Physics 214 Midterm Review Notes

Anne C. Hanna
ahanna@uiuc.edu

June 6, 2005

Contents

1 Disclaimers and usage notes

These midterm notes are not necessarily complete, correct, or even useful. And even if they were, they shouldn't be your only review materials. Look through the text and the lecture notes, and try some practice exams too, please! Anything you've had to apply in this course, whether it be in homework, discussion packets, quizzes, prelabs, labs, practice exams, or lecture ACTs should be considered fair game for the exam.

Also, I would advise not becoming too dependent on the equation sheet during your studying. It has valuable information on it, but it provides this information without context. It is *much* better to have a solid understanding of the concepts and the derivations which stand behind the equations, because then you can easily handle unfamiliar problems, or derive the correct equation for your purposes on the fly.

Okay, rant over.

2 Everyday waves: the wave equation

2.1 Basics

Many of the waves we see in the everyday world and use as little toy physics lab problems can be described by a relatively "simple" differential equation called the wave equation. This equation describes a disturbance $\Psi(x,t)$ whose size varies depending on where and when you measure it, and is commonly written as follows:

$$\frac{\partial^2 \Psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \Psi}{\partial t^2} \tag{1}$$

(Don't be intimidated by the $\partial/\partial x$ things if you haven't seen them before — they just mean you should take the derivative of the function Ψ only with respect to, *eg.*, x, and treat all other variables, such as t, as if they were constant. So if $\Psi(x,t) = 3x + 4t$ then $\partial\Psi/\partial x = 3$ and $\partial\Psi/\partial t = 4$.) The parameter v is a constant and is usually computed from the specific characteristics of the system in question. Examples of systems which satisfy this wave equation include waves on a string, sound waves, and light waves (also known as photons).

Any function of the form $f(x - vt)$ can be a solution to this equation, where f will be determined by the initial conditions. If we know that a $t = 0$ snapshot of the system looks like $f(x)$, then at any later time, it will look like:

$$\Psi(x,t) = \begin{cases} f(x - vt) & \text{if the wave propagates to the right} \\ f(x + vt) & \text{if the wave propagates to the left} \end{cases} \tag{2}$$

So the solutions to this equation actually look like solid objects moving along at some velocity v, which is called the propagation velocity. If we were able to somehow mark a particular point on a wave, we would discover that as time moves forward, the mark simply moves in the propagation direction at the propagation velocity. For example, suppose my system was set up at $t = 0$ as $\Psi(x,0) = e^{-x^2}$, and it propagated to the right at a velocity v. The initial shape of this graph will be a bell curve with its peak at $x = 0$ and a height of 1. At later times the wave would look like

$\Psi(x,t) = e^{-(x-vt)^2}$, and would be an identical bell curve with its peak located at $x = vt$ and a height of 1. The only difference from the original state would be the position of the peak.

The propagation velocity is typically a property of the specific medium, and all waves in that medium will have the same propagation velocity. So, for example, if I specify a particular mass per unit length and tension for a string, or a particular density and composition for air, I will know the speed of waves on the string, or the speed of sound propagation in the air. Light in a vacuum always travels at the same speed, $c = 3 \cdot 10^8$ m/s. (Light in a medium will travel more slowly, depending on the refractive index of the medium, but see Ph212 for that stuff.)

2.2 General properties of sinusoidal solutions

A special type of solution to this equation is one shaped like a sine or cosine function. These solutions constitute what we normally think of as waves, and have the following form:

$$\Psi_c(x,t) = A\cos(\pm kx \pm \omega t + \phi_0) \tag{3}$$

$$\Psi_s(x,t) = A\sin \pm kx \pm \omega t + \phi_0) \tag{4}$$

Since a sine function is just a cosine function with a phase shift ($\sin(x) = \cos(x - \pi/2)$) we can ignore the sines and simply think of all of these solutions as cosines, which I will do from here forward. A solution of this sort has several important parameters:

the amplitude A This measures the "size" of the disturbance – in particular, this is the height of the peaks (and the depth of the valleys) in the cosine function. For a vibrating string, for example, this is the maximum height any part of the string will ever be at (as compared to the equilibrium state of the string, which is just a straight line). So for strings (unlike sound and light waves) the amplitude is an easily-measurable physical parameter.

the intensity $I = A^2$ For some systems, such as light and sound, we are more interested in the energy transmitted by the waves than in the amplitude of the waves themselves. Specifically, we will often have a detector of a certain area, and we will want to know how much energy it receives per second as a function of its size. So we measure the energy per second per unit area ($\mathrm{J/s/m^2} = \mathrm{W/m^2}$). For light and sound waves, this corresponds to the wave intensity, which is equal to the square of the amplitude of the wavefunction. Unfortunately, this means that the units of amplitude for sound and light waves are the extremely awkward $\sqrt{\mathrm{W/m^2}}$ which just doesn't come out to anything nice. But this is okay, because amplitude for light and sound waves is not something we can easily measure anyway.

the wavelength λ The wavelength is defined as the spatial extent of a single full "cycle" of the sinusoid. A full wavelength is the distance from one peak to the next, or one trough to the next. The peak-to-trough distance (or the distance between two adjacent zeros of the wavefunction) is half of a wavelength. Wavelength has units of distance, sometimes written as meters per cycle.

the wavenumber k The wavenumber is *by definition* related to the wavelength as $k = 2\pi/\lambda$. One cycle of the wave is treated as being 2π radians, so k is notionally the number of radians per meter for this particular wave. It has units of rad/m, sometimes written as 1/m or m^{-1}.

the period τ If I stand a fixed point in space, the period is the time it takes from just after one wave peak passes me until the next reaches me. The units of period are seconds, or sometimes seconds/cycle.

the frequency f The frequency tells me how many cycles flow past me per second as I stand at a fixed point in space. It is defined as $f = 1/\tau$ and has units of s^{-1}, also sometimes called Hertz (Hz) or cycles/second.

the angular velocity (sometimes angular frequency or just frequency) ω This frequency is similar to the wavenumber, in that it relates radians to physical coordinates — since one cycle is 2π radians, it tells me how many radians flow past me every second. It is defined as $\omega = 2\pi f$ and has units of radians per second, sometimes written simply as s^{-1}.

the phase shift ϕ_0 If $\phi_0 = 0$ then at $t = 0$ this function will simply be a straight $A\cos(kx)$ shape. Otherwise, the phase shift tells us how far offset the wavefunction is from a normal cosine. For example, $\phi_0 = -\pi/2$ turns it into a sine-like wave. Be careful when solving for phase shifts — often two different phase shifts (eg $\phi_0 = \pm\pi/3$) will result in the same value for the wavefunction at $x = 0$, but only one will describe the actual shape of the graph. It's safest to solve for phase shifts of peaks or troughs, since these will provide unique solutions (well, modulo 2π). ϕ_0 generally has units of radians, but occasionally it will be given to you in degrees. You will usually be best off if your calculator is in radians mode all the time, and you should work in radians as much as possible to avoid later confusion. So at the start of a problem, convert any degrees given to radians, and work the whole problem in radians. At the end, if you need an answer in degrees, convert back. This will save you an incredible amount of hassle. (And besides, all the cool scientists work in radians. And you want to be cool, don't you?)

the overall phase $\phi(x,t)$ If you stand at the point x at time t the overall phase tells you exactly where on the wave you will be located. For example, if there's a peak at that point at that time, then you know $\phi(x,t) = 0$ (or 2π or 4π or whatever). $\phi(x,t)$ is the argument of the cosine function, that is $\phi(x,t) = kx \pm \omega t + \phi_0$. As before, the overall phase is in radians, or should be.

the relative sign of the temporal and spatial dependence If the wavefunction has the form $\Psi(x,t) = A\cos(kx + \omega t + \phi_0)$ or $\Psi(x,t) = A\cos(-kx - \omega t + \phi_0)$ (that is, the space and time terms have the same sign), then the wave is propagating to the left. If the wavefunction has the form $\Psi(x,t) = A\cos(kx - \omega t + \phi_0)$ or $\Psi(x,t) = A\cos(-kx + \omega t + \phi_0)$ (the space and time terms have opposite signs), then the wave is propagating to the right.

Provided two solutions satisfy the exact same linear differential equation (*eg.* both are solutions of the wave equation with identical values of v), then if the corresponding disturbances appear in the same system at the same time, we can find the system's state by simply adding the wavefunctions together. This is is because the wave equation (and the Schroedinger equation and other linear differential equations) allow simple linear superposition of their solutions. The easiest example of this is adding together two cosine waves with the same ω and k values and identical amplitudes A, but possibly different phase shifts (ϕ_1 and ϕ_2), which gives the following wave:

$$\begin{aligned}
\Psi(x,t) &= \Psi_1(x,t) + \Psi_2(x,t) = A\cos(kx - \omega t + \phi_1) + A\cos(kx - \omega t + \phi_2) \\
&= 2A\cos\left(\frac{\phi_2 - \phi_1}{2}\right)\cos\left(kx - \omega t + \frac{\phi_1 + \phi_2}{2}\right)
\end{aligned}$$

This corresponds to, for example, playing two different sounds of the same frequency and amplitude, but out of phase with each other. Notice that the wavenumber and angular frequency for the combined sound remain the same as those for the original sounds, but the amplitude becomes:

$$A' = 2A\cos(\left(\frac{\phi_2 - \phi_1}{2}\right) \quad (5)$$

and the phase shift becomes:

$$\phi_0' = \frac{\phi_1 + \phi_2}{2} \quad (6)$$

2.3 Properties specific to wave equation sinusoids

The above properties apply to any sine or cosine type solution, whether it arises from the wave equation or not. However, there is one further property which is important and related to these but is not general. If you plug a sinusoidal solution into the wave equation, you discover that the wave equation becomes $-k^2\Psi(x,t) = -(\omega^2/v^2)\Psi(x,t)$. So in order for a sinusoid to be a valid wave equation solution, it must be true that $\omega^2 = v^2k^2$. With the assumption that ω, v, and k are all positive, we can then write what is called the *dispersion relation* for this wave equation:

$$\omega = vk \quad (7)$$

A dispersion relation tells us how ω (which is related to the system's time dependence) depends on k (which is related to the system's spatial dependence). Dispersion relations are typically unique to their differential equation and should not be expected to apply to solutions of any other differential equation. In particular, $\omega = vk$ is only true for things like sound waves, light waves, and waves on a string, *not* for massive particles (electrons, protons, neutrons, baseballs, *etc.*), since massive particles satisfy the Schroedinger equation, not the wave equation.

We can use the wave equation dispersion relation to see that the left-propagating sinusoids are of the form $\Psi(x,t) = A\cos(\pm k(x+vt)+\phi_0) = f(x+vt)$ while the right-propagating waves are of the form $\Psi(x,t) = A\cos(\pm k(x-vt)+\phi_0) = f(x-vt)$, where $f(x) = A\cos(\pm kx+\phi_0)$. So with the appropriate ω-k relationship, these solutions do indeed match what we know about the wave equation. Note that by substituting $\omega = 2\pi f$ and $k = 2\pi/\lambda$ into the dispersion relation, we can get the more familiar form

$$v = \lambda f \quad (8)$$

3 Interference

Interference is what happens when at least two correlated waves which satisfy the same differential equation interact. The only interference we will discuss in this course is interference between simple sinusoidal waves. The results of this kind of interference depend only on the amplitudes and frequencies of the waves we are interfering, and on their relative phase at the point where they interfere. It does not matter what type of system we are using, as long as the waves we are interfering are of the same type. Typical interference systems include sound waves, light waves (photons), electrons, protons, and even large molecules like buckyballs.

3.1 Two sources

The simplest version of this situation is as follows: We have two identical point sources (usually two speakers driven by the same amplifier or light from one laser going through two pinholes), which are driven to emit waves with the same initial phase and the same amplitude. It is important that these waves be correlated – which means that they have to have the same frequency and they need to have a *specific* relationship between their initial phases. With light, for example, both source pinholes need to get their light from the same laser, because this is the only way that we can ensure that there will be some consistent relationship between the two light beams. If they have no particular connection, all bets are off.

Anyway, we also assume the waves propagate perfectly, so that as long as we're reasonably near the sources the amplitude of the signal we detect from a single source (with the other turned off) does not depend on our location. The sources are separated by a distance d. Then we place a detector at a point a distance r_1 from the first source and r_2 from the second.

If the two sources are driven in phase, then the overall phases at the sources will be

$$\phi_1(0,t) = \phi_2(0,t) = -\omega t \tag{9}$$

while the phases of the two signals at the detector will be:

$$\begin{aligned} \phi_1(r_1,t) &= kr_1 - \omega t \\ \phi_2(r_2,t) &= kr_2 - \omega t \end{aligned}$$

3.1.1 Identical nearby sources

If the sources have identical amplitude $I_0 = A_0^2$, the value of the wavefunction as measured at the detector will be:

$$\psi(t) = 2A_0 \cos\left(\frac{k}{2}(r_2 - r_1)\right) \cos\left(\frac{k}{2}(r_1 + r_2) - \omega t\right) \tag{10}$$

So the wave at the detector will have amplitude $A = 2A_0 \cos(k(r_2 - r_1)/2)$, and the measured intensity at that point will be:

$$I = A^2 = 4I_0 \cos^2\left(\frac{k}{2}(r_2 - r_1)\right) \tag{11}$$

The intensity is usually the variable of interest in these problems as we are usually dealing with light or sound waves (or with probability density for massive particles, to be discussed later).

We can think of this more generically in terms of the wavelength $\lambda = 2\pi/k$ and the path length difference for the signal from the two different sources $\delta = r_2 - r_1$. The different path lengths for the two signals cause the detected signals to have a phase difference

$$\phi = k(r_2 - r_1) = \frac{2\pi\delta}{\lambda} \tag{12}$$

and we can rewrite the detected intensity as

$$I = 4I_0 \cos^2\left(\frac{\phi}{2}\right) \tag{13}$$

Notice that if the path length difference is an integer number of wavelengths ($\delta = n\lambda$), then the phase difference at the detector is $\phi = 2n\pi$ and so the cosine of $\phi/2$ will be ± 1, which raises the detected intensity to its maximum possible value of $4I_0$. Alternately, if the path length difference is a half-integer number of wavelengths ($\delta = (n+1/2)\lambda$) then the phase difference is $\phi = (2n+1)\pi$ and so the cosine of $\phi/2$ will be zero and thus the detected intensity will be zero, which is the minimum possible value. (This can be generalized for sources with different initial phases by simply adding their initial phase offset $\phi_{21} = \phi_2 - \phi_1$ to the phase offset due to differing path length, so that $I = 4I_0 \cos^2((\phi + \phi_{21})/2)$.)

3.1.2 Identical distant sources

If the sources are very far away from the detector, we can make certain simplifying approximations. We arrange our coordinates so that the sources are located at $x = 0$ and $y = \pm d/2$ where d is the distance between the sources. The detector is located at $x = L$ and $y = \Delta y$. A line drawn straight from the origin to the detector makes an angle θ with the x axis. For very distant sources, we can treat the path length difference δ as being very small compared to the path lengths, so we can make the approximation that the it satisfies:

$$\delta \approx d \sin(\theta) \tag{14}$$

As before, the phase difference (for sources with the same initial phase) will be:

$$\phi = \frac{2\pi\delta}{\lambda} \tag{15}$$

so the maximum intensities will occur where

$$n\lambda = \delta \approx d \sin(\theta_{\mathrm{max}}) \tag{16}$$

while the minimum intensities will occur halfway in between the maxima, at angles:

$$\left(n + \frac{1}{2}\right)\lambda = \delta \approx d \sin(\theta_{\mathrm{min}}) \tag{17}$$

If in addition, θ is very small, we may further approximate that $\cos(\theta) \approx 1$ and $\sin(\theta) \approx \theta$ (where θ is measured in radians). This gives us the following relationships:

$$\begin{aligned} \delta &\approx d\theta && (18) \\ \Delta y &= L\tan(\theta) \approx L\sin(\theta) \approx L\theta && (19) \\ n\lambda &\approx d\sin(\theta_{\mathrm{max}}) \approx d\theta_{\mathrm{max}} && (20) \end{aligned}$$

where the last equation describes the locations of the maximum intensities.

3.1.3 Non-identical sources (phasors and the law of cosines)

If our sources do not have the same amplitude, we must use a more general method, that of phasors. In this method, the signal from a particular source at a particular location and point in time is represented by a vector whose length is equal to the *amplitude* of the source (the *square root* of its intensity), and whose angular position is equal to the phase of the sound from that source at that position and time. So a source with intensity $I = A^2$ and phase ϕ at the location and time of interest would be represented by the following phase vector:

$$\vec{A} = A\cos(\phi)\hat{x} + A_1 \sin(\phi)\hat{y} \tag{21}$$

Phasors rotate counterclockwise about the origin in one full oscillation period of the wave. That is, if the phasor is pointing straight to the right at $t = 0$, then at $t = T = 1/f = 2\pi/\omega$ it will again point straight right. If you have light from two different sources interfering at a particular point, then the total intensity of light at that point may be found by performing vector addition on the light from the two sources and squaring the result:

$$\begin{aligned}
\vec{A}_{\mathrm{TOT}} &= \vec{A}_1 + \vec{A}_2 = \left(A_1\cos\left(\phi_1\right) + A_2\cos\left(\phi_2\right)\right)\hat{x} + \left(A_1\sin\left(\phi_1\right) + A_2\sin\left(\phi_2\right)\right)\hat{y} \\
I_{\mathrm{TOT}} &= \vec{A}_{\mathrm{TOT}} \cdot \vec{A}_{\mathrm{TOT}} \\
&= A_{\mathrm{TOT}}^2 = A_1^2 + A_2^2 + 2A_1A_2\cos\left(\phi_2 - \phi_1\right) \\
&= I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left(\phi_2 - \phi_1\right)
\end{aligned}$$

The second line of the I_{TOT} formula is actually a modified version of the law of cosines as applied to the triangle whose sides are the vectors $\vec{A}_1$, $\vec{A}_2$, and $\vec{A}_{\mathrm{TOT}}$. Recalling that the usual law of cosines is $c^2 = a^2 + b^2 - 2ab\cos C$, where a, b, and c are the side lengths and C is the internal angle opposite side c, we see that if we choose $a = A_1$, $b = A_2$, and $c = A_{\mathrm{TOT}}$ then the internal angle C is actually π minus the angle between the vectors $\vec{A}_1$ and $\vec{A}_2$. So $C = \pi - (\phi_2 - \phi_1)$, and using $\cos(\pi - \theta) = -\cos(\theta)$ gives the formula we found above.

3.2 More than two sources

3.2.1 Non-identical sources

With several non-identical sources, the method is very similar to that for two non-identical sources and is (conceptually, at least) very simple. Simply add all the phase vectors together (making sure to use amplitudes as the vector lengths and *not* intensities). Then square the total phase vector to get the final intensity. You can also use successive applications of our modified law of cosines.

The only thing to worry about is getting the relative angles of successive phasors correct. The relative phase of two phasors is defined as the phase of the second (ϕ_2) minus the phase of the first (ϕ_1). Sometimes you will be told the relative phase of two phasors ($\phi_2 - \phi_1$) and sometimes you will be told their individual phases (ϕ_1 and ϕ_2). Be aware of which you've been given!

Two phasors which have a *relative* phase of zero will point in the same direction, with the point of the first touching the tail of the second. So the total amplitude of their sum will simply be $A_1 + A_2$. Two phasors with relative phase of π (180°) will point in opposite directions, again head-to-tail (*not* tail-to-tail), and the total length of the sum will be $A_1 - A_2$. Phasors with a relative phase of $\pi/2$ (90°) will be perpendicular. If the first one points straight to the right, the second will point straight up. Their sum will have a length of $\sqrt{A_1^2 + A_2^2}$ and will point up and to the right. Phasors with a relative phase of $3\pi/2$ or $-\pi/2$ (270° or −90°) will also be perpendicular, but if the first points straight to the right, the second will point straight down. Their sum will have a length of $\sqrt{A_1^2 + A_2^2}$ and will point down and to the right.

Remember that relative phases, just like absolute phases, should always be measured counterclockwise. The difference is that absolute phases are always measured starting from the *direction* of the positive x-axis (not necessarily the axis itself), while relative phases are measured starting from the *direction* of the first phasor (not from the first phasor itself).

3.2.2 Identical sources

The phasor methods described above can also be applied to sources of identical amplitude which are arranged in an unusual set of positions, or which have differing initial phases. But if we have N identical sources with the same initial phase arranged in a regularly-spaced row (distance between adjacent sources d) and we place the observer a long distance L away from the row of sources, then the approximations from section 3.1.2 above apply and we can write some relatively simple formulas to describe the observed interference pattern.

In this case, we define ϕ to be the relative phase of the signal from any pair of adjacent sources in the row, at the observation point. I_N is the total intensity of the signal seen at that point. $I_{\max} = A_{\max}^2$ is the maximum possible value of $I_N = A_N^2$ for the setup in question, and $I_1 = A_1^2$ is the peak intensity of the light from a single source with all the other sources turned off. The maximum possible intensity for this system occurs when all the sources are in phase, so:

$$I_{\max} = A_{\max}^2 = (NA_1)^2 = N^2 I_1 \tag{22}$$

So the total intensity of the light from this arrangement seen at a particular point will be:

$$I_N = I_1 \left(\frac{\sin(N\phi/2)}{\sin(\phi/2)} \right)^2 = \frac{I_{\max}}{N^2} \left(\frac{\sin(N\phi/2)}{\sin(\phi/2)} \right)^2 \tag{23}$$

As before, the relative phase ϕ will be related to the geometric angle θ at which the observer is located by

$$\phi = \frac{2\pi\delta}{\lambda} \approx 2\pi \frac{d}{\lambda} \sin(\theta) \tag{24}$$

and so the observer location corresponding to a particular phase difference will be $x = L$, $y = \Delta y \approx L\sin(\theta) \approx L \cdot \phi\lambda/2\pi d$. So, for small geometric angles, the shape of the intensity graph is the same whether we're graphing intensity *vs.* phase difference, intensity *vs.* observer geometric angle, or intensity *vs.* observer y-axis location. Only the scale changes.

The intensity graph has several interesting features. The most prominent are the so-called major maxima, large regularly-spaced peaks which occur at phase differences of $\phi = 2n\pi$, where n is an integer. These peaks are usually labelled by their value of n, so that the central ($\phi = 0$, $n = 0$) peak is the zeroth maximum, the $n = 1$ and $n = -1$ peaks are the first maxima, and so forth. The geometric angles for these maxima satisfy:

$$\sin(\theta_n) \approx \frac{n\lambda}{d} \tag{25}$$

and we can make the usual approximation for small geometric angles that $\theta \approx n\lambda/d$. In this small-angle approximation, the major maxima will be evenly spaced. The maxima occur when the waves from all of the sources are in phase, so these peaks have height $I_{\max} = N^2 I_1$, N^2 times the intensity of the signal from a single source.

Another important feature of the intensity graph is the minor maxima. Unlike the major maxima, which correspond to the signal from all of the sources being in phase, the minor maxima occur when all of the signals cancel out except for an amount equivalent to the signal intensity from one source. So these minor maxima have a height of I_1. For a system with N sources, there are $N-2$ minor maxima between every adjacent pair of major maxima.

Finally, there are the minima of the intensity graph, the points where an observer will see no signal intensity at all. These occur centered between every adjacent pair of minor maxima, and (not centered) between each major maximum and the nearest minor maximum on either side. The locations of the minima may be found by setting the numerator of the intensity equation equal to zero, which tells us that they occur when the phase difference between adjacent sources is:

$$\phi = \frac{2\ell\pi}{N} \tag{26}$$

where ℓ is any integer. Notice, however, that this equation overlaps with that for the major maxima, whenever ℓ is divisible by N. But at these locations the denominator of the intensity formula is also zero, and so the intensity is not simply zero there. (It must be determined by L'Hôpital's rule, which gives $I_N = I_{\max} = N^2 I_1$.) So we can say, more simply, that the minima are at $\phi = 2\ell\pi/N$ unless that value of ϕ corresponds to a major maximum. Notice that these minima and the major maxima, as a set, are evenly spaced. So the peak associated with a major maximum will have a width of $\Delta\phi = 4\pi/N$ in phase space ($\Delta\theta \approx 2\lambda/Nd$ in geometric angle, $2\lambda L/Nd$ as measured on a viewing screen), which is twice the width of a minor maximum. Also note that more sources (larger N) means narrower peaks, while longer wavelength signal (larger λ) means wider peaks.

The even spacing of the minima and major maxima also tells us that the minor maxima (and other points where $I_N = I_1$) are located at phase shifts $\phi = (2\ell+1)\pi/N$, which makes sense, since these are the maxima of the numerator of the intensity function.

Diffraction gratings If we take a bit of plastic and make many thin closely-spaced scratches on it, it creates an object confusingly referred to as a diffraction grating. Shining a laser beam through the *diffraction* grating results in an *interference* pattern exactly like the one described above. (Sadly, a "difraction" grating actually exhibits interference. I'm not sure why it has this dumb incorrect name.) The scratches act as "sources" for the pattern, so the number of sources, N, is equal to the number of scratches illuminated by the beam. The distance between adjacent sources, d, is the spacing between the scratches. Often gratings are described in terms of the number of scratches per centimeter, s. So the spacing between adjacent scratches will be $d = 1/s$ centimeters. A beam of width w centimeters incident on this grating will thus result in a pattern with $N = w/d = ws$ sources.

If we shine a source of red light and a source of blue light on the diffraction grating simultaneously, we will see their two interference patterns overlapping each other. Both will have major maxima at the center ($\phi = 0$), and the usual complement of other major maxima spreading out to both sides. The red light maixima, corresponding to light of a longer wavelength (a larger spatial scale) will have a greater angular spacing, with this spacing depending only on the precise wavelength of the light. The first-order maxima of the blue light will fall slightly outside those of the red light, the second-order blue maxima will deviate a little more from the second-order red maxima, and as you go to higher and higher orders the peaks will be better and better separated.

This increasing separation of the peaks is important because they are not simply lines, they have finite width. So if the center of a particular blue peak is not far enough from the center of a particular red peak, we will not be able to see two distinct peaks. All we will observe is a white peak with a bluish tint on one side and a reddish tint on the other. In order to be able to see that there are two different peaks there, the peaks must be separated by more than their angular width. The angular width of a peak depends not only on the wavelength of its light, but also on the number of slits in the grating — as discussed above, more slits there are, the narrower these peaks will be. And narrower peaks means that the lower order (less-separated) ones start to become more distinguishable. In particular, for two closely spaced wavelengths of light, with average value $\lambda = (\lambda_1 + \lambda_2)/2$ and difference $\Delta\lambda = \lambda_2 = \lambda_1$, we will be able to distinguish their nth-order peaks if:

$$\frac{\Delta\lambda}{\lambda} \geq \frac{1}{Nn} \tag{27}$$

For extremely closely spaced light frequencies (for example, the yellow sodium lines, at 589.9950 nm and 589.5924 nm) this criterion allows us to determine how far out we need to go in a particular setup's diffraction pattern to see that there are actually two different colors of light. For most well-spaced wavelengths, we will be able to separate the colors in the first order.

Limitations Most of the discussion so far has referred to situations where the small-angle approximation may be applied to determine the actual interference pattern as viewed on a screen (or by a detector) a distance L from the sources. However, as the geometric angle gets larger, the small-angle approximation ($\sin(\theta) \approx \theta$) no longer applies, and we must take into account the effects of the sine-like relation between the geometric angle and the phase angle.

As we move to larger and larger geometric angles, the peaks we see at those angles will be wider and wider and are more and more separated. And finally, when we get to $\theta = 90^\circ$, things break down entirely. We cannot observe any portion of the interference pattern which corresponds to phase differences larger than that which produces $\theta = 90^\circ$, both due to the geometry of the system, and due to the fact that $\sin(\theta)$ must not be greater than 1, which causes the relation $\phi = (2\pi d/\lambda)\sin(\theta)$ to break down for larger values of ϕ. So the phase shifts we observe in the system must satisfy:

$$\phi = \frac{2\pi d}{\lambda}\sin(\theta) \leq \frac{2\pi d}{\lambda} \tag{28}$$

and the highest-order major maximum we will be able to observe with a particular signal wavelength and source spacing will satisfy;

$$n_{\text{max}} = \left\lfloor \frac{d}{\lambda} \right\rfloor \tag{29}$$

(*ie.* $n \leq d/\lambda$). This may affect our ability to distinguish two different colors of light with a particular grating — if the spacing is too small, the peak order we need to attain to split those wavelengths may not exist in the interference pattern.

4 Diffraction

The sources referred to in the interference section are *point* sources, with no spatial extent. But most real-world sources are not like this — speakers are often several centimeters across, and slits which light passes (or telescope lenses used to receive light) will also have a certain width. Signal from a single source of finite width will interfere with itself, as if it were infinitely many point sources with infinitesimally small spacing between them. This effect is called diffraction.

4.1 Basics

Pure diffraction is observed with a single source, a slit or speaker of uniform width a in the dimension of interest. The source emits a total intensity I_0 at wavelength λ, and the observer is positioned a distance L from the front of the speaker, displaced by Δy from its centerline. (The geometric angle between the centerline and the line from the center of the speaker to the observer's position is, as before, $\theta \approx \Delta y/L$.) This observer will see that the phase difference between light received from the "bottom" edge of the source and light received from the "top" edge of the source is $\phi_{\text{bottom}} - \phi_{\text{top}} = \beta$. This phase angle β is related to the geometric angle θ by:

$$\beta \approx \frac{2\pi a}{\lambda} \sin(\theta) \approx \frac{2\pi a}{\lambda} \theta \tag{30}$$

where the first approximation is for $a \ll L$ and the second is for small geometric angles ($\theta \ll \pi/2$). Notice that this is very similar to the ϕ-d-θ relation for interference. The intensity the observer sees, as a function of β, is:

$$I_1 = I_0 \left(\frac{\sin(\beta/2)}{\beta/2} \right)^2 \tag{31}$$

This intensity graph is distinctly different from that for interference problems. Its central peak at $\beta = 0$ (and $\theta = 0$) is the tallest, having height $I_{\text{max}} = I_0$ (a very different relationship between intensity formula and maximum height than the one for interference patterns). The pattern is symmetric about this peak, and has a series of smaller peaks to either side. These peaks are only half as wide as the central peak. As with the intensity graph, for small geometric angles the shape of the graph is not strongly affected by choosing to graph I_1 *vs.* β, *vs.* θ, or *vs.* Δy.

The other interesting feature of this graph is the locations of the zero-points in between adjacent peaks. Unlike interference patterns, for diffraction we ar typically more interested in knowing the location of these minima than in knowing the location of the peaks. The minima are located where the numerator of the sine factor in the intensity equation becomes zero, that is, where $\beta = 2m\pi$, where m is an integer. The value of m corresponding to a particular minimum is used to label that minimum, so that when we refer to the first-order minima we mean $m = \pm 1$. The geometric angles where these minima occur will satisfy:

$$\sin(\theta_m) \approx \frac{m\lambda}{a} \tag{32}$$

Again, note the similarity to the equation for the locations of the interference *maxima*.

The central peak is usually considered to be the primary output "beam" of the source, since it will contain the large majority of the energy. So for a slit-like source, we can say that it has a beam angular width of:

$$\theta_{\text{beam}} \approx 2\theta_1 \approx 2\frac{\lambda}{a} \tag{33}$$

while the physical width of the beam as seen by an observer a long distance L away will be:

$$W_{\text{beam}} \approx L\theta_{\text{beam}} \approx = 2L\frac{\lambda}{a} \tag{34}$$

4.2 Circular Apertures

If instead of being a slit of width a, the source is a circular object (hole or speaker) of diameter D, then the relation between the phase angle and the geometric angle is somewhat modified:

$$\beta \approx \frac{1}{1.22}\left(\frac{2\pi D}{\lambda}\sin(\theta)\right) \tag{35}$$

which changes the locations of the minima to:

$$\sin(\theta_m) \approx 1.22\frac{m\lambda}{D} \tag{36}$$

and the beam width equations become:

$$\begin{aligned} \theta_{\text{beam}} &\approx 2\theta_1 \approx 2.44\frac{\lambda}{D} \\ W_{\text{beam}} &\approx L\theta_{\text{beam}} \approx 2.44L\frac{\lambda}{D} \end{aligned}$$

4.3 Diffraction limitation

We've already seen how diffraction due to the finite size of the source affects the angular size of the region where substantial signal may be observed (the "beam width"). But we can also turn this around and suppose that the device we are using to detect a particular signal has a finite size.

It then turns out that we can determine the angular or linear size of the smallest observable source by using the size of the detector in the beam width equation. For example, a telescope with a circular lens of diameter D a distance L from earth observing light at a wavelength λ could see objects on earth of linear size $W \approx 2.44\lambda/D$.

A slightly distinct case is the question of how far apart two objects must be before they can be distinguished by this telescope. The way to imagine this is to think of each of the objects as fuzzy blotches, each one appearing to the observer behind the telescope as a single diffraction peak. If the centers of the diffraction peaks are too near to each other they will simply resemble a single peak. But if we put one of the diffraction peaks at the first minimum of the other peak, then we'll see significant amplitude in an unexpected location, suggesting that there's actually two peaks. So minimum angular separation of the two objects is simply the angle between the centerline and the first diffraction minimum, that is:

$$\begin{aligned} \theta_{\text{sep}} &\approx \theta_1 \approx \begin{cases} \frac{\lambda}{a}, & \text{slit} \\ 1.22\frac{\lambda}{D}, & \text{circular aperture} \end{cases} \\ \Delta y_{\text{sep}} &\approx L\theta_{\text{sep}} \approx \begin{cases} L\frac{\lambda}{a}, & \text{slit} \\ 1.22L\frac{\lambda}{D}, & \text{circular aperture} \end{cases} \end{aligned}$$

4.4 Phasors

Finally, you should know that there's also a completely cockeyed phasor representation for diffraction systems. The signal seen by the observer can be thought of, as previously mentioned, as signal from infinitely many point sources with infinitesimally close spacing. So each of these notional sources has a very very short phasor which is turned very very slightly counterclockwise compared to its predecessor.

In the infinite limit, all of these phasors will form a circular arc. The length of this arc (the "distance" you would travel walking along the curve) is equal to the total amplitude of the signal output, which is $A_0 = \sqrt{I_0}$. The angle subtended by this arc (the angle between the two radii drawn from the center of the circle to the edges of the arc) will be the phase angle β between the signal coming from the bottom and top edges of the slit. So, for example, if the arc is a quarter circle, then $\beta = \pi/2 = 90°$. Finally, the chord length (the length of a line that goes from one tip of the arc to the other) will be the observed amplitude at that location A_1.

For a given setup, the arc length A_0 will always be the same, but depending on the location of the observer, the subtended angle β will change. So the radius of the circle from which the arc is taken will depend on the observer's location. For a circle of radius r an arc subtending an angle θ will have a length $s = r\theta$, so the circle radius for any particular phasor will be $A_r = A_0/\beta$.

Finally, do not try to do the geometry to calculate the observed intensity from the phasor diagram — simply use the intensity formula on the equation sheet. It will save you a lot of grief!

5 Mixed interference and diffraction

If we have an arrangement of multiple sources, each of which is of finite size, then the signal seen by an observer will exhibit both interference and diffraction. We consider only the simplest case — a row of N identical slit-shaped sources, each of width a, with a distance d between adjacent sources ($a < d$). Each individual source has a total output intensity I_0, so that the maximum possible signal intensity we can see from this arrangement is $N^2 I_0$, which will be observed when all the sources are exactly in phase.

Each source will output a diffraction pattern $I_1 = I_0(\sin(\beta/2)/(/beta/2))^2$, and all of these diffraction patterns will interfere according to the interference formula $I_N = I_1(\sin(N\phi/2)/\sin(\phi/2))^2$. So the total intensity will have the following formula:

$$I_N = I_0 \left(\frac{\sin(\beta/2)}{\beta/2}\right)^2 \left(\frac{\sin(N\phi/2)}{\sin(\phi/2)}\right)^2 = \frac{I_{\max}}{N^2} \left(\frac{\sin(\beta/2)}{\beta/2}\right)^2 \left(\frac{\sin(N\phi/2)}{\sin(\phi/2)}\right)^2 \quad (37)$$

Essentially, what happens is we take the interference pattern (where all the major interference peaks are the same height) and change the heights of the major peaks so that they match the shape of the diffraction pattern at that point. The diffraction pattern forms a kind of "envelope" controlling the overall intensity, while the interference pattern controls the details of the shape.

For example, you might see that there is a grouping of seven narrow peaks in the center of the pattern with the central one the tallest. Its height will be

$$I_{n=0} = I_{\max} = N^2 I_0 \tag{38}$$

The three to either side of the center will have decreasing heights, according to:

$$I_n = I_{\max} \left(\frac{\sin(\beta_n/2)}{\beta_n/2} \right)^2 \tag{39}$$

where n indicates the count away from the center and $\beta_n = (2\pi a/\lambda)\sin(\theta_n) = (a/d)\phi_n = (a/d)2\pi n$ is the phase difference between signals from either edge of a single source at the location of the relevant interference peak. Just outside this central grouping, in both directions, you might then see a minimum where you expected the next peak to be, and then an additional grouping of three peaks (heights also controlled by the I_n formula), another minimum with a missing peak, another grouping of three, and so forth.

These narrow peaks are all *interference* peaks, and they are broken into small groupings by the constraint of the *diffraction* peaks — each grouping is contained inside a diffraction peak. The minima where there should be interference peaks are the diffraction minima, which (in this case, and almost every situation you will see) happen to coincide perfectly with the usual location of an interference peak.

Using the formulae for the location of the interference peaks and diffraction minima we can determine the relationship between the slit sizes and their spacing. Since the $n = 4$ interference peak is located at $\sin(\theta) = n\lambda/d = 4\lambda/d$ and it coincides with the $m = 1$ diffraction minimum, which is located at $\sin(\theta) = m\lambda/a = \lambda/a$, we may conclude that $n\lambda/d = m\lambda/a$, or, in other words:

$$\frac{d}{a} = \frac{n}{m} \tag{40}$$

So in this situation, $d = 4a$.

We can also determine the number of sources used in the problem by inspection. As before with simple interference, we merely count the number of minor peaks between each adjacent pair of major peaks, and add two to get the number of sources. The number and location of the minor peaks is not affected by the presence of diffraction (although they will most likely be shrunk even further). And, conversely, the number of sources does not affect the overall shape of the interference-diffraction pattern except by adding the minor peaks and increasing the maximum peak height.

6 Particles as waves

6.1 The photoelectric effect

So far we have described the light as if it travels like a wave, and in many circumstances this is a perfectly reasonable thing to do, as can be proven by setting up any of the systems described above. Sending coherent light through different arrangements of slits will produce interference and diffraction patterns (depending on the arrangement), and these patterns are characteristic of wavelike signals. So in this kind of problem, we can think of light as acting exactly like water waves and sound waves, which have *only* wavelike nature.

But there are other situations in which it becomes clear that light can also come in discrete packets, instead of continuous waves. These packets will each have a specific amount of energy, and a specific location at any given time. They move along definite trajectories, like little bullets. Experiments with the photoelectric effect can be used to demonstrate this particlelike nature of light.

6.2 Setup

The photoelectric effect is simply the observation that shining light on a metal surface can cause electrons to be ejected from the surface. If another metal plate is placed some distance away, some of the ejected electrons will be able to fly across the gap to hit the second plate, causing a current to flow between the two plates. The maximum possible attainable current flow is going to be directly proportional to the number of ejected electrons — if N_{e^-} electrons per second are ejected, and all of them make it across the gap, then a current $I_{\max} = -N_{e^-}e$ will flow, since each electron carries a charge $-e$. (Notice that this current actually flows from the second plate to the first, since current is the flow of *positive* charge.)

6.2.1 Variables

We may determine the energy of the ejected electrons by putting an electric potential difference $V = V_{\text{source}} - V_{\text{destination}}$ across the gap between the two plates. (This potential difference is sometimes called a "bias voltage".) If the destination plate is at a lower electric potential than the source plate, then the potential energy of an electron at the destination, $U_{\text{destination}} = -eV_{\text{destination}}$ will be higher than it was at the source. So the electron will have to lose an amount of kinetic energy $\Delta\text{KE} = U_{\text{source}} - U_{\text{destination}} = eV$ in order to cross this gap. If an electron has less energy than this, it will not be able to cross. So if we increase the voltage difference between the two plates just until no more current flows, then that means that none of the electrons had any more kinetic energy than it took to cross the gap at that voltage. So their maximum kinetic energy is:

$$\text{KE}_{e^-,\max} = eV_{\text{stop}} \tag{41}$$

It is also interesting to notice that as we increase the applied voltage (before reaching V_{stop}, the current flow decreases steadily. So many of the electrons must have less than the maximum kinetic energy.

At the other end of the system, we have three parameters we can vary — the color of the light (frequency), the intensity of the light (power in Watts = Joules/s hitting the source plate, which may not be the same as the power emitted by the source, depending on the geometery of the system), and the type of metal from which the plate is constructed.

If we keep the light source the same and change the type of metal for the plate, we discover that the maximum kinetic energy of the ejected electrons depends on the type of the metal. This implies that the electrons must have to sacrifice a certain minimum amount of the energy they receive from the light in order to escape the metal (although they may, through random chance, sacrifice more), and that the amount of energy sacrificed (which can be thought of as a sort of binding energy) depends on the metal type. We call this sacrificed energy the "work function" of the metal, Φ. So the total kinetic energy of an ejected electron will be equal to the energy it received from the light source (E_γ) minus the energy it sacrificed to escape the metal:

$$\mathrm{KE}_{e^-} = E_\gamma - \Phi \tag{42}$$

An electron which receives energy from the light which is less than the work function cannot be ejected.

If we keep the light frequency and the plate metal constant and slowly decrease the light intensity, we discover that the maximum electron kinetic energy does not change, but the maximum current flow (with no voltage difference) decreases steadily, in direct proportion to the decreased light intensity. The current vanishes only when the light is turned off completely. Conversely, if we hold the metal and light intensity constant and decrease the light frequency, we discover that the maximum electron kinetic energy decreases, at a rate propotional to the rate of decrease in light frequency, until it eventually reaches zero. The current flow (at $V = 0$), however, remains constant, until the point where $\mathrm{KE}_{e^-,\max} = 0$, then it abruptly drops to zero.

6.2.2 Conclusions

So, it seems that the light frequency affects the energy each electron receives, while the light intensity effects the number of electrons that receive energy. This suggests that the light is acting as a particle in this situation — each light particle (photon) has a precise, discrete amount of energy, which turns out to be $E_\gamma = hf = \hbar\omega$ (to be explained later), so that:

$$\mathrm{KE}_{e^-,\max} = hf - \Phi \tag{43}$$

Some of the electrons bang around a little bit extra and lose some more energy, so not all of them will have the maximum possible kinetic energy when they escape or are detected. But none of them will have more energy than this. If the lihgt frequency is low enough that the photon energies are less than the work function ($f < f_0 = \Phi/h$) then no electrons will be able to gain enough energy to escape the metal and no current will flow.

The light source, which emits a power of P_{emit} Joules per second is simply emitting a certain number of these energy packets per second:

$$N_{\text{emit}} = \frac{P_{\text{emit}}}{E_\gamma} \tag{44}$$

A certain fraction of these packets are received by the metal plate, where that fraction depends on the plate's distance from the light source r and its area A. If the source is emitting N_{emit} photons per second uniformly in all directions, then these photons will, at a distance r from the source, be distributed evenly over the surface of a sphere of radius r. So each square meter of this surface will emit $N_{\text{emit}}/4\pi r^2$ photons. A plate covering A of these one-square-meter regions will receive A times as many photons as this. Each of these photons can, in principle, eject a single electron (although not all of them do). So the maximum number of electrons ejected per second is:

$$N_{e^-,\max} = \frac{A}{4\pi r^2} N_{\text{emit}} = \frac{A}{4\pi r^2}\left(\frac{P}{hf}\right) \tag{45}$$

which means that the maximum current flow with the bias voltage set to zero is:

$$I_{\max} = -N_{e^-,\max} e = -\frac{A}{4\pi r^2}\left(\frac{P}{hf}\right) e \tag{46}$$

So the principle things to note about the photoelectric effect are:

- It demonstrates that light can act like a particle, as an alternative to its more familiar wave behavior.
- The kinetic energy of the electrons, and the required bias voltage to stop the flow of photoelectric current (also called the stopping voltage) depend on the frequency of the light and the type of metal.
- The size of the photoelectric current (with bias voltage 0 and $f > f_0$) depends only on the light intensity, and the percentage of the emitted light that actually hits the metal plate.

6.2.3 The Duane-Hunt law

An interesting side note is that we can also turn the photoelectric effect around — we can shoot electrons of a known energy at a metal plate to cause photons to be ejected. This is known as the Duane-Hunt effect. An individual electron becomes bound in the metal, gaining energy equal to the work function of the metal, and then rids itself of this excess energy (plus any kinetic energy) by emitting a photon. The maximum energy of emitted photons is found by simply rearranging the photoelectric effect equation ever so slightly:

$$E_{\gamma,\max} = E_{e^-} + \Phi \tag{47}$$

Typically electrons in electron beams have energies much higher than the work function (especially when this setup is being used to produce X-rays for an X-ray machine), so we can usually neglect Φ and write:

$$E_{\gamma,\max} \approx E_{e^-} \tag{48}$$

6.2.4 A note on electron-Volts

This is kind of a freaky-looking unit, but it's not really so bad once you get used to it. An electron-Volt is the amount of energy it takes to move a single electron across an electric potential difference of 1 Volt. So it's equal to the change in the electron's potential energy, which is just e, the charge on the electron, times the voltage difference, 1V. (Recall, $U = qV$ is the potential energy of a charge q in an electric potential V.) So:

$$1 \text{ eV} = e \cdot (1 \text{ V}) = \left(1.6022 \cdot 10^{-19} \text{ C}\right)\left(1 \ \frac{\text{J}}{\text{C}}\right) = 1.6022 \cdot 10^{-19} \text{ J} \tag{49}$$

It's convenient to notice that if an electron crosses a potential difference equal to a particular number of volts, then its energy will change by exactly that number of electron volts. So, an electron which crosses a potential difference of 3V will have its energy change by 3eV.

Also, if you find electron-Volts completely terrifying and incomprehensible, you can simply convert them to Joules using the conversion above, work the problem entirely in SI units, and then convert back to the desired answer units at the end.

6.3 The two-slit experiment

We've seen that light, which we initially described as a wave, can also have the characteristics of particles. But it is also true that anything we think of as a particle can also be described as a wave! Electrons, protons, neutrons, buckyballs, baseballs, and skyscrapers all can be thought of as having wavelike properties as well, and (in principle) subjected to interference and diffraction experiments to prove their wave status.

The simplest of these experiments is the two-slit experiment. In this case we set up a "particle gun" which shoots particles at a distant wall which has a pair of slits in it. The particles go through either one slit or the other and strike a detector which is far beyond the wall. We already know what the detector will see if these things we're shooting behave like waves — we'll simply get the interference pattern for two identical sources, exactly like we would if we were shining a laser on the slits (a laser being a "photon gun"). And it turns out that you can get this pattern with electrons, protons, neutrons, buckyballs, and other small particles.

However, as the particles start to get bigger (larger molecules, baseballs, skyscrapers, *etc.*), it starts to become more difficult to get this nice interference pattern, and not just because of how hard it is to make a "skyscraper gun". There are two aspects to this problem. The first is that it's very difficult to make the "skyscraper wave" coming out of your gun be correlated (*ie.* have a consistent phase), because massive objects like skyscrapers have extremely short wavelengths. Without a correlated beam, there is no guaranteed phase relationship between particles which went through the top slit and particles which went through the bottom slit. And that guaranteed phase relation is what creates the interference pattern. Without it, the strange interference oscillations vanish, and we're left with simply two "piles" of particles, corresponding to the endpoints of the ballistic trajectories of bullet-like particles travelling through either the top or the bottom slit.

The other aspect is that even if you have an initially coherent beam, any attempt to measure which slit a particular particle went through will disturb the phase relationship between the two sets of particles, destroying the interference pattern. Even if you don't deliberately set up the experiment to allow you to differentiate between the two sets of particles, there may still be environmental elements that allow you to tell. For example, if you're flinging skyscrapers around, you'll be able to *see* which slit it's gone through, hear the wind of its passage, and so forth. And these environmental factors will disturb the phase relationship just as thoroughly as, say, adding a particle-counter, or a dense mist, or a spin-flipper to one of the slits on a smaller-scale experiment might. The reason for this is that any possible method of measuring the particles requires the particles to *interact* with something else. And the interaction necessarily changes not only the measurement device, but also the particle (think of Newton — every action produces an equal and opposite reaction), which change breaks the phase relationship.

Since it is practically impossible to eliminate either of the above effects in the everyday macroscopic world, we simply don't see ordinary objects behaving as waves. It's only with very tiny (low-mass) objects, like photons, elementary particles, and small molecules that we can set up sufficiently pristine experimental conditions to observe the wavelike behavior, even though it exists in principle for larger objects as well. But the two-slit experiment *has* been successfully performed even with buckyballs, which are relatively large (60 carbon atoms) and nicely symmetrical molecules.

6.4 Wave-particle formulas and factoids

To make the wave-particle relationship concrete, it's time to introduce some formulas. First, recall that the oscillation period, frequency and angular velocity of a wave are related by $T = 1/f = 2\pi/\omega$, and the wavelength and wavenumber are related by $\lambda = 2\pi/k$. Since the only wave parameter that affects the interference and diffraction patterns is the wavelength, particles with the same wavelength will, when put through (geometrically) identical arrangements of slits, produce identical interference and diffraction patterns, regardless of particle type. So, for example, a 300nm photon and a 300nm electron, both sent through a 1mm slit, will produce identical diffraction patterns.

If a particle has momentum p and total energy E, then we can find the wavelength and frequency of the corresponding wave using the following equations:

$$p = \frac{h}{\lambda} = \hbar k \tag{50}$$

$$E = hf = \hbar\omega \tag{51}$$

So all particles with the same wavelength have the same momentum, and all particles with the same frequency have the same energy. However, due to different relationships between frequency and wavelength for different types of particles, particles with the same momentum will not have the same energy unless they're the same type of particle. In general, though it's not demonstrated here, more massive particles with the same energy will have shorter wavelengths. So if you have a photon, an electron, and a proton which all have energies of 3eV, then the photon will have the longest wavelength, followed by the electron, and the proton will have the shortest wavelength.

The wavelength derived from the first formula is sometimes called the de Broglie wavelength of the particle. The wavelength can be thought of as a notional size for the particle, suggesting that we can't know its position to more than within about one de Broglie wavelength when it's travelling at a speed corresponding to the momentum p. Using this sort of handwavy explanation, we can transform the momentum relation above into what's called the Heisenberg Uncertainty Principle:

$$\Delta x \, \Delta p \gtrsim \hbar \tag{52}$$

(This is, of course, not a precise derivation.) Δx is the "uncertainty" in the position — the average amount by which our measured location can be expected to deviate from the particle's mean location, and Δp is the uncertainty in the momentum.

It is important to realize that in the quantum mechanical world, before measurement the particle doesn't have an "actual" position or momentum, only an average position and momentum. Before we check, the particle doesn't have a particular location or momentum at all, just a probability distribution giving the likelihood of finding it with any particular value of p or x.

If we intend to measure, say, the position, and we know its spatial probability distribution in advance, then we can calculate the expected (average) value of x where we'll find the particle,

$\langle x \rangle_{\text{pre}}$, and the average size of x as well, $\sqrt{\langle x^2 \rangle_{\text{pre}}}$. Comparing these values helps us determine what we can expect (before making the measurement) the deviation from the mean to be:

$$(\Delta x_{\text{pre}})^2 = \left\langle x^2 \right\rangle_{\text{pre}} - \langle x \rangle_{\text{pre}}^2 \tag{53}$$

(If we instead chose to measure this particle's momentum, then we would expect to find an average deviation of at least $\Delta p_{\text{pre}} \gtrsim \hbar / \Delta x_{\text{pre}}$, as given by the Uncertainty Principle.)

When we actually measure the particle's position, our imperfect measuring equipment will cause us to be uncertain about the result, so that we only know its position to within $\pm \Delta x_{\text{post}}$. This will result in a change in the lower limit on the uncertainty in the momentum — it will now become $\Delta p_{\text{post}} \gtrsim \hbar / \Delta x_{\text{post}}$. Notice that if we were somehow able to measure the particle's position perfectly so that $\Delta x_{\text{post}} = 0$, then we would have $\Delta p_{\text{post}} \gtrsim \infty$ — in other words, we would have absolutely no idea what the particle's momentum was. And, conversely, if we measure the momentum perfectly, we will have no idea where the particle is.

Interestingly, if we have a massive particle and we know that its average momentum is zero, then knowing the momentum uncertainty can actually tell us its average kinetic energy:

$$\left\langle \text{KE}_{\text{massive},\langle p \rangle = 0} \right\rangle = \left\langle \frac{p^2}{2m} \right\rangle = \frac{\left\langle p^2 \right\rangle}{2m} = \frac{1}{2m} \left((\Delta p)^2 + \langle p \rangle^2 \right) = \frac{(\Delta p)^2}{2m} \tag{54}$$

If the particle is also in a harmonic oscillator potential well ($U(x) = kx^2/2$) and has an average position $< x >= 0$, then we can pull a similar trick with the position uncertainty and the expected potential energy, so that $< U >= k \left(\Delta x\right)^2 / 2$.

6.5 Photons and the wave equation

Recall that photons obey the wave equation with propagation velocity $v = c = 3 \cdot 10^8$ m/s:

$$\frac{\partial^2 \Psi}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \Psi}{\partial t^2} \tag{55}$$

Plugging in a sine or cosine solution for the wavefunction Ψ gives, as discussed above, the following relations between wavelength and frequency (or wavenumber and angular velocity):

$$\omega = ck \tag{56}$$

$$f = \frac{c}{\lambda} \tag{57}$$

Combining these equations with the energy and momentum equations of the previous section gives:

$$p = \frac{h}{\lambda} = \hbar k = \frac{hf}{c} = \frac{\hbar \omega}{c} = \frac{E}{c} \tag{58}$$

$$E = hf = \hbar \omega = \frac{hc}{\lambda} = \hbar c k = pc \tag{59}$$

If the energy is in units of electron-Volts (eV) and the wavelength is in nanometers (nm) then you can use the following data to compute the energy:

$$hc = 1240 \text{ eV} \cdot \text{nm} \tag{60}$$

$$E = \frac{1240 \text{ ev} \cdot \text{nm}}{\lambda} \tag{61}$$

6.6 Massive particles and the Schroedinger equation

Particles which have mass, unlike the photon, are described by a different differential equation, the time-dependent Schroedinger equation. This equation describes the variation of a wavefunction $\Psi(x,t)$ of a particle of mass m whose potential energy U depends only on its location x:

$$-i\hbar\frac{\partial\Psi}{\partial t} = U(x)\Psi(x,t) - \frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} \tag{62}$$

Notice that this equation, unlike the wave equation, involves only a first derivative with respect to time. In general, the solutions to this equation will depend on what exactly $U(x)$ is, as well as on the boundary conditions of the system. However, for a free particle ($U = 0$), the solutions are sinusoidal, and have exactly the form described in section 2. The dispersion equation can be derived by plugging a sinusoidal solution into the Schroedinger equations:

$$\omega = \frac{\hbar k^2}{2m} = \frac{2\pi h}{2m\lambda^2} \tag{63}$$

$$f = \frac{hk^2}{2m} = \frac{h}{2m\lambda^2} \tag{64}$$

Combining this with the momentum and energy equations, as well as some basic mechanics, gives:

$$p = \frac{h}{\lambda} = \hbar k = \sqrt{\frac{2mf}{h}} = \sqrt{\frac{2m\omega}{\hbar}} = mv = \sqrt{2mE} \tag{65}$$

$$E = \mathrm{KE} = hf = \hbar\omega = \frac{h^2}{2m\lambda^2} = \frac{\hbar^2k^2}{2m} = \frac{1}{2}mv^2 = \frac{p^2}{2m} \tag{66}$$

For computational purposes, it can be helpful to remember that, for electrons:

$$\frac{h^2}{2m_e} = 1.505\ \mathrm{eV}\cdot\mathrm{nm}^2 \tag{67}$$

which allows us to write the energy equation for *any* massive particle as:

$$E = \frac{1.505\ \mathrm{eV}\cdot\mathrm{nm}^2}{\lambda^2}\left(\frac{m_e}{m}\right) \tag{68}$$

Finally, note that if U is constant but *non*zero, then we must make a distinction between the total energy $E = hf = \hbar\omega$ and the kinetic energy $\mathrm{KE} = h^2/2m\lambda^2 = \hbar^2k^2/2m$. In this case:

$$\begin{aligned} E &= \mathrm{KE} + U \\ hf = \hbar\omega &= \frac{h^2}{2m\lambda^2} + U = \frac{\hbar^2k^2}{2m} + U \end{aligned}$$

7 Wavefunctions and probability densities for massive particles

The next important question is what exactly the wavefunction signifies for massive particles. For water waves and waves on a string, the "wavefunction" is simply a measure of the height (in meters) of a visible disturbance. For sound waves, it's a measure of how much the air pressure deviates from the mean value. For light waves, it's related to the size of the electric and magnetic fields at the point in question.

With massive particles, it turns out that the wavefunction is something called the probability amplitude for finding the particle at that location. If we're given this wavefunction, this probability amplitude $\Psi(x,t)$, then there are a couple steps required to make it into something we can observe in a physical system. First, we square it, to get the probability density:

$$P(x,t) = |\Psi(x,t)|^2 \tag{69}$$

For one-dimensional systems, like we've been discussing so far, this probability density is the probability per unit length of finding the particle in the region near x at the time t. It is *not* the probability of finding it *at* the location x. In order to find a total probability, we need to integrate the probability density over the region we're interested in:

$$P_x^{x+\Delta x}(t) = \int_x^{x+\Delta x} P(x',t)\,dx' = \int_x^{x+\Delta x} \left|\Psi(x',t)\right|^2\,dx' \tag{70}$$

This gives the total probability of finding the particle anywhere in the region $x < x' < x + \Delta x$. If Δx is very small, then we may make the approximation:

$$P_x^{x+\Delta x}(t) \approx P(x + \frac{1}{2}\Delta x, t)\Delta x \tag{71}$$

Essentially, we approximated this region of the $P(x,t)$ graph as a very narrow rectangle with width equal to Δx and height equal to the height of the graph at the center of the region, $P(x+\Delta x/2,t)$. The probability of finding the particle in that region is simply equal to the area of the rectangle.

If we integrate the probability density over the entire x-axis (or over the entire set of possible particle locations), then we know we *must* find the particle somewhere in that region, so the total probability will be 1. So we can write the following normalization integral for the wavefunction:

$$1 = \int_{-\infty}^{\infty} P(x',t)\,dx' = \int_{-\infty}^{\infty} \left|\Psi(x',t)\right|^2\,dx' \tag{72}$$

These normalization integrals can be used to help us find the exact form of the wavefunction for a particular system. Another similarly convenient property of the wavefunction is that it is continuous (that is, it can be drawn without lifting your pencil from the paper), and it also has a continuous first derivative (except at points where the potential energy becomes infinite). Once we've determined all the possible solutions to a particular system's Schroedinger equation, these properties allow us to find which of the solutions are actually attainable states of the system, and to link them up at boundaries, and so forth.

However, even without solving the Schroedinger equation for a particular potential energy, we can make some guesses about the shapes of its wavefunctions. A particular Schroedinger solution will have a fixed total energy E. In a classical system, this total energy would be equal to its potential energy U plus its kinetic energy KE:

$$E = U + \mathrm{KE} \tag{73}$$

In a classical system, the particle could only exist in regions where its kinetic energy was zero or positive — it could never have negative kinetic energy. So we would never find the particle in regions where $E < U$. But it turns out that, quantum mechanically, particles *are* able to enter this classically forbidden region. However, their qualitative behavior changes. In classically allowed regions, the particle's wavefunction is, well, a wave, much like the ones we've described above. However, as the particle enters classically forbidden regions, the probability of finding it at a particular location decreases exponentially as it penetrates deeper — the wavefunction is equal to something like $Ae^{-x/L}$ — and the descent is faster the more forbidden the region is. So classically allowed regions are oscillatory, classically forbidden regions are exponential decay.

There is also another kind of region which is forbidden even in quantum mechanics. You will never ever ever find a particle in a region where its potential energy would be infinite ($U = \infty$), no matter how hard you look. So if you're given a system with $U = \infty$ in some region, then you know immediately that $\Psi(x,t) \equiv 0$ in that region, and continuity will require that $\Psi(x,t)$ also be zero at the boundaries of that region.

For this particle with fixed total energy E we can also deduce a couple other facts about its wavefunction. First, if the potential energy U is higher in a particular location, then the total kinetic energy must be lower there ($\mathrm{KE} = E - U$, and a lower kinetic energy means a longer wavelength ($\lambda = \sqrt{h^2/2m\,\mathrm{KE}}$). So the particle's oscillations will have a longer wavelength in regions of higher potential energy, and the wavefunction peaks will be less closely spaced there.

In addition, a lower kinetic energy means a lower particle velocity ($v = \sqrt{2\,\mathrm{KE}/m}$), so the particle will leave these regions of lower kinetic energy more slowly, and will thus spend more time in them. So the probability of finding the particle in these low-KE regions is greater than the probability of finding it in a high-KE region. Thus the wavefunction peaks will be taller (and the valleys deeper) in a region of high potential energy than in a region of low potential energy.