III. Light is a Wave (Physical Optics)

III.F Diffraction of Light

1. The Huygens-Fresnel Principle

Simply put, *diffraction is the tendency of a wave emitted from a finite source or passing through a finite aperture to spread out as it propagates.* To understand diffraction, we should recognize

Diffraction = Interference!

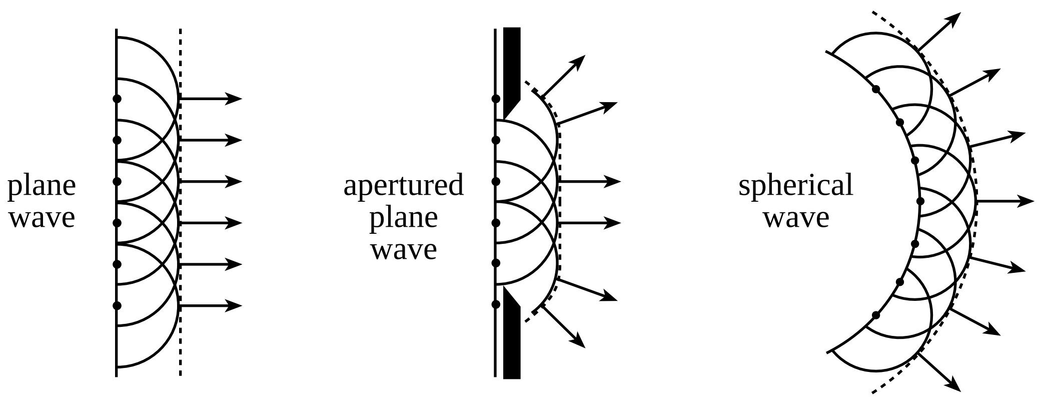
Unlike the interference we have seen thus far, which always occurs between 2 or more separate but discrete waves, diffraction results from the interference of an infinite number of waves emitted by a *continuous distribution of source points*.

Huygens suggested a way of understanding the propagation and diffraction of waves, and Fresnel later amended Huygens principle as follows.

Huygens → "Every point on a wavefront of light can be considered to be a secondary source of spherical wavelets."

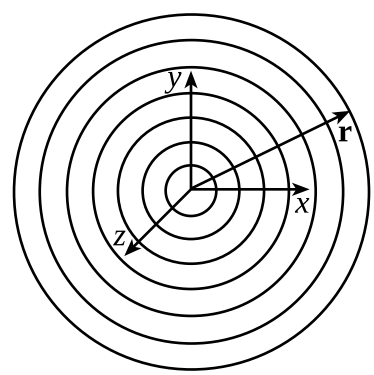
Fresnel → "The wave amplitude at any point beyond the wavefront is a superposition of the amplitudes of all of the wavelets, accounting for both their magnitude and phase."

These concepts are illustrated in the drawings below.



This way of understanding diffraction is based on a description of spherical waves, which we have not looked at yet. As you might guess, a spherical wave is just a wave that travels along the radius of a sphere, rather than along a fixed direction like x, y, or z.

Since everywhere in space the wavevector is parallel to the position vector, or $\mathbf{k} \parallel \mathbf{r}$, then $\mathbf{k}\cdot\mathbf{r} = k\,r$, where k is the wavenumber and r is just the (scalar) distance along any radial direction.



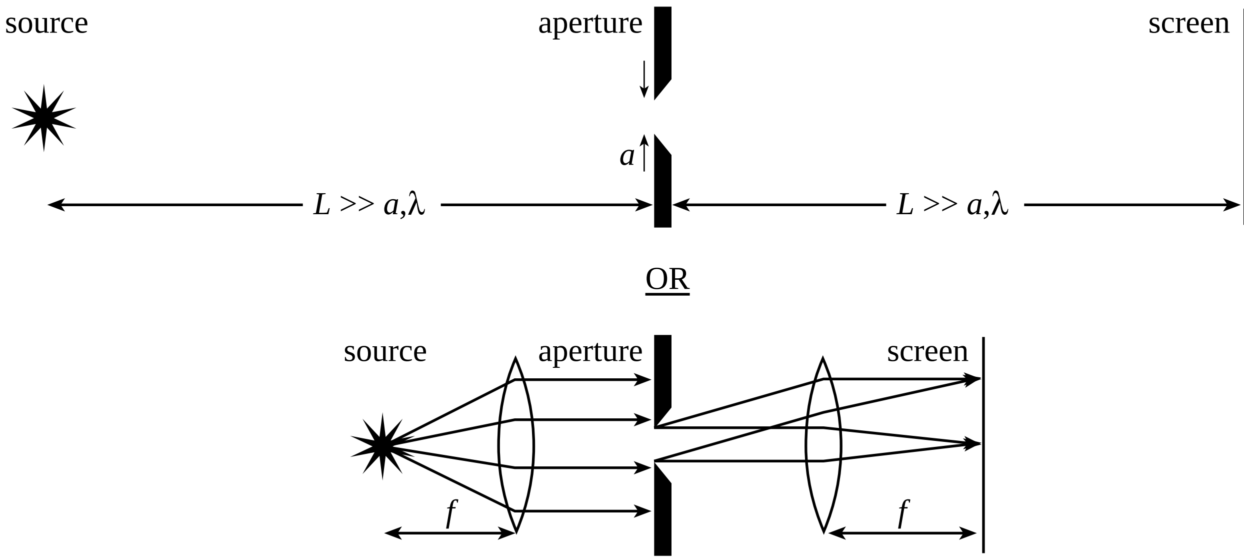
Also, since the wavefront area grows as $4\pi r^2$, but the *power* emitted by the source at the center is *constant*, then the Intensity (which is power/area) must decrease in proportion to $1/r^2$. Since $I = \psi^2$, then the amplitude ψ must be proportional to $1/r$. Thus a spherical wave can be described mathematically by

$$\boxed{\psi(r) = \frac{A}{r}\cos(kr - \omega t)}\ .$$

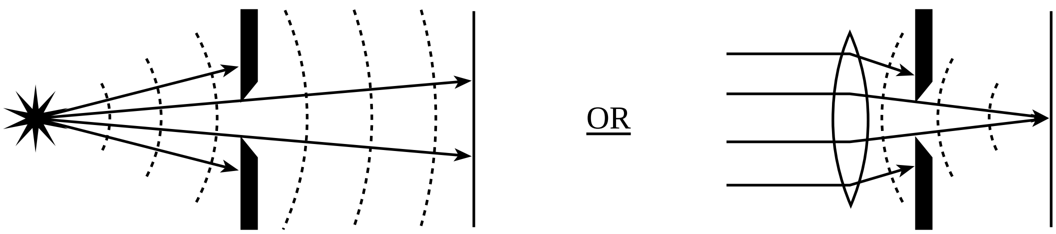
2. Fraunhofer (Far-Field) vs. Fresnel (Near-Field) Diffraction

One rarely attempts to calculate *exactly* the propagation of a light wave through an optical system and the space around it. There are 2 standard approximations to the exact formulation of wave propagation and diffraction: the Fraunhofer and Fresnel approximations. Given the mathematical complexity of the general formulations of these approximate theories (they rely heavily on multi-dimensional calculus), we only outline the regimes of their validity here, and give some specific examples in the next sections.

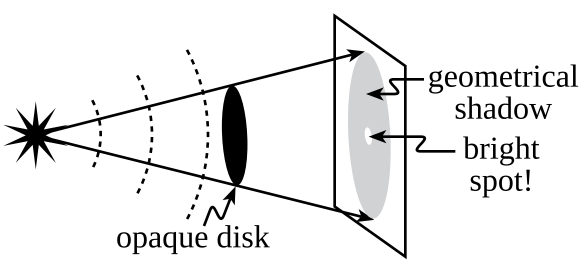
- *Fraunhofer (Far-Field) Regime:* In this regime the wavefronts are essentially plane waves at both the aperture (obstruction causing diffraction) and the screen (location we view the wave). The Fraunhofer description is mathematically simple and very intuitive, but it is only valid when the source, aperture, and screen are all very far apart or lenses are used to make plane waves.



- *Fresnel (Near-Field) Regime:* In this regime the wavefronts may be curved (e.g., spherical) at the aperture and at the screen — the math is thus more complicated. But, as a result we can describe the diffracted wave more accurately very close to an aperture.

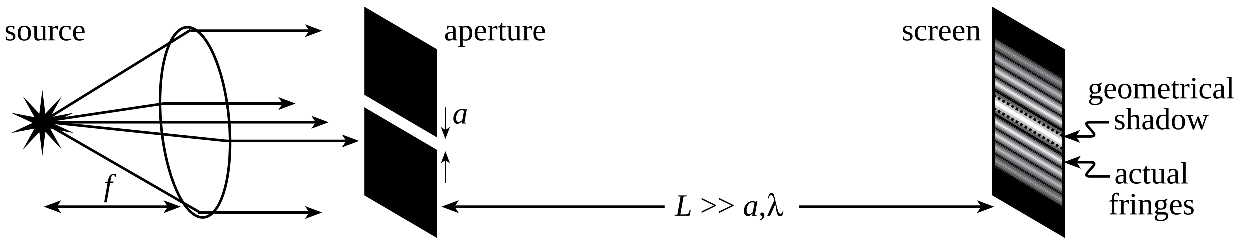


A very historically important example of Fresnel Diffraction is Poisson's Spot.

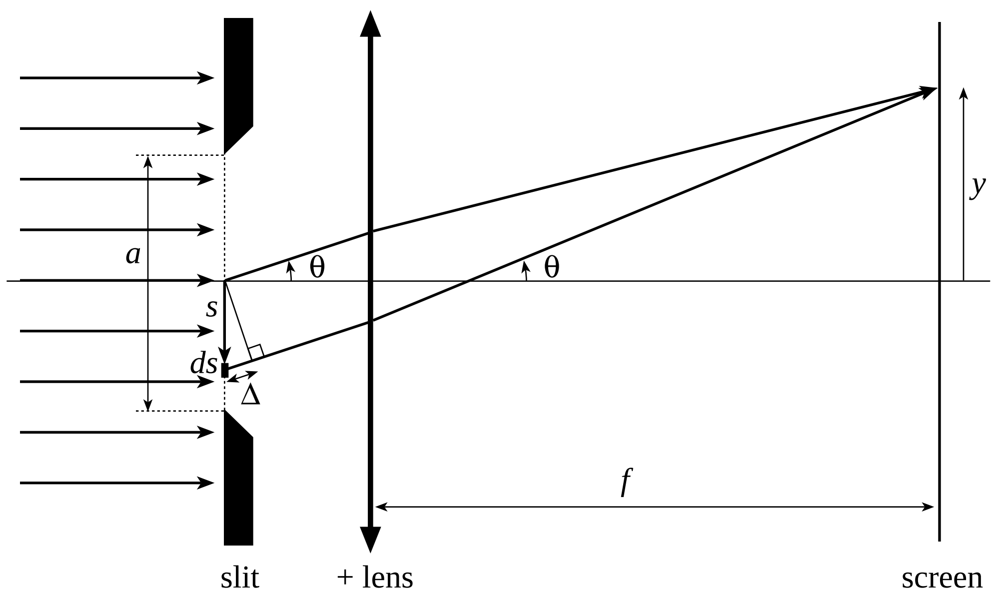


3. Diffraction of Light Passing Through a Narrow Slit

What do we expect to see if we try to "squeeze" light through a very narrow slit? First, the light should *spread out* on the other side of the slit (along the direction we tried to squeeze it), and second, because diffraction is equivalent to interference, we should see *interference fringes*.



Now let's try to predict this pattern more precisely using some mathematics. Look at a close-up.



According to the Huygens-Fresnel Principle, the total wave amplitude at a point y on the screen is the superposition of waves that originate from an infinite number of infinitesimally small point sources in the aperture region. We can think of each point s on the wavefront inside the aperture (where $-a/2 \le s \le a/2$) like a spherical-wave source with amplitude $A_s ds$ (where A_s is the amplitude per unit width, and ds is the width of each infinitesimally small source point).

If r_0 is the distance from the point $s = 0$ on the optical axis to a point y on the screen, then the contribution $d\psi$ to the total amplitude on the screen from the point at $s = 0$ is

$$d\psi(y) = \frac{A_s ds}{r_0} \cos(k r_0 - \omega t).$$

For off-axis points for which $s \neq 0$, the distance is longer or shorter than r_0 by an amount Δ, so

$$d\psi(y) = \frac{A_s ds}{r_0 + \Delta} \cos\left[k(r_0 + \Delta) - \omega t\right].$$

To find the total amplitude $\psi(y)$ we just add up all of the contributions from every point s on the aperture. But because there is an infinite number of infinitesimally small points (of size ds), the sum is actually an *integral*. That is

$$\psi(y) = d\psi_1 + d\psi_2 + d\psi_3 + \cdots = \int d\psi(y) = \int_{s=-a/2}^{s=a/2} \frac{A_s}{r_0 + \Delta(s)} \cos\left[k\left(r_0 + \Delta(s)\right) - \omega t\right] ds.$$

If we recognize that $\sin\theta = \Delta/s$, then we see

$$\Delta(s) = s\sin\theta.$$

Also, since $\Delta << r_0$, then we may approximate the first factor in the integrand as

$$\frac{1}{r_0 + \Delta} \cong \frac{1}{r_0}.$$

However, we should keep the Δ inside the cosine factor, since

$$\cos\left[k\left(r_0 + \Delta\right) - \omega t\right] = \cos\left(k r_0 - \omega t\right)\cos\left(k\Delta\right) - \sin\left(k r_0 - \omega t\right)\sin\left(k\Delta\right);$$

Thus regardless of how large Δ is relative to r_0, as long as it is large enough such that $k\Delta \sim 2\pi$, or $\Delta \sim \lambda$, then it can effect the size of the cosine term as much as r_0! Using these intermediate results, we rewrite the integral expression for $\psi(y)$ as

$$\psi(y) \cong \frac{A_s}{r_0} \int_{-a/2}^{a/2} \cos\left[k\left(r_0 + s\sin\theta\right) - \omega t\right] ds = \frac{A_s}{r_0} \int_{-a/2}^{a/2} \cos\left[\left(k\sin\theta\right)s + \left(k r_0 - \omega t\right)\right] ds.$$

All we need to do now is perform this integral. In case you are little rusty (or unfamiliar) with derivatives and integrals of trigonometric functions, notice that

$$\frac{d}{dx}\left[\frac{1}{a}\sin(ax+b)\right] = \cos(ax+b).$$

Therefore, from the Fundamental Theorem of Calculus, it must be true that

$$\int \cos(ax+b)\,dx = \frac{1}{a}\sin(ax+b) + \textit{some constant}.$$

Here $x = s$, $a = k\sin\theta$, and $b = kr_0 - \omega t$. Therefore we find for $\psi(y)$ the following

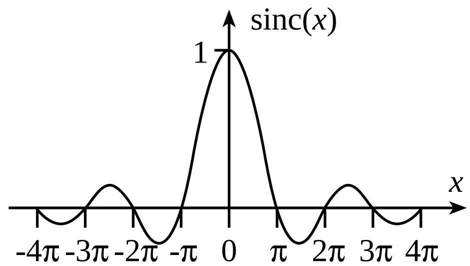
$$\begin{aligned}
\psi(y) &= \frac{A_s}{r_0 k \sin\theta} \sin\left[(k\sin\theta)s - \left(k r_0 - \omega t\right)\right]\Bigg|_{s=-a/2}^{s=a/2} \\
&= \frac{A_s}{r_0 k \sin\theta}\left(\sin\left[k a \sin\theta/2 - \left(k r_0 - \omega t\right)\right] + \sin\left[k a \sin\theta/2 + \left(k r_0 - \omega t\right)\right]\right) \\
&= \frac{\sin\left(k a \sin\theta/2\right)}{k a \sin\theta/2} \frac{A_s a}{r_0} \cos\left(k r_0 - \omega t\right)
\end{aligned}$$

where we have recognized that $\sin(-\theta) = -\sin(\theta)$ and used the trigonometric identity

$$\sin(\theta_1 + \theta_2) + \sin(\theta_1 - \theta_2) = 2\sin(\theta_1)\cos(\theta_2).$$

Before calculating the intensity, it is worth pointing out an interesting function that has naturally occurred in the analysis of diffraction from a slit: the first factor in the expression for $\psi(y)$ is called the sinc function, where

$$\boxed{\mathrm{sinc}(x) \equiv \frac{\sin(x)}{x}}.$$ This function looks like:



The time-averaged intensity is just the time-average of ψ^2, or

$$\langle I \rangle = \langle \psi^2(y) \rangle = \frac{\sin^2(ka\sin\theta/2)}{(ka\sin\theta/2)^2}\frac{A_s^2 a^2}{r_0^2}\langle \cos^2(kr_0 - \omega t)\rangle.$$

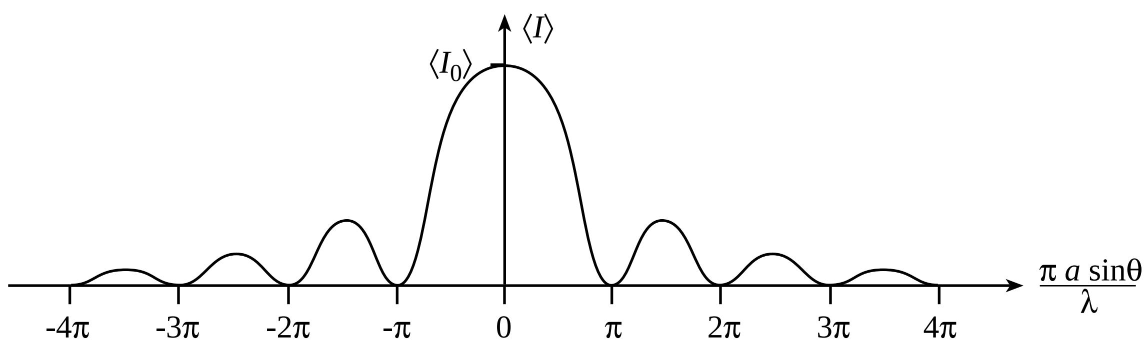
If we recognize the time-average of the $\cos^2$ function is simply 1/2, and that $ka\sin\theta/2 = \pi a\sin\theta/\lambda$, and we define the incident time-averaged intensity to be

$$\langle I_0 \rangle \equiv \frac{A_s^2 a^2}{2r_0^2},$$

where $\langle I_0 \rangle$ is just the intensity of the plane wave incident on the aperture, then the intensity is just

$$\boxed{\langle I \rangle = \langle I_0 \rangle \frac{\sin^2(\pi a\sin\theta/\lambda)}{(\pi a\sin\theta/\lambda)^2} = \langle I_0 \rangle\, \mathrm{sinc}^2(\pi a\sin\theta/\lambda)}.$$

Thus the time-averaged intensity has a peak in the center with smaller fringes on the sides.

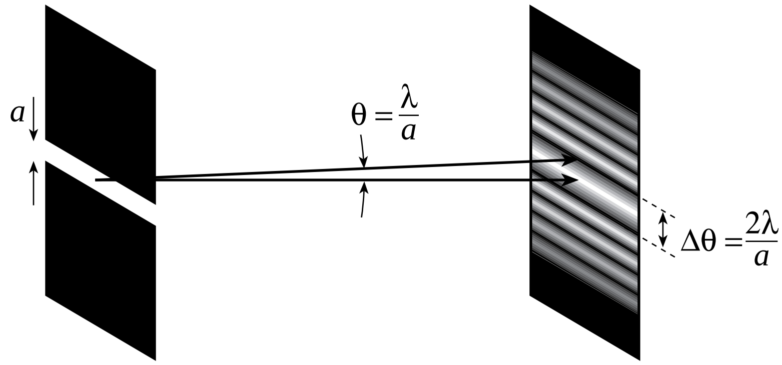


For small angles θ (i.e., the screen is far from the aperture and we look close to the optical axis), we may approximate $\sin\theta \cong \theta$. Then the first zeros on the sides of the central peak occur when

$$\frac{\pi a\sin\theta}{\lambda} \cong \frac{\pi a\theta}{\lambda} = \pi, \text{ or}$$

The angle (relative to optical axis) at which the first zeros occur is:	$\theta = \lambda/a$.

So on the screen we see a pattern that looks like the following.



If we use a lens to ensure that we are in the "far field," then $\theta \cong y/f$, where f is the focal length of the lens, and the zeros in the intensity pattern $\langle I(y)\rangle$ occur when

$$\frac{\pi a y}{\lambda f} = m\pi \quad (m = \pm 1, \pm 2, \ldots), \text{ or}$$

When the screen is in the focal plane of a lens the zeros occur at: $\quad y_m = m\frac{\lambda f}{a} \quad (m = \pm 1, \pm 2, \ldots).$

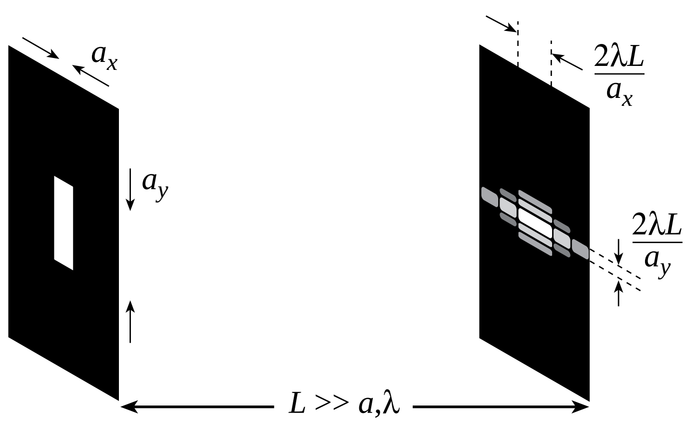
If we just look on a screen a distance L behind the aperture where $L >> a,\lambda$, then $\theta \cong y/L$, and so

When the screen is a distance $L >> a,\lambda$ *away the zeros occur at:* $\quad y_m = m\frac{\lambda L}{a} \quad (m = \pm 1, \pm 2, \ldots).$

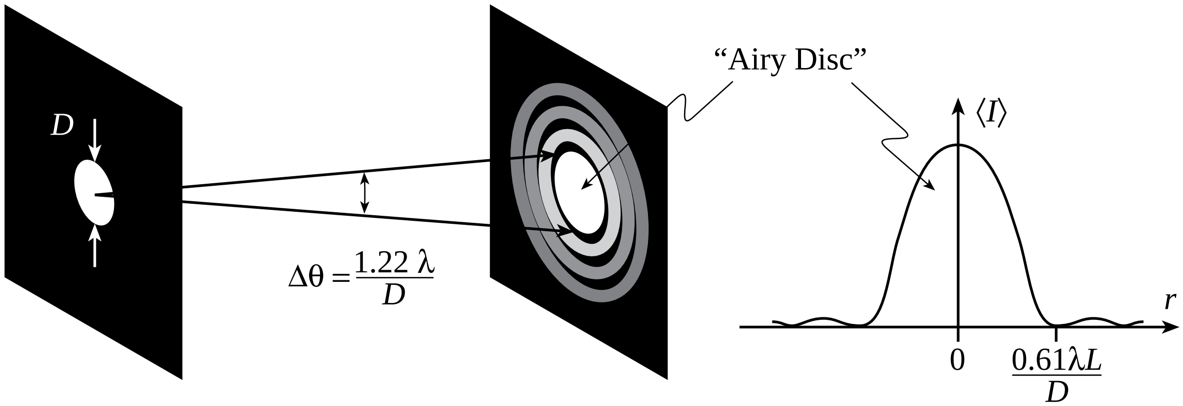
4. Diffraction from 2-Dimensional Apertures

a. Some Common 2-D Apertures:

Two common apertures that illustrate the characteristics of diffraction in 2 dimensions are the rectangular and circular apertures. Diffraction of light through a rectangular aperture is a rather straightforward extension of 1-dimensional diffraction from a slit, as shown in the diagram below.

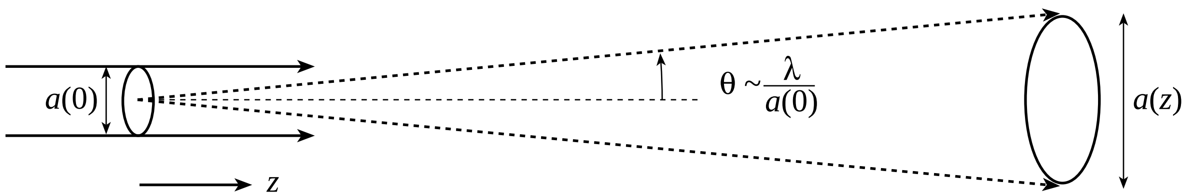


While the circular aperture is qualitatively similar, an accurate quantitative treatment of the pattern requires more complicated mathematics (using beasts called "Bessel Functions"). However, the main result is pretty simple, as shown in the diagram below.



b. Spreading of a Laser Beam by Diffraction:

Because a laser beam is inherently an attempt to confine light in the directions transverse to the propagation direction, the light naturally tries to spread out, just as if it were forced through an aperture!



As you might expect, the angle at which the light spreads is still given by

$$\boxed{\theta \sim \frac{\lambda}{a(0)}}, \text{ such that } \boxed{a(z) \sim \frac{\lambda z}{a(0)}}.$$

Because the laser beam diameter is typically much greater the wavelength of light, or $a(0) >> \lambda$, then the angle at which the beam spreads, θ, is quite small!

As an example, consider the HeNe laser, for which λ = 633 nm and a typical beam diameter is $a(0)$ ~ 0.5-1.0 mm ~ 0.633 mm. Therefore we find

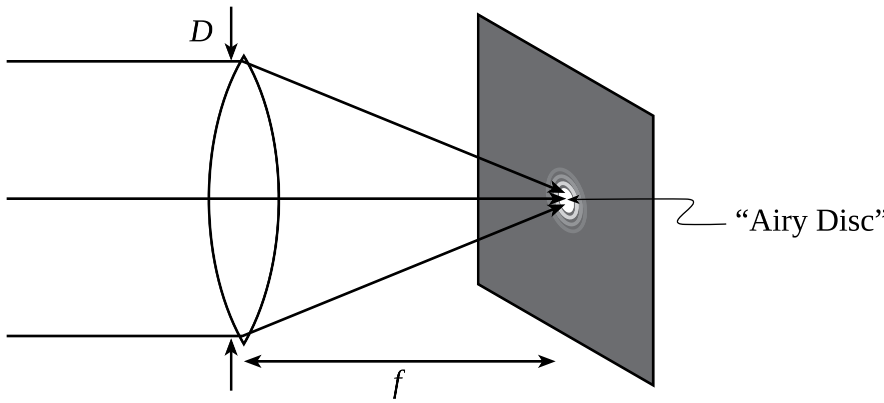
$$\theta \sim \frac{633 \times 10^{-9} m}{0.633 \times 10^{-3} m} = 10^{-3} \text{ radians} = 1 \text{ milliradian}.$$

If we ask how far the beam must propagate before the diameter increases by a factor of 10, we find

$$z \sim \frac{a(0)}{\lambda} a(z) = \frac{a(0)}{\lambda} 10\, a(0) = \frac{10\, a(0)^2}{\lambda} = 6.33 \text{ meters}.$$

c. Diffraction-Limited Spot at the Focal Point of a Lens:

In geometrical optics we assume that an ideal (aberration-free) lens focuses parallel rays to a single point one focal length away from the lens. In fact this is not true, since the finite diameter of the lens itself acts like an aperture for the incident light. Thus as soon as the light passes through the lens it begins to spread out, yielding a blurred spot at the focal point.



Because this phenomenon is equivalent to diffraction of light passing through a circular aperture, light around the focal point will exhibit an Airy Disc pattern. The only difference is that the size of the Airy Disc is now determined by the focal length f and diameter D of the lens:

The width w of the Airy Disc at the focal point of a lens is given by: $w = \dfrac{1.22\,\lambda\, f}{D}$.

If ray aberrations in an optical system can be controlled such that all of the rays leaving a given object point land inside of the Airy Disc associated with the corresponding image point, we say we have Diffraction-Limited Imaging. (This is the absolute best you can do for an optical system that has lenses of a finite diameter!)