OPTICS

Bold assumptions, big results

An exercise in starting with a few simple assumptions to see how far we can get just by thinking and not getting lost with fancy mathematics.

Introduction

A little bit about where I'm coming from. I spent about twenty some years in academia at the University of California Santa Barbara. In the late 60s my teaching was to sophomore engineers and then in the 70's I started a master's degree program in scientific instrumentation in the physics department. That program started out in the standard away of doing fixed labs and lectures with homework. But it then evolved, quite by accident, into a project-oriented course where the students would invent and build instrumentation for researchers on campus, in all academic departments and sometimes with companies off campus. It became very clear that students working on projects where they had to think their way to the end, and which there could be many solutions, would learn a lot faster than students doing labs and lectures where you were just taking instructions and trying to remember them, something many people call "learning". No, it is not learning, it is teaching. On top of this surprise there was also another surprise. We started accepting students into the program from outside engineering and physics, for instance we had a religious studies major and a biologist (who is now a cardiologist) in the program and found that as far as solving problems goes, it didn't seem to make much difference what their undergraduate education was, as long as they were smart and knew some math. Project oriented education is messy though. It even gets hard to grade people who are not doing the same thing and who start with different backgrounds. How do you give a test? Anyway, the lectures went away and it became completely project learning where all the thinking and learning was done. But not learning about facts, learning how to think your way through a problem and get to the end with some reasonable outcome. Learning how to function. I must admit that as part of my history I went to a vocational high school in Iowa where my major was machine shop, where everything was learning by doing.

I quit academia and started a company that made scanning probe microscopes. I started the company with the best student that I had in all the years of the scientific instrumentation program (Hey, I needed lots of help). From my experience in the instrumentation program, we would hire employees who had talent, not experience, unless they came in the same package. It was fun because we quickly became the world's leader in our field.   We learned as we went along. I spoke with a friend from long ago, actually a high school friend, so it was really long ago. He said "wow you must have read a lot about how to do business". Then I realized that he believes that everything you know is what you read or what someone told you, whereas what I believe is that you can learn more by doing something than by reading about it.

After I had been out of academia for a while I had strange idea to go back and teach a course at UC Santa Barbara to involve thinking rather than reading and memorizing. This course went by different names, Learning by thinking, Thinking out-of-the-box, and How do things work? The course went okay for a few years then slowly went downhill. Two things happened, I became probably more senile and the students didn't seem to care about thinking but only about getting a grade. I finally gave up, believing that academia was training students as you would train seals, that is, here's a trick, repeat the treat, here's your fish, but not how do you think up new tricks? Trained seals don't think up new tricks. Part of my "teaching" was to pick subjects where you could start with a couple of simple assumptions and then work your way through the entire (almost) subject. My two favorites were optics and relativity. I thought I would put my thinking on the subjects down in writing. People who want to do it more formally would probably think this is nonsense but some people may enjoy it and find it an easy way to learn the subject. One thing about teaching is that I not only learned a lot both about teaching, but also about the subject matter, even more when I went to put it down on paper.

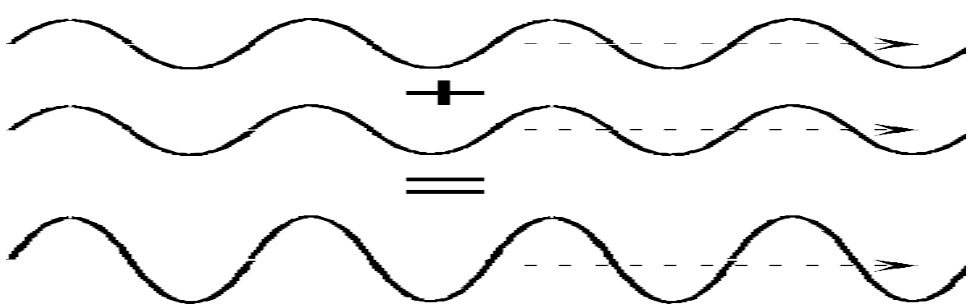
OPTICS

The assumptions:

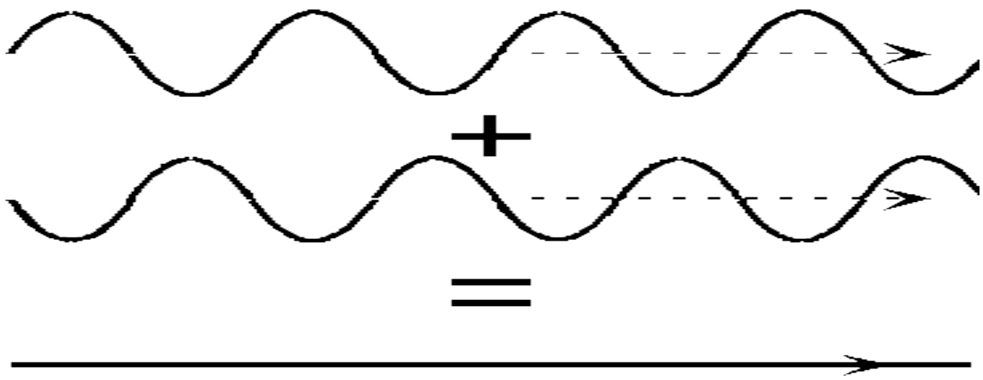
1. Light is a wave (not very bold)
2. Light can take any and all paths from A to B.

I am going to blame this second assumption on Richard Feynman, which is where I think I ran across it (summing over paths). Maybe, though, it was Huygens several centuries ago. We will not bother with the fact that light is an electromagnetic wave but just a wave. When I studied optics, it was divided into two sections, geometrical optics (light is a ray) and physical optics (light is a wave). In our discussions, light is a wave.

Now people usually think light travels in a straight line and so what is this business about taking any path? But light is a wave and, as you know, two waves out of phase can cancel each other, so it may turn out that a lot of these paths cancel each other.

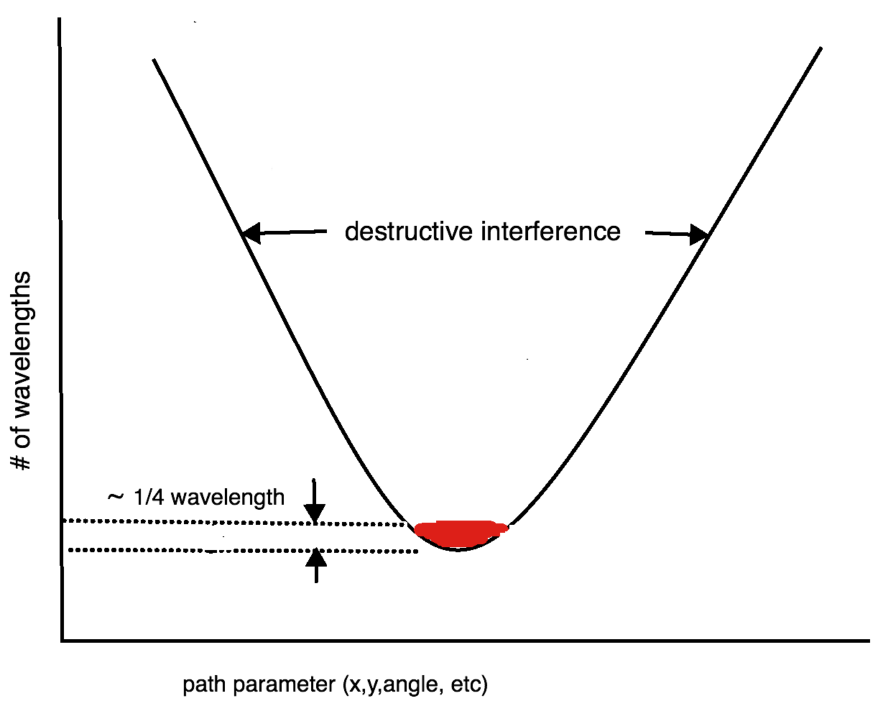


Constructive interference



Destructive interference

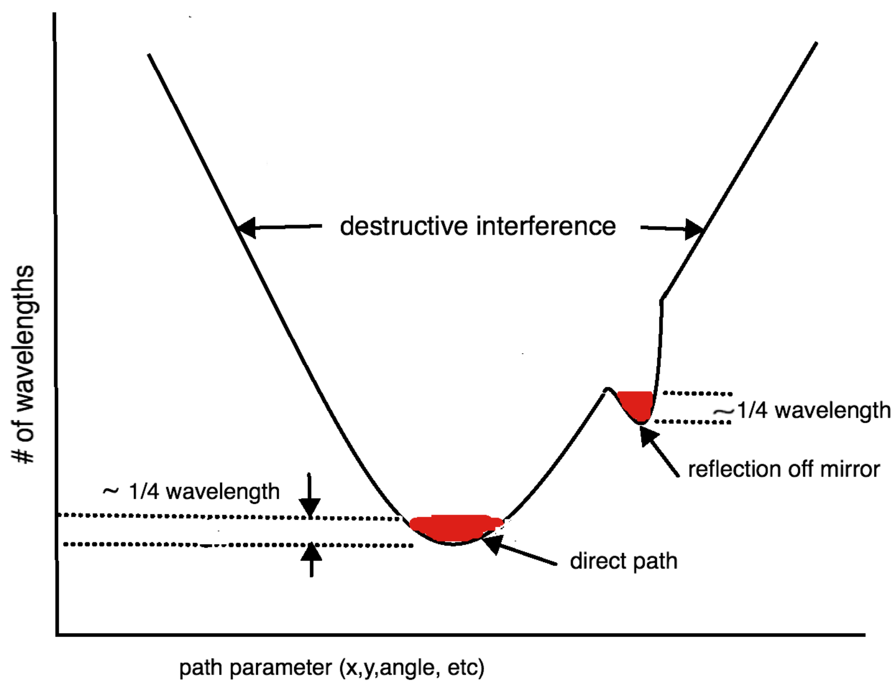
Of course if they all cancel each other, then no light goes from A to B. If we take the long path from A to B, there is always a shorter path, shorter by one half a wavelength that cancels that path and so on down until we get to the path of the minimum number of wavelengths from A to B. When we get to that point, the paths cease to cancel each other and around one quarter of a wavelength of the path of minimum wavelengths is the net result of where the light travels, almost a straight line but with a little slop in it. The slop will be important later.



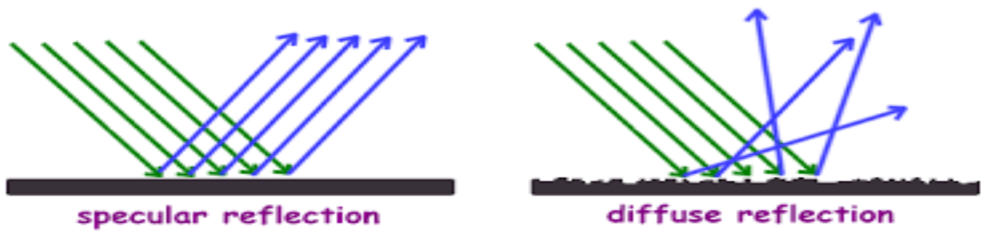
A lot of our results will be similar to Fermat's Principle that light travels on the path of least time from A to B. I can see where this interference of waves can result in light going on the path of the minimum number wavelengths but it's hard for me to understand how light can know which path takes the shortest time without trying all the other paths (at the same time). Also, we don't get the slop.

Flat Mirrors

There may be other paths that have a local minimum in the number of wavelengths. Take for instance the example of a flat mirror. The shortest path from A to B is just a straight line, but then there're many paths that reflect off the mirror. These paths have a minimum number of wavelengths when the angle of incidence equals the angle of reflection, known as "the law of reflection".



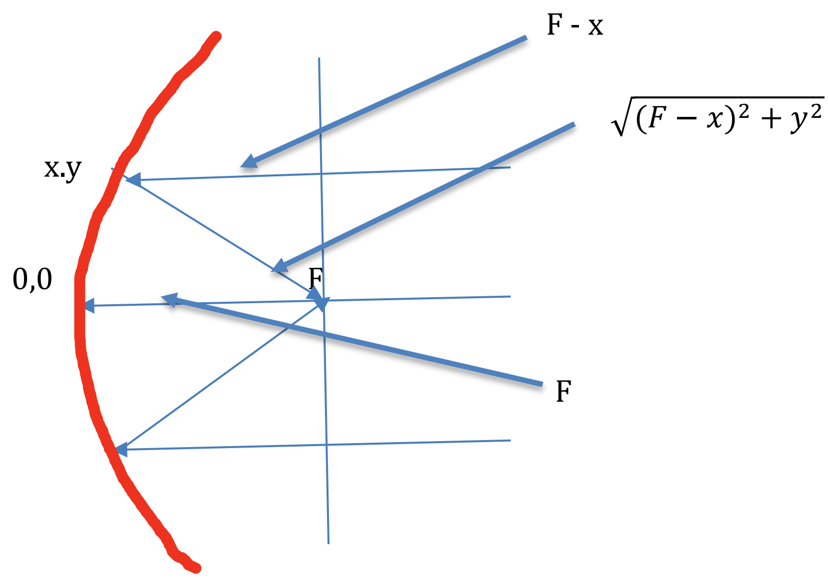
Now most books say that a smooth surface gives specular reflection whereas a rough surface gives diffuse reflection. Anyone who has used an atomic force microscope knows that every surface, except a crystal plane, is rough.



So how rough is rough? Mother's metal polish gives a pretty good mirror-like finish on aluminum, so you might wonder what the grit size is in the polish. A rough surface is like having several surfaces at random heights around some average. Each one of these surfaces has a path of minimum number of wavelengths, the problem is that these paths when they arrive at B, all have a somewhat random phase and so when they add they just give some average intensity independent of the angle. But, if the surface roughness is less than, say, a quarter of a wavelength, our slop allows that there will still be a path of minimum number of wavelengths with only a little interference between surfaces. Given that the wavelength of visible light is about .5 μ, this means that the grit size should be .1 μ or less. The people that seem to care about this on the Internet are the people who sharpen and shave with straight razors. They guess that the grit size is about .1 μ. By the way, using the principle of the path of least time won't get you very far in this argument. Notice that a mirror that may be pretty good in the red may not be so great in the blue and for sure won't be great in the ultraviolet. Ultraviolet mirrors need to be very smooth. Similar reflections happen with sound waves. You probably have all seen the reflector behind a microphone at a football game where they're trying to focus sounds from a distance. Here you want a good reflector. In a concert hall you want just the opposite. You want walls that give diffuse reflection. The problem is complicated by the fact that the wavelength of audible sound goes from centimeters to meters and also the speed of sound is much lower than light, so you get reflections off the walls that start interfering both constructively and destructively causing local variations in the sound intensity which depend on frequency. It is a mess!

Curved Mirrors, Focusing of Light

The subject of curved mirrors is usually approached by using the law of reflection. But since this "law" came about while we were thinking about waves, let's stick with waves. We usually want a curved mirror to focus light and so in our analysis what we want to happen is that the different paths reflecting off the mirror will arrive at some point (the focal point) with the same phase and will therefore interfere constructively



Assuming the waves are in phase at the vertical line through F,

to be in phase at F we need F – x + $\sqrt{(F-x)^2+y^2}$ = 2F

We show the light as rays coming in because we know that light travels essentially in a straight line in air but remember, these are waves.

So: F + x = $\sqrt{(F-x)^2+y^2}$

Squaring both sides and reducing, we get $x = y^2/4$F

So, the shape of the mirror is a parabola. Does this obey the "law of reflection"? Check for yourself, but that is baked into our approach.

A spherical shape does not focus very well and so, as is pointed out in the books, the parabola gets rid of what otherwise would be spherical distortion. But hold on. The parabola is fine for an object, which is far from the mirror, but it is not magic. Let's say that

we bring the object closer and closer to the mirror. In this case, the focus moves further and further from the mirror. At some point we get to where the object and the focus are at the same point in which case the light goes out from the object to all points on the mirror and then returns on the same path to the focus. The shape of a mirror that does this is a spherical shape, not a parabolic shape, with the focus at the center. In this case the sphere gets rid of the "parabolic distortion". So, one shape does not work for all. For a shallow mirror, say much shallower then the focal length, a parabolic mirror is very similar to a spherical mirror which has a radius of two times the focal length of the parabolic mirror.
Now how good does the mirror need to be? If we want light from different parts of the mirror to arrive at the focal point in phase, the mirror has got to be accurate to well less than a quarter of a wavelength,, probably around 1/10 to 1/20 of a wavelength. For visible light this is 25 to 50 nm. As we have already shown, the mirror cannot be rough compared to a wavelength.

So let's go back to objects at infinity. Consider the TV satellite dish. Mine picks up two satellites, one at 110° and the other at 119°, a 9° separation in the horizontal plane. It has separate detectors near the focal point for these two satellites. A parabolic reflector is a good reflector on axis but when you get waves coming in off axis, as is true from these two satellites, its focusing deteriorates. One property of the spherical mirror is that it has no axis, All angles have this same property and the focus just moves around. So one might want a TV dish which is spherical in the horizontal plane and parabolic in the vertical plane. Now take a good look at your satellite dish. You will notice that it is neither round (broader in the horizontal direction) nor axially symmetric (more curved in the vertical direction). It looks to me like a hybrid mirror which is parabolic vertically and spherical horizontally. By the way, these reflectors are quite shallow. Mine has a depth to focal length of about 1/10. It has a vertical width to focal length of 1.0. Large radio telescopes, which cannot move, are spherical so that all incoming directions are treated the same and the "pointing" is done by moving the detector to different focal points.

Now we dig deeper into mirrors. The question is how well a perfect parabolic mirror, which is perfectly smooth, can focus light from infinity. We mentioned that our straight line has some slop, so we might expect the image size to be of the order of a fraction of a wavelength. But life is more complicated than that. We need to get separate paths to interfere coherently at the focal point and so we must consider the geometry of these separate paths. Looking down from the top of the mirror, if we move off to the left of the focal point, the paths for the waves from the right side mirror are now longer and the paths from the left side of the mirror are shorter. At some distance to the left of the focal point the average of the waves from the right side of the mirror and the left side of the mirror will interfere destructively and there'll be no light at that point. This would be the half-width of the focus spot. If the mirror were square, it is easy to see how an element on the right side could interfere with a corresponding element on the left side, spaced apart by half the width of the mirror and so on across the whole mirror. The distance from the axis at which this occurs we can call the half-width of the focal spot.

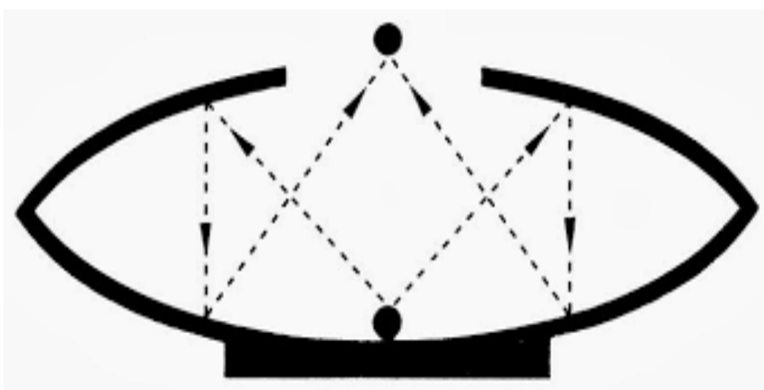
One can also argue for this square mirror that this total interference occurs when the light path from the centroid of the right side of the mirror is one-half a wavelength longer than the path from the centroid of the left side of the mirror. The angle between the axis of the mirror and the line from the point of total interference to the center of the mirror is equal to half the wavelength of light divided by the distance between the two centroids. That is, θ

= λ/D, where D is the width of the mirror. If you multiply this by the distance to the mirror, the focal length, you then get the half-width of the spot focused at the focal point.

Now a circular mirror is a little harder to analyze. What we will use as an estimate, is to take the distance between the centroids of the right and left sides of the mirror and again calculate the half-width of the focus spot, in this case the radius of the spot. The centroid of a semicircular disc is 4R/3 π from the flat side. This gives the distance between the centroids as 0.85 R. This is less than the square mirror because the round mirror is more weighted toward the center of the mirror. All this gives the radius of the focus spot as 1.18· f·λ·/2R. OK, the real answer is 1.22 when you integrate the phase over the surface of the mirror, but, as they say back home, this is close enough for government work and we said we are going to try and stay away from any fancy mathematics. This spot radius in terms of the diameter of the mirror, D, is 1.18· λ·f/D. f/D is called the f-number of the mirror. We will run into the f-number again when we do lenses. A formal analysis shows that 84% of all the light from the mirror falls within this spot. The rest of the light falls in circular rings around the central spot but decreases in intensity with increasing distance from the axis.

We notice a couple of things from our result. The first is that if the diameter of the mirror is just a few wavelengths, then the spot size is of the order of the focal length of the mirror, which means the mirror didn't focus at all. Don't expect those reflectors behind the microphones at sporting events to focus the sound from the bass drum of the marching band. If we want a spot size close to the wavelength of light, then we need a mirror whose f-number is about one. By the way, my TV dish has an f-number of one. My guess is that the wavelength of the TV signal is about 100th of the diameter of the dish, or about 7mm. Not too bad a guess because the actual wavelength is about 20 mm. The second is that we notice that the spot size depends on the wavelength. Red light would have a larger spot size than blue light. So if we had white light, the spot would become redder as we go toward the outside of it, yet some people say that a mirror has no chromatic aberration. But of course it does, light is a wave and the waves "have color", depending on their wavelength. What they mean is that the law of reflection works the same for all wavelengths and the focal length does not depend on the wavelength. The sound mirror behind the microphone treats high frequencies and low frequencies much differently.

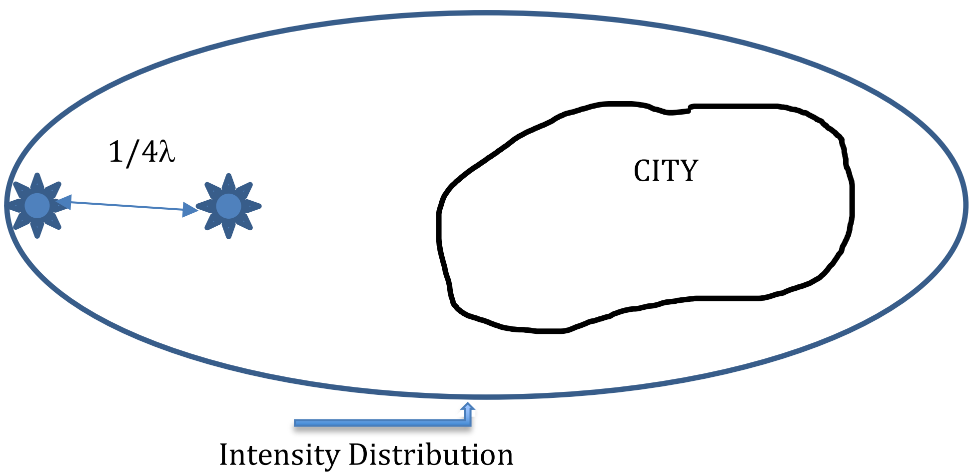
I was the co-author of a patent for a neat optical illusion which involved two parabolic mirrors. These mirrors were such that the focal point of each one of the mirrors was at the surface of the other mirror when the two mirrors were put face-to-face. Below is a picture of the illusion. The screw is down in the bottom of the device.



From the diagram you can see that an object on the bottom mirror, when illuminated, will appear as a real image at the hole in the top mirror. In fact, the surface of the bottom mirror makes an image at the hole so the hole does not look like a hole, it looks like a mirror. The illusion was discovered from two mirrors sitting on top of each other, which, oddly enough, had just the right focal lengths, and someone noticed that dust was collecting on the hole.

These parabolic mirrors can operate in reverse. If we place a light source at the focal point the mirror should produce a parallel beam going out to infinity. Well, almost parallel, it's those damn waves again. Using the same analysis as we did for the focus we find that the beam spreads with a half angle of: $\emptyset = 1.22\lambda/D$. This applies to light from any circular aperture whether it be a laser, a searchlight, or just a hole. For instance if we take a beam coming out of a red laser which has a ½ inch opening, then the spot on the moon will be about 300 kilometers wide. Not such a tiny spot!

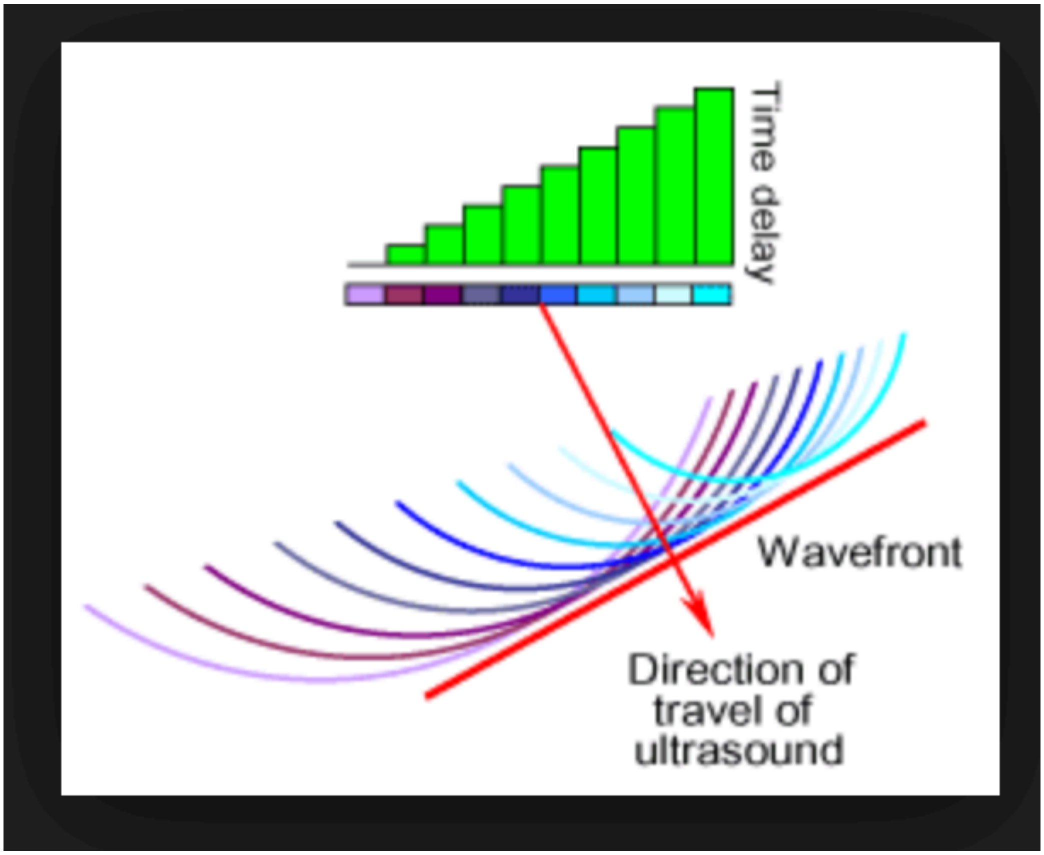
This business of getting rays to arrive at a spot in phase has other applications than just mirrors, especially if one has control of the phase of the source of the wave. I remember traveling in Iowa in the old days out in the countryside, I would come across a radio transmission station which had two antennas. Why two antennas, isn't one good enough? The single antenna radiates radio waves in all directions equally, but the radio station would like to beam the energy toward the city which is a few miles away. Consider the following diagram. Let's say we placed the antennas a quarter of a wavelength apart and operated them a quarter of a cycle out of phase with the tower on the left advanced in phase. When a wave from the left tower reaches the right tower, both waves would be in phase toward the city with constructive interference. In the direction opposite the city, though, when the wave from the right-hand tower gets to the left hand tower the waves are now out of phase and so in the opposite direction of the city we get destructive interference. Shown is an approximation of the intensity pattern, given what tools I have in Word. As a mental exercise, think of what the intensity distribution is like as the phase goes from one quarter of a wavelength advanced all away to one wavelength, which of course is equal to zero wavelengths.



.

One can add more antennas to refine the intensity distribution.

Another use of varying the phase of the emitters is in ultrasound. Here an array of piezoelectric emitters operate at the same frequency but with the phase between the various emitters being varied to move the focal point around in 2 dimensions to produce an image of, say, a liver.

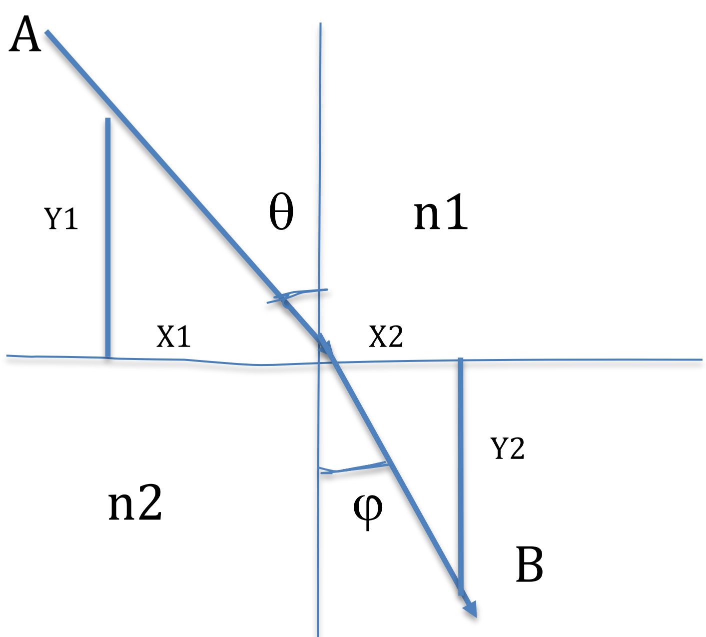


Here is an example of beam steering. the phase can also be used to focus the beam at a certain distance away from the emitter. This phased array idea can also be used in detection to determine in what direction a source is located. For a telescope, the angular resolution is the wavelength of the radiation being detected divided by the diameter of the receiver. For a 1 m optical telescope, this resolution is 0.1 arcseconds. For a radio telescope operating at a wavelength of about 1 cm, to get the same resolution would require a receiver with a diameter of 20km. This can be done with the phased array of radio telescopes in which the detected phase can be corrected so that the signals arrive with the same phase at some point, the point being in the electronics. From these phase corrections one can then determine the direction that the radiation is coming from. An example of such a phased array is the radio telescope array in New Mexico which the reader can look up on the Internet. This array can be made to have a size of about 30 kilometers giving an angular resolution of 0.2 to 0.04 arcseconds.

Another example is with submarines. One can detect the direction and position of other submarines by sending out a Ping and looking for reflections. This would be a bad idea since you're just telling the other submarine where you are. A better idea would be to detect passively noise from the other submarine. You would probably want to look for frequencies which are well below 1 kHz, maybe even 60 cycles. A problem is that this velocity of sound in water is about 1500 meters per second which makes the wavelength in water quite large. For the frequencies mentioned, the wavelength would be one and a half meters or longer. If one wants an angular resolution of 1°, then the size of the array would need to be over 100 m. This is about the length of a submarine and so maybe a better idea is to tow a long boom of detectors behind the submarine.

Light traveling in a transparent media

The interaction of light with electrons in a transparent media and the subsequent re-radiation from the electrons causes the speed of the light to slow down in the media from what it was in the vacuum. The ratio of the speed of light in the vacuum to the speed of light in the media is call the index of refraction and is greater than one. The frequency of the wave stays the same and so the slower speed means a shorter wavelength. If the light is going to go on a path of the minimum number of wavelength then the light will tend to stay away from areas of a short wavelength and prefer areas of longer wavelength. The path may be physically longer but still will be a path of a fewer number of wavelengths. Let's take the standard case with light going from a medium with an index of n1 into a medium with index n2 with n2>n1



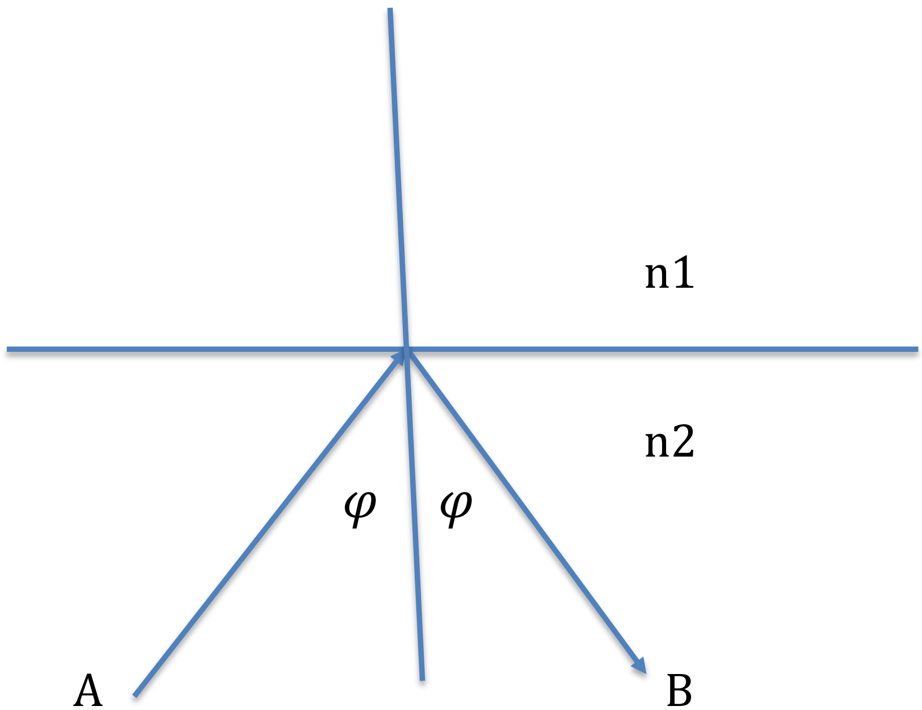
We will calculate the number of wavelengths from A to B and then minimize that number by moving X1 around to get the path. The wavelength in region one is λ/n1 and in region two, λ/n2, where λ is the wavelength in vacuum.

# of wavelengths = $\sqrt{X1^2 + Y1^2}\ n1/\lambda + \sqrt{X2^2 + Y2^2}\ n2/\lambda$ with X1+X2 = constant=C

take the derivative wrt X1 and set it to zero, we get, with dX2/dX1 = -1

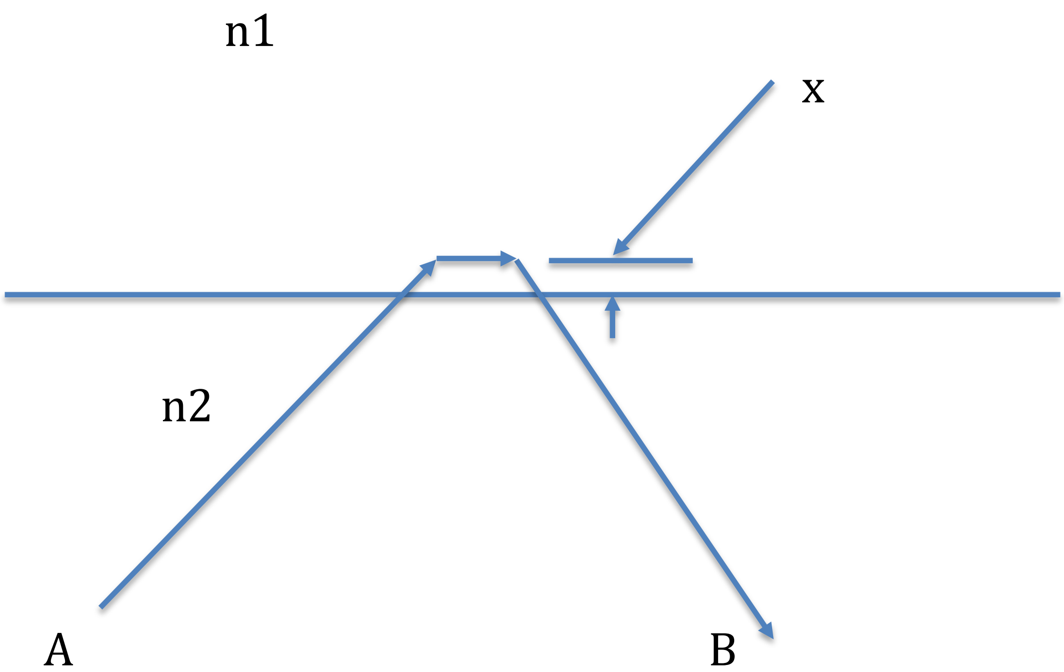
$$n1\sin\theta = n2\sin\varphi \quad \text{Snell's Law}$$

Yes, it is also the path of minimum time.
Now if the light goes from n2 into n1, then the angle φ will reach a point where θ goes to 90° and sinθ will be greater than1, which is not allowed. In this case the light will reflect off the surface between n2and n1. This phenomenon is called total internal reflection.



TOTAL INTERNAL REFLECTION

This may not be the path of minimum number of wavelengths between A and B. The wavelength of light is greater in n1 than in n2 so the path of minimum number of wavelengths that light travels along is one where light travels some distance in n1 very near the interface and then back into n2 to point B. The physical path is longer but the number of wavelengths is less.

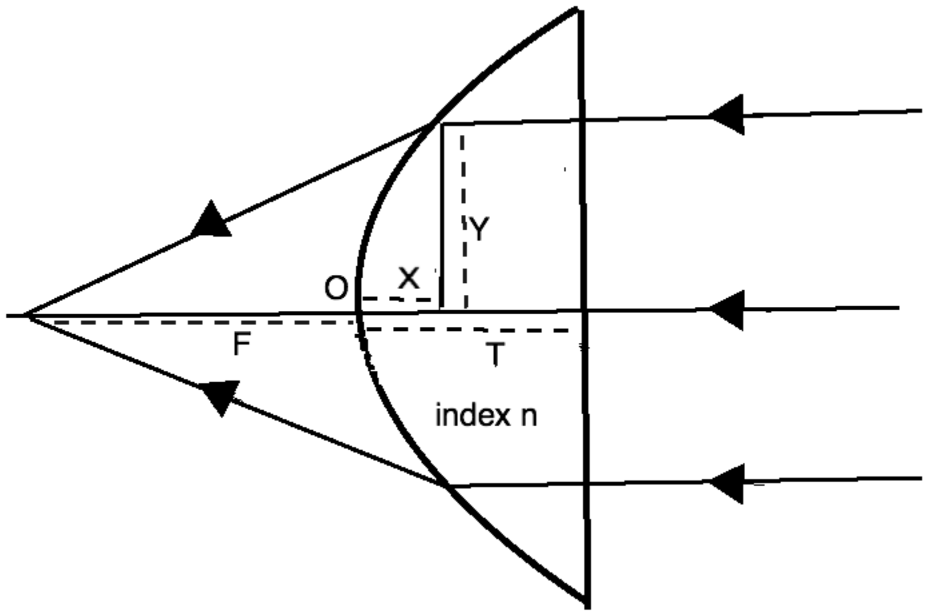


The light cannot go very far into n1 or the path will not be within 1/4 of a wavelength of the minimum number wavelengths. We know that if X is 1/8 of a wavelength then the light would be 1/4 of a wavelength out of phase with the path that goes right along the interface and so the intensity of light in n1 will drop very quickly with X. From our previous arguments, then, X will be no more than about 1/8 of a wavelength. This light in n1 is called the evanescent wave. A more complete analysis shows that the light in n1 drops off exponentially with a negative exponent of 2π times the wavelength.

LENSES

Lenses, just as mirrors, can focus light from different paths to a single point. Again, the idea is to make the number of wavelengths on each of the paths the same so that the light on the different paths will arrive at the focal point in phase. Below I have laid out a Plano Convex lens, meaning the lens is flat on one side and convex on the other. I put the origin at the surface of the convex side and have indicated other distances with dashed lines. What we're going to do is calculate what the shape of the surface of the lens should be to make the different paths arrive at the focal point with the same phase. We want Y as a function of X. The waves are in phase before they enter the flat side of the lens i.e. a plane wave.

Canceling out the wavelength in air, we have



$$Tn + F = (T - X)n + \sqrt{((F + X)^2 + Y^2))}$$

This reduces to:

$$2FX(n - 1) + X^2(n^2 - 1) = Y^2$$

This gives Y as a function of X for the lens surface of this aspheric lens.

For X<<F (a thin lens) we have

$$2FX(n - 1) = Y^2$$

This is the equation of a circle (ignoring terms of x squared) with the origin at the edge of the circle on the X axis, with:

$$R = F(n - 1) \quad F = R/(n - 1)$$

where R is the radius of the circle. So for a thin lens, the shape in three dimensions is just a spherical lens.

For X>>F we have

$$X^2(n^2-1)=Y^2$$

These are 2 straight lines passing through the origin with slopes of:

$$\pm\sqrt{(n^2-1)}$$

This correct lens is called an aspheric lens because it is not spherical (no kidding), but then again, it's the lens that does the correct job. You will now see cameras advertised with aspheric lenses. Spherical lenses, which are easier to make, have spherical aberrations just as the spherical mirror did.

One may think that we should check whether Snell's Law is obeyed. Remember that Snell's Law was a result of our approach, not the other way around. Snell's Law is baked in. So how good (or bad) are these lenses? Let's start with the spherical lens. We said that it had aberrations but let's be quantitative and find out how bad they are. We will turn the problem around. Instead of finding the shape that makes the focal length constant for all values of X (the aspheric lens), let's take the shape to be spherical with radius R and find out how the focal length varies with the thickness of the lens. Consider the same drawing with the flat side of the lens going through the center of the circle. Here we have, with the wavelength in air being canceled out:

$$Rn+F=(R-X)n+\sqrt{((F+X)^2+Y^2)}$$

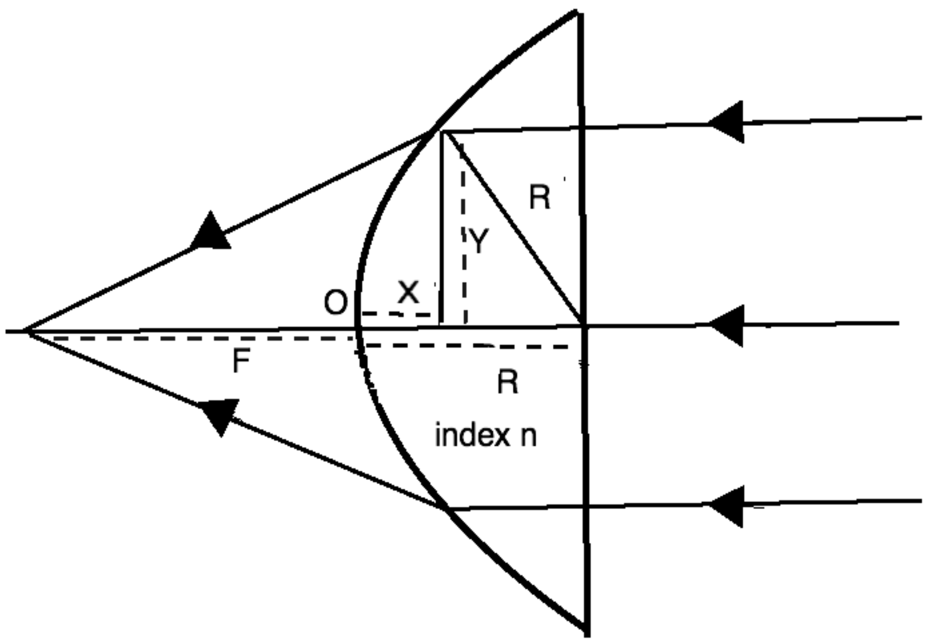
But:

$$Y^2=R^2-(R-X)^2=2RX-X^2$$

Substitute and do the algebra, and:

$$F=R/(n-1)-Xn^2/2(n-1)$$

We get the standard result for the focal length but only when X is zero, otherwise the focal length gets shorter as the thickness increases. For an index of 1.5 we get that dF/ dX is equal to about two. Not a trivial amount. As the thickness of the lens changes by dX, the focal length for that part of the lens decreases by twice that amount.

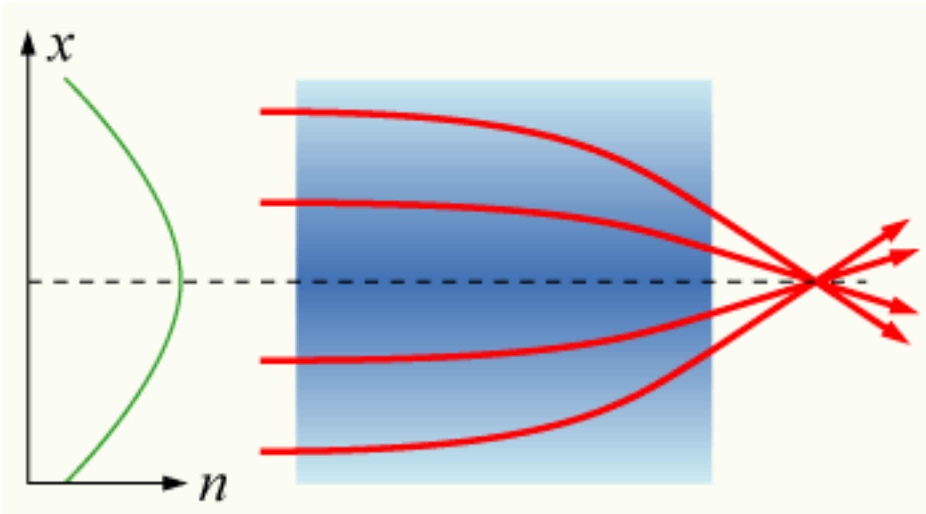


The focal length decreases as we go from the center of the lens to the outer parts of the lens and so the focus gets spread out along the axis of the lens.

How good is the aspheric lens? As with the circular mirror, we have the problem of the diffraction pattern of light passing through a circular aperture. The radius of the focal spot is $1.22\lambda f/D$ where D is the diameter of the lens, the same as the mirror. The smaller the F number, f/D, the smaller the spot size. Let's put in some numbers. For a wavelength of half a micron and an F number of 4 we get a spot diameter of 14 microns, about the pixel pitch of the sensor of a SLR camera.

The idea of the lens was to make several paths the path of the minimum number of wavelengths. They're also the paths of minimum time. These two approaches, the path of minimum time and the path of the minimum number of wavelengths, have tracked each other so far, but when we get to general relativity, where time and space get messed up, but the speed of light stays constant, these two approaches will part company. Guess which one will be right?

There is another way to make the lens and that is not by making a shape, but by varying the index of refraction of the lens. The path of the minimum number of wavelengths will shy away from the area of the high index of refraction, which is the area of shorter wavelength. Although the physical path is longer than a straight line, the number of wavelengths is shorter. The path will be concave toward the higher index of refraction. Such a lens is called a GRIN lens which stands for GRaded INdex of Refraction lens.



These lenses are sometimes used to get light from an LED into an optical fiber and are typically about 1 mm in diameter.

Here is another example of graded index which I will call the light merry-go-round. Let us say that we have a circular piece of glass in which the index of refraction decreases linearly toward the outside diameter. Let us say the index of refraction looks like the following.

$$n = (r_0 - r)/a$$

If we calculate the number of wavelengths in a diameter with radius r then we get:

$$\#\lambda = 2\pi r(r_0 - r)/\lambda_0 a$$

Minimizing this number by taking the derivative and setting it equal to zero, we get:

$$r = r_0/2 \implies n = r_0/2a$$

At this radius the light will go around in a circle. This leaves the following questions:

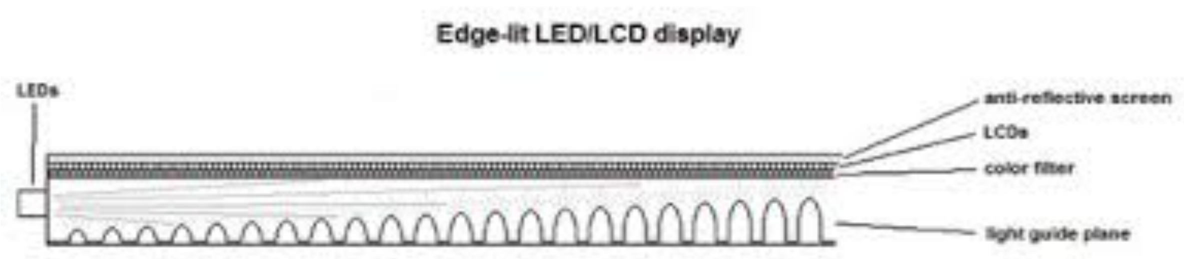
1. Will it work?
2. Can you pump it up?
3. Will the light beam separate into filaments in which each filament has an integral number of wavelengths as it goes around the diameter?
4. Will the filaments interfere destructively?

I don't know the answers, only the questions.

LIGHT PIPES, OPTICAL FIBERS, AND THE LIKE

Edge lit LCD TV

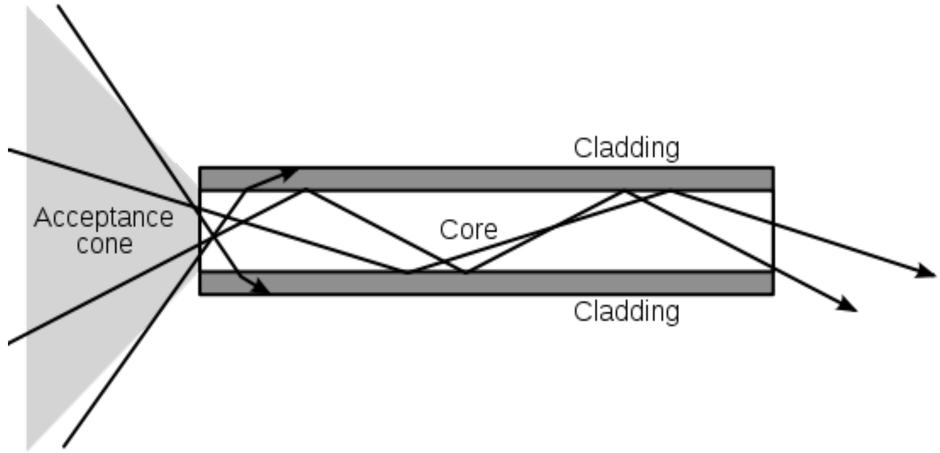
In a medium with an index of refraction greater than one, you can use total internal reflection to guide the light down through the medium as long as the light is at shallow angles. A common light pipe is one that is in almost all households and is the one that creates the back lighting for an LCD television screen. This light guide is a sheet of plastic (or glass) which has LEDs on one or two edges which direct light into the plastic. The plastic acts as an internal reflecting light guide which of course does no good because that doesn't illuminate the TV screen which is on one face of the light pipe. So, what one wants to create is a "leaky" light guide which leaks light out of the front face toward the LCD elements. One way to do this is to put bumps or depressions on the light guide which cause the light to scatter at various angles and some of it will go out the front face. Some will go out the back face but you can put a reflector there to redirect the light to the front face. This is shown in the drawing below.



As the light goes from left to right with some of it scattering out through the front face, the intensity of light in the light pipe diminishes. Therefore in this illustration the bumps get bigger and bigger as one goes from left to right. One can also do this by keeping the height of the bumps the same and increase the number of bumps toward the right. Various shapes of bumps and depressions have been used. One can also print dots of a diffuse reflective material on the back face of the light pipe to scatter the light out through the front face. One can put LEDs on both edges of the light pipe and then the largest bumps would be in the center of the screen.

Light pipes for data transmission

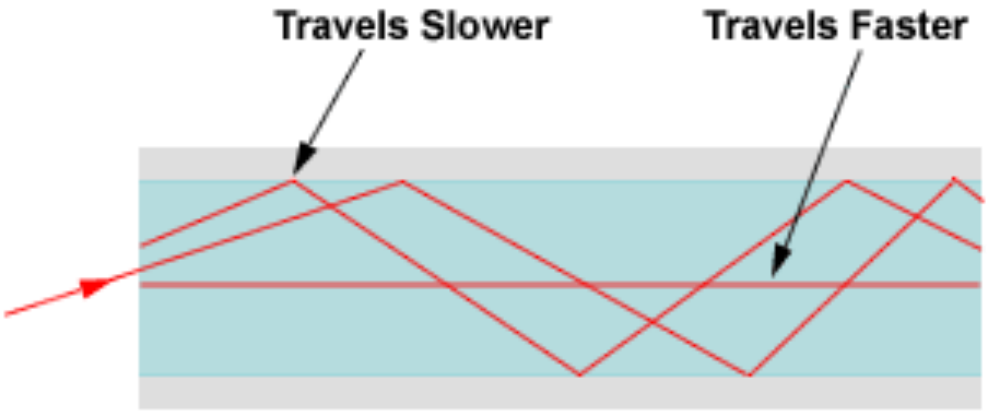
The most popular use of light pipes is an optical fiber to transmit data. These fibers have the capability of moving data at higher rates and much further than copper wire, which is typically used for telephones. An optical fiber consists of a small cylindrical strand of glass which had a fairly high index of refraction. Any imperfections in the surface, though, would cause the internal reflection to get messed up a little bit and some of the light would escape. To avoid this, a cladding is put on the outside of the fiber of a material that has a lower index of refraction. The total internal reflection therefore occurs at the interface of the core of the fiber and the cladding. The cladding is typically pure silica and the core is silica which has been doped to have a higher index of refraction. To analyze an optical fiber properly, one must treat light as an electromagnetic wave in a wave guide. This is well beyond our simple approach and so we are going to forge ahead with "light is a wave" and see how things go.



Typical fibers operate in the infrared with the largest angle of light traveling through the core, with respect to the straight-through path, of about 0,1 to 0.15 radians, a fairly small angle. The largest accepted angle into the fiber is the index times the maximum angle inside. This maximum accepted angle is called the numerical aperture, NA, of the fiber. The wavelength in the fiber is typically 0.5 to 1 micron and the index in the core is typically 1.5. Fibers come in basically two kinds. One is a multi-mode fiber which has a core with a radius of several wavelengths of light and the other is a single-mode fiber where the core radius is only a few wavelengths of light. We will consider the differences of these two fibers in the following.

Multi-mode Optical Fiber

A typical multi-mode fiber has a core which is either 50 or 62.5 μm in diameter. As shown below, there are many paths in a multi-mode fiber, some longer than others. The shortest path, of course, is straight down the core. The other paths which are reflected off of the cladding are longer and therefore take more time to travel down the fiber. A path that is it at an angle θ with respect to the axis of the fiber will have, in a length L of fiber, a path which is longer than the path down the axis by the amount $\Delta = L(1/\cos\theta - 1) \cong L\theta^2/2$. For 1km of fiber and $\theta = 0.1$(the maximum angle) this difference is 5mm, independent of the diameter of the fiber. If a square wave or sine wave signal were sent down the fiber with peaks spaced at 5 mm, at the end of 1 km the signal would be wiped out by this dispersion.



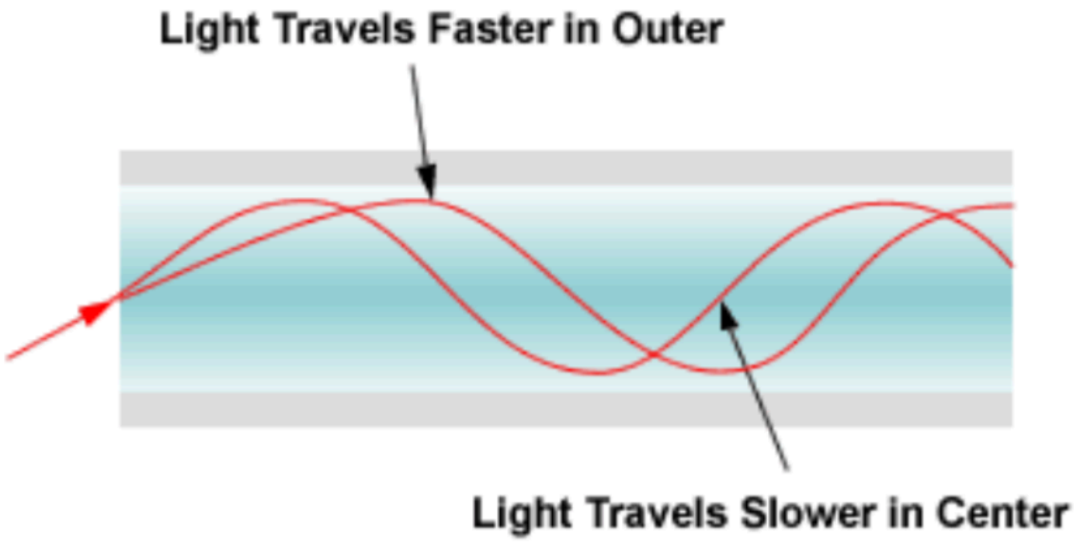
Step-Index Multimode Fiber

Let's calculate what frequency that signal would be. The speed of a sine wave is the frequency times the wavelength which is just equal to the velocity of light divided by the index of refraction for the fiber.

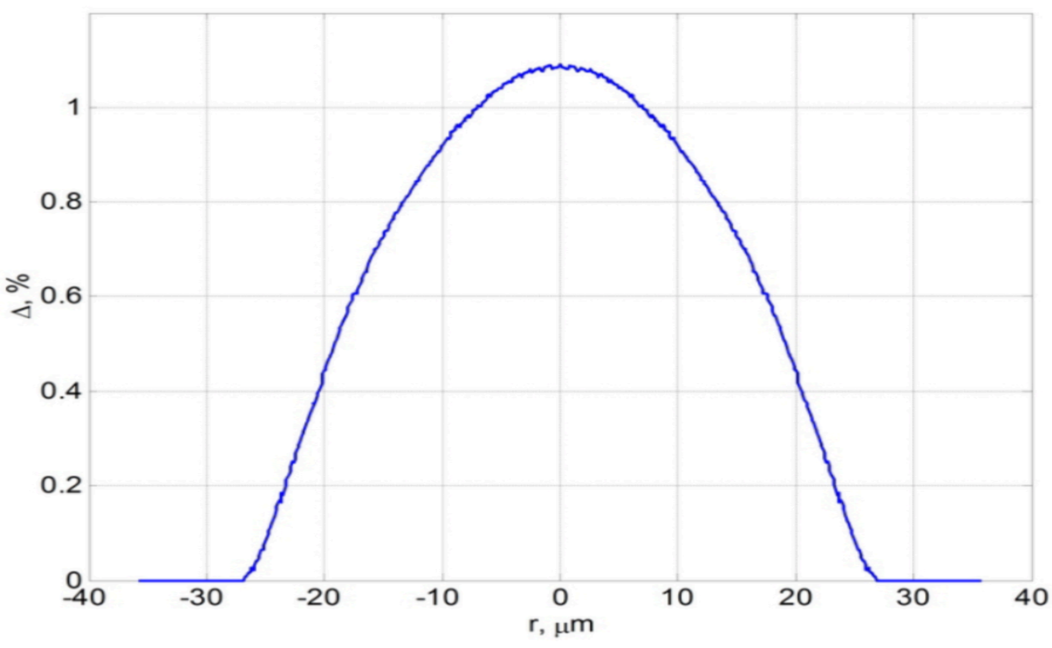
$f \cdot 5mm = c/n$ which gives $f = 40$ Mhz

For a distance of 1 km, this fiber can handle only signals below 40 MHz, called a bandwidth of 40 Mhz-km, not great for sending lots of information long distances but OK just for transporting light, such as in medical applications. This result is independent of the diameter of the fiber, but varies as the inverse of the square of the numerical aperture. Another problem with this multi-mode fiber is that these different paths can interfere constructively at various places in the fiber. This causes a speckle pattern of bright spots coming out of the end of the fiber, each traveling at a different velocity. These spots can number in the hundreds for a 60 µm step-index fiber. This speckle pattern is dependent on any slight bending of the fiber, which alters the different optical paths and therefore the position of the spots. You can make an intrusion alarm by using this effect to detect slight motion in a footpath.

I notice that you can buy 50 µm fiber which has a range of bandwidths up to 2000 MHz-km (OM3), so what is going on? Well, what is going on is a graded-index optical fiber, much like the GRIN lens that we talked about earlier. The idea in the GRIN lens was to have all the optical paths have the same number of wavelengths, so that they would arrive at a point all in phase. It is the same idea with the graded index optical fiber, having the longer paths traverse an area where they have longer wavelengths. A lower index of refraction means a longer wavelength, so the idea is to have the index of refraction decrease as you go towards the outer edge of the core where the longer paths travel. The index profile of a Corning 50 µm OM3 fiber is shown below. Notice that there is not a large change in the index of refraction from the center to the edge of the core, the difference being a little over one percent.



Graded-Index Multimode Fiber



It seems that the manufactures have put all of their efforts of graded index into 50 µm fibers, so, the 50 µm fibers have a much higher bandwidth than the 62.5 µm fibers. This difference is not caused by the different diameters of the fibers. Nonetheless, the constructive interference in the fiber still remains but in a diminished amount.

Single-mode optical fiber

The idea here is to make a fiber where there is none of that constructive interference going on and, in waveguide language, consists of only one electromagnetic mode in the core. We are not going to do waveguide modes and will continue on our “light is a wave” theme and use reasonable assumptions in order to make some calculations (I call it reasonable, but others call it hand waving). We want to calculate the diameter needed for such a fiber, and a parameter called the waveguide chromatic dispersion. We will also talk a little about dispersion-compensated fibers. Since we are going to do some calculations, we need to be careful with our numbers. I will take specifications from a Corning SMF-28 optical fiber which I understand to be a fairly standard single-mode fiber.

Characterized parameters are typical values.

Core Diameter	8.2 μm
Numerical Aperture	0.14 NA is measured at the one percent power level of a one-dimensional far-field scan at 1310 nm.
Effective Group Index of Refraction (N_{eff})	1310 nm: 1.4676 1550 nm: 1.4682

Cable Cutoff Wavelength (λ_{cc})

$\lambda_{cc} \leq 1260$ nm

Mode-Field Diameter

Wavelength (nm)	MFD (μm)
1310	9.2 ± 0.4
1550	10.4 ± 0.5

Maximum Attenuation

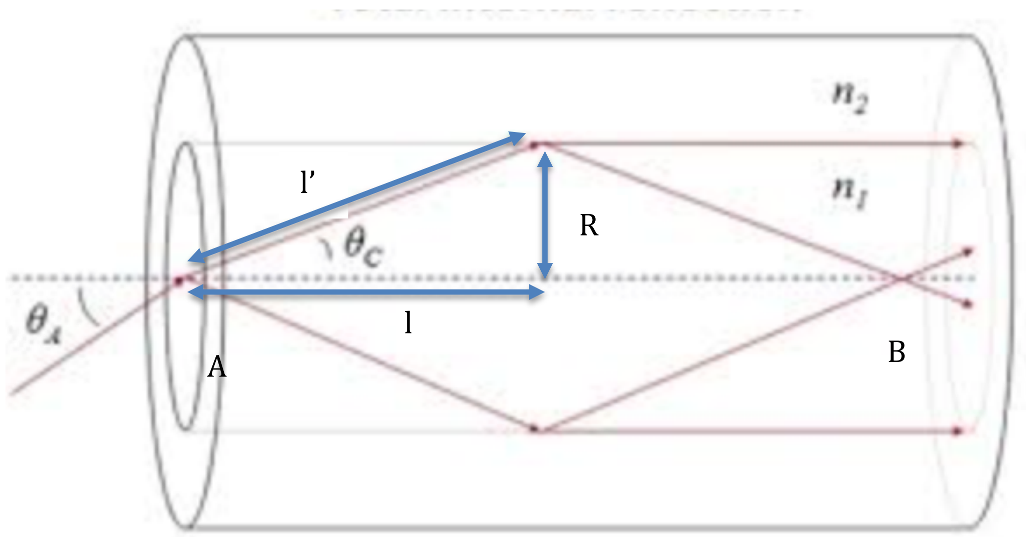
Wavelength (nm)	Maximum Value* (dB/km)
1310	≤ 0.32
1383**	≤ 0.32
1490	≤ 0.21
1550	≤ 0.18
1625	≤ 0.20

Fiber diameter

We want a fiber that, given the different paths traveling down the core and recombining, keeps the phase shifts small. Consider the drawing below.

We treat the fiber as a step-index fiber It shows the path for light at the critical angle θ_c in the core. We want light traveling along this path from A to B to not vary in phase very much from the path down the axis . At most, we want the phase of the two paths to not vary by more than half of a wavelength. Any more than that runs the risk of getting into the area of constructive interference between the two paths, that is, the fiber is no longer single mode. We can do this by making the core of the fiber small. Using the ½ wavelength criteria, the path at the critical angle will not even exist, because, when it reflects back down to the axis of the core, it will interfere destructively with light traveling along the axis (remember our argument that light can go on any path but most paths are canceled). Paths at angles near the critical angle will also be attenuated because of this destructive interference. Therefore, the light will be concentrated along the axis of the fiber with little near the edge of the core.

We want the path l' to be equal to l +λ/4, where λ is the wavelength in the core. This will determine the radius R of the core needed to do this. Smaller radii are OK because they give a smaller phase shift but will have light at the edge of the core.



For small angles, $\theta_c = \theta_A / n_1$ = NA/n_1, where NA is the numerical aperture of the fiber and λ = λ_{out}/n_1 , is the wavelength outside the fiber divided by the index of the core.

l' = l/cosθ_c ≅ l·$(1 + \theta_c^2/2)$; with R ≅ l·θ_c, one gets D = 2R = λ_{out}/NA

For the SMF-28 fiber this gives a maximum diameter of 1.310/.14 = 9.4 μm
The actual core diameter is 8.2 μm and so they are on the safe side. Notice that our diameter calculation is essentially equal to something called the mode-field diameter at 1310nm which they give as 9.2 μm. If the light distribution across the diameter of the fiber is represented by a Gaussian distribution, then the mode-field diameter is the diameter at which the intensity has fallen to $1/e^2$ of the maximum intensity on the axis. This is essentially what we calculated, the diameter of the light distribution out to where the distribution falls to near zero. So we might say MFD ≅ λ_{out}/NA. This is also true at 1550nm for this fiber.
For a given core diameter, as the wavelength decreases, the phase shift increases and the fiber will cease to be single mode. This is called the cut-off wavelength. Since our calculation gives the maximum diameter for a given wavelength, then it also gives the minimum wavelength, the cut-off wavelength, for a fixed diameter. For the SMF-28 fiber we get the following. λ_{cc} = core diameter · NA = 8.2· 0.14 = 1150nm. This is close to the 1260nm in their specs. They may be on the conservative side.

Waveguide chromatic dispersion

It is clear from what we have done, and the fiber specs, that as the wavelength increases, the mode-field diameter also increases. This means that the average path length down the fiber also increases, which decreases the group velocity, which means the effective group index of refraction <u>increases</u>. This effect is called waveguide chromatic dispersion and, interestingly enough, has the opposite sign of the chromatic dispersion that we are familiar with. Usually, as light travels through a transparent material, the index of refraction decreases as the wavelength increases. This is called material dispersion.

We want to calculate a value for this waveguide dispersion, dn_{eff}/dλ using the longest path. For small changes, the index of refraction(n), the velocity of light(v), and the path length(l) are all related;

dn/n = -dv/v = dl'/l , we use R= $\lambda_{out}/2 \cdot$NA and θ_c= NA/n_1; Working through the algebra gives;

$$dn_{eff}/d\lambda = (NA)^2/2n_1 \cdot \lambda_{out} = 0.005/\text{micron} \quad \text{with } \lambda_{out} = 1310 \text{ nm.}$$

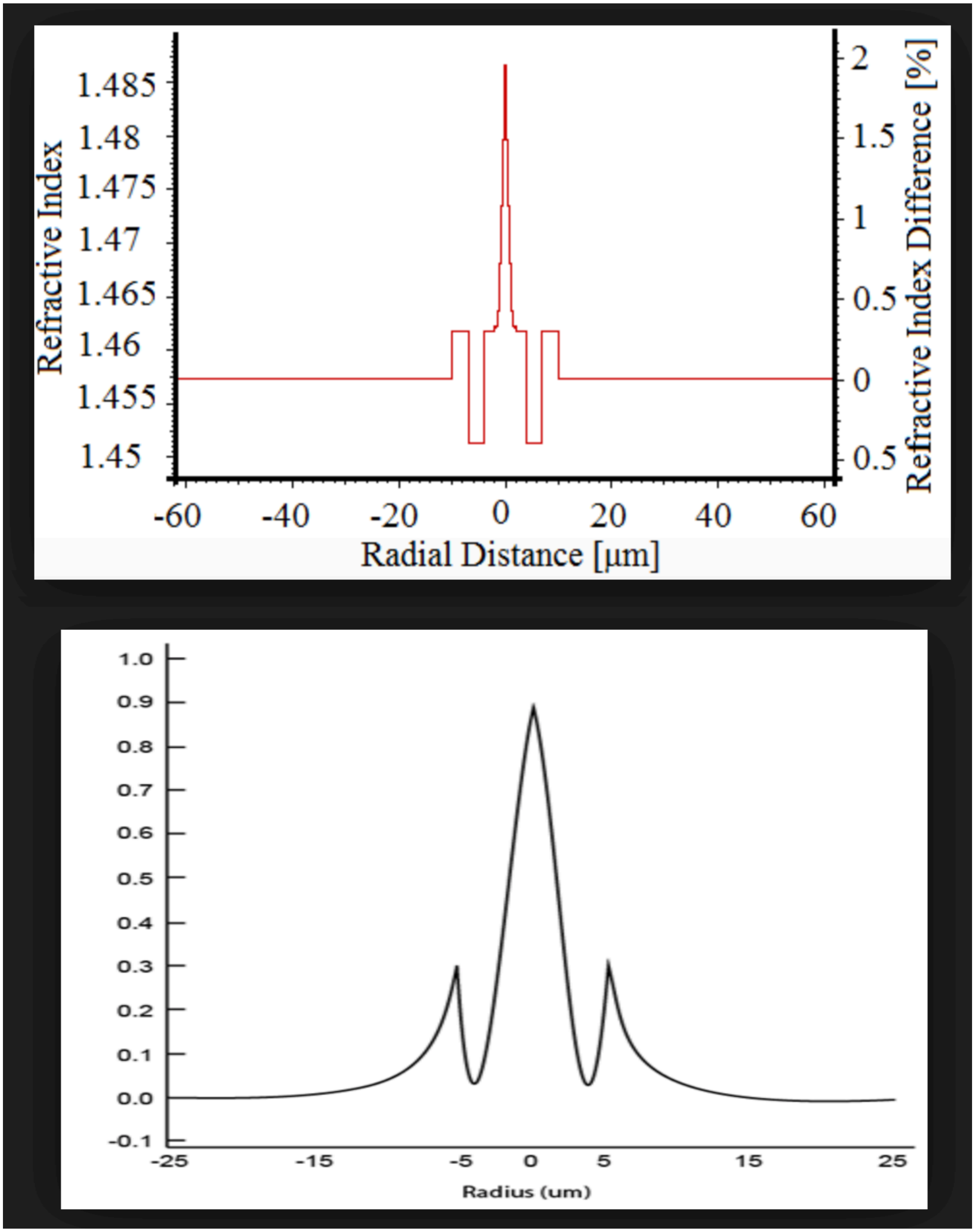
We missed on this one, but we got the sign right, the waveguide index increases with increasing wavelength. The real answer is 0.011.

As we know from our previous discussions on internal reflection, some of the light travels in the cladding very near the surface of the core. I have seen in many places the argument that the waveguide dispersion is caused by the fact that when the wavelength increases, the mode-field diameter increases, and more light goes into the cladding. This argument is wrong for the step index fiber because if more light goes into the cladding, which has a lower index than the core, then the effective index will decrease, not increase, as is the real case. The argument has the wrong sign.

What is interesting about this waveguide dispersion is that at 1310 nm it is equal to and opposite the material dispersion, so that the fiber has no chromatic dispersion at 1310 nm.

Dispersion Compensated Fiber.

From the attenuation specifications on the fiber, it's clear that you would like to operate at 1550nm. The problem is that waveguide index decreases because of the light in the cladding, caused by the larger mode diameter, and so the dispersion is not compensated as it is at 1310nm. The game is to create an index profile across the diameter such that when the mode-field diameter increases, and more light goes into the cladding, the effective index increases and not decreases. This can be done by putting a ring of high index material just outside of the core-cladding interface. Below are shown a couple of instances to illustrate this. As the mode-field diameter increases, more of the light goes into the high-index ring to increase the effective index. One can adjust the parameters such that this waveguide dispersion will be equal and opposite to the material dispersion at or near 1550 nm.



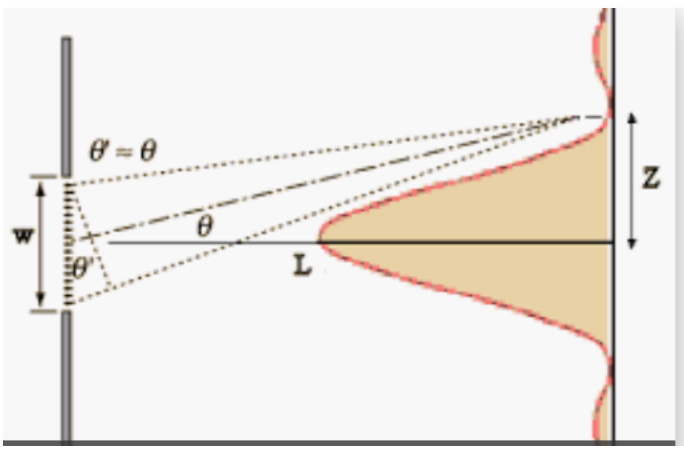
Refractive Index
1.485
1.48
1.475
1.47
1.465
1.46
1.455
1.45
2
1.5
1
0.5
0
0.5
Refractive Index Difference [%]
-60
-40
-20
0
20
40
60
Radial Distance [μm]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-25
-15
-5
0
5
15
25
Radius (um)

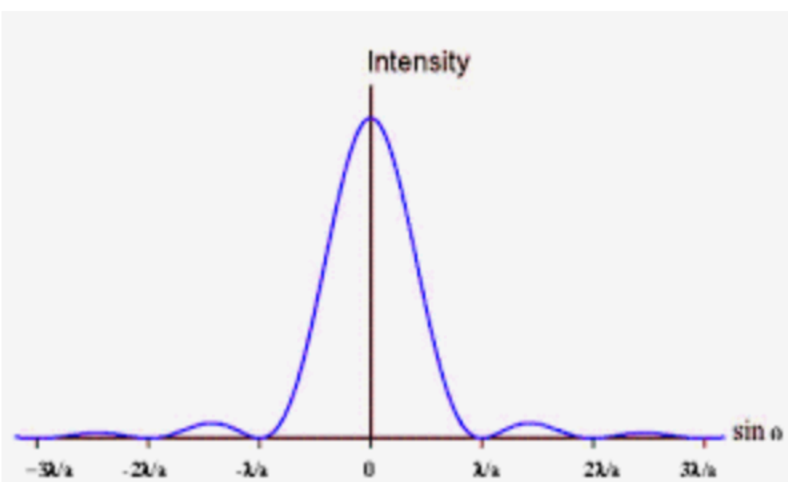
Particles, Waves, and the Heisenberg Uncertainty Principle

In 1905, to explain the photo electric effect, Einstein said that light interacts as if it's a particle called a photon with an energy hf, where h is Planck's constant and f is the frequency of the light. The momentum of this massless particle is, therefore:

$$p = E/c = hf/c = hc/\lambda c = h/\lambda = \text{momentum}$$

The shorter the wavelength, the more momentum the photon has.
Let's consider the case of light shining through a single slit onto a screen. There is a diffraction pattern which spreads the beam out by the angle $\theta = \frac{\lambda}{w}$. About 90% of the entire intensity is in the central peak which I will consider Gaussian shaped.





Now let's consider that the light consists of photons going through the slit. The diffraction pattern is made of many, many photons. Before they go through the slit they have a momentum which is perpendicular to the slit with no transverse momentum. But ,clearly, after they go through the slit, there seems to be some uncertainty in the transverse momentum because the photons on the average make a pattern which widens out. They have picked up transverse momentum by going through the slit but in an uncertain way for each photon. Just after the photons have went through the slit, let's calculate the average uncertainty in their position times the average uncertainty in transverse momentum. We consider the axis that is across the width of the slit..
The <u>maximum</u> uncertainty in the position of the photon is half the width of the slit, w/2 but 50% of the time, the photon is within w/4 of the center of the slit, so we take the average uncertainty in position to be w/4.
The uncertainty in the transverse momentum is a little harder to calculate, but we will make some "educated" guesses. For the central peak, if it is somewhat like a Gaussian, then more than half of the photons will fall within 1/3 the maximum width of the peak, that is within an angle $\lambda/3w$ from the axis of the incoming beam. The contribution of the center peak to the transverse momentum average is then $0.9\lambda/3w$ times the momentum, h/λ, of the photon. I have assumed that 90% of the photons fall in the center peak. The remaining peaks fall very rapidly in intensity as they go out from the axis. For instance the second peak is 4% of the main peak and the next peak is 1.6%. So, I will take the remaining 10% of the photons and put them out at an angle which is three times the angle of the first minimum. Their contribution to the transverse momentum therefore is $0.1\cdot3\lambda/w$ times the momentum of the incoming photon. Adding all this together we get;

$$\Delta x= w/4 \quad \Delta p= (0.9/3 +0.1\cdot3)\lambda/w\cdot h/\lambda = 0.6h/w \text{ so:}$$

$$\Delta x\cdot\Delta p = .15h \cong h/2\pi,$$ which is the Heisenberg uncertainty principle.

I know, you think I cheated, but what would you use for numbers? As an exercise you might analyze the diffraction through a circular aperture and see what you get in an analysis like this one.
A couple of things to note. When thinking about light traveling as a bunch of photons, they travel through the slit without interacting with the slit. You could say that merely the <u>knowledge</u> that the photons went through the slit is enough to create uncertainty in their transverse momentum. It's probably best to think of light traveling as a wave and interacting as a particle. The other thing is that we just showed that the uncertainty principle for photons could've been done in 1905 when Einstein said that light acted both as a wave and a particle. It wasn't until 22 years later that Heisenberg came up with the uncertainty principle. Maybe what was required was that de Broglie postulated that even massive particles travel as a wave. He postulated this in 1924, again about 20 years after Einstein. de Broglie said that massive particles have a wavelength which is equal to Planck's constant divided by their momentum, $\lambda = h/p$, the same as for a photon.

Compared to visible photons, this wavelength is very short. For instance visible photons have an energy of about 0.5 electron volts . Let's compare that to the wavelength of a "slow moving" electron, one which is going 1/100 the speed of light. The rest mass energy of an electron is 0.5 million electron volts.

For a visible photon λ = h/p = hc/E = hc/0.5ev

For this slow moving electron λ = h/p = h/mv = $hc^2/mc^2 \cdot v$ = hc/0.5Mev(c/v) = hc/5000ev

So, this electron has a wavelength which is 10000 times shorter than the visible photon, unnoticed in everyday life. How about the uncertainty principal for these massive particles like the electron? Well, same as the photon. Basically, if something travels as a wave, it cannot be localized unless it's a wave packet. But a wave packet is the addition of waves of different wavelength and in this case, different momenta. The wider the spread in the momenta, the more the wave packet can be localized, that is, the more the uncertainty in momentum, the smaller the uncertainty in position.

In 1927, experiments of firing electrons through crystals and metal foils gave a diffraction pattern just as de Broglie predicted. But, again, Einstein was 20 years ahead of all this. He was putting the finishing touches on General Relativity when others were generalizing his photon idea of 1905.

We seem to of come full circle. Our initial assumption was that waves (light) can travel on any and all paths and then interfere to give the net result. Does that mean that particles like electrons, which travel as waves, can travel on all paths, and then these paths interfere to give the net result? Of course, that's how the photons can make the diffraction pattern. It seems like we are back to Feynman's summing over paths, which is where we started.

OK, that is our tour through optics. We could probably go further but I think we've done enough. If you have comments or corrections please send them to optics @ virgilelings.com