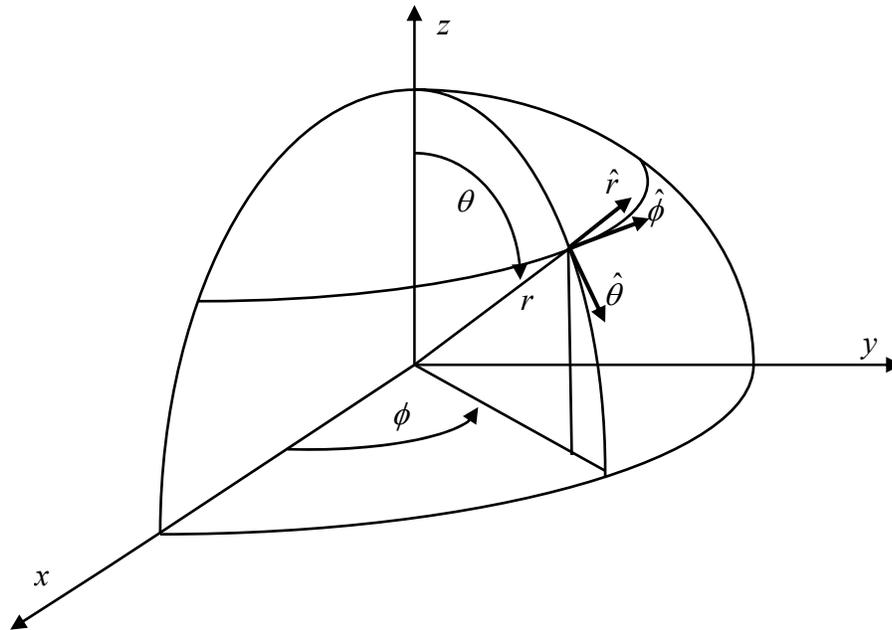


Figure 2.3 Spherical co-ordinate system for a Hertzian dipole

The far field radiation from a Hertzian dipole can be conveniently explained with the help of the spherical co-ordinate system shown in Figure 2.3. The z axis is taken to be the vertical

direction and the xy plane is horizontal. θ denotes the elevation angle and ϕ denotes the azimuthal angle. The xz plane is the elevation plane ($\phi = 0$) or the E-plane which is the plane containing the electric field vector and the direction of maximum radiation. The xy plane is the azimuthal plane ($\theta = \pi/2$) or the H-plane which is the plane containing the magnetic field vector and the direction of maximum radiation [5].

The far field radiation can be explained with the help of the Hertzian dipole or infinitesimal dipole which is a piece of straight wire whose length L and diameter are both very small compared to one wavelength. A uniform current $I(0)$ is assumed to flow along its length. If this dipole is placed at the origin along the z axis, then as given by [5], we can write:

$$E_{\theta} = j\eta \frac{kI(0)Le^{-jkr} \sin \theta}{4\pi r} \left[1 + \frac{1}{jkr} - \frac{1}{(kr)^2} \right] \quad (2.1)$$

$$E_r = \eta \frac{I(0)Le^{-jkr} \cos \theta}{2\pi r^2} \left[1 + \frac{1}{jkr} \right] \quad (2.2)$$

$$H_{\phi} = j \frac{kI(0)Le^{-jkr} \sin \theta}{4\pi r} \left[1 + \frac{1}{jkr} \right] \quad (2.3)$$

$$H_r = 0 \quad (2.4)$$

$$H_{\theta} = 0 \quad (2.5)$$

$$E_{\phi} = 0 \quad (2.6)$$

For far field radiation, terms in r^2 and r^3 can be neglected, hence we can modify the above equations to write:

$$E_{\theta} = j\eta \frac{kI(0)Le^{-jkr}}{4\pi r} \sin \theta \quad (2.7)$$

$$H_{\phi} = j \frac{kI(0)Le^{-jkr}}{4\pi r} \sin \theta \quad (2.8)$$

$$E_r = 0 \quad (2.9)$$

where η = intrinsic free space impedance

$k = 2\pi / \lambda$ = wave propagation constant

r = radius for the spherical co-ordinate system.

In all the above equations, the phase term $e^{j\omega t}$ has been dropped and it is assumed that all the fields are sinusoidally varying with time. It is seen from the above equations that the only

non-zero fields are E_θ and H_ϕ , and that they are transverse to each other. The ratio $E_\theta / H_\phi = \eta$, such that the wave impedance is 120π and the fields are in phase and inversely proportional to r . The directions of E , H and r form a right handed set such that the Poynting vector is in the r direction and it indicates the direction of propagation of the electromagnetic wave. Hence the time average poynting vector given by [5] can be written as:

$$W_{av} = \frac{1}{2} \text{Re}[E \times H^*] \quad (\text{Watts} / \text{m}^2) \quad (2.10)$$

where E and H represent the peak values of the electric and magnetic fields respectively.

The average power radiated by an antenna can be written as:

$$P_{rad} = \oiint W_{rad} ds \quad (\text{Watts}) \quad (2.11)$$

where ds is the vector differential surface = $r^2 \sin \theta d\theta d\phi \hat{r}$

W_{rad} is the magnitude of the time average poynting vector ($\text{Watts} / \text{m}^2$)

The radiation intensity is defined as the power radiated from an antenna per unit solid angle and is given as:

$$U = r^2 W_{rad} \quad (2.12)$$

where U is the radiation intensity in Watts per unit solid angle.

2.5 Antenna Performance Parameters

The performance of an antenna can be gauged from a number of parameters. Certain critical parameters are discussed below.

2.5.1 Radiation Pattern

The radiation pattern of an antenna is a plot of the far-field radiation properties of an antenna as a function of the spatial co-ordinates which are specified by the elevation angle θ and the azimuth angle ϕ . More specifically it is a plot of the power radiated from an antenna per unit solid angle which is nothing but the radiation intensity [5]. Let us consider the case of an isotropic antenna. An isotropic antenna is one which radiates equally in all directions. If the total power radiated by the isotropic antenna is P , then the power is spread over a sphere of radius r , so that the power density S at this distance in any direction is given as:

$$S = \frac{P}{\text{area}} = \frac{P}{4\pi r^2} \quad (2.13)$$

Then the radiation intensity for this isotropic antenna U_i can be written as:

$$U_i = r^2 S = \frac{P}{4\pi} \quad (2.14)$$

An isotropic antenna is not possible to realize in practice and is useful only for comparison purposes. A more practical type is the directional antenna which radiates more power in some directions and less power in other directions. A special case of the directional antenna is the omnidirectional antenna whose radiation pattern may be constant in one plane (e.g. E-plane) and varies in an orthogonal plane (e.g. H-plane). The radiation pattern plot of a generic directional antenna is shown in Figure 2.4.

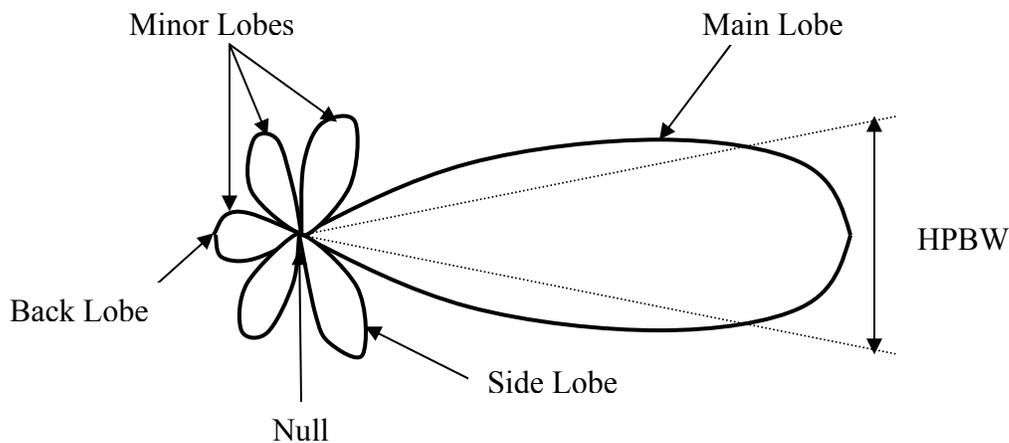


Figure 2.4 Radiation pattern of a generic directional antenna

Figure 2.4 shows the following:

- HPBW: The half power beamwidth (HPBW) can be defined as the angle subtended by the half power points of the main lobe.
- Main Lobe: This is the radiation lobe containing the direction of maximum radiation.
- Minor Lobe: All the lobes other than the main lobe are called the minor lobes. These lobes represent the radiation in undesired directions. The level of minor lobes is

usually expressed as a ratio of the power density in the lobe in question to that of the major lobe. This ratio is called as the side lobe level (expressed in decibels).

- Back Lobe: This is the minor lobe diametrically opposite the main lobe.
- Side Lobes: These are the minor lobes adjacent to the main lobe and are separated by various nulls. Side lobes are generally the largest among the minor lobes.

In most wireless systems, minor lobes are undesired. Hence a good antenna design should minimize the minor lobes.

2.5.2 Directivity

The directivity of an antenna has been defined by [5] as “the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions”. In other words, the directivity of a nonisotropic source is equal to the ratio of its radiation intensity in a given direction, over that of an isotropic source.

$$D = \frac{U}{U_i} = \frac{4\pi U}{P} \quad (2.15)$$

where D is the directivity of the antenna

U is the radiation intensity of the antenna

U_i is the radiation intensity of an isotropic source

P is the total power radiated

Sometimes, the direction of the directivity is not specified. In this case, the direction of the maximum radiation intensity is implied and the maximum directivity is given by [5] as:

$$D_{\max} = \frac{U_{\max}}{U_i} = \frac{4\pi U_{\max}}{P} \quad (2.16)$$

where D_{\max} is the maximum directivity

U_{\max} is the maximum radiation intensity

Directivity is a dimensionless quantity, since it is the ratio of two radiation intensities. Hence, it is generally expressed in dBi. The directivity of an antenna can be easily estimated from the radiation pattern of the antenna. An antenna that has a narrow main lobe would have better directivity, then the one which has a broad main lobe, hence it is more directive.

2.5.3 Input Impedance

The input impedance of an antenna is defined by [5] as “the impedance presented by an antenna at its terminals or the ratio of the voltage to the current at the pair of terminals or the ratio of the appropriate components of the electric to magnetic fields at a point”. Hence the impedance of the antenna can be written as:

$$Z_{in} = R_{in} + jX_{in} \quad (2.17)$$

where Z_{in} is the antenna impedance at the terminals

R_{in} is the antenna resistance at the terminals

X_{in} is the antenna reactance at the terminals

The imaginary part, X_{in} of the input impedance represents the power stored in the near field of the antenna. The resistive part, R_{in} of the input impedance consists of two components, the radiation resistance R_r and the loss resistance R_L . The power associated with the radiation resistance is the power actually radiated by the antenna, while the power dissipated in the loss resistance is lost as heat in the antenna itself due to dielectric or conducting losses.

2.5.4 Voltage Standing Wave Ratio (VSWR)

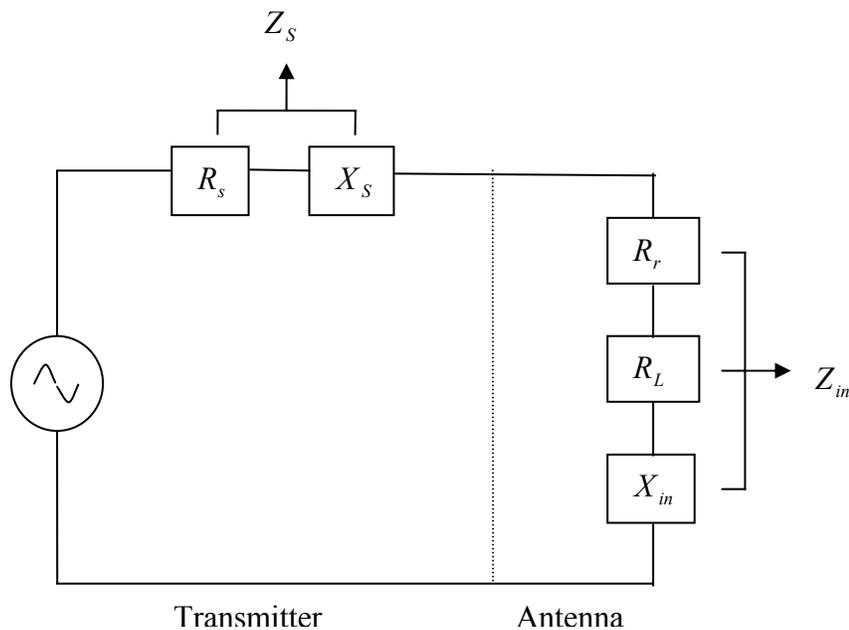


Figure 2.5 Equivalent circuit of transmitting antenna

In order for the antenna to operate efficiently, maximum transfer of power must take place between the transmitter and the antenna. Maximum power transfer can take place only when the impedance of the antenna (Z_{in}) is matched to that of the transmitter (Z_S). According to the maximum power transfer theorem, maximum power can be transferred only if the impedance of the transmitter is a complex conjugate of the impedance of the antenna under consideration and vice-versa. Thus, the condition for matching is:

$$Z_{in} = Z_S^* \quad (2.18)$$

where $Z_{in} = R_{in} + jX_{in}$

$Z_S = R_S + jX_S$ as shown in Figure 2.5

If the condition for matching is not satisfied, then some of the power maybe reflected back and this leads to the creation of standing waves, which can be characterized by a parameter called as the Voltage Standing Wave Ratio (VSWR).

The VSWR is given by Makarov [6] as:

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \quad (2.19)$$

$$\Gamma = \frac{V_r}{V_i} = \frac{Z_{in} - Z_S}{Z_{in} + Z_S} \quad (2.20)$$

where Γ is called the reflection coefficient

V_r is the amplitude of the reflected wave

V_i is the amplitude of the incident wave

The VSWR is basically a measure of the impedance mismatch between the transmitter and the antenna. The higher the VSWR, the greater is the mismatch. The minimum VSWR which corresponds to a perfect match is unity. A practical antenna design should have an input impedance of either 50Ω or 75Ω since most radio equipment is built for this impedance.

2.5.5 Return Loss (RL)

The Return Loss (RL) is a parameter which indicates the amount of power that is “lost” to the load and does not return as a reflection. As explained in the preceding section, waves are reflected leading to the formation of standing waves, when the transmitter and antenna

impedance do not match. Hence the RL is a parameter similar to the VSWR to indicate how well the matching between the transmitter and antenna has taken place. The RL is given as by [6] as:

$$RL = -20 \log_{10} |\Gamma| \quad (\text{dB}) \quad (2.21)$$

For perfect matching between the transmitter and the antenna, $\Gamma = 0$ and $RL = \infty$ which means no power would be reflected back, whereas a $\Gamma = 1$ has a $RL = 0$ dB, which implies that all incident power is reflected. For practical applications, a VSWR of 2 is acceptable, since this corresponds to a RL of -9.54 dB.

2.5.6 Antenna Efficiency

The antenna efficiency is a parameter which takes into account the amount of losses at the terminals of the antenna and within the structure of the antenna. These losses are given by [5] as:

- Reflections because of mismatch between the transmitter and the antenna
- $I^2 R$ losses (conduction and dielectric)

Hence the total antenna efficiency can be written as:

$$e_t = e_r e_c e_d \quad (2.22)$$

where e_t = total antenna efficiency

$e_r = (1 - |\Gamma|^2)$ = reflection (mismatch) efficiency

e_c = conduction efficiency

e_d = dielectric efficiency

Since e_c and e_d are difficult to separate, they are lumped together to form the e_{cd} efficiency which is given as:

$$e_{cd} = e_c e_d = \frac{R_r}{R_r + R_L} \quad (2.23)$$

e_{cd} is called as the antenna radiation efficiency and is defined as the ratio of the power delivered to the radiation resistance R_r , to the power delivered to R_r and R_L .

2.5.7 Antenna Gain

Antenna gain is a parameter which is closely related to the directivity of the antenna. We know that the directivity is how much an antenna concentrates energy in one direction in preference to radiation in other directions. Hence, if the antenna is 100% efficient, then the directivity would be equal to the antenna gain and the antenna would be an isotropic radiator. Since all antennas will radiate more in some direction than in others, therefore the gain is the amount of power that can be achieved in one direction at the expense of the power lost in the others as explained by Ulaby [7]. The gain is always related to the main lobe and is specified in the direction of maximum radiation unless indicated. It is given as:

$$G(\theta, \phi) = e_{cd} D(\theta, \phi) \quad (\text{dBi}) \quad (2.24)$$

2.5.8 Polarization

Polarization of a radiated wave is defined by [5] as “that property of an electromagnetic wave describing the time varying direction and relative magnitude of the electric field vector”. The polarization of an antenna refers to the polarization of the electric field vector of the radiated wave. In other words, the position and direction of the electric field with reference to the earth’s surface or ground determines the wave polarization. The most common types of polarization include the linear (horizontal or vertical) and circular (right hand polarization or the left hand polarization).

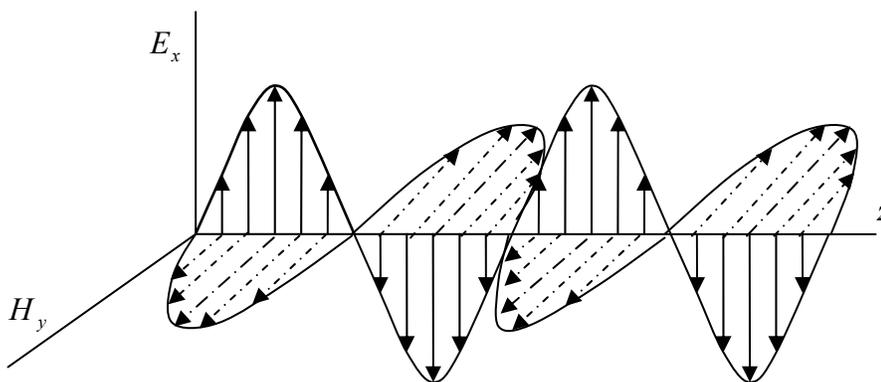


Figure 2.6 A linearly (vertically) polarized wave

If the path of the electric field vector is back and forth along a line, it is said to be linearly polarized. Figure 2.6 shows a linearly polarized wave. In a circularly polarized wave, the electric field vector remains constant in length but rotates around in a circular path. A left hand circular polarized wave is one in which the wave rotates counterclockwise whereas right hand circular polarized wave exhibits clockwise motion as shown in Figure 2.7.

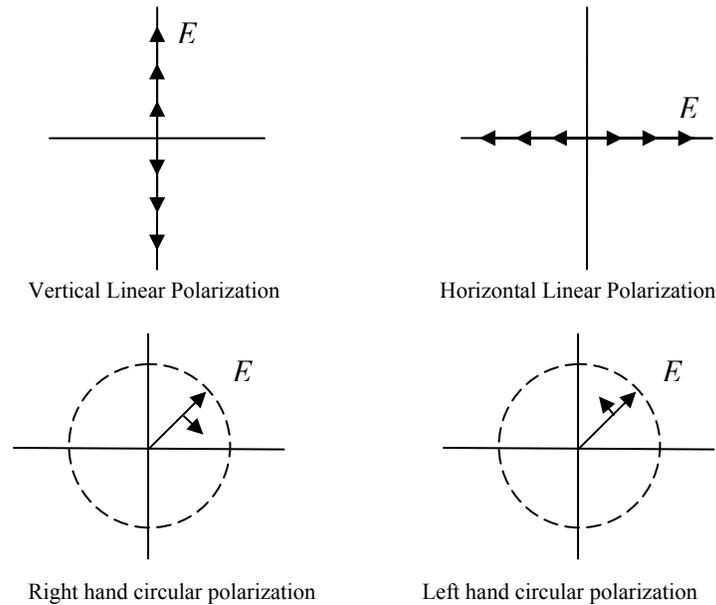


Figure 2.7 Commonly used polarization schemes

2.5.9 Bandwidth

The bandwidth of an antenna is defined by [5] as “the range of usable frequencies within which the performance of the antenna, with respect to some characteristic, conforms to a specified standard.” The bandwidth can be the range of frequencies on either side of the center frequency where the antenna characteristics like input impedance, radiation pattern, beamwidth, polarization, side lobe level or gain, are close to those values which have been obtained at the center frequency. The bandwidth of a broadband antenna can be defined as the ratio of the upper to lower frequencies of acceptable operation. The bandwidth of a narrowband antenna can be defined as the percentage of the frequency difference over the center frequency [5]. According to [4] these definitions can be written in terms of equations as follows:

$$BW_{broadband} = \frac{f_H}{f_L} \quad (2.25)$$

$$BW_{narrowband} (\%) = \left[\frac{f_H - f_L}{f_C} \right] 100 \quad (2.26)$$

where f_H = upper frequency

f_L = lower frequency

f_C = center frequency

An antenna is said to be broadband if $f_H/f_L = 2$. One method of judging how efficiently an antenna is operating over the required range of frequencies is by measuring its VSWR. A $VSWR \leq 2$ ($RL \geq -9.5dB$) ensures good performance.

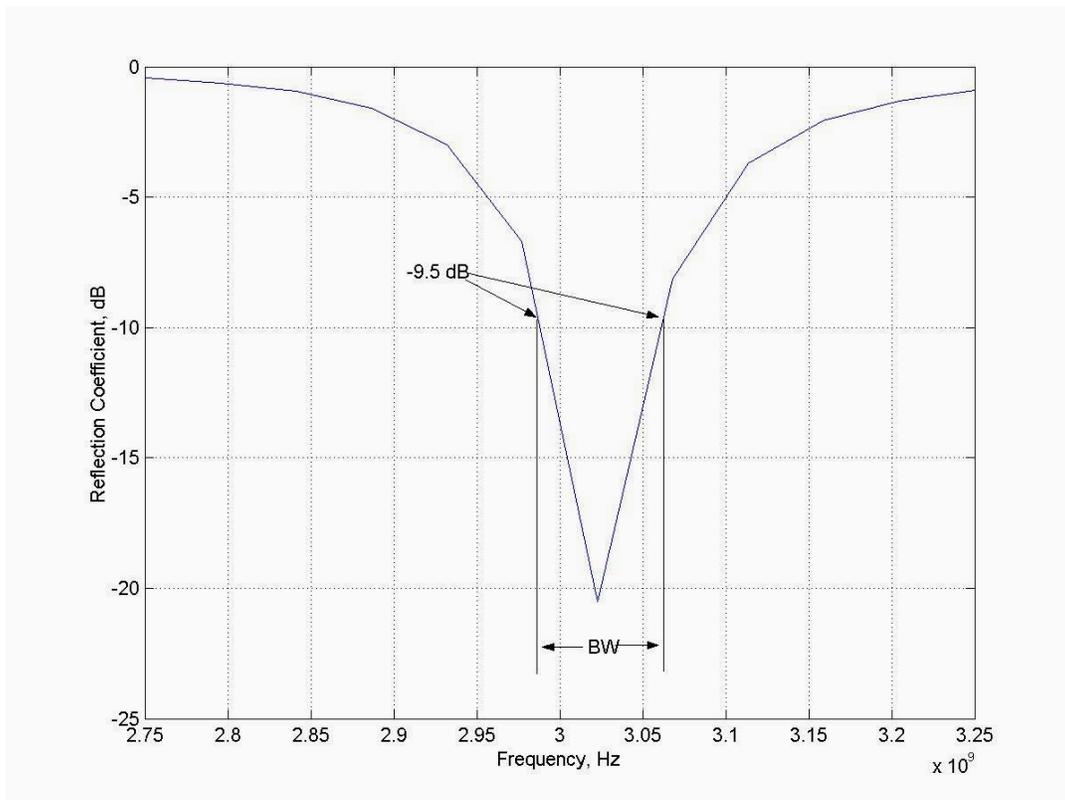


Figure 2.8 Measuring bandwidth from the plot of the reflection coefficient

2.6 Types of Antennas

Antennas come in different shapes and sizes to suit different types of wireless applications. The characteristics of an antenna are very much determined by its shape, size and the type of material that it is made of. Some of the commonly used antennas are briefly described below.

2.6.1 Half Wave Dipole

The length of this antenna is equal to half of its wavelength as the name itself suggests. Dipoles can be shorter or longer than half the wavelength, but a tradeoff exists in the performance and hence the half wavelength dipole is widely used.

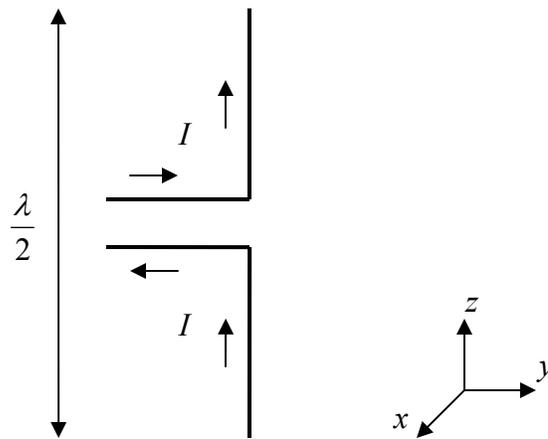


Figure 2.9 Half wave dipole

The dipole antenna is fed by a two wire transmission line, where the two currents in the conductors are of sinusoidal distribution and equal in amplitude, but opposite in direction. Hence, due to canceling effects, no radiation occurs from the transmission line. As shown in Figure 2.9, the currents in the arms of the dipole are in the same direction and they produce radiation in the horizontal direction. Thus, for a vertical orientation, the dipole radiates in the horizontal direction. The typical gain of the dipole is 2dB and it has a bandwidth of about 10%. The half power beamwidth is about 78 degrees in the E plane and its directivity is 1.64 (2.15dB)

with a radiation resistance of 73Ω [4]. Figure 2.10 shows the radiation pattern for the half wave dipole.

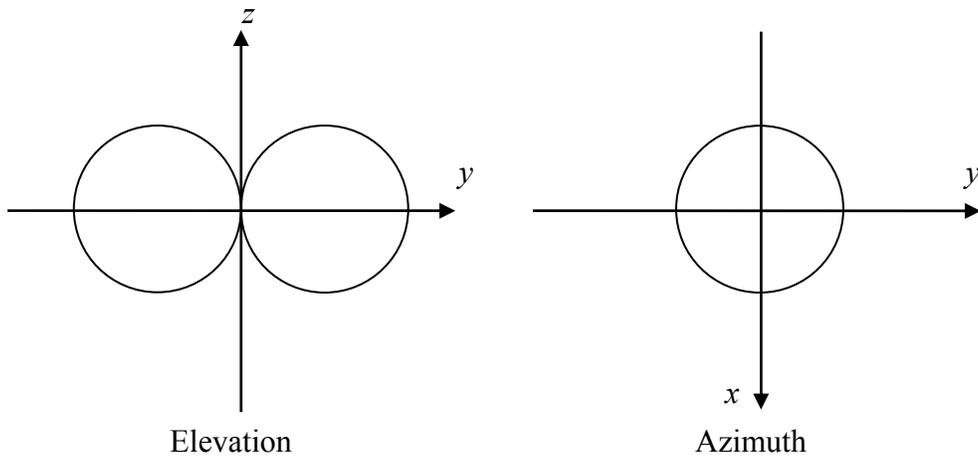


Figure 2.10 Radiation pattern for Half wave dipole

2.6.2 Monopole Antenna

The monopole antenna, shown in Figure 2.11, results from applying the image theory to the dipole. According to this theory, if a conducting plane is placed below a single element of length $L/2$ carrying a current, then the combination of the element and its image acts identically to a dipole of length L except that the radiation occurs only in the space above the plane as discussed by Saunders [8].

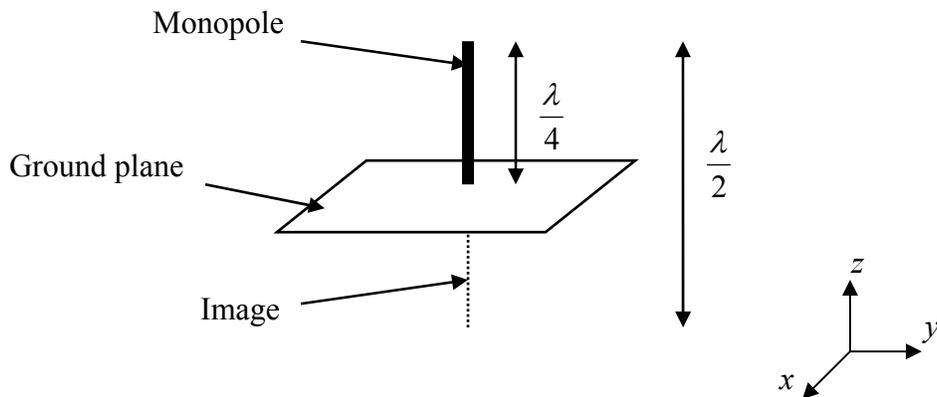


Figure 2.11 Monopole Antenna

For this type of antenna, the directivity is doubled and the radiation resistance is halved when compared to the dipole. Thus, a half wave dipole can be approximated by a quarter wave monopole ($L/2 = \lambda/4$). The monopole is very useful in mobile antennas where the conducting plane can be the car body or the handset case. The typical gain for the quarter wavelength monopole is 2-6dB and it has a bandwidth of about 10%. Its radiation resistance is 36.5Ω and its directivity is 3.28 (5.16dB) [4]. The radiation pattern for the monopole is shown below in Figure 2.12.

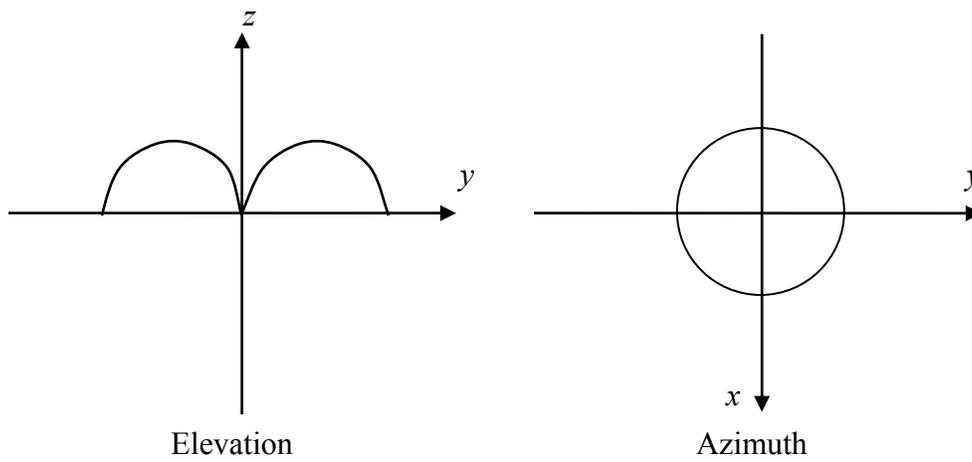


Figure 2.12 Radiation pattern for the Monopole Antenna

2.6.3 Loop Antennas

The loop antenna is a conductor bent into the shape of a closed curve such as a circle or a square with a gap in the conductor to form the terminals as shown in Figure 2.13. There are two types of loop antennas-electrically small loop antennas and electrically large loop antennas. If the total loop circumference is very small as compared to the wavelength ($L \ll \lambda$), then the loop antenna is said to be electrically small. An electrically large loop antenna typically has its circumference close to a wavelength. The far-field radiation patterns of the small loop antenna are insensitive to shape [4].

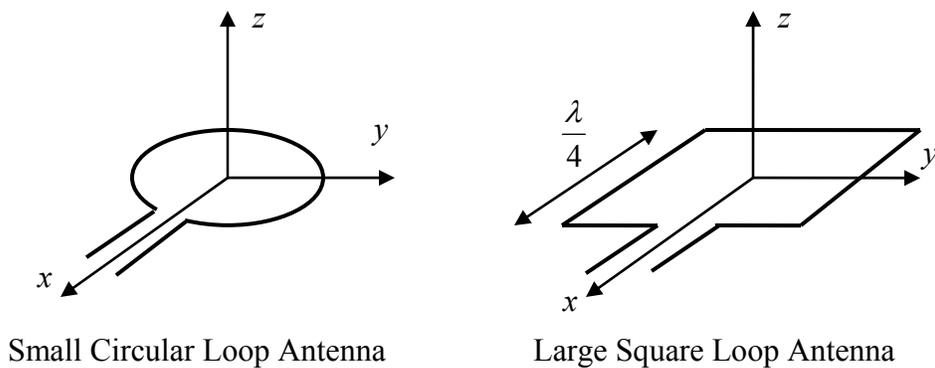


Figure 2.13 Loop Antenna

As shown in Figure 2.14, the radiation patterns are identical to that of a dipole despite the fact that the dipole is vertically polarized whereas the small circular loop is horizontally polarized.

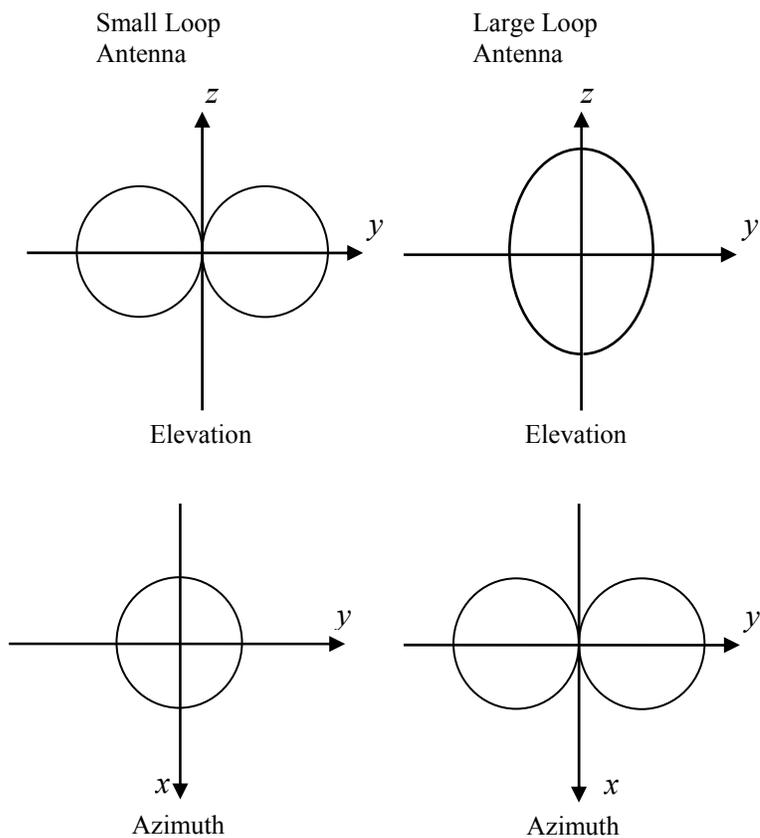


Figure 2.14 Radiation Pattern of Small and Large Loop Antenna

The performance of the loop antenna can be increased by filling the core with ferrite. This helps in increasing the radiation resistance. When the perimeter or circumference of the loop antenna is close to a wavelength, then the antenna is said to be a large loop antenna.

The radiation pattern of the large loop antenna is different than that of the small loop antenna. For a one wavelength square loop antenna, radiation is maximum normal to the plane of the loop (along the z axis). In the plane of the loop, there is a null in the direction parallel to the side containing the feed (along the x axis), and there is a lobe in a direction perpendicular to the side containing the feed (along the y axis). Loop antennas generally have a gain from -2dB to 3dB and a bandwidth of around 10%. . The small loop antenna is very popular as a receiving antenna [4]. Single turn loop antennas are used in pagers and multiturn loop antennas are used in AM broadcast receivers.

2.6.4 Helical Antennas

A helical antenna or helix is one in which a conductor connected to a ground plane, is wound into a helical shape. Figure 2.15 illustrates a helix antenna. The antenna can operate in a number of modes, however the two principal modes are the normal mode (broadside radiation) and the axial mode (endfire radiation). When the helix diameter is very small as compared to the wavelength, then the antenna operates in the normal mode. However, when the circumference of the helix is of the order of a wavelength, then the helical antenna is said to be operating in the axial mode.

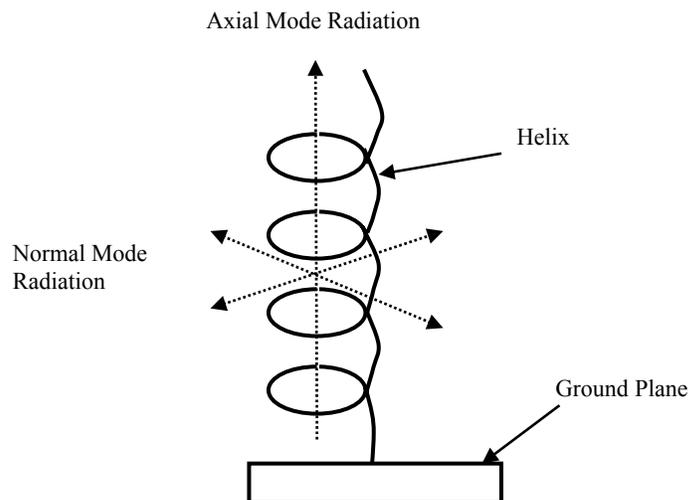


Figure 2.15 Helix Antenna

In the normal mode of operation, the antenna field is maximum in a plane normal to the helix axis and minimum along its axis. This mode provides low bandwidth and is generally used for hand-portable mobile applications [8].

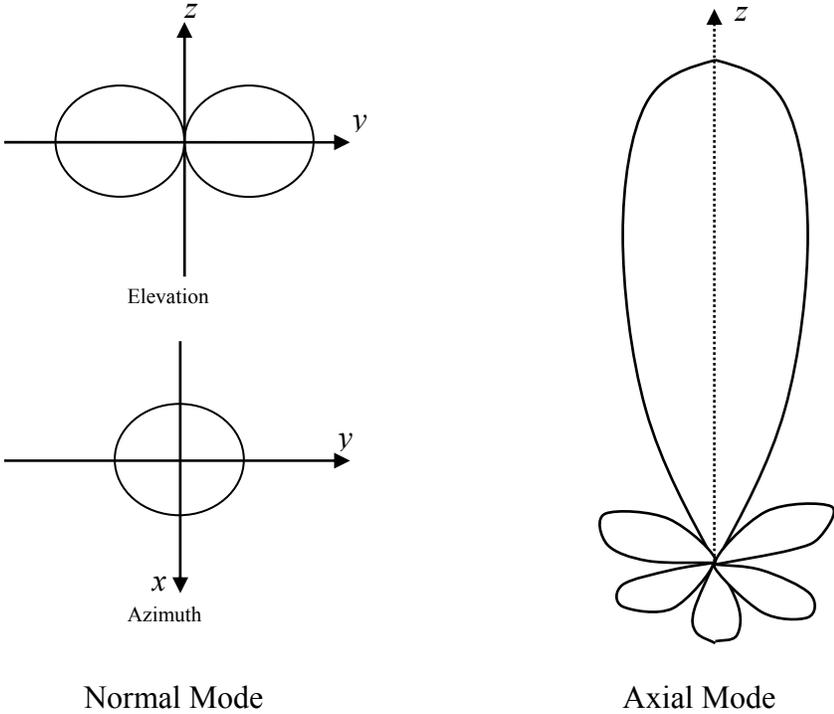


Figure 2.16 Radiation Pattern of Helix Antenna

In the axial mode of operation, the antenna radiates as an endfire radiator with a single beam along the helix axis. This mode provides better gain (upto 15dB) [4] and high bandwidth ratio (1.78:1) as compared to the normal mode of operation. For this mode of operation, the beam becomes narrower as the number of turns on the helix is increased. Due to its broadband nature of operation, the antenna in the axial mode is used mainly for satellite communications. Figure 2.16 above shows the radiation patterns for the normal mode as well as the axial mode of operations.

2.6.5 Horn Antennas

Horn antennas are used typically in the microwave region (gigahertz range) where waveguides are the standard feed method, since horn antennas essentially consist of a waveguide whose end walls are flared outwards to form a megaphone like structure.

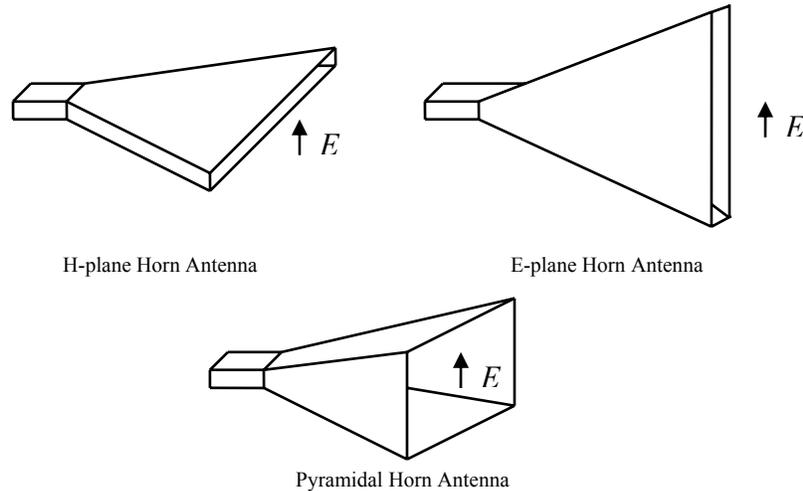


Figure 2.17 Types of Horn Antenna

Horns provide high gain, low VSWR, relatively wide bandwidth, low weight, and are easy to construct [4]. The aperture of the horn can be rectangular, circular or elliptical. However, rectangular horns are widely used. The three basic types of horn antennas that utilize a rectangular geometry are shown in Figure 2.17. These horns are fed by a rectangular waveguide which have a broad horizontal wall as shown in the figure. For dominant waveguide mode excitation, the E-plane is vertical and H-plane horizontal. If the broad wall dimension of the horn is flared with the narrow wall of the waveguide being left as it is, then it is called an H-plane sectoral horn antenna as shown in the figure. If the flaring occurs only in the E-plane dimension, it is called an E-plane sectoral horn antenna. A pyramidal horn antenna is obtained when flaring occurs along both the dimensions. The horn basically acts as a transition from the waveguide mode to the free-space mode and this transition reduces the reflected waves and emphasizes the traveling waves which lead to low VSWR and wide bandwidth [4]. The horn is widely used as a feed element for large radio astronomy, satellite tracking, and communication dishes.

In the above sections, several antennas have been discussed. Another commonly used antenna is the Microstrip patch antenna. The aim of this thesis is to design a compact microstrip patch antenna to be used in wireless communication and this antenna is explained in the next chapter.