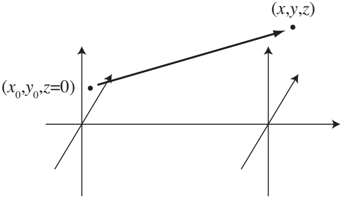
Adding up point sources

- Recall that we start with a wavefield at an input plane $(x_0, y_0, z = 0)$. We treat it as a bunch of Huygens point sources, each with magnitude and phase modulated by $\tilde{g}(x_0, y_0)$.
- To get the field at a downstream position (x, y, z), add up the contribution from the input plane sources:

$$\psi(x,y,z) = \psi_0 \frac{\lambda}{A} \int_{x_0} \int_{y_0} \tilde{g}(x_0, y_0) \frac{\exp[-ikr]}{r} \cos\theta. \qquad (1)$$

- The radius r is given by

$$\begin{aligned} r &= \sqrt{z^2 + (x - x_0)^2 + (y - y_0)^2} \\ &= z\sqrt{1 + \frac{(x - x_0)^2}{z^2} + \frac{(y - y_0)^2}{z^2}} \end{aligned} \qquad (2)$$



Fresnel and Fraunhofer

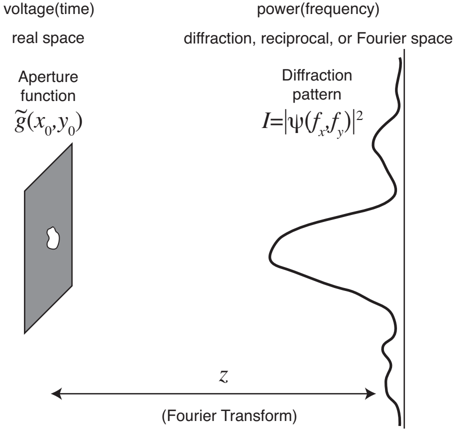
- The *Fresnel approximation* involved discarding terms like $(x-x_0)^4/(8z^4)$ in phase, and saying $1/r \simeq 1/z$ in magnitude.
- The *Fraunhofer approximation* discards the additional terms of $(x_0^2+y_0^2)/(\lambda z)$ in phase.
- This led to the Fresnel-Kirchoff diffraction integral, where in the Fraunhofer approximation we leave out the Fresnel term and where the out-of-integral phase can be ignored when calculating intensities:

$$\psi(x,y,z) = \psi_0\, e^{-i\pi\frac{x^2+y^2}{\lambda z}} \int_{x_0}\int_{y_0} \tilde{g}(x_0,y_0)\, e^{-i\pi\frac{x_0^2+y_0^2}{\lambda z}}\, e^{i2\pi(f_x x_0+f_y y_0)}\, dx_0\, dy_0 \tag{3}$$

where $f_x = x/(\lambda z)$ and $f_y = y/(\lambda z)$ are spatial frequencies.

Fraunhofer=Fourier transform

In the Franhofer approximation of Eq. 3, the far-field intensity $|\psi(f_x, f_y)|^2$ is just a Fourier transform of the input-plane wavefront modification $\tilde{g}(x_0, y_0)$:

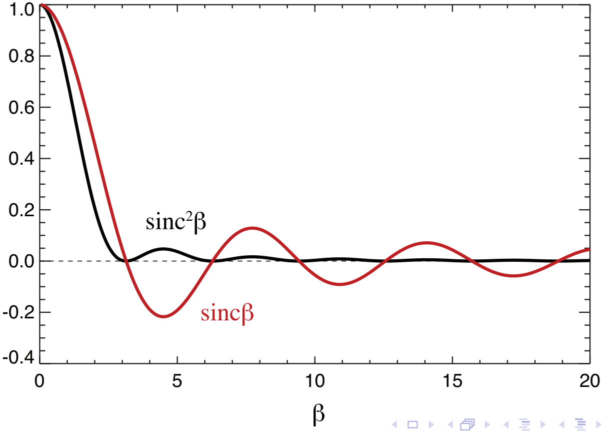


Diffraction from a slit

For a slit of width b, and defining $\beta \equiv \pi b f_x x$, we found in the Fraunhofer approximation

$$\psi = \psi_0 \frac{\sin(\beta)}{\beta} = \psi_0 \mathrm{sinc}(\beta) \tag{4}$$

leading to $I \propto \mathrm{sinc}^2(\beta)$.



Multiple slits

- For two slits, each of width b and separated by a, we found

$$I = I_0 \mathrm{sinc}^2(\beta) \cos^2(\alpha) \tag{5}$$

  with $\alpha \equiv \pi a f_x$.
- For N slits we found

$$I = I_0 \mathrm{sinc}^2(\beta) \left(\frac{\sin(N\alpha)}{\sin(\alpha)} \right)^2 \tag{6}$$

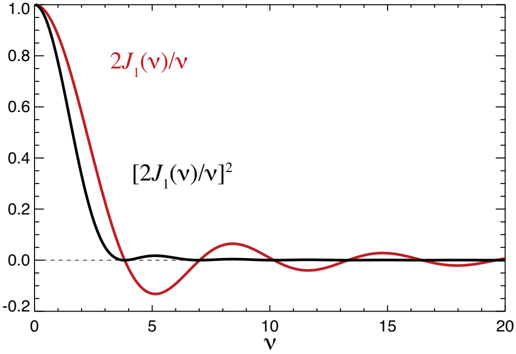
  By varying the width of the slits, certain interference fringes will be enhanced and others will be cancelled out.

Pinhole

For a pinhole with radius a, the Fraunhofer diffraction pattern is described by

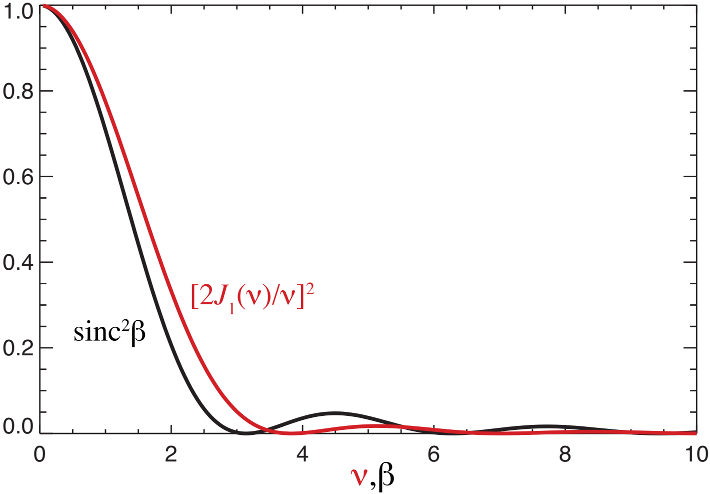
$$\psi = \psi_0 \frac{2J_1(\nu)}{\nu} \tag{7}$$

with $\nu \equiv 2\pi a f_r$, where $2J_1(\nu)/\nu$ is known as an Airy function. The diffraction intensity goes like $[2J_1(\nu)/\nu]^2$, with a first minimum at $\nu = 1.22\pi = 3.83$.



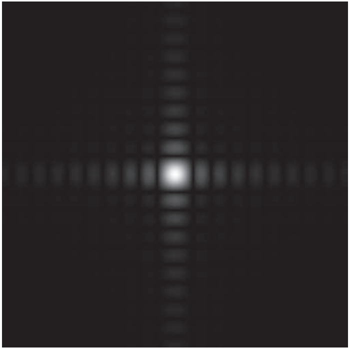
Airy and sinc

Here's a comparison of a pinhole's Airy $[\nu = \pi(2a)f]$ and a slit's sinc $[\beta = \pi bf]$ diffraction intensities:



Airy and sinc: 2D

Here's a comparison of the 2D diffraction intensity from square/sinc and circular/Airy apertures:



Square aperture

Circular aperture

Pinhole aperture diffraction

- Consider the focus of a plane wave by a lens with a radius r_0 and focal length f.
- The lens does two things:
  1. Provides a radial dependent phase shift (thicker glass in the middle slows the wave down more)
  2. Imposes a hard-edged circular aperture (called the *pupil function*)
- The aperture of the lens is like a pinhole from which you have diffraction. If the lens aperture (its pupil function) has a radius a, then a plane wave incident on the lens will not be focused to a point but rather an Airy pattern $[2J_1(\nu)/\nu]$ with $\nu = 2\pi ar/(\lambda f)$. The intensity $[2J_1(\nu)/\nu]^2$ will therefore have a first minimum at $\nu = 1.22\pi$, or

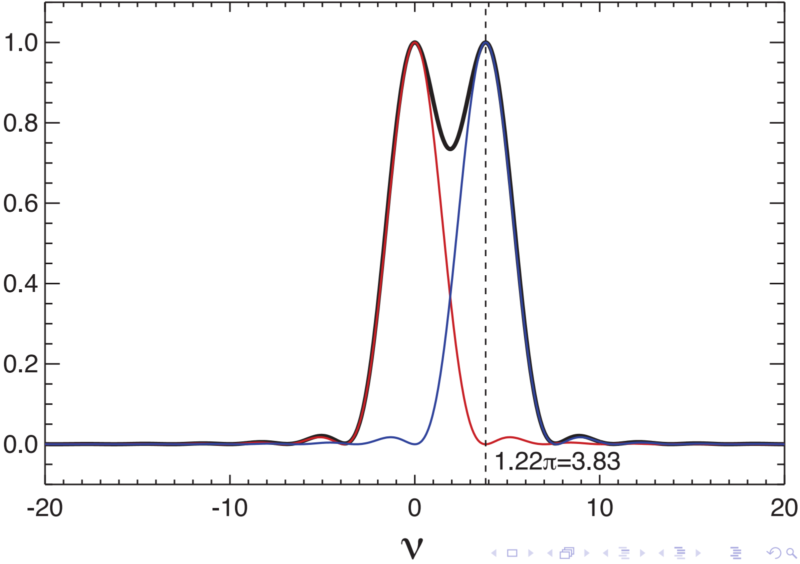
$$\begin{aligned}
\frac{2\pi a r_{\text{first min}}}{\lambda f} &= 1.22\pi \\
2\frac{a}{f}\frac{r_{\text{first min}}}{\lambda} &= 1.22 \\
r_{\text{first min}} &= 0.61\frac{\lambda}{\text{N.A.}} \quad \text{where} \quad \text{N.A.} = \frac{a}{f} \qquad (8)
\end{aligned}$$

Rayleigh's resolution criterion

- Again, we had $r_{\text{first min}} = 0.61\lambda/\text{N.A.}$ as being the position of the first minimum of the diffraction pattern of a circular lens.
- Now consider imaging two nearby stars with a telescope. The light from each star is emitted incoherently with respect to its neighbor (after all, the two stars may in fact be separated by gazillions of light-years in their distance from earth). Therefore the image that the telescope delivers is just the addition of the two intensities.
- Lord Rayleigh (John William Strutt) decided that a sensible criterion for deciding that you could tell if you saw two stars rather than one was to have the second star be centered at a position no closer than the first minimum of the diffraction pattern of the telescope's image of the first star.
- This same criterion applies to imaging two fluorescent molecules in a cell.
- Conclusion: the bigger the lens aperture angle r_0/f, the finer detail you can see.
- Of course aberrations increase as you open up the lens aperture, so there's often a balancing game to play for best resolution.

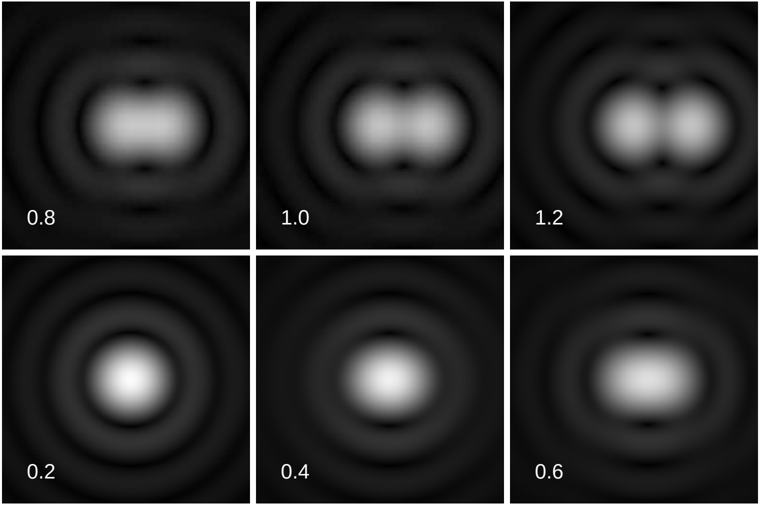
Rayleigh resolution illustrated

Here's Rayleigh's resolution criterion illustrated for two objects separated by at least a distance of $r_{\text{first min}}$:



Rayleigh resolution: 2D images

Here is what two spots look like as their separation is increased (the separation is expressed as a fraction of $r_{\text{first min}}$ of $0.61\lambda/\text{N.A.}$).



Rayleigh resolution example

- Let's estimate the resolution of the human eye. The eye length is 20 mm, and let's say the pupil diameter is 3 mm. It turns out that the numerical aperture also involves the index of any refractive medium which is about 1.3 for the vitreous humor (the fluid in the eye), so we have

$$\text{N.A.} = n \sin\theta = 1.3 \frac{0.5 \cdot 3 \text{ mm}}{20\text{mm}} = 0.0975$$

- The resolution on the retina is then $0.61 \cdot (500 \text{ nm})/0.0975$ or 3 μm. This is about the same size as the cones in the eye; evolution knows about the Rayleigh resolution!
- Look at an object at a near point of 25 cm. The magnification will be $(s/s') = (25/2) = 12.5$, so 3 μm at the retina corresponds to 38 μm at the piece of paper.
- Note that 38 μm corresponds to about $1/670^{\text{th}}$ of an inch, so now you see why 600 dots per inch or dpi is sufficient resolution for high quality inkjet or laser printers.
- Photographs can be represented by a series of dots below the resolution of the eye, with the density or diameter of the spots determining the reflectance within the eye's resolution limit.

Propagating wavefields

The expression of Eq. 3 was what we had for the Fresnel-Kirchoff diffraction integral in the case where we expanded the $(x-x_0)^2+(y-y_0)^2$ term in the approximation for r. Still within the Fresnel approximation [meaning we ignore terms in $(x/z)^4$], we can write an expression for r where we *do not* expand out the squares of the differences:

$$r \simeq z\left[1+\frac{(x-x_0)^2}{2z^2}+\frac{(y-y_0)^2}{2z^2}\right] \tag{9}$$

We will use Eq. 9 for the phase term $\exp[-ikr]$, again write $r \rightarrow z$ for the $1/r$ amplitude decrease-with-distance term, and again ignore the $\cos\theta$ obliquity term. In this case, we can write the Fresnel-Kirchoff diffraction equation as

$$\begin{aligned}\psi(x,y,z) &= \psi_0\frac{\lambda}{A}\int_{x_0}\int_{y_0}\tilde{g}(x_0,y_0)\frac{\exp[-ikr]}{r}\cos\theta \\ &= \psi_0\frac{\lambda}{z}\frac{1}{A}e^{-i2\pi\frac{z}{\lambda}} \\ &\qquad \int_{x_0}\int_{y_0}\tilde{g}(x_0,y_0)e^{-i\pi\frac{(x-x_0)^2}{\lambda z}}\, e^{-i\pi\frac{(y-y_0)^2}{\lambda z}}\end{aligned} \tag{10}$$

Propagating wavefields II

Let's look at the integral of Eq. 10, and consider only the 1D version for the moment:

$$\int \tilde{g}(x_0) e^{-i\pi \frac{(x-x_0)^2}{\lambda z}} \, dx_0$$

This should look familiar!!! Can anyone name a case where we've seen an integral of one function multiplied by another with a shift, integrating over all shifts?

Propagating wavefields III

Again, the integral from Eq. 10 that we want to consider is

$$\int \tilde{g}(x_0) e^{-i\pi \frac{(x-x_0)^2}{\lambda z}}, dx_0$$

You should recognize this as a convolution integral:

$$j(x) = \int g(a)h(x-a)\,da = g(x) * h(x), \tag{11}$$

And what do we know about convolution integrals? That if we take a Fourier transform of each of the functions, we can multiply their transforms and invert to do the convolution calculation, because

$$J(f) = G(f) \cdot H(f). \tag{12}$$

And what do we know about numerical evaluation of Fourier transforms? That they involve only $N \log_2(N)$ rather than N^2 calculation steps; that's why they're called Fast.

Propagating wavefields IV

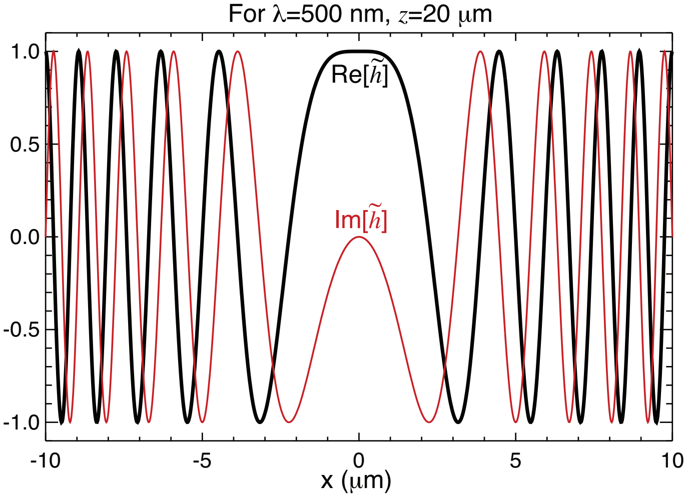
This is great news! We can re-write the Fresnel-Kirchoff diffraction integral from Eq. 10:

$$\begin{aligned}\psi(x,y,z) &= \psi_0 \frac{\lambda}{z}\frac{1}{A} e^{-i2\pi\frac{z}{\lambda}} \\ &\quad \int_{x_0}\int_{y_0} \tilde{g}(x_0,y_0) e^{-i\pi\frac{(x-x_0)^2}{\lambda z}}\, e^{-i\pi\frac{(y-y_0)^2}{\lambda z}} \\ &= \psi_0 \frac{\lambda}{z}\frac{1}{A} e^{-i2\pi\frac{z}{\lambda}} \left\{\tilde{g}(x_0,y_0) * \tilde{h}(x_0,y_0)\right\} \qquad (13) \\ &\text{with} \qquad \tilde{h}(x_0,y_0) \equiv e^{-i\pi\frac{(x_0^2+y_0^2)}{\lambda z}} \qquad (14)\end{aligned}$$

That is, the wavefield at a downstream plane is just given by the wavefield at an upstream plane [which is an a plane wave modulated by $\tilde{g}(x_0,y_0)$] convolved with a propagator function $\tilde{h}(x_0,y_0)$. This propagator function has a magnitude of 1 everywhere, so it only provides a phase modification of the object.

The propagator function

Again, the propagator function has a magnitude of 1 everywhere. Here's how its real and imaginary part varies:



Fourier transform of the propagator

Since convolution of a propagator involves multiplication with the Fourier transform of a propagator, it's useful to know what the Fourier transform of a propagator is:

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} \exp\left[-i\pi\frac{x_0^2}{\lambda z}\right] \exp[i\,2\pi x_0 f]\, dx_0 \\ &= \sqrt{\lambda z}\,\exp\left[i\pi\lambda z f^2\right] = \sqrt{\lambda z}\,\exp\left[i\pi\lambda z\left(\frac{x}{\lambda z}\right)^2\right] \\ &= \sqrt{\lambda z}\exp\left[i\pi\frac{x^2}{\lambda z}\right] \end{aligned} \tag{15}$$

where we have used the fact that spatial frequencies go like $f = x/(\lambda z)$. That is, the Fourier transform of a propagator is a propagator! This is not so surprising, for we said that the Fourier transform of a Gaussian is a Gaussian. . .

Wavefield propagation: final result

Whew! Now that we know that $H(f) = \sqrt{\lambda z} e^{i\pi\lambda z f^2}$, we can re-write the Fresnel-Kirchoff diffraction integral in the Fresnel approximation (Eq. 13) as

$$
\begin{aligned}
\psi &= \psi_0 \frac{\lambda}{z} \frac{1}{A} e^{-i2\pi\frac{z}{\lambda}} \left\{ \tilde{g}(x_0, y_0) * \tilde{h}(x_0, y_0) \right\} \\
&= \psi_0 \frac{\lambda^3}{A} \exp\left[-i2\pi\frac{z}{\lambda}\right] \mathcal{F}^{-1}\left\{ \mathcal{F}\left\{\tilde{g}(x_0, y_0)\right\} \cdot e^{i\pi\lambda z(f_x^2+f_y^2)} \right\} \qquad (16) \\
&= \psi_0 (\lambda^3/A) \exp[-i2\pi z/\lambda] \mathcal{F}^{-1}\left\{ \mathcal{F}\left\{\tilde{g}(x_0, y_0)\right\} \cdot e^{i\pi\lambda z(f_x^2+f_y^2)} \right\}
\end{aligned}
$$

It now becomes something we can calculate numerically:

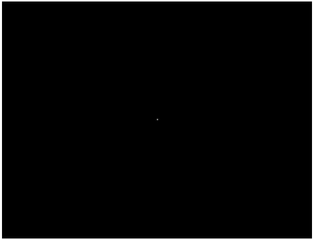
1. Take the Fourier transform of the input wavefield $\mathcal{F}\{\tilde{g}(x_0, y_0)\}$
2. Multiply by the Fourier transform of a propagator $\exp[i\pi\lambda z(f_x^2 + f_y^2)]$
3. Inverse transform the result $\mathcal{F}^{-1}\{\}$

Example: numerical wavefield propagation

Let's illustrate this by doing some numerical Fourier transforms on a digitized image of Jon Stewart, host of *The Daily Show*:

1. Take the square root of the image to convert it from intensity to amplitude
2. Take the Fourier transform (remember that we shift by $N/2$ on the input and output, as we illustrated in 1D!)
3. Square the result to look at the resulting far-field intensity

$\Rightarrow$



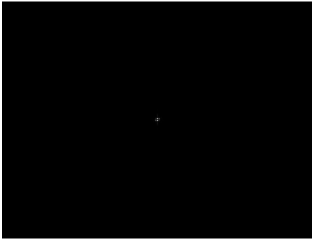
What the $#%?? Just a bright dot in the center? The Fourier transform of what gives a spike?

Removing the DC value

Aha! The Fourier transform of a flat function gives a delta function. So let's do this:

1. Take the square root of the image to convert it from intensity to amplitude
2. Subtract the average value of the image from the image, to give it a "DC" (direct current in circuits) value of zero.
3. Now take the Fourier transform and square it for intensity

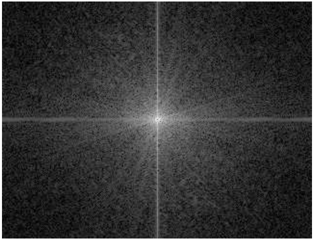
$\Rightarrow$



Well, this is still not terribly informative. Maybe the intensity scaling needs tweaking?

We need a log scale

Examining the values of our image, we find that the intensity varies from $10^{-1.1}$ to $10^{10.7}$! Therefore maybe it's better to display the image on a logarithmic scale, and show eight decades worth of intensity:

 $\Rightarrow$ 

This is looking better—but what about those horizontal and vertical streaks?

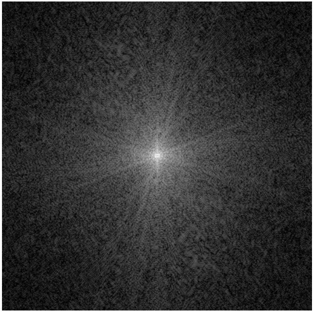
Repeated objects and edge ringing

It turns out that the discrete Fourier transform, with its finite "integration" bounds, has a property of assuming that the object repeats in an array:

Notice the discontinuities at the edges? They will diffract strongly!

Smoothing edges and embedding in a larger array

Solution: roll the edges off with a gaussian (here with $\sigma = 4$ pixels), and embed the image in a larger array (here we go from 322×246 to 512×512 since FFTs are fast with $N = 2^n$ pixels):

 $\Rightarrow$ 

OK, now we have a version of our image that is well-behaved for numerical Fourier transform calculations!