# PHYSICS COURSE – YEAR 12

**MODULE 7: THE NATURE OF LIGHT**

Prior to the 20th century, physicists, including Newton and Maxwell, developed theories and models about mechanics, electricity and magnetism and the nature of matter. These theories and models had great explanatory power and produced useful predictions. However, the 20th century saw major developments in physics as existing theories and models were challenged by new observations that could not be explained. These observations led to the development of quantum theory and the theory of relativity. Technologies arising from these theories have shaped the modern world. For example, the independence of the speed of light on the frame of observation or the motion of the source and observer had significant consequences for the measurement, and concepts about the nature, of time and space.

Throughout this module, students explore the evidence supporting these physical theories, along with the power of scientific theories to make useful predictions.

As always, teachers are required to provide students with opportunities to engage with all the Working Scientifically skills throughout the course.

I will include some extension sections in these notes. These extensions are not necessary for the completion of this Module.

**ELECTROMAGNETIC SPECTRUM**

**Inquiry Question:** What is light?

**James Clerk Maxwell** – main reference – Ref.24

James Clerk Maxwell was born in Scotland in 1831. This was the same year that Michael Faraday settled on the idea of explaining electric and magnetic effects in terms of **lines of force**, something that eventually led to our modern-day concept of a **field**.

Maxwell was a brilliant mathematician and physicist. In his career, Maxwell made huge contributions to science in a number of fields – optics, the theory of colour vision, kinetic theory, astrophysics and his greatest contribution – the theory of electromagnetism. He is considered in the same category of great scientists as Sir Isaac Newton.

In the 1850’s, based on his understanding of the work of Faraday and others, Maxwell developed a theory of how electric and magnetic forces were conveyed using the concept of vortices. A **vortex** was a whirlpool that spun in a fluid that filled all of space. It was a concept that predated the idea of fields. Maxwell believed that the physical image that we use to imagine or contemplate a physical phenomenon was not as important as the mathematical equations that describe what was going on. This is a belief that applies in all science today. Conceptual models are extremely important and very helpful; but they are not the “truth”. In so far as there is scientific truth, it resides in the equations.

The French physicist Armand Fizeau made the first accurate ground-based measurement of the speed of light in the late 1840s. In 1850, Fizeau also shown that light travels more slowly through water than through air, a key prediction of the wave model of light. The corpuscular (particle) theory of light predicted that light would travel faster in water than in air. Fizeau’s work showed that this was not the case. By 1862 Leon Foucault had improved the accuracy of the speed of light measurement to be within 1% of the accepted modern value. This accurate measurement of the speed of light was invaluable in the context of the implications of Maxwell’s theory of electromagnetism.

In 1861 and 1862 Maxwell published a set of four papers “On Physical Lines of Force”, still using vortices, to describe **how electromagnetic waves could propagate**. He found that the speed of these waves depended on the properties of the medium through which they were travelling. Using the then known properties of space for electricity and magnetism, Maxwell determined that **electromagnetic waves would travel at the speed of light**.

In one of his 1862 papers Maxwell comments: “We can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena.”

As he refined the mathematics of his theory, Maxwell found that he could dispense with the idea of vortices and a medium to enable the waves to travel. In 1864, Maxwell published his most famous paper, “A Dynamical Theory of the Electromagnetic Field”. This summed up everything that it is possible to say about classical electricity and magnetism, in a set of four equations that have become known simply as Maxwell’s equations. Every problem involving electricity and magnetism can be solved using these equations, except for certain quantum phenomena. Maxwell’s equations and theory **unified the two concepts of electricity and magnetism into one concept of the electromagnetic field**.

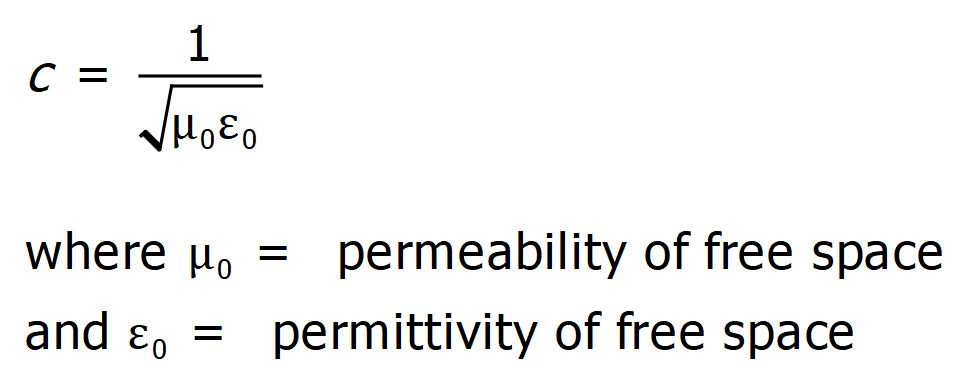
Maxwell’s equations are expressed in a mathematical formalism that is beyond the scope of Stage 6 Physics. So, instead of presenting the equations themselves (which you can [google](https://simple.wikipedia.org/wiki/Maxwell%27s_equations) if you wish) I will simply state what each of the four equations says in plain English (7). You already know these results from your studies of electromagnetism.

* Equation 1 – describes the relationship between a static electric field and the electric charges that cause it: a static electric field points away from positive charges and towards negative charges, and the net outflow of the electric field through any closed surface is proportional to the charge enclosed by the surface. The bigger the charge, the higher the number of electric field lines emanating from it, the stronger the field. This is called Gauss's law for electric fields.
* Equation 2 – states that there are no "magnetic charges" (also called magnetic monopoles), analogous to electric charges.  Instead, the magnetic field due to materials is generated by a dipole, a north pole and south pole in combination and the net outflow of the magnetic field through any closed surface is zero. In other words, a magnet always has a north & south pole and the field lines run from the north, outside the magnet to the south, and then back to the north through the material of the magnet. This is called Gauss’s law for magnetic fields.
* Equation 3 – states that an electric field can be created (induced) by a time varying magnetic field. This is called Faraday’s law of induction.
* Equation 4 – this is called Ampere’s law with Maxwell’s addition. It states that a magnetic field can be created either by a current (Ampere’s law) or by a time varying electric field (Maxwell’s addition).

The final two equations explain how self-sustaining EM waves can travel through free space. I will explain this shortly.

The names given to Maxwell’s equations respect the physicists who originally discovered the individual experimental results. The beauty and power of Maxwell’s theory is that it unites separate ideas we had about electricity and magnetism into one cohesive theory and provides a consistent mathematical formalism to the phenomenon of electromagnetism. From the four equations, every other classical equation in electromagnetism can be derived.

As brilliant and useful as Maxwell’s theory is in itself, it also suggested a further important understanding. Maxwell’s equations contain a constant **c**, which represents the speed at which the EM waves propagate. This constant is related as follows to measurable electric and magnetic properties of matter:



Using the known values of these constants, the calculated value of **c** was, within experimental error, the same as by then, the well-determined speed of light. Maxwell wrote:

“This velocity is so nearly that of light that it seems we have strong reason to conclude that light itself (including radiant heat and other radiations, if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.” (24)

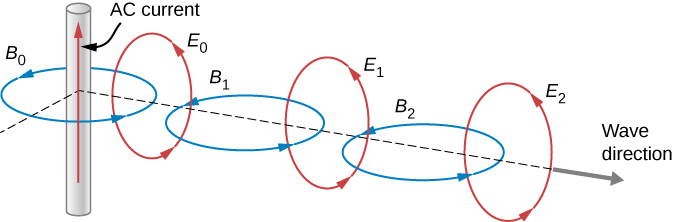
The reference to **other radiations** shows that Maxwell expected that there could be forms of EM waves with wavelengths beyond those of visible light. This was eventually shown to be true.

**The Production and Propagation of EM Waves**

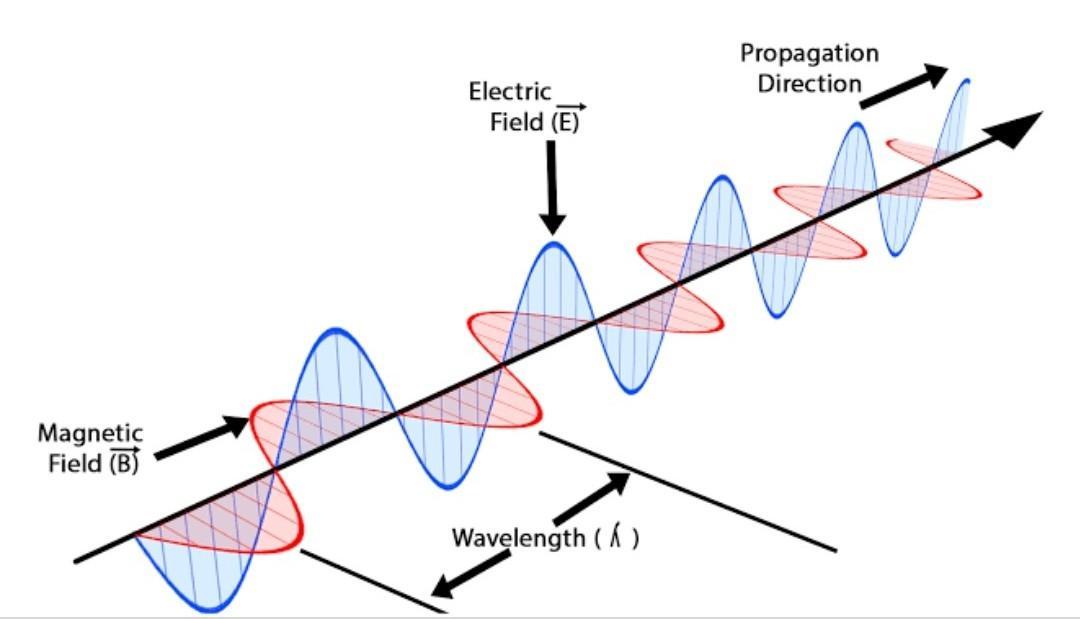
As mentioned above, Maxwell predicted the existence of electromagnetic waves. Equations 3 and 4 used together explain this. Faraday’s law indicates that a changing magnetic field will generate an electric field and Ampere’s law with Maxwell’s addition indicates that a changing electric field or a current in a wire will generate a magnetic field.

So, if we produce an AC current in a long wire or an antenna system, it will generate a changing magnetic field around the wire or antenna. At any point in and around the wire or antenna, this changing magnetic field must by its very nature induce a changing electric field at right angles to the magnetic field. This changing electric field must generate another changing magnetic field at right angles to the electric field and so on. In this way a series of electric and magnetic oscillations at right angles to each other is set up and propagates as an EM wave in a direction at right angles to both the electric and magnetic field directions. The speed of propagation is 3 x 108 m/s (approx). (24)

The following is a depiction showing how a series of changing electric and magnetic fields propagate through space as an EM wave. Diagram taken without change from [Openstax – Maxwell’s Equations and Electromagnetic Waves page](https://cnx.org/contents/dP0ocxV9@5.52:-LQJwSUO@3/33-1-Maxwell-s-Equations-and-Electromagnetic-Waves).



Another depiction of an EM Wave – diagram taken without alteration from: <https://commons.wikimedia.org/wiki/File:Electromagnetic_waves.png>



**Experimental Verification of Maxwell’s Theory of EM**

Heinrich Hertz, a German physicist, achieved the first experimental demonstration of electromagnetic waves in 1887.Hertz used an induction coil to produce oscillating electric sparks between two brass balls connected to two brass plates. The brass plates acted as an aerial system. He used a small loop of wire with a tiny gap in it as the receiver. See diagram below. (25)



As sparks jumped across the gap between the balls, sparks were also observed jumping the gap in the receiver. Hertz reasoned that the spark discharge oscillating backwards and forwards between the brass balls set up changing electric and magnetic fields that propagated as an electromagnetic wave, as postulated by Maxwell. When these waves arrived at the receiver, the changing electric field component caused charges in the loop to oscillate, thus producing the spark across the gap in the receiver. (25)

Hertz carried out a thorough investigation of these waves and showed that they did indeed possess properties like light – reflection, refraction, interference, diffraction and polarisation. By setting up an experiment in which he allowed the waves to reflect from a metal sheet and interfere with themselves to produce standing waves, Hertz was able to determine their wavelength. He calculated the frequency of oscillation of the sparks in his transmitter from knowledge of the parameters of the circuit. Then using v = f  he calculated the speed of the waves as 3 x 108 m/s, as predicted by Maxwell. Thus, Hertz’s experiment confirmed Maxwell’s prediction of EM waves and provided strong experimental support for the idea that light was a form of transverse EM wave. (25)

The waves produced by Hertz eventually became known as radio waves and his research led to the development of radio communications. As Hertz suspected it was indeed oscillating charges that produced the EM waves. Radio waves were some of those “other radiations” that Maxwell had predicted.

**The EM Spectrum**

The wide range of wavelengths (and corresponding frequencies) over which EM waves exist in nature is called the **electromagnetic (EM) spectrum**. This spectrum is as follows (25):

Visible

light

**(m)**

108 106  104  102  100  10-2 10-4 10-6 10-8 10-10 10-12 10-14



100 102  104  106 108 1010  1012 1014 1016  1018 1020  1022

**f (x 3 Hz)**

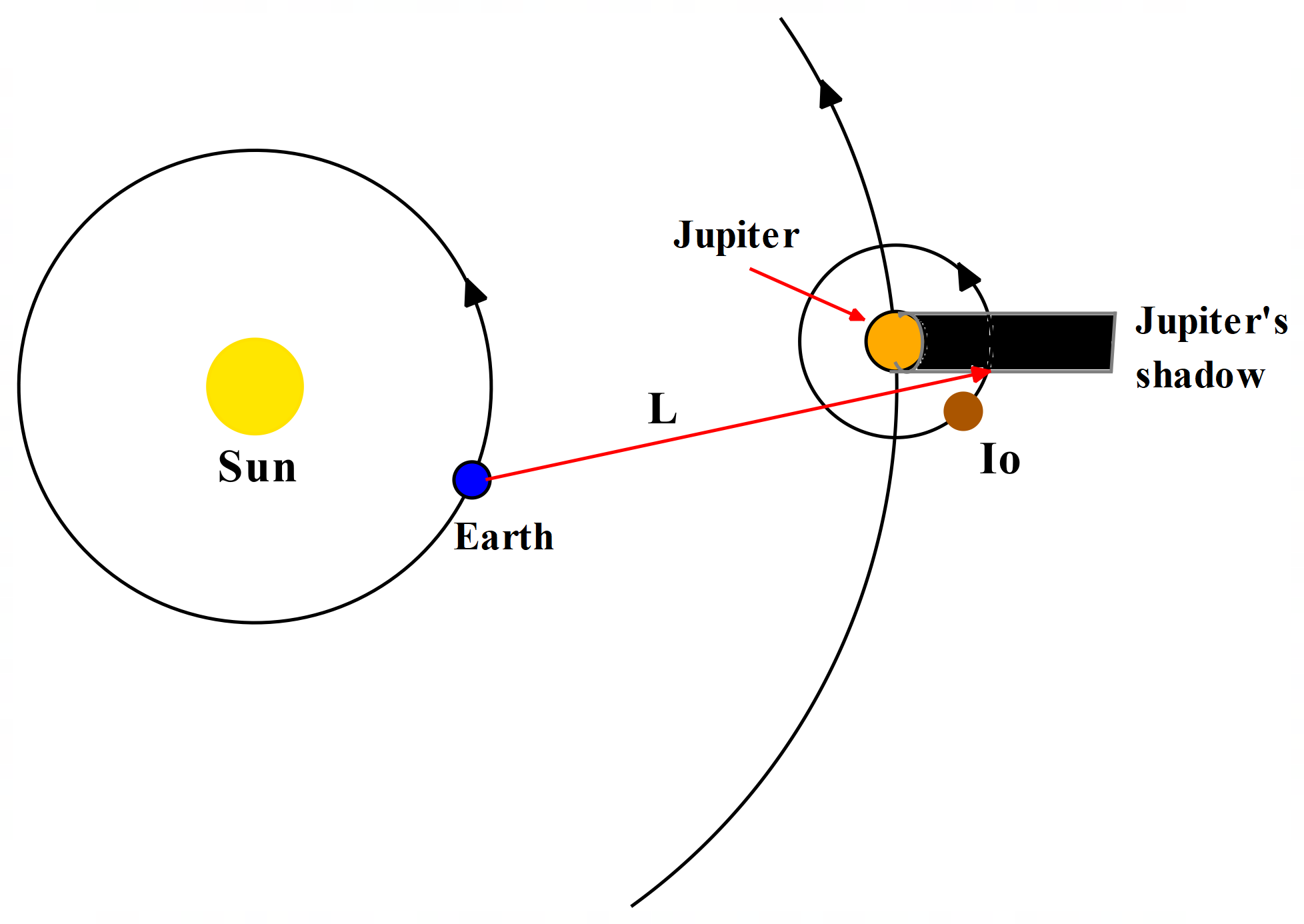
The cut-off wavelengths or frequencies for each of the different types of EM radiation are not precise. There is some overlap. Some types of EM radiation can be further broken down into sub-types. The radio wave band of the spectrum contains the AM radio communications band at its higher wavelength end, followed by the TV band and then the radar and microwave bands at the lower wavelength end. The very narrow visible light band contains all the visible colours: red, orange, yellow, green, blue, indigo and violet, in order from higher to lower wavelength. The visible light band occupies the position between about 780 nm and 380 nm wavelength.

**Experimental Verification of the Speed of Light**

One of the first serious attempts at measuring the speed of light was made by Galileo in 1638 (4). Galileo and an assistant each climbed to the tops of hills separated by about a kilometre. Each carried a lantern covered by a cloth. Galileo had a time piece of some kind with him. Galileo uncovered his lantern and when his assistant saw the light from Galileo’s lantern, he uncovered his own lantern. Galileo attempted to time the round trip of the light. All he managed to do was to determine that the speed of light was extremely high, much faster than the speed of sound.

**Ole Roemer (1644-1710) – Ref. (4)**

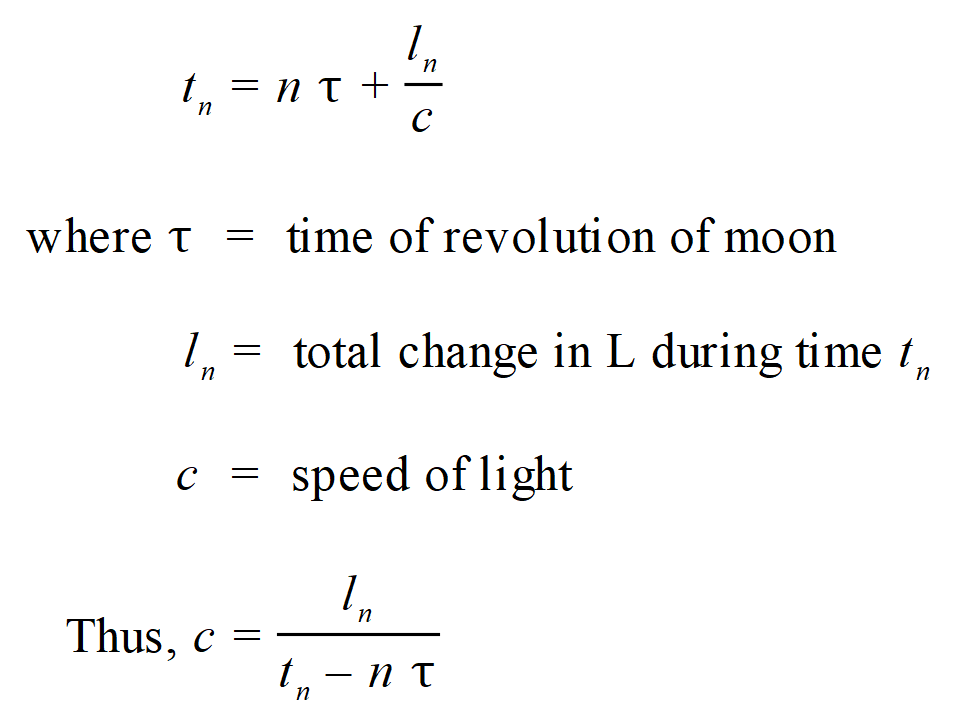
The Danish astronomer, Ole Roemer (also spelled Römer or Rømer & Ole can be spelled Olaf) was the first to calculate the speed of light from astronomical observations, from the eclipses of the satellites of Jupiter. See Diagram below (not to scale).



In the diagram, **L** is the distance from Earth to the satellite’s orbit, which is effectively the distance from Earth to Jupiter.

An eclipse occurs each time a moon, such as Io, enters the shadow of Jupiter. The time for one complete revolution of Io around Jupiter should be constant. Roemer noticed, however, that the average period of revolution of Io appeared to change depending on the time of the year. He reasoned that this time difference could be explained if the speed of light was finite. Then, as the distance between Earth and Jupiter changed as Earth moved around the Sun in its orbit, the light coming from Io would take different amounts of time to make the journey to Earth. **Roemer is credited as being the first person to establish that the speed of light is finite.**

Roemer reasoned that the time for **n** revolutions of Io observed from Earth would be (4):



Roemer counted the number of eclipses, **n**, during 1 year. Then, **tn**, **n** and **ln** are known with **ln** = 0, since in 1 year the change in **L** is zero, as Earth has returned to the same point in orbit as when the count began. This allowed **τ** to be determined.

Next Roemer counted the number of eclipses in half a year beginning with the point of closest approach of Earth to Jupiter. Then, **tn**, **n** and **τ** are known and **ln** = diameter of Earth’s orbit around the Sun. Then **c** can be calculated.

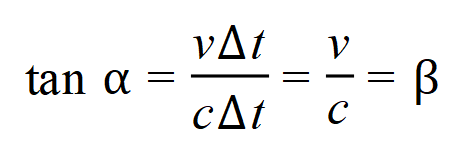
Roemer’s value for **c**, using the most accurately known at the time value for the size of Earth’s orbit around the Sun, was just over 2 x 108 m/s, a remarkable achievement.

**James Bradley (1693-1762) – Refs. (4) & (5)**

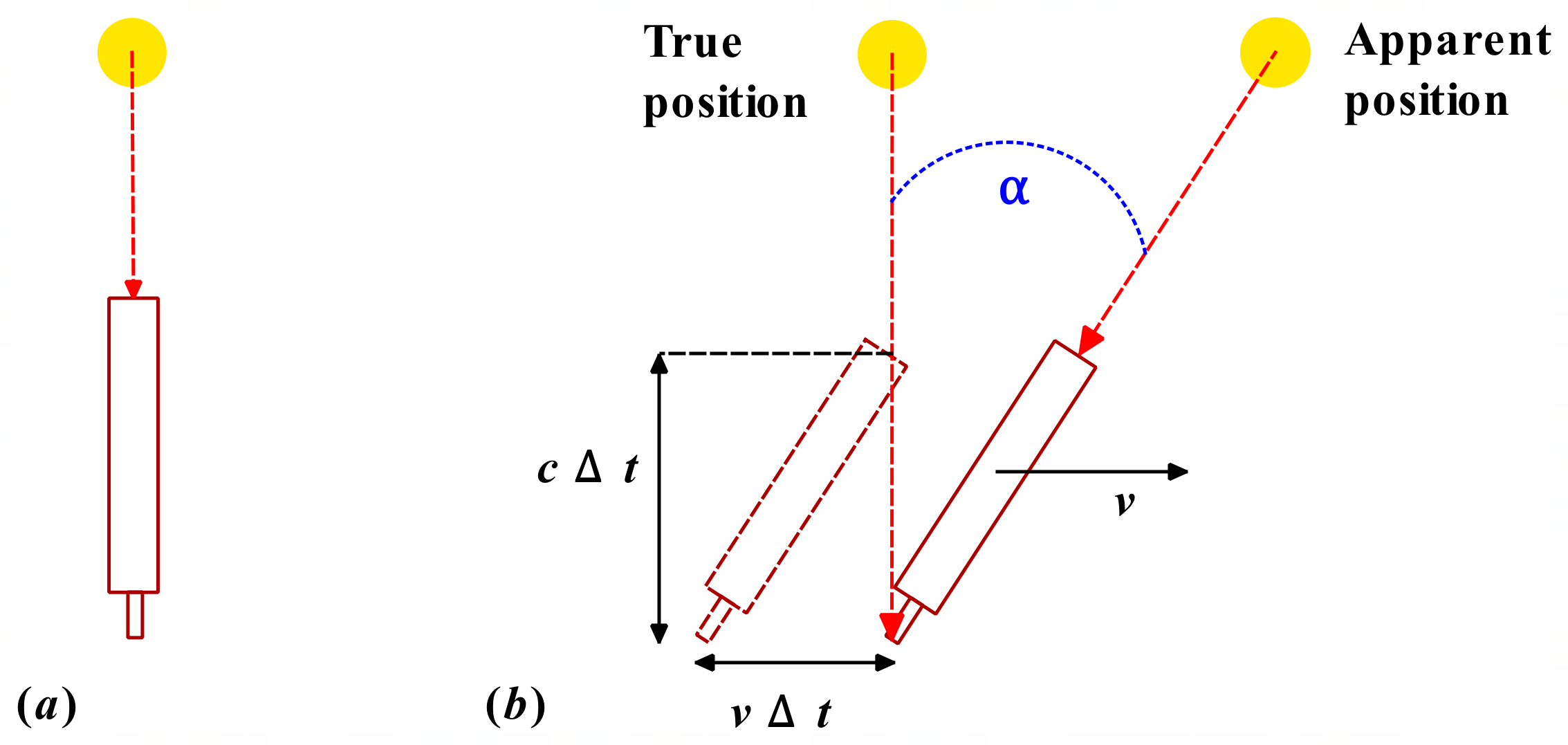
In 1727-29, James Bradley, English priest and astronomer, was using the principle of parallax to measure the distance to stars. During his work, he accidentally discovered an effect he called the aberration of light and which has also been called the **Bradley Aberration** or **Stellar Aberration**. This describes the apparent movement of stars in circles of about 41 seconds of arc.

After much observation and thought, Bradley explained this effect in a similar way to the following. If the Earth was stationary and an astronomer wished to observe a star directly overhead, the astronomer would point the telescope straight up and the light from the star would proceed down the telescope tube and be observed. See diagram (a) below. If, however, the Earth was moving to the right at speed **v** relative to the star, the astronomer would have to tilt the telescope as shown in diagram (b) below in order to see the star.

The light from the star proceeds straight down but during the time **Δt** that the light travels the vertical distance **l = c Δt** from the objective lens to the eyepiece, the telescope has moved a distance **v Δt** to the right. The eyepiece, at the time the ray leaves the telescope, is on the same vertical line as the objective lens was at the time the ray entered the telescope. From the point of view of the telescope, the ray travels along the axis from the objective lens to the eyepiece. The angle of tilt of the telescope, ****, is given by:



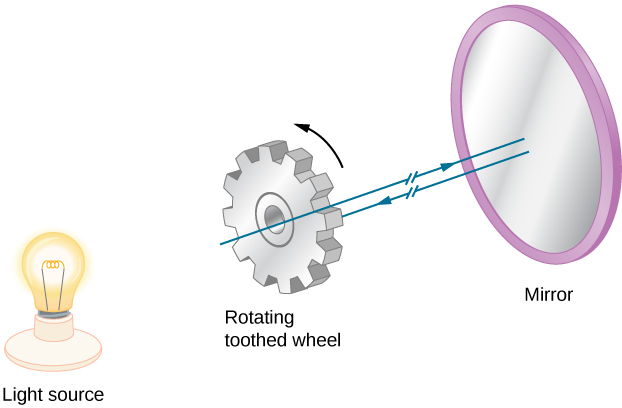
**β** is the aberration constant and from Bradley’s observations has a value of about 1 x 10-4. It was known from astronomical data that the Earth goes around the Sun at a speed of about 30 km/s. Bradley’s value for the speed of light was therefore 3 x 108 m/s. Bradley’s work verified the earlier work of Roemer and improved the accuracy of the measurement of the speed of light.



**Armand Fizeau (1819–1896) – Refs. (7) & (26)**

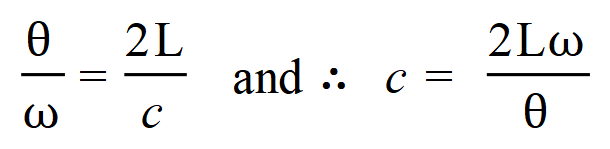
The first successful terrestrial measurement of the speed of light was made by Armand Fizeau in 1849. He placed a **toothed wheel** that could be rotated very rapidly on one hilltop and a mirror on a second hilltop 8 km away. An intense light source was placed behind the wheel, so that when the wheel rotated, it chopped the light beam into a succession of pulses. The speed of the wheel was increased from zero until no light returned to the observer located behind the wheel. This could only happen if the wheel rotated through an angle, ****, from the centre of a gap to the centre of a tooth, while the pulse travelled down to the mirror and back. Knowing the angular speed of the wheel, ****, and the distance to the mirror, **L**, Fizeau determined the speed of light to be 3.13 × 108 m/s, which is only 4% higher than today’s accepted value.

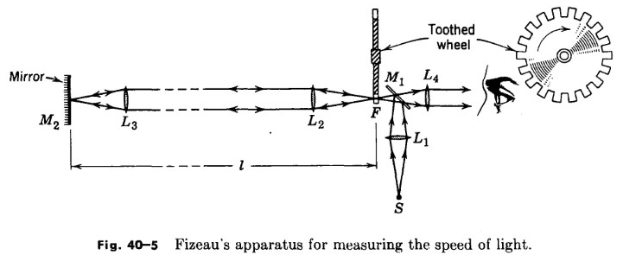
The diagram below is a sketch of the basic idea. A more detailed diagram of the apparatus used by Fizeau follows and is also shown in the Speed of Light Determination – Summary document on The Nature of Light webpage.



**Diagram above:** Fizeau’s method for measuring the speed of light. The teeth of the wheel block the reflected light upon return when the wheel is rotated at a rate that matches the light travel time to and from the mirror. Diagram taken without alteration from [Openstax – The Propagation of Light page](https://cnx.org/contents/5I39byUz@3.3:qL5s9hwt@7/1-1-The-Propagation-of-Light#CNX_UPhysics_34_01_FizeauMeth)

The time needed for the wheel to rotate the required angular distance ****, is the round-trip travel time **2L/c**:

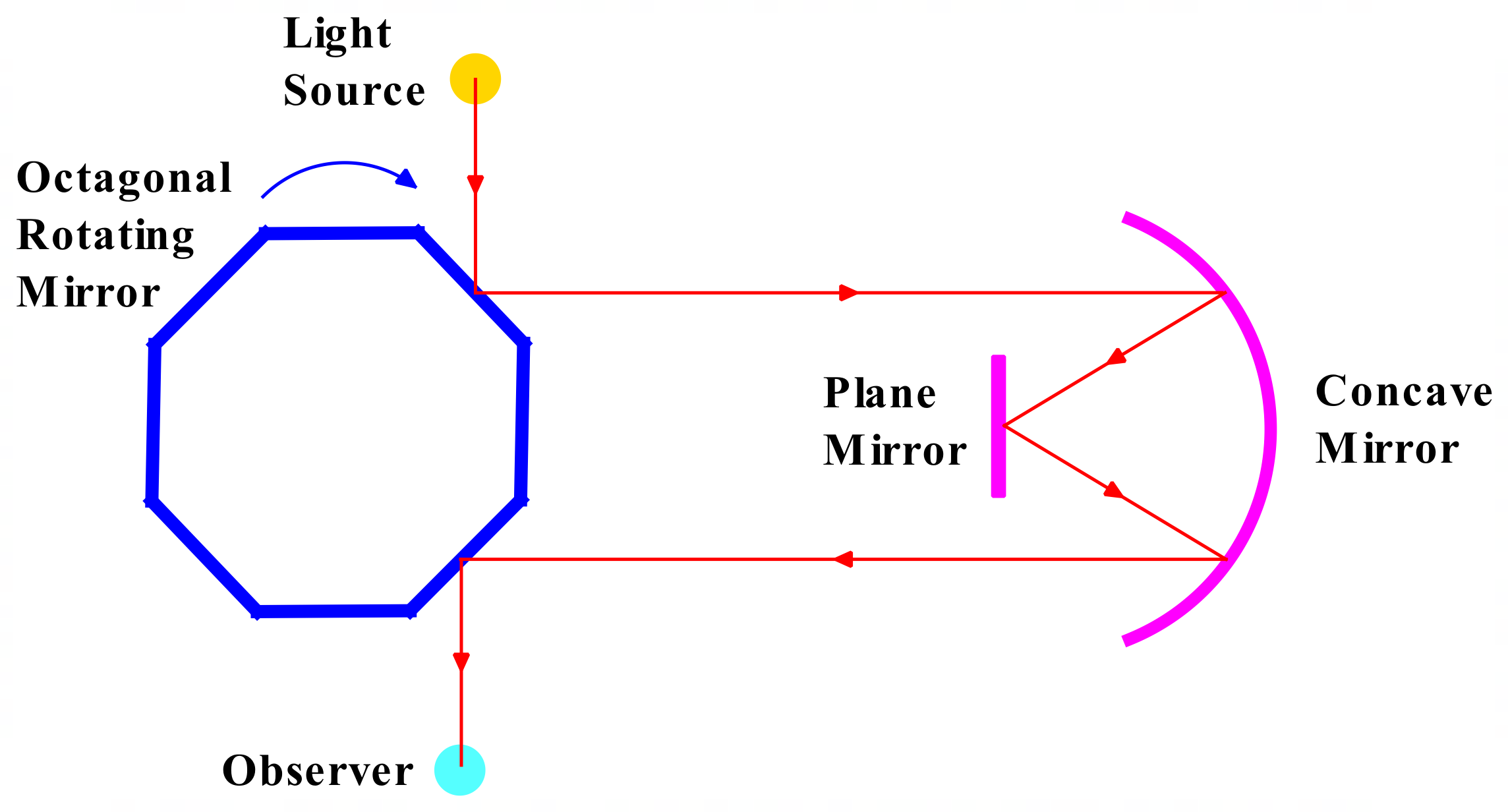




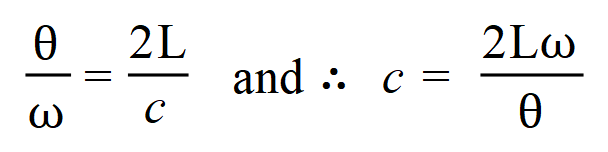
**Foucault (1819–1868) & Michelson (1852-1931) – Refs. (7), (26) & (27)**

The French physicist Jean Bernard Léon Foucault modified Fizeau’s apparatus by replacing the toothed wheel with a rotating mirror. In 1862, he measured the speed of light to be 2.98 × 108 m/s, which is within 0.6% of the presently accepted value.

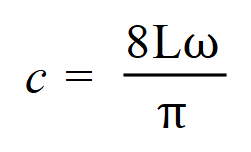
Albert Michelson also used Foucault’s method, with variations, on many occasions to measure the speed of light. His first experiments were performed in 1878; by 1926, he had refined the technique so well that he found **c** to be (2.99796 ± 0.00004) × 108 m/s. The rotating mirror technique works in the following way.



The image of the source can be seen by the observer through a telescope. When the mirror starts rotating the image disappears. The speed of rotation is gradually increased until the image of the source reappears at a particular angular speed, ****. At this speed the rotating mirror is moving through exactly the right angular distance, ****, so that the next face of the rotating mirror takes the place of the previous face exactly, in the time it takes light to make the round-trip distance **2L**, where **L** is the distance between the rotating and concave mirrors. So, as for the Fizeau case, we have:



For an octagonal mirror, **** radians, so we have



Michelson is still held in awe by scientists for his great skill and meticulous planning in obtaining extremely precise and accurate values for **c**. One of those measurements (1923-24) for instance involved a distance between the rotating and concave mirrors of 35 km measured to an accuracy of less than 2.5 cm.

**Importance of Speed of Light for Standards of Measurement – Ref. (28)**

In order to make meaningful measurements in science we need standards of commonly measured quantities, such as those of length, time and mass.

In 1983 the definition of the metre was updated. The metre is defined as "the length of the path travelled by **light** in vacuum during a time interval of ​1⁄299792458  of a second", fixing the value of the **speed of light** at 2.99792458 x 108 m/s by definition. Fixing the definition of the metre in terms of a constant like the speed of light ensures that the fundamental unit for length is derivable from an unchanging universal natural phenomenon. It remains the same over time, it is always accessible and it is easily reproducible – three essential features for any standard of measurement.

The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom. This definition has no relationship to the speed of light – not sure why that was implied in the Syllabus. The speed of light, however, does assist in the latest definition of the standard of mass.

So, for completeness (and not required by the syllabus), the kilogram, was redefined in 2019. The kilogram is defined by taking the fixed numerical value of the Planck constant **h** to be 6.62607015×10−34 when expressed in the unit J⋅s, which is equal to kg⋅m2⋅s−1, where the metre and the second are defined as above.

**The Nature of Spectra**

Isaac Newton in his book *Opticks* in 1704 noted that when a beam of sunlight is shone through a triangular glass prism, the white light is **dispersed**, producing a rainbow of colours, which can be displayed on a screen. The rainbow of colours is called a **spectrum**. In 1814 the German optician Joseph von Fraunhofer discovered that the spectrum of sunlight contains hundreds of fine dark lines, now called **spectral lines**. Fifty years later, chemists found that they could produce spectral lines in the laboratory and use these lines to analyse the kinds of atoms of which substances were made.

There are three types of spectrum: a continuous spectrum, an emission spectrum and an absorption spectrum. Let us now examine each in turn.

**Continuous Spectrum**

**Incandescence** is the effect of dense objects emitting electromagnetic radiation due to their temperature. A hot, glowing solid or liquid, or a hot, glowing, dense gas produces a spectrum consisting of a continuous series of coloured bands ranging from violet on one end to red on the other(16). This is the spectrum produced by the **tungsten filament in an incandescent light globe**, for instance. Examples of other objects that produce continuous spectra include the inner layers of stars and galaxies.

**Emission Spectrum – Refs. (19)**

This is a series of bright, coloured lines on a black background, produced by a hot, glowing, diffuse gas. For example, the visible spectrum of hydrogen when heated to incandescence by passing an electric discharge through the gas is a series of four spectral lines (violet **Hd**, blue **Hg**, green **Hb** & red **Ha**) seen on a black background. See diagram below.

Application

Description automatically generated with low confidence

Spectrogram of visible lines in the Balmer series of hydrogen as obtained with a

constant-deviation spectrograph – Diagram from [Wikimedia Commons](https://www.google.com/search?q=visible+lines+in+the+hydrogen+spectrum+wikimedia+commons&safe=strict&rlz=1C1GCEA_enAU773AU773&tbm=isch&source=iu&ictx=1&fir=RHGN2ZQyoz__pM%252C-lYwtG8m45BHVM%252C_&vet=1&usg=AI4_-kSCMaHRG8gNyCHt6Aska89DaN7nww&sa=X&ved=2ahUKEwiRypC0xbHwAhXEzzgGHSHiDy8Q9QF6BAgaEAE#imgrc=Op1wXNVWz45RGM&imgdii=Dg8lzx0OeodNSM)

Note that the emission lines in the above diagram are not representative of the exact colour or line thickness of the real hydrogen emission lines. The line thickness varies, with red normally being the broadest line and violet the thinnest.

An emission spectrum can be produced using a **gas discharge tube**. This is a glass tube from which most of the air has been removed by vacuum pump. An electrode is placed at each

end of the tube and a large DC voltage is applied between these electrodes. When the applied voltage is high enough, atoms in the gas remaining in the tube are ionised and a

stream of electrons flows from the negative electrode (cathode) to the positive electrode (anode). Such electrical discharges were first studied in detail by William Crookes in 1875 and led to the discovery of the electron. The discharge was originally called **cathode rays**. The following diagram shows a cathode ray tube with the various glows and dark spaces labeled. The voltage applied between the cathode and anode is usually supplied by an induction coil. The diagram is not to scale.



Today we know that as electrons in the atoms of the gas change energy levels, they fluoresce, that is they emit light. Different gases can be used in discharge tubes to produce different colours. Fluorescent lights and neon signs are examples of discharge tubes. Fluorescent lights emit in the ultraviolet part of the spectrum, but this part of the spectrum is invisible to our eyes. To get around this problem, the glass tubes are coated with a material that

absorbs the UV light and re-emits it in the visible part of the spectrum.

Also, the higher the voltage applied between the plates, the more energetic is the discharge and the higher its frequency. X-rays can be produced using discharge tubes.

The light produced by a discharge tube is known as an **emission spectrum**. Only colours corresponding to particular wavelengths of light are produced. In the hydrogen emission spectrum, the overall colour of the gas discharge as seen by our eyes is blue-violet. Neon, on the other hand appears red. Krypton glows lavender. Each element has its own specific emission spectrum.

**Absorption Spectrum – Ref. (19)**

As mentioned previously, in 1814 Fraunhofer discovered that the spectrum of sunlight contains hundreds of fine dark lines. These are referred to as **Fraunhofer lines**. Such a spectrum, consisting of a series of dark spectral lines among the colours of the continuous spectrum, is called an absorption spectrum.

These dark spectral lines represent wavelengths that are missing from an otherwise continuous spectrum. Absorption spectra are produced when light from a hot source of continuous spectrum passes through a cooler, non-luminous, diffuse gas. The cooler gas absorbs certain wavelengths, leaving dark spaces in their place in the spectrum. The spectra of normal cool stars such as the Sun fall into this category. Note that the dark lines in the absorption spectrum of a gas occur at the same wavelengths as the bright lines in the emission spectrum of the same gas.

The top half of the diagram below shows part of the hydrogen absorption spectrum. The dark absorption lines are clearly visible amongst the colours of the continuous spectrum. The visible hydrogen emission spectrum is supplied immediately below the absorption spectrum, showing that the absorption lines occur at the same wavelengths as the bright lines in the emission spectrum. Note too, the variation in line thickness with the alpha-line being the broadest and the delta-line being the thinnest in both spectra.

Shape, rectangle

Description automatically generated

Part of the hydrogen absorption spectrum – Diagram From: [Wikimedia Commons](https://www.google.com.au/search?q=wikimedia+commons+hydrogen+absorption+spectrum&tbm=isch&source=iu&ictx=1&fir=g7vNOZPggmteCM%252CbrtB8TZPksyKyM%252C_&vet=1&usg=AI4_-kRHTwfGCLqbws0HOTkzI1BkKMmoYw&sa=X&ved=2ahUKEwjgrfGw16zyAhUSfisKHWxvDxcQ9QF6BAgfEAE#imgrc=WVbBjh0gD2OLmM)

Absorption spectra are also produced by **reflected sunlight**. We see planets and moons in our solar system, not because they produce their own light but because of the sunlight reflected from them. When sunlight hits the planet or moon, some light is absorbed by the atmosphere and/or surface and the rest is reflected. Examination of the reflected sunlight provides information on the composition of the planet’s or moon’s atmosphere or surface. However, since a planet’s atmosphere tends to contain more complicated molecules rather than simple atoms, it is more difficult to identify the chemical composition from the absorption lines.

The Bohr Model of the Atom, which we will study in Module 8, provides an explanation of the atomic processes that produce the emission and absorption spectra of hydrogen.

**Spectroscopy**

Spectroscopyis the systematic study of spectra and spectral lines. Spectral lines are extremely important in many branches of science, especially in astrophysics. Spectral lines provide very reliable evidence about the chemical composition of distant objects, the temperature of objects, the density of objects and the motion through space of objects. Let us now examine how spectra can be used to obtain such information.

**Comparison of Emission and Absorption Spectra with the Blackbody Spectrum**

A **blackbody** is a hypothetical, ideal body whose surfaces absorb all the thermal radiation incident upon them and allow none to be reflected. All blackbodies at the same temperature emit thermal radiation with the same spectrum, independent of their composition. The intensities of the colours in the spectrum depend only on the temperature (17). Blackbodies have been used as models to help determine the relationship between the colour of light emitted by a hot body and the temperature of the body. The physics of blackbodies is well understood.

A great deal of information about stars and other celestial objects can be obtained by comparing their spectra to blackbody spectra. For example, by comparing the intensity versus wavelength curves for real stars with those of a blackbody we can determine the surface temperature of stars. By comparing a star’s absorption spectrum with a blackbody spectrum and noting the differences we can determine the chemical composition of the star. We will return to some more specifics of blackbodies later in this module.

**Measuring Astronomical Spectra**

The technology required to measure astronomical spectra is an instrument called the **spectrograph**. This is an optical device that is mounted at the focus of a telescope (19). Its purpose is to diffract light into a spectrum so that the intensity at each wavelength can be recorded by a detector. Spectrographs have been designed for use with various regions of the spectrum, with emphasis on the UV, visible and IR regions of the EM spectrum.

In a modern **spectrograph**, light from the telescope objective is focussed onto the entrance slits of the spectrograph. It is then collimated by a mirror to produce a parallel beam and directed onto a **diffraction grating**. A diffraction grating is a piece of glass onto which thousands of parallel, evenly spaced lines per millimetre have been ruled. The diffraction grating causes the light from the telescope objective to be diffracted into a spectrum by the way in which light waves leaving different parts of the grating interfere with one another. This spectrum is then passed through a corrector lens and focussed onto a charge-coupled device (CCD) that records the image. A CCD is a silicon chip containing an array of light sensitive diodes used for capturing images. The recorded spectrum is called a **spectrogram**. (16)

Older spectrographs used a **prism** to **disperse** the light and produce a spectrum in place of the diffraction grating. They also used **photographic plates** rather than CCD’s to record the spectrum (19).

If a photographic plate is used, then the result is a spectrum similar to the examples of the hydrogen emission and absorption spectra shown earlier. If a CCD is used the spectrum produced is in the form of a graph of intensity versus wavelength, in which absorption lines appear as depressions on the graph and emission lines appear as peaks. (19)

**Temperature**

Using a modern spectrograph to record the absorption spectrum of the star in question, a plot of **intensity versus wavelength** can be obtained for the star. The wavelength, max, at which the energy output of the star is a maximum, can then be determined and **Wien’s Displacement Law (below)** used to calculate the temperature of the star. (19)



This measurement would also suggest the spectral class and composition of the star.

**Density**

To determine the abundance or number densities of the atoms of elements present in a star, astronomers examine the **line strengths** of each spectral line. This is done from an **intensity versus wavelength** plot of the star’s absorption spectrum. Such a plot allows astronomers to study **the shapes of individual spectral lines**. Basically, the line strength of a spectral line depends on the number of atoms in the star’s atmosphere capable of absorbing the wavelength in question. For a given temperature, the more atoms there are, the stronger and broader the spectral line appears. This effect is called pressure broadening – the greater the atmospheric density and pressure, the greater the broadening of the spectral lines. So, by careful analysis of the star’s spectrum, an astronomer can determine the number density of atoms of a particular element in the star’s atmosphere. (16 & 19)

**Chemical Composition**

**Each element produces its own unique pattern of spectral lines (19).** To identify the elements present in the outer layers of a star, astronomers scan the absorption spectrum for that star for the unique patterns of spectral lines that correspond to particular elements. Thus, by analysing a star’s spectrum an astronomer can determine the star’s chemical composition. (19)

**The Doppler Effect for Light – Ref. (19)**

Before looking at the velocity information that is contained in stellar spectra, it is necessary to understand the Doppler Effect for light. This is the apparent change in the wavelength or frequency of a light wave when there is relative motion between the observer and the source of light. Imagine that a source, S, of light waves is moving to the right as shown in the following diagram:



The circles represent the crests of light waves emitted from various positions as the source moves along. Notice that as the source moves, the light waves emitted become crowded together in front of the source and spread out behind the source. Hence, Observer 2 sees more wavelengths reach her in a set time period than would be the case if the source were stationary. So, Observer 2 sees a higher frequency and shorter wavelength than if the source were stationary. This means that the light seen by Observer 2 is bluer in colour than if the source were stationary. So, the spectrum of light from an approaching source is blue shifted, that is all lines in the spectrum are shifted towards the short wavelength (blue) end of the spectrum.

By a similar argument Observer 1 sees a lower frequency and longer wavelength for the light than would be the case if the source were stationary. This means that the light seen by Observer 1 is redder in colour than if the source were stationary. So, the spectrum of light from a receding source is red shifted, that is all lines in the spectrum are shifted towards the long wavelength (red) end of the spectrum.

**Translational Velocity**

Stars can move through space in any direction. To calculate the **translational velocity** of a star through space (called its **space velocity**) astronomers measure two quantities. Consider the following diagram as you read on.



Firstly, astronomers determine the star’s **radial velocity** vr **parallel to our line of sight**. This is calculated from measurements of the Doppler shifts of the star’s spectral lines. Secondly, astronomers determine the star’s **tangential velocity** vt **perpendicular to our line of sight** – that is across the plane of the sky. This is calculated from knowledge of the distance, d, to the star and the star’s proper motion, which is the number of arcseconds the star appears to move per year on the celestial sphere. (19)

Once the radial and tangential components of the star’s motion are known, the star’s **space velocity v** can be calculated by simple addition of vectors (19). Note that the space velocity of a star is, by definition, its velocity relative to the Sun. To calculate an accurate value for this, the component of the Earth’s velocity around the Sun that is parallel to our line of sight to the star must be subtracted from the star’s measured radial velocity (16).

**Rotational Velocity – Ref. (19)**

To measure the rotational velocity of a star it is first necessary to obtain an intensity versus wavelength plot of the star’s spectrum. As mentioned previously, this enables astronomers to study **the shapes of individual spectral lines**.

If a star is rotating, light from the side approaching us is slightly **blue shifted**, while light from the receding side is slightly **red shifted**. As a result, **the star’s spectral lines are broadened** in a characteristic fashion. By measuring the shape of the spectral lines, astronomers can calculate the **speed of rotation of the star**.

A different case of rotation involves binary star systems. Where two stars in a binary system have their orbital plane edge on to our line of sight, their speeds of rotation about the centre of mass of the system can be determined from their spectra using the Doppler Effect. As the two stars move around, they periodically approach and recede from us. Hence, the spectral lines of the two stars will be alternately blue shifted and red shifted. From the Doppler shifts, the velocities of approach and recession can be calculated and then used to calculate the orbital period and velocities of rotation of the two stars about the centre of mass of the system.

**LIGHT: WAVE MODEL**

**Inquiry Question:** What evidence supports the classical wave model of light and what predictions can be made using this model?

**Two Competing Models of Light**

Light has fascinated human beings from the earliest times. The first recorded theory on the nature of light comes from Aristotle in the fourth century BCE (6). Centuries later, Lucretius, who, like Democritus before him, believed that matter consisted of indivisible "atoms", thought that light must be a particle given off by the sun. The **particle model of light** became the preferred model and retained that position up until the 17th Century (29).

During the late 1660’s and into the 1670’s, **Isaac Newton** did many excellent experimental studies of light (24). He favoured the particle theory and thought of light as a stream of particles which he called **corpuscles**. Newton published a very persuasive paper on the refraction of light through a prism in 1672 (30). In his work *Optiks* in 1704, he provided a full explanation of his **corpuscular (particle) theory of light** (6). Newton was able to explain light, colour, vision, reflection, refraction and dispersion.

One of the earliest references to **light having a wave nature** came from Leonardo da Vinci (1452-1519). This concept was further developed by the brilliant English scientist Robert Hooke (1635-1702), a contemporary of, and often at odds with, Isaac Newton. Hooke, in his work “Micrographia” (1665) introduced the idea of the “wavefront” of a light wave. Hooke’s pressure wave theory of light meant that light would be a longitudinal wave.

**Christiaan Huygens (1629-1695)**, a Dutch scientist, who would have been the greatest scientist of his generation, if not for living at the same time as Newton, further developed the wave theory of light (24). In 1678, he proposed his **wave model of light**. His Treatise on Light, published in 1690, was a masterful work which successfully explained the same features of light as Newton’s theory.

Both the particle and wave models of light could be used to explain the known behaviour of light up until the end of the 18th Century. Both suggested the same result for reflection, that the angle of incidence equals the angle of reflection. The wave model suggested that in refraction, light would travel more slowly in a denser medium and bend toward the normal in the denser medium. The particle model suggested that in refraction, light would travel more quickly in a denser medium and bend toward the normal in the denser medium. So, both theories predicted the same observation, that light would bend toward the normal, but they differed on the speed of light in the denser medium. Unfortunately, there was no way at that time to measure the speed of light in something like water, to compare it to the speed of light in air.

Mainly due to Newton’s prestige, the particle theory of light held sway with most scientists during the 18th century. At the turn of the 19th Century work on **interference**, **diffraction** and **polarisation** gave clear experimental evidence that light was indeed a wave motion. The particle theory of light could not explain these phenomena. The work on the polarisation of light showed that light was a **transverse wave motion**, as longitudinal waves cannot be polarised. Further clear evidence in favour of the wave model came from the measurement of the speed of light in water, accomplished by Foucault in 1850. Foucault, and a little while later Fizeau, both showed that light travelled more slowly in water than in air. (24)

Let us now investigate the interference, diffraction and polarisation of light.

**Huygens’ Principle and Diffraction**

**Huygens’ Principle** states that every point on a wavefront acts a source of circular secondary wavelets that spread out in the forward direction at the speed of the wave itself. The new wavefront is the envelope of all the secondary wavelets – that is, the tangent to all of them. (32)

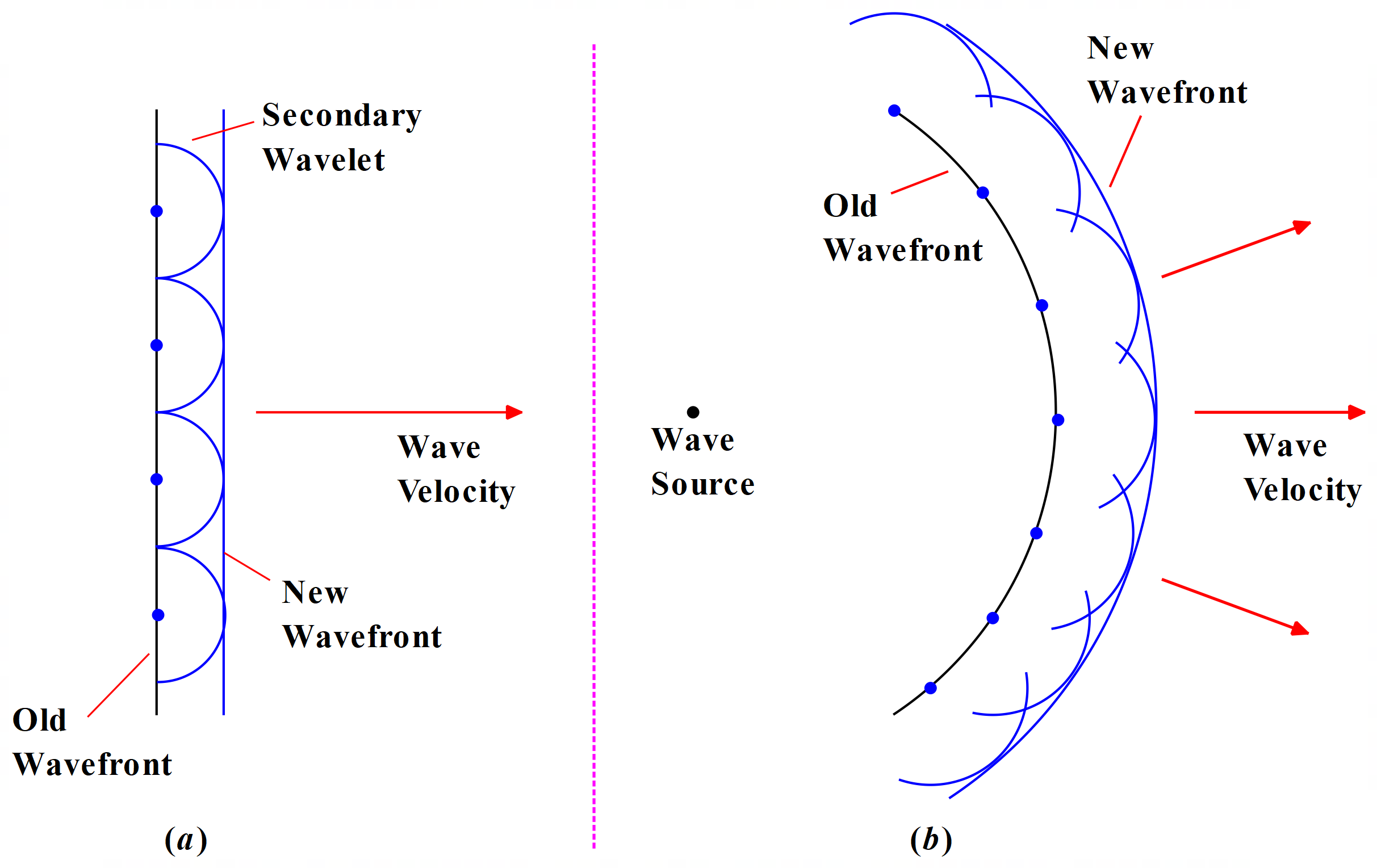


Diagram (a) shows the formation of a new plane wavefront. Diagram (b) shows the formation of a new circular wavefront.

Huygens’ Principle can be used to derive the laws of reflection and refraction. It is also extremely useful in explaining the diffraction of waves around an obstacle, as it predicts that waves will bend around an obstacle.

You will recall from Module 3, that **diffraction** is the spreading of waves as they pass through a small aperture (hole) or move around the edges of an obstacle. Diffraction is most prominent when the size of the aperture or obstacle is of the same order as the wavelength of the wave. If the aperture or obstacle is much larger than the wavelength, diffraction goes unnoticed. The diffraction of light was first recorded by the Jesuit priest Francesco Grimaldi (1618-1663). He noticed that when sunlight entered a darkened room through a tiny hole, the spot on the opposite wall was larger than would have been expected from geometric rays and surrounded by coloured fringes. He named the phenomenon diffraction. (24)

**The Interference of Light**

You will also remember from Module 3 that we discussed the **interference** of waves in relation to the superposition principle. When two or more waves move through the same medium at the same time, they interfere with each other so that the amplitude of the resultant wave at each point is the sum of the of the amplitudes of the individual waves at each point.

Constructive interference occurs where two wave crests or two wave troughs coincide with each other and add together to produce the maximum possible amplitude of the resultant wave. Destructive interference occurs where a crest and a trough coincide with each other to produce the minimum possible amplitude of the resultant wave. If the two interfering waves are identical with each other, then the maximum amplitude of the resultant will be twice the amplitude of each individual wave and the minimum amplitude will be zero. (6)

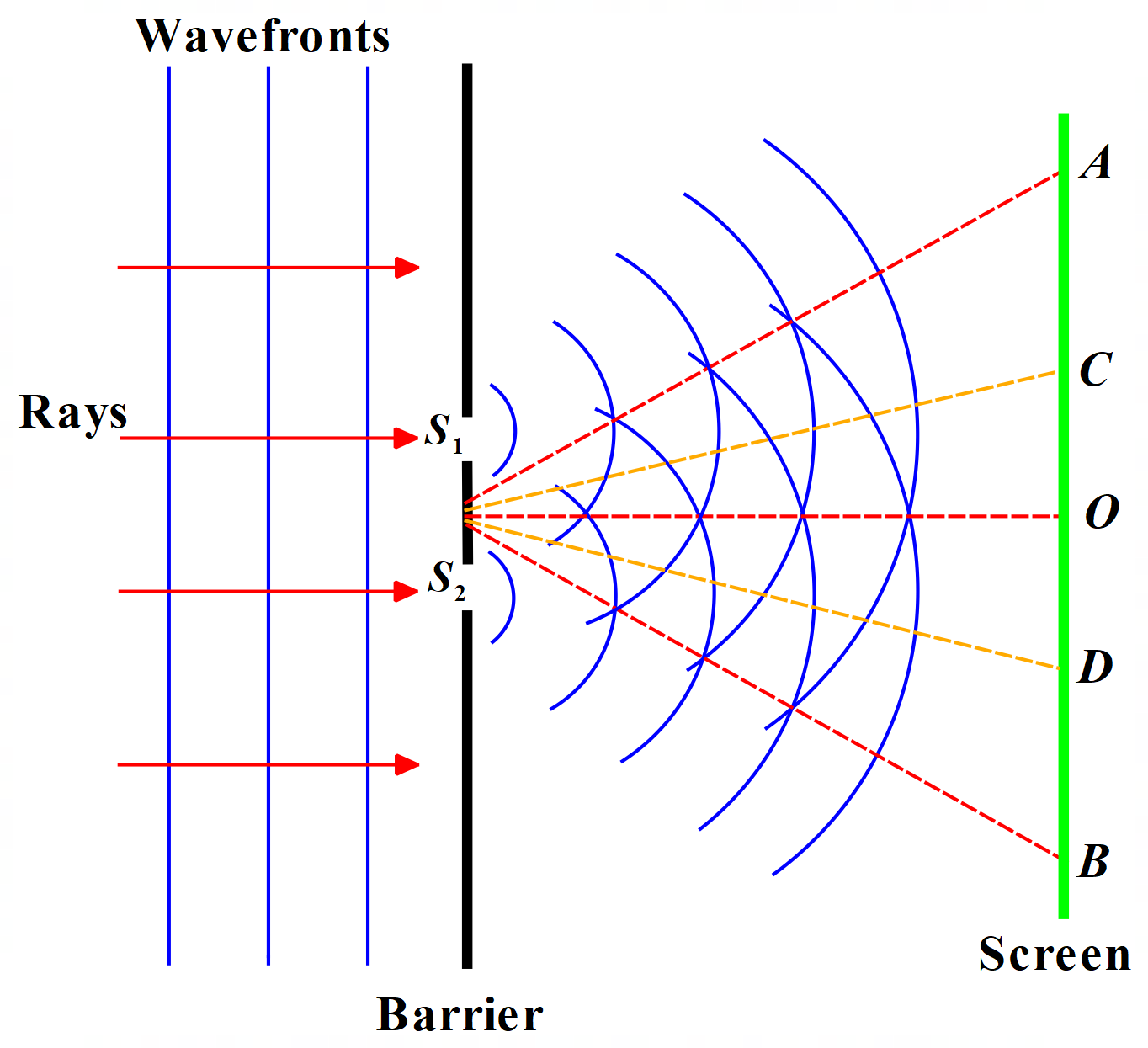
If the interference of any two waves is to be detected, the following conditions must apply (25):

* The two sources must be coherent – that is, they must emit waves that maintain a constant phase relationship with each other and have the same frequency (and therefore, wavelength).
* The sources should emit waves of about the same amplitude, so that almost complete constructive and destructive interference occurs, thus producing maximum contrast.

**Young’s Double Slit Experiment**

In 1801, the Englishman Thomas Young (1773-1829) performed his famous **double-slit experiment**. This provided convincing evidence that light had a wave nature. Young created a single source of light by having sunlight pass through a tiny slit, S. This light then fell on a screen containing two closely spaced, tiny slits, S1 and S2, equidistant and a long distance from S. The light falling on S1 and S2 could be considered to consist of parallel rays of light due to the distance of S from S1 and S2. (7)

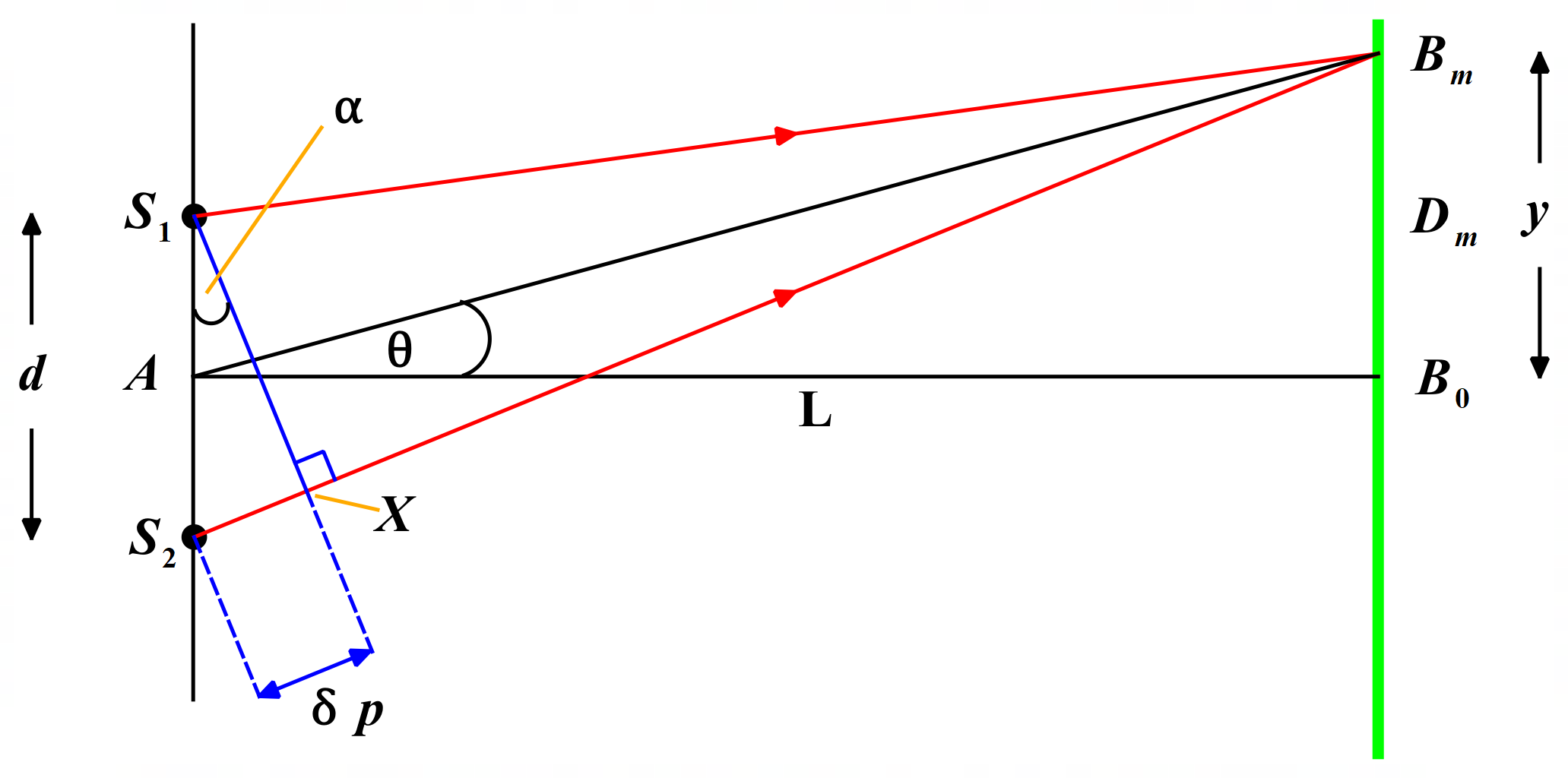
The slits and their separation were very narrow, not much wider than the wavelength of light. If light consisted of particles, we would expect to see at the most two bright lines on a screen placed behind the slits. However, Young observed a series of bright and dark lines (fringes or bands) on the screen. Young could only explain these fringes as the result of wave interference. Hence, light must have a wave nature. (7)



Today we tend to use light of a single wavelength, called monochromatic (single colour) light as the source, rather than sunlight. The diagram above shows light rays from a monochromatic source impinging on two very narrow slits. As the plane waves pass through the apertures, they diffract and then interfere with each other. Circular wavefronts spread out in phase from S1 and S2, which act as coherent sources. The circular wavefronts shown represent crests. Where a crest from S1 coincides with a crest from S2, we get constructive interference. The amplitudes of the two waves add together to give a higher amplitude, and we observe a bright band of light on the screen. This occurs at O, the central maximum, and at A and B, called first order maxima. (32)

Where a crest from S1 coincides with a trough from S2, or vice versa, we get destructive interference. The amplitudes of the two waves add together to give zero amplitude, and we observe a dark band on the screen. This occurs at C and D, called first order minima. (32)

Let us now analyse this situation mathematically. See diagram below.



In the diagram above the two slits are separated by a distance d, which in reality is very small compared to the distance to the screen, L. In real situations, the rays from each slit arriving at a particular point on the screen will be parallel, with  being the angle they make with the horizontal. B0 is the central maximum (central bright band on the screen). Dm represents the m-th minimum (dark band) counting from the central maximum and Bm, the m-th maximum (bright band). The distance from B0 to Bm is y. So, y is the distance from the central maximum to the m-th bright fringe. (25)

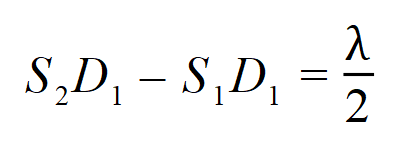
**Central Maximum (B0)**

The waves from S1 and S2 arrive at B0 in phase, their paths S1B0 and S2B0 being equal in length, and reinforce one another. A bright band results. This bright band is often called the zeroth-order maximum. All bright bands are regions of constructive interference (antinodal lines). The central bright band is the most intense. The intensity of the bright bands drops off the further they are from the central maximum. The amount by which the intensity drops off as we move away from the central maximum depends on the width of each slit. (25)

Note the rays for S1B0 and S2B0 are not shown above in order to keep the diagram uncluttered.

**First Order Minimum (D1)**

On either side of the central maximum a dark band occurs. This first order minimum is caused by the waves from S1 and S2 being out of phase by half a wavelength. A crest from one source coincides with a trough from the other when they hit the screen. Thus, annulment occurs. The difference in their path lengths\*, p, must equal half a wavelength. Note the rays for S1D1 and S2D1 are not shown above in order to keep the diagram uncluttered. If you need to, check out the wavefronts diagram on the previous page to see the arrangement of wavefronts that produce the minima (C & D) in that diagram.

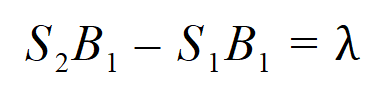


All dark bands are regions of destructive interference (nodal lines). (25)

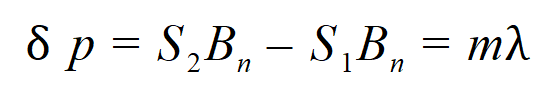
\*The **path length** of a ray is simply the distance the ray travels in moving from one point to another.

**First Order Maximum (B1)**

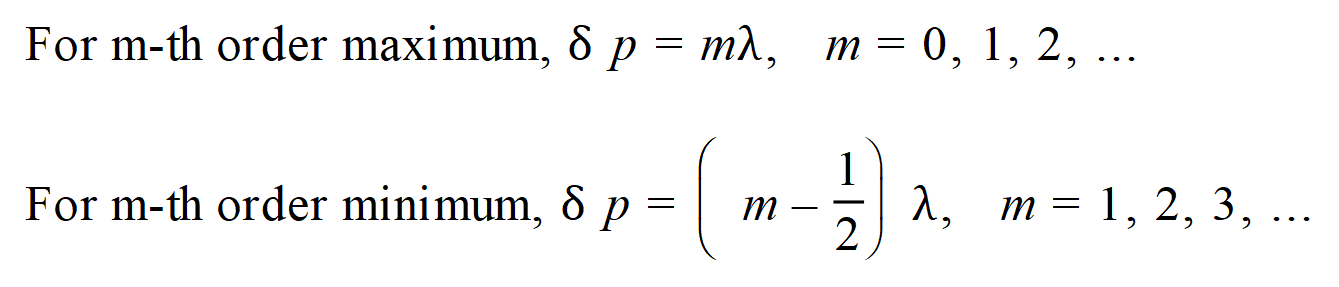
The light waves from S1 and S2 reach B1 in phase and cause maximum reinforcement. Thus, the path length difference between S2B1 and S1B1 must be one wavelength.



Clearly, if Bm is the position of the m-th order maximum, the path length difference, p, between the two light waves which reach Bm in phase equals m wavelengths. (25)



**Path Length Difference Equations**



So, for instance, the second order **bright** band, m = 2, would result from a path difference of:

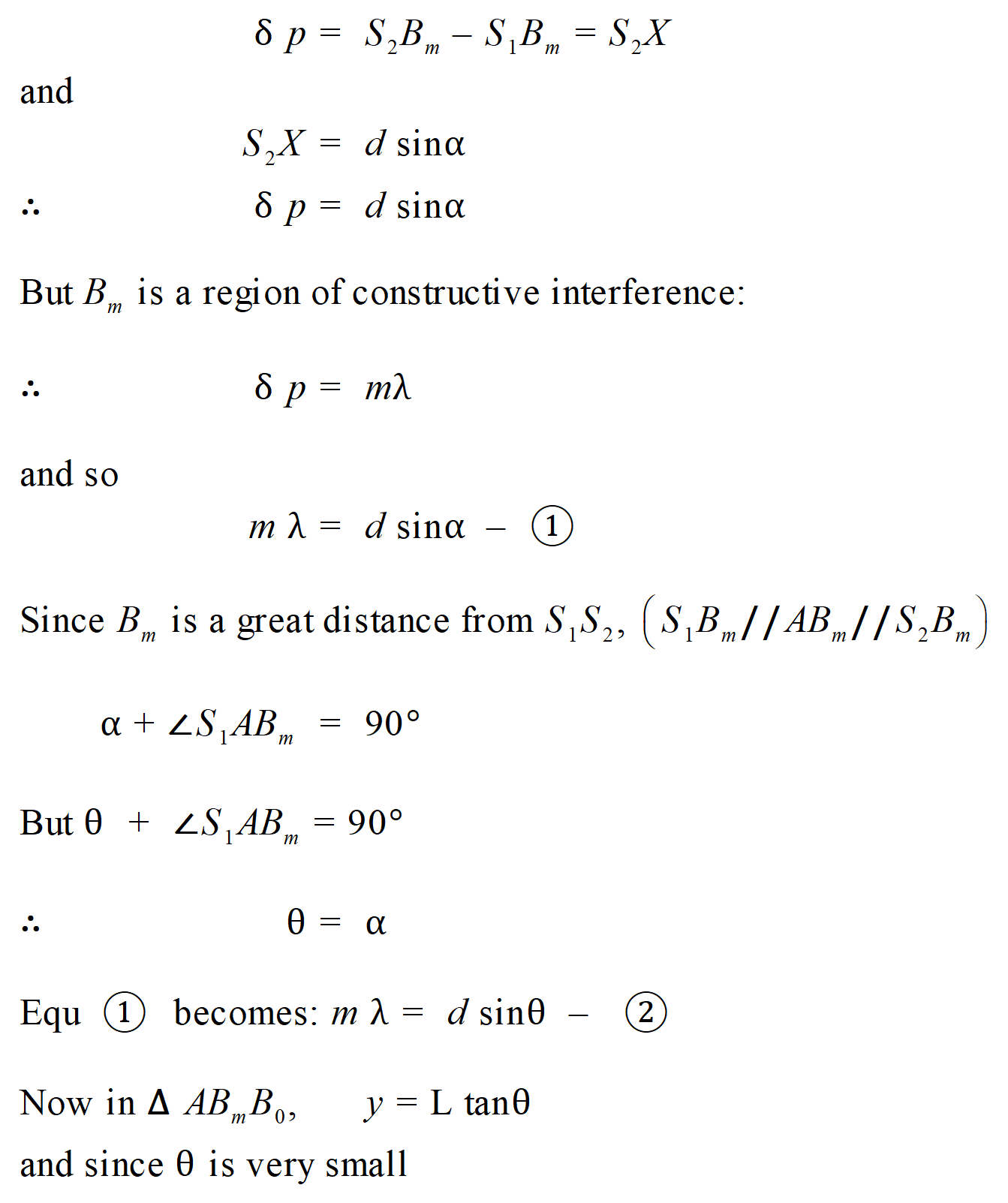
p = 2. The second order **dark** band, m = 2, results from a path difference of: p = (2 - ½) , so p = 3/2 . (25)

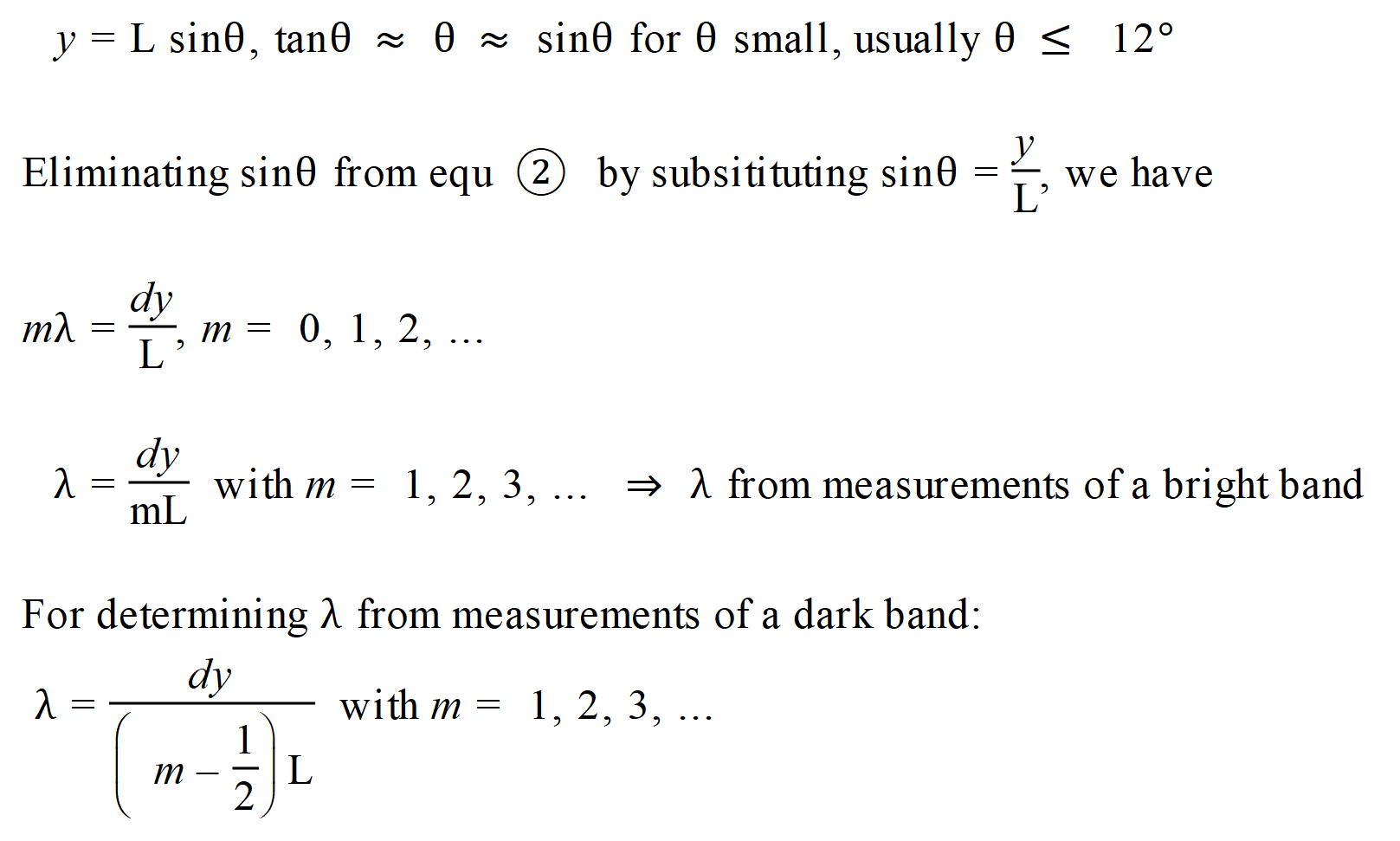
Just a word of caution – don’t panic if you find variation between textbooks on the formula for the path length difference for minima (dark bands). Some books use (m + ½). This means you have to remember that the 1st order minimum occurs for m = 0. The second order minimum occurs for m = 1. I prefer to use the (m - ½) formula. That way the numbers are easier to remember. The first order minimum occurs for m = 1. And so on.

**Determination of the Wavelength of Light – Ref. (25)**

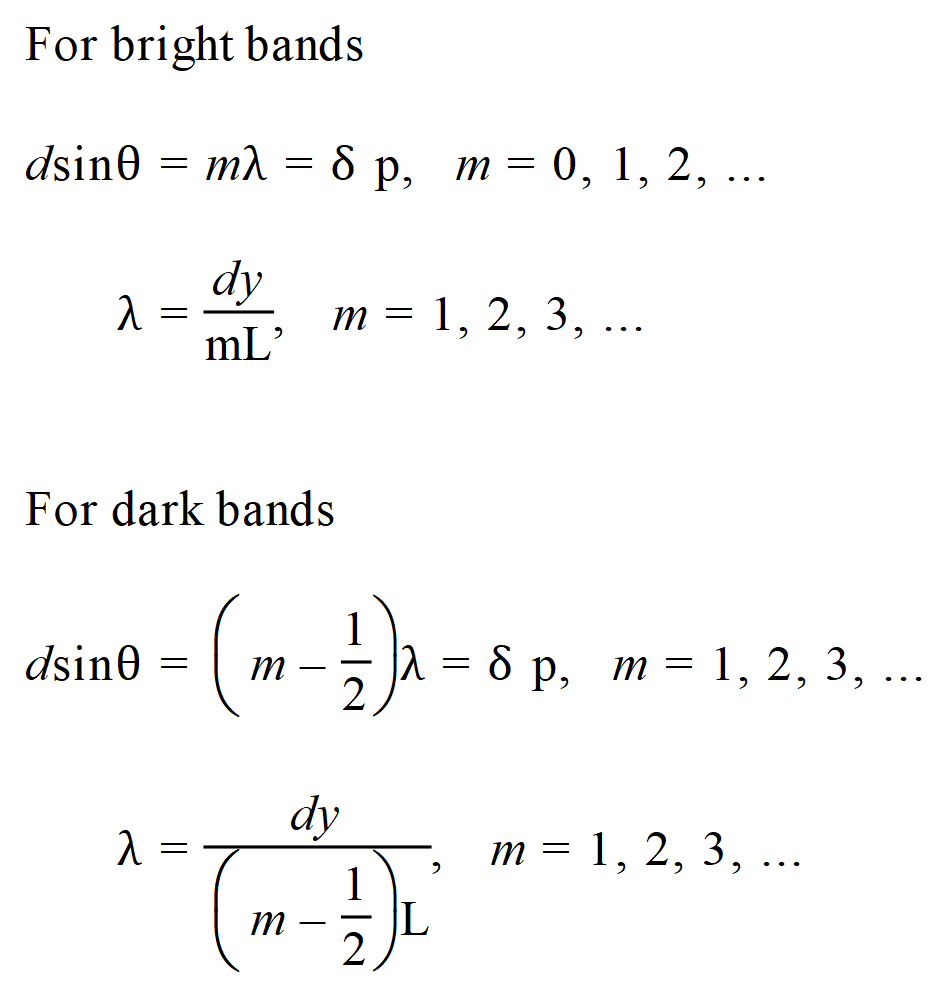
Mathematics is a beautiful thing and here we are going to see yet another example of how wonderful mathematics is at modelling our world. Let us analyse Young’s Double Slit experiment a little further to discover how it can be used to measure the wavelength of light. This is an amazing achievement when you think of how small that really is.

Study the diagram on the previous page. The path difference, p, between the rays reaching and Bm from S1 and S2 respectively is:





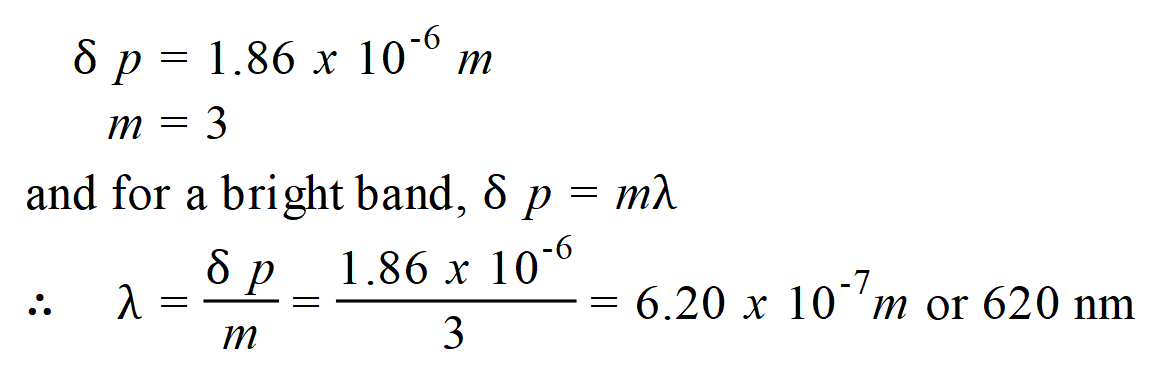
From the above analysis it is clear that the following equations can be used to determine the wavelength, , of a source of light using Young’s double-slit experiment. Which equation you use depends upon the data obtained in the experiment.



Another term to be familiar with in interference work is **fringe separation** also called **fringe width** or **band width**. This is the distance between successive bright fringes. If you are given this as data or asked to find it from other data, you can easily do so using the equations containing y above. Remember y is the distance from the central maximum to the m-th bright fringe. The fringe separation is essentially uniform until the approximation sin  ≈  is no longer valid. So, close to the centre of the interference pattern, the bands will certainly be uniformly spaced.

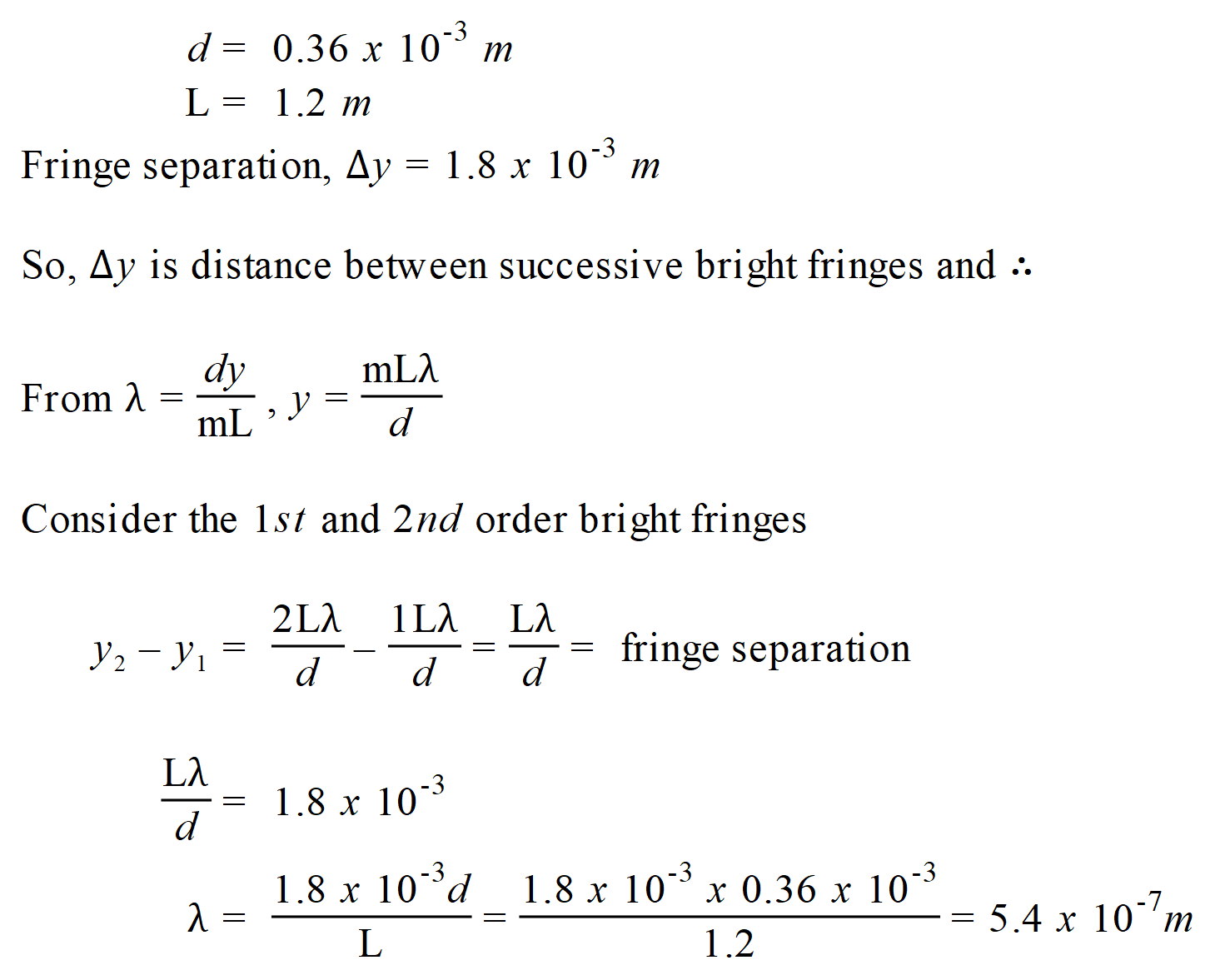
Example Question (1) – In Young’s experiment, the difference in path lengths from the slits to the third bright fringe was found to be 1.86 m. What was the wavelength of the light?

Solution



Example Question (2) – Light from a narrow slit illuminates two narrow slits of separation 0.36 mm and forms an interference pattern on a screen 1.2 m away. If the bright fringes are 1.8 mm apart, calculate the wavelength of the light.

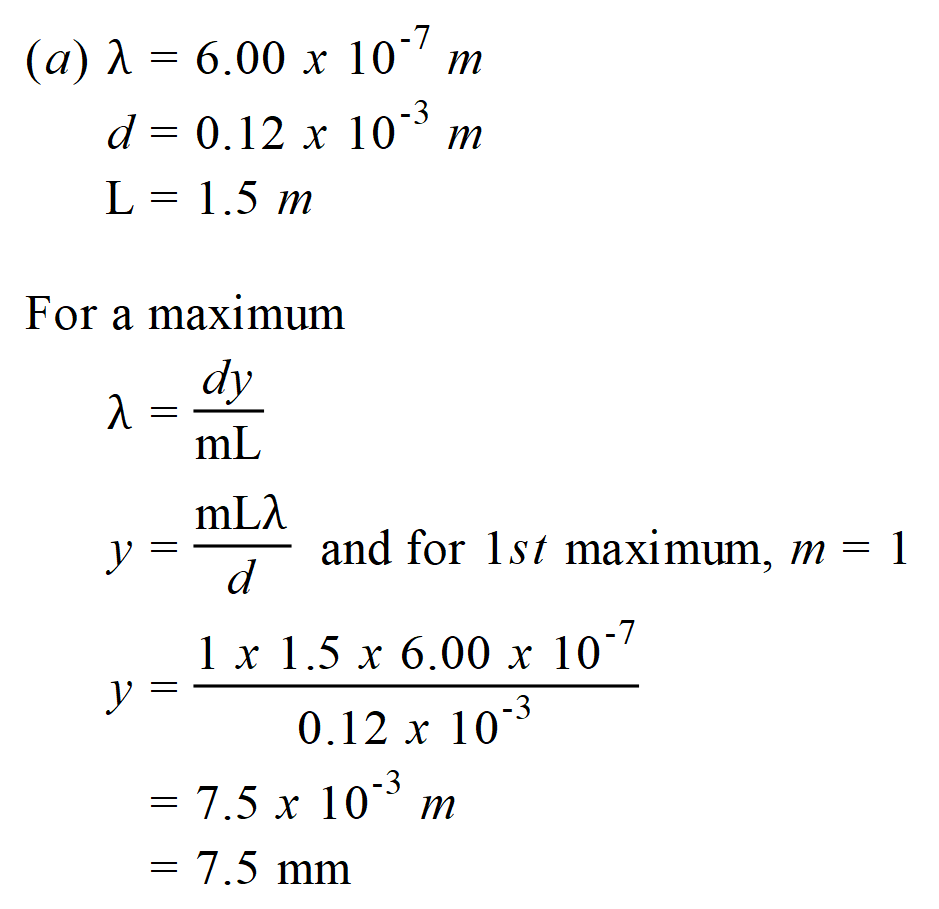
Solution

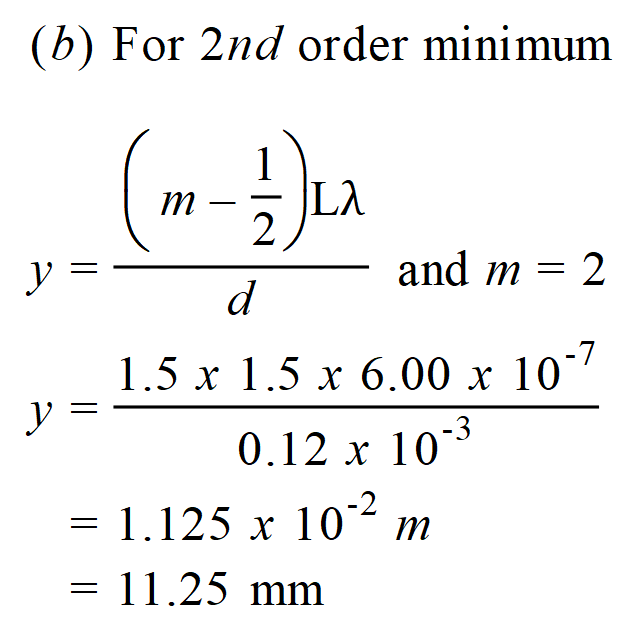


You can remember the formula for fringe separation if you wish, but clearly you can easily work it out when you need it from what you already know.

Example Question (3) – The wavelength of red light is about 600 nm. Two parallel slits 0.12 m apart are illuminated with red light and an interference pattern falls on a screen 1.5 m from the slits. Calculate: (a) The distance from the central maximum to the first bright band. (b) The distance to the second dark band.

Solution





Example Question (4) – See this video from [Khan Academy](https://www.khanacademy.org/science/physics/light-waves/interference-of-light-waves/v/youngs-double-slit-problem-solving).

For a good introduction to the performance of the experiment itself, see [Khan Academy](https://www.youtube.com/watch?v=nuaHY5lj2AA).

**Diffraction Gratings – Ref. (32)**

When there is a need for high resolution separation of different wavelengths, a **diffraction grating** is most often the optical tool used. Diffraction gratings can be either transmission gratings or reflection gratings. We will only consider transmission gratings here.

A diffraction grating (transmission) consists of thousands of equally spaced scratches on a flat transparent material, commonly glass or plastic. The scratches are opaque to light and the spaces between are transparent, giving rise to many coherent sources of light – eg 10,000 per centimetre.

The many parallel slits act as Young’s slits. The slit width is small, resulting in a wide diffraction central maximum. The slit separation is also small, resulting in widely spaced interference fringes.

When monochromatic light falls on the grating, a central maximum is observed with further maxima on either side. These are not uniformly spaced. The maxima are referred to as first order, second order and so on. If white light is used, a spectrum is observed at the position of each of the maxima. If a laser is used as the light source, the interference pattern will consist of very bright dots separated by complete darkness.

The defining equation for a diffraction grating is: **d sin = m**, where m = order of the fringe,  = wavelength of light used, d = separation of slits in the grating and  = angle between the central maximum and the m-th order bright fringe. This is the same equation as for Young’s double slit experiment. A diffraction grating would be better called an interference grating, but there you go.

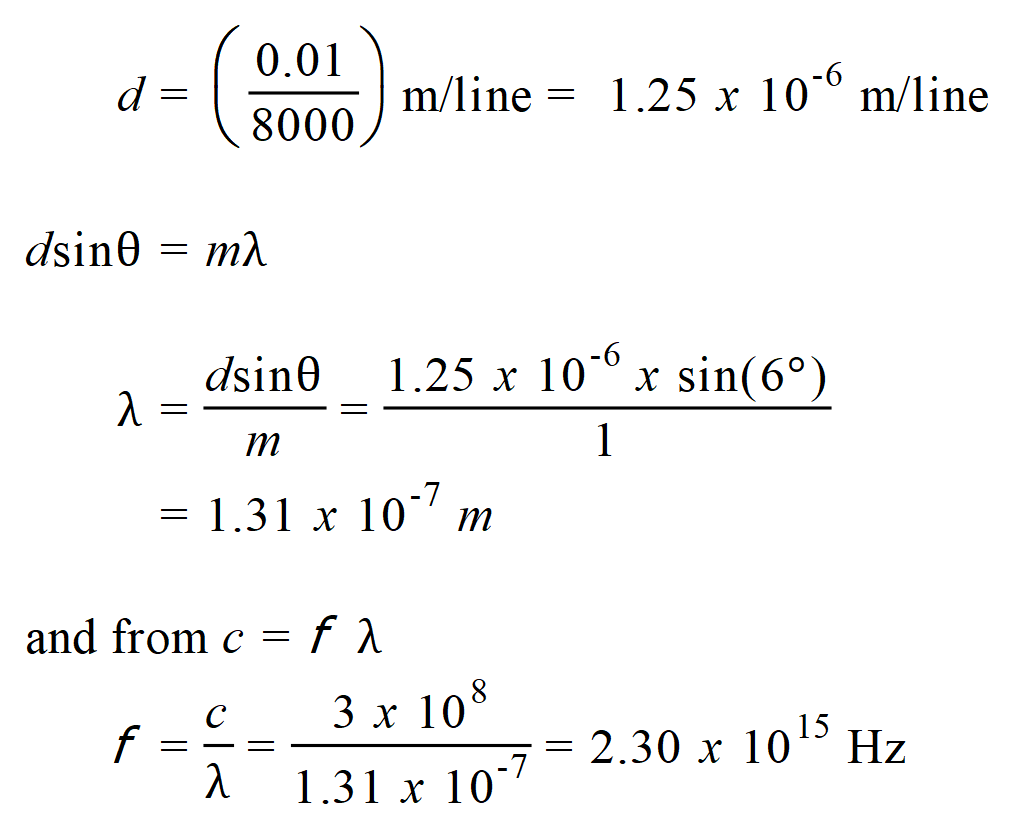
View the excellent video on the diffraction grating by the [Khan Academy](https://www.khanacademy.org/science/physics/light-waves/interference-of-light-waves/v/diffraction-grating). Bear with it. The explanation gets a bit heavy at one point but it’s worth watching all the way through.

Clearly, if you need to do calculations using a dark band, you use the same equation but with (m – ½) in place of m.

There are some good photos & diagrams of diffraction gratings and spectra produced, [here](http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/grating.html). Follow the various links.

Example Question – A diffraction grating has 8000 lines per centimetre and is used to determine the wavelength of light. It is observed that there is a bright line 6° either side of the central maximum. What is the wavelength and frequency of the light being observed?

Solution



Hopefully, your teacher will involve you in some practical work using a **variable deviation spectrometer**. This can be used with a diffraction grating and a laser light source to produce excellent interference patterns which can be used to determine the wavelength of the laser light. The variable deviation spectrometer can also be used with emission lamps of various elements and either a diffraction grating or a triangular glass prism to measure the wavelengths of light comprising the emission spectra of those elements.

**Diffraction of Light Through A Single Slit & Around An Obstacle – Ref. (25)**

Watch this brief [Institute of Physics](https://www.youtube.com/watch?v=71Rp-jG6Eek) video which shows the diffraction of laser light around a thin wire.

In the early 1800’s, as scientists began studying the effects of passing light through small apertures and around the edges of obstacles, it was noticed that the **diffraction pattern** produced by light passing through or around such objects was more complex than was predicted. In the case of light passing through a single slit for instance, instead of showing just a single image of the slit broadened (made wider) by diffraction, the pattern was similar to the interference pattern produced from two slits, with the following exceptions:

* The central maximum is twice the width of the outer maxima; and
* The intensity drops off very rapidly as we move away from the centre.

As surprising as this pattern was to the early investigators, it has a logical explanation. Huygens’ Principle supplies the answer. Each point on a wavefront acts as a source of circular secondary wavelets. As the light wave moves through the gap, typically 0.1 mm in width, it is these secondary wavelets that interfere with each other to produce the interference pattern on the screen.

Watch the first 5 minutes of the video from [Khan Academy](https://www.khanacademy.org/science/physics/light-waves/interference-of-light-waves/v/single-slit-interference) on single slit diffraction, before moving on. It will aid your understanding. You **do not** need to watch the derivation of the equation for single slit diffraction. It is not required by the Syllabus.

There are two different groups of diffraction effects with light. **Fraunhofer diffraction** (named after Joseph von Fraunhofer) occurs when both the light source and the screen are far away from the diffracting aperture. The rays of light involved are essentially parallel. Fraunhofer diffraction is relatively easy to treat mathematically. This is the mathematical treatment that would be used at Stage 6 level if the Syllabus required use of the single slit diffraction equation.

**Fresnel diffraction** (named after Augustin Fresnel) occurs when the light source or screen or both are a short distance from the obstacle causing diffraction. An example of this is diffraction at a straight edge, where light falling on the edge spreads into the geometric shadow zone. An example would be laser light striking the edge of a razor blade. The diffraction of the laser light around the wire that you watched at the start of this section would also be Fresnel diffraction. The mathematics of Fresnel diffraction is more complicated than that of Fraunhofer diffraction. Both types of diffraction are explained using Huygens’ Principle.

**Analysis of Experimental Evidence Supporting the Theories of Newton & Huygens – Refs. (29) & (33) to (39)**

Just a quick comment – I have noticed that some HSC texts on The Nature of Light seem to misunderstand the syllabus requirement here. We are asked to “analyse” the experimental evidence supporting these theories. That means we need to identify the components of the experiments and the relationship between them and draw out and relate the implications of these. Most texts seem to compare & contrast or simply describe.

**Newton’s Corpuscular Theory of Light**

Newton’s theory states that light consists of a stream of discrete, extremely small particles called corpuscles. These corpuscles are perfectly elastic, rigid and light. Every luminous source emits these corpuscles. The theory was successful in explain several features of light in a way that was supported by experimental observations.

**Propagation** – observation suggests that light travels in straight lines. Sunlight forms sharp shadows behind objects it falls upon. This supports Newton’s model. Newton explained that light travels so quickly that the effect of gravity on a horizontal beam of corpuscles is negligible. Hence although the beam does move in a parabolic arc, like a ball thrown through the air, the arc is too small to see and so the beam appears to travel in a straight line.

**Colour** – observation shows that white light can be split into different colours by a prism. This supports Newton’s theory, which explains that colour is a property of light produced by the different sizes of corpuscles within the stream. Even more, Newton’s own experimental genius showed that a beam of white light could be split into a coloured spectrum and then recombined into white light. Clearly, colour was a property of light itself not something that happened as a result of light’s interaction with matter. Note, however, Newton never came up with an experiment to verify that it was indeed the size of corpuscles that produced colour.

**Reflection** – observation shows that when light reflects, the angle of incidence equals the angle of reflection. This supports Newton’s model which proposes that when a corpuscle hits a solid surface, it is repelled by the reaction force from that surface in a symmetrical, elastic way, such that the angle of incidence equals the angle of reflection. This is the same as we observe for an elastic ball thrown at an angle to a hard surface such as a floor or wall.

**Refraction** – observation shows that when light moves from a less dense to a more dense medium, from air into water for instance, the light bends toward the normal to the interface. Snell’s law explains the relationship between the angle of incidence and the angle of refraction. This observation originally supported Newton’s theory, which explained that refraction occurred due to a change in the velocity of the stream of corpuscles as it moved into the denser medium. Newton argued that close to the interface, the concentration of mass in the denser material produced a gravitational attraction on the corpuscles, thus speeding them up in the direction of the interface. No change in velocity perpendicular to the direction of travel occurs. Thus, the light travels more distance through the denser medium in a shorter period of time than when in the air, and therefore bends toward the normal to the interface. Note that Newton’s theory predicts an increase in velocity in the denser medium. This was eventually proven to be incorrect. As mentioned already, in 1850, Foucault and Fizeau showed experimentally that light travels more slowly in a denser medium than in a less dense medium. Thus, Newton’s explanation of refraction was incorrect.

**Polarisation** – observation shows that light can be polarised. Newton did propose that this could be explained if the corpuscles had different shapes that enabled some of them to pass through a polarising filter but not others. He did not come up with a viable experimental demonstration in support of this proposal. Most scientists consider, therefore, that his theory could not explain polarisation in a scientific manner.

**Diffraction and Interference** – observation shows that light undergoes diffraction and interference. Young’s double slit experiment indicates light undergoes interference. Fresnel’s and Fraunhofer’s theories and supporting experiments, done by them and many others, demonstrate that light undergoes diffraction. Both diffraction & interference are wave effects and cannot be explained by Newton’s corpuscular theory of light.

**Huygens’ Wave Theory of Light**

Huygens’ theory states that light travels in the form of waves. Wavefronts are surfaces over which the light wave has constant phase. Every point on a wavefront may be considered a source of secondary spherical wavefronts which spread out in the forward direction at the speed of light. The new wavefront is the tangential surface to all those secondary wavelets.

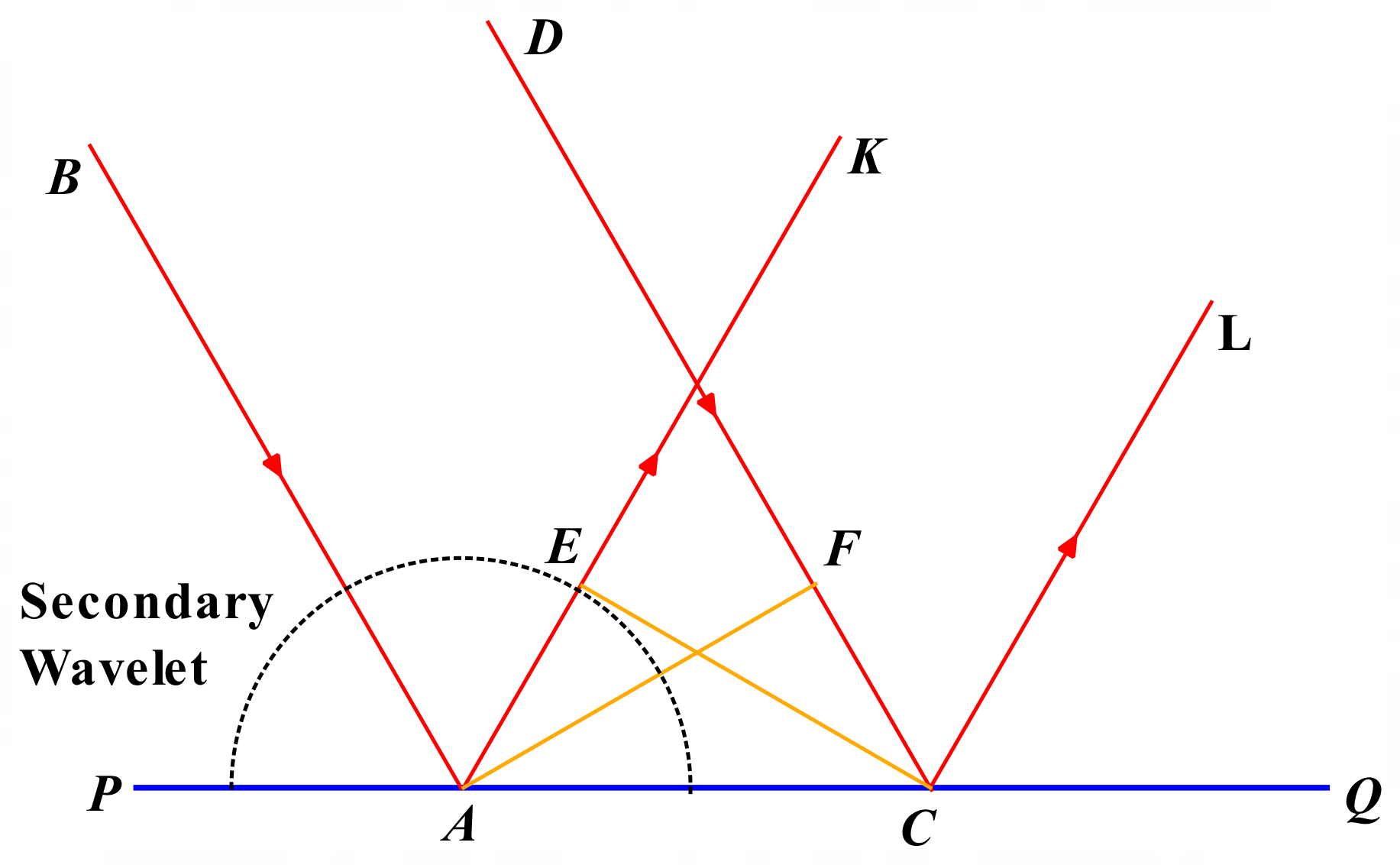
Crucial to his wave theory was the result recently obtained by Ole Roemer (1675) that the speed of light is finite. He considered light waves propagating through an aether, just as sound waves propagate through air. The mystical aether was believed to be a weightless substance, which exists as an invisible entity throughout air and space. He explained the high but finite speed of light by the elastic collisions of a succession of spheres that composed the aether. Light waves, according to Huygens, were thus longitudinal waves. It was his reliance on the aether, a very strange medium indeed, that caused most scientists of the day to find Huygens’ model of light less attractive than Newton’s.

**Propagation** – observation suggests that light travels in straight lines. This supports Huygens’ model, since Huygens could explain how a plane wavefront would move forward by using his principle. The envelope of secondary wavelets forms the new wavefront in the direction of travel of the wave. The direction of travel of a wave at any point is the direction of the normal to the wavefront at that point. Newton and others had issues with the whole idea of secondary wavelets and dismissed this explanation as nonsense.

**Colour** – observation shows that white light can be split into different colours by a prism. Huygens’ theory did not include colour as part of the nature of light. Light was thought of as white light. Advocates of the wave theory argued that colour was produced due to corruption within the glass. So, colour was an external feature not a property of light itself. Today, of course, we know that the colour of light is due to its wavelength (or frequency). So, the modern wave theory of light does explain the colour of light.

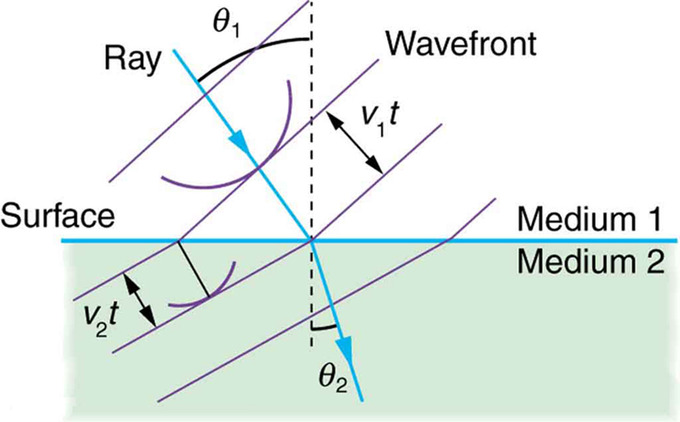
**Reflection** – observation shows that when light reflects, the angle of incidence equals the angle of reflection. This supports Huygens’ theory, which explains reflection using the Huygens’ Principle. The principle can be used to show how when a wavefront strikes a barrier, the secondary wavelets moving back off the barrier produce a reflected wavefront such that the angle of incidence equals the angle of reflection.

The following diagram (adapted from Ref. 33) shows **Huygens’ construction** for reflection. AB and CD are rays of light and AF is a wavefront striking the reflecting surface PQ at A. A secondary wavelet can be seen reflecting from A. A series of such wavelets from each point on AF eventually form the reflected wavefront CE, travelling in the direction indicated by rays AK and CL. The law of reflection can be derived using the geometry of the diagram. We will not do that derivation here.



**Refraction –** observation shows that when light moves from a less dense to a more dense medium, from air into water for instance, the light bends toward the normal to the interface. This supports Huygens’ theory. Huygens believed that the velocity of light in any substance was inversely proportional to its refractive index. So, light would travel more slowly in a denser medium. Using this supposition, Huygens’ Principle can be used to show that secondary wavelets that spread out from a wavefront striking an interface between two media of different densities and which travel more slowly in the denser medium, produce a new wavefront that has changed its direction, such that the angle of refraction is less than the angle of incidence. Snell’s law explains the relationship between the angle of incidence and the angle of refraction. Huygens’ supposition that the speed of light is lower in a denser medium was verified by both Foucault and Fizeau in separate experiments in 1850.

See diagram below showing Huygens’ construction for refraction (from Ref.31).



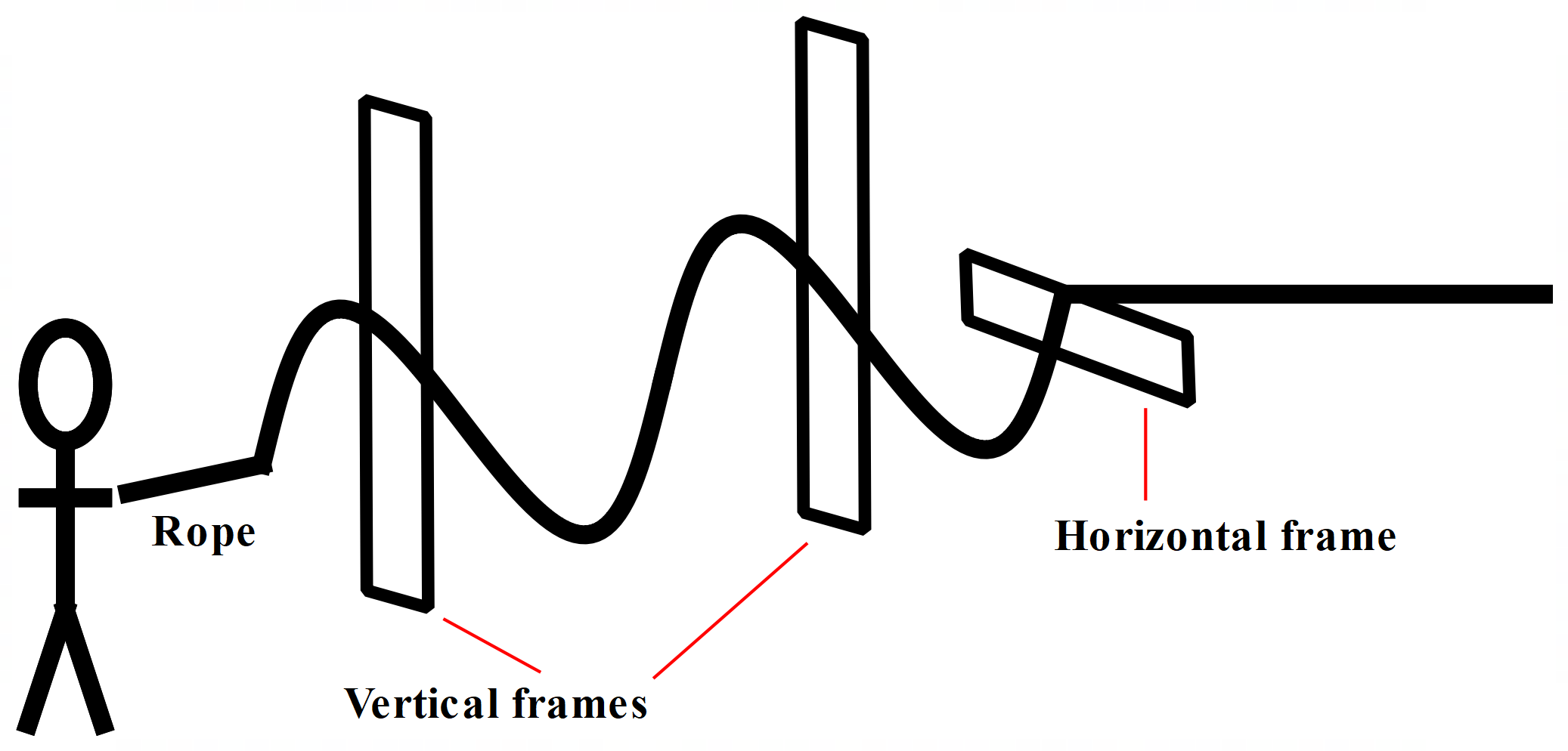
**Huygens’s Refraction**: Huygens’s principle applied to a straight wavefront traveling from one medium to another where its speed is less. The ray bends toward the perpendicular, since the wavelets have a lower speed in the second medium. Diagram taken without alteration from: <https://courses.lumenlearning.com/boundless-physics/chapter/diffraction/>;

link to license: <https://creativecommons.org/licenses/by-sa/4.0/legalcode>

**Diffraction, Interference & Polarisation** – observation shows that light undergoes diffraction, interference and polarisation. Young’s double slit experiment indicates light undergoes interference. Fresnel’s and Fraunhofer’s theories and supporting experiments, done by them and many others, demonstrate that light undergoes diffraction. Both diffraction & interference are wave effects and can be explained using Huygens’ Principle. For instance, the diffraction pattern formed when light passes through a single tiny slit can be explained by the diffraction and interference of secondary wavelets as the light passes through the slit. Polarisation cannot be explained by Huygens’ theory, since it can only occur with transverse waves and Huygens believed light to be a longitudinal wave.

**Polarisation**

In the figure below, a thick rope has been threaded between two vertical frames and a horizontal frame. The frames act as slots through which the rope will oscillate. When a vertical transverse wave is sent along the rope, the wave passes easily through the two vertical slots but is stopped by the horizontal slot. The wave is said to be **polarised** in the vertical plane. That is, its only plane of oscillation is vertical. Waves that are polarised so that they only vibrate in one plane are said to be **plane polarised** waves.

****

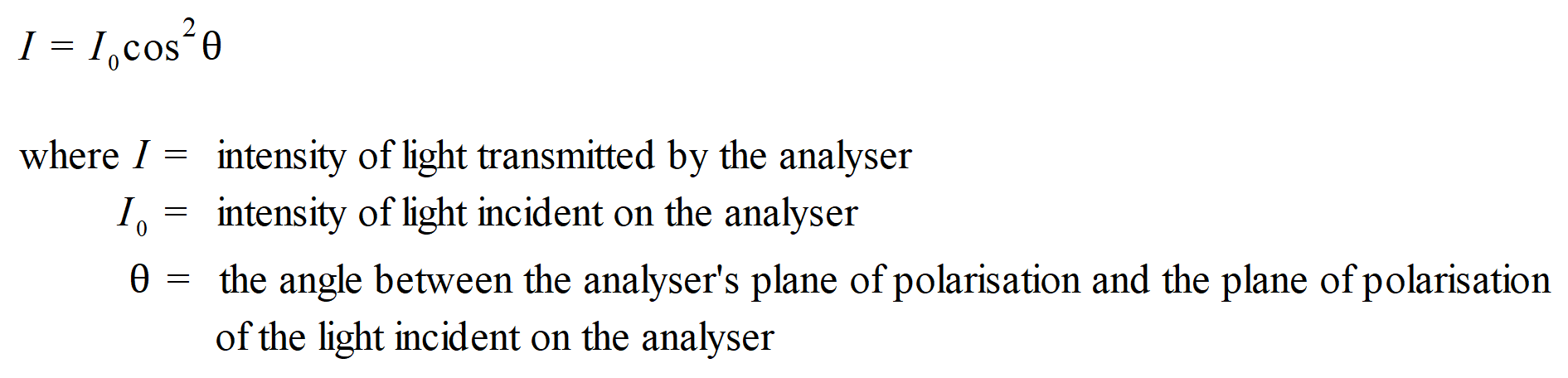
**Polarisation and Malus’ Law – Refs. (24) & (25)**

That light could be polarised was known probably since the mid-17th century (the work of Bartholin with calcite crystal). In the early 1800’s, Etienne Louis Malus (1775-1812) did some meticulous work on the polarisation of light. He is responsible for naming the phenomenon “polarisation” and he also determined the relationship between the intensity of light passing through two polarisers and the angle between their planes of polarisation. Many others did important work on the polarisation of light: especially Sir David Brewster (1812), Francois Arago (1812) and Augustin Fresnel (1817 onward). It was not until Maxwell produced his laws of electromagnetism, however, and Hertz performed his experimental confirmation of EM waves, that the nature of the polarisation of light was fully understood.

A beam of light contains an immense number of waves vibrating in every possible plane. Each wave can be resolved into vertical and horizontal components, usually of the electric field. Thus, it averages out as if half of the waves vibrate vertically and half vibrate horizontally. If a filter called a **polariser**, such as a pair of polaroid sunglasses, is placed in front of a beam of unpolarised light, only those waves that vibrate parallel to the **plane of polarisation** of the sunglasses can pass through. Thus, half of the light rays are eliminated.

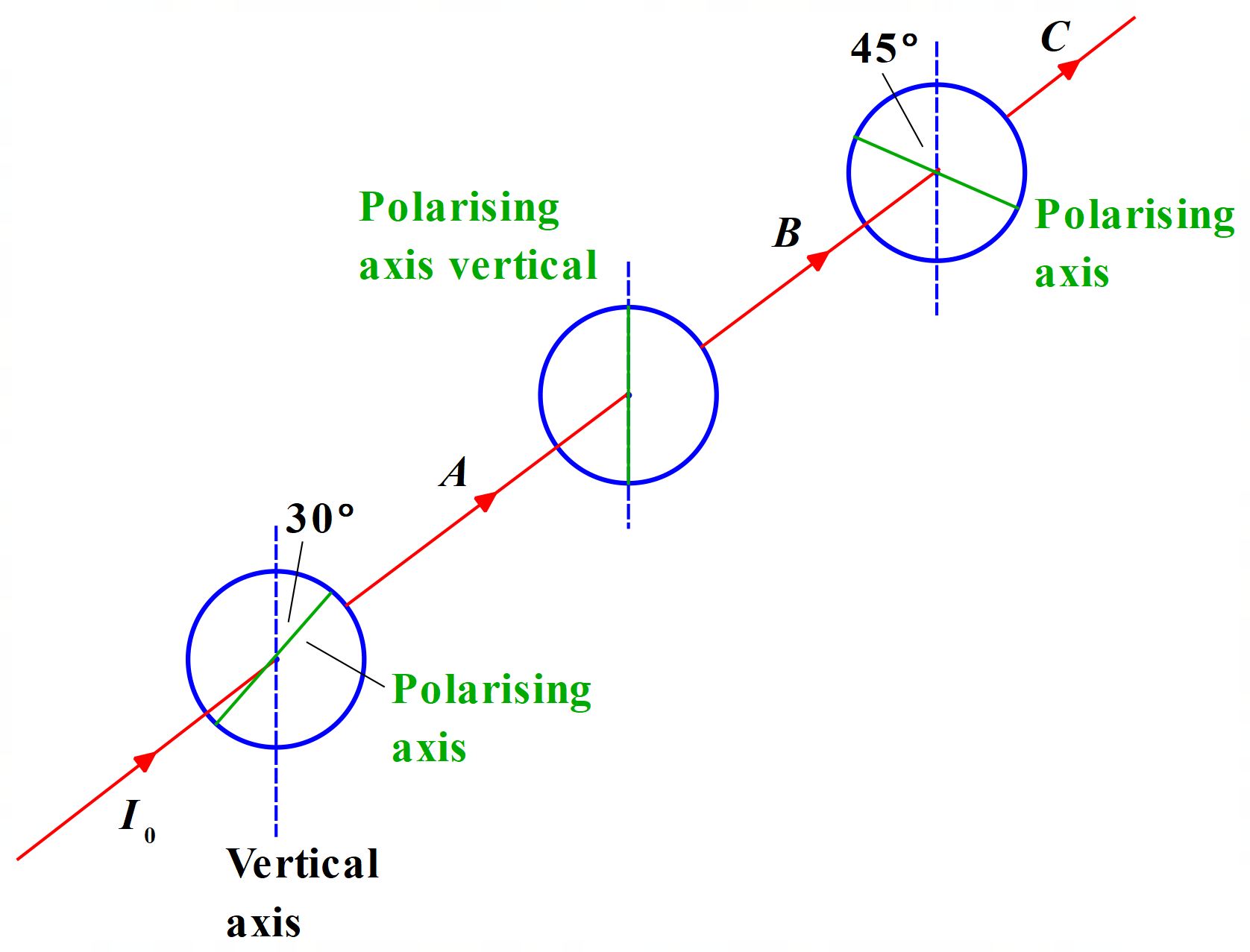
Suppose a second sheet of polaroid material, called an **analyser**, is placed in the path of the already polarised light. If its plane of polarisation is perpendicular to the plane of polarisation of the light that passed through the polariser, almost no light will pass through the analyser.

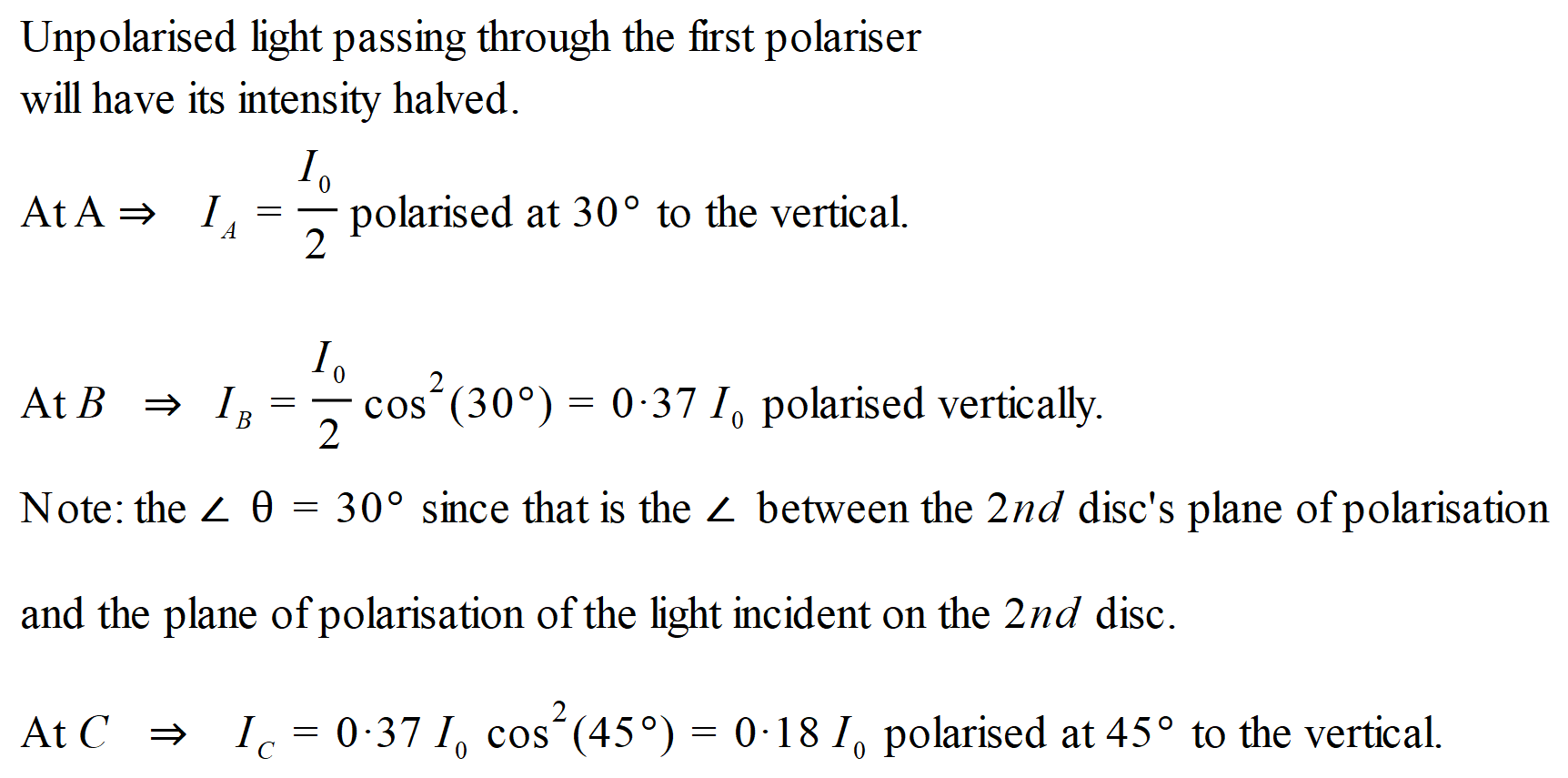
The intensity of light transmitted by an analyser is given by Malus’ Law equation:



Note that for unpolarised light, as mentioned above, all values of **** are equally probable. Therefore, the transmission factor of an ideal polariser for unpolarised light is just the average of the **cos2** function – that is, **½**.

**Example Question** – Unpolarised light of intensity **I0** falls on three polarising discs with polarising axes as shown. Determine the intensity and state of polarisation of the light at points A, B and C.





**Polarisation – Its Significance in Developing A Model for Light**

The fact that light can be polarised strongly supports two ideas about light. Firstly, light consists of waves. Particles would be unaffected by a polariser. Secondly, light waves are transverse waves. Longitudinal waves would not be affected by a polariser. For example, sound waves cannot be polarised.

**The fact that light can be polarised is extremely valuable support to the wave theory of light.** Any theory of light must be able to explain all aspects of light and its interaction with matter. Polarisation along with diffraction and interference are aspects of light’s interaction with matter that can only be explained by a wave model of light. From the turn of the 19th Century, these phenomena were increasingly viewed by scientists as experimental proof that light had a wave nature rather than a particle nature. The fact that light could be polarised was viewed as **especially significant**, since it also identified the type of wave motion. Light had to be a **transverse wave**.

It is interesting to note that Hertz’s experiment that confirmed Maxwell’s prediction of EM waves and provided strong experimental support for the idea that light was a form of transverse EM wave, also demonstrated the effect that became known as the photoelectric effect. Efforts to explain this effect led to further developments in our understanding of the nature of light, way beyond considering light as simply a wave motion. We will consider these developments in the next section of this module.

**LIGHT: QUANTUM MODEL**

**Inquiry Question:** What evidence supports the particle model of light and what are the implications of this evidence for the development of the quantum model of light?

**Blackbody Radiation**

If an iron bar is heated over an extended period, several interesting observations can be made. While still at relatively low temperature, the iron bar radiates heat, but no difference in the colour of the bar is noted. With increasing temperature, the amount of radiation that the bar emits increases very rapidly and visible effects are noted. The iron bar assumes a dull red, and then a bright red colour. As it approaches its melting point the bar becomes a bright yellow colour. If it could be raised to even higher temperatures without melting, the bar would become a blue-white colour. Clearly, with increasing temperature the bar emits more thermal radiation and the frequency of the most intense radiation becomes higher.

The relationship between the colour of light emitted by a hot body and the temperature of the body was first recorded by Thomas Wedgewood in 1792. The porcelain-maker noticed that all of his ovens became red hot at the same temperature, regardless of their shape, size and construction. Experiments by many physicists have since shown that any object with a temperature above absolute zero (zero kelvin or 0 K) emits light of all wavelengths with varying degrees of efficiency (16).

The detailed form of the spectrum of the thermal radiation emitted by most real hot bodies depends to a certain extent upon the composition of the bodies. Physicists found a particularly useful class of hot bodies to assist with the study of the relationship between colour and temperature. This type of body is called a **“blackbody”**, defined as a body whose surfaces absorb all the thermal radiation incident upon them and allow none to be reflected. **All blackbodies at the same temperature emit thermal radiation with the same spectrum, independent of their composition.** **The intensities of the colours in the spectrum depend only on the temperature (17).**

Obviously, blackbodies are hypothetical, ideal bodies. A perfect blackbody does not reflect any light at all. This is the reason why any radiation that it emits is entirely due to its temperature (18). Examples of real bodies which approximate blackbodies include: an object coated with a diffuse layer of black pigment (such as carbon black); an object containing a cavity connected to the outside by a very small hole – such an object is called a **cavity radiator** (17); stars (including our Sun) and planets (19). Note that a blackbody does not necessarily appear black. The Sun does not look black because its temperature is high (around 5800 K) and so it glows brightly. A room-temperature (around 300 K) blackbody, however, would appear very black (19).

The energy density of blackbody radiation **inside** a cavity radiator at various temperatures as a function of wavelength is shown on page 39. Note that the intensity versus wavelength plot for the radiation **emitted** from the hole connecting the cavity to the outside has the same shape. **The radiation inside a cavity whose walls are at temperature T has the same character as the radiation emitted by the surface of a blackbody at temperature T (17).** This provides a useful means of studying blackbody radiation, since cavity radiators are convenient to handle both experimentally and theoretically.

The shape of these blackbody radiation curves was determined experimentally as early as 1899. Physicists used spectrometers to determine the intensity of light emitted at each wavelength. The shape of the curves, however, posed a real problem. **Physicists could not satisfactorily explain the shape of these curves using classical electromagnetic theory.** The wave theory of light predicted, that as the wavelength of emitted radiation becomes shorter, the radiation intensity would increase without limit. Based on the wave model of light, Rayleigh and Jeans derived an equation for black body radiation that suggested that at low wavelengths (high frequencies) the energy density approaches infinity. Historically, the grossly unrealistic prediction at high (ultraviolet) frequencies became known as the **“ultraviolet catastrophe”**.

**Wien’s Approximation and Displacement Law**

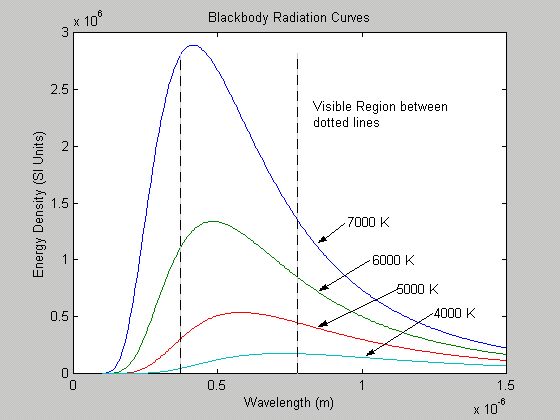
As mentioned above, the intensity versus wavelength plot for the radiation **emitted** by a blackbody has the same shape as the energy density versus wavelength plot shown below. Such a plot indicates that a blackbody of temperature T emits a continuous spectrum with some energy at all wavelengths and that this blackbody spectrum peaks at a wavelength **max**, which becomes shorter with increasing temperature (16).

In 1893, the German physicist Wilhelm Wien used ideas about heat and electromagnetism to develop a quantitative expression, relating wavelength (or colour) of radiation emitted by a hot body to the temperature of the body (19). He achieved this before plots such as that below had been obtained experimentally. The relationship he derived is known as **Wien’s Displacement Law:**



**Wien’s Displacement law indicates that the wavelength of maximum emission of a blackbody is inversely proportional to its temperature in kelvin.** Wien’s Displacement law is particularly useful for determining the surface temperatures of stars (19).

Wien went on studying thermal radiation and in 1896 proposed what became known as **Wien’s approximation** (also sometimes called **Wien's Law** or the **Wien distribution law** – not to be confused with **Wien’s displacement law**) to describe the complete spectrum of thermal radiation. Unfortunately, it failed to accurately fit the experimental data for long wavelengths (low frequency) emission (40 & 41).

****

**Planck’s Hypothesis**

Clearly, a new approach was needed. This came in 1900 when a German physicist, **Max Planck**, suggested a revolutionary idea. **Planck suggested that radiation was emitted or absorbed by a black body in discrete quanta (packets of energy) rather than continuously, as suggested by classical physics.** This daring hypothesis led to the successful explanation of the shape of the energy density curves for black body radiation. With time and the contributions of many physicists, Planck’s hypothesis led to the development of a new model of light and a whole new branch of Physics called Quantum Physics.

# Mathematically, Planck expressed the quantisation of energy emitted from a black body as:

**E = n h f**

Where **E** = the energy of the radiation emitted, **h** = a constant, now called **Planck’s constant = 6.626 x 10-34 Js**, **f** = the frequency of the radiation emitted and **n** = 0, 1, 2, 3, ….., to represent the different multiples of allowed energy coming from the black body at a particular temperature. Conversely, the energy inside a black body cavity is **quantised**, the allowed energy states are called **quantum states** and the integer n is called the **quantum number**.

Although Planck is usually thought of as the father of quantum physics, he believed his solution to the ultraviolet catastrophe was simply a mathematical trick that worked. It was Albert Einstein who really established the quantum model of light with his work on the photoelectric effect in 1905. (23)

Planck’s relation **E = hf**, can be used to determine the energy or frequency (and ⸫ wavelength) of a quantum of light, now called a photon.

(As an aside, the formula that Planck obtained for the energy density in the black body spectrum is:



# Cute, eh? This can be expressed in terms of by using c = . This is how the plots of energy density shown above were drawn.)

**Einstein and The Photoelectric Effect**

As mentioned previously, in 1887 Hertz stumbled across a curious effect of light when conducting his EM wave experiments. Sparks jumping the gap in his receiver were more vigorous when the receiver was exposed to ultraviolet light. Because both light and electricity were involved in this phenomenon, it was called the **photoelectric effect**.

In 1899, J J Thomson verified experimentally that ultraviolet light caused the emission of particles from zinc metal. He subsequently showed that these particles were the same as those in cathode rays (now called electrons). (23)

# In 1902, Philipp Lenard showed that the photoelectric effect is the emission of electrons from the surface of material when the material is illuminated by light of high frequency. In a series of experiments Lenard found that:

* The number of electrons released (the photocurrent) is proportional to the light intensity.
* The emission of photoelectrons was virtually instantaneous (if it occurred).
* Emission was frequency dependent. There is a certain **threshold frequency** below which no photoelectrons were emitted.
* As the intensity of the light increased, the maximum kinetic energy of emitted electrons remained constant. The maximum kinetic energy of emitted electrons was found to depend on the frequency of the light used and the type of surface.

**The last three of these experimental results could not be explained by the classical wave theory of electromagnetism.** Classical theory for instance predicted that electrons in a surface absorbing low intensity radiation of any frequency should accumulate energy for several seconds and then have sufficient energy to be ejected. Electrons absorbing higher intensity radiation should be ejected more quickly. Experimental results showed, however, that emission was almost immediate for all frequencies above the threshold frequency and was independent of intensity.

Inspired by the work of Planck, Albert Einstein proposed the radical idea that light energy is transmitted in discrete packets of energy rather than as a spreading wave. The amount of energy in each packet is a quantum and represents the smallest quantity of light energy of a particular frequency. Einstein called this a light quantum. Eventually, the name **photon** was applied to a quantum of radiant energy. The energy, **E**, of a photon is:

**E = h f**

where **h** = **Planck’s constant = 6.626 x 10-34 Js** & **f** = the frequency of the radiation emitted. **This model of light can be referred to as the particle model or the photon model or the quantum model of light.**

Einstein used his particle model of light to explain the photoelectric effect in the following way: Light striking a surface consists of photons. Each photon carries an energy hf into the surface. When a photon collides with an electron in the surface, it gives up all its energy to this electron, in accordance with the law of conservation of energy. Part of that energy () is used in causing the electron to pass through the metal surface. The rest of the energy (hf - ) is given to the electron as kinetic energy. This is the kinetic energy the electron will have outside the surface if it does not suffer any internal collisions on the way out. In other words, **(hf - ) is the maximum kinetic energy, Kmax, of the photoelectron**.

This explanation is summarised in **Einstein’s photoelectric equation**:

**(hf - ) = Kmax**

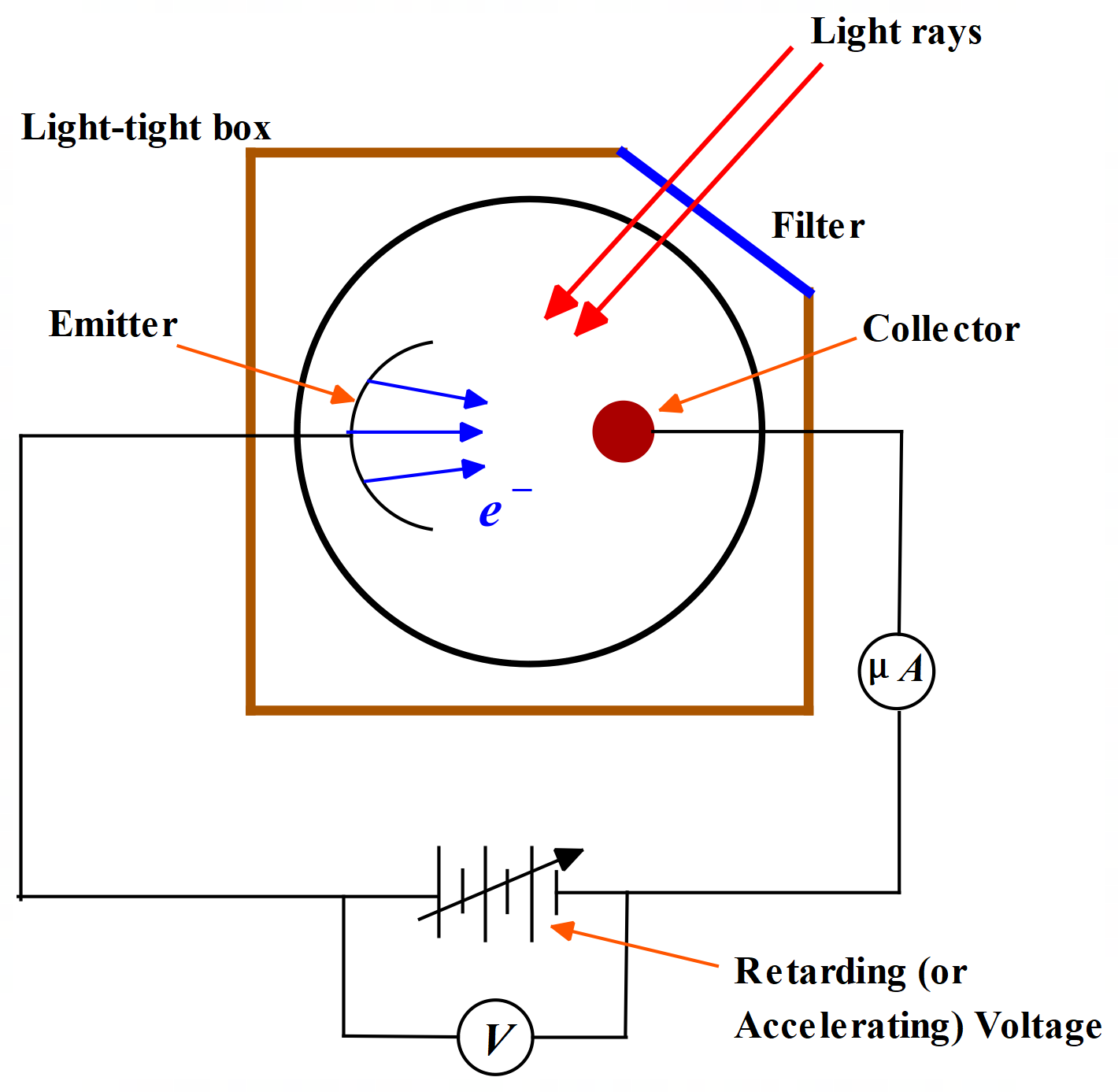
where  is called the **work function** for the surface and is the minimum energy required to remove an electron from the surface. Thus, ** = hf0** where **f0** is the **threshold frequency** below which no photoelectrons are emitted. So, the equation above becomes:

**h (f - f0) = Kmax**

**Robert Millikan**, an American physicist, experimentally verified Einstein’s explanation of the photoelectric effect in 1916. Einstein received the 1921 Nobel Prize in Physics for his work on this effect.

**Photoelectric Effect Practical**

Let us examine how you would investigate the photoelectric effect in various metals. Clearly, the equation **Kmax = (hf - )** is of the form y = mx + b, the equation of a straight line, with gradient h and y-intercept – . So, a plot of **Kmax** versus **f** would be a straight line with those features. But how do we measure **Kmax** experimentally? The circuit design below is one that is commonly used for such an experiment. The method below is taken from reference 21.



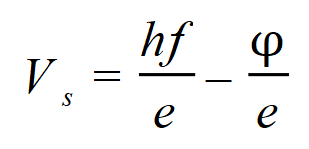
EM radiation (eg light) is passed through a series of filters to cause distinct frequencies of radiation to fall on the emitter of the photocell. The emitter is made of the metal to be tested. The photocell is a vacuum tube to ensure that air molecules do not impede the emitted photoelectrons reaching the collector. The photocell is housed in a light-tight box to exclude unwanted frequencies of radiation.

When the EM radiation above the threshold frequency falls on the emitter, photoelectrons are emitted and make their way to the collector. The photoelectric current is measured by the microammeter. As the collector is made more positive, more photoelectrons are attracted across, increasing the collector current until a saturation current is reached. This is simply the maximum current that can flow. You cannot have more electrons reaching the collector than are liberated at the emitter.

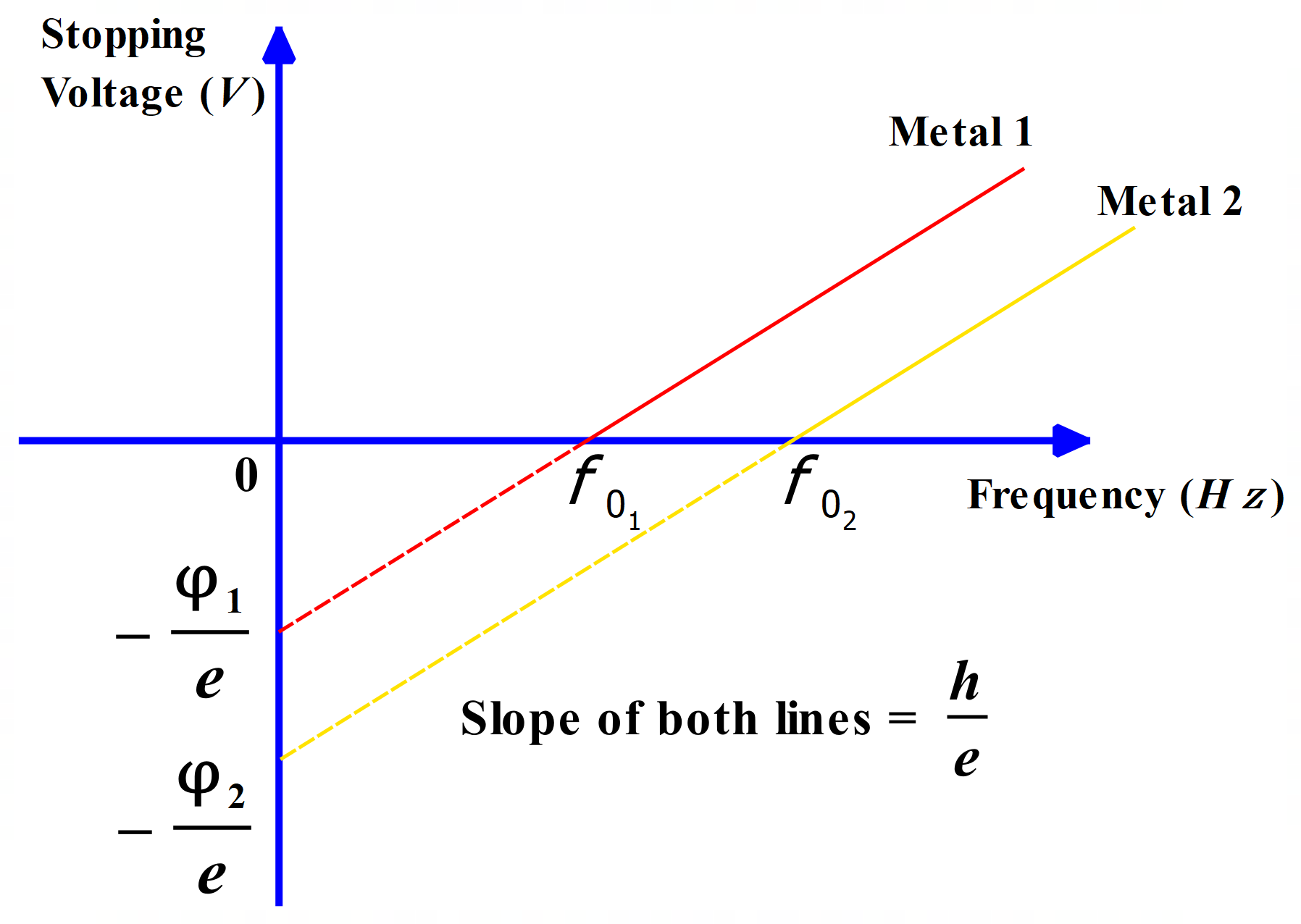
If the collector is now made progressively more negative, the electrons begin to be repelled and eventually there is no collector current. At this point all the photoelectrons, even the most energetic, are repelled back to the emitter. When the collector current just drops to zero, the applied voltage is called the **stopping voltage**. It is a measure of the maximum kinetic energy of the electrons emitted as a result of the photoelectric effect. At that point we have:

**Kmax = (hf - ) = Vs e**, where **Vs** = stopping voltage and **e** = charge on electron.

Clearly, **Vs e** is the energy of the fastest photoelectrons in joules. The above equation can be re-written as:



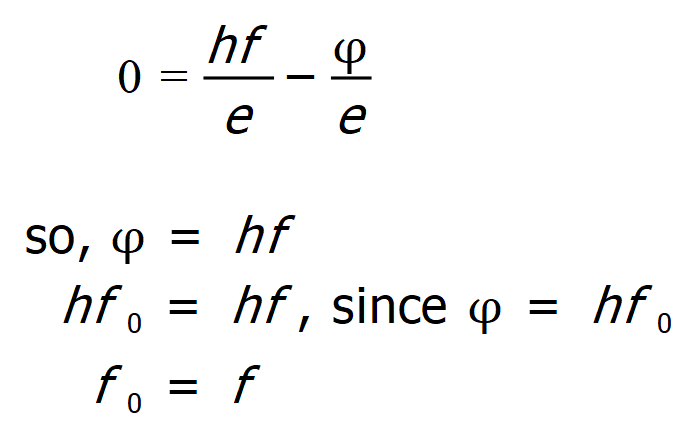
So, a plot of **Vs** versus **f** would be a straight line of gradient **h/e** and y-intercept **-e**. For each metal being studied, the values of stopping voltage versus frequency for the various frequencies tested are plotted on a graph such as that below. The line of best fit is calculated statistically or drawn by eye depending on the accuracy and precision required. This line is then extrapolated to intercept the y-axis (stopping voltage axis).



From the graph, we can then determine Planck’s constant, the threshold or cut-off frequency for the metal and the work function for the metal, within the error limits of the experiment.

Planck’s constant, **h = e x slope of line**

Threshold or cut-off frequency, **f0**, occurs when the stopping voltage is zero (at the x-intercept) since then:



The work function, , is calculated from the y-intercept: ** = - e x y-intercept**.

**Try this PhET** [**photoelectric effect simulation**](https://phet.colorado.edu/en/simulation/photoelectric)**.** It allows you to play around with the variables involved and see the effects produced.

**Comment on the Dual Nature of Light**

Today, we view light as having a dual character. Light behaves like a wave under some circumstances and like a particle, or photon, under others. When we consider the propagation of light or its interaction with other EM radiation, we usually find it convenient to consider light as a wave motion. When we deal with the interaction of light with matter, we usually find it convenient to think of light as a stream of particles.

On some occasions we need to consider the wavelength and frequency properties of light, on others we need to consider the energy or momentum aspects. The energy is related to the frequency, the frequency is related to the speed of light.

**c = f  and E = h f**

**Clearly, these equations can be rearranged as needed.**

**LIGHT AND SPECIAL RELATIVITY**

**Inquiry Question:** How does the behaviour of light affect concepts of time, space and matter?

# Introduction to Relativity

Relativity is the study of the relative motions of objects. Einstein’s **Theory of Relativity** is one of the greatest intellectual achievements of the 20th Century. **Special Relativity**, developed by Einstein in 1905, deals with systems that are moving at constant velocity (no acceleration) with respect to each other. **General Relativity** proposed in 1916 deals with systems that are accelerating with respect to each other. Before commencing our study of Relativity some preliminary definitions are necessary.

**Reference Frames**

A reference frame is defined as a set of axes with respect to which distance measurements can be made. A set of recording clocks is considered to be embedded in the frame to specify time.

An inertial reference frame is defined as one in which Newton’s First Law (his law of inertia) is valid. In other words, an inertial reference frame is one that is not accelerating.

A **non-inertial reference frame** is one that is **accelerating**.

# An Event

A physical event is defined as something that happens independently of the reference frame used to describe it – eg lightning flashes. An event can be characterized in a Cartesian reference frame by stating its coordinates x, y, z and t.

# Brief History of Relativity Before Einstein

The phenomenon of motion has been studied for thousands of years. To the ancient Greek philosopher **Aristotle,** it was obvious that objects would assume a preferred state of rest unless some external force propelled them. He also believed in the concepts of **Absolute Space** and **Absolute Time** – that is that both space and time exist, independently of each other and of other material things (1, 2 & 3). Thus, to Aristotle it was possible to assign absolute values of position and time to events. Aristotle’s work was held in such high regard that it remained basically unchallenged until the end of the sixteenth century, when **Galileo** showed that it was incorrect.

The view that motion must be **relative** – that is, it involves displacements of objects relative to some reference system – had its beginnings with Galileo. Galileo’s experiments and “thought experiments” led him to state what is now called the Principle of Galilean Relativity: the laws of mechanics are the same for a body at rest and a body moving at constant velocity.

Using Galileo’s measurements as a starting point Isaac Newton developed his Laws of Motion and his Law of Universal Gravitation. Newton showed that it is only possible to determine **the relative velocity** of one reference frame with respect to another and not the absolute velocity of either frame. So, as far as mechanics is concerned, no preferred or absolute reference frame exists. The Principle of Newtonian Relativity may be stated as: the laws of mechanics must be the same in all inertial reference frames.

Thus, due to Galileo and Newton, the concept of Absolute Space became redundant since there could be no absolute reference frame with respect to which mechanical measurements could be made. However, Galileo and Newton retained the concept of Absolute Time, or the ability to establish that two events that happened at different locations occurred at the same time (1). In other words, if an observer in one reference frame observed two events at different locations as occurring simultaneously, then all observers in all reference frames would agree that the events were simultaneous.

The Newtonian concept of the structure of space and time remained unchallenged until the development of the electromagnetic theory in the nineteenth century, principally by Michael Faraday and James Clerk Maxwell. As we have seen, Maxwell showed that electromagnetic waves in a vacuum ought to propagate at a speed of c = 3 x 108 m/s, the speed of light (1). To 19th Century physicists this presented a problem. If EM waves were supposed to propagate at this fixed speed c, what was this speed measured relative to? How could you measure it relative to a vacuum? Newton had done away with the idea of an absolute reference frame (2).

Quite apart from the relativity problem, it seemed inconceivable to 19th Century physicists that light and other EM waves, in contrast to all other kinds of waves, could propagate without a medium. It seemed to be a logical step to postulate such a medium, called the **aether (or ether)**, even though it was necessary to assume unusual properties for it, such as zero density and perfect transparency, to account for the fact that it was undetectable. This aether was assumed to fill all space and to be the medium with respect to which EM waves propagate with the speed c. It followed, using Newtonian relativity, that an observer moving through the aether with velocity **u** would measure a velocity for a light beam of **(c + u)** (5). So theoretically, if the aether exists, an observer on earth should be able to measure changes in the velocity of light due to the earth’s motion through the aether. The **Michelson-Morley experiment** attempted to do this.

**The Michelson-Morley Experiment**

In 1887 Albert Michelson and Edward Morley of the USA carried out a very careful experiment at the Case School of Applied Science in Cleveland. The aim of the experiment was to measure the motion of the earth relative to the aether and thereby demonstrate that the aether existed. Their method involved using the interference of light to detect small changes in the speed of light due to the earth’s motion through the aether (5).



The Michelson-Morley interferometer shown above was mounted on a solid stone block floating in a bath of mercury. The earth, together with the apparatus was assumed to be travelling through the aether with a uniform velocity **u** of about 30 km/s. This is equivalent to the earth at rest with the aether streaming past it at a velocity **–u**.

In the experiment, a beam of light from the source **S** is split into two beams by a half-silvered mirror **K** as shown. One half of the beam travels from **K** to **M1** and is then reflected to **K**, while the other half is reflected from **K** to **M2** and then reflected from **M2** back to **K**. At **K** part of the beam from **M1** is reflected to the observer **O** and part of the beam from **M2** is transmitted to **O**.

Although the mirrors **M1** and **M2** are the same distance from **K**, it is virtually impossible to have the distances travelled by each beam exactly equal, since the wavelength of light is so small compared with the dimensions of the apparatus. Thus, the two beams would arrive at **O** slightly out of phase and would produce an interference pattern at **O**.

There is also a difference in the time taken by each beam to traverse the apparatus and arrive at **O**, since one beam travels across the aether stream direction while the other travels parallel and then anti-parallel to the aether stream direction (see the note below). This difference in time taken for each beam to arrive at **O** would also introduce a phase difference and would thus influence the interference pattern.

Now if the apparatus were to be rotated through **90o**, the phase difference due to the path difference of each beam would not change. However, as the direction of the light beams varied with the direction of flow of the aether, their relative velocities would alter and thus the difference in time required for each beam to reach **O** would alter. **This would result in a change in the interference pattern as the apparatus was rotated.**

The Michelson-Morley apparatus could detect a phase change of as little as 1/100 of a fringe. The expected phase change was 4/10 of a fringe. However, no such change was observed.

Thus, the result of the Michelson-Morley experiment was that no motion of the earth relative to the aether was detected. Since the experiment failed in its objective, the result is called a **null** result. The experiment has since been repeated many times and the same null result has always been obtained. (5)

**NOTE:** This time difference mentioned above comes about from classical vector work. After the original beam is split at K the half transmitted to M1 travels with velocity **(c + u)** relative to the “stationary” earth, as it is travelling in the direction of “flow” of the aether. When it is reflected from M1 it travels towards K with a velocity relative to the earth of **(c – u)** against the motion of the aether stream. Thus, the time taken for the total journey of this beam from K to M1 and back again is:



However, the other beam travels with velocity **√ (c2 – u2)** towards M2 and then with the same speed in the opposite direction away from M2 after reflection.



Thus, the time for the total journey of the beam from K to M2 and back again is:



Clearly, t1 and t2 are different.

**The Role of The Michelson-Morley Experiment**

The Michelson-Morley experiment is an excellent example of a critical experiment in science. The fact that no motion of the earth relative to the aether was detected suggested quite strongly that the aether hypothesis was incorrect and that no aether (absolute) reference frame existed for electromagnetic phenomena. This opened the way for a whole new way of thinking that was to be proposed by Albert Einstein in his Theory of Special Relativity.

It is worth noting that the null result of the Michelson-Morley experiment was such a blow to the aether hypothesis in particular and to theoretical physics in general that the experiment was repeated by many scientists over more than 50 years. A null result has always been obtained.

Although many attempts were made to save the aether hypothesis, eventually, physicists like Lorentz (1899), Larmor (1900) and Poincare (1905) showed that the changes needed to make the aether hypothesis consistent with the null result of the Michelson-Morley experiment implied that the aether (absolute) reference frame was impossible. The aether ceased to exist as a plausible substance (4).

**Principle of Relativity**

A relativity principle is a statement of what the invariant quantities are between different reference frames. It says that for such quantities the reference frames are equivalent to one another, no one having an absolute or privileged status relative to the others. So, for example, Newton’s relativity principle tells us that all inertial reference frames are equivalent with respect to the laws of mechanics.

As we have seen, for quite a while in the 19th Century it looked as if there was a preferred or absolute reference frame (the aether) as far as the laws of electromagnetism were concerned. However, in 1904 Henri Poincare proposed his Principle of Relativity: “The laws of physics are the same for a fixed observer as for an observer who has a uniform motion of translation relative to him”. Note that this principle applies to mechanics as well as electromagnetism. Although his principle acknowledged the futility in continued use of the aether as an absolute reference frame, Poincare did not fully grasp the implications. Poincare still accepted the Newtonian concept of absolute time. Einstein abandoned it. (5)

**Einstein’s Theory of Special Relativity**

In 1905, Albert Einstein published his famous paper entitled: **“On the Electrodynamics of Moving Bodies”**, in which he proposed his two postulates of relativity and from these derived his Special Relativity Theory. **Einstein’s postulates are:**

1. **The Principle of Relativity** – All the laws of physics are the same in all inertial reference frames – all inertial reference frames are equivalent.
2. **The Principle of the Constancy of the Speed of Light** – The speed of light in free space has the same value c, in all inertial frames, regardless of the velocity of the observer or the velocity of the source emitting the light.

The significance of the first postulate is that it extends Newtonian Relativity to all the laws of physics not just mechanics. It implies that all motion is relative – no absolute reference frame exists. The significance of the second postulate is that it denies the existence of the aether and asserts that light moves at speed **c** relative to all inertial observers. It also predicts the null result of the Michelson-Morley experiment, as the speed of light along both arms of the interferometer will be **c**.

Perhaps the greatest significance of the second postulate, however, is that it forces us to re-think our understanding of space and time. In Newtonian Relativity, if a pulse of light were sent from one place to another, different observers would agree on the time that the journey took (since time is absolute), but would not always agree on how far the light travelled (since space is not absolute). Since the speed of light is just the distance travelled divided by the time taken, different observers would measure different speeds for light**. In Special Relativity, however, all observers must agree on how fast light travels.** They still do not agree on the distance the light has travelled, so they must therefore now also disagree over the time it has taken. **In other words, Special Relativity put an end to the idea of absolute time** (2)**.**

Clearly, **since c must remain constant, both space and time must be relative quantities**.

**Simultaneity**

Einstein often used **“thought experiments” (**in German: *Gedankenexperiment***)** to assist him in his thinking on relativity. Let us use one of his thought experiments now to illustrate that time is relative. Imagine two observers **O** and **O′**standing at the midpoints of their respective trains (reference frames) **T** and **T′**. **T′** is moving at a constant speed **v** with respect to **T**. Just at the instant when the two observers **O** and **O′** are directly opposite each other, two lightning flashes (events) occur simultaneously in the **T** frame, as shown below. The question is, will these two events appear simultaneous in the **T′** frame?



From our **T** reference frame, it is clear that observer **O′** in the **T′** frame moves to the right during the time the light is travelling to **O′** from **A′** and **B′**. At the instant that **O** receives the light from **A** and **B**, the light from **B′** has already passed **O′**, whereas the light from **A′** has not yet reached **O′**. **O′** will thus observe the light coming from **B′** before receiving the light from **A′**. Since the speed of light along both paths **O′A′** and **O′B′** is **c** (according to the second postulate), **O′** must conclude that the event at **B′** occurred before the event at **A′**. The two events are not simultaneous for **O′**, even though they are for **O**.

We can thus conclude that two events that are simultaneous to one observer are not necessarily simultaneous to a second observer. Moreover, since there is no preferred reference frame, either description is equally valid. It follows that simultaneity is not an absolute concept but depends on the reference frame of the observer. (5)

Length Contraction

When measuring the length of an object it is necessary to be able to determine the exact position of the ends of the object simultaneously. If, however, observers in different reference frames may disagree on the simultaneity of two events, they may also disagree about the length of objects.

In fact, using Special Relativity theory, it is possible to show mathematically and to demonstrate experimentally that **the length of a moving rod appears to contract in the direction of motion relative to a “stationary” observer. This is described by the Lorentz-Fitzgerald Contraction Equation:**



where **l** is the moving length, **l0** is the rest length (or proper length) and **v** is the velocity of the rod relative to the stationary observer. **Note that this contraction takes place in the direction of motion only.** (5) So, for example, an observer on earth watching a rectangular spacecraft move past the earth in the horizontal plane would observe the horizontal length of the craft to be contracted but the vertical width of the craft to remain the same as seen by the observer on the rocket. (Note that this is an over-simplification. Three dimensional objects travelling at relativistic speeds relative to a given reference frame will appear to be distorted in other ways as well, to an observer at rest in that frame. This is outside the scope of this course.)

**Time Dilation**

Let us consider another of Einstein’s thought experiments. Imagine a “light clock”, as shown below. Time is measured by light bouncing between two mirrors. This clock ticks once for one complete up and down motion of the light.



The light clock is placed in a rocket that travels to the right at a constant speed v with respect to a stationary observer on earth. When viewed by an observer travelling with the clock, the light follows the path shown in (a) above. To the stationary observer on earth, who sees the clock moving past at a constant speed, the path appears as in (b) above.

From (a), the time taken for light to make one complete trip up and down, **t0**, is

**t0 = 2.L / c** - (1)

Remember that this represents one tick or one second on the light clock as seen by the observer moving with the clock. From (b), the distance the light moves between **A** and **B** is **c.tAB**, and the distance moved by the whole clock in time **tAB** is **v. tAB**.

So, by Pythagoras’ Theorem:

**(c.tAB )2 = (v.tAB )2 + L2**

and therefore: **tAB 2 = L2 / (c2 – v2)**

which can then be re-arranged (divide throughout RHS by **c2** and take the square root) to give:

**tAB = (L/c) / √ 1 – (v2/c2)**

and thus, the total time taken by the light for one complete up and down motion is:

**tABC = (2L/c) / √ 1 – (v2/c2)**

But from (1) above: **t0 = 2.L / c**

And so, we have:



Clearly, the time interval corresponding to one tick of the light clock is larger for the observer on earth than for the observer on the rocket, since the denominator on the RHS of the above equation is always less than 1.

The above equation may be interpreted as meaning that the time interval **t** for an event to occur, measured by an observer moving with respect to a clock is longer than the time interval **t0** for the same event, measured by an observer at rest with respect to the clock. An alternative way of stating this is that clocks moving relative to an observer are measured by that observer to run more slowly than clocks at rest with respect to that observer. **That is, time in a moving reference frame appears to go slower relative to a “stationary” observer. This result is called time dilation.** The time interval **t0** is referred to as the proper time. The proper time, **t0**, is always the time for an event as measured by the observer in the moving reference frame (Ref 5 pp.63-64).

An example is probably a good idea at this stage. Consider a rocket travelling with a speed of 0.9c relative to the earth. If an observer on the rocket records a time for a particular event as 1 second on his clock, what time interval would be recorded by the earth observer?

From our time dilation equation, we have:

t = 1 / **√ 1 – [(0.9c)2/c2]**

**t**  = 2.29 s

So, to an observer on earth, the time taken for the event is 2.29s. **The earth observer sees that the rocket clock has slowed down.** A second on the rocket is longer than a second on earth. One second on the rocket is equivalent to 2.29 seconds on Earth.

It is essential that you understand that time dilation is not an illusion. It makes no sense to ask ***which of these times is the “real” time***. Since no preferred reference frame exists both times are as real as each other. They are the real times seen for the event by the respective observers.

Remember too, all motion is relative. There is no preferred or absolute reference frame. So, in the example above, to the observer on the rocket, it is he or she who is at rest and the Earth flies past him or her. So, the observer on the rocket will claim that it is the Earth observer’s clock that is running slow. It is only when both clocks are brought together at rest in the same inertial reference frame that it is found that it is indeed the clock that went for the trip in the rocket that has lost time.

Time dilation tells us that a moving clock runs slower than a clock at rest by a factor of

1/**√ 1 – (v2/c2).** This result, however, can be generalised beyond clocks to include all physical, biological and chemical processes. The Theory of Special Relativity predicts that all such processes occurring in a moving frame will slow down relative to a stationary clock.

Note that the factor 1/**√ 1 – (v2/c2)** is called the relativistic (or Lorentz) factor and is often designated by the symbol ****gamma).

The first time you meet the concept of time dilation or length contraction can be a little bit confronting. It does not seem like it could be true. That is only because in our day to day lives, we neither travel fast enough nor observe things that travel fast enough to experience these effects. Have a look at the equations. It is only when the velocity **v** is an appreciable fraction of the speed of light that any significantly noticeable effects occur. For example, in 2016 the Juno space craft clocked 266,000 km/h as it slipped into orbit around Jupiter to become our fastest ever space craft. Even at that speed, the fraction **v/c** is just 0.0002, which makes the denominator of the time dilation equation equal to 1.

You need to sit with time dilation and length contraction for a while and do a few questions. The evidence that both time dilation and length contraction are part of the reality of our universe is overwhelming. We will now spend some time looking at some of the experimental evidence that validates these concepts.

**Observations of Cosmic-Origin Muons at the Earth’s Surface**

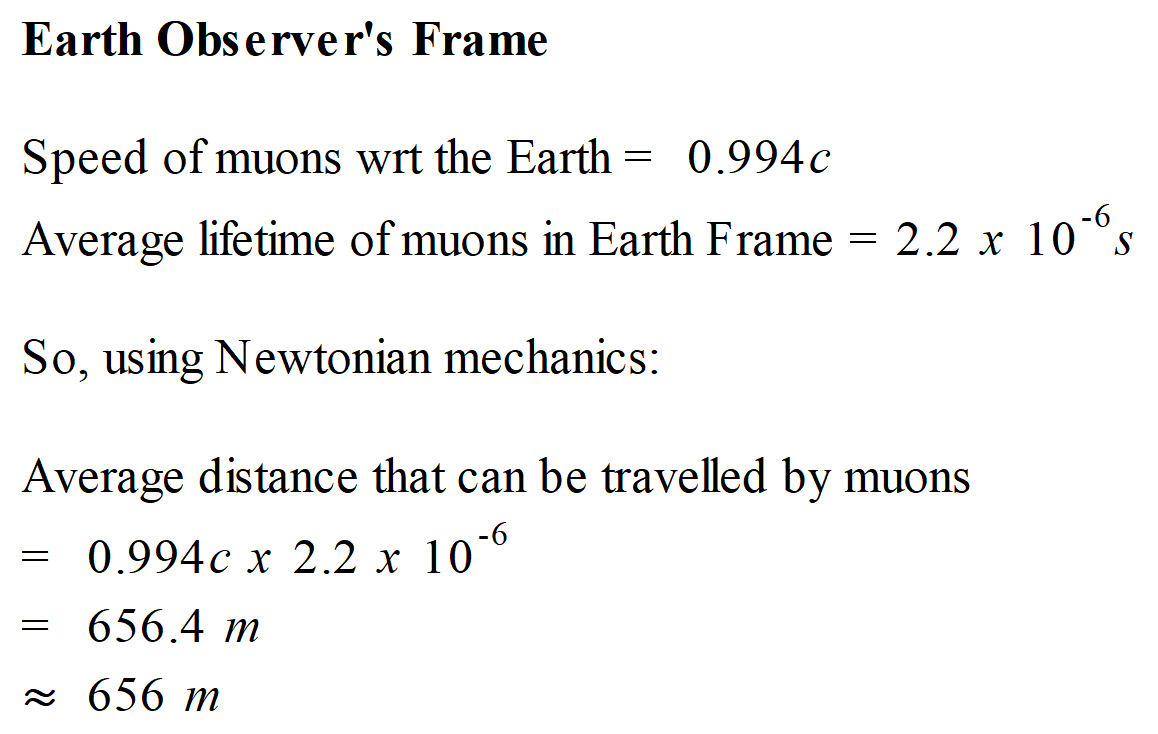
Cosmic rays are high-energy protons and atomic nuclei which move through space at nearly the speed of light and originate outside the solar system. When cosmic rays enter the Earth’s atmosphere, they interact with air molecules in the upper atmosphere creating a cosmic ray shower of particles, including **muons** that reach the Earth’s surface. Muons, **μ-**, commonly have a speed as high as **0.99c** in these cosmic ray showers. At low speeds in the laboratory muons have an average lifetime of about 2.2 x 10-6 s, after which they decay into other particles.

Since the early 1940’s many experiments have been conducted measuring the number of muons at certain heights in the Earth’s atmosphere and the number reaching ground level. In every case it was found that the number of muons reaching the ground significantly exceeded the expected number based on how far they should be able to travel in their lifetimes. These results have been taken as strong evidence for the reality of time dilation and length contraction.

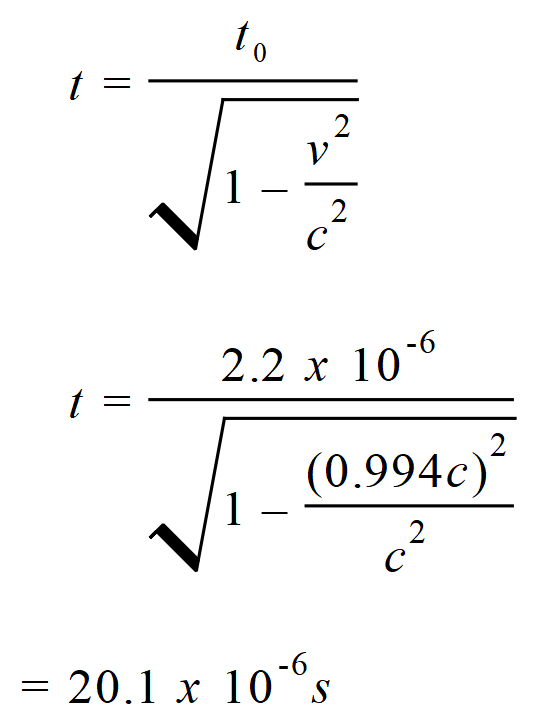
In 1941, Bruno Rossi and David B. Hall conducted experiments at different altitudes in Colorado that demonstrated that the lifetimes of muons travelling near the speed of light appear longer, in qualitative agreement with the prediction of the special relativity.

In 1963, at Mount Washington Observatory, Frisch of MIT and Smith of the University of Illinois performed more precise experiments that were able to confirm this agreement within their margin of error by comparing the rate of detection of muons with that observed near sea level at Cambridge, Massachusetts. Their experiment is a classic and elegant demonstration of the experimental validity of the theory of special relativity. (12) Versions of their experiment continue to be done to this day by physics students who want to see for themselves how special relativity can be verified in the laboratory. There is even a film they made of their experiment. It is available today on [You Tube](https://www.youtube.com/watch?v=rbzt8gDSYIM). Worth a watch.

The analysis that follows uses data from the Frisch & Smith experiment (10) but is common to all these types of experiments. The analysis will be done from both the Earth Observer’s reference frame and the Muon’s reference frame.



The muons are travelling so fast that it is reasonable to suspect that relativistic effects come into play. The earth observer will see the muon’s clock slowed down. The muon’s mean lifetime in the earth observer’s reference frame will be increased. So, applying Special Relativity, we have:

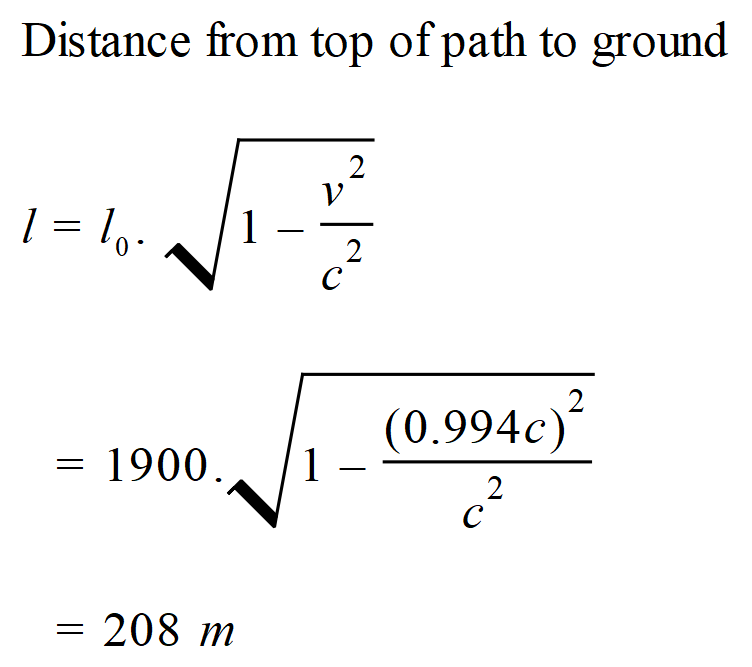


At 0.994c, muons in cosmic rays “live” 9 times longer as seen by the earth observer, than their lifetime as measured at low speed in a laboratory.

⸫ Average distance muon can travel in its extended lifetime = 0.994c x 20.1 x 10-6 = 5994 m, about 9 times as far as expected using Newtonian mechanics.

**Muon’s Reference Frame**

Now, we shall analyse the situation using length contraction, from the perspective of the muon’s reference frame. The muon is travelling fast enough to “see” a contracted distance from the top of its path to the ground. In the Frisch & Smith experiment this was 1900 m – the top of a mountain peak (Mt Washington, New Hampshire, USA) to the base of the mountain.



So, the muons experience a distance of 208 m from the top of their path to the ground, instead of the 1900 m as seen by the earth observer. That is, they experience a distance of about 1/9 that seen by the earth observer.

The time dilation calculation indicates that the average lifetime of the muons is increased in the earth observer’s frame. The length contraction calculation indicates that the distance from top to bottom of their path is shortened in the muon’s frame of reference.

In the various experiments conducted over many years, the muon flux (number) at a particular height is measured and compared to that at ground level. The expected number of muons surviving to reach the ground can be calculated using the mathematics of exponential decay and the known half-life of a muon. The observed number of muons reaching the surface is always more than the expected number. The measured number of muons reaching ground level can only be explained, using time dilation and length contraction analysis similar to that shown above, to accurately describe the mechanics of the muon’s journey.

The results of these experiments provide extremely strong evidence in support of both time dilation and length contraction. This in turn is strong evidence in support of the postulates of special relativity. (8)

**Atomic Clocks – The Hafele-Keating Experiment**

An atomic clock is a device that uses a transition frequency in the EM spectrum of atoms as a frequency standard for its timekeeping element. Atomic clocks are the most accurate time and frequency standards known and are used as primary standards for international time distribution services, to control the wave frequency of television broadcasts, and in global navigation satellite systems such as GPS. Atomic clocks can measure time accurately to many decimal places. That comes in handy when trying to measure the tiny differences in time due to time dilation effects that can be produced at velocities low compared to the speed of light. Today the best cesium clocks are accurate to 16 decimal places.

In 1971, Joseph Hafele and Richard Keating, took four cesium-beam atomic clocks aboard commercial airliners. They flew twice around the world, first eastward, then westward, and compared the clocks against others that remained at the United States Naval Observatory. When reunited, the three sets of clocks were found to disagree with one another, and their differences were consistent with the predictions of special and general relativity. (3)

One set of four atomic clocks remained on the ground and rotated with the Earth’s velocity. One set flew eastward with the Earth’s rotation and thus increased its velocity relative to the clocks on the ground. The final set of clocks flew westward, against the Earth’s rotation and therefore decreased its velocity relative to the clocks on the ground. (3)

For both sets of clocks that went on a world trip, theoretically time dilation would occur. This time dilation had two components. One was due the changed speed of travel (kinematic time dilation). The other was due to the height at which the airliners flew – reducing the acceleration due to gravity (gravitational time dilation). (3)

The mathematical analysis of this situation is slightly beyond the scope of the current course, as it involves binomial expansion of the kinematic time dilation equation and the use of a different time dilation equation for the gravity effect (from General Relativity). There is also the problem of measuring the time differences between a ground clock and a flying clock when neither location is actually an inertial reference frame. Suffice it to say, the maths gets a little more complicated than we need to deal with here. See this [webpage](http://hyperphysics.phy-astr.gsu.edu/hbase/Relativ/airtim.html#c1) if you wish to investigate further.

In the experiment, when the clocks were brought back together at the Naval Observatory, the set of clocks that went eastward (highest velocity) had lost time relative to the ground clocks. The set of clocks that travelled westward had gained time relative to the ground clocks. The amounts of time gain and time loss were as predicted by Special & General Relativity, within acceptable error limits. The experiment had successfully supported the reality of time dilation.

There is a good, brief video explaining this experiment available on [You Tube](https://www.youtube.com/watch?v=flfqLgSV0iA).

I also provide below a table showing the experimental time change results obtained. Note that the time changes are measured in nanoseconds (ns). No wonder we don’t see these effects in our daily lives.

|  |  |  |
| --- | --- | --- |
|  | Eastward Journey | Westward Journey |
| Predicted | -40 +/- 23 ns | + 275 +/- 21 ns |
| Measured | -59 +/- 10 ns | + 273 +/- 7 ns |

\*This table was taken from the excellent [HyperPhysics](http://hyperphysics.phy-astr.gsu.edu/hbase/Relativ/airtim.html#c3) site.

**Evidence From Particle Accelerators**

[A particle accelerator is a machine that uses electromagnetic fields to propel charged particles to very high speeds and energies, and to contain them in well-defined beams.](https://en.wikipedia.org/wiki/Particle_accelerator)

Today, both time dilation and length contraction of particles is routinely confirmed in particle accelerators alongside tests of relativistic energy and momentum, and their consideration is obligatory in the analysis of particle experiments at relativistic velocities.

In 2014, an international team of physicists including Nobel laureate Theodor Hänsch, director of the Max Planck optics institute, conducted one of the most stringent tests of time dilation ever performed. To test the time-dilation effect, physicists need to compare two clocks — one that is stationary and one that moves. To do this, the researchers used the Experimental Storage Ring, where high-speed particles are stored and studied at the GSI Helmholtz Centre for heavy-ion research in Darmstadt, Germany. (13)

The scientists made the moving clock by accelerating lithium ions to one-third the speed of light. Then they measured a set of transitions within the lithium as electrons hopped between various energy levels. The frequency of the transitions served as the ‘ticking’ of the clock. Transitions within lithium ions that were not moving served as the stationary clock.

The researchers measured the time-dilation effect more precisely than in any previous study. (13)

The lifetime of particles produced in particle accelerators appears longer due to time dilation. In such experiments the "clock" is the time taken by processes leading to particle decay, and these processes take place in the moving particle at its own "clock rate", which is much slower than the laboratory clock. This is routinely considered in particle physics, and many dedicated measurements have been performed. For instance, in the muon storage ring at CERN the lifetime of muons circulating with γ = 29.327 was found to be dilated to 64.378 μs, confirming time dilation to an accuracy of 0.9 ± 0.4 parts per thousand. (14)

Various features of collision dynamics in particle accelerators can only be explained using time dilation and /or length contraction. For example, heavy ions that are spherical when at rest should assume the form of "pancakes" or flat disks when travelling nearly at the speed of light. And in fact, the results obtained from particle collisions can only be explained when the increased nucleon density due to length contraction is considered. (15)

**Evidence From Cosmological Studies**

Recall the Doppler Effect for sound which was studied in Module 3. The Doppler Effect is the change in the observed frequency (& wavelength) of a sound wave when there is relative motion between the source and the observer.

Likewise, there is a Doppler Effect for light, sometimes referred to as the **relativistic Doppler Effect**. When there is relative motion between a source of light and an observer, the observer will note a change in the frequency (& wavelength) of the light. If the source approaches the observer, the frequency will increase. If the source recedes from the observer, the frequency will decrease. The relativistic Doppler Effect is different to the Doppler Effect for sound because the equation includes the **time dilation** effect of special relativity and does not involve the medium of propagation as a reference point.

The relativistic Doppler Effect is very important in astronomy. When the emission or absorption spectra of light from distant stars and galaxies are analysed, the results show that the frequencies of spectral lines are shifted compared to the same elements in a laboratory on Earth. This is attributed to the motion of stars and galaxies relative to Earth. A spectral line absorbed by a star moving away from Earth will appear to have a lower frequency and thus a greater wavelength than on Earth. This is referred to as **red shift**, as the spectral line appears to have moved toward the red end of the spectrum by a small amount. If a star approaches Earth, the spectral line moves slightly toward the blue end of the spectrum and is said to be blue shifted. The size of the frequency shift depends only on the relative velocity between the source and observer. Thus, the relative velocity between the source and observer can be determined from measurements of the frequency shift.

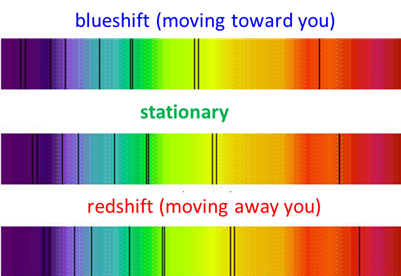


Diagram: Absorption lines in the visible spectrum showing the shift in frequency due to motion of the source. Molecules absorb energy at discrete frequencies which are unique to the atoms that constitute the molecules. The above diagram taken from:

<http://www.physics.usyd.edu.au/teach_res/hsp/sp/mod72/mod72_DopplerLight.htm>

There are two components to the relativistic Doppler Effect. One is a longitudinal effect, which applies when the relative motion of source and observer is toward or away from one another. The other is the **transverse Doppler Effect for light**, which applies when the source moves at right angles to (transverse to) the line joining the source and observer. This is a purely relativistic effect. There is no transverse Doppler Effect for sound. (5)

Experimental verification of the existence of this effect was first performed by Ives & Stilwell in 1938 and the verification has been repeated many times since with increased precision. The transverse Doppler Effect has a simple time dilation interpretation. The moving source of light can be thought of as a moving clock, ticking out electromagnetic oscillations. In our observer reference frame, we measure less oscillations per second for the light coming from the source than would be measured in the reference frame of the source. To us, the moving clock runs slow. We observe a lower frequency (higher wavelength) than the proper frequency. The light is red shifted. (5)

The transverse Doppler Effect is another physical example confirming the relativistic time dilation. (5)

**The Twin Paradox (Extension Topic)**

The Twin Paradox is another example of a thought experiment in relativity. Consider two twins. Twin A takes a trip in a rocket ship at constant speed v relative to the earth to a distant point in space and then returns, again at the constant speed v. Twin B remains on earth the whole time. According to Twin B, the travelling twin will have aged less, since his clock would have been running slowly relative to Twin B’s clock and would therefore have recorded less time than Twin B’s clock. However, since no preferred reference frame exists, Twin A would say that it is he who is at rest and that the earth twin travels away from him and then returns. Hence, Twin A will predict that time will pass more slowly on earth, and hence the earth twin will be the younger one when they are re-united. Since they both cannot be right, we have a paradox.

To resolve the paradox, we need to realise that it arises because we assume that the twins’ situations are symmetrical and interchangeable. On closer examination we find that this assumption is not correct. The results of Special Relativity can only be applied by observers in inertial reference frames. Since the earth is considered an inertial reference frame, the prediction of Twin B should be reliable. Twin A is only in an inertial frame whilst travelling at constant velocity v. During the intervals when the rocket ship accelerates, to speed up or slow down, the reference frame of Twin A is non-inertial. The predictions of the travelling twin based on Special Relativity during these acceleration periods will be incorrect. General Relativity can be used to treat the periods of accelerated motion. When this is done, it is found that the travelling twin is indeed the younger one. (See the section on Atomic Clocks – The Hafele-Keating Experiment on page 56.)

Note that the only way to tell whose clock has been running slowly is to bring both clocks back together, at rest on earth. It is then found that it is the observer who goes on the round trip whose clock has slowed down relative to the clock of the observer who stayed at home. (5)

Watch the YouTube video called [“Complete Solution to the Twins Paradox”](https://www.youtube.com/watch?v=0iJZ_QGMLD0) if you are interested in a more advanced explanation (beyond the scope of the current syllabus).

**Relativity and Space Travel (Extension Topic)**

Time dilation and length contraction have raised considerable interest in regard to space travel. Consider the following thought experiment. Imagine that adventurous Ruby goes on an excursion to Alpha Centauri in a spaceship at 0.9c. Her friend Ava stays behind on earth. Ava knows that –Centauri is 4.3 light years away and so calculates the time for the trip as 4.8 years. Allowing for a brief stopover (shopping etc) when Ruby gets there, Ava expects that Ruby will be back in about 10 years.

Travelling at 0.9c, Ruby measures the distance between earth and –Centauri to be contracted to 1.87 light years and thus calculates the time for the trip as 2.1 years. Thus, she expects to be back on earth in a little over 4 years.

Clearly, these 2.1 years of rocket time must be equivalent to 4.8 years of earth time, since both observers must observe the laws of physics to be the same. (Note: We are ignoring the brief periods of acceleration required by Ruby.) This equivalence can be checked using the time dilation equation.

When Ruby arrives back on Earth, she finds that she has indeed aged a little over four years, whilst poor Ava is nearly 10 years older than when she left. (Perhaps the rare, carnivorous, –Centaurian wolfhound that Ruby has bought for Ava will soothe the upset.)

Seriously, though, the closer v gets to c, the closer the distance to –Centauri and the time required to get there get to zero as seen by Ruby. Obviously, the minimum time for the journey as seen by Ava is 4.3 years. So, if Ruby travels the distance in 1 s, then 1 s of her time is equivalent to 4.3 years of Ava’s (Earth) time. If Ruby travelled for 1 hour at this very high speed, (3600 x 4.3) years or 15480 years would elapse on earth. If Ruby travelled for a whole year on the rocket at this high speed, 135 million years would pass on earth.

While time dilation and length contraction overcome one of the great difficulties of space travel, problems obviously remain in producing such high speeds.

**Relativistic Mass**

Another aspect of the Special Relativity theory is that the mass of a moving object is greater than when it is stationary. In fact, the higher the velocity of the object, the more massive it becomes. This is called **Mass Dilation** and is represented mathematically as:



where m = relativistic mass of particle, m0 = rest mass of particle, v is the velocity of the particle relative to a stationary observer and c = speed of light.

The interesting question is of course, from where does this extra mass come?

Using relativistic mechanics, it can be shown that **the kinetic energy of a moving body contributes to the mass of the body**. (5)

The increase in mass at relativistic velocities has been confirmed many times by extremely precise measurements in particle accelerator experiments.

**The Equivalence of Mass and Energy**

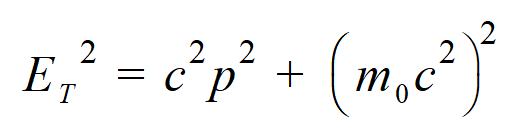
It turns out that mass = energy/c2 or in a more recognizable form:

**E = mc2**

This is Einstein’s most famous equation. Einstein originally derived this equation by using the idea that radiation exerts a pressure on an absorbing body.

This equation states **the equivalence of mass and energy**. Mass is a form of energy. It establishes that energy can be converted into mass and vice versa. For example, when a particle and its antiparticle collide, all the mass is converted into energy. Mass is converted into energy in nuclear fission. When a body gives off energy E in the form of radiation, its mass decreases by an amount equal to **E/c2**. This has been experimentally verified to a very high degree of precision. We will return to these thoughts later.

I need to emphasize that the equation **E = mc2** gives the rest energy, usually called the **internal energy**, of an object. It is the energy possessed by an object by virtue of its mass. If an object is in motion, the total relativistic energy **ET**, is given by the equation:

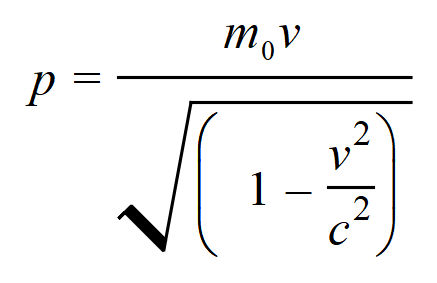


The derivation of this equation is beyond the scope of this course. The equation is not one mentioned in the Syllabus. I include it here for completeness.

In Special Relativity, the Law of Conservation of Energy and the Law of Conservation of Mass have been replaced by the **Law of Conservation of Mass-Energy**. (5)

**Relativistic Momentum**

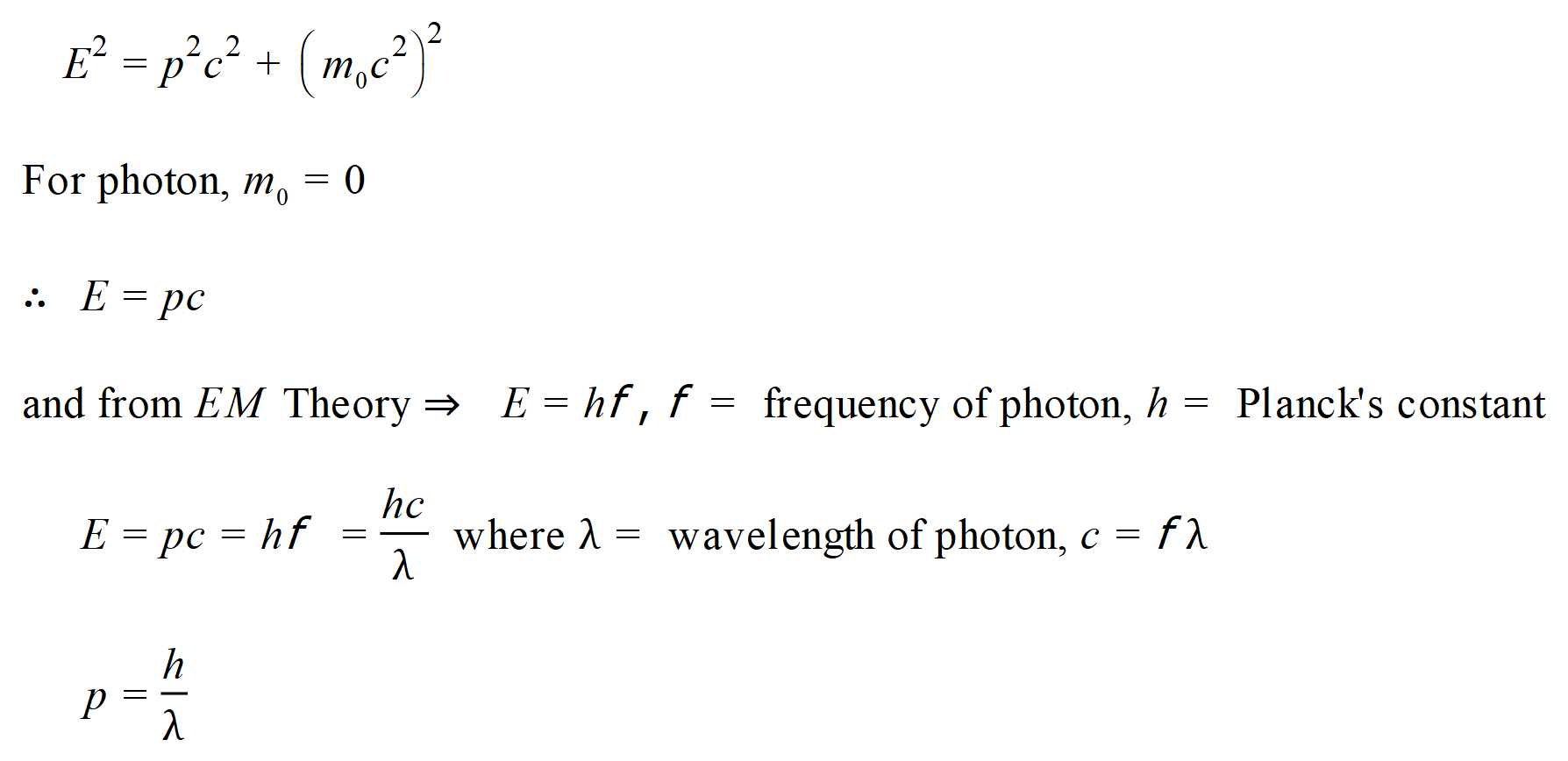
If mass increases with increased velocity, so too must momentum. The relativistic momentum equation is:



where p = relativistic momentum of particle, m0 = rest mass of particle, v is the velocity of the particle relative to a stationary observer and c = speed of light. (5)

Physicists routinely accelerate subatomic particles to velocities near the speed of light, and the momentum of these particles is found to increase precisely in this way. In accelerators allowance must be made for the increase in mass of the particles as their speed increases toward the speed of light. In a synchrotron for instance, magnetic field strength must be increased steadily to counteract the increase in mass and ensure the particles continue to accelerate.

Clearly, for a photon, we run into a difficulty calculating the momentum using the above equation. The photon has zero rest mass and the denominator goes to zero because a photon travels at the speed of light in a vacuum. Yet, we know from the photoelectric effect that a photon does indeed possess a finite momentum. The resolution of this conundrum is to realise that we can use the equation for the total relativistic energy **ET**, to calculate the momentum of a photon.



This equation is an extremely important one. We shall see it again.

Limiting Velocity for Matter

The fact that mass increases as a body gains velocity effectively limits all material objects to travel at speeds lower than the speed of light. The closer a body gets to the speed of light, the more massive it becomes. The more massive it becomes, the more energy that must be used to continue to accelerate the particle to higher velocity. As v 🡪 c, m 🡪 ∞. To accelerate the body up to the speed of light would require an infinite amount of energy. Clearly, this places a limit on the speed that can be attained by a particle. The limiting speed is the speed of light, which no material particle can attain.

**Evidence Confirming or Denying The Postulates of Special Relativity**

At present, there is no peer-reviewed, accepted experimental evidence that refutes the Theory of Special Relativity. It is an extremely well supported and established theory. We are now able to summarise the evidence confirming Einstein’s two postulates.

**First Postulate** – the equivalence of inertial reference frames:

* Observations of cosmic-origin muons at the surface of the Earth
* Evidence of time dilation in the Hafele-Keating experiment with atomic clocks and similar experiments
* Evidence from particle accelerators – verification of time & mass dilation and length contraction
* Transverse Doppler Effect of light – predicted by Special Relativity & observed
* Experiments with radioactive elements showing with great precision that the radiation energy emitted is equivalent to the mass lost by the sample multiplied by the square of the speed of light.

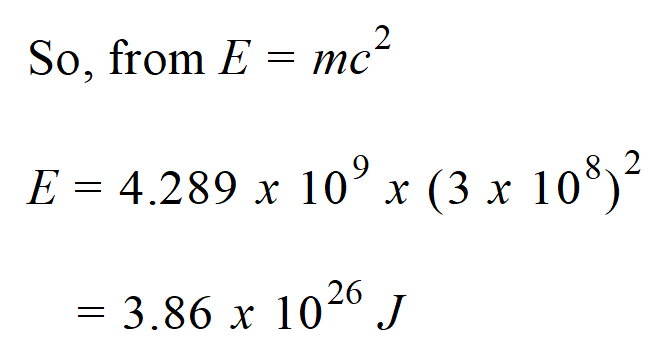
**Second Postulate** – the constancy of the speed of light in a vacuum:

There is no experiment to date that directly confirms the constancy of the speed of light by measuring a one-way journey of light. Most experiments to determine the speed of light depend on light making a round-trip, a two-way journey. (8) However, the Michelson-Morley experiment is usually taken as experimental evidence that the speed of light is constant.

**Examples of Energy Released When Mass Is Converted To Energy**

**Production of Energy by the Sun**

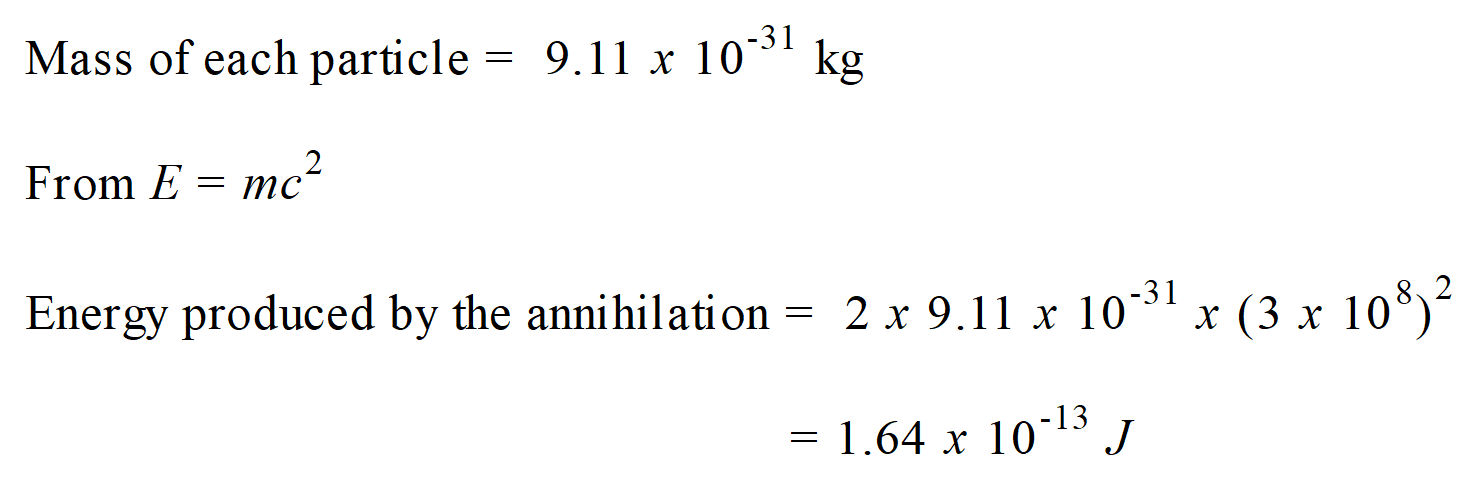
The Sun loses 4.289 x 109 kg per second (11). This mass is converted to energy via the process of nuclear fusion, a process we will study later in this course.



The Sun produces 3.86 x 1026 J of energy every second.

**Particle-Antiparticle Interactions**

The Standard Model of the Atom, which we will study later in this course, tells us that for every matter particle, there exists an antimatter particle equivalent. For instance, for the matter particle we refer to as the electron, the equivalent antimatter particle is the positron. The positron has the same mass and main quantum properties as the electron, but it has a charge of +1. When an electron and positron collide, annihilation occurs and the mass of both particles is converted to energy, which in a low energy collision is carried off by two or more gamma-ray photons.



**Combustion of Conventional Fuel**

Fuels contain chemical energy. When a conventional fuel, such as methane or petrol burns, that energy is released but no mass is converted to energy. It is the energy that is stored in the chemical bonds of the fuel that gets released. The atoms in fuels are simply rearranged from the reactants to the products during combustion. The products may have different properties to the reactants. If you perform a combustion in an isolated system, the total mass of products at the end of the reaction is equal to the total mass of the reactants at the beginning.

**BIBLIOGRAPHY**

1. Hawking, S.W. & Israel, W. (1979). General Relativity – An Einstein Centenary Survey, Cambridge, Cambridge University Press
2. Hawking, S.W. (1988). A Brief History of Time, London, Bantam Press
3. Whitrow, G.J. (1984). The Natural Philosophy of Time, Oxford, Oxford University Press
4. Born, M. (1965). Einstein’s Theory of Relativity, New York, Dover Publications Inc.
5. Resnick, R. (1968). Introduction To Special Relativity, New York, Wiley
6. Bunn, D.J. (1990). Physics for a Modern World, Sydney, The Jacaranda Press
7. Halliday, D. & Resnick, R. (1966). Physics Parts I & II Combined Edition, New York, Wiley
8. Mould, R.A. (2002). Basic Relativity, New York, Springer-Verlag
9. Website: <https://www.aps.org/programs/outreach/history/historicsites/mtwashington.cfm>  
   Used for the Frisch & Smith experiment information
10. L'Annunziata, M.F. (Editor) 2020. Handbook of Radioactivity Analysis: Volume 1: Radiation Physics and Detectors, Academic Press – used in these notes for calculations from Frisch/Smith expt – wrongly attributed (I think) to Rossi & Hall
11. Website: <http://solar-center.stanford.edu/FAQ/Qshrink.html> - mass loss of Sun data
12. Egdall, I.M. (2014). Einstein Relatively Simple: Our Universe Revealed In Everyday Language, Singapore, World Scientific Publishing Co. Pty. Ltd.
13. Website: <https://www.scientificamerican.com/article/einsteins-time-dilation-prediction-verified/>
14. Bailey, J.; et al. (1977). "Measurements of relativistic time dilatation for positive and negative muons in a circular orbit". Nature. **268**: 301. [Bibcode](https://en.wikipedia.org/wiki/Bibcode_(identifier)):[1977Natur.268..301B](https://ui.adsabs.harvard.edu/abs/1977Natur.268..301B). [doi](https://en.wikipedia.org/wiki/Doi_(identifier)):[10.1038/268301a0](https://doi.org/10.1038%2F268301a0).
15. Website: The Physics of RHIC - <https://www.bnl.gov/rhic/physics.asp>
16. Carroll, B.W., & Ostlie, D.A. (1996). An Introduction To Modern Astrophysics, New York, Addison-Wesley Publishing Company Inc.
17. Eisberg, R. & Resnick, R. (1974). Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles, Canada, Wiley
18. Playoust, D.F., & Shanny, G.R. (1991). An Introduction to Stellar Astronomy, Queensland, The Jacaranda Press
19. Kaufmann, W.J. III, & Freedman, R.A. (1999). Universe, (5th Edition), New York, W.H. Freeman & Company
20. Bhathal R., (1993). Astronomy for the Higher School Certificate, Kenthurst, Kangaroo Press Pty Ltd
21. Moyle, D.G., Allan, P.T., Millar, G.L. & Molde, T.A. (1988). Year 12 Senior Physics, Melbourne, Macmillan
22. Gribbin, J., (2015). Einstein’s Masterwork, London, Icon Books Ltd
23. Pais, A., (1982). Subtle is the Lord – The Science and the life of Albert Einstein, New York, Oxford
24. Gribbin, J., (2002). Science A History 1543-2001, London, Penguin
25. Vella, G., (1989). Wave Properties of Light, Milton (Queensland), Brooks Waterloo
26. Webpage: [Openstax – The Propagation of Light page](https://cnx.org/contents/5I39byUz@3.3:qL5s9hwt@7/1-1-The-Propagation-of-Light#CNX_UPhysics_34_01_FizeauMeth)
27. Website: Foucault and the Rotating Mirror Experiments, <http://www.setterfield.org/cx4.html>
28. Website: 2019 redefinition of the SI base units, <https://en.wikipedia.org/wiki/2019_redefinition_of_the_SI_base_units>
29. Website: Light I: Particle or Wave?, <https://www.visionlearning.com/en/library/Physics/24/Light-I/132>
30. Website: From Huygens’ Waves To Einstein’s Photons: Weird Light,  
    <https://www.sciencedirect.com/science/article/pii/S1631070517301032>
31. Website: Lumen – Boundless Physics, <https://courses.lumenlearning.com/boundless-physics/chapter/diffraction/>
32. Giancoli, D.C., (2008). Physics for Scientists & Engineers – Volume II, USA, Pearson Education Inc.
33. Cooper, M., Higher School Physics, Sydney, William Brooks & Co. Pty. Ltd.
34. Website: Light: Particle or Wave, <https://www.olympus-lifescience.com/en/microscope-resource/primer/lightandcolor/particleorwave/>
35. Website: Light Observations, <http://galileo.phys.virginia.edu/classes/609.ral5q.fall04/LecturePDF/L19-LIGHT.pdf>
36. Website: Newton’s Particle Theory of Light, <http://galileo.phys.virginia.edu/classes/609.ral5q.fall04/LecturePDF/L20-LIGHTII.pdf>
37. Website: Newton’s Corpuscular Theory, <https://www.youtube.com/watch?v=uO2uyvf-E3k>
38. Website: Difference Between the Corpuscular and Wave Theory of Light, <https://www.youtube.com/watch?v=FipMEWh-Hns>
39. Website: Newton’s Particle Theory of Light, <https://opik.fyysika.ee/index.php/book/section/8642#/section/8642>
40. Website: Wien Approximation – Wikipedia, <https://en.wikipedia.org/wiki/Wien_approximation>
41. Giancoli, D.C., (2008). Physics for Scientists & Engineers – Volume III, USA, Pearson Education Inc.

**APPENDIX A**

**Statement of Syllabus Content Covered in these Notes**

The following indicates the specific content from the **Stage 6 Physics Syllabus** that has been covered in the notes, worksheets & practicals provided on The Nature of Light Module web page.

The resources on this website are meant to supplement the work you do in class NOT replace it. The notes will always provide you with a comprehensive and accurate set of notes on the Module under study. The worksheets will provide some introduction & practice to appropriate problem-solving skills for the topic. You will need to do much more problem-solving practice than just what is provided on this website. The practicals section will provide some experiments relevant to the topic but again you will need to do more than just what is suggested here. Your teacher should provide you with much more problem-solving & practical experience than you will find on this website.

The content statements that are **ticked** have been covered. Those left without a tick have either not been covered at all or have been only partially covered. These are mainly content statements requiring practical work of some kind.

**Content**

*Electromagnetic Spectrum*

**Inquiry question:** What is light?

Students:

* investigate Maxwell’s contribution to the classical theory of electromagnetism ✓, including:
  + unification of electricity and magnetism ✓
  + prediction of electromagnetic waves ✓
  + prediction of velocity (ACSPH113) Critical and creative thinking icon  Information and communication technology capability icon ✓
* describe the production and propagation of electromagnetic waves ✓ and relate these processes qualitatively to the predictions made by Maxwell’s electromagnetic theory (ACSPH112, ACSPH113) ✓
* conduct investigations of historical and contemporary methods used to determine the speed of light and its current relationship to the measurement of time and distance (ACSPH082) Critical and creative thinking icon  Information and communication technology capability icon ✓
* conduct an investigation to examine a variety of spectra produced by discharge tubes, reflected sunlight or incandescent filaments ✓
* investigate how spectroscopy can be used to provide information about:  Information and communication technology capability icon
  + the identification of elements ✓
* investigate how the spectra of stars can provide information on: Critical and creative thinking icon  Information and communication technology capability icon
  + surface temperature ✓
  + rotational and translational velocity ✓
  + density ✓
  + chemical composition ✓

#### Light: Wave Model

**Inquiry question:** What evidence supports the classical wave model of light and what predictions can be made using this model?

Students:

* conduct investigations to analyse qualitatively the diffraction of light (ACSPH048, ACSPH076)  Information and communication technology capability icon ✓
* conduct investigations to analyse quantitatively the interference of light using double slit apparatus and diffraction gratings (ACSPH116, ACSPH117, ACSPH140)  Information and communication technology capability icon Numeracy icon ✓
* analyse the experimental evidence that supported the models of light that were proposed by Newton and Huygens (ACSPH050, ACSPH118, ACSPH123) Critical and creative thinking icon ✓
* conduct investigations quantitatively using the relationship of Malus’ Law for plane polarisation of light, to evaluate the significance of polarisation in developing a model for light (ACSPH050, ACSPH076, ACSPH120)  Information and communication technology capability icon Numeracy icon ✓

#### Light: Quantum Model

**Inquiry question**: What evidence supports the particle model of light and what are the implications of this evidence for the development of the quantum model of light?

Students:

* analyse the experimental evidence gathered about black body radiation, including Wien’s Law related to Planck's contribution to a changed model of light (ACSPH137) Critical and creative thinking icon  Information and communication technology capability icon Numeracy icon ✓
  + ✓
* investigate the evidence from photoelectric effect investigations that demonstrated inconsistency with the wave model for light (ACSPH087, ACSPH123, ACSPH137) Critical and creative thinking icon  Information and communication technology capability icon ✓
* analyse the photoelectric effect as it occurs in metallic elements by applying the law of conservation of energy and the photon model of light, (ACSPH119)  Information and communication technology capability icon Numeracy icon ✓

#### Light and Special Relativity

**Inquiry question:** How does the behaviour of light affect concepts of time, space and matter?

Students:

* analyse and evaluate the evidence confirming or denying Einstein’s two postulates: ✓
  + the speed of light in a vacuum is an absolute constant ✓
  + all inertial frames of reference are equivalent (ACSPH131) ✓
* investigate the evidence, from Einstein’s thought experiments and subsequent experimental validation, for time dilation ✓and length contraction ✓, and analyse quantitatively situations in which these are observed, for example:
  + observations of cosmic-origin muons at the Earth’s surface  Information and communication technology capability icon Numeracy icon ✓
  + atomic clocks (Hafele–Keating experiment) Critical and creative thinking icon  Information and communication technology capability icon Numeracy icon ✓
  + evidence from particle accelerators Critical and creative thinking icon  Information and communication technology capability icon Numeracy icon ✓
  + evidence from cosmological studies  Information and communication technology capability icon ✓
* describe the consequences and applications of relativistic momentum with reference to:
  +  Information and communication technology capability icon Numeracy icon ✓
  + the limitation on the maximum velocity of a particle imposed by special relativity (ACSPH133) Critical and creative thinking icon ✓
* Use Einstein’s mass–energy equivalence relationship to calculate the energy released by processes in which mass is converted to energy, for example: (ACSPH134)  Information and communication technology capability icon Numeracy icon
  + production of energy by the sun ✓
  + particle–antiparticle interactions, eg positron–electron annihilation ✓
  + combustion of conventional fuel ✓