

Chapter 3

Diffraction

While the particle description of Chapter 1 works amazingly well to understand how radiation is generated and transported in (and out of) astrophysical objects, it falls short when we investigate how we detect this radiation with our telescopes. In particular the *diffraction limit* of a telescope is the limit in spatial resolution a telescope of a given size can achieve because of the wave nature of light. Diffraction gives rise to a “Point Spread Function”, meaning that a point source on the sky will appear as a blob of a certain size on your image plane. The size of this blob limits your spatial resolution. The only way to improve this is to go to a larger telescope.

Since the theory of diffraction is such a fundamental part of the theory of telescopes, and since it gives a good demonstration of the wave-like nature of light, this chapter is devoted to a relatively detailed ab-initio study of diffraction and the derivation of the point spread function.

Literature:

Born & Wolf, “Principles of Optics”, 1959, 2002, Press Syndicate of the Univeristy of Cambridge, ISBN 0521642221. A classic.

Steward, “Fourier Optics: An Introduction”, 2nd Edition, 2004, Dover Publications, ISBN 0486435040

Goodman, “Introduction to Fourier Optics”, 3rd Edition, 2005, Roberts & Company Publishers, ISBN 0974707724

3.1 Fresnel diffraction

Let us construct a simple “Camera Obscura” type of telescope, as shown in Fig. 3.1. We have a point-source at location “a”, and we point our telescope exactly at that source. Our telescope consists simply of a screen with a circular hole called the *aperture* or alternatively the *pupil*. Behind that we place a screen. If we place the source at “a” at a very large distance and make the object small enough that it can effectively be regarded as a point source, then the light waves that enters our aperture can be considered as a plane wave. The finite size of the aperture of our telescope means that the wave front behind the aperture suffers from distortions due to the wave bending around the edges of the aperture. This is called *diffraction* and is a fundamental consequence of the wave nature of light (it happens in fact for any wave phenomenon). According to Huygens’ principle (named after Christiaan Huygens, 1629-1695, a Dutch allrounder in natural sciences), the wave front behind the aperture can be regarded as the sum of infinitely many spherical waves emitted

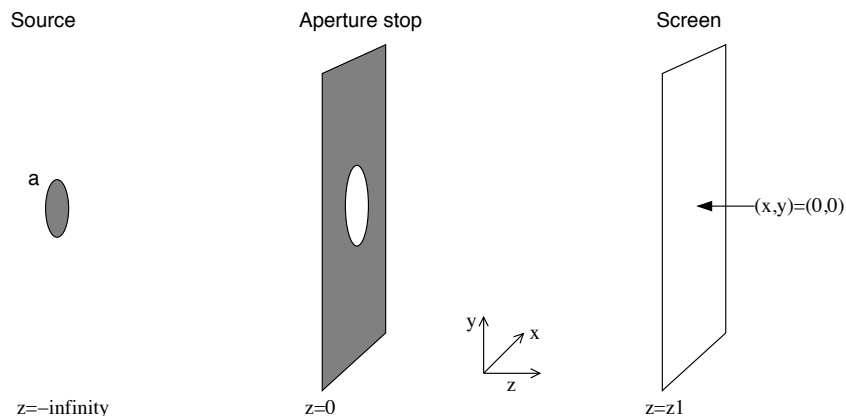


Figure 3.1: The definition of the coordinate system for our derivation of Fresnel diffraction.

in the aperture plane with the appropriate phases. These waves subsequently interfere with each other to form the actual wave pattern. This is pictographically shown in Fig. 3.2. An awesome Java Applet that simulates this, and many more configurations including your own designs, can be found on the web site of Paul Falstad¹: Highly recommended! A mathematical description of Huygens' principle is given by the *Kirchhoff-Fresnel integral*, named after Gustav Robert Kirchhoff (1824-1887), a German physicist, famed for his work with Robert Bunsen on spectroscopy at the University of Heidelberg, and Augustin-Jean Fresnel (1788-1827), a French physicist famed for his practical and mathematical studies of the wave nature of light. We will derive this integral below.

In our derivation we use the setup and coordinate definition as shown in Fig. 3.1. The $z = 0$ plane is the aperture, the screen is at $z = z_1$ while the observed object is at $z = -\infty$.

Let us now ask ourselves what is the electric field at the screen ($z = z_1$) as a function of x and y , given a known electric field configuration in the pupil ($z = 0$). A rigorous derivation of this is not trivial. The answer is, without proof,

$$E(x, y, z_1) = \frac{z_1}{i\lambda} \int_{\text{pupil}} E(x', y', 0) \frac{1}{r^2} \exp \left[-2\pi i \nu \frac{r}{c} \right] dx' dy' \quad (3.1)$$

where the integral is taken over the pupil only. The wavelength is $\lambda = c/\nu$. The distance between the point (x', y') in the pupil plane and the point (x, y) on the screen is

$$r = \sqrt{(x - x')^2 + (y - y')^2 + z_1^2} \quad (3.2)$$

Just as a check: if we put the screen very far away, then $r \simeq z_1$ and indeed $E \propto 1/r$, as we expect from electrodynamics.

Eq. (3.1) is in fact the mathematical formulation of Huygens' principle. The integral can be tricky because of the non-trivial form of Eq. (3.2). A numerical integration can be done relatively easily, but with some approximations to Eq. (3.2) we can also get an answer analytically, which gives us more insight². Let us define a quantity ρ

$$\rho = \sqrt{(x - x')^2 + (y - y')^2} \quad (3.3)$$

¹<http://www.falstad.com/ripple/>

²We follow http://en.wikipedia.org/wiki/Fresnel_diffraction

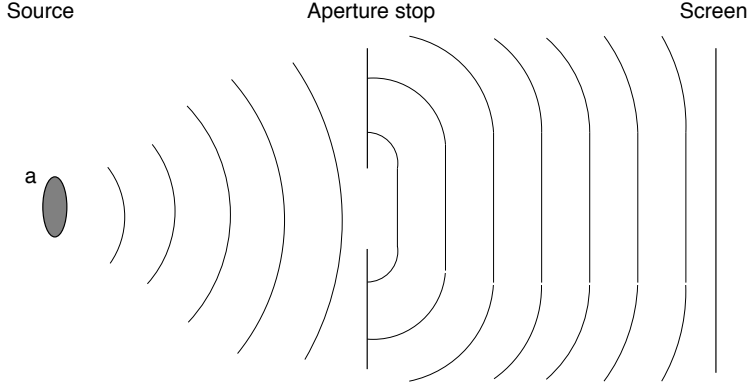


Figure 3.2: Pictogram of a wave propagating through an aperture.

The r then becomes

$$r = \sqrt{\rho^2 + z_1^2} = z_1 \sqrt{1 + \frac{\rho^2}{z_1^2}} \quad (3.4)$$

A Taylor expansion gives

$$r = z_1 + \frac{\rho^2}{2z_1} - \frac{\rho^4}{8z_1^3} + \dots \quad (3.5)$$

If we substitute this into the exponential term in Eq. (3.1) this exponent becomes:

$$\exp \left[-2\pi i \nu \frac{r}{c} \right] = \exp \left[-\frac{2\pi i \nu}{c} \left(z_1 + \frac{\rho^2}{2z_1} - \frac{\rho^4}{8z_1^3} + \dots \right) \right] \quad (3.6)$$

Under certain conditions we can ignore the third and higher terms, which would simplify things considerably. The condition is that it contributes much less than 2π to the phase, i.e.:

$$\frac{2\pi \nu}{c} \frac{\rho^4}{8z_1^3} \ll 2\pi \quad (3.7)$$

or using $\lambda = c/\nu$ we get

$$\frac{\rho^4}{\lambda^4} \ll 8 \frac{z_1^3}{\lambda^3} \quad (3.8)$$

Under this condition we only have the first two Taylor terms left. This is the *Fresnel approximation*. It usually holds for cases in which $\rho \ll z_1$, i.e. when the pupil is much smaller than the distance to the screen.

Let us now return to Eq. (3.1). We use the Fresnel approximation for the r in the exponent. For the r in the denominator it is actually usually safe to approximate it even further by $r \simeq z_1$. To make the following formula simple, let us use the quantity $k = 2\pi\nu/c$ from here on. With the above approximations we can rewrite Eq. (3.1) as

$$E(x, y, z_1) = \frac{e^{-ikz_1}}{i\lambda z_1} \int_{\text{pupil}} E(x', y', 0) \exp \left[-\frac{ik}{2z_1} \{ (x - x')^2 + (y - y')^2 \} \right] dx' dy' \quad (3.9)$$

This is the *Fresnel diffraction integral*. While this expression is clearly more direct and simplified than the original Eq. (3.1), an analytic solution is still only found in special cases.

Let us work out such an example. We follow the lecture by Dr. Daniel Mittleman³. We assume the pupil to be a slit of half-width b . This reduces out 3-D problem to a 2-D problem: we can drop the y -dependence. We thus start from

$$E(x, z_1) = \frac{e^{-ikz_1}}{i\lambda z_1} \int_{-b}^b E(x', 0) \exp \left[-\frac{ik}{2z_1} (x - x')^2 \right] dx' \quad (3.10)$$

Let us, for simplicity, take $E(x', 0)$ constant over the slit, so that it can be put before the integral,

$$E(x, z_1) = E_0 \frac{e^{-ikz_1}}{i\lambda z_1} \int_{-b}^b \exp \left[-\frac{ik}{2z_1} (x - x')^2 \right] dx' \quad (3.11)$$

Change variables: $\xi = x/b$ and $\xi' = x'/b$ to obtain

$$E(\xi, z_1) = E_0 \frac{be^{-ikz_1}}{i\lambda z_1} \int_{-1}^1 \exp \left[-\frac{ikb^2}{2z_1} (\xi - \xi')^2 \right] d\xi' \quad (3.12)$$

Define the *Fresnel number* F

$$F = \frac{kb^2}{2\pi z_1} = \frac{b^2}{z_1 \lambda} \quad (3.13)$$

and we can write

$$E(\xi, z_1) = E_0 \frac{be^{-ikz_1}}{i\lambda z_1} \int_{-1}^1 \exp [-i\pi F (\xi - \xi')^2] d\xi' \quad (3.14)$$

This integral is not solvable analytically, but it can be expressed in terms of known and tabulated transcendental functions: the *Fresnel integrals*:

$$S(t) = \int_0^t \sin(x^2) dx \quad (3.15)$$

$$C(t) = \int_0^t \cos(x^2) dx \quad (3.16)$$

These integrals can be found in tabulated form in books such as the book by *Abramowitz & Stegun*, but you can also tabulate them yourself by computing them numerically. The functions are shown in Fig. 3.3.

Let us focus now just on the exponential in Eq. (3.14):

$$\int_{-1}^1 \exp [-i\pi F (\xi - \xi')^2] d\xi' = \int_{-1-\xi}^{1-\xi} \exp [-i\pi F (\xi' - \xi)^2] d(\xi' - \xi) \quad (3.17)$$

$$= \frac{1}{\sqrt{\pi F}} \int_{-(1+\xi)\sqrt{\pi F}}^{(1-\xi)\sqrt{\pi F}} \exp [-i\eta^2] d\eta \quad (3.18)$$

$$= \frac{1}{\sqrt{\pi F}} \int_{-(1+\xi)\sqrt{\pi F}}^{(1-\xi)\sqrt{\pi F}} [\cos(\eta^2) - i \sin(\eta^2)] d\eta \quad (3.19)$$

We leave the last step, to express this in terms of $S(t)$ and $C(t)$, to you as an exercise. The numerical results of the *absolute value squared* of the above integral for different Fresnel number F are shown in Fig. 3.4.

³www-ece.rice.edu/~daniel/262/pdf/lecture27.pdf

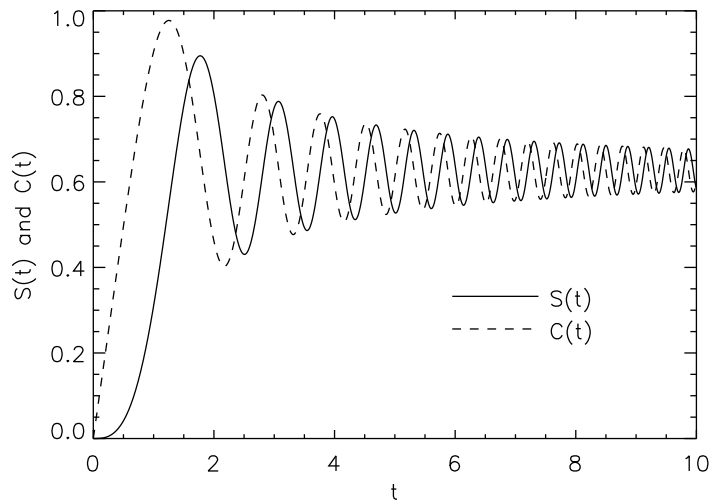


Figure 3.3: The Fresnel integrals $S(t)$ and $C(t)$.

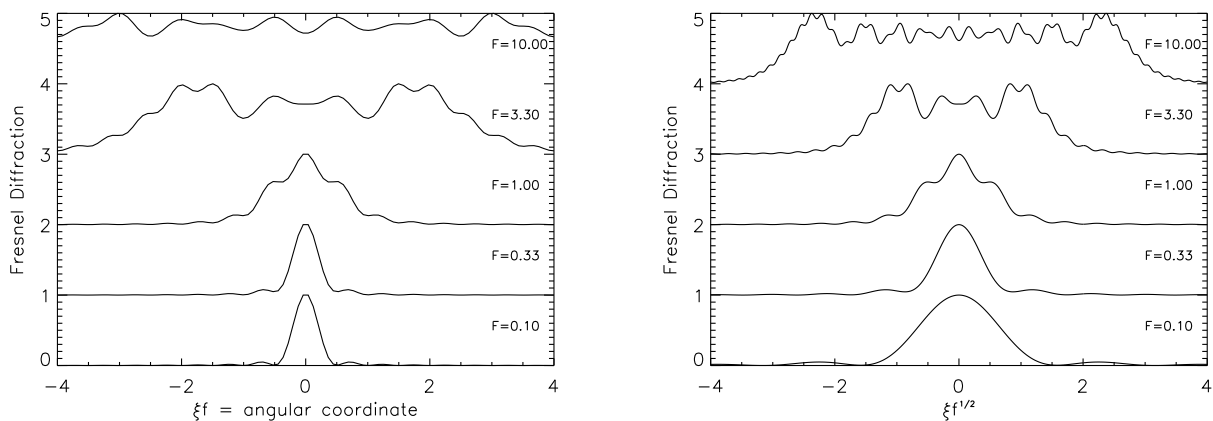


Figure 3.4: The Fresnel diffraction patterns for different values of the Fresnel number F . The vertical scaling is always normalized to 1, and the curves are simply offset by 1 between each other. The left and the right panel show the same results, just with different scaling of the horizontal coordinate. Left: scaling such that the horizontal coordinate is proportional to the angle x/z_1 . There you see that, going from top ($F = 10$) to bottom ($F = 0.1$) the wave is converging to a sinc function. Right: to get the entire width of the pattern on the plot for all our values of F , the horizontal coordinate has been changed (somewhat arbitrarily) to $\xi\sqrt{F}$.

3.2 Fraunhofer diffraction

Fraunhofer diffraction, named after Joseph von Fraunhofer (1787-1826), a German optician, is actually Fresnel diffraction in the limit of small Fresnel number F . It is the next level of approximation of diffraction. Let us start with Eq. 3.14, but writing out the $(\xi - \xi')^2$:

$$E(\xi, z_1) = E_0 \frac{be^{-ikz_1}}{i\lambda z_1} \int_{-1}^1 \exp[-i\pi F(\xi^2 + (\xi')^2 - 2\xi\xi')] d\xi' \quad (3.20)$$

$$= E_0 \frac{be^{-ikz_1 - i\pi F\xi^2}}{i\lambda z_1} \int_{-1}^1 \exp[-i\pi F((\xi')^2 - 2\xi\xi')] d\xi' \quad (3.21)$$

For $F \ll 1$ the $(\xi')^2$ term in the exponent is always small compared to unity (keep in mind that $-1 \leq \xi' \leq 1$ by definition!), while the $-2\xi\xi'$ can still be of the order of unity or larger, if $\xi \gg 1$. So for $F \ll 1$ we can drop the $(\xi')^2$ term. This is the *Fraunhofer approximation*. It obviates the need for the Fresnel integrals. We obtain

$$E(\xi, z_1) = E_0 \frac{be^{-ikz_1 - i\pi F\xi^2}}{i\lambda z_1} \int_{-1}^1 \exp[2\pi i F\xi\xi'] d\xi' \quad (3.22)$$

In other words: the electric field at ξ is proportional to the Fourier transform of a block function between -1 and 1 . This block function describes the transmission of the aperture mask. One can therefore say that, in the far-field limit ($F \ll 1$), what you see on the screen is the Fourier transform of the aperture! The conjugate variable to ξ' is then $F\xi = bx/\lambda z_1$.

For a block function, the Fourier transform is the *sinc function*:

$$\int_{-\infty}^{\infty} \text{box}(t)e^{-2\pi xt} dx = \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \quad (3.23)$$

where the function $\text{box}(x)$ is 1 for $-1/2 \leq x \leq 1/2$ and 0 elsewhere. In other words: for our slit experiment we see a sinc function on the screen.

We can now go back to full 3-D, i.e. re-introducing the y -coordinate. Fresnel diffraction integral Eq. (3.9) then becomes in the limit of $F \ll 1$:

$$E(x, y, z_1) = \frac{e^{-ik(z_1 + (x^2 + y^2)/2z_1)}}{i\lambda z_1} \int_{\text{pupil}} E(x', y', 0) \exp\left[\frac{ik}{2z_1}(xx' + yy')\right] dx' dy' \quad (3.24)$$

So we see that what we observe on the screen is the 2-D Fourier transform of the electric field pattern in the pupil. If we again set, for simplicity, $E(x', y', 0) = E_0$ like we did before, then the image on the screen is the Fourier transform of the pupil again. For a circular pupil of radius b we can use the Fourier transform of the circ-function (Appendix A):

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{circ}\left(\frac{\sqrt{(x')^2 + (y')^2}}{b}\right) e^{-2\pi i(ux' + vy')} dx' dy' = \frac{J_1(2\pi b\sqrt{u^2 + v^2})}{\sqrt{u^2 + v^2}} \quad (3.25)$$

where $u = -x/\lambda z_1$ and $v = -y/\lambda z_1$, the $\text{circ}(x)$ function is 1 for $x \leq 1$ and 0 for $x > 1$, and $J_1(x)$ is the Bessel function of the first kind of order one. Eq. (3.24) thus becomes

$$E(x, y, z_1) = -iE_0 e^{-ik(z_1 + (x^2 + y^2)/2z_1)} \lambda z_1 \frac{J_1\left(\frac{2\pi b}{\lambda z_1} \sqrt{x^2 + y^2}\right)}{\sqrt{x^2 + y^2}} \quad (3.26)$$

What we see on the screen is the flux E^*E :

$$F(x, y) = E_0^* E_0 \lambda^2 z_0^2 \left[\frac{J_1 \left(\frac{2\pi b}{\lambda z_1} \sqrt{x^2 + y^2} \right)}{\sqrt{x^2 + y^2}} \right]^2 \quad (3.27)$$

This is the *Airy pattern*, after Sir George Biddell Airy (1801-1892), Astronomer Royal at Greenwich Observatory, who studied the diffraction of light through circular apertures.

It is important to note that this pattern scales linearly with z_1 : if you put the screen twice as far, the shape of the pattern stays the same, but spread out over twice the size. In other words: the pattern is actually a pattern as a function of the *angles* (θ_x, θ_y) $\equiv (x/z_1, y/z_1)$.

3.3 The “point spread function” of a telescope

So far we just had an aperture and a screen. There was no lens or mirror involved. As a result, the radiation on the screen was not really an image of the sky unless you put the screen at a huge distance (small F number) and use a colossal screen. This is unpractical. Now let us introduce a lens in the pupil. An easy way to mimic a lens is to introduce a phase shift in the pupil plane of $\Delta\phi = h(x^2 + y^2)/\lambda$. We want to fine-tune h such that the wave behind the pupil acts as a circular wave converging on the screen at a distance z_1 . This can be achieved by choosing $h = \pi/z_1$. We get a phase shift:

$$\Delta\phi = \frac{\pi(x^2 + y^2)}{z_1 \lambda} = \frac{k(x^2 + y^2)}{2z_1} \quad (3.28)$$

(with again $k = 2\pi/\lambda$). We can thus write

$$E(x', y', 0) = E_0 \exp \left[i \frac{k((x')^2 + (y')^2)}{2z_1} \right] \quad (3.29)$$

Insert this into the Fresnel diffraction integral Eq. (3.9) and we obtain

$$\begin{aligned} E(x, y, z_1) &= E_0 \frac{e^{-ikz_1}}{i\lambda z_1} \int_{\text{pupil}} \exp \left[-\frac{ik}{2z_1} \{ (x - x')^2 + (y - y')^2 - (x')^2 - (y')^2 \} \right] dx' dy' \\ &= E_0 \frac{e^{-ik(z_1 + (x^2 + y^2)/2z_1)}}{i\lambda z_1} \int_{\text{pupil}} \exp \left[\frac{ik}{z_1} \{ xx' + yy' \} \right] dx' dy' \end{aligned} \quad (3.30)$$

We again arrive at the Fourier transform of the pupil! It is in fact the same as for Fraunhofer diffraction, Eq. (3.24) and likewise the flux seen on the screen is an Airy pattern: What we see on the screen is the flux E^*E :

$$F(x, y) = E_0^* E_0 \lambda^2 z_0^2 \left[\frac{J_1 \left(\frac{2\pi b}{\lambda z_1} \sqrt{x^2 + y^2} \right)}{\sqrt{x^2 + y^2}} \right]^2 \quad (3.31)$$

The difference is that the Fraunhofer result is only valid for very small Fresnel number $F \ll 1$, i.e. in the very far field limit. In practice this means that if you would like to make

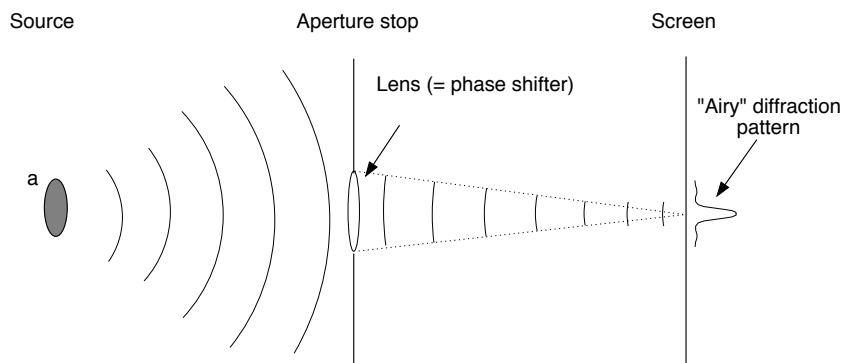


Figure 3.5: By including a lens in the pupil, which shifts the phases such that the waves converge at a distance $z_1 = f$, the focal distance, we find that the wave does not converge to a delta-function but instead converges to an Airy pattern. This is the “point spread function” of the telescope due to diffraction.

an image of an object on the sky without a lens, you would have to put the screen at a very large distance (and consequently have a very large screen!). This is utterly impractical. By introducing the lens we find that if we put the screen at the focal distance (for which in general $F \gtrsim 1$ or even $F \gg 1$) we still get the *Airy* diffraction pattern (see Fig. 3.5 for a pictographic representation). The radius of the first few *nulls* of this pattern on this screen can be derived from the first few nulls of the $J_1(x)$ Bessel function, which are at $x = 3.8317, 7.0156$ and 10.1735 . So for the first, second and third null we set

$$\frac{2\pi br}{\lambda z_1} = \begin{cases} 3.8317 \\ 7.0156 \\ 10.1735 \end{cases} \quad (3.32)$$

where $r = \sqrt{x^2 + y^2}$, which gives, if we replace the aperture radius b conventionally with the aperture diameter $D = 2b$:

$$\frac{Dr_1}{\lambda z_1} = 1.2196 \quad \frac{Dr_2}{\lambda z_1} = 2.2332 \quad \frac{Dr_3}{\lambda z_1} = 3.2384 \quad (3.33)$$

The locations of the peaks of the first and second bump (the *Airy rings*) are at $Dr/\lambda z_1 = 1.64$ and 2.67 , respectively.

If we put the screen more toward the aperture or away from the aperture, the radiation would be spread in a wider area. In fact, for the simple analog of a slit with a cylindrical lens (instead of a circular aperture with a parabolic lens), we already calculated the wave pattern slightly in front and behind the focal plane: it is what is shown in Fig. 3.4. The case when we are in the focal plane corresponds to $F \ll 1$ (bottom curve) where indeed the curve is very close to a sinc function (the slit analog to the Airy function), while the case when we are in front or behind the focal plane corresponds to curves with larger F .

The Airy pattern on the screen, when the screen is at the focal distance, is in fact the most compact one can focus the radiation, as the slit-analog in Fig. 3.4-left shows. In other words: a point source on the sky produces an Airy pattern on the screen. This is often called the *point spread function* (PSF) of diffraction of the telescope. The larger the aperture, the narrower this PSF is.

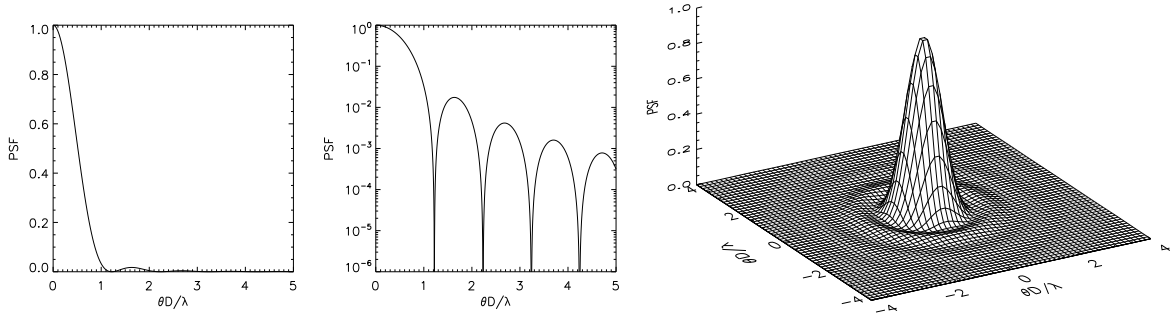


Figure 3.6: The Airy PSF depicted in three ways. Here D is the telescope diameter, λ the wavelength and θ the angle away from the center in radian. In all cases the function is normalized such that at the peak it is 1.

The finite size of the PSF can be translated to a corresponding size of the PSF in angular units on the sky: it determines the maximum spatial resolution of your telescope. This is easily determined, because if we again use (n_x, n_y) as angular position on the sky, then on the screen these correspond to $(x, y) = -(n_x, n_y)z_1$. As a final result one can say that the PSF in terms of angular coordinates on the sky is:

$$I(\theta) \propto \left(\frac{J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda} \right)^2 \quad (3.34)$$

where $D = 2b$ is the diameter of the pupil and where $\theta \equiv \sqrt{n_x^2 + n_y^2}$ is in radian. The square is because the intensity is proportional to the square of the electric field. The first null is at an angle (see above)

$$\theta_{\text{null}} = 0.6098 \frac{\lambda}{b} = 1.22 \frac{\lambda}{D} \quad (3.35)$$

where $D = 2b$ is the telescope aperture diameter. This roughly determines the angular resolution of the telescope.

Any point-like star will be projected on the screen as an Airy pattern, i.e. in addition to the peak, the image also features a *diffraction ring* around the peak, and in principle an infinite succession of diffraction rings of ever increasing radius and decreasing amplitude. The Airy PSF is shown in Fig. 3.6 in three visualizations (lin, log, 2D).

Exercise: Right now the Airy pattern is centered on $x = 0, y = 0$ on the screen ($z = z_1$). How must one modify the phases of $E(x, y, 0)$ over the pupil to get the Airy pattern at another location on the screen? How does this relate to the direction of the point source? And finally, how can one explain this in terms of one of the standard properties of Fourier transforms (see Appendix A)?

3.4 Image smearing: convolution with the PSF

The PSF degrades the spatial resolution of an image. We can express the measured image on the image plane $I_{\text{obs}}(x, y)$ in terms of the image $I_{\text{inf}}(x, y)$ which one would obtain if one

would have infinite spatial resolution:

$$I_{\text{obs}}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I_{\text{inf}}(x', y') \text{PSF}(x - x', y - y') dx' dy' \quad (3.36)$$

where we assume that the function $\text{PSF}(x, y)$ is normalized to unity:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \text{PSF}(x, y) dx dy = 1 \quad (3.37)$$

Eq. 3.36 is a *convolution* of the original image $I_{\text{inf}}(x', y')$ with the PSF. The PSF is the *convolution kernel* of Eq. 3.36.

Question: If we want to simplify our convolution kernel, can we use a Gaussian? Answer: partly yes, partly no. A good approximation of the main peak of the Airy PSF is:

$$\left(\frac{J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda} \right)^2 \simeq e^{-2.5(\theta D/\lambda)^2} \quad (3.38)$$

However, this approximates only the main peak. If you are interested in detecting a faint companion around a bright star (for instance when searching for *extrasolar planets*) the Gaussian approximation wildly underestimates the Airy rings. From Fig. 3.6 one can see that the first Airy ring is 0.0175 times as bright as the peak, but the second Airy ring is 0.00416 times as bright as the peak, i.e. a mere 0.237 times as weak as the first ring. In other words: The first ring is much weaker than the peak, but each successive ring is only a bit weaker still. These *wings of the PSF* extend very far away from a star and may hinder extremely sensitive searches for ultra-faint stellar companions. As seen in from Fig. 3.6 even at $\theta \simeq 5\lambda/D$ a companion with luminosity 10^{-3} of the stellar luminosity is possibly swamped by the PSF of the star. The typical contrast one has to overcome for direct imaging of planets is considerably higher.

In reality we do not have just the spherical aperture: we also have a secondary mirror hanging in the middle of the aperture, as well as a set of support beams to keep the secondary mirror in place. The PSF, being the absolute value squared of the Fourier transform of the aperture, will thus not be precisely an Airy pattern, but a distorted Airy pattern. One can very simply simulate what such a PSF would look like by applying a 2-D FFT () to a 2-D “image” of the aperture. For instance, make a 2048×2048 array of real numbers representing an “image” of the aperture: Take these numbers 0 outside of the aperture or there where the secondary mirror or the support structures are. For the rest it is 1. Make sure that the entire size of the aperture is much smaller than the size of the image. Then apply the FFT operation and you get the PSF shape. This is shown in Fig. 3.7. One sees that the support beams create *diffraction spikes*.

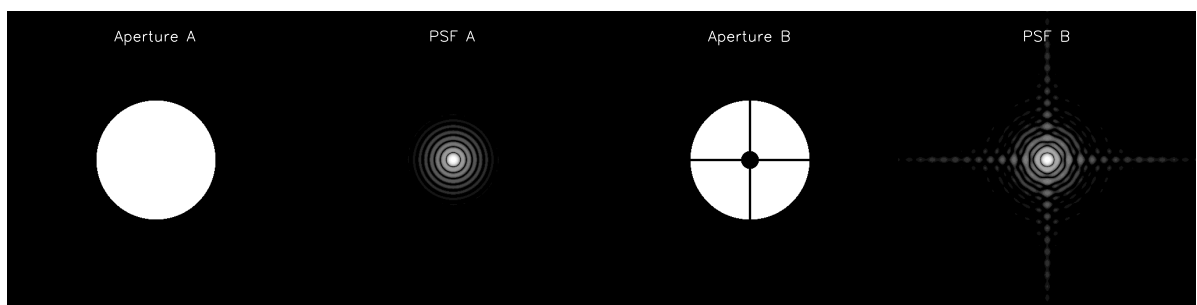


Figure 3.7: A 2-D model of a diffraction PSF. Left two panels: the case of a perfectly circular aperture (you get an Airy PSF). Right two panels: the case of an aperture with a secondary mirror and supporting structures. These models were calculated using a map of 2048×2048 pixels, where in all panels only the central $1/4$ of the map is shown. FFT is used to compute the PSFs from the aperture images.