18

Radiation from Apertures

18.1 Field Equivalence Principle

The radiation fields from aperture antennas, such as slots, open-ended waveguides, horns, reflector and lens antennas, are determined from the knowledge of the fields *over* the aperture of the antenna.

The aperture fields become the *sources* of the radiated fields at large distances. This is a variation of the Huygens-Fresnel principle, which states that the points on each wavefront become the sources of secondary spherical waves propagating outwards and whose superposition generates the next wavefront.

Let $\boldsymbol{E}_a, \boldsymbol{H}_a$ be the *tangential fields* over an aperture A, as shown in Fig. 18.1.1. These fields are assumed to be known and are produced by the sources to the left of the screen. The problem is to determine the radiated fields $\boldsymbol{E}(\boldsymbol{r}), \boldsymbol{H}(\boldsymbol{r})$ at some far observation point.

The radiated fields can be computed with the help of the *field equivalence principle* [1288–1295,1685], which states that the aperture fields may be replaced by *equivalent electric and magnetic surface currents*, whose radiated fields can then be calculated using the techniques of Sec. 15.10. The equivalent surface currents are:

$$\boxed{\begin{aligned} \boldsymbol{J}_s &= \hat{\mathbf{n}} \times \boldsymbol{H}_a \qquad \text{(electric surface current)} \\ \boldsymbol{J}_{ms} &= -\hat{\mathbf{n}} \times \boldsymbol{E}_a \qquad \text{(magnetic surface current)} \end{aligned}} \tag{18.1.1}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the surface and on the side of the radiated fields.

Thus, it becomes necessary to consider Maxwell's equations in the presence of magnetic currents and derive the radiation fields from such currents.

The screen in Fig. 18.1.1 is an arbitrary infinite surface over which the tangential fields are assumed to be zero. This assumption is not necessarily consistent with the radiated field solutions, that is, Eqs. (18.4.9). A consistent calculation of the fields to the right of the aperture plane requires knowledge of the fields over the entire aperture plane (screen plus aperture.)

However, for large apertures (with typical dimension much greater than a wavelength), the approximation of using the fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ only over the aperture to calculate the radiation patterns is fairly adequate, especially in predicting the main-lobe behavior of the patterns.

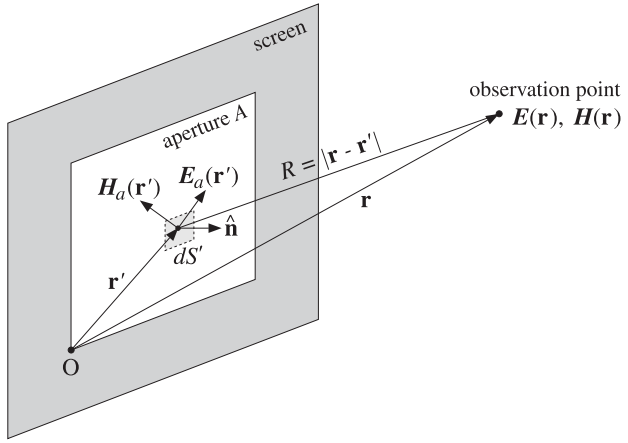


Fig. 18.1.1 Radiated fields from an aperture.

The screen can also be a perfectly conducting surface, such as a ground plane, on which the aperture opening has been cut. In reflector antennas, the aperture itself is not an opening, but rather a reflecting surface. Fig. 18.1.2 depicts some examples of screens and apertures: (a) an open-ended waveguide over an infinite ground plane, (b) an open-ended waveguide radiating into free space, and (c) a reflector antenna.

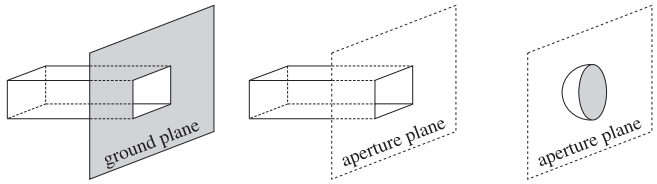


Fig. 18.1.2 Examples of aperture planes.

There are two alternative forms of the field equivalence principle, which may be used when only one of the aperture fields $\boldsymbol{E}_a$ or $\boldsymbol{H}_a$ is available. They are:

$$\boxed{\begin{aligned} \boldsymbol{J}_s &= 0 \\ \boldsymbol{J}_{ms} &= -2(\hat{\mathbf{n}} \times \boldsymbol{E}_a) \end{aligned}} \qquad \text{(perfect electric conductor)} \tag{18.1.2}$$

$$\boxed{\begin{aligned} \boldsymbol{J}_s &= 2(\hat{\mathbf{n}} \times \boldsymbol{H}_a) \\ \boldsymbol{J}_{ms} &= 0 \end{aligned}} \qquad \text{(perfect magnetic conductor)} \tag{18.1.3}$$

They are appropriate when the screen is a *perfect electric conductor* (PEC) on which $\boldsymbol{E}_a = 0$, or when it is a *perfect magnetic conductor* (PMC) on which $\boldsymbol{H}_a = 0$. On the aperture, both $\boldsymbol{E}_a$ and $\boldsymbol{H}_a$ are non-zero.

Using image theory, the perfect electric (magnetic) conducting screen can be eliminated and replaced by an image magnetic (electric) surface current, doubling its value over the aperture. The image field causes the total tangential electric (magnetic) field to vanish over the screen.

If the tangential fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ were known over the entire aperture plane (screen plus aperture), the three versions of the equivalence principle would generate the same radiated fields. But because we consider $\boldsymbol{E}_a, \boldsymbol{H}_a$ only over the aperture, the three versions give slightly different results.

In the case of a perfectly conducting screen, the calculated radiation fields (18.4.10) using the equivalent currents (18.1.2) are consistent with the boundary conditions on the screen.

We provide a justification of the field equivalence principle (18.1.1) in Sec. 18.10 using vector diffraction theory and the Stratton-Chu and Kottler formulas. The modified forms (18.1.2) and (18.1.3) are justified in Sec. 19.2 where we derive them in two ways: one, using the plane-wave-spectrum representation, and two, using the Franz formulas in conjunction with the extinction theorem discussed in Sec. 18.11, and discuss also their relationship to Rayleigh-Sommerfeld diffraction theory of Sec. 19.1.

18.2 Magnetic Currents and Duality

Next, we consider the solution of Maxwell's equations driven by the ordinary electric charge and current densities $\rho, \boldsymbol{J}$, and in addition, by the magnetic charge and current densities $\rho_m, \boldsymbol{J}_m$.

Although $\rho_m, \boldsymbol{J}_m$ are fictitious, the solution of this problem will allow us to identify the equivalent magnetic currents to be used in aperture problems, and thus, establish the field equivalence principle. The generalized form of Maxwell's equations is:

$$\boxed{\begin{aligned} \nabla\times\boldsymbol{H} &= \boldsymbol{J} + j\omega\epsilon\boldsymbol{E} \\ \nabla\cdot\boldsymbol{E} &= \frac{1}{\epsilon}\rho \\ \nabla\times\boldsymbol{E} &= -\boldsymbol{J}_m - j\omega\mu\boldsymbol{H} \\ \nabla\cdot\boldsymbol{H} &= \frac{1}{\mu}\rho_m \end{aligned}} \tag{18.2.1}$$

There is now complete symmetry, or duality, between the electric and the magnetic quantities. In fact, it can be verified easily that the following *duality transformation* leaves the set of four equations *invariant*:

$$\boxed{\begin{array}{lll} \boldsymbol{E} \longrightarrow \boldsymbol{H} & \boldsymbol{J} \longrightarrow \boldsymbol{J}_m & \boldsymbol{A} \longrightarrow \boldsymbol{A}_m \\ \boldsymbol{H} \longrightarrow -\boldsymbol{E} & \rho \longrightarrow \rho_m & \varphi \longrightarrow \varphi_m \\ \epsilon \longrightarrow \mu & \boldsymbol{J}_m \longrightarrow -\boldsymbol{J} & \boldsymbol{A}_m \longrightarrow -\boldsymbol{A} \\ \mu \longrightarrow \epsilon & \rho_m \longrightarrow -\rho & \varphi_m \longrightarrow -\varphi \end{array}} \quad \text{(duality)} \tag{18.2.2}$$

where $\varphi, \boldsymbol{A}$ and $\varphi_m, \boldsymbol{A}_m$ are the corresponding scalar and vector potentials introduced below. These transformations can be recognized as a special case (for $\alpha = \pi/2$) of the following *duality rotations*, which also leave Maxwell's equations invariant:

$$\begin{bmatrix} \boldsymbol{E}' & \eta\boldsymbol{J}' & \eta\rho' \\ \eta\boldsymbol{H}' & \boldsymbol{J}'_m & \rho'_m \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \boldsymbol{E} & \eta\boldsymbol{J} & \eta\rho \\ \eta\boldsymbol{H} & \boldsymbol{J}_m & \rho_m \end{bmatrix} \tag{18.2.3}$$

Under the duality transformations (18.2.2), the first two of Eqs. (18.2.1) transform into the last two, and conversely, the last two transform into the first two.

A useful consequence of duality is that if one has obtained expressions for the electric field $\boldsymbol{E}$, then by applying a duality transformation one can generate expressions for the magnetic field $\boldsymbol{H}$. We will see examples of this property shortly.

The solution of Eq. (18.2.1) is obtained in terms of the usual scalar and vector potentials $\varphi, \boldsymbol{A}$, as well as two new potentials $\varphi_m, \boldsymbol{A}_m$ of the magnetic type:

$$\begin{aligned} \boldsymbol{E} &= -\nabla\varphi - j\omega\boldsymbol{A} - \frac{1}{\epsilon}\nabla\times\boldsymbol{A}_m \\ \boldsymbol{H} &= -\nabla\varphi_m - j\omega\boldsymbol{A}_m + \frac{1}{\mu}\nabla\times\boldsymbol{A} \end{aligned} \tag{18.2.4}$$

The expression for $\boldsymbol{H}$ can be derived from that of $\boldsymbol{E}$ by a duality transformation of the form (18.2.2). The scalar and vector potentials satisfy the Lorenz conditions and Helmholtz wave equations:

$$\begin{aligned} &\nabla\cdot\boldsymbol{A} + j\omega\epsilon\mu\,\varphi = 0 && \nabla\cdot\boldsymbol{A}_m + j\omega\epsilon\mu\,\varphi_m = 0 \\ &\nabla^2\varphi + k^2\varphi = -\frac{\rho}{\epsilon} \quad \text{and} && \nabla^2\varphi_m + k^2\varphi_m = -\frac{\rho_m}{\mu} \\ &\nabla^2\boldsymbol{A} + k^2\boldsymbol{A} = -\mu\boldsymbol{J} && \nabla^2\boldsymbol{A}_m + k^2\boldsymbol{A}_m = -\epsilon\boldsymbol{J}_m \end{aligned} \tag{18.2.5}$$

The solutions of the Helmholtz equations are given in terms of $G(\boldsymbol{r}-\boldsymbol{r}') = \dfrac{e^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|}$:

$$\begin{aligned} \varphi(\boldsymbol{r}) &= \int_V \frac{1}{\epsilon}\rho(\boldsymbol{r}')G(\boldsymbol{r}-\boldsymbol{r}')\,dV', \qquad \varphi_m(\boldsymbol{r}) = \int_V \frac{1}{\mu}\rho_m(\boldsymbol{r}')G(\boldsymbol{r}-\boldsymbol{r}')\,dV' \\ \boldsymbol{A}(\boldsymbol{r}) &= \int_V \mu\boldsymbol{J}(\boldsymbol{r}')G(\boldsymbol{r}-\boldsymbol{r}')\,dV', \qquad \boldsymbol{A}_m(\boldsymbol{r}) = \int_V \epsilon\boldsymbol{J}_m(\boldsymbol{r}')G(\boldsymbol{r}-\boldsymbol{r}')\,dV' \end{aligned} \tag{18.2.6}$$

where V is the volume over which the charge and current densities are nonzero. The observation point $\boldsymbol{r}$ is taken to be outside this volume. Using the Lorenz conditions, the scalar potentials may be eliminated in favor of the vector potentials, resulting in the alternative expressions for Eq. (18.2.4):

$$\boxed{\begin{aligned} \boldsymbol{E} &= \frac{1}{j\omega\mu\epsilon}\left[\nabla(\nabla\cdot\boldsymbol{A}) + k^2\boldsymbol{A}\right] - \frac{1}{\epsilon}\nabla\times\boldsymbol{A}_m \\ \boldsymbol{H} &= \frac{1}{j\omega\mu\epsilon}\left[\nabla(\nabla\cdot\boldsymbol{A}_m) + k^2\boldsymbol{A}_m\right] + \frac{1}{\mu}\nabla\times\boldsymbol{A} \end{aligned}} \tag{18.2.7}$$

These may also be written in the form of Eq. (15.3.9):

$$\boxed{\begin{aligned} \boldsymbol{E} &= \frac{1}{j\omega\mu\epsilon}\left[\nabla\times(\nabla\times\boldsymbol{A}) - \mu\boldsymbol{J}\right] - \frac{1}{\epsilon}\nabla\times\boldsymbol{A}_m \\ \boldsymbol{H} &= \frac{1}{j\omega\mu\epsilon}\left[\nabla\times(\nabla\times\boldsymbol{A}_m) - \epsilon\boldsymbol{J}_m\right] + \frac{1}{\mu}\nabla\times\boldsymbol{A} \end{aligned}} \tag{18.2.8}$$

Replacing $\boldsymbol{A}, \boldsymbol{A}_m$ in terms of Eq. (18.2.6), we may express the solutions (18.2.7) directly in terms of the current densities:

$$\begin{aligned}\boldsymbol{E} &= \frac{1}{j\omega\epsilon}\int_V [k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G]\,dV' \\ \boldsymbol{H} &= \frac{1}{j\omega\mu}\int_V [k^2\boldsymbol{J}_m G + (\boldsymbol{J}_m\cdot\nabla')\nabla' G + j\omega\mu\,\boldsymbol{J}\times\nabla' G]\,dV'\end{aligned} \tag{18.2.9}$$

Alternatively, if we also use the charge densities, we obtain from (18.2.4):

$$\begin{aligned}\boldsymbol{E} &= \int_V [-j\omega\mu\,\boldsymbol{J}G + \frac{\rho}{\epsilon}\nabla' G - \boldsymbol{J}_m\times\nabla' G]\,dV' \\ \boldsymbol{H} &= \int_V [-j\omega\epsilon\,\boldsymbol{J}_m G + \frac{\rho_m}{\mu}\nabla' G + \boldsymbol{J}\times\nabla' G]\,dV'\end{aligned} \tag{18.2.10}$$

18.3 Radiation Fields from Magnetic Currents

The radiation fields of the solutions (18.2.7) can be obtained by making the far-field approximation, which consists of the replacements:

$$G(\boldsymbol{r}-\boldsymbol{r}') = \frac{e^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} \simeq \frac{e^{-jkr}}{4\pi r}e^{j\boldsymbol{k}\cdot\boldsymbol{r}'} \quad\text{and}\quad \nabla \simeq -j\boldsymbol{k} \tag{18.3.1}$$

where $\boldsymbol{k} = k\hat{\mathbf{r}}$. Then, the vector potentials of Eq. (18.2.6) take the simplified form:

$$\boldsymbol{A}(\boldsymbol{r}) = \mu\frac{e^{-jkr}}{4\pi r}\boldsymbol{F}(\theta,\phi)\,, \qquad \boldsymbol{A}_m(\boldsymbol{r}) = \epsilon\frac{e^{-jkr}}{4\pi r}\boldsymbol{F}_m(\theta,\phi) \tag{18.3.2}$$

where the radiation vectors are the Fourier transforms of the current densities:

$$\boxed{\begin{aligned}\boldsymbol{F}(\theta,\phi) &= \int_V \boldsymbol{J}(\boldsymbol{r}')e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dV' \\ \boldsymbol{F}_m(\theta,\phi) &= \int_V \boldsymbol{J}_m(\boldsymbol{r}')e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dV'\end{aligned}} \quad\text{(radiation vectors)} \tag{18.3.3}$$

Setting $\boldsymbol{J} = \boldsymbol{J}_m = 0$ in Eq. (18.2.8) because we are evaluating the fields far from the current sources, and using the approximation $\nabla = -j\boldsymbol{k} = -jk\hat{\mathbf{r}}$, and the relationship $k/\epsilon = \omega\eta$, we find the radiated $\boldsymbol{E}$ and $\boldsymbol{H}$ fields:

$$\boxed{\begin{aligned}\boldsymbol{E} &= -j\omega[\hat{\mathbf{r}}\times(\boldsymbol{A}\times\hat{\mathbf{r}}) - \eta\,\hat{\mathbf{r}}\times\boldsymbol{A}_m] = -jk\frac{e^{-jkr}}{4\pi r}\hat{\mathbf{r}}\times[\eta\boldsymbol{F}\times\hat{\mathbf{r}} - \boldsymbol{F}_m] \\ \boldsymbol{H} &= -\frac{j\omega}{\eta}[\eta\,\hat{\mathbf{r}}\times(\boldsymbol{A}_m\times\hat{\mathbf{r}}) + \hat{\mathbf{r}}\times\boldsymbol{A}] = -\frac{jk}{\eta}\frac{e^{-jkr}}{4\pi r}\hat{\mathbf{r}}\times[\eta\boldsymbol{F} + \boldsymbol{F}_m\times\hat{\mathbf{r}}]\end{aligned}} \tag{18.3.4}$$

These generalize Eq. (15.10.2) to magnetic currents. As in Eq. (15.10.3), we have:

$$\boldsymbol{H} = \frac{1}{\eta}\hat{\mathbf{r}}\times\boldsymbol{E} \tag{18.3.5}$$

Noting that $\hat{\mathbf{r}}\times(\boldsymbol{F}\times\hat{\mathbf{r}}) = \hat{\boldsymbol{\theta}}F_\theta + \hat{\boldsymbol{\phi}}F_\phi$ and $\hat{\mathbf{r}}\times\boldsymbol{F} = \hat{\boldsymbol{\phi}}F_\theta - \hat{\boldsymbol{\theta}}F_\phi$, and similarly for $\boldsymbol{F}_m$, we find for the polar components of Eq. (18.3.4):

$$\boxed{\begin{aligned}\boldsymbol{E} &= -jk\frac{e^{-jkr}}{4\pi r}[\hat{\boldsymbol{\theta}}(\eta F_\theta + F_{m\phi}) + \hat{\boldsymbol{\phi}}(\eta F_\phi - F_{m\theta})] \\ \boldsymbol{H} &= -\frac{jk}{\eta}\frac{e^{-jkr}}{4\pi r}[-\hat{\boldsymbol{\theta}}(\eta F_\phi - F_{m\theta}) + \hat{\boldsymbol{\phi}}(\eta F_\theta + F_{m\phi})]\end{aligned}} \tag{18.3.6}$$

The Poynting vector is given by the generalization of Eq. (16.1.1):

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\mathrm{Re}(\boldsymbol{E}\times\boldsymbol{H}^*) = \hat{\mathbf{r}}\frac{k^2}{32\pi^2\eta r^2}[|\eta F_\theta + F_{m\phi}|^2 + |\eta F_\phi - F_{m\theta}|^2] = \hat{\mathbf{r}}\,\mathcal{P}_r \tag{18.3.7}$$

and the radiation intensity:

$$U(\theta,\phi) = \frac{dP}{d\Omega} = r^2\mathcal{P}_r = \frac{k^2}{32\pi^2\eta}[|\eta F_\theta + F_{m\phi}|^2 + |\eta F_\phi - F_{m\theta}|^2] \tag{18.3.8}$$

18.4 Radiation Fields from Apertures

For an aperture antenna with effective surface currents given by Eq. (18.1.1), the volume integrations in Eq. (18.2.9) reduce to surface integrations over the aperture A:

$$\begin{aligned}\boldsymbol{E} &= \frac{1}{j\omega\epsilon}\int_A [(\boldsymbol{J}_s\cdot\nabla')\nabla' G + k^2\boldsymbol{J}_s G - j\omega\epsilon\,\boldsymbol{J}_{ms}\times\nabla' G]\,dS' \\ \boldsymbol{H} &= \frac{1}{j\omega\mu}\int_A [(\boldsymbol{J}_{ms}\cdot\nabla')\nabla' G + k^2\boldsymbol{J}_{ms}G + j\omega\mu\,\boldsymbol{J}_s\times\nabla' G]\,dS'\end{aligned} \tag{18.4.1}$$

and, explicitly in terms of the aperture fields shown in Fig. 18.1.1:

$$\boxed{\begin{aligned}\boldsymbol{E} &= \frac{1}{j\omega\epsilon}\int_A [(\hat{\mathbf{n}}\times\boldsymbol{H}_a)\cdot\nabla'(\nabla' G) + k^2(\hat{\mathbf{n}}\times\boldsymbol{H}_a)G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E}_a)\times\nabla' G]\,dS' \\ \boldsymbol{H} &= \frac{1}{j\omega\mu}\int_A [-(\hat{\mathbf{n}}\times\boldsymbol{E}_a)\cdot\nabla'(\nabla' G) - k^2(\hat{\mathbf{n}}\times\boldsymbol{E}_a)G + j\omega\mu(\hat{\mathbf{n}}\times\boldsymbol{H}_a)\times\nabla' G]\,dS'\end{aligned}} \tag{18.4.2}$$

These are known as *Kottler's formulas* [1293–1298,1287,1299–1302,1324]. We derive them in Sec. 18.12. The equation for $\boldsymbol{H}$ can also be obtained from that of $\boldsymbol{E}$ by the application of a duality transformation, that is, $\boldsymbol{E}_a \to \boldsymbol{H}_a$, $\boldsymbol{H}_a \to -\boldsymbol{E}_a$ and $\epsilon \to \mu$, $\mu \to \epsilon$.

In the far-field limit, the radiation fields are still given by Eq. (18.3.6), but now the radiation vectors are given by the two-dimensional Fourier transform-like integrals over the aperture:

$$\boxed{\begin{aligned}\boldsymbol{F}(\theta,\phi) &= \int_A \boldsymbol{J}_s(\boldsymbol{r}')e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dS' = \int_A \hat{\mathbf{n}}\times\boldsymbol{H}_a(\boldsymbol{r}')e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dS' \\ \boldsymbol{F}_m(\theta,\phi) &= \int_A \boldsymbol{J}_{ms}(\boldsymbol{r}')e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dS' = -\int_A \hat{\mathbf{n}}\times\boldsymbol{E}_a(\boldsymbol{r}')e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dS'\end{aligned}} \tag{18.4.3}$$

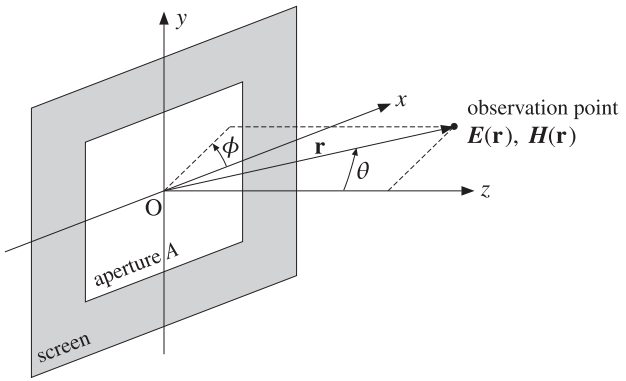


Fig. 18.4.1 Radiation fields from an aperture.

Fig. 18.4.1 shows the polar angle conventions, where we took the origin to be somewhere in the middle of the aperture A.

The aperture surface A and the screen in Fig. 18.1.1 can be arbitrarily curved. However, a common case is to assume that they are both flat. Then, Eqs. (18.4.3) become ordinary 2-d Fourier transform integrals. Taking the aperture plane to be the xy-plane as in Fig. 18.1.1, the aperture normal becomes $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, and thus, it can be taken out of the integrands. Setting $dS' = dx'dy'$, we rewrite Eq. (18.4.3) in the form:

$$\begin{aligned} \boldsymbol{F}(\theta,\phi) &= \int_A \boldsymbol{J}_s(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dx'dy' = \hat{\mathbf{z}} \times \int_A \boldsymbol{H}_a(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dx'dy' \\ \boldsymbol{F}_m(\theta,\phi) &= \int_A \boldsymbol{J}_{ms}(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dx'dy' = -\hat{\mathbf{z}} \times \int_A \boldsymbol{E}_a(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dx'dy' \end{aligned} \tag{18.4.4}$$

where $e^{j\boldsymbol{k}\cdot\boldsymbol{r}'} = e^{jk_x x' + jk_y y'}$ and $k_x = k\cos\phi\sin\theta$, $k_y = k\sin\phi\sin\theta$. It proves convenient then to introduce the *two-dimensional Fourier transforms* of the aperture fields:

$$\boxed{\begin{aligned} \boldsymbol{f}(\theta,\phi) &= \int_A \boldsymbol{E}_a(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dx'dy' = \int_A \boldsymbol{E}_a(x',y')\, e^{jk_x x' + jk_y y'}\, dx'dy' \\ \boldsymbol{g}(\theta,\phi) &= \int_A \boldsymbol{H}_a(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dx'dy' = \int_A \boldsymbol{H}_a(x',y')\, e^{jk_x x' + jk_y y'}\, dx'dy' \end{aligned}} \tag{18.4.5}$$

Then, the radiation vectors become:

$$\begin{aligned} \boldsymbol{F}(\theta,\phi) &= \hat{\mathbf{z}} \times \boldsymbol{g}(\theta,\phi) \\ \boldsymbol{F}_m(\theta,\phi) &= -\hat{\mathbf{z}} \times \boldsymbol{f}(\theta,\phi) \end{aligned} \tag{18.4.6}$$

Because $\boldsymbol{E}_a, \boldsymbol{H}_a$ are tangential to the aperture plane, they can be resolved into their cartesian components, for example, $\boldsymbol{E}_a = \hat{\mathbf{x}} E_{ax} + \hat{\mathbf{y}} E_{ay}$. Then, the quantities $\boldsymbol{f}, \boldsymbol{g}$ can be resolved in the same way, for example, $\boldsymbol{f} = \hat{\mathbf{x}} f_x + \hat{\mathbf{y}} f_y$. Thus, we have:

$$\begin{aligned} \boldsymbol{F} &= \hat{\mathbf{z}} \times \boldsymbol{g} = \hat{\mathbf{z}} \times (\hat{\mathbf{x}} g_x + \hat{\mathbf{y}} g_y) = \hat{\mathbf{y}} g_x - \hat{\mathbf{x}} g_y \\ \boldsymbol{F}_m &= -\hat{\mathbf{z}} \times \boldsymbol{f} = -\hat{\mathbf{z}} \times (\hat{\mathbf{x}} f_x + \hat{\mathbf{y}} f_y) = \hat{\mathbf{x}} f_y - \hat{\mathbf{y}} f_x \end{aligned} \tag{18.4.7}$$

The polar components of the radiation vectors are determined as follows:

$$F_\theta = \hat{\boldsymbol{\theta}} \cdot \boldsymbol{F} = \hat{\boldsymbol{\theta}} \cdot (\hat{\mathbf{y}} g_x - \hat{\mathbf{x}} g_y) = g_x \sin\phi\cos\theta - g_y \cos\phi\cos\theta$$

where we read off the dot products $(\hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{x}})$ and $(\hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{y}})$ from Eq. (15.8.3). The remaining polar components are found similarly, and we summarize them below:

$$\begin{aligned} F_\theta &= -\cos\theta\,(g_y \cos\phi - g_x \sin\phi) \\ F_\phi &= g_x \cos\phi + g_y \sin\phi \\ F_{m\theta} &= \cos\theta\,(f_y \cos\phi - f_x \sin\phi) \\ F_{m\phi} &= -(f_x \cos\phi + f_y \sin\phi) \end{aligned} \tag{18.4.8}$$

It follows from Eq. (18.3.6) that the radiated E-field will be:

$$\boxed{\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{4\pi r}\,[\,(f_x \cos\phi + f_y \sin\phi) + \eta\cos\theta\,(g_y \cos\phi - g_x \sin\phi)\,] \\ E_\phi &= jk\,\frac{e^{-jkr}}{4\pi r}\,[\cos\theta\,(f_y \cos\phi - f_x \sin\phi) - \eta\,(g_x \cos\phi + g_y \sin\phi)\,] \end{aligned}} \tag{18.4.9}$$

The radiation fields resulting from the alternative forms of the field equivalence principle, Eqs. (18.1.2) and (18.1.3), are obtained from Eq. (18.4.9) by removing the g- or the f-terms and doubling the remaining term. We have for the PEC case:

$$\boxed{\begin{aligned} E_\theta &= 2jk\,\frac{e^{-jkr}}{4\pi r}\,[f_x \cos\phi + f_y \sin\phi] \\ E_\phi &= 2jk\,\frac{e^{-jkr}}{4\pi r}\,[\cos\theta\,(f_y \cos\phi - f_x \sin\phi)\,] \end{aligned}} \tag{18.4.10}$$

and for the PMC case:

$$\boxed{\begin{aligned} E_\theta &= 2jk\,\frac{e^{-jkr}}{4\pi r}\,[\eta\cos\theta\,(g_y \cos\phi - g_x \sin\phi)\,] \\ E_\phi &= 2jk\,\frac{e^{-jkr}}{4\pi r}\,[-\eta\,(g_x \cos\phi + g_y \sin\phi)\,] \end{aligned}} \tag{18.4.11}$$

In all three cases, the radiated magnetic fields are obtained from:

$$\boxed{H_\theta = -\frac{1}{\eta}\,E_\phi\,, \qquad H_\phi = \frac{1}{\eta}\,E_\theta} \tag{18.4.12}$$

We note that Eq. (18.4.9) is the average of Eqs. (18.4.10) and (18.4.11). Also, Eq. (18.4.11) is the dual of Eq. (18.4.10). Indeed, using Eq. (18.4.12), we obtain the following H-components for Eq. (18.4.11), which can be derived from Eq. (18.4.10) by the duality transformation $\boldsymbol{E}_a \to \boldsymbol{H}_a$ or $\boldsymbol{f} \to \boldsymbol{g}$, that is,

$$H_\theta = 2jk\,\frac{e^{-jkr}}{4\pi r}\,[g_x\cos\phi + g_y\sin\phi]$$
$$H_\phi = 2jk\,\frac{e^{-jkr}}{4\pi r}\,[\cos\theta\,(g_y\cos\phi - g_x\sin\phi)\,] \tag{18.4.13}$$

At $\theta = 90^o$, the components E_ϕ, H_ϕ become tangential to the aperture screen. We note that because of the $\cos\theta$ factors, E_ϕ (resp. H_ϕ) will vanish in the PEC (resp. PMC) case, in accordance with the boundary conditions.

18.5 Huygens Source

The aperture fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ are referred to as Huygens source if at all points on the aperture they are related by the uniform plane-wave relationship:

$$\boxed{\boldsymbol{H}_a = \frac{1}{\eta}\,\hat{\mathbf{n}}\times\boldsymbol{E}_a} \qquad \text{(Huygens source)} \tag{18.5.1}$$

where η is the characteristic impedance of vacuum.

For example, this is the case if a uniform plane wave is incident normally on the aperture plane from the left, as shown in Fig. 18.5.1. The aperture fields are assumed to be equal to the incident fields, $\boldsymbol{E}_a = \boldsymbol{E}_{\rm inc}$ and $\boldsymbol{H}_a = \boldsymbol{H}_{\rm inc}$, and the incident fields satisfy $\boldsymbol{H}_{\rm inc} = \hat{\mathbf{z}}\times\boldsymbol{E}_{\rm inc}/\eta$.

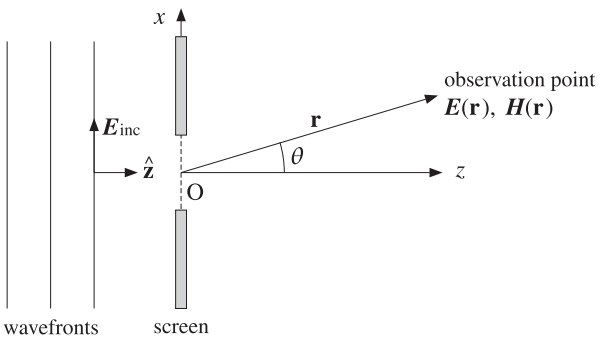


Fig. 18.5.1 Uniform plane wave incident on an aperture.

The Huygens source condition is not always satisfied. For example, if the uniform plane wave is incident obliquely on the aperture, then η must be replaced by the transverse impedance η_T, which depends on the angle of incidence and the polarization of the incident wave as discussed in Sec. 7.2.

Similarly, if the aperture is the open end of a waveguide, then η must be replaced by the waveguide's transverse impedance, such as η_{TE} or η_{TM}, depending on the assumed waveguide mode. On the other hand, if the waveguide ends are flared out into a horn with a large aperture, then Eq. (18.5.1) is approximately valid.

The Huygens source condition implies the same relationship for the Fourier transforms of the aperture fields, that is, (with $\hat{\mathbf{n}} = \hat{\mathbf{z}}$)

$$\boxed{\boldsymbol{g} = \frac{1}{\eta}\,\hat{\mathbf{n}}\times\boldsymbol{f}} \quad\Rightarrow\quad \boxed{g_x = -\frac{1}{\eta}\,f_y\,, \quad g_y = \frac{1}{\eta}\,f_x} \tag{18.5.2}$$

Inserting these into Eq. (18.4.9) we may express the radiated electric field in terms of $\boldsymbol{f}$ only. We find:

$$\boxed{\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,\frac{1+\cos\theta}{2}\,[f_x\cos\phi + f_y\sin\phi] \\ E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,\frac{1+\cos\theta}{2}\,[f_y\cos\phi - f_x\sin\phi] \end{aligned}} \tag{18.5.3}$$

The factor $(1+\cos\theta)/2$ is known as an *obliquity factor*. The PEC case of Eq. (18.4.10) remains unchanged for a Huygens source, but the PMC case becomes:

$$\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,\cos\theta\,[f_x\cos\phi + f_y\sin\phi] \\ E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,[f_y\cos\phi - f_x\sin\phi] \end{aligned} \tag{18.5.4}$$

We may summarize all three cases by the single formula:

$$\boxed{\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,c_\theta\,[f_x\cos\phi + f_y\sin\phi] \\ E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,c_\phi\,[f_y\cos\phi - f_x\sin\phi] \end{aligned}} \qquad \text{(fields from Huygens source)} \tag{18.5.5}$$

where the obliquity factors are defined in the three cases:

$$\begin{bmatrix} c_\theta \\ c_\phi \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1+\cos\theta \\ 1+\cos\theta \end{bmatrix}, \begin{bmatrix} 1 \\ \cos\theta \end{bmatrix}, \begin{bmatrix} \cos\theta \\ 1 \end{bmatrix} \qquad \text{(obliquity factors)} \tag{18.5.6}$$

We note that the first is the average of the last two. The obliquity factors are equal to unity in the forward direction $\theta = 0^o$ and vary little for near-forward angles. Therefore, the radiation patterns predicted by the three methods are very similar in their mainlobe behavior.

In the case of a *modified Huygens source* that replaces η by η_T, Eqs. (18.5.5) retain their form. The aperture fields and their Fourier transforms are now assumed to be related by:

$$\boxed{\boldsymbol{H}_a = \frac{1}{\eta_T}\,\hat{\mathbf{z}}\times\boldsymbol{E}_a} \quad\Rightarrow\quad \boxed{\boldsymbol{g} = \frac{1}{\eta_T}\,\hat{\mathbf{z}}\times\boldsymbol{f}} \tag{18.5.7}$$

Inserting these into Eq. (18.4.9), we obtain the *modified obliquity factors*:

$$c_\theta = \frac{1}{2}\,[1 + K\cos\theta]\,, \quad c_\phi = \frac{1}{2}\,[K + \cos\theta]\,, \qquad K = \frac{\eta}{\eta_T} \tag{18.5.8}$$

18.6 Directivity and Effective Area of Apertures

For any aperture, given the radiation fields E_θ, E_ϕ of Eqs. (18.4.9)–(18.4.11), the corresponding radiation intensity is:

$$U(\theta,\phi) = \frac{dP}{d\Omega} = r^2\mathcal{P}_r = r^2\,\frac{1}{2\eta}\left[|E_\theta|^2 + |E_\phi|^2\right] = r^2\,\frac{1}{2\eta}\,|\boldsymbol{E}(\theta,\phi)|^2 \tag{18.6.1}$$

Because the aperture radiates only into the right half-space $0 \le \theta \le \pi/2$, the total radiated power and the effective isotropic radiation intensity will be:

$$P_{\rm rad} = \int_0^{\pi/2}\int_0^{2\pi} U(\theta,\phi)\,d\Omega\,, \qquad U_I = \frac{P_{\rm rad}}{4\pi} \tag{18.6.2}$$

The directive gain is computed by $D(\theta,\phi) = U(\theta,\phi)/U_I$, and the normalized gain by $g(\theta,\phi) = U(\theta,\phi)/U_{\rm max}$. For a typical aperture, the maximum intensity $U_{\rm max}$ is towards the forward direction $\theta = 0^{\rm o}$. In the case of a Huygens source, we have:

$$U(\theta,\phi) = \frac{k^2}{8\pi^2\eta}\left[c_\theta^2|f_x\cos\phi + f_y\sin\phi|^2 + c_\phi^2|f_y\cos\phi - f_x\sin\phi|^2\right] \tag{18.6.3}$$

Assuming that the maximum is towards $\theta = 0^{\rm o}$, then $c_\theta = c_\phi = 1$, and we find for the maximum intensity:

$$\begin{aligned} U_{\rm max} &= \frac{k^2}{8\pi^2\eta}\left[|f_x\cos\phi + f_y\sin\phi|^2 + |f_y\cos\phi - f_x\sin\phi|^2\right]_{\theta=0} \\ &= \frac{k^2}{8\pi^2\eta}\left[|f_x|^2 + |f_y|^2\right]_{\theta=0} = \frac{k^2}{8\pi^2\eta}\,|\boldsymbol{f}|^2_{\rm max} \end{aligned}$$

where $|\boldsymbol{f}|^2_{\rm max} = \left[|f_x|^2 + |f_y|^2\right]_{\theta=0}$. Setting $k = 2\pi/\lambda$, we have:

$$U_{\rm max} = \frac{1}{2\lambda^2\eta}\,|\boldsymbol{f}|^2_{\rm max} \tag{18.6.4}$$

It follows that the *normalized gain* will be:

$$g(\theta,\phi) = \frac{c_\theta^2|f_x\cos\phi + f_y\sin\phi|^2 + c_\phi^2|f_y\cos\phi - f_x\sin\phi|^2}{|\boldsymbol{f}|^2_{\rm max}} \tag{18.6.5}$$

In the case of Eq. (18.4.9) with $c_\theta = c_\phi = (1+\cos\theta)/2$, this simplifies further into:

$$\boxed{g(\theta,\phi) = c_\theta^2\,\frac{|f_x|^2 + |f_y|^2}{|\boldsymbol{f}|^2_{\rm max}} = \left(\frac{1+\cos\theta}{2}\right)^2\frac{|\boldsymbol{f}(\theta,\phi)|^2}{|\boldsymbol{f}|^2_{\rm max}}} \tag{18.6.6}$$

The square root of the gain is the (normalized) field strength:

$$\boxed{\frac{|\boldsymbol{E}(\theta,\phi)|}{|\boldsymbol{E}|_{\rm max}} = \sqrt{g(\theta,\phi)} = \left(\frac{1+\cos\theta}{2}\right)\frac{|\boldsymbol{f}(\theta,\phi)|}{|\boldsymbol{f}|_{\rm max}}} \tag{18.6.7}$$

The power computed by Eq. (18.6.2) is the total power that is radiated outwards from a half-sphere of large radius r. An alternative way to compute $P_{\rm rad}$ is to invoke energy

conservation and compute the total power that flows *into* the right half-space through the aperture. Assuming a Huygens source, we have:

$$P_{\rm rad} = \int_A \mathcal{P}_z\,dS' = \frac{1}{2}\int_A \hat{\mathbf{z}}\cdot{\rm Re}\left[\boldsymbol{E}_a\times\boldsymbol{H}_a^*\right]dS' = \frac{1}{2\eta}\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|^2 dS' \tag{18.6.8}$$

Because $\theta = 0$ corresponds to $k_x = k_y = 0$, it follows from the Fourier transform definition (18.4.5) that:

$$|\boldsymbol{f}|^2_{\rm max} = \left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\,dS'\right|^2_{k_x=k_y=0} = \left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,dS'\right|^2$$

Therefore, the maximum intensity is given by:

$$U_{\rm max} = \frac{1}{2\lambda^2\eta}\,|\boldsymbol{f}|^2_{\rm max} = \frac{1}{2\lambda^2\eta}\left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,dS'\right|^2 \tag{18.6.9}$$

Dividing (18.6.9) by (18.6.8), we find the directivity:

$$\boxed{D_{\rm max} = 4\pi\,\frac{U_{\rm max}}{P_{\rm rad}} = \frac{4\pi}{\lambda^2}\,\frac{\left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,dS'\right|^2}{\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|^2 dS'} = \frac{4\pi A_{\rm eff}}{\lambda^2}} \quad \text{(directivity)} \tag{18.6.10}$$

It follows that the maximum effective area of the aperture is:

$$\boxed{A_{\rm eff} = \frac{\left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,dS'\right|^2}{\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|^2 dS'} \le A} \quad \text{(effective area)} \tag{18.6.11}$$

and the *aperture efficiency*:

$$\boxed{e_a = \frac{A_{\rm eff}}{A} = \frac{\left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,dS'\right|^2}{A\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|^2 dS'} \le 1} \quad \text{(aperture efficiency)} \tag{18.6.12}$$

The inequalities in Eqs. (18.6.11) and (18.6.12) can be thought of as special cases of the Cauchy-Schwarz inequality. It follows that equality is reached whenever $\boldsymbol{E}_a(\boldsymbol{r}')$ is uniform over the aperture, that is, independent of $\boldsymbol{r}'$.

Thus, *uniform* apertures achieve the *highest* directivity and have effective areas *equal* to their geometrical areas.

Because the integrand in the numerator of e_a depends both on the magnitude and the phase of $\boldsymbol{E}_a$, it proves convenient to separate out these effects by defining the *aperture taper* efficiency or *loss*, $e_{\rm atl}$, and the *phase error* efficiency or *loss*, $e_{\rm pel}$, as follows:

$$e_{\rm atl} = \frac{\left|\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|\,dS'\right|^2}{A\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|^2 dS'}\,, \qquad e_{\rm pel} = \frac{\left|\int_A \boldsymbol{E}_a(\boldsymbol{r}')\,dS'\right|^2}{\left|\int_A |\boldsymbol{E}_a(\boldsymbol{r}')|\,dS'\right|^2} \tag{18.6.13}$$

so that e_a becomes the product:

$$\boxed{e_a = e_{\text{atl}}\, e_{\text{pel}}} \tag{18.6.14}$$

We note that Eq. (18.6.10) was derived under the assumption that the aperture fields were Huygens sources, that is, they were related by, $\eta \boldsymbol{H}_a = \hat{\mathbf{z}} \times \boldsymbol{E}_a$. This assumption is not necessarily always true, but may be justified for electrically large apertures, i.e., with dimensions that are much larger than the wavelength λ.

The correct expression for the directivity is given by Eq. (19.7.10) of Sec. 19.7 and compared with (18.6.10). Directivities that are much larger (even infinite) than those of the uniform case are theoretically possible, but almost impossible to realize in practice. The issues associated with such superdirective apertures are discussed in Sec. 20.22.

18.7 Uniform Apertures

In uniform apertures, the fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ are assumed to be constant over the aperture area. Fig. 18.7.1 shows the examples of a rectangular and a circular aperture. For convenience, we will assume a Huygens source.

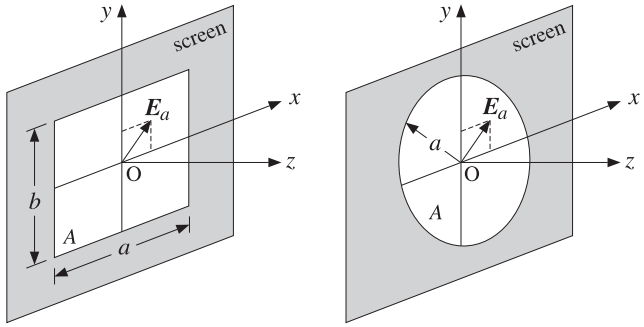


Fig. 18.7.1 Uniform rectangular and circular apertures.

The field $\boldsymbol{E}_a$ can have an arbitrary direction, with constant x- and y-components, $\boldsymbol{E}_a = \hat{\mathbf{x}} E_{0x} + \hat{\mathbf{y}} E_{0y}$. Because $\boldsymbol{E}_a$ is constant, its Fourier transform $\boldsymbol{f}(\theta,\phi)$ becomes:

$$\boldsymbol{f}(\theta,\phi) = \int_A \boldsymbol{E}_a(\boldsymbol{r}')\, e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dS' = \boldsymbol{E}_a \int_A e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dS' \equiv A\, f(\theta,\phi)\, \boldsymbol{E}_a \tag{18.7.1}$$

where we introduced the normalized scalar quantity:

$$\boxed{f(\theta,\phi) = \frac{1}{A}\int_A e^{j\boldsymbol{k}\cdot\boldsymbol{r}'}\, dS'} \qquad \text{(uniform-aperture pattern)} \tag{18.7.2}$$

The quantity $f(\theta,\phi)$ depends on the assumed geometry of the aperture and it, alone, determines the radiation pattern. Noting that the quantity $|\boldsymbol{E}_a|$ cancels out from the ratio in the gain (18.6.7) and that $f(0,\phi) = (1/A)\int_A dS' = 1$, we find for the normalized gain and field strengths:

$$\boxed{\frac{|\boldsymbol{E}(\theta,\phi)|}{|\boldsymbol{E}|_{\max}} = \sqrt{g(\theta,\phi)} = \left(\frac{1+\cos\theta}{2}\right)|f(\theta,\phi)|} \tag{18.7.3}$$

18.8 Rectangular Apertures

For a rectangular aperture of sides a, b, the area integral (18.7.2) is separable in the x- and y-directions:

$$f(\theta,\phi) = \frac{1}{ab}\int_{-a/2}^{a/2}\int_{-b/2}^{b/2} e^{jk_x x' + jk_y y'}\, dx'dy' = \frac{1}{a}\int_{-a/2}^{a/2} e^{jk_x x'}\, dx' \cdot \frac{1}{b}\int_{-b/2}^{b/2} e^{jk_y y'}\, dy'$$

where we placed the origin of the $\boldsymbol{r}'$ integration in the middle of the aperture. The above integrals result in the sinc-function patterns:

$$\boxed{f(\theta,\phi) = \frac{\sin(k_x a/2)}{k_x a/2}\,\frac{\sin(k_y b/2)}{k_y b/2} = \frac{\sin(\pi\nu_x)}{\pi\nu_x}\,\frac{\sin(\pi\nu_y)}{\pi\nu_y}} \tag{18.8.1}$$

where we defined the quantities ν_x, ν_y:

$$\begin{aligned} \nu_x &= \frac{1}{2\pi}\, k_x a = \frac{1}{2\pi}\, ka\sin\theta\cos\phi = \frac{a}{\lambda}\,\sin\theta\cos\phi \\ \nu_y &= \frac{1}{2\pi}\, k_y b = \frac{1}{2\pi}\, kb\sin\theta\sin\phi = \frac{b}{\lambda}\,\sin\theta\sin\phi \end{aligned} \tag{18.8.2}$$

The pattern simplifies along the two principal planes, the xz- and yz-planes, corresponding to $\phi = 0^\circ$ and $\phi = 90^\circ$. We have:

$$\begin{aligned} f(\theta, 0^\circ) &= \frac{\sin(\pi\nu_x)}{\pi\nu_x} = \frac{\sin\big((\pi a/\lambda)\sin\theta\big)}{(\pi a/\lambda)\sin\theta} \\ f(\theta, 90^\circ) &= \frac{\sin(\pi\nu_y)}{\pi\nu_y} = \frac{\sin\big((\pi b/\lambda)\sin\theta\big)}{(\pi b/\lambda)\sin\theta} \end{aligned} \tag{18.8.3}$$

Fig. 18.8.1 shows the three-dimensional pattern of Eq. (18.7.3) as a function of the independent variables ν_x, ν_y, for aperture dimensions $a = 8\lambda$ and $b = 4\lambda$. The x, y separability of the pattern is evident. The essential MATLAB code for generating this figure was (note MATLAB's definition of $\text{sinc}(x) = \sin(\pi x)/(\pi x)$):

```
a = 8; b = 4;
[theta,phi] = meshgrid(0:1:90, 0:9:360);
theta = theta*pi/180;  phi = phi*pi/180;

vx = a*sin(theta).*cos(phi);
vy = b*sin(theta).*sin(phi);

E = abs((1 + cos(theta))/2 .* sinc(vx) .* sinc(vy));

surfl(vx,vy,E);
shading interp;  colormap(gray(16));
```

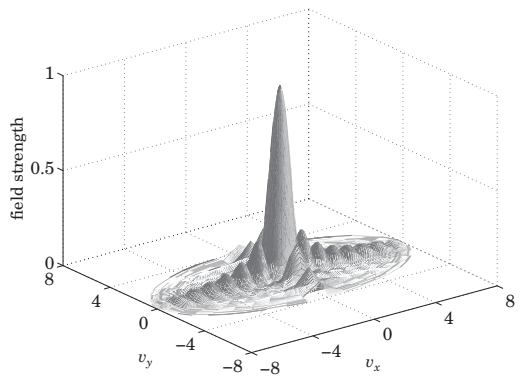


Fig. 18.8.1 Radiation pattern of rectangular aperture ($a = 8\lambda$, $b = 4\lambda$).

As the polar angles vary over $0 \le \theta \le 90^o$ and $0 \le \phi \le 360^o$, the quantities v_x and v_y vary over the limits $-a/\lambda \le v_x \le a/\lambda$ and $-b/\lambda \le v_y \le b/\lambda$. In fact, the *physically realizable* values of v_x, v_y are those that lie in the ellipse in the $v_x v_y$-plane:

$$\boxed{\frac{v_x^2}{a^2} + \frac{v_y^2}{b^2} \le \frac{1}{\lambda^2}} \qquad \text{(visible region)} \tag{18.8.4}$$

The realizable values of v_x, v_y are referred to as the *visible region*. The graph in Fig. 18.8.1 restricts the values of v_x, v_y within that region.

The radiation pattern consists of a narrow mainlobe directed towards the forward direction $\theta = 0^o$ and several sidelobes.

We note the three characteristic properties of the sinc-function patterns: (a) the 3-dB width in v-space is $\Delta v_x = 0.886$ (the 3-dB wavenumber is $v_x = 0.443$); (b) the first sidelobe is down by about 13.26 dB from the mainlobe and occurs at $v_x = 1.4303$; and (c) the first null occurs at $v_x = 1$. See Sec. 22.7 for the proof of these results.

The 3-dB width in angle space can be obtained by linearizing the relationship $v_x = (a/\lambda)\sin\theta$ about $\theta = 0^o$, that is, $\Delta v_x = (a/\lambda)\Delta\theta\cos\theta\,|_{\theta=0} = a\Delta\theta/\lambda$. Thus, $\Delta\theta = \lambda\Delta v_x/a$. This ignores also the effect of the obliquity factor. It follows that the 3-dB widths in the two principal planes are (in radians and in degrees):

$$\Delta\theta_x = 0.886\,\frac{\lambda}{a} = 50.76^o\,\frac{\lambda}{a}\,, \quad \Delta\theta_y = 0.886\,\frac{\lambda}{b} = 50.76^o\,\frac{\lambda}{b} \tag{18.8.5}$$

The 3-dB angles are $\theta_x = \Delta\theta_x/2 = 25.4^o\,\lambda/a$ and $\theta_y = \Delta\theta_y/2 = 25.4^o\,\lambda/b$. Fig. 18.8.2 shows the two principal radiation patterns of Eq. (18.7.3) as functions of θ, for the case $a = 8\lambda$, $b = 4\lambda$. The obliquity factor was included, but it makes essentially no difference near the mainlobe and first sidelobe region, ultimately suppressing the response at $\theta = 90^o$ by a factor of 0.5.

The 3-dB widths are shown on the graphs. The first sidelobes occur at the angles $\theta_a = \text{asin}(1.4303\lambda/a) = 10.30^o$ and $\theta_b = \text{asin}(1.4303\lambda/b) = 20.95^o$.

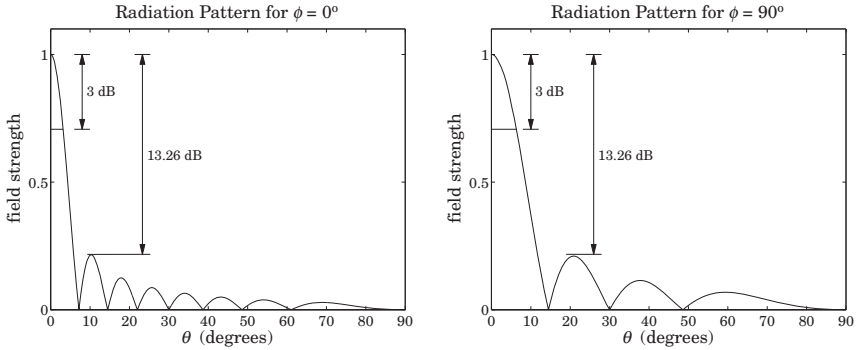


Fig. 18.8.2 Radiation patterns along the two principal planes ($a = 8\lambda$, $b = 4\lambda$).

For aperture antennas, the gain is approximately equal to the directivity because the losses tend to be very small. The gain of the uniform rectangular aperture is, therefore, $G \simeq D = 4\pi(ab)/\lambda^2$. Multiplying G by Eqs. (18.8.5), we obtain the *gain-beamwidth product* $p = G\,\Delta\theta_x\,\Delta\theta_y = 4\pi(0.886)^2 = 9.8646\ \text{rad}^2 = 32\,383\ \text{deg}^2$. Thus, we have an example of the general formula (16.3.14) (with the angles in radians and in degrees):

$$G = \frac{9.8646}{\Delta\theta_x\,\Delta\theta_y} = \frac{32\,383}{\Delta\theta_x^o\,\Delta\theta_y^o} \tag{18.8.6}$$

18.9 Circular Apertures

For a circular aperture of radius a, the pattern integral (18.7.2) can be done conveniently using cylindrical coordinates. The cylindrical symmetry implies that $f(\theta,\phi)$ will be independent of ϕ.

Therefore, for the purpose of computing the integral (18.7.2), we may set $\phi = 0$. We have then $\boldsymbol{k}\cdot\boldsymbol{r}' = k_x x' = k\rho'\sin\theta\cos\phi'$. Writing $dS' = \rho'\,d\rho'\,d\phi'$, we have:

$$f(\theta) = \frac{1}{\pi a^2}\int_0^a\int_0^{2\pi} e^{jk\rho'\sin\theta\cos\phi'}\,\rho'\,d\rho'\,d\phi' \tag{18.9.1}$$

The ϕ'- and ρ'-integrations can be done using the following integral representations for the Bessel functions $J_0(x)$ and $J_1(x)$ [1790]:

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{jx\cos\phi'}\,d\phi' \quad \text{and} \quad \int_0^1 J_0(xr)\,r\,dr = \frac{J_1(x)}{x} \tag{18.9.2}$$

Then Eq. (18.9.1) gives:

$$\boxed{f(\theta) = 2\,\frac{J_1(ka\sin\theta)}{ka\sin\theta} = 2\,\frac{J_1(2\pi u)}{2\pi u}}\,, \qquad u = \frac{1}{2\pi}\,ka\sin\theta = \frac{a}{\lambda}\,\sin\theta \tag{18.9.3}$$

This is the well-known Airy pattern [638] for a circular aperture. The function $f(\theta)$ is normalized to unity at $\theta = 0^o$, because $J_1(x)$ behaves like $J_1(x) \simeq x/2$ for small x.

Fig. 18.9.1 shows the three-dimensional field pattern (18.7.3) as a function of the independent variables $v_x = (a/\lambda)\sin\theta\cos\phi$ and $v_y = (a/\lambda)\sin\theta\sin\phi$, for an aperture radius of $a = 3\lambda$. The obliquity factor was not included as it makes little difference near the main lobe. The MATLAB code for this graph was implemented with the built-in function `besselj`:

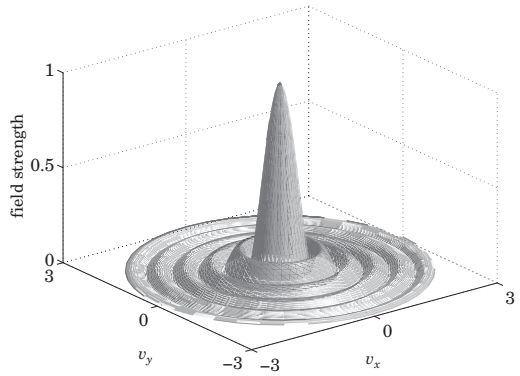


Fig. 18.9.1 Radiation pattern of circular aperture ($a = 3\lambda$).

```
a = 3;
[theta,phi] = meshgrid(0:1:90, 0:9:360);
theta = theta*pi/180;  phi = phi*pi/180;

vx = a*sin(theta).*cos(phi);
vy = a*sin(theta).*sin(phi);
u = a*sin(theta);

E = ones(size(u));
i = find(u);
E(i) = abs(2*besselj(1,2*pi*u(i))./(2*pi*u(i)));

surfl(vx,vy,E);
shading interp;  colormap(gray(16));
```

The visible region is the circle on the $v_x v_y$-plane:

$$v_x^2 + v_y^2 \le \frac{a^2}{\lambda^2} \tag{18.9.4}$$

The mainlobe/sidelobe characteristics of $f(\theta)$ are as follows. The 3-dB wavenumber is $u = 0.2572$ and the 3-dB width in u-space is $\Delta u = 2\times 0.2572 = 0.5144$. The first null occurs at $u = 0.6098$ so that the first-null width is $\Delta u = 2\times 0.6098 = 1.22$. The first sidelobe occurs at $u = 0.8174$ and its height is $|f(u)| = 0.1323$ or 17.56 dB below the mainlobe. The beamwidths in angle space can be obtained from $\Delta u = a(\Delta\theta)/\lambda$, which gives for the 3-dB and first-null widths in radians and degrees:

$$\Delta\theta_{3\text{dB}} = 0.5144\,\frac{\lambda}{a} = 29.47^\text{o}\,\frac{\lambda}{a}\,, \qquad \Delta\theta_\text{null} = 1.22\,\frac{\lambda}{a} = 70^\text{o}\,\frac{\lambda}{a} \tag{18.9.5}$$

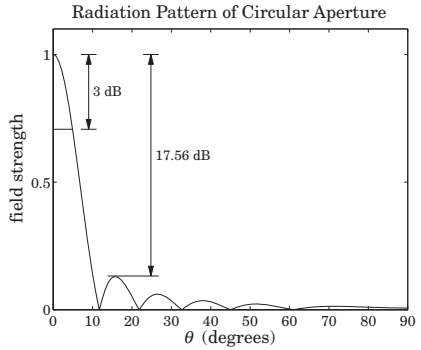


Fig. 18.9.2 Radiation pattern of circular aperture ($a = 3\lambda$).

The 3-dB angle is $\theta_{3\text{dB}} = \Delta\theta_{3\text{dB}}/2 = 0.2572\lambda/a = 14.74^\text{o}\,\lambda/a$ and the first-null angle $\theta_\text{null} = 0.6098\lambda/a$. Fig. 18.9.2 shows the radiation pattern of Eq. (18.7.3) as a function of θ, for the case $a = 3\lambda$. The obliquity factor was included.

The graph shows the 3-dB width and the first sidelobe, which occurs at the angle $\theta_a = \text{asin}(0.817\lambda/a) = 15.8^\text{o}$. The first null occurs at $\theta_\text{null} = \text{asin}(0.6098\lambda/a) = 11.73^\text{o}$, whereas the approximation $\theta_\text{null} = 0.6098\lambda/a$ gives 11.65^o.

The gain-beamwidth product is $p = G(\Delta\theta_{3\text{dB}})^2 = \big[4\pi(\pi a^2)/\lambda^2\big](0.514\lambda/a)^2 = 4\pi^2(0.5144)^2 = 10.4463\ \text{rad}^2 = 34\,293\ \text{deg}^2$. Thus, in radians and degrees:

$$G = \frac{10.4463}{(\Delta\theta_{3\text{dB}})^2} = \frac{34\,293}{(\Delta\theta^\text{o}_{3\text{dB}})^2} \tag{18.9.6}$$

The first-null angle $\theta_\text{null} = 0.6098\lambda/a$ is the so-called *Rayleigh diffraction limit* for the nominal angular resolution of optical instruments, such as microscopes and telescopes. It is usually stated in terms of the diameter $D = 2a$ of the optical aperture:

$$\boxed{\Delta\theta = 1.22\,\frac{\lambda}{D} = 70^\text{o}\,\frac{\lambda}{D}} \qquad \text{(Rayleigh limit)} \tag{18.9.7}$$

18.10 Vector Diffraction Theory

In this section, we provide a justification of the field equivalence principle (18.1.1) and Kottler's formulas (18.4.2) from the point of view of vector diffraction theory. We also discuss the Stratton-Chu and Franz formulas. A historical overview of this subject is given in [1302,1324].

In Sec. 18.2, we worked with the vector potentials and derived the fields due to electric and magnetic currents radiating in an unbounded region. Here, we consider the problem of finding the fields in a volume V bounded by a closed surface S and an infinite spherical surface S_∞, as shown in Fig. 18.10.1.

The solution of this problem requires that we know the current sources within V *and* the electric and magnetic fields tangential to the surface S. The fields $\boldsymbol{E}_1, \boldsymbol{H}_1$ and

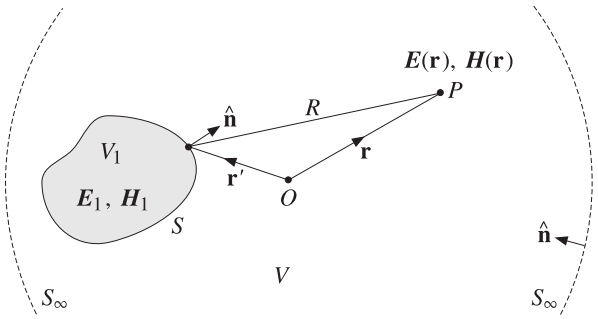


Fig. 18.10.1 Fields outside a closed surface S.

current sources inside the volume V_1 enclosed by S have an effect on the outside only through the tangential fields on the surface.

We start with Maxwell's equations (18.2.1), which include both electric and magnetic currents. This will help us identify the effective surface currents and derive the field equivalence principle.

Taking the curls of both sides of Ampère's and Faraday's laws and using the vector identity $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{E}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{E}) - \nabla^2\boldsymbol{E}$, we obtain the following inhomogeneous Helmholtz equations (which are duals of each other):

$$\begin{aligned} \nabla^2\boldsymbol{E} + k^2\boldsymbol{E} &= j\omega\mu\,\boldsymbol{J} + \frac{1}{\epsilon}\boldsymbol{\nabla}\rho + \boldsymbol{\nabla}\times\boldsymbol{J}_m \\ \nabla^2\boldsymbol{H} + k^2\boldsymbol{H} &= j\omega\epsilon\,\boldsymbol{J}_m + \frac{1}{\mu}\boldsymbol{\nabla}\rho_m - \boldsymbol{\nabla}\times\boldsymbol{J} \end{aligned} \tag{18.10.1}$$

We recall that the Green's function for the Helmholtz equation is:

$$\nabla'^2 G + k^2 G = -\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}'), \quad G(\boldsymbol{r}-\boldsymbol{r}') = \frac{e^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} \tag{18.10.2}$$

where $\boldsymbol{\nabla}'$ is the gradient with respect to $\boldsymbol{r}'$. Applying Green's second identity given by Eq. (C.27) of Appendix C, we obtain:

$$\int_V \left[G\nabla'^2\boldsymbol{E} - \boldsymbol{E}\nabla'^2 G\right] dV' = -\oint_{S+S_\infty}\left[G\frac{\partial \boldsymbol{E}}{\partial n'} - \boldsymbol{E}\frac{\partial G}{\partial n'}\right] dS', \qquad \frac{\partial}{\partial n'} = \hat{\mathbf{n}}\cdot\boldsymbol{\nabla}'$$

where G and $\boldsymbol{E}$ stand for $G(\boldsymbol{r}-\boldsymbol{r}')$ and $\boldsymbol{E}(\boldsymbol{r}')$ and the integration is over $\boldsymbol{r}'$. The quantity $\partial/\partial n'$ is the directional derivative along $\hat{\mathbf{n}}$. The negative sign in the right-hand side arises from using a unit vector $\hat{\mathbf{n}}$ that is pointing *into* the volume V.

The integral over the infinite surface is taken to be zero. This may be justified more rigorously [1295] by assuming that $\boldsymbol{E}$ and $\boldsymbol{H}$ behave like radiation fields with asymptotic form $E \to \text{const.}\, e^{-jkr}/r$ and $\boldsymbol{H} \to \hat{\mathbf{r}}\times\boldsymbol{E}/\eta$.† Thus, dropping the S_∞ term, and adding and subtracting $k^2 G\,\boldsymbol{E}$ in the left-hand side, we obtain:

$$\int_V \left[G(\nabla'^2\boldsymbol{E} + k^2\boldsymbol{E}) - \boldsymbol{E}(\nabla'^2 G + k^2 G)\right] dV' = -\oint_S\left[G\frac{\partial \boldsymbol{E}}{\partial n'} - \boldsymbol{E}\frac{\partial G}{\partial n'}\right] dS' \tag{18.10.3}$$

†The precise conditions are: $r|\boldsymbol{E}| \to$ const. and $r|\boldsymbol{E} - \eta\boldsymbol{H}\times\hat{\mathbf{r}}| \to 0$ as $r \to \infty$.

Using Eq. (18.10.2), the second term on the left may be integrated to give $\boldsymbol{E}(\boldsymbol{r})$:

$$-\int_V \boldsymbol{E}(\boldsymbol{r}')\,(\nabla'^2 G + k^2 G)\,dV' = \int_V \boldsymbol{E}(\boldsymbol{r}')\,\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}')\,dV' = \boldsymbol{E}(\boldsymbol{r})$$

where we assumed that $\boldsymbol{r}$ lies in V. This integral is zero if $\boldsymbol{r}$ lies in V_1 because then $\boldsymbol{r}'$ can never be equal to $\boldsymbol{r}$. For arbitrary $\boldsymbol{r}$, we may write:

$$\int_V \boldsymbol{E}(\boldsymbol{r}')\,\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}')\,dV' = u_V(\boldsymbol{r})\,\boldsymbol{E}(\boldsymbol{r}) = \begin{cases} \boldsymbol{E}(\boldsymbol{r}), & \text{if} \quad \boldsymbol{r}\in V \\ 0, & \text{if} \quad \boldsymbol{r}\notin V \end{cases} \tag{18.10.4}$$

where $u_V(\boldsymbol{r})$ is the characteristic, or indicator, function of the volume region V:†

$$u_V(\boldsymbol{r}) = \begin{cases} 1, & \text{if} \quad \boldsymbol{r}\in V \\ 0, & \text{if} \quad \boldsymbol{r}\notin V \end{cases} \tag{18.10.5}$$

We may now solve Eq. (18.10.3) for $\boldsymbol{E}(\boldsymbol{r})$. In a similar fashion, or, performing a duality transformation on the expression for $\boldsymbol{E}(\boldsymbol{r})$, we also obtain the corresponding magnetic field $\boldsymbol{H}(\boldsymbol{r})$. Using (18.10.1), we have:

$$\boxed{\begin{aligned} \boldsymbol{E}(\boldsymbol{r}) &= \int_V \left[-j\omega\mu\,G\,\boldsymbol{J} - \frac{1}{\epsilon}G\,\boldsymbol{\nabla}'\rho - G\,\boldsymbol{\nabla}'\times\boldsymbol{J}_m\right] dV' + \oint_S\left[\boldsymbol{E}\frac{\partial G}{\partial n'} - G\frac{\partial \boldsymbol{E}}{\partial n'}\right] dS' \\ \boldsymbol{H}(\boldsymbol{r}) &= \int_V \left[-j\omega\epsilon\,G\,\boldsymbol{J}_m - \frac{1}{\mu}G\,\boldsymbol{\nabla}'\rho_m + G\,\boldsymbol{\nabla}'\times\boldsymbol{J}\right] dV' + \oint_S\left[\boldsymbol{H}\frac{\partial G}{\partial n'} - G\frac{\partial \boldsymbol{H}}{\partial n'}\right] dS' \end{aligned}} \tag{18.10.6}$$

Because of the presence of the particular surface term, we will refer to these as the *Kirchhoff diffraction formulas*. Eqs. (18.10.6) can be transformed into the so-called *Stratton-Chu formulas* [1293–1298,1287,1299–1302,1324]:‡

$$\boxed{\begin{aligned} \boldsymbol{E}(\boldsymbol{r}) &= \int_V \left[-j\omega\mu\,G\,\boldsymbol{J} + \frac{\rho}{\epsilon}\,\boldsymbol{\nabla}'G - \boldsymbol{J}_m\times\boldsymbol{\nabla}'G\right] dV' \\ &\quad + \oint_S \left[-j\omega\mu\,G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\boldsymbol{\nabla}'G + (\hat{\mathbf{n}}\times\boldsymbol{E})\times\boldsymbol{\nabla}'G\right] dS' \\ \boldsymbol{H}(\boldsymbol{r}) &= \int_V \left[-j\omega\epsilon\,G\,\boldsymbol{J}_m + \frac{\rho_m}{\mu}\,\boldsymbol{\nabla}'G + \boldsymbol{J}\times\boldsymbol{\nabla}'G\right] dV' \\ &\quad + \oint_S \left[j\omega\epsilon\,G\,(\hat{\mathbf{n}}\times\boldsymbol{E}) + (\hat{\mathbf{n}}\cdot\boldsymbol{H})\,\boldsymbol{\nabla}'G + (\hat{\mathbf{n}}\times\boldsymbol{H})\times\boldsymbol{\nabla}'G\right] dS' \end{aligned}} \tag{18.10.7}$$

The proof of the equivalence of (18.10.6) and (18.10.7) is rather involved. Problem 18.4 breaks down the proof into its essential steps.

Term by term comparison of the volume and surface integrals in (18.10.7) yields the effective surface currents of the field equivalence principle:*

$$\boxed{\boldsymbol{J}_s = \hat{\mathbf{n}}\times\boldsymbol{H}, \quad \boldsymbol{J}_{ms} = -\hat{\mathbf{n}}\times\boldsymbol{E}} \tag{18.10.8}$$

†Technically [1301], one must set $u_V(\boldsymbol{r}) = 1/2$, if $\boldsymbol{r}$ lies on the boundary of V, that is, on S.

‡See [1289,1296,1302,1324] for earlier work by Larmor, Tedone, Ignatowski, and others.

*Initially derived by Larmor and Love [1302,1324], and later developed fully by Schelkunoff [1288,1290].

Similarly, the effective surface charge densities are:

$$\rho_s = \epsilon\,\hat{\mathbf{n}}\cdot\boldsymbol{E}, \qquad \rho_{ms} = \mu\,\hat{\mathbf{n}}\cdot\boldsymbol{H} \tag{18.10.9}$$

Eqs. (18.10.7) may be transformed into the *Kottler formulas* [1293–1298,1287,1299–1302,1324], which eliminate the charge densities ρ, ρ_m in favor of the currents $\boldsymbol{J}, \boldsymbol{J}_m$:

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{r}) &= \frac{1}{j\omega\epsilon}\int_V \left[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\right] dV' \\
&\quad + \frac{1}{j\omega\epsilon}\oint_S \left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS' \\
\boldsymbol{H}(\boldsymbol{r}) &= \frac{1}{j\omega\mu}\int_V \left[k^2\boldsymbol{J}_m G + (\boldsymbol{J}_m\cdot\nabla')\nabla' G + j\omega\mu\,\boldsymbol{J}\times\nabla' G\right] dV' \\
&\quad + \frac{1}{j\omega\mu}\oint_S \left[-k^2 G(\hat{\mathbf{n}}\times\boldsymbol{E}) - \big((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla'\big)\nabla' G + j\omega\mu(\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\right] dS'
\end{aligned} \tag{18.10.10}$$

The steps of the proof are outlined in Problem 18.5.

A related problem is to consider a volume V bounded by the surface S, as shown in Fig. 18.10.2. The fields inside V are still given by (18.10.7), with $\hat{\mathbf{n}}$ pointing again into the volume V. If the surface S recedes to infinity, then (18.10.10) reduce to (18.2.9).

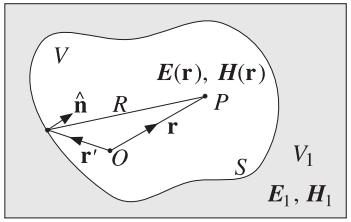


Fig. 18.10.2 Fields inside a closed surface S.

Finally, the Kottler formulas may be transformed into the *Franz formulas* [1298,1287,1299–1301], which are essentially equivalent to Eq. (18.2.8) amended by the vector potentials due to the equivalent surface currents:

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{r}) &= \frac{1}{j\omega\mu\epsilon}\left[\nabla\times\big(\nabla\times(\boldsymbol{A}+\boldsymbol{A}_s)\big) - \mu\,\boldsymbol{J}\right] - \frac{1}{\epsilon}\nabla\times(\boldsymbol{A}_m+\boldsymbol{A}_{ms}) \\
\boldsymbol{H}(\boldsymbol{r}) &= \frac{1}{j\omega\mu\epsilon}\left[\nabla\times\big(\nabla\times(\boldsymbol{A}_m+\boldsymbol{A}_{ms})\big) - \epsilon\,\boldsymbol{J}_m\right] + \frac{1}{\mu}\nabla\times(\boldsymbol{A}+\boldsymbol{A}_s)
\end{aligned} \tag{18.10.11}$$

where $\boldsymbol{A}$ and $\boldsymbol{A}_m$ were defined in Eq. (18.2.6). The new potentials are defined by:

$$\begin{aligned}
\boldsymbol{A}_s(\boldsymbol{r}) &= \oint_S \mu\,\boldsymbol{J}_s(\boldsymbol{r}')\,G(\boldsymbol{r}-\boldsymbol{r}')\,dS' = \oint_S \mu\,[\hat{\mathbf{n}}\times\boldsymbol{H}(\boldsymbol{r}')]\,G(\boldsymbol{r}-\boldsymbol{r}')\,dS' \\
\boldsymbol{A}_{ms}(\boldsymbol{r}) &= \oint_S \epsilon\,\boldsymbol{J}_{ms}(\boldsymbol{r}')\,G(\boldsymbol{r}-\boldsymbol{r}')\,dS' = -\oint_S \epsilon\,[\hat{\mathbf{n}}\times\boldsymbol{E}(\boldsymbol{r}')]\,G(\boldsymbol{r}-\boldsymbol{r}')\,dS'
\end{aligned} \tag{18.10.12}$$

Next, we specialize the above formulas to the case where the volume V contains no current sources ($\boldsymbol{J} = \boldsymbol{J}_m = 0$), so that the $\boldsymbol{E}, \boldsymbol{H}$ fields are given only in terms of the surface integral terms.

This happens if we choose S in Fig. 18.10.1 such that all the current sources are inside it, or, if in Fig. 18.10.2 we choose S such that all the current sources are outside it, then, the Kirchhoff, Stratton-Chu, Kottler, and Franz formulas simplify into:

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{r}) &= \oint_S \left[\boldsymbol{E}\frac{\partial G}{\partial n'} - G\frac{\partial \boldsymbol{E}}{\partial n'}\right] dS' \\
&= \oint_S \left[-j\omega\mu\,G(\hat{\mathbf{n}}\times\boldsymbol{H}) + (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS' \\
&= \frac{1}{j\omega\epsilon}\oint_S \left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS' \\
&= \frac{1}{j\omega\epsilon}\nabla\times\Big(\nabla\times\oint_S G(\hat{\mathbf{n}}\times\boldsymbol{H})\,dS'\Big) + \nabla\times\oint_S G(\hat{\mathbf{n}}\times\boldsymbol{E})\,dS'
\end{aligned} \tag{18.10.13}$$

$$\begin{aligned}
\boldsymbol{H}(\boldsymbol{r}) &= \oint_S \left[\boldsymbol{H}\frac{\partial G}{\partial n'} - G\frac{\partial \boldsymbol{H}}{\partial n'}\right] dS' \\
&= \oint_S \left[j\omega\epsilon\,G(\hat{\mathbf{n}}\times\boldsymbol{E}) + (\hat{\mathbf{n}}\cdot\boldsymbol{H})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\right] dS' \\
&= \frac{1}{j\omega\mu}\oint_S \left[-k^2 G(\hat{\mathbf{n}}\times\boldsymbol{E}) - \big((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla'\big)\nabla' G + j\omega\mu(\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\right] dS' \\
&= -\frac{1}{j\omega\mu}\nabla\times\Big(\nabla\times\oint_S G(\hat{\mathbf{n}}\times\boldsymbol{E})\,dS'\Big) + \nabla\times\oint_S G(\hat{\mathbf{n}}\times\boldsymbol{H})\,dS'
\end{aligned} \tag{18.10.14}$$

where the last equations are the Franz formulas with $\boldsymbol{A} = \boldsymbol{A}_m = 0$.

Fig. 18.10.3 illustrates the geometry of the two cases. Eqs. (18.10.13) and (18.10.14) represent the vectorial formulation of the Huygens-Fresnel principle, according to which the tangential fields on the surface can be considered to be the sources of the fields away from the surface.

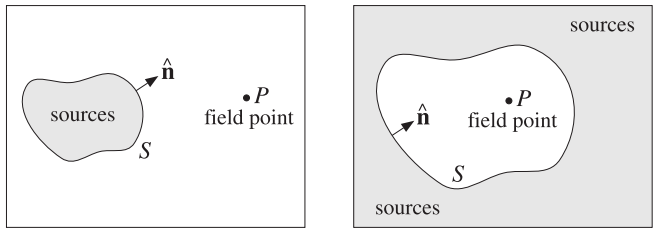


Fig. 18.10.3 Current sources are outside the field region.

18.11 Extinction Theorem

In all of the equivalent formulas for $\boldsymbol{E}(\boldsymbol{r})$, $\boldsymbol{H}(\boldsymbol{r})$, we assumed that $\boldsymbol{r}$ lies within the volume V. The origin of the left-hand sides in these formulas can be traced to Eq. (18.10.4), and therefore, if $\boldsymbol{r}$ is not in V but is within the complementary volume V_1, then the left-hand sides of all the formulas are *zero*. This does not mean that the fields inside V_1 are zero—it only means that the sum of the terms on the right-hand sides are zero.

To clarify these remarks, we consider an imaginary closed surface S dividing all space in two volumes V_1 and V, as shown in Fig. 18.11.1. We assume that there are current sources in both regions V and V_1. The surface S_1 is the same as S but its unit vector $\hat{\mathbf{n}}_1$ points into V_1, so that $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}$. Applying (18.10.10) to the volume V, we have:

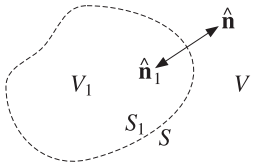


Fig. 18.11.1 Current sources may exist in both V and V_1.

$$\frac{1}{j\omega\epsilon}\oint_S \left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS'$$

$$+\frac{1}{j\omega\epsilon}\int_V \Big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}\boldsymbol{E}(\boldsymbol{r}), & \text{if } \boldsymbol{r}\in V\\ 0, & \text{if } \boldsymbol{r}\in V_1\end{cases}$$

The vanishing of the right-hand side when $\boldsymbol{r}$ is in V_1 is referred to as an *extinction theorem*.† Applying (18.10.10) to V_1, and denoting by $\boldsymbol{E}_1, \boldsymbol{H}_1$ the fields in V_1, we have:

$$\frac{1}{j\omega\epsilon}\oint_{S_1} \left[k^2 G(\hat{\mathbf{n}}_1\times\boldsymbol{H}_1) + \big((\hat{\mathbf{n}}_1\times\boldsymbol{H}_1)\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}_1\times\boldsymbol{E}_1)\times\nabla' G\right] dS'$$

$$+\frac{1}{j\omega\epsilon}\int_{V_1} \Big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}0, & \text{if } \boldsymbol{r}\in V\\ \boldsymbol{E}_1(\boldsymbol{r}), & \text{if } \boldsymbol{r}\in V_1\end{cases}$$

Because $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}$, and on the surface $\boldsymbol{E}_1 = \boldsymbol{E}$ and $\boldsymbol{H}_1 = \boldsymbol{H}$, we may rewrite:

$$-\frac{1}{j\omega\epsilon}\oint_S \left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS'$$

$$+\frac{1}{j\omega\epsilon}\int_{V_1} \Big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}0, & \text{if } \boldsymbol{r}\in V\\ \boldsymbol{E}_1(\boldsymbol{r}), & \text{if } \boldsymbol{r}\in V_1\end{cases}$$

Adding up the two cases and combining the volume integrals into a single one, we obtain:

$$\frac{1}{j\omega\epsilon}\int_{V+V_1} \Big[(\boldsymbol{J}\cdot\nabla')\nabla' G + k^2 G\boldsymbol{J} - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}\boldsymbol{E}(\boldsymbol{r}), & \text{if } \boldsymbol{r}\in V\\ \boldsymbol{E}_1(\boldsymbol{r}), & \text{if } \boldsymbol{r}\in V_1\end{cases}$$

†In fact, it can be used to prove the Ewald-Oseen extinction theorem that we considered in Sec. 15.6.

This is equivalent to Eq. (18.2.9) in which the currents are radiating into unbounded space. We can also see how the sources within V_1 make themselves felt on the outside only through the tangential fields at the surface S, that is, for $\boldsymbol{r}\in V$:

$$\frac{1}{j\omega\epsilon}\int_{V_1}\Big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV'$$

$$= \frac{1}{j\omega\epsilon}\oint_S \left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS'$$

18.12 Vector Diffraction for Apertures

The Kirchhoff diffraction integral, Stratton-Chu, Kottler, and Franz formulas are equivalent only for a *closed* surface S.

If the surface is open, as in the case of an aperture, the four expressions in (18.10.13) and in (18.10.14) are no longer equivalent. In this case, the Kottler and Franz formulas remain equal to each other and give the correct expressions for the fields, in the sense that the resulting $\boldsymbol{E}(\boldsymbol{r})$ and $\boldsymbol{H}(\boldsymbol{r})$ satisfy Maxwell's equations [1289,1287,1302,1324].

For an open surface S bounded by a contour C, shown in Fig. 18.12.1, the Kottler and Franz formulas are related to the Stratton-Chu and the Kirchhoff diffraction integral formulas by the addition of some line-integral correction terms [1296]:

$$\begin{aligned}\boldsymbol{E}(\boldsymbol{r}) &= \frac{1}{j\omega\epsilon}\int_S \left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS'\\ &= \frac{1}{j\omega\epsilon}\,\nabla\times\Big(\nabla\times\int_S G(\hat{\mathbf{n}}\times\boldsymbol{H})\,dS'\Big) + \nabla\times\int_S G(\hat{\mathbf{n}}\times\boldsymbol{E})\,dS'\\ &= \int_S \left[-j\omega\mu\, G(\hat{\mathbf{n}}\times\boldsymbol{H}) + (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\right] dS' - \frac{1}{j\omega\epsilon}\oint_C (\nabla' G)\,\boldsymbol{H}\cdot d\mathbf{l}\\ &= \int_S \left[\boldsymbol{E}\frac{\partial G}{\partial n'} - G\frac{\partial \boldsymbol{E}}{\partial n'}\right] dS' - \oint_C G\,\boldsymbol{E}\times d\mathbf{l} - \frac{1}{j\omega\epsilon}\oint_C (\nabla' G)\,\boldsymbol{H}\cdot d\mathbf{l}\end{aligned} \tag{18.12.1}$$

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{r}) &= \frac{1}{j\omega\mu}\int_S \left[-k^2 G(\hat{\mathbf{n}}\times\boldsymbol{E}) - \big((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla'\big)\nabla' G + j\omega\mu(\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\right] dS'\\ &= -\frac{1}{j\omega\mu}\,\nabla\times\Big(\nabla\times\int_S G(\hat{\mathbf{n}}\times\boldsymbol{E})\,dS'\Big) + \nabla\times\int_S G(\hat{\mathbf{n}}\times\boldsymbol{H})\,dS'\\ &= \int_S \left[j\omega\epsilon\, G(\hat{\mathbf{n}}\times\boldsymbol{E}) + (\hat{\mathbf{n}}\cdot\boldsymbol{H})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\right] dS' + \frac{1}{j\omega\mu}\oint_C (\nabla' G)\,\boldsymbol{E}\cdot d\mathbf{l}\\ &= \int_S \left[\boldsymbol{H}\frac{\partial G}{\partial n'} - G\frac{\partial \boldsymbol{H}}{\partial n'}\right] dS' - \oint_C G\,\boldsymbol{H}\times d\mathbf{l} + \frac{1}{j\omega\mu}\oint_C (\nabla' G)\,\boldsymbol{E}\cdot d\mathbf{l}\end{aligned} \tag{18.12.2}$$

The proof of the equivalence of these expressions is outlined in Problems 18.7 and 18.8. The Kottler-Franz formulas (18.12.1) and (18.12.2) are valid for points off the aperture surface S. The formulas are not consistent for points on the aperture. However, they have been used very successfully in practice to predict the radiation patterns of aperture antennas.

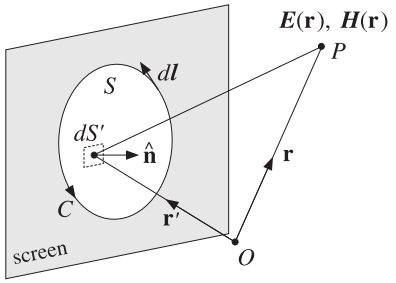


Fig. 18.12.1 Aperture surface S bounded by contour C.

The line-integral correction terms have a minor effect on the mainlobe and near sidelobes of the radiation pattern. Therefore, they can be ignored and the diffracted field can be calculated by any of the four alternative formulas, Kottler, Franz, Stratton-Chu, or Kirchhoff integral—all applied to the *open* surface S.

18.13 Fresnel Diffraction

In Sec. 18.4, we looked at the radiation fields arising from the Kottler-Franz formulas, where we applied the *Fraunhofer approximation* in which only linear phase variations over the aperture were kept in the propagation phase factor e^{-jkR}. Here, we consider the intermediate case of *Fresnel approximation* in which both linear and quadratic phase variations are retained.

We discuss the classical problem of diffraction of a spherical wave by a rectangular aperture, a slit, and a straight-edge using the Kirchhoff integral formula. The case of a plane wave incident on a conducting edge is discussed in Problem 18.11 using the field-equivalence principle and Kottler's formula and more accurately, in Sec. 18.15, using Sommerfeld's exact solution of the geometrical theory of diffraction. These examples are meant to be an introduction to the vast subject of diffraction.

In Fig. 18.13.1, we consider a rectangular aperture illuminated from the left by a point source radiating a spherical wave. We take the origin to be somewhere on the aperture plane, but eventually we will take it to be the point of intersection of the aperture plane and the line between the source and observation points P_1 and P_2.

The diffracted field at point P_2 may be calculated from the Kirchhoff formula applied to any of the cartesian components of the field:

$$E = \int_S \left[E_1 \frac{\partial G}{\partial n'} - G \frac{\partial E_1}{\partial n'} \right] dS' \tag{18.13.1}$$

where E_1 is the spherical wave from the source point P_1 evaluated at the aperture point P', and G is the Green's function from P' to P_2:

$$E_1 = A_1 \frac{e^{-jkR_1}}{R_1}, \quad G = \frac{e^{-jkR_2}}{4\pi R_2} \tag{18.13.2}$$

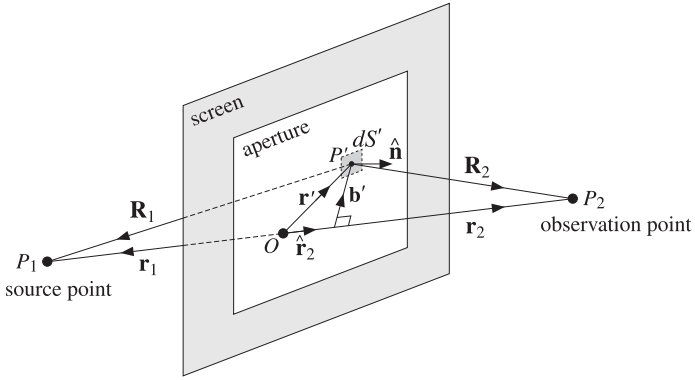


Fig. 18.13.1 Fresnel diffraction through rectangular aperture.

where A_1 is a constant. If $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the vectors pointing from the origin to the source and observation points, then we have for the distance vectors $\mathbf{R}_1$ and $\mathbf{R}_2$:

$$\begin{aligned} \mathbf{R}_1 &= \boldsymbol{r}_1 - \boldsymbol{r}', \quad R_1 = |\boldsymbol{r}_1 - \boldsymbol{r}'| = \sqrt{r_1^2 - 2\boldsymbol{r}_1 \cdot \boldsymbol{r}' + \boldsymbol{r}' \cdot \boldsymbol{r}'} \\ \mathbf{R}_2 &= \boldsymbol{r}_2 - \boldsymbol{r}', \quad R_2 = |\boldsymbol{r}_2 - \boldsymbol{r}'| = \sqrt{r_2^2 - 2\boldsymbol{r}_2 \cdot \boldsymbol{r}' + \boldsymbol{r}' \cdot \boldsymbol{r}'} \end{aligned} \tag{18.13.3}$$

Therefore, the gradient operator $\boldsymbol{\nabla}'$ can be written as follows when it acts on a function of $R_1 = |\boldsymbol{r}_1 - \boldsymbol{r}'|$ or a function of $R_2 = |\boldsymbol{r}_2 - \boldsymbol{r}'|$:

$$\boldsymbol{\nabla}' = -\hat{\mathbf{R}}_1 \frac{\partial}{\partial R_1}, \quad \boldsymbol{\nabla}' = -\hat{\mathbf{R}}_2 \frac{\partial}{\partial R_2}$$

where $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ are the unit vectors in the directions of $\mathbf{R}_1$ and $\mathbf{R}_2$. Thus, we have:

$$\begin{aligned} \frac{\partial E_1}{\partial n'} &= \hat{\mathbf{n}} \cdot \boldsymbol{\nabla}' E_1 = -\hat{\mathbf{n}} \cdot \hat{\mathbf{R}}_1 \frac{\partial E_1}{\partial R_1} = (\hat{\mathbf{n}} \cdot \hat{\mathbf{R}}_1) \left(jk + \frac{1}{R_1} \right) A_1 \frac{e^{-jkR_1}}{R_1} \\ \frac{\partial G}{\partial n'} &= \hat{\mathbf{n}} \cdot \boldsymbol{\nabla}' G = -\hat{\mathbf{n}} \cdot \hat{\mathbf{R}}_2 \frac{\partial G}{\partial R_2} = (\hat{\mathbf{n}} \cdot \hat{\mathbf{R}}_2) \left(jk + \frac{1}{R_2} \right) \frac{e^{-jkR_2}}{4\pi R_2} \end{aligned} \tag{18.13.4}$$

Dropping the $1/R^2$ terms, we find for the integrand of Eq. (18.13.1):

$$E_1 \frac{\partial G}{\partial n'} - G \frac{\partial E_1}{\partial n'} = \frac{jkA_1}{4\pi R_1 R_2} \left[(\hat{\mathbf{n}} \cdot \hat{\mathbf{R}}_2) - (\hat{\mathbf{n}} \cdot \hat{\mathbf{R}}_1) \right] e^{-jk(R_1+R_2)}$$

Except in the phase factor $e^{-jk(R_1+R_2)}$, we may replace $\mathbf{R}_1 \simeq \boldsymbol{r}_1$ and $\mathbf{R}_2 \simeq \boldsymbol{r}_2$, that is,

$$E_1 \frac{\partial G}{\partial n'} - G \frac{\partial E_1}{\partial n'} = \frac{jkA_1}{4\pi r_1 r_2} \left[(\hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_2) - (\hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_1) \right] e^{-jk(R_1+R_2)} \tag{18.13.5}$$

Thus, we have for the diffracted field at point P_2:

$$\boxed{E = \frac{jkA_1}{4\pi r_1 r_2} \left[(\hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_2) - (\hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_1) \right] \int_S e^{-jk(R_1+R_2)} \, dS'} \tag{18.13.6}$$

The quantity $[(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_1)]$ is an obliquity factor. Next, we set $r = r_1 + r_2$ and define the "free-space" field at the point P_2:

$$E_0 = A_1\,\frac{e^{-jk(r_1+r_2)}}{r_1+r_2} = A_1\,\frac{e^{-jkr}}{r} \tag{18.13.7}$$

If the origin were the point of intersection between the aperture plane and the line P_1P_2, then E_0 would represent the field received at point P_2 in the *unobstructed* case when the aperture and screen are absent.

The ratio $D = E/E_0$ may be called the *diffraction coefficient* and depends on the aperture and the relative geometry of the points P_1, P_2:

$$\boxed{D = \frac{E}{E_0} = \frac{jk}{4\pi F}\,[(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_1)]\int_S e^{-jk(R_1+R_2-r_1-r_2)}\,dS'} \tag{18.13.8}$$

where we defined the "focal length" between r_1 and r_2:

$$\boxed{\frac{1}{F} = \frac{1}{r_1}+\frac{1}{r_2}} \quad\Rightarrow\quad \boxed{F = \frac{r_1 r_2}{r_1+r_2}} \tag{18.13.9}$$

The Fresnel approximation is obtained by expanding R_1 and R_2 in powers of r' and keeping only terms up to second order. We rewrite Eq. (18.13.3) in the form:

$$R_1 = r_1\sqrt{1-\frac{2\hat{\mathbf{r}}_1\cdot\boldsymbol{r}'}{r_1}+\frac{\boldsymbol{r}'\cdot\boldsymbol{r}'}{r_1^2}}\,,\quad R_2 = r_2\sqrt{1-\frac{2\hat{\mathbf{r}}_2\cdot\boldsymbol{r}'}{r_2}+\frac{\boldsymbol{r}'\cdot\boldsymbol{r}'}{r_2^2}}$$

Next, we apply the Taylor series expansion up to second order:

$$\sqrt{1+x} = 1+\frac{1}{2}x-\frac{1}{8}x^2$$

This gives the approximations of R_1, R_2, and $R_1+R_2-r_1-r_2$:

$$R_1 = r_1 - \hat{\mathbf{r}}_1\cdot\boldsymbol{r}' + \frac{1}{2r_1}[\boldsymbol{r}'\cdot\boldsymbol{r}' - (\hat{\mathbf{r}}_1\cdot\boldsymbol{r}')^2]$$

$$R_2 = r_2 - \hat{\mathbf{r}}_2\cdot\boldsymbol{r}' + \frac{1}{2r_2}[\boldsymbol{r}'\cdot\boldsymbol{r}' - (\hat{\mathbf{r}}_2\cdot\boldsymbol{r}')^2]$$

$$R_1+R_2-r_1-r_2 = -(\hat{\mathbf{r}}_1+\hat{\mathbf{r}}_2)\cdot\boldsymbol{r}' + \frac{1}{2}\left[\left(\frac{1}{r_1}+\frac{1}{r_2}\right)\boldsymbol{r}'\cdot\boldsymbol{r}' - \frac{(\hat{\mathbf{r}}_1\cdot\boldsymbol{r}')^2}{r_1} - \frac{(\hat{\mathbf{r}}_2\cdot\boldsymbol{r}')^2}{r_2}\right]$$

To simplify this expression, we now assume that the origin is the point of intersection of the line of sight P_1P_2 and the aperture plane. Then, the vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are anti-parallel and so are their unit vectors $\hat{\mathbf{r}}_1 = -\hat{\mathbf{r}}_2$. The linear terms cancel and the quadratic ones combine to give:

$$R_1+R_2-r_1-r_2 = \frac{1}{2F}[\boldsymbol{r}'\cdot\boldsymbol{r}' - (\hat{\mathbf{r}}_2\cdot\boldsymbol{r}')^2] = \frac{1}{2F}\,|\boldsymbol{r}'-\hat{\mathbf{r}}_2(\boldsymbol{r}'\cdot\hat{\mathbf{r}}_2)|^2 = \frac{1}{2F}\,\mathbf{b}'\cdot\mathbf{b}' \tag{18.13.10}$$

where we defined $\mathbf{b}' = \boldsymbol{r}' - \hat{\mathbf{r}}_2(\boldsymbol{r}'\cdot\hat{\mathbf{r}}_2)$, which is the perpendicular vector from the point P' to the line-of-sight P_1P_2, as shown in Fig. 18.13.1.

It follows that the Fresnel approximation of the diffraction coefficient for an arbitrary aperture will be given by:

$$\boxed{D = \frac{E}{E_0} = \frac{jk\,(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)}{2\pi F}\int_S e^{-jk\,(\mathbf{b}'\cdot\mathbf{b}')/(2F)}\,dS'} \tag{18.13.11}$$

A further simplification is obtained by assuming that the aperture plane is the xy-plane and that the line P_1P_2 lies on the yz plane at an angle θ with the z-axis, as shown in Fig. 18.13.2.

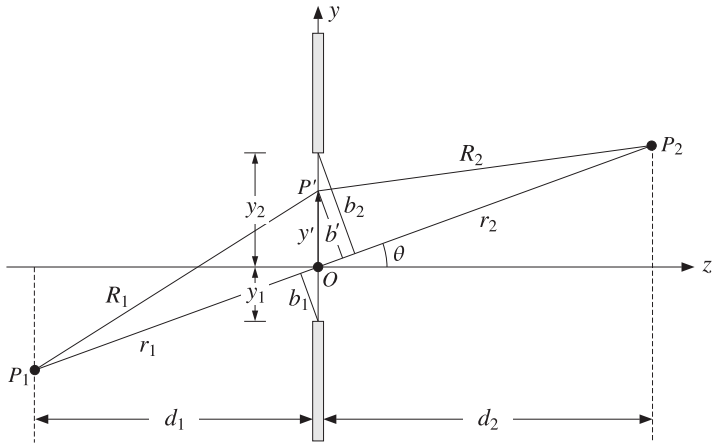


Fig. 18.13.2 Fresnel diffraction by rectangular aperture.

Then, we have $\boldsymbol{r}' = x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}}$, $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, and $\hat{\mathbf{r}}_2 = \hat{\mathbf{z}}\cos\theta + \hat{\mathbf{y}}\sin\theta$. It follows that $\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2 = \cos\theta$, and the perpendicular distance $\mathbf{b}'\cdot\mathbf{b}'$ becomes:

$$\mathbf{b}'\cdot\mathbf{b}' = \boldsymbol{r}'\cdot\boldsymbol{r}' - (\hat{\mathbf{r}}'\cdot\hat{\mathbf{r}}_2)^2 = x'^2 + y'^2 - (y'\sin\theta)^2 = x'^2 + y'^2\cos^2\theta$$

Then, the diffraction coefficient (18.13.11) becomes:

$$D = \frac{jk\cos\theta}{2\pi F}\int_{-x_1}^{x_2}\int_{-y1}^{y_2} e^{-jk(x'^2+y'^2\cos^2\theta)/2F}\,dx'\,dy' \tag{18.13.12}$$

where we assumed that the aperture limits are (with respect to the new origin):

$$-x_1 \le x' \le x_2\,,\quad -y_1 \le y' \le y_2$$

The end-points y_1, y_2 are shown in Fig. 18.13.2. The integrals may be expressed in terms of the Fresnel functions $C(x)$, $S(x)$, and $\mathcal{F}(x) = C(x) - jS(x)$ discussed in Appendix F. There, the complex function $\mathcal{F}(x)$ is defined by:

$$\boxed{\mathcal{F}(x) = C(x) - jS(x) = \int_0^x e^{-j(\pi/2)u^2}\,du} \qquad \text{(Fresnel function)} \tag{18.13.13}$$

We change integration variables to the normalized Fresnel variables:

$$u = \sqrt{\frac{k}{\pi F}}\, x' , \quad v = \sqrt{\frac{k}{\pi F}}\, y' \cos\theta \tag{18.13.14}$$

where $b' = y' \cos\theta$ is the perpendicular distance from P' to the line P_1P_2, as shown in Fig. 18.13.2. The corresponding end-points are:

$$u_i = \sqrt{\frac{k}{\pi F}}\, x_i , \quad v_i = \sqrt{\frac{k}{\pi F}}\, y_i \cos\theta = \sqrt{\frac{k}{\pi F}}\, b_i , \qquad i = 1, 2 \tag{18.13.15}$$

Note that the quantities $b_1 = y_1 \cos\theta$ and $b_2 = y_2 \cos\theta$ are the perpendicular distances from the edges to the line P_1P_2. Since $du\, dv = (k\cos\theta/\pi F)\, dx'\, dy'$, we obtain for the diffraction coefficient:

$$D = \frac{j}{2} \int_{-u_1}^{u_2} e^{-j\pi u^2/2}\, du \int_{-v_1}^{v_2} e^{-j\pi v^2/2}\, dv = \frac{j}{2} [\mathcal{F}(u_2) - \mathcal{F}(-u_1)]\, [\mathcal{F}(v_2) - \mathcal{F}(-v_1)]$$

Noting that $\mathcal{F}(x)$ is an odd function and that $j/2 = 1/(1-j)^2$, we obtain:

$$\boxed{D = \frac{E}{E_0} = \frac{\mathcal{F}(u_1) + \mathcal{F}(u_2)}{1-j}\, \frac{\mathcal{F}(v_1) + \mathcal{F}(v_2)}{1-j}} \quad \text{(rectangular aperture)} \tag{18.13.16}$$

The normalization factors $(1-j)$ correspond to the infinite aperture limit $u_1, u_2, v_1, v_2 \to \infty$, that is, no aperture at all. Indeed, since the asymptotic value of $\mathcal{F}(x)$ is $\mathcal{F}(\infty) = (1-j)/2$, we have:

$$\frac{\mathcal{F}(u_1) + \mathcal{F}(u_2)}{1-j}\, \frac{\mathcal{F}(v_1) + \mathcal{F}(v_2)}{1-j} \longrightarrow \frac{\mathcal{F}(\infty) + \mathcal{F}(\infty)}{1-j}\, \frac{\mathcal{F}(\infty) + \mathcal{F}(\infty)}{1-j} = 1$$

In the case of a *long slit* along the x-direction, we only take the limit $u_1, u_2 \to \infty$:

$$\boxed{D = \frac{E}{E_0} = \frac{\mathcal{F}(v_1) + \mathcal{F}(v_2)}{1-j}} \quad \text{(diffraction by long slit)} \tag{18.13.17}$$

18.14 Knife-Edge Diffraction

The case of *straight-edge* or *knife-edge* diffraction is obtained by taking the limit $y_2 \to \infty$, or $v_2 \to \infty$, which corresponds to keeping the lower edge of the slit. In this limit $\mathcal{F}(v_2) \to \mathcal{F}(\infty) = (1-j)/2$. Denoting v_1 by v, we have:

$$\boxed{D(v) = \frac{1}{1-j} \left(\mathcal{F}(v) + \frac{1-j}{2} \right)}, \qquad v = \sqrt{\frac{k}{\pi F}}\, b_1 \tag{18.14.1}$$

where,

$$\mathcal{F}(v) = \int_0^v e^{-j\pi u^2/2}\, du , \qquad D(v) = \frac{1}{1-j} \int_{-\infty}^{v} e^{-j\pi u^2/2}\, du \tag{18.14.2}$$

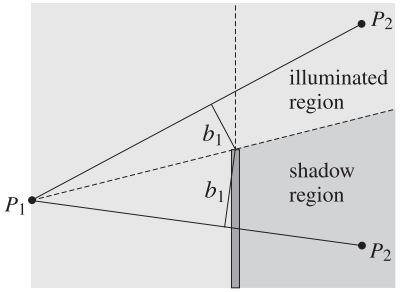


Fig. 18.14.1 Illuminated and shadow regions in straight-edge diffraction.

Positive values of v correspond to positive values of the clearance distance b_1, placing the point P_2 in the illuminated region, as shown in Fig. 18.14.1. Negative values of v correspond to $b_1 < 0$, placing P_2 in the *geometrical shadow* region behind the edge.

The magnitude-square $|D|^2$ represents the intensity of the diffracted field relative to the intensity of the unobstructed field. Since $|1-j|^2 = 2$, we find:

$$\boxed{|D(v)|^2 = \frac{|E|^2}{|E_0|^2} = \frac{1}{2} \left| \mathcal{F}(v) + \frac{1-j}{2} \right|^2} \tag{18.14.3}$$

or, in terms of the real and imaginary parts of $\mathcal{F}(v)$:

$$\boxed{|D(v)|^2 = \frac{1}{2} \left[\left(C(v) + \frac{1}{2} \right)^2 + \left(S(v) + \frac{1}{2} \right)^2 \right]} \tag{18.14.4}$$

The quantity $|D(v)|^2$ is plotted versus v in Fig. 18.14.2. At $v = 0$, corresponding to the line P_1P_2 grazing the top of the edge, we have $\mathcal{F}(0) = 0$, $D(0) = 1/2$, and $|D(0)|^2 = 1/4$ or a 6 dB loss. The first maximum in the illuminated region occurs at $v = 1.2172$ and has the value $|D(v)|^2 = 1.3704$, or a gain of 1.37 dB.

The asymptotic behavior of $D(v)$ for $v \to \pm\infty$ is obtained from Eq. (F.4). We have for large positive x:

$$\mathcal{F}(\pm x) \to \pm \left(\frac{1-j}{2} + \frac{j}{\pi x}\, e^{-j\pi x^2/2} \right)$$

This implies that:

$$\boxed{D(v) = \begin{cases} 1 - \dfrac{1-j}{2\pi v}\, e^{-j\pi v^2/2}, & \text{for} \quad v \to +\infty \\ -\dfrac{1-j}{2\pi v}\, e^{-j\pi v^2/2}, & \text{for} \quad v \to -\infty \end{cases}} \tag{18.14.5}$$

We may combine the two expressions into one with the help of the unit-step function $u(v)$ by writing $D(v)$ in the following form, which defines the asymptotic diffraction coefficient $d(v)$:

$$\boxed{D(v) = u(v) + d(v)\, e^{-j\pi v^2/2}} \tag{18.14.6}$$

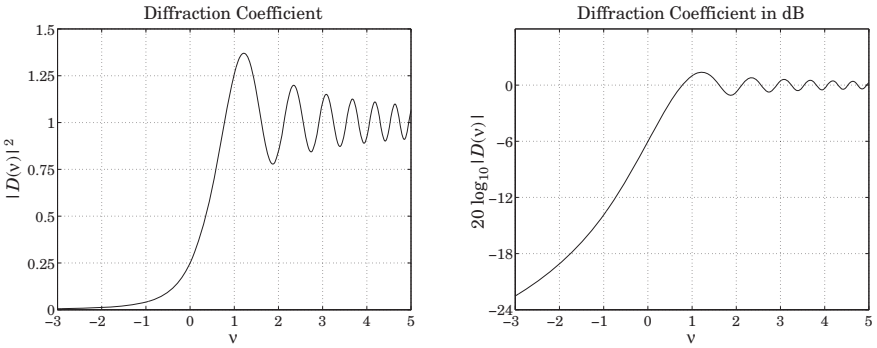


Fig. 18.14.2 Diffraction coefficient in absolute and dB units.

where $u(\nu)=1$ for $\nu \geq 0$ and $u(\nu)=0$ for $\nu < 0$.

With $u(0)=1$, this definition requires $d(0)=D(0)-\nu(0)=0.5-1=-0.5$. But if we define $u(0)=0.5$, as is sometimes done, then, $d(0)=0$. The asymptotic behavior of $D(\nu)$ can now be expressed in terms of the asymptotic behavior of $d(\nu)$:

$$\boxed{d(\nu) = -\frac{1-j}{2\pi\nu}}, \qquad \text{for} \quad \nu \to \pm\infty \tag{18.14.7}$$

In the illuminated region $D(\nu)$ tends to unity, whereas in the shadow region it decreases to zero with asymptotic dB attenuation or loss:

$$L = -10\log_{10}|d(\nu)|^2 = 10\log_{10}(2\pi^2\nu^2), \quad \text{as} \quad \nu \to -\infty \tag{18.14.8}$$

The MATLAB function `diffr.m`, mentioned in Appendix F, calculates the diffraction coefficient (18.14.1) at any vector of (real) values of ν. It has usage:

```
D = diffr(v);    % knife-edge diffraction coefficient D(v)
```

For values $\nu \leq 0.7$, the diffraction loss can be approximated very well by the following function [1309]:

$$L = -10\log_{10}|D(\nu)|^2 = 6.9 + 20\log_{10}\left(\sqrt{(\nu+0.1)^2+1} - \nu - 0.1\right) \tag{18.14.9}$$

Example 18.14.1: *Diffraction Loss over Obstacles.* The propagation path loss over obstacles and irregular terrain is usually determined using knife-edge diffraction. Fig. 18.14.3 illustrates the case of two antennas communicating over an obstacle. For small angles θ, the focal length F is often approximated in several forms:

$$F = \frac{r_1 r_2}{r_1 + r_2} \simeq \frac{d_1 d_2}{d_1 + d_2} \simeq \frac{l_1 l_2}{l_1 + l_2}$$

These approximations are valid typically when d_1, d_2 are much greater than λ and the height h of the obstacle, typically, at least ten times greater. The clearance distance can

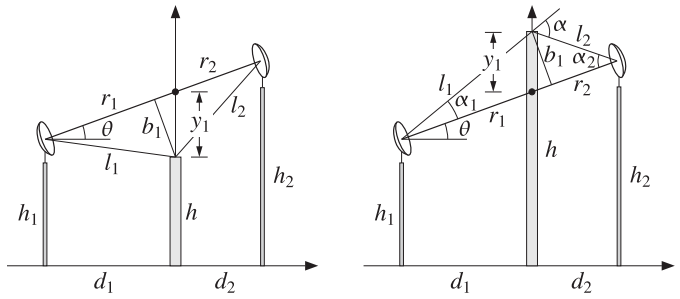


Fig. 18.14.3 Communicating antennas over an obstacle.

be expressed in terms of the heights:

$$b_1 = y_1\cos\theta = \left(\frac{h_1 d_2 + h_2 d_1}{d_1 + d_2} - h\right)\cos\theta$$

The distance b_1 can also be expressed approximately in terms of the subtended angles α_1, α_2, and α, shown in Fig. 18.14.3:

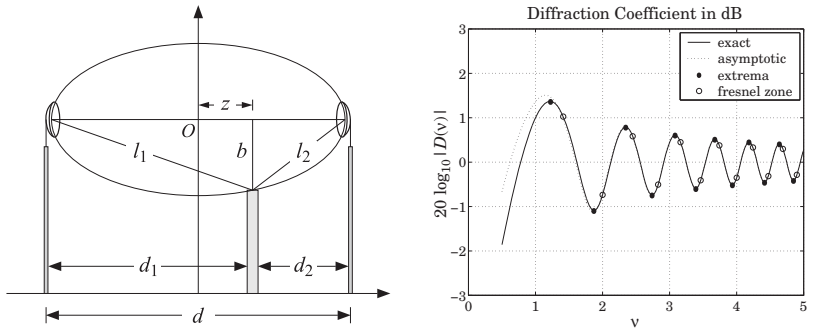
$$b_1 \simeq l_1\alpha_1 \simeq l_2\alpha_2 \quad \Rightarrow \quad b_1 = \sqrt{l_1 l_2 \alpha_1 \alpha_2} \tag{18.14.10}$$

and in terms of α, we have:

$$\alpha_1 = \frac{\alpha l_2}{l_1 + l_2}, \quad \alpha_2 = \frac{\alpha l_1}{l_1 + l_2} \quad \Rightarrow \quad b_1 = \alpha F \quad \Rightarrow \quad \nu = \alpha\sqrt{\frac{2F}{\lambda}} \tag{18.14.11}$$

The case of multiple obstacles has been studied using appropriate modifications of the knife-edge diffraction problem and the geometrical theory of diffraction [1398–1413]. □

Example 18.14.2: *Fresnel Zones.* Consider two antennas separated by a distance d and an obstacle at distance z from the midpoint with clearance b, as shown below. Fresnel zones and the corresponding Fresnel zone ellipsoids help answer the question of what the minimum value of the clearance b should be for efficient communication between the antennas.



The diffraction coefficient $D(\nu)$ and its asymptotic form were given in Eqs. (18.14.1) and (18.14.5), that is,

$$D(\nu)=\frac{1}{1-j}\left(\mathcal{F}(\nu)+\frac{1-j}{2}\right),\quad \nu=\sqrt{\frac{k}{\pi F}}\,b=\sqrt{\frac{2}{\lambda F}}\,b,\quad F=\frac{d_1 d_2}{d_1+d_2} \tag{18.14.12}$$

and for positive and large clearance b, or equivalently, for large positive ν,

$$D_{\text{as}}(\nu)=1-\frac{1-j}{2\pi\nu}\,e^{-j\pi\nu^2/2}=1-\frac{1}{\sqrt{2}\pi\nu}\,e^{-j\pi(\nu^2/2+1/4)} \tag{18.14.13}$$

As can be seen in the above figure on the right, the diffraction coefficients $D(\nu)$ and $D_{\text{as}}(\nu)$ agree closely even for small values of ν. Therefore, the extrema can be obtained from the asymptotic form. They correspond to the values of ν that cause the exponential in (18.14.13) to take on its extremal values of ± 1, that is, the ν's that satisfy $\nu^2/2+1/4=n$, with integer n, or:

$$\nu_n=\sqrt{2n-0.5},\quad n=1,2,\dots \tag{18.14.14}$$

The corresponding values of $D(\nu)$, shown on the figure with black dots, are given by

$$D_{\text{as}}(\nu_n)=1-\frac{1}{\sqrt{2}\pi\nu_n}\,e^{-j\pi n}=1-\frac{1}{\sqrt{2}\pi\nu_n}\,(-1)^n \tag{18.14.15}$$

An alternative set of ν's, also corresponding to alternating almost extremum values, are those that define the conventional Fresnel zones, that is,

$$u_n=\sqrt{2n},\quad n=1,2,\dots \tag{18.14.16}$$

These are indicated by open circles on the graph. The corresponding $D(\nu)$ values are:

$$D_{\text{as}}(u_n)=1-\frac{e^{-j\pi/4}}{\sqrt{2}\pi u_n}\,(-1)^n \tag{18.14.17}$$

For clearances b that correspond to ν's that are too small, i.e., $\nu<0.5$, the diffraction coefficient $D(\nu)$ becomes too small, impeding efficient communication. The smallest acceptable clearance b is taken to correspond to the first maximum of $D(\nu)$, that is, $\nu=\nu_1$ or more simply $\nu=u_1=\sqrt{2}$.

The locus of points (b,z) corresponding to a fixed value of ν, and hence to a fixed value of the diffraction coefficient $D(\nu)$, form an ellipsoid. This can be derived from (18.14.12) by setting $d_1=d/2+z$ and $d_2=d/2-z$, that is,

$$\nu=\sqrt{\frac{2}{\lambda F}}\,b\quad\Rightarrow\quad b^2=\frac{\lambda F}{2}\,\nu^2=\frac{\lambda\,(d^2/4-z^2)}{2d}\,\nu^2,\quad \text{because } F=\frac{d_1 d_2}{d_1+d_2}=\frac{d^2/4-z^2}{d}$$

which can be rearranged into the equation of an ellipse:

$$\left(\frac{8}{\nu^2\lambda d}\right)b^2+\left(\frac{4}{d^2}\right)z^2=1$$

For $\nu=u_1=\sqrt{2}$, this defines the first Fresnel zone ellipse, which gives the minimum acceptable clearance for a given distance z:

$$\left(\frac{4}{\lambda d}\right)b^2+\left(\frac{4}{d^2}\right)z^2=1 \tag{18.14.18}$$

If the obstacle is at midpoint ($z=0$), the minimum clearance becomes:

$$b=\frac{1}{2}\sqrt{\lambda d} \tag{18.14.19}$$

For example, for a distance of $d=1$ km, using a cell phone frequency of $f=1$ GHz, corresponding to wavelength $\lambda=30$ cm, we find $b=\sqrt{\lambda d}/2=8.66$ meters.

A common interpretation and derivation of Fresnel zones is to consider the path difference between the rays following the straight path connecting the two antennas and the path getting scattered from the obstacle, that is, $\Delta l=l_1+l_2-d$. From the indicated triangles, and assuming that $b\ll d_1$ and $b\ll d_2$, we find:

$$l_1=\sqrt{d_1^2+b^2}\simeq d_1+\frac{b^2}{2d_1},\quad l_2=\sqrt{d_2^2+b^2}\simeq d_2+\frac{b^2}{2d_2}$$

which leads to the following path length Δl, expressed in terms of ν:

$$\Delta l=l_1+l_2-d=\frac{b^2}{2}\left(\frac{1}{d_1}+\frac{1}{d_2}\right)=\frac{b^2}{2F}=\frac{\lambda}{4}\,\nu^2$$

The corresponding phase difference between the two paths, $e^{-jk\Delta l}$, will be then:

$$e^{-jk\Delta l}=e^{-j\pi\nu^2/2} \tag{18.14.20}$$

which has the same form as in the diffraction coefficient $D_{\text{as}}(\nu)$. The values $\nu=u_n=\sqrt{2n}$ will make the path difference a multiple of $\lambda/2$, that is, $\Delta l=n\lambda/2$, resulting in the alternating phase $e^{-jk\Delta l}=(-1)^n$.

The discrepancy between the choices ν_n and u_n arises from using $D(\nu)$ to find the alternating maxima, versus using the plain phase (18.14.20). □

The Fresnel approximation is not invariant under shifting the origin. Our choice of origin above is not convenient because it depends on the observation point P_2. If we choose a fixed origin, such as the point O in Fig. 18.14.4, then, we must determine the corresponding Fresnel coefficient.

We assume that the points P_1, P_2 lie on the yz plane and take P_2 to lie in the shadow region. The angles θ_1, θ_2 may be chosen to be positive or negative to obtain all possible locations of P_1, P_2 relative to the screen.

The diffraction coefficient is still given by Eq. (18.13.8) but with r_1, r_2 replaced by the distances l_1, l_2. The unit vectors towards P_1 and P_2 are:

$$\hat{\mathbf{l}}_1=-\hat{\mathbf{z}}\cos\theta_1-\hat{\mathbf{y}}\sin\theta_1,\qquad \hat{\mathbf{l}}_2=\hat{\mathbf{z}}\cos\theta_2-\hat{\mathbf{y}}\sin\theta_2 \tag{18.14.21}$$

Since $\boldsymbol{r}'=x'\hat{\mathbf{x}}+y'\hat{\mathbf{y}}$ and $\hat{\mathbf{n}}=\hat{\mathbf{z}}$, we find:

$$\hat{\mathbf{l}}_1\cdot\boldsymbol{r}'=-y'\sin\theta_1,\quad \hat{\mathbf{l}}_2\cdot\boldsymbol{r}'=-y'\sin\theta_2,\quad \hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_1=-\cos\theta_1,\quad \hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_2=\cos\theta_2$$

The quadratic approximation for the lengths R_1, R_2 gives, then:

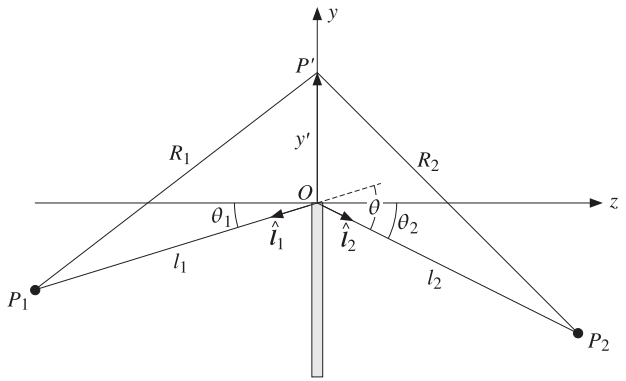


Fig. 18.14.4 Fresnel diffraction by straight edge.

$$R_1+R_2-l_1-l_2=-(\hat{\mathbf{l}}_1+\hat{\mathbf{l}}_2)\cdot\boldsymbol{\rho}'+\frac{1}{2}\left[\left(\frac{1}{l_1}+\frac{1}{l_2}\right)(\boldsymbol{\rho}'\cdot\boldsymbol{\rho}')-\frac{(\hat{\mathbf{l}}_1\cdot\boldsymbol{\rho}')^2}{l_1}-\frac{(\hat{\mathbf{l}}_2\cdot\boldsymbol{\rho}')^2}{l_2}\right]$$

$$=y'(\sin\theta_1+\sin\theta_2)+\left(\frac{1}{l_1}+\frac{1}{l_2}\right)\frac{x'^2}{2}+\left(\frac{\cos^2\theta_1}{l_1}+\frac{\cos^2\theta_2}{l_2}\right)\frac{y'^2}{2}$$

$$=\frac{1}{2F}x'^2+\frac{1}{2F'}[y'^2+2F'y'(\sin\theta_1+\sin\theta_2)]$$

$$=\frac{1}{2F}x'^2+\frac{1}{2F'}(y'+y_0)^2-\frac{1}{2F'}y_0^2$$

where we defined the focal lengths F, F' and the shift y_0:

$$\frac{1}{F}=\frac{1}{l_1}+\frac{1}{l_2},\quad \frac{1}{F'}=\frac{\cos^2\theta_1}{l_1}+\frac{\cos^2\theta_2}{l_2},\quad y_0=F'(\sin\theta_1+\sin\theta_2) \tag{18.14.22}$$

Using these approximations in Eq. (18.13.6) and replacing r_1, r_2 by l_1, l_2, we find:

$$E=\frac{jkA_1e^{-jk(l_1+l_2)}}{4\pi l_1l_2}[(\hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_1)]\int_S e^{-jk(R_1+R_2-l_1-l_2)}\,dS'$$

$$=\frac{jkA_1e^{-k(l_1+l_2)}}{4\pi l_1l_2}(\cos\theta_1+\cos\theta_2)e^{jky_0^2/2F'}\int e^{-jkx'^2/2F-jk(y'+y_0)^2/2F'}\,dx'dy'$$

The x'-integral is over the range $-\infty < x' < \infty$ and can be converted to a Fresnel integral with the change of variables $u = x'\sqrt{k/(\pi F)}$:

$$\int_{-\infty}^{\infty}e^{-jkx'^2/2F}\,dx'=\sqrt{\frac{\pi F}{k}}\int_{-\infty}^{\infty}e^{-j\pi u^2/2}\,du=\sqrt{\frac{\pi F}{k}}(1-j)$$

The y'-integral is over the upper-half of the xy-plane, that is, $0 \le y' < \infty$. Defining the Fresnel variables $u = (y'+y_0)\sqrt{k/(\pi F')}$ and $v = y_0\sqrt{k/(\pi F')}$, we find:

$$\int_0^{\infty}e^{-jk(y'+y_0)^2/2F'}\,dy'=\sqrt{\frac{\pi F'}{k}}\int_v^{\infty}e^{-j\pi u^2/2}\,du=\sqrt{\frac{\pi F'}{k}}(1-j)D(-v)$$

where the function $D(v)$ was defined in Eq. (18.14.1). Putting all the factors together, we may write the diffracted field at the point P_2 in the form:

$$\boxed{E=E_{\text{edge}}\frac{e^{-jkl_2}}{\sqrt{l_2}}D_{\text{edge}}}\qquad\text{(straight-edge diffraction)} \tag{18.14.23}$$

where we set $ky_0^2/2F' = \pi v^2/2$ and defined the incident field E_{edge} *at the edge* and the overall edge-diffraction coefficient D_{edge} by:

$$E_{\text{edge}}=A_1\frac{e^{-jkl_1}}{l_1},\qquad D_{\text{edge}}=\sqrt{\frac{FF'}{l_2}}\left(\frac{\cos\theta_1+\cos\theta_2}{2}\right)e^{j\pi v^2/2}D(-v) \tag{18.14.24}$$

The second factor $(e^{-jkl_2}/\sqrt{l_2})$ in (18.14.23) may be interpreted as a *cylindrical wave* emanating from the edge as a result of the incident field E_{edge}. The third factor D_{edge} is the angular gain of the cylindrical wave. The quantity v may be written as:

$$v=\sqrt{\frac{k}{\pi F'}}\,y_0=\sqrt{\frac{kF'}{\pi}}(\sin\theta_1+\sin\theta_2) \tag{18.14.25}$$

Depending on the sign and relative sizes of the angles θ_1 and θ_2, it follows that $v > 0$ when P_2 lies in the shadow region, and $v < 0$ when it lies in the illuminated region. For large positive v, we may use Eq. (18.14.5) to obtain the asymptotic form of the edge-diffraction coefficient D_{edge}:

$$D_{\text{edge}}=\sqrt{\frac{FF'}{l_2}}\,\frac{\cos\theta_1+\cos\theta_2}{2}\,e^{j\pi v^2/2}\,\frac{1-j}{2\pi v}\,e^{-j\pi v^2/2}=\sqrt{\frac{FF'}{l_2}}\,\frac{\cos\theta_1+\cos\theta_2}{2}\,\frac{1-j}{2\pi v}$$

Writing $\sqrt{F/l_2}=\sqrt{l_1/(l_1+l_2)}$ and replacing v from Eq. (18.14.25), the $\sqrt{F'}$ factor cancels and we obtain:

$$\boxed{D_{\text{edge}}=\sqrt{\frac{l_1}{l_1+l_2}}\,\frac{(1-j)(\cos\theta_1+\cos\theta_2)}{4\sqrt{\pi k}(\sin\theta_1+\sin\theta_2)}} \tag{18.14.26}$$

This expression may be simplified further by defining the overall diffraction angle $\theta = \theta_1 + \theta_2$, as shown in Fig. 18.14.4 and using the trigonometric identity:

$$\frac{\cos\theta_1+\cos\theta_2}{\sin\theta_1+\sin\theta_2}=\cot\left(\frac{\theta_1+\theta_2}{2}\right)$$

Then, Eq. (18.14.26) may be written in the form:

$$D_{\text{edge}}=\sqrt{\frac{l_1}{l_1+l_2}}\,\frac{(1-j)}{4\sqrt{\pi k}}\cot\frac{\theta}{2} \tag{18.14.27}$$

The asymptotic diffraction coefficient is obtained from Eqs. (18.14.26) or (18.14.27) by taking the limit $l_1 \to \infty$, which gives $\sqrt{l_1/(l_1+l_2)} \to 1$. Thus,

$$D_{\text{edge}}=\frac{(1-j)(\cos\theta_1+\cos\theta_2)}{4\sqrt{\pi k}(\sin\theta_1+\sin\theta_2)}=\frac{(1-j)}{4\sqrt{\pi k}}\cot\frac{\theta}{2} \tag{18.14.28}$$

Eqs. (18.14.27) and (18.14.28) are equivalent to those given in [1302].

The two choices for the origin lead to two different expressions for the diffracted fields. However, the expressions agree near the forward direction, $\theta \simeq 0$. It is easily verified that both Eq. (18.14.1) and (18.14.27) lead to the same approximation for the diffracted field:

$$E = E_{\text{edge}} \frac{e^{-jkl_2}}{\sqrt{l_2}} \sqrt{\frac{l_1}{l_1 + l_2}} \frac{1 - j}{2\sqrt{\pi k}\,\theta} \tag{18.14.29}$$

18.15 Geometrical Theory of Diffraction

Geometrical theory of diffraction is an extension of geometrical optics [1398–1413]. It views diffraction as a local edge effect. In addition to the ordinary rays of geometrical optics, it postulates the existence of "diffracted rays" from edges. The diffracted rays can reach into shadow regions, where geometrical optics fails.

An incident ray at an edge generates an infinity of diffracted rays emanating from the edge having different angular gains given by a diffraction coefficient D_{edge}. An example of such a diffracted ray is given by Eq. (18.14.23).

The edge-diffraction coefficient D_{edge} depends on (a) the type of the incident wave, such as plane wave, or spherical, (b) the type and local geometry of the edge, such as a knife-edge or a wedge, and (c) the directions of the incident and diffracted rays.

The diffracted field and coefficient are usually taken to be in their asymptotic forms, like those of Eq. (18.15.25). The asymptotic forms are derived from certain exactly solvable canonical problems, such as a conducting edge, a wedge, and so on.

The first and most influential of all such problems was Sommerfeld's solution of a plane wave incident on a conducting half-plane [1287], and we discuss it below.

Fig. 18.15.1 shows a plane wave incident at an angle α on the conducting plane occupying half of the xz-plane for $x \geq 0$. The plane of incidence is taken to be the xy-plane. Because of the cylindrical symmetry of the problem, we may assume that there is no z-dependence and that the fields depend only on the cylindrical coordinates ρ, ϕ.

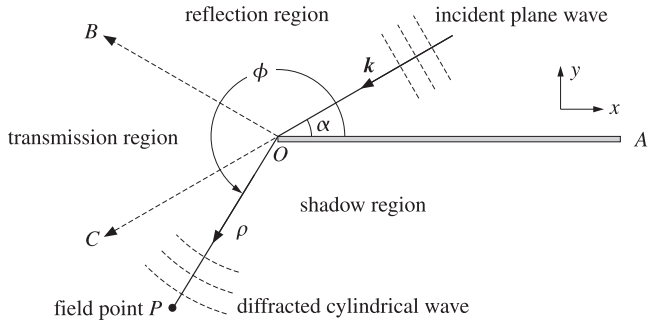


Fig. 18.15.1 Plane wave incident on conducting half-plane.

Two polarizations may be considered: TE, in which the electric field is $\boldsymbol{E} = \hat{\mathbf{z}}\, E_z$, and TM, which has $\boldsymbol{H} = \hat{\mathbf{z}}\, H_z$. Using cylindrical coordinates defined in Eq. (E.2) of Appendix E, and setting $\partial/\partial z = 0$, Maxwell's equations reduce in the two cases into:

$$\begin{aligned} &\text{(TE)} \quad \nabla^2 E_z + k^2 E_z = 0, \quad H_\rho = -\frac{1}{j\omega\mu}\frac{1}{\rho}\frac{\partial E_z}{\partial \phi}, \quad H_\phi = \frac{1}{j\omega\mu}\frac{\partial E_z}{\partial \rho} \\ &\text{(TM)} \quad \nabla^2 H_z + k^2 H_z = 0, \quad E_\rho = \frac{1}{j\omega\epsilon}\frac{1}{\rho}\frac{\partial H_z}{\partial \phi}, \quad E_\phi = -\frac{1}{j\omega\epsilon}\frac{\partial H_z}{\partial \rho} \end{aligned} \tag{18.15.1}$$

where $k^2 = \omega^2 \mu\epsilon$, and the two-dimensional ∇^2 is in cylindrical coordinates:

$$\nabla^2 = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial \phi^2} \tag{18.15.2}$$

The boundary conditions require that the tangential electric field be zero on both sides of the conducting plane, that is, for $\phi = 0$ and $\phi = 2\pi$. In the TE case, the tangential electric field is E_z, and in the TM case, $E_x = E_\rho \cos\phi - E_\phi \sin\phi = E_\rho = (1/j\omega\epsilon\rho)(\partial H_z/\partial \phi)$, for $\phi = 0, 2\pi$. Thus, the boundary conditions are:

$$\begin{aligned} &\text{(TE)} \quad E_z = 0, \quad \text{for } \phi = 0 \text{ and } \phi = 2\pi \\ &\text{(TM)} \quad \frac{\partial H_z}{\partial \phi} = 0, \quad \text{for } \phi = 0 \text{ and } \phi = 2\pi \end{aligned} \tag{18.15.3}$$

In Fig. 18.15.1, we assume that $0 \leq \alpha \leq 90^\circ$ and distinguish three wedge regions defined by the half-plane and the directions along the reflected and transmitted rays:

$$\begin{aligned} &\text{reflection region } (AOB): && 0 \leq \phi \leq \pi - \alpha \\ &\text{transmission region } (BOC): && \pi - \alpha \leq \phi \leq \pi + \alpha \\ &\text{shadow region } (COA): && \pi + \alpha \leq \phi \leq 2\pi \end{aligned} \tag{18.15.4}$$

The case when $90^\circ \leq \alpha \leq 180^\circ$ is shown in Fig. 18.15.2, in which α has been redefined to still be in the range $0 \leq \alpha \leq 90^\circ$. The three wedge regions are now:

$$\begin{aligned} &\text{reflection region } (AOB): && 0 \leq \phi \leq \alpha \\ &\text{transmission region } (BOC): && \alpha \leq \phi \leq 2\pi - \alpha \\ &\text{shadow region } (COA): && 2\pi - \alpha \leq \phi \leq 2\pi \end{aligned} \tag{18.15.5}$$

We construct the Sommerfeld solution in stages. We start by looking for solutions of the Helmholtz equation $\nabla^2 U + k^2 U = 0$ that have the factored form: $U = ED$, where E is also a solution, but a simple one, such as that of the incident plane wave. Using the differential identities of Appendix C, we have:

$$\nabla^2 U + k^2 U = D(\nabla^2 E + k^2 E) + E\nabla^2 D + 2\,\nabla E \cdot \nabla D$$

Thus, the conditions $\nabla^2 U + k^2 U = 0$ and $\nabla^2 E + k^2 E = 0$ require:

$$E\nabla^2 D + 2\,\nabla E \cdot \nabla D = 0 \quad \Rightarrow \quad \nabla^2 D + 2(\nabla \ln E)\cdot \nabla D = 0 \tag{18.15.6}$$

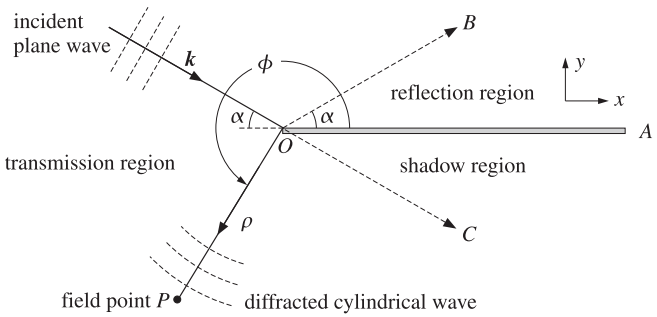


Fig. 18.15.2 Plane wave incident on conducting half-plane.

If we assume that E is of the form $E = e^{jf}$, where f is a real-valued function, then, equating to zero the real and imaginary parts of $\nabla^2 E + k^2 E = 0$, we find for f:

$$\nabla^2 E + k^2 E = E(k^2 - \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} f + j\nabla^2 f) = 0 \quad \Rightarrow \quad \nabla^2 f = 0\,, \quad \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} f = k^2 \tag{18.15.7}$$

Next, we assume that D is of the form:

$$D = D_0 \int_{-\infty}^{\nu} e^{-jg(u)}\, du \tag{18.15.8}$$

where D_0 is a constant, ν is a function of ρ, ϕ, and $g(u)$ is a real-valued function to be determined. Noting that $\boldsymbol{\nabla} D = D_0 e^{-jg}\, \boldsymbol{\nabla}\nu$ and $\boldsymbol{\nabla} g = g'(\nu)\, \boldsymbol{\nabla}\nu$, we find:

$$\boldsymbol{\nabla} D = D_0 e^{-jg}\, \boldsymbol{\nabla}\nu\,, \quad \nabla^2 D = D_0 e^{-jg}\,(\nabla^2\nu - jg'(\nu)\, \boldsymbol{\nabla}\nu \cdot \boldsymbol{\nabla}\nu)$$

Then, it follows from Eq. (18.15.6) that $\nabla^2 D + 2(\boldsymbol{\nabla} \ln E)\cdot \boldsymbol{\nabla} D = \nabla^2 D + j\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} D$ and:

$$\nabla^2 D + j\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} D = D_0 e^{-jg}\,[\nabla^2\nu + j(2\,\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla}\nu - g'\boldsymbol{\nabla}\nu \cdot \boldsymbol{\nabla}\nu)\,] = 0$$

Equating the real and imaginary parts to zero, we obtain the two conditions:

$$\nabla^2\nu = 0\,, \quad \frac{2\,\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla}\nu}{\boldsymbol{\nabla}\nu \cdot \boldsymbol{\nabla}\nu} = g'(\nu) \tag{18.15.9}$$

Sommerfeld's solution involves the Fresnel diffraction coefficient of Eq. (18.14.1), which can be written as follows:

$$\boxed{D(\nu) = \frac{1}{1-j}\left[\frac{1-j}{2} + \mathcal{F}(\nu)\right] = \frac{1}{1-j}\int_{-\infty}^{\nu} e^{-j\pi u^2/2} du} \tag{18.15.10}$$

Therefore, we are led to choose $g(u) = \pi u^2/2$ and $D_0 = 1/(1-j)$. To summarize, we may construct a solution of the Helmholtz equation in the form:

$$\nabla^2 U + k^2 U = 0\,, \qquad \boxed{U = ED = e^{jf} D(\nu)} \tag{18.15.11}$$

where f and ν must be chosen to satisfy the four conditions:

$$\begin{aligned} &\nabla^2 f = 0, \quad \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} f = k^2 \\ &\nabla^2 \nu = 0, \quad \frac{2\,\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla}\nu}{\boldsymbol{\nabla}\nu \cdot \boldsymbol{\nabla}\nu} = g'(\nu) = \pi\nu \end{aligned} \tag{18.15.12}$$

It can be verified easily that the functions $u = \rho^a \cos a\phi$ and $u = \rho^a \sin a\phi$ are solutions of the two-dimensional Laplace equation $\nabla^2 u = 0$, for *any* value of the parameter a. Taking f to be of the form $f = A\rho^a \cos a\phi$, we have the condition:

$$\boldsymbol{\nabla} f = Aa\rho^{a-1}[\hat{\boldsymbol{\rho}} \cos a\phi - \hat{\boldsymbol{\phi}} \sin a\phi] \quad \Rightarrow \quad \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} f = A^2 a^2 \rho^{2(a-1)} = k^2$$

This immediately implies that $a = 1$ and $A^2 = k^2$, so that $A = \pm k$. Thus, $f = A\rho\cos\phi = \pm k\rho\cos\phi$. Next, we choose $\nu = B\rho^a \cos a\phi$. Then:

$$\begin{aligned} \boldsymbol{\nabla} f &= A(\hat{\boldsymbol{\rho}} \cos\phi - \hat{\boldsymbol{\phi}} \sin\phi) \\ \boldsymbol{\nabla}\nu &= Ba\rho^{a-1}[\hat{\boldsymbol{\rho}} \cos a\phi - \hat{\boldsymbol{\phi}} \sin a\phi] \\ \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla}\nu &= ABa\rho^{a-1}[\cos\phi \cos a\phi + \sin\phi \sin a\phi] = ABa\rho^{a-1}\cos(\phi - a\phi) \\ \boldsymbol{\nabla}\nu \cdot \boldsymbol{\nabla}\nu &= B^2 a^2 \rho^{2(a-1)} \end{aligned}$$

Then, the last of the conditions (18.15.12) requires that:

$$\frac{1}{\pi\nu}\,\frac{2\,\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla}\nu}{\boldsymbol{\nabla}\nu \cdot \boldsymbol{\nabla}\nu} = \frac{2A\rho^{1-2a}\cos(\phi - a\phi)}{\pi a B^2 \cos a\phi} = 1$$

which implies that $a = 1/2$ and $B^2 = 2A/\pi a = 4A/\pi$. But since $A = \pm k$, only the case $A = k$ is compatible with a real coefficient B. Thus, we have $B^2 = 4k/\pi$, or, $B = \pm 2\sqrt{k/\pi}$.

In a similar fashion, we find that if we take $\nu = B\rho^a \sin a\phi$, then $a = 1/2$, but now $B^2 = -4A/\pi$, requiring that $A = -k$, and $B = \pm 2\sqrt{k/\pi}$. In summary, we have the following solutions of the conditions (18.15.12):

$$\begin{aligned} f &= +k\rho\cos\phi\,, \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2} \\ f &= -k\rho\cos\phi\,, \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2} \end{aligned} \tag{18.15.13}$$

The corresponding solutions (18.15.11) of the Helmholtz equation are:

$$\boxed{\begin{aligned} U(\rho,\phi) &= e^{jk\rho\cos\phi}\, D(\nu)\,, \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2} \\ U(\rho,\phi) &= e^{-jk\rho\cos\phi}\, D(\nu)\,, \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2} \end{aligned}} \tag{18.15.14}$$

The function $D(\nu)$ may be replaced by the equivalent form of Eq. (18.14.6) in order to bring out its asymptotic behavior for large ν:

$$U(\rho,\phi) = e^{jk\rho\cos\phi}\left[u(\nu)+d(\nu)\,e^{-j\pi\nu^2/2}\right], \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2}$$

$$U(\rho,\phi) = e^{-jk\rho\cos\phi}\left[u(\nu)+d(\nu)\,e^{-j\pi\nu^2/2}\right], \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2}$$

Using the trigonometric identities $\cos\phi = 2\cos^2(\phi/2)-1 = 1-2\sin^2(\phi/2)$, we find for the two choices of ν:

$$k\rho\cos\phi - \frac{1}{2}\pi\nu^2 = k\rho\left[\cos\phi - 2\cos^2\frac{\phi}{2}\right] = -k\rho$$

$$-k\rho\cos\phi - \frac{1}{2}\pi\nu^2 = -k\rho\left[\cos\phi + 2\sin^2\frac{\phi}{2}\right] = -k\rho$$

Thus, an alternative form of Eq. (18.15.14) is:

$$\boxed{\begin{aligned} U(\rho,\phi) &= e^{jk\rho\cos\phi}\,u(\nu)+e^{-jk\rho}\,d(\nu)\,, \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2} \\ U(\rho,\phi) &= e^{-jk\rho\cos\phi}\,u(\nu)+e^{-jk\rho}\,d(\nu)\,, \qquad \nu = \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2} \end{aligned}} \tag{18.15.15}$$

Shifting the origin of the angle ϕ still leads to a solution. Indeed, defining $\phi' = \phi \pm \alpha$, we note the property $\partial/\partial\phi' = \partial/\partial\phi$, which implies the invariance of the Laplace operator under this change. The functions $U(\rho,\phi\pm\alpha)$ are the elementary solutions from which the Sommerfeld solution is built.

Considering the TE case first, the incident plane wave in Fig. 18.15.1 is $\boldsymbol{E} = \hat{\mathbf{z}}E_i$, where $E_i = E_0 e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}$, with $\boldsymbol{r} = \hat{\mathbf{x}}\rho\cos\phi + \hat{\mathbf{y}}\rho\sin\phi$ and $\boldsymbol{k} = -k(\hat{\mathbf{x}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)$. It follows that:

$$\begin{aligned} \boldsymbol{k}\cdot\boldsymbol{r} &= -k\rho(\cos\phi\cos\alpha + \sin\phi\sin\alpha) = -k\rho\cos(\phi-\alpha) \\ E_i &= E_0 e^{-j\boldsymbol{k}\cdot\boldsymbol{r}} = E_0 e^{jk\rho\cos(\phi-\alpha)} \end{aligned} \tag{18.15.16}$$

The image of this electric field with respect to the perfect conducting plane will be the reflected field $E_r = -E_0 e^{-j\boldsymbol{k}_r\cdot\boldsymbol{r}}$, where $\boldsymbol{k}_r = k(-\hat{\mathbf{x}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)$, resulting in $E_r = -E_0 e^{jk\rho\cos(\phi+\alpha)}$. The sum $E_i + E_r$ does vanish for $\phi = 0$ and $\phi = 2\pi$, but it also vanishes for $\phi = \pi$. Therefore, it is an appropriate solution for a full conducting plane (the entire xz-plane), not for the half-plane.

Sommerfeld's solution, which satisfies the correct boundary conditions, is obtained by forming the linear combinations of the solutions of the type of Eq. (18.15.14):

$$\boxed{E_z = E_0\left[e^{jk\rho\cos\phi_i}\,D(\nu_i) - e^{jk\rho\cos\phi_r}\,D(\nu_r)\right]} \quad \text{(TE)} \tag{18.15.17}$$

where

$$\begin{aligned} \phi_i &= \phi - \alpha\,, \qquad \nu_i = 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi_i}{2} \\ \phi_r &= \phi + \alpha\,, \qquad \nu_r = 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi_r}{2} \end{aligned} \tag{18.15.18}$$

For the TM case, we form the sum instead of the difference:

$$\boxed{H_z = H_0\left[e^{jk\rho\cos\phi_i}\,D(\nu_i) + e^{jk\rho\cos\phi_r}\,D(\nu_r)\right]} \quad \text{(TM)} \tag{18.15.19}$$

The boundary conditions (18.15.3) are satisfied by both the TE and TM solutions. As we see below, the choice of the positive sign in the definitions of ν_i and ν_r was required in order to produce the proper diffracted field in the shadow region. Using the alternative forms (18.15.15), we separate the terms of the solution as follows:

$$\boxed{E_z = E_0 e^{jk\rho\cos\phi_i}\,u(\nu_i) - E_0 e^{jk\rho\cos\phi_r}\,u(\nu_r) + E_0 e^{-jk\rho}\left[d(\nu_i) - d(\nu_r)\right]} \tag{18.15.20}$$

The first two terms correspond to the incident and reflected fields. The third term is the diffracted field. The algebraic signs of ν_i and ν_r are as follows within the reflection, transmission, and shadow regions of Eq. (18.15.4):

reflection region:	$0 \le \phi < \pi - \alpha$,	$\nu_i > 0$,	$\nu_r > 0$
transmission region:	$\pi - \alpha < \phi < \pi + \alpha$,	$\nu_i > 0$,	$\nu_r < 0$
shadow region:	$\pi + \alpha < \phi \le 2\pi$,	$\nu_i < 0$,	$\nu_r < 0$

(18.15.21)

The unit-step functions will be accordingly present or absent resulting in the following fields in these three regions:

reflection region:	$E_z = E_i + E_r + E_d$
transmission region:	$E_z = E_i + E_d$
shadow region:	$E_z = E_d$

(18.15.22)

where we defined the incident, reflected, and diffracted fields:

$$\begin{aligned} E_i &= E_0 e^{jk\rho\cos\phi_i} \\ E_r &= -E_0 e^{jk\rho\cos\phi_r} \\ E_d &= E_0 e^{-jk\rho}\left[d(\nu_i) - d(\nu_r)\right] \end{aligned} \tag{18.15.23}$$

The diffracted field is present in all three regions, and in particular it is the only one in the shadow region. For large ν_i and ν_r (positive or negative), we may replace $d(\nu)$ by its asymptotic form $d(\nu) = -(1-j)/(2\pi\nu)$ of Eq. (18.14.7), resulting in the asymptotic diffracted field:

$$\begin{aligned} E_d &= -E_0 e^{-jk\rho}\,\frac{1-j}{2\pi}\left(\frac{1}{\nu_i} - \frac{1}{\nu_r}\right) \\ &= -E_0 e^{-jk\rho}\,\frac{1-j}{2\pi 2\sqrt{k/\pi}\,\rho^{1/2}}\left(\frac{1}{\cos(\phi_i/2)} - \frac{1}{\cos(\phi_r/2)}\right) \end{aligned}$$

which can be written in the form:

$$\boxed{E_d = E_0\,\frac{e^{-jk\rho}}{\rho^{1/2}}\,D_{\text{edge}}} \tag{18.15.24}$$

with an edge-diffraction coefficient:

$$\boxed{D_{\text{edge}} = -\frac{1-j}{4\sqrt{\pi k}}\left(\frac{1}{\cos\dfrac{\phi_i}{2}} - \frac{1}{\cos\dfrac{\phi_r}{2}}\right)} \tag{18.15.25}$$

Using a trigonometric identity, we may write D_{edge} as follows:

$$D_{\text{edge}} = -\frac{1-j}{4\sqrt{\pi k}}\left(\frac{1}{\cos\dfrac{\phi-\alpha}{2}} - \frac{1}{\cos\dfrac{\phi+\alpha}{2}}\right) = -\frac{1-j}{\sqrt{\pi k}}\,\frac{\sin\dfrac{\phi}{2}\sin\dfrac{\alpha}{2}}{\cos\phi+\cos\alpha} \tag{18.15.26}$$

Eqs. (18.15.22) and (18.15.24) capture the essence of the geometrical theory of diffraction: In addition to the ordinary incident and reflected geometric optics rays, one also has diffracted rays in all directions corresponding to a cylindrical wave emanating from the edge with a directional gain of D_{edge}.

For the case of Fig. 18.15.2, the incident and reflected plane waves have propagation vectors $\boldsymbol{k} = k(\hat{\mathbf{z}}\cos\alpha - \hat{\mathbf{y}}\sin\alpha)$ and $\boldsymbol{k}_r = k(\hat{\mathbf{z}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)$. These correspond to the incident and reflected fields:

$$E_i = E_0 e^{-j\boldsymbol{k}\cdot\boldsymbol{r}} = E_0 e^{-jk\rho\cos(\phi+\alpha)}\,, \quad E_r = -E_0 e^{-j\boldsymbol{k}_r\cdot\boldsymbol{r}} = -E_0 e^{-jk\rho\cos(\phi-\alpha)}$$

In this case, the Sommerfeld TE and TM solutions take the form:

$$\boxed{\begin{aligned} E_z &= E_0\left[e^{-jk\rho\cos\phi_i}\,D(\nu_i) - e^{-jk\rho\cos\phi_r}\,D(\nu_r)\right] \\ H_z &= H_0\left[e^{-jk\rho\cos\phi_i}\,D(\nu_i) + e^{-jk\rho\cos\phi_r}\,D(\nu_r)\right] \end{aligned}} \tag{18.15.27}$$

where, now:

$$\begin{aligned} \phi_i &= \phi + \alpha\,, \qquad \nu_i = 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi_i}{2} \\ \phi_r &= \phi - \alpha\,, \qquad \nu_r = -2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi_r}{2} \end{aligned} \tag{18.15.28}$$

The choice of signs in ν_i and ν_r are such that they are both negative within the shadow region defined by Eq. (18.15.5). The same solution can also be obtained from Fig. 18.15.1 and Eq. (18.15.17) by replacing α by $\pi - \alpha$.

18.16 Problems

18.1 Show that Eq. (18.4.9) can be written in the compact vectorial form:

$$\boldsymbol{E} = -jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{z}}\times\boldsymbol{f} - \eta\,\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{g})]\,, \quad \boldsymbol{H} = -\frac{jk}{\eta}\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{f}) + \eta\,\hat{\mathbf{z}}\times\boldsymbol{g}]$$

Similarly, show that Eqs. (18.4.10) and (18.4.11) can be written as:

$$\begin{aligned} \boldsymbol{E} &= -2jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{z}}\times\boldsymbol{f}]\,, & \boldsymbol{H} &= -\frac{2jk}{\eta}\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{f})] \\ \boldsymbol{E} &= 2jk\eta\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{g})]\,, & \boldsymbol{H} &= -2jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{z}}\times\boldsymbol{g}] \end{aligned}$$

18.2 Prove the first pair of equations for $\boldsymbol{E}, \boldsymbol{H}$ of the previous problem by working exclusively with the Kottler formulas (18.4.2) and taking their far-field limits.

18.3 Explain in detail how the inequality (18.6.12) for the aperture efficiency e_a may be thought of as an example of the Schwarz inequality. Then, using standard properties of Schwarz inequalities, prove that the maximum of e_a is unity and is achieved for uniform apertures. As a reminder, the Schwarz inequality for single-variable complex-valued functions is:

$$\left|\int_a^b f^*(x)\,g(x)\,dx\right|^2 \le \int_a^b |f(x)|^2\,dx \cdot \int_a^b |g(x)|^2\,dx$$

18.4 To prove the equivalence of the Kirchhoff diffraction and Stratton-Chu formulas, (18.10.6) and (18.10.7), use the identities (C.29) and (C.32) of Appendix C, to obtain:

$$\begin{aligned} \int_V \left[j\omega\mu\,G\,\boldsymbol{J} + \frac{1}{\epsilon}G\,\nabla'\rho + G\,\nabla'\times\boldsymbol{J}_m\right]dV' &= \int_V \left[j\omega\mu\,G\,\boldsymbol{J} - \frac{\rho}{\epsilon}\,\nabla' G + \boldsymbol{J}_m\times\nabla' G\right]dV' \\ &\quad - \oint_S \left[\hat{\mathbf{n}}\,\frac{\rho}{\epsilon}\,G + \hat{\mathbf{n}}\times\boldsymbol{J}_m G\right]dS' \end{aligned}$$

Then, using the identity (C.33), show that Eq. (18.10.6) can be rewritten in the form:

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{r}) = &-\int_V \left[j\omega\mu\,G\,\boldsymbol{J} - \frac{\rho}{\epsilon}\,\nabla' G + \boldsymbol{J}_m\times\nabla' G\right]dV' \\ &+ \oint_S \left[\hat{\mathbf{n}}\,\frac{\rho}{\epsilon}\,G + \hat{\mathbf{n}}\times\boldsymbol{J}_m G\right]dS' \\ &- \oint_S \left[\hat{\mathbf{n}}\,G\,\nabla'\cdot\boldsymbol{E} - (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G - G\,\hat{\mathbf{n}}\times(\nabla'\times\boldsymbol{E}) - (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla' G\right]dS' \end{aligned}$$

Finally, use $\rho/\epsilon = \nabla'\cdot\boldsymbol{E}$ and $\nabla'\times\boldsymbol{E} + \boldsymbol{J}_m = -j\omega\mu\boldsymbol{H}$ to obtain (18.10.7).

18.5 Prove the equivalence of the Stratton-Chu and Kottler formulas, (18.10.7) and (18.10.10), by first proving and then using the following dual relationships:

$$\begin{aligned} \int_V [j\omega\rho\nabla' G - (\boldsymbol{J}\cdot\nabla')\nabla' G]\,dV' &= \oint_S [((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla')\nabla' G - j\omega\epsilon(\hat{\mathbf{n}}\cdot\boldsymbol{E})\nabla' G] \\ \int_V [j\omega\rho_m\nabla' G - (\boldsymbol{J}_m\cdot\nabla')\nabla' G]\,dV' &= -\oint_S [((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla')\nabla' G + j\omega\mu(\hat{\mathbf{n}}\cdot\boldsymbol{H})\nabla' G] \end{aligned}$$

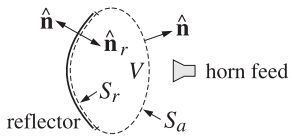
To prove these, work component-wise, use Maxwell's equations (18.2.1), and apply the divergence theorem on the volume V of Fig. 18.10.1.

18.6 Prove the equivalence of the Kottler and Franz formulas, (18.10.10) and (18.10.11), by using the identity $\nabla\times(\nabla\times\boldsymbol{A}) = \nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A}$, and by replacing the quantity $k^2 G(\boldsymbol{r}-\boldsymbol{r}')$ by $-\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}') - \nabla'^2 G$. Argue that the term $\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}')$ makes a difference only for the volume integrals, but not for the surface integrals.

18.7 Prove the equivalence of the modified Stratton-Chu and Kirchhoff diffraction integral formulas of Eq. (18.12.1) and (18.12.2) by using the identity (C.42) of Appendix C and replacing $\nabla'\cdot\boldsymbol{E} = 0$ and $\nabla'\times\boldsymbol{E} = -j\omega\mu\boldsymbol{H}$ in the source-less region under consideration.

18.8 Prove the equivalence of the Kottler and modified Stratton-Chu formulas of Eq. (18.12.1) and (18.12.2) by subtracting the two expressions, replacing $j\omega\epsilon\boldsymbol{E} = \nabla'\times\boldsymbol{H}$, and using the Stokes identity (C.38) of Appendix C.

18.9 Consider a reflector antenna fed by a horn, as shown on the right. A closed surface $S = S_r + S_a$ is such that the portion S_r caps the reflector and the portion S_a is an aperture in front of the reflector. The feed lies outside the closed surface, so that the volume V enclosed by S is free of current sources.



Applying the Kottler version of the extinction theorem of Sec. 18.10 on the volume V, show that for points $\boldsymbol{r}$ outside V, the field radiated by the induced surface currents on the reflector S_r is equal to the field radiated by the aperture fields on S_a, that is,

$$\begin{aligned}\boldsymbol{E}_{\text{rad}}(\boldsymbol{r}) &= \frac{1}{j\omega\epsilon}\int_{S_r}\left[k^2 G\,\boldsymbol{J}_s + (\boldsymbol{J}_s\cdot\boldsymbol{\nabla}')\boldsymbol{\nabla}' G\right] dS' \\ &= \frac{1}{j\omega\epsilon}\int_{S_a}\left[k^2 G(\hat{\mathbf{n}}\times\boldsymbol{H}) + \left((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\boldsymbol{\nabla}'\right)\boldsymbol{\nabla}' G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\boldsymbol{\nabla}' G\right] dS'\end{aligned}$$

where the induced surface currents on the reflector are $\boldsymbol{J}_s = \hat{\mathbf{n}}_r\times\boldsymbol{H}$ and $\boldsymbol{J}_{ms} = -\hat{\mathbf{n}}_r\times\boldsymbol{E}$, and on the perfectly conducting reflector surface, we must have $\boldsymbol{J}_{ms} = 0$.

This result establishes the equivalence of the so-called aperture-field and current-distribution methods for reflector antennas [1686]. See also Sec. 21.9.

18.10 Consider an x-polarized uniform plane wave incident obliquely on the straight-edge aperture of Fig. 18.14.4, with a wave vector direction $\hat{\mathbf{k}}_1 = \hat{\mathbf{z}}\cos\theta_1 + \hat{\mathbf{y}}\sin\theta_1$. First show that the tangential fields at an aperture point $\boldsymbol{r}' = x'\,\hat{\mathbf{x}} + y'\,\hat{\mathbf{y}}$ on the aperture above the straight-edge are given by:

$$\boldsymbol{E}_a = \hat{\mathbf{x}}\,E_0 e^{-jky'\sin\theta_1}, \qquad \boldsymbol{H}_a = \hat{\mathbf{y}}\,\frac{E_0}{\eta_0}\cos\theta_1 e^{-jky'\sin\theta_1}$$

Then, using Kottler's formula (18.12.1), and applying the usual Fresnel approximations in the integrand, as was done for the point source in Fig. 18.14.4, show that the diffracted wave below the edge is given by Eqs. (18.14.23)–(18.14.25), except that the field at the edge is $E_{\text{edge}} = E_0$, and the focal lengths are in this case $F = l_2$ and $F' = l_2/\cos^2\theta_2$

Finally, show that the asymptotic diffracted field (when $l_2 \to \infty$), is given near the forward direction $\theta \simeq 0$ by:

$$E = E_{\text{edge}}\,\frac{e^{-jkl_2}}{\sqrt{l_2}}\,\frac{1-j}{2\sqrt{\pi k\,\theta}}$$

18.11 Assume that the edge in the previous problem is a perfectly conducting screen. Using the field-equivalence principle with effective current densities on the aperture above the edge $\boldsymbol{J}_s = 0$ and $\boldsymbol{J}_{ms} = -2\hat{\mathbf{n}}\times\boldsymbol{E}_a$, and applying the usual Fresnel approximations, show that the diffracted field calculated by Eq. (18.4.1) is still given by Eqs. (18.14.23)–(18.14.25), except that the factor $\cos\theta_1 + \cos\theta_2$ is replaced now by $2\cos\theta_2$, and that the asymptotic field and edge-diffraction coefficient are:

$$E = E_0\,\frac{e^{-jkl_2}}{\sqrt{l_2}}\,D_{\text{edge}}, \qquad D_{\text{edge}} = \frac{(1-j)\,2\cos\theta_2}{4\sqrt{\pi k}\,(\sin\theta_1 + \sin\theta_2)}$$

Show that this expression agrees with the exact Sommerfeld solution (18.15.26) at normal incidence and near the forward diffracted direction.

19

Diffraction – Plane-Wave Spectrum

This chapter continues the previous one on radiation from apertures. The emphasis is on Rayleigh-Sommerfeld diffraction theory, plane-wave spectrum representation for scalar and vector fields, radiated and reactive power of apertures, integral equations for apertures in conducting screens, revisiting the Sommerfeld half-plane problem using Wiener-Hopf factorization techniques, the Bethe-Bouwkamp model of diffraction by small holes, and the Babinet principle.

19.1 Rayleigh-Sommerfeld Diffraction Theory

In this section, we recast Kirchhoff's diffraction formula in a form that uses a Dirichlet Green's function (i.e., one that vanishes on the boundary surface) and obtain the Rayleigh-Sommerfeld diffraction formula. In the next section, we show that this reformulation is equivalent to the *plane-wave spectrum* approach to diffraction, and use it to justify the modified forms (18.1.2) and (18.1.3) of the field equivalence principle. In Chap. 20, we use it to obtain the usual Fresnel and Fraunhofer approximations and discuss a few applications from Fourier optics.

We will work with the scalar case, but the same method can be used for the vector case. With reference to Fig. 19.1.1, we consider a scalar field $E(\boldsymbol{r})$ that satisfies the source-free Helmholtz equation, $(\nabla^2 + k^2)E(\boldsymbol{r}) = 0$, over the right half-space $z \geq 0$.

We consider a closed surface consisting of the surface S_∞ of a sphere of very large radius centered at the observation point $\boldsymbol{r}$ and bounded on the left by its intersection S with the xy plane, as shown in the Fig. 19.1.1. Clearly, in the limit of infinite radius, the volume V bounded by $S + S_\infty$ is the right half-space $z \geq 0$, and S becomes the entire xy plane. Applying Eq. (18.10.3) to volume V, we have:

$$\int_V \left[G(\nabla'^2 E + k^2 E) - E\,(\nabla'^2 G + k^2 G)\right] dV' = -\oint_{S+S_\infty}\left[G\,\frac{\partial E}{\partial n'} - E\,\frac{\partial G}{\partial n'}\right] dS' \quad (19.1.1)$$

The surface integral over S_∞ can be ignored by noting that $\hat{\mathbf{n}}$ is the negative of the radial unit vector and therefore, we have after adding and subtracting the term $jkEG$:

$$-\int_{S_\infty}\left[G\,\frac{\partial E}{\partial n'} - E\,\frac{\partial G}{\partial n'}\right] dS' = \int_{S_\infty}\left[G\left(\frac{\partial E}{\partial r} + jkE\right) - E\left(\frac{\partial G}{\partial r} + jkG\right)\right] dS'$$