Wave Primer

Senior Physics Challenge
Isaac Physics Project
Cavendish Laboratory, Cambridge

Contents

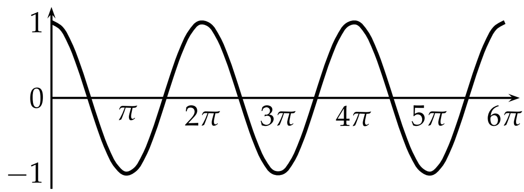
Many wonderful and crucial physical phenomena can be explained using the model of a wave.

In this course we will bring together the mathematical and physical tools you will have been given during a pre-university course to establish a more comprehensive description of wave motion. This will enable you to access the methodology of more advanced texts explaining, for example, vibration and sound, electromagnetism, alternating current and the wave formulation of quantum mechanics.

1 Phase

The phase ϕ of a wave measures the stage of the oscillation at a place and a time: is it a peak, a trough or somewhere in between. In this article, we will take a phase of 0 to refer to a peak. We usually measure ϕ in units of angle, where 360° would represent a full oscillation. Therefore $\phi = 180^\circ$ would refer to a trough, and $\phi = 90^\circ$ and $\phi = 270^\circ$ would refer to positions where the oscillator is at its equilibrium position (on the axis).

However, we will want to use calculus. The standard results for differentiating trigonometric functions (e.g. $\frac{d}{d\theta}\sin\theta = \cos\theta$) only apply when angles are measured in radians, and so we shall use radians from now on. There are 2π rad in one circle, and therefore we convert between radians and degres using the fact that $180^\circ = \pi$ rad.



Peaks correspond to phases of $\phi = 0\,\text{rad}, 2\pi\,\text{rad}, 4\pi\,\text{rad}$; troughs are at $\phi = \pi\,\text{rad}, 3\pi\,\text{rad}$, and the oscillator is at the equilibrium position when $\phi = \frac{1}{2}\pi\,\text{rad}, \frac{3}{2}\pi\,\text{rad}$ and so on.

2 Angular frequency

The frequency f of a wave is measured in hertz Hz. This gives the number of complete waves which pass us each second. Each whole wave is worth 2π rad of phase. Therefore the phase change in one second will be $2\pi f$. This is called the **angular frequency** ω and is measured in rad s^{-1}.

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

3 Phase and time

If an oscillation begins at a peak, then $\phi = 0$ when $t = 0$. The additional gain in phase during time t will be ωt, and so the phase at time t is ωt. This enables us to write the oscillation as

$$y = A\cos(\omega t). \tag{2}$$

A wave does not have to start at a peak. If its phase when $t = 0$ is ϕ_0, then its phase at time t will be $\phi = \phi_0 + \omega t$ and our oscillation is written

$$y = A\cos(\omega t + \phi_0). \tag{3}$$

Using the trigonometric identity $\cos(a+b) = \cos a \cos b - \sin a \sin b$, this oscillation can also be written

$$y = A\cos\phi_0 \cos\omega t - A\sin\phi_0 \sin\omega t, \tag{4}$$

where we see that our oscillation can be thought of as having two independent oscillating parts. One is an even (cosine) function with amplitude $A\cos\phi_0$, while the other is an odd (sine) function with amplitude $-A\sin\phi_0$. As these amplitudes are proportional to a cosine and sine respectively of the same angle, it is natural to consider the two parts as components of the whole oscillation. Indeed we will often draw this as an amplitude vector with components $(A\cos\phi_0, A\sin\phi_0)$. However it is important to realise that it is not a vector: the two components are not independent of each other in the way that the components of a vector are, but represent the odd and even parts of the same oscillation.

Sometimes it is preferable to give the amplitudes of the sine and cosine components rather than the overall amplitude A and the starting phase

ϕ_0. Either way of doing it gives the two pieces of information needed to be able to work out the displacement of the oscillator at any time in the future.

If we were shown an oscillation $y = C\cos\omega t - S\sin\omega t$, then we would use equation 4 and equate

$$\begin{aligned} C &= A\cos\phi_0 \\ S &= A\sin\phi_0 \end{aligned}$$

and then work out the overall amplitude and phase from

$$\begin{aligned} A &= \sqrt{C^2 + S^2} \\ \phi_0 &= \tan^{-1}\left(\frac{S}{C}\right). \end{aligned} \tag{5}$$

Notice that when we describe an oscillation using sines and cosines, it is not possible to eyeball the phase and the overall amplitude. We shall come back to this point when we introduce complex numbers for describing waves.

4 Phase difference and Interference

Often a wave is produced by one source, but then reaches a sensor by more than one route. The waves which arrive all have the same frequency, but may have different amplitudes and phases. Let us suppose that the first wave has amplitude A and phase ϕ, while the second has B and θ. We wish to know the amplitude of the combined wave.

The combined oscillation is

$$\begin{aligned} y &= A\cos(\omega t + \phi) + B\cos(\omega t + \theta) \\ &= A\cos\phi\cos\omega t - A\sin\phi\sin\omega t + B\cos\theta\cos\omega t - B\sin\theta\sin\omega t \\ &= (A\cos\phi + B\cos\theta)\cos\omega t - (A\sin\phi + B\sin\theta)\sin\omega t. \end{aligned} \tag{6}$$

This is in the form $y = C\cos\omega t - S\sin\omega t$ if we write

$$\begin{aligned} C &= A\cos\phi + B\cos\theta \\ S &= A\sin\phi + B\sin\theta. \end{aligned} \tag{7}$$

The overall amplitude R will be

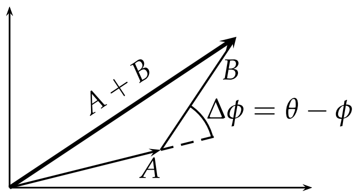
$$\begin{aligned} R &= \sqrt{C^2 + S^2} \\ &= \sqrt{(A\cos\phi + B\cos\theta)^2 + (A\sin\phi + B\sin\theta)^2} \\ &= \sqrt{A^2 + B^2 + 2AB\cos\phi\cos\theta + 2AB\sin\phi\sin\theta}, \end{aligned} \tag{8}$$

where we remember that $A^2\cos^2\phi + A^2\sin^2\phi = A^2$, and a similar relationship holds for B. Using the trigonometric relationship $\cos(a-b) = \cos a\cos b + \sin a \sin b$, the amplitude can be written

$$R = \sqrt{A^2 + B^2 + 2AB\cos(\phi - \theta)}. \tag{9}$$

The largest oscillations are when $\cos(\phi-\theta) = 1$, when $R = \sqrt{A^2 + B^2 + 2AB} = A + B$. The smallest oscillations are when $\cos(\phi - \theta) = -1$, when $R = \sqrt{A^2 + B^2 - 2AB} = |A - B|$. If $A = B$, the maximum amplitude is $2A = 2B$ while the minimum is 0 (complete cancellation).

The difference in phases $\Delta\phi = \phi - \theta$ is called the **phase difference**. This parameter is vital if we want to work out the way the waves will interfere. The condition for the largest resultant amplitude is that $\cos\Delta\phi = 1$, so $\Delta\phi = 2n\pi$. This leads to a peak of one wave lining up with a peak of the other.



Many students prefer to visualise this rather than rely on the equations. If we use our language of C and S begin components of something similar to a vector, then we see in this case the two individual waves having (C, S) of $(A\cos\phi, A\sin\phi)$ and $(B\cos\theta, B\sin\theta)$, which add vectorially to make a wave with components $(A\cos\phi + B\cos\theta, A\sin\phi + B\sin\theta)$. This can be drawn in the diagram above as a vector sum of a vector A in direction ϕ and a vector B in direction θ, with angle $\pi - \Delta\phi$ opposite the resultant. You can then use the cosine rule to obtain $R^2 = A^2 + B^2 - 2AB\cos(\pi - \Delta\phi)$. As $\cos(\pi - a) = -\cos a$, we have $R^2 = A^2 + B^2 + 2AB\cos\Delta\phi$ as before.

5 Beats

When two musical notes with very little difference in frequency (pitch) are played, a pulsing sound is heard. This is called 'beats'. This phenomenon can be explained in terms of oscillating functions. Suppose the two oscillations have frequencies of f and $f + g$. The angular frequencies will be ω and $\omega + \sigma$ where $\omega = 2\pi f$ and $\sigma = 2\pi g$. If both oscillations have $\phi_0 = 0$, then at time t, the phases of the two oscillations will be ωt and $(\omega + \sigma)\,t = \omega t + \sigma t$. The phase difference between the oscillations is therefore σt.

We will have a strong signal when σt represents a whole number of complete oscillations, as then the two sounds will peak at the same time. After $t = 0$, the next time when the two sounds will constructively interfere is when $\sigma t = 2\pi$. This occurs when $t = 2\pi/\sigma$.

The number of beats each second will be $1/t = \sigma/2\pi$. This is called the beat frequency. As $\sigma = g/2\pi$, the beat frequency $1/t = g$ is equal to the difference in the original frequencies.

The situation can also be analysed by writing the waves as cosine functions. Again, we use the trigonometric identiy $\cos(a + b) = \cos a \cos b - \sin a \sin b$.

$$\begin{aligned} y &= A\cos\omega t + B\cos(\omega t + \sigma t) \\ &= A\cos\omega t + B\cos\sigma t\cos\omega t - B\sin\sigma t\sin\omega t \end{aligned} \tag{10}$$

This is an oscillation in the form $y = C\cos\omega t - S\sin\omega t$ where

$$\begin{aligned} C &= A + B\cos\sigma t \\ S &= B\sin\sigma t. \end{aligned} \tag{11}$$

The overall amplitude is given by equation 5:

$$\begin{aligned} R &= \sqrt{C^2 + S^2} \\ &= \sqrt{(A + B\cos\sigma t)^2 + B^2\sin^2\sigma t} \\ &= \sqrt{A^2 + B^2 + 2AB\cos\sigma t}, \end{aligned} \tag{12}$$

where we remembered that $\sin^2\sigma t + \cos^2\sigma t = 1$.

We will get loud sounds when the amplitude is greatest. This is when $\cos\sigma t = 1$, which is when $\sigma t = 2n\pi$, so $t = 2n\pi/\sigma$. The time between loud sounds is therefore $2\pi/\sigma$, which agrees with our earlier result that the beat frequency will be $1/t = \sigma/2\pi = g$.

6 Waves

So far, we have looked at oscillations. Next we want to look at waves. Our first wave is moving along the x axis from left to right at speed c. We wish to know $y(x, t)$, in otherwords the displacement caused by the wave at any position and time. To do this, we remember that whatever is at x at time t was at the origin ($x = 0$) at an earlier time. It took time x/c for this wave to travel from the origin, and therefore whatever is at x at time t was at the origin at time $t - x/c$. Let the oscillation at the origin $y(x = 0, t) = A\cos(\omega t + \phi_0)$. We can then write

$$y_{\rightarrow}(x,t) = A\cos\left\{\omega\left(t - \frac{x}{c}\right) + \phi_0\right\} = A\cos\left(\omega t - \frac{\omega x}{c} + \phi_0\right). \quad (13)$$

We simplify this important equation by defining a new quantity k (the wave number):

$$k = \frac{\omega}{c} = \frac{2\pi f}{c} = 2\pi\frac{f}{c} = 2\pi\frac{1}{\lambda} = \frac{2\pi}{\lambda}, \quad (14)$$

where we remember that $c = f\lambda$ for a wave where λ is the wavelength.

Then our wave equation becomes

$$y_{\rightarrow}(x,t) = A\cos(\omega t - kx + \phi_0). \quad (15)$$

For a wave moving to the left, whatever is at $-x$ at time t was at the origin at time $t - |x|/c = t + x/c$. This enables us to write

$$\begin{aligned} y_{\leftarrow}(x,t) &= A\cos\left\{\omega\left(t + \frac{x}{c}\right) + \phi_0\right\} \\ &= A\cos(\omega t + kx + \phi_0). \quad (16) \end{aligned}$$

The difference between $y_{\rightarrow}$ and $y_{\leftarrow}$ is in the sign of the kx term. Changing the sign of k effectively changes the direction of the wave. When we come to waves in more than one dimension, this becomes more apparent, and we then refer to a **wave vector $\mathbf{k}$** of magnitude $2\pi/\lambda$ which points in the direction of wave propagation.

7 Waves and interference

Let us imagine a wave from a source $y(x=0,t) = A\cos(\omega t + \phi_0)$ which arrives at a point by two routes. One route has a length L_1, the other L_2 where we will define the **path difference** $\Delta L = |L_2 - L_1|$. For each wave, we will use x to label how far that wave is along its path. By the time the waves combine, the first wave will have $x_1 = L_1$, while the second will have $x_2 = L_2$ even though those waves are in the same place. The two waves will not necessarily have the same amplitude by the time they recombine, so let us write the amplitudes of the two parts A_1 and A_2.

If we use the trigonometric identity $\cos(a-b) = \cos a \cos b + \sin a \sin b$, and use the substitution $a = \omega t + \phi_0$, the combined oscillation will be

$$\begin{aligned} y &= A_1 \cos(\omega t - kL_1 + \phi_0) + A_2 \cos(\omega t - kL_2 + \phi_0) \\ &= A_1 \cos kL_1 \cos(\omega t + \phi_0) + A_1 \sin kL_1 \sin(\omega t + \phi_0) \\ &\quad + A_2 \cos kL_2 \cos(\omega t + \phi_0) + A_2 \sin kL_2 \sin(\omega t + \phi_0) \\ &= (A_1 \cos kL_1 + A_2 \cos kL_2)\cos(\omega t + \phi_0) \\ &\quad + (A_1 \sin kL_1 + A_2 \sin kL_2)\sin(\omega t + \phi_0) \end{aligned} \tag{17}$$

As in section 4 the overall amplitude can be evaluated from the amplitudes of the sine and cosine components $C = A_1 \cos kL_1 + A_2 \cos kL_2$ and $S = -(A_1 \sin kL_1 + A_2 \sin kL_2)$:

$$\begin{aligned} R &= \sqrt{(A_1 \cos kL_1 + A_2 \cos kL_2)^2 + (A_1 \sin kL_1 + A_2 \sin kL_2)^2} \\ &= \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos kL_1 \cos kL_2 + 2A_1A_2 \sin kL_1 \sin kL_2} \\ &= \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(kL_1 - kL_2)} \end{aligned} \tag{18}$$

This is analagous to equation 9, in which the resultant amplitude was determined by a phase difference $\Delta\phi$. Here the phase difference,

$$\Delta\phi = kL_1 - kL_2 = k(L_2 - L_1) = \frac{2\pi}{\lambda}(L_2 - L_1), \tag{19}$$

is related to the path difference $\Delta L = |L_2 - L_1|$.

When we solve problems in wave interference, the first task is to calculate the path difference ΔL. From this, we can work out the phase difference

$\Delta\phi = k\Delta L$. The resultant amplitude can then be evaluated from

$$R = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos(k\Delta L)}. \tag{20}$$

This is not too difficult to work out when there are just two routes. Things get much more complicated once multiple routes are involved (such as when handling a diffraction grating or wide slit), when a large number of cosine functions need to be added.

8 Standing waves

Let's see what happens in a region where a left-moving and right-moving wave co-exist. We add the functions $y_{\leftarrow}$ and $y_{\rightarrow}$ from equations 15 and 16 to gain

$$\begin{aligned} y_{\text{total}} &= y_{\leftarrow} + y_{\rightarrow} = A\cos(\omega t + kx) + A\cos(\omega t - kx) \\ &= 2A\cos\left(\frac{\omega t + kx + \omega t + kx}{2}\right)\cos\left(\frac{\omega t + kx - (\omega t - kx)}{2}\right) \\ &= 2A\cos(\omega t)\,\cos(kx)\,, \end{aligned} \tag{21}$$

where we used the trigonometric identity we shall prove later as equation 26.

Each part of the wave oscillates up and down, but the amplitude of the oscillation depends on x. When $kx = 0, \pi, 2\pi, 3\pi$ and so on, $|\cos kx| = 1$ and the total wave has amplitude $2A$. These positions are antinodes. When $kx = \frac{1}{2}\pi, \frac{3}{2}\pi, \frac{5}{2}\pi$ and so on, $\cos kx = 0$ and there is no oscillation at all. These points are called nodes.

An adjacent node and antinode are separated by a phase $k\,\Delta x = \frac{1}{2}\pi$, which means that

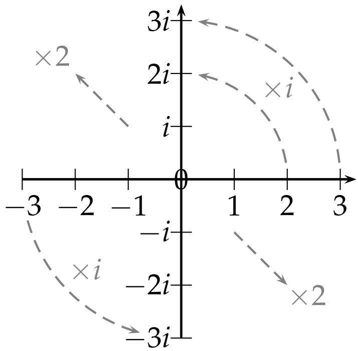
$$\Delta x = \frac{\pi}{2k} = \frac{\pi\lambda}{4\pi} = \frac{\lambda}{4}. \tag{22}$$

Adjacent antinodes (or adjacent nodes) are separated by a phase of $k\,\Delta x = \pi$ and therefore a distance $\Delta x = \frac{1}{2}\lambda$.

9 Complex numbers

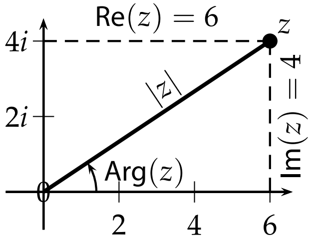
Sines and cosines are useful functions for handling problems involving waves and oscillations, but they do suffer from drawbacks. It is not easy to tell the overall amplitude and phase from a wave written as a sine plus a cosine component. Furthermore, when differentiated they swap over (eg. $\frac{d}{d\theta}\sin\theta = \cos\theta$) making analysis of differential equations more difficult.

We can avoid these difficulties if we use complex numbers. Complex numbers use a 2-dimensional number line called an **Argand diagram**. The horizontal axis is called the **real** axis and contains all of the normal numbers like 2, -3.5 and π. The vertical axis is called the **imaginary** axis, and components of a number in this direction are written with i. So the equivalent of 1 on the vertical axis is labelled i, and the equivalent of -3 is written $-3i$.

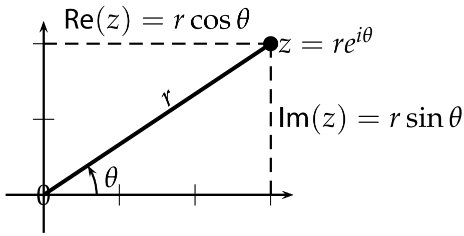


On this diagram, if you multiply a number by 2 (eg. $1 \to 2$ or $i \to 2i$) it gets twice as far from the origin. If you multiply a real number by i, then this rotates its location on the diagram by $\pi/2$ anticlockwise (eg. $2 \to 2i$ or $-3 \to -3i$). We define $i^2 = -1$ so that multiplication by i rotates the location of **any** point on the diagram by $\pi/2$ anticlockwise (eg $2i \to 2i \times i = 2i^2 = -2$ and $-3i \to -3i \times i = -3i^2 = +3$). This means that $-i \times i = 1$ and therefore $i^{-1} = -i$.

Complex numbers can be written in two ways. The first way, which is analagous to writing vectors in Cartesian form $(3,4)$ is $z = 6 + 4i$. This number has a real part Re$(z) = 6$ and an imaginary part Im$(z) = 4$. This is shown on an Argand diagram below.



You can also write a complex number with its distance from the origin and the angle it makes to the real axis. The distance to the origin is called the **modulus** of the number. Here $|z| = \sqrt{6^2 + 4^2} = \sqrt{52}$. The angle to the real axis is called the **argument** of the number and is always given in radians. Here Arg$(z) = \tan^{-1}(4/6) = 0.588$ rad to three significant figures.



A remarkable and beautiful formula named after Euler[1] states that $e^{i\theta} = \cos\theta + i\sin\theta$. On the Argand diagram, this means that $e^{i\theta}$ is on the unit circle with argument θ. Thought about vectorially, this means that we can write the complex number z whose modulus is r and argument θ as $z = re^{i\theta} = r\cos\theta + ir\sin\theta$, as shown in the diagram above.

If we multiply two complex numbers $z = re^{i\theta}$ and $w = ue^{i\phi}$, then we will get

$$zw = re^{i\theta}\, ue^{i\phi} = rue^{i\theta}e^{i\phi} = rue^{i(\theta+\phi)}. \tag{23}$$

In other words, when complex numbers are multiplied together, the modulus of the product is the product of the moduli $|zw| = ru = |z||w|$, and the argument is the sum of the original arguments Arg$(zw) = \theta + \phi =$ Arg$(z) +$ Arg(w).

[1] To prove this relationship, you can use the Taylor expansion of $\cos\theta + i\sin\theta$ and show that it is the same as the expansion of $e^{i\theta}$ if you assume that $i^2 = -1$. Alternatively, take $S = e^{-i\theta}(\cos\theta + i\sin\theta)$ and differentiate it with respect to θ. $S = 1$ when $\theta = 0$, and as you will have shown that $dS/d\theta = 0$ for all θ, this means that $S = 1$ always, and thus that $e^{i\theta} = \cos\theta + i\sin\theta$.

We define the **complex conjugate** of a complex number $z = x + iy$ as $z^\star = x - iy$. You can think of this as a reflection in the real axis: we switch the sign of the imaginary part, but keep the real part the same. An alternative view is that we switch the sign of the argument while keeping the modulus the same $\left(re^{i\theta}\right)^\star = re^{-i\theta}$.

Multiplying a number $z = re^{i\theta}$ by its complex conjugate $z^\star = re^{-i\theta}$ gives r^2, which is the square of the modulus. This is known as the **modulus square** of the number. If you are given a complex number in Cartesian form $z = x + iy$, then the easiest way of calculating the modulus is to multiply by the complex conjugate and then take the square root. As $i^2 = -1$, we see that this fits with what we would expect from Pythagoras' Theorem:

$$zz^\star = (x + iy)(x - iy) = x^2 + ixy - ixy - i^2y^2 = x^2 + y^2. \tag{24}$$

We shall apply our knowledge to oscillations in the next section, but before we do, let's note how easy it is to work out trigonometric identities using Euler's equation:

$$\begin{aligned}
\cos(a+b) &= \text{Re}\left(e^{i(a+b)}\right) \\
&= \text{Re}\left(e^{ia}e^{ib}\right) \\
&= \text{Re}\left\{(\cos a + i\sin a)(\cos b + i\sin b)\right\} \\
&= \text{Re}\left(\cos a\cos b + i\cos a\sin b + i\sin a\cos b + i^2\sin a\sin b\right) \\
&= \text{Re}\left\{\cos a\cos b - \sin a\sin b + i(\cos a\sin b + \sin a\cos b)\right\} \\
&= \cos a\cos b - \sin a\sin b \qquad (25) \\
\cos A + \cos B &= \text{Re}\left(e^{iA} + e^{iB}\right) \\
&= \text{Re}\left(e^{i(A+B)/2}e^{i(A-B)/2} + e^{i(A+B)/2}e^{i(B-A)/2}\right) \\
&= \text{Re}\left\{e^{i(A+B)/2}\left(e^{i(A-B)/2} + e^{-i(A-B)/2}\right)\right\} \\
&= \text{Re}\left\{e^{i(A+B)/2} \times 2\cos\left(\frac{A-B}{2}\right)\right\} \\
&= 2\cos\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right), \qquad (26)
\end{aligned}$$

where we have used the fact that $e^{i\theta} + e^{-i\theta} = (\cos\theta + i\sin\theta) + (\cos\theta - i\sin\theta) = 2\cos\theta$.

10 Oscillations with complex numbers

Using complex numbers, we take equation 3

$$y = A\cos(\omega t + \phi_0), \tag{27}$$

and rewrite it

$$y = \mathrm{Re}\left(Ae^{i(\omega t + \phi_0)}\right) = \mathrm{Re}\left(Ae^{i\omega t}e^{i\phi_0}\right). \tag{28}$$

For convenience, we often don't bother to write the Re at the front. Effectively, we define

$$z = Ae^{i\phi_0}e^{i\omega t}, \tag{29}$$

and remember that the actual displacement $y = \mathrm{Re}(z)$. To see how we analyse a practical situation, let us look at the addition of a wave $A\cos(\omega t + \phi)$ to a second wave $B\cos(\omega t + \theta)$. Written using complex numbers, this becomes

$$\begin{aligned} z &= Ae^{i\omega t}e^{i\phi} + Be^{i\omega t}e^{i\theta} \\ &= \left(Ae^{i\phi} + Be^{i\theta}\right)e^{i\omega t}. \end{aligned} \tag{30}$$

The part multiplying the $e^{i\omega t}$, namely $Ae^{i\phi} + Be^{i\theta}$, is called the complex amplitude as it gives all of the information of the strength and phase of the wave. If we want to know how strong the signal will be, we find the modulus square of the wave and find

$$\begin{aligned} z^\star z &= \left(Ae^{-i\phi} + Be^{-i\theta}\right)e^{-i\omega t} \times \left(Ae^{i\phi} + Be^{i\theta}\right)e^{i\omega t} \\ &= \left(Ae^{-i\phi} + Be^{-i\theta}\right)\left(Ae^{i\phi} + Be^{i\theta}\right)e^{-i\omega t}e^{i\omega t} \\ &= A^2 + B^2 + ABe^{i(\theta - \phi)} + ABe^{i(\phi - \theta)} \\ &= A^2 + B^2 + 2AB\cos(\phi - \theta). \end{aligned} \tag{31}$$

This is the same as the square of the resultant amplitude R of equation 8. We therefore have a way to work out the amplitude of a combined oscillation. We write each oscillation as a complex number, add them, find the modulus square $z^\star z = R^2$ and finally take the square root.

An extra simplification can be made if we note that the oscillatory part $e^{i\omega t}$ vanished from the expression at the second stage of equation 31 because it was multiplied by its reciprocal $e^{-i\omega t}$. Thus this part can be neglected in situations like this one.

In fact there is an even easier way. Let us work with the complex amplitude of the combined wave, and remembering that we define $\Delta\phi = \phi - \theta$, we can write the complex amplitude

$$\begin{aligned} Z &= Ae^{i\phi} + Be^{i\theta} = Ae^{i\phi} + Be^{i(\phi+\Delta\phi)} \\ &= e^{i\phi}\left(A + Be^{i\Delta\phi}\right) \\ R^2 &= z^\star z = \left(A + Be^{-i\Delta\phi}\right)\left(A + Be^{i\Delta\phi}\right) \\ &= A^2 + B^2 + ABe^{i\Delta\phi} + ABe^{-i\Delta\phi} \\ &= A^2 + B^2 + 2AB\cos\Delta\phi. \end{aligned} \tag{32}$$

So far, you might not consider this approach worth the trouble. However even here, we did not need to remember any trigonometric identities. When we end up adding lots of oscillations, as in section 14, we will find that the complex numbers make things much easier. We also find things easier in differential equations, such as the one governing a forced, damped oscillation.

11 Forced, damped oscillations

A damped oscillator is a mass m attached to a spring of constant k and a shock absorber of constant r. The two forces acting on the mass are $F = -kx$ from the spring, pulling it back into the equilibrium position $x = 0$; and the force resisting motion in the shock absorber $F = -rv = -r\dot{x}$. We assume that the shock absorber gives a force proportional to the speed of the mass, acting to slow it down. Newton's second law for this situation is written $F = ma = -rv - kx$, or $m\ddot{x} + r\dot{x} + kx = 0$ where we use dots to represent differentiation with respect to time.

Here we will consider the situation where there is a periodic force $F = F_0 \cos\omega t$ also acting on the mass. This situation is a good model for many practical situations from radio tuners to charged particles in an electromagnetic wave to buildings in an earthquake. Our equation now becomes $ma = -rv - kx + F_0 \cos\omega t$. We say that x will be the real part of a complex variable z, and we notice that $F_0 \cos\omega t$ is the real part of $F_0 e^{i\omega t}$. Our equation for z then becomes

$$m\ddot{z} = -r\dot{z} - kz + F_0 e^{i\omega t}. \tag{33}$$

Furthermore, we assume that (after any transient oscillations have died away) there can only be oscillations at the driving frequency ω. Therefore, we write $z = Ae^{i\omega t}$ and we try this out as a solution of the equation. Using the chain rule, $\dot{z} = i\omega e^{i\omega t} = i\omega z$. Similarly $\ddot{z} = d\dot{z}/dt = i^2\omega^2 z = -\omega^2 z$. Putting these into equation 33, we get

$$-m\omega^2 z = -i\omega r z - kz + F_0 e^{i\omega t}, \tag{34}$$

and as $z = Ae^{i\omega t}$, this is equal to

$$-m\omega^2 A = -i\omega r A - kA + F_0. \tag{35}$$

This can be rearranged to make A the subject

$$A = \frac{F_0}{k - m\omega^2 + ir\omega}. \tag{36}$$

If the mass were just pushed by a steady force F_0, it would move a distance $x_0 = F_0/k$. In terms of x_0, the complex amplitude of the oscillation becomes

$$A = \frac{x_0}{1 - \frac{m\omega^2}{k} + \frac{ir\omega}{k}}. \tag{37}$$

The amplitude can be worked out by evaluating the modulus square, however hopefully you can see that the peak amplitude is going to be when the denominator is as small as possible. If the damping is small, and the r term is ignored, the denominator goes to zero when $k = m\omega^2$, which is when $\omega = \sqrt{k/m}$. You may be aware that this is the **natural frequency** of the system: the frequency at which the mass will oscillate by itself on the spring. We can see that as $\omega \to 0$, $A \to x_0$, and that as $\omega \to \infty$, $A \to 0$, but A can be quite large in between. This explains resonance excellently.

If you had tried doing this without the complex numbers, you would have had sine **and** cosine functions in the differential equation. This would have been much more cumbersome to solve.

The phase of the oscillation can also be worked out from the argument of A. Let us get a flavour of this without doing the calculation. When $\omega = 0$,

A is real, and has the same sign as z_0, so the phase is 0. When $\omega \rightarrow \infty$, only the ω^2 term is significant, and $A = -kx_0/m$. Here A is real, but of opposite sign to x_0, so the oscillation is exactly out of phase with respect to the driving force. Finally, when $\omega^2 = k/m$, A is entirely imaginary (it has no real part) which tells you that the oscillation will be $\pi/2$ radians (one quarter of a period) out of phase with respect to the force.

12 Waves with complex numbers

For oscillations, we wrote $y = A\cos(\omega t + \phi_0)$ as the real part of $z = Ae^{i\phi_0}e^{i\omega t}$. Equation 15 describes a wave travelling to the right thus:

$$y_{\rightarrow} = A\cos(\omega t - kx + \phi_0)\,.$$

The equivalent using complex numbers is $y_{\rightarrow} = \mathrm{Re}\,(z_{\rightarrow})$ where

$$z_{\rightarrow} = Ae^{i(\omega t - kx + \phi_0)} = Ae^{i\phi_0}e^{-ikx}e^{i\omega t}. \tag{38}$$

Similarly for a wave moving to the left $y_{\leftarrow} = \mathrm{Re}\,(z_{\leftarrow})$ where

$$z_{\leftarrow} = Ae^{i(\omega t + kx + \phi_0)} = Ae^{i\phi_0}e^{ikx}e^{i\omega t}. \tag{39}$$

Let us now think about two waves (travelling to the right) which have travelled two different distances $x_1 = L$ and $x_2 = L + \Delta L$, and which have amplitudes A_1 and A_2. The total displacement will be $y_T = \mathrm{Re}\,(z_T)$ where

$$\begin{aligned}
z_T &= A_1e^{i\phi_0}e^{-ikx_1}e^{i\omega t} + A_2e^{i\phi_0}e^{-ikx_2}e^{i\omega t} \\
&= A_1e^{i\phi_0}e^{-ikL}e^{i\omega t} + A_2e^{i\phi_0}e^{-ik(L+\Delta L)}e^{i\omega t} \\
&= A_1e^{i\phi_0}e^{-ikL}e^{i\omega t} + A_2e^{i\phi_0}e^{-ikL}e^{ik\Delta L}e^{i\omega t} \\
&= \left(A_1 + A_2e^{-ik\Delta L}\right)e^{i\phi_0}e^{-ikL}e^{i\omega t}.
\end{aligned} \tag{40}$$

From here, we refer to the part in brackets as the complex amplitude

$$Z_T = A_1 + A_2e^{-ik\Delta L}. \tag{41}$$

The resultant amplitude R will be the square root of the modulus square of z_T. In symbols $R^2 = z_T^\star z_T$. We have

$$\begin{aligned}
R^2 = z_T^\star z_T &= Z_T^\star e^{-i\phi_0}e^{+ikL}e^{-i\omega t} \times Z_Te^{i\phi_0}e^{-ikL}e^{i\omega t} \\
&= Z_T^\star Z_T \times e^{-i\phi_0}e^{+i\phi_0}e^{+ikL}e^{ikL}e^{-i\omega t}e^{+i\omega t} \\
&= Z_T^\star Z_T.
\end{aligned} \tag{42}$$

The value of R^2 is therefore given by

$$\begin{aligned}
R^2 &= Z_T^\star Z_T = \left(A_1 + A_2 e^{+ik\Delta L}\right)\left(A_1 + A_2 e^{-ik\Delta L}\right) \\
&= A_1^2 + A_2^2 + A_1 A_2 \left(e^{+ik\Delta L} + e^{-ik\Delta L}\right) \\
&= A_1^2 + A_2^2 + 2A_1 A_2 \cos k\Delta L, \qquad (43)
\end{aligned}$$

in agreement with equation 20.

In future we will make things more straightforward by writing the complex amplitude Z_T directly from the knowledge of ΔL using equation 41, and then evaluate R as its modulus. In other words, we will not need to worry about the $e^{i\omega t}$, e^{-ikL} or $e^{i\phi_0}$ terms.

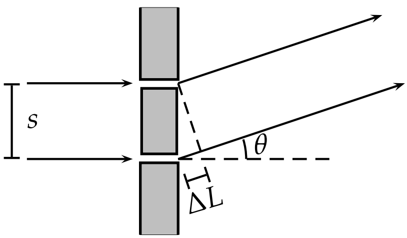
The emergence of standing waves is clear when using complex numbers too. If we use equations 38 and 39, we find

$$\begin{aligned}
z_\leftarrow + z_\rightarrow &= A e^{i\phi_0} e^{+ikx} e^{i\omega t} + A e^{i\phi_0} e^{-ikx} e^{i\omega t} \\
&= A e^{i\phi_0} e^{i\omega t}\left(e^{+ikx} + e^{-ikx}\right) \\
&= A e^{i\phi_0} e^{i\omega t} \times 2\cos kx. \qquad (44)
\end{aligned}$$

and the amplitude of the oscillation is given by the factor multiplying the $e^{i\phi_0 + i\omega t}$ term, namely $2A\cos kx$ as we found in equation 22.

13 Two sources

In the diagram below, we have two identical sources of wave (in phase with each other) illuminating a distant screen on the right. The source separation is s.



Given that the rays reaching the screen will be very nearly parallel, the path difference $\Delta L = s\sin\theta$ where θ is the angle of the rays to the horizontal axis. Given that the two rays will have travelled virtually identical distances, we will assume that their amplitudes are the same A. Using equation 41 we write the complex amplitude

$$Z = A + Ae^{-ik\Delta L} = A\left(1 + e^{-ik\Delta L}\right). \tag{45}$$

The intensity on the screen will be related to the amplitude squared, and thus to $R^2 = Z^\star Z$. We could work this out using equation 43 setting $A_1 = A_2 = A$ and getting $R^2 = 2A^2(1 + \cos k\Delta L)$. A trignometric identity $1 + \cos 2a = 2\cos^2 a$ then allows us to simplify our expression to $R^2 = 4A^2\cos^2(k\Delta L/2)$.

Another method, which can be quicker, is to take a mean phase factor outside of the bracket of equation 45. Here the two phases are 0 and $-ik\Delta L$. The mean of these is $-ik\Delta L/2$. If we take this factor outside the bracket, we are left with

$$\begin{aligned} Z &= Ae^{-ik\Delta L/2}\left(e^{ik\Delta L/2} + e^{-ik\Delta L/2}\right) \\ &= Ae^{-ik\Delta L/2} \times 2\cos\left(\frac{k\Delta L}{2}\right), \end{aligned} \tag{46}$$

from which we can ignore the phase factor at the front to see that $R = 2A\cos(k\Delta L/2)$ and thus R^2 will be as calculated above, but without needing to know the trigonometric identity.

Our reasoning has produced

$$R^2 = 4A^2\cos^2\left(\frac{k\Delta L}{2}\right) = 4A^2\cos^2\left(\frac{2\pi}{\lambda}\frac{s\sin\theta}{2}\right). \tag{47}$$

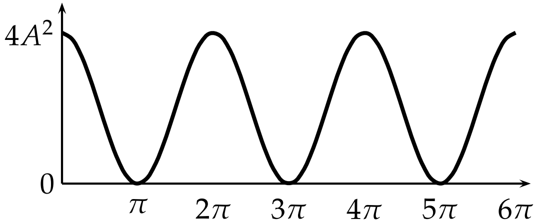
The function $\cos^2 x$ has its maxima when $x = n\pi$ where n is an integer. Thus constructive interference occurs when

$$\frac{2\pi}{\lambda}\frac{s\sin\theta}{2} = n\pi,$$

which is when

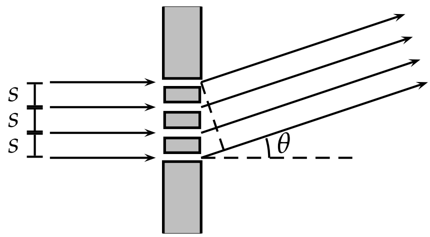
$$s\sin\theta = n\lambda, \tag{48}$$

a fact which you were probably already aware of. However our new formula enables you to calculate the total amplitude at any angle. You can also see that there will be fully destructive interference when the cosine is zero, which is when, for example $s \sin\theta = \lambda/2$. The graph of intensity as a function of $ks \sin\theta$ is shown below.



14 Multiple sources

Here we consider N slits, as in the diagram below.



If we take the distance from the top slit to the screen as L, each slit below it has an extra distance $\Delta L = s \sin\theta$ to travel. Thus the path difference for the pth slit will be $p\Delta L = ps \sin\theta$, where p runs from 0 for the first slit to $N-1$ for the last. If we take the amplitudes as all being equal, the total complex amplitude will be

$$
\begin{aligned}
Z &= \sum_{p=0}^{N-1} A e^{-ikp\Delta L} \\
&= A \frac{1 - e^{-ikN\Delta L}}{1 - e^{-ik\Delta L}},
\end{aligned}
$$

where we use the usual result for summing geometric series[2].

[2] If you need a proof, let us set $S = \sum_{n=0}^{N-1} a^n$. Then $aS = \sum_{n=0}^{N-1} a^{n+1} = \sum_{n=1}^{N} a^n$. We can then write $aS - S = a^N - 1$, which means $S(a-1) = a^N - 1$ so $S = \left(1 - a^N\right) / (1 - a)$.

To find the resultant amplitude, we use our trick of taking a mean phase factor out of the brackets:

$$\begin{aligned} Z &= A\frac{e^{-ikN\Delta L/2}}{e^{-ik\Delta L/2}} \times \frac{e^{ikN\Delta L/2} - e^{-ikN\Delta L/2}}{e^{ik\Delta L/2} - e^{-ik\Delta L/2}} \\ &= A\frac{e^{-ikN\Delta L/2}}{e^{-ik\Delta L/2}} \times \frac{\sin(kN\Delta L/2)}{\sin(k\Delta L/2)}, \end{aligned} \tag{49}$$

where we note that $e^{ia} - e^{-ia} = (\cos a + i\sin a) - (\cos a - i\sin a) = 2i\sin a$. The resultant square amplitude (which will be related to the intensity) is equal to the modulus square of Z. Calculating the modulus square removes the phase factors at the front, just as it did in equation 42. The result is

$$R^2 = \frac{\sin^2(kN\Delta L/2)}{\sin^2(k\Delta L/2)}. \tag{50}$$

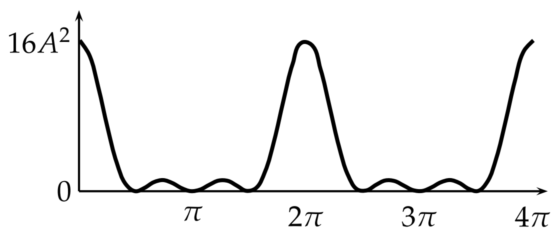
From this, given that $\sin n\pi = 0$ if n is an integer, we can see we get fully destructive interference when the numerator is zero, and hence when

$$\frac{2\pi}{\lambda}\frac{Ns\sin\theta}{2} = n\pi,$$

which is when

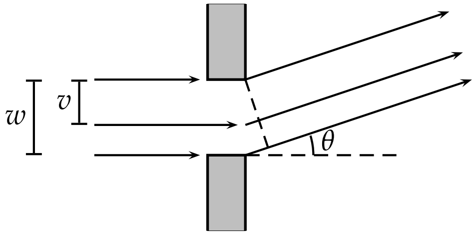
$$s\sin\theta = \frac{n\lambda}{N}. \tag{51}$$

The fully constructive points are when top and bottom both tend to zero at the same rate, and are at the same values of θ as in equation 48. We therefore see that as θ increases, there are $N-1$ fully destructive points between each pair of fully constructive interferences. The graph of intensity as a function of $ks\sin\theta$ is shown in the graph below for $N = 4$. Fully constructive interference occurs when $s\sin\theta = 0, \lambda, 2\lambda$ and so on, while destructive interference occurs when $s\sin\theta = \frac{1}{4}\lambda$, $\frac{2}{4}\lambda$, $\frac{3}{4}\lambda$, (not $\frac{4}{4}\lambda$ as that is constructive), $\frac{5}{4}\lambda$, $\frac{6}{4}\lambda$, $\frac{7}{4}\lambda$ and so on.



15 Wide source

Here we consider a wide source of width w, as in the diagram. We assume that each part of the wide source acts as a separate source of waves. Effectively, we consider an infinite number of little sources next to each other in the gap. We label each part of the gap with its distance from the top v. The top of the gap has $v = 0$, the bottom has $v = w$.

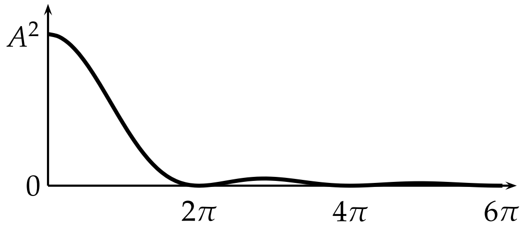


If we take the distance to the screen from the top of the slit as L, then the path difference ΔL relative to that ray will be $v \sin\theta$. Now that we have an infinite number of mini sources, we use an integral to add up their contributions. Let us write A/w as the amplitude of wave per unit width of the slit, so that each dv of slit width contributes an amplitude $A\,dv/w$. This means that

$$\begin{aligned}
Z &= \int_{v=0}^{w} \frac{Ae^{-ikv\sin\theta}}{w} dv \\
&= \left[\frac{Ae^{-ikv\sin\theta}}{-ikw\sin\theta}\right]_{v=0}^{w} \\
&= \frac{Ae^{-ikw\sin\theta} - A}{-ikw\sin\theta} \\
&= A\frac{1 - e^{-ikw\sin\theta}}{ikw\sin\theta} \\
&= Ae^{-ikw\sin\theta/2}\frac{e^{ikw\sin\theta/2} - e^{-ikw\sin\theta/2}}{ikw\sin\theta} \\
&= Ae^{-ikw\sin\theta/2}\frac{2i\sin\left(kw\sin\theta/2\right)}{ikw\sin\theta} \\
&= Ae^{-ikw\sin\theta/2}\frac{2\sin\left(kw\sin\theta/2\right)}{kw\sin\theta} && (52) \\
&= Ae^{-ikw\sin\theta/2}\frac{\sin\left(kw\sin\theta/2\right)}{kw\sin\theta/2} && (53)
\end{aligned}$$

$$R^2 = Z^\star Z = A^2 \frac{\sin^2(kw\sin\theta/2)}{(kw\sin\theta/2)^2} \tag{54}$$

We have fully destructive interference when $kw\sin\theta/2 = n\pi$, which happens when $w\sin\theta = n\lambda$. The graph of intensity as a function of $kw\sin\theta$ is shown below.



16 Partial differentation and Eigenfunctions

Ordinary differentiation concerns function y of a single variable, like x or t (but not both). The ordinary derivative dy/dx measures the change of the function y per unit change in the variable x. When a function, like a wave function, depends on x and t, we can not differentiate with respect to either without knowing what the other is doing.

To simplify matters, we define the **partial derivatives** $\partial y/\partial x$ and $\partial y/\partial t$. The derivative $\partial y/\partial x$ means the derivative of y with respect to x treating the other independent variable t as a constant. To give an example, suppose $y = 2x^2 + 3xt + 4t$. Then $\partial y/\partial x = 4x + 3t$, while $\partial y/\partial t = 3x + 4$.

Let us find the partial derivatives of our wave function $y = Ae^{i\omega t}e^{-ikx}$. We find

$$\frac{\partial y}{\partial t} = i\omega A e^{i\omega t} e^{-ikx} = i\omega y \tag{55}$$

$$\frac{\partial y}{\partial x} = -ikAe^{i\omega t}e^{-ikx} = -iky. \tag{56}$$

This means that

$$-i\frac{\partial y}{\partial t} = \omega y \tag{57}$$

$$i\frac{\partial y}{\partial x} = ky. \tag{58}$$

When an operator (such as $\partial/\partial t$) is applied to a function, and gives a multiple of the original function, we say that the function is an **eigenfunction** of the operator, and the multiple is the **eigenvalue**.

For the way we have defined our wavefunctions, $-i\partial/\partial t$ is a **frequency** operator, since if this is applied to our wavefunction y, the result will be ωy. Similarly $i\partial/\partial x$ is a **wavenumber** operator, because when it is applied to y the result is ky.

Before leaving this section, we can differentiate equations 55 and 56 again to yield

$$\frac{\partial^2 y}{\partial t^2} = i^2\omega^2 A e^{i\omega t}e^{-ikx} = -\omega^2 y \tag{59}$$

$$\frac{\partial^2 y}{\partial x^2} = (-i)^2 k^2 A e^{i\omega t}e^{-ikx} = -k^2 y. \tag{60}$$

It therefore follows that

$$\frac{\partial^2 y}{\partial x^2} = \frac{k^2}{\omega^2}\frac{\partial^2 y}{\partial t^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}, \tag{61}$$

since equation 14 tells us that $\omega = ck$ where c is the speed of the wave. Equation 61 is known as the **wave equation**. Wherever we find it, the situation concerned allows for waves to travel through at a speed c.

When James Clerk Maxwell unified the equations of electricity and magnetism, he noticed that both electric and magnetic fields obeyed an equation like 61 with $c = \sqrt{1/\epsilon_0\mu_0} = 3.00 \times 10^8$m/s. This was the first clue as to what was waving in a light wave: electric and magnetic fields.

17 Caution: Power

Up until now, when we have been interested in the strength of a wave, we have calculated its amplitude. In our calculations, the square of the amplitude, which is equal to the modulus square of the complex amplitude has often been given as this is often tidier mathematically.

The square of the amplitude is also important physically, as it is proportional to the intensity of a wave (or the energy stored in an oscillation). However, we do have to take care when calculating energies or intensities because anything proportional to y^2 is proportional to $(\mathrm{Re}\, z)^2$ and this is not the same as $z^\star z$ nor $\mathrm{Re}\,(z^2)$.

Let us first look an oscillation $z = Ae^{i(\omega t - kx)}$. Suppose the intensity is proportional to the square of $y = \mathrm{Re}\, z$. The intensity will be proportional to $(A\cos(\omega t - kx))^2 = A^2 \cos^2(\omega t - kx)$. In time, this averages to $\frac{1}{2}A^2$, as the $\cos^2$ function has an average value of $\frac{1}{2}$. This factor of one half is our recognition that the imaginary component of z is quite that — imaginary. Only the real [cosine] component can contribute energy to the system.

When many electrical components (such as capacitors and inductors) are connected to an alternating voltage, the current through them is not in phase with the voltage across them. Using complex notation, we can write the current as $I = I_0 e^{i\omega t}$ and the voltage as $V = V_0 e^{i(\omega t + \phi)} = V_0 e^{i\omega t} e^{i\phi}$ where ϕ is the phase difference between current and voltage.

The power dissipated in this component can not be $VI = V_0 I_0 e^{i(2\omega t + \phi)}$ as the power is measurable and must be real.

However the power is also not to be obtained as the modulus of VI as this would be $V_0 I_0$. Not only would this not take the time averaging into account, it would also imply that the phase ϕ were irrelevant.

Calculating the time average of $\mathrm{Re}\, VI = V_0 I_0 \cos(2\omega t + \phi)$ gives zero. Surely that can't be right either.

We obtain the actual power by remembering that only the real parts of V and I are physical. Then, using the relationship in equation 25, we have the power

$$
\begin{aligned}
P &= \mathrm{Re}\, V \times \mathrm{Re}\, I \\
&= V_0 \cos(\omega t + \phi) \times I_0 \cos(\omega t) \\
&= V_0 (\cos \omega t \cos \phi - \sin \omega t \sin \phi) \times I_0 \cos \omega t \\
&= V_0 I_0 \cos \phi \cos^2(\omega t) - V_0 I_0 \sin \phi \, \sin(\omega t) \cos(\omega t) \\
&= V_0 I_0 \cos \phi \cos^2(\omega t) - \tfrac{1}{2} V_0 I_0 \sin \phi \, \sin(2\omega t)
\end{aligned}
\tag{62}
$$

The time averaged power will then be given [correctly] as $\frac{1}{2} V_0 I_0 \cos \phi$. This is because $\cos^2 \omega t$ averages to $\frac{1}{2}$ and $\sin 2\omega t$ averages to zero.

18 Contrast with matter waves

Our studies so far have all been with real waves. y could be the displacement of a violin string from its usual position, the electric field at a point in a light wave, or the voltage across a component in an alternating current circuit. Since it was easier for us to solve the problem using complex numbers, we introduced z such that $\mathrm{Re}\,(z) = y$, where we didn't really mind what the imaginary part was, as it was not important to us. Only the real part of z had any physical meaning.

When we apply wave methods to quantum mechanics (wave mechanics), things are different. If we used our old approach to model a freely moving electron using $y = A\cos(\omega t - kx)$, with the intensity y^2 somehow related to the number of electrons at the point, we would have a pulsing of electron density along the line as the cosine squared function crescendos and then fades away repeatedly. This does not meet with observation, and is therefore false.

However, we do get a very good agreement with experiment if we let the electron wave function be a true complex number where both real and imaginary parts 'really' do mean something. Then we have $z = Ae^{i(\omega t - kx)}$. The equivalent of an intensity now truly is the modulus square $z^\star z = A^2$ which is steady and does not pulse if only one wavefunction is present in the system.

You may say, "What does the imaginary part of the wave actually represent?" It is a good question, but the answer is that as only real numbers can come out of scientific measurements, only functions of the wavefunction which produce real numbers (such as the modulus square) are physically meaningful in quantum mechanics.

The modulus square of the wavefunction represents the probability of finding the particle at that point.

To find other useful functions, we first have to own up to something unpleasant. Deep in the history of quantum mechanics, it was decided that our $z^\star = Ae^{-i(\omega t - kx)}$ would be used as the standard wavefunction not our z. We will write that as

$$\psi = Ae^{i(kx - \omega t)}. \tag{63}$$

Notice that our wave eigenfunctions (with the change in sign of exponent) now become

$$i\frac{\partial}{\partial t}\psi = \omega\psi \tag{64}$$

$$-i\frac{\partial}{\partial x}\psi = k\psi. \tag{65}$$

Let us define $\hbar = h/2\pi$. Using this, the Einstein equation $E = hf$ with $\omega = 2\pi f$ becomes $E = \hbar\omega$. De Broglie typothesis $\lambda = h/p$ means that $2\pi/k = h/p$ and so $p = \hbar k$. We can then write

$$p\psi = -i\hbar\frac{\partial}{\partial x}\psi \tag{66}$$

$$E\psi = i\hbar\frac{\partial}{\partial t}\psi. \tag{67}$$

The kinetic energy of a particle is given by $p^2/2m$ (where here p^2 means the p function applied twice), and if this is added to the potential energy V we would get the total energy E. It therefore follows that

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi + V\psi = E\psi = i\hbar\frac{\partial}{\partial t}\psi. \tag{68}$$

That is the Schrödinger Equation my friends — but alas our journey together ends there, and the story of that momentous equation and how it may be used is for another to tell on another day.

A Appendix: General diffraction problem

Here, let us define x, y co-ordinates marking points in the plane of our object, and X, Y marking points on the screen where the diffraction pattern will appear. Let the screen be distance Z from the object, and let us define $L^2 = X^2 + Y^2 + Z^2$. The distance from (x, y) to (X, Y) will be similar to L given that we assume that $x \ll X$ and $y \ll Y$. We calculate the distance using Pythagoras' Theorem, and write it as $L + \Delta L$. The path difference ΔL

can then be determined:

$$\begin{aligned}
\Delta L &= \sqrt{(X-x)^2+(Y-y)^2+Z^2}-L \\
&= \sqrt{X^2-2xX+x^2+Y^2-2yY-y^2+Z^2}-L \\
&= \sqrt{L^2-2xX-2yY+x^2+y^2}-L \\
&= L\left(1-2\frac{xX+yY}{L^2}+\frac{x^2+y^2}{L^2}\right)^{1/2}-L \qquad (69)\\
&= L-\frac{xX+yY}{L}+O\left(\frac{x^2}{L}\right)-L \\
&\approx -\frac{xX+yY}{L}. \qquad (70)
\end{aligned}$$

This approximation is good providing that Z is sufficiently larger than the width of the object so that the ignored terms do not constitute an appreciable fraction of a wavelength[3]. Let us use angle α to represent the deflection in the X direction, so $\sin\alpha = X/L$. Similarly, let β be the deflection in the Y direction, so $\sin\beta = Y/L$. Our path difference then becomes

$$\Delta L = -x\sin\alpha - y\sin\beta. \qquad (71)$$

To obtain the amplitude of wave at the position (X, Y) on the screen, we integrate over the object. Let the amplitude transmitted (per unit area) at (x, y) be $A(x, y)$. The amplitude transmitted by a small $dx\,dy$ area of the object will then be $dA = A(x, y)\,dx\,dy$, and we can obtain the complex amplitude arriving at (X, Y) using an equivalent of 41:

$$R(X,Y) = \int\int A(x,y)e^{-ik(x\sin\alpha+y\sin\beta)}\,dx\,dy. \qquad (72)$$

In later study, you will see that this equation has the same form as the equation used to produce two dimensional Fourier Transforms, and that there is therefore a strong similarity between the Fourier Transform of the object and its diffraction pattern.

[3]In other words, that $x^2/L \ll \lambda$ for all x values where the object transmits light. This is known as the Fraunhofer condition for diffraction. If you want to analyse things closer to the sources, a more complex analysis is needed — Fresnel diffraction.