

Introduction for Astronomy 61 ...

These are lecture notes lifted from Physics 4, from the spring of 2013. They're useful here because they provide a short summary of some concepts in optics, which many of you will not have seen, and which will be helpful for Astronomy 61.

This starts with the 16th lecture of Physics 4, and by that time we have explored the basic laws of electromagnetism in some detail; some of the explanations (e.g., how refraction and reflection work microscopically) build on this knowledge, but it's not essential that you understand these.

When we get to diffraction and interference, you *will* want to remember that light is a propagating disturbance in the electromagnetic field, in which the electric field and magnetic field are perpendicular both to each other, and to the direction of propagation. When two waves are superposed, the resulting EM fields are the sums of the fields of the two waves, because the electric and magnetic fields are simply vectors, that obey the laws of vector addition. This superposition principle is the key to understanding interference, and that in turn is the principle used in nearly all modern spectrographs.

In any case, read through these for interest and background.

Lecture 16: Reflection and Refraction, and a Simple Lens

Think of an electromagnetic wave propagating along. At any instant, the crest of the wave forms a 2-dimensional surface that's perpendicular to the direction of propagation. This is called a *wave front*. The wave front moves along at the speed of light.

If you're close in to the source of waves – say, a few wavelengths away from the source – the wave fronts will be appreciably spherical. This is sometimes called the *near field*, and understanding what happens there can require a full solution of Maxwell's equations to match the local conditions. You also need to take careful account of the wave nature of the radiation when the radiation interacts with things that are not too many wavelengths in size; for example, if you're designing an antenna for FM radio (which has a wavelength around 3 m), you'd better be sure to remember that you're trying to capture a wave.

Visible light has a wavelength of about 1/2000 mm. This certainly isn't zero, and we'll soon study *physical optics*, in which we explore *diffraction* and *interference* effects that depend on the wave nature of light.

However, there are many problems for which the 1/2000 mm wavelength of light is unimportant. In that case we can simply use the concept of a *light ray*, which is a line perpendicular to the wavefront, parallel to the direction of travel, and use *geometrical optics* to see what happens.

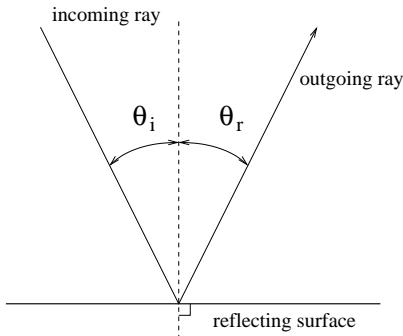
Before proceeding, we need just a couple of simple rules.

First, we need to understand *specular reflection*, which is what happens in a mirror. A light wave comes into the mirror, at some angle θ from the perpendicular. A mirror consists of a layer of good conductor*. The electric fields in the light wave 'shake' the

* A good conductor, that is, for the very high frequencies at which visible light oscillates. Silver and aluminum are excellent conductors at these frequencies, but copper is kind of mediocre.

charges in the conductor, and then – rather miraculously – they immediately re-radiate an exact copy of the original wave, so that it departs along the angle θ , but on the other side of the perpendicular. Geometrically, this is incredibly simple:

$$\text{Angle of incidence } \theta_i = \text{Angle of reflection } \theta_r.$$



Second, we need to understand *refraction*, which is more complicated. Obviously, some materials are transparent – glass and water come to mind. These materials are generally *dielectrics*, and you'll remember that a dielectric will electrically polarize in response to an imposed field. As the wave sweeps through the material, the electric field of the wave leads to a rapidly changing electrical polarization of the material, which *itself* generates a copy of the original wave; this then interacts with the original wave in such a way as to slow down the propagation. Just why this should happen is not intuitive, but it does; *in effect, the light travels slower in a medium than it does in a vacuum*. The *frequency* of the light remains the same (so the photon energy stays the same, too), but the speed goes down. Since $\lambda = v/f$ for wave speed v , it's the *wavelength* that gets shorter.

The amount by which the wave slows down in the medium – and hence shortens in wavelength – is quantified by the

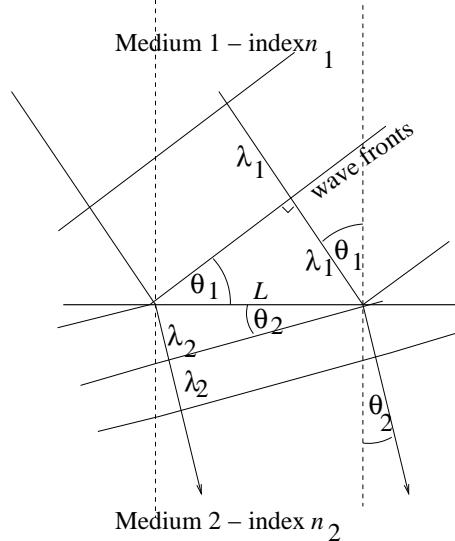
$$\text{index of refraction } n = \frac{c}{v},$$

where v is the speed of the wave in the medium and c is (as usual) the vacuum speed of light. For example, liquid water has $n \approx 4/3$, which means that light propagates at roughly $3c/4$ in water, and wavelengths in water are about $3/4$ of what they would be in a vacuum. Glass has an index of refraction around 1.5, with some variation with the type of glass. Air also has a very small index of refraction, about 1.0003, which varies with density. The index of refraction of the vacuum is obviously exactly 1.

Now, suppose light is propagating in a medium with index n_1 and it encounters a boundary to another medium, index n_2 , and suppose the light hits the interface at an angle θ_1 away from the normal (perpendicular). Let λ_1 and λ_2 be the wavelengths in the two media. The wavefronts will *bend* as they cross the boundary, because they have to remain continuous across the boundary and the change of wavelength forces them to bend. We can use the figure to determine how much they bend. Two rays are seen striking the boundary; I've chosen them so they cross the boundary at the location where two successive wave fronts intersect the boundary. The distance along the boundary between

successive wave fronts is L , as shown, and θ_2 is the outgoing angle. If you look at the triangles formed by L and θ_1 , and by L and θ_2 , you can see that

$$\sin \theta_1 = \frac{\lambda_1}{L} \quad \text{and} \quad \sin \theta_2 = \frac{\lambda_2}{L}.$$



Since $\lambda_1 = \lambda_{\text{vac}}/n_1$, and similarly for λ_2 , this becomes

$$n_1 \sin \theta_1 = \frac{\lambda_{\text{vac}}}{L} = n_2 \sin \theta_2,$$

which gives us our desired result, *Snell's law of refraction*,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Snell's law means that *rays bend toward the normal when they enter a higher-index medium, and away from the normal when they enter a lower-index medium*. As an example, a ray in air ($n_1 = 1$, very nearly) striking a water surface ($n_2 = 4/3$, very nearly) at 30 degrees from the vertical will travel down into the water at about 22 degrees from the vertical:

$$\theta_2 = \sin^{-1} \left(\frac{n_1}{n_2} \sin \theta_1 \right) = \sin^{-1} \left(\frac{3}{4} \times 0.5 \right) \approx 22 \text{ degrees.}$$

Now, notice how I had to take an arcsine(that is, \sin^{-1}) to find θ_2 . The domain of the arcsine is ± 1 , because the sine cannot exceed this range. If $n_1 < n_2$ if the ray is moving into a higher-index medium then θ_2 will always be computable, but what if $n_1 > n_2$ and θ_1 is large? Then $(n_1/n_2) \sin \theta_1$ can exceed 1, so there's no solution for θ_2 !

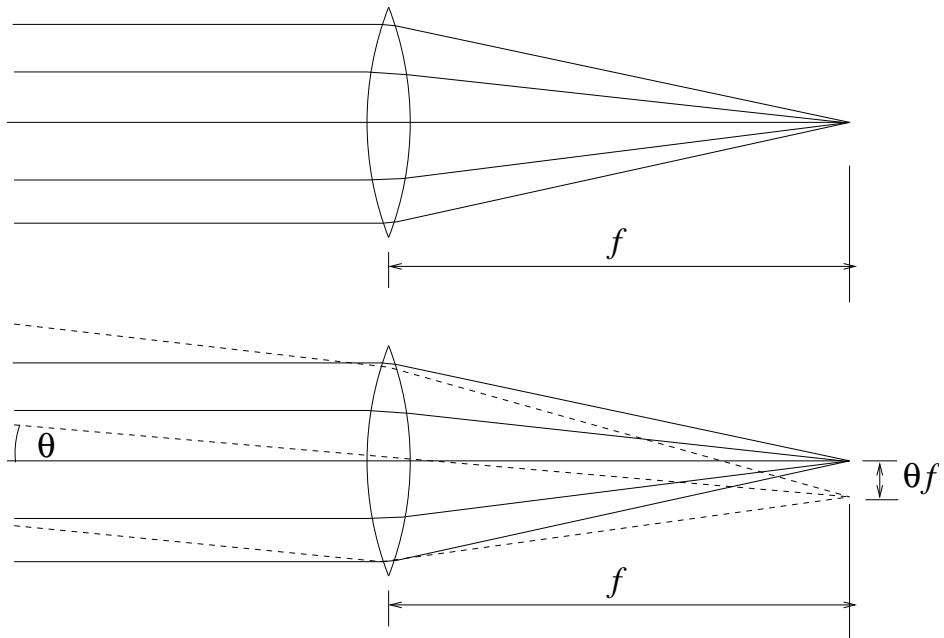
In that case, the ray comes up to the boundary, and undergoes *total internal reflection*. You can see this if you're swimming underwater - if you look at the water surface from below at a grazing angle, it looks like a mirror, because the rays of light coming up from the bottom of the pool are bouncing off the surface.

With these two laws, we can begin to understand a few things.

Suppose we make a glass lens, which is circular and fatter in the middle than it is around the edges (this makes it a *convex lens*). We carefully fashion the surfaces to be smooth curves. Lenses are easiest to fabricate if their surfaces are, mathematically, sections of spheres; this makes a lens that works pretty well in most cases, but not perfectly. One can fuss with lens design almost forever to get just the properties you want modern camera lenses are amazing bits of technology but for our purposes, a simple thin lens' approximation is fine.

Consider rays coming in parallel to the axis of the lens (refer to the diagram on the next page). If the ray is coincident with the axis, it passes straight through the lens, because the surface of the lens is perpendicular to the ray at the center. But if the ray is toward the edge, the surface is not perpendicular to the ray, so the ray bends toward the axis when it enters the lens, and bends toward the axis again when it leaves the lens (convince yourself of this with Snell's law). The lens crosses the axis a distance f behind the lens.

If a ray is not so far toward the edge, then it, too will bend toward the axis, but - because the surface is not so strongly slanted at that point - the ray won't bend as much. If the lens is properly designed and fabricated, the result will be that this second ray also crosses the axis a distance f behind the lens; in fact, all rays parallel to the axis will meet at a point a distance f behind the lens. This distance f is called the lens' *focal length*.



Now let's consider another bundle of rays, parallel to themselves, but not parallel to the lens axis. Let the direction of this new bundle of rays differ from the direction of rays along the axis by an angle θ . Once again, the ray that strikes the center of the lens will go pretty much straight through, and once again the other rays in this bundle (shown as dashed lines) converge a distance f behind the lens. Notice that if θ is small, and measured in radians, the spot where these rays converge is a distance θf from where the rays parallel

to the axis converge. (This is the great convenience of small angles trigonometry becomes very simple, because $\sin \theta \approx \tan \theta \approx \theta$).

Consider a distant scene say, a flagpole that's 1 km away. You align a convex lens so that its axis points directly toward the bottom of the flagpole. Now, where you're standing, *the rays that originate at the bottom of the flagpole are almost exactly parallel*. If you were close up, this wouldn't be true, but its far away, so its very nearly true the lens is only a few centimeters across, and the flagpole is 1 km away. So all those rays that start at the bottom of the flagpole and strike the lens end up passing through the convergent point, right on the axis.

Consider now the top of the flagpole. Where you're standing, those rays are parallel to *each other*, but they're *not* parallel to the rays from the bottom of the flagpole. They're just like the dashed lines in the figure, and as we've seen, they converge at a distance θf from the where the rays from the bottom of the flagpole converge.

Every point on a very distant scene will, like the bottom and the top of the flagpole, produce a bundle of essentially parallel rays again, rays parallel to *themselves*, but *not* to the rays from other points. Each of those bundles will focus to a different point, a distance f behind the lens. This surface on which the bundles focus is called the *focal plane* (or, more generally, the *focal surface*, since it need not be flat). If you were to build a box behind the lens to shut out stray light, and put a white card in the focal plane, you'd see a perfect little image of the distant scene, projected onto the card. This is called a *real image* - an image where real rays converge, the kind where, in principle, you could put a white card and project the image.

This is, of course, a camera. Replace the white card with some kind of light-sensitive recorder film in an old camera, or a digital detector in a modern one. This simply registers how much light falls on each pixel, and you have your image.

It's also the eye. The cornea and lens of the eye form, in effect, a convex lens system, which focuses light on the retina, which in turn has a huge number of light-sensitive cells.

Notice that the image of the top of the flagpole is below the image of the bottom the image appears upside-down! We don't perceive the world as being upside down, because we actually see with our brains. Our eyes just collect the data.

You can see that what the lens is doing, for this infinitely distant case, is creating a mapping (in the mathematical sense) between directions in space - the directions of the ray bundles - and points on the focal surface.

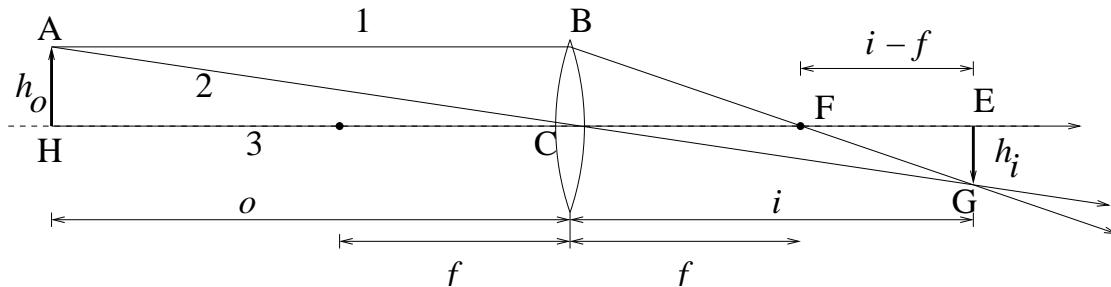
Next, we'll consider what happens when the source is not arbitrarily far away, so that the incoming rays are not parallel.

Lecture 17: Image Formation with Lenses; The Eye

Last time we saw how a convex lens takes parallel bundles of light coming from different directions – each of which originates at a different point in a *distant* scene – and forms an image at the focal plane. This is exactly what a camera lens does, and also what your eye does when you’re looking at things that are far away.

But not everything is far away, so we consider now how to form images of objects that are not infinitely far away *.

Have a look at this illustration. A simple convex lens sits at the middle; its focal length is f , and I’ve drawn a dot on *either* side of the lens to rerepresent the on-axis focal points. (Light can pass through the lens either from left-to-right or right-to-left, so there’s a focal point on either side). Over to the left I’ve drawn an upward-pointing arrow to represent the *object* we wish to image; it is a distance o (oh, not zero) away from the lens. We want to know: (1) How far from the lens *image* form on the right? We’ll call his distance i , as shown. (2) Will the image be rightside-up (‘erect’) or upside-down (‘inverted’)? (3) How large will the image be compared to the original object?



We can figure this out by *ray-tracing*. Before we start this, note that *every* ray that leaves, say, the tip of the arrow, and strikes the lens *anywhere*, will focus at the same point on the right-hand side, and *every* ray that leaves, say, the tail of the arrow and strikes the lens, will focus at some *other* point on the right-hand side. Our job here is to select a few rays that will be useful in telling where these points are and hence answering our questions.

First, let’s draw a ray (labeled ‘1’) that starts at the tip of the arrow at A, and then goes parallel to the axis. The ray is bent by the lens at B, and then passes through the focal point F, because *any* ray parallel to the axis must go through the focal point. Next, we draw ray 2, which goes through the center of the lens C. Any ray going through the middle of the (thin) lens is *not deflected*, so it passes through and intersects ray 1 some distance beyond the focal point (point G). *This defines the location of the image*. Finally, we draw a ray from the bottom of the object arrow through the middle of the lens – this goes straight through C and defines the axis of the lens. The bottom of the image arrow

* Many textbooks develop this using concave mirrors, which behave equivalently to convex lenses.

will have to be on ray 3. Point G already defines the location of the image, so we draw the bottom of the arrow at point E, a distance i from the lens.

Remember that these are only three of an infinity of rays that go through the lens at different places.

Now, let's do some geometry. Because ray 1 starts out parallel to the axis, $BC = AH = h_o$. Also, triangles BFC and EFG are similar, so

$$\frac{h_o}{f} = \frac{h_i}{i-f} \text{ so that } h_i = \frac{h_o(i-f)}{f}, \quad [17.1]$$

Triangles ACH and GCE are *also* similar, so

$$\frac{BC}{CF} = \frac{h_o}{o} = \frac{EG}{EF} = \frac{h_i}{i}.$$

Substituting for h_i from eqn. 17.1 then gives

$$\frac{h_o}{o} = \frac{1}{i} \times \frac{h_o(i-f)}{f}.$$

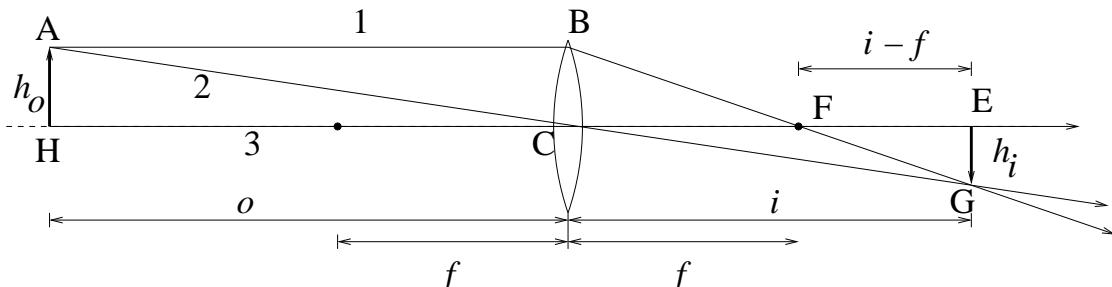
Cancelling h_o and rearranging, we find that

$$\frac{1}{o} = \frac{(i-f)}{if} = \frac{1}{f} - \frac{1}{i}.$$

so that

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i}. \quad (\text{image-object distance}) \quad [17.2]$$

[Diagram repeated from first page for easy reference.]



This makes sense. If the object is very far away, $o \rightarrow \infty$, so that $i = f$ – this is just what we found earlier, namely that the image of a distant scene forms a distance f behind the lens. As we bring the object in closer, the image moves farther from the lens – slowly at first, and then more quickly. When the object at $o = 2f$, then $i = 2f$ also; you can see that 17.2 will be satisfied then, because $1/o + 1/i = 1/(2f) + 1/(2f) = 1/f$. (The diagram, as drawn, is close to this case, but with o a little larger than $2f$, and i a little less.) As you keep moving the object in closer, then i recedes off into the distance to the right, and

as you bring the object up to f on the left, the image recedes off toward infinity on the right. (Note the symmetry of the situation – in the end, it doesn’t matter which side you call ‘ i ’ and which side you call ‘ o ’. The deep reason for this is that the path of a ray is *time-reversible* – the ray could be moving left to right, or equally well be moving right to left.)

The *magnification* of the system will be the size of the image divided by the size of the object; from the similar triangles ACH and ECG, this is just

$$\frac{h_i}{h_0} = -\frac{i}{o},$$

where I’ve prepended a minus sign to indicate that the image is *inverted*.

Even this simple behavior is rich with real-world examples. If you have perfect vision, your relaxed eye should focus correctly on objects at ‘infinity’, i.e., the focal length of your cornea-lens system should be equal to the length of your eyeball, so $i = f$. What if you try to look at something closer? You don’t stretch your eyeball, so i remains the same. Instead, you change the focal length of your lens; rearranging 17.2 we have

$$\frac{1}{i} = \frac{1}{f} - \frac{i}{o} = \text{constant}.$$

If o decreases, that increases the second term, which tends to make the expression too negative, so the first term must increase to compensate, meaning that you have to make f shorter. This is accomplished by muscles around the lens, that squeeze the lens a little, increasing its curvature. The ability to vary the eyes’ focus is called *accommodation*.

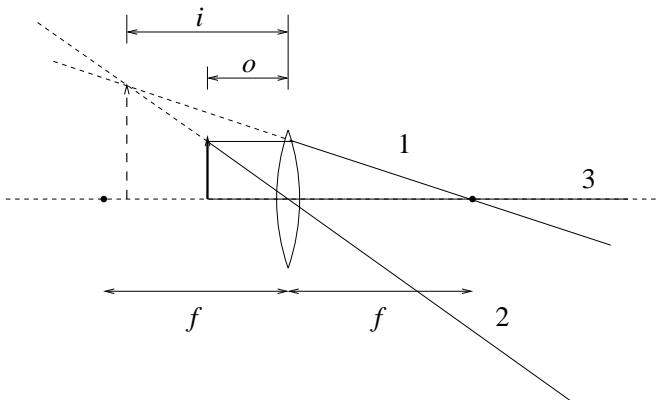
Somewhere between the ages of 40 and 50, the lens of the eye inevitably loses flexibility – there’s no way to prevent it. People who’ve had perfect vision all their lives find that their eyes no longer focus close, so they can’t read. This condition is called *presbyopia*. All that’s needed to compensate is a convex lens, to shorten the effective focal length. Put another way, the convex lens takes the diverging rays from the page and makes them more-or-less parallel again, so that the eye can focus.

Another condition, only slightly less common, is *myopia*, in which f is too short to focus a distant scene. Since f is too short, the image of the distant scene forms in front of the retina. If you trace the rays from any spot in the distant scene, they converge to the image, and then continue to *diverge* behind the image, until they hit the retina and form a circular spot, so a person with myopia sees a distant streetlight as a bright, filled-in circle. Since all the circles from all the spots on a distant scene are superposed, they form a blurry image on the retina. However, as we’ve seen, as the object distance o grows smaller, the image moves backward away from f . If you move o in close enough, i will be equal to the length of the eyeball, and so equation 17.2 is satisfied. People with myopia can therefore see well at distances less than some critical distance, so myopia is commonly known as *nearsightedness*. The correction for nearsightedness is simply to interpose a negative (concave) lens, so that f becomes longer. As we’ll see, negative lenses are minifying, so glasses for people with myopia make their eyes look smaller.

Much less common is a condition called *hyperopia*, in which the lens is too weak, so f is too long. Then the sharp image forms *behind* the retina. This is devastating, because

now there is *no* distance for which the eye can form an image; the eye is focused ‘beyond infinity’. This can be corrected by wearing a convex lens, or positive lens, which makes the eyes look bigger. Hyperopia is sometimes called *farsightedness*. A mildly hyperopic patient with good accomodation would be able to compensate enough to focus at infinity, but not at short working distances.

Virtual images. What if we move our object *inside* the focal length f ? To analyze this, we again choose ray 1 to be from the tip, parallel to the axis; it again goes through the focus point on the right. Ray 2 goes straight through the middle of the lens – but now, it *never intersects* ray 1, so there is no *real* image formed. Ray 3 simply goes along the axis.



The actual light in ray 1 does not intercept the actual light in ray 2. However, ray 1 and ray 2 are not parallel, so the straight lines they define on the right *do* intercept on the left side of the lens. Here, they form what is called a *virtual image*, shown by the dashed arrow in the figure. If you look into the lens from the right, you’ll see this virtual image – it looks like a bigger arrow, farther away than the real one.

This is the principle behind a *magnifier*. Suppose you want to look at something little. You hold it close to your eye – and it gets all blurry, because your eye cannot accomodate that close. Looking through a positive lens will show you a virtual image that looks enlarged, and perhaps more importantly, looks farther away, where your eye can actually focus.

As it turns out, equation 17.2 still holds, *provided* we get our signs right. Suppose we make $o = f/2$. Then eqn. 17.2 becomes

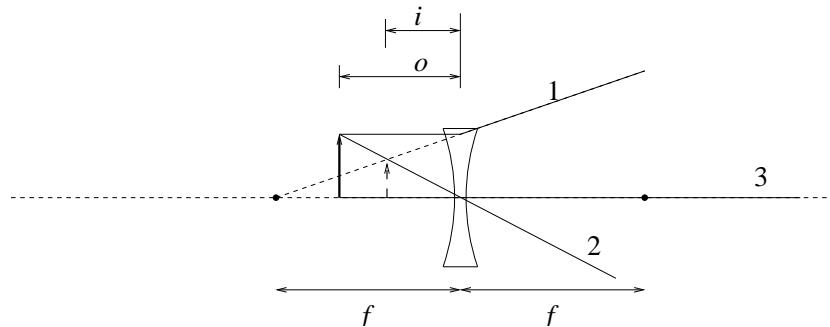
$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i} = \frac{2}{f} + \frac{1}{i},$$

from which we conclude that $i = -f$. All we need to do is interpret i as *negative* when it’s on the same side of the *lens* as o . If the object is held right up to the lens, the lens makes essentially no difference, and the virtual image becomes pretty much the same as the object; if you pull the object back toward f , the virtual image recedes toward infinity. The magnification is again $-i/o$, but since i is negative, the magnification is positive, indicating that the image is erect.

Diverging (negative) lenses. A *concave lens* – which is thinner in the middle than at the edges – *diverges* light, as one can readily see by considering parallel rays striking the

lens. If parallel light comes in from the left, it will keep going right, diverging as if it had come from a point a distance f to the *left* of the lens. We can keep the same formalism if we assign a *negative* focal length to a diverging lens.

A negative lens cannot form a *real* image – the images are always virtual. The figure illustrates this. As usual, ray 1 starts at the arrowhead, runs parallel to the axis, but now when it strikes the lens, it bends *away* from the axis, so that the line it takes traces back to f on the left side of the lens. Ray 2 goes straight through the axis. As you can see, you're left with a *virtual* image that is *minified and erect*.



With a positive (convex) lens, you can, in principle, focus an image of the sun to a fairly small spot, and conceivably start a fire. You won't be able to do this with a negative (concave) lens.

I've drawn the concave lens as being concave on both sides. Many diverging lenses are convex on one side, but more sharply concave on the other; the negative lenses in eyeglasses are ground this way, for reasons that are a bit beyond our scope here. If you're myopic, the front surface of your glasses will usually be convex, but extra concavity in the back surface more than makes up for it.

Diopters. If a lens creates a strong convergence or divergence of light, it has a short focal length f ; a weak lens will have a long focal length, which goes to infinity in the limit of a flat sheet of glass. The ‘strength’ of a lens is therefore inversely related to the focal length. This motivates the *diopter*, which is a measure of lens strength defined as

$$\text{strength in diopters} = \frac{1}{f \text{ in meters}}.$$

For example, a +10-diopter lens has a focal length of $1/10 \text{ m} = 10 \text{ cm}$, and a -5 diopter lens would have a focal length of -20 cm , the negative sign indicating a *diverging* lens.

Diopters are used in lens prescriptions. If you're myopic, here's an incredibly simple test you can use to find your rough prescription. Without any corrective lenses, measure the farthest distance at which you can get sharp focus, perhaps by looking at your finger and bringing it in until it's sharp. Convert this distance to meters (in my case, it's about 0.17 m), and invert it to find the diopters (in my case, about 6 diopters, or actually minus 6 since I'm myopic – that's actually pretty close to my prescription.) Basically, you've found the first point for which eqn. 17.2 holds for your unaided eye, and from this you can find how much extra negative f you need to add to your eye so that you get convergence for $o = \infty$.

Eyes can also be *astigmatic*. This means that the cornea and lens are not symmetric around the optical axis. To correct this, one makes a *cylindrical* lens – it forms a lens along one dimension, but is uniform along the other. A convex cylindrical lens would focus parallel light to a *line* rather than a point. In reality, glasses are seldom pure cylinders, but have a cylindrical component. You can tell if a pair of glasses include an astigmatism component by holding them out, looking at a distant scene through the lenses, and rolling them clockwise or counterclockwise. If the scene stretches like rubber, then the wearer is astigmatic.

Prescriptions for corrective lenses have lines labeled ‘O.D.’ (Latin for right eye) and ‘O.S.’ (for left eye); each line has three numbers. The first is the *sphere*, the converging power of the lens in diopters. For most people this is negative. The second is the *cylinder*, also in diopters, and the third is the *angle* of the long axis of the cylinder. If you’re older and suffer from presbyopia, you may also have a lower section of the glasses in which the sphere is not as strong – it’s reduced by somewhat over a diopter – so that you can still see close by. Glasses that are ground that way are called *bifocals*.

Mirrors behave much the same as lenses do, but (in some ways) in reverse.

Consider a concave, spherical* mirror, with radius of curvature R . The mirror might be made of glass, but the front surface is coated with a thin layer of silver or aluminum, so the glass does not refract – it simply supports a reflective layer. If we put a light source at the center of curvature of the sphere, every ray that strikes the surface will reflect straight back where it came from; if we instead illuminate the mirror with parallel light, it will focus pretty well on a spot that is *half* as far from the mirror as R . Therefore, $f = R/2$.

The *concave* mirror behaves very much like the *convex* lens. Objects out beyond f form *real, inverted* images, but they’re on the *same side* of the mirror as the object. Objects *inside* f form *virtual, magnified* images that appear on the far side of the mirror. A make-up mirror works exactly that way; you see a magnified, virtual image of yourself. (Indeed, you can think of a flat mirror as a limiting case of a spherical mirror, with $f = \infty$; in that case, your image i appears to be as far behind the mirror as you are in front of it.)

A *convex* mirror, on the other hand, is much the same as a concave *lens*. All the images are virtual – they appear to be ‘inside’ the mirror. The images are also *minimized*. These are often used when a large field of view is needed, as in the right wing mirrors on a car, for which “objects in mirror are closer than they appear”.

On a personal note, I have to say I’m very fond of large concave mirrors. All big telescopes today are made with mirrors, rather than lenses, for many reasons. The telescope Dartmouth shares in Arizona has a concave mirror 2.4 meters in diameter, which weighs nearly two tons; we also have a share of a telescope in South Africa for which the mirror is *10 meters* in diameter; it’s actually made of 91 segments which are carefully aligned to

* A spherical mirror is one in which the surface is a section of a sphere. It usually looks like a flat plate with a slight curvature on one side. Spherical mirrors do not focus perfectly except for rays that are close to the axis – they suffer from *spherical aberration*. A parabolic mirror focuses better, but over a narrow field of view. This is a rich and bottomless subject – for simplicity’s sake, we’ll just think about spherical mirrors.

form a single spherical mirror. The spherical aberration in this mirror is eliminated using a sophisticated 4-mirror spherical aberration corrector. But this is all fodder for another course.

Lecture 18: Optical Instruments; Diffraction and Interference

Last time we saw how lenses and mirrors can form real and virtual images of objects at finite distances; before that, we considered objects that are arbitrarily far away.

Let's extend this somewhat and explore the principles behind some optical instruments.

We already touched on the *magnifier*, which is a positive lens. Suppose you have a small object that you want to examine in detail. If you hold it close to your eye, to make it look big, it's a blur – your eye can't focus that close. Put another way, the closeness of the object makes its rays diverge rapidly, so your eye can't re-converge them to form an image on the retina. If you hold a magnifier close to the object, it takes the diverging rays from the object and makes them diverge more slowly. Tracing the rays back through the lens, you find a magnified *virtual* image, which appears far enough away from your eye to focus. You can see plenty of magnified detail in this image.

The magnifier features as a component in two devices, the *telescope* and the *microscope*, arguably the most important scientific instruments ever devised.

To see things far away, you of course want a *telescope*. To do this, you start with a positive lens with a fairly long focus, called the *objective* lens. If you point such a lens at an object far away, it will produce a real, inverted image of the object; if the object is at infinity, the real, inverted image coincides with the focal plane a distance f behind the lens.

You could put a detector in the focal plane and capture this image, and indeed that's what modern astronomers do nearly all the time. However, you might want to look at the image instead. You can do this, because if you don't put anything in the focal plane, the rays simply pass through the focal plane and continue diverging. From behind the focal plane, it looks like there's a little copy of the distant scene, hanging there in space. The problem is that this is usually inconveniently small; if you try to look at this image with your unaided eye, it's too small to see well, so the point of having a telescope is lost. To see the image in detail, simply put a magnifier (called an *eyepiece*) behind the focal plane. Looking through the magnifier at the real image is exactly like looking at a real object located at the focal plane. This simple telescope creates an inverted (upside-down) image – which, if you're looking at a terrestrial scene, is a bit disconcerting. To correct this, binoculars pass the light through prisms, in which the light undergoes total internal reflection; the prisms are arranged so that, after an even number of bounces, the image has been flipped rightside-up.

As I noted at the end of the previous write-up, all large astronomical telescopes use concave mirrors to focus the light; these are precisely equivalent optically to large lenses, but are much more practical to make.

A *microscope* is somewhat similar to a telescope, but uses a very short focal-length objective lens. The subject is put just a little farther than f from the objective, which leads to a highly-magnified real image on the other side, much farther from the lens, many times f away. Again, an eyepiece (magnifier) is used to view this real image.

Dispersion and Chromatic Aberration. Camera lenses, eyepieces and the like are complicated systems, often with many lenses in them. The complications exist in part to provide a wide field of view, and to correct for the fact that spherical surfaces don't quite focus correctly. However, another reason they're so complicated is to correct for *chromatic aberration* – the way in which simple lenses don't handle all colors the same way.

Chromatic aberration exists because glass, like most other substances, is *dispersive* – the index of refraction n depends somewhat on the wavelength of light. In most substances, n is greatest for violet light, and least for red light.

Dispersion accounts for the familiar colors produced by sunlight passing through a prism, or by sunlight passing through the spherical water droplets that make up a rainbow. The violet light gets refracted the most, and the red light the least, and the rest is spread out in between.

In a simple lens, just the same thing happens, so the focal length is longest for the red light and shortest for the violet. An image taken with a simple lens will only be in focus for a narrow range of wavelengths. One way to correct this problem is to use two different kinds of glass; one has a higher index, but greater dispersion; the other has a lower index, but less dispersion. By combining a stronger positive lens of the lower-dispersion glass with a weaker negative lens of the higher-dispersion glass, you can get a lens that focuses all the light at the same place, very nearly. This is called an *achromat*. By using variations on this scheme, modern camera lenses cancel chromatic aberration very accurately. The eye suffers chromatic aberration, but it's too small to notice.

Physical optics. There are many phenomena for which the ray picture of light isn't sufficient. To understand these phenomena, you need to remember that light is a *wave*.

The two aspects we'll be most concerned with are *interference* and *diffraction*.

Interference occurs because the waves are made of vector fields (electric and magnetic), so the net wave field in any point is the vector sum of the fields of individual waves arriving at that point; this is simply the principle of *superposition*. If one wave points up, and the other points down, they tend to cancel, and *destructive interference* occurs; if both point up at the same time, then they reinforce each other, and you get *constructive interference*.

Suppose you take light from a single source, and make two copies of it. Take one copy and delay it slightly, and then mix it with the other copy. If you delay it by a half-wavelength, then one copy 'zigs' as the other copy 'zags', and the interference is destructive. If you delay it by a full wavelength, then you get constructive interference. It's really that simple.

Here's an example of how this can be done. Take two glass microscope slides. Support them so that they are separated by an air gap a few microns thick. Illuminate them with the yellow-orange light from a sodium lamp, which is in a very narrow range of wavelengths around $590 \text{ nm} = 0.590 \mu\text{m}$. Note that the gap is just a few wavelengths thick.

When light passes from air to glass or from glass to air, some of it is reflected back. The light reflected from the top of the gap and the light reflected from the bottom of the gap are two 'copies' of the same light. Because the light reflected from the lower surface has a

little farther to travel, it's delayed slightly compared to the light reflected from the upper surface. If the delay amounts to an integer number of wavelengths (1, 2, 3 ...), then the interference is constructive, and the reflected light is brighter; if the delay is a half-integer ($1/2, 3/2, 5/2 \dots$) then the interference is destructive, and the reflected light is fainter. If the signals from the top and bottom are equal strength, the cancellation is perfect.

Suppose the two slides are perfectly flat, and you make the gap thicker on one end than the other – perhaps by putting a spacer on one end of the slides, and letting them touch on the other. Then, as you go from end that touches to the end where the spacer is, the gap will gradually open up, alternating from integer to half-integer spacing (and all values in between). If you look at the slides, you'll see bright bands where there's constructive interference, and dark bands where there's destructive interference. These are called *interference fringes*. This characteristic light-dark-light *fringe pattern* is ubiquitous in physical optics.

In principle, the reflection from the top layer of the top slide should interfere with the signal from the bottom layer of the slide, and so on, but in practice this tends not to be important, because the slides are so many wavelengths thick that the signal ‘loses count’ over the relatively long delay. This occurs because ordinary light, even in a spectral line, is not especially *coherent*. To get a rough idea of what ‘coherence’ means, consider a perfect sinusoid, continued forever. If you're told when the peak of one wave occurs, you can predict exactly when any peak occurs, arbitrarily far into the future. But if you start randomly altering the parameters of the sinusoid, you won't be able to do this. Coherence is a measure of how much randomness is present, and in particular the length of time over which one can predict the location of the peaks. Natural light is highly incoherent; light from a sodium lamp, having a narrow range of frequencies, is more coherent. Laser light has the special property of being *extremely* coherent, because it contains a very large number of exact copies of the same photon; we'll explain how that works a bit later.

Because natural light tends to have low coherence, you only see interference effects in natural light when the delays are fairly short. Furthermore, in order to see *any* interference, the two signals being mixed have to come from the *same* source. Two different light bulbs produce signals that have nothing to do with each other, so they don't create an interference pattern.

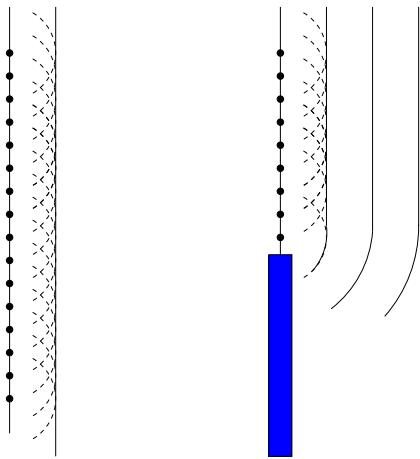
Oil slicks. You'll recall that different wavelengths of visible light appear as different colors. White light is a mixture of all the different wavelengths.

If a drop of oil falls on water, it will quickly spread out to be very thin, and you'll get reflections from the air-oil interface and from the oil-water interface just a wavelength or two below it. You get constructive interference when the extra travel corresponds to an integer number of wavelengths; therefore, when (say) green light interferes constructively, other wavelengths will not be as bright. This accounts for the rather pale color fringes you see in an oil slick.

Diffraction occurs whenever a wave passes an obstacle of some kind. Rather than proceeding in a straight line as the ray picture would suggest, the wave ‘wraps’ around the obstacle, and it may even change direction.

To understand diffraction, it's useful to introduce an aspect of wave behavior called *Huygens' principle*, named after a 17th-century Dutch mathematician. Consider a flat

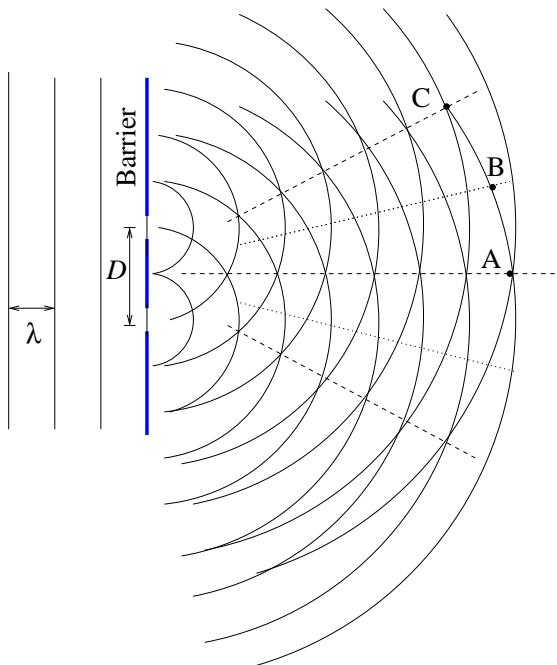
wavefront. Now, think of every point on that wavefront emitting a spherical wave. Propagate those waves forward for a bit, and you'll find that the waves all cancel, *except* in the forward direction – so the wave simply propagates as a flat wavefront. But, if you place an obstacle in the path, some of the spherical wavefronts are cut out, leading to incomplete cancellation among the other spherical wavefronts. The net effect is that the wavefront is no longer flat, but has been bent.



The *double-slit experiment* involves both diffraction and interference effect. You have a barrier, and in it you cut two narrow slits a short distance D apart. Illuminate these slits with light from the *same source*, so that the light entering both slits has the same pattern of zigs and zags. To make it simple, suppose this light has a single wavelength λ . As light passes through a slit, it diffracts, because the slit is narrow. Beyond the slit, the waves spread out in two circular patterns – one pattern centered on one slit, and the other pattern centered on the other.

Consider a point beyond the slits, where waves from both slits are arriving. If the point is equidistant from both slits (like point A in the diagram), then the waves from the two slits arrive in synch, so the interference is constructive. But if you move to a position where the distance to one slit differs by a half-wavelength from the distance to the other (point B), the waves from the two slits are out of synch by a half-cycle, and you get destructive interference. Still farther from the center line (point C), the difference becomes a whole cycle, and you get constructive interference again – this is sometimes called the *first order* peak. The pattern continues outward. If you put a white card off to the right, you will see interference fringes – bright where the interference is constructive, and dark where it's

destructive.



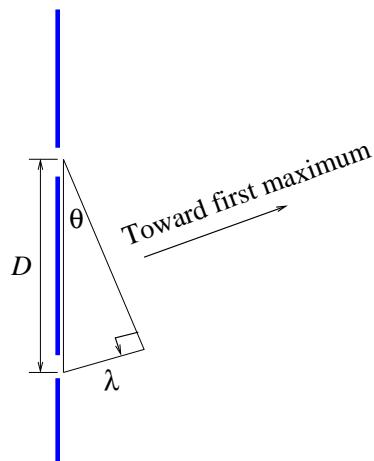
It's not too hard to figure out the angle at which the first maximum leaves the slit. At the first maximum, the distance to the far slit is exactly one wavelength greater than the distance to the near slit. The construction in the figure shows that this occurs if

$$\sin \theta = \frac{\lambda}{D}.$$

Another maximum will occur where the extra distance is 2λ , and so on (and of course there's a maximum in the middle, where the extra distance is zero), so the maxima occur at

$$\sin \theta = n\lambda/D, \text{ where } n = 0, 1, 2, \dots$$

It's clear from this that there's a limit on n , since you must have $n\lambda/D \leq 1$ in order to have it be equal to a sine.



Lecture 19: Diffraction and Interference Devices; Polarization; Starting Relativity

Last time we looked at the two-slit experiment. This was first done by Thomas Young in 1801, and it's historically very important because it proved for the first time that light is a wave. Before that, Newton's idea that light consists of 'corpuscles' had been ascendant. Ironically, almost exactly one century later, Max Planck proposed the idea of the photon, which was fleshed out nicely by Einstein a few years later; we now view light as *both* a wave *and* a particle. This is not exactly intuitive, but it appears to be the way things work. More later!

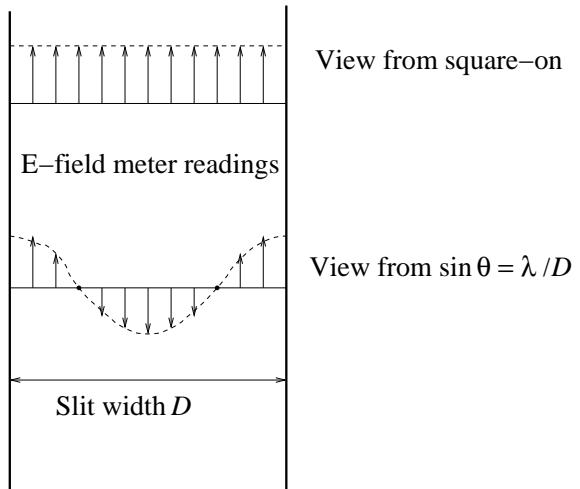
Single-slit diffraction. If you pass light through a narrow slit, it diffracts. What's less obvious is that it also *interferes* with itself.

Here's a *gedankenexperiment** to help you see why this happens. Imagine taking a single slit, and stringing across the slit a row of tiny 'E-field meters'. Each meter reads the instantaneous electric field and indicates the strength and direction of the field by displaying a tiny arrow on a tiny little screen.

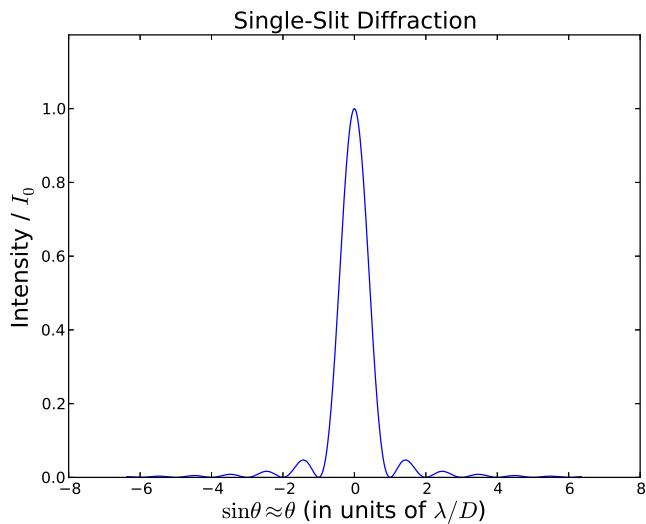
For what follows, it's important to note that if you're looking at this from somewhere beyond the slit, the size of the wave field arriving at your location at any instant is proportional to the sum of all the readings on all the meters. That's basically Huygens' principle – each point on the slit radiates its own wave, and these waves interfere with each other at points beyond the slit.

Now, suppose you view the slit square-on. Then the distance between you and all the meters is the same, so you see the E-fields all varying up and down in unison. These add to give a strong wave at your location. But as you move away from the center line, the distance to one side grows larger and the other grows smaller, so the meters get out of step, simply because they're no longer all the same distance from you, so the different meters suffer different time delays. Once you've moved far enough so that the near side of the slit is closer than the far side by exactly one wave, your view of the meters at any instant covers an entire cycle of a sine wave. The sum of an entire cycle of a sine wave is zero, so there is a *null* at that point. If you go farther, so that the difference in path lengths to the two sides of the slit is $3\lambda/2$, then $2/3$ of the slit's meters form a whole cycle of a sine and cancel each other, but the other $1/3$ of the slit is half of a cycle of the sine, and it *doesn't* cancel. You see a maximum, but because there's cancellation over most of the slit, it's lower than the central maximum.

* This is German for 'thought experiment'. In this you imagine a hypothetical situation and use it to infer novel and sometimes surprising consequences of physical laws. Einstein was the all-time master of the *gedankenexperiment*, so we'll soon have cause to use the term quite a bit.



When you work this out mathematically, it turns out that the function that describes the diffraction pattern is proportional to $(\sin(\theta)/\theta)^2$, which looks like the graph. You get a bright central maximum, and then a minimum where $\sin \theta \approx \theta = \lambda/D$, that is, where there's a delay of one wavelength from one cycle of the slit to the other.



Interestingly, you get some diffraction passing a wave through any aperture, even if the aperture is many wavelengths across. The Hubble Space Telescope has a mirror 2.4 m in diameter, which is about 5 million wavelengths of visible light. The diffraction of the light passing through this aperture amounts to about a 5-millionth of a radian, and sets the limit of how sharp the images can be.

A *diffraction grating* actually employs both diffraction and interference. To see the principle, imagine that you had a very large number of slits (instead of just two), with precisely equal spacing D ; this is called a *grating*. Illuminate the grating with a monochromatic (i.e., all the same wavelength) plane wave, and look at it from an angle θ away

from the perpendicular, a good distance from the grating. Each slit diffracts the light. Suppose the wavelength is *exactly* right to get constructive interference in first order, so that $\sin \theta = \lambda/D$. Then *all* the slits will contribute to the light. But if λ is just just *little bit off*, then you'll see the slits at one side of the grating significantly out of phase with the rest of the slits, and you'll get destructive interference. Consequently, the light at this one wavelength comes out at a *very* sharply defined angle. If you add in other wavelengths, each of those comes out at some other, sharply defined angle, and you form a spectrum, much like the spectrum formed by a prism.

A few details can flesh this out. I've assumed that the grating is illuminated square-on by the incoming waves, but one can relax this condition and the grating still forms a spectrum – the math just gets slightly more complicated. To diffract optical light, gratings need to have a very close spacing, comparable to a couple of wavelengths. It's obviously impractical to make literal slits this fine, but one can fabricate plastic sheeting that has a nice pattern of fine grooves that work nicely. Fine gratings are often made by etching grooves in glass and then coating with a reflective substance – the many tiny mirror segments form a fine ‘reflection grating’.

Even a compact disk makes a nice grating – the grooves burned into the reflective layer are nicely spaced, and they reflect different wavelengths at different angles. This is why CDs look so colorful.

Gratings can be made that have tremendous dispersive power – they can be used in instruments to discriminate wavelengths very, very finely, so that wavelengths that differ by one part in, say, 50000 can be spit easily. Gratings are at the heart of nearly all modern spectrometers, or (as we call them in astronomy) spectrographs.

Polarization. In most natural light, the electric field vector shakes around in random directions perpendicular to the line of sight.

In some circumstances, though, the field vector will oscillate in a preferred direction (though, of course, always perpendicular to the direction of propagation). This condition is called *linear polarization*. (Electric field vectors have direct effects on charges, so the E-field is the field chosen to represent this; the B-field is still along for the ride, of course).

Linear polarization arises naturally in several circumstances. For example, when light passes an air molecule, the charges in the molecules oscillate in response to the electric field, and of course this oscillation is perpendicular to the direction of propagation. Sometimes the oscillating charge will *scatter* the ambient light – send it off in another direction. If the light is scattered by 90 degrees from its original direction, it will be polarized, in the plane perpendicular to the original direction – which is the same plane in which the charges in the molecule were oscillating. For this reason, the clear blue sky 90 degrees from the sun is highly polarized; all you're seeing there is light that has scattered off air molecules, becoming polarized in the process.

The American inventor and entrepreneur Edwin Land invented a cheap way of making plastic *polarizers*, which pass only one plane of linear polarization and reject the rest. Polaroid sunglasses are made of this stuff – by tilting your head back and forth with polaroid sunglasses, you can easily see the polarization of the world around you, including the polarization in a clear sky. The motivation for using polaroids in sunglasses is that a

lot of the scattered light from the ground tends to be horizontally polarized, so you can reduce glare by wearing glasses that pass only vertical polarization. One reason that a sunlit scene tends to have some polarization is that light is polarized when it is reflected from a refractive-index jump (e.g., an air-glass or air-water surface). If the surface is horizontal, then so is the polarization of the reflected light. (It's possible to understand polarization on reflection using a rather tricky argument, which I'll skip in the interest of time).

Linear polarization is easiest to visualize, but it's also possible to have *circular polarization*, in which the electric field vector describes a corkscrew, with a full rotation per cycle. There are two possible helicities for this which are called *left circularly polarized* and *right circularly polarized*.

Many aspects of polarization are interesting and useful in other parts of science. Polarization and organic chemistry interact in interesting ways, with different stereoisomers of some compounds rotating the plane of polarization in different directions. Some transparent materials rotate the plane of polarization when they're under stress, which makes it a great tool for mechanical engineers.

Finally, it's worth noting that the amount of polarization is continuously variable. If the field shakes around randomly, the light is unpolarized, but if there's just a partial tendency to shake in one plane, then it's *partially polarized*; if it all shakes one way, then it's *completely polarized*.