Waves Notes[1]

Produced for Eduquas content

15 January 2014

[1] Including standing waves, definitions and all sorts of other wavey whatnot

Please forward errors and corrections to `joerowing@exeterms.ac.uk`

Contents

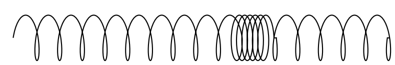
1 Preliminary ideas

1.1 Kinds of waves

Waves exist as transverse and longitudinal waves[2]: The *transverse wave* is the one we are so very familiar with from drawings in textbooks and from our experience of water waves. It is by far the easiest to draw and as such I used transverse waves to illustrate my descriptions of phase above. In a transverse wave the particles all oscillate perpendicularly to the direction in which the energy is moving.

[2] Specification statement: distinguish between transverse and longitudinal waves

Longitudinal waves are somewhat harder to draw, and are less commonly represented because they consist of oscillations in the same axis as the direction of energy transfer



A longitudinal wave is a series of compressions and rarefactions, such as we see on the picture of a slinky above

. Note that I have been careful with my terminology in the above text to try and make it clear that the wave transfers only energy, not matter, from one place to another.

To understand the work in this section you will need to know some definitions from previous work.[3] These are given here:

[3] Specification statement: explain the terms displacement, amplitude, wavelength, frequency, period and velocity of a wave

Frequency [4]This is the number of waves passing a point per second (or the number of vibrations per second). Symbol: f; unit: Hz (Hz).

[4] The frequency is related to the period Frequency Hz $= \frac{1}{\text{Periods}^{-1}}$

Wavelength The distance from any point on a wave to the corresponding point on the next wave. Symbol: λ; unit: m.

Displacement The distance of a point on a wave from its undisturbed position. It can be positive or negative. Symbol: s; unit: m.

Amplitude The maximum displacement from the undisturbed position. Symbol: a; unit: m.

Phase This measures how far out of step two waves or the oscillations of two points on an individual wave are. It is measured in cycles (i.e. number of oscillations) or in degrees. In the diagram waves A and B are out of phase by one quarter cycle or 90°.

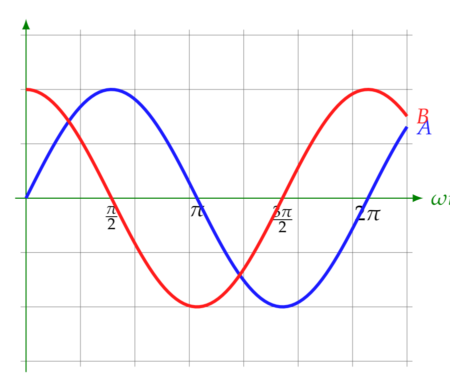


Figure 1: Two waves, 90° out of phase

Wave speed Wave speed[5] is related to frequency and wavelength by the formula:

$$\text{Velocity} = \text{frequency} \times \text{wavelength}$$

$$c = f\lambda$$

[5] Specification statement: recall and use the equation $c = f\lambda$

Phase Transverse waves are able to oscillate in any plane perpendicular to the direction of their motion. We can restrict a wave to one plane of oscillation, either through filtering out all others, or careful conditions when we create the wave. A wave oscillating in one plane is said to be polarised. In the image below we are filtering out all of the components of the incident light that are not polarised in the 'y' direction

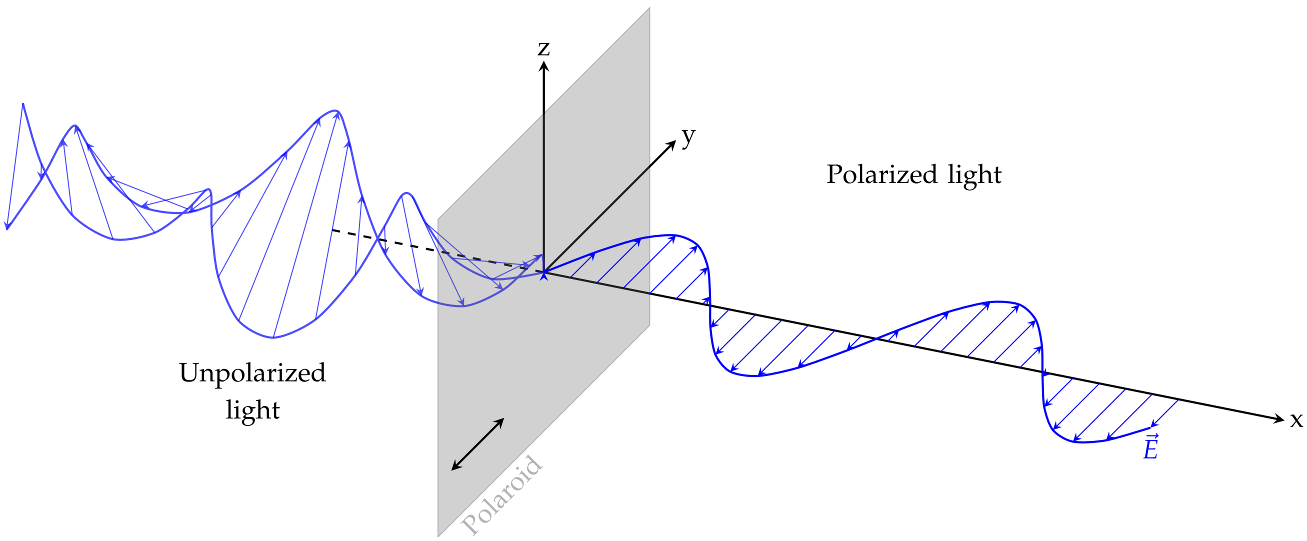


Figure 2: Light passes through a polarizing filter device. Only one incident wavelength is shown (monochromatic light). The magnetic field, perpendicular to the electric one, is not drawn.

2 Superposition

[6] When two or more waves arrive at the same place at the same time they combine with each other according to the principle of superposition

. This states that the resultant displacement at any point is equal to the sum of the displacements of the individual waves. Note that a displacement may be negative so addition of two displacements may result in a resultant, which is smaller than either of the original displacements.// When two waves of equal amplitude and wavelength meet in phase they interfere constructively to produce a new wave of twice the amplitude but the same wavelength[7] .

[6] Specification statement: state, explain and use the principle of superposition

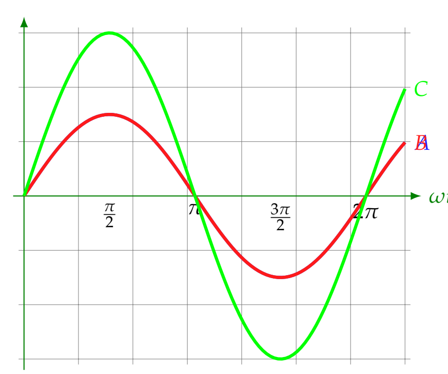


Figure 3: Constructive interference

[7] Specification statement: state the conditions necessary for two-source interference to be observed, i.e. constant phase difference, vibrations in the same line

When two waves of equal amplitude and wavelength meet in antiphase they interfere destructively to cancel each other out

.

When two waves of equal amplitude and wavelength meet with any other phase difference the resultant will have an intermediate amplitude.

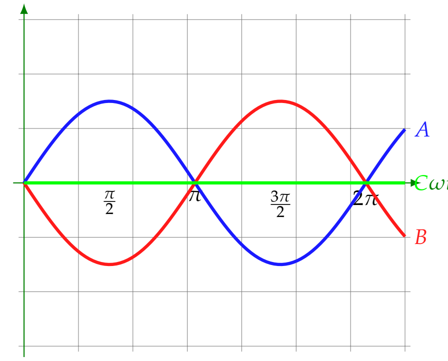


Figure 4: destructive interference

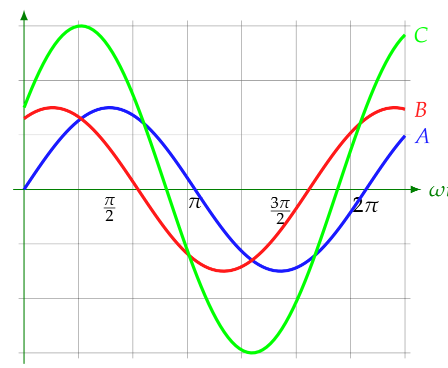


Figure 5: Two waves, out of phase

3 Coherence

Coherence[8] is an essential condition for the interference of waves. Two sources of waves are described as coherent if they emit waves with a constant phase difference (note that the waves do not necessarily have to be in phase). Two waves arriving at a point are said to be coherent if there is a constant phase difference between them as they pass that point.

[8] Specification statement: give examples of coherent and incoherent sources,

4 Interference

Interference cannot be observed with light - or any form of electromagnetic radiation - from two separate light sources. This is because the phase difference between light waves from the two sources changes randomly so the points of cancellation and reinforcement move about at random. The two sets of waves need to be produced from a single source, either:

- by dividing the wavefront and arranging for the divided wavefronts to overlap, as in the double slit experiment (see below) or
- dividing the amplitude of the waves using a partial reflector and arranging for the separated waves to overlap, as in thin film interference.

4.1 Path difference and Interference

One way of demonstrating interference effects is to take a single source of waves (s) and allow the waves to travel by two different paths to a detector (D).

This can be done as shown below by having one path directly from S to D and another path via a reflector (R). The path difference[9] is the difference between the distances travelled, SD and SRD.

[9] Specification statement: show an understanding of path difference, phase difference, and coherence

As R is moved perpendicular to the direction $\vec{SD}$ a point will be reached where the direct wave and the reflected wave arrive at D in phase. At this point constructive interference occurs at D and a maximum amplitude is detected. (Note that this occurs when the

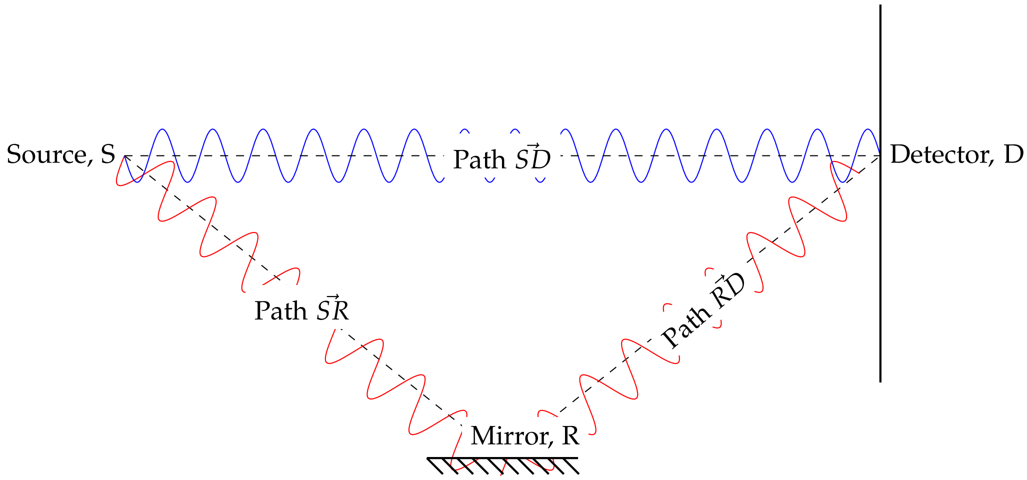


Figure 6: Illustrating path difference.

path difference has a value like $\frac{1}{2}$ or $\frac{3}{2}$ wavelengths as a phase change of $\frac{1}{2}$ cycle occurs on reflection.)

If R is moved away from the line $\vec{SD}$ the path difference increases and a point is reached where the direct and reflected waves arrive in antiphase. A minimum amplitude is now detected. The amplitude will not in general be zero as the amplitude of the reflected wave will be less than that for the direct wave as it has travelled further so complete cancellation will not occur.

Further movement of R away from $\vec{SD}$ will restore the maximum amplitude as the two waves will again be in phase. The increase in the length $S\vec{R}D$ that occurs in moving from one maximum to the next is one wavelength and this provides a method for measuring wavelength.

5 *Single slit diffraction*

In general it is considered that light only travels in straight lines - evidence for example is the production of shadows. When water waves pass through a gap or past an object they spread into the "shadow" region as shown below (A). This is known as "diffraction". The width of the opening controls the degree of diffraction. A narrow opening causes more diffraction (B).

5.1 *Diffraction through a single narrow slit:*

If the slit is more than a few millimetres wide, a sharp rectangular shadow is formed (C). If the slit is narrowed (D) the pattern begins to change - a wide central band is observed with a series of bright fringes either side. The central band is twice the width of the outer fringes. As the slit approaches its narrowest, the pattern broadens and becomes less bright (due to cutting down of light able to pass through slit).

If white light is used the central band is white, however, the fringes are coloured. If monochromatic (single wavelength/colour) light is used the central band and fringes are the same colour. Red light has a more spread out pattern than blue - it also has a larger wavelength than blue - diffraction increases with wavelength.

Intensity distribution of pattern[10]: The diagram below (fig: 7) shows a typical diffraction pattern obtained when monochromatic light passes through a single narrow slit. Also plotted is the intensity of the pattern.

[10] Specification statement: recall the shape of the intensity pattern from a single slit and its effect on double-slit and diffraction grating patterns

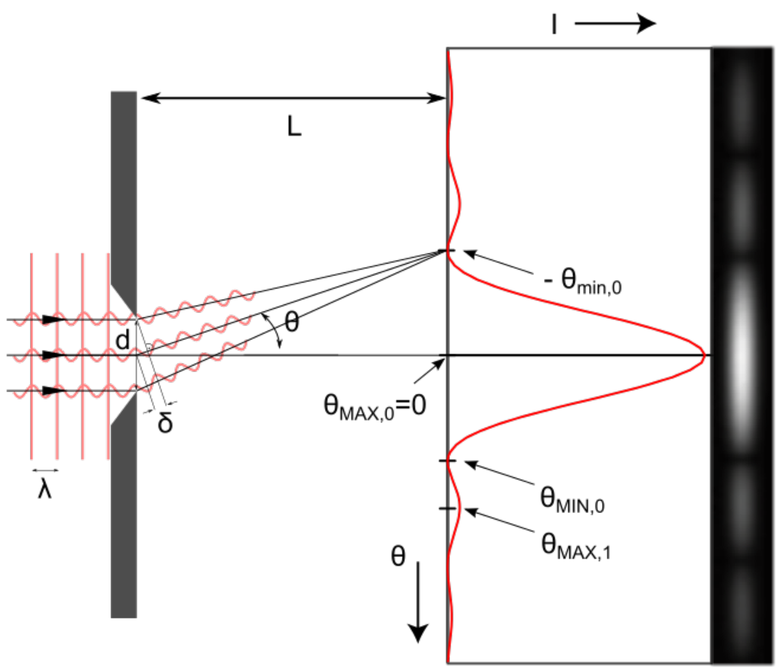


Figure 7: Single slit diffraction pattern and geometry.

The central peak has the greatest intensity, with decreasing intensity with each fringe either side. Each band/fringe has its greatest intensity in its centre. This is the point where light is most likely to fall in its range. The fringes are caused by the edges of the slit acting like two wave sources, the waves from the two wave "sources" interfere superimposing an interference pattern over the diffraction pattern.

6 Youngs experiment

The modern double-slit experiment is a demonstration that light and matter can display characteristics of both classically defined waves and particles; moreover, it displays the fundamentally probabilistic nature of quantum mechanical phenomena. A simpler form of the double-slit experiment was performed originally by Thomas Young in 1801. He believed it demonstrated that the wave theory of light was correct and his experiment is sometimes referred to as Young's experiment or Young's slits. The experiment belongs to a general class of "double path" experiments, in which a wave is split into two separate waves that later combine into a single wave. Changes in the path lengths of both waves result in a phase shift, creating an interference pattern.

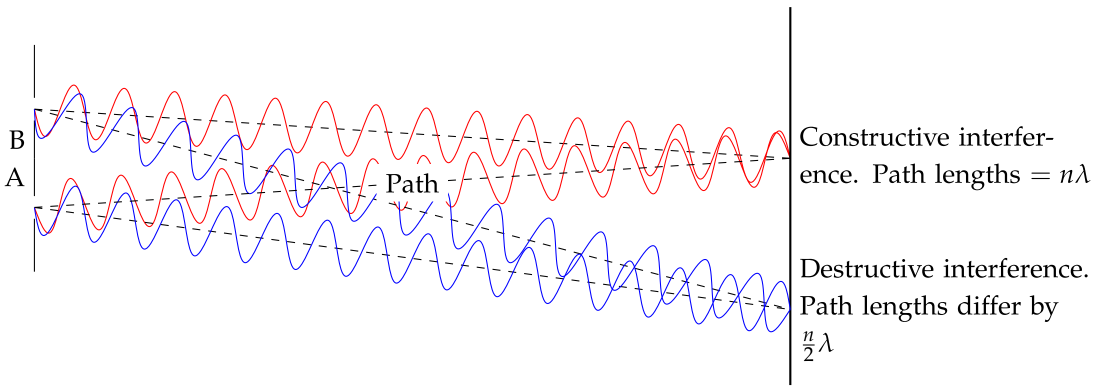


Figure 8: Constructive and destructive interference as a result of path differences.

The position of the bright and dark fringes is determined by the geometry of the experimental set-up.The distance from center to bright spot n is:

$$y = \frac{n\lambda D}{d}$$

We know that the maxima will be at angles such that the path difference will be whole number of wavelengths. The geometric derivation is shown below.[11]

[11] Specification statement: describe an experimental demonstration of two-source interference for light, appreciating the historical importance of Young's experiment, and be familiar with experiments which demonstrate two source interference for water waves, sound waves and microwaves;

Note that this makes use of the small angle approximation, a useful simplification of the basic trigonometric functions which is approximately true in the limit where the angle approaches zero.

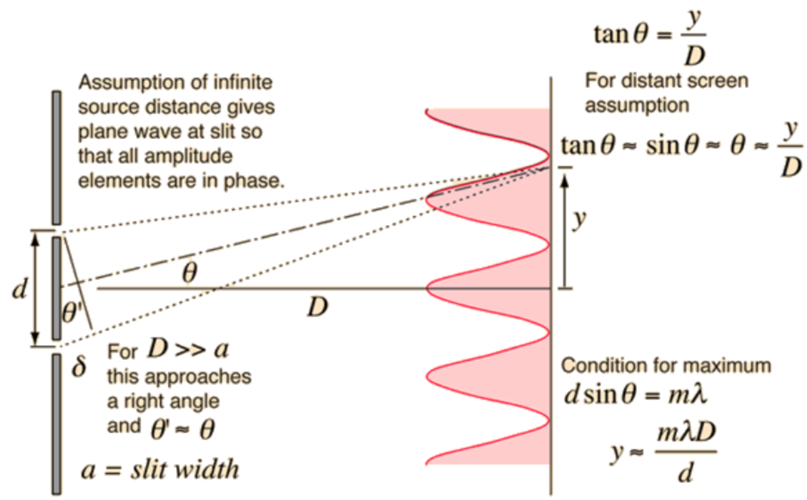


Figure 9: Beam pattern and geometric workings for Young's slits

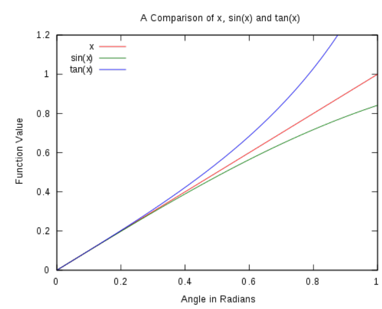


Figure 10: Illustrating the small angle approximation

They are truncations of the Taylor series for the basic trigonometric functions to a second-order approximation.

7 Diffraction grating

A diffraction grating is usually a piece of glass or plastic with closely-spaced lines on it. A transmission grating has thin clear spaces/slits where light can pass through, while a reflection gratings have shiny surfaces between the lines that reflect light. The tracks on a compact disc are very close together - the shiny gaps between the lines act as a reflection grating, leading to a colourful light spectrum being formed when light reflects off the surface. Transmission gratings are used to measure wavelengths of light to a very high degree of accuracy (spectrometer) - all substances have characteristic absorption/emission spectra.

7.1 Explaining diffraction gratings:

The individual lines act like wave sources or thin slits - they produce diffracting waves which will then interfere. Destructive interference leads to cancellation in nearly all directions - there are a few very clearly defined directions where constructive interference occurs, leading to bright fringes. The direction of these fringes depends on the wavelength of light - the wavelength can thus be determined!

Diffraction grating formula Parallel rays of monochromatic light of wavelength "λ" fall on a diffraction grating with slit separation/grating spacing "d".

If the grating has N lines per metre, the grating spacing is given by;

$$d = \frac{1}{N}$$

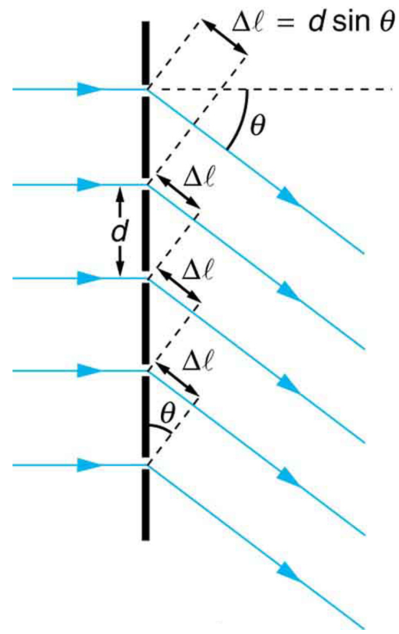
[12]

For constructive interference light from adjacent slits must be in phase (with light from all slits/lines).

$\Delta l = dsin\theta$, where θ is the angle of diffraction

Hence

$$dsin\theta = n\lambda$$



[12] Specification statement: use the equation $dsin\theta = n\lambda$ for a diffraction grating

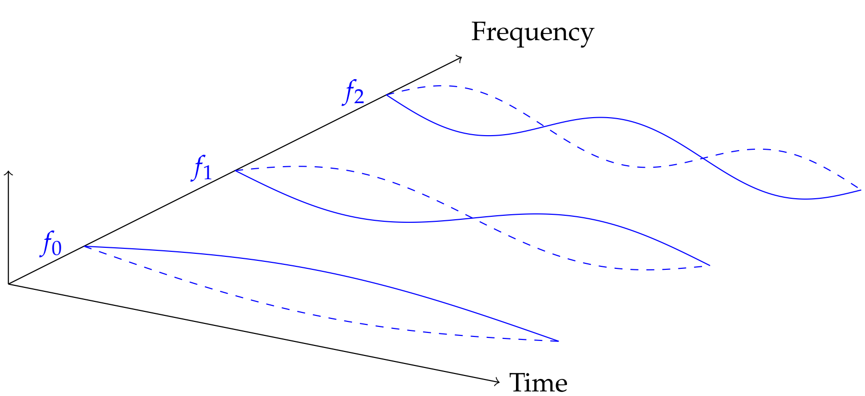
8 Standing Waves

Standing or stationary waves[13] are a superposition effect, which occurs when two identical waves (implying waves with a similar amplitude and frequency) travelling in opposite directions interfere. The most common example of this is where incident and reflected waves interfere, for example on a spring or in a pipe in a musical instrument.

With a string, spring or pipe of fixed length there will be certain frequencies where a standing wave pattern is produced as shown below. These frequencies will be those which give an integral number of half wavelengths within the fixed length. The standing wave is produced by interference of the waves travelling in both directions caused by reflection at the ends. See the diagram on the right.

It is characterised by nodes, which are points of zero displacement, and antinodes, which are points of maximum displacement. The first few possible patterns for transverse waves on a string or spring are shown below. The separation of adjacent nodes or antinodes is

[13] Specification statement: understand that a stationary wave can be regarded as a superposition of two progressive waves of equal amplitude and frequency,travelling in opposite directions and that the internodal distance is $\frac{\lambda}{2}$



half a wavelength. The different numbers of half wavelengths on the spring or string correspond to different allowed frequencies. The lowest of these (on the left) is called the fundamental. The others are whole number multiples of the fundamental frequency and are called harmonics. Each subsequent frequency is known as an overtone. Because of the phase change on reflection at the fixed ends the ends are both nodes.

The wavelength can be found by measuring the node separation and doubling it. This gives a method for finding the speed of sound or electromagnetic waves if the frequency is known, using v = f. For example, if a standing wave is set up using 1 GHz radio waves and the node separation is found to be 0.15 m the wavelength will be 0.30 m and the wave speed is given by:

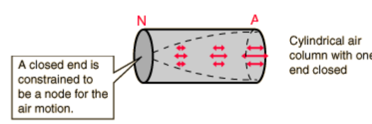
$$c = f \times \lambda = (1 \times 10^9 Hz) \times 0.30m = 3 \times 10^8 ms^{-1}$$

9 *Standing Waves in Musical Instruments*

Stringed instruments such as violins and pianos produce standing waves as described above. Hitting, plucking or bowing the string produces waves, which travel off in both directions. The pitch of the fundamental and therefore of the harmonics may be increased by reducing the length of the string, by increasing the tension in it or by using a lighter string.

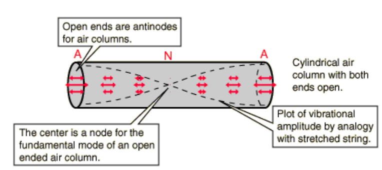
Wind instruments work by setting up longitudinal standing waves in tubes. Examples are organ pipes, trumpets and clarinets.

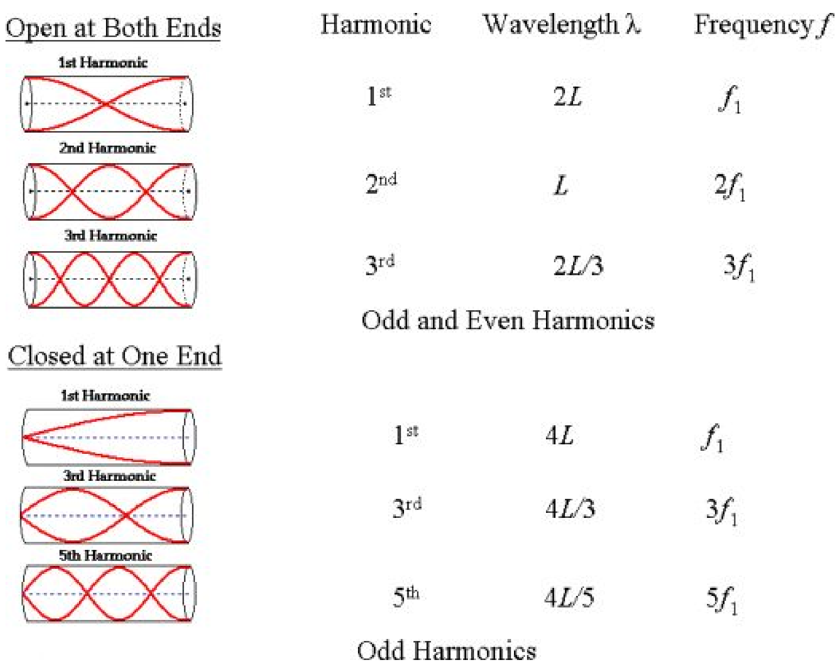
If the tube is open at one end and closed at the other there will be a node at the closed end and an antinode at the open end. At the fundamental frequency the length of the tube is therefore 1/4 of a wavelength. Harmonics occur at frequencies where the length of the tube is 3/4, 5/4, 7/4 etc wavelengths.closed



When any musical instrument is played the note you hear is at the fundamental frequency but a number of harmonics are also present. The relative intensity of the different harmonics determines the quality of the note you hear. It is this, which enables us to distinguish the same note played on different instruments.

If the tube is open at both ends both ends have an antinode and the fundamental occurs when the tube length is half a wavelength. Harmonics are at whole number multiples of the fundamental frequency.



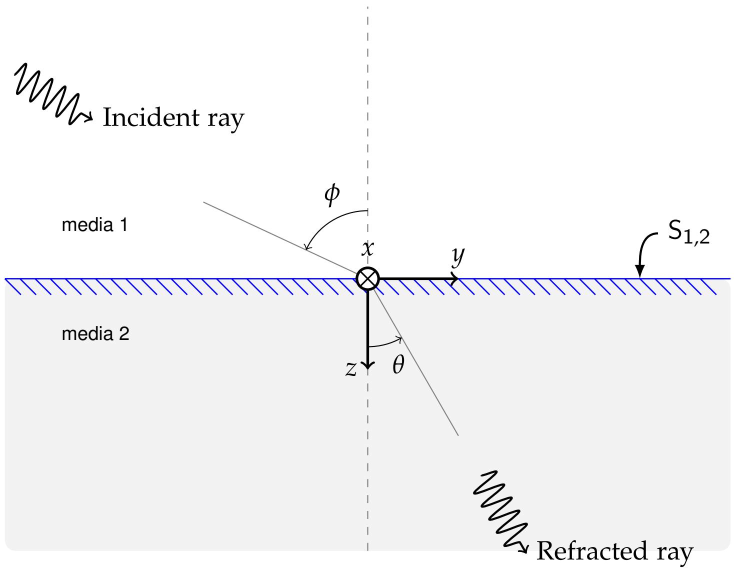


10 Refraction of Light

The Snell's Law equation is written in such a way as to emphasise symmetry. It is used to predict the change in direction that results from a change in velocity.

Looking at the equation for snell's law we see that there must be a point at which the incident ray causes a refraction at, or above 90 degrees. This is called the ***critical angle***. At shallower angles of incidence, total internal reflection occurs. The only application of total internal reflection required to be learnt is the step-index multimode ('thick core') optical fibre. Step-index means simply that the core is glass of one refractive index and the cladding is glass of a lower index (students need to know why it has to be lower), with an abrupt change in index at the interface. While such fibres are fine for conveying light for illumination, students need to know that they can't be used for transmitting a rapid sequence of data over a long distance. Multimode dispersion is the problem.

Light travelling at an angle to the axis of the fibre will travel further for a given axial length of fibre, than light travelling parallel to, or at a smaller angle to, the axis, and so will arrive later. Thus the arrival time of an element of data encoded in the light is smeared out. The element could start to arrive (by the shortest route) earlier than the previous element has finsished arriving by its longest route. Even worse confusion can occur. There are two ways round the problem of multimode dispersion…



$$\text{Snell's Law} = \frac{\sin\phi}{\sin\theta} = \frac{v_1}{v_2} = \frac{n_2}{n_1}$$

Figure 11: Refraction from low refractive index to higher refractive index (eg. Air to glass)

$$n_1 \sin\theta_i = n_2 \sin\theta_r$$

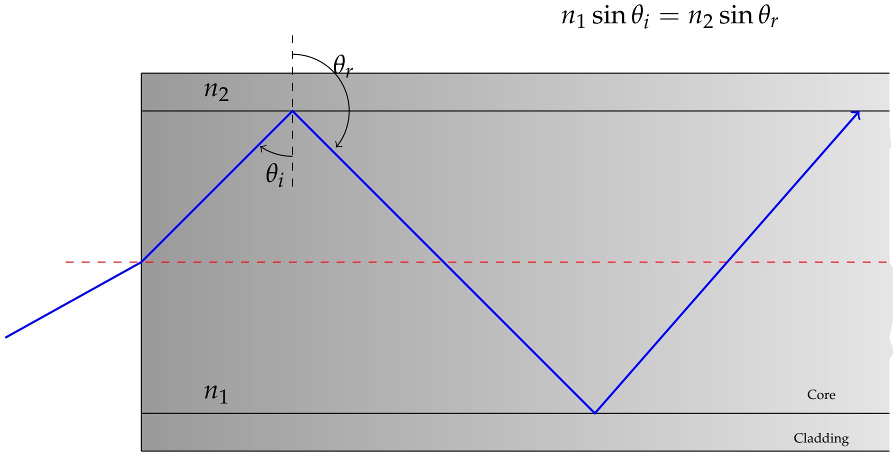


Figure 12: Total internal reflection in a single fibre

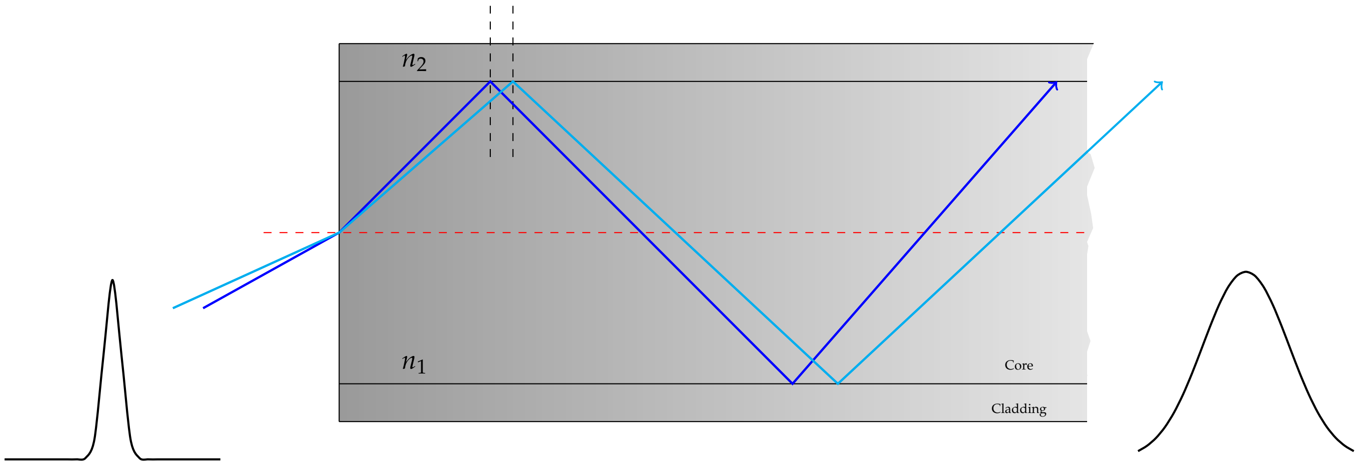


Figure 13: Multi-mode dispersion in a fibre resulting from the two different length paths

- Make fibres with graded index cores. This means cores which have a progressively lower index as we go out from the axis towards the interface with the cladding. The lower the index the faster the light travels so, if the grading is correctly calculated, the longer, more zigzaggy, paths cash in on the 'faster medium' and take no longer than the short, axial, route. Clever stuff, but note that graded index fibres are not in the WJEC specification.

- Make the core very thin. Its diameter must be no more than a few wavelengths of the light (or infrared) being carried. Such fibres are monomode. Light travels parallel to the axis. There are no zigzag modes. Students are required to know this. You are not required to know why very thin fibres are monomode. This is just as well, because it cannot be shown by ray optics, nor even by simple application of Huygens Principle. Electromagnetic wave theory is needed.

The website

`http://www.techoptics.com/pages/Fiber%20Optics%20-%20Optical%20Fiber.html`

gives an excellent summary of fibres for data transmission, with some facts and figures.

11 Photons

Everything in nature seems to come in lumps or quanta (singular: quantum). For example, ordinary matter is made of atoms, and electric charge comes in units of e. This lumpiness was only becoming fully accepted a hundred years ago. But in 1905 Einstein made the bold suggestion that light, too, was 'lumpy'. Light quanta are now called photons. A photon is a discrete packet of electromagnetic radiation energy. The energy of a photon is given by

$$E = hf$$

In which f is the frequency of the light and h is a constant called Planck's constant. $h = 6.6 \times 10^{-34}$Js [14]

Einstein suggested some experiments in which the quantisation of light should reveal itself. The simplest to understand involved the photoelectric effect, a phenomenon known about since the late 1880s.

[14] For interest only ... The constant h had first arisen in the earlier (1900) work of Max Planck on the radiation inside a cavity with hot walls. Planck had shown that the energies of oscillating particles in the wall seemed to be quantised.

12 *The Photo-electric Effect*

When electromagnetic radiation of high enough frequency falls on a metal surface, electrons are emitted from the surface. For most metals, ultraviolet is needed. For some (including sodium, potassium, caesium), light towards the violet end of the spectrum) will release electrons.

12.1 *Demonstrating the Photo-electric Effect*

Either using a zinc plate with a gold leaf electroscope (or a coulombmeter)

- Clean a zinc plate with fine emery paper or steel wool.
- Attach the plate to the top disc on a gold leaf electroscope, so there is good electrical contact.
- Charge the zinc plate and inner assembly of the electroscope negatively, e.g. by rubbing the zinc plate with a polythene rod which has been rubbed with wool or fur. The leaf should now be raised, because the leaf and the back plate are both charged negatively and repel each other. The leaf should temporarily rise further if the charged polythene rod is brought near the zinc plate.
- Place an ultraviolet lamp near the zinc plate. Switch it on. The leaf should be seen to fall. [Safety note: Don't look at the ultraviolet lamp (when it's turned on!)] Clearly the plate (and inner assembly of electroscope) is losing charge.
- Repeat the procedure, but charging the zinc plate and inner assembly of the electroscope positively, e.g. by rubbing the plate with a charged perspex rod. This time the ultraviolet does not affect the leaf. Charge is not lost.

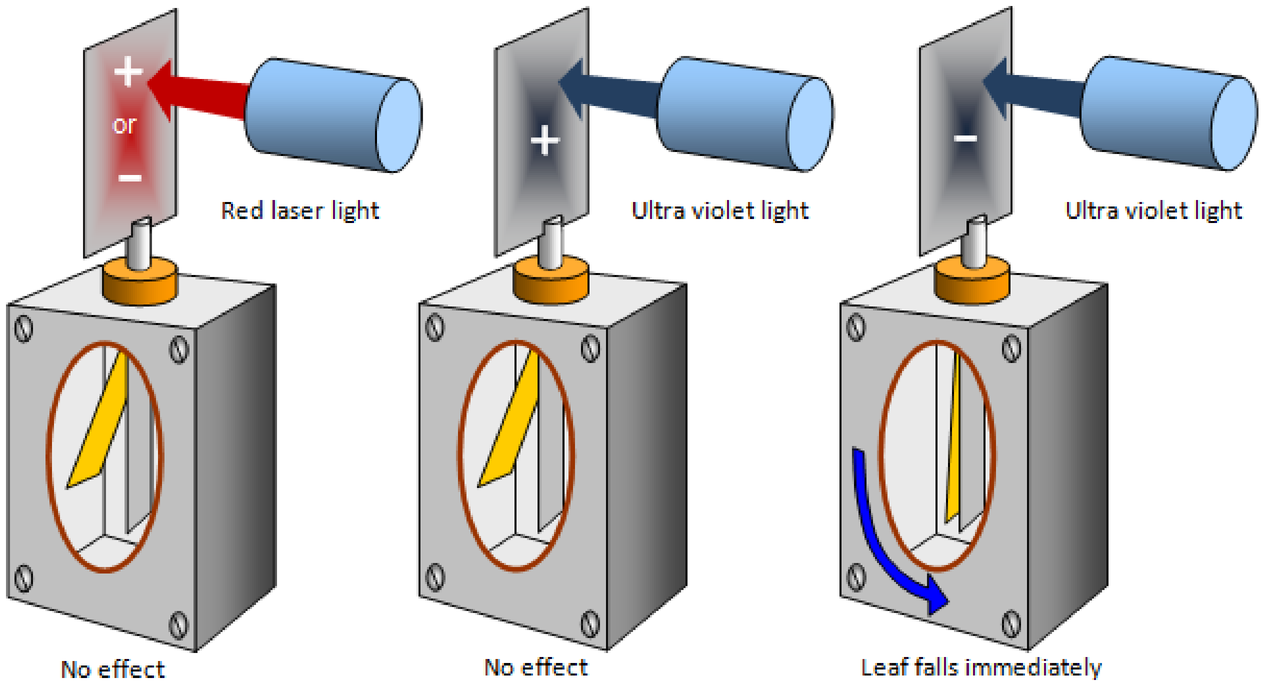


Figure 14: Photoelectric effect demonstrated with electroscope

The simplest explanation is the correct one... The ultraviolet causes electrons to be emitted from the zinc plate. If the plate is charged positively, the electrons are attracted back again. If the plate is charged negatively the emitted electrons are repelled and lost from the plate for ever.

12.2 *Photo-electric Puzzles*

Before 1905, the energy of a beam of light was thought of as distributed uniformly across broad wavefronts. Calculations showed that it should take some time before an electron in a metal surface could absorb enough energy from the light to escape from the surface. Yet emission is observed as soon as the light falls on the surface. Another puzzle was why, for a given surface, we find that light of frequency below a certain value (the threshold frequency) causes no electron emission at all. Einstein's theory of the photo-electric effect solves both these problems...

12.3 *Einstein's Photo-electric Equation*

Although the free electrons in a metal have no allegiance to particular atoms, there are forces 'bonding' them to the lattice of ions as a whole. In order to escape from the metal an electron has to do work against these forces. Some have to do more work than others, but there is a certain minimum quantity of work to be done, so no electron can escape unless it is given a certain minimum energy. The work function, ϕ, of a metal is the minimum energy needed by an electron in order to escape from the surface. Einstein's key idea was that any electron which leaves the surface is ejected by the action of a single photon. Photons don't co-operate in the process. Recall that a photon of light of frequency f has energy $= hf$. Suppose that a photon gives its energy hf to an electron, and that the electron is able to escape. The minimum energy used in escaping is ϕ, so the maximum kinetic energy the escaped electron can have is what's left over of the photon's energy. So we have the simple equation:

$$E_{Kinetic} = hf - \phi$$

This assumes that the photon energy is greater than (or equal to) the work function; in other words that $hf > \phi$ or $f = \phi/h$. If $f < \phi/h$, the photon energy will be less than the work function so no electrons at all can escape - a simple explanation of the phenomenon of threshold frequency.

The threshold frequency, f_0, for a metal is the minimum frequency of electromagnetic radiation needed to produce electron emission from the surface. From the argument just given,

$$f_0 = \phi/h.$$

This relationship can also be deduced from Einstein's equation; At the threshold frequency even the most energetic electron will only just manage to escape, so $KE_{max} = 0$, and therefore $f0 = \phi/h$. Provided that the light is above the threshold frequency, as soon as it falls on the metal surface electrons will start to be emitted, as emission results from individual photon 'hits', and is not a cumulative process as supposed before Einstein.

12.4 *Experimental test of Einstein's Equation*

We use the arrangement with the vacuum photocell shown below for demonstrating the photoelectric effect.

- Use white light with a coloured filter, or a light emitting diode, to illuminate the metal surface with approximately monochromatic light. [Its wavelength can be found using a diffraction grating, hence its frequency, using $f = c/\lambda$.]

- Increase the p.d. between the collecting electrode and the metal surface until the current drops to zero. At this point the p.d. is

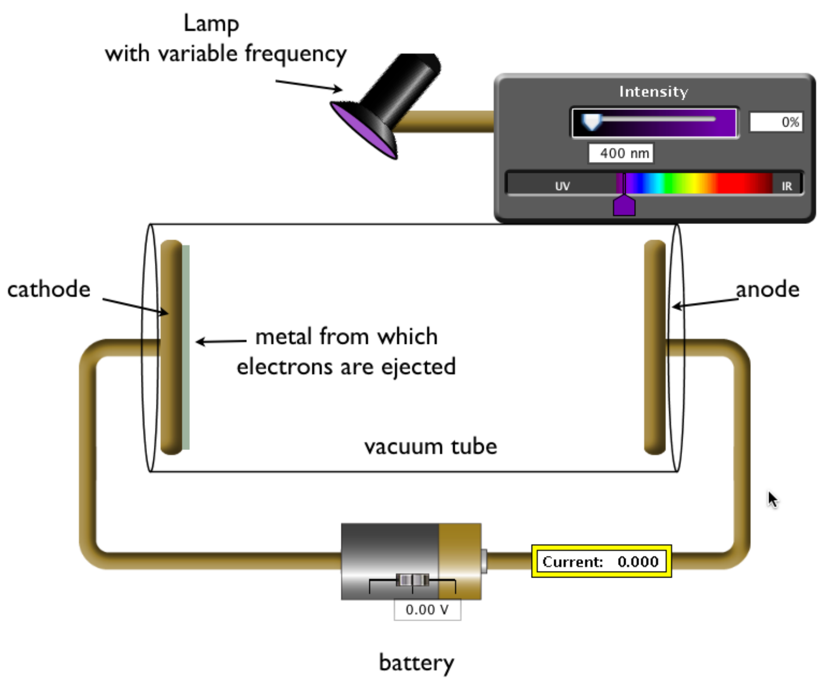


Figure 15: Photoelectric effect with vacuum cell. Note: Picture comes from the fantastic simulation at Phet.

called the stopping voltage, V_{stop}, because it stops all emitted electrons, even those with the most K.E., from reaching the collector electrode.

- The maximum K.E. of the emitted electrons is simply given by [15]

$$E_k = V_{Stop} \times 1.6 \times 10^{-19}$$

- Repeat the process using two or three more frequencies of light.
- Plot a graph of KE_{max} against frequency, f. If Einstein's equation is correct it should have a positive slope equal to h and a negative intercept, equal to ϕ. We can see this by comparing Einstein's equation with $y = mx + c$.

$$\begin{aligned} E = &\quad hf- &\quad \phi \\ y = &\quad mx+ &\quad c \end{aligned}$$

A sample graph is presented below.
It is useful practice to find from the graph:

- a value of Planck's constant,
- the threshold frequency for the metal,
- the threshold wavelength for the metal

[15] How do we justify this? Because of the applied voltage, emitted electrons are subject to repulsion by the positive collector electrode and attraction by the emitting surface, hence to a resultant force towards the emitting surface. The electrons therefore get slower and slower as they cross the gap. When the stopping voltage is applied, even the most energetic of emitted electrons have no K.E. left when they have made it across the gap. The K.E. lost is equal to the P.E. gained for these electrons. That's what the equation states. It is just like finding the K.E. of a ball thrown upwards in the Earth's gravitational field by measuring the its maximum height, and using the energy conservation equation K.E. lost = mgh

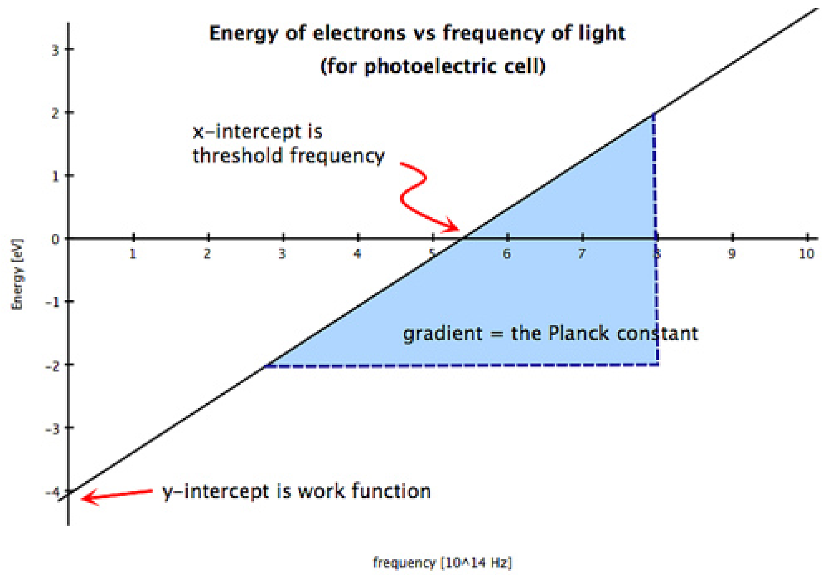


Figure 16: A graph of Frequency vs Kinetic energy

- the work function for the metal

16

12.5 *Effect of changing the Light Intensity*

If we bring a monochromatic light source towards a surface we increase the light energy falling on the surface, per m^2, per s. We are said to be increasing the intensity of the light. Clearly we can apply the same idea to ultraviolet or any other electromagnetic radiation. We find that:

1. For light or ultraviolet of a given frequency, changing the intensity has no effect on the maximum K.E. of the emitted electrons. This is exactly what Einstein's theory predicts. The energy given to individual electrons comes from individual photons, and a photon's energy, hf, depends only on the frequency (or, equivalently, the wavelength) of the radiation. It doesn't, then, depend on its intensity (provided we don't change the frequency).

2. For light or ultraviolet of a given frequency, increasing the intensity increases the number of electrons emitted per second.

Again, this is just what we'd expect from Einstein's theory. Increasing the intensity means increasing the number of photons arriving at the surface, per m2, per s. Naturally this means that more electrons will be emitted. [17] We can show the effect with the same vacuum photocell arrangement used for demonstrating the photoelectric

[16] In practice it is difficult to obtain a good value for h using a commercially available vacuum photocell. Slight impurities on the metal surface (e.g. a thin oxide film), and unwanted electron emission from the collector electrode, both affect the stopping voltage. The first convincing verification of Einstein's photo-electric equation, leading to an accurate value of the Planck constant was completed in 1916 by R.A. Millikan, working in the United States. The secret of his success was a remotely operated knife working in the vacuum to skim off surface layers from the caesium surface as they became contaminated.

[17] Each identical photon has the same probability of emitting an electron.

effect. Note that the polarity of the power supply is arranged to encourage electrons to cross the gap.

- Use a monochromatic light source to illuminate the caesium surface.
- Check that increasing the p.d. does not affect the current, I. This means that all the electrons emitted per second are being collected.
- Bring the light source closer and observe the effect on I.
- I is the charge flowing per second, so the number of electrons emitted per second is I/e in which e is the charge on each electron.

13 Laser Physics

Lasers, by now, are in nearly every household as they were once in every self-respecting science fiction film. Although, the death ray mystique will appeal to most A-level students there is real value in studying the physics which forms the foundation of laser construction. Unlike other subjects such as relativity or nuclear physics which can only be touched upon within an A-level course, the essential physics underpinning lasers can be taught reasonably well in a few lessons. Also, these fundamentals of lasers can be taught at the right level while not oversimplifying the subject.

LASER is an acronym and stands for Light Amplification by Stimulated Emission of Radiation. Which leads us nicely onto what is stimulated emission?

13.1 The Three Important Atomic Processes

These three processes are:

1. Absorption of light
2. Spontaneous emission of light
3. Stimulated emission of light

Absorption of light by an atom is shown in the diagram below - a photon of the correct energy is absorbed by the atom and an electron gains enough energy to move from the ground state to the excited state (Note: for the moment we are only considering the ground state and the first excited state only).

Spontaneous emission is the reverse process - an electron drops spontaneously (and randomly) from the excited state to the ground state and emits a photon of the same energy. These photons have random phase and random direction.

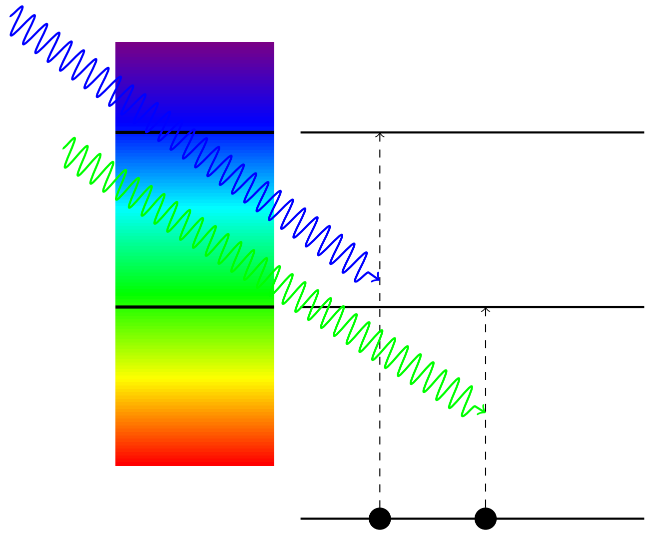


Figure 17: Absorption of light. The photon energy has to correspond to a possible energy level transition for the energy to be absorbed by an electron.

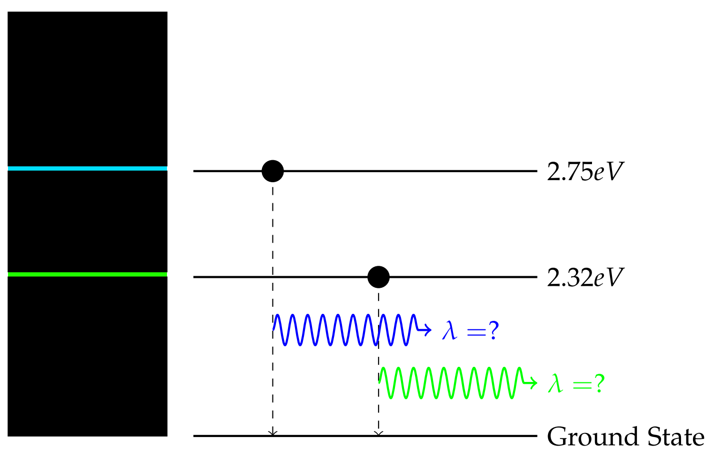


Figure 18: Emission spectra and corresponding energy levels. Note: the frequency and wavelength can be calculated from the change in energy of the electron using $E = hf$

However, there is also a third process [which was originally proposed by Einstein in 1917]. This process is known as stimulated emission - an electron is 'stimulated' to drop from its excited state by an incoming photon.

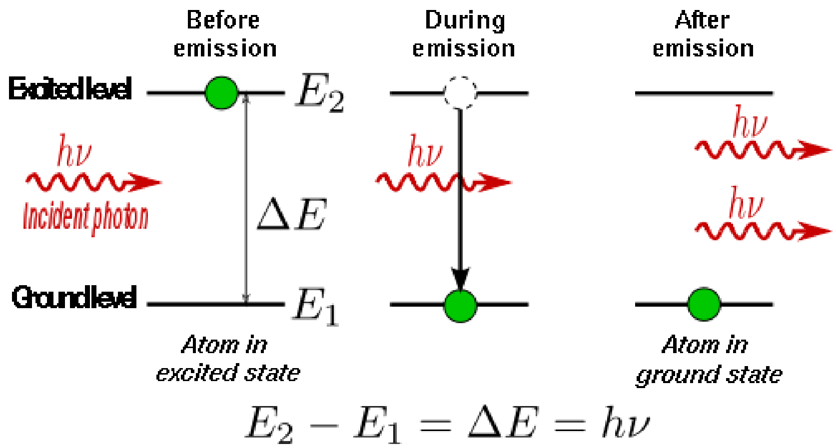


Figure 19: Stimulated emission - an incoming photon of an energy matching the transition can cause an electron to drop to a lower energy level. In order to do this it must release a photon identical to the first.

The reason that the electron is stimulated to drop is that the incoming photon is an electromagnetic wave and its e-m field will exert an oscillating force on the excited electron. If the incoming photon is of the correct frequency, this oscillating force will cause the excited electron to drop and both photons will exit with the same frequency, phase and direction. Note: again, the incoming photon needs to be of the correct energy.

13.2 *Inverting The Population*

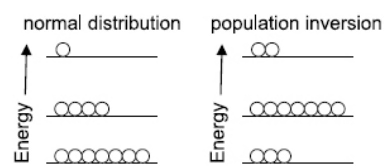


Figure 20: Population inversion - when more electrons are in an excited state than in the ground state. This is the opposite of a 'normal' situation

In order to get as much light out of a system as is possible we need to get as many atoms excited as is possible.

Obviously, the more electrons we have in an excited state the more will drop and emit photons (either spontaneously or through stimulation). However, there is one serious problem that arises when we produce a lot of light - the very photons that we produce are the actual photons that can be absorbed (they have the correct energy to produce both effects). If we have photons being absorbed all the time then our laser beam isn't getting any stronger.

Forget, for the moment about spontaneous emission (we are allowed to but we'll explain why later). When a photon arrives at an atom one of three things can happen:

1. It can pass by and do nothing.

2. It can be absorbed (if the atom is in the ground state).

3. It can cause stimulated emission (if the atom is in the excited state).

When it comes to producing a laser beam with a high intensity the three options above will have the following effect on the beam.

1. No change in the beam.

2. Net loss of one photon from the beam.

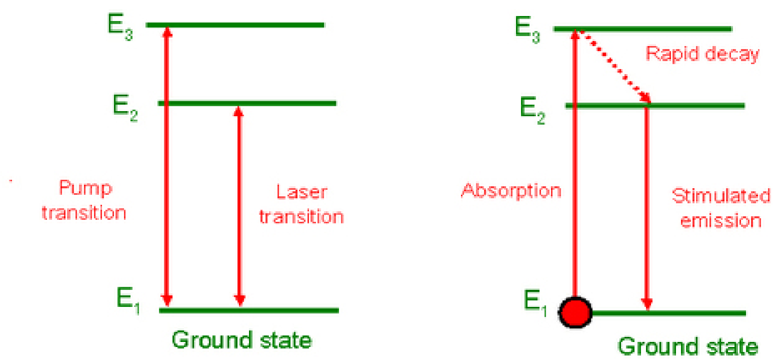
3. Net gain of one photon in the beam.

We need to arrive at a situation where stimulated emission is more likely than absorption so that the laser beam increases in intensity. Since stimulated emission occurs if the electrons are in the upper level and absorption when electrons are in the lower level we need to get more electrons into the upper, excited level. This is called population inversion (or $N_2 > N_1$ as stated in the syllabus, where N_2 and N_1 are the number of electrons in the excited state and the ground state respectively).

Unfortunately, this goes against what happens in nature - lower energy levels are always more heavily populated than higher energy levels when we have thermal equilibrium. There's only one thing for it - get rid of this thermal equilibrium. How do we do this? We continue to pump energy into exciting electrons to higher energy levels to maintain a population inversion and to break the conditions of thermal equilibrium.

Population inversion is not usually possible if we only have two energy levels (if pumping is carried out by light).

13.3 The 3 Energy Level Laser System



[18]

1. Pumping. Electrons are promoted from the ground state (E1) to E3 usually by using an external light source or by electron collisions.

2. Electrons drop quickly (because E3 is chosen to have a short lifetime of the order of nanoseconds) to the metastable (E2). Calling E2 metastable means that it has a long lifetime and electrons stay there for a long time (not that long really around a millisecond but that's a very long time for an electron).

3. This is the transition that produces the laser photons so we must have N2 > N1. Note that, although stimulated emission still reduces our population inversion, the pumping is at a different wavelength. We have to make sure that the pumping [1] exceeds the stimulated emission [3] to maintain a population inversion.

[18] Note:

- E3 (to E2) has to have a short lifetime because E3 cannot start to fill up - pumping won't then be possible. Also, we don't want the electrons to stay in E3 and have them stimulated to drop back to E1 by the pumping light - that's back to the 2-level system again which wasn't quite good enough.
- More than half the electrons from E1 must be pumped to E2 (via E3) in order to obtain a population inversion - that's a lot of electrons!

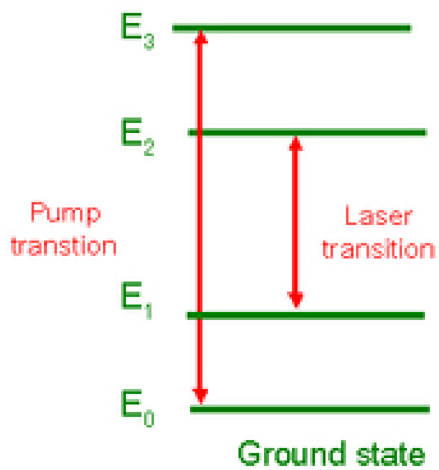
13.4 The 4 Energy Level Laser System

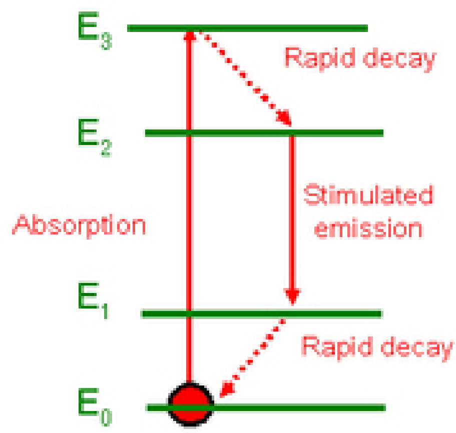
1. Pumping again.

2. Quick drop to the metastable state E2.

3. This is the laser light producing transition so this time N2 > N1. However, because E0 is the ground state, E1 is practically empty initially so obtaining population inversion is far, far easier (definitely no need to pump half the electrons!).

4. Another quick transition so E1 has a short lifetime. This is because we want E1 to be empty so that we have a population inversion (if N1 is small it's easier for N2 to be larger than N1).

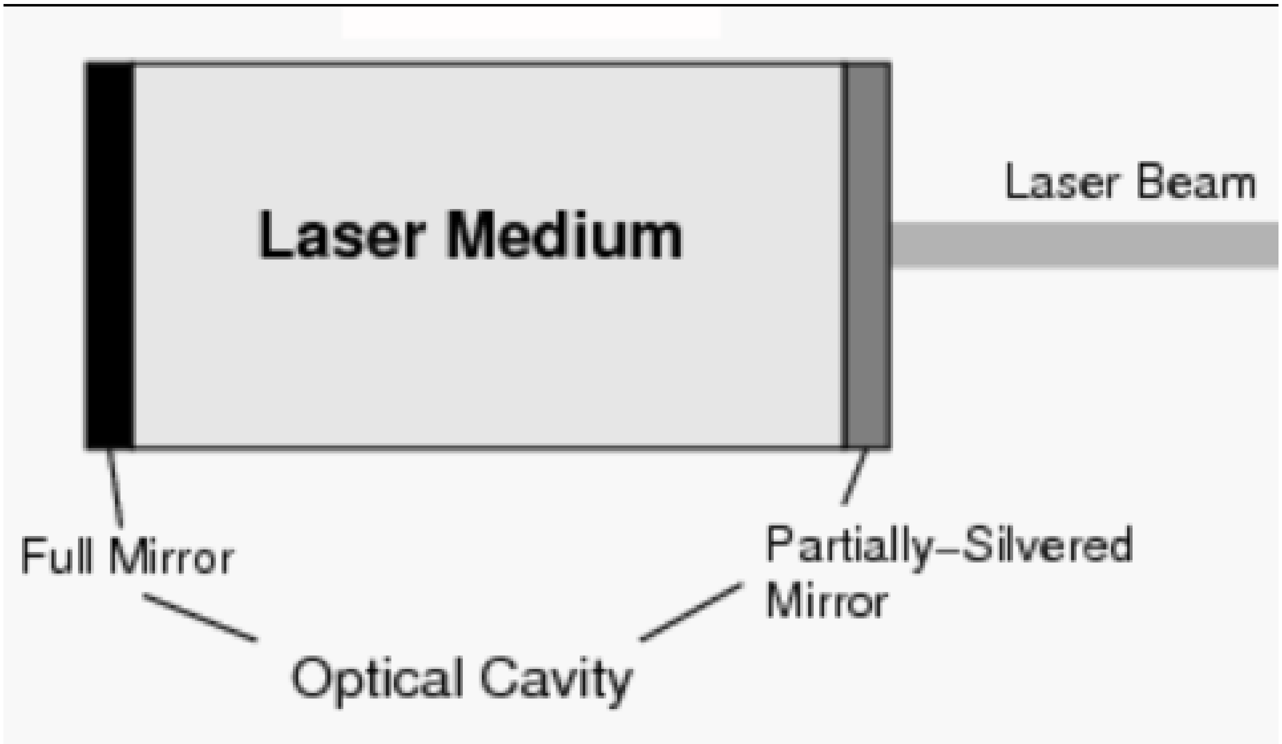




13.5 Laser Construction

In order to ensure that the laser produces light of a high enough intensity, the above set up is used. The amplifying medium is the region where the population inversion exists. This means that the conditions are right in the amplifying medium for stimulated emission. Under these conditions one photon has the potential to produce two photons and these can produce 4 photons, then 8 photons etc. Like a chain reaction, this process will lead to an exponential increase in output energy. Laser physicists aren't happy with this, they go even further - they use mirrors to ensure that this exponential increase happens many times. Because only 1% of the light exits each time it reflects back and forth between the mirrors, on average, the beam will pass through the amplifying medium a hundred times before it exits. Now, considering that each time the beam passes through the amplifying medium it is increasing exponentially, this factor of 100 makes an enormous difference.

This all leads to very high light intensities inside the amplifying medium and this is why (as was said earlier) we can forget about spontaneous emission. Imagine that you're an excited electron sitting happily in your higher energy level. Normally, you'll just drop down spontaneously when your time is up. But, inside a laser, there's so much light that you never drop spontaneously because before your time's up you've been disturbed by another photon, stimulated to join in with all the other light and join in coherently as well!



13.6 *Efficiency*

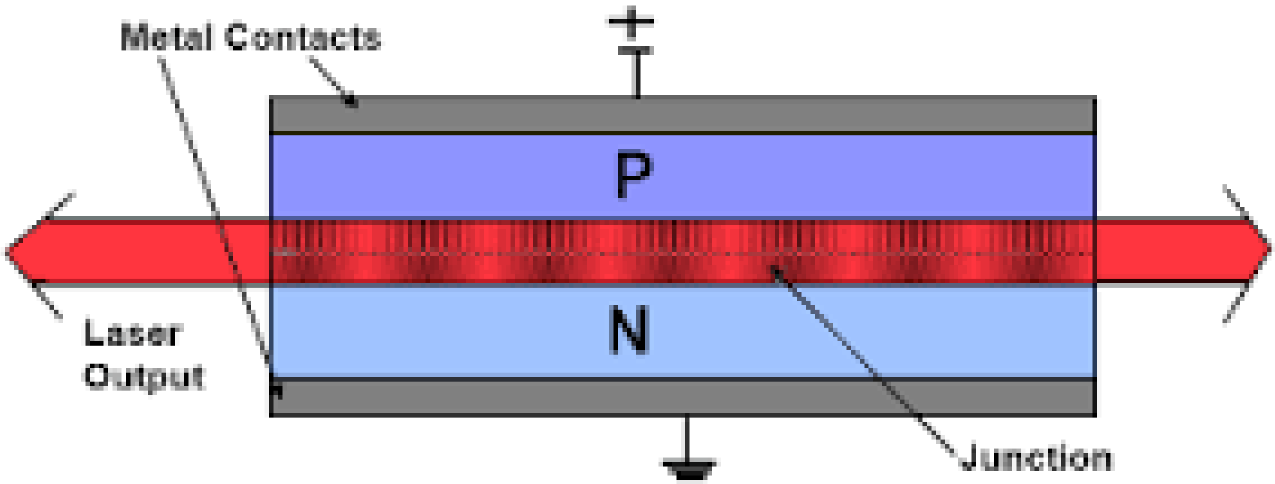
Usually, lasers are very inefficient beasts. Because of the large energies required to maintain a population inversion, their efficiencies are generally far below 1%. Some reasons for this:

- The pumping energy is considerably larger than the output photon energy.
- High intensity pumping combined with the high intensity of the laser beam means that the amplifying medium will get very hot. So, there will be large heat losses. To make this matter worse, we need to cool the amplifying medium usually so that it, or its container, doesn't melt. By cooling the system we just transfer more heat and increase our losses but better this than destroy a £50 000 laser!

13.7 *Semiconductor Lasers*

The basic structure of a standard 'edge emitting' semiconductor laser is shown below. The whole block shown below is a semiconductor chip with dimensions approximately 0.5mm $\times$ 0.5 mm $\times$ 1 mm.

The above laser fits the basic shape of a normal laser, the mirrors, however, are far from the 100% and 99% reflecting ideals discussed



earlier. The mirrors are simply due to the semiconductor-air boundary at the edges of the chip. [This in fact gives 40% reflection only (at both sides).] This would be disastrous for highly inefficient gas lasers but not for our semiconductor laser. The reason why:

- The population inversion inside the semiconductor sandwich area is millions of times higher than in gas lasers [10^{25} electrons/m^3].
- The exponential increase in light intensity (i.e. 1 photon becoming two, becoming four etc.) occurs far more quickly because of the higher population inversion.
- So the fact that we lose 60% of the light at each reflection is compensated for by having huge gains between the mirrors.

13.8 *Advantages and Uses of Laser Diodes*

These are straightforward and can be summarised as follows:
Advantages:

- Cheaper
- Smaller
- More efficient
- Easy to mass produce

Some uses:

- Inside DVD and CD players
- Bar-code readers
- Telecommunications (via optical fibres)

- Image scanning
- Laser surgery

The usefulness of laser diodes is 'reflected' in the number of them produced annually - around 1 billion (109) laser diodes are produced worldwide per year!

For more than enough further reading see:

- `http://members.aol.com/WSRNet/tut/ut1.htm`
- Wikipedia `http://en.wikipedia.org` then type 'laser' or 'semiconductor laser'
- Google search 'laser theory'