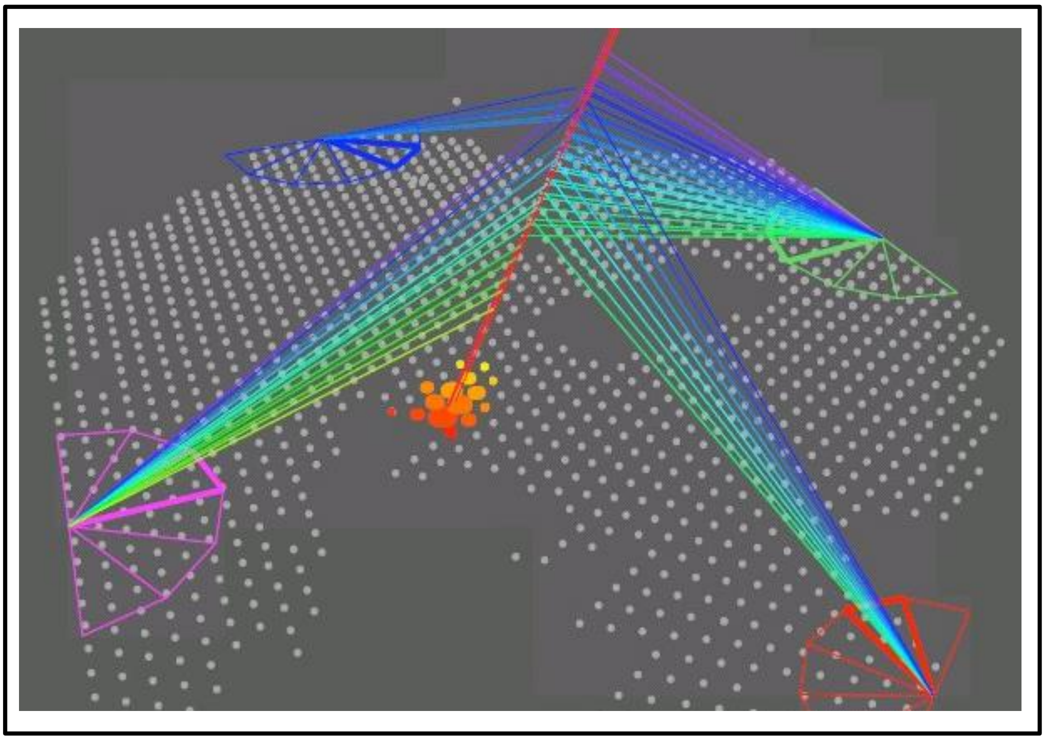
Grove Academy
Advanced Higher Physics

Quanta and Waves
Summary Notes



Contents

Quantum Theory

An Introduction to Quantum Theory
Quantum mechanics was developed in the early twentieth century to explain experimental observations that could not be explained by classical physics. In many cases various quantisation 'rules' were proposed to explain these experimental observations but these did not have any classical justification. This results in physics that is difficult to reconcile with everyday experience or with normal intuition.

This led to two of the most famous physicists of the twentieth century to make the following statements:

'If quantum mechanics hasn't profoundly shocked you, you haven't understood it yet.'
Neils Bohr

'I think it safe to say that no one understands quantum mechanics.'
Richard Feynman

However, quantum mechanics has been one of the most successful theories in physics explaining many experimental observations such as blackbody radiation, the photoelectric effect, spectra, atomic structure and electron diffraction.

Blackbody Radiation

Towards the end of the nineteenth century there was interest in the frequencies (or wavelengths) emitted by a 'black body' when the temperature is increased. When an object is heated it can radiate large amounts of energy as infrared radiation. We can feel this if we place a hand near, but not touching, a hot object. As an object becomes hotter it starts to glow a dull red, followed by bright red, then orange, yellow and finally white (white hot). At extremely high temperatures it becomes a bright blue-white colour.

Measurements were made of the intensity of the light emitted at different frequencies (or wavelengths) by such objects. In addition measurements were made at different temperatures. In order to improve the experiment and avoid any reflections of the radiation, a cavity was used with a small hole, which emits the radiation: a black body.

A surface that absorbs all wavelengths of electromagnetic radiation is also the best emitter of electromagnetic radiation at any wavelength. Such an ideal surface is called a **black body**. The continuous spectrum of radiation it emits is called black-body radiation.

It was found that the amount of black-body radiation emitted at any frequency depends **only** on the temperature, not the actual material.

Specific intensity is a measure of the radiation *emitted* by a body. Irradiance is a measure of the radiation *received* by a surface.

Graphs of specific intensity against wavelength (or frequency) are shown in Figure 1. As the temperature increases, each maximum shifts towards the higher frequency (shorter wavelength).

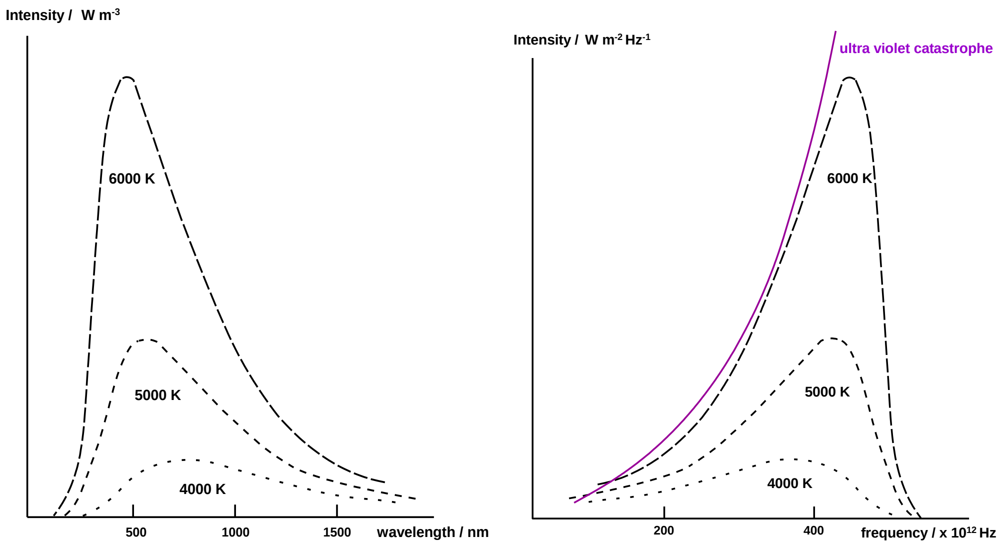


Figure 1 *Graphs of specific intensity against wavelength and against frequency.*

Attempts to obtain theoretically the correct black-body graph using classical mechanics failed. Wien obtained an equation that 'fitted' observations at high frequencies (low wavelengths). Later Lord Rayleigh obtained an equation that 'fitted' at low frequencies but tended off to infinity at high frequencies (see line on the above frequency graph). This divergence was called the ultraviolet catastrophe and puzzled many leading scientists of the day.

In 1900 Planck looked at the two equations and produced a 'combined' relationship, which gave excellent agreement with the experimental curve. However, initially this relationship could not be derived from first principles. It was a good mathematical 'fudge'!

Planck studied his relationship and the theory involved and noticed that he could resolve the problem by making the assumption that the absorption and emission of radiation by the oscillators could only take place in 'jumps' given by:

$$E = nhf \qquad (1)$$

where E is energy, f is frequency, h is a constant and n = 0, 1, 2, 3, ….

Using this assumption he could derive his equation from first principles. The constant of proportionality h was termed Planck's constant. (The word quantum, plural quanta, comes from the Latin 'quantus', meaning 'how much'.)

It must be emphasised that Planck did this in a mathematical way with no justification as to why the energy should be quantised – but it worked! To Planck the oscillators were purely theoretical and radiation was not actually emitted in 'bundles', it was just a 'calculation convenience'. It was some years before Planck accepted that radiation was really in energy packets.

The Photoelectric Effect

In 1887 Hertz observed that a spark passed between two plates more often if the plates were illuminated with ultraviolet light. Later experiments by Hallwachs and Lenard gave the unexpected results we are familiar with, namely:

(a) the non-emission of electrons with very bright but low frequency radiation on a metal surface, e.g. very bright red light, and
(b) the increase in the speed of the emitted electron with frequency but not with intensity. Increasing the intensity only produced more emitted electrons.

These results were unexpected because energy should be able to be absorbed continuously from a wave. An increase in the intensity of a wave also means an increase in amplitude and hence a larger energy.

In 1905 Einstein published a paper on the photoelectric effect entitled *On a Heuristic Viewpoint Concerning the Production and Transformation of Light*. He received the Nobel Prize for Physics for this work in 1921. The puzzle was why energy is not absorbed from a continuous wave, e.g. any electromagnetic radiation, in a cumulative manner. It should just take more time for energy to be absorbed and an electron emitted but this does not happen. Einstein proposed that electromagnetic radiation is emitted and absorbed in small packets. (The word ‘photon’ was introduced by Gilbert in 1926.) The energy of each packet is given by:

$$E = hf \qquad (2)$$

where E is the energy of a ‘packet’ of radiation of frequency f.

This proposal also explained why the number of electrons emitted depended on the irradiance of the electromagnetic radiation and why the velocity of the emitted electrons depended on the frequency. It did not explain the ‘packets’ or why they should have this physical ‘reality’.

Models of the Atom

Rutherford's scattering experiment indicated that the majority of the mass of the atom was in a small nucleus, with the electrons 'somewhere' in the atomic space. He and his assistants could not 'see' the electrons. A picturesque model of the atom, similar to a small solar system, came into fashion. This model had some features to commend it. Using classical mechanics, an electron in an orbit could stay in that orbit, the central force being balanced by electrostatic attraction. However, the electron has a negative charge and hence it should emit radiation, lose energy and spiral into the nucleus.

The current theory was insufficient. Why do the electrons 'remain in orbit'? Do they in fact 'orbit'?

In the late nineteenth century attempts were made to introduce some 'order' to the specific frequencies emitted by atoms. Balmer found, by trial and error, a simple formula for a group of lines in the hydrogen spectra in 1885.

$$\frac{1}{\lambda} = R(\frac{1}{2^2} - \frac{1}{n^2}) \qquad (3)$$

where λ is the wavelength, R the Rydberg constant, n is an integer 2, 3, 4,…

Other series were then discovered, e.g. Lyman with the first fraction $1/1^2$ and Paschen with the first fraction $1/3^2$. However, this only worked for hydrogen and atoms with one electron, e.g. ionised helium, and moreover did not provide any theoretical reason why the formula should work.

In 1913 Bohr introduced the idea of **energy levels**. Each atom has some internal energy due to its structure and internal motion but this energy cannot change by any variable amount, only by specific discrete amounts. Any particular atom, e.g. an atom of gold say, has a specific set of energy levels. Different elements each have their own set of levels. Experimental evidence of the day provided agreement with this idea and energy level values were obtained from experimental results. Transitions between energy levels give the characteristic line spectra for elements.

In order to solve the problem that an electron moving in a circular orbit should continuously emit radiation and spiral into the nucleus, Bohr postulated that an electron can circulate in certain permitted, stable orbits without emitting radiation. He made the assumption that the normal electromagnetic phenomena did not apply at the atomic scale! Furthermore he made an intuitive guess that angular momentum is quantised. It is said that he noticed that the units of Planck's constant ($J\ s^{-1}$) are the same as those of angular momentum ($kg\ m^2\ s^{-1}$). The allowed orbit, of radius r, of an electron must have angular momentum of an integral multiple of $h/2\pi$:

$$mvr = \frac{nh}{2\pi} \qquad (4)$$

where n is an integer 1, 2, 3,…

The angular momentum of a particle, of mass m, moving with tangential speed v, is mvr.

Thus for any specific orbit n we can calculate the radius of that orbit given the tangential speed or vice versa.

Theoretical aside

For the hydrogen atom with a single electron, mass m_e revolving around a proton (or more correctly around the centre of mass of the system), we can assume the proton is stationary since it is ~2000 times bigger. Hence, equating the electrostatic force and centripetal force $m_e v^2/r$:

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r_n{}^2} = \frac{m_e v_n{}^2}{r_n}$$ where ε_0 is the permittivity of free space (5)

for the nth orbit.

Equations (4) and (5) can be solved simultaneously to give:

$$r_n = \varepsilon_0 \frac{n^2 h^2}{\pi m_e e^2} \text{ and } v_n = \frac{1}{\varepsilon_0}\frac{e^2}{2\pi n h} \quad (6)$$

for the nth orbit.

Calculating r_1 for the radius of the first Bohr orbit uses data given in assessments, namely h, m_e, e, ε_0, and gives $r_1 = 5.3 \times 10^{-11}$ m.

These equations give the values of the radii for the **non-radiating** orbits for hydrogen and the value of n was called the **quantum number** of that orbit.

Example problem

For the hydrogen atom, calculate the velocity of an electron in the first Bohr orbit of radius 5.3×10^{-11} m.

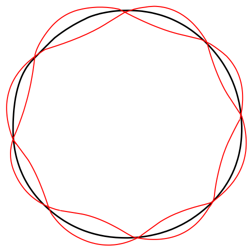
Using $mvr = \frac{nh}{2\pi}$ with $n = 1$ we obtain $v = 2.2 \times 10^6$ m s^{-1}.

Note: Bohr's theory only applies to an atom with *one* electron, e.g. the hydrogen atom or ionised helium atom.

However, as mentioned above, the idea of energy levels can be extended to all atoms, not just hydrogen.

So this theory is not complete since it did not allow any **prediction** of energy levels for any specific element nor did it explain why angular momentum should be quantised or why electrons in these orbits did not radiate electromagnetic energy! Another question was, 'what happens during a transition?'

An aside
An extension by de Broglie suggested that electron orbits are standing waves. The electron, now behaving like a wave, forms a standing wave of an integral number of wavelengths that just 'fits' into the circumference of an orbit.



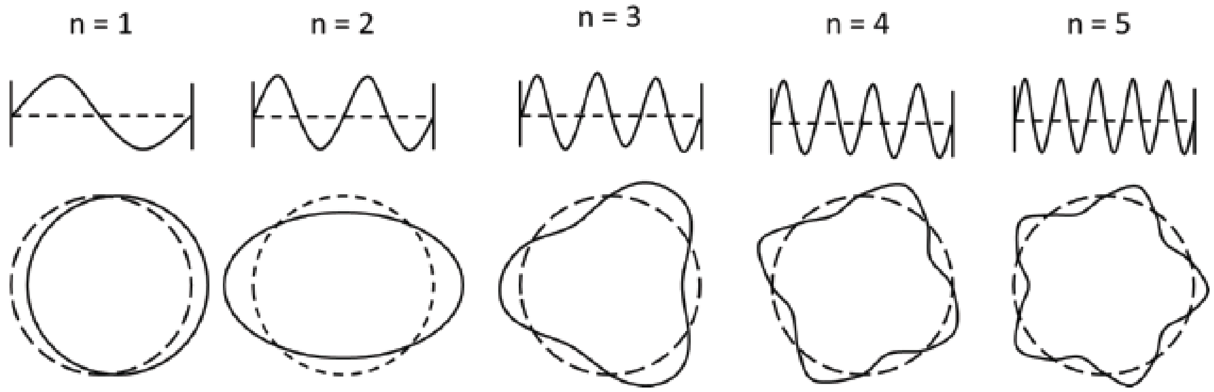
The red standing wave has an integral number of full waves 'fitting' into the circumference. (The central nucleus is not shown.)

Figure 2 Electron standing waves

This is a picturesque idea but somewhat old-fashioned since an electron does not take a 'wiggly' path around the nucleus. This is an example of a model that should not be taken too seriously and could lead to poor understanding. Quantum mechanics shows that we cannot describe the motion of an electron in an atom in this way.

Electrons as "Standing Waves"
An extension by Louis de Broglie suggested that electron orbits are standing waves.

The electron, now behaving like a wave, forms a standing wave of an integer number of wavelengths that 'fits' into the circumference of an orbit. This is represented here:



The dashed lines represent the circumference, while the solid line represents the hypothesised wave-like path of the electron.
This is a picturesque idea but somewhat old-fashioned since an electron does not take a 'wiggly' path around the nucleus. This is an example of a model that should not be taken too seriously and could lead to poor understanding. Quantum mechanics goes on to show that we cannot describe the motion of an electron in an atom in this way.

De Broglie Wavelength

We use the word ‘particle’ to describe localised phenomena that transport mass and energy, and the word ‘wave’ to describe delocalised (spread out) phenomena that carry energy but no mass.
Experimental observations seem to suggest that both electromagnetic radiation and electrons can behave like particles *and* like waves. They exhibit both wave phenomena, such as interference and diffraction, and particle phenomena, for example photons causing electron emission in the photoelectric effect or electron ‘billiard ball type’ collisions.

An electron can show wave-like phenomena. In the mid-1920s G P Thomson, in Aberdeen, bombarded a thin metal with an electron beam and obtained a diffraction ring. In 1927 Davisson and Germer directed a beam of electrons onto the surface of a nickel crystal and observed the reflected beam. They had expected to see diffuse reflection since even this smooth surface would look ‘rough’ to the tiny electrons. To their surprise, they observed a similar pattern to X-ray diffraction from a surface. Thomson and Davisson were awarded the Nobel Prize in 1937 for demonstrating the wave-like properties of electrons. (Thomson’s father, J J Thomson, won the Nobel Prize in 1906 for discovering the electron as a particle.)

In 1923 de Broglie suggested that since light had particle-like properties, perhaps nature was dualistic and particles had wave-like properties.

From relativity theory, the energy of a particle with zero rest mass, e.g. a photon, is given by $E = pc$ and we know that $E = hf$, hence $p = h/\lambda$.

Thus the wave and particle are related through its momentum.

For a particle $p = mv$ and for a wave $p = h/\lambda$, giving a relationship $mv = h/\lambda$ or:

$\lambda = \frac{h}{p}$ where p is the momentum and h Planck’s constant (7)

Thus we can calculate the **de Broglie wavelength** of a particle of velocity v.

Example problems

1. A neutron and an electron have the same speed.
   Which has the longer de Broglie wavelength?
   The electron, since the neutron has the larger mass. (The mass is in the denominator.)

2. An electron microscope uses electrons of wavelength of 0.01 nm. What is the required speed of the electrons?
   Using $\lambda = \frac{h}{p}$ for electrons and $p = mv$ gives:
   $0.04 \times 10^{-9} = (6.63 \times 10^{-34})/9.11 \times 10^{-31} \times v)$ and $v = 1.8 \times 10^{-7}$ m s^{-1}

Notice that this wavelength of 0.04 nm is very much smaller than that of blue light. Hence the use of electrons in microscopes can improve the resolution of the image compared to optical microscopes.

Quantum Mechanics

'The more you see how strangely Nature behaves, the harder it is to make a model that explains how even the simplest phenomena actually works.'
Richard Feynman

Matter was thought to be 'atomistic' with 'particles' making basic interactions and the properties of the particles continually changing smoothly from place to place. Waves moved continuously from place to place.

Classical mechanics could not explain the various 'quantisation rules', which attempted to give some limited agreement between observation and theory. The apparent dual wave particle nature of matter could not be explained. With quantum theory these ideas needed to be revised.

There are various forms of quantum mechanics: Heisenberg's matrix mechanics, Erwin Schrödinger's wave mechanics, Dirac's relativistic field theory and Feynman's sum over histories or amplitude mechanics.

In essence quantum mechanics provides us with the means to calculate **probabilities** for physical quantities. Exact physical quantities, e.g. position or velocity, do not have unique values at each and every instant.

'Balls in quantum mechanics do not behave like balls in classical mechanics … an electron between release and detection does not have a definite value for its position. This does not mean that the electron has a definite position and we don't know it. It means the electron just does not have a position just as love does not have a colour.'
Strange World of Quantum Mechanics, D Styer

Quantum theories incorporate the following concepts:

(i) Transitions between stationary states are **discrete**. There is no meaning to any comment on a system in an intermediate state.

Depending on the experiment, matter or waves may behave as a wave or a particle. However, in a certain way they act like both together. It is just not a sensible question in quantum mechanics to ask if matter is a wave or a particle.

(ii) Every physical situation can be characterised by a wavefunction (or other mathematical formalism). This wavefunction is not directly related to any actual property of the system but is a description of the potentialities or possibilities within that situation. The wavefunction provides a statistical ensemble of similar observations carried out under the specified conditions. It does not give the detail of what will happen in any particular individual observation. The probability of a specific observation is obtained from the square of the wavefunction. This is an important and non-intuitive idea.

The quantum probability aspect is very different from classical physics, where we consider there is an actual state and any 'probability' comes from our inadequate measuring or statistical average.

In the quantum domain we can only calculate **probabilities**. For example, we cannot state when a particular nucleus will decay (although we can measure a half-life) but we could calculate the probability of a particular nucleus decaying after a certain time. This is typical of the rules of quantum mechanics – the ability to calculate probabilities.

Quantum mechanics has enjoyed unprecedented practical success. Theoretical calculations agree with experimental observations to very high precision.

Quantum mechanics also reminds us that there is discreteness in nature and there are only probabilities.

Double-slit experiment
The double-slit experiment with light shows an interference pattern. This is a standard experiment to demonstrate that light is a wave motion.

There is a central maximum opposite the central axis between the two slits as shown in Figure 3.

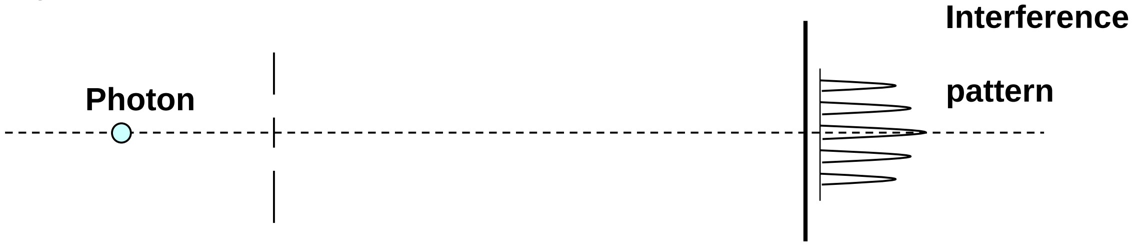


Figure 3 *Double slit experiment*

In more recent years this experiment has been performed with **single** photons and a detector screen. Each photon reaches the screen and the usual interference pattern is gradually built up.

The question is how does a single photon 'know about' the slit it does *not* pass through? Let us place a detector near each slit as shown in Figure 4. In this diagram the detectors are switched off and *not* making any measurements.

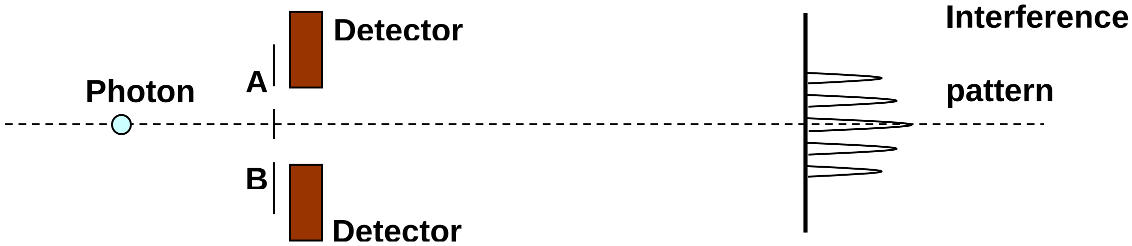


Figure 4 *Detectors switched off*

Let us now switch on detector A as shown in Figure 5.

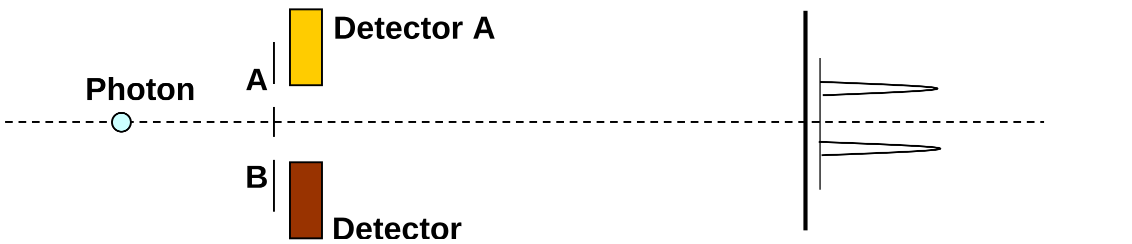


Figure 5 *Detector switched on*

We lose the interference effect and simply obtain a pattern for particles passing through two slits. We would get the same pattern if we switched on detector B instead of detector A or if we switched on both detectors.

It seems if we ask the question ‘Where is the photon?’ or ‘Which slit does the photon pass through?’ and set up an experiment to make a measurement to answer that question, e.g. determine which slit the photon passed through, we do observe a ‘particle’ with a position but lose the interference effect.

It appears that the single photon in some way does ‘know about’ both slits. This is one of the non-intuitive aspects of quantum mechanics, which suggests that a single particle can pass through *both* slits.

A very similar double-slit experiment can be performed with electrons. Again we can arrange for only one electron to ‘pass through’ the slits at any one time. The position of the electrons hitting the ‘screen’ agrees with our familiar interference pattern. However, as soon as we attempt to find out which slit the electron passes through we lose the interference effect. For electron interference the spacing of the ‘slits’ must be small. Planes of atoms in a crystal can be used to form the slits since electrons have a very small associated wavelength, the de Broglie wavelength.

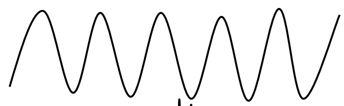
These observations are in agreement with quantum mechanics. We cannot measure wave and particle properties at the same time.

<u>The Uncertainty Principle</u>

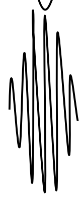
A theoretical introduction

Using the wave theory of quantum mechanics outlined above we can produce a wavefunction describing the ‘state’ of the system, e.g. an electron. However, we find that it is not possible to determine with accuracy *all* the observables for the system. For example, we can compute the likelihood of finding an electron at a certain position, e.g. in a box. The wavefunction may then be effectively zero everywhere else and the uncertainty in its position may be very small inside the box. If we then consider its momentum wavefunction we discover that this is very spread out, and there is nothing we can do about it. This implies that *in principle*, if we ‘know’ the position, the momentum has a very large uncertainty.

Consider a wave with a single frequency. Its position can be thought of as anywhere along the wave but its frequency is uniquely specified.



Now consider a wave composed of a mixture of slightly different frequencies, which when added together produces a small ‘wave packet’. The position of this wave can be quite specific but its frequency is conversely non-unique.



Heisenberg’s Uncertainty Principle should more appropriately be called Heisenberg’s Indeterminacy Principle since we can measure either x or p_x with very low ‘uncertainty’ but we cannot measure both. If one is certain, the other is indeterminate.

Theoretical considerations also shows that the energy E and time t have this dual indeterminacy.

A thought experiment to illustrate Heisenberg's Uncertainty Principle

In classical physics it was assumed that all the attributes, such as position, momentum, energy etc, could be measured with a precision limited only by the experiment.

In the atomic domain is this still true?

Let us consider an accurate method to determine the position of an electron in a particular direction, for example in the x direction. The simplest method is to use a 'light gate', namely to allow a beam of electromagnetic radiation to hit the electron and be interrupted in its path to a detector. To increase the accuracy we can use radiation of a small wavelength, e.g. gamma rays. However, we note that by hitting the electron with the gamma rays the velocity of the electron will alter (a photon-electron collision). Now the velocity or momentum of the electron in the x direction will have changed. Whatever experiment we use to subsequently measure the velocity or momentum cannot determine the velocity before the electron was 'hit'. To reduce the effect of the 'hit' we can decrease the frequency of the radiation, and lose some of the precision in the electron's position. We just cannot 'win'!

Heisenberg's Uncertainty Principle is stated as

$$\Delta x \Delta p_x \geq \frac{h}{4\pi} \qquad (8)$$

where Δx is the uncertainty in the position, Δp_x is the uncertainty in the component of the momentum in the x direction and h is Planck's constant.

Quantum mechanics can show that there are other pairs of quantities that have this indeterminacy, for example energy and time:

$$\Delta E \Delta t \geq \frac{h}{4\pi} \qquad (9)$$

where ΔE is the uncertainty in energy and Δt is the uncertainty in time.

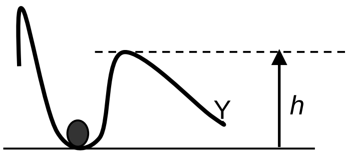
We notice that the pairs of quantities in these relationships (termed conjugate variables) have units that are the same as those of h, namely J s. For energy and time this is obvious. For position and momentum we have m kg m s^{-1}, which we can adjust as kg m^2 s^{-2} s, multiplying by s^{-1} and s. The kg m^2 s^{-2} is J, giving the required total unit of J s.

The question 'Does the electron have a position and momentum before we look for it?' can be debated. Physicists do not have a definitive answer and it depends on the interpretation of quantum mechanics that one adopts. However, this is not a useful question since the wavefunctions give us our information and there is a limit on what we can predict about the quantum state. We just have to accept this. It is worth reiterating that quantum mechanics gives superb agreement with experimental observations.

Using quantum mechanics the spectral lines for helium and other elements can be calculated and give excellent agreement with experimental observations. More importantly, quantum mechanics provides a justification for the ad hoc quantisation 'rules' introduced earlier and gives us a very useful tool to explain theoretically observed phenomena and make quantitative and accurate predictions about the outcomes of experiments.

Potential wells and quantum tunnelling

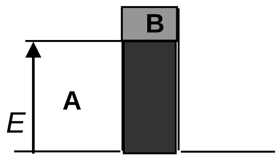
Imagine a ball in a 'dip'. The shape of the ups and downs is irrelevant.
The ball cannot get to position Y unless it receives energy $E = mgh$.



The ball is in a 'potential well' of 'height' mgh. This means that the ball needs energy E equal to or greater than mgh in order to 'escape' and get to position Y.

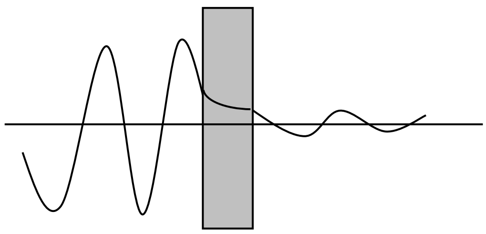
In the quantum world things are a touch different, although the concept of a 'potential well' or 'potential barrier' is useful.

Now let us consider an electron with some energy E on the left-hand side of a barrier of energy greater than E. The electron is thus confined to side A. It does not have enough energy to 'get over' the barrier and 'escape'.



Not so according to quantum theory.

The wavefunction is continuous across a barrier. The amplitude is greater in region A but it is finite, although much smaller, outside region A to the right of the barrier. Although the probability of finding the electron in region A is very high, there is a **finite** probability of finding the electron beyond the barrier.



The probability depends on the square of the amplitude. Hence it appears that the electron can 'tunnel out'. This is called **quantum tunnelling** and has some interesting applications.

Examples of quantum tunnelling

Alpha decay

For some radioactive elements, e.g. polonium 212, the alpha particles are held in the nucleus by the residual strong force and do not have enough energy to escape. However, because of quantum tunnelling they do escape and quantum mechanics can calculate the half-life. In 1928 George Gamow used quantum mechanics (Schrödinger wave equation) and the idea of quantum tunnelling to obtain a relationship between the half-life of the alpha particle and the energy of emission. Classically the alpha particle should not escape.

Scanning tunnelling microscope

A particular type of electron microscope, the scanning tunnelling microscope, has a small stylus that scans the surface of the specimen. The distance of the stylus from the surface is only about the diameter of an atom. Electrons 'tunnel' across the sample. In this way the profile of the sample can be determined. Heinrich Rohrer and Gerd Binnig were awarded the Nobel Prize for their work in this field in 1986.

Virtual particles

Another interesting effect of the Uncertainty Principle is the 'sea' of virtual particles in a vacuum. We might expect a vacuum to be 'empty'. Not so with quantum theory.

A particle can 'appear' with an energy ΔE for a time less than Δt where

$$\Delta E \Delta t \geq \frac{h}{4\pi}.$$

Do these virtual particles 'exist'? This is not really a sensible question for quantum mechanics. We cannot observe them in the short time of their existence. However, they are important as 'intermediate' particles in nuclear decays and high energy particle collisions and if they are omitted theoretical agreement with observations may not be obtained. Virtual particles are important when using Feynman diagrams to solve problems.

Particles from Space

Cosmic Rays

The term cosmic ray is not precisely defined, but a generally accepted description is ‘high energy particles arriving at the Earth which have originated elsewhere’. In the early 1900s, radiation was detected using an electroscope. However, radiation was still detected in the absence of known sources. This was background radiation.

Austrian physicist Victor Hess made measurements of radiation at high altitudes from a balloon, to try and get away from possible sources on Earth. He was surprised to find the measurements actually **increased** with altitude. At an altitude of 5000 m the intensity of radiation was found to be five times that at ground level.

Hess named this phenomenon cosmic radiation (later to be known as cosmic rays).

It was thought this radiation was coming from the Sun, but Hess obtained the same results after repeating his experiments during a nearly complete solar eclipse (12 April 1912), thus ruling out the Sun as the (main) source of radiation. In 1936 Hess was awarded the Noble Prize for Physics for the discovery of cosmic rays.

Tracks produced by cosmic rays can be observed using a cloud chamber. Charles T R Wilson is the only Scot ever to be awarded the Nobel Prize for Physics. He was awarded it in 1927 for the invention of the cloud chamber.

Robert Millikan coined the phrase ‘cosmic rays’, believing them to be electromagnetic in nature.

By measuring the intensity of cosmic rays at different latitudes (they were found to be more intense in Panama than in California), Compton showed that they were being deflected by the Earth’s magnetic field and so must consist of electrically charged particles, i.e. electrons or protons rather than photons in the form of gamma radiation.

Cosmic rays come in a whole variety of types, but the most common are protons, followed by helium nuclei. There is also a range of other nuclei as well as individual electrons and gamma radiation (see Table 1).

Table 1 Composition of cosmic rays

Nature	**Approximate percentage of all cosmic rays**
Protons	89
Alpha particles	9
Carbon, nitrogen and oxygen nuclei	1
Electrons	less than 1
Gamma radiation	less than 0.1

The energies of cosmic rays cover an enormous range, with the most energetic having energies much greater than those capable of being produced in current particle accelerators.

The highest energies produced in particle accelerators are of the order of 1 teraelectronvolt (10^{12} eV). Cosmic rays have been observed with energies ranging from 10^9 to 10^{20} eV. Those with energies above 10^{18} eV are referred to as ultra-high-energy cosmic rays (UHECRs).

The ‘Oh-my-God’ (OMG) particle with energy of 3×10^{20} eV was recorded in Utah in 1991.

Converting to joules (J), 3×10^{20} eV = $3 \times 10^{20} \times 1.6 \times 10^{-19}$ J = 48 J, ie ~50 J.

That is enough energy to throw a throw a 25 kg mass (e.g. a bag of cement) 2 m vertically upwards. It is also approximately equal to the kinetic energy of a tennis ball served at about 100 mph by Andy Murray.

Order of magnitude **open-ended question** opportunity here:
mass = 60 g = 0.06 kg, speed = 45 m s^{-1},
kinetic energy = $0.5 \times 0.06 \times 45 \times 45 = 60$ J

The OMG particle was probably a proton and as such had about **40 million** times the energy of the most energetic protons ever produced in an Earth-based particle accelerator.

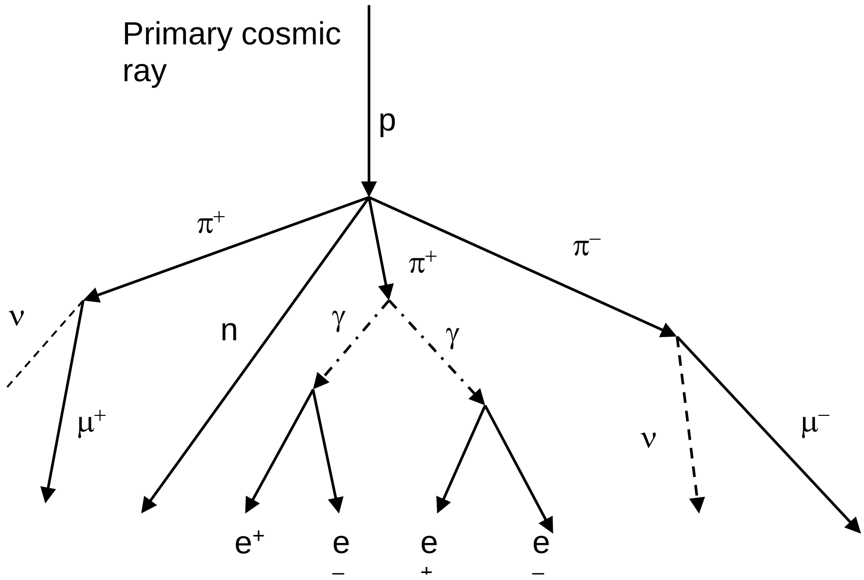
Such UHECRs are thought to originate from fairly local (in cosmological terms) distances, i.e. within a few hundred million light years. This is because were they to originate from further away it would be hard to understand how they get all the way here at all, since the chances are they would have interacted with Cosmic Microwave Background Radiation (CMBR) photons along the way, producing pions.

The lowest energy cosmic rays come from the Sun and the intermediate energy ones are presumed to be created within our galaxy, often in connection with supernovae. The main astrophysics (rather than particle physics) to come from the study of cosmic rays concerns supernovae since they are believed to be the main source of cosmic rays.

However, the origin of the highest energy cosmic rays is still uncertain. Active galactic nuclei (AGN) are thought to be the most likely origin for UHECRs. A group of cosmologists (including Martin Hendry from the University of Glasgow) are working on the statistical analysis of apparent associations between the incoming direction of the highest energy cosmic rays and active galaxies.

Interaction with the Earth's atmosphere
When cosmic rays reach the Earth, they interact with the Earth's atmosphere, producing a chain of reactions resulting in the production of a large number of particles known as a **cosmic air shower** (Figure 1). Air showers were first discovered by the French scientist Pierre Auger in 1938. Analysing these showers allows the initial composition and energies of the original (primary) cosmic rays to be deduced.

When cosmic rays from space (primary cosmic rays) strike particles in the atmosphere they produce secondary particles, which go on to produce more collisions and particles, resulting in a shower of particles that is detected at ground level. The primary cosmic rays can usually only be detected directly in space, for example by detectors on satellites, although very high energy cosmic rays, which occur on rare occasions, can penetrate directly to ground level.



p, proton; e^-, electron; e^+ positron;
ν neutrino;

Figure 1 *Air shower.*

Detection
Consequently there are two forms of detector: those that detect the air showers at ground level and those located above the atmosphere that detect primary cosmic rays.

Cherenkov radiation
Air shower particles can travel at relativistic speeds. Although relativity requires that nothing can travel faster than the speed of light in a vacuum, particles may exceed the speed of light in a particular medium, for example water. Such particles then emit a beam of Cherenkov radiation – the radiation that causes the characteristic blue colour in nuclear reactors. (This is a bit like the optical equivalent of a sonic boom.)

Atmospheric fluorescence
When charged particles pass close to atoms in the atmosphere, they may temporarily excite electrons to higher energy levels. The photons emitted when the electrons return to their previous energy levels can then be detected.

The Pierre Auger observatory in Argentina was set up to study high-energy cosmic rays. It began operating in 2003 and at that time was the largest physics experiment in the world. It is spread over several thousand square miles and uses two basic types of detectors:

- 1600 water tanks to detect the Cherenkov effect
- four detectors of atmospheric fluorescence.

The Solar Wind and Magnetosphere

Structure of the Sun
The interior of the Sun consists of three main regions:
1. the **core**, within which nuclear fusion takes place
2. the **radiative zone**, through which energy is transported by photons
3. the **convective zone**, where energy is transported by convection.

The extended and complex solar atmosphere begins at the top of the convective zone, with the **photosphere**.

The photosphere is the visible surface of the Sun and appears smooth and featureless, marked by occasional relatively dark spots, called sunspots. Moving outwards, next is the **chromosphere**. Sharp spicules and prominences emerge from the top of the chromosphere.

The **corona** (from the Greek for *crown*) extends from the top of the chromosphere. The corona is not visible from Earth during the day because of the glare of scattered light from the brilliant photosphere, but its outermost parts are visible during a total solar eclipse.

The depth of each layer relative to the radius of the Sun (R_S) is shown in Figure 2. The photosphere is about 330 km deep ($0.0005R_S$) and the chromosphere is about 2000 km ($0.003R_S$) deep.

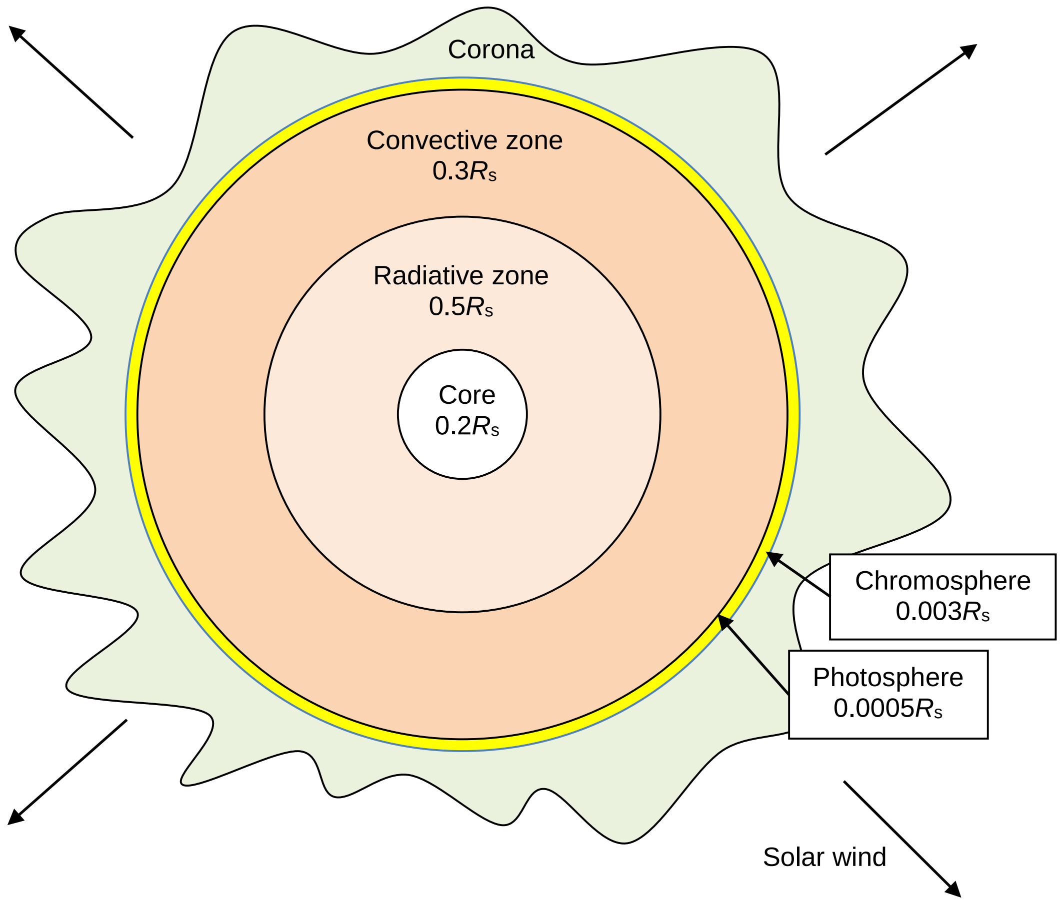


Figure 2 *Structure of the Sun.*

Coronagraphs are special telescopes that block out the light from the photosphere to allow the corona to be studied. These are generally used from mountain tops (where the air is thin) and from satellites. They have been able to detect the corona out beyond 20R_S, which is more than 10% of the way to Earth.

The corona is permeated by magnetic fields. In particular there are visible loops along which glowing ionised gaseous material can be seen to travel. They have the shape of magnetic field lines and begin and end on the photosphere.

Information about magnetic fields in the corona has come from the study of emitted X-rays, obtained from satellites and space stations. The corona is the source of most of the Sun's X-rays because of its high temperature, which means it radiates strongly at X-ray wavelengths. The corona's X-ray emission is not even, however, with bright patches and dark patches. The dark areas hardly emit any X-rays at all and are called **coronal holes**.

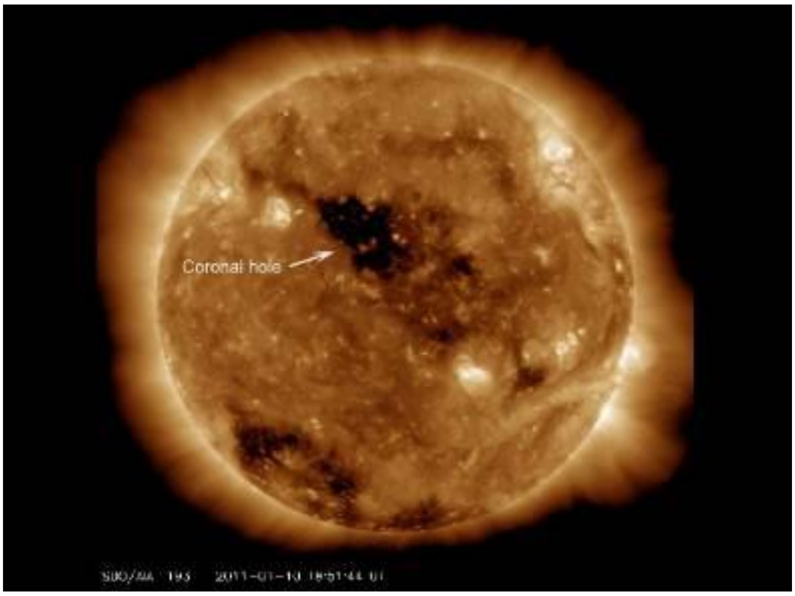


Figure 3 *Coronal hole.*

The Solar and Heliospheric Observatory (SOHO) is a project of international collaboration between the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA) to study the Sun from its deep core to the outer corona and the solar wind.

In 2006 a rocket was launched from Cape Canaveral carrying two nearly identical spacecraft. Each satellite was one half of a mission entitled Solar TErrestrial RElations Observatory (STEREO) and they were destined to do something never done before – observe the whole of the Sun simultaneously. With this new pair of viewpoints, scientists will be able to see the structure and evolution of solar storms as they blast from the Sun and move out through space.

The solar wind
There is a continual flow of charged particles emanating from the Sun because of the high temperature of the corona. This gives some particles sufficient kinetic energy to escape from the Sun's gravity. This flow is called the **solar wind** and is plasma composed of approximately equal numbers of protons and electrons (i.e. ionised hydrogen). It can be thought of as an extension of the corona itself and as such reflects its composition. The solar wind also contains about 8% alpha particles (i.e. helium nuclei) and trace amounts of heavy ions and nuclei (C, N, O, Ne, Mg, Si, S and Fe).

The solar wind travels at speeds of between 300 and 800 kms^{-1}, with gusts recorded as high as 1000 km^{-1} (2.2 million miles per hour).

Comet tails
Although it was known that solar eruptions ejected material that could reach the Earth, no-one suspected that the Sun was continually losing material regardless of its apparent activity. It had been known for a long time that comet tails always pointed away from the Sun, although the reason was unknown. Ludwig Biermann (of the Max Planck Institute for Physics in Göttingen) made a close study of the comet Whipple-Fetke, which appeared in 1942. It had been noted that comet tails did not point directly away from the Sun. Biermann realised this could be explained if the comet was moving in flow of gas streaming away from the Sun. The comet's tail was acting like a wind-sock, In the early 1950s Biermann concluded that even when the Sun was quiet, with no eruptions or sunspots, there was still a continuous flow of gas from it.

In 1959, the Russian space probe Luna 1 made the first direct observation and measurement of the solar wind. The probe carried different sets of scientific devices for studying interplanetary space, including a magnetometer, Geiger counter, scintillation counter and micrometeorite detector. It was the first man-made object to reach escape velocity from Earth.

Coronal holes
The magnetic field lines from coronal holes don't loop back onto the surface of the corona. Instead they project out into space like broken rubber bands, allowing charged particles to spiral along them and escape from the Sun. There is a marked increase in the solar wind when a coronal hole faces the Earth.

Solar flares
Solar flares are explosive releases of energy that radiate energy over virtually the entire electromagnetic spectrum, from gamma rays to long wavelength radio waves. They also emit high-energy particles called **solar cosmic rays**. These are composed of protons, electrons and atomic nuclei that have been accelerated to high energies in the flares. Protons (hydrogen nuclei) are the most abundant particles followed by alpha particles (helium nuclei). The electrons lose much of their energy in exciting radio bursts in the corona. These generally occur near sunspots, which leads to the suggestion they are magnetic phenomena. It is thought that magnetic field lines become so distorted and twisted that they suddenly snap like rubber bands. This releases a huge amount of energy, which can heat nearby plasma to 100 million kelvin in a few minutes or hours. This generates X-rays and can accelerate some charged particles in the vicinity to almost the speed of light.

The energies of solar cosmic ray particles range from millielectronvolts (10^{-3} eV) to tens of gigaelectronvolts (10^{10} eV).

The highest energy particles arrive at the Earth within half an hour of the flare maximum, followed by the peak number of particles 1 hour later.

Particles streaming from the Sun after solar flares or other major solar events can disrupt communications and power delivery on Earth.

A major solar flare in 1989:
- caused the US Air Force to temporarily lose communication with over 2000 satellites
- induced currents in underground circuits of the Quebec hydroelectric system, causing it to be shut down for more than 8 hours.

The solar cycle
All solar activities show a cyclic variation with a period of about 11 years.

Sunspots
When an image of the sun is focussed on a screen dark spots called sunspots are often visible. By observing over a number of days they will be seen to move (due to the rotation of the Sun) and also change in size, growing or shrinking. The sunspots look dark because they are cooler than the surrounding photosphere. A large group of sunspots is called an **active region** and may contain up to 100 sunspots.

The general pattern of the movement of sunspots (individual sunspots may appear and disappear – short-lived ones only lasting a few hours whereas others may last for several months) shows that the Sun is rotating with an average period of about 27 days with its axis of rotation tilted slightly to the plane of the Earth's orbit.

Unlike the Earth, the Sun does not have a single rotation period. The period is 25 days at the Sun's equator and lengthens to 36 days near the poles. Sections at different latitudes rotate at different rates and so this is called **differential rotation**.

The three main features of the solar cycle are:

1. the number of sunspots
2. the mean latitude of sunspots
3. the magnetic polarity pattern of sunspot groups.

The number of sunspots increases and decreases with an 11-year cycle, the mean solar latitude at which the sunspots appear progresses towards the solar equator as the cycle advances and the magnetic polarity pattern of sunspot groups reverses around the end of each 11-year cycle (making the full cycle in effect 22 years).

The magnetosphere
The magnetosphere is the part of the Earth's atmosphere dominated by the Earth's magnetic field. This region also contains a diffuse plasma of protons and electrons. The magnetic field resembles that of a bar magnet, tipped at about 11° to the Earth's rotational axis. However, the magnetic field is believed to be generated by electric currents in conducting material inside the Earth, like a giant dynamo. Geological evidence shows that the direction of the Earth's magnetic field has reversed on several occasions, the most recent being about 30,000 years ago. This lends evidence for the 'dynamo' model as the reversal can be explained in terms of changes in the flow of conducting fluids inside the Earth.

Interaction of the solar wind with the Earth's magnetic field

The solar wind interacts with the magnetosphere and distorts its pattern from the simple bar magnet model outlined above. The Earth's magnetic field also protects it from the solar wind, deflecting it a bit like a rock deflecting the flow of water in a river. The boundary where the solar wind is first deflected is called the bow shock. The cavity dominated by the Earth's magnetic field is the **magnetosphere**, see Figures 4 and 5.

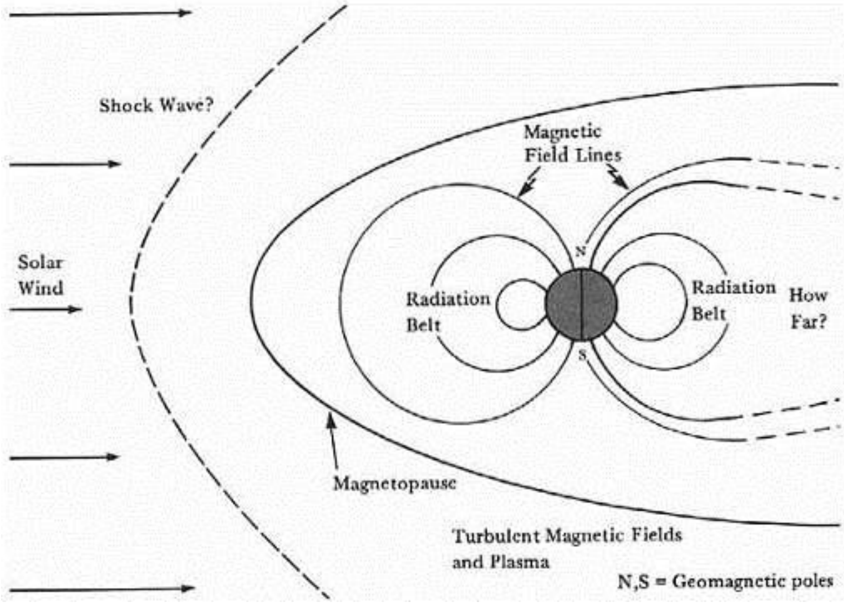


Figure 4 *The magnetosphere as visualised in 1962.*

High-energy particles from the solar wind that leak into the magnetosphere and become trapped to form the Van Allen belts of radiation. These are toroidal in shape and concentric with the Earth's magnetic axis. There are two such belts: the inner and the outer.

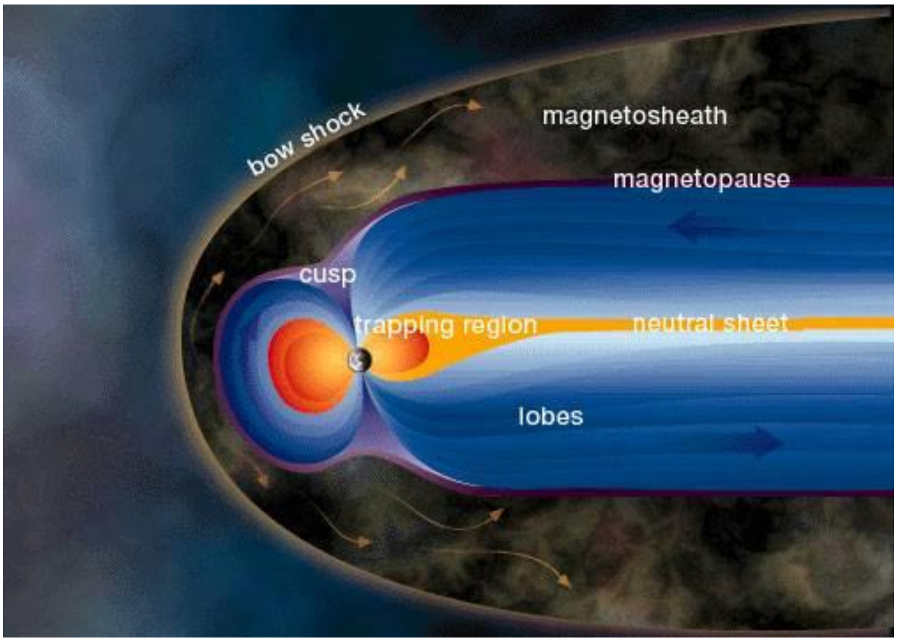


Figure 5 *Current perception of the Earth's magnetosphere.*

The inner Van Allen belt lies between one and two Earth radii from its axis, (R_E< inner belt < $2R_E$) then there is a distinct gap followed by the outer belt lying between three and four Earth radii ($3R_E$ < outer belt < $4R_E$). The inner belt traps protons with energies of between 10 and 50 MeV and electrons with energies greater than 30 MeV. The outer belt contains fewer energetic protons and electrons.

The charged particles trapped in the belts spiral along magnetic field lines and oscillate back and forth between the northern and southern mirror points with periods between 0.1 and 3 seconds as shown in Figure 6

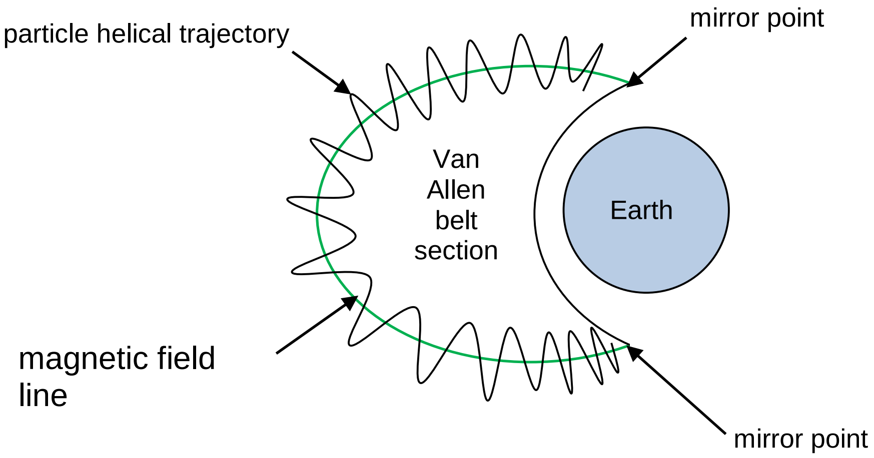


Figure 6 *Van Allen belt.*

Particles in the inner belt may interact with the thin upper atmosphere to produce the aurorae. These result from the excitation of different atoms in the atmosphere, each of which produces light with a characteristic colour due to the different energy associated with that atomic transition.

Charged particles in a magnetic field

The force acting on a charge q, moving with velocity v through a magnetic field B is given by:

$$F = qvB$$

where F, v and B are all mutually at right angles to each other.

Circular motion

As F is always at right angles to v, the particle will move with uniform motion in a circle, where F is the central force (assuming any other forces are negligible), so:

$$F = \frac{mv^2}{r}$$

Equating the magnetic force to the central force we get:

$$\frac{mv^2}{r} = qvB$$

so

$$r = \frac{mv}{qB}$$

Helical motion

If a charged particle crosses the magnetic field lines at an angle, then its velocity can be resolved into two orthogonal components: one perpendicular to the field and the other parallel to it.

The perpendicular component provides the central force, which produces uniform circular motion as shown above. The component parallel to the magnetic field does not cause the charge to experience a magnetic force so it continues to move with constant velocity in that direction, resulting in a helical path.

This can be illustrated using a dual-beam electron tube, with a coil positioned at the front of the tube so it produces an axial magnetic field, see Figures 7 and 8.

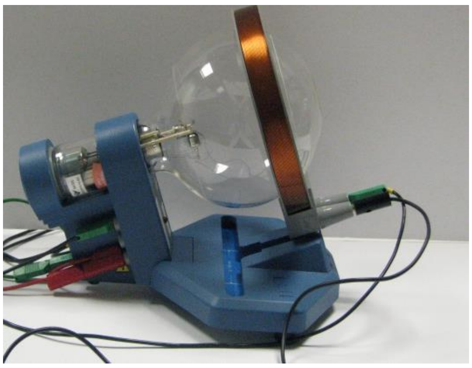


Figure 7 *Coil at front of dual beam tube.*

Figure 8 *Helical motion of electron beam.*

Aurorae

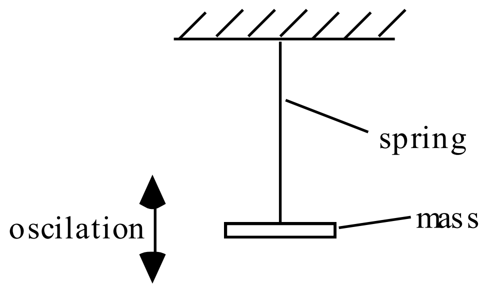
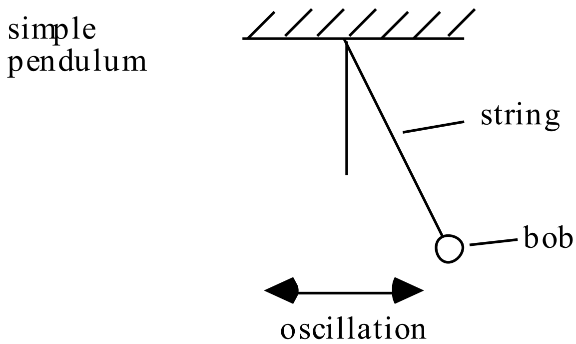
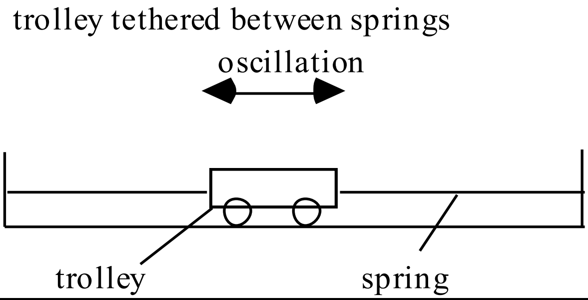
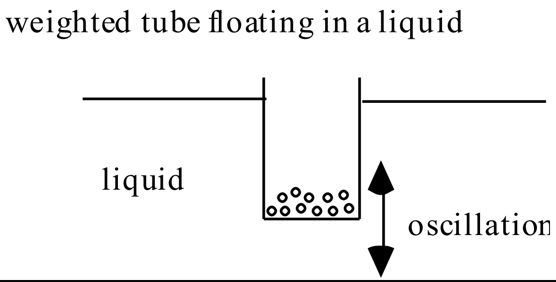
The aurora (aurora borealis in the Northern hemisphere – the northern lights; aurora australis in the southern hemisphere – southern lights) are caused by solar wind particles which penetrate the Earth's upper atmosphere, usually within 20° of the north or south poles. Between 80 and 300 km above the Earth's surface (aircraft fly at around 10 km altitude) these particles strike nitrogen molecules and oxygen atoms, causing them to become excited and subsequently emit light in the same way as happens in electric discharge lamps. The most common colours, red and green, come from atomic oxygen, and violets come from molecular nitrogen.

Figure 9 *Aurora borealis.*

Simple Harmonic Motion (SHM)
If an object is subject to a linear restoring force, it performs an oscillatory motion termed 'simple harmonic'. Before a system can perform oscillations it must have (1) a means of storing potential energy and (2) some mass which allows it to possess kinetic energy. In the oscillating process, energy is continuously transformed between potential and kinetic energy.

Note: any motion which is periodic and complex (i.e. not simple!) can be analysed into its simple harmonic components (Fourier Analysis). An example of a complex waveform would be a sound wave from a musical instrument.

Examples of SHM

Example and Diagram	E_p stored as:	E_k possesed by moving:
mass on a coil spring 	elastic energy of spring	mass on spring
	potential energy (gravitational) of bob	mass of the bob
	elastic energy of the springs	mass of the trolley
	potential energy (gravitational) of the tube	mass of the tube

Note that for the mass oscillating on the spring, there is always an **unbalanced** force acting on the mass and this force is always **opposite** to its direction of motion. The unbalanced force is momentarily zero as the mass passes through the rest position.

To see this, consider the following: when the mass is moving upwards past the rest position, the gravitational force (**downwards**) is greater than the spring force. Similarly when moving downwards past the rest position, the spring force (**upwards**) is greater than the gravitational force downwards.

This situation is common to all SHMs. The force which keeps the motion going is therefore called the **restoring** force.

Definition of Simple Harmonic Motion

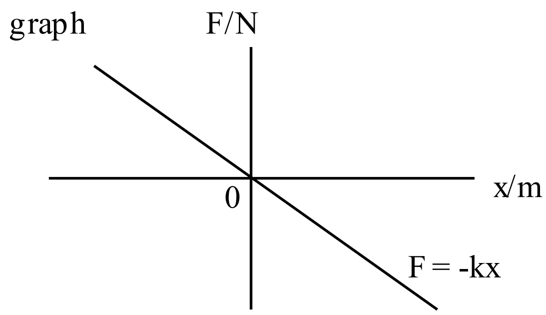
When an object is displaced from its equilibrium or at rest position, and the unbalanced force is proportional to the displacement of the object and acts in the opposite direction, the motion is said to be simple harmonic.

Graph of Force against displacement for SHM

$$F = -kx$$

F is the restoring force (N)
k is the force constant (N m^{-1})
x is the displacement (m)

The negative sign shows the direction of vector F is always opposite to vector x.



If we apply Newton's Second Law in this situation the following alternative definition in terms of acceleration as opposed to force is produced.

$$F = ma = m\frac{d^2x}{dt^2} = -kx \qquad a = -\frac{k}{m}x \quad \text{thus} \quad \frac{d^2x}{dt^2} = -\frac{k}{m}x$$

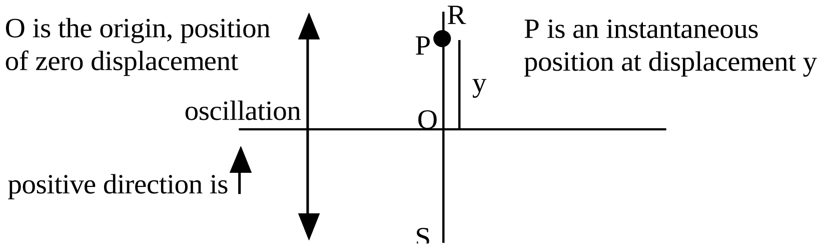
Remember that k is a force constant which relates to the oscillating system.
The constant, $\frac{k}{m}$ is related to the period of the motion by $\omega^2 = \frac{k}{m}$, $\omega = \frac{2\pi}{T}$

This analysis could equally well have been done using the y co-ordinate.

Thus an equivalent expression would be $\frac{d^2y}{dt^2} = -\omega^2 y$.

Kinematics of SHM

Point P is oscillating with SHM between two fixed points R and S. The amplitude of the oscillation is therefore ½ RS and this is given the symbol a. The displacement y is the vector OP.



The period, T, of the motion is the time taken to complete one oscillation, e.g. path O->R->O ->S->O.

The frequency, f, is the number of oscillations in one second.

$\boxed{f = \frac{1}{T}}$ and because $\omega = \frac{2\pi}{T}$ $\boxed{\omega = 2\pi f}$

Solutions of Equation for SHM

The equation $\frac{d^2y}{dt^2} = -\omega^2 y$ could be solved using integration to obtain equations for velocity v and displacement y of the particle at a particular time t. However, the calculus involves integration which is not straightforward. We will therefore start with the solutions and use differentiation. The possible solutions for the displacement y at time t depend on the **initial conditions** and are given by:

y = a cos ωt, if y = 0 at t = 0	and	y = a sin ωt, if y = a at t = 0

Acceleration

$$\frac{dy}{dt} = \frac{d}{dt} a\cos\omega t = -a\omega\sin\omega t$$

$$\frac{d^2y}{dt^2} = \frac{d}{dt} - a\omega\sin\omega t = -a\omega^2\cos\omega t$$

As $y = a\cos\omega t ... so ... \frac{d^2y}{dt^2} = -\omega^2 y$

$$\frac{dy}{dt} = \frac{d}{dt} a\sin\omega t = a\omega\cos\omega t$$

$$\frac{d^2y}{dt^2} = \frac{d}{dt} a\omega\cos\omega t = -a\omega^2\sin\omega t$$

As $y = a\sin\omega t ... so ... \frac{d^2y}{dt^2} = -\omega^2 y$

Velocity

$$\frac{dy}{dt} = v = \frac{d}{dt} a\cos\omega t = -a\omega\sin\omega t$$

$v^2 = a^2\omega^2\sin^2\omega t$ and $y^2 = a^2\cos^2\omega t$

$\sin^2\omega t + \cos^2\omega t = 1$ so $\frac{v^2}{a^2\omega^2} + \frac{y^2}{a^2} = 1$

Thus $v^2 = \omega^2(a^2 - y^2)$

So $v = \pm\omega\sqrt{a^2 - y^2}$

$$\frac{dy}{dt} = v = \frac{d}{dt} a\sin\omega t = a\omega\cos\omega t$$

$v^2 = a^2\omega^2\cos^2\omega t$ and $y^2 = a^2\sin^2\omega t$

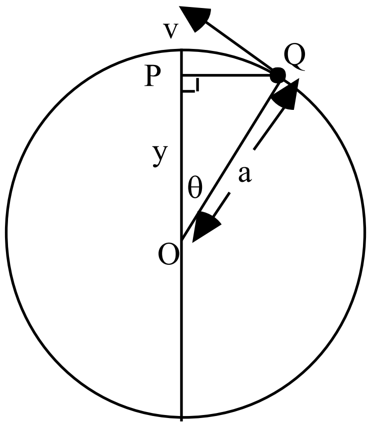
$\sin^2\omega t + \cos^2\omega t = 1$ so $\frac{v^2}{a^2\omega^2} + \frac{y^2}{a^2} = 1$

Thus $v^2 = \omega^2(a^2 - y^2)$

So $v = \pm\omega\sqrt{a^2 - y^2}$

Linking SHM with Circular Motion

This allows us to examine the mathematics of the motion and is provided for interest. If the point Q is moving at constant speed, v, in a circle, its projection point P on the y axis will have **displacement** y = a cos ωt



radius OQ sweeps out ω rad s^{-1}

positive direction of y is upwards

note that $\sin\theta = \dfrac{QP}{OQ}$

$\sin\theta = \dfrac{\sqrt{a^2 - y^2}}{a}$

The **velocity** of point P is: $v_p = \dfrac{dy}{dt} = \dfrac{d}{dt} a\cos\omega t = -a\omega\sin\omega t$ (assume P is moving down)

when y = 0, θ = $\dfrac{\pi}{2}$ and sin θ = 1 $\boxed{v_{\max} = \pm a\omega}$

This occurs as P goes through the origin in either direction.

when y = ± a, θ = 0 or π and sin θ = 0 $\boxed{v_{\min} = 0}$

This occurs as P reaches the extremities of the motion.

The **acceleration** of point P is: $a_p = \dfrac{d^2y}{dt^2} = \dfrac{d}{dt} - a\omega\sin\omega t = -a\omega^2\cos\omega t$

when y = 0, θ = $\dfrac{\pi}{2}$ and cos θ = 0 $\boxed{a_{\min} = 0}$

This occurs as P goes through the origin in either direction.

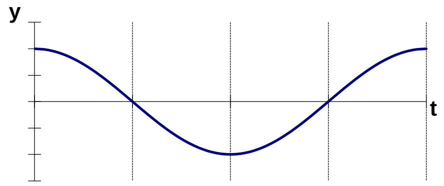
when y = ±a, θ = 0 or π and cos θ = 1 $\boxed{a_{\max} = \pm a\omega^2}$

This occurs as P reaches the extremities of the motion.

The acceleration is negative when the displacement, y, is positive and vice versa; i.e. they are out of phase, see graphs of motion below. Knowledge of the positions where the particle has maximum and minimum acceleration and velocity **is required**

To understand these graphs it is helpful if you see such graphs being generated using a motion sensor. In particular, pay close attention to the phases of the graphs of the motion and note that the basic shape is that of the sine/cosine graphs.

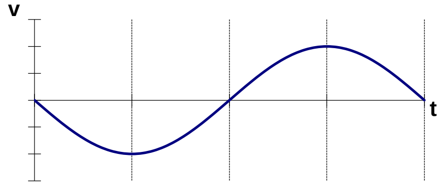
Displacement-time



Summary of Equations

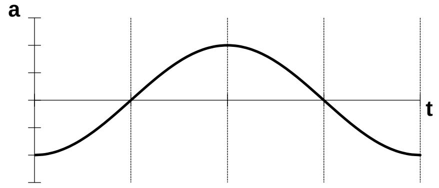
$$\boxed{y = a\cos\omega t}$$

Velocity-time



$$\boxed{v = \pm\omega\sqrt{a^2 - y^2}}$$

Acceleration-time



$$\boxed{a = \frac{d^2y}{dt^2} = -\omega^2 y}$$

Note that this form, acceleration = - ω^2 y, is consistent with our definition of SHM ω^2 is a positive constant. This implies that the sine and cosine equations must be solutions of the motion.

Compare this constantly **changing** acceleration with situation where only **uniform** acceleration was considered.
The equation used in a particular situation **depends on the initial conditions.**

Thus: if y = 0 at time t = 0 use $y = a \sin \omega t$
if y = a at time t = 0 use $y = a \cos \omega t$

Another possible solution for SHM is: $y = a \sin(\omega t + \phi)$ where ϕ is known as the phase angle.

Example

An object is vibrating with simple harmonic motion of amplitude 0.02 m and frequency 5.0 Hz. Assume that the displacement of the object, y = 0 at time, t = 0 and that it starts moving in the positive y-direction.

(a) Calculate the maximum values of velocity and acceleration of the object.

(b) Calculate the velocity and acceleration of the object when the displacement is 0.008 m.

(c) Find the time taken for the object to move from the equilibrium position to a displacement of 0.012 m.

Solution

Initial conditions require; $y = a \sin \omega t$; $v = a\omega \cos \omega t$; and $acc = -\omega^2 x$

$f = 5$ Hz $\quad \omega = 2\pi f = 31.4$ rad s^{-1}

(a) $v_{max} = \omega a = 31.4 \times 0.02 = 0.63$ m s^{-1}

$acc_{max} = -\omega^2 a = -(31.4)^2 \times 0.02 = -19.7$ m s^{-2}

(b) $v = \pm \omega \sqrt{a^2 - y^2} = \pm 31.4 \sqrt{0.02^2 - 0.008^2} = \pm 0.58$ m s^{-1}

$acc = -\omega^2 y = -31.4^2 \times 0.008 = -7.9$ m s^{-2}

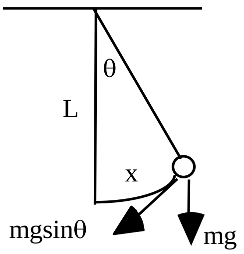
(c) use $y = a \sin \omega t$; $0.012 = 0.02 \sin 31.4t$ (when $y = 0.012$ m)

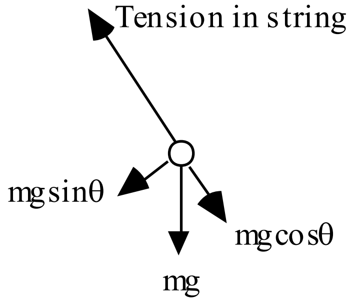
$\sin 31.4\,t = \frac{0.012}{0.02} = 0.6$ giving $31.4\,t = 0.644$ and $t = \frac{0.644}{31.4}$

Thus $t = 0.0205$ s (Remember that angles are in radians)

Proof that the Motion of a Simple Pendulum approximates to SHM

The sketches below show a simple pendulum comprising a point mass, m, at the end of an inextensible string of length L. The string has negligible mass.





The restoring force F on the bob is $F = -mg \sin\theta$

If the **angle θ is small** (less than about 10°) then $\sin\theta = \theta$ in radians and $\theta = \frac{x}{L}$

Then $F = -mg\theta = -mg\frac{x}{L}$ Thus $F = -\frac{mg}{L}x$

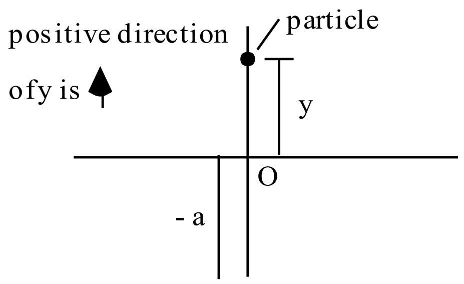
The restoring force therefore satisfies the conditions for SHM for small displacements.

Then acceleration is $a = -\frac{g}{L}x$ which if compared with $a = -\omega^2 x$ gives $\omega^2 = \frac{g}{L}$ ($\omega = 2\pi f$)

$f = \frac{1}{2\pi}\sqrt{\frac{g}{L}}$ and the period of the pendulum $\boxed{T = 2\pi\sqrt{\frac{L}{g}}}$

Energy Equations for SHM

Consider the particle moving with simple harmonic motion below. The particle has maximum amplitude a and period T = $\frac{2\pi}{\omega}$.



Kinetic energy equation for the particle

$E_k = \frac{1}{2}$ m v^2 $= \frac{1}{2}$ m $[\pm \omega \sqrt{a^2 - y^2}\ ]^2$ => $\boxed{E_k = \frac{1}{2} m\ \omega^2 (a^2 - y^2)}$

Potential energy equation for the particle

When at position O the potential energy is zero, (with reference to the equilibrium position) and the **kinetic energy is a maximum**.

The kinetic energy is a maximum when y = 0: $E_{k_{max}} = \frac{1}{2}$ m $\omega^2 a^2$

At point O, total energy E = $E_k + E_p = \frac{1}{2}$ m $\omega^2 a^2 + 0$

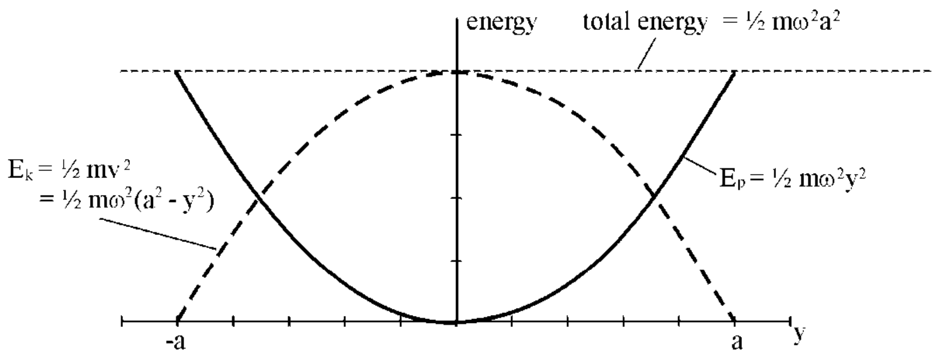
$$E = \frac{1}{2} m\ \omega^2 a^2 \qquad \text{or} \quad E = \frac{1}{2} k\ a^2 \qquad \text{because } \omega^2 = \frac{k}{m}$$

The **total energy E** is the **same at all points** in the motion.

Thus for any point on the swing: as above $E = E_k + E_p$

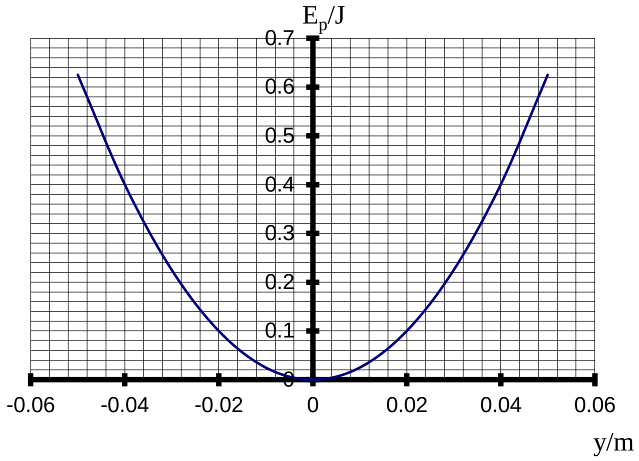
$\frac{1}{2}$ m $\omega^2 a^2 = \frac{1}{2}$ m $\omega^2 (a^2 - y^2) + E_p$ => $\boxed{E_p = \frac{1}{2} m\ \omega^2 y^2}$

The graph below shows the relation between potential energy, E_p, kinetic energy E_k, and the total energy of a particle during SHM as amplitude y changes from - a to + a.



Example on energy and SHM

The graph below shows how the potential energy, E_p, of an object undergoing SHM, varies with its displacement, y. The object has mass 0.40 kg and a maximum amplitude of 0.05 m.



(a) (i)Find the potential energy of the object when it has a displacement of 0.02 m.

(ii)Calculate the force constant, k for the oscillating system. (k should have unit N m^{-1}).

(b) Find the amplitude at which the potential energy equals the kinetic energy.

Solution

(a) (i) *From graph* $E_p = 0.10$ J

(ii) $E_p = \frac{1}{2} k y^2$

$0.1 = \frac{1}{2} k (0.02)^2$

$k = \frac{0.2}{(0.02)^2} = 500 \text{ N m}^{-1}$

(b) $E_p = E_k$

$\frac{1}{2} k y^2 = \frac{1}{2} m \omega^2 (a^2 - y^2)$

$= \frac{1}{2} k (a^2 - y^2)$ since $\omega^2 = \frac{k}{m}$

$y^2 = a^2 - y^2$ or $2 y^2 = a^2$

$y = \frac{a}{\sqrt{2}}$ when $E_p = E_k$

$y = \frac{0.05}{\sqrt{2}} = 0.035$ m

Damping of Oscillations

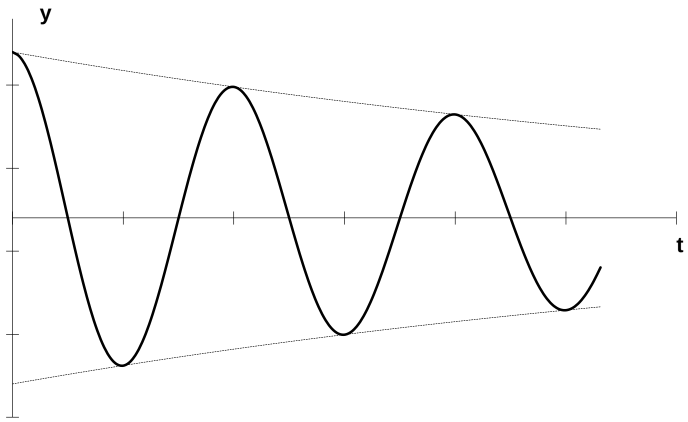
Oscillating systems, a mass on a spring, a simple pendulum, a bobbing mass in water, all come to rest eventually. We say that their motion is **damped**. This means that the amplitude of the motion decreases to zero because energy is transformed from the system. A simple pendulum takes a long time to come to rest because the frictional effect supplied by air resistance is small - we say that the pendulum is lightly damped. A tube oscillating in water comes to rest very quickly because the friction between the container and the water is much greater - we say that the tube is heavily damped.

If the damping of a system is increased there will be a value of the frictional resistance which is just sufficient to prevent any oscillation past the rest position - we say the system is **critically** damped. Systems which have a very large resistance, produce no oscillations and take a long time to come to rest are said to be **overdamped**. In some systems overdamping could mean that a system takes longer to come to rest than if underdamped and allowed to oscillate a few times.

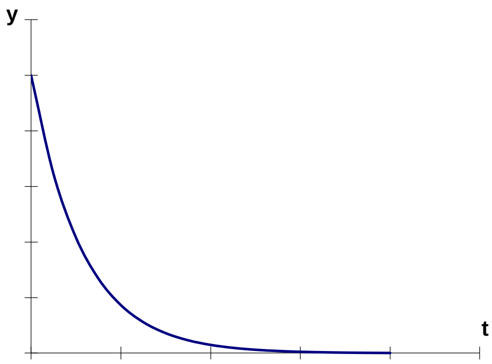
An example of damped oscillations is a car shock absorber which has a very thick oil in the dampers. When the car goes over a bump, the car does not continue to bounce for long. Ideally the system should be critically damped. As the shock absorbers get worn out the bouncing may persist for longer.

The graphs below give a graphical representation of these different types of damping.

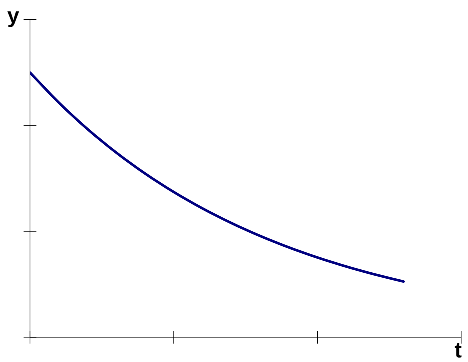
Damped oscillations



Critically damped



Overdamped



Waves

Wave Motion

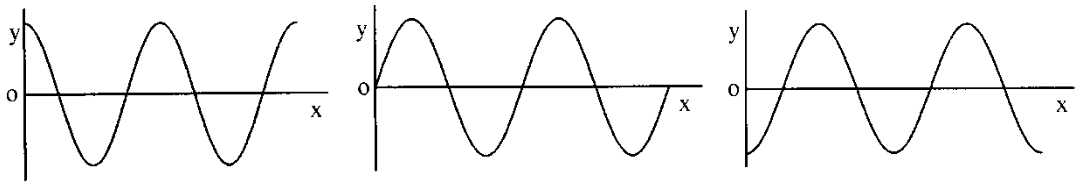
In a wave motion energy is transferred from one position to another with no net transport of mass.

Consider a water wave where the movement of each water particle is at right angles (transverse) to the direction of travel of the wave. During the wave motion each particle, labelled by its position on the x-axis, is displaced some distance y in the transverse direction. In this case, "no net transfer of mass" means that the water molecules themselves do not travel with the wave - the wave energy passes over the surface of the water, and in the absence of a wind/tide any object on the surface will simply bob up and down.

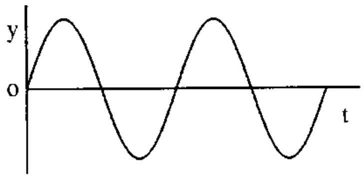
The Travelling Wave Equation

The value of the displacement y depends on which particle of the wave is being considered. It is dependent on the x value, and also on the time t at which it is considered. Therefore y is a function of x and t giving $y = f(x,t)$. If this function is known for a particular wave motion we can use it to predict the position of any particle at any time.

Below are 'snapshots' of a transverse wave taken at different times showing how the displacement of different particles varies with position x.



The following diagram shows the movement of **one** particle on the wave as a function of time.



Initial condition at the origin:
$y = 0$ when $t = 0$.

For a wave travelling from **left to right** with speed v, the particle will be performing SHM in the y-direction.

The equation of motion of the particle will be:

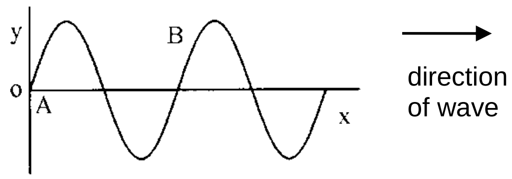
$y = a \sin \omega t$ where a is the amplitude of the motion.

The displacement of the particle is simple harmonic. The sine or cosine variation is the simplest description of a wave.

When $y = 0$ at $t = 0$ the relationship for the wave is $y = a \sin \omega t$, as shown above.
When $y = a$ at $t = 0$ the relationship for the wave is $y = a \cos \omega t$.

Deriving the travelling wave equation

Consider a snapshot of the wave as shown below.



The time, t, for the wave disturbance to travel from A (x = 0) to B (x = x) is $\frac{x}{v}$.

Consider particle A at position x = 0.
The equation of motion of particle A is given by $y = a \sin \omega t$
where t is the time at which the motion of particle A is observed.

Now consider particle B at position x = x and the time t = t.
Since wave motion is a repetitive motion:

motion of particle B (x = x, t = t) = motion of particle A (x = 0, t = $\frac{x}{v}$),

[i.e. the motion of particle B = motion of particle A at the **earlier** time of t = $\frac{x}{v}$].

General motion of particle A is given by $y = a \sin \omega t$, but in this case $t = t - \frac{x}{v}$

hence $y = a \sin \omega(t - \frac{x}{v})$.

Motion of particle B (x = x, t = t) is also given by $y = a \sin \omega(t - \frac{x}{v})$.

In general: $y = a \sin \omega(t - \frac{x}{v})$ also $\omega = 2\pi f$ and $v = f\lambda$

$y = a \sin 2\pi f(t - \frac{x}{f\lambda})$ which gives

$$\boxed{y = a \sin 2\pi \left(ft - \frac{x}{\lambda}\right)}$$

for a wave travelling from left to right in the **positive** x-direction.

The equation of a wave travelling right to left in the **negative** x-direction is

$$y = a \sin 2\pi\left(ft + \frac{x}{\lambda} \right).$$

The Intensity of a Wave

The intensity of a wave is directly proportional to the square of its amplitude.

$$\boxed{\text{intensity} \propto a^2}$$

Longitudinal and transverse waves

With transverse waves, as in water waves, each particle oscillates at right angles to the direction of travel of the wave. In longitudinal waves, such as sound waves, each particle vibrates along the direction of travel of the wave.

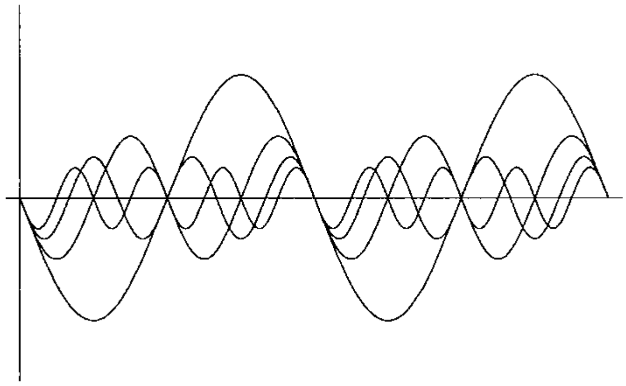
Principle of Superposition of Waveforms

Travelling waves can pass through each other without being altered. If two stones are dropped in a calm pool, two sets of circular waves are produced. These waves pass through each other. However at any point at a particular time, the disturbance at that point is the algebraic sum of the individual disturbances. In the above example, when a 'trough' from one wave meets a 'crest' from the other wave the water will remain calm.

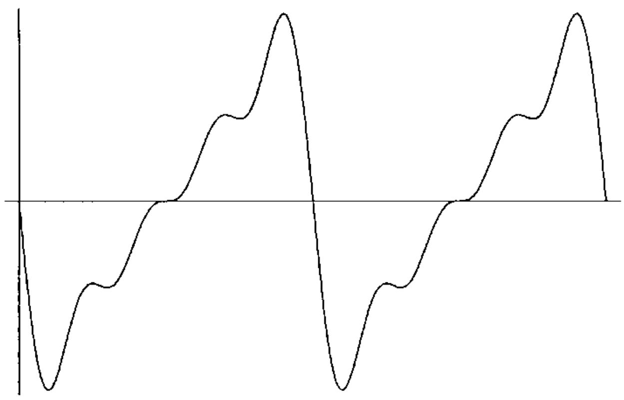
A **periodic** wave is a wave which repeats itself at regular intervals. All periodic waveforms can be described by a mathematical series of sine or cosine waves, known as a Fourier Series. For example a saw tooth wave can be expressed as a series of individual sine waves.

$$y(t) = -\frac{1}{\pi}\sin \omega t - \frac{1}{2\pi}\sin 2\omega t - \frac{1}{3\pi}\sin 3\omega t - \ldots\ldots\ldots\ldots$$

The graph below shows the first four terms of this expression.



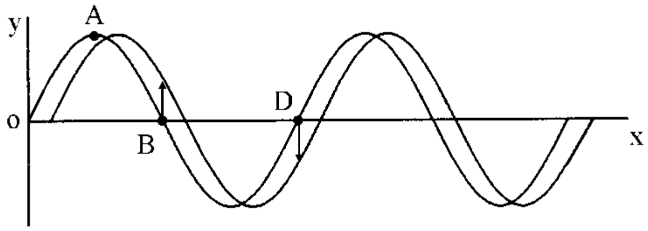
When all these terms are superimposed (added together) the graph below is obtained. Notice that this is tending to the sawtooth waveform. If more terms are included it will have a better saw tooth form.



Phase Difference

A phase difference exists between two points on the same wave.

Consider the snapshots below of a wave travelling to the right in the positive x-direction.



Points O and D have a phase difference of 2π radians.
They are both at zero displacement and will next be moving in the negative direction.
They are separated by one wavelength (λ).

Points O and B have a phase difference of π radians.
They both have zero displacement but B will next be going positive and O will be going negative. They are separated by $\lambda/2$. Notice that points A and B have a phase difference of $\lambda/2$.

The table below summarises phase difference and separation of the points.

Phase difference	Separation of points
0	0
$\pi/2$	$\lambda/4$
π	$\lambda/2$
2π	λ

Notice that $\dfrac{\text{phase difference}}{\text{separation of points}} = \dfrac{2\pi}{\lambda}$ = constant.

If the phase difference between two particles is ϕ when the separation of the particles is x, then $\dfrac{\phi}{x} = \dfrac{2\pi}{\lambda}$.

In general, for two points on a wave separated by a distance x the phase difference is given by:

$$\boxed{\phi = 2\pi\frac{x}{\lambda}}$$

where ϕ is the phase angle in radians

Example

A travelling wave has a wavelength of 60 mm. A point P is 75 mm from the origin and a point Q is 130 mm from the origin.

(a) What is the phase difference between P and Q?

(b) Which of the following statements best describes this phase difference:
'almost completely out of phase'; 'roughly ¼ cycle out of phase';
'almost in phase'.

Solution

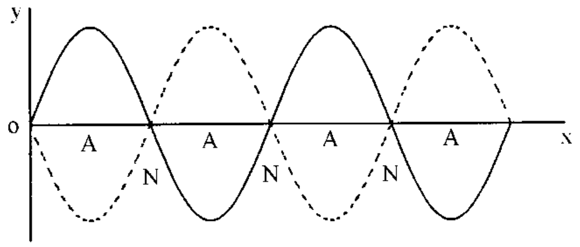
(a) separation of points = 130 - 75 = 55 mm = 0.055 m

phase difference = $2\pi \frac{0.055}{0.060}$ = 5.76 radians

(b) P and Q are separated by 55 mm which is almost one wavelength, hence they are 'almost in phase'. Notice that 5.76 radians is 330°, which is close to 360°.

Stationary Waves
A stationary wave is formed by the interference between two waves, of the **same** frequency and amplitude, travelling in **opposite** directions. For example, this can happen when sound waves are reflected from a wall and interfere with the waves approaching the wall.

A stationary wave travels neither to the right nor the left, the wave 'crests' remain at fixed positions while the particle displacements increase and decrease in unison.



A - antinodes

N - nodes

There are certain positions which always have **zero amplitude** independent of the time we observe them; these are called **nodes**.

There are other points of **maximum amplitude** which are called **antinodes**.

Note that the distance between each node and the next node is $\frac{\lambda}{2}$ and, that the distance between each antinode and the next antinode is $\frac{\lambda}{2}$.

Use of standing waves to measure wavelength
Standing waves can be used to measure the wavelength of waves. The distance across a number of minima is measured and the distance between consecutive nodes determined and the wavelength calculated. This method can be used for sound waves or microwaves.

Formula for standing waves

Consider the two waves y_1 and y_2 travelling in the opposite direction, where

$$y_1 = a \sin 2\pi(ft - \frac{x}{\lambda}) \quad \text{and} \quad y_2 = a \sin 2\pi(ft + \frac{x}{\lambda})$$

When these two waves meet the resultant displacement y is given by

$$y = y_1 + y_2 = a \sin 2\pi(ft - \frac{x}{\lambda}) + a \sin 2\pi(ft + \frac{x}{\lambda})$$

$$y = 2 a \sin 2\pi ft \cos \frac{2\pi x}{\lambda} \qquad \text{(using } a \sin P + a \sin Q = 2a \sin\frac{P+Q}{2} \cos\frac{P-Q}{2}\text{)}$$

$$\text{Giving} \qquad y = 2 a \sin \omega t \cos \frac{2\pi x}{\lambda}$$

Notice that the equation is a function of two trigonometric functions, one dependent on time *t* and the other on position *x*. Consider the part which depends on position. We can see that there are certain fixed values of x for which $\cos \frac{2\pi x}{\lambda}$ is equal to zero.

These are $x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}$, etc.

This shows that there are certain positions where y = 0 which are independent of the time we observe them - the nodes.

The positions at which the amplitude of the oscillation is maximum are given by

$\cos \frac{2\pi x}{\lambda} = 1$, that is $x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}$, etc.

These are points of maximum amplitude - the antinodes.

Interference - Division of Amplitude

Producing interference
Interference of waves occurs when waves overlap. There are two ways to produce an interference pattern for light: division of amplitude and division of wavefront. Both of these involve splitting the light from a single source into two beams. We will consider division of amplitude first and division of wavefront in the next section.

Division of amplitude
This involves splitting a single light beam into two beams, a reflected beam and a transmitted beam, at a surface between two media of different refractive index. In some cases multiple reflections can occur and more than two beams are produced. Before we consider specific examples we need to consider some general properties of interference.

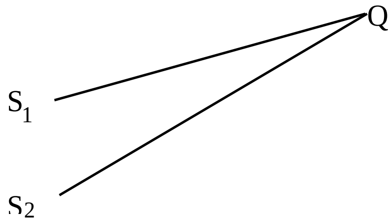
Coherent sources
Two coherent sources must have a **constant phase difference**. Hence they will have the **same frequency**.

To produce an interference pattern for light waves the two, or more, overlapping beams always come from the **same single source**. When we try to produce an interference pattern from two separate light sources it does not work because light cannot be produced as a continuous wave. Light is produced when an electron transition takes place from a higher energy level to a lower energy level in an atom. The energy of the photon emitted is given by $\Delta E = hf$ where ΔE is the difference in the two energy levels, f is the frequency of the photon emitted and h is Planck's constant. Thus a source of light has continual changes of phase, roughly every nanosecond, as these short pulses of light are produced. Two sources of light producing the same frequency will not have a constant phase relationship so will not produce clear interference effects.

This is not the case for sound waves. We can have two separate loudspeakers, connected to the same signal generator, emitting the same frequency which will produce an interference pattern.

Path Difference and Optical Path Difference

Sources S_1 and S_2 are two coherent sources **in air**.



The path difference is (S_2Q - S_1Q). For constructive interference to take place at Q, the waves must be in phase at Q. Hence the path difference must be a whole number of wavelengths.

$$(S_2Q - S_1Q) = m\lambda \qquad \text{where } m = 0, 1, 2, 3, ...$$

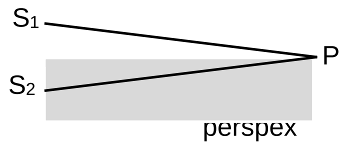
(**Note**: the letter m is used to denote an integer, not n, since we use n for refractive index.)

Similarly, for destructive interference to take place the waves must be out of phase at Q by $\lambda/2$ (that is a 'crest' from S_1 must meet a 'trough' from S_2).

$$(S_2Q - S_1Q) = (m + \frac{1}{2})\lambda$$

Optical path difference

In some situations the path followed by one light beam is inside a transparent material of refractive index, n. Consider two coherent beams S_1 and S_2 where S_1P is in air and S_2P is in perspex of refractive index n = 1.5. We will consider the point P itself to be in air.



The geometrical path difference $S_1P - S_2P$ is zero.

But will there be constructive interference at P?

The wavelength inside the perspex is **less** than that in air $\lambda_{perspex} = \frac{\lambda_{air}}{1.5}$. Hence the waves from S_1 and S_2 may **not** arrive at P in phase. For example, if there were exactly Z whole waves between S_1P, there will be 1.5 x Z waves between S_2P which may or may not be a whole number of wavelengths.

The **optical path length** must be considered not the geometrical path length.

Optical path length	=	refractive index × geometrical path length

Thus the relationships for constructive and destructive interference must be considered for **optical path lengths**, S_2P and S_1P.

For constructive interference	$(S_2P - S_1P) = m\lambda$	where m is an integer
For destructive interference	$(S_2P - S_1P) = (m + \frac{1}{2})\lambda$	where m is an integer

Phase difference and optical path difference

The optical path difference is the difference in the two optical path lengths, namely $(S_2P - S_1P)$ in our general example.

The phase difference is related to the optical path difference:

$$\text{phase difference} = \frac{2\pi}{\lambda} \times \text{optical path difference}$$

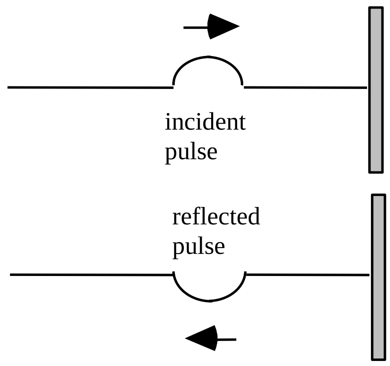
where λ is the wavelength in vacuum.

Notice that when the optical path difference is a whole number of wavelengths, the phase difference is a multiple of 2π, i.e. the waves are in phase.

Phase Change on Reflection

To understand interference caused by multiple reflections it is necessary to consider what happens when a light wave moving in air hits a material such as glass.

On a large scale we can see what happens to the wave when a pulse on a rope or 'slinky' reflects off a dense material such as a wall.



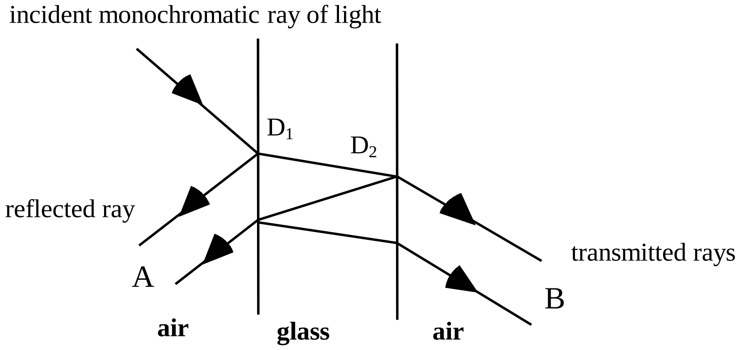
The reflected pulse is said to undergo a phase change of 180° or π radians. The reflected pulse is 180° out of phase with the incident pulse. If these two pulses were to meet they would momentarily cancel as they passed one another.

There is a similar phase change when a light wave is reflected off a sheet of glass.

In general for **light** there is a **phase change of π on reflection** at an interface where there is an **increase** in optical density, e.g. a higher refractive index such as light going from air to glass. There is **no** phase change on reflection where there is a decrease in optical density, e.g. a lower refractive index such as light going from glass to air.

Thin parallel sided film

Interference by division of amplitude can be produced by thin films as shown below.

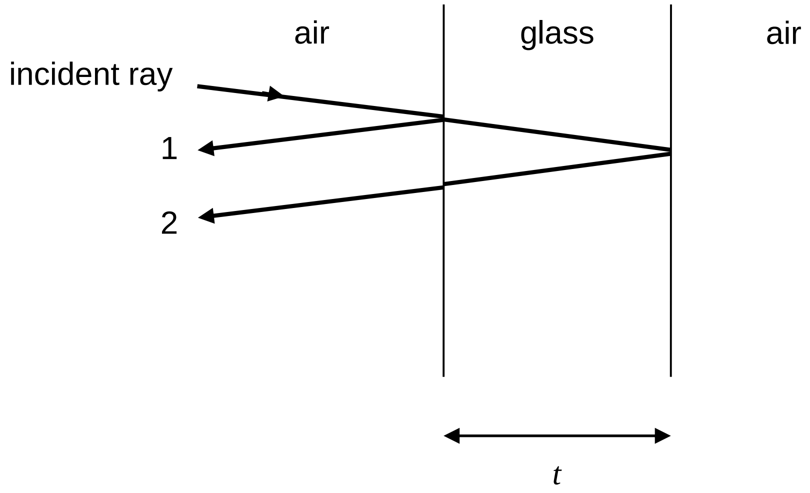


Notice that an **extended** source can be used. The amplitude of the beam is divided by reflection and transmission at D_1, and again by reflection and transmission at D_2 at the back of the glass sheet.

An eye, at A, will focus the reflected beams and an eye at B will focus the transmitted beams. Thus interference patterns can be observed in both the reflected and transmitted beams.

Condition for maxima and minima in the fringes formed in a thin film
The following explanations are for light incident normally on a thin film or sheet of glass. The diagrams only show light paths at an angle to distinguish clearly the different paths.

Reflected light



The ray following path 1 reflects off the glass which has a higher refractive index than air. It therefore experiences a π phase change.

The ray following path 2 reflects off air so experiences no phase change on reflection. However, it travels through the glass twice so has an optical path difference compared to ray 1 of $2nt$, where n is the refractive index of the glass.

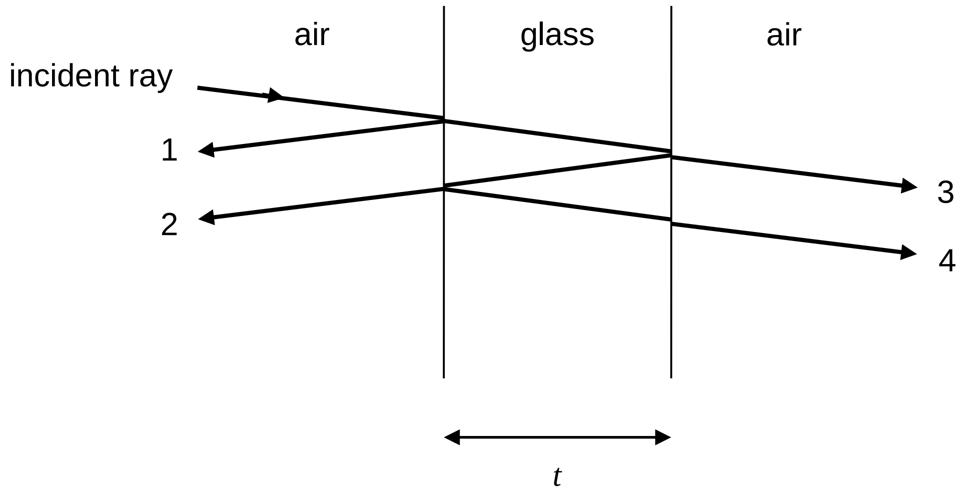
Therefore for constructive interference for the reflected light, i.e. for rays 1 and 2 to be in phase, then the optical path difference $2nt$ must give a π phase change. Therefore:

$2nt = (m + ½)\lambda$ where m is an integer.

For constructive interference for the reflected light, i.e. for rays 1 and 2 to be exactly out of phase, then the optical path difference $2nt$ must give zero phase change. Therefore:

$2nt = m\lambda$ where m is an integer.

Transmitted light



The ray following path 3 passes through the glass with zero phase change.

The ray following path 4 reflects off air twice so experiences no phase change on reflection. However, it travels through the glass twice more than path 3 so has an optical path difference compared to ray 3 of $2nt$, where n is the refractive index of the glass.

Therefore for constructive interference for the transmitted light, i.e. for rays 3 and 4 to be in phase, then the optical path difference $2nt$ must give zero phase change. Therefore:

$2nt = m\lambda$ where m is an integer.

For constructive interference for the transmitted light, i.e. for rays 3 and 4 to be exactly out of phase, then the optical path difference $2nt$ must give a π phase change. Therefore:

$2nt = (m + ½)\lambda$ where m is an integer.

Note
For a certain thickness of thin film the conditions are such that the reflected light and transmitted light have opposite types of interference. Therefore energy is conserved at all times.

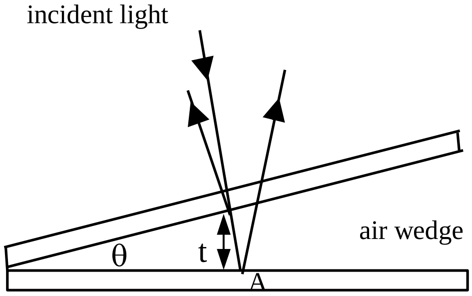
Example
A sheet of mica is 4.80 μm thick. Light of wavelength 512 nm is shone onto the mica. When viewed from above, will there be constructive, destructive, or partial destructive interference? The refractive index of mica is 1.60 for light of this wavelength.

Solution
For destructive interference

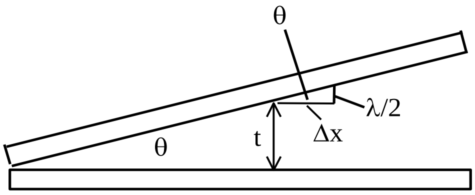
$$2nt = m\lambda$$
$$2 \times 1.60 \times 4.80 \times 10^{-6} = m \times 512 \times 10^{-9}$$
$$m = 30$$

This is an integer. Hence destructive interference is observed.

Wedge Fringes
Two glass slides are arranged as shown below.
Division of amplitude takes place at the *lower* surface of the top glass slide.



Enlarged view showing the geometry



When viewed from above the optical path difference = 2t
There is a phase difference of π on reflection at A. Hence the condition for a dark fringe is $2t = m\lambda$ assuming an air wedge.

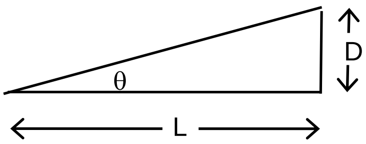
For the next dark fringe t increases by $\frac{\lambda}{2}$ (see right hand sketch above).

Thus the spacing of fringes, Δx, is such that $\tan\theta = \frac{\lambda}{2\,\Delta x}$ giving

$$\boxed{\Delta x = \frac{\lambda}{2\tan\theta}}$$

A wedge of length L and spacing D

$\tan\theta = \frac{D}{L}$.



The fringe spacing is given by $\boxed{\Delta x = \frac{\lambda L}{2D}}$ where λ is the wavelength of light in air.

In practice the distance across a number of fringes is measured and □x determined.

Notice that the fringes are formed inside the wedge, and that the two reflected rays are diverging. The eye, or a microscope, must be focussed between the plates for viewing the fringes.

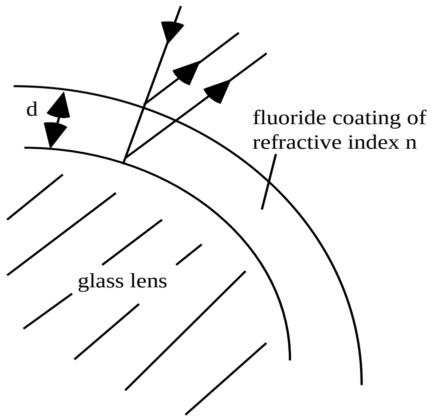
A wedge can be formed by two microscope slides in contact at one end and separated by a human hair or ultra thin foil at the other end. In this way the diameter of a human hair can be measured.

Similarly, if a crystal is placed at the edge and heated, the thermal expansion can be measured by counting the fringes as the pattern changes.

Non-reflecting Coating

Good quality lenses in a camera reflect very little light and appear dark or slightly purple. A thin coating of a fluoride salt such as magnesium fluoride on the surface of the lens allows the majority of the light falling on the lens to pass through.

The refractive index, n, of the coating is chosen such that $1 < n < n_{glass}$.



Notice that there is a phase change of π at both the first and second surfaces.

For cancellation of reflected light:

optical path difference = $\frac{\lambda}{2}$.

Optical path in fluoride = 2nd

thus 2nd = $\frac{\lambda}{2}$ and

$$\boxed{d = \frac{\lambda}{4n}}$$

Complete cancellation is for one particular wavelength only. Partial cancellation occurs for other wavelengths.

The wavelength chosen for complete cancellation is in the yellow/green (i.e. middle) of the spectrum. This is why the lens may look purple because the reflected light has no yellow present. The red and blue light are partially reflected to produce the purple colour observed.

Colours in thin films

When a soap film is held vertically in a ring and is illuminated with monochromatic light it initially appears coloured all over. However when the soap drains downwards a wedge shaped film is produced, with the top thinner than the bottom. Thus horizontal bright and dark fringes appear. When illuminated by white light, colours are formed at positions where the thickness of the film is such that constructive interference takes place for that particular colour. Just before the soap film breaks, the top appears black because the film is so thin there is virtually no path difference in the soap. Destructive interference occurs because of the phase change on reflection.

Similar colours are observed when a thin film of oil is formed on water.

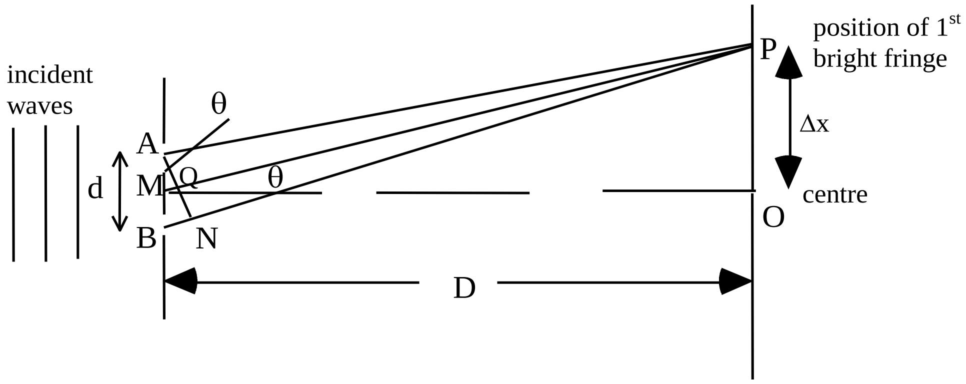
Interference - Division of Wavefront

Division of Wavefront
When light from a single point source is incident on two small slits, two coherent beams of light can be produced. Each slit acts as a secondary source due to diffraction.

If an extended source is used, each part of the wavefront will be incident on the slit at a different angle. Each part of the source will then produce a fringe pattern, but slightly displaced. When the intensity of all the patterns is summed the overall interference pattern may be lost. However a line source parallel to the slits is an exception.

Compare this with the use of an **extended** source in 'division of amplitude'.

Young's Slits Experiment
The diagram below shows light from a single source of monochromatic light incident on a double slit. The light diffracts at each slit and the overlapping diffraction patterns produce interference.



A bright fringe is observed at P. Angle PMO is θ.
N is a point on BP such that NP = AP. Since P is the first bright fringe BN = λ
For **small** values of θ AN cuts MP at almost 90^0 giving angle MAQ = θ and hence angle BAN = θ.
Again providing θ is **very small**, sin θ = tan θ = θ in radians
From triangle BAN: $\theta = \frac{\lambda}{d}$ also from triangle PMO: $\theta = \frac{\Delta x}{D}$

Thus $\frac{\Delta x}{D} = \frac{\lambda}{d}$ or $\Delta x = \frac{\lambda D}{d}$

Giving the fringe separation between adjacent fringes Δx

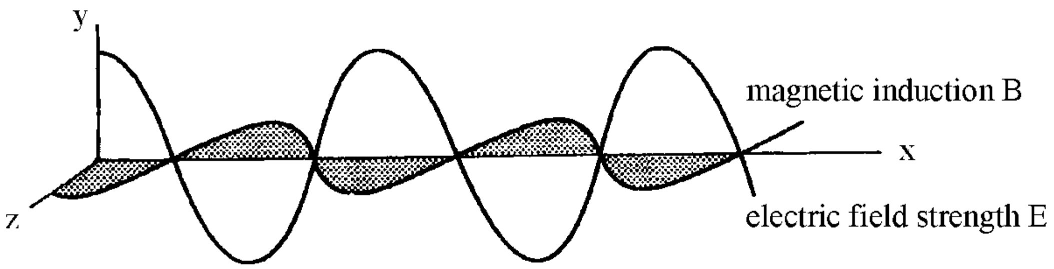
$$\boxed{\Delta x = \frac{\lambda D}{d}}$$

Note
This formula only applies if x<<D, which gives θ small. This is likely to be true for light waves but not for microwaves.

The position of the fringes is dependent on the wavelength. Thus if white light is used we can expect overlapping colours either side of a central white maximum. The red, with the longer wavelength, will be the furthest from this white maximum ($\Delta x_{red} > \Delta x_{violet}$ since $\lambda_{red} > \lambda_{violet}$).

Polarisation

Polarised and unpolarised waves
Light is a wave motion, and is part of the electromagnetic spectrum. In all **electromagnetic waves** the electric field and magnetic field vary. The diagram below shows a 3-dimensional picture of such a wave.

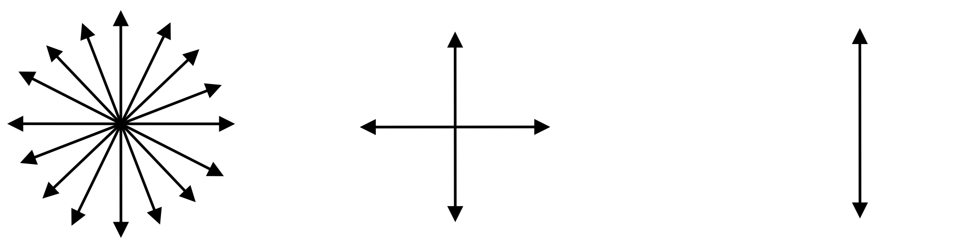


The above diagram shows the variation of the electric field strength, E, in the x-y plane and the variation of the magnetic induction, B, in the x-z plane. In this example the electric field strength is only in one plane. The wave is said to be plane polarised, or **linearly polarised**. For example, in Britain this is the way that T.V. waves are transmitted. Aerials are designed and oriented to pick up the vertical electric field strength vibrations. These vibrations contain the information decoded by the electronic systems in the television.

Light from an ordinary filament lamp is made up of many separate electromagnetic waves produced by the random electron transitions in the atoms of the source. So unlike the directional T.V waves, light waves from a lamp consist of many random vibrations. This is called an **unpolarised** wave.

When looking at an **unpolarised** wave coming towards you the direction of the electric field strength vector would appear to be vibrating in all direction, as shown in the diagram (i) on the left below. The magnetic induction vector would be perpendicular to the electric field strength vector, hence this too would be vibrating in all directions However when discussing polarisation we refer to the electric field strength vector only.

All the individual electric field strength vectors could be resolved in two mutually perpendicular direction to give the other representation of a **unpolarised** wave, as shown below in the centre diagram (ii).



(i) unpolarised wave (ii) unpolarised wave (iii) polarised wave

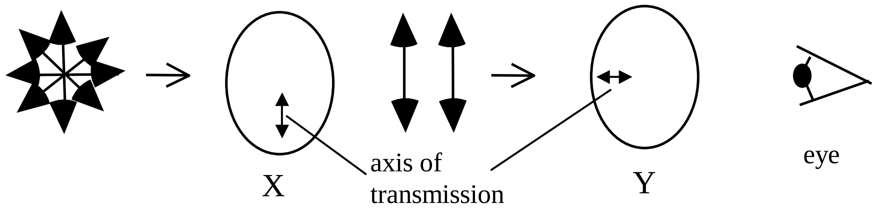
The right hand diagram (iii) above represents a **polarised** wave.

Longitudinal and transverse waves
Note that only transverse waves can be polarised. Longitudinal waves, e.g. sound waves, cannot be polarised.

Polarisation using Filters
We can produce a **linearly polarised** wave if we can somehow absorb the vibrations in all the other directions except one.

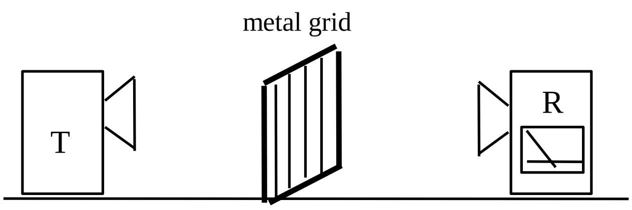
In 1852 Herapath discovered that a crystal of iodo-quinine sulphate transmitted one plane of polarisation, other planes being absorbed. In 1938 Land produced the material 'Polaroid', which has a series of parallel long hydrocarbon chains. Iodine atoms impregnate the long chains providing conduction electrons. Light is only *transmitted* when the electric field strength vector is *perpendicular* to the chain.

The arrangement below shows a polaroid filter at X producing linearly polarised light. The polaroid at X is called a **polariser.** Vibrations of the electric field strength vector at right angles to the axis of transmission are absorbed.



A second polaroid at Y is placed perpendicular to the first one, as shown above. This is called an **analyser.** The analyser absorbs the remaining vibrations because its axis of transmission is at right angles to the polariser at X and no light is seen by the eye. The light between X and Y is called **linearly** or **plane polarisation**.

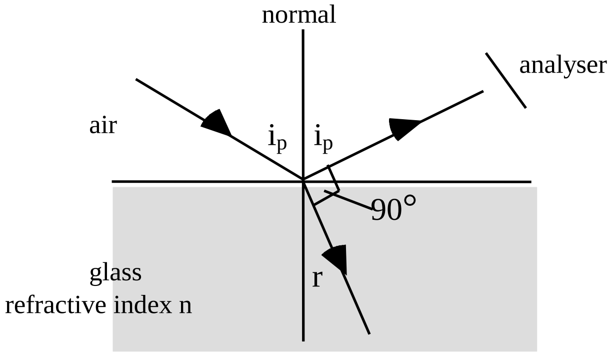
These effects also can be demonstrated using microwaves and a metal grid.



The microwaves emitted by the horn are plane polarised. In this example the electric field strength vector is in the *vertical plane*. The waves are **absorbed** by the rods and re-radiated in all directions. Hence there will be a very low reading on the receiver, R. When the metal grid is rotated through 90^o the waves will be transmitted, and the reading on the receiver will rise. Notice that the microwaves are *transmitted* when the plane of oscillation of the electric field strength vector is *perpendicular* to the direction of the rods.

Polarisation by Reflection
Plane polarised waves can be produced naturally by light reflecting from any electrical insulator, like glass. When refraction takes place at a boundary between two transparent materials the components of the electric field strength vector parallel to the boundary are largely reflected. Thus reflected light is partially plane polarised.

Plane polarisation at the Brewster angle



Consider a beam of unpolarised light incident on a sheet of smooth glass. This beam is partially reflected and partially refracted. The angle of incidence is varied and the reflected ray viewed through an analyser, as shown above. It is observed that at a certain angle of incidence i_p the reflected ray is plane polarised. No light emerges from the analyser at this angle.

The **polarising angle** i_p or **Brewster's angle** is the angle of incidence which causes the reflected light to be linearly polarised.

This effect was first noted by an experimenter called Malus in the early part of the nineteenth century. Later Brewster discovered that **at** the polarising angle i_p the refracted and reflected rays **are separated by 90°.**
Consider the diagram above, which has this 90° angle marked:

$$n = \frac{\sin i_p}{\sin r}$$

but $r = (90 - i_p)$ thus $\sin r = \sin(90 - i_p) = \cos i_p$

$$\text{thus} \quad n = \frac{\sin i_p}{\cos i_p} = \tan i_p$$

$$\boxed{n = \tan i_p}$$

Example
Calculate the polarising angle for glycerol, n = 1.47.
What is the angle of refraction inside the glycerol at the Brewster angle?

Solution
Using the equation $n = \tan i_p$ $1.47 = \tan i_p$ giving $i_p = 56^o$.
At the Brewster angle, which is the polarising angle,
angle of refraction + $i_p = 90^o$ thus angle of refraction = 44°.

Reduction of Glare by Polaroid sunglasses
When sunlight is reflected from a horizontal surface, e.g. a smooth lake of water, into the eye, eyestrain can occur due to the glare associated with the reflected light. The intensity of this reflected beam can be reduced by wearing polaroid sunglasses. These act as an analyser and will cut out a large part of the reflected polarised light.