Prof. Raghuveer Parthasarathy
University of Oregon
Physics 352 – Winter 2008

2 DIFFRACTION

For centuries, debate raged over whether light is a wave or a particle – an interesting history that we won't go into. It turns out that light behaves as both a wave and a particle – it travels as a wave, as described in Section 1, but conveys energy as a particle, which won't concern us. The wave nature of light becomes important and apparent when considering spatial disturbances that are not large compared to the wavelength (λ), for example the sharp edge of a barrier, around which electromagnetic waves **diffract** (Figure 2.1). Like ripples in a pond, the waves can bend around the barrier, so the intensity of the light reaching a screen beyond the barrier does not reveal a perfectly sharp shadow, but rather a fuzzy one. Diffraction refers to phenomena like this, in which the wave nature of light and, as we'll see, its **interference** with itself determine its intensity profile. In general, the regime in which the wave nature of light is important is called **Physical Optics**. (The regime in which the system size is much greater than the wavelength of light, and hence wave properties are relatively unimportant, is called **Geometric Optics**, or **Ray Optics**.)

Diffraction is a general property of waves, and the phenomena we'll explore in this section apply to water waves, sound waves, etc., in addition to light.

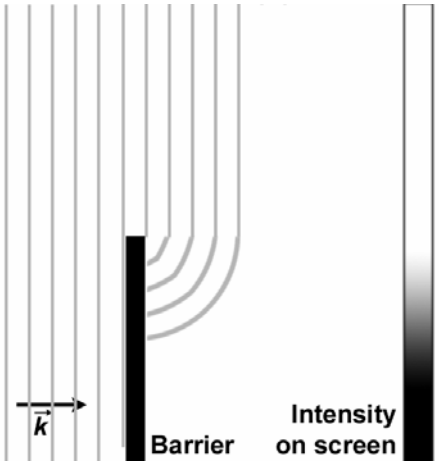

Figure 2.1. The wave nature of light is important when dealing with spatial features, for example the sharp edge of a barrier, that are comparable or smaller in size than the wavelength of light, for example the sharp edge of a barrier. Waves can bend around the barrier – think of ripples in a pond – and so the pattern of light reaching a screen intensity does not show a sharp "black or white" division but rather a "fuzzy" edge. Shadows are never sharp!

2.1 Two-Slit Interference

Consider a plane wave incident on a barrier with two slits, separated by a distance D (Figure 2.2). (Imagine the slits themselves to have negligible width – we'll come back to this later.) Each slit acts as a point-source for waves, which continue propagating to the right in the Figure. Far to the

right is a screen. We want to know: what is the **intensity** , I , of the light hitting the screen as a function of θ, the angle relative to a line perpendicular to the barrier (see Figure 2.2)?

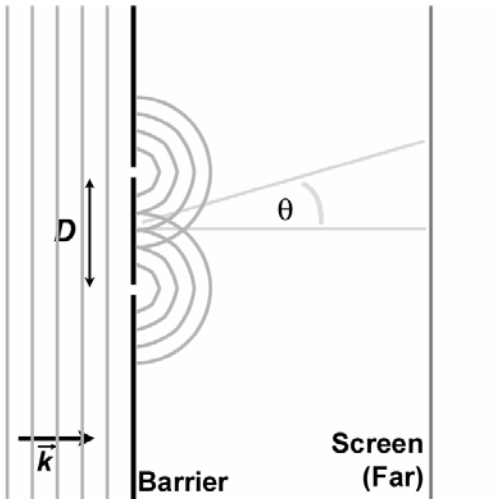

Figure 2.2. Two-slit interference: A plane wave is incident from the left on two slits of negligible width separated by distance *D*. Each slit acts like a point source for waves continuing to the right; the two resulting waves interfere with one another. This interference is manifested in the pattern of light intensity observed on a distant screen, and is a function of the wavelength, *D*, and the angle θ.

The electric field of the incident wave is

$$\vec{E} = \vec{E}_0 \exp\left[j\left(kx - \omega t\right)\right],$$

with $k = 2\pi/\lambda$, as usual (see Section 1). We could add any phase offset to this – it doesn't matter, as you'll see shortly. We're concerned with the light hitting a far off screen, at angle θ. If the screen were close by, a ray would have to leave slit #1 at some angle θ_1 and slit #2 at some angle θ_2, where θ_1 and θ_2 may be different, to both reach the screen at angle θ. However, as the screen moves farther and farther away, both θ_1 and θ_2 approach θ – try drawing this if you don't understand. So, to consider $I(\theta)$ we need to consider rays leaving each slit at angle θ. Let's define our coordinates so that the barrier is at $x = 0$.

The two rays that travel at angle θ are indicated in Figure 2.3; their fields are

$$\vec{E}_1 = \vec{E}_0 \exp\left[j\left(ks - \omega t\right)\right]$$

$$\vec{E}_2 = \vec{E}_0 \exp\left[j\left(k\left(s + \delta\right) - \omega t\right)\right]$$

where we've defined s as the coordinate in the "tilted" θ direction, and we've indicated the extra distance that ray 2 has to travel by δ. Note that $\vec{E}_1(s = 0)$ and $\vec{E}_2(s = -\delta)$ have the same phase, as they should since they come from the same incident wave.

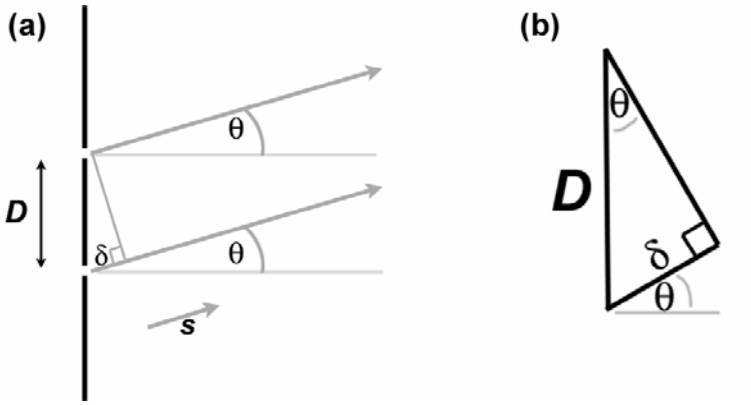

Figure 2.3. The geometry of light propagation for two-slit interference (see Figure 2.2). (a) For the angle θ illustrated, light traveling from the lower slit (slit #2) travels a greater distance than light from slit #1. The extra path length is denoted δ, and is the reason for a phase difference between the two waves. (b) A "zoomed in" view of the geometry relating *D*, δ, and θ.

Graphically, we can see that if δ is an integer multiple of λ, the two waves will add constructively (Figure 2.4).

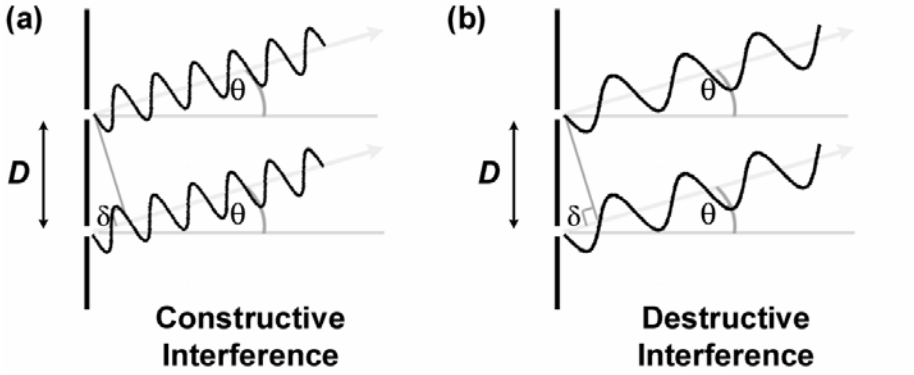

Figure 2.4. Interference. (a) If the extra path length, δ, between the two paths is an **integer** multiple of wavelengths, the two waves will constructively interfere, leading to high intensity at the screen. (b) If the extra path length δ between the two paths is a **half-integer** multiple of wavelengths, the two waves will destructively interfere, leading to zero intensity at the screen – note that when wave #1 is "up," wave #2 is "down" and vice versa.

If δ is a half- integer multiple of λ, the two waves will add destructively, and give zero light intensity.

Let's examine this mathematically. The superposition of the two electric fields:

$$\vec{E} = \vec{E}_1 + \vec{E}_2 = \vec{E}_0 \exp\left[j\left(ks - \omega t\right)\right]\left\{1 + \exp\left[jk\delta\right]\right\}.$$

From geometry, $\delta = D\sin\theta$ (Figure 2.3b), so

$$\vec{E} = \vec{E}_0 \exp\left[\, j\left(ks - \omega t\right)\right]\left\{1 + \exp\left[\, jkD\sin\theta\right]\right\}$$
$$\vec{E} = \vec{E}_0 \exp\left[\, j\left(ks - \omega t\right)\right]\left\{1 + \exp\left[2\pi\, jD\sin\theta/\lambda\right]\right\}.$$

The intensity (see Sections 1.1.3 and 1.2.3) is given by $I \propto \left|\vec{E}\right|^2 = \vec{E}\cdot\vec{E}^*$. Therefore

$$I \propto \left|\vec{E}_0\right|^2 (1)\left\{1 + \exp\left[2\pi\, jD\sin\theta/\lambda\right]\right\}\left\{1 + \exp\left[-2\pi\, jD\sin\theta/\lambda\right]\right\}$$
$$I \propto \left|\vec{E}_0\right|^2 \left[2 + 2\cos\left(\frac{2\pi}{\lambda} D\sin\theta\right)\right],$$

making use of the Euler relation $\cos(x) = \frac{1}{2}\left(\exp(jx) + \exp(-jx)\right)$. Via the identity $2\left[1 + \cos(2x)\right] = \cos^2(x)$, the intensity becomes

$$\boxed{I \propto 4\left|\vec{E}_0\right|^2 \cos^2\left(\pi \frac{D\sin\theta}{\lambda}\right)}$$

Note that *without* interference – just considering the incident plane wave, for example, $I \propto \left|\vec{E}_0\right|^2$, with the same constant of proportionality (c's etc.) – we'll define this intensity as I_0. Therefore:

$$\boxed{I = 4I_0 \cos^2\left(\pi \frac{D\sin\theta}{\lambda}\right)}.$$

As we saw graphically, if $D\sin\theta = m\lambda$, where m is an integer, the $\cos^2$ factor is maximal and we have **constructive** interference. If $D\sin\theta = \frac{m}{2}\lambda$, where m is an odd integer (i.e. $\frac{m}{2} = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$), the $\cos^2$ factor is zero and we have **destructive** interference. The intensity pattern we see on the screen, therefore, is **not uniform** but rather has a sequence of maxima and minima. This is plotted in Figure 2.5.

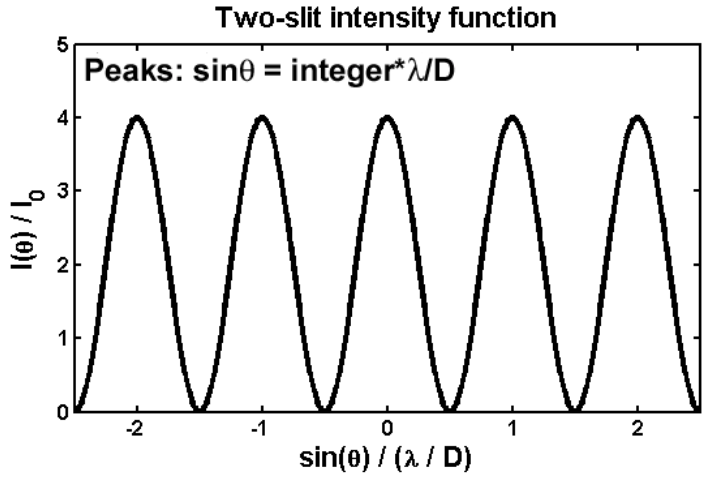

Figure 2.5. The two-slit intensity function: $I = 4I_0 \cos^2\left(\pi \frac{D\sin\theta}{\lambda}\right)$.

Note that the maximal value of the intensity is 4 times that of a single wave. If interference "didn't exist" we'd have light from the two slits combining to simply give twice the single-wave intensity. With interference, we have bright peaks with 4 times the intensity and dark minima with zero intensity.

2.2 *N*-Slit Interference

Now consider *N* slits, *each* separated by distance *D* (drawn in Figure 2.6) for *N*=5.

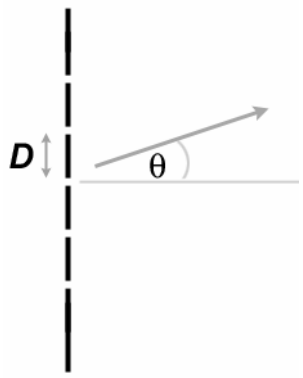

Figure 2.6. *N*-slit interference – each slit is of negligible width and is separated from its neighbor by distance D. In the example drawn, *N*=5.

Building on our *N*=2 analysis in Section 2.1, we can write the total electric field as

$$\vec{E} = \vec{E}_0 \exp\left[j\left(ks - \omega t\right)\right]\left\{1 + \exp\left[j\alpha\right] + \exp\left[j2\alpha\right] + \exp\left[j3\alpha\right] + ... + \exp\left[j(N-1)\alpha\right]\right\},$$

where for convenience we've defined $\alpha \equiv \frac{2\pi}{\lambda} D \sin\theta$. Note that this is

$$\vec{E} = \vec{E}_0 \exp\left[j\left(ks - \omega t\right)\right]\left\{1 + \left(\exp\left[j\alpha\right]\right) + \left(\exp\left[j\alpha\right]\right)^2 + \left(\exp\left[j\alpha\right]\right)^3 + ... + \left(\exp\left[j\alpha\right]\right)^{(N-1)}\right\},$$

i.e. the terms in the braces form a finite geometric series, since each term is equal to the preceding one times $e^{j\alpha}$.

As we all should know, the summation of a geometric series is:

$$\sum_{k=0}^{\infty} r^k = 1 + r + r^2 + r^3 + ... = \frac{1}{1-r}.$$

In case you've forgotten the proof, here it is: Define *S* as $S \equiv \sum_{k=0}^{\infty} r^k$. Therefore $S - 1 = r + r^2 + r^3 + ... = r(1 + r + r^2 + r^3 + ...) = rS$. Since $S - 1 = rS$, $S = \frac{1}{1-r}$.

The truncated sum $\sum_{k=0}^{N-1} r^k = 1 + r + r^2 + r^3 + \ldots r^{N-1} = \sum_{k=0}^{\infty} r^k - r^N \sum_{k=0}^{\infty} r^k$, as you can easily verify by writing out the terms. Therefore $\sum_{k=0}^{N-1} r^k = \frac{1}{1-r} - r^N \frac{1}{1-r} = \frac{1-r^N}{1-r}$.

Applying this to our electric field expression:

$$\vec{E} = \vec{E}_0 \exp\left[\, j\left(ks - \omega t\right)\right]\left(\frac{1-\exp(jN\alpha)}{1-\exp(j\alpha)}\right).$$

We can simplify the expression in the parentheses by factoring out exponentials from the numerator and denominator:

$$\frac{1-\exp(jN\alpha)}{1-\exp(j\alpha)} = \frac{\exp\left(-jN\alpha/2\right)\left[\exp\left(jN\alpha/2\right) - \exp\left(-jN\alpha/2\right)\right]}{\exp\left(-j\alpha/2\right)\left[\exp\left(j\alpha/2\right) - \exp\left(-j\alpha/2\right)\right]} = \exp\left(-j\left(N-1\right)\alpha/2\right)\frac{\sin\left(\frac{N\alpha}{2}\right)}{\sin\left(\frac{\alpha}{2}\right)},$$

using the Euler relation $\sin(x) = \frac{1}{2j}\left(\exp(jx) - \exp(-jx)\right)$.

Therefore: $\vec{E} = \vec{E}_0 \exp\left[\, j\left(ks - \omega t\right)\right]\exp\left(-j\left(N-1\right)\alpha/2\right)\frac{\sin\left(N\alpha/2\right)}{\sin\left(\alpha/2\right)}$.

The intensity $I \propto \left|\vec{E}\right|^2$ – it's easy to take the complex conjugate of the above expression and multiply, since each of the exponential terms simply yields 1, from which we find:

$$\boxed{I = I_0 \frac{\sin^2\left(N\alpha/2\right)}{\sin^2\left(\alpha/2\right)}}.$$

Explicitly writing the α's:

$$\boxed{I = I_0 \frac{\sin^2\left(N\pi D \sin\theta/\lambda\right)}{\sin^2\left(\pi D \sin\theta/\lambda\right)}}.$$

This is plotted in Figures 2.7 and 2.8.

Maxima and minima. We see that the *numerator* of $I(\theta)$ is zero when $N\pi D \sin\theta/\lambda = m\pi$, i.e. $D\sin\theta/\lambda = m/N$, where m is an integer – but note that **both** numerator and denominator are zero if m is an integer multiple of N. We see that the *denominator* is zero when $\pi D \sin\theta/\lambda = m'\pi$, i.e. $D\sin\theta/\lambda = m'$, where m' is an integer – in this case, however, the numerator must also be zero since $N\pi D \sin\theta/\lambda = Nm'\pi$, and N is an integer. If both the numerator and denominator are zero you can verify (L'Hopital's rule) that $I \to I_0 N^2$. I'll leave a more detailed summary of the locations of maxima and minima as an exercise.

Let's illustrate $I(\theta)$ with a plot; we'll choose $N = 6$ slits (Figures 2.7 and 2.8). There are large maxima separated in angle by $\sin\theta = \lambda/D$. From your analysis of $I(\theta)$, you should find that this **angular spacing** between the peaks is independent of the number of slits. The **angular width** of the large peaks is approximately $\Delta\sin\theta \approx \lambda/ND$ – half the distance in angle to the first local

minimum – which gets sharper as we increase the number of slits – a very useful feature, as we'll see shortly.

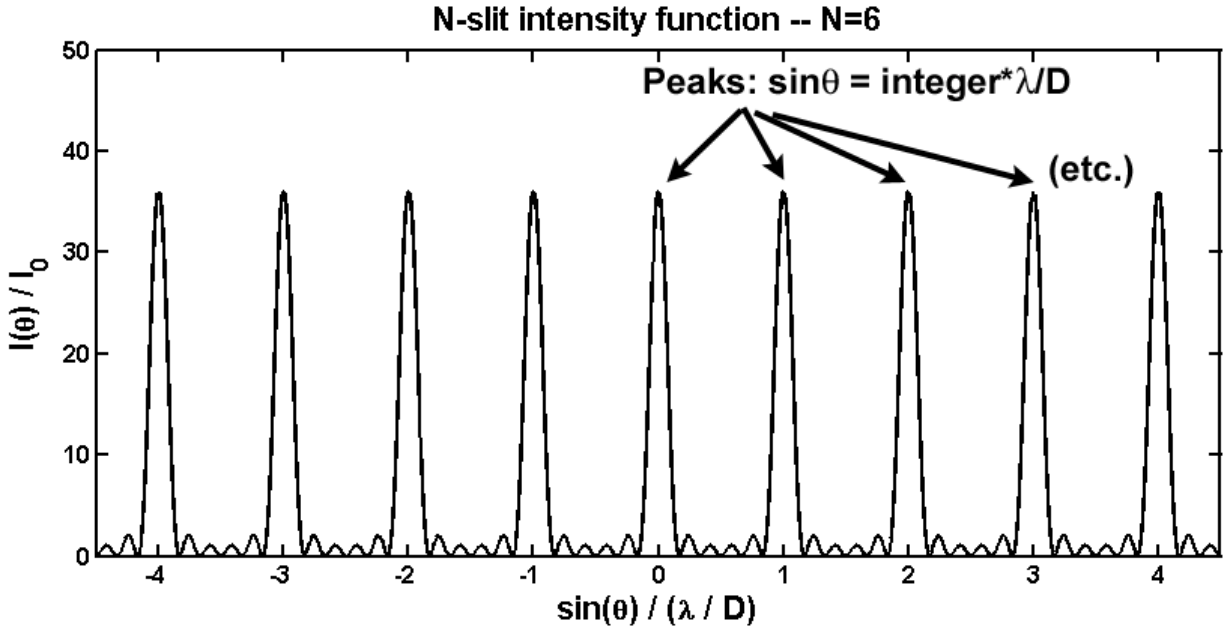

Figure 2.7. The *N*-slit intensity function, $I = I_0 \dfrac{\sin^2\left(N\pi D \sin\theta / \lambda\right)}{\sin^2\left(\pi D \sin\theta / \lambda\right)}$, plotted for *N*=6.

Note that there are infinitely many large maxima separated in angle by λ / D. Between each pair of these large peaks are *N*-2 smaller maxima and *N*-1 zeros.

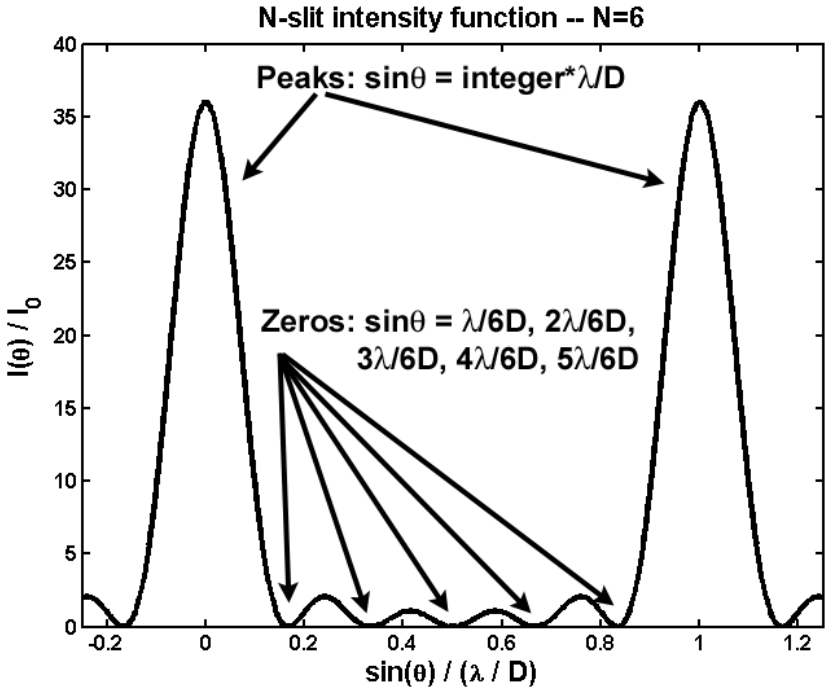

Figure 2.8. The same *N*-slit interference pattern as plotted in the previous figure, "zoomed in" to show the "zeroth" and first peaks.

2.3 The Diffraction Grating as a Monochromator

Suppose we have a telescope that collects light from a star, and we want to measure the star's spectrum – i.e. the intensity as a function of wavelength, $I(\lambda)$. How can we do this? Our detector (like most good detectors, at least over some range of wavelengths) simply measures intensity, regardless of the wavelength of the light hitting it.

We can pass the light though an N-slit grating, or, equivalently, reflect it off a surface with N mirrors – a **diffraction grating**. How does this help? Light of wavelength λ_1 is deflected to angle λ_1 / D. By this we mean that the maximal intensity peak for light of this "color" is at the angle given by $\sin\theta_1 = \lambda_1 / D$, and integer multiples, as in Section 2.2; typically, the angles involved are small, so $\sin\theta \approx \theta$. Light of wavelength λ_2 is deflected to angle λ_2 / D, etc. So moving our detector to various positions on the screen and measuring the intensity as a function of *angle* on the screen reveals the intensity as a function of wavelength! (In other words $I(\lambda_1) = I(\theta_1)$, $I(\lambda_2) = I(\theta_2)$, etc.).

The *sharper* the diffraction peaks (high N) the *finer* the resolution in λ – see the end of Section 2.2.

Extrasolar planets. The discovery (within the past ≈ 10 years) of planets outside our solar system – one of the most remarkable discoveries of recent history – used exactly the approach outlined above to measure tiny shifts in stellar spectra due to the influence of the orbiting planets. The N of the diffraction grating was around 100,000 !

2.4 Single-Slit Interference

In our initial discussion of two-slit interference we neglected the finite width of the diffraction grating. This finite width is important – just as waves from each slit interfere with one another, waves traversing various paths through a *single* slit will interfere with one another, and lead to diffraction. Fortunately, it is easy to analyze single-slit interference – it is simply the limit of the N-slit case discussed in Section 2.2 as $N \to \infty$, $D \to 0$, and the product $ND \to a$, where a is the width of the slit.

You'll examine this in your homework, and show that

$$\boxed{I(\theta) = I_0 \left(\frac{\sin\beta}{\beta} \right)^2}, \text{ where } \boxed{\beta = \frac{\pi a \sin\theta}{\lambda}}, \text{ as plotted in Figure X.}$$

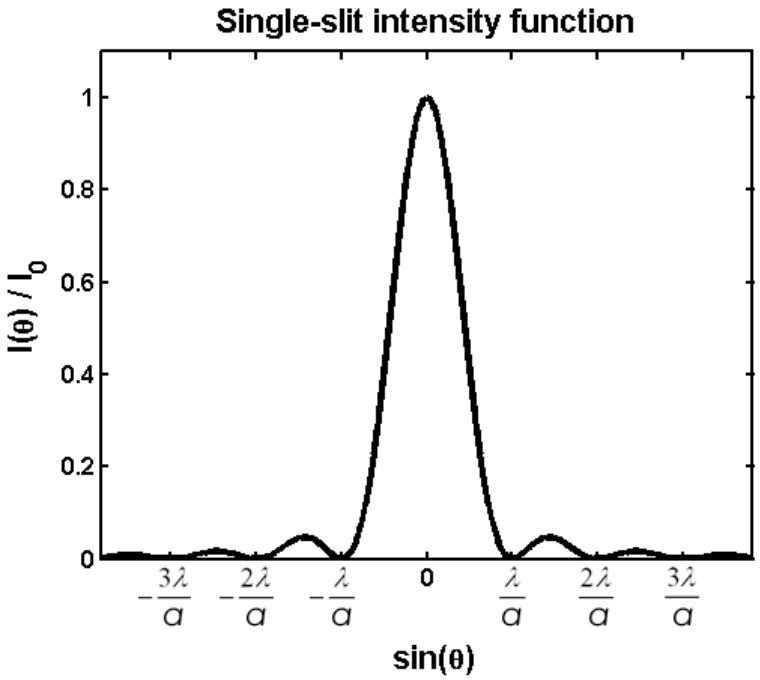

Figure 2.9. The intensity function of a single slit of width *a*. Note that the angular width of the peak is approximately λ / a.

2.5 Diffraction and Resolution

This single-slit diffraction pattern is exceptionally important. Any optical element – the pupil of your eye, a telescope mirror, a microscope lens, etc – is an aperture, and the $I(\theta)$ above describes how light travels through it. Why?

We've been considering light leaving an aperture, i.e. being "transmitted," and reaching a screen, where it is "received." But look carefully at Figures 2.2, 2.4, or 2.6 – the setup of our wave interference scenarios. Our analysis didn't invoke at all the *direction* the waves were traveling, only the path length difference between various paths. So we would get the same interference effects if light were *transmitted* from a point source at angle θ on the screen, passed through aperture(s), and were detected at the left.

Consider light from a point source (e.g a star) located at the screen (e.g. the sky). We observe the point source by detecting the intensity passing through a single-slit aperture of width a (e.g. a telescope lens plus intensity detector). We tilt the barrier containing our slit (e.g. our telescope) so that the angular position of the star of interest is $\theta_1 = 0$; this angle gives the maximum of the single slit intensity function, and we happily detect light from the star. We tilt the telescope; at the new $\theta_2 = 0$ there is no star; we see no light. We tilt further; at this third $\theta_3 = 0$ there is light again. "Aha!," we say, "We have seen two stars!"

Now suppose there were two stars very close to one another in angular position – let's say the difference in $\sin\theta$, is just $0.1\lambda / a$. (We typically deal with small angles, by the way, so $\sin\theta \approx \theta$.) Since the width of our interference function is $\approx \lambda / a$, no matter how precisely we

point at one star, we'll be detecting a sizeable fraction of the intensity of the other – *there is no way we can tell that we're looking at two stars rather than only one*!

The **angular limit of resolution**, often just referred to as the **resolution**, of our single-slit aperture – the minimum angular separation that two objects must have in order to be able to distinguish them – is $\boxed{\theta_{res} \approx \lambda / a}$, where a is the aperture size. (It's an "approximately equals" sign because there are different ways of defining criteria for distinguishability that won't concern us; most commonly, one uses the "Rayleigh criterion" $\theta_{res} = 1.22\,\lambda / a$.) Note that smaller θ_{res} means that we can more finely distinguish objects – we can "see" better – and that this can be achieved by increasing the size of our aperture. This is why one builds big telescopes. (Big telescopes have another, *unrelated*, advantage: they collect more light.)

This issue of diffraction sets the fundamental limit on the performance of telescopes, microscopes, etc. (Though as we'll see later, the past decade or so has seen clever ways around this "diffraction limit" in certain contexts in microscopy.)