Physics Factsheet

January 2001 Number 07

Electrical Current, Voltage and Resistance

What is current?
Electricity can seem very abstract and difficult to understand. The key to grasping the subject, like so many in Physics, is to build up a picture of what is happening and follow the concepts through logically. The aim of this Factsheet is to help you do just that by explaining in simple terms what current, voltage and resistance are and how they all play a part in an electrical circuit.

An **electric current** is nothing more than a net movement or flow of charge in a certain direction. In a conducting metal the charge carriers are free electrons; these electrons originate from the rigidly bonded metal atoms that form the structure of the conductor. Their outer electrons are only weakly bonded to the atom and so many escape and are free to move throughout the structure of the metal. As the metal atoms have lost electrons they are no longer neutral but are now positively charged. Good metallic conductors include silver and copper.

Metals are not the only materials that conduct; semiconductors are a group of materials whose resistance lies somewhere between that of metallic conductors and insulators. A semiconductor is made from covalently bonded materials. Electrons in the outermost orbits only have a small 'jump' to make to move to the next orbit a little further away. When an electron does this it has two implications. As there are not many electrons in these higher levels it can jump into a vacant site in the adjacent atom and in so doing move through the element. Secondly, when it jumps up it leaves a vacant site below it known as a hole, these holes act as though they were positively charged and move the opposite way through the metal, so we get double the current we would expect. Silicon and Germanium are examples of semiconductors.

Liquids can also conduct as they long as they contain charged particles. For instance, impure water will conduct as the impurities in it exist in the form of ions, which move through the liquid. Pure water will not as the H_2O molecules are neutral.

*Current is the **rate** of **flow** of **charge***
*If the current is **constant**, we have:*

$$I = \frac{Q}{t}$$

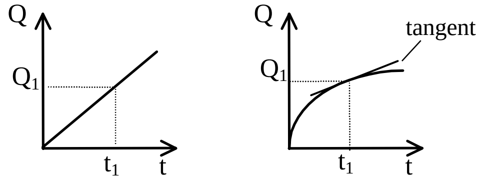
I = current (amps, A) Q = charge flowing past a point (coulombs, C)
t = time taken for the amount of charge Q to flow (seconds, s).
*This formula will also give **average current** if the current is variable.*

*If the current is **not** constant, then the gradient of a charge (y-axis) against time (x-axis) graph would give the current at a particular time*

Current and electrons
We now know that current is rate of flow of charge. The flow of charge in solids depends on the movement of charge carriers; these are generally **electrons**.

Current therefore depends on the number of free (meaning "able to move") charge carriers in the material and how quickly they move - the more charge carriers there are and the faster they move, the higher the current will be.

Finding current and average current from a graph



In both of the above graphs, we need to find the current at time t_1
In both cases, we use the **gradient** of the graph to find the current.

- In the first graph the gradient is constant and so therefore is the current. So dividing Q_1 by t_1 will give the right answer as you are taking the gradient of the graph.
- In example 2 dividing Q_1 by t_1 will give the **average** current up untilthat point. The gradient has been falling so the current at the time $t_1$will be less than the average value. So instead we must use a **tangent** to the line at t_1. Watch out for the distinction between average and instantaneous values.

Drift velocity
All free electrons move around due to their thermal energy even if there is no current present. However as this motion is completely random then the net effect is no overall movement. To be part of a current they have to exhibit a drift velocity in a given direction

Consider the two electron paths below:

Both display random motion as their thermal kinetic energy causes them to move, colliding with the fixed positive ions that make up a metal's structure as they go. The first electron shows very little change in displacement, the second has moved about the same distance as the first but shows a definite displacement to the right - it has a drift superimposed on its random motion. Therefore we can say the second electron is probably part of an electrical current and the first is not.

When the overall effect on all the electrons is taken into account this small drift shown by each electron provides a current, whilst when all electrons in the first example are considered the net movement of charge is zero.

Current can be calculated from:

$$I = nAQv$$

I = current (amps, I)
n = number of free charge carriers per m^3
Q: Charge on each charge carrier. (coulombs, C)
A: Cross-sectional area (m^2)
v : Drift velocity (ms^{-1})

***Exam hint**: Many students lose marks by not using the correct units in the equation above. In particular, note that cross-sectional area must be in m^2.*
To convert cm^2 to m^2, divide by 100^2 = 10 000.
To convert mm^2 to m^2, divide by 1000^2 = 1000000

If I, A and Q are constant, then v is inversely proportional to n. In other words, to carry the same current, if there are fewer charge carriers they must move at a higher speed.

Metallic conductors have a far higher value of n; for a metal and semiconductor of the same dimensions, carrying the same current, v must be higher in the semiconductor typically around m/s, as opposed to mm/s for a metal.

In an insulator there are no free charges available to carry current, therefore n = 0m^{-3} and therefore, from the equation, I = 0A.

Typical Exam Question:
Calculate an average value for the drift velocity of free electrons moving through a wire of area 1.5mm^2, when they form a current of 6.1A. Copper contains 1.0 × 10^{29} free electrons per m^3. The charge on an electron is 1.6 × 10^{-19} C **[3]**

We are going to use the equation I = nAvQ, so we need to convert the cross-sectional area into the correct units:

A in m^2 = 1.5× 1÷ 10^{-6} = 1.5×10^{-6}m^2 ✓

Substitute in to our equation:

v = I/(nAQ)

= 6.1/(1 × 10^{29} × 1.5× 10^{-6} × 1.6× 10^{-19}) ✓

= 0.00025 ms^{-1} (= 0.25 mms^{-1}) ✓

Why does current flow?

By now you should have a picture of a conductor, for instance a metal, as structure of fixed positive ions surrounded by a sea of free electrons, colliding and rebounding with these ions as they flow, but gradually making their way from one end of the wire to the other. These electrons behave much as an incompressible liquid which explains why current starts to flow immediately when a switch is closed and the rate of flow is the same throughout, just like water flowing along a pipe.

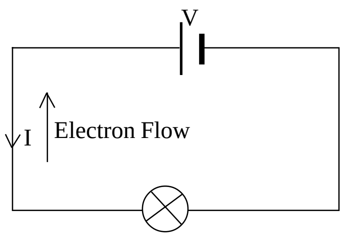
The question is; what causes charged particles to flow? The answer – **electromotive force (EMF) or voltage**. Voltage is one of most fundamental yet widely misunderstood quantities in electricity. It is a difference in potential between two points in an electrical circuit.

You can compare this to gravitational potential energy. Imagine a ball held above the ground. The ball has more potential energy because it is above the ground. It tends to fall towards the point with lower potential energy.

Similarly, a **positive** charge will "fall" from the higher (more positive) potential to the lower (more negative) potential. **Negative** charges - like **electrons** - behave in the opposite way, and move from a lower potential to a higher potential (this is where the analogy with a ball falling fails to work any more).

Current is conventionally said to be in the direction that a **positive** charge would move. So **electrons move in the opposite direction to the conventional current**.

To try and understand voltage further let's consider a simple electrical circuit.



The cell imposes an EMF on the circuit - its negative terminal contains extra electrons and its positive terminal lacks them. As soon as a complete circuit is formed the free electrons in the wire are attracted towards the positive terminal and repelled from the negative so they start to drift and are replaced by the extra electrons from the cell.

As the electrons approach the positive terminal they lose potential energy, in much the same way a mass loses potential energy as it approaches the earth. This potential energy cannot disappear - it must have been converted to a different form. In a vacuum the electrons would have gained kinetic energy as they would have accelerated towards the higher potential. However we know that electrons in a metal move with a constant (drift) velocity and therefore do not gain kinetic energy. In fact the energy is transferred to the bulb which converts it into heat and to a lesser extent light.

The light bulb does not use electrons - this is why the current either side and indeed all the way around the circuit is the same; rather it converts the loss of PE of the electrons into heat and light. This is an example of the transfer of electrical energy. This is essentially what all components in an electrical circuit are designed to do. The electrons transfer energy from the voltage source (eg. a battery) to components as they are pushed around the circuit. When a cell has lost its extra electrons and the positive terminal has become neutral it is flat and must be recharged.

Potential Difference and Electromotive Force

Although potential difference is related to energy, strictly speaking it is defined as the amount of electrical energy dissipated by a unit charge when it moves between two points in a circuit. A p.d. of 1 volt between two points means that a charge of 1 coulomb will dissipate 1 joule of energy when it moves between them.

$$V = \frac{W}{Q}$$

V: Potential difference between two points.
W: Energy dissipated in moving between those two points.
Q: Total charge that has moved between the two points.

An EMF is also a difference in potential but it causes the current to flow. For the EMF, W would be electrical work done on the charge by the supply (i.e. energy supplied) instead of a measure of work done by the charge on a component (i.e. energy dissipated).

Remember p.d. is relative: two points that are at 10,000 and 9998V and two points that are at 4 and 2V both have p.d.s of 2V across them. A bird standing on a high voltage power line does not get electrocuted as both its feet are at the same, albeit very high potential. If there is no p.d. then no current flows and it survives to fly another day. When dealing with circuits we are able define one point - usually the negative terminal of the battery -as having 0V potential and then measure all other potentials relative to this.

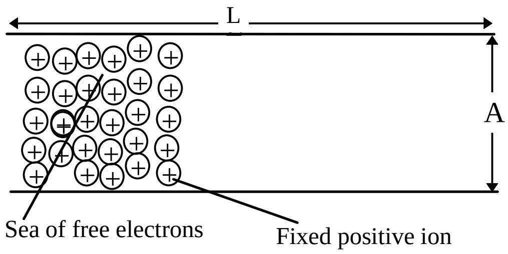
Remember current is a flow and therefore is always **through** something. Potential difference is a difference between two points and is therefore always **across** something.

Now you should have a picture of free electrons travelling through the structure of the metal delivering energy to any component they pass through. Imagine the electrons as a traffic jam in which the traffic is nose to tail; if there is a break in the circuit no current can flow as the front electrons have nowhere to flow to, meaning electrons throughout the circuit are stationary. If the traffic moves we will get the same number of cars past every point in the road, as they all travel as fast as the cars at the front allow. In more Physical terms the sea of free electrons behaves as an incompressible liquid, with any EMF providing a push but the liquid only flows if the pipe it is travelling in is not blocked.

Resistance

Resistance in metals has its origins in the atoms that make up the material. In solid form the atoms are tightly bound into a lattice structure. It is their outermost electrons that escape to form a sea of free electrons. When the electrons escape they leave their atoms as fixed positive ions.

When we apply an emf, free electrons are pushed away from the negative terminal and towards the positive, so they try to flow from one end of the wire to the other. As they try to get past the fixed positive ions they collide repeatedly, rebounding after each collision, this generates the random component of their motion, with the applied emf imposing the drift. When they collide with the positive ions they transfer some of their kinetic energy to it. This explains why their kinetic energy does not increase as they lose potential energy - it is transferred to the wire, and also how energy is transferred to a component or wire by a current. The more the ions get in the way of the electrons, the higher the resistance and the lower the current that will flow.



If the current transfers energy to a component faster than it can dissipate the heat to the surroundings, its temperature will increase. As the fixed positive ions gain thermal KE their vibration around their fixed equilibrium positions increases and they get in the way more. The electrons find it harder to get through the metal hence the resistance increases. In all metals resistance goes up as temperature increases.

We define resistance by

$R = \frac{V}{I}$

R: Resistance in ohms (Ω)
V: potential difference across component (V)
I: current across component (A).

Ohm's law states that for a metallic conductor I α V as long as its temperature is constant, in other words if we double the voltage then the current will also double.

As a semiconductor's temperature increases more electrons are promoted so they are free to conduct. This increase in the number of charge carriers more than compensates for the increased vibrations in the material. We say it has a negative coefficient of resistance - as its temperature increases its resistance falls. This can be seen from I = nAvQ. If the number of charge carriers per unit volume increases, than the size of current for a given p.d. increases.

Resistivity

Different metals will inhibit currents by different amounts due to differences in their structures. We call this **resistivity**. We use resistivity for the same reason we use the Young Modulus - it is independent of a material's dimensions.

Using our previous diagram, if the length of the conductor is increased then its resistance must also increase as the electrons have further to travel and therefore have more ions to get past. If we increase the cross-sectional area, then resistance falls because there are more gaps for the electrons to pass through. The relationships are directly and inversely proportional respectively, i.e:

$$R \; \alpha \; L/A$$

The constant of proportionality is resistivity and it is constant for a any amount of a given metal. For a conductor **it is defined as the product of resistance and cross-sectional area per unit length**.

$R = \frac{\rho L}{A}$

R: Resistance (Ω)
ρ: Resistivity(Ωm)
A: Cross-sectional area (m^2)
L: Length of conductor (m)

Remember whilst resistance depends on dimensions resisitivity, depends only on the type of metal. Sometimes you will be asked to calculate conductivity, this is just the inverse of resisitivity.

Typical Exam Question:
A heating coil made from 11 metres of wire is connected to a 240V mains supply. The wire has diameter of 0.25mm and resistivity of $1.0\times10^{-6}\Omega m$.
(a) Calculate its resistance **[4]**
(b) How much current will flow in it ? **[1]**

(a) As we are going to use R = ρl/A, we calculate the cross-sectional area:
The radius is half the diameter
r = 0.25/2 = 0.125mm ✓
$A = \pi r^2 = \pi(0.125\times10^{-3})^2 = 4.9\times10^{-8}m^2$ ✓
Now we substitute into our equation involving resistivity:
$R = \rho l/A = (1\times10^{-6}\times11)/(4.9\times10^{-8})$ ✓ $= 224\Omega$ ✓
(b) To calculate current we have a voltage and resistance:
So using I = V/ R = 240/224 = 1.07Ω ✓

Now you should be able to understand current as a flow of charged particles under the influence of a potential difference. In a metal electrons drift from one end of the wire to the other with fixed positive ions impeding their progress. It is this resistance to their motion that causes the electron's loss in potential energy to be transferred to the component the current moves through. Ohm's law links the quantities of potential difference, current and resistance and, at a constant temperature, R remains constant.

Typical Exam Question
A 12 cm length of copper wire of area $4\times10^{-7}m^2$ is connected across a potential difference of 2V. A current of 4A is measured flowing through the wire. Charge carrier density for copper = $1.0\times10^{29}m^{-3}$.
Calculate
(a) The resistivity of copper. **[4]**
(b) The drift velocity of the electrons in the wire. **[3]**
(c) If the area of the wire is doubled what effect will this have on thedrift velocity, provided the current is unchanged? **[3]**
(d) The current in the wire is increased to a point where the wire begins to heat up. What effect does this have on the resistance of the wire and why? **[3]**

(a) Firstly find the resistance using Ohm's law.
R = V/I = 2/4 = 0.5Ω ✓
Rearrange the formula and substitute in
$\rho = RA/l$ ✓ $= (0.5\times4\times10^{-7})/12 = 1.7\times10^{-8}\Omega m$ ✓
(b) Rearrange the equation containing drift velocity and substitute the values
$v = I/(nAQ)$ ✓ $= 4/(1\times10^{29}\times4\times10^{-7}m^2\times1.6\times10^{-19})$ ✓
$= 6.25\times10^{-4}ms^{-1}$ ✓
(c) As v = I/(nAQ) ✓ *and I, n and Q are constant* ✓
then if A doubles v will halve ✓*as v α 1/A.*

(d) If the wire heats up then the resistance of the wire increases ✓*.*
This because the fixed positive ions in the wire vibrate more ✓
and so impede the flow of electrons more ✓*.*

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's answer is given below.

(a) Define potential difference. [2]

Potential difference is the energy lost ✓ between two points.

1/2

> Student has forgotten about unit charge and answer should imply moving charge.

(b)i) A constantan wire has a diameter of 0.4mm and a length of 150cm. Find its resistance given the resisitivity of constantan is $5\times10^{-7}\Omega$m. [5]

$A = \pi r^2 = \pi \times 0.2^2$ ✓ $= 0.125m^2$ ✗

$R = \rho l/A = (5\times10^{-7}\times 1.5)/0.125$ ✗ $= 6\times10^{-6}\Omega$ ✗

1/4

> Student has remembered to halve the diameter but not to convert mm^2 to m^2. Equation is correct but the incorrect answer is obtained as area is wrong.

ii) What would happen to the resistance if the diameter and length were both doubled? [4]

$R = \rho l/A$

If length is doubled resistance doubles ✓

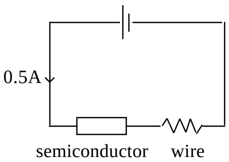
If diameter is doubled then resistance halves ✗

So overall resistance is unaffected ✗

1/4

> Student has not realised that if the radius/diameter is doubled then the area increases by four times because we square radius to get area. Student's logic is correct with the answers worked out but does not give the correct answer.

(c) A semiconductor of the same dimensions is connected into a circuit



with the wire as shown below:

Find the ratio of drift velocities of charges in the wire compared to the semiconductor if the semiconductor has 7.2×10^{25} charge carriers per m^3 and the wire $1\times10^{29}m^{-3}$. [3]

For the semiconductor:

$v = I/(nAQ) = 0.5/(7.2\times10^{25}\times0.125\times1.6\times10^{-16}) = 3.5\times10^{-7}ms^{-1}$ ✓ecf

For the wire:

$v = I/(nAQ) = 0.5/(1\times10^{29}\times0.125\times1.6\times10^{-16}) = 2.5\times10^{-10}ms^{-1}$ ✓ecf

2/3

> Answer is wrong due to the area but as student has already been penalised student attains marks due too error carried forward. Student has forgotten to take a ratio.

Examiner's answers

(a) *Potential difference is defined as the amount of energy transferred ✓ by unit charge ✓ when it moves between two points.*

(b) i) *$A = \pi r^2 = \pi\times(0.2$ ✓ $\times10^{-3}$ ✓$)^2 = 0.125\times10^{-6}m^2$ ✓*

$R = \rho l/A = (5\times10^{-7}\times1.5)/(0.125\times10^{-6})$ ✓ $= 6\Omega$ ✓

ii) *$R = \rho l/A$*

$R \propto$ length therefore if length is doubled resistance doubles ✓

$A \propto r^2$ therefore if diameter/radius is doubled then the area increases by four times. ✓

$R \propto 1/A$ and so resistance falls to a quarter. ✓

So overall resistance falls by one half ✓

(c) *For the semiconductor*

$v = I/(nAQ) = 0.5/(7.2\times10^{25}\times0.125\times10^{-6}\times1.6\times10^{-16}) = 0.35ms^{-1}$ ✓

For the wire:

$v = I/(nAQ) = 0.5/(1\times10^{29}\times0.125\times10^{-6}\times1.6\times10^{-16}) = 2.5\times10^{-4}ms^{-1}$ ✓

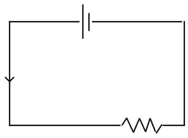
$v_{semiconductor}:v_{conductor}$ $2.5\times10^{-4}: 0.35$ 1:1400

Timed test 19 marks – 20minutes

1. (a) Define the following electrical terms and state the relevant SI units.
   i) Current [2]
   ii) Resitivity [3]
   iii) Potential difference [2]

   (b) What is the difference between the concepts of potential difference and EMF? [2]

2. (a) A p.d. of 1V is applied across the copper wire shown below.



wire of length 70cm and cross-sectional area $7\times10^{-8}mm^2$

   i) Find its resistance. [3]
   ii) Find the current flowing in the wire. [1]
   iii) Find the drift velocity [3]

   (b) Sketch the path you would expect an electron to follow when part of a current flowing in a metallic conductor. Show the polarity of your supply. [3]

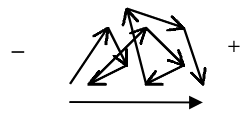
Answers

1. (a) i) Current is rate of flow of charge ✓. SI unit is the amp/ampere. ✓
   ii) Resistivity is defined as the product of resistance and area ✓ per unit length ✓. SI unit ohm metre/ Ωm ✓
   iii) Potential difference is the energy transferred per unit charge when a charge moves between two points in circuit ✓.
   SI unit is the volt ✓.

   (b) Emf is energy transferred to charge carriers ✓ whereas potential difference is energy transferred from charge carriers to components ✓.

2. (a) i) $R = \rho l/A$ ✓ $= 1.7\times10^{-8}\times0.7/(7\times10^{-8})$ ✓ $= 0.17\Omega$ ✓
   ii) $I = V/R = 1/0.17 = 5.9A$ ✓
   iii) $v = I/(nAQ)$ ✓ $=5.9/(1.0\times10^{29}\times7\times10^{-8}\times1.6\times10^{-19})$ ✓ $=5.3mms^{-1}$ ✓

   (b)



Drift element of motion shown ✓
Random element of motion shown ✓
Correct + and - shown. ✓

Acknowledgements: *This Physics Factsheet was researched and written by* ***Alan Brooks.*** *The Curriculum Press, Unit 305B, The Big Peg, 120 Vyse Street, Birmingham, B18 6NF. Physics Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber. No part of these Factsheets may be reproduced, stored in a retrieval system, or transmitted, in any other form or by any other means, without the prior permission of the publisher.* ***ISSN 1351-5136***

Physics Factsheet

www.curriculum-press.co.uk Number 111

Electrons as Charge Carriers

Mobile charges are necessary for the flow of current, or for the build-up of electrostatic charge. The most common charge carriers we deal with are electrons carrying current around a circuit. But there are a wide variety of ways in which charge can flow.

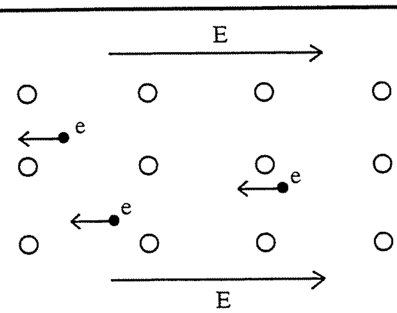
Examples are electrons flowing through a conducting material, electrons flowing from one material to another, electron beams through a vacuum, electron-ion pairs through a low pressure gas, and ion-ion pairs through an electrolyte (although this example is outside the remit of this factsheet).

In this factsheet, we will look at some of these examples, and also look at the forces causing the flow of current and the relevant calculations.

Conduction electrons in a metal

In any metal, the one or two outermost electrons are only weakly bound to their nucleus. Within the lattice structure of the body of the metal, these electrons separate from their atoms (which become ions) and are free to carry electric current. We call them **conduction electrons**.

There are two ways in which these conduction electrons can be persuaded to flow in an organised manner, and carry a current. The first is through an applied voltage.

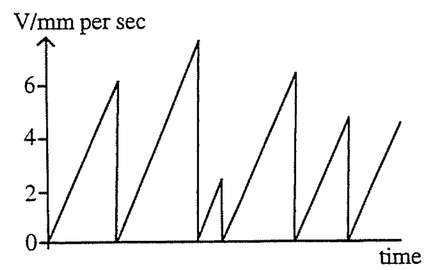


The applied voltage sets up an electric field within the metal, according to the equation:

$E = \dfrac{V}{d}$ Units: Vm^{-1} or NC^{-1}

This field then exerts a force on the conduction electrons:
$F = Ee$

This causes the electrons to accelerate towards the positive terminal of the supply, as opposite charges attract. However resistive processes, such as collisions with the lattice ions, cause the electrons to travel around the circuit at a predictable average velocity.



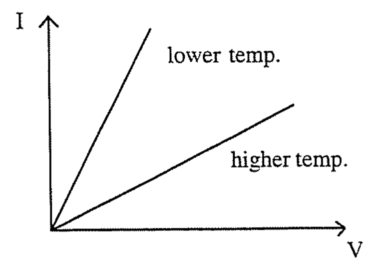
Example: A 20cm length of conductor has a p.d. of 12 volts applied across it. Find the force exerted on each conduction electron by the electric field.

Answer: $E = \dfrac{12}{0.20} = 60Vm^{-1}$

$F = Ee = 60 \times 1.6 \times 10^{-19} = 9.6 \times 10^{-18} N$

Temperature effects in a metal

The number of conduction electrons in the metal is fixed. The temperature has no effect. However increased temperature in the metal means an increase in the amplitude of vibration of the lattice ions. This increases the resistance of the metal as the conduction electrons suffer more collisions, slowing their progress around the circuit.



For a metal, the only significant effect on current flow is through the increased lattice vibration.

Conduction electrons in a semiconductor

At absolute zero, semiconductors are insulators. All electrons would be bound to their atoms. But at room temperature, a small proportion of the most weakly bound (outermost) electrons have enough energy to become free, and are available to carry current.

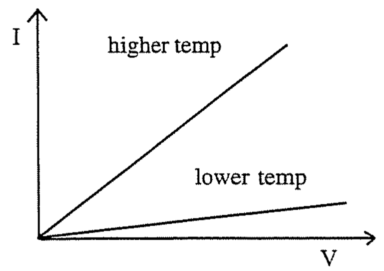
The theory for the flow of electrons through a semiconductor when a p.d. is applied is similar to that for a metal. However the number of available electrons is very much smaller. (We will discuss this when we look at drift velocity.)

When dealing with a metal, we can think in terms of current in amperes. With a semiconductor, we often think in terms of microamps.

Temperature effects in a semiconductor

Once again, increased lattice vibration with rising temperature will cause more collisions and slow down the progress of an individual electron. However, this is a secondary effect in a semiconductor.

The rising temperature gives many more weakly bound electrons enough energy to escape from their atoms. They are now available to carry current. This effect far outweighs the lattice vibration effect, and the resistance of the semiconductor decreases quickly with increasing temperature:



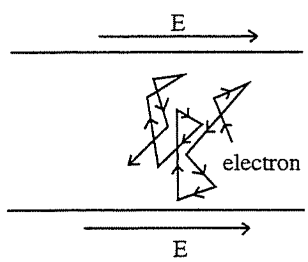
The current flow itself causes a heating effect, causing more electrons to become free, leading to even greater current flow. This effect is made use of in a thermistor.

> *For a semiconductor, the most important effect of increased temperature is the increase in charge carriers. Increased lattice vibration is relatively unimportant.*

Thermal velocity and drift velocity

So far, we have only talked about electron movement along the line of the applied p.d.

However the conduction electrons have thermal energy of their own (depending on the ambient temperature) and are in continuous random motion whether there is applied voltage or not. This is called their **thermal velocity**. The organised motion related to the current flow is called **drift velocity**.



In fact the thermal velocity is far greater than the drift velocity. The thermal velocity may be in the order of 1Mms^{-1} while the drift velocity might be 1mms^{-1}.

> **Example: Find the ratio of thermal velocity to drift velocity for the values given.**
>
> **Answer:** $\dfrac{\textit{thermal velocity}}{\textit{drift velocity}} = \dfrac{10^6}{10^{-3}} = 10^9$

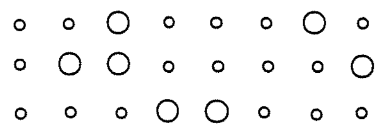
> The equation for calculating the drift velocity is: $I = nAve$
> where n is the number of conduction electrons per m^3, A is the cross-sectional area of the conductor, v is the drift velocity, and e is the charge on an electron.

> **Example: For an identical current flow in a metal and a semiconductor of identical dimensions, where would you expect the greatest drift velocity?**
>
> *Answer: Rearranging the equation:* $v = \dfrac{I}{nAe}$
>
> *This tells us that the metal (with a much greater **number density** of conduction electrons, **n**) will have a much smaller drift velocity than the semiconductor. This makes sense, as to get the same current flow with far fewer charge carriers, the carriers must be moving through the material much more quickly.*
>
> *Or conversely, if the drift velocity were the same, we would expect the current flow to be much less in the semiconductor.*

Numerical calculations on drift velocity appear in the questions at the end.

Alloys and Impurities

The progress of conduction electrons through a lattice is impeded by the collisions which occur with the lattice ions. Impurity atoms in the lattice cause irregularities which increase the collisions. The result is reduced current flow.



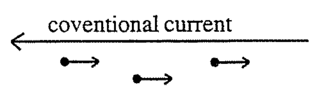
impurity atoms in metal lattice

Reduced current flow implies a greater resistance for the material. So pure metals tend to be the best conductors. And impure metals and alloys tend to have greater resistances. Pure copper is a good conductor. Brass (a copper and zinc alloy) is far less conductive.

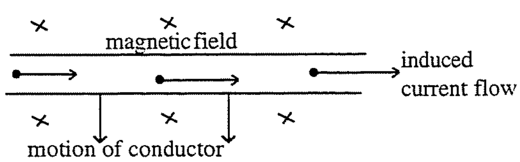
Induced Currents

So far we have assumed that all current flow is due to applied voltage. But electromagnetic induction is caused by the effect of a magnetic field on a moving charge carrier. The charge carrier (in this factsheet an electron) must have a component of its motion perpendicular to the applied magnetic field. The force on the moving electron will be at right angles to the plane of its motion and the magnetic field. Fleming's Right Hand Rule predicts the direction of the induced current.

> ***Exam Hint:*** *Fleming's rules refer to current flow. Remember that conventional current flow is defined by the motion of a positive charge. So electrons moving to the **right** are considered to be a current flowing to the **left**.*

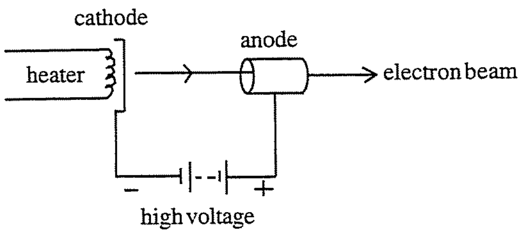


In a generator, as the coil is turned by the turbine, the electrons in the coil are moved across an applied magnetic field. The magnetic force on these electrons causes them to move along the conductor. An induced current flows around the coil.



Electron beams

In a cathode ray tube, electrons freed from the cathode (by thermionic emission) are accelerated into a fluorescent screen by an applied voltage. The vacuum in the tube means that their progress is not impeded. The energy of each electron reaching the screen is: $E = eV$



Example: Suppose the light energy emitted by a television screen is 5 watts, and that the accelerating p.d. is 5000 volts. Let us assume the process is 100% efficient. Find the number of electrons reaching the screen each second.

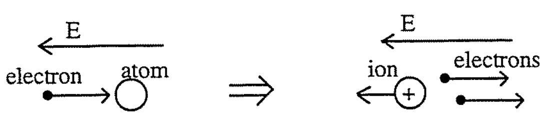
Answer: The energy of each electron will be
$E = eV = 1.6 \times 10^{-19} \times 5000 = 8.0 \times 10^{-16}$ *joules.*
So the number of electrons required to provide an energy of $5Js^{-1}$ *is* $\frac{5}{8.0 \times 10^{-16}}$ *or* 6.25×10^{15} *electrons.*

A very high energy electron beam is used in an X-ray tube to cause the target metal to emit high energy e.m. radiation (the X-rays).

Gas discharge tubes

Tube such as strip lights work because electrons flowing between the electrodes in the tube collide with atoms of a low pressure gas (e.g. neon, argon, mercury) and excite the atoms. When the atoms of the gas drop back to their ground state, they emit light.

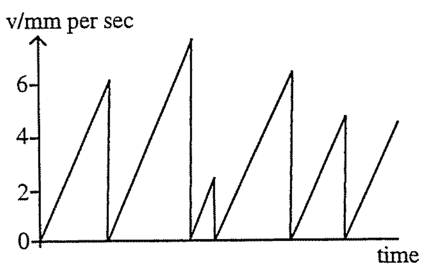
In some tubes, the electrons are accelerated by such high voltages that they actually ionise gas atoms. The positive ions and negative electrons then flow in opposite directions, colliding with and exciting other gas atoms along the way. The current in these tubes is carried by both the ions and the electrons.



Before and after collision

Practice Questions

1. Here is a graph of drift speed of an electron as it travels through a metal lattice.



   (a) Estimate the average drift speed of this electron (and all of the others).
   (b) Sketch a possible graph of a similar electron if the metal is heated.
   (c) Explain how a similar graph for an electron in a semiconductor would be different.

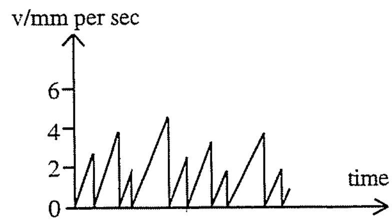
2. A copper wire in a circuit has a cross-sectional radius of 1.0mm and a length of 1.5 metres. The current flow is 200mA. The number density of copper is $8 \times 10^{28} m^{-3}$.
   (a) Find the average drift velocity of an electron through the copper wire.
   (b) Find the length of time it takes an electron to travel the length of the wire.
   (c) Find the number of conduction electrons in motion in the wire at any time.

3. Two wires are both made of copper. Wire A is twice as long as wire B and has double the radius of B. Both wires carry the same current. Find the ratio of the electron drift velocities in both wires.

4. In a cathode ray tube, the electron beam is accelerated through 2000 volts. The power reaching the screen is 1.5 watts.
   (a) Find the energy given to each electron.
   (b) Find the number of electrons reaching the screen every second.
   (c) Use two different methods to find the current flow.

Answers

1. (a) The average drift speed might be estimated at $3mms^{-1}$.
   (b) If the metal is heated, the lattice ions will vibrate more energetically causing more frequent collisions with the conduction electrons and a lower drift speed:



   (c) The graphs might be very similar to those for the metal. The difference with the semiconductor is the tiny number of conduction electrons, and the increase in their number with increasing temperature.

2. (a) $v = \frac{1}{nAe} = \frac{0.200}{8 \times 10^{28} \times 3.14 \times 10^{-6} \times 1.6 \times 10^{-19}} = 5.0 \times 10^{-6}\ ms^{-1}$

   (b) $t = \frac{s}{v} = \frac{1.5}{5.0 \times 10^{-6}} = 3.0 \times 10^{5}s$

   (c) no. = volume × number density = $A \times l \times n$
   $= 3.14 \times 10^{-6} \times 1.5 \times 8 \times 10^{28} = 3.8 \times 10^{23}$

3. The length of the wire is not relevant, as it does not appear in the equation for drift velocity. As I, n, and e are the same for both situations, we can write: $v \propto \frac{1}{A}$

If A has double the radius of B, then it has 4 times the area. So the ratio of the drift velocity of A:B is 1:4.

4.(a) $E = QV = eV = 1.6 \times 10^{-19} \times 2000 = 3.2 \times 10^{-16}$ J

(b) number = $\frac{\text{total power}}{\text{energy of each electron}} = \frac{1.5}{3.2 \times 10^{-16}} = 4.7 \times 10^{15}$

(c) (i) $I = \frac{P}{V} = \frac{1.5}{2000} = 7.5 \times 10^{-4}$ A

(ii) $I = \frac{Q}{t} = \frac{4.7 \times 10^{15} \times 1.6 \times 10^{-19}}{1} = 7.5 \times 10^{-4}$

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU
Physics Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber.

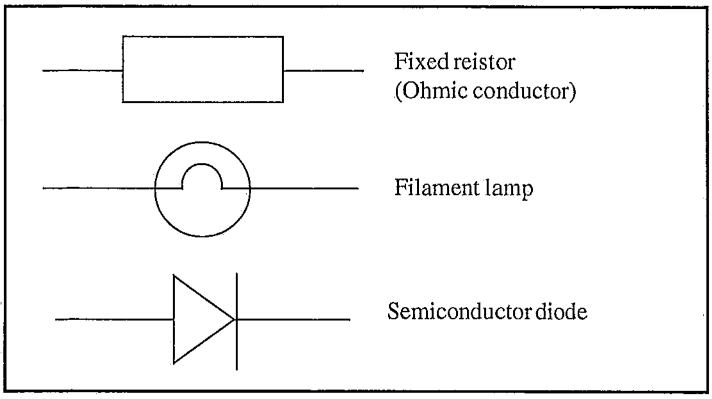
ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 136

Interpreting and Using V-I Graphs

A V-I characteristic is a graph of how the current through a component varies with the potential difference across it. You need to recall and explain the shape of three V-I characteristics: filament lamp (light-bulb), semiconductor diode and a conductor that obeys Ohm's law.



A V-I characteristic is graph of current against potential difference for a component.

Exam Hint: *You will need to recall and explain V-I characteristics for filament lamps, semiconductor diodes and Ohmic conductors.*

Ohmic conductor

An Ohmic conductor obeys Ohm's law. In exam questions, Ohmic conductors are typically fixed resistors or simply a piece of wire in a circuit.

Ohm's law states that the current passing through a conductor is proportional to the potential difference across the conductor: $I \propto V$

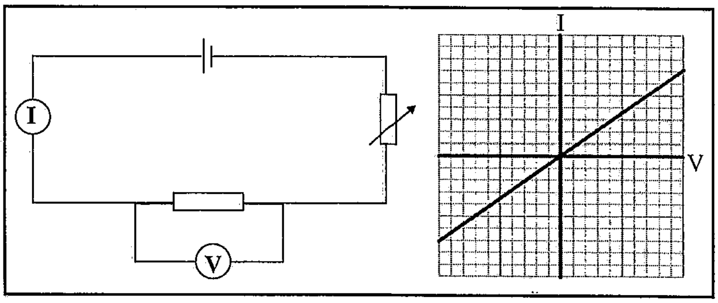
This can only be true if the resistance of the component does NOT vary. As a conductor heats up, it's resistance increases. So a conductor that obeys Ohm's law must stay at the same temperature.

Exam Hint: *An Ohmic conductor must remain at constant temperature for the resistance to remain constant.*

A circuit that could be used to measure the V-I characteristic for a component is shown below. The current through the fixed resistor is measured for a given potential difference.

Then the variable resistor is altered and the current is measured again, producing the V-I characteristic for an Ohmic conductor (Fig 1).

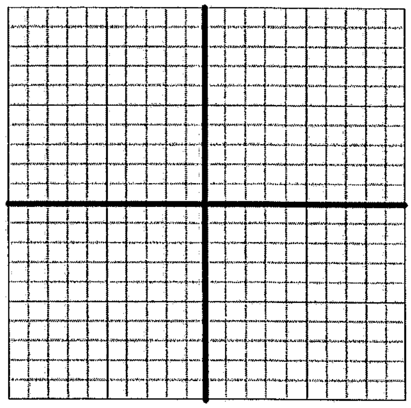
Fig 1. Ohmic conductor



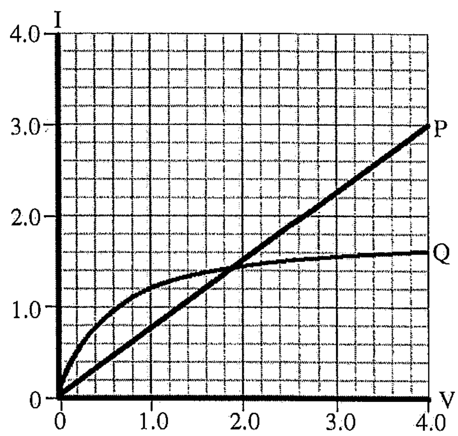
The V-I characteristic for an Ohmic conductor (like a fixed resistor or piece of wire) is a straight line through the origin

Worked example 1

(a) Draw the V-I characteristic for two Ohmic conductors, X and Y, on the axes. Y has a lower resistance than X. *(2 marks)*



(b) State and explain which component, P or Q, is an Ohmic conductor. *(2 marks)*

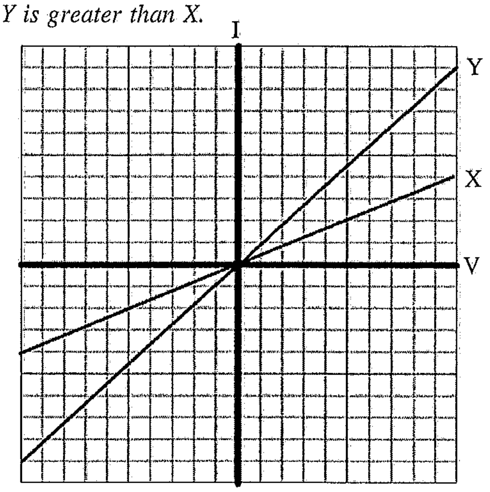


(c) (i) Calculate the resistance of component Q at 0.5V and 4.0V. (2 marks)
(ii) Explain the differences in your resistance values.(2 marks)
(iii)Calculate the resistance of component P. (1 mark)

Worked example 1

Answers

(a) Both X and Y are straight lines through the origin. Gradient of Y is greater than X.



Explanation: *As the resistance of Y is lower than X, more current can pass through Y for the same potential difference.*

(b) Component P. The resistance does not vary.

Explanation: *The graph of an Ohmic conductor is a straight line through the origin. This is because the current through an Ohmic conductor is directly proportional to the potential difference across it.*

(c) (i) For Q, when V=0.5V, I = 0.9A   V/I=R 0.5V/0.9A = 0.6Ω
when V=4V, I = 1.6A V/I=R 4V/1.6A = 2.5Ω

(ii) The bulb filament is hotter at 4V than 0.5V. The metal ions vibrate more rapidly in their lattice at higher temperature and collide more often with conduction electrons, leading to higher resistance.

(iii) V/I=R   4V/3A = 1.3Ω

Filament lamp

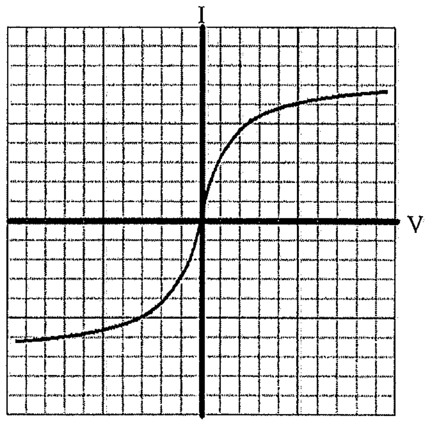
An Ohmic conductor should remain at constant temperature.

What about a conductor that changes temperature?

A filament lamp consists of a very thin wire which becomes hot as current passes through it. It becomes so hot that it glows, giving off light (and significant wasted heat) Fig 2.

The resistance of a bulb increases with temperature and voltage.

Fig 2. Filament lamp

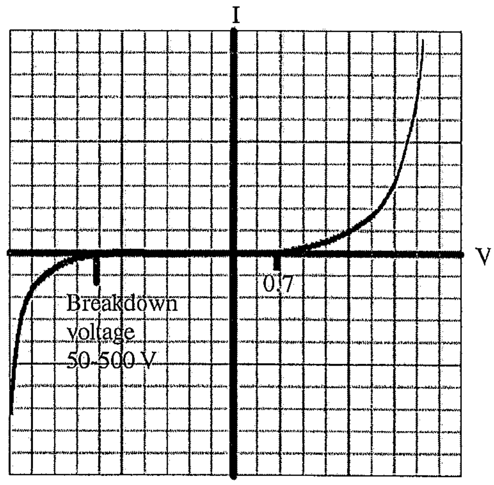


The metal filament, usually tungsten, reaches about 2000°C. This causes a ten-fold increase in resistance compared to lower current. A metal consists of a lattice of vibrating positive ions, surrounded by a "sea" of electrons free to move. As a conductor gets warmer, the positive ions vibrate more, making collisions with conduction electrons more likely, increasing the resistance.

Semiconductor diodes

A diode allows current to pass in only one direction when a potential difference is applied across it (forward bias). If potential difference is applied in the opposite direction, no current flows (reverse bias).

Fig 3. Semiconductor diode



A diode allows current to pass in one direction when a p.d. is applied. If the reverse p.d. is applied, little or no current flows.

A diode does not conduct until a threshold p.d. is passed, usually about 0.6-0.7V.

Exam Hint: *Remember that a semiconductor diode does **not** conduct if less than about 0.7V is applied across it.*

If a much larger reverse bias p.d. is applied, the diode will eventually break down and conduct. This breakdown p.d. could be from 50-500V depending on the individual component.

*A diode has **very high** resistance when an applied forward bias potential difference is below 0.6V or if a reverse bias potential difference is applied.*

Worked example 3

(a) Explain the term 'reverse bias' for a semiconductor diode (2 marks)

(b) (i) State how the resistance of a semiconductor varies with applied potential difference in both directions (3 marks)

(ii) Draw the V-I characteristic for a diode when a reverse bias potential difference is applied.

Answers

(a) Reverse bias occurs when the negative diode terminal (cathode) is connected to a positive potential and the positive diode terminal (anode) is connected to a negative potential.

(b) (i) Forward bias: resistance is very high until a potential difference of 0.7V is reached, when it becomes very low. Reverse bias: resistance is very high until break down potential difference of 50-500V is reached.

(ii) See Fig 3 page 2.

Practice Questions

1. (a) Draw the V-I characteristic for a semiconductor diode. Label important voltage values. Include forward and reverse bias (4 marks).

   (b) Explain the shape of the V-I characteristic for a filament lamp (4 marks).

   (c) (i) When the current passing through the lamp is 110mA, the power is 25W. Calculate the resistance of the lamp. (2 marks)

   (ii) Calculate the potential difference across the lamp at this current and label the graph accordingly (1 mark).

2. (a) Two components, X and Y, obey Ohm's law. If Y has a lower resistance than X, draw their appropriate V-I characteristics. (2 marks)

   (b) Draw the V-I characteristic for a semiconductor diode and add approximate labels to the voltage axis. Explain how the resistance of this diode varies with the applied voltage. (7 marks)

Answers

1 (c) (i) 2066 Ohms
(ii) 227V

Acknowledgements:
This Physics Factsheet was researched and written by J Carter
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

September 2001 Number 23

Circuit Electricity I

In this Factsheet we will:
- **briefly explain what current, voltage and resistance are;**
- **investigate how and why electrical circuits function;**
- **look at examples of questions on electrical circuits.**

Vital to an understanding of electrical circuits is a basic knowledge of what current, voltage and resistance are. A more full explanation can be found in Factsheet No. 07 - Electrical Current, Voltage and Resistance.

Introduction

An **electrical current** is a net movement or flow of charge in a given direction. In a metal, this charge is negative and is due to the movement of free electrons within the structure of the metal. You can imagine current as behaving in a similar way to cars in a nose-to-tail traffic queue or as molecules in an incompressible fluid. Cars at the rear of a traffic queue can only move forwards if the cars at the front do - otherwise they have no space to move in to. In a similar way you cannot depress the plunger of a water-filled, plastic syringe if the other end is blocked - the water molecules have nowhere to go. For this reason it is necessary to have a complete circuit loop, with no breaks, for a current to flow. We will examine circuit loops in more detail later on. This also means that current will be the same at any point around a series circuit - again more later on.

Exam Hint: *- Remember conventional current (the direction we assume the charge travels in a current) is in the opposite sense to the direction of electron travel (the direction they actually travel).* ***Conventional current flows from positive to negative, electrons travel from negative to positive.*** *You should always mark conventional current on diagrams unless specifically asked to do otherwise.*

- A v**oltage** is a measure of how much potential energy a unit charge has at a point, specifically here in an electric circuit.
- A **potential difference** exists between two points if a charge has a differing value of potential energy at each of the points. A **voltage drop** means the charge has lost energy; an **emf** (electromotive force) means the charge has been supplied with energy.

The easiest type of p.d. to understand is that produced by a cell – which contains a surplus of electrons in the negative terminal and a lack in the positive terminal. Electrons in the negative side of the terminal are repelled by the like charge surrounding them and pushed out into the circuit to fill the gaps left by electrons drawn towards and into the positive terminal. Just as a mass falling to earth loses gravitational potential energy, electrons moving towards a positive terminal lose electrical potential energy, and it is this eletrical potential energy that is transferred into heat.

Resistance in a metal is due to the presence of fixed, positive ions which inhibit the flow of electrons through the metal. As electrons collide with the positive ions, the electrons give up energy and it is transferred to the structure of the metal, normally resulting in an increase in temperature, which causes heat to be transferred to the surroundings. Fundamentally this is how an electric heater and filament lamp function.

In brief, an electromotive force causes electrons to move around a complete circuit giving up energy as they progress. The current transfers energy from the cell to components in the circuit.

Kirchoff's Laws

As already mentioned, current in a wire is composed of electrons. When this flow comes to a junction, unless they leak out of the wire (which they don't) the number of electrons that flows in must equal the number which flows out. This is **Kirchoff's First Law.**

Kirchoff's First Law states that the sum of currents flowing into a junction must equal the sum of currents flowing out of a junction. It can be written using the shorthand $\Sigma I = 0$ and it is fundamentally a statement of conservation of charge.

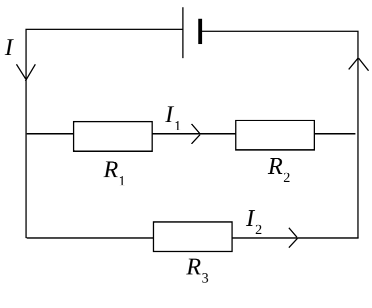
As you already know energy is always conserved.
- An emf is a measure of energy transformed into electrical energy per unit charge;
- voltage drop is a measure of the energy transformed from electrical energy to other forms per unit charge.

Bearing in mind conservation of energy it stands to reason that the total energy supplied to each electron will equal the amount of energy lost. In other works the sum of emfs must equal the sum of voltage drops.

Kirchoff's Second Law states that the sum of the potential differences across components in any complete loop around the circuit must equal the sum of the electromotive forces supplying it. In other words $\Sigma E = IR$. This is fundamentally a statement of the conservation of energy.

Remember - *choose a direction around the circuit. If the emf opposes you then it must be given a negative sign.*

For example:



From K1: $I = I_1 + I_2$

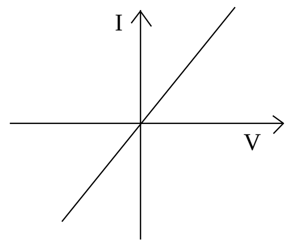
From K2: $E = V_1 + V_2$ and also from the other loop $E = V_3$.

Even if you do not meet Kirchoff's laws by name you will us them implicitly in circuit calculations, as you will see later in this Factsheet.

Different components respond differently to different voltages and currents.

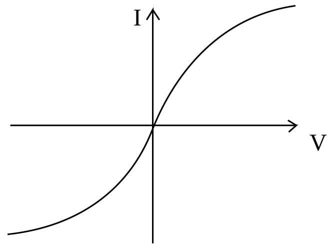
I-V characteristics

1. Current-carrying wire at a constant temperature.



For this component I ∝ V. When a component behaves in this way it is said to obey Ohm's law. **Ohm's law states that the current through an ohmic conductor is directly proportional to the voltage across it provided its temperature is constant.** The resistance is constant and is given by the inverse of the gradient (R = 1/m). If we double the voltage, the current will also double.

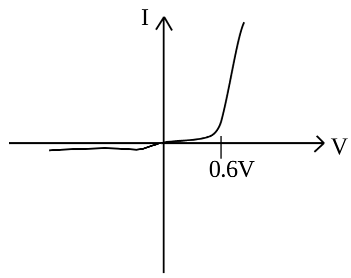
2. Filament lamp.



Here, resistance is not constant. As current goes up, V/I increases and the resistance therefore gets bigger. This is because filaments are very narrow and so heat up appreciably as current increases.

We can still apply $V = IR$ but only at a given voltage and current - i.e. if we double our value of voltage the current will increase by less than double because the filament's resistance has increased. So if the resistance of a filament bulb is 6Ω at 12V it will be less than this at a lower voltage; it will <u>not</u> be constant.

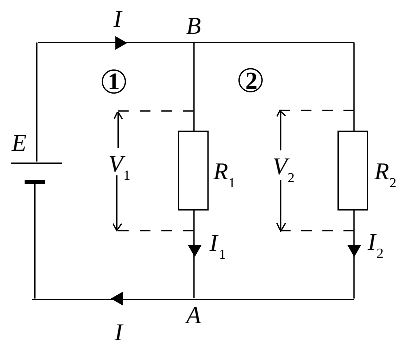
3. Diode



A diode only lets current through when it is flowing in one direction. When the diode is facing the direction of current flow it is said to be in **forward bias**. Even in forward bias it will not conduct until a voltage of around 0.6V is across it; this is called the "switch on" voltage. After the switch-on voltage is reached, resistance decreases as current rises. When the diode opposes the attempted direction of current flow then it is in **reverse bias** and has a very high (effectively infinite) resistance, allowing no current to flow.

We are now in a position to tackle questions based on electrical circuits. Firstly we will discuss different types of circuit and apply what we covered so far. The rest of this Factsheet is merely application of what we have learnt so far.

Parallel circuits



The emf present in the circuit supplies each charge with the same amount of energy. Using conventional current, charges flow through the circuit until they reach B where they choose loop ① or ②. They then proceed to travel through R_1 or R_2 and meet at point A to travel to the negative terminal of the cell.

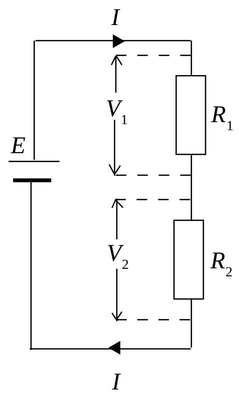
If a charge follows loop ① then it must lose an equal amount of energy in R_1 as it gained from the cell. Therefore $V_1 = E$. We can justify this by applying Kirchoff 's Second Law.

If a charge follows loop ② then the only place it can lose the energy it gained is R_2 therefore $V_2 = E$ – again a consequence of Kirchoff 's Second Law.

As charge does not disappear, the current travelling from the cell will be shared between the two pathways, so $I = I_1 + I_2$. This is justifiable using Kirchoff's First Law.

Parallel components have the **same p.d.** across them but **shared current** through them.

Series circuits

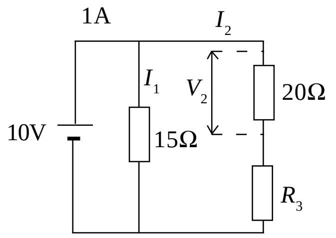


The current through both components will be the same, as every charge only has one pathway to follow. We can justify this using Kirchoff's First Law.

As charges travel through both resistances then they will give up energy. As they must give up the same amount of energy the cell gave them, $E = V_1 + V_2$. If we consider the circuit loop, we can apply Kirchoff's Second Law as justification.

Series components have the **same current** through them but **shared** p.d. across them.

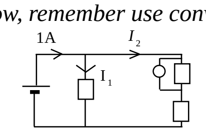
Typical Exam Question



a) Mark on the diagram above the direction of all the labelled currents. [1]

b) Using the numerical values given find values for I_1, V_2 and R_3. [5]

a) Answer marked below, remember use conventional current positive to negative. ✓



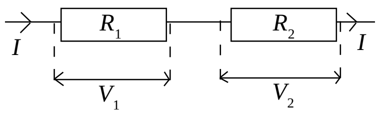
b) I_1*: Using V = IR*
$I_1 = 10/15 = 0.67A$ ✓
V_2*: Firstly we need to find the current and then we can apply V=IR.*
We find I_2 *using Kirchoff's first law:* $I = I_1 + I_2$.
So $I_2 = I - I_1 = 1 - 0.67 = 0.33A$ ✓
Now we can apply V=IR
$V_2 = 0.33 \times 20 = 6.6V$ ✓
R_3*: Now we need to find the voltage across* R_3 *using Kirchoff's second law.*
$V_3 = 10 - 6.6 = 3.4V$ ✓
And again using V = IR
$R_3 = 3.4/0.33 \approx 10\Omega$ ✓

Exam hint: *A lot of students realise they have to use V =IR but then apply it wrongly. E.g. in the above question we have been very careful to use the value of V and I that applies to the component we are dealing with. When calculating* R_3 *we made sure we were using the voltage across* R_3 *(3.3V not 10V) only and the current through it (0.33A not 1A).*

Total effective resistance

When components with resistance are placed in series or parallel then we can derive a formula to find their total effective resistance. In other words the value of the one resistor that could replace them to give the same resistance as they are both giving together.

Resistance in series



The total p.d. is shared so that: $V = V_1 + V_2$ (Kirchoff 2).

Using $V = IR$ and I is the same in both $V = IR_1 + IR_2$ (Kirchoff 1).

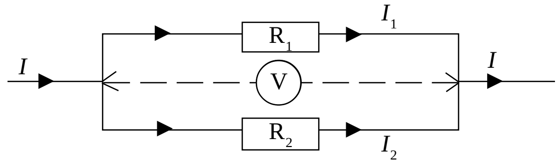
Then $V = IR_{eff}$ where R_{eff} is the total effective resistance of the two resistors.
This gives: $IR_{eff} = IR_1 + IR_2$
Cancelling the I's gives $\boldsymbol{R_{eff} = R_1 + R_2}$

This tells us that even a very small resistance placed in series with another resistance increases the circuit resistance. The current now has to travel through the original resistor and a new resistance and so finds it harder to flow. **A resistor in series always increases the effective resistance.**

Ammeters are always connected in series with the component we are interested in, so that the same current passes through the ammeter and the component. If they have a resistance then they will decrease the amount of current flowing through the component which will change the curcuit.

It is important ammeters have a low resistance and in the ideal case they should have zero resistance.

Resistance in parallel



We know that the current is shared by both resistors.

$I = I_1 + I_2$ (Kirchoff 1).

Using $I = V/R$ we get $I = V/R_1 + V/R_2$.

The total current = V/R_{eff} where again R_{eff} is the effective resistance of the two resistors. So using this; $V/R_{eff} = V/R_1 + V/R_2$

Cancelling the V's, as they are the all equal (Kirchoff 2)

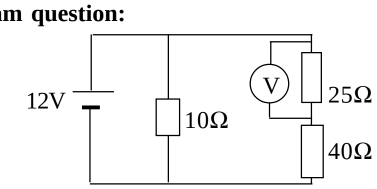
$$\frac{1}{R_{eff}} = \frac{1}{R_1} + \frac{1}{R_2}$$

Now, even if we substitute a very large resistance into R_2 we still find the overall effect is to decrease the circuit resistance. This is because the current not only has the path through R_1 it had before but also a second path through R_2, therefore the resistance to its passage is reduced. **A resistor in parallel always decreases the effective resistance.**

Voltmeters are always connected in parallel. This means they will decrease the effective resistance of the component they are measuring and therefore lower the voltage across it. **It is necessary to use a voltmeter with a high resistance relative to the component it is across to gain accurate readings. Ideally a voltmeter should have an infinite resistance - this is effectively the case with a cathode ray oscilloscope.**

You may not need to learn these derivations but you should certainly understand them. These equations can be extended to more resistors.

Typical exam question:



Given the voltmeter shown in the above diagram can be taken to have an infinite resistance find:
a) the total resistance of the circuit. [3]
b) the current drawn from the cell. [2]
c) the reading on the voltmeter. [4]

*a) First work out series combination, then combine with 10Ω resistance. Work in this order as the 10Ω is in parallel with **both** 25Ω **and** 40Ω*
. $R_{eff} = R_1 + R_2 = 25 + 40 = 65\Omega$ ✓
Parallel combination:

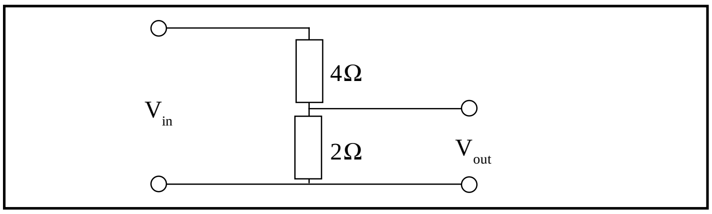
$$\frac{1}{R_{eff}} = \frac{1}{R_1} + \frac{1}{R_2} \Rightarrow \frac{1}{R_{eff}} = \frac{1}{65}✓ + \frac{1}{10} \Rightarrow R_{eff} = 8.7\Omega ✓$$

b) $I = V/R_{total} = 12/8.7$ ✓ $= 1.38A$ ✓
c) The current through the 25Ω resistor is the same at any point along that branch of the circuit. $I = 12/65$ ✓ $= 0.18A$ ✓
Now we have the current through the 25Ω resistor and using its resistance. $V = IR = 0.18 \times 25$ ✓ $= 4.5V$ ✓

Potential dividers

We looked at series circuits earlier and saw how their components share the voltage applied to the circuit. One device which makes use of this is the **potential divider**. (Fig 1)

Fig 1. Potential divider



Potential dividers can supply a p.d. of any value up to the value of the supply p.d. by varying the size and arrangement of the resistors. This means we can tap off different p.d.'s by placing our components across one of the resistors being supplied by the fixed source.

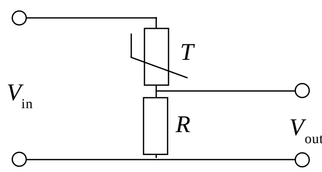
Potential dividers make use of the fact that the amount of p.d. across resistances in series is proportional to their resistance; $V \propto R$ due to Ohm's law. *I* is constant because the resistances are in series and therefore current is the same in both. So if we know the proportion one of the resistances is relative to the resistance of the whole circuit then we can work out the fraction of the p.d. it will take. This is easier to grasp if you consider the circuit above: The potential difference is shared across the 2Ω and 4Ω resistors. The total resistance is 6Ω so the 4Ω resistor will take 4/6 or 2/3 of the supply voltage (V_{in}), and the 2Ω will take 2/6 or 1/3 of V_{in}.

If we reconsider the previous exam question, we were asked to find the p.d. across a 25Ω resistor when it was connected in a series combination with a 40Ω resistor, the combination having 12V across it.

Now we can say the resistor is 25/65 = 5/13 of the total resistance and so will take $5/13 \times 12V = 4.62V$ - in agreement with our previous answer which had been rounded. Using this approach is not compulsory but it can speed up calculations or allow you to check your answers.

Thermistors

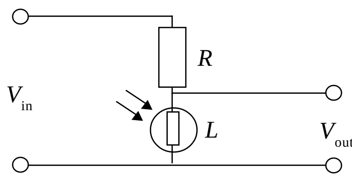
A thermistor can be used in a potential divider to provide a device that responds to temperature variation.



As temperature increases then the resistance of *T* falls. This means *R* takes a greater proportion of the voltage drop. If a fan were connected to V_{out} as the temperature increased, it would receive a greater and greater voltage.

Light Dependent Resistor

In the LDR circuit below as the light levels increase then the resistance of the LDR decreases and so does the p.d. across it.

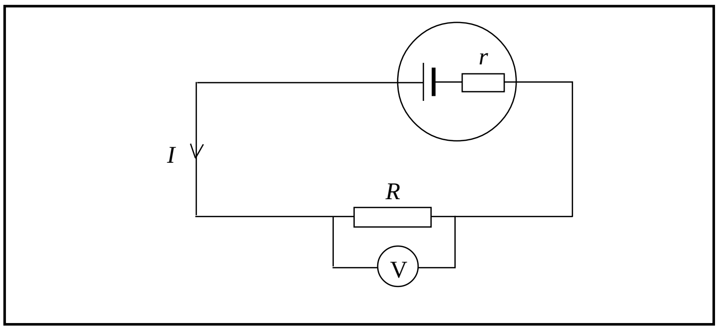


If a lamp were placed across V_{out} then it would be illuminated when dark and switch off when light.

Internal resistance

A cell itself has some resistance - known as **internal resistance**. A electron leaving the negative terminal will be involved in collisions within the cell itself and will encounter resistance. Some of the energy that the emf supplies to the moving charge is used in overcoming the internal resistance rather than in the rest of the circuit itself and so the circuit receives a lower voltage than it would do if there was no internal resistance. The voltage in the external circuit is less than the emf of the battery and as the current gets bigger, this effect becomes more pronounced. The energy dissipated in the cell due to the internal resistance is what makes it become hot when in use. We represent the internal resistance with a resistor in series with the battery labelled r (Fig 2).

Fig 2. Internal resistance



Obviously we can not get inside the cell to measure r directly so we have to use other means of finding how big it is.

Notice firstly that the current will be the same in both resistances as they are in series. Also, by Kirchoff's second law, the emf will be shared across *r* and *R*. So as $V = IR$ then: $E = IR + Ir$

and $E = I(R + r)$

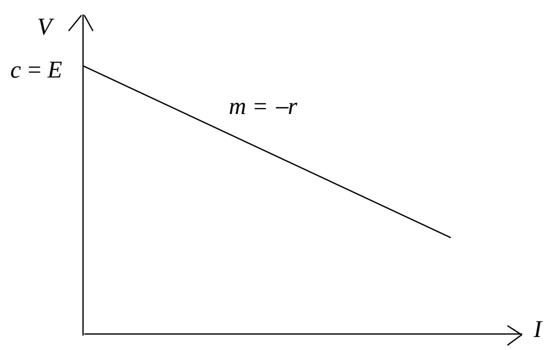
Ir is referred to as the **lost volts** because it is not available to the rest of the circuit. *IR* is the **terminal p.d.** as this is the voltage across the terminals of the cell when a current is being drawn. As *IR* is also the p.d. we measure across *R* we give it the symbol *V*. So we get

$$V = E - Ir.$$

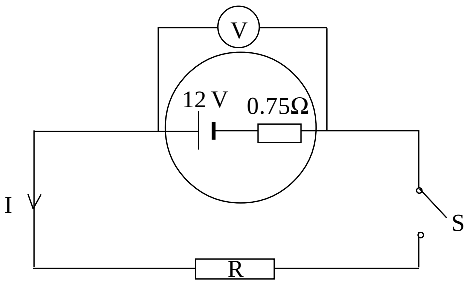
The emf is the p.d. across the terminals of a battery (or cell) when no current is allowed to flow from it.

We could measure the emf directly with an oscilloscope, as they have effectively infinite resistance; failing this we can employ a graphical approach. If we replace *R* with a variable resistor we can obtain a set of values for *V* and *I* using an appropriately placed voltmeter and ammeter. This gives us *V* as our *y* variable and *I* as our *x* variable.

Compared with y = mx + c we get a straight line for a plot of V vs I with a gradient of -r and a y-intercept of E.



Typical Exam Question



With the switch S open the voltmeter, which has a very high resistance, reads 12V. After it is closed the reading falls to 9.5V and then rises shortly afterwards to 10V. Given the 10Ω resistor in the diagram is initially at 0°C answer the following:

a) Why does the voltmeter read 12V with S open? [2]
b) What is the value of R initially? [3]
c) What is the value of the current after a short time [1]
d) What has happened to the value of the R? Explain why. [2]

a) *It reads 12V as this is the emf ✓ and the emf is present across the terminals when no current flows ✓*

b) *To find the value of R we need the current through it, we can find the current through r and, as they are in series, use this to find R.*
$V = 12 - 9.5 = 2.5V$
$I = 2.5/0.75 = 3.33A$ ✓
Now use: $E = I(R + r)$ ✓ $\Rightarrow 12 = 3.33(R + 0.75)$
$R = 2.85\Omega$ ✓

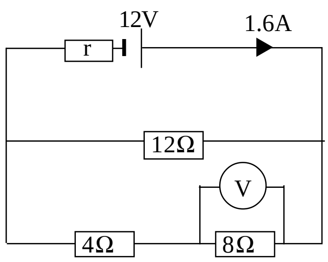
c) $V = 12 - 10 = 2V$
$I = 2/0.75 = 2.67A$ ✓

d) *If I has decreased then R must have increased ✓ this is because it has increased in temperature ✓ as it started at 0°C and the resistance of a resistor increases with temperature.*

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's answer is given below.

A circuit is set up as below. The cell does <u>not</u> have negligible internal resistance. For the first part of the question you can assume that the voltmeter has negligible resistance.



a) (i) What is meant by the internal resistance of a cell? Explain the effect it has on the terminal p.d. of the cell compared to its emf. [2]

The resistance the current inside has to overcome before it leaves the cell. ✓ *It lowers it.* ✗

> Student has not explained the effect although the statement is correct.

a) (ii) What is the value of the internal resistance? [3]

$E = I(R + r)$ ✓
$1/R = 1/12 + 1/4 + 1/8 \therefore R = 2.18\Omega$ ✗
$12 = 1.6(2.18 + r) \quad r = 5.32\Omega$ ✗

> Even though the student's method is essentially correct they have misunderstood the arrangement of the resistors by trying to combine all the resistances in one equation

The voltmeter is now replaced with another one having a resistance of 20Ω.

b) (i) What is the total resistance of the circuit now? [3]

$1/R = 1/20 + 1/8 \qquad R = 5.7\Omega$ ✓
$R = 5.71 + 4 = 9.71\Omega$
$R = 9.71 + 5.32 = 15.03\Omega$ ✗
$1/R = 1/15.03 + 1/12 = 6.67\Omega$ ✗

> The student was correct until s/he mistakenly assumed r was in series with three resistors in the bottom branch of the circuit – not true. In fact r is in series with these three and the 12Ω, so the correct course of action was to use the parallel formula and then add in r. Note the student would not have been penalised for using 5.32Ω. S/he would have received error carried forward marks.

b) (ii) Find the reading on the voltmeter. [4]

$V = IR$
$V = 1.6 \times 5.17 = 9.13V$ ✗✗✗✗

> The student has used the correct resistance but the current is wrong for two reasons. The current splits up and so it does not flow through the combination and, as the voltmeter's resistance has changed then the current will have changed too. Notice that the question is worth four marks – it is highly unlikely an answer this short would gain so many marks.

Examiner's answers

a) (i) *The resistance a current must overcome within the cell* ✓
The terminal p.d. is lower than the emf as the current loses some energy in overcoming the internal resistance ✓

a) (ii) $E = I(R + r)$ ✓
$1/R = 1/12 + 1/(4 + 8) \qquad \therefore R = 6\Omega$ ✓
$12 = 1.6(6 + r) \qquad r = 1.5\ \Omega$ ✓

b) (i) $1/R = 1/20 + 1/8 \qquad R = 5.71\Omega$ ✓ — *For voltmeter and 8Ω*
$R = 5.71 + 4 = 9.71\Omega$ — *Now the 4Ω*
$1/R = 1/9.71 + 1/12 \qquad R = 5.37\Omega$ — *Now 12Ω*
$R = 5.37 + 1.5 = 6.87\ \Omega$ ✓ — *Now r*

b) (ii) *We have divided total voltage by total resistance to find the total current which must all pass through r.*
$I = 12/6.87 = 1.75A\Omega$ ✓
$V = 1.75 \times 1.5 = 2.62\ V$ ✓

Now the voltage across the series combination must be
$V = 12 - 2.62 = 9.38V$

The voltmeter and the resistor it is across have a resistance of 5.71Ω from a total of 5.71 + 4 = 9.71Ω for that branch.
$(5.71/9.71) \times 9.38 = 5.52V$ ✓✓

Questions

1. (a) Explain the following terms below:

   (i) internal resistance; [1]

   (ii) "lost volts"; [1]

   (iii) terminal p.d.; [1]

   (iv) electromotive force; [1]

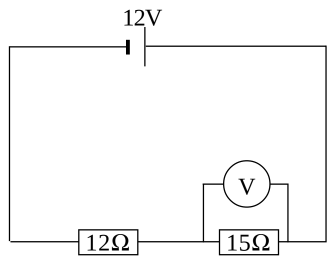
   (b) State an equation that relates emf, currrent and internal resistance [2]

   (c) (i) Under what conditions are the emf and terminal p.d. equal? [1]

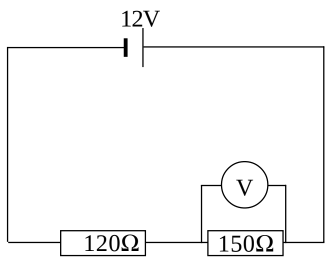
   (ii) Which fundamental laws are Kirchoff's 1st and 2nd laws based upon? [2]

   (d) Two circuits are shown below. The resistance of the voltmeter is 200Ω. You may assume the cell has negligible internal resistance.

A



B



   (i) Find the reading on the voltmeter in **A**. [3]

   (ii) Find the reading on the voltmeter in **B**. [2]

   (iii) What would be the p.d. across the 15Ω resistor in **A** and the 150Ω resistor in **B** if the voltmeter were not present? [2]

   (iv) Explain why the voltmeter is much more appropriate to use in circuit **A** compared with circuit **B**. [2]

Answers

1. (a) (i) The internal resistance is the opposition to current flow charge experiences inside a cell. ✓

   (ii) The "lost volts" are the voltage across the internal resistance. (So called because as they are not available to the circuit). ✓

   (iii) Terminal p.d. is the p.d. across the part of the circuit external to the cell. ✓

   (iv) Emf is the total voltage the cell supplies to the external circuit and the internal resistance. ✓

   (b) $E = I(R + r)$ ✓ where E = emf, I =current, R = external resistance and r = internal resistance ✓

   (c) (i) The emf and terminal p.d. are equal when no current is drawn from the cell. ✓

   (ii) Kirchoff's 1st Law: conservation of charge. ✓
   Kirchoff's 2nd Law: conservation of energy. ✓

   (d) (i) 1/R = 1/15 + 1/200 R = 13.95Ω ✓
   (13.95/(13.95 + 12)) × 12 ✓ = 6.45V ✓
   For V and 15Ω using potential divider idea

   (ii) 1/R = 1/150 + 1/200
   R=85.7 Ω ✓
   (85.7/ (85.7 + 120) x 12) = 5V ✓

   (iii) (15/(12 + 15)) × 12 = 6.67V ✓ for *A*
   (150/(120 + 150)) × 12 = 6.67 ✓ for *B*

   (iv) The answer for circuit *A* is much closer to 6.67 than that of circuit *B* because the voltmeter resistance is significantly larger than 12Ω or 15Ω. ✓
   In *B* the voltmeter resistance is comparable with 120Ω or 120Ω or 150Ω and will give an inaccurate reading. ✓

Acknowledgements:
This Physics Factsheet was researched and written by Alan Brooks
The Curriculum Press,Unit 305B, The Big Peg,120 Vyse Street, Birmingham, B18 6NF

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 68

Energy in Electric Circuits

The purpose of this Factsheet is to bring together ideas of energy transfer in electric circuits. Before studying the Factsheet you should make sure that you are familiar with the ideas of types of energy and ideas of current, potential difference, and resistance from your GCSE course.

You will also find the following Factsheets useful to consolidate some ideas:
Factsheet 5: for ideas of work energy and power.
Factsheet 7: for ideas of current, voltage and resistance.
Factsheet 23: for ideas of currents in circuits, internal resistance and Kirchhoff's Laws.
Factsheet 31: which defines working and heating, for thermodynamics considerations.

In questions at AS Level, you are likely to find that you are required to use ideas of energy transfer combined with other ideas such as internal resistance. In the synoptic paper at A2, you are likely to find the ideas combined with the concept of heating for thermodynamics questions.

Energy transfer in the cell.
At GCSE level you learned that in a cell, chemical energy is transformed into electrical energy. At A level you should recognise this process as working, as opposed to heating, because it is an ordered process and not due to a temperature difference.

The e.m.f., ε, of the cell is defined to be the energy transformed in moving a unit charge across the cell between the plates. So the e.m.f, ε (in Volts) is the work done (in Joules) per unit charge (in coulombs).

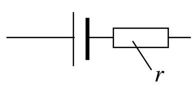
$$\varepsilon = \frac{\Delta W}{q}$$

where: ε = *e.m.f of the cell (V)*
ΔW = *work done (J)*
q = *unit charge (Q)*

Worked Example
1. (a) What is the energy change when a charge of 5μC is transferred across a cell of e.m.f. 3V?
Energy change = e.m.f × charge = $3 \times 5 \times 10^{-6} = 15\mu$ *J*

(b) 9μJ of work is done when a charge of 6μC is moved across a cell. Wwhat is its e.m.f?
$$e.m.f = \frac{9 \times 10^{-6}\ J}{6 \times 10^{-6} C} = \frac{1.5J}{C} = 1.5V$$

If the e.m.f drives a current, I, then I coulombs of charge are moved across the cell per second, and if the current continues for t seconds, then the work done = $\varepsilon \times I \times t$. This is the chemical energy transformed, however, the cell offers some resistance to the flow of charge, so not all of the energy is transformed into electrical energy. The potential difference appearing across the terminals of the cell (the terminal P.D. V) is less than the e.m.f. and the difference is described as the "lost volts". The cell is usually drawn as if it had a resistor in series with it, though the resistance is actually within the body of the cell. This resistance is described as the "internal resistance" (r) of the cell.



The "lost volts" depends on the current drawn from the cell, since the lost volts will be $I \times r$. This leads to the equation:

$$V = \varepsilon - Ir$$

where V = terminal p.d (Volts)
ε = e.m.f of cell (Volts)
I = curremt (amps)
r = internal resistance (Ω)

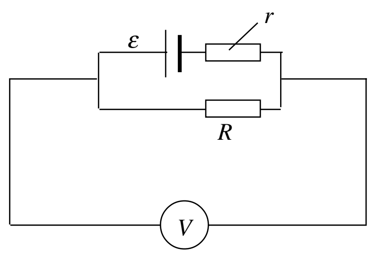
Energy transfer in a load resistor
The P.D, V across a resistor gives the energy transfer per unit charge. The current, I, through the resistor gives the number of coulombs of charge per s, so $V \times I$ gives the energy change per second. – the **power**. Combining this with Ohm's Law $V = I \times R$ gives $P = I \times R \times R = I^2R$. This is the energy per s dissipated as heat in the resistor.

$$P = I \times R \times R = I^2R$$

Energy wasted by the internal resistance
Applying these ideas to the internal resistance gives the energy per second wasted by the internal resistance as I^2r.

Energy wasted per second by the internal resistance = I^2r

Energy transfered in the circuit



In the circuit shown, an external resistor, R is connected across a cell of e.m.f., ε and internal resistance, r. The voltmeter measures the P.D. across the resistor, which is also the terminal P.D. of the cell.

The power dissipated in the resistor (energy per second) is $V \times I$. If I is large, V will be small, because drawing more current from the cell increases the lost volts. For V to be large, I will be small, so to get maximum power from the resistor, a compromise is needed. The value of R to give maximum power in the load resistor can be calculated. Students also doing A Level maths might like to do this calculation by getting an expression for the power ($V \times I$) in the load in terms of ε, R and r, differentiating it, and setting the differential to zero. [Don't worry if you have not done enough maths to do this.] It turns out that the maximum power delivered is when R = r; the external load resistance is the same as the internal resistance of the cell.

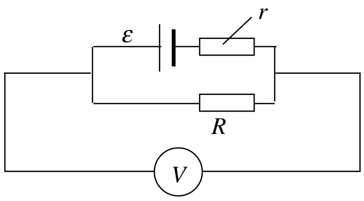
Kirchhoffs 2nd Law

Factsheet 23 deals with Kirchhoff's laws in detail. You should appreciate that Kirchhoff's 2nd Law is really a statement of conservation of energy. The $\Sigma\varepsilon$ is the sum of the energy being transformed into electrical energy by the cells and the ΣIR is the sum of the electrical energy being dissipated as heat by the resistors. One being taken as positive and the other negative means that we can write: $\Sigma\varepsilon = \Sigma IR$

Kirchhoff's 2nd Law: $\Sigma\varepsilon = \Sigma IR$, is a statement of conservation of energy

Practice Questions

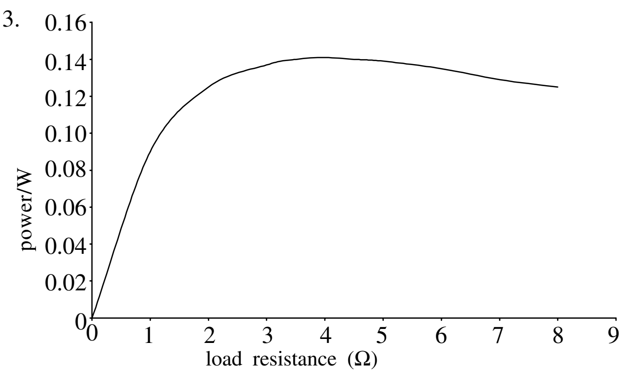
1. Define the e.m.f of a cell.
2. (a) What is the energy transfer per second for a cell of e.m.f. 6V when it is delivering a current of 0.2A?
   (b) If it has internal resistance of 0.3Ω, how much energy is wasted per second?
   (c) What will be the terminal P.D. of the cell, in these conditions?
   (d) What will be its terminal P.D. when it is delivering a current of 0.5A?
3. A resistor is used in the circuit shown.



(a) If the value of r is 4Ω, and ε = 1.5V, calculate values of the power dissipated in the load for values of R between 0 and 8Ω and draw up a table.
(b) Plot power against R, including more values around R = 4Ω as appropriate, so that you can draw a representative curve.
(c) Hence show that the maximum value of the power is when R = r

Answers

1. The e.m.f of a cell is the work done in moving a unit charge across the cell.
2. (a) Energy transfer 6J/C × 0.2C/s = 1.2J/s
   (b) Wasted energy = I^2r = 0.2 × 0.2 x 0.3 = 1.2 × 10^{-3}J
   (c) V = ε - Ir = 6 – (0.2 × 0.3) = 5.94V
   (d) V = ε - Ir = 6 – (0.5 × 0.3) = 5.85V
3.



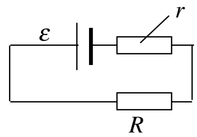
Use $I = \frac{\varepsilon}{R+r} = \frac{1.5}{R+4}$

R	0	1	2	3	4	5	6	7	8	3.5	4.5
R+4	4	5	6	7	8	9	10	11	12	7.5	8.5
I	0.375	0.3	0.25	0.214	0.188	0.167	0.15	0.136	0.125	0.2	0.1765
I^2R	0	0.09	0.125	0.137	0.141	0.139	0.135	0.129	0.125	0.14	0.14

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they could be improved. The examiner's mark scheme is given below.

The diagram shows a cell, of e.m.f ε and internal resistance, *r*, driving a current I through a load resistor as shown.



(a) Using these symbols, write down a formula for

(i) the power dissipated in the load resistor. (1)

power = V × I *0/1*

The candidate has merely written down the standard formula, and not used the symbols given in the question, ε, I, R and r

(ii) the power dissipated by the internal resistance (1)

power = V × I *0/1*

Again the candidate has failed to apply a formula to the context of the question.

(iii) the rate of conversion of energy in the cell (1)

Energy = ε *0/1*

The candidate knows that energy conversion has something to do with the e.m.f. but not what.

(b) Using these formulae, write down an equation for the conservation of energy in the circuit, and hence show that

$$I = \frac{\varepsilon}{(R + r)}$$ **(2)**

0/2

Since the candidate has failed to obtain the correct expressions, s/he is unable to put them together.

Examiner's answers

(a) (i) I^2R ✓

(ii) I^2r ✓

(iii) $\varepsilon \times I$ ✓

(b) Conservation of energy requires that the energy per second transformed in the cell ε x I equals the energy transformed per second in the circuit, so

$\varepsilon \times I = I^2R + I^2r$ ✓

dividing by I gives

$\varepsilon = I(R + r)$ *so* ✓

$$I = \frac{\varepsilon}{(R + r)}$$ ✓

Acknowledgements:
This Physics Factsheet was researched and written by Janice Jones. The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

www.curriculumpress.co.uk Number 62

Answering AS Questions on Circuits

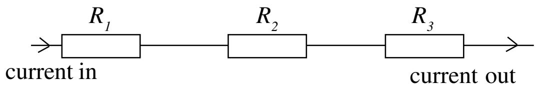
This Factsheet looks at some common types of questions involving electrical circuits that appear in AS level examination papers. Several ideas and equations will be used that have already been covered in previous Factsheets, however this Factsheet will concentrate on practising how to use the equations when answering questions.

Calculating the equivalent resistance

The equivalent resistance of a combination of resistors is the size of the single resistor that could be used to replace the entire combination and still offer the same resistance to the flow of charge.

(i) Resistors in series

Resistors that are connected in series offer only one path of conducting material for the charge in the circuit to follow.

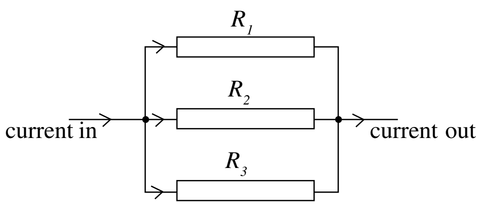


The charge is forced to flow through all of the resistors and the equivalent resistance of the combination of resistors is given by simply adding up the individual resistors. For example, in the circuit above with three resistors the equivalent resistance is given by:

> $R_{total} = R_1 + R_2 + R_3$
> R_{total} = *Equivalent resistance of all three resistors* (Ω)
> R_1 = *Resistance 1* (Ω)
> R_2 = *Resistance 2* (Ω)
> R_3 = *Resistance 3* (Ω)

(ii) Resistors in parallel

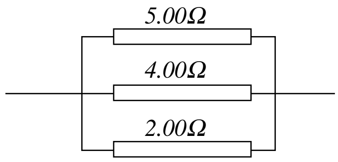
Resistors that are connected in parallel offer several paths of conducting material for the charge in the circuit to follow.



The charge only flows through one of the resistors in the above circuit. The equivalent resistance of this combination is given by a slightly more complex equation. For example, in the circuit above with three resistors the equivalent resistance is given by:

> $\frac{1}{R_{total}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$
> R_{total} = *Equivalent resistance of all three resistors* (Ω)
> R_1 = *Resistance 1* (Ω)
> R_2 = *Resistance 2* (Ω)
> R_3 = *Resistance 3* (Ω)

Example
Calculate the equivalent resistance of the following combination of resistors.



$$\frac{1}{R_{total}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$
$$= \frac{1}{5} + \frac{1}{4} + \frac{1}{2} = 0.2 + 0.25 + 0.5 = 0.95$$
$$R_{total} = \frac{1}{0.95} = 1.05\Omega$$

> ***Exam Hint – Avoiding careless errors***
> *The most common careless error in using this equation is calculating the answer for* $\frac{1}{R_{total}}$ *and forgetting to take the reciprocal of this number in order to quote the answer for* R_{total}.

Using equivalent resistance calculations in AS questions

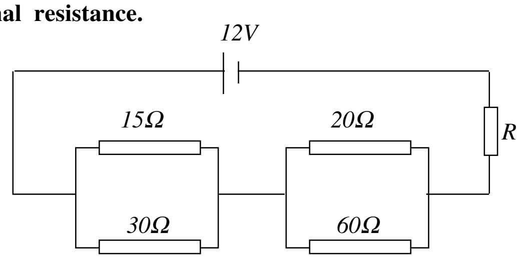
Calculating equivalent resistances of a combination of resistors usually forms part of a question that involves calculating the current through or voltage across a component. This calculation can be made using the following equation:

> $V = IR$ *where* V = *Voltage across component* (V)
> I = *current through component* (A)
> R = *resistance of component* (Ω)

> ***Current through components in series***
> *The current flowing through all components that are in series with each other will be the same.*

> ***Exam Hint - using V = IR***
> *When substituting values into this equation it is important that the values relate to the same component. You cannot use the emf of the battery with the resistance of a single resistor in the circuit, for example. If the total emf of the circuit is used then the total resistance of the circuit must be used to calculate the total current in the circuit.*

Typical Exam Question
A battery of emf 12V is connected to a resistor network as shown in the diagram below. It can be assumed that the battery has no internal resistance.



(a) Show that the resistance of the single equivalent resistor that could replace the four resistors between points A and B is 25Ω.(4)
(b) If R = 50Ω, calculate the current flowing through R. (2)

Answer
(a) *The combination of resistors consists of 2 sets of resistors in parallel. Let us first calculate the equivalent resistance of the 15Ω and 30Ω resistors, which are in parallel, we'll call this resistance R_1.*

$$\frac{1}{R_1} = \frac{1}{15} + \frac{1}{30} = 0.0667 + 0.0333 = 0.10$$

$$R_1 = \frac{1}{0.1} = 10\Omega \checkmark$$

Secondly, we can calculate the equivalent resistance of the 20Ω and 60Ω resistors, which are also in parallel. We'll call this R_2.

$$\frac{1}{R_2} = \frac{1}{20} + \frac{1}{60} = 0.05 + 0.01667 = 0.06667$$

$$R_2 = \frac{1}{0.06667} = 15\Omega \checkmark$$

The total resistance of the combination will be $R_1 + R_2$, as these two equivalent resistances are in parallel with each other.
Total Resistance = $R_1 + R_2$ = 10 + 15 ✓ = 25Ω ✓

(b) *The current in R can be calculated using the equation $I = \frac{V}{R}$.*
We have to be careful, however, as we do not know the voltage across R. The question only tells us the emf of the battery.
We can still use

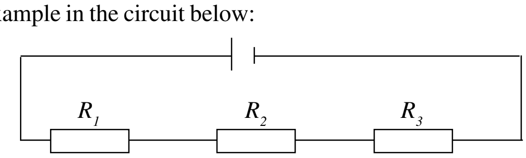
$$I = \frac{V}{R}$$

if we use the emf of the battery and the total resistance of the circuit. This will calculate the total current through the battery. R is in series with the battery and so will have the same current flowing through it as the battery.

$$I = \frac{V}{R} = \frac{12}{(50 + 25)} \checkmark = 0.16A \checkmark$$

The potential divider equation

This equation allows the voltage to be calculated across a single resistor in a series arrangement of several resistors.
For example in the circuit below:

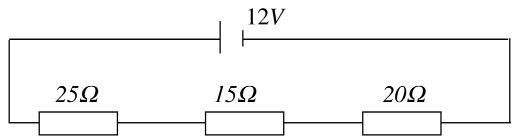


$$V_1 = \left(\frac{R_1}{R_{total}}\right) \times E$$

V_1 = Resistance across R_1 (Ω)
R_1 = Resistance (Ω).
R_{total} = Total resistance of the circuit (= $R_1 + R_2 + R_3$ in this circuit)
E = emf of the battery (V)

One advantage of using this equation is that no current is involved in the equation.

Example
Calculate the voltage across the 25Ω resistor in the circuit below.



$$V = \left(\frac{25}{(25 + 15 + 20)}\right)12 = 5.0V$$

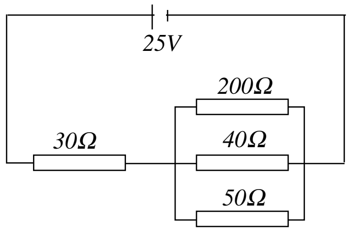
Note that the answer must be given as 5.0 to show it is correct to 2SF

Although the potential divider equation would seem to be limited to series circuits, it can also be used with parallel combinations. In this case the equivalent resistance of the parallel combinations would have to be calculated first and this equivalent resistance would then be used in the potential divider equation.

Voltage across parallel components
The voltage across all components that are in parallel is always equal.

The total voltage around a circuit
The voltage supplied by the battery in a series circuit is equal to the total voltage across all of the components in the circuit.

Typical Exam Question
A battery of 25V and negligible internal resistance is connected to a resistor network as shown in the circuit diagram below.



(a) Show that the resistance of the single equivalent resistor that could replace the three resistors in parallel is 20Ω. (2)
(b) What is the voltage across the 30Ω resistor? (2)
(c) What is the voltage across each of the three resistors in parallel? (3)

Answer
(a) *This is simply a matter of substituting values into the equation that we have already used.*

$$\frac{1}{R_{total}} = \frac{1}{200} + \frac{1}{40} + \frac{1}{50} = 0.0050 + 0.0250 + 0.0200 = 0.0500 \checkmark$$

$$R_{total} = \frac{1}{0.05} = 20\Omega \checkmark$$

(b) *We can now use the potential divider equation as long as we use the equivalent resistance for the parallel combination.*

$$\text{Voltage across } 30\Omega \text{ resistor} = \left(\frac{30}{(30+20)}\right)25 \checkmark = 15V \checkmark$$

(c) *All of these resistors are in parallel and so will have the same voltage across them. This voltage will also be the same as the voltage across the equivalent resistance. This can be calculated using the potential divider equation once more.*

$$\text{Voltage across all} \checkmark \text{ resistors in parallel} = \left(\frac{20}{(30+20)}\right)25 \checkmark = 10V \checkmark$$

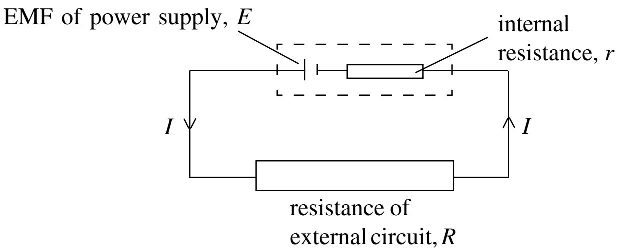
This answer could also have been calculated by knowing that the voltage across all of the resistors in the circuit must be the same as the voltage supplied by the battery.

Voltage supplied = voltage across parallel combination + voltage across 30Ω resistor

Voltage across parallel combination = voltage supplied − voltage across 30Ω resistor = 25 – 15 = 10Ω

Internal Resistance Calculations

Some questions about electrical circuits involve power supplies that have an internal resistance. Internal resistance is the name given to the resistance to the flow of charge inside the terminals of the power supply. The power supply is best thought of as a perfect emf in series with this internal resistance inside the confines of the power supply as in the diagram below.



By considering the power supply as a separate emf and resistor the internal resistance can be treated as just another resistor in series in the circuit.

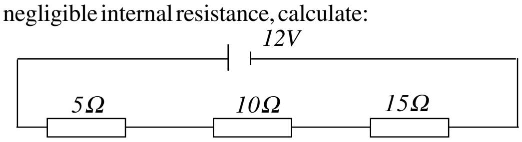
The last key concept of knowing that the voltage supplied by the battery is equal to the voltage across all of the components leads us to the following expression:

> *Emf of power supply* = *voltage across external resistance* + *voltage across internal resistance*
>
> $E = V + Ir$
> E = *emf of power supply*
> V = *voltage across external resistance*
> Ir = *current × internal resistance*
> = *voltage across internal resistance*

Quantitative (Calculation) Test

1. Calculate the equivalent resistance of a *10Ω* resistor and a *15Ω* resistor in that are connected in parallel.(2)

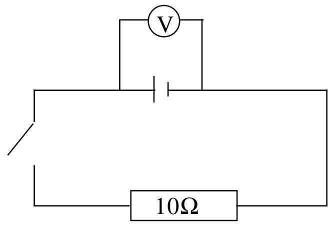
2. In the circuit shown below, assuming that the power supply has negligible internal resistance, calculate:



   (a) The voltage across the 5.0 *Ω* resistor.(2)
   (b) The current passing through the 5.0*Ω* resistor.(2)
   (c) The current passing through the power supply.(1)

3. In the circuit below the battery has an emf of 6.0 V and an internal resistance of 10 *Ω*.
   Calculate
   (a) The current flowing through the battery.(2)
   (b) The voltage between points A and B.(2)

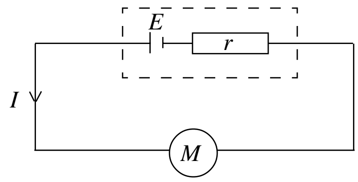
4. A battery is connected to a 10 Ω resistor as shown in the diagram below.



   When the switch is open the voltmeter reads 12.0 *V* and when it is closed it reads 11.5*V*
   (a) Explain why the readings are different. (2)
   (b) Calculate the internal resistance of the battery. (3)

Typical Exam Question

A car battery has an e.m.f. of 12.0*V*. When a car is started the battery supplies a current of 95.0 *A* to the starter motor. The voltage across the terminals of the battery drops at this time to 10.6*V* due to the internal resistance of the battery. The circuit diagram is shown below:



(a) Calculate the internal resistance of the battery. (2)
(b) The manufacturer warns against short-circuiting the battery. Calculate the current that would flow if the terminals of the battery were short-circuited and comment on your answer.(3)
(c) When completely discharged, the battery can be fully recharged by a current of 2.6*A* for 15 hours.
(i) How much charge is stored by the battery. (2)
(ii) How long could the motor be switched on for, continuously.(2)

Answer

(a) The question gives enough information to substitute straight into our equation involving the internal resistance of the battery. 10.6V is the voltage across the battery terminals, but this also represents the voltage across the external resistance. In this case the external resistance is provided by the motor.
$E = V + Ir$
Rearranging this equation to make internal resistance the subject of the equation gives:
$$r = \frac{(E-V)}{I} = \frac{(12-10.6)}{95✓} = 0.015\Omega\ ✓$$

(b) Short circuiting the battery implies that the two terminals of the battery are connected together with a wire and nothing else. This means that the circuit would consist of the e.m.f. of the battery and the internal resistance and nothing else.
The e.m.f. would be dropped across the internal resistance.✓
$$I = \frac{\text{voltage across internal resistance}}{\text{internal resistance}} = \frac{12}{0.014737} = 810\ A\ ✓$$
This is a huge current and very dangerous. It would create a great deal of heat energy.✓

NB The value for internal resistance used in this question has included all of the significant figures calculated form part (a) of the question.

(c) (i) The question has given us a time and a current. We have been asked to calculate a charge. The equation that links these three quantities is: Charge = Current × time
= (2.6)(15 × 3600)✓
= 140400 C = 140000C✓
Note how the time, given in the question in hours, has been turned into seconds. There are 3600 seconds in one hour.
Also note how the final answer has been rounded up to 2 significant figures, the same as the figures quoted in the question.

(ii) From part (b) of the question we have calculated the current that the motor uses. We have calculated from part (c)(i) of this question the charge that the battery has stored. We can now calculate a time using the same equation as in the last part of the equation.
Note how all significant figures from these previous calculations have been carried forward.
$$\text{Time} = \frac{\text{charge}}{\text{current}} = \frac{140400}{814.286✓} = 172\ \text{seconds}✓$$

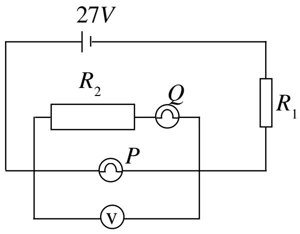
Exam Hint – Carrying forward answers to subsequent calculations
If a question requires you to carry out a calculation that uses a value that you have given as the answer to a previous part of the question, you must carry forward and use as many sig. figs. for this value as possible.

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's answer is given below.

In the diagram below *P* and *Q* are two lamps. *P* offers a resistance of 6.0Ω when it has its normal operating voltage of 12 *V* across it. *Q* offers a resistance of 2.0Ω when it has its normal operating voltage of 6.0 *V* across it.

The two lamps are connected in the circuit shown below. The battery has an emf of 27 *V* and negligible internal resistance. The resistors R_1 and R_2 have been chosen so that the lamps are operating at their normal operating voltage.



(a) What is the reading on the voltmeter? (1)

27V

> The student has assumed that all of the voltage supplied by the battery will be dropped across the parallel combination. There will be some voltage dropped across the resistor R_1. The question explains that the lamps are operating at their normal operating voltage. The voltmeter is directly across lamp *P* and so it will read the normal operating voltage of lamp *P*.

(b) Calculate the resistance of R_2. (3)

Resistance = voltage/current = (27-6)/?

> The student has correctly identified the correct equation to use but has got stuck on what value to use for current. The current through R_2 is the same as the current through lamp *Q*, which we can calculate. The student has also correctly noted that the voltage across R_2 will be the voltage across lamp *P* minus the voltage across lamp *Q* as lamp *Q* and R_2 are in parallel with lamp *P*.

(c) Calculate the current through R_1. (3)

$I = \frac{V}{R} = \frac{27}{(6+2)} = 3.4\ A$

> The student has simply picked any value of voltage and resistance from the question and substituted it into the equation. No consideration has been made of what the voltage and resistance refer to and whether they are across the same components or not.
> R_1 is in series with the parallel arrangement of bulbs and resistors and so it will have the same current through it as the total current through *P* and *Q*. We can work out the current through *P* and *Q* by using *I = V/R* for each bulb and then adding these values together.

(d) Calculate the voltage across R_1. (2)

$V = \frac{27}{2} = 13.5\ V$

> The student has now assumed that the 27 volts of the battery will be shared equally between R_1 and the parallel combination. This would only be the case if R_1 had the same equivalent resistance as the parallel combination.
> The reading across the voltmeter has been calculated in part (a). The voltage across R_1 will be the voltage of the power supply minus the reading on the voltmeter as the total voltage supplied to a circuit is equal to the voltage across all of the components of the circuit.

(e) Calculate the resistance of R_1.(2)

$R = \frac{V}{I} = \frac{13.5}{3.4} = 4.0\ \Omega$

> The student has followed the correct procedure by substituting values in for *V* and *I* from the previous parts of the question. Unfortunately these are the incorrect values and in direct contradiction to what the student did in part (c) of the question.

Examiner's answers

(a) Voltage = voltage across lamp P = 12 V

(b) Current through R_2 = Current through lamp Q = voltage /resistance = 6/2 = 3.0 A

Voltage across R_2 = 12 – voltage across lamp Q = 12 – 6 = 6.0V

$R_2 = \frac{voltage}{current} = \frac{6}{3} = 2.0\ \Omega$

(c) Current through lamp P = voltage/resistance = 12/6 = 2.0 A

Current through lamp Q = 3.0 A

Current through R_1 = current through lamp P + current through lamp Q = 2+3= 5.0O

(d) Voltage across R_1 = voltage of power supply – voltage across parallel section = 27 – 12 = 13 V

(e) $R_1 = \frac{V}{I} = \frac{13}{5} = 2.6\ \Omega$

Answers

1. $\frac{1}{R_{total}} = \frac{1}{10} + \frac{1}{15} = 0.1 + 0.06667 = 0.16667$

   $R_{total} = \frac{1}{0.16667} = 6.0\ \Omega$

2. (a) $V = \frac{5}{(5+10+15)}\ 12 = 2.0$ V

   (b) $I = \frac{V}{R} = \frac{2}{5} = 0.40$ A

   (c) 0.40 A

3. (a) $I = \frac{V}{R} = \frac{6}{(15+15+10)} = 0.15$ A

   (b) $V = E - Ir = 6 - (0.15 \times 10) = 4.5$ V

4. (a) When the switch is open no current flows around the circuit and the voltmeter reads the emf of the battery.
   When the switch is closed a current flows and a voltage is dropped across the internal resistance of the battery and so the voltmeter reads the 'terminal voltage of the battery'.

   (b) Current through circuit = current through 10 Ω resistor

   $= \frac{11.5}{10} = 1.15$ A

   $V = E - Ir$

   Therefore $r = \frac{-(V - E)}{I} = \frac{(11.5 - 12)}{1.15} = 0.43\ \Omega$

Acknowledgements: *This Physics Factsheet was researched and written by* ***Jason Slack***
The Curriculum Press,Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU.
 ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 90

Answering Exam Questions: Series and Parallel Calculations

Series and parallel calculations appear not only in electricity (resistors and capacitors), but also in mechanics (elasticity and oscillations).

The first difficulty that occurs is that the relationships involved are not consistent e.g. series and parallel resistor and capacitor systems work in opposite ways. A second difficulty that catches out some students is that these relationships occur within a square-root sign for simple harmonic motion.

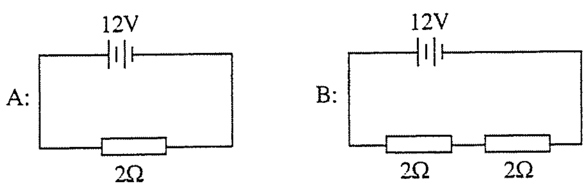
There are other examples of series and parallel relationships such as capacitors and inductors in tuning circuits, but they are beyond the scope of this Factsheet.

Combinations of resistors

Resistors in **series** are simple to deal with. If we add a second resistor in series with an existing one, it is more difficult to push the current through. The effective resistance increases.

Series: $R = R_1 + R_2 + ...$ (ohms, Ω)

Example 1: The diagram shows two circuits, A and B:

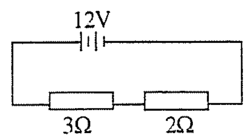


Find:
(a) the effective resistance of each circuit.
(b) the current through each circuit.
(c) the power dissipated in each resistor.
(d) the power drawn from each supply.

Answers
(a) A: $R = 2\Omega$
B: $R = 4\Omega$
(b) A: $I = \frac{V}{R} = \frac{12}{2} = 6A$
B: $I = 3A$
(c) A: $P = I^2R = 36 \times 2 = 72W$
B: $P = 9 \times 2 = 18W$ for each resistor.
(d) A: supply power, $P = VI = 12 \times 6 = 72W$
B: supply power, $P = 36W$

Adding resistors in series not only reduces the power dissipated in each resistor, it also reduces the power (and current) drawn from the supply.

Example2 : Calculate the power dissipated in each resistor in this circuit.



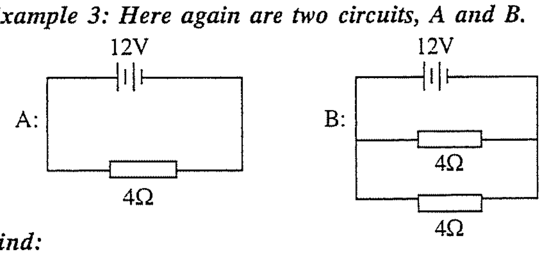
Answer: $I = \frac{V}{R} = \frac{12}{5} = 2.4A$.
For 3Ω resistor, $P = I^2R = 2.4^2 \times 3 = 17.3W$.
For 2Ω resistor, $P = 11.5W$.

In a series set-up, more power is dissipated in the larger resistor.

Adding resistors in **parallel** increases the possible routes the current can take, making it easier for the current to flow. The effective resistance is smaller than any of the individual resistances.

Parallel: $\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + ...$ (ohms, Ω)

Example 3: Here again are two circuits, A and B.



Find:
(a) the effective resistance of each circuit.
(b) the current drawn from each supply.
(c) the power dissipated in each resistor.
(d) the power drawn from each supply.

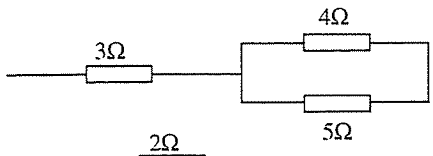
Answers:
(a) A: $R = 4\Omega$
B: $\frac{1}{R} = \frac{1}{4} + \frac{1}{4}$, $R = 2\Omega$
(b) A: $I = \frac{12}{4} = 3A$
B: $I = \frac{12}{2} = 6A$
(c) A: $V = 12V$, $P = \frac{V^2}{R} = \frac{144}{4} = 36W$
B: $P = 36W$ (again) in each resistor
(d) A: $P = VI = 36W$
B: $P = VI = 72W$

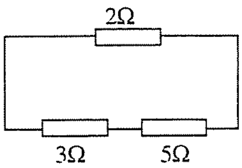
Adding resistors in parallel increases the power (and current) drawn from the supply.

Exam Hints:
(a) Notice from the series and parallel examples that power can be calculated from current through or p.d. across each resistor. Use whichever relationship is most convenient.
(b) The most common error in calculating the effective parallel resistance is forgetting to invert your calculation e.g. 1/R = 1/6, R = 6Ω.

For combinations of series and parallel resistors, just work through logically:

***Example 4:** Find the effective resistance in each case:*

A:


B:


Answer:
(a) 4Ω and 5Ω in parallel: $\frac{1}{R} = \frac{1}{4} + \frac{1}{5}$, *R = 2.2Ω*
3Ω and 2.2Ω in series, R = 5.2Ω

(b) 3Ω and 5Ω in series: R = 8Ω
2Ω and 8Ω in parallel: R = 1.6Ω

Combinations of capacitors

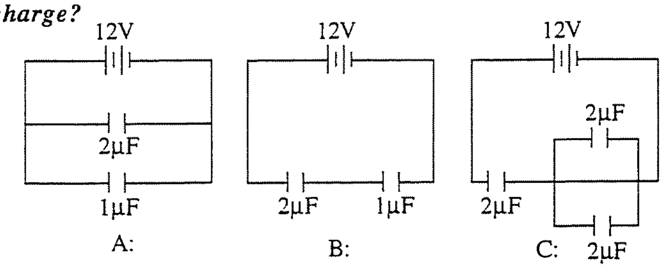
The methods used here are very similar to those for resistors. But remember, capacitors in parallel allow more charge to be stored – the effective capacitance increases. Capacitors in series mean a smaller p.d. across each capacitor, reducing the charge that can be stored – the effective capacitance decreases.

Series: $\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} ...$ (farads, F)

Parallel: $C = C_1 + C_2 + ...$ (farads, F)

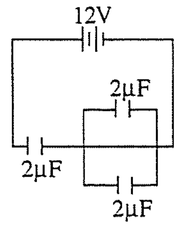
***Exam hint:** Be very careful with powers. Capacitance is often measured in μF (10^{-6} F).*

***Example 5:** Which of these combinations would hold the most charge?*

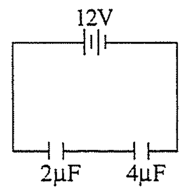


Answer:
Circuit A: $C = 2 + 1 = 3\mu F$.
Circuit B: $\frac{1}{C} = \frac{1}{2} + \frac{1}{1} = \frac{3}{2}$, $C = \frac{2}{3}$ *or 0.67μF*
Circuit C: Parallel C = 2 + 2 = 4μF,
then series $\frac{1}{C} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, $C = 1.33\mu F$.
So circuit A holds the most charge.

***Example 6:** Find the energy stored in each capacitor.*



***Answer:** The parallel combination leads to this equivalent circuit.*



As both capacitors in a series circuit hold the same charge, Q, then from Q = CV, we have
$C_1V_1 = C_2V_2$, *or* $\frac{V_1}{V_2} = \frac{C_2}{C_1} = \frac{4}{2} = \frac{2}{1}$
So the 2μF capacitor has 8 volts across it, and the 4μF capacitor (or the parallel combination) has only 4 volts across it.

The energy stored:
$E_1 = ½ CV^2 = ½ \times 2 \times 10^{-6} \times 64 = 6.4 \times 10^{-5}$ *J.*
Similarly, $E_2 = E_3 = 1.6 \times 10^{-5}$ *J*

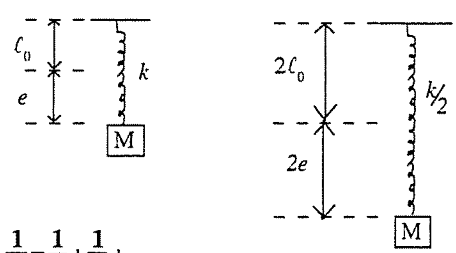
Using the effective capacitance of 1.33μF found earlier, we find the total energy stored is $E = 9.6 \times 10^{-5}$ *J (as expected).*

***Exam hint:** Notice that the total energy stored by the effective capacitance must be the same as that stored by the individual capacitors in total.*

Combinations of springs

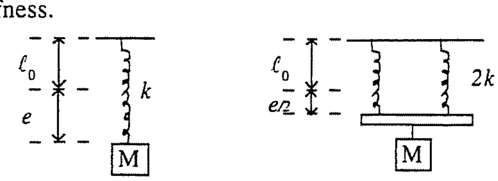
Again, common sense leads you to the correct results. You would expect a long bungee rope to stretch more than a short one. So combining springs in series reduces the stiffness i.e. the **spring constant, k.**

k = F/e (Nm^{-1}), where **e** is the extension of the spring system.



Series: $\frac{1}{k} = \frac{1}{k_1} + \frac{1}{k_2} + ...$

And if you have ever used a muscle building apparatus involving springs, you will know that combining springs in parallel increases the stiffness.



Parallel: $k = k_1 + k_2 + ...$

***Exam hint:** Unlike resistors and capacitors, the springs you might be asked to combine are almost certain to be identical.*

Example:
(a) For three identical springs of spring constant k_o = 400Nm^{-1}, find the effective spring constant when they are combined in series and in parallel.
(b) If each spring has an initial length of 20cm, find the extension in each case if a 30N force is applied to the combination.

Answer:

(a) In series, $\frac{1}{k} = \frac{1}{k_o} + \frac{1}{k_o} + \frac{1}{k_o} = \frac{3}{k_o}$, $k = \frac{k_o}{3} = 133\text{Nm}^{-1}$

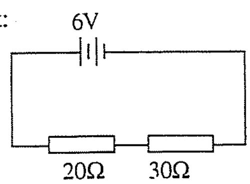
In parallel, $k = k_o + k_o + k_o = 3k_o = 1200\text{Nm}^{-1}$

(b) In series, $F = ke$, $e = \frac{F}{k} = \frac{30}{133} = 0.23\text{m}$

In parallel, $e = \frac{30}{1200} = 0.025\text{m}$

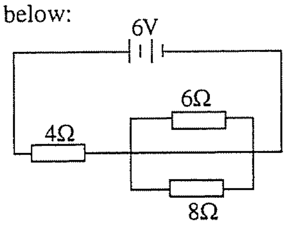
Questions

1. For this circuit:



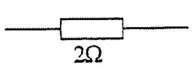
Find
(a) the effective resistance
(b) the current drawn from the supply
(c) the p.d. across each resistor
(d) the power dissipated in each resistor
(e) the power drawn from the supply.

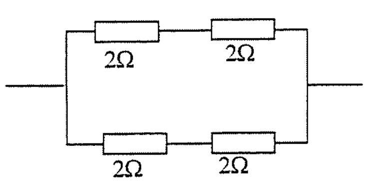
2. For the circuit below:



Find
(a) the effective resistance
(b) the current and power drawn from the supply
(c) the p.d. dropped across each resistor
(d) the power dissipated in each resistor.

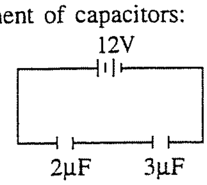
3. A resistance of 2Ω can be constructed from identical resistors in both of these ways:





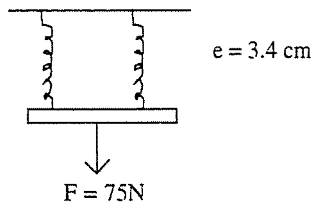
Why might you choose the second arrangement?

4. In this arrangement of capacitors:



(a) Find the effective capacitance and the charge stored in the system.
(b) What is the charge stored on each capacitor?
(c) Find the voltage across each capacitor.
(d) Find the energy stored on each capacitor.
(e) Compare this to the energy provided by the supply. Explain.

5. For this arrangement, find the spring constant for each of these identical springs if a 75N force causes an extension, e, of 3.4cm.



Answers

1 (a) 50Ω.
(b) $6/50 = 0.12A$
(c) 20Ω: $V = 0.12 \times 20 = 2.4V$
30Ω: $V = 3.6V$
(d) 20Ω: $I^2R = 0.29W$
30Ω: $0.43W$
(e) $P = VI = 6 \times 0.12 = 0.72W$ $(= 0.29 + 0.43)$

2. (a) $\frac{1}{R_p} = \frac{1}{6} + \frac{1}{8}$, $R_p = 3.4\Omega$
$R = 4 + 3.4 = 7.4\Omega$
(b) $I = 0.81A$, $P = 4.9W$
(c) 4Ω: $V = IR = 0.81 \times 4 = 3.2V$
Parallel pair: $V = 2.8V$
(d) 4Ω: $P = \frac{V^2}{R} = 2.6W$
6Ω: $P = \frac{V^2}{R} = 1.3W$
8Ω: $P = 0.98W$
Total = 4.9W as expected.

3. *In a high power circuit, it might be preferable to share the power among the four resistors, to prevent a single resistor overheating.*

4. (a) $\frac{1}{C} = \frac{1}{2 \times 10^{-6}} + \frac{1}{3 \times 10^{-6}}$, $C = 1.2 \times 10^{-6}$ F
$Q = CV = 1.44 \times 10^{-5}$ C.
(b) The same in a series circuit: $Q = 1.44 \times 10^{-5}$C
(c) $2\mu F$: $V = Q/C = 7.2V$
$3\mu F$: $V = 4.8V$
(d) $2\mu F$: $E = \frac{1}{2}CV^2 = 5.2 \times 10^{-5}$ J
$3\mu F$: $E = 3.5 \times 10^{-5}$ J
(e) Supply: $E = QV = 1.73 \times 10^{-4}$ J
Total for capacitors, $E = 8.7 \times 10^{-5}$ J
Half of the energy drawn from the supply is dissipated in Joule heating.

5. For the combination, $k = \frac{F}{e} = \frac{75}{3.4 \times 10^{-2}} = 2205\ \text{Nm}^{-1}$.
Parallel springs, spring constants add.
For each spring, $k = 1103\ \text{Nm}^{-1}$

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

www.curriculumpress.co.uk Number 56

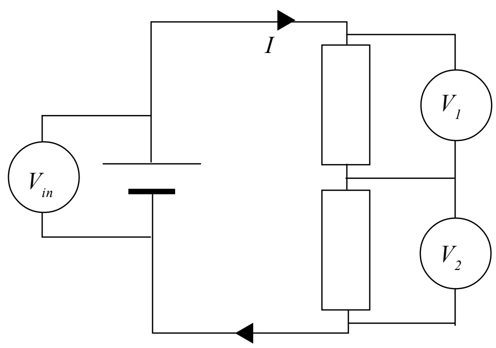
The Potential Divider and its Uses

> *The potential divider is an electrical device that supplies a range of potential differences (voltage drops) of anywhere between 0V to a maximum of the supply potential difference by using a combination of resistances.*

In order to understand how the potential divider works it is necessary to remind ourselves, briefly, of some more basic ideas in electricity.

The simpliest form of the potential divider uses two fixed resistors in series.

Fig. 1: A simple potential divider circuit



$V_{in} = V_1 + V_2$ I is constant throughout circuit

Any resistance in a circuit will have a voltage drop across it, we know this from $V = IR$. In series the total of the voltage drops across all components in the circuit must equal the potential difference supplied, this is simply a statement of conservation of energy but is also expressed as Kirchoff's second law. The electrons in the current have no option but to flow through both as there are no alternative routes for them, so current is the same for both resistances.

> *Series components will have the **same current** through them whereas they will **share the supply potential difference** across them.*

> ***Exam Hint:** Remember potential differences and voltage drops are always **across** components, this is because they are the difference in energy per unit charge between when they enter and then when they emerge from a component. This is why voltmeters are always attached across components. Voltages **never** flow or travel **through** components – this makes no sense and is therefore wrong.*

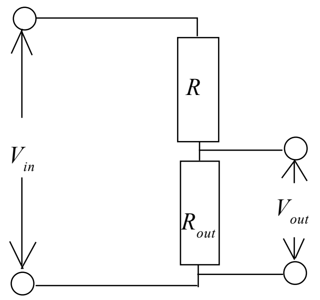
Potential dividers make use of the fact that the amount of potential difference across resistances in series is directly proportional to their resistances - $V \propto R$ (from Ohm's law) – as current is constant for both components. We can then tap off a voltage by placing our component across only one of the resistors in series.

> *The larger the resistance of resistor in a potential divider the greater its share of the voltage. The voltages will always add up to the supply potential difference.*

Potential dividers can supply a potential difference of any value up to the value of the supply potential difference by varying the size and arrangement of the resistors. This means we can tap off varying potential differences from a fixed supply.

The fraction of the supply voltage a resistor will take is equal to its fraction of the total resistance. This is expressed by the equation below (Fig 2):

Fig. 2: A typical potential divider showing input and output voltages



$$\frac{V_{out}}{V_{in}} = \left(\frac{R_{out}}{R + R_{out}}\right)$$

$$V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in}$$

Worked Example: Using the above circuit,find the value of V_{out}, given that $V_{in} = 12V$, $R = 6k\Omega$, $R_{out} = 3k\Omega$.

We can (i) Use the formula:

$$V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{3000}{6000 + 3000}\right) \times 12 = 4V$$

Or (ii) use ratios

R_{out} forms $\frac{1}{3}$ of the total resistance $\frac{3000}{9000}$

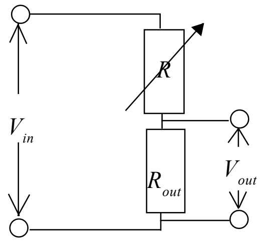
so it will take $\frac{1}{3}$ of the supply voltage.

$V_{out} = \frac{1}{3} \times 12 = 4V$

More complicated potential dividers use other components in conjunction with fixed resistances.

A variable resistor or rheostat can be used to change the portion of the supply voltage being used. One example is the volume control on a stereo system.

Fig 3: Simplified diagram of the potential divider (volume control)



This time one of the fixed resistors has been replaced with a rheostat. By varying the resistance of R, because the voltage drop depends on the resistance, we can change the value of the output signal across R_{out}. If the resistance of the rheostat is increased this increases its total share of the potential difference, this leaves less for R_{out} and its share falls, as their total has to remain equal to the supply. Because V_{out} falls then the volume will also fall.

Typical Exam Question
A potential divider is formed with a rheostat connected in series with a fixed resistor of 4kΩ. Both are connected across a 12V supply. Initially the rheostat is set at 10KΩ.
(a) Calculate the potential difference across R_{out} (V_{out}). [2]
(b) Explain without further calculation what happens to the value of the output signal if the resistance of R is decreased. [2]

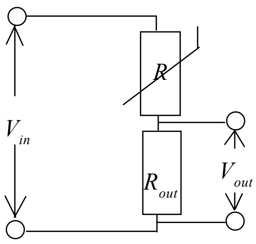
$$(a)\ V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{4000}{10000 + 4000}\right) \times 12 \checkmark = 3.4V \checkmark$$

(b) If R is increased then the voltage across it decreases ✓
This means the voltage across R_{out}, V_{out}, increases ✓

There are various other components that can be placed in a potential divider but they all work on the same principle as the simple volume control above. Following are some possible arrangements.

A thermistor is a resisting device made of a semiconductor, as its temperature increases its resistance decreases as more charge carriers are liberated and can form a larger current. A thermistor can be used in a potential divider to provide a device that responds to temperature variation.

Fig 4. A temperature dependent potential divider using a fixed resistor and thermistor.

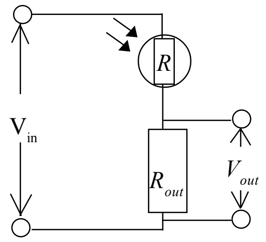


As temperature increases, then the resistance of the thermistor decreases, this in turn lowers the voltage across it. Hence the voltage across R_{out} (V_{out}) increases. If we connect an LED or a buzzer across R_{out} then the rise in temperature could trigger a warning. Alternatively a fan could be triggered to cool the object in question, for example the radiator in a car.

Notice that if we tap the voltage drop across the thermistor itself rather than the fixed resistor then the opposite effect occurs – the colder something is the higher V_{out}; possibly triggering a relay linked to a heater or an LED showing an ice warning in a car.

A light dependent resistor (LDR) as the name suggests has a resistance that varies with light intensity. As the intensity of light falling on it increases its resistance decreases as more energy is supplied, enabling more charge carriers to be released and form a larger current.

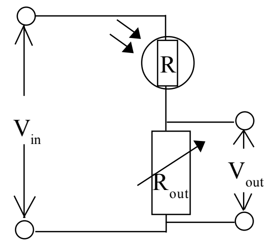
Fig 5: A light sensitive potential divider using a LDR and a resistor



In the LDR circuit above, as the light levels increase then the resistance of the LDR decreases and so does the potential difference across it. V_{out} would increase and could trigger a light sensitive alarm in a safe, for example. As before with the thermistor, if we swap V_{out} so it is across the LDR then as light levels fall our output voltage would increase, possibly triggering a nightlight.

It is important to realise that the above circuits are simplified versions of the circuits used in applications discussed. Often the output voltage is connected across a transistor or relay rather then the output device directly, this allows it to trigger a definite switch on rather than a gradual increase in voltage and allows a larger current to supply the component. One alteration that can be made is using a rheostat in series with an LDR or Thermistor, as show below.

Fig. 6: Use of a rheostat to change sensitivity of potential divider



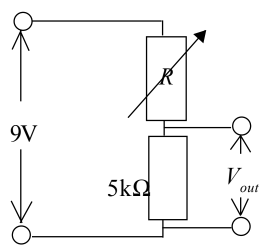
If the rheostat's resistance is decreased then more voltage will drop across the LDR at a given light intensity, this increases the light intensity needed to give the same value of V_{out} across R_{out}. In this way circuits can be fine tuned to respond in the desired way.

Practice questions

1. A potential divider is used to detect changes in temperature using a thermistor with a resistance that varies between 1000Ω and 20000Ω.
   (a) Apart from some form of power supply what other component is required?
   (b) Suggest a reasonable resistance for it to be set to.

2. An LDR varies between 500Ω and 12kΩ according to the level of light intensity on it. If it is used in a potential divider in conjunction with 15V power supply and 2.0kΩ fixed resistor calculate the range of voltages that can be tapped off across the fixed resisitor.

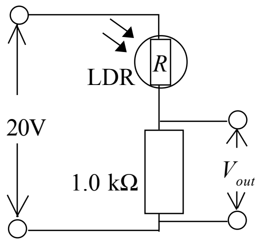
3. A variable resistor is used in a potential divider as shown below.



R is set at 7kΩ.
(a) Find the reading on a voltmeter used to measure V_{out}.
(b) If the value of R is doubled what is the reading on V_{out} now?

Exam Workshop
This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's mark scheme is given below.

The following question relates to the potential divider shown below.



The LDR is exposed to a reasonably high level of light intensity. The reading on a voltmeter attached across the terminals labelled V_{out} is 10V.

(a) What is the resistance of the LDR at this intensity? **[1]**
1kΩ ✓ 1/1

> The student has correctly realised that for the two resistances to share the supply p.d. equally they must have the same resistance.

(b) (i) Find the reading on the meter when the LDR has a resistance of 0.7KΩ. **[2]**

$$V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{0.7}{1.0 + 0.7}\right) \times 20 = 8.2V \qquad 0/2$$

> The student has mixed up the two resistances using the LDR resistance in place of the fixed resistance. Student has also forgotten that resistances are quoted in kΩ; although they cancel in this particular case it is still good practice to indicate you are aware of this

(ii) What is the voltage drop across the LDR at this intensity?[1]
20 – 8.23 = 11.8V ✓ ecf 1/1

> the correct technique has been used but the wrong answer is obtained due to the error in (b) (i). This has already been penalised so student gains an error carried forwards mark.

(c) The light intensity is altered and an increase in the value of V_{out} is observed. State and explain whether this represents an increase or decrease in light intensity. **[3]**

The light intensity has decreased as the voltage through the LDR has decreased because the voltage through the resistor is larger. *0/3*

> Voltage is always across and never through. The light intensity must increase to lower its resistance and therefore the voltage drop across it.

Examiner's answers
(a) 1.0kΩ ✓

(b) (i) $V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{1k}{0.7k + 1k}\right) \times 20$ ✓ *= 11.76V* ✓

(ii) 20 – 11.76 = 8.24V ✓

(c) The light intensity increases ✓
because the voltage across the LDR must fall ✓
so its resistance must fall, for V_{out} *to increase* ✓

Answers

1. (a) Fixed resistor, rheostat/variable resistor
   (b) 5kΩ – 10kΩ – must give a reasonable variation between max and min thermistor resistance

2. Greatest V_{out} when LDR resistance is lowest.

$$V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{2000}{500 + 2000}\right) \times 15 = 12V.$$

Lowest when LDR resistance is highest.

$$V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{2000}{12000 + 2000}\right) \times 15 = 2.1V$$

Range is 2.1V – 12V.

3. (a) $V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{5k}{7k + 5k}\right) \times 9 = 3.75V$

   (b) $V_{out} = \left(\frac{R_{out}}{R + R_{out}}\right) \times V_{in} = \left(\frac{5k}{14k + 5k}\right) \times 9 = 2.4V$

Acknowledgements: This Physics Factsheet was researched and written by ***Alan Brooks.***
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU.
Physics Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber.

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 67

The Negative Temperature Coefficient (NTC) Thermistor and the Light-Dependent Resistor

This Factsheet will explain:
- how the resistance of an NTC Thermistor varies with temperature;
- how the resistance of a Light-Dependent Resistor varies with light level;
- how the effects may be demonstrated experimentally;
- how LDRs or thermistors may be used to vary voltage.

Before studying this Factsheet, you should ensure that you are familiar with:
- the concept and definition of resistance from your GCSE course – that Ohm found the ratio of the P.D. across a component to the current through it to be a constant, which he called "resistance";
- that this simple concept does not hold for all materials or components; that "resistance" is not a universal constant, but is still a useful concept in prescribed circumstances. Hence, a definition of "resistance" as the ratio of V/I in a given set of conditions;
- that in metals resistance increases with temperature.

You may also find it useful to look at Factsheet 7 – Current, voltage and resistance and at Factsheet 28 – Graphs in physics.

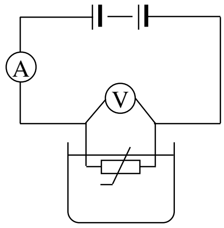
Resistance

> ***Resistance is defined as the ratio of the voltage across a component to the current through it.***

The concept of resistance varying with changing conditions is usually presented in one of two ways:
- either as a graph of resistance against a particular factor such as temperature or light intensity,
- or as current/voltage characteristics. In this case, the resistance is related to the gradient of the graph.
- The graph of the current/voltage relationship is often presented with the other 3 quadrants of the graph shown, not just the usual positive/positive one. This is because for some components there is a difference if the current is reversed.
- The graphs are sometimes presented as voltage/current and sometimes as current/voltage. **Always** look carefully to see which way round the graph is plotted.

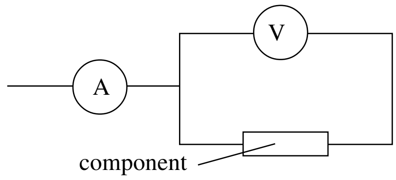
> *Always check to see which way round a graph is plotted. The resistance is the gradient of a voltage/current graph, but the **inverse** gradient of a current/voltage graph.*

Investigating the voltage/current characteristics for a thermistor with changing temperature.

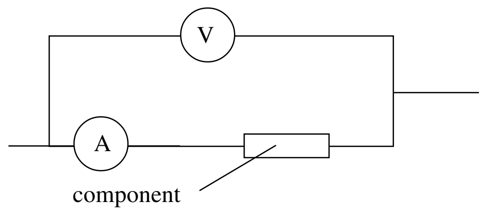


Apply a suitable voltage across a thermistor to give a reasonable reading on the ammeter. Take pairs of values of *V* and *I* for changing temperatures from 0^0C (by immersing in ice) to 100^0C (by heating the water to boiling). Ensure that the temperature is even throughout the water at each reading by stirring. Repeat each reading and average them to increase accuracy. Calculate R for each pair of values and plot *R* against *t*.

This investigation can be done very conveniently using datalogging equipment. Advantages include a large number of values due to the possible frequency of sampling, and automatic graph plotting facilities. An ohmmeter can be used in place of the voltmeter and ammeter. Using an ohmmeter reduces the errors involved in using ammeter/voltmeter. When using the ammeter/voltmeter, either the ammeter measures the current through the voltmeter as well as that through the component:

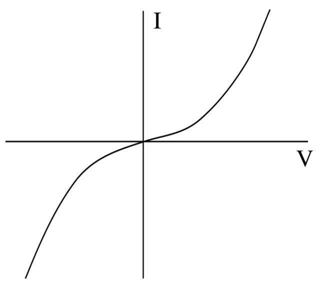


or the voltmeter measures the P.D. across the ammeter as well as that across the component.



Either method introduces an error; though this should be small if the meters are close to being "ideal" i.e. the ammeter has a very small resistance and the voltmeter a very high resistance.

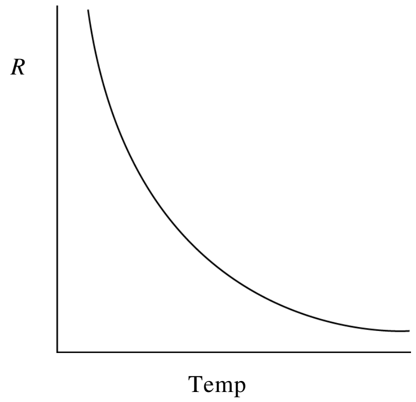
The Current/Voltage characteristic of a NTC Thermistor
The graph obtained is of the form shown below:



On the I/V plot the resistance is the inverse of the gradient, i.e. the resistance is **decreasing**. This is because as V increases, the thermistor gets hotter.

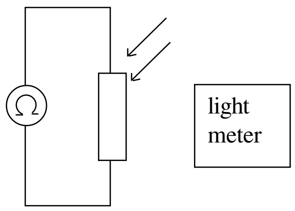
> *For the NTC Thermistor the resistance **decreases** as temperature increases.*

A plot of resistance against temperature would look as below:



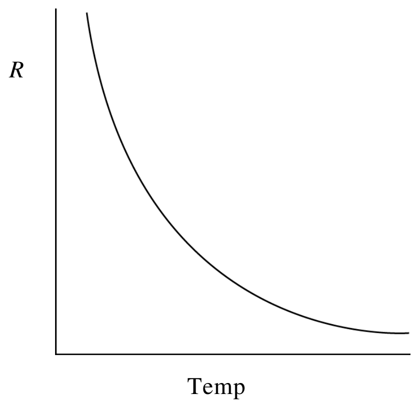
The Light Dependent Resistor

The dependence of the resistance of an LDR on the intensity of light can be investigated using the apparatus shown below:



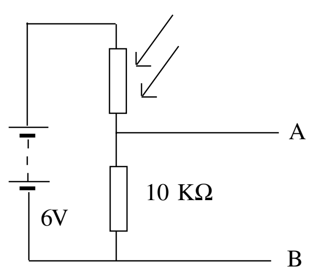
The ambient light level is varied and read by the light meter. For each light level, the resistance of the LDR is measured with a digital ohmmeter. The light level may be varied by altering the source of light, or by having a strong source of light moved further back in stages, or by putting successive sheets of paper over a source. Obviously, it is important that the light meter is reading the light intensity at the position of the LDR, so care must be taken to achieve this.

The graph of the dependence of the resistance on the light intensity is a below:



Using a Thermistor or an LDR to control voltage

A thermistor or an LDR can be used as part of a potential divider to control voltage.



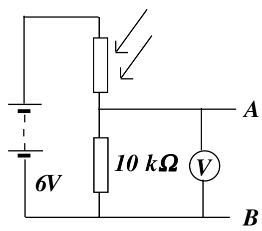
The resistance of the bottom resistor is fixed at, say, 10 kΩ. The LDR in the top part of the potential divider has a resistance that will vary depending on the light intensity. It may be from the order of 1MΩ in low light, to 1kΩ in higher light levels. In strong light, the resistance of the LDR will be about 1kΩ, hence the P.D. between A and B will be about (10k/11k) 6V i.e. not far off 6V.

When the LDR is in low light, its resistance will be around 1MΩ, so the P.D. between A and B will be about: $\frac{10 \times 10^3}{1001 \times 10^3} \times 6\text{V}$ i.e. 0.06V.

Thus the P.D. between A and B will vary from 6V to almost nothing depending on the light level incident on the LDR. This arrangement can be used to trigger a change from on to off in a logic gate, which can turn various devices on or off. This concept is used extensively in automatic electronic control systems, where the system is required to respond to changes in light level. The value of the fixed resistor in the bottom half of the potential divider can be set to determine the precise light level at which the logic gates switch.

An NTC thermistor can be used in exactly the same way to respond to changes in temperature, rather than light intensity. This arrangement is used extensively in devices such as fire alarms and automatic sprinkler systems.

Worked Example
The following circuit can be used as a lightmeter.



(a) The maximum value of the LDR is 900kΩ. What is the reading on the voltmeter for this resistance (assuming that the voltmeter has a very high resistance)? ***(2)***
(b) The minimum value of the LDR is 1kΩ. What is the reading on the voltmeter for this resistance? ***(2)***
(c) Explain how the readings on the voltmeter enable this circuit to be used as a lightmeter. ***(2)***
(d) Explain what difference it would make to the readings if the voltmeter did not have a very high resistance. ***(2)***

Answers

(a) The reading on the voltmeter will be:

$$\frac{10 \times 10^3}{910 \times 10^3} \times 6V = 6.6 \times 10^{-2} V$$

(b) The reading on the voltmeter will be:

$$\frac{10 \times 10^3}{11 \times 10^3} \times 6V = 5.45V$$

(c) The readings on the voltmeter are related to the light intensity and so can be used as a measure of it. Since it is not a linear relationship, a calibration curve would be drawn up to show the light intensity corresponding to each meter reading.

(d) If the voltmeter did not have a very high resistance, it would affect the voltage between A and B, the reading would be:

$$\frac{R_{AB}}{R_t} \times 6V$$

where R_{AB} is the parallel combined resistance of the 10kΩ and the voltage of the voltmeter, i.e.

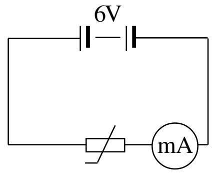
$$\frac{1}{R_{AB}} = \frac{1}{10 \times 10^3} + \frac{1}{R_{voltmeter}}$$

and R_t is the series combined resistance of R_{AB} and the resistance of the LDR.

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's mark scheme is given below.

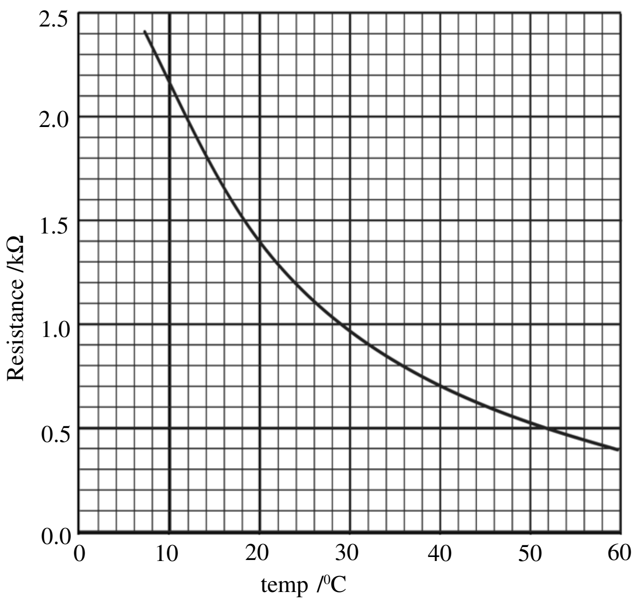
An NTC thermistor is used in the following circuit to make a temperature sensor.



(a) Explain how the circuit works. (3)
The reading on the mA is related to the temperature. *1/3*

> The candidate has scored a mark for knowing how the reading is achieved, but shows no understanding of the principles behind the device.

(b) The graph shows dependence of the resistance on temperature.



What will be the reading on the milliameter when the thermistor is at a temperature of 20°C? **(3)**
1.4kΩ *0/3*

> The pupil has merely read the value of resistance from the axis, and not used V = IR to calculate the current.

Examiner's Answers

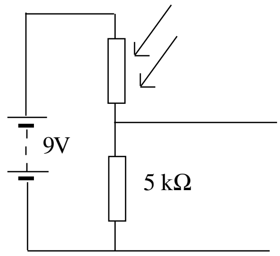
(a) The resistance of the NTC thermistor decreases with temperature, so the current increases with temperature. A calibration curve can be drawn up, which shows the corresponding temperature for each current value.

(b) From the graph, the resistance of the thermistor at 20°C will be 1.4kΩ, so

$$I = \frac{V}{R} = \frac{6}{1.4 \times 10^3}$$

Questions

1. Define "resistance" for a circuit component.
2. What is meant by
   (a) An NTC thermistor
   (b) An LDR?
3. Describe how you would obtain results to plot a graph to show the temperature dependence of an NTC thermistor.
4.



In the arrangement shown in the diagram, what is the potential difference across the 5kΩ resistor when the resistor of the LDR is
(a) 1kΩ
(b) 100kΩ?

5. Describe how an NTC thermistor could be used in a potential divider arrangement for use in a fire alarm.

Answers

1. The resistance of a component is the ratio of the potential difference across it to the current through it.
2. (a) An NTC thermistor is a resistor whose resistance decreases as temperature increases.
   (b) An LDR is a resistor whose resistance varies with the light intensity incident on it. The resistance is lower for greater intensity.
3. See text
4. (a) The P.D across the 5kΩ is
   $\frac{5}{6} \times 9V = 7.5V$
   (b) The P.D across the 5kΩ in this case is
   $\frac{5}{100} \times 9V = 0.45V$
5. An NTC Thermistor could be used in an arrangement like that in question 4, with the thermistor in place of the LDR. When the temperature rises, the resistance of the thermistor will go down, so the P.D. across the 5kΩ resistor will rise. This could trigger a logic gate at a set value, switching on an alarm or an automatic sprinkler. The value of the fixed resistor should be chosen to determine the temperature at which the gate switches on.

Acknowledgements:
This Physics Factsheet was researched and written by Janice Jones.
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

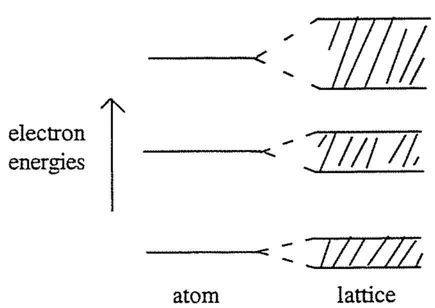
www.curriculum-press.co.uk Number 102

The Physics of Semiconductor Diodes

Electricity used to be simple. Some materials conducted electricity, and others were insulators. Unfortunately someone then discovered another type of material – the semiconductor.

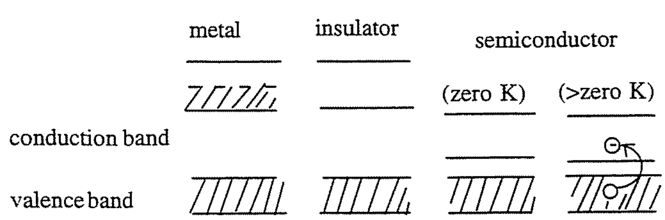
These materials sometimes act as insulators, and sometimes as conductors. And we have learned to alter and combine them to form a large number of devices that can control electrical circuits in many ways. So perhaps the discovery of semiconductors wasn't all bad.

Intrinsic Semiconductors:
For an atom, we talk about allowed energy levels for the electrons. However, for a solid, these energy levels broaden out into bands:



Conductors have their outer band (the conduction band) partially occupied. These electrons are available to carry current. With an insulator, the conduction band is empty.

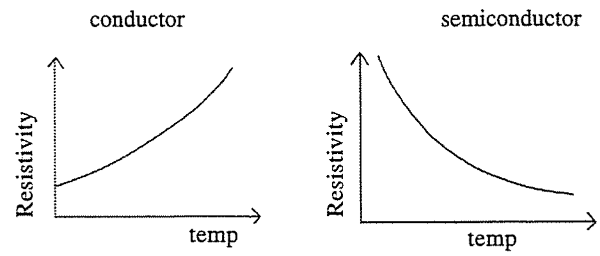
A semiconductor (at absolute zero of temperature) is similar to an insulator. However the gap between the two bands is quite small. As the temperature rises, more and more electrons gain enough thermal energy to jump the gap into the conduction band, and the semiconductor can carry a small amount of current.



Each **electron** in the conduction band is matched by a **hole** in the valence band. This **hole** acts as a mobile positive charge, adding to the conductivity of the material.

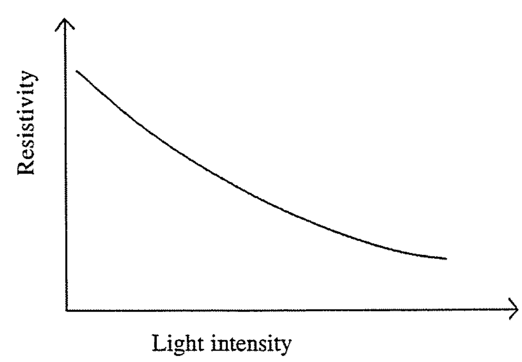
> *Example*
> ***As the temperature rises in the semiconductor, why don't all the electrons gain a tiny bit of energy – not enough to jump the gap?***
> ***Answer:*** *Things don't work that way. Random collisions, etc, cause a spread of energies. The most energetic electrons will jump into the conduction band.*

This thermal effect leads to very different resistance effects in conductors and semiconductors as the temperature rises.



> *Example*
> ***In one sentence each, explain the trends noticed in these two graphs.***
>
> *Answer*
> *With a metal, increased lattice vibration makes it more difficult for the conduction electrons to move through the lattice. With a semiconductor, a rise in temperature enables more and more electrons to jump into the conduction band, and the resistivity falls.*

With some semiconductors, shining visible light onto the material gives some electrons enough energy to jump into the conduction band. The more light, the more electron-hole pairs are formed. This causes an effect on resistivity similar to that of rising temperature.

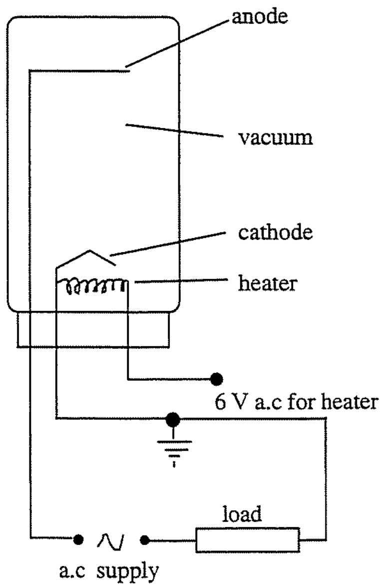


> *Example*
> ***Name the two semiconductor devices that make use of these effects of temperature and light on resistivity.***
>
> ***Answers:*** *Thermistor and light dependent resistor (LDR).*

> ***Exam Hint:*** *Be prepared to draw graphs for* ***resistance v temperature*** *(as above) and* ***current v voltage*** *for a thermistor. Remember that increased voltage causes increased current, which leads to a heating effect. Be prepared to draw relevant graphs for an LDR as well.*

Diode Valves

A diode is a device that only allows a current to flow through it in one direction. We had diodes long before semiconductors were developed;



In this diode valve, a 6V supply causes the heater to glow and make the metal cathode hot. Electrons are freed from the heated cathode and attracted through the vacuum to the positive anode, completing the circuit.

However, when the anode goes negative, it repels the electrons freed from the heated cathode. There is no current flow around the circuit.

This device worked well, but there were many weaknesses compared to the semiconductor diode. Think of some - there is a question about this at the end of the Factsheet.

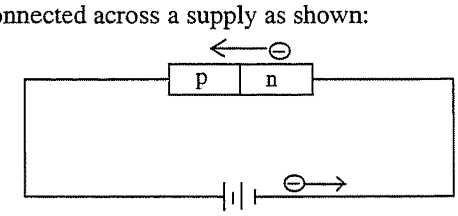
Semiconductor (p-n) diode:

Intrinsic semiconductors can be **doped** with impurity atoms which give them either many extra **conduction electrons** (n-type) or **holes** (p-type). They are now called **extrinsic** semiconductors, and can carry far more current.

> ***Example : What do the n and the p stand for?***
> ***Answer:*** *Negative and positive obviously. A totally unnecessary question.*

However the real magic happens when the two types of semiconductors are bonded together. At the junction, electrons from the n-type cross over to fill holes in the p-type. A potential difference is formed across this **depletion layer**, stopping any further flow of charge.

The only way that a continuous current can flow through this device is if it is connected across a supply as shown:

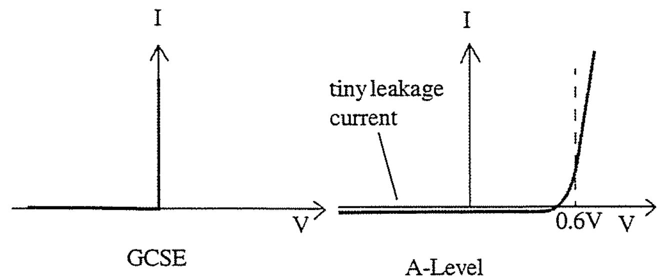


The applied p.d. can force electrons across the depletion layer. The p.d. that must be overcome to enable a current flow is about 0.6 volts for silicon semiconductors.

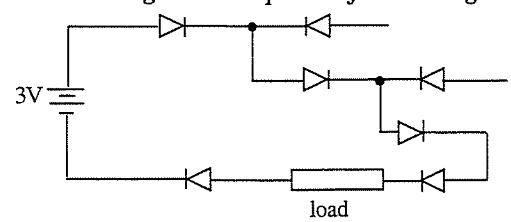
If the supply is reversed, only a tiny leakage current flows. This leakage current is caused by continuous production of electron-hole pairs from ambient thermal energy. We call the two set-ups **forward biased** and **reverse-biased.**

> ***Example: What necessary component has been left out of this circuit?***
> ***Answer:*** *A series resistor to protect the diode. If the applied p.d. is above 0.6 volts, the diode basically acts as a short circuit.*

> *Notice from the diagram below the difference between the simplistic understanding of the junction diode from GCSE Physics, and the required understanding at A-level.*



> ***Example***
> ***Antonio and Esmerelda design a circuit using diodes to direct the current along a certain path to flow through the load:***



> ***What is the difficulty here?***
> ***Answer:*** *They have forgotten that there is a drop of 0.6 volts across each diode. The supply won't be able to drive any current through the load.*

Specially designed diodes

In addition to the standard junction diode, other diodes have been designed for specific purposes.

Light emitting diode

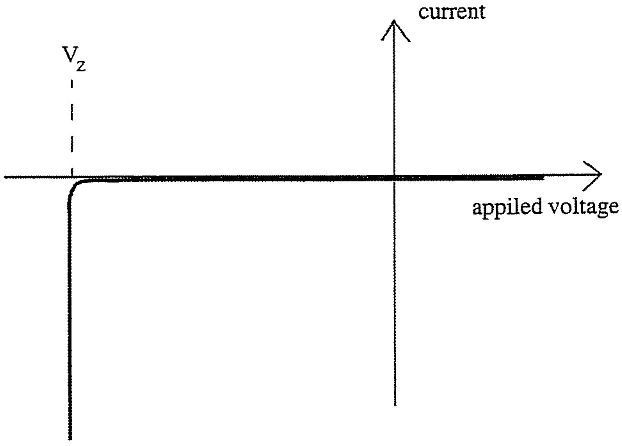
LED's are composed of selected semiconductor materials where the energy released when electrons and holes combine at the junction is given off as visible light. The colour of the light emitted depends on the semiconductor material chosen.

Photodiode

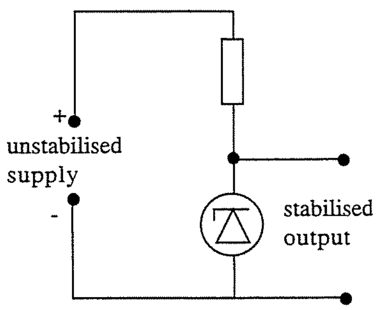
We have mentioned that thermal energy causes a tiny current when the diode is reverse-biased. If light of the right frequency is directed onto the depletion layer, it can also produce extra electron-hole pairs. So the current through a **reverse-biased** photodiode will depend on the amount of light shining onto it (assuming the light is of the correct frequency).

Zener diode

Only the tiny leakage current should flow through a reverse-biased diode. But at a certain critical voltage (V_Z), the diode breaks down:



This voltage is sometimes called the Zener voltage, and diodes can be manufactured to have different values for this. A reverse-biased Zener diode provides a stabilised output voltage from a supply which may not be quite so stable.



Practice Questions

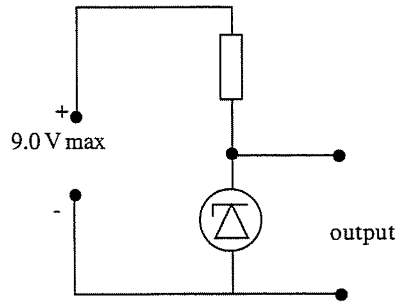
1. Complete this table for a metal and a semiconductor **as the temperature rises**:

	Metal	**Semiconductor**
Number of conduction electrons		Increases
Amplitude of lattice vibrations	Increases	
Overall effect on resistance		Decreases

2. Name several advantages of semiconductor diodes over diode valves.

3. An LED is used in a circuit as a warning light. the circuit has a 12 volt supply and the LED requires 8mA to operate at its intended brightness.
   (a) What p.d. is dropped across the LED itself (assume the same properties as a silicon junction diode)?
   (b) What voltage must be dropped across the series protective resistor?
   (c) What is the value of this resistor?

4. A reverse-biased Zener diode provides a stable voltage of 2.7 volts. The unstabilised power supply across it has a maximum p.d. of 9.0 volts.



   (a) Find the value of the protective resistor, R, if the Zener diode cannot safely dissipate a power of more than 0.10 watts.
   (b) Find the current through a 500 ohm load attached across the output.

Answers

1.

	Metal	**Semiconductor**
Number of conduction electrons	Fixed	Increases
Amplitude of lattice vibrations	Increases	Increases (but this effect is outweighed by a massive increase in conduction electrons and holes)
Overall effect on resistance	Increases	Decreases

2. Use less power, portable devices, physically smaller, less fragile, longer lasting, cheaper to make and use, faster response time (higher frequencies), etc.

3. (a) 0.6 volts
   (b) 11.4 volts
   (c) $R = V / I = 11.4 / 0.008 = 1400$ ohms.

4. (a) Diode: $P = 2.7 \times I$, $I = 0.10 / 2.7 = 0.037A$
   Resistor: $V_{max} = 9.0 - 2.7 = 6.3$ volts
   $R = V / I = 6.3 / 0.037 = 170$ ohms

   (b) $I = V / R = 2.7 / 500 = 0.0054A$ (this will reduce the current through the Zener diode, but won't affect the output voltage)

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU
Physics Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber.

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 91

Accuracy in Electrical Investigations

This Factsheet is not meant to be a survey of practical investigations concerning electricity, or to provide complete methods for these investigations.

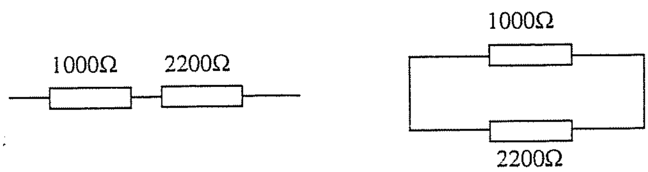
It is instead intended to look at the causes of inaccuracy in electrical work, and to suggest some techniques for minimising and measuring inaccuracy.

Resistors and Combinations of Resistors

Standard resistors (e.g. carbon film) are manufactured to stated tolerances (maximum percentage errors) often described by a coloured band on the resistor. A gold band means 5% maximum error.

> *Example 1: What is the maximum possible error, in ohms, for a 22kΩ resistor with a stated tolerance of 5%?*
>
> *Answer:* $\frac{5}{100} \times 22000 = 1100\Omega$.

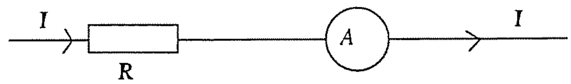
Often we must combine resistors to achieve required values. It is worth seeing the effect on the percentage error of such combinations.

> ***Example 2: A 1000Ω resistor and a 2200Ω resistor (each of 5% tolerance) are wired in series, then in parallel. Find the maximum percentage error in each combination.***
>
> 
>
>
> *Answer:*
> *Series:*
> *Maximum value* $= (1000 \times 1.05) + (2200 \times 1.05) = 3360\Omega$.
> *Nominal value* $= 1000 + 2200 = 3200\Omega$.
> *Percentage error* $= \frac{160}{3200} \times 100 = 5\%$.
>
> *Parallel: Maximum value:*
> $\frac{1}{1050} + \frac{1}{2310} = 1.385 \times 10^{-3}$, $R = 722\Omega$.
> *Nominal value:* $\frac{1}{1000} + \frac{1}{2200} = 1.455 \times 10^{-3}$, $R = 688\Omega$.
> *Percentage error = 5% again.*

> *Combining resistors with the same tolerance should lead to an effective resistance with the same tolerance. It should be noted that these are maximum percentage errors. In practice, some of the errors might well cancel each other. The effective resistance of the combination would probably be closer to the nominal value. But there is no guarantee of this.*

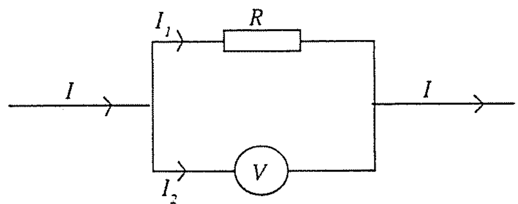
Real and Ideal Meters

An ammeter is placed in series with a resistor. An ideal ammeter has **zero resistance**, and does not reduce the current flow.

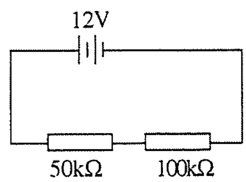
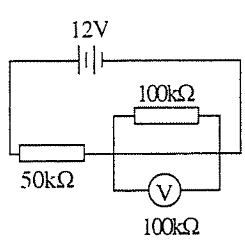


In practise, the ammeter will have some resistance. But in modern electronics, current flow is tiny (implying a large circuit resistance), so the resistance of the ammeter is not too important.

However, voltmeters are placed in parallel to the component being measured, leading to real problems at times.



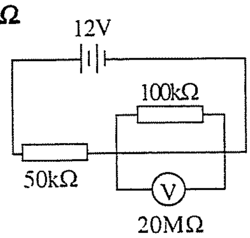
The ideal voltmeter would have **infinite resistance**, drawing no current away from the main circuit. But a moving coil meter requires a current through it to produce a deflection. If the current flow through R is very small, the current sidetracked to the parallel voltmeter could be significant, and lead to a measurable error in the voltage reading.

> *Example 3 (a): Look at this circuit:*
>
> 
>
>
> ***Find the p.d. across the 100kΩ resistor.***
>
> ***(b) Compare this second circuit.***
>
> 
>
>
> ***Now find the p.d. across the 100kΩ resistor.***
>
> **Answer:**
> *(a)* $I = \frac{12}{150\,000} = 8.0 \times 10^{-5}$ A.
> $V(\text{across } 100k\Omega) = IR = 8.0 \times 10^{-5} \times 1.0 \times 10^{5} = 8V$.
> *(b) Parallel:* $\frac{1}{100\,000} + \frac{1}{100\,000} = \frac{2}{100\,000}$
> $R = 50k\Omega$ *(same as other resistor)*
> *V(across parallel combination) = 6V.*

Introducing the moving coil voltmeter across the 100kΩ resistor has changed the p.d. across it from 8V to 6V.

The solution to this voltmeter problem is to use a **digital voltmeter** in electronic circuits. These meters can have effective resistances up in the megohm region.

Example 4: This circuit uses a digital voltmeter with an input resistance of 20MΩ.



Find the p.d. across the 100kΩ resistor.

Answer:

Parallel: $\frac{1}{1.0\times10^5} + \frac{1}{2.0\times10^7} = 1.005\times10^{-5}$, $R = 9.95\times10^4\ \Omega$

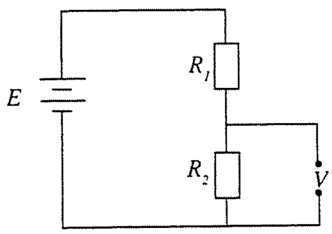
$I = \frac{12}{(5.0\times10^4 + 9.95\times10^4)} = 8.03\times10^{-5}$ A.

V(across parallel combination) = IR = $8.03\times10^{-5}\times9.95\times10^4 = 8.0$ V.
(the same as without the meter in the circuit)

Exam hint: *When doing practical work, always consider the effects of the measuring instruments on the values being obtained. High input impedance devices, like digital voltmeters and cathode ray oscilloscopes, can significantly increase accuracy.*

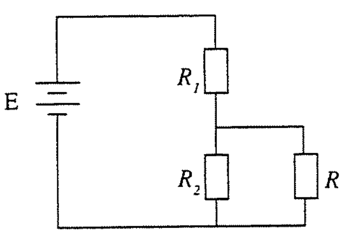
Potential Dividers

A potential divider allows us to produce an adjustable output p.d. which is less than the fixed input p.d. from the supply.



The output p.d. is given by $V = \frac{R_2}{R_1 + R_2}\times E$.

However the current drawn by the load can affect the size of the output voltage.

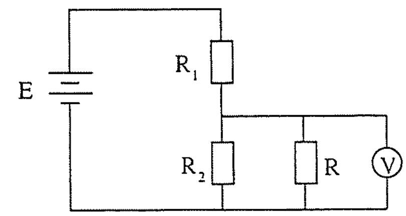


The load, R, now forms a parallel branch with R_2, lowering the output voltage. (This theory is exactly the same as in the previous situation where the voltmeter is in parallel with the 100kΩ resistor.)

To minimise this effect, R_2 must be much smaller than the load, R. However, if R_2 is very small, then the current and power dissipation through R_1 and R_2 will be large.

This can lead to Joule heating problems and is very wasteful of power.

Calculations with potential dividers are only accurate if the load is much larger than the resistances in the divider itself. In practical situations, a voltmeter must be used to provide an accurate value for the output p.d.



Resistance and Temperature

We tend to assume that resistors have fixed resistances. But their resistance values increase with temperature. As an example, for copper the temperature coefficient of resistance can be approximately given as:

$$\alpha = \frac{\Delta R}{(R_0 \Delta\theta)}$$

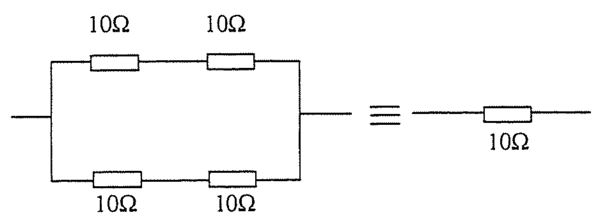
where R_0 is the resistance at 0°C
Δθ is the temperature rise.

Example 5: What is the percentage increase in the resistance of a length of copper wire when its temperature goes up by 25°C? (α = 0.0043°C⁻¹)

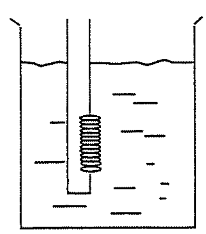
Answer: $\frac{\Delta R}{R_0} = \alpha\Delta\theta = 0.0043 \times 25 = 0.11$
The percentage increase is 11%.

Joule heating causes significant changes in resistance in metallic conductors.

There are various ways of minimising Joule heating effects:
(a) keep the current very small.
(b) ensure adequate ventilation or use a cooling fan
(c) use physically large components, or combinations to share the heating effect.

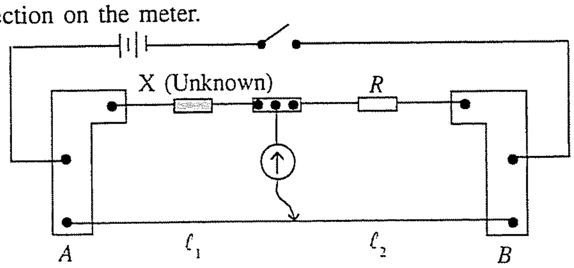


(d) use cooling e.g. immersing a resistance wire in water.



Metre bridges

Wheatstone bridges and metre bridges are very useful in making accurate determinations of resistance and of the emf of cells. Their accuracy relies on the fact that at the balance point, there is no deflection on the meter.



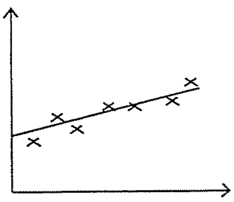
Null deflection means no errors due to current flowing through the meter. In our diagram, $\frac{X}{R} = \frac{l_1}{l_2}$

However there can still be sources of inaccuracy. At *A* and *B* there can be contact resistances. These add to the resistances of wire lengths l_1 and l_2. These end errors are most significant if l_1 or l_2 is very small.

> *When using metre bridges, make R approximately equal to X. This will give a balance point near the middle of the wire, reducing the effect of end errors on l_1 and l_2.*

Graphical work

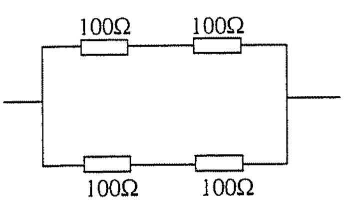
Quite often, we can use graphical work to improve accuracy, making use of best fit lines.



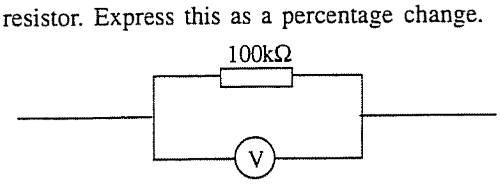
This is, of course, applicable to all fields of physics and other sciences. But it is always best to improve the accuracy of the investigation itself. Don't rely on graphical and other techniques to sort out experimental weaknesses.

Questions

1. Use the method of maximum and minimum values to find the maximum percentage error for this combination. The tolerance of each resistor is 5%.



2. (a) A voltmeter draws a current of 0.1mA when a p.d. of 12V is applied across it. Find its resistance.
   (b) Assume the resistance it exerts is a constant. Find the change in effective resistance when it is connected across a 100kΩ resistor. Express this as a percentage change.

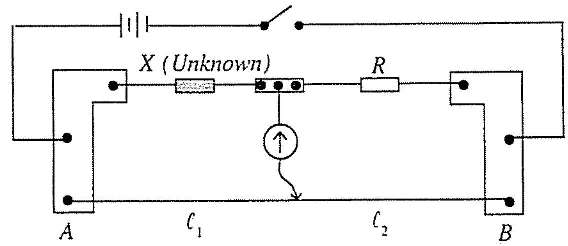


   (c) Find the percentage change in effective resistance when the voltmeter is connected across a 100Ω resistor.

3. Assume the temperature coefficient of resistance of copper is 0.0043^oC^{-1}.
   (a) Find the increase in resistance when a length of copper wire of resistance 0.24Ω at 0°C is heated from 10°C to 45°C.
   (b) Express this as a percentage change.

4. If you decided to reduce Joule heating in a circuit by using very thin resistance wire to reduce the current flow, suggest two other factors you would have to consider (in terms of accuracy).

5. Here is another look at a metre bridge.



The wire has a total resistance of 6.0Ω. There is a contact resistance of 0.010Ω at A and at B. The wire is 1.00m long.
   (a) If the balance point is at the centre of the wire, find the resistances of l_1 and l_2.
   (b) Find the percentage errors in Rl_1 and Rl_2 due to the contact resistances.
   (c) Find the percentage error in $\frac{Rl_1}{Rl_2}$.
   (d) If the unknown resistance is changed, and the balance point is just 1.0cm from A, find the resistances of l_1 and l_2.
   (e) Again find the percentage errors in Rl_1 and Rl_2.
   (f) Now find the percentage error in $\frac{Rl_1}{Rl_2}$

Answers:

1. Nominal value: Series 200Ω in each branch, then Parallel gives R = 100Ω.
   Maximum value: Increase each resistor to 105Ω. Find that effective R = 105Ω.
   Maximum percentage error is 5Ω for the combination, as well as for each resistor.

2. (a) R = V/I = 120 000Ω.
   (b) 1/R = 1 / 100 000 + 1 / 120 000. R = 54 500Ω.
   Percentage change = 46%.
   (c) 1/R = 1 / 100 + 1 / 120 000. R = 99.9Ω
   Percentage change = 0.1%.

3. (a) $DR = \alpha R_0 \theta = 0.036\Omega$
   (b) Percentage change = (0.036 / 0.24) × 100 = 15%.

4. Examples might be (a) increased percentage errors from electrical measuring instruments when dealing with very small currents, and (b) increased percentage errors in measuring the diameter of the wire.

5. (a) $l_1 = 3.0\Omega$, $l_2 = 3.0\Omega$.
   (b) % error l_1 = % error (l_2 = (0.010 / 3.0) × 100 = 0.33%
   (c) % error l_1 / l_2 = 0.66%
   (d) l_1 = 1/100 × 6.0 = 0.060Ω
   l_2 = 99/100 × 6.0 = 5.94Ω
   (e) % error l_1 = (0.010 / 0.060) × 100 = 16.7%
   % error l_2 = 0.17%
   (f) % error l_1 / l_2 = 16.9%.

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU
Physics Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber. ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 73

Why Students Lose Marks- Current Electricity

Shall we start with this question?
A 24 volt a.c. supply provides power at an efficiency of 37% to an immersion heater running for 3 hours per day and a freezer unit active for 3.5 hours per day except Sunday. Given that the supply provides a total energy of 1040 kWh over the period of a normal year and has an internal resistance of 6.4 ohms, what advantage would be gained by switching to a 48 volt supply with greater efficiency but a larger internal resistance?

Don't fancy it? (Neither do I.)

This Factsheet will not teach you how to gain full marks on impossible problems (or even very difficult ones). We all struggle with work we find difficult. But what we don't want to do is lose marks on easy questions.

We are going to concentrate in this Factsheet on two types of question commonly found within this topic – calculations and graphical work.

Let's look at calculations first. And because we are not concerning ourselves with covering the syllabus, but with our approach to mathematical problems, we will restrict ourselves to straightforward work on standard electrical circuits.

Calculations:

We can maximise our results by attacking these problems in a straightforward manner, proceeding step by step through the calculation.

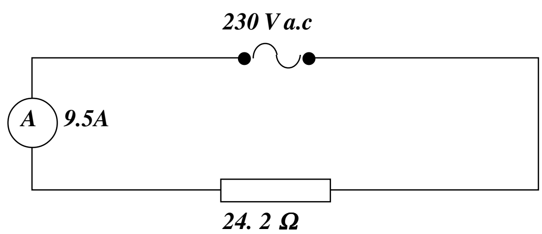
1. Make a list of the information given, and the information you are asked to work out. Draw a circuit diagram if you think that will help.
2. Write down the equations which might be useful.
3. Identify the equation you think will work. Write it down.
4. Make sure the data is in the correct units. Now solve the problem.
5. Finally, state the final answer to a sensible number of significant figures. And write down the correct unit.

Now take a look at your answer. Is it reasonable? (Does your CD player really draw a current of 500 amps? Perhaps you had better look at your calculation again.)

And remember, even if your calculation has gone wrong, if you have written down the equation, transformed it, substituted the data, stated the correct unit, etc, you may gain almost all of the marks anyway.

If all you write down is the final answer, you are taking a big chance.

Example 1 A mains (230V a.c.) electric heater draws a current of 9.5 amps. The heating element has a resistance of 24.2 ohms. If all of the electrical power is transformed into heat, what thermal energy will be produced in 12 minutes?



Solution:
What do we know?
$V = 230V$ (a.c.)
$I = 9.5$ amps
$R = 24.2$ ohms
$t = 12$ minutes
$E = ?$
$P = ?$ (but may not be needed – energy, not power, is asked for)

Relevant equations:
$V = IR$
$P = VI$ (or $P = I^2R$)
$E = Pt$
$E = VIt$

We are given V, I and t, so it looks like the equation to use is E=VIt. It doesn't need rearranging, but notice that the time has been given in minutes.

$E = VIt = 230 \times 9.5 \times (12 \times 60) = 1\ 573\ 200$ joules

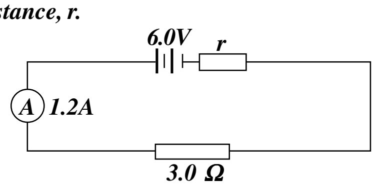
$E = 1.6 \times 10^6$ joules (to 2 s.f.)

A joule is a tiny amount of energy, so this is reasonable.

Our answer is correct, earning full marks. But even if we had, for example, forgotten to change minutes into seconds, we could still end up gaining most of the marks available for this calculation.

Sometimes you can't do everything in one step. But the same logical approach will still get you through to the answer.

Example 2: A cell of e.m.f. 6.0 volts sends a current of 1.2 amps through an external resistor, R, of 3.0 ohms. Find the power dissipated in the internal resistance, r.



Solution:

Write down the knowns and unknowns.
E = 6.0 volts
R = 3.0 ohms
I = 1.2 amps
r = ?
P = ? (power dissipated in internal resistance, r)

Relevant equations:
$E = IR + Ir$
$V = Ir$ *(voltage across r)*
$P = VI = I^2r$ *(power dissipated in r)*

We don't know V or r. But we should be able to find r from the first equation, then find P from the third equation.

$E = IR + Ir$
$6.0 = 1.2 \times 3.0 + 1.2 \times r$
This works out to r = 2.0 ohms.

Using $P = I^2r$
$P = 1.2^2 \times 2.0 = 2.88$ *watts.*

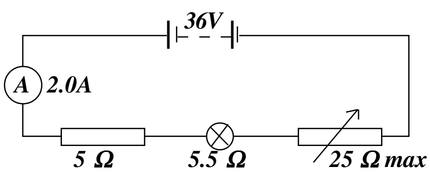
P = 2.9 watts (to 2s.f. and using the correct unit)

This seems a reasonable value for the power dissipated in the internal resistor.

And again we have shown all of our working, checked s.f. and units, and looked at the answer to see if it is reasonable.

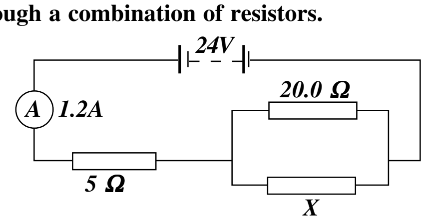
Here are a couple of questions for you to try. Work through them in this same way. Then check your solutions against those at the end of the Factsheet.

Problem 1:
A lamp requires a current of 2.0 amps to operate at its recommended brightness. Its resistance, X, is 5.5 ohms when it is operating normally. It is in series with a 36 volt d.c. power supply, a 5 ohm fixed resistor, R1, and a 25 ohm variable resistor, R2. The circuit is shown below:



(a) At what resistance should the variable resistor be set in order for the lamp to work normally?
(b) And what is the rated power of this lamp?

Problem 2
The circuit below shows a 24 volt battery supplying a current of 1.2 amps through a combination of resistors.



(a) Find the voltage across the parallel resistors.
(b) Find the current through unknown resistor X.
(c) Find the value for the resistance of X.

Now let's look at graphs.

Graphical Work

There are a lot of easy marks available for graphical work, especially on practical exams. But it's also very easy to lose these marks, as the standard demanded is far higher at A-level than at GCSE.

Here are some ideas to keep in mind:

1. Read the instructions and construct the axes the right way round. Do you think A-level students would never make such a simple mistake? You'd be in for a shock. Strange things can happen in an exam situation.
2. Don't forget to label the axes with quantities and units. **It's especially easy to forget the units.** And remember, logarithms of quantities don't have units. You can write **log (I/A)**, but not **log I/A**.
3. Use the range of the data compiled to work out the scales. If you can double the scale and still fit all the points on the graph, or if you use multiples of 3 or 7, for example, you could end up losing the scale mark and perhaps the plotting marks as well.

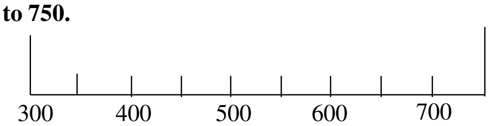
Example: Your graph paper has 90 small squares (probably 2 mm each) across the x – axis. Your data points range from 320 to 560. What scale would you choose?
(a) 0 to 900.

Not acceptable. Your range of values covers much less than half of this.

(b) 300 to 600.

Not acceptable. Thirty small squares on the paper would cover a range of values of one hundred. And plotting the data would be a nightmare.

(c) 300 to 750.



A good choice. The range of data really requires scale points from 300 to 600. This easily covers over half of your chosen scale, and the data would be easy to plot.

4. Plotting the points with small crosses is perhaps most suitable. Blobs will lose you marks. Small dots can easily be lost under best-fit lines. And circles are best saved for displaying anomalous data points.

5. Don't be careless when drawing best-fit lines. They may well be curves, rather than straight lines. Precision is demanded at A-level. And make sure gradient triangles cover at least half of your best-fit line.

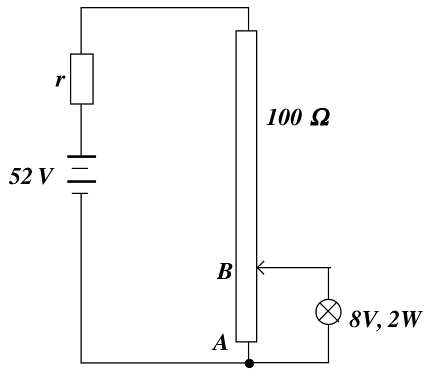
Now the final test. Here is a problem for you. Use graph paper of 9 squares against 11 squares to plot this (use part of a larger sheet if necessary). And check your graph against the suggested solution at the end of the Factsheet. But remember that there is more than one acceptable way of setting out a graph.

Problem 3: Use this data to plot the current through resistor R (y-axis) against the voltage across the resistor (x-axis). Then find the resistance of R.

Voltage/V	*Current/A*
1.0	*0.07*
2.8	*0.22*
4.2	*0.30*
6.4	*0.48*
8.1	*0.63*
10.1	*0.75*
12.2	*0.94*

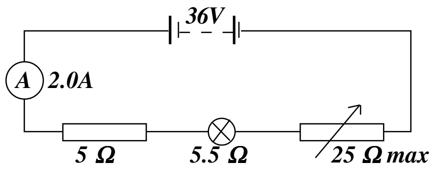
Finally, for those of you who have decided that Physics problems are little more than trivial, here's a parting question for you:

Problem 4: A potential divider of length 20 cm has a resistance of 100 ohms. It is used to illuminate a lamp designed to operate at a power of 2 watts when the potential difference across it is 8 volts. The supply is fixed at 52 volts d.c.



If the distance from A to B is 5 cm, what is the internal resistance, r, of the supply?

Problem 1 Solution:



(a) What do we know (and what would we like to know)?
V=36V
I=2.0 amps
X=5.5 ohms
R1=5 ohms
R2=? (25 ohms maximum)

Relevant equations:
V=IR
R=R1 + R2 + X (for series circuit)

To find the value of the variable resistor, X, we must first find the total resistance R (from Ohm's Law).

R=V/I=36/2.0=18 ohms.

Substituting, we have 18=5 + R2 + 5.5

So R2 = 7.5 ohms. (To 2s.f.) And we have remembered the unit.

This is within the range for the variable resistor.

(b) We now know that R2 is set at 7.5 ohms, in addition to the previous data.

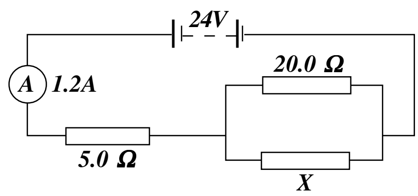
Relevant equations:
P=VI
P=I^2R

We have to be careful here. The current I is the same everywhere around a series circuit. But we don't know the voltage across the lamp itself. The first equation might not be useful. However we do know the lamp's resistance. This is 5.5 ohms. So we can use the second equation:

P = $I^2R = (2.0^2 \times 5.5)$ *= 22 watts. (Again 2s.f. is reasonable, and the unit is correct.)*

And once again, a simple logical approach, writing down each step, and checking significant figures and units, will mean we don't throw away any easy marks.

Problem 2 Solution:



(a) What do we know?
V=24 volts (supply)
I=1.2 amps (through ammeter)
Resistors of 5.0 ohms and 20.0 ohms
X=?

Relevant equations:
R = R1 + R2 (series)
1/R = 1/R1 + 1/R2 (parallel)
V = V1 + V2 (series)
V = IR

We should be able to do this by just adding the voltages across the two sets of resistors. We can use Ohm's Law to find the voltage across the 5.0 ohm resistor.

V = IR = 1.2 × 5.0 = 6.0 volts

Then the third equation will let us find the voltage across the parallel section of the circuit:

V (parallel) = 24.0 – 6.0 = 18.0 volts.

This seems reasonable. Units and s.f. are OK.

(b) What do we know? (concentrate on the parallel section)
V (supply) = 24.0 volts
V (parallel) = 18.0 volts
R = 20.0 ohms (in parallel branch)
I = 1.2 amps (through ammeter)
(With Ohm's Law, this might be enough.)

Relevant equations:
V=IR
I = I1 + I2 (in parallel circuit)

We should be able to use Ohm's Law to find the current through the 20.0 ohm resistor, then use the second equation to find the current through X.

I = V/R = 18.0/20.0 = 0.9 amps (through the 20.0 ohm resistor)

I = 1.2 – 0.9 = 0.3 amps through resistor X.

This seems reasonable. We can't really add any more s.f. to this value, as the current has only been given to one decimal place.

(c) What do we know? (again concentrate on the parallel section)
V(parallel) = 18.0 volts
I = 0.3 amps (through X)

We can obviously jump straight to Ohm's Law for our answer.

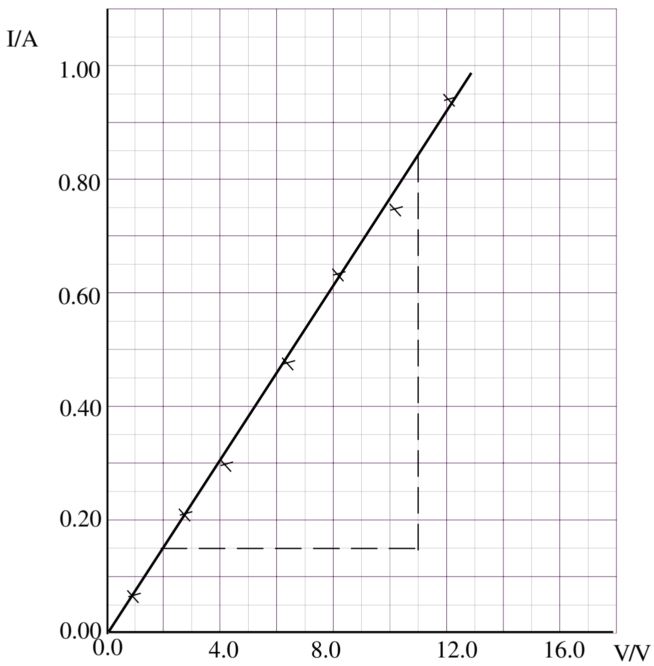
R = V/I = 18.0/0.3 = 60 ohms.

We could check this answer by using our series and parallel resistor formulae to find the effective value of R for the whole circuit, then seeing that V = IR works for the circuit. It does. And we have remembered to include the unit.

We can't really add any more s.f. to our answer as we have been using a value of current calculated to just one s.f. in our solution.

And once again, by using a simple approach, writing everything down along the way, we have given ourselves the best chance of maximising our score.

Problem 3 Solution:

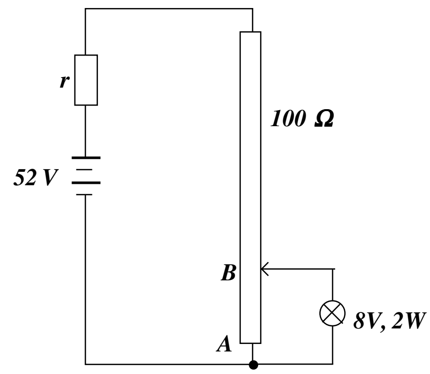


$$\text{Gradient} = \frac{(I_2 - I_1)}{(V_2 - V_1)} = \frac{(0.84 - 0.15)}{(11.0 - 2.0)} = 0.077$$

$$R = \frac{1}{\text{Gradient}} = 13.0\ \Omega.$$

What would you have scored out of ten?
Axes correct? 1
Quantities and units? 1
Acceptable scale? 1
Points correct? 2
Best-fit line? 1
Good presentation? 1
Gradient triangle? 1
Gradient correct? 1
Value and unit? 1
Anything less than 9.5 is unacceptable.

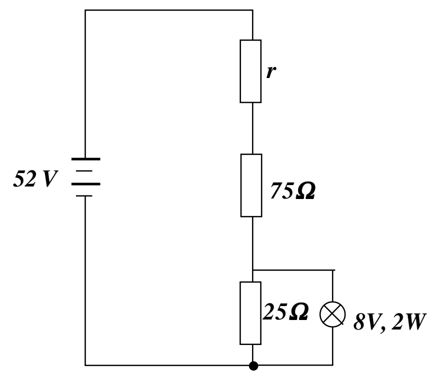
Problem 4 Solution:



Parts of this problem can be solved quickly by common sense. We can see straightaway that AB is one-quarter of the total length of the potential divider, so the resistance across AB will be 25 ohms.

And using $P = V^2/R$, we can find the resistance of the lamp to be 32 ohms (try this yourself).

It might now be useful to redraw the circuit. Don't be afraid of redrawing circuits or diagrams when you have determined values that might be usefully displayed. An up-to-date diagram can help you carry on with the problem:



Also for the lamp:
P = 2 watts
V = 8 volts

From this we know that the top section has a voltage of 52 – 8 = 44 volts across it.

It looks like one way to solve this will be by linking the current through the top section to the sum of the currents in the parallel branches below:

Top section: $I = \frac{V}{R} = \frac{44}{(75 + r)}$ *amps*

Left branch: $I = \frac{V}{R} = \frac{8}{25} = 0.32$ *amps*

Right branch: $I = \frac{V}{R} = \frac{8}{32} = 0.25$ *amps*

As the currents in the parallel branches must add up to the total current:

$$\frac{44}{(75 + r)} = 0.32 + 0.25 = 0.57$$

$44 = 0.57 \times (75 + r)$
$44 = 42.8 + 0.57r$
r = 2.1 ohms (to 2 s.f.)

And so by looking at the diagrams and using some common sense, added to some basic Physics relationships, we can find the internal resistance of the supply.

Acknowledgements: *This Physics Factsheet was researched and written by Paul Freeman*
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU. Ph

Physics Factsheet

www.curriculum-press.co.uk Number 116

Graphical Work in Electricity

Principal Examiners have identified weaknesses in graphical work within Physics – especially in the interpretation of sketch graphs, and calculations based upon these. In general, students prefer to work with numbers rather than general ideas, but often the general principle is more important. In this Factsheet, we will concentrate on graphical work within electricity.

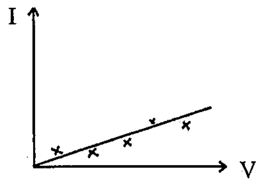
Electrical components

Much of this work goes back to GCSE Physics, but it can still lead to problems at A-level.

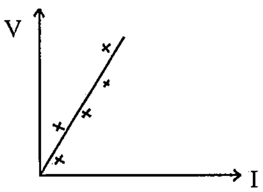
> *Exam Hint: Don't assume that you remember in detail the properties of components, etc, that you have studied in previous years. Your knowledge will almost certainly be a bit fuzzy, and, without careful revision, you could be caught out in an exam situation.*

(a) Metallic resistors at constant temperature

The difficulty here is that graphs can be presented in two ways. In a typical investigation, the p.d. is the independent variable, so a typical graph might resemble:



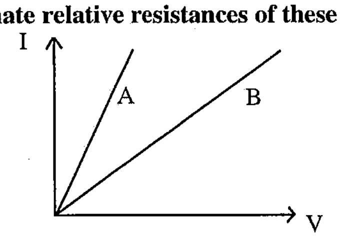
However these graphs are often presented with the axes reversed.



The reason for this is that the gradient of this graph is the most accurate way of determining the resistance - from Ohm's Law (R=V/I).

> *Exam Hint: A voltage-current graph can be presented with either quantity along the X-axis. Take a close look before you start doing calculations or drawing conclusions.*

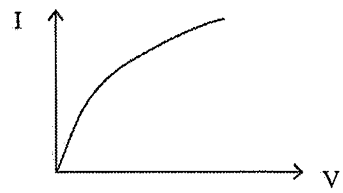
Example 1: The V-I graphs for resistors A and B are sketched below. Estimate relative resistances of these two components.



Answer: Notice that these graphs have V along the X-axis. For a fixed current, the p.d. across B is about three times the p.d. across A. So the resistance of component B is about three times the resistance of A.

(b) Metallic resistors at varying temperature

The typical graph here is that of a lamp filament as the current through it is increased (and its temperature rises).



You should be able to make some observations about what this graph shows, and use your knowledge of Physics to explain these effects.

Notice:
(i) For very small values of current, the graph is a straight line.

(ii) As the current becomes greater, the line curves and starts to level out.

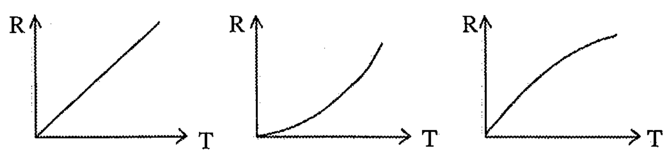
(iii) The line **never** bends back down again.

What can we deduce from these observations?
(i) For very small values of current, the resistance of the filament remains constant. The filament stays at constant temperature.

(ii) As the current increases, the resistance of the filament increases. The filament's temperature is increasing from Joule heating.

(iii) The temperature (and resistance) continue to rise as the current increases.

Example 2: From the V-I graphs of the lamp filament, determine which of these graphs best represents how the resistance changes as the temperature of the filament increases.



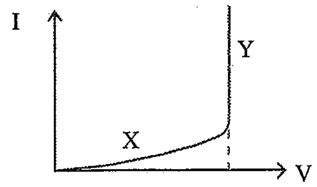
Answer: The V-I graph doesn't give us the information to determine this. We can't tell from the V-I graph how quickly the temperature is rising. You would need to monitor the temperature as well as the current, as you increased the p.d.

From your knowledge of lattice vibrations and resistance, you might be able to hazard a guess as to whether the resistance would level off or head towards infinity.

> *Exam Hint: You can only draw a certain amount of information from a simple graph. In an exam situation, only make inferences that you can justify.*

(c) Thermistor (NTC)

The V-I graph for a negative temperature coefficient thermistor might resemble this:



We can see:

(i) In region X, the current increases at an increasing rate. The resistance must be slowly decreasing.

(ii) In region Y, the current increases dramatically (with no increase in p.d.). The resistance must suddenly drop very quickly.

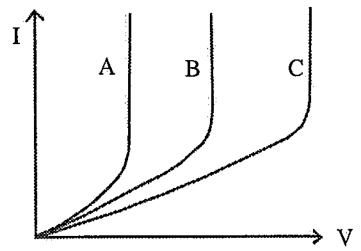
How can we explain this?

(i) With small currents, there is a small rise in temperature due to Joule heating. The shape of the graph shows the resistance is falling as the temperature rises. However, at any voltage, the temperature stays stable.

(ii) When we reach a critical point, the increasing temperature causes the resistance to drop and the current to rise. However the ambient cooling of the air around the thermistor is unable to stabilise the temperature. Joule heating causes the temperature to keep rising, causing the resistance to keep dropping, and causing the current to rise even more, leading to an avalanche effect.

Example 3: Sketch possible V-I graphs, on the same set of axes, for this thermistor when placed:
(a) in a vacuum, (b) in air, and (c) in water. Briefly explain your graphs.

Answer:

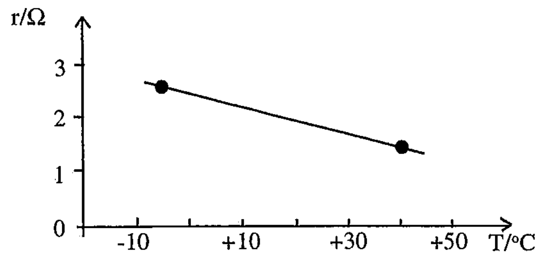


There is most cooling in water, and least in a vacuum. The critical temperature would be reached at a higher current (higher voltage) when the cooling is greater.

Internal Resistance in Cells

All cells have an internal resistance. The value depends on many factors – the type of cell, how much of its chemical constituents have been used up, its temperature, etc. We will have a brief look at the sort of thing that can be deduced from graphs.

The internal resistance of a lead-acid battery depends on its temperature.



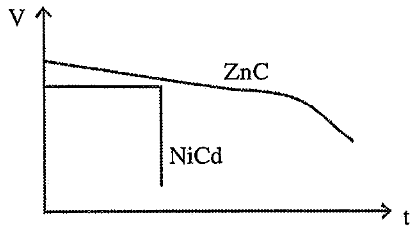
Example 4: Given an emf of 12 volts for this battery, find the voltage dropped, and power dissipated, across the internal resistance when a current of 2A is drawn
(a) at –5°C
(b) at 40°C.

Answer:
(a) $r = 2.5$ *ohms,* $V = IR = 2 \times 2.5 = 5V$, $P = I^2R = 10W$.
(b) $r = 1.5$ *ohms,* $V = IR = 2 \times 1.5 = 3V$, $P = I^2R = 6W$.

Example 5: Describe what you would expect to happen, as time passes, when a load of 3.5ohms is placed across this battery at –5°C.

Answer: Total resistance is 6.0 ohms. So current I = 2A. But as time passes, the power dissipated would warm up the battery. The internal resistance would fall, and the current would rise.

The sketch below displays the output voltages against time for a ZnC cell and a NiCd cell (rechargeable) as the same steady current is drawn from each.



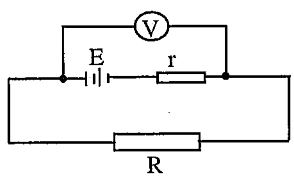
We can see the ZnC cell has a greater capacity, but the NiCd cell gives a steadier output voltage.

Example 6:
(a) How can we tell that the ZnC cell has a greater capacity?
(b) Can you compare the emf's of the cells?
(c) Describe how the internal resistance of each cell changes as current is drawn from it.

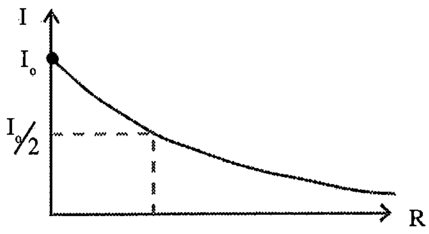
Answers
(a) The ZnC cell provides the same current for a longer time.
(b) Not without knowing the internal resistances. Only the output voltages are given.
(c) For the NiCd cell, the internal resistance remains constant, then rises sharply as the cell becomes totally discharged. For the ZnC cell, the internal resistance gradually increases as the cell is used up.

Exam Hint: *When comparing cell types, besides considering just the environmental and cost benefits of rechargeable cells, keep in mind the other aspects illustrated by the graphs, such as poor capacity of rechargeables, and lack of warning that they are running out.*

Internal Resistance in simple circuits

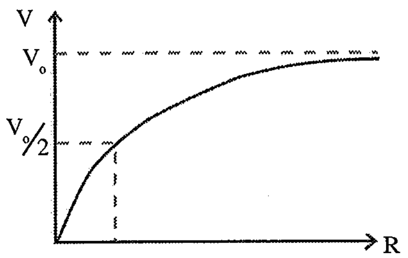


A number of different graphs can be drawn for this setup.
If we increase the load resistance, R:



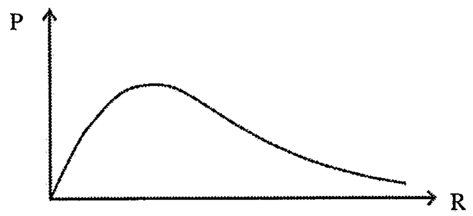
How could we find internal resistance, r, from this graph?
In 2 different ways:
(a) For R=0, $I = I_o = E/r$.
So we find r from $r = E/I_o$.
(b) For $I = I_o/2$, the total resistance is 2r, so r = R at this current flow.

Similarly, if we look at a graph of V against R:



We can determine the emf and internal resistance.
(a) As R approaches infinity (open circuit), no current flows, so no p.d. is dropped across r. The emf $E = V_o$.
(b) When $V = V_o/2$, equal p.d. is dropped across r and R. As the current through both is the same, then r = R when $V = V_o/2$.

A graph of power dissipated in the load, R, is interesting.



When R=0, the power dissipated in R will be zero ($P = I^2R$).
As R approaches infinity, the current drops to zero, so the power dissipated in R will be zero ($P = I^2R$).
It turns out that the maximum power is dissipated in the load, R, when R = r.

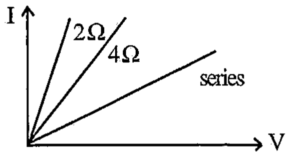
Power transfer is most efficient when the load is matched to the internal resistance. The maximum efficiency must be only 50%.

Prcatice Questions

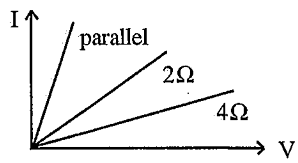
1. A 2 ohm resistor and a 4 ohm resistor are connected in series. Sketch a V-I graph (with V on the X-axis) for each resistor and for the combination, as the p.d. across the circuit is increased. Use the same set of axes.
2. The same two resistors are now connected in parallel. Repeat the graphs.
3. A resistor and a negative temperature coefficient thermistor have the same resistance at room temperature. Sketch a V-I graph for the combination when they are connected (a) in series, and (b) in parallel. Explain your graphs.
4. A simple circuit has a supply of emf E and internal resistance r, connected to a load R.
   (a) Find an expression for the power dissipated in load R when:
   (i) R=0 (ii) R=0.5r (iii) R=r (iv) R=1.5r (v) R=5r.
   (b) Sketch a graph of power dissipated, P, against load resistance, R.
5. For the same setup, sketch a graph of the power, P, dissipated in the internal resistance against load resistance, R, as R is increased. Briefly explain your graph.

Answers

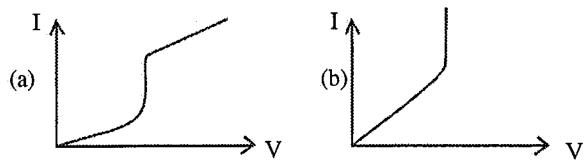
1.



2.



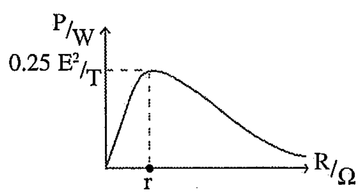
3.



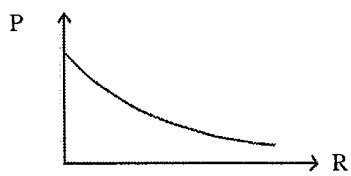
In series, there is a sudden increase in current due to the thermistor (as its resistance plummets), then the usual linear relationship due to the resistor in series.
In parallel, the graph is steeper at the beginning, as the resistors in parallel have a smaller combined resistance than those in series. Then the sudden drop in resistance of the thermistor effectively short circuits the parallel resistor.

4. (a) (i) $P = I^2R = 0$.
(ii) $I = E/(R+r) = E/1.5r$
Then P dissipated in load is $I^2R = 0.22E^2/r$ (watts)
(iii) $I = E/2r$, $P = 0.25E^2/r$ (watts)
(iv) $I = E/2.5r$, $P = 0.24E^2/r$ (watts)
(v) $I = E/6r$, $P = 0.14E^2/r$ (watts)

(b)



5.



As R increases from zero, the current decreases from its maximum value (E/r). As r is constant, $P = I^2 r$ will also decrease.

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU
Physics Factsheets may be copied free of charge by teaching staff or students, provided that their school is a registered subscriber. ISSN 1351-5136

Physics Factsheet

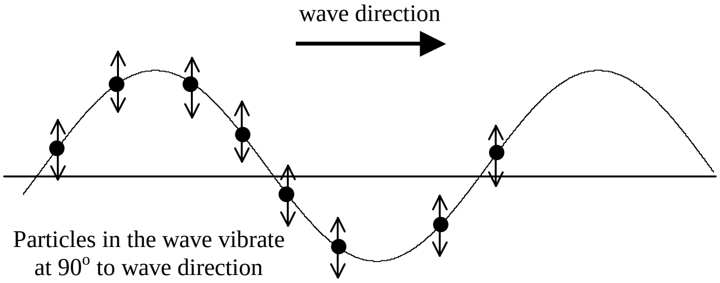
April 2001 Number 17

Waves Basics

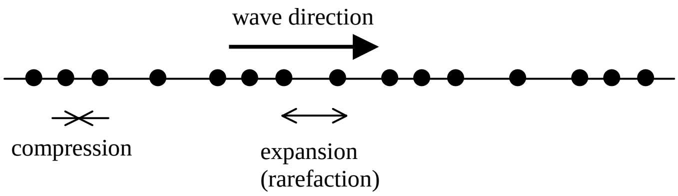
This Factsheet will introduce wave definitions and basic properties.

Types of wave
Waves may be **mechanical** (i.e. they require a medium such as air or water to propagate) or **electromagnetic** (which propagate in a vacuum)
Waves can be classified as:

a) **Transverse** – here the disturbance is at right angles to the direction of the wave. Examples include water waves and light (electromagnetic).



b) **Longitudinal** – here the disturbance is in the same direction, parallel to the direction of the waves. Examples include sound and seismic P waves.



Particles of the wave vibrate in the same direction as the wave

Both longitudinal and transverse waves can be represented graphically as shown below.

> ***Exam Hint:*** *You will be expected to know examples of longitudinal and transverse waves and to describe the differences between them.*

Wave pulses and continuous waves

- A wave pulse involves a short or single disturbance of the medium it is travelling in. For example, dropping an object in water may produce a wave pulse.
- Continuous waves involve repeated disturbances of the medium. For example, to produce continuous waves in a ripple tank, the dipper would have to be dipped into the tank at regular intervals.

The rest of the Factsheet will focus on continuous waves.

The wave formula

$$v = \lambda \times f$$

v = velocity (ms^{-1}) λ = wavelength (m) f = frequency (Hz)

To see where this formula comes from, consider how far the wave moves in one second.

- We know (from the definition of frequency) that there are f waves each second.
- Each wave is of length λ.
- So the total distance moved in one second is $f \times \lambda$
- But velocity = distance÷ time, so $v = f\lambda$

Typical Exam Question

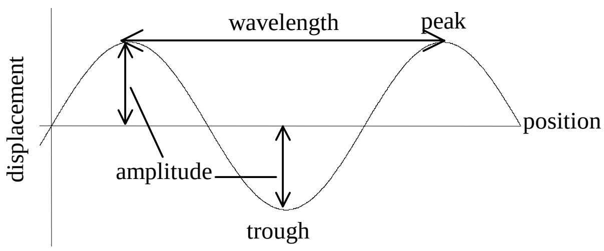
a) What is the velocity of a wave with wavelength 25cm and frequency 12 Hz? [2]
0.25×12 ✓= = $3ms^{-1}$ ✓

b) A wave has a velocity of 1.25 ms^{-1}. Eight waves are observed to pass a fixed point in 2 seconds. Find
i) the period of the wave [1]
2÷8 = 0.25 ✓

ii) the wavelength of the wave [2]
f = 1/T = 4 ✓ *λ = v/f = 1.25/4 = 0.3125 m.* ✓

Glossary

- Amplitude: the maximum displacement of a wave particle from its undisturbed position.
- Wavelenth: the distance between two similar points on a wave. Units: metres
- Frequency: the number of waves that pass a point in one second. Units: Hertz (Hz)
- Period: the time taken to complete one wave cycle. Units: second (s)
- The period (T) and the frequency (f) are related by $T = \frac{1}{f}$
- Peak(or crest): the point of maximum displacement – the "highest point"
- Trough: the point of minimum displacement – the "lowest point"



Wave properties

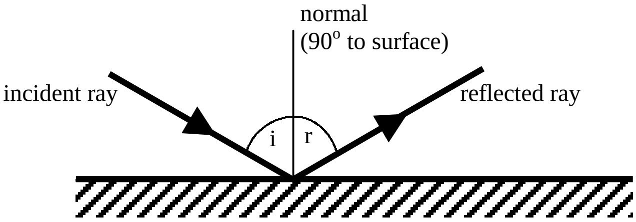
All waves will undergo the following processes according to the same laws:

- reflection
- refraction
- interference
- diffraction

Reflection and its laws

There are two laws of reflection. They apply to both plane (flat) and curved mirrors (or other reflecting surface for waves other than light).

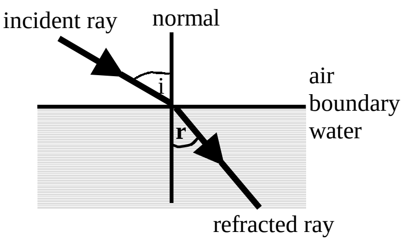
- the incident ray and reflected ray lie in a single plane which is perpendicular to the surface at the point of incidence
- angle of incidence (i) is equal to angle of reflection (r)



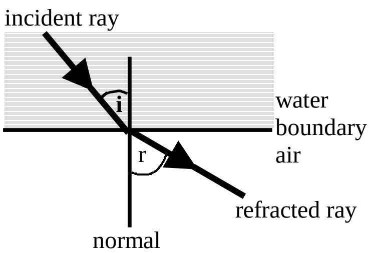
NB: The angles are **always** measured to the **normal**, **NOT** the **reflecting surface**

Refraction

Refraction is the change in direction of a wave as it crosses the boundary between two materials (eg air and water).



If the wave crossed the boundary in the opposite direction (i.e. water to air) then the wave direction would change in the opposite way:

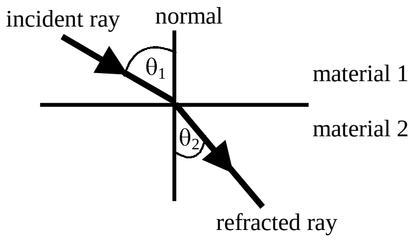


Refraction is governed by two laws:

- the incident ray and reflected ray lie in a single plane which is perpendicular to the surface at the point of incidence
- "**Snell's Law**": at the boundary between any two given materials, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is constant for rays of any particular wavelength.

Refractive index

Consider a wave passing from material 1 to material 2. In material 1 its angle to the normal is θ_1 and in material 2 it is θ_2.



From Snell's Law, we know that

$$\frac{\sin\theta_1}{\sin\theta_2} = \text{a constant}$$

This constant is the **refractive index of material 2 with respect to material 1**, written $_1n_2$.

So Snell's Law can be written as:

$$\frac{\sin\theta_1}{\sin\theta_2} = {}_1n_2$$

Now suppose the ray is instead travelling from material 2 to material 1. Using the above equation, we would obtain:

$$\frac{\sin\theta_2}{\sin\theta_1} = {}_2n_1$$

This gives:

$$_2n_1 = \frac{1}{_1n_2}$$

Wave speed, wavelength and refractive index

Refractive index is also related the the wave speed and wavelength in the two materials. This is used to **define the refractive index**.

$$_1n_2 = \frac{v_1}{v_2}$$

where v_1 = velocity of wave in material 1
v_2 = velocity of wave in material 2

This defines the refractive index.

Also, since **frequency does not change** during refraction:

$$_1n_2 = \frac{\lambda_1}{\lambda_2}$$

where λ_1 = wavelength of wave in material 1
λ_2 = wavelength of wave in material 2

The refractive index is often given relative to **air**. So if, for example, a question tells you that the refractive index of glass is 1.50, it means that $_{air}n_{glass} = 1.50$.

***Exam Hint**: Defining the refractive index is commonly asked. Make sure you use the equation involving wave speeds, and define all the terms.*

Typical Exam Question

The refractive index of water is *1.33*.
The speed of light in air is $3\times 10^{8}\ ms^{-1}$.

a) Calculate the speed of light in water. [2]

$$1.33 = \frac{v_{air}}{v_{water}} = \frac{3\times 10^8}{v_{water}} \checkmark$$

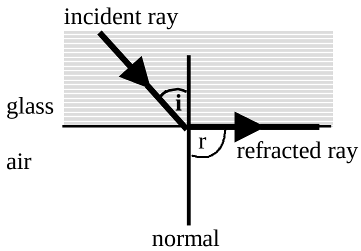
$$v_{water} = \frac{3\times 10^8}{1.33} = 2.26\times 10^8\ ms^{-1} \checkmark$$

b) State the effect of the refraction on the frequency [1]
none – it is unchanged ✓

***Tip**: The speed of light in any other medium should always be **lower** than in vacuum (or air). Use this to check your answer – if you have got too high a speed, you have probably got the v_{air} and $v_{material}$ the wrong way up in the equation*

Critical angle and total internal reflection

When light travels from a material with a higher refractive index to one with a lower refractive index (eg from glass to air), it is possible for the angle of refraction to be 90°.



The angle at which this occurs is called the **critical angle, c**.

Using the equation $\frac{\sin\theta_1}{\sin\theta_2} = {}_1n_2$, we find: $\frac{\sin c}{\sin 90} = {}_{glass}n_{air}$

But since we are usually given ${}_{air}n_{glass}$, not ${}_{glass}n_{air}$, it is more useful to write this as:

$$\frac{\sin c}{\sin 90} = \frac{1}{{}_{air}n_{glass}}$$

Since sin90° = 1, we have:

$$\sin c = \frac{1}{{}_{air}n_{glass}}$$

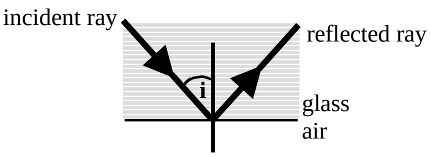
Given that the refractive index of glass is 1.50, we can calculate:

$$\sin c = \frac{1}{1.50} = 0.667 \Rightarrow c = 41.8^o \text{ (3 SF)}$$

Note: this cannot occur for light travelling from air to glass, since the angle to the normal decreases when travelling in this direction.

What happens for angles larger than the critical angle?

If the angle of incidence is greater than the critical angle, then the ray cannot be refracted – instead it is **totally internally reflected**



In fact, a certain amount of reflection will always occur at the interface, but for angles greater than the critical angle, **only reflection** can occur. The incident and reflected ray obey the laws of reflection.

Interference

When two sets of waves combine, we use:

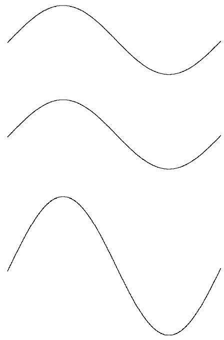
Principle of Superposition
The resultant displacement at a point is equal to the vector sum of the individual displacements at that point.

To see how this works, we will look at some examples:

1. Constructive Interference

Wave 1
+
Wave 2

Produces

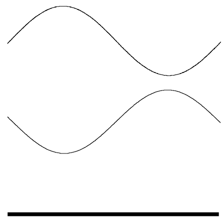


In this case, the peaks and troughs in the two waves coincide, and hence reinforce each other.

2. Destructive Interference

Wave 1
+
Wave 2

Produces

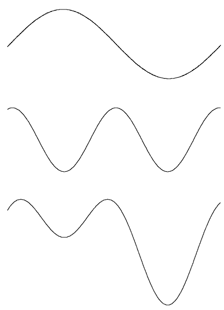


In this case, the peaks in one wave coincide with the troughs in the other, to produce no resultant displacement – the two cancel.

3.

Wave 1
+
Wave 2

Produces



In this case, peaks and troughs do not exactly coincide.

Phase difference

Phase difference is a way of measuring how far ahead one wave is of another (eg half a wavelength, quarter of a wavelength etc). It is given as an angle, with a whole wavelength corresponding to 360°. So if one wave leads another by a quarter of a wavelength, this is a phase difference of ¼ × 360° = 90°.

If the phase difference is zero (example 1 above), they are **in phase.**
If the phase difference is 180° (= half a wavelength), they are **completely out of phase** (example 2).

- Sources of waves are **coherent** if they maintain a constant phase difference and have the same frequency (eg lasers).
- A **wavefront** is a line or surface in the path of a wave motion on which all the disturbances are in phase. It is perpendicular to the direction travel of the wave.

Typical Exam Question

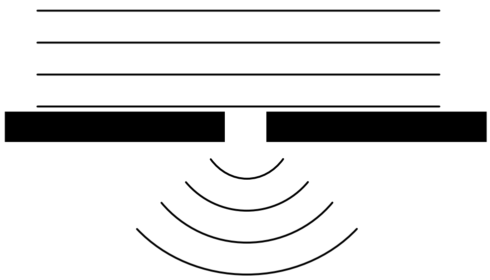
a) Explain what is meant by superposition of waves. [2]
Waves coinciding at a point in space ✓
Disturbances add together, i.e. 'superpose' ✓

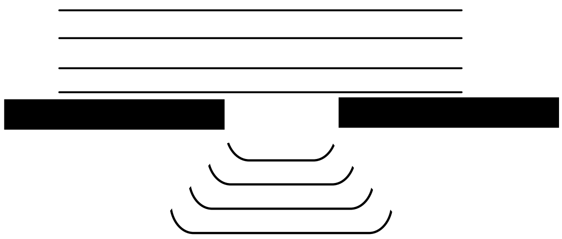
b) Distinguish between constructive and destructive interference. [4]
Constructive: waves in phase as they superpose ✓
Disturbances add to give a larger amplitude ✓
Destructive: waves 180^o out of phase ✓
Disturbances cancel to give zero amplitude ✓

c) State the conditions necessary for sources of waves to be coherent.[2]
Same frequency ✓ Constant phase relationship ✓

Diffraction

When waves pass an edge of an obstacle, or through a gap, they spread out and change shape. The wavelength, frequency and velocity remains constant. The extent of the spreading depends on the size of the gap, as shown below:





Exam Hint*: When drawing diagrams of diffraction, make sure you keep the spacing between the wavefronts the same – this shows the frequency of the waves is unchanged.*

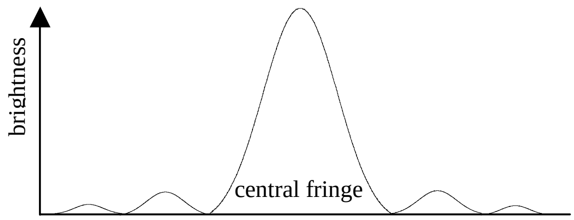
Appreciable diffraction only occurs if the gap is no bigger than the wavelength of the wave.

- The narrower the slit, the greater the diffraction for a particular wavelength
- The longer the wavelength for a constant slit width, the greater the diffraction

Diffraction of light and the single slit

Diffraction of light through a slit onto a screen leads to the production of light and dark fringes.

The brightness and width of the fringes can be represented graphically:

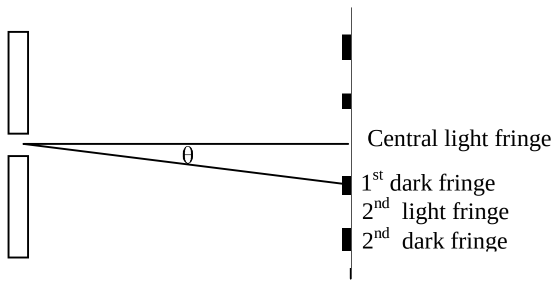


The fringe pattern has the following properties:
- It is symmetrical
- The central bright fringe is much brighter than the other fringes
- It is twice as wide as the other fringes
- The brightness (intensity of light) decreases with distance from the central fringe – so the outer fringes are the faintest.

The position of the dark fringes can be calculated using:

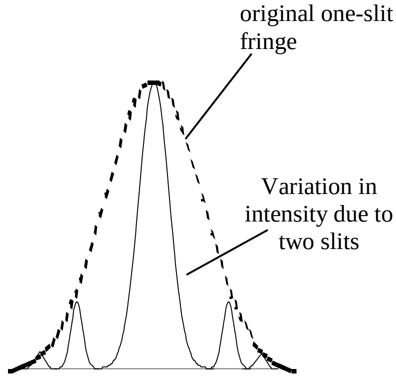
$$\sin\theta = \frac{a\lambda}{w}$$

where θ = angle subtended in the centre (see diagram)
λ = wavelength
w = slit width
a = 1, 2, 3…. (fringe number)



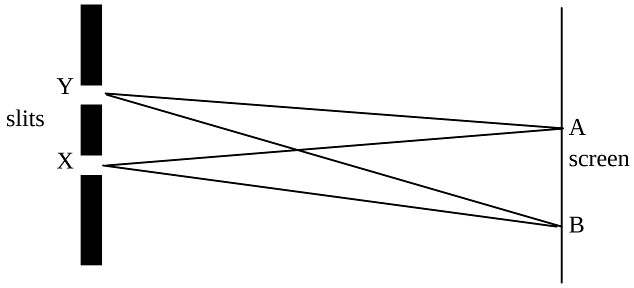
Double slit diffraction

This produces a difference in intensity within each bright fringe seen in the single slit pattern. This results from interference between light from one slit the other



This interference results from the fact that light from the different slits travels a different distance to reach a given point on the screen; this is referred to as the **path difference**.

At the centre of the screen (point A), waves from slits X and Y have travelled the same distance and therefore are in phase, and hence interfere constructively leading to a bright fringe.



As the distance from the centre changes, so does the path difference between the waves from each slit. At point B, for example, the waves from slit Y have travelled substantially further than those from slit X.

When the path difference becomes half the wavelength, then destructive interference occurs, producing a dark fringe.

Further increases in distance from the screen increases the path difference until it becomes a whole wavelength – this makes the waves in phase again, so constructive interference occurs. Further increases produce a path difference of 1½ wavelengths – giving destructive interference again.

So the bright fringes are produced from path difference $m\lambda$, and the dark fringes from path difference $(m + ½)\lambda$, where λ is the wavelength and m is any whole number.

The fringe spacing between two adjacent bright fringes is given by

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

D = distance to screen; λ = wavelength; d = slit spacing

In between the light and dark fringes, the interference is not perfectly constructive or destructive, so the intensity of the light changes gradually.

Polarisation

Normally, the oscillations in a transverse wave may be in many different directions. For example, for a wave travelling out of this page towards you, the oscillations will be in the plane of the page, and could be left to right, up and down, diagonally etc.

Transverse waves may undergo **polarisation**. A polarised wave oscillates in one direction only. Longitudinal waves cannot be polarised because the oscillations are already in one direction only.

> ***Exam Hint:*** *This is a key difference between longitudinal and transverse waves, and is often asked.*

Polarised light is most easily produced using a piece of Polaroid (as used in sunglasses). Polaroid works by only allowing through light which oscillates in a particular direction

If the light is passed through a vertical piece of Polaroid, then the emerging ray will be polarised vertically. It will have half the intensity of the original beam – this is why sunglasses work.

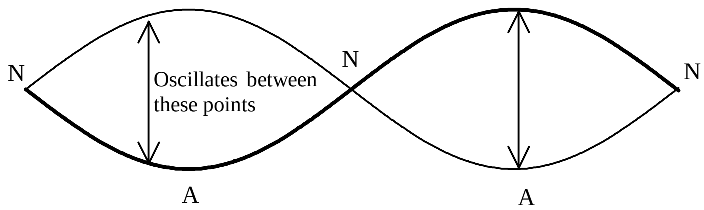
If this ray then meets a horizontal piece of Polaroid, no light will pass through.

Progressive and stationary waves

The waves discussed so far have been **progressive** i.e. they move in a particular direction, transferring energy along the direction the wave is travelling. In a **stationary** (or standing) wave, the wave does not move in a particular direction and energy is **stored** by the wave.

Stationary waves are the result of two progressive waves of the same frequency travelling in opposite directions along the same line. The diagram below shows an example of a stationary wave.

Each point on the wire can oscillate between the two positions shown.



- The **nodes** (marked N) never move. These occur where the two original waves interfere destructively
- The **antinodes** (marked A) are points of maximum displacement. They occur where the two original waves interfere constructively.

Table 1 below compares progressive and stationary waves.

Table 1. Progressive and stationary waves

Stationary wave	**Progressive wave**
Stores vibrational energy	Transmits vibrational energy
Amplitude varies	Amplitude is constant
All points between any two adjacent nodes are in phase	Phase varies smoothly with distance along the path of the wave
Nodes are half a wavelength apart; antinodes are midway between nodes	No nodes or antinodes

Typical Exam Question

a) Explain the terms node and antinode [2]
Node: point of no vibration ✓
Antinode: point of maximum vibration ✓

b) Two identical progressive waves are travelling along the same straight line in opposite directions.
(i) Explain how a stationary wave pattern is formed [3]
Stationary wave is formed by the superposition ✓
of the two waves
Nodes are created by destructive interference ✓
and antinodes are created by constructive interference. ✓
(ii) Compare the amplitude and phase of particles along a stationary wave with those of a progressive wave. [2]
All points on a stationary wave are in phase, points on a progressive wave are out of phase with each other ✓
All points on a progressive wave have the same amplitude, different points on stationary wave have different amplitudes ✓

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's answer is given below.

a) (i) Explain the term 'wave front' [1]
it's at right-angles to the wave direction 0/1

> This statement is true, but does not explain wavefront. When the student read the next part of the question, s/he should have realised that this was not an adequate answer.

(ii) State the relationship between the orientation of a wave front and the direction in which the wave is travelling [1]
at right angles ✓ *1/1*

b) A longitudinal wave of frequency 30 kHz has a speed of 340ms^{-1} when travelling in air. Its wavelength when travelling in water is 0.05m.

(i) Calculate the minimum distance between two points on the wave that differ in phase by 60^o when it is travelling through air.[3]
60^o = one sixth of wavelength ✓
0.05 ÷ 6 = 0.008m *1/3*

> *The first part of the method is correct, but the student has used the wavelength for water waves. Read the question! This shows the advantage of showing working – without it, no marks would have been awarded.*

(ii) Calculate the speed of the wave in water [2]
1.5ms^{-1} *0/2*

> The student has probably failed to convert kHz to Hz. If s/he had shown working, one mark might have been awarded, since s/he could have demonstrated the knowledge that frequency is unchanged. The answer should have worried him/her!

(iii) A pulse of the wave lasts for 10ms. Calculate the number of complete waves that it contains. [2]
30 000 × 0.01 ✓*= 300* ✓ *2/2*

> The student has evidently now realised that the frequency is 30 kHz, not 30 Hz, but has neglected to change the earlier answer! Always make time to check.

Examiner's Answers

a) (i) A surface in which all oscillations are in phase ✓
(ii) They are perpendicular ✓

b) (i) λ = v/f
=340/30000
=0.0113m ✓
Phase difference of 60^o corresponds to λ/6 ✓
= 1.89mm ✓
(ii) speed in water = f × wavelength in water ✓
= 30 000 × 0.05 =1500ms^{-1} ✓
(iii) Number of waves = time of pulse × frequency ✓
= 0.01 × 30000 =300 ✓

Questions

1. a) Explain the difference between transverse and longitudinal waves.
   b) Give two examples of transverse waves and two examples of longitudinal waves.

2. Explain the difference between a wave pulse and a continuous wave.

3. Explain what is meant by the following terms:
   Amplitude
   Wavelength
   Peak
   Trough
   Frequency

4. Define the refractive index for a wave travelling from material 1 to material 2.

5. Explain what is meant by the Principle of Superposition.

6. Explain what is meant by diffraction.

7. Explain why sound waves cannot be polarised.

8. Give two differences between a stationary wave and a progressive wave.

9. A wave has speed 50ms^{-1} and wavelength 2m. Calculate its period.

10. The refractive index of glass is 1.50.
    a) A ray of light passes from air to glass. It makes an angle of 20^o to the normal just before entering the glass.
    Calculate the angle the refracted ray makes with the normal.
    b) The speed of light in air is 3 × 10^{8}ms^{-1}.
    Calculate the speed of light in glass.
    c) Calculate the critical angle for glass.

Answers

1. See page 1
2. See page 1
3. See page 1
4. See page 2
5. See page 2
6. See page 4
7. See page 5
8. See page 5
9. f = 50/2 = 25Hz
   T = 1/f = 0.04 s

10. a) sin20/sinr = 1.5
    sinr = sin20/1.5 = 0.228
    r = 13^o

    b) c_{air} / c_{glass} = 1.50
    c_{glass} = 3 × 10^8/ 1.50 = 2 × 10^8 ms^{-1}

    c) sinc = 1/1.5 = 0.667
    c = 42^o (2 SF)

Acknowledgements: *This Factsheet was researched and written by Nirinder Hunjan*
Curriculum Press, Unit 305B The Big Peg, 120 Vyse Street, Birmingham B18 6NF*.*
 ISSN 1351-5136

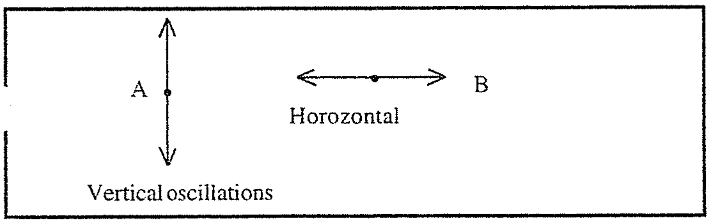
Physics Factsheet

www.curriculum-press.co.uk Number 93

Polaroids and Polarisation

Polarisation is a wave property. But not all waves can exhibit this property.

In a transverse wave, the oscillations are at right angles to the direction of travel. The plane of this oscillation defines the polarisation of the wave. Here are two waves travelling towards us. We say that wave **A** is vertically plane-polarised, and wave **B** is horizontally plane-polarised.



With longitudinal (compression) waves, such as sound, polarisation is not possible, as every wave oscillates back-and-forth along the line of travel.

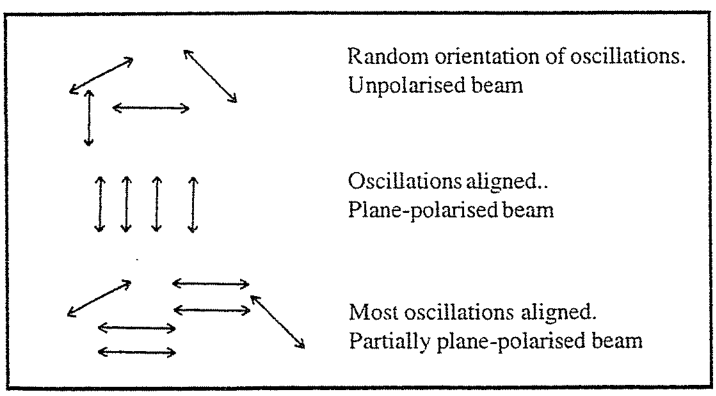
> ***Example: Light is an e.m. wave and can be polarised. Sound is a mechanical wave and cannot be polarised. Does this mean that all e.m. waves can be polarised and all mechanical waves cannot be polarised?***
>
> ***Answer:*** *It is true that all e.m. waves can be polarised. This is because they are all transverse waves. But transverse mechanical waves can also be polarised. Water waves on the sea always oscillate in a vertical plane.*

> ***Exam Hint:*** *You often see questions concerning wave properties. Remember that reflection, refraction, and diffraction are properties of* ***all*** *waves. But polarisation only occurs with* ***transverse*** *waves.*

Types of Polarisation

If every transverse wave oscillates in a fixed plane (in general), how can a beam of light ever by unpolarised?

A beam consists of a great number of waves. The polarisation of the beam depends on the alignment of all of these waves.

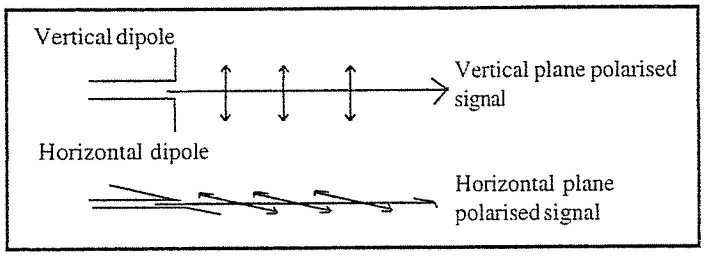


> ***Example: Can the oscillations on the surface of water ever by horizontally polarised?***
>
> ***Answer:*** *Unlikely. Horizontal disturbances tend to set up turbulent (rotating) motion. The Great Whirlpool below Niagara Falls is an example of this.*

How Polarisation is set up

Polarisation can be set up intentionally. It can also happen just through the properties of waves and materials.

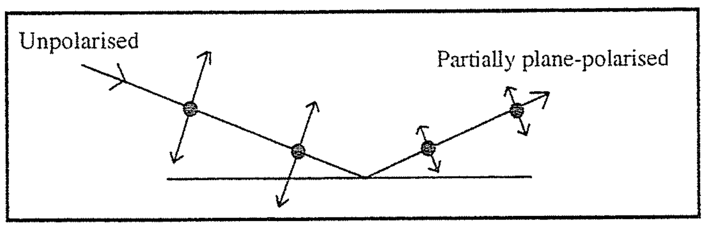
With microwaves, the plane of polarisation of the beam depends purely on the **alignment of the dipole aerial** transmitting the signal.



We will later see how this can be made use of.

With light, polarisation can be achieved by reflection, scattering, and filtering (polaroid sheets).

A beam of unpolarised light **reflects** as a partially plane-polarised beam from a shiny surface:

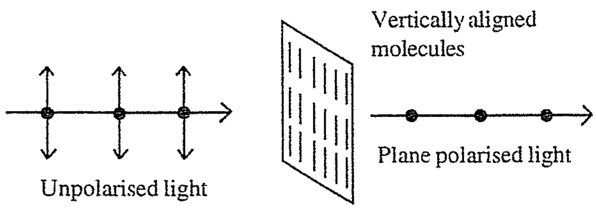


The oscillations parallel to the surface are least affected by reflection.

> ***Example: What has happened to the oscillations perpendicular to the surface?***
>
> ***Answer:*** *They have been absorbed by the material. The wave energy is transformed to thermal energy.*

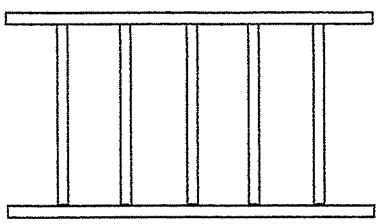
When a light beam is **scattered** as it travels through a medium, the scattered beam is partially plane-polarised. Any light scattered through 90° is completely plane-polarised. Looking at the blue sky through a polarising filter shows different intensities at different angles.

When a light beam travels through a **polarising filter** (a polaroid sheet), it emerges plane-polarised. The filter has long molecules in alignment with each other. These absorb the oscillations in this plane.



It is the horizontal oscillations that get through the filter in this diagram.

Example: Which microwave oscillations are stopped by the metal grid shown?



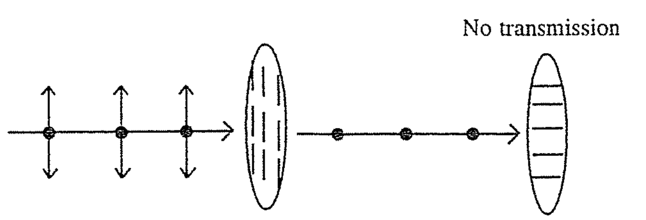
Answer: *The* ***vertical*** *oscillations are absorbed by the* ***vertical*** *metal rods.*

Be clear how this effect works for microwaves and for light. Many students think that the vertical oscillations would fit between the rods, and that it is the horizontal oscillations which would be blocked. But this is not the case.

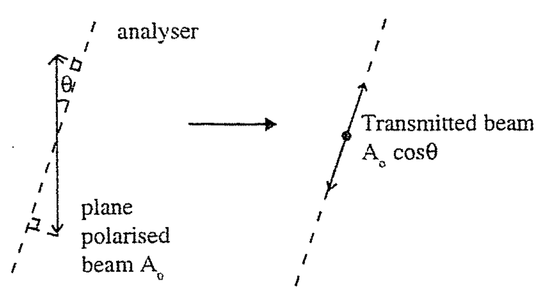
Polarisation Calculations

Mathematical calculations are generally limited to working out transmitted components and energy loss (or transmission) in partially plane-polarised waves.

Two **crossed** polaroids block out all of the light (or let us assume they do). The first polaroid produces plane-polarised light. This oscillation is blocked by the second polaroid (sometimes called the **analyser**).



But if the second polaroid was rotated to an angle of θ° from the first (rather than through 90°), some of the plane-polarised light would get through:



The intensity of the transmitted beam is reduced from $\mathbf{A_o}$ to $\mathbf{A_o \cos\theta}$

Notice that not only has the amplitude of the transmitted beam been reduced, but the plane of polarisation has been rotated through angle θ.

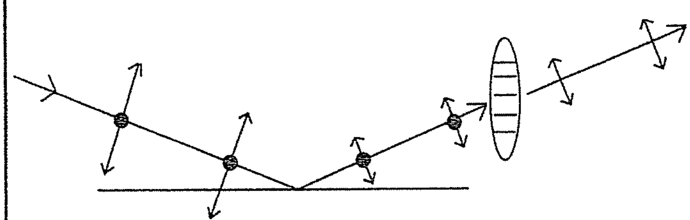
In general, the **intensity** of a wave is proportional to the **square of the amplitude**

$$\mathbf{I \propto A^2}$$

This means that in our example, the intensity of the transmitted beam, $\mathbf{I = I_o \cos^2}\theta$. A calculation in the problems at the end of the Factsheet will illustrate this point.

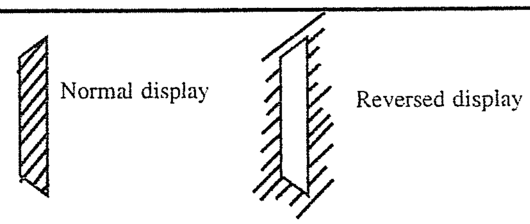
Uses of Polarisation

1. A problem we have all noticed is **glare** reflected from shiny surfaces like water. This glare is partially plane-polarised in the horizontal plane. Using polaroid filters (sunglasses) aligned to block the horizontally polarised light, a great deal of the glare can be eliminated without other light sources being affected to anywhere near the same extent.



2. Liquid crystal displays make use of polarisation. A voltage across a region of the cell causes this region to act like a polarising filter. Crossed polaroids above and below ensure no light can go through, then reflect back out. The region appears dark.

   If you rotate one of these polaroids through 90°, the display on the screen reverses.



3. Most television signals use horizontally polarised microwaves. Weak signal areas (e.g. valleys) may need a booster transmitter. This aerial emits vertically polarised microwaves. There should be no interference between the two signals.

4. Polarisation is used for stress analysis in structures. No light should emerge from crossed polaroids. But some materials (e.g. perspex) under stress can twist the plane of polarisation of the light, causing some of the light to get through the second polaroid.

   This effect depends on the amount of stress and on the wavelength of the light being used. A perspex model of a structure placed between the polaroids can produce a coloured pattern indicating weaknesses in the structure.

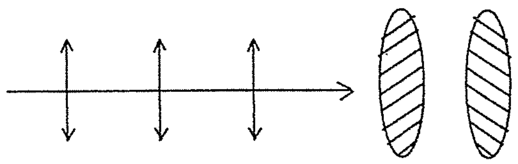
Questions

1. (a) How would you show that the radiation from a microwave transmitter was vertically polarised?
   (b) Why would you expect light from a filament lamp to be unpolarised?

2. If the polarisation vectors in a light beam are all random, why don't they cancel each other out (destructive interference)?

3. A vertically polarised beam of light of amplitude **A** is incident on a polaroid filter with its polarising direction at 20° to the vertical.
   (a) Find the magnitude and plane of polarisation of the transmitted beam.
   (b) Find the percentage reduction in intensity in the transmitted beam.

4. A vertically polarised beam of light of amplitude **A** passes through a filter with its polarising direction 23° clockwise from vertical, then through a filter with its polarising direction 15° anticlockwise from vertical.


polarising directions

   (a) Find the amplitude and intensity of the transmitted beam.
   (b) What is its plane of polarisation?

5. When the Sun is low in the sky, its light can be reflected from vertical glass windows into your eyes. Discuss how polaroid sunglasses would cope with this glare.

Answers

1. (a) Insert a vertical metal grid and see if it reduces the signal at the receiver to zero.
   (b) The individual atoms emitting the waves are all acting independently. No particular plane of polarisation is favoured.

2. At any instant the random spread of vectors would cause a resultant plane of polarisation. But this would be continuously (and very quickly) changing, resulting in an unpolarised beam. From energy considerations the rays could not cancel each other, or energy that had been emitted would disappear.

3. (a) $\mathbf{A = A_0\cos20° = 0.94A_0}$
   Plane of polarisation of transmitted beam is 20° from the vertical.

   (b) Transmitted intensity $\mathbf{I = I_0\cos^2\theta = 0.88I_0}$
   So the percentage reduction in intensity is **12%**.

4. (a) First filter rotates plane of polarisation 23° clockwise.
   $\mathbf{A_1 = A_0\cos23° = 0.92A_0}$

   Second filter rotates plane of polarisation (23 + 15 =) 38° anticlockwise.
   $\mathbf{A_2 = A_1\cos38° = 0.79A_1 = 0.73A_0}$

   So $A_2 = \mathbf{0.73A_0}$ and the plane of polarisation is **15° anticlockwise** (from vertical).
   $\mathbf{I \propto A^2,\ \ I_2 = 0.73^2 I_0 = 0.53I_0}$

5. In this set-up, it is the vertically polarised oscillations (parallel to the plane of the window) that would reflect into your eyes. The polaroid glasses are aligned to block horizontally polarised rays. The sunglasses would not help reduce the glare.

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press,Bank House, 105 King Street,Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

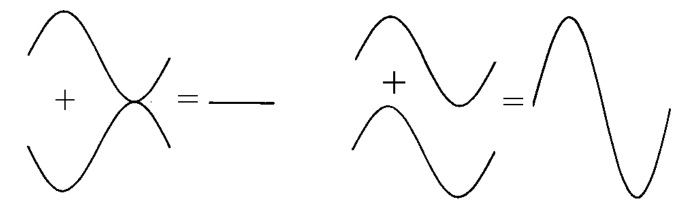
January 2003 Number 47

Interference

This Factsheet will expain:
- what is meant by interference
- interference patterns
- double slit interference

Before working through this Factsheet you should make sure you understand the basic ideas in waves (Factsheet 17)

What is interference?
When two waves of a similar type meet they **interfere** with each other. The intensity at that point is the sum of the intensities of the two waves at that point. This is called the **principle of superposition** of waves. It works for all types of waves and for two waves or more.

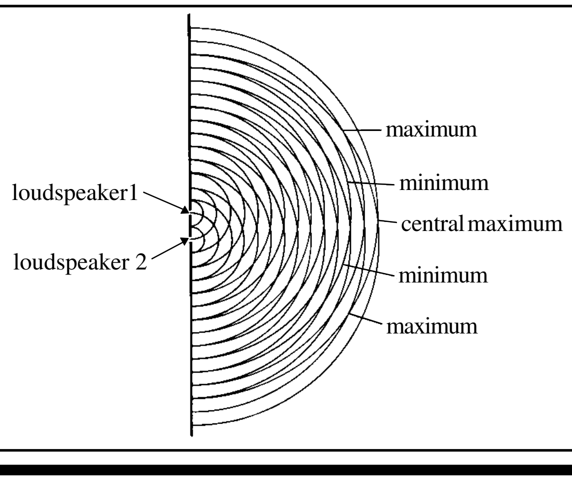


Two waves out of phase interfere to cancel one another out

Two waves in phase interfere to produce a bigger wave

Adding waves of different wavelengths and amplitudes results in complex waveforms. However, there are some interesting effects when two waves with the *same* wavelength overlap.

Constructive and destructive interference
Imagine, for example, a situation where the same pure musical note is fed to two loudspeakers a short distance apart and facing in the same direction. You might expect that the sound you would hear in front of the speakers would be louder than if there were only one speaker. To a certain extent this is true – but it is only part of the story. If you move around in the area in front of the speakers then you would notice that in some places the sound is louder, in others it is quieter. These are **interference effects**. The diagram shows what is happening. The positions of the loudspeakers are labelled 1 and 2 and the diagram shows where maxima (positions where the sound is loudest) and minima (positions where the sound is quietest) occur.



At every position in front of the speakers, waves are arriving from both sources. In some places the waves will be in phase with one another; that is, the peaks of both waves arrive at exactly the same point at the same time. In these places, the principle of superposition predicts that the resultant wave has double the amplitude – and you would hear a louder sound. This is called **constructive interference** – each wave interferes with (or is superposed upon) the other to produce a bigger wave. However, there will be other points nearby where the peak of the wave arriving from one speaker coincides exactly with the trough of the wave arriving from the other speaker.

If the two waves are of equal amplitude at this point, then they interfere with one another to produce a wave of zero amplitude – cancelling one another out. At this point you would expect to hear absolute silence! (In practice, echoes from the surrounding surfaces mean it is unlikely it would actually be silent – but the sound heard at this point would certainly be much quieter.) This is known as **destructive interference** – and is the more surprising effect of the principle of superposition. Remember that it is not quiet at this position because *no* wave is arriving – but because TWO waves are arriving, out of phase, and each one is cancelling out the other. If one of the speakers were to be turned off, the sound at this point would get louder.

Constructive interference
- *occurs when two waves are in phase - their peaks coincide*
- *amplitude of resulting wave is sum of original amplitudes*

Destructive interference
- *occurs when two waves are totally out of phase/phase difference of $180^o = \pi$ radians)*
- *amplitude of resulting wave is the difference of original amplitudes - which will be zero if the original waves had the same amplitude*

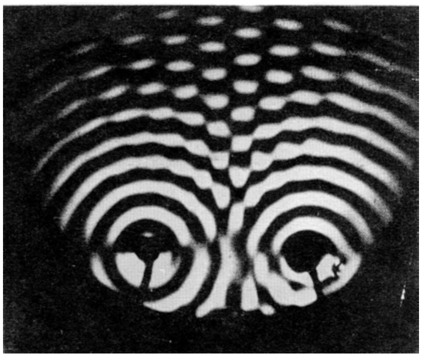
Interference occurs for all types of wave – so it is possible to observe exactly the same effect with two light sources – as long as they are both **coherent** and of similar amplitudes. (Coherent waves have the same frequency and a constant phase relationship.) In this case, the effect of constructive interference would be brighter light and the effect of destructive interference would be no light – or darkness. It is even harder to believe that an area is in darkness because two lights are shining on it simultaneously – but just because something is surprising doesn't make it untrue! In this case, switching or blocking one of the light sources off would make it light again at that point!

Interference patterns of this kind in microwaves are the reason that the food in a microwave oven is placed on a rotating turntable. If it were kept still, there would be hot spots and cold spots in the food – in most foodstuffs, the conduction is too slow to spread the heat around evenly in the cooking time.

In radio waves, interference patterns are set up when the same signal is simultaneously transmitted from two aerials at a distance apart. If you had a moveable receiver (such as a car radio) and were tuned into this frequency as you travelled across the region where both of these signals overlapped, then you would experience the received signal getting weaker and stronger – and the radio getting quieter and louder - as you moved through the interference pattern.

Demonstrating intereference patterns

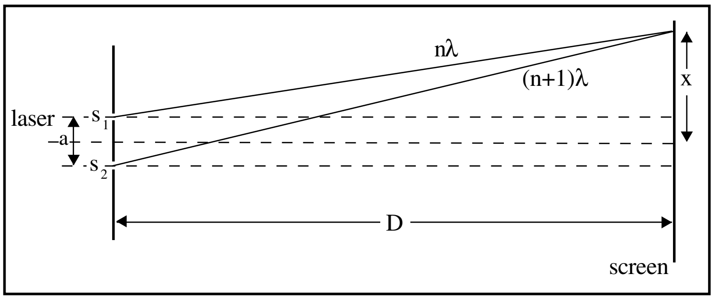
The effect can be clearly demonstrated in a ripple tank. Two coherent wave sources are produced by attaching two dippers to a single beam suspended by rubber bands above the water in the ripple tank. The beam is vibrated by a motor with an eccentric mass (i.e. a mass which is not mounted at its centre) so that it wobbles as the motor spins. The height is adjusted so that both dippers touch the surface of the water producing coherent waves from points several centimeters apart. The resulting wave pattern is shown in the photograph below. The pattern of constructive and destructive interference, as waves coincide in phase in some places and out of phase in others, produces bands of interference that appear to fan out from the space between the sources.



Double slit interference

To demonstrate the same effect as described above using light waves requires a little more sophistication. In practice, the only way to ensure two coherent wave sources is to use a single light source of just one wavelength – best achieved by using a laser – and to split that using an arrangement known as **Young's slits** (after Thomas Young, who designed this experiment in 1801 to measure the wavelength of light). This is a slide with two narrow, parallel slits very close together. The laser beam falls onto both slits and passes through each, spreading out slightly by diffraction as it does so. The light waves emerging from each slit act as two separate beams of light and (since they come from the same source) these two waves have the same wavelength and leave the slits in phase with one another – they are coherent.

A screen is placed some distance from the laser (usually several metres away – so the room needs to be dark if they are to be seen clearly) and interference bands (or **fringes**) can be seen. They appear as alternate areas of light and darkness, where the two waves interfere with one another constructively and destructively. The diagram below shows the relationship between the various quantities in this arrangement.



- Where the two waves have travelled the same distance from both slits to the screen (i.e. right in the middle of the interference pattern) then they will be in phase and there will be a bright fringe – this central maximum is the brightest of all the fringes.
- Moving to either side of the central fringe, there will be another bright fringe at the point where the wave from one of the slits is once again in phase with the wave from the other slit. This occurs when the wave from the 'further' of the two slits has travelled exactly one wavelength more than the other – we say that the **path difference** is equal to one wavelength.
- There will be further bright fringes at each position on the screen where the path difference is exactly equal to two wavelengths, three wavelengths and so on. The bright fringes will appear to be equally spaced on the screen.
- In between these bright fringes will be dark fringes of destructive interference. At each of these positions, the path difference will be exactly equal to ½ a wavelength, 1½ wavelengths, 2½ wavelengths and so on – making the waves exactly 180° *out* of phase.

The equation: $\lambda = \frac{ax}{D}$ describes this relationship

where: λ = wavelength
a = separation of the slits
(distance between the centres of the two slits)
x = distance between successive fringes
D = distance from the slits to the screen
(all measurements in metres)

Questions

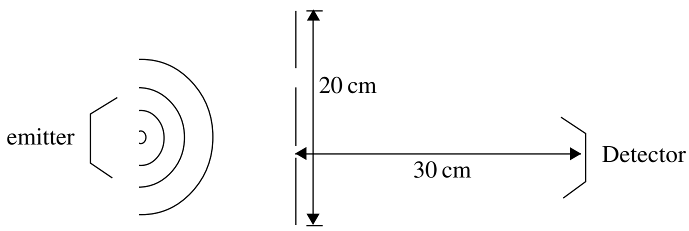
1. (a) Explain what is meant by superposition of waves. [2]
   (b) Distinguish between constructive and destructive interference. [4]
   (c) State <u>two</u> conditions necessary for sources of waves to be coherent. [2]
   (d) Why is it necessary to have two coherent sources to produce an interference pattern. [2]

2 (a) A student sets up two loudspeakers to perform an experiment on the interference of sound. The loudspeakers produce sound waves with the same frequency, which is known to be below 1kHz. She finds a point of destructive interference, i.e. a point where the noise level is very low, at 3m from one speaker and 2m from the other.

Determine all possible frequencies at which the speakers could have been oscillating. Take the speed of sound in air as $340ms^{-1}$. [4]

(b) She now stands midway between a different pair of speakers, which are oscillating at 400 Hz, and 401Hz respectively. The sound waves produced by these speakers have equal amplitudes. Describe and explain what she hears. [4]

3. Microwaves with a wavelength of 6cm are directed towards a metal plate. The plate is 20cm wide and has two parallel slits in it. An interference pattern is formed on the far side of the plate.



(a) A microwave detector is placed 30 cm directly in front of one of the slits. The detector gives a zero reading.
   (i) Comment on the phase difference between the waves from each slit at this point and use this to explain why the detector gave a reading of zero. [2]
   (ii) Calculate the distance between the two slits. [4]

(b) What would happen to the interference pattern if the plate were rotated through an angle of 90°? Explain your answer. [3]

Answers

1. (a) Waves coinciding at a point in space ✓
   Disturbances add together, i.e. 'superpose' ✓ (or diagram)

   (b) Constructive: waves in phase as they superpose ✓
   Disturbances add to give a larger amplitude ✓
   Destructive: waves 180^o out of phase ✓
   Disturbances cancel to give zero amplitude ✓ (or diagram)

   (c) Coherent sources have:
   Same frequency ✓
   Constant phase relationship ✓

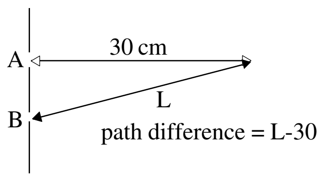
   (d) Without coherent sources the phase relationship between the waves would be continuously changing ✓
   No steady interference pattern. ✓

2. (a) Destructive interference or cancellation implies the path difference is an odd number of half wavelengths, i.e. $\lambda/2, 3\lambda/2, 5\lambda/2, 7\lambda/2$ etc✓
   The path difference = 1m
   Hence: λ = 2m, 2/3m, 2/5m, 2/7m, etc. ✓
   Using $f = v/\lambda$ gives the following possible frequencies:
   170Hz, 510Hz, 850Hz, 1190Hz, etc. ✓
   As we know that the speakers were oscillating below 1kHz, the only possible frequencies for the sound are: 170Hz, 510Hz, 850Hz ✓

   (b) The student will hear a periodic variation in the loudness of the sound ('beats') ✓
   The phase difference between the two waves constantly, and periodically changes due to the slight difference in frequency ✓

   When the waves are in phase they will reinforce each other, increasing the loudness of the sound. As the two waves move into anti-phase they cancel each other and the loudness of the sound decreases ✓
   The frequency of the beating effect will be: 401 – 400 = 1Hz ✓

3. (a) (i) The waves from the two slits must be in anti-phase at this point✓ i.e. phase difference of π radians. Giving destructive interference✓
   (ii) The path difference from each of the slits must be an odd number of half wavelengths ✓



   Smallest possible path difference = $\lambda/2$ = 3cm ✓
   Then $AB^2 = 33^2 - 30^2$, AB = 13.75cm ✓
   The next possible path difference, $3\lambda/2$ gives AB = 25cm which is greater than the width of the plate ✓

   (b) The pattern would vanish ✓
   Microwaves are transverse and polarised ✓
   With the plate turned through 90^o, the microwaves cannot pass through the slits ✓

Exam Workshop

This is a typical weak student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiners' mark scheme is given below.

A pair of parallel slits are illuminated with light from a sodium vapour lamp of frequency 5.09×10^{14}Hz. A series of light and dark fringes is projected on to a screen 1m from the slits.

(a) Explain why a bright fringe is always found at the centre of the pattern. [3]

There is constructive interference 1/3

The student has demonstrated that s/he knows what fringes are caused by, but has failed to explain why it occurs at the centre. S/he should have realised that this was not adequate for 3 marks

(b) The distance between the central bright fringe and the adjacent bright fringe is 1mm. How far apart are the slits? ($c = 3 \times 10^8$ ms^{-1}) [4]

$\lambda = \frac{ax}{D}$ ✓

$\lambda = \frac{a \times 1}{1} = a$ ✗

$f\lambda = c$

✓ecf

$a = \frac{c}{f} = \frac{3 \times 10^8}{5.09 \times 10^{14}} = 5.9 \times 10^7$ 2/4

The student has not changed 1 mm into 10^{-3}m when using the formula. The rest of the working is correct, but the answer should be given to 3 significant figures

Examiner's answers

(a) At the centre of the pattern, waves from each slit have travelled equal distances.i.e. path difference is zero ✓
This implies that the phase difference is zero ✓
Thus constructive interference will always occur at the centre of the pattern, producing a bright fringe ✓

(b) wavelength $\lambda = \frac{c}{f} = \frac{3 \times 10^8}{5.09 \times 10^{14}} = 5.89 \times 10^{-7}$m ✓

$\lambda = \frac{ax}{D}$

Let x = *distance between the bright fringes*
a = *distance between slits*
D = *distance between slits and screen*

$a = \frac{\lambda D}{x}$ ✓

$a = \frac{5.89 \times 10^{-7} \times 1}{(1.0 \times 10^{-3})}$ ✓

$a = 5.89 \times 10^{-4}$m ✓

Acknowledgements: This Physics Factsheet was researched and written by **Keith Penn.**
The Curriculum Press,Unit 305B, The Big Peg,120 Vyse Street, Birmingham, B18 6NF.

ISSN 1351-5136

Physics Factsheet
www.curriculum-press.co.uk Number 100

Stationary Waves on Strings and in Air Columns

This factsheet will:
- **Explain how stationary (standing) waves are formed**
- **Consider the different modes of vibration and the effect of this on musical notes**
- **Consider the similarities and differences between stationary waves on strings (eg guitar) and those in air columns (eg organ pipes)**
- **Give you practice and guidance on doing exam-style questions**

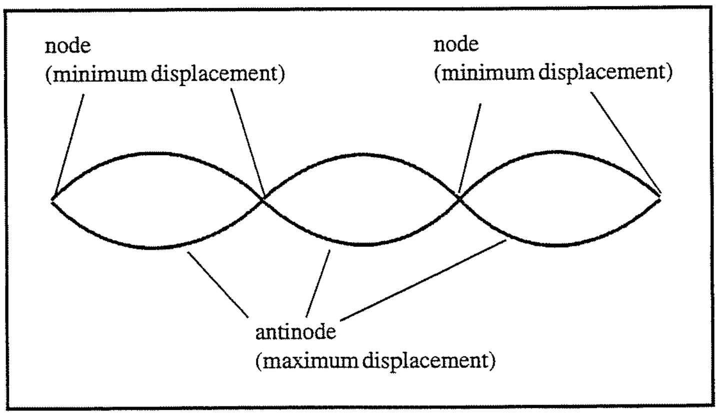
A stationary wave is formed when two waves travelling in opposite directions interfere. To make a stationary wave, the two waves must have:
- the same **speed**
- the same **frequency**
- equal or nearly equal **amplitudes**

> **One easy way for this to happen is for a wave to interfere with its own reflection.** This is what happens in stringed and woodwind instruments. Before we look at strings and air columns, though, here are some general points about stationary waves.

There are points where the displacement is always zero – these are called **nodes**.

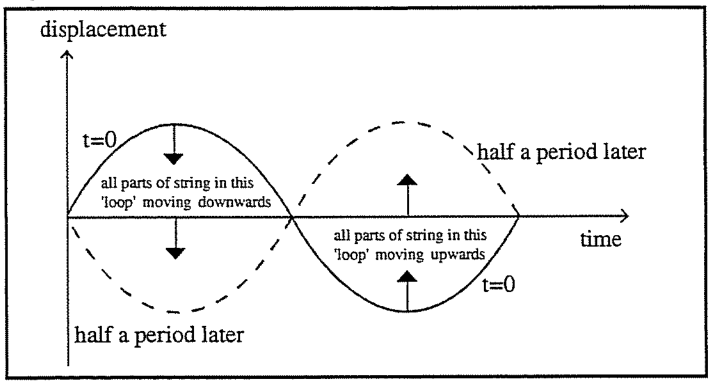
Midway between the nodes are **antinodes**, where the displacement is greatest (Fig 1).

Fig 1



At any one time all the particles in one "loop" are **in phase**, and each loop is exactly out of phase with adjacent loops (Fig 2).

Fig 2

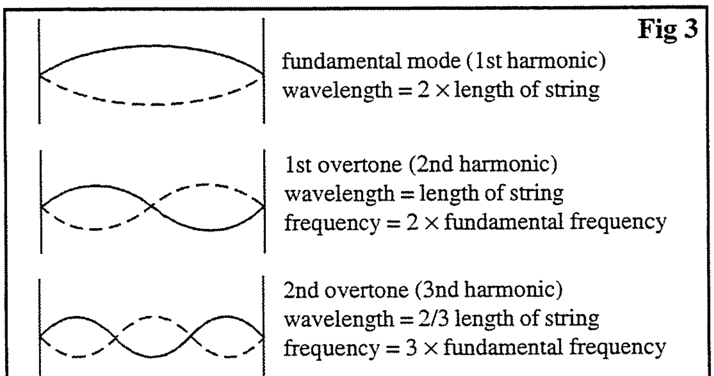


> ***Example 1***
> ***A stationary wave is set up on a wire of length 0.93 m so that it vibrates at 120 Hz. The fundamental frequency of the wire is 40 Hz.***
> ***(a) Draw a sketch of the stationary wave obtained. [1 mark]***
> ***(b) Calculate the speed of the wave. [3 marks]***
>
> ***Answer***
> *The frequency is 3 times the fundamental frequency, so we have the third harmonic* [1]
>
> 
>
> *The wavelength is the length of two loops*
> *= 2/3 × 0.93 m = 0.62 m* [1]
>
> *For the speed,* $v = f\lambda$
> $= 120\ \text{Hz} \times 0.62\ \text{m}$ [1]
> $= 74\ \text{m s}^{-1}$ [1]

Stationary Waves on Strings

When one end of a stretched string is vibrated, a travelling wave moves along the string, and reflects from the other end. *The wave interferes with its own reflection* and so a stationary wave is set up. The ends of the string *can not move*, and so must be **nodes**. This one fact allows us to draw all the possible modes of vibration.

Fig 3

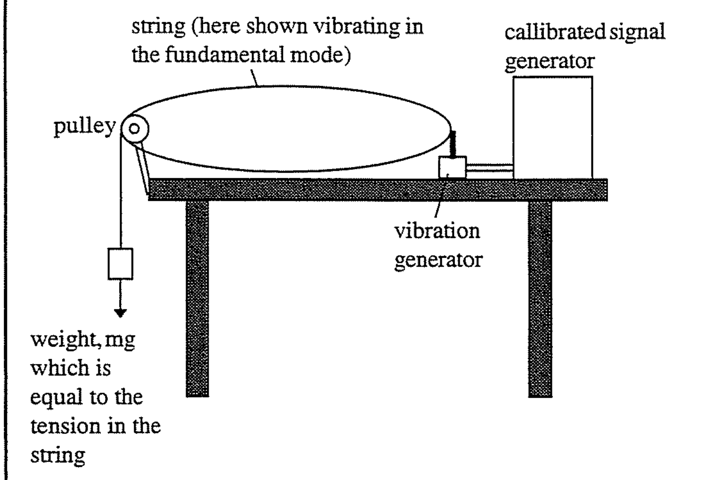


Notice that two adjacent nodes (or antinodes) are half a wavelength apart.

Exam Hint
You can always find the wavelength if you know the length of the string:
- *the wavelength is the "length of two loops".*
- *So for the ninth harmonic, which has nine loops, the wavelength is two ninths of the length of the string*

You can *force* the string to vibrate in a given mode by using a vibration generator oscillating at the required frequency Fig 4.

Fig 4



However, if simply plucked, the string will vibrate in all modes at once. The **fundamental mode**, will, in general have the largest amplitude, and the frequency of the fundamental determines the **pitch** obtained. **Superposed** on the fundamental are the other **harmonics** in varying 'strengths' – it is the relative amplitudes of these other frequencies that make middle 'C' on a guitar, say, sound different from middle 'C' on a piano. The quality of a sound, due to 2nd and higher harmonics is sometimes called its **timbre**.

The profile of a stationary wave does not move along the wire, but the two travelling waves that produce it have a speed v equal to:

$v = \sqrt{\frac{T}{\mu}}$ where T is the **tension** in the string, and μ is its **mass per unit length**

Since $v = f\lambda$ and the **wavelength** of the fundamental mode is $2L$, then the **fundamental frequency** f_0 for a stretched string is given by

$$f_0 = \frac{1}{2L}\sqrt{\frac{T}{\mu}}$$

Notice that **length**, **tension** and **mass per unit length** are the only factors affecting the pitch of the note. You knew this already: to make a guitar play a higher pitch, you:
(a) shorten the string by holding a string down
(b) tighten the string
(c) play a thinner string

Example 2
A signal generator is set at 152 Hz, 10 loops fit the length of the vibrating length of string exactly. The string is of length 2.0 m and the mass on the end of it is 0.72 kg. Calculate the mass of the string. **[5 marks]**

Answer
The wavelength is the length of two loops = 2/10 x 2.0 m = 0.4 m

For the speed, $v = f\lambda$
$= 152 \text{ Hz} \times 0.4 \text{ m}$
$= 60.8 \text{ m s}^{-1}$ *[1]*

We can now use this in
$$v = \sqrt{\frac{T}{\mu}}$$
$$\Rightarrow \mu = \frac{T}{v^2}$$ *[1]*
(check you can rearrange this correctly)

The tension, T, is equal to the weight hanging from it $= 0.72 \text{ kg} \times 9.8 \text{ N kg}^{-1} = 7.056 \text{ N}$ *[1]*

Therefore, $\mu = \frac{7.056}{60.8^2}$
$= 1.9 \times 10^{-3} \text{ kg m}^{-1}$ (1.9 g m^{-1}) *[1]*

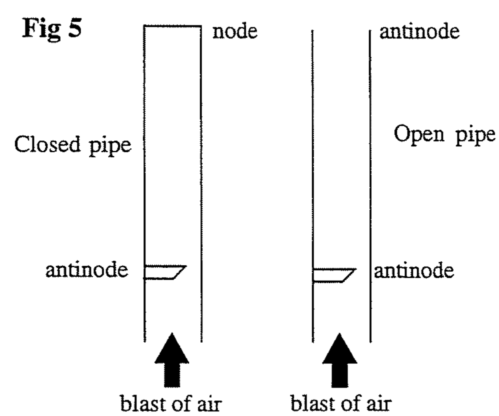
And for the 2.0 m the mass is therefore 3.8 g [1]

Stationary Waves in Air Columns
These are set up when the air at one end of a pipe is caused to vibrate. (In an organ this is done by blasting air at an edge). The resulting sound wave *interferes with its own reflection* from the other end of the pipe to produce the stationary wave.

These are very similar to stationary waves on strings. One important difference is that the stationary waves are longitudinal rather than transverse because they are formed from sound waves.

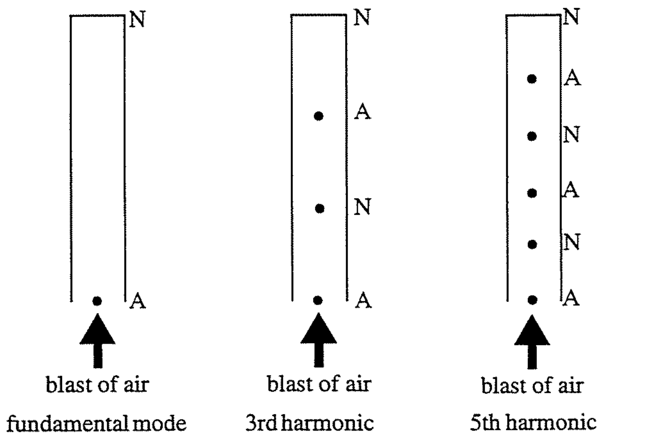
Open pipes and **closed pipes** behave slightly differently.
- In both cases at the end where the air is blasted in, the air is free to move, and so there is an antinode.
- In a closed pipe, the closed end prevents air moving and so there must be a node here.
- In an open pipe, the sound wave reflects off the free air at the end of the pipe, and since the air is free to move here, then there is an antinode at the open end.

Fig 5



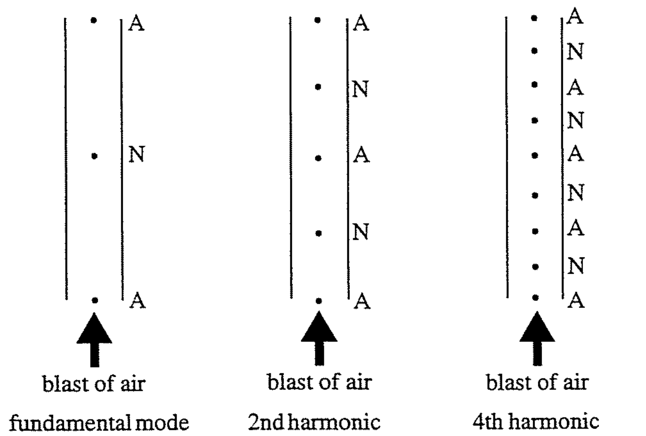
Because of this, the modes of vibration for open and closed pipes are different Fig 6.

Fig 6



Because there must be an antinode at the open end and a node at the closed end, a closed pipe can only produce odd harmonics

Fig 7



Notice that, because there is an antinode at each end, an open pipe can produce both odd and even harmonics. This gives it a richer sound than a closed pipe.

Notice also that the wavelength of the fundamental mode in the open pipe is half that of the closed pipe, and thus the fundamental frequency of an open pipe is twice that of a closed pipe.

Practice Questions

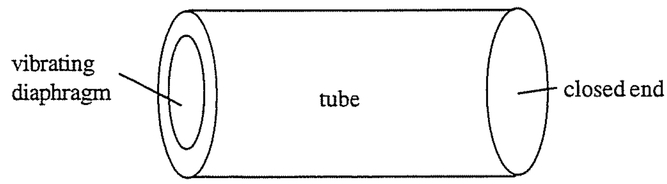
1. A horizontal string of length 0.65 m has a mass of 5.5 g and is put under a tension of 110 N. It is plucked. (a) Calculate the speed of the transverse waves on the string. (b) Calculate the wavelengths and frequencies of the fundamental and second harmonic.

2. The fundamental frequency of vibration of a stretched wire is 120 Hz. Calculate the new fundamental frequency if (a) the tension in the wire is doubled, the length remaining constant, (b) the length is doubled with constant tension, (c) the tension in the wire is doubled and the length of the wire is also doubled.

3. A guitar string is 0.70 m long. The string is tuned so that when its full length is plucked it vibrates at a frequency of 384 Hz. To play a higher note, the string is pressed so that the length free to vibrate is shorter. A fret (ridge) on the neck of the guitar ensures that the correct length is produced when the string is pressed. A certain fret is positioned so that when it is used, the frequency of the note obtained is 427 Hz. What length of string is vibrating now?

4. What is the frequency of the sound emitted by an open-ended organ pipe 1.7 m long when sounding its fundamental frequency, if the speed of sound in air is 340 m s^{-1}?

5.



In the diagram above, when the diaphragm vibrates at 2000 Hz a stationary wave pattern is set up, and the distance between adjacent nodes is 8.0 cm. When the frequency is gradually reduced, the stationary wave pattern disappears, but then reappears at a frequency of 1600 Hz. Calculate:
(a) the speed of sound in air [2 marks]
(b) the distance between adjacent nodes at 1600 Hz [2 marks]
(c) the next lower frequency at which a stationary wave is obtained [1 mark]
(d) the length of the tube [2 marks]

6. A piece of glass tubing is closed at one end by covering it with a sheet of metal. The fundamental frequency is found to be 280 Hz. Calculate the length of the tube. If the metal sheet is now removed, calculate the wavelengths and frequencies of the fundamental and second harmonic of the resulting open pipe.

Answers

1. 114 ms^{-1},
fundamental: wavelength = 1.30 m, f = 87.7 Hz
second harmonic: wavelength = 0.65 m, f = 175 Hz

2. 170 Hz, 60 Hz, 84.9 Hz

3. 0.63 m

4. 100 Hz

5. (a) 320 m s^{-1}
(b) 0.10 m apart
(c) 1200 Hz
(d) wavelength is 0.80 m, the tube is 0.40 m long

6. 0.304 m
0.607 m, 560 Hz. 0.304 m, 1120 Hz

Acknowledgements:
This Physics Factsheet was researched and written by Michael Lingard
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

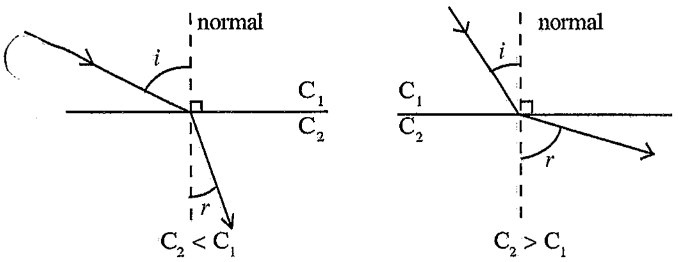
www.curriculum-press.co.uk Number 131

Refraction Calculations

Examiners' reports have drawn attention to weaknesses candidates have displayed in calculations concerning light rays. Calculations involving refraction caused most difficulties. This Factsheet will summarise the theory involved, but will concentrate on examples and problems.

Refraction

A light ray changes direction at a boundary between two media, if its speed changes.



(For the sake of this Factsheet, we are ignoring any reflected ray from the surface.)

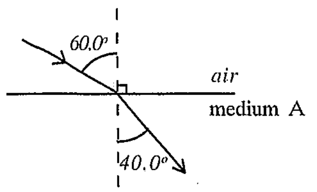
> *A light ray refracts* ***away*** *from the normal when it speeds up, and* ***towards*** *the normal when it slows down.*

The equation linking angle and speed:

$$\frac{\sin i}{\sin r} = \frac{C_1}{C_2}$$

where i and r are the angles of incidence and refraction measured to the normal to the surface.

Example 1: The speed of light in air is $3.00 \times 10^8 ms^{-1}$. Calculate the speed of light in medium A.

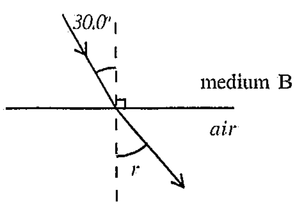


Answer:

$$\frac{\sin 60.0}{\sin 40.0} = \frac{3.00 \times 10^8}{C_2}$$

$$C_2 = 2.23 \times 10^8 ms^{-1}$$

Example 2: The speed of light in medium B is $2.40 \times 10^8 ms^{-1}$. Calculate the angle of refraction.



Answer:

$$\frac{\sin 30.0}{\sin r} = \frac{2.40 \times 10^8}{3.00 \times 10^8}$$

$$r = 38.7°$$

> *Exam Hint: Check that you have not reversed the data in your calculations by matching greater speed with larger angle (measured to the normal) in your results.*

Index of refraction (refractive index)

A number is assigned to every transparent material according to the speed of light through the material. This value, n, is the refractive index of the material. n = 1.00 for air or a vacuum. For all other materials, n is greater than one.

A few examples:

Material	Refractive index, n
air / vacuum	1.00
glass	1.50
water	1.33

Different types of glass will have different refractive indices. The value of the refractive index can also depend on the wavelength of the light. We will use the values in this table for calculations.

The previous equation can now be written:

$$\frac{\sin i}{\sin r} = \frac{C_1}{C_2} = \frac{n_2}{n_1}$$

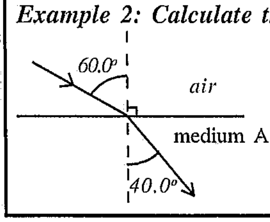
Example 1: Find the speed of light through water and glass.

Answer:

Water: $\frac{c}{3.0 \times 10^8} = \frac{1.00}{1.33}$ $c = 2.26 \times 10^8 ms^{-1}$

Glass: $\frac{c}{3.0 \times 10^8} = \frac{1.00}{1.50}$ $c = 2.00 \times 10^8 ms^{-1}$

Example 2: Calculate the refractive index of our medium A.

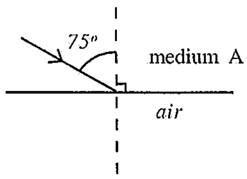


Answer:

$$\frac{n}{1.00} = \frac{\sin 60}{\sin 40}$$

$$n = 1.35$$

Example 3: A ray of light through this medium A is incident on a glass surface at an angle of 75° to the normal. Calculate the angle of refraction.

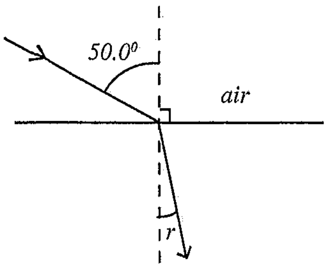


Answer: The sine of the angle of refraction works out to be greater than one. This is impossible. No refraction occurs.

> *When light speeds up at a boundary, and the angle of incidence is large, refraction may not occur. Total Internal Reflection occurs. See Factsheet 86 on Optical Fibres for a detailed discussion.*

It is interesting to see what happens when light refracts through two layers. First of all, suppose a ray of light travels from air into glass.

Example 1: Find the angle of refraction in the situation shown.



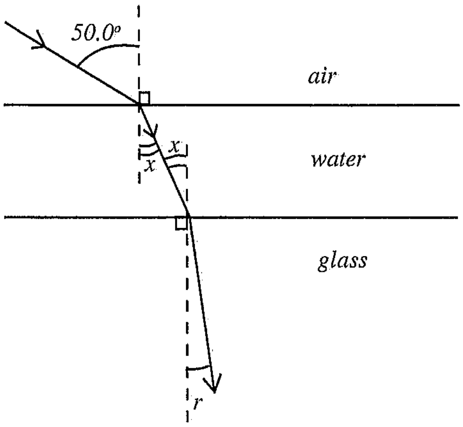
Answer:

$$\frac{\sin r}{\sin 50.0} = \frac{1.00}{1.50}$$

$$r = 30.7^{o}$$

Now suppose a layer of water is lying on the glass.

Example 2: Find the final angle of refraction.



Answer:

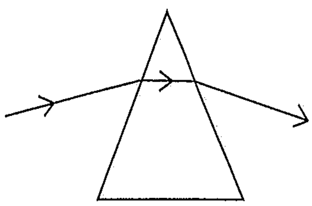
First boundary $\frac{\sin x}{\sin 50.0} = \frac{1.00}{1.33}$ $x = 35.2^{o}$

Second boundary $\frac{\sin r}{\sin 35.2} = \frac{1.33}{1.50}$ $r = 30.7^{o}$

It turns out that the intervening layer of water has no effect on the direction of the final refracted ray.

Non-parallel surfaces

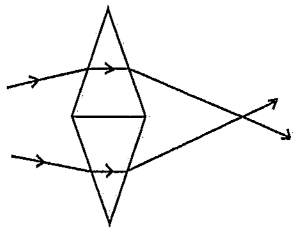
Parallel surfaces e,g, a glass block, result in the incident ray emerging parallel to its original direction. But non-parallel surfaces e.g. a triangular prism, can cause an increased deviation in the path.



(So far we have ignored the fact that different wavelengths – different colours – of light travel at different speeds through media like glass. This causes different angles of refraction, leading to a spectrum being produced from white light.)

Lenses

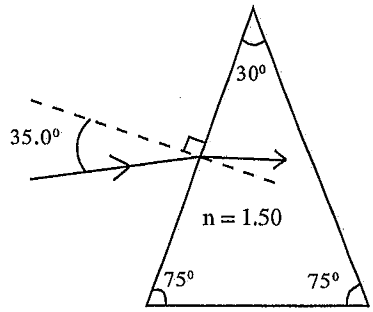
Combining two prisms shows how non-parallel surfaces can be used to focus light rays.



It is not a major step to see how curved, rather than straight, surfaces can focus all the rays from an object at one image point.

Exam-style question

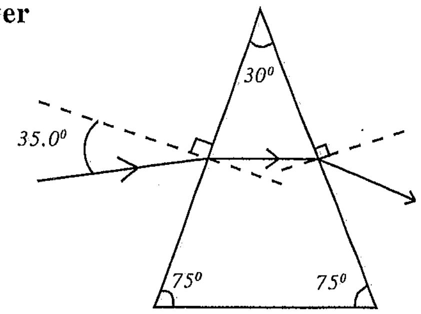
A monochromatic ray of light is incident on the surface of the triangular prism shown in the diagram.



(a) Sketch the path of the ray through the glass and emerging into the air.
(b) Calculate the angle of refraction at the first surface.
(c) Calculate the angle of incidence at the second surface.
(d) Calculate the angle of refraction at the second surface.
(e) Calculate the deviation of the emergent ray from its original direction.

Answer

(a)



(b) $r = \sin^{-1}\left(\frac{1.00}{1.50} \times \sin 35.0\right) = 22.5^{0}$

(c) $30 - 22.5 = 7.5^{0}$

(d) $r = \sin^{-1}\left(\frac{1.50}{1.00} \times \sin 7.5\right) = 11.3^{o}$

(e) deviation $= (35.0 - 22.5) + (11.3 - 7.5) = 16.3^{0}$

Practice Questions

1. An intervening layer has no effect on the direction of the final refracted ray. What effect does it have on the ray?

2. With an intervening layer, the equations could be written:

$$\frac{\sin x}{\sin y} = \frac{n_y}{n_x} \quad \text{and} \quad \frac{\sin y}{\sin z} = \frac{n_z}{n_y}$$

Show mathematically that the intervening layer should not affect the final angle z.

3. Total internal reflection can occur when light travels from our medium A (n=1.35) into air. Find the direction of the largest incident angle where refraction will occur.

4. The speed of light through air is $3.00 \times 10^8 ms^{-1}$, and the refractive index of air is 1.00. Use the equation to calculate the speed of light through the other materials in the table.

Material	Refractive index	Speed of light (ms^{-1})
Air	1.00	3.00×10^8
Lead fluoride	1.76	
Diamond	2.42	
Paraffin	1.44	
Ice	1.31	
Glycerol	1.47	

Answers

1. It will displace the final ray to the side i.e. parallel to the ray without the intervening layer.

2. $$\frac{\sin x}{\sin y} \times \frac{\sin y}{\sin z} = \frac{n_y}{n_x} \times \frac{n_z}{n_y}$$

Cancelling identical terms leads to:

$$\frac{\sin x}{\sin z} = \frac{n_z}{n_x}$$

This is the same equation that you would use if you had only the media x and z, with no intervening layer.

3. The refracted angle has a maximum value of 90° (sin r = 1.0).

$$\frac{\sin i}{\sin 90} = \frac{1.00}{1.35} \quad \text{so } i = 47.8^0$$

4.

Material	Refractive index	Speed of light (ms^{-1})
Air	1.00	3.00×10^8
Lead fluoride	1.76	**1.70×10^8**
Diamond	2.42	**1.24×10^8**
Paraffin	1.44	**2.08×10^8**
Ice	1.31	**2.29×10^8**
Glycerol	1.47	**2.04×10^8**

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press,Bank House, 105 King Street,Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 86

Optical Fibres in Communication

We generally see optical fibres used in exciting products such as fibre optic table lamps and lights that change colour on artificial Christmas trees.

However optical fibres are extremely important these days in the communications industry, and are becoming more important all the time. In America they are introducing fibre optic cables into homes with the capability of bringing in the signals for high definition interactive television, streaming video, telephones, and internet broadband (ten times faster than what we presently regard as broadband). And you can still have an artificial Christmas tree in the corner of the room.

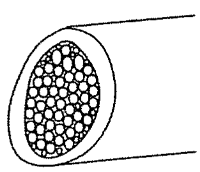
Of course there are other important uses of optical fibres outside of communications. An example would be the endoscope used in medicine to see inside the body. However we will concentrate on communications in this Factsheet.

Advantages and disadvantages of Optical Fibres

The only real alternatives to optical fibres at the moment are copper wires or coaxial cable. For optical fibre:

Advantages

1. Optical fibres use light. There is no danger of fire hazard or electrical shock. Adjacent fibres won't interfere with the signals being carried by each other. There is no electrical or magnetic field surrounding the fibre, meaning little danger of signals being "tapped". They are also immune from outside interference.
2. An optical fibre can carry more information per second than a copper wire. In addition the signal travels faster and further.
3. Glass won't corrode like metals.
4. The fibres are lighter, thinner, and easier to handle than electrical wiring. A great many fibres can be bundled into one cable.

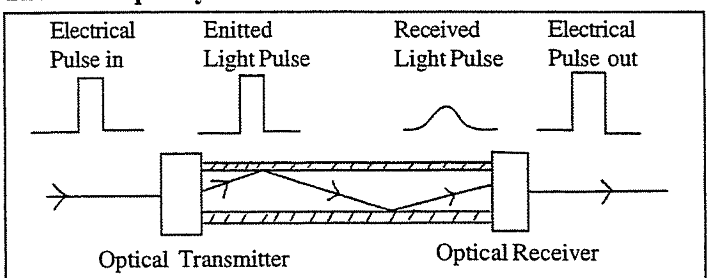


5. As there is less signal degradation in optical fibres, only very low power is required, and there are fewer "repeaters" to boost the signal. This all saves money.

Disadvantages

1. The fibres are more expensive to manufacture. But the fewer repeaters and lower running costs outweigh this disadvantage.
2. Optical fibres require more expensive signal sources and detectors. Also it is more difficult to join lengths of fibre together or construct junctions.

The Fibre Optic System



An electric pulse is fed into an **optical transmitter**. This produces a light pulse which mirrors the electrical pulse received.

Some of the words or phrases highlighted will be required. However these differ between syllabuses and modules. Find out what you need to know.

This light pulse is then transmitted along the optical fibre. Some degradation occurs in the signal.

The **optical receiver** changes the received light pulse into a better defined electrical signal, matching the signal from the source. (This accuracy is only really possible with digital pulses.)

Because of **attenuation** (or energy loss) in the signal due to the scattering and absorption of light by the glass, **repeaters** are inserted periodically in long systems to boost the signal.

In the rest of this Factsheet we will be looking at the types of fibres used, their advantages and disadvantages in signal transmission, and the types of transmitters and receivers used.

Typical exam question

(a) What do we mean by <u>attenuation</u> in optical fibres?.

(b) State two ways in which energy is lost along the length of an optical fibre.

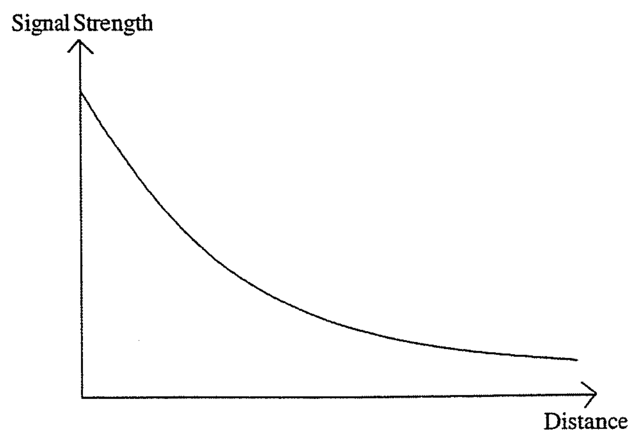
(c) If a fibre loses 5% of its signal strength per km, how much of its strength would be left after 20 km?

Answers

(a) Attenuation is a measure of the rate of loss of signal strength along the length of the fibre.

(b) Energy is lost by the scattering and absorption of the light rays as they travel through the glass fibre.

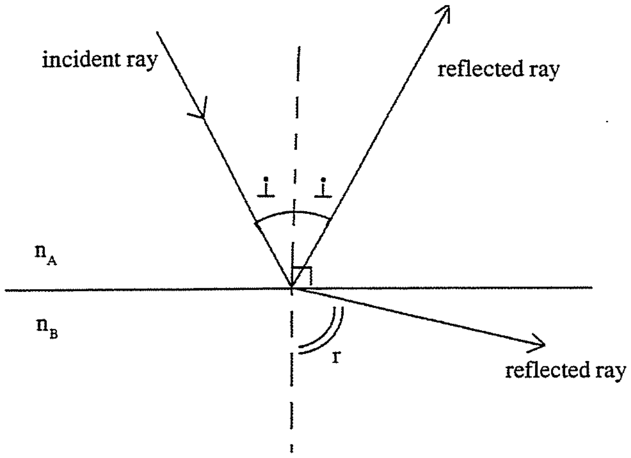
(c) After each km it retains 95% of the signal strength it had at the beginning of that km. This leads to an exponential decay curve.



After 20km: Signal strength = $0.95^{20} \times A = 0.36\,A$ (where A is the original signal strength).

How optical fibres work

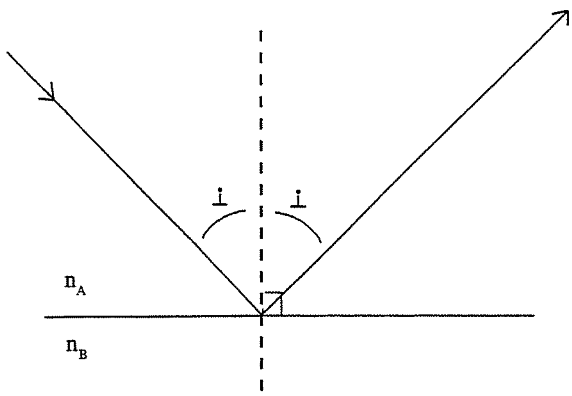
Optical fibres are designed to make use of **total internal reflection.**



When light rays travelling through medium A are incident on a boundary with medium B, we expect some reflection and some refraction to occur. The refraction obeys the law:

$(\sin i) / (\sin r) = n_B / n_A$

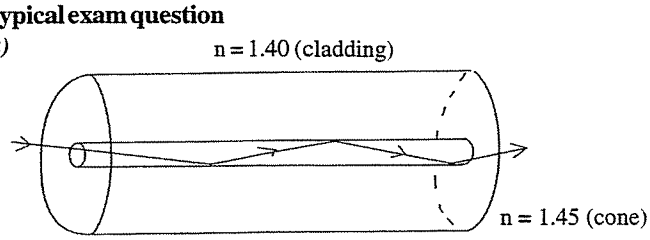
If light travels more quickly through medium B, then at a certain angle of incidence, the angle of refraction will be 90 degrees (no light escapes). We call this angle of incidence the **critical angle, c.** For all angles of incidence greater than the critical angle, total internal reflection occurs.



In an optical fibre, a glass core is surrounded by glass or plastic cladding with a smaller index of refraction. This leads to total internal reflection.

Typical exam question

(a)



If the indices of refraction are as shown in the diagram, find the critical angle.

(b) ***For this same set-up, what value would be required for the index of refraction of the cladding if the critical angle is to be 80 degrees? (The index of refraction of the core is unchanged.)***

Answers

(a) $\sin c / 1 = 1.40 / 1.45$ $\quad c = 75$ *degrees.*

(b) $\sin 80 / 1 = n / 1.45$ $\quad n = 1.43$

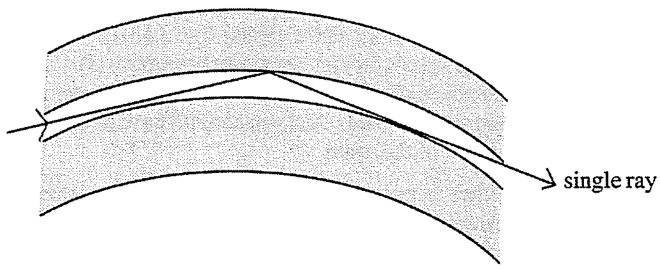
Types of Optical Fibres

There is more than one type of fibre. Each has its own advantages and disadvantages. Each is designed for a specific purpose.

> ***Exam Hint:*** *Learn the advantages and uses of the different types of optical fibre.*

1.Single Mode

This is the simplest to understand. Although the outside diameter of the cladding is about 120 microns (1.2×10^{-4} m), the light-carrying core has a diameter of less than 10 microns. This means that with a laser diode source, only a single ray of light travels along the core.



We call this single mode transmission. There is very little deterioration in the signal. With a high quality glass core and a powerful source laser emitting a single wavelength of light, the effective range for this fibre can be several hundred km. These fibres are used for telephony and cable television.

Single mode cables are more expensive to manufacture than multimode, and need a highly directional, powerful light source.

Typical exam question

(a) ***Estimate the length of time it would take a fibre optic system to carry a signal from the UK to the USA under the Atlantic. (Take $c = 2 \times 10^8$ ms^{-1} in the cable. Estimate the length of the cable under the sea.)***

(b) ***Estimate the length of time it would take a microwave signal to travel from the UK to the USA via a satellite link. (Geosynchronous satellites orbit at a height of about 36 000 km above the Earth's surface.)***

(c) ***Which would give less delay in a telephone conversation?***

Answers

(a) Length of cable about 5000 km

$$t = \frac{s}{v} = \frac{(5 \times 10^6)}{(2 \times 10^8)} = 0.025 \text{ s}$$

(b) Path of microwave about 75 000 km

$$t = \frac{(7.5 \times 10^7)}{(3 \times 10^8)} = 0.25 \text{ s}$$

(c) The delay using the optical fibre is not noticeable. Remembering that the signal delay there and back would be double the value estimated, this would be noticeable with a satellite link.

2. Multimode

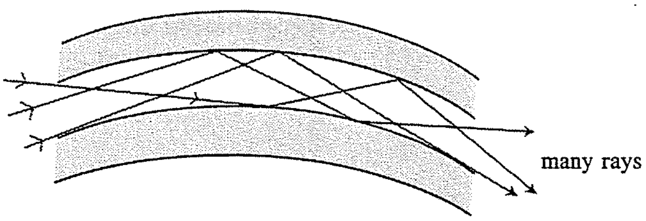
These cables have a much wider core – perhaps 60 microns in diameter. They can be used with cheaper, less directional LED sources. However they have other problems.

There are two sorts of multimode cable:

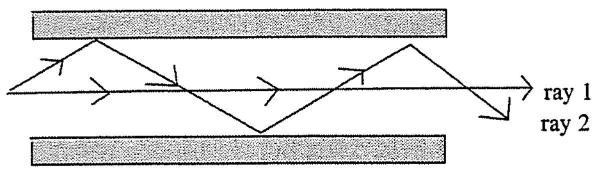
- **Step-Index Multimode**
- **Graded-Index Multimode**

Step-Index Multimode

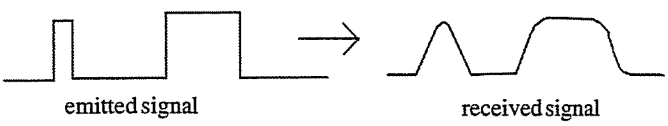
This is a simple cable with a wide core of uniform refractive index, surrounded by a glass cladding. The sudden change of refractive index causes total internal reflection. This leads to rays reflecting along inside the core at a variety of angles. Each of these rays is called a mode.



Let us look at two modes:



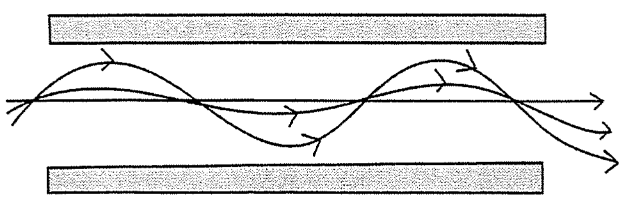
It is obvious that ray 2 travels further than ray 1 as it moves along the core. When the LED emits a pulse of light into the fibre, some rays will reach the far end of the cable before others. This causes a blurring of the signal.



This effect is called **modal dispersion.** It limits the **bandwidth** (to be discussed later) and means that the fibre has a short range, as the amount of dispersion increases with distance from the source. Repeaters are required to "clean up" and re-emit the signal.

Graded Index Multimode

The core has variations in its composition to largely counteract the different path lengths of the modes. The index of refraction of the core gradually decreases from its centre to its outside.



The rays that travel the greatest distance spend more of their time in the outer regions of the core, where they can travel more quickly.

An LED source can be used. Modal dispersion is decreased, leading to greater bandwidth and increased range.

Multimode fibres are often used in local area networks where a great range and very high bandwidth are not essential. The effective length may be only a few km.

Limitations

Let's look again at step-index fibres.

Example:

In a step-index fibre, one ray may travel 75 metres further than another while carrying a signal along a 5 km length of cable. And suppose we are transmitting narrow pulses at a frequency of 1 Mhz. If the speed of light through glass is about 2×10^8 ms^{-1}, how far apart are these pulses?

Distance = speed x time

The time between pulses is 1×10^{-6} s.

Distance between pulses = $2 \times 10^8 \times 1 \times 10^{-6}$ m = 200 m.

The 75 metre spreading of the signal is perhaps acceptable.

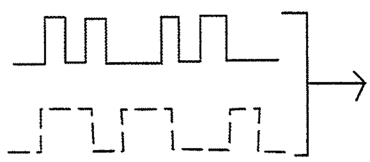
But if we tried to transmit these pulses at a frequency of 10 Mhz, the distance between pulses would be just 20 m. the modal dispersion of 75 m would completely blur the signal.

The **bandwidth** of a fibre is the maximum frequency at which a signal can be transmitted. To increase the bandwidth available for this fibre, we would have to shorten it. It is not suitable for long distance transmission.

> The higher the bandwidth of a fibre, the more data can be transmitted per second.

Multiplexing

One way to increase the quantity of data transmitted through a fibre is to superimpose signals. This is called **multiplexing.** With optical fibres, Dense Wavelength Division Multiplexing (DWDM) can be used. Several signals, each transmitted in a different colour (different wavelengths) are sent along the fibre together.



At the far end, coloured filters could separate the signals.

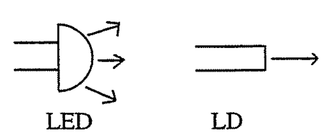
This is equivalent, in an electrical system, to superimposing signals on different carrier frequencies. An electronic tuner could then select the required signal.

Optical Transmitters and Receivers

There are two different types of light emitters used with optical fibres.

(a) LED. With multimode fibres, where the core is quite wide and rays can spread out, a **light emitting diode** is a suitable source. Usually wavelengths in the infrared region are used. An LED is an inexpensive and reliable source.

(b) LD. With single mode fibres, where the core is very thin, a **laser diode** is required. It emits a strong single ray of light that can be aimed straight along the fibre. However laser diodes are more expensive and require more complex circuitry.



The standard optical receiver is a **photodiode.** Light falls onto the junction of a reverse-biased diode (no current flow). The light gives electrons enough energy to cross the junction and complete the circuit. The optical pulse becomes an electrical signal.

Questions

1. Give three advantages of optical fibres over copper wire in communications systems.

2. Define **critical angle** and **total internal reflection**.

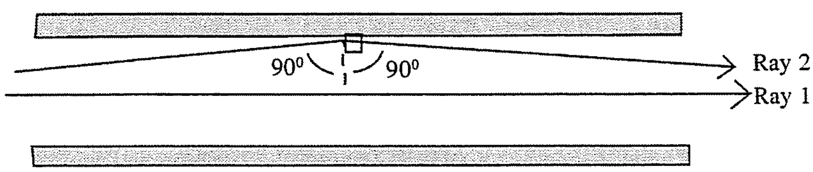
3. An optical fibre has a glass core (with index of refraction n = 1.46) surrounded by a glass cladding (with index of refraction n = 1.43). Find the critical angle for this fibre.

4. A high quality optical fibre loses 2% of its signal strength over a distance of 10 km. What percentage of its signal strength remains after 200 km?

5. (a) Name the 2 modes of optical fibre used.

   (b) Give one advantage and disadvantage of each.

6. In a step-index multimode fibre where the critical angle is 80 degrees, how much further would ray 2 travel than ray 1 over a 5 km length of fibre?



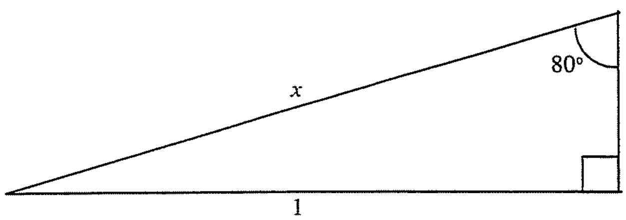
Answers

1, 2 see text for answers

3. sin c = 1.43 / 1.46, c = 78.4 degrees

4. $0.98^{20} = 0.67 = 67\%$

6.



sin 80deg = 1 / x

So x = 1.015

Over a 5 km fibre, the distance travelled is 1.015 × 5000 = 5075 m.

The increase in distance is 75 m.

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

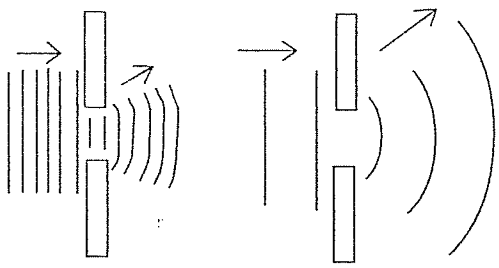
Physics Factsheet

www.curriculum-press.co.uk Number 78

Diffraction Effects

Diffraction seems an unusual effect to most of us. We tend to think of waves or rays travelling in straight lines. They may reflect off a surface or change direction when entering a different medium, but we expect them to travel in a straight line otherwise.

The curving of waves within a medium caused by diffraction goes against our usual assumptions, and can be quite a problem at times. However we can also make use of this wave property.



Diffraction causes waves to bend around edges, or spread out through gaps. We often say that diffraction is a maximum when the wavelength is similar to the width of the aperture. If the aperture becomes smaller than the wavelength, diffraction will, in fact, increase. However the total wave energy passing through the gap decreases, making the diffraction effect less noticeable. So perhaps it is reasonable to think in terms of wavelength matching aperture size. It certainly makes calculations easier.

> *Strong diffraction occurs when the wavelength is equal to the width of the aperture.*

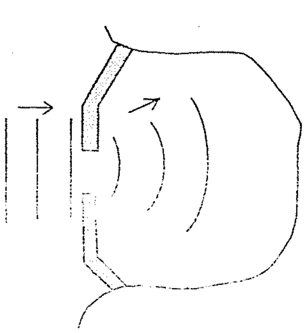
What we are going to look at in this Factsheet is diffraction, and its effects in a variety of wave motions (noting similarities and differences, uses and difficulties). We will also consider some basic mathematical relationships involving diffraction.

Answers to the questions posed will be given at the end of the Factsheet. It will prove useful to attempt each question as it occurs, rather than all together at the end.

A more in-depth look at diffraction with light, including more complex maths, will feature in a follow-up Factsheet.

Water waves

We build harbour walls with narrow entrances to shelter boats from large waves. But diffraction causes the waves to spread out within the harbour, limiting the value of the harbour walls:



Problem 1

(a) A wider opening would reduce diffraction within the harbour, but what are the problems with this?

(b) A narrower opening would increase diffraction. Is this necessarily counter-productive?

Another interesting diffraction effect can be seen with boats. A small boat often has a strong wave disturbance directly behind it, where a larger, wider boat does not.

Problem 2

Using ideas of diffraction and interference, can you explain this effect?

Sound waves

Sound travels at about 330 ms^{-1} through air (depending on temperature). This means that the wavelengths of audible sound tend to match aperture or obstacle size around us in nature (and in buildings). So diffraction effects are very large with sound waves.

> *Note that diffraction occurs both for transverse and longitudinal oscillations, and also for both mechanical and electromagnetic waves.*

Problem 3

If we can hear sounds between 50 Hz and 20 000 Hz, find the range of wavelengths (in air) to which these frequencies correspond. (Notice how this range spans the size of many things around us, leading to diffraction.)

Owl hoots are at a relatively low frequency compared to songbirds. This longer wavelength leads to greater diffraction effects, bending the sound around trees and hillsides. The owl can communicate its presence over considerable distances. However the diffraction makes it very difficult for any prey to successfully locate the exact position of the owl.

Foghorns operate at low frequencies (long wavelengths). Diffraction means ships can hear the sound around (or over) islands and headlands that may be between the ships and the foghorn. However, again, locating the direction to the sound source is difficult.

Problem 4

Suppose a sound system is playing in the house. You are in the back garden with the door open. When you are in line of sight to the speakers, you hear more treble; when you are off to the side you hear more bass. Explain this.

In a concert hall, it is important that the sound from the speakers fans out across the hall. But it is also important to limit the sound intensity reaching the ceiling, as the reflections will increase echoing (reverberation) within the hall.

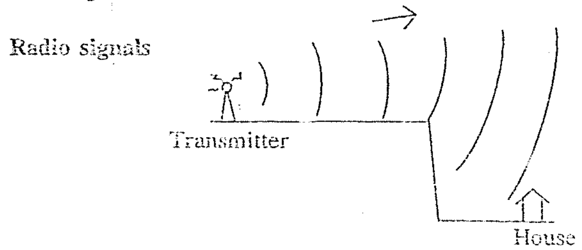
Problem 5:

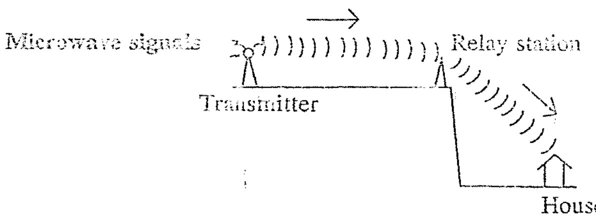
What shape would you choose for the loudspeaker apertures within the concert hall? (Should they be tall and thin, or short and wide?)

Radio Waves and Microwaves

Diffraction is useful for radio transmission, especially for long and medium wave signals. These longer wavelengths will diffract over hills and around buildings, making the signal accessible in 'awkward' areas.

However FM radio and television make use of shorter wavelengths for their carrier frequencies, extending into the microwave region. Shorter wavelengths mean less diffraction. Relay stations are needed to maintain 'line of sight' transmission.





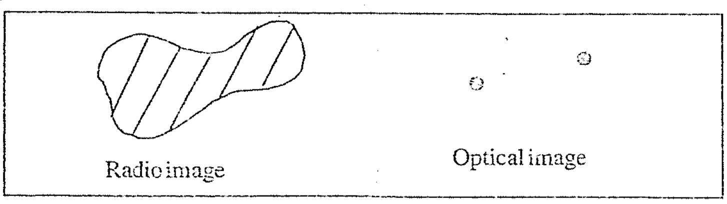
Radio telescopes are used to study radio transmissions from space. We use the term **resolution** to describe the ability of a device to separate the images of two sources which have only a small angular separation.

The resolution of a telescope (optical or radio) depends on the ratio of the wavelength to the aperture. To minimise diffraction (and improve resolution), the aperture should be as large as possible, compared to the wavelength being observed.

Limit of resolution $\propto \lambda/a$ *where a is the diameter of the aperture.*

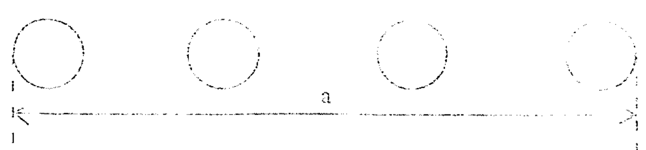
> **Problem 6**
> Suppose an optical telescope of lens diameter 1m is used to observe light of wavelength 5×10^{-7} m. What diameter must the reflector of a radio telescope be in order to produce the same resolution when observing radio waves of wavelength 10 cm?

Obviously we cannot build radio dishes with diameters measured in kilometres (see solution to Problem 6). Resolution is much less with radio telescopes than optical telescopes. Radio sources appear as blurs, while the same sources can be observed as pinpoints optically.



It might then seem pointless studying radio emissions, when the images from optical observations are so much sharper. However the radio data provides different information, which might be useful in its own right. Sometimes optical observations are made to positively identify the source of the radio waves.

One way in which resolution can be improved with radio telescopes is by linking the signals received from a line of small dishes:



The effective aperture, a, is very much increased using this setup, improving the resolution. However the total energy collected (the 'brightness' of the signal received) depends on the area of the dishes. So this system increases resolution significantly, but does not do much for the intensity of the signal received.

> **Problem 7**
> A line of 9 dishes, each of diameter 1.0m with a 50.0m gap between adjoining dishes, is used to detect radio emissions from a source in space. Compare the energy collected by these dishes with that which would be collected by a dish with a real diameter equal to the effective diameter of this system.

One place you may have seen microwave diffraction in the laboratory is to model X-ray or electron diffraction in crystallography. X-rays and electrons can have wavelengths similar to the spacing between atomic planes. We can simulate the diffraction effects observed with model atomic lattices of plane spacing similar to the wavelength of the microwaves used (about 3 cm).

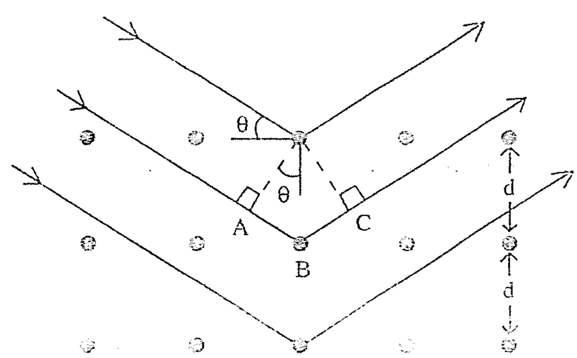
X-rays and Electrons

It may seem strange to group these topics, but both X-rays and electrons are used to study atomic structure.

X-rays are, of course, high energy electromagnetic radiation. They are produced by bombarding a metal target with a beam of accelerated electrons. The X-rays produced have wavelengths spanning those required to undergo diffraction by crystal planes.

We usually think of electrons as particles, but the wave-particle duality theory tells us that they have a wavelength associated with them, and electrons of the right energy will display diffraction effects when directed into crystals.

The diffraction theory is similar for both:



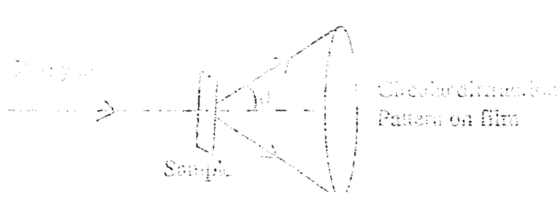
The receiver will measure an interference maximum when the extra distance travelled by the second ray is a whole number of wavelengths:

$n\lambda = 2d \sin \theta$, where n=1, 2, 3, etc.

(The "2" in this equation comes from the path difference between two adjacent rays being AB + BC in the diagram.)

If we know the wavelength and measure the angle, we can work out the plane spacing for the crystal. (This is usually of the order of 10^{-10} m.)

If we select one specific known X-ray wavelength (using filters), then the crystal would have to be oriented at exactly the right angle to give a diffraction pattern. But if we use a powder (or an equivalent), where the crystal grains are randomly at all possible orientations, we will always get a diffraction pattern.



Problem 8
If a beam of X-rays of wavelength 2.1×10^{-10} m is directed at a thin metal sheet (composed of a large number of randomly oriented metal grains), a diffraction pattern is produced. It is found that rays deflected through 36 degrees produce the first interference maximum. Find the spacing between the atomic planes in the metal.

An alternative to these transmission patterns, where the X-ray beam penetrates the sample, is the back reflection method. Diffraction patterns are obtained from X-rays reflected back from the surface of a sample.

With electrons, we select the wavelength by choosing the voltage through which we accelerate them.

$E = eV$ (usually quoted in electron-volts)

$p = mv = h/\lambda$ (de Broglie's equation)

h is Planck's constant
m and e are the rest mass and charge of an electron.

De Broglie's equation links wave and particle properties of matter.

By combining these equations (and remembering that $E = \frac{1}{2}mv^2$), we reach our final relationship:

$$\lambda = h / \sqrt{(2meV)}$$

Problem 9
(a) If we accelerate an electron through 200V, what energy do we give it (stated in joules)?
(b) What wavelength would this energy translate to for an electron?
(c) Would this be suitable for studying crystal structure using electron diffraction?

Problem 10
What voltage would produce electrons with a wavelength of exactly 3.0×10^{-10} m?

Electron diffraction can also be used to study the nucleus. Here we need wavelengths of about 10^{-15} m for diffraction patterns. If you use the above equations, you will find that the speeds required of the electrons are greater than the speed of light.

However it is possible to produce these wavelengths. As the electrons approach the speed of light, relativistic calculations replace the above equations with the simple expression:

$$E = eV = hc/\lambda$$

A quick calculation will show you that if you can accelerate the electrons through several million volts, you can achieve very short wavelengths.

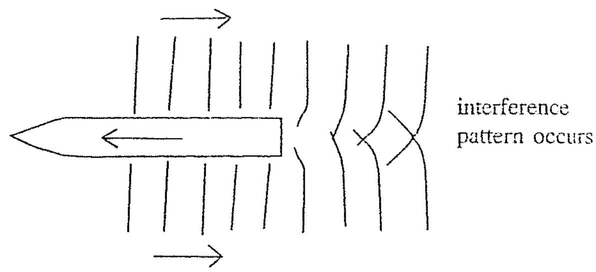
Visible Light
A separate Factsheet will deal with diffraction effects with visible light. Factsheet 81 will include:

(a) Single slit diffraction (through gaps and holes)
(b) Multiple slits
(c) Transmission and reflection gratings
(d) Crossed gratings
(e) Resolution with optical devices
(f) Mathematical calculations with the above topics
(g) Difficulties caused by diffraction
(h) Uses of diffraction e.g. spectroscopy

Answers

Problem 1 Solution
(a) More wave energy would enter the harbour. Reflections from inner harbour walls could still spread the wave energy throughout the harbour. So increasing the harbour opening is not a sensible option.
(b) Although diffraction would increase, the total wave energy entering the harbour would decrease, reducing the danger to the boats sheltering in the harbour. However a narrower harbour entrance would make navigation into the harbour more difficult in poor weather conditions.

Problem 2 Solution
With a narrower boat, waves diffracting around the back corners can meet behind the boat, leading to a strong interference effect.



(My apologies to readers if there is a nautical term for 'back corners'.)

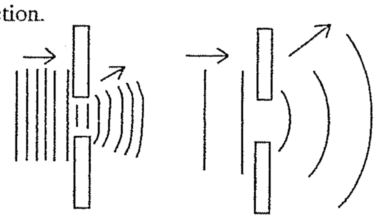
Problem 3 Solution:

$$\lambda = \frac{v}{f} = \frac{330}{50} = 6.6\text{m}$$

$$\lambda = \frac{v}{f} = \frac{330}{20\,000} = 0.017\text{m} = 1.7\text{cm}$$

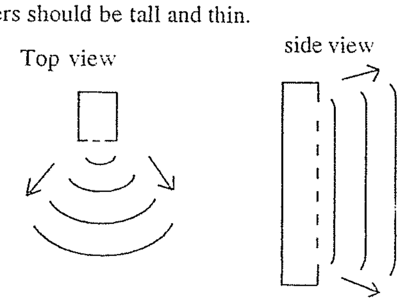
This range of wavelengths for audible sound makes diffraction very noticeable.

Problem 4 Solution:
Through the same aperture, longer wavelengths (bass notes) experience greater diffraction.



A greater proportion of the energy from the bass notes will be diffracted through greater angles. The 'line of sight' sound will have lost more energy from the bass, sounding more treble. Off to the side, the music will sound more bass (deeper).

Problem 5 Solution:
The speakers should be tall and thin.



Diffraction from the narrow horizontal aperture fans the sound out across the hall.

The large vertical aperture reduces upward diffraction. This minimises sound energy reflecting from the ceiling.

Several other measures are taken to optimise the acoustics within the concert hall, sometimes by manipulating other properties of waves. But

Problem 6 solution

$(\lambda/a)_{optical} = (\lambda/a)_{radio}$

$$\frac{5\times10^{-7}}{1} = \frac{0.1}{a}$$

$$a = \frac{0.1}{5\times10^{-7}} = 200\ 000\ m = 200\ km.$$

So the radio dish required to produce the same resolution would have a diameter of 200 km.

Problem 7 solution

Diameter of big dish = $(9 \times 1.0) + (8 \times 50.0) = 409$ m

Area = $\pi \times r^2 = \pi \times 205 \times 205 = 130\ 000$ sq.m.

For small dishes:

Area = $9 \times \pi \times 0.5 \times 0.5 = 7.1$ sq.m.

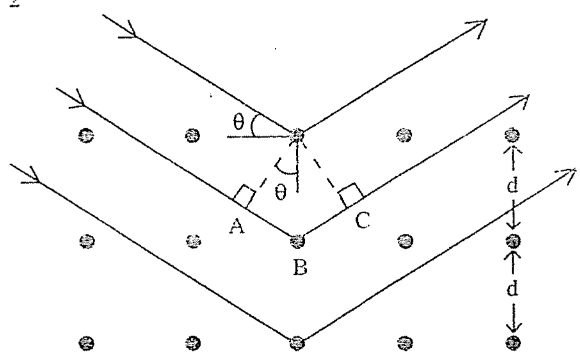
The ratio of energies collected will be the same as the ratio of areas = 130 000 : 7.1 = 18 000 : 1

Problem 8 Solution:

$n\lambda = 2d \sin\theta$

$n=1,\ \lambda=2.1 \times 10^{-10}$ m

$\theta = \frac{36}{2} = 18$ deg (see diagram)



$$d = \frac{\lambda}{(2\sin\theta)} = 3.4 \times 10^{-10}\ m.$$

Problem 9 Solution:

(a) $E = 200\ eV = 200 \times 1.6 \times 10^{-19}\ J = 3.2 \times 10^{-17}\ J$

(b) $$\lambda = \frac{6.6\times10^{-34}}{\sqrt{(2\times9.1\times10^{-31}\times1.6\times10^{-19}\times200)}} = 8.6\times10^{-11}\ m.$$

(c) This seems reasonable. The spacing of planes in crystals is typically 3×10^{-10} m.

Problem 10 Solution:

Rearranging the given equation:

$$V = \frac{h^2}{(2me\lambda^2)} = 16.6V$$

Typical exam question

We will conclude with a (fairly simple) typical exam question for you to attempt:

(a) Define diffraction (in words).

(b) What relationship between wavelength and aperture size leads to maximum diffraction?

(c) Wave motion of wavelength, λ, gives a diffraction pattern when it goes through an aperture of width, a. What happens to this pattern if we:
(i) Double the wavelength (only)
(ii) Double the aperture width (only)
(iii) Double both wavelength and aperture width?

(d) For multiple slit diffraction, the fringe spacing, y, is given by the equation:
$y = \lambda D / d$, where D is the distance from slits to screen, and d is the slit spacing.

(i) Microwaves of wavelength 3.0 cm travel through a double slit of spacing 6.5 cm, and the detector is placed 0.8 m beyond the slits. Find the distance between adjacent maxima in the interference pattern.

(ii) The microwave source and detector are replaced by an infrared source and detector. What would you notice about the diffraction pattern?

Answers

(a) Diffraction is the bending of a wavefront when it travels through an aperture or past an edge.

(b) Wavelength and aperture size should be approximately equal.

(c) (i) The pattern spreads out.
(ii) The pattern closes in.
(iii) The pattern is unaffected.

(d) (i) $y = \frac{\lambda D}{d} = \frac{0.03\times0.8}{0.065} = 0.37\ m$

(ii) The wavelength is much shorter for infrared. The fringe spacing would therefore be much smaller. You probably would not be able to see a diffraction pattern.

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman.
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU.

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 81

Diffraction and Light

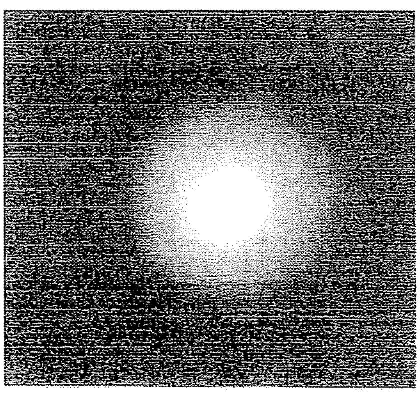
n Factsheet 78 we looked at diffraction effects across a wide range of wave notions – both electromagnetic and mechanical. This time we will focus on liffraction specifically with visible light.

Why? Because our eyes are more sensitive to this range of wavelengths han any other. In addition, diffraction causes limitations in the resolution of optical instruments, and can also be used in the study of atomic energy evels (through spectroscopy). Also, some diffraction effects can be seen by the naked eye, and demand explanation.

Atmospheric Diffraction

An example of a diffraction effect we see in nature is atmospheric diffraction. Light can bend around water droplets in thin clouds.

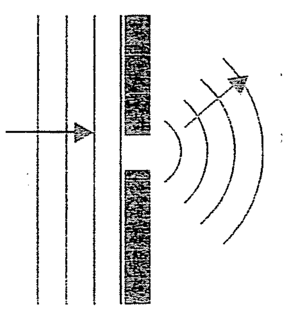
The black-and-white photograph shows a coloured diffraction ring which can appear when the Sun is low in the sky on a misty morning. Similar liffraction rings can be seen around the Moon.



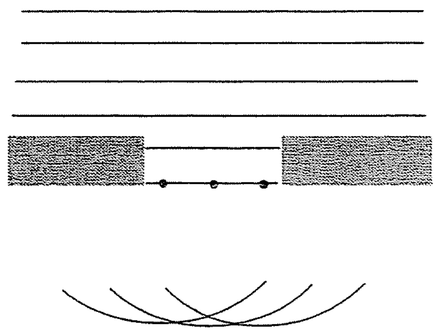
The "silver lining" we can see around dark clouds is also a diffraction effect.

Single Slit Diffraction

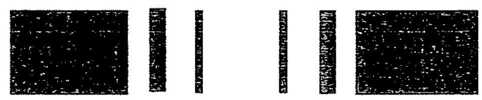
We know that light will spread out (diffract) on going through a gap, and hat the diffraction effect is greater for a longer wavelength or narrower gap.



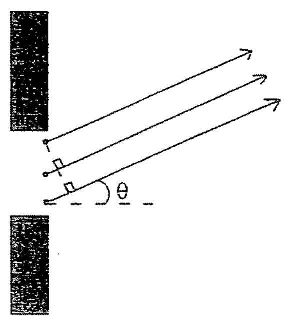
Huygen's Theory says that we can think of each point on a wavefront as a new source of waves. This gives us a nice visual idea of how diffraction occurs when we send light through a gap. The resultant wavefront curves at the ends.



This wavefront spreads out as it travels forwards. When it reaches a screen, we see a diffraction pattern displayed.



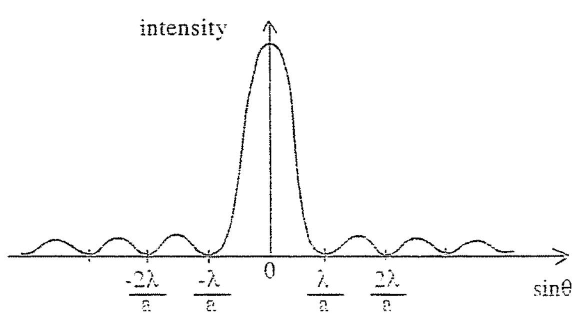
This is an interference effect, caused by light rays from various parts of the gap travelling different distances to reach a point on the screen. They may arrive in or out of phase, causing constructive or destructive interference.



The dark fringes (destructive interference) occur when:

$\sin\theta = n\lambda/a$, where n = 1,2,3.... and where a is the width of the gap

A graph of the intensity across the screen might look like this:

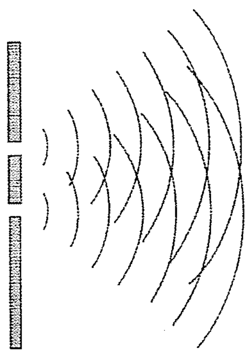


Problem 1:
(a) For light of wavelength 5.0×10^{-7}m incident on a gap of width 7.5×10^{-6}m, find the angles for the first two dark fringes.
(b) If the screen is 1.5m from the aperture, find the width of the central maximum.
(c) Find the width of the central (zeroth order) maximum if the width of the gap is doubled.

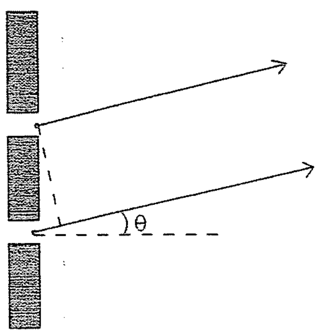
Notice that making the aperture wider causes the interference pattern to close in.

Two Slit Diffraction

If we send coherent wavefronts through two narrow gaps separated by a spacing of 'd' metres, an interference pattern is again formed. Waves spreading out (diffracting) from the two slits overlap to form a set of bright and dark equally spaced fringes on the screen.



Thinking of each slit as a point source, we can draw a ray diagram.

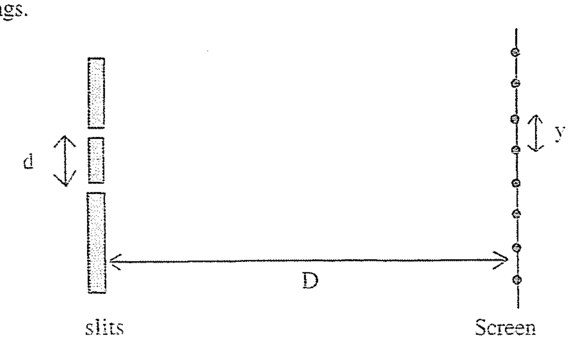


And the maths tells us that maxima will be seen when:

$\sin\theta = n\lambda / d$, where n = 0,1,2,3....

Notice that this expression gives the position of the maxima. Compare this to the single slit diffraction equation.

There is another useful expression for multi-slit diffraction and diffraction gratings.



If the screen is at a distance of D metres from the slits, and the fringe spacing observed on the screen is y metres, then we can show that:

$\sin\theta = \tan\theta$ (approximately) $= y/D$ (for small angles)

So we can write the fringe spacing, y, as:

$y = \lambda D / d$

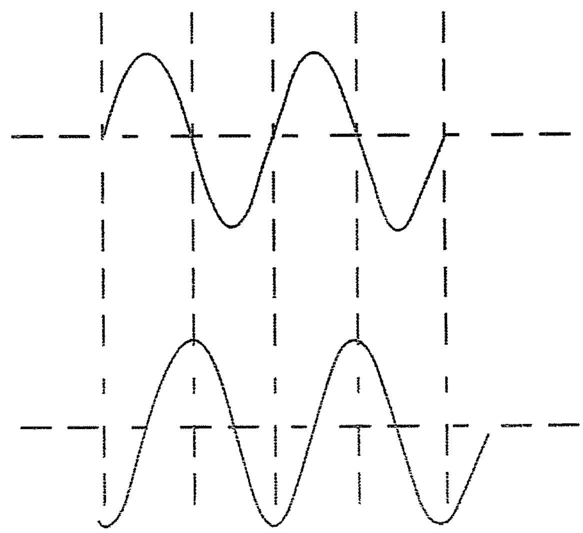
Problem 2:
With a plane wavefront of light of wavelength 6.5×10^{-7}m incident on a twin slit of spacing 5.2×10^{-5}m, find the fringe spacing on a screen at a distance of 2.5m from the slits.

Problem 3:
Consider the set-up in Problem 2 again. What would the effect be in each case of:
(a) doubling the wavelength?
(b) doubling the spacing between the slits?
(c) doubling both at the same time?

Exam Hint: *Be prepared for questions concerning the effect of changing wavelength, aperture size, slit spacing, monochromatic light to polychromatic light, etc.*

Phase and Coherence:

The **phase difference** between 2 waves travelling together depends on the fraction of a wavelength which one lags behind the other.



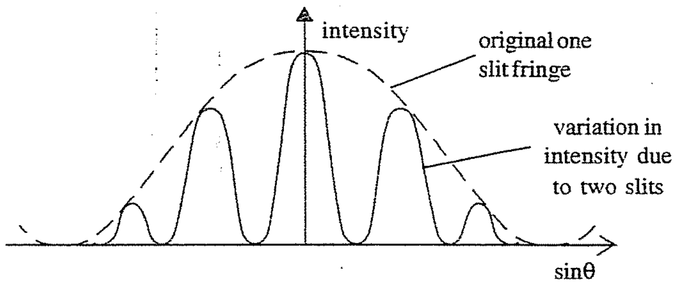
The phase difference here is one-quarter of a wavelength (or 90 degrees).

We say that two wave sources are **coherent** if there is a constant phase difference between them.
For our double slit diffraction experiment, as long as the waves reaching each slit are coherent, our interference theory still works. The only difference evident would be that the whole interference pattern would be slightly displaced one way or another along the screen (compared to the waves being completely in phase).

Multiple slit diffraction (or a diffraction grating) requires coherent wavefronts. The light must come from a single source.

Combination Effects

We looked at double slit theory on the assumption that we could treat each slit as a point source. However each slit has an aperture of width 'a'. The double slit pattern we have observed is actually superimposed on the relevant single slit pattern.

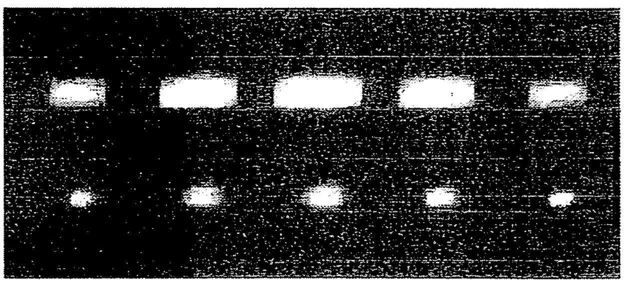


The double slit pattern will fade in and out as it fits within the envelope of the single slit pattern.

Problem 4:
For a double slit diffraction set-up, it is noticed that the fifth order maximum (n=5) disappears, as it is superimposed on the first order minimum of the single slit pattern. If the slit spacing, d, is 4.0×10^{-5}m, find the aperture width, ā, for each of the slits.

Multiple Slits

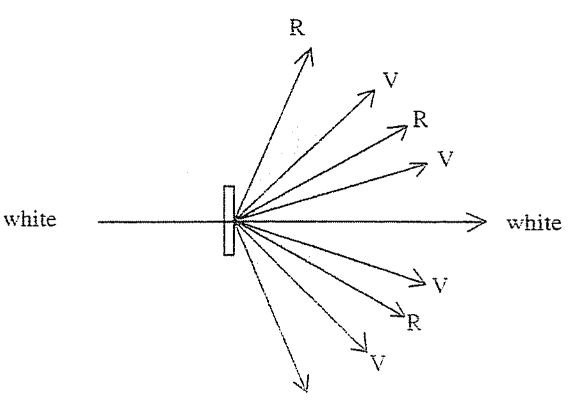
If we have more than two slits, but the values of slit aperture, a, and slit spacing, d, remain unchanged, then our mathematics remains the same. What we notice is that the bright fringes are **narrower** and more **intense**.



Monochromatic and Polychromatic Diffraction

So far we have assumed a single wavelength (monochromatic). If we use a white light source, we will get spectra, rather than bright bands, for our maxima.

As red light has the longest wavelength, by looking at our theory we can see that red will be on the outside of each spectrum formed, and violet on the inside. The central maximum will be white as no path difference is involved.



For larger angles, adjacent maxima for different colours will overlap each

Problem 5:
Suppose we take the wavelength of red and blue light to be 7×10^{-7}m and 4×10^{-7}m. For multiple slit diffraction, for what order, n, would the red maximum be outside the blue maximum for order, (n+1)?

Diffraction Gratings

A diffraction grating is a plate with many close, regularly spaced slits (or rulings). There may be thousands per millimetre. Nowadays gratings need no longer be physical scorings on a glass plate. They can be produced on photographic film and created from a holographic interference pattern.

The patterns from a grating are the same (in theory) as those from multiple slits. However the greatly increased number of slits lets much more light through. And increased interference sharpens the maxima observed.

A diffraction grating produces brighter and sharper maxima. Spectra produced from polychromatic sources are better defined using a grating. The theory is the same as for multiple slits, or even just two slits.

Typical exam question

(a) A diffraction grating has 850 lines per mm. Find the slit spacing for this grating.
(b) Find the angle for the first maximum for a wavelength of 5.0×10^{-7}m.
(c) The diffraction pattern obtained using a source of this wavelength is displayed on a screen. State any changes you would observe in the position or appearance of this pattern if the grating was blacked out except for two adjacent slits.
(d) Describe the pattern observed if a white light source was used.

Answers

(a) 850 lines/mm = 8.5×10^5 lines / m.

Slit spacing, $d = \dfrac{1}{8.5 \times 10^5} = 1.2 \times 10^{-6}$ m.

(b) $\sin\theta = \dfrac{n\lambda}{d} = \dfrac{1 \times 5.0 \times 10^{-7}}{1.2 \times 10^{-6}}$

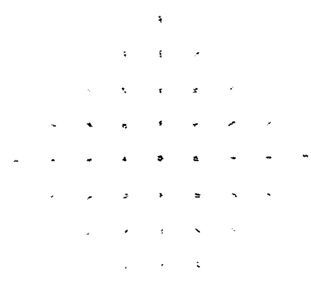
$\theta = 25$ degrees.

(c) The pattern would be less bright, as less light would reach the screen. The pattern would not be as well defined (sharp) with two slits as with a grating.

(d) You would see a central (zero order) white maximum. The other maxima would be spectra with the violet on the inside of each spectrum (closer to the straight-through direction.

Reflection gratings work just as well as **transmission** gratings. If you hold a CD up to the light you can see interference spectra resulting from the diffraction of white light incident on the grooves. This is often more interesting than the music on the CD.

With **crossed** gratings we can produce a two-dimensional interference pattern.

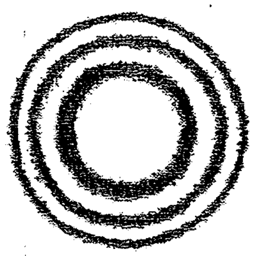


This pattern can also be obtained if we look at a bright point source

Resolution

Stars can be considered as point sources of light. But even the most powerful and optically perfect telescopes cannot distinguish two stars if the angle separating them is too small. This problem is caused by single slit diffraction at the aperture of the telescope.

A circular aperture produces a circular diffraction pattern with a central bright spot.



If two patterns are closely superimposed, it may be impossible to separate them. The two stars may appear as one blur. We have previously discussed this problem with radio astronomy.

Example:
Explain why astronomers prefer larger aperture telescopes, and may use blue or violet filters to observe the light from stars.

Answer:
Larger apertures and shorter wavelengths reduce the diffraction effects, increasing the resolving power. These changes may allow us to resolve the two stars. The equation for resolving two point sources says that their angular separation must be great enough so that:

$\sin\theta > 1.22\lambda/a$

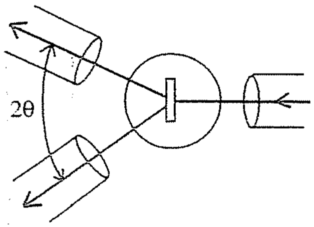
A smaller wavelength and a larger aperture lead to a smaller angle being resolved.

In astrophysics / cosmology you may be required to solve numerical problems on resolution.

> **Problem 6:**
> What angular separation would be required to just resolve 2 stars observed through a telescope of aperture 30 cm using a filter allowing through light of wavelength 4.5×10^{-7}m?

Spectrometer:

Factsheet 75 (Line Spectra) discusses how emission spectra are produced by viewing a discharge tube through a spectrometer. The diffraction grating in the spectrometer gives rise to a series of discrete coloured lines diffracted through different angles according to the wavelength of the light.



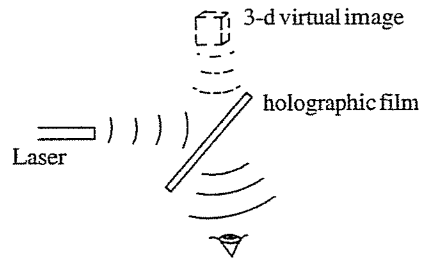
> **Problem 7:**
> For a certain line in the spectrum from a gas discharge lamp, the angular separation of the first order maxima (each side of the straight through direction) is 72 degrees. The diffraction grating has 1200 rulings per mm.
> (a) Find the wavelength of the light emitted.
> (b) What would the angle of diffraction be for the second order line for this wavelength of light?

Holography

Holography could easily be a topic on its own. We will just mention it for completeness in this study of optical diffraction.

A beam splitter allows a reference beam from a laser (coherent light at a single frequency) to interfere with a beam reflected from the object. Photographic film is used to store this interference pattern.

If the laser is then directed onto this interference pattern, diffraction occurs.

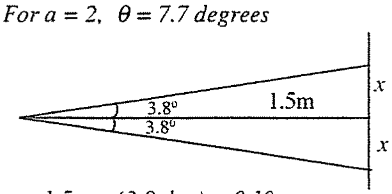


The viewer sees a virtual image of the object behind the film. Reflection holographs can also be used to recreate the image.

Solutions

Problem 1 solution:
(a) $\sin\theta = n\lambda/a$
For $a = 1$, $\theta = 3.8$ degrees
For $a = 2$, $\theta = 7.7$ degrees

(b)



$x = 1.5 \tan(3.8 \text{ deg}) = 0.10m$
The width of the central maximum is 0.20m.

(c) $n=1$
$\sin\theta = (5.0 \times 10^{-7}) / (1.5 \times 10^{-5}) = 1.9$ degrees.
The central maximum then works out to be 0.10m.

Problem 2 Solution: $y = \lambda D/d = 0.031m$

Problem 3 Solution:
(a) *This doubles the fringe spacing (spreads out the pattern).*
(b) *This halves the fringe spacing (closes in the pattern).*
(c) *The two changes should cancel each other.*

Problem 4 Solution: $\sin\theta = \lambda / a$ coincides with $\sin\theta = n\lambda / d$
$\lambda / a = n\lambda / d$
$a = d / n$, $a = 8.0 \times 10^{-6}m$.

Problem 5 Solution:
If the two maxima coincide with each other:
$\sin\theta = n\lambda(red)/d = (n+1)\lambda(blue)/d$
$n\lambda(red) = (n+1)\lambda(blue)$
$7 \times 10^{-7}n = 4 \times 10^{-7}(n+1)$
$n = 4/3 = 1.33$
So the red maximum would be outside the next blue maximum for $n=2$.

Problem 6 Solution: $\sin\theta = 1.22\lambda / a = 1.22 \times 4.5 \times 10^{-7} / 0.30$
$\theta = 1.0 \times 10^{-4}$ degrees.

Problem 7 Solution:
(a) $\sin\theta = n\lambda/d$ where $n = 1$ and $\theta = 36$ degrees.
$d = 1 / 1200 = 8.3 \times 10^{-4}mm = 8.3 \times 10^{-7}m$.
$\lambda = d \sin\theta/n = 4.9 \times 10^{-7}m$.

(b) $\sin\theta = (2 \times 4.9 \times 10^{-7})/(8.3 \times 10^{-7}) > 1$. *There is no second order spectrum obtainable. You would have to use a coarser grating (increase d) to make θ less than 90 degrees.*

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman.
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

Physics Factsheet

www.curriculum-press.co.uk Number 69

Experiments With Waves

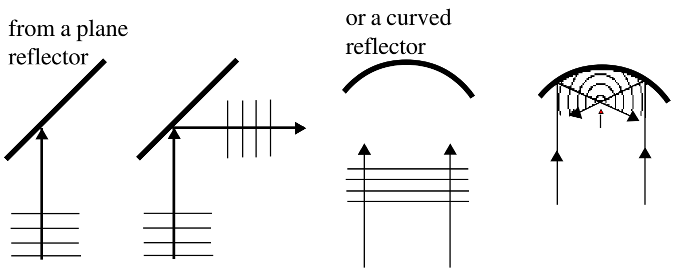
A **ripple tank** is a simple and common way of demonstrating properties of waves, including reflection, refraction, diffraction and interference. The tank is shallow with sloping sides (to cut down on reflection when waves hit the sides) and with a transparent bottom so that a light source can be mounted below the tank, projecting a magnified image of the water waves onto a screen (usually the ceiling) above.

The movement of the waves can be 'stopped' for certain observations and measurements using a stroboscope – either a simple hand wheel with regularly spaced slits or an electronic stroboscope in place of the projection lamp. The stroboscope doesn't actually stop the wave movement, of course – it gives a series of 'snapshot' views of the wave timed at such a frequency that each snapshot is exactly one wave period after the previous one. This means that the viewer sees each wave crest in exactly the same position as the preceding wave – making it appear to be standing still. (Running the stroboscope slightly slower or slightly faster can give the appearance of slow forward or reverse movement of the waves – a fact well-known to cinema-goers used to seeing wagon wheels appear to spin backwards as the stagecoach slows down and the wheel movement 'strobes' with the frame frequency of the cine camera.)

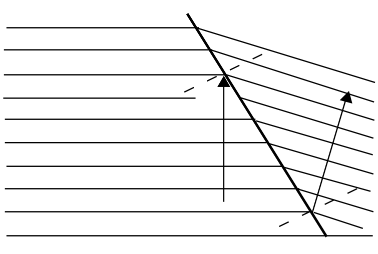
Waves are usually generated by a small electric motor mounted on a wooden bar. The bar hangs by two rubber bands from a stable cross frame above one end of the tank. The motor has a weight attached to the axle, but the weight is mounted eccentrically (i.e. off-centre) so that it wobbles as it spins – and so, therefore, does the wooden bar. The bar can be lowered so that it touches the surface of the water in the tank – causing parallel plane waves to be generated at the frequency of rotation of the motor. Alternatively, one or more small plastic dippers – in the shape of small balls – can be fixed into holes in the bar and the height adjusted so that the dipper(s) just touch the surface. This produces a point source of waves which spread out as a series of concentric circles.

A range of accessories may then be placed in the path of the waves to demonstrate the various properties of waves.

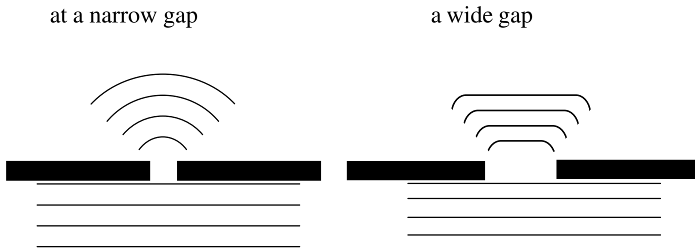
Reflection



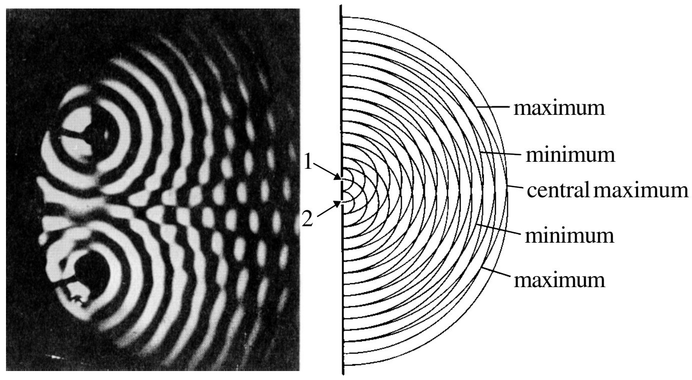
Refraction



Diffraction



Interference



The ripple tank demonstrates wave properties in water waves – which have wavelengths large enough to be seen clearly. However, once the effects of wave properties have been understood, those *effects* can be recognised, observed and investigated using other types of waves even when the wavelengths involved are either too large or too small for the individual *waves* to be seen.

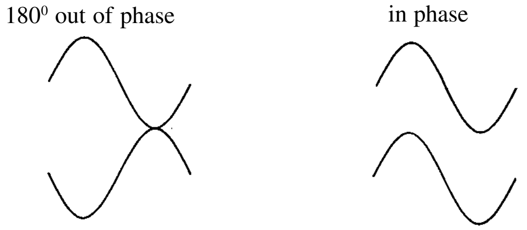
Wave properties in light waves - using a laser

A laser produces a convenient narrow, strong, sharply focussed beam of light – and so in reflection and refraction experiments it can replace the simple ray which uses a low voltage electric lamp. However, the laser really comes into its own when used in diffraction and interference experiments.

Safety note: *The typical popular (science fiction) understanding of a laser is that it will cut through a sheet of metal. Although such lasers exist, you will be pleased or disappointed to find that the one you will use does not produce anything like this sort of power. It is, however, powerful enough to do serious and permanent damage to the retina of the eye if you look directly into the beam. Don't try it – and make sure the experimental set-up doesn't allow anybody else to try it either, intentionally or otherwise! If the beam reflects off a highly reflective surface (like a mirror or glass window) it can be just as dangerous.*

For interference between two wave sources to be observed, those sources need to be **coherent**. Two waves are said to be coherent when they have the same wavelength and frequency and a *constant phase relationship* to one another.

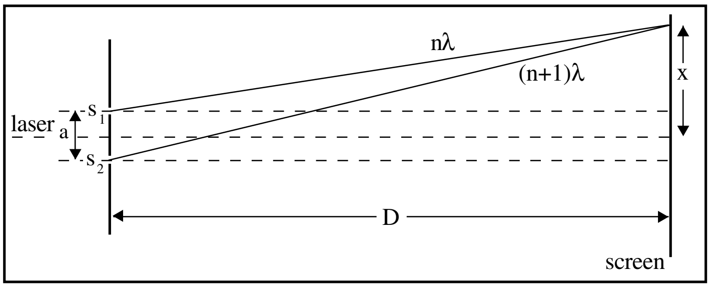
The phase relationship refers to the relative position of peaks and troughs of two waves. If the peaks of both waves occur at the same position, then the waves are said to be *in phase*. If the peaks of one wave coincide with the troughs of the other, then they are exactly *out of phase*, or *180° out of phase*.



Producing two coherent sources of water waves is not too difficult – using the ripple tank arrangement described above, the wooden bar simply has two plastic dippers placed in it, both touching the surface of the water. When the motor wobbles, so does the bar and so do the dippers – all together. Producing two coherent sound waves is not too difficult either – just connect two loudspeakers to the same amplifier output (i.e. mono, not stereo!). However, two coherent light sources are a different problem altogether. An ordinary light bulb of the type used in a laboratory ray box emits a wide range of light frequencies. Even passing the light through a colour filter still leaves a range of frequencies too wide for interference effects to be seen. The effect of interference at one frequency is alongside the same effect for a slightly different frequency – and each effect would be masked by all of the others.

To make a comparison with sound waves, the light bulb is like hitting all the notes on a piano at the same time – producing a noisy chord (or discord!) of lots of different frequencies all at once. To hear clear sound effects, it is necessary to pick out a single frequency – like a single note on the piano, or better still a tuning fork. The laser is rather like a tuning fork for light waves – it emits a very pure, narrow frequency of light. Even so, buying two similar lasers and putting them alongside one another would not guarantee coherent waves. A simple but clever trick is to shine one laser beam through two narrow, parallel slits, usually mounted in a projector slide (the arrangement is known as *Young's slits*, named after the inventor). The light waves diffract (spread out slightly) as they pass through the slits and become two beams and, since they have come from the same single source, they **must be** coherent.

The coherent waves are projected onto a screen several metres away from



the laser – which could be the opposite wall – and maxima and minima are observed as clear patches of bright light and darkness. These represent the positions where the waves are exactly in phase or exactly out of phase respectively – just as seen with the ripple tank interference experiment above. This equation describes the relationship between the various measurements which may now be taken, where:

The equation: $\lambda = \dfrac{ax}{D}$ *describes this relationship*

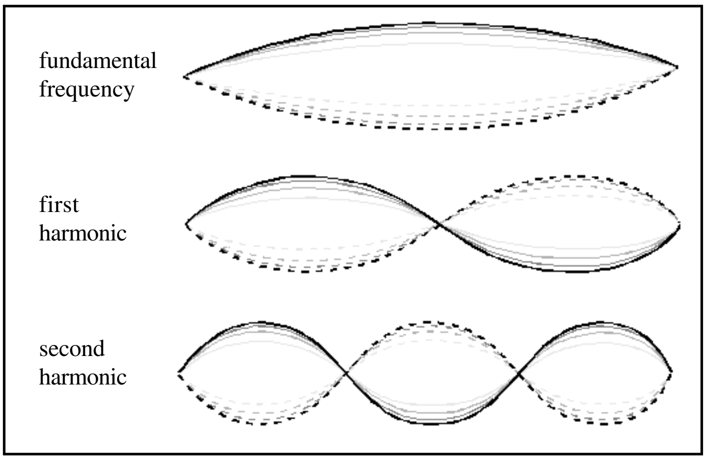
where: λ = wavelength
a = *separation of the slits (m)*
(distance between the centres of the two slits)
x = *distance between successive fringes*
D = *distance from the slits to the screen (m)*

Stationary waves experiments

Stationary waves offer a number of interesting experimental possibilities. They enable certain key measurements to be taken with a degree of accuracy difficult or impossible when dealing with travelling waves. A stationary wave can arise when a wave travelling in one direction along, or through, a particular medium, interferes with a second wave, similar in every respect to the first but travelling in the opposite direction. In practice, the 'second' wave is often in fact the first wave returning after reflection from an end or fixed point of the wave guide or wave medium.

The stationary wave has fixed nodes – positions of zero amplitude – and antinodes – positions of maximum amplitude. Often, we have a situation where only certain clearly defined frequencies can exist in a stationary wave.

Example – a vibrating string, fixed at both ends (like a guitar string)



A simple experimental way to investigate standing waves with two fixed ends is to set up a stretched cord between two fixed points several metres apart. This can be thin rubber cord, square or circular in cross-section and of about 5mm diameter fixed to two rigid supports (which might be clamp stands clamped tightly, to the benchtop); or it might be thin string, in which case at least one end needs to run over a pulley and then be loaded with a small mass – the mass may then be adjusted to change the tension in the string, one of the factors affecting the fundamental frequency of vibration.

Close to one of its ends, the string needs to pass through the top of a vibration generator which is driven by a signal generator producing a sine wave output of variable frequency.

The string will resonate (vibrate with large amplitude – as shown in the diagram showing 'fundamental frequency) when the driving frequency is equal to the fundamental frequency of the string set-up. In this condition, the stationary wave has only two nodes – the two ends of the string – and one central antinode. The wavelength is easily calculated, since the full length of the string is exactly equal to half of one wavelength. The frequency of the vibrations may be measured by feeding the output of the signal generator into an oscilloscope or a frequency meter; or by the use of a calibrated stroboscope. (The flash rate of the stroboscope is increased slowly until the vibrations appear to 'stand still'.)

The wave equation (speed = frequency × wavelength) may be used to calculate the speed of wave travel along the string. A number of investigations are possible with this set-up, by changing the tension in the string, or the mass per unit length of the string, or the length of the string.

Doubling the driving frequency will produce a stationary wave of the first harmonic as shown in the diagram above – the full length of the string now corresponds to one full wavelength. (This corresponds to playing a note one octave higher than the fundamental on a musical string instrument.) Three times the frequency gives the second harmonic – with three half-wavelengths fitting into the length of the string, and so on. It is relatively easy to get up to the tenth or even the fifteenth harmonic with a light string or elastic cord. Experiment with different coloured backgrounds and lighting to see the stationary waves most clearly – it is sometimes easier to see the wave pattern whilst looking along the string from one end.

Measuring the speed of sound

The speed of sound may be measured by setting up a small region of stationary sound waves inside the laboratory. Sound is produced by a signal generator driving a loudspeaker – at a fairly high audible frequency, say 3000Hz. A wooden board or convenient wall about 1.5m from the loudspeaker is used to reflect (echo) the sound waves back. A standing wave pattern now exists between the loudspeaker and the reflecting surface. A microphone is mounted in a clamp stand so that it can be moved and rested in positions between the loudspeaker and the reflector.

An oscilloscope is used to display the output from the microphone. If the microphone is slowly moved towards the loudspeaker it will be noticed that the amplitude of received sound goes through maxima and minima of amplitude. (It is worth pushing the clamp stand holding the microphone with a metre rule to cut down on reflection of sound from your arm and body.) The distance between successive maxima is one half-wavelength. (Moving the microphone through about ten successive maxima and dividing that distance to find one wavelength can reduce the error.) Frequency of the sound waves can be measured directly from the oscilloscope. The wave equation may now be used to calculate the speed of the sound waves.

Resonance in air columns

Stationary waves in air columns (such as organ pipes or other wind instruments) differ from those set up in stretched strings because different sets of harmonics are 'allowed' by the physical constraints of the medium. This is the main reason why a particular note played on a string instrument sounds different from the *same note* played on a wind instrument – the fundamental frequency is the same but the pattern of harmonics which overlay that note (at lower amplitudes) are different. This pattern of harmonics is known as the 'quality' or sometimes the 'timbre' of the note. A hollow cylinder containing a column of vibrating air is normally closed at one end and open at the other – so, whilst one end (the closed end) corresponds to a node of the stationary wave pattern, the open end corresponds to an antinode. This means that the lowest (fundamental) frequency allowed is when one *quarter* of a wavelength is contained within the air column. The first harmonic is when three-quarters of a wavelength is equal to the length of the air column, and subsequent harmonics correspond to five-quarter wavelengths, seven-quarter wavelengths and so on.

In a wind instrument, the air column is normally set vibrating by resonance from a small oscillator at one end – which may be a reed (as in some woodwind instruments) or even the musician's lips (brass instruments). This may be modelled by a tuning fork, or a small loudspeaker driven by a signal generator, placed at the open mouth of the air column.

One way to set up the air column with variable length is to have two lengths of glass tubing, one of which fits *inside* the other. The wide-bore tube is mounted vertically (supported by a clamp) and the bottom end plugged with a rubber bung or cork. The tube is then filled – not quite to the top – with water. The narrower tubing is then lowered inside the water-filled tube and the top end held with a clamp and clamp stand.

The air column to be used is then inside the narrow tube – the bottom surface being the surface of the water. The length of the air column can now be easily adjusted by simply adjusting the height of the narrow tube and re-clamping. A metre rule may be mounted outside the tubing for easy measurement of the air column length. The tuning fork or loudspeaker is placed at the top (open) mouth of the air column and the length of the column adjusted until the sound gets noticeably louder – this means that the air column is resonating. Alternatively, the air column may be left at the same length and the frequency of sound from the loudspeaker changed until resonance occurs. Once again, a variety of investigations are possible with the arrangement.

Microwave experiments

Most laboratory microwave transmitters and receivers work at a wavelength of around 3cm – which is a very convenient and measurable size for many experiments.

Receivers usually come in two forms – one with a directional collector (rather like a square funnel) and a probe type which is good for finding the strength of a signal at a point. Both types of receiver usually have an output which can be connected to a sensitive microammeter so that relative signal strength can be measured.

The transmitter and receiver arrangement can be used for many kinds of wave experiment – including investigation of reflection (microwaves will reflect well off a metal sheet), refraction, diffraction and interference. The convenient wavelength of the microwaves usually means that measurements of suitable accuracy may be made with simple apparatus such as a metre rule. Apparatus suitable for diffracting and refracting microwaves are usually available from the supplier of the 3cm wave transmitting and receiving equipment.

Practice Quetions

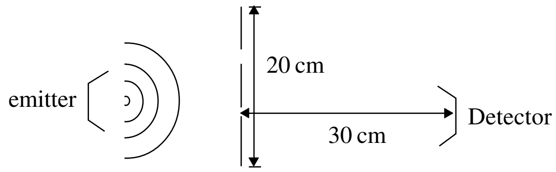
1. A pair of parallel slits are illuminated with light from a sodium vapour lamp of frequency 5.09×10^{14}Hz. A series of light and dark fringes is projected on to a screen 1m from the slits.
   (a) Explain why a bright fringe is always found at the centre of the pattern [3]

   (b) The distance between the central bright fringe and the adjacent bright fringe is 1mm. How far apart are the slits? ($c = 3 \times 10^8$ ms^{-1}) [4]

2. (a) 'Waves from two sources can only combine to form a stable interference pattern if they are coherent.' What does 'coherent' mean? [2]

   (b) A student sets up two loudspeakers to perform an experiment on the interference of sound. The loudspeakers produce sound waves with the same frequency, which is known to be below 1kHz. She finds a point of annulment, i.e. a point where the noise level is very low, at 3m from one speaker and 2m from the other. Determine all possible frequencies at which the speakers could have been oscillating. Take the speed of sound in air as *340ms^{-1}* [4]

3. Microwaves with a wavelength of 6cm are directed towards a metal plate. The plate is 20cm wide and has two parallel slits in it. An interference pattern is formed on the far side of the plate.



   (a) A microwave detector is placed 30cm directly in front of one of the slits. The detector gives a zero reading.
   (i) Comment on the phase difference between the waves from each slit at this point and hence explain why the detector gave a reading of zero [2]

   (ii) Calculate the distance between the slits [4]

   (b) What would happen to the interference pattern if the plate was rotated through an angle of 90^0? Explain your answer [3]

4. (a) Explain the difference in terms of energy, between progressive (travelling) and stationary (standing) waves [2]

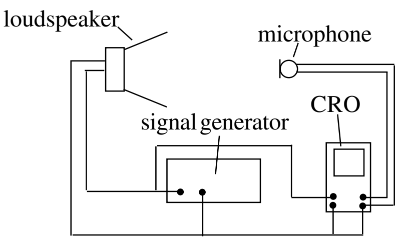
   (b) The velocity v of a transverse wave on a stretched string is given in the formula: $v^2 = \frac{T}{\mu}$
   where T is the tension in the string and μ is the mass per unit length. Use this equation to derive an expression for the fundamental frequency of a vibrating wire in terms of *T*, μ and its length, *L*[2]

   (c) A violin string is 35 cm long and has a mass of 2.25 g. It produces a note of frequency 256 Hz when sounding its first overtone. Find the tension in the string [4]

Exam Workshop

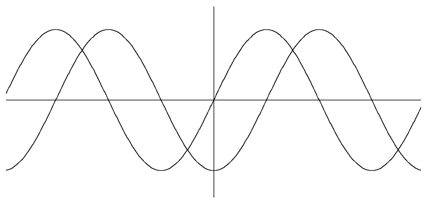
This is a typical weak student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiners' mark scheme is given below.

The equipment shown in the diagram is set up for an experiment.



(a) The CRO displays two traces: one directly from the signal fed to the loudspeaker, the other from the microphone.

(i) Draw a labelled diagram of a typical display you would expect to see on the CRO [3]



2/3

Correct! but the student hasn't *labelled* the diagram – a common mistake!

(ii) State two aspects of this display that would change as the distance between the loudspeaker and microphone is increased [2]

the phase difference between the waves increases
the amplitude of microphone signal decreases 2/2

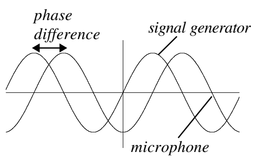
Correct - 2 marks

(b) Explain how the CRO measurements can be used to find the wavelength of the sound [5]

Gradually move microphone away for the speaker Each time traces are in phase the microphone has moved one wavelength. Count the number of wavelengths moved. Measure distance moved and calculate wavelength 4/5

Correct method, but the student did not state that the traces must be in phase before the microphone is moved

Examiner's answers



(a) (i) Two traces drawn, out of phase ✓✓
Labels shown ✓

(ii) Phase difference increases ✓
Amplitude of microphone signal decreases ✓

(b) Place microphone so traces are in phase ✓
Gradually move microphone away ✓
Each time traces are back in phase, microphone has moved one wavelength✓ Count the number of wavelengths moved ✓
Measure distance moved and calculate wavelength ✓

Acknowledgements:
This Physics Factsheet was researched and written by Keith Penn.
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

Answers

1. (a) At the centre of the pattern, waves from each slit have travelled equal distances. i.e. path difference is zero ✓
This implies that the phase difference is zero ✓
Thus constructive interference will always occur at the centre of the pattern, producing a bright fringe ✓ [3]

(b) $\lambda = \frac{c}{f} = \frac{3\times10^8}{5.09\times10^{14}} = 5.89\times10^{-7}$m ✓
Let y = distance between the bright fringes
d = distance between slits
D = distance between slits and screen
m = order of the fringe counting outwards from the central bright fringe

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

When $m = 1$, $d = \frac{\lambda D}{y}$ ✓

$d = \frac{5.89\times10^{-7}\times1}{1.0\times10^{-3}}$ ✓

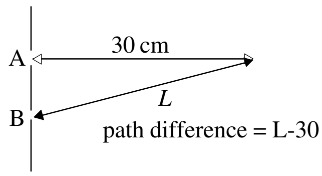
$d = 5.89\times10^{-4}$m ✓ [4]

2. (a) Waves from two sources are coherent if:
Their frequencies are equal ✓
The phase difference between them is constant ✓ [2]

(b) 'Annulment' or cancellation implies the path difference is an odd number of half wavelengths, i.e. $\lambda/2$, $3\lambda/2$, $5\lambda/2$, $7\lambda/2$, etc.; ✓
The path difference = 1m
Hence: λ = 2m, 2/3m, 2/5m, 2/7m, etc. ✓
Using $f = v/\lambda$, gives the following possible frequencies:
170Hz, 510Hz, 850Hz, 1190Hz, etc. ✓
As we know that the speakers were oscillating below 1kHz, the only possible frequencies for the sound are: 170Hz, 510Hz, 850Hz ✓ [4]

3. (a) (i) The waves from the two slits must be in anti-phase at this point✓ i.e. phase difference of π radians, giving destructive interference✓ [2]

(ii) The path difference from each of the slits must be an odd number of half wavelengths ✓



Smallest possible path difference = $\lambda/2$ = 3cm ✓
Then $AB^2 = 33^2 - 30^2$, AB = 13.75cm ✓
The next possible path difference, $3\lambda/2$ gives AB = 25cm which is greater than the width of the plate ✓ [4]

(b) The pattern would vanish. ✓ Microwaves are transverse and polarised. ✓ With the plate turned through 90°, the microwaves cannot pass through the slits ✓ [3]

4. (a) Progressive waves show net displacement of energy from one place to another✓ Stationary waves maintain energy within a boundary and hence no net displacement energy occurs ✓ [2]

(b) In fundamental mode $L = \frac{\lambda}{2}$

$\lambda = 2L$ ✓

$f = \frac{c}{\lambda} = \frac{1}{2}L\sqrt{\frac{T}{\mu}}$ ✓ [2]

(c) First overtone $L = \lambda$✓

$T = v^2u = f^2\lambda^2\mu$ ✓

$T = f^2L^2\frac{m}{L} = f^2Lm$ ✓

$T = 256^2\times0.35\times2.25\times10^{-3}$

$T = 52$N ✓ [4]

Physics Factsheet

www.curriculum-press.co.uk Number 122

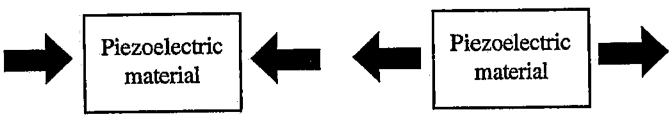
Uses of Ultrasound

Humans can hear sound across a wide range of frequencies, typically from 20 to 20,000 Hertz (waves per second). Younger people can usually hear even higher pitched sounds, possibly up to 22,000Hz. Ultrasound is simply sound that is of too a high frequency for humans to hear. Many animals can hear ultrasound and we can detect it electronically.

The only difference between sound and ultrasound is frequency. The definition of ultrasound is based on the hearing range of **humans.**

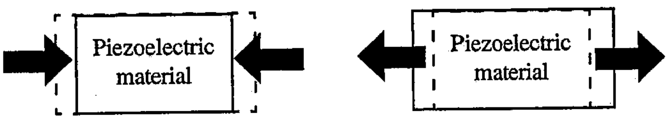
Ultrasound has a variety of medical and industrial uses and is commonly used in the fishing industry. Several different types of animals, including bats, whales, dolphins, porpoises and two species of birds, use it for navigation and feeding.We will look at the production and use of ultrasound by humans and animals.

How do humans produce ultrasound? We must use artificial means as we cannot produce such sounds naturally! We use a group of materials that have ***piezoelectric*** properties. A voltage (potential difference) can be generated across a material by squeezing or stretching it (applying mechanical stress).



1) squeezing the material produces a voltage

2) stretching the material produces the opposite voltage



3) applying a voltage in one direction, the material contracts

3) applying a voltage in the opposite direction, the material expands

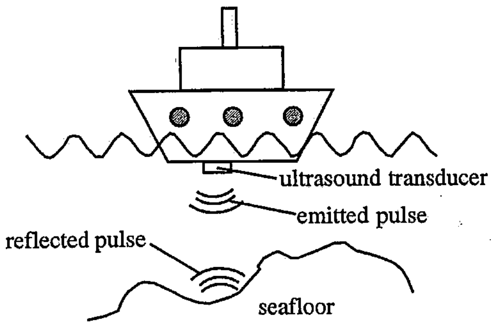
The reverse is also true; you can make the material change shape by applying a potential difference across it. The material will expand when a pd is applied in one direction and contract when the opposite pd is applied. Introduce a rapidly alternating pd and the material expands and contracts rapidly. Ideally, you need to cause vibrations at the resonant frequency for the material. This gives large amplitude vibrations which can pass through to any connecting material. To produce ultrasound, shape your piezoelectric material so the resonant frequency is well above 20,000Hz.

A material which can produce mechanical vibrations (and therefore sound) from an applied electrical voltage is known as a transducer: it can convert between different forms of energy. Quartz is one example of a piezoelectric material, often used in clocks and watches due to the precise frequency produces when a voltage is applied. More commonly for medical and industrial uses is lead zirconium nitrate (PZT).

Exam Hint: *learn the definition of a transducer: a device that can convert between forms of energy for a useful output, e.g. a measurement.*

Now we know how ultrasound is produced artificially, but how is this signal used? Most uses of ultrasound involve sending a signal which reflects from a distant object and this echo is detected. The time difference between the emitted pulse and the received echo tells us how far the sound has travelled, giving us a distance measurement.

Many ships use an echo sounder for sea-depth measurement. A sound is emitted, reflects at the sea bed and the echo is detected. The time difference, taken with the speed of sound in sea water, tells us the distance to the sea bed. Why use ultrasound? A ship, particularly the engine, is quite noisy. Echo-sounding uses a 50,000Hz frequency, far higher than ship noise which avoids any confusion. The other benefit is that these very high frequency waves are not diffracted much and they keep to a tight beam: preventing energy loss and allowing a clear echo. Fishing ships can also use ultrasound echo location to detect shoals of fish.



Example calculation

A ship emits an ultrasound pulse and receives the echo 0.43 seconds later. What is the depth of the sea at this point? Assume the speed of sound in water is 1500m/s.

Distance = Speed × Time

Distance = 1500m/s × 0.43s

Distance = 645m or 0.65km

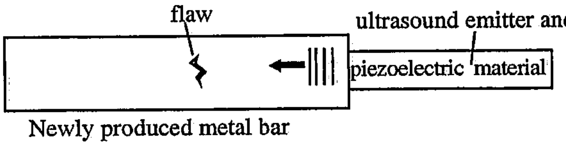
However, the ultrasound travels from the ship to the sea bed AND BACK. The sea is actually 645m / 2 = 322.5m deep!

Exam Hint: *Do not forget that ultrasound distance measurement relies on ECHOES. The ultrasound pulse travels TWICE as far the required distance measurement, so divide by two.*

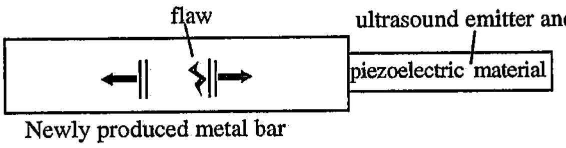
Defect detection

Ultrasound is also commonly used for detecting flaws inside materials without having to cut them apart: non-destructive evaluation. Once a metal bar has been cast, there could be an internal flaw, a gap where the material is weaker. Depending on the purpose, this could cause serious problems. An ultrasound transducer is placed against the material with a jelly-like material to allow the sound waves to pass easily into the bar. Only one signal should be detected: the reflection from the end of the bar. Another signal could indicate a flaw, reflecting some sound waves within the bar.

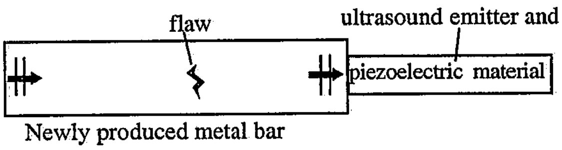
1) Pulses of high frequency sound waves are produced by the piezoelectric material and pass into metal bar



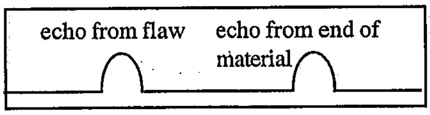
2) Some reflection occurs when the ultrasound passes through the flaw



3) An echo is detected by the ultrasound transducer. The remaining ultrasound pulse has now reflected from the end of the metal bar



4) The ultrasound transducer detects two signals, one from the flaw and one from the end of the material



Example calculation:
Non-destructive testing is carried out on a steel bar. The speed of sound in steel is approximately 5100m/s. Two ultrasound echoes are detected, after 0.0021 and 0.0034 seconds. How long is the bar and where is the flaw?

It must take longer for the ultrasound to reach the end of the bar, so 0.0034s is the time taken to travel the length of the whole bar AND BACK.

Distance = speed × time
Distance = 5100m/s × 0.0034s
Distance = 17.34m

The bar must be 17.34m / 2 = 8.67m long

The time for the ultrasound to reach the flaw and travel BACK is 0.0021s.
Distance = speed × time = 5100m/s × 0.0021s = 10.71m
The distance to the flaw must be 10.71m / 2 = 5.36m from the sensor.

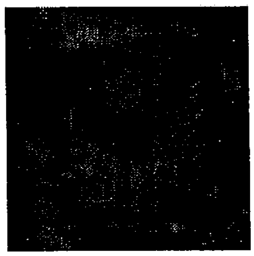
Medical uses

1. Physiotherapy

High intensity pulses of ultrasound can be used for physiotherapy. The heat generated in tissue can help alleviate muscle aches. For medical imaging, much lower intensity ultrasound radiation is used. There are several advantages for using ultrasound in medical diagnoses. During pregnancy, the foetus is very sensitive to ionising radiation like x-rays and the risk of harm to the unborn baby is high. This is basically because the foetus is undergoing a rapid cell division and growth which could be disturbed by x-rays, causing foetal abnormalities. Another major advantage of ultrasound is that ANY density change causes some reflection. So quite subtle differences between tissue types can be detected via ultrasound whereas an x-ray would be insensitive to the differences. This means that ultrasound has a wide range of uses for pregnancy alone: finding the size of the foetus and therefore time of pregnancy, detecting a foetal heartbeat, the sex of the baby, physical abnormalities and more.

2. Foetal scan

When ultrasound is used to examine a foetus, the ultrasound transducer is usually placed against the woman's abdomen. The method is essentially similar to the distance ranging methods discussed so far but in practice is far more complex. The transducer emits ultrasound pulses. These pass through to the abdomen of the pregnant woman. A jelly-like liquid is used to ensure most of the wave passes into the patient. Without this, most of the wave would reflect from the skin due to the large density difference between tissue and air.

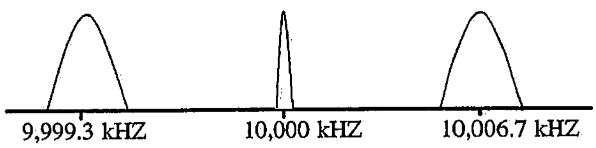


Ultrasound techniques have improved: 3D imaging

Reflections occur at each tissue density change throughout the body and foetus so getting a good angle is essential. The reflected pulses are received by the piezoelectric ultrasound transducer. This signal is amplified and passed to a computer where significant processing is required to present an image on the screen for diagnosis

3. Blood flow

Doppler ultrasound is also an essential medical tool. The Doppler effect is noticeable when a fire engine drives past. As the vehicle drives towards you, the siren sounds high pitched. As it passes you and moves away, it sounds lower pitched. In the same way, ultrasound is affected by the fluid flow in, for example, blood vessels. If the blood is moving away from the transducer, the reflected signal is lower frequency than it should have been. If the blood is moving towards the transducer, the reflected signal is higher frequency than it would otherwise have been. The shift in frequency tells us how fast the blood is flowing and is useful for any cardiovascular diagnosis. No Doppler shift indicates little or no blood flow, which would require immediate treatment.

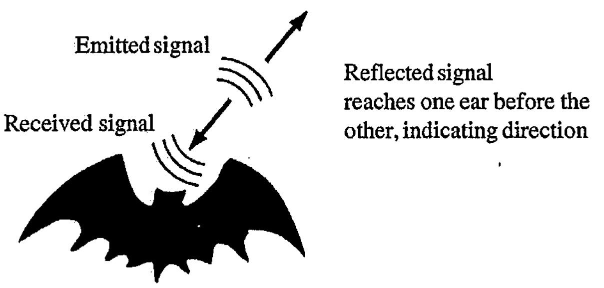


Received ultrasound signal: blood flowing away from transducer | Transmitted Ultrasound signal | received ultrasound signal: blood flowing towards transducer

4. Other medical uses
Medical uses of ultrasound also include echo-cardiography, where the ultrasound is used to study the beating heart, essential for determining the correct function of each chamber, blood vessel or artery.

> *Exam Hint: Be prepared to discuss the advantages and disadvantages of different medical techniques, e.g. ultrasound and x-rays for diagnosing foetal abnormalities or heart disease.*

Natural ultrasound
A variety of animals use ultrasound for finding their way and catching prey. Ultrasound is used as an alternative to sight in low light conditions: either dark or murky. Some species of bat emit ultrasound in the larynx, just as humans ourselves produce audible sounds. These sounds are emitted, usually through the mouth. Different species of bat produce different frequency sounds, ranging from 14,000 to 100,000 Hz. Therefore, a small fraction is audible to humans as a high-pitched click. Bats usually emit about 10-20 clicks per second, using ultrasound pulses to identify objects as well as their position and distance. Bats have TWO detectors: their ears. There is a time difference between for the reflected pulse to arrive at each ear which helps identify location as well as distance.



Whales, dolphins and porpoises have probably the most sophisticated form of ultrasound echolocation. These are all mammals and have evolved from land based mammals around 50 million years ago. The shape of dolphins and porpoises have evolved to make these creatures very effective at emitting and receiving ultrasound. Dolphins produce all sound in a region just behind the bulge at the front of their head, the equivalent of our larynx and passes into the ***melon***. The melon is the fatty bulge of the forehead of the dolphin and allows dolphins to focus and direct their clicks of ultrasound. They receive the echo vibrations from objects through their long jaw bones and these vibrations are passed backwards to the inner ear of the dolphin.

One area of interest is the ability of dolphins to detect and treat human illness or disabilities. There are some anecdotes about dolphins apparently locating tumours in swimmers, but whilst their highly efficient ultrasound could possibly function in this way, there is little evidence to support it.

Several organisations encourage swimming with dolphins for a range of illnesses and there is some evidence of depression being alleviated by swimming with dolphins. It is not yet known whether this is because of some special ability of dolphins or the well recognised benefit of being around any animal generally. It is also believed that dolphin and porpoises may use large amplitude pulses of ultrasound to physically stun prey, although this is still inconclusive.

Practice Questions
1. Give your definition of ultrasound.

2. Define a transducer.

3. Describe, in detail, how ultrasound is artificially produced using piezoelectric materials.

4. State two industrial uses of ultrasound and explain the relevant advantages.

5. Describe two medical uses of ultrasound.

6. What is the advantage of bats having two ultrasound receivers (ears)?

7 Sketch a diagram and explain how dolphins produce and receive ultrasound for echolocation.

8. A fishing ship receives ultrasound echoes 0.24s and 0.38s after transmission. How deep is the shoal of fish and the sea-bed?

9. An ultrasound transducer receives an echo after 8×10^{-5}s and 1.6×10^{-4}s during a pregnancy scan. Approximating the speed of sound in all tissue as 2500m/s and assuming that the signal is reflecting only from the head of the foetus, how large is its head?

10. Calculate the wavelength of the following ultrasound signals in air: 22,000Hz, 50,000Hz, 100kHz, 10MHz. The speed of sound in air is around 340m/s.

11. Calculate the wavelength of a 20MHz ultrasound signal in air and water. The speed of sound in water is 1500m/s.

Numerical answers
8. 180m and 285m

9. 10cm

10. 1.5×10^{-2}m, 6.8×10^{-3}m, 3.4×10^{-3}m, 3.4×10^{-5}m

11. 1.7×10^{-5}m, 7.5×10^{-5}m

Acknowledgements:
This Physics Factsheet was researched and written by Jeremy Carter
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 159

Non Ionising Medical Imaging: MRI, PET, Ultrasound and Thermography

Ultrasound

> *Ultrasound is exactly like ordinary sound except that it has a very high frequency.*

Examining boards often ask how ultrasound is different from or similar to X-rays, so remember:

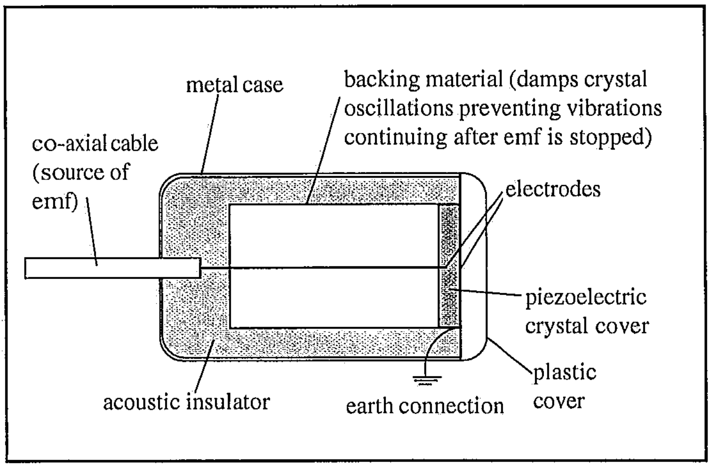
Comparison	Ultrasound (similar to sound)	X-rays (electromagnetic radiation)
Difference	Longitudinal wave	Transverse wave
Difference	Non-ionising (safe)	Ionising
Similar	Superposition of waves occurs	Superposition of waves occurs
Wave Equation (useful for problems)	Speed = wavelength x frequency	Speed = wavelength x frequency

Ultrasound is generated using piezoelectric crystals, which act as transducers. The crystal is deformed when an alternating potential difference is applied across it, causing the crystal to vibrate at the same frequency as the applied potential difference. This vibration can then be used to generate ultrasound.

When the crystal receives ultrasound it will develop an alternating potential difference across it.

> *A transducer converts one form of energy into another.*

Cross-section through a single transducer ultrasound generator probe:



There are two types of ultrasound scans:

A-scans (or amplitude scans)

Short pulses of ultrasound are sent into the body and a detector scans for reflected pulses. A time base is used to find the time the echo (or reflected pulse) takes to return. This enables the distance between structures in the body to be calculated. This type of scan is most often used to find the precise depth of a structure. (e.g. detecting tumours, which would give an abnormal depth scan).

Only **one transducer** is needed but the frequency must be arranged such that each reflected pulse arrives back at the probe before the next pulse is emitted.

B-scans (or brightness scans)

The reflected pulse is displayed by the brightness of a spot. By using **an array** of transducers a 2-D image can be built up. (e.g. monitoring the growth of a prenatal foetus and helping to detect any abnormalities)

Acoustic Impedance This determines how much energy is reflected at a boundary between two materials. If there is little change in acoustic impedance there will be little energy (i.e. ultrasound signal) reflected, but if there is a large change in acoustic impedance there will be a lot of reflection.

> *Air has a very different acoustic impedance to skin (or other soft tissue). To prevent most of the ultrasound being reflected back from the skin, a gel or oil with a similar acoustic impedance to skin is first applied the patient, so that there is no longer a thin layer of air between the ultrasound probe and the skin.*

Acoustic impedance, symbol Z, is defined by the equation:

$Z = \rho v$ where ρ is the density of the medium and v is the speed of sound in the medium.

The fraction of reflected intensity may be calculated by using the equation

$$I_r \div I_o = (Z_2 - Z_1)^2 \div (Z_2 + Z_1)^2$$

Where: I_r is the intensity of the reflected ultrasound signal,
I_o is the intensity of the incident ultrasound signal,
Z_1 is the acoustic impedance of tissue 1
Z_2 is the acoustic impedance of tissue 2.

The Doppler Ultrasound Probe

> *There is a shift in frequency when a wave is received by a moving object, and a similar frequency shift when a moving object sends out a wave. This is called the Doppler Effect.*

A **continuous** ultrasound signal is sent out and its echo is received back from the moving object of interest in the body. Two transducers are needed, one to produce and emit the ultrasound and one to receive it. For example blood cells are moving and so when they receive the ultrasound its frequency is Doppler shifted, and when they transmit the ultrasound back it is again Doppler shifted giving a doubling of the Doppler shift.

The shift in wavelength $\Delta\lambda$ is analysed and compared to the initial wavelength emitted by the probe, and the velocity, v calculated using the equation:

$$\Delta\lambda \div \lambda = 2 f \div c$$

where c is the speed of the ultrasound wave.

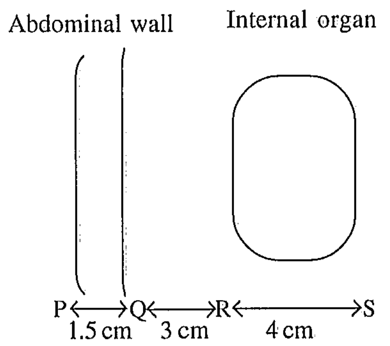
This can be used to detect clots or thrombosis, as these cause changes in the blood flow through a vein or an artery.

> ***Extract from examiners' report:*** *"There were 10ⁿerrors in converting from microseconds."*

Units

These can be tricky, so try the following **worked example**:

A schematic diagram of a cross section through part of the abdomen of a patient is shown below:



(a) If the speed of ultrasound in the abdominal wall is 1500 ms^{-1} find the time interval between a pulse being sent out and its echo being received from the inside surface at Q. (3 marks)

Solution: *Use time = distance ÷ speed*
Distance = 1.5 × 2 = 3 cm,
then time = 3 ÷ 100 ÷ 1500 = 2 × 10^{-5} s or 20 μs

> ***Extract from Examiners' Report:*** *"Many candidates failed to include the ×2 factor".*

(b) The time between the pulses emitted from the transducer is 50 μs. Find the frequency of emission of the pulses. (2 marks)

Solution: *Frequency = inverse of the time period*
$f = 1 \div (50 \times 10^{-6}) = 20000$ Hz

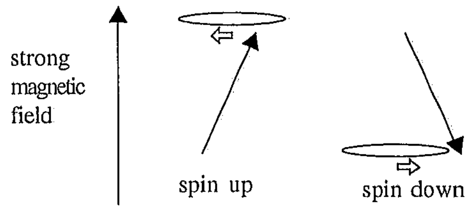
(c) The time taken for the ultrasound echo to return from surface at S is 60 μs.
(i) What problems can this cause (2marks)
(ii) Give the solution to the problems you described in part (i) of this question.

Solution (i): *Time taken for the pulse to return is longer than the time between the pulses, and the reflections need to reach the transducer before the next pulse is transmitted, otherwise the image will not be accurate.*

Solution (ii): *Decrease the frequency of the ultrasound.*

Magnetic Resonance Imaging or MRI

Hydrogen nuclei consist of just one proton each. They are also spinning about an axis and, because a moving charge acts like an electric current, they each produce a tiny magnetic field. Normally, their magnetic fields will cancel each other out, as they are randomly arranged. However, if a strong magnetic field is applied, they will align themselves with their tiny magnetic fields either parallel or antiparallel to the field direction. Then all **the nuclear spins will precess around the field lines** (as shown below), **producing a net magnetic field of their own.** They all do this at the same characteristic frequency, which is called the Larmor frequency.



> *The Larmor frequency is a resonant frequency of the protons and it is directly proportional to the strength of the strong applied magnetic field.*

Radio waves of the Larmor frequency are applied and some nuclei are excited into the higher energy spin state. The radio waves are switched off and the nuclei revert to their original spin state giving off electromagnetic radiation of the Larmor frequency . This electromagnetic radiation dies away with time and its duration equals the time taken for all the nuclei to switch back, and is measured by the scanner. It is called the **relaxation time.** It is characteristic of the type of tissue that the nuclei are in and the strength of the strong applied magnetic field.

The radiofrequency signals detected and relaxation times are processed by a computer together with the other various properties of the MRI signal to build up a detailed image of the human body.

> ***Extracts from Examiners' Report:***
> *"Many candidates had no knowledge of the regions of the electromagnetic spectrum used in CT and MRI scans" " Candidates had little idea of the role of the e.m. radiation in MRI imaging"*

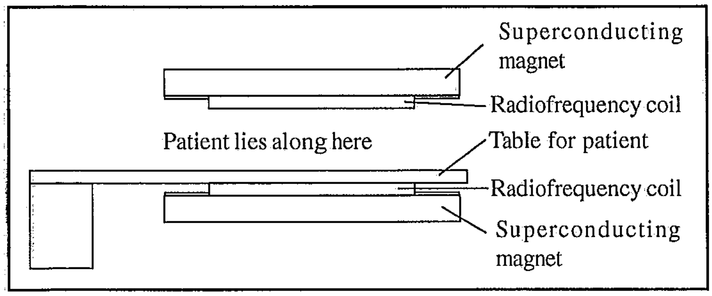
Good images of cross section are obtained for soft tissue, for example the brain. However the images of harder tissue such as bone have poor resolution.

> *Certain metals have a temperature below which they are superconductors. Their resistivity is zero below this transition temperature.*

The magnet used is an extremely powerful electromagnet. Coils of wire are arranged around a cylinder, and carry a huge electric current. This current would cause a large heating effect, but the coils are cooled with liquid Helium to a temperature at which the metal carrying the current becomes 'superconducting'. The current through the coils flows indefinitely and there is no heating of the coil. Superconducting electromagnets produce an exceptionally stable and large magnetic field, and so are suitable for use in MRI scanners. The magnetic field produced must be as uniform as possible over the area of the patient's body which is being scanned, if good resolution is to be achieved.

The radiofrequency coils are situated within the magnetic field and around the area of the patient's body which is to be scanned.

Cross Section of Arrangement for MRI Scan



Other arrangements are possible, for example the use of gradient radiofrequency coils for small areas of the human body, such as the knee, where the coils are around the area to be scanned. The patient may then sit or lie within the superconducting magnet.

Patients with metal implants are not suited to MRI scanning as the implant will become dislodged, move through tissue and harm the patient.

Great care must be taken to remove or secure any metal objects situated in the scanning room, which will be attracted with great force by the magnet and can be accelerated by the magnetic field sufficiently to severely harm apparatus and patient.

Metal objects, such as rings, worn by the patient must be removed and stored outside the scanning room, as they may move through the patient's tissue and will certainly cause the magnetic field of the superconducting magnet to become inhomogeneous, reducing the resolution obtained when scanning.

Resolution may be defined as the smallest discernable distance.

(PET) Positron EmissionTomography Scans

Positrons are positive electrons.

Positrons are emitted by certain isotopes, for example oxygen-15, carbon-11 and nitrogen-13. These isotopes are used to label compounds which are known to accumulate in specific areas of the human body. [For example glycogen can be labelled with oxygen-15 and glycogen is known to accumulate in the most active parts of the brain.] The labelled compound is then injected into the patient's body.

When a positron collides with an electron they annihilate each other, emitting a pair of gamma rays which move off in opposite directions so conserving momentum.

The positrons emitted collide with electrons, giving off pairs of gamma rays. The gamma rays are detected by gamma detectors (scintillation counters give a flash of light which is detected and amplified by photomultiplier tubes) and the output from these is fed into a computer. The computer uses the data to produce real-time images on a screen.

The detectors are often arranged in a circular array, with the patient at the centre and the length of his body perpendicular to the circle, so that he can be positioned to produce images of cross section of the organ of interest. These are used by the computer to build up a picture of the organ and the level of activity of each area of it. The more active areas will absorb more of the labelled positron emitting compound and so give a stronger signal.

Examination questions often ask which technique might be used for a certain diagnostic situation and why, so it is a good idea to compare them:

Technique	Advantages	Disadvantages
X-rays	X-rays produce good images of dense tissue (e.g. bone). They can produce images of organs filled with air (e.g. lungs).	High dose of ionising radiation received by the patient. Ionising radiation received by people working with X-rays must be monitored and limited.
Ultrasound	Safe (no known side effects). Good images of soft tissue. Can produce moving images. Portable machines can be taken to the patient.	Ultrasound does not pass through air (e.g. cannot study lungs). Ultrasound cannot penetrate bone so organs covered by bone (e.g. brain) cannot be studied. Low resolution.
MRI	Safe (no known side effects). Very good images of soft tissue. Good (and safe) images of brain and other delicate areas.	May cause claustrophobia in patients. Poor images of hard tissue (e.g. bone) Expensive. Needs a special room.
PET	Can study the uptake of a substance by an organ (i.e. organ function).	Injection of radioactive substance required.. A cyclotron is needed to produce positron emitting isotopes and the isotopes have very short half lives, so this is very expensive.

Thermography

This uses thermal radiation emitted by the patient over the area of interest. The radiation is detected using infrared detectors, and the output fed into a computer. For example both halves of the patient may be compared to detect differences in surface blood flow. A higher level of blood flow gives off more thermal radiation. However, time must be allowed for the surface temperature of the patient to stabilise in the room in which the scan is to take place.

Practice Questions

1. (a) State two physical properties of human tissue which determine its acoustic impedance. (2 marks)
   (b) What are the conditions under which ultrasound is strongly reflected at a boundary between two different types of tissue. (1 mark)
   (c) Why is a gel used between the patient and the transducer when an ultrasound scan is taken. (3 marks)

2. (a) Give a difference between the probe used in an A-scan and that used in a B-scan. (1 mark)
   (b) How is a piezoelectric crystal used to generate waves of ultrasound? (4 marks)
   (c) In order to make the crystal pulses short, short voltage pulses are applied. However the assembly must have an additional modification if the crystals are to emit short pulses. Explain the modification used and why it is necessary. (2 marks)

3. (a) Thermography can be used to detect circulation problems in patients. What part of the electromagnetic spectrum is used in thermography? (1 mark)
   (b) Why are the detectors used in thermography often cooled in liquid nitrogen? (3 marks)
   (c) Give an example of a non-medical use of thermography. (1 mark)

4. (a) Give a hazard to a patient which could be associated with MRI and describe an associated safety precaution.(2 marks)
   (b) State two advantages of X-ray imaging compared to MRI. (2 marks)
   (c) Name the region of the electromagnetic spectrum used in MRI scans. (1 mark)
   (d) The commonest elements present in the human body are oxygen-16 (atomic number 8), carbon-12 (atomic number 6), and hydrogen-1 (atomic number 1). Explain why it is justifiable to say that the nuclei imaged in an MRI scan are those of hydrogen? (3 marks)

5. (a) In a PET scan an isotope emits a positron which travels a distance ~1mm before losing its kinetic energy and being annihilated by an electron. Two gamma photons of the same energy are emitted in opposite directions. Give the reason why the ã photons must be emitted in opposite directions. (1 mark)
   (b) Calculate the energy of each photon in question 14. Give your answer in MeV. (3 marks)

Answers

1. (a) The speed of sound in the tissue (or elasticity of the tissue) and the density of the tissue.
   (b) When there is a large difference in acoustic impedance between the two tissue types.
   (c) The gel excludes air. Air has a very different acoustic impedance to soft tissue and so most of the ultrasound would be reflected. Gel has a similar acoustic impedance to soft tissue and so transmission of ultrasound is maximised.

2. (a) A B-scan uses a multi-transducer whereas an A-scan uses a single transducer probe.
   (b) Two surfaces of a thin slice of crystal are used as electrodes. These electrodes are connected to a very high frequency (MHz) e.m.f. source. The applied e.m.f. causes a rapidly alternating electric field across the crystal between the electrodes. The crystal expands and contracts at the same frequency as the e.m.f. which is applied. The vibrations of the faces of the crystal produce ultrasound pressure waves.
   (c) A sheet of vibration-absorbing backing material is applied behind the crystal to stop the vibrations continuing after each electrical signal is stopped.

3. (a) Infra-red.
   (b) Cooling the detectors reduces unwanted signals. This is important in thermography as the current from the detectors is low. If the surroundings are cooled they produce less noise.
   (c) To detect survivors in collapsed buildings/ to image people or animals at night/ burglar alarms/ weapons systems/ etc.

4. (a)

Hazard	Safety Precaution
Noise and ear damage	Wear ear plugs
Magnetism and metal	Remove earrings, watches etc.
Confined space and claustrophobia	Counselling or sedation.

   (b) Any two from X-rays are cheaper, quicker, relatively portable, can be used if patients have metal implants, better imagesof bones.
   (c) Radio waves.
   (d) The nucleons in carbon and oxygen are paired and so the tiny magnetic moments due to spin cancel out. Thus resonant signals are only emitted from nuclei of odd nucleon number. Hydrogen has a single proton.

5 (a) Momentum must be conserved.
   (b) Mass of positron = mass of electron. Thus 2 × mass of electron is converted into energy of 2 photons.
   $E=mc^2$ gives energy of each γ-photon = $9.109\times10^{-31} \times 3\times10^{8} \times 3\times10^{8} = 8.1981\times10^{-14}$ J $= 5.124\times10^{5}$ eV = 0.51 MeV
   (rest mass of electron = 9.109×10^{-31} kg and 1eV = 1.60×10^{-19} J)

Acknowledgements:
This Physics Factsheet was researched and written by Christine Collier
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 75

Line Spectra

The purpose of this Factsheet is to explain the appearance, origin and significance of Line Spectra. Before studying the Factsheet, you should make sure that you are familiar with the idea of a spectrum from your GCSE course, and with the ideas of quantization of energy levels in an atom.

These ideas of Line Spectra are an important introduction to the understanding of the ideas of Factsheet 51 – The Electromagnetic Doppler Effect and the Expanding Universe.

Questions on line spectra are likely to occur on the PHY 4 Paper and also on PHY 6, the Synthesis Paper.

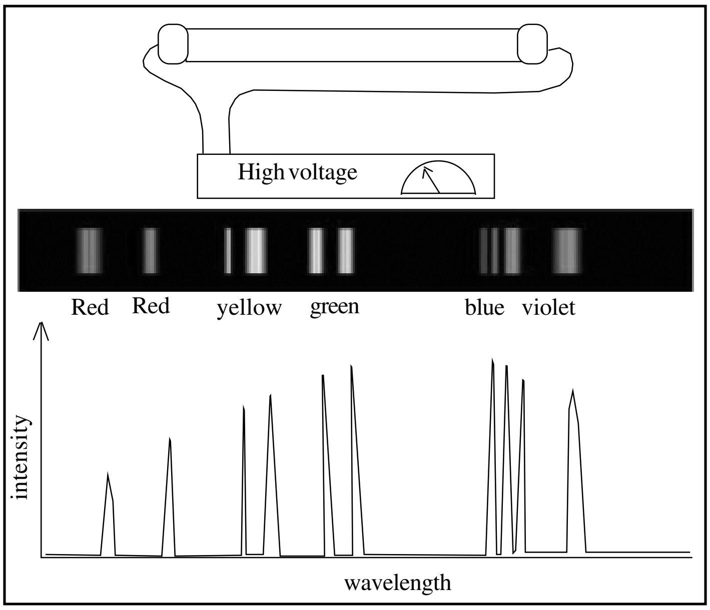
Appearance of Line Spectra

You will be familiar with the idea of a continuous spectrum, as produced by a prism. As the name suggests, line spectra consist not of a continuous band of different colours, but of discrete, separate lines. Each line is of a given frequency (wavelength), representative of a particular element.

> *A line spectrum consists of discrete lines, each typical of the element concered.*

Observing Line Spectra in the laboratory

Emission line spectra may be observed in the laboratory by viewing a discharge tube with a diffraction grating. The tube contains a particular element, e.g., sodium, which, when operated at appropriate voltage, vaporises. The hot vapour emits light. The diffraction grating allows the spectrum to be viewed, rather like the prism does for a continuous spectrum.



> *Line spectra are observed by viewing a disharge tube with a diffraction grating*

Origin of line spectra

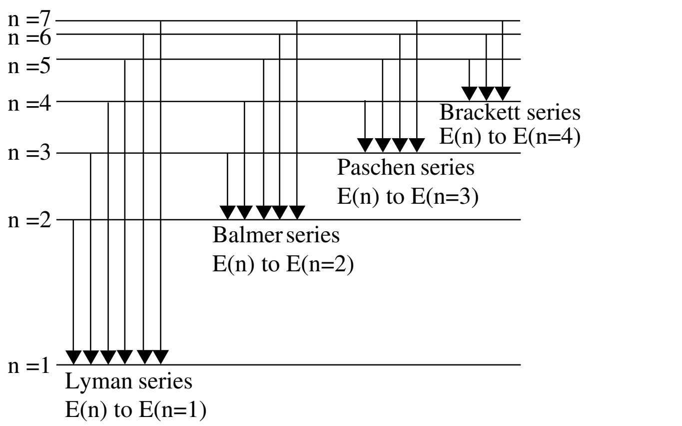
An emission line spectrum is produced by a hot gaseous element. At high temperatures the electrons of the atoms are excited into higher energy levels. When they drop back to a lower level, the energy is emitted at a specific frequency (wavelength) depending on the energy gap between the levels, determined by the equation:

> $\Delta E = hf$ *Where* ΔE = *energy level gap (J),*
> f = *frequency of the observed spectral line (Hz)*
> h = *Planck's constant.*
>
> *The observed lines are of frequency f, determined by the energy level gap and derived from the equation* $\Delta E = hf$
> ***N.B.*** *The energy levels may well be given in eV rather than J, so the figure must be multiplied by* 1.6×10^{-19} *to convert eV to J.*

Significance of Line spectra

The existence of line spectra provides evidence for the existence of quantized energy levels in the atom. Observations of the line spectra for hydrogen – the Lyman, Balmer, Paschen and Bracket Series enabled the energy levels of the single electron of the hydrogen atom to be established and agreement between experimental evidence and predictions of the theory is very good, thus giving confidence in the theory.

Electron transitions for the hydrogen atom



Calculations of the frequencies of lines in spectra

> ***Example: The ground state for the electron in a hydrogen atom is –13.6 eV, the first excited state is –3.4eV. Calculate the frequency of the line in the spectrum representing this transition.***
>
> $\Delta E = hf$, so $(13.6 - 3.4) \times 1.6 \times 10^{-19} = 6.63 \times 10^{-34} \times f$
>
> $f = \dfrac{10.2 \times 1.6 \times 10^{-19}}{6.63 \times 10^{-34}} = 2.46 \times 10^{15}$ Hz

Absorption Spectra

If white light (i.e. radiation of all frequencies – this term is used for the whole EM spectrum, not just visible light) passes through a cooler vapour, then the spectrum observed is of a continuous coloured spectrum with dark lines in the positions where the bright lines would have been in an emission spectrum. This is known as an absorption spectrum.

The absorption spectrum is formed as the vapour absorbs specific frequencies from white light. These frequencies are linked to the energies required for the electrons to jump into higher excited states. When the electrons later drop back into lower states, they emit radiation at these same frequencies. However this radiation is emitted in all directions, so the intensity in the orginal direction is reduced.

An absorption spectrum consists of dark lines where the bright lines would have been in an emission spectrum. It is formed when white light passes through a cooler vapour.

Importance of absorption spectra

Observations of the absorption spectra of the light from distant stars has been used to identify which elements are present in the star. The white light passes through the cooler vapours of the outer layers of the star, forming the absorption spectrum. The positions of the dark lines can be compared with emission spectra of known elements on Earth.

Absorption spectra indicate what elements are present in stars.

Doppler Shift of line spectra from distant stars.

Factsheet 51 deals with this important aspect of the use of line spectra. Known spectral lines are found to be shifted slightly towards lower frequencies (Red Shift). One explanation of this is that the source is moving away and the speed of recession can be calculated from the theory. This provides evidence to support the Big Bang theory of the origin of the Universe.

Red Shift of line spectra suggests that galaxies are moving away from each other, and this supports the Big Bang theory of the origin of the Universe.

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's mark scheme is given below.

(a) Describe how you would produce and observe an emission line spectrum in the laboratory. (2)

You would look at a discharge bulb through a slit 0/2

The candidate has incorrectly described it as a bulb rather than a discharge lamp, and has indicated a slit rather than a diffraction grating.

(b) Describe what the spectrum would look like. (1)

A series of lines 1/1

While this is a very simplistic description, it is adequate for 1 mark. If more marks were allocated for the question, then a better description would be: "a series of discrete bright lines against a dark background."

(c) Explain the origin of emission line spectra (3)

When the electrons in an atom change energy the line is emitted. 1/3

The candidate has shown understanding of energy change , but given no idea of quantized levels.

(d) Explain how observation of absorption spectra helps to determine the elemental make-up of a star. (4)

Each line represents a particular element, so you can tell which elements are there. 1/4

The candidate has some understanding of the link between the lines and the element, but this is obviously insufficient for 4 marks

Examiner's Answers

(a) You would view a gas discharge lamp with a diffraction grating.

(b) The spectrum is a series of discrete bright lines against a dark background.

(c) Energy levels for the electrons in the atom are quantized. When an excited electron drops back from a higher level to a lower level, the energy is emitted as a photon of frequency given by: $\Delta E = hf$.

(d) Absorption spectra are produced when white light passes through a cooler vapour (such as the atmosphere of a star). Dark lines appear where bright ones would have been in the emission spectrum for a particular element, so comparison with known element emission spectra identifies the elements.

Typical Exam Question

(a) Decribe the appearance and origin of an emission line spectrum. (5)

(b) Discus the similarities and differences between emission spectra and absorption spectra. You may be awarded a mark for the clarity of your answer. (5)

(a) An emission line spectrum consists of discrete bright lines against a dark background. It arises because electrons in the atom cannot take any value of energy, only certain allowable levels. In a hot vapour, the electrons are excited into higher energy levels and when they drop back to the ground state in stages, the energy difference between levels is emitted as a photon of frequency given by $\Delta E = hf$, where ΔE is the energy difference, h is Planck's constant and f the frequency.

(b) Both emission and absorption spectra consist of discrete lines at certain frequencies, but the absorption spectrum has bright lines against a dark background at the same frequencies which the emission spectrum has as bright lines against a dark background.
Both occur because of electron movement between allowable levels, but emission are due to excited electrons dropping down into lower levels, whereas absorption are due to electrons being excited into higher levels. Emission occurs from hot vapours, whereas absorption occurs when white light (all frequencies) passes through cooler vapour.

Questions

1. What is an emission line spectrum?

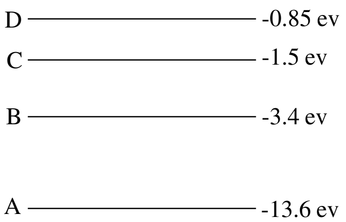
2. Why does an absorption line spectrum occur when white light passes through a cooler vapour?

3. How can observation of absorption spectra be used to identify the elements present in the atmosphere of a star?

4. (a) What is meant by "Red Shift" of line spectra?
   (b) Explain how the red shift of the line spectra of distance galaxies provides evidence to support the Big Bang theory.

5. Explain how emission spectra support the idea of quantization of energy levels.

6. The diagram shows some of the energy levels for atomic hydrogen.



   (a) Calculate the wavelength of the line which would appear in a line spectrum for the transition between the levels marked A and C.

   (b) Which transition represents a photon **absorbed** with the shortest wavelength.

   (c) Which transition represents a photon **emitted** with the longest wavelength.

Answers

1. See text.

2. When white light passes through a cooler vapour, frequencies appropriate to allowable transitions between energy levels in the atom are absorbed and electrons go into higher energy levels. Later, they drop back, but the photons are emitted in all directions, whereas the original energy was absorbed from the forward direction, so the specific frequency appears darker.

3. The frequencies of absorption lines in the spectra from stars can be compared with the emission lines in spectra of known elements and the elements in the atmosphere of the star identified.

4. (a) Red Shift is the slight shifting of known frequencies in the absorption spectra of distant stars to lower values (longer wavelength).

   (b) The most likely explanation for the red shift is that the source is moving away (Doppler effect). This implies that galaxies are now moving away from one another, i.e. they were originally all in the same place – Big Bang theory.

5. From a study of the Lyman series of lines in a hydrogen spectrum, possible values for the energy levels can then be worked out. These values can then be used to predict other series which should exist. Good agreement between theory and observation confirms the theory.

6. (a) $\Delta E = (13.6 - 1.5) \times 1.6 \times 10^{-19}$J

   so, $f = \dfrac{12.1 \times 1.6 \times 10^{-19}}{6.63 \times 10^{-34}}$, $c = f\lambda$

   so $\lambda = \dfrac{3 \times 10^{8} \times 6.63 \times 10^{-34}}{(12.1 \times 1.6 \times 10^{-19})} = 1.03 \times 10^{-7}$m

   (b) A $\rightarrow$ D . In absorption, the electron is excited to a higher level. Shortest wavelength is the highest frequency, therefore the largest energy gap.

   (c) D $\rightarrow$ C . In emission, the electron drops back to a lower level. Longest wavelength is lowest frequency, therefore the smallest energy gap

Acknowledgements:
This Physics Factsheet was researched and written by Janet Jones
The Curriculum Press,Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

www.curriculum-press.co.uk Number 134

Atomic Calculations from the Bohr Model

Bohr knew that the hydrogen atom consisted of a negatively charged electron orbiting a positively charged proton but... although atoms are small there was a very BIG problem.

The electron was orbiting in a circle. That meant it was accelerating. But accelerating charges always give out electromagnetic radiation. So the electron should lose energy.

Bohr took the dramatic step of saying that the electron in an atom could only be in certain orbits and couldn't be anywhere else. In addition, when the electrons were in these orbits they didn't radiate any energy at all. Furthermore, he said he could calculate the radius of the closest possible orbit. It was $r = 5.3 \times 10^{-11}$ m, which became known as the Bohr Radius.

It might seem obvious - the electron is a small particle. But in the case of the atom the electron didn't behave like a particle. For a start when an electron fell towards a proton and ended up at the Bohr radius it did not turn its potential energy into kinetic energy, it gave out a photon of energy equal to the loss in potential energy (13.6eV)

A Prince to the Rescue!

Prince Louis de Broglie, a dozen years later, found himself thinking that waves and particles were really the same thing. de Broglie (pronounced 'de Brooey') came up with an equation which linked particles and waves. It is:

$\lambda = h/p$

λ is the wavelength of the 'wave' in metres (m)
h is Planck's constant, (6.63×10^{-34} Js)
p is the momentum of the 'particle' ($kgms^{-1}$)

Worked Example 1
What is the 'wavelength' of the electron in the Bohr orbit?

Momentum = mass × velocity. So $\lambda = h/mv$

As h, m, and v were known:

$$\lambda = \frac{6.63 \times 10^{-34}}{(9.1 \times 10^{-31} \times 2.19 \times 10^{6})} = \mathbf{3.32 \times 10^{-10}\ m}$$

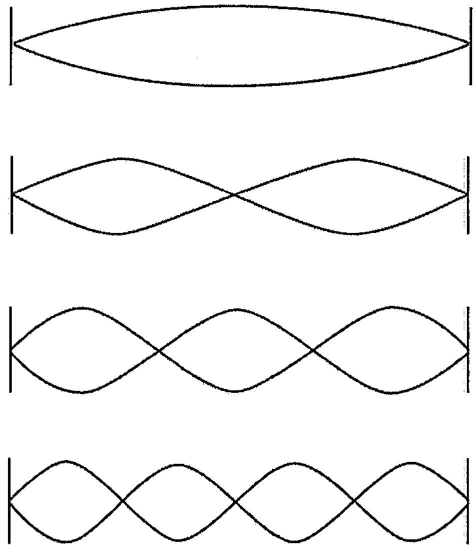
Now consider this:
Take the Bohr radius. Multiply it by 2π. This gives the circumference of the Bohr orbit. Put the numbers in:

$$2\pi r = 2 \times 3.14 \times 5.3 \times 10^{-11} = \mathbf{3.32 \times 10^{-10}\ m}$$

But this was exactly the same as the 'wavelength' of the electron that was in that orbit! The conclusion was inescapable.

The electron was behaving like a wave once it was part of an atom. Even better, to satisfy Bohr's condition that the electron didn't radiate energy away as it 'orbited', the electron wave had to be a standing or stationary wave.

Exam Hint:
These are the shapes of the first few standing waves. It is a good idea to learn them.



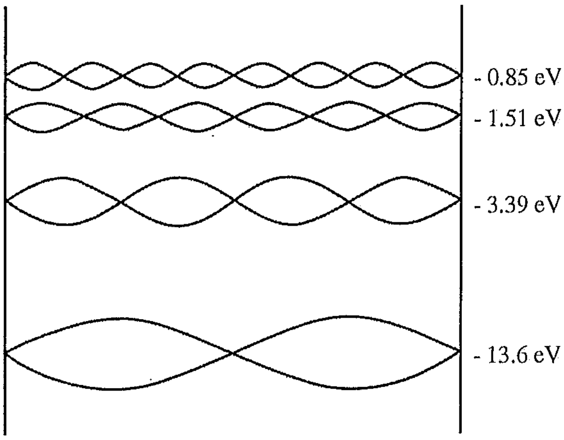
Bohr had calculated other radii where the electron could be orbiting and still not give out any energy. These radii turned out to be where the next possible standing waves could wrap themselves around the proton.

Exam Hint: *Be careful when you say exactly which standing wave shapes occupy the Bohr orbits.*

Only the waves with whole numbers of wavelengths are stable. So only the second and fourth shapes shown above correspond to actual orbits.

It was only standing waves that could exist. If the waves weren't standing waves they would not be stable and so would give out their energy and cease to exist.

The standing waves corresponded to the energy levels found by experiment as shown below.



Now it was clear. The electron wasn't a particle at all once it was orbiting around the proton. It was acting as a standing wave.

Worked Example 2

Remember that an electron falling towards a proton from a large distance away will finally settle into an orbit whose radius is the Bohr radius. Or perhaps we should say it will become a standing wave in its simplest allowed pattern. Either way it still has to get rid of 13.6eV of potential energy. (see diagram above). It does this by radiating away a photon of exactly this energy.

$E = hc/\lambda = 13.6eV = 13.6 \times 1.6 \times 10^{-19} = 2.18 \times 10^{-18}$ J

so $\lambda = \frac{hc}{2.18 \times 10^{-18}} = \frac{(6.63 \times 10^{-34} \times 3 \times 10^{8})}{2.18 \times 10^{-18}}$

Answer $= 9.1 \times 10^{-8}$ m

This is in the ultraviolet part of the electromagnetic spectrum. By experiment on hydrogen atoms, light of exactly this wavelength was observed to be given out when a far-away electron fell towards a proton to make a complete hydrogen atom.

*The electron, or wave, is not stable at higher energy levels. So it quickly, in a fraction of a second, returns to its lowest possible energy level. The energy **difference** is radiated away from the atom in the form of a photon.*

Exam Hint: *Be careful which way the electron moves. If it drops down energy levels then a photon is emitted. But an electron can also absorb a photon of exactly the right energy and move up energy levels. Make sure you make clear to the examiner that you know whether the photon is being **absorbed** or **emitted**.*

Worked example 3

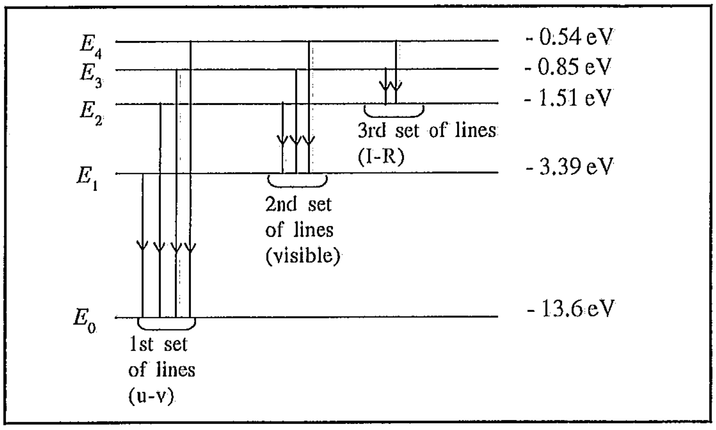
If the electron fell from the -3.39eV level (otherwise known as the first excited state) back to the -13.6eV level (the ground state) then the energy **difference** between these levels is 13.6 – 3.39 = 10.21eV. This is:

$10.21 \times 1.6 \times 10^{-19} = 1.63 \times 10^{-18}$ J and this is the exact energy of the photon emitted.

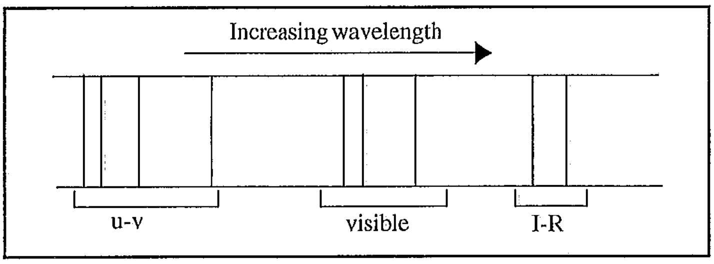
Using $E = hc/\lambda$ we can calculate the wavelength of this photon. It is:

$\lambda = hc/E = (6.63 \times 10^{-34} \times 3 \times 10^{8}) / 1.63 \times 10^{-18} = 1.22 \times 10^{-7}$ m.

The result was that the emission spectrum of hydrogen consisted of lines which showed a definite pattern. This pattern could be explained by using the idea of electrons falling back to the ground state from various higher energy levels (or of standing waves returning to their simplest allowed form).

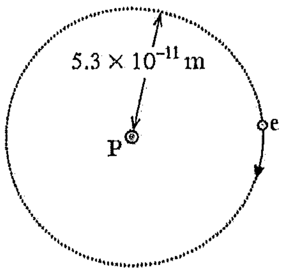


Taken directly from the above diagram, the emission spectrum would look like this.



Practice Questions

1. The Bohr model of a hydrogen atom assumes that an electron **e** is in a circular orbit around a proton **P**.



In the ground state the orbit has a radius of 5.3×10^{-11} m. At this separation the electron is attracted to the proton by a force of 8.1×10^{-8} N.

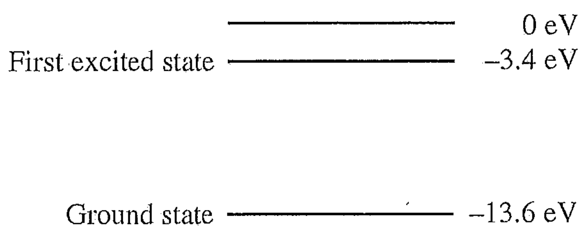
(a) State what is meant by *the ground state.*

(b) (i) Show that the speed of the electron in this orbit is about $2.2 \times 10^{6}\,ms^{-1}$ (mass of an electron = 9.1×10^{-31} kg)

(ii) Calculate the de Broglie wavelength of an electron travelling at this speed. (Planck constant = 6.63×10^{-34} Js)

(iii) How many waves of this wavelength fit the circumference of the electron orbit?

2. The energy level diagram shown below is a very simplified diagram for the hydrogen atom.



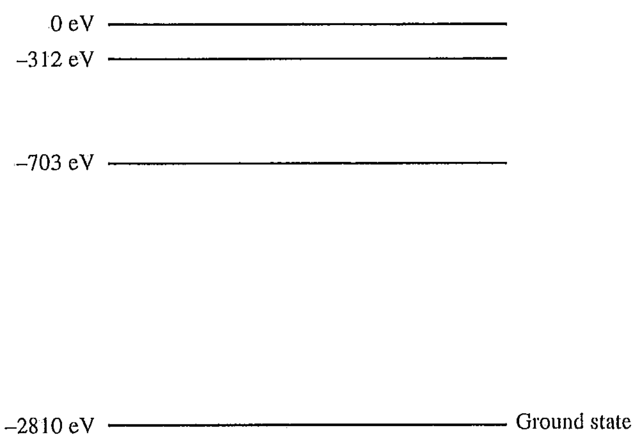
A moving electron with a kinetic energy of 12.5 eV collides with an atom of hydrogen and causes the electron in the ground state to be raised to the first excited state.

(a) Calculate the kinetic energy of the moving electron after the collision (in eV)

(b) Calculate the wavelength of the photon emitted when the electron in the first excited state returns to the ground state.

3. A muon is a particle which was discovered in a particle accelerator. It is a 'heavy electron'. It has exactly the same properties as an electron but its *mass* is 208 times greater.

Scientists have made an 'atom' in which the muon orbits a proton. The muon energy levels for this 'atom' has been discovered.



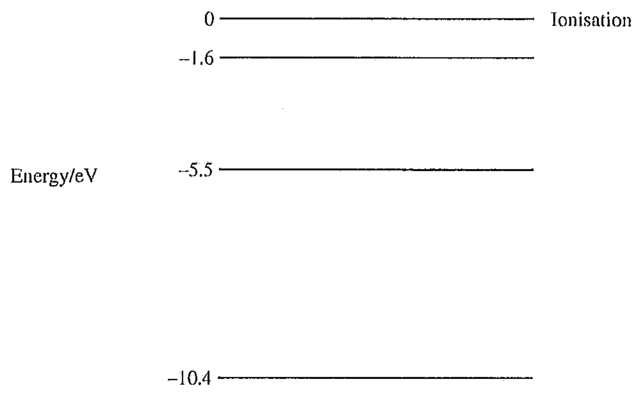
(a) Write down the ionisation energy of this atom.

(b) Hence calculate the *maximum possible* wavelength of a photon which, when absorbed, would be able to ionise this 'atom'.

(c) To which part of the electromagnetic spectrum does this wavelength belong?

(d) Calculate the de Broglie wavelength of a muon travelling at 10% of the speed of light.

4. The diagram shows some of the energy levels of a Thallium atom.



(a) Calculate the ionisation energy in Joules for an electron in the −10.4 eV ground state.

(b) A neutron of kinetic energy 9.4 eV collides with a Thallium atom. As a result, an electron moves from the −10.4 eV level to the −1.6 eV level. What is the kinetic energy in eV of the neutron after the collision?

(c) A transition between which two energy levels in the Thallium atom will give rise to an emission line of wavelength close to 320 nm? Draw this transition on the diagram above.

Answers

1. (b) (i) 2.17×10^6 ms^{-1}
   (ii) 3.3×10^{-10} m
   (iii) Exactly One!

2. (a) 2.3 eV
   (b) 1.22×10^{-7} m

3. (a) 2810 eV
   (b) 4.42×10^{-10} m
   (d) 1.17×10^{-13} m

4. (a) 1.664×10^{-18} J
   (b) 0.6 eV
   (c) photon energy = 3.88 eV

Acknowledgements:
This Physics Factsheet was researched and written by Paul Dutton
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

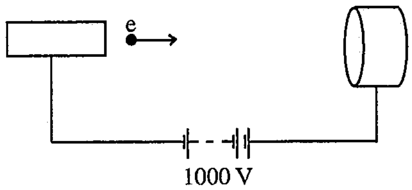
www.curriculum-press.co.uk Number 119

Calculations on the Quantum Atom

Examiners' reports have highlighted some common errors made by students in attempting energy level calculations. In this Factsheet we shall concentrate on getting the calculations correct rather than developing the theory.

Joules and the electron-volt

Energy can always be expressed in Joules. But at the atomic level, a Joule is a very large unit for energy.
From the equation **E = QV**, we can see that accelerating a single electron through a p.d. of perhaps 1000 volts would give the electron a kinetic energy:



$KE = QV = 1.6 \times 10^{-19} \times 1000 = 1.6 \times 10^{-16} J$

It is considerably more convenient to just express this as 1000 electron-volts (eV). To convert energy in electron-volts into joules, you just multiply by the electron charge (1.6×10^{-19} C).

> Electron-volts may be convenient, but Joules must be used in all calculations using formulae based on the SI system of units.

Example: A helium nucleus is accelerated from rest through 5000V.
(a) Find its kinetic energy in eV
(b) Convert this energy into joules
(c) Calculate the velocity (using $m = 1.7 \times 10^{-27}$ kg for protons and neutrons)

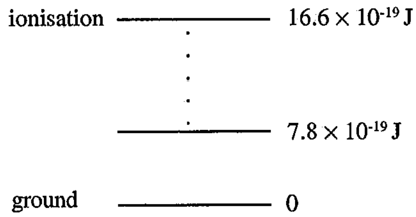
Answer:
(a) $KE = QV = 2 \times 5000 = 10\,000$ eV, or 10 keV.
(b) $KE = 1.6 \times 10^{-19} \times 10\,000 = 1.6 \times 10^{-15}$ J
(c) $1.6 \times 10^{-15} = \frac{1}{2}(4 \times 1.7 \times 10^{-27}) \times v^2$, $v = 6.9 \times 10^5$ ms^{-1}.

Exam Hint: *Be prepared to convert between eV and joules in either direction. To avoid confusion, remember that the value in eV will be a much larger number than the value in joules. In addition be careful to spot prefixes on the unit. It is just as common to write 2.6 MeV as 2.6×10^6 eV.*

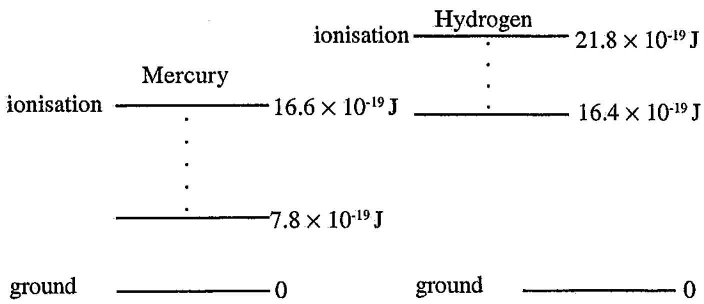
Example: Write 62 MeV in standard form, then convert to joules.
Answer: *62 MeV = 6.2×10^7 eV = 9.9×10^{-12}J.*

"Negative" Energy Levels

An atom is in its *ground state* when it cannot lose any energy. However it can gain energy to jump to higher allowed energy states, as its outer electrons become *excited.* It would seem to make sense to call the ground state zero energy, and make higher energy levels positive in value. For a mercury atom

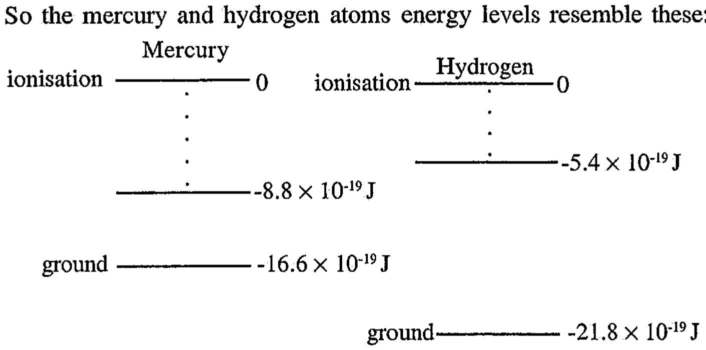


But if we put this next to a hydrogen atom:



This would mean that two identical electrons, freed by ionisation from the two atoms, would appear to have different energies. It wouldn't make sense. The ionisation levels must have the same energy. So we define the ionisation energy levels as zero, and make the lower (trapped) levels negative.

So the mercury and hydrogen atoms energy levels resemble these:



This means that we must be prepared to work with negative numbers.

Example:
(a) Find the energy required to lift an electron in a hydrogen atom from its ground state to the first excited state.
(b) Express this in eV.
Answer:
(a) $-5.4 \times 10^{-19} - (-21.8 \times 10^{-19}) = 16.4 \times 10^{-19}$ J
(b) 10.3 eV

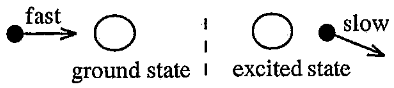
> ***Exam Hint:*** *Make certain that you know how to work with negative numbers. And check that you know how to enter calculations like that above into your calculator. All calculators do not work in the same way.*

Transitions between levels

The actual energy of a level is less important than the transitions between levels. You must be prepared to calculate energies needed to excite atoms to higher levels, and also energies emitted (as photons) when atoms drop to lower levels.

> ***Example: A high-speed electron collides with a hydrogen atom and gives it enough energy to jump from its ground state to its first excited state. What is the minimum speed of this electron? (m = 9.1 × 10^{-31}kg for an electron)***
>
> ***Answer:***
> *Energy gained = 16.4 × 10^{-19}J (as before)*
> *KE = ½mv^2, v = 1.9 × 10^6 ms^{-1}.*
> *This is the minimum speed that the electron must have to cause this transition.*

Of course the moving electron mentioned may be going faster than this. If so, it can transfer 16.4 × 10^{-19}J to the atom in the collision, and retain the rest of its kinetic energy:



> ***Example: The high-speed electron strikes the hydrogen atom with a speed of 2.2 × 10^6 ms^{-1}. If the atom jumps from its ground state to its first excited state, find:***
> ***(a) the original KE of the electron***
> ***(b) the KE after the collision***
> ***(c) its speed after the collision.***
>
> ***Answer:***
> *(a) KE = 2.2 × 10^{-18}J.*
> *(b) KE = 2.2 × 10^{-18} – 1.64 × 10^{-18} = 5.6 × 10^{-19}J*
> *(c) v = 1.1 × $10^6$$ms^{-1}$*

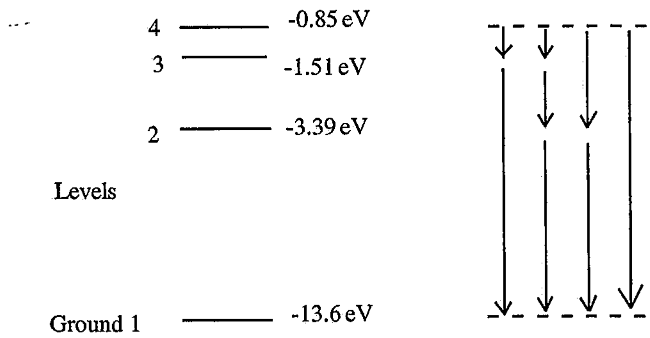
> When an incident particle has enough KE to actually *ionise* an atom, the freed electron may well have some KE itself, so we cannot be certain how much KE the incident particle will have remaining.

After an atom is excited into a higher state, it will eventually drop back to its ground state, emitting energy as a quantum of e.m. radiation. The relationship **E = hf = hc/λ** allows us to calculate the wavelength of the radiation emitted. (h = 6.6 × 10^{-34}Js)

> ***Example:***
> ***(a) Find the wavelength emitted by a hydrogen atom when it drops back from its first excited state to its ground state.***
> ***(b) In which part of the e.m. spectrum is this?***
>
> ***Answer:***
> *Energy lost = 16.4 × 10^{-19}J*
> *So λ = hc / E = 1.2 × 10^{-7}m. This is in the ultra-violet region.*

> ***Example: What is the shortest wavelength radiation that could be emitted by a hydrogen atom?***
>
> ***Answer:*** *The shortest wavelength implies the greatest energy transition in the atom.*
> *This is the drop from the ionisation level to the ground state (21.8 × 10^{-19}J).*
> *λ = hc / E = 9.08 × 10^{-8}m. This is in the ultraviolet region again.*

Of course an electron can drop down to the ground level in more than one step. This diagram shows the bottom four energy levels of the hydrogen atom (with the energy levels expressed in eV), and possible sets of transitions down to the ground level:



> ***Example: How many wavelengths of e.m. radiation could be emitted by hydrogen atoms at level 4 returning to the ground state?***
> ***Answer:*** *There are 6 different energy transitions shown: 4 to 3, 4 to 2, 4 to 1, 3 to 2, 3 to 1, and 2 to 1.*

> ***Exam Hint:*** *When working out possible sets of transitions, use a logical approach as in the diagram. It is easy to miss one out when putting them in a random order.*

Practice Questions

1. An electron has a velocity of 4.0 × 10^6 ms^{-1}.
   (a) Find the kinetic energy in joules (m = 9.1 × 10^{-31} kg)
   (b) Find its kinetic energy in eV.
   (c) What p.d. would an electron at rest have to be accelerated through to gain this amount of energy?
2. Write these energies in standard form, then convert them to Joules:
   (a) 3.4 MeV (b) 26.2 keV (c) 95 meV
3. (a) Find the energy required to excite a mercury atom in its ground state to its first excited state.
   (b) Convert this to eV.
   (c) What velocity for a moving electron is matched to this energy?
4. A proton travelling at a speed of 1000ms^{-1} strikes a mercury atom in its ground state. Could it raise the atom into an excited state? (m = 1.67 × 10^{-27}kg)
5. If you were given the bottom five energy levels for an atom, how many different energy transitions are possible?
6. Using the hydrogen atom energy levels given in eV, find the longest wavelength that could be emitted, and identify its position in the e.m. spectrum.

Answers

1. (a) KE = ½mv^2 = 7.28 × 10^{-18}J
   (b) KE in eV = 7.28 × 10^{-18} / 1.6 × 10^{-19} = 45.5 eV.
   (c) p.d. = 45.5 V.
2. (a) 3.4 MeV = 3.4 × 10^6 eV = 5.4 × 10^{-13} J
   (b) 26.2 keV = 2.62 × 10^4 eV = 4.2 × 10^{-15} J
   (c) 95 meV = 9.5 × 10^{-2} eV = 1.5 × 10^{-20} J
3. (a) 7.8 × 10^{-19} J (b) 4.9 eV (c) 1.3 × 10^6 ms^{-1}.
4. KE = ½mv^2 = 8.35 × 10^{-22} joules. This is not nearly sufficient energy.
5. Ten unique transitions.
6. Longest wavelength implies smallest energy transition.
   The smallest energy transition is -0.85 – (-1.51) = 0.66 eV = 1.06 × 10^{-19} J
   So λ = hc / E = 18.6 × 10^{-7} m. This is in the infrared region.

Acknowledgements:
This Physics Factsheet was researched and written by Paul Freeman

Physics Factsheet

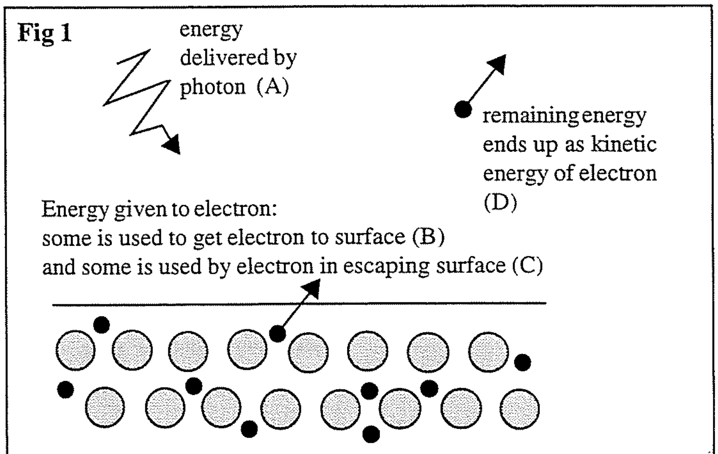
www.curriculum-press.co.uk Number 105

Calculations on the Photoelectric Effect

When a metal surface is illuminated with light, electrons can be emitted from the metal surface, provided the frequency of the light is above a certain value. This is called the photoelectric effect.

> *The* ***photoelectric effect*** *is one of the key pieces of evidence that light can behave as a stream of particles called* ***photons****, rather than as a wave. If light were behaving as a wave, then the emission of electrons would behave very differently. This is beyond the scope of this factsheet.*

Fig 1



As you can see in Fig 1, behind the photoelectric effect is a really simple idea:

Energy delivered by the photon = energy delivered to the electron (A in the diagram)

Of the energy delivered to the electron:
- some is used in getting electron *to* the metal surface (B in the diagram)
- some is used in getting electron to *escape* the metal surface (C)
- whatever energy remains appears as *kinetic energy* of the electron (D)

So, very simply: A = B + C + D

A : is the energy of the photon. This is proportional to the frequency of the light, such that Energy of photon = hf
where h is Planck's constant = 6.63×10^{-34} Js

C: is the energy needed for an electron to escape the surface. This is known as the **work function**, and is a constant for a given metal. For example, the work function of zinc is 4.24 eV, and that of caesium is 1.35 eV. The symbol for work function W or ϕ .

D is the kinetic energy of the electron and is thus given by:
$KE = \frac{1}{2} m v^2$

So: **Energy delivered by the photon = energy delivered to the electron** now turns into:

$$hf = \text{energy to get to surface, B} + W + \tfrac{1}{2} m v^2$$

The speeds (and energies) of the electrons will depend upon how much of their energy they used getting to the surface. The fastest electrons will be the ones that are already at the surface, so that B is zero. This gives us our final formula, relating only to the fastest electrons, and the one we will use:

$$hf = W + KE_{max}$$

or, $hf = W + \frac{1}{2}mv^2_{max}$

> ***Exam Hint:*** *when calculating speeds from kinetic energies as we will do later, the KE must be in Joules. However, as you have already seen, the work function and photon energies are often given in electronvolts (eV). It is therefore vital that you:*
> - *remember at all times which units you are using*
> - *are able to convert between them: 1 eV = 1.6×10^{-19} J*

Worked example

A metal surface with a work function of 4.0 electronvolts is illuminated by light of wavelength 200 nm. What is the maximum velocity of the photoelectrons produced?

speed of light **= 3×10^8 m s^{-1}**
Planck's constant **= 6.6×10^{-34} J s**
mass of electron **= 9.1×10^{-31} kg**

The first step is to calculate the frequency of the light from its wavelength
v = fλ, so

$$f = \frac{v}{\lambda} = \frac{3 \times 10^8 \text{ m s}^{-1}}{200 \times 10^{-9} \text{ m}} = 1.5 \times 10^{15} \text{ Hz}$$

(don't forget the powers of ten when dealing with nano, micro, milli etc)

Now we can use either of the formulas above.
The first ($hf = W + KE_{max}$) will work but requires a two step process – you will have to use $KE = \frac{1}{2} m v^2$ afterwards to convert KE into speed, so we will use the second formula.

$$hf = W + \tfrac{1}{2}mv^2_{max}$$

$$(6.6\times10^{-34}\text{ J s})\times(1.5\times10^{15}\text{ s}^{-1}) = (4.0\times1.6\times10^{-19}\text{ J}) + (0.5\times9.1\times10\text{-}31\text{ kg}\times v^2)$$

(Notice that I changed the work function into Joules from electronvolts)

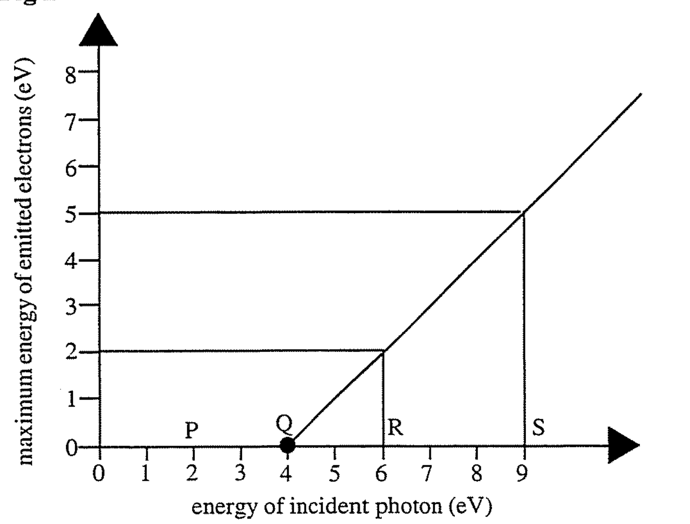
giving $v = 8.8 \times 10^5$ m s^{-1}

> ***Exam Hint:*** *Check you can get this result with your calculator. It is very common for students to do all the hard work and then forget to square-root v^2 at the end. Don't fall into this trap!*

Graphs in the photoelectric effect

The formula $hf = W + KE_{max}$ results in the following graph

Fig 2



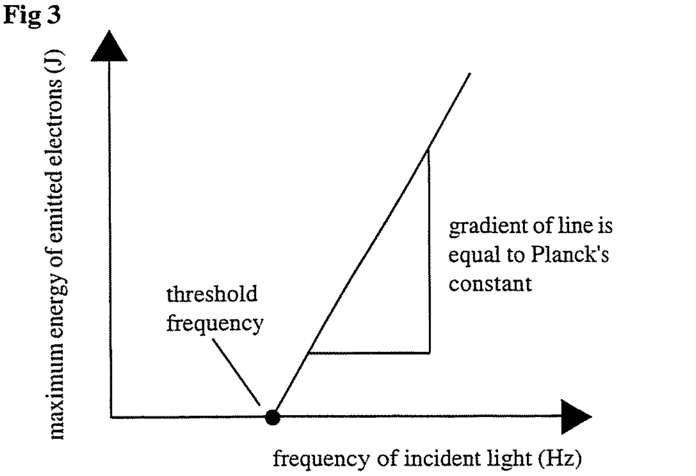
Consider 4 incident photons P, Q, R and S, which have different energies (and frequencies)

- Photon **P** does not have enough energy to cause any **photoemission** of electrons.
- Photon **Q** has *just* enough energy to cause photoemission. There is no 'spare energy' left over, so the photoelectrons have no KE – all the energy went into leaving the surface. This is of course the work function, so the work function of this particular metal is 4 eV.
- Photon **R** had 6 eV. 4 eV of those were used in getting from the surface, so the maximum KE possible is 2 eV.
- Photon **S** had 9 eV. 4 eV of those were used in getting from the surface, so the maximum KE possible is 5 eV
- By considering R and S, you should be able to see that the gradient of the graph is 1, or '45 degrees'.

> *Note that if the energy of the photon is not great enough for an electron to escape the surface, then no electrons will be emitted. This is why there is a threshold frequency: above that frequency electrons are emitted, below it they are not.*

Sometimes this graph is plotted with the *frequency* of the light on the x-axis rather than the *photon energy*. In this case the gradient of the line is not 1.

Fig 3



The equation of a straight line is: $y = m\ x + c$

The photoelectric equation can be written $KE = hf - W$

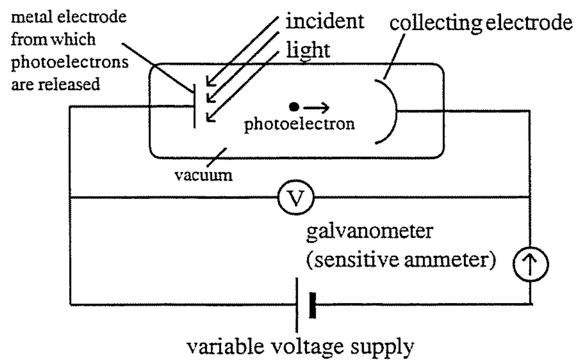
Comparing the two, and looking at Fig 3 shows that:

- the gradient of the line gives a value for Planck's constant
- the threshold frequency, f_0, can be read off the x-axis
- the work function can be obtained by continuing the line and reading off the negative y-intercept. Notice that the work function can also be obtained from the threshold frequency: $W = hf_0$

> A photoelectron is not a different particle from an electron. It is just an electron that has been emitted from a metal by a photon. It is still a normal electron!

Stopping potential

If a negative electrode is placed near the emitting metal surface, then it will repel the photoelectrons. If it is negative enough, then the photoelectrons will not reach it.



The circuit above shows a **photocell**. As the voltage of the supply is increased, more and more electrons are repelled back, and the galvanometer reading falls. Only the fastest electrons reach the collecting electrode. When the supply voltage reaches a value known as the **stopping potential**, even the fastest electrons can not reach the collecting electrode and the current falls to zero.

In energy terms, the electrical potential energy of an electron near the collecting electrode has risen higher than the kinetic energy of the photoelectrons, which therefore cannot reach the electrode. (This is the electrical equivalent of a very simple gravitational idea: if you place a target above you on a wall you may be able to hit it by throwing a ball, but if you raise the target there will come point where you can not give the ball enough KE to reach.)

The electrical energy transferred to an electron crossing from one electrode to the other is qV, where q is the charge on an electron and V is the potential difference between the electrodes. At the stopping potential V_{stop}, the fastest electrons are *just* prevented from making it, so:

$$qV_{stop} = \frac{1}{2}mv^2_{max}$$

which equals $hf - W$

Practice Questions

For the following questions, you may use the data below:

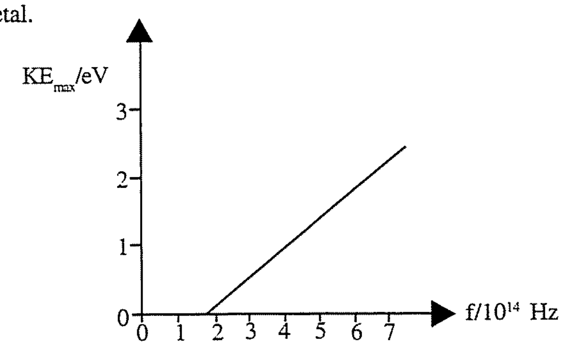
speed of light = $3 \times 10^8\ m\ s^{-1}$
mass of electron = 9.1×10^{-31} kg
charge on electron = 1.6×10^{-19} C
Planck's constant = 6.6×10^{-34} Js

1. Light of frequency 9.2×10^{14} Hz falls upon a metal with a work function of 2.5 eV.
   (a) Calculate the maximum kinetic energy of the resulting photoelectrons
   (b) Calculate the maximum speed of the resulting photoelectrons
   (c) What is the threshold frequency for photoelectric emission?
   (d) If a nearby electrode is made negative using a potential difference V, what value of V is required to stop any photoelectrons reaching it?

2. The work function of zinc is 4.2 eV. What is the maximum wavelength of light that will cause photoemission of electrons from zinc?

3. Light of wavelength 0.60 μm incident upon a metal surface ejects photoelectrons with kinetic energies up to a maximum value of 2.0×10^{-19} J.
   (a) What is the work function of the metal?
   (b) If a beam of light causes no photoelectrons to be emitted, what can you say about its wavelength?

4. The graph below shows how the maximum KE of photoelectrons depends on the frequency of incident radiation for a certain metal.



   (a) Use the graph to estimate the work function of the metal
   (b) Use the graph to obtain a value for Planck's constant (harder)

5. A sample of magnesium is used as an electrode in a photocell. It is illuminated with u-v light of wavelength 225 nm, and a current flows in the photocell. The current can be reduced to zero by making the other electrode negative using a potential difference of 1.4 V. Calculate the work function of magnesium.

6. The work function of caesium is 1.35 eV. A photocell contains a caesium surface that is illuminated with light of wavelength 380 nm. What potential difference must be applied the the cell to just prevent a current passing through it?

7. A metal surface with a work function of 2.9 eV is illuminated with light of wavelength 400 nm. What will be the stopping potential for the photoelectrons?

Answers

1. (a) 2.1×10^{-19} J
   (b) $6.7 \times 10^5\ m\ s^{-1}$
   (c) 6.1×10^{14} Hz
   (d) 1.3 V

2. 2.9×10^{-7} m (295 nm)

3. (a) 1.3×10^{-19} J (0.81 eV)
   (b) must be greater than 1.5×10^{-6} m

4. (a) approx 0.7 eV
   (b) approx 6.7×10^{-34} Js

5. 6.6×10^{-19} J (4.1 eV)

6. 1.9 V

7. 0.19 V

Acknowledgements:
This Physics Factsheet was researched and written by Michael Lingard
The Curriculum Press, Bank House, 105 King Street, Wellington, Shropshire, TF1 1NU

ISSN 1351-5136

Physics Factsheet

September 2000 Number 01

The Quantum Nature of Light

This Factsheet covers one of the fundamental - and hard to grasp - concepts of modern physics; wave-particle duality. To ensure this important topic is clear, the Factsheet is longer than usual.

What is light?

Well, there's no simple answer, because it behaves differently in different situations. Usually it behaves as a wave – it can be diffracted, reflected and refracted – so we consider it to be a wave. However, sometimes this 'wave model' does not explain what we see – for example, when we shine light onto a metal surface, electrons can be released (the photoelectric effect). This requires a new explanation; that light comes in lumps (quanta), or particles called photons. So depending on what experiment we do, light behaves as *either* a 'wave' or a 'particle'. Neither explanation by itself can describe all that we observe.

> ***'Wave-particle duality'** means:*
> *the ability of something to show both wave-like and particle-like properties, depending on how we look at it (i.e. what experiment is performed). It is just one aspect of **quantum physics**.*

Light is part of the electromagnetic spectrum, which we commonly explain as consisting of waves with different frequencies (e.g. gamma rays, X-rays, radio waves, infra-red radiation and microwaves). We find that all parts of the electromagnetic spectrum show the same 'wave-particle duality' as light, so we commonly speak of '**electromagnetic radiation'** (EM radiation) rather than 'light'.

What is the particle model of EM radiation?

By Plank's theory, EM radiation (e.g. light) is considered to be composed of a stream of small lumps or packets (quanta), which are called **photons**. Each photon travels at the speed of light and has no mass and no charge. Each photon has a fixed amount of energy (E) which depends on the frequency (f) of the EM radiation: the higher the frequency, the higher the energy of each photon:

$$E = hf$$

E: Photon energy (J)
h: Planck's constant (6.63×10^{-34} Js)
f: frequence of EM radiation (Hz)

The more photons (i.e. packets of energy) there are, then the more energy there is arriving at a point each second. Thus, the **power** (defined as energy per second) of the EM radiation depends on the number of photons arriving per second. If the power is measured over an area (e.g. $1m^2$), then we call it **intensity** (Wm^{-2}), which in the case of light is the same as its **brightness**.

> *To summarise the Particle Model of EM radiation:*
> - *EM radiation comes in a stream of packets, known as **photons**.*
> - *Each photon has no charge and no mass*
> - *The **Energy** of each photon depends on frequency only*
> - *The **Power** or **Intensity** of EM radiation of a fixed frequency depends on the number of photons arriving each second (where Intensity = Power/area).*

> **Exam Hint :** If you are asked to explain "Wave Particle Duality" with examples, ensure you give examples of light acting both as a particle and as a wave

The differences between the particle and wave views of light

The fundamental differences between the wave and particle views of light are in their explanations of (i) the energy of the light and (ii) how often the 'light' arrives.

The *particle theory* says the energy depends on frequency; and the number of photons arriving per second depends on the power (see above for explanations).

The *wave theory* says the energy depends on the amplitude of the wave, which is related to the power (power $\propto$ amplitude^2); and the number of waves arriving per second is defined as the frequency.

In summary: Differences between particle & wave views of light

Property of Light	Depends on	
	Particle Theory	Wave Theory
Energy	**Frequency** $E = hf$	**Power** $P \propto A^2$
How often it arrives	**Power** #photons/sec	**Frequency** #waves/sec

The Photoelectric Effect

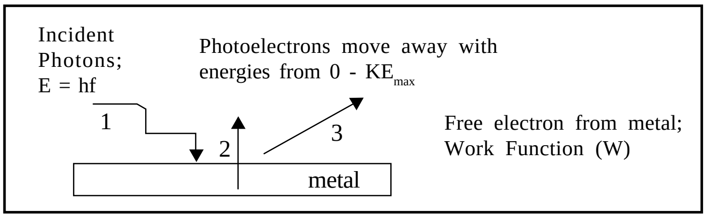
> ***What is the Photoelectric Effect?***
> *When EM radiation falls on a metal surface, electrons are emitted so long as the frequency of the EM radiation is above a certain 'threshold' value. The emitted electrons, called photoelectrons, have kinetic energies ranging from zero to a maximum value.*

How is the Photoelectric Effect explained?

- Light consists of *particles* (photons) of energy; the higher the frequency, the higher their energy.
- One ***single*** photon collides with *one* electron and gives all its energy to that electron.
- If this energy is enough to free the electron from the metal surface, then the electron will be released, otherwise it won't be released.
- *The minimum energy required to release an electron from a metal surface* is called the **Work Function (W)**. This differs for different metals, as each metal surface binds the electrons with varying 'strengths'.
- Whether the electron is released or not depends on the energy of the photon (E) in comparison to the metal work function (W):
  - **Electron just released (E=W).** If the energy of a *single* photon (E) is equal to the work function (W), it can *just* release *one* electron from the *surface* of the metal. We say the photon is at the **threshold frequency, (f_0)**, *the minimum frequency of EM radiation required to just release electrons from a metal surface,* where $E = hf_0 = W$; or rearraged $\mathbf{f_0 = W/h}$.
  - **Electron not released (E<W).** If the EM radiation is below the threshold frequency, one photon does not have enough energy to release one electron and no electrons are emitted.
  - **Electron released with energy to spare (E>W).** A photon above the threshold frequency not only has enough energy to free the electron from the metal, but has excess energy which is given to the electron as kinetic energy (KE). The largest value this can take is given by Einstein's Equation (see below).

By using the principle of conservation of energy, we can also write an equation for the transfer of energy from the photon to the electron (pictorially represented in Fig 1).

Fig 1. The Photoelectric Effect



Photon energy → Energy to free electron from metal surface + Electron KE
hf = W + KE_{max}

Re-arranging this, we obtain;

Einstein's Equation

$KE_{max} = hf - w$

KE_{max}: *maximum electron kinetic energy (J)*
hf : photon energy (J)
W : metal work function (J)

Exam Hints
- **Einstein's Equation** may be written in any of the following forms:
  $KE_{max} = hf - hf_0$ (by $f_0 = W/h$)
  $KE_{max} = hc/\lambda - hc/\lambda_0$ (by $c = f\lambda$)
  $= hc(1/\lambda - 1/\lambda_0)$
- Often energies are given in electron-volts (eV). To convert eV into joules, multiply by the charge on an electron, 1.6×10^{-19}C, as 1 eV = 1.6×10^{-19}J. *E.g.* $2eV \rightarrow 2\times(1.6\times10^{-19}) = 3.2\times10^{-19}$ *J*.

Why is the kinetic energy shown as the *maximum* value (KE_{max})? As W is the *minimum* energy required to release an electron from the *surface* of the metal, it follows that the equation gives the largest electron kinetic energy possible, thus it is shown as $\mathbf{KE_{max}}$. Electrons that are deeper in the metal will be more strongly bound and so take more energy to be freed, leaving less energy for the electron's KE. Thus a **range** of electron kinetic energies are observed, ranging from zero up to a maximum value given by the above equation.

Typical Exam Question

(a) Explain what each of the terms in Einstein's Equation represents: $hf = KE_{max} + \phi$ [3]

A metal surface of work function 3.0 eV is illuminated with radiation of wavelength 350nm.

(b) Calculate the threshold frequency and wavelength. [3]

(c) Calculate the maximum kinetic energy of the emitted photoelectrons, in Joules. [3]

Answer

(a) hf – the energy of each light particle or photon✓ *; KE_{max} – the maximum kinetic energy of the emitted photoelectrons* ✓ *; φ - the work function of the metal* ✓

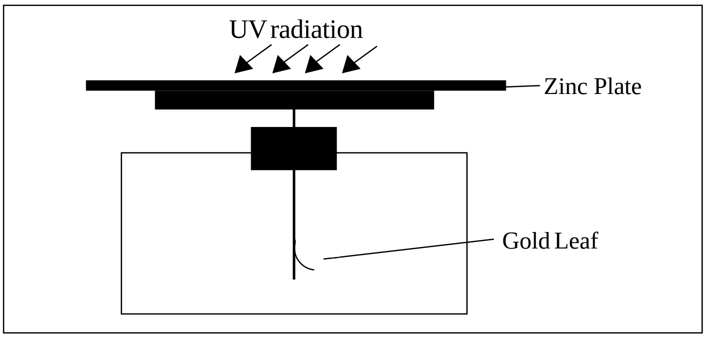
(b) The work function in <u>*joules*</u>, $\phi = 3.0eV \times (1.6\times10^{-19}C) = 4.8\times10^{-19}J$ ✓ *; The threshold frequency (f_0) is defined as* $f_0 = W/h = 7.2\times10^{14}Hz$ ✓ *; The threshold wavelength is given by the wave equation* $c = f\lambda$, $\lambda_0 = c/f_0 = (3.0\times10^8 ms^{-1})/(7.2\times10^{14}Hz) = 4.1\times10^{-7}m$ ✓

(c) Re-arrange the equation in (a) to give $KE_{max} = hf - \phi$ ✓ *; where* $f = c/\lambda = (3.0\times10^8 ms^{-1})/(350\times10^{-9} m) = 8.6\times10^{14}Hz$ ✓ *, and* $\phi = 4.8\times10^{-19}J$ *(N.B. this must be in joules); Thus* $KE_{max} = 9.0\times10^{-20}J$ ✓

A simple demonstration of the Photoelectric Effect

This uses a **gold leaf electroscope** (Fig 2). The electroscope is initially charged either positively or negatively, so the gold leaf is raised.

Fig 2. Gold Leaf Electroscope



When ultra-violet (UV) light is shone on the clean zinc plate, then:
- If the electroscope is initially **negatively** charged, the gold leaf rapidly falls, showing the electroscope is discharging.
- If the electroscope is initially **positively** charged, the gold leaf does not move from its raised position: no effect is observed.

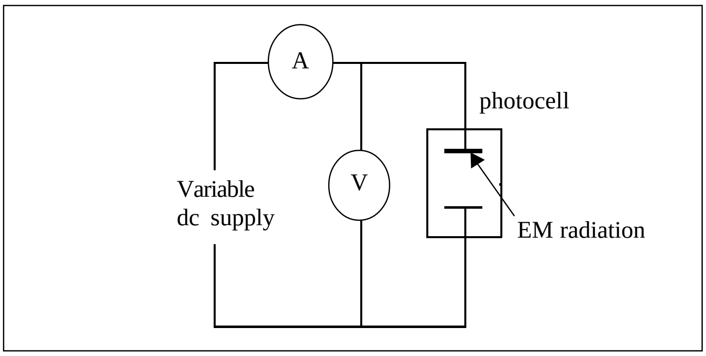
These observations are explained by the Photoelectric Effect, as electrons are released from the zinc plate when UV radiation falls on it:
- When the plate is **negatively** charged, any electrons emitted from the zinc plate will be repelled from the plate as 'like charges (negative) repel'. Thus the electroscope 'loses' charge and discharges.
- When the plate is **positively** charged, any emitted electrons are immediately attracted back to the plate as 'unlike charges attract'. So photoemission does occur, but the electroscope does not discharge.

Important points to remember
- The UV light is not charged
- A positively charged plate is one which lacks electrons.
- Only electrons are emitted from the metal surface

Measuring the Photoelectric Effect and checking Einstein's Equation

Experimental measurements can be made to confirm the Photoelectric Effect (or to determine Einstein's Equation) using the apparatus shown in Fig 3.

Fig 3. Photoelectric Effect apparatus



Apparatus

This essentially consists of a photocell and a variable d.c. power supply. The photocell consists of two metal plates sealed in an evacuated quartz container. Quartz is used for the photocell, in preference to glass, as it does not absorb higher frequencies of EM radiation (e.g. ultra violet) that are above the threshold frequencies of common metals. In addition, the photocell is evacuated to prevent the photoemitted electrons colliding with air molecules. The variable d.c. supply is used to alter the voltage across the metal plates in the photocell. Measurements are taken from the ammeter and voltmeter for different supply voltages.

Photoemission

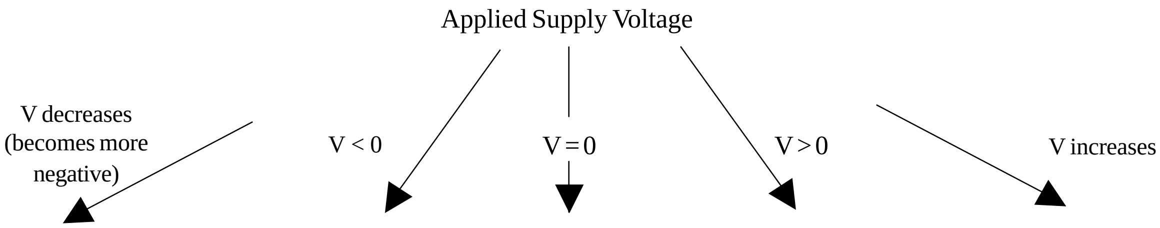
Electromagnetic radiation of a known frequency is directed on to one of the photocell plates.

- If it is above the threshold **frequency** (i.e. if it has sufficient energy), then photoelectrons are emitted.
- The number of photoelectrons emitted per second depends on the number of photons arriving per second, i.e. the Power, or **Intensity**, of the EM radiation.

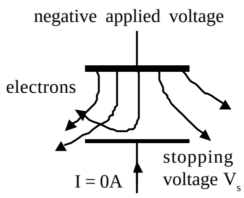
Electron flow in the circuit

What happens to the electrons after they are emitted from the plate depends on the potential (i.e voltage) of the metal plates in the photocell. This is controlled by the applied voltage (V) from the variable d.c. power supply. Consider the following cases, starting with V = 0:

Effect of Supply Voltage on photocell current.

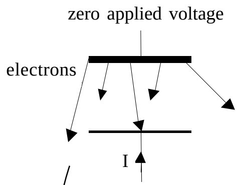


V decreases (becomes more negative): The current continues to drop, until even the most energetic electrons (with KE_{max}) are unable to overcome the potential energy of the electric field. At this point, no electrons will bridge the gap between the plates and so the current falls to zero. *The reverse potential required to reduce the current to zero is termed the* **stopping voltage** V_S.



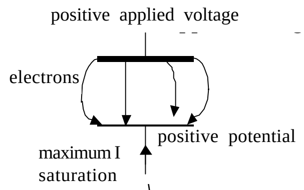
V < 0: A *reverse potential* gives the receiving plate a negative potential (relative to the battery). Thus, some electrons are **repelled** from the plate, leading to a reduced current flowing through the circuit

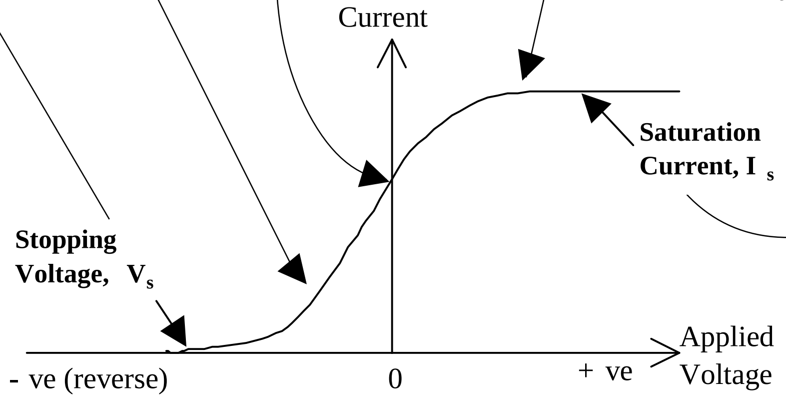
V = 0: Some electrons hit the other metal plate, thus completing the electrical circuit and allowing a current to flow.



V > 0: The 'receiving plate' is given a *positive potential* by the battery, **attracting** more of the photoelectrons towards the plate. Thus the current in the circuit increases.

V increases: The current will continue to rise as the potential of the plate is made more positive, until all the emitted photoelectrons are attracted and captured. At this point the current reaches its maximum value, called the **saturation current, I_s.**





↑ EM Energy ⇒ ↑V_S **↑ EM Intensity ⇒ ↑I_S**

Saturation Current and Stopping Voltage

The **saturation current** depends only on the number of photons arriving per second (i.e **Intensity**), as each photon releases one electron (assuming 100% efficiency).

The **stopping voltage** depends only on the energy of the photoelectrons, which is determined by the photon energy (i.e. **frequency**) for a given metal.

In Summary: **stopping voltage and saturation current**

Photoelectric Effect	Depends on **electron**	Alter by changing **photon**
Stopping Voltage, V_s	*Energy*	*Frequency (Photon Energy)*
Saturation Current, I_s	*number emitted per second*	*Intensity (number of photons/sec)*

Dependence of Stopping Voltage on Photon Frequency

At the stopping voltage, the most energetic electrons just cannot overcome the potential energy of the electric field between the plates. Thus;

Max. Electron energy= Electrical Potential Energy

$$KE_{max} = eV_s$$

Substituting this in Einstein's Equation gives:

$$hf - W = eV_s;$$

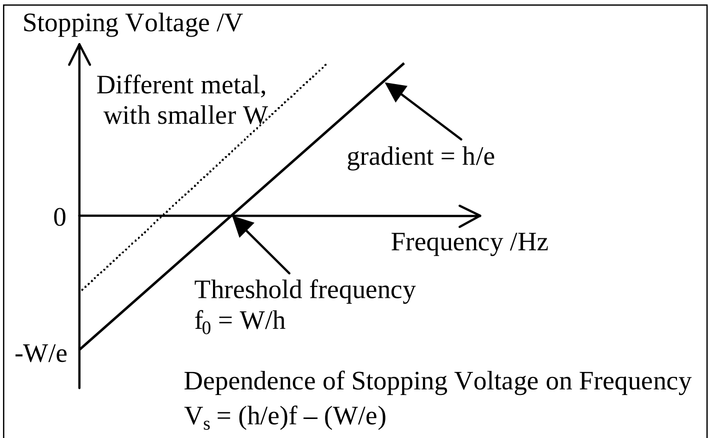
Graphically, this is represented by a straight line if V_s is plotted against f

Dependence of stopping voltage on EM frequency

$$V_s = (h/e)f - (W/e)$$

V_s : stopping voltage (V)
h : Plank's constant (6.63×10^{-34}) Js
e : charge on an electron (1.6×10^{-19} C)
f : EM radiation frequency (Hz)
W : metal work function (J)

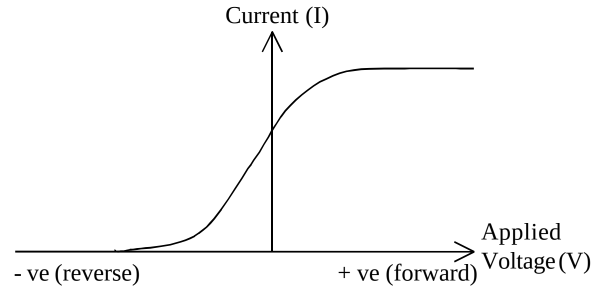
Fig. 4. Dependence of Stopping Voltage on EM frequency



- The gradient is (h/e) regardless of the metal being tested. It can be used to experimentally determine a value for Plank's constant (h) if the charge on an electron (e) is known.
- The y-intercept is (-W/e). It depends on the work function of the metal plate used and is always negative. Thus different metals will have different y-intercepts (work functions), but always the same gradient.
- The frequency axis intercept is equal to the threshold frequency (f_0 = W/h).
- A graph of maximum electron kinetic energy against frequency has exactly the same form, but the gradient becomes h and the y-intercept

Typical Exam Question

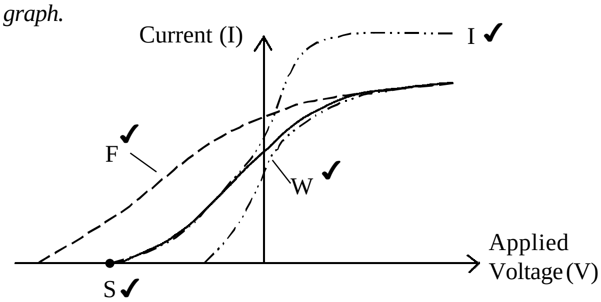
A photoelectric cell is illuminated. The graph below shows how the current (I) through the cell varies with the applied voltage (V) across it.



(a) Why does a current flow for positive voltages? [3]
(b) Why does the current reach a constant value for large positive voltages? [2]
(c) Why does the current reach 0A for a large negative voltage? [2]
(d) Add the following to the graph:
(i) A point, labelled S, to show the stopping potential [1]
(ii) A curve, labelled F, showing what you would expect if only the frequency of the light illuminating the cell were increased. [1]
(iii) A curve, labelled I, showing what you would expect if only the intensity of the light illuminating the cell were increased. [1]
(iv) A curve, labelled W, showing what you would expect if only a metal of a slightly larger work function were used. [1]

Answer

(a) Electrons are released from one of the metal plates in the photocell by the photoelectric effect. ✓ *They are then attracted towards the other metal plate which is positively charged by the battery.* ✓ *Therefore the circuit is completed and a current flows.* ✓
(b) The current is limited by the number of photoelectrons released each second. ✓ *Increasing the applied voltage varies how many of the photoelectrons are collected by the positively charged metal plate in the photocell. At large enough potentials, all the electrons are collected* ✓ *and so the current cannot increase further.*
(c) When the collecting plate is negatively charged, it repels the photoelectrons. At large enough voltages, even the most energetic electrons are repelled. ✓ *Thus, no electrons 'bridge' the gap between the plates to complete the circuit and no current flows.* ✓
(d) See graph.



Typical Exam Question

Light of wavelength 420nm with a power of 10mW is incident on a metal. (Charge on an electron is 1.6×10^{-19} C).
(a) Calculate the energy of each photon [3]
(b) Calculate the number of photons arriving per second. [2]
(c) Assuming 50% of the incident photons release electrons, what current is produced? [3]

Answer

(a) $E = hf = hc/\lambda$ ✓ *$= (6.63\times10^{-34})(3\times10^{8})/(420\times10^{-9})$* ✓
$= 4.74\times10^{-19}$ J ✓
(b) Number/second = Power/ photon energy = $0.01/4.74\times10^{-19}$ ✓
$= 2.11\times10^{16}$ s^{-1} ✓
(c) Number of electrons emitted/sec = $0.5 \times 2.11\times10^{16} = 1.05\times10^{16} s^{-1}$ ✓
As each electron has charge e, the current produced = (N/s)e ✓
$= (1.05\times10^{16})(1.6\times10^{-19}) = 1.68$ mA ✓

Why does the Photoelectric Effect provide *evidence for particle nature of light*, rather than its wave nature?
Before studying this, make sure that you understand the differences between the wave and particle theories of light (see page 1). Also, remember that intensity is defined as the power per unit area. The observations made in the Photoelectric Effect are unable to be explained using the wave theory of light. Therefore, the particle theory was developed in order to explain what was observed. These are summarised in Table 1.

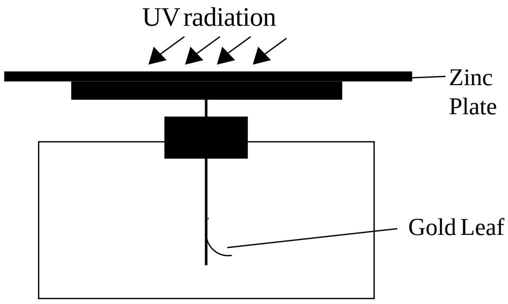
Table 1. Photoelectric Effect - Observations and Explanations

What is Observed?	What does Wave Theory predict?	What does Particle Theory predict?
For a given metal, electrons are only emitted above a certain **threshold frequency** of the EM radiation, irrespective of its intensity.	The greater the intensity (amplitude) of the EM radiation, the greater the energy that arrives at the metal surface. So a high enough intensity should cause electrons to be emitted regardless of the frequency. The frequency should have no effect here.	• Light consists of photons with energy **E = hf** • Minimum energy required to release an electron is the work function (W) • To release electrons, E>W • Threshold frequency (f_0) occurs when E=W; **hf_0 = W**
The maximum KE of the emitted electrons depends only on **frequency** of the EM radiation.	The greater the intensity (amplitude) of the EM radiation, the greater the energy that arrives at the metal surface. So, more energy is given to each electron. The frequency should have no effect here.	• For frequencies above the threshold frequency ($f>f_0$), Einstein's Equation gives: **KE_{max} = hf – W** • Thus KE_{max} depends on the frequency
The **number** of photoelectrons emitted per second depends only on **intensity** of the EM radiation, for a single frequency.	The intensity of the EM radiation relates to energy, not the number of waves arriving per second. So, the number of emitted electrons depends on the number of waves arriving (i.e. the frequency) and not the wave energy (i.e. intensity). The intensity should have no effect here.	• The number of photons arriving per second depends on the intensity of the EM radiation • One photon can release one electron (assuming 100% efficiency) • Thus the number of electrons emitted per second depends on intensity
Low intensity EM radiation (above the threshold frequency) results in **immediate emission** of electrons.	Low intensity EM radiation has low energy. So it will take some time before enough energy builds up on the metal surface to free one electron.	• Intensity only relates to how many photons arrive per second, so few arrive per second for low intensities • But, each photon has enough energy to release an electron ($f>f_0$) so immediate electron emission occurs

Exam Workshop

This is a typical poor student's answer to an exam question. The comments explain what is wrong with the answers and how they can be improved. The examiner's answer is given below.

The photoelectric effect can be simply demonstrated by using an electroscope which is negatively charged, as shown in the diagram.



When ultra violet light is shone on the metal plate, the gold leaf gradually returns to the vertical. However, when red light alone is used, the foil remains displaced and does not return to the vertical.

(a) **Explain why the gold leaf is initially displaced from the vertical [2]**
It's Repelled ✓

- Needs more detail for 2 marks

(b) **Explain why the gold leaf gradually returns to the vertical when UV is shone on to the metal plate? [3]**
It becomes less negatively charged. ✓
The light takes the extra charge away. ✗

- Not enough detail to gain 3 marks.
- Be precise - light does not 'take the extra charge away'.

(c) **Use the particle theory of light to explain why red light does not have the same effect as ultra-violet light. [5]**
Red light doesn't free electrons as it doesn't have as much energy as UV ✓. *Energy comes in lumps that depends on the frequency* ✓.

- Did not mention what the particle theory of light is (i.e. photons).
- Did not explain why electrons are released when UV light shines on the plate.
- Needs to be specific about red light being below the threshold frequency – too vague here to gain mark.

(d) **If the metal has a work function of 2.3 eV, calculate the maximum kinetic energy of the emitted electrons, in Joules, when light of 200 nm is used. [3]**
$KE_{max} = hf - W = h(200\times10^{-9})/(3.0\times10^{8}) - 2.3 = -2.3eV$ ✗

- Used λ not f in equation
- Calculated energy in eV, not Joules as asked

Examiner's Answers

(a) *Both the foil and rod are negatively charged.* ✓ *Like charges repel, so the foil moves away from the rod.* ✓

(b) *The UV light frees electrons so they can escape from the metal plate (the photoelectric effect)* ✓ *so the negative charge on the metal plate and rod decrease* ✓. *So, the repulsion between the foil and the rod decrease and the foil drops due to gravity.* ✓

(c) *Light consists of packets (quanta) called photons* ✓. *The energy of each photon depends on its frequency, by E = hf* ✓. *As red light has a lower frequency than UV light, it has less energy* ✓. *One photon gives its energy to one electron* ✓. *In order to release an electron, the photon energy must be greater than or equal to the metal work function: red light photons have less energy than this and so do not provide enough energy to free an electron* ✓.

(d) *Use Einstein's Equation:* $KE_{max} = hf - W$,
where $f = c/\lambda = (3.0\times10^{8})/(200\times10^{-9}) = 1.5\times10^{15}Hz$ ✓
and $W = 2.3eV \times (1.6\times10^{-19}C) = 3.7\times10^{-19}J$ ✓.
Substitute in the equation to get $KE_{max} = 6.2\times10^{-19}J$ ✓.

Qualitative (Concept) Test
Look in the text for the answers to the questions and use the hints!

(1) Explain **wave – particle duality**. Include an example in your answer.

(2) What is a **photon**?

(3) Describe the **photoelectric effect**.

(4) How does the photoelectric effect provide evidence for the **particle nature** of electromagnetic radiation?

(5) What is the **work function** of a metal? Explain how this relates to **the threshold frequency** for the metal. (***Hint:*** *See "How is the Photoelectric Effect Explained"*)

(6) What is **Einstein's equation**? What does it describe? Explain all symbols used.

(7) What is the **stopping voltage**? How can it be decreased for a given metal?

(8) Describe an **experiment** to verify Einstein's photoelectric effect. Explain how values for the metal work function and Plank's constant can be determined form this experiment.

(9) If visible light is incident on zinc, no electrons are emitted, irrespective of how intense the light is. Explain this observation.

(10) Monochromatic electromagnetic radiation is incident on a metal surface, resulting in the emission of electrons.
Describe the effect on the maximum kinetic energy and on the number of electrons ejected per second for the following changes:
(a) increasing/decreasing the frequency of the incident radiation
(b) increasing/decreasing the intensity of the incident radiation
(c) changing the metal for one with a greater/smaller work function
(***Hint:*** *See "saturation current and stopping voltage"*)

Quantitative (Calculation) Test
Time for test:35 mins Total Marks:34

Data: Plank's Constant $h = 6.63 \times 10^{-34}$ Js
electron mass $m_e = 9.1 \times 10^{-31}$ kg
electron charge $e = 1.6 \times 10^{-19}$ C
Speed of EM in a vacuum $c = 3.0 \times 10^{8}$ ms^{-1}

(1) Calculate the energy of a photon of:
(i) wavelength 1.2 μm; (ii) frequency 2.0×10^{18}Hz.

(2) Calculate the frequency and wavelength of a photon of energy 60 MeV.

(3) Light of frequency 3×10^{15}Hz is incident on a metal of work function 5.0×10^{-19} J causing electrons to be emitted with maximum kinetic energy 1.46×10^{-18}J.
(a) Calculate a value for Plank's constant.
(b) Calculate the threshold frequency for this metal
(c) Calculate the minimum frequency of radiation required to achieve electron kinetic energies above 4×10^{-18}J.

(4) Laser light of 490nm with a power of 7.5×10^{-2}W is incident on a metal of work function 2 eV.
(a) Calculate the number of photons emitted per second.
(b) If 10% of the photons result in an electron being emitted, what current is produced?
(c) Calculate the energy of each photon in electron volts.
(d) (i) Calculate the maximum kinetic energy of the photoelectrons in electron volts. (ii) What speed would these electrons have (neglect relativistic effects)?
(e) If the incident power is doubled, what will be the effect on (i) the maximum energy of the emitted electrons; and (ii) the number of electrons emitted per second
(f) Calculate the threshold wavelength for the metal.
(g) Will light of 1.5×10^{15}Hz result in the emission of electrons?

Further Reading
- 'The Quantum World' by J.C. Polkinghorne published by Pelican ISBN 0140226532. A good overview of the subject with minimal mathematical content.

Related Factsheets
- Quantum Nature of Particles
- Basic Wave Properties
- Applied Wave Properties

Quantitative Test Answers

(1) (i) $E = hf = h(c/\lambda)$ ✓ $= (6.63\times10^{-34}\text{Js})(3\times10^{8}\text{ms}^{-1})/(1.2\times10^{-6}\text{m})$ ✓
$=$ **1.66×10^{-19}J** ✓;
(ii) $E = hf = (6.63\times10^{-34}\text{Js})(2.0\times10^{18}\text{Hz}) =$ **3.14J** ✓

(2) $f = E/h = (60\times10^{6})(1.6\times10^{-19}\text{ J})/(6.63\times10^{-34}\text{Js})$ ✓ $=$ **1.45×10^{22}Hz** ✓;
$\lambda = c/f = (3.0\times10^{8}\text{ms}^{-1})/(1.45\times10^{22}\text{Hz})$ ✓ $=$ **2.07×10^{-14}m** ✓

(3) (a) by Einstein's Equation $h = W/f + KE_{max}/f$ ✓
$= (5.0\times10^{-19}\text{J})/(3\times10^{15}\text{Hz}) + (1.46\times10^{-18}\text{J})/(3\times10^{15}\text{Hz})$ ✓
$=$ **6.5×10^{-34}Js** ✓.
(b) threshold frequency, $f_o = W/h = (5.0\text{x}10^{-19}\text{J})/(6.63\text{x}10^{-34}\text{Js})$ ✓
$=$ **7.5×10^{14}Hz** ✓
(c) by Einstein's Equation; $f = W/h + KE_{max}/h$ ✓ $= f_o + KE_{max}/h$ ✓
$= 7.5\times10^{14} + (1.46\times10^{-18}\text{J})/(6.63\times10^{-34}\text{Js}) =$ **6.8×10^{15}Hz** ✓

(4) (a) # photons/sec = (emitted power)/(energy per photon)
$=$ Power/(hc/λ) ✓
$=\{(7.5\times10^{-2}\text{W})(490\times10^{-9}\text{m})\}/\{(6.63\times10^{-34}\text{Js})(3.0\times10^{8}\text{ms}^{-1})\}$ ✓
$=$ **1.85×10^{17} s^{-1}** ✓
(b) Current = (# electrons emitted/sec) × (charge on each electron)
$= 0.1\times$(# photons/sec) x e ✓
$= 0.1(1.85\times10^{17}\text{ s}^{-1})(1.6\times10^{-19}\text{C}) =$ **3.0mA** ✓
(c) $E = hf = hc/\lambda = (6.63\times10^{-34}\text{Js})(3.0\times10^{8}\text{ms}^{-1})/(490\times10^{-9}\text{m})$
$= 4.06\times10^{-19}$J ✓;
As 1eV= $1.6\text{x}10^{-19}$J, this corresponds to $(4.06\times10^{-19}\text{J})/(1.6\times10^{-19}\text{J})$
$=$ **2.54eV** ✓
(d) (i) by Einstein's Equation; $KE_{max} = hf - W = E - W$ ✓
$= 2.54 - 2.0 =$ **0.54eV** ✓
(ii) $KE = ½ m_e v^2$; $v = \sqrt{\{2(KE_{max})/m_e\}}$
$= \sqrt{\{2(0.54\times1.6\times10^{-19}\text{J})/(9.1\times10^{-31}\text{kg})\}}$ ✓ $=$ **4.3×10^{5}ms^{-1}** ✓
(e) (i) **no change** as energy is independent of Intensity; ✓
(ii) **doubled, i.e. 3.7×10^{16}s^{-1}** ✓ as doubling the intensity doubles the number of electrons arriving per second and thus also of the electrons emitted per second.
(f) $W = hf_o = h(c/\lambda_o)$; $\lambda_o = hc/W$ ✓
$= (6.63\times10^{-34}\text{Js})(3\times10^{8}\text{ms}^{-1})/(2\times1.6\times10^{-19}\text{J})$ ✓ $=$ **155nm** ✓
(g) **No,**✓
as this frequency is below the threshold frequency calculated from part (f);
$f_o = c/\lambda_o = (3.0\times10^{8}\text{ms}^{-1})/(155\times10^{-9}\text{m}) = 1.9\times10^{15}$Hz ✓

Acknowledgements:
This Physics Factsheet was researched and written by ***Thane Gilmour***
The Series editor for physics is Thane Gilmour
The Curriculum Press,Unit 305B, The Big Peg,120 Vyse Street, Birmingham, B18 6NF

ISSN 1351-5136