CHAPTER 3: ANTENNAS

Antennas couple propagating electromagnetic waves to and from circuits and devices, typically using wires (treated in Section 3.2) or apertures (treated in Section 3.3). In practice complicated solutions of Maxwell's equations for given boundary conditions are usually not required for system design and analysis because the antenna properties have already been specified by the manufacturer, and must only be understood. Section 3.1 characterizes these general transmitting and receiving properties of antennas, which are derived in subsequent sections.

3.1 BASIC ANTENNA PROPERTIES

Most antennas reversibly link radiation fields to currents flowing in wires at frequencies ranging from sub-audio through the far-infrared region. Aperture antennas that link radiation fields to materials can operate in microwave, infrared, visible, ultraviolet, X-ray, gamma ray, and even higher energy regimes. The design of lens and mirror systems for coupling radiation directly to materials is generally called "optics", and the use of these optical techniques for coupling radiation to wires or waveguides is often called "quasi-optics". For the most part, all of these systems can be characterized by the definitions that follow.

A transmitter with available power P_T will radiate P_{TR}, where the antenna *radiation efficiency* $\eta \triangleq P_{TR}/P_T \le 1$, and typically is close to unity for most antennas. Ideally, this power would be radiated exclusively in the intended direction, but in practice some fraction is radiated instead to the side into *sidelobes*, or to the rearward 2π steradians in the form of *backlobes.*

The ability of an antenna to radiate energy in a desired direction is characterized by its *antenna directivity*, $D(f,\theta,\phi)$, which is the ratio of power actually transmitted in a particular direction to that which would be transmitted had the power P_{TR} been radiated isotropically; therefore, directivity is sometimes called "directivity over isotropic". The directivity is defined as:

$$D(f,\theta,\phi) \triangleq \frac{P(f,\theta,\phi)}{P_{TR}/4\pi} \tag{3.1.1}$$

where $P(f,\theta,\phi)$ (watts/steradian) integrated over 4π steradians equals P_{TR}. Therefore

$$\int_{4\pi} D \, d\Omega = 4\pi \tag{3.1.2}$$

A more common characterization is *antenna gain* $G(f,\theta,\phi)$, which is the ratio of the power actually transmitted in the direction θ,ϕ to that which would be radiated if the entire available transmitter power P_T were radiated isotropically. That is:

$$G(f,\theta,\phi) \triangleq \eta_R D(f,\theta,\phi) = \frac{P(f,\theta,\phi)}{P_T/4\pi} \tag{3.1.3}$$

Sometimes the term "gain over isotropic" is used to facilitate memorization of the definition.

In cases where the absolute values of the gain or directivity are no of interest the *antenna pattern* $T(f,\theta,\phi)$ is often used, where it is the same as the gain or directivity normalized so its maximum value is unity.

These simple definitions already permit us to calculate interesting quantities. For example, consider the MIT Haystack radar antenna which has an approximate gain of 73 dB at a wavelength of one centimeter. The flux density Wm^{-2} incident on various targets when one-megawatt radar pulses are transmitted can be calculated using:

$$P\left(Wm^{-2}\right) = G(\theta,\phi) \bullet \frac{P_T}{4\pi R^2} \tag{3.1.4}$$

Radar pulses directed toward the moon, Jupiter, and the star Antares at distances $R = 3\times10^8$, 10^{12}, and 3×10^{16} meters yield fluxes of 10^{-5}, 10^{-12}, and 10^{-21} watts/meter^2, respectively. Note that if we have a receiver near Antares with a noise temperature T_R = 10K and our bandwidth is 1 Hz, then its noise equivalent power is $kTB \cong 10^{-22}$ watts, which is one tenth the power of the incoming Haystack radar signal incident upon a one-square-meter patch of receiving aperture. As we shall see, receiving apertures capturing far greater areas are readily built, and so interstellar communications is quite practical provided we are content with a propagation delay of a year or more and the costs of installing the remote station.

The receiving properties of antennas are characterized by the *antenna affective area* $A(f,\theta,\phi)$, where the available power at the output of the receiving antenna $P_r(f)$ is the product of the effective area of the receiving antenna in direction θ,ϕ and the flux density $S(Wm^{-2}Hz^{-1})$ incident from that direction; i.e.:

$$P_r(f) = A(f,\theta,\phi)S(f,\theta,\phi) = \int_{4\pi} A(f,\theta,\phi) \bullet \left[I(f,\theta,\phi) \bullet d\Omega\right] \left(WHz^{-1}\right) \tag{3.1.5}$$

For our Antares example, the received power would be 10^{-17} W on Antares if the antenna there had an effective area $A(f,\theta,\phi)$ of $10^4 m^2$ in the direction of Earth. If $kTB = 10^{-22}$ watts then the

1-Hz SNR $= 10^5$ and the 10^4-Hz SNR = 10, which, as we shall see, is adequate for communicating good voice and acceptable music signals.

As we shall prove for the case where reciprocity applies, the receiving and transmitting properties of antennas are simply related: they both have the same antenna pattern. This can be derived from the principle of detailed balance, which applies for reciprocal antenna structures. Consider an arbitrary antenna radiating power $P(f,\theta,\phi)$ within the small solid angle $d\Omega$, as suggested in Figure 3.1-1.

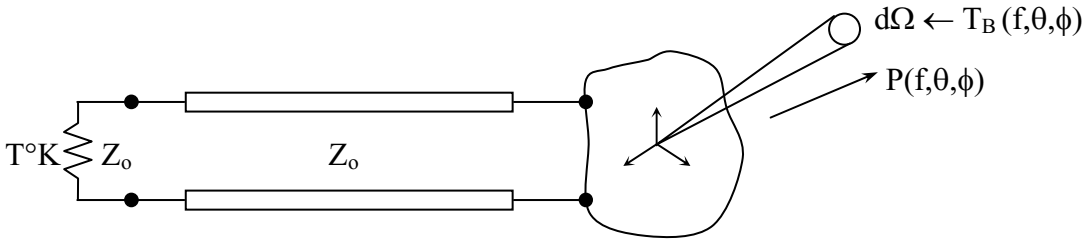

Figure 3.1-1 Antenna system exhibiting the principle of detailed balance

The *principle of detailed balance* says that the power received within each $d\Omega$ matches the power radiated by an antenna into the same solid angle $d\Omega$ when that antenna is driven by a matched load in thermal equilibrium with a radiation field at temperature T°K. This power radiated into $d\Omega$ is:

$$P_{rad}(f,\theta,\phi)\,d\Omega\,df = (kTdf/4\pi)\;Gd\Omega \tag{3.1.6}$$

Using the definition of antenna effective area, the power the antenna receives within the solid angle $d\Omega$ is:

$$P_r = \frac{1}{2}\frac{2kT_B(\theta,\phi)}{\lambda^2}df\;A(f,\theta,\phi)\;d\Omega \tag{3.1.7}$$

where the factor of 1/2 results because only one polarization can received by an antenna connected to a single-mode TEM transmission line, and the total blackbody radiation associated with the brightness temperature $T_B = T$ in the Rayleigh-Jeans limit is given by (2.1.28). If we equate (3.1.6) and (3.1.7) in accord with the principle of detailed balance, we obtain the very important and general equation:

$$G(f,\theta,\phi) = \frac{4\pi}{\lambda^2}A(f,\theta,\phi) \tag{3.1.8}$$

This expression assumes that $hf << kT$ and that signals arriving from different directions are uncorrelated so that their powers superimpose. Detailed balance also assumes system reciprocity or the absence of magnetized plasmas or ferrites.

These expressions for the receiving and transmitting properties of antennas can be combined to yield the performance of radio links such as the one illustrated in Figure 3.1-2

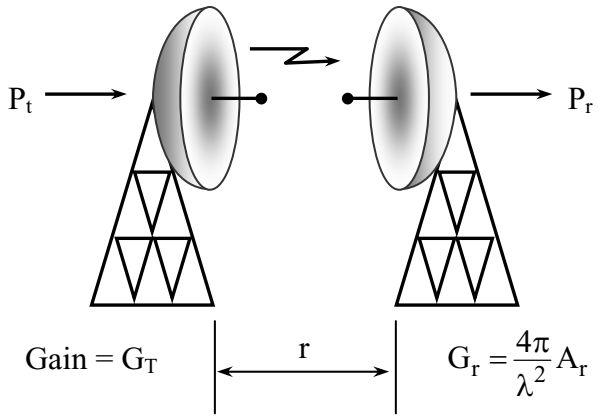

Figure 3.1-2 Radio link

Our received power P_r is simply the product of the effective area A_r of the receiving antenna and the transmitted flux [Wm^{-2}] incident upon it; i.e.:

$$P_r = \frac{P_t}{4\pi r^2} \bullet G_t \bullet A_r \quad \text{Watts} \tag{3.1.9}$$

The system designer maximizing P_r might be tempted to let the separation r between the two antennas approach zero, which would suggest received powers approaching infinity. The solution to this paradox is that the distance r must be greater than some minimum for (3.1.9) to be valid. For example, let $P_r = P_t$ at this minimum distance r_{min}, and let the effective areas of the two antennas be equal to $D^2\left(m^2\right)$. Then follows from (3.1.9) that:

$$\frac{G_t A_r}{4\pi r_{min}^2} = 1 = \frac{A_t A_r}{\lambda^2 r_{min}^2} = \frac{D^4}{\lambda^2 r_{min}^2} \tag{3.1.10}$$

where we have also used (3.1.8). This suggests $r_{min} = D^2/\lambda$, although in practice we usually require that $r > 2D^2/\lambda$ in order to use (3.1.9) with confidence, as explained later.

Although power is usually the parameter of interest in communications or radar systems, antenna temperature is often preferred when passive sensors for radio astronomy or remote sensing are of interest. Equating the power received by an antenna (3.1.5) to the Rayleigh-Jeans power in a transmission line (2.1.16), and integrating over all angles, we obtain:

$$kT_A\left(W\ Hz^{-1}\right) = \int_{4\pi} A(\theta,\phi)I(\theta,\phi)\ d\Omega \tag{3.1.11}$$

where the intensity $I(\theta,\phi)$ and effective area $A(\theta,\phi)$ both apply to the same specific polarization, and the units of intensity are $Wm^{-2}Hz^{-1}ster^{-1}$. If the antenna can intercept both polarizations, then each polarization would yield its own separate T_A. Substituting the Rayleigh-Jeans expression for intensity $I = 2kT_B/\lambda^2$, divided by two to reflect the fact that only one polarization is involved, yields the desired expressions for antenna temperature:

$$T_A = \frac{1}{\lambda^2}\int_{4\pi} A(\theta,\phi)T_B(\theta,\phi)d\Omega \tag{3.1.12}$$

$$T_A = \frac{1}{4\pi}\int_{4\pi} G(\theta,\phi)T_B(\theta,\phi)d\Omega \tag{3.1.13}$$

These expressions assume that the power arriving from different directions can be superimposed, which requires that signals arriving from different directions be statistically independent. Typical exceptions to this might include coherent transmitted signals that arrive at the receiving antenna by different paths and interfere constructively or destructively, depending on frequency-dependent phase differences.

Often the target of interest subtends a solid angle small compared to the beamwidth of the antenna. Then the integral over angle in (3.1.13) responds principally to the average brightness temperature $\overline{T}_B$ of the source:

$$T_{A_S} = \frac{1}{4\pi}\int_{\Omega_S} G(\theta,\phi)T_B(\theta,\phi)d\Omega = \Omega_s\frac{G_o\overline{T}_B}{4\pi} = \Omega_s\frac{A_o\overline{T}_B}{\lambda^2} \tag{3.1.14}$$

where Ω_s is the source solid angle as viewed from the antenna; G_o and A_o are the on-axis antenna gain and effective area, respectively. This configuration is illustrated in Figure 3.1-3.

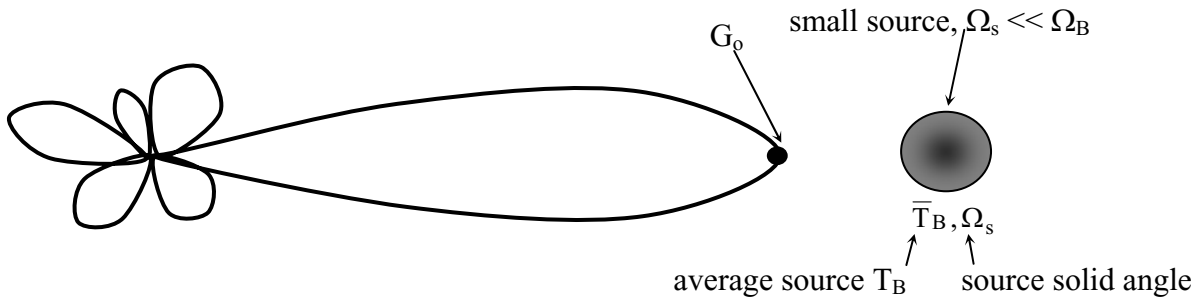

Figure 3.1-3 Antenna viewing a small thermal source

Equation 3.1.14 for source antenna temperature can usefully be interpreted as the average brightness temperature of the source $\overline{T}_B$ times a coupling coefficient less than unity. A simple expression for this coupling coefficient results if we define the useful concept of *beam solid angle* Ω_A so that :

$$G_o \Omega_A \triangleq \int_{4\pi} G \, d\Omega = 4\pi\eta_r \tag{3.1.15}$$

where this relation between $G(\theta,\phi)$ and Ω_A is suggested by Figure 3.1-4(a). The beam solid angle Ω_A (steradians) usually approximates the solid angle contained within the antenna beam's half-power contour. The solid angle Ω_s of an unresolved source is related to Ω_A as suggested in Figure 3.1-4(b), so that (3.1.15) becomes:

$$T_A \cong \left(\frac{\Omega_s}{\Omega_A} \eta_R \right) \overline{T}_B \tag{3.1.16}$$

where the factor of $\eta_R \frac{\Omega_s}{\Omega_A}$ in (3.1.16) is the coupling coefficient between source brightness $\overline{T}_B$ and T_A .

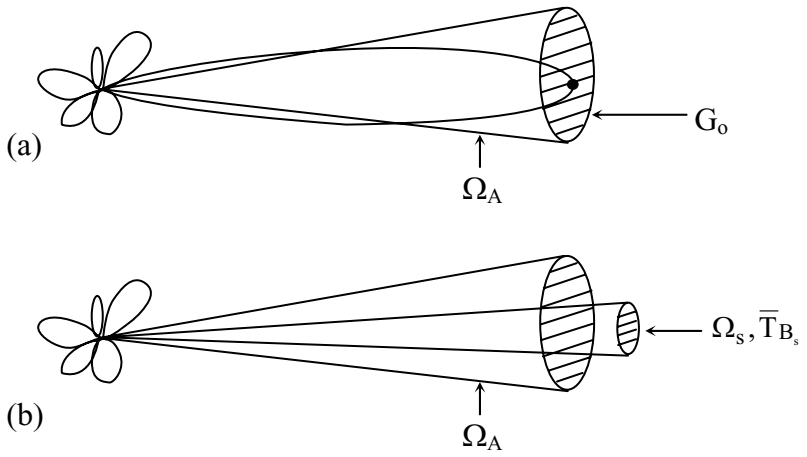

Figure 3.1-4 Antenna beam solid angle in relation to (a) the antenna pattern and (b) source solid angle

There are three standard ways to characterize small thermal sources. The most complete description is $T_B(\theta,\phi,f)$ for each of two polarizations, which may be reduced to two average brightness temperatures $\overline{T}_B$. The third characterization is the *source flux density* $S(f)\left[Wm^{-2}Hz^{-1}\right]$ for each polarization, which is the integral of the source intensity over the source solid angle:

$$S(f)\left[Wm^{-2}Hz^{-1}\right] = \int_{\Omega_s} I(f,\theta,\phi)d\Omega \tag{3.1.17}$$

The utility of the source flux density is that it is independent of the antenna used to observe it. The international scientific unit for flux density is the *Jansky*, which is defined as 10^{-26} $Wm^{-2}Hz^{-1}$.

3.2 WIRE ANTENNAS

To accurately compute the electromagnetic properties of wire antennas it is necessary to solve Maxwell's equations subject to boundary conditions imposed by the wires and their surrounding environment. Since this problem is often complicated, it is customary and usually adequate to assume the wire antenna is located in free space and to ignore to first order any radiative interactions between one part of the wire array and another. This approximation works well for most wire antennas less than a wavelength across. Since Maxwell's equations are linear, the radiation from such wire antennas can be approximated as the linear superposition of the radiation produced by each infinitesimal sub-element of such an antenna. Such sub-elements act

as *short dipole antennas*, the properties of which can be derived directly from Maxwell's equations, as follows.

First it is useful to review Maxwell's equations, which govern the electric field $\overline{E}(\text{volts/meter})$ and the magnetic field $\overline{H}(\text{amperes}/\text{meter})$. Their related variables are the electric displacement $\overline{D} = \varepsilon\overline{E}$ (Coulombs/meter^2) and the magnetic flux density $\overline{B} = \mu\overline{H}$ (Teslas), where ε is the permittivity of the medium and in vacuum is $\varepsilon_o = 8.8542\times10^{-12}$ farads/meter, and μ is the permeability and in vacuum is $\mu_o = 4\pi\times10^{-7}$ henries/meter. One Tesla is defined as one $\text{Weber}/\text{meter}^2 = 10^4$ gauss. Maxwell's equations and their static limits are:

Maxwell's equations		Statics	Complex Maxwell's equations	
$\nabla\times\overline{E} = -\dfrac{\partial\overline{B}}{\partial t}$	$\rightarrow$	$=0$	$\nabla\times\underline{E} = -j\omega\underline{B}$	(3.2.1)
$\nabla\times\overline{H} = \overline{J} + \dfrac{\partial\overline{D}}{\partial t}$	$\rightarrow$	$= \overline{J}\left(a\ m^{-2}\right)$	$\nabla\times\underline{H} = \underline{J} + j\omega\underline{D}$	(3.2.2)
$\nabla\bullet\overline{D} = \rho$	$\rightarrow$	$= \rho\left(C\ m^{-3}\right)$	$\nabla\bullet\underline{D} = \underline{\rho}$	(3.2.3)
$\nabla\bullet\overline{B} = 0$	$\rightarrow$	$= 0$	$\nabla\bullet\underline{B} = 0$	(3.2.4)

where $\overline{J}$ is the current density (amperes/meter^2) and the ρ is the charge density (Coulombs/meter^3). The *dell operator* is defined as:

$$\nabla \triangleq \hat{x}\partial/\partial x + \hat{y}\partial/\partial y + \hat{z}\partial/\partial z \tag{3.2.5}$$

where $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit vectors in Cartisian coordinates.

Consider the problem suggested by Figure 3.2-1, where a source within volume V_q produces fields observed by a person at point p, which is separated from any particular point q in the source by the distance r_{pq}.

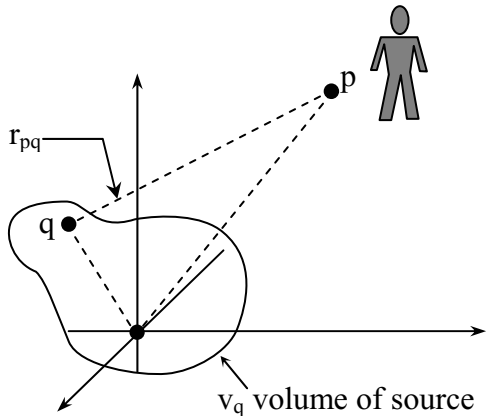

Figure 3.2-1 Source and observer of electromagnetic radiation

Since the radiation expressions closely resemble the static solutions to Maxwell's equations, it is useful to begin with the static solutions themselves.

In the static limit, since $\nabla \times \overline{E} = 0$:

$$\overline{E} = -\nabla\phi \tag{3.2.6}$$

where ϕ is the electrostatic potential, which is a position-dependent scalar variable. Similarly, since $\nabla \bullet \overline{B} = 0$,

$$\overline{B} = \nabla \times \overline{A} \tag{3.2.7}$$

where $\overline{A}$ is the vector potential.

The electrostatic potential ϕ_p at point p may be found by combining (3.2.3) and (3.2.6) in several steps to yield:

$$\phi_p = \frac{1}{4\pi\varepsilon} \int_{V_q} \frac{\rho_q}{r_{pq}} dv_q \text{ [volts]} \tag{3.2.8}$$

Similarly, (3.2.2) and (3.2.7) can be combined in the static limit to show:

$$\overline{A}_p = \frac{\mu}{4\pi} \int_{V_q} \frac{\overline{J}_q}{r_{rq}} dv_q \text{ [Tesla m]} \tag{3.2.9}$$

Thus in the static limit it is simple to solve for the electromagnetic fields $\overline{E}$ and $\overline{H}$ by first using (3.2.8) and (3.2.9) to relate the source charge density distribution ρ_q and current density distribution $\bar{J}_q$ to their corresponding potentials, and then to differentiate these potentials by using (3.2.6) and (3.2.7).

The dynamic solutions to Maxwell's equations can be derived in a similar manner by noting the divergence of $\overline{B}$ is zero so that, as before:

$$\overline{B} = \nabla \times \overline{A} \tag{3.2.10}$$

It can then be shown that the dynamic solutions to Maxwell's equations for the vector potential at point p are:

$$\overline{A}_p(t) = \frac{\mu}{4\pi} \int_{V_q} \frac{\bar{J}_q\left(t - r_{pq}/c\right)}{r_{pq}}\, dv_q \tag{3.2.11}$$

where we integrate over the entire volume of the source region V_q. It is interesting to note that the dynamic solution for vector potential (3.2.11) is the same as the static solution (3.2.9) except that the result depends not on the instantaneous current density, but rather on the current density as it existed r_{pq}/c in the past, where c is the velocity of light. That is, the dynamic solution is the same as the static solution except that we must wait while the information about the current density at any point q propagates to the observer at point p. The velocity of light in vacuum is $c = 1/\sqrt{\mu_o \varepsilon_o} \cong 3\times 10^8 \text{ ms}^{-1}$. In the sinusoidal steady state (3.2.11) reduces to:

$$\underline{\overline{A}}_p = \frac{\mu}{4\pi} \int_{V_q} \frac{\underline{\bar{J}}_q e^{-jkr_{pq}}}{r_{pq}} dv_q \tag{3.2.12}$$

where the propagation constant k is:

$$k = \omega\sqrt{\mu_o \varepsilon_o} = \omega / c = 2\pi / \lambda \tag{3.2.13}$$

Determining the electromagnetic fields radiated by dynamic charges and currents is then straight-forward. We may first use (3.2.12) to calculate the vector potential $\underline{\overline{A}}_p$ from the given source current distribution $\underline{\bar{J}}_q$. The magnetic field $\underline{\overline{B}}(\bar{r})$ can be found from the vector potential distribution $\underline{\overline{A}}(\bar{r})$ using (3.2.10). The electric field distribution $\underline{\overline{E}}(\bar{r})$ is:

$$\underline{\overline{E}} = \left(\nabla \times \underline{\overline{H}}\right) \Big/ j\omega t \tag{3.2.14}$$

which follows from (3.2.2).

We may now use this simple approach to find the electromagnetic fields radiated by an *elementary dipole antenna,* often called a *Herzian dipole*. The typical Herzian dipole appears in Figure 3.2-2 where a sinusoidal current $\underline{I}_o$ in a short wire segment of length d radiates fields detected by an observer at point r >> d.

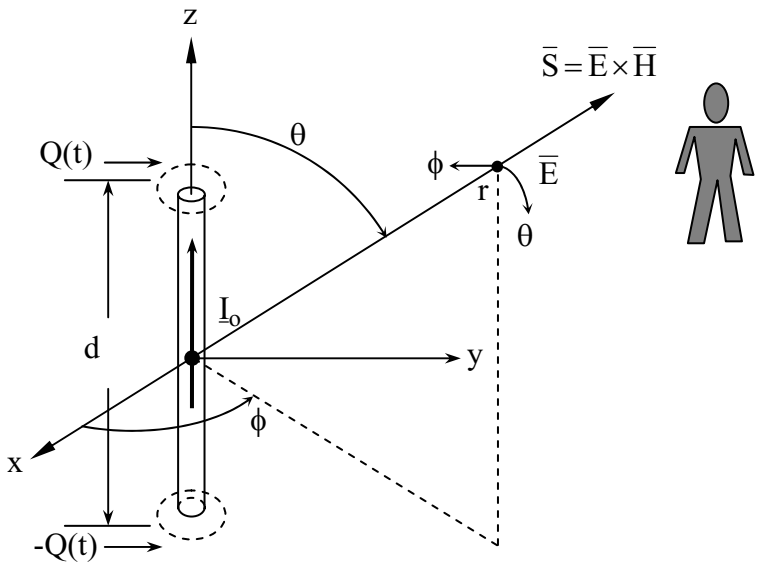

Figure 3.2-2 Herzian dipole antenna

Conservation of charge uniquely relates the current $\underline{I}_o$ to the oscillating charges $\underline{Q}$ at the ends of the short dipole, where we assume that these two charges have a zero sum. Simple expressions are obtained in the spherical coordinate system where the dipole is aligned with the z-axis at the origin and the angles θ and ϕ are defined to be zero along the z- and x-axes, respectively. The electromagnetic fields produced close to the dipole in the *near field*, where:

$$r << \frac{\lambda}{2\pi} \tag{3.2.15}$$

are called *quasistatic fields* because they approximate the electromagnetic fields produced by static current and charge distributions. They vary slowly in time and are proportional to $\underline{I}_o$ and $\underline{Q}$, respectively, which are necessarily 90° out of phase. Superimposed on these quasistatic electromagnetic fields are the radiated fields, which dominate only in the *far field* where:

$$r >> \frac{\lambda}{2\pi} \tag{3.2.16}$$

In the near field of a Herzian dipole the time-average electric energy storage dominates the average magnetic energy stored. In the far-field region where $r >> \frac{\lambda}{2\pi}$, the radiated fields dominate the quasistatic fields. This far field is distinct from the "far-field" region associated with antenna arrays or aperture antennas where r is sufficiently great that the antenna pattern is no longer a function of distance from the antenna; there $r > 2D^2/\lambda$, where D is the maximum dimension of the antenna or antenna array, as discussed further later.

In the far field of a Herzian dipole the radiated electric field is approximately:

$$\underline{\overline{E}} \cong \hat{\theta} j\eta_o \frac{k\underline{I}d \sin\theta}{4\pi r} e^{-jkr} \tag{3.2.17}$$

where $\eta_o = \sqrt{\mu_o/\varepsilon_o} = 377\Omega$, and is called the *characteristic impedance of free space.* Using (3.2.1) we can readily find:

$$\underline{\overline{H}} \cong \hat{\phi} j \frac{k\underline{I}_o d \sin\theta}{4\pi r} e^{-jkr} \tag{3.2.18}$$

The radiated power density (Wm^{-2}) is generally characterized by the *Poynting vector,* which is defined as:

$$\overline{S}(t) \triangleq \overline{E}(t) \times \overline{H}(t) \tag{3.2.19}$$

$$\underline{\overline{S}} \triangleq \underline{\overline{E}} \times \underline{\overline{H}}^* \tag{3.2.20}$$

where the time-average power density is:

$$\left\langle \overline{S}(t) \right\rangle = \frac{1}{2} \mathrm{Re}\left\{ \underline{\overline{S}} \right\} \left(Wm^{-2} \right) \tag{3.2.21}$$

For a short dipole antenna:

$$\left\langle \overline{S}(t,\theta) \right\rangle = \hat{r} \frac{\eta_o}{2} \left| \frac{\underline{I}_o d \sin\theta}{2\lambda r} \right|^2 \tag{3.2.22}$$

which corresponds to a doughnut-shaped radiation pattern with a null along the z-axis, as suggested in Figure 3.2-3.

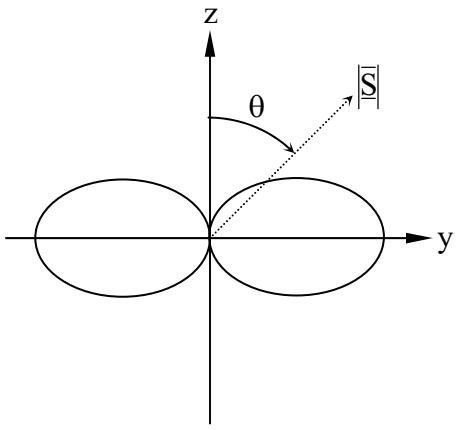

Figure 3.2-3 Antenna pattern of a z-oriented Hertzian dipole

By integrating over all angles θ,ϕ we can find the total power transmitted:

$$P_t = \frac{1}{2} R_e \left\{ \int_{4\pi} \underline{\bar{S}} \bullet \hat{r}\, d\Omega \right\} = \frac{\pi}{3} \eta \left| \frac{\underline{I_o} d}{\lambda} \right|^2 \text{ watts} \tag{3.2.23}$$

Combining the definition of antenna gain $G = \langle \bar{S}(t) \rangle \Big/ \left(P_t / 4\pi r^2 \right)$ with (3.2.22) and (3.2.23) yields the gain of a short-dipole antenna:

$$G = \frac{3}{2} \sin^2 \theta \tag{3.2.24}$$

Any antenna coupled to a linear reciprocal electromagnetic environment exhibits a complex impedence $\underline{Z}(f)$ as seen from the antenna terminals where:

$$\underline{Z}(f) = R_r(f) + jX(f) \tag{3.2.25}$$

where R_r is the *radiation resistance* of the antenna and jX is its reactance, as illustrated in Figure 3.2-4.

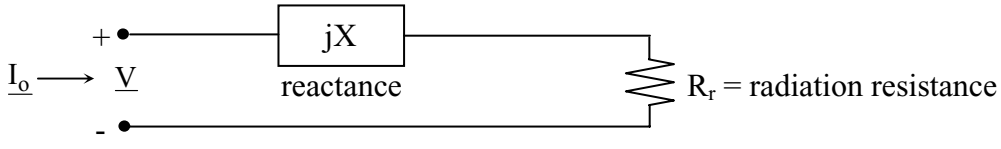

Figure 3.2-4 Equivalent circuit for an antenna

The radiation resistance is simply related to the drive current $\underline{I}_o$ and the resulting total transmitted power P_t, which can be found from (3.2.23):

$$P_t = \frac{1}{2}|\underline{I}_o|^2 R_r \tag{3.2.26}$$

Thus any antenna as seen from the circuit to which it is connected, has a resistive component $R_d + R_r$, where R_d consists of any dissipative losses in the antenna structure itself, and the radiation resistance R_r corresponds to power which is radiated away and dissipated remotely. The reactance of the antenna jX is positive imaginary if the near fields of the structure are predominantly inductive, i.e. if magnetic stored energy exceeds electric stored energy, or negative imaginary if the structure is predominantly capacitive.

The equivalent circuit of a short-dipole antenna can be readily computed. Consider the short-dipole illustrated in Figure 3.2-5, where the antenna is connected to the external circuit by wires joined at its center.

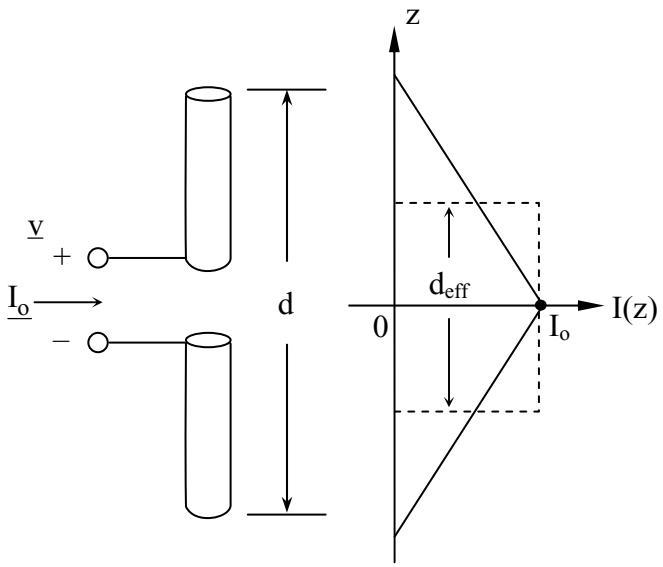

Figure 3.2-5 Short-dipole antenna and its current distribution

The wires connect to the circuit by a small TEM line that does not radiate or significantly alter the electromagnetic fields produced by the short-dipole antenna itself, because those fields are perpendicular to the TEM line. The current distribution on a physical short-dipole antenna must necessarily go to zero at the ends of the wire. The reasons for the approximately triangular current distribution I(z) will be explained shortly. The expression for a Herzian dipole (3.2.17)

assumed a uniform current $\underline{I}_o$ over the length d of the the dipole and therefore must be modified when the integral (3.2.12) for $\underline{\overline{A}}_p$ is computed; that is:

$$\underline{I}_o d \rightarrow \underline{I}_o d_{eff} = \int_{-\infty}^{\infty} \underline{I}(z)dz \tag{3.2.27}$$

where the *dipole effective length* $d_{eff} = d/2$ because of the triangular current distribution which results when $d << \lambda/2\pi$.

When we substitute into (3.2.26) the expression for the total power transmitted P_t by a short-dipole antenna, we obtain the radiation resistance for a short-dipole antenna:

$$R_r = \frac{2\pi}{3}\eta_o \left(d_{eff}/\lambda\right)^2 \text{ ohms} \tag{3.2.28}$$

Simple examples illustrate some of the problems associated with the practical use of short-dipole antennas. Consider an AM radio station broadcasting at 1 MHz (λ = 300 m) with an antenna of effective length $d_{eff} = 1\,\text{meter}$; then (3.2.28) yields $R_r \cong 0.01$ ohms, which is very badly mismatched to typical amplifier impedances of 50-100 ohms. Equation 3.2.28 suggests the effective length should be somewhat greater than one-quarter wavelength in order to achieve high radiation efficiencies, but at these lengths (3.2.28) is beginning to lose validity because we assumed $d << \lambda/2\pi$. To mitigate this impedance matching problem most sensitive long-wavelength radio receivers employ resonant structures or transformers to match the antenna to the input amplifier.

Wire antennas of arbitrary shape can be analyzed if we assume their current distributions are known. Often this is a good approximation because the current distribution is approximately that of a TEM transmission line. When the current distribution $\underline{I}\left(\overline{r}\right)$ is known we can use (3.2.17) to find the far-field $\underline{\overline{E}}(\theta,\phi)$.

Consider the arbitrary antenna in Figure 3.2-6, which radiates towards a person at point p at an angle $\theta(\ell)$ which is a function of position ℓ along the wire antenna. The antenna current $I(\ell)$ and distance $r(\ell)$ to the observer are also functions of position along the wire.

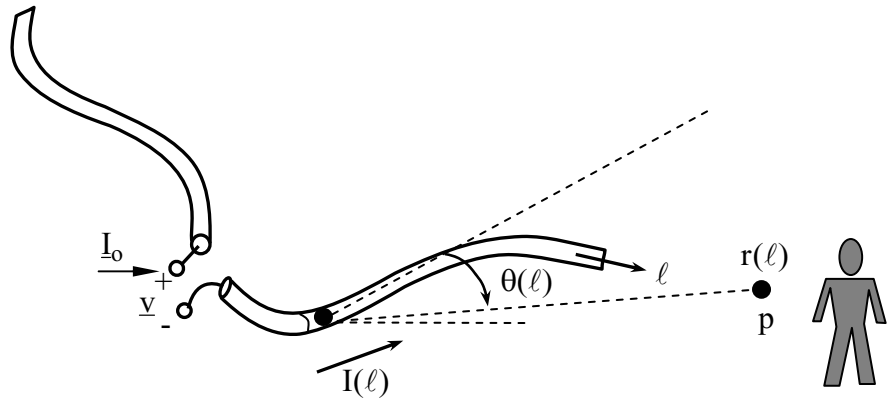

Figure 3.2-6 General long-wire antenna

The far field produced by a general long-wire antenna of length L is:

$$\overline{\underline{E}}_{ff} \cong \frac{jk\eta}{4\pi r}\int_L \hat{\theta}(\ell)\ \underline{I}(\ell)e^{-jkr(\ell)}\sin\ \theta(\ell)d\ell \tag{3.2.29}$$

from which the total transmitted power and radiation resistance can be computed.

The circumstances under which the current distribution on wire antennas can be reasonably estimated using the TEM approximation is suggested by the observation that the electric and magnetic fields in the immediate vicinity of a wire antenna contain most of the energy associated with the current $\underline{I}$, and therefore control the velocity of propagation and the ratio of current to voltage.

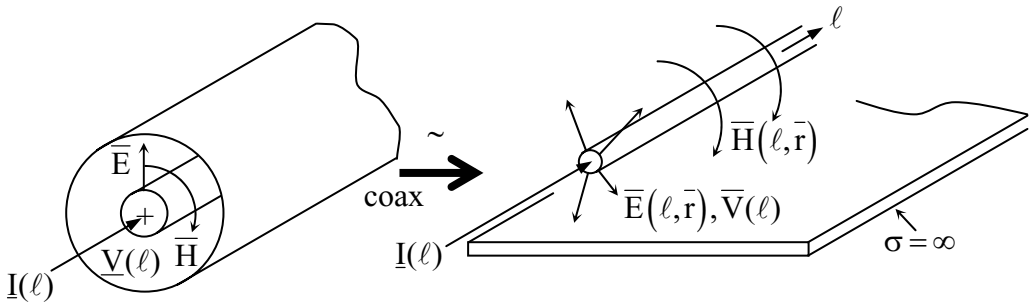

Figure 3.2-7 Similarity between the TEM fields in a coaxial cable and those surrounding free-standing wire antennas

Consider the similarity between a coaxial cable and an isolated wire over a ground plane, as illustrated in Figure 3.2-7. In a coaxial geometry both the electric and magnetic fields decay as 1/r and the stored electric and magnetic energies decay as $1/r^2$; therefore most of the energy near a wire antenna is stored within a few wire radii. Farther from the wire the fields weaken and

bend to satisfy boundary conditions imposed by the surrounding environment. These more remotely stored energies are too small to control the current distribution within the antenna itself, which closely approximates that of a TEM transmission line.

Perhaps the most popular wire antenna is the *half-wave dipole,* which is a center-fed straight wire antenna of length λ/2, as illustrated in Figure 3.2-8(a). Since the currents at the ends must go to zero it follows that they have a cosine form.

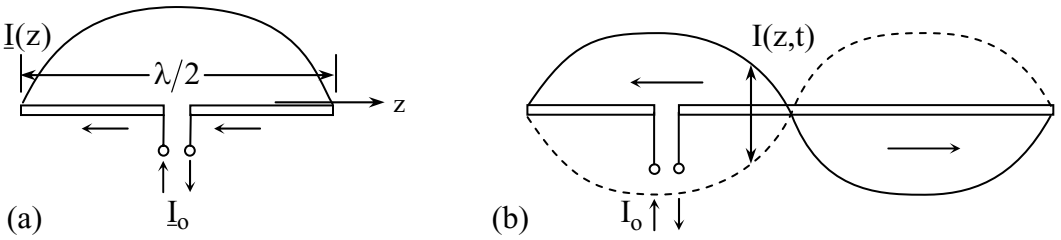

Figure 3.2-8 Current distributions on (a) half-wave and (b) full-wave antennas

Provided the antenna is isolated in space and the wire diameter is much smaller than the antenna length L, then $R_r \cong 73\Omega$ and the reactance X is approximately zero for $L \cong \lambda/2$. The reactance X of a full-wave antenna for which L ≅ λ, as illustrated in Figure 3.2-8(b), is also approximately zero, while the radiation resistance is somewhat greater than 73Ω because it radiates slightly more total power into two conical lobes.

The general expression for a center-fed wire of length L is:

$$\underline{\bar{E}}_{ff} \cong \hat{\theta}\frac{j\eta I_o e^{-jkr}}{2\pi r \sin\theta}\left[\cos\left(\frac{kL}{\alpha}\cos\ \theta\right) - \cos\ \frac{kL}{\alpha}\right] \tag{3.2.30}$$

which, for L = λ/2, reduces to:

$$\underline{\bar{E}}_{ff} \cong \hat{\theta}\frac{j\eta I_o e^{-jkr}}{2\pi r \sin\theta}\cos\left(\frac{\pi}{2}\cos\ \theta\right) \tag{3.2.31}$$

The antenna pattern for the half-wave dipole is suggested in Figure 3.2-9. The radiation from a half-wave dipole is slightly more directive than that of a short dipole because signals radiated closer to the z-axis tend to cancel, particularly when contributions from the two ends one-half-wavelength apart are considered.

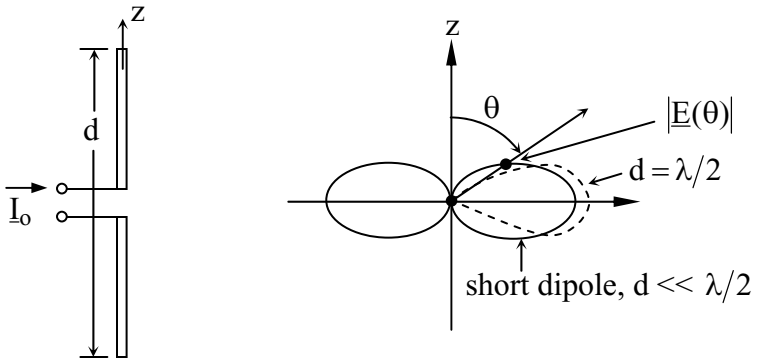

Figure 3.2-9 Antenna pattern for a half-wave dipole

The TEM model for current distribution on wire antennas permits approximate understanding of other standard configurations such as the off-center-fed wire antenna of Figure 3.2-10.

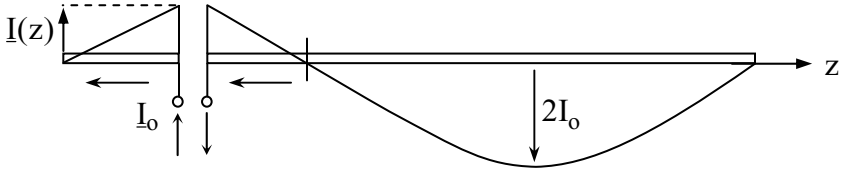

Figure 3.2-10 Current distribution on an off-center-fed wire antenna ~0.7 λ long

The antenna in Figure 3.2-10 is fed a small fraction of a wavelength from one end so that the input current $\underline{I}_o$ is much smaller than the peak current distribution $2\underline{I}_o$ one-quarter wavelength from the open-circuit end of the long wire. This current distribution follows directly from the TEM-current-distribution assumption and the location of the current feed relative to the two ends of the antenna. By placing the feed point even closer to the left-hand end of the antenna, the ratio between the peak current and the drive current can be made even greater than the illustrated ratio of two. Thus this off-center-fed antenna exhibits transformer-like qualities. Because most of the power in this antenna will be radiated as if from a half-wavelength dipole carrying $2\underline{I}_o$, its radiation resistance is greater than that of a half-wave dipole, although its reactance X is capacitive as a result of the excessive electric energy stored to the left of the current null near the feed point.

The TEM model for current distribution breaks down when different parts of a wire antenna system interact strongly, or when radiation itself begins to alter the current distribution. For example, the long-wire antenna illustrated in Figure 3.2-11 radiates sufficiently well that the TEM wave traveling from the feed point to the right toward the long end of the antenna decays approximately exponentially. The result is a standing wave in the one-half wavelength segment

at the left end, a traveling wave portion to the right, and a standing wave portion near the right-hand end where the decaying wave is reflected.

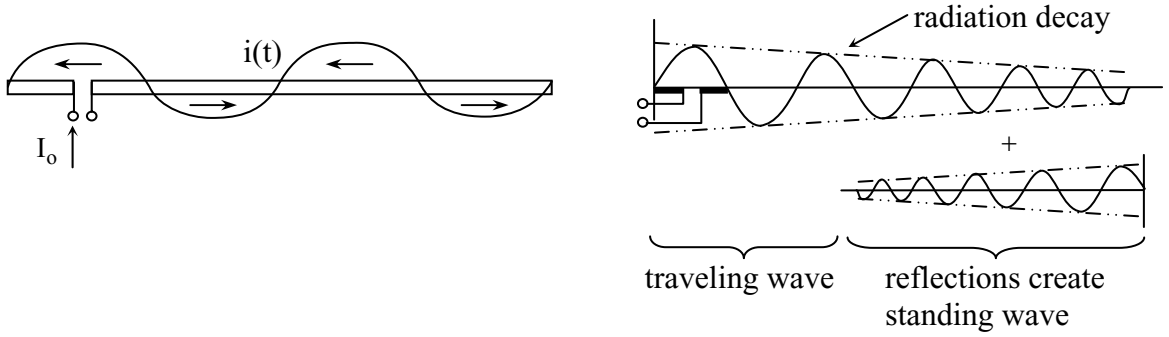

Figure 3.2-11 Currents on a long-wire antenna

An *antenna array* consists of N antenna elements for which the element antenna patterns in isolation are identical and similarly oriented, as suggested in Figure 3.2-12. Sometimes antenna arrays are comprised of sub-arrays which may be oriented or radiate differently. Assume the phase lag between the i^{th} antenna element and the observer at point p is $2\pi r_i/\lambda$, and that the electric field contribution at p from element i in the θ,ϕ direction is $\underline{A}_i \underline{E}_o(\theta,\phi)e^{j\varphi_o(\theta,\phi)-j\varphi_i(\theta,\phi)}$; that is:

$$\overline{\underline{E}}_p \cong \underline{E}_o(\theta,\phi)e^{-j\varphi_o(\theta,\phi)}\sum_{i=1}^{N}\underline{A}_i e^{-j\varphi_i(\theta,\phi)} \tag{3.2.32}$$

where the current feeding the ith antenna is $\underline{A}_i \underline{I}_o$. The gain of the antenna array is proportional to the square of (3.2.32) or:

$$G(\theta,\phi) \propto \left|\overline{\underline{E}}\right|^2 = \left|\overline{\underline{E}}_o(\theta,\phi)\right|^2 \bullet \left|\sum_i \underline{A}_i e^{-j\varphi_i(\theta,\phi)}\right|^2 \tag{3.2.33}$$

This *antenna array equation* (3.2.33) shows how (3.2.32) can be factored into an *element factor* $\left|\overline{\underline{E}}_o(\theta,\phi)\right|^2$ that characterizes the presumably identical directional properties of the elements, and an *array factor* that characterizes the directional effects of the relative positions and current amplitudes $\underline{A}_i$ associated with each of the separate elements i. This useful expression simplifies understanding the radiation properties of groups of antenna elements, and also of single elements for which the current distribution has repetitive qualities.

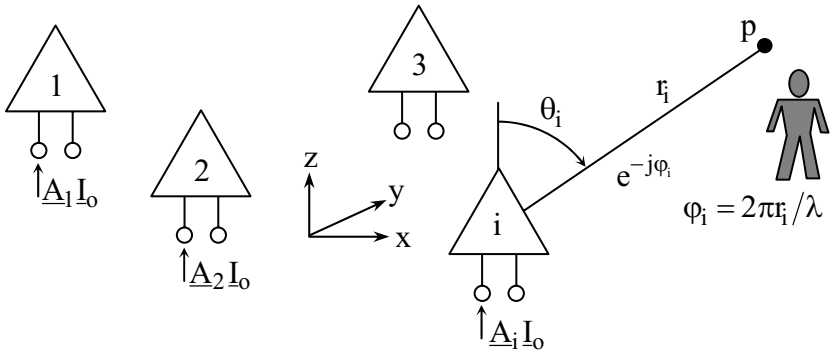

Figure 3.2-12 General antenna array

Although the array factor is often designed to produce only one main lobe, sometimes the antenna elements are spaced more than $\lambda/2$ apart so as to produce multiple grating lobes. For example, consider the case of two identically oriented antenna elements spaced λ apart and driven in phase. In this case the array factor will be maximum along the axis joining the two elements and also in the plane perpendicular to that axis; in all of these directions the radiated contributions from the two elements are in phase. However, these two equal radiated contributions exactly cancel along two conical surfaces at angles of $\pm 60°$ with respect to the main axis.

Wire antennas often operate in the vicinity of reflectors such as the ground, ocean, or planar metallic surface. Consider a radiating current $\underline{\bar{J}}$ terminated in two charge reservoirs, as illustrated in Figure 3.2-13.

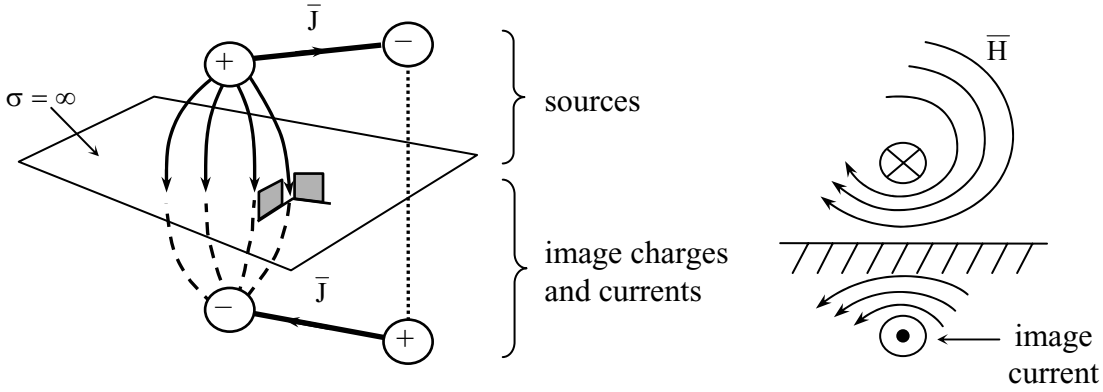

Figure 3.2-13 Image currents and charges

The boundary condition imposed by a perfectly conducting sheet ($\sigma = \infty$) is that the electric fields intersecting it must be perpendicular, and magnetic fields at the boundary must be parallel to the surface, as suggested in Figure 3.2-13. But these boundary conditions for the electric

fields are equivalent to those imposed by having image charges and currents on the opposite side of the conductor, as suggested in the figure. Since the electric fields satisfying the boundary conditions imposed by the conducting sheet are identical to those produced by the boundary conditions imposed by having image charges and currents instead, the electric and magnetic fields associated with the image charges and currents must be identical above the ground plane. Replacing mirror surfaces with equivalent image charges and currents greatly simplifies computing the total radiated fields because those fields can be found by superimposing the fields radiated by both the original and image sources. Superposition of the effects of such similar sources is often handled by treating the separate elements as components in an antenna array.

An example of the application of the antenna array equation (3.2.33) to long-wire antennas is suggested in Figure 3.2-14. We may consider the full-wavelength antenna as composed of two half-wavelength antennas having identical radiating properties, one excited positively and the other negatively, or 180° out of phase. The array factor exhibits maxima along the positive and negative z-axis, and nulls in the x-y plane where the contributions from the two elements cancel because they are out of phase. The element factor is simply that of a half-wavelength dipole, and is shaped like a doughnut with a vanishingly small hole. When the dumbbell-shaped array factor is multiplied by the doughnut-shaped element factor the result is the four-lobed pattern suggested in Figure 3.2-14. Arriving at this result is much quicker using the antenna array factorization approach than it would be by direct computation.

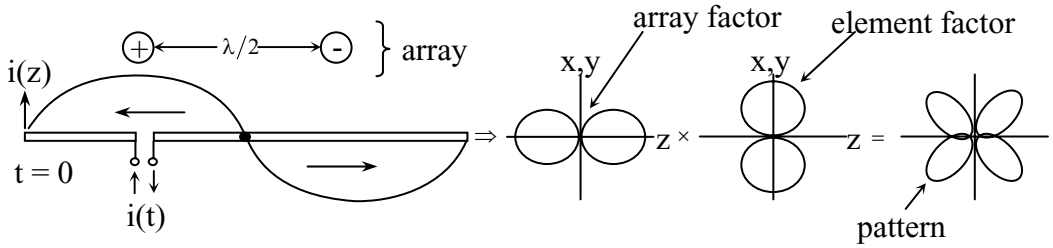

Figure 3.2-14 Full-wavelength antenna considered as a two-element array

A more elaborate long-wire antenna that can be modeled as an antenna array is the antenna used by Jansky in 1927 when discovering cosmic radiation and opening the field of radioastronomy. The basic concept is illustrated in Figure 3.2-15, where a long wire composed of 13 half-wavelength straight segments is driven from a single central driving point excited by I_o.

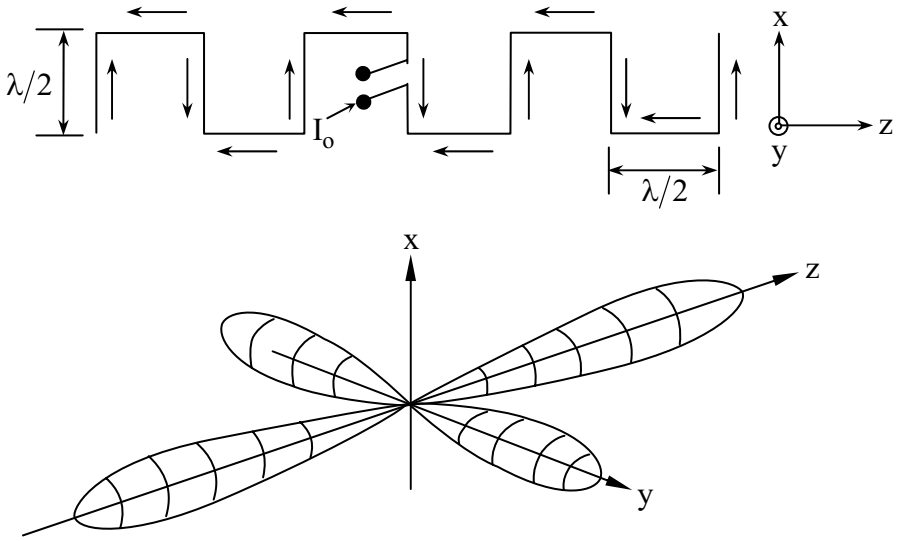

Figure 3.2-15 Jansky long-wire antenna

If we consider the bent wire to be an open-circuited TEM transmission line, we would expect current nulls spaced at intervals of λ/2 relative to the ends of the wire; these current nulls would occur at the bends in the wire, and currents in the straight segments would periodically have the directions illustrated. For this instantaneous current distribution, the radiated phasors from each of the vertical x-oriented wires would add in phase for radiation in the ±z directions. Because the electric field in the ±z directions would be 7 times the field contributed by one element alone, the antenna gain in these directions would be 7^2 times greater than the gain for only one element radiating in that direction. Radiation in the ±y directions has both an x- and a z-polarized component. There are six z-directed currents in the antenna, and so the z-polarized gain in the ±y directions is 6^2 times the gain available from one such current element. In contrast, the x-directed currents radiate phasors in the ± y-directions which cancel, all but one, and so the x-polarized gain in the ± y-direction is exactly that of a single current element. Calculation of the exact antenna gain would require integration of the radiated power over 4π steradians, and comparison to isotropic radiation. The radiation pattern of the Jansky antenna is also suggested in Figure 3.2-15. The actual Jansky antenna employed two such bent wires parallel to one another spaced λ/2 apart so that the y-radiated signals canceled, eliminating all but the two z-directed main lobes.

The long-wire antenna of Figure 3.2-11 can also be considered as an array antenna for which the contributions of each resonant half-wavelength segment tend to cancel in directions perpendicular to the wire and along the main axis, so the main lobe is conical, radiating at an angle of perhaps 20° to the main axis of the wire, as suggested in Figure 3.2-16.

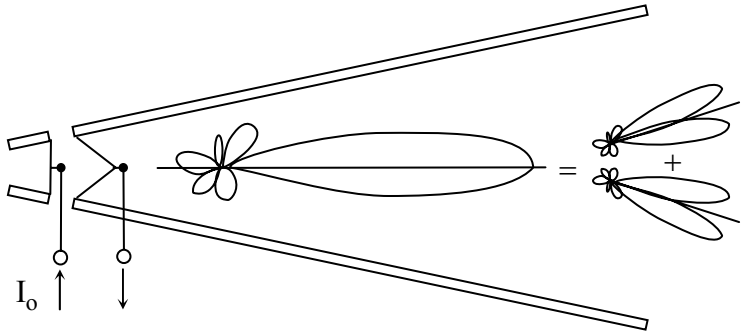

Figure 3.2-16 Vee antenna composed of two long-wire antenna elements

The antenna pattern created by the two wires at an angle is dominated by the coherent addition of the signals radiated by the two cones in their overlapping central direction, as suggested by the figure. Although such an antenna has many sidelobes, it does have the virtue of a very wide bandwidth, simple construction, and substantial directivity for very long wavelengths. Because such antennas are typically suspended from poles above a ground plane, the image of the vee antenna doubles the effective size of the array and produces an enhancement at an angle above the horizon that depends on the height of the antenna and the wavelength of interest. Thus such antennas are effective at launching long wavelengths toward the ionosphere for long distance communications.

Another common use of mirrors is shown in Figure 3.2-17. In this case the current $\underline{I}_o$ drives a half-wavelength dipole located λ/4 in front of a parallel reflecting mirror. The mirror produces an out-of-phase image current λ/2 from the driving current so that the two radiated components add coherently in the forward z direction. The overall pattern can be understood as the product of the array factor, which has a broad lobe in the +z direction, and the element factor, which is the doughnut-shaped antenna pattern of a single half-wavelength antenna with a vanishingly small hole oriented along the axis of the radiating dipole. the product resembles the array factor, which is non-zero only in +z direction.

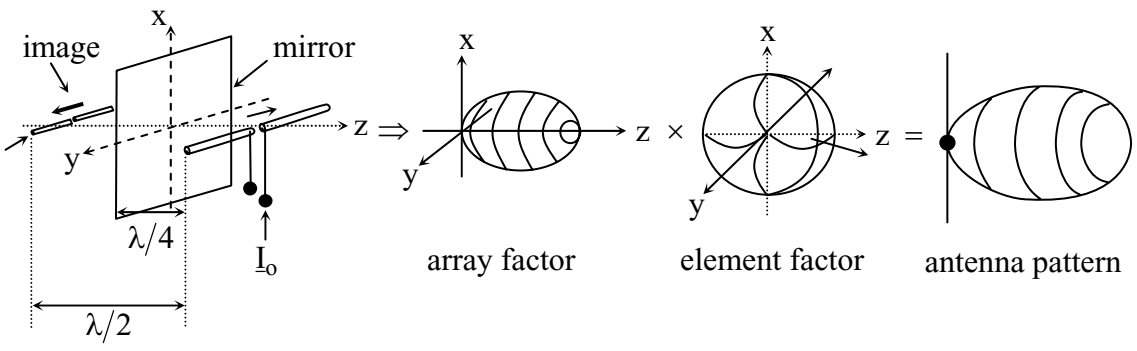

Figure 3.2-17 Half-wavelength dipole positioned λ/4 in front of a mirror

Another common use of a mirror is with a short-dipole wire antenna mounted perpendicular to the conducting surface of an automobile, as illustrated in Figure 3.2-18. In this case the vehicle surface approximates a ground plane through which a wire of length D extends at right angles; a radio receiver is connected to the base of the wire antenna and to the ground plane. The image current corresponding to the effects of the ground plane is in the same direction as the current flow above the ground plane, as suggested in the figure, so that the combined currents are those of a short-dipole antenna of physical length 2D and effective length D, yielding a traditional doughnut-shaped short-dipole antenna pattern. Such antennas are commonly used for automobile radios in the broadcast band near 1 MHz (λ=300m) and have a length of ~1 meter. In the FM band a one-meter wire is approximately $\lambda/2$ long for channels near 100 MHz. The irregular shape of most cars and their windows effectively "blurs" the equivalent image charge with limited consequences.

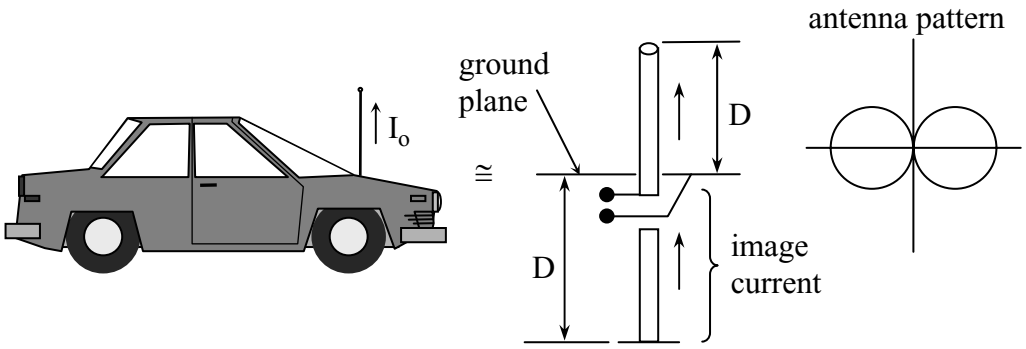

Figure 3.2-18 Short-dipole automobile antenna

Most wire antennas are very simple structures for which analytic methods give acceptable approximate solutions. Recent computer tools have advanced in power so that they are generally used instead. These software tools now permit reasonably complex antennas to be analyzed with useful accuracy. In general, the more highly reactive and complex the antenna structure and the greater the total length of wire to be studied, the less accurate are the resulting computed antenna pattern and impedance predictions, particularly in the nulls or low-amplitude sidelobes.

A representative modern antenna design technique is that of genetic algorithms which can produce antennas of almost random configuration. Genetic algorithms optimize with respect to some desired performance metric, which typically includes a cost for deviations from the desired target pattern and impedance, plus a cost for the mass of the resulting structure. Current commercial software computes antenna patterns with sufficient accuracy to permit genetic design of wire antennas of a few wavelengths if they are not too reactive. Genetic algorithms randomly vary a vector that characterizes the antenna so as to optimize the performance metric.

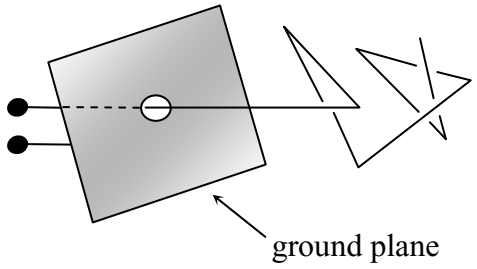

Figure 3.2-19 Example of a wire antenna designed using a genetic algorithm

Vector parameters might include, for example, the lengths of each straight segment of the antenna quantized in units of $\lambda/16$, and the angles between successive elements, similarly quantized. Experiments typically yield rats-nest configurations similar to that of Figure 3.2-19 that agree reasonably well with physical verification units. In general, constraining the configuration to design families or symmetries known to perform well can accelerate this optimization process.

Antennas can also be constructed of multiple wires, one combination of elements being connected to the transmitter or receiver, and the others being freestanding elements that coherently scatter radiation to produce the overall antenna pattern. Scattering can be understood as reception by a short-circuited dipole, followed by re-radiation by the induced current, as discussed later

The receiving properties of dipole antennas are also extremely important in their own right, and are derived below, beginning with the open-circuited short-dipole antenna illustrated in Figure 3.2-20. The results can be generalized to any reciprocal antenna The figure assumes a uniform plane wave is incident from $\theta = 90°$ with linear polarization in the z direction.

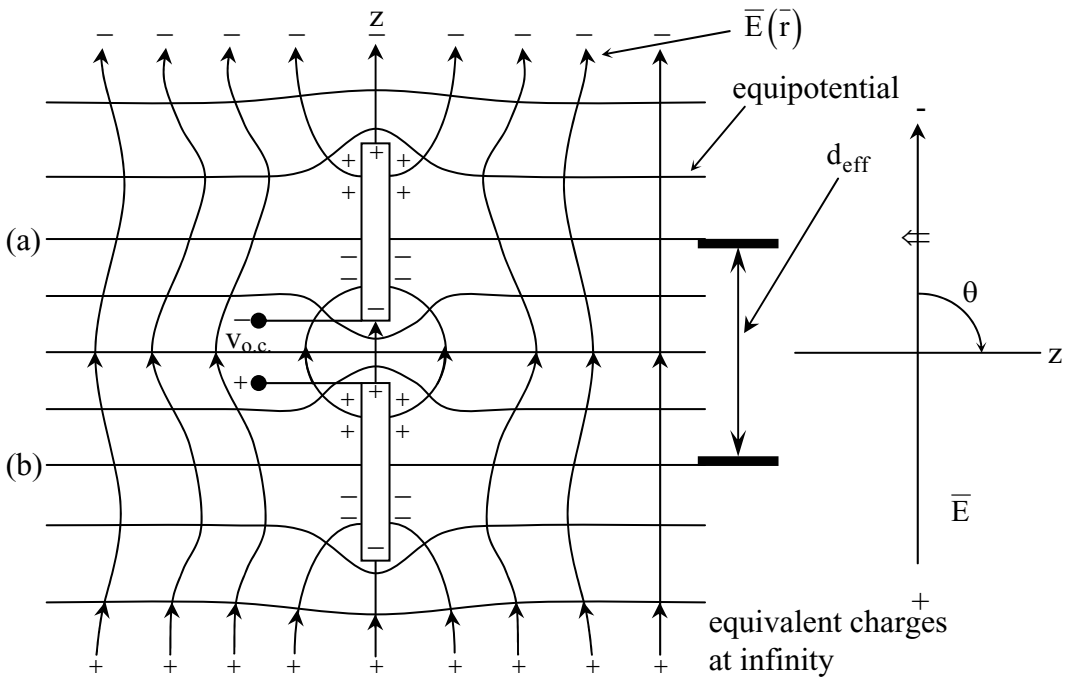

Figure 3.2-20 Electromagnetic fields and potentials near a short-dipole antenna intercepting a plane wave

Because this structure is much smaller than a wavelength, the electromagnetic field solution will be quasistatic, free from the complexities imposed by dynamics or radiation. That is, we can assume that a uniform electric field applied to the short-dipole originates from positive and negative charges at infinity, as illustrated. The symmetry of this configuration mandates a planar equipotential surface at the midpoint between the two halves of the dipole and perpendicular to the applied electric fields, as illustrated. If the wires are sufficiently thin, similar planar equipotentials located at (a) and (b) bisect each of the two wire elements. These two equipotentials differ by the open-circuit voltage $v_{oc.}$ at the antenna terminals. Other equipotential surfaces are parallel to the first, but must deflect around the conductors, also as illustrated. Once these equipotential surfaces are established, it is easy to visualize the associated electric fields, which must be everywhere perpendicular to the equipotentials. Far from the dipole antenna the electric field is straight and parallel to the z-axis, while closer to the dipole it becomes more sinuous. Some of the electric fields lines actually intersect the metallic elements, but always perpendicular to their perfectly conducting surfaces. If the distance between the planes (a) and (b) is d_{eff}, then it is easy to see from the figure that:

$$v_{o.c.} = -\bar{E} \bullet \hat{z}\, d_{eff} = -E \sin\theta\, d_{eff} \tag{3.2.34}$$

where θ is the angle between the z-axis and direction of arrival for the incidental plane wave. Thus $v_{o.c.}$ is maximum when the incident electric field is parallel to the dipole axis.

Equation 3.2.34 can be combined with the equivalent circuit for a passive short-dipole antenna to yield the maximum power that can be extracted from the incident wave. Figure 3.2-21 illustrates how such an antenna equivalent circuit might be matched to receivers so that the maximum power can be extracted.

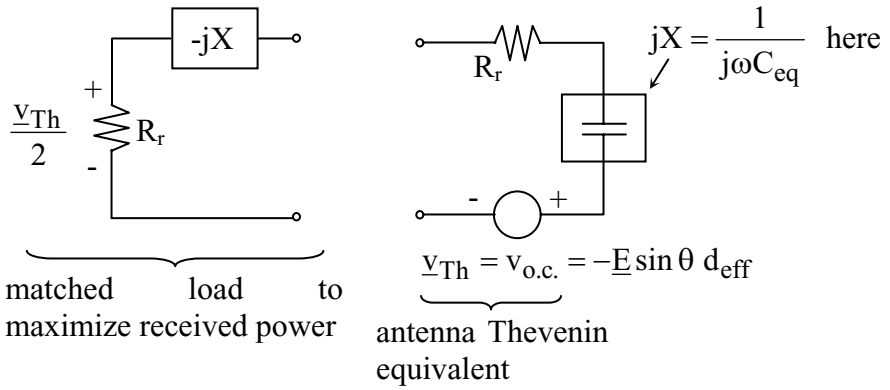

Figure 3.2-21 Equivalent circuit for a short dipole antenna matched to maximize received power

The receiver reactance should cancel that of the short-dipole antenna, and the real part of the impedance should equal the radiation resistance R_r of the antenna as given by (3.2.28). When the load is matched to the antenna the maximum power received is:

$$P_{rec} = \left|\frac{\underline{v}_{o.c.}}{2}\right|^2 \Big/ 2R_r = \frac{|\underline{E}|^2 d_{eff}^2}{8\, R_r}\sin^2\theta \tag{3.2.35}$$

$$P_{rec} = \frac{|\underline{E}|^2}{2\eta_o}\left(\frac{\lambda^2}{4\pi}\bullet\frac{3}{2}\sin^2\theta\right) \tag{3.2.36}$$

where the open circuit voltage is divided across the two resistances. Equation 3.2.36 follows from 3.2.35 by substituting the radiation resistance expression (3.2.28) for a short-dipole antenna and utilizing the fact that the effective length d_{eff} for a transmitting short-dipole antenna is the same d/2 obtained here in the thin-wire limit. Equation 3.2.36 represents the product of the incident flux density (W m^{-2}) and a factor that is a function only of the antenna and incident wave geometry; this second factor is the effective area $A_e(\theta,\phi)$ of the short-dipole antenna:

$$A_e(\theta,\phi) = \frac{\lambda^2}{4\pi}G(\theta,\phi) \tag{3.2.37}$$

where $G(\theta,\phi) = (3/2)\sin^2\theta$ (3.2.24). This very important result, which relates the gain and effective area of a short-dipole antenna by a simple factor $\lambda^2/4\pi$, can be extended to all reciprocal antennas, as follows.

Equation 3.2.37 applies only if the antenna is reciprocal, which is generally the case if there are no magnetized plasmas or ferrites in its vicinity. Equivalently, if we regard a set of antennas linked by a reciprocal electromagnetic medium as a single N-port network characterized by the impedance matrix $\underline{\underline{Z}}$ or an equivalent scattering matrix $\underline{\underline{S}}$, then these matrices must equal their own transposes. This is true if the permeability, permittivity, and conductivity matrices characterizing all media in the vicinity of the antenna also equal their own transposes.

To prove that (3.2.37) applies to an arbitrary antenna of effective area A_1 and gain G_1, consider it as radiating towards a receiving short-dipole antenna characterized by A_2 and G_2 in the direction of the transmitting antenna, as illustrated in Figure 3.2-22. The figure also illustrates equivalent circuits for both the general antenna and the short-dipole antenna, where the mutual impedance coupling these two antennas is $\underline{Z}_{21} = \underline{Z}_{12}$, assuming reciprocity.

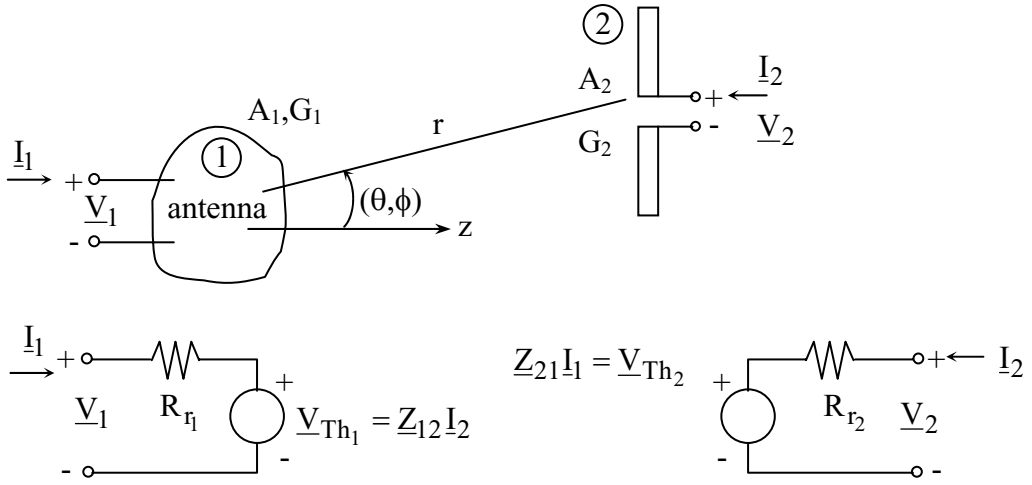

Figure 3.2-22 Equivalent circuits for two coupled antennas

The power P_{r_1} received by antenna 1 from antenna 2 can be characterized in terms of their equivalent circuits and also in terms of the gains and effective areas of the two antennas, as follows:

$$P_{r_1} = |\underline{Z}_{12}\underline{I}_2|^2 / 8R_{r_1} = P_{t_2}\frac{G_2}{4\pi r^2} \bullet A_1 \qquad (3.2.38)$$

$$P_{r_2} = \left|\underline{Z}_{21}\underline{I}_1\right|^2 \Big/ 8R_{r_2} = P_{t_1}\frac{G_1}{4\pi r^2}\bullet A_2 \tag{3.2.39}$$

where P_{t_i} is the power transmitted by antenna i. By dividing (3.2.38) by (3.2.39) we obtain:

$$P_{r_2}/P_{r_1} = \frac{G_1 A_2 P_{t_1}}{G_2 A_1 P_{t_2}} \tag{3.2.40}$$

$$\frac{A_1}{G_1} = \frac{A_2}{G_2}\bullet\frac{P_{t_1}}{P_{t_2}}\bullet\frac{P_{r_1}}{P_{r_2}} \tag{3.2.41}$$

If we again divide (3.2.39) by (3.2.38), but this time using the circuit descriptions, we obtain:

$$\frac{P_{r_1}}{P_{r_2}} = \frac{\left|\underline{Z}_{12}\underline{I}_2\right|^2 R_{r_2}}{\left|\underline{Z}_{21}\underline{I}_1\right|^2 R_{r_1}} = \frac{P_{t_2}}{P_{t_1}} \tag{3.2.42}$$

if $\underline{Z}_{12} = \underline{Z}_{21}$. Substituting (3.2.42) into (3.2.41) we obtain:

$$\frac{A_1}{G_1} = \frac{A_2}{G_2} = \frac{\lambda^2}{4\pi} \tag{3.2.43}$$

Thus, if all media in or near an antenna are reciprocal so that the permeability, permitivity, and conductivity matrices equal their own transposes, then (3.2.43) applies. This is an extremely useful result because, as we have seen, computing the radiating properties and gains G of most antennas can be straight-forward if their current distributions are known.

As an example consider the effective area of a short-dipole antenna that is matched to its receiver. Since the maximum gain of a short-dipole antenna is 1.5, (3.2.37) implies the effective area must be less than $3\lambda^2/8\pi$, which is approximately $(\lambda/3)^2$, independent of the effective length of the short-dipole antenna. This implies that the capture cross-section or effective area of a short-dipole antenna can be vastly larger than the physical cross-section of the antenna itself, as suggested in Figure 3.2-23. In the 1-MHz AM radio band where $\lambda \cong 300$ m, this maximum theoretical effective area is larger than a football field even though most radios and their antennas fit in the palm of a hand. In practice the effective areas of these antennas are much smaller than this maximum, but still substantially larger than the radio itself. One reason these extremely large effective areas can seldom be achieved in practice is that we have assumed the antenna is matched to the receiver. But (3.2.28) yields very low values of radiation resistance for such antennas, and therefore any matching circuits must be extremely reactive and narrow band. Such perfect matches are generally not obtained in practice. On the other hand, a cell phone

operating at 900 MHz can use a half-wavelength dipole six inches long and be perfectly matched with an effective area up to 4 inches in diameter.

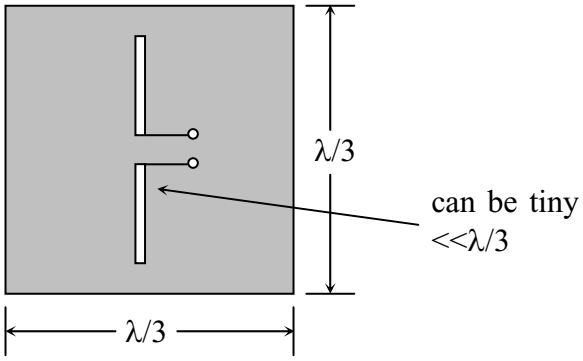

Figure 3.2-23 Nominal effective area of a matched short-dipole antenna

In multi-conductor antennas often only one wire is connected to the receiver or transmitter, and the others are *parasitic wires*. These wires scatter radiation in their immediate vicinity. Consider a short-circuited dipole and its equivalent circuit, as illustrated in Figure 3.2-24.

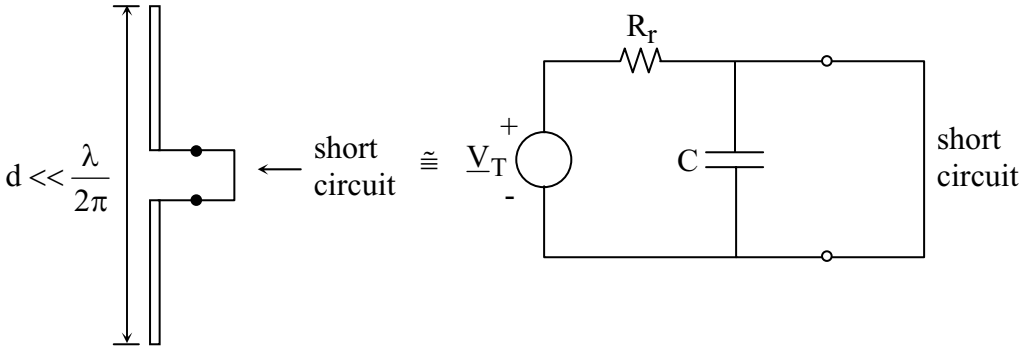

Figure 3.2-24 Equivalent circuit of a short-circuited short-dipole antenna at frequency f

The equivalent circuit of a short-dipole antenna has a shunt capacitance associated with electric charges that aggregate on the wire surfaces with increasing charge density towards the central antenna terminals. Although the illustrated circuit lumps this capacitor at the output terminals, in fact it is distributed across the radiation resistance and is difficult to model accurately. If we restrict ourselves to a half-wavelength dipole for which the reactance is zero, then the scattered power might be related to the power dissipated in the radiation resistance R_r in Figure 3.2-24. That scattered power is:

$$P_{scat} = |\underline{V}_T|^2 / 2R_r = S_{inc} \bullet A_{scat} \tag{3.2.44}$$

where the incident flux is:

$$S_{inc} = |\underline{E}|^2 / 2\eta_o \tag{3.2.45}$$

This can be compared to the power received by the same short-dipole if it were matched:

$$P_r = S_{inc} \bullet A = |\underline{V}_T|^2 / 8R_r \tag{3.2.46}$$

Comparison of (3.2.44) and (3.2.46) reveals that the scattering cross-section is approximately 4 times that of a matched half-wave dipole or $3\lambda^2/2\pi$, corresponding to an approximate diameter of 0.7λ.

A half-wave dipole stores most of its average electric and magnetic energy, which balance at resonance, within a few wire radii. The *resonator Q* of the half-wavelength resonance can be computed using the standard formula:

$$Q = \frac{\omega_o W_T}{P_d} \tag{3.2.47}$$

where the resonant frequency $\omega_o = 2\pi c/\lambda$, W_T is the total stored energy in the vicinity of the resonant dipole, and P_d is the power dissipated in the resonator or, in this case, the power radiated. For typical wire diameters and lengths the computed Q is approximately 10, and is defined as the number of radians before the average un-driven resonator current drops to 1/e of its starting value.

One simple example where this phenomena was used involved placement in low earth orbit of large numbers of half-wavelength needles designed to scatter centimeter-wavelength communication signals should the normal ionosphere become more highly ionized in a nuclear conflict so as to block traditional wavelengths. A surprisingly small payload of needles is adequate to provide a reflective belt circling the earth along the trajectory of a single launch rocket.

Resonant wires may also be used to substitute for a reflecting mirror, as suggested in Figure 3.2-25.

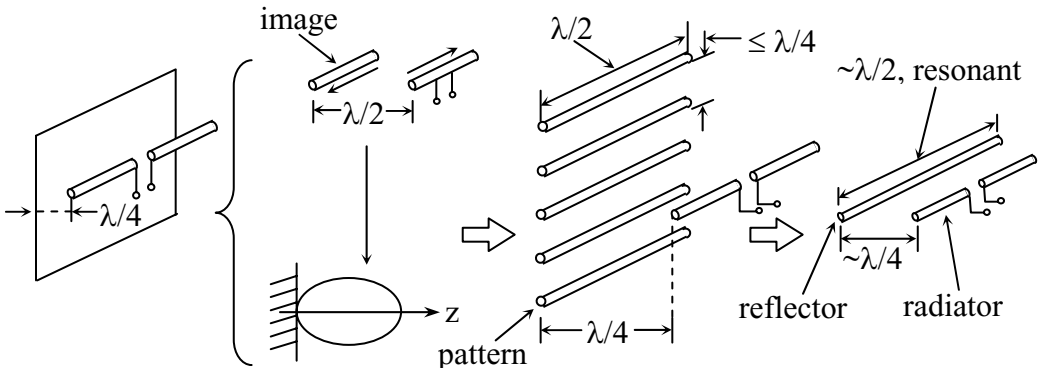

Figure 3.2-25 Resonant half-wave dipoles used as reflectors in antenna arrays

Although multiple resonant wires reflect better than a single resonant wire, they cannot equal the reflectivity of a solid planar surface. Nonetheless, useful reflections can be obtained close to the resonant frequency of an isolated reflecting half-wavelength dipole. By slightly lengthening or shortening that isolated free-standing wire, its current will slightly lead or lag the phase of the current that would flow at resonance, as can be seen for a RLC resonator:

$$\underline{V} = \underline{I}\underline{Z} = \underline{I}\left(R + Ls + 1/Cs\right) = \underline{I}|\underline{Z}|e^{j\phi_z} \tag{3.2.48}$$

This equation corresponds to the circuit and phase response illustrated in Figure 3.2-26. If we lengthen the wire to reduce its resonant frequency ω_o, then the phase lag of the induced current can be increased up to a maximum of 90°, whereas shortening the wire permits a phase lead up to a maximum of 90°. Reasonable fractions of these maximum phase shifts are achieved when the frequency is offset by ω_o/Q. The resonator Q and the sensitivity of phase shift to antenna length can be increased by increasing the ratio of stored energy W_t to power radiated by using thinner wires.

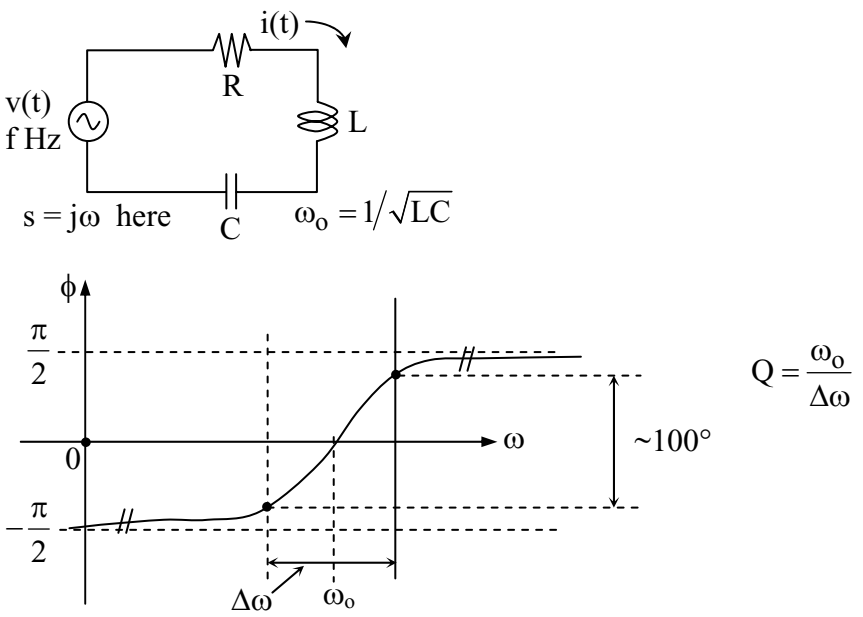

Figure 3.2-26 Resonator circuit and its frequency-dependent phase relationships

In some cases the phase relationships between the driven and parasitic currents are readily determined. For example, consider the antenna pictured in Figure 3.2-27a for which the parasitic wire is a short distance from the driven current. The two closely spaced wires form a TEM line which is open-circuited at the ends so that the parasitic current $\underline{I}_2 \cong \underline{I}_1$ and their radiated fields approximately cancel in all directions. The impedance at the antenna terminals is approximately a short circuit, which is the TEM impedance $\lambda/2$ from an open circuit, consistent with the inability of this antenna to radiate strongly. Figure 3.2-27b portrays the situation when the resonant parasitic $\lambda/2$ dipole carries the induced current $\underline{I}_2$, which radiates to the left in phase with the signal radiated in that direction by $\underline{I}_1$. This maximizes the gain in that direction relative to the gain in the rear direction; such an isolated wire behind the driven element is called a *parasitic reflector*.

Elements called *parasitic directors* are typically positioned a distance D in front of the driven element, as illustrated Figure 3.2-27c. The phase shifts introduced by the distance D and by the detuning of the parasitic element, accomplished by shortening it slightly, act together to partially cancel the two signals radiated to the left, and to superimpose them positively for radiation to the right.

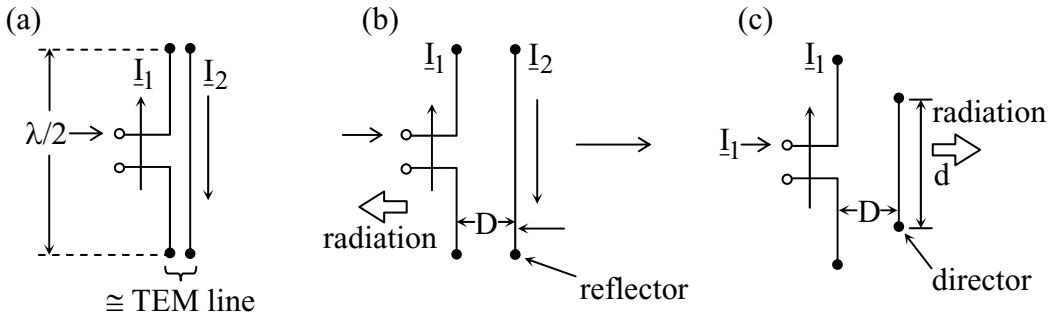

Figure 3.2-27 Half-wave dipoles driving parasitic wires: (a) insufficient radiator, (b) reflector, (c) director

Figure 3.2-28 illustrates a popular configuration called a Yagi antenna, which typically has multiple reflectors and directors, all cut to slightly different lengths to maximize the front-to-back gain ratio over the desired bandwidth. Originally the design of Yagi antennas was largely an art reinforced by experiment. Today computers can solve Maxwell's equations with sufficient accuracy that Yagi antennas can more readily be optimized with respect to their directivity, sidelobes, and reactance over the desired bandwidth.

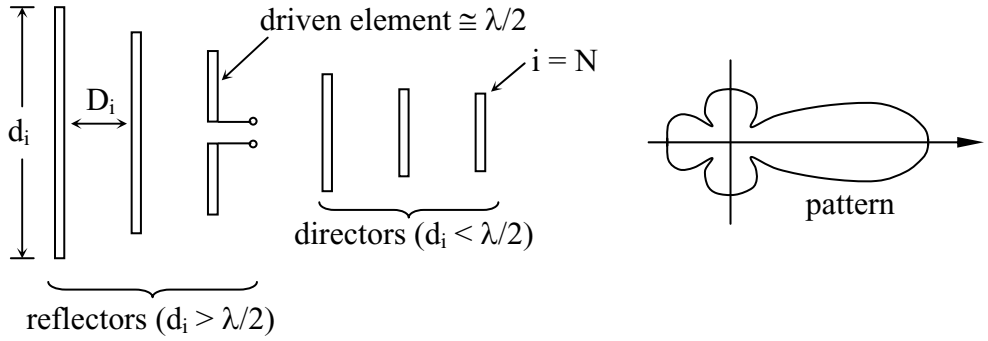

Figure 3.2-28 Yagi antenna

Another very popular wire antenna is the half-wave *folded dipole antenna* illustrated in Figure 3.2-29, which has two parallel wires very close together that behave as a TEM line. The line is short circuited at both ends so that at the antenna terminals both λ/4 sections of antenna behave as open circuits and no equal-and-opposite TEM currents can exist on the antenna. That is, $\underline{I}_A = \underline{I}_B$, which is the "common mode" of propagation for a TEM structure, and involves no TEM currents. Because the relationship between the terminal current $\underline{I}_o$ and the current $\underline{I}_A$ is the same as for the standard half-wave dipole, and since the radiated power now is quadrupled in every direction because the total drive current equals $\underline{I}_A + \underline{I}_B = 2\underline{I}_A$, the total radiated power and input impedance is quadrupled relative to that for a half-wave dipole driven by $\underline{I}_o$. Thus the

input impedance of a half-wave folded dipole is approximately 300 ohms, where this higher impedance is sometimes desired in practice.

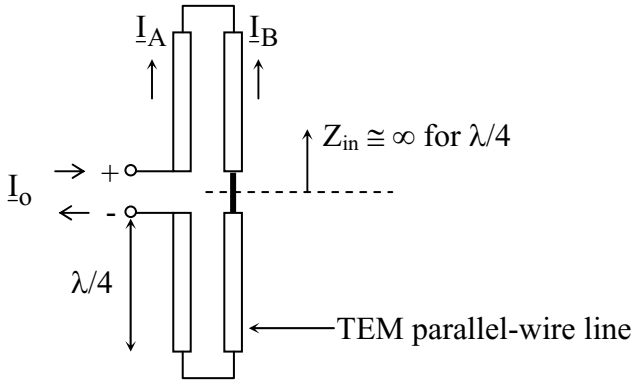

Figure 3.2-29 Half-wave folded dipole antenna

Small alterations are often made to folded dipole antennas to compensate for the fact that the propagation velocities for the TEM mode and the common mode are different because the balance between electric energy storage inside and outside the dielectric which supports the two wires is different for these two modes. The electric field distributions for the TEM and common modes are shown in Figure 3.2-30a and b, respectively. If the dielectric constant of the plastic supporting the two wires is greater than unity, then the TEM mode, which has more of its electrical energy inside the plastic and therefore interacts more strongly with it, will propagate more slowly than the common mode, for which the electric field exits the plastic more rapidly. As a result the length of the resonator for the TEM mode will be slightly less than the free-space resonant wavelength corresponding to the common mode currents. This discrepancy can be resolved by making both the TEM and common mode structures simultaneously resonant through use of a "pig-tail" extension at the ends of the antenna, as illustrated in Figure 3.2-30c.

Another widely used wire antenna with useful directivity is the *helical antenna*, illustrated in Figure 3.2-31. The dimensions of the helix--D, L, d-- are chosen so that the waves radiated by each turn of the helix will add in phase in the forward direction because their contributions differ in phase by an amount just equal to the propagation time between turns. If we assume the antenna current propagates along the helix in a nominal TEM mode at velocity c, then we want the propagation time for one turn, related to the hypotenuse of a triangle, to differ from the distance between turns by an integral number of wavelengths, i.e.:

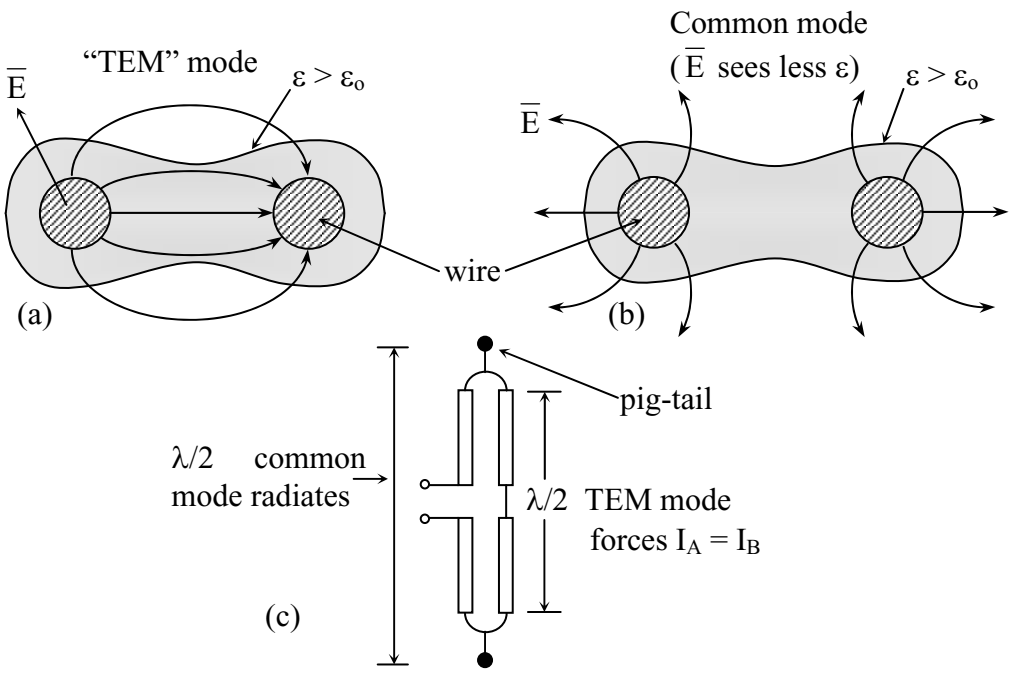

Figure 3.2-30 Half-wave folded dipole antenna with pig-tail.

$$\sqrt{(2\pi R)^2 + d^2} - d = n\lambda \tag{3.2.49}$$

For the special case where D=2d, then d must equal $n\lambda\Big/\left[(4\pi^2+1)^{1/2}-1\right]$, or ~$\lambda$/2.3. When the antenna is viewed from the front, as suggested in Figure 3.2-31b, all the contributions add in phase from each portion of the circle. Currents on one side of the circle will be 180° out of phase with those on the other, thus they point in the same direction, creating an effective dipole source that rotates in the direction of the helix at f Hz. If the helix is too short then the reflections of $\underline{I}(\ell)$ from the open end of the wire will produce standing waves that weaken the intended radiation pattern; for this reason more than 5 or 10 turns are commonly employed so that the currents reaching the end of the helix are relatively weak and less disruptive. The mirror effectively doubles the length of the helix, and the image currents radiate in phase with the directly radiated signals. Because of interwire interactions the optimum values to D, L, and d may vary from their idealized values.

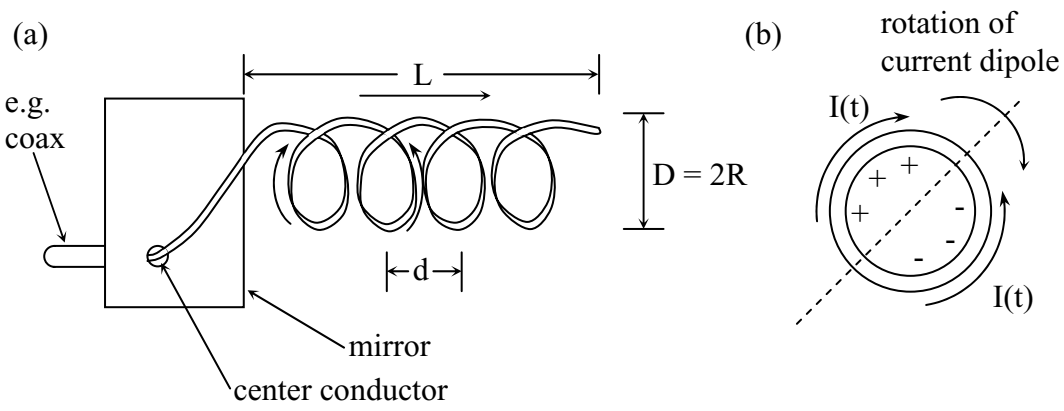

Figure 3.2-31 Helical antenna (a) side view (b) view from front

Very broad bandwidth wire antennas can be obtained with log-periodic structures, such as the dipole array illustrated in Figure 3.2-32. The log-periodic half-wave dipole antenna is driven from one end with spacings between adjacent dipoles so that the waves radiated in the forward direction add in phase while those radiated in the back direction more nearly cancel. Because significant radiation occurs with only two or three such wires fully excited, the dipoles can have slowly varying length, as illustrated. In general, for a mid-band wavelength the dipoles near the input are too short to radiate effectively and so they contribute a reactance, and dipoles past the resonant active portion of the antenna indicated by the dashed circle in the figure are not sufficiently excited to radiate significantly. Because the active part of the antenna shifts toward shorter dipoles at high frequencies and towards longer dipoles at low frequencies, it can be shown that the radiation pattern and impedance are approximately frequency independent in the limit of a continuum of dipoles. The resulting pattern is approximately that suggested in the figure. The same log-periodic concept can be applied to helical antennas, which also would be driven from the small outward-pointing end; such helical antennas resemble that in Figure 3.2-31a, but with a conical shape having a base near the mirror.

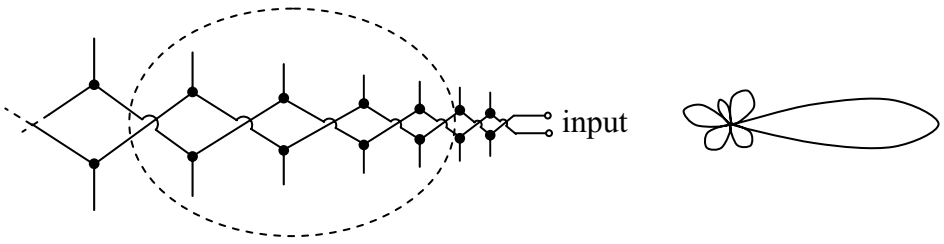

Figure 3.2-32 Log-periodic antenna

Because half-wave dipoles, folded dipoles, helical antennas, and other radiators are often fed by TEM lines and are located in front of a mirror, they often employ additional structures called baluns which permit the mirror to be grounded to one of the two feed wires without significantly disturbing the radiated pattern, as suggested in Figure 3.2-33c. A half-wave dipole with an

ungrounded mirror is shown in Figure 3.2-33a, and the same antenna simply soldered to the mirror is shown in Figure 3.2-33b. The problem with this soldered configuration is that the current which should flow into wire B will flow instead down the outside of the coaxial conductor C to the mirror, effectively halving the length of the dipole radiator. One standard solution is to add the wire D as illustrated in Figure 3.2-33c, which runs parallel and close to the conductor C, forming a TEM line with it which is shorted at the mirror and therefore presents an open circuit $\lambda/4$ away at the wire A. This open circuit precludes any TEM currents flowing toward the mirror on the outside surface of the coaxial tube C and forces those currents to travel instead outwards along conductor B.

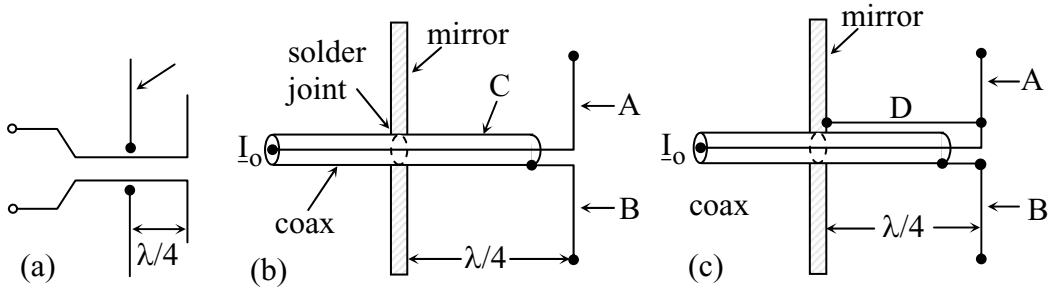

Figure 3.2-33 Balun configurations to permit grounded mirrors on wire antennas

3.3 APERTURE ANTENNAS

Aperture antennas are often characterized by a presumed known electric field distribution $\underline{\overline{E}}(x,y)$ within a defined aperture, in contrast with wire antennas that are characterized by a presumed known current distribution. The linear character of Maxwell's equations suggests that the radiated fields can be computed by superimposing contributions from each incremental source comprising the radiator, as is generally done for wire antennas. It is easily shown that a uniform surface current can radiate a uniform electromagnetic plane wave. By extension, a non-uniform plane wave in an aperture can be replaced by an equivalent surface current which locally matches the given aperture distribution, and this current sheet can be used to calculate the radiated fields in the same manner employed for wire antennas. The approximate result of this calculation relating the x-polarized aperture fields $\underline{E}_x(x,y)$ to the far-field solution $\underline{\overline{E}}(\theta,\phi,R)$ is *Huygen's superposition integral*:

$$\underline{\overline{E}}(\theta,\phi,R) \cong \frac{j}{2R\lambda}(1+\cos\,\theta)(-\hat{\alpha}_x)\int_A \underline{E}_x(x,y)\, e^{-j(2\pi/\lambda)r(x,y)}dxdy \qquad (3.3.1)$$

where R is the distance between the observer and aperture origin, and r(x,y) is the distance between any radiating patch and the observer; the corresponding phase lag is $e^{-j(2\pi/\lambda)r(x,y)}$. The coordinates are suggested in Figure 3.3-1. Equation 3.3.1 is written in terms of the x-

polarized plane wave in the aperture which radiates electric fields polarized in the $\hat{\alpha}_x$ direction. In general, the x-polarized and y-polarized aperture fields produce orthogonally polarized far fields that can be superimposed; i.e. they would be computed separately for each polarization and the results would be summed.

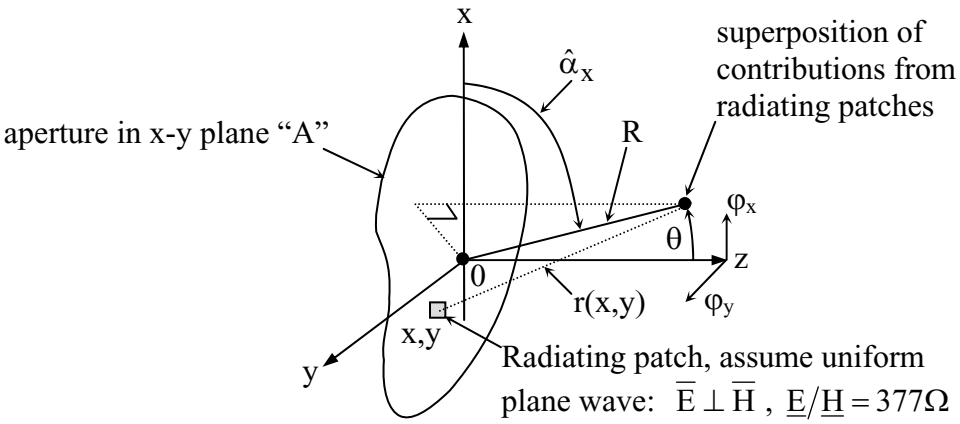

Figure 3.3-1 Geometry for aperture antenna

The radiation geometry and Huygen's integral (3.3.1) are simplified for radiation directions close to the z-axis (see Figure 3.3-2) such that the phase lag associated with r(x,y) can be approximated as:

$$r(x,y) \cong R - x\,\sin\varphi_x - y\,\sin\varphi_y \cong R - x\varphi_x - y\varphi_y \tag{3.3.2}$$

where $\varphi_x, \varphi_y << 1$.

In this far-field limit where (3.3.2) applies, (3.3.1) becomes:

$$\overline{\underline{E}}(\theta,\phi,R) \cong -\frac{e^{-j2\pi R/\lambda}}{2R\lambda}(1+\cos\,\theta)\hat{\alpha}_x \bullet \int_A \underline{E}_x(x,y)e^{+j2\pi(x\varphi_x+y\varphi_y)/\lambda}dxdy \tag{3.3.3}$$

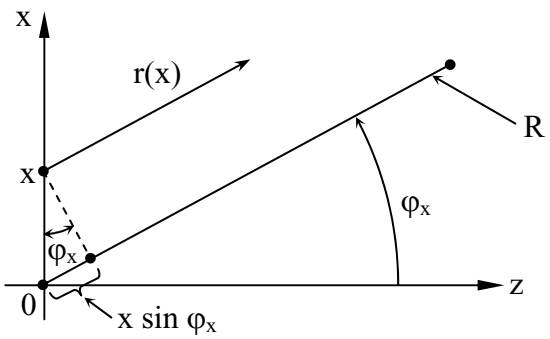

Figure 3.3-2 Geometry for aperture far-field radiation

where the fixed phase lag $e^{-j2\pi R/\lambda}$ has been factored out front. It simplifies notation to define the entire factor in front of the integral sign as the complex vector $\underline{\overline{K}}$, in which case Huygen's principle (3.3.3) becomes:

$$\underline{\overline{E}}\left(\varphi_x,\varphi_y\right)\left(vm^{-1}\right) \cong \underline{\overline{K}} \int_A \underline{E}_x(x,y) e^{+j2\pi(x\varphi_x + y\varphi_y)/\lambda} dxdy \tag{3.3.4}$$

which approximates a Fourier transform. The inverse approximate Fourier transform is:

$$\hat{x}\underline{E}_x(x,y)\left(vm^{-1}\right) \cong \frac{\hat{x}}{\underline{\overline{K}}} \int_{2\pi} \underline{E}_x\left(\varphi_x,\varphi_y\right) e^{-j2\pi(x\varphi_x + y\varphi_y)/\lambda} d\varphi_x d\varphi_y \tag{3.3.5}$$

The Fourier relationship defined in (3.3.4) and (3.3.5) is extremely useful for understanding aperture antennas, and can be represented diagramatically as:

$$\begin{array}{ccc} \underline{E}_x(x,y)\left(vm^{-1}\right) & \tilde{\leftrightarrow} & \underline{E}_x\left(\varphi_x,\varphi_y\right)\left(vm^{-1}\right) \\ \downarrow & & \downarrow \end{array} \tag{3.3.6}$$

$$R_{\underline{E}_x}\left(\tau_{\lambda_x},\tau_{\lambda_y}\right)\left(vm^{-1}\right)^2 \tilde{\leftrightarrow} \left|\underline{E}_x\left(\varphi_x,\varphi_y\right)\right|^2\left(vm^{-1}\right)^2 = 2\eta_o S\left(\varphi_x,\varphi_y\right) \tag{3.3.7}$$

The double-headed arrows in (3.3.6) and (3.3.7) indicate the approximate Fourier relationship between the two variables, and the downward-pointing arrows indicate the irreversible nature of those relationships. The units of radiated flux $S\left(\varphi_x,\varphi_y\right)$ are watts/meter^2. Note how the Fourier relationship of (3.3.6) permits us to relate immediately the autocorrelation function of the aperture field distribution $R_{\underline{E}_x}\left(\tau_{\lambda_x},\tau_{\lambda_y}\right)$ to the squared magnitude of the far-field $\underline{E}_x\left(\varphi_x,\varphi_y\right)$, where we define aperture displacement $\tau_{\lambda_x} = (x_1 - x_2)/\lambda$.

We can compute the directivity of an aperture antenna by substituting (3.3.4) into the definition of directivity (3.1.2):

$$D(\theta,\phi) \triangleq \frac{P(\theta,\phi,f,R)}{P_{TR}\big/4\pi R^2} \tag{3.3.8}$$

where P_{TR} is the total power radiated by the aperture. This yields, under the far-field pariaxial approximation of (3.3.4), the result:

$$\begin{aligned} D\left(\varphi_x,\varphi_y\right) &\cong \frac{(1+\cos\ \theta)^2}{2\eta_o\,(2R\lambda)^2}\,\frac{\left|\int_A \underline{E}_x(x,y)e^{j2\pi\left(x\varphi_x+y\varphi_y\right)/\lambda}dxdy\right|^2}{\frac{1}{2\eta_o}\int_A|\underline{E}_x(x,y)|^2dxdy\ \Big/4\pi R^2} \\ &= \frac{\pi(1+\cos\ \theta)^2}{\lambda^2}\,\frac{\left|\int_A \underline{E}_x(x,y)e^{j2\pi\left(x\varphi_x+y\varphi_y\right)/\lambda}dxdy\right|^2}{\int_A|\underline{E}_x(x,y)|^2dxdy} \end{aligned} \tag{3.3.9}$$

where the total power radiated P_{TR} has been computed by integrating over the aperture A the flux density passing through it, rather than integrating the far-field power over 4π steradians. Because we have assumed the aperture has no reactive power elements, they do not contribute to this integral.

A useful upper bound on aperture directivity $D\left(\varphi_x,\varphi_y\right)$ can be found by using the Schwartz inequality:

$$\left|\int fg\ dx\right|^2 \le \left(\int|f|^2\ dx\right)\left(\int|g|^2\ dx\right) \tag{3.3.10}$$

Applying (3.3.10) to (3.3.9) yields:

$$\left|\int_A \underline{E}_x e^{j[\ ]}dxdy\right|^2 \le \left(\int_A|\underline{E}_x|^2\,dxdy\right)\left(\int_A 1^2\,dxdy\right) \tag{3.3.11}$$

$$D(\varphi_x,\varphi_y) \le \frac{4\pi}{\lambda^2} \bullet \frac{\int_A |\underline{E}_x|^2 \, dxdy \bullet A_o}{\int_A |\underline{E}_x|^2 \, dxdy} = \frac{4\pi A_o}{\lambda^2} \tag{3.3.12}$$

where A_o is the physical area of the aperture. But the effective area of any aperture antenna is given by (3.1.8) and (3.1.3):

$$D(\varphi_x,\varphi_y) = \frac{4\pi}{\lambda^2} \bullet \frac{A_e(\varphi_x,\varphi_y)}{\eta_R} \tag{3.3.13}$$

therefore the bound on directivity (3.3.12) becomes:

$$A_e(\varphi_x,\varphi_y) \le \eta_R A_o \tag{3.3.14}$$

where the radiation efficiency $\eta_R \tilde{<} 1.0$.

The effective area of an aperture antenna is less than the physical area for two principal reasons. Mismatch and dissipative losses reduce the radiation efficiency η_R below unity, and if the aperture is not illuminated with uniform intensity and phase, then the aperture efficiency η_A is also less than unity, where we define η_A as:

$$\eta_A \triangleq \frac{A_e(\max)}{\eta_R A_o} \tag{3.3.15}$$

As we shall see, aperture efficiencies typically range between 0.6 and 0.7 for many aperture antennas because uniform illumination and unity efficiency are achieved at the expense of higher side-lobes, which usually are undesirable.

A few aperture distributions lend themselves to simple analytic solutions for antenna directivity. For example, a uniformly illuminated circular aperture antenna of diameter D has directivity:

$$D(f,\theta,\phi) = \left[\frac{\pi D}{\lambda}(1+\cos\ \theta)\right]^2 \Lambda_1^2\left(\frac{\pi D}{\lambda}\sin\ \theta\right) \tag{3.3.16}$$

where the lambda function $\Lambda_1(q)$ is defined as $J_1(q)/q$ and $J_1(q)$ is a Bessel function of the first kind. Because this is an ideal uniformly illuminated aperture with uniform phase, its directivity at $\theta = 0$ is $4\pi A_o/\lambda^2$.

A more general and typical aperture electric field distribution is tapered toward the edges and can often be approximated as:

$$E_x(r) = \left[1 - \left(\frac{2r}{D}\right)^2\right]^P \tag{3.3.17}$$

where r is the radius measured from the center of the aperture, D is the aperture diameter, and the power P characterizes the degree of tapering, as suggested in Figure 3.3-3.

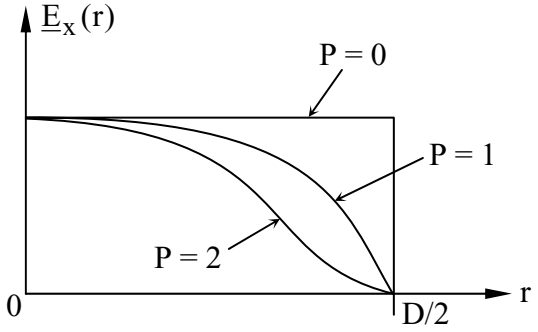

Figure 3.3-3 Idealized aperture excitation functions for circular apertures

The antenna pattern for uniform illumination (P = 0) is illustrated in Figure 3.3-4, which shows quantitatively the sidelobe structure.

Such antenna patterns are often characterized by a few key parameters. The full width at half power $\Theta_{B1/2}$ is 1.02 λ/D for P = 0. The first null θ_{N1} is 1.22 λ/D off-axis for P=0, and the first sidelobe has an amplitude of 17.6 dB below on-axis directivity. These parameters are tabulated in Table 3.3.1 for three values of P.

Table 3.3.1 Antenna pattern parameters for circular apertures

P	$\Theta_{B1/2}$	θ_{N1}	first sidelobe	η_A
0	1.02 λ/D	1.22 λ/D	17.6 dB	1.00
1	1.27 λ/D	1.63 λ/D	24.6 dB	0.75
2	1.47 λ/D	2.03 λ/D	30.4 dB	0.56

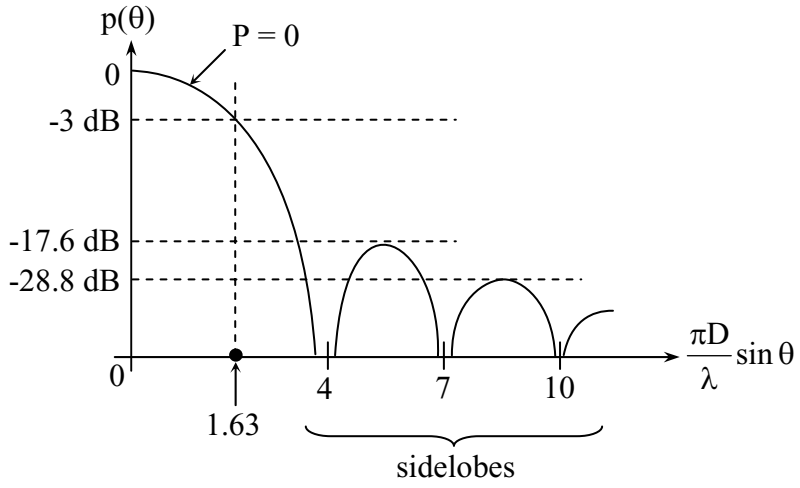

Figure 3.3-4 Antenna pattern for a uniformly illuminated circular aperture

Figure 3.3-5 schematically suggests how the main lobe and sidelobes would originate and be distributed in space for a circular parabolic reflector having a feed located at the parabolic focus. The sidelobes illustrated in Figure 3.3-5 originate primarily from the diffraction pattern of the circular aperture, whereas the back lobes generally originate from spillover of energy radiated by the feed beyond the lip of the reflector, and additional back lobes are produced when the spillover diffracts at the lip of the reflector. Additional sidelobes may be created by scattering from any feed support structure or from other objects in the path of the beam. A compromise often must be made between under-illuminating the reflector (increasing P) to reduce side and back lobes, and over-illuminating the reflector to reduce the beamwidth of the main lobe and increase the gain. In practice this compromise leads to aperture illumination functions close to those produced for P=1. In practice $\theta_{B1/2}$ is 1.2-1.3 λ/D, and the first sidelobe is typically ~20dB below the main beam.

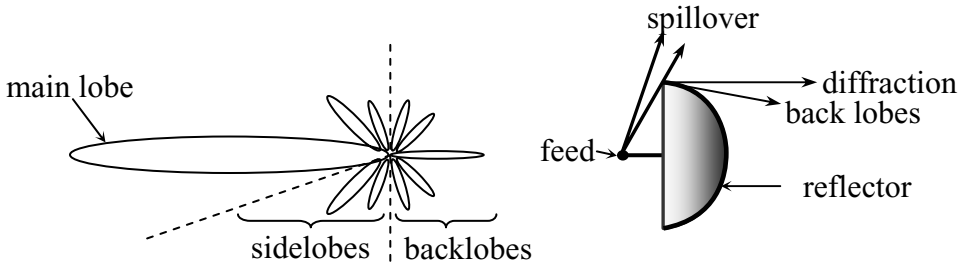

Figure 3.3-5 Sidelobe structure and origin for a parabolic reflector antenna

Some common types of fixed parabolic antennas are shown in Figure 3.3-6. Most parabolic antennas are circularly symmetric with a point focus that produces a uniform phase front after reflection. The cylindrical parabola has a line focus, which has the added virtue that the antenna

can be scanned laterally if the line feed is a phased array. The off-axis paraboloid is simply a segment of a circularly symmetric parabolic reflector, but with a feed positioned so that the reflected radiation does not intercept and scatter from it. This type of antenna has lower sidelobes, particularly for compact structures, and is currently widely used for consumer reception of direct-broadcast satellite television.

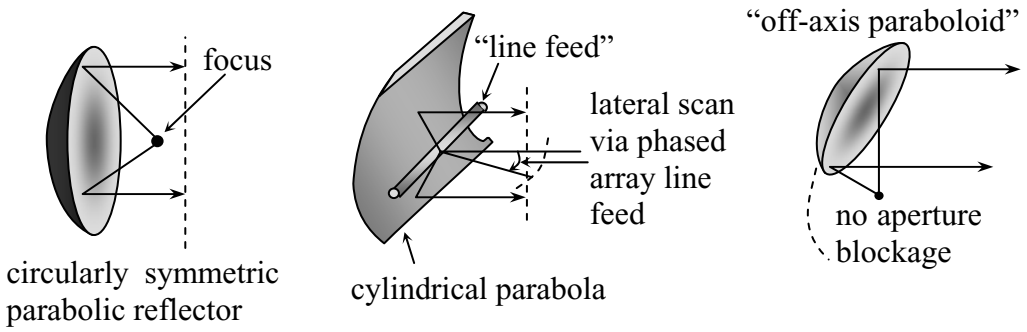

Figure 3.3-6 Common fixed parabolic reflector antennas

Parabolic antennas can be slightly scanned by moving the feed to one side, as suggested in Figure 3.3-7, but at the expense of reduced gain and increased sidelobes. The angle over which a parabolic reflector can be squinted depends strongly on the ratio f/D, where f is the focal length of the parabola, which is the minimum distance between the focus and the parabola. For f/D = 0.5, approximately 3-5 beams with useable gain and sidelobes can be achieved; the gain degradation might be ~1 dB. This assumes the beams cross at their 3-dB points. The number of beams achievable in any direction is $n \cong 12(f/D)^2$, and the number available in two directions θ,ϕ is $N \cong 120(f/D)^4$. Thus parabolic mirrors used to produce images usually have longer focal lengths, as in standard parabolic reflector optical telescopes. If shorter focal lengths are desired for imaging, good lens systems enable many more independent points to be resolved simultaneously.

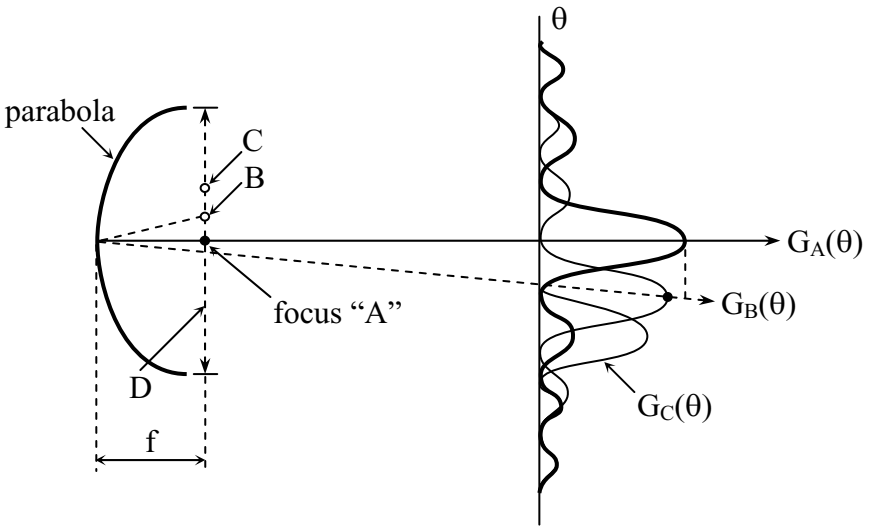

Figure 3.3-7 Antenna patterns for off-axis fed parabolic reflector

Very large apertures can be scanned without moving the reflector itself. One approach is to use a spherical reflector in combination with a line feed that compensates for the phase errors which would be introduced by a point feed, as suggested in Figure 3.3-8. A spherical reflector approximates a parabola only over small angles, where the corresponding focal point is located exactly halfway between the reflector and its center of curvature. Uniform plane waves arriving near the center of the aperture are concentrated at this focal point, whereas those arriving near the perimeter of a spherical reflector are focused at a point somewhat closer to the reflector itself. It is therefore possible to illuminate the entire spherical reflector in phase by using a line feed, as suggested in the figure, that directs rays from its top end, located near the classic focus, downward toward the center of the reflector, while array elements positioned closer to the reflector are phased to transmit or receive radiation more laterally. For the Puerto Rican Aricebo radio telescope, which has a diameter of 1000 feet, the illuminated portion of the reflector for any position of the line feed is approximately 600 feet in diameter. The illuminated portion can be repositioned by physically moving the line feed so that it points in the desired direction. This antenna can scan approximately ±15° in any direction relative to zenith.

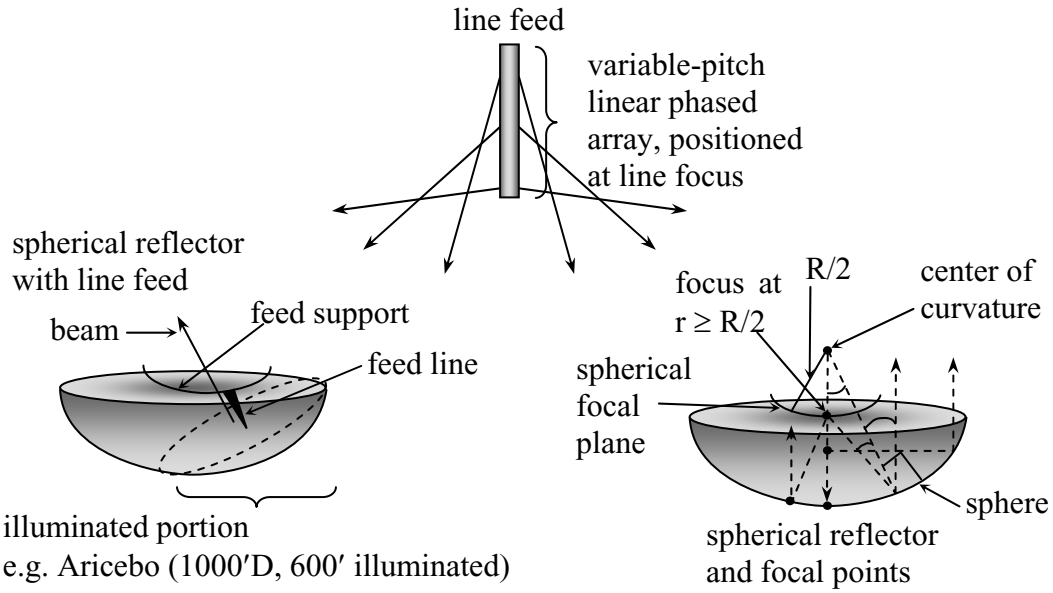

Figure 3.3-8 Spherical reflector antenna with line feed

Sometimes rapid scanning of a stationary large parabolic reflector is desired. One approach is suggested in Figure 3.3-9, where a toroidal parabolic reflector is employed which has a circular focal line along which a feed can be spun at high speed.

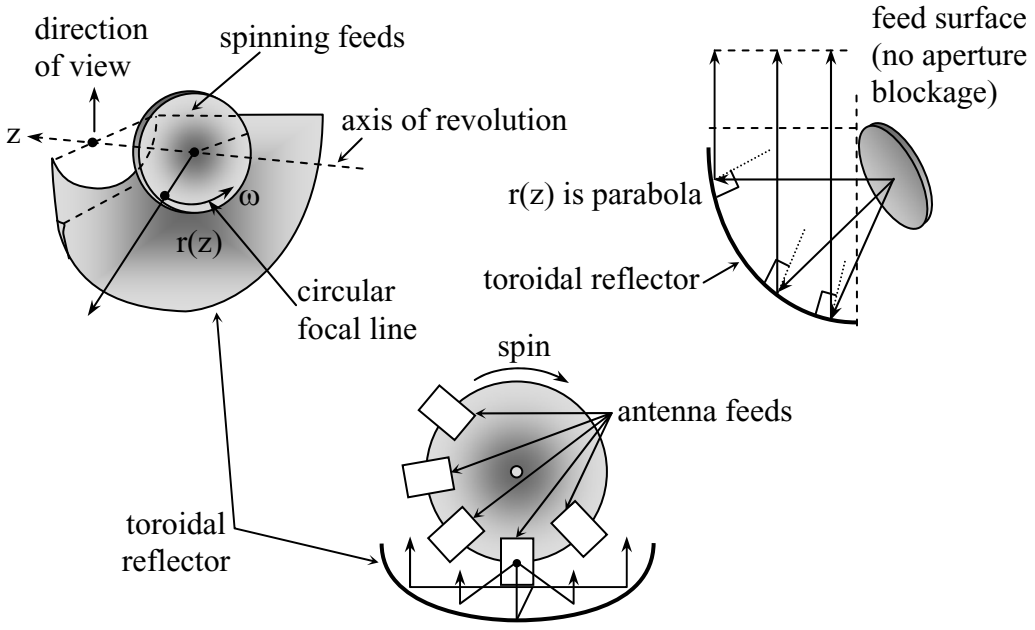

Figure 3.3-9 Scanned parabolic torus reflector antenna

Even when viewing a fixed direction such an antenna does not have a focal point. Instead the focus is distributed over a small aperture that must be phased accordingly using a lens or a specially shaped sub-reflector. A feed illuminating this small aperture and located on the circular locus of focal points can be spun on a rotating table so as to scan the beam in a plane perpendicular to the axis of revolution z. Because such a spinning plate is typically large, many antenna feeds can be positioned around it so that many frequencies can be observed simultaneously. This structure has the additional advantage that there is no aperture blockage.

Parabolic antennas are often fed with a *waveguide horn*, as illustrated in Figure 3.3-10.

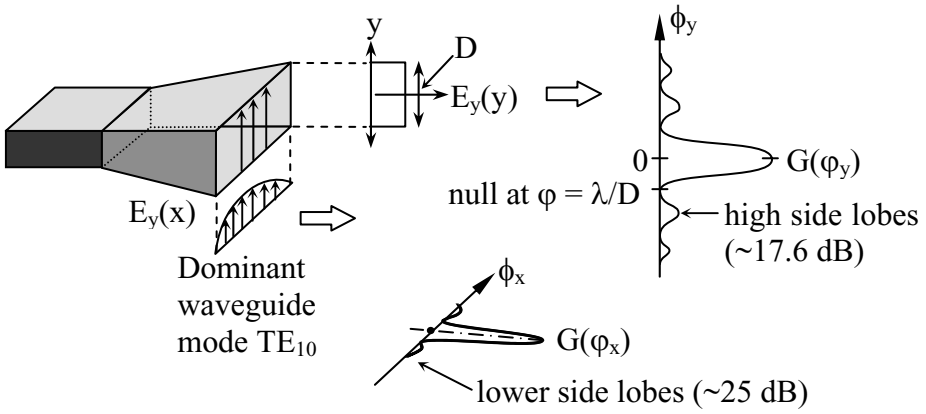

Figure 3.3-10 Pyramidal horn antenna and its pattern

The figure illustrates how a rectangular waveguide propagating the dominant mode can be gently flared so as to preserve the electric field distribution of the dominant mode in the open aperture, which then diffracts as governed by Huygen's principle. If the transition to the flare is not sufficiently gentle, additional modes may be generated, some of which may reflect energy back down the waveguide, introducing loss. Because the electric field strength in the aperture is uniform along the y-axis, and abruptly becomes zero at the top and bottom edges of the horn, as illustrated, high sidelobes are incurred because the fields radiated by a boxcar illumination pattern approximate a sinc function, which is the Fourier transform of a boxcar. The sidelobes in the x-direction are much lower, ~25dB, corresponding to the Fourier transform of a cosine function, which characterizes the electric field distribution in the x direction for the dominant waveguide mode TE_{10}. Carefully designed obstacles can also be placed in the throat of such horns to introduce additional waveguide modes which add coherently so as to produce a more tapered aperture distribution.

An alternate design, the *scalar feed*, has found wide acceptance. In this case the waveguide is coupled to a conical feed horn that has circular grooves cut in the wall. These grooves are ~$\lambda/4$ deep and are spaced approximately $\lambda/4$ apart, as suggested in the side view of Figure 3.3-11. The short circuit at the bottom of each groove is transformed by the $\lambda/4$ section into an open

circuit at the top that precludes current flow radially along the horn surface at that frequency. By precluding such currents, the electric field that otherwise would terminate abruptly at the wall is forced to approach zero there, thus tapering the aperture field distribution in that direction. The result is a circular aperture field distribution that is tapered in both the x and y directions. Because the grooves themselves will produce echoes that return energy to the throat of the horn, they are spaced ~λ/4 apart so that adjacent echoes cancel. In practice, other variations can be used.

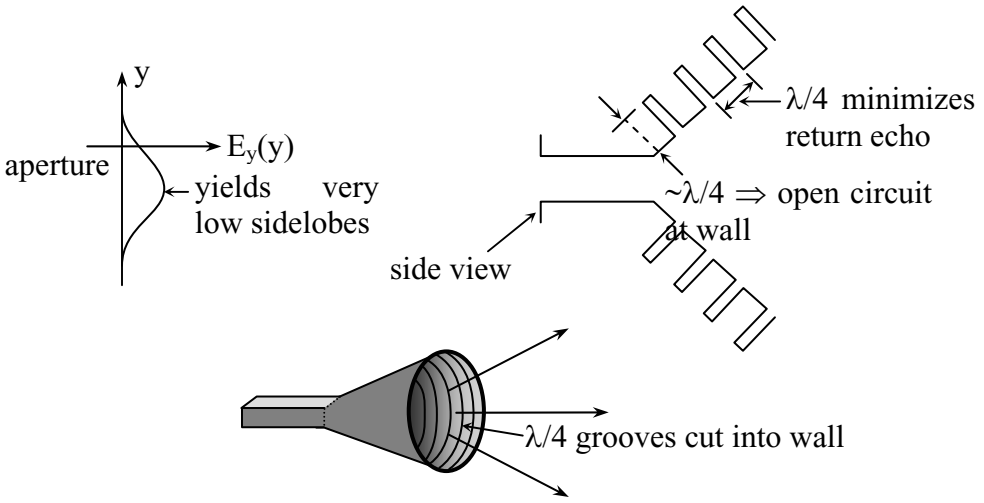

Figure 3.3-11 Scalar feed horn

Practical design considerations for these feed horns preclude spacing them sufficiently close together to achieve the full benefit of multifeed arrays, such as those illustrated in Figure 3.3-7. The problem is suggested in Figure 3.3-12 where it is clear that the aperture excitation functions for feeds A and B do not overlap, and therefore their Fourier transforms must also be orthogonal, implying a crossover point below 3 dB between the two far-field patterns A and B.

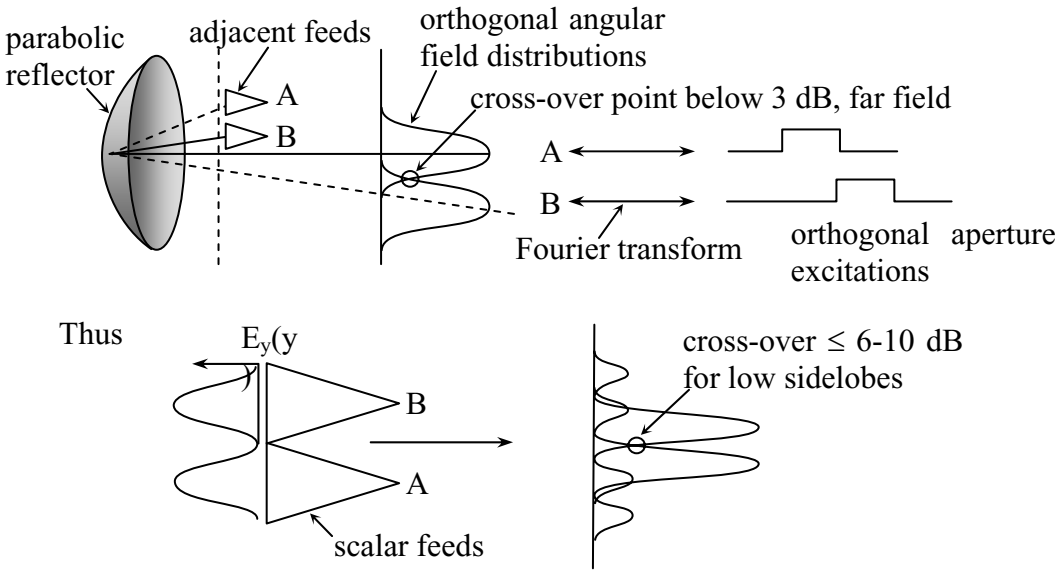

Figure 3.3-12 Multiple-horn antenna feeds

That is, the Fourier transforms of two orthogonal functions must also be orthogonal. Scalar feeds pose an even more serious problem because their aperture excitation functions are not only orthogonal, but they also approach zero sooner so that the cross-over point between the associated far-field antenna beams can be diminished 6-10 dB or more. Solutions to this problem are suggested in Figure 3.3-13. For example, the four-array solution uses four close-packed feed arrays (A, B, C, and D) and their beams are interleaved.

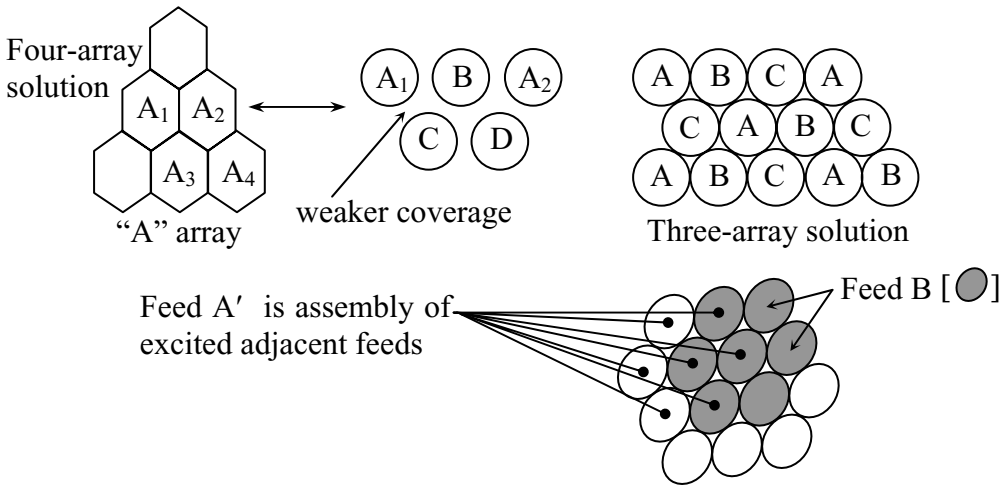

Figure 3.3-13 Feed-horn arrays

One approach to reducing nulls between adjacent beams is to drive a cluster of antenna feed horns in phase so as to synthesize a larger equivalent feed A'. Each individual feed-horn might then be coupled to more than one output waveguide, and the larger synthesized feeds, e.g. A' and B', would overlap so as to produce cross-over levels substantially better than 3 dB. Alternatively, different feed arrays can be focused on the same region in space so that one set of antenna patterns fills in the voids of the other. Figure 3.3-13 suggests both a four-array solution and a three-array solution which produce such interdigitated patterns with no deep unobserved nulls.

Sometimes antenna apertures are directed toward one another at distances too close for the antenna link equation (3.1.9) to be valid. The problem is suggested in Figure 3.3-14, which illustrates how the uniform-phase-front model conflicts with the radial emission model at a distance r such that $r(\lambda/D) \cong D$. Only for distances beyond $r = D^2/\lambda$ does the radiated phase front begin to become spherical. For most purposes the far field of an aperture antenna is defined as occurring at distances r such that:

$$r \gtrsim \frac{2D^2}{\lambda} \tag{3.3.18}$$

Beyond this distance the far-field antenna pattern rapidly becomes independent of distance r.

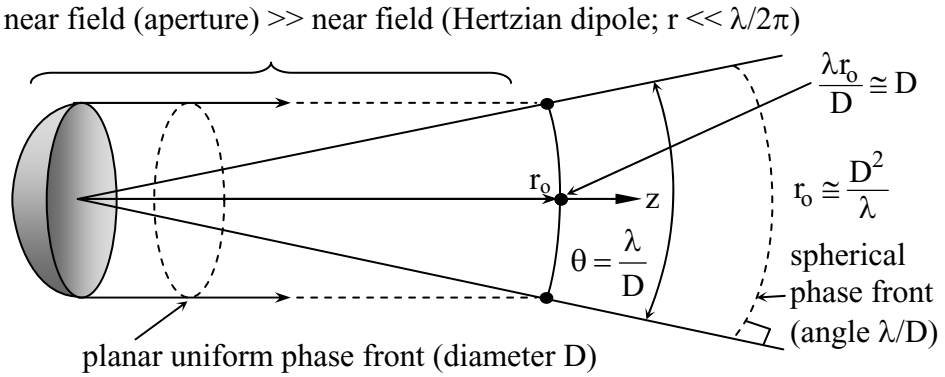

Figure 3.3-14 Far-field of an aperture antenna

Caution must be used when calculating the power transfer between two antennas located within the far field distance. Consider the example in Figure 3.3-15 where an antenna of aperture A_o transmits toward a similar aperture of area $A_o/3$ with a uniform phase front so that two-thirds of the power misses the second antenna and propagates past it.

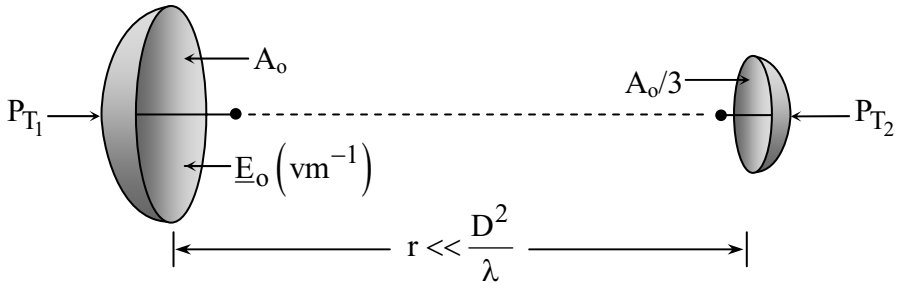

Figure 3.3-15 Parabolic antennas coupled in their near fields

The second antenna clearly intercepts one-third of the power that is transmitted. By reciprocity, it follows that the first antenna would receive only one-third the power transmitted by the second, which conflicts with a naive photon model that suggests that all of the photons from the smaller antenna would fall within the aperture A_o and be received. This paradox can be easily resolved.

Consider a horn antenna that is uniformly illuminated when it is transmitting, as illustrated in Figure 3.3-16.

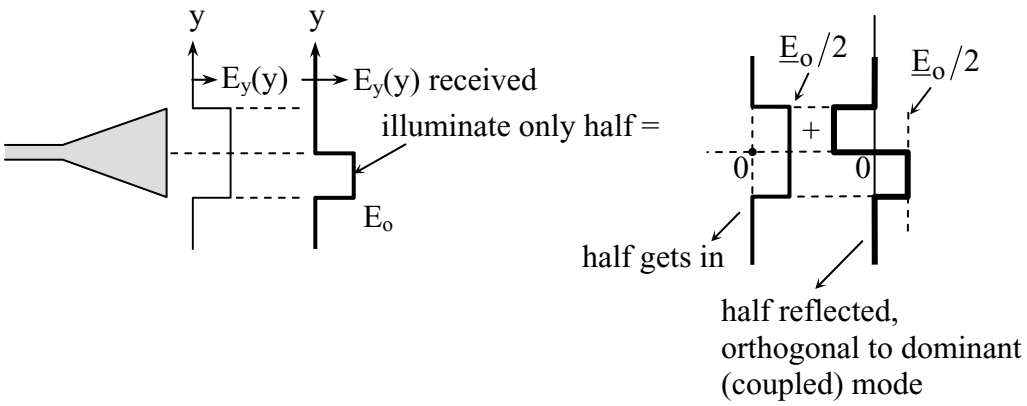

Figure 3.3-16 Mode orthogonality and near-field antenna coupling

If this horn is externally illuminated with an electric field strength E_o only over its lower half, as illustrated, then we can represent the aperture excitation as the sum of a uniform field of strength $\underline{E}_o/2$ plus a second component that resembles a square wave of amplitude $\underline{E}_o/2$. The uniform component is perfectly received, and conveys half the power in the incident wave because its flux density is one-quarter that of the original wave $(\underline{E}_o/2)^2$, but over twice the truncated aperture area. Thus a uniform wave incident upon only a portion α of a uniformly illuminated aperture has a coupling coefficient equal to α. Thus the principle of reciprocity introduced to explain the coupling in Figure 3.3-15 also applies to Figure 3.3-16.

Both antennas and uniform plane waves in space can be characterized by their polarization. If the radiation is purely monochromatic, any electromagnetic field is partially characterized by the elliptical path traced by either the electric or magnetic field vectors. This polarization ellipse can be characterized as suggested in Figure 3.3-17, where the ellipse is defined by the trajectory of the tip of the electric vector $\overline{E}(t)$.

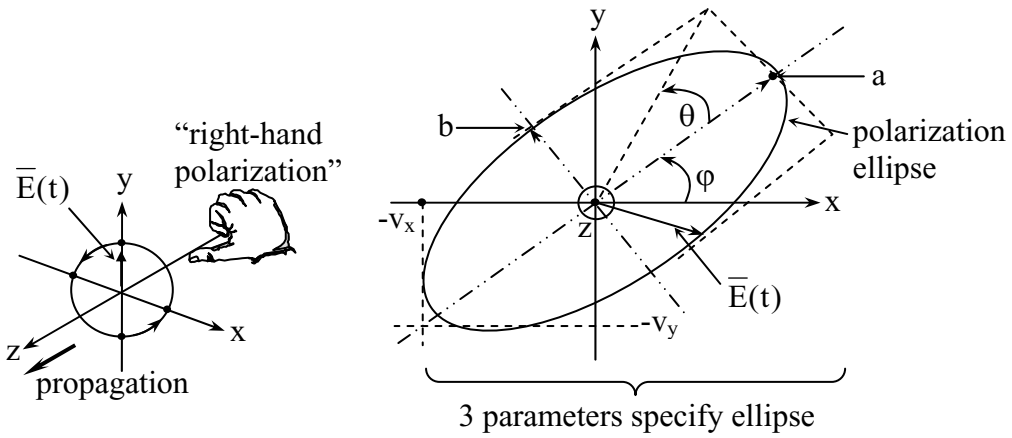

Figure 3.3-17 Polarization ellipse for monochromatic radiation

For a uniform plane wave propagating in the ± z-direction the polarization ellipse must be in the x-y plane. For a general electromagnetic field, monochromatic radiation must necessarily produce an ellipse in some plane, which we define as the x-y plane. The ellipse can be specified by various groups of three parameters, such as (a,b,ϕ), (a,ϕ,θ), (v_x, v_y, θ), or others; we also need to specify whether the ellipse is being traced in the right-hand or left-hand direction. The parameters (a,b) correspond to the major and minor axes of the polarization ellipse, respectively, while (v_x, v_y) correspond to the maximum amplitudes of the electric vector along the x- and y-axes, respectively. The angles ϕ and θ refer to the tilt of the ellipse and its eccentricity, as suggested in the figure. For a uniform plane wave propagating the + z-direction the electrical engineering community defines right-hand polarization as suggested in Figure 3.3-17, where the thumb of the right hand points in the direction of propagation and the fingers point in the direction of electrical vector rotation. The physics community tends to use the opposite definition, although the engineering community has embedded their definition in standards with legal force.

Most useful signals have finite bandwidth, and so the polarization description requires a fourth parameter for completeness. A narrow-band signal conveyed by a uniform plane wave characterized by $\overline{E}(t)$ at a given point in space can alternatively be characterized by independent amplitudes and phases along each of the x- and y-axes:

$$\bar{E}(t) = \hat{x}\ v_x(t)\ \cos\ (\omega t+\phi(t)) + \hat{y}\ v_y(t)\ \cos\ (\omega t+\phi(t)+\delta(t)) \tag{3.3.19}$$

For narrowband signals the electric field amplitudes along the x- and y-axes, $v_x(t)$ and $v_y(t)$, are slowly varying and random, and the mean values of δ, v_x, and v_y may be non-zero.

One of the earliest and most widely used methods for characterizing polarization is the set of *Stokes' parameters*. These four scalar parameters roughly represent the total power in the wave, the degree to which it is x-polarized, the degree to which it is polarized at +45°, and the degree to which it is circularly polarized, as defined below:

$$I \equiv S_o \equiv \left[\left\langle v_x^2(t)\right\rangle + \left\langle v_y^2(t)\right\rangle\right] \Big/ 2\eta_o \tag{3.3.20}$$

$$Q \equiv S_1 \equiv \left[\left\langle v_x^2(t)\right\rangle - \left\langle v_y^2(t)\right\rangle\right] \Big/ 2\eta_o \tag{3.3.21}$$

$$U \equiv S_2 \equiv 2\left\langle v_x(t) \bullet v_y(t)\ \cos\delta(t)\right\rangle \Big/ 2\eta_o \tag{3.3.22}$$

$$V \equiv S_3 \equiv 2\left\langle v_x(t) \bullet v_y(t)\ \sin\delta(t)\right\rangle \Big/ 2\eta_o \tag{3.3.23}$$

If such a narrowband wave is perfectly polarized, then both $\delta(t)$ and the ratio $v_x(t)/v_y(t)$ are constants so that the ellipse has a fixed shape and a slowly varying size. In this case it is easy to show using equations (3.3.20-23) that:

$$S_0^2 = S_1^2 + S_2^2 + S_3^2 \tag{3.3.24}$$

Therefore any three Stokes' parameters specify the polarization of a perfectly polarized wave. In contrast, for a completely unpolarized wave it is clear from this definition that $S_1 = S_2 = S_3 = 0$. Moreover, if two uncorrelated uniform plane waves are superimposed, the Stokes' parameters for the combination are the sums of the Stokes' parameters associated with each wave separately.

In general, narrowband electromagnetic waves tend to be partially polarized so that all four Stokes' parameters are non-zero. Using the fact that uncorrelated waves permit superposition of Stokes' parameters, we can represent an arbitrary wave as a superposition of an unpolarized wave of intensity S_u with one that is perfectly polarized:

$$[S_0, S_1, S_2, S_3] = [S_u, 0, 0, 0] + [S_0 - S_u, S_1, S_2, S_3] \tag{3.3.25}$$

where:

$$(S_0 - S_u)^2 = S_1^2 + S_2^2 + S_3^2 \tag{3.3.26}$$

The unpolarized power S_u can be found by measuring the four Stokes' parameters and then using (3.3.26). The fractional polarization is defined as $(S_0 - S_u)/S_0$.

Another commonly used characterization of polarization is the *coherency matrix* $\underline{\underline{J}}$, where:

$$\underline{\underline{J}} \triangleq \frac{1}{\eta_o}\begin{bmatrix} \langle \underline{E}_x \underline{E}_x^* \rangle & \langle \underline{E}_x \underline{E}_y^* \rangle \\ \langle \underline{E}_x^* \underline{E}_y \rangle & \langle \underline{E}_y \underline{E}_y^* \rangle \end{bmatrix} \tag{3.3.27}$$

where:

$$\overline{E}(t) = \hat{x}\ R_e\left\{\underline{E}_x(t)e^{j\omega t}\right\} + \hat{y}\ R_e\left\{\underline{E}_y(t)e^{j\omega t}\right\} \tag{3.3.28}$$

and where the fields are narrowband because $\overline{E}(t)$ varies slowly. It is easily seen from this definition that the coherency matrices for x-polarized, right-circular, and unpolarized radiation are:

$$\underline{\underline{J}}_x = 2\ S_0\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{3.3.29}$$

$$\underline{\underline{J}}_{RC} = S_0\begin{bmatrix} 1 & -j \\ j & 1 \end{bmatrix} \tag{3.3.30}$$

$$\underline{\underline{J}}_u = S_0\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.3.31}$$

Coherency matrices for two superimposed co-propagating plane waves superimpose, just as do Stokes' parameters. As a result, when the coherency matrices for two equal-strength orthogonally polarized waves are superimposed, the result is twice the coherency matrix for an unpolarized wave. That is,

$$\underline{\underline{J}}_x + \underline{\underline{J}}_y = \underline{\underline{J}}_{RC} + \underline{\underline{J}}_{LC} = \underline{\underline{J}}_A + \underline{\underline{J}}_B = 2\underline{\underline{J}}_u \tag{3.3.32}$$

where A and B are any two arbitrary orthogonal polarizations and the coherency matrices have equal traces. Thus, given any arbitrary coherency matrix $\underline{\underline{J}}_A$, we can readily find the coherency matrix for the orthogonal polarization using (3.3.32).

In general the polarization transmitted by an arbitrary antenna is different in each direction θ,ϕ, although antennas are usually designed to exhibit pure and uniform polarization characteristics over as wide a solid angle as possible. We may define the polarization characteristics of an antenna in terms of the far fields they would radiate in two chosen orthogonal polarizations $\underline{E}_i$ and $\underline{E}_j$:

$$\frac{G_{ij}(\theta,\phi)}{G(\theta,\phi)} = \frac{A_{ij}(\theta,\phi)}{A(\theta,\phi)} = \frac{\underline{E}_i \underline{E}_j^*}{\underline{E}_i \underline{E}_i^* + \underline{E}_j \underline{E}_j^*} \tag{3.3.33}$$

where the orthogonal polarizations $\{i,j\}$ could be $\{x,y\}$, $\{r,\ell\}$, or any $\{a,b\}$ where $b \perp a$. We claim without proof that the total power received by any antenna can be simply represented as the trace of the matrix product between the effective area matrix $\underline{\underline{A}}$ and the coherency matrix $\underline{\underline{J}}_{inc}$ of the incident wave:

$$P_{rec} = \frac{1}{2} T_r \left[\underline{\underline{A}} \, \underline{\underline{J}}_{inc}^{t} \right] [\text{Watts}] \tag{3.3.34}$$

$$P_{rec} = \frac{1}{2} \left[A_{11}J_{11} + A_{12}J_{12} + A_{21}J_{21} + A_{22}J_{22} \right] \tag{3.3.35}$$

In most cases only the $A_{11}J_{11}$ term is significant. Since antennas receive uncorrelated signals from many directions, the total power received is represented as an integral:

$$P_{rec} = \frac{1}{2} \int_{4\pi} T_r \left[\underline{\underline{A}}(\theta,\phi) \underline{\underline{J}}^{t}(\theta,\phi) \right] d\Omega \tag{3.3.36}$$

One way to measure the unknown polarization of an incoming wave is to use four antennas making four scalar power measurements M_a, M_b, M_c, M_d where:

$$\begin{bmatrix} M_a \\ M_b \\ M_c \\ M_d \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \underline{A}_{11_a} & \underline{A}_{12_a} & \underline{A}_{21_a} & \underline{A}_{22_a} \\ \underline{A}_{11_b} & \underline{A}_{12_b} & \bullet & \bullet \\ \underline{A}_{11c} & \bullet & \bullet & \bullet \\ \underline{A}_{11d} & \bullet & \bullet & \underline{A}_{22_d} \end{bmatrix} \begin{bmatrix} \underline{J}_{11} \\ \underline{J}_{12} \\ \underline{J}_{21} \\ \underline{J}_{22} \end{bmatrix} \tag{3.3.37}$$

in accord with (3.3.34). If the 4 x 4 matrix $\underline{\underline{A}}_4$ is non-singular then we can solve (3.3.37) for the unknown coherency matrix:

$$\underline{\hat{J}} = 2\underline{\underline{A}}_4^{-1}\underline{M} \tag{3.3.38}$$

To determine whether a set of four polarization measurements is complete, we may determine if $\underline{\underline{A}}_4$ is singular. Consider the case where we measure the powers in x, y, right-circular, and left-circular polarization, for which:

$$\underline{\underline{A}}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -j & j & 1 \\ 1 & j & -j & 1 \end{bmatrix} \tag{3.3.39}$$

It is easily seen that the determinant of this matrix (3.3.39) is zero, which we might suspect because it cannot distinguish between two otherwise identical elliptical polarizations, one of which is tilted to the right and one to the left. In contrast, if we measure x, 45°, right-circular, and left-circular polarization, we obtain:

$$\underline{\underline{A}}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & -j & j & 1 \\ 1 & j & -j & 1 \end{bmatrix} \tag{3.3.40}$$

for which the determinant is non-zero and all polarizations can be distinguished and measured correctly.

A wide variety of polarization measurement systems and circuits can be employed, one of which was used by Cohen (*Proc. IRE*, 1, 1958), to measure the instantaneous and rapidly changing polarization of solar radio bursts.

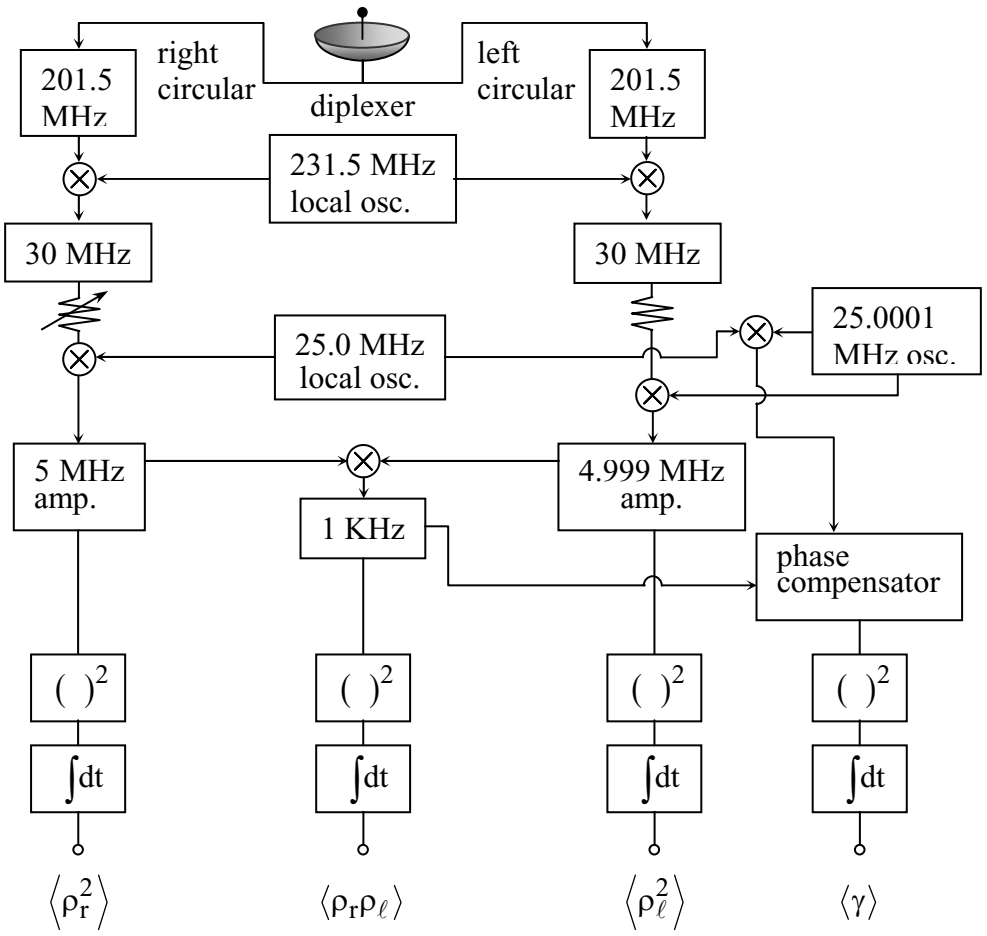

Figure 3.3-18 Polarimeter circuit

This system extracts the right- and left-circular polarizations from a single antenna feed and beats them down to 30-MHz i.f. frequencies, where their powers are balanced by use of a potentiometer on one arm. They are down-converted a second time at frequencies spaced 1-kHz apart, where the powers of the right- and left-circular components are measured separately. The two signals are also mixed together and filtered at 1 kHz to extract the correlation between right and left $\langle \rho_r \rho_\ell \rangle$. In addition, the phase difference between the product of the right and left circular components and the phase of the 1-kHz local oscillator yields the estimated average phase difference $\langle \gamma \rangle$ between the left- and right-circular components. These four measurements are uniquely related to the four Stokes' parameters and the coherency-matrix description.

As signals propagate from a transmitter toward a receiver their polarization and intensity distribution can be disturbed by phase errors introduced along the propagation path. Some such phase differences are large scale and systematic, such as those induced in waves propagating parallel to the terrestrial surface in an atmosphere where the humidity and temperature, and therefore permitivity, vary with height, thus gently bending the direction of propagation. Waves

generally bend toward layers of slower phase velocity. Similar phenomena occur in the ionosphere. However, in both the atmosphere and ionosphere there are often sufficient inhomogeneities that random phase and amplitude gradients are introduced across the phase front. If these rms phase variations are much less than $\lambda/2\pi$, the fluctuations are weak and can be characterized as discussed later. If the variations are large compared to $\lambda/2\pi$, these strong fluctuations can introduce some degree of cancellation between rays arriving at a given point with differing phases, and therefore produce amplitude variations across the phase front. If this scattering process results in a phase front that is non-uniform but exhibits constant amplitude across the region of interest, a thin-screen model for phase variation is often employed.

Our expressions for the power received by antennas assume that the rays arriving from different directions are uncorrelated, but this may not be true if multipath is present, in which case the intensity distribution across the aperture is non-uniform. In this case both the phase and amplitude receiving characteristics for an antenna as a function of angle must be considered when calculating the integrated strength of the electric field at the receiver itself; that is, we must integrate arriving complex electric field vectors (phasors) over all angles, weighted by the complex antenna receiving characteristics, before integrating and squaring, rather than integrating the received power over all angles as we have done previously.

Most receiving antennas introduce phase errors of their own because of small mechanical errors in the structure defining the phase front of the antenna. These may arise from systematic design and fabrication errors, from the transient effects of gravity, wind, and thermal gradients, and from systematic offsets in an antenna feed. Antenna errors may result from machining tolerances, natural surface roughness, errors in antenna structural adjustment, and the presence of boltheads or uneven wire mesh.

An example of these problems is the history of the original 300-ft parabolic antenna at the National Radioastronomy Observatory in Greenbank, West Virginia. This very low cost antenna was constructed of wire mesh over a lightweight metal frame that sagged slightly in a systematic way. After many years of use this frame backup structure was stiffened so that the antenna could be operated at shorter wavelengths. Unfortunately workmen left footprints on the wire mesh in the form of small indentations, producing phase errors that introduced new sidelobes to the antenna pattern. To remove the effects of the footprints the wire mesh was removed from the antenna and steamrolled on a flat concrete surface. When reinstalled, the previous broad footprint-based sidelobes had disappeared, only to be replaced by large narrow ones because the steamrolled mesh had expanded in two dimensions, producing long waves in the surface. The ultimate solution was to install new panels with superior tolerances. To better understand how phase errors of different size and shape affect antenna patterns, it is useful to address the problem analytically.

Consider the aperture illustrated in Figure 3.3-19, for which we wish to compute the effects of small aperture phase errors on the antenna pattern in directions close to the z-axis where

$\phi_x, \phi_y << 1$. The relationship between the aperture excitation function $\underline{E}_x(x,y)$ and the antenna directivity $D(\bar{\varphi})$ can be simply represented:

$$\begin{array}{ccc} \underline{E}_x(x,y) & \tilde{\leftrightarrow} & \underline{E}_x(\varphi_x,\varphi_y) \\ \downarrow & & \downarrow \\ R_{\underline{E}_x}(\bar{\tau}) & \tilde{\leftrightarrow} & \left|\underline{E}_x(\bar{\varphi})\right|^2 \propto D(\bar{\varphi}),\ G(\bar{\varphi}) \end{array} \tag{3.3.41}$$

following (3.3.6-7). To simplify notation we represent the angles $\bar{\varphi}$ and displacements $\bar{\tau}$ as 2-dimensional vectors rather than writing the two components explicitly.

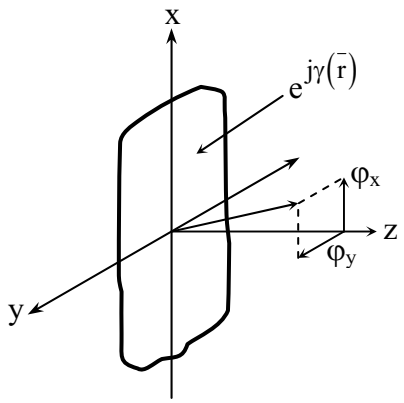

Figure 3.3-19 Aperture antenna with small phase errors $\gamma(\bar{r})$

If we compute the directivity using the approximate Fourier transform of the aperture autocorrelation function $R_{\underline{E}_x}(\bar{\tau})$, we obtain:

$$D(f,\bar{\varphi}) = \left[\pi(1+\cos\ \theta)^2 / \lambda^2\right] \bullet \frac{\int_A \underline{R}_{E_x}(\bar{\tau}) e^{-j\frac{2\pi}{\lambda}(\bar{\varphi}\bullet\bar{\tau})} d\tau_x d\tau_y}{\int_A |\underline{E}_x(x,y)|^2\, dxdy} \tag{3.3.42}$$

which has the expected value over an ensemble of such antennas:

$$E\{D(f,\bar{\varphi})\} = \frac{\pi(1+\cos\ \theta)^2}{\lambda^2 \int_A |\underline{E}_x(x,y)|^2\, dxdy} \bullet \int_A E\{R_{\underline{E}_x}(\bar{\tau})\} e^{-j\frac{2\pi}{\lambda}(\bar{\varphi}\bullet\bar{\tau})} d\tau_x d\tau_y \tag{3.3.43}$$

where:

$$E\left\{R_{\underline{E}_x}\left(\bar{\tau}\right)\right\} = R_{\underline{E}_o}\left(\bar{\tau}\right) E\left\{e^{j\gamma\left(\bar{r}\right) - j\gamma\left(\bar{r}-\bar{\tau}\right)}\right\} \tag{3.3.44}$$

$R_{\underline{E}_o}\left(\bar{\tau}\right)$ is the autocorrelation function of the aperture field distribution in the absence of phase errors and:

$$\underline{E}_x\left(\bar{r}\right) = E_o\left(\bar{r}\right) e^{j\gamma\left(\bar{r}\right)} \tag{3.3.45}$$

where $\gamma\left(\bar{r}\right)$ is the random phase at position $\bar{r}$. Under the assumption of spatial stationarity in the phase errors, it follows that:

$$E\left\{e^{j\gamma\left(\bar{r}\right) - j\gamma\left(\bar{r}-\bar{\tau}\right)}\right\} = E\left\{e^{j\gamma(o) - j\gamma\left(\bar{\tau}\right)}\right\} \tag{3.3.46}$$

To evaluate (3.3.46) it is useful to recall the definition of *characteristic function*, which is the Fourier transform of a probability distribution p(x); it is also called the *moment-generating function*. For example, the characteristic function of p(x) is:

$$E\left[e^{j\omega x}\right] \equiv \int_{-\infty}^{\infty} p(x) e^{j\omega x} dx \triangleq \Gamma(\omega; x) \tag{3.3.47}$$

Characteristic functions are useful, for example, when we seek the probability distribution of the sum of a series of random variables x_i. In this case the probability distribution of the sum is the convolution of the probability distribution for each variable separately, assuming they are uncorrelated. That is,

$$p\left\{\sum_i x_i\right\} = p\left(x_1\right) * p\left(x_2\right) * \ldots * p\left(x_n\right) = F\left\{\prod_{i=1}^{n} F\left[p\left(x_i\right)\right]\right\} \tag{3.3.48}$$

where "F" is the Fourier transform operator. Thus,

$$E\left\{e^{j\omega_1\gamma(0) + j\omega_2\gamma\left(\bar{\tau}\right)}\right\} = \Gamma\left(\omega_1, \omega_2; \gamma(0), \gamma\left(\bar{\tau}\right)\right) \tag{3.3.49}$$

In general, if $\underline{\gamma}_1, \underline{\gamma}_2$ are jointly Gaussian random variables, then:

$$\Gamma\left(\omega_1,\omega_2;\underline{\gamma}_1,\underline{\gamma}_2\right)=e^{-\frac{1}{2}[\omega_1\ \omega_2]\begin{bmatrix}\underline{\gamma}_1\underline{\gamma}_1^* & \underline{\gamma}_1\underline{\gamma}_2^*\\ \underline{\gamma}_2\underline{\gamma}_1^* & \underline{\gamma}_2\underline{\gamma}_2^*\end{bmatrix}\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}} \tag{3.3.50}$$

Thus, to evaluate (3.3.46) we may use (3.3.50), where γ_1 and γ_2 are defined as $\gamma(0)$ and $\gamma\left(\bar{\tau}\right)$, respectively, and ω_1 and ω_2 are defined as 1 and -1, respectively. Therefore,

$$E\left\{e^{j\gamma(0)-j\gamma\left(\bar{\tau}\right)}\right\}=\Gamma\left(\omega_1=1,\ \omega_2=-1;\ \gamma(0),\gamma\left(\bar{\tau}\right)\right) \tag{3.3.51}$$

Evaluating (3.3.50) we see:

$$\Gamma\left(\omega_1=1,\ \omega_2=-1;\ \gamma(0),\gamma\left(\bar{\tau}\right)\right)=e^{-\frac{1}{2}[1\ -1]\begin{bmatrix}\underline{\gamma}(0)\underline{\gamma}^*(0) & \gamma(0)\underline{\gamma}^*\left(\bar{\tau}\right)\\ \underline{\gamma}^*(0)\underline{\gamma}\left(\bar{\tau}\right) & \underline{\gamma}\left(\bar{\tau}\right)\underline{\gamma}^*\left(\bar{\tau}\right)\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix}} \tag{3.3.52}$$

By stationarity $\underline{\gamma}\left(\bar{\tau}\right)\underline{\gamma}^*\left(\bar{\tau}\right)=\underline{\gamma}(0)\underline{\gamma}^*(0)$, and (3.3.52) becomes:

$$\Gamma\left(\omega_1=1,\ \omega_2=-1;\ \gamma(0),\gamma\left(\bar{\tau}\right)\right)=e^{\underline{\gamma}(0)\underline{\gamma}^*\left(\bar{\tau}\right)-\underline{\gamma}(0)\underline{\gamma}^*(0)} \tag{3.3.53}$$

therefore:

$$E\left\{R_{\underline{E}_x}\left(\bar{\tau}\right)\right\}=R_{\underline{E}_o}\left(\bar{\tau}\right)\bullet e^{\underline{\gamma}(0)\underline{\gamma}^*\left(\bar{\tau}\right)-\underline{\gamma}(0)\underline{\gamma}^*(0)} \tag{3.3.54}$$

We may now use (3.3.54) to determine the expected value of the directivity of an antenna with random phase errors:

$$E\left\{D(f,\bar{\varphi})\right\}=\left[\frac{\pi(1+\cos\ \theta)^2}{\lambda^2\int\limits_A\left|E\left(\bar{r}\right)\right|^2 da}\right]\bullet\int\limits_A\left[e^{\underline{\gamma}(0)\underline{\gamma}^*\left(\bar{\tau}\right)-\underline{\gamma}(0)\underline{\gamma}^*(0)}R_{\underline{E}_o}\left(\bar{\tau}\right)\right]e^{-j\frac{2\pi}{\lambda}\bar{\varphi}\bullet\bar{\tau}}d\tau_x d\tau_y \tag{3.3.55}$$

The nature of the first exponential term in the integral of (3.3.55) is suggested by the sketches in Figure 3.3-20.

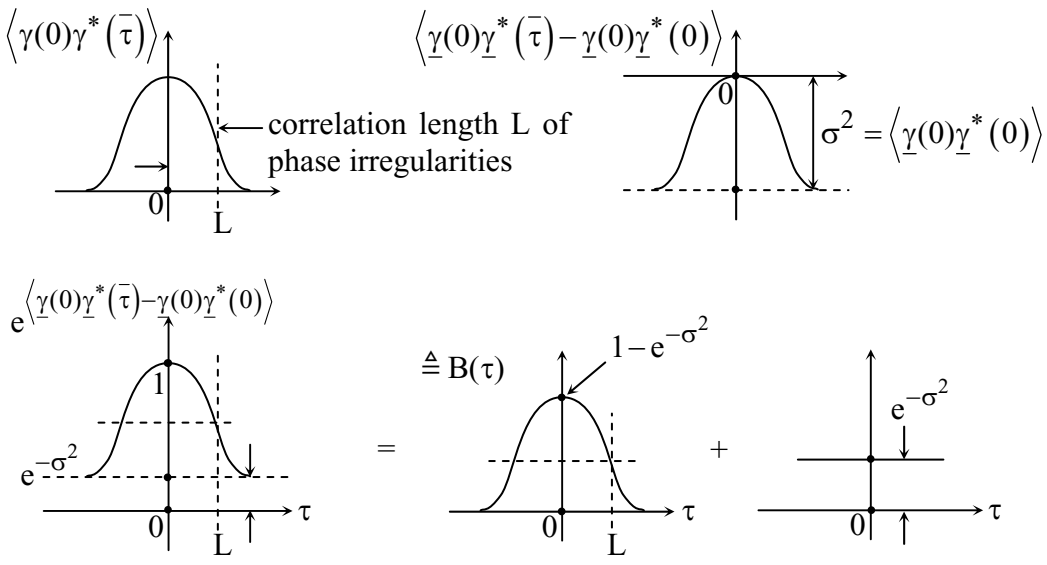

Figure 3.3-20 Functions used to calculate the expected directivity of an aperture with phase errors

The top left-hand curve characterizes the autocorrelation function of the phase errors across the aperture as a function of aperture displacement $\bar{\tau}$, where we assume the nominal correlation length is L. The mean-square phase deviation σ^2 is subtracted to give the exponent desired. Using this exponent yields the curve in the lower left-hand figure, which resembles a Gaussian sitting on a plateau of value $e^{-\sigma^2}$; this can be decomposed into a nominal Gaussian of amplitude $1-e^{-\sigma^2}$ and correlation length L, plus a uniform term $e^{-\sigma^2}$ as suggested in the lower right-hand side of the figure.

Using the results developed in Figure 3.3-20 and (3.3.55) we obtain:

$$E\left\{D(f,\bar{\varphi})\right\} = \left[\frac{\pi(1+\cos\,\theta)^2}{\lambda^2 \int\limits_A \left|\underline{E}(\bar{r})\right|^2 da}\right] \bullet \int_A \left[e^{-\sigma^2} + B(\bar{\tau})\right] \bullet R_{\underline{E}_o(\bar{\tau})} \bullet e^{-j\frac{2\pi}{\lambda}\bar{\varphi}\bullet\bar{\tau}}\, d\tau_x d\tau_y \qquad (3.3.56)$$

$$E\left\{D(f,\theta,\phi)\right\} \cong e^{-\sigma^2} D_o(f,\theta,\phi) + B(\bar{\varphi}) * D_o(f,\theta,\phi) \qquad (3.3.57)$$

The first term in (3.3.57) corresponds to the expected gain degradation due to random phase errors across an ensemble of antennas, and it preserves the original shape of the antenna pattern. The second term corresponds to an increased sidelobe level due to random scattering from the

phase errors, where the angular spread of the new sidelobe pattern is approximately λ/L, as suggested in Figure 3.3-21.

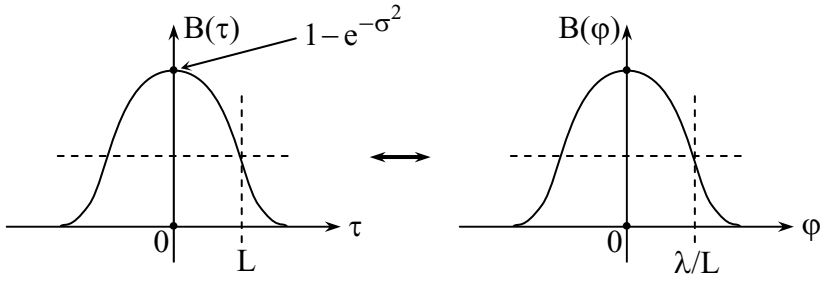

Figure 3.3-21 Transform relationship between the phase-error autocorrelation function and the resulting sidelobes

First consider the gain degradation associated with random phase errors. Let b be the rms surface tolerance of a reflector antenna. The on-axis gain of the average antenna with such phase errors is:

$$G_o' = G_o e^{-\sigma^2} = G_o e^{-(2b \bullet 2\pi/\lambda)^2} = G_o e^{-(b4\pi/\lambda)^2} \tag{3.3.58}$$

which suggests that if b = λ/4π, $G_o^{'} = G_o e^{-1}$, which is a substantial degradation. This degradation diminishes rapidly as b decreases; b = λ/16 yields $0.54\,G_o$, and b = λ/32 yields $0.9\,G_o$. Power no longer directed into the main lobe shifts instead into the new sidelobes, as suggested in Figure 3.3-22.

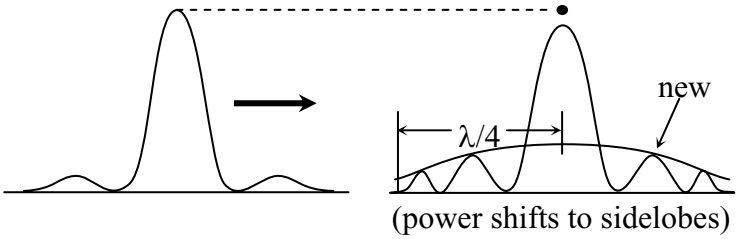

Figure 3.3-22 Power reduction and sidelobe creation effects of phase errors

Since b is fixed for any antenna, the wavelength dependence of gain degradation is defined by (3.3.58), and has the form suggested in Figure 3.3-23.

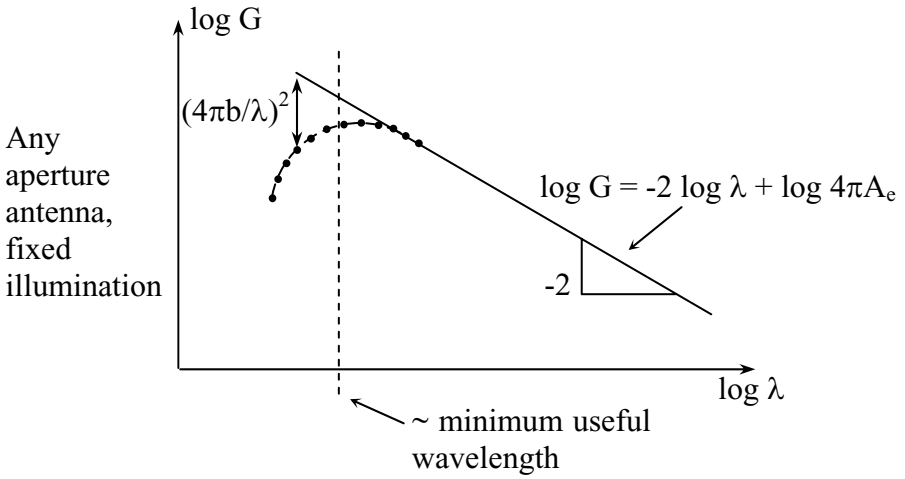

Figure 3.3-23 Standardized gain degradation as a function of wavelength for antennas with phase errors

The slope of the relation between log G and log λ is -2, as shown. The effect of phase errors is to establish a minimum useful wavelength for any antenna, below which the antenna gain G_o' no longer increases. In practice most antennas are not used at such short wavelengths because antenna flexing and other considerations make the antenna pattern less reliable. It is standard practice to evaluate the surface tolerance b of new antennas by measuring their gain as a function of wavelength in the vicinity of the minimum useful wavelength. Examining the growth of sidelobes at these same wavelengths also gives an indication of the correlation length of the phase errors which are responsible for the degradation.

3.4 WAVE PROPAGATION

For most applications of interest the medium through which electromagnetic waves propagate can be characterized by a position-dependent complex dielectric constant or permittivity $\underline{\varepsilon}$ where spatial variations of the real part are associated with refraction and reflection, while the imaginary part is generally associated with loss, which produces absorption and thermal emission.

Most gases are not ionized, and so electromagnetic waves interact principally with the electric or magnetic dipole moments of the gas molecules. When these interactions cause the molecules to spin up or change state, gaining energy, we have absorption, and when the molecules lose rotational or other energy they enhance the strength of the electromagnetic field by emitting radiation. Normally such emission is random and thermal in character, although this emission can be stimulated by incoming radiation producing coherent emission and maser (microwave amplification by stimulated emission of radiation) action.

The two most plentiful gases in the terrestrial atmosphere are nitrogen N_2 and oxygen O_2, both of which are diatomic molecules. Nitrogen in its ground state has no permanent electric or magnetic dipole moment and therefore interacts very little with electromagnetic radiation. Only at very high pressures where the transient induced electric dipole moments become important does N_2 absorb significantly. The oxygen molecule, however, has a magnetic dipole moment which interacts strongly near a band of resonant frequencies stretching 50-70 GHz and near an isolated resonance at 118.31 GHz plus a series of lines at still higher frequencies. In addition, pressure-induced electric dipole moments in oxygen plus the pressure-broadened wings of its magnetic-dipole transitions produce continuum absorption that ranges down to 0 Hz and is approximately proportional to the square of frequency.

The permanent electric dipole moment of H_2O absorbs weakly near its resonance at 22.235 GHz, and quite strongly near its resonance at 183.31 GHz, as well as at a series of resonances at still higher frequencies. Water vapor also has a non-resonant absorption extending to 0 Hz which is approximately proportional to the square of frequency. The shapes of these common molecules are indicated in Figure 3.4-1.

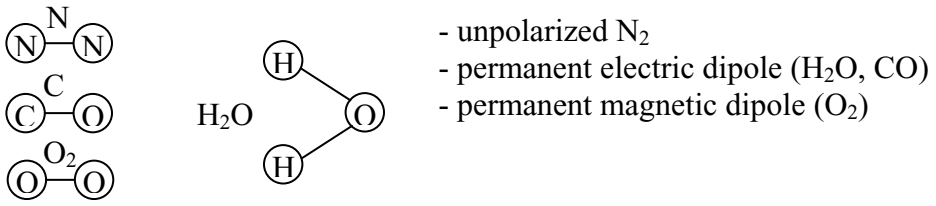

Figure 3.4-1: Molecular configurations of common atmospheric molecules

The absorption spectrum of the terrestrial atmosphere is presented in Figure 3.4-2 in the form of zenith attenuation, from earth to space, as a function of frequency; only major molecule constituents contribute significantly in this frequency band.

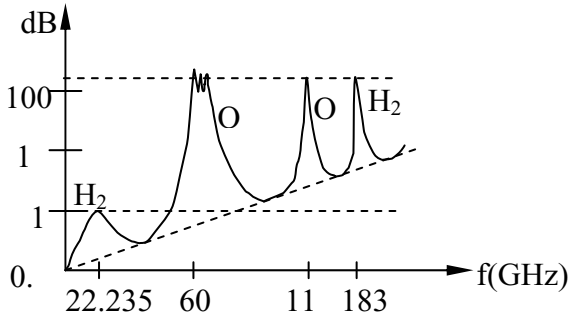

Figure 3.4-2: Atmospheric absorption spectrum of radiowaves at zenith (clear air) same as 20-6

Each of these spectral lines has a resonant frequency f which is associated with the transition of a molecule from state i, in which it had energy E_i, to another molecular state having energy E_j, where the energy difference (joules) is:

$$E_i - E_j = \pm hf \tag{3.4.1}$$

If the molecular energy decreases as a result of the transition $\left(E_j < E_i\right)$, then the excess energy is radiated as a photon at frequency f. The probability of this happening per second is:

$$P_r = A + B\rho_f \tag{3.4.2}$$

where A is called the "Einstein A" coefficient, and corresponds to the probability of spontaneous emission. B is often called the "Einstein B" coefficient and is proportional to the probability of stimulated emission as a result of the ambient radiation intensity ρ_f $\left(\text{joules/meter}^3\right)$. The probability of absorption is the same as the probability of stimulated emission, i.e. $B\rho_f$. A simple ratio links A to B:

$$A/B = 8\pi hf^3/c^3 \tag{3.4.3}$$

The significance of this ratio A/B is that spontaneous emission occurs much more rapidly for higher energy gaps hf, i.e. in proportion to f^3. The implications of this are best understood in terms of the energy states of typical atoms and molecules.

Figure 3.4-3 illustrates the hierarchical nature of energy transitions in typical molecules.

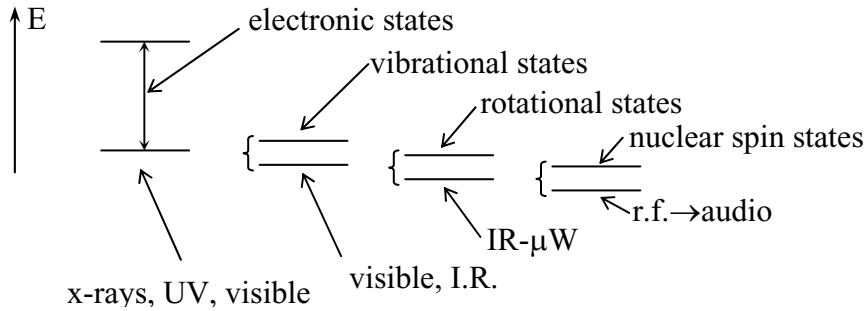

Figure 3.4-3: Typical energy level hierarchy

The greatest energy gaps involve transitions between electronic states, where these gaps may correspond to energies of tens of electron volts or more. Within any electronic state the

molecule may be in a variety of possible vibrational states, between which the energy gaps might be on the order of a fraction of an electron volt to a few electron volts, corresponding to photons in the visible or infrared region. A molecule in any particular electronic state and vibrational state could then be in any of several rotational states, the energy gaps between which correspond to photons in the infrared to microwave region. For any one of these states, there may be any one of many possible nuclear spin states, the energy gaps between which fall in the radio frequency to audio frequency range, as suggested in Figure 3.4-3.

In the radio region the ratio A/B is so small that the lifetime of most molecular states is determined by intermolecular collisions, not spontaneous emission. The effects of such collisions on the rotation of molecules and their associated dipole moments is suggested in Figure 3.4-4.

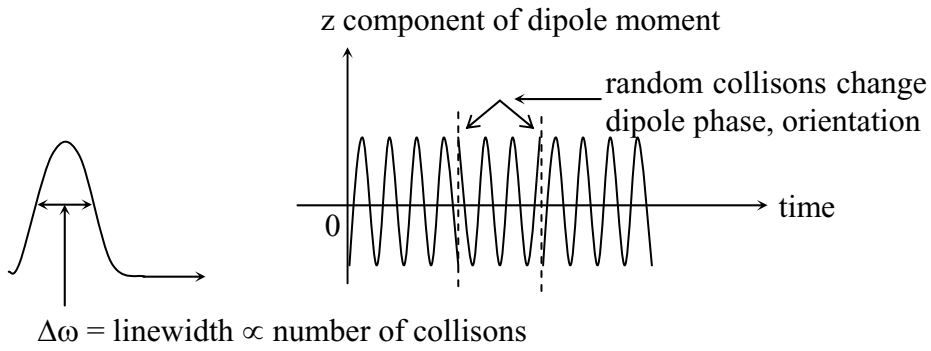

Figure 3.4-4: Effect of molecular collisions on molecular rotation and spectral linewidth

The sinusoidal behavior in Figure 3.4-4 is associated with the regular rotation of a molecule with an electrical dipole moment, and in a classical sense it would radiate at the rotational frequency. However, random (poisson) collisions will interrupt that motion producing a random phase shift, as suggested in the figure at the two events marked by dotted lines. If we now Fourier transform this interrupted sinusoid, we get a Lorentz line shape, resembling the line shape shown in Figure 3.4-4. The linewidth $\Delta\omega$ is proportional to the number of collisions per second, which is proportional to the pressure P if the molecules of interest are a trace gas present in small amounts, and is proportional to P^2 if there is only a single gas present.

Typical linewidths for the atmospheric spectral lines illustrated in Figure 3.4-2 are 2 GHz at standard temperature and pressure for oxygen and water vapor resonances below 1 THz. At low pressures these random collisions become sufficiently rare that other physical mechanisms determine the spectral linewidth. If the gas is warm or turbulant, doppler effects will broaden the line and control the linewidth at low pressures.

If the gas is sufficiently cool and quiet, then the spontaneous emission randomly terminates the sinusoid illustrated in Figure 3.4-4, yielding the *intrinsic linewidth* of ~1/T, where T is the average state lifetime as determined by the Einstein A coefficient. Doppler broadening by the

random thermal motion of molecules can be determined by noting that the translational energy of the molecule is kT/2 in each of three degrees of freedom, or 3kT/2 total, where $3kT/2 = m\overline{v^2}/2$, where v is the random thermal motion of any molecule. Solving this equation for the rms molecular velocity v_{rms} yields the rms doppler shift $f_{D_{rms}} = v_{rms}/\lambda = \left(3kT/m\lambda^2\right)^{0.5}$.

The total absorption at any frequency in a gas is normally the linear superposition of the absorption associated with all the molecular lines in the gas effective at the frequency of interest. In some cases, however, the spectral lines overlapping in some region of the spectrum may be physically linked. An important example of this is provided by the oxygen molecule in the lower terrestrial atmosphere where each of the 30 most important spectral lines are associated with different rotational states of the same O_2 molecule. Classical analysis suggests a phenomena which is observed experimentally, namely that two physically coupled overlapping spectral lines tend to have enhanced absorption between their resonant frequencies and diminished absorption outside them, as suggested in Figure 3.4-5. Coupling is introduced if collisions tend to favor certain molecular state changes over others.

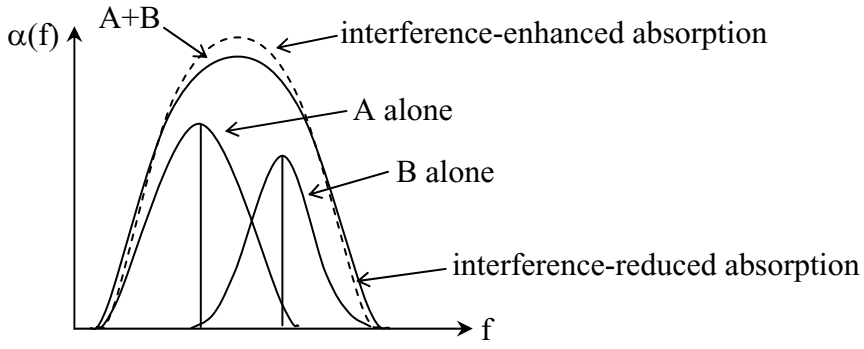

Figure 3.4-5: Absorption spectrum of physically coupled spectral lines exhibiting interference-enhanced and interference–reduced absorption relative to spectral lines A and B in isolation

Developing accurate models predicting the total absorption of physically coupled spectral line complexes remains an active field of research, although numerous molecules such as O_2 have been largely solved and characterized.

Sinusoidal electromagnetic waves propagate along a given z-axis as:

$$\underline{\overline{E}} = \underline{\overline{E}}_o e^{-j\omega\sqrt{\mu\varepsilon}z} \tag{3.4.4}$$

where ε in the atmosphere typically has a real part that controls refraction and an imaginary part that controls loss. The index of refraction n of air at radio frequencies is approximately given by:

$$(n-1)10^6 = (79/T)(p + 4800\,e/T) \tag{3.4.5}$$

where T is °K, and p and e are total and partial water vapor pressures (mb). In the optical region of the spectrum thermal and gravitationally induced gradients in atmospheric density are responsible for most refractive variations, whereas in the radio region variations in tropospheric refraction are controlled primarily by inhomogeneities in the water vapor distribution.

Refraction, which is the redirection of a uniform plane wave at a planar boundary between two media, is characterized by Snell's law:

$$\sin\theta_i / \sin\theta_t = \sqrt{\mu_t \varepsilon_t / \mu_i \varepsilon_i} \tag{3.4.6}$$

where θ_i is the angle of incidence (measured between the surface normal and the direction of propagation) and θ_t is the angle of transmission; ε_t and ε_i are the corresponding permittivities. We assume $\mu_t = \mu_i$ in the atmosphere. Refractive effects are commonly seen in the turbulant images near candle flames, campfires, hot roadways, and wiggling stars. Optical telescopes on mountains typically experience one-arc-sec "seeing" on good nights, where the blur is composed partly of rapid (~10-100 Hz) small lateral displacements and distortions in apparent shape, which may be of the same magnitude. Ground-based telescopes at less favorable sites might typically experience seeing of 2-10 arc-seconds on good nights and the very best mountains at the highest altitudes may approach ~0.4 arc-second seeing. The absolute refraction, which is dominated by the gravitational gradient in atmospheric density, is many times larger and must be compensated, particularly when observing targets near the horizon.

In the radio region refractive effects are often manifest as seasonal variations in the signal strength associated with radio or TV broadcasts, and with the occasional trapping or *ducting* of radio waves over long distances by a layer of moist air near the surface which acts like a waveguide, permitting radiowaves to bounce back and forth between the ocean and the upper boundaries of the moist layer at angles beyond the critical angle associated with total reflection. Ducting can also occur in cold or humid layers of air, or in under-ionized ionospheric layers. Acoustic ducting can occur in cold air layers next to lakes or in cool or salty layers of the ocean.

The atmospheric water vapor scale height is approximately 2 km, which is one-quarter the scale height for atmospheric density, so humidity-based refractive effects are mostly a lower tropospheric phenomenon. They can be quite strong, however, and can displace the apparent radio position of the sun near the horizon by half a solar diameter or more. Thermal inhomogeneities due to turbulance are most prevalent near convective instabilities and in the boundary layer, which is the first few hundred meters of atmosphere or more, depending on surface winds and thermal gradients. Thermal refraction at high altitudes is typically more stable and layered. Thermal layering is particularly evident when radio waves propagate parallel to the surface of the earth over long distances so that rays arriving along different paths might reasonably cancel, producing strong random and rapidly varying fading.

Ionospheric and space plasmas, which typically have both neutral and ionized components, also introduce important refractive effects. For example, the ionosphere has a free electron density n_e of approximately 10^7 to $10^{12}\left(m^{-3}\right)$ between ~50 and ~5000 km altitude, where the maximum free electron density typically falls in the main ionospheric layers 100 – 400 km. The ionosphere plays a particularly important role at radio frequencies below 10 – 30 MHz, which approximates the maximum plasma frequency of the ionosphere, depending on location, season, and solar wind conditions. Radio wave propagation in the ionosphere is principally governed by its permittivity:

$$\varepsilon = \varepsilon_o\left(1-\omega_p^2/\omega^2\right) \tag{3.4.7}$$

where the *plasma frequency* ω_p is:

$$\omega_p = \sqrt{\frac{n_e q^2}{m\varepsilon_o}}\left(rs^{-1}\right) \tag{3.4.8}$$

where m is the reduced mass of the electron, $m = m_e m_i/(m_e + m_i) \cong m_e$. For frequencies above the plasma frequency the phase velocity is:

$$v_p = c\Big/\sqrt{1-\left(\omega_p/\omega\right)^2} > c \tag{3.4.9}$$

and the group velocity is:

$$v_g = c\sqrt{1-\left(\omega_p/\omega\right)^2} < c \tag{3.4.10}$$

Below the plasma frequency radio waves are perfectly reflected from the ionosphere.

The principal effects of the ionosphere on radio wave propagation is to trap radio waves between itself and the surface of the earth so that at frequecies below ~10 – 30 MHz radio links can be established at continental or intercontinental distances. These links, particularly at the highest frequencies, are subject to strong fading. Observing sources above the ionosphere is difficult at frequencies below 10 – 30 MHz, although such observations can occur near ionospheric minima at frequencies as low as ~5 MHz. The ionosphere can also refract waves, causing fading, at frequencies many times higher than the plasma frequency. At frequencies above ~1 – 2 GHz, the principal effect of the ionosphere is to introduce a small amount of Faraday rotation into the polarization of the arriving radio wave.

The plasma frequency of the interplanetary medium is generally below a few kilohertz. Despite these low plasma frequencies, and the even lower plasma frequencies of interstellar space,

interstellar radio waves are nonetheless impacted by refractive effects which can cause frequency-selective fading at frequencies below ~1 GHz. The plasmas generated around spacecraft re-entering the earth's atmosphere can also impede radio communications between the spacecraft and ground for brief intervals.

Plasmas can also absorb and emit radio waves as a result of the transient electric dipoles created as electrons electrostatically ricochet off positive ions. These interactions are proportional to the number of collisions per unit volume, which is proportional to the product of the number n_e of electrons per cubic meter and the number of ions, and therefore is proportional to n_e^2. Although this is not an important effect in the ionosphere, the greater size of the solar corona and photosphere yields larger optical depths and radio brightness temperatures up to $\sim 10^6$ K, not including much brighter emissions associated with coherent radio effects. Such non-thermal radio emissions produced by coherent motions in plasmas can radiate gigawatts in planetary magnetospheres such as that of Jupiter, or substantially greater powers in solar flares.

Terrestrial communication links are also strongly affected by hydrometeors in the form of fog, clouds, rain, sleet, and snow. These interactions are often modeled by the interactions of a plane wave with a dielectric sphere of diameter D and uniform dielectric constant ε. If $D << \lambda\sqrt{\varepsilon_o/\varepsilon}/2\pi$, then we are in the Rayleigh regime where the dielectric sphere has an induced electric dipole that re-radiates some of the incident radiation in the same pattern as would a Hertzian dipole. The scattering cross-section $\sigma_s\left(m^2\right)$ is much smaller than the geometric cross-section in this regime. If D >> λ/2π, then the scattering-cross section approaches the geometric area of the sphere. For frequencies intermediate between the Rayleigh and geometric regimes, Mie scattering characterizes the interaction which typically favors forward scattering accompanied by sidelobes in various other directions.

In the Mie scattering regimes the cross-section oscillates in frequency by a maximum of ~12 dB, as illustrated in Figure 3.4-6.

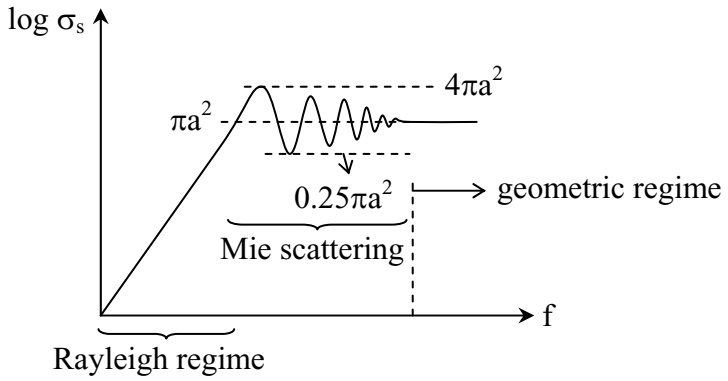

Figure 3.4-6: Scattering cross-section of a nominal dielectric sphere

Because rain-drop diameters statistically vary over an order of magnitude (~0.5-5 mm), the oscillitory character of σ_s is blurred and not evident in radio spectral observations of rain. Non-precipitating cloud-particle diameters are typically less than 50 microns.

In the Rayleigh regime the absorption cross-section of a dielectric sphere is:

$$\sigma_D \propto (D/\lambda)^3 \lambda \tag{3.4.11}$$

so that the absorption coefficient $\alpha(\mathrm{dBm}^{-1})$ is proportional to Mf^2, where M is the mass density $(\mathrm{kg/m}^3)$. The absorption of clouds also has a temperature dependence associated with that of the permittivity, so that cloud absorption in the Rayleigh regime is:

$$\alpha(\mathrm{neperes/m}) \cong \frac{M \times 10^{[0.0122(291-T)-6]}}{\lambda^2} \tag{3.4.12}$$

where $M = \mathrm{kg/m}^3$ liquid water, λ is the wavelength (m), and T is cloud temperature (°K).

The scattering cross-section in the Rayleigh regime:

$$\sigma_s \propto (a/\lambda)^6 \lambda^2 \tag{3.4.14}$$

so that the absorption coefficient associated with scattering (dBm^{-1}) is proportional to V^2f^4, where V is the typical drop volume. That is, if the same cloud mass density is concentrated in small drops it has a much smaller impact on scattering cross-section than if it were reassembled into large drops. Thus precipitation-measuring radars, which respond to scattering cross-section, are disproportionately sensitive to the large-droplet population, which can lead to errors in measurements of m. These errors may be reduced by using multiple wavelengths, one or more of which lie in the Mie or geometric regimes. The total fraction of incoming radiation which is back-scattered by water clouds is only ~10% and reaches a maximum at frequencies near ~100 GHz for precipitation. Most precipitation aloft has, however, elements that are frozen or partially frozen, for which the absorption is much less and the effective scattered power is increased. The total power scattered by ice well above the freezing level of the atmosphere can exceed 65 percent at millimeter-wave frequencies. Attenuation of radio waves by rain becomes important above ~1 GHz, and can exceed 30 dB above 10 GHz in strong precipitation.