Waves for people who know all about rocks

Thomas DeGrand

February 18, 2012

Contents

Chapter 1

Introduction

What these notes are all about

Perhaps a good start would be to explain the title. This book is a little introduction to the classical physics (and mathematics) of wave motion. The intended audience is sophomore physics majors at the University of Colorado, who are members of a Modern Physics course (actually not so modern; it is mostly about early twentieth century physics). We want to be able to talk at some low level about quantum physics. Everybody has probably heard the story, "the physics of the very small, say electrons bound to nuclei to make atoms, is not like the physics of classical point particles in a potential well, evolving under Newton's laws. It is more like the physics of classical wave systems..." The problem with this story is that if the audience has no intuition for what a classical wave system really is, these sentences make no sense. Maybe it would be a good idea to spend a few weeks talking about waves?

That still doesn't explain the title. Presumably, if you are in my target audience, you have just completed the equivalent of a year of freshman physics. If the subject hasn't changed much since I had to take that class, you are now an expert on the physics of rocks: point particles sitting in some potential well (or equivalently, feeling some instructor-chosen force), and you know how to find their motion, by solving $\vec{F} = m\vec{a}$. After the year of hard work you have put in doing this, I have to tell you something unpleasant: only a tiny corner of physics can be understood by solving $\vec{F} = m\vec{a}$. The really interesting and important parts of physics involve fields (see below for definition) and their associated wave phenomena. You spend years in physics and mathematics classes learning how to deal with them. However, most of those years are spent learning all the technical details. The essential facts are very simple.

Fields vs. rocks

Let's begin. How do we describe "rock problems"? Typically, we focus on a set of coordinates describing some localized region of space, and ask how that region changes with time. More simply, how does the rock fly through the air? A highbrow way of describing these problems is that they involve a small number of degrees of freedom: for motion in three dimensions, the three components of the vector $\vec{x}(t)$ which label the location of the center of mass of the particle. Perhaps you might want to add the three components of the momentum, maybe (if you are dealing with a rigid body) some components which describe its rotational motion. There is always one independent variable, time, and your solution at later time depends on a small number of variables defined at some earlier time (think about solving $\vec{F} = m\vec{a}$, and fixing the location and the velocity of the particle at $t = 0$.)

In contrast, the problems we want to solve have structure everywhere in space. Think of the temperature in this room, as a function of time. It is characterized by one number (the temperature, of course), but that number must be specified for every location in this room, for any time. Mathematically, we have a function $T(\vec{x}, t)$, whose properties we want to describe. Physicists call quantities like $T(\vec{x}, t)$ "fields." The equations which describe them are going to be more complicated than the rock's $\vec{F} = m\vec{a}$. That makes sense, physically. Imagine that you are sitting next to a radiator. $T(\vec{x}, t)$ near you is going to evolve differently than it would if you were sitting next to an open window on a cold day. Typically, when you ask questions about field variables, your solution depends on an infinite number of initial conditions, for example, the values of $T(\vec{x}, t)$ at all the $\vec{x}$ points on the surface of some region, evaluated at some (initial?) time.

Mathematicians say that the difference between rock problems and field problems is that rock problems are encoded by "ordinary differential equations," while field problems involve the solution of "partial differential equations."

This sounds scary, but there is some good news: In freshman physics, I don't remember spending much time on any one kind of "rock problem." They were all different. But when one studies field problems, one encounters the same problem over and over: either the temporal evolution of the field variables, or their spatial dependence, or both, is repetitive. The variables oscillate in time, $T(\vec{x}, t) \sim f(\vec{x}) \cos(\omega t)$ where $f(\vec{x}) \sim \cos(\vec{k} \cdot \vec{x})$. Even when the real solution does not take this particular form, it can often be usefully (and profitably) written as a sum of such terms. All of the interesting wave behavior I'll describe – interference, diffraction – can be understood by focusing on this one particular kind of behavior.

The reader objects!

At this point you might say – this does not sound like physics. I signed up for a physics class, and I want to talk about atoms. Haven't you tricked me?

Maybe I have! But, remember, I am a theoretical physicist. Mathematics is the language I use to explain physical phenomena. If you want to to write poetry in German, you have to know German grammar. If you want to create new physics, you need mathematics to explain what you are doing.

Theorists have their own version of "mathematical phenomenology." Part of it is recognizing simplicity in complicated equations. The only way I know to do this is to learn all the different behavior which can arise from a simple system. Then you can recognize (if you are lucky or bright) that the complicated system you are studying is actually a simple system in disguise. The only way to acquire this knowledge is to study simple systems.

Remember, too: language is slippery. You can invent a story which seems to explain anything. But it has a chance of being true, only if it is logically consistent, both internally and in its connection to other physical phenomena. Mathematics is the tool we use to ensure that logical consistency. And we are not doing string theory. What I am going to tell you about is (or should) be part of the tool kit of every physicist.

What we are going to do

It turns out that most of the wave phenomena that every physicist knows begins and nearly ends with the study of the simple harmonic oscillator, so that is what we will look at, first. The mathematics we need is almost completely the mathematics of trigonometric identities. Nobody can remember these identities, so there are a whole set of tricks which replace trigonometry by the algebra (and calculus) of complex numbers. This sounds quite improbable, but the reason we do it is that the mathematics becomes much simpler! (Differentiation is replaced by multiplication, for example.) I will introduce the physics of wave motion hand in hand with the mathematics you need. And, it will happen, as we get deeper into the subject of waves, we discover all sorts of beautiful and improbable behavior.

Where this will be used

"The physics of the very small, say electrons bound to nuclei to make atoms, is not like the physics of classical point particles in a potential well, evolving under Newton's laws. It is more like the physics of classical wave systems…"

That's a nice story. But it sounds fantastic, is it true? (Suppose it were 1925, it would seem even more fantastic.) If it is true, then experiments involving electrons will show phenomena which look like classical wave phenomena. One of those phenomena is diffraction:

a beam of electrons illuminating a crystalline solid will scatter in a particular way, the same way that light scatters off a diffraction grating. Detailed properties of the diffraction pattern will depend on the wavelength of the electron and on the spacing of atoms in the solid (which we could already know from measuring it via X-ray scattering).

We will also play around a bit with "wave equations," which are the partial differential equations which describe how fields (like the temperature $T(\vec{x}, t)$) depend on their values at earlier time or at nearby locations. The analog for quantum mechanics is an equation called the Schrödinger equation. Its solutions allow one to predict most of the properties of matter on the smallest distance scales.

And, of course, you will encounter wave behavior when you study electromagnetism, fluid dynamics, acoustics, strength of materials, vibrations in structures, solid state physics, optics, heat transport, explosions, and the motion of flocks of birds. We won't have time to go into any of these topics in any detail.

A suggested reading list:

- Boas, "Mathematical methods in the physical sciences, 2nd ed," particularly chapters 2 (complex numbers) and 7 (Fourier series)
- French "Vibrations and Waves"
- Feynman lectures, vol 1, chapters 21-25

As the author of these notes I am absolutely confident that they are full of typos. So if you think you have read them very carefully, and you have not either told me about or emailed me a list of typos, then you have not read them carefully enough.

(Thanks to C. Yarrish for corrections.)

Chapter 2

The simple harmonic oscillator

The simple harmonic oscillator – a mass on a spring

Hopefully, one of the rock problems you saw last year was a mass m attached to a Hooke's law spring with spring constant k. I am no good with forces, so I prefer to think about the potential energy

$$V(x) = \frac{1}{2}kx^2. \tag{2.1}$$

However you want to approach the problem, you end up with Newton's equation, which in this case is

$$m\frac{d^2x}{dt^2} = -kx \tag{2.2}$$

Just for bookkeeping purposes, we define the quantity $\omega^2 = k/m$. (Check the units, ω is a frequency.) Then our equation become

$$\frac{d^2x}{dt^2} = -\omega^2 x. \tag{2.3}$$

We could solve this honestly, but instead let us cheat and just recall a formula from calculus class,

$$\frac{d^2\cos(\omega t)}{dt^2} = -\omega^2\cos(\omega t). \tag{2.4}$$

The same equations have the same solutions, so we know that a solution to Eq. 2.3 is

$$x(t) = A\cos(\omega t). \tag{2.5}$$

The cosine is not the only function whose second derivative is its negative, so in fact the most general solution to our problem is

$$x(t) = A\cos(\omega t) + B\sin(\omega t). \tag{2.6}$$

It involves two arbitrary quantities A and B How can we find them? Remember how you solve mechanics problems: you have to specify both the position and velocity at some time, say at $t = 0$: this is easy, because $\cos\ 0 = 1$ and $\sin\ 0 = 0$, so

$$x(0) = A \tag{2.7}$$

and (take the derivatives yourself)

$$v(0) = \frac{dx}{dt}|_{t=0} = \omega B. \tag{2.8}$$

In words, our two quantities are determined by two initial conditions.

We can write the answer more compactly, at the cost of knowing trigonometric identities: recall that

$$\cos(a - b) = \cos(a)\cos(b) + \sin(a)\sin(b) \tag{2.9}$$

and so if we make the substitutions $b = \omega t$ and

$$A = C\cos(\phi); \qquad B = C\sin(\phi) \tag{2.10}$$

we have

$$x(t) = C\cos(\omega t - \phi) \tag{2.11}$$

where (work it out!) $\tan\phi = B/A$. We still have two unknowns, C and ϕ, which have to be determined from the initial conditions. These quantities have names: C is called the "amplitude" and ϕ is called the "phase." Physically, the amplitude is the size of the motion, $x(t)$ oscillates between $+C$ and $-C$. The phase ϕ tells you when the amplitude is maximum ($\omega t_{max} = \phi$). The quantity ω is called the "angular frequency" (sometimes just "the frequency," but this is slightly sloppy language). Note that the motion is repetitive and at times t, $t + 2\pi/\omega$, $t + 4\pi/\omega$, $\ldots$, the particle returns to the same location. (It's more useful to say that when $\omega t = 2\pi$, 4π, 6π, $\ldots$, the particle returns to its original location.) Equivalently, the period of oscillation is $T = (2\pi)/\omega$.

The amplitude also has a role in determining the energy of the oscillator. Recall that the energy E is the sum of the kinetic and potential energy,

$$E = \frac{1}{2}m(\frac{dx}{dt})^2 + \frac{1}{2}kx^2. \tag{2.12}$$

Also recall that if $x(t) = C\cos(\omega t - \phi)$, $dx/dt = -\omega C\sin(\omega t + \phi)$ and (with $k = m\omega^2$)

$$E = \frac{1}{2}m\omega^2C^2(\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi)) = \frac{1}{2}m\omega^2C^2. \tag{2.13}$$

Note several more apparently simple things:

- The energy is proportional to the square of the amplitude
- The amplitude can take on any value, it only depends on the initial conditions, so the energy could be any (positive) value
- The frequency is fixed, if the mass and spring constant k are fixed

The analogy with circular motion

For the next step we turn to a problem which seems superficially quite different: a particle undergoing uniform circular motion, like a stone tied to a rope. Let's suppose the stone moves counterclockwise in a circle in the $x-y$ plane. If we look at the picture, the motion can be described in terms of an angle which increases uniformly with time, $\theta = \omega t$, (until it returns to the start and repeats). We assume that at time $t = 0$ the stone is located at an angle ϕ with respect to the x axis. Also the length of the rope, R, is a constant. Projecting this motion onto the x and y axis we have

$$x(t) = R\cos(\theta) = R\cos(\omega t + \phi) \tag{2.14}$$

and

$$y(t) = R\sin(\theta) = R\sin(\omega t + \phi) \tag{2.15}$$

Notice the similarity of Eqs. 2.14 and 2.11! If you project circular motion onto a line, say by putting a flashlight in the plane of circular motion and projecting an image on the wall, the projection moves as a simple harmonic oscillator. R is the amplitude of the oscillation and ϕ is the phase angle.

Actually, a better use of the analogy is to run it backwards: to think of harmonic motion as a reduced form of uniform circular motion.

Lightning review of complex numbers

Please pause and carefully read Chapter 22 of Volume 1 of the Feynman lectures on physics, before going on.

Do you recall learning about complex numbers in your high school algebra class? I hope so! So here is a lightning review. A complex number is a pair of real numbers joined together with the square root of minus one, i,

$$z = a + ib. \tag{2.16}$$

We say that "a is the real part of the complex number z" and "b is the imaginary part of z," and we write this expressions as $a = \mathrm{Re}z$ and $b = \mathrm{Im}z$ for short.

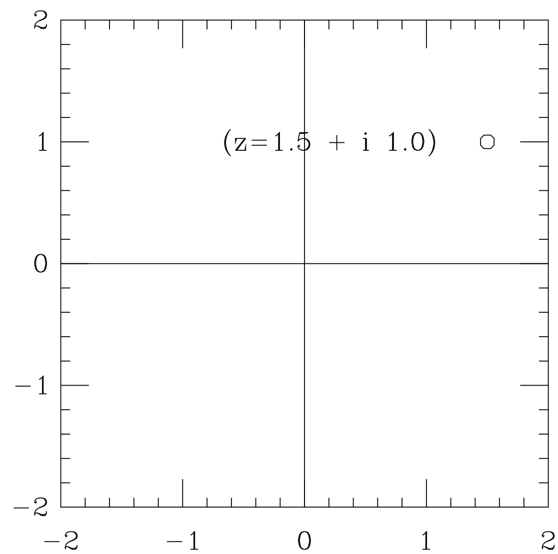

Figure 2.1: A generic complex number ($z = 1.5 + i1.0$) plotted in the complex plane.

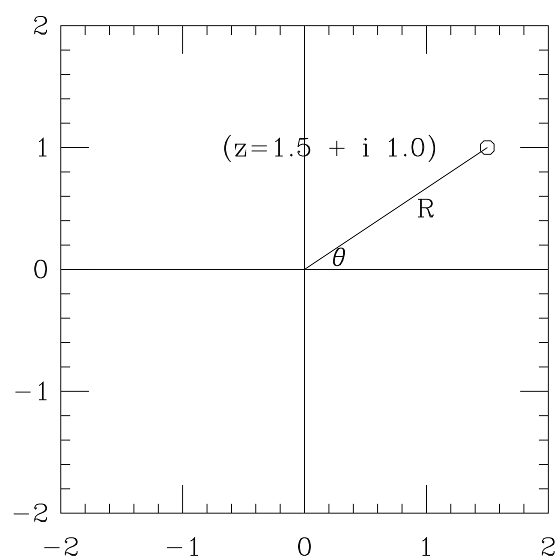

Figure 2.2: Our generic complex number ($z = 1.5 + i1.0$) plotted in the complex plane with its polar angle (here $\tan\theta = y/x = 2/3$ and length ($R = \sqrt{1.5^2 + 1^2}$).

The addition of two complex numbers, $z_1 = a_1 + ib_1$ and $z_2 = a_2 + ib_2$ is just usual vector addition (you add components)

$$z_1 + z_2 = a_1 + ib_1 + a_2 + ib_2 = (a_1 + a_2) + i(b_1 + b_2). \tag{2.17}$$

Multiplication is also straightforward,

$$(a_1 + ib_1) \times (a_2 + ib_2) = (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + a_2 b_1). \tag{2.18}$$

Notice that the product of two complex numbers is also a complex number.

The "complex conjugate" of a complex number $z = a + ib$ is gotten by flipping the sign of the imaginary part: $z^* = a - ib$. Note that

$$z^* z = (a - ib)(a + ib) = a^2 + b^2. \tag{2.19}$$

As a simple check, if $z = 3 + 4i$, then the complex conjugate of z is $z^* = 3 - 4i$ and can you multiply $(3 - 4i) \times (3 + 4i) = 25$?

Notice that $z^* z = z z^*$ (commutativity), and note that one often writes $z^* z = |z|^2$.

Complex conjugation has other uses. One is in division.

$$\frac{a + ib}{c + id} = \frac{a + ib}{c + id}\left(\frac{c - id}{c - id}\right) = \frac{ac + bd + i(bc - ad)}{c^2 + d^2} \tag{2.20}$$

which is now written in our standard form (real + $i\times$imaginary).

Hopefully your instructor will give you a long problem set to practice all this!

There is a natural visualization of a complex number $z = a + ib$ as a point in a plane. The point is labeled with its x-coordinate, which we take to be a, and its y-coordinate, b. This is exactly like using rectangular coordinates for a two-dimensional vector. There is another way to characterize the vector, of course – in polar coordinates. The squared length of the vector is $R^2 = a^2 + b^2 = z^* z$. The angle the ray makes with the $\hat{x}$ axis, θ, is given by $\tan\theta = b/a$ or $a = R\cos\theta$ and $b = R\sin\theta$.

Note that if you want to add two complex numbers, it is like adding two vectors – you just add the components. In polar form, adding the vectors is identical to the graphical methods you know for adding vectors.

Finally, there are complex functions of a complex variable $z = x + iy$, just written $f(z)$. They are not mysterious, just write the function and substitute:

$$z^2 = (x + iy)^2 = x^2 - y^2 + 2ixy$$

$$
\begin{aligned}
1-z &= 1-x-iy \\
\frac{1}{z}z &= \frac{x-iy}{x^2+y^2}(x+iy)=1 \\
\frac{z+z^*}{2} &= x \\
\frac{z-z^*}{2} &= iy
\end{aligned}
\tag{2.21}
$$

Complicated functions (not powers) often only have meaning as power series.

Now consider more carefully the polar form

$$z = R\cos\theta + iR\sin\theta \tag{2.22}$$

and suppose $R = 1$ so $z = \cos\theta + i\sin\theta$. What is z, more simply? Amazingly, it is

$$\exp(i\theta) = \cos\ \theta + i\sin\ \theta. \tag{2.23}$$

If you don't believe this, expand both sides in a Taylor series about $\theta = 0$ and compare

$$\exp(i\theta) = 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} = (1 - \frac{\theta^2}{2!} + \ldots) + i(\theta - \frac{\theta^3}{3!} + \ldots). \tag{2.24}$$

(Actually I can never remember the series for the sine and cosine beyond the first two terms, so I use this formula backwards to fill in my ignorance.)

Note that $e^{i\pi} = \cos\pi + i\sin\pi = -1$, $\exp(i\pi/2) = i$, and so on.

Also note that if we have determined R and θ from the rectangular form of z, then we know $z = R\exp(i\theta)$ as a compact polar expression for z.

What is z^*? If $z = a + ib$ then $z^* = a - ib$, so if $z = R\exp(i\theta)$, then $z^* = R\exp(-i\theta)$! $z^*z = (R\exp(-i\theta))(R\exp(i\theta)) = R^2\exp(-i\theta + i\theta) = R^2$.

Powers are simple in polar form:

$$
\begin{aligned}
z^n &= [Re^{i\theta}]^n = R^n e^{in\theta} \\
z^{-1} &= \frac{1}{z} = \frac{1}{R}e^{-i\theta} \\
\log z &= \log Re^{i\theta} = \log\exp(\log R + i\theta) = \log R + i\theta.
\end{aligned}
\tag{2.25}
$$

Note $\log(-1) = \log e^{i\pi} = i\pi$.

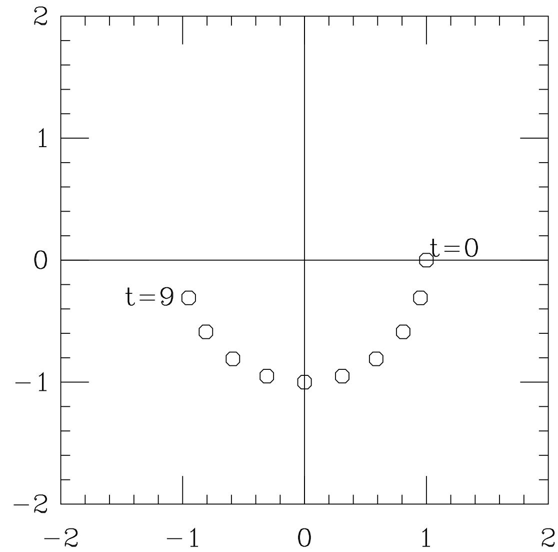

Figure 2.3: The function $z(t) = \exp(-i(2\pi/20)t)$ plotted as a function of t for $t = 0$ to 9 in the complex plane.

Complex exponentials for oscillators

Now think about the function

$$z(t) = R\exp(-i\omega t). \tag{2.26}$$

At $t = 0$ this is a real number, R, that is the complex number is located on the real axis. As t increases, the complex number rotates clockwise (at $\omega t = \pi/2$, $z = -iR$, at $\omega t = \pi$, $z = -R$, at $\omega t = 3\pi/2$, $z - iR$, at $\omega t = 2\pi$ we are back at the real axis, $z = R$). The function $z(t) = R\exp(-i\omega t)$ is a mathematical realization for uniform circular motion.

Now let us run the analogy (rotation in the complex plane) → (rotation) → harmonic motion to its logical conclusion and think of the motion of the mass in a harmonic oscillator as the real part of complex motion, that is, $x(t)$ is actually a complex number $z(t)$, which evolves in time as

$$z(t) = x(0)\exp(-i\omega t), \tag{2.27}$$

that is,

$$x(t) = \text{Re } x(0)\exp(-i\omega t) \tag{2.28}$$

The way you use this formula is to replace the cosine by the complex exponential at the start of the calculation, work everything out, and then at the end take the real part.

This probably seems either absurd or obvious (maybe both). Why do this? Pure expediency! You can describe classical wave motion entirely in terms of sines and cosines. However, there is a price for doing this: you have to be absolutely fluent in the algebra and calculus of trigonometric identities. This is knowledge which most people do not want to retain in their minds. Instead, people prefer to remember only simple things. Here the simple thing is "how to differentiate an exponential." Of course, you know the answer: Differentiating an exponential is done by multiplication,

$$\frac{d \exp(ax)}{dx} = a \exp(ax). \tag{2.29}$$

For the complex exponential in Eqn. 2.28

$$\frac{d \exp(-i\omega t)}{dt} = -i\omega \exp(-i\omega t) \tag{2.30}$$

and of course if

$$x(t) = x_0 \exp(-i\omega t) \tag{2.31}$$

then

$$v(t) = \frac{dx(t)}{dt} = -i\omega x_0 \exp(-i\omega t). \tag{2.32}$$

Furthermore, the acceleration is

$$a(t) = \frac{dv(t)}{dt} = -\omega^2 x_0 \exp(-i\omega t) = -\omega^2 x(t) \tag{2.33}$$

which is precisely Newton's equation for the oscillator. What is the $-i$ in the velocity? Recall the convention, take the real part (and let's assume that x_0 is real for now): $\mathrm{Re}(-i\omega x_0 \exp(-i\omega t)) = \mathrm{Re}(-i\omega x_0(\cos(\omega t) - i\sin(\omega t)) = -x_0 \sin(\omega t)$ which is $v(t)$ if $x(t) = x(0)\cos(\omega t)$.

The desire not to think when it is not necessary permeates all theoretical physics, and so you will find that any physics text which is not pandering to you will adopt this trick. It is also widely used by engineers. Signal processing is done using either complex exponentials or real exponentials. The second choice takes us into Laplace transforms, which you can learn about in sophomore electrical engineering courses.

As an added feature, integrating an exponential is done by division,

$$\int dy \exp(ay) = \frac{\exp(ax)}{a} \tag{2.34}$$

Of course, nothing is free...the problem with doing this is that you are forced to think more abstractly about the science. I think about oscillation as motion in the complex plane:

the particle rotating in a circle of radius x_0, the directin depending on whether I am thinking about $\exp(i\omega t)$ or $\exp(-i\omega t)$, rather than oscillating up and down in a two dimensional (x and t) picture.

Why the minus sign in Eq. 2.26 ? It is a pure convention, and you always have to check whatever book you are reading, to make sure that it uses this convention and not the opposite one with a plus sign. ($\cos(\omega t)$ is the real part of both $\exp(i\omega t)$ and and $\exp(-i\omega t)$.) We could have done the whole calculation with $x_0 \exp(+i\omega t)$. It is useful to adopt this $-i\omega t$ convention for yourself, and to live with it, because the same $\exp(-i\omega t)$ factor appears in quantum mechanics, and there it is NOT a convention. (The next few sentences may be more than you want to know right now.) The rough analog of the oscillator's $x(t)$ in quantum mechanics is a field called the "wave function" or "probability amplitude," $\psi(x,t)$. It is an intrinsically complex object – a complex number at every point of space and time. Its absolute square (in the sense of a complex number, $|\psi(x,t)|^2 = (\text{Re}\psi(x,t))^2 + (\text{Im}\psi(x,t))^2$ has the interpretation of being proportional to the probability of finding the particle near the location x at time t. Anyway, the Schrödinger equation for the time evolution of a state with energy E is

$$i\hbar \frac{\partial \psi(x,t)}{\partial t} = E\psi(x,t) \tag{2.35}$$

so

$$\psi(x,t) = \exp(-iEt/\hbar)\psi(x,0) \tag{2.36}$$

gives the time dependence of these special states.

In the formula, the constant $\hbar$ (pronounced h-bar) is Planck's constant h divided by 2π. Its numerical value is $\hbar c \sim 1970$ electron volt- Angstroms or equivalently 197 MeV-fm (mega-electron volts-Fermis; an electron volt is the energy gained by a particle with the electron's charge falling through a potential difference of one volt. An Angstrom is 10^{-8} cm and atoms are about an Angstrom in diameter. A Fermi is 10^{-13} cm and this is roughly the radius of a proton. I've given you the useful units for practical calculations. Nobody I know remembers what is the value of $\hbar$ is in MKS!) More about this, later.

Finally, one more thing: The quantity $x(0)$ does not have to be real. Suppose it is not: suppose $x(0) = A + iB$. Then the complex coordinate is

$$\begin{aligned} x(t) &= (A+iB)\exp(-i\omega t) \\ &= (A\cos(\omega t) + B\sin(\omega t)) + i(B\cos(\omega t) - A\sin(\omega t)) \end{aligned} \tag{2.37}$$

Our convention is to take the real part of the expression, which is $x(t) = A\cos(\omega t) + B\sin(\omega t)$, the general solution we originally found. A is the value of x at time zero, ωB is the velocity at time zero.

The driven oscillator

Let us look at a slightly different problem. Suppose we have a harmonic oscillator, and we drive it by applying some external force $F(t)$ to it. It is easy to write down Newton's equation,

$$m\frac{d^2x}{dt^2} = -kx + F(t) \tag{2.38}$$

but now what? Let us retreat slightly, and consider a special force, one which is harmonic,

$$m\frac{d^2x}{dt^2} = -kx + F_0\cos(\omega t). \tag{2.39}$$

This time we can find a solution: we guess $x(t) = x_0\cos(\omega t)$, and try it out:

$$m\frac{d^2x}{dt^2} = -m\omega^2 x_0\cos(\omega t) = -kx_0\cos(\omega t) + F_0\cos(\omega t) \tag{2.40}$$

The cosine terms are common and we can divide them out. This leaves an equation for the amplitude,

$$-m\omega^2 x_0 = -kx_0 + F_0 \tag{2.41}$$

which we solve (pausing to introduce the natural frequency of the oscillator, $\omega_0^2 = k/m$)

$$x_0 = -\frac{F_0}{m}\left(\frac{1}{\omega^2 - \omega_0^2}\right) \tag{2.42}$$

The mass follows the driving force (pretty much by design), and the driving force completely determines the amplitude of the oscillator. Notice what happens when the driving frequency ω, approaches the natural frequency of the oscillator, ω_0: The amplitude becomes very large! This is an example of what is called "resonance." Physically, it is very easy to make the oscillator move at a frequency that it would be happy to move at, in the absence of any extra force.

Right at $\omega = \omega_0$, bad things happen – the amplitude tries to diverge. Presumably, to accurately model what is going on there, we must worry about additional physics effects, not included in our equation. Let us leave this problem until next semester.

Let's repeat this with $\exp(-i\omega t)$:

$$m\frac{d^2x}{dt^2} = -kx + F_0\exp(-i\omega t) \tag{2.43}$$

Again, we look for a solution in which x tracks the force, so we write $x(t) = x_0\exp(-i\omega t)$. We drop this in, and find

$$m\frac{d^2x}{dt^2} = -m\omega^2 x_0\exp(-i\omega t) = -kx_0\exp(-i\omega t) + F_0\exp(-i\omega t) \tag{2.44}$$

and again

$$x_0 = -\frac{F_0}{m}\left(\frac{1}{\omega^2-\omega_0^2}\right). \tag{2.45}$$

It does not look like we have gained much. However, suppose our driving force was $F(t) = F_1\cos(\omega t - \phi)$? This is the real part of $F_1\exp(-i(\omega t-\phi))$ which is (rearranging parentheses) $(F_1\exp(i\phi))\exp(-i\omega t)$. We do not have to re-solve our problem, using trigonometric identities. Just treat F_0 as complex: $F_0 = F_1\exp(i\phi)$ and drop it into the solution we just found.

$$x_0 = -\frac{F_1\exp(i\phi)}{m}\left(\frac{1}{\omega^2-\omega_0^2}\right). \tag{2.46}$$

Complex $x(t)$ just means that the solution, real $x(t)$, has a phase factor,

$$x(t) = -\frac{F_1}{m}\left(\frac{1}{\omega^2-\omega_0^2}\right)\cos(\omega t-\phi). \tag{2.47}$$

How almost everything is almost an oscillator

Suppose you are given a mechanical system, which is already pre-assembled, and all of its parts are at rest. You tap it lightly on the side with a hammer, and it, and all of its internal parts, begin to move. Perhaps it begins to ring. How do you describe all this motion?

To answer this question, look closely at some part of the mechanical system, say at a mass point (a bolt or a glob of solder). Before you struck the side of the system, the mass point was not moving. Because it was not moving, there was net force acting on the system. Said differently, the potential energy of the mass point, evaluated at its equilibrium location, $V(x_0)$, was a minimum. Now (gently) hit the side of the system with a hammer. You slightly displace the mass point from x_0 to x. The potential is now $V(x)$ which is probably very close to $V(x_0)$. How close? Taylor's theorem tells us:

$$V(x) = V(x_0) + (x-x_0)\frac{dV(x)}{dx}|_{x=x_0} + \frac{1}{2}(x-x_0)^2\frac{d^2V(x)}{dx^2}|_{x=x_0} + \ldots \tag{2.48}$$

However, $V(x_0)$ was a minimum! This means that $\frac{dV(x)}{dx}|_{x=x_0} = 0$ both from mathematics (the minimum of a smooth function always occurs at a place where the first derivative vanishes) or physically (this term is the force at location x_0, but we were in equilibrium there, so the force was zero). Thus we are left with

$$V(x) = V(x_0) + \frac{1}{2}(x - x_0)^2 \frac{d^2V(x)}{dx^2}|_{x=x_0} + \ldots \tag{2.49}$$

If we neglect the ..., then we have a potential proportional to $(x - x_0)^2$, which is what we mean by a harmonic oscillator potential, or equivalently, that the force is proportional to $x - x_0$, which is Hooke's law. The "spring constant" is just $k = \frac{d^2V(x)}{dx^2}|_{x=x_0}$.

You might object: this is too simple! the potential that your mass point sees, at x_0, is what it is because there are other mass points connected to it by springs. When you tap the side of the box, you also displace the location of the other mass points. What happens then?

What happens then, is that the system still vibrates, but the vibrational motion of a single mass point is more complicated than pure $\exp(-i\omega t)$, instead, it is a linear combination of several $\exp(-i\omega_j t)$'s. Motion involving several mass points can be simple, a pure harmonic oscillator. These separate collections of masses evolve independently, each with its own $\exp(-i\omega_j t)$. We call them the "normal modes" of the system. In that sense, every classical system is almost a collection of harmonic oscillators.

Rather than derive this result, let us cheat and work backwards: Suppose we have two equal mass one-dimensional particles located at x_1 and x_2, and suppose that they move as follows:

$$\begin{aligned} x_1(t) + x_2(t) &= A\exp(-i\omega_1 t) \\ x_1(t) - x_2(t) &= B\exp(-i\omega_2 t) \end{aligned} \tag{2.50}$$

Clearly, the individual motion is

$$\begin{aligned} x_1(t) &= \frac{A}{2}\exp(-i\omega_1 t) + \frac{B}{2}\exp(-i\omega_2 t) \\ x_2(t) &= \frac{A}{2}\exp(-i\omega_1 t) - \frac{B}{2}\exp(-i\omega_2 t) \end{aligned} \tag{2.51}$$

What physical system could produce such motion? The $x_1(t) - x_2(t)$ motion has an acceleration

$$\frac{d^2(x_1 - x_2)}{dt^2} = -\omega_2^2(x_1 - x_2) \tag{2.52}$$

which corresponds to a potential $V = m\omega_2^2(x_1 - x_2)^2$ – this is a spring connecting particles 1 and 2. The "sum" term is a harmonic oscillator acting on the center of mass. We can build that by attaching another spring to the center of the first spring. You will (hopefully) spend endless hours doing the reverse of this problem in your classical mechanics class – given a set of springs and masses, what are the normal modes?

This is an extremely powerful observation: if the motion of a complicated mechanical system is actually given by a superposition of harmonic oscillator motion, then we have a general method for understanding the system: Find the normal modes (the analogs of $x_1 \pm x_2$ in Eq. 2.50 which evolve as pure exponentials, not sums of exponentials), and the natural frequencies (the analogs of ω_1 and ω_2) and you can completely characterize the behavior of the system. Often, you only need to know the natural frequencies. For example, when you tapped the side of the box, you heard it ring. The ringing is just high frequency vibration which couples to sound waves (compressional waves) in the air. If you measure the frequency spectrum of the ring, you can determine some of the vibrational frequencies of the box, and vice versa – if you know the spectrum, you know what sound you will hear. Of course, you don't know the amplitude of the various modes, until you figure out the initial conditions, but that is also a story for later.

Actually, everything is NOT almost an oscillator. You already know of at least one counter example – a pendulum. Only for very small angle oscillations is the pendulum a simple harmonic oscillator. This is because the potential is $V(\theta) = -mg\cos(\theta)$, so the equation of motion is (it's actually a torque equation, if you are being picky)

$$ml^2\frac{d^2\theta}{dt^2} = -mgl\sin(\theta) \tag{2.53}$$

For small angles, we Taylor expand $\sin(\theta) = \theta$ and then we have our oscillator

$$ml^2\frac{d^2\theta}{dt^2} = -mgl\theta, \tag{2.54}$$

but you also probably remember the Taylor expansion's next term from some calculus exercise, $\sin(\theta) = \theta - (1/6)\theta^3$. The equation of motion is different from an oscillator and it is no surprise that the solution is also different. (Maybe you also did an experiment at some point, measuring the period and discovering it is not quite $T = 2\pi\sqrt{l/g}$?) If the initial angle gets too great, all kind of amazing things happen to the pendulum – think about starting it with the weight nearly vertical and initial conditions which cause it to shoot over the top.

But, if you are handed a pendulum in which the plumb bob is hanging vertically below its support, and you are also given a very small hammer, all this discussion is irrelevant. You have a simple harmonic oscillator.

Chapter 3

Traveling waves

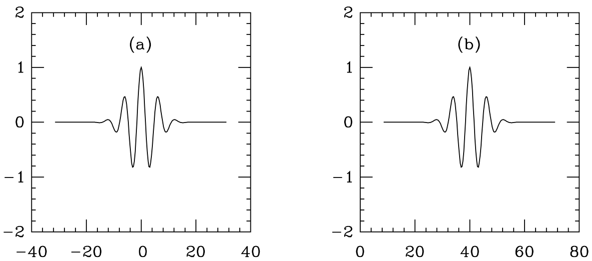

Figure 3.1: Profile of a traveling wave at two different times (note the x axis.)

Some general statements about wave motion

So what is a wave? You throw a stone in a lake and a ripple runs out from the impact point. I strike a gong and the room fills with its sound.

Mathematically, the way that we describe waves is as follows: Recall that a "field" is a function defined at every point of space and time. Wave motion is an excitation of some field. In these two cases the field is the local height of the water at some point on the surface of the lake, or the magnitude of the pressure of the air at some point in the room.

People generally talk about two kinds of waves, traveling waves and standing waves.

The first kind of wave (the ripple) is an example of a "traveling wave:" in words, the ripple is the peak in a function, and the location of the peak moves with the velocity of the wave. Let's imagine this situation, but in one dimension. We imagine that we have a slack line tied rather tightly between two trees. At time $t = 0$ we strike one end of the line with a stick. This deforms the line by putting a pulse on it. If we parameterize the local height of the slack line above an imaginary line parallel to the ground by a function $f(x, t)$, then before we hit the line, $f(x, t) = 0$, and afterward, $f(x, t)$ is some function. We see the pulse run down the line. In our imaginary world, the pulse does not change its shape with time, but its center moves with some velocity, which I'll call c. (Technically, this is called a "non-dispersive wave," and of course c has nothing to do with the speed of light.) If that is the case, then $f(x, t) = g(x - ct)$ where g is some function. What this equation means is that at some time t_1 the value of the function at $x_i = ct_1$ is identical with the value of the function at a later time t_2, but at a location $x_2 = ct_2$. The $x - ct$ just parameterizes how the wave travels.

The second kind of wave (the sound of the gong) is an example of a "standing wave."

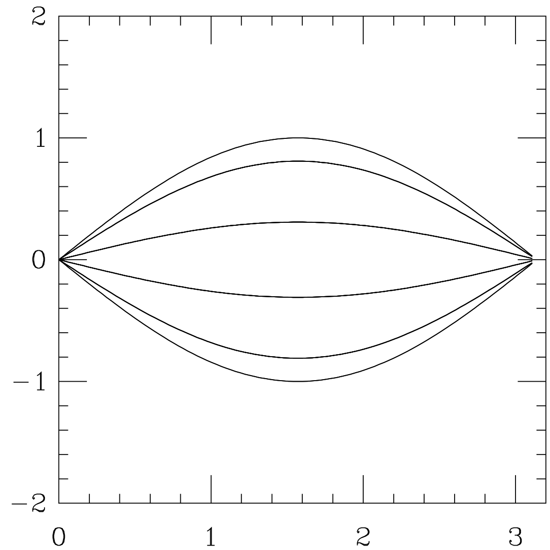

Figure 3.2: A standing wave, $\psi(x,t) = \cos(2\pi(t/20))\sin(x)$, plotted for all x and for various times.

Here the wave fills the room. If we stand in any particular place in the room, we hear a single tone – the time dependence is proportional to $\exp(-i\omega t)$ – but the amplitude of the wave, the pressure on our ear drum, will vary from place to place in the room. We might write $\psi(x,t) = f(x)\exp(-i\omega t)$ to capture this behavior.

Clearly, we could have more complicated solutions, both for traveling waves and standing waves. Two people could strike the slack line, giving (perhaps) a left moving and a right moving wave: $f(x,t) = g_2(x - ct) + g_2(x + ct)$. Our gong (think of being inside an organ pipe) could have overtones, $\psi(x,t) = f_1(x)\exp(-i\omega_1 t) + f_2(x)\exp(-i\omega_2 t)$. Indeed, there are many cases in which the whole idea of a traveling wave versus a standing wave has no meaning: as we will see, it is possible –and often convenient – to express the traveling wave as a superposition of standing waves.

Wave equations

There is a famous story about Schrödinger and one of his colleagues. You probably know, de Broglie wrote a thesis in 1923 or so, in which he argued that, just as the electromagnetic field shows particle like properties, so particle systems, like electrons bound to atoms, should show wave like properties. Anyway, Schrödinger gave a journal club talk in his department about de Broglie's work.. At the end of this talk a skeptical member of the audience (Peter Debye, look him up) said "Well this is all well and good, but if there are matter waves, what is their wave equation?" and this apparently stung Schrödinger...

"Derivation" starting with the answer

Let's return to the traveling wave solution and try to guess an equation. Recall that the wave moved away from the source at a constant velocity, and its shape did not change: the amplitude of the wave was $f(x,t) = g(x - ct)$ for some function g. Now invoke the chain rule twice:

$$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 f(x,t)}{\partial t^2} \tag{3.1}$$

This is called the "one dimensional wave equation." To go to three dimensions, we can sneak a look in a vector calculus book and write

$$\nabla^2 f(\vec{x},t) = \frac{1}{c^2}\frac{\partial^2 f(\vec{x},t)}{\partial t^2} \tag{3.2}$$

where the expression

$$\nabla^2 f(\vec{x},t) = \frac{\partial^2 f(\vec{x},t)}{\partial x^2} + \frac{\partial^2 f(\vec{x},t)}{\partial y^2} + \frac{\partial^2 f(\vec{x},t)}{\partial z^2} \tag{3.3}$$

is called the "Laplacian."

Pause for a question: Why not just one derivative? Clearly

$$\frac{\partial f(x,t)}{\partial x} = -\frac{1}{c}\frac{\partial f(x,t)}{\partial t}. \tag{3.4}$$

Answer: Eq. 3.1 also has a solution $f(x,t) = g(x + ct)$. This corresponds to a traveling wave moving to the left: at some late time, the ripple peaks at $x = -ct$. The one-derivative equation does not have this as a solution. In more than one dimension, we want to be able to describe waves going in any direction. We need something rotationally invariant to give us that. The Laplacian will work.

Our "derivation" of Eqs. 3.1-3.2 seems incredibly naive, but in fact, these wave equations are probably the ones you will encounter most often in your research. Maxwell's equations can be manipulated to give Eq. 3.2 as the defining equation for electromagnetic radiation – the function $f(\vec{x},t)$ could be one of the components of the electric field, or of the magnetic field. There c really is the velocity of light. The wave equation also describes sound waves in air. Waves in fluids are complicated, but the wave equation is usually a good starting point to describe water waves or sound waves in air.

Derivation starting with masses and springs

This is a real derivation of a wave equation. If it is too much math for you, don't worry, it won't be on the exam. But I just dislike fake derivations (like the one giving Eq. 3.1), so here goes.

Suppose we have a set of mass points sitting along a line (so it is a one-dimensional system) which are connected by springs. Look at one of the mass points, say, the ith one. It feels a potential $V(x_i) = \frac{1}{2}m\omega^2[(x_i - x_{i-1})^2 + (x_i - x_{i+1})^2]$, so it feels a force $F_i = m\omega^2[x_{i-1} - 2x_i + x_{i+1}]$, that is,

$$\frac{d^2x_i}{dt^2} = \omega^2[x_{i-1} - 2x_i + x_{i+1}]. \tag{3.5}$$

Of course, there is an equation for each mass point i. Now take a step back and imagine that in equilibrium all the mass points are spaced close together, a distance a apart. This means that that the label i corresponds to a distance $r = i$ times a. The quantity x_i tells how much the mass point is displaced from its equilibrium value. Let's call this displacement $x_i = y(r)$. Finally, if the displacements are all small, then we can just Taylor expand the displacements to write

$$x_{i+1} = y(r) + a\frac{\partial y(r)}{\partial r} + \frac{1}{2}a^2\frac{\partial^2 y(r)}{\partial r^2} \tag{3.6}$$

and

$$x_{i-1} = y(r) - a\frac{\partial y(r)}{\partial r} + \frac{1}{2}a^2\frac{\partial^2 y(r)}{\partial r^2}. \tag{3.7}$$

Notice how the first derivatives cancel in the force: we have

$$\frac{d^2y(r,t)}{dt^2} = c^2\frac{d^2y(r,t)}{dx^2} \tag{3.8}$$

where now the prefactor is $c^2 = \omega^2 a^2$. Notice how the units work out, ω is an inverse time and a is a length. The excitations of a set of masses connected by springs obeys the wave equation.

Presumably this is not a surprise – in the freshman physics halls there are long contraptions, made of point masses and springs, and it is easy to excite traveling waves down them.

A short list of examples

While you may not want to know this, there are many wave equations which you could encounter in your physics education. What you might want to know is that most of us only deal with two of them, the wave equation (Eq. 3.2, above) and the Schrödinger equation, which is (for a particle of mass m in a potential $V(x)$ in one dimension)

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = (-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x))\psi(x,t). \tag{3.9}$$

Dare I also mention the diffusion equation, the Dirac equation, the Klein-Gordon equation? Something else you might want to know: if you understand the behavior of waves that obey one particular wave equation, you probably will understand the behavior of waves which obey some other wave equation.

General features of waves

There are two particularly important features of waves, which will get you most of the way to a qualitative understanding of any situation involving wave motion.

The first point is the relation between the intensity of a wave and its amplitude. "Intensity" is a shorthand for many related quantities: the brightness of the light in the room, the apparent loudness of a sound...usually "intensity" is related to the local energy density in the wave. The rule is this: Intensity is proportional to the square of the amplitude of the wave.

Do you remember your electrostatics? Even though the static electric field is not (strictly speaking) a wave, it still obeys the intensity rule: the electrostatic energy contained in the field is proportional to the square of the electric field. (In MKS it is $(\epsilon_0/2)E^2$.) Did they teach you about the Poynting vector? The Poynting vector $\vec{S} = \vec{E} \times \vec{H}$ is a measurement of energy flux (energy per area per unit time) – never mind the cross product, it is the product of two amplitudes, one for the electric field and one for the magnetic field.

In our mechanical spring and mass example, it is even easier to see the connection. Look at one of the masses, It is (part of) a harmonic oscillator. For an oscillator, the energy is proportional to the square of the amplitude of oscillation.

The second feature is the "superposition principle:" generally, the sum of two solutions to the wave equation is also a solution. That is easy to see from our joke derivation using traveling waves. Just write $f(x,t) = g_1(x-ct) + g_2(x-ct) + \ldots$ and use the chain rule individually on each of the g_i's:

$$\frac{\partial f(x,t)}{\partial x^2} - \frac{1}{c^2}\frac{\partial f(x,t)}{\partial t^2} = 0 + 0 + \ldots \tag{3.10}$$

Like all good laws, both the amplitude-intensity rule and the superposition rule have exceptions. If the systems you are studying contain the adjective "non-linear," be prepared for surprises. But for most of us, superposing amplitudes and then squaring the result is a way of life. Let's see how to do that.

Superposing sinusoidal waves

Just writing $f(x,t) = g_1(x-ct) + g_2(x-ct)$ and computing the intensity seems rather innocuous:

$$I(x,t) = f(x,t)^2 = g_1(x-ct)^2 + g_2(x-ct)^2 + 2g_1(x-ct)g_2(x-ct) \tag{3.11}$$

But if we look closely at this expression, we see something which is not so banal: while the amplitude of the wave is just a sum, the intensity of the superposition solution is not just the sum of the individual intensities. The individual intensities are the first two terms, $g_1(x-ct)^2$ and $g_2(x-ct)^2$. But there is the cross term, the last term $2g_1(x-ct)g_2(x-ct)$. Generally, terms like this are called "interference terms." They can be quite important.

Their presence makes wave physics (superficially) very different from rock physics. In the physics of point particles, intensity is generally additive: if "intensity" means energy, and you have five particles, just add up the five energies. Of course, this is an over-simplification; our classical system of springs and masses does support wave solutions, and superposition, but it is a useful over-simplification: when you are working with waves, always be alert for the possibility of interference.

Let's look at the slack line example. Give the slack line a length L. Suppose we put a left moving pulse and a right moving pulse with the same shape on each end of the line. Then $f(x,t) = g(x-ct) + g(L-x-ct)$. The intensity is of course the square. Most of the time, an intensity profile just consists of two bumps: the interference term is zero, and the maximum intensity is just $g(0)^2$ for either bump. However, when the two pulses meet in the center, the intensity jumps to $(g(0)+g(0))^2 = 4g(0)^2$. There is one big blob of energy (temporarily) at the center of the line as the pulses pass through each other. Contrast this with the case when the two pulses have opposite sign: at one end the line is struck from above, at the other end, from below. Then $f(x,t) = g(x-ct) - g(L-x-ct)$. Again, as long as the pulses do not coincide there are two blobs of intensity $g(0)^2$ moving down the line. But look what happens when the pulses meet at $x = L/2$ – the intensity vanishes (!) and then reappears after the pulses pass through each other.

The first case is called "constructive interference" and the second case is called "destructive interference." Clearly, by choosing the two pulses to be identical, I have produced a maximally dramatic effect. If one pulse had an amplitude different from the other, there would not be complete constructive nor destructive interference.

The physics of interference is often combined with sinusoidal solutions of the wave equation, so let us specialize to that case. We also stay in one dimension.

Plane waves

The simplest example of a traveling wave is the "plane wave." This is a wave moving in a fixed direction. In one dimension, a plane wave moving to the right is described by the function

$$f(x,t) = A\cos(kx - \omega t - \phi). \tag{3.12}$$

At a fixed location in space, the wave oscillates in time with angular frequency ω, or equivalently, with a period $T = 2\pi/\omega$. At a fixed time, the wave is "everywhere," for example, at $t = 0$, $f(x,t) = A\cos(kx - \phi)$. To see it is a wave moving to the right, at time $t = 0$ put a drop of ink on the top of the wave form nearest the origin (at $x = \phi/k$). Wait a short time t. The ink drop is now at $x = (\phi + \omega t)/k$. This solution is called a "plane wave" because if it exists in more than one dimension, the wave is completely uniform in the directions transverse to the direction of its motion. For example, in three dimensions, we could write a wave function

$$f(x,y,z) = A\cos(kx - \omega t - \phi) \tag{3.13}$$

which corresponds to a wave moving in the positive x direction. Promoting the number k into a vector, a wave moving in a direction parallel to $\vec{k}$ is

$$f(\vec{r},t) = A\cos(\vec{k}\cdot\vec{r} - \omega t - \phi) \tag{3.14}$$

Since the plane wave exists everywhere, it exists for all time. If you are too literal-minded, this is obviously unphysical, but (once again) we allow ourselves the luxury of thinking about the impossible in order to simplify our lives. Obviously, a very long wave is well approximated by a plane wave.

The planarity of the plane wave seems a bit peculiar, when we think about the stone going into the water. We might imagine that the only way to make a plane wave is to have some giant antenna extending over a wall. However, there is an easier way to make a plane wave: put the source infinitely far away. With the water analogy, the wave generated by disturbing the surface of the water is an outgoing circle. Far away from the source, to any observer whose size is small compared to the distance from the source, the circle is effectively a straight line – a plane wave. To make the infinitely long wave train, we have to substitute some repetitive source for the stone we toss in.

What is k? Look again at the $t = 0$ solution: $f(x,t) = A\cos(kx - \phi)$. This is a pattern which repeats every interval λ in x of $k\lambda = 2\pi$ or (how we remember this fact) $k = 2\pi/\lambda$. λ is called the "wave length" and k is called the "wave number."

Finally, k and ω are usually related to each other, and the relation comes from whatever wave equation which the wave obeys. If that equation is "the" wave equation,

$$\frac{d^2 f(r,t)}{dt^2} = c^2 \frac{d^2 f(r,t)}{dx^2} \tag{3.15}$$

we use the chain rule and discover that $\omega = ck$. (Relations like this are called "dispersion relations.") Recalling that $\omega = 2\pi/T = 2\pi\nu$ where ν is the usual (radio dial) frequency, this is the familiar relation between frequency and wavelength $\nu\lambda = c$ where c is the velocity of the wave.

The intensity is just $A^2\cos^2(kx - \omega t - \phi)$. This is just a constant plus another oscillation, $A^2(1 + \cos(2(kx - \omega t - \phi))/2$.

Often, one's detector is not fast enough to see the time evolution in the intensity. Then we will see the time-averaged intensity,

$$I(x) = \frac{1}{T}\int_0^T dt A^2 \cos^2(kx - \omega t - \phi). \tag{3.16}$$

In this expression, T is supposed to be very long. But we can just average the intensity over one period: then the period is $T = 2\pi/\omega$. You an either grind out the integral (to build character) or just recognize: $\cos^2\ y = (1 + \cos\ 2y)/2$, and $\cos(2(kx - \omega t - \phi))$ averages to zero over one period. The time-averaged intensity is just half the square of the amplitude.

$$I(x) = \frac{A^2}{2} \tag{3.17}$$

Now suppose that our wave form consists of a superposition of two plane waves, and to make the point I want to make as sharp as possible, I will give them equal amplitude, but assume that they have a relative phase, ϕ. Just to simplify the trigonometric identities in the exciting math to follow, I'll split the phase between the two waves:

$$f(x,t) = A(\cos(kx - \omega t - \phi/2) + \cos(kx - \omega t + \phi/2)) \tag{3.18}$$

Let us look at the wave at time $t = 0$:

$$f(x,t) = A(\cos(kx - \phi/2) + \cos(kx + \phi/2)). \tag{3.19}$$

Do you recall, $\cos(a + b) = \cos\ a \cos\ b - \sin\ a \sin\ b$? Do the math, admire the cancellation,

$$f(x,t) = 2A\cos(\frac{\phi}{2})\cos(kx). \tag{3.20}$$

What have we discovered? If $\phi = 0$, the waves add constructively; the amplitude just doubles (and the intensity goes up by a factor of four). If $\phi = \pi$, the waves destructively interfere, the amplitude is zero. Since the angle corresponds to a relative phase difference between the waves is 180 degrees, all that is happening is that one wave is rising while the other one is falling. Notice as a curiosity that if $\phi = \pi/2$ (45 degrees) the amplitude of the superposition is the same as the amplitude for either wave, but the overall phase of the solution is shifted: peaks of the first wave occur at $kx - \phi/2 = 0,\ 2\pi, \ldots$, while the peaks of the superposed wave occur at $kx = 0,\ 2\pi, \ldots$. You can also give the superposition an amplitude with the opposite sign of either of the original waves.

Spherical waves

Our solution $f(x,t) = g(x \pm ct)$ describes plane waves moving to the left or right in the $\hat{x}$ direction. Often, the physical situation in three dimensions is different: we have an isolated point source producing our wave. It is easy to see what the answer is, without much mathematics. Imagine exploding a bomb at location $\vec{r} = 0$ at time $t = 0$. A pulse of energy (basically a compression wave) goes out from the detonation point, carrying all the energy of the explosion. After a time t the pulse is a giant spherical shell, of radius $r = ct$. All of the energy is contained in the shell. Usually, in this case, when we talk about "intensity" we mean something like "'energy per unit area" or "energy per unit volume." If we are a creature whose size is small compared to the distance of the wave, this will give us a quantity proportional to the pressure which which the wave will strike us. In either case, because the energy is spread out over the shell, and the surface area of the shell scales like

r^2, the intensity must fall off like $1/r^2$. The intensity is still the square of an amplitude, so this says that the amplitude of an outgoing spherical wave is

$$f(r,t) = \frac{g(r-ct)}{r} \tag{3.21}$$

Parenthetically, this is still a solution to the wave equation because in spherical coordinates the radial part of the Laplacian is

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}[rf(r,t)]. \tag{3.22}$$

Even more parenthetically, you don't see this formula in most math books, but it is still true, you can derive it.

In electromagnetism, you might recall that the electric field of a point charge is $E(r) = q/r^2$. The energy density in the electric field is proportional to E^2, but this result shows that this static electric field has nothing to do with electromagnetic radiation, the power of r is wrong. Instead, you will find that the radiation field of a moving charge scales roughly like $E \sim qa/(rc^2)$ where a is the acceleration of the charge. (Check for yourself that the units are consistent with q/r^2.) Why the acceleration and not the velocity? Basically it's a relativity thing, if E were proportional to the velocity, one could transform to a frame where the charge would be at rest, and then we would have to get back to Coulomb's law.

Question to ponder: how does the amplitude of a circular (stone in the water) wave vary with r?

Usually, when one studies interference effects, this overall factor of $1/r$ is completely unimportant even though you might think it is important: interference usually comes from radiation from two nearby antennas, which means that they are at slightly different distances from the observer. So we are interfering something like

$$f(r,t) = \frac{g_1(r_1-ct)}{r_1} + \frac{g_2(r_2-ct)}{r_2} \tag{3.23}$$

The reason it is not important is that usually interference effects happen when the separation of the sources is a few wavelengths, and the distance between the sources and observers is usually much much greater. So the difference between the two $1/r$'s falls off like $1/r^2$ and we can work with

$$f(r,t) = \frac{g_1(r_1-ct) + g_2(r_2-ct)}{r} \tag{3.24}$$

where r is the average of r_1 and r_2.

Two-slit diffraction

Let's look at interference in a two dimensional problem. Suppose you have a pair of sources radiating coherently. You could make this in several ways: you could imagine two radio antennas, fed from a common transmitter. A simpler version of this is to imagine shining a plane wave of light on a screen, in which you have cut two small holes. The light coming out of the holes came from the same source, so the two sources are in phase. Let the plane of the screen define the $x-y$ axis, and put another screen some long distance away in z, as shown in the figure. If each of the holes is tiny, it will emit basically as a point source. This means that it will emit a spherical wave, uniform in intensity in all directions (in front of the screen, of course). The amplitude from the ith source is

$$f_i(r,t) = \cos(kr_i - \omega t) \tag{3.25}$$

where r_i is the distance from the hole to the screen. So for our two holes,

$$f(r,t) = \cos(kr_1 - \omega t) + \cos(kr_2 - \omega t). \tag{3.26}$$

From the picture, if we go sufficiently far away, the rays r_1 and r_2 are parallel for all practical purposes. However, if we are looking at rays emitted at an angle θ from the normal to the surface, and if we look in the plane of the holes, the picture plus a little high school geometry tells us that $r_1 = r_2 + d\sin\theta$ where d is the spacing of the holes. This is exactly the situation we had in the last section:

$$f(r,t) = \cos(kr_1 - \omega t) + \cos(kr_1 - \omega t + kd\sin\theta) \tag{3.27}$$

we will get constructive interference if $kd\sin\theta$ is a multiple of 2π and destructive interference if $kd\sin\theta = \pi$, 3π, 5π and so on. Since $k = 2\pi/\lambda$, we get constructive interference when

$$n\lambda = d\sin\theta \tag{3.28}$$

and destructive interference in between, when

$$(n+1/2)\lambda = d\sin\theta \tag{3.29}$$

These are called the "Bragg conditions." You would observe them on the screen as a set of alternating bright and dark patches.

We can of course compute the intensity. (Somehow, nobody remembers this formula while everybody remembers the Bragg formula). Rather than squaring sines and cosines, let us pop into complex exponentials and write the wave function as the real part of

$$f = \exp(ikr_1 - \omega t)) + \exp(i(kr_1 - \omega t + kd\sin\theta)). \tag{3.30}$$

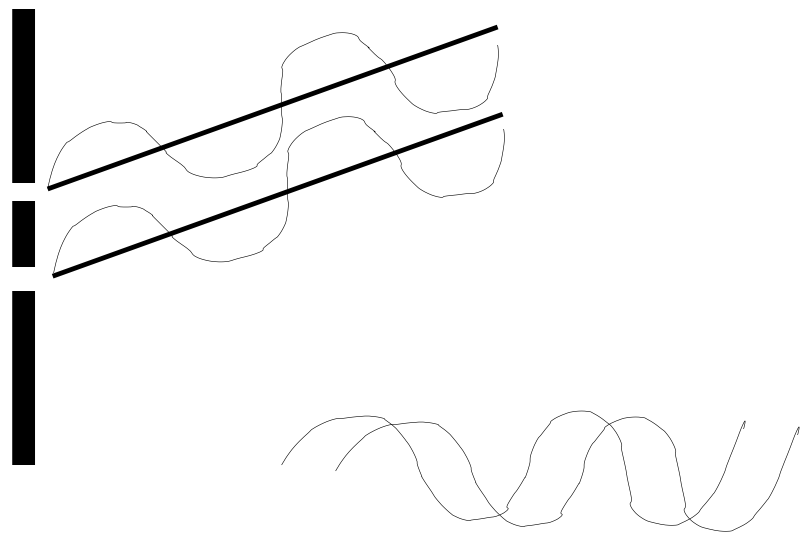

Figure 3.3: The "claassic picture" of two-slit diffraction. The wave trains from the two apertures interfere at the detector.

We can factorize this into

$$f = e^{i(kr-\omega t)}[1 + e^{i\phi}] \tag{3.31}$$

where $\phi = kd \sin\theta$. The intensity is the absolute square of f. In the expression, the first term is a pure complex exponential, so its square is one, and the second term is $[1+e^{-i\phi}][1+e^{i\phi}] = 1 + e^{i\phi} + e^{-i\phi} + 1] = 2(1 + \cos\phi)$. Again we see (it is always good to check) that when ϕ is a multiple of 2π, we get constructive interference.

If you have a gloomy mind, you have probably realized that the interference pattern will change as we move out of the plane of the two holes. The math gets annoying enough not to want to think about this. Instead, replace the holes by two slits extending out of the paper, and only look in the direction perpendicular to their orientation. Not the math and the physics is simple: we have a set of alternating bright and dark stripes on the screen.

More complicated examples

From two slits, why not many slits:

$$f = e^{i(kr-\omega t)}[1 + e^{i\phi} + e^{2i\phi} + e^{3i\phi} + e^{4i\phi} + \ldots] \tag{3.32}$$

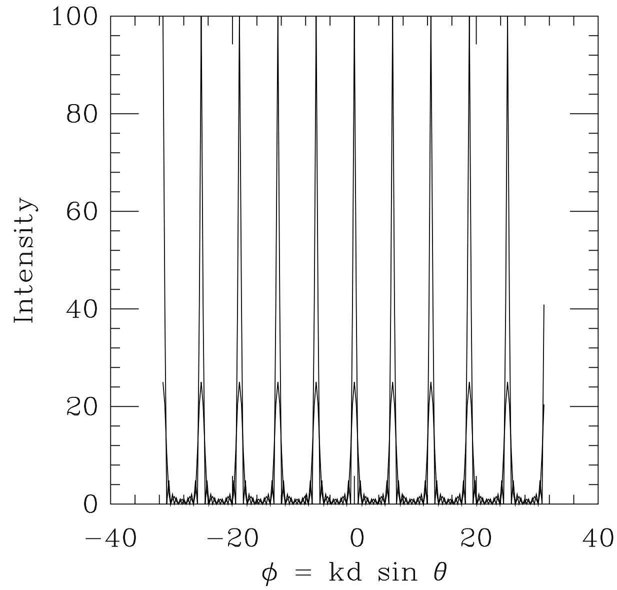

Figure 3.4: Interference pattern for 5 and 10 slits. You cannot see the 5-slit pattern under the giant spikes of the 10-slit pattern.

where $\phi = kd\sin\theta$. Call the object in the square brackets A. Then the intensity is (again) just $|A|^2$. The radiation from all the sites adds coherently when $e^{i\phi} = 1$ or $kd\sin\theta = 2\pi n$ or (once again) $n\lambda = d\sin\theta$, the Bragg formula again. What is different is the intensity. The easiest case to study is when there are an infinite number of sources. In that case, you can do the sum: define $x = e^{i\phi}$ and the series is a geometric series $1 + x + x^2 + \ldots = 1/(1-x)$. The intensity is the absolute square,

$$I = |\frac{1}{1-e^{i\phi}}|^2 = (\frac{1}{1-e^{-i\phi}})(\frac{1}{1-e^{i\phi}}) = \frac{1}{2-2\cos\phi} \tag{3.33}$$

Now when we have constructive interference, the intensity becomes very large. In this example it actually goes to infinity, but in practice, you never have an infinite number of slits. For "infinite," read "large." (You can work it out: if $\phi = 0$ and there are N slits, A in Eq. 3.32 is N and the intensity is N^2.) What you would see on the screen would be a set of sharp maxima.

I can think of three physical examples of this situation:

- Very directional radio antennas can be constructed with equally spaced arrays of wires
- We could of course have a piece of paper with many slits cut in it. Less prosaically, we could take a piece of metal, polish it, and cut parallel grooves in its surface. Then

illuminate it at an angle. Light scattering from the grooves interferes coherently. We have just built a "diffraction grating." Of course, different wavelengths show peaks at different angles ($n\lambda = d \sin\theta$ again), so by selecting an angle, you select a particular wavelength of light.

- We can replace the slits by a set of scattering sites – for example, the atoms in a regular crystal. Light (actually X-rays) could illuminate the crystal. Radiation from the different scatterers interferes. In practice, you have to consider the crystal in three dimensions, so the answer is more complicated. This is the real Bragg scattering. In this case you are presumed to know the wavelength of your radiation, and you use $n\lambda = d \sin\theta$ to measure d – to measure the lattice spacing of the atoms in the crystal.

So far we have considered only very narrow slits. One wide slit is a little more complicated, but a few moment's thought gives us the answer: Suppose the slit has a width L. Treat the slit as a whole collection of individual little slits (or sources): the scattering amplitude, the analog of $[1 + e^{i\phi} + e^{2i\phi} + e^{3i\phi} + e^{4i\phi} + \ldots]$ is

$$A = \int_{-L/2}^{L/2} \exp(iqx)dx \tag{3.34}$$

where $q = k \sin\theta$. I chose to put my origin in the center of the aperture to get a pretty integral, so I get

$$A = 2\frac{\sin qL/2}{q}. \tag{3.35}$$

You recall 'l Hospital's rule – as you take q to zero, A goes to L because $\sin\ x \sim x$. The intensity is the square of this "sinc" function. This pattern is shown in the figure: again we have Bragg minima and maxima, but they are modulated by the $1/q$.

Notice that the size of the spot is given by $qL \sim 1$ or $Lk \sin\theta \sim 1$. $k \sin\theta$ is the projection of the wave number parallel to the barrier. Calling this direction "x" means that $k \sin\theta = \Delta k_x$ is the wave number induced in the $\hat{x}$ direction by the barrier's length $L = \Delta x$. That is, $\Delta k_x \Delta x \sim 1$. For classical waves, this formula has the interpretation that restricting the size occupied by a traveling wave to be a distance Δx smears its wave number by an amount inversely proportional to Δx. A long wave train, of length L cannot have a sharp wave number; instead it must be a mixture of waves within $\Delta k \sim 1/L$ of its nominal wave number. A similar statement applies to the frequency properties of short temporal pulses: only an infinitely long wave is a pure tone.

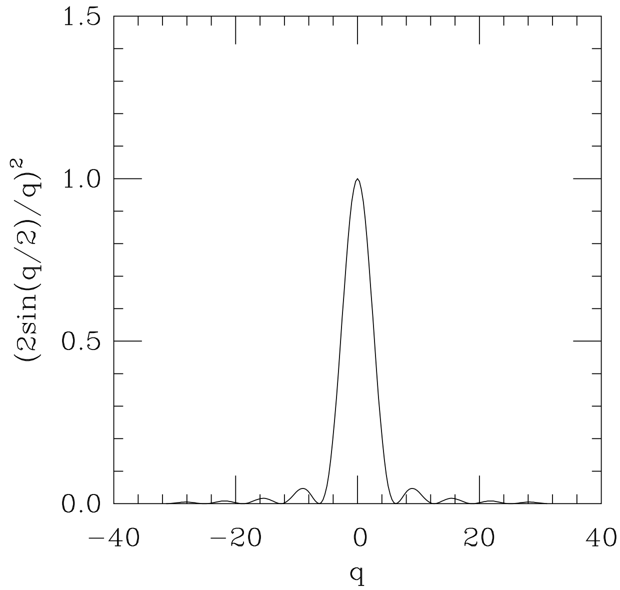

Figure 3.5: Interference pattern from difrraction through a single wide slit (L=1 in Eq. 3.35).

The same result for a the spot would occur for scattering about a solid body. We speak of "diffraction scattering" when we illuminate an object whose size is larger than, but of the same order of magnitude as, the wavelength of our light.

This beautiful subject rapidly turns into an exercise in calculus, as we turn to ever more complicated examples. The general results are basically unchanged:

- Waves arising from multiple sources typically show interference effects: the intensity of the radiation shows a set of maxima and minima.
- The location of the maxima and minima depends on the geometry of the source, but typically occurs when some characteristic size of the source is related to the wavelength of the radiation, "$n\lambda = d \sin\theta$" plus bells and whistles.
- If you know the wavelength of your radiation you can determine properties of the scatterer, and vice versa.

Chapter 4

Standing waves

Standing waves on a string

Let us consider yet another superposition situation: a pair of equal-amplitude traveling waves with the same frequency, but moving in the opposite directions:

$$f(x,t) = A(\cos(kx - \omega t) - \cos(kx + \omega t)) \tag{4.1}$$

From the usual trigonometric identities we can write the as

$$f(x,t) = 2A\sin(kx)\sin(\omega t). \tag{4.2}$$

This is an example of a "standing wave" – the wave crests do not move to the left or to the right as time evolves, instead, they remain in place, at $kx = 0$, π, 2π and so on. At every location in space the amplitude of the wave oscillates from positive to negative with frequency ω.

These waves extend infinitely far in all directions. However, we can use this solution to model a "trapped wave." Imagine that a string is extended along the x axis from $x = 0$ to $x = L$, and at these points it is pinned down. The string has a profile $y(x,t)$. Because the string is secretly a set of point masses connected by springs, the equation for the profile is actually the wave equation, so excitations of the string are amplitudes which are solutions of the wave equation. In addition, we require that our solutions have a single time dependence which is purely sinusoidal. A musician would say that the string vibrates in "pure modes." These solutions are given by Eq. 4.2 – if we arrange for $\sin(kL)$ to be equal to zero. This can only happen for special values of k, equal to π/L, $2\pi/L$, and so on. We say that k is "quantized," $k = \pi m/L$ for integer $m > 0$. Because k is quantized, and $\omega = ck$ where c is the velocity of waves on the string, the vibrational frequency is also quantized, $\omega = mc\pi/L$. Our musician would say that $\omega = c\pi/L$ is the "fundamental frequency" and that $\omega = c\pi/L$, $\omega = 2c\pi/L$, $\omega = 3c\pi/L$, are the first, second, third (and so on) harmonics. We have discovered something rather interesting:

- The pure frequencies of the string are a set of discrete levels.
- The values of those frequencies depends on the geometry of the string (in this case, its length L), "internal fundamental constants" (like the speed of sound c), plus constraints we impose (in this case, that the string is pinned at its ends).
- Of course, the amplitude can be anything – A in Eq. 4.2 is completely undetermined.

While the discussion we have given only applied to the stretched string, it is completely general – essentially all wave equations (or systems which support wave motion) which are confined to finite range regions of space have pure-frequency solutions which share these properties. Keeping the music analogy, a two-dimensional example would be the surface of a kettle drum, where the surface is pinned down, and a three-dimensional example would be a (closed) organ pipe. Finding these solutions is, of course, a complicated but standard problem which you will learn all about if you take a good course on partial differential equations. But let's look ahead a bit and see how it is done (by example).

Sometimes, there are several solutions corresponding to the same frequency, which had different spatial dependence. An example of such a solution would be the vibrational modes of a Hula Hoop. Here the rest state is just a flat circle, and the excitations would be a function $f(\theta, t)$, where θ is an angular variable which parameterizes a location around the hoop. We could imagine a mode $f(\theta, t) = A\cos(\theta)\sin(\omega t)$: the Hula Hoop warps up on one half of the hoop $(-\pi/2 < \theta < \pi/2)$ when it is moving down on the other half, $(\pi/2 < \theta < 3\pi/2)$. This is just like the stretched string, if the string has a circumference L. Therefore $\omega = 2\pi c/L$. But there is another mode with the same frequency, $f(\theta, t) = A\sin(\theta)\cos(\omega t)$, where the $(0 < \theta < \pi)$ and $(\pi < \theta < 2\pi$ modes move up and down, out of phase.

An even simpler example are the two stretched-string modes $f(x, t) = 2A\sin(kx)\cos(\omega t)$ and $f(x, t) = 2A\sin(kx)\sin(\omega t)$. The technical term for these kinds of modes is to say that they are "degenerate."

A serious solution – separation of variables

A mathematician would say that we want to solve the one-dimensional wave equation

$$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 f(x,t)}{\partial t^2}, \tag{4.3}$$

subject to a set of "boundary conditions:" imposed values of $f(x, t)$ at the surfaces of our system. In this case the "surface" is just the two points at $x = 0$ and $x = L$, plus some statement about the value of the solution and its derivative at all x at $t = 0$.

The only way I (and most people) know how to solve this problem without going to a computer involves a trick called "separation of variables." The game is to guess a solution which has a particular form, then to show that the solution works. The goal is to convert the partial differential equation Eq. 4.3 into a set of ordinary differential equations, differential equations involving a single variable, which (hopefully) can then be solved one at a time. The "separation of variables" guess is to write

$$f(x,t) = X(x)T(t) \tag{4.4}$$

i.e. a product of two solutions. We boldly plug our guess into Eq. 4.3. Because the partial derivative with respect to x only hits the function of x, and likewise for t, we have

$$\frac{\partial^2 f(x,t)}{\partial x^2} = T(t)\frac{d^2X(x)}{dx^2} \tag{4.5}$$

and

$$\frac{\partial^2 f(x,t)}{\partial t^2} = X(x)\frac{d^2T(t))}{dt^2} \tag{4.6}$$

or

$$T(t)\frac{d^2X(x)}{dx^2} = \frac{1}{c^2}X(x)\frac{d^2T(t))}{dt^2}. \tag{4.7}$$

Cross multiply to find

$$\frac{1}{X(x)}\frac{d^2X(x)}{dx^2} = \frac{1}{c^2}\frac{1}{T(t)}\frac{d^2T(t))}{dt^2}. \tag{4.8}$$

Now stare at this equation. Do you notice that the left hand side is the same for any value of t, while the right hand side is independent of x? And yet, the equation is supposed to be true for ALL x and t. The only way that can be correct, is if both sides of the equation are in fact equal to a constant. In that case

$$\frac{1}{X(x)}\frac{d^2X(x)}{dx^2} = -k^2. \tag{4.9}$$

and

$$\frac{1}{c^2}\frac{1}{T(t)}\frac{d^2T(t))}{dt^2} = -k^2. \tag{4.10}$$

Hmm...the X equation is

$$\frac{d^2X(x)}{dx^2} = -k^2X(x). \tag{4.11}$$

Where have we seen this before? Of course, it is an oscillator equation, the solution is

$$X(x) = A\sin(kx) + B\cos(kx) \tag{4.12}$$

where A and B are arbitrary constants. To find them, we have to think about our boundary conditions:

- The string is pinned to $X = 0$ at $x = 0$. We can only keep the sine solution, because $\cos\ 0 = 1$. So
$$X(x) = A\sin(kx) \tag{4.13}$$
- The string is also pinned to the origin at $x = L$. That means that $X(L) = A\sin(kL) = 0$. This can only be true if kL is a multiple of π – just as we found at the top of the chapter.

The time equation is also an oscillator equation,
$$\frac{d^2T(t)}{dt^2} = -k^2c^2T(t) \tag{4.14}$$
Its solution is $T(t) = C\sin(ckt) + D\cos(ckt)$, or $T(t) = C\sin(\omega t) + D\cos(\omega t)$ where of course ω takes on the special values $\omega_n = n\pi c/L$. So our solution is
$$f(x,t) = \sin(n\pi x/L)[A'\sin(n\pi ct/L) + B'\cos(n\pi ct/L)]. \tag{4.15}$$

That's great, we have constructed a solution in a methodical way. One thing might puzzle you, why did I call the separation constant $-k^2$? Couldn't it be anything, positive or negative? This is the "artistic" part of separation of variables. When you are comfortable with it, you can run along, mindlessly making the guess and taking derivatives. But when you hit your version of Eq. 4.11, you have to see if your solution would solve the boundary conditions. If you guessed that the separation constant was positive, not negative, you would write
$$\frac{d^2X(x)}{dx^2} = k^2X(x), \tag{4.16}$$
which has the solution $X(x) = A\cosh(kx) + B\sinh(kx)$ or $X(x) = C\exp(kx) + D\exp(-kx)$. You could get a solution which satisfies $X(x = 0) = 0$ by setting $A = 0$, or $C = -D$, but you could not then force $X(x = L)$ to be zero. So the boundary conditions force you to take $-k^2$ for the separation constant – in this case. After a while, you develop an intuition for guessing the separation constant which will work – sometimes.

Notice also, that without additional information about the wave, we have found a linear combination of two degenerate modes, the cosine (in time) one and the sine one.

Combining standing waves

Any musician will tell you, that the typical sound which comes out of a stretched string is not a pure tone. As in the case of traveling waves, a superposition of standing waves is also a solution (providing that it satisfies all the boundary conditions). For the stretched string, we cold have a mixture of the fundamental and first overtone, $k = \pi/L$ and $2\pi/L$,

$$f(x,t) = A_1 \sin(\pi x/L) \sin(c\pi t/L) + A_2 \sin(2\pi x/L) \sin(2c\pi t/L). \tag{4.17}$$

And we don't have to stop there, we could combine any number of normal modes,

$$f(x,t) = \sum_n c_n f_n(x,t) \tag{4.18}$$

with arbitrary coefficients c_n.

There are two complementary ways of dealing with Eq. 4.18:

- Reading the right hand first, if you know all the $f_n(x,t)$'s, it appears that you can construct an almost arbitrarily shaped wave by using a carefully chosen set of c_n's. So, if we know the c_n's, what is $f(x,t)$?
- Reading the left hand first, we can imagine that our string has some almost arbitrarily shape imposed on it, like the shapes we got when we stretch the slack line with a stick. But, in fact, that arbitrary shape is really composed of a linear combination of standing waves – the amount of standing wave mode $f_n(x,t)$ in the amplitude is given by the coefficient c_n. So, if we know $f(x,t)$, what are the c_n's?

To sharpen this question, let us reword it a bit. Also, let us drop down to functions of a single variable, which I will take to be the coordinate, x. Clearly, you could do this for t as well. Or x and t together. This gets messy. So, sticking with one variable, x,

$$f(x) = \sum_n c_n f_n(x). \tag{4.19}$$

Suppose the functions $f_n(x)$ were as different as they could possibly be. Then Eq. 4.19 says in words: each term in the sum contributes a piece to $f(x)$ which is unique, the piece with one n is completely different from the piece with a different n. The whole thing smells like writing a vector as the sum of its components, $\vec{F} = \hat{x} f_x + \hat{y} f_y + \hat{z} f_z$. Each term in the vector sum contributes something unique, which the other terms cannot – the amount of x-component in the vector. The final answer, $\vec{F}$, is a vector, pointing some direction in space. We can think about it (mostly) without having to specify a coordinate system. And, of course, if we want to know how much of $\vec{F}$ comes from its x-component, we can just dot $\vec{F}$ into $\hat{x}$: $f_x = \hat{x} \cdot \vec{F}$.

For the vector case "as different as it could possibly be" means that the unit vectors pointing in different directions are orthogonal, $\hat{x} \cdot \hat{y} = 0$. What does this mean for a function $f(x)$? The only possible choice has to be related to an integral over the allowed range of our functions, something like

$$\int_0^L f_n(x) f_m(x) dx = 0 \tag{4.20}$$

if $m \neq n$. (In fact, there are many possible choices like this, one can include some additional function of x in the integrand.) For if that is the case, we can take Eq. 4.19, multiply both sides by one of the $f_n(x)$'s, say $f_m(x)$, and integrate:

$$\int_0^L f_m(x) f(x) = \sum_n c_n \int_0^L f_m(x) f_n(x) dx \tag{4.21}$$

But on the right hand side, the integral is zero unless $n = m$. This collapses the sum to a single term, and we can determine c_m, it is

$$c_m = \frac{1}{I_m} \int_0^L f_m(x) f(x) \tag{4.22}$$

where $I_m = \int_0^L f_m(x)^2 dx$. This means that, given an arbitrary wave form, we can decompose it into its components.

This is a great miracle. And unlike many miracles, it is actually useful: If you have a set of magic functions, the $f_m(x)$'s, you can study ANY wave form using them, via

$$f(x) = \sum_n c_n f_n(x) \tag{4.23}$$

by doing all your manipulations on the right hand side. This is another example of the "knowing less is better" approach to physics, you simply learn all the properties of the f_n's and then you can do anything.

There is a further miracle, the f_n's generally correspond to "pure tone" solutions of wave equations. For example in our situation, we could study the equation

$$\frac{d^2 f_n(x)}{dx^2} = -k^2 f_n(x). \tag{4.24}$$

subject to the boundary conditions that $f(x) = 0$ if $x = 0$ or $x = L$. These solutions are (as we found several times already) $f_n(x) = A \sin k_n x$ with $k_n = n\pi/L$. Observe that if $n \neq m$,

$$\int_0^L \sin(\frac{n\pi x}{L}) \sin(\frac{m\pi x}{L}) = 0. \tag{4.25}$$

This says that the pure tones form a set of functions which are as different as they can possibly be, according to the criterion of Eq. 4.20.

As a final miracle (possibly the most practical comment I can make): people have been solving problems this way for the better part of 200 years, so essentially every miraculous function you will ever need has been worked out, its properties tabulated, and cataloged. Apart from the sine and cosine, these functions are named after dead white male nineteenth century mathematicians and physicists (Bessel functions, Legendre polynomials, Gegenbauer polynomials ...). To me, they are as similar as different Romance languages are similar; when I need one, I learn about it and then forget what I knew until next time. To join the church of miracles, you have to learn the holy language: the miraculous functions are called "orthogonal polynomials," and the miraculous relation Eq. 4.20 is called an "orthogonality relation." (Here "orthogonal" carries the same meaning that it does for two vectors whose dot product is zero.) Often times, it is convenient to define the constant in front of the orthogonal polynomial so that the function is "normalized:" This means that

$$\int dx f_m(x) f_m(x) = 1 \tag{4.26}$$

and

$$\int dx f_m(x) f_n(x) = 0 \tag{4.27}$$

if $n \neq m$.

Equations like Eq. 4.24 are called "eigenvalue equations." The word "eigenvalue" is a mixed German-English construction, "eigen" meaning "characteristic." Notice that Eq. 4.24 only has solutions for special ("characteristic") values of $-k^2$. In our case, we had $k = \pi m/L$ for any $m > 0$. These characteristic values are called (unsurprisingly) "eigenvalues," and their solutions are also called "eigenfunctions" or "eigenmodes." Numbers which label the eigenvalues (the subscripts n and m in this case) are usually called "quantum numbers."

Our example of the modes of a stretched string is an example of a mathematical construction called a "Fourier series." Let's list what we want to do for that special case. We are interested in describing an arbitrary wave form $f(x)$ which is designed to be zero at $x = 0$ and $x = L$. We take our eigenfunctions to be the sine waves,

$$f_n(x) = A \sin(n\pi x/L) \tag{4.28}$$

It is convenient to normalize them, so let us find A:

$$1 = A^2 \int_0^L dx \sin^2(n\pi x/L) \tag{4.29}$$

The integral is $L/2$ by inspection, or write $\sin^2(n\pi x/L) = (1-\cos(2\pi n x/L))/2$. The integral from zero to LL always goes over an integer number of full wavelengths of the second term and gives zero. So $1 = A^2L/2$, $A = \sqrt{2/L}$. So any arbitrary wave form can be written as

$$f(x) = \sum_n c_n \sqrt{\frac{2}{L}} \sin(n\pi x/L) \tag{4.30}$$

You could at this point encounter either of our two examples. First, you secretly know all the c_n's, and you want the wave form $f(x)$. Plug in the c_n's, do the sum (a knowledge of programming might help) and there you are.

Alternatively, you know $f(x)$ but you do not want to work with it. You prefer dealing with sine waves. Multiply both sides of Eq. 4.30 by one of the eigenmodes, and integrate

$$\int_0^L f(x)\sqrt{\frac{2}{L}}\sin(m\pi x/L)dx = \sum_n c_n \int_0^L dx\sqrt{\frac{2}{L}}\sin(n\pi x/L)\sqrt{\frac{2}{L}}\sin(m\pi x/L) = c_m \tag{4.31}$$

If you can do the integral, you can find each c_m and reconstruct the series.

In our example, $f(x)$ is the amplitude ot the wave at location x. Since we are dealing with one variable, we could also imagine a problem of time evolution: $f(t)$ could be the location in time of a particle under some arbitrary force, which we have engineered to return to the origin after a time T. The eigenfunctions are sine waves in time. What do they correspond to, physically? Harmonic oscillator motion, again. Everything is a sum of simple harmonic oscillators!

All miracles have a dark side: in practice, this elegant story is only useful if a small number of c_m's are large, and the rest are small. This may mean that you have a miraculous formal story, which does not help much in real life situations. But like all purveyors of miracles, I choose not to dwell on that point.

Also, in the last chapter, we talked about traveling waves, but never thought about putting any restriction on the frequency or wave number. We can connect that discussion with the present one if we simply imagine that everything in the last chapter happened in a really big box. Remember, the wave number is $k = n\pi/L$ so the difference in wave numbers between successive eigenvalues is $\Delta k = \pi/L$. This obviously vanishes as L goes to infinity. In the mathematics literature, in this limit, Fourier series turn into Fourier transforms, and overall, there are small changes in all the formulas, like sums turning into integrals. But these are all just technical details which you can learn about in a math class – or on your own.

Chapter 5

All the important points, summarized at the end

Let's return to the question I asked at the beginning: When we say that a system exhibits classical wave behavior, what do we really mean?

- Usually, the system is characterized by dynamical variables which are defined for all locations and times, bundled together into an object called a "field" and labeled (generically) as $\psi(\vec{x}, t)$.

- Usually, the spatial and temporal behavior of ψ is given by the solution of a partial differential equation, a "wave equation.'

- Usually, the sum of two or more solutions to the wave equation is also a solution.

- Usually, the solutions of the wave equation, $\psi(\vec{x}, t)$, are not the final objects measured in an experiment: that final object, the intensity, is usually the square of $\psi(\vec{x}, t)$. When this occurs, you should expect to see interference effects: If $\psi(\vec{x}, t) = c_1 f_1(x, t) + c_2 f_2(x, t)$, then $I(x, t) = c_1^2 I_1(x, t) + c_2^2 I_2(x, t) + c_1 c_2 I_{int}(x, t)$ where $I_1(x, t) = f_1^2(x, t)$, $I_2(x, t) = f_2^2(x, t)$, and $I_1(x, t) = 2 f_1(x, t) f_2(x, t)$.

- Usually, the combination of wave equation plus boundary conditions means that there will be a set of solutions which are "pure tones" or eigenfunctions, and these solutions will only occur for particular values of the frequency or wave number, the eigenvalues. Typically, eigenfunctions with different quantum numbers are orthogonal, and any arbitrary wave form can be decomposed into a linear superposition of eigenfunctions.

- Usually, the intensity of the wave can take on any value, but boundaries generally force the values of the eigenvalues to be restricted. "Eigenvalues are quantized."

One last comment: these notes were written to describe generic wave behavior. As a physics student, you are certain to encounter waves in the vibrational modes of mechanical systems, in classical electrodynamics, and in quantum mechanics. When you get down to the details, all these systems are slightly different one from each other and in many cases the details matter. For example, in the first two cases, the physical degrees of freedom, the coordinates of the parts of the mechanical systems or the electric and magnetic fields, are real. We might study them using complex variables just to simplify formulas, but "the imaginary part of the complex electric field" is just a mathematical convenience. In quantum mechanics, the physical wave function is complex. This actually makes the equations simpler, at the cost of making the physics more abstract. But that's just life. Again, it is like knowing

Latin and having to speak Spanish. You may want to pay close attention to the endings on your verbs!