

Oscillations and Waves

Dr. Andrew McKinley

2022-01-18

Contents

Welcome	7
About these notes	7
Error reporting	8
Course history	8
Changelog	8
Report an error	8
1 Simple Harmonic Motion	9
1.1 A simple example of SHM	9
1.2 Positioning in SHM	10
1.3 Velocity in SHM	11
1.4 Acceleration in SHM	11
1.5 Comparing displacement, velocity and acceleration	12
1.6 Initial conditions	13
1.7 Frequency and angular frequency	13
1.8 SHM and circular motion	14
1.9 Energy in SHM	16
2 SHM in Real Systems	19
2.1 General motion near equilibrium	19
2.2 Example: a diatomic molecule	20
2.3 Example: Mass on a vertical spring	22
2.4 The Simple Pendulum	24
2.5 The Physical Pendulum	25
3 Complex Numbers	29
3.1 Overview of complex numbers	29
3.2 Useful results for complex numbers	30
3.3 The Argand Diagram	31
3.4 Polar representation of complex numbers	33
3.5 Exponential representation of complex numbers	34
3.6 Complex representation of oscillations	34
3.7 Take-home points	35

4 Damped oscillations	37
4.1 The general case of damping	37
4.2 Light Damping	40
4.3 Critical damping	40
4.4 Overdamping	41
4.5 Quality factor and energy in damped SHM	41
5 Forced oscillations	43
5.1 The Transient Solution	44
5.2 The Steady State solution	44
5.3 Steady state behaviour - Resonance	44
5.4 Full solution of the forced oscillator	46
5.5 Energy in driven oscillators	49
5.6 Impedance	52
6 Simple wave motion and the Wave Equation	55
6.1 Wave pulses	57
6.2 Deriving the wave equation	59
6.3 The wave equation - proof by substitution	61
6.4 The Phase Velocity - the velocity of waves	62
6.5 Summary	64
7 Harmonic Waves	65
7.1 Transverse sine and cosine waves	65
7.2 Travelling waves	67
7.3 Complex representation of waves	68
7.4 Energy carried by waves on a string	68
7.5 Summary	72
8 Reflection and Transmission at boundaries	75
8.1 Power transmitted and reflected at a boundary	77
8.2 Example of reflection and transmission	79
8.3 The impedance of a piece of string	79
8.4 Reflection and transmission revisited	81
8.5 Impedance - Miscellaneous cases	82
9 Sound waves and the Doppler effect	85
9.1 Energy of sound waves	85
9.2 Wave intensity	86
9.3 Levels of intensity	87
9.4 The Doppler Effect (non-relativistic)	87
10 Superposition and Standing Waves	93
10.1 Superposition of harmonic waves	93
10.2 Two waves with same amplitude and frequency	94
10.3 Standing waves	96
10.4 Wave function for a standing wave	96

CONTENTS	5
10.5 Waves on strings fixed at both ends	98
10.6 Organ pipes and other wind instruments	100
10.7 Pipe closed at one end	101
11 Fourier analysis	103
11.1 Example: the square wave	104
11.2 The harmonic makeup of any wave	105
11.3 Fourier Transforms	108
11.4 Some important Fourier transforms	108
11.5 The Bandwidth Theorem	109
12 Wave packets and Dispersion	111
12.1 Two waves, same amplitude, different frequency (Beats)	111
12.2 Dispersion	114
12.3 Water waves	116
12.4 Sound waves in a crystal (phonons)	117
13 Interference and Diffraction	121
13.1 Conditions for interference	121
13.2 Propagation of Light - Huygens's Principle	122
13.3 The two-slit interference pattern	122
13.4 Diffraction from a single wide slit	125
13.5 Diffraction from two wide slits	127
13.6 Multiple slits with equal spacing	129
13.7 Fourier transforms in diffraction	130
13.8 Resolution limit of optical instruments	130
14 Properties of Light	135
14.1 Reflection and Refraction	135
14.2 Total internal reflection	138
15 Mirrors	141
15.1 Plane mirror: object and image	141
15.2 A spherical mirror	142
16 Lenses	147
16.1 Refraction at a spherical surface	147
16.2 Refraction through a thin lens	150
16.3 Types of lens	152
16.4 Ray diagrams	154
16.5 Combining lenses	155
17 The Wave-Particle duality	159
17.1 The de Broglie hypothesis	160
17.2 The wavefunction and its interpretation	161
17.3 Heisenberg's Uncertainty Principle	162
17.4 The "Particle in a box" problem	163

18 The Schrödinger Equation	167
18.1 Particle in a square potential well	168
18.2 Reflection and transmission of particle waves	171
18.3 Barrier penetration - ‘tunnelling’	172

Welcome

Welcome to the course notes for the Oscillations and Waves component of PHYS10005, Core Physics II.

The most recent version of these notes can be found at <https://awmckinley1.gitub.io/oscwaves>

About these notes

These notes have been prepared in a format which maintains compatibility with screen-readers, while also allowing the facility for both PDF and EPUB downloads for those who wish to use them.

The HTML also has themes to facilitate easier reading (font colours, serif/sans serif fonts etc.). Please explore the top bar of the web environment to explore this.

The HTML notes allow for embedding of rich content (videos, animations, etc.), and for this reason I recommend accessing the course using the web-links provided, as these cannot be included in static PDF or EPUB documents. However, I will ensure links are included in these formats as far as possible.

These notes are a “live document”, and as such they can (and will) be updated should any erroneous explanations be found or additional explanations be needed. For this reason, please do email me if you have any questions or if anything is not clear.

The most recent version can always be found at the link at the top of this section, and any changes are reflected in the change-log ()

Please note that as this is a “live document”, a downloaded version can become obsolete. Please refer back regularly and refer to the changelog for any updates.

Error reporting

As much as we try, sometimes errors creep in. At best, it is little more than a careless typo which makes the reader tut with disappointment, at worst it risks seriously misleading the learner. For this reason I have created a feedback form to report errors; this is accessible at the bottom of this page.

Course history

This iteration of the course has been written by Andrew McKinley from 2021. This has been based extensively on previous course notes by Ben Maughan, Simon Hanna and Massimo Antognozzi.

Changelog

2022-01-18 Updating the permalink and github repository address

2021-12-01 Fixing typos across the board

2021-03-16 Updated to fix images in chapters 17 and 18.

2021-03-16 Updated to address image problems in chapters 14-16. These should (hopefully!) now be working!

2021-03-07 Updated to address error reports, added additional detail for figures in Chapter 11, and including images for chapter 13

2021-02-14 Updated to address error reports, and to include images and additional explanations for Chapters 10 and 11.

2021-02-12 Updated to address error reports 1-13. Thank you all who sent them in.

2021-01-27 Initial commit to GitHub; all text for the course complete, some images pending.

2020-09-16 Initial versioning and inclusion of course material.

Report an error

Direct link to error reporting form (opens in new window)

Embedded form:

Chapter 1

Simple Harmonic Motion

Simple harmonic motion (SHM) is a simple and common type of oscillatory motion. It is a model which is widely used in modelling systems due to its simplicity.

In general, an object will move under SHM where its acceleration is:

1. proportional to its displacement, but
2. in the opposite direction.

The force causing this acceleration is often termed a *restoring force* as it acts to push the object back to its starting point.

1.1 A simple example of SHM

Consider a block on a spring (Figure 1.1)

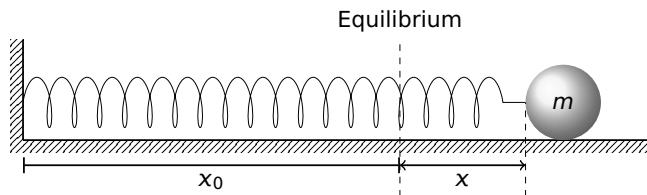


Figure 1.1: A mass on a spring, stretched distance x past its equilibrium length x_0

By Hooke's law, the spring exerts a force on the block proportional to its displacement x , but in the opposite direction, pushing the block back to its equilibrium position, shown mathematically in Equation (1.1):

$$F_x = -kx \quad (1.1)$$

In this example, F_x is considered a **restoring force**, while k is the force constant of the spring.

Applying Newton's Second Law to this problem, we can obtain the mathematical description of the system:

$$\begin{aligned} F_x &= ma_x \\ -kx &= m \frac{d^2x}{dt^2} \end{aligned} \quad (1.2)$$

... and through rearrangement and combination with Equation (1.1) we obtain the description of how this mass will move (Equation (1.3)):

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x \quad (1.3)$$

The general form of this expression for any system can be considered as shown in Equation (1.4)

$$\frac{d^2x}{dt^2} = -Cx \quad \text{or} \quad \ddot{x} = -Cx \quad (1.4)$$

...where C is a positive constant which depends on the system and represents a ratio of the elastic (k) and inertial (m) contributions within the system.

1.1.1 Key terms

- **Period:** The time T for one complete oscillation back and forth (units s)
- **Frequency:** The reciprocal of the period; $f = \frac{1}{T}$, units s^{-1} or Hz.

1.2 Positioning in SHM

SHM can be described by a general equation of motion, defining the position (x) of the oscillating mass using a cosine function (Equation (1.5)):

$$x = A \cos(\omega t + \delta) \quad (1.5)$$

The parameters in this equation are:

- A : The amplitude of the oscillation
- $\omega t + \delta$: Phase of motion
- δ : Phase constant

For any single oscillator, the time origin can always be chosen so that $\delta = 0$. For two or more oscillators there will generally be a phase difference between them (*i.e.* they will not always be at the same ‘zero’ position at time zero - Figure 1.2):

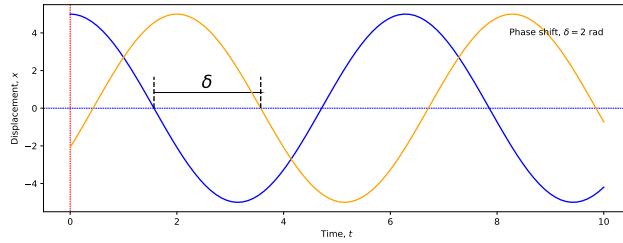


Figure 1.2: The changing position of two oscillators with respect to time, with a relative phase shift of 2 radians

1.3 Velocity in SHM

To find the velocity of the oscillating mass, we can simply find the first derivative of its position with respect to time (Equation (1.6)):

$$\begin{aligned} v &= \frac{dx}{dt} \\ &= -A\omega \sin(\omega t + \delta) \end{aligned} \quad (1.6)$$

A quick inspection of this shows that the velocity v is maximised when x is at a minimum; *i.e.* as the object passes through its equilibrium position.

1.4 Acceleration in SHM

Again, to find the acceleration, we find the second derivative of its position with respect to time (Equation (1.7)):

$$\begin{aligned} a &= \frac{dv}{dt} = \frac{d^2x}{dt^2} \\ &= -A\omega^2 \cos(\omega t + \delta) \end{aligned} \quad (1.7)$$

...or, to use the Newtonian “dot” notation (Equation (1.8)):

$$a = \ddot{x} = -\omega^2 x \quad (1.8)$$

If we now compare this with Equation (1.3) we can see that we have an expression for ω for the oscillating mass m on a spring of force constant k (Equation (1.9)):

$$\begin{aligned}\omega^2 &= \frac{k}{m} \\ \omega &= \sqrt{\frac{k}{m}}\end{aligned}\tag{1.9}$$

1.5 Comparing displacement, velocity and acceleration

When we now compare the displacement, velocity and acceleration we make a number of observations. Firstly, they are all sinusoidal functions; variously sine and cosine functions. However, when we overlay these we have a better indication of how they interrelate (Figure 1.3)

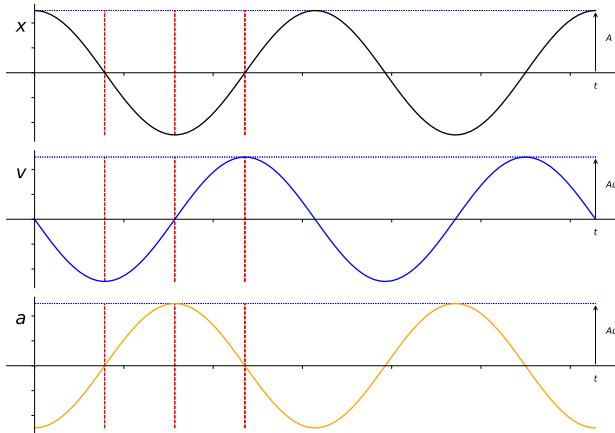


Figure 1.3: Comparing the changes of position, velocity and acceleration with time for a harmonic oscillator. Note that when x is at zero, v is maximised, while a is at a maximum when v is zero. The relative amplitudes of each of the waves is given.

We can now make a few more observations:

- When the displacement x is at a maximum (x_{\max}), the velocity v is zero while the acceleration is at its maximum **but negative with respect to displacement** ($a = -a_{\max}$)
- When the displacement x is zero, the velocity v is at its maximum value ($v = \pm v_{\max}$) and the acceleration is zero.
- The pattern repeats with each period; namely x_0 (displacement at time $t = 0$) is equal to the displacement x_T (displacement after one period of oscillation, T), and the same for the acceleration and velocity.
- In general, $x_t = x_{t+T}$; the displacement at time t is equal to the displacement at the time t plus one period of oscillation, T .

This allows us to compare the displacement, velocity and acceleration at four points in the oscillation (Table 1.1):

Table 1.1: Relating the displacement, velocity and acceleration at different times in the oscillation for a simple harmonic oscillator.

Time	Displacement, x	Velocity, v	Acceleration, a
$t = 0$	$x_0 = A$	$v_0 = 0$	$a_0 = -a_{\max}$
$t = \frac{T}{4}$	$x_{\frac{T}{4}} = 0$	$v_{\frac{T}{4}} = -v_{\max}$	$a_{\frac{T}{4}} = 0$
$t = \frac{T}{2}$	$x_{\frac{T}{2}} = -A$	$v_{\frac{T}{2}} = 0$	$a_{\frac{T}{2}} = a_{\max}$
$t = \frac{3T}{4}$	$x_{\frac{3T}{4}} = 0$	$v_{\frac{3T}{4}} = v_{\max}$	$a_{\frac{3T}{4}} = 0$
$t = T$	$x_T = x_0 = A$	$v_T = v_0 = 0$	$a_T = a_0 = -a_{\max}$

1.6 Initial conditions

We mentioned in Section 1.5 that the displacement, velocity and acceleration expressions were based on sinusoidal functions, and each function had a scaling factor A (the **amplitude** of the oscillation) and a **phase** component δ . In most problems, we wish to determine the value of these constants. In order to determine these, we establish the initial conditions of the oscillation.

In Figure 1.3 we defined our displacement at $+A$ which set up the rest of the problem. However, we will not always be so fortunate. For a general case, we then need to be more discerning.

We can establish expressions for both the amplitude and the phase component by setting $t = 0$ in our general expressions (Equation (1.10)):

$$\begin{aligned} x_t &= A \cos(\omega t + \delta) & \rightarrow x_0 &= A \cos(\delta) \\ v_t &= -A\omega \sin(\omega t + \delta) & \rightarrow v_0 &= -A\omega \sin(\delta) \end{aligned} \quad (1.10)$$

We now treat these as simultaneous equations to find δ and A (Equation (1.11))¹:

$$\tan \delta = \frac{\sin \delta}{\cos \delta} = -\frac{v_0}{\omega x_0} \quad \text{and} \quad A^2 = x_0^2 + \frac{v_0^2}{\omega^2} \quad (1.11)$$

1.7 Frequency and angular frequency

In Section 1.5 we stated that the nature of the oscillation meant that it repeats after every oscillation; mathematically $x(t) = x(t+T)$; the position x at time t is equal to the position at time $(t+T)$.

¹Note that we use the trigonometric identity $\cos^2 \alpha + \sin^2 \alpha = 1$ to find A

When we apply this to the velocity, we obtain the following expression:

$$\begin{aligned} v(t) &= v(t+T) \\ A \cos(\omega t + \delta) &= A \cos(\omega(t+T) + \delta) \\ &= A \cos([\omega t + \delta] + \omega T) \end{aligned} \quad (1.12)$$

Due to the cyclic nature of a cosine function, $\cos(\alpha) = \cos(\alpha + 2\pi)$, this must therefore mean (Equation (1.13)):

$$\omega T = 2\pi \quad \text{or} \quad \omega = \frac{2\pi}{T} \quad (1.13)$$

This gives us a way to think about ω ; its connection to circular motion (the clue is the 2π !). It can be thought of as the **angular frequency**, with units radians s⁻¹, and an oscillation of 2π radians corresponds to one period of oscillation.

Additionally, since the frequency of the oscillation f is the reciprocal of the period of oscillation ($f = \frac{1}{T}$), the angular frequency can be rewritten as $\omega = 2\pi f$, and $f = \frac{\omega}{2\pi}$.

For the spring system we discussed in Section 1.1, we stated that the angular frequency $\omega = \sqrt{\frac{k}{m}}$. Therefore we can obtain an expression for the frequency of our oscillator (Equation (1.14)):

$$f = \frac{1}{T} = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (1.14)$$

Inspection of this equation reveals the behaviour of our oscillator:

- If we have a stiffer spring (larger k), we expect the frequency f to increase,
- If we use an oscillator with larger mass (larger m), we would expect the frequency (f) to decrease.
- The frequency (and therefore period) of simple harmonic oscillation is independent of amplitude.²

1.8 SHM and circular motion

We mentioned an “angular frequency” for SHM; this would appear to suggest behaviour akin to circular motion. It is therefore worth exploring our descriptions of circular motion.

Imagine a point mass moving in a circle (Figure 1.4). For convenience, we imagine this using Cartesian $x - y$ axes, shown in Figure 1.4.

²A caveat to this is for large amplitudes where other factors start to affect the behaviour. But this is then no longer *simple* harmonic motion!

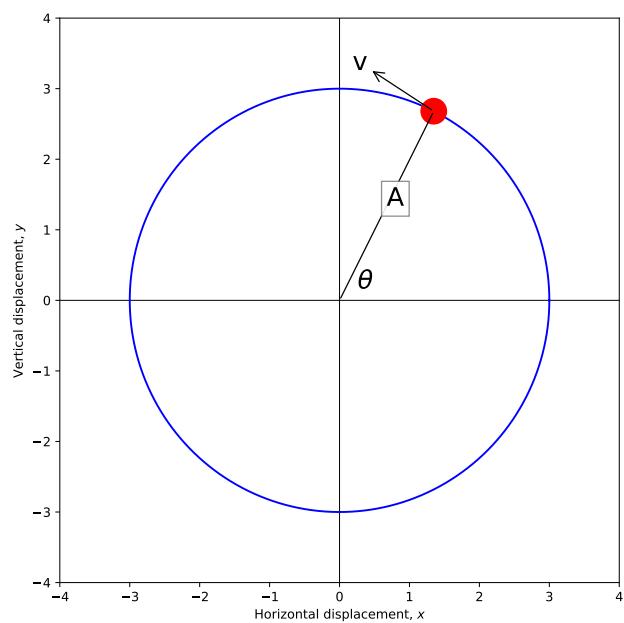


Figure 1.4: A particle moving in a circle of radius A can be assumed to have an instantaneous linear velocity v . The x and y components of the motion are found from trigonometry of the radius A and the angle θ .

The particle of mass m is moving in a circle of radius A with instantaneous linear velocity v ; the radius makes an angle θ with the x -axis. We now look at how its position maps onto each of the axes:

- The angular velocity of the particle is ω ; found *via* $\frac{v}{A}$
 - We can then describe θ in terms of ω :
 - $\theta = \omega t + \delta$ (δ is the angle at time $t = 0$)
- The particle's position on the x -axis is therefore found *via*:
 - $x = A \cos \theta = A \cos(\omega t + \delta)$
 - This corresponds with the expression for SHM for a particle moving in a linear fashion (Equation (1.5)).

We can also consider how its position maps onto the y -axis:

- The position on the y -axis is found *via*:
 - $y = A \sin \theta = A \sin(\omega t + \delta) \equiv A \cos(\omega t + [\delta - \frac{\pi}{2}])$
 - This once again corresponds with the expression for SHM for a particle moving in a linear fashion.
 - The y -component of the motion is $\frac{\pi}{2}$ out of phase with the x -component

This illustrates that circular motion is a combination of two perpendicular SHM oscillations of the same frequency and amplitude, but a relative phase of $\frac{\pi}{2}$.

1.9 Energy in SHM

As with all isolated systems, the total energy E of the simple harmonic oscillator is constant, however the contributions from potential energy (U) and KE vary with time.

$$E = KE + U = \text{constant} \quad (1.15)$$

Let's go back to the condition for SHM; there is a restoring force proportional to the displacement:

$$F = -kx \quad (1.16)$$

Knowing that the potential energy is the first derivative of the force, we can integrate this force expression (with respect to x) to get back to the energy statement:³

$$\begin{aligned} U &= \int F dx \\ &= \frac{1}{2} kx^2 \quad [+C] \end{aligned} \quad (1.17)$$

³The constant of integration will evaluate to zero from the starting condition $U = 0$ at zero displacement.

However, we already have an expression for how the displacement, x , varies with time (Equation (1.5)); let's now substitute this into the result from Equation (1.17):

$$U = \frac{1}{2}kA^2 \cos^2(\omega t + \delta) \quad (1.18)$$

We can also generate an expression for the kinetic energy; remember that kinetic energy can be found from $\frac{1}{2}mv^2$; so we use the expression for v given in Equation (1.6):

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}mA^2\omega^2 \sin^2(\omega t + \delta) \end{aligned} \quad (1.19)$$

We can simplify this using Equation (1.7) for a particle on a spring, where $\omega^2 = \frac{k}{m}$:

$$KE = \frac{1}{2}kA^2 \sin^2(\omega t + \delta) \quad (1.20)$$

Combining the result of Equations (1.18) and (1.20) we find the result in Equation (1.21):

$$\begin{aligned} E_{\text{total}} &= U + KE \\ &= \frac{1}{2}kA^2 \cos^2(\omega t + \delta) + \frac{1}{2}kA^2 \sin^2(\omega t + \delta) \\ &= \frac{1}{2}kA^2 [\cos^2(\omega t + \delta) + \sin^2(\omega t + \delta)] \\ &= \frac{1}{2}kA^2 \end{aligned} \quad (1.21)$$

This result tells us that the total energy in a simple harmonic oscillation is **proportional to the square of the amplitude**.

1.9.1 Points to bear in mind

- $U = U_{\text{max}}$ at $x = \pm x_{\text{max}}$
- $KE = KE_{\text{max}}$ at $x = 0$
- $U_{\text{average}} = KE_{\text{average}} = \frac{1}{2}E_{\text{total}}$

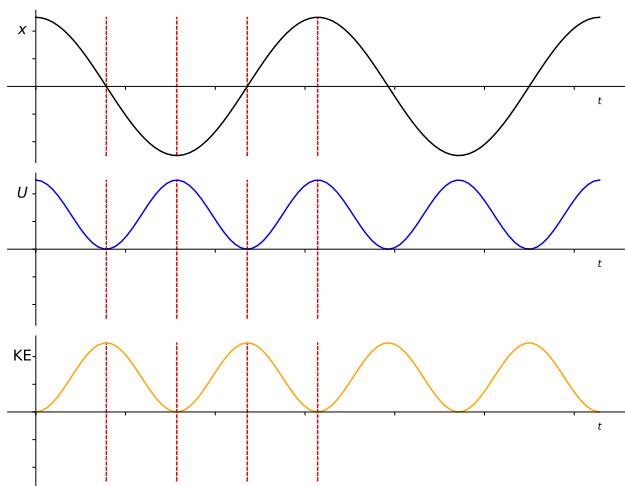


Figure 1.5: Variation of the kinetic energy (KE) and potential energy (U) of an harmonic oscillator with displacement x about the equilibrium position.

Chapter 2

SHM in Real Systems

Textbook link: Tipler and Mosca: Ch 14.2 to 14.4

2.1 General motion near equilibrium

A way of thinking about SHM is that it is a point mass oscillating within a potential energy field. As with any potential energy field, the force on this particle is given by the gradient of the potential energy and is directed down the potential energy slope. Mathematically, for a potential energy field, the force may be found as follows (Equation (2.1)):

$$F = -\frac{dU}{dr} \quad (2.1)$$

In a one-dimensional system, this is expressed as follows (Equation (2.2)):

$$F_x = -\frac{dU}{dx} \quad (2.2)$$

As mentioned in Section 1.1, under SHM the force is proportional to the displacement from the equilibrium position and in the opposite direction; *i.e.*:

$$F_x = -kx \quad (2.3)$$

Applying the principle from Equation (2.1) we can therefore integrate this expression with respect to x to obtain the expression for our potential energy. We covered this in Section 1.9, and we found the result (Equation (2.4); remember that, due to initial conditions, the constant of integration reduces to zero).

$$U = \frac{1}{2}kx^2 \quad (2.4)$$

Simple inspection and recall of our mathematics knowledge tells us that this simple equation represents a parabola. This gives us a series of useful points to bear in mind:

- A parabolic potential energy function implies SHM and *vice versa*;
- For small amplitudes of oscillation, many potential energy functions may be *approximated* by a parabola (*e.g.* a pendulum, vibrating molecules)
- A system undergoing SHM is called a **harmonic oscillator**.

The simplicity of the **simple harmonic oscillator** model is what makes it such a generally useful system to consider.

2.2 Example: a diatomic molecule

A diatomic molecule is a useful system to consider as an example because it can be approximated to a harmonic oscillator at small displacements about the equilibrium. The potential energy curve for a vibrating diatomic molecule (in this case the H₂ hydrogen molecule) is shown in Figure 2.1.

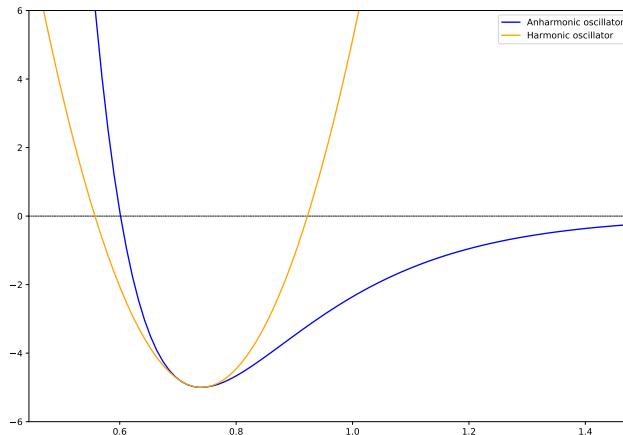


Figure 2.1: Comparing the Morse potential of the anharmonic oscillator with the harmonic approximation. Near equilibrium, the harmonic oscillator model approximates diatomic behaviour, however this rapidly deviates from reality.

The potential of a vibrating diatomic is known as the Morse potential; the form of this is outwith this discussion, however it is useful to think that, for small displacements around the equilibrium separation the potential energy curve approximates a parabola. We can therefore re-draw our potential energy curve as

such, and show this in Figure 2.2.¹

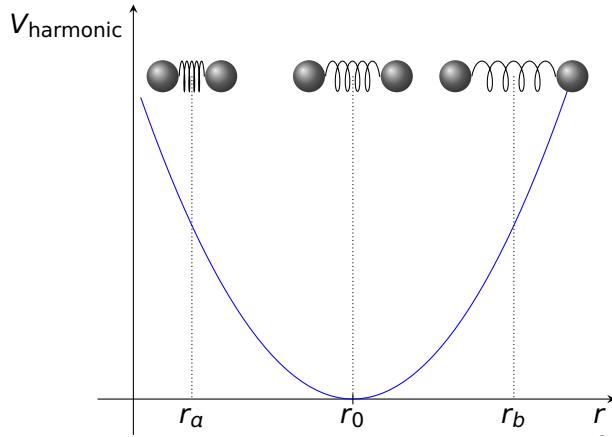


Figure 2.2: The parabolic approximation of a diatomic molecule, showing the potential varying with compression or extension from equilibrium separation, r_0 .

The equation of the parabola shown in 2.2 takes the following form:

$$U(r) = A + B(r - r_0)^2 \quad (2.5)$$

...where A and B are constants relating to the molecular system under consideration, and r_0 is the equilibrium bond length.

The force on the bond can then be found from the first derivative of the bond potential described in Equation (2.5):

$$\begin{aligned} F_r &= -\frac{dU}{dr} \\ &= -2B(r - r_0) \end{aligned} \quad (2.6)$$

Since the term $(r - r_0)$ is the displacement from the equilibrium position, we see that the force, F_r is a **restoring force** and is proportional to the displacement (and in the opposite direction!), telling us that the motion is SHM. In this example however, the parabolic approximation fails at larger amplitudes.

¹Simplistically, the deviation is due to nuclear repulsion at high compression, while at large extension the bond eventually breaks - the “zero potential” point.

2.3 Example: Mass on a vertical spring

In Section 1.1 we considered a mass on a horizontal spring; there was only a single force acting on the mass (the force from the spring), however we are now considering a vertical spring and must consider the effects of gravity (Figure 2.3).

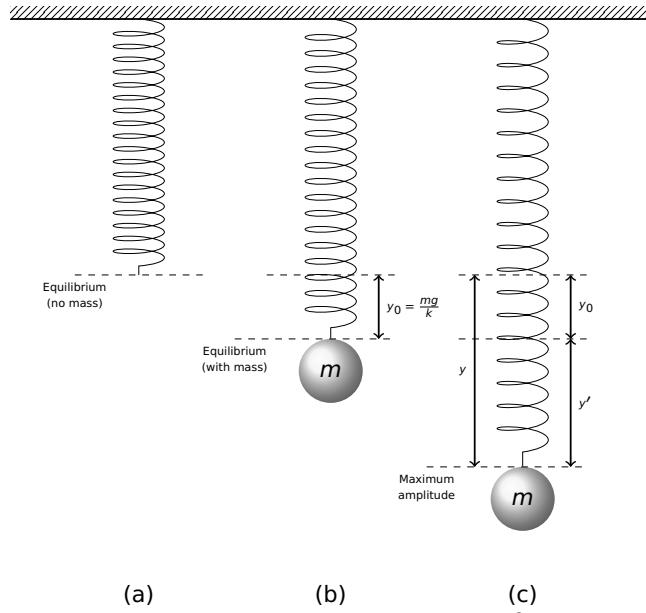


Figure 2.3: A particle oscillating on a vertical spring. There are two equilibrium positions corresponding to the equilibrium position of the unladen spring (a) and the equilibrium position of the mass loaded on spring (b), where the weight of the load is balanced by the force from the extended spring.

In this case we need to work through a slightly different process to find the equation of motion, chiefly because the equilibrium displacement of the mass (y_0) is different from the equilibrium position of the spring.

The equilibrium position of the mass y_0 is lower than the equilibrium extension of the spring (gravity on the mass causes the spring to stretch). This is found by relating the force from the spring and the force of gravity acting on the mass (Equation (2.7)):

$$\begin{aligned} \text{Force due to spring extension} &= \text{Gravity acting on mass} \\ ky_0 &= mg \\ y_0 &= \frac{mg}{k} \end{aligned} \tag{2.7}$$

We now apply Newton's Second Law to obtain an expression for the acceleration

on the mass due to the forces acting on it (the spring force and gravity).

$$\begin{aligned} m\ddot{y} &= \text{spring force + gravity} \\ m\frac{d^2y}{dt^2} &= -ky + mg \end{aligned} \quad (2.8)$$

In this expression, y is the total extension of the spring (the extension to the mass equilibrium point, y_0 plus the displacement from this point, y').

We now substitute a variable; since $y = y_0 + y'$, we substitute y for y' :

- $y = y_0 + y'$
- $y' = y - y_0$

Since y_0 is a constant:

$$\frac{dy'}{dt} = \frac{dy}{dt} \quad \text{and} \quad \frac{d^2y'}{dt^2} = \frac{d^2y}{dt^2} \quad (2.9)$$

Replacing y for $(y_0 + y')$ in Equation (2.8), we obtain Equation (2.10):

$$m\frac{d^2y'}{dt^2} = -k(y_0 + y') + mg \quad (2.10)$$

Since $ky_0 = mg$ (Equation (2.7)), we can therefore eliminate these terms from Equation (2.10), and rewrite as Equation (2.11):

$$m\frac{d^2y'}{dt^2} = -ky' \quad (2.11)$$

This means that, in reference to Figure 2.3, we still have SHM centered on the equilibrium position of the mass. This may seem like a self-evident result, however it is useful to recognise the role of gravity; its effect is to shift the equilibrium position of the oscillation from $y = 0$ (the equilibrium position of the spring) to $y = y_0$ ($y' = 0$).

Let's now consider the energy in this system. The system already contains some elastic energy as the spring is already stretched to y_0 by the gravity acting on the mass:

$$\text{elastic potential energy} = \frac{1}{2}ky^2 - \frac{1}{2}ky_0^2 \quad (2.12)$$

The gravitational potential energy (relative to the starting position y_0) is given by:

$$\text{gravitational potential energy} = mg(y - y_0) \quad (2.13)$$

The total potential energy is therefore given by Equation (2.14):

$$U = \frac{1}{2}ky^2 - \frac{1}{2}ky_0^2 - mg(y - y_0) \quad (2.14)$$

We can then show that the total potential energy expression in Equation (2.14) can be simplified to that shown in Equation (2.15):

$$U = \frac{1}{2}ky'^2 \quad (2.15)$$

You should ensure you understand how this simplification is done; this is left as an exercise.

Overall, the expression for the total potential energy shown in Equation (2.15) will still yield a parabola and as such the oscillation is still a simple harmonic oscillation as before.

2.4 The Simple Pendulum

A pendulum is a mechanically simple oscillating system; its oscillation can be considered SHM for small displacements. A simple pendulum is shown in Figure 2.4, and it is worth quickly revisiting this example as we will extend this in our further examples.

For small displacements (small ϕ), the equation of motion can be considered as in Equation (2.16):

$$\frac{d^2\phi}{dt^2} = -\frac{g}{L} \sin \theta \approx -\frac{g\phi}{L} \text{ for small } \phi \quad (2.16)$$

This is SHM, with angular frequency ω and period T found as in Equation (2.17):

$$\omega^2 = \frac{g}{L} \quad \text{and} \quad T = 2\pi\sqrt{\frac{L}{g}} \quad (2.17)$$

The solution for the equation of motion of this system then becomes (Equation (2.18)):

$$\phi = \phi_0 \cos(\omega t + \delta) \quad (2.18)$$

...where ϕ_0 is the amplitude of the system and δ is the phase constant. Notice that, for the pendulum we express the amplitude in terms of the angle of the string rather than an absolute distance (shown as s in Figure 2.4).

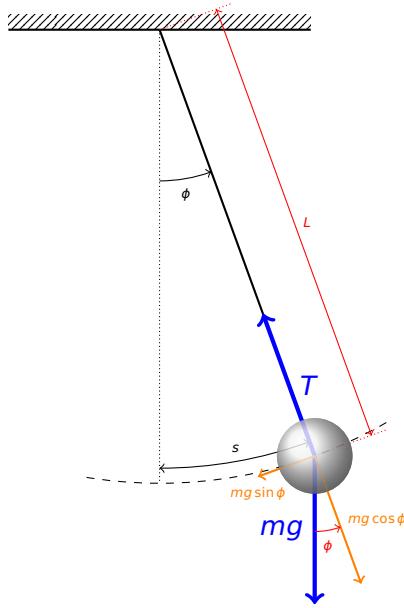


Figure 2.4: A simple pendulum, consisting of a mass m swinging on a string of length L . At angle ϕ , the weight of the mass mg can be resolved into components to determine the mechanics of the system.

2.5 The Physical Pendulum

Many oscillating systems demonstrate rotational oscillations under gravity akin to the simple pendulum. In this case, the system rotates around a pivot, P , and this can then be considered as a pendulum with the centre of mass acting as the ‘bob’. An example of a general system is shown in Figure 2.5.

Now, as a restoring force, we consider the **torque** of the centre of mass around the pivot. Remember that the torque is defined as the product of the force and the **perpendicular distance of the force’s line of action from the pivot**. If using vectors, this is considered as the **cross product** of the force vector \mathbf{F} with the position vector of the centre of mass from the pivot, \mathbf{D} (Equation (2.19)):

$$\begin{aligned} \text{Torque about pivot } P &= \mathbf{F} \times \mathbf{D} \\ &= \hat{\mathbf{n}} |F||D| \sin \phi \\ &= \hat{\mathbf{n}} mgD \sin \phi = \tau \hat{\mathbf{n}} \end{aligned} \quad (2.19)$$

In this case, the unit vector $\hat{\mathbf{n}}$ is perpendicular to the plane of rotation and is included for completeness’ sake. What we are interested in is the magnitude of this torque vector, τ .

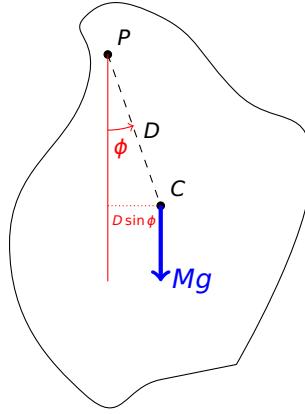


Figure 2.5: A physical pendulum, consisting of a physical object of mass M swinging on a pivot through point P of length L . At angle ϕ , the weight of the mass Mg can again be resolved into components to determine the mechanics of the system.

Recall from rotational motion that the angular acceleration α and the torque τ are connected via the moment of inertia I (Equation (2.20)):²

$$\begin{aligned} I \frac{d^2\phi}{dt^2} &= I\alpha = \tau \\ &= -mgD \sin \phi \end{aligned} \quad (2.20)$$

We now approximate this for small ϕ :

$$\frac{d^2\phi}{dt^2} = -\frac{mgD\phi}{I} = -\omega^2\phi \quad (2.21)$$

This allows us to identify expressions again for the angular frequency, ω , and the period T (Equation (2.22)):

$$\omega^2 = \frac{mgD}{I} \quad \text{and} \quad T = 2\pi \sqrt{\frac{I}{mgD}} \quad (2.22)$$

We can compare this result with that for the simple pendulum shown in Equation (2.17); if we remember that the moment of inertia I is defined as $I = mD^2$, we can substitute this into Equation (2.22) and see that this is a general case for any rotating body:

²Note that the appearance of the negative sign indicates that the torque force is opposite to the direction of increasing ϕ

$$T_{\text{simple pendulum}} = 2\pi \sqrt{\frac{I_{\text{simple pendulum}}}{mgD}} = 2\pi \sqrt{\frac{mD^2}{mgD}} = 2\pi \sqrt{\frac{D}{g}} \quad (2.23)$$

...giving us our expected result (where the general term D can be replaced for the length of the simple pendulum, L).

Chapter 3

Complex Numbers

Textbook link: Jordan and Smith, Ch. 6

As we move on with our descriptions of oscillations and, in particular, wave behaviour, the mathematics can become increasingly involved. We can greatly simplify the mathematics by using **complex numbers** in our derivations. For now, it is useful to quickly revisit the complex number concept.

While the idea of a “complex number” sounds ... *complex*, the use of these numbers becomes straightforward as we apply our familiar mathematical techniques. In the context of mathematics, the term **complex** simply means ‘more than one part’; therefore, a **complex number** is a number with more than one part. It is this two-component nature of a complex number which makes them so useful in many aspects of Physics, and particularly when describing wave behaviour.

3.1 Overview of complex numbers

The general form of a complex number z is shown in Equation (3.1):

$$z = a + ib \quad (3.1)$$

The symbol z is a general term for a complex number, and has two components, a “real” component a and an “imaginary” component, b . The imaginary number, i , is defined using the process shown in Equation (3.2):

$$\begin{aligned} x^2 &= -1 \\ x &= \pm i \\ i^2 &= -1 \end{aligned} \quad (3.2)$$

The terms ‘real’ and ‘imaginary’ are nothing more than labels. Neither is any more or less “realistic” than the other nor is it any less valid. Some may claim that the number i is a ‘pretend’ number; however were this to be true, it would not be as useful as it is!¹.

The next useful concept to recall is the **complex conjugate**, z^* . This is defined as in Equation (3.3):

$$\begin{aligned} z &= a + ib \\ z^* &= a - ib \\ zz^* &= a^2 + b^2 \end{aligned} \tag{3.3}$$

In general, for any complex number of the form $z = a \pm ib$, there exists its complex conjugate, $z^* = a \mp ib$ such that zz^* is a wholly real number and equal to $a^2 + b^2$.

The complex conjugate is particularly useful when finding fractions of complex numbers as it is used to make the denominator of the fraction wholly “real”.

3.2 Useful results for complex numbers

There are a number of useful results worth remembering when handling complex numbers, and these are listed here.

For a pair of complex numbers, z_1 and z_2 :

$$\begin{aligned} z_1 &= a_1 + ib_1 \\ z_2 &= a_2 + ib_2 \end{aligned} \tag{3.4}$$

...we can establish the following principles:

- Equality:

$$\text{If } a_1 = a_2 \quad \text{and} \quad b_1 = b_2 \quad \text{then} \quad z_1 = z_2 \tag{3.5}$$

- Addition and subtraction:

$$\begin{aligned} z_1 + z_2 &= (a_1 + a_2) + i(b_1 + b_2) \\ z_1 - z_2 &= (a_1 - a_2) + i(b_1 - b_2) \end{aligned} \tag{3.6}$$

- Products:

$$\begin{aligned} z_1 \times z_2 &= (a_1 + ib_1)(a_2 + ib_2) \\ &= (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + a_2 b_1) \end{aligned} \tag{3.7}$$

¹Remember that negative numbers were once seen as ‘pretend numbers’, as you could not have negative eight apples. They have since become indispensable in many applications, not least financial transactions!

- Reciprocal:

$$\frac{1}{z} = \frac{z^*}{zz^*} = \frac{a - ib}{a^2 + b^2} \quad (3.8)$$

- Division:

$$\frac{z_1}{z_2} = \frac{z_1 z_2^*}{z_2 z_2^*} = \frac{(a_1 a_2 + b_1 b_2) - i(a_1 b_2 - a_2 b_1)}{a_2^2 + b_2^2} \quad (3.9)$$

Other useful principles which involve the complex conjugate are listed in Equation (3.10) below:

$$\begin{aligned} (z_1 + z_2)^* &= z_1^* + z_2^* \\ (z_1 z_2)^* &= z_1^* z_2^* \\ \left(\frac{z_1}{z_2}\right)^* &= \frac{z_1^*}{z_2^*} \\ a = \operatorname{Re}(z) &= \frac{1}{2}(z + z^*) \\ b = \operatorname{Im}(z) &= \frac{i}{2}(z - z^*) \end{aligned} \quad (3.10)$$

3.3 The Argand Diagram

Since a complex number consists of two independent components, we have another way to describe these numbers. Complex numbers can be plotted on a graph, with the ‘real’ component plotted on one axis (the x -axis) and the ‘imaginary’ component plotted on the other axis (the y -axis). This is the basis of the Argand diagram (Figure 3.1).

This allows us to define a complex number in terms of a **modulus** (radial distance from the origin) and an **argument** (angle from the ‘real’ axis). Useful properties of the modulus are listed in Equation (3.11):

$$\begin{aligned} |z^*| &= |z| \\ zz^* &= |z|^2 \\ |z_1 z_2| &= |z_1| |z_2| \\ \left|\frac{z_1}{z_2}\right| &= \frac{|z_1|}{|z_2|} \end{aligned} \quad (3.11)$$

$$|z_1 + z_2| \neq |z_1| + |z_2|$$

The Argand diagram is a representation of the **complex plane**, through which it becomes possible to visualise properties of complex numbers. One example of this is the addition of complex numbers; these can be considered to behave as vectors (Figure 3.2)

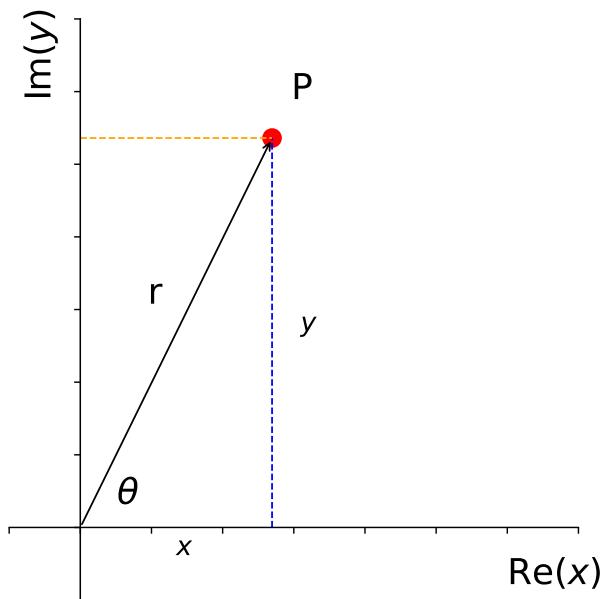


Figure 3.1: A typical Argand diagram, showing the Real ('Re') axis and the Imaginary ('Im') axis. The point P can be defined in ' x, y ' terms (the 'complex number'), or can be defined as polar ' r, θ ' terms (termed 'modulus' and 'argument')

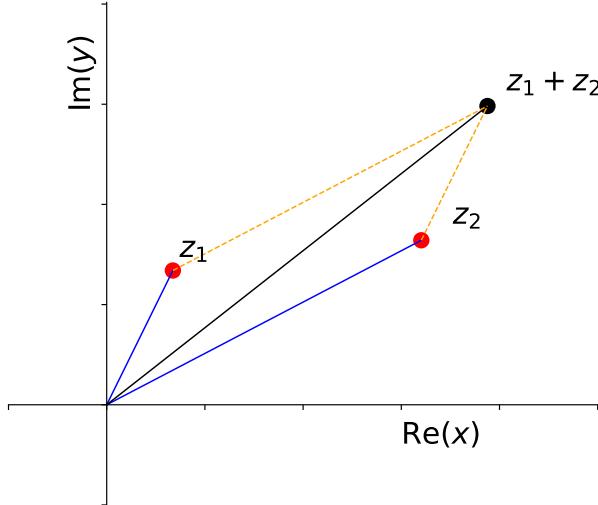


Figure 3.2: Addition of complex numbers z_1 and z_2 can be shown graphically on an Argand diagram; the separate consideration of the ‘real’ and ‘imaginary’ components is analogous to the separate consideration of vector components.

3.4 Polar representation of complex numbers

As well as the **Cartesian interpretation** of the Argand diagram, we can also consider a **polar representation** of a complex number; where instead of “real” and “imaginary” components acting as (x, y) coordinates, we define the position of the complex number on the complex plane as a radius and an angle, θ . We have already illustrated this in Figure 3.1

In this representation, the complex number can be expressed a different way:

$$\begin{aligned}
 z &= a + ib \\
 r &= |z| = \sqrt{a^2 + b^2} \\
 a &= r \cos \theta & b &= r \sin \theta & \theta &= \arg(z) = \arctan\left(\frac{b}{a}\right) \\
 z &= r \cos \theta + i|z| \sin \theta
 \end{aligned} \tag{3.12}$$

Normally, θ will lie in the range such that $-\pi < \theta \leq \pi$, meaning that our complex number representation is now shown in Equation (3.13):

$$\begin{aligned}
 z &= a + ib = r \cos \theta + ir \sin \theta \\
 z &= r (\cos \theta + i \sin \theta)
 \end{aligned} \tag{3.13}$$

3.5 Exponential representation of complex numbers

The exponential representation of a complex number takes the general form of $z = Ae^{i\theta}$. This is based on series expansions of $\cos \theta$ and $i \sin \theta$, which shows *De Moivre's theorem*. Key results from this are shown in Equation (3.14):

$$\begin{aligned} e^{i\theta} &= \cos \theta + i \sin \theta \\ (e^{i\theta})^n &= (\cos \theta + i \sin \theta)^n = e^{in\theta} \\ (\cos \theta + i \sin \theta)^n &= \cos n\theta + i \sin n\theta \end{aligned} \quad (3.14)$$

This means that we obtain the following representations for complex numbers:

$$\begin{aligned} z &= r(\cos \theta + i \sin \theta) = re^{i\theta} \\ z^* &= r(\cos \theta - i \sin \theta) = re^{-i\theta} \\ \text{where: } r &= |z| \quad \theta = \arg(z) \end{aligned} \quad (3.15)$$

Combining these with Equation (3.10) we also note the following useful results (Equation (3.16))

$$\begin{aligned} \cos \theta &= \frac{1}{2}(e^{i\theta} + e^{-i\theta}) \\ \sin \theta &= \frac{1}{2i}(e^{i\theta} - e^{-i\theta}) \end{aligned} \quad (3.16)$$

3.6 Complex representation of oscillations

Having quickly readdressed our understanding of complex numbers, we now turn our attention to the application of these in the context of oscillations and waves.

Consider the general equation of SHM (Equation (3.17), derived from Equation (1.8))

$$\frac{d^2u}{dt^2} + \omega^2 u = 0 \quad (3.17)$$

As has been previously discussed, sinusoidal functions can form the basis of solutions to this differential equation; so both $\cos \omega t$ and $\sin \omega t$ are solutions to this equation. Therefore, any linear combination of these solutions will also be a solution, *i.e.* the linear combination shown here:

$$u = c_1 \cos \omega t + c_2 \sin \omega t \quad (3.18)$$

... will also satisfy Equation (3.17). This can be extended using De Moivre's theorem (Equation (3.14)) allowing an exponential representation of an oscillation as shown in Equation (3.19):

$$u = A(\cos \omega t + i \sin \omega t) \equiv Ae^{i\omega t} \quad (3.19)$$

Therefore the solution $u = Ae^{i\omega t}$ represents an oscillation with amplitude A and frequency ω

3.7 Take-home points

- We can **always** represent an oscillation using a complex exponential function
- To obtain the actual physical displacement of the system we simply examine either the real or the imaginary part of the solution:

$$\begin{aligned} \text{Either: } \text{displacement} &= \operatorname{Re}(u) = A \cos \omega t \\ \text{or: } \text{displacement} &= \operatorname{Im}(u) = A \sin \omega t \end{aligned} \quad (3.20)$$

The main advantage of working with complex exponentials is that they are considerably easier to manipulate than the trigonometric functions sine and cosine. In general it is far easier to use this exponential notation when multiplying oscillations (such as you will explore in electrical circuits later). However, when adding oscillations or waves you may find it easier using a trigonometric identity.

You should be comfortable using either approach to represent an oscillation.

Chapter 4

Damped oscillations

Textbook link

The oscillations we have looked at so far make the assumption that the oscillation will continue indefinitely and that no energy is gained or lost by the system. Such perpetually oscillating systems are extremely unusual, and almost every oscillation you encounter in the real world will lose energy to its surroundings, either requiring constant input of energy to maintain the oscillation or it will eventually dissipate all of its energy and the oscillation will stop. The energy is dissipated through a process known as ‘damping’.

The change in amplitude of a damped oscillation is illustrated in Figure 4.1. Both amplitude and energy decrease by a **constant percentage** in each cycle; this is an **exponential decrease**.

4.1 The general case of damping

For a simply damped system, the damping force is proportional to the velocity of the oscillating mass and **opposes** the direction of motion (Equation (4.1)):

$$F = -bv \equiv -b \frac{dx}{dt} \quad (4.1)$$

This expression for the damping force can then be included with the equation for the force from the spring (Equation (1.1)) to consider the overall acceleration of a mass undergoing damped harmonic motion. (Equation (4.2)).

$$\begin{aligned} F = -kx - bv &= -kx - b \frac{dx}{dt} = ma \\ m \frac{d^2x}{dt^2} &= -kx - b \frac{dx}{dt} \end{aligned} \quad (4.2)$$

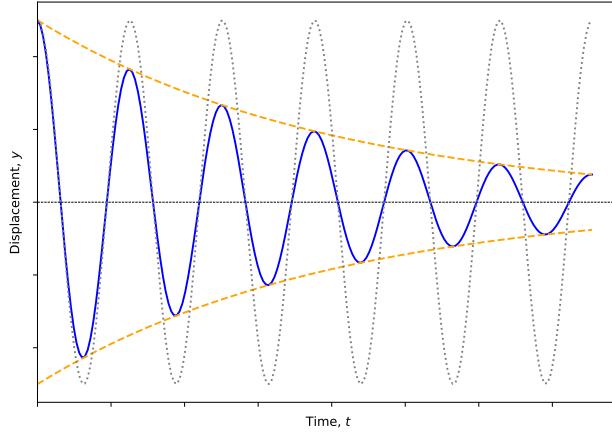


Figure 4.1: The amplitude of a damped oscillation decays exponentially with time. The observed position of the oscillation is shown in blue, while the maximum possible amplitude (related to energy stored in the system) is illustrated by the orange dotted line. The undamped oscillation is shown in grey for comparison.

This can be written as a differential equation for **damped SHM** as shown in Equation (4.3):

$$m \frac{d^2x}{dt^2} + b \frac{dx}{dt} + kx = 0 \quad (4.3)$$

Let's now try to find a solution for this differential equation. Firstly, let's make a couple of assumptions:

1. The amplitude of the damped oscillation is subject to an exponential decay over a timescale of 2τ (don't worry about the factor of 2 for now), and
2. The damped oscillation has a frequency ω' which may be different from the natural frequency of the undamped oscillator, ω_0 .

Our exponential decay factor then becomes $e^{-\frac{t}{2\tau}}$, and the exponential form of the wave equation becomes $e^{i(\omega' t + \delta)}$ (combination of Equations (1.5) and (3.19)). When these are combined we obtain a form of the solution shown in Equation (4.4)

$$x = A_0 e^{-\frac{t}{2\tau}} e^{i(\omega' t + \delta)} \quad (4.4)$$

We can now find the first and second derivatives (make sure you are able to do this; you will need the product rule for differentiation) of this expression to substitute into the differential equation (Equation (4.3)):

$$\begin{aligned}\frac{dx}{dt} &= \left(-\frac{1}{2\tau} + i\omega'\right)x \\ \frac{d^2x}{dt^2} &= \left(-\frac{1}{2\tau} + i\omega'\right)^2 x\end{aligned}\tag{4.5}$$

We can now take our derivatives shown in Equation (4.5) and substitute into the differential equation (Equation (4.3)):

$$\begin{aligned}m \frac{d^2x}{dt^2} + b \frac{dx}{dt} + kx &= 0 \\ m \left(-\frac{1}{2\tau} + i\omega'\right)^2 x + b \left(-\frac{1}{2\tau} + i\omega'\right) x + kx &= 0\end{aligned}\tag{4.6}$$

Multiplying this expression out, we obtain:

$$m \left(\frac{1}{4\tau^2} - \frac{i\omega'}{\tau} - \omega'^2 \right) x + b \left(-\frac{1}{2\tau} + i\omega' \right) x + kx = 0\tag{4.7}$$

...and we can now combine the **Real** and **Imaginary** components:

- **Imaginary:**

$$-\frac{\omega' m}{\tau} + \omega' b = 0\tag{4.8}$$

$$\tau = \frac{m}{b}\tag{4.9}$$

- **Real:**

$$\left(\frac{1}{4\tau^2} - \omega'^2 \right) m - \frac{b}{2\tau} + k = 0\tag{4.10}$$

Therefore (using result from Equation (4.9)):

$$\omega'^2 = \frac{k}{m} - \left(\frac{b}{2m} \right)^2\tag{4.11}$$

However, we already know that $\frac{k}{m} = \omega_0^2$ (equation (1.9)), so:

$$\omega'^2 = \omega_0^2 - \left(\frac{b}{2m} \right)^2\tag{4.12}$$

Having found this result, we can now say that a general solution to damped SHM is as shown in Equation (4.4), or in trigonometric notation as shown in Equation (4.13):

$$x = A_0 e^{-\frac{t}{2\tau}} \cos(\omega' t + \delta) \quad (4.13)$$

...where

- A_0 = initial (maximum) amplitude
- $\tau = \frac{m}{b}$ is the characteristic decay time or time constant
- $\omega' = \sqrt{\omega_0^2 - \left(\frac{b}{2m}\right)^2}$, where ω_0 is the frequency of the undamped oscillator
($\omega_0^2 = \frac{k}{m}$ for a mass on a spring)

Having derived and defined a general expression for damped oscillations, we will now turn to look at different modes of damping.

4.2 Light Damping

In a system which is lightly damped, we can make a number of assumptions:

1. That the frequency of the damped oscillator (ω') is approximately equal to that of the undamped oscillator; $\omega' \approx \omega_0$
2. That the damping factor, $\frac{b}{2m\omega_0}$ is significantly less than one; i.e. $\frac{b}{2m\omega_0} \ll 1$

When we apply these assumptions to Equation (4.4) we obtain the “standard” SHM oscillation (tending towards $e^{i(\omega' t + \delta)}$ as ω' tends to ω_0) with an exponential decay on its amplitude (Equation (4.14)):

$$A = A_0 e^{-\frac{t}{2\tau}} \quad (4.14)$$

The effect of this light damping is that if b (the damping coefficient on the velocity, Equation (4.1)) increases, the damped frequency ω' will decrease, and the characteristic decay time, τ will also decrease.

4.3 Critical damping

In a critically damped system, the **system does not oscillate**; rather, it returns to equilibrium in the shortest possible time. This can be imagined as the damping required to **exactly stop** the vibration and no more. The damped oscillation frequency ω' then, by definition, is equal to zero ($\omega' = 0$) and, by placing this condition into Equation (4.15), we obtain the result for the damping coefficient, b (Equation (4.16))

$$\omega' = \sqrt{\omega_0^2 - \left(\frac{b}{2m}\right)^2} \quad (4.15)$$

$$b = b_c = 2m\omega_0 \quad (4.16)$$

4.4 Overdamping

An overdamped system is one which has so much damping applied that the system returns to equilibrium *even more slowly* than in the critically damped case (Section 4.3). This could be imagined as a mass on a spring which is allowed to return to equilibrium within an extremely viscous medium (honey, or treacle!) and takes a considerable time to slowly return to equilibrium. A comparison of the displacement/time curve between a critically damped system and an overdamped system is shown in Figure 4.2.

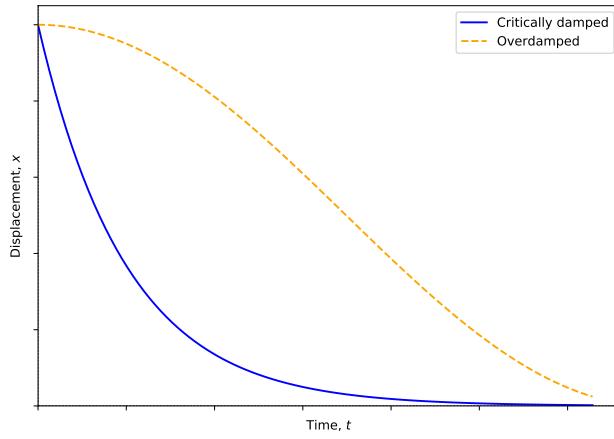


Figure 4.2: A critically damped system returns to its equilibrium position in the shortest possible time, while an overdamped system works against an overbearing damping force slowly returning it to its equilibrium position.

In the overdamped case, the damping coefficient, b is greater than the critical damping coefficient, b_c :

$$b > b_c \quad (4.17)$$

4.5 Quality factor and energy in damped SHM

Recall that we determined energies within SHM in Equations (1.20) and (1.21); remembering that $\omega^2 = \frac{k}{m}$, we can write the overall energy of the system as in Equation (4.18):

$$E = \frac{1}{2}m\omega^2 A^2 \quad (4.18)$$

Within the damped regimes we have established the equations for damping, in particular how to calculate the damped amplitude at a given time, A (Equation

(4.14)). By combining this with Equation (4.18) we can obtain an expression for the damped energy, E , as a proportion of the initial energy of the system, E_0 (Equation (4.19)):

$$\begin{aligned} E &= \frac{1}{2}m\omega^2 \left(A_0 e^{-\frac{t}{2\tau}} \right)^2 \\ &= E_0 e^{-\frac{t}{\tau}} \end{aligned} \quad (4.19)$$

The decay time τ can now be considered the time taken for the energy to decrease to $\frac{1}{e}$ of its original value (see? There was a reason we considered the timescale 2τ in Section 4.1).

A useful measure of the **persistence** of an oscillation is the **quality factor**, Q . This is defined as shown in Equation (4.20):

$$Q = \omega_0 \tau = \frac{\omega_0 m}{b} \quad (4.20)$$

Generally:

- Large Q represents lighter damping, persistent oscillation. (think **high quality oscillation**)
- Small Q represents heavier damping, oscillation stops rapidly. (think **low quality oscillation**)

Q can relate to the energy loss per cycle of oscillation; firstly define the rate of change of energy (Equation (4.21), from Equation (4.18)):

$$\frac{dE}{dt} = -\frac{E_0}{\tau} e^{-\frac{t}{\tau}} = -\frac{E}{\tau} \quad (4.21)$$

If we are considering finite changes, we can adapt our calculus to allow $\Delta E \approx dE$ and $\Delta t \approx dt = T$. Applying this to Equation (4.21) and rearranging gives:

$$\frac{|\Delta E|}{E} = \frac{T}{\tau} = \frac{2\pi}{\omega_0 \tau} = \frac{2\pi}{Q} \quad (4.22)$$

This gives us the result for Q shown in Equation (4.23):

$$Q = \frac{2\pi}{\left(\frac{|\Delta E|}{E}\right)_{\text{cycle}}} \quad (4.23)$$

This gives us one more way to consider the quality factor: it is inversely proportional to the fractional energy loss per cycle of the oscillation. A low Q therefore corresponds to a higher fractional energy loss per cycle than a high Q factor.

Chapter 5

Forced oscillations

Textbook link: Tipler and Mosca 14.5

In Chapter 4 we explored the effect of damping on a system and we said that every system in the real world is, to a greater or lesser extent, a damped system in which energy is lost (dissipated) to the surroundings. In order to maintain the amplitude of any oscillation we must supply energy to the system at the same rate as it is lost to the surroundings; for example, pushing a child on a swing.

The equation for forced SHM is given in Equation (5.1):

$$m \frac{d^2x}{dt^2} + b \frac{dx}{dt} + kx = F_0 e^{i\omega t} \quad (5.1)$$

Compare this to Equation (4.3) for damped SHM; we have now applied an oscillating force represented as $F_0 e^{i\omega t}$, with an amplitude of F_0 and frequency ω . Note that this driving frequency ω is different to the natural *undamped* frequency of the oscillator, $\omega_0 = \sqrt{\frac{k}{m}}$ and different again to the frequency of the damped *unforced* oscillator, $\omega' = \sqrt{\omega_0^2 - \frac{b^2}{4m^2}}$.

Any forced oscillation consists of two distinct regimes:

1. An initial transient period during which the oscillations are established, and
2. A steady state period during which the oscillations have constant amplitude and a frequency equal to the driving frequency, ω .

The general solution of the equation for forced SHM shown in Equation (5.1) is a combination of these two regimes.

5.1 The Transient Solution

The solution to the transient component of the forced oscillator is identical to the solution of damped SHM (*i.e.* Equation (5.1) with the right-hand side set to zero; identical to Equation (4.3)). Its solution is shown in Equation (5.2):

$$\begin{aligned} x &= A_0 e^{-(\frac{b}{2m})t} e^{i(\omega' t + \delta')} \\ \text{or: } x &= A_0 e^{-(\frac{b}{2m})t} \cos(\omega' t + \delta') \end{aligned} \quad (5.2)$$

This contribution to motion establishes the oscillation, but rapidly decays with time constant $\tau = \frac{m}{b}$. The term δ' is simply the phase constant for this transient oscillation.

5.2 The Steady State solution

Once in the steady state, the energy which is put into the system during each cycle is equal to the energy dissipated per cycle due to the damping in the system. If there is no damping of the system, energy keeps being added to the system and the amplitude will increase indefinitely. This is **not** a steady state, and is an **unphysical result**. The frequency then of this ‘driven’ oscillator in the steady state is equal to the driving frequency.

The amplitude (and hence the energy) of the system in the steady state depends on both the **amplitude** and the **frequency** of the driving force. For a steady state, the solution to Equation (5.1) is in the form shown in Equation (5.3):

$$x = A e^{i(\omega t - \delta)} \quad (5.3)$$

The terms A and δ are defined in Equations (5.4) and (5.5) below:

$$A = \frac{F_0}{\sqrt{m^2 (\omega_0^2 - \omega^2)^2 + b^2 \omega^2}} \quad (5.4)$$

$$\delta = \arctan \left(\frac{b\omega}{m(\omega_0^2 - \omega^2)} \right) \quad (5.5)$$

5.3 Steady state behaviour - Resonance

When we vary the frequency of the driving frequency we find that the response of the driven system varies. If we examine the power transferred to the system as we vary the driving frequency, we see that there is a peak around the natural frequency, ω_0 , of the driven system (Figure 5.1). This is a phenomenon known as **resonance**.

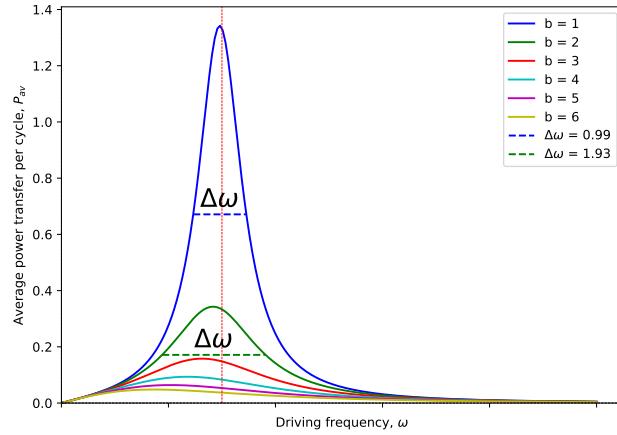


Figure 5.1: The power transferred to a system varies with the driving frequency. The power transferred to a system is proportional to ωA^2 . The full-width at half-maximum value is designated $\Delta\omega$. Note that as the damping coefficient b increases, there is a shift in maximum power transmission away from the natural frequency ω_0 .

If this system is damped, we also see a change in the ‘full-width at half-maximum’ value.¹ This FWHM is designated by $\Delta\omega$ in Figure 5.1 and is related to the Q -factor of the system via Equation (5.6):

$$\frac{\Delta\omega}{\omega_0} = \frac{1}{Q} \quad (5.6)$$

This allows us to determine the Q -factor of a system through measurement of resonance of the system, as a lightly-damped system (high Q) will give a sharp resonance with low $\Delta\omega$, while a more heavily damped system will lead to a more broad resonance profile.

This use of the Q -factor is important as it give us a measure of the ‘sharpness’ (or quality) of the resonance. It may be applied to many systems including electronic circuits.

We can also examine the phase of the oscillator’s displacement x relative to that of the driving force, δ , shown in Figure 5.2.

We can see that at low driving frequencies, there is very little phase shift between the two (the force is at a minimum when the displacement is at a minimum and is in the same direction), while at very high driving frequencies we approach

¹The ‘full-width at half maximum’ (FWHM) is a term widely used in signals processing to describe the ‘spread’ of a signal. A large FWHM indicates a broad signal over a range of frequencies, while a small FWHM indicates a sharp signal over a narrow band of frequencies.

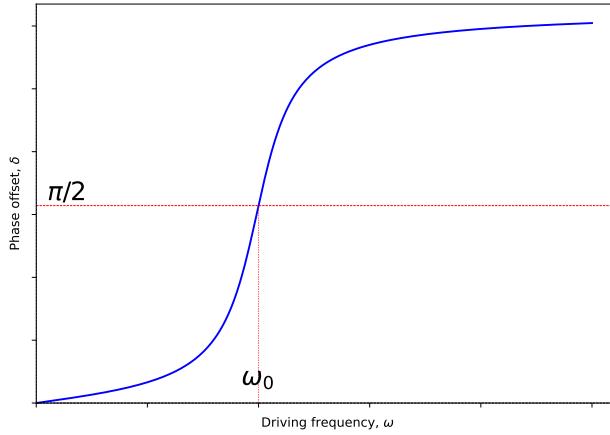


Figure 5.2: The phase shift δ of the observed oscillation varies relative to the driving frequency ω .

a maximum phase shift of $\frac{\pi}{2}$ between the two (force still at a minimum at minimum displacement, but now the force is directed in the opposite direction to the displacement; akin to SHM). The phase change is at its most rapid when the frequency of the driving force ω is similar to the natural frequency of the oscillator (ω_0). At this point, the driving force are “in quadrature” (90° out of phase), with the force at a maximum when x is changing most rapidly.

5.4 Full solution of the forced oscillator

As with the damped oscillator explored in Chapter 4, the full solution of the forced oscillator is found through the combination of the “transient state” (a free, damped system) and the “steady state” (fully forced oscillations). This is illustrated in Equation (5.7):

$$\begin{aligned} x &= \text{transient} + \text{steady state} \\ \text{i.e. } x &= A_0 e^{-(\frac{b}{2m})t} e^{i(\omega't+\delta')} + A e^{i(\omega t-\delta)} \end{aligned} \quad (5.7)$$

The terms in Equation (5.7) are as follows:

- δ is given by Equation (5.5);
- A given by Equation (5.4);
- A_0 and δ' depend on the initial conditions;
- ω is the driving frequency;
- ω' is the frequency of the damped (transient) oscillations.

We can simplify our view of Equation (5.7) by applying initial conditions whereby the displacement $x = 0$ at time $t = 0$. Therefore:

$$0 = A_0 e^{i(\delta')} + A e^{-i\delta} \quad (5.8)$$

...or, to rearrange, we obtain the form shown in Equation (5.9):

$$A_0 = -A e^{-i(\delta+\delta')} \quad (5.9)$$

As with any complex representation, we now compare the **real** and **imaginary** parts of the solution. Remember that Equation (5.9) expands *via* De Moivre's theorem to:

$$A_0 = -A (\cos(\delta + \delta') + i \sin(\delta + \delta')) \quad (5.10)$$

The imaginary component of Equation (5.10) reduces to zero (the solution, A_0 is fully real):

$$0 = -A \sin(\delta + \delta') \quad (5.11)$$

We can therefore relate δ and δ' as follows:

$$\delta + \delta' = \pi \quad i.e. \quad \delta' = \pi - \delta \quad (5.12)$$

We now examine the 'real' component of Equation (5.10), equal to A_0 :

$$A_0 = -A \cos(\delta + \delta') \quad (5.13)$$

...and doing the same analysis, now that we know that, under our initial conditions, $\delta + \delta' = \pi$:

$$A_0 = -A \cos \pi \quad i.e. \quad A_0 = A \quad (5.14)$$

We can now revisit Equation (5.7), now that we have values for δ , δ' and A_0 under our starting conditions ($x = 0$ when $t = 0$):

$$\begin{aligned} x &= A \left[e^{i(\omega t - \delta)} + e^{-(\frac{b}{2m})t} e^{i(\omega' t + \delta')} \right] \\ &= A \left[e^{i(\omega t - \delta)} + e^{-(\frac{b}{2m})t} e^{i(\omega' t + \pi - \delta)} \right] \\ &= A \left[e^{i(\omega t - \delta)} + e^{-(\frac{b}{2m})t} e^{i(\omega' t - \delta)} e^{i\pi} \right] \\ &= A \left[e^{i(\omega t - \delta)} - e^{-(\frac{b}{2m})t} e^{i(\omega' t - \delta)} \right] \end{aligned} \quad (5.15)$$

Finally, we can find the actual displacement of the oscillator by examining the real component of Equation (5.15), summarised in Equation (5.16):

$$\operatorname{Re}(x) = A \left[\cos(\omega t - \delta) - e^{-(\frac{b}{2m})t} \cos(\omega' t - \delta) \right] \quad (5.16)$$

We can visualise the amplitude of a forced oscillator, as shown in Figure 5.3. At very low frequencies, the oscillation has the same amplitude as the driving oscillation. This rapidly increases under resonance to a peak, but then drops rapidly to a value towards zero.

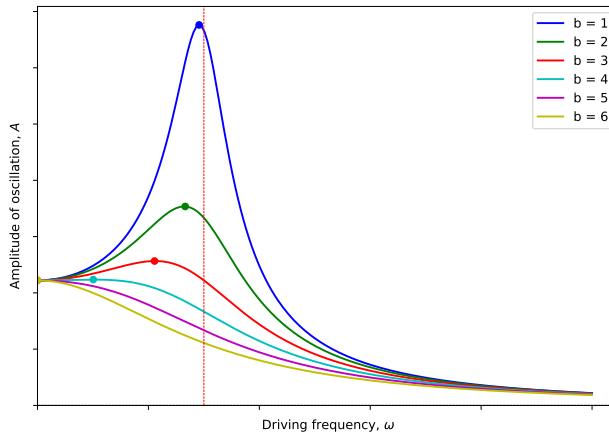


Figure 5.3: The amplitude of a forced oscillation varies with driving frequency; close to the driving amplitude at very low frequencies, close to zero at very high frequencies, with a peak occurring under resonant conditions. Note that the position of this peak varies with the damping coefficient b .

Note that the position of the resonant peak amplitude also varies with the damping coefficient b .

5.4.1 Special cases of forced oscillations

There are three special cases of forced oscillations to consider:

1. $\omega \ll \omega_0$
2. $\omega \gg \omega_0$
3. $\omega = \omega_0$

In the special case when $\omega = \omega_0$ and the damping coefficient b is small (allowing the damped frequency ω' to be approximately equal to the driving frequency ω), we can rewrite the expression of the displacement (Equation (5.16)) as shown in Equation (5.17):

$$\operatorname{Re}(x) = A \cos(\omega_0 t - \delta) \left[1 - e^{-(\frac{b}{2m})t} \right] \quad (5.17)$$

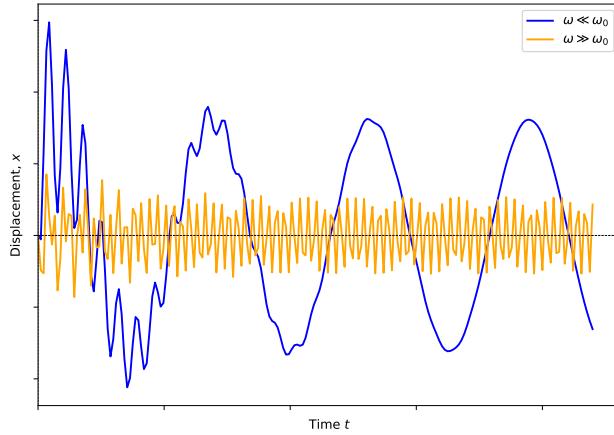


Figure 5.4: When the frequency of the driving oscillation ω is very different to the natural frequency of the system we see two regimes; firstly, where $\omega \ll \omega_0$ (blue), the natural frequency decays and we are left to observe the driving frequency at a large amplitude; when $\omega \gg \omega_0$ (orange), we see a much smaller amplitude oscillation at close to the driving frequency.

We can see that the oscillation will converge to a maximum value, at which point the energy put into the system becomes equal to the energy dissipated by the system through damping, as shown in Figure 5.5

5.5 Energy in driven oscillators

In a driven oscillator, energy is continually added to the system. When it reaches its steady state, the rate of loss of energy in each cycle due to damping of the system is equal to the work done by the driving force.

We can demonstrate this through integration; remember that (simplistically!) work done is “force \times distance”, so under varying force (as we have here), we can integrate the force with respect to displacement x to find an expression for the work done.

For the sake of convenience, we will consider the “real” component of the equation for forced SHM shown in Equation (5.1) (the left-hand-side is entirely “real”, so the “imaginary” component reduces to zero anyway):

$$m \frac{d^2x}{dt^2} + b \frac{dx}{dt} + kx = F_0 \cos \omega t \quad (5.18)$$

We also know that the steady state solution for the displacement x in SHM is as given in Equation (1.5). We can then find an expression for the velocity v by

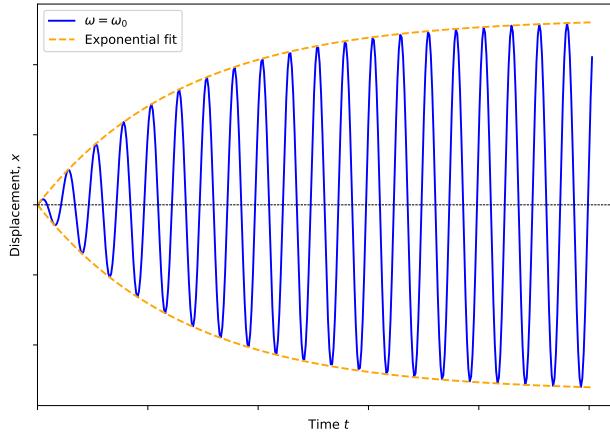


Figure 5.5: When the frequency of the driving oscillation $\omega = \omega_0$ (blue), we see the amplitude grow to a plateau at which point the energy put into the system is equal to the energy losses from the system.

finding the first derivative of this expression (Equation (5.19)):

$$\begin{aligned} x &= A \cos(\omega t - \delta) \\ v = \frac{dx}{dt} &= -\omega A \sin(\omega t - \delta) \end{aligned} \quad (5.19)$$

5.5.1 Energy input for driven oscillators

To find the energy which is put into the system over one cycle, we therefore need to find the integral of the force acting over one cycle of the oscillation, from $t = 0$ to $t = T$ (remember that T is the period of the oscillation). We do this integration by substituting our variable dx for vdt (as defined in Equation (5.19)):

$$\begin{aligned}
E_{\text{in}} &= \int_0^T \mathbf{F} \cdot d\mathbf{x} \equiv \int_0^T \mathbf{F} \cdot v dt \\
&= \int_0^T F_0 \cos \omega t \cdot [-\omega A \sin(\omega t + \delta)] dt \\
&= -\omega A F_0 \int_0^T \cos \omega t \cdot \sin(\omega t + \delta) \\
&= -\omega A F_0 \int_0^T \cos \omega t \cdot [\sin \omega t \cos \delta + \cos \omega t \sin \delta] \\
&= -\omega A F_0 \int_0^T [\cos \omega t \sin \omega t \cos \delta + \cos \omega t \cos \omega t \sin \delta] \\
&= -\omega A F_0 \int_0^T [\frac{1}{2} \sin 2\omega t \cos \delta + (\frac{1}{2} \cos 2\omega t + \frac{1}{2}) \sin \delta] \\
&= -\omega A F_0 [-\frac{1}{4\omega} \cos 2\omega t \cos \delta + (\frac{1}{4\omega} \sin 2\omega t + \frac{t}{2}) \sin \delta]_0^T \\
&= -\omega A F_0 \left(\left[-\frac{1}{4\omega} \cos 2\omega T \cos \delta + \left(\frac{1}{4\omega} \sin 2\omega T + \frac{T}{2} \right) \sin \delta \right] - \left[-\frac{1}{4\omega} \cos 0 \cos \delta + \left(\frac{1}{4\omega} \sin 0 + \frac{0}{2} \right) \sin \delta \right] \right)
\end{aligned} \tag{5.20}$$

Evaluation of this integral gives us the end result shown in Equation(5.21):

$$E_{\text{in}} = \frac{1}{2} \omega A F_0 T \sin \delta \tag{5.21}$$

If we recall that $\delta = \frac{\pi}{2}$ at resonance, substitution of this value into Equation (5.21) tells us that the maximum energy is transferred to the oscillator when driven at a resonant frequency - in line with our expectations.

If we now use the definition of δ in Equation (5.5), we can use trigonometry and the expression for A in Equation (5.4) to give the expression $\sin \delta = \frac{Ab\omega}{F_0}$; we can also use the fact that the period of oscillation T and the angular frequency ω are related by Equation (1.13) ($T = \frac{2\pi}{\omega}$), and simplify the expression for the energy input for the forced oscillator as shown in Equation (5.22):

$$E_{\text{in}} = \pi b \omega A^2 \tag{5.22}$$

5.5.2 Energy lost in driving oscillators

To now determine the amount of energy, we follow a similar process as followed in Section 5.5.1. This time however we need to determine the work done by the oscillator on the damping force during one cycle. This integral is set up as follows:

$$E_{\text{lost}} = \int_0^T F v dt \tag{5.23}$$

...but this time the force is the damping force, $F_{\text{damp}} = bv$.

$$E_{\text{lost}} = \int_0^T bv^2 dt \quad (5.24)$$

As before, we have the result that $v = -\omega A \sin(\omega t - \delta)$ (Equation (5.19)), and following through the integration in a similar manner as in the previous section we are led to the result in Equation (5.25) for a **driven damped oscillator at steady state**.

$$E_{\text{lost}} = \pi b \omega A^2 = E_{\text{in}} \quad (5.25)$$

This result should not be a surprise; at a steady state, we expect the energy lost to be equivalent to the energy put into the system, however it is nice that this is vindicated through the mathematics!

A similar result may be obtained for electrical systems in resonance.

Note: When $b = 0$ the above result appears to fail (no energy is lost, but none is put in??); in fact, in an undamped system there can be no steady state (as no energy is lost!), so the amplitude of vibration $A \rightarrow \infty$, so this result does not apply.

5.6 Impedance

As you learn about fields, you will examine electrical circuits and see that circuits containing capacitors and inductors are analogous to mechanical oscillators, with the electrical charge oscillating within the circuit. These properties of circuits have many important applications and you will find that much of the analysis we have done here can be applied to those electrical systems.

In the context of an electrical circuit it is useful to define the term **impedance** a measure of the opposition to the flow of current, and is defined as the ratio of the voltage to current ($\frac{V}{I}$) for a particular circuit component. For a resistor this is simply the resistance, however capacitors and inductors also possess impedance.

It turns out to be helpful to write this impedance as a complex number with components in both the real and imaginary plane. This works well with the complex representation of oscillations and naturally takes care of any differences in phase of the current and voltage in different components.

Bringing the analogy back to mechanical oscillations, a mechanical system can also be considered to have an impedance. We define the mechanical impedance as the **force required to produce unit velocity**, *i.e.* $Z_m = \frac{F}{v}$, or $F = vZ_m$. This is the mechanical equivalent of Ohm's law; force corresponding to the voltage, and velocity corresponding to current.

This idea of mechanical impedance will be useful in discussion of wave propagation.

Chapter 6

Simple wave motion and the Wave Equation

Textbook link: Tipler and Mosca, Section 15.1

A wave is a means by which energy and momentum are carried through space, without transporting matter. When we consider a medium which carries a wave, the particles of that material oscillate about a mean position, but have an average displacement of zero; *i.e.* they always return to their starting position. A wave can be of any shape - there is no requirement for a wave to be sinusoidal, though this is the simplest shape which we can consider mathematically.

We also consider waves to be either **transverse**; where the displacement of the medium is perpendicular to the direction of wave propagation (*e.g.* a wave travelling along a string), or they can be **longitudinal**; where the displacement of the medium is parallel to the direction of propagation (*e.g.* a sound wave passing through air).

A note on transverse waves

In contrast with longitudinal waves, the medium carrying a transverse wave is displaced perpendicular to the direction of travel. This gives rise to the phenomenon of **polarisation**. A *plane of polarisation* is defined as a plane containing the displacement direction and the direction of propagation. For any given transverse wave, two orthogonal independent polarisations are possible. All other polarisations may be constructed from weighted combinations of these two basic polarisations. For electromagnetic waves, it is the **electric field vector** which defines the plane of polarisation in combination with the direction of propagation. We will revisit these properties of transverse waves as we go through our discussion.

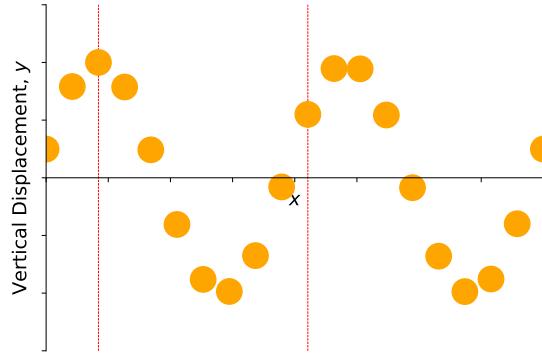


Figure 6.1: An animation to show the propagation of a wave through a material arising from an organised transverse oscillation of many particles. The two marked particles only move up and down, but the net effect of the organised motion is a propagation of energy to the right.

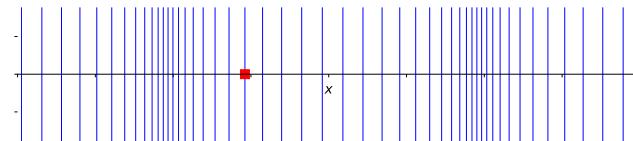


Figure 6.2: An animation to show the propagation of a wave through a material arising from an organised longitudinal oscillation of many particles. The marked particle now moves in the direction of propagation, but oscillates around a fixed point; the net effect of the organised motion is again a propagation of energy to the right.

6.1 Wave pulses

Where a wave is a sustained periodic disturbance which propagates energy through a medium, a **wave pulse** in contrast is any localised non-periodic disturbance propagating an energy pulse through the medium. A typical pulse is shown in Figure 6.3.

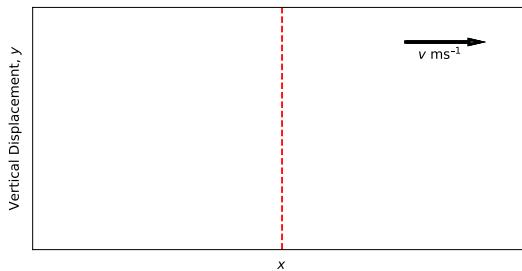


Figure 6.3: A wave pulse propagating through a system at velocity v ; this causes a temporary transverse disturbance of molecules in the system from their origin.

We illustrate a pulse graphically as any function defined as $y = f(x)$, where the $+x$ direction is the direction of propagation of the pulse. If the pulse propagates without changing shape, it becomes convenient to consider a moving reference frame within which the pulse is stationary; *i.e.* rather than imagining the pulse moving to the right along fixed axes, we keep the pulse stationary in our view and move the axes to the left.

In the moving frame then, the pulse is described as $y' = f(x')$ for all times, because the pulse does not change its shape.

We can inter convert between the two frames of reference by the relation:

$$x = x' + vt \quad (6.1)$$

...where v is the velocity of the pulse. This allows us to convert the position in the moving reference frame, x' , back to the position in the fixed reference frame by adding the distance vt .

If the shape of the pulse in the moving frame is defined as $y' = f(x')$ we can use Equation (6.1) to find the shape of the pulse in the static frame, y :

$$\begin{aligned} y &= f(x') \\ &= f(x - vt) \end{aligned} \quad (6.2)$$

The relation described in Equation (6.2) describes a wave moving to the **RIGHT**; for a pulse moving to the **left**, v becomes negative, and hence:

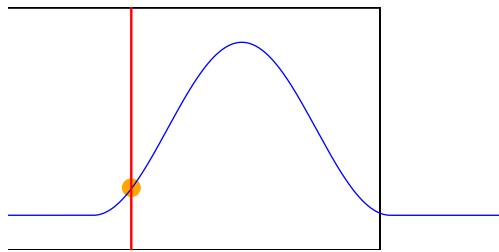


Figure 6.4: Rather than imagining a moving wave pulse, we can consider it from the reference frame of the wave pulse; in this example the wave pulse is stationary, but the space moves past it.

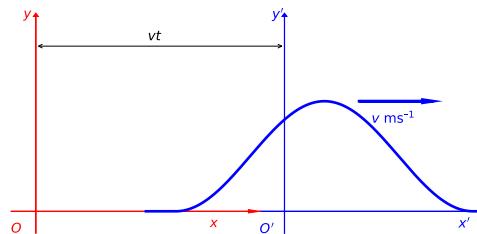


Figure 6.5: Another way of visualising the reference frames is to consider the x', y' space moving relative to the stationary x, y space; the offset between them can then be quantified as vt , where v is the velocity of the wave pulse and t is the time elapsed.

$$y = f(x + vt) \quad (6.3)$$

The function $y = f(x \pm vt)$ is known as the **wave function**; it describes the displacement of the medium, whether the transverse displacement of a string or the longitudinal displacement of air molecules in a sound wave.¹.

The *wave function* is a solution of the **wave equation** (Equation (6.4)):²

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} \quad (6.4)$$

Any function in the form $y = f(x \pm vt)$ is a solution of this wave equation; *i.e.* the wave equation describes the uniform propagation of **any** displacement, provided it does not change shape as it travels. There are numerous examples of such functions, including:

- $y = \exp(x - vt)^2$
- $y = \frac{\sin(x-vt)}{x-vt}$
- $y = \cos(x + vt)$

6.2 Deriving the wave equation

To better understand the wave equation, it is useful to know its derivation. To do this, we shall first consider a segment of string from a curved part of a wave pulse (Figure 6.6):

The length of this segment is $\approx \Delta x$ (for small angles), and the mass of this segment of string $m = \mu \Delta x$ (where μ is the mass per unit length). As we are considering a transverse wave, this segment of string will move vertically, and all forces acting on the segment arise from the tension force within the string. We can determine these forces by resolving the horizontal and vertical components of this force F using the parameters laid out in Figure 6.6.

By considering the net vertical force acting on this segment:

$$\begin{aligned} \sum F &= F \sin \theta_2 - F \sin \theta_1 \\ &= F (\tan \theta_2 - \tan \theta_1) \quad \text{for small angles} \end{aligned} \quad (6.5)$$

We can also determine an expression for the slope, S , of the segment of string at a given point:

¹It may seem counter intuitive that, for a wave moving with positive v in the $+x$ direction, we subtract vt ; but try thinking this way: after time t , the wave will have advanced further to the right. If, after one second, the shape of the wave has its peak at x_a , after three seconds, the peak will have advanced to $x_b = x_a + vt$; **however**, the value of the wavefunction at x_a at this new time (x'_a) will have come from the value of the wavefunction three seconds earlier, to the **left** of the point; *i.e.* $x'_a = x_a - vt$

²The ∂ symbol refers to *partial differentiation*.

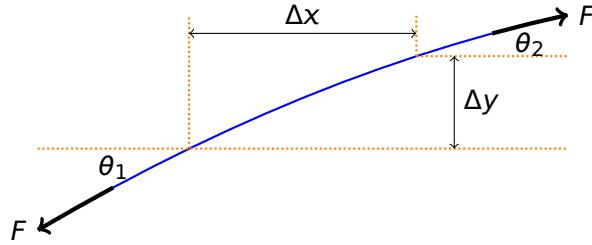


Figure 6.6: Here we consider a small portion of a string carrying a wave pulse; the tension in the string F is constant throughout, but resolving the forces via the angles θ_1 and θ_2 allows us to determine the equation of motion on this segment.

$$S = \tan \theta = \frac{\partial y}{\partial x} \quad (6.6)$$

This means we can now express the overall force acting on the segment in terms of the slope of the segment at its start and end points (assuming small angles!) by combining Equations (6.5) and (6.6) :

$$\sum F = F(S_2 - S_1) = F\Delta S \quad (6.7)$$

The quantity $F\Delta S$ in Equation (6.7) is the net force acting on the segment; so we can now apply Newton's second law:

$$F\Delta S = ma = \mu\Delta x \frac{\partial^2 y}{\partial t^2} \quad (6.8)$$

...or, to rearrange:

$$\frac{\Delta S}{\Delta x} = \frac{\mu}{F} \frac{\partial^2 y}{\partial t^2} \quad (6.9)$$

Since the slope ΔS of the segment is the 'rate of change of y with respect to x ' (Equation (6.6)), we can redefine the term in Equation (6.8); in the limit, as $\Delta x \rightarrow 0$:

$$\frac{\Delta S}{\Delta x} \approx \frac{\partial S}{\partial x} = \frac{\partial}{\partial x} \left(\frac{\partial y}{\partial x} \right) = \frac{\partial^2 y}{\partial x^2} \quad (6.10)$$

Finally, by combining this result in Equation (6.10) with the result in Equation (6.9), we obtain the wave equation for a stretched string (Equation (6.11)):

$$\frac{\partial^2 y}{\partial x^2} = \frac{\mu}{F} \frac{\partial^2 y}{\partial t^2} \quad (6.11)$$

As we mentioned in Section 6.1, any function of the form $y = f(x \pm vt)$ will be a solution to the wave equation (6.11). In the case of the string, the solution will solve the wave equation provided that $v^2 = \frac{F}{\mu}$.

6.3 The wave equation - proof by substitution

In this section, we will show the proof of the wave equation which we determined graphically in Section 6.2. Here we will use *partial differentiation*, explaining the ∂ notation we have already seen in this work.³

Let's work through this step by step:

1. Consider the general function $y = \cos(x - vt)$. Differentiate this first with respect to x :

$$\frac{\partial y}{\partial x} = -\sin(x - vt) \quad (6.12)$$

The second derivative therefore becomes:

$$\frac{\partial^2 y}{\partial x^2} = -\cos(x - vt) \quad (6.13)$$

2. Now we differentiate with respect to time t :

$$\frac{\partial y}{\partial t} = v \sin(x - vt) \quad (6.14)$$

... and find the second derivative:

$$\frac{\partial^2 y}{\partial t^2} = -v^2 \cos(x - vt) \quad (6.15)$$

3. We can now combine Equations (6.13) and (6.15) to eliminate the term $\cos(x - vt)$, we obtain the wave equation as required:

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} \quad (6.16)$$

³Partial differentiation is needed when we have multivariate expressions. We can only differentiate one variable at a time, so *partial* differentiation indicates that we are holding all other variables constant. In our wave equations, we have position and time variables, so when differentiating x with respect to y , we are holding the time t constant.

Through unpacking of Equation (6.16) we can uncover how it might be adapted to obtain wave equations for other systems. The term x represents the direction of propagation of the wave, while y represents the disturbance of the particle carrying the wave (in the case of a transverse wave, this is perpendicular to the direction of propagation).

For longitudinal sound waves, Equation (6.16) becomes:

$$\frac{\partial^2 s}{\partial x^2} = \frac{1}{v_s^2} \frac{\partial^2 s}{\partial t^2} \quad (6.17)$$

In this example, s is the displacement of molecules of the medium parallel to the direction of the propagation, while v_s is the velocity of sound in the medium.

For electromagnetic waves, Equation (6.16) becomes either:

$$\frac{\partial^2 E_z}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 E_z}{\partial t^2} \quad (6.18)$$

...or

$$\frac{\partial^2 B_y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 B_y}{\partial t^2} \quad (6.19)$$

...where E_z and B_y are the transverse components of the electric and magnetic field vectors (we will discuss this further later), and c is the speed of light.

Note: The speed of propagation cannot be deduced from the wave equation, instead it must be obtained from a model of the system concerned.

6.4 The Phase Velocity - the velocity of waves

We define the phase velocity as the speed of propagation of any particular point on a wave. Since the wave propagates without changing shape, it does not matter which point we pick.

The phase velocity will depend on a combination of the elastic and inertial terms.

For waves travelling on a string the phase velocity can be found from the tension in the string (F) and the linear mass density (mass per unit length) of the string (μ):

$$v = \sqrt{\frac{F}{\mu}} \quad (6.20)$$

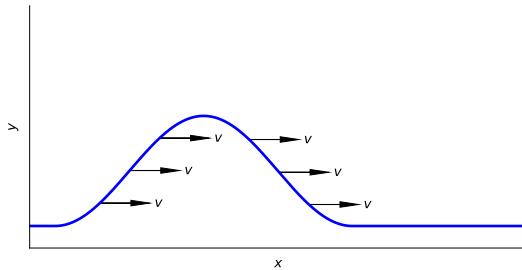


Figure 6.7: A reminder that every point in a wave pulse here moves at the same velocity v , such that the shape of the pulse is not changed.

This has parallels for sound waves in a fluid, with phase velocity found from the bulk modulus B (a factor describing the fluid's resistance to compression) and the equilibrium density ρ of the fluid:

$$v_s = \sqrt{\frac{B}{\rho}} \quad (6.21)$$

We can go further to look at the propagation of sound waves in an isotropic solid (solid of constant composition) - however we can only use this to consider a thin section, not a bulk solid (this is an acceptable model for considering earth tremors in the Earth's crust). This time the phase velocity depends on the density of the solid ρ and either the Young's modulus (Y , for longitudinal P waves) or the shear modulus (G , for transverse S waves):

$$v_s = \sqrt{\frac{Y}{\rho}} \text{ or } \sqrt{\frac{G}{\rho}} \quad (6.22)$$

These equations describe how wave phase velocities can be determined in solids and liquids; in a gas, we need to use other parameters, however the core relationship is still familiar:

$$v_s = \sqrt{\frac{\gamma RT}{M}} = \sqrt{\frac{\gamma k_B T}{m}} \quad (6.23)$$

Here, the terms R , T and M refer to the gas constant (units $\text{J K}^{-1} \text{ mol}^{-1}$), absolute temperature (units K) and the molar mass of the gas (units kg mol^{-1}) respectively⁴, while k_B and m refer to the Boltzmann constant (units J K^{-1})

⁴Note that the molar mass is expressed in kg mol^{-1} here, rather than the more common g mol^{-1} ; the reason for this is to ensure continuity and present everything in SI units.

and the mass of an individual gas molecule (units kg).

Finally, we can obtain a similar expression for the phase velocity of light waves:

$$c = \sqrt{\frac{1}{\mu_0 \epsilon_0}} \quad (6.24)$$

...where μ_0 and ϵ_0 are the permeability of free space and the permittivity of free space respectively.

6.5 Summary

In this chapter, we have introduced the principle of waves and the wave equation. Waves carry energy and momentum through space by localised organised oscillations without net transport of matter. They can be transverse (oscillation perpendicular to direction of propagation) or longitudinal (oscillation parallel to direction of propagation), or can have a more complex displacement pattern. Waves do not have to be sinusoidal, and can have any shape.

Any travelling wave can be described by the function in the form of $y = f(x \pm vt)$ which satisfies the wave equation, relating the displacement of the medium (y) to the position of the wave (x) and the phase velocity (v).

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} \quad (6.25)$$

Remember that the phase velocity cannot be predicted from the wave equation, but depends on the physics of the system. It generally results from a consideration of the elastic and inertial properties of the system. For waves on a string, $v = \sqrt{\frac{F}{\mu}}$, where F is the tension in the string and μ is the mass per unit length (linear mass density).

Chapter 7

Harmonic Waves

Textbook link: Tipler and Mosca, Section 15.2

A **harmonic wave** is a general term for a wave which, at some instant of time, can be described by a sinusoidal function (*i.e.* it is a sine or a cosine function). They are the simplest of waves to consider, and we will devote this chapter to exploring their properties.

7.1 Transverse sine and cosine waves

If we consider a string which is excited by a tuning fork or other object undergoing simple harmonic motion (SHM), we can imagine the shape of the save on the string at some instant of time appearing as a sine or a cosine wave (depending on our choice of origin) (Figure 7.1):

This sinusoidal appearance is known as a **harmonic wave**. Each point on the string oscillates up and down with the same frequency as the driving frequency. During the period T , the wave moves through distance λ :

$$v = \frac{\lambda}{T} = f\lambda \quad (7.1)$$

...where v is the phase velocity of the wave and f is the frequency in hertz (Hz), and λ defined as the **wavelength** *i.e.* the spacial repeat distance of the wave.

A harmonic wave has a unique frequency and wavelength, and other waves (*e.g.* wave pulses) may be regarded as a superposition of many harmonic waves of different frequencies (we will discuss the Fourier analysis of this later).

At any instant in time, the wave can be described by the relation in Equation (??):

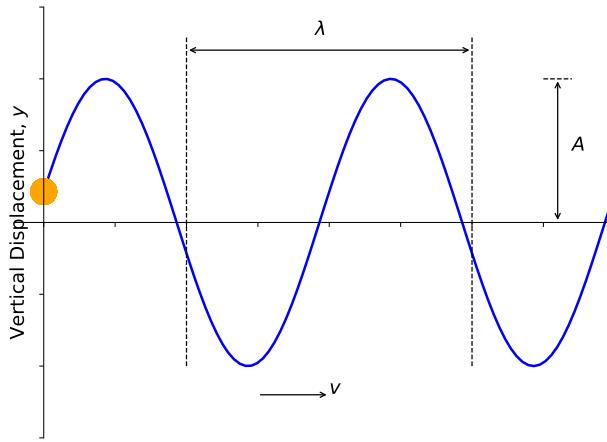


Figure 7.1: A sine wave in a string created by an oscillating particle. The wavelength λ is shown for one complete cycle, while the amplitude A is defined as the maximum deviation from the origin point.

$$y = A \sin(kx + \delta) \quad (7.2)$$

...where A is the amplitude of the wave and δ is the phase constant. Let's now choose the origin so that δ is equal to zero (*i.e.* there is no phase constant). Now we can show the periodic condition in Equation (7.3):

$$\sin kx = \sin k(x + \lambda) \quad (7.3)$$

Here we specify the periodic condition; the amplitude (and phase!) at position x is equal to the amplitude (and phase!) at position x **plus one wavelength**. Because this is a sine function, we know that, in order to achieve this condition, the value $k\lambda$ must be equal to 2π :

$$\begin{aligned} \sin \theta &= \sin (\theta + 2\pi) \\ \sin kx &= \sin (kx + k\lambda) = \sin (kx + 2\pi) \\ k\lambda &= 2\pi \end{aligned} \quad (7.4)$$

This allows us to give a value for k (Equation (7.5)):

$$k = \frac{2\pi}{\lambda} \quad (7.5)$$

The parameter k is defined as the **wave number** of the wave, in units radians per metre (rad m^{-1}).

7.2 Travelling waves

The wave we showed in Section 7.1 was a snapshot in time, so the wave was, in effect, static. We instead wish to consider a travelling wave. To do this, instead of writing x as in Equation (7.2) with $x - vt$ (see Section 6.1). Equation (7.2) then becomes Equation (7.6):

$$y = A \sin k(x - vt) = A \sin(kx - kvt) \quad (7.6)$$

As we have described, T is the period of the wave, so any point on the wave will oscillate up and down also with period T . This means that, for the wave position at time t , it will return to the same state at time $(t + T)$. We can therefore write Equation (7.6) in terms of this period:

$$\begin{aligned} y = A \sin(kx - kvt) &= A \sin(kx - kv(t + T)) \\ &= A \sin(kx - kvt - kvT) \end{aligned} \quad (7.7)$$

Applying the same reasoning as shown in Equation scheme (7.4), we are led to the result:

$$kvT = 2\pi \quad (7.8)$$

If we revisit our discussions on SHM (Chapter 1.7) we defined the relationship between T and 2π (Equation (1.13)) as $T = \frac{2\pi}{\omega}$ (and also $\omega = 2\pi f$). This allows us to define a number of factors as follows:

- **Phase velocity:**

$$v = \frac{2\pi}{kT} = \frac{\omega}{k} \quad (7.9)$$

- **Travelling wave to the right:**

$$y = A \sin(kx - kvt) = A \sin(kx - \omega t) \quad (7.10)$$

- **Travelling wave to the left:**

$$y = A \sin(kx + kvt) = A \sin(kx + \omega t) \quad (7.11)$$

It is worth noting for each of these factors:

1. The sign of ω :
 - negative for waves travelling to right;
 - positive for waves travelling to left
2. The wave expressions can be shown to satisfy the wave equation by substitution.

7.3 Complex representation of waves

We introduced complex numbers for descriptions of oscillations; we can use the same treatment for our wave equations. Just as $\sin(kx - \omega t)$ and $\cos(kx - \omega t)$ satisfy the wave equation, so too will $e^{i(kx - \omega t)}$. It is often convenient to write the wave expression in a complex form as shown in Equation (7.12):

$$y = Ae^{i(kx - \omega t)} \quad (7.12)$$

In this expression, the sine wave is given by the ‘imaginary’ component of y , while the cosine part is given by the ‘real’ component.

7.4 Energy carried by waves on a string

Suppose we now have a string attached to an oscillating driver at one end. As the driver oscillates, it imparts energy to the string at $x = 0$ by:

1. Stretching the string to give it potential energy, and
2. imparting transverse speed to the string to increase its kinetic energy.

As the waves move along the string, so the energy is transported along the string.

7.4.1 Potential energy of string segment

We now consider the same string segment as we showed in Figure 6.6, but now we consider it stretched (Figure 7.2). We can picture this as the string is “relaxed” in its ‘horizontal’ orientation (length of segment is Δx), but when a wave passes along it, the string elongates to accommodate the curve of the wave form. This means our segment now takes on a new length Δl .

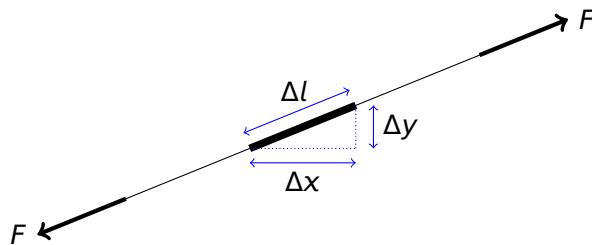


Figure 7.2: As the wave passes along a stretched string, we can consider it to be ‘stretched’; considering a string element of length Δx , it is stretched to new length Δl as the wave passes by.

The work done (U) in stretching the segment Δx can be expressed as:

$$\Delta u = F(\Delta l - \Delta x) \quad (7.13)$$

...where F is the tension in the string and the extension is given by $(\Delta l - \Delta x)$. We can apply Pythagoras to relate Δx and Δl :

$$\begin{aligned} \Delta l^2 &= \Delta x^2 + \Delta y^2 \\ &= \Delta x^2 \left[1 + \left(\frac{\Delta y}{\Delta x} \right)^2 \right] \end{aligned} \quad (7.14)$$

Therefore we can isolate Δl :

$$\Delta l = \Delta x \left[1 + \left(\frac{\Delta y}{\Delta x} \right)^2 \right]^{\frac{1}{2}} \quad (7.15)$$

We can approximate this expression by using the Taylor series expansion for $\sqrt{1+n}$, where $n = \left(\frac{\Delta y}{\Delta x} \right)^2$; assuming that the fraction is significantly less than one we can write this as an approximation and disregard terms past the first two terms:¹

$$\Delta l \approx \delta x \left[1 + \frac{1}{2} \left(\frac{\Delta y}{\Delta x} \right)^2 + \dots \right] \quad (7.16)$$

Rearranging this expression to obtain the expression for the extension of the string $\Delta l - \Delta x$:

$$\Delta l - \delta x \approx + \frac{\delta x}{2} \left(\frac{\Delta y}{\Delta x} \right)^2 + \dots \quad (7.17)$$

This now allows us to obtain an expression for the work done in stretching the spring solely in terms of the x and y displacement (from Equation (7.13))

$$\Delta u = \frac{F \Delta x}{2} \left(\frac{\Delta y}{\Delta x} \right)^2 \quad (7.18)$$

As we have an expression for y in terms of x (Equation (7.2)), we can differentiate this with respect to x to approximate $\frac{\Delta y}{\Delta x}$:

$$\frac{\Delta y}{\Delta x} \simeq \frac{dy}{dx} = kA \cos(kx - \omega t) \quad (7.19)$$

¹The Taylor expansion for this expression is $\sqrt{1+n} = 1 + \frac{n}{2} - \frac{n^2}{8} + \frac{n^3}{16} - \dots$

We also know that $v^2 = \frac{F}{\mu}$ (from the wave equation), and that $v = \frac{\omega}{k}$ (Equation (7.9)), therefore:

$$\Delta U = \frac{1}{2} \left(\frac{\mu \omega^2}{k^2} \right) \Delta x (kA \cos(kx - \omega t))^2 \quad (7.20)$$

Tidying up and cancelling, we obtain the expression for the potential energy stored in an element of string of length Δx (Equation (7.21)):

$$\Delta U = \frac{1}{2} \mu \omega^2 A^2 \Delta x \cos^2(kx - \omega t) \quad (7.21)$$

7.4.2 Kinetic energy of string segment

We will again consider the segment of string discussed in Section 7.4.1; an element Δx of the string, of mass Δm (Figure 7.3).

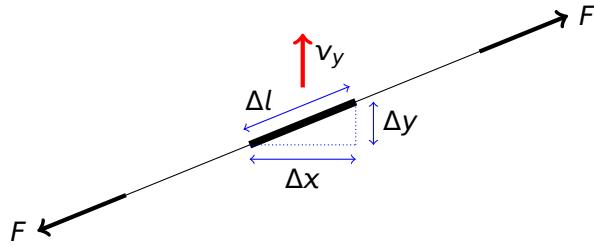


Figure 7.3: Similar to the situation shown above, the same string segment will have a kinetic energy associated with its velocity v_y .

As before, the segment is stretched to new length Δl , but the mass is still given by:

$$\Delta m = \mu \Delta x \quad (7.22)$$

...where μ is the mass per unit length. We now use the **transverse velocity** of the segment (*i.e.* **not** the wave velocity) to determine the kinetic energy of the segment:

$$\Delta KE = \frac{1}{2} \Delta m v_y^2 = \frac{1}{2} \mu \Delta x \left(\frac{dy}{dt} \right)^2 \quad (7.23)$$

Again, we know the expression for the vertical displacement y in terms of x and t (Equation (7.2)), so we now differentiate with respect to t :

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

- $v_y = \frac{dy}{dt} = -\omega A \cos(kx - \omega t)$

Therefore our expression for the kinetic energy becomes:

$$\Delta KE = \frac{1}{2} \mu \omega^2 A^2 \Delta x \cos^2(kx - \omega t) \quad (7.24)$$

A quick comparison of Equations (7.21) and (7.24) shows that these expressions are identical; *i.e.* the KE stored in the string is the **same** as the PE stored in the spring.

7.4.3 Total energy of wave on a string

We can therefore find the total energy of the string segment carrying a harmonic wave as the total of the kinetic and potential energies:

$$\Delta E = \Delta KE + \Delta U \quad (7.25)$$

i.e.:

$$\Delta E = \mu \omega^2 A^2 \Delta x \cos^2(kx - \omega t) \quad (7.26)$$

Note that the energy of the segment varies with time with twice the frequency of the wave (since $\cos^2 \theta = \frac{1}{2}(1 + \cos 2\theta)$).

We can also define the average energy at any point (Equation (7.27)) using the time-average definition $\langle \cos^2 \theta \rangle = \frac{1}{2}$:

$$\Delta E_{av} = \frac{1}{2} \mu \omega^2 A^2 \Delta x \quad (7.27)$$

...and we can define the average *energy density* (per unit length) as:

$$\varepsilon = \frac{\Delta E_{av}}{\Delta x} = \frac{1}{2} \mu \omega^2 A^2 \quad (7.28)$$

There are several things to note from this derivation:

- KE is at a maximum when displacement is zero
- At this point the string is most stretched, so PE is at a maximum also
- PE and KE are **in phase** (unlike in a pendulum)

These points are illustrated in Figure 7.4.

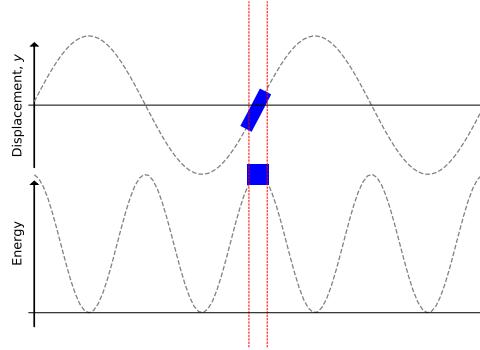


Figure 7.4: The kinetic energy and potential energy of a displaced string element have maxima and minima at the same points in the oscillation; as the element passes through the origin (top plot), it is at its most stretched (PE maximum), and it is traveling at its fastest (KE maximum).

7.4.4 Transport of energy and power

As the wave propagates along the string, energy is transported by the moving wavefront at speed v . The average energy passing a point on the string in time Δt is the average energy in the segment of length $\Delta x = v\Delta t$. This means we can rewrite Equation (7.27):

$$\Delta E_{av} = \frac{1}{2}\mu\omega^2 A^2 v \Delta t \quad (7.29)$$

Since the power transmitted is a *rate of change of energy*, i.e. $\frac{\Delta E}{\Delta t}$, we can obtain an expression for the average power transmitted:

$$\Delta P_{av} = \frac{dE_{av}}{dt} \approx \frac{\Delta E_{av}}{\Delta t} = \frac{1}{2}\mu\omega^2 A^2 v \quad (7.30)$$

From this result, we can see that both the average energy and average power transmitted are both proportional to A^2 ; a similar observation as in SHM (Chapter 1).

7.5 Summary

We have covered a large amount of derivations in this chapter, however the take-home points are the following:

When describing sine waves travelling through a medium, the following statements apply:

- For a wave travelling to the **right**: $y = A \sin(kx - \omega t)$
- For a wave travelling to the **left**: $y = A \sin(kx + \omega t)$
- The **phase velocity**: $v = \frac{\omega}{k}$ m s⁻¹

...where:

- k = wavenumber = $\frac{2\pi}{\lambda}$
- λ = wavelength /m
- ω = angular frequency = $2\pi f$ /rad m⁻¹
- f = frequency /Hz = $\frac{1}{T}$
- T = period /s
- A = amplitude /m

The average energy carried by a wave (per unit length) is given by:

$$\varepsilon = \frac{\Delta E_{av}}{\Delta x} = \frac{1}{2}\mu\omega^2 A^2 \quad (7.31)$$

...and the average power transmitted by the wave is given by:

$$\Delta P_{av} = \frac{\Delta E_{av}}{\Delta t} = \frac{1}{2}\mu\omega^2 A^2 v \quad (7.32)$$

Chapter 8

Reflection and Transmission at boundaries

We now turn our attention to what happens to waves and wave pulses when they encounter boundaries. We define a boundary as the dividing line between regions with different phase velocity. In the context of imagining our waves moving along strings, a boundary can exist either between the string and a rigid anchoring point, or at a point where the two strings join, each with a different mass density (*i.e.* a thick string joining to a thin string).

Any wave incident on such a boundary between regions with different phase velocities will be partly reflected back from the boundary and partly transmitted through the boundary.

In the case of a boundary between a thick string and a thin string, the phase velocity v is related to the tension F and the mass density μ via $v^2 = \frac{F}{\mu}$; *i.e.* heavier strings will have a lower phase velocity.

Considering reflection/transmission across such a boundary in a purely qualitative manner:

1. A wave propagating along a thin string towards a boundary with a thicker string will be reflected from the boundary **with inversion**, as well as a proportion of the energy transmitted as a wave into the thicker string.
2. A wave propagating along a thick string towards a boundary with a thinner string will be reflected from the boundary **without inversion**, as well as a proportion of the energy transmitted as a wave into the thinner string.

This qualitative outcome is illustrated in Figures 8.1 and 8.2.

If instead we fix a uniform string to an immovable anchor rather than another, thicker, string, we will get complete reflection of the wave pulse with inversion. This is akin to saying the anchor has an infinite mass density, μ .

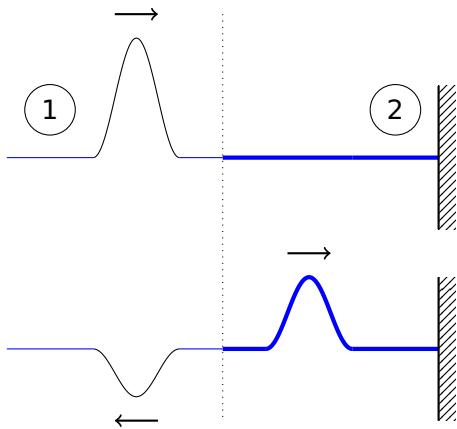


Figure 8.1: When a wavepulse originates in string 1 (where $\mu_1 \ll \mu_2$), the wave will be both reflected with inversion and partially transmitted into string 2 of greater μ

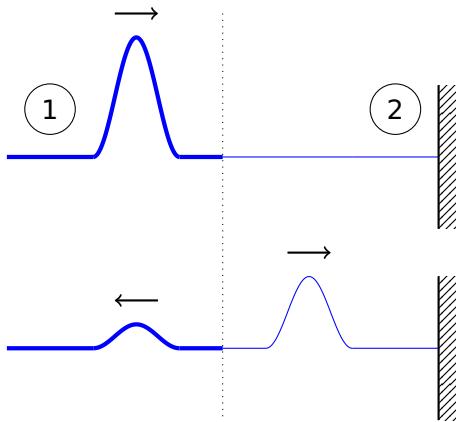


Figure 8.2: If we reverse the situation with the wavepulse originating in the string with greater μ (now $\mu_1 \gg \mu_2$), the wave is still partially transmitted into string 1, but the reflection within string 2 is no longer inverted.

To quantitatively assess these outcomes, we can obtain the reflected and transmitted amplitudes for harmonic waves by considering the power transmitted through the string as the wave propagates.

8.1 Power transmitted and reflected at a boundary

The principles of energy conversion state that, as a wave encounters a boundary, the energy in the incident wave must equal the total energy of the reflected and transmitted waves. This principle therefore also applies to the total power of the system, *i.e.*:

$$P_i = P_r + P_t \quad (8.1)$$

We can therefore use Equation (7.30) to derive an expression for the power before and the powers after the wave encounters the boundary (Equation (8.2)):

$$\frac{1}{2}\mu_1\omega^2A_i^2v_1 = \frac{1}{2}\mu_1\omega^2A_r^2v_1 + \frac{1}{2}\mu_2\omega^2A_t^2v_2 \quad (8.2)$$

...or, since $\omega = v_n k_n$ and $F = \mu_n v_n^2$ (rearrangements of equations seen previously):

$$\frac{1}{2}Fk_1\omega A_i^2 = \frac{1}{2}Fk_1\omega A_r^2 + \frac{1}{2}Fk_2\omega A_t^2 \quad (8.3)$$

Note that: * A_i , A_r and A_t are the amplitudes of the incident, reflected and transmitted waves respectively; * μ_1 , v_1 , μ_2 , v_2 refer to the mass per unit length and the phase velocities for strings 1 and 2 respectively; * k_1 and k_2 are the wavenumbers for each of the two strings; * ω is the **same** for all waves - this depends only on the source; * λ will be **different** on each string; since $\lambda = \frac{v}{f} = \frac{2\pi v}{\omega}$, *i.e.* λ will be smaller on the heavier string; * Conversely $k = \frac{2\pi}{\lambda}$ will be larger on the heavier string; * $\mu = \frac{F}{v^2}$; the tension F will be the same in both strings.

When we compare the proportion of the incident power which is reflected, we can show that:

$$\frac{\text{Reflected power}}{\text{Incident power}} = \frac{k_1 A_r^2}{k_1 A_i^2} = \left(\frac{k_1 - k_2}{k_1 + k_2} \right)^2 \quad (8.4)$$

We can also show the proportion of the incident power which is transmitted:

$$\frac{\text{Transmitted power}}{\text{Incident power}} = \frac{k_2 A_t^2}{k_1 A_i^2} = \frac{4k_1 k_2}{(k_1 + k_2)^2} \quad (8.5)$$

Note These results **only** hold for waves on strings where the tension, F , is the same in both strings. We will discuss a more general result in the next section.

Proof of power ratios

From our expression of the conservation of powers (Equation (8.3)), we can cancel the common terms:

$$k_1 A_i^2 = k_1 A_r^2 + k_2 A_t^2 \quad (8.6)$$

However, at the interface, the wave must be continuous on both sides; therefore the amplitude on each side must be the same:

$$A_i + A_r = A_t \quad (8.7)$$

We now substitute Equation (8.7) into (8.6) to eliminate A_r :

$$k_1 A_i^2 = k_1 (A_t - A_i)^2 + k_2 A_t^2 \quad (8.8)$$

We can then expand and rearrange this:

$$\begin{aligned} k_1 A_i^2 &= k_1 (A_t - A_i)^2 + k_2 A_t^2 \\ &= k_1 (A_t^2 - 2A_t A_i + A_i^2) + k_2 A_t^2 \\ 0 &= k_1 A_t^2 - 2k_1 A_t A_i + k_2 A_t^2 \\ &= k_1 \frac{A_t^2}{A_i^2} - 2k_1 \frac{A_t}{A_i} + k_2 \frac{A_t^2}{A_i^2} \\ 0 &= \left(\frac{A_t}{A_i} \right)^2 (k_1 + k_2) - 2k_1 \frac{A_t}{A_i} \end{aligned} \quad (8.9)$$

This is a quadratic equation, and we can then say the following; either:

$$\frac{A_t}{A_i} = 0 \quad (8.10)$$

...which represents total reflection (zero transmission), and not what we would expect to have with two joined strings, or:

$$\left(\frac{A_t}{A_i} \right) = \frac{2k_1}{k_1 + k_2} \quad (8.11)$$

...from which we obtain:

$$\frac{k_2}{k_1} \left(\frac{A_t}{A_i} \right)^2 = \frac{4k_1 k_2}{(k_1 + k_2)^2} \quad (8.12)$$

...as required in Equation (8.5) for the ratio of transmitted to incident power.

We can obtain the result for Equation (8.4) for ratio of reflected to incident power in a similar manner by substitution of Equation (8.7) into (8.6) to instead eliminate A_t .

8.2 Example of reflection and transmission

Consider a wave travelling from a light string to a heavy string, where $\mu_2 = 4\mu_1$.

Remember that:

$$\mu = \frac{F}{v^2} = F \frac{k^2}{\omega^2} \quad (8.13)$$

i.e.

$$k \propto \sqrt{\mu} \quad (8.14)$$

We can therefore determine:

- $k_2 = 2k_1$ via the square-root relationship
- The fraction of power reflected will be:

$$\frac{A_r^2}{A_i^2} = \left(\frac{1-2}{1+2} \right)^2 = \frac{1}{9} \quad (8.15)$$

- The fraction of power transmitted:

$$\frac{k_2 A_t^2}{k_1 A_i^2} = \left(\frac{4 \times 2 \times 1}{(1+2)^2} \right) = \frac{8}{9} \quad (8.16)$$

8.3 The impedance of a piece of string

The examples we considered in Section 8.1 were a particular result under constant tension. We can generalise the result by considering the impedances of the media on either side of the boundary.

We discussed the concept of impedance (both electrical and mechanical) in Section 5.6; however the salient points are:

- “Impedance” describes the property of a system which resists motion, either mechanical or motion of charge;
- Any material through which waves propagate presents impedance to those waves;

- In general, impedance depends on inertia and elasticity;
- For a string, we define impedance as:

$$Z = \frac{\text{transverse force}}{\text{transverse velocity}} \quad (8.17)$$

- i.e. for a given force, a large Z implies a small velocity and *vice versa*;
- Similar definitions can be written for longitudinal waves.

We now consider a string which is driven by an oscillating force (Figure 8.3)

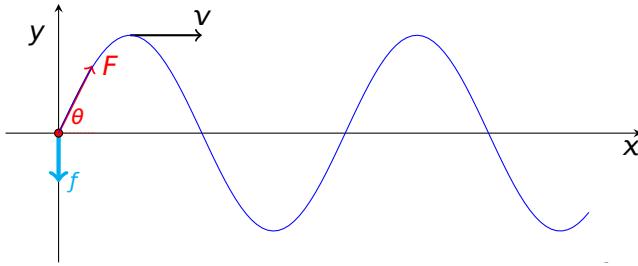


Figure 8.3: When a wave is driven by an oscillating force at its origin (f), the element of the string indicated by the dot experiences the tension within the string, F , at angle θ from the direction of propagation.

The driving force on this string is given by the relationship:

$$f = -f_0 \cos \omega t = -f_0 e^{i\omega t} \quad (8.18)$$

This driving force is negative because f points downwards at time $t = 0$.

As this is a wave, the vertical displacement at any point is given by the relationship:

$$y = A e^{i(kx - \omega t)} \quad (8.19)$$

We now consider the tension in the string, F , and resolve this in the transverse direction. At $x = 0$, we assume small angle of θ :

$$\begin{aligned} f &= -F \sin \theta \\ &\simeq -F \tan \theta \quad \text{for small angles} \\ &= -F \frac{\partial y}{\partial x} \end{aligned} \quad (8.20)$$

We can now use the definition of impedance given above ($Z = \frac{\text{force}}{\text{velocity}}$) and the expressions for the force and velocity:

$$\begin{aligned} Z &= \frac{f}{v_y} \\ &= -F \frac{\partial y}{\partial x} \div \frac{\partial y}{\partial t} \end{aligned} \quad (8.21)$$

We already have expressions for $\frac{\partial y}{\partial x}$ and $\frac{\partial y}{\partial t}$ by differentiating the wave equation $y = A e^{i(kx - \omega t)}$:

$$\frac{\partial y}{\partial x} = ikA e^{i(kx - \omega t)} \quad (8.22)$$

$$\frac{\partial y}{\partial t} = -i\omega A e^{i(kx - \omega t)} \quad (8.23)$$

Therefore:

$$\begin{aligned} Z &= \frac{-F \times ikA e^{i(kx - \omega t)}}{-i\omega A e^{i(kx - \omega t)}} \\ &= \frac{Fk}{\omega} \end{aligned} \quad (8.24)$$

Remember also that the phase velocity $v = \frac{\omega}{k}$, so we can also express the impedance as $Z = \frac{F}{v}$ and $Z = \mu v$ (because $v^2 = \frac{F}{\mu}$), where μ is the mass per unit length of the string and F is the tension within the string.

8.4 Reflection and transmission revisited

We can now express our previous result of the reflection and transmission coefficients (derived in Equation (8.9)) more generally:

Reflection coefficient:

$$\frac{A_r}{A_i} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad (8.25)$$

Transmission coefficient:

$$\frac{A_t}{A_i} = \frac{2Z_1}{Z_1 + Z_2} \quad (8.26)$$

Writing these in terms of the power (the approach used in Section 8.1), we obtain the following expressions:

$$\frac{\text{Reflected power}}{\text{Incident power}} = \frac{Z_1 A_r^2}{Z_1 A_i^2} = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2} \right)^2 \quad (8.27)$$

and:

$$\frac{\text{Transmitted power}}{\text{Incident power}} = \frac{Z_2 A_t^2}{Z_1 A_i^2} = \frac{4Z_1 Z_2}{(Z_1 + Z_2)^2} \quad (8.28)$$

These are general expressions which apply in mechanical, electrical and optical systems. From them, we can see that:

- If $Z_2 > Z_1$, the reflected wave is inverted;
 - From Equation (8.25), the term $Z_1 - Z_2$ is negative under these conditions, leading to a negative amplitude
- If the second string is a “wall” (*i.e.* immovable, infinite μ)
 - $Z_2 \rightarrow \infty$;
 - $A_r = -A_i$ (by energy conservation);
 - $A_t = 0$;
 - Therefore the wave is fully reflected and inverted.
- If $Z_2 = Z_1$ we have:
 - Impedance matching;
 - No reflection;
 - Maximum power transfer.

8.5 Impedance - Miscellaneous cases

For **longitudinal (sound) waves**, generally we expect the impedance Z to be described by $Z = \rho_0 v_p$, where:

- ρ is the mean density of the medium
- v_p is the phase velocity of the wave

Example values for longitudinal sound waves are: * Air: $\sim 400 \text{ kg m}^{-2} \text{ s}^{-1}$ * Water: $1.45 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}$ * Steel: $3.9 \times 10^7 \text{ kg m}^{-2} \text{ s}^{-1}$

For **transverse waves on a string**, the impedance Z is described by $Z = \mu v_p$, where:

- μ is the mass per unit length
- v_p is the phase velocity of the wave

For **electromagnetic waves**, the impedance depends on the medium under consideration:

- In a dielectric medium:

$$Z = \sqrt{\frac{\mu\mu_0}{\epsilon\epsilon_0}} \quad (8.29)$$

where in this case μ and ϵ are the permittivity and the permeability of the medium.

- In free space:

$$Z = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.6\Omega \quad (8.30)$$

- For a light wave in a dielectric medium:

$$Z = \frac{1}{n} \quad (8.31)$$

where n is the refractive index of the medium (see later).

Chapter 9

Sound waves and the Doppler effect

Textbook link: Tipler and Mosca, Section 15.2-4 As has been mentioned already, sound waves are a longitudinal wave propagated by the localised displacement of air molecules in the direction of propagation. This displacement of air molecules within sound waves can be described by the function shown in Equation(9.1).

$$s(x, t) = s_0 \sin(kx - \omega t) \quad (9.1)$$

In contrast to the transverse waves previously discussed, there are now only two dimensions to this function; while we considered the transverse displacement y of an element of the medium carrying the transverse wave, in this longitudinal wave the longitudinal displacement s is in the x direction, *i.e.* the same direction as the propagation of the wave.

This displacement of the molecules leads to changes in both the density (ρ) and the pressure (p) of the medium. It is worth recognising that both p and ρ are out of phase with the displacement (when the displacement is at a maximum, the pressure and density are at a minimum):

$$p = p_0 \sin(kx - \omega t - \frac{\pi}{2}) \quad (9.2)$$

... where the initial pressure $p_0 = \rho\omega vs_0$ and v is the phase velocity.

9.1 Energy of sound waves

When considering the energy of sound waves, we can examine the expression we already have for transverse waves and adapt this for our longitudinal waves.

Recall that the expression for the energy within a transverse wave is given in Equation (7.27):

$$\Delta E_{\text{av}} = \frac{1}{2} \mu \omega^2 A^2 \Delta x \quad (9.3)$$

To adapt this for longitudinal sound waves, we perform the following substitutions:

- We replace the linear mass density μ (units kg m^{-1}) with the density of the medium, ρ (units kg m^{-3})
- We replace our transverse amplitude A with the longitudinal displacement s_0
- To keep units congruent, we replace the change in segment length caused by the wave Δx with the change in segment volume caused by the wave ΔV

Our result is therefore:

$$\Delta E_{\text{av}} = \frac{1}{2} \rho \omega^2 s_0^2 \Delta V \quad (9.4)$$

9.2 Wave intensity

We have mentioned that with sound waves we are now considering a three-dimensional volume. This means we need to consider the effect of this on the energy of the wave at a distance r from its origin. The energy at a given distance will be spread uniformly over a spherical surface; therefore we need to consider the power per unit area of this surface. This is the **intensity** of the wave.

$$I = \frac{P_{\text{av}}}{4\pi r^2} \quad (9.5)$$

We already know that the average power is defined as the rate of change of the average energy:

$$P_{\text{av}} = \frac{\Delta E_{\text{av}}}{\Delta t} \quad (9.6)$$

We can therefore rewrite Equation (9.5) taking this into account.

$$\begin{aligned} I &= \frac{\Delta E_{\text{av}}}{4\pi r^2 \Delta r} \frac{\Delta r}{\Delta t} \\ &= \frac{\Delta E_{\text{av}}}{\Delta V} v \end{aligned} \quad (9.7)$$

In Equation (9.7) we identify the term $\frac{\Delta r}{\Delta t}$; this is the speed at which the wave travels from the centre of the sphere, so is the phase velocity v . We can also say that the term $4\pi r^2 \Delta r$ is the rate of change of volume, ΔV . We now use our expression for the average energy (Equation (9.4)) to simplify this expression:

$$I = \frac{1}{2} \rho \omega^2 s_0^2 v = \frac{p_0^2}{2\rho v} \quad (9.8)$$

In other words, the intensity of the sound wave travelling at constant speed v through a medium of constant density ρ at a point in space is proportional to the square of the amplitude of the wave, p_0 .

9.3 Levels of intensity

The human ear perceives sounds according to the logarithm of their intensity - not the absolute value of the intensity.¹ To represent an **intensity level** we use the term **decibel** (dB). This intensity level, β is represented in Equation (9.9) as follows:

$$\beta = 10 \log_{10} \left(\frac{I}{I_0} \right) \quad (9.9)$$

The term I_0 is the absolute intensity considered to be at the absolute limit of human hearing, where $I_0 = 10^{-12} \text{ W m}^{-2}$. A description of approximate intensity levels is shown in Table 9.1.

Table 9.1: A description of the approximate decibel level of particular sounds.

β /dB	Description
0	Hearing threshold
40	Library
70	Busy traffic
120	Pain threshold

9.4 The Doppler Effect (non-relativistic)

You have already met the Doppler effect for light in the context of the Special Relativity course; here we will briefly revisit it in the context of non-relativistic

¹When you encounter electronics you will find potentiometers labelled “audio taper”; this describes their use in audio applications in which the resistance is a logarithmic response to accommodate our logarithmic perception of sound.

cases.

The general principles of the Doppler effect are unchanged, namely:

- If the source and observer move relative to each other, the observed frequency is different from the emitted frequency;
- When the source and observer move towards each other, $f_{\text{obs}} > f_{\text{source}}$;
- When the source and observer move away from each other, $f_{\text{obs}} < f_{\text{source}}$;
- The frequency change, Δf , depends on whether the source or observer move relative to each other.

Consider a source moving relative to its surrounding medium at a speed of u_s . We can visualise this as a moving ‘dipper’ in a pool of water (Figure 9.1)

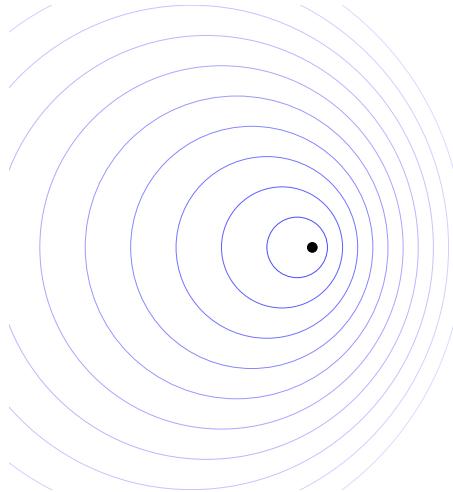


Figure 9.1: When a dipper (marked by the black dot) moves relative to the water, we see that the wavefronts ‘bunch’ in the direction of motion and diverge behind the motion of the dipper.

This is a visual representation of a number of key statements:

- The speed of waves v in the medium is independent of the movement of the source;
- The source produces waves at a frequency f_0 ;
- In a given time frame, Δt , the source will emit N_s wavefronts, where:

$$N_s = f_0 \Delta t \quad (9.10)$$

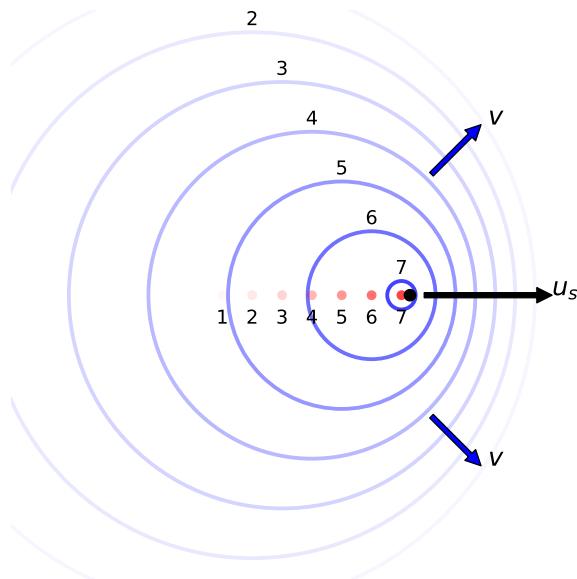


Figure 9.2: We can look at this in more detail by showing where the dipper was for each of the spreading wavefronts. Here the dipper is moving forward at speed u_r , while the wavefronts spread out from their point of origin at speed v .

From these statements, we can calculate the observed wavelength, λ' , by considering the relative distance traveled by the wave in a given time-frame and the number of waves produced by the source in that time-frame (Equation (9.11)):

$$\lambda' = \frac{\text{relative distance}}{\text{no. of waves}} = \frac{(v \pm u_s)\Delta t}{f_0\Delta t} \quad (9.11)$$

We can now consider two extremes; the observed wavelength in front of the source (to the right in the diagram), and the observed wavelength behind the source (to the left in the diagram):

$$\lambda'_{\text{behind}} = \frac{v + u_s}{f_0} \quad \lambda'_{\text{in front}} = \frac{v - u_s}{f_0} \quad (9.12)$$

We can now determine the frequency observed by using these expressions for the wavelength. Firstly, we determine the number of wavefronts passing the observer in the time frame Δt :

$$N_{\text{obs}} = \frac{v_{\text{obs}}\Delta t}{\lambda'} \quad (9.13)$$

...where v_{obs} is the speed of the waves relative to the observer; i.e. $v_{\text{obs}} = v \pm u_{\text{obs}}$; where u_{obs} corresponds to the observer moving to the right in the diagram (as this reduces the relative velocity between the observer and the wave). We can now rewrite the expression in Equation (9.13) in terms of the phase velocity v and the relative speed of the observer, u_{obs} :

$$N_{\text{obs}} = \frac{v \pm u_{\text{obs}}\Delta t}{\lambda'} \quad (9.14)$$

...and finally we have the observed frequency f' of a source in motion:

$$f' = \frac{N_{\text{obs}}}{\Delta t} = \frac{v \pm u_{\text{obs}}}{\lambda'} \quad (9.15)$$

When we combine this with the equation for λ' , we obtain the general result (Equation @ref{eq:ch9-dopplereffect7})

$$f' = \left(\frac{v \pm u_{\text{obs}}}{v \pm u_s} \right) f_0 \quad \text{or} \quad f' = \left(\frac{1 \pm \frac{u_{\text{obs}}}{v}}{1 \pm \frac{u_s}{v}} \right) f_0 \quad (9.16)$$

It is important to pay attention to the signs when using the above equations; it can be easy to confuse the positive directions. The sign convention can help with this:

- The direction from the observer towards the source is positive

You can check your results using the principles:

- The observed frequency f' **increases** when the source and observer approach each other;
- The observed frequency f' **decreases** when the source and observer move apart.

Chapter 10

Superposition and Standing Waves

Textbook link: Tipler and Mosca, Section 15.1

10.1 Superposition of harmonic waves

In Section 6.1 we showed that the wave equation (eq. (6.4)) is satisfied by any function in the form $y = f(x \pm vt)$. We can go further than this, and specify general expressions which can satisfy the wave equation.

Specifically, if we have two functions, y_1 and y_2 (Equation (10.1)), which satisfy the wave equation, then their sum (Equation (10.2), including scaling constants C_1 and C_2) must also satisfy the wave equation.

$$\begin{aligned} y_1 &= A_1 \sin [k_1(x \pm vt)] \\ y_2 &= A_2 \sin [k_2(x \pm vt)] \end{aligned} \tag{10.1}$$

$$y_3 = C_1 y_1 + C_2 y_2 \tag{10.2}$$

Using double angle formulae we can demonstrate that y_3 can also be written in the form $y = f(x \pm vt)$, thus satisfying the wave equation.¹ Therefore, the new wave y_3 is a linear superposition of the original waves y_1 and y_2 .

It is worth remembering that the wave equation was derived in the case of small amplitude disturbances; therefore if the vibration amplitudes become too large,

¹When including scaling factors, this can become less than trivial, but can still be done.

the principle of superposition may fail. This can lead to some very interesting effects, including **non-linear optics**²

10.2 Two waves with same amplitude and frequency

Let's now consider two waves with the same amplitude and frequency; but differing in phase (Equation (10.3)):

$$\begin{aligned} y_1 &= y_0 \sin(kx - \omega t) \\ y_2 &= y_0 \sin(kx - \omega t + \delta) \end{aligned} \quad (10.3)$$

If the two waves are superimposed, for example if they are travelling through the same medium, the resultant wavefunction is just the sum of y_1 and y_2 (Equation (10.4)):

$$\begin{aligned} y &= y_1 + y_2 \\ &= y_0 \sin(kx - \omega t) + y_0 \sin(kx - \omega t + \delta) \end{aligned} \quad (10.4)$$

In many situations it is mathematically simpler to use the complex exponential notation to treat waves and oscillations. However, in the case of a linear addition of two waves, it is simpler to just use trigonometric identities.

Here we will use the identity:

$$\sin \theta_1 + \sin \theta_2 = 2 \sin\left(\frac{\theta_1 + \theta_2}{2}\right) \cos\left(\frac{\theta_1 - \theta_2}{2}\right) \quad (10.5)$$

... we obtain:

$$y = 2y_0 \sin\left(kx - \omega t + \frac{\delta}{2}\right) \cos\left(\frac{-\delta}{2}\right) \quad (10.6)$$

As an aside, we can do the same thing with complex notation; we can write our wave superposition as follows:

$$y = y_1 + y_2 = y_0 e^{i(kx - \omega t)} + y_0 e^{i(kx - \omega t + \delta)} \quad (10.7)$$

Remembering that we are interested in the ‘imaginary’ component at the end since we started with y_1 and y_2 as sine waves, we can write this superposition as:

²Related to the non-linear response of the electromagnetic interactions with the medium, rather than light travelling in straight lines!

$$y = y_0 e^{i(kx - \omega t)} (1 + e^{i\delta}) \quad (10.8)$$

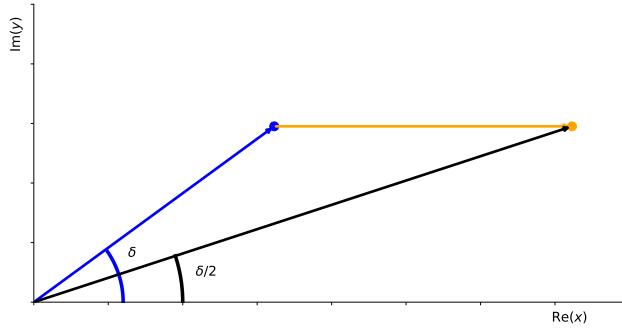


Figure 10.1: We can illustrate addition of complex numbers on an Argand diagram; here we show the sum $(1 + e^{i\delta})$; the result of this is the argument is halved from δ to $\frac{\delta}{2}$.

We can show using an Argand diagram (Figure 10.1) that the term $(1 + e^{i\delta})$ is equal to $2 \cos\left(\frac{\delta}{2}\right) e^{i\frac{\delta}{2}}$, and hence:

$$y = 2y_0 \cos(\delta/2) e^{i(kx - \omega t + \delta/2)} \quad (10.9)$$

When we expand this expression using De Moivre's theorem, we obtain the same result as in Equation (10.6):

$$y = \underbrace{2y_0 \cos\left(\frac{\delta}{2}\right)}_{\text{New amplitude}} \underbrace{\sin\left(kx - \omega t + \frac{\delta}{2}\right)}_{\text{Travelling wave}} \quad (10.10)$$

This tells us that the resulting wave has the same frequency as the component waves, but a different amplitude and phase. This new amplitude is given by the expression in Equation (10.11):

$$A_{\text{new}} = 2y_0 \cos\left(\frac{\delta}{2}\right) \quad (10.11)$$

There are three special cases to be aware of:

1. If $\delta = 0$, the waves are exactly in phase and the waves add together (constructive interference), with amplitude $A = 2y_0$;

2. If $\delta = \pi$, the waves are exactly out of phase and the waves subtract (destructive interference), and the amplitude $A = 0$;
3. If $\delta = \frac{2\pi}{3}$, the resultant wave has exactly the same amplitude as the input waves.

All three of these cases can be derived by substituting the relevant value of δ into Equation (10.11).

10.3 Standing waves

A standing wave is a specific outcome which occurs when a wave is confined to space (for example on a piano string) and it reflects at the boundaries and travels back along its original path. This leads to waves travelling in both directions which combine by superposition. Only certain frequencies can exist in a standing wave, as the superposition leads to a stationary pattern called a **standing wave**.

10.4 Wave function for a standing wave

In order to fully consider a standing wave, we need to derive the form of its wavefunction. To do this, we consider two waves travelling in opposite directions along a string (Equation (10.12)). Because one is a reflection of the other (they each reflect from the boundaries), they will have the same frequency and phase.

$$\begin{aligned} y_1 &= y_0 \sin(kx - \omega t) \\ y_2 &= y_0 \sin(kx + \omega t) \end{aligned} \quad (10.12)$$

The resultant vertical displacement of the string is then the sum of these two waves (Equation (10.13)):

$$\begin{aligned} y &= y_1 + y_2 \\ y &= y_0 \sin(kx - \omega t) + y_0 \sin(kx + \omega t) \end{aligned} \quad (10.13)$$

We can add these directly using a trigonometric identify or we can work in the complex notation:

$$\begin{aligned} y &= y_0 \sin(kx - \omega t) + y_0 \sin(kx + \omega t) \\ &= y_0 e^{i(kx-\omega t)} + y_0 e^{i(kx+\omega t)} \\ &= y_0 e^{ikx} (e^{-i\omega t} + e^{i\omega t}) \end{aligned} \quad (10.14)$$

We can use either an Argand diagram or De Moivre's theorem to show that the term $(e^{-i\omega t} + e^{i\omega t}) = 2 \cos \omega t$, and hence:

$$y = 2y_0 \cos \omega t e^{ikx} \quad (10.15)$$

We can now expand the complex exponent using De Moivre's theorem again and, remembering we are interested in the 'imaginary' component (as this contains the desired sine function), and we obtain the result in Equation (10.16):

$$y = \underbrace{2y_0 \cos \omega t}_{\text{time-dependent amplitude}} \underbrace{\sin kx}_{\text{static wave}} \quad (10.16)$$

This result describes a static wave, $y = \sin kx$ whose amplitude varies in time as $A = 2y_0 \cos \omega t$. Note that it is possible to obtain this result via a trigonometric identity also.

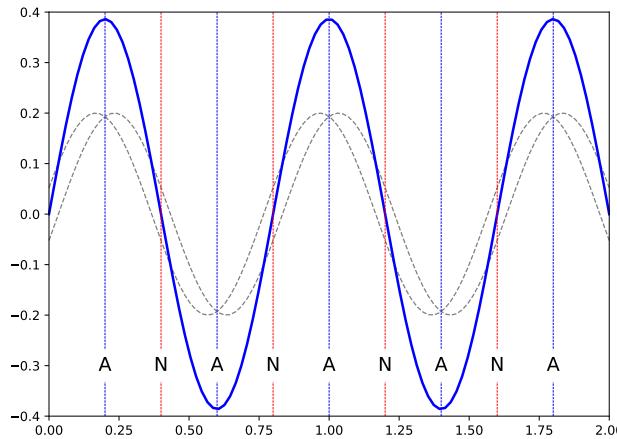


Figure 10.2: A standing wave is the result of two waves equal in frequency and amplitude moving past each other. This forms a static wave whose amplitude varies in time, and has nodes (N) and antinodes (A) present in the waveform. Here we show the fifth harmonic; 5 antinodes and a wavelength $\lambda = 2/5 = 0.8$.

This standing wave is illustrated in Figure 10.2; but we notice that there are boundary conditions enforced; namely that the ends of the string are fixed at a constant, zero displacement; *i.e.*:

- $y = 0$ at $x = 0$, and:
- $y = 0$ at $x = L$ at all times t

...where L is the length of the string. From this we deduce that $\sin kL = 0$ and therefore there are a family of solutions for k and λ , known as harmonics.

- If $\sin kL = 0$, then $kL = 0, \pi, 2\pi, \dots$, or $k_n L = n\pi$
- Additionally, $\lambda_n = \frac{2\pi}{k_n} = \frac{2L}{n}$ where $n = 1, 2, 3, \dots$

In other words, the vibrational modes are **quantised* due to the boundary conditions.

10.5 Waves on strings fixed at both ends

We have described the mathematics of standing waves; let's now apply this to a wave travelling on a string which is fixed at both ends. Figure 10.3 illustrates the fundamental wavelength of the string, which corresponds to twice the length of the string.

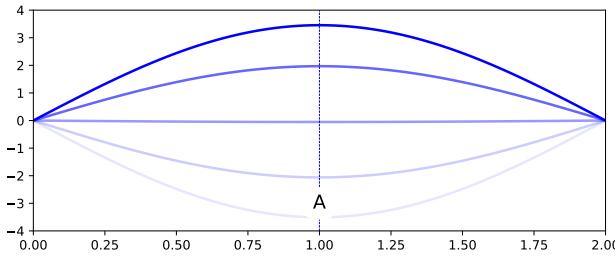


Figure 10.3: The fundamental frequency, or first harmonic. This is one half-wavelength enclosed in the boundaries and has a single antinode.

We can then visualise the harmonics within the standing wave on this fixed string (Figure 10.4)

The modes of vibration (resonances) shown in Figure 10.4 illustrate the occurrence of **nodes** (points which do not move) and **antinodes** (points with the maximum vibration amplitudes). Note also that the end-points of the string must be nodes as well, as these points are fixed.

In general, the n th harmonic will have $\frac{1}{n} \times$ wavelength and $n \times$ the frequency of the fundamental vibration shown in Figure 10.3.

Table 10.1: Showing the variation of wavelength and frequency of each harmonic with respect to the first (fundamental) wave.

Harmonic	λ	f
Fundamental, first	$2L$	f_1
Second	L	$2f_1$
Third	$\frac{2L}{3}$	$3f_1$
Fourth	$\frac{L}{2}$	$4f_1$
Fifth	$\frac{2L}{5}$	$5f_1$
n th	$\frac{2L}{n}$	nf_1

In general, for the n th harmonic:

- Wavelength is given by $\lambda_n = \frac{2L}{n}$
- Frequency is given by $f = nf_1 = \frac{nv}{\lambda_1} = \frac{nv}{2L}$

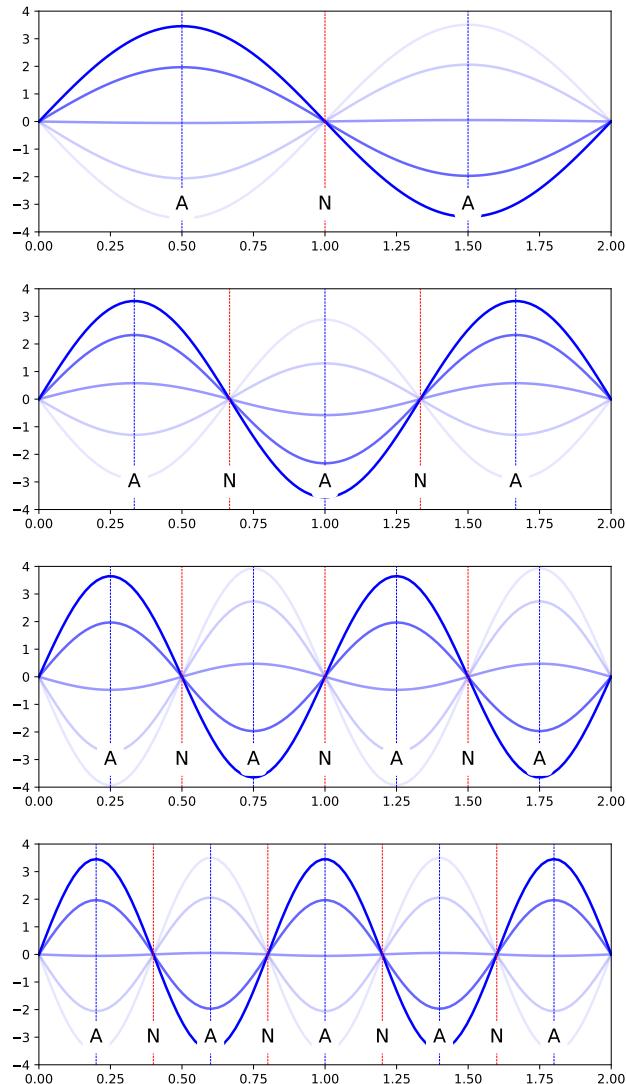


Figure 10.4: The second, third, fourth and fifth harmonics of a standing wave on a string. Notice that the n th harmonic has $\frac{n}{2}$ wavelengths contained in the space.

...where v is the phase velocity (the speed of propagation of the wave along the string).

The resonant frequencies, or harmonics, of the string are known as its natural frequencies. Any string will resonate with maximum amplitude when excited with these frequencies, and this set of harmonics are known as a harmonic series. The actual harmonics heard when the string is excited will depend on the manner of its excitation; *e.g.* a string plucked at its centre will *only* display the odd harmonics; *i.e.* those with an anti-node in the centre.

In stringed instruments (violin, piano, guitar etc.) the vibration of the string is amplified by a mechanical resonator; a soundboard in the case of the piano, or resonant cavities for a guitar or violin. These resonators must be carefully designed to resonate equally well over a wide range of frequencies.

10.6 Organ pipes and other wind instruments

In contrast to a vibrating string, wind instruments rely on a resonance within a column of air. We can model these pipes as a simple pipe, resonating at its natural frequencies when air is blown into (or across) an opening at one end. The resonant behaviour will differ depending on whether the other end of the pipe is open or closed. We will consider each of these cases in turn.

10.6.1 Pipes open at both ends

In this model, the column of air is able to vibrate at its ends, so we have a similar set of harmonics as for a string, but with displacement antinodes at its ends (the air can vibrate freely at the ends of the tube). There is a second set of nodes/antinodes corresponding to the pressure; these do not align with the displacement nodes/antinodes; rather a **pressure node** aligns with a **displacement antinode** and vice versa. (if an air molecule does not move, we have a displacement node, but it is continuously ‘squashed’ from both sides by the oscillating air molecules, so experiences the biggest pressure change).

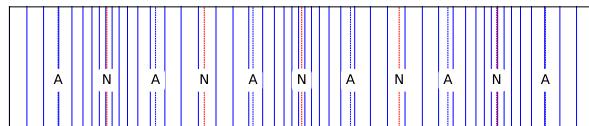


Figure 10.5: This image shows the node/antinode structure of a standing wave in a **closed tube**. Note the presence of a displacement node at either end where molecules are compressed against the end of the tube

We observe **all** harmonics in this system; there are no concerns about ‘position of plucking’ that there is for the string. The displacement of the air molecules

extends a little beyond the ends of the tube, so the effective length is given by $L_{\text{eff}} = L + \Delta L$, where ΔL is a small end correction. Therefore:

$$\lambda_n = \frac{2L_{\text{eff}}}{n} \quad \text{and} \quad f_n = \frac{nv}{2L_{\text{eff}}} \quad (10.17)$$

10.7 Pipe closed at one end

We now have a different situation with different boundary conditions:

- There must be a displacement node at the closed end
- There must be a displacement antinode at the open end

This now gives us a fundamental wavelength four times longer than the effective length of the tube (the shortest node-antinode separation is a quarter wavelength).

This means that we only observe the **odd** harmonics (the even harmonics would not allow the boundary conditions for this tube).

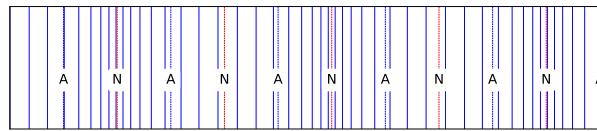


Figure 10.6: For a tube open at one end, the standing waves now have a node at the closed end, and an antinode at the open end. This changes the available harmonics within the tube.

$$\lambda_n = \frac{4L_{\text{eff}}}{n} \quad n = 1, 3, 5, \dots \quad (10.18)$$

$$f_n = \frac{nv}{4L_{\text{eff}}} \quad n = 1, 3, 5, \dots \quad (10.19)$$

Chapter 11

Fourier analysis

Any vibrating string such as those in stringed musical instruments contain a range of different harmonics all drawn from the harmonic series. The exact combination of harmonics will vary depending on the way the string is set in motion; *e.g.* a violin string is bowed, while a piano string is struck. The resulting sound wave will be a complicated periodic function, and not obviously a sine wave. The fundamental theorem of Fourier analysis is that **any** periodic function can be constructed by a linear superposition of sine and cosine waves. A corollary to this is that any periodic waveform can therefore be *constructed* using a linear superposition of sine and cosine waves. The set of sines and cosines used to represent a function in Fourier analysis is referred to as a *Fourier series*, and the set of constants in the series are known as *Fourier coefficients*.

The standard form for a Fourier series representation of a function $P(t)$ with period T is:

$$\begin{aligned} P(t) = & \frac{1}{2}a_0 + a_1 \cos \omega t + b_1 \sin \omega t \\ & + a_2 \cos 2\omega t + b_2 \sin 2\omega t \\ & + a_3 \cos 3\omega t + b_3 \sin 3\omega t \end{aligned} \tag{11.1}$$

...or, to express as a summation:

$$P(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t) \tag{11.2}$$

...where $\omega = \frac{2\pi}{T}$

The presence of both sine and cosine terms allows the relative phase of each component to be controlled:

- If $P(t)$ is an **even function**, only **cosine** terms are required;
- Similarly, only **sine** terms are required if the function is **odd**

The Fourier coefficients a_0, a_1, a_2 etc. are given by:

$$a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} P(t) \cos n\omega t \, dt \quad n = 0, 1, 2, 3 \dots \quad (11.3)$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} P(t) \sin n\omega t \, dt \quad n = 1, 2, 3 \dots \quad (11.4)$$

It can be seen that $\frac{a_0}{2}$ is equal to the average value of $P(t)$ over one cycle.

11.1 Example: the square wave

Consider a square wave defined on the interval $[-1, 1]$ by:

$$P(t) = \begin{cases} 1 & -\frac{1}{2} \leq t \leq \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases} \quad (11.5)$$

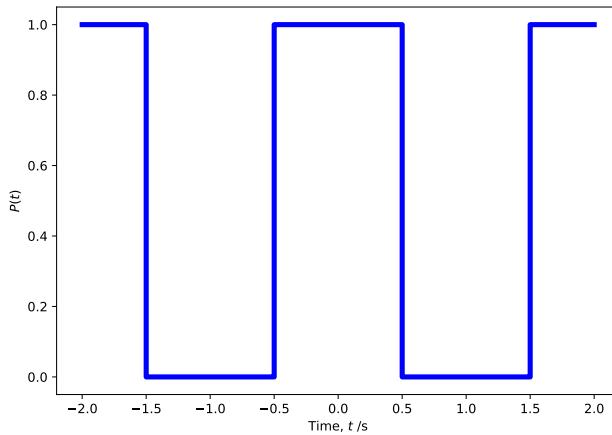


Figure 11.1: The form of a square wave; often termed a 'top hat' function. This is interpreted in electronics as a digital signal which is either on or off.

The period of this function is 2 s, so that $\omega = \pi$.

Since $P(t)$ is an even function (*i.e.* it has symmetry about the y axis), it follows from symmetry that $b_1 = b_2 = b_3 = \dots = 0$. On the other hand, the cosine coefficients are given by:

$$\begin{aligned}
 a_0 &= \int_{-1}^1 P(t) \cos 0 dt = \int_{-\frac{1}{2}}^{\frac{1}{2}} dt = 1 \\
 a_n &= \int_{-1}^1 P(t) \cos n\pi t dt \\
 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P(t) \cos n\pi t dt \\
 &= \frac{2}{n\pi} \sin\left(\frac{n\pi}{2}\right) \quad n = 1, 2, 3, \dots
 \end{aligned} \tag{11.6}$$

Thus, we find the following:

- $a_0 = 1$
- $a_1 = \frac{2}{\pi}$
- $a_2 = 0$
- $a_3 = -\frac{2}{3\pi}$
- $a_4 = 0$
- $a_5 = \frac{2}{5\pi}$
- $a_6 = 0$
- ...

...and the Fourier series is given by:

$$P(t) = \frac{1}{2} + \frac{2}{\pi} \left(\cos \pi t - \frac{\cos 3\pi t}{3} + \frac{\cos 5\pi t}{5} - \dots \right) \tag{11.7}$$

As we increase the number of terms in the series, we see our wave comes closer and closer to the square wave we sketched earlier.

A square wave can be thought of as the most extreme deviation from a sinusoidal function; yet we have demonstrated that we can reconstruct a square wave from a linear superposition of sinusoidal wave forms. Using these principles, we can build any waveform from a linear superposition of sinusoidal waves.

11.2 The harmonic makeup of any wave

We have shown using the example of the square wave that any wave can be made of the sum of harmonic sinusoidal components, each with a characteristic amplitude. Once again, let's consider the example of musical instruments. When two different wind instruments play the same note, (concert A, 440 Hz), they naturally sound different; the oboe sounds distinctly different to a clarinet, however they are, ostensibly, playing a note with the same frequency. The reason they sound different is because the wave which comes out of each has a fundamentally different form; the **tone quality**.

The wave-forms for the oboe and clarinet are shown in Figure 11.3, where, despite having the same fundamental frequency, the waveform is considerably

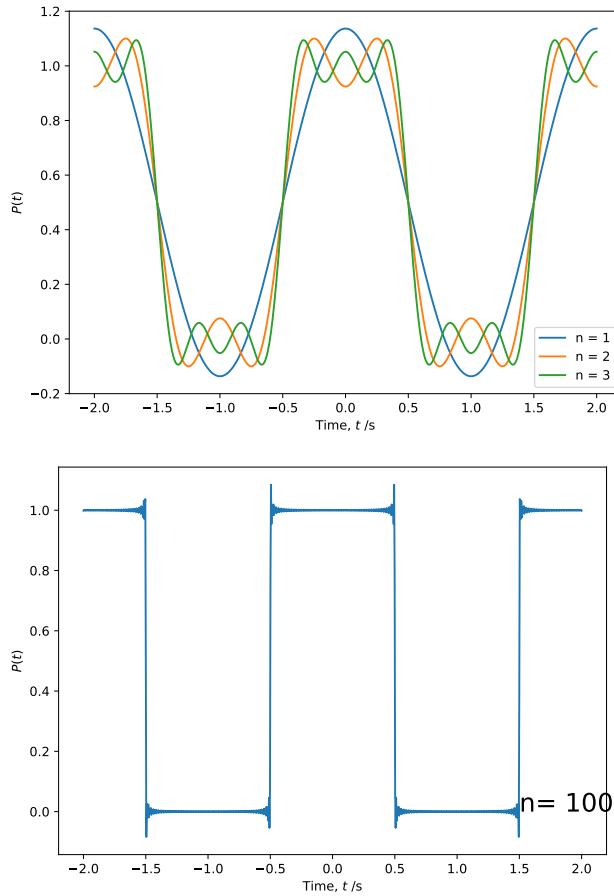


Figure 11.2: As we add increasing numbers of terms to the Fourier summation, our collection of sinusoidal waves converges on the form of the 'top hat' function.

more complex than that of a tuning fork (a ‘nearly pure’ wave form with no additional harmonics).

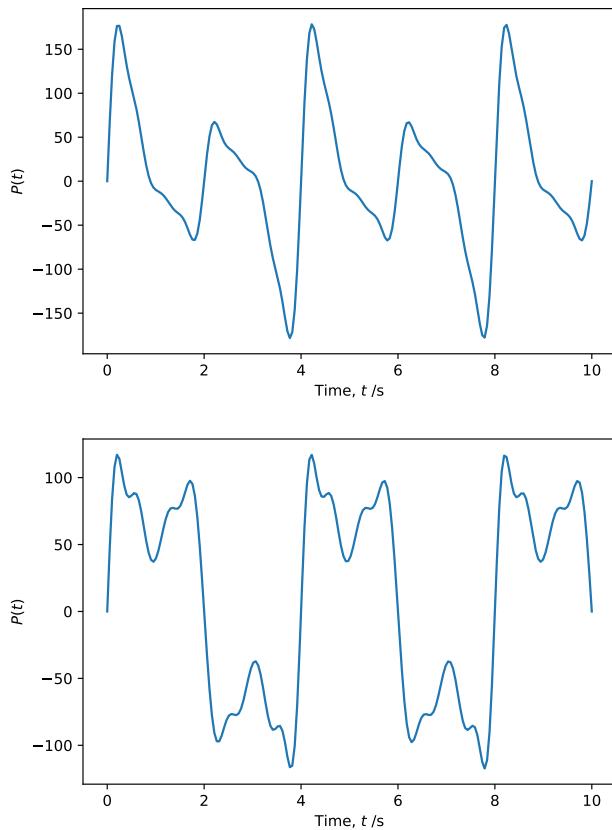


Figure 11.3: We have an illustration of the waveforms for an oboe (top) and clarinet (bottom); these are both playing the same frequency, but we see a very different harmonic makeup.

These waveforms can be analysed in terms of the relative contributions from each of the harmonics in the harmonic series, and we can see in Figure ?? that the contributions from each of the harmonics in the harmonic series is quite different between the two. The shapes of the tubes for each of the instruments leads to different contributions from the harmonics in each instrument, and this shapes the distinctive sound we hear. The real trick of course is identifying what this harmonic makeup is from the presented waveform. This is the essence of Fourier analysis; identifying the harmonic makeup of a given waveform, and has applications way beyond musical instruments!

11.3 Fourier Transforms

A Fourier transform is a mathematical tool which allows for the identification of the component frequencies and corresponding amplitude of a given waveform. It is a tool which is used in almost area of instrumental analysis, and its importance cannot be understated. The principles of the Fourier transform are as follows:

- If a period of oscillation is allowed to become infinite (*i.e.* the wave becomes non-periodic), we can calculate the spectral density (the amplitude of each contributing harmonic wave) of any non-periodic function *e.g.* a wave pulse.
- In this limit the Fourier coefficients become continuous functions *i.e.* all frequencies are allowed;
- The (infinite) array of Fourier coefficients is referred to as the **Fourier Transform** of the function
- Thus, we can show that **any** arbitrary wave pulse can be produced from a superposition of harmonic waves, which is why we attach such importance to understanding their behaviour.

11.4 Some important Fourier transforms

11.4.1 A cosine wave

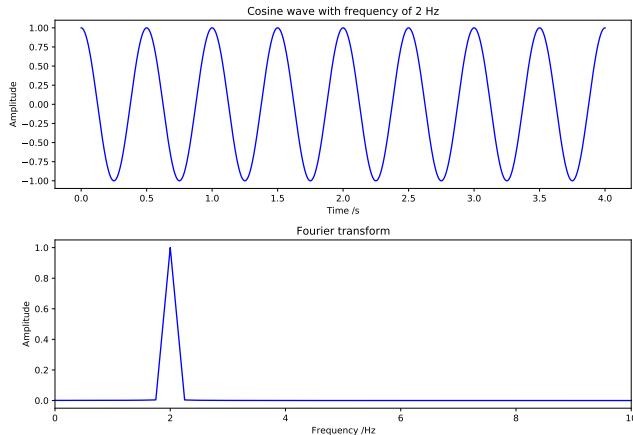


Figure 11.4: A periodic cosine function (top) made up of a single frequency gives a single 'spike' in the Fourier transform (bottom) at the appropriate frequency. Note that this Fourier transform has finite width because it is a genuine 'fast Fourier transform' performed in Python using data from a cosine plot; so is not the idealised 'infinitesimally narrow delta function' which it would otherwise be.

A pure cosine wave has a single frequency; the Fourier transform yields a single spike (a single Fourier coefficient) at this frequency. Strictly, there is also a spike

at $-\omega_0$ because the transform is symmetric.

11.4.2 The ‘top hat’ function

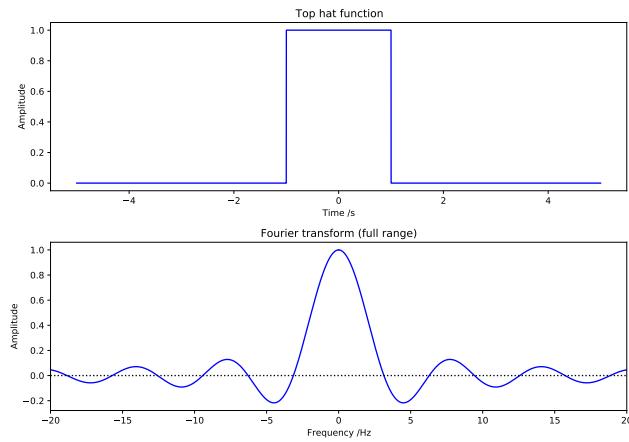


Figure 11.5: A non-periodic ‘top hat’ function is made up of an infinite Fourier series, involving all frequencies. We see this reflected in the Fourier transform (bottom) where the transform function extends to infinity.

A top-hat function is a wavepulse for a square wave, defined as a pulse width of τ . The Fourier transform yields a continuous series of Fourier coefficients, with defined positions along the frequency axis.

11.4.3 A Gaussian function

For a Gaussian function with full-width at half-maximum (FWHM) τ , its Fourier transform yields another Gaussian function, with FWHM = $\frac{2\pi}{\tau}$. A Gaussian function is a good approximation to a laser pulse; the shorter the laser pulse, the more frequencies are required to generate it. This has ramifications on the colour of the pulse! (this is a subject for another course!)

11.5 The Bandwidth Theorem

Tipler and Mosca, 16.3

The relationship between the widths of the two Gaussian functions shown in Section 11.4.3 is important and has far-reaching consequences. If the Gaussian function represents a wave pulse with width (in time) Δt , the Fourier transform tells us it contains a spread of frequencies, $\Delta\omega$. However, the product $\Delta t\Delta\omega$ is a constant:

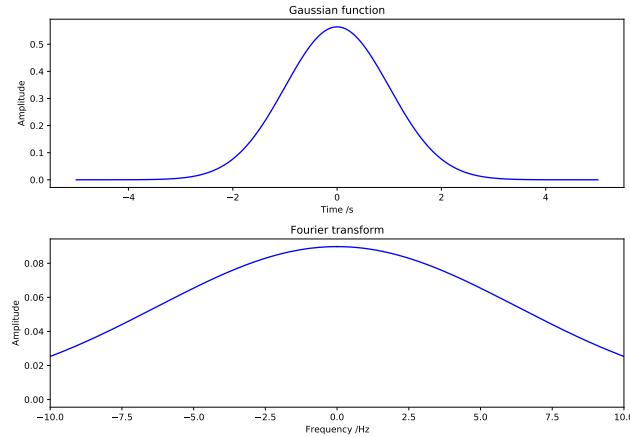


Figure 11.6: The Fourier transform of a finite Gaussian function is another Gaussian function, but the width is a reciprocal of the original function.

$$\Delta t \Delta \omega = 2\pi \quad (11.8)$$

This is known as the “Bandwidth theorem” and, if we examine the spatial width of the pulse (Δx rather than Δt) we see a similar relationship:

$$\Delta x \Delta k = 2\pi \quad (11.9)$$

For a more general pulse shape, the determination of the width of the pulse becomes more arbitrary, and so the bandwidth theorem becomes:

$$\Delta t \Delta \omega \approx 2\pi \quad \text{and} \quad \Delta x \Delta k \approx 2\pi \quad (11.10)$$

There are a number of implications of the bandwidth theorem:

- A **narrow** wave pulse will contain a **wide** range of frequencies;
- A **broad** wave pulse will contain a **narrow** range of frequencies;
- $\Delta \omega = 0$ implies that $\Delta t = \infty$; *i.e.* an infinite sine wave with a single frequency - a perfect wave;
- Similarly, $\Delta k = 0$ implies $\Delta x = \infty$.

Chapter 12

Wave packets and Dispersion

Textbook link: Tipler and Mosca, Section 16.3

A **wave packet** is exactly what it sounds like; a “parcel” containing a group of waves all of different frequencies. How these wave packets travel will vary depending on the properties of the medium through which they travel. If all the frequency components of the wave packet (or wave pulse) travel with the same phase velocity, v_p , the resulting disturbance propagates without changing shape and the medium is known as a **non-dispersive medium**. If, however, frequency components travel at different speeds through the medium, we have a **dispersive** medium. You will have likely observed the effects of a dispersive medium; if you have ever stood at a quiet railway platform (usually underground), you can hear the tell-tale sound of an approaching train from the “pinging” sound in the rails long before you hear the actual vehicle itself as the high-frequencies travel faster through the rail than the lower frequencies.

The velocity of propagation of the packet is known as the **group velocity** v_g , the velocity at which energy is carried through the medium. In a non-dispersive medium, this is equal to the phase velocity v_p . If the medium is **dispersive** however, different frequencies travel with different values of v_p , and the wavepacket spreads as it travels; this is the origin of the ‘pinging’ sound we hear in railway tracks.

12.1 Two waves, same amplitude, different frequency (Beats)

Let’s consider two waves with the same amplitude but differing in frequency (*i.e.* amplitude y_0 is common to both, each has independent frequency ω_1 and ω_2).

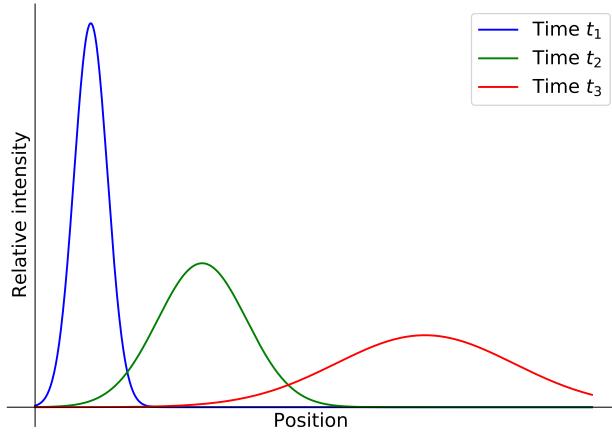


Figure 12.1: As a wave packet travels through a dispersive medium in which higher frequencies travel faster than lower frequencies, the wave packet 'spreads out' in space.

For convenience we will assume that the phase difference between the two is zero at time $t = 0$, and we will then consider their displacements at an arbitrary x coordinate (e.g. $x = 0$).

Firstly, let's lay out the mathematical description of each wave:

$$y_1 = y_0 \sin \omega_1 t \quad \text{and} \quad y_2 = y_0 \sin \omega_2 t \quad (12.1)$$

We can now obtain an expression for the overall observed amplitude, $y = y_1 + y_2$, and use the standard trigonometric identities¹ to combine the two equations:

$$\begin{aligned} y &= y_1 + y_2 \\ &= y_0 \sin \omega_1 t + y_0 \sin \omega_2 t \\ &= 2y_0 \cos\left(\frac{\omega_1 - \omega_2}{2}t\right) \sin\left(\frac{\omega_1 + \omega_2}{2}t\right) \end{aligned} \quad (12.2)$$

As an aside, we can achieve the same result by using complex exponentials where the sum of the two waves is given in Equation (12.1), where we are interested in the imaginary component since we started with sine waves:

$$y = y_0 e^{i\omega_1 t} + y_0 e^{i\omega_2 t} \quad (12.3)$$

These can be added on an Argand diagram to show the result in Equation (12.4)

¹Here we use the sum-to-product identity $\sin A + \sin B = 2 \cos\left(\frac{A-B}{2}\right) \sin\left(\frac{A+B}{2}\right)$

$$y = y_0 r e^{i\theta} \quad (12.4)$$

...where:

$$r = 2 \cos \left(\frac{\omega_1 t - \omega_2 t}{2} \right) \quad \text{and} \quad \theta = \frac{\omega_1 t + \omega_2 t}{2} \quad (12.5)$$

Our end result if we expand this complex representation is:

$$y = 2y_0 \cos \left(\frac{\omega_1 t - \omega_2 t}{2} \right) e^{i(\frac{\omega_1 t + \omega_2 t}{2})} \quad (12.6)$$

As we said above, we are interested in the ‘imaginary’ (sine) component, and we end up with the same result as shown in Equation (12.2). We can simplify the expression as in Equation (??).

$$y = \underbrace{2y_0 \cos \left(\frac{\Delta\omega}{2} t \right)}_{\text{slowly varying amplitude}} \underbrace{\sin \omega_{av} t}_{\text{wave with average frequency}} \quad (12.7)$$

...where $\Delta\omega = \omega_1 - \omega_2$.

The frequency f_0 of the resulting wave is the average of the two input waves, while the amplitude oscillates with frequency $\frac{\Delta\omega}{2}$. This is known as **beating**. The actual frequency of the beats themselves is twice this, as shown in Figure 12.2.

Therefore, if the frequency of the slowly varying amplitude is $\frac{\Delta\omega}{2}$, the observed frequency of the ‘beats’ will be twice this; *i.e.* the difference in frequency between the two sources.

Repeating the analysis above but with the full expressions for the two sine waves (*i.e.* $y_n = y_0 \sin(kx - \omega t)$), we obtain the expression in Equation (12.8)

$$y = 2y_0 \cos \left(\frac{\Delta kx - \Delta\omega t}{2} \right) \cdot \sin(k_{av}x - \omega_{av}t) \quad (12.8)$$

From Equation (12.8) we find there are two velocities to consider; the phase velocity of the “average” wave, and the phase velocity of the “envelope” (the waveform which describes the amplitude; see Figure 12.2). These are laid out below:

The phase velocity of the “average” wave:

$$v_{av} = \frac{\omega_{av}}{k_{av}} \quad (12.9)$$

The phase velocity of the envelope:

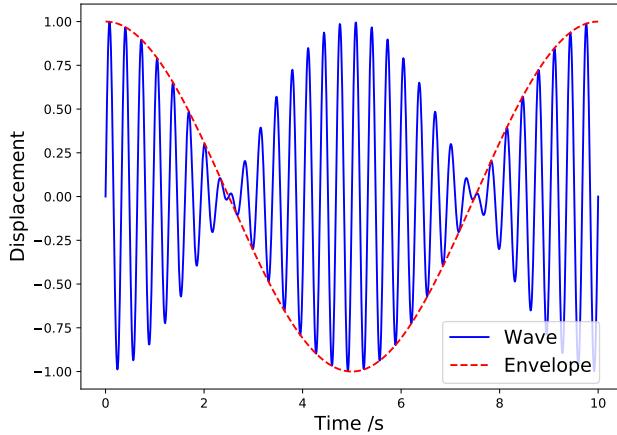


Figure 12.2: As two waves interfere with each other ($f_1 = 3.0$ Hz; $f_2 = 3.2$ Hz), we see the interference structure as a waveform of frequency 3.1 Hz with a slowly varying amplitude at frequency 0.1 Hz (the ‘envelope’). This visualises the ‘beats’ in the interference structure.

$$v_{env} = \frac{\Delta\omega}{\Delta k} \quad (12.10)$$

In a *non-dispersive* medium (*i.e.* all frequencies travel with the same phase velocity), it can be shown that $v_{av} = v_{env} = \frac{\omega_1}{k_1} = \frac{\omega_2}{k_2}$ = the phase velocity of the medium.

However, in a dispersive medium, the velocities are not equal, *i.e.* $v_{av} \neq v_{env}$, and the “envelope” propagates at a different speed to the individual components. We associate v_{env} with the group velocity v_g and in this case would be represented as:

$$v_g = \frac{\Delta\omega}{\Delta k} \quad (12.11)$$

12.2 Dispersion

As has already been mentioned, in a **dispersive** medium, v_p is different for every frequency component, and the group velocity is not equal to the initial phase velocity v_p ; *i.e.* $v_g \neq v_p$. For any dispersive medium, we can write a relationship between ω and k so that ω varies as a function of k ; *i.e.* $\omega = \omega(k)$. This is called the **dispersion relation** and depends on the physics of the particular wave phenomena being observed.

In the general case, the group velocity is given by the derivative of the dispersion relation (Equation (12.12)). The proof of this relation is available in some advanced textbooks, but it is not necessary for this course.

$$v_g \equiv \frac{\partial w(k)}{\partial k} \quad (12.12)$$

For a non-dispersive medium, ω is directly proportional to k (Equation (12.13))

$$w(k) = v_p k \quad \text{with } v_p = \text{constant} \quad (12.13)$$

The consequence of this is that, when the function is differentiated according to the principle in Equation (12.12), we find that $v_g = v_p$ (Equation (12.14)) as has been mentioned previously.

$$v_g \equiv \frac{\partial w(k)}{\partial k} = \frac{d(v_p k)}{dk} = v_p \quad (12.14)$$

For a dispersive medium however, v_p becomes a function of k . If we consider a simple case where ω has a linear dependence on k (Equation (12.15)) with constant terms a and b :

$$\omega = ak + b \quad (12.15)$$

As before, the phase velocity is still defined as $v_p = \frac{\omega}{k} = a + \frac{b}{k}$, while the group velocity is found by differentiation (Equation (12.16)):

$$v_g = \frac{\partial w(k)}{\partial k} = a \quad (12.16)$$

We see that these expressions for the phase velocity $v_p = a + \frac{b}{k}$ and the group velocity $v_g = a$ are different; therefore the envelope of the wave packet will move at a different speed from the phase velocity of the wave. However, the group velocity is still the same for all wavenumbers k .

When ω is a more general function of k , the group and phase velocities will each depend differently on k . The dependence of the group velocity on k means that envelope doesn't move at a single velocity and its components of different wavenumber (or wavelength) have different velocities, distorting the envelope.

So-called *normal dispersion* occurs when the group velocity is lower than the phase velocity ($v_g < v_p$); this is the most common situation, however it is possible to have *anomalous dispersion* in which the group velocity, and hence the energy, travels faster than the individual waves.

Examples of dispersion include:

- Splitting of light by prism
- Formation of a rainbow
- Phonons propagating through a crystalline solid
- Spreading of light pulses in fibre-optic cables. The dispersion limits the maximum length of cable before signal reconditioning is needed.

12.3 Water waves

The dispersion relation for water is given in Equation (12.17), which includes the acceleration due to gravity (g) and the depth of the water (h):

$$\omega^2 = gk \tanh(kh) \quad (12.17)$$

For shallow water and/or long wavelengths, $kh \ll 1$ and $\tanh(kh) \simeq kh$. This allows Equation (12.17) to be simplified to:

$$\omega^2 \simeq gk \cdot kh \quad \text{and} \quad \omega \simeq k\sqrt{gh} \quad (12.18)$$

We can then determine the phase velocity p in this situation as:

$$v_p = \frac{\omega}{k} = \sqrt{gh} \quad (12.19)$$

...and there is **no dispersion** (the phase velocity has no variable dependence). This applies in the case of the tsunami, where the water depth is small compared to the wavelength, even when crossing oceans. This means that these waves are able to propagate over global distances without significantly changing shape.

For intermediate depths, we might use the expansion of $\tanh(x)$ to write this term as:

$$\tanh(kh) = kh - \frac{(kh)^3}{3} + \dots \quad (12.20)$$

...from which we obtain:

$$\omega^2 \simeq gk \cdot \left(kh - \frac{(kh)^3}{3} \right) \quad (12.21)$$

We can now use this to determine the phase velocity of the wavepacket in water (Equation (12.22)):

$$v_p = \frac{\omega}{k} = \frac{\sqrt{gh}}{\sqrt{1 + \frac{(kh)^2}{3}}} \quad (12.22)$$

Equation (12.22)) tells us that the dispersion of the wave increases with increasing water depth relative to the wavelength.

12.4 Sound waves in a crystal (phonons)

When we consider sound-waves in a crystal, we can consider both longitudinal sound waves as well as transverse sound waves (as we are not in a fluid, a transverse wave is now possible due to the intermolecular forces within the crystal).

We will first consider the one-dimensional case, as the three dimensional case can be derived from the principles laid out here. Firstly, let's consider what a wave might look like when propagating along a one-dimensional crystal. We imagine our one-dimensional lattice to be a line of molecules, all connected by a "spring" (to model the molecular bond). In this model, we can either have a longitudinal wave, where the springs are stretched and compressed as the molecules oscillate in the direction of the crystal axis, or we can have a transverse wave, in which the 1-D line of molecules behaves as a string. These are illustrated in Figure 12.3.

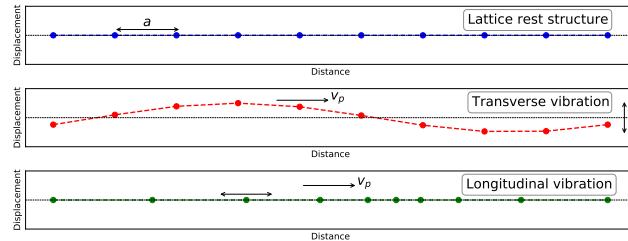


Figure 12.3: A 1 dimensional lattice can carry sound waves (phonons, also known as lattice vibrations) via both transverse and longitudinal vibrational modes.

From this model, it can be shown that the phase velocity can be given by the relationship shown in Equation (12.23), relating the tension in the 1-D lattice (F), the linear density of the 1-D lattice (μ) and the periodicity of the lattice, a .

$$v_p = \frac{\omega}{k} = \frac{F \sin\left(\frac{ka}{2}\right)}{\mu \frac{ka}{2}} \quad (12.23)$$

In the long-wavelength limit, $k \rightarrow 0$ and $v_p \rightarrow \frac{F}{\mu}$.

Additionally, we can rewrite Equation (12.23) to gain an expression for ω :

$$\omega = \frac{kF}{\mu} \frac{\sin\left(\frac{ka}{2}\right)}{\frac{ka}{2}} = \frac{2F}{a\mu} \sin\left(\frac{ka}{2}\right) \quad (12.24)$$

If we plot the angular frequency ω against the wavenumber k , we obtain the **dispersion curve**, shown in Figure 12.4.

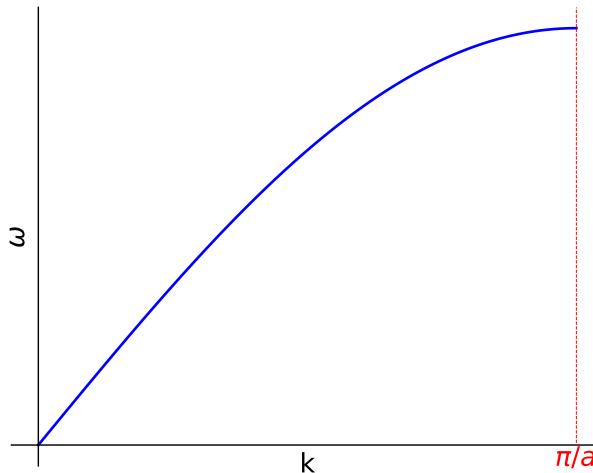


Figure 12.4: The variation in angular frequency ω with wavenumber k in a dispersive medium. There is an upper limit to the wavenumber which the medium can sustain, namely π/a

From Equation (12.24), we can find the group velocity v_g by differentiation:

$$v_g = \frac{d\omega}{dk} = \frac{F}{\mu} \cos\left(\frac{ka}{2}\right) \quad (12.25)$$

Once again, examination of Equation (12.25) shows:

- The group velocity, v_g tends to $\frac{F}{\mu}$ as $k \rightarrow 0$; and:
- The group velocity also drops to zero when $k = \frac{\pi}{a}$

At the point when $k = \frac{\pi}{a}$, alternate atoms move in exact anti-phase, and we have the appearance of a standing wave (Figure 12.5). Therefore there is an effective cut-off at $k = \frac{\pi}{a}$; waves with a greater wave number than this (shorter wavelength, higher frequency) cannot be supported by the lattice.²

Note that the periodicity of the lattice, a (also known as the molecular spacing) is typically around $10^{-9} - 10^{-10}$ m, so this effect will only be observed for **very** high frequencies.

²This is a similar effect to the aliasing observed in sample signals in Experiment S

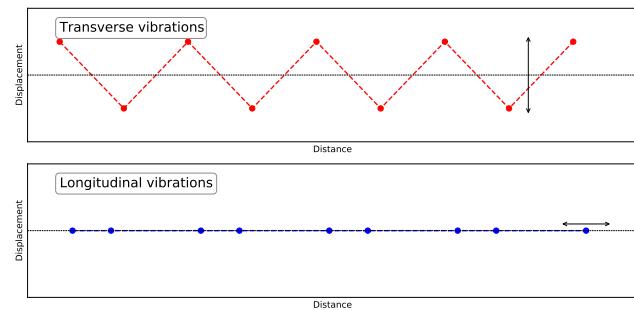


Figure 12.5: When the wavenumber $k = \pi/a$, the atoms move in perfect anti-phase and the movement appears as a standing wave. The lattice can no longer support a higher wavenumber than this.

Chapter 13

Interference and Diffraction

13.1 Conditions for interference

In Section 10.2, we discussed the principle of superposition of two waves with the same frequency and wavelength in the context of forming a standing wave. Specifically, we found that the resultant wave has an amplitude which depends on the phase difference.

Mathematically, for two waves described by y_1 and y_2 , we obtain the result shown below (from Equation (10.10)):

$$y_1 = A_0 \sin(kr - \omega t) \quad \text{and} \quad y_2 = A_0 \sin(kr - \omega t + \delta) \quad (13.1)$$

$$y = \underbrace{2A_0 \cos\left(\frac{\delta}{2}\right)}_{\text{New amplitude}} \underbrace{\sin\left(kr - \omega t + \frac{\delta}{2}\right)}_{\text{Travelling wave}} \quad (13.2)$$

Therefore the amplitude A of the resultant wave may be described by:

$$A = 2A_0 \cos\left(\frac{\delta}{2}\right) \quad (13.3)$$

This amplitude A is a maximum when the waves are *in phase* ($\delta = 2m\pi$), and at a minimum when the waves are *out of phase* ($\delta = (2m + 1)\pi$).

Typically, a phase difference between identical waves results from a path difference; for example two light waves travelling by different routes. For such a path difference, Δr , the phase difference is given by $\delta = \frac{2\pi\Delta r}{\lambda}$. In the case of **constructive** interference, $\delta = 2m\pi$ and $\Delta r = m\lambda$, while for **destructive** interference $\delta = (2m + 1)\pi$ and $\Delta r = (m + \frac{1}{2})\lambda$.

For any two waves to interfere, they must be **coherent**; *i.e.* there must be a fixed phase relationship between the two sources. This is usually achieved by splitting a single light source. Light sources such as tungsten filament bulbs are **not** coherent; the emission results from thermal excitation and the different components of the bulb radiate independently of one another. A LASER on the other hand **is** coherent; the beam is produced by stimulated emission, a process which only occurs if all photons are in phase.

13.2 Propagation of Light - Huygens's Principle

Huygens' Principle is summarised as follows:

Each point on a primary wavefront serves as a source of spherical secondary wavelets, which advance with a speed and frequency equal to those of the primary wave. The primary wavefront at some later time is the envelope of these wavelets.

While the propagation of light is in fact governed by Maxwell's equations, Huygens' construction provides an extremely valuable way of analysing diffraction problems. An illustration of Huygens' principle is shown in Figure 13.1; we only consider the *forward moving* portion of the wavelet, not the whole spherical wavelet.

13.3 The two-slit interference pattern

Tippler and Mosca, 33.3

Consider a double slit experiment, where uniform coherent waves approach a pair of slits in a barrier (Figure 13.2). According to Huygens' Principle, each slit then becomes a source of wavelets, which then radiate outwards from this point and, since the wavelets from each source are coherent, they then interfere.

If we examine the path of waves travelling in an arbitrary direction from each slit which converge on the observation screen (Figure #ref(fig:ch13-doubleslit2)), we can start to consider the behaviour of our wavelets.

If we assume that the distance to the screen L is considerably larger than the distance between the slits d then, since $L \gg d$, then we can say that the angles θ_1 , θ_2 and θ are approximately equal (*i.e.* $\theta_1 \approx \theta_2 \approx \theta$); the Fraunhofer diffraction condition.

Note that when $L \simeq d$, this approximation no longer applies and the diffraction becomes very complex, and is termed **Fresnel conditions**. For the purposes of this discussion we will only consider the Fraunhofer case (*i.e.* $L \gg d$).

We can obtain an expression for the path difference between the waves; if we are assuming that the distance to the screen is so large the rays may be considered parallel, then the path difference may be determined from Figure 13.3:

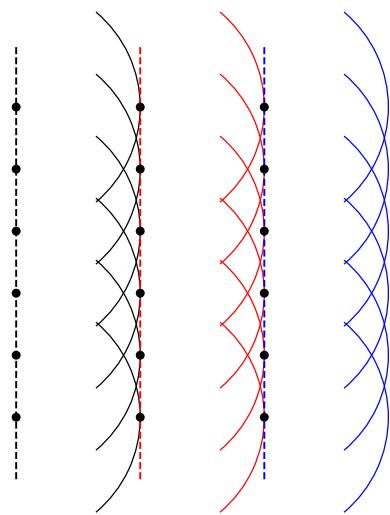


Figure 13.1: The Huygens' wavelet construction. A wavefront is considered to be a source of secondary wavelets; each subsequent wavefront is the envelope of the forward moving component of the wavelets.

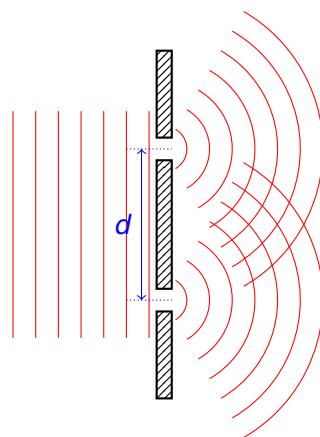


Figure 13.2: The configuration of the double slit experiment. Two narrow slits are separated at distance d , each of which becomes a source of secondary wavelets which can then create the interference pattern.

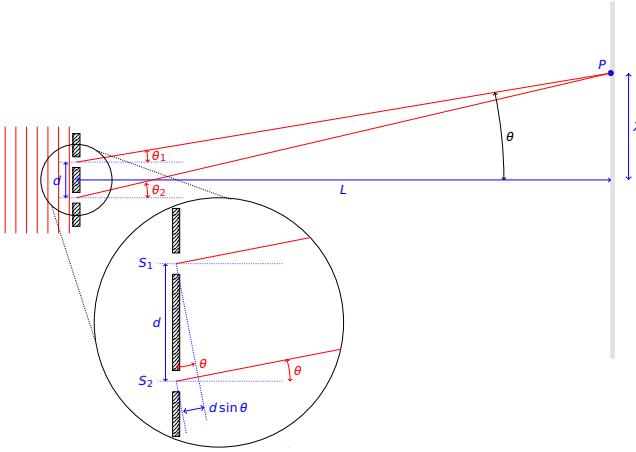


Figure 13.3: Looking at the whole double slit experiment, we can determine the angles within the experiment and define the observation position X on the screen as a function of these angles. Under the Fraunhofer condition (long field interference), the rays can be considered parallel, allowing the path difference and hence phase offset to be determined.

$$\begin{aligned} d \sin \theta &= m\lambda ; & m &= 0, 1, 2, \dots \text{ for maxima} \\ d \sin \theta &= \left(m + \frac{1}{2}\right)\lambda ; & m &= 0, 1, 2, \dots \text{ for minima} \end{aligned} \quad (13.4)$$

At point P on the observation screen we can then determine the phase difference, δ :

$$\delta = \frac{2\pi}{\lambda} d \sin \theta \quad (13.5)$$

However, remember that, as $L \gg d$, then $\sin \theta \approx \tan \theta = \frac{X}{L}$ for small θ . Therefore, at points where interference maxima are observed (*i.e.* $d \sin \theta = m\lambda$):

$$d \sin \theta = \frac{dX_m}{L} = m\lambda \quad \text{and} \quad X_m = \frac{m\lambda L}{d} \quad (13.6)$$

Equation (13.6) tells us that the distance of the “ m ”th maximum from the midpoint of the diffraction pattern scales only with the value of m ; all other values being constant for the particular experimental setup; *i.e.* the maxima will be **equally spaced**.

We can also consider the intensity of the fringes; from our work in Section 7.4.4, the intensity (the transmitted power) is proportional to the square of the amplitude. Using the result shown in Equation (13.3), we can derive the expression in Equation (13.7):

$$I = A^2 = 4A_0^2 \cos^2\left(\frac{\delta}{2}\right) \quad (13.7)$$

However, since we have the expression $\delta = \frac{2\pi d}{\lambda} \sin \theta \approx \frac{2\pi d X}{\lambda L}$ the phase difference δ is proportional to the distance of the fringe from the midpoint of the interference pattern, X . Therefore, since $\delta \propto X$, we expect to see fringes following the intensity of a \cos^2 function (Figure 13.4).

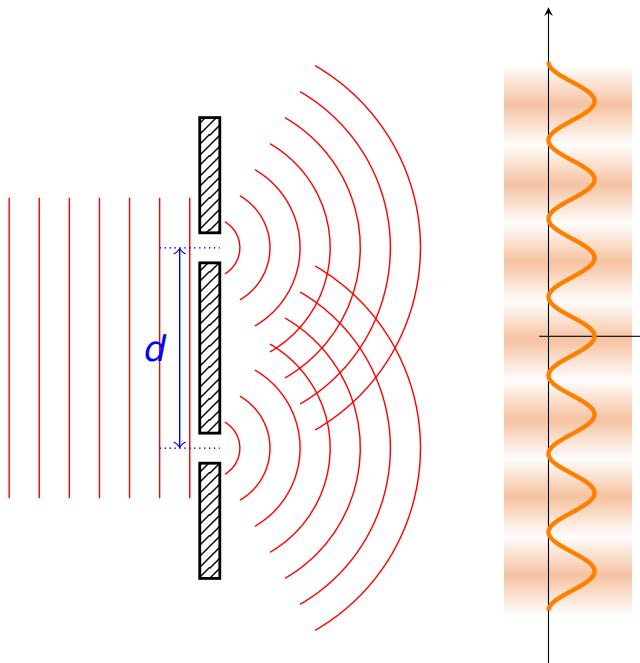


Figure 13.4: The interference pattern from two narrow slits gives an intensity pattern proportional to a \cos^2 function.

13.4 Diffraction from a single wide slit

Let's now turn to the effect of coherent plane wavefronts incident on a single wide slit (Figure 13.5).

In this situation, we imagine dividing the aperture into a set of secondary sources and use the Huygens' construction to consider secondary wavelets. These 'diffracted waves' will appear to spread. To find the diffracted intensity in a particular direction, we add together the secondary waves travelling in that direction, taking into account the path (and hence phase) differences.

We will conduct our analysis in two stages:

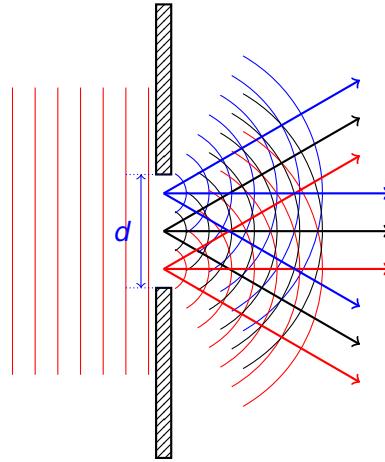


Figure 13.5: A single wide slit of width d can be thought of as a source of many different wavelets; here we have shown three rays to give an indication of the complexity of the interference pattern.

1. We consider the superposition of waves from two secondary sources, symmetrically disposed about the centre of the slit, similar to the “two narrow slit” example above;
2. We then add together the contributions from all pairs of sources across the wide slit.

The workthrough for this example is slightly involved, so is shown elsewhere, however we show the elements of the solution here and, as for the two narrow slits example above, use it to determine the amplitude of the superposition as a function of X , the distance on screen.

The result of the derivation is that the superposition observed as a function of X is a travelling wave with amplitude as:

$$A \propto \text{sinc} \left(\frac{\pi a X}{\lambda L} \right) \quad (13.8)$$

...where $\text{sinc}(x) = \frac{\sin(x)}{x}$.¹ Therefore the intensity ($\propto A^2$) observed on the screen is given by:

$$I \propto A^2 \propto \text{sinc}^2 \left(\frac{\pi a X}{\lambda L} \right) \quad (13.9)$$

The positions of the dark fringes can then be determined by setting this equal to zero:

¹The “sinc” function is used often in signal processing, and its name comes from its name the ‘sine cardinal’ function, and is widely used in Fourier transforms.

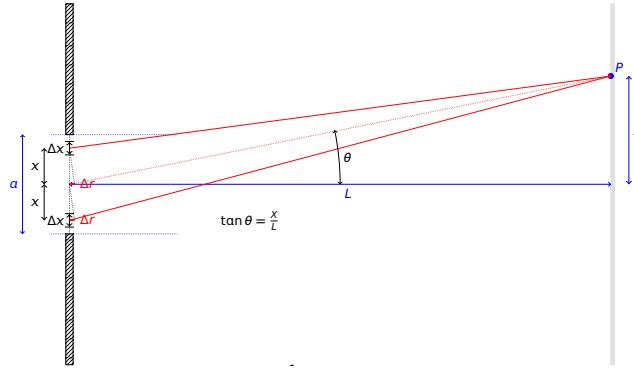


Figure 13.6: Construction of the diffraction pattern from a single wide slit involves looking at pairs of wavelets Δx ; through quantifying the phase difference between these two wavelets and summing over the slit we can reconstruct the diffraction pattern.

$$\begin{aligned}
 \text{sinc}^2\left(\frac{\pi a X}{\lambda L}\right) &= 0 \\
 \equiv \frac{\sin^2\left(\frac{\pi a X}{\lambda L}\right)}{\sin^2\left(\frac{\pi a X}{\lambda L}\right)} &= 0 \\
 \frac{\pi a X}{\lambda L} &= \pi, 2\pi, 3\pi, \dots \equiv m\pi \\
 \frac{a X}{\lambda L} &= m, \text{ where } m \neq 0
 \end{aligned} \tag{13.10}$$

From Equation (13.10), we can determine that our dark fringes occur when the output of the sinc function in Equation (13.9) is equal to zero; *i.e.* dark fringes occur at $X_m = \frac{m\lambda L}{a}$ where $m = 1, 2, 3, \dots$, but $m \neq 0$. Therefore the separation of the dark fringes, X_m is inversely proportional to the slit width a , *i.e.* $X_m \propto \frac{1}{a}$.

13.5 Diffraction from two wide slits

Once again, we have a series of coherent planar wavefronts approaching the barrier, this time with two wide slits present (Figure 13.7).

We can use the same method as was used previously, but with different limits of integration, and we obtain a travelling wave with amplitude as given in Equation (13.11):

$$A = aA_0 \cos\left(\frac{\pi b X}{\lambda L}\right) \text{sinc}\left(\frac{\pi a X}{\lambda L}\right) \tag{13.11}$$

The corresponding intensity observed of the fringes observed on the screen is then as shown in Equation (13.12):

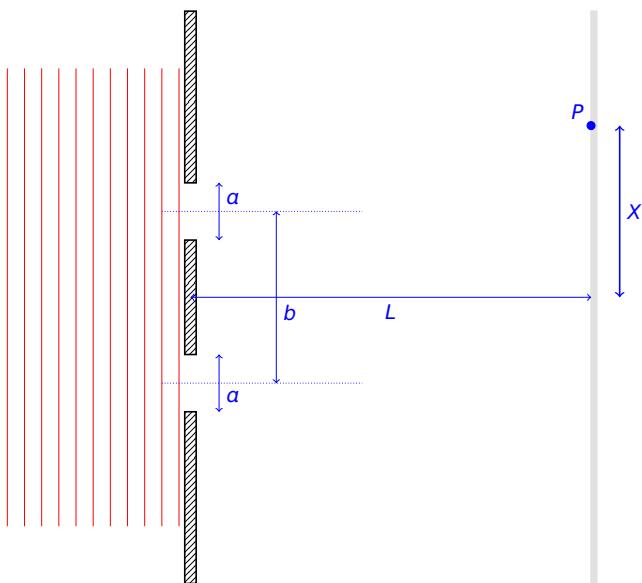


Figure 13.7: The configuration for two wide slits of width a at separation b appears similar to the construction for narrow slits, except now each slit is a source of a number of wavelets, rather than a single wavelet. This has profound implications on the diffraction pattern observed.

$$I \propto A^2 = a^2 A_0^2 \cos^2 \left(\frac{\pi bX}{\lambda L} \right) \operatorname{sinc}^2 \left(\frac{\pi aX}{\lambda L} \right) \quad (13.12)$$

In other words, the intensity is the product of the diffraction pattern from the two “narrow” slits with that from the single wide slit:

$$I \propto A^2 = \underbrace{a^2 A_0^2}_{\text{Constant scaling factors}} \underbrace{\cos^2 \left(\frac{\pi bX}{\lambda L} \right)}_{\text{Double slit intensity}} \underbrace{\operatorname{sinc}^2 \left(\frac{\pi aX}{\lambda L} \right)}_{\text{Wide slit intensity}} \quad (13.13)$$

When we have two functions combined in this manner, we say that one is the ‘envelope’ of the other - it effectively “contains” and controls the intensity of the other. In general, the lower frequency function is the ‘envelope’, so here the sinc^2 function is the envelope of the \cos^2 function. The intensities of this combined function are shown in Figure 13.8

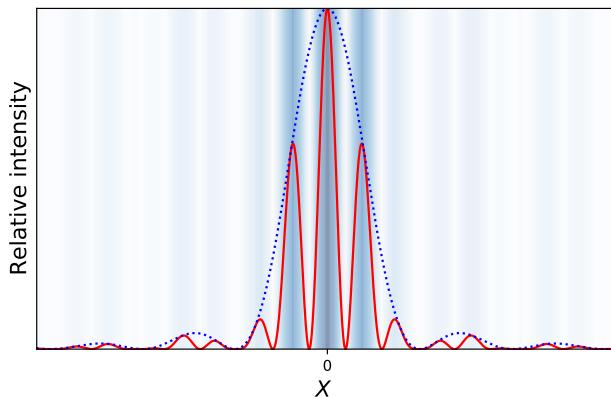


Figure 13.8: The diffraction pattern from two wide slits is a combination of the result of a single wide slit and that of two narrow slits. The sinc^2 function becomes the envelope shaping the intensities of the peaks from the narrow slit solution.

13.6 Multiple slits with equal spacing

Over the past few sections we have laid out the principles of diffraction and interference of light from a pair of narrow slits, a single wide slit, and a pair of wide slits. We have shown how the relative intensity functions are combined and how this gives us our intensity patterns for the combination of the two more simple examples. We can apply these principles to determine the expected interference patterns we would observe for multiple slit systems.

13.7 Fourier transforms in diffraction

The method we have used to determine the intensities of fringes in diffraction/interference patterns may be generalised for any aperture by introducing a general aperture function $f(x)$ which describes the shape of the aperture (*e.g.* if we have two narrow slits, we use two delta functions to describe them). We can then sum the phase components from all elements Δx across this aperture. The amplitude of the constant wave is then given by Equation (13.14)

$$A = \sum_{\text{aperture}} f(x) e^{i\delta(x)} \Delta x \quad (13.14)$$

Equation (13.14) is telling us that we can scale our wave (described by the complex representation) originating from aperture element Δx by the aperture function $f(x)$, and find out the overall result by summing up all these components.

If we substitute $\delta(x) = 2\pi(\frac{x}{\lambda}) \sin \theta$ and let $\Delta x \rightarrow dx$, we can then replace the summation with the integral shown in Equation (13.15):

$$A = \int_{-\infty}^{\infty} f(x) e^{i2\pi(\frac{x}{\lambda}) \sin \theta} dx \quad (13.15)$$

This integral is an example of a Fourier transform *i.e.* the *amplitude* of a (far-field, Fraunhofer) diffraction pattern is given by the Fourier transform of the aperture. The diffracted intensity is then given by the square of this Fourier transform.

Example In Section 11.4.2 we demonstrated a ‘top-hat’ function, which could represent a broad slit. The Fourier transform of this gave us the diffracted amplitude which, when squared, gives us the expected intensity pattern derived above.

13.8 Resolution limit of optical instruments

Textbook link Tipler and Mosca, 33.7

Light from a point source will be diffracted by a circular aperture (*e.g. the objective lens of a telescope, microscope or even your eye) to give a pattern similar to that of a single wide slit, but with circular symmetry. This is illustrated in Figure 13.9.

Knowing that this intensity is based on the sinc^2 function described in Equation (13.9), we can determine the angular position of the first dark fringe (first zero point) is given by the angular position:

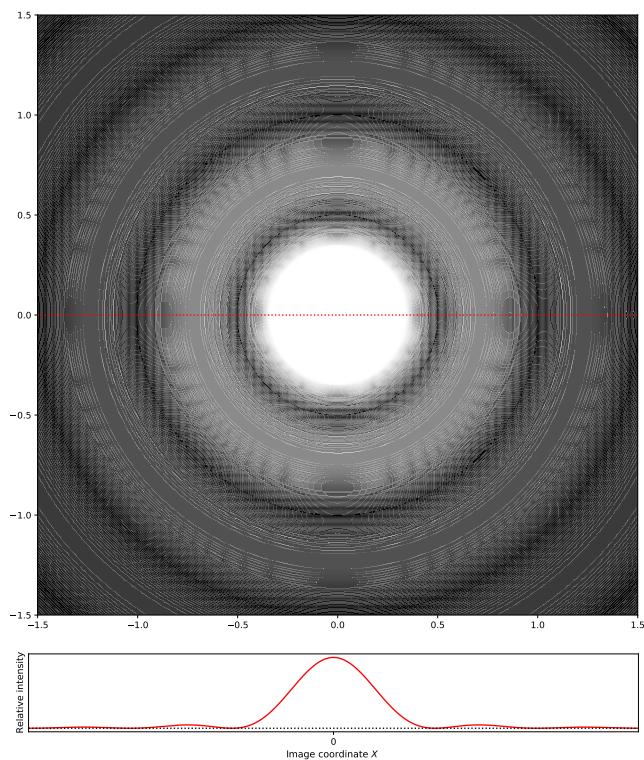


Figure 13.9: A simulation of the diffraction pattern of a point source through an aperture. The defined disc in the center is known as the Airy disc image. The intensity across the diffraction pattern is as we have seen for the one dimensional slit, derived from the sinc function.

$$\sin \theta = \frac{1.22\lambda}{D} \quad (13.16)$$

...where D is the diameter of the aperture.

This is the end result for a single incoherent source passing through an aperture; however what happens if we consider more than one incoherent source? At what point is it impossible to determine the number of sources? It is this which determines the limit of optical resolution.

Let's now consider two incoherent point sources with an angular separation α at the aperture.

The sources are at the limit of resolution ("just resolved") if the maximum of the diffraction pattern of one point source (caused by the aperture) corresponds to the first minimum of the diffraction pattern of the other; in other words; we see a 'bright spot' where we would expect to see darkness. In terms of the angular separation α , the two sources are just resolved if:

$$\alpha = \frac{1.22\lambda}{D} \quad (13.17)$$

(note the use of the small angle approximation here)

This is simulated in the intensity profile shown in Figure 13.10.

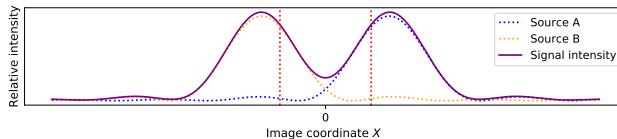


Figure 13.10: The limit of optical resolution occurs if the maximum of one source aligns with the minimum of the second; this creates a 'dip' between the observed signal intensity allowing the sources to be separately resolved. Any closer however and these will merge into a single bright peak in the middle.

Figure 13.11 shows two point sources which are clearly resolved ($\alpha > \frac{1.22\lambda}{D}$), while Figure 13.12 shows two point sources which are just at the limit of resolution ($\alpha = \frac{1.22\lambda}{D}$)

It is easier to visualise this limit by plotting a graph of the expected intensities. We can resolve the two sources provided there is a clear "dip" in intensity of the resulting intensity sum between the peaks of the two interference patterns, however if they are close enough together such that the result has no dip in the middle we can no longer resolve the sources.

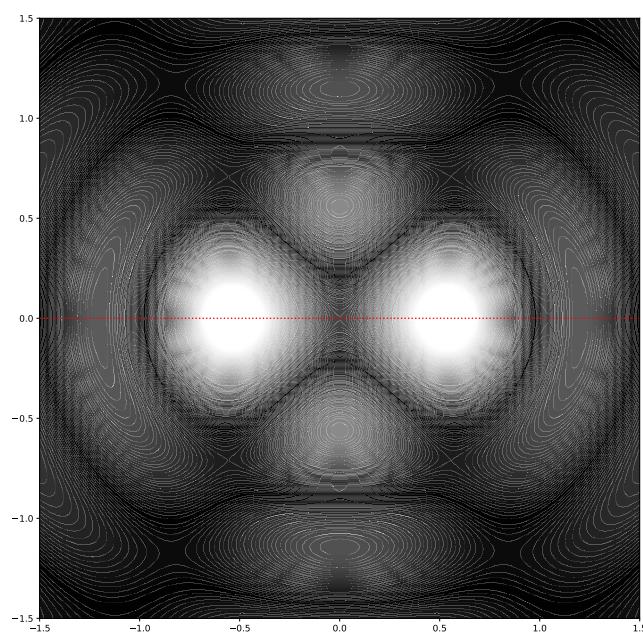


Figure 13.11: Sources can be clearly defined up to a certain limit; but diffraction patterns cause interference around the image plane, but each source still has its defined Airy disc.

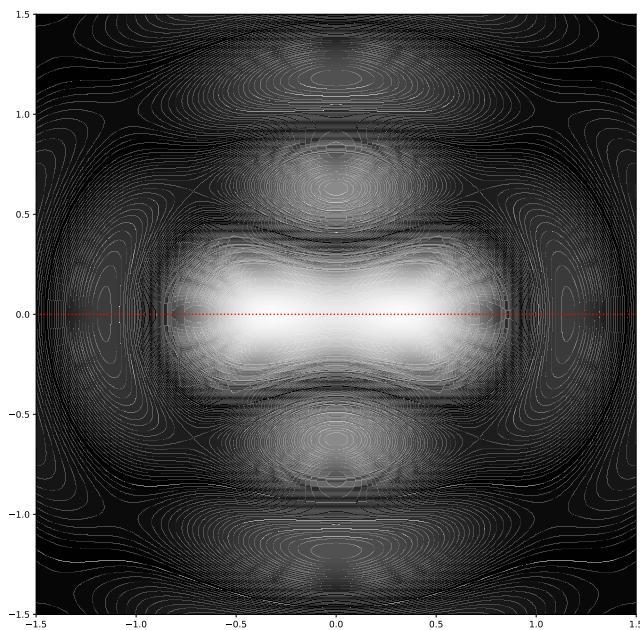


Figure 13.12: At the resolution limit, a 'dimming' can still be perceived between the sources. At closer separation however, the two Airy discs will blend into one bright source.

Chapter 14

Properties of Light

The consideration of the wave nature of light offers a way to explain part of its behaviour; it is worth therefore giving this some consideration.

14.1 Reflection and Refraction

Refraction is a phenomenon arising from the differing speeds at which light travels in different media. The refractive index n of a material is a way to describe this variation and may be determined by the ratio of the speed of light in a vacuum (c) to its speed v in the medium concerned:

$$n = \frac{c}{v} \quad (14.1)$$

Refractive indices are however approximate; different wavelengths of light travel at different speeds in a given material - we covered this phenomenon in Section 12.2. Refractive indices do not account for these differences, so values reported are usually at a standardised wavelength (usually that of sodium d-line emission, $\lambda = 589$ nm).

Table 14.1: A list of refractive indices of common materials

Material	$n_{589\text{nm}}$
vacuum	1
air	1.003
water	1.33
common glass	1.52
diamond	2.42

It is worth noting that the **frequency** of light remains unchanged while passing through different media; rather it is the **wavelength** which varies:

$$\lambda_n = \frac{\lambda}{n} \quad (14.2)$$

When a beam of light encounters an interface between two transparent media of different refractive indices, part of the light is reflected from the interface, and part of the beam is refracted through the interface (Figure 14.1).

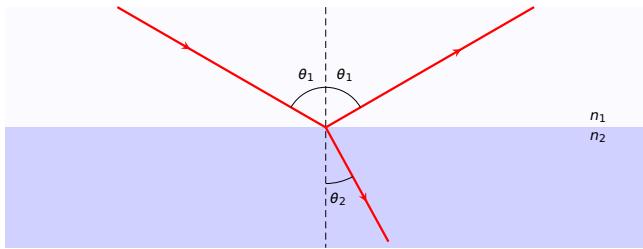


Figure 14.1: The path of a light ray passing into a medium of higher refractive index n_2 . The angles of incidence and reflection are equal(θ_1), while the angle of refraction (θ_2) is determined from the ratio of the refractive indices via Snell's law.

From Figure 14.1, we note the following:

- The reflected beam has an angle of reflection equal to the angle of incidence
- The angle of incidence θ_1 is the angle between the incident beam and the normal to the surface
- The angle of reflection θ'_1 is the angle between the reflected beam and the normal to the surface
- The angle of refraction θ_2 is the angle between the refracted beam and the same normal.

The paths that the light rays take through either reflection or refraction are described by **Fermat's principle**, namely that the light travels along the path of least time; however due to the variation of the speed of light, this path is not necessarily the shortest path between the two points. Let's first consider the case of reflection (Figure 14.2).

Consider a light ray travelling from A to B by reflection from the interface at point P . The distance APB is at a minimum when the angles θ_1 and θ'_1 are equal, where the light ray reflects at point P_{\min} .

For a light ray travelling from A to C by refraction through the interface at point P , a similar principle may be used. Since the light travels more slowly in the bottom material, there is a balance to be found in minimising the time taken to travel from A to C . This is akin to a lifeguard racing to save a swimmer

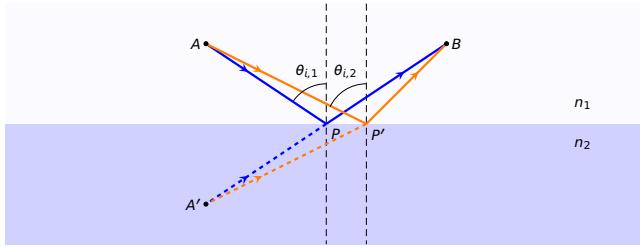


Figure 14.2: Fermat’s principle states that a light ray takes the route (whether refraction or reflection) which is the shortest time. When the light ray remains in the same medium through reflection, the shortest path APB is one where the line $A'PB$ is straight; *i.e.* the angle of reflection is equal to the angle of incidence.

in difficulty; the lifeguard can run faster over sand than swimming through the sea, but the quickest path to the swimmer is not a direct line.

Through following Fermat’s principle, we arrive at **Snell’s law of refraction** (Equation (14.3)), allowing us to calculate the angles of reflection and refraction of a light ray passing from one material into another.

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (14.3)$$

We can also consider the intensity of light reflected from an interface; this depends on:

1. the angle of incidence of the light ray;
2. the polarisation of the light; and
3. the refractive indices of the two media.

When light is perpendicular to the interface (as is often the case for optical instruments), the reflectivity R is defined as the ratio between the intensities of the reflected light I and the incident light I_0 :

$$R = \frac{I}{I_0} \quad \text{or} \quad I = RI_0 \quad (14.4)$$

...where:

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 \quad (14.5)$$

Similarly, there is also a **transmission coefficient**, defined as $T = 1 - R$, which defines the proportion of light transmitted through the interface. Complete reflection/transmission behaviour is described by Fresnel’s equations.

14.2 Total internal reflection

When a beam of light passes through an interface from a larger to a smaller refractive index ($n_1 > n_2$), the angle of refraction θ_2 is larger than the angle of incidence θ_1 . As θ_1 reaches a particular value (the ‘critical angle’, θ_c), $\theta_2 = 90^\circ$, and the refracted beam disappears. At this point the beam is reflected from the interface back into the first medium.

The conditions for the critical angle are found from Snell’s law (Equation (14.3)), setting $\theta_1 = \theta_c$ and $\theta_2 = 90^\circ$:

$$\sin \theta_c = \frac{n_2}{n_1} \sin(90^\circ) = \frac{n_2}{n_1} \quad (14.6)$$

14.2.1 Evanescence

A full solution of the wave equations for light undergoing total internal reflection reveals a solution within the low refractive index region. This is a consequence of the continuity equations *i.e.* the wave amplitude must be continuous across the interface.

The solution to these equations is a decaying exponential, known as an **evanescent “wave”** (Figure 14.3)

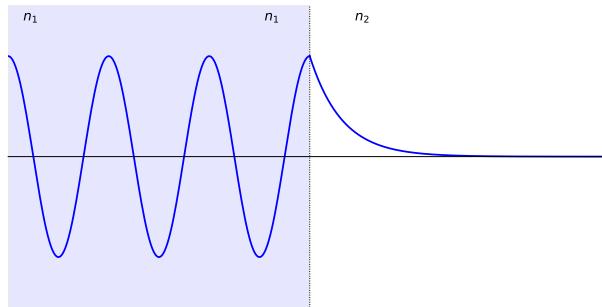


Figure 14.3: The intensity of an evanescent ‘wave’ decays exponentially with distance from the surface of the interface on the same approximate scale as the wavelength. It can be thought of as a ‘spreading of energy’ rather than an actual wave.

The decay length of this evanescent “wave” is approximately equal to the wavelength of the incident beam, λ . The reason for terming this a “wave” is that it does not propagate - *i.e.* it is not a travelling wave. The evanescent “wave” stores energy and, if a second glass block is placed “close” to the first (*i.e.* within one micron), the energy may be taken from the evanescent wave and a travelling wave produced in the second block (albeit with a smaller amplitude to the incident wave).

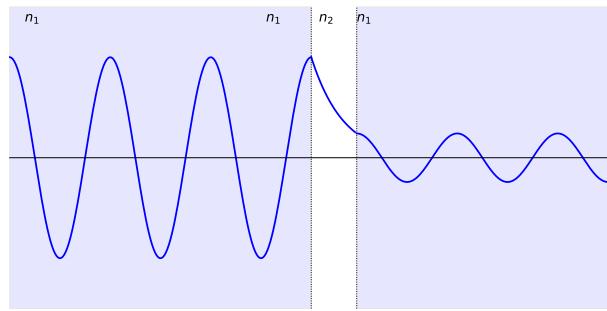


Figure 14.4: If a medium of sufficiently high refractive index is brought close to the surface where total internal reflection is taking place, energy can be extracted from the evanescent ‘wave’; this is the basis of fingerprint scanners.

This example of extracting energy from an evanescent wave (Figure 14.4) is known as ‘frustrated total internal reflection’; the photons have ‘tunneled’ across the air gap.¹

14.2.2 Applications of total internal reflection

Total internal reflection has many applications; we mention some here, however the list is not exhaustive!

Optical fibres: A light beam which is coupled to a very thin glass fibre is almost parallel to the axis of the fibre. In this condition light will hit the fibre walls with an angle larger than the critical angle resulting in total internal reflection and very high transmission.

Touch screen technology and fingerprint acquisition: Light is coupled into a thin glass screen. As a fingertip touches the screen it scatters the evanescent wave above the screen and can be detected.

¹We often use the term “tunneling” to describe quantum phenomena passing through a barrier; you will meet the term frequently!

Chapter 15

Mirrors

We have introduced the properties of light as it interacts with different materials. In the next two sections we will now use this to show how simple optical elements work and how to determine factors describing how light rays behave when interacting with these simple optical elements. IN this chapter, we discuss **mirrors**.

15.1 Plane mirror: object and image

Consider an object placed at point P , shown in Figure 15.1; light rays coming from this object are reflected by a mirror to reach the observer.

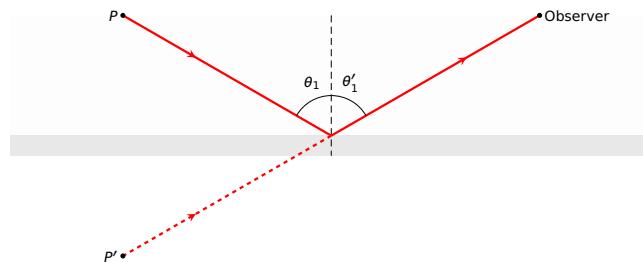


Figure 15.1: Reflection from a plane mirror occurs when the angle of incidence from P is equal to the angle of reflection to the observer. To the observer, the light appears to come from point P' behind the mirror; a *virtual image* which cannot be projected onto a screen.

As shown in Figure 15.1 the observer perceives the light rays as though they are coming from point P' ; we call this point the **image** of the object P . This image is called a **virtual** image because the rays coming from the object do not converge onto P' . When rays from the object converge on the image, the

image is called **real**. A real image can be observed on a screen or recorded on a film/CCD sensor without the use of lenses.

15.1.1 Construction of a ray diagram for a plane mirror

To locate the the image of an object produced by a plane mirror, we follow a series of steps to draw the ray diagram:

1. From a point on the object, trace a ray perpendicular to the mirror. This ray will be reflected along the same direction;
2. Trace a second ray from the same point on the object and forming an angle with the normal to the mirror;
3. Extend these two rays behind the mirror;
4. They will meet in one point that is the virtual image of the object point.

15.2 A spherical mirror

We follow a similar process as we did for the plane mirror (Section 15.1) in considering the nature of the image. Once again, we place an object at point P and consider how the light rays from this object are reflected from a concave mirror (Figure 15.2).

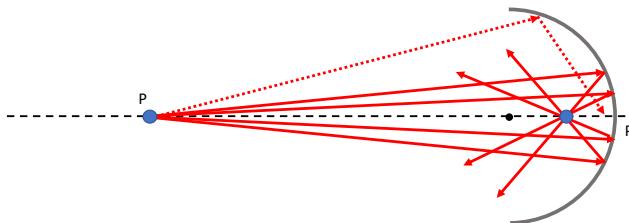


Figure 15.2: Rays coming from point P are focussed at point P' provided they do not deviate too much from the optical axis (paraxial rays); Wide deviations result in imperfect image formation (spherical aberrations).

This time, rays coming from point P , located on the optical axis of the concave mirror, are reflected by the mirror and focused to a point P' (a ‘real’ image, as the rays converge). Any rays which are far from the optical axis (non-paraxial) are not focused on P' ; a term known as *spherical aberration*.

15.2.1 The Mirror equation

The ‘mirror equation’ is a tool to calculate the position of an image of a spherical concave mirror relative to the object position and the centre point of the mirror. Figure 15.3 shows all the relevant distances and angles which are needed for considering the mirror equation.

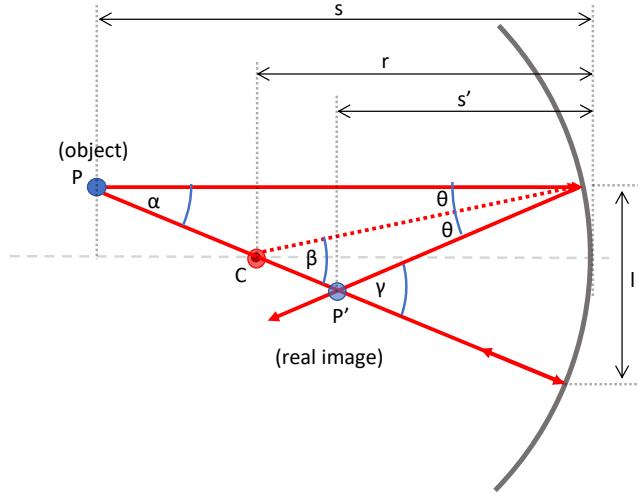


Figure 15.3: Building the mirror equation requires the measurements s , s' , r and l , as well as the angles α , β , γ and θ .

We can use geometric relations in Figure 15.3 to determine relationships between the labelled angles:

$$\begin{aligned}\beta &= \alpha + \theta \\ \gamma &= \alpha + 2\theta \\ \alpha + \gamma &= 2\beta\end{aligned}\tag{15.1}$$

We are now going to make a few assumptions regarding the angles. Assuming that all rays are paraxial and that the object is sufficiently far away that the condition $\sin X \approx X$:

$$\alpha \approx \frac{l}{s} \quad \beta \approx \frac{l}{r} \quad \gamma \approx \frac{l}{s'}\tag{15.2}$$

Substituting these approximations into the expression $\alpha + \gamma = 2\beta$, we get the **mirror equation**:

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{r}\tag{15.3}$$

...and, as $s \rightarrow \infty$ (*i.e.* we have parallel rays), then:

$$s' = \frac{r}{2} = f\tag{15.4}$$

...where f is the focal length of the mirror. This is the distance of the focal point F from the optical surface of the mirror along the optical axis. Light from distant objects forms an image at $s' = f$; i.e. the image of a distant field is formed in the focal plane, perpendicular to the optical axis and intersects it at F .

An alternate representation of the mirror equation therefore becomes:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \quad (15.5)$$

15.2.2 Constructing a ray diagram for a spherical mirror

As before with the plane mirror, there is a procedure to construct a ray diagram for a spherical mirror:

1. A ray from the object at P parallel to the mirror axis will be reflected towards the focal point of the mirror, F ;
2. The focal point F is on the mirror axis, half-way between the centre of the mirror and the mirror surface;
3. The ray from the object at P and passing through the mirror centre will be reflected back on itself;
4. The point at which the reflected rays intersect will define the position of the image of an object at point P .

When we follow these steps, we find differing results depending on where point P is relative to the mirror itself:

1. When P is beyond the centre of the mirror, C , the image formed is **real, inverted** and ***smaller*** than the object;
2. When P lies between C and F , the image is **real, inverted** and **larger** than the object;
3. When P is between F and the optical surface of the mirror, the image is **virtual, erect** and **larger** than the object.

15.2.3 Considering a convex spherical mirror

All of the principles we introduced for a concave spherical mirror apply for a convex mirror, only now the rays need to be projected back through the optical surface to the focal point and centre of the mirror. This time however, now that the object at point P lies past the optical surface, the rays appear to diverge from a point behind the mirror, forming a virtual image. We use the same rays as before to find the image as **virtual, erect** and **smaller** than the object.

15.2.4 Lateral magnification

Spherical mirrors can be considered to have magnification properties; anyone seeing personal vanity mirrors (these used to be known as ‘shaving’ or ‘makeup’

mirrors); or even a spoon - will have seen the magnification effect of such mirrors. This magnification ratio is found from the ratio of the distance of the image from the optical axis and the object from the optical axis:

$$m = \frac{y'}{y} = -\frac{s'}{s} \quad (15.6)$$

If the image is inverted, then the magnification is negative.

It is useful to consider the **sign conventions** when using these equations:

1. s is:

- **positive** when the object (real) is in front of the mirror;
- **negative** when the object (virtual) is behind the mirror;

2. s' is:

- **positive** when the image (real) is in front of the mirror;
- **negative** when the image (virtual) is behind the mirror;

3. r, f are:

- **positive** when the centre/focus are in front of the mirror (concave);
- **negative** when the centre/focus is behind the mirror (convex).

Chapter 16

Lenses

Textbook link: Tipler and Mosca, Section 15.1

Having examined the principles of reflection in respect of curved mirrors, we now look at the case of a curved interface between two optically transparent materials. We introduced the principle of refraction in Section 14.1, where a light ray will slow down on entering a more optically dense medium (higher refractive index) and its path will change as described by Snell's law of refraction (Equation (14.3)). With a plane interface, the behaviour is straightforward, however now that we have a curved surface, the angle of incidence (and hence the angle of refraction) will vary depending on which part of the surface a parallel light ray interacts with. The theory introduced here describes how lenses are used to manipulate light rays.

16.1 Refraction at a spherical surface

We will first consider how a light ray refracts when encountering a single curved interface between two media. In the example shown in Figure 16.1, the image is shown inside the more optically dense medium.

Examining the rays in Figure 16.1, we can use Snell's law and geometry to establish the relationships between the refractive indices and the image properties.

Firstly, via Snell's law:

$$\begin{aligned} n_1 \sin \theta_1 &= n_2 \sin \theta_2 \\ n_1 \theta_1 &\approx n_2 \theta_2 \quad (\text{for small angles}) \end{aligned} \tag{16.1}$$

We use the equivalency that, at small angles (in radians), $\sin \theta \approx \theta$. This allows us to study the geometry and establish relationships between the angles:¹

¹It is worth ensuring you can follow through the geometric arguments which lead to these

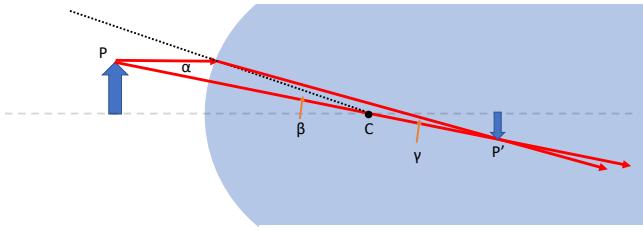


Figure 16.1: Refraction at a spherical surface involves the focussing of the rays to a real image. Note that this is not (yet) a lens as only a single refraction is considered. Rays through the center of curvature C are perpendicular to the surface and are not refracted, while rays which are not perpendicular to the surface are subject to Snell's law and refract as expected. These rays converge with the central rays to create an image at P' .

$$\begin{aligned}\beta &= \theta_2 + \gamma = \frac{n_1}{n_2} \theta_1 + \gamma \\ \theta_1 &= \alpha + \beta\end{aligned}\quad (16.2)$$

We can then substitute our expression for θ_1 to eliminate this from the first expression and rearrange:

$$\begin{aligned}\beta &= \frac{n_1}{n_2}(\alpha + \beta) + \gamma \\ (n_2 - n_1)\beta &= n_1\alpha + n_2\gamma\end{aligned}\quad (16.3)$$

As we are assuming small angles, we are also assuming that the rays are paraxial, we can then establish relations for the angles in relation to Figure 16.2:

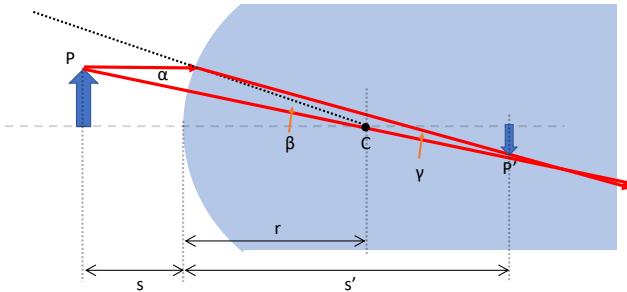


Figure 16.2: We can determine the position of the image at s' from the refractive indices and the object position s through determining the angles and assuming small values.

results.

Assuming paraxial rays, then:

$$\alpha \approx \sin \alpha = \frac{l}{s} \quad \beta \approx \sin \beta = \frac{r}{l} \quad \gamma \approx \frac{l}{s'} \quad (16.4)$$

We can now combine the results in Equation (16.4) with that of Equation (16.3) to obtain the general equation for a curved interface (Equation (16.5)), relating the image position to the refractive indices of the two materials and the radius of curvature, r :

$$\frac{n_1}{s} + \frac{n_2}{s'} = \frac{(n_2 - n_1)}{r} \quad (16.5)$$

16.1.1 Sign conventions

As with the mirror, there are a number of sign conventions to be aware of for a reflecting surface:

1. s is:
 - **positive** when the object (real) is in front of the surface (incident side);
 - **negative** when the object (virtual) is behind the surface (transmission side) *e.g.* formed by another lens or mirror;
2. s' is:
 - **positive** when the image (real) is behind the surface (transmission side);
 - **negative** when the image (virtual) is in front of the surface (incident side);
3. r, f are:
 - **positive** when the centre/focus are behind the surface (transmission side)
 - **negative** when the centre/focus are in front of the surface (incident side);

Note that these conventions are different to the mirror conventions.

16.1.2 Magnification

As for the spherical mirror, the curved interface can be considered to have magnification properties. We can consider these in terms of the angles of incidence and refraction along the optical axis of the surface (Figure 16.3).

In this construction, we have defined two triangles, each of which give us expressions for the angles of incidence and refraction; the angle of incidence can be determined by the height of the object (y) and its distance from the surface (s), while the angle of refraction is defined by the height of the image (y') and its distance behind the surface (s').

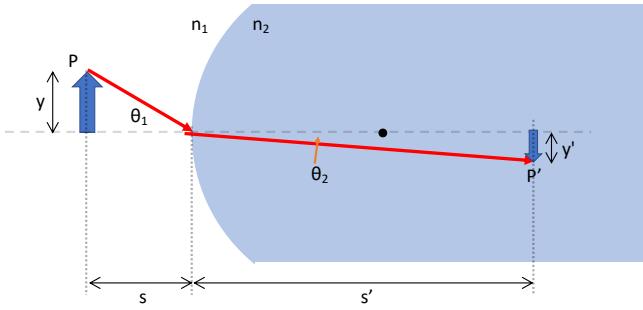


Figure 16.3: The magnification ratio between the object and image size is related to the angles θ_1 and θ_2 .

This allows us to then define the relationships between these angles:

$$\begin{aligned}\tan \theta_1 &= \frac{y}{s} & \theta_1 &\approx \frac{y}{s} \quad (\text{at small angles}) \\ \tan \theta_2 &= \frac{y'}{s'} & \theta_2 &\approx \frac{y'}{s'} \quad (\text{at small angles})\end{aligned}\tag{16.6}$$

Once again, we define the magnification as the ratio between the image size (y') and the object size (y); therefore using Snell's law ($n_1\theta_1 \approx n_2\theta_2$ for small angles) we can now relate these statements of y and s to find the magnification, m :

$$m = \frac{y'}{y} = -\frac{n_1 s'}{n_2 s}\tag{16.7}$$

The significance of the negative sign in Equation (16.7) is in indicating the inverted image (see Figure 16.3)

16.2 Refraction through a thin lens

We now consider a thin lens. The term “thin” is a way of describing the lens such that it is sufficiently thin that an image of an object is not able to form inside the more dense medium as described in Section 16.1. Rather, light rays refract once passing into the more dense medium, and then are refracted again back into the less dense medium before an image is formed.

Refraction occurs at two surfaces of radius r_1 and r_2 respectively. Generally we assume the lens is in air, so the refractive index of the material is n and we assume the refractive index of air to be ~ 1 . We consider the refraction at each surface separately in order to build the overall picture of the light path shown in Figure 16.4.

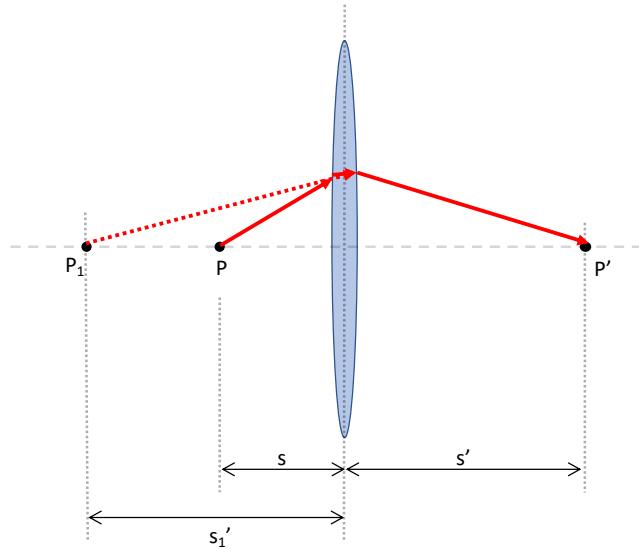


Figure 16.4: A thin lens now has two refracting surfaces (some lenses are flat on one side; they still have refraction at these surfaces!). The first refraction from the object at point P creates an apparent virtual object at P' for the second refraction.

The best way to understand what is happening in Figure 16.4 is to look at each interface in turn:

- At the first surface, we have the same process as we encountered in Section 16.1 with the light refracting from the air into the lens material. This creates a light path within the lens appearing to originate from a virtual object image at P_1 . We can determine its position by adapting Equation (16.5) by replacing our values of $n_1 = 1$ and $n_2 = n$ (Equation (16.8)).

$$\frac{1}{s} + \frac{n}{s'_1} = \frac{(n-1)}{r_1} \quad (16.8)$$

This virtual object formed at P_1 now becomes the object image for refraction at the second surface, and lies at a distance s'_1 from the lens. Note that:

- s'_1 is **negative** in the first refraction as P_1 is a virtual image for refraction at the first surface (image on same side as the object);
- s'_1 is **positive** for the second refraction as the “object image” at P_1 is on the incident side for the second surface.

We can now again adapt Equation (16.5) for the second refraction, again adapting our values of $n_1 = n$ and $n_2 = 1$.²

²Note that n_1 and n_2 are now reversed because we are now moving from the more dense

$$-\frac{n}{s'_1} + \frac{1}{s'} = \frac{(1-n)}{r_2} \quad (16.9)$$

We combine Equations (16.8) and (16.9) to eliminate the position of the virtual object at P_1 to obtain the overall equation for a light ray from an object refracting through a thin lens (Equation (16.10))

$$\frac{1}{s} + \frac{1}{s'} = (n-1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \quad (16.10)$$

The consequence of this is that, if we take the object distance $s = \infty$ (*i.e.* parallel rays), then $s' = f$, the focal length of the lens; therefore we can find the focal length of the lens via Equation (16.11):

$$\frac{1}{f} = (n-1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \quad (16.11)$$

Equation (16.11) is known as the *lens maker's equation*, allowing the creation of a given focal length f by shaping the radii of curvature on either side. Additionally, by equating Equations (16.10) and (16.11), we obtain the *thin lens equation*, Equation (16.12):

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \quad (16.12)$$

Note that this is the same form as the mirror equation (Equation (15.5)), but using the differing sign conventions, reflecting the position of the image.

16.3 Types of lens

Having derived the lens equation, we now look at the different forms which lenses can take.

16.3.1 The converging lens

Any lens (with material of n greater than that of the surrounding medium) which is **thicker** in the middle than at the edges will be a **converging lens** *i.e.* parallel rays from any direction are brought to a focus on the focal plane (Figure 16.5).

F and F' are the first and second focal points respectively, each at a distance f from the lens.

material into the less dense material.

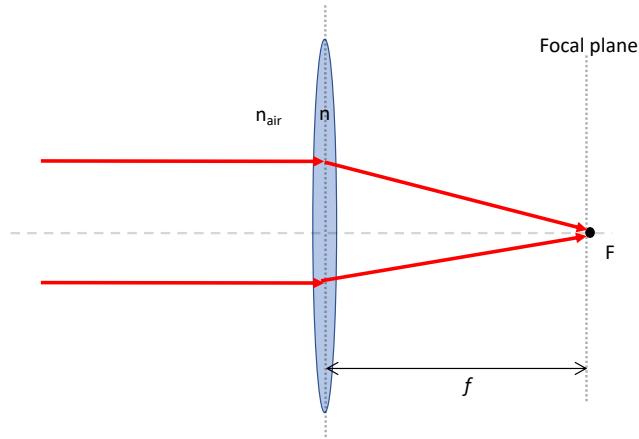


Figure 16.5: A converging lens is shown to be thicker in the middle than at the edges. This causes light rays to converge(!), creating a real image. In this example, rays from infinity (parallel rays) are focussed onto the focal point F . Any rays from 'infinity' at any angle will be focussed onto this focal plane.

16.3.2 The diverging lens

In contrast to the converging lens (Section 16.3.1), a **diverging lens** will be any lens (where n is greater than the surrounding medium) which is **thinner** in the middle than at the edges. In this case, parallel rays from any direction are diverged as though originating *from* the focal point F on the incident side of the lens (Figure 16.6)

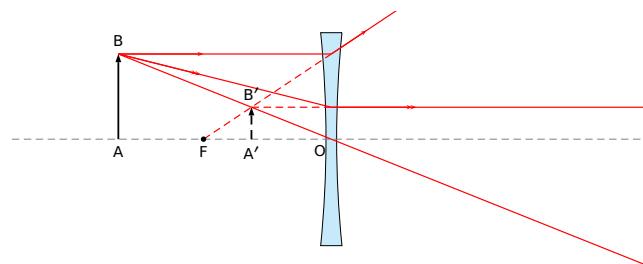


Figure 16.6: A diverging lens is shown to be thinner in the middle than at the edges. This causes light rays to diverge(!), however the rays do not converge to form an image. Instead, rays are projected backwards to the point at which they would converge (where the light rays appear to come from); this is a virtual image.

16.4 Ray diagrams

We now need to set a series of rules for constructing ray diagrams. This allows us to understand how lenses form their images, where those images lie, and allow us to ensure we use the correct sign conventions within the lens equations.

We consider three principal rays:

1. Parallel ray - drawn parallel to the lens axis, which is bent by the lens to pass through the focus (converging lens) or is bent to appear as though it comes *from* the focus (diverging lens);
2. Central ray - this passes through the centre of the lens and is undeviated by the lens;³
3. Focal ray - this is a ray originating from the focal point of a converging lens, and the emerging ray will be parallel to the lens axis. Generally this ray is not needed as the other two rays are usually sufficient to identify the location of the image.

16.4.1 Ray diagram for a converging lens

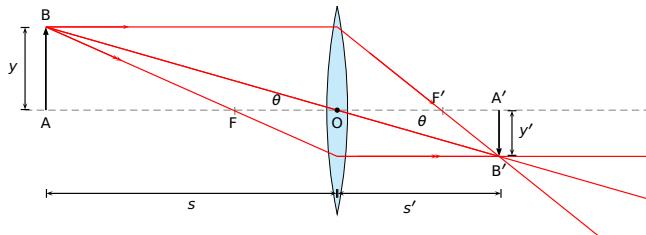


Figure 16.7: The rays in a converging lens shown in this example converge to create a real image. The distances s , s' can be used to determine the magnification ratio y'/y for the image.

Using Figure 16.7 we can see that, for an object between ∞ and F , the image is **real** and **inverted**. We can then determine the magnification by using geometry:

Since:

$$\tan \theta = \frac{y}{s} = \frac{-y'}{s'} \quad (16.13)$$

(remember that since the image is inverted, y' is negative). We can then determine the magnification m by finding the ratio of the object and image height:

³Actually, the central ray will be slightly affected, but the emerging ray is parallel to the incident ray, and the offset is negligible in most circumstances.

$$m = \frac{y'}{y} = -\frac{s'}{s} \quad (16.14)$$

The negative term in the magnification tells us that the image is inverted.

16.4.2 Ray diagram for a diverging lens

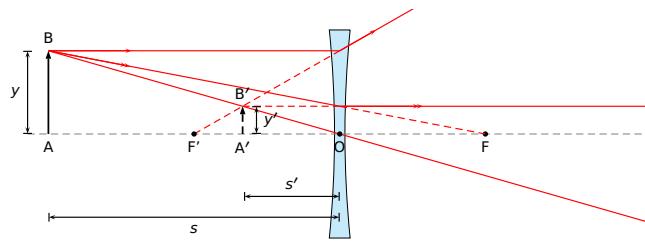


Figure 16.8: The rays in a diverging lens shown in this example do not converge; instead we look at the point where rays appear to come from to create a virtual image. The distances s , s' can again be used to determine the magnification ratio y'/y for the image.

Using Figure 16.8 we can see that, for an object between ∞ and F , the image is **virtual** and **erect**. We can then determine the magnification m by using geometry as before:

$$m = \frac{y'}{y} = -\frac{s'}{s} \quad (16.15)$$

16.5 Combining lenses

When two lenses are put together, the image formed by the first lens becomes the object for the second lens; this then becomes the basis for constructing a ray diagram. We can use these diagrams to determine the effective focal length of the combined lens system.

16.5.1 Lenses side-by-side, zero distance

In this example we imagine the two lenses sitting exactly next to each other along the principal axis such that there is no distance between them. Firstly, we write the thin lens equation for lens 1 (derived from Equation (16.12)), remembering that s_1 is the distance of the object from the lens, s'_1 is the distance of the image from the lens, and f_1 is the focal length of the lens.

$$\frac{1}{s_1} + \frac{1}{s'_1} = \frac{1}{f_1} \quad (16.16)$$

We can now write the thin-lens equation for lens 2 - this time, remember that the *image* formed by lens 1 becomes the *object* for lens 2; *i.e.* the distance of “object 2” (the image from lens 1) from lens 2 is $s_2 \equiv -s'_1$ (since this new “object” is a virtual object on the transmission side of the lens):

$$\frac{1}{s_2} + \frac{1}{s'_2} = -\frac{1}{s'_1} + \frac{1}{s'_2} = \frac{1}{f_2} \quad (16.17)$$

We can now add together Equations (16.16) and (16.17) to gain an expression for the effective focal length of the two-lens system, f_{eff} :

$$\begin{aligned} \frac{1}{f_{\text{eff}}} &= \frac{1}{f_1} + \frac{1}{f_2} \\ &= \frac{1}{s_1} + \frac{1}{s'_1} - \frac{1}{s'_1} + \frac{1}{s'_2} \\ &= \frac{1}{s_1} + \frac{1}{s'_2} \end{aligned} \quad (16.18)$$

16.5.2 Lenses adjacent, but separated

We can carry out the same analysis as in Section 16.5.1 and obtain the same result, however it will be more instructive this time to look at ray diagrams as this will help to explain what is happening, particularly in respect of the “virtual object” phenomenon.

Step One: consider the first lens (Figure 16.9)

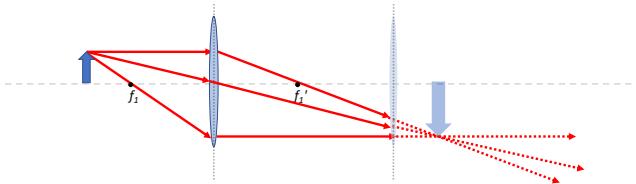


Figure 16.9: For a pair of converging lenses, the image from the first lens creates the object for the second. In this example, the second lens blocks formation of the otherwise ‘real’ image from the first lens.

We draw the ray diagram for the first lens, following the conventions for ray paths as laid out in Section 16.4; we can see that the image formed by the first lens in this example lies past the second lens.

Step Two: consider the second lens (Figure 16.10)

As the “object” which the second lens focuses is a “virtual object” (*i.e.* it lies on the transmission side of the lens), we need to treat the ray diagram for this second lens slightly differently; we need to project the rays *backwards* through the lens to the focal point and lens centre:

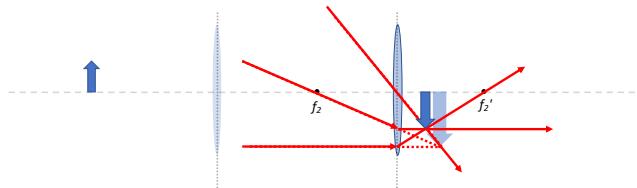


Figure 16.10: The now virtual image from the first lens creates a virtual object for the second lens; rays are projected backwards to imagine where they come from in order to set up the refractions for the second lens and create the final object.

1. The Parallel ray is projected back from the virtual object, parallel to the principal axis. As this ray existed in the first ray diagram, it is parallel to the principal axis before it encounters Lens 2; therefore this needs to be refracted to pass through the focal point of lens 2, F'_2 ;
2. The central ray is projected back from the virtual object and passes through the centre of lens 2. Where this intersects with the refracted parallel ray we define the position of the final image.
3. The focal ray is projected back from the virtual object to pass through F_2 ; in contrast to the parallel ray, it is refracted to be parallel to the principal axis, once again intersecting with the first two rays to form the final image.

This two-lens system ultimately yields a final image which is **real** (on the transmission side of the lens system), **inverted** and, for this example, smaller than the object.

Compound lens systems such as this are the basis of *optical trains* in instrumentation; a series of optical elements carrying and shaping a light beam through an instrument.

16.5.3 The Astronomical Telescope

A basic telescope consists of two optical elements; for example two lenses, two mirrors, or one lens and one mirror. The aim of using a telescope is to produce an image of a distant object which appears “close” to the eye. A consequence of this is that the image is much smaller than the object, but is closer to the observer and therefore appears larger. However, the lateral magnification is small. For this reason, we tend to think of telescopes in terms of **angular magnification** - the increase in angle subtended by the image at the eye. The larger this angle, the larger the apparent magnification of the object.

We can determine the angular magnification of a telescope through construction of an appropriate ray diagram. A telescope is constructed by placing the lenses so that the image of the objective lens is formed at the focal plane of the objective lens, with the eyepiece lens being placed so that the image *also* lies on its focal plane. This way the appearance of the light-rays coming from infinity (parallel

rays) is preserved for the observer.

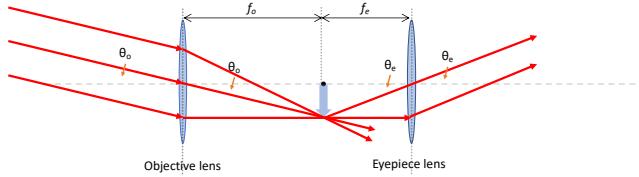


Figure 16.11: The two lenses in an astronomical telescope are positioned so their focal points are coincident. This means that parallel rays from infinity come into the telescope and parallel rays come from the eyepiece. This allows the observer to focus on an image which is much smaller than the object but, because it is closer to the observer, it has a large angular magnification.

The angular magnification is defined in a similar way to the linear magnification described in Section 16.1.2; namely the ratio of the eyepiece observation angle (θ_e) and the objective observation angle (θ_o), *i.e.*:

$$M = \frac{\theta_e}{\theta_o} \quad (16.19)$$

Via trigonometry on Figure 16.11, we can determine expressions for the angles in terms of the distances within the telescope, namely:

$$\tan \theta_o = -\frac{y'}{f_o} \approx \theta_o \quad \text{and} \quad \tan \theta_e = \frac{y'}{f_e} \approx \theta_e \quad (16.20)$$

Remember that the image height y' is negative as the image is inverted relative to the source. We can then write the magnification as:

$$M = \frac{\theta_e}{\theta_o} = -\frac{f_o}{f_e} \quad (16.21)$$

This expression for the magnification is negative telling us that the final image we observe through the eyepiece is inverted relative to the source.

Chapter 17

The Wave-Particle duality

Textbook link: Tipler and Mosca, Chapter 34

So far, we have spoken about light as an electromagnetic wave; indeed it was demonstrated to have wave behaviour by Thomas Young in 1801 with his double-slit experiment. We covered the background to this experiment in Section 13.3 and demonstrated how this shows the wave nature of light using our knowledge of superposition. Prior to Young's demonstration however there had been much debate about whether light was a wave or a particle.

In 1905, Einstein reintroduced the idea that light might behave as a particle, suggesting that light is quantised into photons of discrete energies; this offered an explanation for the photoelectric effect (the emission of electrons from a metal surface by light of certain frequencies or higher).

In this argument, the energy of a photon is related to the frequency ν (Greek letter ν) via Einstein's relation (Equation (17.1)):

$$E = h\nu = \frac{hc}{\lambda} = \hbar\omega \quad (17.1)$$

In this relationship, h is Planck's constant ($h = 6.626 \times 10^{-34}$ J s) and \hbar is the reduced Planck constant, $\hbar = \frac{h}{2\pi}$.

As we have two experiments, each of which demonstrates that light behaves as either a wave or as a particle depending on the circumstances, but neither wave nor particle behaviour adequately describes the behaviour of light in all situations. Generally however, we can say that the interaction between light and matter can be described by its particle properties, while the propagation of light can be described by its wave properties.

17.1 The de Broglie hypothesis

In 1924, Louis de Broglie suggested that if something which was previously understood to be a wave (light) can have particle properties, could something already understood to be a particle (*i.e.* matter) also have wave properties? He therefore postulated that particles might possess a wavelength related to its linear momentum p , described in Equation (17.2):

$$\lambda = \frac{h}{p} \quad (17.2)$$

Conversely, determination of the linear momentum of a wave could be determined by Equation (17.3):

$$p = \frac{h}{\lambda} = \hbar k \quad (17.3)$$

This creates the relationship between the linear momentum of a wave p and its wave number k (see Equation (7.5) for more on the wave number).

De Broglie went further, using the Einstein relation (Equation (17.1)) to define the frequency ν as in Equation (17.4), where E is the **total** energy; *i.e.* not just the kinetic energy, but also the potential energy and any other energy terms (including rest mass energy):

$$\nu = \frac{E}{h} \quad (17.4)$$

However, in a non-relativistic case (the only case we are considering), then the rest mass can be subtracted from this, resetting the energy scale to a new zero point. Of course, this will also change the frequency ν compared to a particle where we included the rest-mass energy. This is OK; we simply think of the frequency of the associated wave as being another measure of the energy of the particle for whatever energies we are considering.

In non-relativistic cases, the rest-mass energy is often just ignored without comment and, for our purposes, we will do the same. This means we consider E to simply consist of the sum of the kinetic and potential energies of the particle and ignore any other energies in the system. For particles, their momentum is not ambiguous and its relationship with wavelength is clear; this is therefore more useful than thinking about energy and frequency for most cases.

These equations are thought to apply to all particles, **including photons**. Therefore, via de Broglie's equation, photons must therefore carry momentum (Equation (17.5)), leading to the idea of radiation pressure *i.e.* that a beam of light exerts a pressure on any surface it strikes.

$$p = \frac{h}{\lambda} = \frac{E}{c} \quad (17.5)$$

For macroscopic objects the de Broglie wavelength is vanishingly small *e.g.* for a golf ball $\lambda \approx 10^{-34}$ m, and the macroscopic object behaves distinctly like a particle. However, for subatomic particles, the wavelength can have relatively large values in the nanometre range, leading to observable wave effects.

Today, the wave-nature of particles is well established, with electron and neutron diffraction experiments becoming routine. However, a question remains:

- If a particle is a wave, what is the nature of the wave, and what is the wave equation which governs it?

17.2 The wavefunction and its interpretation

Tipler and Mosca 34.5

The **wavefunction** is a concept which we use to describe the wave behaviour of particles; it can be thought of as a ‘*guiding wave*’, and from it we can determine a full mathematical description of the particle’s behaviour and properties. It uses the Greek letter ψ (psi) and is dependent on a particle’s position in space; $\psi(x)$ in one dimension, up to $\psi(x, y, z)$ in three dimensions.

If a particle does behave like a wave, what does the wavefunction, $\psi(x)$, look like and what does it actually mean? There are a number of interpretations depending on what we are considering:

- For waves on a string, $\psi(x)$ is the displacement of the string;
- For a sound wave, it is either the lateral displacement of the molecules, or the pressure at a given point;
- For an electromagnetic wave (light), it is either the electric or the magnetic field vector.

Recall that, for any wave, its power is proportional to the square of the amplitude, $\text{power} \propto \text{amplitude}^2$ (Section 7.4.4). In the context of light, we additionally said that the power was proportional to the intensity (Section 13.3) and that the intensity was proportional to the number of photons. It therefore follows that the number of photons is proportional to the square of the amplitude of the wavefunction; *i.e.:*

$$N \propto \psi^2(x) \quad (17.6)$$

The probability of finding a **single** photon in a small volume element, dx , is written as $P(x)dx$. Since $P(x)$ is proportional to the number of photons, it follows that:

$$P(x) \propto \psi^2(x) \quad (17.7)$$

In fact, we write this as a direct equality, and then apply a normalisation condition to the wavefunction, so that we can say that all of the probabilities add to one, *i.e.* the particle must be somewhere!

$$\int_{-\infty}^{\infty} \psi^2(x) dx = 1 \quad (17.8)$$

We can therefore interpret the square of a particle's wavefunction as a measure of the probability of finding the particle at a particular location.

The wavefunction is also a solution of the **Schrödinger wave equation**; a fundamental equation describing the propagation and evolution of the wavefunction of a particle. It effectively tells us how the probability of finding a particle at any point in space changes with time; we will come back to this concept presently.

17.3 Heisenberg's Uncertainty Principle

Tipler and Mosca 34.6

The central idea within the wave-particle duality is that if a particle's behaviour can be fully described using a wave representation, it will be governed by all of the physics we already know for waves. In particular, the wavefunction will be subject to the bandwidth theorem (Section 11.5), reproduced in Equation (17.9):

$$\Delta x \Delta k \approx 2\pi \quad (17.9)$$

The de Broglie relation introduced in Equation (17.3) can be rearranged to its form in Equation (17.10)

$$k = \frac{p}{\hbar} \quad (17.10)$$

Combining Equations (17.9) and (17.10) gives us the famous Heisenberg relation, Equation (17.11):

$$\Delta x \Delta p \approx h \quad (17.11)$$

The relationship we show in Equation (17.11) is only an approximation because the precise form of the Heisenberg relation depends on the definition of the width of the wavefunction; *e.g.* if Δx and Δp are defined as standard deviations in measurements, we obtain:

$$\Delta x \Delta p \geq \frac{1}{2} \hbar \quad (17.12)$$

For the purposes of this course, unless explicitly asked, either formulation shown in Equation (17.11) or (17.12) is acceptable, but be clear which you are using.

If we choose to express the uncertainty principle in words, we can state it as:

It is not possible to measure both the position of a particle and its linear momentum simultaneously to an arbitrary position.

In other words: the precision with which we measure one characteristic affects the precision with which we are able to measure the other.

Heisenberg’s uncertainty principle is a natural consequence of the bandwidth theorem as it is a natural consequence of the wave-type behaviour of a particle.

We have considered position and linear momentum as intrinsically “particulate behaviour”; how does this then relate to the measurement of a wave, in which we measure frequency and time? We now revert to the other representation of the bandwidth theorem:

$$\Delta t \Delta \omega \approx 2\pi \quad (17.13)$$

We can then make some substitutions to consider the energy of particles. Since:

$$E = h\nu = h \times \frac{\omega}{2\pi} = \hbar\omega \quad (17.14)$$

Therefore, via substitution:

$$\Delta t \Delta E \approx h \quad (17.15)$$

...or, if we are considering standard deviations:

$$\Delta t \Delta E \geq \frac{1}{2} \hbar \quad (17.16)$$

17.4 The “Particle in a box” problem

The “Particle in a box” model is a useful model for considering how a particle behaves when it is confined within spatial limits; for example an electron trapped within an atom, a proton confined within a nucleus, and so on. There are numerous discussions of the model, all of varying complexity, however they all rely on the same basic assumptions:

1. The “box” is a defined region of space, whether in one, two or three dimensions;
2. Inside the “box”, the potential is zero (so it can only carry kinetic energy)
3. Outside the “box”, the potential is infinite; thereby constraining the particle within the box
4. At the boundary of the box, the wavefunction drops to zero:
 - Wavefunctions must be continuous; the wavefunction is zero outside the box, so a discontinuity would result if the wavefunction did not drop to zero towards the boundary.

These conditions constrain the wavefunction to specified wavelengths within the box; if we are required to have a node at the edges of the box, we can only fit a “whole number of half-wavelengths” in the box; if we consider a one dimension box we consider the shapes of the wave within the box to be akin to the form of a standing wave on a stretched string. This means only certain frequencies - and hence certain energies - are permitted within the box, and we have **quantised energies**.

We can arrive to this result by another discussion:

1. Particles are represented by wave-packets; these propagate through a system as a “fuzzy cloud” of probabilities. We know the particle exists somewhere within the envelope of the wavepacket, but we don’t know exactly where.
2. The “box” is a region of space where the potential confines the particles (and hence their wavefunction).
3. Imagining a one-dimensional box, a classical particle would simply bounce back and forth between the box, having any value of energy or momentum.
4. A quantum particle also bounces back and forth, but now it is the wavepacket which is reflected by the walls of the box. As with waves on a string, constructive and destructive interference occurs between incident and reflected waves.
5. Only certain wavelengths interfere constructively, forming a standing wave pattern.

Whichever explanation we follow, we end with the same result; that the wavefunction (and hence the energy) of a constrained particle is **quantised** by the existence of the boundary conditions.

As we have mentioned, the wavefunction ψ is zero at the edges of the box, meaning that any wave we observe must have a “whole number of half-wavelengths”; *i.e.* the possible wavelengths which will fit in a box of length L are given by Equation (17.17):

$$\lambda_n = \frac{2L}{n} \quad \text{or} \quad L = n \frac{\lambda_n}{2} \quad \text{where} \quad n = 1, 2, 3, \dots \quad (17.17)$$

As we said that the potential within the box is zero, the total energy of the

particle in the box is equal to its kinetic energy, *i.e.*:

$$E = \frac{p^2}{2m} \quad (17.18)$$

Applying the de Broglie relation for momentum, we can find the energy of the n th harmonic (where n is the quantum number for the system):

$$E_n = \frac{p_n^2}{2m} = \frac{\left(\frac{h}{\lambda_n}\right)^2}{2m} \quad (17.19)$$

We can then substitute the expression for λ_n from Equation (17.17) to give:

$$E_n = \frac{h^2}{2m\lambda_n^2} = \frac{n^2h^2}{8mL^2} \quad (17.20)$$

In other words, Equation (17.17) shows how only energy levels E_n are allowed when the particle is constrained in a one-dimensional box. This model can be extended into two and then three dimensions to describe the allowed energy levels within an atom. We can see that if we have a large box, the energy levels get closer and closer together which, at macroscopic dimensions, will resemble a continuum. For small box sizes however (size of atoms, molecules etc.) the energy levels become widely spaced and energy quantisation dominates behaviour.

If the energy of the particle does not fit the E_n formulation, then the wavefunction will instead be given by a superposition of the allowed states:

$$\psi = c_1\psi_1 + c_2\psi_2 + c_3\psi_3 + \dots \quad (17.21)$$

The position of the particle cannot be exactly determined in a quantum mechanical system. While a classical particle could be anywhere within the box, the quantum particle’s position is given when we determine its probability of existence *via* $\psi^2(x)$ (see Section 17.2). When we plot the square of the wavefunction we see that a quantum mechanical particle is more likely to be in certain regions in the box, and there will be nodes in the box where the particle will not reside (Figure 17.1).

For a large n , the maxima and minima of the quantity $\psi^2(x)$ become very close together and the particle is then likely to be found more-or-less anywhere within the box, just as for a classical particle.

This is an example of the **correspondence principle**:

In the limit of large quantum numbers, the classical and quantum mechanical calculations must yield the same results.

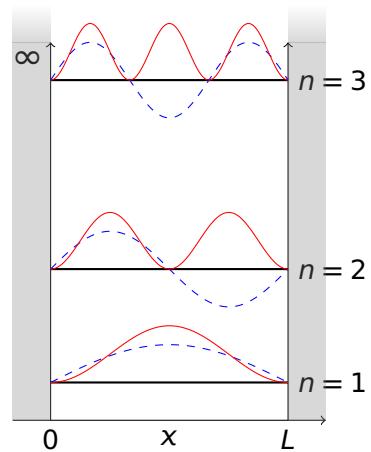


Figure 17.1: The relative energies of the first three quantum numbers $n = 1, 2, 3$ for the particle in a box. The wave function of the particle ($\psi(x)$; dotted blue line) is shown along with the probability of finding the particle at each energy state ($\psi^2(x)$; solid red line).

Chapter 18

The Schrödinger Equation

Textbook link: Tipler and Mosca, Section 35.1

The *time dependent* Schrödinger equation in one dimension is given in Equation (18.1):

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + U\psi(x, t) = i\hbar \frac{\partial \psi(x, t)}{\partial t} \quad (18.1)$$

...where U is the potential energy function which might vary with position and time. Note the factor i and that the right-hand side only has a first derivative with respect to time. Additionally, the wavefunction itself may be complex, so we modify our expression for the probability of finding a particle in a region of space dx to:

$$P(x, t)dx = |\psi(x, t)|^2 dx = \psi^* \psi dx \quad (18.2)$$

...where ψ^* is the complex conjugate of ψ .

If U is independent of time, we can separate out the time-dependence of the wavefunction *i.e.* the oscillatory time-dependent amplitude of the standing wave as follows:

$$\psi(x, t) = \psi(x)e^{-i\omega t} \quad (18.3)$$

The probability distribution for this wavefunction is then given by:

$$P(x, t) = \psi^* \psi = \psi^*(x)e^{+i\omega t} \times \psi(x)e^{-i\omega t} \quad (18.4)$$

The exponential terms cancel, giving the result:

$$P(x, t) = |\psi(x)|^2 \quad (18.5)$$

...in other words, the probability distribution is static with time. If we take the time-independent expression for the wavefunction in Equation (18.3) and substitute this into Equation (18.1), differentiating and cancelling through by a factor of $e^{-i\omega t}$, we get the *time-independent* form of the equation, Equation (18.6):

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + U(x)\psi(x) = E\psi(x) \quad (18.6)$$

...where $E = \hbar\omega$ is the energy of the particle.

Since we are only working in the non-relativistic case, the rest-mass energy is not included in E here (*i.e.* we are discussing energies in addition to the rest-mass energy). As discussed in the last chapter (17), this is fine as it simply offsets all energies by a constant amount without changing any of the physics. In this case, E in the time-independent Schrödinger equation is again just the sum of the kinetic and potential energies.

The time-independent Schrödinger equation may be solved for a variety of systems, provided that the resultant wavefunction, $\psi(x)$ satisfies any boundary conditions, as well as satisfying the normalisation condition:

$$\int_{-\infty}^{\infty} |\psi|^2 dx = 1 \quad (18.7)$$

...which implies also that we must have:

$$\psi(x) \rightarrow 0 \quad \text{as} \quad x \rightarrow \pm\infty \quad (18.8)$$

18.1 Particle in a square potential well

18.1.1 The infinite potential well

We have already introduced an example of this; the particle in a box problem (Section 17.4), and the formal solutions for a particle in a square potential well were deduced there. There are parallels with electrons in an atom, however going from a one dimensional “box” to a three dimensional spherical surface is slightly involved.¹

¹The wavefunctions which fit onto the surface of a sphere describe the shapes of atomic orbitals, a collection of waveforms known as *spherical harmonics*. These same waveforms are widely used in other applications, including (among other things) acoustic engineering.

The general principle with the infinite well is that the particle cannot ever escape the box; inside the box the potential energy U is equal to zero, and the time-independent equation reduces to the form shown in Equation (18.9) (for one dimension):

$$\frac{d^2\psi(x)}{dx^2} = -k^2\psi(x) \quad (18.9)$$

...where:

$$k^2 = \frac{2mE}{\hbar^2} = \frac{p^2}{\hbar^2} \quad \text{as expected} \quad (18.10)$$

This equation should look very familiar to us; it looks like a simple harmonic oscillator with the oscillation instead being in space rather than in time (see Section 1.1). The solutions to this equation then are any sine or cosine wave, leading to the expected standing wave solutions when the boundaries are taken into account.

18.1.2 The finite potential well

We now adapt the infinite well to a finite well; the potential inside the well is still zero, however we now have a finite energy barrier at the edges.

In a classical consideration, if the energy E of the particle is less than the depth of the well (U_0), then the particle is trapped inside the well. For a quantum particle however the picture is rather different. When we examine the Schrödinger equation we now have to consider what happens with the potential energy term. Recall that the general form of the time-independent Schrödinger equation is as follows:

$$-\frac{\hbar^2}{2m} \frac{\partial^2\psi(x)}{\partial x^2} + U(x)\psi(x) = E\psi(x) \quad (18.11)$$

Inside the well, $U = 0$, so we would expect to have the same sinusoidal functions as the infinite well, however outside the well we need to consider the fact that U has a finite value (even if $E < U_0$). We rearrange the Schrödinger equation as follows:

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{\partial^2\psi(x)}{\partial x^2} + U(x)\psi(x) &= E\psi(x) \\ \frac{\partial^2\psi(x)}{\partial x^2} &= \frac{2m(U_0 - E)}{\hbar^2}\psi(x) \end{aligned} \quad (18.12)$$

It can be shown that a simple sinusoidal function such as $\psi = \sin(kx)$ will not be a solution to this equation; however an exponential will work as a solution in the form $\psi = e^{\pm\kappa x}$, where κ is given by:

$$\kappa = \frac{\sqrt{2m(U_0 - E)}}{\hbar} \quad (18.13)$$

The \pm solutions are exponential increase at $x < 0$ and decrease at $x > L$ (*i.e.* there is an exponential ‘tail-off’ at either boundary of the box). If the energy of the particle is less than U_0 , the particle will be bound by the potential well, as expected, **however**, the wavefunctions will actually ‘spill over’ the edges of the box. This implies that the particle will spend some of its time **outside the well**, contrary to classical expectations!

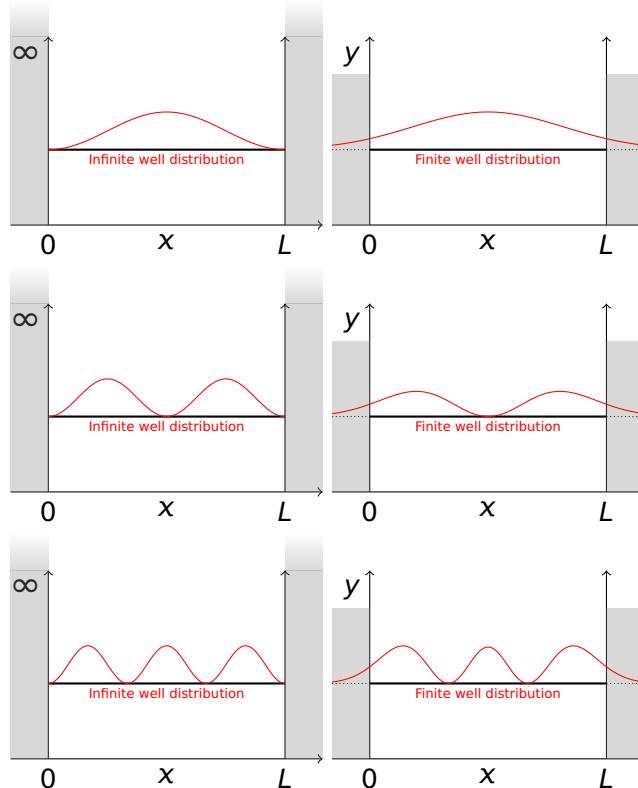


Figure 18.1: In a finite potential well, there is an exponential ‘tail off’ in the probability of finding the particle at the edges of the box. In these circumstances, the wavefunction decays exponentially outside the boundaries, but there is a finite probability of finding the particle at these locations.

18.2 Reflection and transmission of particle waves

Let's now consider what happens as a particle of energy E travels towards a potential energy step function of height U_0 (Figure 18.2).

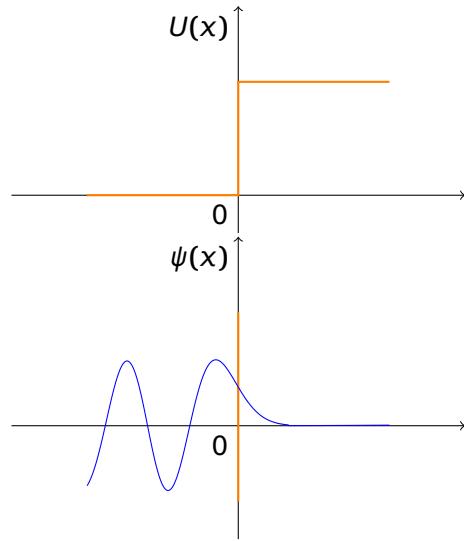


Figure 18.2: A potential energy step function such as that which might bound a finite well is shown here; a sudden step up in potential from one region of space to another. If the wavefunction does not have sufficient energy to pass the step, it will tunnel into the step and decay exponentially (lower diagram).

In a classical situation, we can consider the following situations:

- If $E < U_0$, the particle would be completely reflected
- If $E > U_0$, the particle would be completely transmitted, but with a reduction in velocity corresponding to the loss of kinetic energy in overcoming the potential barrier.

In a quantum mechanical situation however things are rather different:

- If $E < U_0$, the wavefunction penetrates a short distance beyond the step, into the classically forbidden region; however the particle is totally reflected.
- If $E > U_0$, the particle is sometimes reflected, sometimes transmitted.

Whether the particle is transmitted or reflected is a probabilistic process; the probability of reflection is given by:

$$R = \frac{(k_1 - k_2)^2}{(k_1 + k_2)^2} \quad (18.14)$$

...where k_1 and k_2 are the wavenumbers in each of the two regions. This result should, again, be familiar to us as the same result we obtained for waves reflecting on strings as the wave moves from one string to another of a different density (Section 8.1). Once again, remember that a given wavenumber k is defined as:

$$k^2 = \frac{2mE}{\hbar^2} = \frac{p^2}{\hbar^2} \quad (18.15)$$

18.3 Barrier penetration - ‘tunnelling’

Now, instead of a step potential, consider a potential barrier in the path of the particle, at a height such that the energy of the particle E is slightly less than U_0 (Figure 18.3):

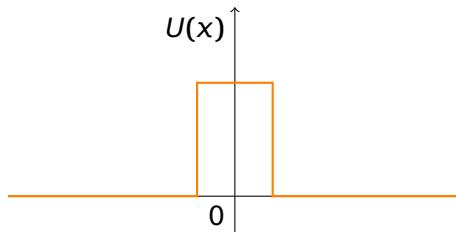


Figure 18.3: A potential barrier has finite width as shown above. This can stop a particle passing however a particle can tunnel through.

As in Section 18.2, we expect the wavefunction to penetrate into the barrier, **however** if the barrier is sufficiently narrow, the wavefunction will have a finite amplitude when it reaches the end of the barrier, and the wavefunction returns from exponential behaviour to oscillating sinusoidal behaviour once again, though with a reduced amplitude (Figure 18.4):

In other words, there is a probability that the particle will penetrate, or “tunnel through”, the barrier. The smaller amplitude on the other side of the barrier indicates that there is a smaller probability of finding the particle on the transmission side of the barrier. In classical mechanics this would be a completely forbidden behaviour; though classical tunneling of waves is possible (via evanescent modes when a wave is undergoing total internal reflection; see Section 14.2.1).

Quantum mechanical tunneling has been proposed as the mechanism by which α particles are spontaneously and unpredictably ejected from radioactive nuclei, and the phenomenon sees application in the scanning tunneling microscope (STM) to obtain atomic resolution images of certain surfaces by monitoring the current as electrons tunnel across a gap between a probe and the surface.

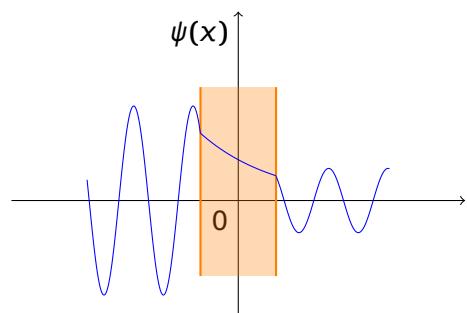


Figure 18.4: A waveparticle approaching a potential barrier can tunnel through the barrier if it is sufficiently narrow. It will appear the other side of the barrier with a decayed amplitude, suggesting a smaller probability of finding the particle on the other side.