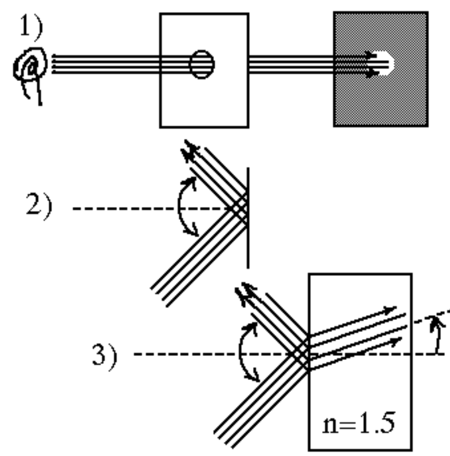
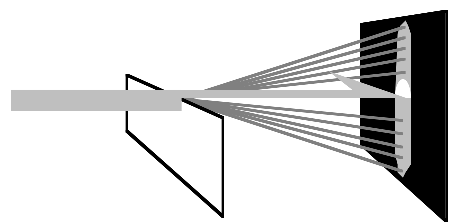
Chapter 5: Diffraction

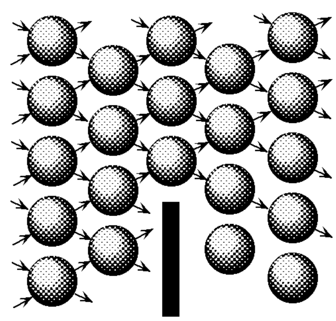
Most of our intuition about light is based on *ray* or *geometrical* optical concepts. These are based, in turn, on three basic premises: 1) that light "particles" travel in straight lines (what we normally think of "rays") until they hit something, 2) that when they hit a reflector, the angle of reflection equals the angle of incidence, and 3) when they hit a material of different refractive index, part of the light reflects (called the Fresnel reflection) and part continues at a different angle (measured to the perpendicular) determined by Snell's Law[1]. These three "laws" account for 99+% of what we see in everyday life, but they are only an approximation. When we deal with highly coherent light, such as from lasers, diffraction and interference effects become much more prominent than usual. These are usually described by *wave* or *physical* optical concepts, and are more complex and accurate than geometrical-optical concepts. Of course, in the limit of low coherence light the two approaches must both agree to within acceptable accuracy, and they do.

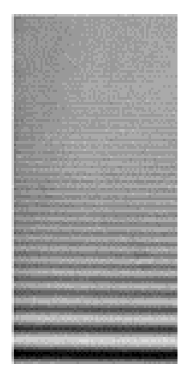


A simple experiment will show the limitations of the first premise, that light travels in completely straight lines. After this, you will believe that almost anything is possible! Consider an undiverged laser beam headed toward a white wall; it forms a single spot on the wall, perhaps 2 mm in diameter. Bring a razor blade slowly up into the beam, about a meter from the wall; mostly, you will see the spot being cut off, going through a half-round phase, and then being extinguished as the beam reflects off the solid blade. This is the "geometrical shadow" of the blade in the beam. But if you look closely while the blade edge is within the beam, you will see a streak of light above and below the geometrical shadow of the blade that carries a small percentage of the incident light. Now, you might think that the light <u>above</u> the shadow comes from a reflection from the razor blade's edge, as though it were a half-cylinder, but that doesn't explain the dark fringes in that light, and it certainly doesn't explain the light found <u>below</u> the shadow, which seems to veer around the edge as though deflected by some strange attractive force. If you put your eye in the streak (be careful to avoid the straight-through beam!), you will see that the light comes only from the edge of the razor blade.



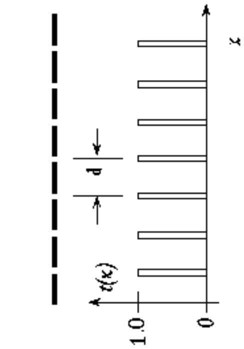
This "non-straight-line" behavior of light is a simple wave-optical phenomenon called *diffraction*. An explanation was first offered by Huyghens around 1678 [2]. He said that it was reasonable to consider air (and also vacuum) as a volume filled with imaginary spheres, like closely packed marbles, and that light was like a "nudge" from one of those spheres, which would nudge all the adjacent spheres, which would nudge all their neighbors, and so on and so forth (something like a pan full of "klackers"). If a sphere was equally nudged by neighbors to the left, left-up, and left-down, it would move to the right (a vector addition of the nudges), and nudge only that neighbor. Thus a wide nudge wave would propagate in a single direction. But, if a wall is stuck in among the spheres, a sphere just to the right of the top of the wall gets no nudges from below, and thus gets a net downward nudge from what is left of its neighbors. Thus a nudge starts propagating downward and forward into the "shadow" that should be cast by the wall. The story was fleshed out by Fresnel in the 1820's [3]. He proposed that the nudges were periodic and even sinusoidal. Thomas Young had earlier anticipated some of the implications of periodicity, and argued that the "nudges" were actually side-to-side vibrations (lateral, not longitudinal excitations) of the medium [4]. Maxwell then showed (in the 1870's) that these are lateral oscillations of coupled electrical and magnetic fields. It is all a fascinating story about the slow overcoming of a set of very strongly held beliefs in the particle theory of light established by Isaac Newton in the early 1700's, and I recommend browsing a book like Ronchi's *The Nature of Light* for more of that history [5].





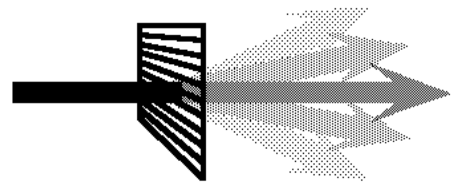
Diffraction pattern formed by a horiz. knife edge.

Diffraction by periodic structures:



For discussing holography, we will concentrate on diffraction effects caused by repetitive or periodic structures, such as evenly-spaced slits in an opaque screen (like a picket fence). We usually describe these by a transmittance function, $t(x,y)$, and periodicity in the x-direction with spacing d means, in mathematical terms, that $t(x\text{-}md\,y) = t(x,y)$, where m is any integer. Note that, because light will generally be moving from left to right across the page (for a while, at least), along the z-axis, the x-axis is drawn vertically here, as it will be in most of our sketches. For these examples, there will be no y-dependence of the transmittance pattern.

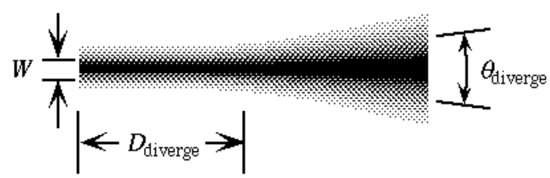
When an undiverged laser beam hits such a periodic structure, it breaks up into several laser beams deflected upwards and downwards by multiples of a certain angle determined by the spacing and the wavelength of the light. Actually, the trigonometric sines of the angles are multiples of the sine of a certain angle, up to the limit of ±1.0. The relationship is described by the simple form of the *diffraction equation* (which we will soon have to prove):



$$\sin\theta_{out,m} = m\sin\theta_{+1} = m\frac{\lambda}{d}, \quad m = 0, \pm 1, \pm 2, \ldots \; . \tag{1}$$

Single-slit diffraction:

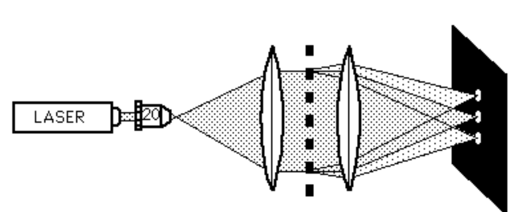
The situation is actually more complex than it might seem at first glance. Even if the laser beam goes through an empty frame, it eventually starts to diverge, to expand at some constant diameter increase per distance, which is to say that it diverges with some constant angle, θ_{diverge}. Even if the wavefronts are carefully collimated when they come out of the laser, by some distance downstream they will have become spherical, diverging from a point at the front of the laser. This is all due to the fact that the laser beam is like a uniform and infinite plane wave that has suddenly passed through a circular aperture or window in order to get out of the laser. That constriction at one end causes the beam to spread out at its other end, and the smaller the constriction width, W, the larger the angle of divergence, θ_{diverge}. This effect is called *single-slit diffraction*, although in this case the "slit" is a smoothly tapering circular aperture, producing a beam with a *Gaussian* intensity profile (uniquely, the beam remains Gaussian as it diverges!).



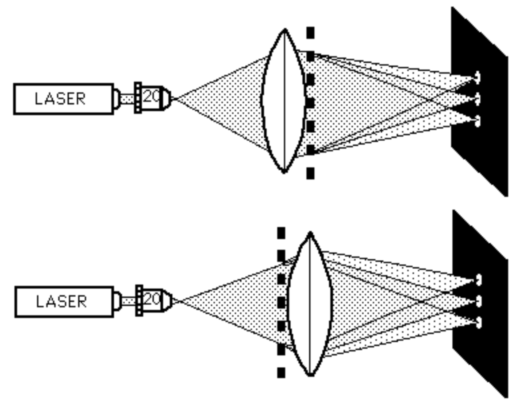
If the beam from the laser passes through a nearby diffraction grating, all of the downstream orders eventually start to expand, and all with the same divergence angle. Thus the downstream spot pattern evolves at a large enough distance into a pattern that no longer changes shape, but only expands uniformly with distance. That pattern is called the *Fraunhofer*, or *far-field* diffraction pattern[6]. Within some distance (the "far-field distance," naturally), the diffraction is characterized as *Fresnel* or *near-field* diffraction. That pattern changes mysteriously as a function of distance, generally requiring mathematical techniques that go beyond the scope of this course. As we approach the grating itself, we expect to see the geometrical shadow of the grating emerge. But not far downstream from there, we also find a negative image of the grating pattern!

The use of lenses:

It is inconvenient to traipse far down a hallway to look at far-field diffraction patterns, so we often use lenses to bring them into focus at much closer distances. First, imagine that a microscope objective has been attached to the laser to produce a point source of diverging spherical waves. A first lens is then placed its own focal distance away from the point source, so as to produce collimated or plane wavefronts. Then comes the grating, which breaks up the incident plane wave into a series of plane waves at various angles. Then comes a second lens, which causes each of those plane waves to curve inward toward a focus one focal length behind the lens. If we put a white cardboard screen there, we will observe the same pattern that would appear in the far field of the grating, except scaled down in the ratio of the focal length to the distance to the far-field pattern.



From a wave- optical point of view, a lens simply multiplies the amplitude of the wavefront by a factor that varies as the square of the distance from the center of the lens (it is a complex phase-only multiplication, which adds a phase that varies as r^2). The diffracting pattern also produces a multiplication of the wavefront, often by an amplitude-only function of x and y, and the second lens is again a phase-only multiplication. Now, because the results of these multiplications are invariant under interchange (or commutation) of the operations, it doesn't matter which happens first, or second, or third. Also, the effect of two lenses being exactly in tandem is the same as for a single lens of twice the thickness variation. So, all three of the sketched optical setups produce the same intensity pattern in the back focal plane of the lens!



Viewing diffraction patterns with the eye:

An important outcome of this discussion is the realization that the naked human eye can also be used to view diffraction patterns. If a point source of light is viewed in sharp focus, then a grating placed just in front of the eye will produce a "far field" diffraction pattern on the retina, which will appear as an array of spots in the same plane as the point light source, surrounding it like a halo. This is very useful for "quick and dirty" checks of diffraction theory with materials at hand.

Styles of diffraction analysis:

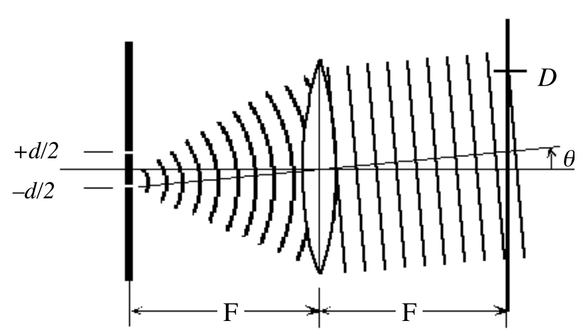
"Every problem in optics becomes easy if you look at it the right way," the old maxim goes, and there are dozens of ways of looking at diffraction and trying to understand its effects. We will take a passing glance at two very different approaches, and then simply accept the mathematical rules that result without many further questions. There are plenty of other books to refer to for other approaches, so be sure to hunt for one that appeals most to your sense of physically-reasonable explanations if you have further questions. They are all simplifications in one sense or another, as even the most basic problem in diffraction (by an opaque single edge) is not yet completely solved!

Graphical analysis:

Here, we will look at the addition of contributions to the far-field intensity pattern as more and more equally-spaced narrow slits are opened, showing that the pattern converges to a series of distinct spots in the limit of many, many slits. First we consider the contribution of a single isolated slit. As that slit narrows to an idealized line source, the transmitted wavefront becomes an idealized cylindrical wave with an amplitude that is equal in all directions (there may be a gradual cosine-theta falloff, which we will ignore). Now the question is: what happens if we open an identical slit parallel to the first and separated by a distance "d"? We consider, for simplicity's sake, the geometry shown here, with the slits one focal length of the lens (called "F") in front of the lens, and the observation plane one focal length behind the lens, so that parallel "rays" can be considered wherever possible.

Two-slit diffraction, one at a time:

Here we show the wave from the lower slit only; the wave from the upper slit will be symmetrical. Let the two slits be at equal distances from the z-axis, at $+d/2$ and $-d/2$, so that their equal contributions will arrive IN phase at x=0 in the back focal plane of the lens. First we open the lower slit by itself. We want to know the location, x=D, at which the wave from the lower slit will be exactly one-half cycle out of phase with (lagging behind) its value at x=0



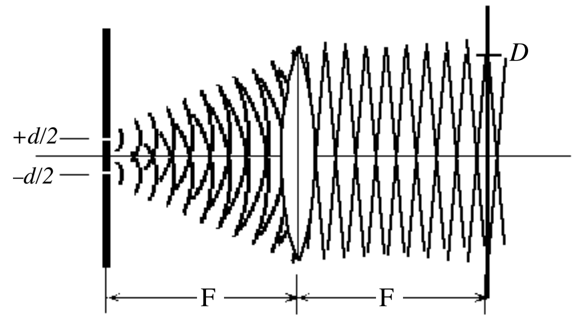
$$D\sin\theta = \frac{\lambda}{2},$$

$$D\sin\left(\tan^{-1}\frac{d/2}{F}\right) = \frac{\lambda}{2}, \qquad (2)$$

$$D \approx \lambda\frac{F}{d}.$$

That is, the lower slit emits a cylindrical wave that the lens transforms into a tilted plane wave with an angle $\theta = \tan^{-1}(d/2F) \approx \sin^{-1}(d/2F)$. This creates a phase increase of $\phi = 2\pi\,(x/\lambda)\sin\theta$ in any plane behind the lens, including the back focal plane at F. The distance D then follows from $\pi = 2\pi\,(D/\lambda)\sin\theta$, giving our result, $D \approx \lambda F/d$.

<u>Two-slit diffraction, two at a time</u>:



Now we open the upper slit too. The two plane waves overlap, and are in-phase along the z-axis (at x=0) because of symmetry. Because the waves are equally but oppositely tilted, they become increasingly out of phase as x increases, until they are so far out of phase that they are back <u>in</u> phase again (the wavefronts cross peak over peak). The height at which the phase difference between the waves from the two slits is 2π is given by $D = \lambda/\sin\theta$. Note that the phase difference (and thus the interference pattern) is independent of the distance along the z-axis, but we will stay at the z=F plane for this discussion.

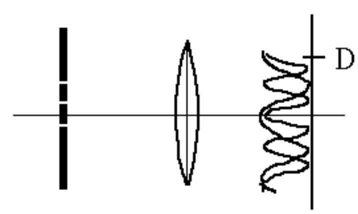
As the observation location moves between 0 and D, the waves from the two slits arrive increasingly OUT of phase, with the phase difference, ϕ_1-ϕ_2, passing π radians or 180° at D/2and continuing on the way to being 2π radians or 360° out of phase, which is back IN phase, at x=D. The phase difference is a linear function of the height in the back focal plane. Thus the intensity of the interference pattern formed by the two equal-intensity waves varies according to our familiar "interference equation,"

$$I_{\text{total}}(x) = 2\,I_0\left(1 + \cos\left(2\pi\frac{x}{D}\right)\right). \qquad (3)$$

As the observation location moves onward from D to 2D, the phase difference further increases from 2π to 4π, and the cosinusoidal fringe pattern continues through equally-spaced maxima and minima until the angles become so large that our paraxial approximations break down.

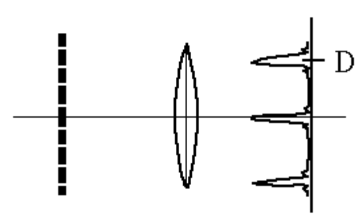
If we look below the axis in the back focal plane, the same phase variation happens, but with the opposite sign. Thus the cosinusoidal pattern extends for many cycles above and below the axis, producing a series of parallel bright "fringes" in the x-y plane that look a lot like furrows in a plowed field.

The pattern we have been talking about is usually called "Young's double-slit fringes," honoring their first explanation by Thomas Young in 1801. Young based his arguments for the wave theory of light on those patterns, leading to the work of Fresnel and Maxwell, and finally overcoming hundreds of years of domination by Newton's particle theory. For us, they are also the building blocks of a theory, this time of holographic imaging! Fortunately, we don't have to contradict any giants in the field; Gabor, Leith, and Denisyuk have all agreed with these ideas about waves and light, especially where lasers are concerned.



<u>Multiple-slit diffraction, *N* at a time</u>:

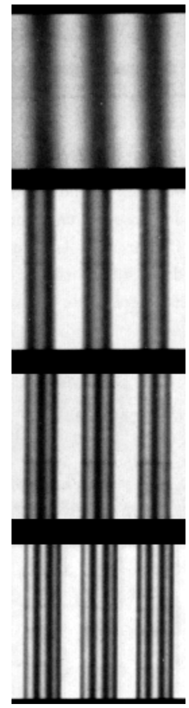
The really interesting part begins when we introduce a third slit, spaced a distance d above the first slit. There are now <u>three</u> interference patterns formed, one for every possible pair of slits, and one of those patterns (between the first and third slits) has *two* intensity maxima between zero and D on the x-axis. And, one of those maxima is centered at x = 0, right on top of the maximum formed by the first two slits.(and on top of the maximum formed by the second and third slits). That is, every pair of adjacent slits produces a pattern that has a maximum at 0, D, 2D, and so forth, but the patterns that are formed by slits further apart have other maxima in between. The sum of all the patterns has principal peaks that are much narrower than for the two-slit pattern, with a weaker "secondary maximum" in between.



$$D = F\tan\theta_{+1}$$

$$\sin\theta_{+1} = \frac{\lambda}{d} = \lambda\cdot f$$

As even more equally-spaced slits are opened above and below the first two slits, further components of the intensity pattern are introduced. We state without proof that the interference pattern formed by even more slits is equal to the interference patterns formed by all possible pairs of the slits, minus a constant equal to the sum of the intensities of the individual slits[7]. The pattern from the furthest-apart slits (let's say that they are $N \cdot d$ apart) has the finest fringes, with N-2 maxima between 0 and D. As more and more slits are opened, and further fringe patterns are added to the overall pattern, the overall intensity pattern converges to a characteristic shape, with principal maxima that are N^2 as high and $1/N$ as wide as for the N=2 case, and with N-2 smaller maxima in between. As N becomes a few hundred, the light concentrates almost entirely in the peaks, one right on axis and others spaced equally up and down on the output plane, separated by the distance D.



Diffraction patterns produced by 2, 3, 4, and 5 slits.

A mathematical proof of the properties of diffraction gratings is certainly in order here, but we will save our effort for a different approach. Various analyses can be found in optics textbooks that emphasize one or another point of view. For the time being, let's explore the general properties we have described so far.

The grating can be considered as an optical component that breaks an incoming plane wave into a set of outgoing plane waves at roughly-equally-spaced angles, a "fan" of beams or rays. The <u>amount</u> of energy in each member of the fan of beams depends on the details of the shape of the grating slits, whether they are hard-edged or soft-edged, or just slow down the wave a bit. But the <u>angle</u> of deflection of each component of the fan depends only on the spacing of the slits, and that is the aspect that we will explore in this chapter.

The grating equation:

The previous discussion showed that an incoming plane perpendicular to the grating results in one plane wave angled up at an angle "theta" (θ) given by $\sin(\theta) = \lambda/d$, and another wave angled downward by the same angle, at $-\theta$. And above and below those are more waves deflected by larger angles, given by $\sin(\theta) = \pm 2\lambda/d$. If the slits have a suitable shape, much larger deflection angles (up to 90°) can be produced. We describe these beams as various "orders" of deflected or diffracted waves, with the "first order" beams being those closest to the straight-through or "zero order" beam. The next set, if those beams exist at all, are the "second order" beams, and so forth, through the third, fourth, fifth, and higher orders.

To begin, let's look at the on-axis grating equation, which describes what happens to a plane wave that comes in perpendicular to the grating. A "fan" of plane waves emerges, consisting of pairs of waves deflected or diffracted through equal but opposite angles. These angles are given by the first or "on-axis" form of the *grating equation:*

$$\sin\theta_{\text{out},m} = m\frac{\lambda}{\text{d}} \ , \qquad (4)$$

where m is the "order number" of a particular beam of interest. The relationship breaks down when the sine of the diffracted angle goes beyond unity (for sufficiently large m, for example), and there is no wave that propagates corresponding to that order. Instead, a so-called "evanescent" wave leaks slightly to the right of the grating, turns around, and re-enters the grating to contribute to the specularly reflected light.

One of the most important observations from the grating equation is that long-wavelength or "red" light is deflected through a larger angle than mid-wavelength or "green" light, which is in turn deflected through a larger angle than short-wavelength or "blue" light. This qualitative fact is important to remember in more complicated situations too, and one mnemonic for it is to remember "the three $\mathbb{R}$s": "Red Rotates Radically!" Landscape paintings occasionally show upside-down rainbows, and holographer's sketches sometimes show upside-down spectra! So it is important to have an easy way to remember which way is "up" in diffraction.

Red
Rotates
Radically!

Spatial frequency:

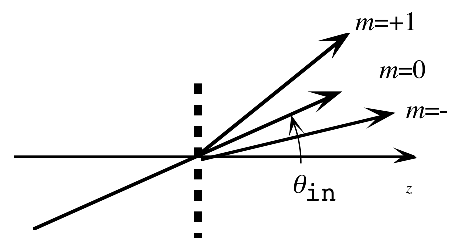
We have been describing diffraction gratings so far in terms of their repeat distance, d, which gives a kind of concreteness to the discussion. However, from now on we will almost always describe them instead in terms of their "spatial frequency" in cycles per millimeter. This allows the analogies with temporal frequency in electrical engineering to become more obvious, and we describe the spatial frequency by the variable $f = 1/d$. We can also include the orientation of the grating by letting f become a two-dimensional vector, $\boldsymbol{f}$, oriented perpendicularly to the grating's grooves, with magnitude $f = |\boldsymbol{f}| = 1/d$. However, the orientation of the grating will always be clear in our discussions (in the x-direction, unless otherwise noted), and we will try to stick with the scalar f wherever possible, in the spirit of "shop math" calculations.

Grating frequency, ray trace (5 orders):

As an example, let's consider a grating of spatial frequency $f = 450$ cycles/mm, or $d = 2.2\ \mu$m. A He-Ne laser beam ($\lambda = 633$ nm) incident at $\theta = 0°$ produces seven output beams: $\theta_0 = 0°$, $\theta_{\pm 1} = \pm 16.5°$, $\theta_{\pm 2} = \pm 34.7°$, $\theta_{\pm 3} = \pm 58.7°$, $\theta_{\pm 4}$ = evanescent.

Off-axis grating equation:

When the incident beam comes in at an angle to the perpendicular, the output fan of beam *roughly* follows it around, staying centered about the continuing beam—but upon closer look the angles between some of the beams increase and others decrease, sometimes significantly. Also, some beams may try to come out at angles beyond 90°, becoming evanescent in the process, and others may "emerge" from the grating on the other side of the beam fan. The details are described by the "off-axis grating equation" below, in which θ_{in} is the angle of the incident beam normal to the grating perpendicular, as shown in the diagram.



$$\sin\theta_{\text{out},m} = m\cdot\lambda\cdot f + \sin\theta_{\text{in}}$$

$$\sin\theta_{\text{out},m} = m\frac{\lambda}{d} + \sin\theta_{\text{in}} \ . \tag{5}$$

Or restated in spatial frequency terms,

$$\sin\theta_{\text{out},m} = m\cdot\lambda\cdot f + \sin\theta_{\text{in}} \ . \tag{6}$$

We will "prove" this relationship in combination with an interesting result in diffraction theory in the next section.

Diffraction by a sinusoidal grating:

Our arguments for physical reasonableness have been built on a model of the grating as a series of narrow slits, with assurances that the angles depend only on the spatial frequency of the slits and not on their width or other properties. Now we will examine this premise for a special kind of slit, one that attenuates the wavefront according to a smoothly varying (specifically, sinusoidal) function. Such a pattern might be produced by exposing a piece of photo film to a two-beam interference pattern, for example.

<u>Transmittance of a grating</u>:

We describe the sinusoidal transmittance pattern of the grating/film as

$$t_{\text{amp}}(x,y) = a + b\cos(2\pi f x) \ , \tag{7}$$

where t_{amp} is the ratio of the amplitude of the electric field of the light wave just after and just before the grating. That is, at every point,

$$E_{\text{out}}(x,y) = t_{\text{amp}}(x,y)\cdot E_{\text{in}}(x,y) \ . \tag{8}$$

Note that in photography we usually consider the transmittance of the *intensity* of the light wave (or its negative logarithm, the *optical density*), which would be the square of the amplitude transmittance we are considering here (a point to consider when attempting to measure transmittances directly). If we assume that the amplitude transmittance must always be between zero and unity (a non-amplifying medium, you might say), there are limitations on the size of a compared to b; for example:

$$\begin{aligned} 0 \le a + b \le 1 \\ 0 \le a - b \le 1 \end{aligned} \ . \tag{9}$$

In a more elaborate view of amplitude transmittance, the phase of the output wavefront can also be manipulated, leading to the description of the amplitude transmittance by complex numbers. For example, if the wave is delayed by one-half of a cycle, the amplitude transmittance is effectively minus one. Further discussion will be deferred to the chapter concerning the diffraction efficiency of such gratings; the physical principles can be illustrated by positive-real amplitude transmittances.

Effect of illumination:

The grating is illuminated by a plane wave inclined at an angle, $\theta_{in,}$ as shown in the sketch. The mathematical description of that wave is, as shown in Chap. 2, Eqns. 2 and 11, where now z=0 at the observation plane:

$$E_{\text{in}}(x,y,t) = \sin\left(2\pi\nu t - \frac{2\pi}{\lambda} x\cdot\sin\theta_{\text{in}}\right) \ . \tag{10}$$

The output wave is then simply given by

$$\begin{aligned} E_{\text{out}}(x,y,t) &= E_{\text{in}}(x,y,t)\cdot t_{\text{amp}}(x,y) \\ &= \sin\left(2\pi\nu t - \frac{2\pi}{\lambda} x\cdot\sin\theta_{\text{in}}\right)\cdot\left(a + b\cos(2\pi f x)\right) . \end{aligned} \tag{11}$$

Application of a familiar trig identity,

$$\sin\alpha\cdot\cos\beta = \frac{1}{2}\left[\sin(\alpha+\beta) + \sin(\alpha-\beta)\right] \ , \tag{12}$$

to the product provides the output wave as the sum of three components:

$$\begin{aligned} E_{\text{out}}(x,y,t) = {} & a\cdot\sin\left(2\pi\nu t - \frac{2\pi}{\lambda} x\cdot\sin\theta_{\text{in}}\right) \\ & + \frac{b}{2}\sin\left(2\pi\nu t - \left(\frac{2\pi}{\lambda} x\cdot\sin\theta_{\text{in}} + 2\pi f x\right)\right) \\ & + \frac{b}{2}\sin\left(2\pi\nu t - \left(\frac{2\pi}{\lambda} x\cdot\sin\theta_{\text{in}} - 2\pi f x\right)\right) \ . \end{aligned} \tag{13}$$

We identify these (by phase-footprint inspection) as three plane waves, which we distinguish by order number m equal to zero, plus one, and minus one:

$$\begin{aligned} E_{\text{out}}(x,y,t) = {} & E_0 \sin\left(2\pi\nu t - \frac{2\pi}{\lambda} x\cdot\sin\theta_0\right) \\ & + E_{+1}\sin\left(2\pi\nu t - \left(\frac{2\pi}{\lambda} x\cdot\sin\theta_{+1}\right)\right) \\ & + E_{-1}\sin\left(2\pi\nu t - \left(\frac{2\pi}{\lambda} x\cdot\sin\theta_{-1}\right)\right) \ . \end{aligned} \tag{14}$$

where the output angles are given by

$$\begin{aligned} \sin\theta_0 &= \sin\theta_{\text{in}},\ \theta_0 = \theta_{\text{in}} \ , \\ \sin\theta_{+1} &= \lambda f + \sin\theta_{\text{in}} \ , \\ \sin\theta_{-1} &= -\lambda f + \sin\theta_{\text{in}} \ . \end{aligned} \tag{15}$$

This motivates/justifies the generalization of the *grating equation* to the off-axis case:

$$\sin\theta_{\text{out},m} = m\lambda f + \sin\theta_{\text{in}} \ . \tag{16}$$

However, a more far-reaching observation is that the amplitudes of the two first orders of diffracted waves are $b^2/4$, and those of all the higher order waves are zero! That is, a purely sinusoidal transmittance grating diffracts light only into the first orders (and the zero or "straight-through" order). More complex gratings (such as slit-like gratings) will diffract light into many orders, but these gratings can be thought of as combinations of mathematically-simpler sinusoidal gratings, each having a frequency that is a multiple of the first-order frequency, and each giving rise to a single pair of diffracted orders. This is equivalent to analyzing the complex transmittance pattern as a Fourier series, which has led to a whole new field of optical theories based on communication theory.

Conclusions:

What happens when light hits a picket fence structure (that is, a “grating”) is truly amazing, and raises all kinds of doubts about our real-world physical intuition. Suppose we let the light trickle through one photon at a time? Which direction does an individual photon take? Such questions are the meat of quantum physics courses, and this is a classical optics course, so shelve them for now and try to learn to accept the rules described by the equations we have proven.

What happens? The beam breaks up into a number of distinct beams that go in very distinct and well-defined directions given by the “grating equation.” That equation involves the “spatial frequency” of the grating (in cycles per millimeter, analogous to the cycles/second of radio and TV signals) and the trigonometric sines of all the angles involved. Luckily, the sine is nearly a linear function of the angle, for small enough angles, and we can get a fairly simple general idea before relegating the calculations to a computer program.

The amount of energy in the various beams is another interesting story that will occupy an entire chapter of its own. Suffice to say that the simplest type of grating to analyze is a sinusoidal variation of transmittance between zero and unity, and it sends all its light into the plus and minus first orders, plus the straight-through zero order (and light absorbed in the grating). None of the possible higher orders are stimulated! Enter Fourier…

Soon it will be time to combine the stories of interference and diffraction to learn about “holography,” after a digression about *how much* light goes into these various beams.

References:

1. Willebrod Snell van Roijen, 1591-1626 (of Leyden, Holland). He succeeded in giving an exact form to the law of refraction, as did Descartes shortly afterward.

2. Huyghens or Huygens (hī'génz, hoi'-), Christiaan, 1629-1695, Dutch mathematician, physicist, and astronomer who discovered Saturn's rings (1655), pioneered the use of the pendulum in clocks (1657), and formulated Huyghens' principle (ca. 1678).

3. Jean Pierre Augstin Fresnel (fré nel', frá-), 1788-1827, French physicist and government civil engineer. First memoir on diffraction submitted on October 15, 1815 (at age 27).

4. Thomas Young, 1773-1829, English physicist, physician, and Egyptologist. He was professor of natural philosophy (1801-03) at the Royal Institution of Great Britain, where he presented the modern physical concept of energy, and was elected (1811) a staff member of St. George's Hospital, London. In 1807 he stated a theory of color vision known as the Young-Helmholtz theory (the 3-primary-color theory) and described the vision defect called *astigmatism*. Young conducted experiments in diffraction and interference (1801) that could only be explained by the wave theory of light, finally overturning Newton’s corpuscular theory. He also established a coefficient of elasticity (Young's modulus) and helped to decipher the Rosetta Stone. He was hounded out of physics by a hostile journal editor, and spent most of his life as a medical doctor.

5. Vasco Ronchi, *The Nature of Light,* Harvard Univ. Press, Cambridge, 1970.

6. Joseph von Fraunhofer, 1787-1826, German physicist.

7. I hope that this is true - you should try to check it out for yourself!