20

Diffraction – Fourier Optics

This chapter continues the discussion of diffraction and emphasizes Fourier optics topics, Fresnel and Fraunhofer approximations, Talbot effect, Fourier properties of lens systems, 1-D and 2-D apodizer design, aperture synthesis concentrating on narrow beamwidths and low sidelobe designs, and concluding with a discussion of superresolution, superdirectivity, and superoscillations.

20.1 Fresnel Approximation

The Fresnel approximation for planar apertures is obtained from the Rayleigh-Sommerfeld formula (19.1.6). Using (19.1.7), we have:

$$E(\boldsymbol{r}_\perp, z) = \int_S E(\boldsymbol{r}'_\perp, 0)\, \frac{2z}{R}\left(jk + \frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R}\, d^2\boldsymbol{r}'_\perp \tag{20.1.1}$$

where $R = \sqrt{(x-x')^2+(y-y')^2+z^2} = \sqrt{|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2 + z^2}$. The Fresnel approximation assumes that z is large enough such that $|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp| \ll z$, which can be realized if the aperture has dimension d so that $|\boldsymbol{r}'_\perp| < d$, and one assumes that the observation point $\boldsymbol{r}_\perp$ remains close to the z-axis (the paraxial approximation) such that $|\boldsymbol{r}_\perp| < d$, and z is chosen such that $z \gg d$. Then, we can approximate R as follows:

$$R = \sqrt{|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2 + z^2} = z\sqrt{1 + \frac{|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2}{z^2}} \simeq z\left[1 + \frac{1}{2}\frac{|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2}{z^2}\right] = z + \frac{|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2}{2z}$$

where we used the Taylor series expansion $\sqrt{1+x} \simeq 1 + x/2$. Assuming also that R or z is much greater than the wavelength of the wave, $z \gg \lambda$, so that $k \gg 1/R$ we obtain,

$$\frac{2z}{R}\left(jk + \frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R} \simeq jk\frac{e^{-jk(z+|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2z)}}{2\pi z} = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2z} \tag{20.1.2}$$

where we set $R \simeq z$ in the amplitude factors, but kept the quadratic approximation in the phase e^{-jkR}. The Fresnel approximation is finally:

$$\boxed{E(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz} \int_S E(\boldsymbol{r}'_\perp, 0)\, e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2z}\, d^2\boldsymbol{r}'_\perp} \quad \text{(Fresnel)} \tag{20.1.3}$$

This amounts to replacing the propagator impulse response $h(\boldsymbol{r}_\perp, z)$ of Eq. (19.2.9) by the approximation of Eq. (20.1.2):

$$\boxed{h(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z} = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk(x^2+y^2)/2z}} \quad \text{(Fresnel)} \tag{20.1.4}$$

Noting that $k = 2\pi/\lambda$, the constant factor in front is often written as:

$$\frac{jk}{2\pi z} = \frac{j}{\lambda z}$$

The accuracy of the Fresnel approximation can be quantified by considering the higher-order terms in the expansion of the square root in,

$$kR = k\sqrt{z^2 + |\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2} \approx kz + \frac{k|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2}{2z} - \frac{k|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^4}{8z^3} + \cdots$$

The approximation will be accurate if the third term is small. For an aperture of typical size D, the quantity $|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|$ remains of the order of D, thus, we obtain the following condition on the distance z at which the Fresnel approximation is accurate,

$$\frac{kD^4}{8z^3} \ll 1 \quad \Rightarrow \quad z^3 \gg \frac{1}{8}kD^4 \quad \Rightarrow \quad kz \gg \frac{1}{2}(kD)^{4/3} \tag{20.1.5}$$

The Fresnel approximation can also be understood from the plane-wave spectrum point of view. The Fourier transform of (20.1.4) is obtained from the following Fourier integral, which is a special case of Eq. (3.5.18):

$$\sqrt{\frac{jk}{2\pi z}}\int_{-\infty}^{\infty} e^{jk_x x}\, e^{-jkx^2/2z}\, dx = e^{jk_x^2 z/2k} \tag{20.1.6}$$

where $\sqrt{j}$ should be understood as $e^{j\pi/4}$. More generally,

$$\sqrt{\frac{jk}{2\pi z}}\int_{-\infty}^{\infty} e^{jk_x x}\, e^{-jk(x-x_0)^2/2z}\, dx = e^{jk_x x_0}\, e^{jk_x^2 z/2k} \tag{20.1.7}$$

Applying (20.1.6) with respect to the x and y integrations, we obtain the two-dimensional spatial Fourier transform of $h(\boldsymbol{r}_\perp, z)$:

$$\hat{h}(\boldsymbol{k}_\perp, z) = \int_{-\infty}^{\infty} h(\boldsymbol{r}_\perp, z)\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, d^2\boldsymbol{r}_\perp = \frac{jk}{2\pi z}\, e^{-jkz}\int_{-\infty}^{\infty} e^{-jk|\boldsymbol{r}_\perp|^2/2z}\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, d^2\boldsymbol{r}_\perp\,, \quad \text{or,}$$

$$\boxed{\hat{h}(\boldsymbol{k}_\perp, z) = e^{-jkz}\, e^{j|\boldsymbol{k}_\perp|^2 z/2k} = e^{-jkz}\, e^{j(k_x^2+k_y^2)z/2k}} \tag{20.1.8}$$

Then, Eq. (20.1.3) can be written in the wavenumber domain as,

$$E(\boldsymbol{r}_\perp, z) = \int_{-\infty}^{\infty} \hat{E}(\boldsymbol{k}_\perp, z)\; e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} = \int_{-\infty}^{\infty} \hat{E}(\boldsymbol{k}_\perp, 0)\, \hat{h}(\boldsymbol{k}_\perp.z)\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2}$$

with,

$$E(\boldsymbol{r}_\perp, 0) = \int_{-\infty}^{\infty} \hat{E}(\boldsymbol{k}_\perp, 0)\; e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \quad \Leftrightarrow \quad \hat{E}(\boldsymbol{k}_\perp, 0) = \int_{-\infty}^{\infty} E(\boldsymbol{r}_\perp, 0)\; e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, d^2\boldsymbol{r}_\perp$$

Thus, we may write the Fresnel diffraction formula convolutionally and in its plane-wave spectrum form, where we expanded, $|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2 = |\boldsymbol{r}_\perp|^2 + |\boldsymbol{r}'_\perp|^2 - 2\boldsymbol{r}_\perp \cdot \boldsymbol{r}'_\perp$, in the third expression,

$$\begin{aligned} E(\boldsymbol{r}_\perp, z) &= e^{-jkz} \int_{-\infty}^{\infty} \hat{E}(\boldsymbol{k}_\perp, 0)\, e^{j|\boldsymbol{k}_\perp|^2 z/2k}\, e^{-j\boldsymbol{k}_\perp \cdot \boldsymbol{r}_\perp} \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \\ &= \frac{jk}{2\pi z}\, e^{-jkz} \int_S E(\boldsymbol{r}'_\perp, 0)\, e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2z}\, d^2\boldsymbol{r}'_\perp \\ &= \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z} \int_S E(\boldsymbol{r}'_\perp, 0)\, e^{-jk|\boldsymbol{r}'_\perp|^2/2z}\, e^{jk\boldsymbol{r}_\perp \cdot \boldsymbol{r}'_\perp/z}\, d^2\boldsymbol{r}'_\perp \end{aligned} \quad \text{(Fresnel)} \tag{20.1.9}$$

The last integral is recognized as the two-dimensional Fourier transform of the product, $E(\boldsymbol{r}'_\perp, 0)\, e^{-jk|\boldsymbol{r}'_\perp|^2/2z}$, evaluated at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/z$.

Eq. (20.1.8) can also be obtained from the exact form $\hat{h}(\boldsymbol{k}_\perp, z) = e^{-jk_z z}$ by assuming that for large z the evanescent modes will have already decayed and assuming the approximation $k_x^2 + k_y^2 \ll k^2$ for the propagating modes. Then, we can write:

$$k_z = \sqrt{k^2 - |\boldsymbol{k}_\perp|^2} = k\sqrt{1 - \frac{|\boldsymbol{k}_\perp|^2}{k^2}} \simeq k\left[1 - \frac{1}{2}\frac{|\boldsymbol{k}_\perp|^2}{k^2}\right] = k - \frac{|\boldsymbol{k}_\perp|^2}{2k}$$

and, hence

$$e^{-jk_z z} \simeq e^{-jkz}\, e^{j|\boldsymbol{k}_\perp|^2 z/2k}$$

Because of the assumption $|\boldsymbol{k}_\perp| < k$, the maximum transverse wavenumber will be $|\boldsymbol{k}_\perp| = k = 2\pi/\lambda$, and correspondingly the smallest achievable transverse spatial resolution will be $\Delta r_\perp \sim 1/|\boldsymbol{k}_\perp| \sim \lambda$, that is, about one wavelength. This is the basic diffraction limit of optical instruments, such as lenses and microscopes. Implicit in Eq. (20.1.9) then, is that the range of integration in $\boldsymbol{k}_\perp$-space must be restricted to the propagating range, $|\boldsymbol{k}_\perp| \le k$.

Near-field optics methods [534–553], where the evanescent modes are not ignored, overcome this limitation and can achieve much higher, subwavelength, resolutions. Although ordinary lenses are diffraction-limited, it has been shown recently [398] that "superlenses" made from metamaterials having negative refractive index can achieve perfect resolution.

Circularly Symmetric and One-Dimensional Cases

In the special case when the fields have circular symmetry, as it happens often when working with circular apertures and scalar fields, the above expressions can be written as 0th-order Hankel transforms, i.e., with respect to the J_0 Bessel function.

Introducing cylindrical coordinates in the space and wavenumber domains, similar to Eqs. (19.11.1) and (19.12.8) of Chap. 19,

$$\begin{aligned} r_\perp &= |\boldsymbol{r}_\perp| = \sqrt{x^2 + y^2} \\ \phi &= \text{atan2}(y, x) \\ k_\perp &= |\boldsymbol{k}_\perp| = \sqrt{k_x^2 + k_y^2} \\ \psi &= \text{atan2}(k_y, k_x) \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} x &= r_\perp \cos\phi \\ y &= r_\perp \sin\phi \\ k_x &= k_\perp \cos\psi \\ k_y &= k_\perp \sin\psi \end{aligned} \tag{20.1.10}$$

and assuming that $E(\boldsymbol{r}_\perp, 0)$ depends only on, $r_\perp = |\boldsymbol{r}_\perp|$, all angular integrations can be done with the help of the following integral of Eq. (19.12.22),

$$\int_0^{2\pi} e^{jk_\perp r_\perp \cos(\psi - \phi)}\, \frac{d\psi}{2\pi} = J_0(k_\perp r_\perp) \tag{20.1.11}$$

then, $\hat{E}(\boldsymbol{k}_\perp, 0)$ will depend only on $k_\perp = |\boldsymbol{k}_\perp|$, and will satisfy the following Hankel transform relationships,

$$\begin{aligned} \hat{E}(k_\perp, 0) &= \int_0^\infty E(r_\perp, 0)\, J_0(k_\perp r_\perp)\, 2\pi\, r_\perp dr_\perp \\ E(r_\perp, 0) &= \int_0^\infty \hat{E}(k_\perp, 0)\, J_0(k_\perp r_\perp)\, \frac{k_\perp dk_\perp}{2\pi} \end{aligned} \tag{20.1.12}$$

Then, the cylindrical symmetry is preserved by the propagation process, and (20.1.9) can be written as follows, where we replaced, $\boldsymbol{r}_\perp \cdot \boldsymbol{r}'_\perp = r_\perp r'_\perp \cos(\phi - \phi')$, and the ϕ' integration was done with (20.1.11),

$$\begin{aligned} E(r_\perp, z) &= e^{-jkz} \int_0^\infty \hat{E}(k_\perp, 0)\, e^{jk_\perp^2 z/2k}\, J_0(k_\perp r_\perp)\, \frac{k_\perp dk_\perp}{2\pi} \\ &= \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jkr_\perp^2/2z} \int_0^\infty E(r'_\perp, 0)\, e^{-jkr'^2_\perp/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp \end{aligned} \tag{20.1.13}$$

Another special case is when the aperture field $E(x', y', 0)$ depends only on one transverse coordinate, say, $E(x', 0)$. Then, the dependence of (20.1.3) on the y direction can be integrated out using the integral

$$\sqrt{\frac{jk}{2\pi z}} \int_{-\infty}^{\infty} e^{-jk(y-y')^2/2z}\, dy' = 1 \tag{20.1.14}$$

and we obtain the following 1-D version of the propagation impulse response,

$$h(x, z) = \int_{-\infty}^{\infty} h(\boldsymbol{r}_\perp, z)\, dz = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jkx^2/2z} \int_{-\infty}^{\infty} e^{-jky^2/2z}\, dy, \quad \text{or,}$$

$$h(x, z) = \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz}\, e^{-jkx^2/2z} \quad \Leftrightarrow \quad \hat{h}(k_x, z) = e^{-jkz}\, e^{jk_x^2 z/2k} \tag{20.1.15}$$

and the corresponding one-dimensional Fresnel formula, written convolutionally and in its plane-wave spectrum form,

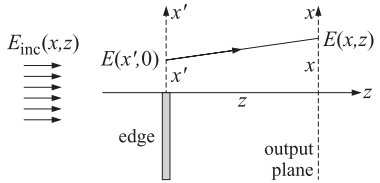
$$\begin{aligned} E(x,z) &= e^{-jkz}\int_{-\infty}^{\infty}\hat{E}(k_x,0)\,e^{jk_x^2 z/2k}\,e^{-jk_x x}\,\frac{dk_x}{2\pi} \\ &= \sqrt{\frac{jk}{2\pi z}}\,e^{-jkz}\int_{-\infty}^{\infty}E(x',0)\,e^{-jk(x-x')^2/2z}\,dx' \\ &= \sqrt{\frac{jk}{2\pi z}}\,e^{-jkz}\,e^{-jkx^2/2z}\int_{-\infty}^{\infty}E(x',0)\,e^{-jkx'^2/2z}\,e^{jkxx'/z}\,dx \end{aligned} \tag{20.1.16}$$

where again, the k_x-integration may be restricted to $|k_x| \le k$ for distances z that are large enough so that the evanescent modes have decayed.

Example 20.1.1: *Knife-Edge Diffraction.* We revisit the problem of knife-edge diffraction using the Fresnel formula (20.1.16). The infinite edge is along the y direction and it occupies the region $x < 0$, as shown in the figure below. The incident plane-wave field and the diffracted field at distance z are:

$$E_{\text{inc}}(x,z) = E_0 e^{-jkz}$$

$$E(x,z) = \sqrt{\frac{jk}{2\pi z}}\,e^{-jkz}\int_{-\infty}^{\infty}E(x',0)\,e^{-jk(x-x')^2/2z}\,dx'$$



At the input plane, $E(x',0) = E_0$, for $x' \ge 0$, and $E(x',0) = 0$, for $x' < 0$. Then, the above integral becomes:

$$E(x,z) = E_0 e^{-jkz}\sqrt{\frac{jk}{2\pi z}}\int_0^{\infty}e^{-jk(x-x')^2/2z}\,dx'$$

Making the change of variables,

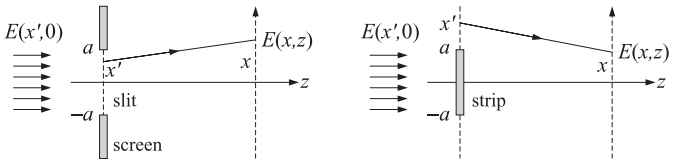
$$\sqrt{\frac{k}{2z}}\,(x'-x) = \sqrt{\frac{\pi}{2}}\,u\,, \qquad \nu = \sqrt{\frac{k}{\pi z}}\,x$$

the above integral can be reduced to the Fresnel integral $\mathcal{F}(x)$ of Appendix F:

$$E(x,z) = E_0 e^{-jkz}\sqrt{\frac{j}{2}}\int_{-\nu}^{\infty}e^{-j\pi u^2/2}\,du = E_0 e^{-jkz}\,\frac{1}{1-j}\left[\mathcal{F}(\nu)+\frac{1-j}{2}\right]$$

This is identical (up to the paraxial assumption) to the case discussed in Sec. 18.14. When $x < 0$, the observation point lies in the shadow region. □

Example 20.1.2: *Diffraction by an infinite slit.* Consider an infinite slit on an opaque screen. The y-dimension of the slit is infinite and its x-size is $|x| \le a$, as shown on the left in the figure below. The same figure also shows an opaque strip of the same size.



The incident field is a uniform plane wave, $E_{\text{inc}}(x,z) = E_0 e^{-jkz}$, whose values on the slit are $E(x',0) = E_0$. The diffracted field at distance z is given by Eq. (20.1.16):

$$E(x,z) = \sqrt{\frac{jk}{2\pi z}}\,e^{-jkz}\int_{-\infty}^{\infty}E(x',0)\,e^{-jk(x-x')^2/2z}\,dx' = E_0\sqrt{\frac{jk}{2\pi z}}\,e^{-jkz}\int_{-a}^{a}e^{-jk(x-x')^2/2z}\,dx'$$

The integral can be reduced to the Fresnel integral $\mathcal{F}(x)$ of Appendix F by making the change of variables:

$$\sqrt{\frac{k}{2z}}\,(x'-x) = \sqrt{\frac{\pi}{2}}\,u\,, \quad \nu_{\pm} = \sqrt{\frac{k}{\pi z}}\,(\pm a - x)$$

so that

$$\sqrt{\frac{jk}{2\pi z}}\int_{-a}^{a}e^{-jk(x-x')^2/2z}\,dx' = \sqrt{\frac{j}{2}}\int_{\nu_-}^{\nu_+}e^{-j\pi u^2/2}\,du = \frac{\mathcal{F}(\nu_+)-\mathcal{F}(\nu_-)}{1-j} \equiv D(x,z)$$

where we used $\sqrt{j/2} = 1/(1-j)$. Thus, $E(x,z)$ becomes:

$$E(x,z) = e^{-jkz}\,D(x,z) \tag{20.1.17}$$

For the case of the strip, the limits of integration are changed to:

$$\sqrt{\frac{jk}{2\pi z}}\left(\int_a^{\infty}+\int_{-\infty}^{-a}\right)e^{-jk(x-x')^2/2z}\,dx' = \frac{\mathcal{F}(\infty)-\mathcal{F}(\nu_+)+\mathcal{F}(\nu_-)-\mathcal{F}(-\infty)}{1-j} = 1 - D(x,z)$$

where we used $\mathcal{F}(\infty) = -\mathcal{F}(-\infty) = (1-j)/2$. Thus, the diffracted field in the strip case will be given by the complementary expression

$$E(x,z) = e^{-jkz}\left[1 - D(x,z)\right] \tag{20.1.18}$$

This result is an example of the scalar Babinet principle discussed in Sec. 19.13,

$$E_{\text{slit}}(x,z) + E_{\text{strip}}(x,z) = e^{-jkz}$$

Fig. 20.1.1 shows the diffracted patterns in the two cases. The graphs plot the quantities $|D(x,z)|$ and $|1 - D(x,z)|$ versus x in the two cases.

The slit was chosen to be four wavelengths wide, $a = 4\lambda$, and the diffracted patterns correspond to the near, medium, and far distances $z = a$, $z = 20a$, and $z = 100a$. The latter case corresponds to the Fraunhofer pattern having a small ratio $a^2/\lambda z = 1/25$. For example, for the slit case, the corresponding pattern approximates (but it is not quite there yet) the typical sinc-function Fourier transform of the rectangular slit distribution $E(x',0) = E_0$, for $-a \le x' \le a$:

$$\hat{E}(k_x,0) = \int_{-a}^{a}E_0\,e^{jk_x x'}\,dx' = 2aE_0\,\frac{\sin(k_x a)}{k_x a}$$

where this is to be evaluated at $k_x = kx/z$ for the diffraction pattern $E(x,z)$. The property that at the center of the strip, $x = 0$, the diffracted pattern is not zero is an example of the so-called *Poisson's spot*, or, *Arago's spot* [638]. □

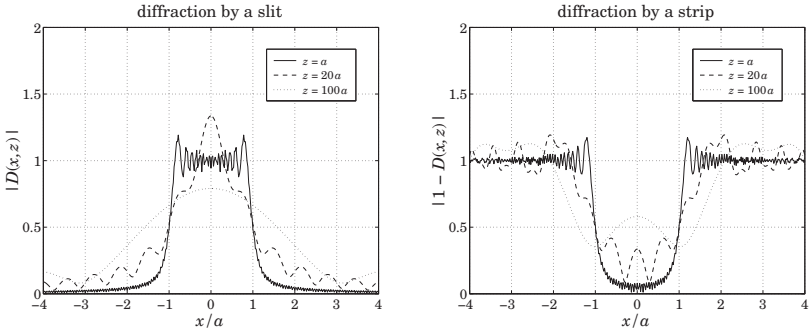


Fig. 20.1.1 Fresnel diffraction by a slit and a strip

Example 20.1.3: *Gaussian beam propagation.* In laser applications one often assumes a beam with a gaussian profile at the launch position ($z = 0$), which broadens as it propagates to $z > 0$. In the Fresnel approximation, the propagated field can be worked out in closed form. For simplicity, we consider a 1-D version with an initial beam profile of width w,

$$E(x,0) = E_0 \exp\left[-\frac{x^2}{2w^2}\right] \tag{20.1.19}$$

Using (20.1.16) the propagated field to distance z can be obtained by the following steps,

$$\hat{E}(k_x,0) = \int_{-\infty}^{\infty} E(x,0)\, e^{jk_x x}\, dx = E_0\sqrt{2\pi w^2}\, \exp\left[-\frac{1}{2}k_x^2 w^2\right]$$

$$\hat{E}(k_x,z) = e^{-jkz}\, e^{jk_x^2 z/2k}\, \hat{E}(k_x,0) = E_0 e^{-jkz}\sqrt{2\pi w^2}\, \exp\left[-\frac{1}{2}k_x^2\left(w^2 - j\frac{z}{k}\right)\right]$$

$$E(x,z) = \int_{-\infty}^{\infty} \hat{E}(k_x,z)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = E_0\, e^{-jkz}\sqrt{\frac{w^2}{w^2 - jz/k}}\, \exp\left[-\frac{x^2}{2(w^2 - jz/k)}\right]$$

where the integrals were done with the help of the integral of Eq. (3.5.18). Defining the distance, $z_0 = kw^2$, and its inverse, $\alpha_0 = 1/z_0 = 1/(kw^2)$, we may rewrite $E(x,z)$ as,

$$\begin{aligned} E(x,z) &= E_0\, e^{-jkz}\, (1 - j\alpha_0 z)^{-1/2} \exp\left[-\frac{x^2}{2w^2(1 - j\alpha_0 z)}\right] \\ &= E_0\, e^{-jkz}\, (1 - j\alpha_0 z)^{-1/2} \exp\left[-\frac{x^2(1 + j\alpha_0 z)}{2w^2(1 + \alpha_0^2 z^2)}\right] \end{aligned} \tag{20.1.20}$$

with magnitude,

$$|E(x,z)| = E_0 (1 + \alpha_0^2 z^2)^{-1/4} \exp\left[-\frac{x^2}{2w^2(1 + \alpha_0^2 z^2)}\right] \tag{20.1.21}$$

It is evident that the gaussian beam is getting wider and shorter as it propagates, with an effective width and effective height given as functions of z by,

$$w(z) = w(1 + \alpha_0^2 z^2)^{1/2}, \quad E_0(z) = E_0(1 + \alpha_0^2 z^2)^{-1/4} \tag{20.1.22}$$

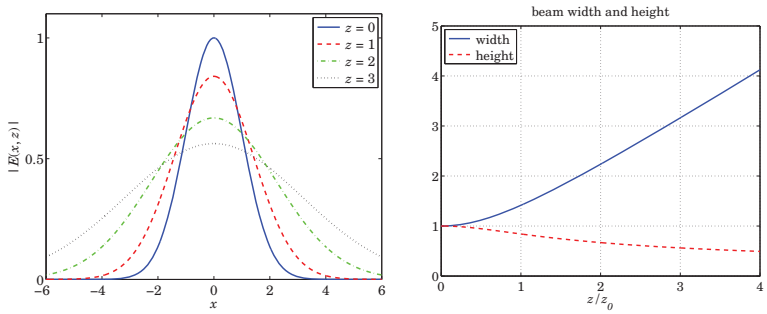


Fig. 20.1.2 Gaussian beam propagation.

Fig. 20.1.2 plots Eq. (20.1.21) as a function of x at four distances $z = 0, z_0, 2z_0, 3z_0$; it also plots Eqs. (20.1.22) over the range $0 \le z \le 4z_0$. The parameters were, $E_0 = w = z_0 = 1$.

Mathematically, this example is the same as the gaussian pulse propagation example discussed in Eqs. (3.6.1)–(3.6.6) of Sec. 3.6. The analogies between pulse compression and dispersion compensation on the one hand, and Fresnel approximation and lenses on the other, have been discussed by Papoulis [1416,1431]. □

20.2 Self-Imaging of Periodic Structures – Talbot Effect

The *Talbot or self-imaging* effect, first observed by Talbot in 1836, arises when a periodic grating is illuminated by a normally-incident plane wave. In the Fresnel approximation regime, the original grating image appears to be reproduced exactly at multiples of a certain distance called the Talbot length, and furthermore, there is a complicated structure of subimages at fractions of the Talbot length. Fig. 20.2.1 shows an example of the resulting fractal-type structure, termed a *Talbot carpet*. References [1459–1478] include some of the original papers, some recent reviews on applications, and the related Lau effect which involves a double grating. Our discussion is based on [1474–1477].

There are certain classes of fields, such as plane-waves, for which the plane-wave spectrum representation can be evaluated exactly. For example, consider a linear combination of TE plane waves (with TE defined relative to the xz plane) incident at different angles on the $z = 0$ plane. The y-component of the total E-field at $z = 0$ is given by a sum of, say, N terms,

$$E(x,0) = \sum_{n=1}^{N} E_n e^{-jk_{nx}x}, \quad k_{nx} = k\sin\theta_n$$

where θ_n are the angles of incidence measured with respect to the z-axis. The corresponding 1-D Fourier transform is,

$$\hat{E}(k_x,0) = \int_{-\infty}^{\infty} E(x,0)\, e^{jk_x x}\, dx = \sum_{n=1}^{N} E_n\, 2\pi\delta(k_x - k_{nx})$$

Then, the exact and the Fresnel plane-wave spectrum representations at $z \geq 0$ are,

$$E(x,z) = \int_{-\infty}^{\infty} \hat{E}(k_x,0)\, e^{-jk_z z}\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = \sum_{n=1}^{N} E_n\, e^{-jk_{nz} z}\, e^{-jk_{nx} x}\,, \quad k_{nz} = k\cos\theta_n$$

$$E(x,z) = \int_{-\infty}^{\infty} \hat{E}(k_x,0)\, e^{-jkz}\, e^{jk_x^2 z/2k}\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = e^{-jkz} \sum_{n=1}^{N} E_n\, e^{jk_{nx}^2 z/2k}\, e^{-jk_{nx} x} \tag{20.2.1}$$

where the Fresnel case can also be obtained from the exact one by the following approximation that effectively assumes that all the angles θ_n are small,

$$k_{nz} = \sqrt{k^2 - k_{nx}^2} \approx k - \frac{k_{nx}^2}{2k}$$

Another class of fields that can be computed exactly are *periodic gratings* that are periodic in x with some period, say, d, so that they can be expanded in a finite or infinite Fourier series of harmonics in the form,

$$E(x,0) = \sum_{n=-\infty}^{\infty} E_n\, e^{-nk_d x}\,, \quad k_d = \frac{2\pi}{d} = \text{fundamental harmonic} \tag{20.2.2}$$

with Fourier coefficients, for $-\infty < n < \infty$,

$$E_n = \frac{1}{d}\int_{-d/2}^{d/2} E(x,0)\, e^{jnk_d x}\, dx = \frac{1}{d}\int_{-d/2}^{d/2} E(x,0)\, e^{2\pi jnx/d}\, dx \tag{20.2.3}$$

This case is similar to (20.2.1) if we identify, $k_{nx} = nk_d$. Thus, we have,

$$\begin{aligned} E(x,z) &= \sum_{n=-\infty}^{\infty} E_n\, e^{-jz\sqrt{k^2 - n^2 k_d^2}}\, e^{-jnk_d x} \qquad \text{(exact)} \\ E(x,z) &= e^{-jkz} \sum_{n=-\infty}^{\infty} E_n\, e^{jn^2 k_d^2 z/2k}\, e^{-jnk_d x} \qquad \text{(Fresnel)} \end{aligned} \tag{20.2.4}$$

We will concentrate on the Fresnel case. Let us define the Talbot length by,

$$\boxed{z_T = \frac{2\pi k}{k_d^2} = \frac{d^2}{\lambda}} \quad \text{(Talbot length)} \tag{20.2.5}$$

so that the Fresnel expression in (20.2.4) can be written as,

$$\boxed{E(x,z) = e^{-jkz} \sum_{n=-\infty}^{\infty} E_n\, e^{j\pi n^2 z/z_T}\, e^{-2\pi jnx/d}} \tag{20.2.6}$$

At $z = 2z_T$, the quadratic phase factor becomes unity, indeed,

$$e^{j\pi n^2 z/z_T}\Big|_{z=z_T} = e^{j\pi n^2 (2z_T)/z_T} = e^{2\pi jn^2} = 1\,, \quad \text{for all } n$$

It follows from (20.2.6) that the initial field reproduces itself at this distance,

$$E(x,2z_T) = e^{-2jkz_T}\, E(x,0) \quad \Rightarrow \quad \big|E(x,2z_T)\big| = \big|E(x,0)\big|$$

which implies that $\big|E(x,z)\big|$ is *periodic* in z with period *twice* the Talbot length, $2z_T$.

Similarly, at $z = z_T$, we have, $e^{j\pi n^2 z/z_T} = e^{j\pi n^2} = e^{j\pi n}$, for all n, with the latter identity following from the observation that $e^{j\pi n^2} = e^{j\pi n} = 1$ if n is even, and when n is odd, i.e., $n = 2m+1$, we have, $e^{j\pi n^2} = e^{j\pi(2m+1)n} = e^{2\pi jnm} e^{j\pi n} = e^{j\pi n}$. In this case, the two exponentials in (20.2.6) can be combined as follows,

$$E(x,z_T) = e^{-jkz_T} \sum_{n=-\infty}^{\infty} E_n\, e^{j\pi n}\, e^{-2\pi jnx/d} = e^{-jkz_T} \sum_{n=-\infty}^{\infty} E_n\, e^{-2\pi jn(x-d/2)/d}\,, \quad \text{or,}$$

$$E(x,z_T) = e^{-jkz_T}\, E(x - \tfrac{1}{2}d,\, 0) \tag{20.2.7}$$

Thus, the initial field is reproduced exactly but is laterally shifted by half-period $\frac{1}{2}d$. To summarize, we have relative to an arbitrary distance z_0,

$$\begin{aligned} E(x,z_0+z) &= e^{-jkz} E(x - \tfrac{1}{2}d, z_0)\,, \quad \text{when } z \text{ an odd multiple of } z_T \\ E(x,z_0+z) &= e^{-jkz} E(x,z_0)\,, \quad \text{when } z \text{ is an even multiple of } z_T \end{aligned} \tag{20.2.8}$$

Of course, the former implies the latter by applying it twice, and invoking the periodicity in x, e.g., $\big|E(x,2z_T)\big| = \big|E(x-\frac{1}{2}d, z_T)\big| = \big|E(x-d,0)\big| = \big|E(x,0)\big|$.

The *fractional* Talbot effect corresponds to subimages that are generated at the fractional distances $z_{pq} = (p/q)z_T$, where p, q are coprime integers.

To z_{pq} we can add any multiple of z_T. Indeed, if we set, $z = z_0 + z_{pq}$, with $z_0 = (2s+r)z_T$, where $r = 0, 1$, we will have, $E(x, z_0 + z_{pq}) = e^{-jkz_0} E\big(x - \frac{1}{2}rd,\, z_{pq}\big)$. Setting now $z = z_{pq} = (p/q)z_T$ in (20.2.6), we obtain,

$$E(x,z_{pq}) = e^{-jkz_{pq}} \sum_{n=-\infty}^{\infty} E_n\, e^{j\pi n^2 p/q}\, e^{-2\pi jnx/d} \tag{20.2.9}$$

The fractional Talbot subimages are a consequence of this expression. We note that if p is even (so q must be odd), then, the factor, $a_n = e^{j\pi n^2 p/q}$, is periodic in n with period q, as can be verified explicitly,

$$a_{n+q} = e^{j\pi(n+q)^2 p/q} = e^{j\pi n^2 p/q} \cdot e^{2\pi jnp} \cdot e^{j\pi pq} = e^{j\pi n^2 p/q} \cdot 1 \cdot 1 = a_n$$

On the other hand, if p is odd (now q can be even or odd), then the sequence, $e^{j\pi n^2 p/q}$ is no longer periodic with period q. But the following sequence is, $a_n = e^{j\pi n^2 p/q} e^{-j\pi n}$, indeed since $p-1$ is even, we have,

$$a_{n+q} = e^{j\pi(n+q)^2 p/q} e^{-j\pi(n+q)} = e^{j\pi n^2 p/q} e^{-j\pi n} \cdot e^{j\pi(p-1)q} = a_n$$

Combining the two cases, we denote by r the remainder of the division of p by 2, that is, $r = \text{rem}(p,2)$, and define the sequence, $a_n = e^{j\pi n^2 p/q} e^{-j\pi nr}$, which will now be periodic with period q, for all p.

This periodicity implies that a_n can be expanded in a *discrete Fourier series* [48], i.e., as an inverse DFT consisting of a linear combination of q discrete sinusoids at the q DFT frequencies, $\omega_m = 2\pi m/q$, $m = 0, 1, \ldots, q-1$, that is,

$$\boxed{a_n = e^{j\pi n^2 p/q}\, e^{-j\pi n r} = \frac{1}{q}\sum_{m=0}^{q-1} A_m\, e^{2\pi j m n/q}} \quad -\infty < n < \infty \tag{20.2.10}$$

where the q coefficients A_m are recognized as the q-point DFT of the discrete sequence a_n, defined as follows for $m = 0, 1, \ldots, q-1$,

$$\boxed{A_m = \sum_{n=0}^{q-1} a_n\, e^{-2\pi j n m/q} = \sum_{n=0}^{q-1} e^{j\pi n^2 p/q}\, e^{-j\pi n r}\, e^{-2\pi j n m/q}} \tag{20.2.11}$$

The DFT coefficients A_m are known as *quadratic Gauss sums* and play a prominent role in Number Theory. Closed-form expressions for A_m in terms of p, q have been given in [1474-1476]. Adapted to our case, they are for $m = 0, 1, \ldots, q-1$,

$$A_m = \begin{cases} \sqrt{q}\left(\dfrac{p}{q}\right) \exp\left[-j\pi\left(\dfrac{q-1}{4} + \dfrac{p}{q}\,(p\backslash q)^2\, m^2\right)\right], & p, q = \text{even, odd} \\ \sqrt{q}\left(\dfrac{q}{p}\right) \exp\left[j\pi\left(\dfrac{p}{4} - \dfrac{p}{4q}\,(p\backslash q)^2\,(2m+q)^2\right)\right], & p, q = \text{odd, even} \\ \sqrt{q}\left(\dfrac{p}{q}\right) \exp\left[-j\pi\left(\dfrac{q-1}{4} + \dfrac{2p}{q}\,(2\backslash q)\,(2p\backslash q)^2\,(2m+q)^2\right)\right], & p, q = \text{odd, odd} \end{cases} \tag{20.2.12}$$

where $\left(\frac{p}{q}\right)$ is the Jacobi symbol taking on the values ± 1, and $(p\backslash q)$ denotes the integer or modular inverse of p relative to q.[†]

It is evident that the coefficients A_m have constant magnitude, $|A_m| = \sqrt{q}$, for $m = 0, 1, \ldots, q-1$. The MATLAB function **talbot**, computes A_m from Eq. (20.2.12). It requires that p, q be coprime and has usage,

```
A = talbot(p,q)
```
% Gauss sums for the fractional Talbot effect

In our simulation examples below, we find it simpler to compute them numerically by a DFT matrix transformation. Example 20.2.2 compares the two methods. Replacing

[†]The definitions of these concepts may be found in Wikipedia, `https://www.wikipedia.org`

now, $e^{j\pi n^2 p/q} = a_n\, e^{j\pi n r}$, into Eq. (20.2.9) and using (20.2.10), we obtain,

$$\begin{aligned} E(x, z_{pq}) &= e^{-jkz_{pq}} \sum_{n=-\infty}^{\infty} E_n\, e^{j\pi n^2 p/q}\, e^{-2\pi j n x/d} \\ &= e^{-jkz_{pq}} \sum_{n=-\infty}^{\infty} E_n\, a_n\, e^{j\pi n r}\, e^{-2\pi j n x/d} \\ &= e^{-jkz_{pq}} \sum_{n=-\infty}^{\infty} E_n \frac{1}{q}\sum_{m=0}^{q-1} A_m\, e^{2\pi j m n/q}\, e^{j\pi n r}\, e^{-2\pi j n x/d} \\ &= e^{-jkz_{pq}} \frac{1}{q}\sum_{m=0}^{q-1} A_m \sum_{n=-\infty}^{\infty} E_n\, e^{-2\pi j n (x - md/q - rd/2)/d}, \quad \text{or,} \end{aligned}$$

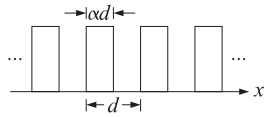
$$\boxed{E(x, z_{pq}) = e^{-jkz_{pq}} \frac{1}{q}\sum_{m=0}^{q-1} A_m\, E\left(x - m\frac{d}{q} - r\frac{d}{2},\, 0\right)} \tag{20.2.13}$$

which shows that $E(x, z_{pq})$ is a linear combination of q overlapping copies of the original grating $E(x, 0)$, displaced laterally at multiples of d/q, with an overall displacement by $d/2$ when p is odd.

Example 20.2.1: This example from [1477] illustrates the generation of a Talbot carpet. Figure 20.2.1 plots the magnitude, $|E(x, z)|$, as a function of x, z, over the ranges, $-2d \le x \le 2d$ and $-0.5z_T \le z \le 3.5z_T$. One can observe the periodicity at $z = 2z_T$, relative to $z = 0$, and the lateral shift at $z = z_T$ and $z = 3z_T$, as well as some of the fractional Talbot subimages at distances $(p/q)z_T$.

The periodic grating was taken to be a square wave with a duty cycle, $\alpha = 0.1$, defined as follows over one period, $-d/2 \le x \le d/2$,

$$E(x, 0) = \begin{cases} 1, & |x| \le \frac{1}{2}\alpha d \\ 0, & \frac{1}{2}\alpha d < |x| \le \frac{1}{2} d \end{cases}$$



(20.2.14)

The bright spots correspond to the narrow square pulses of the grating. The duty cycle was chosen to be small enough so that the laterally repeated subimages at spacings d/q are clearly visible. The Fourier coefficients were calculated from Eq. (20.2.3),

$$E_n = \frac{\sin(\alpha\pi n)}{\pi n}, \quad -\infty < n < \infty \qquad (E_0 = \alpha) \tag{20.2.15}$$

The field $E(x, z)$ was computed using the following truncated version of the Fourier series (20.2.6), where the coefficients E_n were multiplied by a Hamming window w_n in order to reduce the Gibbs ripples arising from the truncation,

$$E(x, z) = e^{-jkz} \sum_{n=-M}^{M} w_n E_n\, e^{j\pi n^2 z/z_T}\, e^{-2\pi j n x/d} \tag{20.2.16}$$

with, $w_n = 0.54 + 0.46\cos(\pi n/M)$. The following MATLAB code segment illustrates the generation of this graph,

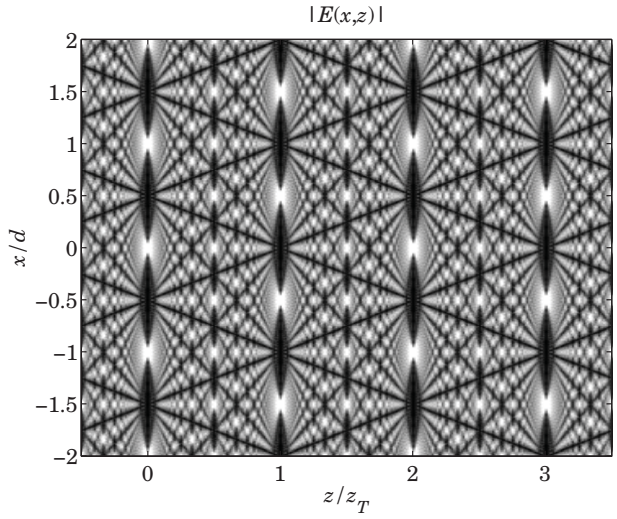


Fig. 20.2.1 Talbot carpet.

```
d = 1; a = 0.1;                         % d = period, a = duty cycle
M = 100; n = -M:M;                      % no. Fourier series coefficients

w  = @(n) 0.54 + 0.46*cos(pi*n/M);      % Hamming window
En = @(n) a*sinc(a*n).*w(n);            % windowed coefficients

E = @(x,z) sum(En(n).*exp(j*pi*n.^2*z).*exp(-2*pi*j*n*x));    % E(x,z)

x = linspace(-2, 2, 801);               % x in units of d
z = linspace(-0.5, 3.5, 801);           % z in units of z_T

for i=1:length(x)                       % calculate |E(x,z)|
   for j=1:length(z)
      F(i,j) = abs(E(x(i),z(j)));
   end
end

Emag = 300*F/max(max(F));               % normalized magnitude

figure; image(z,x,Emag);                % plot image
colormap(gray(128));
```

Example 20.2.2: This example from [1474] illustrates the computation of the fractional subimages. The grating was again a square wave, but with a duty cycle of $\alpha = 0.5$, known as a Ronchi grating. Fig. 20.2.2 plots the magnitudes $|E(x, z_{pq})|$ at $z_{pq} = (p/q)z_T$, for the choices, $q = 20$, and, $p = 0, 1, 2, 3, 4, 5, 6, 7, 8$.

The typical MATLAB code for generating these graphs is as follows, shown for the specific case $p = 3$. The code also verifies that the two methods of calculating $E(x, z_{pq})$ from Eqs. (20.2.9) and (20.2.13) produce the same result to within MATLAB's numerical accuracy,

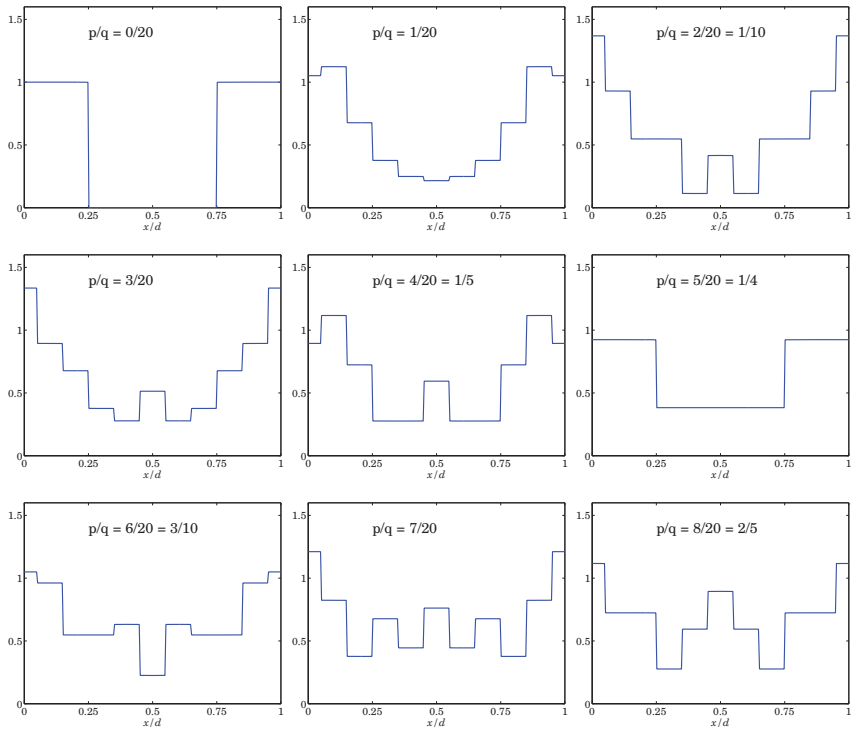


Fig. 20.2.2 Magnitude $|E(x, z_{pq})|$ versus x, over one period, $0 \le x \le d$.

```
d = 1; a = 0.5;                         % d = period, a = duty cycle
M = 1000; n = -M:M;                     % no. Fourier series coefficients

w  = @(n) 0.54 + 0.46*cos(pi*n/M);      % Hamming window
En = @(n) a*sinc(a*n).*w(n);            % Fourier series coefficients

E = @(x,z) sum(En(n).*exp(j*pi*n.^2*z).*exp(-2*pi*j*n*x));

x = linspace(0,1,251);                  % x in units of d

p=3; q=20; r = rem(p,2);
zpq = p/q;

for i=1:length(x),                % calculate E(x,zpq) using Eq.(20.2.9)
   F(i) = E(x(i),zpq);
end

figure; plot(x,abs(F),'b-');    % plot magnitude |E(x,z_pq)|

n = (0:q-1)'; D = exp(-2*pi*j*n*n'/q);      % qxq DFT matrix
a = exp(j*pi*n.^2*p/q).*exp(-j*pi*n*r);     % qx1 sequence a(n)
A = D*a;                                    % qx1 DFT coefficients A(m)
```

```
for i=1:length(x)              % calculate E(x,zpq) using Eq.(20.2.12)
   S = 0;
   for m=0:q-1
      S = S + A(m+1)*E(x(i)-m*d/q-r*d/2,0);
   end
   Fpq(i) = S/q;
end

Ediff = norm(F-Fpq)            % compare methods, Ediff = 8.9675e-13

B = talbot(p,q);               % here, p=3,q=20 are coprime

Err = norm(A-B);               % compare DFT with TALBOT, Err = 2.7907e-11
```

For $q = 20$ and the values $p = 0, 2, 4, 5, 6, 8$, the pairs p, q are not coprime. In these cases the ratios p/q are reduced to their coprime versions, i.e., $p/q = [0, 2, 4, 5, 6, 8]/20 = [0/1,\ 1/10,\ 1/5,\ 1/4,\ 3/10,\ 2/5]$. The q-term expansion (20.2.13) is still valid, but the q-point DFT A_m has zero entries that effectively reduce it to the equivalent coprime value of q. The table below shows the ($q = 20$) values of A_m, $m = 0, 1, \ldots, 19$, rounded to two digits for display purposes, for the cases of $p = 0, 1, 2, 3, 4, 5$.

In the coprime cases, $p = 1$ and $p = 3$, the array A_m is full. In the case $p = 2$, which reduces to $p/q = 2/20 = 1/10$ with a reduced q of 10, the array A_m effectively reduces to a 10-dimensional one. In the $p = 4$ case with $p/q = 4/20 = 1/5$, A_m effectively reduces to a 5-dimensional one. For $p = 5$ with $p/q = 5/20 = 1/4$, it reduces to a 4-dimensional one, and in the trivial $p = 0$ case, to 1-dimensional. We note also that in these non-coprime cases, the magnitude of the non-zero A_m's is $|A_m| = \sqrt{gq}$, where g is the greatest common divisor of p, q, where g can be calculated in MATLAB with the command, `g = gcd(p,q)`.

m	p = 0	p = 1	p = 2	p = 3	p = 4	p = 5
0	20.00	-3.16-3.16i	4.47+4.47i	-3.16+3.16i	0	-7.07-7.07i
1	0	3.62+2.63i	0	1.38+4.25i	0	0
2	0	-4.42-0.70i	5.64+2.87i	-2.03-3.98i	7.24+5.26i	0
3	0	3.62-2.63i	0	1.38-4.25i	0	0
4	0	0.70+4.42i	5.64-2.87i	-3.98-2.03i	0	0
5	0	-4.47+0.00i	0	4.47+0.00i	0	10.00
6	0	-0.70-4.42i	-2.87-5.64i	3.98+2.03i	7.24-5.26i	0
7	0	3.62-2.63i	0	1.38-4.25i	0	0
8	0	4.42+0.70i	-2.87+5.64i	2.03+3.98i	0	0
9	0	3.62+2.63i	0	1.38+4.25i	0	0
10	0	3.16+3.16i	4.47-4.47i	3.16-3.16i	-8.94+0.00i	7.07+7.07i
11	0	3.62+2.63i	0	1.38+4.25i	0	0
12	0	4.42+0.70i	-2.87+5.64i	2.03+3.98i	0	0
13	0	3.62-2.63i	0	1.38-4.25i	0	0
14	0	-0.70-4.42i	-2.87-5.64i	3.98+2.03i	7.24-5.26i	0
15	0	-4.47+0.00i	0	4.47+0.00i	0	10.00
16	0	0.70+4.42i	5.64-2.87i	-3.98-2.03i	0	0
17	0	3.62-2.63i	0	1.38-4.25i	0	0
18	0	-4.42-0.70i	5.64+2.87i	-2.03-3.98i	7.24+5.26i	0
19	0	3.62+2.63i	0	1.38+4.25i	0	0

For the particular coprime case, $p = 3$, $q = 20$, the above code segment compares also the calculation of the A_m coefficients using the DFT matrix approach and the closed-form expressions (20.2.12) as implemented by the function TALBOT.

In the non-coprime cases, after p, q are reduced by their greatest common divisor, $g = \gcd(p, q)$, so that, $p = gp_0$, $q = gq_0$, and $p/q = p_0/q_0$, then one can apply TALBOT to the reduced pair p_0, q_0 and obtain the q_0-point DFT coefficients, which correspond to the non-zero entries in the above columns, up to a circular shift by $q/2$ and an overall factor g, noting that these columns have magnitudes, $\sqrt{gq} = \sqrt{g^2 q_0} = g\sqrt{q_0}$.

For a more thorough test, we have computed A_m for all possible coprime pairs in the ranges, $1 \le p \le 30$ and $1 \le q \le 30$, using the DFT method and the closed-form expressions (20.2.12), implemented by the following code segment,

```
P=30; Q=30;
Err=[];

for p=1:P
   for q=1:Q
      if gcd(p,q)~=1, continue; end            % skip non-coprime pairs

      r = rem(p,2);
      n = (0:q-1)'; D = exp(-2*pi*j*n*n'/q);   % qxq DFT matrix
      a = exp(j*pi*n.^2*p/q).*exp(-j*pi*n*r);  % q-point discrete signal a(n)
      A = D*a;                                 % q-point DFT of a(n)

      B = talbot(p,q);                         % q-point DFT from TALBOT

      Err = [Err,norm(A-B)];                   % Euclidean norm error
   end
end

Emax = max(Err)                                % overall maximum error

figure; semilogy(Err, 'b.-')                   % plot error
```

The error, plotted in Fig. 20.2.3, is measured by the Euclidean norm between the two DFT vectors, $\|A - B\|$. Its maximal value was $E_{\max} = 5.7348\times10^{-7}$. For certain p, q pairs, the error is zero and such points appear as gaps in the semilogy plot. □

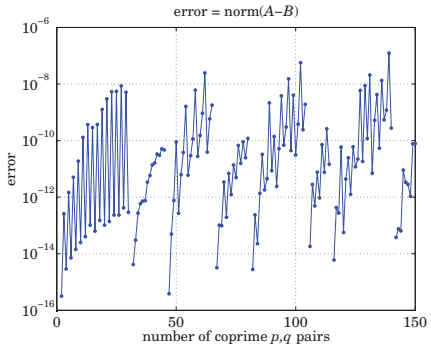


Fig. 20.2.3 Computational error between DFT method and closed-form expressions (20.2.12).

20.3 Fraunhofer Approximation

The *Fraunhofer approximation* is a limiting case of the Fresnel approximation when the distance z is even larger than that in the Fresnel case. More precisely, it is obtained in the far-field limit when $k|\boldsymbol{r}_\perp'|^2 \ll z$, or, $kz \gg (kD)^2$, where D is the size of the aperture.

In this approximation, the field $E(\boldsymbol{r}_\perp, z)$ becomes proportional to the *Fourier transform* $\hat{E}(\boldsymbol{k}_\perp, 0)$ of the field at the input plane. It is similar to the far-field approximation of Eq. (19.3.3) discussed in Sec. 19.3, except now we are further assuming the paraxial approximation in which z is much larger than the transverse directions $|\boldsymbol{r}_\perp|$.

A direct way of deriving the Fraunhofer approximation is by applying the stationary-phase approximation—Eq. (H.6) of Appendix H—to the evaluation of the plane-wave spectrum integral (20.1.9). Define the phase function

$$\phi(\boldsymbol{k}_\perp) = \frac{|\boldsymbol{k}_\perp|^2 z}{2k} - \boldsymbol{k}_\perp \cdot \boldsymbol{r}_\perp = \left[\frac{k_x^2 z}{2k} - k_x x\right] + \left[\frac{k_y^2 z}{2k} - k_y y\right] \equiv \phi_x(k_x) + \phi_y(k_y)$$

Then, the stationary-point with respect to the k_x variable is,

$$\phi_x'(k_x) = \frac{k_x z}{k} - x = 0 \quad \Rightarrow \quad k_x = \frac{xk}{z}, \quad \phi_x''(k_x) = \frac{z}{k}$$

and similar expressions for $\phi_y(k_y)$. Thus, vectorially, the stationary point is at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/z$. Using Eq. (H.6) (with diagonal matrix Φ), we obtain:

$$\int_{-\infty}^{\infty} \hat{E}(\boldsymbol{k}_\perp, 0)\, e^{j\phi(\boldsymbol{k}_\perp)}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \simeq \sqrt{\frac{2\pi j}{\phi_x''(k_x)} \cdot \frac{2\pi j}{\phi_y''(k_y)}} \left[\hat{E}(\boldsymbol{k}_\perp, 0)\, e^{j\phi(\boldsymbol{k}_\perp)}\, \frac{1}{(2\pi)^2}\right]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{z}}$$

Noting that $\phi(\boldsymbol{k}_\perp) = -k|\boldsymbol{r}_\perp|^2/2z$ at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/z$, we finally find:

$$\boxed{E(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z} \left[\hat{E}(\boldsymbol{k}_\perp, 0)\right]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{z}}} \quad \text{(Fraunhofer)} \qquad (20.3.1)$$

A simpler way of deriving (20.3.1) is by using (20.1.9) and noting that the factor $e^{-jk|\boldsymbol{r}_\perp'|^2/2z}$ can be ignored if we assume that $k|\boldsymbol{r}_\perp'|^2 \ll z$, which leads to:

$$E(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z} \int_{-\infty}^{\infty} E(\boldsymbol{r}_\perp', 0)\, e^{jk\boldsymbol{r}_\perp \cdot \boldsymbol{r}_\perp'/z}\, d^2\boldsymbol{r}_\perp' \qquad (20.3.2)$$

and the last integral factor is recognized as $\hat{E}(\boldsymbol{k}_\perp, 0)$ evaluated at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/z$. The Fraunhofer condition, $k|\boldsymbol{r}_\perp'|^2 \ll z$, implicitly assumes that the aperture is finite so that the transverse integration over $\boldsymbol{r}_\perp'$ is limited to a finite range. If we denote by D the typical extent of the aperture, then, the condition reads, $kD^2 \ll z$, or, $kz \gg (kD)^2$.

In the *circularly symmetric* case, the Fourier integral in (20.3.2) reduces to a Hankel transform,

$$\boxed{\begin{aligned} E(r_\perp, z) &= \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jkr_\perp^2/2z} \int_{-\infty}^{\infty} E(r_\perp', 0)\, J_0\left(\frac{kr_\perp r_\perp'}{z}\right) 2\pi\, r_\perp' dr_\perp' \\ &= \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jkr_\perp^2/2z} \left[\hat{E}(k_\perp, 0)\right]_{k_\perp = \frac{kr_\perp}{z}} \end{aligned}} \qquad (20.3.3)$$

The *one-dimensional* version of (20.3.1) is obtained in a similar way from Eq. (20.1.16),

$$\boxed{E(x, z) = \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz}\, e^{-jkx^2/2z} \left[\hat{E}(k_x, 0)\right]_{k_x = \frac{kx}{z}}} \quad \text{(Fraunhofer)} \qquad (20.3.4)$$

Example 20.3.1: *Exact, Fresnel, and Fraunhofer diffraction by circular apertures* [1433–1445]. Consider a circular aperture of radius a, and assume a uniform field, $E(\boldsymbol{r}_\perp, 0) = u(a - |\boldsymbol{r}_\perp|)$, at $z = 0$, where $u(\cdot)$ is the unit-step function. The on-axis field at distance $z \geq 0$ can be calculated in closed-form and compared for the three cases of using: (a) the exact Rayleigh-Sommerfeld integral (20.1.1), (b) the circularly-symmetric Fresnel formula (20.1.13), and (c) the Fraunhofer formula (20.3.1).

Setting $E(\boldsymbol{r}_\perp', 0) = 1$ and $\boldsymbol{r}_\perp = 0$ in (20.1.1), we have $R = \sqrt{|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2 + z^2} = \sqrt{r_\perp'^2 + z^2}$. The angular integration in $d^2\boldsymbol{r}_\perp' = r_\perp' dr_\perp' d\phi'$ can be done immediately, replacing it by $2\pi r_\perp' dr_\perp'$, and the remaining integral over $r_\perp'$ can be done in closed-form,

$$\boxed{\begin{aligned} E(0, z) &= \int_0^a \frac{2z}{R}\left(jk + \frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R}\, 2\pi\, r_\perp' dr_\perp' \\ &= e^{-jkz} - \frac{z}{\sqrt{z^2 + a^2}}\, e^{-jk\sqrt{z^2+a^2}} \end{aligned}} \quad \text{(exact)} \qquad (20.3.5)$$

Setting $r_\perp = 0$ in the Fresnel formula (20.1.13) and noting that $J_0(0) = 1$, we obtain,

$$\boxed{\begin{aligned} E(0, z) &= \frac{jk}{2\pi z}\, e^{-jkz} \int_0^a e^{-jkr_\perp'^2/2z}\, 2\pi\, r_\perp' dr_\perp' \\ &= e^{-jkz} - e^{-jkz}\, e^{-jka^2/2z} \end{aligned}} \quad \text{(Fresnel)} \qquad (20.3.6)$$

This can also be obtained from Eq. (20.3.5) by making the approximation in the exponential,

$$\sqrt{z^2 + a^2} \approx z + \frac{a^2}{2z}$$

while setting, $\sqrt{z^2 + a^2} \approx z$, in the amplitude factors. For the Fraunhofer case, we must determine the Fourier transform of $E(\boldsymbol{r}_\perp', 0) = 1$ over the aperture, that is, from (20.1.12),

$$\hat{E}(k_\perp, 0) = \int_0^a J_0(k_\perp r_\perp)\, 2\pi\, r_\perp dr_\perp = \pi a^2\, \frac{2J_1(k_\perp a)}{k_\perp a} \qquad (20.3.7)$$

which evaluates to $\hat{E}(0, 0) = \pi a^2$, at $k_\perp = kr_\perp/z = 0$, so that (20.3.1) gives,

$$\boxed{E(0, z) = \frac{jk}{2\pi z}\, e^{-jkz} \left[\hat{E}(k_\perp, 0)\right]_{k_\perp = 0} = e^{-jkz}\, \frac{jka^2}{2z}} \quad \text{(Fraunhofer)} \qquad (20.3.8)$$

and this can also be obtained from (20.3.6) by assuming $z \gg ka^2$ and using the Taylor series approximation of the exponential, $e^{jka^2/2z} \approx 1 + jka^2/2z$.

Fig. 20.3.1 plots the exact, Fresnel, and Fraunhofer fields for the case $a = 10\lambda$. The Fresnel and Fraunhofer distances, shown as the dotted vertical lines, were $z_{\text{fresnel}} = 39.8\lambda = 3.98\,a$ and $z_{\text{fraunh}} = 628.3\lambda = 62.83\,a$, and were computed from the expressions,

$$kz_{\text{fresnel}} = (ka)^{4/3}, \quad kz_{\text{fraunh}} = (ka)^2$$

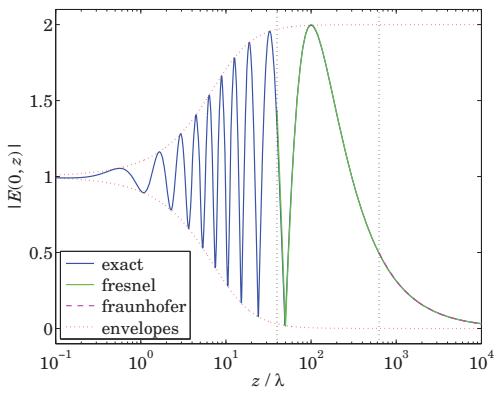


Fig. 20.3.1 Magnitude plots, $|E(0,z)|$ vs. z.

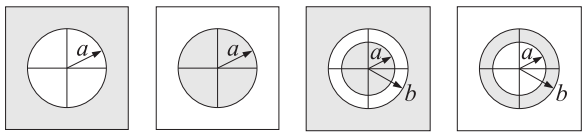
The exact expression (20.3.5) was evaluated and plotted over the range, $10^{-1} \le z/\lambda \le 10^4$. The Fresnel expression (20.3.6) was plotted only for $z \ge z_{\text{fresnel}}$, and is barely distinguishable on the graph from the exact expression (see color graph online). Similarly, the Fraunhofer expression was plotted for $z \ge z_{\text{fraunh}}$, and is also barely distinguishable from the exact case.

The envelopes of the oscillating exact field were defined by setting, $e^{-jk\sqrt{z^2+a^2}} = \pm e^{-jkz}$, in the magnitude of (20.3.5), that is, they are the curves,

$$E_{\text{env}}(z) = 1 \pm \frac{z}{\sqrt{z^2+a^2}}$$

This example demonstrates how the fields transition gradually from the exact expression to Fresnel, and then to Fraunhofer, as z is gradually increased.

The case of the complementary aperture consisting of an opaque disk of radius a, as well the cases of an annular aperture of inner and outer radii a, b, and its complementary ring aperture are depicted below.



The integral (20.3.5) can be done exactly in all cases, resulting in,

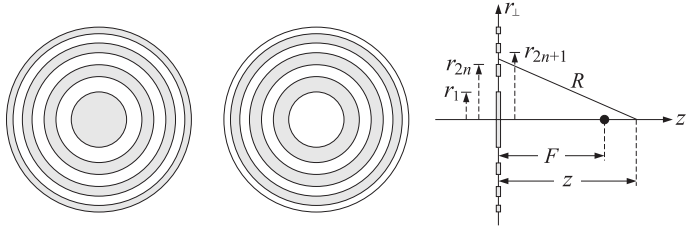
$$E(0,z) = \frac{z}{\sqrt{z^2+a^2}}\, e^{-jk\sqrt{z^2+a^2}} \qquad \text{(opaque disk)}$$

$$E(0,z) = \frac{z}{\sqrt{z^2+a^2}}\, e^{-jk\sqrt{z^2+a^2}} - \frac{z}{\sqrt{z^2+b^2}}\, e^{-jk\sqrt{z^2+b^2}} \qquad \text{(annular)} \qquad (20.3.9)$$

$$E(0,z) = e^{-jkz} - \frac{z}{\sqrt{z^2+a^2}}\, e^{-jk\sqrt{z^2+a^2}} + \frac{z}{\sqrt{z^2+b^2}}\, e^{-jk\sqrt{z^2+b^2}} \qquad \text{(annular ring)}$$

We note also that the scalar Babinet principle of Eq. (19.13.17) is satisfied for both the disk and the annular cases, that is, $E(0,z) + E_{\text{compl}}(0,z) = e^{-jkz}$ □

Example 20.3.2: *Fresnel zone-plate lens* [1435].† An aperture consisting of alternating transparent and opaque annular apertures, as shown below, can act as a lens with a prescribed focal length, say, F, provided the radii of the transparent zones are properly chosen so that the diffracted waves from these zones interfere constructively at distance F.



For N transparent zones, the n-th zone is defined by, $r_{2n} \le r_\perp \le r_{2n+1}$, $n = 0, 1, \ldots, N-1$, that is, they are the intervals, $[r_0, r_1], [r_2, r_3], [r_4, r_5], \ldots$, starting with $r_0 = 0$. Similarly, the opaque zones are the intervals, $[r_1, r_2], [r_3, r_4], [r_5, r_6]$, etc. For the complementary aperture shown above the groups of subintervals are interchanged.

Assuming a uniform incident field $E_0 = 1$ from the left side and using the results of the previous example for a single annular aperture, the net field at distance z along the axis arising from all N annuli will be given by the sum,

$$E(0,z) = \sum_{n=0}^{N-1} \left[\frac{z}{\sqrt{z^2+r_{2n}^2}}\, e^{-jk\sqrt{z^2+r_{2n}^2}} - \frac{z}{\sqrt{z^2+r_{2n+1}^2}}\, e^{-jk\sqrt{z^2+r_{2n+1}^2}} \right] \qquad (20.3.10)$$

In order for the sum (20.3.10) to exhibit constructive interference at $z = F$, the zone radii must be chosen such that the relative path lengths from the zone edges to the focal point F be multiples of half-wavelength, that is,

$$\sqrt{F^2+r_m^2} - F = m\frac{\lambda}{2} \quad \Rightarrow \quad r_m^2 = m\lambda\left(F + \frac{m\lambda}{4}\right), \quad m = 0, 1, 2, \ldots \qquad (20.3.11)$$

Thus, at $z = F$ we have, $\sqrt{F^2+r_m^2} = F + m\lambda/2$, and,

$$E(0,F) = \sum_{n=0}^{N-1} \left[\frac{F}{F+(2n)\lambda/2}\, e^{-jk(F+(2n)\lambda/2)} - \frac{F}{F+(2n+1)\lambda/2}\, e^{-jk(F+(2n+1)\lambda/2)} \right]$$

But, $e^{-jkn\lambda} = e^{-2\pi jn} = 1$ and $e^{-jk(n\lambda+\lambda/2)} = e^{-2\pi jn} e^{-j\pi} = -1$, so that,

$$E(0,F) = e^{-jkF} \sum_{n=0}^{N-1} \left[\frac{F}{F+n\lambda} + \frac{F}{F+n\lambda+\lambda/2} \right] \qquad (20.3.12)$$

Typically, we have $F \gg N\lambda$, and hence,

$$E(0,F) = e^{-jkF} \sum_{n=0}^{N-1} \left[\frac{F}{F+n\lambda} + \frac{F}{F+n\lambda+\lambda/2} \right] \approx e^{-jkF} \sum_{n=0}^{N-1} \left[\frac{F}{F} + \frac{F}{F} \right] = 2N\, e^{-jkF}$$

† see also, `http://zoneplate.lbl.gov/theory`

The half-wavelength choice in (20.3.11) is now evident: it causes the presence of the factor, $e^{-j\pi} = -1$, which changes the sign of the second term in (20.3.10) from negative to positive. Had the right-hand side in (20.3.11) been $m\lambda$, then, the two terms would have canceled causing destructive interference. The opaque zones block the contribution of such terms, thus enhancing the constructive interference. For any F, the summation (20.3.12) can be expressed in terms of the *digamma* function $\psi(x) = \Gamma'(x)/\Gamma(x)$,[†] which is a built-in function in MATLAB. Setting, $f = F/\lambda$, we obtain,

$$E(0,F) = e^{-jkF} f \left[\psi(f+N) - \psi(f) + \psi(f+1/2+N) - \psi(f+1/2) \right] \tag{20.3.13}$$

Fig. 20.3.2 plots the normalized magnitude square, $|E(0,z)/E(0,F)|^2$, over the range $0 \le z/F \le 2$. The focal length was chosen to be $F = 50\lambda$ and two cases, $N = 10$ and $N = 20$, were considered. The following MATLAB code illustrates the computation,

```
la = 1; k = 2*pi/la; F = 50*la;

z = linspace(0,2,901);      % z in units of F

for N = [10,20]
   [Z,n] = meshgrid(z*F, 0:N-1);

   Ra = sqrt(Z.^2 + 2*n*la.*(F+2*n*la/4));
   Rb = sqrt(Z.^2 + (2*n+1)*la.*(F+(2*n+1)*la/4));
   E = sum(Z./Ra.*exp(-j*k*Ra) - Z./Rb.*exp(-j*k*Rb));

   f = F/la;
   Ef = f*(psi(f+N)-psi(f) + psi(f+1/2+N)-psi(f+1/2))    % psi(x)
   Dz = 0.886/N; z1 = 1-Dz/2; z2 = 1+Dz/2;               % Dz in units of F

   figure; plot(z,abs(E/Ef).^2,'b-', [z1,z2],[1,1]/2,'r-');
end
```

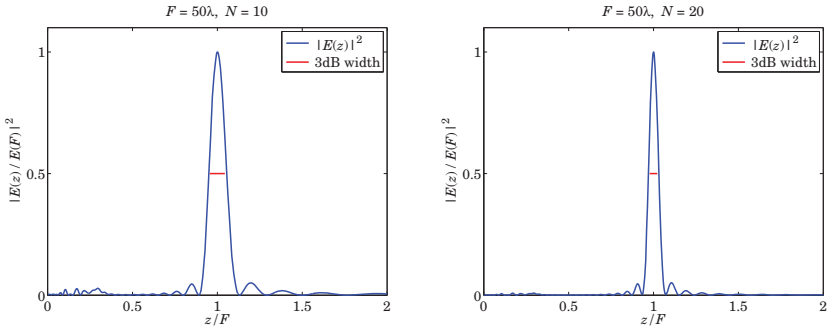


Fig. 20.3.2 Focusing by Fresnel zone plate.

[†]Eq. (20.3.13) was derived from the relationship, $\psi(x+N) - \psi(x) = \sum_{n=0}^{N-1} \frac{1}{x+n}$.

The peak at $z = F$ is the result of the constructive interference, and gets narrower with the number of zones N. The following approximate expression may be derived [1435] for its 3-dB width (i.e., full-width-at-half-maximum),

$$\Delta z = 0.886 \frac{F}{N} \tag{20.3.14}$$

The graphs display Δz as the short (red) horizontal line at the one-half level. To derive (20.3.14), we may use the Fresnel approximation of Eq. (20.3.10), as in Eq. (20.3.6),

$$E(0,z) = e^{-jkz} \sum_{n=0}^{N-1} \left[e^{-jkr_{2n}^2/2z} - e^{-jkr_{2n+1}^2/2z} \right] \tag{20.3.15}$$

We may also set approximately, $r_m^2 = m\lambda(F + m\lambda/4) \approx m\lambda F$, to obtain,

$$E(0,z) = e^{-jkz} \sum_{n=0}^{N-1} \left[e^{-2\pi jnF/z} - e^{-2\pi jnF/z} e^{-j\pi F/z} \right] = e^{-jkz} \left(1 - e^{-j\pi F/z}\right) \sum_{n=0}^{N-1} e^{-2\pi jnF/z}$$

$$= e^{-jkz} \left(1 - e^{-j\pi F/z}\right) \frac{1 - e^{-2\pi jNF/z}}{1 - e^{-2\pi jF/z}} = e^{-jkz} \frac{1 - e^{-2\pi jNF/z}}{1 + e^{-\pi jF/z}}$$

Since z is centered around F, we may set $z = F \pm \delta z$, for small δz, and write approximately, $F/z \approx 1 \mp \delta z/F$, and use the approximate value, $E(0,F) = 2Ne^{-jkF}$, for normalization,

$$\left| \frac{E(0,z)}{E(0,F)} \right| = \frac{1}{2N} \left| \frac{1 - e^{-2\pi jN(1\mp\delta z/F)}}{1 + e^{-\pi j(1\mp\delta z/F)}} \right| = \frac{1}{2N} \left| \frac{1 - e^{\pm 2\pi jN\delta z/F}}{1 - e^{\pm\pi j\delta z/F)}} \right| = \frac{\sin(\pi N\delta z/F)}{2N\sin(\pi\delta z/2F)}$$

We anticipate that $\delta z = xF/N$ with some constant x to be determined from the 3-dB condition, which reads after replacing $\delta z/F = x/N$,

$$\left| \frac{E(0,z)}{E(0,F)} \right| = \frac{\sin(\pi x)}{2N\sin\left(\frac{\pi x}{2N}\right)} \approx \frac{\sin(\pi x)}{2N \cdot \frac{\pi x}{2N}} = \frac{\sin(\pi x)}{\pi x} = \frac{1}{\sqrt{2}}$$

where we approximated $\sin(\pi x/2N) \approx \pi x/(2N)$ for large N. We note now that the solution of the equation, $\sin(\pi x)/(\pi x) = 1/\sqrt{2}$, is $x = 0.443$, so that $\delta z = 0.443F/N$, which leads to (20.3.14) for the full width, $\Delta z = 2\delta z$. □

20.4 Cascading of Optical Elements

In Fourier optics applications, one considers the passage of light through various optical elements that perform certain functions, such as Fourier transformation using lenses. For example, Fig. 20.4.1 shows an input field starting at aperture plane a, then propagating a distance z_1 to a thin optical element where it is modified by a transmittance function, and then propagating another distance z_2 to an aperture plane b.

Assuming that the input/output relationship of the optical element is multiplicative, $E_+(\boldsymbol{r}_\perp) = T(\boldsymbol{r}_\perp) E_-(\boldsymbol{r}_\perp)$, the relationship between the output field at plane b to the input

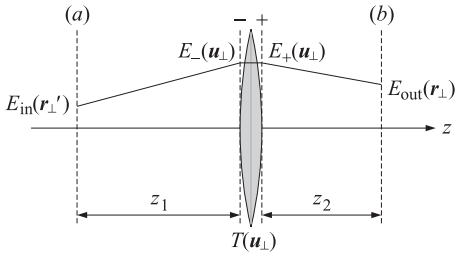


Fig. 20.4.1 Field propagated from plane a to plane b through a thin optical element.

field at plane a is obtained by successively applying the propagation equation (19.5.1), where in the optics context, the Fresnel form (20.1.4) is used for $h(\boldsymbol{r}_\perp, z)$,

$$\begin{aligned} E_{\text{out}}(\boldsymbol{r}_\perp) &= \int_S h(\boldsymbol{r}_\perp - \boldsymbol{u}_\perp, z_2)\, E_+(\boldsymbol{u}_\perp)\, d^2\boldsymbol{u}_\perp = \int_S h(\boldsymbol{r}_\perp - \boldsymbol{u}_\perp, z_2)\, T(\boldsymbol{u}_\perp)\, E_-(\boldsymbol{u}_\perp)\, d^2\boldsymbol{u}_\perp \\ &= \int_S h(\boldsymbol{r}_\perp - \boldsymbol{u}_\perp, z_2)\, T(\boldsymbol{u}_\perp)\, h(\boldsymbol{u}_\perp - \boldsymbol{r}'_\perp, z_1)\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{u}_\perp\, d^2\boldsymbol{r}'_\perp \\ &= \int_S h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \end{aligned}$$

where the overall transfer function from plane a to plane b will be:

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \int_S h(\boldsymbol{r}_\perp - \boldsymbol{u}_\perp, z_2)\, T(\boldsymbol{u}_\perp)\, h(\boldsymbol{u}_\perp - \boldsymbol{r}'_\perp, z_1)\, d^2\boldsymbol{u}_\perp \tag{20.4.1}$$

and we labeled the spatial x, y coordinates by $\boldsymbol{r}'_\perp, \boldsymbol{u}_\perp$, and $\boldsymbol{r}_\perp$ on the planes (a), the optical element, and plane (b).

In a similar fashion, one can work out the transfer function of more complicated configurations. For example, passing through two transmittance elements as shown in Fig. 20.4.2, we will have:

$$E_{\text{out}}(\boldsymbol{r}_\perp) = \int_S h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \tag{20.4.2}$$

where

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \int_S h(\boldsymbol{r}_\perp - \boldsymbol{u}_\perp, z_2)\, T_2(\boldsymbol{u}_\perp)\, h(\boldsymbol{u}_\perp - \boldsymbol{v}_\perp, z_0)\, T_1(\boldsymbol{v}_\perp)\, h(\boldsymbol{v}_\perp - \boldsymbol{r}'_\perp, z_1)\, d^2\boldsymbol{u}_\perp\, d^2\boldsymbol{v}_\perp \tag{20.4.3}$$

20.5 Lenses – Transmittance Properties

Lenses are probably the most important optical elements. Their interesting properties arise from their transmittance function, which has the quadratic phase:

$$\boxed{T(\boldsymbol{r}_\perp) = e^{jk|\boldsymbol{r}_\perp|^2/2F} = e^{jk(x^2+y^2)/2F}} \quad \text{(lens transmittance)} \tag{20.5.1}$$

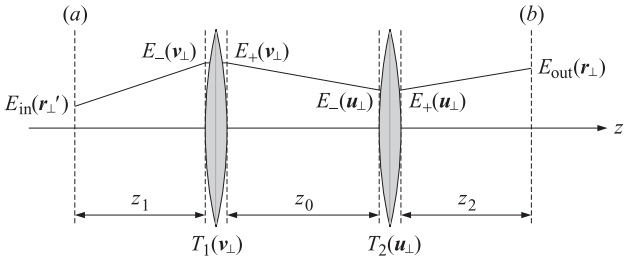


Fig. 20.4.2 Field propagated from plane a to plane b through multiple optical elements.

where $F > 0$ is the focal length. This applies to *convex* lenses. For *concave* ones, F must be replaced by $-F$ in (20.5.1).

Because the Fresnel propagation factor $e^{-jk|\boldsymbol{r}_\perp|^2/2z}$ also has the same type of quadratic phase, but with the *opposite* sign, it is possible for lenses to act as spatial "dispersion compensation" elements, much like the dispersion compensation and pulse compression filters of Chap. 3. Lenses have many uses, such as compensating propagation effects and focusing the waves on appropriate planes, or performing spatial Fourier transforms.

The transmittance function (20.5.1) can be derived with the help of Fig. 20.5.1, which shows a ray entering a (convex) spherical glass surface from the left at a distance x from the axis.

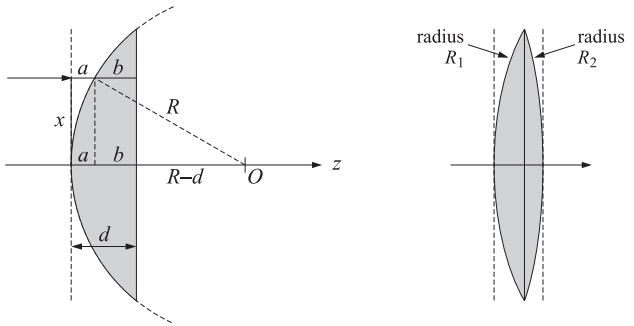


Fig. 20.5.1 Transmittance of a thin spherical lens.

Let R and d denote the radius of the spherical element, and its maximum width along its axis to the flat back plane, and let n be its refractive index. The wave travels a distance a in air and a distance b in the glass. If k is the free-space wavenumber, then in the glass it changes to $k_g = kn$. Therefore, the wave will accumulate the following phase as it propagates from the front plane to the back plane:

$$e^{-j\phi(x)} = e^{-jka}\, e^{-jk_g b}$$

where we are assuming a thin lens, which allows us to ignore the bending of the ray arising from refraction. Because, $a + b = d$, we have for the net phase:

$$\phi(x) = ka + k_g b = ka + nk(d-a) = nkd - (n-1)ka$$

The distance a is easily seen from the above figure to be:

$$a = R - \sqrt{R^2 - x^2}$$

Assuming that $x \ll R$, we can expand the square root to get:

$$a = R - R\sqrt{1 - \frac{x^2}{R^2}} \simeq R - R\left[1 - \frac{1}{2}\frac{x^2}{R^2}\right] = \frac{x^2}{2R}$$

Thus, the phase $\phi(x)$ is approximately,

$$\phi(x) = knd - (n-1)ka = knd - \frac{(n-1)kx^2}{2R}$$

If we make up a convex lens by putting together two such spherical lenses with radii R_1 and R_2, as shown in Fig. 20.5.1, then the net phase change between the front and back planes will be, ignoring the constant nkd terms:

$$\phi(x) = -(n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)\frac{kx^2}{2} \equiv -\frac{kx^2}{2F} \tag{20.5.2}$$

where we defined the focal length F of the lens through the "lensmaker's equation,"

$$\frac{1}{F} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{20.5.3}$$

In a two-dimensional description, we replace x^2 by $|\boldsymbol{r}_\perp|^2 = x^2 + y^2$. Thus, the phase change and corresponding transmittance function will be:

$$\phi(\boldsymbol{r}_\perp) = -\frac{k|\boldsymbol{r}_\perp|^2}{2F} \quad \Rightarrow \quad T(\boldsymbol{r}_\perp) = e^{-j\phi(\boldsymbol{r}_\perp)} = e^{jk|\boldsymbol{r}_\perp|^2/2F}$$

Some examples of the various effects that can be accomplished with lenses can be obtained by applying the configurations of Figs. 20.4.1 and 20.4.2 with appropriate choices of the aperture planes and focal lengths. We will use the Fresnel approximation (20.1.4) for $h(\boldsymbol{r}_\perp, z)$ in all of the examples and assume that the transmittance (20.5.1) extends over the entire xy plane—in effect, we are replacing the lens with the ideal case of an infinitely thin transparency with transmittance (20.5.1).

The main property of a lens is to turn an incident plane wave from the left into a spherical wave converging on the lens focus on the right, and similarly, if a source of a spherical wave is placed at the focus on the left, then the diverging wave will be converted into a plane wave after it passes through the lens. These cases are shown in Fig. 20.5.2.

The case on the left corresponds to the choices $z_1 = 0$ and $z_2 = F$ in Fig. 20.4.1, that is, the input plane coincides with the left plane of the lens. The incident wave has a

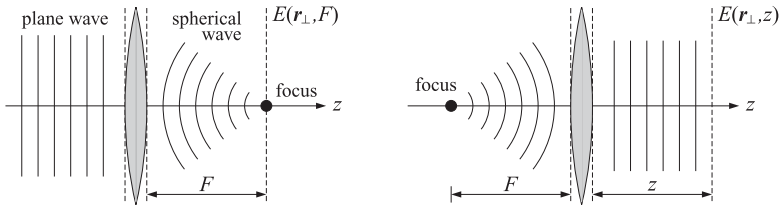


Fig. 20.5.2 Spherical waves converging to, or diverging from, a lens focal point.

constant amplitude on the plane $E_{\text{in}}(\boldsymbol{r}_\perp) = E_0$. Noting that $h(\boldsymbol{r}'_\perp - \boldsymbol{u}_\perp, 0) = \delta(\boldsymbol{r}'_\perp - \boldsymbol{u}_\perp)$,† we obtain from Eq. (20.4.1) with $z_1 = 0$ and $z_2 = F$,

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = T(\boldsymbol{r}'_\perp)\, h(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp, F) = \frac{jk}{2\pi F}\, e^{-jkF}\, e^{jk|\boldsymbol{r}'_\perp|^2/2F}\, e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2F}$$

the quadratic phase terms combine as follows:

$$e^{jk|\boldsymbol{r}'_\perp|^2/2F}\, e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2F} = e^{-jk|\boldsymbol{r}_\perp|^2/2F}\, e^{jk\boldsymbol{r}_\perp \cdot \boldsymbol{r}'_\perp/F}$$

and result in the following transfer function from the left plane of the lens to its right *focal* point,

$$\boxed{h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = T(\boldsymbol{r}'_\perp)\, h(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp, F) = \frac{jk}{2\pi F}\, e^{-jkF}\, e^{-jk|\boldsymbol{r}_\perp|^2/2F}\, e^{jk\boldsymbol{r}_\perp \cdot \boldsymbol{r}'_\perp/F}} \tag{20.5.4}$$

Its integration with the constant input results in:

$$E(\boldsymbol{r}_\perp, F) = \int_S h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp = \frac{jk}{2\pi F}\, e^{-jkF}\, E_0 e^{-jk|\boldsymbol{r}_\perp|^2/2F} \int_{-\infty}^{\infty} e^{jk\boldsymbol{r}_\perp \cdot \boldsymbol{r}'_\perp/F}\, d^2\boldsymbol{r}'_\perp$$

The integral is equal to the Dirac delta, $(2\pi)^2\delta(k\boldsymbol{r}_\perp/F) = (2\pi)^2\delta(\boldsymbol{r}_\perp)F^2/k^2$.† Thus,

$$E(\boldsymbol{r}_\perp, F) = \frac{j2\pi F}{k}\, e^{-jkF}\, E_0\, \delta(\boldsymbol{r}_\perp) \tag{20.5.5}$$

which is sharply focused onto the focal point $\boldsymbol{r}_\perp = 0$ and $z = F$. This represents the ideal case of a lense of infinite extent. The case of a lens with a finite aperture is discussed later. For the second case depicted on the right in Fig. 20.5.2, we first note that the paraxial approximation for a spherical wave of a point source placed at the origin is obtained by setting, $E(\boldsymbol{r}'_\perp, 0) = E_0\, \delta(\boldsymbol{r}'_\perp)$, in Eq. (20.1.1) and expanding for large z,

$$E_0\, \frac{2z}{r}\left(jk + \frac{1}{r}\right)\frac{e^{-jkr}}{4\pi r} \simeq E_0\, \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z}, \quad r = \sqrt{z^2 + |\boldsymbol{r}_\perp|^2} \simeq z + \frac{|\boldsymbol{r}_\perp|^2}{2z}$$

If this source is placed at the left focal point of the lens, then, the diverging paraxial spherical wave, after traveling distance $z = F$, will arrive at the left plane of the lens as,

$$E_{\text{in}}(\boldsymbol{r}'_\perp, 0) = E_0\, \frac{jk}{2\pi F}\, e^{-jkF}\, e^{-jk|\boldsymbol{r}'_\perp|^2/2F} \equiv E_1\, e^{-jk|\boldsymbol{r}'_\perp|^2/2F}$$

†Here, $\delta(\boldsymbol{r}_\perp) = \delta(x)\,\delta(y)$, denotes the two-dimensional delta function.

†where we used the property, $\delta(a\boldsymbol{r}_\perp) = \delta(\boldsymbol{r}_\perp)/a^2$, for 2-D delta functions.

The transmittance of the lens will compensate this propagation phase resulting into a constant field at the output plane of the lens, which will then propagate to the right as a plane wave:

$$E(\boldsymbol{r}_\perp', 0) = T(\boldsymbol{r}_\perp') E_{\text{in}}(\boldsymbol{r}_\perp', 0) = e^{jk|\boldsymbol{r}_\perp'|^2/2F} E_1 e^{-jk|\boldsymbol{r}_\perp'|^2/2F} = E_1$$

The propagated field to distance z is obtained from Eq. (20.1.3):

$$E(\boldsymbol{r}_\perp, z) = e^{-jkz} \frac{jk}{2\pi z} \int_{-\infty}^{\infty} E_1 e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2/2z} d^2\boldsymbol{r}_\perp' = E_1 e^{-jkz} \frac{jk}{2\pi z} \frac{2\pi z}{jk} = E_1 e^{-jkz}$$

where the integral was evaluated using twice the result of Eq. (20.1.14). Thus, the transmitted wave is a uniform plane wave propagating along the z-direction.

20.6 Magnification Properties of Lenses

Next, we look at the magnifying properties of a lens. Fig. 20.6.1 shows an image placed at distance z_1 from the left and its image formed at distance z_2 on the right. It is well-known that the distances z_1, z_2 must be related by:

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{F} \tag{20.6.1}$$

The magnification law relates the size of the image to the size of the object:

$$M = \frac{x_2}{x_1} = \frac{z_2}{z_1} \qquad \text{(magnification ratio)} \tag{20.6.2}$$

These properties can be derived by tracing the rays emanating from the top of the object. The ray that is parallel to the lens axis will bend to pass through the focal point on the right. The ray from the top of the object through the left focal point will bend to become parallel to the axis. The intersection of these two rays defines the top point of the image. From the geometry of the graph one has:

$$\frac{x_1}{z_1 - F} = \frac{x_2}{F} \quad \text{and} \quad \frac{x_2}{z_2 - F} = \frac{x_1}{F}$$

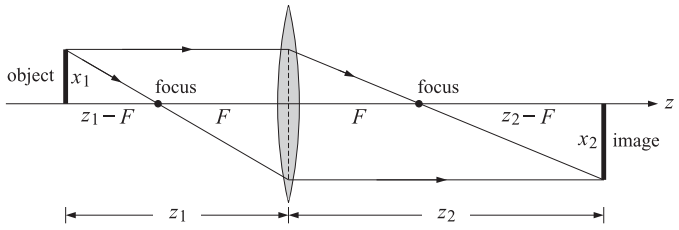


Fig. 20.6.1 Lens law of magnification.

The consistency of these equations requires the condition, $(z_1 - F)(z_2 - F) = F^2$, which is known as Newton's relationship and is equivalent to (20.6.1). Then, Eq. (20.6.2) follows by replacing F from (20.6.1) into the ratio, $x_2/x_1 = (z_2 - F)/F$.†

To understand (20.6.1) and (20.6.2) from the point of view of Fresnel diffraction, we note that the transfer function (20.4.1) involves the following quadratic phase factors, with the middle one being the lens transmittance:

$$e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{u}_\perp|^2/2z_2} e^{jk|\boldsymbol{u}_\perp|^2/2F} e^{-jk|\boldsymbol{u}_\perp - \boldsymbol{r}_\perp'|^2/2z_1}$$

$$= e^{-jk|\boldsymbol{r}_\perp|^2/2z_2} e^{-jk|\boldsymbol{r}_\perp'|^2/2z_1} e^{-jk(1/z_1 + 1/z_2 - 1/F)|\boldsymbol{u}_\perp|^2/2} e^{jk\boldsymbol{u}_\perp \cdot (\boldsymbol{r}_\perp/z_2 + \boldsymbol{r}_\perp'/z_1)}$$

Because of Eq. (20.6.1), the term that depends quadratically on $\boldsymbol{u}_\perp$ cancels and one is left only with a linear dependence on $\boldsymbol{u}_\perp$. This integrates into a delta function in (20.4.1), resulting in

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}_\perp') = \frac{jke^{-jkz_1}}{2\pi z_1} \frac{jke^{-jkz_2}}{2\pi z_2} e^{-jk|\boldsymbol{r}_\perp|^2/2z_2} e^{-jk|\boldsymbol{r}_\perp'|^2/2z_1} (2\pi)^2 \delta\left(\frac{k\boldsymbol{r}_\perp}{z_2} + \frac{k\boldsymbol{r}_\perp'}{z_1}\right) \tag{20.6.3}$$

The delta function forces $\boldsymbol{r}_\perp = -(z_2/z_1)\boldsymbol{r}_\perp'$, which is the same as (20.6.2). The negative sign means that the image is upside down. Noting that

$$\delta\left(\frac{k\boldsymbol{r}_\perp}{z_2} + \frac{k\boldsymbol{r}_\perp'}{z_1}\right) = \frac{z_1^2}{k^2} \delta\left(\boldsymbol{r}_\perp' + \frac{z_1}{z_2}\boldsymbol{r}_\perp\right)$$

we obtain for the field at the output plane:

$$E_{\text{out}}(\boldsymbol{r}_\perp) = -\frac{z_1}{z_2} E_{\text{in}}\left(-\frac{z_1}{z_2}\boldsymbol{r}_\perp\right) e^{-jk(z_1+z_2)} e^{-jk|\boldsymbol{r}_\perp|^2(z_1+z_2)/2z_2^2} \tag{20.6.4}$$

which represents a scaled and reversed version of the input. The overall negative sign comes from the two j factors in (20.6.3), which are usually ignored in the literature. The *geometrical* image, i.e., the image according to geometrical optics, will be denoted by,

$$E_g(\boldsymbol{r}_\perp) = -\frac{z_1}{z_2} E_{\text{in}}\left(-\frac{z_1}{z_2}\boldsymbol{r}_\perp\right) \tag{20.6.5}$$

therefore, ignoring the phase factors, $e^{-jk(z_1+z_2)} e^{-jk|\boldsymbol{r}_\perp|^2(z_1+z_2)/2z_2^2}$,

$$E_{\text{out}}(\boldsymbol{r}_\perp) = -\frac{z_1}{z_2} E_{\text{in}}\left(-\frac{z_1}{z_2}\boldsymbol{r}_\perp\right) \quad \Leftrightarrow \quad -\frac{z_2}{z_1} E_{\text{out}}\left(-\frac{z_2}{z_1}\boldsymbol{r}_\perp\right) = E_{\text{in}}(\boldsymbol{r}_\perp) \tag{20.6.6}$$

20.7 Point-Spread Function of a Lens

In the previous section, we considered a lens with infinite extent, which led to the delta-function transfer function (20.6.3) and the geometric-optics result (20.6.4).

A thin lens with a finite aperture is characterized by its so called *exit pupil* function $P(\boldsymbol{u}_\perp)$ of the transverse coordinates $\boldsymbol{u}_\perp$ on the lens plane, which vanishes outside the extent of the lens. In its simplest form, $P(\boldsymbol{u}_\perp)$ is the characteristic or indicator function

† here, x_1, x_2 are positive, representing the geometrical lengths shown in Fig. 20.6.1.

of the aperture, which is equal to unity for $\boldsymbol{u}_\perp$ in the aperture, and zero outside it. For example, an ordinary circular lens of radius a, will have a pupil function defined in terms of the unit-step function $u(x)$ as follows,

$$P(\boldsymbol{u}_\perp) = u(a - |\boldsymbol{u}_\perp|) = \begin{cases} 1, & |\boldsymbol{u}_\perp| < a \\ 0, & |\boldsymbol{u}_\perp| > a \end{cases}$$

The overall transfer function of the lens, from its entry plane to its exit plane, will be $P(\boldsymbol{u}_\perp)\, e^{jk|\boldsymbol{u}_\perp|^2/2F}$. More generally, as shown in Fig. 20.7.1, an appropriate filter may be placed at the exit pupil, which may be designed to achieve certain properties, such as improving resolution or reducing the sidelobes. In such case, the effective pupil function will be the product of the filter and the aperture's indicator function.

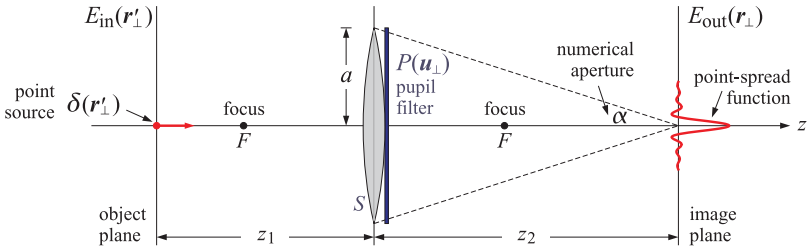


Fig. 20.7.1 Lens imaging system and its point-spread function.

The transfer relationship between the fields at the input and output planes will have the usual linear form,

$$E_{\text{out}}(\boldsymbol{r}_\perp) = \int h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \tag{20.7.1}$$

but with the transfer function integral being restricted over the lens aperture S only,

$$\begin{aligned} h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) &= \int_S h(\boldsymbol{r}_\perp - \boldsymbol{u}_\perp, z_2)\, P(\boldsymbol{u}_\perp)\, e^{jk|\boldsymbol{u}_\perp|^2/2F}\, h(\boldsymbol{u}_\perp - \boldsymbol{r}'_\perp, z_1)\, d^2\boldsymbol{u}_\perp \\ &= \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2} \int_S e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{u}_\perp|^2/2z_2}\, P(\boldsymbol{u}_\perp)\, e^{jk|\boldsymbol{u}_\perp|^2/2F}\, e^{-jk|\boldsymbol{u}_\perp - \boldsymbol{r}'_\perp|^2/2z_1}\, d^2\boldsymbol{u}_\perp \end{aligned}$$

Assuming that z_1, z_2 satisfy the lens law (20.6.1), the phase factors that are quadratic in $\boldsymbol{u}_\perp$ would still cancel as before, resulting in,

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z_2}\, e^{-jk|\boldsymbol{r}'_\perp|^2/2z_1} \int_S P(\boldsymbol{u}_\perp)\, e^{jk\boldsymbol{u}_\perp\cdot(\boldsymbol{r}_\perp/z_2 + \boldsymbol{r}'_\perp/z_1)}\, d^2\boldsymbol{u}_\perp \tag{20.7.2}$$

The last integral is recognized as the 2-D spatial Fourier transform of the pupil function $P(\boldsymbol{u}_\perp)$, that is,

$$\boxed{\hat{P}(\boldsymbol{k}_\perp) = \int_S P(\boldsymbol{u}_\perp)\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{u}_\perp}\, d^2\boldsymbol{u}_\perp} \quad \text{(pupil's 2-D Fourier transform)} \tag{20.7.3}$$

where the finite support of $P(\boldsymbol{u}_\perp)$ restricts the integration to be over the aperture only. Thus, (20.7.2) becomes,

$$\boxed{\begin{aligned} h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) &= \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z_2}\, e^{-jk|\boldsymbol{r}'_\perp|^2/2z_1}\, \Big[\hat{P}(\boldsymbol{k}_\perp)\Big]_{\boldsymbol{k}_\perp = k\left(\frac{\boldsymbol{r}_\perp}{z_2} + \frac{\boldsymbol{r}'_\perp}{z_1}\right)} \\ &= \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z_2}\, e^{-jk|\boldsymbol{r}'_\perp|^2/2z_1}\, \hat{P}\left(\frac{k\boldsymbol{r}_\perp}{z_2} + \frac{k\boldsymbol{r}'_\perp}{z_1}\right) \end{aligned}} \tag{20.7.4}$$

This is the *point-spread function* (PSF) also known as the *impulse response* of the lens system. The name is justified by considering a delta-function point source at the origin, $E_{\text{in}}(\boldsymbol{r}'_\perp) = \delta(\boldsymbol{r}'_\perp)$, then, Eq. (20.7.1) gives,

$$E_{\text{out}}(\boldsymbol{r}_\perp) = \int h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)\, \delta(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp = h(\boldsymbol{r}_\perp, 0)$$

Similarly, if the point source is centered at $\boldsymbol{r}^0_\perp$, that is, $E_{\text{in}}(\boldsymbol{r}'_\perp) = \delta(\boldsymbol{r}'_\perp - \boldsymbol{r}^0_\perp)$, then, $E_{\text{out}}(\boldsymbol{r}_\perp) = h(\boldsymbol{r}_\perp, \boldsymbol{r}^0_\perp)$.

Example 20.7.1: For a circular uncoated lens aperture of radius a, the 2-D transform of the pupil reduces to the usual Airy pattern, $2J_1(x)/x$,

$$\hat{P}(k_\perp) = \int_S e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{u}_\perp}\, d^2\boldsymbol{u}_\perp = \int_0^a J_0(k_\perp\rho)\, 2\pi\rho d\rho = \pi a^2\, \frac{2J_1(k_\perp a)}{k_\perp a} \tag{20.7.5}$$

where $k_\perp = |\boldsymbol{k}_\perp|$. The PSF is then, with $r_\perp = |\boldsymbol{r}_\perp|$,

$$h(r_\perp, 0) = \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2}\, e^{-jkr_\perp^2/2z_2}\, \pi a^2\, \frac{2J_1\left(\frac{kr_\perp a}{z_2}\right)}{\frac{kr_\perp a}{z_2}} \tag{20.7.6}$$

This is depicted in Fig. 20.7.1. As seen in that figure, the geometrical optics cone from the lens to the origin on the image plane is defined by the *numerical aperture* ratio, $N_A = a/z_2 = \tan\alpha$, which for small angles α can be written approximately as $N_A = a/z_2 = \tan\alpha \approx \sin\alpha \approx \alpha$. Thus, the PSF reads,

$$h(\boldsymbol{r}_\perp, 0) = \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2}\, e^{-jkr_\perp^2/2z_2}\, \pi a^2\, \frac{2J_1(kr_\perp N_A)}{kr_\perp N_A}$$

A related point of view is to introduce the angle θ defined by, $\tan\theta = r_\perp/z_2$, subtended from the origin of the lens to the point $r_\perp$ on the image plane. For small such angles, we can approximate again, $\sin\theta \approx \tan\theta = r_\perp/z_2$, and

$$h(\boldsymbol{r}_\perp, 0) = \frac{jke^{-jkz_1}}{2\pi z_1}\, \frac{jke^{-jkz_2}}{2\pi z_2}\, e^{-jkr_\perp^2/2z_2}\, \pi a^2\, \frac{2J_1(ka\sin\theta)}{ka\sin\theta}$$

This is essentially the same as the result of Eq. (18.9.3), the only difference being that here we are using the paraxial Fresnel approximation to describe the propagation through the lens system, whereas in Sec. 18.9 we worked with the radiation-field approximation. □

20.8 Cylindrically-Symmetric and One-Dimensional Lenses

More generally, for a circularly-symmetric pupil function, Eq. (20.7.3) reduces to the pupil's Hankel transform, that is, with $k_\perp = |\boldsymbol{k}_\perp|$,

$$\hat{P}(k_\perp) = \int_0^a P(\rho) J_0(k_\perp \rho)\, 2\pi\rho\, d\rho \tag{20.8.1}$$

and the PSF becomes,

$$h(r_\perp, 0) = \frac{jke^{-jkz_1}}{2\pi z_1} \frac{jke^{-jkz_2}}{2\pi z_2} e^{-jkr_\perp^2/2z_2}\, \hat{P}\Big(\frac{kr_\perp}{z_2}\Big) \tag{20.8.2}$$

For the case of a 1-D cylindrical lens, we will have the input/output relationship,

$$E_{\text{out}}(x) = \int h(x,x')\, E_{\text{in}}(x')\, dx' \tag{20.8.3}$$

The 1-D spatial Fourier transform of the pupil,

$$\hat{P}(k_x) = \int_S P(u)\, e^{jk_x u}\, du \qquad \text{(pupil's 1-D Fourier transform)} \tag{20.8.4}$$

And, the transfer function between input and output planes,

$$\begin{aligned} h(x,x') &= \sqrt{\frac{jk}{2\pi z_1}} \sqrt{\frac{jk}{2\pi z_2}}\, e^{-jk(z_1+z_2)} e^{-jkx^2/2z_2} e^{-jkx'^2/2z_1} \Big[\hat{P}(k_x)\Big]_{k_x = k\left(\frac{x}{z_2}+\frac{x'}{z_1}\right)} \\ &= \sqrt{\frac{jk}{2\pi z_1}} \sqrt{\frac{jk}{2\pi z_2}}\, e^{-jk(z_1+z_2)} e^{-jkx^2/2z_2} e^{-jkx'^2/2z_1}\, \hat{P}\Big(\frac{kx}{z_2} + \frac{kx'}{z_1}\Big) \end{aligned} \tag{20.8.5}$$

Example 20.8.1: As an example of an uncoated lens of extent, $-a \le u \le a$, we have,

$$\hat{P}(k_x) = \int_{-a}^{a} e^{jk_x u}\, du = 2a\, \frac{\sin(k_x a)}{k_x a} \tag{20.8.6}$$

This is the 1-D version of (20.7.5). The 1-D PSF is then,

$$h(x,0) = \sqrt{\frac{jk}{2\pi z_1}} \sqrt{\frac{jk}{2\pi z_2}}\, e^{-jk(z_1+z_2)} e^{-jkx^2/2z_2}\, 2a\, \frac{\sin(kxa/z_2)}{kxa/z_2}$$

20.9 Shift-Invariance and Coherent Transfer Function

As it stands, the transfer function $h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)$ of Eq. (20.7.4) does not depend on the difference, $\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp$, and therefore, the input/output transfer relationship (20.7.1) is linear but not shift-invariant. However, under certain conditions, shift-invariance can be argued. First, let us define the geometrical-optics images of the PSF and the input projected onto the image plane, as in (20.6.5),

$$\begin{aligned} h_g(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) &= -\frac{z_1}{z_2}\, h\Big(\boldsymbol{r}_\perp, -\frac{z_1}{z_2}\, \boldsymbol{r}'_\perp\Big) \\ E_g(\boldsymbol{r}_\perp) &= -\frac{z_1}{z_2}\, E_{\text{in}}\Big(-\frac{z_1}{z_2}\, \boldsymbol{r}_\perp\Big) \end{aligned} \tag{20.9.1}$$

Changing integration variables, the input/output equation (20.7.1) can be rewritten as,

$$E_{\text{out}}(\boldsymbol{r}_\perp) = \int h(\boldsymbol{r}_\perp, \boldsymbol{r}''_\perp) E_{\text{in}}(\boldsymbol{r}''_\perp)\, d^2\boldsymbol{r}''_\perp = \int h_g(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) E_g(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \tag{20.9.2}$$

The scaled PSF, $h_g(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)$, comes close to being shift-invariant,

$$\begin{aligned} h(\boldsymbol{r}_\perp, \boldsymbol{r}''_\perp) &= \frac{jke^{-jkz_1}}{2\pi z_1} \frac{jke^{-jkz_2}}{2\pi z_2} e^{-jk|\boldsymbol{r}_\perp|^2/2z_2} e^{-jk|\boldsymbol{r}''_\perp|^2/2z_1}\, \hat{P}\Big(\frac{k\boldsymbol{r}_\perp}{z_2} + \frac{k\boldsymbol{r}''_\perp}{z_1}\Big) \\ h_g(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) &= -\frac{z_1}{z_2} \frac{jke^{-jkz_1}}{2\pi z_1} \frac{jke^{-jkz_2}}{2\pi z_2} e^{-jk|\boldsymbol{r}_\perp|^2/2z_2} e^{-jk|\boldsymbol{r}'_\perp|^2 z_1/2z_2^2}\, \hat{P}\Big(\frac{k}{z_2}(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp)\Big) \end{aligned}$$

where we replaced $\boldsymbol{r}''_\perp = -z_1\boldsymbol{r}'_\perp/z_2$. The phase factor, $e^{-jk|\boldsymbol{r}_\perp|^2/2z_2}$ can be ignored since it only affects the phase of the output, not its magnitude. The other phase factor $e^{-jk|\boldsymbol{r}''_\perp|^2/2z_1} = e^{-jk|\boldsymbol{r}'_\perp|^2 z_1/2z_2^2}$ could be ignored [1422,1446] if the size of the input is small enough such that $k|\boldsymbol{r}''_\perp|^2 \ll z_1$. Under these circumstances, the effective PSF becomes shift-invariant,

$$\boxed{h_g(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = h_g(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp) = e^{-jk(z_1+z_2)}\, \frac{k^2}{(2\pi)^2 z_2^2}\, \hat{P}\Big(\frac{k}{z_2}(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp)\Big)} \tag{20.9.3}$$

The linearity and shift-invariance of the input/output relationship allows one to define the *spatial frequency response* of the lens system, also called the *coherent transfer function* (CTF) or the *amplitude transfer function*, defined as the 2-D spatial Fourier transform of the PSF, that is,

$$\hat{h}_g(\boldsymbol{k}_\perp) = \int_{-\infty}^{\infty} h_g(\boldsymbol{r}_\perp)\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, d^2\boldsymbol{r}_\perp \tag{20.9.4}$$

The transfer relationship (20.9.2) can now be expressed either convolutionally in the space domain or multiplicatively in the wavenumber domain in terms of the corresponding 2-D Fourier transforms,

$$E_{\text{out}}(\boldsymbol{r}_\perp) = \int h_g(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp) E_g(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \quad \Leftrightarrow \quad \hat{E}_{\text{out}}(\boldsymbol{k}_\perp) = \hat{h}_g(\boldsymbol{k}_\perp)\, \hat{E}_g(\boldsymbol{k}_\perp) \tag{20.9.5}$$

Inserting (20.9.3) into (20.9.4) we obtain,

$$\hat{h}_g(\boldsymbol{k}_\perp) = e^{-jk(z_1+z_2)} \int_{-\infty}^{\infty} \hat{P}\Big(\frac{k\boldsymbol{r}_\perp}{z_2}\Big)\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{k^2}{z_2^2}\, \frac{d^2\boldsymbol{r}_\perp}{(2\pi)^2} \tag{20.9.6}$$

But the pupil $P(\boldsymbol{u}_\perp)$ is related to $\hat{P}(\boldsymbol{k}_\perp)$ by the inverse Fourier transform,

$$P(\boldsymbol{u}_\perp) = \int_{-\infty}^{\infty} \hat{P}(\boldsymbol{k}_\perp) e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{u}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2}$$

If in (20.9.6) we make the replacements, $\boldsymbol{u}_\perp = -\boldsymbol{k}_\perp z_2/k$, and $\boldsymbol{k}_\perp = \boldsymbol{r}_\perp k/z_2$ and change variables of integration, we recognize that $\hat{h}_g(\boldsymbol{k}_\perp)$ is essentially the pupil function itself,

$$\boxed{\hat{h}_g(\boldsymbol{k}_\perp) = e^{-jk(z_1+z_2)} \Big[P(\boldsymbol{u}_\perp)\Big]_{\boldsymbol{u}_\perp = -\frac{\boldsymbol{k}_\perp z_2}{k}} = e^{-jk(z_1+z_2)}\, P\Big(-\frac{\boldsymbol{k}_\perp z_2}{k}\Big)} \tag{20.9.7}$$

Because $P(\boldsymbol{u}_\perp)$ is space-limited, Eq. (20.9.7) implies that the frequency response will be *bandlimited*. For example, for a pupil of radius a, bandlimiting restricts the range of wavenumbers to be less than a maximum value,

$$|\boldsymbol{u}_\perp| < a \quad \Rightarrow \quad \left|\frac{\boldsymbol{k}_\perp z_2}{k}\right| < a \quad \Rightarrow \quad \boxed{|\boldsymbol{k}_\perp| < k_{\max} = \frac{ka}{z_2} = kN_A} \tag{20.9.8}$$

Thus, the lens system acts as a *spatial low-pass filter* that reduces the resolution of the image. The smallest transverse distance that can be resolved is roughly of the order of $1/k_{\max}$ or of the order of λ/N_A. This is effectively the so-called *Abbe diffraction limit* of the resolving power of optical systems, such as a microscope.

We mention also that similar results hold in the 1-D case, where now one must define the scaled geometric-image quantities by,

$$E_g(x) = e^{j\pi/2}\sqrt{\frac{z_1}{z_2}}\,E_{\text{in}}\left(-\frac{z_1}{z_2}x\right), \quad h_g(x,x') = e^{-j\pi/2}\sqrt{\frac{z_1}{z_2}}\,h\left(x,-\frac{z_1}{z_2}x'\right) \tag{20.9.9}$$

with input/output relationship,

$$E_{\text{out}}(x) = \int_{-\infty}^{\infty} h(x,x')E_{\text{in}}(x')\,dx' = \int_{-\infty}^{\infty} h_g(x,x')E_g(x')\,dx' \tag{20.9.10}$$

Under the same type of approximations, we have the shift-invariant versions,

$$\begin{aligned}
h(x,x'') &= \sqrt{\frac{jk}{2\pi z_1}\frac{jk}{2\pi z_2}}\,e^{-jk(z_1+z_2)}\,e^{-jkx^2/2z_2}e^{-jkx''^2/2z_1}\,\hat{P}\left(\frac{kx}{z_2}+\frac{kx''}{z_1}\right)\\
h_g(x,x') &= -j\sqrt{\frac{z_1}{z_2}\frac{jk}{2\pi z_1}\frac{jk}{2\pi z_2}}\,e^{-jk(z_1+z_2)}\,e^{-jkx^2/2z_2}e^{-jkx'^2z_1/2z_2^2}\,\hat{P}\left(\frac{k}{z_2}(x-x')\right)\\
h_g(x,x') &\approx h_g(x-x') = \frac{k}{2\pi z_2}\,e^{-jk(z_1+z_2)}\,\hat{P}\left(\frac{k}{z_2}(x-x')\right)
\end{aligned}$$

with coherent transfer function,

$$\hat{h}_g(k_x) = e^{-jk(z_1+z_2)}\int_{-\infty}^{\infty}\hat{P}\left(\frac{kx}{z_2}\right)e^{jk_xx}\,\frac{k}{z_2}\,\frac{dx}{2\pi} = e^{-jk(z_1+z_2)}\,P\left(-\frac{k_xz_2}{k}\right) \tag{20.9.11}$$

20.10 Fourier Transformation Properties of Lenses

To see the Fourier transformation property of lenses, consider the left picture in Fig. 20.5.2 with the output plane placed at the right focal point $z_2 = F$, but take an arbitrary field $E_{\text{in}}(\boldsymbol{r}_\perp')$ incident at the left plane of the lens, as shown in Fig. 20.10.1.

Assuming initially an infinite lens, the overall transfer function is still the same as in Eq. (20.5.4), thus, giving:

$$\begin{aligned}
E(\boldsymbol{r}_\perp,F) &= \int_{-\infty}^{\infty} h(\boldsymbol{r}_\perp,\boldsymbol{r}_\perp')E_{\text{in}}(\boldsymbol{r}_\perp')\,d^2\boldsymbol{r}_\perp'\\
&= \frac{jk}{2\pi F}\,e^{-jkF}\,e^{-jk|\boldsymbol{r}_\perp|^2/2F}\int_{-\infty}^{\infty}E_{\text{in}}(\boldsymbol{r}_\perp')\,e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}_\perp'/F}\,d^2\boldsymbol{r}_\perp'
\end{aligned} \tag{20.10.1}$$

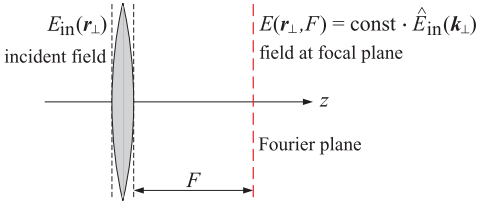


Fig. 20.10.1 Fourier transformation at focal plane of a lens.

The last integral factor is recognized as the Fourier transform of the input, $\hat{E}_{\text{in}}(\boldsymbol{k}_\perp)$, evaluated at wavenumber, $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/F$. Thus, we obtain:

$$\boxed{E(\boldsymbol{r}_\perp,F) = \frac{jk}{2\pi F}\,e^{-jkF}\,e^{-jk|\boldsymbol{r}_\perp|^2/2F}\left[\hat{E}_{\text{in}}(\boldsymbol{k}_\perp)\right]_{\boldsymbol{k}_\perp=\frac{k\boldsymbol{r}_\perp}{F}}} \tag{20.10.2}$$

This result is similar to the Fraunhofer case of Eq. (20.3.1), but it is valid at the much shorter Fresnel distance $z = F$, instead of the far-field distances. It is analogous to the output of the pulse compression filter in chirp radar discussed in Chap. 3, see for example Eq. (3.10.14). Another difference is that (20.10.2) is valid only at the focal distance $z = F$, whereas the Fraunhofer equation (20.3.1) is valid at any (large) z. In the 1-D case, Eq. (20.10.2) reads,

$$\begin{aligned}
\boxed{\begin{aligned}
E(x,F) &= \sqrt{\frac{jk}{2\pi F}}\,e^{-jkF}\,e^{-jkx^2/2F}\int_{-\infty}^{\infty}E_{\text{in}}(x')\,e^{jkxx'/F}\,dx'\\
&= \sqrt{\frac{jk}{2\pi F}}\,e^{-jkF}\,e^{-jkx^2/2F}\left[\hat{E}_{\text{in}}(k_x)\right]_{k_x=\frac{kx}{F}}
\end{aligned}}
\end{aligned} \tag{20.10.3}$$

The focusing result of Eq. (20.5.5) is a special case of (20.10.2); indeed, a constant input, $E_{\text{in}}(\boldsymbol{r}_\perp) = E_0$, has the 2-D Fourier transform $\hat{E}_{\text{in}}(\boldsymbol{k}_\perp) = \delta(\boldsymbol{k}_\perp)$, which leads to (20.5.5) when evaluated at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/F$.

When the finite extent of the lens is taken into account with a pupil function $P(\boldsymbol{r}_\perp')$, then, the output $E(\boldsymbol{r}_\perp,F)$ will be proportional to the 2-D Fourier transform of the product $E_{\text{in}}(\boldsymbol{r}_\perp')P(\boldsymbol{r}_\perp')$, with the integration restricted over the lens aperture, that is,

$$E(\boldsymbol{r}_\perp,F) = \frac{jk}{2\pi F}\,e^{-jkF}\,e^{-jk|\boldsymbol{r}_\perp|^2/2F}\int_S E_{\text{in}}(\boldsymbol{r}_\perp')\,P(\boldsymbol{r}_\perp')\,e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}_\perp'/F}\,d^2\boldsymbol{r}_\perp' \tag{20.10.4}$$

Fig. 20.10.2 shows two additional configurations of interest. One is when the input plane lies at some distance d to the left of the lens but the output plane is still the focal plane on the right at $z = F$. The other is when the input plane coincides with the left plane of the lens at $z = 0$, but the output plane is slightly axially shifted to $z \neq F$, referred to as axial defocusing.

In the first case, we may apply the result of Eq. (20.10.2) to the field $E_a(\boldsymbol{r}_\perp)$ at the left plane of the lens. Its Fourier transform is obtained by propagating the input's Fourier

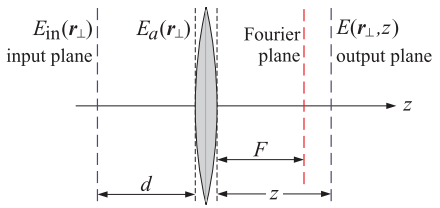


Fig. 20.10.2 Fourier transformation and axial defocusing.

transform by a distance d, using the Fresnel propagation filter (20.1.4), that is,

$$\hat{E}_a(\boldsymbol{k}_\perp) = \hat{h}(\boldsymbol{k}_\perp, d)\,\hat{E}_{\rm in}(\boldsymbol{k}_\perp) = e^{-jkd}\, e^{j|\boldsymbol{k}_\perp|^2 d/2k}\,\hat{E}_{\rm in}(\boldsymbol{k}_\perp)$$

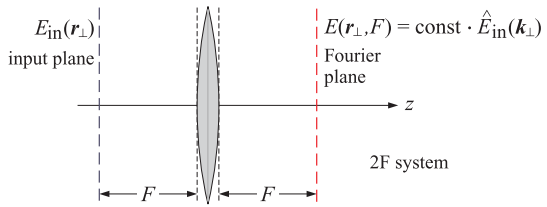
thus we find at the output plane $z = F$,

$$\begin{aligned} E(\boldsymbol{r}_\perp, F) &= \frac{jk}{2\pi F}\, e^{-jkF}\, e^{-jk|\boldsymbol{r}_\perp|^2/2F} \left[\hat{E}_a(\boldsymbol{k}_\perp)\right]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{F}} \\ &= \frac{jk}{2\pi F}\, e^{-jkF}\, e^{-jk|\boldsymbol{r}_\perp|^2/2F} \left[e^{-jkd}\, e^{j|\boldsymbol{k}_\perp|^2 d/2k}\,\hat{E}_{\rm in}(\boldsymbol{k}_\perp)\right]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{F}}, \quad \text{or,} \end{aligned}$$

$$\boxed{\begin{aligned} E(\boldsymbol{r}_\perp, F) &= \frac{jk}{2\pi F}\, e^{-jk(d+F)}\, e^{-jk|\boldsymbol{r}_\perp|^2(F-d)/2F^2} \left[\hat{E}_{\rm in}(\boldsymbol{k}_\perp)\right]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{F}} \\ &= \frac{jk}{2\pi F}\, e^{-jk(d+F)}\, e^{-jk|\boldsymbol{r}_\perp|^2(F-d)/2F^2} \int_{-\infty}^{\infty} E_{\rm in}(\boldsymbol{r}'_\perp)\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/F}\, d^2\boldsymbol{r}'_\perp \end{aligned}} \tag{20.10.5}$$

The result is still proportional to the Fourier transform of the input. In the special case when $d = F$, known as a $2F$-system, the quadratic phase factor is eliminated,

$$\boxed{\begin{aligned} E(\boldsymbol{r}_\perp, F) &= \frac{jk}{2\pi F}\, e^{-2jkF} \int_{-\infty}^{\infty} E_{\rm in}(\boldsymbol{r}'_\perp)\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/F}\, d^2\boldsymbol{r}'_\perp \\ &= \frac{jk}{2\pi F}\, e^{-2jkF} \left[\hat{E}_{\rm in}(\boldsymbol{k}_\perp)\right]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{F}} \end{aligned}} \quad (2F \text{ system}) \tag{20.10.6}$$



The 1-D version of (20.10.5) reads,

$$\boxed{\begin{aligned} E(x, F) &= \sqrt{\frac{jk}{2\pi F}}\, e^{-jk(d+F)}\, e^{-jkx^2(F-d)/2F^2} \left[\hat{E}_{\rm in}(k_x)\right]_{k_x = \frac{kx}{F}} \\ &= \sqrt{\frac{jk}{2\pi F}}\, e^{-jk(d+F)}\, e^{-jkx^2(F-d)/2F^2} \int_{-\infty}^{\infty} E_{\rm in}(x')\, e^{jkxx'/F}\, dx' \end{aligned}} \tag{20.10.7}$$

For a lens with finite pupil $P(\boldsymbol{r}_\perp)$, we may apply the result (20.10.4) to the field $E_a(\boldsymbol{r}_\perp)$ at the left plane of the lens, and then back-propagate $E_a(\boldsymbol{r}_\perp)$ to the input plane at distance d, that is, we have the two-step process,

$$\begin{aligned} E_a(\boldsymbol{u}_\perp) &= \frac{jk\, e^{-jkd}}{2\pi d} \int_{-\infty}^{\infty} e^{-jk|\boldsymbol{u}_\perp - \boldsymbol{r}'_\perp|^2/2d}\, E_{\rm in}(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \\ E(\boldsymbol{r}_\perp, F) &= \frac{jk\, e^{-jkF}}{2\pi F}\, e^{-jk|\boldsymbol{r}_\perp|^2/2F} \int_S E_a(\boldsymbol{u}_\perp)\, P(\boldsymbol{u}_\perp)\, e^{jk\boldsymbol{u}_\perp\cdot\boldsymbol{r}_\perp/F}\, d^2\boldsymbol{u}_\perp \end{aligned}$$

Combining the two and changing variables from $\boldsymbol{u}_\perp$ to $\boldsymbol{v}_\perp = \boldsymbol{u}_\perp - \boldsymbol{r}'_\perp - \frac{d}{F}\boldsymbol{r}_\perp$, we obtain after some algebra,

$$E(\boldsymbol{r}_\perp, F) = \frac{jk\, e^{-jk(d+F)}}{2\pi F}\, e^{-jk|\boldsymbol{r}_\perp|^2(F-d)/2F^2} \int_{-\infty}^{\infty} E_{\rm in}(\boldsymbol{r}'_\perp)\, P_d\left(\boldsymbol{r}'_\perp + \frac{d}{F}\boldsymbol{r}_\perp\right) e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/F}\, d^2\boldsymbol{r}'_\perp \tag{20.10.8}$$

where P_d is a weighted version of P defined as,

$$P_d(\boldsymbol{r}_\perp) = \frac{jk}{2\pi d} \int_{-\infty}^{\infty} e^{-jk|\boldsymbol{v}_\perp|^2/2d}\, P(\boldsymbol{v}_\perp + \boldsymbol{r}_\perp)\, d^2\boldsymbol{v}_\perp \tag{20.10.9}$$

where $P(\boldsymbol{v}_\perp + \boldsymbol{r}_\perp)$ restricts the integration range over $\boldsymbol{v}_\perp$ such that $\boldsymbol{v}_\perp + \boldsymbol{r}_\perp$ lies in the lens aperture. For an infinite lens that has $P(\boldsymbol{r}_\perp) = 1$ for all $\boldsymbol{r}_\perp$, we find, $P_d(\boldsymbol{r}_\perp) = 1$, and we recover (20.10.5). For a finite lens, the following approximation has been used [1422],

$$P_d(\boldsymbol{r}_\perp) = \frac{jk}{2\pi d} \int_{-\infty}^{\infty} e^{-jk|\boldsymbol{v}_\perp|^2/2d}\, P(\boldsymbol{v}_\perp + \boldsymbol{r}_\perp)\, d^2\boldsymbol{v}_\perp \approx P(\boldsymbol{r}_\perp) \tag{20.10.10}$$

which is justified for small d relative to the lens size. In this approximation, referred to as *vignetting*, we have,

$$E(\boldsymbol{r}_\perp, F) = \frac{jk\, e^{-jk(d+F)}}{2\pi F}\, e^{-jk|\boldsymbol{r}_\perp|^2(F-d)/2F^2} \int_{-\infty}^{\infty} E_{\rm in}(\boldsymbol{r}'_\perp)\, P\left(\boldsymbol{r}'_\perp + \frac{d}{F}\boldsymbol{r}_\perp\right) e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/F}\, d^2\boldsymbol{r}'_\perp \tag{20.10.11}$$

In the 1-D case, we have,

$$E(x, F) = \sqrt{\frac{jk}{2\pi F}}\, e^{-jk(d+F)}\, e^{-jkx^2(F-d)/2F^2} \int_{-\infty}^{\infty} E_{\rm in}(x')\, P_d\left(x' + \frac{d}{F}x\right) e^{jkxx'/F}\, dx' \tag{20.10.12}$$

$$P_d(x) = \sqrt{\frac{jk}{2\pi d}} \int_{-\infty}^{\infty} e^{-jkv^2/2d}\, P(v + x)\, dv \tag{20.10.13}$$

One explanation of the approximation (20.10.10) is that for small d, the phase factor $e^{-jk\nu_\perp^2/2d}$ oscillates rapidly, except in the vicinity of $\nu_\perp = 0$, and effectively only the $\nu_\perp = 0$ value of the integrand contributes to the integral (20.10.9). In fact, Eq. (20.10.10) follows from the stationary-phase approximation to this integral. If we apply Eq. (H.6) of Appendix H to the phase function $\phi(\nu_x,\nu_y) = -k(\nu_x^2+\nu_y^2)/2d = -k|\boldsymbol{\nu}_\perp|^2/2d$, which has a stationary point at $\boldsymbol{\nu}_\perp = 0$, and a diagonal second-derivative matrix, $\Phi = \text{diag}(-k/d, -k/d)$, then the stationary-phase approximation to the integral is,

$$\frac{jk}{2\pi d}\int_{-\infty}^{\infty} e^{-jk|\boldsymbol{\nu}_\perp|^2/2d}\, P(\boldsymbol{\nu}_\perp + \boldsymbol{r}_\perp)\, d^2\boldsymbol{\nu}_\perp \approx \frac{jk}{2\pi d}\cdot e^{-j\pi/2}\,\frac{2\pi}{\sqrt{k^2/d^2}}\, P(\boldsymbol{r}_\perp) = P(\boldsymbol{r}_\perp)$$

Example 20.10.1: To check the approximation (20.10.10) numerically, consider the 1-D version of the above, for a uniform pupil $P(x) = u(a-|x|)$,

$$P_d(x) = \sqrt{\frac{jk}{2\pi d}}\int_{-\infty}^{\infty} e^{-jk\nu^2/2d}\, u(a-|\nu+x|)\, d\nu = \sqrt{\frac{jk}{2\pi d}}\int_{-a-x}^{a-x} e^{-jk\nu^2/2d}\, d\nu \quad (20.10.14)$$

This integral can be expressed in terms of the Fresnel integral $\mathcal{F}(x)$ defined in Eq. (F.2) of Appendix F, and evaluated by the MATLAB function **fcs**, that is, after changing integration variables, we have,

$$P_d(x) = \sqrt{\frac{j}{2}}\left[\mathcal{F}\left(\sqrt{\frac{k}{\pi d}}(a-x)\right) - \mathcal{F}\left(\sqrt{\frac{k}{\pi d}}(-a-x)\right)\right] \quad (20.10.15)$$

For the values, $k = 1$, $a = 10$, Fig. 20.10.3 plots $P_d(x)$ for the following values, $d = a/5$, $d = a/10$, $d = a/100$, and $d = a/1000$, over the range $-2a \le x \le 2x$. As d gets smaller, $P_d(x)$ tends to the unit-step pupil $P(x)$. The MATLAB code used to generate these graphs was as follows,

```
k=1; a=10;
x = linspace(-2,2,401);                % x in units of a

for M = [5, 10, 100, 1000]
   d = a/M;

   t1 = sqrt(k/pi/d) * (-x-1) * a;     % limits of Fresnel function
   t2 = sqrt(k/pi/d) * (-x+1) * a;

   Pd = sqrt(j/2) * (fcs(t2) - fcs(t1));

   figure; plot(x,real(Pd),'b-', x,imag(Pd),'r--')
end
```

In the second configuration depicted in Fig. 20.10.2, the propagation response from the left plane of the lens to distance z on the right is as follows, where we also allow for a pupil function for the lens,

$$h(\boldsymbol{r}_\perp,\boldsymbol{r}'_\perp) = h(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp, z)\, e^{jk|\boldsymbol{r}'_\perp|^2/2F} P(\boldsymbol{r}'_\perp) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp-\boldsymbol{r}'_\perp|^2/2z}\, e^{jk|\boldsymbol{r}'_\perp|^2/2F} P(\boldsymbol{r}'_\perp)$$

$$= \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z}\, e^{jk|\boldsymbol{r}'_\perp|^2(z-F)/2Fz}\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/z} P(\boldsymbol{r}'_\perp)$$

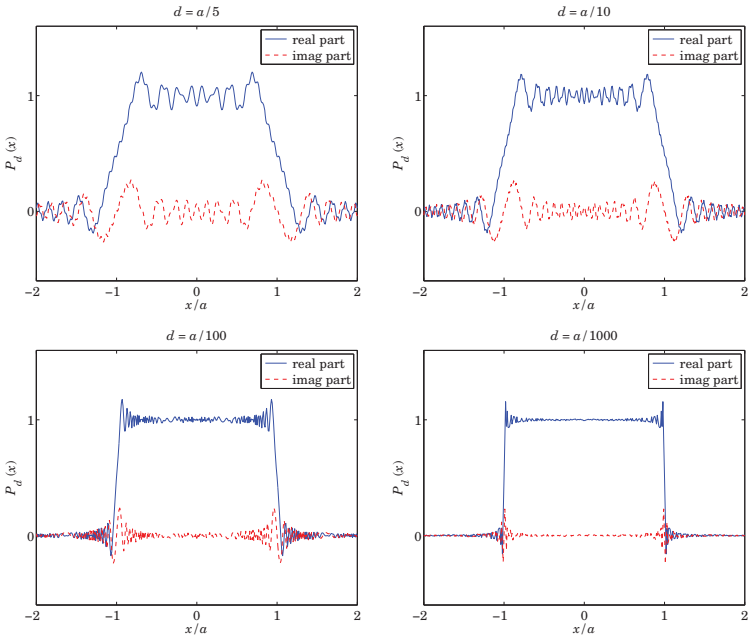


Fig. 20.10.3 Pupil stationary-phase approximation.

so that the output field at distance z is,

$$E(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z}\int_S e^{jk|\boldsymbol{r}'_\perp|^2(z-F)/2Fz}\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/z}\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, P(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \quad (20.10.16)$$

and is proportional to the Fourier transform of, $e^{jk|\boldsymbol{r}'_\perp|^2(z-F)/2Fz}\, E_{\text{in}}(\boldsymbol{r}'_\perp)P(\boldsymbol{r}'_\perp)$. In the circularly symmetric and 1-D cases, (20.10.16) reduces to,

$$E(r_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jkr_\perp^2/2z}\int_0^{\infty} e^{jkr'^2_\perp(z-F)/2Fz}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) E_{\text{in}}(r'_\perp)\, P(r'_\perp)\, 2\pi r'_\perp dr'_\perp$$

$$E(x, z) = \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz}\, e^{-jkx^2/2z}\int_{-\infty}^{\infty} e^{jkx'^2(z-F)/2Fz}\, e^{jkxx'/z}\, E_{\text{in}}(x')\, P(x')\, dx' \quad (20.10.17)$$

For distances near the focus, we may set, $z = F + \delta z$, and retain δz only in the quadratic exponent, resulting in the following approximation that is used to quantify axial defocusing [638],

$$E(\boldsymbol{r}_\perp, F+\delta z) = \frac{jk}{2\pi F}\, e^{-jkF} e^{-jk|\boldsymbol{r}_\perp|^2/2F}\int_{-\infty}^{\infty} e^{jk|\boldsymbol{r}'_\perp|^2\delta z/2F^2}\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/F}\, E_{\text{in}}(\boldsymbol{r}'_\perp)\, P(\boldsymbol{r}'_\perp)\, d^2\boldsymbol{r}'_\perp \quad (20.10.18)$$

Summary

The two-dimensional Fourier transform of the input field appears in the output field in several optical contexts (in some cases, the input may be modulated by a quadratic phase factor, such as $e^{-jk|\boldsymbol{r}_\perp|^2/2z}$).

a. at the focal plane of a lens under the Fresnel approximation, Eqs. (20.10.2)
b. at the focal plane in a $2F$-system, Eqs. (20.10.5) and (20.10.6)
c. at a large distance z under the Fraunhofer approximation, Eq. (20.3.1)
d. at large radial distance r under the far-field approximation, Eq. (19.3.3)
e. at a large distance z under the Fresnel approximation, Eq. (20.1.9)
f. near the focal point of a lens in an axially defocused system, Eq. (20.10.18)
g. the PSF of a lens system is the 2-D Fourier transform of the exit pupil, Eq. (20.7.4)

Moreover as we see below, the 2-D Fourier transform of the input becomes an essential part of the so-called $4F$ imaging system in which a finite aperture at the Fourier plane causes the spatial filtering of the input, thus decreasing the resolution, but also allowing more specialized filtering, such as deblurring.

20.11 4F Optical Processor

A $4F$ imaging system consists of two $2F$ systems in cascade as shown in Fig. 20.11.1. Assuming large lens pupils and ignoring vignetting effects, the first $2F$ system will produce the 2-D Fourier transform of the input at the middle focal plane (the Fourier plane). After multiplication by an appropriate spatial filter with transfer function, say, $H(\boldsymbol{k}_\perp)$, the second $2F$ system produces the Fourier transform of the product at the output plane.

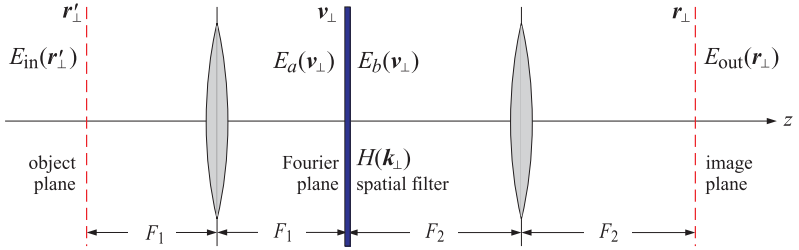


Fig. 20.11.1 4F optical processor.

But the Fourier transform of a product of Fourier transforms becomes a convolution in the space domain, thus, the operation of the $4F$ system can be described symbolically as follows, where the hat denotes Fourier transformation and the dot and asterisk denote ordinary multiplication and convolution, respectively.

$$E_{\text{in}} \quad \Rightarrow \quad H \cdot \hat{E}_{\text{in}} \quad \Rightarrow \quad E_{\text{out}} = \widehat{H \cdot \hat{E}_{\text{in}}} = \hat{H} * E_{\text{in}} \tag{20.11.1}$$

In absence of any processing at the Fourier plane (that is, $H \equiv 1$), the second $2F$ system will produce the original input up to a reversal and magnification. This is the expected result on the basis of geometrical optics, as shown in Fig. 20.11.2. This follows essentially from the property that the Fourier transform of a Fourier transform is the original field reversed. Thus, we expect in this case the same geometrical-optics result as in Eq. (20.6.6), up to a $4F$ propagation delay,

$$E_{\text{out}}(\boldsymbol{r}_\perp) = -\frac{F_1}{F_2}\, E_{\text{in}}\left(-\frac{F_1}{F_2}\,\boldsymbol{r}_\perp\right) \quad \Leftrightarrow \quad -\frac{F_2}{F_1}\, E_{\text{out}}\left(-\frac{F_2}{F_1}\,\boldsymbol{r}_\perp\right) = E_{\text{in}}(\boldsymbol{r}_\perp) \tag{20.11.2}$$

where the first represents the geometric-optics projection of the object onto the image plane, and the second, the projection of the image back onto the object plane.

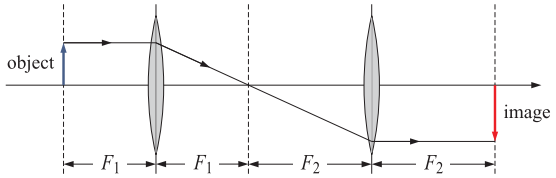


Fig. 20.11.2 Geometric image in 4F optical processor.

To derive the operations of Eq. (20.11.1) in some detail, let us denote the transverse coordinates on the Fourier plane by $\boldsymbol{v}_\perp$ and assume that the filter shown in Fig. 20.11.1 has a given transmittance function, say, $F(\boldsymbol{v}_\perp)$, so that the fields across the plane will be related by, $E_b(\boldsymbol{v}_\perp) = F(\boldsymbol{v}_\perp)E_a(\boldsymbol{v}_\perp)$. Applying Eq. (20.10.6) in succession then gives,

$$\begin{aligned} E_a(\boldsymbol{v}_\perp) &= G_1\, e^{-2jkF_1} \int_{-\infty}^{\infty} E_{\text{in}}(\boldsymbol{r}'_\perp)\, e^{jk\boldsymbol{v}_\perp\cdot\boldsymbol{r}'_\perp/F_1}\, d^2\boldsymbol{r}'_\perp \\ E_b(\boldsymbol{v}_\perp) &= F(\boldsymbol{v}_\perp)E_a(\boldsymbol{v}_\perp) \\ E_{\text{out}}(\boldsymbol{r}_\perp) &= G_2\, e^{-2jkF_2} \int_{-\infty}^{\infty} E_b(\boldsymbol{v}_\perp)\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{v}_\perp/F_2}\, d^2\boldsymbol{v}_\perp \end{aligned} \tag{20.11.3}$$

where, for convenience, we defined,

$$G_1 = \frac{jk}{2\pi F_1}, \quad G_2 = \frac{jk}{2\pi F_2} \tag{20.11.4}$$

Remapping the output to the object plane by the geometric-optics projection of Eq. (20.11.2), we have,

$$\begin{aligned} E_a(\boldsymbol{v}_\perp) &= G_1\, e^{-2jkF_1} \int_{-\infty}^{\infty} E_{\text{in}}(\boldsymbol{r}'_\perp)\, e^{jk\boldsymbol{v}_\perp\cdot\boldsymbol{r}'_\perp/F_1}\, d^2\boldsymbol{r}'_\perp \\ E_{g,\text{out}}(\boldsymbol{r}_\perp) \equiv -\frac{F_2}{F_1}\, E_{\text{out}}\left(-\frac{F_2}{F_1}\,\boldsymbol{r}_\perp\right) &= -G_1\, e^{-2jkF_2} \int_{-\infty}^{\infty} F(\boldsymbol{v}_\perp)E_a(\boldsymbol{v}_\perp)\, e^{-jk\boldsymbol{r}_\perp\cdot\boldsymbol{v}_\perp/F_1}\, d^2\boldsymbol{v}_\perp \end{aligned} \tag{20.11.5}$$

In order to write the first equation as a Fourier transform and the second as an inverse Fourier transform, let us redefine the functions $F(\boldsymbol{v}_\perp)$ and $E_a(\boldsymbol{v}_\perp)$ as functions

of the wavenumber, $\boldsymbol{k}_\perp = k\boldsymbol{\nu}_\perp/F_1$, or, $\boldsymbol{\nu}_\perp = F_1\boldsymbol{k}_\perp/k$,

$$H(\boldsymbol{k}_\perp) = F(\boldsymbol{\nu}_\perp)\Big|_{\boldsymbol{\nu}_\perp=\frac{\boldsymbol{k}_\perp F_1}{k}}\,, \qquad E(\boldsymbol{k}_\perp) = \frac{1}{G_1}\,E_a(\boldsymbol{\nu}_\perp)\Big|_{\boldsymbol{\nu}_\perp=\frac{\boldsymbol{k}_\perp F_1}{k}} \tag{20.11.6}$$

where we also rescaled E_a by the gain factor G_1. Then, after changing integration variables from $\boldsymbol{\nu}_\perp$ to $\boldsymbol{k}_\perp$, we obtain,

$$\boxed{\begin{aligned} E(\boldsymbol{k}_\perp) &= e^{-2jkF_1}\int_{-\infty}^{\infty} E_{\text{in}}(\boldsymbol{r}'_\perp)\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}'_\perp}\, d^2\boldsymbol{r}'_\perp = e^{-2jkF_1}\,\hat{E}_{\text{in}}(\boldsymbol{k}_\perp) \\ E_{g,\text{out}}(\boldsymbol{r}_\perp) &= e^{-2jkF_2}\int_{-\infty}^{\infty} H(\boldsymbol{k}_\perp)E(\boldsymbol{k}_\perp)\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \end{aligned}} \tag{20.11.7}$$

These represent the actions of the two $2F$ stages of the $4F$ system. Inserting the first equation into the second, the result may be written as a space-domain convolution that represents the overall input/output relationship,

$$\boxed{\begin{aligned} E_{g,\text{out}}(\boldsymbol{r}_\perp) &= e^{-2jk(F_1+F_2)}\int_{-\infty}^{\infty} H(\boldsymbol{k}_\perp)\,\hat{E}_{\text{in}}(\boldsymbol{k}_\perp)\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \\ &= e^{-2jk(F_1+F_2)}\int_{-\infty}^{\infty} h(\boldsymbol{r}_\perp-\boldsymbol{r}'_\perp)\,E_{\text{in}}(\boldsymbol{r}'_\perp)\,d^2\boldsymbol{r}'_\perp \end{aligned}} \tag{20.11.8}$$

where $h(\boldsymbol{r}_\perp)$ is the impulse response of the filter, i.e., the inverse Fourier transform,

$$\boxed{h(\boldsymbol{r}_\perp) = \int_{-\infty}^{\infty} H(\boldsymbol{k}_\perp)\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2}} \tag{20.11.9}$$

Thus, the $4F$ system acts as a *linear shift-invariant spatial filter* with a prescribed transfer function $H(\boldsymbol{k}_\perp)$ or impulse response $h(\boldsymbol{r}_\perp)$. We note that $H(\boldsymbol{k}_\perp)$ of Eq. (20.11.6) is essentially the same as that of Eq. (20.9.7) if we think of $F(\boldsymbol{\nu}_\perp)$ as a pupil function.

In the one-dimensional case, the analogous relationship to the geometrical optics result (20.11.2) is according to Eq. (20.9.9),

$$E_{\text{out}}(x) = e^{j\pi/2}\sqrt{\frac{F_1}{F_2}}\,E_{\text{in}}\left(-\frac{F_1}{F_2}\,x\right) \quad\Leftrightarrow\quad e^{-j\pi/2}\sqrt{\frac{F_2}{F_1}}\,E_{\text{out}}\left(-\frac{F_2}{F_1}\,x\right) = E_{\text{in}}(x) \tag{20.11.10}$$

Using Eq. (20.10.7) with $d = F$, the 1-D versions of Eqs. (20.11.5) will be then,

$$\begin{aligned} E_a(\nu) &= G_1\, e^{-2jkF_1}\int_{-\infty}^{\infty} E_{\text{in}}(x')\, e^{jk\nu x'/F_1}\,dx' \\ E_{\text{out}}(x) &= G_2\, e^{-2jkF_2}\int_{-\infty}^{\infty} F(\nu)E_a(\nu)\, e^{jkx\nu/F_2}\,d\nu \\ E_{g,\text{out}}(x) &\equiv e^{-j\pi/2}\sqrt{\frac{F_2}{F_1}}\,E_{\text{out}}\left(-\frac{F_2}{F_1}\,x\right) \\ &= e^{-j\pi/2}\,G_1\, e^{-2jkF_2}\int_{-\infty}^{\infty} F(\nu)E_a(\nu)\, e^{-jkx\nu/F_1}\,d\nu \end{aligned} \tag{20.11.11}$$

where now we have the definitions,

$$G_1 = e^{j\pi/4}\sqrt{\frac{k}{2\pi F_1}}\,, \quad G_2 = e^{j\pi/4}\sqrt{\frac{k}{2\pi F_2}} \tag{20.11.12}$$

We may redefine again $F(\nu)$ and $E_a(\nu)$ as functions of the wavenumber, $k_x = k\nu/F_1$,

$$H(k_x) = F(\nu)\Big|_{\nu=\frac{k_x F_1}{k}}\,, \qquad E(k_x) = \frac{1}{G_1}\,E_a(\nu)\Big|_{\nu=\frac{k_x F_1}{k}} \tag{20.11.13}$$

Then, after changing integration variables from ν to k_x, the two-stage operations of the $4F$ system will read,

$$\boxed{\begin{aligned} E(k_x) &= e^{-2jkF_1}\int_{-\infty}^{\infty} E_{\text{in}}(x')\, e^{jk_x x'}\,dx' = e^{-2jkF_1}\,\hat{E}_{\text{in}}(k_x) \\ E_{g,\text{out}}(x) &= e^{-2jkF_2}\int_{-\infty}^{\infty} H(k_x)\,E(k_x)\, e^{-jk_x x}\,\frac{dk_x}{2\pi} \end{aligned}} \tag{20.11.14}$$

which combine into the single filtering operation,

$$\boxed{\begin{aligned} E_{g,\text{out}}(x) &= e^{-2jk(F_1+F_2)}\int_{-\infty}^{\infty} H(k_x)\,\hat{E}_{\text{in}}(k_x)\, e^{-jk_x x}\,\frac{dk_x}{2\pi} \\ &= e^{-2jk(F_1+F_2)}\int_{-\infty}^{\infty} h(x-x')\,E_{\text{in}}(x')\,dx' \end{aligned}} \tag{20.11.15}$$

where $h(x)$ is the 1-D impulse response of the filter $H(k_x) = F(\nu)$,

$$\boxed{h(x) = \int_{-\infty}^{\infty} H(k_x)\, e^{-jk_x x}\,\frac{dk_x}{2\pi}} \tag{20.11.16}$$

Fig. 20.11.3 depicts three examples: (a) a lowpass spatial filter consisting simply of a finite aperture that lets through only low frequencies, (b) a highpass filter consisting of an aperture stop that blocks low frequencies, and (c) a cascaded example of two $4F$ systems of transfer functions H_1, H_2, which acts as the equivalent filter, $H_1(\boldsymbol{k}_\perp)\cdot H_2(\boldsymbol{k}_\perp)$, indeed, the essential operations at each of the outputs of the five image planes are as follows, denoting the inverse Fourier transforms by, $h_1 = \hat{H}_1$ and $h_2 = \hat{H}_2$,

$$E_{\text{in}} \quad\Rightarrow\quad H_1\cdot\hat{E}_{\text{in}} \quad\Rightarrow\quad h_1 * E_{\text{in}} \quad\Rightarrow\quad H_2\cdot H_1\cdot\hat{E}_{\text{in}} \quad\Rightarrow\quad (h_2 * h_1) * E_{\text{in}}$$

Example 20.11.1: For the lowpass filter shown in Fig. 20.11.3, we assume a circular aperture of radius b. Then, the filter's transmittance function will be defined in terms of the unit-step function, $F(\boldsymbol{\nu}_\perp) = u(b - |\boldsymbol{\nu}_\perp|)$. The filter's transfer function will be,

$$H(\boldsymbol{k}_\perp) = F(\boldsymbol{\nu}_\perp)\Big|_{\boldsymbol{\nu}_\perp=\frac{\boldsymbol{k}_\perp F_1}{k}} = u\left(b - \frac{|\boldsymbol{k}_\perp|F_1}{k}\right) = u(k_{\max} - |\boldsymbol{k}_\perp|)\,, \quad k_{\max} = k\,\frac{b}{F_1}$$

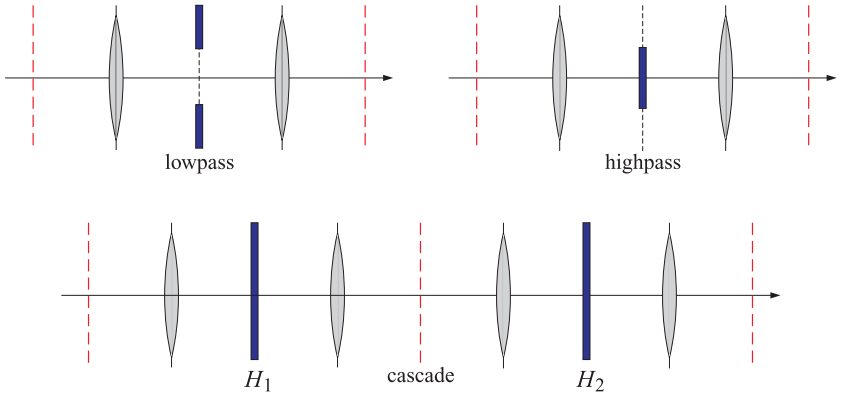


Fig. 20.11.3 Lowpass, highpass, and cascaded filter examples.

Thus the filter acts as a bandlimiting filter, cutting off all higher spatial frequencies that are above $k_{\max}$ and letting through only those for which, $|\boldsymbol{k}_\perp| \le k_{\max}$. The corresponding impulse response has the usual $2J_1(x)/x$ form,

$$\begin{aligned} h(\boldsymbol{r}_\perp) &= \int_{-\infty}^{\infty} H(\boldsymbol{k}_\perp)\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} = \int_{|\boldsymbol{k}_\perp|\le k_{\max}} e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \\ &= \int_0^{k_{\max}} J_0(k_\perp r_\perp)\, \frac{2\pi k_\perp dk_\perp}{(2\pi)^2} = k_{\max}\, \frac{J_1(k_{\max} r_\perp)}{2\pi r_\perp} \\ &= |G_1|^2 \cdot \pi b^2 \cdot \frac{2J_1(k_{\max} r_\perp)}{k_{\max} r_\perp} \end{aligned} \tag{20.11.17}$$

where we set, $k_\perp = |\boldsymbol{k}_\perp|$ and $r_\perp = |\boldsymbol{r}_\perp|$, and G_1 was defined in (20.11.4).

For the one-dimensional case, we assume a slit opening with support $-b \le \nu \le b$, defined in terms of the unit-step, $F(\nu) = u(b - |\nu|)$. Then, the filter's transfer function will be,

$$H(k_x) = u\left(b - \frac{|k_x| F_1}{k}\right) = u(k_{\max} - |k_x|)\,, \quad k_{\max} = k\,\frac{b}{F_1}$$

Its impulse response has the usual $\sin x/x$ form,

$$\begin{aligned} h(x) &= \int_{-\infty}^{\infty} H(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = \int_{-k_{\max}}^{k_{\max}} e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= \frac{\sin(k_{\max} x)}{\pi x} = |G_1|^2 \cdot 2b \cdot \frac{\sin(k_{\max} x)}{k_{\max} x} \end{aligned} \tag{20.11.18}$$

where G_1 is defined in Eq. (20.11.12). Using (20.11.14), the two-stage filtering operations will have the explicit forms,

$$\begin{aligned} E(k_x) &= e^{-2jkF_1} \int_{-\infty}^{\infty} E_{\text{in}}(x')\, e^{jk_x x'}\, dx' = e^{-2jkF_1}\, \hat{E}_{\text{in}}(k_x) \\ E_{g,\text{out}}(x) &= e^{-2jkF_2} \int_{-\infty}^{\infty} u(k_{\max} - |k_x|) E(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= e^{-2jkF_2} \int_{-k_{\max}}^{k_{\max}} E(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= e^{-2jk(F_1+F_2)} \int_{-k_{\max}}^{k_{\max}} \hat{E}_{\text{in}}(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= e^{-2jk(F_1+F_2)} \int_{-\infty}^{\infty} \frac{\sin(k_{\max}(x-x'))}{\pi(x-x')}\, E_{\text{in}}(x')\, dx' \end{aligned} \tag{20.11.19}$$

It is evident that the output, $E_{g,\text{out}}(x)$, is bandlimited to the range, $|k_x| \le k_{\max}$, with a bandlimited Fourier transform,

$$\hat{E}_{g,\text{out}}(k_x) = e^{-2jk(F_1+F_2)}\, \hat{E}_{\text{in}}(k_x) \cdot u(k_{\max} - |k_x|) \tag{20.11.20}$$

Fig. 20.11.4 plots the normalized impulse responses, that is, the functions,

$$\begin{aligned} &\text{(2-D case)} \qquad h(u) = \frac{2J_1(\pi u)}{\pi u}\,, \qquad \pi u = k_{\max} r_\perp \\ &\text{(1-D case)} \qquad h(u) = \frac{\sin(\pi u)}{\pi u}\,, \qquad \pi u = k_{\max} x \end{aligned}$$

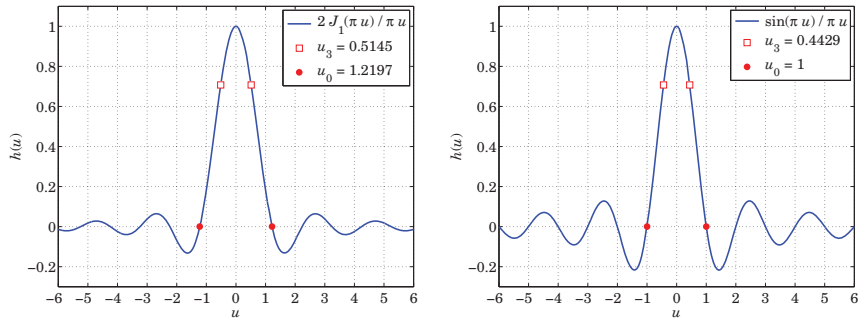


Fig. 20.11.4 2-D and 1-D lowpass inpulse responses.

The first nulls and 3dB points are also shown. The first nulls correspond to the Rayleigh and Abbe resolution criteria that give the smallest resolvable transverse separations, i.e.,

$$\begin{aligned} k_{\max} r_{\min} = \pi u_0 = 1.2197\pi \quad &\Rightarrow \quad r_{\min} = 1.2197\,\frac{\lambda}{2N_A} \qquad \text{(2-D case)} \\ k_{\max} x_{\min} = \pi u_0 = \pi \quad &\Rightarrow \quad x_{\min} = \frac{\lambda}{2N_A} \qquad \text{(1-D case)} \end{aligned} \tag{20.11.21}$$

where in both cases, $D = 2b$ and $N_A = b/F_1 = D/2F_1$, The same criteria can also be expressed in terms of the smallest angular separations, $\theta_{\min} = r_{\min}/F_1$ or $\theta_{\min} = x_{\min}/F_1$, assuming $F_1 \gg \lambda$,

$$\theta_{\min} = \frac{r_{\min}}{F_1} = 1.2197\,\frac{\lambda}{D} = 70^{\circ}\,\frac{\lambda}{D}\,, \qquad \theta_{\min} = \frac{x_{\min}}{F_1} = \frac{\lambda}{D} = 57^{\circ}\,\frac{\lambda}{D} \tag{20.11.22}$$

Example 20.11.2: *Lyot coronagraph.*[†] The purpose of a coronagraph is to block the intense light from the sun when a telescope is pointed towards it, allowing a nearby object to be seen that lies at a small angular distance. It is an example of a $6F$ system consisting of a highpass blocking filter $H(\boldsymbol{u}_\perp)$, called the Lyot mask, at the first Fourier plane ($\boldsymbol{u}_\perp$ plane), followed by a lowpass filter $L(\boldsymbol{w}_\perp)$, called the Lyot stop, at the second Fourier plane ($\boldsymbol{w}_\perp$ plane), which is simply an aperture opening, as shown in Fig. 20.11.5.

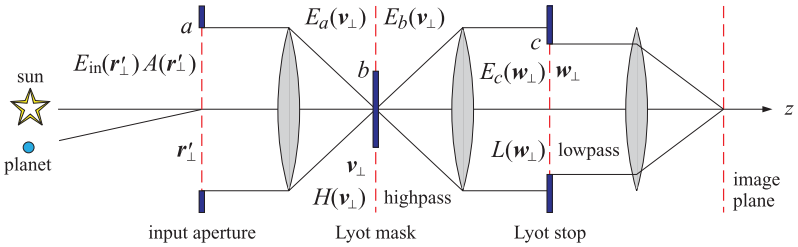


Fig. 20.11.5 Lyot coronagraph.

The Lyot mask blocks most of the sunlight, but similar to eclipses of the sun, enough light still diffracts from the edges of the mask. The purpose of the second Lyot stop is to remove that diffracted light. In this example, we will assume that the input pupil of the telescope has radius a, the Lyot mask has radius $b < a$, and the Lyot stop, radius $c \lessapprox a$. Thus, the transmittance functions at the input and v, w planes are in terms of the unit-step function,

$$\begin{aligned} A(\boldsymbol{r}'_\perp) &= u(a - |\boldsymbol{r}'_\perp|) \\ H(\boldsymbol{v}_\perp) &= u(|\boldsymbol{v}_\perp| - b) \\ L(\boldsymbol{w}_\perp) &= u(c - |\boldsymbol{w}_\perp|) \end{aligned}$$

The fields at the outputs of the three planes will be symbolically,

$$A \cdot E_{\rm in} \quad \Rightarrow \quad H \cdot (\hat{A} * \hat{E}_{\rm in}) \quad \Rightarrow \quad L \cdot [\hat{H} * (A \cdot E_{\rm in})\,]$$

The input field $E_{\rm in}$ from the sun may be assumed to be uniformly filling the entire input aperture, so that at the output of the input aperture we may set, $E_{\rm in}(\boldsymbol{r}'_\perp) = u(a - |\boldsymbol{r}'_\perp|)$.

For simplicity, we will work with the 1-D version and assume that all lenses have the same focal length, F. Denoting by $E_a(v)$ the field at the left input of the Lyot mask, and by, $E_b(v) = H(v)E_a(v)$, the field at its output, and denoting the propagation gain by $G_0 = e^{-2jkF}e^{j\pi/4}\sqrt{k/2\pi F}$, we have applying (20.11.11),

[†] http://www.lyot.org
http://www. lyot.org/background/coronagraphy.html

$$\begin{aligned} E_a(v) &= G_0 \int_{-\infty}^{\infty} E_{\rm in}(x')\,e^{jkvx'/F}\,dx' \\ &= G_0 \int_{-\infty}^{\infty} u(a - |x'|)\,e^{jkvx'/F}\,dx' = G_0 \int_{-a}^{a} e^{jkvx'/F}\,dx' \\ &= G_0 \cdot 2a \cdot \frac{\sin(kav/F)}{kav/F} \\ E_b(v) &= H(v)E_a(v) = G_0 \cdot 2a \cdot \frac{\sin(kav/F)}{kav/F} \cdot u(|v| - b) \end{aligned}$$

Then, the field at the left input of the Lyot spot at the w-plane, will be,

$$\begin{aligned} E_c(w) &= G_0 \int_{-\infty}^{\infty} E_b(v)\,e^{jkwv/F}\,du = G_0^2 \int_{|v|\geq b} 2a \cdot \frac{\sin(kav/F)}{kav/F}\,e^{jkwv/F}\,dv \\ &= G_0^2 \left[\int_{-\infty}^{\infty} - \int_{-b}^{b}\right] 2a \cdot \frac{\sin(kav/F)}{kav/F}\,e^{jkwv/F}\,dv \\ &= G_0^2\,\frac{2\pi F}{k}\left[u(a - |w|) - \frac{1}{\pi}\int_0^b \frac{2\sin(kav/F)\cos(kwv/F)}{v}\,dv\right] \\ &= G_0^2\,\frac{2\pi F}{k}\left[u(a - |w|) - \frac{1}{\pi}\int_0^b \frac{\sin(ku(a+w)/F) + \sin(ku(a-w)/F)}{u}\,du\right] \\ &= G_0^2\,\frac{2\pi F}{k}\left[u(a - |w|) - \frac{1}{\pi}\,{\rm Si}\left(\frac{kb(a+w)}{F}\right) - \frac{1}{\pi}\,{\rm Si}\left(\frac{kb(a-w)}{F}\right)\right] \end{aligned}$$

where ${\rm Si}(z)$ is the sine-integral function defined in Eq. (G.2) of Appendix G, and computed by our MATLAB function, **Si**, or, the built-in function **sinint**,

$${\rm Si}(z) = \int_0^z \frac{\sin t}{t}\,dt$$

We note that the overall gain factor is simply a phase, $G_0^2\,2\pi F/k = e^{-j\pi/2}e^{-4jkF}$, and ignoring it, we may write,

$$E_c(w) = u(a - |w|) - \frac{1}{\pi}\left[{\rm Si}\left(\frac{kb(a+w)}{F}\right) + {\rm Si}\left(\frac{kb(a-w)}{F}\right)\right] \tag{20.11.23}$$

This represents the diffracted light from the tails of the main lobe of $E_a(v)$, and is concentrated near the edges of the entrance pupil. The field at the output of the Lyot spot will be then, $E_c(w)L(w) = E_c(w)\cdot u(c - |w|)$, whose job is to further remove such edge contributions.

Fig. 20.11.6 plots the fields $E_a(v), E_b(v), E_c(w)$ and $L(w)E_c(w)$. The parameter values, chosen only for display convenience, were,

$$k = 1\,, \quad a = 20\,, \quad F = 20\,, \quad b = 0.5a = 10\,, \quad c = 0.85a = 17$$

The scales of the top two and bottom two graphs are not the same—the field $E_a(v)$ was normalized to its maximum $2a = 40$, so that the real scale of the bottom two graphs should be $1/40$ of what is shown. The MATLAB code for generating these graphs was,

```
k=1; a=20; F=20; b=0.5*a; c = 0.85*a;

v = linspace(-2,2,1001)*a;
w = v;

Ea = sinc(k*a*v/F/pi);      % normalized to unity at u=0

H = (abs(v)>=b);            % Lyot mask

Eb = H.*Ea;                 % output of Lyot mask

Ec = (abs(w)<a) - (Si(k*b*(a+w)/F) + Si(k*b*(a-w)/F))/pi;

L = (abs(w)<=c);            % Lyot spot

Ed = L.*Ec;                 % output of Lyot spot

figure; plot(v/a,abs(Ea),'b-', v/a,H,'r--');
figure; plot(v/a,abs(Eb),'b-', v/a,H,'r--');
figure; plot(w/a,abs(Ec),'b-', w/a,L,'r--');
figure; plot(w/a,abs(Ed),'b-', w/a,L,'r--');
```

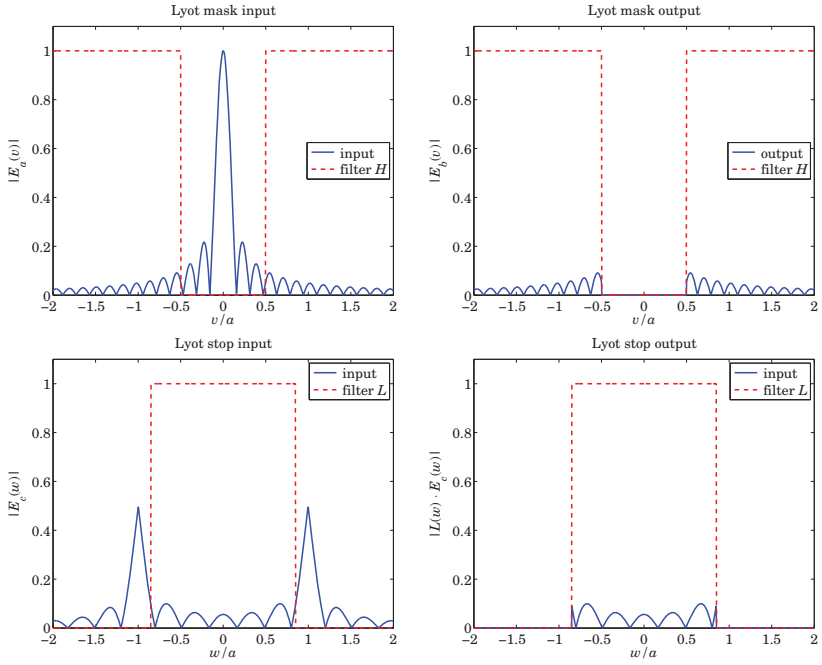


Fig. 20.11.6 Coronagraph operations.

20.12 Apodization Design and Aperture Synthesis

We have seen that at far distances from an input plane under the Fraunhofer approximation, or, under the Fresnel approximation in a $2F$ system such as the one shown in Fig. 20.12.1, the radiated field is proportional to the spatial Fourier transform of the input. For example, in the one-dimensional case, we have,

$$\begin{aligned} E(v,z) &= \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz}\, e^{-jkv^2/2z}\, \Big[\hat{E}_{\text{in}}(k_x)\Big]_{k_x=\frac{kv}{z}} \qquad \text{(Fraunhofer)} \\ E(v,F) &= \sqrt{\frac{jk}{2\pi F}}\, e^{-2jkF}\, \Big[\hat{E}_{\text{in}}(k_x)\Big]_{k_x=\frac{kv}{F}} \qquad (2F \text{ system}) \end{aligned} \tag{20.12.1}$$

where v is the transverse coordinate at the Fourier plane, with the Fourier transform defined as follows, where x is the transverse coordinate on the input plane,

$$\hat{E}_{\text{in}}(k_x) = \int_{-\infty}^{\infty} E_{\text{in}}(x)\, e^{jk_x x}\, dx \tag{20.12.2}$$

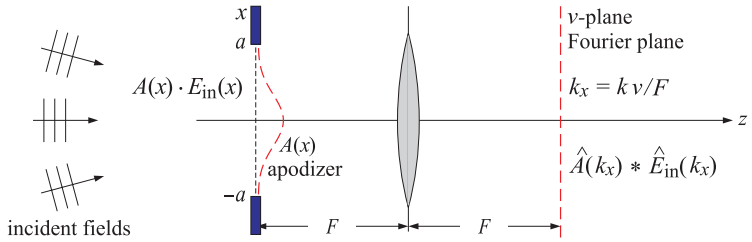
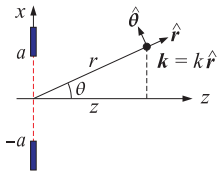


Fig. 20.12.1 Apodization or windowing.

Similarly, in the far-field approximation of Eq. (19.4.9), the radiation field in the direction θ at large radial distances r from the input plane is,



$$E(r,\theta) \approx e^{j\frac{\pi}{4}} \sqrt{\frac{k}{2\pi r}}\, e^{-jkr}\, \cos\theta\, \Big[\hat{E}_{\text{in}}(k_x)\Big]_{k_x = k\sin\theta} \tag{20.12.3}$$

where $\boldsymbol{k}$ is in the radial direction, that is, $\boldsymbol{k} = k\hat{\mathbf{r}} = \hat{\mathbf{x}}k_x + \hat{\mathbf{z}}k_z = \hat{\mathbf{x}}k\sin\theta + \hat{\mathbf{z}}k\cos\theta$.

The effect of a finite aperture at the input plane is to multiply the incident field by a transmittance function, known as an *apodizing* or *windowing* function, say, $A(x)$, which in its simplest form is the characteristic or indicator function of the aperture. The resulting field, $A(x)\cdot E_{\text{in}}(x)$, then propagates to the Fourier plane where its Fourier transform becomes the convolution of the transforms of the two factors,

$$A(x)\cdot E_{\text{in}}(x) \quad \Rightarrow \quad \widehat{A\cdot E_{\text{in}}}(k_x) = (\hat{A} * \hat{E}_{\text{in}})(k_x) = \int_{-\infty}^{\infty} \hat{A}(k_x - k'_x)\hat{E}_{\text{in}}(k'_x)\,\frac{dk'_x}{2\pi} \tag{20.12.4}$$

The convolution operation causes the smearing of $\hat{E}_{\text{in}}(k_x)$ and may result in loss of resolution and the inability to detect weak sources. For example, suppose the input consists of two incident plane waves arriving at the $z = 0$ plane at angles θ_1, θ_2, and with relative strengths, E_1, E_2. Then, the input, its Fourier transform, and the smeared output at the Fourier plane will be as follows, where $k_{x1} = k \sin\theta_1$ and $k_{x2} = k \sin\theta_2$,

$$\begin{aligned} E_{\text{in}}(x) &= E_1 e^{-jk_{x1}x} + E_2 e^{-jk_{x2}x} \\ \hat{E}_{\text{in}}(k_x) &= 2\pi E_1 \delta(k_x - k_{x1}) + 2\pi E_2 \delta(k_x - k_{x2}) \\ \hat{A}(k_x) * \hat{E}_{\text{in}}(k_x) &= E_1 \hat{A}(k_x - k_{x1}) + E_2 \hat{A}(k_x - k_{x2}) \end{aligned}$$

A typical $\hat{A}(k_x)$ has a central peak (main lobe) at $k_x = 0$ of a certain width, surrounded by sidelobes of a certain height, as shown on the left in Fig. 20.12.2. The two sharp (delta-function) spectral lines at k_{x1}, k_{x2} of the above field $\hat{E}_{\text{in}}(k_x)$ are smeared and replaced at the Fourier plane by shifted copies of $\hat{A}(k_x)$ centered at k_{x1}, k_{x2}, as depicted on the right in Fig. 20.12.2.

If the peaks are too close to each other, they will overlap and tend to appear as a single peak. Their resolvability requires that their separation be greater than about one mainlobe width, that is, $|k_{x_2} - k_{x_1}| \gtrsim \Delta k$. This translates to the smallest angular separation that can be resolved (near the forward direction), $|\theta_2 - \theta_1| \gtrsim \Delta k/k$. But, even when they do meet this condition, if one of the signals is too weak, it may be lost under the sea of sidelobes from the stronger one. For example, in the search for planets of distant suns, the expected light intensity from the planet is about 100 dB below that of the sun, i.e., 10^{-10} times weaker in power, or 10^{-5} times in amplitude. Therefore, to be able to "see" such planets, the sidelobes must be down by at least 100 dB.

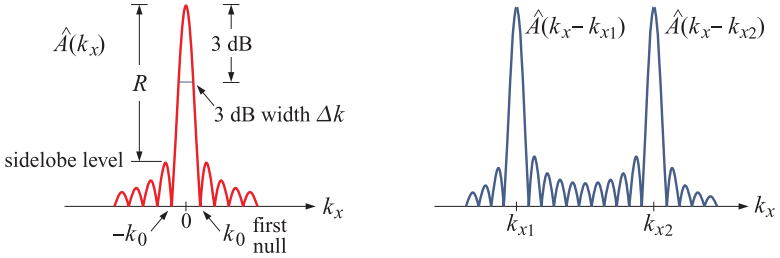


Fig. 20.12.2 Basic parameters of an apodizing function.

Thus, the important parameters of any apodizing function that determine its resolution and detectability properties are its mainlobe width and its sidelobe level. The ideal apodizing window $A(x)$ will have the narrowest mainlobe and the smallest sidelobes.[†]

The mainlobe width, Δk, can be measured either at the 3-dB level (half-power level), or, as half of the mainlobe base measured from null to null.

The Rayleigh resolution criterion takes the first null as the minimum resolvable separation. If we denote that null by k_0 and the 3-dB wavenumber by k_3, i.e., the points

[†]The word "apodize" comes from the greek word for "without feet", i.e., no sidelobes.

at which $\hat{A}(k_0) = 0$ and, $|\hat{A}(k_3)/\hat{A}(0)|^2 = 1/2$, then, the 3-dB width (full-width at half-maximum) will be $2k_3$. The two possible choices for the mainlobe width will be $\Delta k = k_0$ and $\Delta k = 2k_3$, leading to the resolution criteria,

$$\begin{aligned} |k_{x_2} - k_{x_1}| &\gtrsim k_0 \qquad \text{(Rayleigh criterion)} \\ |k_{x_2} - k_{x_1}| &\gtrsim 2k_3 \qquad \text{(3-dB criterion)} \end{aligned} \tag{20.12.5}$$

For example, a uniform rectangular aperture, $A(x) = u(a - |x|)$, will have a sinc-function transform,

$$A(x) = u(a - |x|) \quad \Rightarrow \quad \hat{A}(k_x) = \int_{-a}^{a} A(x) e^{jk_x x}\, dx = 2a \cdot \frac{\sin(k_x a)}{k_x a} \tag{20.12.6}$$

Its first-null and 3-dB wavenumbers are,

$$k_0 = \frac{\pi}{a}, \quad k_3 = \frac{0.443\pi}{a}, \quad 2k_3 = \frac{0.886\pi}{a} \tag{20.12.7}$$

The corresponding Rayleigh and 3-dB criteria expressed in terms of the smallest resolvable wavenumber separation Δk, or the smallest angular separation $\Delta\theta = \Delta k/k$, will be as follows, where we replaced, $k = 2\pi/\lambda$,

$$\begin{aligned} \text{(Rayleigh)} \qquad & \Delta k = \frac{\pi}{a}, \quad \Delta\theta = \frac{\lambda}{2a} \\ \text{(3-dB)} \qquad & \Delta k = \frac{0.886\pi}{a}, \quad \Delta\theta = 0.886\,\frac{\lambda}{2a} \end{aligned} \tag{20.12.8}$$

The highest sidelobe lies at 13.26 dB below the main lobe and occurs at a wavenumber $k_{\text{side}} = 1.43\pi/a$. The numerical values were calculated with the MATLAB code,

```
u3 = fzero(@(u) sinc(u)-1/sqrt(2), 0.4);     % 2*u3 = 0.8859
us = fminbnd(@(u) -abs(sinc(u)), 1,2);       % us = 1.4303
R  = -20*log10(abs(sinc(us)));               % R = 13.26 dB
```

For typical apodizing functions, the mainlobe width is inversely proportional to the aperture size a, thus, the resolution capability improves as a increases. Indeed, in the limit of large a, the transform $\hat{A}(k_x)$ in the above example becomes increasingly sharper, converging to a delta function $\delta(k_x)$. This can be seen from the limit,

$$\lim_{a\to\infty} \hat{A}(k_x) = \lim_{a\to\infty} \frac{2\sin(k_x a)}{k_x} = 2\pi\delta(k_x)$$

In optics and antenna applications, there is of course a practical limit to the physical size of the apertures. By contrast, in time-series spectrum estimation applications, it is often (but not always) possible to improve resolution by increasing the length of the time-window by simply measuring a longer time signal. Next, we discuss design methods for apodization functions and aperture distributions.

There is a large literature in signal processing, antenna array design, optics, and astronomy on designing apodization functions that achieve narrow beamwidths and low sidelobes, including the design of apodized coronagraphs, star-shaped masks, and petal-shaped starshade occulters, as well as designs for superresolving and superdirective apertures [1241–1284,1479–1618].

We discuss several windowing functions in Chap. 23 with the emphasis on discrete arrays. Here, we concentrate on their continuous versions, considering $A(x)$ to be a function of the continuous variable x.

In the context of a radiating aperture, as depicted for example in (20.12.3), we note that $A(x)$ is not an apodization function, but rather it represents the actual field $E_{\rm in}(x)$ at the aperture, and is referred to as the *aperture distribution*. Here, the objective is to design $A(x)$ so that the aperture will radiate a desired radiation pattern $\hat{A}(k_x)$.

Mathematically, the two problems of designing an apodization function and synthesizing an aperture distribution are the same. Since $A(x)$ is aperture-limited to an interval $-a \le x \le a$, its Fourier transform $\hat{A}(k_x)$ cannot be bandlimited,

$$\hat{A}(k_x) = \int_{-a}^{a} A(x)\, e^{jk_x x}\, dx \quad \Leftrightarrow \quad A(x) = \int_{-\infty}^{\infty} \hat{A}(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \tag{20.12.9}$$

The total "power" contained in $A(x)$ may be defined via the Parseval identity,

$$\int_{-a}^{a} |A(x)|^2\, dx = \int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, \frac{dk_x}{2\pi} \tag{20.12.10}$$

Typically, the function $A(x)$ is characterized by certain properties: It is even in x and normalized to unity at its center, $A(0) = 1$. Its transform $\hat{A}(k_x)$ is real-valued and highly peaked in the forward direction, $k_x = 0$, with a mainlobe assumed to be concentrated in an interval, $|k_x| \le k_0$, such as the null-to-null interval. Moreover, it follows from Eq. (20.12.9) that $\hat{A}(k_x)$ is an entire function of k_x in the complex k_x-plane, which imposes certain restrictions on the types of apodization functions that can be designed.

Several design parameters have been used to assess apodizing functions, and we collect them here.

- The mainlobe width, Δk, measured either at the 3-dB level or at first-null.
- The sidelobe level R in dB relative to the mainlobe peak. Denoting by $|\hat{A}|_{\rm side}$ the maximum sidelobe height of $|\hat{A}(k_x)|$, we have the definition,

$$R = 10 \log_{10} \left[\frac{|\hat{A}(0)|^2}{|\hat{A}|^2_{\rm side}} \right] = \text{sidelobe level} \tag{20.12.11}$$

- The "encircled energy" [1504], also known as the "beam gain", or the "beam efficiency," defined as the fraction of the total power that resides within the mainlobe, that is, within the range, $|k_x| \le k_0$,

$$\mathcal{E}(k_0) = \frac{\displaystyle\int_{-k_0}^{k_0} |\hat{A}(k_x)|^2\, dk_x}{\displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x} \tag{20.12.12}$$

- The "Strehl ratio", defined as the ratio of the mainlobe peak value at $k_x = 0$, relative to the corresponding value of a uniform, unapodized, aperture of the same size $A_{\rm unif}(x) = u(a - |x|)$, so that $\hat{A}_{\rm unif}(0) = 2a$, as implied by Eq. (20.12.6),

$$S = \frac{|\hat{A}(0)|^2}{|\hat{A}_{\rm unif}(0)|^2} = \left| \frac{1}{2a} \int_{-a}^{a} A(x)\, dx \right|^2 \tag{20.12.13}$$

- The "transmission" factor, defined as the ratio of the total power to that of the uniform unapodized aperture $A_{\rm unif}(x)$,

$$\tau = \frac{\displaystyle\int_{-a}^{a} |A(x)|^2\, dx}{\displaystyle\int_{-a}^{a} |A_{\rm unif}(x)|^2\, dx} = \frac{1}{2a} \int_{-a}^{a} |A(x)|^2\, dx = \frac{1}{2a} \int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, \frac{dk_x}{2\pi} \tag{20.12.14}$$

In the context of aperture distributions, the quantity (20.12.10) is not quite the radiated power. Assuming for simplicity that $A(x)$ is a TE or a TM field relative to the xz plane, that is, $E_y(x)$ or $\eta H_y(x)$, we recall from Eq. (19.7.14) that the radiated power is given by an integral over the visible region only, $-k \le k_x \le k$,

$$2\eta P_{\rm rad} = \int_{-k}^{k} \hat{A}(k_x)|^2\, \frac{k_z}{k}\, \frac{dk_x}{2\pi} \tag{20.12.15}$$

where $k_x = k \sin\theta$, $k_z = k \cos\theta$. A similar integral over the invisible region, $|k_x| > k$, represents the reactive power stored in the vicinity of the aperture. The directivity, assumed to be maximum towards the forward direction, is given by (19.7.16),

$$\frac{D}{2ak} = \frac{|\hat{A}(0)|^2}{2a \displaystyle\int_{-k}^{k} |\hat{A}(k_x)|^2\, \frac{k_z}{k}\, \frac{dk_x}{2\pi}} \tag{20.12.16}$$

where we normalized D by the dimensionless variable $2ak$. Following Taylor [1262,1602], one can argue that since $\hat{A}(k_x)$ is highly peaked in the forward direction $k_x = 0$, or, $\theta = 0$, one may replace the quantity $k_z/k = \cos\theta$ by unity and arrive at the following approximate definitions of power and directivity,

$$2\eta P_{\rm rad} = \int_{-k}^{k} \hat{A}(k_x)|^2\, \frac{dk_x}{2\pi}\,, \qquad \frac{D}{2ak} = \frac{|\hat{A}(0)|^2}{2a \displaystyle\int_{-k}^{k} |\hat{A}(k_x)|^2\, \frac{dk_x}{2\pi}} \tag{20.12.17}$$

With Taylor we may also define the limiting directivity D_∞ integrated over the entire k_x range, and the so-called *supergain* or *superdirectivity* ratio γ,

$$\begin{aligned} \frac{D_\infty}{2ak} &= \frac{|\hat{A}(0)|^2}{2a \displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, \frac{dk_x}{2\pi}} = \frac{\left| \displaystyle\int_{-a}^{a} A(x)\, dx \right|^2}{2a \displaystyle\int_{-a}^{a} |A(x)|^2\, dx} \\ \gamma = \frac{D}{D_\infty} &= \frac{\displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x}{\displaystyle\int_{-k}^{k} |\hat{A}(k_x)|^2\, dk_x}\,, \quad \gamma - 1 = \frac{\displaystyle\int_{|k_x|>k} |\hat{A}(k_x)|^2\, dk_x}{\displaystyle\int_{|k_x|\le k} |\hat{A}(k_x)|^2\, dk_x} \end{aligned} \tag{20.12.18}$$

We note that the maximum value of $D_\infty/(2ak)$ is unity and is realized for uniform apertures. The superdirectivity ratio γ is a measure of the proportion of the reactive power to the radiated power, and is similar to (the inverse of) the encircled energy $\mathcal{E}(k)$,

but is defined over the entire visible region, not just the main beam. The quantities in Eqs. (20.12.17) and (20.12.18) can be used as additional measures to assess the properties of apodizing functions $A(x)$. We note also that $D/2ak$ and $D_\infty/2ak$ are referred to as "specific gains," or "aperture or taper efficiencies."

Since $A(x)$ is aperture-limited in the interval $-a \le x \le a$, one may apply the theory of bandlimited functions discussed in Appendix J. One needs to make the following mapping between the time-frequency variables (t, ω) of that Appendix and the spatial variables (k_x, x) here, so that a function that is frequency-limited in $[-\omega_0, \omega_0]$ becomes space-limited in $[-a, a]$, and a function that is time-concentrated in an interval $[-t_0, t_0]$ becomes wavenumber-concentrated in some desired wavenumber interval $[-k_0, k_0]$,

$$t_0\,,\ \omega_0\,,\ t\,,\ \omega \quad \Rightarrow \quad k_0\,,\ a\,,\ k_x\,,\ x \tag{20.12.19}$$

Eqs. (J.48)–(J.54) of Appendix J summarize three function bases for describing bandlimited functions: (a) the prolate spheroidal wave functions (PSWF), (b) the spherical Bessel functions, and (c) the sinc functions. Under the above mapping, the basis functions are mapped as follows:

$$\begin{aligned}
&\text{(PSWF)} && \psi_n(t_0,\omega_0,t) \quad \Rightarrow \quad \psi_n(k_0,a,k_x) = \int_{-a}^{a} \frac{1}{a\mu_n}\,\psi_n\Big(k_0,a,\frac{x}{a}k_0\Big)\,e^{jk_x x}\,dx \\
&\text{(Bessel)} && j_n(\omega_0 t) \quad \Rightarrow \quad j_n(ak_x) = \int_{-a}^{a} \frac{1}{i^n\,2a}\,P_n\Big(\frac{x}{a}\Big)\,e^{jk_x x}\,dx \\
&\text{(sinc)} && \frac{\sin(\omega_0 t-\pi n)}{\omega_0 t-\pi n} \quad \Rightarrow \quad \frac{\sin(ak_x-\pi n)}{ak_x-\pi n} = \int_{-a}^{a} \frac{1}{2a}\,e^{-j\pi n x/a}\,e^{jk_x x}\,dx
\end{aligned} \tag{20.12.20}$$

where we have also expressed them as spatial Fourier transforms so that one can determine by inspection the corresponding aperture-limited basis function in the x-domain. The index n runs over $0 \le n < \infty$ for the prolate and Bessel cases, and over $-\infty < n < \infty$ for the sinc basis. The function $P_n(\xi)$ is the standard un-normalized nth-order Legendre polynomial defined over $|\xi| \le 1$. The parameter μ_n is given by $\mu_n = i^n\sqrt{2\pi\lambda_n/c}$, where $c = ak_0$ is the *space-bandwidth product*, and λ_n is the nth eigenvalue characterizing the prolate functions, as discussed in Appendix J.

To clarify the notation, we have defined $\psi_n(k_0,a,k_x)$ in terms of the standard scaled PSWF function $\phi_n(c,\eta)$ of Eq. (J.7) as follows,

$$\psi_n(k_0,a,k_x) = \frac{1}{\sqrt{k_0}}\,\phi_n\left(c,\frac{k_x}{k_0}\right), \quad c = ak_0 \tag{20.12.21}$$

The bases satisfy the following orthogonality properties, adapted from Appendix J,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \psi_n(k_0,a,k_x)\,\psi_m(k_0,a,k_x)\,dk_x = \delta_{nm} \\
&\int_{-k_0}^{k_0} \psi_n(k_0,a,k_x)\,\psi_m(k_0,a,k_x)\,dk_x = \lambda_n\,\delta_{nm} \\
&\int_{-\infty}^{\infty} j_n(ak_x)\,j_m(ak_x)\,dk_x = \frac{\pi}{a}\,\frac{\delta_{nm}}{2n+1} \\
&\int_{-\infty}^{\infty} \frac{\sin(ak_x-\pi n)}{ak_x-\pi n}\cdot\frac{\sin(ak_x-\pi m)}{ak_x-\pi m}\,dk_x = \frac{\pi}{a}\,\delta_{nm}
\end{aligned} \tag{20.12.22}$$

where the prolate basis satisfies dual orthogonality properties over the infinite interval and over the finite interval $[-k_0, k_0]$.

Because these bases are complete in the space of (square-integrable) aperture-limited functions, it follows that $\hat{A}(k_x)$ and $A(x)$ can be expanded as linear combinations of the basis functions, with the expansion coefficients determined with the help of the above orthogonality properties.

Denoting the indicator function of the interval $[-a,a]$ by, $\chi_a(x) = u(a-|x|)$, we have the following expansions with respect to the PSWF basis,

$$\begin{aligned}
\hat{A}(k_x) &= \sum_{n=0}^{\infty} \hat{A}_n\,\psi_n(k_0,a,k_x) \\
A(x) &= \sum_{n=0}^{\infty} \frac{\hat{A}_n}{a\mu_n}\,\psi_n\Big(k_0,a,\frac{x}{a}k_0\Big)\cdot\chi_a(x) \\
\hat{A}_n &= \int_{-\infty}^{\infty} \hat{A}(k_x)\,\psi_n(k_0,a,k_x)\,dk_x = \frac{2\pi}{a\mu_n^*}\int_{-a}^{a} A(x)\,\psi_n\Big(k_0,a,\frac{x}{a}k_0\Big)dx \\
&= \frac{1}{\lambda_n}\int_{-k_0}^{k_0} \hat{A}(k_x)\,\psi_n(k_0,a,k_x)\,dk_x
\end{aligned} \tag{20.12.23}$$

The spherical Bessel function basis is not very common, however, see [1678,1679] for some applications to communications and antenna design. The expansions are,

$$\begin{aligned}
\hat{A}(k_x) &= \sum_{n=0}^{\infty} \hat{A}_n\,j_n(ak_x) \\
A(x) &= \sum_{n=0}^{\infty} \frac{\hat{A}_n}{i^n\,2a}\,P_n\Big(\frac{x}{a}\Big)\cdot\chi_a(x) \\
\hat{A}_n &= \frac{a(2n+1)}{\pi}\int_{-\infty}^{\infty} \hat{A}(k_x)\,j_n(ak_x)\,dk_x = i^n(2n+1)\int_{-a}^{a} A(x)\,P_n\Big(\frac{x}{a}\Big)dx
\end{aligned} \tag{20.12.24}$$

In the sinc-function basis the expansion is very simple, with the coefficients $\hat{A}_n$ being simply the samples of $\hat{A}(k_x)$ at the discrete wavenumbers $k_x = nk_s$, where $k_s = \pi/a$ is the sampling interval,

$$\hat{A}(k_x) = \sum_{n=-\infty}^{\infty} \hat{A}_n \frac{\sin(ak_x - \pi n)}{ak_x - \pi n} = \sum_{n=-\infty}^{\infty} \hat{A}_n \frac{\sin(a(k_x - nk_s))}{a(k_x - nk_s)}$$

$$A(x) = \frac{1}{2a} \sum_{n=-\infty}^{\infty} \hat{A}_n e^{-j\pi nx/a} \cdot \chi_a(x) = \frac{1}{2a} \sum_{n=-\infty}^{\infty} \hat{A}_n e^{-jnk_s x} \cdot \chi_a(x) \tag{20.12.25}$$

$$\hat{A}_n = \hat{A}(nk_s) = \hat{A}\left(\frac{\pi n}{a}\right) = \int_{-a}^{a} A(x)\, e^{j\pi nx/a}\, dx$$

Eq. (20.12.25) is a consequence of Shannon's sampling theorem, as applied to the aperture-limited function $A(x)$ and is known in the antenna design context as the *Woodward-Lawson method* [1600,1601]. Its discrete version for array design is discussed in Sec. 23.5. The method is useful if the sum over n is finite, as it is in the case of Taylor's n-bar method considered below.

It is often convenient to think of $\hat{A}(k_x)$ as a function of the normalized wavenumber variable, $u = k_x/k_s = ak_x/\pi$, or, $k_x = uk_s = \pi u/a$, that is, define,

$$F(u) = \hat{A}(k_x)\Big|_{k_x = \frac{\pi u}{a}} \tag{20.12.26}$$

Then, the Fourier transform relationships (20.12.9) read,

$$F(u) = \int_{-a}^{a} A(x)\, e^{j\pi ux/a}\, dx \quad \Leftrightarrow \quad A(x) = \frac{1}{2a} \int_{-\infty}^{\infty} F(u)\, e^{-j\pi ux/a}\, du \tag{20.12.27}$$

The sampled wavenumbers $k_x = nk_s$ in the Woodward-Lawson case then become integers in u-space, $u = k_x/k_s = nk_s/k_s = n$, and Eqs. (20.12.25) read,

$$F(u) = \sum_{n=-\infty}^{\infty} F(n) \frac{\sin(\pi(u-n))}{\pi(u-n)}$$

$$A(x) = \frac{1}{2a} \sum_{n=-\infty}^{\infty} F(n)\, e^{-j\pi nx/a} \cdot \chi_a(x) \tag{20.12.28}$$

$$F(n) = \int_{-a}^{a} A(x)\, e^{j\pi nx/a}\, dx$$

The quantity u represents the observation angle θ (or, rather, $\sin\theta$), in the standard beamwidth units of $\lambda/(2a)$, indeed, since $k_x = k\sin\theta$, we have with $k = 2\pi/\lambda$,

$$\boxed{\sin\theta = \frac{k_x}{k} = u\frac{\lambda}{2a}} \tag{20.12.29}$$

If the mainlobe of $\hat{A}(k_x)$ is concentrated in $|k_x| \le k_0$, then, in u-units, $F(u)$ will be concentrated in $|u| \le u_0$, where $u_0 = ak_0/\pi$. The space-bandwidth product will be $c = ak_0 = \pi u_0$, and the maximum angle of the mainlobe,

$$\sin\theta_0 = \frac{k_0}{k} = u_0 \frac{\lambda}{2a} \tag{20.12.30}$$

Thus, to be able to resolve a weaker signal from a stronger one, the weaker one must have $u > u_0$, falling in the sidelobe range of the stronger one, and its strength must be above the sidelobe level R.

Next, we discuss three highly effective apodization functions that are optimum in some sense and have narrow mainlobes and controllable sidelobes: (a) the prolate apodization function, which is optimal in the sense that it maximizes the encircled energy, (b) Taylor's I_0-sinh one-parameter distribution, which is an excellent closed-form approximation to the optimum prolate, and (c) Taylor's n-bar distribution, which is a compromise between the optimum prolate and the optimum Chebyshev window.

The Chebyshev window has equiripple sidelobes and is optimum in a different sense, that is, it has the narrowest mainlobe for a given sidelobe level. To understand this intuitively, we note that as one pushes the sidelobes down, the mainlobe gets wider, and conversely, if one allows the sidelobes to go up, then the mainlobe will get narrower. Thus, for a prescribed sidelobe level, if all sidelobes have the same height, the mainlobe will be the narrowest possible.

20.13 Prolate Window

The prolate apodization function is up to a scale factor the zero-order PSWF function ψ_0 considered above, and was originally derived by Slepian for both the 1-D and 2-D problems [1505]. It is defined by the parameters k_0, a, where the interval $[-k_0, k_0]$ is a measure of the width of the mainlobe (but k_0 is not a null point). The window is optimum in the sense that $\hat{A}(k_x)$ maximizes the power contained in the interval $[-k_0, k_0]$, that is, the encircled energy ratio,

$$\mathcal{E}(k_0) = \frac{\int_{-k_0}^{k_0} |\hat{A}(k_x)|^2\, dk_x}{\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x} = \max \tag{20.13.1}$$

This problem was solved in Eq. (J.57) of Appendix J where it was shown that the solution is the zero-order PSWF function, therefore, keeping only the $n = 0$ term in (20.12.23) we have in the k_x and x domains,

$$\hat{A}(k_x) = \hat{A}_0 \psi_0(k_0, a, k_x)\,, \quad A(x) = \frac{\hat{A}_0}{a\mu_0} \psi_0\left(k_0, a, \frac{x}{a}k_0\right) \cdot \chi_a(x)$$

with the maximized value of the encircled energy given by the λ_0 eigenvalue, $\mathcal{E}(k_0) = \lambda_0$. The constant $\hat{A}_0$ can be determined by requiring that $A(0) = 1$, so that,

$$\frac{\hat{A}_0}{a\mu_0} \psi_0(k_0, a, 0) = 1$$

Thus,

$$\hat{A}(k_x) = a\mu_0 \frac{\psi_0(k_0, a, k_x)}{\psi_0(k_0, a, 0)}\,, \quad A(x) = \frac{\psi_0\left(k_0, a, \frac{x}{a}k_0\right)}{\psi_0(k_0, a, 0)} \cdot \chi_a(x) \tag{20.13.2}$$

There is no obvious relationship between the wavenumber k_0 that defines the extent of the mainlobe and the corresponding sidelobe level R in dB. One would like to be able to specify R and determine the proper value of k_0 to be used in the design.

There is a roundabout solution to this. Because Taylor's one-parameter I_0-sinh window discussed in the next section is a close approximation to the prolate ψ_0 window, one can use the sidelobe-width calculation for that window and slightly tweak it to make it applicable here. Taylor's window depends on a single parameter B which is related to the sidelobe level R in dB by the relationship,

$$R = R_0 + 20\log_{10}\left(\frac{\sinh(\pi B)}{\pi B}\right) \tag{20.13.3}$$

where $R_0 = 13.26$ dB is the maximum sidelobe level of the rectangular window that we calculated in Sec. 20.12. Given R, one can solve (20.13.3) for B numerically. Then, one could take the space-bandwidth product to be approximately, $c = ak_0 \approx \pi B$. However, as we showed in Sec. 23.11, a better choice is,

$$c = ak_0 = (0.95B + 0.14)\pi \tag{20.13.4}$$

This works well over the range, $15 \le R \le 150$ dB. The MATLAB function, **tbw**, described in the next section can be used to solve (20.13.3) for B, Then, k_0 is calculated from (20.13.4). Alternatively, one can use the slightly less accurate Kaiser-Schafer [1266] empirical expression of Eq. (20.14.5). The directivity and other window parameters are worked out in Problem 20.6. For large k_x or $u = ak_x/\pi$, it follows from Eq. (J.10) that $\hat{A}(k_x)$ decays like $\text{sinc}(u)$,

$$\hat{A}(k_x) \to \text{const.}\,\frac{\sin(ak_x)}{ak_x} = \text{const.}\,\frac{\sin(\pi u)}{\pi u}, \quad \text{as} \quad |k_x| \to \infty$$

Example 20.13.1: Fig. 20.13.1 shows two prolate windows designed for $R = 60$ and $R = 100$ dB. The aperture half-width was $a = 1$. The following MATLAB code segment illustrates the calculation. The function **pswf** from Appendix J was used,

```
a=1;
x = linspace(-1,1,401)*a;
u = linspace(-20,20,2001); kx = pi*u/a;

for R = [60,100]
   [B,Du] = tbw(R);              % Du = 3-dB width
   c = (0.96*B+0.14)*pi;         % c parameter for prolate window
   k0 = c/a;                     % k0 parameter
   u0 = c/pi;                    % k0 in units of u

   Ak = pswf(k0,a,0,kx);         % A(kx), kx domain, PSWF from Appendix J
   [A0,la0] = pswf(k0,a,0,0);    % A(0) and lambda0
   Ax = pswf(k0,a,0,x*k0/a);     % A(x), x-domain

   Ak = abs(Ak/A0);     % normalized to unity at kx=0 for plotting in dB
   Ax = Ax/A0;          % normalized to unity at x=0 by convention

   figure; plot(u, 20*log10(Ak));
   figure; plot(x, Ax);
end
```


The wavenumber plots were in units of $u = ak_x/\pi$. The space-bandwidth products were in the two cases, $c = 8.2880$ and $c = 12.1736$, and the corresponding u-parameters, $u_0 = ak_0/\pi = 2.6382$ and $u_0 = 4.1933$. □

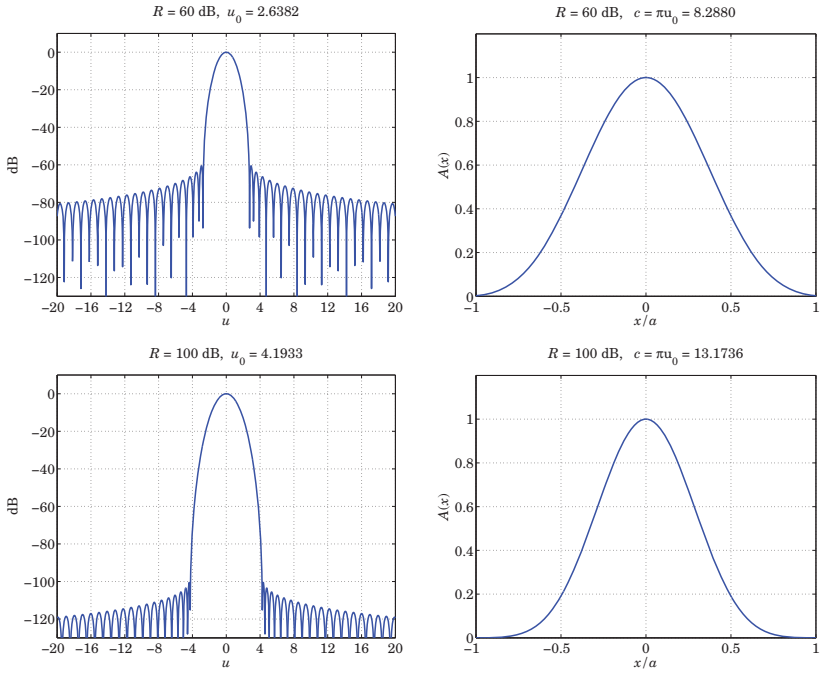


Fig. 20.13.1 Prolate window, $|\hat{A}(k_x)|$ in dB versus $u = ak_x/\pi$, and $A(x)$ versus x.

Example 20.13.2: To examine the feasibility of detecting exo-planets with the $R = 100$ dB window that has $u_0 = 4.2$, consider a Mercury-type planet at a distance of 5×10^{10} meters (about one-third A.U.) from its sun, and relative intensity 10^{-10}, at a distance of 10^{17} meters (about 11 light years) from earth. The subtended angle will be $\theta \approx \sin\theta = 5\times10^{10}/10^{17} = 5\times10^{-7}$ radians, and the corresponding value of u for visible light at, say, 500 nm, using a 4-meter telescope,

$$u = \frac{ak_x}{\pi} = \frac{2a}{\lambda}\sin\theta = \frac{4}{500\times10^{-9}}\cdot 5\times10^{-7} = 4$$

which is at the threshold of visibility since $u \lessapprox u_0$. On the other hand, a Jupiter-like planet with intensity 10^{-10}, at 5 A.U. from its sun, and at a distance of 16 light years (10^6 A.U.) from earth would have,

$$u = \frac{2a}{\lambda}\sin\theta = \frac{4}{500\times10^{-9}}\cdot\frac{5}{10^6} = 40$$

and would be easily resolvable with such a window. □

20.14 Taylor's One-Parameter Window

Taylor's one-parameter window function [1262] is based on the Fourier transform pair,

$$\frac{\sinh(a\sqrt{k_0^2-k_x^2})}{a\sqrt{k_0^2-k_x^2}} = \frac{1}{2a}\int_{-a}^{a} I_0\big(k_0\sqrt{a^2-x^2}\big)\, e^{jk_x x}\, dx \tag{20.14.1}$$

where I_0 is the order-0 modified Bessel function of the first kind and k_0 is a parameter playing the same role as in the prolate window. Thus, we define,

$$\hat{A}(k_x) = \frac{\sinh(a\sqrt{k_0^2-k_x^2})}{a\sqrt{k_0^2-k_x^2}} \quad\Leftrightarrow\quad A(x) = \frac{1}{2a}\, I_0\big(k_0\sqrt{a^2-x^2}\big)\cdot\chi_a(x) \tag{20.14.2}$$

Its discrete version is the Taylor-Kaiser window used in Sec. 23.10 for array design. Introducing the variable, $\pi B = ak_0$, and the normalized wavenumber, $u = ak_x/\pi$, we may rewrite (20.14.2) in the notation of Eq. (20.12.26),

$$\begin{aligned} F(u) &= \frac{\sinh\big(\pi\sqrt{B^2-u^2}\big)}{\pi\sqrt{B^2-u^2}} = \frac{\sin\big(\pi\sqrt{u^2-B^2}\big)}{\pi\sqrt{u^2-B^2}} \\ A(x) &= \frac{1}{2a}\, I_0\big(\pi B\sqrt{1-x^2/a^2}\big)\cdot\chi_a(x) \end{aligned} \tag{20.14.3}$$

If so desired, $A(x)$ may be normalized to unity at $x=0$. The pattern function $F(u)$ switches from its sinh to its sine form at $u = B$, and behaves like $\sin(\pi u)/\pi u$ for large u. The first null occurs at $\sqrt{u^2-B^2} = 1$, or, $u_1 = \sqrt{B^2+1}$, with the sidelobe structure beginning after that. The highest sidelobe level is the maximum value of the ratio $|\sin(\pi u)/\pi u|$ other than at $u = 0$. That was determined in Sec. 20.12 to be, $R_0 = 13.26$ dB. The sidelobe level R is defined as the relative ratio of the value at $u = 0$, that is, $F(0) = \sinh(\pi B)/\pi B$, to the highest sidelobe level R_0, thus, we obtain the following relationship between B and R, which was used in (20.13.3),

$$R = 20\log_{10}\left[\frac{|F(0)|}{|F|_{\max\,\text{side}}}\right] = R_0 + 20\log_{10}\left(\frac{\sinh(\pi B)}{\pi B}\right) \tag{20.14.4}$$

This equation may be solved numerically for B. Alternatively, one can calculate B using the slightly less accurate Kaiser-Schafer [1266] empirical expression discussed in Sec. 23.10 in the context of array design,

$$\pi B = \begin{cases} 0.76609\,(R-R_0)^{0.4} + 0.09834\,(R-R_0), & R_0 < R \le 60 \\ 0.12438\,(R+6.3), & 60 < R < 120 \end{cases} \tag{20.14.5}$$

The 3-dB width in u-units is given by $\Delta u = 2u_3$, where u_3 is the solution of the half-power condition, $|F(u_3)/F(0)|^2 = 1/2$,

$$\frac{\sin\big(\pi\sqrt{u_3^2-B^2}\big)}{\pi\sqrt{u_3^2-B^2}} = \frac{1}{\sqrt{2}}\,\frac{\sinh(\pi B)}{\pi B} \tag{20.14.6}$$

The beam efficiency or encircled energy may be defined with respect to $k_0 = \pi B/a$, or, with respect to the first null, $k_1 = \pi u_1/a = \pi\sqrt{B^2+1}/a$. We choose the former,

$$\mathcal{E}(k_0) = \frac{\displaystyle\int_{-k_0}^{k_0} |\hat{A}(k_x)|^2\, dk_x}{\displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x} = \frac{\displaystyle\int_{-B}^{B} |F(u)|^2\, du}{\displaystyle\int_{-\infty}^{\infty} |F(u)|^2\, du} \tag{20.14.7}$$

in which the denominator can be expressed in terms of $A(x)$, using Parseval's identity,

$$\mathcal{E}(k_0) = \frac{\displaystyle\int_{-B}^{B} |F(u)|^2\, du}{2a\displaystyle\int_{-a}^{a} |A(x)|^2\, dx} = \frac{\displaystyle\int_{-B}^{B} \Big|\operatorname{sinc}\Big(\sqrt{u^2-B^2}\Big)\Big|^2\, du}{\displaystyle\int_0^1 \Big|I_0\Big(\pi B\sqrt{1-x^2}\Big)\Big|^2\, dx} \tag{20.14.8}$$

Similarly, the normalized directivity can be expressed as,

$$\frac{D_\infty}{2ak} = \frac{|F(0)|^2}{\displaystyle\int_{-\infty}^{\infty} |F(u)|^2\, du} = \frac{\big|\sinh(\pi B)\big|^2}{(\pi B)^2\displaystyle\int_0^1 \Big|I_0\Big(\pi B\sqrt{1-x^2}\Big)\Big|^2\, dx} \tag{20.14.9}$$

The numerical calculation of $B, \Delta u, D_\infty, \mathcal{E}$ from R is implemented by the MATLAB function, **tbw**,[†] with usage,

```
[B,Du,D,E] = tbw(R)      % Taylor's B-parameter and beamwidth
```

Eqs. (20.14.4) and (20.14.6) are solved with the built-in function **fzero**, while the finite integrations in (20.14.8) and (20.14.9) are implemented with the help of the tanh-sinh quadrature function **quadts**. To clarify the calculation, the essential MATLAB code of the **tbw** function is,

```
F = @(u,B) abs(sinc(sqrt(u.^2-B.^2)));   % pattern function

u0 = fminbnd(@(u) -F(u,0), 1,2);         % u0 = 1.430292
R0 = -20*log10(F(u0,0));                 % R0 = 13.261459 dB
B0 = 0.04*R + 0.06;                      % initial search estimates
u3 = 0.005*R + 0.5;                      % initial search estimates

B = fzero(@(B) 20*log10(F(0,B)) + R0 - R, B0);
Du = 2*fzero(@(u) F(u,B) - F(0,B)/sqrt(2), u3);

[wx,x] = quadts(0,1);              % weights, quadrature points for x-integration
[wu,u] = quadts(0,B);              % weights, quadrature points for u-integration
%[wu,u]=quadts(0,sqrt(B^2+1));     % if mainlobe is defined null-to-null

P = wx' * besseli(0,pi*B*sqrt(1-x.^2)).^2;     % total power
P0 = 2 * wu' * F(u,B).^2;                      % power in mainlobe

D = F(0,B)^2 / P;                  % directivity
E = P0 / P;                        % beam efficiency
```

[†] which supplants the older function, **taylorbw**

Example 20.14.1: The following MATLAB code segment shows how to calculate the aperture distribution and the wavenumber pattern of the Taylor window, for the same specifications as in Example 20.13.1.

```
a=1;
x = linspace(-1,1,401)*a;
u = linspace(-20,20,2001); kx = pi*u/a;

F = @(u,B) abs(sinc(sqrt(u.^2-B.^2)));

for R = [60,100]
   [B,Du] = tbw(R);              % Du = 3-dB width

   Fk = 20*log10(F(u,B) / F(0,B));
   Fx = besseli(0,pi*B*sqrt(1-x.^2/a^2)) / besseli(0,pi*B);

   figure; plot(u,Fk,'b-');
   figure; plot(x,Fx,'b-');
end
```

The graphs are not shown since they are virtually indistinguishable from those of Fig. 20.13.1.

20.15 Taylor's N-bar Window

It is evident from Fig. 20.13.1 that the prolate and Taylor's one-parameter windows have their largest sidelobes near the mainlobe, meeting the design specification for R. The sidelobes that are further away have attenuations higher than R (with an envelope decaying like $1/u$). Thus, the sidelobe behavior of these windows is better than it needs to be, and this comes at the expense of a wider mainlobe. If the farther sidelobes could be raised to the same specified level R, then the mainlobe would narrow.

This is the idea behind the Dolph-Chebyshev window, whose application to discrete arrays is discussed in Sec. 23.9. Van der Maas [1250] has shown that in the continuous-space limit (i.e., for large number of equally-spaced array elements with infinitesimal spacings spanning the aperture of length $2a$), the Dolph-Chebyshev window reduces to the ideal pattern,

$$\hat{A}_0(k_x) = \cosh\left(a\sqrt{k_0^2 - k_x^2}\right) = \cos\left(a\sqrt{k_x^2 - k_0^2}\right) \tag{20.15.1}$$

or, in units of $u = ak_x/\pi$, defining the parameter A through, $\pi A = ak_0$,

$$F_0(u) = \hat{A}_0(k_x)\Big|_{k_x=\frac{\pi u}{a}} = \cosh\left(\pi\sqrt{A^2 - u^2}\right) = \cos\left(\pi\sqrt{u^2 - A^2}\right) \tag{20.15.2}$$

The mainlobe corresponds to the range, $|k_x| \le k_0$, or, $|u| \le A$. The sidelobe range begins at $u = A$, where the hyperbolic cosine switches to ordinary cosine, which has equal ripples. Since the maximum value of the cosine ripples is unity, and the value at $u = 0$ is, $F_0(0) = \cosh(\pi A)$, it follows that the attenuation R will be in terms of A,

$$\begin{aligned} R &= 20\log_{10}\left[\frac{|F_0(0)|}{1}\right] = 20\log_{10}[\cosh(\pi A)] \quad \text{(dB)} \\ A &= \frac{1}{\pi}\operatorname{acosh}(10^{R/20}) \end{aligned} \tag{20.15.3}$$

Unfortunately, the ideal pattern $\hat{A}_0(k_x)$ is not realizable because its inverse Fourier transform $A(x)$ has spikes at its end points at $x = \pm a$, and is not bounded. To see this, consider the Fourier transform pair (20.14.2) of Taylor's one-parameter window,

$$\frac{\sinh(a\sqrt{k_0^2 - k_x^2})}{\sqrt{k_0^2 - k_x^2}} \quad \Leftrightarrow \quad \frac{1}{2} I_0(k_0\sqrt{a^2 - x^2}) \cdot \chi_a(x) \tag{20.15.4}$$

It is evident that $\hat{A}_0(k_x) = \cosh(a\sqrt{k_0^2 - k_x^2})$ is the derivative of the left-hand side of the above with respect to the parameter a. Thus, taking the a-derivatives of both sides in (20.15.4), we obtain the Fourier pair,[†]

$$\cosh\left(a\sqrt{k_0^2 - k_x^2}\right) \quad \Leftrightarrow \quad \frac{ak_0}{2}\,\frac{I_1(k_0\sqrt{a^2 - x^2})}{\sqrt{a^2 - x^2)}} \cdot \chi_a(x) + \frac{1}{2}[\delta(x-a) + \delta(x+a)] \tag{20.15.5}$$

where we used the modified Bessel function property $I_0'(z) = I_1(z)$, and $I_0(0) = 1$. The delta functions arise from the differentiation of the indicator function. Indeed, noting that $\chi_a(x)$ can be written as the product of unit steps, $\chi_a(x) = u(a - |x|) = u(a - x)\cdot u(a + x)$, and using the differentiation property, $u'(z) = \delta(z)$, we have,

$$\frac{\partial}{\partial a}\chi_a(x) = \delta(a - x)\cdot u(a + x) + \delta(a + x)\cdot u(a - x) = \delta(x - a) + \delta(x + a)$$

In Taylor's n-bar window method [1602] only the first few nearest sidelobes are equiripple at level R, while the rest decay like $\operatorname{sinc}(u)$ as in the prolate and one-parameter cases. To explain the method, we start with the standard infinite product expansions of the ideal response and the sinc response in terms of their zeros [1791],

$$\begin{aligned} \cos\left(\pi\sqrt{(u/\sigma)^2 - A^2}\right) &= \cosh(\pi A)\cdot\prod_{n=1}^{\infty}\left[1 - \frac{u^2}{\sigma^2[A^2 + (n - \frac{1}{2})^2]}\right] \\ \operatorname{sinc}(u) = \frac{\sin(\pi u)}{\pi u} &= \prod_{n=1}^{\infty}\left[1 - \frac{u^2}{n^2}\right] \end{aligned} \tag{20.15.6}$$

where a dilation factor $\sigma \gtrsim 1$ has been introduced in the ideal pattern resulting in a slightly wider mainlobe. The zeros of the ideal pattern are at $u_n = \pm\sigma\sqrt{A^2 + (n - \frac{1}{2})^2}$, and those of the sinc, $u_n = \pm n$. A modified pattern can be made up whose first $\bar{n} - 1$ zeros are taken from the equiripple pattern, and the rest from the sinc, that is,

$$F(u) = \cosh(\pi A)\cdot\prod_{n=1}^{\bar{n}-1}\left[1 - \frac{u^2}{\sigma^2[A^2 + (n - \frac{1}{2})^2]}\right]\cdot\prod_{n=\bar{n}}^{\infty}\left[1 - \frac{u^2}{n^2}\right] \tag{20.15.7}$$

In order to pass smoothly from the first group of zeros to the second, the dilation factor σ is chosen such that two groups match at $n = \bar{n}$, that is,

$$\sigma\sqrt{A^2 + (\bar{n} - \tfrac{1}{2})^2} = \bar{n} \quad \Rightarrow \quad \boxed{\sigma = \frac{\bar{n}}{A^2 + (\bar{n} - \frac{1}{2})^2}} \tag{20.15.8}$$

[†] Previously, we encountered this pair in a somewhat different form in Sec. 3.3.

In addition, Taylor imposes the condition that $\partial\sigma/\partial\bar{n} < 0$, which requires that $\bar{n} > 2A^2 + \frac{1}{2}$. However, this requirement can be relaxed if one is interested in satisfying the sidelobe specification R only in a neighborhood near the mainlobe and is willing to let farther sidelobes rise above R, which will result in an even narrower mainlobe.

As $\bar{n}$ increases, $F(u)$ resembles more and more the ideal pattern of Eq. (20.15.6). The second product in $F(u)$ consisting of the sinc zeros can be expressed more conveniently in terms of gamma functions,[†]

$$\prod_{n=\bar{n}}^{\infty}\left[1-\frac{u^2}{n^2}\right]=\frac{\Gamma^2(\bar{n})}{\Gamma(\bar{n}+u)\,\Gamma(\bar{n}-u)} \tag{20.15.9}$$

Thus, $F(u)$ takes the simpler form,

$$F(u)=\cosh(\pi A)\cdot\frac{\Gamma^2(\bar{n})}{\Gamma(\bar{n}+u)\,\Gamma(\bar{n}-u)}\cdot\prod_{n=1}^{\bar{n}-1}\left[1-\frac{u^2}{\sigma^2\left[A^2+(n-\frac{1}{2})^2\right]}\right] \tag{20.15.10}$$

Since Eq. (20.15.9) vanishes at all integers $u = |n| \geq \bar{n}$, the samples of $F(u)$ will also vanish at those integers, that is, $F(n) = 0$, for $|n| \geq \bar{n}$. One can then express $F(u)$, and the corresponding aperture distribution $A(x)$, in an alternative form using the Woodward-Lawson expansions of Eq. (20.12.28) as finite sums involving the surviving samples, $F(n)$, $|n| \leq \bar{n}-1$,

$$\begin{aligned} F(u) &= \sum_{n=-\bar{n}+1}^{\bar{n}-1} F(n)\,\frac{\sin\big(\pi(u-n)\big)}{\pi(u-n)} \\ A(x) &= \frac{1}{2a}\sum_{n=-\bar{n}+1}^{\bar{n}-1} F(n)\,e^{-j\pi nx/a}\cdot\chi_a(x) \end{aligned} \tag{20.15.11}$$

The required values of $F(n)$ are computed from (20.15.10), for $-(\bar{n}-1)\leq n\leq\bar{n}-1$,

$$F(n)=\cosh(\pi A)\cdot\frac{\Gamma^2(\bar{n})}{\Gamma(\bar{n}+n)\,\Gamma(\bar{n}-n)}\cdot\prod_{m=1}^{\bar{n}-1}\left[1-\frac{n^2}{\sigma^2\left[A^2+(m-\frac{1}{2})^2\right]}\right] \tag{20.15.12}$$

The center value is $F(0) = \cosh(\pi A)$. The 3-dB width can be estimated accurately by finding the 3-dB frequency of the ideal pattern, that is, the solution of the equation,

$$\cosh\left(\pi\sqrt{A^2-(u_3/\sigma)^2}\right)=\frac{1}{\sqrt{2}}\cosh(\pi A)$$

which implies for the width, $\Delta u = 2u_3$,

$$\Delta u = 2\sigma\sqrt{A^2-\frac{1}{\pi^2}\operatorname{acosh}^2\left(\frac{1}{\sqrt{2}}\cosh(\pi A)\right)} \tag{20.15.13}$$

The directivity/specific gain is given by the finite sum,

$$\frac{D_\infty}{2ak}=\frac{|F(0)|^2}{\displaystyle\int_{-\infty}^{\infty}|F(u)|^2\,du}=\frac{|F(0)|^2}{\displaystyle\sum_{n=-\bar{n}+1}^{\bar{n}-1}|F(n)|^2} \tag{20.15.14}$$

[†]Following from the representation, $1/\Gamma(z)=z\,e^{\gamma z}\prod_{n=1}^{\infty}e^{-z/n}\left(1+\frac{z}{n}\right)$, γ is Euler's constant.

The transition point between the cosh and cosine behavior, $\sqrt{A^2-(u/\sigma)^2}=0$, or, $u_0=\sigma A$, can be used as an estimate of the mainlobe width. Alternatively, one can use the first null, $u_1=\sigma\sqrt{A^2+\frac{1}{4}}$. The beam efficiency/encircled energy can be computed relative to the interval $[-u_0,u_0]$ or $[-u_1,u_1]$. We choose the former,

$$E_0=\frac{\displaystyle\int_{-u_0}^{u_0}|F(u)|^2\,du}{\displaystyle\int_{-\infty}^{\infty}|F(u)|^2\,du}=\frac{\displaystyle\int_{-u_0}^{u_0}|F(u)|^2\,du}{\displaystyle\sum_{n=-\bar{n}+1}^{\bar{n}-1}|F(n)|^2} \tag{20.15.15}$$

Given $R, \bar{n}$, the MATLAB function, **tnb1**, calculates the parameters $A, \Delta u, D_\infty, E_0$, as well as the coefficient vector, $F(n)$, $|n| \leq \bar{n}-1$. It has usage,

```
[A,Fn,Du,D,E] = tnb1(R,nb);    % Taylor's n-bar window (1-D)
```

In **tnb1**, the integral in (20.15.15) is computed numerically using the tanh-sinh quadrature function **quadts**. Although, Eq. (20.15.12) is convenient for the evaluation of $F(n)$, the numerical evaluation of the gamma-functions can lead to extremely large values for larger $\bar{n}$, for example, $\Gamma(100) = 9.3326\times10^{155}$, and the ratios of such large values in (20.15.12) might lead to inaccuracies. An alternative way to calculate $F(n)$ is by writing (20.15.7) in the form,

$$F(u)=\cosh(\pi A)\cdot\frac{\sin(\pi u)}{\pi u}\cdot\frac{\displaystyle\prod_{n=1}^{\bar{n}-1}\left[1-\frac{u^2}{\sigma^2\left[A^2+(n-\frac{1}{2})^2\right]}\right]}{\displaystyle\prod_{n=1}^{\bar{n}-1}\left[1-\frac{u^2}{n^2}\right]} \tag{20.15.16}$$

The ratio of the sinc function and the product in the denominator involves pole/zero cancellations at $u = \pm n$. By taking the limit $u \to n$ and allowing for such cancellations, we arrive at the following alternative expression for $F(n)$, which is implemented in the function **tnb1**,

$$F(n)=\cosh(\pi A)\cdot\frac{(-1)^{n+1}\displaystyle\prod_{m=1}^{\bar{n}-1}\left[1-\frac{n^2}{\sigma^2\left[A^2+(m-\frac{1}{2})^2\right]}\right]}{2\displaystyle\prod_{\substack{m=1\\m\neq n}}^{\bar{n}-1}\left[1-\frac{n^2}{m^2}\right]} \tag{20.15.17}$$

for $1 \leq n \leq \bar{n}-1$, then setting, $F(-n) = F(n)$, and, $F(0) = \cosh(\pi A)$.

Example 20.15.1: This example illustrates how to use the output of **tnb1** to calculate $F(u)$ and $A(x)$ in a vectorized fashion (in u and x). The specifications were $R = 100$ dB and $\bar{n} = 30$. Fig. 20.15.1 plots $F(u)$ over the range $0 \leq u \leq 40$, and in an expanded view on the right, and compares it to the one-parameter design. The MATLAB code used to generate these graphs was as follows,

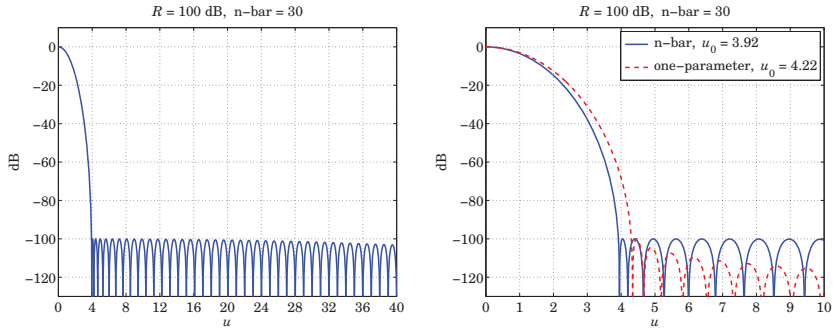


Fig. 20.15.1 Comparison of Taylor's n-bar and one-parameter windows.

```
R = 100; nb = 30;

[A,Fn,Du,D,E] = tnb1(R,nb);

sigma = nb / sqrt(A^2 + (nb-0.5)^2);
u0 = sigma * A;                    % turning point
% u0 = sigma * sqrt(A^2 + 1/4);    % first null

u = linspace(0,40,4001);
x = linspace(-1,1,401);

F=0; Ax=0;
for n = -(nb-1) : nb-1
   F = F + Fn(nb+n)*sinc(u-n);
   Ax = Ax + Fn(nb+n)*cos(pi*n*x)/2;
end

Fabs = abs(F); FdB = 20*log10(Fabs/max(Fabs));
Amax = max(abs(Ax)); Ax = Ax/Amax;

f = @(u,B) abs(sinc(sqrt(u.^2-B.^2)));
[B,Dub] = tbw(R);
v0 = B;                  % turning point
% ub = sqrt(B^2+1);      % first null

GdB = 20*log10(f(u,B) / f(0,B));

figure; plot(u,FdB,'b-', u,GdB,'r--')
```

The calculated transition points were $u_0 = \sigma A = 3.9173$ for the $\bar{n}$ case, and $u_0 = B = 4.2222$ for the one-parameter design. The values of A, σ were $A = 3.8853$ and $\sigma = 1.0082$, and the minimum value of $\bar{n} > 2A^2 + 1/2 = 30.67$. If $\bar{n}$ is chosen less than that, then as we mentioned, the far sidelobes would become higher than the specified R and the mainlobe a little narrower.

Fig. 20.15.2 shows the pattern resulting from $\bar{n} = 6$ and a somewhat higher $R = 110$ dB in order to still meet the 100 dB requirement in a narrow range, $u_0 \le u \le 5$, where for this design the mainlobe transition point was $u_0 = 3.67$. □

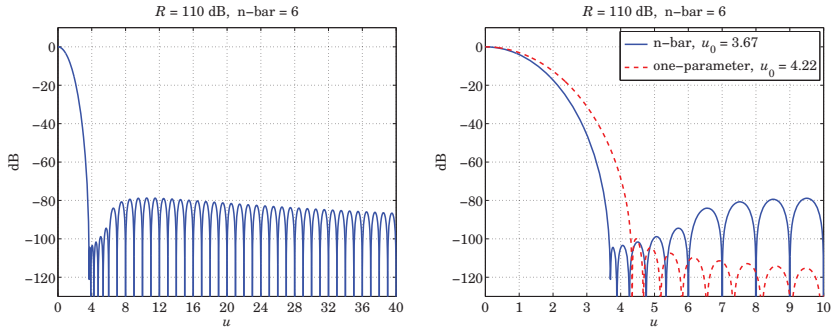


Fig. 20.15.2 Using a smaller $\bar{n}$.

20.16 Circularly Symmetric Apodization Functions

For a general two-dimensional planar aperture S the apodization function $A(\boldsymbol{r}_\perp)$ would be a tapered version of the indicator function $\chi_S(\boldsymbol{r}_\perp)$ of the aperture. The wavenumber pattern $\hat{A}(\boldsymbol{k}_\perp)$ would be the 2-D Fourier transform of $A(\boldsymbol{r}_\perp)$. For a circularly symmetric apodization function over a circular aperture of radius a, the Fourier relationship reduces to a Hankel transform. Thus, we have the Fourier pair in the general case,

$$\begin{aligned} \hat{A}(\boldsymbol{k}_\perp) &= \int_S A(\boldsymbol{r}_\perp)\, e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, d^2\boldsymbol{r}_\perp \\ A(\boldsymbol{r}_\perp) &= \int_{-\infty}^{\infty} \hat{A}(\boldsymbol{k}_\perp)\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, \frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \end{aligned} \tag{20.16.1}$$

In the circularly symmetric case, $A(\boldsymbol{r}_\perp), \hat{A}(\boldsymbol{k}_\perp)$, are functions of the radial coordinates, $r_\perp = |\boldsymbol{r}_\perp| = \sqrt{x^2+y^2}$, and, $k_\perp = |\boldsymbol{k}_\perp| = \sqrt{k_x^2+k_y^2}$. Then, Eqs. (20.16.1) can be written as the self-reciprocal order-0 Hankel transforms,[†]

$$\begin{aligned} \hat{A}(k_\perp) &= \int_0^a A(r_\perp)\, J_0(k_\perp r_\perp)\, 2\pi r_\perp dr_\perp \\ A(r_\perp) &= \int_0^\infty \hat{A}(k_\perp)\, J_0(k_\perp r_\perp)\, \frac{k_\perp dk_\perp}{2\pi} \end{aligned} \tag{20.16.2}$$

For two functions, $A(r_\perp), B(r_\perp)$, that are both aperture-limited in $0 \le r_\perp \le a$, and have Fourier/Hankel transforms, $\hat{A}(k_\perp), \hat{B}(k_\perp)$, we have the Parseval identity,

$$\int_0^a A^*(r_\perp) B(r_\perp)\, 2\pi r_\perp dr_\perp = \int_0^\infty \hat{A}^*(k_\perp) B(k_\perp)\, \frac{k_\perp dk_\perp}{2\pi} \tag{20.16.3}$$

Depending on the implementation, the wavenumber $k_\perp$ is proportional to the transverse displacement on the image plane, for example, $k_\perp = k\nu_\perp/F$, in a 2-D version of Fig. 20.12.1, or, in a far-field/radiation-field interpretation, since $k_x = k\sin\theta\cos\phi$,

[†] actually, the pair $A(r_\perp)$ and $\hat{A}(k_\perp)/2\pi$ are self-reciprocal.

$k_y = k\sin\theta\sin\phi$, we have, $k_\perp = \sqrt{k_x^2 + k_y^2} = k\sin\theta$. As in the 1-D case, it will be useful to think of $\hat{A}(k_\perp)$ as a function of the variable $u = ak_\perp/\pi$, which represents the angle direction in the standard beamwidth units of $\lambda/2a$,

$$\boxed{\sin\theta = \frac{k_\perp}{k} = u\,\frac{\lambda}{2a}} \tag{20.16.4}$$

As an example, for a uniform aperture, the function $A(r_\perp)$ is simply the indicator function, $A(r_\perp) = \chi_a(r_\perp) = u(a - r_\perp)$, and we obtain the usual Airy pattern (see also Sec. 18.9, where it is plotted versus angle θ),

$$\begin{aligned} A(r_\perp) &= \chi_a(r) = u(a - r_\perp) \\ \hat{A}(k_\perp) &= \pi a^2 \cdot \frac{2J_1(ak_\perp)}{ak_\perp} = \pi a^2 \cdot \frac{2J_1(\pi u)}{\pi u} \end{aligned} \tag{20.16.5}$$

The function, $2J_1(\pi u)/\pi u$, is the two-dimensional version of the sinc, $\sin(\pi u)/\pi u$, and is sometimes referred to as the "jinc" function. It is implemented by the MATLAB function, **jinc**, in the EWA toolbox, and is depicted in Fig. 20.16.1.

The first null is at $u_0 = 1.2197$, corresponding Rayleigh's resolution criterion, and the first sidelobe maximum is at $u_{\text{side}} = 1.6347$ with a value, $R_{\text{side}} = 0.1323$, or, in dB, $R_0 = -20\log_{10}(R_{\text{side}}) = 17.570150$ dB. The 3-dB point is at $u_3 = 0.5145$ and the full width, $\Delta u = 2u_3 = 1.0290$. These values can be determined easily by the MATLAB code,

```
u0 = fzero(@jinc,1);
us = fminbnd(@(u) -abs(jinc(u)), 1,2);
R0 = -20*log10(abs(jinc(us)));
u3 = fzero(@(u) abs(jinc(u))-1/sqrt(2), 1);
```

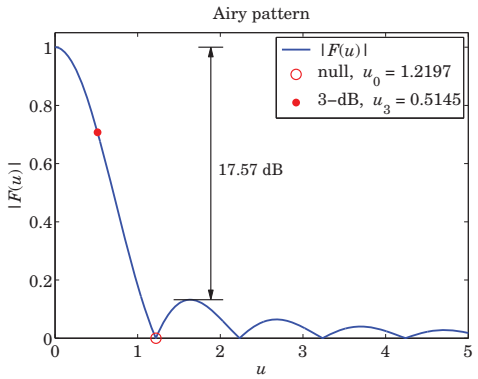


Fig. 20.16.1 The Airy pattern, $|F(u)| = |2J_1(\pi u)/\pi u|$.

The definitions of the various 1-D design parameters [1504] carry over to the 2-D case. The encircled energy is defined relative to an interval, $0 \le k_\perp \le k_0 = \pi u_0/a$,

$$\mathcal{E}(k_0) = \frac{\int_0^{k_0} |\hat{A}(k_\perp)|^2\, k_\perp dk_\perp}{\int_0^{\infty} |\hat{A}(k_\perp)|^2\, k_\perp dk_\perp} = \frac{\int_0^{u_0} |F(u)|^2\, u\, du}{(2a)^2 \int_0^a |A(r_\perp)|^2\, r_\perp dr_\perp} \tag{20.16.6}$$

The Strehl ratio and transmission factor, normalized relative to the corresponding values of the uniform aperture, $A_{\text{unif}}(r_\perp) = u(a - r_\perp)$, are given by,

$$\begin{aligned} S &= \frac{|\hat{A}(0)|^2}{|\hat{A}_{\text{unif}}(0)|^2} = \left| \frac{1}{\pi a^2} \int_0^a A(r_\perp)\, 2\pi r_\perp dr_\perp \right|^2 \\ \tau &= \frac{\int_0^a |A(r_\perp)|^2\, 2\pi r_\perp dr_\perp}{\int_0^a |A_{\text{unif}}(r_\perp)|^2\, 2\pi r_\perp dr_\perp} = \frac{1}{\pi a^2} \int_0^a |A(r_\perp)|^2\, 2\pi r_\perp dr_\perp \end{aligned} \tag{20.16.7}$$

Similarly, the two-dimensional limiting directivity of Eq. (19.7.11), normalized to its uniform value, $D_{\text{unif}} = 4\pi(\pi a^2)/\lambda^2 = (ka)^2$, is given by,

$$\frac{D_\infty}{(ka)^2} = \frac{|\hat{A}(0)|^2}{\pi a^2 \int_0^\infty |\hat{A}(k_\perp)|^2 \frac{k_\perp dk_\perp}{2\pi}} = \frac{\left| \int_0^a A(r_\perp)\, 2\pi r_\perp dr_\perp \right|^2}{\pi a^2 \int_0^a |A(r_\perp)|^2\, 2\pi r_\perp dr_\perp} \tag{20.16.8}$$

There is a large literature on circularly symmetric apodization functions, reviewed in [1504]. One widely used family of functions in both optics and aperture antennas, such as dish antennas, are the so-called lambda functions or Sonine functions, defined as follows for integer n, or real-valued n, such that $n > -1$,

$$\begin{aligned} A(r_\perp) &= \left(1 - \frac{r_\perp^2}{a^2}\right)^n \cdot \chi_a(r_\perp) \\ \hat{A}(k_\perp) &= \pi a^2 \cdot \frac{\Gamma(n+1)\, 2^{n+1} J_{n+1}(k_\perp a)}{(k_\perp a)^{n+1}} \end{aligned} \tag{20.16.9}$$

where $A(r_\perp)$ and $\hat{A}(k_\perp)$ are Hankel pairs as in (20.16.2). The case $n = 0$ corresponds to the usual Airy pattern. The cases $n = 1, 2$ are used in dish antennas. Linear combinations of such functions have also been used to represent more general apodization functions [1502]. However, we do not pursue these further since we are more interested in designs with narrow beams and low sidelobes.

Among the latter are the *generalized* prolate spheroidal wave functions derived by Slepian [1505,1646], which are the 2-D versions of the 1-D prolates that we discussed above. Hansen's one-parameter window discussed in the next section provides a good approximation to the generalized prolates, just like Taylor's one-parameter window is a good approximation to the optimum 1-D case. Besides Hansen's one-parameter window, we also discuss below the Fourier-Bessel and Dini series representations of circular apodization functions, as well as the 2-D version of Taylor's n-bar window.

20.17 Hansen One-Parameter Window

The generalization of Taylor's one-parameter window to two dimensions was derived by Hansen [1606]. It is based on the following Fourier/Hankel transform pair,† which is the 2-D version of the 1-D pair of Eq. (20.14.2),

$$\begin{aligned} A(r_\perp) &= I_0\left(k_0\sqrt{a^2-r_\perp^2}\right)\cdot\chi_a(r_\perp) \\ \hat{A}(k_\perp) &= \pi a^2\cdot\frac{2I_1\left(a\sqrt{k_0^2-k_\perp^2}\right)}{a\sqrt{k_0^2-k_\perp^2}} = \pi a^2\cdot\frac{2J_1\left(a\sqrt{k_\perp^2-k_0^2}\right)}{a\sqrt{k_\perp^2-k_0^2}} \end{aligned} \tag{20.17.1}$$

where k_0 is a positive parameter that defines the extent of the mainlobe. Introducing the parameter H through, $\pi H = ak_0$, we may rewrite the above in u-units,

$$\begin{aligned} A(r_\perp) &= I_0\left(\pi H\sqrt{1-(r_\perp/a)^2}\right)\cdot\chi_a(r_\perp) \\ F(u) &= \pi a^2\cdot\frac{2I_1\left(\pi\sqrt{H^2-u^2}\right)}{\pi\sqrt{H^2-u^2}} = \pi a^2\cdot\frac{2J_1\left(\pi\sqrt{u^2-H^2}\right)}{\pi\sqrt{u^2-H^2}} \end{aligned} \tag{20.17.2}$$

where $F(u) = \hat{A}(k_\perp)$ at $k_\perp = \pi u/a$. At $u = H$, the function $F(u)$ switches from I_1 to J_1, and for large u it decays like a sinc, $\sin(\pi u)/\pi u$,

The first null occurs at $\sqrt{u_0^2-H^2} = \mu_1$, or, $u_0 = \sqrt{H^2+\mu_1^2}$, where $\mu_1 = 1.2197$ is the first null of $2J_1(\pi u)/\pi u$, or, the first non-zero null of J_1, that is, $J_1(\pi\mu_1) = 0$. The sidelobe structure begins after that. As we saw above, the highest sidelobe of the Airy pattern is $R_0 = 17.57$ dB below the main peak, therefore, the analogous relationship to (20.13.3) becomes,

$$R = R_0 + 20\log_{10}\left(\frac{2I_1(\pi H)}{\pi H}\right) \tag{20.17.3}$$

The 3-dB width in u-units is given by $\Delta u = 2u_3$, where u_3 is the solution of the half-power condition, $|F(u_3)/F(0)|^2 = 1/2$,

$$\frac{J_1\left(\pi\sqrt{u_3^2-H^2}\right)}{\pi\sqrt{u_3^2-H^2}} = \frac{1}{\sqrt{2}}\,\frac{I_1(\pi H)}{\pi H} \tag{20.17.4}$$

The beam efficiency/encircled energy and the normalized directivity given by (20.16.6) and (20.16.8), can be written in closed form,

$$\begin{aligned} \mathcal{E}(u_0) &= \frac{\int_0^{u_0}|F(u)|^2\,u\,du}{(2a)^2\int_0^a|A(r_\perp)|^2 r_\perp dr_\perp} = 1-\frac{J_0^2\left(\pi\sqrt{u_0^2-H^2}\right)+J_1^2\left(\pi\sqrt{u_0^2-H^2}\right)}{I_0^2(\pi H)-I_1^2(\pi H)} \\ \frac{D_\infty}{(ka)^2} &= \frac{\left|\int_0^a A(r_\perp)\,2\pi r_\perp dr_\perp\right|^2}{\pi a^2\int_0^a|A(r_\perp)|^2\,2\pi r_\perp dr_\perp} = \frac{\left[\frac{2I_1(\pi H)}{\pi H}\right]^2}{I_0^2(\pi H)-I_1^2(\pi H)} \end{aligned} \tag{20.17.5}$$

† see Gradshtein [1791], entry 6.596.6.

where the expression for $\mathcal{E}(u_0)$ is valid for both $u_0 \geq H$ and $u_0 < H$, where in the latter case, we have,

$$J_0^2\left(\pi\sqrt{u_0^2-H^2}\right)+J_1^2\left(\pi\sqrt{u_0^2-H^2}\right) = I_0^2\left(\pi\sqrt{H^2-u_0^2}\right)-I_1^2\left(\pi\sqrt{H^2-u_0^2}\right)$$

Example 20.17.1: Fig. 20.17.1 plots the functions $|F(u)|$ and $A(r_\perp)$ for the two cases of $R = 60$ and $R = 100$ dB, and demonstrates how to compute H and the 3-dB width Δu from R. The MATLAB code below illustrates the computation,

```
us = fminbnd(@(u) -abs(jinc(u)), 1,2);
R0 = -20*log10(abs(jinc(us)));              % find R0 = 17.57 dB

f = @(u,H) abs(jinc(sqrt(u.^2-H.^2)));

G = @(u,H) real(besseli(0,pi*sqrt(H^2-u^2))^2 - besseli(1,pi*sqrt(H^2-u^2))^2);

r = linspace(0,1,201);      % units of a
u = linspace(0,20,2001);

for R = [60,100]

   H0 = 0.04*R + 1.07;
   H = fzero(@(H) R0 + 20*log10(f(0,H)) - R, H0);

   u30 = 0.031*R + 0.46;
   Du = 2*fzero(@(u) f(u,H)-f(0,H)/sqrt(2), u30);    % 3-dB width
   m0 = fzero(@(u) besselj(1,pi*u),1.25);
   u0 = sqrt(H^2 + m0^2);                            % first null

   E0 = 1 - G(u0,H)/G(0,H);                    % encircled to u0
   D = f(0,H)^2 / G(0,H);                      % normalized directivity

   F = 20*log10(f(u,H) / f(0,H));              % normalized to unity at u=0

   A = besseli(0,pi*H*sqrt(1-r.^2)) / besseli(0,pi*H);   % normalized to A(0)=1

   figure; plot(u,F,'b-');
   figure; plot(r,A,'b-');

end
```

Because of the log vertical scales, the first-null $u_0 = \sqrt{H^2+\mu_1^2}$ was placed on each graph at the level of $-R$ dB. In the limit $u_0 \to \infty$, $\mathcal{E}(u_0)$ in (20.17.5) tends to unity as it should. However, for large R, most of the energy resides in the mainlobe and we expect $\mathcal{E}(u_0)$ to be practically unity. The computed values are shown below,

R	H	Δu	D	u_0	E_0
60	2.6548	1.6669	0.4209	2.9216	0.999996097701
100	4.3503	2.0611	0.2710	4.5180	0.999999999749

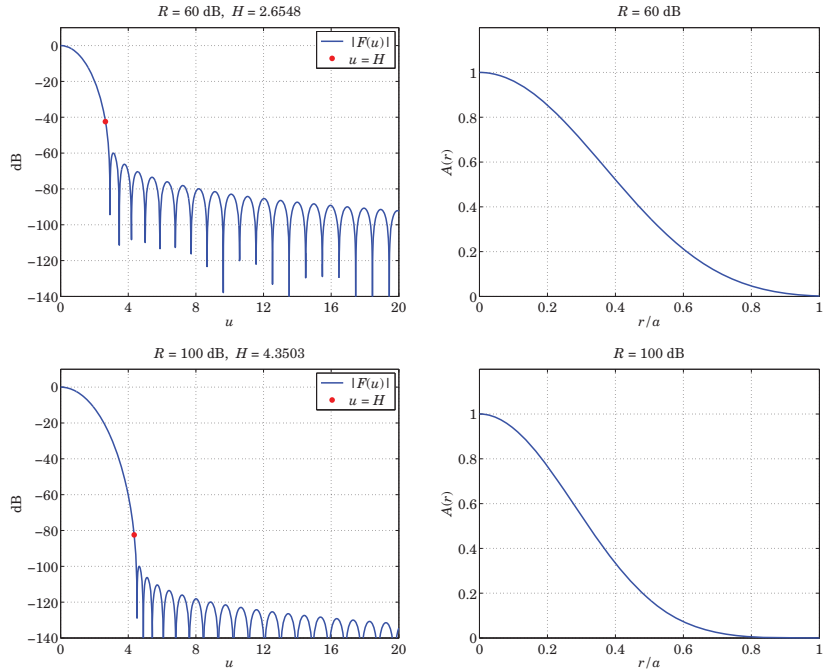


Fig. 20.17.1 Hansen window, $|\hat{A}(k_\perp)|$ in dB versus $u = ak_\perp/\pi$, and $A(r_\perp)$ versus $r_\perp/a$.

20.18 Fourier-Bessel and Dini Series Expansions

In the two-dimensional circularly-symmetric case, the role of the Woodward-Lawson sinc-basis expansions of Eq. (20.12.25), is played by the Fourier-Bessel and Dini series expansions [1641,1642].

The Fourier-Bessel expansion is with respect to the zeros of $J_0(z)$, whereas the Dini expansion uses the zeros of $J_1(z)$. The following Lommel-type integrals [1641,1642] can be used to derive all the results in this section,

$$\begin{aligned}\frac{1}{\pi a^2}\int_0^a J_0\Big(\alpha\frac{\rho}{a}\Big)J_0\Big(\beta\frac{\rho}{a}\Big)2\pi\rho\,d\rho &= \frac{2\alpha J_1(\alpha)J_0(\beta)-2\beta J_1(\beta)J_0(\alpha)}{\alpha^2-\beta^2}\\ \frac{1}{\pi a^2}\int_0^a J_0^2\Big(\alpha\frac{\rho}{a}\Big)2\pi\rho\,d\rho &= J_0^2(\alpha)+J_1^2(\alpha)\end{aligned} \tag{20.18.1}$$

In particular, if we set $\alpha = \alpha_n$ equal to the nth root of $J_0(z)$, that is, $J_0(\alpha_n) = 0$, then Eqs. (20.18.1) reduce to a form suitable for the Fourier-Bessel expansions,

$$\begin{aligned}\frac{1}{\pi a^2}\int_0^a J_0\Big(\alpha_n\frac{\rho}{a}\Big)J_0\Big(\beta\frac{\rho}{a}\Big)2\pi\rho\,d\rho &= \frac{2\alpha_n J_1(\alpha_n)J_0(\beta)}{\alpha_n^2-\beta^2}\\ \frac{1}{\pi a^2}\int_0^a J_0^2\Big(\alpha_n\frac{\rho}{a}\Big)2\pi\rho\,d\rho &= J_1^2(\alpha_n)\end{aligned} \tag{20.18.2}$$

Similarly, if α_n is a zero of $J_1(z)$, that is, $J_1(\alpha_n) = 0$, then,

$$\begin{aligned}\frac{1}{\pi a^2}\int_0^a J_0\Big(\alpha_n\frac{\rho}{a}\Big)J_0\Big(\beta\frac{\rho}{a}\Big)2\pi\rho\,d\rho &= \frac{2\beta J_1(\beta)J_0(\alpha_n)}{\beta^2-\alpha_n^2}\\ \frac{1}{\pi a^2}\int_0^a J_0^2\Big(\alpha_n\frac{\rho}{a}\Big)2\pi\rho\,d\rho &= J_0^2(\alpha_n)\end{aligned} \tag{20.18.3}$$

It is straightforward to show that if β tends to a zero, $\beta \to \alpha_m$, then, the right-hand sides of the top equations in (20.18.2) or (20.18.3) become zero if $\alpha_m \neq \alpha_n$, and they reduce to the bottom equations if $\alpha_m = \alpha_n$.

The required roots of $J_0(z)$ and $J_1(z)$ can be computed easily by MATLAB's built-in function, **fzero**. The table below shows the first 10 roots α_n, as well as the values, α_n/π,

n	J0, a(n)	J1, a(n)	n	a(n)/pi	a(n)/pi
0	-	0	0	-	0
1	2.4048	3.8317	1	0.7655	1.2197
2	5.5201	7.0156	2	1.7571	2.2331
3	8.6537	10.1735	3	2.7546	3.2383
4	11.7915	13.3237	4	3.7534	4.2411
5	14.9309	16.4706	5	4.7527	5.2428
6	18.0711	19.6159	6	5.7522	6.2439
7	21.2116	22.7601	7	6.7519	7.2448
8	24.3525	25.9037	8	7.7516	8.2454
9	27.4935	29.0468	9	8.7514	9.2459
10	30.6346	32.1897	10	9.7513	10.2463

Note that $J_1(z)$ has an extra zero at $z = 0$. The table was generated by the following MATLAB code, which can easily be modified to compute as many roots as desired,

```
for n=1:10,

    a0(n) = fzero(@(z) besselj(0,z), pi*(n-1/4));         % zeros of J0
    a1(n) = fzero(@(z) besselj(1,z), pi*(n+1/4));         % zeros of J1

    fprintf('%2d  |  %7.4f  |  %7.4f', n, a0(n),a1(n));
    fprintf('    %2d  |  %7.4f  |  %7.4f\n', n, a0(n)/pi,a1(n)/pi);

end
```

For large n, the zeros are given approximately by,

$$\begin{aligned}\alpha_n &= \pi\Big(n-\frac{1}{4}\Big) \qquad \text{(for } J_0\text{)}\\ \alpha_n &= \pi\Big(n+\frac{1}{4}\Big) \qquad \text{(for } J_1\text{)}\end{aligned} \tag{20.18.4}$$

which also serve to initialize the search argument of **fzero**. This trend is already evident in the above table.

Fourier-Bessel Series

Using (20.18.2) and reverting to our original transverse space and wavenumber variables, $r_\perp, k_\perp$, we define the following aperture-limited and mutually orthogonal basis functions and their Hankel transforms, for $n = 1, 2, \ldots$,

$$\begin{aligned} \Phi_n(r_\perp) &= \frac{J_0(k_n r_\perp)}{\pi a^2 J_1^2(k_n a)} \cdot \chi_a(r_\perp) \\ \hat{\Phi}_n(k_\perp) &= \int_0^a \Phi_n(r_\perp)\, J_0(k_\perp r_\perp)\, 2\pi r_\perp dr_\perp = \frac{2k_n J_0(k_\perp a)}{a J_1(k_n a)\,(k_n^2 - k_\perp^2)} \end{aligned} \tag{20.18.5}$$

where k_n are related to the roots of $J_0(z)$ via, $k_n a = \alpha_n$, so that,

$$J_0(k_n a) = J_0(\alpha_n) = 0\,, \quad n = 1, 2, \ldots \tag{20.18.6}$$

The scale factors have been chosen so that the Hankel transforms satisfy the following property, which is equivalent to the mutual orthogonality of $\Phi_n(r_\perp)$,

$$\hat{\Phi}_n(k_m) = \delta_{nm}\,, \quad n, m = 1, 2, \ldots, \tag{20.18.7}$$

This can be proven by taking the limit $k_\perp \to k_m$ in (20.18.5), and using the Taylor expansion, $J_0(k_\perp a) \approx J_0'(k_m a)\,(k_\perp a - k_m a) = -J_1(k_m a)\,(k_\perp - k_m)\,a$. The orthogonality property follows by noting that,

$$\begin{aligned} \int_0^a \Phi_n(r_\perp)\Phi_m(r_\perp)\, 2\pi r_\perp dr_\perp &= \frac{1}{\pi a^2 J_1^2(k_m a)} \int_0^a \Phi_n(r_\perp)\, J_0(k_m r_\perp)\, 2\pi r_\perp dr_\perp \\ &= \frac{\hat{\Phi}_n(k_m)}{\pi a^2 J_1^2(k_m a)} = \frac{\delta_{nm}}{\pi a^2 J_1^2(k_m a)} \end{aligned} \tag{20.18.8}$$

The Fourier-Bessel series is an expansion of an aperture-limited apodization function $A(r_\perp)$ in the $\Phi_n(r_\perp)$ basis, which also implies the corresponding expansion of its Hankel transform $\hat{A}(k_\perp)$ in the $\hat{\Phi}_n(k_\perp)$ basis,

$$\boxed{\begin{aligned} A(r_\perp) &= \sum_{n=1}^{\infty} c_n \Phi_n(r_\perp) \\ \hat{A}(k_\perp) &= \sum_{n=1}^{\infty} c_n \hat{\Phi}_n(k_\perp) \end{aligned}} \quad \text{(Fourier-Bessel series)} \tag{20.18.9}$$

The reason for the normalization choice (20.18.7) now becomes apparent: it immediately implies that the expansion coefficients are the *sampled values* of the Hankel transform at the Bessel function zeros, that is,

$$\boxed{c_n = \hat{A}(k_n) = \int_0^a A(r_\perp)\, J_0(k_n r_\perp)\, 2\pi r_\perp dr_\perp}\,, \quad n = 1, 2, \ldots \tag{20.18.10}$$

Dini Series

Similar expansions can be written for the Dini series, which uses the zeros of $J_1(z)$. Now, the zero-wavenumbers k_n must be defined by $k_n a = \alpha_n$, which also include the zero case, $k_0 a = \alpha_0 = 0$, so that,

$$J_1(k_n a) = J_1(\alpha_n) = 0\,, \quad n = 0, 1, 2, \ldots \tag{20.18.11}$$

Using (20.18.3), we define the Dini basis functions and their Hankel transforms by,

$$\begin{aligned} \Phi_n(r_\perp) &= \frac{J_0(k_n r_\perp)}{\pi a^2 J_0^2(k_n a)} \cdot \chi_a(r_\perp) \\ \hat{\Phi}_n(k_\perp) &= \int_0^a \Phi_n(r_\perp)\, J_0(k_\perp r_\perp)\, 2\pi r_\perp dr_\perp = \frac{2k_\perp J_1(k_\perp a)}{a J_0(k_n a)\,(k_\perp^2 - k_n^2)} \end{aligned} \tag{20.18.12}$$

Again, the normalization factors have been chosen so that,

$$\hat{\Phi}_n(k_m) = \delta_{nm}\,, \quad n, m = 1, 2, \ldots, \tag{20.18.13}$$

from which the mutual orthogonality property follows, as in (20.18.8),

$$\int_0^a \Phi_n(r_\perp)\Phi_m(r_\perp)\, 2\pi r_\perp dr_\perp = \frac{\delta_{nm}}{\pi a^2 J_0^2(k_n a)}\,, \quad n, m = 0, 1, 2, \ldots \tag{20.18.14}$$

The function $\hat{\Phi}_n(k_\perp)$ may be rewritten as follows, and can be evaluated at any vector of $k_\perp$ and any Bessel root $\alpha_n = k_n a$ with the MATLAB function **jinc**,

$$\hat{\Phi}_n(k_\perp) = \frac{2k_\perp J_1(k_\perp a)}{a J_0(k_n a)\,(k_\perp^2 - k_n^2)} = \frac{2J_1(k_\perp a)}{k_\perp a} \cdot \frac{k_\perp^2}{k_\perp^2 - k_n^2} \cdot \frac{1}{J_0(k_n a)} \tag{20.18.15}$$

The $n = 0$ case, reduces to the usual Airy pattern, with $\hat{\Phi}_0(0) = 1$,

$$\hat{\Phi}_0(k_\perp) = \frac{2J_1(k_\perp a)}{k_\perp a}$$

The Dini series expansion of an apodization function and its Hankel transform will be,

$$\boxed{\begin{aligned} A(r_\perp) &= \sum_{n=0}^{\infty} c_n \Phi_n(r_\perp) \\ \hat{A}(k_\perp) &= \sum_{n=0}^{\infty} c_n \hat{\Phi}_n(k_\perp) \end{aligned}} \quad \text{(Dini series)} \tag{20.18.16}$$

where, again, c_n are the sampled values of the Hankel transform,

$$\boxed{c_n = \hat{A}(k_n) = \int_0^a A(r_\perp)\, J_0(k_n r_\perp)\, 2\pi r_\perp dr_\perp}\,, \quad n = 0, 1, 2, \ldots \tag{20.18.17}$$

The Dini series is the 2-D version of the Woodward-Lawson sampling expansion in the 1-D sinc basis. To compare the two, we write (20.18.16) explicitly, and also rewrite Eq. (20.12.25) by defining $k_n = n\pi/a$ in that case,

$$\hat{A}(k_\perp) = \sum_{n=0}^{\infty} \hat{A}(k_n)\,\frac{2J_1(k_\perp a)}{k_\perp a}\cdot\frac{k_\perp^2}{k_\perp^2 - k_n^2}\cdot\frac{1}{J_0(k_n a)}$$
$$\hat{A}(k_x) = \sum_{n=-\infty}^{\infty} \hat{A}(k_n)\,\frac{\sin(k_x a - k_n a)}{a(k_x - k_n)} \tag{20.18.18}$$

Noting that $\sin(k_x a - k_n a) = \sin(k_x a - \pi n) = (-1)^n \sin(k_x a)$, and folding the sum over n into non-negative values of n, assuming that $\hat{A}(k_x)$ is even in k_x, we can rewrite the Dini and Woodward-Lawson cases in a very similar way,

$$\hat{A}(k_\perp) = \hat{A}(0)\,\frac{2J_1(k_\perp a)}{k_\perp a} + \sum_{n=1}^{\infty} \hat{A}(k_n)\,\frac{2J_1(k_\perp a)}{k_\perp a}\cdot\frac{k_\perp^2}{k_\perp^2 - k_n^2}\cdot\frac{1}{J_0(k_n a)}$$
$$\hat{A}(k_x) = \hat{A}(0)\,\frac{\sin(k_x a)}{k_x a} + \sum_{n=1}^{\infty} \hat{A}(k_n)\,\frac{2\sin(k_x a)}{k_x a}\cdot\frac{k_x^2}{k_x^2 - k_n^2}\cdot(-1)^n \tag{20.18.19}$$

where we note that the factor $J_0(k_n a)$ alternates in sign exactly like $(-1)^n$.

The expansions (20.18.9) and (20.18.16) are useful only if the number of terms is finite. In this respect, Dini's series is more useful in practice. Indeed, a finite number of terms is needed in Taylor's n-bar window discussed in the next section. A finite Dini series has also been used in the design of super-resolving pupils [1497,1498].

20.19 Taylor's Two-Dimensional N-bar Window

Taylor's two-dimensional generalization of his n-bar window [1603] is implemented with the help of the Dini series. As in the 1-D case, it proves convenient to work in u-units, defined as follows,

$$u = \frac{k_\perp a}{\pi}, \qquad \mu_n = \frac{k_n a}{\pi} = \frac{\alpha_n}{\pi}, \quad n = 0, 1, 2, \ldots \tag{20.19.1}$$

where α_n are the zeros of $J_1(z)$. Defining the aperture pattern by $F(u) = \hat{A}(k_\perp)$ at $k_\perp = \pi u/a$, then, $F(\mu_n) = \hat{A}(k_n)$, and the Dini series expansions (20.18.16) will read,

$$A(r_\perp) = \sum_{n=0}^{\infty} F(\mu_n)\,\frac{J_0(\pi\mu_n r_\perp/a)}{\pi a^2 J_0^2(\pi\mu_n)}\cdot\chi_a(r_\perp)$$
$$F(u) = \sum_{n=0}^{\infty} F(\mu_n)\,\frac{2J_1(\pi u)}{\pi u}\cdot\frac{u^2}{u^2 - \mu_n^2}\cdot\frac{1}{J_0(\pi\mu_n)} \tag{20.19.2}$$

As in the 1-D case, the function $F(u)$ of the n-bar distribution is built out of $\bar{n} - 1$ zeros of the ideal, slightly dilated, Chebyshev pattern, with the rest of the zeros taken from the Airy jinc-function pattern, whose zeros are the μ_n for $n > 0$. Thus, the 2-D version of Eq. (20.15.6) will be,

$$\cos\left(\pi\sqrt{(u/\sigma)^2 - A^2}\right) = \cosh(\pi A)\cdot\prod_{n=1}^{\infty}\left[1 - \frac{u^2}{\sigma^2[A^2 + (n - \frac{1}{2})^2]}\right]$$
$$\text{jinc}(u) = \frac{2J_1(\pi u)}{\pi u} = \prod_{n=1}^{\infty}\left[1 - \frac{u^2}{\mu_n^2}\right] \tag{20.19.3}$$

and $F(u)$ is constructed as,

$$F(u) = \cosh(\pi A)\cdot\prod_{n=1}^{\bar{n}-1}\left[1 - \frac{u^2}{\sigma^2[A^2 + (n - \frac{1}{2})^2]}\right]\cdot\prod_{n=\bar{n}}^{\infty}\left[1 - \frac{u^2}{\mu_n^2}\right] \tag{20.19.4}$$

Since the value at $u = 0$ is, $F(0) = \cosh(\pi A)$, it follows that the sidelobe attenuation R (in dB) will be given in terms of the parameter A by the same equation as (20.15.3),

$$R = 20\log_{10}[\cosh(\pi A)] \quad\Leftrightarrow\quad A = \frac{1}{\pi}\,\text{acosh}(10^{R/20}) \tag{20.19.5}$$

As in the 1-D case, the dilation factor σ is chosen such that two groups of zeros match at $n = \bar{n}$, that is,

$$\sigma\sqrt{A^2 + (\bar{n} - \tfrac{1}{2})^2} = \mu_{\bar{n}} \quad\Rightarrow\quad \sigma = \frac{\mu_{\bar{n}}}{A^2 + (\bar{n} - \frac{1}{2})^2} \tag{20.19.6}$$

Since $F(u)$ vanishes by construction at all Bessel zeros μ_n with $n \geq \bar{n}$, that is, $F(\mu_n) = 0$, for $n \geq \bar{n}$, it follows that the Dini representation (20.19.2) will consist of a finite sum only over the surviving samples, $F(\mu_n)$, for $0 \leq n \leq \bar{n} - 1$,

$$\boxed{\begin{aligned} A(r_\perp) &= \sum_{n=0}^{\bar{n}-1} F(\mu_n)\,\frac{J_0(\pi\mu_n r_\perp/a)}{\pi a^2 J_0^2(\pi\mu_n)}\cdot\chi_a(r_\perp) \\ F(u) &= \sum_{n=0}^{\bar{n}-1} F(\mu_n)\,\frac{2J_1(\pi u)}{\pi u}\cdot\frac{u^2}{u^2 - \mu_n^2}\cdot\frac{1}{J_0(\pi\mu_n)} \end{aligned}} \tag{20.19.7}$$

In order to calculate the required values of $F(\mu_n)$, we follow the same procedure as in Eq. (20.15.16) and rewrite (20.19.4) in the pole/zero cancellation form,

$$F(u) = \cosh(\pi A)\cdot\frac{2J_1(\pi u)}{\pi u}\cdot\frac{\prod_{n=1}^{\bar{n}-1}\left[1 - \frac{u^2}{\sigma^2[A^2 + (n - \frac{1}{2})^2]}\right]}{\prod_{n=1}^{\bar{n}-1}\left[1 - \frac{u^2}{\mu_n^2}\right]} \tag{20.19.8}$$

By taking the limit $u \to \mu_n$ and allowing for such cancellations,† we obtain the following expression [1603] for $F(\mu_n)$, for $1 \leq n \leq \bar{n} - 1$,

$$F(\mu_n) = -\cosh(\pi A)\cdot\frac{J_0(\pi\mu_n)\prod_{m=1}^{\bar{n}-1}\left[1 - \frac{\mu_n^2}{\sigma^2[A^2 + (m - \frac{1}{2})^2]}\right]}{\prod_{\substack{m=1\\ m\neq n}}^{\bar{n}-1}\left[1 - \frac{\mu_n^2}{\mu_m^2}\right]} \tag{20.19.9}$$

and $F(0) = 1$. The 3-dB width can be calculated by (20.15.13), that is,

$$\Delta u = 2\sigma\sqrt{A^2 - \frac{1}{\pi^2}\,\text{acosh}^2\left(\frac{1}{\sqrt{2}}\cosh(\pi A)\right)} \tag{20.19.10}$$

† using the Taylor expansion around the μ_n zero, $J_1(\pi u) \approx J_1'(\pi\mu_n)\,\pi(u - \mu_n) = J_0(\pi\mu_n)\,\pi(u - \mu_n)$.

The encircled energy/beam gain can be defined by (20.16.6), relative to the first-null interval $[0, u_0]$, where, $u_0 = \sigma\sqrt{A^2 + \frac{1}{4}}$,

$$\mathcal{E}(u_0) = \frac{\int_0^{u_0} |F(u)|^2\, u\, du}{(2a)^2 \int_0^a |A(r_\perp)|^2 r_\perp dr_\perp} = \frac{\frac{\pi^2}{2}\int_0^{u_0} |F(u)|^2\, u\, du}{\pi a^2 \int_0^a |A(r_\perp)|^2 2\pi r_\perp dr_\perp} \tag{20.19.11}$$

The numerator can be evaluated numerically, while the denominator can be expressed as a (finite, in this case) sum over the pattern sampled values $\hat{A}(k_n)$, as follows from the Dini series expansion and orthogonality property (20.18.14),

$$\pi a^2 \int_0^a |A(r_\perp)|^2 2\pi r_\perp dr_\perp = \sum_{n=0}^{\bar{n}-1} \left| \frac{\hat{A}(k_n)}{J_0(k_n a)} \right|^2 \tag{20.19.12}$$

Similarly, the two-dimensional limiting directivity, normalized to its uniform value, $D_{\text{unif}} = 4\pi(\pi a^2)/\lambda^2 = (ka)^2$, is given by Eq. (20.16.8),

$$\frac{D_\infty}{(ka)^2} = \frac{|\hat{A}(0)|^2}{\pi a^2 \int_0^a |A(r_\perp)|^2 2\pi r_\perp dr_\perp} = \frac{|\hat{A}(0)|^2}{\sum_{n=0}^{\bar{n}-1} \left| \frac{\hat{A}(k_n)}{J_0(k_n a)} \right|^2} \tag{20.19.13}$$

Given $R, \bar{n}$, the MATLAB function, **tnb2**, calculates the parameters $A, \Delta u, D_\infty, E_0$, and the coefficient and Bessel zero vectors, $F(\mu_n), \mu_n$, $n = 0, 1, \ldots, \bar{n}-1$. It has usage,

```
[A,Fn,mu,Du,D,E] = tnb2(R,nb);    % Taylor's n-bar window (2-D)
```

Example 20.19.1: This example demonstrates how to use **tnb2** to calculate $F(u)$ and $A(r_\perp)$ in a vectorized fashion. Fig. 20.19.1 plots the functions $|F(u)|$ and $A(r_\perp)$ for the two cases of $R = 60$ dB, $\bar{n} = 10$, and, $R = 100$, $\bar{n} = 30$. The MATLAB code below illustrates the computation,

```
R = 100; nb = 30;

[Ap,Fn,mu,Du,D,E] = tnb2(R,nb);

r = linspace(0,1,201);                % units of a
u = linspace(0,20,2001);

F = 0; A = 0;
for n=1:nb,                           % calculate F(u) and A(r)
    F = F + Fn(n) * jinc(u,mu(n));
    A = A + Fn(n) * besselj(0,pi*mu(n)*r)/besselj(0,pi*mu(n))^2/pi;
end

sigma = mu(nb+1) / sqrt(Ap^2 + (nb-0.5)^2);
u0 = sigma * sqrt(Ap^2 + 1/4);              % first null

A = A/max(A); F = abs(F); Fdb = 20*log10(F/max(F));

figure; plot(u,Fdb,'b-', u0,-R,'r.');
figure; plot(r,A,'b-');
```

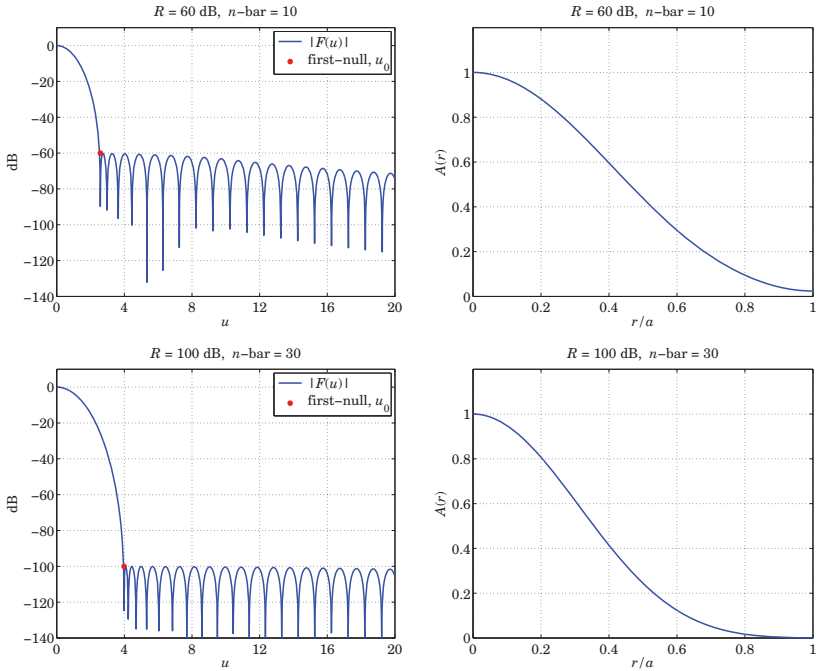



Fig. 20.19.1 Taylor's n-bar window, $|F(u)|$ in dB versus u, and $A(r_\perp)$ versus $r_\perp/a$.

As in Example 20.17.1, because of the log vertical scales, the first-null u_0 was placed on each graph at the level of $-R$ dB. Because R is large, most of the energy resides in the mainlobe and $\mathcal{E}(u_0)$ is practically unity. The computed values are shown below,

R	A	Δu	D	u_0	E_0
60	2.4194	1.5098	0.5180	2.5822	0.999887111028
100	3.8853	1.8691	0.3310	3.9824	0.999999948865

The results of this example can be compared directly with those of Example 20.17.1 on the Hansen window with the same specifications. One observes that the n-bar design produces a narrower mainlobe because the sidelobes remain almost equiripple for a while. □

20.20 Star-Shaped Masks, Starshade Occulters

The realization and manufacture of a smoothly varying circularly symmetric apodization function $A(r_\perp)$ is a challenging problem, requiring special materials and transparencies with precise tolerances to achieve the design objectives. The realization would be more straightforward if $A(r_\perp)$ were binary, consisting simply of either opaque $(A = 0)$ or fully transparent $(A = 1)$, open, parts.

In an interesting paper, Vanderbei, Spergel, and Kasdin [1531] have shown that it is possible to replace *any*[†] smooth circularly symmetric apodizer $A(r_\perp)$ with an equivalent binary mask consisting of a number of star-shaped vanes, as shown in Fig. 20.20.1.

For a given number of vanes N, the equivalence is valid over a finite wavenumber range, $0 \le k_\perp \le k_{\max}$, with the outer limit $k_{\max}$ increasing with N. The shape of the vanes depends on the shape of $A(r_\perp)$ and the vanes converge to a tip at the edge of the star-shaped aperture provided $A(r_\perp)$ is a decreasing function of $r_\perp$, as is typically (but not always) the case.

To see this equivalence, consider a set of N opaque vanes, with the surface openings between the vanes defined as the following set of cylindrical coordinates $(r_\perp, \phi)$,

$$\begin{aligned} S &= \{(r_\perp,\phi) \;\big|\; 0 \le r_\perp \le a\,, \quad \phi \in \Phi(r_\perp)\} \\ \Phi(r_\perp) &= \bigcup_{n=0}^{N-1} \left[\frac{\pi(2n+1)}{N} - \frac{\pi}{N} A(r_\perp)\,, \;\; \frac{\pi(2n+1)}{N} + \frac{\pi}{N} A(r_\perp) \right] \end{aligned} \tag{20.20.1}$$

where $A(r_\perp)$ is the equivalent smooth apodization function to be realized by the star-shaped mask, and $[\alpha,\beta]$ denotes the angle interval, $\alpha \le \phi \le \beta$. Using cylindrical coordinates for the transverse wavenumber $\boldsymbol{k}_\perp$ as defined in Eq. (20.1.10), it follows that the Fourier pattern (20.16.1) over the open aperture S, will be,

$$\hat{A}(k_\perp,\psi) = \int_S e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, d^2\boldsymbol{r}_\perp = \int_S e^{jk_\perp r_\perp \cos(\psi-\phi)}\, r_\perp\, dr_\perp\, d\phi \tag{20.20.2}$$

The exponential may be replaced by a sum over Bessel functions using the identity of Eq. (19.12.24), that is,

$$e^{jk_\perp r_\perp\cos(\psi-\phi)} = \sum_{m=0}^{\infty} j^m J_m(k_\perp r_\perp)\, e^{jm(\psi-\phi)} \tag{20.20.3}$$

Then, Eq. (20.20.2) becomes,

$$\hat{A}(k_\perp,\psi) = \sum_{m=0}^{\infty} j^m e^{jm\psi} \int_0^a J_m(k_\perp r_\perp) \left(\int_{\Phi(r_\perp)} e^{-jm\phi}\, d\phi \right) r_\perp\, dr_\perp \tag{20.20.4}$$

The integral over $\Phi(r_\perp)$ is calculated by assuming initially that m is any real number,

$$\begin{aligned} \int_{\Phi(r_\perp)} e^{-jm\phi}\, d\phi &= \sum_{n=0}^{N-1} \int_{[2n+1-A(r_\perp)]\pi/N}^{[2n+1+A(r_\perp)]\pi/N} e^{-jm\phi}\, d\phi \\ &= 2\, e^{j\pi m/N}\, \frac{\sin(\pi m A(r_\perp)/N)}{m} \sum_{n=0}^{N-1} e^{2\pi jnm/N} \\ &= 2\, e^{j\pi m/N}\, \frac{\sin(\pi m A(r_\perp)/N)}{m}\, \frac{1-e^{2\pi jm}}{1-e^{2\pi jm/N}} \end{aligned} \tag{20.20.5}$$

and taking appropriate limits at integer values of m, we have more explicitly,

[†] assumed to be real-valued and satisfying, $0 \le A(r_\perp) \le 1$.

$$\int_{\Phi(r_\perp)} e^{-jm\phi}\, d\phi = \begin{cases} 2\pi A(r_\perp)\,, & \text{if } m = 0 \\ 2\, e^{j\pi p}\, \dfrac{\sin(\pi p A(r_\perp))}{p}\,, & \text{if } m = pN,\ p \ne 0 \\ 0\,, & \text{other } m \end{cases} \tag{20.20.6}$$

It follows that,

$$\begin{aligned} \hat{A}(k_\perp,\psi) &= \int_0^a J_0(k_\perp r_\perp)\, A(r_\perp)\, 2\pi r_\perp\, dr_\perp \\ &\quad + 2\sum_{p=1}^{\infty} e^{jpN\psi}\, j^{pN}\, \frac{e^{j\pi p}}{p} \int_0^a J_{pN}(k_\perp r_\perp)\, \sin(\pi p A(r_\perp))\, r_\perp\, dr_\perp \end{aligned} \tag{20.20.7}$$

For $k_\perp a < 2\sqrt{N+1}$, the $p \ge 1$ terms can be shown [1531] to converge rapidly to zero with increasing N. Therefore, effectively only the first term survives, which demonstrates the equivalence of the star-shaped mask with the given circularly-symmetric apodization function $A(r_\perp)$, satisfying the Hankel transform relationship (20.16.2).

Example 20.20.1: Fig. 20.20.1 shows the star-shaped masks with $N = 20$, corresponding to the two designed cases of Example 20.19.1 with sidelobe levels $R = 60$ dB and $R = 100$ dB. The MATLAB code for producing these figures is as follows (shown for $R = 100$),

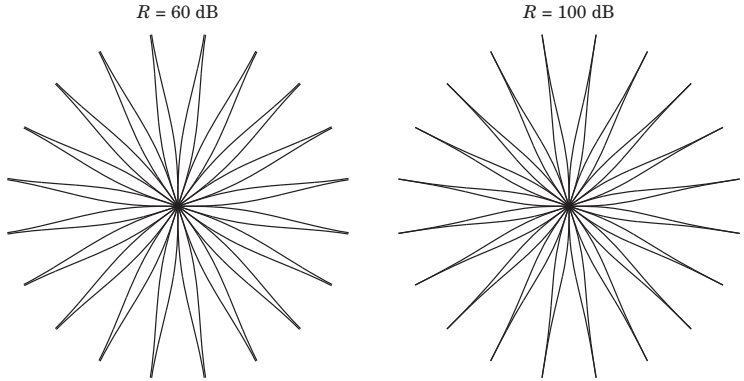


Fig. 20.20.1 Star-shaped mask realizations of Example 20.19.1.

```
r = linspace(0,1,501);      % units of a

u = linspace(0,20,2001);    % wavenumber in u-units

R = 100; nb = 30;

[Ap,Fn,mu] = tnb2(R,nb);
```

```
F = 0; A = 0;
for n=1:nb,
    F = F + Fn(n) * jinc(u,mu(n));
    A = A + Fn(n) * besselj(0,pi*mu(n)*r)/besselj(0,pi*mu(n))^2/pi;
end

A = A/max(A);
F = abs(F); Fdb = 20*log10(F/max(F));

figure; plot(u,Fdb,'b-');                % wavenumber pattern
figure; plot(r,A,'b-');                  % apodization function A(r)

N = 20;

figure; hold on
for n=0:N-1,                             % draw N vanes
   phi = (2*n+1-A)*pi/N;                 % lower half of a vane
   x = r.*cos(phi); y = r.*sin(phi);
   plot(x,y,'k-');

   phi= (2*n+1+A)*pi/N;                  % upper half of a vane
   x = r.*cos(phi); y = r.*sin(phi);
   plot(x,y,'k-');
end
axis equal; axis square;
```

To see what happens when $A(r_\perp)$ is not monotonically decreasing, run the same code with $R = 50$ dB and $\bar{n} = 20$. □

Starshade Occulters

The search for extrasolar Earth-like planets orbiting nearby stars is a challenging task for ordinary telescopes—the light from the star blots out the weak light from the planet. The light intensity from such a typical planet is expected to be 100 dB (i.e., 10^{-10} times) weaker than that of its parent star.

The use of an external occulter, or starshade, that blocks the starlight from reaching the telescope has been proposed as possible solution [1516–1544]. However, a plain opaque disk would not work as an occulter because of the spot of Arago caused by light diffraction around the edges of the disk. The shape of the occulter must be properly designed to suppress such diffracted light. A typical starshade is depicted in Fig. 20.20.2.

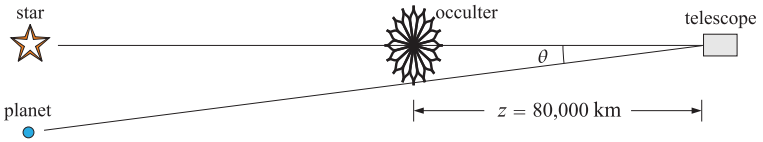


Fig. 20.20.2 Starshade occulter.

If the occulter has radius R and is positioned at a distance z from the telescope, the planet to be detected must lie beyond the edge of the occulter, that is, at an angular separation of at least, $\theta \approx \tan\theta = R/z$. This can be solved for z to determine the proper

placement of the occulter for a given R and desired θ, that is, $z = R/\theta$. For example, for an occulter of radius of 25 meters and a nominal angle of 65 mas (milliarcsec), we must have,

$$z = \frac{R}{\theta} = \frac{25}{(0.065/3600)\,\pi/180} = 79332.6 \text{ km}$$

Such distances z are large enough so that the Fresnel approximation can be applied at optical wavelengths, but are too short for the simpler Fraunhofer approximation. For example, for $R = 25$ meters and $\lambda = 500$ nm, we obtain the following estimates of the Fresnel and Fraunhofer distances beyond which the respective approximations can be applied (see also Example 20.3.1),

$$z_{\text{fresnel}} = \frac{1}{k}(kR)^{4/3} = 17 \text{ km}, \qquad z_{\text{fraunh}} = \frac{1}{k}(kR)^2 = 7.85\times 10^6 \text{ km}$$

To apply the Fresnel approximation to this problem, we will assume that the telescope and starshade are aligned with the star under observation and choose coordinates so that the starshade is at $z = 0$ and the telescope at a distance z to the right as shown above. The light from the star is assumed to be a plane wave of amplitude E_0 incident normally on the starshade. Thus, the plane wave measured at distance z will be $E_0 e^{-jkz}$.

The starshade may be designed to apply a circularly symmetric apodization function, say, $A(r_\perp)$, on the incident plane wave. The complementary aperture to the starshade will have a corresponding apodization function $A_c(r_\perp) = 1 - A(r_\perp)$. Using the Fresnel approximation of Eq. (20.1.13), the occulted field measured at the plane of the telescope at distance z can be expressed [1535,1543] either in terms of $A(r_\perp)$, or through the Babinet principle, in terms of $A_c(r_\perp)$ (see also Problem 20.9),

$$\begin{aligned} E(r_\perp, z) &= E_0\, e^{-jkz}\, \frac{jk}{2\pi z}\, e^{-jkr_\perp^2/2z} \int_0^\infty A(r'_\perp)\, e^{-jkr'^2_\perp/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp \\ &= E_0\, e^{-jkz}\left[1 - \frac{jk}{2\pi z}\, e^{-jkr_\perp^2/2z} \int_0^\infty A_c(r'_\perp)\, e^{-jkr'^2_\perp/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp\right] \end{aligned}$$

The $r'_\perp$ integration in the latter expression can be restricted to be over the effective radius, say, R, of the complementary hole aperture, so that,

$$E(r_\perp, z) = E_0\, e^{-jkz}\left[1 - \frac{jk}{2\pi z}\, e^{-jkr_\perp^2/2z} \int_0^R A_c(r'_\perp)\, e^{-jkr'^2_\perp/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp\right] \tag{20.20.8}$$

The design requirement for $A(r'_\perp)$ is that, $|E(r_\perp, z)|^2 \le 10^{-10}|E_0|^2$, over an appropriate range, $0 \le r_\perp \le r_{\max} \le R$. Once $A(r'_\perp)$ is found, Eq. (20.20.8) may be computed numerically, and may be realized as a star-shaped mask as we discussed above. As an example, we consider the "offset hyper-gaussian" apodization function introduced by Cash for the New Worlds Observer (NWO) platform [1534,1540,1543],

$$A(r_\perp) = \begin{cases} 0\,, & r_\perp \le a \\ 1 - \exp\left[-\left(\dfrac{r_\perp - a}{b}\right)^m\right], & r_\perp > a \end{cases} \tag{20.20.9}$$

and its complementary function, $A_c(r_\perp) = 1 - A(r_\perp)$,

$$A_c(r_\perp) = \begin{cases} 1\,, & r_\perp \le a \\ \exp\left[-\left(\dfrac{r_\perp - a}{b}\right)^m\right], & r_\perp > a \end{cases} \tag{20.20.10}$$

These functions are plotted in Fig. 20.20.3 for the proposed values [1534] of the parameters, $a = b = 12.5$ meters, $m = 6$, with the overall radius of the occulter $R = 31.25$ meters, so that $a = b = 0.4R$, and telescope-occulter distance of $z = 80,000$ km. We note that the inflection point occurs at $r_\perp = a + b(1 - 1/m)^{1/m}$, which is approximately equal to $a + b$ for large m.

The corresponding star-shaped realization using $N = 16$ petals is also shown in the same figure. Such petal realizations, but using polynomial functions $A(r_\perp)$, were first considered by Marchal [1517].

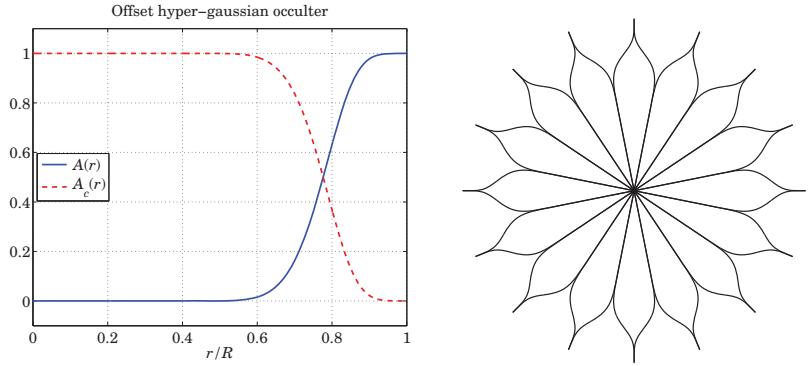


Fig. 20.20.3 Occulting apodizer and its star-shaped realization.

Although the definition (20.20.9) contains no upper cutoff for $r_\perp$, the value $R = 31.25$ is a practical cutoff beyond which the exponential term becomes effectively zero. In computing the integral in (20.20.8) numerically, we found it beneficial to extend the integration range to $1.5R$.

Fig. 20.20.4 plots the occulted field intensity, $|E(r_\perp, z)|^2$ versus transverse distance $r_\perp$ at the plane of the telescope at distance $z = 80,000$ km. The left graph shows the same four wavelength cases as in [1543], that is, $\lambda = 2000, 1000, 500, 250$ nm. We observe that the desired 100-dB suppression is well satisfied within a 5-meter radius for the optical band.

The right graph shows the case of a plain opaque circular disk of radius R that has $A(r_\perp) = u(r_\perp - R)$ or $A_c(r_\perp) = u(R - r_\perp)$, at wavelength $\lambda = 500$ nm. The spot of Arago phenomenon is evident at the center $r_\perp = 0$. It is precisely the diffracted components near $r_\perp = 0$ that are being eliminated by the apodized occulter.

The MATLAB code used to generate the above graphs is as follows, in which the integrations in (20.20.8) are implemented with the tanh-sinh quadrature function, **quadts**,

```
R = 31.25;                      % meters
a = 0.4; b = 0.4; m = 6;        % a,b in units of R
z = 80000e3;                    % meters, z = 80000 km
N = 16;                         % number of petals

A  = @(r) (r>=a) .* (1 - exp(-((r-a)/b).^m));   % occulter
Ac = @(r) 1 - A(r);                             % complementary
```

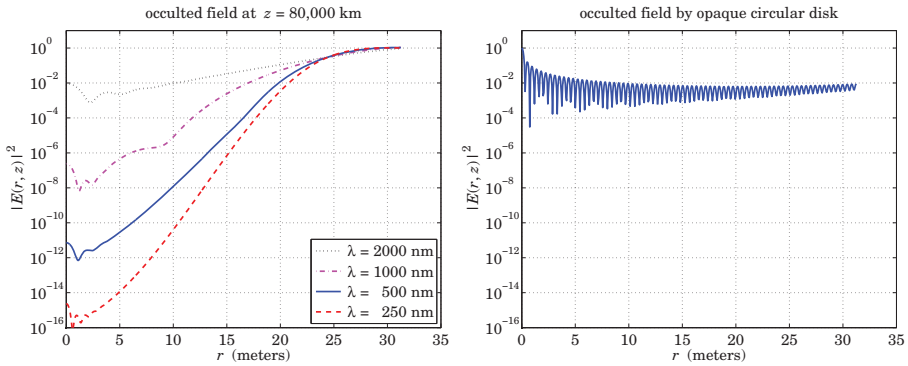



Fig. 20.20.4 Occulted field intensity $|E(r_\perp, z)|^2$ versus $r_\perp$, at plane of telescope.

```
r = linspace(0,1,501);

figure; plot(r,A(r),'b-', r,Ac(r),'r--');

figure; hold on                         % generate star-shaped mask
for n=0:N-1,                            % Ac(r) could also be used
   phi = (2*n+1-A(r))*pi/N;
   x = r.*cos(phi); y = r.*sin(phi);
   plot(x,y,'k-');
   phi= (2*n+1+A(r))*pi/N;
   x = r.*cos(phi); y = r.*sin(phi);
   plot(x,y,'k-');
end
axis equal; axis square;

Rq = 1.5;                               % extended integration range
[wq,rq] = quadts(0,Rq,8);               % quadts weights and evaluation points

E = [];

for  la = [250, 500, 1000, 2000]        % nanometers

   k = 2*pi/la * 1e9;                   % rads/meter
   F = k*R^2/z;                         % Fresnel number = F/(2*pi) = R^2/(lambda*z)

   for i=1:length(r)                    % Eq.(20.20.11)
     I = wq' * (besselj(0,F*r(i)*rq) .* exp(-j*F*rq.^2/2) .* Ac(rq) .* rq);
     Ek(i) = 1 - j*F * exp(-j*F*r(i)^2/2) * I;
   end

   E = [E; abs(Ek).^2];                 % intensities

end

figure; semilogy(r*R, E(4,:), 'k:', r*R, E(3,:), 'm-.', ...
                 r*R, E(2,:), 'b-', r*R, E(1,:), 'r--');
```

For the purpose of computing Eq. (20.20.8), we may rewrite it in terms of the normalized variables, $\rho = r_\perp/R$, $\rho' = r'_\perp/R$, and the Fresnel number, $F = kR^2/z$,†

$$E(\rho,z) = E_0\, e^{-jkz}\left[1 - jF\, e^{-jF\rho^2/2}\int_0^1 A_c(\rho')\, e^{-jF\rho'^2/2}\, J_0(F\rho\rho')\, \rho'\, d\rho'\right] \qquad (20.20.11)$$

The computed Fresnel numbers were as follows for the four wavelength cases,

λ (nm)	F	$F/(2\pi)$
250	306.80	48.83
500	153.40	24.41
1000	76.70	12.21
2000	38.35	6.10

References [1516–1544] contain further details on the design of optimum apodizers for occulters and coronagraphs, star-shaped and spiderweb masks. See also the NWO web site, `http://newworlds.colorado.edu/`.

20.21 Superresolving Apertures

The emphasis of the apodization designs that we discussed in Sections 20.12–20.20, was to suppress the sidelobes in order to be able to detect weak sources. We mentioned that as one tries to suppress the sidelobes, the mainlobe gets wider, thus, potentially decreasing the resolution in Rayleigh's sense that two sources of approximately equal strength cannot be distinguished if they both fall within the range of the mainlobe.

The question then arises whether it is possible to make the mainlobe narrower than the diffraction limit of the uniform aperture while at the same time sufficiently suppressing the sidelobes—at least within a *finite* field of view (FOV), i.e., within a finite distance from the mainlobe. More generally, is it possible to obtain any desired wavenumber pattern, even infinitely narrow, from an aperture or source of finite size?

The answer is that "in principle, yes," but "in practice, no." The question has been addressed by Bouwkamp and de Bruijn [1599] for linear antennas, and by Woodward and Lawson [1601] for apertures, as well as by Toraldo di Francia [1481,1482] for circular apertures. See also the additional references on superdirective antennas [1599–1618] and [1479–1500] on extensions of Toraldo di Francia's approach and its experimental demonstration [1494].

Even though such superresolving designs work well within the FOV, they are usually accompanied by incredibly large sidelobes outside the FOV, where most of the energy is concentrated, or, in the case of aperture antennas they are accompanied by very large reactive energy stored in the vicinity of the antennas. Moreover, because of the widely ranging values of the design parameters, the manufacturing tolerances become extremely difficult to meet. These remarks will be illustrated below.

† actually the Fresnel number is $F/(2\pi) = R^2/(\lambda z)$.

Toraldo Rings and Annuli

We recall that a uniform circular aperture of radius a, has the standard, $2J_1(v)/v$, wavenumber pattern, that is, if $A(r_\perp) = u(a - r_\perp)$, then

$$\hat{A}(k_\perp) = \int_0^\infty A(r_\perp) J_0(k_\perp r_\perp)\, 2\pi r_\perp\, dr_\perp = \pi a^2 \cdot \frac{2J_1(k_\perp a)}{k_\perp a} \qquad (20.21.1)$$

whose first null at $k_\perp a = v_1 = 3.8317$ determines the extent of the mainlobe. Introducing the variable $v = k_\perp a$, we may define $F(v) = \hat{A}(k_\perp)$ and write (20.21.1) as,

$$F(v) = \pi a^2 \cdot \frac{2J_1(v)}{v} \qquad (20.21.2)$$

The integral (20.21.1) is a linear combination of $J_0(k_\perp r_\perp)$ ring patterns of different radii $r_\perp$. As noted by Rayleigh and Airy [1479,1480], a single such narrow ring, say at radius a, would have $A(r_\perp) = \delta(r_\perp - a)$ and $\hat{A}(k_\perp) = 2\pi a \cdot J_0(k_\perp a)$, or, $F(v) = 2\pi a J_0(v)$, and would exhibit better resolution than the uniform case, having a narrower mainlobe since the first null of the pattern is now at $k_\perp a = v_0 = 2.4048$.† However, this comes at the expense of higher sidelobes. The ring and the uniform cases are compared in the top-left graph of Fig. 20.21.2.

The "superresolution gain", defined as the ratio of the first null of the uniform case to the first null of the superresolved mainlobe, would be in this case, $g = v_1/v_0 = 3.8317/2.4048 = 1.5934$.

Toraldo di Francia's idea [1481] was to try to diminish the sidelobes of J_0 without altering the first null at v_0, by forming a linear combination of a finite number of rings such that additional zeros are introduced near the sidelobes of J_0. Fig. 20.21.1 depicts the case of four rings. Alternatively, in order to allow more light through, he also proposed the use of annuli [1482], as shown on the right in this figure.

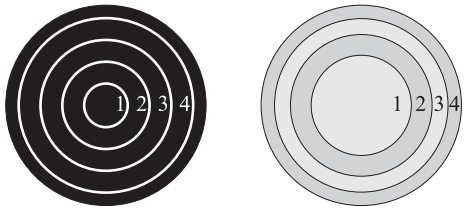


Fig. 20.21.1 Toraldo rings and annuli ($N = 4$).

The simplest example from [1481] is a pattern formed by three rings at radii, $a_1 = a/3$, $a_2 = 2a/3$, and $a_3 = a$,

$$\hat{A}(k_\perp) = c_1 J_0(k_\perp a_1) + c_2 J_0(k_\perp a_2) + c_1 J_0(k_\perp a_3)\,, \quad \text{or,}$$

$$F(v) = c_1 J_0\left(\frac{v}{3}\right) + c_2 J_0\left(\frac{2v}{3}\right) + c_3 J_0(v)$$

† v_1, v_0 are the smallest non-zero roots satisfying, $J_1(v_1) = J_0(v_0) = 0$.

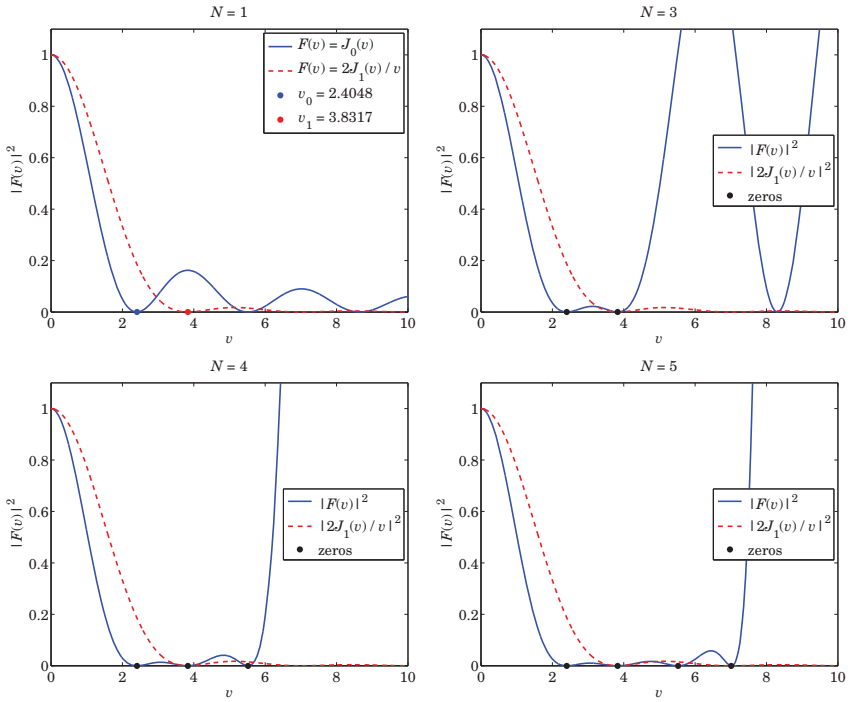


Fig. 20.21.2 Toraldo di Francia examples with $N = 1, 3, 4, 5$ rings, gain $g = 1.59$.

Choosing the extra zero to be at $\nu = \nu_1 = 3.8317$, the coefficients are then determined by the three conditions, $F(0) = 1$, $F(\nu_0) = F(\nu_1) = 0$,

$$\begin{aligned} F(0) &= c_1 J_0(0) + c_2 J_0(0) + c_3 J_0(0) = c_1 + c_2 + c_3 = 1 \\ F(\nu_0) &= c_1 J_0\left(\frac{\nu_0}{3}\right) + c_2 J_0\left(\frac{2\nu_0}{3}\right) + c_3 J_0(\nu_0) = 0 \\ F(\nu_1) &= c_1 J_0\left(\frac{\nu_1}{3}\right) + c_2 J_0\left(\frac{2\nu_1}{3}\right) + c_3 J_0(\nu_1) = 0 \end{aligned}$$

The solution of these equations gives, $[c_1, c_2, c_3] = [0.9505, -1.7723, 1.8218]$, and the resulting pattern is plotted in the top-right graph of Fig. 20.21.2. The suppression of the first sidelobe of J_0 has resulted in a finite field of view, but also in very large sidelobes further away. The FOV can be further widened by increasing the number of rings and imposing additional zeros at further sidelobes.

In general, using N rings at radii a_n, $n = 1, 2, \ldots, N$, the aperture distribution and pattern are constructed as follows,

$$\begin{aligned} A(r_\perp) &= \sum_{n=1}^{N} c_n \frac{\delta(r_\perp - a_n)}{2\pi a_n} \\ \hat{A}(k_\perp) &= \sum_{n=1}^{N} c_n J_0(k_\perp a_n)\,, \quad F(\nu) = \sum_{n=1}^{N} c_n J_0(\alpha_n \nu) \end{aligned} \tag{20.21.3}$$

where $\alpha_n = a_n/a$, $n = 1, 2, \ldots, N$. The N coefficients c_n are determined by imposing the N conditions, $F(0) = 1$ and $F(\nu_i) = 0$ at $N-1$ appropriately chosen zeros, ν_i, $i = 1, 2, \ldots, N-1$, resulting in the $N \times N$ linear system of equations,

$$\begin{aligned} F(0) &= \sum_{n=1}^{N} c_n = 1 \\ F(\nu_i) &= \sum_{n=1}^{N} c_n J_0(\alpha_n \nu_i) = 0\,, \quad i = 1, 2, \ldots, N-1 \end{aligned} \tag{20.21.4}$$

where the ring radii $a_n = \alpha_n a$ can be arbitrarily chosen. Following Toraldo di Francia, we choose them to be equally-spaced in $[0, a]$, that is, $\alpha_n = n/N$, $n = 1, 2, \ldots, N$.

Two further examples from [1481] shown in the lower two graphs in Fig. 20.21.2, correspond to the cases of $N = 4$ and $N = 5$ rings, with the $N-1$ zeros chosen to alternate between the zeros of $J_0(\nu)$ and $J_1(\nu)$,

$$\begin{aligned} N = 4, \quad \nu_i &= [2.4048,\ 3.8317,\ 5.5201] \\ N = 5, \quad \nu_i &= [2.4048,\ 3.8317,\ 5.5201,\ 7.0156] \end{aligned}$$

and resulting in the coefficients c_n,

$$\begin{aligned} N = 4, \quad c_n &= [-2.8130,\ 7.6524,\ -7.5827,\ 3.7433] \\ N = 5, \quad c_n &= [13.6152,\ -37.3952,\ 43.8297,\ -27.7134,\ 8.6636] \end{aligned}$$

We observe again the widening of the FOV, but at the expense of large values beyond. The above examples had the same superresolution gain of $g = \nu_1/\nu_0 = 3.8317/2.4048 = 1.5934$. The next example, from [1488,1489], has gain $g = 2$ and attempts to create a mainlobe that closely follows the compressed Airy mainlobe, that is,

$$F(\nu) = \frac{2J_1(g\nu)}{g\nu} \tag{20.21.5}$$

Five zeros are chosen ($N = 6$) to coincide with the first five non-zero roots of $J_1(g\nu) = 0$, with the six constraint equations, $F(0) = 1$ and $F(\nu_i) = 0$, $i = 1, 2, \ldots, 5$. The resulting pattern is plotted on the top-left in Fig. 20.21.3, and on bottom-left in semilogy units. We observe that the mainlobe closely follows that of Eq. (20.21.5). The computed zeros ν_i and coefficients c_i are,

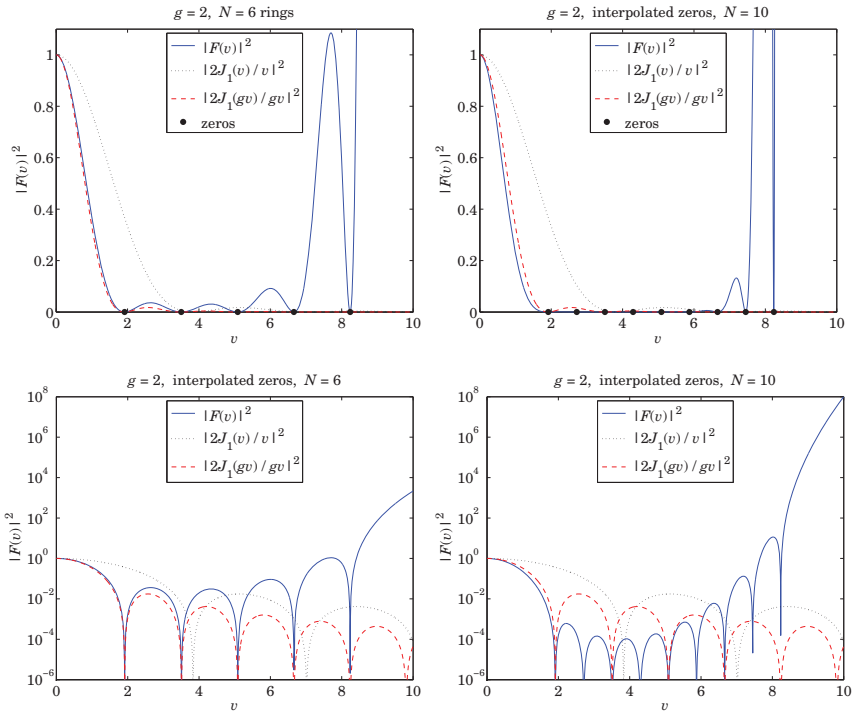


Fig. 20.21.3 Superresolution gain $g = 2$, with 5 zeros of $J_1(g\nu) = 0$, and interpolated zeros.

i	ν_i	c_i
1	1.915853	-2.291191×10^{2}
2	3.507793	6.545794×10^{2}
3	5.086734	-8.196087×10^{2}
4	6.661846	6.027091×10^{2}
5	8.235315	-2.618712×10^{2}
6		5.431055×10^{1}

In order to improve the FOV and further reduce the sidelobes, we may introduce additional zeros between the above zeros. For simplicity, we take them to be the midpoint averages between the original zeros. The resulting pattern is plotted on the top-right in Fig. 20.21.3 and on the bottom-right in semilogy units. The FOV has been effectively extended from $\nu = 5$ to about $\nu = 7$, but the price is to introduce incredibly large sidelobes outside the FOV. Contrary to intuition, suppressing the FOV sidelobes actually resulted in a slightly narrower mainlobe than that of (20.21.5), shifting the extra energy to the far sidelobes. The zeros and coefficients were as follows in this case,

i	ν_i	c_i
1	1.915853	-2.803205×10^{9}
2	2.711823	8.622638×10^{9}
3	3.507793	-1.244824×10^{10}
4	4.297264	1.173322×10^{10}
5	5.086734	-7.909577×10^{9}
6	5.874290	3.905794×10^{9}
7	6.661846	-1.396901×10^{9}
8	7.448580	3.457084×10^{8}
9	8.235315	-5.332753×10^{7}
10		3.884012×10^{6}

The extremely large values of the coefficients and their wide range of variation by four orders of magnitude would be very difficult, if not impossible, to realize in practice. The MATLAB code used to generate these graphs is as follows,

```
g = 2;                          % superresolution gain
N = 6;                          % number of rings, N-1 zeros

an = (1:N)/N;                   % radii
vi = zeros(N,1);                % vi(1)=0 is not a zero, but corresponds to F(0)=1

for i=1:N-1,
   vi(i+1) = fzero(@(v) besselj(1,v*g), (i+1/4)*pi/g);    % zeros of J1(g*v)=0
end

Vi = vi * an;                   % NxN matrix [vi/N, 2*vi/N, 3*vi/N, ..., N*vi/N]
Fi = [1; zeros(N-1,1)];         % Fi(1)=1 corresponds to vi(1)=0

c = besselj(0,Vi) \ Fi;         % expansion coefficients, besselj is vectorized

v = linspace(0,10,2001)';       % column vector

V = v * an;                     % matrix [v/N, 2*v/N, 3*v/N, ..., N*v/N]

F = abs(besselj(0, V) * c).^2;      % evaluate F(v), same size as v

figure; plot(v,F,'b-'); ylim([0, 1]);

% --- interpolated case ---

y = filter([1,1]/2,1,vi);       % compute averages between zeros
vi(2:2:2*N-2) = vi(2:N);        % keep original zeros
vi(3:2:2*N-3) = y(3:N);         % add interpolated zeros, new vi has length 2*N-2

N = 2*N-2;                      % new N
an = (1:N)/N;                   % new radii

Vi = vi * an;
Fi = [1; zeros(N-1,1)];

c = besselj(0,Vi) \ Fi;         % expansion coefficients
```

```
V = v * an;

F = abs(besselj(0, V) * c).^2;

figure; plot(v,F,'b-'); ylim([0, 1]);
```

Toraldo di Francia's alternative method of using annuli instead of rings is as easily implemented. The nth annulus is defined by the radial range, $a_{n-1} \le r_\perp \le a_n$, for $n = 1, 2, \ldots, N$, where by convention, $a_0 = 0$, so that the first annulus is the circle of radius a_1. The aperture distribution (normalized by πa^2) and pattern are in this case,

$$A(r_\perp) = \frac{1}{\pi a^2} \sum_{n=1}^{N} c_n [u(a_n - r_\perp) - u(r_\perp - a_{n-1})]$$

$$\hat{A}(k_\perp) = \frac{1}{\pi a^2} \sum_{n=1}^{N} c_n \left[\pi a_n^2 \frac{2J_1(k_\perp a_n)}{k_\perp a_n} - \pi a_{n-1}^2 \frac{2J_1(k_\perp a_{n-1})}{k_\perp a_{n-1}} \right] \qquad (20.21.6)$$

$$F(v) = \sum_{n=1}^{N} c_n \left[\alpha_n^2 \frac{2J_1(\alpha_n v)}{\alpha_n v} - \alpha_{n-1}^2 \frac{2J_1(\alpha_{n-1} v)}{\alpha_{n-1} v} \right]$$

where as before, $\alpha_n = a_n/a$, $n = 0, 1, \ldots, N$, and the values are chosen to be $\alpha_n = \sqrt{n/N}$, so that all annuli have the same area, that is, $1/N$ the total πa^2 of the aperture, indeed, we have, $\alpha_n^2 - \alpha_{n-1}^2 = n/N - (n-1)/N = 1/N$. The N coefficients c_n are determined by imposing N conditions at $v = 0$ and at $N-1$ zeros.

Fig. 20.21.4 shows the same example of Fig. 20.21.3 using annuli instead of rings. The essential features remain the same. The computed zeros are the same as before, and the coefficients c_n are now,

n	c_n	c_n, interpolated
1	-3.880891×10^{2}	-4.814752×10^{8}
2	2.547225×10^{3}	4.711524×10^{9}
3	-6.716727×10^{3}	-2.049879×10^{10}
4	8.910165×10^{3}	5.204489×10^{10}
5	-5.950388×10^{3}	-8.497935×10^{10}
6	1.603814×10^{3}	9.254088×10^{10}
7		-6.721158×10^{10}
8		3.139471×10^{10}
9		-8.558137×10^{9}
10		1.037337×10^{9}

The MATLAB code for generating these graphs was as follows,

```
g = 2;
N = 6;

a = sqrt((0:N)/N);      % radii
vi = zeros(N,1);

f = @(v,a,b) b^2 * jinc(b*v/pi) - a^2 * jinc(a*v/pi);   % pattern of [a,b] annulus
```

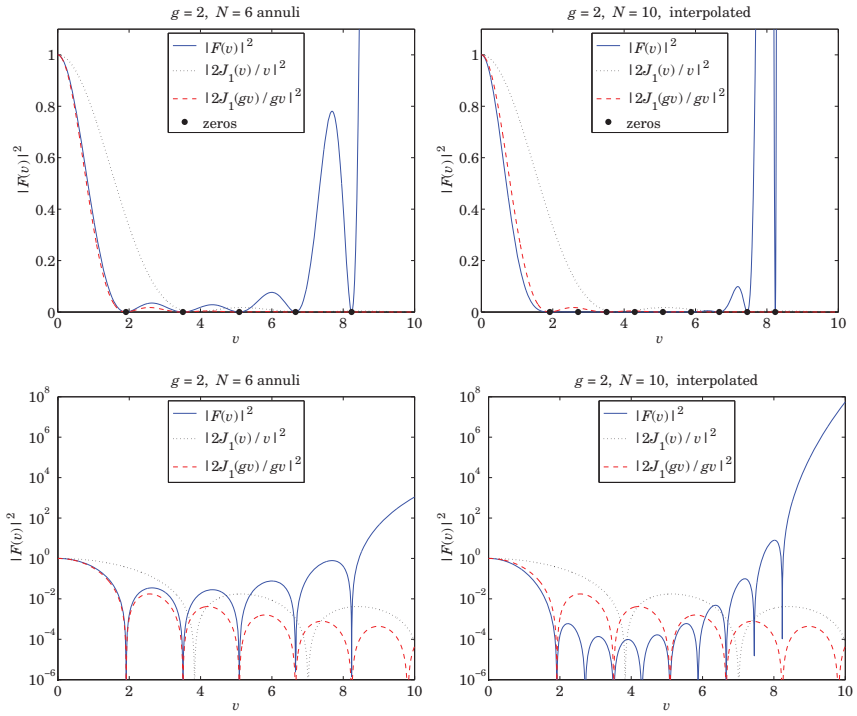



Fig. 20.21.4 Superresolution gain $g = 2$, with 5 zeros of $J_1(gv) = 0$, and interpolated zeros.

```
for i=1:N-1,
   vi(i+1) = fzero(@(v) besselj(1,v*g), (i+1/4)*pi/g);    % zeros of J1(g*v)=0
end

v = linspace(0,10,2001)';

Gi = []; G = [];
for i=1:N,                                  % construct pattern matrices
   Gi = [Gi, f(vi, a(i),a(i+1))];           % for coefficients c
   G = [G, f(v, a(i),a(i+1))];              % for pattern F(v)
end

Fi = [1; zeros(N-1,1)];

c = Gi \ Fi;                                % expansion coefficients

F = abs(G * c).^2;                          % evaluate F(v)

figure; plot(v,F,'b-'); ylim([0, 1]);
```

```
% --- interpolated zeros ---

N = length(vi);
y = filter([1,1]/2,1,vi);
vi(2:2:2*N-2) = vi(2:N);
vi(3:2:2*N-3) = y(3:N);

N = 2*N-2;              % new N
a = sqrt((0:N)/N);      % new radii

Gi = []; G = [];
for i=1:N,                                  % construct pattern matrices
   Gi = [Gi, f(vi, a(i),a(i+1))];           % for coefficients c
   G = [G, f(v, a(i),a(i+1))];              % for pattern F(v)
end

Fi = [1; zeros(N-1,1)];

c = Gi \ Fi;                                % expansion coefficients

F = abs(G * c).^2;                          % evaluate F(v)

figure; plot(v,F,'b-'); ylim([0, 1]);
```

Woodward-Lawson Superresolving 1-D Apertures

Perhaps the first example of a superresolving aperture was by Woodward and Lawson [1601], who considered a 1-D aperture of length $2a = \lambda$ and designed an aperture distribution that has a superresolution gain of $g = 2$.

In the 1-D case, we recall that a uniform aperture of dimensions $[-a, a]$ will have the standard sinc-function pattern, that is, with $A(x) = u(a - |x|)$, and, $v = k_x a$,

$$\hat{A}(k_x) = \int_{-\infty}^{\infty} A(x)\, e^{jk_x x} dx = 2a \cdot \frac{\sin(k_x a)}{k_x a} = 2a \cdot \frac{\sin(v)}{v} \tag{20.21.7}$$

Since in the antenna context, $k_x = k \sin\theta$, where $k = 2\pi/\lambda$, it follows that the visible region corresponding to real angles θ is the range, $|k_x| \le k$ or, $|v| \le ka$, in v-units. Since $2a = \lambda$ in the Woodward-Lawson example, we have $ka = 2\pi a/\lambda = \pi$, so that the visible region becomes, $|v| \le \pi$. A compressed sinc-pattern by a superresolution factor g, will have response,

$$F_g(v) = \frac{\sin(gv)}{gv} \tag{20.21.8}$$

The first nulls of the standard and the compressed patterns, which define the extents of the corresponding mainlobes, will be the smallest non-zero roots of, $\sin(v) = 0$, and, $\sin(gv) = 0$, respectively, that is, $v = \pi$, and, $v = \pi/g$. Thus, for the compressed pattern, the interval $|v| \le \pi/g$ defines the mainlobe, and $\pi/g \le |v| \le \pi$, is that portion of the sidelobe range that lies in the visible region. Woodward-Lawson's idea was to introduce additional zeros in the visible range $[\pi/g, \pi]$ in order to further suppress those sidelobes. The aperture distribution and pattern were constructed as linear combinations of shifted sinc-functions, as follows,

$$\begin{aligned} A(x) &= \frac{1}{2a} \sum_{n=-M}^{M} c_n e^{-jk_n x} \cdot \chi_a(x) \\ \hat{A}(k_x) &= \sum_{n=-M}^{M} c_n \frac{\sin\big((k_x - k_n)a\big)}{(k_x - k_n)a}, \quad F(v) = \sum_{n=-M}^{M} c_n \frac{\sin(v - v_n)}{v - v_n} \end{aligned} \tag{20.21.9}$$

where $v_n = k_n a$, $n = \pm 1, \pm 2, \ldots, \pm M$ are the chosen zeros. The $2M+1$ linear combination coefficients c_n, $-M \le n \le M$, are determined by imposing the $2M+1$ conditions,

$$\begin{aligned} F(0) &= 1 \\ F(v_n) &= 0, \quad n = \pm 1, \pm 2, \ldots, \pm M \end{aligned}$$

The zeros are symmetrically placed about $n = 0$ and as a result, the coefficients c_n, and $F(v)$ and $A(x)$ are symmetric and real-valued.

Such expansions are reminiscent of the Woodward-Lawson orthogonal basis expansions of Eq. (20.12.25), however, here they are different because v_n do hot have to be chosen at the multiples, $v_n = \pi n$. With arbitrarily chosen v_n, the above shifted sinc functions are linearly independent, but not necessarily mutually orthogonal.

Fig. 20.21.5 shows the patterns $|F(v)|$ and aperture distributions $A(x)$ constructed via (20.21.9), for the cases, $2a = \lambda$, $g = 2$, and the following choices of M zeros in the interval $[\pi/g, \pi]$, the first being the original Woodward-Lawson example [1601],

$$M = 3, \qquad v_n = \frac{\pi}{2} + (n-1)\frac{\pi}{4}, \quad n = 1, 2, 3 \qquad \text{(Woodward-Lawson)}$$

$$M = 4, \qquad v_n = \frac{\pi}{2} + (n-1)\frac{\pi}{6}, \quad n = 1, 2, 3, 4$$

We observe similar features as in the 2-D case, namely, (a) the solution follows closely the mainlobe of the compressed sinc, (b) the sidelobes are suppressed in the visible region, e.g., $|v| \le \pi$, but are huge in the invisible region, (c) the coefficients c_n and aperture distribution $A(x)$ undergo huge variations over the range $[-a, a]$. The computed coefficients c_n were as follows in the two cases,

n	c_n	c_n
−4		39396219.62
−3	−16678.51	−192113502.42
−2	58231.42	347572126.67
−1	−66582.91	−241677043.29
0	49825.86	93654887.13
1	−66582.91	−241677044.03
2	58231.42	347572128.11
3	−16678.51	−192113503.43
4		39396219.88

The MATLAB code used to generate the graphs and coefficients c_n is as follows, where the values of g, M can be changed to try different cases,

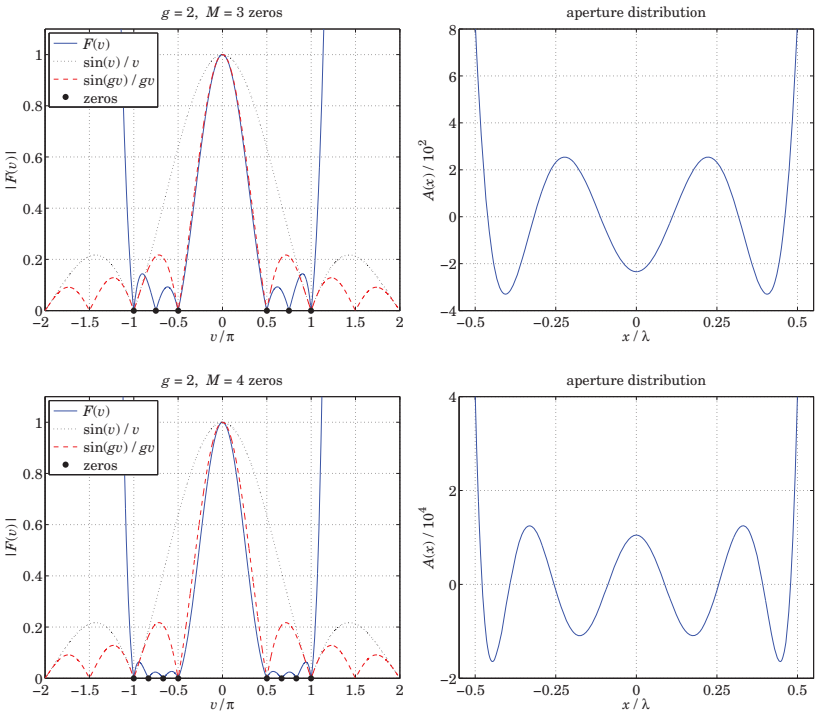


Fig. 20.21.5 Superresolution gain $g = 2$, with $M = 3, 4$ zeros, visible region $-\pi \le v \le \pi$.

```
a = 1/2;                            % units of lambda, here, k*a = pi
g = 2;                              % gain
M = 3;                              % M zeros in [1/g, 1]
un = linspace(1/g, 1, M);           % vn = pi*un

Fn = [zeros(M,1); 1; zeros(M,1)];   % response at zeros and u=0
un = [-flip(un), 0, un]';           % symmetrize for convenience
N = length(un);                     % N = 2*M+1

u = linspace(-2,2,8001)';           % column vector, v = pi*u

Gn = []; G = [];
for i=1:N                           % data matrices
   Gn = [Gn, sinc(un - un(i))];     % for coefficients c
    G = [G,  sinc(u - un(i))];      % for F(v) evaluation
end

c = Gn\Fn;                          % expansion coefficients

F = abs(G*c);                       % evaluate |F(v)|
F1 = abs(sinc(u));                  % uniform sinc
Fg = abs(sinc(g*u));                % compressed sinc
```



```
figure; plot(u,F,'b-', u,F1,'k:', u,Fg,'r--'); ylim([0,1]);

kn = pi*un/a;                       % wavenumbers at un
x = linspace(-1,1,1001)*a;          % x in [-a,a]
A = c' * exp(-j*kn*x) / (2*a);      % aperture distribution

figure; plot(x, real(A), 'b-')
```

20.22 *Superdirectivity, Superresolution, Superoscillations*

The concepts of superdirectivity or supergain, superresolution, and superoscillations have played a prominent role in recent years. In this section, we discuss their interrelationships using prolate spheroidal wave functions (PSWF) as our primary tool. A comprehensive review of superresolution concepts may be found in [1545].

Superdirectivity

It was alluded in the beginning of the previous section that it is possible, at least theoretically, to achieve any desired wavenumber pattern from a finite aperture, even one that has infinite directivity. For example, in the 1-D case, such ideal pattern would be a delta-function, $\hat{A}(k_x) = \delta(k_x)$, at least within the visible region.

To accomplish such goal, one may expand the desired pattern $\hat{A}(k_x)$ into a complete and orthogonal set of basis functions in k_x space, determine the expansion coefficients, and through an inverse spatial Fourier transform, construct the corresponding aperture-limited distribution $A(x)$. The prolate spheroidal wave functions (and the generalized ones for the 2-D case), mentioned in Sec. 20.12, are a convenient set of basis functions.

Examples of this procedure can be found in [1571,1487,1489,1653] for 2-D apertures, in [1615,1616] for 1-D apertures, and in [1447,1609–1613] for optimum aperture distributions under reactive-energy constraints.

Here, we consider only the 1-D case over a finite aperture $[-a, a]$. With $k_x = k \sin\theta$, the visible region is the wavenumber interval $[-k, k]$. We recall from Sec. 20.12 (see also Appendix J) that the PSWF basis functions, $\psi_n(k, a, k_x)$, are completely defined relative to the two intervals $[-a, a]$ in x-space, and $[-k, k]$ in k_x-space, with the space-bandwidth product, $c = ka$.[†] They are self-invariant under Fourier transformation,

$$\psi_n(k,a,k_x) = \int_{-a}^{a} \frac{1}{a\mu_n}\, \psi_n\Big(k, a, \frac{x}{a}k_0\Big)\, e^{jk_x x}\, dx\,, \qquad \mu_n = i^n \sqrt{\frac{2\pi\lambda_n}{c}} \tag{20.22.1}$$

for $n \ge 0$, and satisfy the dual orthogonality and completeness properties,

$$\begin{aligned} \int_{-\infty}^{\infty} \psi_n(k,a,k_x)\, \psi_m(k,a,k_x)\, dk_x &= \delta_{nm} \\ \int_{-k}^{k} \psi_n(k,a,k_x)\, \psi_m(k,a,k_x)\, dk_x &= \lambda_n \delta_{nm} \end{aligned} \tag{20.22.2}$$

[†] in this section, we take, $k_0 = k = 2\pi/\lambda$.

$$\sum_{n=0}^{\infty} \frac{1}{\lambda_n} \psi_n(k,a,k_x)\, \psi_n(k,a,k'_x) = \delta(k_x - k'_x)\,, \quad \text{for } k_x, k'_x \in [-k,k] \tag{20.22.3}$$

Any (square-integrable) aperture-limited function $A(x)$ over $[-a,a]$ and its Fourier transform $\hat{A}(k_x)$, can be expanded as linear combinations of the ψ_n functions, with the coefficients determined from the orthogonality properties. Thus, as in Eq. (20.12.23),

$$\begin{aligned} \hat{A}(k_x) &= \int_{-a}^{a} A(x)\, e^{jk_x x}\, dx = \sum_{n=0}^{\infty} c_n\, \psi_n(k,a,k_x) \\ A(x) &= \int_{-\infty}^{\infty} \hat{A}(k_x)\, e^{jk_x x}\, \frac{dk_x}{2\pi} = \sum_{n=0}^{\infty} \frac{c_n}{a\mu_n}\, \psi_n\Big(k,a,\frac{x}{a}k\Big) \cdot \chi_a(x) \\ c_n &= \int_{-\infty}^{\infty} \hat{A}(k_x)\, \psi_n(k,a,k_x)\, dk_x = \frac{1}{\lambda_n} \int_{-k}^{k} \hat{A}(k_x)\, \psi_n(k,a,k_x)\, dk_x \end{aligned} \tag{20.22.4}$$

where $\chi_a(x) = u(a - |x|)$ is the indicator function of the interval $[-a,a]$.

We see right away that if we wish to obtain a sharp delta-function wavenumber pattern , $\hat{A}(k_x) = \delta(k_x)$, over the visible region $-k \le k_x \le k$, then the last expression in (20.22.4) gives for the coefficients,

$$c_n = \frac{1}{\lambda_n} \int_{-k}^{k} \hat{A}(k_x)\, \psi_n(k,a,k_x)\, dk_x = \frac{1}{\lambda_n} \int_{-k}^{k} \delta(k_x)\, \psi_n(k,a,k_x)\, dk_x\,, \quad \text{or,}$$

$$c_n = \frac{1}{\lambda_n}\, \psi_n(k,a,0)\,, \qquad n \ge 0 \tag{20.22.5}$$

which when inserted back into (20.22.4) gives,

$$\begin{aligned} \hat{A}(k_x) &= \sum_{n=0}^{\infty} \frac{1}{\lambda_n}\, \psi_n(k,a,0)\, \psi_n(k,a,k_x) = \delta(k_x)\,, \quad \text{for } |k_x| \le k \\ A(x) &= \sum_{n=0}^{\infty} \frac{1}{a\mu_n\lambda_n}\, \psi_n(k,a,0)\, \psi_n\Big(k,a,\frac{x}{a}k\Big) \cdot \chi_a(x) \end{aligned} \tag{20.22.6}$$

This is also a consequence of the completeness property (20.22.3) by setting $k'_x = 0$. The remarkable feature of (20.22.6) is that it expresses the inverse Fourier transform of a delta-function not as a constant, but as a sum of aperture-limited terms. This is because $\hat{A}(k_x)$ is not equal to $\delta(k_x)$ for all k_x, but only over the visible region. In fact, outside the visible region, $\hat{A}(k_x)$ has incredibly large values, just like the designs of the previous section.

Eq. (20.22.6) is an example of superoscillation [1619–1631]. Indeed, if we think of k_x as "time" and x as "frequency", then $A(x)$ represents a "bandlimited spectrum," with maximum frequency $x = a$, whereas $\hat{A}(k_x) = \delta(k_x)$ represents an extremely fast, instantaneous, "time" variation over a limited "time" interval $[-k,k]$, which is infinitely faster than the maximum "frequency" contained in $A(x)$. See also Example J.3 of Appendix J for a related discussion.

Since the eigenvalues λ_n decrease rapidly to zero, to make numerical sense of (20.22.6) we must regularize it, for example, by keeping only a finite number of terms, say, $M+1$ terms, (see Appendix J, Eq. (J.44) for alternative regularizations),

$$\begin{aligned} \hat{A}_M(k_x) &= \sum_{n=0}^{M} \frac{1}{\lambda_n}\, \psi_n(k,a,0)\, \psi_n(k,a,k_x) \equiv \delta_M(k_x)\,, \quad \text{for } |k_x| \le k \\ A_M(x) &= \sum_{n=0}^{M} \frac{1}{a\mu_n\lambda_n}\, \psi_n(k,a,0)\, \psi_n\Big(k,a,\frac{x}{a}k\Big) \cdot \chi_a(x) \end{aligned} \tag{20.22.7}$$

We show next that the coefficients c_n given by (20.22.5) are also those that maximize the directivity [1615,1616], which in fact becomes infinite in the limit $M \to \infty$. We recall from Sec. 20.12 that the true, approximate, and limiting directivity, and Taylor's superdirectivity ratio γ, are defined in terms of $\hat{A}(k_x)$ by,

$$\begin{aligned} \frac{D_{\text{true}}}{2\pi k} &= \frac{|\hat{A}(0)|^2}{\displaystyle\int_{-k}^{k} |\hat{A}(k_x)|^2\, \frac{k_z}{k}\, dk_x} \\ \frac{D}{2\pi k} &= \frac{|\hat{A}(0)|^2}{\displaystyle\int_{-k}^{k} |\hat{A}(k_x)|^2\, dk_x}\,, \qquad \frac{D_\infty}{2\pi k} = \frac{|\hat{A}(0)|^2}{\displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x} \\ \gamma &= \frac{D}{D_\infty} = \frac{\displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x}{\displaystyle\int_{-k}^{k} |\hat{A}(k_x)|^2\, dk_x} \end{aligned} \tag{20.22.8}$$

As we argued in Sec. 20.12, D_{true} is approximately equal to D for highly forward-peaked patterns. Therefore, we will work with D. Optimizations based on D_{true} have been considered by Rhodes [1447,1609–1613]. Inserting the general expansion (20.22.4) in these expressions and using the orthogonality properties, we may rewrite D, D_∞, γ in terms of the coefficients c_n,

$$\begin{aligned} \frac{D}{2\pi k} &= \frac{\Big|\displaystyle\sum_{n=0}^{\infty} c_n\, \psi_n(k,a,0)\Big|^2}{\displaystyle\sum_{n=0}^{\infty} \lambda_n |c_n|^2}\,, \qquad \frac{D_\infty}{2\pi k} = \frac{\Big|\displaystyle\sum_{n=0}^{\infty} c_n\, \psi_n(k,a,0)\Big|^2}{\displaystyle\sum_{n=0}^{\infty} |c_n|^2} \\ \gamma &= \frac{D}{D_\infty} = \frac{\displaystyle\sum_{n=0}^{\infty} |c_n|^2}{\displaystyle\sum_{n=0}^{\infty} \lambda_n |c_n|^2} \end{aligned} \tag{20.22.9}$$

It is shown in Problem 20.4 that, as expected, the maximization of D_∞ leads to the maximum value $D_\infty = 2\pi(2a)/\lambda = 2ak$, and the uniform distribution $A(x) = u(a - |x|)$, and sinc-function pattern, $\hat{A}(k_x) = 2a\,\sin(k_x a)/(k_x a)$.

On the other hand, the maximization of D leads to the delta-function pattern $\delta(k_x)$. Starting with a finite number of terms M and later taking the limit $M \to \infty$, we have the maximization problem for D,

$$\frac{D}{2\pi k} = \frac{\left| \sum_{n=0}^{M} c_n \psi_n(k,a,0) \right|^2}{\sum_{n=0}^{M} \lambda_n |c_n|^2} = \max$$

Applying the Cauchy-Schwarz inequality,

$$\left| \sum_{n=0}^{M} f_n g_n^* \right|^2 \le \left(\sum_{n=0}^{M} |f_n|^2 \right) \left(\sum_{n=0}^{M} |g_n|^2 \right) \tag{20.22.10}$$

with the choices, $f_n = \lambda_n^{1/2} c_n$, and, $g_n = \lambda_n^{-1/2} \psi_n^*(k,a,0) = \lambda_n^{-1/2} \psi_n(k,a,0)$, noting that the ψ_n functions are real, we have,

$$\frac{D}{2\pi k} = \frac{\left| \sum_{n=0}^{M} c_n \psi_n(k,a,0) \right|^2}{\sum_{n=0}^{M} \lambda_n |c_n|^2} \le \sum_{n=0}^{M} \frac{1}{\lambda_n} \psi_n^2(k,a,0) \tag{20.22.11}$$

with equality reached in (20.22.10) when $f_n = g_n$, or, in our case,

$$\lambda_n^{1/2} c_n = \lambda_n^{-1/2} \psi_n(k,a,0) \quad \Rightarrow \quad c_n = \frac{1}{\lambda_n} \psi_n(k,a,0)\,, \quad 0 \le n \le M \tag{20.22.12}$$

These agree with (20.22.5), and the maximum value of Eq. (20.22.11) is the value of $\hat{A}_M(k_x) = \delta_M(k_x)$ at $k_x = 0$. In the limit $M \to \infty$ we obtain infinite directivity since, $\delta(k_x) = \infty$ at $k_x = 0$. For a finite M, since the eigenvalues are in decreasing order, $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_M$, it follows that the superdirectivity ratio γ will be bounded by,

$$\gamma_M = \frac{\sum_{n=0}^{M} |c_n|^2}{\sum_{n=0}^{M} \lambda_n |c_n|^2} \quad \Rightarrow \quad \frac{1}{\lambda_0} \le \gamma_M \le \frac{1}{\lambda_M} \tag{20.22.13}$$

The upper bound, $1/\lambda_M$, tends to infinity as $M \to \infty$. The lower bound, $1/\lambda_0$, is realized when $c_0 = 1$ and $c_n = 0$, for $n \ge 1$, that is, when $\hat{A}(k_x) = \psi_0(k,a,k_x)$, which is recognized as the optimum prolate apodizer of Sec. 20.13. On the other hand, in the optimum superdirective case with c_n given by (20.22.12), we have,

$$\gamma_M = \frac{\sum_{n=0}^{M} \frac{1}{\lambda_n^2} \psi_n^2(k,a,0)}{\sum_{n=0}^{M} \frac{1}{\lambda_n} \psi_n^2(k,a,0)} \tag{20.22.14}$$

Example 20.22.1: This is a variation of Example J.3 of Appendix J. Figure 20.22.1 plots the approximations (20.22.7) for the cases $M = 20$ and $M = 30$. The space-bandwidth product was $c = ka = 4\pi$, where for convenience we chose $k = 1$ and $a = 4\pi$. We observe how $\hat{A}_M(k_x)$ resembles $\delta(k_x)$ within $[-k,k]$ as M increases. Outside the visible region the values of $\hat{A}_M(k_x)$ are extremely large. This is also reflected in the aperture distribution $A_M(x)$ which undergoes huge variations over $[-a,a]$, just like the Woodward-Lawson example of Fig. 20.21.5. The MATLAB code for this example was,

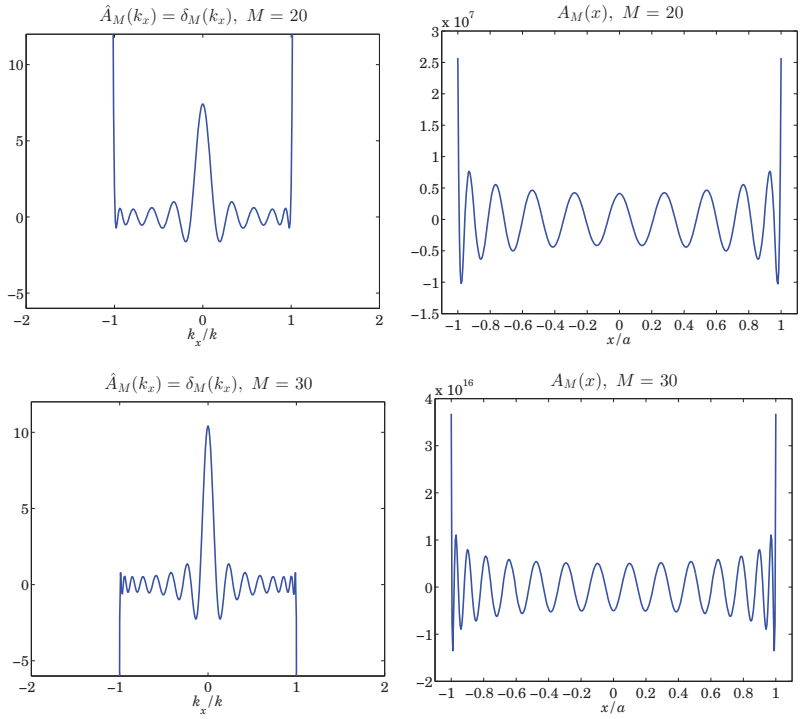


Fig. 20.22.1 Optimum superdirective aperture, with $M = 20, 30$.

```
c=4*pi;
k=1; a=c;

kx = linspace(-2,2,401);     % kx in units of k
x = linspace(-1,1,401);      % x in units of a

for M = [20,30]

   [Psi,La] = pswf(k,a,M,kx);

   Psi0 = Psi(:,kx==0);

   F = Psi0' * diag(1./La) * Psi;
```

```
    figure; plot(kx, F, 'b-');  ylim([-10,10]);

    % ---- aperture distribution ----

    [Psi,La] = pswf(k,a,M,x*k);

    Psi0 = Psi(:,x==0);

    n = (0:M)';
    Mu = i.^n .* sqrt(2*pi*La/c);

    A = Psi0' * diag(1./(a*Mu.*La)) * Psi;        % calculate A(x)

    figure; plot(x, A, 'b-');

end
```

The maximum directivity and ratio γ_M were calculated from Eqs. (20.22.7) and (20.22.14), resulting in $D/2\pi k = 7.4187$ and 10.4232, and $\gamma_M = 3.28\times10^{14}$ and 3.19×10^{32}, for $M = 20$ and 30, respectively. □

Example 20.22.2: Here, we show how to design the Woodward-Lawson example of Fig. 20.21.5 using the PSWF approach. We choose the same parameters, $a = \lambda/2$, gain $g = 2$, and for convenience, we set $k = 1$ again, so that, $c = ak = 2\pi a/\lambda = \pi$. The desired wavenumber pattern is taken to be a compressed sinc-pattern by the superresolution factor g, but only over the extent of its mainlobe, setting it to zero outside the mainlobe, that is,

$$\hat{A}(k_x) = \frac{\sin(gk_x a)}{gk_x a}\cdot u(\pi - |gk_x a|) \quad\Rightarrow\quad F_g(v) = \frac{\sin(gv)}{gv}\cdot u(\pi - |gv|) \tag{20.22.15}$$

where the unit-step restricts the non-zero values to the range $|gk_x a| \le \pi$. The graph on the left of Fig. 20.22.2 shows the designed pattern using $M = 10$. The designed mainlobe is virtually indistinguishable from that of Eq. (20.22.15).

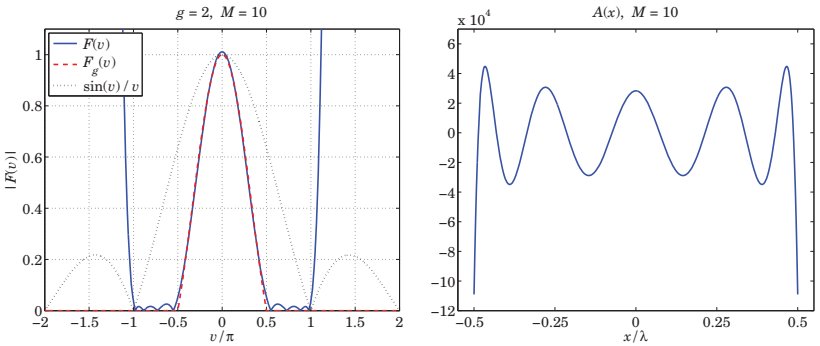


Fig. 20.22.2 Compressed sinc, with gain $g = 2$, and $M = 10$.

The values outside the visible region $[-k, k]$ are huge. The graph on the right shows the aperture distribution, which exhibits the same type of large variations like the ones

in Fig. 20.21.5. The $M+1$ expansion coefficients c_n were calculated numerically from Eq. (20.22.4), in which (20.22.15) was used, for $n = 0, 1, \ldots, M$,

$$c_n = \frac{1}{\lambda_n}\int_{-k}^{k}\hat{A}(k_x)\,\psi_n(k,a,k_x)\,dk_x = \frac{1}{\lambda_n}\int_{-\pi/ga}^{\pi/ga}\frac{\sin(gk_x a)}{gk_x a}\,\psi_n(k,a,k_x)\,dk_x$$
$$\hat{A}_M(k_x) = \sum_{n=0}^{M} c_n\,\psi_n(k,a,k_x)\,, \qquad A_M(x) = \sum_{n=0}^{M}\frac{c_n}{a\mu_n}\,\psi_n\Big(k,a,\frac{x}{a}k\Big)\cdot\chi_a(x) \tag{20.22.16}$$

The integrations in c_n were done with the tanh-sinh quadrature function **quadts**. The following MATLAB code segment illustrates the computation,

```
k = 1; lambda = 2*pi/k;
a = lambda/2;
c = k*a;                          % c = pi
g = 2;
M = 10;

[wq,kq] = quadts(-1/g, 1/g);      % quadts weights and evaluation points

[Psq,La] = pswf(k,a,M,kq);        % PSWFs evaluated at kq

for n=1:M+1
   cn(n) = 1/La(n) * (sinc(g*a*kq/pi)'.*Psq(n,:)) * wq;    % coefficients c_n
end

u = linspace(-2,2,2001); v = pi*u; kx = v/a;        % use v-units for display

[Psi,La] = pswf(k,a,M,kx);              % evaluate PSWFs at kx = pi*v

F = abs(cn * Psi);                      % evaluate pattern F(v)

Fg = abs(sinc(g*u).*(abs(u)<=1/g));     % compressed truncated sinc pattern

figure; plot(u,F,'b-', u,Fg,'r--'); ylim([0,1]);

% ---- aperture distribution ----

x = linspace(-1,1,401)*a;               % x in units of a

[Psi,La] = pswf(k,a,M,x*k/a);           % PSWFs for aperture distribution

n = (0:M)';
Mu = i.^n .* sqrt(2*pi*La/c);

A = cn * diag(1./(a*Mu)) * Psi;         % evaluate A(x)

figure; plot(x/lambda, A, 'b-');
```

The computed coefficients c_n were,

n	c_n	n	c_n
0	5.4139×10^{-1}	6	-1.8987×10^{2}
1	0	7	0
2	-5.9918×10^{-1}	8	5.2521×10^{3}
3	0	9	0
4	7.3039×10^{0}	10	1.5445×10^{5}
5	0		

Because (20.22.15) is even in k_x and the ψ_n are odd when n is odd, it follows that all the c_n vanish for odd n. □

In the present aperture antenna context, the field of view was taken to be the entire visible region $[-k,k]$. The giant sidelobes residing in the invisible region $|k_x|>k$ are associated with large reactive as opposed to radiated power and are responsible for the extremely large values of the superdirectivity ratio γ, which provides a measure of the proportion of reactive to radiated power,

$$\gamma-1=\frac{\displaystyle\int_{-\infty}^{\infty}|\hat{A}(k_x)|^2\,dk_x}{\displaystyle\int_{-k}^{k}|\hat{A}(k_x)|^2\,dk_x}-1=\frac{\displaystyle\int_{|k_x|>k}|\hat{A}(k_x)|^2\,dk_x}{\displaystyle\int_{|k_x|\le k}|\hat{A}(k_x)|^2\,dk_x}=\frac{\text{reactive}}{\text{radiated}}$$

The visible region in, $u=k_xa/\pi$, units is $|u|\le 2a/\lambda$. For typical microwave antenna applications where the antenna size is of the order of λ, the upper limit $2a/\lambda$ is of the order of unity. On the other hand, in astronomical applications involving telescopes with radii of orders of meters and visible light in nanometers, the upper limit $2a/\lambda$ is a very large. In such cases, the Toraldo di Francia type designs of Sec. 20.21 have a much more restrictive field of view than the visible region.

Spatial Superresolution

The terms "resolution" and "superresolution" were used in the apodization and superdirectivity contexts to mean "frequency resolution" with the goal of obtaining sharp wavenumber spectral peaks in $\boldsymbol{k}_\perp$ space from a *space-limited* aperture. The same terms are also used in a dual sense in the literature to mean "spatial resolution" with the goal of obtaining sharp spatial features from a *wavenumber-limited* system.

All diffraction-limited optical systems are characterized by a finite wavenumber bandwidth, say, k_B, which effectively limits the spatial details to be greater than approximately $1/k_B$. Superresolution refers in this context to the attempt to extrapolate and recover those spatial frequencies higher than k_B that were present in the input before passing through the optical system so that sharper spatial details can be recovered.

There is a vast literature on this subject and our discussion in this section cannot possibly do justice to it. The reader is referred to the reviews [1545,1548] for further references.

A prototypical example of a diffraction-limited optical system is the 4F system discussed in Sec. 20.11. Its one-dimensional version is depicted below in Fig. 20.22.3. The operation of the system is described by Eqs. (20.11.13)–(20.11.15), or,

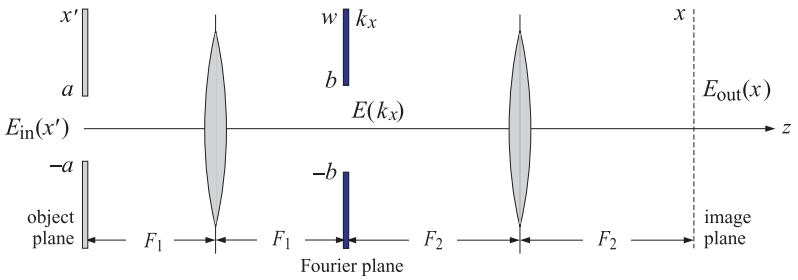


Fig. 20.22.3 Bandlimited optical system.

$$E(k_x)=e^{-2jkF_1}\hat{E}_{\text{in}}(k_x)=e^{-2jkF_1}\int_{-\infty}^{\infty}E_{\text{in}}(x')\,e^{jk_xx'}\,dx'\Big|_{k_x=\frac{kw}{F_1}}$$

$$\begin{aligned}E_{g,\text{out}}(x)&=e^{-2jk(F_1+F_2)}\int_{-\infty}^{\infty}H(k_x)\,\hat{E}_{\text{in}}(k_x)\,e^{-jk_xx}\,\frac{dk_x}{2\pi}\\&=e^{-2jk(F_1+F_2)}\int_{-\infty}^{\infty}h(x-x')\,E_{\text{in}}(x')\,dx'\end{aligned}\qquad(20.22.17)$$

They describe the outputs at the Fourier plane and at the image plane, where w denotes the x-coordinate on the Fourier plane, and the filter $H(k_x)$ is evaluated at the wavenumber, $k_x=kw/F_1$. The impulse response $h(x)$ is the inverse Fourier transform,

$$h(x)=\int_{-\infty}^{\infty}H(k_x)\,e^{-jk_xx}\,\frac{dk_x}{2\pi}\qquad(20.22.18)$$

and, $E_{g,\text{out}}(x)$ is the geometric-optics projected output $E_{\text{out}}(x)$ back onto the object plane as defined in Eq. (20.11.11), that is,

$$E_{g,\text{out}}(x)=e^{-j\pi/2}\sqrt{\frac{F_2}{F_1}}\,E_{\text{out}}\Big(-\frac{F_2}{F_1}\,x\Big)\qquad(20.22.19)$$

If at the Fourier plane we assume a slit opening with support $-b\le w\le b$, then the transfer function of the aperture can be defined in terms of the unit-step, $F(w)=u(b-|w|)$. In k_x units the filter transfer function will represent an *ideal lowpass filter* with cutoff, $k_B=kb/F_1$, indeed,

$$H(k_x)=F(w)\Big|_{k_x=\frac{kw}{F_1}}=u\Big(b-\frac{|k_x|F_1}{k}\Big)=u\big(k_B-|k_x|\big)\,,\qquad\boxed{k_B=k\,\frac{b}{F_1}=k\cdot N_A}$$

where $N_A=b/F_1$ is the numerical aperture of the opening. Thus, k_B represents the maximum spatial frequency let through by the Fourier plane. The corresponding impulse response $h(x)$ will have the usual sinc form of an ideal lowpass filter,

$$h(x)=\int_{-\infty}^{\infty}H(k_x)\,e^{-jk_xx}\,\frac{dk_x}{2\pi}=\int_{-k_B}^{k_B}e^{-jk_xx}\,\frac{dk_x}{2\pi}=\frac{\sin(k_Bx)}{\pi x}\qquad(20.22.20)$$

We will also assume that the input $E_{\rm in}(x)$ has *finite* extent, say, $-a \le x \le a$, vanishing outside that interval, as depicted by the input slit $[-a,a]$ in Fig. 20.22.3. Under this assumption, Eqs. (20.22.17) take the simpler form,

$$\begin{aligned} E(k_x) &= \hat{E}_{\rm in}(k_x) = \int_{-a}^{a} E_{\rm in}(x')\, e^{jk_x x'}\, dx' \\ E_{g,\rm out}(x) &= \int_{-k_B}^{k_B} \hat{E}_{\rm in}(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = \int_{-a}^{a} \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')}\, E_{\rm in}(x')\, dx' \end{aligned} \tag{20.22.21}$$

where we have also ignored the phase factors e^{-2jkF_1}, e^{-2jkF_2} which represent the propagation delays across the two $2F$ planes, It is evident that the output, $E_{g,\rm out}(x)$, is bandlimited to the range, $|k_x| \le k_B$, with a Fourier transform,

$$\hat{E}_{g,\rm out}(k_x) = u(k_B - |k_x|) \cdot \hat{E}_{\rm in}(k_x) = \begin{cases} \hat{E}_{\rm in}(k_x)\,, & |k_x| \le k_B \\ 0\,, & |k_x| > k_B \end{cases} \tag{20.22.22}$$

Recovering the full $\hat{E}_{\rm in}(k_x)$, for all $-\infty < k_x < \infty$, from the knowledge of $\hat{E}_{g,\rm out}(k_x)$ would appear to be impossible. And it is, in general. However, because we assumed a finite object, $E_{\rm in}(x) \neq 0$ only for $|x| \le a$, its Fourier transform $\hat{E}_{\rm in}(k_x)$ given in (20.22.21) will be an analytic function of k_x, and as such, it can be recovered in principle from any finite segment of k_x, such as, $|k_x| \le k_B$.

The practical implementation of this approach is difficult since the analytic continuation process is very sensitive to noise. A more promising approach would be to try to invert the integral equation in (20.22.21). The PSWF basis functions provide an ideal tool for this purpose since they are eigenfunctions of that integral operator. However, even this approach represents an ill-posed problem that requires some sort of regularization for its implementation.

To apply the PSWF functions, we must make the following mapping between the time-frequency variables (t,ω) of Appendix J and the spatial variables (x,k_x), essentially interchanging the roles of x and k_x in the definitions of Sec. 20.12,

$$\begin{aligned} t_0\,,\ \omega_0\,,\ t\,,\ \omega \quad &\Rightarrow \quad a\,,\ k_B\,,\ x\,,\ k_x \\ c = t_0\omega_0 \quad &\Rightarrow \quad c = ak_B \\ \psi_n(t_0,\omega_0,t) \quad &\Rightarrow \quad \psi_n(a,k_B,x) \end{aligned} \tag{20.22.23}$$

Specifically, the functions $\psi_n(a,k_B,x)$ are defined in terms of the standard scaled PSWF function $\phi_n(c,\eta)$ of Eq. (J.7) as follows,

$$\psi_n(a,k_B,x) = \frac{1}{\sqrt{a}}\, \phi_n\left(c, \frac{x}{a}\right)\,, \quad c = ak_B \tag{20.22.24}$$

For the rest of this section, we will denote $\psi_n(a,k_B,x)$ by $\psi_n(x)$. From Appendix J, we know that the functions $\psi_n(x)$ are eigenfunctions of the sinc integral operator in (20.22.21), and satisfy the following dual orthogonality properties on the finite interval $[-a,a]$ and over the infinite interval,

$$\begin{aligned} &\int_{-a}^{a} \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')}\, \psi_n(x')\, dx' = \lambda_n\, \psi_n(x)\,, \quad \text{for all } x \\ &\int_{-\infty}^{\infty} \psi_n(x)\, \psi_m(x)\, dx = \delta_{nm}\,, \quad \int_{-a}^{a} \psi_n(x)\, \psi_m(x)\, dx = \lambda_n \delta_{nm} \end{aligned} \tag{20.22.25}$$

for $n,m \ge 0$. Moreover, they satisfy the following completeness properties over the finite and infinite intervals,

$$\begin{aligned} &\sum_{n=0}^{\infty} \frac{1}{\lambda_n}\, \psi_n(x)\, \psi_n(x') = \delta(x-x')\,, \quad \text{for } x,x' \in [-a,a] \\ &\sum_{n=0}^{\infty} \psi_n(x)\, \psi_n(x') = \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')}\,, \quad \text{for all } x,x' \end{aligned} \tag{20.22.26}$$

The bandlimited spatial Fourier transform of $\psi_n(x)$ is,

$$\hat{\psi}_n(k_x) = \int_{-\infty}^{\infty} \psi_n(x)\, e^{jk_x x}\, dx = \frac{2\pi}{k_B \mu_n^*}\, \psi_n\left(\frac{k_x}{k_B}\, a\right) \cdot \chi_{k_B}(k_x) \tag{20.22.27}$$

where $\mu_n = i^n \sqrt{2\pi\lambda_n/c}$, and the conjugate μ_n^* appears in order to conform with the opposite exponential sign convention used in spatial as opposed to time Fourier transforms, and, $\chi_{k_B}(k_x) = u(k_B - |k_x|)$, is the indicator function of the interval $[-k_B,k_B]$.

An arbitrary square-integrable function $E(x)$ with a Fourier transform $\hat{E}(k_x)$ that is bandlimited over $[-k_B,k_B]$ can be expanded in the basis functions $\psi_n(x)$,

$$\begin{aligned} E(x) &= \sum_{n=0}^{\infty} c_n \psi_n(x) \\ \hat{E}(k_x) &= \sum_{n=0}^{\infty} c_n\, \frac{2\pi}{k_B \mu_n^*}\, \psi_n\left(\frac{k_x}{k_B}\, a\right) \cdot \chi_{k_B}(k_x) \end{aligned} \tag{20.22.28}$$

with expansion coefficients computed over the finite or infinite intervals,

$$c_n = \frac{1}{\lambda_n} \int_{-a}^{a} \psi_n(x) E(x)\, dx = \int_{-\infty}^{\infty} \psi_n(x) E(x)\, dx\,, \quad n \ge 0 \tag{20.22.29}$$

Because $E(x)$ is bandlimited, it can also be expanded in the sinc-basis of the sampling theorem, where, $d = \pi/k_B$ denotes the Nyquist sampling space interval,

$$\begin{aligned} E(x) &= \frac{\pi}{k_B} \sum_{n=-\infty}^{\infty} E(nd)\, \frac{\sin\big(k_B(x-nd)\big)}{\pi(x-nd)} \\ \hat{E}(k_x) &= \frac{\pi}{k_B} \sum_{n=-\infty}^{\infty} E(nd)\, e^{-jk_x nd} \cdot \chi_{k_B}(k_x) \end{aligned} \tag{20.22.30}$$

Since the function $E(x)$ is completely characterized by its space samples $E(nd)$ at multiples of d, we may think of d as defining the spatial resolution for $E(x)$. We note also that d represents the *Rayleigh or Abbe resolution* distance, that is, the distance corresponding to the first null of the impulse response (20.22.20), which is the solution of, $\sin(k_B x) = 0$, or, $k_B d = \pi$, or, $d = \pi/k_B = \lambda/2N_A$,

$$\boxed{d = \frac{\lambda}{2N_A}} \qquad \text{(Abbe diffraction limit)} \tag{20.22.31}$$

The inversion of the integral operation in (20.22.21), its formulation using PSWFs, and its application to the image restoration problem was considered first by Barnes [1569]. We discussed it in some detail in Appendix J, including several regularization schemes and the sensitivity to noise. The inverse is given formally by,

$$\begin{aligned} E_{g,\text{out}}(x) &= \int_{-a}^{a} \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')}\, E_{\text{in}}(x')\, dx' \quad \Rightarrow \\ E_{\text{in}}(x) &= \int_{-a}^{a} K(x,x')\, E_{g,\text{out}}(x')\, dx' \end{aligned} \tag{20.22.32}$$

where $K(x,x')$ is the formal inverse of the sinc kernel over the finite interval $[-a,a]$,

$$K(x,x') = \sum_{n=0}^{\infty} \frac{1}{\lambda_n^2}\, \psi_n(x)\, \psi_n(x')\,, \quad \text{for } x,x'' \in [-a,a] \tag{20.22.33}$$

satisfying,

$$\int_{-a}^{a} K(x,x'')\, \frac{\sin\big(k_B(x''-x')\big)}{\pi(x''-x')}\, dx'' = \delta(x-x')\,, \quad \text{for } x,x' \in [-a,a] \tag{20.22.34}$$

A regularized version is obtained by truncating the summation (20.22.33) to $n = M$, where M is of the order of the Shannon number $N_c = 2c/\pi$.

An alternative way to invert (20.22.32) is to express the input in terms of the expansion coefficients of the output. Since $E_{g,\text{out}}(x)$ is bandlimited, we may apply (20.22.28), calculate its expansion coefficients, and recover $E_{\text{in}}(x)$ as follows,

$$\begin{aligned} b_n &= \frac{1}{\lambda_n} \int_{-a}^{a} \psi_n(x) E_{g,\text{out}}(x)\, dx = \int_{-\infty}^{\infty} \psi_n(x) E_{g,\text{out}}(x)\, dx\,, \quad n \geq 0 \\ E_{g,\text{out}}(x) &= \sum_{n=0}^{\infty} b_n \psi_n(x) \quad \Rightarrow \quad E_{\text{in}}(x) = \sum_{n=0}^{\infty} \frac{1}{\lambda_n} b_n \psi_n(x) \end{aligned} \tag{20.22.35}$$

Eq. (20.22.35) can be verified easily, using the eigenvalue equation (20.22.25),

$$\int_{-a}^{a} \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')}\, E_{\text{in}}(x')\, dx' = \int_{-a}^{a} \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')} \left[\sum_{n=0}^{\infty} \frac{1}{\lambda_n} b_n \psi_n(x)\right] =$$

$$= \sum_{n=0}^{\infty} \frac{1}{\lambda_n} b_n \cdot \int_{-a}^{a} \frac{\sin\big(k_B(x-x')\big)}{\pi(x-x')} \psi_n(x)\, dx' = \sum_{n=0}^{\infty} \frac{1}{\lambda_n} b_n \cdot \lambda_n \psi_n(x) = E_{g,\text{out}}(x)$$

Since the eigenvalues λ_n decrease rapidly to zero beyond the Shannon number N_c, the only numerically meaningful coefficients b_n/λ_n are those over the range $0 \leq n \leq M$, where M is of the order of N_c. Thus, the effective number of *degrees of freedom* of the image can be taken to be the Shannon number N_c [1485]. Truncating, the expansions (20.22.35), we obtain the approximate inverse,

$$E_{\text{in}}(x) \approx \sum_{n=0}^{M} \frac{1}{\lambda_n} b_n \psi_n(x) \tag{20.22.36}$$

As an example, consider an impulsive input, $E_{\text{in}}(x) = \delta(x)$, for $-a \leq x \leq a$. Then, the output would be the bandlimited impulse response $E_{g,\text{out}}(x) = h(x)$ of (20.22.18). The required expansions can be derived directly from (20.22.35), or as special cases of Eq. (20.22.26), setting $x' = 0$, and recognizing that the expansion coefficients are,

$$b_n = \psi_n(0)\,, \quad n \geq 0$$

Thus, we obtain the following expressions for the output $h(x)$, the input $\delta(x)$, and the bandlimited Fourier transform $H(k_x)$ of the output,

$$\begin{aligned} h(x) &= \frac{\sin(k_B x)}{\pi x} = \sum_{n=0}^{\infty} \psi_n(0)\, \psi_n(x) \\ \delta(x) &= \sum_{n=0}^{\infty} \frac{1}{\lambda_n}\, \psi_n(0)\, \psi_n(x)\,, \quad \text{for } x \in [-a,a] \\ H(k_x) &= \chi_{k_B}(k_x) = \sum_{n=0}^{\infty} \frac{2\pi}{k_B \mu_n^*}\, \psi_n(0)\, \psi_n\Big(\frac{k_x}{k_B}\, a\Big) \cdot \chi_{k_B}(k_x) \end{aligned} \tag{20.22.37}$$

The regularized, truncated, approximations are then,

$$\begin{aligned} h_M(x) &= \sum_{n=0}^{M} \psi_n(0)\, \psi_n(x) \\ \delta_M(x) &= \sum_{n=0}^{M} \frac{1}{\lambda_n}\, \psi_n(0)\, \psi_n(x) \\ H_M(k_x) &= \sum_{n=0}^{M} \frac{2\pi}{k_B \mu_n^*}\, \psi_n(0)\, \psi_n\Big(\frac{k_x}{k_B}\, a\Big) \cdot \chi_{k_B}(k_x) \end{aligned} \tag{20.22.38}$$

Example 20.22.3: Fig. 20.22.4 plots Eqs. (20.22.38) for the cases of $M = 5$ and $M = 20$. The PSWF parameters were chosen to be $a = 2\pi$, $k_B = 1$, $c = ak_B = 2\pi$. The MATLAB code for generating these graphs was,

```
a = 2*pi; kB = 1; c = a*kB

x  = linspace(-3,3,601) * a;
kx = linspace(-1.5,1.5,601);

for M = [5,20]
   [Psi,La] = pswf(a,kB,M,x);
   Psi0 = Psi(:,x==0);                        % psi_n(0) values

   h = Psi0' * Psi;                           % output h(x)
   delta = Psi0' * diag(1./La) * Psi;         % estimated input, delta(x)

   figure; plot(x/a,h,'b-', x/a,sinc(kB*x/pi)*kB/pi,'r--');

   figure; plot(x/a,delta, 'b-'); ylim([-0.5,1.5]);

   [Psi,La] = pswf(a,kB,M,kx*a/kB);           % Fourier transform calculation
```

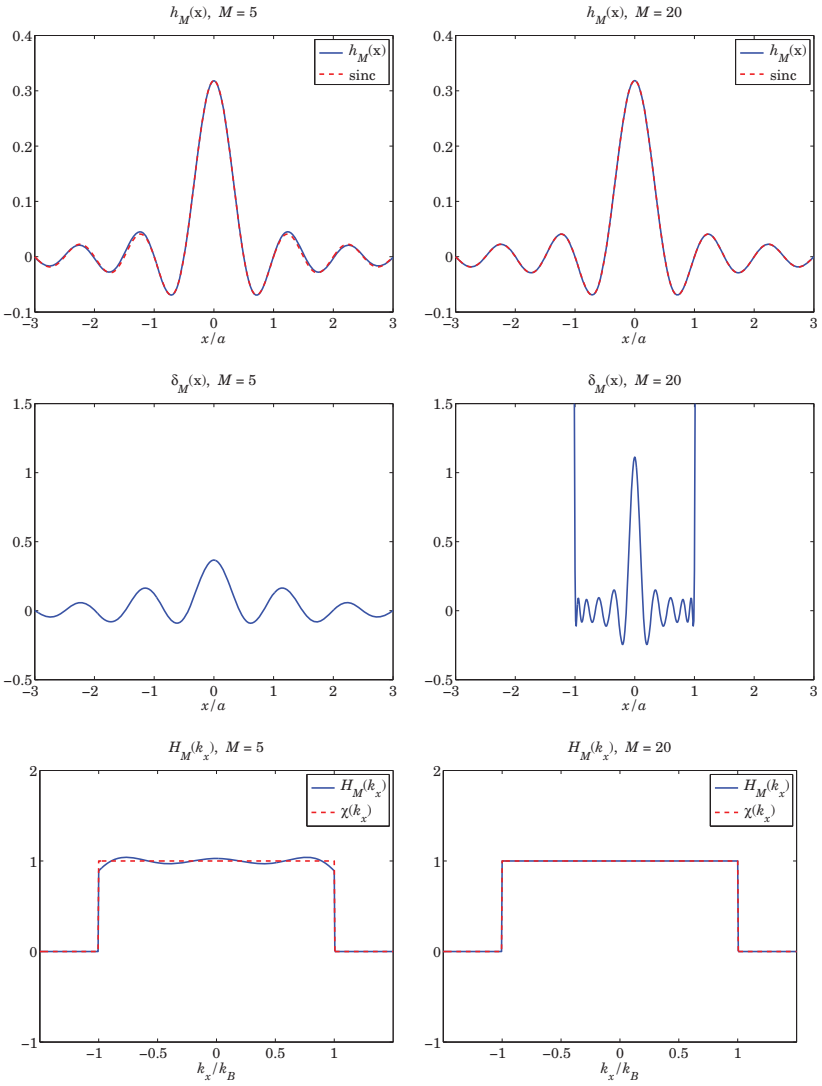


Fig. 20.22.4 Impulse response approximations, for $M = 5$ and $M = 20$.

```
Psi0 = Psi(:,kx==0);

n = (0:M)';
Mu = i.^n .* sqrt(2*pi*La/c);

chi = (abs(kx)<=kB);                % indicator function of [-kB,kB]
```



```
   H = (2*pi/kB) * Psi0' * diag(1./conj(Mu)) * Psi .* chi;     % FT of h(x)

   figure; plot(kx,H, 'b-',  kx,chi,'r--');
end
```

Because the expansion of $h_M(x)$ does not involve divisions by λ_n, it converges very quickly to the theoretical sinc, and the two are barely distinguishable on the graphs. The same is true of the estimated Fourier transform $H_M(k_x)$.

On the other hand, while the values of estimated input $\delta_M(x)$ within the interval $[-a, a]$ resemble a delta-function spike as M increases, its values outside $[-a, a]$ are enormous. Within $[-a, a]$, $\delta_M(x)$ exhibits *superoscillatory* behavior. Indeed, since the maximum spatial frequency contained in $\delta_M(x)$ is k_B, the fastest oscillating sinusoid in $\delta_M(x)$ will be, $\cos(k_B x)$, which cycles by an amount of $k_B a/2\pi = 1$ cycles in the range $[0, a]$ meters. However, one observes four cycles in the range $[0, a]$ in $\delta_M(x)$ for the case $M = 20$. □

All superresolution designs are typically accompanied by extremely large values outside the field of view, and exhibit superoscillatory behavior inside it.

Superoscillations—defined as bandlimited signals that over a particular time period can oscillate faster than their fastest frequency—are typically characterized by similar features, namely, weak values within the superoscillating interval and huge values outside it. More details on superoscillations and their construction and applications may be found in Refs. [1619–1631].

Focusing of Plane Waves

We saw in Sec. 19.2 that the plane-wave spectrum representation is equivalent to Rayleigh-Sommerfeld diffraction theory, and that propagation is equivalent to lowpass filtering, which at far distances (i.e., in the Fresnel or Fraunhofer regimes) has the effect of removing the low-frequency evanescent modes, with only the propagating modes surviving. Thus, similar questions arise as in the $4F$ case whether it is possible to recover lost spatial details from observing the propagated/diffracted waveform.

In the one-dimensional version, the propagation filter for propagating by a distance $z = L \geq 0$ along the z-direction, has transfer function, $e^{-jk_z L}$, where k_z is the evanescent square root,

$$k_z = \begin{cases} \sqrt{k^2 - k_x^2}\,, & \text{if} \quad |k_x| \leq k \\ -j\sqrt{k_x^2 - k^2}\,, & \text{if} \quad |k_x| > k \end{cases} \tag{20.22.39}$$

It follows that for the evanescent modes $|k_x| > k$, the filter response decays exponentially, that is, $e^{-jk_z L} = \exp\left(-L\sqrt{k_x^2 - k^2}\right)$. Thus, the filter acts effectively as a lowpass filter, becoming a better filter with increasing L.

The fields $E_a(x)$ and $E_b(x)$ shown in Fig. 20.22.5 at the two planes $z = z_a$ and $z = z_b$, separated by distance $L = z_b - z_a$, are related by the propagation filter $e^{-jk_z L}$. To determine the output $E_b(x)$ from the input $E_a(x)$, one can first calculate the Fourier transform of $E_a(x)$, then propagate it by the filter, and then reconstruct $E_b(x)$ by inverse Fourier transformation, as outlined in the following steps,

$$\hat{E}_a(k_x) = \int_{-\infty}^{\infty} E_a(x)\, e^{jk_x x}\, dx$$

$$\hat{E}_b(k_x) = \hat{E}_a(k_x)\, e^{-jk_z L} \qquad \text{(forward propagation)}$$

$$E_b(x) = \int_{-\infty}^{\infty} \hat{E}_b(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = \int_{-\infty}^{\infty} \hat{E}_a(k_x)\, e^{-jk_z L} e^{-jk_x x}\, \frac{dk_x}{2\pi}$$

These can be inverted in order to answer the question: what should be the input $E_a(x)$ that would be reshaped into a particular desired output $E_b(x)$ upon propagating by a distance L? This is illustrated in Fig. 20.22.5 where the input is being reshaped into a narrower focal spot.

Inverting, $\hat{E}_b(k_x) = \hat{E}_a(k_x)\, e^{-jk_z L}$, we have, $\hat{E}_a(k_x) = \hat{E}_b(k_x)\, e^{jk_z L}$, so that starting with the desired $E_b(x)$, we may reconstruct $E_a(x)$ as follows,

$$\boxed{\begin{aligned} \hat{E}_b(k_x) &= \int_{-\infty}^{\infty} E_b(x)\, e^{jk_x x}\, dx \\ \hat{E}_a(k_x) &= \hat{E}_b(k_x)\, e^{jk_z L} \\ E_a(x) &= \int_{-\infty}^{\infty} \hat{E}_b(k_x)\, e^{jk_z L} e^{-jk_x x}\, \frac{dk_x}{2\pi} \end{aligned}} \qquad \text{(inverse diffraction)} \qquad (20.22.40)$$

This is the problem of *inverse diffraction*, which has been widely studied in the literature [1584–1590]. Not every $E_b(x)$ can be designed. Since for evanescent modes the inverse filter increases exponentially, $e^{jk_z L} = \exp\left(L\sqrt{k_x^2 - k^2}\right)$, the desired spectrum $\hat{E}_b(k_x)$ must be such that to result into a convergent inverse Fourier integral for $E_a(x)$. For example, a perfect focal spot, $E_b(x) = \delta(x)$, for all x, cannot be designed.

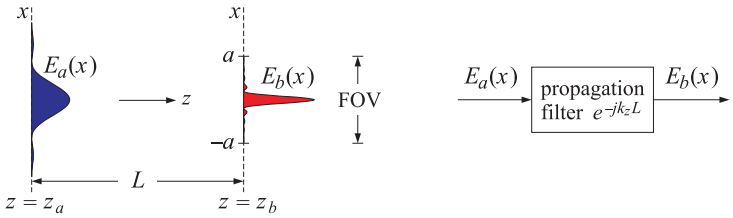


Fig. 20.22.5 Focusing onto a narrow spot.

For arbitrary L, Eqs. (20.22.40) can be implemented numerically. However, certain simplifications take place in the two special cases of a very large L in the far-field, and a very small L in the near-field.

In the first case, since L is large, by the time $E_a(x)$ propagates to $E_b(x)$, the evanescent modes will have died out and would not contribute to $\hat{E}_b(k_x)$, so that the inverse Fourier integral for $E_a(x)$ need only be integrated over the propagating modes, $|k_x| \le k$. In the second case, since L is small and the evanescent modes cannot be ignored, one could try to limit the integration to be only over the evanescent modes by choosing

$\hat{E}_b(k_x)$ to exist only in the evanescent range, while ensuring the convergence of the inverse integral for $E_a(x)$. Below, we consider examples of both cases.

For the fist case, if we assume that L is large enough to ignore all the evanescent modes, then the propagation filter may be thought of as an ideal lowpass filter with cutoff $k_x = k$, and the propagated spectrum $\hat{E}_b(k_x)$ may be taken to be bandlimited over the interval $[-k, k]$. As first proposed in [1554], the machinery of the PSWF functions can then be used to design any desired $E_b(x)$, specified within a finite field of view, say, $-a \le x \le a$, even a narrow $\delta(x)$ restricted over $[-a, a]$.

The design steps require two numerical integrations, one to determine the expansion coefficients into the $\psi_n(x)$ functions, and another one to perform the inverse Fourier transform for $E_a(x)$, with the integration restricted to be only over the $[-k, k]$ interval because of the bandlimited nature of $\hat{E}_b(k_x)$. We start by specifying $E_b(x)$ over $[-a, a]$, then, using $M+1$ basis functions, we have,

$$\boxed{\begin{aligned} &1.\quad b_n = \frac{1}{\lambda_n} \int_{-a}^{a} \psi_n(x) E_b(x)\, dx, \quad n = 0, 1, \ldots, M \\ &2.\quad E_b(x) = \sum_{n=0}^{M} b_n \psi_n(x) \\ &3.\quad \hat{E}_b(k_x) = \sum_{n=0}^{M} b_n \frac{2\pi}{k\mu_n^*} \psi_n\left(\frac{k_x}{k} a\right) \cdot \chi_k(k_x) \\ &4.\quad E_a(x) = \int_{-k}^{k} \hat{E}_b(k_x)\, e^{jk_z L} e^{-jk_x x}\, \frac{dk_x}{2\pi} \end{aligned}} \qquad (20.22.41)$$

where $\psi_n(x)$ are the functions $\psi_n(a, k_B, x)$ with $k_B = k$. Step two provides a bandlimited approximation to $E_b(x)$, and step three calculates its Fourier transform. Step four constructs the input $E_a(x)$, also bandlimited in $[-k, k]$, that would propagate into $E_b(x)$ at distance L.

Example 20.22.4: We consider the same example from [1554], in which the desired focal spot is taken to be a compressed sinc with superresolution gain, $2\pi g = 40$, over the field of view $[-a, a]$ with $a = 0.6\lambda$, that is,

$$E_b(x) = \frac{\sin(gkx)}{gkx} = \frac{\sin(40x/\lambda)}{40x/\lambda}, \quad \text{for } |x| \le a = 0.6\lambda \qquad (20.22.42)$$

where, by contrast, the Abbe diffraction limit corresponds to $g = 1$. Fig. 20.22.6 displays the results of the four design steps in (20.22.41), for $M = 6$ and $M = 20$. The propagation distance was chosen to be $L = 20\lambda$.

The output $E_b(x)$ was plotted only over twice the FOV, $-2a \le x \le 2a$. Outside the FOV, $E_b(x)$ takes enormous values as is typical of all superresolution examples. The input $E_a(x)$, being symmetric in x, was plotted over the range, $0 \le x \le 120\lambda$, and again, we note its enormous values.

Because of the symmetry of $E_b(x)$ the coefficients b_n are non-zero only for even values of n and are listed below, where those for $0 \le n \le 6$ correspond to the case $M = 6$, and continuing for $7 \le n \le 20$ for the case $M = 20$. We note the enormous dynamic range of values, ranging over 18 orders of magnitude,

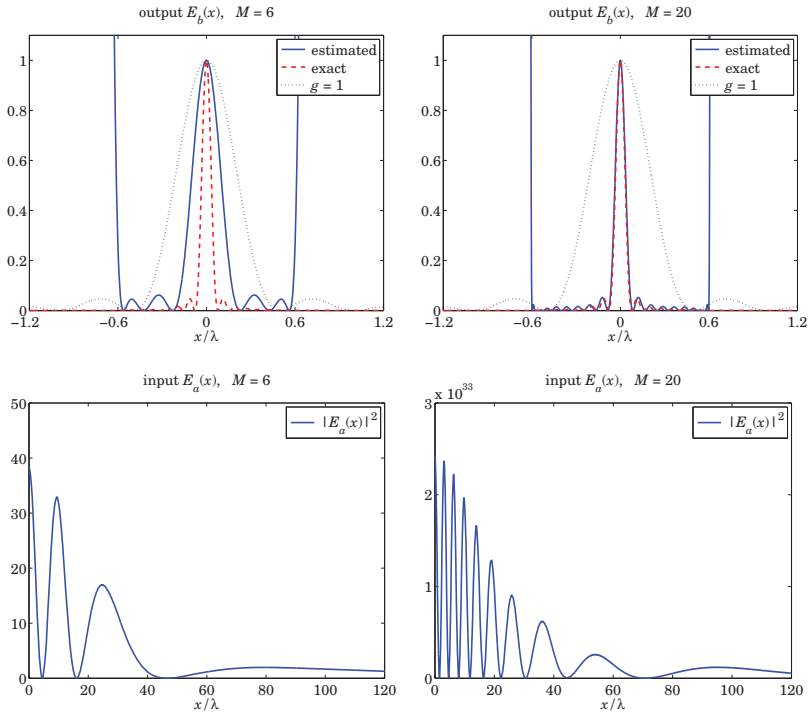


Fig. 20.22.6 Subwavelength focusing without evanescent waves, $M = 6$ and $M = 20$.

n	b_n	n	b_n
0	1.0260×10^{-1}	10	$-2.3540\times10^{+5}$
2	-1.0044×10^{-1}	12	$3.4742\times10^{+7}$
4	9.5813×10^{-1}	14	$-5.5154\times10^{+9}$
6	$-3.7193\times10^{+1}$	16	$1.8705\times10^{+12}$
8	$1.9365\times10^{+3}$	18	$-7.8910\times10^{+14}$
		20	$2.7855\times10^{+17}$

The MATLAB code used to generate these graphs was as follows, where the required integrations were done with the tanh-sinh quadrature function **quadts**,

```
lambda = 1;
k = 2*pi/lambda; a = 1.2*lambda/2; c = a*k
L = 20*lambda;                                  % propagation distance

g = 40/(2*pi);                                  % superresolution gain

Eb = @(x) sinc(g*k*x/pi);                       % desired Eb(x)
```



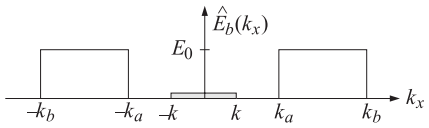
```
E1 = @(x) sinc(k*x/pi);                         % Rayleigh case, g=1

x = linspace(-2,2, 1201)*a;                     % plotting range for Eb(x)
xa = linspace(0,120,1201) * lambda;             % plotting range for Ea(x)

[wq,xq] = quadts(-a,a,8);          % quadts weights and evaluation points
[wk,kq] = quadts(-k,k,8);


for M = [6,20]
   [Pq,La] = pswf(a,k,M,xq);                    % psi_n(xq)

   m = (0:M)';
   Mu = (i.^m) .* sqrt(2*pi*La/c);

   b = (Pq * (wq.*Eb(xq)))./La;                 % b_n coefficients
   b(2:2:end)=0;                                % b_n = 0 at odd n

   Pb = pswf(a,k,M,x);                          % psi_n(x) at x

   Est = b' * Pb;                               % estimated Eb(x)

   E0 = abs(Est(find(x==0)));
   F = Est / E0;                                % normalize to unity at x=0

   F2 = abs(F).^2; Eb2 = abs(Eb(x)).^2; E2 = abs(E1(x)).^2;

   figure; plot(x,F2,'b-', x,Eb2,'r--', x,E2,'k--');
   xlim([-2,2]*a); ylim([0,1]);

   % ---- calculate Ea(x) ----

   Pk = pswf(a,k,M,kq*a/k);                     % psi_n(kx*a/k) at kx=kq

   Ebk = (b' * diag(1./conj(Mu)) * Pk) / k;     % Eb(kx), row vector

   kz = sqrt(k^2 - kq.^2);                      % kz at quadts points kq

   for p=1:length(xa)
      Ea(p) = Ebk * (exp(j*kz*L) .* exp(-j*kq*xa(p)) .* wk);
   end

   Ea2 = abs(Ea).^2;

   figure; plot(xa,Ea2,'b-')
end
```

The practical limitations of such designs have been discussed in [1556]. □

Next, we consider the other special case of the inverse diffraction geometry shown in Fig. 20.22.5, in which L is in the near field and $E_b(x)$ is chosen to be completely evanescent. The problem is to determine the input field $E_a(x)$ to achieve a highly focused output field $E_b(x)$. The general approach of designing such evanescent focusing fields, referred to as "near-field focusing plates," was first proposed by Merlin [1591] and further developed in [1592–1597].

As a concrete example from [1592,1593], consider an output field that has an ideal bandpass spectrum over the range $k_a \le |k_x| \le k_b$, with $k < k_a$, as shown below,

$$\hat{E}_b(k_x) = \begin{cases} E_0\,, & k_a \le |k_x| \le k_b \\ 0\,, & \text{otherwise} \end{cases}$$



The field $E_b(x)$ is determined by the inverse Fourier transform,

$$\begin{aligned} E_b(x) &= \int_{-\infty}^{\infty} \hat{E}_b(k_x)\, e^{-jk_x x}\, \frac{dk_x}{2\pi} = \int_{k_a \le |k_x| \le k_b} E_0\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= E_0\, \frac{\sin(k_b x) - \sin(k_a x)}{\pi x} = \frac{E_0 k_b}{\pi} \left[\frac{\sin(k_b x)}{k_b x} - \frac{k_a}{k_b}\, \frac{\sin(k_a x)}{k_a x} \right] \end{aligned} \tag{20.22.43}$$

We note that if $k_b \gg k_a > k$, then $E_b(x)$ is effectively given by a narrow sinc function, $\sin(k_b x)/(k_b x)$, with resolution distance at first-null, $d = \pi/k_b \ll \pi/k$, which is much smaller than the diffraction limit $d = \pi/k = \lambda/2$.

The corresponding input field $E_a(x)$ is reconstructed from (20.22.40), in which k_z is replaced by its evanescent version, $k_z = -j\sqrt{k_x^2 - k^2}$, because the integration range lies in the evanescent region, $|k_x| > k$,

$$\begin{aligned} E_a(x) &= \int_{-\infty}^{\infty} \hat{E}_b(k_x)\, e^{jk_z L} e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= \int_{k_a \le |k_x| \le k_b} E_0\, e^{jk_z L} e^{-jk_x x}\, \frac{dk_x}{2\pi} = \int_{k_a \le |k_x| \le k_b} E_0\, e^{L\sqrt{k_x^2 - k^2}}\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \end{aligned} \tag{20.22.44}$$

The last integral can be computed numerically, as shown for example in the MATLAB code below. However, if we further assume that $k_a \gg k$, then, since $|k_x| > k_a \gg k$, the evanescent square root can be further approximated by $k_z = -j\sqrt{k_x^2 - k^2} \approx -j|k_x|$, and the resulting approximate integral can be done in closed form,

$$\begin{aligned} E_a(x) &= \int_{k_a \le |k_x| \le k_b} E_0\, e^{L\sqrt{k_x^2 - k^2}}\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \approx \int_{k_a \le |k_x| \le k_b} E_0\, e^{L|k_x|}\, e^{-jk_x x}\, \frac{dk_x}{2\pi} \\ &= \frac{E_0}{\pi}\, \mathrm{Im}\left[\frac{e^{jk_b(x - jL)} - e^{jk_a(x - jL)}}{x - jL} \right] \\ &= \frac{E_0}{\pi}\, \frac{e^{k_b L}\left[x \sin(k_b x) + L\cos(k_b x)\right] - e^{k_a L}\left[x \sin(k_a x) + L\cos(k_a x)\right]}{x^2 + L^2} \end{aligned} \tag{20.22.45}$$

We note that $E_a(x)$ is much bigger than $E_b(x)$, approximately by the exponential factor $e^{k_b L}$, which dominates the other factor $e^{k_a L}$. Fig. 20.22.7 plots $E_b(x)$ computed from (20.22.43), and $E_a(x)$ from (20.22.45), for the following parameter values,

$$L = \frac{\lambda}{16}\,, \quad k_a = 2k\,, \quad k_b = 10k\,, \quad E_0 = \frac{\pi}{k_b}$$

We observe how the output $E_b(x)$ is narrower but weaker than $E_a(x)$. The exponential factors have the values, $e^{k_a L} = 50.7540$ and $e^{k_a L} = 2.1933$. The MATLAB code used to generate these graphs is given below.

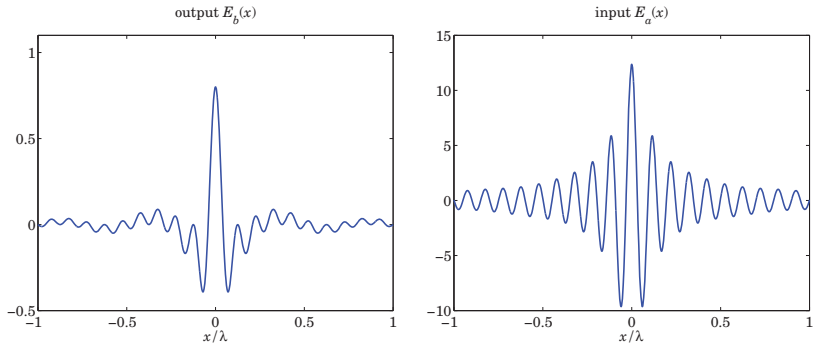


Fig. 20.22.7 Near-field output and input waveforms.

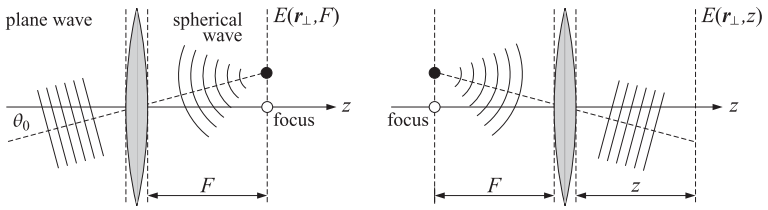
```
lambda = 1; k = 2*pi/lambda;
ka = 2*k; kb = 10*k; L = lambda/16;

Eb = @(x) sinc(kb*x/pi) - ka/kb * sinc(ka*x/pi);
Ea = @(x) 1/kb * imag((exp(j*kb*(x-j*L)) - exp(j*ka*(x-j*L)))./(x-j*L));

x = linspace(-1,1,2001) * lambda;

% [wq,kq] = quadts(ka,kb,8);              % quadts weights and evaluation points
% kz = -j*sqrt(kq.^2 - k^2);              % evanescent kz
%
% for i=1:length(x),
%    Ee(i) = E0 * wq' * (2*cos(kq*x(i)).*exp(j*kz*L))/2/pi;      % exact Ea(x)
% end

figure; plot(x,Eb(x),'b-')
% plot(x,Eb(x),'b-', x,sinc(kb*x/pi),'r--');    % plot exact & approximate Eb(x)

figure; plot(x, Ea(x),'b-')
% plot(x, Ea(x),'b-', x,Ee,'r--') ;             % plot exact & approximate Ea(x)
```

The exact inverse (20.22.44) is not plotted since it essentially coincides with that of the approximation (20.22.45), however, the MATLAB code needed to compute it was included in commented form.

20.23 Problems

20.1 A uniform plane wave, $E(x,z) = E_0 e^{-jk(x\sin\theta_0 + z\cos\theta_0)}$, is incident obliquely on a lens at an

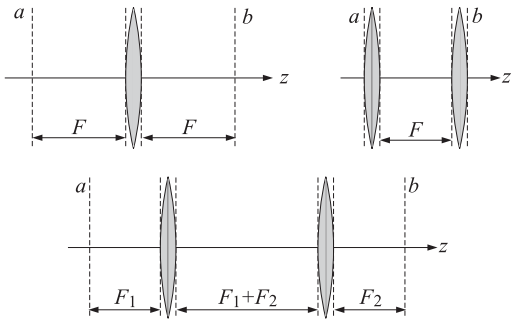
angle θ_0 with the z axis, as shown in the figure below.



Using similar methods as for Fig. 20.5.2, show that after passing through the lens, the wave will converge onto the shifted focal point with coordinates $z = F$ and $x = F\sin\theta_0$ (the above diagram implies that $x = F\tan\theta_0$, which is approximately equal to $F\sin\theta_0$ for small θ_0.)

Conversely, consider a point source of a spherical wave starting at the point $z = -F$ and $x = F\sin\theta_0$. Show that upon passage through the lens, the spherical wave will be converted into the obliquely moving plane wave $E(x,z) = E_1 e^{-jk(x\sin\theta_0 + z\cos\theta_0)}$. What is E_1?

20.2 Consider the three lens configurations shown below. They are special cases of Figs. 20.4.1 and 20.4.2, with appropriate choices for the input and output aperture planes a and b.



Working with Eqs. (20.4.1) and (20.4.3), show that the transfer functions $h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)$ are given as follows for the three cases:

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \frac{jk}{2\pi F}\, e^{-2jkF} e^{jk(\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp)/F}, \quad h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \frac{jk}{2\pi F}\, e^{-jkF} e^{jk(\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp)/F}$$

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = -\frac{F_1}{F_2}\,\delta\left(\boldsymbol{r}'_\perp + \frac{F_1}{F_2}\,\boldsymbol{r}_\perp\right)$$

Show that the first two cases perform a Fourier transformation as in Eq. (20.10.2), but without the quadratic phase factors. Show that the third case, performs a scaling of the input with a magnification factor $M = -F_2/F_1$.

20.3 Show that the encircled energy ratio for a uniform aperture, $A(x) = u(a - |x|)$, is given by,

$$\mathcal{E}(k_0) = \frac{\displaystyle\int_{-k_0}^{k_0} |\hat{A}(k_x)|^2\, dk_x}{\displaystyle\int_{-\infty}^{\infty} |\hat{A}(k_x)|^2\, dk_x} = \frac{2}{\pi}\left[\mathrm{Si}(2ak_0) - \frac{\sin^2(ak_0)}{ak_0}\right] \tag{20.23.1}$$

where $\mathrm{Si}()$ is the standard sine integral defined in Appendix G.

20.4 Consider an aperture-limited apodization function expanded as a sum of PSWF functions as in Eq. (20.12.23). Choose $k_0 = k = 2\pi/\lambda$, and show that the limiting directivity of Eq. (20.12.18) can be expressed in terms of the expansion coefficients in the following form,

$$\frac{D_\infty}{2ak} = \frac{\pi}{a}\,\frac{\left|\displaystyle\sum_{n=0}^{\infty} \hat{A}_n\,\psi_n(k_0, a, 0)\right|^2}{\displaystyle\sum_{n=0}^{\infty} |\hat{A}_n|^2}$$

Using this expression and the Cauchy-Schwarz inequality, show that the maximum value is $D_\infty = 2ak$, and is attained with the coefficients $\hat{A}_n = 2\pi\psi_n(k_0, a, 0)$. Moreover, show that with these coefficients, the expansions of Eq. (20.12.23) lead to the uniform distribution,

$$A(x) = u(a - |x|), \quad \hat{A}(k_x) = 2a\cdot\frac{\sin(k_x a)}{k_x a}$$

Hint: Eqs. (J.28)–(J.31) of Appendix J.

20.5 Consider an aperture-limited apodization function expanded as a sum of spherical Bessel functions as in Eq. (20.12.24), or as a sum of sinc-functions as in (20.12.28). Show that the limiting directivity of Eq. (20.12.18) can be expressed in terms of the corresponding expansion coefficients in the following forms,

$$\frac{D_\infty}{2ak} = \frac{|\hat{A}_0|^2}{\displaystyle\sum_{n=0}^{\infty} |\hat{A}_n|^2/(2n+1)} = \frac{|F(0)|^2}{\displaystyle\sum_{n=-\infty}^{\infty} |F(n)|^2}$$

As in the previous problem, show that the maximum value of the directivity is unity, determine the coefficients that realize this maximum, and show that the resulting aperture distribution is uniform. Moreover, show the following consequences of the Parseval identity for each case, written with respect to the variable $u = ak_x/\pi$,

$$\text{(Bessel)} \qquad \int_{-\infty}^{\infty} |F(u)|^2\, du = 2a\int_{-a}^{a} |A(x)|^2\, dx = \sum_{n=0}^{\infty} \frac{|\hat{A}_n|^2}{2n+1}$$

$$\text{(sinc)} \qquad \int_{-\infty}^{\infty} |F(u)|^2\, du = 2a\int_{-a}^{a} |A(x)|^2\, dx = \sum_{n=0}^{\infty} |F(n)|^2$$

20.6 For the optimum prolate apodization function defined in Eq. (20.13.2), show that the encircled energy, Strehl ratio, transmission coefficient, and normalized directivity, are,

$$\mathcal{E}(k_0) = \lambda_0, \quad S = \frac{\mu_0^2}{4}, \quad \tau = \frac{a\mu_0^2}{4\pi|\psi_0(0)|^2}, \quad \frac{D_\infty}{2ak} = \frac{\pi}{a}\,|\psi_0(0)|^2$$

where $\psi_0(0) = \psi_0(k_0, a, 0)$, $\mu_0 = \sqrt{2\pi\lambda_0/c}$, $c = ak_0$, and λ_0 is the 0-th prolate eigenvalue.

20.7 Prove Eq. (20.15.17) from (20.15.16).
Hint: $\sin(\pi(u-n)) = (-1)^n \sin(\pi u)$, for integer n.

20.8 For the Hansen window of Eq. (20.17.2), show that the encircled energy ratio and limiting directivity are given by the following closed-form expressions,

$$\mathcal{E}(u_0) = \frac{\int_0^{u_0} |F(u)|^2\, u\, du}{(2a)^2 \int_0^a |A(r_\perp)|^2 r_\perp dr_\perp} = 1 - \frac{J_0^2\left(\pi\sqrt{u_0^2 - H^2}\right) + J_1^2\left(\pi\sqrt{u_0^2 - H^2}\right)}{I_0^2(\pi H) - I_1^2(\pi H)}$$

$$\frac{D_\infty}{(ka)^2} = \frac{\left|\int_0^a A(r_\perp)\, 2\pi r_\perp dr_\perp\right|^2}{\pi a^2 \int_0^a |A(r_\perp)|^2\, 2\pi r_\perp dr_\perp} = \frac{\left[\dfrac{2I_1(\pi H)}{\pi H}\right]^2}{I_0^2(\pi H) - I_1^2(\pi H)}$$

20.9 To prove Eq. (20.20.8) without invoking the Babinet principle, first derive the following integrals, where $\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp$ are two-dimensional transverse vectors with magnitudes $r_\perp = |\boldsymbol{r}_\perp|$ and $r'_\perp = |\boldsymbol{r}'_\perp|$,

$$\frac{jk}{2\pi z} \int_{-\infty}^{\infty} e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp|^2/2z}\, d^2\boldsymbol{r}'_\perp = \frac{jk}{2\pi z}\, e^{-jkr_\perp^2/2z} \int_0^\infty e^{-jkr_\perp'^2/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp = 1$$

then, using this result, show the relationship for any $A(r_\perp)$ and $A_c(r_\perp) = 1 - A(r_\perp)$,

$$\frac{jk}{2\pi z}\, e^{-jkr_\perp^2/2z} \int_0^\infty A(r'_\perp)\, e^{-jkr_\perp'^2/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp =$$

$$= 1 - \frac{jk}{2\pi z}\, e^{-jkr_\perp^2/2z} \int_0^\infty A_c(r'_\perp)\, e^{-jkr_\perp'^2/2z}\, J_0\left(\frac{kr_\perp r'_\perp}{z}\right) 2\pi\, r'_\perp dr'_\perp$$

20.10 *Computer Experiment.* Carry out the same experiment using the same numerical values as in Example 20.22.4, but instead of Eq. (20.22.42), take the ideal focal spot within the field-of-view $[-a, a]$ to be a delta function,

$$E_b(x) = \delta(x)\,, \quad \text{for } |x| \le a = 0.6\lambda \tag{20.23.2}$$

Make similar plots as in Fig. 20.22.6. *Hint:* In this case the expansion coefficients are simply, $b_n = \psi_n(a, k, 0)/\lambda_n$, $n = 0, 1, \ldots, M$.

21

Aperture Antennas

21.1 Open-Ended Waveguides

The aperture fields over an open-ended waveguide are not uniform over the aperture. The standard assumption is that they are equal to the fields that would exist if the guide were to be continued [1].

Fig. 21.1.1 shows a waveguide aperture of dimensions $a > b$. Putting the origin in the middle of the aperture, we assume that the tangential aperture fields $\boldsymbol{E}_a$, $\boldsymbol{H}_a$ are equal to those of the TE_{10} mode. We have from Eq. (9.4.3):

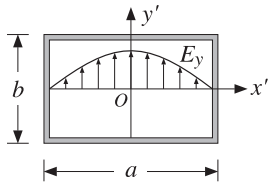


Fig. 21.1.1 Electric field over a waveguide aperture.

$$E_y(x') = E_0 \cos\left(\frac{\pi x'}{a}\right)\,, \quad H_x(x') = -\frac{1}{\eta_{TE}} E_0 \cos\left(\frac{\pi x'}{a}\right) \tag{21.1.1}$$

where $\eta_{TE} = \eta/K$ with $K = \sqrt{1 - \omega_c^2/\omega^2} = \sqrt{1 - (\lambda/2a)^2}$. Note that the boundary conditions are satisfied at the left and right walls, $x' = \pm a/2$.

For larger apertures, such as $a > 2\lambda$, we may set $K \simeq 1$. For smaller apertures, such as $0.5\lambda \le a \le 2\lambda$, we will work with the generalized Huygens source condition (18.5.7). The radiated fields are given by Eq. (18.5.5), with $f_x = 0$:

$$\begin{aligned} E_\theta &= jk\, \frac{e^{-jkr}}{2\pi r}\, c_\theta\, f_y(\theta, \phi) \sin\phi \\ E_\phi &= jk\, \frac{e^{-jkr}}{2\pi r}\, c_\phi\, f_y(\theta, \phi) \cos\phi \end{aligned} \tag{21.1.2}$$