Notes on

Waves

Daniel F. Styer; Schiffer Professor of Physics; Oberlin College

Abstract: Classical waves. More than beach play, this topic covers a lot of territory.

Contents

Chapter 1

Introduction to Waves

What is a wave? When I was a teen I read popular science books saying that "In quantum mechanics, an electron behaves somewhat like a particle and somewhat like a wave." OK, I thought. I know what a particle is. But what's a wave? I had only used the word "wave" at a beach. When the popular science books said "like a wave" they clearly didn't mean "made of salt water". So what did they mean? Here are some possible answers:

- **A function**

  $$y(x,t) = A\sin(kx - \omega t).$$

  There are problems with this definition. First of all, the function $\sin(kx - \omega t)$ extends over all space and all time. This wave started infinitely far in the past and will keep going for ever and ever, amen. Real waves are finite in space (waves on the ocean end when they hit the beach) and of course finite in time.

  ⟦Following songwriters Nickolas Ashford and Valerie Simpson, ("Ain't No Mountain High Enough," 1966, sung most famously by Marvin Gaye and Tammi Terrell) I like to say that there "ain't no ocean wide enough, ain't no string long enough" to carry a pure sine wave.⟧

  Furthermore, there are waves like tsunamis that are one big pulse, not a periodic repetition like a sine wave. This proposed definition of "wave" is too narrow.

- **A moving extended object.** This proposed definition is too broad. Toss a football. That's a moving extended object, but it's not a wave.

  The proposed definition is at the same time too narrow. Suppose there's an underwater boulder on a stream bed. The flowing water will pile up into a mound upstream of the boulder, a mound called a "standing wave". The mound doesn't move: the water moves, but the pattern remains fixed. So waves don't *have* to move.

- **A solution** $y(x,t)$ of the "wave equation"

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}.$$

  This definition is too narrow, because there are wave equations other than this so-called "classical wave equation". Quantum mechanical waves obey the Schrödinger equation. Soliton waves obey the Korteweg–de Vries (KdV) equation. Even water waves don't obey the classical wave equation exactly: In the classical wave equation all wavelengths travel with the same speed, but for real water waves in the deep ocean waves with long wavelength travel faster than those with short wavelength.

- **Any function of both space and time** $y(x,t)$. Too broad: A current produces a magnetic field that is a function of space. If the current changes with time, then the magnetic field does too. No one calls this a wave. (All waves are functions of both space and time, but not all functions of both space and time are waves.)

- **A function of both space and time** *not* of the product form $f(x)g(t)$. Too narrow: Standing waves are of this form.

I confess that I still don't have a good definition of "wave". This is not unusual: in 1964 US Supreme Court Justice Potter Steward wrote that it was impossible to define "pornography", but that "I know it when I see it". The term "wave" is, in this respect, similar. I don't have a definition but I do have a list of things seen as waves...

Examples of waves

Mechanical waves [waves on a string, on a slinky, on an ocean, in a double bed (husband rolls over, waves on mattress wake up wife)].
Sound waves, seismic waves [varieties of mechanical wave].
Electromagnetic waves [optical light, infrared, radio; ultraviolet, X-ray, gamma-ray].
Quantum mechanical waves.
Traffic waves [cars slow down before reaching an accident].
Population waves [lemmings].
Business waves [business boom begins downtown, spreads to suburbs, by the time it reaches far suburbs, downtown has crashed].
The film "Nonrecurrent Wave Fronts" shows more unconventional waves.

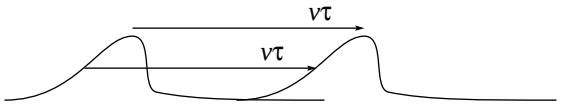
Things that are not waves: A true wave must be a function of both position and time. Some AC generators can produce a potential difference $V(t)$ that is sinusoidal [$\sin(\omega t)$] or triangular or step-like. These are often called "wave generators"

that can produce "sine waves, triangle waves, or square waves" despite the fact that the potential difference is a function of t only, not of x. In medicine, an EEG traces out the potential difference between two points in the brain. This is called a "brain wave" despite the fact that it's a function of time only. Similarly for an EKG, but now it's two points on either side of the heart.

Waves that don't change shape. Send a pulse wave down a long horizontal string. Any piece of the string moves only a little bit up and down, or right and left. But the pulse *shape* moves a long distance. And (to a good approximation) the pulse doesn't change shape as it travels.

Here's a snapshot of such a pulse at some time:

After a time τ, each part of the pulse has moved right by a distance $v\tau$:



If the shape of the initial pulse is $f(x)$, then to find the upward string displacement at time τ, we have to go back left a distance $v\tau$ to find out what the displacement had been at the initial time:

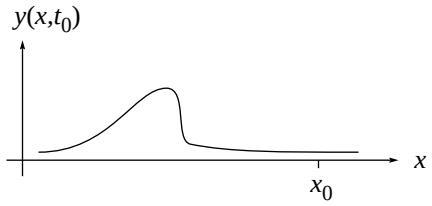
$$\begin{array}{ll} t = 0 & y(x,t) = f(x) \\ t = \tau & y(x,t) = f(x - vt) \end{array}$$

In summary

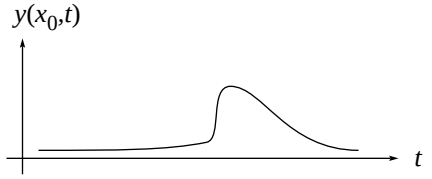
a shape-preserving wave moving right at wave speed v is	$y(x,t) = f(x - vt)$
a shape-preserving wave moving left at wave speed v is	$y(x,t) = f(x + vt)$.

I know this seems wrong — you're used to rightward motion having a + sign and leftward motion having a − sign — but you've just worked it out and it is what it is.

We've produced snapshots showing $y(x,t)$ as a function of x at some given t_0.

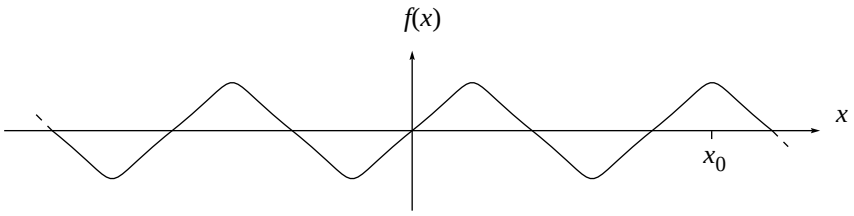


What does $y(x,t)$ look like as a function of t at some given x_0? If you stand at x_0, the wave passes over you with a steep rise and then a gradual fall. The $y(x_0,t)$ curve is the mirror image of the figure above.



Sinusoidal wave on an infinite string. Back on page 3 I disparaged the sinusoidal wave as infinite in space and in time: "starting infinitely far in the past and going for ever and ever, amen." But just as the point particle, the infinite plane of charge, and the infinite solenoid don't exist but can be useful approximations, so the infinite sine wave doesn't exist but can be a useful approximation for a wave that extends over a distance much longer than one wavelength.

Suppose the initial wave happens to be $f(x) = A\sin(kx)$.



This wave is said to have amplitude A and wavelength

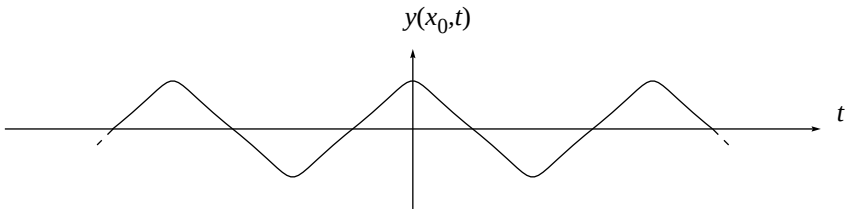
$$\lambda = \frac{2\pi}{k}.$$

(When x increases by an amount λ, kx increases by an amount 2π, and this is one up-and-down cycle of the wave.)

If this is the initial wave that moves to the right, then our general result about initial waves $f(x)$ from page 5 says that

$$\begin{array}{ll} t = 0 & y(x,t) = A\sin(kx) \\ t = \tau & y(x,t) = A\sin(k(x - vt)) \end{array}$$

How does this wave look if we stand at one point (say x_0 in the figure) and watch the wave travel over us?



It looks like a sinusoidal oscillation! One period T of this oscillation corresponds to one wavelength passing over the point x_0. So distance = rate × time becomes

$$\lambda = vT.$$

You will remember that the period T of a sinusoidal oscillation is related to the frequency f and the angular frequency ω through

$$T = \frac{1}{f} = \frac{2\pi}{\omega}.$$

Combining these results gives

$$kv = \omega. \tag{1.1}$$

This relationship is not hard to derive but it's used so often that it's worth memorizing. My former student Afan Ottenheimer thought about this relationship for light, with $v = c$, and memorized it as "$kc = \omega$ . . . Kansas cows eat wheat".

We can now cast the sinusoidal wave into its conventional form,

$$y(x,t) = A\sin(kx - \omega t). \tag{1.2}$$

Power transmitted by a sinusoidal wave. It is clear that waves carry energy (they can undermine the foundations of beach-side condominiums), but how much? It makes sense that for a sinusoidal wave the power transmitted (the "intensity") should increase with amplitude: giant waves deliver more power than ripples. But it's not only that: the wave (5 meters) $\sin(kx - \omega t)$ delivers the same power as the wave $(-5$ meters$)\sin(kx - \omega t)$, so it makes sense that power transmitted should be proportional to A^2.

In addition, a slowly changing wave (say with a period of 12 hours) will deliver less power than a rapidly changing wave (say with a period of 30 seconds). Once again, the result is expected to be independent of the sign of the period, so it makes sense that power transmitted should be proportional to $1/T^2$.

Textbooks[1] prove that our suspicions are correct: the power transmitted by a sinusoidal wave (the "intensity") is proportional to

$$(A/T)^2. \tag{1.3}$$

I'll leave the derivation to the textbooks.

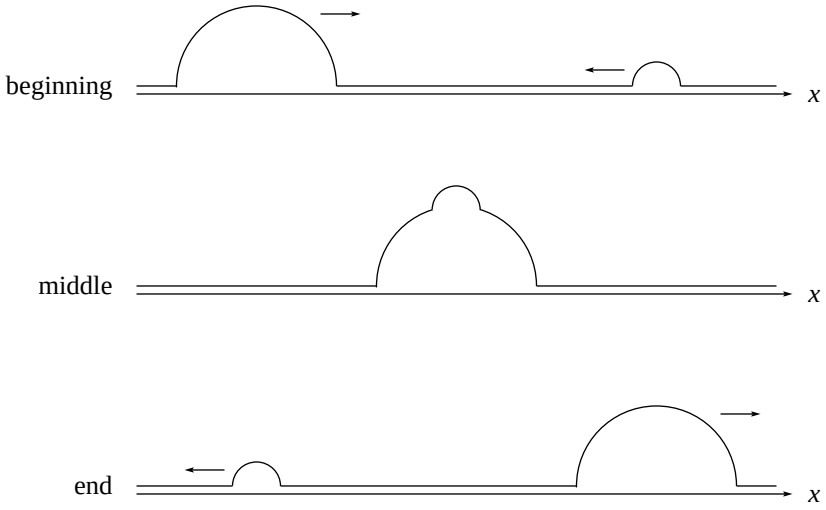
[1] For example Jearl Walker, *Fundamentals of Physics: Halliday & Resnick* (Wiley, tenth edition, 2014) equation (16-33) on page 455.

Chapter 2

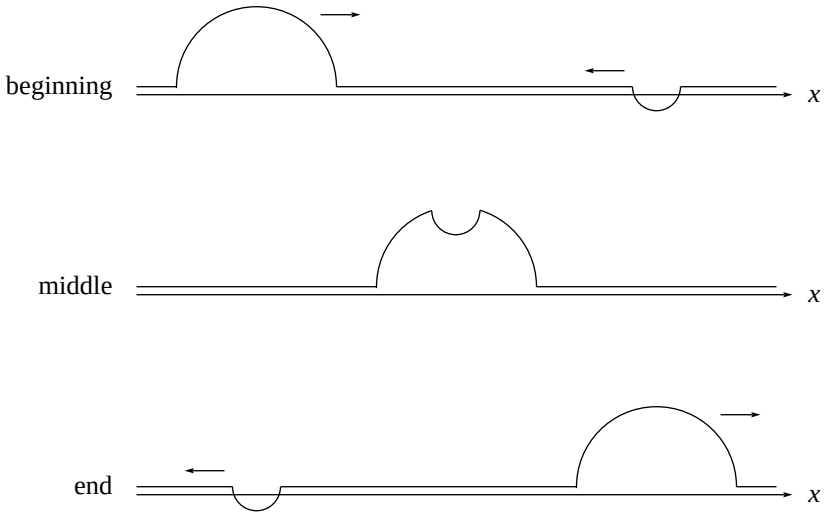
Superposition and Standing Waves

A wave travels to the right: $f(x - vt)$.
Another wave travels to the left: $g(x + vt)$.
The total wave is just the sum of the two waves: $f(x - vt) + g(x + vt)$.
This is called "superposition".
⟦It is not true of all waves, but it's true for the so-called "linear" waves we'll treat in this course.⟧
Check out the videos "When Pulses Collide" and "When Pulses Collide II".

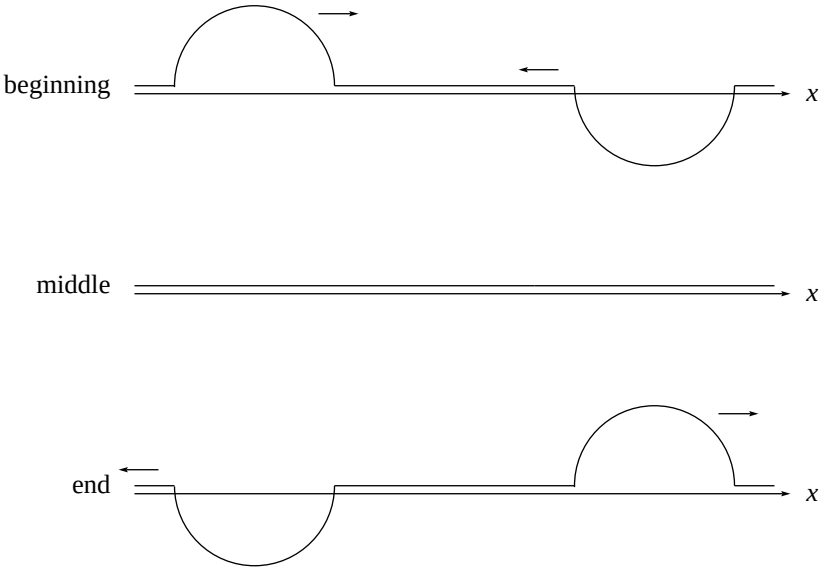
In the example below, a big semicircular wave moves right, a small one moves left. When they cross over each other, the two waves add. Then each continues independently on its own way as if they had never known each other.



The same holds if the small wave moving left happens to have a downward rather than an upward displacement.



What if the two waves are the same size? In this case their displacements cancel out completely as they pass over each other.

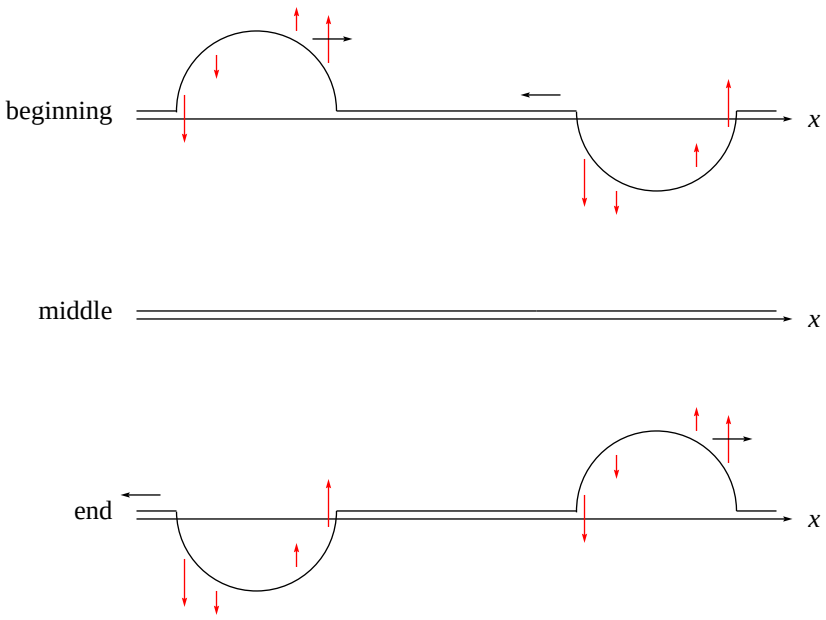


It would be very funny if this were done as a lecture demonstration and you happened to walk into class late just at the time marked "middle". You would see a straight string with no displacement at all, then two semicircular waves would pop into being on the straight string, the "up" wave moving right and the "down" wave moving left! This is a puzzle. How can a straight, unstretched spring just pop two semicircular waves into existence?

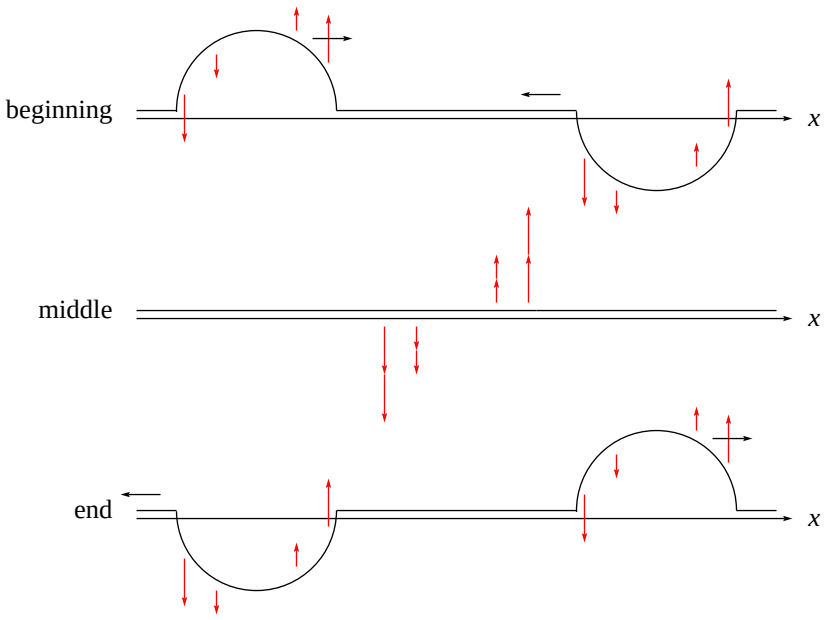
I'll resolve this puzzle on the next page. But think about it for a few moments before turning the page, both to make sure you understand why it's puzzling and to see if you can't resolve the puzzle yourself.

You know from introductory mechanics that to specify the state of a particle you must specify both its position and its velocity. The wave pictures on the previous pages show only the positions of the particles that make up the string, and not their velocities, so they don't specify the state. (There are black velocity arrows, but they signify the velocity of the waveform, not the velocity of the particles on the string.)

Paint a green dot on a single bit of string, and think about the motion of that dot as the rightward moving, "up" wave washes over it. That dot moves first up, then down. When the leftward moving, "down" wave washes over the green dot, it moves first down, then up. For the "beginning" and "end" situations, when the two waves are well-separated, the velocities of representative dots are shown using red arrows in the figure below.



Now think about what happens to a single string element at the "middle" situation. The total displacement of a dot on the string is the sum of the displacement due to the two superposing waves. And the same is true of the velocities. But at the "middle" time, when the string element *displacements* sum they cancel out to zero, while when the string element *velocities* sum they actually increase.



When you walked late into class, you saw the string at an instant, as in a snapshot, and of course a snapshot can't show the velocities. The situation at the middle is indeed a straight, unstretched string, but it's not a straight, unstretched spring at rest. It's the motion of the string (invisible in the snapshot) that enables it to pop two semicircular waves into existence.

Challenge: Can you show that if the waveform is $y(x,t) = f(x - vt)$, and if

$$f'(x) = \frac{df(x)}{dx},$$

then the velocity of the string element at (x,t) is $-vf'(x - vt)$?

Superposition of sine waves. Suppose the two waves superposing are not semicircular pulses, but instead sine waves:

$$\begin{aligned} y_1(x,t) &= A\sin(kx - \omega t) \\ y_2(x,t) &= A\sin(kx + \omega t). \end{aligned}$$

Then the total wave is

$$y(x,t) = y_1(x,t) + y_2(x,t) = A\sin(kx - \omega t) + A\sin(kx + \omega t). \tag{2.1}$$

Okay, so that's the sum, but how can we understand the character of $y(x,t)$? If you knew the trigonometric sum and difference formulas, you might be able to make some progress. But I forgot those formulas the minute I left high school (if not before).

Instead, I like to perform trigonometric manipulations using complex arithmetic and the fact that, for θ real,

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

If you aren't familiar with this fact, see the appendix on "Euler's formula". ⟦The great Swiss mathematician's name is pronounced "Oiler."⟧

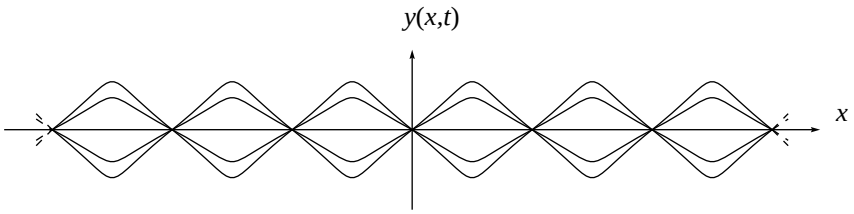
First, establish two consequences of Euler's formula:

$$\sin(\theta) = \Im m\left\{e^{i\theta}\right\} \text{ and } \cos(\theta) = \tfrac{1}{2}(e^{i\theta} + e^{-i\theta}).$$

(The symbol $\Im m\{z\}$ means "imaginary part of z".) Now go at it:

$$\begin{aligned} y(x,t) &= A\sin(kx-\omega t) + A\sin(kx+\omega t) \qquad (2.2)\\ &= A\,\Im m\left\{e^{i(kx-\omega t)} + e^{i(kx+\omega t)}\right\}\\ &= A\,\Im m\left\{e^{ikx}e^{-i\omega t} + e^{ikx}e^{+i\omega t}\right\}\\ &= A\,\Im m\left\{e^{ikx}[e^{-i\omega t} + e^{i\omega t}]\right\}\\ &= A\,\Im m\left\{e^{ikx}[2\cos(\omega t)]\right\}\\ &= 2A\cos(\omega t)\,\Im m\left\{e^{ikx}\right\}\\ &= 2A\cos(\omega t)\sin(kx). \qquad (2.3) \end{aligned}$$

In other words, $y(x,t)$ is just a sine function of x, but with amplitude that varies with time: The amplitude is $2A$ at $t = 0$, then diminishes to 0 at $t = \frac{1}{2}\pi/\omega$, becomes $-2A$ at $t = \pi/\omega$, 0 again at $t = \frac{3}{2}\pi/\omega$, and returns to $2A$ at $t = 2\pi/\omega$.



I don't know about you, but I never would have guessed that this behavior is hidden within the equation (2.2). These are called "standing waves".

The most remarkable thing about this result is that there are points, called nodes, where the two waves, one moving right and one moving left, sum up to no motion at all! Rodolphe Radau[1] wrote of this phenomena in the context of sound waves,

[1] *Wonders of Acoustics; or, The phenomena of sound*, from the French of Rodolphe Radau, the English translated and revised by Robert Ball (New York, Charles Scribner & Co., series: Illustrated Library of Wonders; Marvels of Nature, Science, and Art, 1870) page 212.

saying that "sounds quarrel, fight, and when they are of equal strength destroy one another, and give place to silence." What is the distance between two nodes? It is

$$\tfrac{1}{2}\lambda = \tfrac{1}{2}\frac{v}{f}.$$

You might regard this derivation as so much fiddle-faddle. Where am I going to get an infinitely long string? And once I've gotten it, how can I send two exactly identical waves down it from the two ends? (Especially since the infinitely long string doesn't have ends!) The answer to the second question is that I can mount a piece of string with one end clamped motionless, like a node. Waves will reflect from that clamp and traverse the string in the opposite direction. If the clamp is tight, the reflection will be nearly complete and the two sine waves will be almost exactly identical.

Now I need only a semi-infinite string, which is only half as unrealistic as an infinite string but still unrealistic. But if I clamp another end, that motionless clamped point will also behave like a node. A string of length L between two clamps will support standing waves with n humps between nodes whenever

$$L = n\frac{v}{2f} \quad \text{where} \quad n = 1, 2, 3, \ldots.$$

That is, the string of length L will support standing waves, but not of any frequency: only for frequencies

$$f_n = n\frac{v}{2L} \quad \text{where} \quad n = 1, 2, 3, \ldots. \tag{2.4}$$

To see this phenomenon in action, check out James Dann's video "Standing Waves Part I: Demonstration" at

http://www.youtube.com/watch?v=-gr7KmTOrx0

Challenge: In what sport do standing waves play an essential role? Jump rope.

What about waves in two dimensions? (Say, waves on a drumhead, or on a thin sheet of metal.) The same sort of thing happens, but now the math is more complicated: For example, waves starting at one corner of a square sheet of metal will reflect from the edges opposite that corner. But waves traveling along an edge will arrive back at the starting corner before waves traveling along the diagonal do. This means that the behavior of standing waves on a sheet of metal is richer than the behavior of standing waves on a string. (You could say "more difficult" or you could say "richer". The choice is yours.)

In this context, the sheet of metal is called a "Chladni plate", and this is another favorite demo of mine. Dianna Cowern, who calls herself "Physics Girl", has made a great video of this demo called "Singing plates – Standing Waves on Chladni plates"

http://www.youtube.com/watch?v=wYoxOJDrZzw

(Watch out! At 1:59 she says that a node the string "appears to be not moving at all". There's no "appears" about it... at a node the string is not moving at all.)

What about three dimensions? All musical instruments — violins, organs, drums, even the voice box — work by setting up standing waves and then letting some of that standing wave leak out of instrument so that listeners can hear it. The richness of musical instruments — the sweetness of the violin, the brightness of the oboe, the mellowness of the bassoon, the throatyness of the clarinet, the ethereal eloquence of the Native American flute, and even the expansive range of expression of the human voice — all reflect the richness of standing wave patterns in three dimensions.

Problem

2.1 These notes derive the standing wave product (2.3) from the sum form (2.2) using complex arithmetic. If you don't like complex numbers you might want to do that derivation using trig sum and difference formulas instead. Try it and see which way you think is easier.

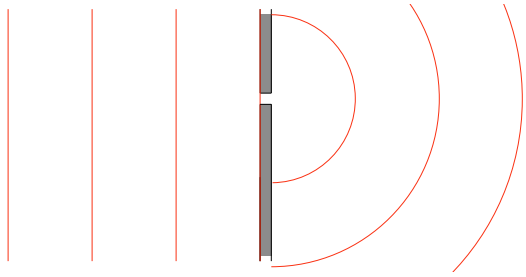
Chapter 3

Two-Slit Interference

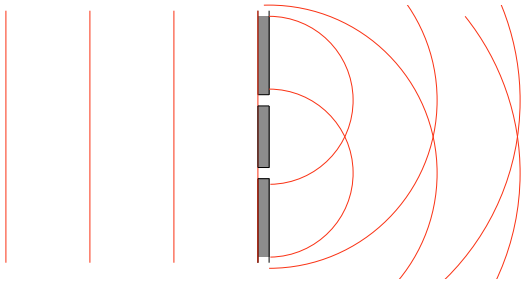
We've seen how to use superposition when two identical waves come from two different directions. I think you realize that this would be an unusual occurrence. A more frequently encountered situation, which exhibits the same fundamental phenomenon, comes when a wave has two different paths from the source to the detector. In this situation the phenomenon is usually called "interference" rather than "superposition", even though it's the same thing. (If there are two or three or a dozen paths from source to detector, the phenomenon is usually called "interference". If there are thousands or an infinite number of paths, it's usually called "diffraction". But the reason I say "usually" is that you'll find violations of this usage rule.)

Here's the setup: Suppose a sinusoidal wave of any type (sound wave, water wave, light wave) approaches an absorbing wall with a tiny hole. (What is tiny? Hole diameter much smaller than the wave's wavelength.) What happens? If light were a ray moving in straight lines then a tiny bright spot would appear on a distant screen. The correct answer, realizing that light is a wave, is that the hole acts as a source of spherical waves, so the distant screen is bright all over. (See figure on next page. The orange lines represent wave crests, and this figure is a snapshot. If it were a movie, then as time went on the wave crests would move to the right.) This answer is not obvious but it goes to the heart of what we mean by "wave". It is called the Huygens construction.

⟦One can derive the Huygens construction from the fundamental principles underlying the particular wave in question — the fluid flow equations for sound or water waves, the Maxwell equations for light waves — but this derivation is both hairy and unilluminating. It is better at this point for you to accept Huygens's construction phenomenologically, just as Christiaan Huygens did in 1678. The Huygens construction was used in this way until 1818 when Augustin-Jean Fresnel gave an explanation from fundamental principles. Fresnel's explanation contained a minor flaw cleared up by David A.B. Miller in 1991.⟧



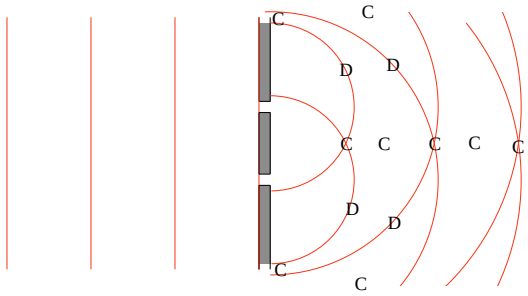
This is interesting (and it's unexpected if you're used to light traveling in straight lines) but it's not an example of interference, which requires two (or more) paths from the source to the detector. So, what happens if there are two identical holes? The answer in ray optics is obvious: on the detector screen are two bright spots. And if the holes are much larger than a wavelength then that's what happens, to high accuracy. But what happens if there are two identical tiny holes? On the right we get the superposition of the spherical waves from one hole and the spherical waves from the second.



There's an important difference between this sketch and the one on the previous page. On the previous page the orange lines represented wave crests. On this page the orange lines on the left represent wave crests, but the orange arcs on the right represent where the wave crests would be if there were only one hole. When there are two holes the total wave is the superposition (the sum) of the wave due to the top hole plus the wave due to the bottom hole. If the wave crest from one hole falls on top of the wave crest from the other, then the two waves superpose to make a very high crest. If the wave trough from one hole falls on top of the wave trough

from the other, then the two waves superpose to make a very deep trough. But if the wave crest from one hole falls on top of the wave trough from the other, then the two waves superpose to make nothing. (The two sound waves "destroy one another, and give place to silence". Or, for the case of light, the two light waves add up to give darkness.)

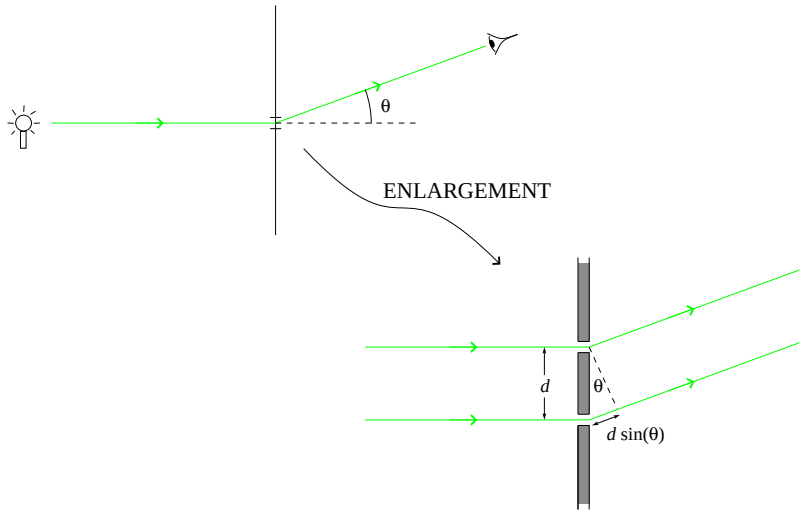
The first two examples are called "constructive interference", the third is called "destructive interference". [The word "interference" has a different meaning in physics and in football. Football players never (or at least never deliberately) run "constructive interference". Although I imagine an episode in which a football coach tells a player to "go out on the field and run interference." The player goes out and scores a goal for the opposing team. When the coach screams "I told you to run interference!" the player replies "I did, coach! I ran *constructive* interference."] The sketch below is the same as the one above, except that I have inserted a C at points of constructive interference, a D at points of destructive interference.



To find which points exhibit constructive interference and which destructive interference, it's a simple matter of finding the distance to each hole. If the two paths differ by an integer number of wavelengths (including the integer 0), then the interference is constructive. If they differ by a half-half integer number of wavelengths, then the interference is destructive. If they differ by anything else, the interference is partially destructive.

It's surprisingly difficult to execute this scheme at an arbitrary point to the right of the two holes (called "the Fresnel case"). For now we consider the "Fraunhofer limit" when the detector is very far from the two holes.

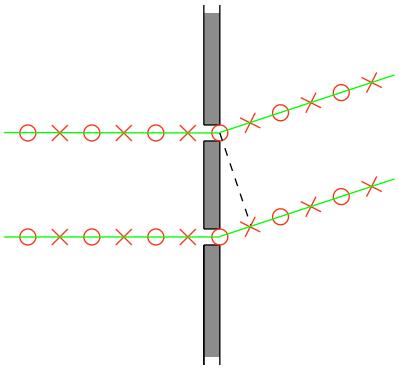
This sketch shows the geometry for the Fraunhofer limit.



The upper left sketch shows an overview of the setup. A distant source (in this case a light bulb) sends light to the two slits (so close together that they can't be resolved at this scale). At some angle θ away is a distant detector (in this case an eye).

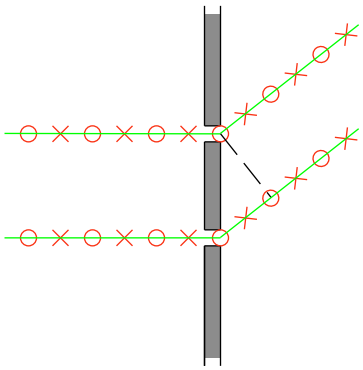
To see that there are actually two slits you have to enlarge the image considerably — this is done in the lower right. At this enlarged scale you can make out the two slits, separated by a distance d; you can also see that there are actually two paths from source to detector. (On the scale of the upper left these paths were so close that they appeared as a single line.) The source and detector are so far away that, at this scale, these two paths are virtually parallel. A little geometry will convince you that the angle θ shown in the lower right is the same as the angle θ shown in the upper left, and that the the leg of the right triangle sketched out has length $d\sin(\theta)$.

Consider first the case where the detector (eye) is situated at the angle shown here:



Light spreads out in all directions from the two tiny holes, but I'm only interested in light going toward the detector, so I show only the electric field along those two green lines. Electric field pointing out of the page ("crest") is represented by a circle, electric field pointing into the page ("trough") is represented by a cross. This sketch is a snapshot: as time goes on the crests and troughs move right along the green lines. The detector in this picture is positioned so that the "extra length" $d\sin(\theta)$ along the bottom path is exactly half a wavelength: $d\sin(\theta) = \frac{1}{2}\lambda$. To the right, the crests on the upper green line always match up with troughs on the lower green line. When the light from these paths comes together far away at the detector, there will be complete destructive interference: The light from the top path will add up with the light from the bottom path to produce darkness.

I can draw the same situation but with the detector at a different angle:



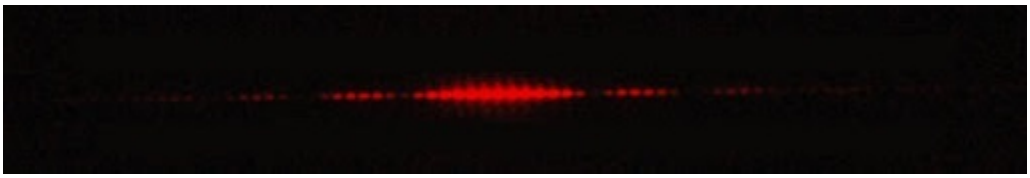
In this case $d\sin(\theta) = \lambda$. To the right, the crests on the upper green line always match up with crests on the lower green line, while troughs match up with troughs.

When the light from these paths comes together far away at the detector, there will be complete constructive interference, so the detector here will experience brightness.

If the detector were moved still further upward, so that $d\sin(\theta) = \frac{3}{2}\lambda$, we would again find the darkness of complete destructive interference. In general (where $m = 0, \pm1, \pm2, \ldots$):

$$\begin{array}{lll} \text{complete destructive interference} & \text{dark} & d\sin(\theta) = (m + \frac{1}{2})\lambda \\ \text{complete constructive interference} & \text{bright} & d\sin(\theta) = m\lambda \end{array} \tag{3.1}$$

It seems absurd in the extreme to suggest that light plus light can add up to darkness, so let's do the experiment. Shine a red laser through two tiny holes and look at the result on a distant screen. If light moved in straight lines, the result would be two tiny bright spots. Here's what really happens, absurd or not:



One last thing: Although I've written about "two tiny holes", these are actually hard to produce. Much easier to make are "two thin slits". So this phenomenon is more often called "two-slit interference" than "two-hole interference".

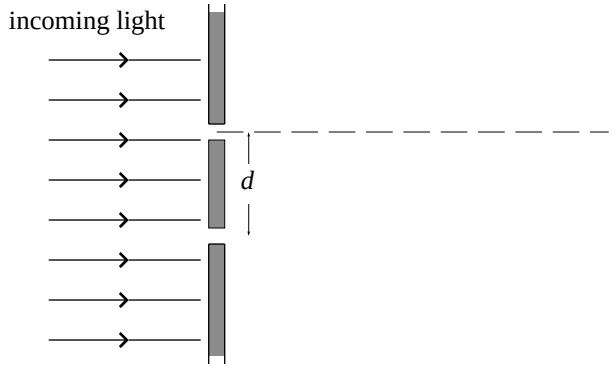
Problems

3.1 In his 1963 book *Strength to Love*, Dr. Martin Luther King, Jr. wrote that:

> Darkness cannot drive out darkness; only light can do that. Hate cannot drive out hate; only love can do that.

We have seen in this chapter that light plus light might add up to darkness. Does this invalidate the first part of Dr. King's metaphor?

3.2 *Fresnel interference.* Two narrow slits, illuminated by light of wavelength λ, are separated by distance $d = 3.00\lambda$. Consider the light intensity (or "brightness") along a line directly behind the top slit. (The dashed line in the figure below.) How far from the top slit is the farthest point of completely destructive interference?

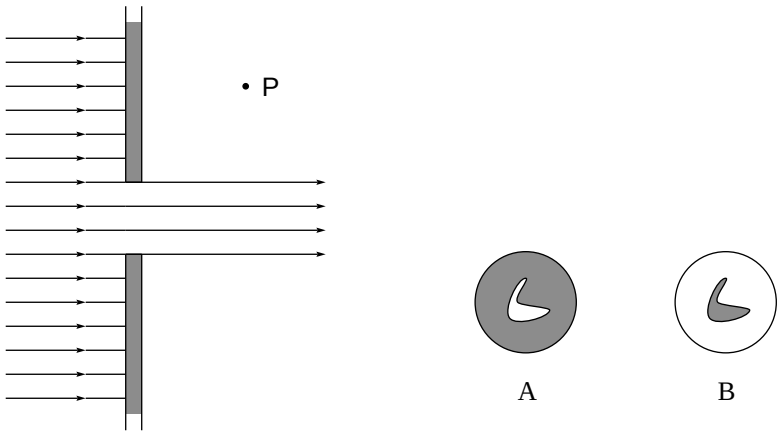


⟦*Problem-solving tip:* Some might try to solve this problem by searching through the text for a suitable "formula to plug into". They would encounter equation (3.1) for the location of dark spots and then be perplexed: "What angle should I use for θ?" This perplexity reflects the fact that equation (3.1) *doesn't apply* to this situation. (Why not?)

Do *not* thumb through your text looking for the silver bullet equation that will solve your problem. That would be like writing an essay for your literature course by thumbing through a dictionary hoping that you'll hit upon the right word. Instead of shoehorning the problem into the Procrustean bed of equations in your text, let the problem speak to you itself. (I have mixed a lot of metaphors in this paragraph, but that mixture illustrates just how very flawed the "find the right equation" approach to problem solving is.)

This backwards approach to problem solving sometimes comes up in the regime of public policy: "I'm a conservative. The tools in my policy toolkit cannot solve the problems of climate change. Hence I conclude that climate change doesn't exist." This makes for bad public policy, and for bad physics as well.⟧

3.3 *Babinet's principle.* Light falls on a barrier which contains a big hole — so much bigger than the wavelength of the light that diffraction effects are negligible and ray optics applies. Thus there's brightness behind the hole but darkness everywhere else behind the barrier. Select some point P within the shadow.



An obstacle such as A is placed over the hole. Because obstacle A has thin pieces, it diffracts light into the shadow, and there is now some light intensity at point P.

Obstacle B is the photographic negative of obstacle A: it blocks light where A passes light and passes light where A blocks light. Because obstacle B also has thin pieces, it too diffracts light into the shadow, and again there is some light intensity at point P.

Because obstacles A and B are complete opposites, you might expect that their diffraction patterns would be opposites also. This expectation is completely wrong: Show that the light intensity at point P is exactly the same whether obstacle A or obstacle B is used.

Clues: (a) Huygens's construction says that when the hole is unobstructed it acts as an infinite number of radiators, each sending waves in all directions, including the direction toward point P. Given that point P receives light from an infinite number of sources, why is it dark there? (b) Use superposition. (c) The solution to this problem is much shorter than its statement.

> Jacques Babinet (1794–1872) was French, so his name is pronounced "Babi-nay." His parents wanted him to become a magistrate, but instead he became a physics professor at age 26 and a member of the prestigious *Académie des Sciences* at age 46. He was an early proponent (after Young and Fresnel) of the wave theory of light, and the first scientist to use diffraction gratings for spectroscopy. He experimented on optical effects in min-

eralogy and meteorology (rainbows, coronas, and the polarization of skylight), and invented a goniometer and the "Babinet compensator," which is still used today to produce and analyze polarized light. He achieved considerable fame as a popularizer of science.

Chapter 4

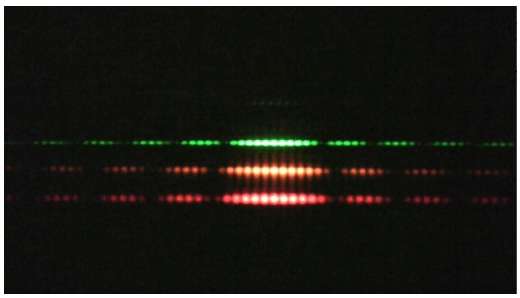
Interference Topics

4.1 Using interference for length measurements

When light of wavelength λ passes between through two very thin slits with separation d, the bright spots (so-called "interference maxima") fall at angle θ where

$$d\sin(\theta) = m\lambda \quad \text{where} \quad m = 0, \pm 1, \pm 2, \ldots . \tag{4.1}$$

When Thomas Young performed the first two-slit interference experiment in 1803, he measured d and θ, and used them to calculate λ. You can see the wavelength effect in the experiment below: light with a shorter wavelength (green) spreads out through a smaller angle.



Today it's more usually done the other way around: knowing the wavelength and the angle, we can calculate a distance. Most of the distances I'd like to measure don't happen to fall between two slits, so this basic idea has been modified in numerous ways to make numerous measurement instruments. The whole field of using the wavelength of light to measure distances is called "interferometry" — we will encounter examples later.

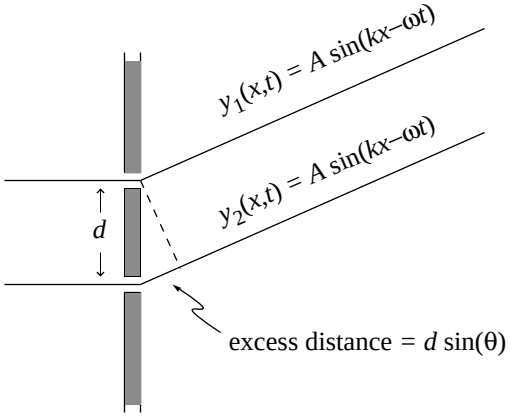
For example, when light passes through a vacuum it has speed c and wavelength λ. When that same light enters a transparent medium with "index of refraction n", it has a slower speed c/n and a shorter wavelength λ/n. Problem 4.1 converts this observation into a tool for measuring the thickness of a flake of mica.

From 1889 to 1960, the meter was defined as the distance between two scratches on a platinum-iridium bar located in Sèvres, France. But from 1960 to 1983 it was defined as 1 650 763.73 wavelenghts of the orange-red light produced by glowing krypton. (Since 1983 the meter has been defined as the distance traveled by light in vacuum in 1/299 792 458 second.)

4.2 Intensity in two-slit interference

We know the location of the interference maxima and zeros, but what about intermediate angles?

Setup. A wave of wavelength λ (wave number $k = 2\pi/\lambda$, angular frequency $\omega = kv$) passes through two narrow slits located a distance d apart:



Place the detector a long distance L from the top slit ("Fraunhofer limit"). Then the detector is a distance $L + d\sin\theta$ from the bottom slit. Hence the wave signal at the detector due to the top slit is

$$y_1(x,t) = A\sin(kL - \omega t),$$

while the wave signal at the detector due to the bottom slit is

$$y_2(x,t) = A\sin(k(L + d\sin\theta) - \omega t).$$

The total wave signal received is

$$y(t) = A\sin(kL - \omega t) + A\sin(kL + kd\sin\theta - \omega t). \tag{4.2}$$

Clearly, this total signal is a function periodic in time with angular frequency ω. We wish to put it into the form

$$y(t) = [\text{amplitude}]\sin([\text{phase}] - \omega t).$$

Once it's in this form, equation (1.3) says that the intensity (or "brightness") is proportional to [amplitude]^2.

Math. To make this algebra easier, define

$$\phi = kd\sin\theta$$

and use Euler's relation

$$e^{it} = \cos t + i\sin t.$$

In these terms,

$$\begin{aligned} y(t) &= A\sin(kL - \omega t) + A\sin(kL + kd\sin\theta - \omega t) \\ &= A\,\Im m\left\{e^{i(kL-\omega t)} + e^{i(kL+\phi-\omega t)}\right\} \\ &= A\,\Im m\left\{e^{i(kL+\phi/2-\phi/2-\omega t)} + e^{i(kL+\phi/2+\phi/2-\omega t)}\right\} \\ &= A\,\Im m\left\{e^{i(kL+\phi/2-\omega t)}[e^{-i\phi/2} + e^{+i\phi/2}]\right\} \\ &= A\,\Im m\left\{e^{i(kL+\phi/2-\omega t)}[2\cos(\phi/2)]\right\} \\ &= 2A\cos(\phi/2)\sin(kL + \phi/2 - \omega t). \end{aligned} \tag{4.3}$$

In terms of the form above,

$$[\text{amplitude}] = 2A\cos(\phi/2).$$

The intensity of this signal is proportional to the amplitude squared. If we define the intensity at $\theta = 0$ to be I_m ("Intensity at the middle" or, as it turns out, "Intensity at the maximum"), then

$$\text{intensity} = I_m\cos^2\left(\frac{\phi}{2}\right) \qquad \text{where} \qquad \phi = \frac{2\pi d}{\lambda}\sin\theta. \tag{4.4}$$

Challenge: Can you show that this intensity function has maxima and zeros as already demonstrated at equation (3.1)?

Our formula passes this test, but there is a problem. The formula predicts that all the intensity maxima are equally bright. A glance at the experimental result illustrated on page 22 shows that this prediction is *not* correct. The problem is that the experimental slits are not infinitely narrow. At equation (7.1) we will rectify this defect.

4.3 Coherence

In the two-slit interference experiment, our source of interfering light was two narrow slits. Would you get the same result using two narrow lightbulbs, each the same shape and width as the slit? No, because of an effect called "coherence".

Suppose you build a picket fence with pickets one inch wide and with two inches between pickets. If you start at the left edge of one picket, and move six inches right, you'll be at the left edge of another picket. But if you move 6000 inches right, then chances are you will *not* be at the left edge of a picket. Sure, if the pickets were all *exactly* one inch wide and *exactly* two inches apart, then after moving 6000 inches you'd be at the left edge of a picket. But tiny imperfections entered when you built the fence. If you move only six inches those imperfections are negligible. But if you move 6000 inches they add up and generate a mismatch.

The same thing holds for light. In, say, a sodium lamp, the source of light is trillions of radiating atoms, each one of which glows for about 10^{-9} second, and then turns off. (At this point in your education, you can't derive that value 10^{-9} second — you'll just have to take my word for it.[1]) Each glow event sends out about 30 centimeters of light.

Graph the electric field of a light beam with wavelength 600 nm as a function of position for a given instant. The electric field at one point is the sum of the electric fields generated by those trillion atomic glows. If you walk 0.3 centimeters down the graph, about one percent of those glow events have stopped and been replaced by new glow events, but 99% of the light is from the same glow events. But if you walk 30 centimeters down the graph all of the atomic glowers have been replaced by other glowers.

If you start at a wave crest and move down by 5000 wavelengths, you'll move a distance of 0.3 cm (600 nm $\times$ 5000 = 0.3 cm), and you'll land on top of another wave crest. But if you do this 100 times to move a distance of 30 centimeters, chances are you *won't* end up on a wave crest, because of the small errors inserted as one glow event turns off and another turns on.

The distance where you cross over from being "pretty sure you'll end up on another wave crest" to being "pretty sure you won't" is called the "coherence length". (The exact length will depend on the exact accuracy demanded of "landing on top of another wave crest".)

The coherence length depends on the source of light: A sodium lamp has a coherence length of about 0.5 mm (shorter than the 30 cm mentioned above because of collisions between sodium atoms as they radiate). The familiar red light from a

[1] David J. Griffiths, *Introduction to Quantum Mechanics*, second edition (Pearson, Upper Saddle River, NJ, 2005) section 9.3.2, "The Lifetime of an Excited State".

helium-neon laser has a coherence length of about 20 cm. Semiconductor lasers emit light with coherence lengths up to 100 m.

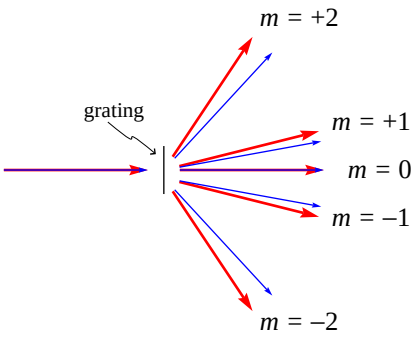
Going back to the thought experiment at the beginning of this section, if you had two lamps rather than two slits there would indeed be the bright and dark bands of an interference pattern, but that pattern would last for only about 10^{-9} seconds. Then it would shift to a different interference pattern, then shift again and again. Our eyes are not quick enough to respond to these jumping patterns, and we see only an average. The two light sources are said to be "incoherent".

4.4 Gratings

What happens if you have not two slits but three? Call the slits A, B, and C, with distance d between adjacent slits. Position your detector at an angle θ where slits A and B would produce an interference maximum, that is, where the light going through slit A interferes constructively with the light going through B: crest atop crest, trough atop trough. Well then the light going through B and C also interferes constructively. And if A and B interfere constructively, and B and C interfere constructively, then A and C must interfere constructively.

In short, the criterion for a maximum in three-slit interference is the same as the criterion for a maximum in two-slit interference. Similarly for four slits, or five, or five thousand. A collection of a vast number of slits is called a "grating".

Gratings are useful for separating colors. We've already seen that short wavelengths (like blue) are spread less than long wavelengths (like red). If you have blue and red light mixed together, a grating can separate them.



These days you are more likely to see white light split into colors using a grating than with a prism.

4.5 Geometrical optics limit

From everyday life, we are used to the idea that light travels in straight lines. This is called "ray optics" or "geometrical optics". But in reality, as demonstrated through interference experiments, light has a wave character. This is called "wave optics" or "physical optics". Aren't these two positions contradictory?

The resolution to this apparent paradox is that wave optics is correct, but that geometrical optics is a good approximation to wave optics in situations where the wavelength is much shorter than the features with which the light interacts. In everyday life light interacts with windows and eyeglasses and so forth, things very large compared to a wavelength of light (400–700 nm). So in everyday life geometrical optics is usually satisfactory for visible light and we say, to high accuracy, that "light travels in straight lines". But FM radio signals have wavelengths of 10–100 meters, and in this case the geometrical optics approximation is usually poor.

Problems

4.1 *Interference with a mica mask.* When light passes through a vacuum it has speed c and wavelength λ. When that same light enters a transparent medium with "index of refraction n", it has a slower speed c/n and a shorter wavelength λ/n. A double-slit interference apparatus is set up and illuminated with light of wavelength $\lambda = 551$ nm (green). A thin flake of mica ($n = 1.58$) is then inserted behind one of the slits. Upon inserting, the seventh bright side maximum ($m = 7$) moves to the very center of the viewing screen. How thick is the mica flake?

4.2 These notes derive the wave (4.3) from the sum (4.2) using complex arithmetic. If you don't like complex numbers you might want to do that derivation using trig sum and difference formulas instead. Try it and see which way you think is easier.

Chapter 5

Interference from Thin Films

All of us are fascinated by the beautiful colors of soap films. But some of us are more eloquent in describing them. In his book *The Deltoid Pumpkin Seed*, John McPhee describes a model airplane competition in which the object is not to fly fast, but to stay in the air for as long as possible. Such model aircraft must be as lightweight as possible, so they are sheathed with a thin, transparent film.

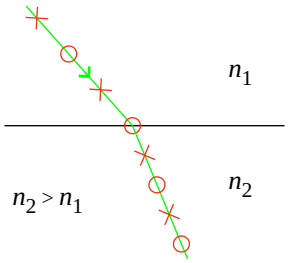
> The film was so thin—one-tenth the thickness of Saran Wrap—that light could not pass through it in the way that light ordinarily goes through a transparent substance. Instead, it refracted, reflected, caromed wildly, and split itself into all the colors of the spectrum in shimmering iridescence. When these airplanes flew, they were fantastically beautiful, slowly circling, climbing, spraying color in all directions. They flew, most notably, in Hangar No. 1 at Lakehurst—a giant barn a thousand feet long, almost two hundred feet high, steel-structured, sheathed in wood. This had been the hangar of the Hindenburg, which had burned just outside. Hangar No. 1 had been built for the big rigid airships, and now, in their continuing absence, the all-day twilight of the hangar was sometimes weirdly alive with eight, nine, or even ten almost invisible airplanes climbing slowly toward the roof, each barely heavier than air.

The rainbow patterns produced by oil slicks are particularly wondrous. Motor oil is dull brown. Put a drop on a rain puddle, and it spreads into colorful shimmering iridescence. How can such an ugly substance become so beautiful?

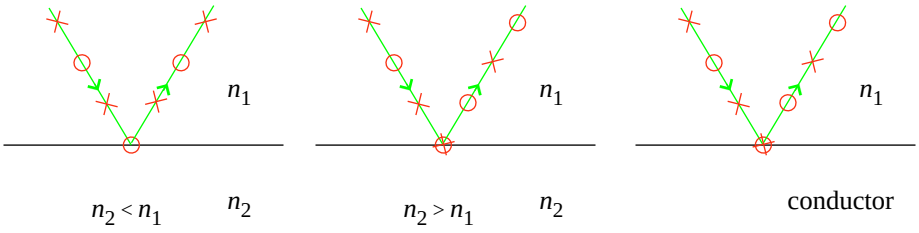
Here are three facts about light passing through a medium (that is, a transparent material like glass or water or air rather than through vacuum). All three can be derived from the Maxwell equations, but in fact all three were discovered experimentally before Maxwell was even born.

1. The speed of light through a medium with index of refraction n is c/n.

2. Transmission through interface. If light of wavelength λ in vacuum enters a medium with index of refraction n, the wavelength in the medium is λ/n.



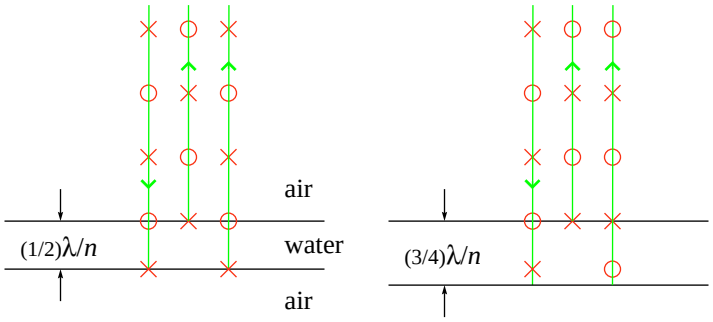
3. Reflection from interface. Reflection is accompanied by a 180° phase change if $n_2 > n_1$ or if the reflecting surface is conducting.



A few words about fact 3. I would have thought that crests and troughs would just continue as always through a reflection. And that is what happens when $n_2 < n_1$. But when $n_2 > n_1$ there's a 180° phase change: when a crest reaches the surface, a trough is reflected; when a trough reaches the surface, a crest is reflected. I can't give you an explanation for this fact, but I can give you an analogy: When light is reflected from a conductive surface (like the silver backing of a mirror) the conductor has zero electric field along the surface. That means an incoming crest has to be canceled by an outgoing trough, and so forth. I can also give you a mnemonic for this strange behavior, in the form of a silly poem:

high to low, phase change no
low to high, phase change pi

Suppose light of a single wavelength shines straight down upon a film of water immersed in air (side view below). The light shines straight down and reflects straight up: some reflects from the top surface, some from the bottom. If I drew it that way, the two upward lights would be directly on top of the downward light, and you wouldn't be able to see what was going up and what was going down. So I'll draw the light reflected from the top surface displaced a bit to the right, and the light reflected from the bottom surface displaced a bit more to the right.



Air has index of refraction $n = 1.00$, water has $n = 1.33$, so: (1) The wavelength of the light in water is *shorter* than the wavelength in air (λ/n). (2) Reflection from the top surface is "low to high, phase change pi" while reflection from the bottom is "high to low, phase change no".

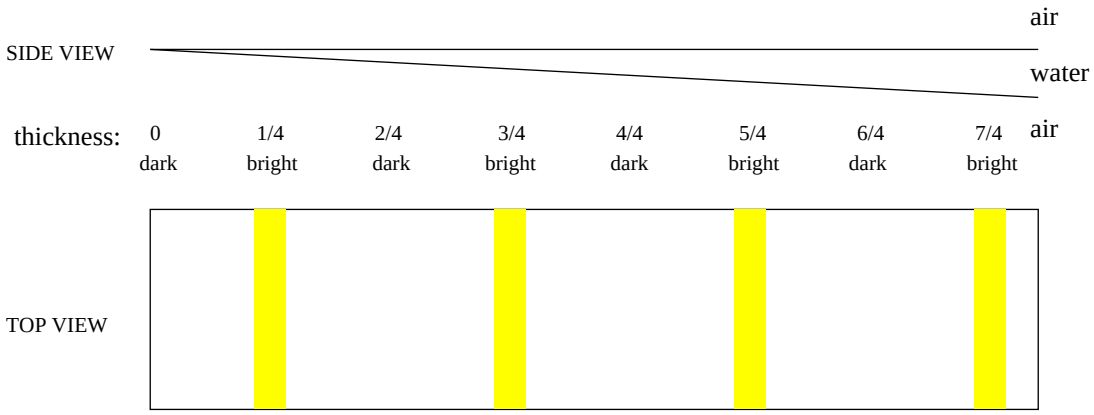
First suppose the film happens to have thickness half a wavelength. The light reflected from the bottom surface has one more wavelength than the light reflected from the top surface, because it traverses a longer length. But the light reflected from the top surface has the phase change, as if half a wavelength had been inserted upon reflection. Examine the two upward light paths: crests arrive on top of troughs — complete destructive interference. If you look down on a film of this thickness, you will see darkness.

What happens if the film happens to have thickness three-quarters of a wavelength? The light reflected from the bottom surface has 1.5 more wavelengths than the light reflected from the top surface, because it traverses a longer length. But the light reflected from the top surface still has the effective 0.5 wavelength inserted. Now crests arrive on top of crests — complete constructive interference. If you look down on a film of this thickness, you will see brightness.

I hope the story is now clear: if you increase (or decrease) the film thickness by a quarter wavelength, you will insert (or remove) half a wavelength in the path traversed by the bottom reflection, but make no change in the top reflection. Such an increase (or decrease) will change the interference from destructive to constructive, or vice versa.

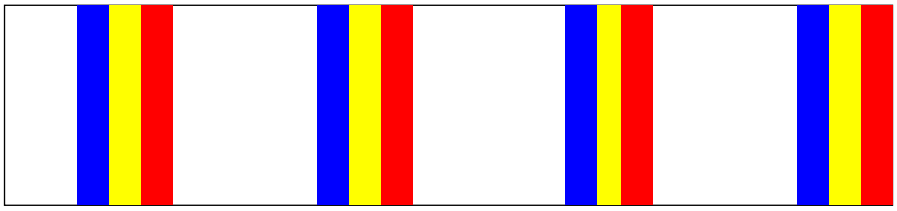
As always, interference involves two paths from source to detector. In this case, the two paths are "reflection from the top surface" and "reflection from the bottom surface".

The figure below shows what happens if the water film varies in thickness. (On the scale of a wavelength of light, the water film is still virtually flat, so the above analysis holds.) If you look down on such a film wedge illuminated vertically with yellow light, you will see bright yellow bands separated by dark spaces. (The figure shows thickness in terms of the number of water wavelengths of yellow light.)



(The yellow bands decrease in brightness gradually as the wedge width changes, but my drawing program shows blocks of color far better than gradations of brightness, so this figure incorrectly draws the bright bands with sharp edges.)

Finally, what if the wedge were illuminated not with yellow light, but with white light, composed of all colors? The short (blue) wavelengths will be bright when the wedge is less thick, and the long (red) wavelengths will be bright when the wedge is thicker. So the bright spots will bring out all the colors of the rainbow. (Once again, this figure incorrectly shows sharp edges to the colors.)



Soap bubbles and oil films are of course not uniform in thickness. We have explained their colorful iridescence.

"Harvard Natural Sciences Lecture Demonstrations" has a great video of this demo at

http://www.youtube.com/watch?v=4I34jA1fDp4

Note that the upper part of the circle — the thinnest part of the wedge — is dark, in accord with our wedge analysis.

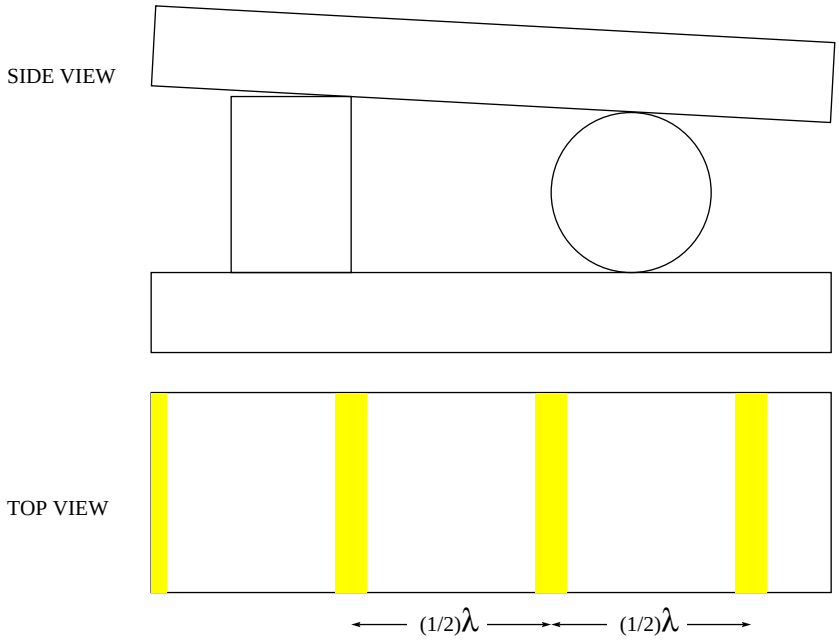
Interferometry. So we've explained the beautiful colors of soap bubbles, and we understand why a blob of oil is ugly but a thin film of oil is beautiful. But in the process we've uncovered something else: a measurement tool of extraordinary accuracy. Just by counting out the dark and yellow bands in the wedge above, we can measure the thickness of the water film to the accuracy of a quarter of the wavelength of yellow light in water — about 100 nanometers.[1]

I would think that to measure something with such extraordinary accuracy, you'd need an expensive, delicate apparatus operated by highly skilled personnel. No! All you need to do is count yellow bands! This measurement technique is called "interferometry".

The trouble is, the technique as described so far measures the thickness of a water wedge. What if I want to measure something else? Say I run a ball-bearing factory, and I want to produce metal spheres that are 1 cm in diameter, plus or minus 600 nm. That is, I want spheres of diameter between 10 000 600 nm and 9 999 400 nm.

I buy a standard block 1 cm tall, and two very flat plates of glass. (Such blocks, invented by the Swedish machinist Carl Edvard Johansson, are called Johansson gauges or "Jo blocks". The plates are called "optical flats".) I test a sphere by putting both the block and the sphere between the two glass plates, and illuminating them from above with yellow light. (Figure on next page.) When I look straight down on the top glass plate, I see yellow bands. The analysis above assures me that over the distance between two yellow bands, the top glass plate has sloped down (or up!) by half a wavelength. There are less than two yellow bands between the block and the sphere, so the sphere is shorter than the block, but by less than one wavelength of yellow light.

[1] Wavelength of yellow light in vacuum: 600 nm. Wavelength of yellow light in water: 450 nm. A quarter of that wavelength: about 100 nm.



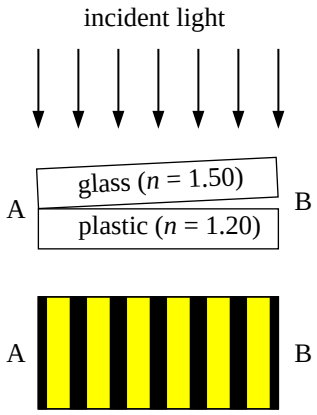
(When I was a boy, you could drop out of high school and get a well-paid although boring job counting bands of yellow light. Now such jobs are done by photodetectors instead of people.)

How do the manufacturers know that the Jo block is exactly 1 cm tall? How do they know that the glass plates are so very flat? They use interferometry!

Problems

5.1 *Oil slick.* A disabled tanker leaks kerosene ($n = 1.45$) into the Persian Gulf, creating a large slick on top of the water ($n = 1.33$). In one region this slick is 461 nm thick. At a certain hour the sun is directly overhead. (a) You fly over the Gulf at this hour and look directly down at the slick. For which wavelength(s) of visible light is the reflection brightest because of constructive interference? (b) At the same time, your friend is scuba diving below the slick. For which wavelength(s) of visible light is the transmission intensity strongest?

5.2 *Interferometry: Using light waves as a ruler.* A perfectly flat piece of glass ($n = 1.50$) is placed over a perfectly flat piece of plastic ($n = 1.20$) as shown below. They touch only at point A. Yellow light of wavelength 600 nm shines down from above. Dark bands in the reflected light are present as shown in the sketch. (a) How thick is the gap between glass and plastic at its widest point B? (b) Water ($n = 1.33$) seeps into the gap. How many dark bands are now present? (The straightness and equal spacing of the bands are accurate tests of the flatness of the glass and plastic.)



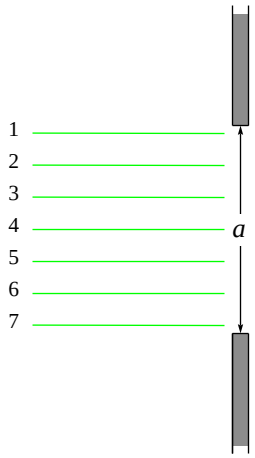
Chapter 6

Single-Slit Diffraction

Okay, we've been talking about "infinitely thin slits" for too long. What happens when light passes through a finite slit?

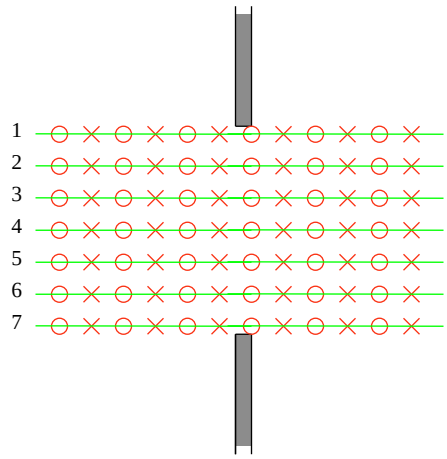
In this case we have not one or two or even thousands of paths from the source to the detector, but an infinite number. So we will have to perform not sums, but integrals. We are obviously striking out into dangerous territory, and I'm going to begin cautiously.

Setup. I'll call the slit width a. Consider that "Fraunhofer limit" in which both the source and the detector are very far from the slit. It would take a long time to draw an infinite number of paths from source to detector, so I'll draw seven and leave the rest to our imaginations.



Path number 4 goes exactly through the middle of the slit... the distance from the top of the slit to path number 4 is $a/2$. I show only the path from the source to the slit... later I'll position the detector and show the paths from the slit to the detector.

Detector at angle $\theta = 0$. If the detector is positioned directly behind the slit, then all seven paths interfere constructively. If you drew in more paths — say one between every pair of green lines shown here — then they would interfere constructively as well.

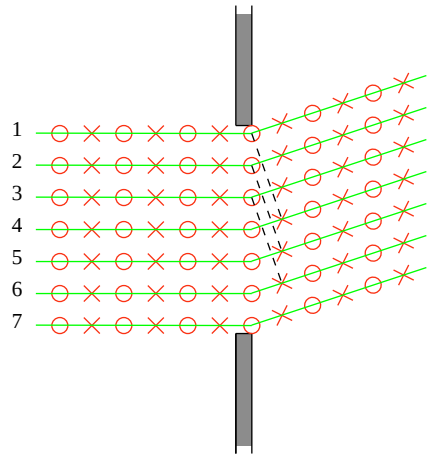


At angle $\theta = 0$ all the infinite number of paths interfere constructively, so the light there will be bright.

Detector at a special angle $\theta > 0$. Next position the detector so that the extra length of the middle path (path number 4) is exactly half a wavelength:

$$\frac{a}{2}\sin(\theta) = \frac{\lambda}{2}.$$

Now the diagram looks like this



You can see that at this angle paths 1 and 4 cancel each other out: the crests from 1 arrive at the same time as the troughs from 4, so there is destructive interference. Similarly, paths 2 and 5 cancel out; paths 3 and 6 cancel out. Only the light from path 7 reaches the detector.

But what if we had drawn more paths? Say we draw in six more paths: one between 1 and 2 (call it 1a), one between 2 and 3 (call it 2a), ..., one between 6 and 7 (call it 6a). Can you see that path 1a cancels out path 4a? Similarly 2a cancels 5a, while 3a cancels 6a. Although we've drawn six more paths, no more light reaches the detector because all six of these paths pair up and kill off. Still only the light from path 7 reaches the detector.

We can draw still more paths, but every path we draw will be canceled out by another. The only reason we said "only the light from path 7 reaches the detector" was that we had considered a finite number of paths. Once you realize that there are an infinite number of paths, you realize that every path passing through the top half of the slit interferes destructively with a path passing through a distance $a/2$ below it, so we have complete destructive interference when

$$a\sin(\theta) = \lambda.$$

Window width that shines darkness. Further: we've show that a slit with the special width

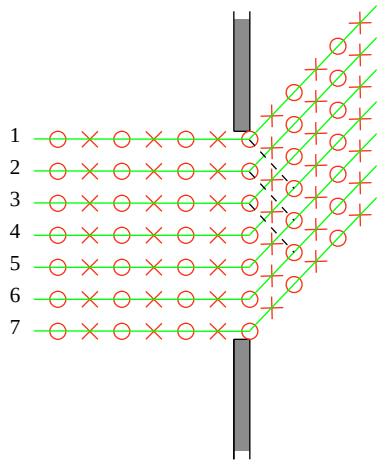
$$w^* = \frac{\lambda}{\sin(\theta)}$$

shines no light in the direction θ. Of course, if we position two, or three, or seventeen windows of this same width adjacent, they will again shine no light in the direction θ. In short, we have complete destructive interference (darkness) whenever $a = mw^*$, that is when

$$a\sin(\theta) = m\lambda \quad \text{for} \quad m = \pm 1, \pm 2, \pm 3, \ldots. \tag{6.1}$$

(Unlike the list in equation (3.1), this list of integers *excludes* zero. When $m = 0$ we in fact have a bright spot.)

How not to find diffraction maxima. You might think at this point: Now we can find diffraction maxima as well! Just use a different special angle were the extra length on path 4 is one wavelength rather than one-half wavelength.



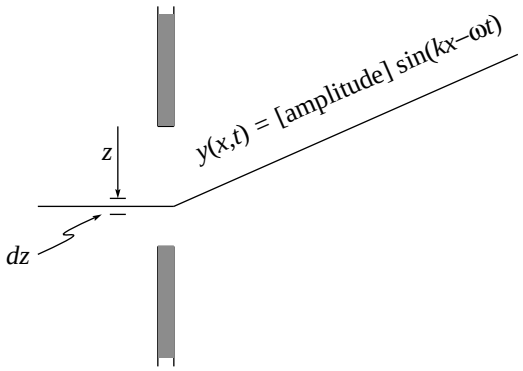
If the extra length for path 4 had been an integer number of wavelengths

$$\frac{a}{2}\sin(\theta) = m\lambda$$

then we would have constructive interference. This construction fails . . . in fact we just found that all such angles (except for $m = 0$) result in interference zeros. *Challenge:* Can you find the argument's flaw?

To correctly locate the diffraction maxima, we must find the intensity at all angles, using techniques like those of section 4.2, "Intensity in two-slit interference", and then analyze that intensity curve.

Intensity in single-slit diffraction: Setup. A wave of wavelength λ passes through a single slit of width a:



According to the Huygens construction, each infinitesimal element of this finite slit acts as a tiny radiator emitting spherical waves. Think about the infinitesimal element of width dz situated a distance z from the top of the slit. We will integrate the contribution from each such infinitesimal source.

Place the detector a long distance L from the top of the slit ("Fraunhofer limit"). Then the detector is a distance $L + z\sin\theta$ along the path shown in the figure. Hence the wave signal at the detector due to the path shown is

$$y(x,t) = [\text{amplitude}]\sin(k(L + z\sin\theta) - \omega t).$$

Because this wave is the wave due to an infinitesimal window of width dz, the amplitude will be very small. We write

$$[\text{amplitude}] = A\frac{dz}{a}$$

(where the division by a insures that "[amplitude]" has the proper dimensions). The total wave signal received by the detector is the signal integrated over all possible paths, from $z = 0$ to $z = a$:

$$y(t) = \int_0^a \frac{A}{a}\sin(kL + kz\sin\theta - \omega t)\,dz.$$

Clearly, this is a function periodic in time with angular frequency ω. We wish to put it into the form

$$y(t) = [\text{amplitude}]\sin([\text{phase}] - \omega t).$$

Once it's in this form, the intensity is proportional to $[\text{amplitude}]^2$.

Intensity in single-slit diffraction: Math. To make this algebra easier, we use Euler's relation

$$e^{it} = \cos t + i \sin t.$$

In these terms,

$$\begin{aligned} y(t) &= \int_0^a \frac{A}{a} \sin(kL + kz\sin\theta - \omega t)\, dz \\ &= \Im m \left\{ \int_0^a \frac{A}{a} e^{i(kL + kz\sin\theta - \omega t)}\, dz \right\} \\ &= \Im m \left\{ \int_0^a \frac{A}{a} e^{i(kL-\omega t)} e^{ikz\sin\theta}\, dz \right\} \\ &= \Im m \left\{ \frac{A}{a} e^{i(kL-\omega t)} \int_0^a e^{ikz\sin\theta}\, dz \right\}. \end{aligned}$$

But

$$\int_0^a e^{ikz\sin\theta}\, dz = \left[\frac{1}{ik\sin\theta} e^{ikz\sin\theta} \right]_{z=0}^{a} = \frac{1}{ik\sin\theta} \left(e^{ika\sin\theta} - 1 \right)$$

so we define

$$\alpha = \tfrac{1}{2} ka \sin\theta$$

and find

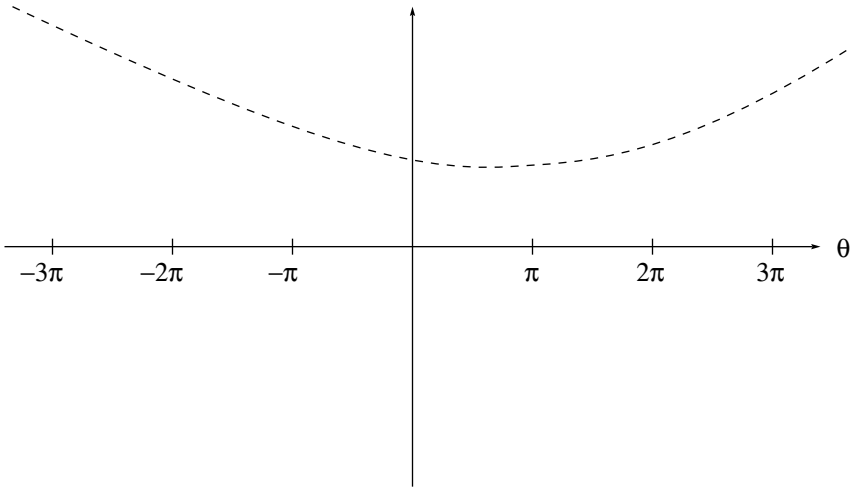
$$\begin{aligned} y(t) &= \Im m \left\{ \frac{A}{2i\alpha} e^{i(kL-\omega t)} \left(e^{i2\alpha} - 1 \right) \right\} \\ &= \Im m \left\{ \frac{A}{2i\alpha} e^{i(kL-\omega t)} e^{i\alpha} \left(e^{+i\alpha} - e^{-i\alpha} \right) \right\} \\ &= \Im m \left\{ \frac{A}{2i\alpha} e^{i(kL+\alpha-\omega t)} \left(e^{+i\alpha} - e^{-i\alpha} \right) \right\} \\ &= \Im m \left\{ \frac{A}{2i\alpha} e^{i(kL+\alpha-\omega t)} \left(2i\sin\alpha \right) \right\} \\ &= \Im m \left\{ A \frac{\sin\alpha}{\alpha} e^{i(kL+\alpha-\omega t)} \right\} \\ &= A \frac{\sin\alpha}{\alpha} \sin(kL + \alpha - \omega t). \end{aligned} \tag{6.2}$$

This expression is in the desired form. Because intensity is proportional to amplitude squared,

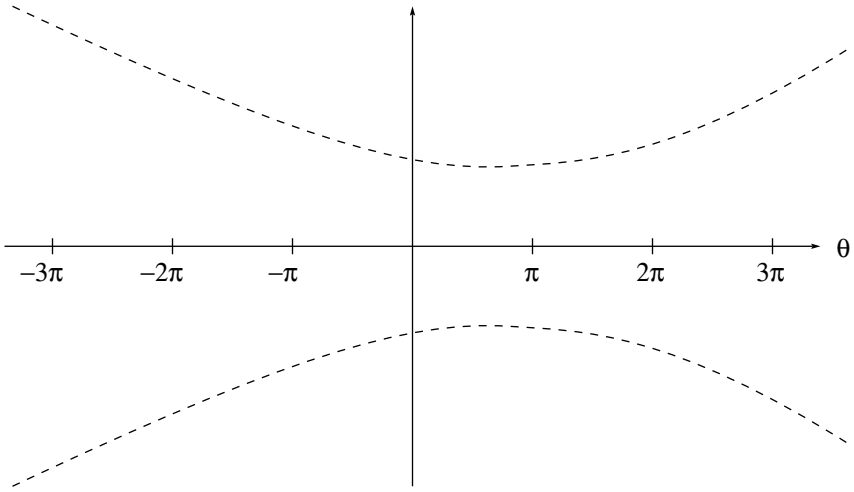
$$\text{intensity} = I_m \left(\frac{\sin\alpha}{\alpha} \right)^2 \quad \text{where} \quad \alpha = \frac{\pi a}{\lambda} \sin\theta. \tag{6.3}$$

Character of the intensity function. Just knowing the formula doesn't help much. What is its character? What is it telling us about nature?

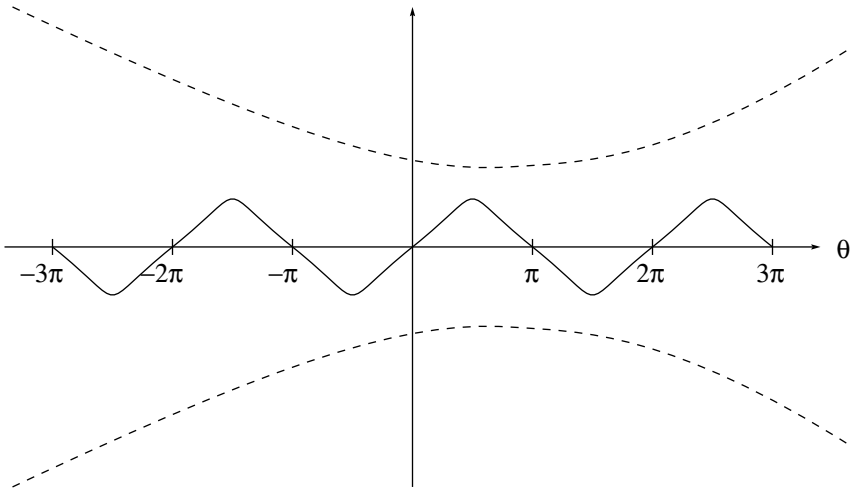
Do you know the trick for plotting $g(\theta)\sin(\theta)$, where $g(\theta)$ varies slowly when θ increases by 2π? First plot $g(\theta)$:



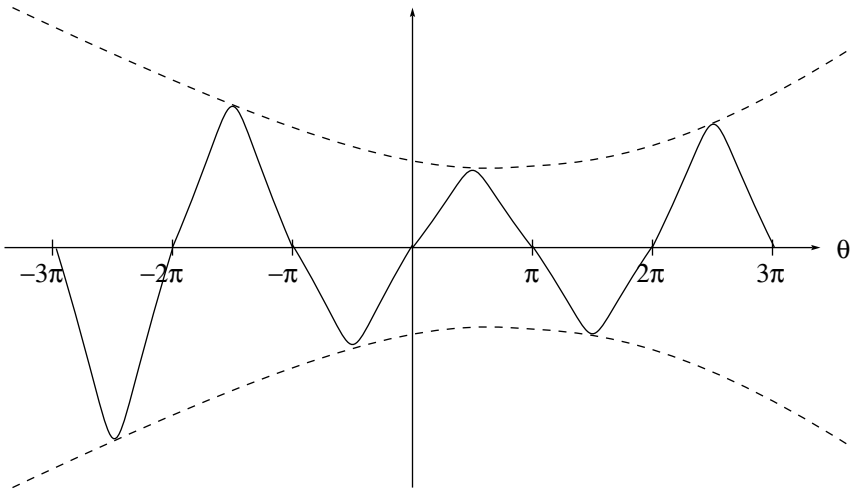
Then on the same graph plot $-g(\theta)$:



Now plot $\sin(\theta)$ as well:



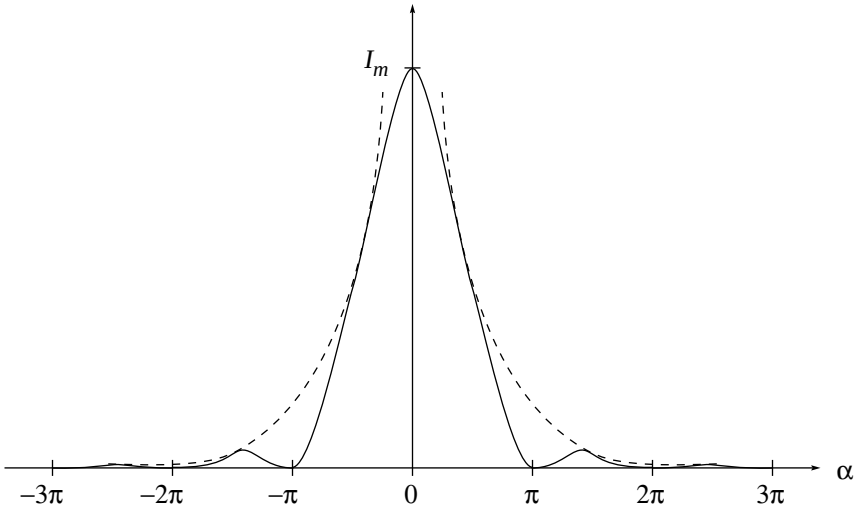
Because $-1 \leq \sin(\theta) \leq +1$, the product function $g(\theta)\sin(\theta)$ always falls within the envelope of $+g(\theta)$ and $-g(\theta)$. At the points where $\sin(\theta) = 0$, the product $g(\theta)\sin(\theta)$ of course equals 0 as well. At the points where $\sin(\theta) = +1$, the product $g(\theta)\sin(\theta)$ of course equals $g(\theta)$ — the product function touches the top of the envelope. And where $\sin(\theta) = -1$, the product function touches the bottom of the envelope.



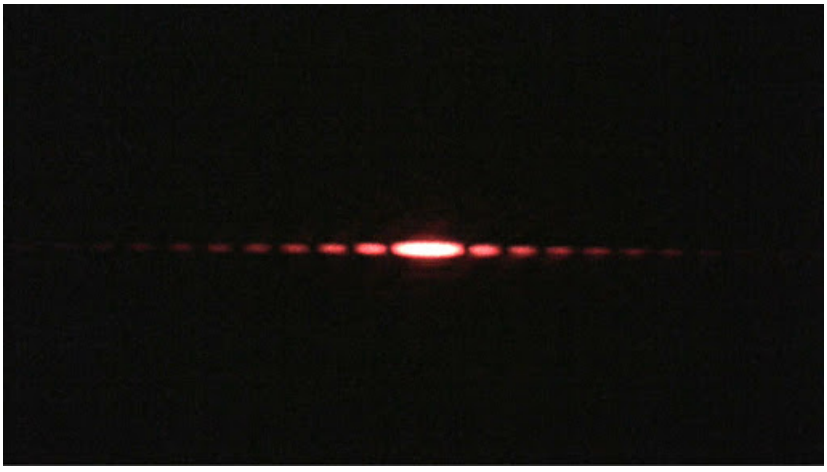
Apply such tricks to the function

$$I_m \left(\frac{\sin \alpha}{\alpha} \right)^2 .$$

The function is of course even and never negative. It bounces up and down between zero and I_m/α^2. A difficulty comes at $\alpha = 0$. But for small α, $\sin(\alpha) \approx \alpha$, so at $\alpha = 0$ the intensity function equals I_m. In short, the intensity function looks like



What does experiment say?



I'm happy.

Problems

6.1 *Maxima in the single-slit diffraction intensity curve.* Diffraction from a slit of width a produces an intensity curve of

$$I(\theta) = I_m \left(\frac{\sin \alpha}{\alpha} \right)^2 \quad \text{where} \quad \alpha = \frac{\pi a}{\lambda} \sin \theta.$$

We have already seen that this results in intensity minima (zeros) when

$$a \sin(\theta) = m\lambda \quad \text{for} \quad m = \pm 1, \pm 2, \pm 3, \ldots.$$

Show that it results in local intensity maxima whenever

$$\tan \alpha = \alpha.$$

(Thus the diffraction maxima are *not* located exactly halfway between the minima.)

6.2 *Width of the single-slit diffraction intensity curve.* The full width at half-maximum (FWHM) of a central diffraction maximum is defined as the angle between the two points in the pattern where the intensity is half that at the center of the pattern. Show that the point of half-maximum occurs when $\sin \alpha = \alpha/\sqrt{2}$.

Chapter 7

A Farewell to Waves

You will never truly say farewell to waves, because waves pervade our understanding of nature. But this is the last chapter of this document.

Waves in the day-to-day world. One of the things I love about waves (as opposed to relativity or quantum mechanics) is that there are immediate connections between the theory of waves and the day-to-day world. The next time you're shipped a package containing a rectangular styrofoam panel for padding, ignore the item you've purchased and pick up the styrofoam panel. With one hand, shake the short side. You've set up a standing wave pattern! What happens if you shake the long side?

If you like this sort of home experiment, I recommend these three books:

- *Waves* by Frank S. Crawford, Jr. (1968).
- *Light and Colour in the Open Air* by M.G.J. Minnaert (1940) [also published under titles *The Nature of Light and Colour in the Open Air* and *Light and Color in the Outdoors*].
- *Rainbows, Halos, and Glories* by Robert Greenler (1980).

I recommend even more these three papers by Frank Crawford, all concerning acoustic echos and all published in the *American Journal of Physics*:

- "Chirped Handclaps" (**38**, 378, March 1970) begins "One morning last December, as I was jogging on the outskirts of Stockholm, my path led me past the wall of a large factory. Following an impulse, I clapped my hands and listened. . . ".

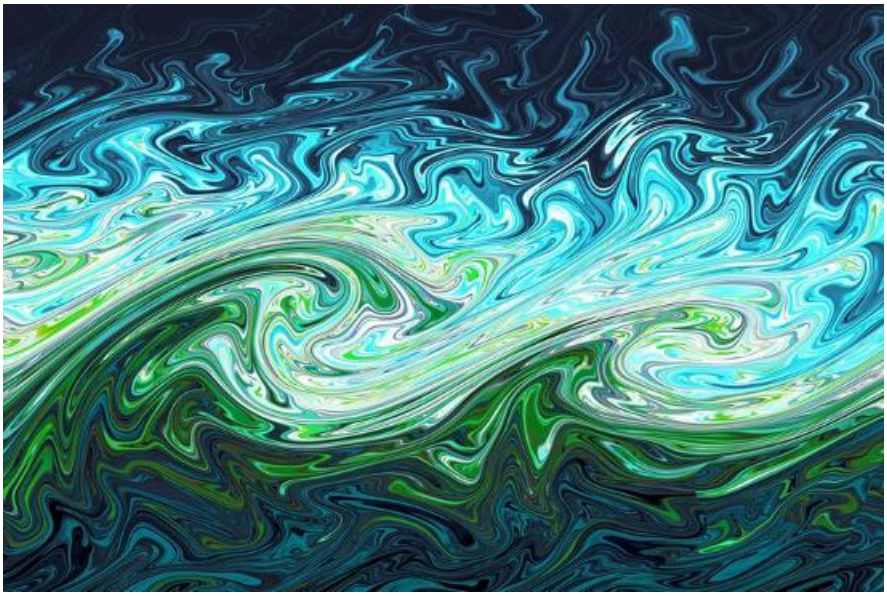
- "Douglas Fir Echo Chamber" (**38**, 1477, December 1970) begins "One afternoon last January I was walking with friends through a forest of magnificent Douglas fir. ... On an impulse I broke the forest silence with a loud whoop."
- "Culvert Whistlers" (**39**, 610–615, June 1971) begins "One morning last May I was romping with my two children on a beach near Bolinas, California, when, during a lapse of my attention, they disappeared."
- See also P.M. Rinard, "Rayleigh, Echoes, Chirps, and Culverts" (*Am. J. Phys.* **40**, 923–924, June 1972).

Just to give you a glimpse of everyday waves and optics, I show one photo that I took while backpacking on the Na Pauli coast of Kauai (part of my effort to go backpacking in each of the fifty states)

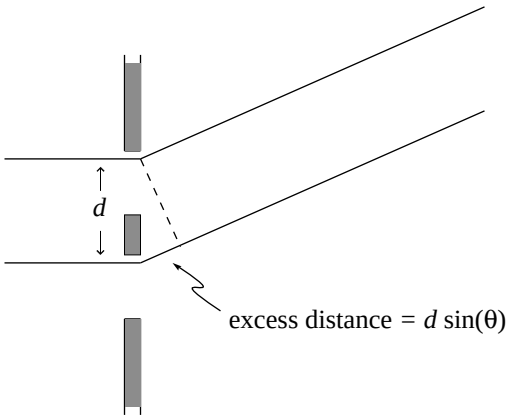
and a photo of so-called "supernumerary rainbows" (which I have never personally witnessed)

Wave research. You might think that the field of waves, initiated by Thomas Young in 1803, would have been so well explored over the previous two centuries that there would be nothing more to learn. You'd be wrong. One of the most important and challenging fields of physics research today is the topic of "turbulence", which is a subfield of waves.

Because the classical wave equation and the quantal Schrödinger equation are mathematically similar, this topic is in turn related to quanal chaos. Eric Heller researches both fields, and his wife thought that the data coming from his computer simulations was so beautiful it should be made into artwork. Here is a copy of Heller's artwork "Storm waves, chaos model"



Intensity in two-slit interference with finite slits. A wave of wavelength λ passes through two slits, each of width a, located a distance d apart:



Place the detector a long distance L from the top of the top slit ("Fraunhofer limit"). What is the intensity there? My first thought is that it's just the sum of the intensity due to the top slit (equation 6.3) plus the intensity due to the bottom slit, but that's wrong! It doesn't take account of interference. To get the right answer, we have to add not the intensities, but the wave signals y.

The wave signal at the detector due to the top slit is just given by equation (6.2). The wave signal at the detector due to the bottom slit is just given by equation (6.2), except with "L" replaced with "$L + d\sin\theta$". Thus the total wave signal is

$$y(t) = A\frac{\sin\alpha}{\alpha}\sin(kL + \alpha - \omega t) + A\frac{\sin\alpha}{\alpha}\sin(kL + kd\sin\theta + \alpha - \omega t).$$

This expression is exactly in the form of equation (4.2), except that A in (4.2) changes to $A(\sin\alpha/\alpha)$ above, and kL in (4.2) changes to $kL + \alpha$ above.

So the intensity result here has exactly the same form as the intensity result (4.4), with these two substitutions. The second substitution has no effect, because the intensity results are independent of L. Thus

$$\text{intensity} \quad = \quad I_m\left(\frac{\sin\alpha}{\alpha}\right)^2\cos^2\left(\frac{\phi}{2}\right) \tag{7.1}$$

$$\text{where} \quad \alpha = \frac{\pi a}{\lambda}\sin\theta \quad \text{and} \quad \phi = \frac{2\pi d}{\lambda}\sin\theta.$$

In short, the intensity for double slit diffraction with two wide slits is the product of the intensities for single slit diffraction with one wide slit and for double slit interference with two narrow slits.

The Poisson (Fresnel) bright spot. This is one of my favorite stories and favorite demos. You can check it out here

http://vanderbei.princeton.edu/images/Questar/PoissonSpot.html

Appendix A

Euler's formula

Where does

$$e^{i\theta} = \cos\theta + i\sin\theta \quad \text{(for } \theta \text{ real)}$$

come from? There are a number of ways to find it. Which way is most natural depends on which definitions you prefer for $e^{a\theta}$, $\cos\theta$, and $\sin\theta$. I prefer these:

The function	$e^{a\theta}$	is the solution to	$\frac{df}{d\theta} = af(\theta)$	with	$f(0) = 1$.
The function	$\cos\theta$	is the solution to	$\frac{d^2f}{d\theta^2} = -f(\theta)$	with	$f(0) = 1$
				and	$f'(0) = 0$.
The function	$\sin\theta$	is the solution to	$\frac{d^2f}{d\theta^2} = -f(\theta)$	with	$f(0) = 0$
				and	$f'(0) = 1$.

Using these definitions, it's clear that $e^{i\theta}$ is defined as the solution to $f'(\theta) = if(\theta)$ with $f(0) = 1$. Writing the complex function $f(\theta)$ as

$$f(\theta) = x(\theta) + iy(\theta), \quad \text{where} \quad x(0) = 1, \quad y(0) = 0,$$

the differential equation $f'(\theta) = if(\theta)$ becomes

$$x'(\theta) + iy'(\theta) = ix(\theta) - y(\theta).$$

The real and imaginary parts of this equation are

$$x'(\theta) = -y(\theta) \quad \text{and} \quad y'(\theta) = x(\theta).$$

To find a differential equation in terms of $x(\theta)$ alone, take the derivative of the left equation and then employ the right equation:

$$x''(\theta) = -x(\theta) \quad \text{with} \quad x(0) = 1 \quad \text{and} \quad x'(0) = 0.$$

This is the definition of $\cos\theta$.

To find a differential equation in terms of $y(\theta)$ alone, take the derivative of the right equation and then employ the left equation:

$$y''(\theta) = -y(\theta) \quad \text{with} \quad y(0) = 0 \quad \text{and} \quad y'(0) = 1.$$

This is the definition of $\sin\theta$.