ANTENNA ENGINEERING

ANTENNA ENGINEERING (3-1-0)

Module-I (14 Hours)

Antenna Definition,Principles of Radiation, Basic antenna parameters, Retarded Vector Magnetic Potential, Radiation field from Current element., Current Distribution on a thin Wire. Half wave dipole and Quarterwave monopole.

Two-element array. Principle of Pattern Multiplication. Linear Array. Broadside and end fire patterns, Balun.

Module-II

(12 Hours)

Folded Dipole, Yagi Antenna. Frequency Independent Antenna. Log Periodic Dipole array, Secondary Source and Aperture Antennas (Basics & applications).

Module-III (10 Hours)

Horn Antennas-Pyramidal & Sectoral Horn. Radiation Pattern and Gain of horn antenna. Parabolic Reflector Antenna -Principle, analysis, Radiation Pattern and Gain. Principles of Cassegrain Antenna.

Module-IV

(08 Hours)

Microstrip Antenna – Basic Characteristics, Rectangular Patch, Radiation principle, Feeding Techniques,Cavity model.

Antenna Measurements – Radiation Pattern, Gain and Input Impedance.

Text Books:

1. Electromagnetic Wave and Radiating system by E. C. Jordan and K.G. Balmain, 10, 11, 12, 13, 14, and 15.

2. Antennas Theory – Analysis and Design by C. Balanis, 2nd Edition, John Willey & Sons Selected portion Ch. 11, 12, 13, 15, and 16.

References Books:

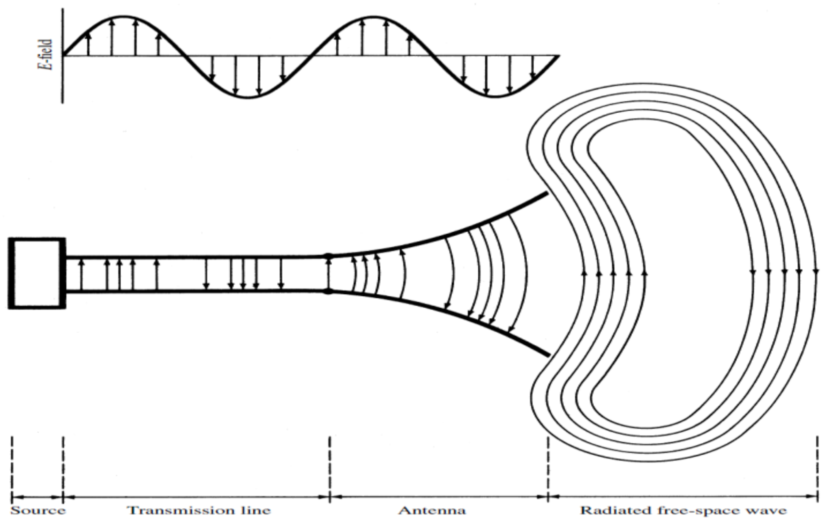
1. Antenna Engineering by J. D. Krauss.
2. Antenna Engineering by W. L. Weeks.
3. Antennas and Wave Propagation by G. S. N. Raju, Pearson Education.
4. Antenna & Wave Propagation by R E. Collins.

ANTENNA DEFINITION:

An antenna is defined by Webster's Dictionary as "a usually metallic device (as a rod or wire) for radiating or receiving radio waves." The IEEE Standard Definitions of Terms for Antennas (IEEE Std 145–1983)∗ defines the antenna or aerial as "a means for radiating or receiving radio waves." In other words the antenna is the transitional structure between free-space and a guiding device, as shown in Figure 1.1. The guiding device or transmission line may take the form of a coaxial line or a hollow pipe (waveguide), and it is used to transport electromagnetic energy from the transmitting source to the antenna, or from the antenna to the receiver. In the former case, we have a transmitting antenna and in the latter a receiving antenna.

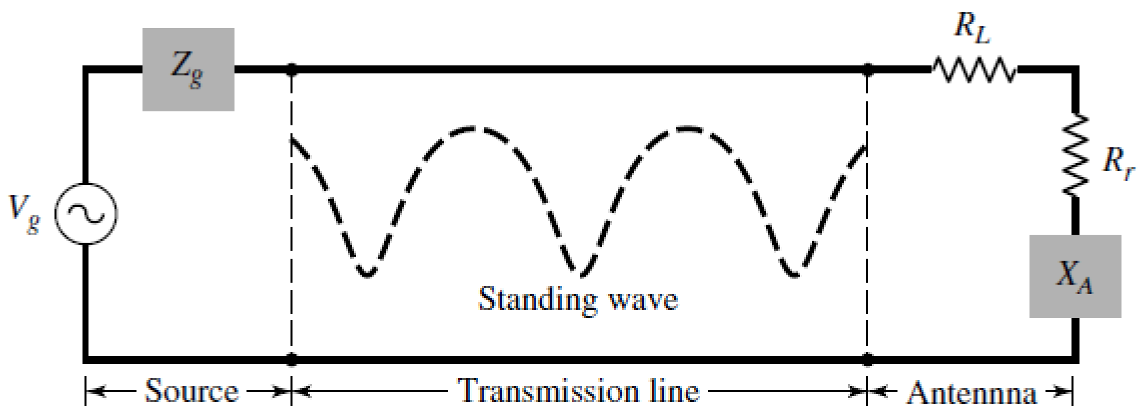
A transmission-line Thevenin equivalent of the antenna system of Figure 1.1 in the transmitting mode is shown in Figure 1.2 where the source is represented by an ideal generator, the transmission line is represented by a line with characteristic impedance Zc, and the antenna is represented by a load ZA [ZA = (RL + Rr) + jXA] connected to the transmission line. The Thevenin and Norton circuit equivalents of the antenna are also shownin Figure 2.27. The load resistance RL is used to represent the conduction and dielectric losses associated with the antenna structure while Rr , referred to as the radiation resistance, is used to represent radiation by the antenna. The reactance XA is used to represent the imaginary part of the impedance associated with radiation by the antenna. This is discussed more in detail in Sections 2.13 and 2.14. Under ideal conditions, energy generated by the source should be totally transferred to the

Radiation resistance Rr , which is used to represent radiation by the antenna. However, in a practical system there are conduction-dielectric losses due to the lossy nature of the transmission line and the antenna, as well as those due to reflections (mismatch) losses at the interface between the line and the antenna. Taking into account the internal impedance of the source and neglecting line and reflection (mismatch) losses, maximum power is delivered to the antenna under conjugate matching.



The reflected waves from the interface create, along with the traveling waves from the source toward the antenna, constructive and destructive interference patterns, referred to as standing waves, inside the transmission line which represent pockets of energy concentrations and storage, typical of resonant devices. A typical standing wave pattern is shown dashed in Figure 1.2, while another is exhibited in Figure 1.15. If the antenna system is not properly designed, the transmission line could act to a large degree as an energy storage element instead of as a wave guiding and energy transporting device. If the maximum field intensities of the standing wave are sufficiently large, they can cause arching inside the transmission lines. The losses due to the line, antenna, and the standing waves are undesirable. The losses due to the line can be minimized by selecting low-loss lines while those of the antenna can be decreased by reducing the loss resistance represented by RL in Figure 1.2. The standing waves can be reduced, and the energy storage capacity of the line minimized, by matching the impedance of the antenna (load) to the characteristic impedance of the line. This is the same as matching loads to transmission lines, where the load here is the antenna, and is discussed more in detail in Section 9.7. An equivalent similar to that of Figure 1.2 is used to represent the antenna system in the receiving mode where the source is replaced by a receiver. All other parts of the transmission-line equivalent remain the same. The radiation resistance Rr is used to represent in the receiving mode the transfer of energy from the free-space wave to the antenna.

In addition to receiving or transmitting energy, an antenna in an advanced wireless system is usually required to optimize or accentuate the radiationen ergy insome directions and suppress it in others. Thus the antenna must also serve as a directional device in addition to a probing device. It must thentake various forms to meet the particular need at hand, and it may be a piece of conducting wire, an aperture, a patch, an assembly of elements (array), a reflector, a lens, and so forth.



$$Z_A = (R_L + R_r) + jX_A$$

PRINCIPLES OF RADIATION:

It simply states that to create radiation, there must be a time-varying current or an acceleration (or deceleration) of charge. We usually refer to currents in time-harmonic applications while charge is most often mentioned in transients. To create charge acceleration (or deceleration) the wire must be curved, bent, discontinuous, or terminated [1], [4]. Periodic charge acceleration
(or deceleration) or time-varying current is also created when charge is oscillating in a time-harmonic motion, as shown in Figure 1.17 for a λ/2 dipole. Therefore:

1. If a charge is not moving, current is not created and there is no radiation.
2. If charge is moving with a uniform velocity:

a. There is no radiation if the wire is straight, and infinite in extent.

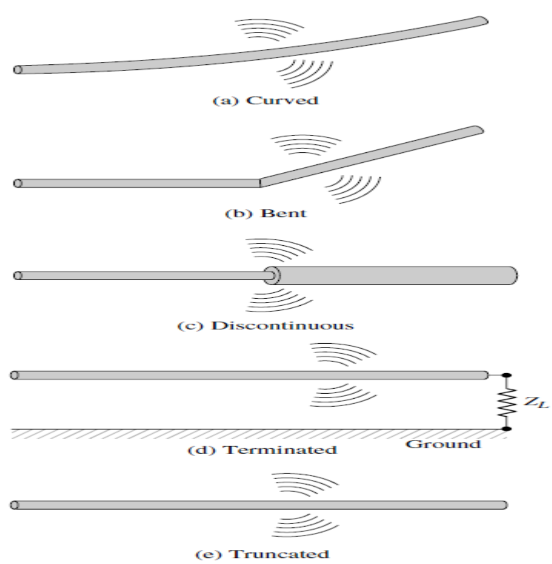
b. There is radiation if the wire is curved, bent, discontinuous, terminated, or truncated, as shown in Figure 1.10.

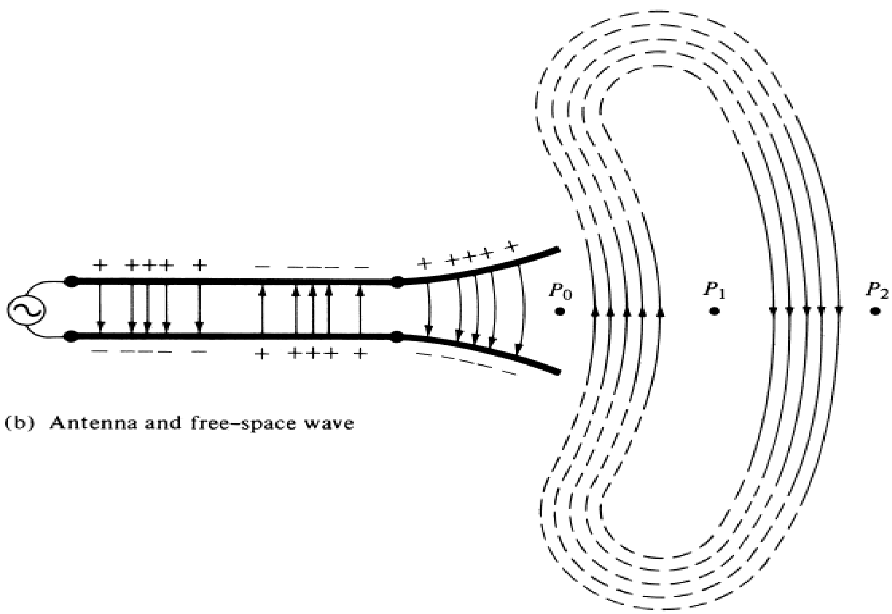
3. If charge is oscillating in a time-motion, it radiates even if the wire is straight.

A qualitative understanding of the radiation mechanism may be obtained by considering a pulse source attached to an open-ended conducting wire, which may be connected to the ground through a discrete load at its open end, as shown in Figure 1.10(d). When the wire is initially energized, the charges (free electrons) in the wire are set in motionby the electrical lines of force created by the source. When charges are accelerated in the source-end of the wire and decelerated (negative acceleration with respect to original motion) during reflection from its end, it is suggested that radiated fields are produced at each end and along the remaining part of the wire, [1], [4]. Stronger radiation with a more broad frequency spectrum occurs if the pulses are of shorter or more compact duration while continuous time-harmonic oscillating charge produces, ideally, radiation of single frequency determined by the frequency of oscillation. The

Acceleration of the charges is accomplished by the external source in which forces set the charges in motion and produce the associated field radiated. The deceleration of the charges at the end of the wire is accomplished by the internal (self) forces associated with the induced field due to the buildup of charge concentration at the ends of the wire.

The internal forces receive energy from the charge buildup as its velocity is reduced to zero at the ends of the wire. Therefore, charge acceleration due to an exciting electric field and deceleration due to impedance discontinuities or smooth curves of the wire are mechanisms responsible for electromagnetic radiation. While both current density (Jc) and charge density (qv) appear as source terms in Maxwell’s equation, charge is viewed as a more fundamental quantity, especially for transient fields. Even though this interpretation of radiation is primarily used for transients, it can be used to explain steady state radiation[4].





(b) Antenna and free–space wave

Let us consider a voltage source connected to a two-conductor transmission line which is connected to an antenna. This is shown in Figure 1.11(a). Applying a voltage across the two-conductor transmission line creates an electric field between the conductors. The electric field has associated with it electric lines of force which are tangent to the electric field at each point and their strength is proportional to the electric field intensity. The electric lines of force have a tendency to act on the free electrons (easily detachable from the atoms) associated with each conductor and force them to be displaced. The movement of the charges creates a current that in turn creates a magnetic field intensity. Associated with the magnetic field intensity are magnetic lines of force which are tangent to the magnetic field. We have accepted that electric field lines start on positive charges and end on negative charges. They also can start on a positive charge and end at infinity, start at infinity and end on a negative charge, or form closed loops neither starting or ending on any charge. Magnetic field lines always form closed loops encircling current-carrying conductors because physically there are no magnetic charges. In some mathematical formulations, it is often convenient to introduce equivalent magnetic charges and magnetic currents to draw a parallel between solutions involving electric and magnetic sources. The electric field lines drawn between the two conductors help to exhibit the distribution of charge. If we assume that the voltage source is sinusoidal, we expect the electric field between the conductors to also be sinusoidal with a period equal to that of the applied source. The relative magnitude of the electric field intensity is indicated by the density (bunching) of the lines of force with the arrows showing the relative direction (positive or negative). The creation of time-varying electric and magnetic fields between the conductors forms electromagnetic waves which travel along the transmission line, as shown in Figure 1.11(a). The electromagnetic waves enter the antenna and have associated with them electric charges and corresponding currents. If we remove part of the antenna structure, as shown in Figure 1.11(b), free-space waves can be formed by “connecting” the open ends of the electric lines (shown dashed). The free-space waves are also periodic but a constant phase point P0 moves outwardly with the speed of light and travels a distance of λ/2 (to P1) inthe time of one-half of a period. It has been shown [6] that close to the antenna the constant phase point P0 moves faster thanthe speed of light but approaches the speed of light at points far away from the antenna (analogous to phase velocity inside a rectangular waveguide). Figure 1.12 displays the creationan d travel of free-space waves by a prolate spheroid with λ/2 interfocal distance where λ is the wavelength. The free-space waves of a center-fed λ/2 dipole, except in the immediate vicinity of the antenna, are essentially the same as those of the prolate spheroid.

BASIC ANTENNA PARAMETERS:

RADIATION PATTERN

An antenna radiation pattern or antenna pattern is defined as “a mathematical function or a graphical representation of the radiation properties of the antenna as a function of space coordinates. In most cases, the radiation pattern is determined in the far field region and is represented as a function of the directional coordinates. Radiation properties include power flux density, radiation intensity, field strength, directivity, phase or polarization.” The radiation property of most concern is the two- or three dimensional spatial distribution of radiated energy as a function of the observer’s position along a path or surface of constant radius. A convenient set of coordinates is shown in Figure 2.1. A trace of the received electric (magnetic) field at a constant radius is called the amplitude field pattern.

On the other hand, a graph of the spatial variation of the power density along a constant radius is called an amplitude power pattern. Often the field and power patterns are normalized with respect to their maximum value, yielding normalized field and power patterns. Also, the power pattern is usually plotted on a logarithmic scale or more commonly in decibels (dB). This scale is usually desirable because a logarithmic scale can accentuate in more details those parts of the Pattern that have very low values, which later we will refer to as minor lobes. For an antenna, the

(a) Field pattern(in linear scale) typically represents a plot of the magnitude of the electric or magnetic field as a function of the angular space.
(b) Power pattern(in linear scale) typically represents a plot of the square of the magnitude of the electric or magnetic field as a function of the angular space.
(c) power pattern(in dB) represents the magnitude of the electric or magnetic field in decibels, as a function of the angular space.

To demonstrate this, the two-dimensional normalized field pattern (plotted in linear scale), power pattern(plotted in linear scale), and power pattern (plotted on a logarithmic dBscale) of a 10-element linear antenna array of isotropic sources, with a spacing of d = 0.25λ betweenthe elements, are shown in Figure 2.2. In this and subsequent patterns, the plus (+) and minus (−) signs in the lobes indicate the relative polarization of the amplitude between the various lobes, which changes (alternates) as the nulls are crossed. To find the points where the pattern achieves its half-power (−3 dB points), relative to the maximum value of the pattern, you set the value of the

a. field pattern at 0.707 value of its maximum, as shown in Figure 2.2(a)
b. power pattern (in a linear scale) at its 0.5 value of its maximum, as shown in Figure 2.2(b)
c. power pattern (in dB) at −3 dBvalue of its maximum, as shown in Figure 2.2(c).

Isotropic, Directional, and Omnidirectional Patterns

An isotropic radiator is defined as “a hypothetical lossless antenna having equal radiation in all directions.” Although it is ideal and not physically realizable, it is often Figure 2.5 taken as a reference for expressing the directive properties of actual antennas. A directional antenna is one “having the property of radiating or receiving electromagnetic waves more effectively in some directions than in others. This term is usually applied to an antenna whose maximum directivity is significantly greater than that of a half-wave dipole.” Examples of antennas with directional radiation patterns are shown in Figures 2.5 and 2.6. It is seen that the pattern in Figure 2.6 is non directional in the azimuth plane [f (φ), θ = π/2] and directional in the elevation plane [g(θ), φ = constant]. This type of a pattern is designated as omni directional, an dit is defined as one "having an essentially nondirectional pattern in a given plane (in this case in azimuth) and a directional pattern in any orthogonal plane (in this case in elevation)." An omnidirectional patternis thena special type of a directional pattern.

Field Regions

The space surrounding an antenna is usually subdivided into three regions: (a) reactive
near-field, (b) radiating near-field (Fresnel) and (c) far-field (Fraunhofer) regions as
shown in Figure 2.7. These regions are so designated to identify the field structure in
each. Although no abrupt changes in the field configurations are noted as the boundaries
are crossed, there are distinct differences among them. The boundaries separating
these regions are not unique, although various criteria have been established and are
commonly used to identify the regions.

Reactive near-field region is defined as "that portion of the near-field region immediately surrounding the antenna wherein the reactive field predominates." For most antennas, the outer boundary of this region is commonly taken to exist at a distanceR < 0.62 _ D3/λ from the antenna surface, where λ is the wavelength and D is the largest dimension of the antenna. "For a very short dipole, or equivalent radiator, the outer boundary is commonly taken to exist at a distance λ/2π from the antenna surface."

Radiating near-field (Fresnel) region is defined as "that region of the field of an antenna between the reactive near-field region and the far-field region wherein radiation fields predominate and wherein the angular field distribution is dependent upon the distance from the antenna. If the antenna has a maximum dimension that is not large compared to the wavelength, this region may not exist. For an antenna focused at infinity, the radiating near-field region is sometimes referred to as the Fresnel region on the basis of analogy to optical terminology. If the antenna has a maximum overall dimension which is very small compared to the wavelength, this field region may not exist." The inner boundary is taken to be the distance R ≥ 0.62

_ D3/λ and the outer boundary the distance R < 2D2/λ where D is the largest∗ dimension of the antenna. This criterion is based on a maximum phase error of π/8. In this region the field pattern is, in general,
a function of the radial distance and the radial field component may be appreciable.

Far-field (Fraunhofer) region is defined as “that region of the field of an antenna where the angular field distribution is essentially independent of the distance from the antenna. If the antenna has a maximum∗ overall dimension D, the far-field region is commonly taken to exist at distances greater than 2D2/λ from the antenna, λ being the wavelength. The far-field patterns of certain antennas, such as multibeam reflector antennas, are sensitive to variations in phase over their apertures. For these antennas 2D2/λ may be inadequate. In physical media, if the antenna has a maximum overall dimension, D, which is large compared to π/|γ |, the far-field region can be taken to begin approximately at a distance equal to |γ |D2/π from the antenna, γ being the
propagation constant in the medium. For an antenna focused at infinity, the far-field region is sometimes referred to as the Fraunhofer region on the basis of analogy to optical terminology.” In this region, the field components are essentially transverse and the angular distribution is independent of the radial distance where the measurements are made. The inner boundary is taken to be the radial distance R = 2D2/λ and the outer one at infinity.

RADIATION POWER DENSITY

Electromagnetic waves are used to transport information through a wireless medium or a guiding structure, from one point to the other. It is then natural to assume that power and energy are associated with electromagnetic fields. The quantity used to describe the power associated with
an electromagnetic wave is the instantaneous Poynting vector defined as

$$\mathcal{W} = \mathcal{E} \times \mathcal{H}$$

Since the Poynting vector is a power density, the total power crossing a closed surface can be obtained by integrating the normal component of the Poynting vector over the entire surface. In equation form

$$\mathcal{P} = \oiint_S \mathcal{W} \cdot d\mathbf{s} = \oiint_S \mathcal{W} \cdot \hat{\mathbf{n}}\, da$$

For applications of time-varying fields, it is often more desirable to find the average power density which is obtained by integrating the instantaneous Poynting vector over one period and dividing by the period. For time-harmonic variations of the form ejωt , we define the complex fields E and H which are related to their instantaneous counterparts _ and _ by

$$\mathscr{E}(x, y, z; t) = \mathrm{Re}[\mathbf{E}(x, y, z)e^{j\omega t}]$$

$$\mathscr{H}(x, y, z; t) = \mathrm{Re}[\mathbf{H}(x, y, z)e^{j\omega t}]$$

The first term of (2-7) is not a function of time, and the time variations of the second are twice the given frequency. The time average Poynting vector (average power density) can be written as

$$\mathscr{W} = \mathscr{E} \times \mathscr{H} = \tfrac{1}{2}\mathrm{Re}[\mathbf{E} \times \mathbf{H}^*] + \tfrac{1}{2}\mathrm{Re}[\mathbf{E} \times \mathbf{H}e^{j2\omega t}]$$

$$\boxed{\mathbf{W}_{\mathrm{av}}(x, y, z) = [\mathscr{W}(x, y, z; t)]_{\mathrm{av}} = \tfrac{1}{2}\mathrm{Re}[\mathbf{E} \times \mathbf{H}^*]} \quad (\mathrm{W/m^2})$$

The ½ factor appears in(2-7) and (2-8) because the E and H fields represent peak values, and it should be omitted for RMS values. A close observationof (2-8) may raise a question. If the real part of $(E \times H^*)/2$ represents the average (real) power density, what does the imaginary part of the same quantity represent? At this point it will be very natural to assume that the imaginary part must represent the reactive (stored) power density associated with the electromagnetic fields. In later chapters, it will be shown that the power density associated with the electromagnetic fields of an antenna in its far-field region is predominately real and will be referred to as radiation density.

$$\boxed{\begin{aligned} P_{\mathrm{rad}} = P_{\mathrm{av}} &= \oint\!\!\!\oint_S \mathbf{W}_{\mathrm{rad}} \cdot d\mathbf{s} = \oint\!\!\!\oint_S \mathbf{W}_{\mathrm{av}} \cdot \hat{\mathbf{n}}\, da \\ &= \frac{1}{2} \oint\!\!\!\oint_S \mathrm{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} \end{aligned}}$$

RADIATION INTENSITY

Radiation intensity ina givendirectionis defined as “the power radiated from an antenna per unit solid angle.” The radiation intensity is a far-field parameter, and it can be obtained by simply multiplying the radiation density by the square of the distance. In mathematical form it is expressed as

$$\boxed{U = r^2 W_{\text{rad}}}$$

where

U = radiation intensity (W/unit solid angle)

Wrad = radiationden sity (W/m2)

The total power is obtained by integrating the radiation intensity, as given by (2-12), over the entire solid angle of 4π. Thus

$$P_{\text{rad}} = \oiint_{\Omega} U \, d\Omega = \int_0^{2\pi} \int_0^{\pi} U \sin\theta \, d\theta \, d\phi$$

For anisotropic source U will be independent of the angles θ and φ, as was the case for Wrad. Thus (2-13) canbe written as

$$P_{\text{rad}} = \oiint_{\Omega} U_0 \, d\Omega = U_0 \oiint_{\Omega} d\Omega = 4\pi U_0$$

or the radiation intensity of an isotropic source as

$$U_0 = \frac{P_{\text{rad}}}{4\pi}$$

BEAMWIDTH

Associated with the pattern of an antenna is a parameter designated as beamwidth. The beamwidth of a pattern is defined as the angular separation between two identical points on opposite side of the pattern maximum. In an antenna pattern, there are a number of beamwidths. One of the most widely used beamwidths is the Half-Power Beamwidth (HPBW), which is defined by IEEE as: “In a plane containing the direction of the maximum of a beam, the angle between the two

directions in which the radiation intensity is one-half value of the beam." This is demonstrated in Figure 2.2. Another important beamwidth is the angular separation between the first nulls of the pattern, and it is referred to as the First-Null Beamwidth (FNBW). Both the HPBW and FNBW are demonstrated for the pattern in Figure 2.11 or the pattern of Example 2.4. Other beamwidths are those where the patternis −10 dB from the maximum, or any other
value. However, inpractice, the term beamwidth, with no other identification, usually refers to HPBW.

DIRECTIVITY

In the 1983 versionof the IEEE Standard Definitions of Terms for Antennas, there has been a substantive change in the definition of directivity, compared to the defin tion of the 1973 version. Basically the term directivity in the new 1983 version has been used to replace the term directive gain of the old 1973 version. In the new 1983 version the term directive gain has been deprecated. According to the authors of the new 1983 standards, "this change brings this standard in line with common usage among antenna engineers and with other international standards, notably those of the International Electrotechnical Commission (IEC)." Therefore directivity of an antenna defined as "the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions. The average radiation intensity is equal to the total power radiated by the antenna divided by 4π. If the direction is not specified, the direction of maximum radiation intensity is implied." Stated more simply, 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. In mathematical form, using (2-15), it can be written as

$$\boxed{D = \frac{U}{U_0} = \frac{4\pi U}{P_{\text{rad}}}}$$

If the direction is not specified, it implies the direction of maximum radiation intensity
(maximum directivity) expressed as

$$\boxed{D_{max} = D_0 = \frac{U|_{max}}{U_0} = \frac{U_{max}}{U_0} = \frac{4\pi U_{max}}{P_{rad}}}$$

D = directivity (dimensionless)
D0 = maximum directivity (dimensionless)
U = radiation intensity (W/unit solid angle)
Umax = maximum radiation intensity (W/unit solid angle)
U0 = radiation intensity of isotropic source (W/unit solid angle)
Prad = total radiated power (W)

ANTENNA EFFICIENCY

Associated with an antenna are a number of efficiencies and can be defined using Figure 2.22. The total antenna efficiency e0 is used to take into account losses at the input terminals and within the structure of the antenna. Such losses may be due, referring to Figure 2.22(b), to

1. reflections because of the mismatch between the transmission line and the antenna
2. I 2R losses (conduction and dielectric)

In general, the overall efficiency can be written as

e0 = er ec ed

where
e0 = total efficiency (dimensionless)
er = reflection(mismatch) efficiency = (1 − |?|2) (dimensionless)
ec = conduction efficiency (dimensionless)
ed = dielectric efficiency (dimensionless)
? = voltage reflection coefficient at the input terminals of the antenna
[? = (Zin − Z0)/(Zin + Z0) where Zin = antenna input impedance,
Z0 = characteristic impedance of the transmission line]
VSWR = voltage standing wave ratio = 1 + |?|/(1 − |?|)

Usually ec and ed are very difficult to compute, but they canbe determined experimentally.

Even by measurements they cannot be separated, and it is usually more convenient to write (2-44) as e0 = erecd = ecd (1 − |?|2)

GAIN

Another useful measure describing the performance of an antenna is the gain. Although the gain of the antenna is closely related to the directivity, it is a measure that takes into account the efficiency of the antenna as well as its directional capabilities. Remember that directivity is a measure that describes only the directional properties of the antenna, and it is therefore controlled only by the pattern. Gain of an antenna (in a given direction) is defined as “the ratio of the intensity, in a given direction, to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically. The radiation intensity corresponding to
the isotropically radiated power is equal to the power accepted (input) by the antenna divided by 4π.” Inequation form this canbe expressed as

$$\text{Gain} = 4\pi \frac{\text{radiation intensity}}{\text{total input (accepted) power}} = 4\pi \frac{U(\theta,\phi)}{P_{in}} \quad \text{(dimensionless)}$$

RETARDED VECTOR MAGNETIC POTENTIAL:

Although magnetic currents appear to be physically unrealizable, equivalent magnetic currents arise when we use the volume or the surface equivalence theorems. The fields generated by a harmonic magnetic current in a homogeneous region, with J = 0 but M _= 0, must satisfy ∇ · D = 0. Therefore, EF canbe expressed as the curl of the vector potential F by

$$\boxed{\mathbf{E}_F = -\frac{1}{\epsilon}\nabla \times \mathbf{F}}$$

Substituting (3-16) into Maxwell’s curl equation

$$\nabla \times \mathbf{H}_F = j\omega\epsilon\mathbf{E}_F$$

reduces it to

$$\nabla \times (\mathbf{H}_F + j\omega\mathbf{F}) = 0$$

From the vector identity of (3-6), it follows that

$$\boxed{\mathbf{H}_F = -\nabla\phi_m - j\omega\mathbf{F}}$$

where φm represents an arbitrary magnetic scalar potential which is a function of position. Taking the curl of (3-16)

$$\nabla \times \mathbf{E}_F = -\frac{1}{\epsilon}\nabla \times \nabla \times \mathbf{F} = -\frac{1}{\epsilon}[\nabla\nabla \cdot \mathbf{F} - \nabla^2\mathbf{F}]$$

and equating it to Maxwell's equation

$$\boxed{\nabla \times \mathbf{E}_F = -\mathbf{M} - j\omega\mu\mathbf{H}_F}$$

leads to

$$\nabla^2\mathbf{F} + j\omega\mu\epsilon\mathbf{H}_F = \nabla\nabla \cdot \mathbf{F} - \epsilon\mathbf{M}$$

$$\nabla^2\mathbf{F} + k^2\mathbf{F} = -\epsilon\mathbf{M} + \nabla(\nabla \cdot \mathbf{F}) + \nabla(j\omega\mu\epsilon\phi_m)$$

By letting

$$\boxed{\nabla \cdot \mathbf{F} = -j\omega\mu\epsilon\phi_m \Rightarrow \phi_m = -\frac{1}{j\omega\mu\epsilon}\nabla \cdot \mathbf{F}}$$

reduces (3-23) to

$$\boxed{\nabla^2\mathbf{F} + k^2\mathbf{F} = -\epsilon\mathbf{M}}$$

and (3-19) to

$$\mathbf{H}_F = -j\omega\mathbf{F} - \frac{j}{\omega\mu\epsilon}\nabla(\nabla \cdot \mathbf{F})$$

RADIATION FIELD FROM CURRENT ELEMENT:

An infinitesimal linear wire (l _ λ) is positioned symmetrically at the origin of the coordinate system and oriented along the z axis, as shown in Figure 4.1(a). Although infinitesimal dipoles are not very practical, they are used to represent capacitor-plate (also referred to as top-hat-loaded) antennas. In addition, they are utilized as building blocks of more complex geometries. The end plates are used to provide capacitive loading in order to maintain the current on the dipole nearly uniform. Since the end plates are assumed to be small, their radiation is usually negligible. The wire, in Addition to being very small (l _ λ), is very thin (a _ λ). The spatial variationof the current is assumed to be constant an d given by

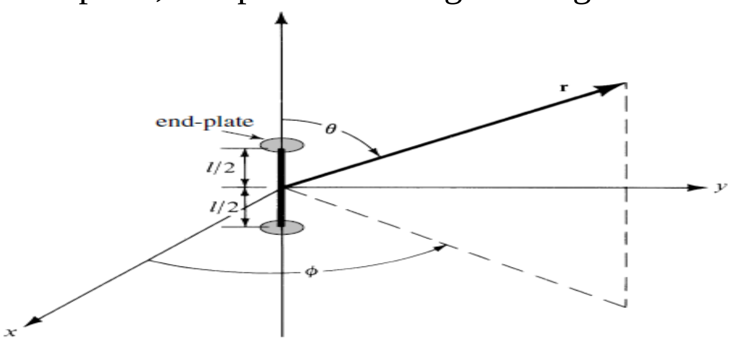
$$\mathbf{A}(x, y, z) = \frac{\mu}{4\pi} \int_C \mathbf{I}_e(x', y', z') \frac{e^{-jkR}}{R}\, dl'$$

Since the source only carries an electric current Ie, Im and the potential function F are zero. To find A we write

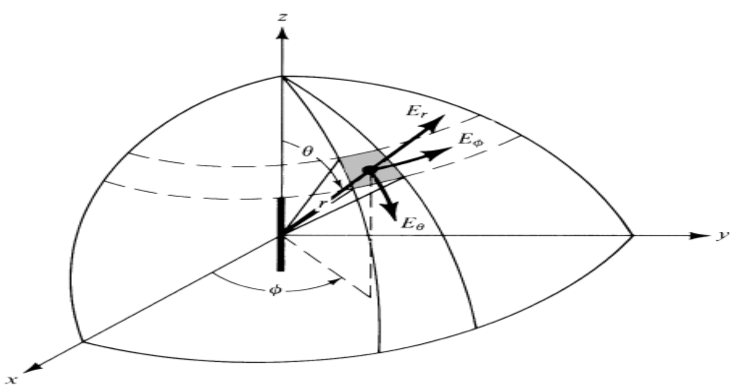
$$\mathbf{I}_e(x', y', z') = \hat{\mathbf{a}}_z I_0$$

$$x' = y' = z' = 0 \text{ (infinitesimal dipole)}$$

where (x, y, z) represent the observation point coordinates, represent the coordinates of the source, R is the distance from any point on the source to the observation point, and path C is along the length of the source.



(a) Infinitesimal dipole



For this problem, Ax = Ay = 0, so (4-5) using (4-4) reduces to

$$\boxed{\mathbf{A}(x, y, z) = \hat{\mathbf{a}}_z \frac{\mu I_0}{4\pi r} e^{-jkr} \int_{-l/2}^{+l/2} dz' = \hat{\mathbf{a}}_z \frac{\mu I_0 l}{4\pi r} e^{-jkr}}$$

$$\begin{bmatrix} A_r \\ A_\theta \\ A_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}$$

$$A_r = A_z \cos\theta = \frac{\mu I_0 l e^{-jkr}}{4\pi r} \cos\theta$$

$$A_\theta = -A_z \sin\theta = -\frac{\mu I_0 l e^{-jkr}}{4\pi r} \sin\theta$$

$$A_\phi = 0$$

$$\mathbf{H} = \hat{\mathbf{a}}_\phi \frac{1}{\mu r} \left[\frac{\partial}{\partial r}(r A_\theta) - \frac{\partial A_r}{\partial \theta} \right]$$

$$\boxed{\begin{gathered} H_r = H_\theta = 0 \\ H_\phi = j \frac{k I_0 l \sin\theta}{4\pi r} \left[1 + \frac{1}{jkr} \right] e^{-jkr} \end{gathered}}$$

$$\mathbf{E} = \mathbf{E}_A = -j\omega\mathbf{A} - j\frac{1}{\omega\mu\epsilon}\nabla(\nabla \cdot \mathbf{A}) = \frac{1}{j\omega\epsilon}\nabla \times \mathbf{H}$$

Power Density and Radiation Resistance

The input impedance of an antenna, which consists of real and imaginary parts, was discussed in Section 2.13. For a lossless antenna, the real part of the input impedance was designated as radiation resistance. It is through the mechanism of the radiation resistance that power is transferred from the guided wave to the free-space wave. To find the input resistance for a lossless antenna, the Poynting vector is formed in terms of the E- an dH-fields radiated by the antenna. By integrating the Poynting vector over a closed surface (usually a sphere of constant radius), the total power radiated by the source is found. The real part of it is related to the input resistance.

$$\mathbf{W} = \tfrac{1}{2}(\mathbf{E} \times \mathbf{H}^*) = \tfrac{1}{2}(\hat{\mathbf{a}}_r E_r + \hat{\mathbf{a}}_\theta E_\theta) \times (\hat{\mathbf{a}}_\phi H_\phi^*)$$

$$= \tfrac{1}{2}(\hat{\mathbf{a}}_r E_\theta H_\phi{}^* - \hat{\mathbf{a}}_\theta E_r H_\phi{}^*)$$

$$W_r = \frac{\eta}{8}\left|\frac{I_0 l}{\lambda}\right|^2 \frac{\sin^2\theta}{r^2}\left[1 - j\frac{1}{(kr)^3}\right]$$

$$W_\theta = j\eta\frac{k|I_0 l|^2 \cos\theta \sin\theta}{16\pi^2 r^3}\left[1 + \frac{1}{(kr)^2}\right]$$

$$P = \oiint_S \mathbf{W} \cdot d\mathbf{s} = \int_0^{2\pi}\int_0^{\pi} (\hat{\mathbf{a}}_r W_r + \hat{\mathbf{a}}_\theta W_\theta) \cdot \hat{\mathbf{a}}_r r^2 \sin\theta \, d\theta \, d\phi$$

$$P = \int_0^{2\pi}\int_0^{\pi} W_r r^2 \sin\theta \, d\theta \, d\phi = \eta\frac{\pi}{3}\left|\frac{I_0 l}{\lambda}\right|^2\left[1 - j\frac{1}{(kr)^3}\right]$$

The transverse component W_θ of the power density does not contribute to the integral. Thus (4-14) does not represent the total complex power radiated by the antenna. Since W_θ , as given by (4-12b), is purely imaginary, it will not contribute to any real radiated power. However, it does contribute to the imaginary (reactive) power which along with the second term of (4-14) can be used to determine the total reactive power of the antenna. The reactive power density, which is most dominant for small values of kr, has both radial and transverse components. It merely changes between outward and inward directions to form a standing wave at a rate of twice per cycle. It also moves in the transverse direction as suggested by (4-12b).

$$P = \frac{1}{2}\iint_S \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{s} = \eta\left(\frac{\pi}{3}\right)\left|\frac{I_0 l}{\lambda}\right|^2\left[1 - j\frac{1}{(kr)^3}\right]$$

$$= P_{\text{rad}} + j2\omega(\tilde{W}_m - \tilde{W}_e)$$

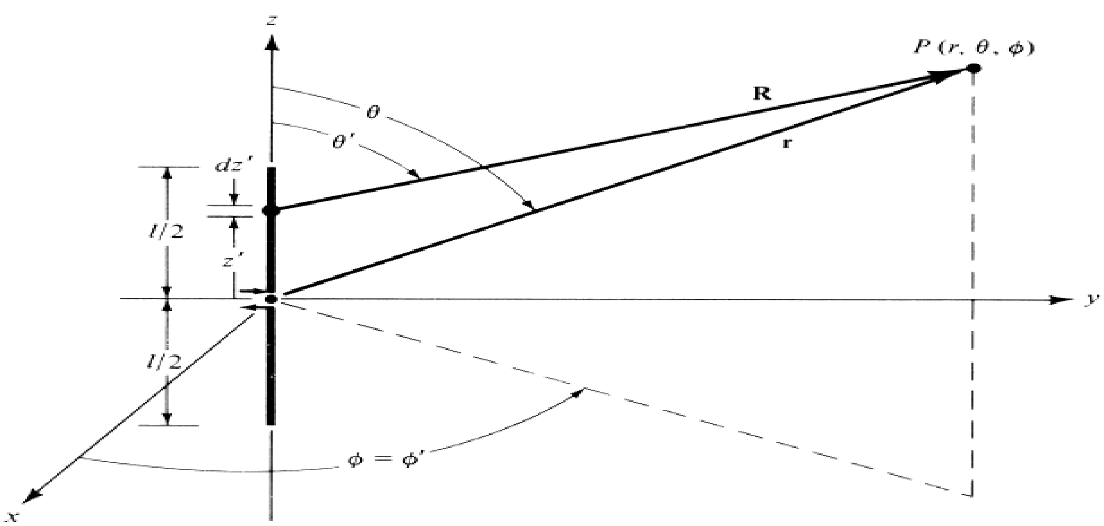
$$P_{\text{rad}} = \eta\left(\frac{\pi}{3}\right)\left|\frac{I_0 l}{\lambda}\right|^2 = \frac{1}{2}|I_0|^2 R_r$$

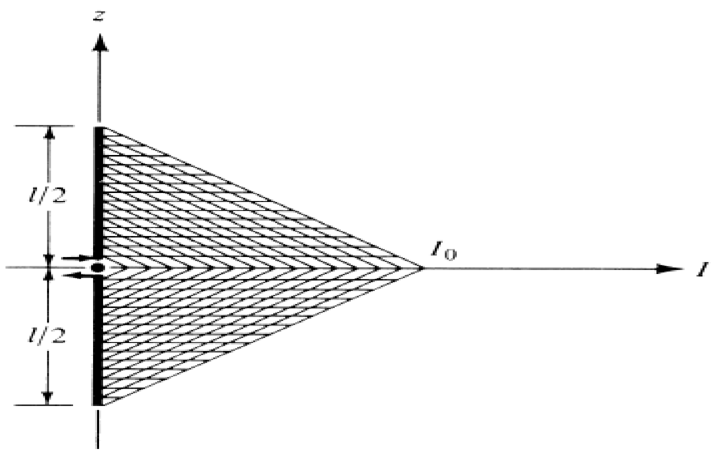
$$R_r = \eta \left(\frac{2\pi}{3}\right)\left(\frac{l}{\lambda}\right)^2 = 80\pi^2 \left(\frac{l}{\lambda}\right)^2$$

CURRENT DISTRIBUTION ON A HALF WAVE DIPOLE:

The creation of the current distribution ona thinwire was discussed inSection 1.4, and it was illustrated with some examples inFigure 1.16. The radiationproperties of an infinitesimal dipole, which is usually taken to have a length $l \leq \lambda/50$, were discussed in the previous section. Its current distribution was assumed to be constant. Although a constant current distribution is not realizable (other than top-hat-loaded elements), it is a mathematical quantity that is used to represent actual current distributions of antennas that have been incremented into many small lengths.

A better approximation of the current distribution of wire antennas, whose lengths are usually $\lambda/50 < l \leq \lambda/10$, is the triangular variation of Figure 1.16(a). The sinusoidalvariations of Figures 1.16(b)–(c) are more accurate representations of the current distribution of any length wire antenna. The most convenient geometrical arrangement for the analysis of a dipole is usually to have it positioned symmetrically about the origin with its length directed along the z-axis, as shown in Figure 4.4(a). This is not necessary, but it is usually the most convenient. The current distribution of a small dipole ($\lambda/50 < l \leq \lambda/10$) is shown in





$$\mathbf{I}_e(x', y', z') = \begin{cases} \hat{\mathbf{a}}_z I_0 \left(1 - \frac{2}{l} z'\right), & 0 \leq z' \leq l/2 \\ \hat{\mathbf{a}}_z I_0 \left(1 + \frac{2}{l} z'\right), & -l/2 \leq z' \leq 0 \end{cases}$$

Following the procedure established in the previous section, the vector potential of(4-2) canbe written using (4-33) as

$$\mathbf{A}(x, y, z) = \frac{\mu}{4\pi} \left[\hat{\mathbf{a}}_z \int_{-l/2}^{0} I_0 \left(1 + \frac{2}{l} z'\right) \frac{e^{-jkR}}{R} \, dz' + \hat{\mathbf{a}}_z \int_{0}^{l/2} I_0 \left(1 - \frac{2}{l} z'\right) \frac{e^{-jkR}}{R} \, dz' \right]$$

Because the overall length of the dipole is very small (usually $l \leq \lambda/10$), the values of R for different values of z _ along the length of the wire ($-l/2 \leq$ z _ $\leq l/2$) are not much different from r. Thus R canbe approximated by R r throughout the integration path. The maximum phase error in(4-34) by allowing R = r for $\lambda/50 < l \leq \lambda/10$, will be $kl/2 = \pi/10$ rad = 18◦ for $l = \lambda/10$. Smaller values will occur for the other lengths. As it will be shown in the next section, this amount of phase error is usually considered negligible and has very little effect on the overall radiation characteristics. Performing the integration, (4-34) reduces to

$$\mathbf{A} = \hat{\mathbf{a}}_z A_z = \hat{\mathbf{a}}_z \frac{1}{2} \left[\frac{\mu I_0 l e^{-jkr}}{4\pi r} \right]$$

which is one-half of that obtained in the previous section for the infinitesimal dipole and given by (4-4). The potential function given by (4-35) becomes a more accurate approximation as

kr →∞. This is also the region of most practical interest, and it has been designated as the far-field region. Since the potential function for the triangular distribution is one-half of the corresponding one for the constant (uniform) current distribution, the corresponding fields of the former are one-half of the latter. Thus we can write the E and H-fields radiated by a small dipole

$$\left.\begin{aligned} E_\theta &\simeq j\eta \frac{kI_0 l e^{-jkr}}{8\pi r}\sin\theta \\ E_r &\simeq E_\phi = H_r = H_\theta = 0 \\ H_\phi &\simeq j\frac{kI_0 l e^{-jkr}}{8\pi r}\sin\theta \end{aligned}\right\} \quad kr \gg 1$$

same as the ones with the constant current distribution given by (4-31) and (4-32), respectively.

TWO ELEMENT ARRAY:

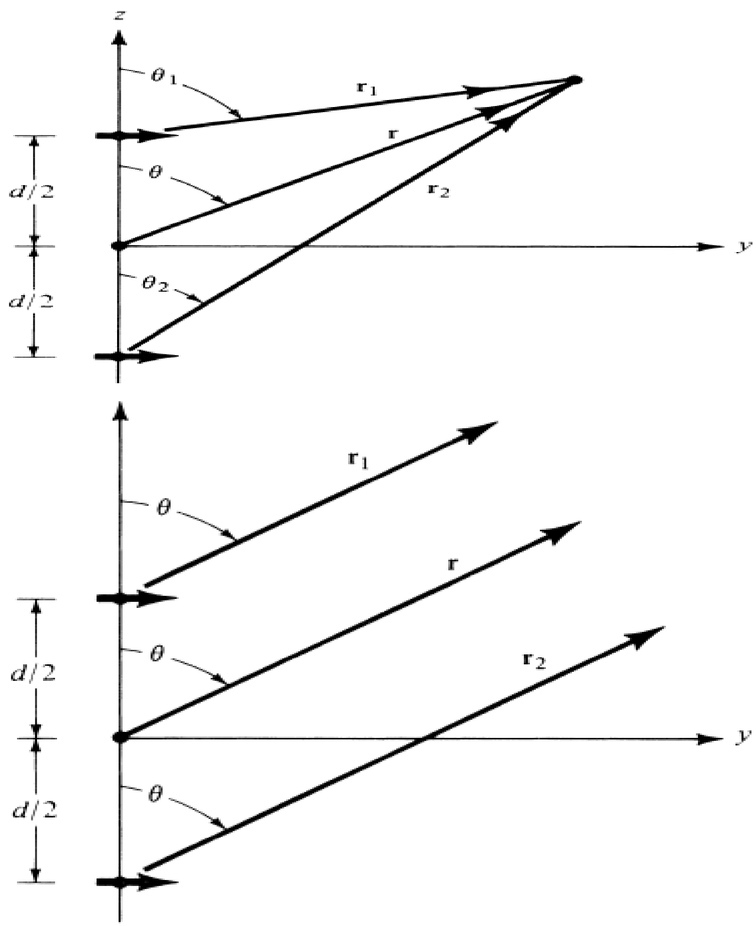
Let us assume that the antenna under investigation is an array of two infinitesimal horizontal dipoles positioned along the z-axis, as showninFigure 6.1(a). The total field radiated by the two elements, assuming no coupling between the elements, is equal to the sum of the two and in the y-z plane it is given by

$$\mathbf{E}_t = \mathbf{E}_1 + \mathbf{E}_2 = \hat{\mathbf{a}}_\theta\, j\eta \frac{kI_0 l}{4\pi}\left\{\frac{e^{-j[kr_1-(\beta/2)]}}{r_1}\cos\theta_1 + \frac{e^{-j[kr_2+(\beta/2)]}}{r_2}\cos\theta_2\right\}$$

where β is the difference in phase excitation between the elements. The magnitude excitation of the radiators is identical. Assuming far-field observations and referring to

$$
\left.
\begin{aligned}
&\theta_1 \simeq \theta_2 \simeq \theta \\
&r_1 \simeq r - \frac{d}{2}\cos\theta \\
&r_2 \simeq r + \frac{d}{2}\cos\theta
\end{aligned}
\right\} \text{ for phase variations}
$$

$$
r_1 \simeq r_2 \simeq r \qquad \text{for amplitude variations}
$$



$$
\mathbf{E}_t = \hat{\mathbf{a}}_\theta j\eta \frac{kI_0 l e^{-jkr}}{4\pi r}\cos\theta[e^{+j(kd\cos\theta+\beta)/2} + e^{-j(kd\cos\theta+\beta)/2}]
$$

$$
\mathbf{E}_t = \hat{\mathbf{a}}_\theta j\eta \frac{kI_0 l e^{-jkr}}{4\pi r}\cos\theta\left\{2\cos\left[\frac{1}{2}(kd\cos\theta+\beta)\right]\right\}
$$

It is apparent from (6-3) that the total field of the array is equal to the field of a single element positioned at the origin multiplied by a factor which is widely referred to as the array factor. Thus for the two-element array of constant amplitude, the array factor is givenby

$$\mathrm{AF} = 2\cos[\tfrac{1}{2}(kd\cos\theta + \beta)]$$

PRINCIPLE OF PATTERN MULTIPLICATION:

$$(\mathrm{AF})_n = \cos[\tfrac{1}{2}(kd\cos\theta + \beta)]$$

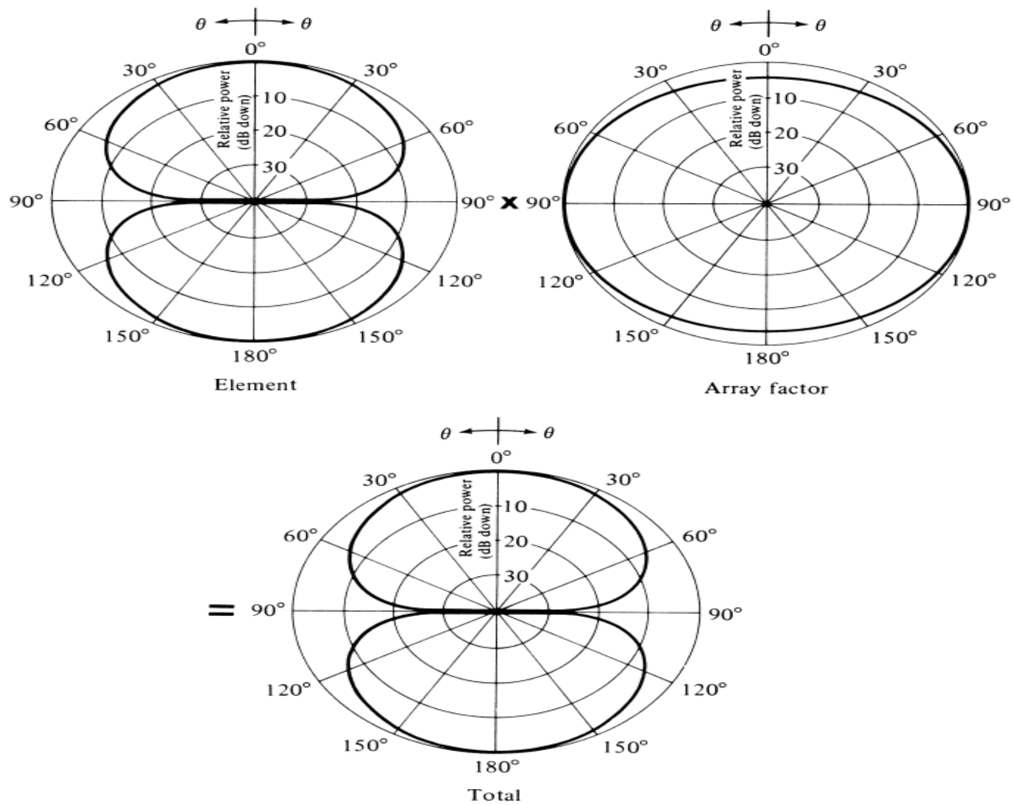
The array factor is a function of the geometry of the array and the excitation phase. By varying the separation d and/or the phase β betweenthe elements, the characteristics of the array factor and of the total field of the array can be controlled. It has been illustrated that the far-zone field of a uniform two-element array of identical elements is equal to the product of the field of a single element, at a selected reference point (usually the origin), and the array factor of that array. That is,

E(total) = [E(single element at reference point)] × [array factor] (6-5)

This is referred to as pattern multiplication for arrays of identical elements, and it is analogous to the pattern multiplication of (4-59) for continuous sources. Although it has been illustrated only for an array of two elements, each of identical magnitude, it is also valid for arrays with any number of identical elements which do not necessarily have identical magnitudes, phases, and/or spacings between them. This will be demonstrated in this chapter by a number of different arrays. Each array has its own array factor. The array factor, in general, is a function of the number of elements, their geometrical arrangement, their relative magnitudes, their relative phases, and their spacings. The array factor will be of simpler form if the elements have identical amplitudes, phases, and spacings. Since the array factor does not depend on the directional characteristics of the radiating elements themselves, it can be formulated by replacing the actual elements with isotropic (point) sources. Once the array factor has been derived using the point-source array, the total field of the actual array is obtained by the use of (6-5). Each point-source is assumed to have the amplitude, phase, and location of the corresponding element it is replacing. In order to synthesize the total pattern of an array, the designer is not only required to select the proper radiating elements but the geometry (positioning) and excitation of the individual elements.

To better illustrate the pattern multiplication rule, the normalized patterns of the single element, the array factor, and the total array for each of the above array examples are showninFigures 6.3, 6.4(a), and 6.4(b). In each figure, the total pattern of the array is obtained by multiplying the pattern of the single element by that of the array factor.

In each case, the pattern is normalized to its own maximum. Since the array factor for the example of Figure 6.3 is nearly isotropic (within 3 dB), the element pattern and the total pattern are almost identical in shape. The largest magnitude difference between the two is about 3 dB, and for each case it occurs toward the direction along which the phases of the two elements are in phase quadrature (90° out of phase). For Figure 6. this occurs along $\theta = 0^\circ$ while for Figures 6.4(a,b) this occurs along $\theta = 90^\circ$. Because the array factor for Figure 6.4(a) is of cardioid form, its corresponding element and total patterns are considerably different. In the total pattern, the null at $\theta = 90^\circ$ is due to the element pattern while that toward $\theta = 0^\circ$ is due to the array factor. Similar results are displayed inFigure 6.4(b).



LINEAR ARRAY:

Now that the arraying of elements has been introduced and it was illustrated by the two-element array, let us generalize the method to include N elements. Referring to the geometry of Figure 6.5(a), let us assume that all the elements have identical amplitudes but each succeeding element has a β progressive phase lead current excitation relative to the preceding one (β represents the phase by which the current in each element leads the current of the preceding element). An array of identical elements all of identical magnitude and each with a progressive phase is referred to as a uniform array. The array factor can be obtained by considering the elements to be point sources. If the actual elements are not isotropic sources, the total field can be formed by multiplying the array factor of the isotropic sources by the field of a single element. This is the pattern multiplication rule of (6-5), and it applies only for arrays of identical elements. The array factor is givenby

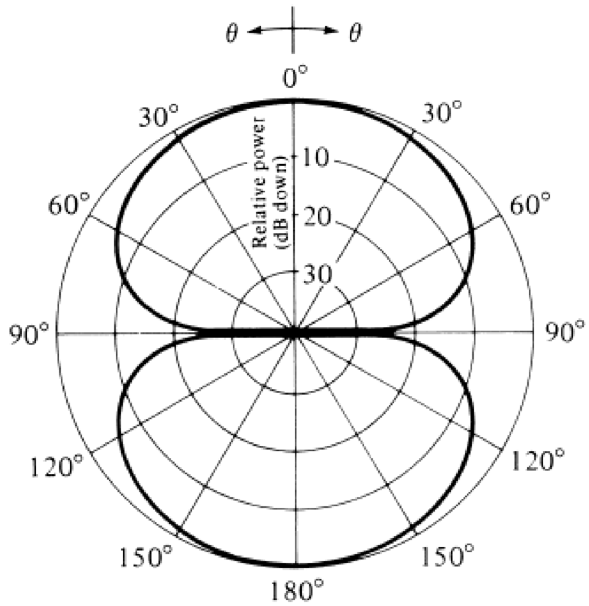
$$\mathrm{AF} = 1 + e^{+j(kd\cos\theta+\beta)} + e^{+j2(kd\cos\theta+\beta)} + \cdots + e^{j(N-1)(kd\cos\theta+\beta)}$$

$$\mathrm{AF} = \sum_{n=1}^{N} e^{j(n-1)(kd\cos\theta+\beta)}$$

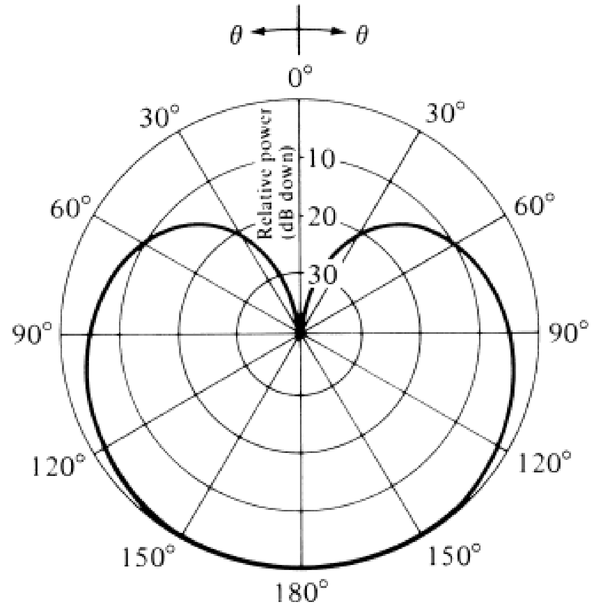
which canbe written as

$$\mathrm{AF} = \sum_{n=1}^{N} e^{j(n-1)\psi}$$

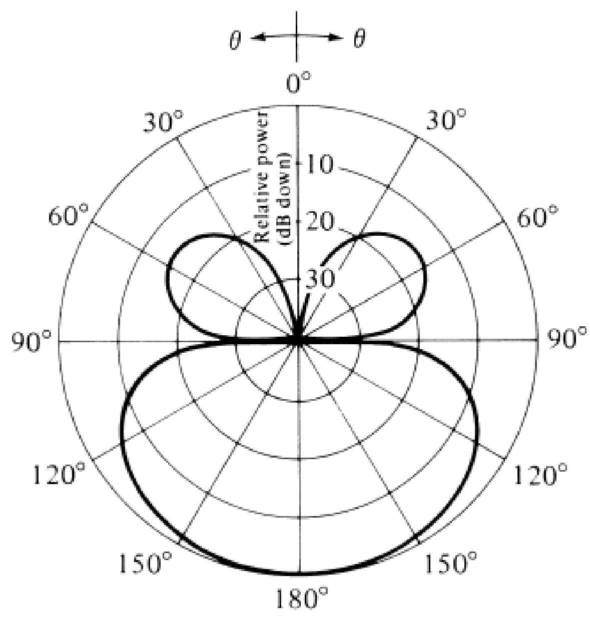
$$\text{where} \quad \psi = kd\cos\theta + \beta$$



Element



Array factor



Total

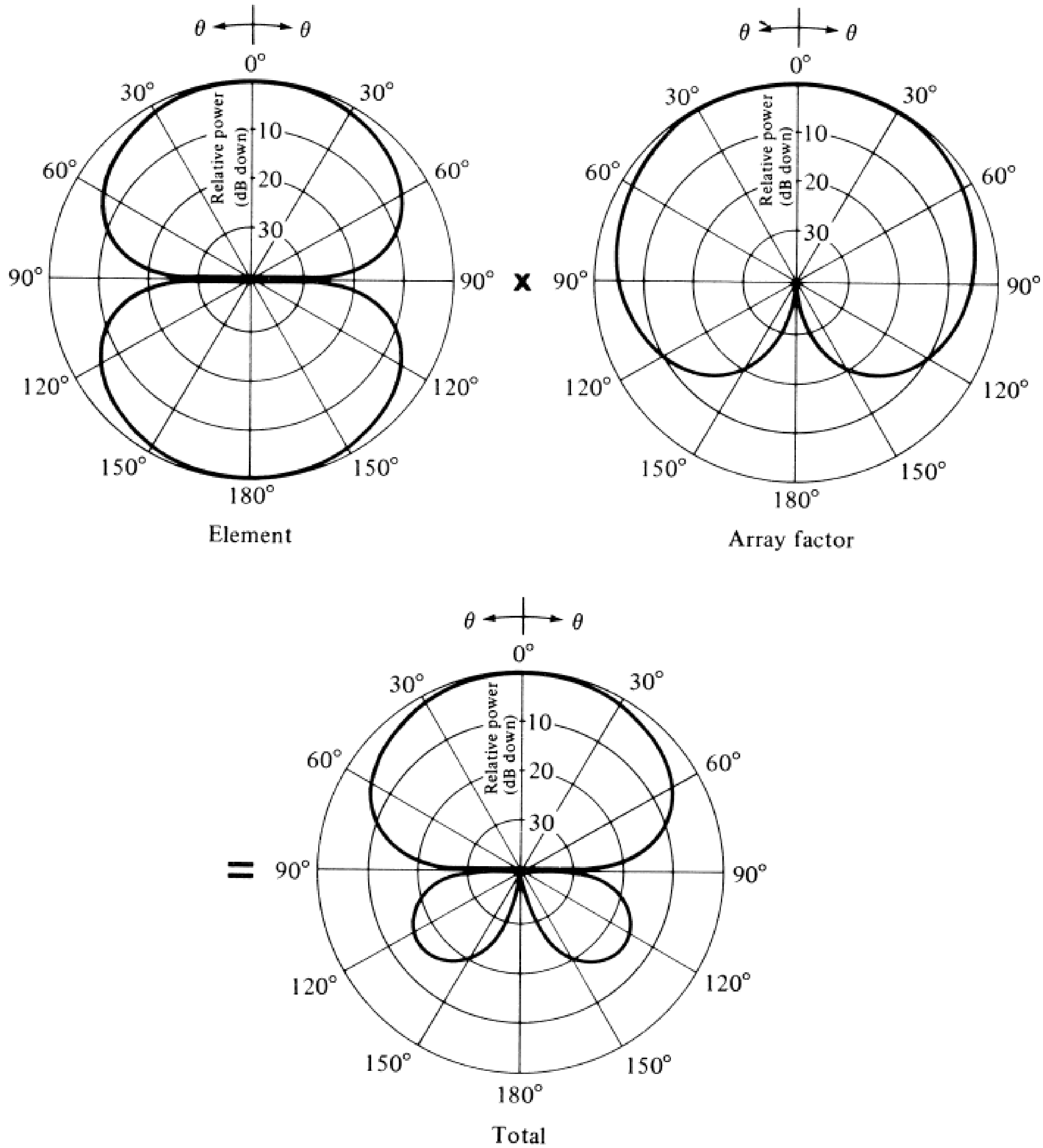
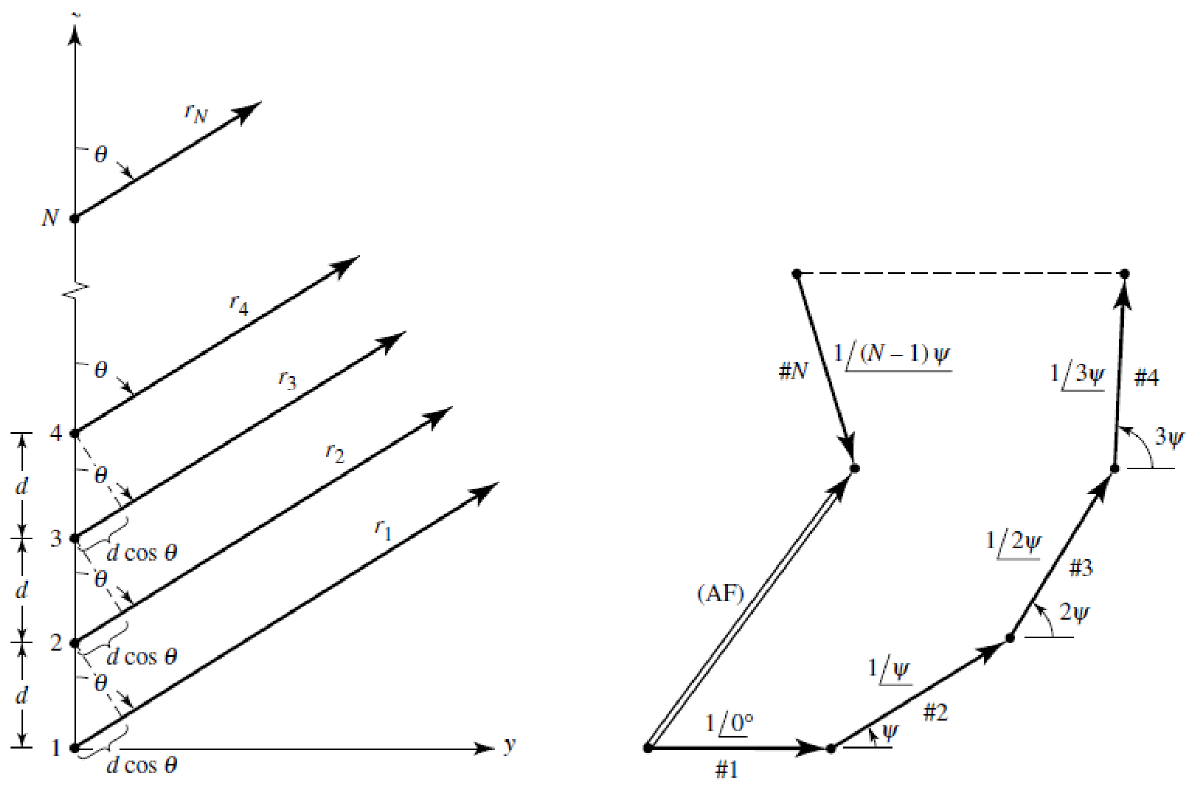


Figure 6.4 (b) $\beta = -90°$, $d = \lambda/4$ (*continued*).

Since the total array factor for the uniform array is a summation of exponentials, it canbe represented by the vector sum of N phasors each of unit amplitude and progressive phase ψ relative to the previous one. Graphically this is illustrated by the phasor diagram inFigure 6.5(b). It is apparent from the phasor diagram that the amplitude and phase of the AF can be controlled in uniform arrays by properly selecting the relative phase ψ between the elements; in nonuniform arrays, the amplitude as well as the phase can be used to control the formation and distribution of the total array factor.The array factor of (6-7) canalso be expressed in analtern ate, compact and closed form whose functions and their distributions are more recognizable. This is accomplished as follows.

Multiplying both sides of (6-7) by ejψ, it canbe written as

$$(\mathrm{AF})e^{j\psi} = e^{j\psi} + e^{j2\psi} + e^{j3\psi} + \cdots + e^{j(N-1)\psi} + e^{jN\psi}$$



(a) Geometry (b) Phasor diagram

Subtracting (6-7) from (6-8) reduces to

$$\mathrm{AF}(e^{j\psi} - 1) = (-1 + e^{jN\psi})$$

which canalso be written as

$$\mathrm{AF} = \left[\frac{e^{jN\psi} - 1}{e^{j\psi} - 1}\right] = e^{j[(N-1)/2]\psi}\left[\frac{e^{j(N/2)\psi} - e^{-j(N/2)\psi}}{e^{j(1/2)\psi} - e^{-j(1/2)\psi}}\right]$$

$$= e^{j[(N-1)/2]\psi}\left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{1}{2}\psi\right)}\right]$$

If the reference point is the physical center of the array, the array factor of (6-10)reduces to For small values of ψ, the above expressioncan be approximated by

$$\mathrm{AF} = \left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{1}{2}\psi\right)}\right]$$

The maximum value of (6-10a) or (6-10b) is equal to N. To normalize the array factors so that the maximum value of each is equal to unity, (6-10a) and (6-10b) are written in normalized form as

$$\mathrm{AF} \simeq \left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\frac{\psi}{2}}\right]$$

$$(\mathrm{AF})_n = \frac{1}{N}\left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{1}{2}\psi\right)}\right]$$

BROADSIDE ARRAY:

In many applications it is desirable to have the maximum radiation of an array directed normal to the axis of the array [broadside; θ0 = 90◦ of Figure 6.5(a)]. To optimize the design, the maxima of the single element and of the array factor should both be directed toward θ0 = 90◦. The requirements of the single elements

can be accomplished by the judicious choice of the radiators, and those of the array factor by the proper separation and excitation of the individual radiators. In this section, the requirements that allow the array factor to "radiate" efficiently broadside will be developed.
Referring to (6-10c) or (6-10d), the first maximum of the array factor occurs when

$$\psi = kd\cos\theta + \beta = 0$$

Since it is desired to have the first maximum directed toward θ0 = 90◦, then

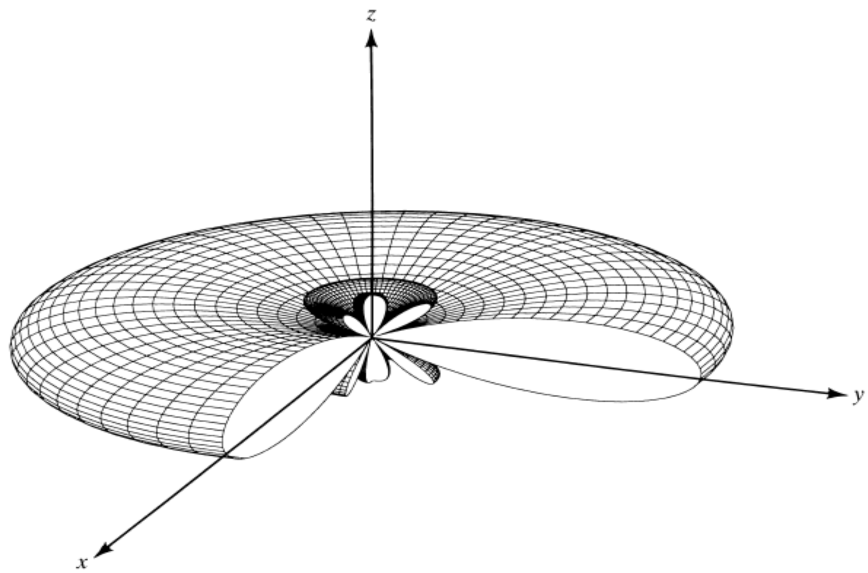
$$\psi = kd\cos\theta + \beta|_{\theta=90^\circ} = \beta = 0$$

Thus to have the maximum of the array factor of a uniform linear array directed broadside to the axis of the array, it is necessary that all the elements have the same phase excitation(in addition to the same amplitude excitation). The separation between the elements can be of any value. To ensure that there are no principal max ma in other directions, which are referred to as grating lobes, the separationbetweenthe elements should not be equal to multiples of a wavelength (d _= nλ, n = 1, 2, 3 . . .) when β = 0.
If d = nλ, n = 1, 2, 3, . . . and β = 0, then

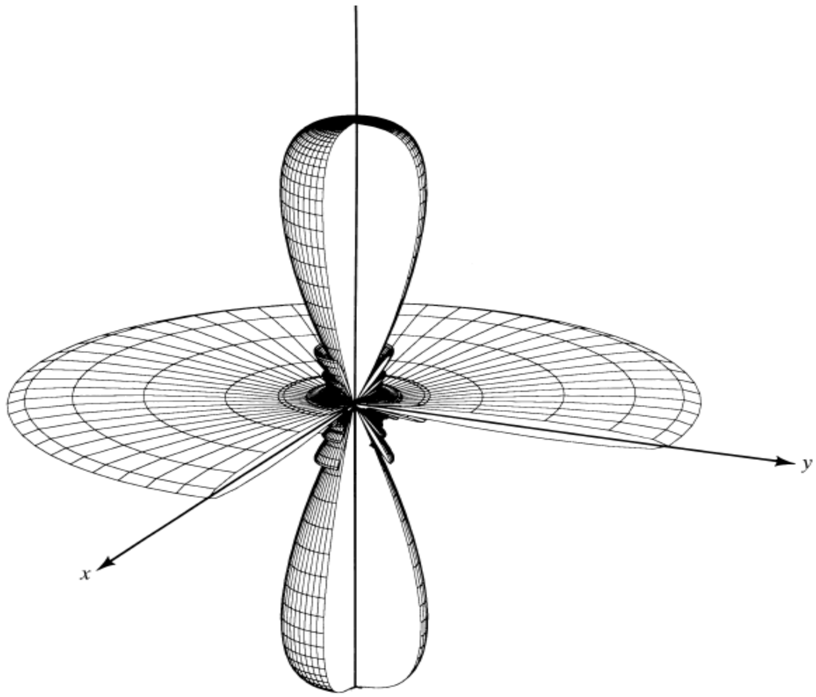
$$\psi = kd\cos\theta + \beta\Big|_{\substack{d=n\lambda \\ \beta=0 \\ n=1,2,3,\ldots}} = 2\pi n\cos\theta\Big|_{\theta=0^\circ,180^\circ} = \pm 2n\pi$$

This value of ψ when substituted in(6-10c) makes the array factor attain its maximum value. Thus for a uniform array with β = 0 an dd = nλ, in addition to having the maxima of the array factor directed broadside (θ0 = 90◦) to the axis of the array, there are additional maxima directed along the axis (θ0 = 0◦ , 180◦) of the array (endfire radiation).
One of the objectives in many designs is to avoid multiple maxima, in addition to the main maximum, which are referred to as grating lobes. Often it may be required to select the largest spacing between the elements but with no grating lobes. To avoid any grating lobe, the largest spacing between the elements should be less than one wavelength (dmax < λ). To illustrate the method, the three-dimensional array factor of a 10-element (N = 10) uniform array with β = 0 an dd = λ/4 is shown plotted in Figure 6.6(a). A 90◦ angular sector has been removed for better view of the pattern distribution in the elevation plane. The only maximum

occurs at broadside (θ0 = 90◦). To form a comparison, the three-dimensional pattern of the same array but with d = λ is also plotted in Figure 6.6(b). For this pattern, in addition to the maximum at θ0 = 90◦, there are
additional maxima directed toward θ0 = 0◦, 180◦. The corresponding two-dimensional patterns of Figures 6.6(a,b) are showninFigure 6.7.
If the spacing between the elements is chosen between λ < d < 2λ, then the maximum of Figure 6.6 toward θ0 = 0◦ shifts toward the angular region 0◦< θ0 < 90◦ while the maximum toward θ0 = 180◦ shifts toward 90◦< θ0 < 180◦. When d = 2λ,there are maxima toward 0◦, 60◦ , 90◦ , 120◦ and 180◦.



(a) Broadside ($\beta = 0$, $d = \lambda/4$)



(b) Broadside/end-fire ($\beta = 0$, $d = \lambda$)

END-FIRE ARRAY:

Instead of having the maximum radiation broadside to the axis of the array, it may be desirable to direct it along the axis of the array (end-fire). As a matter of fact, it may be necessary that it radiates toward only one direction (either θ0 = 0◦ or 180◦ of Figure 6.5).To direct the first maximum toward θ0 = 0◦,

$$\psi = kd\cos\theta + \beta|_{\theta=0^\circ} = kd + \beta = 0 \Rightarrow \beta = -kd$$

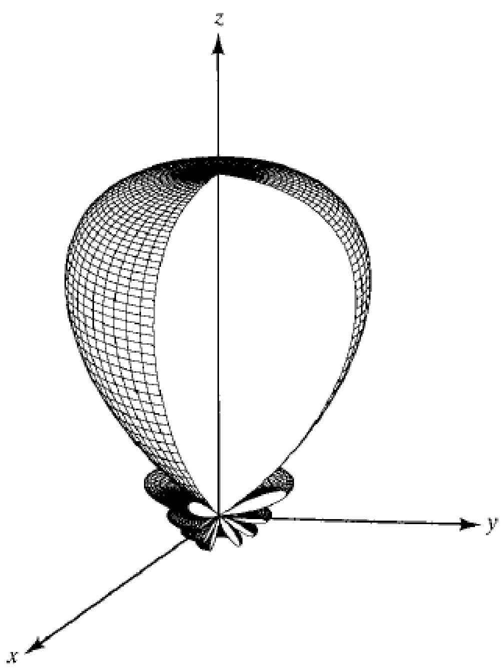
If the first maximum is desired toward θ0 = 180◦, then

$$\psi = kd\cos\theta + \beta|_{\theta=180^\circ} = -kd + \beta = 0 \Rightarrow \beta = kd$$

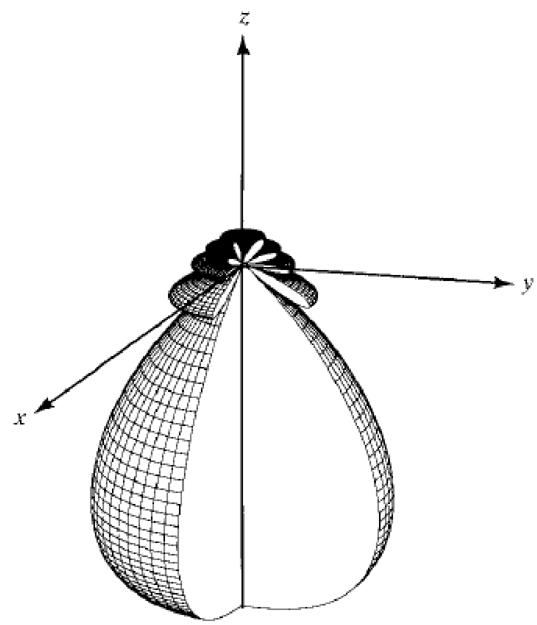
Thus end-fire radiation is accomplished when β = −kd (for θ0 = 0◦) or β = kd (for θ0 = 180◦).

If the element separation is d = λ/2, end-fire radiation exists simultaneously in both directions (θ0 = 0◦ and θ0 = 180◦). If the element spacing is a multiple of a wavelength (d = nλ, n = 1, 2, 3, . . .), then inadditionto having end-fire radiation in both directions, there also exist maxima in the broadside directions. Thus for d = nλ, n = 1, 2, 3, . . . there exist four maxima; two in the broadside directions and two along the axis of the array. To have only one end-fire maximum and to avoid any grating lobes, the maximum spacing between the elements should be less than

dmax < λ/2. The three-dimensional radiation patterns of a 10-element (N = 10) array with d =
λ/4, β = +kd are plotted inFigure 6.8. When β = −kd, the maximum is directed along θ0 = 0◦ and the three-dimensional pattern is shown in Figure 6.8(a). However, when β = +kd, the maximum is oriented toward θ0 = 180◦, and the three-dimensional pattern is shown in Figure 6.8(b). The two-dimensional patterns of Figures 6.8(a,b) are shown in Figure 6.9. To form a comparison, the array factor of the same array (N = 10) but with d = λ and β = −kd has been calculated. Its pattern is identical to that of a broadside array with N = 10, d = λ, and it is shown plotted in Figure 6.7. It is seen that there are four maxima; two broadside and two along the axis of the array.



(a) $\theta_0 = 0°$

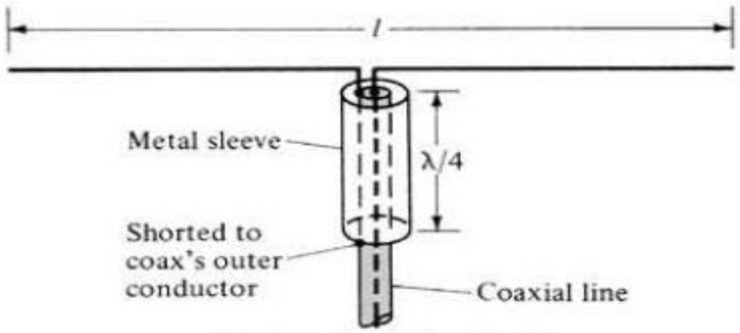


(b) $\theta_0 = 180°$

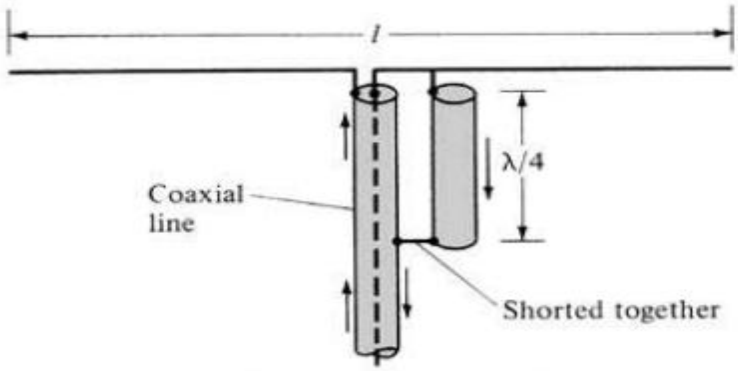
BALUNS:

A twin-lead transmission line (two parallel-conductor line) is a symmetrical line whereas a coaxial cable is inherently unbalanced. Because the inner and outer (inside and outside parts of it) conductors of the coax are not coupled to the antenna in the same way, they provide the unbalance. The result is a net current flow to ground on the outside part of the outer conductor. This is showninFigure 9.25(a) where anelectrical equivalent is also indicated. The amount of current flow I3 onthe outside surface of the outer conductor is determined by the impedance Zg from the outer shield to ground. If Zg canbe made very large, I3 can be reduced significantly. Devices that can be used to balance inherently unbalanced systems, by canceling or choking the outside current, are known as baluns (balance to unbalance).

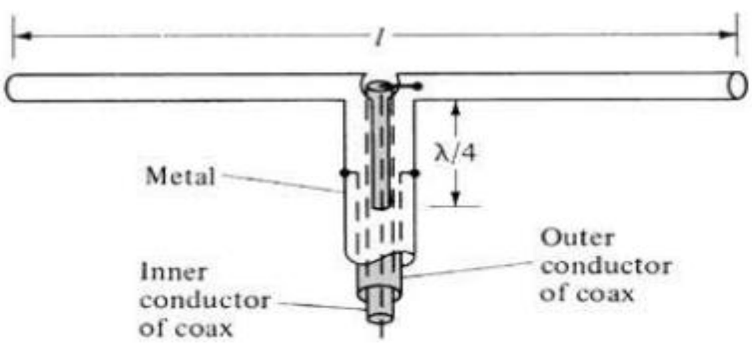
One type of a balunis that shown inFigure 9.25(b), referred to usually as a bazooka balun. Mechanically it requires that a $\lambda/4$ in length metal sleeve, and shorted at its one end, encapsulates the coaxial line. Electrically the input impedance at the open end of this $\lambda/4$ shorted transmission line, which is equivalent to Zg, will be very large (ideally infinity). Thus the current I3 will be choked, if not completely eliminated, and the system will be nearly balanced.



(b) Bazooka balun (1 : 1)



(c) λ/4 coaxial balun (1 : 1)



(d) Coaxial balun (1 : 1)

Another type of a balun is that shown in Figure 9.25(c). It requires that one end of a λ/4 section of a transmission line be connected to the outside shield of the main coaxial line while the other is connected to the side of the dipole which is attached to the center conductor. This balun is used to cancel the flow of I3. The operation of it can be explained as follows: In Figure 9.25(a) the voltages between each side of the dipole and the ground are equal in magnitude but 180◦ out of phase, thus producing acurrent flow on the outside of the coaxial line. If the two currents I1 and I2 are equal in magnitude, I3 would be zero. Since arm #2 of the dipole is connected directly to the shield of the coax while arm #1 is weakly coupled to it, it produces a much larger current I2. Thus there is relatively little cancellation in the two currents.

The two currents, I1 and I2, can be made equal in magnitude if the center conductor of the coax is connected directly to the outer shield. If this connection is made directly at the antenna terminals, the transmission line and the antenna would be short-circuited, thus eliminating any radiation. However, the indirect parallel-conductor connection of Figure 9.25(c) provides the desired current cancellation without eliminating the radiation.

The current flow on the outer shield of the main line is canceled at the bottom end of the λ/4 section(where the two jointogether) by the equal inmagn itude, but opposite inphase, current in the λ/4 sectionof the auxiliary line. Ideally then there is no current flow in the outer surface of the outer shield of the remaining part of main coaxial line. It should be stated that the parallel auxiliary line need not be made λ/4 in length to achieve the balance. It is made λ/4 to prevent the upsetting of the normal operation of the antenna.

A compact constructionof the baluninFigure 9.25(c) is that in Figure 9.25(d). The outside metal sleeve is split and a portionof it is removed onopposite sides. The remaining opposite parts of the outer sleeve represent electrically the two shorted λ/4 parallel transmission lines of Figure 9.25(c). All of the baluns shown in Figure 9.25 are narrowband devices.

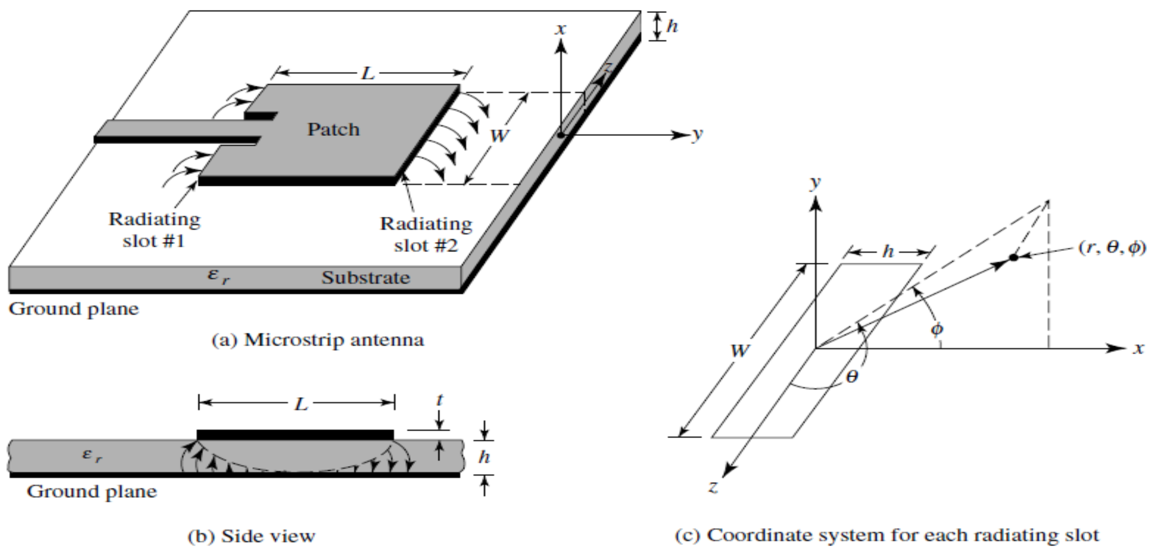
Microstrip Antennas

Introduction

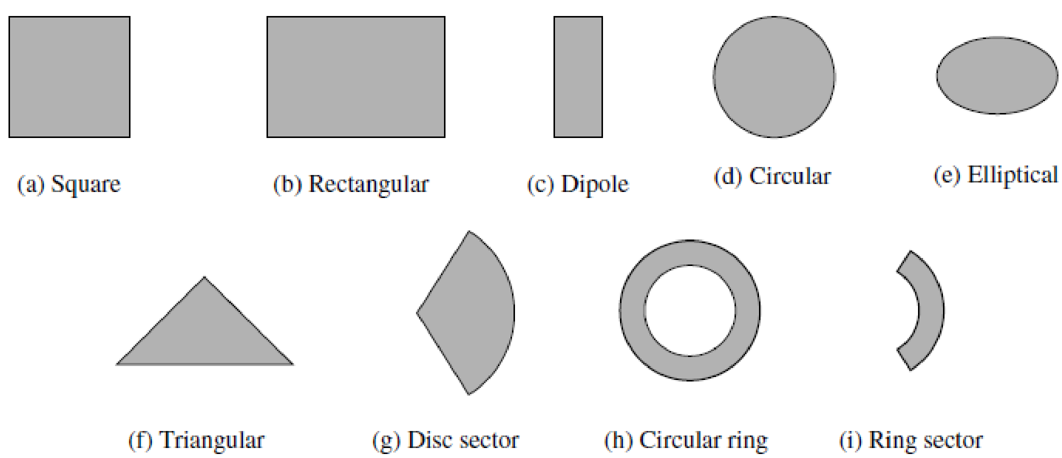
In high-performance aircraft, spacecraft, satellite, and missile applications, where size, weight, cost, performance, ease of installation, and aerodynamic profile are constraints, low-profile antennas may be required. Presently there are many other government and commercial applications, such as mobile radio and wireless communications that have similar specifications. To meet these requirements, microstrip antennas can be used. These antennas are low profile, conformable to planar and nonplanar surfaces, simple and inexpensive to manufacture using modern printed-circuit technology, mechanically robust when mounted on rigid surfaces, compatible with MMIC designs, and when the particular patch shape and mode are selected, they are very versatile in terms of resonant frequency, polarization, pattern, and impedance. In addition, by adding loads between the patch and the ground plane, such as pins and varactor diodes, adaptive elements with variable resonant frequency, impedance, polarization, and pattern can be designed. Major operational disadvantages of microstrip antennas are their low efficiency, low power, high Q (sometimes in excess of 100), poor polarization purity, poor scan performance, spurious feed radiation and very narrow frequency bandwidth, which is typically only a fraction of a percent or at most a few percent.

Basic characteristics

Microstrip antennas received considerable attention starting in the 1970s, although the idea of a microstrip antenna can be traced to 1953 and a patent in 1955. Microstrip antennas consist of a very thin (t _ $\lambda 0$, where $\lambda 0$ is the free-space wavelength) metallic strip (patch) placed a small fraction of a wavelength ($h < \lambda 0$, usually $0.003\lambda 0 \leq h \leq 0.05\lambda 0$) above a ground plane. The microstrip patch is designed so its pattern maximum is normal to the patch (broadside radiator). This is accomplished by properly choosing the mode (field configuration) of excitation beneath the patch. End-fire radiation can also be accomplished by judicious mode selection. For a rectangular patch, the length L of the element is usually $\lambda 0/3 < L < \lambda 0/2$. The strip (patch) and the ground plane are separated by a dielectric sheet (referred to as the substrate). There are numerous substrates that can be used for the design of microstrip antennas, and their dielectric constants are usually in the range of $2.2 \leq I_r \leq 12$. The ones that are most desirable for good antenna performance are thick substrates whose dielectric constant is in the lower end of the range because they provide better efficiency, larger bandwidth, loosely bound fields for radiation into space, but at the expense of larger element size. Thin substrates with higher dielectric constants are desirable for microwave circuitry because they require tightly bound fields to minimize undesired radiation and coupling, and lead to smaller element sizes; however, because of their greater losses, they are less efficient and have relatively smaller bandwidths.



Microstrip antenna and coordinate system.

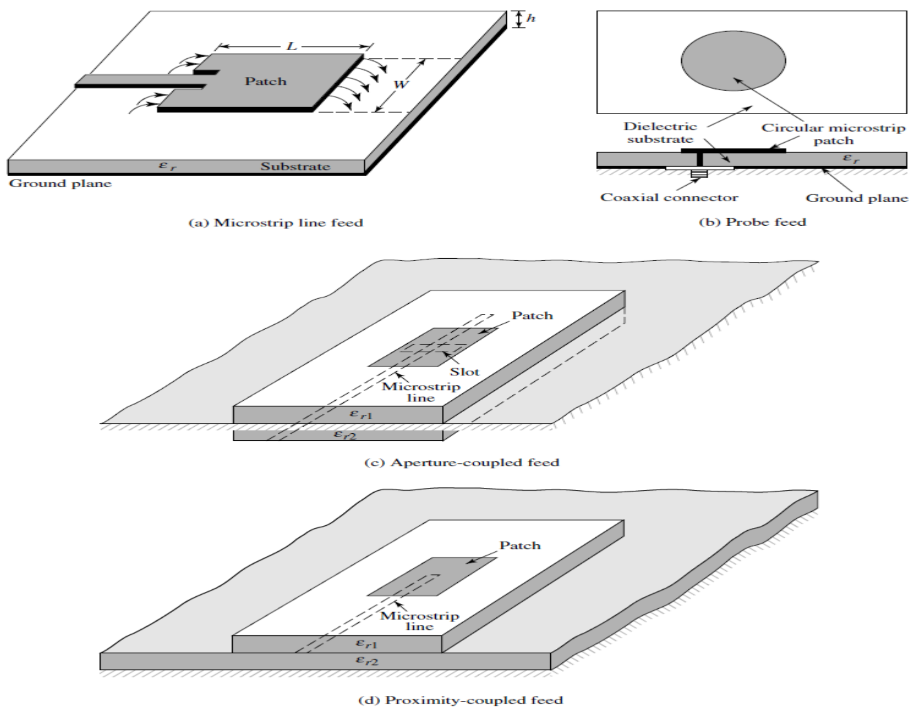


Representative shapes of microstrip patch elements.

Often microstrip antennas are also referred to as *patch* antennas. The radiating elements and the feed lines are usually photo etched on the dielectric substrate. The radiating patch may be square, rectangular, thin strip (dipole), circular, elliptical, triangular or any other configuration. These and others are Square, rectangular, dipole (strip), and circular are the most common because of ease of analysis and fabrication, and their attractive radiation characteristics, especially low cross-polarization radiation.

Feeding methods

There are many configurations that can be used to feed microstrip antennas. The four most popular are the microstrip line, coaxial probe, aperture coupling, and proximity coupling. One set of equivalent circuits for each one of these is shown in Figure 14.4. The microstrip feed line is also a conducting strip, usually of much smaller width compared to the patch. The microstrip-line feed is easy to fabricate, simple to match by controlling the inset position and rather simple to model. However as the substrate thickness increases, surface waves and spurious feed radiation increase, which for practical designs limit the bandwidth (typically 2–5%). Coaxial-line feeds, where the inner conductor of the coax is attached to the radiation patch while the outer conductor is connected to the ground plane, are also widely used. The coaxial probe feed is also easy to fabricate and match, and it has low spurious radiation. However, it also has narrow bandwidth and it is more difficult to model, especially for thick substrates $(h > 0.02\lambda 0)$. Both the microstrip feed line and the probe possesses inherent asymmetries which generate higher order modes which produce cross-polarized radiation. To overcome some of these problems, non contacting aperture-coupling feeds have been introduced. The aperture coupling is the most difficult of all four to fabricate and it also has narrow bandwidth. However, it is somewhat easier to model and has moderate spurious radiation. The aperture coupling consists of two substrates separated by a ground plane. On the bottom side of the lower substrate there is a microstrip feed line whose energy is coupled to the patch through a slot on the ground plane separating the two substrates. This arrangement allows independent optimization of the feed mechanism and the radiating element. Typically a high dielectric material is used for the bottom substrate, and thick low dielectric constant material for the top substrate. The ground plane between the substrates also isolates the feed from the radiating element and minimizes interference of spurious radiation for pattern formation and polarization purity. For this design, the substrate electrical parameters, feed line width, and slot size and position can be used to optimize the design [38]. Typically matching is performed by controlling the width of the feed line and the length of the slot. The coupling through the slot can be modeled using the theory of Bethe which is also used to account for coupling through a small aperture in a conducting plane. This theory has been successfully used to analyze waveguide couplers using coupling through holes. In this theory the slot is represented by an equivalent normal electric dipole to account for the normal component (to the slot) of the electric field, and an equivalent horizontal magnetic dipole to account for the tangential component (to the slot) magnetic field. If the slot is centered below the patch, where ideally for the dominant mode the electric field is zero while the magnetic field is maximum, the magnetic coupling will dominate. Doing this also leads to good polarization purity and no cross-polarized radiation in the principal planes. Of the four feeds described here, the proximity coupling has the largest bandwidth (as high as 13 percent), is somewhat easy to model and has low spurious radiation. However its fabrication is somewhat more difficult. The length of the feeding stub and the width-to-line ratio of the patch can be used to control the match.



Typical feeds for microstrip antennas.

<u>Rectangular patch</u>

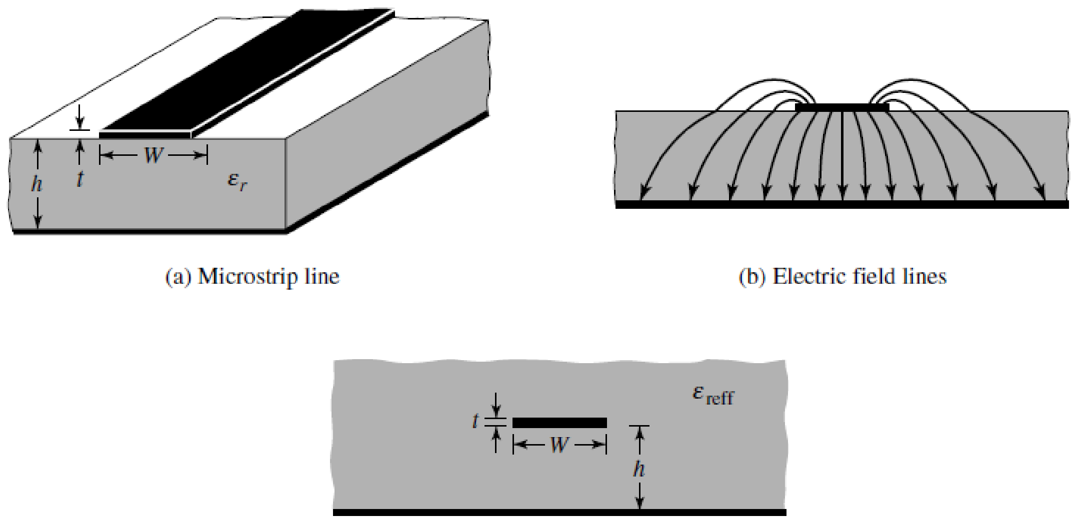
The rectangular patch is by far the most widely used configuration. It is very easy to analyze using both the transmission-line and cavity models, which are most accurate for thin substrates. We begin with the transmission-line model because it is easier to illustrate.

Transmission-Line Model

It was indicated earlier that the transmission-line model is the easiest of all but it yields the least accurate results and it lacks the versatility. However, it does shed some physical insight. A rectangular microstrip antenna can be represented as an array of two *radiating* narrow apertures (slots), each of width *W* and height *h*, separated by a distance *L*. Basically the transmission-line model represents the microstrip antenna by two slots, separated by a low-impedance *Zc* transmission line of length *L*.

<u>Fringing Effects</u>

Because the dimensions of the patch are finite along the length and width, the fields at the edges of the patch undergo fringing. This is illustrated along the length for the two radiating slots of the microstrip antenna. The same applies along the width. The amount of fringing is a function of the dimensions of the patch and the height of the substrate. For the principal *E*-plane (*xy*-plane) fringing is a function of the ratio of the length of the patch *L* to the height *h* of the substrate (*L/h*) and the dielectric constant L_r of the substrate. Since for microstrip antennas *L/h* <1, fringing is reduced; however, it must be taken into account because it influences the resonant frequency of the antenna. The same applies for the width. This is a nonhomogeneous line of two dielectrics; typically the substrate and air. As can be seen, most of the electric field lines reside in the substrate and parts of some lines exist in air. As *W/h*< 1 an d*Ir* <1, the electric field lines concentrate mostly in the substrate. Fringing in this case makes the microstrip line look wider electrically compared to its physical dimensions. Since some of the waves travel in the substrate and some in air, an *effective dielectric constant* L_{reff} is introduced to account for fringing and the wave propagation in the line. To introduce the effective dielectric constant, let us assume that the center conductor of the microstrip line with its original dimensions and height above the ground plane is embedded into one dielectric. The effective dielectric constant is defined *as the dielectric constant of the uniform dielectric material so that the line has identical electrical characteristics, particularly propagation constant, as the actual line of Figure 14.5(a)*. For a line with air above the substrate, the effective dielectric constant has values in the range of $1 < L_{reff} < Ir$. For most applications where the dielectric constant of the substrate is much greater than unity *(*$L_r < 1$*)*, the value of L_{reff} will be closer to the value of the actual dielectric constant *Ir* of the substrate. The effective dielectric constant is also a function of frequency. As the frequency of operation increases, most of the electric field lines concentrate in the substrate. Therefore the microstrip line behaves more like a homogeneous line of one dielectric (only the substrate), and the effective dielectric constant approaches the value of the dielectric constant of the substrate.



(a) Microstrip line

(b) Electric field lines

(c) Effective dielectric constant

Microstrip line and its electric field lines, and effective dielectric constant

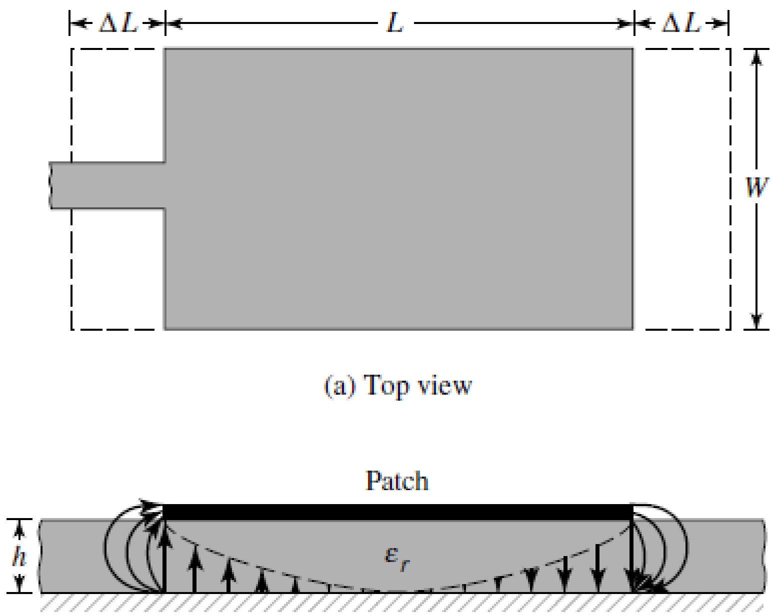
For low frequencies the effective dielectric constant is essentially constant. At intermediate frequencies its values begin to monotonically increase and eventually approach the values of the dielectric constant of the substrate. The initial values (at low frequencies) of the effective dielectric constant are referred to as the *static values*, and they are given by

$$W/h > 1$$

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left[1 + 12\frac{h}{W}\right]^{-1/2}$$

Because of the fringing effects, electrically the patch of the microstrip antenna looks greater than its physical dimensions. For the principal *E*-plane (*xy*-plane), where the dimensions of the patch along its length have been extended on each end by a distance *3L*, which is a function of the effective dielectric constant L_{reff} and the width-to-height ratio (*W/h*). A very popular and practical approximate relation for the normalized extension of the length is

$$\frac{\Delta L}{h} = 0.412\frac{(\varepsilon_{reff} + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\varepsilon_{reff} - 0.258)\left(\frac{W}{h} + 0.8\right)}$$



(a) Top view

(b) Side view

Physical and effective lengths of rectangular microstrip patch.

Since the length of the patch has been extended by *3L* on each side, the effective length of the patch is now ($L = \lambda/2$ for dominant TM010 mode with no fringing)

$$L_{eff} = L + 2\Delta L$$

For the dominant TM_{010} mode, the resonant frequency of the microstrip antenna is a function of its length. Usually it is given by

$$(f_r)_{010} = \frac{1}{2L\sqrt{\varepsilon_r}\sqrt{\mu_0\varepsilon_0}} = \frac{v_0}{2L\sqrt{\varepsilon_r}}$$

where $v0$ is the speed of light in free space. Since (14-4) does not account for fringing, it must be modified to include edge effects and should be computed using

$$(f_{rc})_{010} = \frac{1}{2L_{eff}\sqrt{\varepsilon_{reff}}\sqrt{\mu_0\varepsilon_0}} = \frac{1}{2(\mathrm{L}+2\Delta\,\mathrm{L})\sqrt{\varepsilon_{reff}}\sqrt{\mu_0\varepsilon_0}}$$

$$= q\frac{1}{2L\sqrt{\varepsilon_r}\sqrt{\mu_0\varepsilon_0}} = q\frac{v_0}{2L\sqrt{\varepsilon_r}}$$

$$q = \frac{(f_{rc})_{010}}{(f_r)_{010}}$$

The q factor is referred to as the *fringe factor* (length reduction factor). As the substrate height increases, fringing also increases and leads to larger separations between the radiating edges and lower resonant frequencies.

RADIATION PATTERN

For a linearly polarized antenna, performance is often described in terms of its principal *E*- and *H* -plane patterns. The *E-plane* is defined as "the plane containing the electric-field vector and the direction of maximum radiation," and the *H-plane* as "the plane containing the magnetic-field vector and the direction of maximum radiation." Although it is very difficult to illustrate the principal patterns without considering a specific example, it is the usual practice to orient most antennas so that at least one of the principal plane patterns coincide with one of the geometrical principal planes. For this example, the *x*-*z* plane (elevation plane; $\Phi = 0$) is the principal *E*-plane and the *x*-*y* plane (azimuthal plane; $\Phi = \pi/2$) is the principal *H* -plane. Other coordinate orientations can be selected. The omnidirectional pattern of Figure 2.6 has an infinite number of principal E-planes (elevation planes; $\Phi = \Phi c$) and one principal *H*-plane (azimuthal plane; $\theta = 90^\circ$).

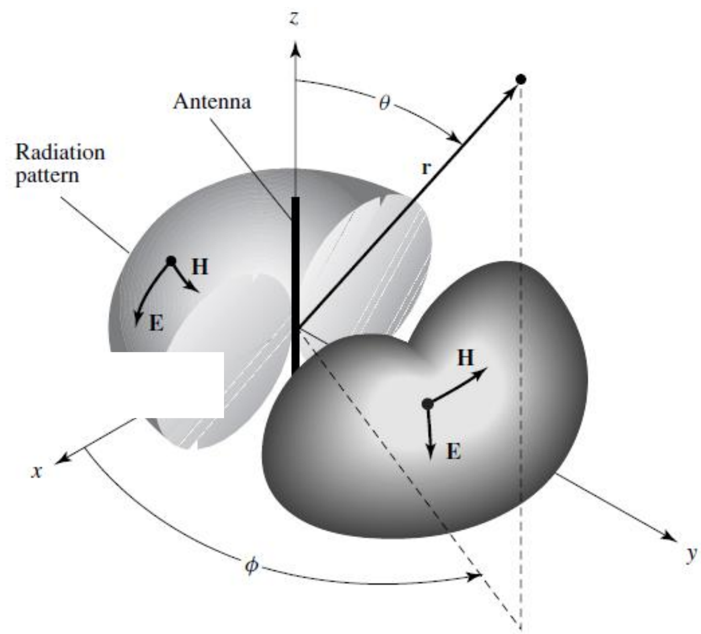


Figure 2.6 Omnidirectional antenna pattern.

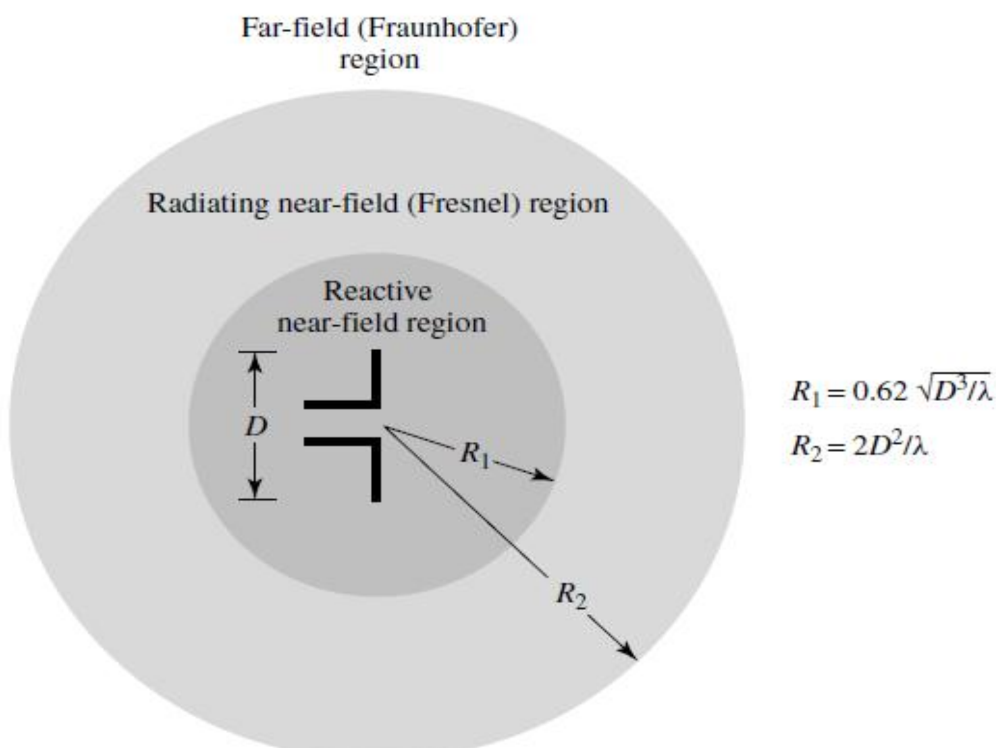


Figure 2.7 Field regions of an antenna.

Cavity Model

Microstrip antennas resemble dielectric-loaded cavities, and they exhibit higher order resonances. The normalized fields within the dielectric substrate (between the patch and the ground plane) can be found more accurately by treating that region as a cavity bounded by electric conductors (above and below it) and by magnetic walls (to simulate an open circuit) along the perimeter of the patch. This is an approximate model, which in principle leads to a reactive input impedance (of zero or infinite value of resonance), and it does not radiate any power. However, assuming that the actual fields are approximate to those generated by such a model, the computed pattern, input admittance, and resonant frequencies compare well with measurements. This is an accepted approach, and it is similar to the perturbation methods which have been very successful in the analysis of waveguides, cavities, and radiators. To shed some insight into the cavity model, let us attempt to present a physical interpretation in to the formation of the fields within the cavity and radiation through its side walls. When the microstrip patch is energized, a charge distribution is established on the upper and lower surfaces of the patch, as well as on the surface of the ground plane.

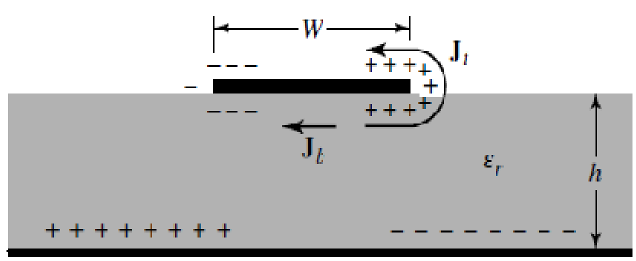


Figure 14.12 Charge distribution and current density creation on microstrip patch.

The charge distribution is controlled by two mechanisms; an *attractive* and a *repulsive* mechanism [34]. The *attractive* mechanism is between the corresponding opposite charges on the bottom side of the patch and the ground plane, which tends to maintain the charge concentration on the bottom of the patch. The *repulsive* mechanism is between like charges on the bottom surface of the patch, which tends to push some charges from the bottom of the patch, around its edges, to its top surface. The movement of these charges creates corresponding current densities $\mathbf{J}_b$ and $\mathbf{J}_t$, at the bottom and top surfaces of the patch, respectively. Since for most practical microstrip the height-to-width ratio is very small, the attractive mechanism dominates and most of the charge concentration and current flow remain underneath the patch. A small amount of current flows around the edges of the patch to its top surface. However, this current flow decreases as the height-to-width ratio decreases. In the limit, the current flow to the top would be zero, which ideally would not create any tangential magnetic field components to the edges of the patch. This would allow the four side walls to be modelled as perfect magnetic conducting surfaces which ideally would not disturb the magnetic field and, in turn, the electric field distributions beneath the patch. Since in practice there is a finite height-to-width ratio, although small, the tangential magnetic fields at the edges would not be exactly zero. However, since they will be small, a good approximation to the cavity model is to treat the side walls as perfectly magnetic conducting. This model produces good normalized electric and magnetic field distributions (modes) beneath the patch. If the microstrip antenna were treated only as a cavity, it would not be sufficient to find the absolute amplitudes of the electric and magnetic fields. In fact by treating the walls of the cavity, as well as the material within it as lossless, the cavity would not radiate and its input impedance would be purely reactive. Also the function representing the impedance would only have real poles. To account for radiation, a loss mechanism has to be introduced. This was taken into account by the radiation resistance R_r and loss resistance R_L. These two resistances allow the input impedance to be complex and for its function to have complex poles; the imaginary poles representing, through R_r and R_L, the radiation and conduction-dielectric losses. To make the microstrip lossy using the cavity model, which would then represent an antenna, the loss is taken into account by introducing an effective loss tangent δ_{eff}. The effective loss tangent is chosen appropriately to represent the loss mechanism of the cavity, which now behaves as an antenna and is taken as the reciprocal of the antenna quality factor Q ($\delta_{eff} = 1/Q$).

Antenna Measurements

Radiation Pattern

An antenna *radiation pattern* or *antenna pattern* is defined as “a mathematical function or a graphical representation of the radiation properties of the antenna as a function of space coordinates. In most cases, the radiation pattern is determined in the farfield region and is represented as a function of the directional coordinates. Radiation properties include power flux density, radiation intensity, field strength, directivity, phase or polarization.” The radiation property of most concern is the two- or three -dimensional spatial distribution of radiated energy as a function of the observer’s position along a path or surface of constant radius. A convenient set of coordinates is shown in Figure 2.1. A trace of the received electric (magnetic) field at a constant radius is called the amplitude field *pattern*. On the other hand, a graph of the spatial variation of the power density along a constant radius is called an amplitude *power pattern*. Often the *field* and *power* patterns are normalized with respect to their maximum value, yielding *normalized field* and *power patterns*. Also, the power pattern is usually plotted on a logarithmic scale or more commonly in decibels (dB). This scale is usually

desirable because a logarithmic scale can accentuate in more details those parts of the pattern that have very low values, which later we will refer to as minor lobes. For an antenna, the

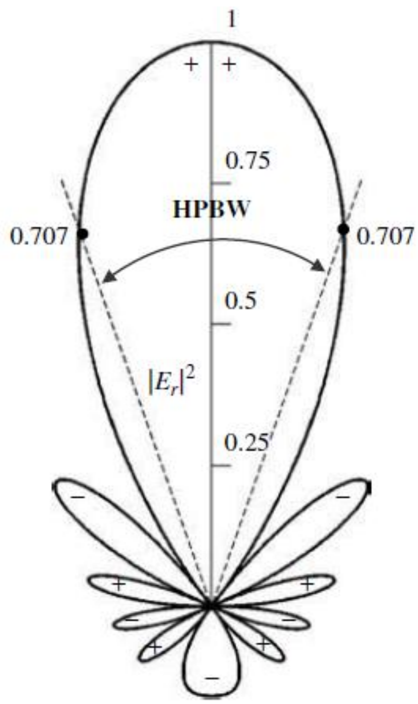
a. *field* pattern(*in linear scale*) typically represents a plot of the magnitude of the electric or magnetic field as a function of the angular space.

b. *power* pattern(*in linear scale*) typically represents a plot of the square of the magnitude of the electric or magnetic field as a function of the angular space.

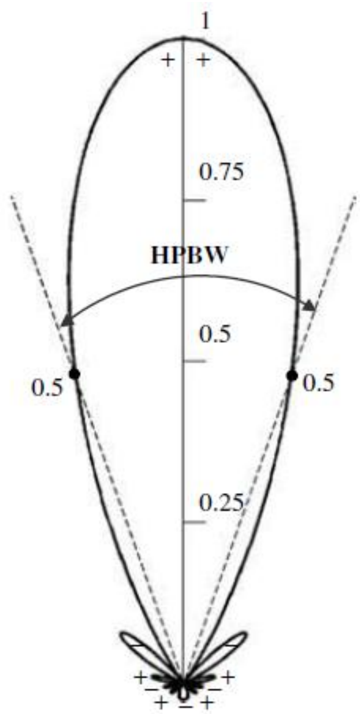
c. *power* pattern(*in dB*) represents the magnitude of the electric or magnetic field, in decibels, as a function of the angular space.

To demonstrate this, the two-dimensional normalized field pattern (*plotted in linear scale*), power pattern(*plotted in linear scale*), and power pattern (*plotted on a logarithmic dB scale*) of a 10-element linear antenna array of isotropic sources, with a spacing of $d = 0.25\lambda$ between the elements, are shown in Figure 2.2. *In this and subsequent patterns, the plus (+) and minus (−) signs in the lobes indicate the relative polarization of the amplitude between the various lobes, which changes (alternates) as the nulls are crossed*. To find the points where the pattern achieves its half-power (−3 dB points), relative to the maximum value of the pattern, you set the value of the

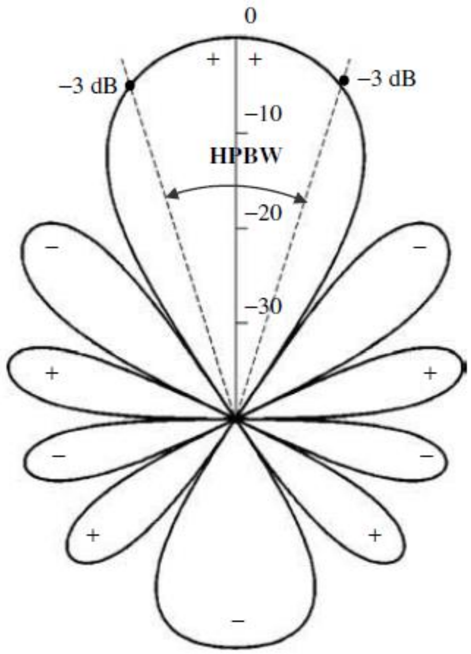
a. field pattern at 0.707 value of its maximum, as shown in Figure (a)
b. power pattern (in a linear scale) at its 0.5 value of its maximum, as shown in Figure (b)
c. power pattern (in dB) at −3 *dB value of its maximum, as shown in Figure (c).*



(a) Field pattern (in linear scale)



(b) Power pattern (in linear scale)



(c) Power pattern (in dB)

GAIN

Another useful measure describing the performance of an antenna is the *gain*. Although the gain of the antenna is closely related to the directivity, it is a measure that takes into account the efficiency of the antenna as well as its directional capabilities. Remember that directivity is a measure that describes only the directional properties of the antenna, and it is therefore controlled only by the pattern. *Gain* of an antenna (in a given direction) is defined as "the ratio of the intensity, in a given direction, to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically. The radiation intensity corresponding to the isotropically radiated power is equal to the power accepted (input) by the antenna divided by 4π." In equation form this can be expressed as

$$Gain = 4\pi \frac{\text{radiation intensity}}{\text{total input (accepted) power}} = 4\pi \frac{U(\theta,\phi)}{P_{in}} \text{(dimensionless)}$$

In most cases we deal with *relative gain*, which is defined as "the ratio of the power gain in a given direction to the power gain of a reference antenna in its referenced direction." The power input must be the same for both antennas. The reference antenna is usually a dipole, horn, or any other antenna whose gain can be calculated or it is known. In most cases, however, the reference antenna is a *lossless isotropic source*. Thus

$$G = \frac{4\pi U(\theta,\phi)}{P_{in}(\text{lossless isotropic source})} \text{(dimensionless)}$$

When the direction is not stated, the power gain is usually taken in the direction of maximum radiation.

Input Impedance

Input impedance is defined as “the impedance presented by an antenna at its terminals or the ratio of the voltage to current at a pair of terminals or the ratio of the appropriate components of the electric to magnetic fields at a point.” In this section we are primarily interested in the input impedance at a pair of terminals which are the input terminals of the antenna. In Figure 2.27(a) these terminals are designated as $a - b$. The ratio of the voltage to current at these terminals, with no load attached, defines the impedance of the antenna as

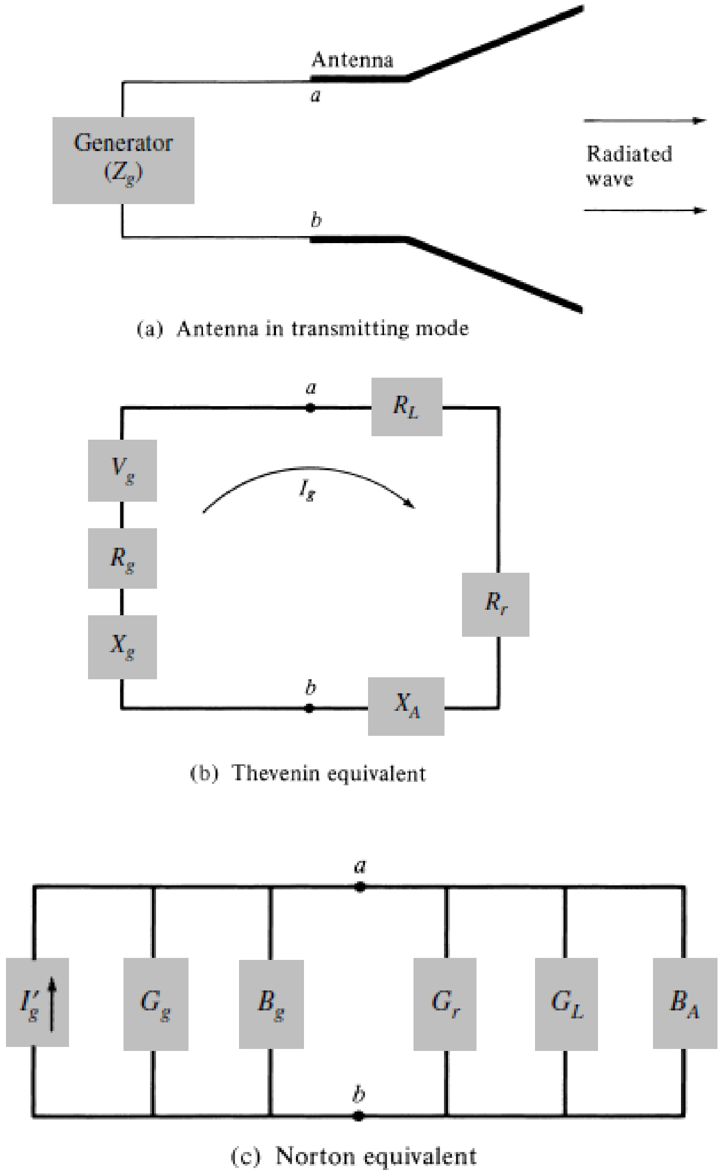
$$Z_A = R_A + jX_A$$

where

ZA = antenna impedance at terminals *a* –*b* (ohms)

RA = antenna resistance at terminals *a* –*b* (ohms)

XA = antenna reactance at terminals *a* –*b* (ohms)



Transmitting antenna and its equivalent circuits.

In general the resistive part consists of two components; that is

$$R_A = R_r + R_L$$

where

Rr = radiation resistance of the antenna

RL = loss resistance of the antenna

If we assume that the antenna is attached to a generator with internal impedance

$$Z_g = R_g + jX_g$$

where

Rg = resistance of generator impedance (ohms)

Xg = reactance of generator impedance (ohms)