Diffraction Theory And Antennas

R. H. CLARK and JOHN BROWN

1980

Contents

Chapter 1

Introduction

Our understanding of the phenomenon of diffraction really begins with Huygens' famous construction, now known as Huygens' principle. This states that each point on a propagating wavefront can be considered as a secondary source radiating a spherical wave. The transverse polarization of the radiated wave was not apparent to Huygens; nor was the principle of wave interference, which had to wait more than a hundred years for Thomas Young to discover it. Augustin Fresnel then combined the ideas of Huygens and Young in order to describe the nature of the light diffracted by various aperture, such as a knife edge, an opaque strip, and a parallel-sided opening. This last example provides a useful introduction to the phenomenon of diffraction and its application to aperture antenna.

1.1 An Elementary Example of Diffraction

Consider a plane electromagnetic wave incident normally on a parallel-sided slit cut in a thin perfectly conducting plane, as in Fig. 1.1. For this two-dimensional situation a typical element of width $\mathrm{d}x$ in the aperture formed by the slit can be supposed, by an obvious modification of Huygens' principle, to radiate a cylindrical wave into the medium to the right of the aperture plane. It is reasonable to assume that the strength of the secondary source is proportional to the magnitude E_0 of the field incident on the aperture from the left, and to the elemental width $\mathrm{d}x$. Then, without bothering with any refinement at this stage, the contribution of the element $E_0\mathrm{d}x$ to the field at a point p a distance r' away is

$$\mathrm{d}E_p = CE_0\mathrm{d}x\frac{1}{\sqrt{r'}}\exp(-jkr'), \tag{1.1}$$

where C is a constant and k is the phase retardation per unit distance suffered by a wave traveling in the medium to the right of the aperture plane at the

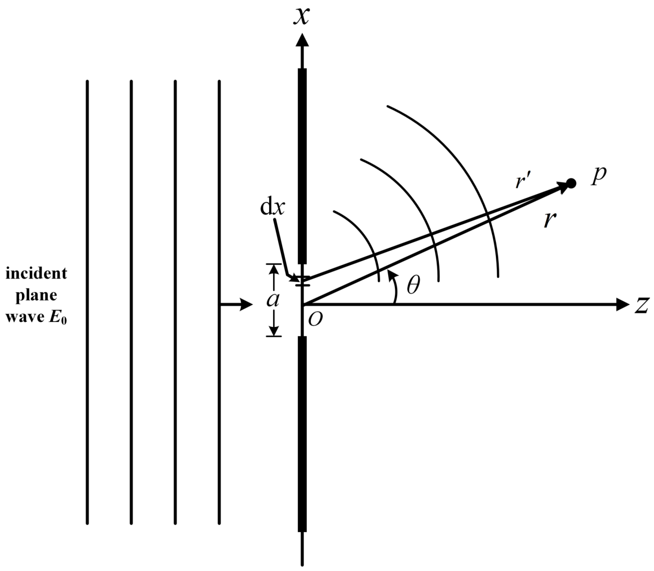

Figure 1.1: Plane wave diffracted by a parallel-sided slit cut in a perfectly conducting plane. Huygens representation.

frequency of the incident field. If the distance r of the point p from the center O of the aperture is very large, in comparison with the aperture width a, it is legitimate to replace r' by r under the square root, and r' by $r - xsin\theta$ in the phase term of equation (1.1). (The direction Op makes and angle θ to the axis normal to the aperture). Then, integrating over the width of the aperture, the total diffracted field at p is

$$E_p = \frac{CE_0}{\sqrt{r}} \exp(-jkr) \int_{-a/2}^{a/2} \exp(jkx \sin\theta) \mathrm{d}x \tag{1.2}$$

so that

$$E_p = \frac{CE_0 a}{\sqrt{r}} \exp(-jkr) \frac{\sin(\dfrac{\pi a \sin\theta}{\lambda})}{(\dfrac{\pi a \sin\theta}{\lambda})} \tag{1.3}$$

Thus Huygens' principle embodied in equation (1.1) gives the field at p as a superposition of cylindrical waves from all parts of the aperture (equation (1.2)) which produces the interference pattern of equation (1.3).

This equation reveals that the field at the distance point p diffracted by the aperture at O is basically a cylindrical waves centered on O, with an angular dependency of the form $(\sin\psi)/\psi$ where $\psi = \pi a \sin\theta/\lambda$. The angular dependence has a characteristic lobe structure, with the main lobe having its maximum in the direction $\theta = 0$, and with zeros at angles given by $\sin\theta = \pm\lambda/a, \pm 2\lambda/a$, and so on.

The uniformly illuminated slit is a useful elementary model of an aperture antenna, which is a class of antennas that is widely used at microwave frequencies for communication and radar surveillance. Thus it is immediately apparent from equation (1.3) that the level of the radiated field depends on the aperture width, increasing as a increases. At the same time the angular width of the main lobe is correspondingly reduced. In antenna parlance (see section 1.4), as the aperture width is increased, the gain of the antenna is increased and its beamwidth is reduced.

1.2 Diffraction Theory By Plane Wave

Developments in diffraction theory have been dominated by the ideas of Huygens and Fresnel. So, although considerable refinements have been achieved (see Stratton (1941), for example) in Kirchioff's scalar theory of diffraction and then in Stratton and Chu's vector theory, the underlying principle of spherical waves radiated by known fields has remained. This is understandable in view of the simplicity and strength of Huygen's original idea. But it has meant that the student of diffraction has been presented with some difficult mathematical hurdles to clear before he can obtain some familiarity with the phenomenon of diffraction.

The difficult seems to arise from the fact that whereas spherical waves are a natural physical entity, they are rather clumsy from a mathematical point of view. In contrast, the plane wave, which is a simple mathematical entity, can never exist as such in the real physical world. However, it has long been known that naturally occurring fields can be *represented* by the superposition of either a discrete set or a continuum of plane waves traveling in different directions. Perhaps the simplest and best know example of this is the field in a rectangular waveguide, which is given precisely by the superposition of two plane waves traveling in directions equally incident to the waveguide axis (see Appendix A.10). In general such a set of plane waves is know as an angular spectrum.

In this book we shall be using the angular plane-wave spectrum concept to develop the theory of diffraction. This approach is not only relatively simple mathematical, as already indicated, but it is also fundamentally more precise than older theories of diffraction. This is because the approximations that must inevitably be made occur at a large stage of the analysis in the case of the plane-wave approach.

One conceptual difficulty concerning the physical plausibility of the angular spectrum method must be dealt with right away. It is that, since plane waves are of infinite lateral extent, it may seem strange at first sight that they can be used to represent realistic fields. But the situation is no different, in principle, from the representation of an arbitrary time waveform $f(t)$ by the superposition of a set of ideal sinusoids. In that case, if the amplitude spectrum as a function of frequency ω is $F(\omega)$ then the waveform is given by the Fourier integral,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\exp(j\omega t)d\omega. \tag{1.4}$$

The individual sinusoids in the integrand of this equation are of infinite duration in time, whereas the waveform $f(t)$ that they represent is usually of finite duration. An immediate advantage of the Fourier integral of equation (1.4) is that it can be inverted to give

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\exp(-j\omega t)dt. \tag{1.5}$$

A similar advantage occurs in using the angular plane-wave spectrum to represent radiated fields, as the following outline analysis shows.

Suppose, instead of the cylindrical waves of Fig. 1.1, the field diffracted by a slit is represented as the superposition of a set of plane waves, of which a typical member is that shown in Fig. 1.2. The angle θ is now a variable describing the directions of the component plane waves constituting the angular spectrum. For technical reasons direction will be specified by $\sin\theta$ rather than θ. Then if the set of plane waves is described by the spectrum function $F(\sin\theta)$, the contribution of this single plane wave of elementary amplitude $F(\sin\theta)\mathrm{d}(\sin\theta)$ to the field at some point p is

$$\mathrm{d}E_p(x,z) = F(\sin\theta)\mathrm{d}(\sin\theta)\exp\{-jk(x\sin\theta + z\cos\theta)\} \tag{1.6}$$

The complete field at p is then obtained by integration as

$$E_p(x,z) = \int_{-\infty}^{\infty} F(\sin\theta)\exp\{-jk(x\sin\theta + z\cos\theta)\}\mathrm{d}(\sin\theta) \tag{1.7}$$

(For present purposes it may be assumed that the range of integration is artificially extended beyond its natural limits of ± 1 for analytical convenience. However, it will be shown in Chapter 2 that waves traveling in directions such that $|\sin\theta| > 1$ are far from artificial). Now, defining the field over the aperture as

$$f(x) = E_p(x,0), \tag{1.8}$$

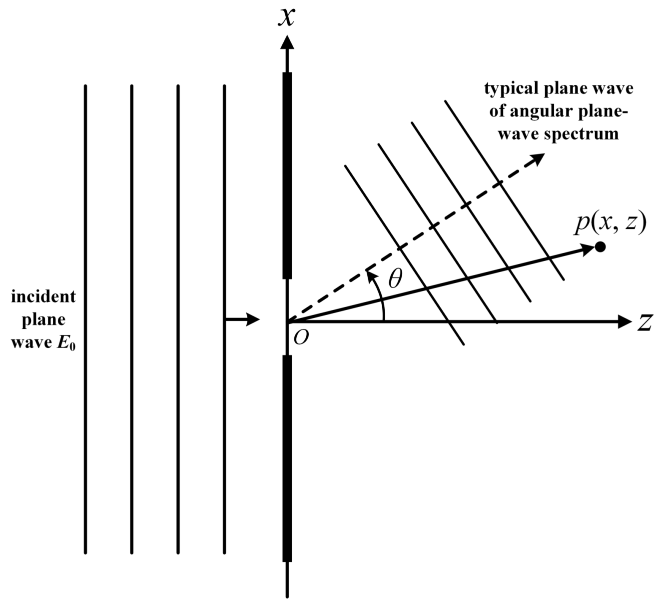

Figure 1.2: A plane wave diffracted by a slit. Plane-wave spectrum representation.

equations 1.7 and 1.8 can be combined to give

$$f(x) = \int_{-\infty}^{\infty} F(\sin\theta)\exp(-jkx\sin\theta)\mathrm{d}(\sin\theta) \tag{1.9}$$

This is a Fourier integral that can be inverted to give the angular spectrum $F(\sin\theta)$ in terms of the aperture field $f(x)$ as

$$F(\sin\theta) = \frac{1}{\lambda}\int_{-\infty}^{\infty} f(x)\exp(jkx\sin\theta)\mathrm{d}x \tag{1.10}$$

In the elementary diffraction example of section 1.1 the aperture field is assumed to be non-zero only across the slit, so that

$$f(x) = \begin{cases} E_0 & \text{for } |x| \leqslant a/2 \\ 0 & \textit{otherwise} \end{cases} \tag{1.11}$$

Substituting this into equation (1.10) yields the angular spectrum

$$\begin{aligned} F(\sin\theta) &= \frac{E_0}{\lambda}\int_{-a/2}^{a/2}\exp(jkx\sin\theta)\mathrm{d}x \\ &= \frac{E_0 a}{\lambda}\frac{\sin(\frac{\pi a\sin\theta}{\lambda})}{(\frac{\pi a\sin\theta}{\lambda})} \end{aligned} \tag{1.12}$$

which has the same angular dependence as the far field of equation (1.3) derived from Huygen's principle. Thus we can seen in this particular example, and will later establish as generally true, that the angular spectrum gives the far field of the radiation. And furthermore, as equation (1.10) shows, the angular spectrum is simply the Fourier transform of the field across the radiation aperture. Once $F(\sin\theta)$ is known the radiated field $E_p(x, z)$ at any point to the right of the aperture plane is determined by the integral of equation (1.7).

The angular plane-wave spectrum representation was popular around the turn of the last century. Rayleigh (1896) used it to describe the fields reflected and transmitted at a corrugated dielectric boundary illuminated by a plane wave, and Debye (1909) used the representation to investigated the fields in the region of a focus. In 1902 Whittaker had proved that the field described by an angular spectrum of plane waves radiating in all directions was a solution of the wave equation. But this form of the angular spectrum leaves the source of radiation unspecified, and is therefore not unique. This limitation was overcome in Weyl's (1919) representation of the field radiating into a half-space bounded by a plane. The source field can be specified over this plane, and so the representation can be unique.

The application of the angular plane-wave spectrum to aperture antenna analysis and synthesis was pioneered by Booker and Clemmow (1950) and Woodward and Lawson (1948), essentially using the Weyl representation in two dimensions. Brown (1958) extended the representation the three dimensions, and introduced a reciprocity theorem for antennas which enabled the concept to be applied to receiving as well as transmitting antennas. This led to a transmitter/receiver coupling formula which applied to near-field coupling as well as to the far field. The angular plane-wave spectrum concept is now used extensively in planar-antenna synthesis (Rhodes, 1974) and in near-field antenna measurements (Paris and Joy, 1978).

In Chapter 2 and 3 of this book the concept of an angular spectrum of plane waves is established for two- and three-dimensional fields. In Chapter 4 the concept is applied to transmitting antennas, receiving antennas, and the coupling between them. Fresnel diffraction is examined in Chapter 5, and reflection from flat and curved conducting surfaces in Chapter 6. Particular examples of planar aperture antennas are examined in detail in Chapter 7. Chapter 8 looks briefly at the problem of radiation from non-planar apertures. Appendices provide a summary of the properties of plane waves and of some theorems derived from Maxwell's equations. The remainder of this introductory chapter will be devoted to a physical description of the planar-aperture antennas to be examined in Chapter 7, and to giving a list of those performance characteristics and terms that are in common use in antenna practice.

1.3 Some Aperture Antennas

The following example of microwave antennas used in communications or radar can be described as planar-aperture antennas.

The Electromagnetic Horn Antenna
At microwave frequencies a simple way of making a radiating antenna is to use an open-ended waveguide with a flared transition section (that is, a horn) added to achieve a reasonable match and increased directivity. (See Fig. 1.3).

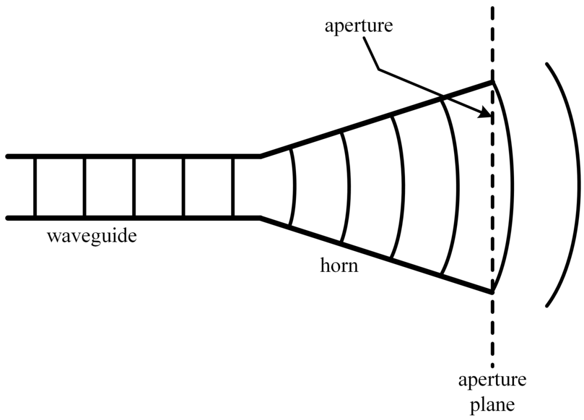

Figure 1.3: The Electromagnetic Horn Antenna. Feint lines indicate equiphase wavefronts of the radiating field.

An obvious starting point for the analysis of such a radiating device is to assume that the field over the mouth of the horn is an expanded form of the front. Thus the aperture of this antenna is the real aperture formed by the mouth of the horn, assumed to be part of the aperture plane. The field distribution over this plane, the aperture field distribution, could be taken approximately to be the expanded waveguide field over the actual aperture and zero over the remaining part of the aperture plane.

The electromagnetic horn is robust and fairly easy to manufacture. It is widely used as a standard in antenna gain measurements. (For a definition of antenna gain see section 1.4). But owing to the curvature of the emerging wavefront its gain is relatively low. The different methods that have been devised to correct the curved wavefront of simple primary sources, such as the electromagnetic horn, into the more desirable planar wavefront have led to a variety of antenna designs, of which the horn-lens combination is perhaps the most obvious.

The Horn-Lens Antenna
Here the curved wavefront of the field emerging from the mouth of the elec-

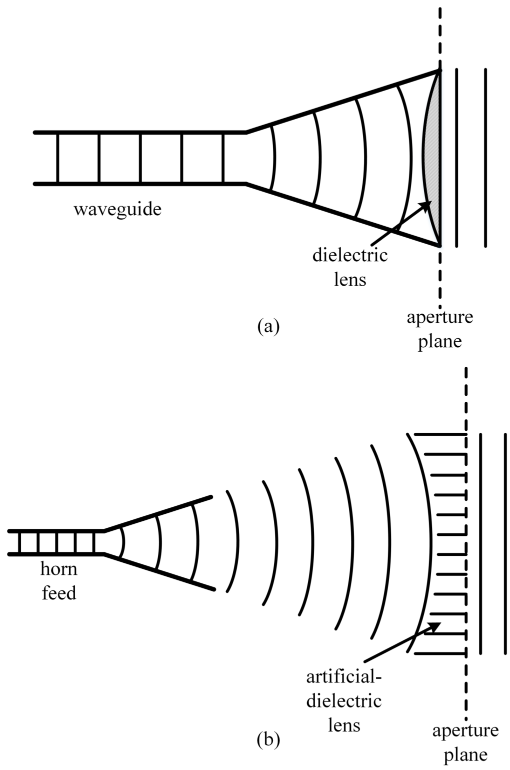

Figure 1.4: The Horn-Lens Antenna. The curved wavefront corrected width (a) a solid dielectric lens, and (b) an artificial-dielectric lens.

tromagnetic horn is corrected by the use of a converging lens. The resulting planar wavefront leads to an improvement in the directivity of the radiated field. Two examples are shown in Fig. 1.4). The solid dielectric lens of Fig. 1.4(a) behaves in exactly the same way as a convex glass lens in optics. The artificial dielectric lens shown schematically in Fig. 1.4(b) consists of a stack of meta-walled waveguides within which the phase velocity of the waves is greater than in air hence the concave construction of the lens. In both cases the natural aperture is just beyond the output face of the lens.

The advantage of the increased directivity and gain are somewhat offset in the horn lens antenna by the reflections that inevitably occur at the front and back surfaces of the lens. The essential robustness of the electromagnetic horn is also diminished in the horn-less combination.

The Horn-Paraboloid Antenna

An ingenious design which achieves the desired wavefront correction of the field emerging from a horn, without introducing partial reflections and without loss of robustness, is the horn-paraboloid assembly shown in Fig. 1.5. The metal cowl is welded on to the horn and has paraboloidal (this is, part of

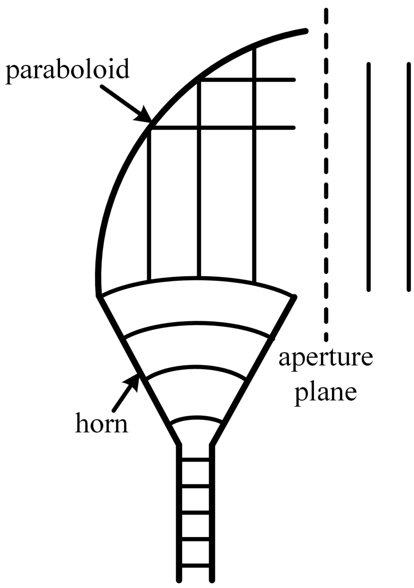

Figure 1.5: The Horn-Paraboloid Antenna.

a parabola of resolution) profile which transforms the spherical wavefront at the mouth of the horn into a planar wavefront which emerges from the side of the assembly. It is convenient in this case to suppose that the radiating aperture of this device lies just beyond the metal structures, as shown in the figure.

However, there is another factor which affects antenna directivity. For maximum directivity the wavefront in the aperture should not only be planar, but as wide as possible. In the antenna designs so far mentioned emphasis has been on correcting the wavefront. We now turn to a series of antenna designs which both correct the wavefront and extend the lateral dimensions of the aperture.

The Paraboloid Reflector Antenna

This antenna consists of a primary feed, which is shown in Fig 1.6. as an electromagnetic horn, positioned at the focus of a paraboloid reflector. The

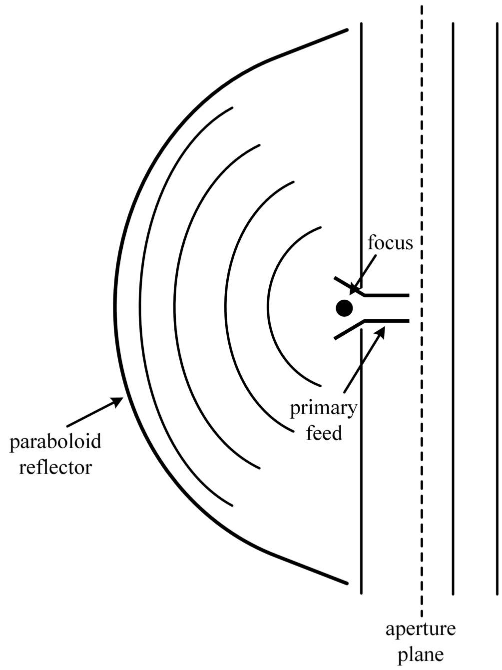

Figure 1.6: The Paraboloid Reflector Antenna.

geometrical properties of the parabola ensure that the spherical wave incident from the feed is transformed into a plane wave on reflection. The lateral extent of the emerging aperture fields, which it is convenient to place beyond the primary feed and its support structure, can be as wide as the lateral extent of the paraboloid.

The simplicity of design (basically that of the searchlight and Newton telescope) and its robustness of construction make the focus-fed paraboloid reflector a very popular choice as a high-gain antenna in communications and radar. However, the physical presence between the reflector and the supposed aperture of the primary feed, its feeder waveguide and its associated support structure, give rise to some deterioration in performance, and to some rather unsatisfactory modifications to the analysis, in comparison with the ideal.

The Cassegrain Double Reflector Antenna

The addition of a second reflector of hyperboloid shape, labeled as the subreflector in Fig. , leads to an even more robust and compact design, known as the Cassegrain. (The Cassgrain system also originated as a design for an optical telescope). The primary feed is inserted through the center of the main

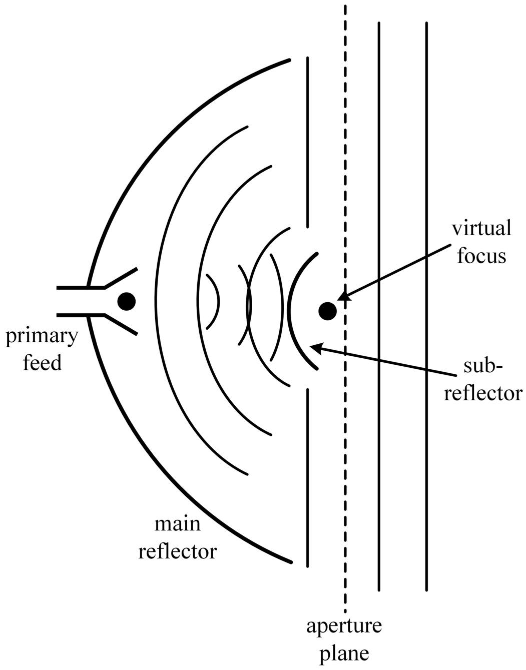

Figure 1.7: The Cassgrain Double Reflector Antenna.

paraboloidal reflector, thus eliminating long waveguides feeders and their association noise problems. There is also the possibility of simple control of the antenna radiation pattern by appropriately shaping the sub-reflector. But the problem aperture blocking, this time by the sub-reflector and its supports, still remains.

The Offset-Fed Paraboloid Reflector Antenna

The problem of blocking of the desired aperture field by the feed can be largely overcome by placing the primary feed in an offset position, as indicated in Fig. 1.8, and appropriately restricting the extent of the main reflector. The price to be paid for the resulting improvement is that such an antenna is awkward to manufacture (at least for mass production), and the skew geometry can give rise to increased cross-polarized radiation, which may be undesirable.

Nonetheless, our main concern here is not with the pros and cons of different antennas, but with the common feature that makes them eligible to be considered as aperture antennas; which is that an analysis of their perfor-

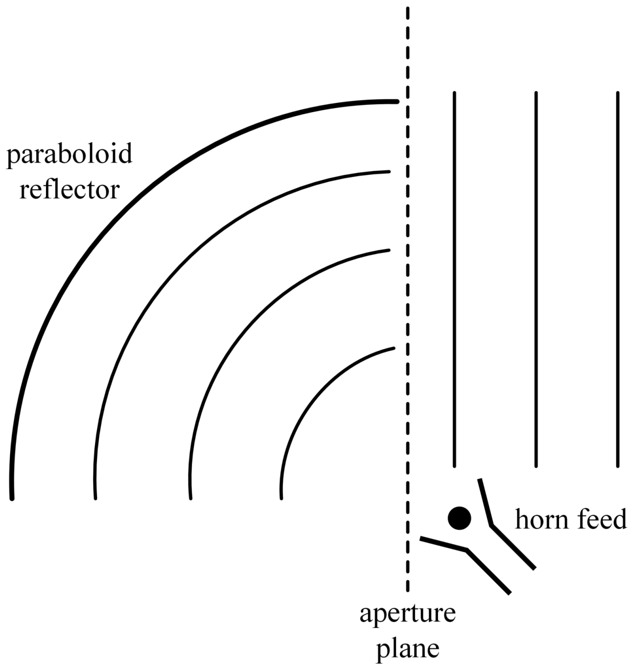

Figure 1.8: The Offset-Fed Paraboloid Reflector Antenna.

mance arises naturally (but not necessarily exclusively) from consideration of a field distribution over some surface in the vicinity of the antenna structures. This is clearly so in the present instance, as suggested in Fig. 1.8.

1.4 Basic Antenna Concepts

The object of antenna analysis is an accuracy description of the radiating and receiving characteristics of an antenna. Over the years certain concepts and definitions have developed and become established as part of the common language among antenna engineers. We will review these concepts and definitions here. They will be developed in detail and discussed at greater length in later chapters.

The Radiated Far Field
At a large distance r from any antenna its electric field can be expressed in the form

$$\mathbf{E}(r,\theta,\phi) = \frac{\exp(-jkr)}{kr}\mathbf{e}(\theta,\phi) \tag{1.13a}$$

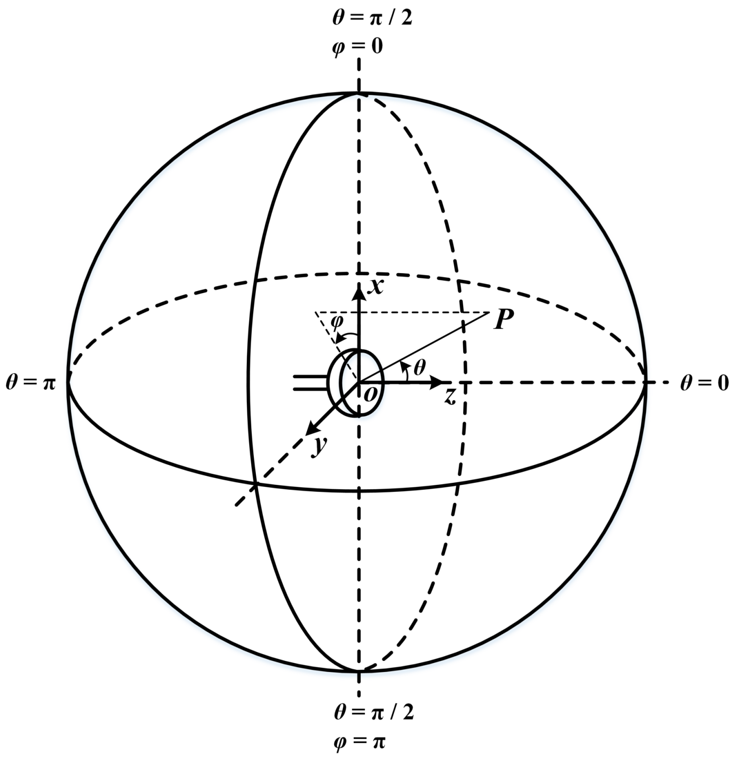

Figure 1.9: Geometry for the radiated field.

such that

$$\mathbf{u}_r \cdot \mathbf{E}(r, \theta, \phi) = 0, \tag{1.13b}$$

and the associated magnetic field is

$$\mathbf{H}(r, \theta, \phi) = \frac{1}{Z}\mathbf{u}_r \times \mathbf{E}(r, \theta, \phi) \tag{1.14}$$

The observation point P (See Fig. 1.9) is assumed to lie on a sphere of radius r, centered on a convenient point O in the vicinity of the antenna. The spherical polar coordinates of the point P are (r, θ, ϕ) where θ, the polar angle, and ϕ, the azimuth angle, together define the direction of P from O. The direction (θ, ϕ) is also that of the unit vector in the radial direction, $\mathbf{u}_r$.

The following features of the antenna far field should be noted:

(a) The distance r of the observation point from the antenna should be at least the Rayleigh distance, which will be shown in section 5.5.3 to be $2a^2/\lambda$, where a is the maximum dimension of the radiating aperture.

(b) The time variation of the fields is understood to be sinusoidal, having the complex from $\exp(j\omega t)$. This meas that the actual space-time behavior for a phasor-vector such as $\mathbf{E}(r, \theta, \phi)$ is obtained from

$$\mathbf{E}(r, \theta, \phi, t) = \mathrm{Re}[\mathbf{E}(r, \theta, \phi)e^{j\omega t}] \tag{1.15}$$

where Re denotes the real part. The convention we are using is that the phasor-vector $\mathbf{E}$ has a complex magnitude E and an absolute magnitude $|E|$, which is also the peak value of the sinusoidal time variation.

(c) The form of the far electric field $\mathbf{E}(r, \theta, \phi)$ of equation (1.13a) is the product of a uniform spherical scalar wave and the vector pattern function $\mathbf{e}(\theta, \phi)$.

(d) The dimension of $\mathbf{e}(\theta, \phi)$ are the same as those of $\mathbf{E}(r, \theta, \phi)$, namely volts per meter, as a consequence of arbitrary multiplying r in the denominator of equation equation (1.13a) by the plane-wave phase constant $k = \omega\sqrt{\mu\epsilon}$.

(e) In a particular direction (θ, ϕ) the amplitude of the field falls off as r^{-1}, and its phase is retarded linearly as kr, with increasing radial distance r.

(f) At a constant radial distance the dependence on direction of the amplitude, phase and polarization (that is, vector direction) of the electric field is given by $\mathbf{e}(\theta, \phi)$, which is the vector pattern function for a particular antenna.

(g) The electric field $\mathbf{E}(r, \theta, \phi)$ is polarized such that it is always tangential to the sphere of radius r, as specified by equation (1.13b). This is illustrated in Fig. 1.10.

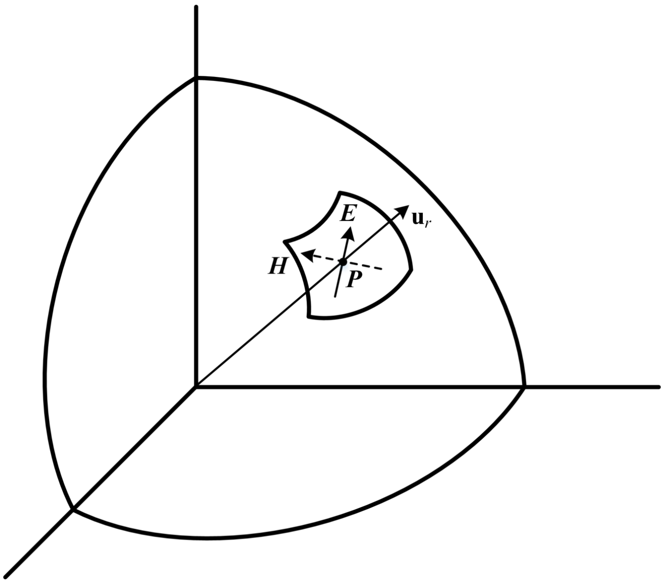

Figure 1.10: Polarization of the far field. $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{u}_r$ form a right-handed set.

(h) The magnetic field $\mathbf{H}(r, \theta, \phi)$ of equation (1.14) is also tangential to the sphere of radius r. It is also orthogonal to $\mathbf{E}(r, \theta, \phi)$. The $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{u}_r$ are mutually orthogonal everywhere in the far field. They form a right-handed set in the stated order (that is, if $\mathbf{E}$ is rotated on to $\mathbf{H}$, a right-handed screw thread would advantage in the direction of $\mathbf{u}_r$) as specified by equation (1.13) and (1.14) and illustrated in Fig. 1.10.

(i) In equation (1.14), which relates the electric and magnetic fields, Z is the plane-wave impedance of the propagating medium. In terms of the permeability μ and permittivity ϵ of the medium,

$$Z = \sqrt{\frac{\mu}{\epsilon}} \tag{1.16}$$

(j) In the neighborhood of any point $\mathbf{P}$ in the field the electric and magnetic fields have the *local* character of a plane wave traveling in the radial direction.

(k) The power flux density, given by half the real part of the complex Poynting vector

$$\mathbf{S} = \frac{1}{2} \text{ Re } \mathbf{E} \times \mathbf{H}^*, \tag{1.17}$$

is directed radially outward (this is, $\mathbf{S} = \mathbf{u}_r S_r$) for the values of $\mathbf{E}$ and $\mathbf{H}$ given above for the far field. The asterisk denotes complex conjugate. The radial component of the power flux density is therefore

$$S_r(r, \theta, \phi) = \frac{\mathbf{e}(\theta, \phi) \cdot \mathbf{e}^*(\theta, \phi)}{2(kr)^2 Z} = \frac{|e(\theta, \phi)^2|}{2(kr)^2 Z} \tag{1.18}$$

Directivity, Gain and Efficiency

The directivity $D(\theta, \phi)$ of a radiating antenna is defined as the ratio:

$$D(\theta, \phi) = \frac{\text{Power flux density (p.f.d) from antenna in direction } (\theta, \phi)}{\text{P.f.d. when same power is radiated uniformly in all directions}}. \tag{1.19}$$

The power flux densities in this definition must be measured at the same radial distance, or equivalently the power flux density must be defined per unit solid angle. In either case the definition refers to the far field.

If P_t is the total power radiated by the antenna, the power flux density at a distance r when this is radiated uniformly in all directions (that is, radiated isotropically) is

$$S_r^{iso}(r, \theta, \phi) = \frac{P_t}{4\pi r^2}, \tag{1.20}$$

and the antenna directivity is

$$D(\theta, \phi) = \frac{S_r(r, \theta, \phi)}{S_r^{iso}(r, \theta, \phi)} = \frac{\lambda^2 |e(\theta, \phi)|^2}{2\pi Z P_t} \tag{1.21}$$

for the far fields defined earlier.

Antenna gain $G(\theta, \phi)$ has a definition similar to that of directivity, the only difference being that the denominator of equation (1.19) is based on the input power P_{in} delivered to the antenna. Thus

$$G(\theta, \phi) = \eta D(\theta, \phi); 0 \leqslant \theta \leqslant 1, \tag{1.22}$$

where η, know as the antenna efficiency, is the ratio of total power radiated to the power input to the antenna, namely,

$$\eta = P_t / P_{in} \tag{1.23}$$

In terms of the far-field pattern function, then,

$$G(\theta,\phi) = \frac{\lambda^2 |e(\theta,\phi)|^2}{2\pi Z P_{in}} \tag{1.24}$$

The maximum value of the gain function $G(\theta,\phi)$ is also often referred to simply as the gain of the antenna.

The polarization of the far field is described by the complex vector $\mathbf{e}(\theta,\phi)$, which is constrained to lie in a plane perpendicular to the radial direction. The vector, $\mathbf{e}$ can therefore be resolved into two components with basis vector $\mathbf{e}_1$ and $\mathbf{e}_2$, so that

$$\mathbf{e} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 \tag{1.25}$$

in which a_1 and a_2 are complex. The basis vectors can be so chosen that they are normalized with

$$\mathbf{e}_1 \cdot \mathbf{e}_1{}^* = \mathbf{e}_2 \cdot \mathbf{e}_2{}^* = 1 \tag{1.26}$$

and orthogonal in the sense that

$$\mathbf{e}_1 \cdot \mathbf{e}_2{}^* = \mathbf{e}_2 \cdot \mathbf{e}_1{}^* = 0. \tag{1.27}$$

This resolution of the field could be into two orthogonal linearly polarized plane waves, or into a combination of right-hand and left-hand circularly polarized plane waves, whichever is more appropriate. Then

$$\mathbf{e} \cdot \mathbf{e}^* = |a_1|^2 + |a_2|^2 = |e|^2 \tag{1.28}$$

and it is clear from equation (1.21) that the directivity can be resolved into two parts, namely,

$$D(\theta,\phi) = D_1(\theta,\phi) + D_2(\theta,\phi) \tag{1.29}$$

based on the chosen resolution of the field. Equation (1.24) shows that the gain $G(\theta,\phi)$ can be similarly resolved.

Radiation Patterns, Beamwidth and Sidelobes

It is usually required of an antenna that it be directive. The simplest measure of its effectiveness as such is the gain, that is, the maximum value of the gain function $G(\theta,\phi)$. The gain is thus the ratio, invariably stated in decibels (dB), of the maximum power flux density produced by the antenna to its value if the power delivered to the antenna had been radiated isotropically.

More information about its directive properties can be obtained from the antenna's radiation patterns. There are plots of radiated field strength or, more usually, power flux density (directivity or gain) as a function of angle.

Since direction is specified by two angles whereas it is only possible to plot against one, antenna radiation patterns consist of a set of section in the θ–ϕ, or some equivalent, plane. For example, suppose that an antenna has its maximum power radiation in the direction $\theta = 0$. Then a plot of the antenna gain, in decibels relative to the maximum value, as a function of θ with ϕ held constant might be that sketched in Fig. 1.11. Two methods of plotting

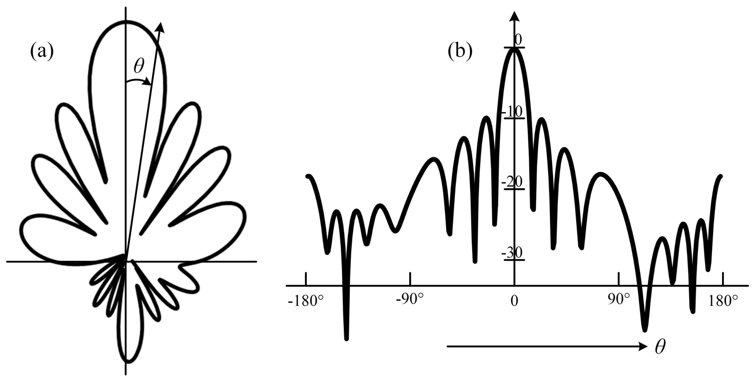

Figure 1.11: Polar and rectangular plts of relative gain in dB versus for constant ϕ.

the same information are shown, one in polar the other in rectangular form.

Antenna radiation patterns often have a lobe structure similar to that shown. The lobe containing the direction of maximum radiated power is the main lobe; all other lobes are referred to as sidelobes. The two sidelobes adjacent to the main lobe are the first sidelobes, and the lobe diametrically opposite the main lobe, if it exists, is called the back lobe. The lobes are separated by nulls, so called because the power radiated in these directions can in theory be zero. The angular width of the main lobe of the antenna, known as its beam width, is a useful measure of the capability of the antenna of resolving the angular position of a distant point source. Two common ways of defining antenna beamwidth are shown in Fig. 1.12. One is the angle between the two points on either side of the main lobe at which the radiated power has fallen to half its maximum value, that is, the -3 dB points, and is known as the 3-dB beamwidth. The other definition is the angle between the first nulls of the pattern, which is usually about twice the 3-dB beamwidth. This fact makes the additional point that if two distant sources are separated in angle by the 3-dB beamwidth and one of them lies in the center of the main beam, then the other will lie in the region of the first null.

Receiving Antennas

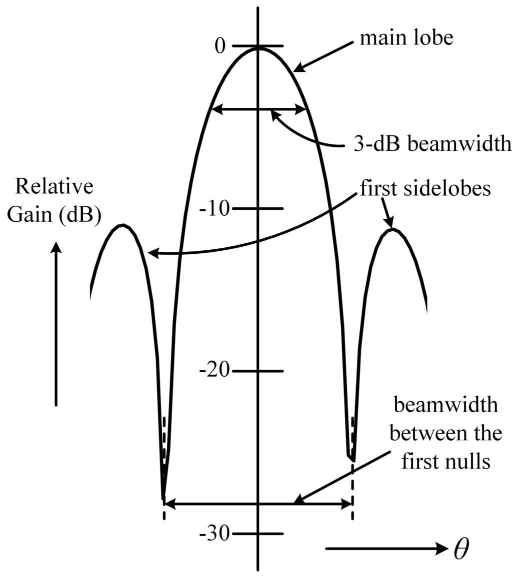

Figure 1.12: Details of the main lobe and first sidelobes of the antenna pattern in Fig. 1.11(b). showing two ways of defining antenna beamwidth.

The directivity of a receiving antenna is defined in terms of its response to an incident plane wave. It takes the form of an effective receiving area:

$$A(\theta,\phi) = \frac{\text{Power delivered to receiver}}{\text{Power density in plane wave incident from direction } (\theta,\phi)} \tag{1.30}$$

It is assumed that the polarization of the incident plane wave is adjusted to deliver maximum power to the receiver: a condition known as **polarization match**.

Consider the plane wave to be $\mathbf{E}'$, incident from the direction (θ',ϕ'), as shown in Fig. 1.13. If the antenna has a vector far-field radiation pattern $\mathbf{e}(\theta,\phi)$, the antenna reciprocity theorem of section 4.2.1 states that the received signal will be proportional to $\mathbf{E}'\cdot\mathbf{e}(\theta',\phi')$. The vector $\mathbf{E}'$ and $\mathbf{e}(\theta',\phi')$ lies in the same plane; the condition of polarization match is therefore when they are coincident in that plane. When this occurs the power delivered to the receiver will be proportional to $|E'|^2\cdot|e(\theta',\phi')|$, which means that it is proportional to the product of the power density in the incident plane wave and the gain of the antenna given by equation (1.24). Thus the effective receiving area of an antenna proportional to its transmitting gain. The precise

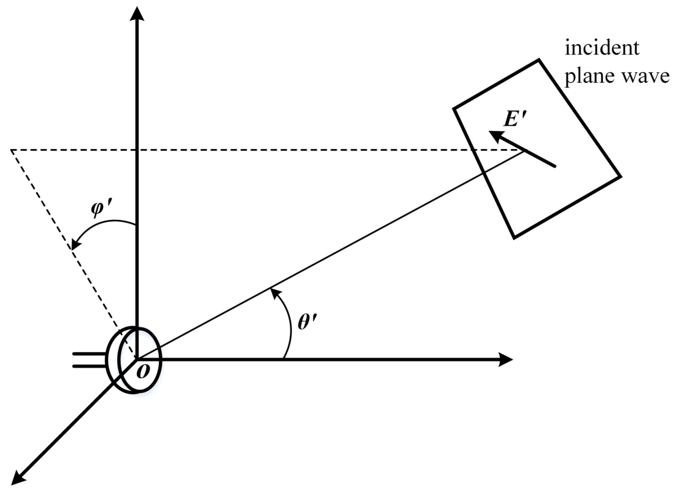

Figure 1.13: Plane wave incident on a receiving antenna.

relationship will be shown in section 4.2.2 to be

$$A(\theta, \phi) = \frac{\lambda^2}{4\pi} G(\theta, \phi) \tag{1.31}$$

Transmitter to Receiver Coupling

An immediate use for the definition we have just given of antenna gain and receiving area is the far-field coupling equation, known as the **Friis transmission formula** which we will now derive. (This is a particular form of a more general result that will be obtained later in section 4.3).

Suppose that the transmitting and receiving antennas are disposed as shown in Fig. 1.14. The distance r between them is presumed to satisfy the far-field criterion that $r \geqslant 2a^2/\lambda$, where a is now the largest dimension of both antenna apertures. The receiver is in the direction θ, ϕ from the transmitter, and the transmitter is in the direction (θ', ϕ') from the receiver; and if the transmitting antenna gain is $G_T(\theta, \phi)$, when P_{in} is delivered to the transmitter, the power flux density incident on the receiving antenna in the form of a locally plane wave will be

$$S_r = P_{in} \frac{G_T(\theta, \phi)}{4\pi r^2} \tag{1.32}$$

If the effective receiving area of the receiving antenna is $A_R(\theta', \phi')$, the power delivered to the receiver will be

$$P_{rec} = P_{in} \frac{G_T(\theta, \phi) A_R(\theta', \phi')}{4\pi r^2}, \tag{1.33}$$

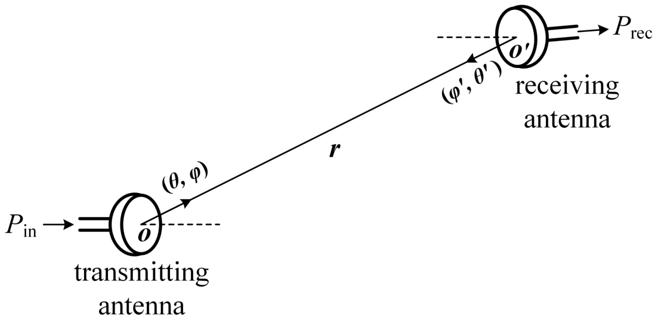

Figure 1.14: Plane wave incident on a receiving antenna.

it being assumed that the receiving antenna is polarization matched to the incident field. An alternative form for the received power, using equation (1.31) is

$$P_{rec} = P_{in} \frac{\lambda^2 G_T(\theta, \phi) G_R(\theta', \phi')}{4\pi^2 r^2} \tag{1.34}$$

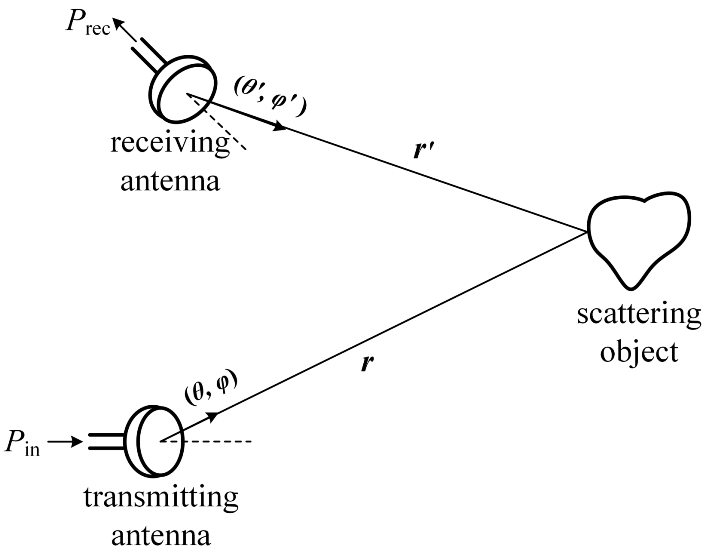

Figure 1.15: Geometry for the indirect coupling between a transmitting and receiving antenna via a scattering object.

Another important form of coupling between a transmitter and a receiver occurs indirectly via a scattering object, as in radar. The scatterer is represented by its **scattering cross-section**, which has the following rather tortuous but basically simple definition. The scattering cross-section σ is that area which, when placed perpendicular to the plane-wave field incident on the scatterer, would intercept that amount of power which, when radiated uniformly in all directions, would give the observed power flux density in a particular direction, it is therefore a function of the direction and polarization of the incident field, and of the direction of observation.

If, with the geometry show in Fig. 1.15, power P_{in} is delivered to the transmitting antenna and the scatterer is in its far field, the power intercepted will be $P_{in}G_T(\theta,\phi)\sigma/4\pi r^2$, and the power delivered to the receiver will be therefore be

$$P_{rec} = P_{in}\frac{\sigma G_T(\theta,\phi)G_R(\theta',\phi')}{4\pi^2 r^2 r'^2}, \tag{1.35}$$

again assuming that the receiving antenna is polarization matched to the field incident upon it. In radar systems the same antenna is often used for transmitting and receiving, in which case the received power is

$$P_{rec} = P_{in}\frac{\lambda^2 G^2(\theta,\phi)}{4\pi^3 r^4}, \tag{1.36}$$

where we have set $G_T(\theta,\phi) = G_R(\theta,\phi) = G(\theta,\phi), r' = r$ and have made use of equation (1.31).

Chapter 2

Plane-wave Representation of Two-dimentional Fields

Plane waves are exact solutions of Maxwell's fundamental field equations. Maxwell's equations are linear, provided the medium itself is linear (that is, the electrical properties of the medium such as its permittivity and permeability are independent of the electric and magnetic field). Hence the principle of linear superposition applies, and the superposition of individual plane waves traveling in different directions in the medium constitutes an exact solution.

Our ultimate goal is to study the diffraction of electromagnetic waves in realistic, three-dimensional situations. But it is much easier to introduce the underlying concepts and techniques in two dimensions. Hence this chapter is devoted to building up a picture of two-dimensional fields in term of individual plane of this representation. In the following chapter it will then to be possible to extend the representation to three dimensions, concentrating on the inevitably increased complexity of the mathematical while taking the underlying concepts for granted.

2.1 Plane Waves

We begin by listing the properties of homogeneous plane electromagnetic waves. These will be familiar to most readers. But if they are not, a derivation based on transmission-line theory can be found in Appendix A.

A plane electromagnetic waves, of single-frequency time dependence $\exp(j\omega t)$, traveling in a uniform, isotropic an lossless open region, has the following properties:

(a) It is characterized by a single direction, such as that denoted by the unit $\mathbf{u}$ in Fig. 2.1. This direction is the direction of propagation of the

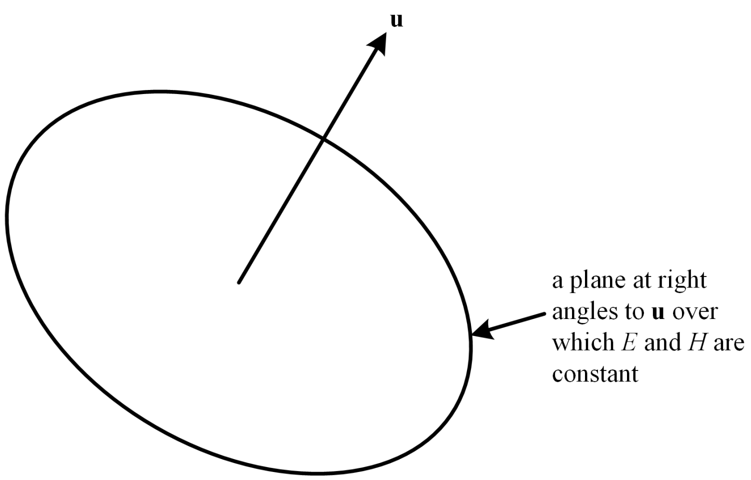

Figure 2.1: A plane wave traveling in the direction of $\mathbf{u}$.

plane wave.

(b) Over planes at right angles to $\mathbf{u}$ the phasor electric and magnetic fields, E and H, are constant in both amplitude and phase.

(c) The complex amplitude E and H are related to each other by the proportionality

$$E = ZH \tag{2.1}$$

in which the constant Z is the characteristic (or plane-wave) impedance of the medium. In the SI system, the units of E are volts per meter and those of H are amperes per meter, so that the units of Z are ohms. In a uniform lossless medium of permeability μ and permittivity ϵ

$$Z = (\mu/\epsilon)^{1/2} \tag{2.2}$$

and is therefore real. This meas that the electric and magnetic fields will be in phase. If the plane wave is traveling in free space, $\mu = 4\pi \times 10^{-7}$ heries per meter and $\epsilon = 8.854 \times 10^{-12}$ farads per meter, and its characteristic impedance is 376.7 ohms.

(d) The directions of the vector electric field $\mathbf{E}$, the vector magnetic field $\mathbf{H}$, and the direction of propagation $\mathbf{u}$ are mutually at right angles, as indicated in Fig. 2.2. The vector $(\mathbf{E}, \mathbf{H}, \mathbf{u})$ form a right-handed set, which means that they bear the same relationship to each other as, for example, the directions $(\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z)$ of the axes of the usual Cartesian

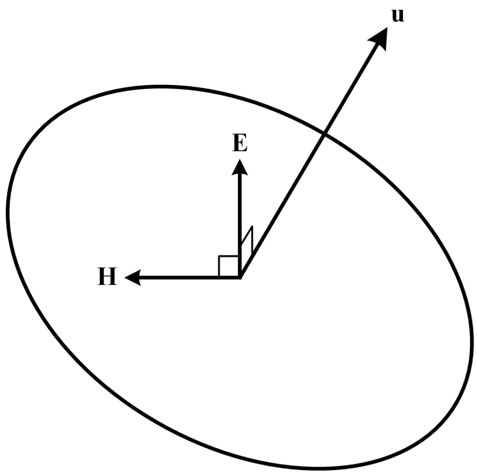

Figure 2.2: Vector fields **E** and **H** form a right-handed set with **u**.

coordinate system.

Exercise

Suppose that the above plane wave is traveling in free space in a direction from the origin of a Cartesian coordinate system towards the point (1,1,1), and that we are told that the x-component of the electric field has a peak value of $10^{-3}Vm^{-1}$, and that the y-component of the electric field is zero. Calculate the remaining field components.

Exercise

Show that for the plane wave described geometrically in Fig. 2.2 the vector magnetic field is given by the vector

$$\mathbf{H} = Z^{-1}\mathbf{u} \times \mathbf{E} \tag{2.3}$$

(e) In a lossless medium the magnitude of the fields are the same everywhere but the phase of both the electric and magnetic fields is retarded in the direction of propagation of the wave. Thus, if the electric field over some reference plane containing the point O in Fig. 2.3. is E_0, then the electric field E over a plane a distance ξ away from the reference plane in the direction of propagation will be

$$E = E_0 \exp(-jk\xi) \tag{2.4}$$

where the time factor $\exp(jwt)$ as is usual for phasor representation of fields, has been suppressed. The phase constant k is the amount by

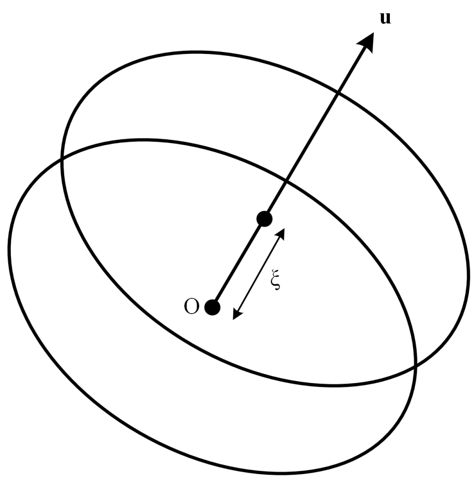

Figure 2.3: Planes separated by a distance ξ in the direction of propagation.

which the phase of the plane wave is retarded in unit distance, and is given by

$$k = \omega(\mu\epsilon)^{1/2} \tag{2.5}$$

for a lossless medium of permeability μ and permittivity ϵ. The wavelength λ of a propagation wave is the distance over which the phase changes by 2π. Hence

$$k = 2\pi/\lambda \tag{2.6}$$

Exercise
Find the characteristic impedance and speed of travel of a plane wave propagating in a uniform lossless dielectric of relative permeability 1 and relative permittivity 2.56.
If now the same dielectric has a small amount of conductivity such that in loss tangent (for a definition see Appendix A.6) is 0.02 instead of zero, how will this change the characteristic impedance and speed of travel of the plane wave? What is the loss in dB/km?

(f) In antenna parlance the term polarization refers to the direction of the vector electric field, which for a plane wave is necessarily confined to planes orthogonal to **u**. If **E** points in a single direction throughout all time and space then the plane wave is said to be **linearly polarized**. Any plane wave traveling in the direction **u** can be represented as the sum of suitable scalar multiples of any two linearly polarized plane waves $\mathbf{E_1}$ and $\mathbf{E_2}$, provided they are not collinear. The most convenient choice is to take $\mathbf{E_1}$ and $\mathbf{E_2}$ to be at right angles, as shown in Fig. 2.4.

A **circularly polarized** plane wave is one which can be represented

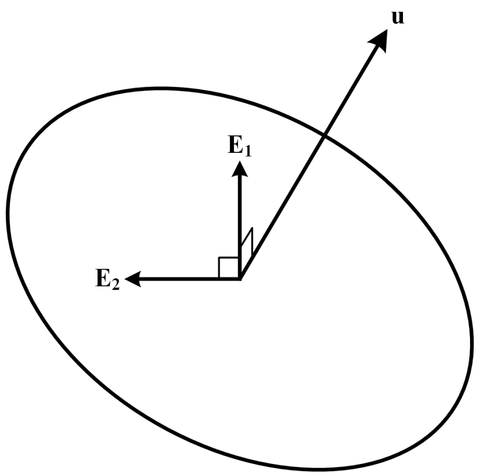

Figure 2.4: Two orthogonal linearly polarized plane waves $\mathbf{E_1}$ and $\mathbf{E_2}$.

as the sum of two orthogonal linear polarized plane waves of equal amplitude but out of phase by $\pi/2$ radians. Let $\mathbf{E}_1 = E_1\mathbf{u}_1$ and $\mathbf{E}_2 = E_2\mathbf{u}_2$, where $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal unit vectors, and put $E_1 = E_0$ and $E_2 = \mp jE_0$. Then the plane wave withe electric field

$$\mathbf{E} = E_0\mathbf{u}_1 \mp jE_0\mathbf{u}_2 \tag{2.7}$$

is a circularly polarized plane wave, the negative sign corresponding to clockwise rotation of the electric vector as the wave propagates (also known as right-handed circular polarization), and the positive sign corresponding to anticlockwise rotation (or left-handed circular polarization).

If the two orthogonal linearly polarized plane waves are combine in phase, then the resultant wave is again linearly polarized. The combination of two such orthogonal linearly polarized plane waves is akin to the formation of Lissajou figures on an oscilloscope screen. When their amplitude and relative phases are arbitrary the tip of the electric-field vector traces out an ellipse, and the plane wave is said to be **elliptically polarized**. (see Appendix A.5 for a fuller treatment).

Exercise

Show that a linearly polarized plane wave can be represented as the sum of two circularly polarized plane waves traveling in the same direction, of the same amplitude but of opposite sense. Hence show that an arbitrary polarized plane wave can be represented as a suitable weighted sum of two circularly polarized waves of opposite sense.

Exercise

The two linearly polarized electric field vector $\mathbf{E}_1$ and $\mathbf{E}_2$ depicted in Fig. 2.4 satisfy the general orthogonal condition for electromagnetic fields, which is that the scalar product

$$\mathbf{E}_1 \cdot \mathbf{E}_2{}^* = 0 \tag{2.8}$$

in which the asterisk denotes the complex conjugate. Form the two circularly polarized plane waves

$$\mathbf{E}_r = E_1\mathbf{u}_1 - jE_1\mathbf{u}_2 \text{ and } \mathbf{E}_l = E_2\mathbf{u}_1 - jE_2\mathbf{u}_2 \tag{2.9}$$

in which $\mathbf{u}_1$ and $\mathbf{u}_2$ are spatially orthogonal, and show that $\mathbf{E}_r$ and $\mathbf{E}_l$ are themselves orthogonal.

(g) The power flow in the plane wave is given, as for any electromagnetic wave, by the vector

$$\mathbf{E} = \frac{1}{2}\ \mathrm{Re}\ \mathbf{E} \times \mathbf{H}^* \tag{2.10}$$

which is the Poynting vector averages over one cycle of the oscillation of frequency $f = 2\pi/\omega$. The asterisk denotes complex conjugate and Re the real part. It is clear from Fig. 2.2 that the Poynting vector $\mathbf{S}$ is in the same direction as $\mathbf{u}$, the direction of propagation of the plane wave. The units of $\mathbf{S}$ are watts per square meter. For the homogeneous plane wave of Fig. 2.2 in a lossless medium, the electric and magnetic fields are not only spatially orthogonal but are also precisely in phase, so the vector power flow is

$$\mathbf{S} = (2Z)^{-1}|E|^2\mathbf{u} = (Z/2)|H|^2\mathbf{u} \tag{2.11}$$

Exercise

Rederive the last formula for the power flow in a plane wave by substituting the vector relationship between $\mathbf{E}$ and $\mathbf{H}$, namely.

$$\mathbf{H} = Z^{-1}\mathbf{u} \times \mathbf{E} \text{ or } \mathbf{E} = Z\mathbf{H} \times \mathbf{u} \tag{2.12}$$

into the formula for the average Poynting vector. Use the vector triple product relations for any three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, that

$$\begin{aligned} \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) &= (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c} \\ (\mathbf{a} \times \mathbf{b}) \times \mathbf{c} &= (\mathbf{c} \cdot \mathbf{a})\mathbf{b} - (\mathbf{c} \cdot \mathbf{b})\mathbf{a} \end{aligned} \tag{2.13}$$

Exercise
Show that the power flow in a plane wave represented as the sum of two orthogonally polarized plane waves is just the sum of the power flows in the two waves.

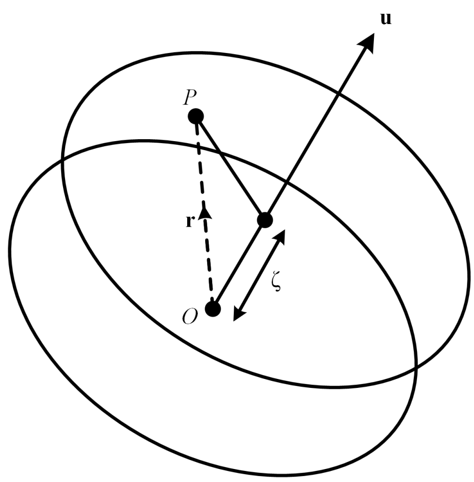

Figure 2.5: Geometry to determine the field at P when a plane wave has electric field $\mathbf{E}_o$ over the plane containing the point O.

(h) Collecting the properties of plane waves given in sections (a) to (g), if a plane wave is propagating in the direction $\mathbf{u}$ in a lossless, uniform isotropic and unbounded medium, and if its vector electric field is $\mathbf{E}_o$ over the plane containing the point O (see Fig. 2.5), then the vector electric field at the point P, whose vector position with respect to the point O is $\mathbf{r}$, will be

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_o \exp(-jk\mathbf{u}\cdot\mathbf{r}) \tag{2.14}$$

since the distance between the planes containing O and P is $\zeta = \mathbf{u}\cdot\mathbf{r}$. The electric field, whatever the polarization, must be orthogonal to $\mathbf{u}$, which is expressed by

$$\mathbf{u}\cdot\mathbf{E}(\mathbf{r}) = 0 \tag{2.15}$$

The vector magnetic field must be orthogonal to both $\mathbf{u}$ and $\mathbf{E}(\mathbf{r})$, and its phasor amplitude is related to that of $\mathbf{E}(\mathbf{r})$ by the characteristic impedance Z of the medium, so that

$$\begin{aligned}\mathbf{H}(\mathbf{r}) &= Z^{-1}\mathbf{u}\times\mathbf{E}(\mathbf{r}) \\ &= Z^{-1}\mathbf{u}\times\mathbf{E}_o \exp(-jk\mathbf{u}\cdot\mathbf{r})\end{aligned} \tag{2.16}$$

The corresponding expression for the power flow in the plane wave is

$$\begin{aligned}\mathbf{S}(\mathbf{r}) = \frac{1}{2}\ \mathrm{Re}\ \mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r}) &= \frac{1}{2}\ \mathrm{Re}\ \mathbf{E}_o \times Z^{-1}(\mathbf{u} \times \mathbf{E}_o{}^*) \\ &= (2Z)^{-1}(\mathbf{E}_o \cdot \mathbf{E}_o^*)\mathbf{u} \\ &= (2Z)^{-1}|E_o|^2\mathbf{u}\end{aligned} \tag{2.17}$$

Some authors prefer to combine the phase constant k and direction $\mathbf{u}$ into the single quantity

$$\mathbf{k} = k\mathbf{u}, \tag{2.18}$$

sometimes known as the **vector wavenumber** of the plane wave. It has the virtue of giving information about the frquency (since $k = \omega(\mu\epsilon)^{1/2}$) and direction of the plane wave in one symbol. The electric field at point P is then

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_o \exp(-j\mathbf{k} \cdot \mathbf{r}) \tag{2.19}$$

and similarly for the magnetic field. This slight simplification in the argument of the exponential term is offset by increased complexity elsewhere: in particular, where $\mathbf{u}$ occurs it must be replaced by $\mathbf{k}/k$. We have preferred to retain the explicit dependence of field expression on the direction $\mathbf{u}$, or something equivalent to it, partly for this reason. But the main reason for our choice is to emphasize the fact that the angular plane-wave spectrum, which we are now in a position to examine in detail, is a function of direction.

2.2 Angular Spectrum for Two-Dimensional Fields

Referring to the Cartesian coordinate system of Fig. 2.6, it will be supposed that the fields are uniform with y, and hence depend on the coordinates (x, z) only. The x-y plane will be taken to be the aperture plane, and our interest will be in the fields diffracted into the supposed uniform, isotropic, source-free and lossless medium to the right of the aperture plane, the half-space $z \geqslant 0$.

To construct the fields in this half-space from a set of plane waves traveling in different directions, it is clear that they must all be traveling parallel to the x-z plane in order that the fields are independent of y. A typical member of the set of plane waves is shown in 2.6, its direction $\mathbf{u}$ making an angle θ to the z axis. Since the sources are to the left of the x-y aperture plane, the plane waves must all travel into the half-space $z \geqslant 0$, which restricts θ to the range

$$-\pi/2 \leqslant \theta \leqslant \pi/2\ , \tag{2.20}$$

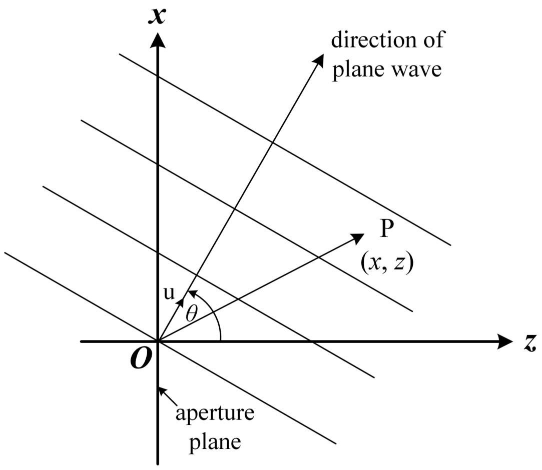

Figure 2.6: Two-dimensional geometry for the field which are uniform with y, which is at right angles to the plane of the figure.

assuming for now that θ is real.

The vector position of the field point P with respect to the origin O is

$$\mathbf{r} = \mathbf{u}_x x + \mathbf{u}_z z \tag{2.21}$$

where $\mathbf{u}_x$ and $\mathbf{u}_z$ are unit vectors in the directions of the axes of x and z. The unit vector in the direction of propagation of the plane wave is

$$\mathbf{u} = \mathbf{u}_x \sin\theta + \mathbf{u}_z \cos\theta \tag{2.22}$$

Abbreviating this by writing

$$s = \sin\theta c \quad = \cos\theta = \sqrt{1 - s^2} \tag{2.23}$$

the direction unit vector becomes

$$\mathbf{u} = \mathbf{u}_x s + \mathbf{u}_z c \tag{2.24}$$

in which s and c are the direction cosines. Hence the projection of OP on the direction of the plane wave is

$$\mathbf{u} \cdot \mathbf{r} = sx + cz \tag{2.25}$$

If the vector electric field, as it passes the origin O, is $\mathbf{E_o}$, the electric field at the point P will be

$$\mathbf{E}(x, z) = \mathbf{E}_o \exp -jk(sx + cz) \tag{2.26}$$

(see equation (2.14)). It will be convenient to resolve the field into two orthogonal linearly polarized plane waves, one with the electric vector parallel to the x-z plane and the other with the electric vector perpendicular to the plane. The first of these linearly polarized plane waves has its vector magnetic field pointing entirely in the transverse y-direction, and will therefore be referred to as the **transverse magnetic** (TM) case. The second has its electric field pointing entirely in the transverse direction, and will be referred to as the **transverse electric** (TE) case. We will treat the TM case first and the TE case, which is simply its dual, later.

2.2.1 Angular spectrum of transverse magnetic (TM) fields

All the plane waves in the angular spectrum which represents a transverse magnetic field in two dimensions will have their electric field lying in the plane of propagation, as depicted in Fig. 2.7. The magnetic field will then be wholly transverse. If E_o is the phasor amplitude of the electric field of

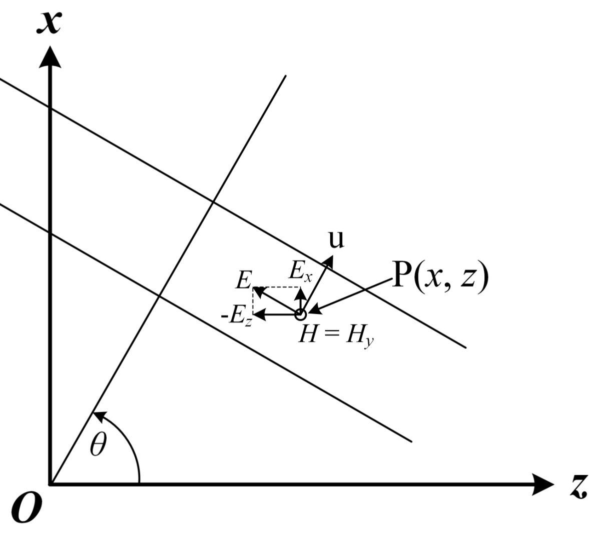

Figure 2.7: Showing the field components at point $\mathrm{P}(x, z)$ for a plane wave whose magnetic field is wholly transverse to the x-z plane.

the plane waves as it passes the origin, then the Cartesian components of the electromagnetic field at $\mathrm{P}(x, z)$ will be

$$\begin{aligned} E_x(x,z) &= E_o \cos\theta \exp -jk(sx+cz) \\ E_z(x,z) &= -E_o \sin\theta \exp -jk(sx+cz) \\ H_y(x,z) &= Z^{-1} E_o \cos\theta \exp -jk(sx+cz) \end{aligned} \tag{2.27}$$

with the remaining component E_y, H_x and H_z all zero. Z is the characteristic impedance and k the phase constant of the medium.

A set of plane waves traveling in different directions is most conveniently represented by a spectrum function $A(\theta)$, such that the electric-field amplitude of the plane wave traveling in the direction θ is $A(\theta)\mathrm{d}\theta$. This spectrum function is completely analogous to the frequency spectrum function of time-series analysis, and like the frequency spectrum the angular spectrum $A(\theta)$ can be continuous, discrete, or a mixture of the two. The field components are obtained by replacing E_o by $A(\theta)\mathrm{d}\theta$ in the above equation, and then integrating over the range of angles for which the angular spectrum is defined. For example the x-component of the electric field will be

$$E_x(x,z) = \int A(\theta) \cos\theta \exp\{-jk(sx+cz)\}\mathrm{d}\theta \tag{2.28}$$

For the reason which will be explained in the next section, it is simpler to express the angular dependence of the spectrum in terms of $s = \sin\theta$, rather than in the terms of θ itself. So replacing $A(\theta)$ by $F(s)$, and noting that $\mathrm{d}s = \cos\theta \mathrm{d}\theta$, equation (2.28) becomes

$$E_x(x,z) = \int F(s) \exp\{-jk(sx+cz)\}\mathrm{d}s \tag{2.29}$$

which gives the x-component of the electric field at the point (x, z) as the integrated effect of all the plane waves in the angular spectrum $F(s)$. Note that $c = \cos\theta$ is retained as a convenient abbreviation for $\sqrt{1-s^2}$.

According to relation (2.20) the range of integration for s would be ± 1. However, it will be shown in the next section that for completeness the s integration has to be extend to cover the whole real line, and hence that the limits of integration for s are $\pm\infty$.

The field components at the point (x, z) for a two-dimensional transverse magnetic field, in terms of the angular plane-wave spectrum function $F(s)$, are therefore

$$\begin{bmatrix} E_x(x,z) \\ E_z(x,z) \\ H_y(x,z) \end{bmatrix} = \int_{-\infty}^{\infty} F(s) \begin{bmatrix} 1 \\ -s/c \\ (Zc)^{-1} \end{bmatrix} \exp\{-jk(sx+cz)\}\mathrm{d}s \tag{2.30}$$

in which $c = \sqrt{1 - s^2}$. The important point to note at this stage is that the fields anywhere on and to the right of the aperture plane, that is, in the half-space $z \geqslant 0$, have been expressed in the TM case in terms of the single spectrum function $F(s)$. It will be shown later that the remaining field components, which are transverse electric, can be represented by a second spectrum function. But before doing so we will examine some of the important features of the representation of fields by an angular spectrum of plane waves by looking at equation (2.30) in rather more detail.

Exercise

Show by substituting into equation (2.30) that the discrete angular plane-wave spectrum $F(s) = E_o c_o \delta(s - s_o)$, where $s_o = \sin\theta_o$, $c_o = \cos\theta_o$ and $\delta()$ is the Dirac delta function, represent a plane wave of amplitude E_o traveling in the direction $\theta = \theta_o$.

Exercise

Find the field component of the two-dimensional TM field whose angular spectrum is

$$F(s) = \frac{E_o C_o}{2}[\delta(s - s_o) + s(s + s_o)]$$

which is in fact the interference pattern of two inclined plane waves of equal amplitude. Sketch the planes of constant amplitude and constant phase. Show, for any of the three field components, that adjacent planes over which its amplitude is zero are separated by a distance $\lambda/(2s_o)$, and that planes of constant phase between which the phase differs by 2π are separated by λ/c_o, where $c_o = (1 - s_o^2)^{1/2}$.

Exercise

The composite field examined in the previous exercise is of a type known as an **inhomogeneous plane wave**, since amplitude and phase are both constant over non-coincident planes. Deduce the propagation constant and wave impedance (defined as the ratio of the orthogonal electric and magnetic field transverse to the direction of propagation) of this inhomogeneous plane wave.

Exercise

Determine where, in the fields examined in the previous two exercises, two infinity thin, perfectly conducting, parallel planes may be introduced without disturbing the fields. (See Appendix A.10, for a fuller discussion, from the point of view of TM modes in a parallel-plate waveguide.)

2.3 Evanescent Waves

In order to get a better idea of the physical meaning of the plane-wave spectrum representation of the two-dimensional TM field of equation (2.30), consider the elemental contribution to the x-component of the electric field

$$\mathrm{d}E_x(x,z) = F(s)\mathrm{d}s\exp\{-jk(sx+cz)\} \tag{2.31}$$

which is the contribution of that plane wave in the spectrum of amplitude $F(s)\mathrm{d}s$ traveling in the direction making an angle $\theta = sin^{-1}s$ to the z axis.

When $|s| \leqslant 1$, θ lies in the range $-\pi/2 \leqslant \theta \leqslant \pi/2$, and the elemental plane wave of equation (2.31) is of the homogeneous type described in the section 2.1. The wave travels with characteristic speed of the medium $(\mu\epsilon)^{-1/2}$, and transfers power into the half-space $z \geqslant 0$.

But when $|s| > 1$, assuming s still to be real, the character of the wave changes because the cosine

$$c = \sqrt{1-s^2} = \pm j\chi \ (\chi \text{ real and positive}) \tag{2.32}$$

is now imaginary. Substituting this into equation (2.31), and noting that we must choose $c = -j\chi$ in order that the fields remain finite as $z \to \pm\infty$,

$$\mathrm{d}E_x(x,z) = F(s)\mathrm{d}s\exp(-jksx)\exp(-k\chi z), \tag{2.33}$$

This is a plane wave of inhomogeneous type, in that the amplitude is no longer constant over planes of constant phase. It has the following properties:

(a) The direction of propagation of the wave is along the x-axis, that is, parallel to the aperture plane, either positive or negative depending on the sign of s. The wavefronts over which the phase is constant are shown in Fig. 2.8.

(b) The amplitudes of the field components decrease exponentially in the $+z$ direction, away from the aperture plane. For this reason they are often called evanescent, which means disappearing. The distance from the aperture plane at which the amplitude of the elemental plane wave of equation (2.33) has decreased to a fraction e^{-1} (=0.3679) of its value at $z = 0$ is

$$z = \frac{\lambda}{2\pi\chi}\ , \tag{2.34}$$

where χ is real and positive. For all but the smallest values of χ evanescent waves will have become negligibly small at distance of more than a few wavelengths from the aperture plane.

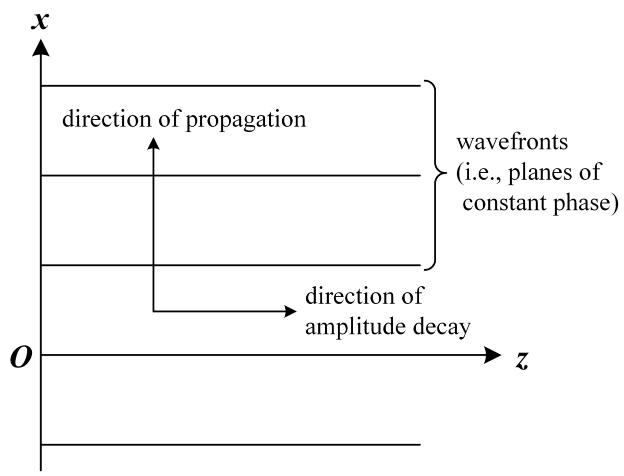

Figure 2.8: When $|s| > 1$ the plane waves of the angular spectrum $F(s)$ are inhomogeneous.

(c) The phase velocity of the evanescent wave, obtained by restoring the time dependence in equation (2.33), is

$$v_p = \frac{1}{s(\mu\epsilon)^{1/2}} \tag{2.35}$$

This velocity is slower than that of the homogeneous plane waves propagating in the medium, since $|s| > 1$. These slow waves do carry power, but it is not propagated into the region $z \geqslant 0$. It merely travels back and forth in the aperture plane, and can be thought of as being stored there. Because of their close association with a particular surface, in this case the aperture plane, these waves are also known as electromagnetic surface waves.

Exercise
Use the Poynting vector of equation (2.10) to determine the power flow in the elemental wave defined by equation (2.31) and its associated field components. Compare the two case when $|s| < 1$ and $|s| > 1$.

Exercise
Define an aperture wave impedance for the elemental wave of equation (2.31) etc. as

$$Z_{ap} = \frac{\mathrm{d}E_x}{\mathrm{d}H_y} \tag{2.36}$$

which is the ratio of the transverse orthogonal components of the electric and magnetic field 'looking out' into the half-space $z \geqslant 0$. Show that Z_{ap} is imaginary when $|s| > 1$, and comment on the significance of this result.

Exercise
Apply the Poynting vector of equation (2.10) directly to the fields of equation (2.30) and show that the total power radiated into the half-space $z \geqslant 0$ is

$$P_{rad} = \frac{\lambda}{2Z} \int_{-1}^{+1} c^{-1} |F(s)|^2 \mathrm{d}s \tag{2.37}$$

It may be useful in achieving this result to note that

$$\int_{-\infty}^{\infty} \exp\{\pm jkx(s - s')\} \mathrm{d}x = \lambda \delta(s - s') \tag{2.38}$$

in which $k = 2\pi/\lambda$ and $\delta()$ is the Dirac delta function.

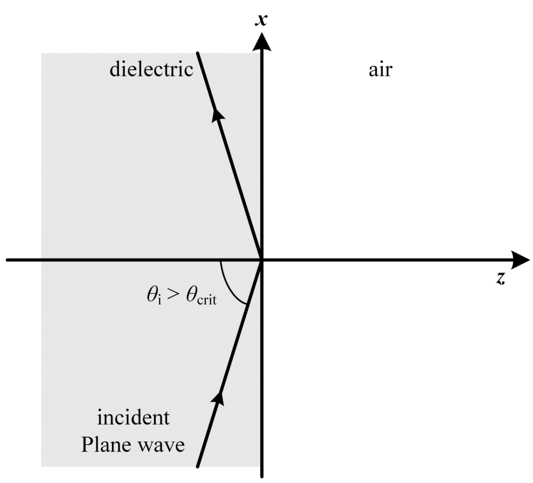

Figure 2.9: A plane wave incident at a dielectric/air boundary at an angle of incident greater than critical.

Example
A very instructive physical example of the type of wave we have been discussing occurs on the far side of a dielectric boundary at which total internal reflecting occurs. (see Fig. 2.9 and Appendix A.9). In this example the aperture plane will be taken to be coincident with the boundary between the dielectric and air. With a plane wave incident from the dielectric side at an

angle of incidence θ_i greater than the critical angle θ_{crit} (for which the plane wave transmitted into the air medium would travel in a direction just parallel to the aperture plane) the incident wave will be totally reflected. However, the fields to the right of the aperture plane cannot be zero, since the boundary conditions (see Appendix B.2.3) require that the tangential components of **E** and **H** be continuous across it. In fact the fields on the air side of the boundary are evanescent waves, which are local to the boundary and carry no energy across it. Thus writing Snell's law as

$$s = \sin\theta_t = (\epsilon_r)^{1/2}\sin\theta_i \tag{2.39}$$

in which θ_t is the angle of transmission corresponding to the angle of incidence θ_i, and ϵ_r is the relative permittivity of the dielectric, it is clear that when $\theta_i > \theta_{crit}(=\sin^{-1}(\epsilon_r)^{-1/2})$ s becomes greater than unity but remains real. These are precisely the conditions for which we deduced the properties of evanescent waves.

Mathematical Comment What is the nature of θ when the magnitude of $\sin\theta$ is greater than unity? The answer is that θ is in general a complex angle. If we apply the two conditions:

(I) s is real and in the rage $-\infty < s < \infty$

(II)

$$\text{when } |s| > 1, c = \sqrt{1-s^2} = -j\chi \text{ is negative imaginary,that is, } \chi \text{ is real and } > 0, \tag{2.40}$$

we can deduce that the contour Γ of θ in the complex-θ plane, corresponding to s traversing the real line from $-\infty$ to $+\infty$, is that given in Fig. 2.10. That par of Γ for which θ is real, namely $-\pi/2 \leqslant \theta \leqslant \pi/2$, corresponds to $-1 \leqslant s \leqslant 1$. The remainder of the contour Γ goes off to infinity in directions which ensure that the fields in the angular plane-wave spectrum representation of equation (2.30) are bounded as $z \longrightarrow \infty$.

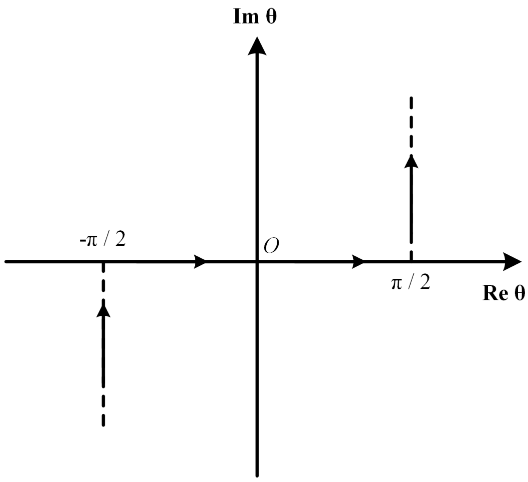

Figure 2.10: Representation of θ in the complex-θ plane: contour Γ corresponds to $s = \sin\theta$ tranversing the real line from $-\infty$ to $+\infty$.

2.4 Aperture Field: Fourier Transform of the Angular Spectrum

2.5 Far Field: Approximated by the Angular Spectrum

2.6 The Complete Two-Dimensional Field: Uniqueness

Chapter 3

Three-Dimensional Fields

3.1 Angular Spectrum for Linearly Polarized Aperture Fields

3.2 Aperture Field: Fourier Transform of the Angular Spectrum

3.2.1 Separable aperture fields

3.2.2 Circularly symmetric aperture fields

3.3 Evanescent Waves and Alternative Representations

3.3.1 Wavenumber-component representation

3.3.2 Spherical-polar angle representation

3.4 Far Field: Approximated by the Angular Spectrum

3.4.1 Stationary-phase evaluation of double integrals

3.4.2 Far field for y-polarized aperture fields

3.4.3 Representation for an arbitrary waveform

3.5 Complete Representation of Three-Dimensional Radiating Fields

3.5.1 Fields in a half space

3.5.2 Huygen's principle

3.5.3 Fields in the whole space

Chapter 4

Transmitting and Receiving Antennas

4.1 The Antenna as Transmitter

4.1.1 Gain function and the angular spectrum

4.1.2 Gain of a uniformly illuminated aperture

4.1.3 Polarization and gain function

4.2 The Antenna as Receive

4.2.1 Antenna reciprocity theorem

4.2.2 Relation between effective receiving area and gain function

4.3 General Antenna Coupling Formula

4.3.1 Near-field antenna measurements

4.3.2 Aperture-field form of the antenna coupling formula

Chapter 5

Fresnel Diffraction

5.1 Fresnel's Diffraction Formula

5.2 Diffraction by a Conducting Half Plane

5.3 Transition to the Far Field

5.4 Detailed Structure of the Radiated Field

5.5 Fresnel Diffraction in Three Dimensions

5.5.1 Fresnel's diffraction formula

5.5.2 Diffraction by a conducting half plane

5.5.3 Transition to the far field

5.5.4 Fresnel zones

Chapter 6

Reflection from Perfectly Conducting Surfaces

6.1 Reflection by an Infinite Plane

6.2 Reflection from Flat, Finite Surface

6.3 Reflection from Curved Surfaces – by an Approximate Method

6.4 Reflection from Curved Surfaces – by a More Precise Method

6.4.1 The method of reflecting Huygens elements

6.4.2 Application to a doubly curved surface

6.4.3 An apex-illuminated convex paraboloidal reflector

Chapter 7

Analysis of Some Aperture Antenna

7.1 Electromagnetic Horns

7.1.1 Pyramidal Horn

7.1.2 Other types of horns

7.2 Paraboloid Reflector Antennas

7.2.1 Geometry of the Paraboloid

7.2.2 Radiation from an axisymmetrically front-fed paraboloid by projected aperture-field method

7.2.3 Validity of the projected aperture-field method

7.2.4 Far field calculated by a surface integration method

7.2.5 Fields in the focal plane of a paraboloid reflector axially illuminated by a plane wave

7.2.6 Aperture blocking

7.2.7 Random reflector-profile effects

7.3 Cassegrain Reflector Antennas

7.4 Offset Paraboloid-Reflector Antennas

7.5 The Horn-Paraboloid Antenna

Chapter 8

Radiation From Non-Planar Apertures

8.1 Derivation of the Near-field to Far-field Formula

8.2 The Kichhoff-Huygens Formula

8.3 Relationship between Formulas for Scattering and Radiation

Appendix A

Transmission lines, plane waves, and simple waveguides

Since

A.1 Plane Waves Derived From The Parallel-plate Transmission Line

Elementary

A.2 Properties of Plane waves

A.3 Power Flow in a Plane Wave

A.4 General Expression for a Plane Wave

A.5 Polarization

A.6 Propagation in a Lossy Medium

A.7 Reflection: Normal Incidence

A.8 Reflection and Refraction

A.9 Total Internal Reflection

A.10 Waveguides

A.11 Concluding remark

Appendix B

Summary of Maxwell's Equations and some of theire Consequences

Since

B.1 Maxwell's Equations

Elementary

B.1.1 Electric Field Divergence of Electric Flux Density

B.1.2 Magnetic Field Divergence of Magnetic Flux Density

B.1.3 Time Variation of Magnetic Field

B.1.4 Time Variation of Electic Field

B.2 Some Consequences and Particular Forms of Maxwell's Equations

B.2.1 Linearity Superposition of Solutions

B.2.2 Poynting's Theorem Flow of Electromagnetic Energy

B.2.3 Boundary Conditions

B.2.4 Maxwell's Equations in Source-Free, Lossless Media

B.2.5 Sinusoidal Time Variation Average Power flow

B.2.6 Lossy Media Complex Permittivity Representation

B.2.7 Duality of Solutions

B.2.8 Reciprocity Theorems

B.2.9 Uniqueness Theorem