Antonine Physics A2

Topic 14E Electronics

How to Use this Book

How to use these pages:

- This book intended to complement the work you do with a teacher, not to replace the teacher.
- Read the book along with your notes.
- If you get stuck, ask your teacher for help.
- The best way to succeed in Physics is to practise the questions.

There are many other resources available to help you to progress:

- Web-based resources, many of which are free.
- Your friends on your course.
- Your teacher.
- Books in the library.

This Topic covers Electronics, which has made a welcome return to the A-level Syllabus. As well as being an interesting subject in its own right, knowledge and skills in Electronics is an important tool in the study of many parts of Physics at university level.

We will look at the use of electronics in analogue systems, which are used in audio applications and simple motor control. We will also look at digital systems that are at the heart of computers. We also consider how electronics is used in data communication.

You need to make sure that you are happy with electrical circuits and properties.

This is quite a long topic. Work through it slowly and carefully.

Topic 14E

Option E Electronics

1. Discrete Semi-conductor Devices

Tutorial 14 E.01 MOSFET

AQA Syllabus

Contents

14E.011 Semi-Conduction

While the models of semi-conduction are NOT on the syllabus, it is worthwhile for us to think about it so that we can understand the action of semi-conductor components. These models are simple. More detailed models are available but need a good understanding of quantum theory and are beyond the level we need.

All transistors depend on **semiconductors**. The most common semiconductor is **silicon**. The resistivity of pure silicon is high at 0.1 Ω m (compared with 49×10^{-8} Ω m for constantan), but it can be reduced by the addition of certain impurities such as **aluminium** or **phosphorous**. The addition of impurities is called **doping**.

Silicon is from **Group 4** in the periodic table, so has 4 outer shell electrons. The atoms share electrons in covalent bonds with 4 other atoms. If we put in a **Group 5** atom like phosphorus, it will share electrons with 4 silicon atoms, but since it has 5 atoms in the outer shell, there is a **free**, unpaired electron. This electron involved in conduction. The idea is modelled in the picture below (*Figure 1*).

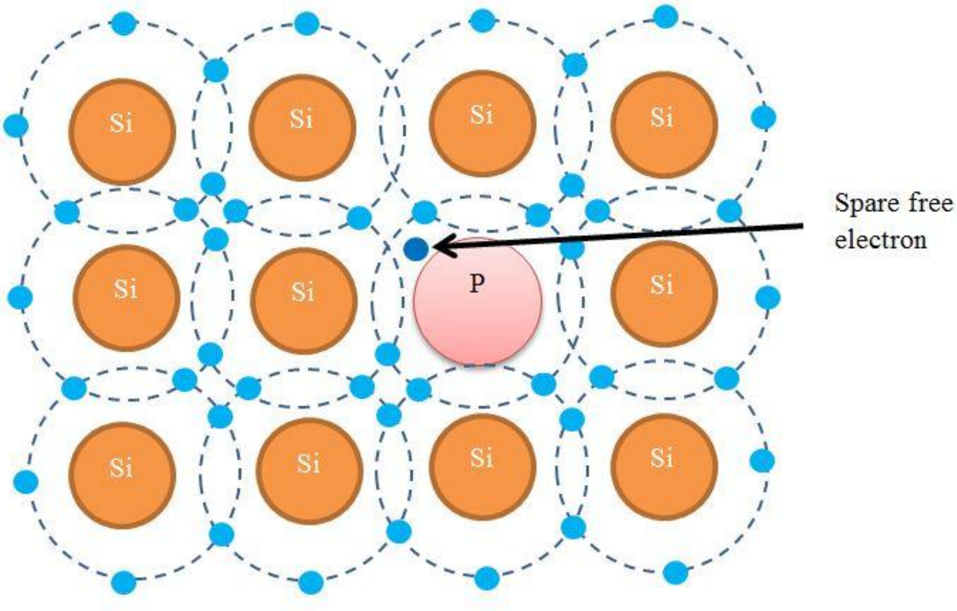


Figure 1 Phosphorus atom in a lattice of silicon atoms

This free electron is available for conduction. Since it is negatively charged, the semiconductor doped with a Group 5 metal is called an **n-type** semiconductor. Arsenic and Antimony are also used as **dopants**.

If the silicon is doped with a Group 3 metal like aluminium, there is an electron **vacancy**. This is because Group 3 elements have 3 outer shell electrons. This is shown below (*Figure 2*):

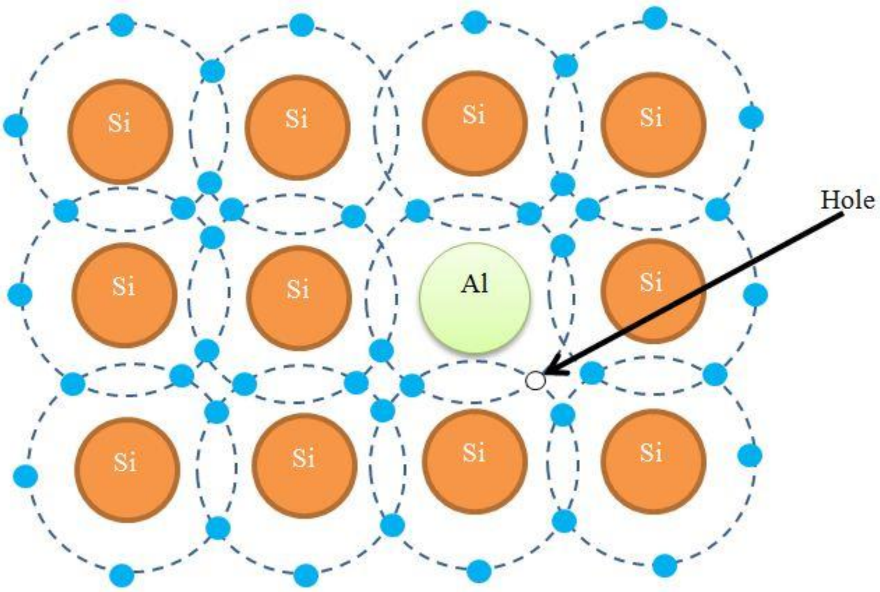


Figure 2 Aluminium atom in a lattice of silicon atom giving a hole

The vacancy or **hole** be replaced by an electron, and the empty space can thus **move about**. Since it's a deficiency of an electron, it results in a **positive charge**. Therefore, the semiconductor doped with a Group 3 metal is called a **p-type** material. Holes move from positive to negative as a conventional current.

14E.012 Conduction Band Model

The conduction mechanism can be explained using the **conduction band** model. Remember that electrons in the quantum world occupy energy levels. These are like the rungs of a ladder. This states that there are bands of energy levels to which electrons can be raised. There are lots of rungs close together in the energy ladder. Normally the electrons occupy the valence band, i.e. their place in the outer shell. If the right amounts of energy is applied, the electrons can be raised into the conduction bands.

In metallic conduction, the conduction band and the valence band overlap (*Figure 3*):

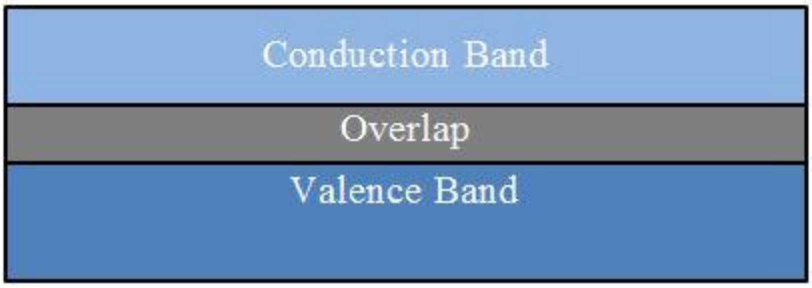


Figure 3 Overlap between the valence and the conduction bands

Since the two bands overlap, many electrons can move into the conduction band, making metals good conductors.

In semi-conduction, the conduction band and valence band are separated by an energy gap. For silicon this is 1.09 eV. For germanium, the gap is 0.72 eV. At 0 K no electrons can cross it (*Figure 4*):

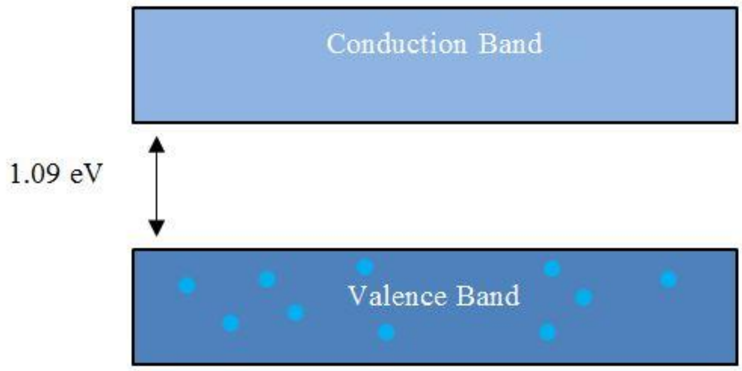


Figure 4 Energy gap between the valence and the conduction band in Si

However, at high temperatures like 300 K, some electrons can cross the energy gap (*Figure 5*).

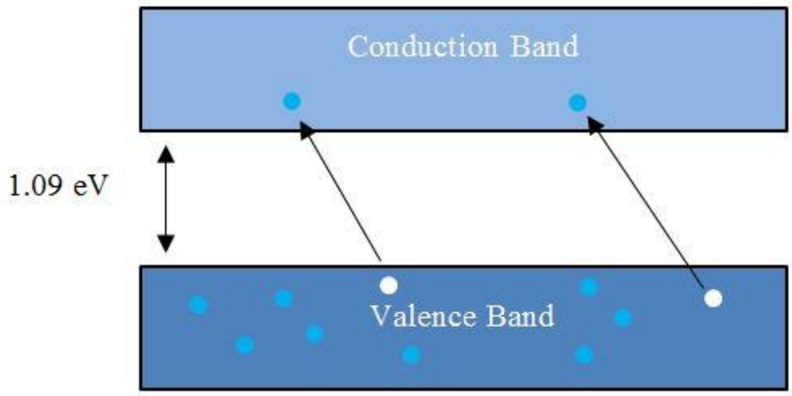


Figure 5 Some electrons can cross the energy gap at higher temperatures

If the semiconductor is doped, then more electrons can cross the gap.

14E.013 Depletion Zone

If a piece of n-type material is placed next to a piece of p-type material, a **p-n junction**, the free electrons are attracted by the holes to fill them. This leaves a region, the **depletion zone**, where there are no spare charge carriers.

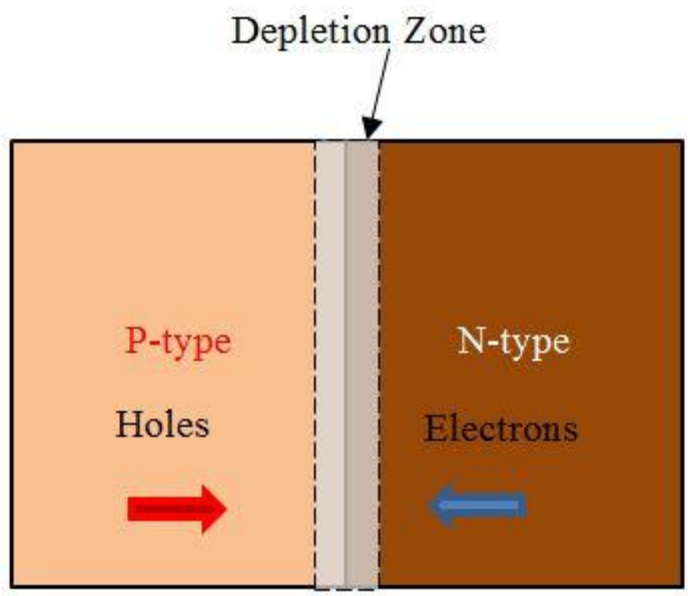


Figure 6 The depletion zone.

The depletion zone acts as an **insulator**. Let us look at this more closely (*Figure 7*). First, we will place a piece of n-type material next to the p-type material.

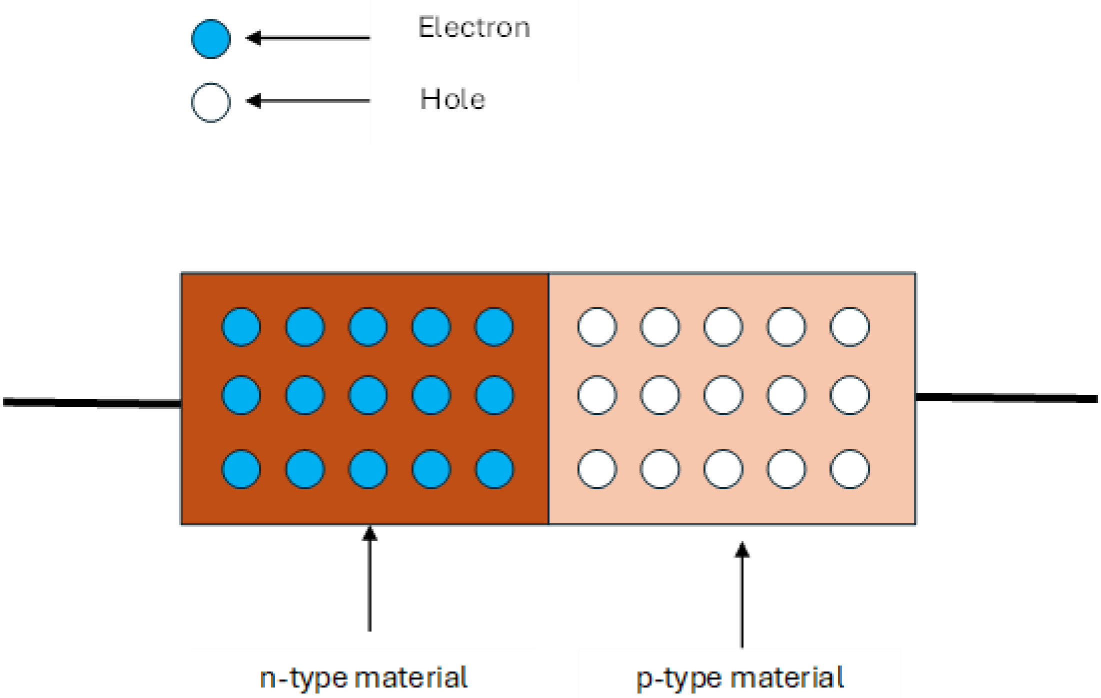


Figure 7 Placing n-type material next to p-type material

We find that some of the electrons next to the boundary migrate across to fill in holes, leaving holes where they were (*Figure 8*).

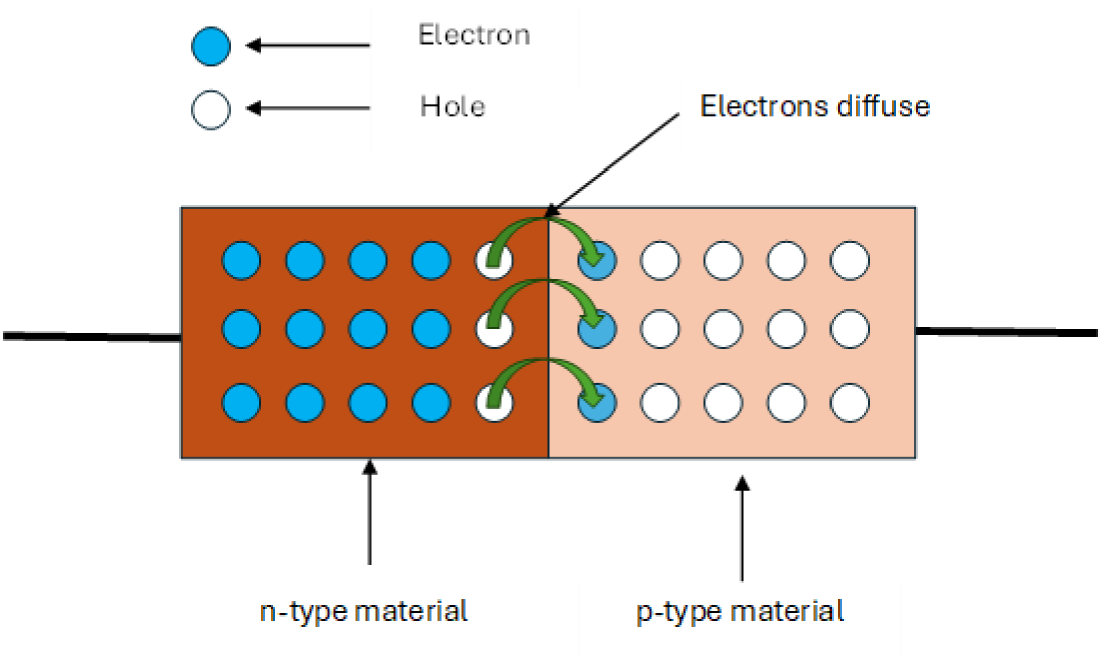


Figure 8 Electrons diffuse across the boundary

This results in an **electric field** being set up across the boundary (*Figure 9*).

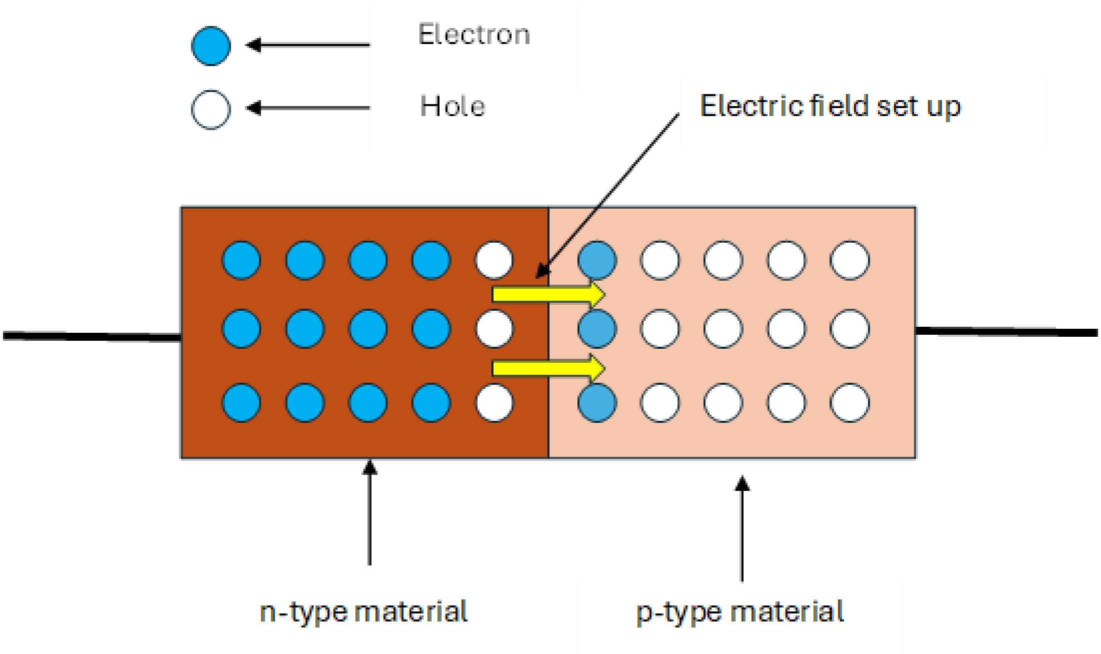


Figure 9 Electric field set up

The result of this electric field is that there is a **depletion layer** that acts as an insulating zone (*Figure 10*).

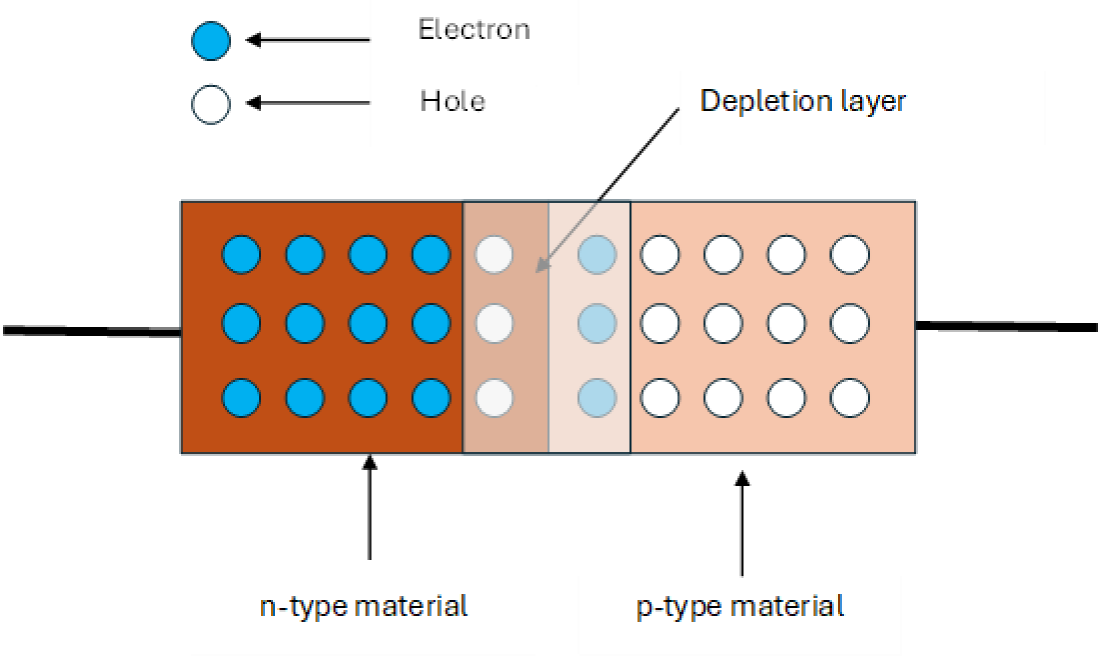


Figure 10 Depletion layer is formed

If a positive voltage that's sufficient to overcome the Coulomb repulsion within the depletion zone, electrons will cross to the p-type material. The depletion zone will disappear. This is **forward bias** (*Figure 11*).

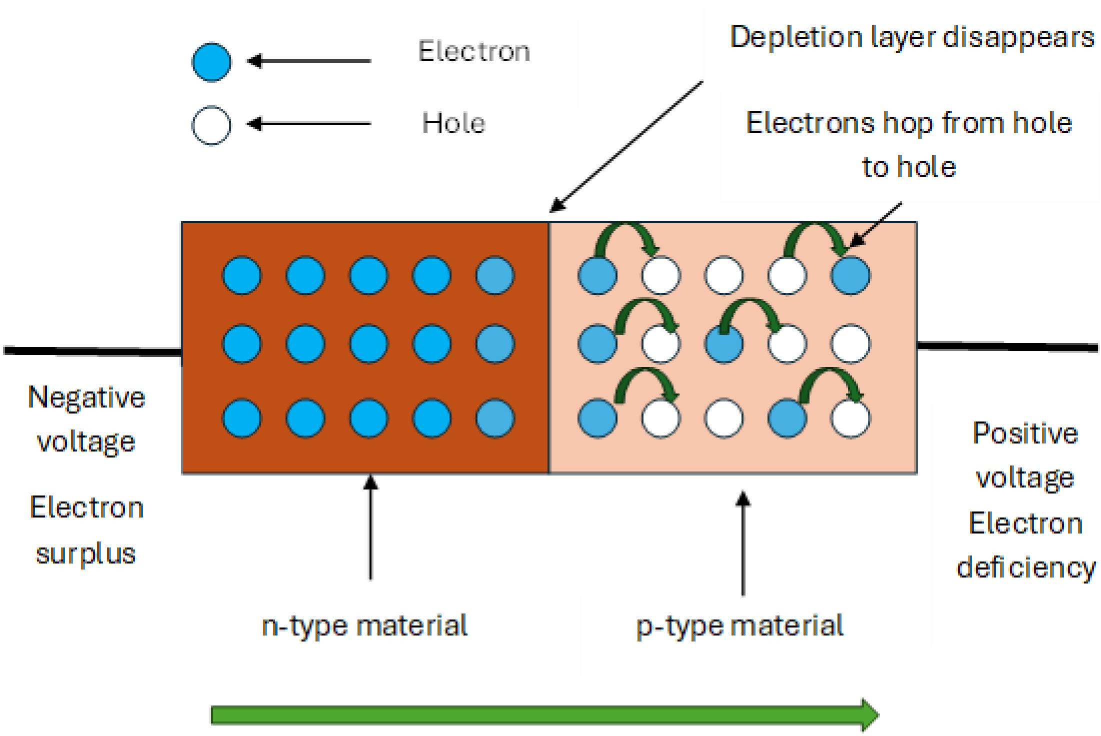


Figure 11 Foward bias

The current can only flow once the potential difference between the positive and negative terminals is greater than 0.6 V. The electrons can diffuse from hole to hole. The diode characteristic graph is like this (*Figure 12*).

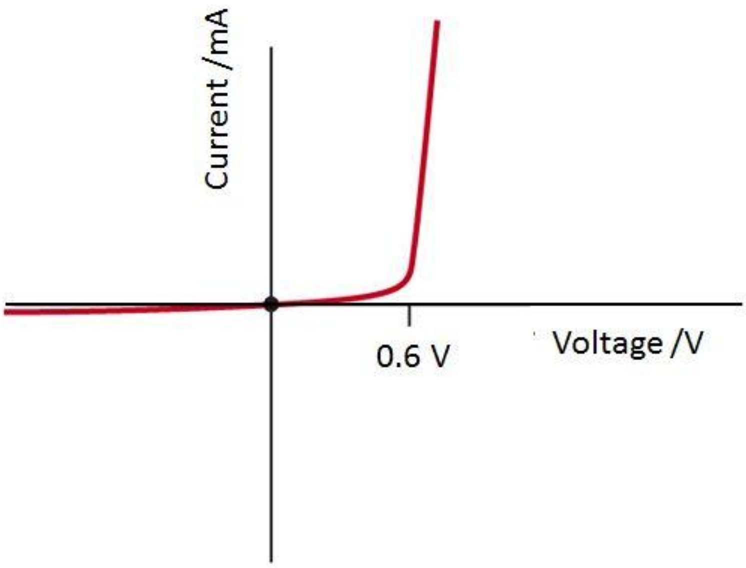


Figure 12 Diode characteristic graph

Note that the holes are **electron vacancies**. That means that there **is no electron present** in the space where we would normally expect to find an electron. The Al^{3+} ion, which gives arise to the hole, itself does not move at all.

If the p-type material is connected to the positive of a battery, electrons from the n-type material will move towards the p-type material that has a greater deficiency of electrons. The depletion zone gets narrower. This is because the overall value of the **electric field** within the junction is reduced. See *Figure 13*.

Electric fields are discussed in detail in Topic 9. All we really need to understand is that the electric field is a vector quantity. It is defined as **force per unit charge**. A positive field goes from positive to negative.

$$E = \frac{F}{Q}$$ .... *Equation 1*

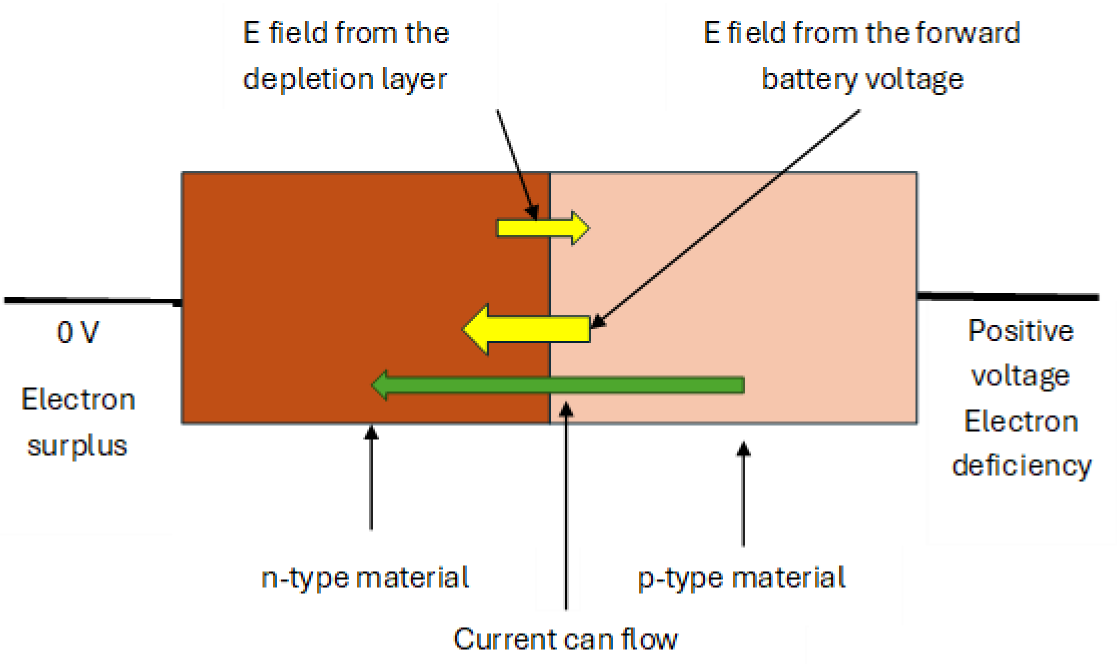


Figure 13 Conduction across a p-n junction

Above a certain voltage (0.6 V), the electric field from the battery is bigger (and in the opposite direction to) than the electric field set up by the depletion zone. So, we can say:

$$E_{\text{total}} = E_{\text{battery}} - E_{\text{depletion}}$$ .... *Equation 2*

In other words, the electrons start to have a sufficient energy to jump the gap, and the junction starts to conduct. A very small increase in the voltage results in a very big increase in the current. The junction is said to be **forward biased**.

Before the junction reaches 0.6 V, there is a very small leakage current. This is because the electrons have **quantum properties** and there is a small probability that they have sufficient energy to leap the boundary.

If the positive voltage is applied to the n-type material (**reverse bias**), the depletion zone will increase, preventing conduction (*Figure 14*).

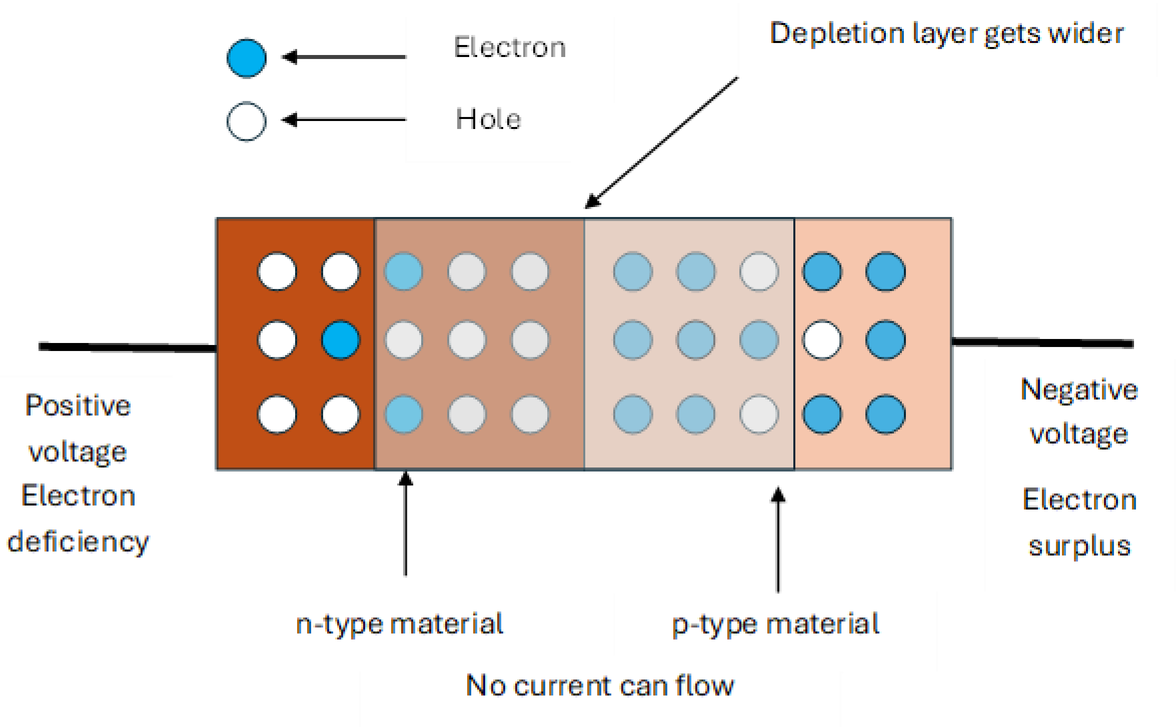


Figure 14 Reverse bias

If the polarity of the battery is reversed, i.e., the positive voltage is applied to the n-type material, electrons are attracted to the positive terminal, and the holes (which act as positive charge carriers) move to the negative terminal. The depletion layer gets wider. The junction is said to be **reverse biased** and will not conduct.

The p-n junction is at the heart of a **diode**, a one-way electrical "valve".

A reverse biased diode has a **breakdown** voltage, above which conduction will occur. This may well destroy the diode. However, there are diodes that are designed to operate in reverse bias, for example, the **Zener diode**. Zener diodes have a typical breakdown voltage of 5.6 V and are used as simple **voltage regulators**.

14E.014 Bipolar Transistors

The most familiar transistor is the **bipolar junction transistor** (BJT) which is NOT on the syllabus, but it is worth looking at it briefly. You will certainly meet it if you do further study of electronics. Here is a typical BJT circuit (*Figure 15*).

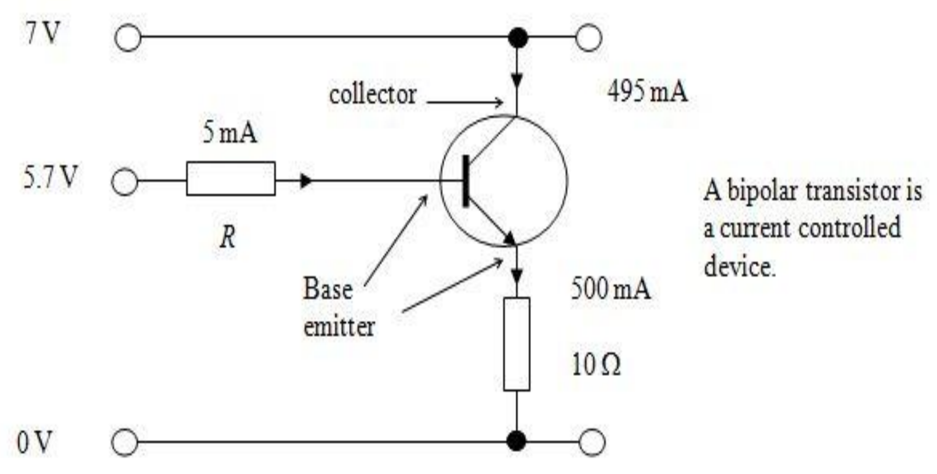


Figure 15 A typical bipolar transistor in circuit

Notice here that

- The current coming out of the emitter (I_e) = 500 mA.
- The current going through the base (I_{be}) = 5 mA.
- The current coming into the collector (I_{ce}) = 495 mA.

We can measure the **current gain** of a transistor, which is the ratio of the collector-emitter current to the base-emitter current.

$$\text{Gain} = I_{ce} \div I_{be}$$ .... *Equation 3*

In catalogues, you may see gain as h_{fe}.

Figure 16 is a typical circuit involving a transistor. It is a light operated switch that uses a relay to turn on a mains-powered bulb.

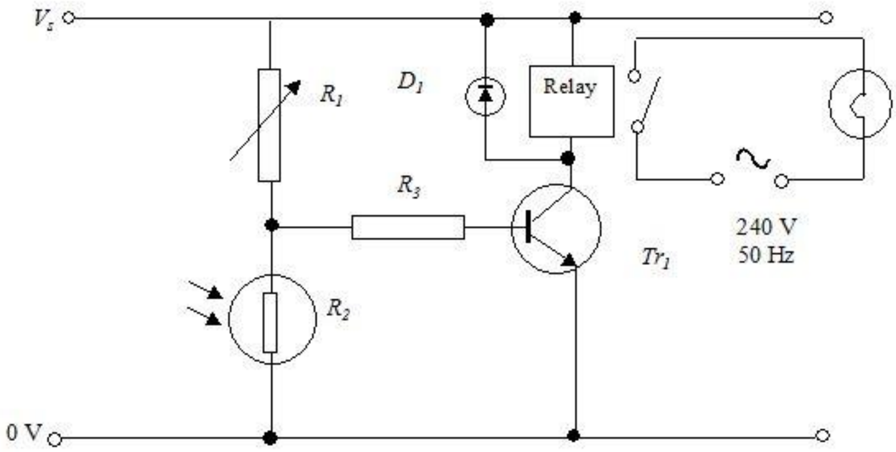


Figure 16 A BJT being used to operate a relay to turn on a mains lamp

The voltage drop from the **base** to the **emitter** is about 0.6 V. This the voltage drop across a typical diode junction. The voltage drop between the **collector,** and the **emitter** is about 1.2 V, as the current has to pass 2 diode junctions.

A bipolar transistor consists of two p-n junctions back to back like this (*Figure 17*):

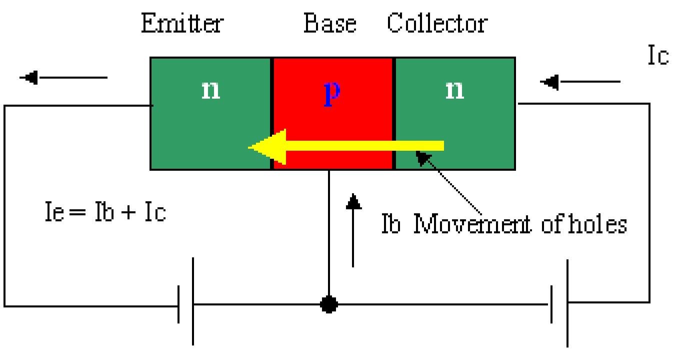


Figure 17 Structure of a bipolar transistor

The arrows represent **conventional** current flowing from positive to negative. Semiconductors are doped so that some have an **excess of electrons**, **n-type material**, while others have a deficiency **of electrons** called **holes**. These are found in **p-type** materials.

Transistors conduct when the base-emitter voltage is about **0.6 V**. Transistors are current controlled devices. The bigger the base emitter current, the bigger the collector emitter current.

The transistor can be used as an amplifier or a solid state switch.

14E.015 Characteristic Graphs for the Bipolar Transistor

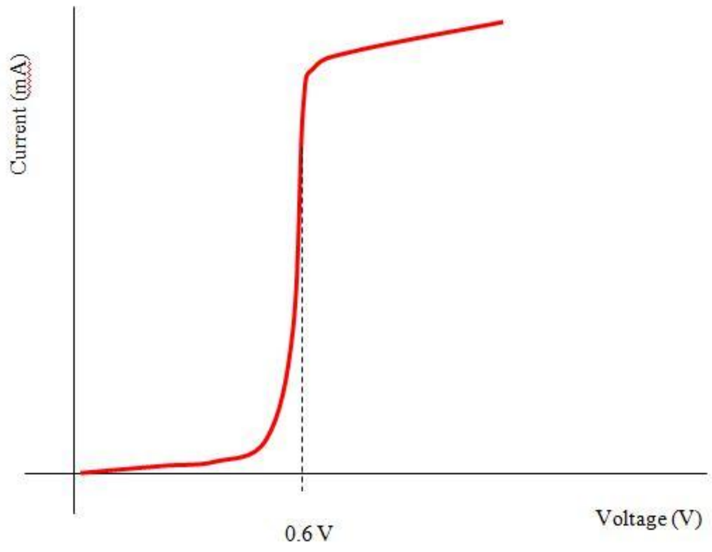


Figure 18 Transconductance

The voltage current characteristic sometimes called the **transconductance** (*Figure 18*).

We can measure the collector-emitter current and plot it against the base-emitter current (*Figure 19*):

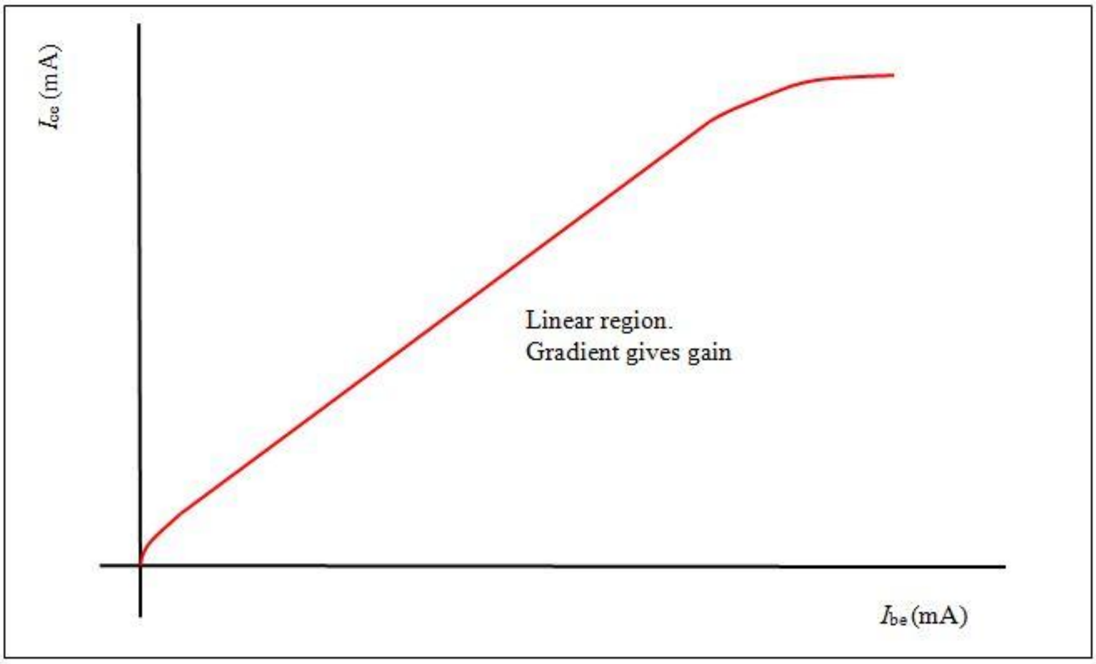


Figure 19 Graph showing the collector-emitter current against the base emitter current

This graph shows us that the collector-emitter current rises **linearly** with the base-emitter current. The **gradient** gives us the **current gain**. The bipolar junction transistor is a **current-controlled** device.

As the base-emitter current increases, the collector-emitter current increases to a **maximum** value which cannot be exceeded. The transistor is **saturated**. The transistor risks being overheated due to the heating effect of the current. **Heatsinks** are provided to reduce the operating temperature of the transistor. However, an overheated transistor can go into **thermal runaway** and can be destroyed.

14E.016 MOSFET

The most familiar transistor is the **bipolar junction transistor** (BJT) which we have looked at in 14E.014 and 14E.015.

The **Metal Oxide Semiconductor Field Effect Transistor** (MOSFET) is quite different to the BJT. Here is the symbol for the MOSFET (*Figure 20*):

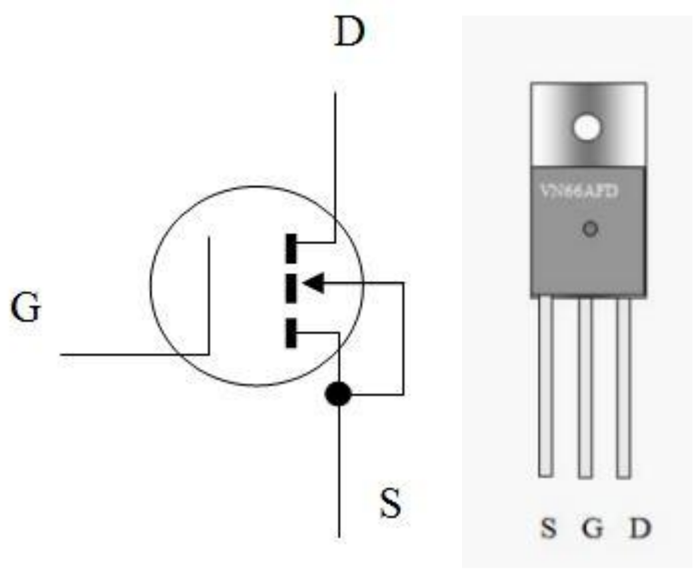


Figure 20 Metal Oxide semiconductor field effect transistor

The terminals are:

- **Gate** - a voltage at the gate turns on the MOSFET.
- **Source** - this is where the electrons come in. The source is the negative terminal.
- **Drain** - this is where the electrons leave. This is connected to the positive.

In some MOSFETs there is a fourth terminal, the **body** (B) which is connected to the 0 V line. In most MOSFETs it's connected to the source.

As electronics uses **conventional** currents (flowing from positive to negative), the current flows from **drain to source**.

The general characteristics for a MOSFET are:

- The input (**gate**) resistance is very high, about $10^{12}\ \Omega$.
- The **output resistance** is about the same as a bipolar transistor. The actual value depends on the type. For a signal MOSFET it would be in the range 10 to 50 kΩ, while in a power MOSFET it would be somewhat lower.
- The switch-on voltage is between 1.0 and 2.5 V.

The MOSFET is a **voltage-controlled** device. This means that a **voltage** alone is needed to turn on the MOSFET. The current needed is about 1 pA (1×10^{-12} A). The picture below shows the way a motor can be turned on by dipping two 4 mm plugs into a cup of coffee (*Figure 21*).

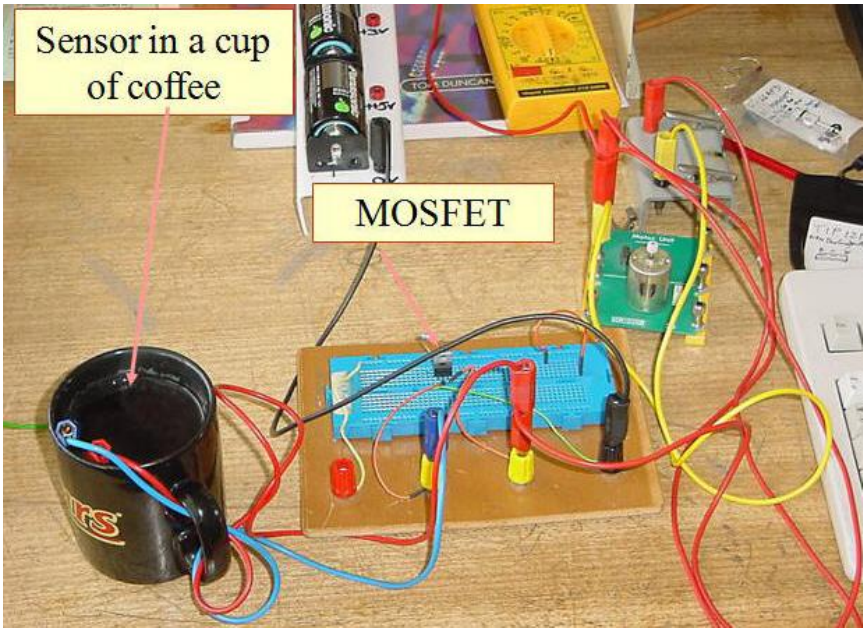


Figure 21 A MOSFET is controlled by a voltage, not a current

14E.017 How a MOSFET Works

The **n-channel enhancement mode** MOSFET is made up like this (*Figure 22*):

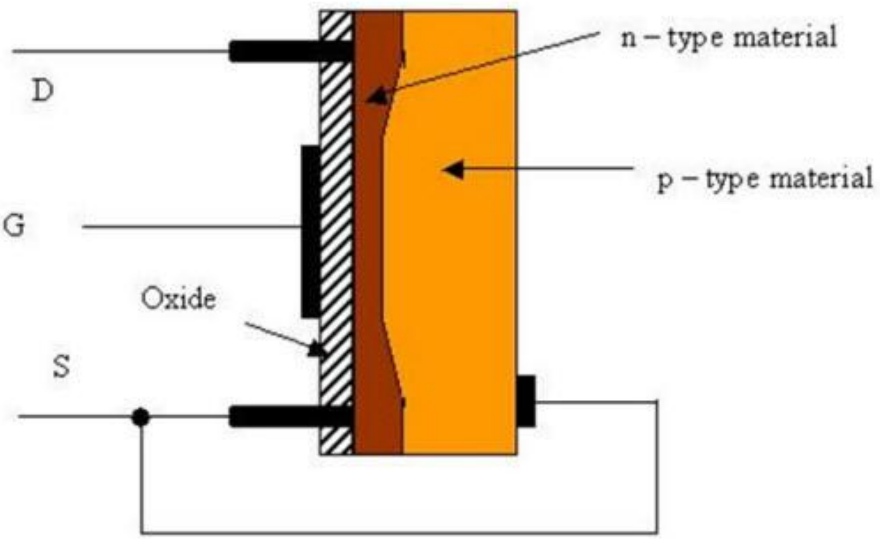


Figure 22 How a MOSFET is made up

The **oxide layer** is an insulator and ensures that there is NO direct connection between the **gate** and the **n-type** material. This is why the MOSFET is a voltage controlled device. Current cannot flow from the gate to the n-type material.

When the gate-source voltage is zero very low, there is a **depletion zone** like this(*Figure 23*):

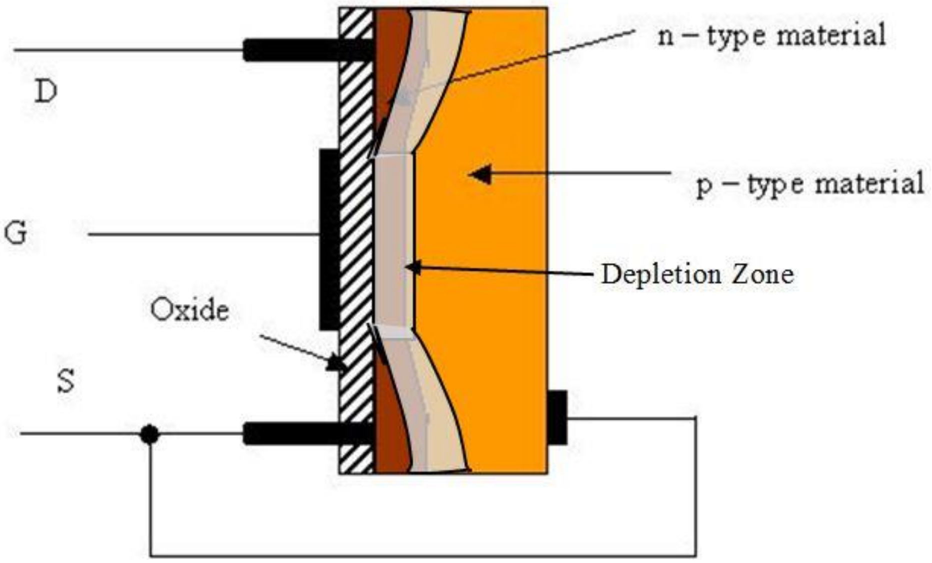


Figure 23 Depletion zone in a MOSFET

The **depletion zone** acts as an insulator. Therefore, no current flows.

The gate voltage makes an electric field across the region of n-material (called the **n-channel**). The positive voltage repels the holes and attracts electrons to the n-channel. The depletion layer is reduced, so that more electrons are available for the n-channel. Electrons can travel from the source to the drain. Remember that conventional current flows from the drain to the source (*Figure 24*).

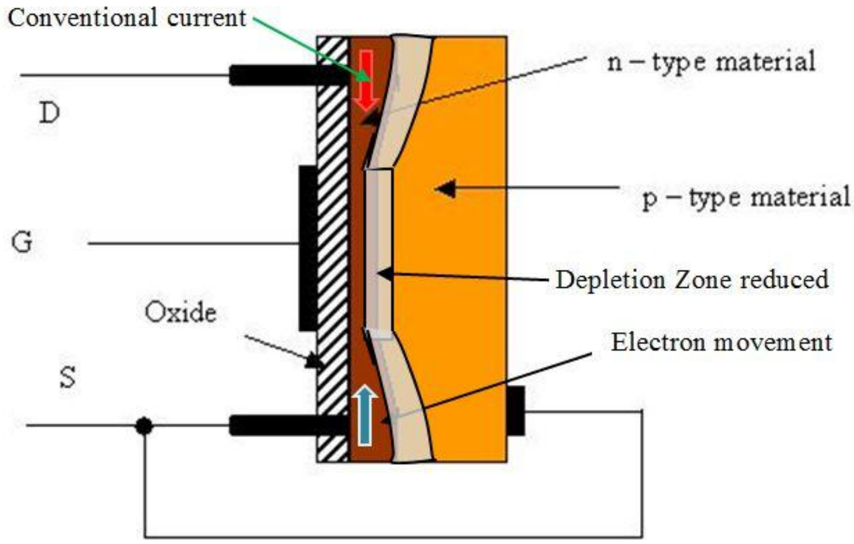


Figure 24 Depletion zone is reduced when the gate voltage is increased

The current in the picture above is less than the maximum current because the n-channel is narrower than it could be. If we increase the gate voltage a bit more, the depletion zone below the gate is reduced even further (*Figure 25*).

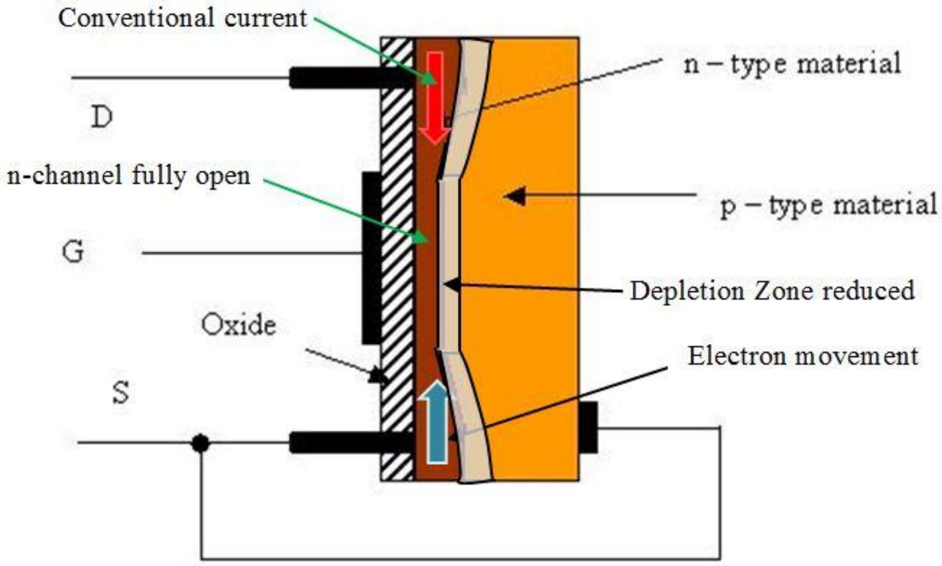


Figure 25 Reducing the depletion zone further by increasing the gate voltage

The n-channel is now fully open so that the MOSFET conducts the biggest current it can. Note that the depletion zone is still there. Current does not flow through the p-type material. The depletion zone can be thought of as like a rubber diaphragm in a water valve. The water can get past it when the valve is open but not round it.

The n-channel starts to open with a gate voltage of about 1.7 V. It is fully open with a gate voltage of about 2.5 V to 3.0 V. Once fully open the current cannot increase, unless the supply voltage is increased. The MOSFET is **saturated**.

This type of MOSFET is said to have an **n-channel enhancement mode**. It acts as a normally open switch (i.e. OFF). This is the one you need to know about.

There are three other types of MOSFET.

- **p-channel enhancement** type - the drain is negative compared to the source. It acts as a normally open switch.
- **n-channel depletion type** - It acts as a normally open switch (ON) and turns off with an increase in gate voltage.
- **p-channel depletion type**.

You are not expected to describe these types in the exam.

In the Exam

This description on MOSFET action may well form a six-mark essay question.

Think the points through in a logical way:

Gate voltage = 0.

Gate voltage just as the current flows.

Full current flow.

14 E.018 Characteristics of a MOSFET

The characteristics of a MOSFET can be measured in a circuit like this (*Figure 26*):

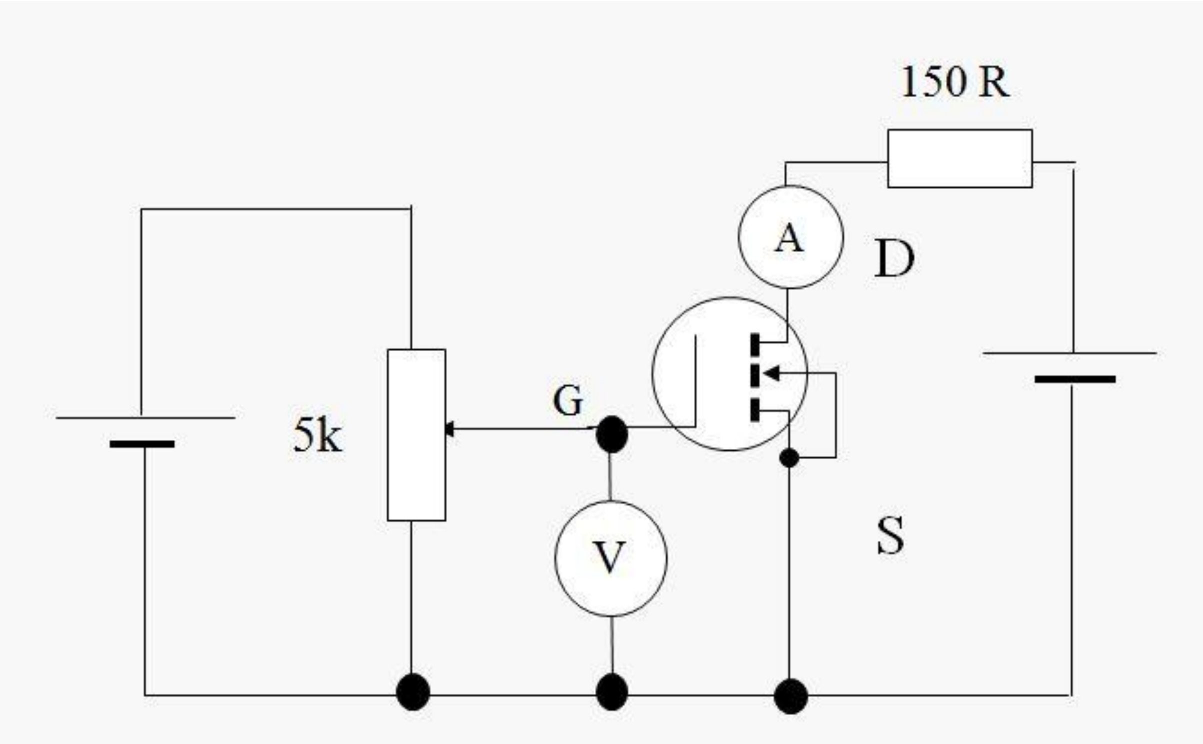


Figure 26 Measuring the characteristics of a MOSFET

We can make the following measurements:

- The **gate - source voltage**, V_{GS}: the voltage between the gate and the source that controls the MOSFET.
- The **drain-source voltage**, V_{DS}: the voltage between the drain (+) and the source (-).
- The **drain source current**, I_{DS}: the current through the MOSFET.
- The **threshold voltage**, V_{th}: the gate source voltage that just creates a conducting path in the n-channel.

The results of this experiment are shown on the graph (*Figure 27*):

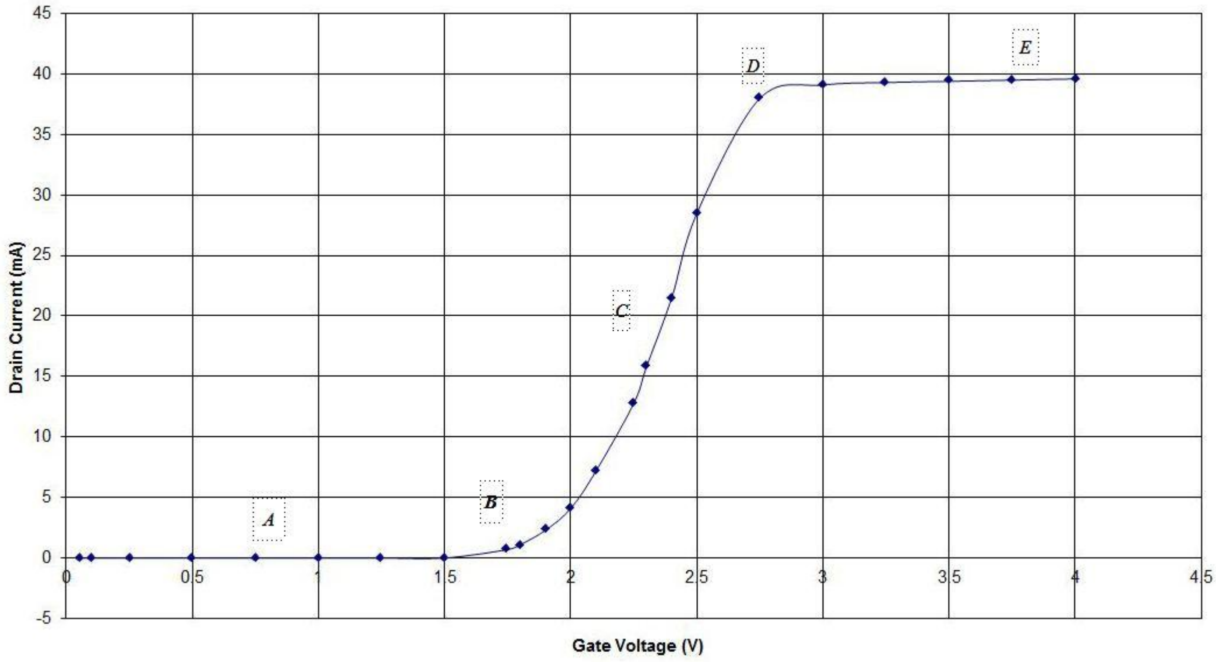


Figure 27 Transconductance of a MOSFET

The characteristics of the MOSFET were explored at a constant drain-source voltage. This is often different to the gate-source voltage, and may be from a separate supply, as was the case in this experiment. We can work out what the voltage was given our answer from Question 14E.01.4 and knowing that there was a 150 Ω current limiting resistor.

The MOSFET can be used as an amplifier or a switch. The amplifier is not on the syllabus.

The circuit below acts as a switch (*Figure 28*).

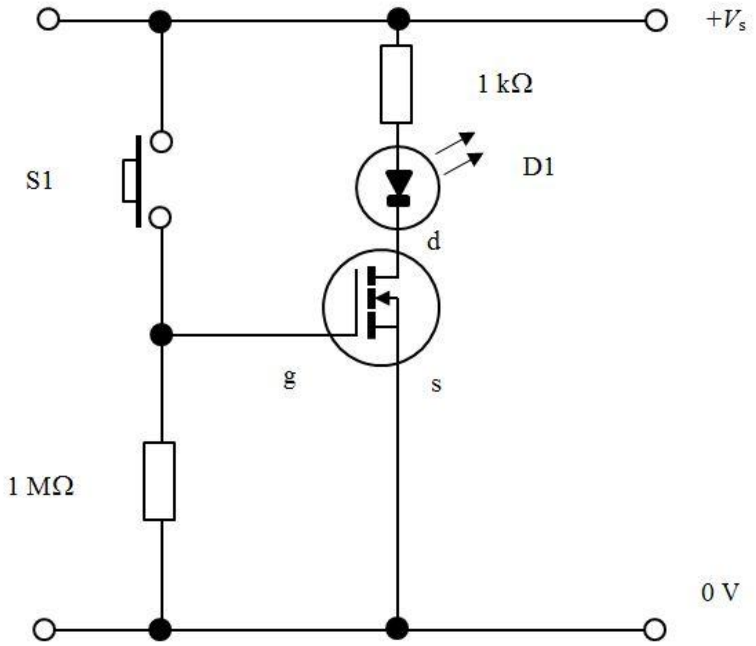


Figure 28 MOSFET used as a switch

When the switch S1 is pressed, the LED D1 turns on. The MOSFET is known as a **driver**.

MOSFETs are used in many digital chips. We will look at digital electronics later

Power MOSFETs can carry a heavy current, for example 10 A. The MOSFET is still turned on by a gate-source voltage of about 2 V, and the gate current is negligible. This contrasts to bipolar transistors that need a driver transistor to drive the power transistor. Power MOSFETs need to have a **heat-sink** to prevent them from overheating and going into thermal runaway. If a motor or any other inductive component is being turned on by a MOSFET, a **reverse biased diode** is needed in parallel with the motor. This prevents damage from a **reverse voltage spike**.

The MOSFET in *Figure 29* is switched on when the output of the potential divider reaches 2.0 V. Note that the motor has a reverse-biased diode to prevent the reverse voltage spike that may occur when the motor is turned off. This can destroy the MOSFET.

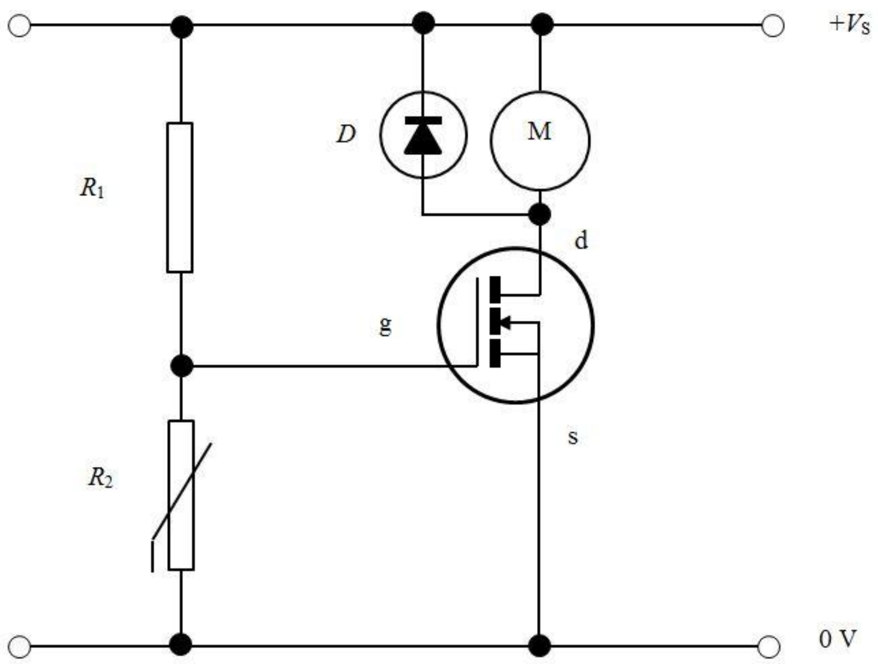


Figure 29 MOSFET used instead with a potential divider

The MOSFET can take the place of a **relay**. It has the advantage of having no moving parts, nor are there any inductive effects from a coil. It can last a lot longer than an electromechanical relay, which has a finite number of operations. The MOSFET can act at high frequency as well.

If you are using an electronic circuit using a high voltage, for example 230 V AC mains, it may be preferable to use a relay for improved isolation for the electronic circuit from the high voltage circuit.

The switch on voltage is determined by the voltage divider made by R1 and R2. The voltage divider equation is:

$$V_{out} = V_{in}\left(\frac{R_2}{R_1 + R_2}\right)$$

.... *Equation 4*

In the Exam

You may well be asked about the input circuit for the MOSFET. It will involve a voltage divider.

Although the MOSFET amplifier is NOT on the syllabus, you may possibly be asked to do a voltage divider calculation on the output stage.

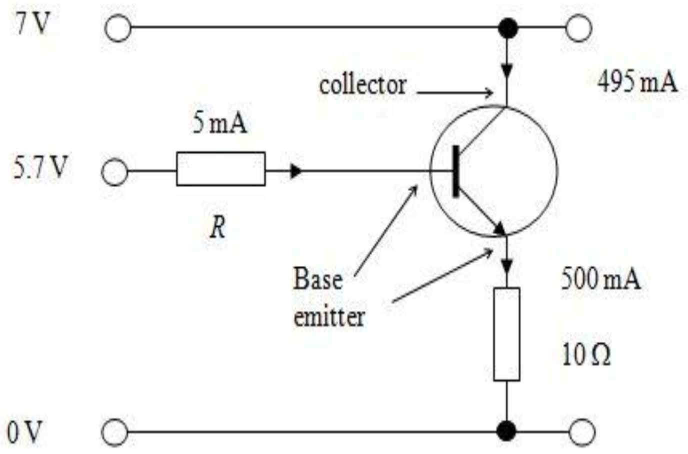
If one of the resistors is a transducer, you may be asked to lift a resistance from a graph, having been given a value for, say, light level.

MOSFETS need to be handled with care. They are stored in conductive packages to prevent **static electricity** building up. A high voltage can punch a hole through the thin oxide layer in the gate. That renders the component useless. Therefore, you need to wear an earthing strap before handling a MOSFET.

Questions

Tutorial 14E.01

14E.01.1



What is the current gain of the transistor shown above?

14E.02.2

What do the letters G, D, and S stand for on a MOSFET?

14E.01.3

Look at *Figure* 27. Write down what is happening at points *A* to *E* in the graph.

14E.01.4

Look at *Figure 27*. Calculate the transconductance of this MOSFET at point *C*. Give your answer to an appropriate number of significant figures and give the correct units. What is the equivalent resistance?

14E.01.5

Calculate the drain-source voltage using data from the graph, circuit, and your answer to Question 14E.01.4.

14E.01.6

Look at *Figure 28*. Explain in terms of what you know about MOSFET why D1 shines when S1 is switched on.

14E.01.7

In the circuit above (*Figure 29*), the MOSFET has a resistance of 22 Ω. The motor takes a current of 0.50 A and has a power of 6.5 W.

Calculate the supply voltage to this circuit.

14E.01.8

The threshold voltage of the MOSFET in *Figure 29* is 2.0 V.

(a) What is the component R_2?

(b) If R_1 has a value of 150 kΩ, calculate the value of R_2 that will give a gate-source voltage of 2.0 V

Tutorial 14 E.02 Zener Diodes	
AQA Syllabus	
Contents	
14E.021 Zener Diode	14E.022 Characteristic of the Zener Diode
14E.023 Voltage Clamp	14E.024 Simple Square Wave Generator

We are familiar with the p-n junction diode from the tutorials in Electricity (Topic 4) and 14E.01. We also looked at the Light Emitting Diode (LED). In this tutorial, we will look at the Zener diode.

14E.021 Zener Diode

The **Zener diode** was named after an American Physicist, Clarence Melvin Zener (1905 - 1993). It is an unspectacular component that looks like this:

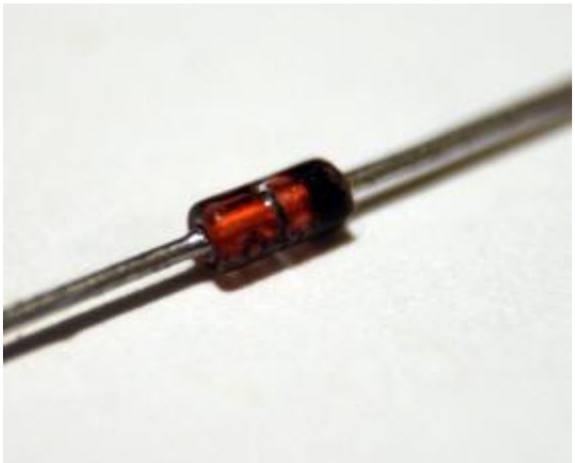


Figure 30 A Zener diode (Image by Teravolt. Wikimedia Commons)

There are several symbols that are used for the Zener diode. This is the one I have always used (*Figure 31*):

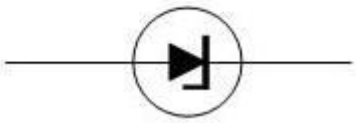


Figure 31 Symbol for a Zener diode

The ordinary diode prevents a reverse-biased voltage until a particular voltage when there is a sudden avalanche of electrons. This voltage is about - 30 V, and when the avalanche happens, a large current, about 500 mA flows, giving a heating effect that will burn the diode out very quickly. The Zener diode behaves in a similar way, but the breakdown voltage is much lower, and the avalanche current is much less.

14E.022 Characteristic of the Zener Diode

We can investigate the characteristics of a Zener diode using the circuit below (*Figure 32*):

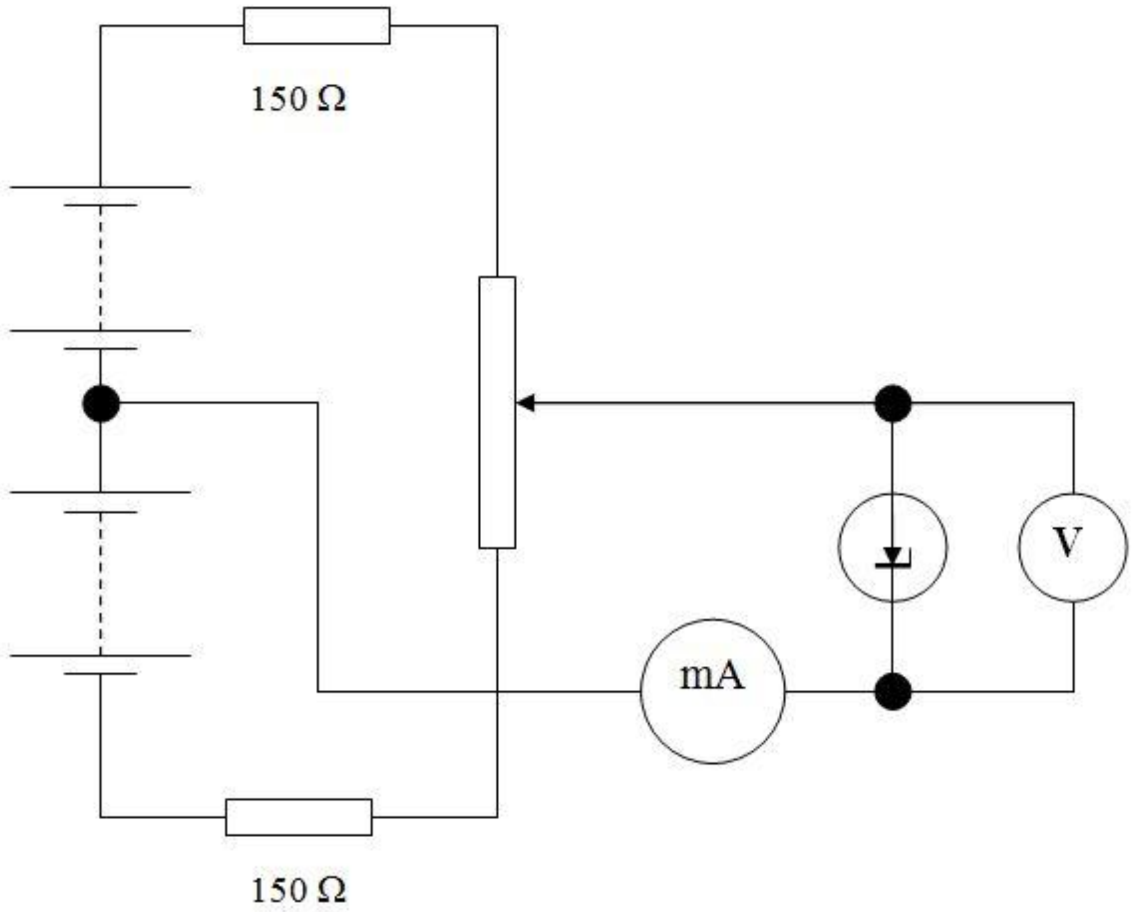


Figure 32 Investigation of a Zener diode

This circuit allows positive and negative voltages to be applied to the diode. Its behaviour is very like an ordinary diode, but a typical reverse-biased breakdown voltage is –5.6 V, and there is a very rapid rise in current.

This is shown in the graph (*Figure 33*):

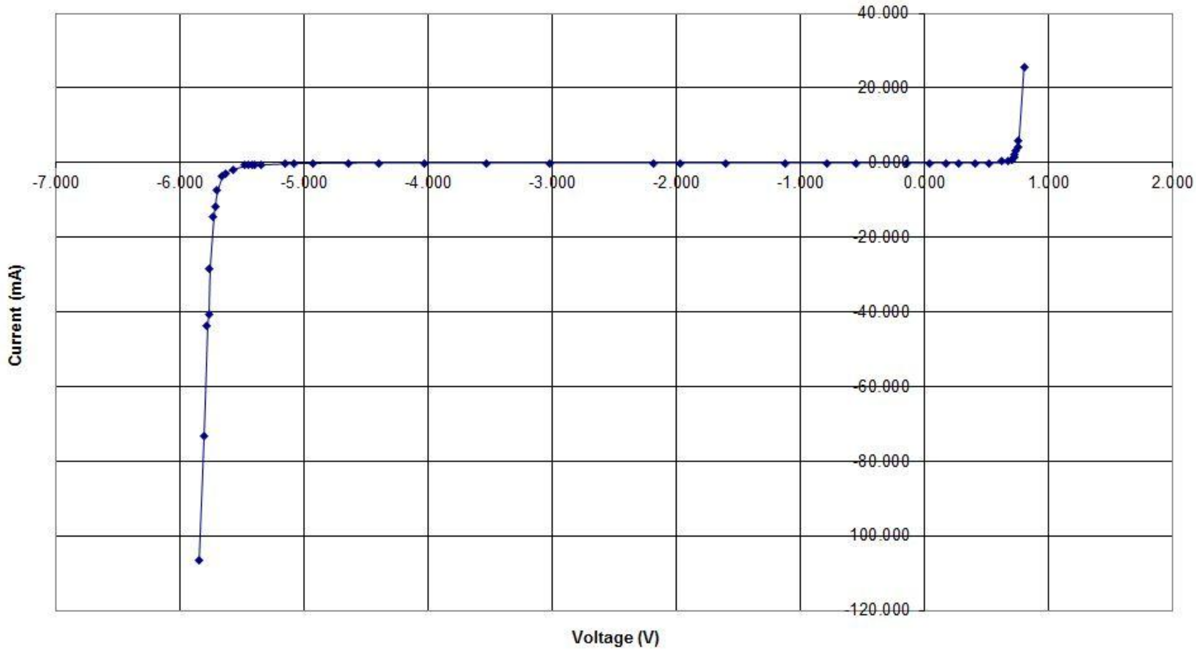


Figure 33 Characteristic of a Zener diode

In forward bias, the Zener diode behaves just like an ordinary diode, with a forward bias voltage of about 0.7 V. In reverse bias, the current is 0 mA until the breakdown voltage of (-)5.6 V is reached. If the reverse voltage is above 5.6 V, a current will flow through the diode. The higher the voltage, the greater the current. There will be a limit to the current that the diode can conduct. The **Zener diode** is designed to be used in a **reverse biased** configuration.

14E.023 Voltage Clamp

In its reverse biased configuration, it will hold the output voltage at a constant 5.6 V. It can be described as a **voltage clamp**. Zener diodes are found in **voltage regulators** (*Figure 34*).

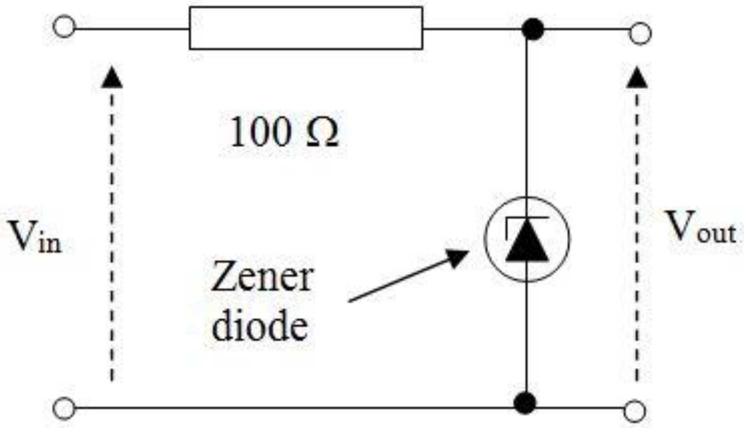


Figure 34 Zener diode being used as a voltage clamp, a very simple voltage regulator

The circuit above (*Figure 34*) shows a Zener diode being used as a very simple voltage regulator.

The graph (*Figure 35*) *shows* how diode limits the voltage, which remains steady at 5.6 volts while a small current is taken. However, the diode can only take a limited current, and the output voltage would fall if the current taken is excessive. The 100 Ω resistor in the circuit above limits the current to 10 mA which will stop the diode being overloaded.

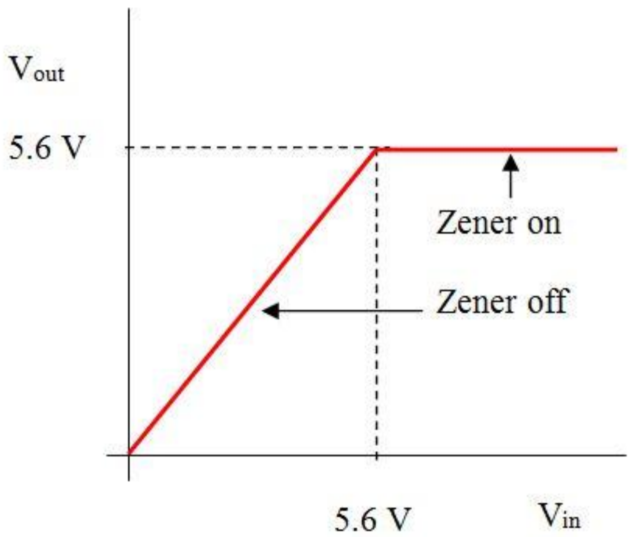


Figure 35 Graph of output voltage against input voltage for a Zener diode

The simple diode clamp is rather limited. The maximum current of 10 mA occurs when there is no load. Let's put a load of 1000 Ω across the output terminals (*Figure 36*):

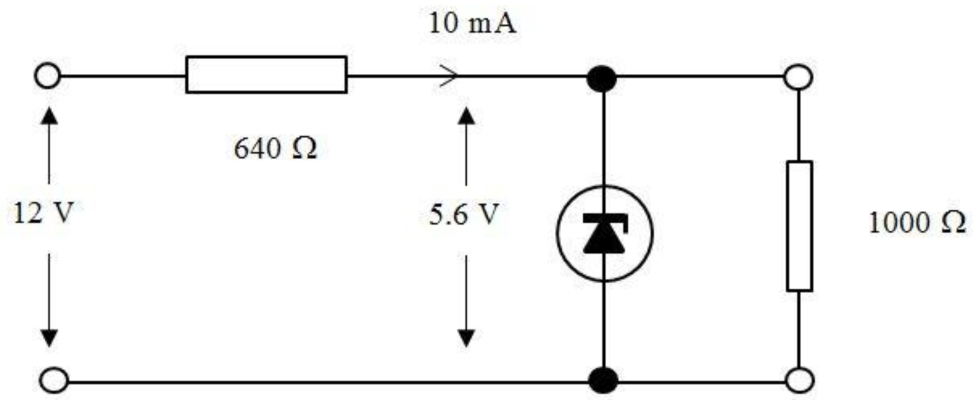


Figure 36 Putting a load onto a simple voltage clamp

A current of 10 mA still flows through the 640 Ω resistor. However, there is less current flowing through the Zener diode.

The minimum current through the Zener diode is 0 mA.

For a lower resistance than that calculated in Question 14E.02.3, the voltage is no longer locked at 5.6 V. The voltage is now determined using the voltage divider equation:

$$V_{out} = V_{in}\left(\frac{R_2}{R_1 + R_2}\right)$$ .... *Equation 5*

For a low resistance load, a more sophisticated voltage regulator is needed.

5.6 V is not the only voltage you can get. There are standard Zener diode voltages other than 5.6 V. They include: 2.4 V; 3.3 V; 4.7 V; 11 V; etc. The highest voltage is 62 V.

Zener diodes can placed in **series** to give a regulated voltage that is the sum of the voltages (*Figure 37*):

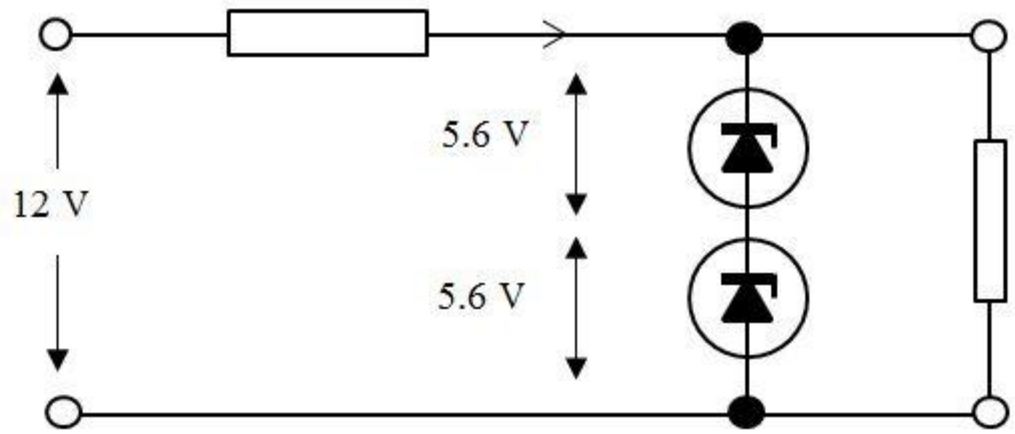


Figure 37 Zener diodes in series

The voltage across the load is 11.2 V.

14.024 Simple Square Wave Generator

Zener diodes can be used to make a very simple square wave generator (*Figure 38*):

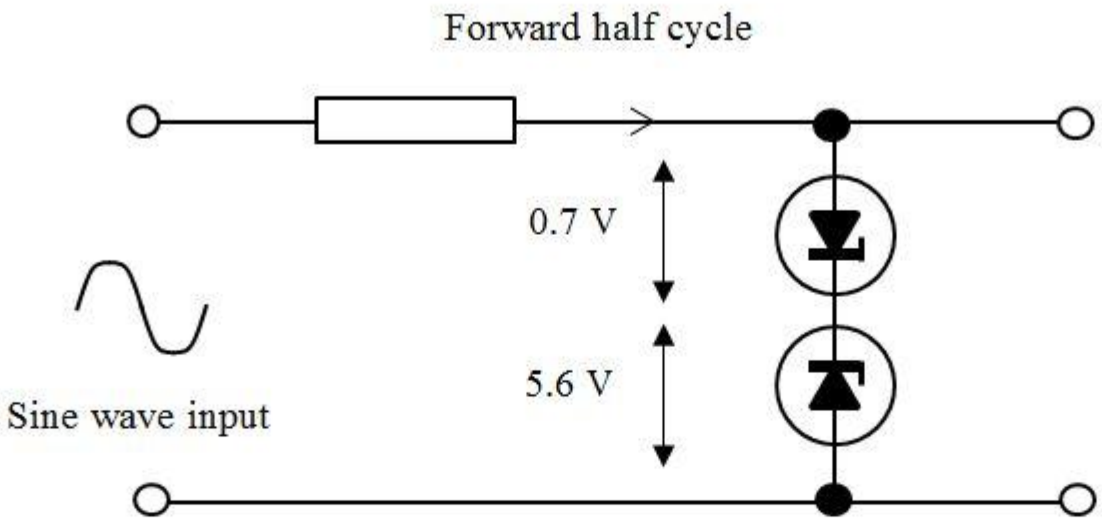


Figure 38 A very simple square wave generator

The sine wave is clipped like this (*Figure 39*):

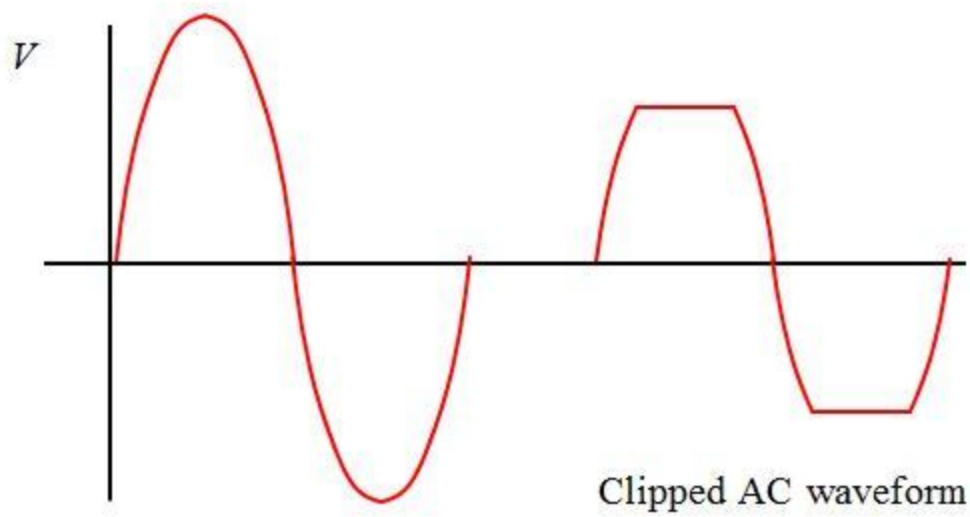


Figure 39 Waveform from a diode clipper

Note that the voltage across the forward biased Zener diode is 0.7 V (the normal diode biasing voltage). Therefore, the voltage will be 0.7 + 5.6 V = 6.3 V. This circuit is sometimes called a **diode clipper**, or even *a poor man's square wave generator*.

Questions

Tutorial 14E.02

14E.02.1

A Zener diode can carry a maximum current of 10 mA at its breakdown voltage of 5.6 V. It is connected to a 12 volt supply. Show that the current limiting resistor has a resistance of 640 W

14E.02.2

Refer to *Figure 36*. Calculate the current flowing through the Zener diode now.

14E.02.3

Refer to *Figure 36*. What is the load resistance for when there is zero current in the Zener diode?

14E.02.4

Refer to *Figure 36*. What is the load voltage if the load is 300 Ω?

Tutorial 14E.03 Photodiode

AQA Syllabus

Contents

14E.031 Photodiode

Any junction between an n-type material and a p-type material can act as a **photodiode**. Indeed, the junctions of any diode can be affected by photons of sufficient energy. Diode junctions are usually encapsulated in opaque materials to prevent light getting at the junction. However, junctions can still be affected by X-ray and gamma ray photons.

A photodiode looks like this (*Figure 40*):

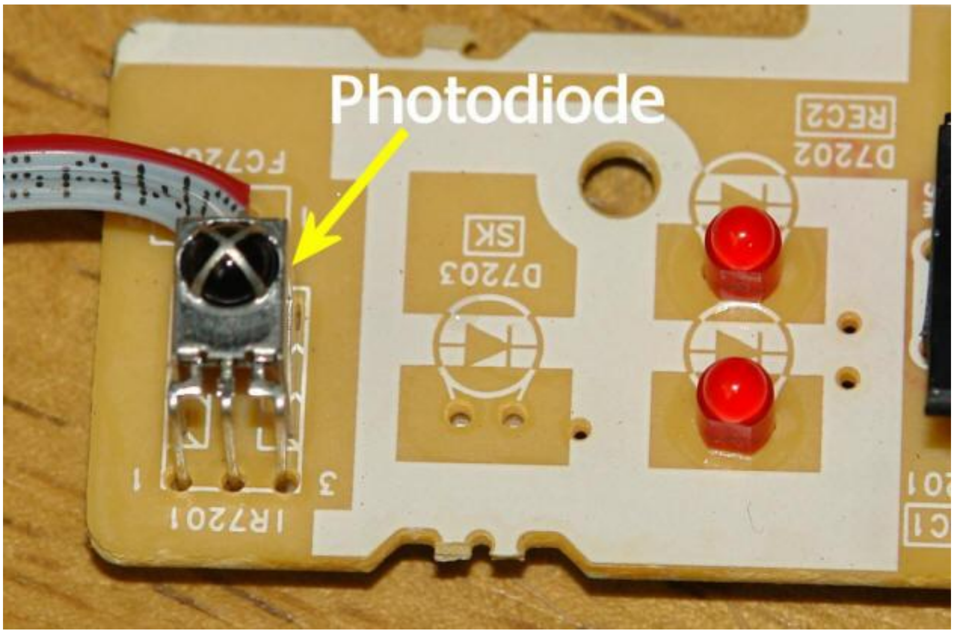


Figure 40 A photodiode

This particular one belonged to a defunct hard disc video recorder (HDD recorder). Its purpose was to pick up the infra-red signals emitted by a hand-held remote controller.

The symbol for a photodiode is:

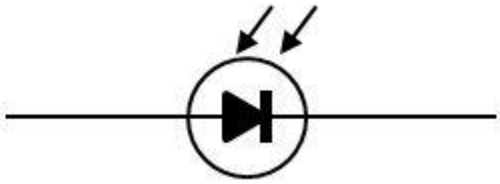


Figure 41 Symbol for a photodiode

If the depletion zone of a junction is struck by a photon of sufficient energy, an **electron-hole pair** is generated (*Figure 42*).

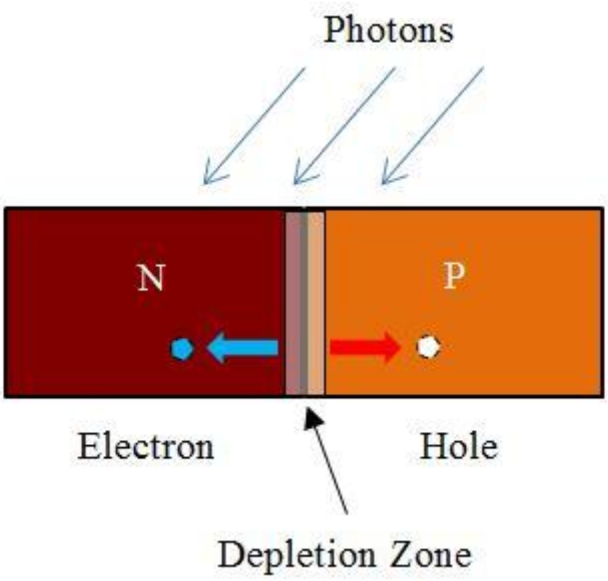


Figure 42 Action of a photodiode

We will look at:

- How Photodiodes work.
- What measurements we can make on photodiodes.
- How we can use photodiodes.

14E.032 Photovoltaic Mode

The photons transfer energy to the semiconductor which results in electrons from the **valence** bands getting sufficient energy to jump to the **conduction** band. They are **free** and are repelled by the electromagnetic force in the depletion zone. They move to the n-type material. When the electrons are removed, a hole is left, and this moves towards the p-type material as a result of the repulsive electromagnetic force in the depletion zone. The field made by the depletion layer moves the holes towards the anode and the electron towards the cathode. The electric field accelerates the electron, so they have a high **drift velocity**. They cross the p-n junction without combining with the atoms there.

If the junction is **zero biased**, a small voltage is generated. If connected to an outside circuit, a current will flow. This is called the **photovoltaic effect** and can be used in a **photovoltaic cell**. A photovoltaic cell uses this idea with large area p-n junctions.

The advantage of using the photodiode in photovoltaic mode is that the dark current is very low. The disadvantage is that the response time is high.

A photovoltaic effect can be observed by shining bright white light onto a yellow, orange, or green LED. It doesn't work with a red LED.

14E.033 Photoconductive Mode

The photodiode is always used in **reverse bias**. When there is a reverse voltage on the photodiode, there is always a small current that passes through the diode. It is **independent** of the light level and occurs even in the dark. It is called the **dark current** and often given the Physics code I_l. The dark current has a typical value of about 0.5 μA (5×10^{-7} A). When the reverse-biased photodiode is exposed to light there is a **reverse current** or **photocurrent** that depends on the light intensity, NOT the reverse biased voltage. Therefore:

Total current = Dark current + photocurrent

The dark current needs to be small for the photodiode to be sufficiently sensitive.

The p-n-junction photodiode has now mostly been superseded by the **PIN-junction photodiode**. Instead of just a p-type material and an n-type material, there is a piece of **intrinsic** semiconductor material placed between the p material and the n material like this (*Figure 43*):

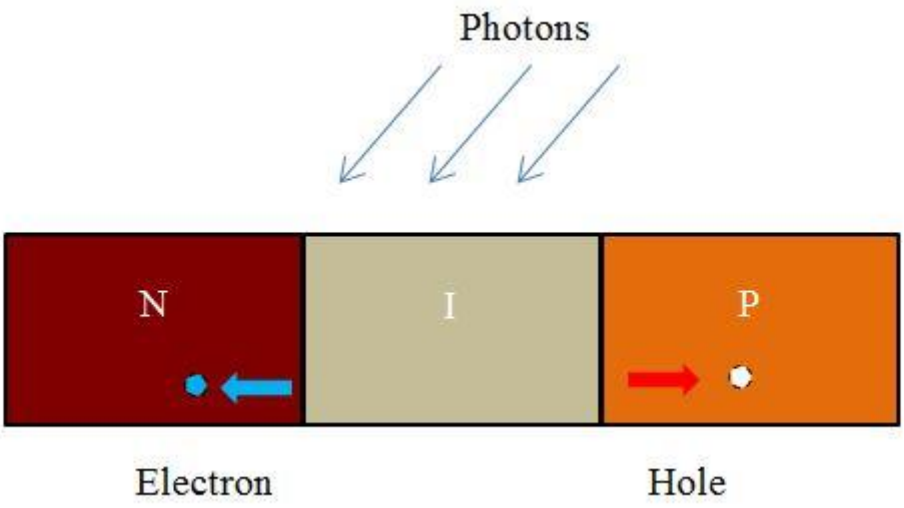


Figure 43 PIN junction diode

While the **n-type** material has doping of **Group 5** elements, and the **p-type** material has doping of **Group 3** elements, the **i-type** material is the pure **Group 4** semiconductor with no doping. Therefore, there are very few charge carriers in the conduction band. Under reverse bias, n-type material is connected to the positive, and the p-type material is connected to the negative. Therefore, the holes will go to the negative end in the p-type material, attracted by the electrons. The electrons in the n-type material will go to the positive end of the n-type material, attracted by the deficiency of electrons, like this (*Figure 44*):

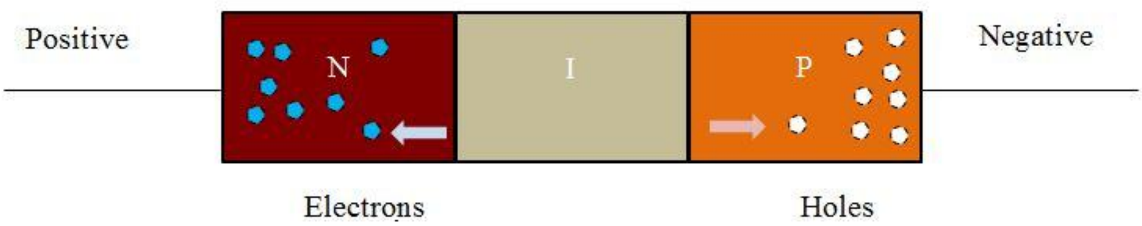


Figure 44 Increasing the depletion layer

This makes the depletion zone very wide indeed. The majority carriers (electrons in the n-type and holes in the p-type) will not carry any current under these circumstances. Minority carriers (electrons in the p-type and holes in the n-type) will carry a small electric current due to the externally applied electric field. This is the **dark** current.

When photons of sufficient energy are applied to the depletion layer, electron-hole pairs are generated. Due to repulsion within the depletion layer, the electrons go to the n-type material, and the holes go to the p-type material. Like the p-n-junction photodiode the total current is:

Total current = Dark current + photocurrent

This might seem much more involved than the simple p-n-junction, but it ensure that there are many more minority carriers to take the current. There are advantages to this:

- A greater bandwidth.
- A higher quantum efficiency (more electrons released by photons);
- Higher response speed.

A further development is the **avalanche photodiode**. A larger reverse-biased voltage gives more energy to the electron-hole pairs.

The higher voltage does NOT increase the number of electron-hole pairs.

The increased energy leads to a higher drift velocity, which will result in electrons being knocked off as the fast-moving free electrons collide with atoms. These too are accelerated by the reverse voltage, and collide with other atoms, knocking off yet more electrons. The result of this is an **avalanche** of electrons. A snow avalanche is a useful way to think this through. A small shock is applied to an unstable snowfield on a slope. Bits of snow are dislodge, which in turn dislodge more snow. Eventually thousands of tonnes of snow hurtle down the hillside and can do a lot of damage.

The advantages of this are:

- Higher gain.
- Higher sensitivity.

The disadvantage is that there is a lot of **noise** (unwanted signals).

14E.034 Measurements on Photodiodes

Like all electronic components, photodiodes have a certain number of **operational parameters**, or measurements that can be made:

- Dark current - leakage current when there is no light. This can be affected by the materials used, and temperature.
- Responsivity - the ratio of the photocurrent to the applied light intensity.
- Quantum efficiency - the ratio of electron-hole pairs generated to the number of photons. If every single photon made one electron-hole pair, the quantum efficiency would be 100 %.
- Response time - the time taken for the photon induced charge carriers to cross the depletion layer.

The measurements can be made using a simple circuit like this (*Figure 45*):

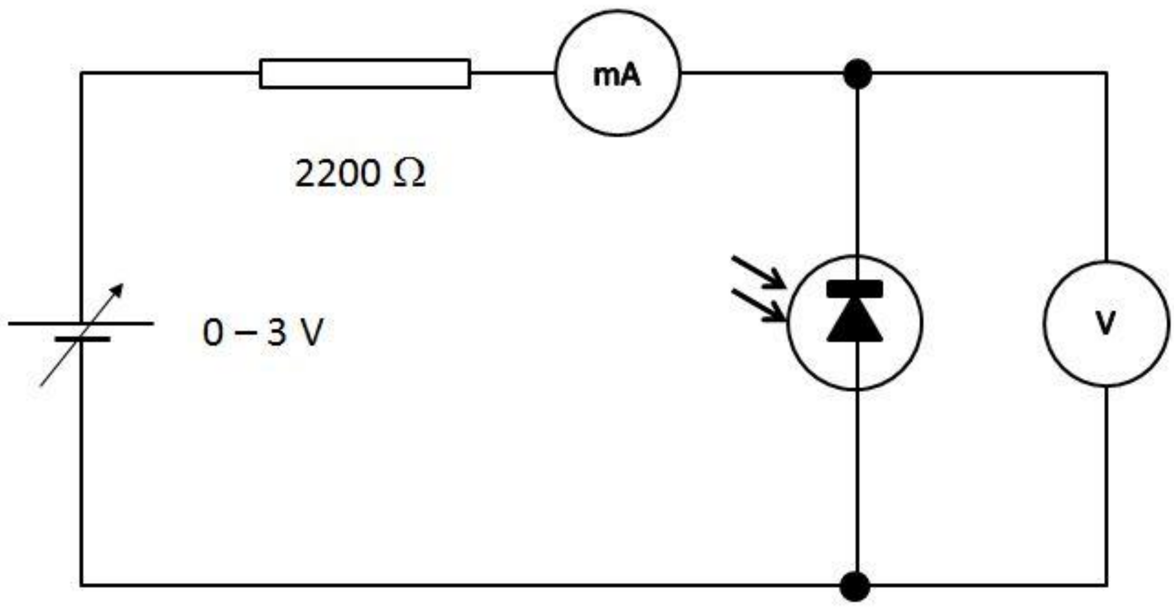


Figure 45 Investigating the photodiode

Dark Current is the leakage current caused by the reverse-bias electric field when there is no light. The number of free charge carriers can be increased with temperature. This is because electrons acquire sufficient energy to go into the conduction band with a higher temperature. The graph of dark current against temperature is shown in *Figure 46*.

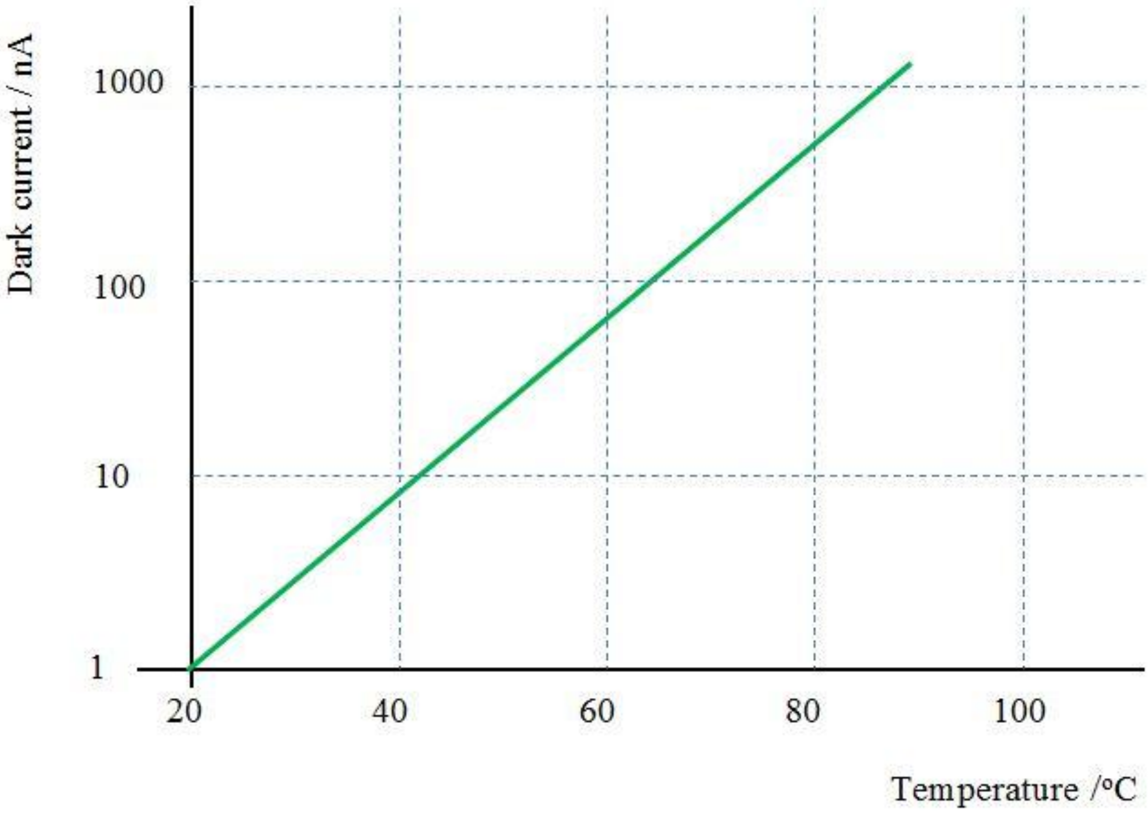


Figure 46 Graph of dark current against temperature

Note that the vertical axis is not a linear scale but goes up in **decades**. It is a **logarithmic** scale. The dark current increases 1000 times as the temperature increases from 20 °C to 90 °C. The implication of this is that photodiodes need to be used in a stable environment where the temperature is constant. The slope of the graph will depend on what materials are used for doping the p-material and the n-material.

Be careful in the way that you read logarithmic scales.

The point that is half-way between 0.1 and 1.0 is NOT 0.5; it is 0.316.

Similarly, the half way point between 1 and 10 is 3.16.

Responsivity is defined as the **ratio** of the **photocurrent** (A) to the **power** (W) of the light that is striking the photodiode. If we know the intensity (brightness) of the light, we can work out the power by:

$$P = IA$$

...........*Equation 6*

[P – power (W), I = intensity (W m^{-2}), A – area (m^2)]

Responsivity has the physics code R_λ and the units are A W^{-1}. The formula is:

$$R_\lambda = \frac{I_{PD}}{P}$$

.... *Equation 7*

Question 14E.03.1 gets you to practise this.

The **quantum efficiency** reflects the number of electron hole pairs generated when photons land on the photodiode. It has the physics code η ("eta" - a Greek letter long 'ē'). A typical infra-red photodiode has a quantum efficiency of 70 % at a wavelength of 950 nm. You may need to revise what you did on photon energy in Particle and Quantum Physics (Topics 2 and 3).

The typical **response time** of a PIN photodiode are in the order of nanoseconds (where 1.0 ns = 1.0 × 10^{-9} s). For a typical IR photodiode, the rise time is 2.5 ns, and the fall time is 2.5 ns.

14E.035 Photodiode Response to Light

The **response to wavelength** of a photodiode is an important consideration for the electronic engineer. If a photodiode is designed to be used in the infra-red region, it will not work at all if visible light is used. Photodiodes have a range of wavelengths to which they are sensitive, and a wavelength to which they are most sensitive. A typical example is shown on the graph (*Figure 47*):

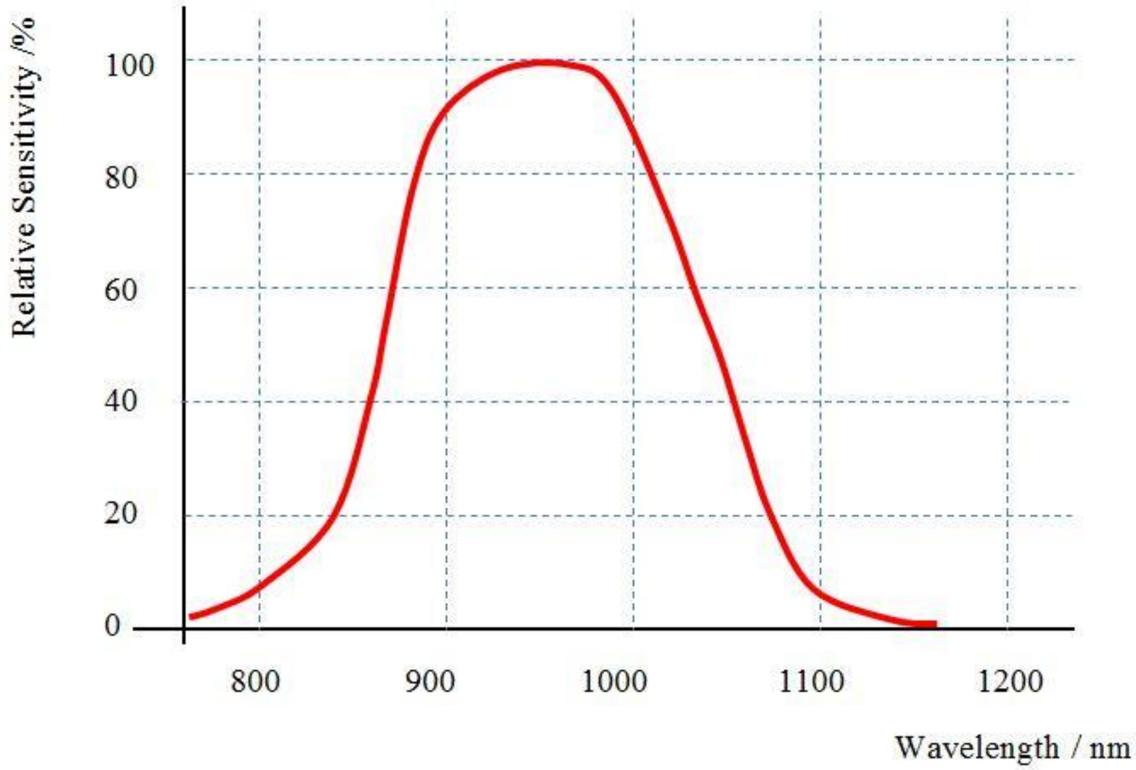


Figure 47 Wavelength sensitivity of a photodiode

The maximum sensitivity for this diode is about 950 nm. This is the in the infra-red region. Photodiodes are available for visible light as well.

We can plot a graph of the photocurrent against the light intensity (*Figure 48*):

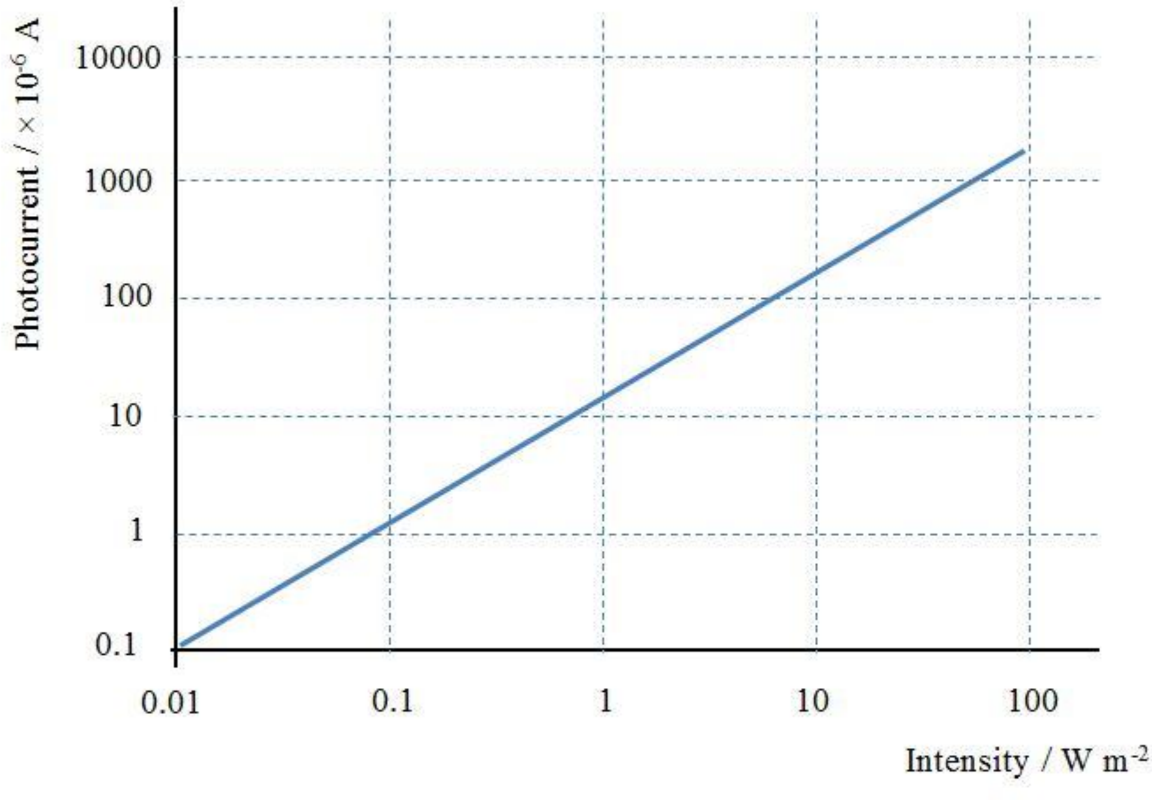


Figure 48 Graph of photocurrent against light intensity

This graph has decade or logarithmic scales on both the horizontal and vertical axes.

Be careful in the way that you read these.

The point that is half-way between 0.1 and 1.0 is NOT 0.5; it is 0.316.

Similarly, the half way point between 1 and 10 is 3.16.

If we measure the photocurrent against the reverse voltage at different light levels, we get a graph like this (*Figure 49*):

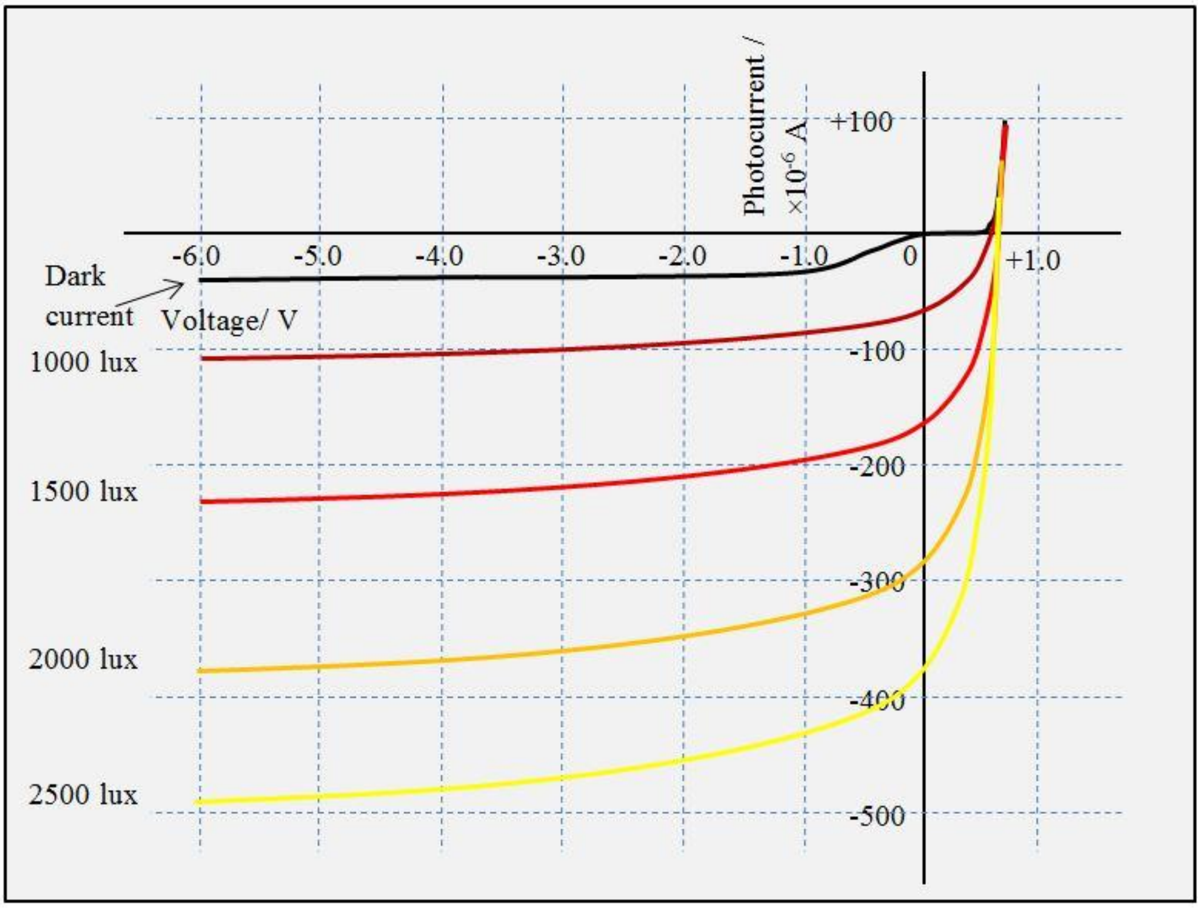


Figure 49 Dark current against voltage at increasing light levels

The voltage required for conduction in forward bias is about 0.7 V regardless of the light level. The dark current at a given reverse biased voltage depends on the light level itself.

Notice that the photocurrent increases with increasing **light level**. The light level is in this case measured in **lux**, where 1 lux = 1 lumen per square metre. It is not possible to do a direct conversion from lux into watts per square metre as there is a different conversion factor for every different wavelength. The human eye has a maximum sensitivity to green light at a wavelength of 555 nm, at which point the conversion factor is:

$$\mathbf{1\ W\ m^{-2} = 683\ lux}$$

For any given light level, the photocurrent does not increase that much as the reverse voltage increases. This is because at a given light level, a maximum amount of electron-hole pairs are released, so the photocurrent is **saturated**. An increased light level (brightness) more charge carriers are released.

A photodiode in forward bias can act as a very inefficient LED. (Turn the voltage up high enough and there is a very bright flash and a puff of smoke.)

14E.036 How the Photodiode is used

Notice from the graph above that in forward bias, the photodiode acts as a normal diode. It conducts as a forward biased diode at about 0.7 V. Light levels make little difference in forward bias. Therefore, the photodiode is **always** used in **reverse bias**. The circuit below is the very simplest way of using the photodiode (*Figure 50*):

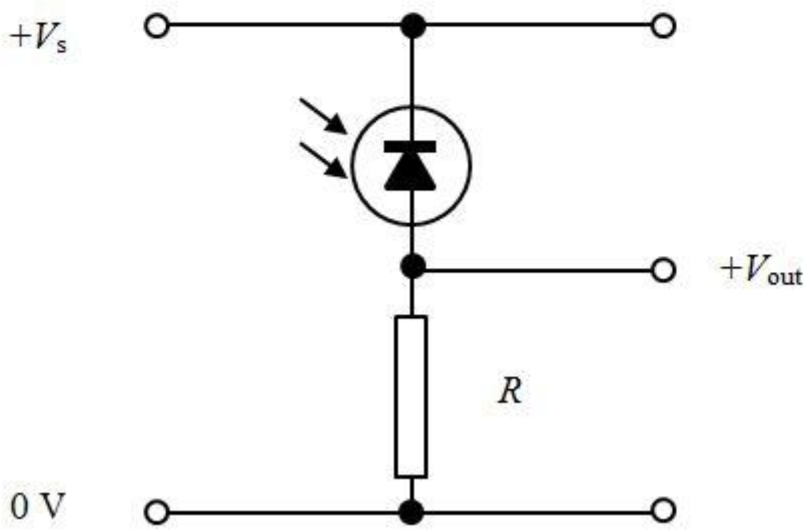


Figure 50 Simple way of using a photodiode

The resistor *R* is there to limit the current through the photodiode which can only take a limited current. It also provides a voltage which depends on the current. So V_{out} is proportional to the total current (which is the sum of the photocurrent and the dark current). To work out the value of resistor you need, you need to know what the light intensity will be and have a graph that will give you the current at that intensity. We have seen this graph before (*Figure 51*):

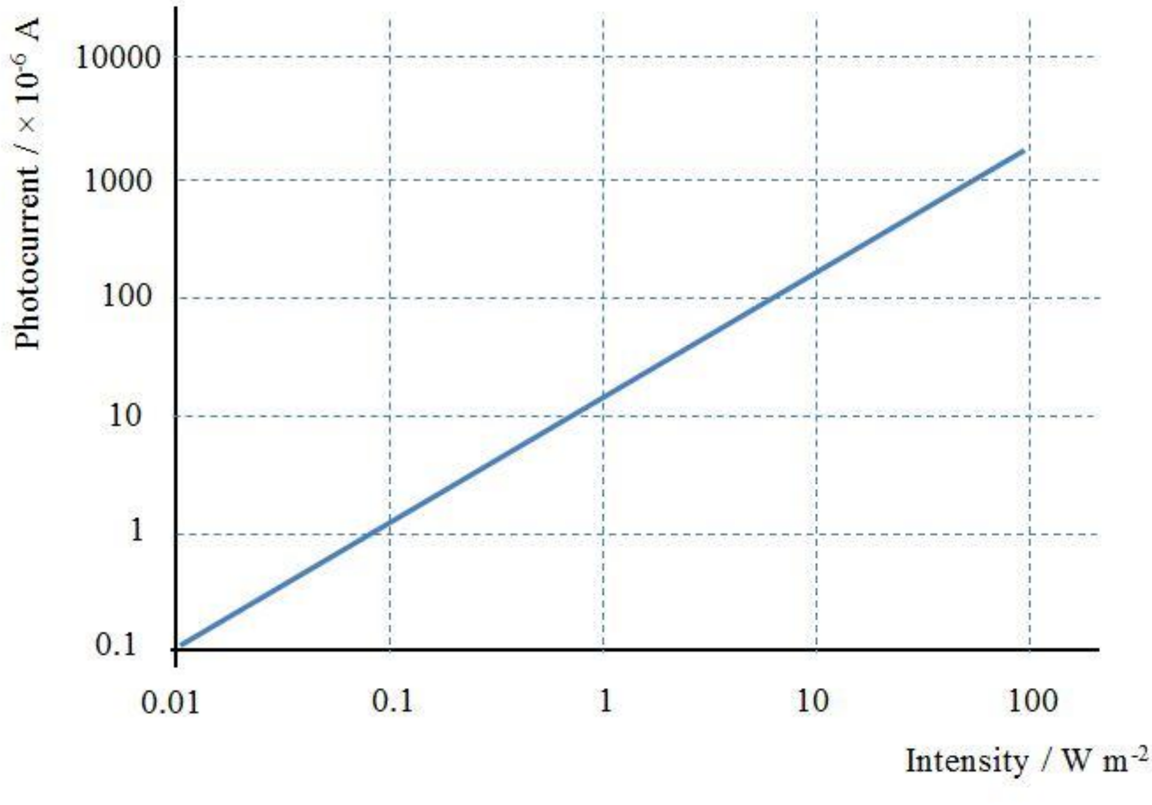
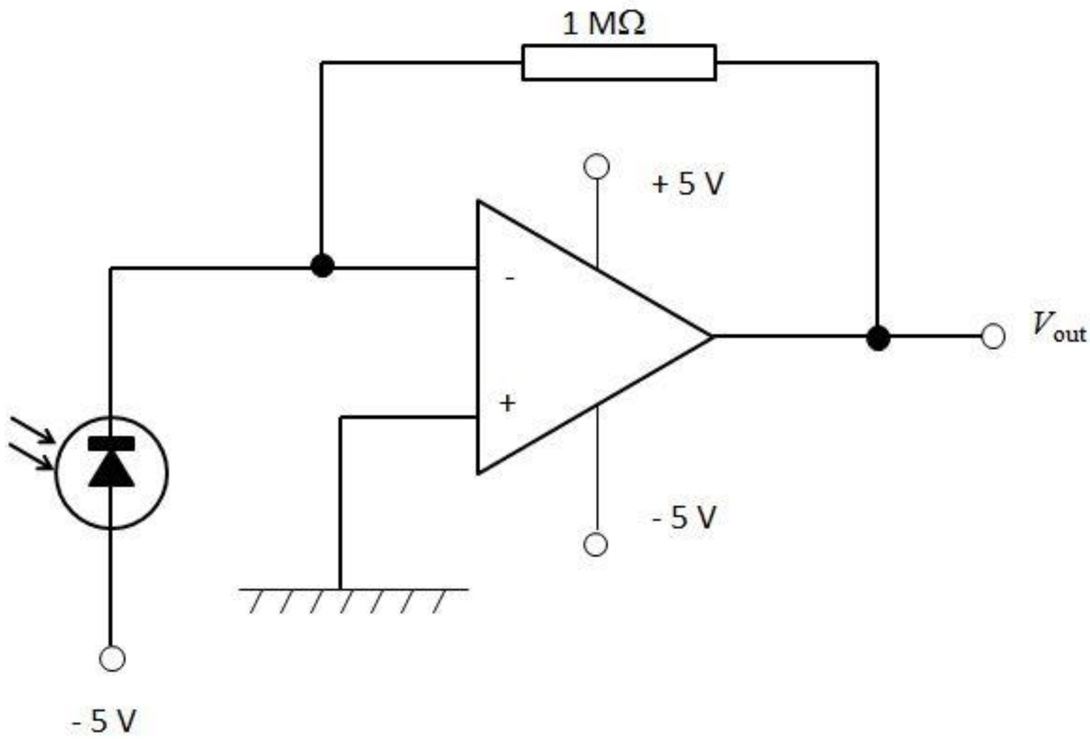


Figure 51 Graph of photocurrent against intensity

In the Exam

The kind of question you are most likely to asked about the photodiode is one that asks you to lift data off a graph. The sketch graphs I have drawn do not show the minor lines spaced out in a logarithmic manner. In the exam, they will be. You will get a mark for the correct reading of the graph. A wrong reading will be penalised there, but the error will be carried forward for correct use of the incorrect value.

It is quite likely that you will be asked about the photodiode in the context of a wider circuit, like this:



This circuit contains an **operational amplifier** (op-amp) which we will look at later.

You will probably be asked a 1 mark question for a use of a photodiode. Some are given below, but it's not an exhaustive list.

14E .037 More Uses for a Photodiode

There are many uses for photodiodes, including:

- Light meters in a camera.
- Automatic flash control in a camera.
- Flame monitors.
- Twilight detector in a car.
- Fibre-optic link.
- Detector for a TV remote controller.

The diagram shows the idea for a fibre optic link (*Figure 52*):

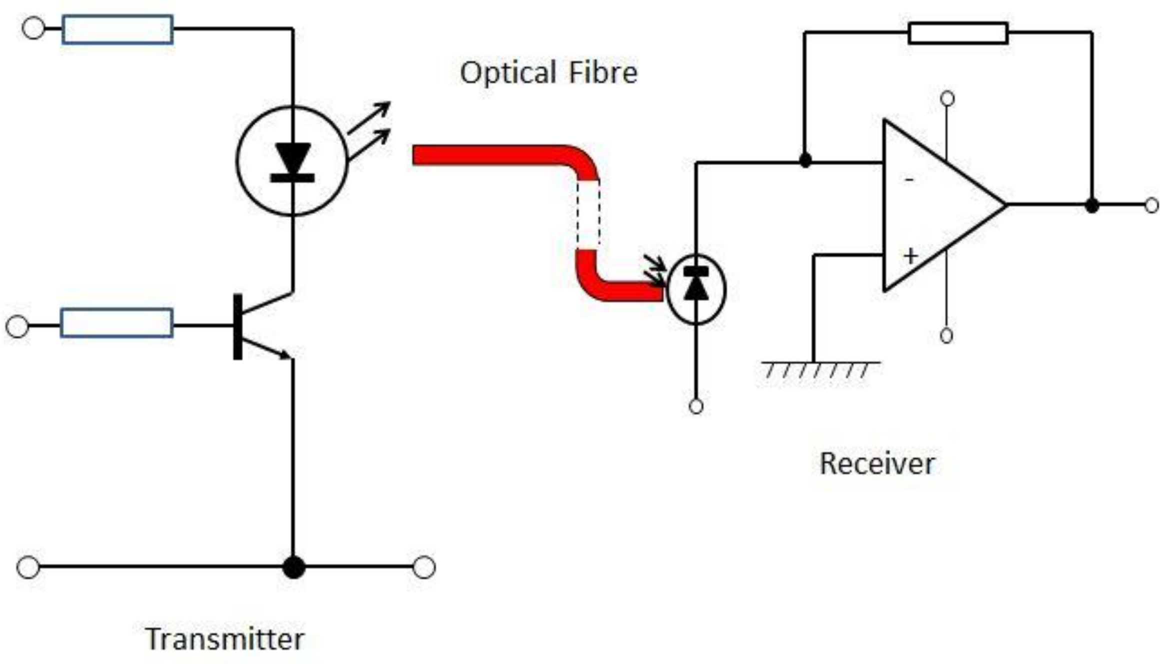


Figure 52 Photodiode in a fibre optic link.

The LED transmits pulses in the infra-red region as optical fibre glass is particularly transparent to infra-red wavelengths. An infra-red photodiode is used on the receiver.

14E.038 Photodiodes with Scintillators

When a high energy particle such as a beta particle or gamma ray photon interacts with certain crystals, a flash of light is given off. This is called **scintillation**. A photomultiplier tube was often used to detect these flashes. Nowadays photodiodes are more likely to be used to detect the flashes in the scintillators. Here is a **scintillator** with a photodiode (*Figure 53*):

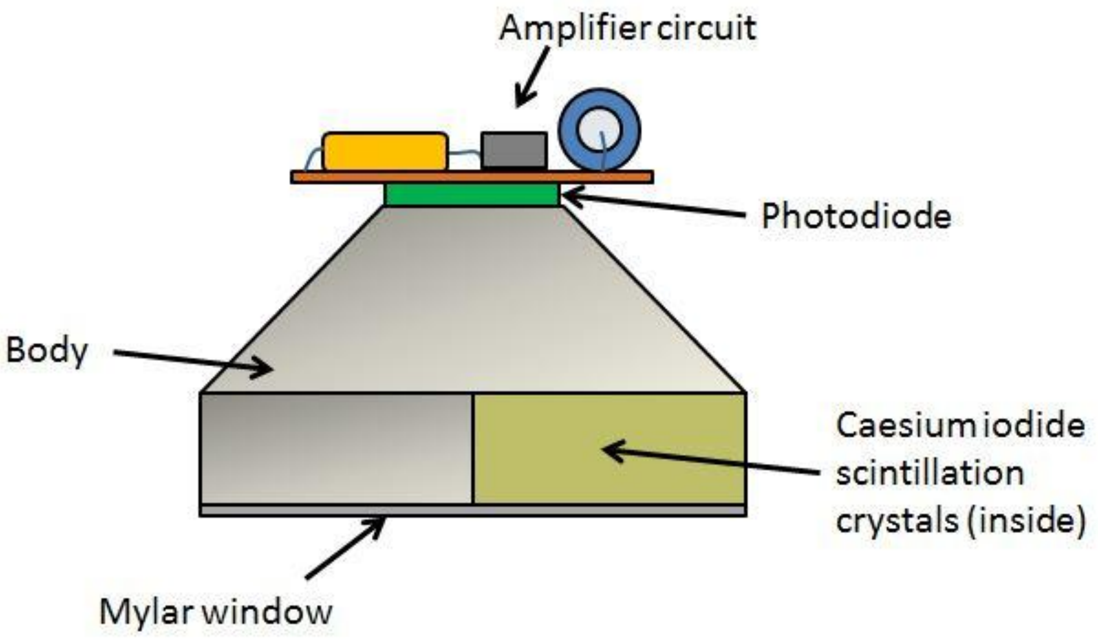


Figure 53 Photodiode being used with a scintillator

They are used where:

- There are no high voltages available or desirable (for example in medicine).
- Stable gain is essential (for example in monitoring the environment).
- High magnetic fields are present.
- Detectors need to be rugged.

These types of photodiode detectors are not much good with low energy particles as the noise (unwanted signals) can swamp the scintillation signals.

Questions

Tutorial 14E.03

14E.03.1

A photodiode passes a total current of 60 mA when light of intensity 10 W m^{-2} is shone on it. The dark current is 1.5 mA. The photodiode has a circular window that is 10 mm in diameter.

Calculate:

(a) the photocurrent.

(b) the power of the light.

(c) the responsivity.

14E.03.2

Use your answer to Question 14E.03.1 (b) to calculate the number of photons falling on the photodiode every second, assuming the wavelength is 510 nm.

Hence work out the quantum efficiency, using your answer to Question 14E.03.1 (a).

Planck's constant, $h = 6.63 \times 10^{-34}$ J s

Electronic charge, $e = 1.6 \times 10^{-19}$ C

14E.03.3

The rise and fall time of a photodiode are both 2.5 ns, and the diode conducts for 1 ns between the rise and the fall. Calculate the frequency of the input signal.

14E.03.4

Refer to the graph in *Figure* 47. The maximum quantum efficiency of the photodiode that gave the graph above is known to be 70 %. Calculate the quantum efficiency if the wavelength is 1000 nm.

14 E.03.5

Refer to page 47. What is 2500 lux in W m^{-2} at 555 nm, green light?

14E.03.6

Refer to Figure 51 in answering this question.

A photodiode is designed to work in a light intensity of 10 W m^{-2}. It is connected to a 3 V supply and a resistor. When the reverse current is flowing, a voltage of 0.50 V is measured across the diode. Calculate the value of the resistor.

Tutorial 14E.04 Hall Sensors	
AQA Syllabus	
Contents	

14E.041 Hall Sensor Basics

The Hall sensor uses the **Hall** Effect. It is often used to detect the **rotary motion** of a magnet set into a wheel or axle. Examples of uses are:

- Tachometers for diesel engines.
- Speed sensors on exercise bikes or rowing machines.
- Contactless measurement of direct currents (i.e. without inserting an ammeter);
- Feedback for motor control.

The symbol for a Hall sensor is this (*Figure 54*):

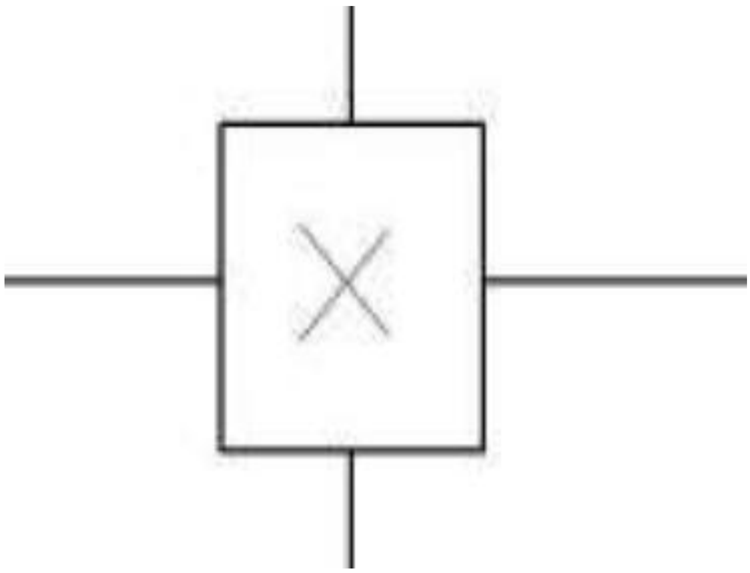


Figure 54 Symbol for a Hall Sensor

14E.042 The Hall Effect

It was first discovered in 1879 by an American Physicist, Edwin Herbert Hall (1855 - 1938). He found that when a current-carrying conductor or a semiconductor is placed in a magnetic field, a voltage occurs that is perpendicular to the flow of the current. This is called the Hall Voltage.

The Hall Effect can be observed in conductors, semi-conductors, ionised gases, and plasmas. We will be looking at the effect of magnetic fields in electrons in semi-conductors, because the nature of semi-conductors allow for a relatively high voltage which is easy to measure. It is also easily reproduced in a school or college Physics laboratory. So, let's have a look at what happens.

The separation of charges leads to a potential difference, or voltage that is called the **Hall voltage** (V_H). This voltage also causes an electric field, *E* which is uniform. The idea is shown in the picture below (*Figure 55*):

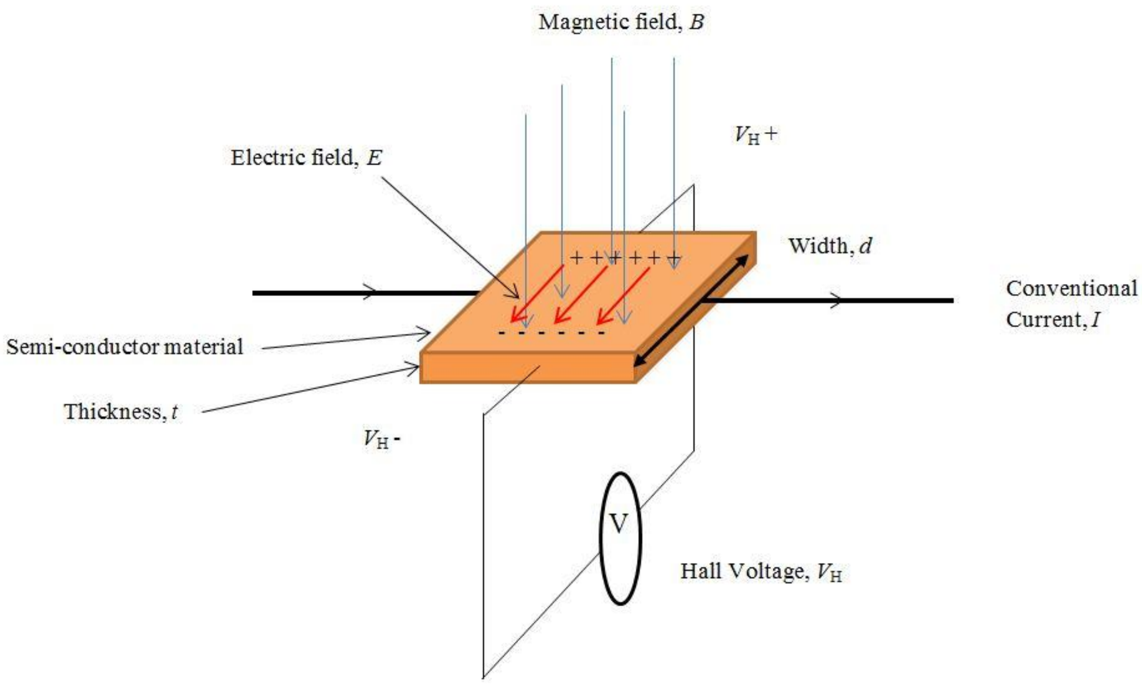


Figure 55 The Hall Voltage

You are NOT expected to use the equation or derive it. However, if you wish to see how the equation is derived and used, please see 14E.043. The output of a Hall sensor is about 0.1 V. Connected to a circuit, it would give a current of about 1 mA. This can be amplified to give a bigger voltage and current.

14E.043 The Hall Equation and its Derivation (Extension only)

Refer to *Figure* 55. The electric field, E, across the Hall slice is given by:

$$E = \frac{V_H}{d}$$ .... *Equation 8*

We also know from the definition of electric field that the force is:

$$F = Eq$$ .... *Equation 9*

So, we can write:

$$F = \frac{V_H q}{d}$$ .... *Equation 10*

We also know that

$$F = Bqv$$ .... *Equation 11*

from previous tutorials, so we can write:

$$Bqv = \frac{V_H q}{d}$$ .... *Equation 12*

The q terms obligingly cancel out, and we can rearrange to give:

$$V_H = Bvd$$ .... *Equation 13*

Measuring the speed of individual electrons is not at all easy, but if we review Topic 11 Tutorial 11.03, we know that:

$$I = nAvq$$ .... *Equation 14*

Where:

- n = number of charge carriers per unit volume (m^{-3}).
- A = area of the conductor (m^{2}).
- v = speed of charge flow (m s^{-1}).
- q = charge (C).
- I = current (A).

So, we rearrange to get:

$$v = \frac{I}{nAq}$$

.... *Equation 15*

and then substitute into *Equation 13* to give:

$$V_{\mathrm{H}} = \frac{BId}{nAq}$$

.... *Equation 16*

Now area

$$A = dt$$ .... *Equation 17*

So, we can write:

$$V_{\mathrm{H}} = \frac{BId}{ndtq}$$

.... *Equation 18*

and the d terms cancel out to give us our final relationship:

$$V_{\mathrm{H}} = \frac{BI}{ntq}$$

.... *Equation 19*

where:

- V_H = Hall voltage.
- B = magnetic flux density (T).
- I = current (A).
- n = number of charge carriers per unit volume (m^{-3}).
- t = thickness of the slice (m).
- q = charge (C).

Note that n is the number of charge carriers per unit volume, not the number of turns per unit length.

The term dt in the argument above was distance × thickness to give area. It was nothing to do with a time interval!

The number of charge carriers per unit volume for typical semi-conductors is shown in the table:

Semi-conductor	n / m^{-3}
Gallium arsenide	1.10×10^{25}
Germanium	2.02×10^{21}
Pure Silicon	1.50×10^{16}

These figures are for a temperature of 300 K. The number of free charge carriers per unit volume rises as the temperature rises. We will use these, as college and university physics labs tend to have a temperature of about 300 K (27 °C). You would be well boiling if the temperature in the lab was 400 K.

For metals:

Metal	n / m^{-3}
Aluminium	6.02×10^{28}
Copper	8.46×10^{28}
Tungsten	3.43×10^{28}

The Hall effect in theory can be observed in metals, but the Hall voltage would be so tiny that it's negligible. A calculation using the data above in a copper wire gives a Hall voltage of about 2×10^{-9} V.

14E.044 Some uses of the Hall Sensor

The picture (*Figure 56*) below shows the **Hall Probe** used with a data-logger.

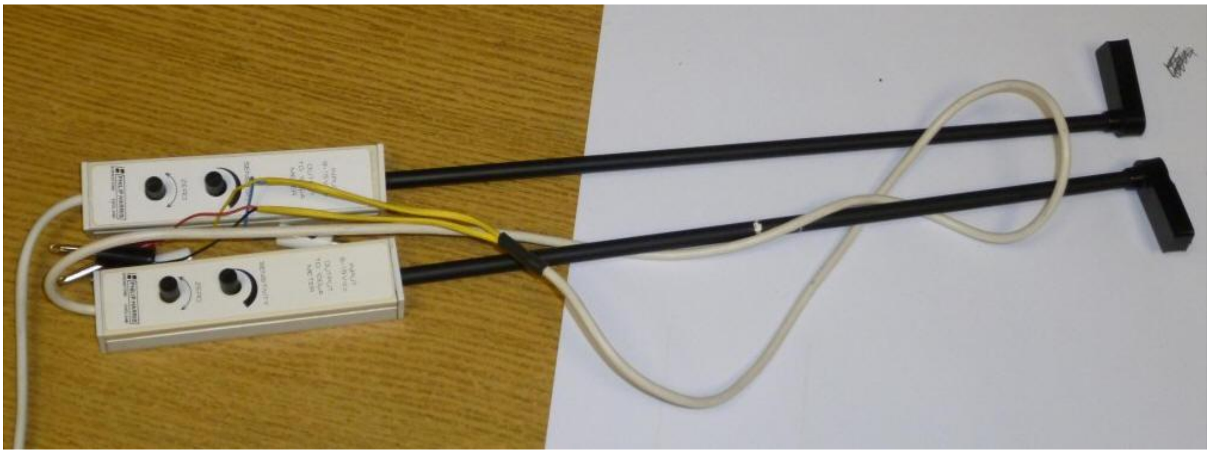


Figure 56 Hall probes for a datalogger

Anti-lock Braking System (ABS)

These devices are fitted to most cars nowadays. While tyres have a high coefficient of friction on tarmac, it is still possible for the wheels to lock. The car goes into a skid. At best this results in marks on the road and a patch of rubber scraped off the tyres. At worst, the results can be tragic.

A Hall sensor is at the heart of the **ABS system**. There is a rotating ring with magnets in on each wheel. The Hall sensor gives out a voltage each time the magnets pass. The voltage is detected by a **microcontroller** or computer that compares the voltage pulses coming from each wheel while braking is happening. The **computer** has an output that is connected to a valve housing that looks like this (*Figure 57*):

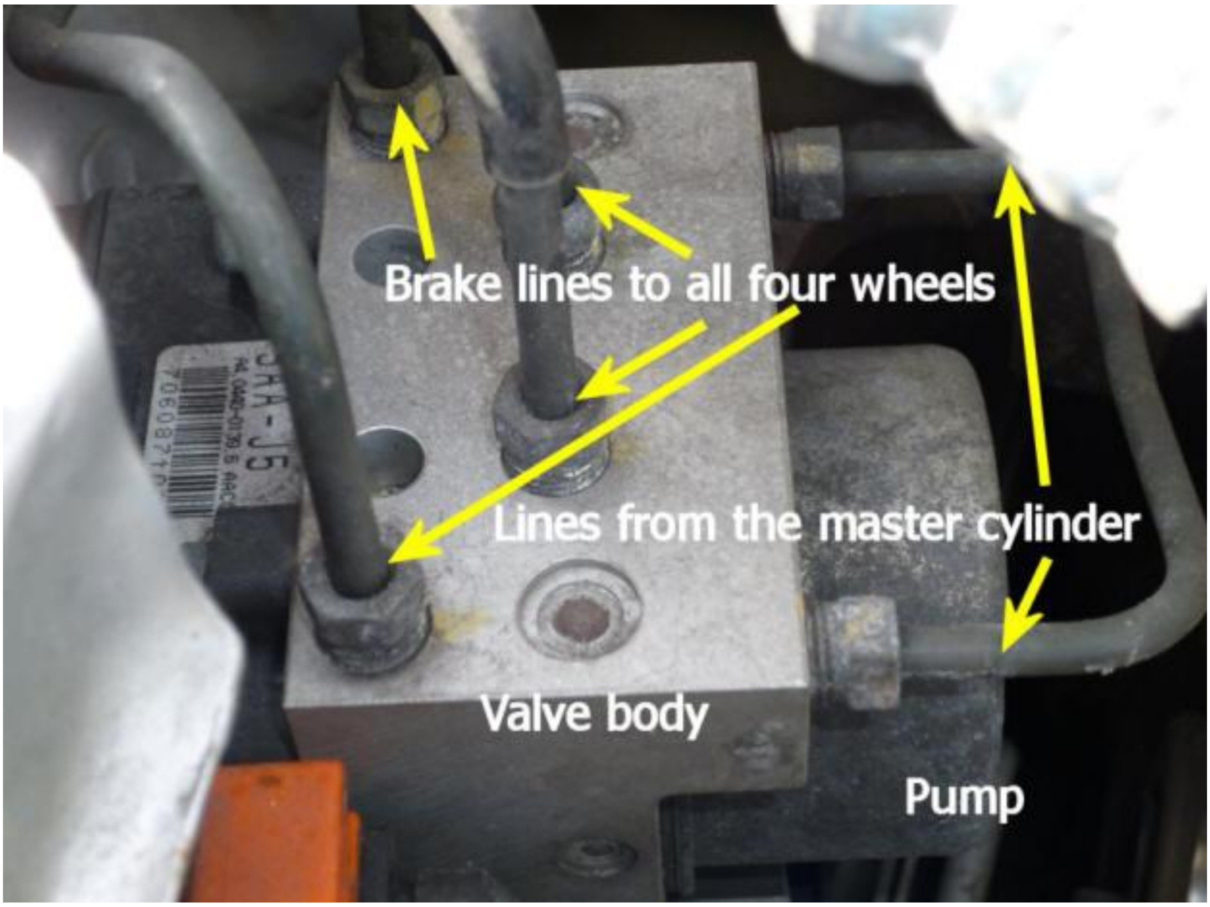


Figure 57 Antilock braking system on a car

When one of the wheels turns slower than the others, the situation is detected by the **microcontroller,** and a valve isolates the brake-line from the master cylinder (which is connected to the brake pedal). If the wheel is still turning slower, the brake line is released. Then a pump is automatically turned on to restore the pressure, so that the wheel resumes its braking action. The ABS can be felt operating by an unpleasant juddering. (If you have experienced this while driving, you will know what I mean.) The thing NOT to do is to pump the brake as you would if the car didn't have ABS. Let the system do the controlling for you.

A good driver should rarely have to brake hard, let alone make a crash stop. But sometimes the unexpected happens to the most experienced driver. (It happened to me while I was driving down a hill in snowy conditions and the ABS brought the car under control, before it ended up as a 1200 kg toboggan.)

Finding the Earth's Magnetic Field Using a Hall Sensor

The small Hall sensor in a data-logger probe is not sensitive enough to detect the Earth's magnetic field. However, it is possible to have a Hall sensor that is large enough to be sufficiently sensitive. The Earth's magnetic field is in three dimensions. The Hall sensor can pick up the value of a perpendicular magnetic field in one dimension only, so you need to take measurements for the x, y, and z components (*Figure 58*).

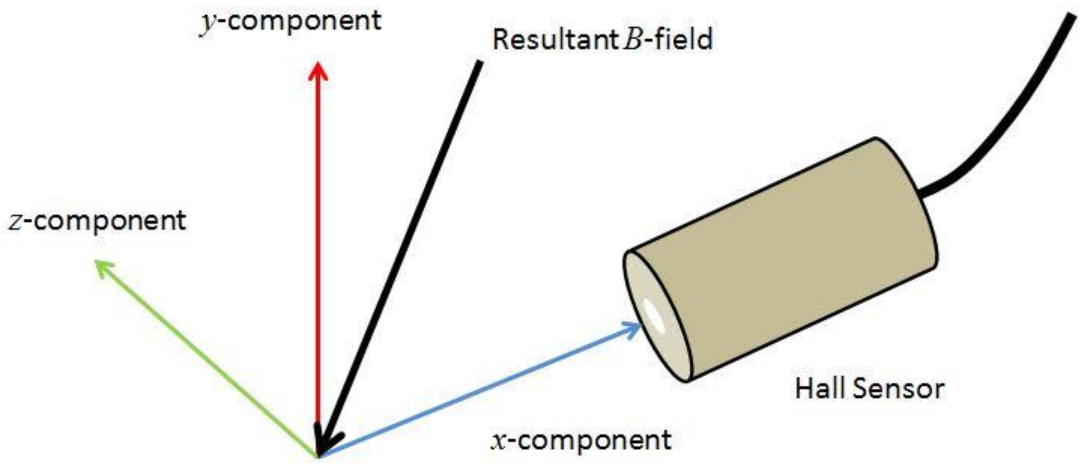


Figure 58 Using a Hall sensor to work out the Earth's magnetic field

The resultant magnetic field is worked out using:

$$B_R^2 = B_x^2 + B_y^2 + B_z^2$$

.... *Equation 20*

The directions and angles can also be worked out.

A Hall probe for a data-logger can be used to investigate a more powerful magnetic field in a similar way.

In some mobile telephones, there are Hall sensors that can give you the value of the Earth's magnetic field. The value of the Earth's magnetic field is about 6.5 × 10^{-5} T but can be half that value in some places.

Hall Sensors in Electronic Circuits

The Hall sensor in an electronic circuit can be connected to a variety of processing devices. In this case (*Figure 59*), it's connected to an **operational amplifier**.

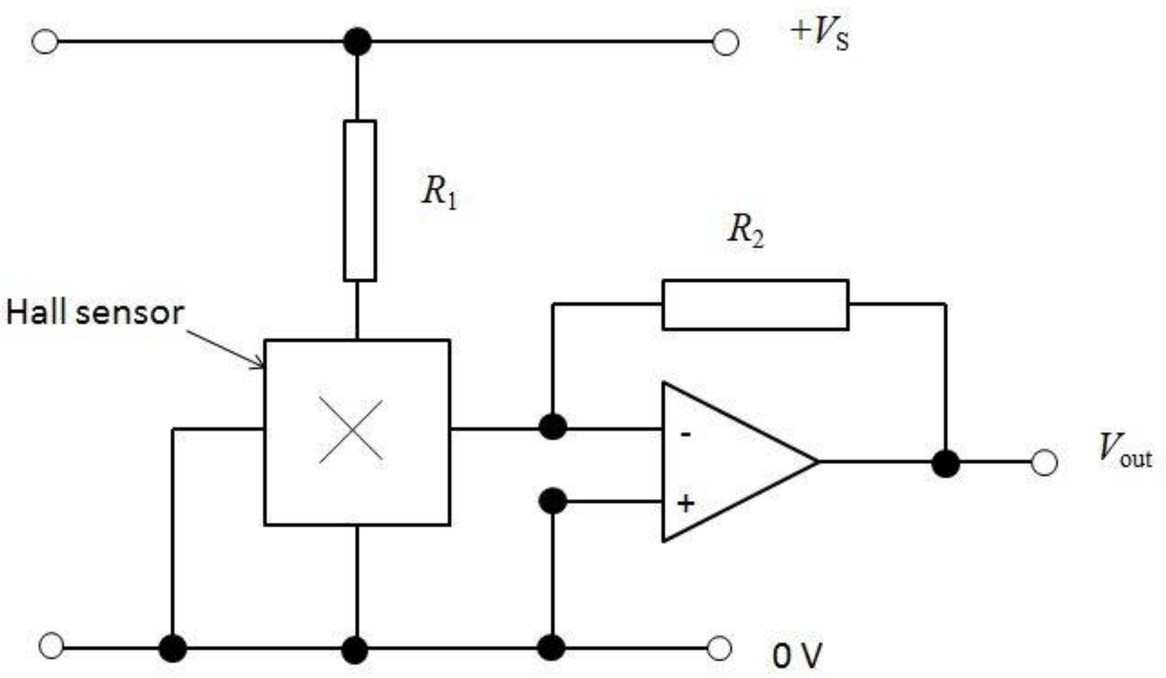


Figure 59 Hall sensor connected to an operational amplifier

We will consider the behaviour of an operational amplifier in a later tutorial, but it will detect the output from the Hall sensor to give an output voltage.

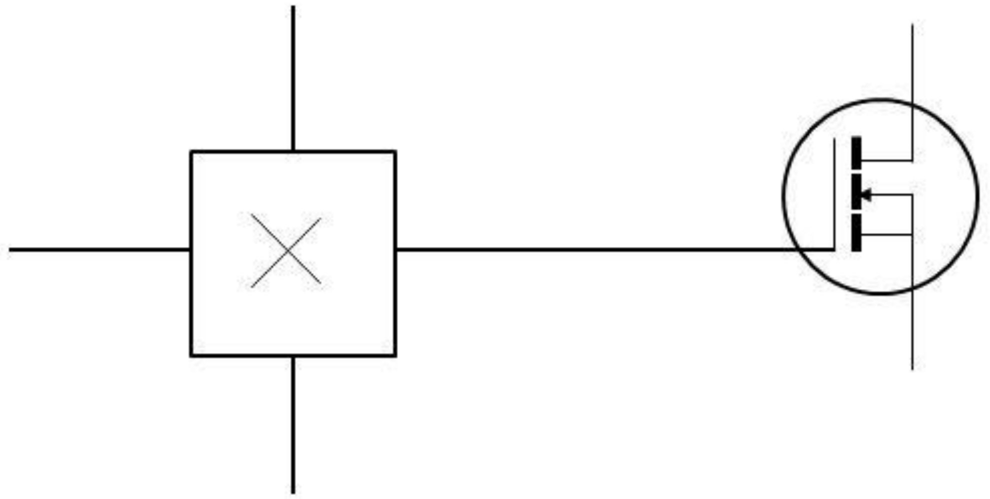
Hall sensors are often packaged with an amplifier in a **chip** to give a higher voltage and current, so that they can be used with a MOSFET or a bipolar transistor.

Questions

Tutorial 14E.04

14E.04.1

A student set up a Hall sensor so that the output was connected to the gate of a MOSFET as shown, instead of an operational amplifier.



Explain whether or not this circuit will work.

2. Analogue and Digital Signals

Tutorial 14 E.05 Digital and Analogue Signals

AQA Syllabus

Contents

14E.051 What is the difference between digital and analogue signals?

We see and hear things in **analogue**. The eye can adjust to any brightness from just about detectable to blinding. It is estimated that there are 10^9 different light levels to which the eye can adapt. Below the range, the light will not be detected; above it, damage will occur. Colours are in a continuous **spectrum** as well. Similarly for sound, the ear can detect an amplitude of about the width of an atom (about 10^{-10} m) at a sound level of 0 dB up to the 120 dB which is the threshold of pain and will cause hearing damage. The intensity difference is 10^{12} times.

A sound signal is picked up by a **microphone**, which converts the sound levels and frequencies to an analogue electrical signal which can have any value between $+V_S$ and $-V_S$. The picture below (*Figure 60*) shows a typical **analogue waveform**.

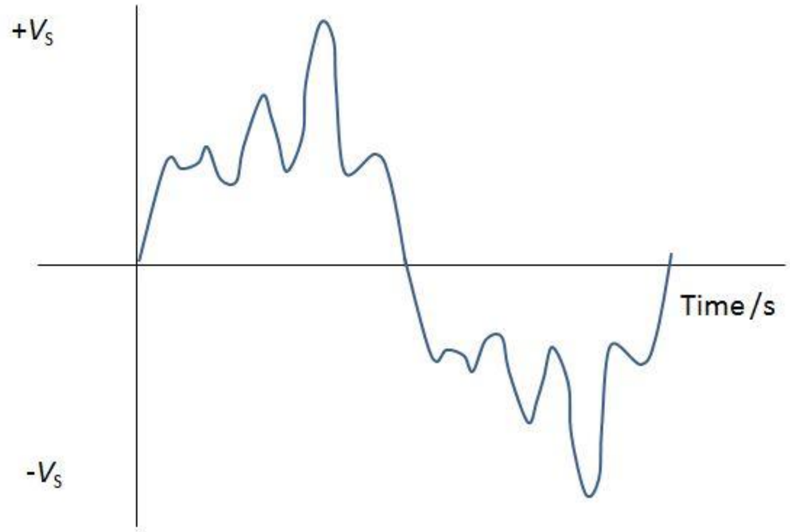


Figure 60 An analogue waveform

Digital signals are simpler. They are either **ON** (1 or HIGH) or **OFF** (0 or LOW). In theory it is possible to be an expert on digital electronics without knowing anything about electricity at all, other than the difference between ON and OFF. In practice, some knowledge does help! Digital signals come in pulses that are 1 (+5.0 V) and 0 (0 V). In practice anything below about 2 V is considered to be LOW. Here is a digital waveform (*Figure 61*):

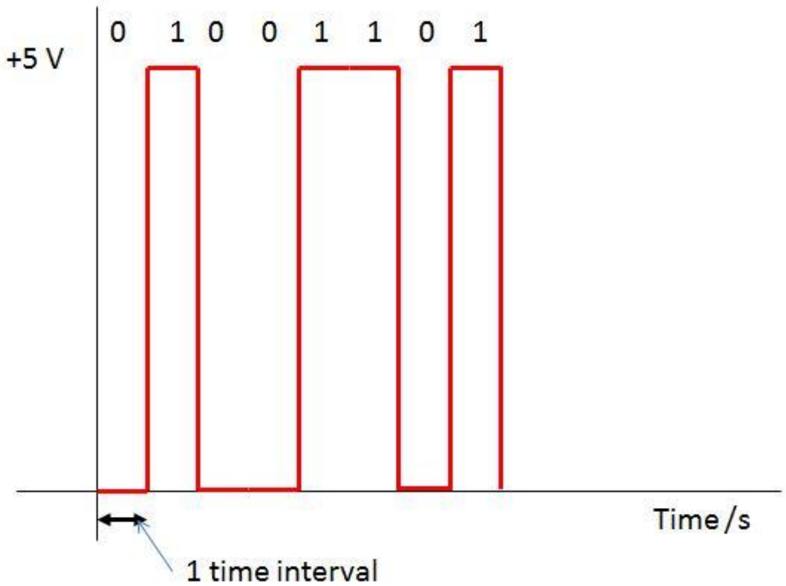


Figure 61 A digital signal

It represents the binary number 01001101 (which is 77).

Binary numbers (numbers to the base 2) are essential for **digital data processing**. Many years ago, attempts were made at analogue computing, but nowadays all computers are digital. Computers are **adding machines**. They can only add up binary numbers. They cannot subtract. Subtraction is done by a process of complementary addition. Multiplication is done by serial addition. Division is done by serial complementary addition. You may be thinking that your computer can do very sophisticated games. It's all done by binary addition. The computer can do it at a very fast rate.

14E.052 Bits and Bytes

Computers use **bits** and **bytes**.

A **bit** is a binary digit, i.e. a 0 or a 1. The binary number 11 is a two-bit number.

2-bit numbers are not much use. This picture (*Figure 62*) is in 2-bit colour:

Figure 62 2-bit colour

The colours in the next picture (*Figure 63*) are 4-bit, so there are 16 different colours:

Figure 63 4-bit colour

The earliest computers used 8-bit numbers, which are called **bytes**. These are from 00000000 to 11111111 (0 to 255).

For an 8-bit number there are 256 (2^8) combinations. If each number represented a single colour, we would get a colour palette that looked like this (*Figure 64*):

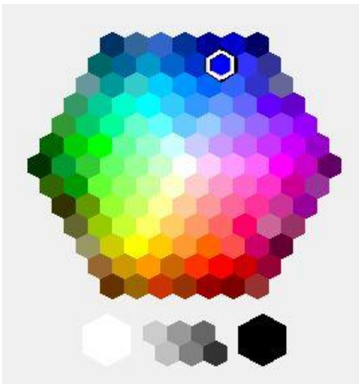


Figure 64 8-bit colour palette

An 8-bit colour picture looks like this (*Figure 65*):

Figure 65 8-bit colour

It's slightly better than the one before. But this picture (the original) is a 32-bit image (*Figure 66*). A 32-bit number is sometimes called a four-byte **word**.

Figure 66 32-bit colour

Modern computers use **64-bit** numbers or 8-byte words. This gives 18446744073709551616 combinations, i.e., 1.845×10^{19}, which is quite a lot.

In the earliest days, memories for computers were in kilobytes (kB). 1 megabyte (1 MB) was considered very large. Now files of up to 100 MB are routine. PC games may have a central program which is 1 gigabyte (1 GB) or more. Hard-disc drives of 1 terabyte (1 TB) are easily available.

File size	***Bytes***	***Bits***	***Power of 2***
1 byte	1	8	0
1 kilobyte	1024	8192	10
1 megabyte	1 048 578	8 388 624	20
1 gigabyte	1 073 741 824	8 589 934 592	30
1 terabyte	1 099 511 627 776	8 796 093 022 208	40

When looking at internet speeds, these are quoted in megabits per second. A 1 megabit per second download speed is 131 kB s^{-1} (not very fast).

14E.053 Using Binary Numbers

Although you are not expected to know more than 1 - 10 in binary for this syllabus, it's worth mentioning that base 10 is not used in the computing world. Instead, **hexadecima**l (base-16) is the counting system that computer experts use. This table below shows the systems in 4-bit and 8-bit numbers:

4-bit binary	*8-bit binary*	*Decimal*	*Hexadecimal*
0000	00000000	0	0
0001	00000001	1	1
0010	00000010	2	2
0011	00000011	3	3
0100	00000100	4	4
0101	00000101	5	5
0110	00000110	6	6
0111	00000111	7	7
1000	00001000	8	8
1001	00001001	9	9
1010	00001010	10	A
1011	00001011	11	B
1100	00001100	12	C
1101	00001101	13	D
1110	00001110	14	E
1111	00001111	15	F

14E.054 Sensors

You will have used data-loggers at some stage in your A-level sciences. In Physics, the obvious sensors are voltmeter and ammeter sensors. You may have used position sensors, force sensors, light sensors, and temperature sensors. In Chemistry, you may have used a pH sensor. The picture below shows an infra-red sensor (*Figure 67*):

Figure 67 Infrared sensor for datalogging

All sensors detect an **environmental factor**. These can be any value, so are analogue signals. They have to be converted to digital signals to be used by the computer. The picture shows a typical **interface box**. This one can be used as a stand-alone data-logger, or it can be connected to a computer (*Figure 68*):

Figure 68 An analogue to digital converter

14E.055 Analogue to Digital Conversion

Suppose we wanted to convert the analogue signal in the graph into digital signals (*Figure 69*):

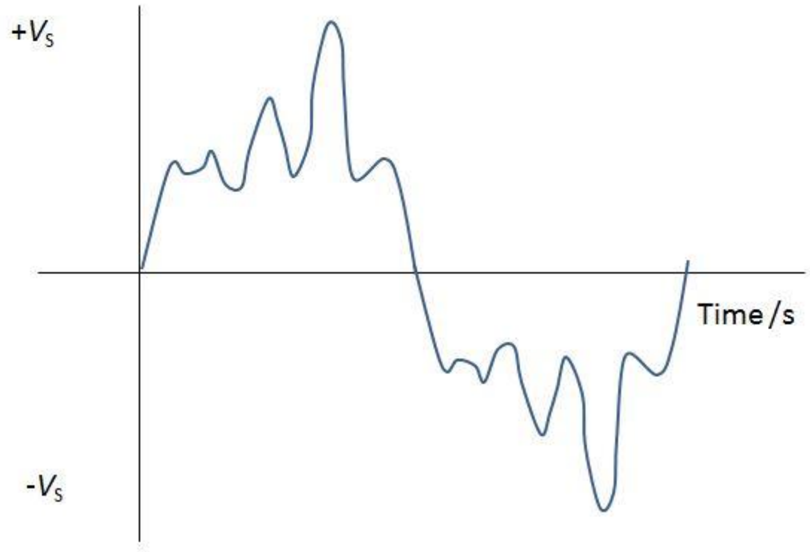


Figure 69 An analogue signal

The first thing we need to do is make sure that the voltage is always positive. We do this by moving the 0 V potential to $+V_s/2$ (*Figure 70*):

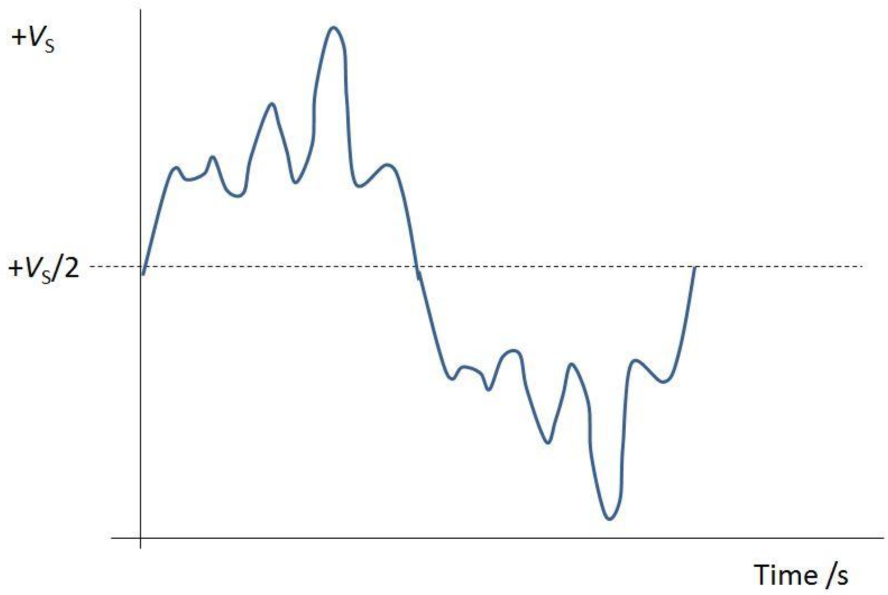


Figure 70 Making the analogue signal positive

We need to represent the voltage as a number of different levels. This is called **quantisation**. 1-bit quantisation gives us 2 levels, 0, and 1 (*Figure 71*):

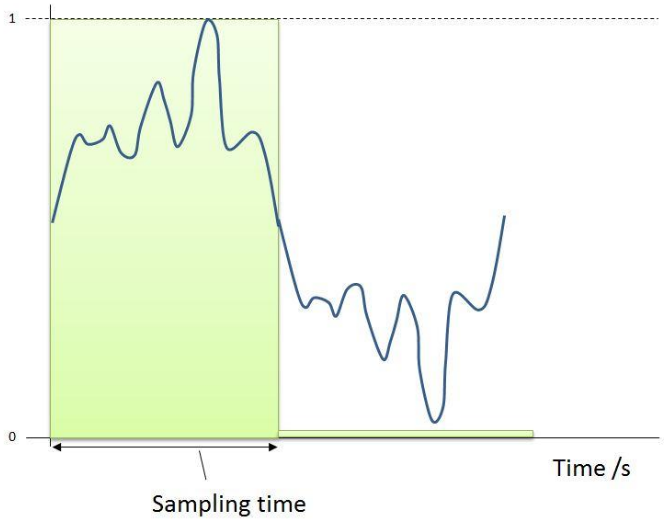


Figure 71 One bit quantisation

This would not give us a particularly useful signal. If we decreased the sampling time, we would get the same result, because there is no level 1/2. Each sample gives out a 0 or a 1. In your answer to Question 14E.05.6, you will have got a line of 0 and 1.

2-bit quantisation gives us 4 levels, 00, 01, 10, and 11. 4-bit quantisation gives us 16 (2^4) levels (*Figure 72*):

Figure 72 4-bit quantisation

We also need to decide the frequency at which we determine the levels. This is called the **sampling rate**. The level is determined to the nearest whole number. We cannot have a sample that gives us 0110 and a half! The sampled signal looks like this (*Figure 73*):

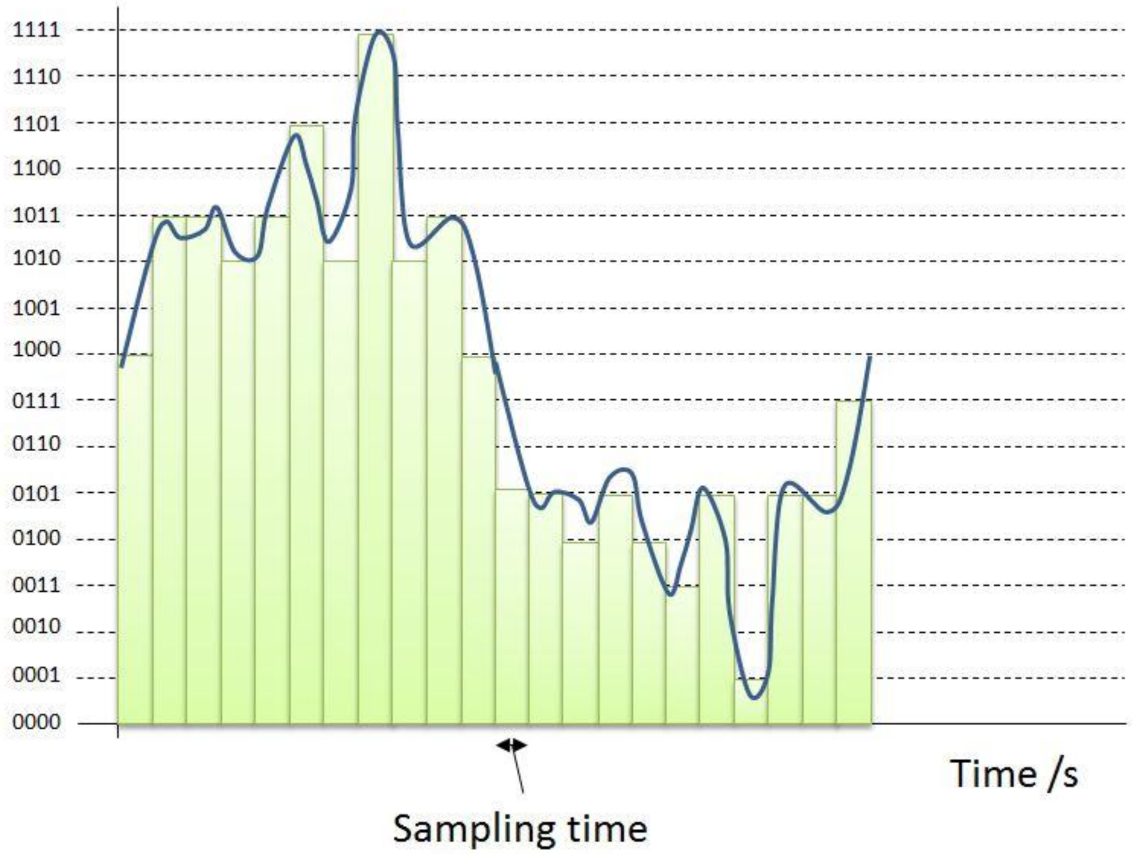


Figure 73 Sampling the analogue wave

In this case, there are many more levels, and the sampling time is shorter. The term for converting analogue to digital signals in this way is **pulse code modulation**.

This digital output would give a very distorted sound when converted back to analogue. We can improve the sound by converting to 16-bit samples. This is sometimes called the **bit depth.** A bit depth of 16 gives 2^{16} levels which results in 65536 levels. We can also increase the sampling rate (reducing the sampling time). The current standards for a CD are 16-bit bit depth and 44 kHz sampling rate. Improved audio performance can be achieved by increasing the sampling rate. Audio engineers reckon that the sampling rate should be twice the highest frequency.

The DVD standard is bit-depth of 24-bits. This gives 16 777 216 levels, resulting in a much higher resolution. The sampling rate is 96 kHz.

14E.056 Noise

In analogue systems, there is always a certain amount of random and unwanted signals, which are called **noise**. They result from the electronic components in the amplifier and are due a number of factors such as temperature, junction reverse breakdown, causing avalanches of electron. Stray electromagnetic fields, such as those caused by lightning are common sources of noise.

Noise is most commonly heard as an intrusive **hiss**, especially when an amplifier is turned up. Poorly shielded mains transformers lead to hum (100 Hz). Random clicks can be the result of lightning discharges. If a signal is amplified at several different stages, the noise can become a significant part of the signal. If the noise is severe, it can detract from the enjoyment of the music. If the signal is weak, it can make it unintelligible.

A well-designed analogue amplifier will reduce the noise to a minimum but not get rid of it completely.

Digital circuits have the advantage of **reducing** the noise, because the additional little waves made by the noise are not quantised.

14E.057 Advantages of Digital Signals

Digital signals can carry **much more information** than an analogue signal. When you use a DAB radio, you tune into one frequency, 225.648 MHz. On this wave, digital signals from a large number of radio stations are carried using a process of **multiplexing**. The precise details are beyond the scope of these notes. Similarly, huge amounts of digital data can be carried along an optical fibre cable. The signals in optical fibres deteriorate less than they do in wires.

Digital signals can be **cleaned up** more easily than analogue signals.

14E.058 Cleaning Digital Signals

A problem with optical fibres is that the signal tends to get **smeared**. The idea is shown below (*Figure 74*):

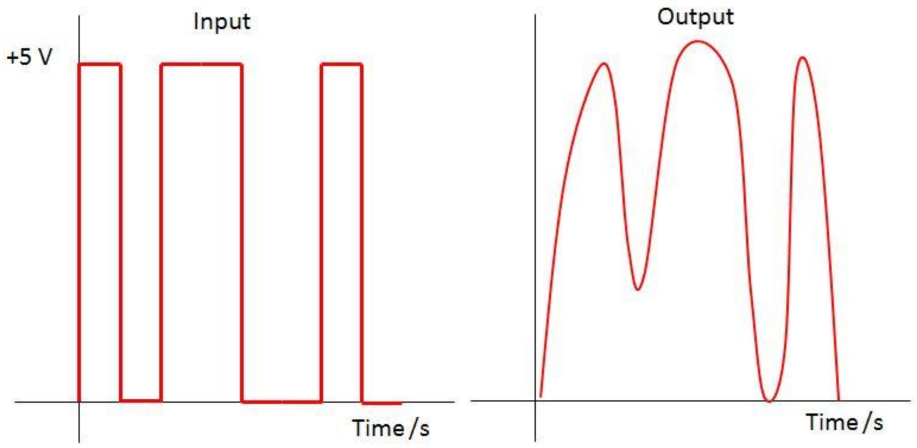


Figure 74 Smearing of square waves

By using clipping circuits and operational amplifiers, the distorted waves are converted back into square waves.

The same train of pulses is shown with some noise that it has picked up (*Figure 75*). Just because the signal is digital, it does NOT mean that it never gets noise. Noise is easier to get rid of.

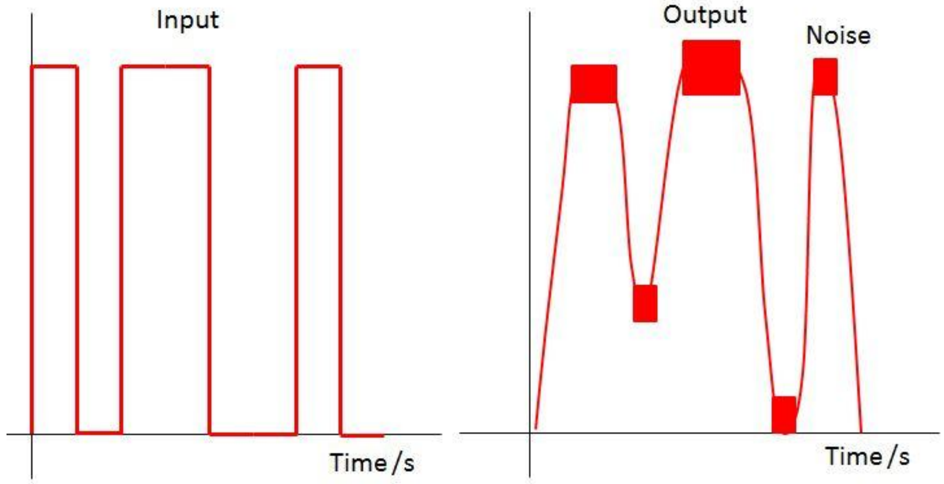


Figure 75 Noise in a square wave

The noise is removed by passing the signal through a **filter** that removes very high frequency signals.

14E.059 Data Storage

A major advantage of digital data processing is **data storage**.

The most obvious storage medium for an **analogue recording** is the **LP record**. This is a 30 cm disc made of polyvinyl chloride, onto which a continuous groove is cut. About 25 minutes of music can be stored on each side. The record has to be turned over. It is easily damaged by scratches and dust. A poor quality player can also damage the track. However, vinyl LPs have recently made a comeback, because many people think they sound better.

An alternative is ferromagnetic tape. The trouble is with tape is that there was noticeable hiss. To get a decent recording, the tape has to move past a recording head at a considerable speed. Studio quality reel-to-reel tape decks are very expensive. The cassette deck that your parents would have had as teens did not allow for high quality recordings, even though a hiss reduction system (Dolby) was present in good quality machines. The cassette was never intended to be a high quality recording medium. It was developed for dictation machines in offices. Often the tape from the cassette could end up inside the machine all chewed up. The cassette deck is now a museum piece.

Your parents will also have used a video cassette recorder to record TV programmes off-air. For video recorders to work, the speed of tape moving past the head had to be several metres per second. In the VCR, this was achieved by an ingenious system of a rotating head. However, this could mis track and the results could be unwatchable.

For digital data storage, ferromagnetic media have been used since the earliest days. In the nineteen seventies Edinburgh University bought the largest possible hard-drive - 100 megabytes. In the earliest days of home computers, programs could be stored on cassette tapes. I even remember programs being broadcast on channels targeted at young people - they sounded like a swarm of angry bees stuck in a glass jar.

In your computer, you have a hard drive of anything from 500 GB to 2 TB. CD media can store up to 700 MB, while DVDs can store 4 GB

Solid state drives have been around for many years in the form of pen drives, which can hold 64 GB of data. The picture below (*Figure* 76) shows a variety of hard drives:

Figure 76 Data storage media

The one on the left is the HDD from and old computer. The one in the middle is an external HDD that I use to back up my data (I am paranoid about losing my data) and the pen-drive on the right is the one I used for the resources I used for teaching. Nowadays data are frequently stored on large central computers, **the cloud**. The data are kept secure.

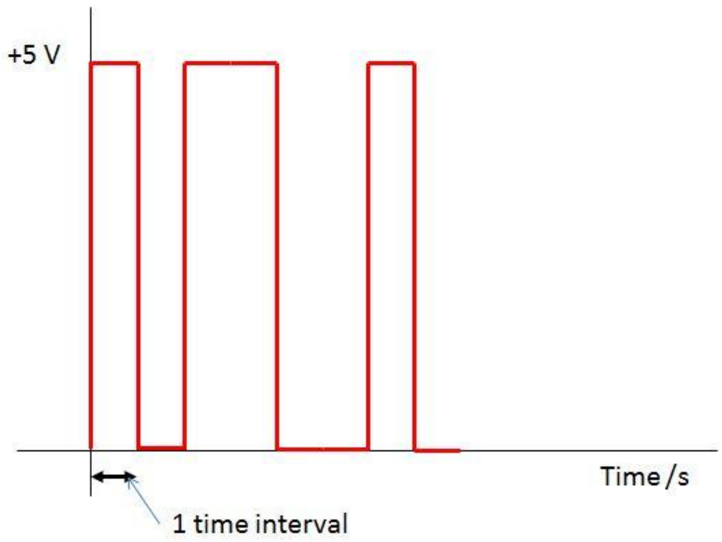
Digital media are not indestructible. CDs can mis track or even get stuck. Data can be **corrupted**. It only takes a 1 to become a 0 in a vital place for the data to be meaningless. There are ways that computer manufacturers have devised to check and repair data, but they may not work in every case. Hard disc drives like the ones above can be easily damaged by heat or by dropping onto a hard surface. Sometimes they just fail. They are complex and precision-made devices. It is a good idea to have your data backed up, so that if the worst happens, you do not lose them.

Questions

Tutorial 14E.05

14E.05.1

What binary number is represented below?



14E.05.2

How many levels can you get from a two-bit number?

14E.05.3

Show that binary 11111111 is the same as decimal 255.

14E.05.4

How many different colours can you get with a 32-bit colour palette?

14E .05.5

Convert the following hexadecimal numbers to binary and decimal:

6; C; 13; 3F

14E.05.6

If the sample time is decreased to 1/10 of what is shown above, what is the output?

14E.05.7

Refer to *Figure 73*. Write down the sequence of binary numbers that would form the analogue wave.

14E.05.8

Refer to *Figure 73*. The analogue wave has a frequency of 512 Hz (C above middle C). Calculate the sampling time.

Tutorial 14 E.06 LC Resonance	
AQA Syllabus	
Contents	
14E.061 Mechanical Resonance	14E.062 Electrical Resonance
14E.063 Series Resonance	14E.064 Parallel Resonance
14E.065 Resonant Frequency of the LC Circuit	14E.066 Q-factor
14E.067	14E.068
14E.069	

To make sense of this tutorial, you may want to review the sections in Topic 11 about alternating currents.

14E.061 Mechanical Resonance

The physics phenomenon of **resonance** can be easily demonstrated in a mechanical system. One example is pushing a child on a swing. You have to push the child at the right point, so that the child swings higher and higher (they like it). If you don't push at exactly the right time, the swings get less. We can demonstrate **mechanical resonance** with a **vibration generator** acting on a mass on a spring (*Figure 77*).

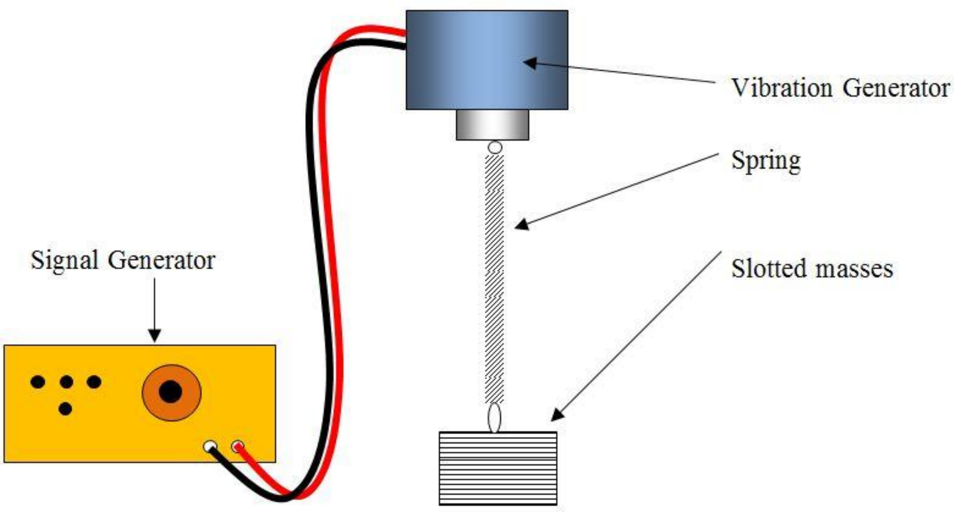


Figure 77 Demonstrating mechanical resonance

If we alter the frequency, we see that the mass bounces with varying **amplitude**. However, at the **resonant frequency**, the amplitude gets very large. It is not unknown for the masses to fly off! A typical value for the resonant frequency of this kind of system is about 1.5 Hz.

The condition for mechanical resonance is for the **natural frequency** (the frequency at which the spring bounces freely) to **equal** the **forcing frequency**.

Mechanical resonance can be very useful, but can also be a nuisance, or even destructive.

14E.062 Electrical Resonance

Electrical circuits can be made to **resonate**. If we have a circuit with an **inductor** and a **capacitor**, we find that at a certain frequency, the current goes to a very high value. This is e**lectrical resonance**. This phenomenon can be used in:

- Electrical filters;
- Tuned circuits, such as those found in a radio receiver.

The **capacitor** is analogous to the **spring**, while the **inductor** is analogous to the **mass** in the mechanical resonance system.

14E.063 Series Resonance

Consider a pure **LC circuit**, i.e. one with no resistance at all. It is connected to a supply voltage V at a frequency of f Hz (*Figure 78*).

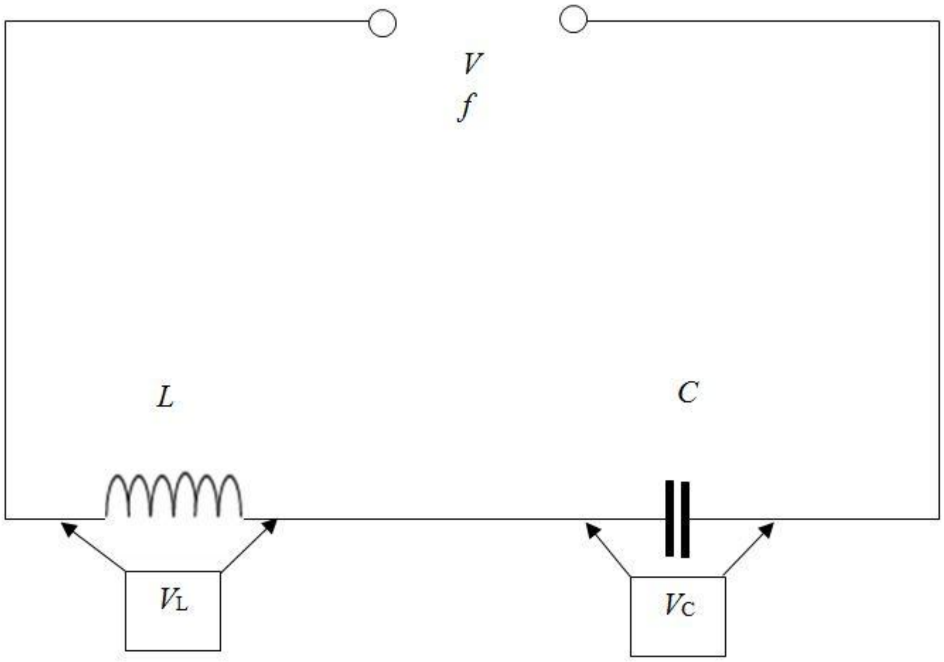


Figure 78 Series LC circuit

The current is the same throughout. The voltage V_L is related to the reactance in the inductor by the equation:

$$V_L = IX_L$$ .... *Equation 21*

The voltage V_C is related to the reactance in the capacitor by the equation:

$$V_C = IX_C$$ .... *Equation 22*

The voltages add up **vectorially**. Since there is no resistive element, there is no resistive voltage.

Therefore, the phasor diagram looks like this (*Figure 79*).

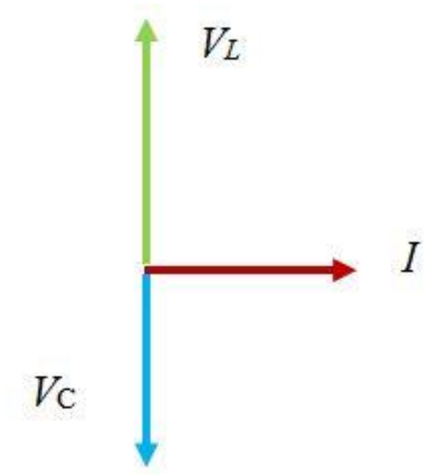


Figure 79 Phasor diagram

Now consider what happens as we change the frequency:
• The reactance of the capacitor gets less as the frequency goes up.
• The reactance of the inductor goes up as the frequency goes up.

This stands to reason as the reactance of the capacitor is given by:

$$X_C = \frac{1}{2\pi f C}$$ .... *Equation 23*

And the reactance of the inductor is given by:

$$X_L = 2\pi f L$$ .... *Equation 24*

Consider a circuit with a pure inductance of **0.15 mH** and a capacitor of **47 μF**. If we use the values in the equations to generate data for a graph, we see (*Figure 80*):

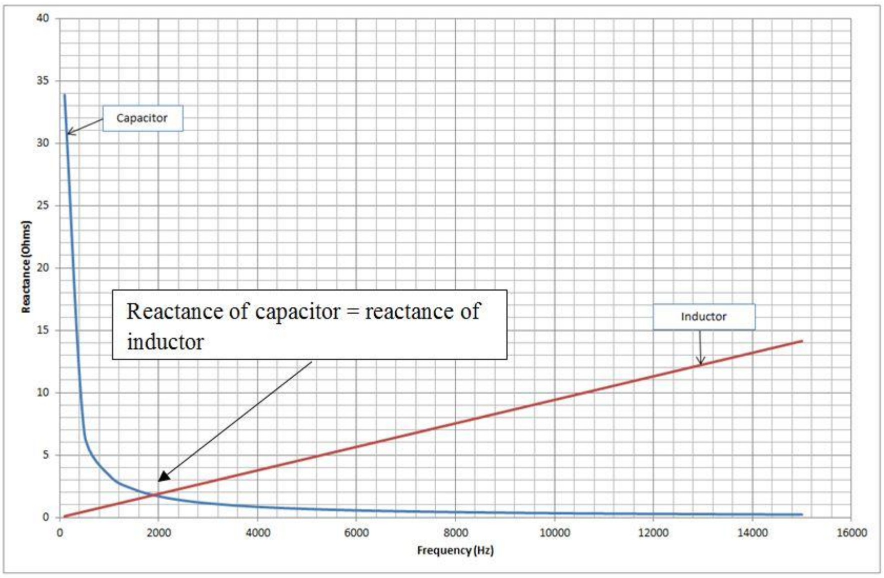


Figure 80 Graph of reactance against frequency

At a certain point, about 1900 Hz in this case, we find that:

the reactance of the capacitor = reactance of the inductor.

If we look at the voltage phase vectors, we see that they are of **equal magnitude** and **opposite directions**. So, they add up to 0.

Since the potential difference is 0, and a current is flowing, we can say that the **impedance** (the vector sum of the reactances) is 0.

The graph of impedance against frequency looks like this (*Figure 81*):

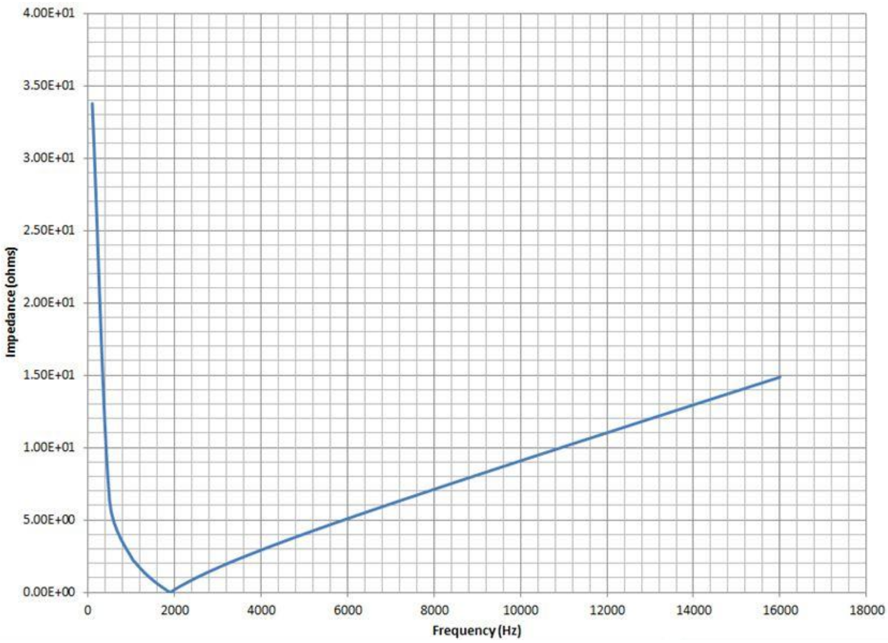


Figure 81 Graph of impedance against frequency

The resonant frequency occurs when the impedance is zero. In this case it is 1900 Hz. If we plot the current against the frequency, we get (*Figure 82*):

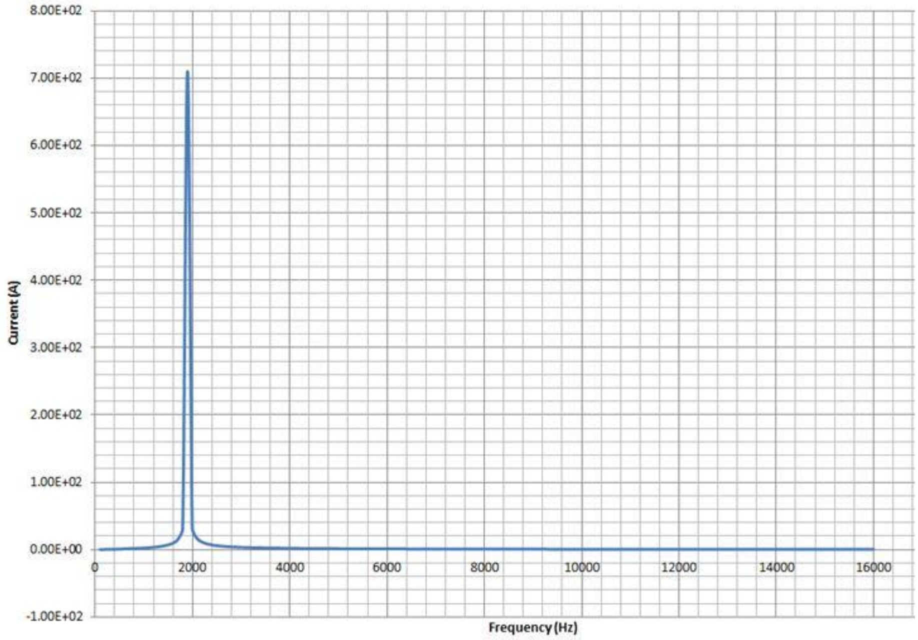


Figure 82 Graph of current against frequency

You can see there is a spike in the current at resonance, somewhere at about 1900 Hz. The peak current is about 710 A. The spike is shown in more detail in this graph (*Figure 83*):

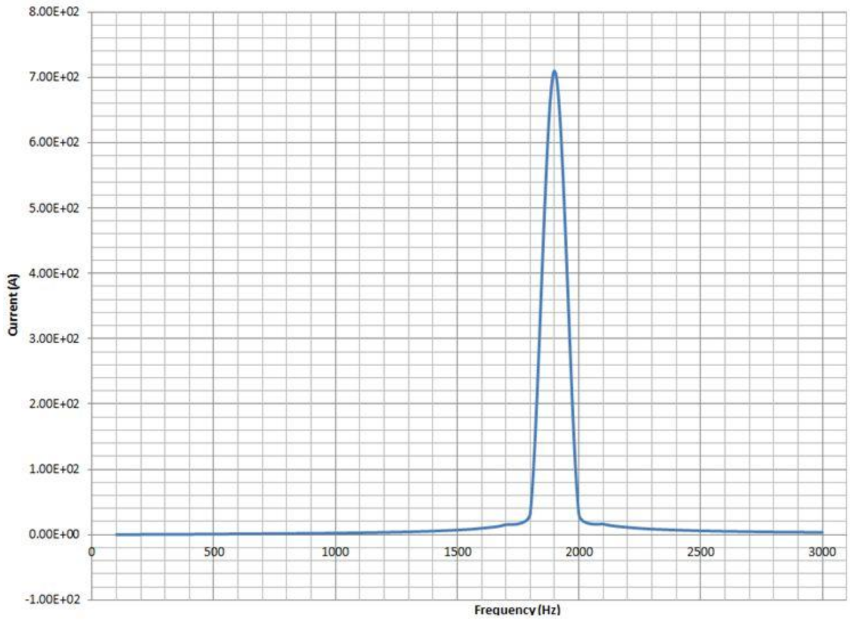


Figure 83 Closer view of the peak current in resonance

The value of the peak shown in the graph was produced by **data modelling**, not by direct observation in the lab. Signal generators that can give out such a current are not available, certainly not in the school or college physics lab. In theory the current should be **infinite** at resonance. In practice this would not occur, as there is **resistance** in the inductor and the wires. The maximum current would be determined by the resistance of the inductor.

14E.064 Parallel Resonance

Consider this circuit (*Figure 84*) that consists of an inductor of inductance L, and capacitor of capacitance C. It is connected to an alternating voltage V that has a frequency of f.

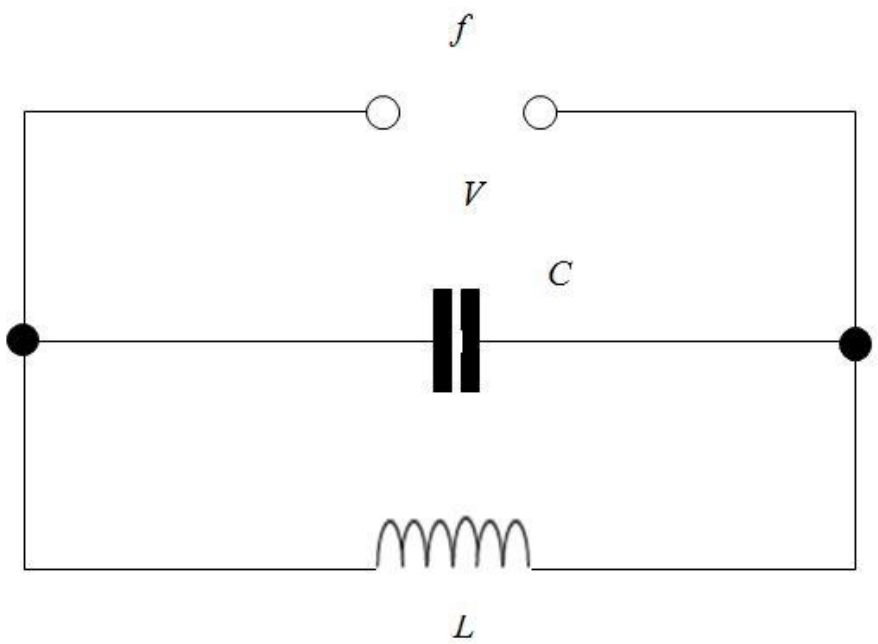


Figure 84 Circuit to show parallel resonance

We know:

- The voltage across L and C is the same (V);
- The currents add up vectorially;
- The current I_C leads the voltage by 90 degrees;
- The current I_L lags the voltage by 90 degrees.

We can show this using the phase vector diagram (*Figure 85*):

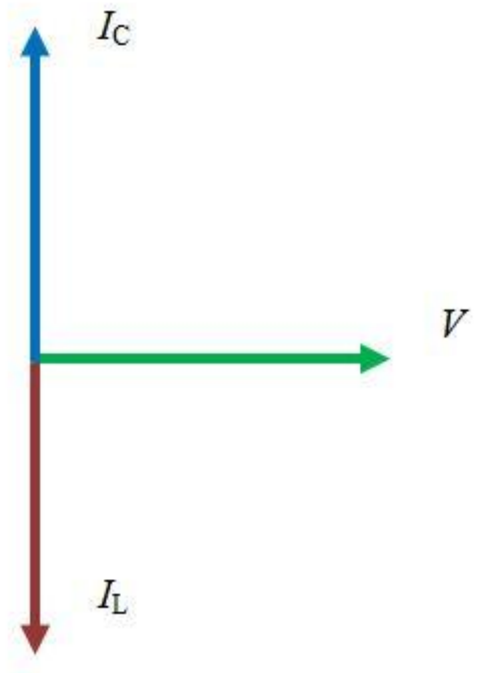


Figure 85 Phasor diagram showing the phase relationships between voltage and currents

We can see that the total current, I_T is the difference between I_C and I_L. This is because the current vectors are pointing in the **opposite direction**. Therefore:

$$I_T = I_C - I_L$$ .... *Equation 25*

We know that:

$$I = \frac{V}{X}$$ .... *Equation 26*

We know that if X is bigger, I will be smaller. We can explain this in terms of the reactances and the frequency. As frequency goes up, the reactance of a capacitor goes down, while the reactance of the inductor goes up. Therefore, as frequency goes up, I_C gets bigger, while I_L gets smaller.

If we change the frequency, we have seen that the currents change with the reactances of the inductor and the capacitor. There comes a point where the reactance of the capacitor and the reactance of the inductor are equal, i.e.:

$$X_C = X_L$$

.... Equation 27

It doesn't take a genius to see that if the voltages and reactances are the same, the currents are the same. However the directions are **opposite**, so the currents cancel out to 0.

We have achieved **parallel resonance**. We can see this in the graph below (*Figure 86*):

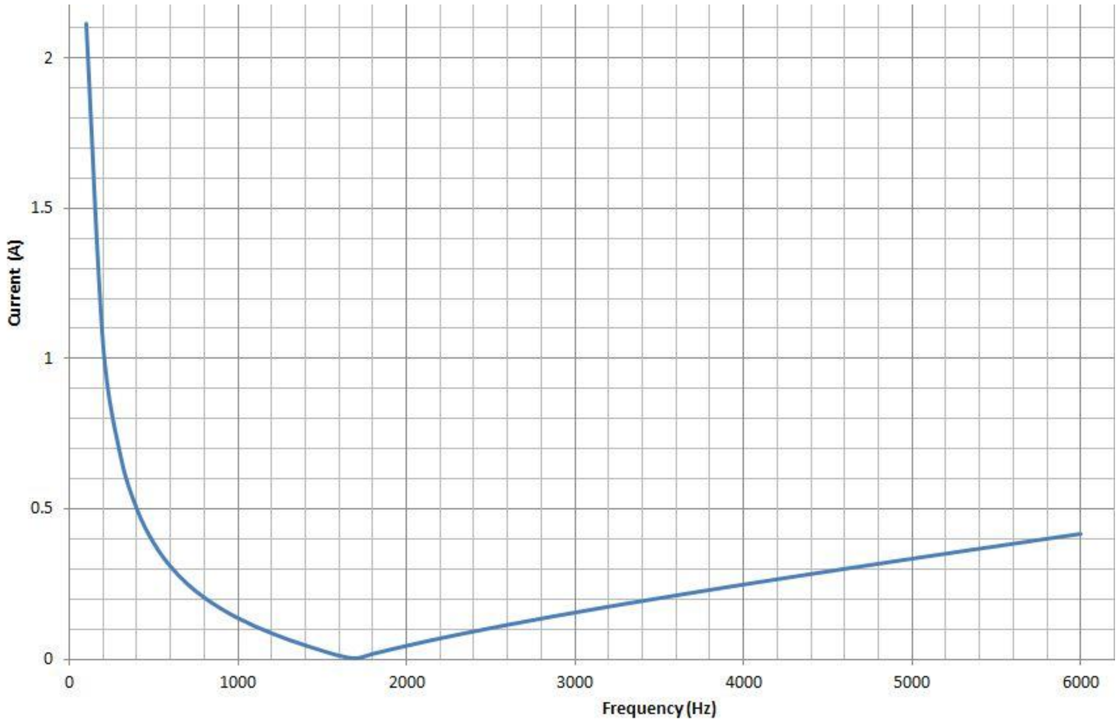


Figure 86 Graph of current against frequency

At the **resonant frequency**, the current falls to (almost) zero. The corresponding impedance graph looks like this (*Figure 87*):

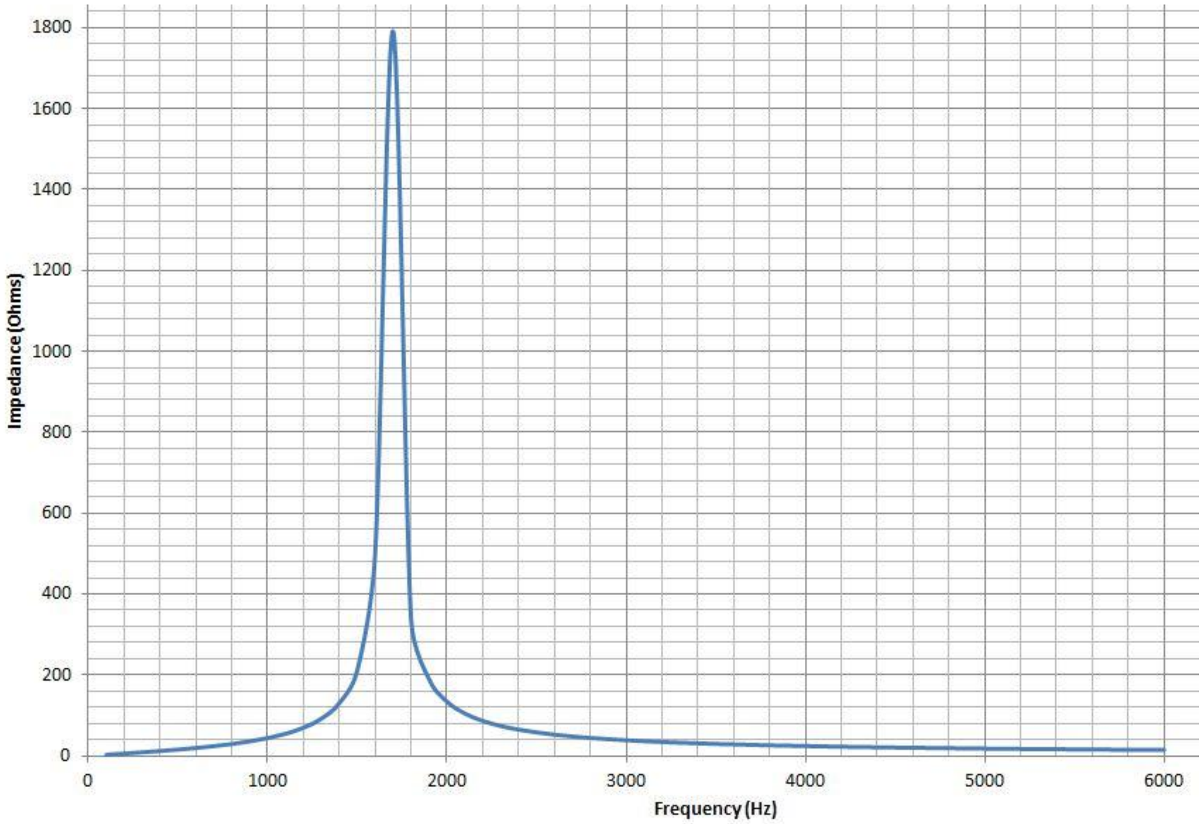


Figure 87 Graph showing variation of impedance with frequency

Here we see that the impedance goes up to a high value at the same frequency as the current drops to a very low value. Although these data were obtained using data-modelling, a school frequency generator can quite easily produce a very low current, so parallel resonance can be studied easily in the school or college Physics lab.

I have included both series and parallel resonance here, although the AQA syllabus requires only parallel resonance. Other syllabuses require series resonance as well. To sum up:

- Series resonance results in the current being a very high value, and the impedance being a very low value.
- Parallel resonance results in the current being almost zero, with the impedance being very high.

14E.065 Resonant Frequency of the LC Circuit

Let's look at how we can find the precise frequency at which **resonance** occurs. We know that:

- the reactance of a **capacitor** is given by:

$$X_C = \frac{1}{2\pi f C}$$

.... *Equation 28*

- and that the reactance of an **inductor** is given by:

$$X_L = 2\pi f L$$

.... *Equation 29*

We know that in resonance:

$$X_L + -X_C = 0$$

.... *Equation 30*

Therefore:

$$2\pi f L = \frac{1}{2\pi f C}$$

.... *Equation 31*

We can rearrange this to give:

$$4\pi^2 f^2 = \frac{1}{LC}$$

.... *Equation 32*

This will give us an expression to give us the **resonant frequency**:

$$f^2 = \frac{1}{4\pi^2 LC}$$

.... *Equation 33*

This can be **square rooted** to give:

$$f = \frac{1}{2\pi\sqrt{(LC)}}$$

.... *Equation 34*

Let's substitute the values (see *Page 85*) to get the true resonant frequency for the case we looked at above:

$$f = \frac{1}{2\pi \times \sqrt{(0.15 \times 10^{-3} \times 47 \times 10^{-6})}} = 1896 \text{ Hz}$$

.........*Equation 35*

This is consistent with what we saw on the graph.

You are not expected to know this derivation for the exam. The same argument can be applied for both series and parallel resonance.

If there is **resistance** in the LC circuit, the resonance is damped. The resistance is analogous to the card you stick onto the slotted mass when you are investigating damped resonance.

14E.066 Q-factor

The voltage magnification at resonance is called the ***Q* factor**, which is defined as:

The ratio between the inductor voltage and the supply voltage

So we write:

$$Q = \frac{\text{voltage across } L}{\text{Supply voltage}}$$

.... *Equation 36*

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

.... *Equation 37*

It's a ratio, so has no units.

Q-factor can also be defined in energy terms:

The ratio between the stored energy per cycle and the supplied energy per cycle

This is entirely consistent with the voltage definition as voltage is **energy per unit charge**. Therefore:

$$Q = \frac{\text{energy stored per cycle}}{\text{energy supplied per cycle}}$$ .... *Equation 38*

We know that:

$$E = Pt$$ .... *Equation 39*

For **one** cycle, we can rewrite this as:

$$E = PT$$ .... *Equation 40*

Since for the period of one cycle, T

$$T = \frac{1}{f}$$ .... *Equation 41*

we can write an expression for the energy per cycle:

$$E = \frac{P}{f}$$ .... *Equation 42*

It can be shown that the Q factor can be related to the **bandwidth** of the system. The bandwidth is defined as:

the range of values in which the energy is above 50 % of the maximum energy per cycle.

Since

$$P = I^2 X_L$$ .... *Equation 43*

the current would be 0.707 of the maximum value. This is shown on the graph (*Figure 88*):

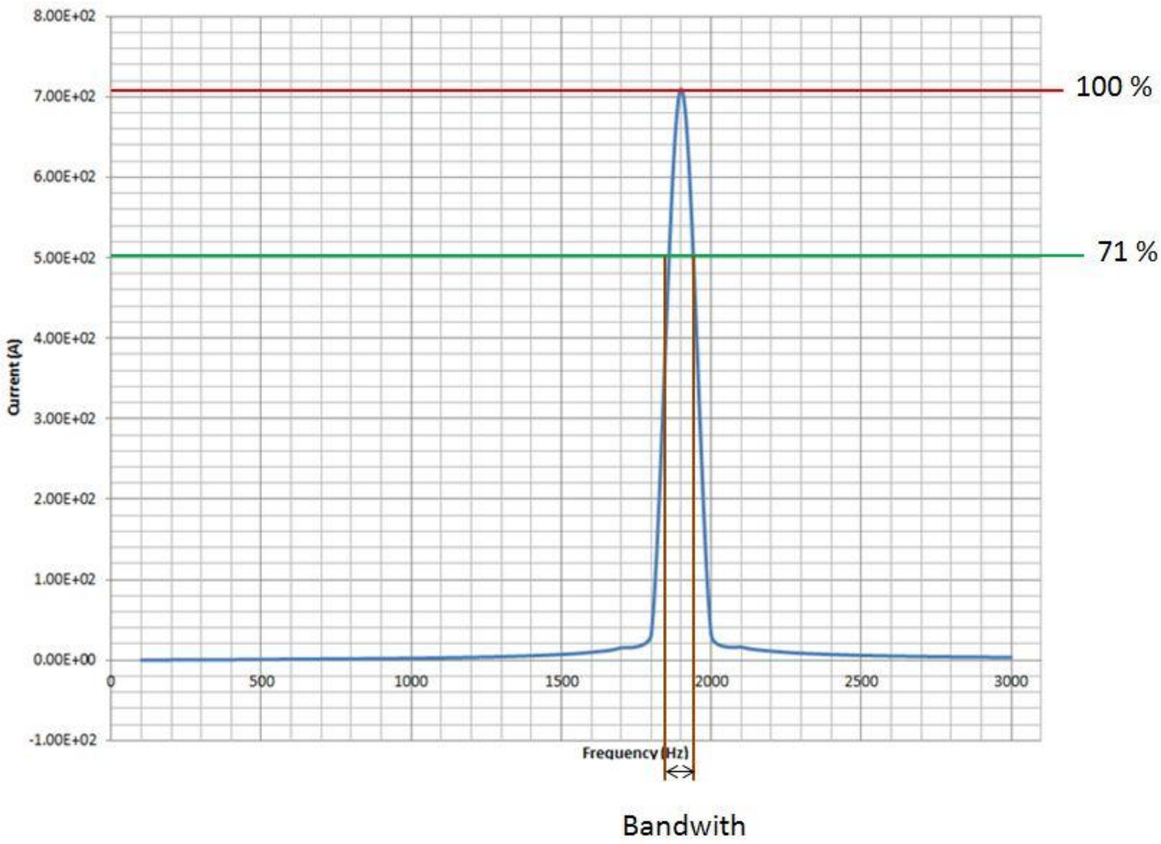


Figure 88 Series resonance graph showing the bandwidth

The Q factor is given by the bandwidth relationship:

$$Q = \frac{f_0}{f_B}$$ .... *Equation 44*

where f_0 is the resonant frequency and the f_B is the frequency range of the bandwidth:

f_B = Highest frequency - lowest frequency .... *Equation 45*

In the Exam

The concept of resonance is quite likely to have interpretation of graphical data, especially when the Q-factor is considered.

Find the lowest and the highest frequency that gives 71 % ($1/\sqrt{2}$) of the maximum voltage or current.

To get the magnified voltage, multiply the supply voltage by the Q factor.

Questions

Tutorial 14E.06

14E.06.1

Refer to *Figure 79*. Explain what the phasor diagram is showing.

14E.06.2

Why is resistance not mentioned?

14E.06.3

A 330 μF capacitor is in series with a perfect inductor of inductance 50 mH. Both are connected to an alternating supply voltage of 6 V.
a. Which data item is irrelevant to the question?
b. Work out the resonant frequency.

14E.06.4

In a circuit, there is an overall pure inductance of 0.141 mH. It is found to resonate at a frequency of 25000 Hz. Calculate the capacitance of the circuit.

14E.06.5

Refer to *Figure 88*. Show that the Q factor in the graph is about 19.

3. Operational Amplifiers

Tutorial 14 E.07 Ideal and Real Operational Amplifiers

AQA Syllabus

Contents

Before we look at the **operational amplifier**, it's worth knowing about some terms that we use with amplifiers.

14E.071 Amplifier Gain

The output voltage from an input device such as a microphone, or tape head in a cassette recorder is very small, in the order of millivolts. The loudspeaker takes a voltage of 20 V. Therefore, there needs to be a way of boosting the voltage to a level at which it can be used. A **transformer** is no good, because, as the voltage is increased, the current is reduced. We have to find a way of increasing the current as well as the voltage. This is done using an **amplifier**.

The extent to which the amplifier increases the voltage, current, or power is called the **gain**. It is the ratio of the output voltage (or current, or power) to the input voltage (or current, or power):

$$\text{Gain} = \frac{\text{Output voltage}}{\text{Input voltage}}$$

In Physics code:

$$\text{Gain} = \frac{V_{\text{out}}}{V_{\text{in}}}$$

.... Equation 46

In this section of work, we will consider the **voltage gain**.

Let us now look at what happens when we apply a sinusoidal voltage to the input of an amplifier and see its effect on the output voltage (*Figure 89*).

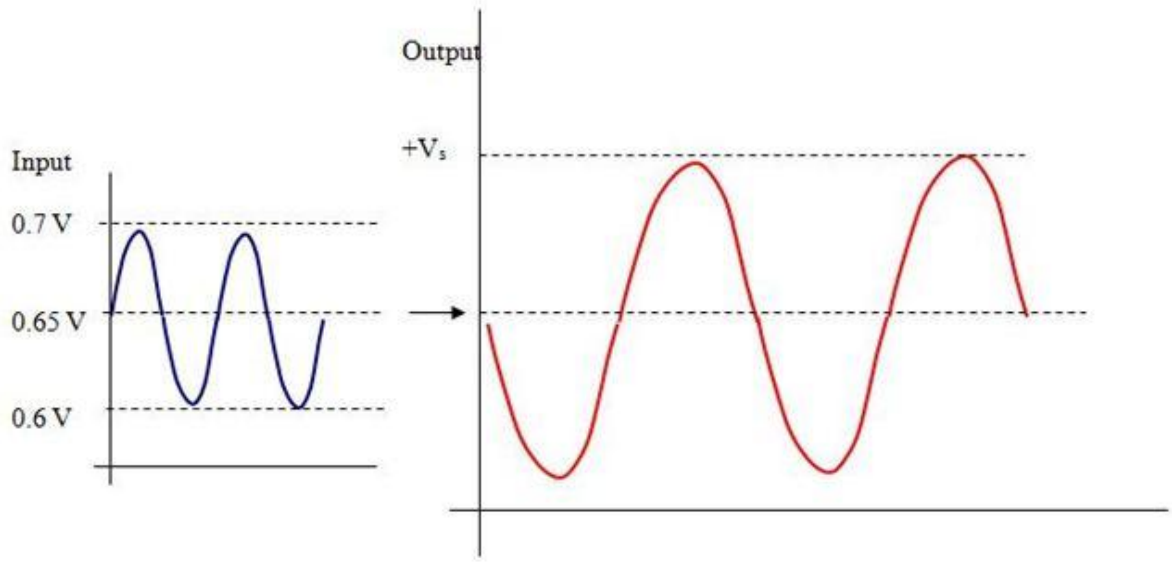


Figure 89 Voltage gain of an amplifier

Notice that the **phase** of the wave is **changed by 180°** (π radians). This means that the output wave is upside down compared to the input wave. The picture (*Figure 90*) shows the output wave on the top trace of the CRO and the input wave on the bottom trace.

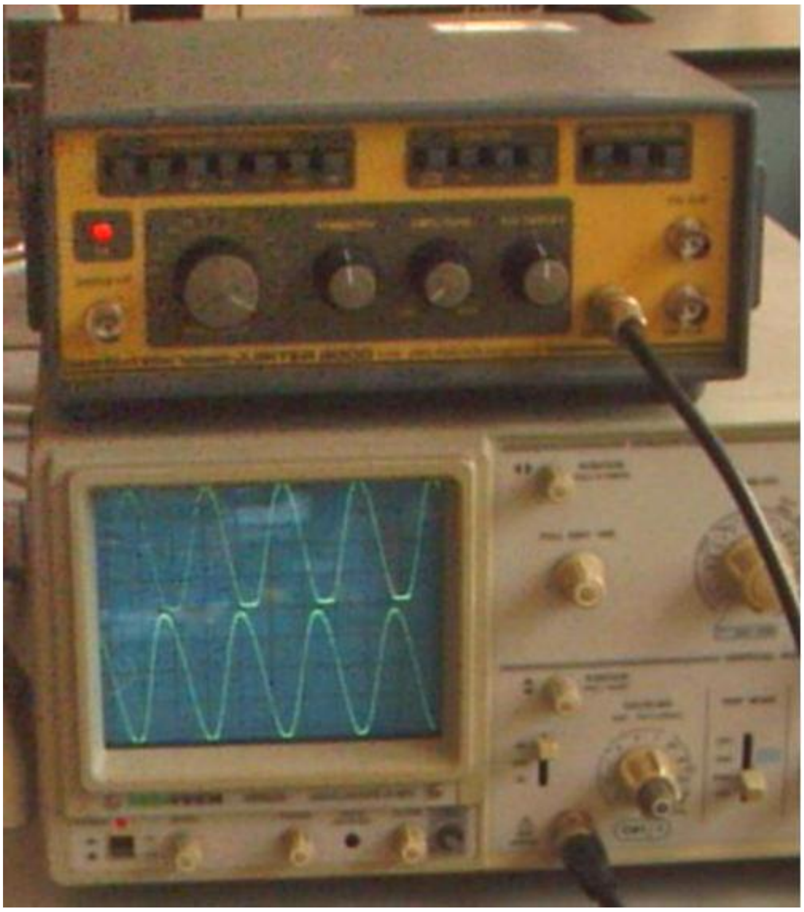


Figure 90 Phase change shown on a CRO

If the average level of the input signal is too large, we will get **distortion** (*Figure 91*)

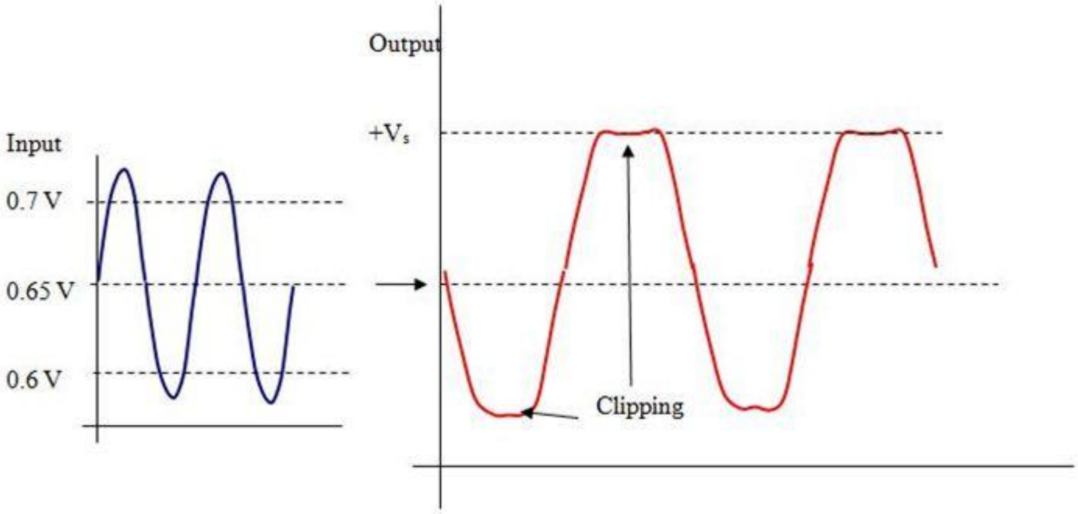


Figure 91 Clipping resulting in distortion

Notice that the peaks and troughs of the waves are cut off. This is called **clipping** and results in a noticeable distortion. Bad clipping can make an audio signal at best unpleasant to listen to, at worst unintelligible.

14E.072 Bandwidth

The **frequency response** of an amplifier is the range of frequencies that an amplifier can amplify. The graph of **power gain** against frequency for many amplifiers is like this (*Figure 92*):

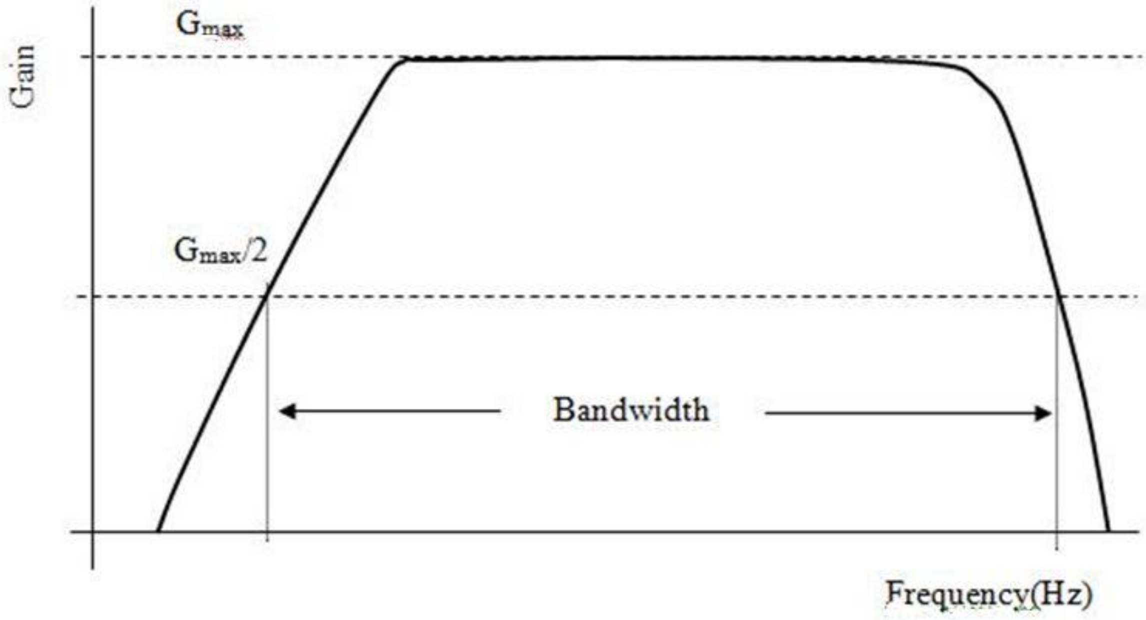


Figure 92 Graph showing the gain of an amplifier against frequency

The **bandwidth** of the amplifier is the range of frequency at which the power is at least half the maximum power. (We saw that in the last tutorial in reference to energy per cycle.)

If a loudspeaker provides a constant load, we can say that a voltage rise of **root 2** times gives a doubling of power. This is because the current will also go up **root 2** times. Therefore, we can define the bandwidth as:

the frequency in which the voltage gain is not less than 1/√2 times the maximum value.

This is about 71 %.

So, the **voltage gain-frequency** graph looks like this (*Figure 93*):

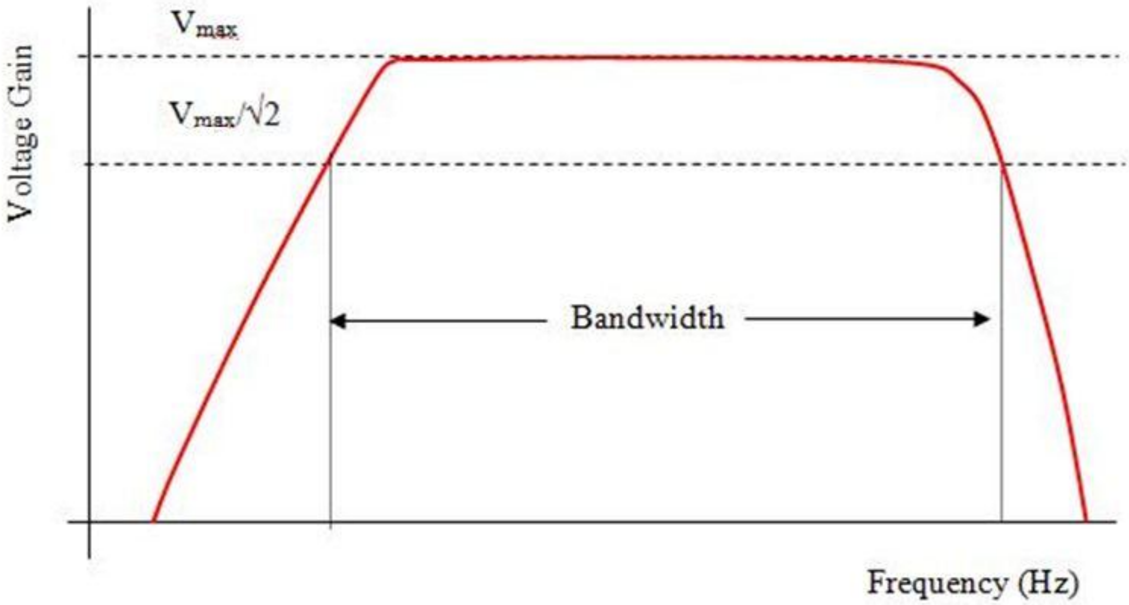


Figure 93 Graph showing the voltage gain against the frequency

14E.073 Feedback

Feedback is a term that electronic engineers use in which a fraction of the output is taken back to the input. Negative feedback is widely used in amplifier circuits as it reduces the gain. It also makes the amplifier more stable. Amplifiers without negative feedback tend to be rather unstable. This can arise due to:

- Temperature differences
- Stray inductance and capacitance effects
- Noise within the components or from poor soldering.
- Fluctuations from the power supply.

The effect of an unstable amplifier is that the output becomes distorted in an unpredictable and random way.

14E.074 Positive Feedback

Positive feedback is not used in amplification. The diagram (*Figure 94*) shows the idea:

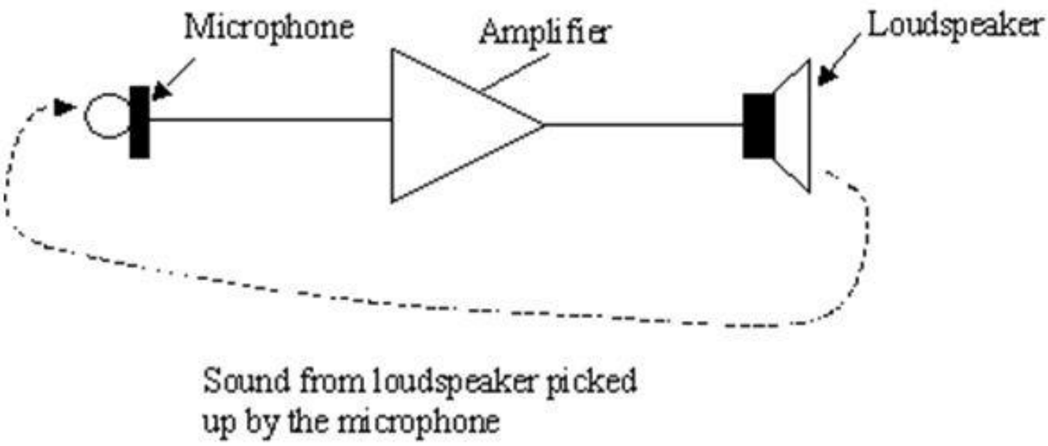


Figure 94 Positive feedback

The microphone picks up some sound, and this goes with the input signal to be amplified. This makes the sound louder, so the input to the microphone gets larger, which gives a larger output... The result of this is a **feedback loop** or **howl round**. Whatever you call it, it sounds the same, an ear shattering boom or screech. Positive feedback is used constructively in circuits such as the **Schmitt Trigger** or **oscillator** circuits.

14E.075 Negative Feedback

Negative feedback reduces the gain but increases the stability by feeding a small fraction of the output to the input. The phase is changed. This reduces the input so that the output is reduced as well. Therefore, the amplifier is much easier to control. The principle of **negative feedback** is best shown with an op-amp circuit (*Figure 95*).

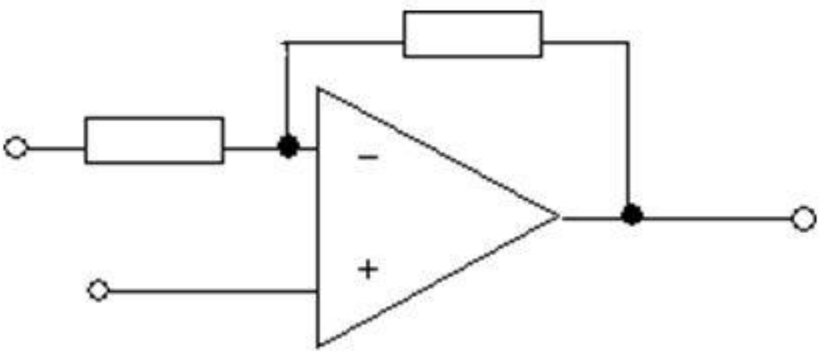


Figure 95 Negative feedback

This is what happens:

- The output is fed into the inverting input,
- The phase will be changed by 180°.
- This will reduce the output.
- This is added to the input wave, reducing the size of the input.

The graph (*Figure 96*) shows this:

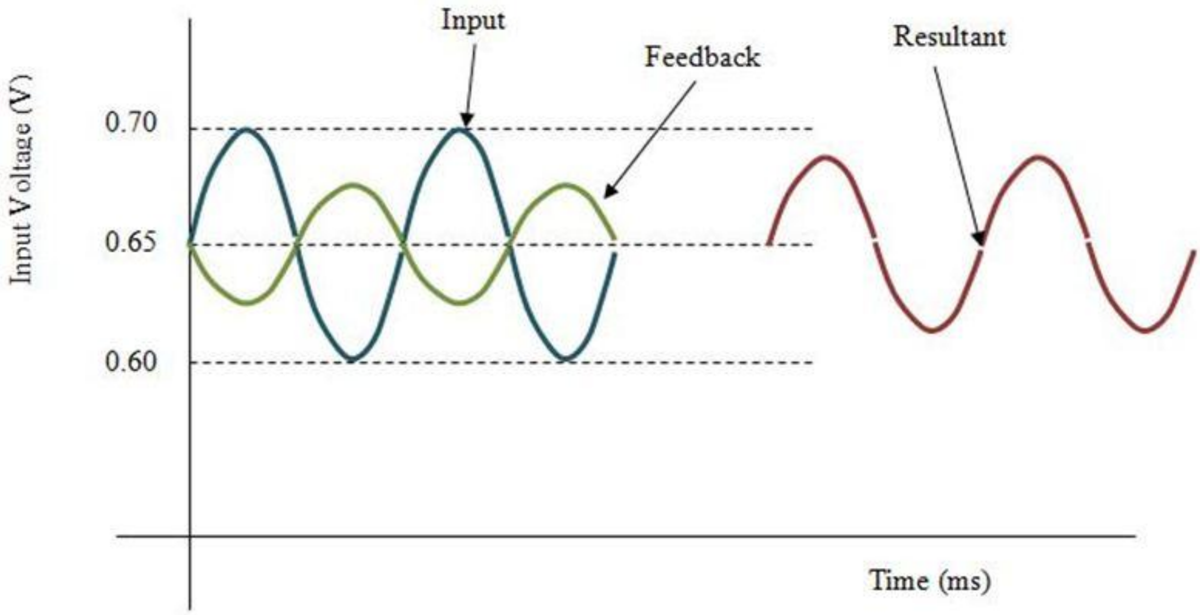


Figure 96 Principle of negative feedback

14E.076 Operational Amplifiers

The **operational amplifier** was originally devised in the 1960s for use in **analogue computers**. These are nowadays almost completely obsolete, although a few have some very special applications. However, the operational amplifier still has many uses in control and instrumentation electronics. Although the original circuits used discrete components and were very expensive, miniaturisation has enabled op-amps to be made as integrated circuits, available for a few pence.

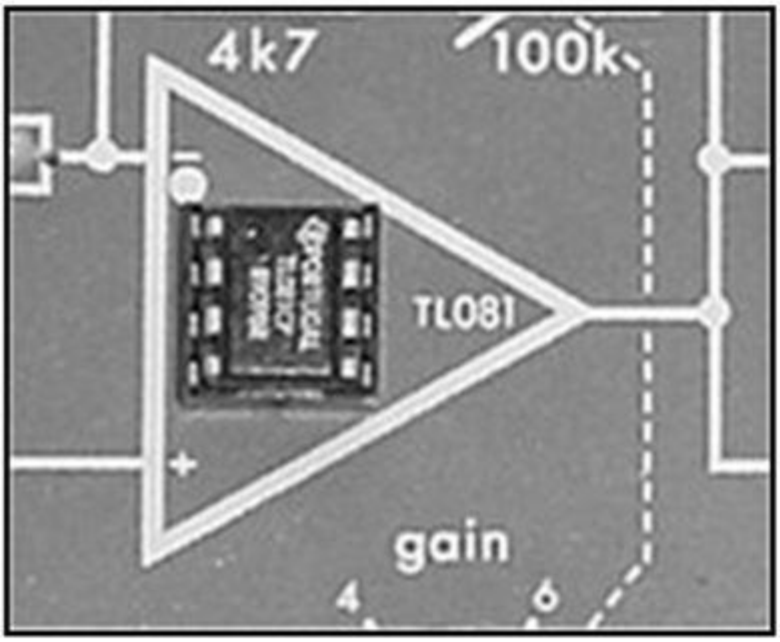


Figure 97 A typical operational amplifier

Operational amplifiers (**op-amps**) require a **dual rail power supply**, which means having a central 0 volts rail, and a + 15 V rail and a – 15 V rail. The full circuit diagram is shown below (*Figure 98*), but generally we will ignore the power supply.

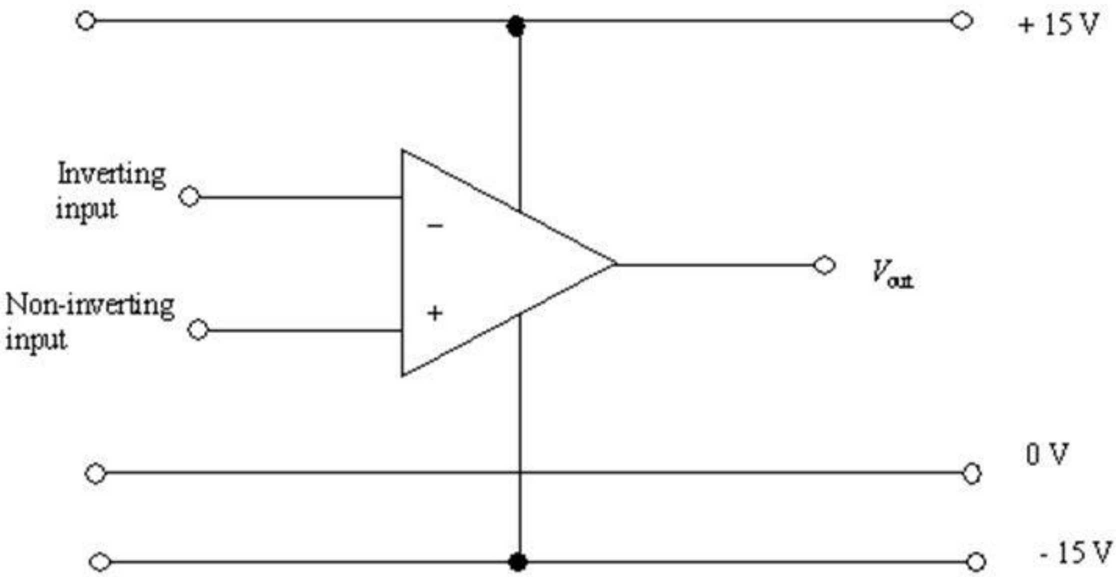


Figure 98 Dual rail power supply for an op-amp

Notice that the op-amp has two inputs and one output. It is a **difference** amplifier and amplifies the difference between the inverting input and non-inverting input. Be careful not to confuse the symbol with a non-inverting gate. We need to be aware of some definitions to do with op-amps:

- **Open-loop voltage gain** – ratio of the input to the output voltage with no feedback applied. It is the D.C. gain of the amplifier or the gain at a frequency of 1 Hz.
- **Closed loop voltage gain** – the voltage gain with feedback.
- **Bandwidth** is the frequency range in which the output does not fall by more than 3 dB from its maximum value.

14E.077 Ideal Op-Amp

The **ideal** op-amp should have the following characteristics:

- Infinite open loop gain
- Infinite input impedance so that no current is drawn
- Zero output impedance so that maximum current can be transferred to the load.
- Very wide bandwidth.

In practice the maximum open loop gain is 200 000. Beyond that limit the amplifier goes into **saturation** which means that the voltage cannot go any higher. The voltage is limited, of course, by the supply voltages. In practice the limits are rather lower than this, about 1.5 to 2 V below the value of the supply. Suppose the supply voltage was 15 V. The maximum output voltage would be 15 – 1.5 = 13.5 V

The behaviour or **characteristic** of the op-amp is shown in this graph (*Figure 99*). Notice that the horizontal axis is marked **difference in voltage**. The op-amp is a **difference amplifier**.

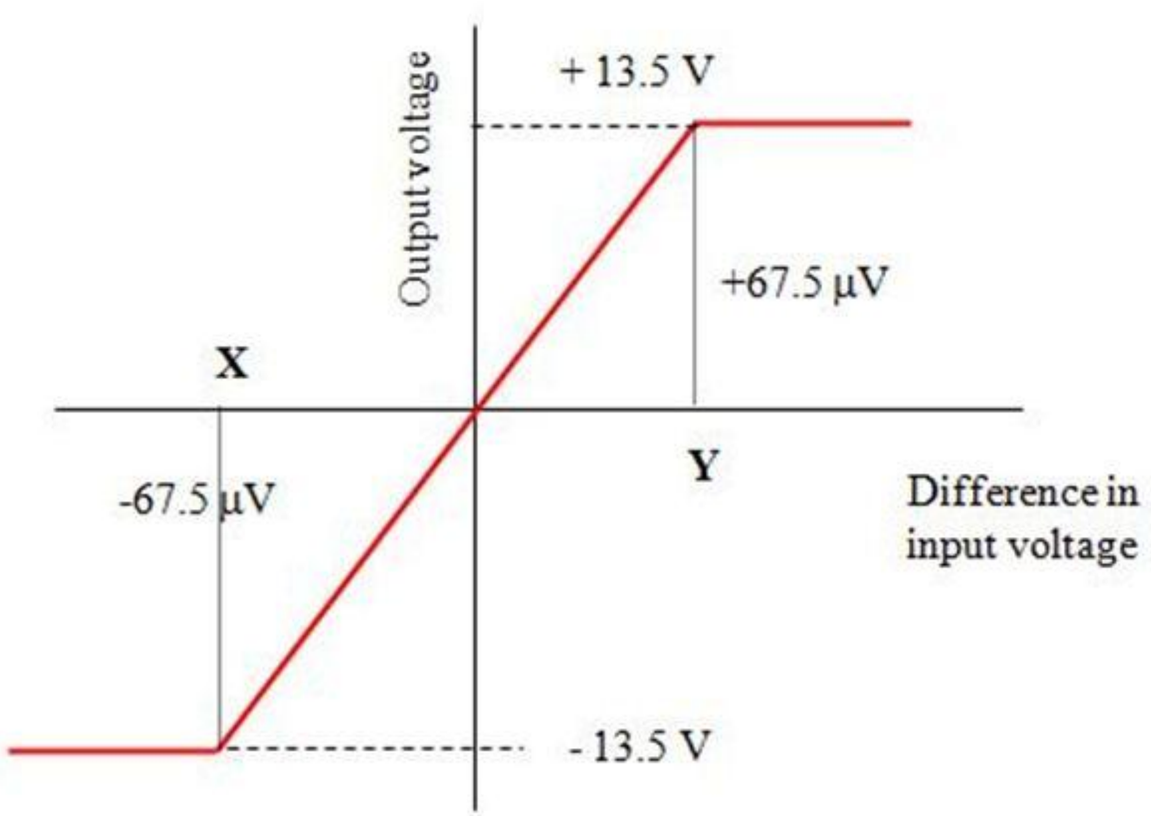


Figure 99 Characteristic of an op-amp

The graph tells us:

- The op-amp amplifies the **difference** between the two input voltages
- When the difference exceeds the limits **X** and **Y**, the output is saturated.
- In between the limits the graph is linear, so there is little distortion in these regions.

14E.078 Real Op-amp

The characteristic of real op-amps makes them unsuitable for use as amplifiers in open loop form, as clipping will occur and this will distort the signal. Therefore, some of the output is returned to the amplifier by a **feedback** loop. This reduces the gain and makes the amplifier more stable. The amplifier can be used in open loop form as a **voltage comparator.**

The equation for the gain of an operational amplifier is:

Output voltage = gain × voltage difference

In Physics code:

$$V_{\text{out}} = A_{\text{OL}}\,[(V+) - (V-)]$$ .... *Equation 47*

Note that the physics code A_{OL} is used here rather than G for gain. It stands for **open loop gain**, which is the gain of the op-amp without feedback.

The open loop **frequency response** of the op-amp is not very good (*Figure 100*):

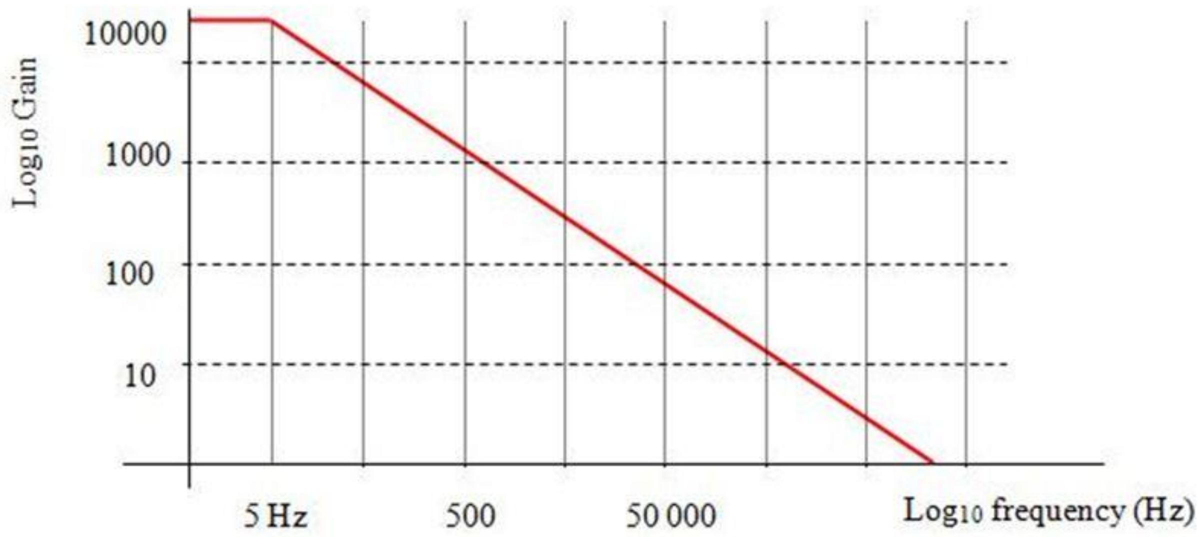


Figure 100 Frequency response of the open loop op-amp

We can see that the gain starts to fall away quite dramatically above a frequency of only 5 Hz, which is not very high. It would be quite useless as an audio amplifier.

However, the gain can be improved by reducing the gain with the use of **negative feedback**. This is shown in *Figure 101*.

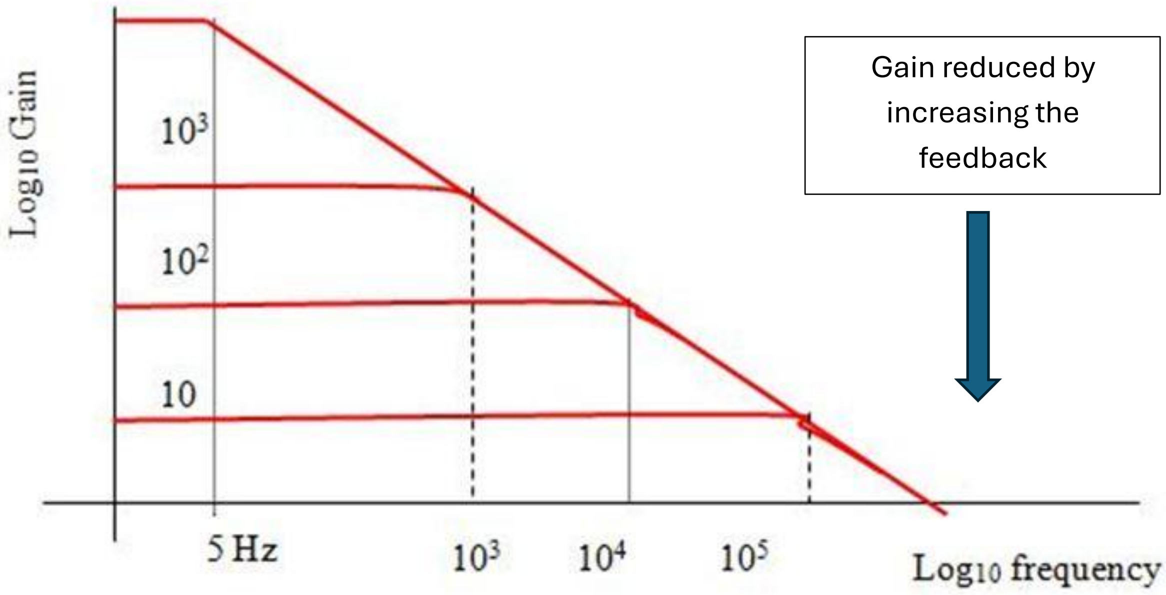


Figure 101Bandwidth is improved by increasing the negative feedback (lowering the gain)

We can see from this graph that the more negative feedback that we apply, the wider the bandwidth. There is a useful relationship for op-amps:

bandwidth × gain = constant .... *Equation 48*

The constant is often called the **unity gain bandwidth**. This is the bandwidth at which the gain of the amplifier is 1. The constant is a property of the op-amp. So, the unity gain bandwidth is the bandwidth of the op-amp when the op-amp is configured to have a gain of 1. We can measure the bandwidth using a set-up like this (*Figure 102*):

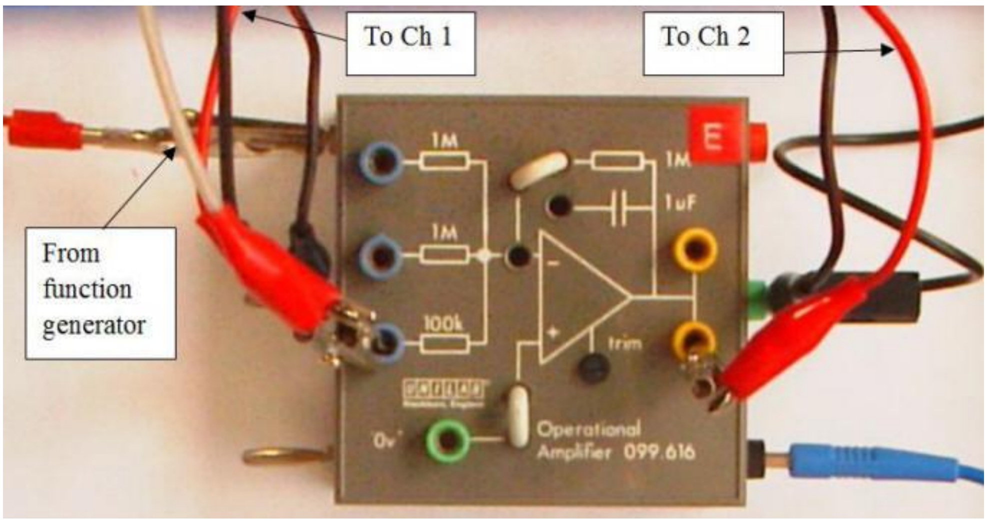


Figure 102 Measuring the unity gain bandwidth

14E.079 Using the Op-Amp as a Voltage Comparator

The op-amp being a differential amplifier is very useful as a **voltage comparator**. It will give a high positive or negative voltage, depending on which voltage is higher:

- If the non-inverting input is higher, the output voltage will be positive
- If the inverting input is higher, the output voltage will be negative.
- The voltage difference must be greater than 135 mV if saturation is to occur. This is not difficult to achieve.

The voltage comparator gives a digital output from an analogue input.

The diagram (*Figure* 103) below shows the voltage comparator used as a light operated switch.

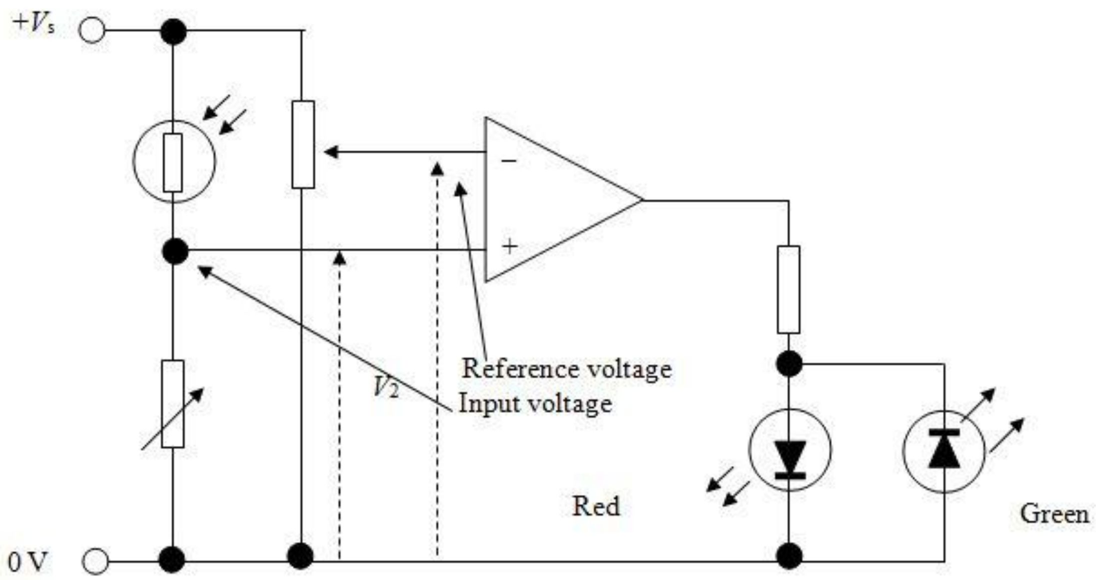


Figure 103 Op-amp used as a voltage comparator

This is how it works:

- The reference voltage is adjusted with the variable resistor so that it is equal to the voltage V_2.
- When light falls on the LDR, V_2 goes up because the resistance of the LDR goes down.
- V_2 will be bigger than the reference voltage, so that the op-amp goes into saturation.
- The LED lights up, showing a high output.
- The reverse-biased diode protects the LED for when there is a large negative voltage.

The most common operational amplifier is the **LM741** device. Others are available.

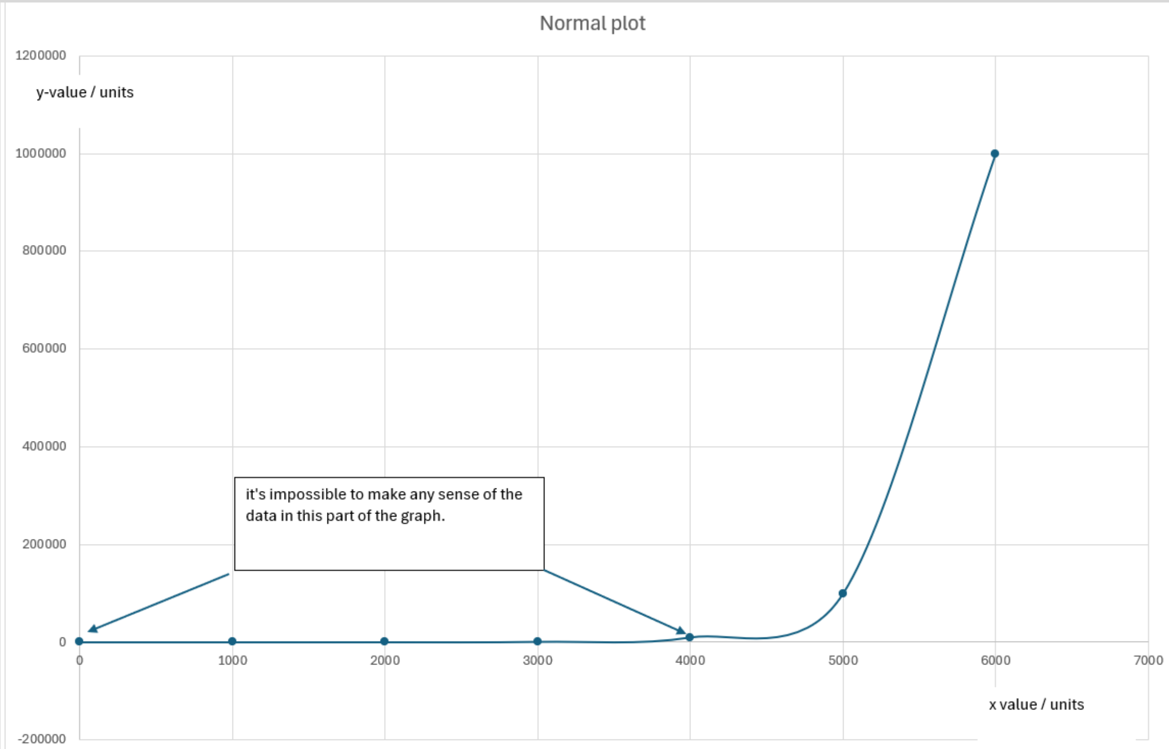
Maths Note

It is important to know how to read data from and to interpret graphs with logarithmic scales.

Here are some data. These data have no physical significance other than being randomly put together to illustrate the point.

x / units	y / units
1	1
1000	10
2000	100
3000	1000
4000	10000
5000	100000
6000	1000000

Let's do a normal plot:



Notice the y-axis has a very large range, so it's very hard to make any sense of what is happening below x = 5000 units. The y-axis needs to be compressed. This is done by taking the logarithm of the y-value.

A logarithm is any number expressed as a power of the base number. You have come across natural logarithms before (log to the base e where e = 2.718...). You can have logarithms to any base you like. In this illustration, we will use a base of 10. We write these as:

$$\log_{10} \text{ or } \lg$$

where

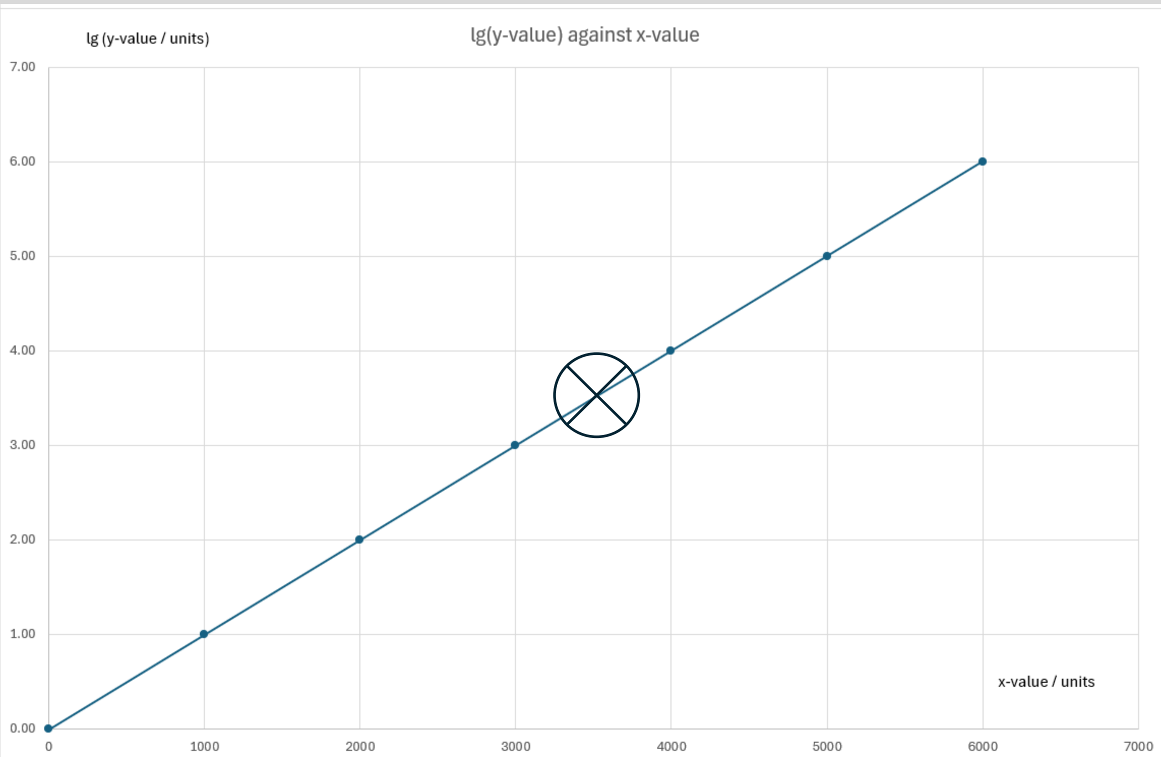
$$\lg(10) = 1, \lg(100) = 2, \text{ and so on.}$$

So, we compress the y-axis by taking logs, where 10000 = 10^4, thus log (10000) = 4.

Now look at the way we have modified the data to compress the y-axis.

x / units	lg (y / units)
1	0.00
1000	1.00
2000	2.00
3000	3.00
4000	4.00
5000	5.00
6000	6.00

And we plot the data:



Notice how the y-axis has been compress and gives easy values like 3,4 and 5. These mean 10^3, 10^4 and 10^5 respectively.

Let's plot a point at x-value = 3500 units. If we read across to the y-value scale, we read the answer as about 3.5 units. But that is NOT our final answer.

We have to take the **antilogarithm**, which is a **power of 10**. It is often written as lg^{-1}.

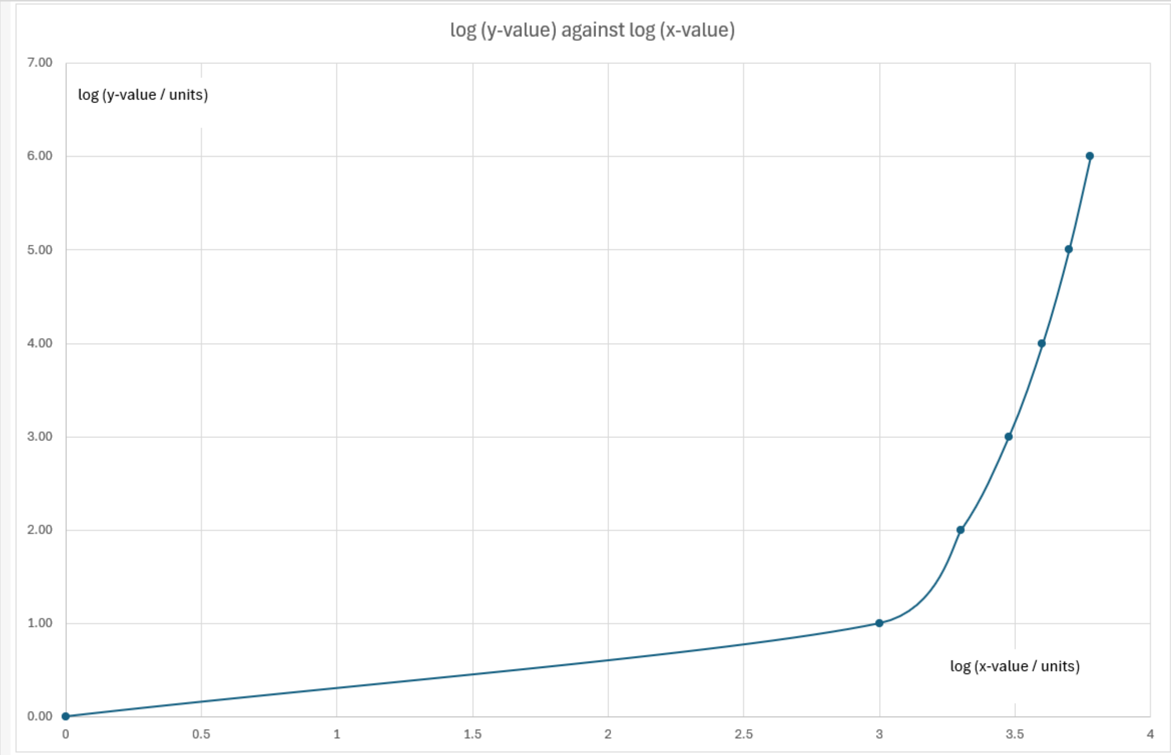
For example:

$$\text{Antilog}\,(2.3010) = lg^{-1}\,(2.3010) = 10^{2.3010} = 200$$

In our case above:

$$\text{Answer} = 10^{3.5} = 3160 \text{ units.}$$

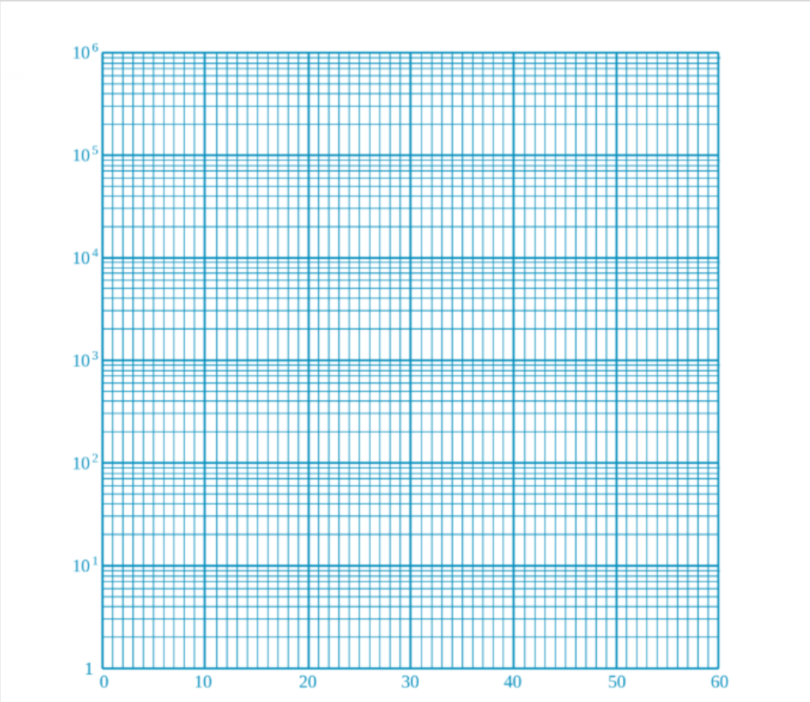
We can have both scales as logarithmic scales as shown below:



While it may be unclear as to what we can do with such a graph, we can look at data where the x-value is less than 1000 units.

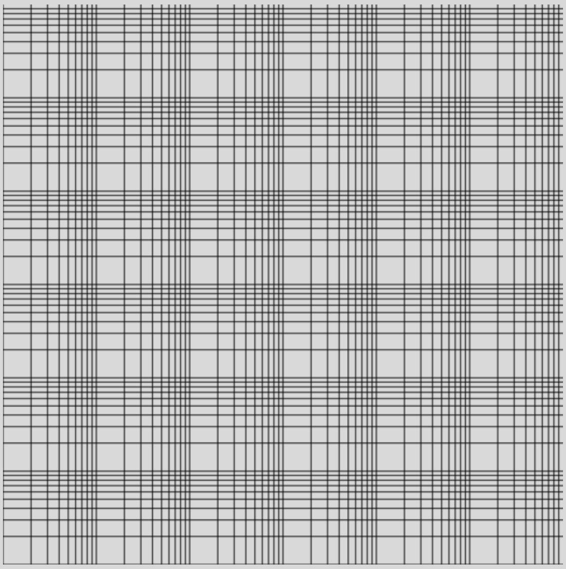
Generating logarithms from harvested data can be tedious. The data I have used here is from an Excel ® Spreadsheet.

But you can plot data directly without having to convert the data points to logs. The data are plotted onto **semi-logarithmic** graph paper as shown below:



Note that the y-axis is logarithmic while the x-axis is linear. Note that the logarithmic axis goes up in **decades**, and the steps within the decades are not even.

We can also get log-log paper which looks like this:



(Images from Wikimedia Commons)

While you will not be required to draw semi-log or log -log plots at A-level, you will certainly come across them in University level Physics, where at least one quantity has a large range of values.

You must make sure how to use $\log_{10}$ on your calculator. Most calculators have '**lg**' on the key for $\log_{10}$ and '**ln**' for natural logarithms ($\log_e$). It is up to you to know what you are working with. Many marks are lost due to such errors.

You cannot use logs with simple negative numbers.

lg(0) is impossible. You calculator will give an error.

lg (1) = 0

For numbers < 1, the logs are negative. For example, lg (0.2) = -0.699. Since $\log_{10}$ (2) = 0.301, $\log_{10}$ (0.2) = -1 + 0.301 = -0.699.

4 × lg (2) = 4 × 0.301 = 1.204. This is not the same as lg (4 × 2) = lg (8) = 0.902

Questions

Tutorial 14E.07

14E.07.1

An amplifier has a gain of 200. The input voltage is 75 mV, what is the output voltage?

14E.07.2

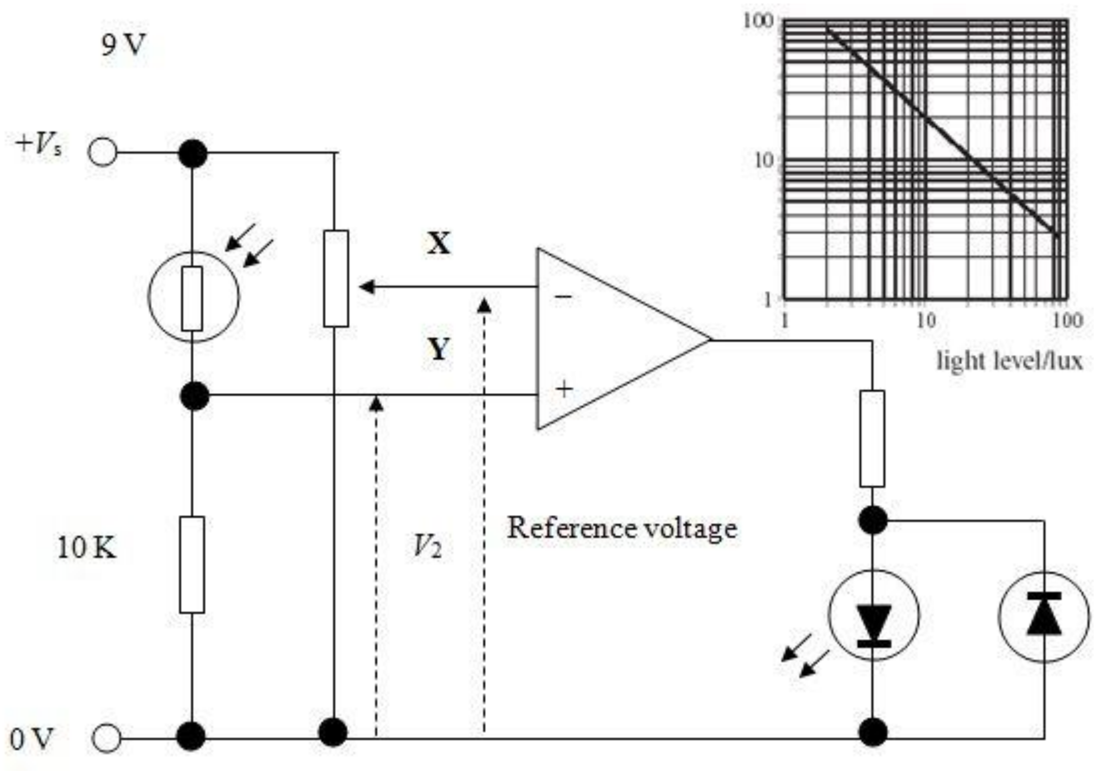
What would the input voltage be to give an output voltage of ± 13.5 V? What is the voltage swing needed to go from negative to positive saturation? The op-amp gain is 100 000.

14E.07.3

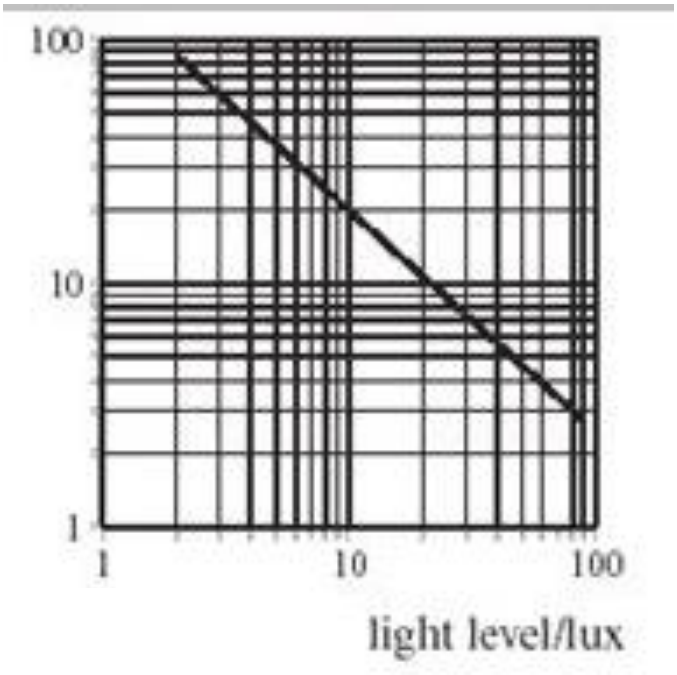
Compare the real op-amp with the ideal op-amp.

14E.07.4

The LDR in this circuit has a light level-resistance characteristic as shown in the graph. A larger version is shown below:



Resistance / kΩ



(a) The potentiometer is set half-way down its track. Calculate the voltage at X and Y.

(b) Work out the resistance of the LDR at which the output of the op-amp just changes.

(c) What is the light level at which this happens? (Resistance in kilohms is on the vertical axis).

Tutorial 14 E.08 Uses of Operational Amplifiers	
AQA Syllabus	
Contents	
14E.081 The Inverting Amplifier	14E.082 The Non-Inverting Amplifier
14E.083 Summing Amplifier	14E.084 Digital to Analogue Conversion
14E.085 The Difference Amplifier	14E.086 Derivation (Extension only)

14E.081 The Inverting Amplifier

Negative feedback is achieved by bringing a fraction of the output signal to the inverting input of the op-amp. The photograph shows an op-amp with negative feedback to give an **inverting amplifier** (*Figure 104*).

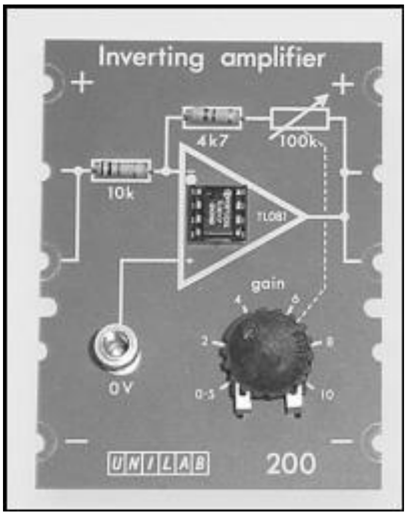


Figure 104 An inverting op-amp

Negative feedback also has the following effects:

- The gain is reduced
- The bandwidth (frequency response) is increased
- The stability is improved.

This allows the op-amp to be used in many applications including audio.

The diagram (*Figure 105*) shows a typical negative feedback arrangement:

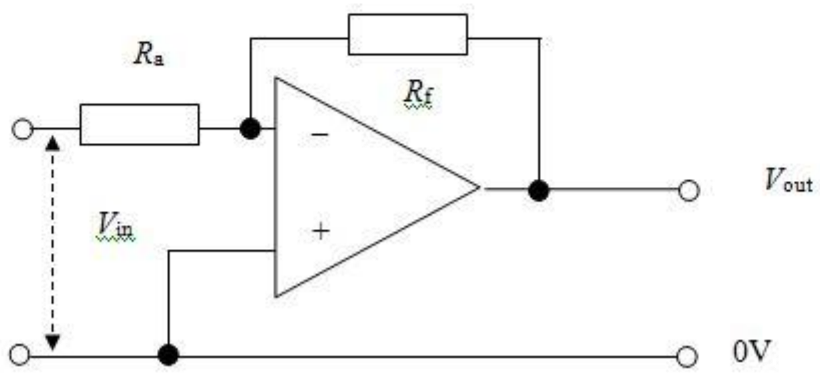


Figure 105 Negative feedback in an op-amp

Note that in these notes I will use R_a for the input resistance instead of R_{in}, which is used in the syllabus. Both are correct.

In this arrangement the non-inverting input is connected to the 0V line. Since there is very little voltage difference between the non-inverting and the inverting input, we can also say that the voltage at the inverting input is almost at 0V as well. So, we say that the inverting input is at a **virtual earth**.

Let us suppose there is a current I_1 flows through R_a and a current I_2 flows through R_f. Ohm's Law allows us to say:

Current through R_a,

$$I_1 = \frac{V_{in}}{R_a}$$

.... *Equation 49*

Current through R_f,

$$I_2 = \frac{V_{out}}{R_f}$$

.... *Equation 50*

Since the resistance at the inverting terminal is very high, no current can flow through the inverting input. At point **X** (according to Kirchhoff I) the current in is the same as the current out. Therefore, the current is the same through R_a and R_f (*Figure 106*).

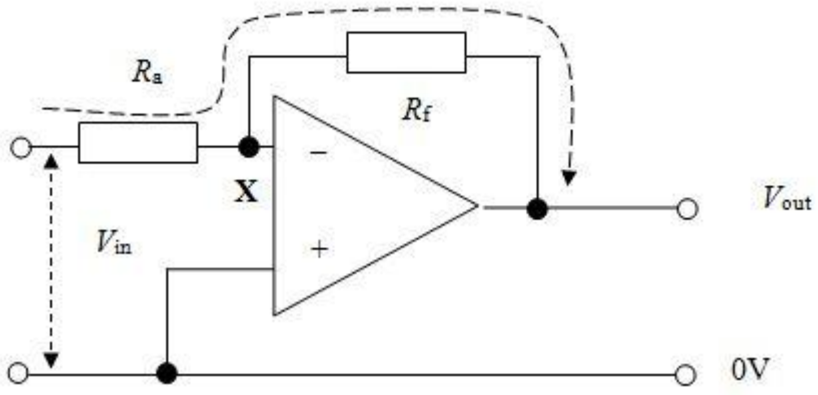


Figure 106 Current flowing through the feedback resistor

Kirchhoff I:

$$I_1 + I_2 = 0$$ .... *Equation 51*

Substituting *Equations 49* and *50* into *51*:

$$\frac{V_{in}}{R_a} + \frac{V_{out}}{R_f} = 0$$ .... *Equation 52*

Rearranging:

$$\frac{V_{in}}{R_a} = -\frac{V_{out}}{R_f}$$ .... *Equation 53*

Therefore:

$$\frac{V_{out}}{V_{in}} = -\frac{R_f}{R_a}$$ .... *Equation 54*

Since point X is at a virtual earth and the voltage of X is zero, and therefore the input resistance is the same as the input resistor. So, the input resistance of this amplifier is 1000 Ω. This will have implications for **load matching** for an input device such as a microphone. The output in all inverting amplifiers relates to the input as shown below (*Figure 107*):

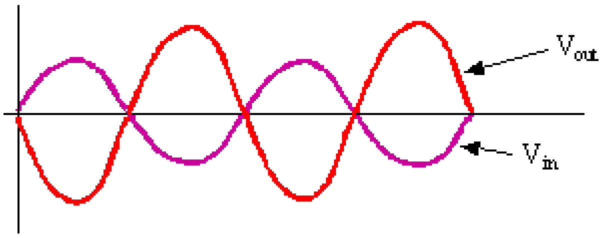


Figure 107 Input and output of an op-amp are 180 degrees out of phase

Note the input and output are **180 degrees** out of phase.

14E.082 The Non-Inverting Amplifier

In this circuit the input voltage is applied to the **non-inverting** input (*Figure 108*).

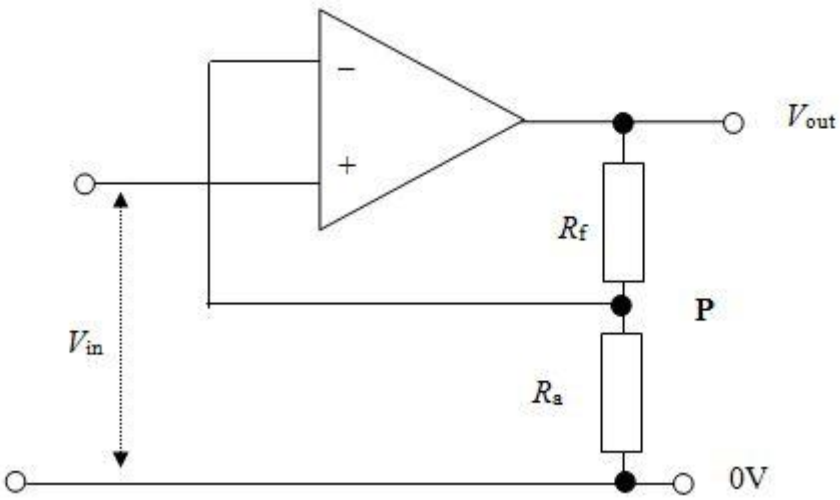


Figure 108 Non-Inverting op-amp

Notice that:

- This amplifier uses negative feedback to the **inverting** input.
- There is no difference in the voltage between the inverting and non-inverting inputs, so we can say that the voltage at the non-inverting and the inverting input is the same. So, we can say that the voltage at **P** is V_{in}.
- Since no current is drawn by the inverting input, the current in R_{a} is the same as the current in R_{f}. So, we can treat the two resistors as a **potential divider** and apply the potential divider equation.

We can therefore write:

$$V_{\text{out}} = V_{\text{in}} \left(\frac{R_{\text{a}}}{R_{\text{f}} + R_{\text{a}}} \right)$$

……….. *Equation 55*

Rearranging gives:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \left(\frac{R_{\text{a}}}{R_{\text{f}} + R_{\text{a}}} \right)$$

………… *Equation 56*

The term $V_{\text{out}}/V_{\text{in}}$ is the gain.

The term:

$$\left(\frac{R_{\text{a}}}{R_{\text{f}} + R_{\text{a}}} \right)$$

can be rewritten as:

$$\frac{R_{\text{a}}}{R_{\text{a}}} + \frac{R_{\text{f}}}{R_{\text{a}}}$$

Therefore:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = 1 + \frac{R_{\text{f}}}{R_{\text{a}}}$$

………… *Equation 57*

If we look at the input, we see that there is no feedback resistor in the input, therefore we can say that the input resistance is that of the op-amp. The input resistance is very high indeed, and very little current is taken.

Extension

The problem with the inverting amplifier used as a **voltage follower** is that the output is at 180° out of phase with the input. A voltage follower can be based on the non-inverting circuit with 100 % negative feedback to the inverting input, and input resistance is very high indeed (*Figure 109*).

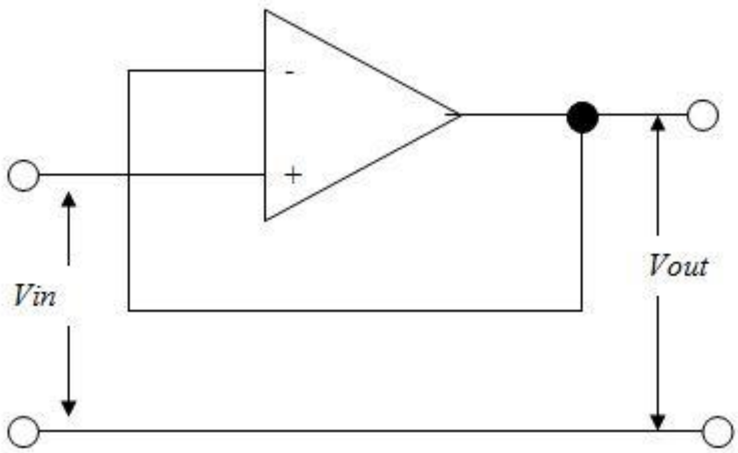


Figure 109 Voltage follower

The voltage gain of the op-amp in this configuration is about 1. This because of the feedback factor (the fraction fed back), given the code β (beta, a Greek letter 'b') is 1.

We can show this by considering the **open loop gain** A_0. The actual gain A is given by:

$$A = \frac{A_0}{1 + \beta A_0}$$

.... *Equation 58*

If β is 1, and A_0 is very large, we can say that A is approximately 1.

The main use of the voltage follower is as a **buffer amplifier**, which matches a high input impedance with a low input load. You would come across such a circuit in the input stage of a digital multimeter, which has a very high input impedance, allowing the voltage read to be the same as the voltage that should be there.

14E.083 Summing Amplifier

This kind of amplifier is used in **digital to analogue conversion**, or as a **mixer** in an audio system (*Figure 110*).

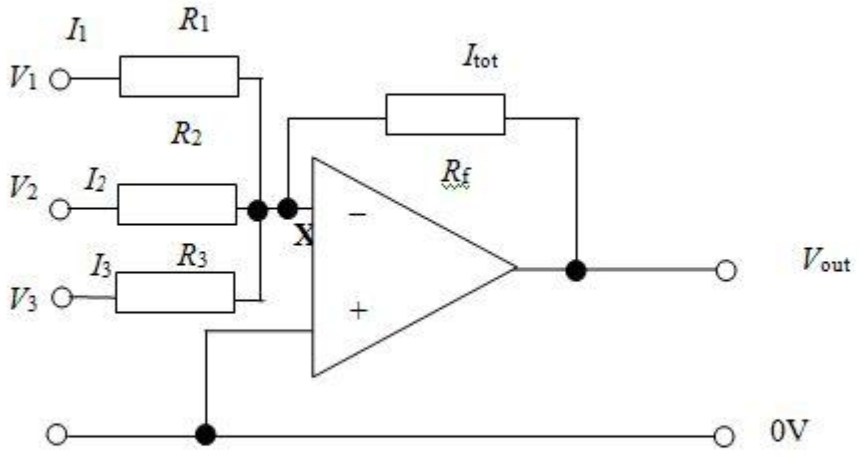


Figure 110 Summing amplifier

This circuit uses negative feedback into the inverting input, but instead of one input, there are three inputs.

Kirchhoff I tells us that:

$$I_{\text{tot}} = I_1 + I_2 + I_3$$ .... *Equation 59*

So, we can use Ohm's law to rewrite this in terms of voltage and resistance.

Notice that in the next step we use the minus sign because **X** is at a virtual earth, so we have to climb the potential difference hill to get to V_{out}. This is Kirchhoff II.

$$-\frac{V_{\text{out}}}{R_{\text{f}}} = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}$$ .... *Equation 60*

Therefore:

$$V_{\text{out}} = -R_{\text{f}}\left(\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}\right)$$

.... *Equation 61*

For any number of resistors, the equation becomes:

$$V_{\text{out}} = -R_{\text{f}}\left(\frac{V_1}{R_1} + \frac{V_2}{R_2} + \cdots + \frac{V_{\text{n}}}{R_{\text{n}}}\right)$$

.... *Equation 62*

If the values of all the resistors are the same:

$$V_{\text{out}} = -(V_1 + V_2 + V_3)$$ .... *Equation 63*

The output is the **sum** of all the input signals but is of **opposite** polarity.

Summing amplifiers are found in **mixing desks** which add together the inputs from several different audio sources.

14E.084 Digital to Analogue Conversion

There are many occasions in which digital signals are converted to analogue. Computers know two states, ON and OFF, but the signals that we pick up and use are analogue. So, the output of a computer (or a CD deck) is pretty meaningless without some kind of **digital to analogue conversion**.

A summing amplifier can be used, which gives an output equal to the sum of the digital signals (*Figure 111*).

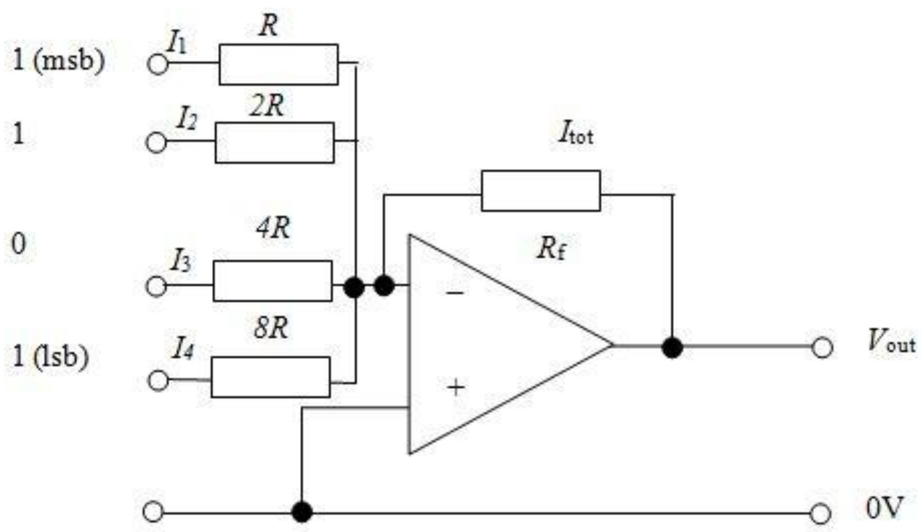


Figure 111 Summing amplifier used in digital to analogue conversion

In the diagram, *msb* means **most significant bit**, while *lsb* means **least significant bit**.

Since the inverting input is a virtual earth, we can say that the total current I_{tot} is given by:

$$V_{out} = -I_{tot} R_f \quad \text{.............. Equation 64}$$

$$V_{out} = -R_f (I_1 + I_2 + I_3 + I_4) \quad \text{.............. Equation 65}$$

$$V_{out} = -R_f \left(\frac{V_1}{R} + \frac{V_2}{2R} + \frac{V_3}{4R} + \frac{V_4}{8R} \right)$$

.... *Equation 66*

For every change in one bit, there is a voltage change of 0.0625 V. For an eight bit word it is 1/256 = 0.00390625 V. For a 16 bit word the smallest voltage change is $1/2^{16}$ = 1/65536 = 0.0000152 V. You can see that the more bits there are, the smaller the voltage change per bit, leading to greater resolution.

The diagram here shows the digital output converted to an analogue wave pattern. See *Figure 112*.

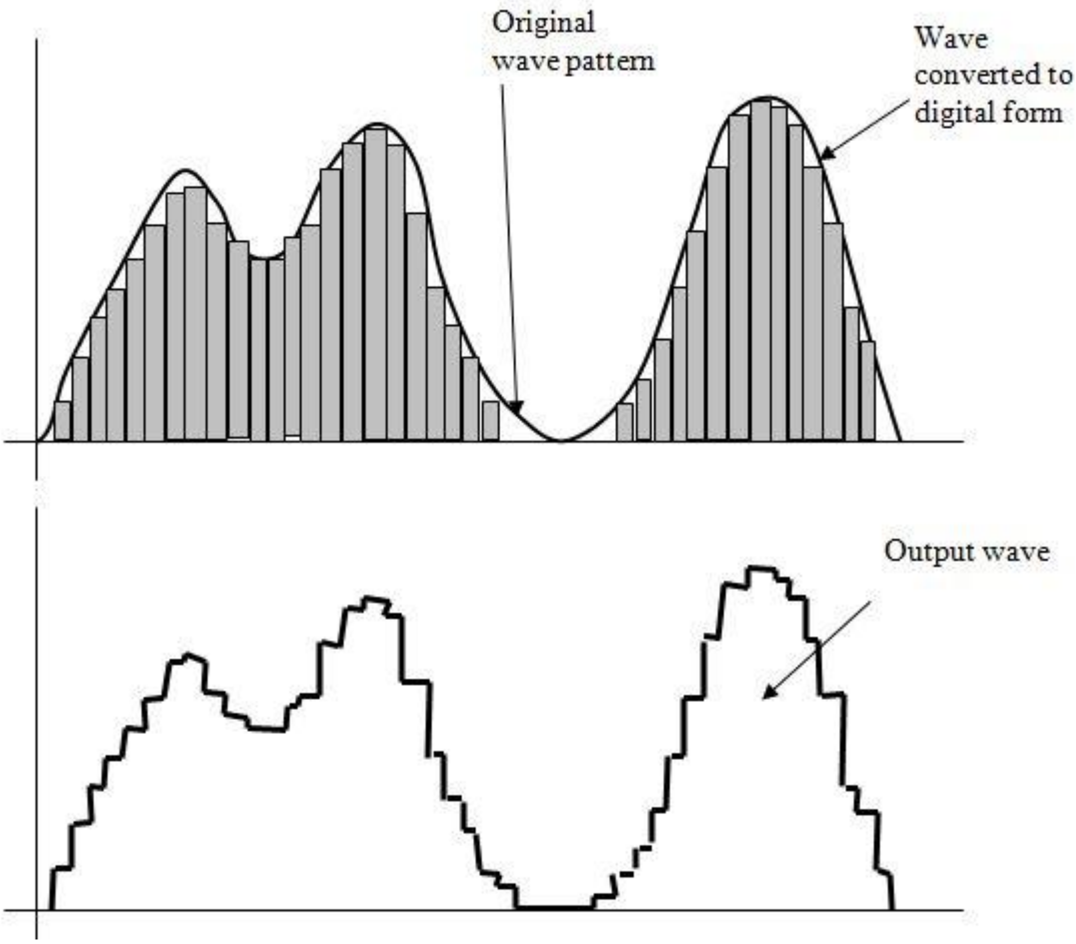


Figure 112 Digital to analogue conversion.

You can see that the output wave is similar to the input but would result in some distortion. This can be reduced markedly by increasing the resolution. However, there will always be an element of this kind of distortion. Audiophiles (people who like listening to music on High-Fidelity music reproduction equipment) reckon that analogue-recorded music sounds more natural than digitally recorded music and argue that a vinyl LP played on a **high quality record deck** sounds better than a compact disc. When these notes were originally written, the vinyl LP had almost completely been superseded by CDs. Now many music companies are re-issuing albums on vinyl, and manufacturers are again making high quality record decks. Why? People prefer the sound!

14E.085 The Difference Amplifier

The **difference amplifier** amplifies the difference between the voltage applied to the inverting input and the non-inverting input as shown in the circuit (*Figure 113)*.

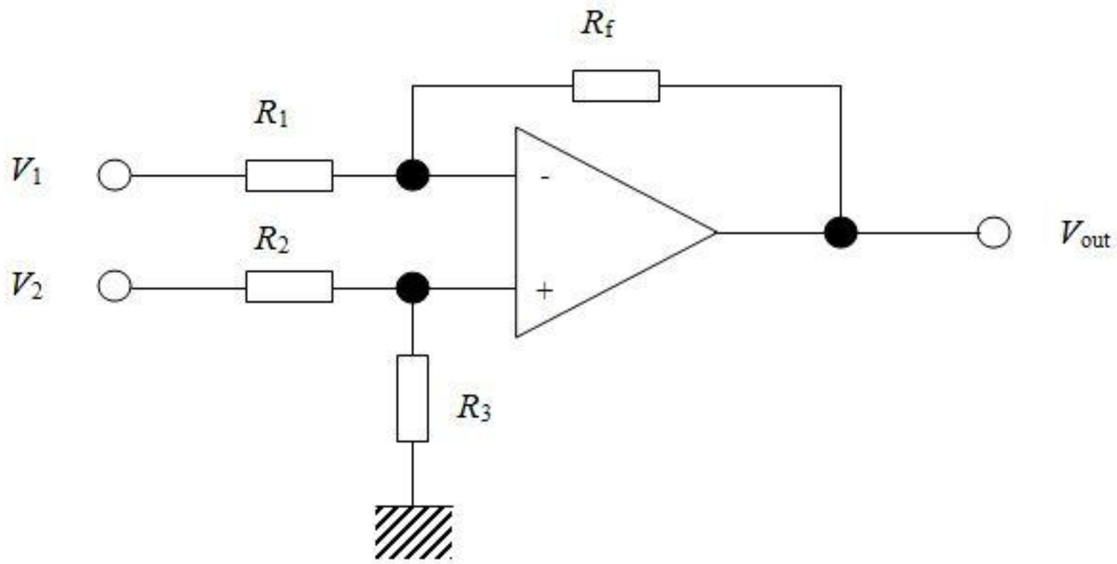


Figure 113 The difference amplifier

The derivation of an equation for the output voltage is long and tedious. It is NOT required for the syllabus. But if you want to see it, see 14E.086.

A simplified difference amplifier equation can be written by making R_1 and R_2 the same, and R_f and R_3 the same. The equation is:

$$V_{out} = (V_2 - V_1)\left(\frac{Rf}{R_1}\right)$$ .... *Equation 67*

In many difference amplifiers, all the resistors have the same value.

The difference amplifier has the disadvantage that the input resistance is quite low and may differ between inputs. Differences in resistance are eliminated if R_1 and R_2 the same, and R_f and R_3 the same. This is the way that the amplifier is normally set up.

The difference amplifier is used to amplify very small differences between two voltages, for example the voltages given off by probes investigating nerve activity in an athlete when he is doing exercise. With the athlete on a cycling machine or treadmill (*Figure 114*), probes are attached to the arms and the legs. The voltage differences (of about 10^{-2} V) are amplified and fed to a computer that can monitor a number of physiological parameters as he exercises.

Figure 114 An athlete's cycling machine (Photo: Steven Fruitsmaak, Wikimedia Commons)

If there are spurious voltages that are affecting both probes due to, for example, mains hum, the difference amplifier blocks these off. This is because each spurious voltage has the same value, and the difference is zero. We say that the spurious signals have been **rejected**, and electronic engineers call this property the **Common Mode Rejection Ratio** (CMRR)

If the values of the input resistances are different, the spurious signals will be amplified.

14E.086 Derivation (Extension only)

This derivation is NOT on the syllabus, and you will NOT be asked about it in the exam. If you study electronics at university level, you may come across it.

Let us apply a voltage of V_1 volts to the inverting input (-) of the amplifier, and zero volts to the non-inverting input (+). We know that the potential at the inverting input is almost zero (virtual earth). We also know that the input resistance of the inverting input is very high indeed, so we can assume that no current goes into the input. So, the current flows like this (*Figure 115*).

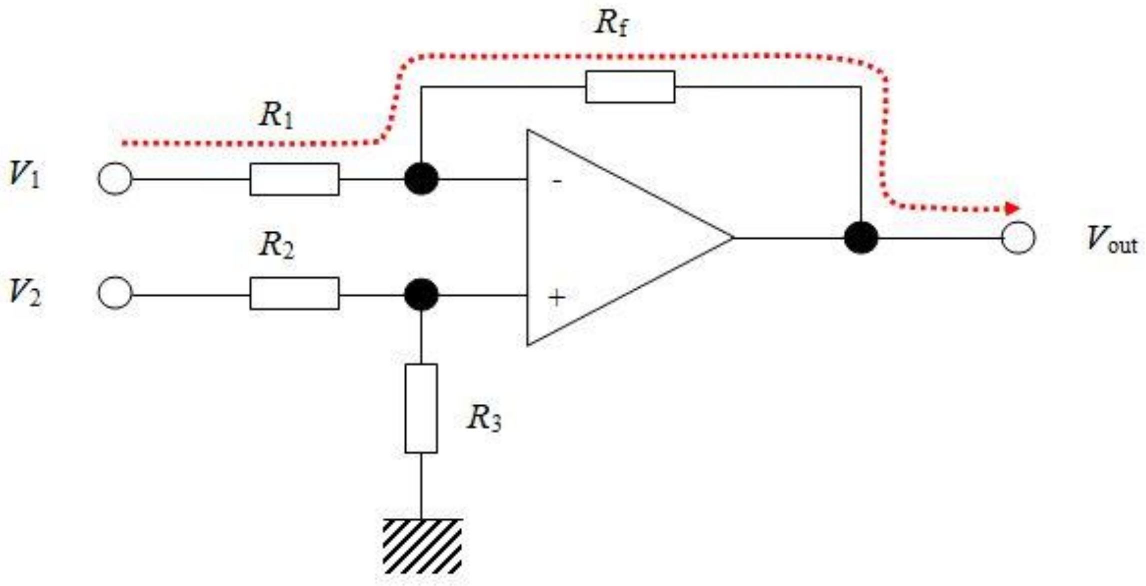


Figure 115 Path of current through the feedback resistor

From Ohm's Law:

$$I = V_1/R_1$$ .... *Equation 68*

Since this is a series circuit, the current through R_f is also I, so we can work out the voltage across R_f. Since the current passes the virtual earth, it is now going up the potential hill, so the voltage across R_f is negative.

$$-V_f = IR_f = R_f(V_1/R_1)$$ .... *Equation 69*

The voltage V_f is the same as the voltage V_{out}, so we can write:

$$V_{out} = -V_1 R_f / R_1$$ .... *Equation 70*

Now we can rewrite this in terms of the gain:

$$\text{Gain} = V_{out} / V_1 = -R_f / R_1$$ .... *Equation 71*

Now let's apply a voltage of V_2 volts to the non-inverting input (+) and zero volts to the inverting input. The current flows like this (*Figure 116*).

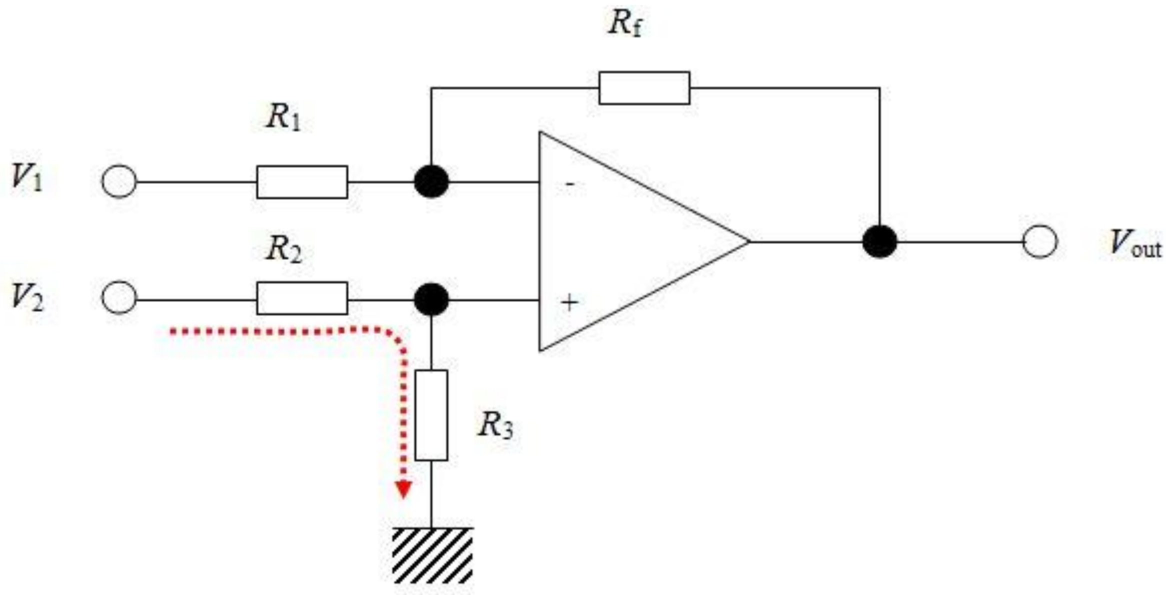


Figure 116 Path of current at the non-inverting input

Now the non-inverting input will NOT be at zero volts. Instead, we have a **voltage divider** which will give us the voltage, V_3, at the non-inverting input.

$$V_3 = V_2 \left(\frac{R_3}{R_2 + R_3} \right)$$ .... *Equation 72*

The voltage difference between the non-inverting input and the inverting input is very little, so we can say that the voltage at the inverting input is V_3 (*Figure 117*).

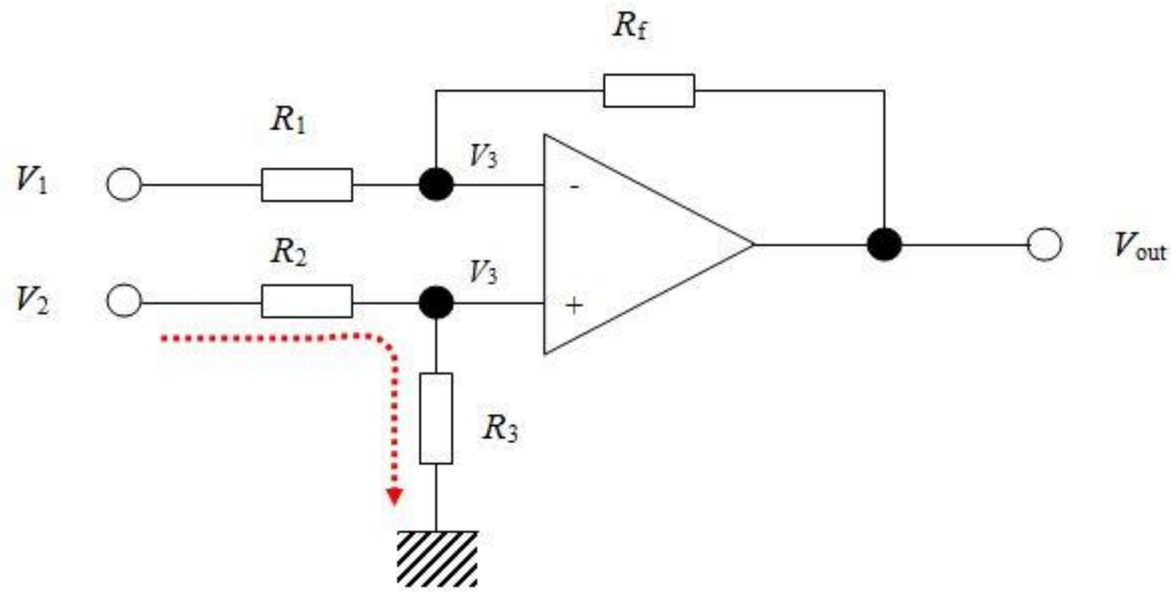


Figure 117 Voltage at each input

Since the voltage across R_1 is going up the potential hill to the inverting input, we can say that the voltage across the R_1 resistor is $-V_3$. So $V_1 = -V_3$.

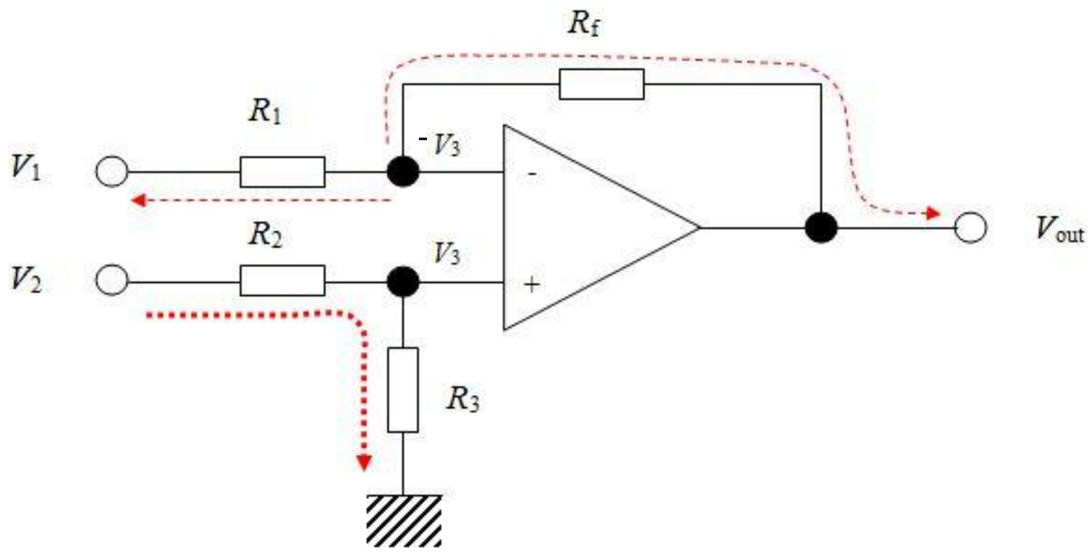


Figure 118 Potentials in the difference op-amp

We know for the non-inverting the V_{out} expression was:

$$V_{out} = -V_1 R_f / R_1 \quad \text{.............. } \textit{Equation 73}$$

The non-inverting input can be considered to be a source of potential difference. Normally this is 0, so we can write:

$$V_{\text{out}} = 0 + -V_1 R_f / R_1$$ .... *Equation 74*

In this case, we have a definite value for potential difference, V_3. Also, we know that $V_1 = -V_3$. This gives us:

$$V_{\text{out}} = V_3 + V_3 R_f / R_1$$ .... *Equation 75*

Now we can write an expression for V_{out}.

$$V_{out} = V_3 \left(1 + \frac{R_f}{R_1}\right)$$ .... *Equation 76*

We also know that:

$$V_3 = V_2 \left(\frac{R_3}{R_2 + R_3}\right)$$ .... *Equation 77*

This gives us:

$$V_{out} = V_2 \left(\frac{R_3}{R_2 + R_3}\right)\left(1 + \frac{R_f}{R_1}\right)$$ .... *Equation 78*

This rearranges to:

$$\frac{V_{out}}{V_2} = \left(\frac{R_3}{R_2 + R_3}\right)\left(1 + \frac{R_f}{R_1}\right)$$

.... *Equation 79*

Now suppose we put a voltage across both inputs. The currents do this (*Figure 119*).

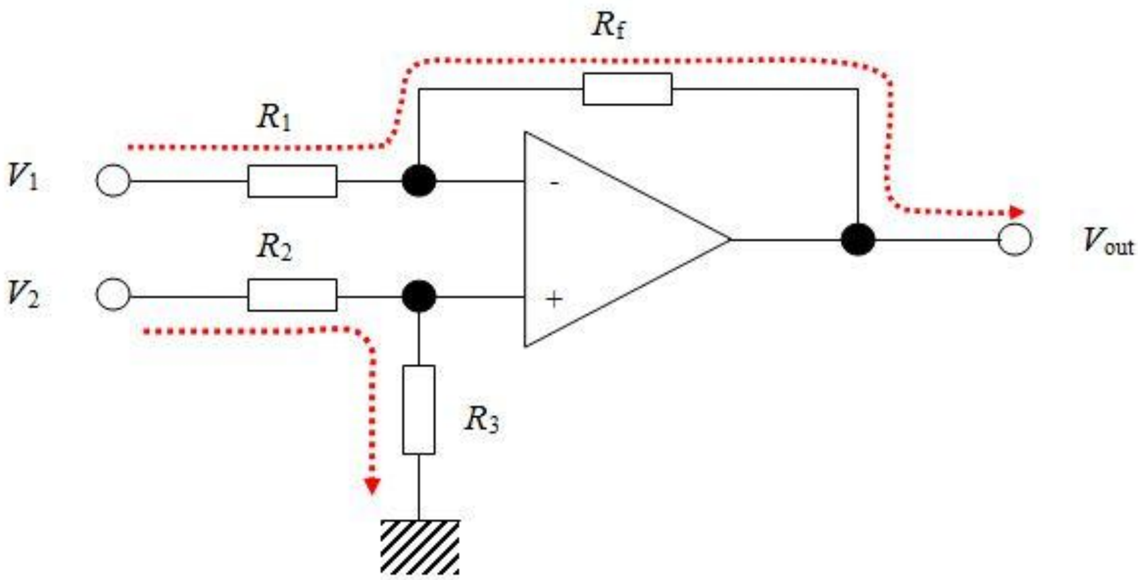


Figure 119 Current flow with a voltage at each input

We can combine the two equations to give us:

$$V_{out} = V_2\left(\frac{R_3}{R_2 + R_3}\right)\left(1 + \frac{R_f}{R_1}\right) - V_1\left(\frac{R_f}{R_1}\right)$$

.... *Equation 80*

Questions

Tutorial 14E.08

14E.08.1

Explain in terms of the difference between the inputs of the amplifier what is meant by a virtual earth.

14E.08.2

If R_a = 1000 Ω and R_f = 100000 Ω, what is the gain of the inverting amplifier?

14E.08.3

Why can we say that the voltage at P is V_{in} for a non-inverting amplifier?

14E.08.4

What is the gain of a non-inverting amplifier with a feedback resistor of 10000 ohms and R_a value of 100 ohms?

14E.08.5

What does the circuit in *Figure 110* remind you of?

14E.08.6

What does *Equation 59* mean in simple terms?

14E.08.7

What is the formula for Ohm's Law?

14E.08.8

Explain why the statement $V_{out} = -(V_1 + V_2 + V_3)$ (*Equation 63*) is true.

14E.08.9

An op-amp has 2 inputs, one having a resistance of 1000 Ω and the other having an input resistance of 5000 Ω. The two inputs have a voltage of + 4 V and + 5 V respectively. The feedback resistance is 2000 Ω. What is the output voltage?

14E.08.10

Suppose V = 5 volts and R = 10 k and R_f = 1k. Use these to work out V_{out} for the binary word 1101.

14E.08.11

A Difference amplifier has input V_1 set at 2 V and input V_2 set at 4 V. The two input resistances are both 50 ohms and the R_f and R_3 are 100 ohms.

Show that the output voltage is 4 V.

14E.08.12

Write an expression for V_{out} if $R_1 = R_2 = R_3 = R_f$.

4. Digital Signal Processing

Tutorial 14 E.09 Simple Logic Systems

AQA Syllabus

Contents

14E.091 Simple Logic Gates

Logic gates are at the heart of **digital electronics**. In digital electronics, we need to know nothing about electricity, other than the difference between **on** (1) and **off** (0).

Digital electronics is widely used in telecommunications, computers, and sound recording. All digital devices are based on these simple building blocks. In a digital camera, there are millions of these gates.

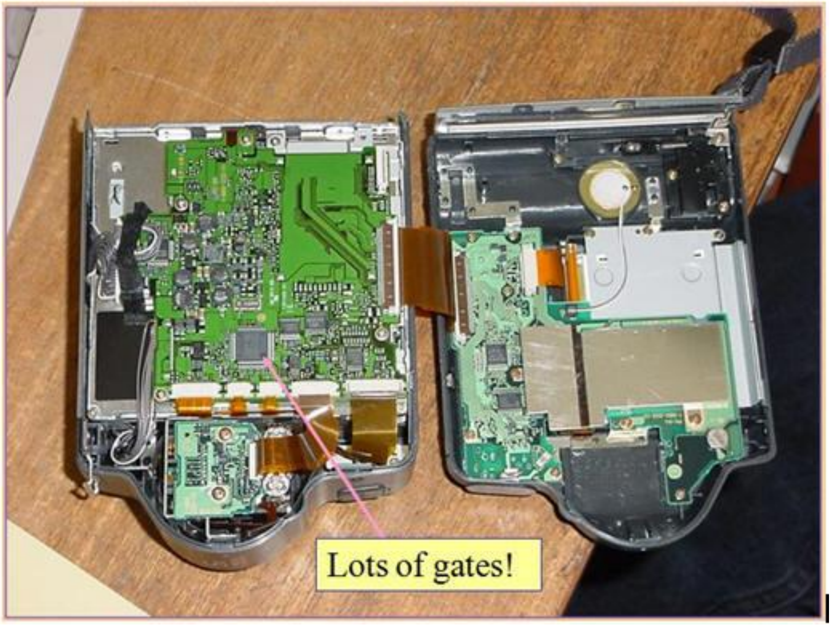


Figure 120 A digital camera has millions of logic gates

14E.092 The NOT gate

The simplest of all the **logic gates** is the **NOT gate**. Notice that when we write the name of a logic gate, we always write it in UPPER CASE letters. **Truth tables** summarise the output condition for a variety of input conditions.

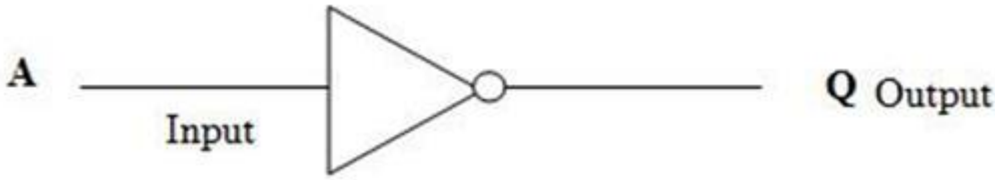


Figure 121 The NOT gate

For the NOT gate the truth table looks like this:

INPUT	*OUTPUT*
0	1
1	0

The NOT gate is often called an **inverter**. Note the circle on the diagram (*Figure 121*). This shows the inverting function. It is made from a single **transistor** that acts as a **switch** (*Figure 122*).

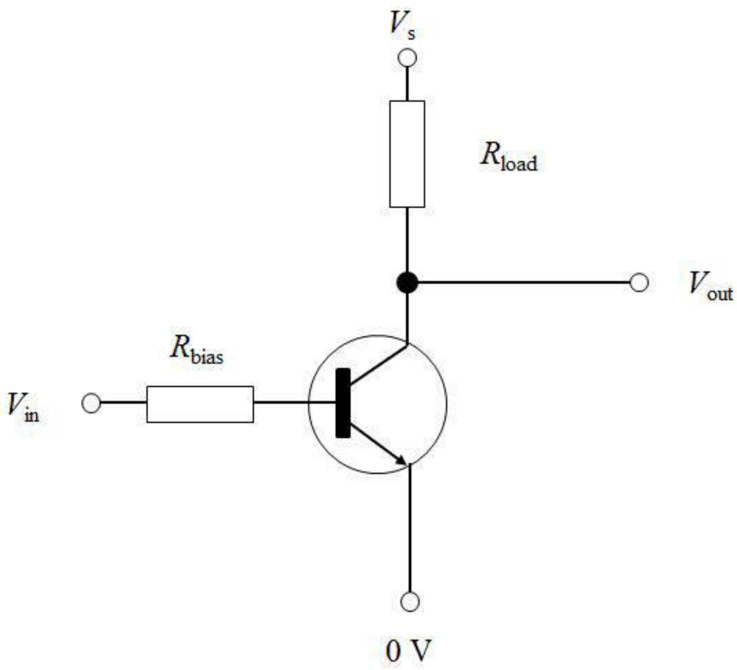


Figure 122 NOT gate from a single transistor

When there is a high voltage at V_{in}, there is a current through the bias resistor. The voltage on the base of the transistor is 0.6 V so that it turns the transistor on. This turns on the current through the load resistor. As the conducting transistor has a very low resistance, we can say that the voltage at V_{out} is very close to zero. Therefore, we can say that the transistor has an **inverting** function.

14E.093 The AND gate

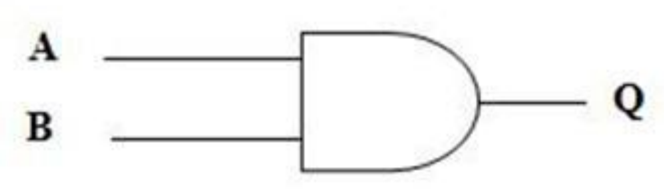


Figure 123 Symbol for an AND gate

For the **AND gate** (*Figure 123*) the truth table looks like this:

A	*B*	*OUTPUT*
0	0	0
1	0	0
0	1	0
1	1	1

We can show the AND gate as a simple circuit of two switches in series. When both switches are closed, the bulb lights up (*Figure 124*).

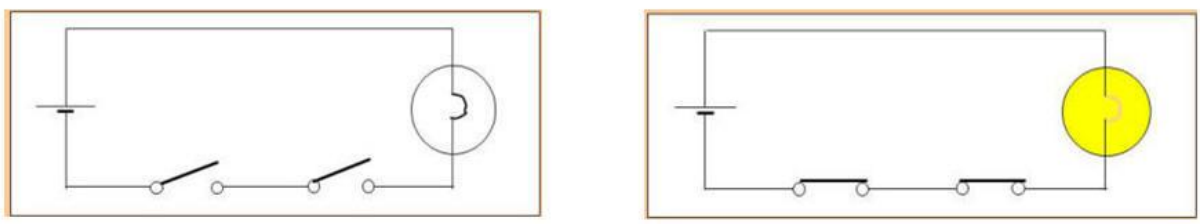


Figure 124 The AND-gate equivalent is two series switches

The AND gate performs the same task as arithmetical **multiplication**. This means that if A = 0, the output is 0, and the same with B. A = B = 1 will give an output of 1.

14E.094 The OR-gate

Figure 125 The OR gate

For the OR gate the truth table looks like this:

A	*B*	*OUTPUT*
0	0	0
1	0	1
0	1	1
1	1	1

The OR gate is the logical equivalent to arithmetical **addition**.

The OR gate can be made from two parallel switches (*Figure 126*).

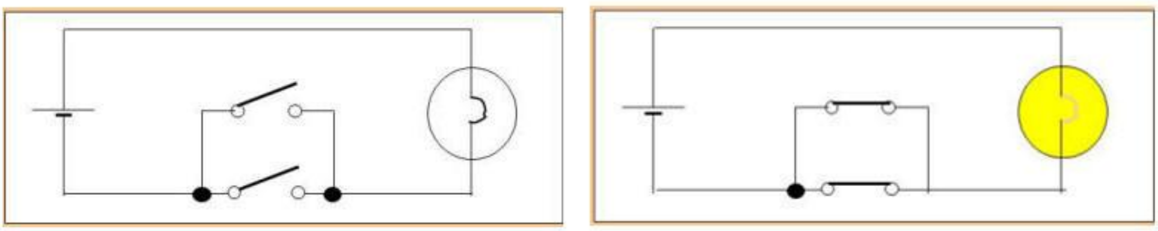


Figure 126 The OR gate is equivalent to two switches in parallel

The light bulb lights up when either of the switches is closed OR both.

Note that we do not get double brightness when both switches are on.

14E.095 The Exclusive OR gate

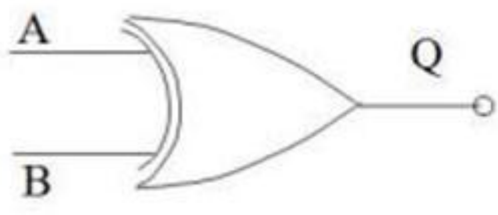


Figure 127 The EX-OR gate

For the EX-OR gate, the truth table looks like this:

A	*B*	*OUTPUT*
0	0	0
1	0	1
0	1	1
1	1	0

The exclusive OR (EX-OR) will only return a high output is either one or the other input is high, but not both.

14E.096 The NAND gate

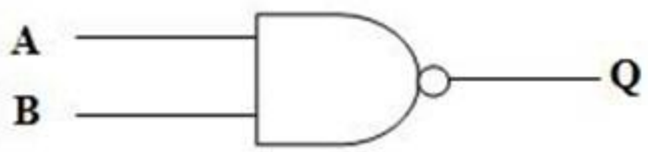


Figure 128 The NAND gate

Note the circle on the output. The small circle on the output line tells us that it has an **inverting** function.

For the NAND gate the truth table looks like this:

A	B	OUTPUT
0	0	1
1	0	1
0	1	1
1	1	0

The NAND gate gets its name from the contraction of NOT and AND. NAND gates are particularly easy to make, and other gates are actually made up of combinations of NAND gates. We will look at this later.

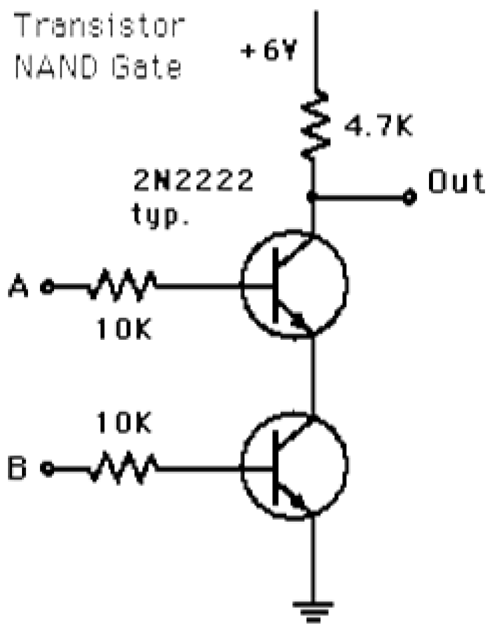


Figure 129 NAND gate made from transistors

14E.097 The NOR gate

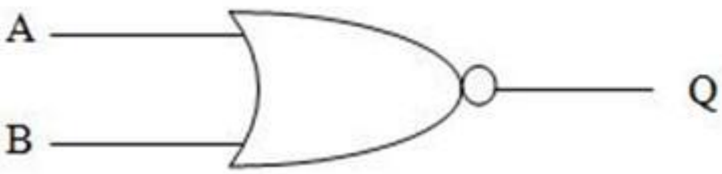


Figure 130 The NOR gate

For the NOR gate the truth table looks like this:

A	*B*	*OUTPUT*
0	0	1
1	0	0
0	1	0
1	1	0

14E.098 Logic Gates and Systems

Logic gates are processor systems. The inputs would be the outputs of two other subsystems, for example a light sensing subsystem and a heat sensing subsystem.

The output of a logic gate provides too little current to drive much more than an LED, even when it is at logic state 1. Therefore, it has to be fed into a **driver** subsystem in order for an output device such as a relay and or a motor to work.

In real world electronics, the logic gates need a voltage of about 5 V to work. Any signal above about 3.5 V is considered to be HIGH. Anything below 3.5 V is LOW. However, you need to be careful, because a LOW output (0 in theory) might give a voltage that is more than enough to turn on a transistor.

Combining Logic Gates

The output of one logic gate can be connected to the input of another. The example is a simple system using a NAND gate with a NOT gate (*Figure 131*).

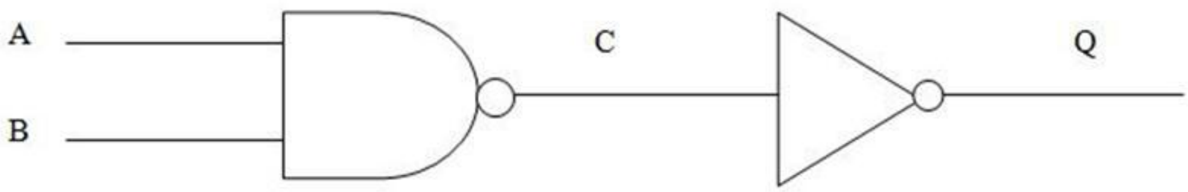


Figure 131 A simple combination of two gates

Answer question 14E.09.2.

You should find that it's the same as an AND gate

Now try something a little more complex (*Figure 132*):

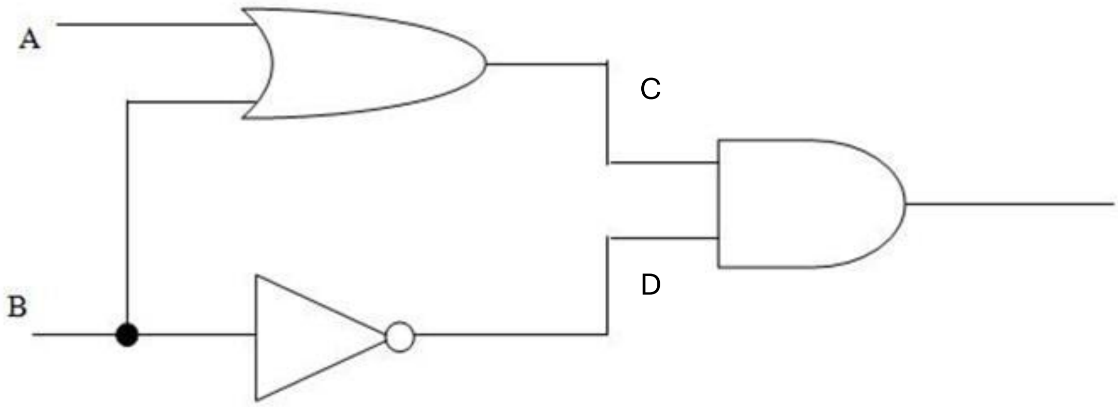


Figure 132 Simple combination of three gates

Answer question 14E.09.3.

14E.099 Making Gates from NAND gates

It is more economical to make circuits from just one kind of chip, and many circuits are made up of just NAND gates. Let us look how (*Figure 133*):

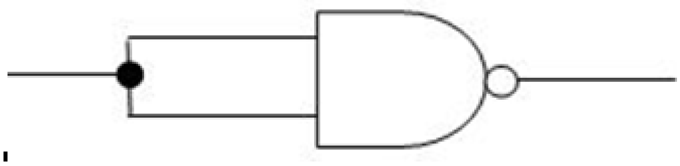


Figure 133 Making a NOT gate from a NAND gate

The two inputs of a NAND gate connected together make the NAND gate into a NOT gate.

In this circuit below the output of a NAND gate is inverted by the NOT gate to produce the output of an AND gate.

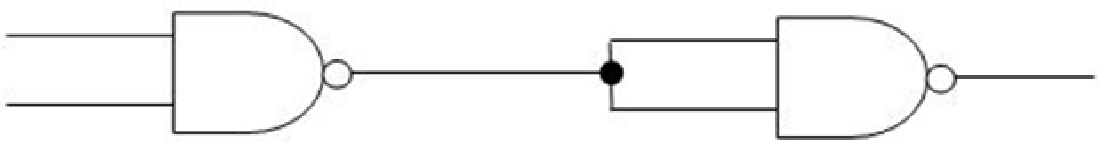


Figure 134 Making an AND gate from two NAND gates

We can show this in a truth table:

A	*B*	*C*	*Q*
0	0	1	0
0	1	1	0
1	0	1	0
1	1	0	1

Look at this circuit (*Figure 135*):

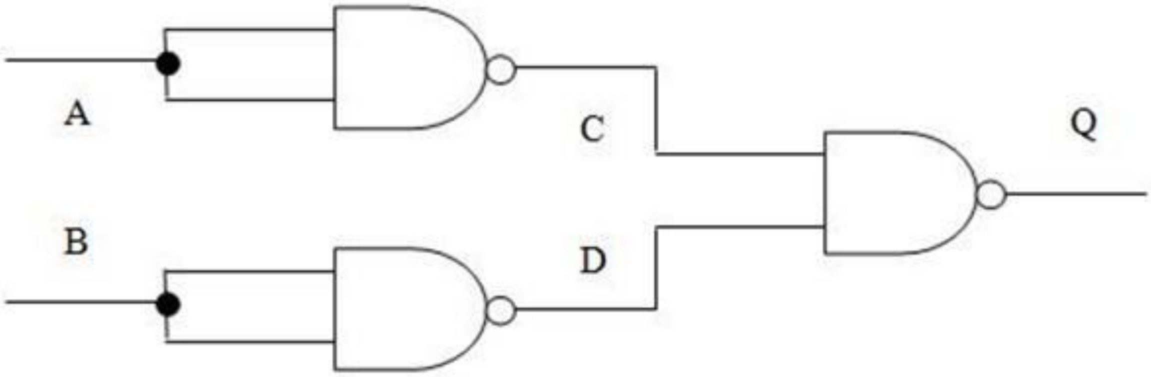


Figure 135 Three NAND gates combined

14E.0910 Boolean Algebra

Boolean algebra was invented by George Boole (1815 - 1864). It was devised for mathematicians working with the (then) new mathematical discipline of **logic**.

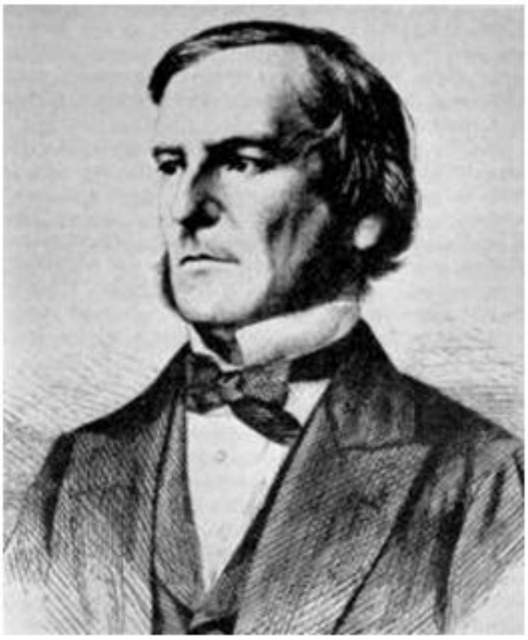


Figure 136 George Boole (1815 - 1864)

In digital electronics it does the same job as a truth table, but with symbols. A Boolean expression tells us **what condition will give an output of 1**.

For a NOT gate the Boolean expression is:

$$Q = \bar{A}$$

.... *Equation 81*

The symbol $\bar{A}$ is pronounced "A-bar", and means that the state Q is opposite to the state Q. So, the statement says, "Q is equal to NOT A".

This means that the output Q is a 1 when A is a 0.

The truth table associated with the statement is:

INPUT	*OUTPUT*
0	1
1	0

For an AND gate the Boolean expression is:

$$Q = A.B$$

.... *Equation 82*

The dot between the A and the B mean that both A AND B have to be 1 for Q to be 1. The expression is pronounced, "Q equals A dot B". The dot is often referred to as "the **dot product**", which is another way of saying multiplication. The equivalent truth table is:

A	B	Q
0	0	0
1	0	0
0	1	0
1	1	1

For the OR gate the Boolean expression is:

$$Q = A + B$$

.... *Equation 83*

This is pronounced, “Q is equal to A OR B” and the truth table is:

A	B	Q
0	0	0
1	0	1
0	1	1
1	1	1

We looked at this example using truth tables. Now we are going to analyse it using Boolean algebra (*Figure 137*).

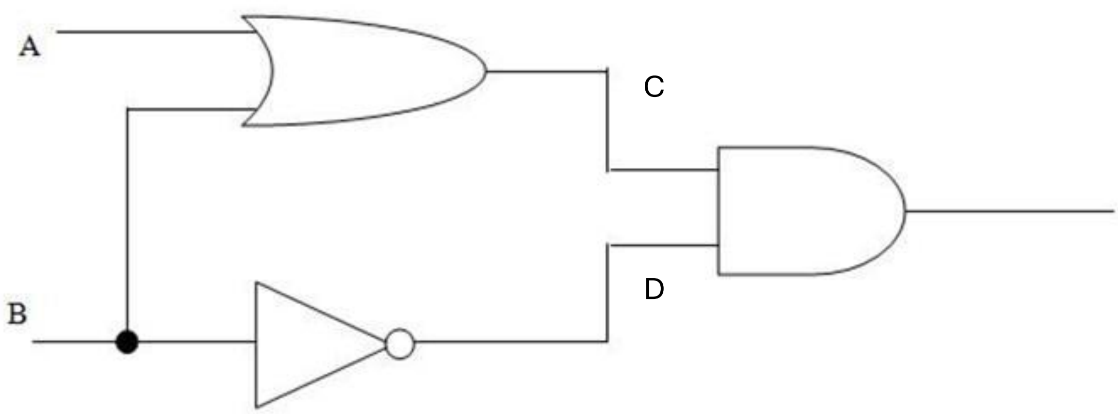


Figure 137 Circuit to use in a Boolean algebra analysis

Analyse this circuit using the questions 14E.095 to 14E.09.8.

Now you will need to look at some rules that will help you to simplify the expression you have just worked out:

Name	Boolean Expression
T4 (Identity Law)	A.A = A
T4	A + A = A
T4	A = A (“NOT NOT A”)
T7	A.0 = 0
T7	A + 0 = A
T8	A + 1 = 1
T9	$A.\bar{A} = 0$
T9	$A + \bar{A} = 1$

The rules help us to simplify a lot of more complex expressions.

We can build a circuit using a Boolean expression, which we will look at now.

A	*B*	*OUTPUT*
0	0	0
1	0	1
0	1	1
1	1	0

1. Look at where the output is 1. It occurs when A is 1 and B is 0, we get a 1, or when A is 0 and B is 1. We can write this in Boolean notation as:

$$Q = \bar{A}.B + A.\bar{B}$$

………….. *Equation 84*

2. This means that we need an OR gate to give us the output Q. So, here is the OR gate with its inputs (*Figure 138*):

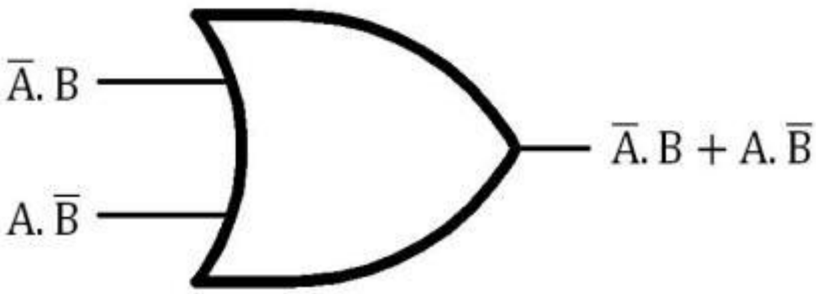


Figure 138 OR gate with inputs in Boolean notation

3. To achieve the inputs, we need to have two AND gates, each of which has ONE of its inputs inverted. This part of the circuit will give out 1 when A = 0 and B = 1.

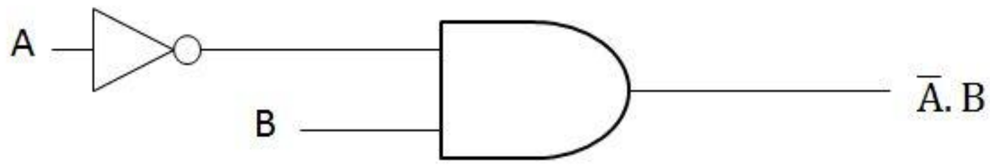


Figure 139 Using an AND gate for input A

4. This part of the circuit will give out 1 when A = 1 and B = 0

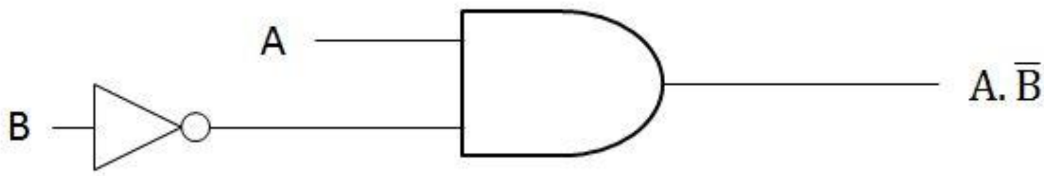


Figure 140 Using an AND gate for input B

5. So now we need to connect each part to the two inputs A and B.

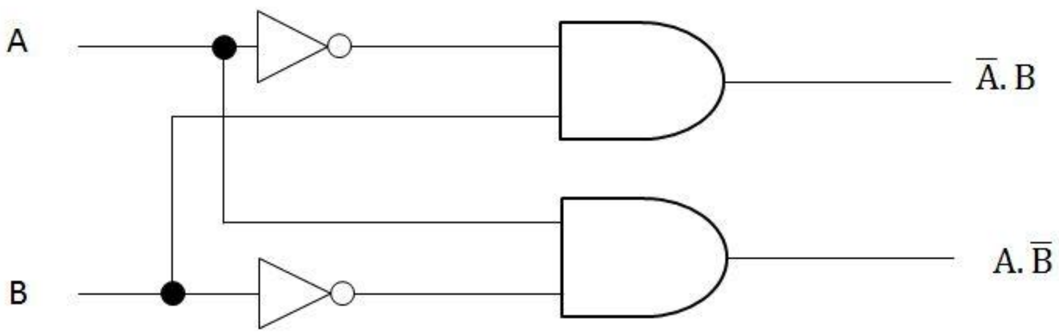


Figure 141 Combining the Inputs

6. Now we need to connect up our circuit to the OR gate:

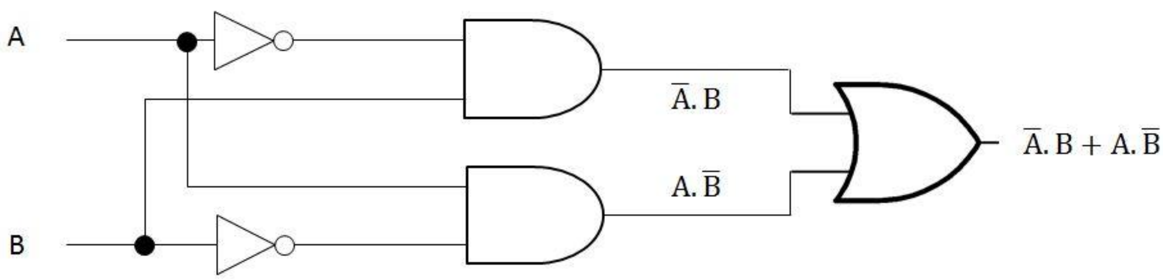


Figure 142 The completed circuit.

When you wire it all up on a logic tutor board (*Figure 143*), you get an impossible tangle of wires!

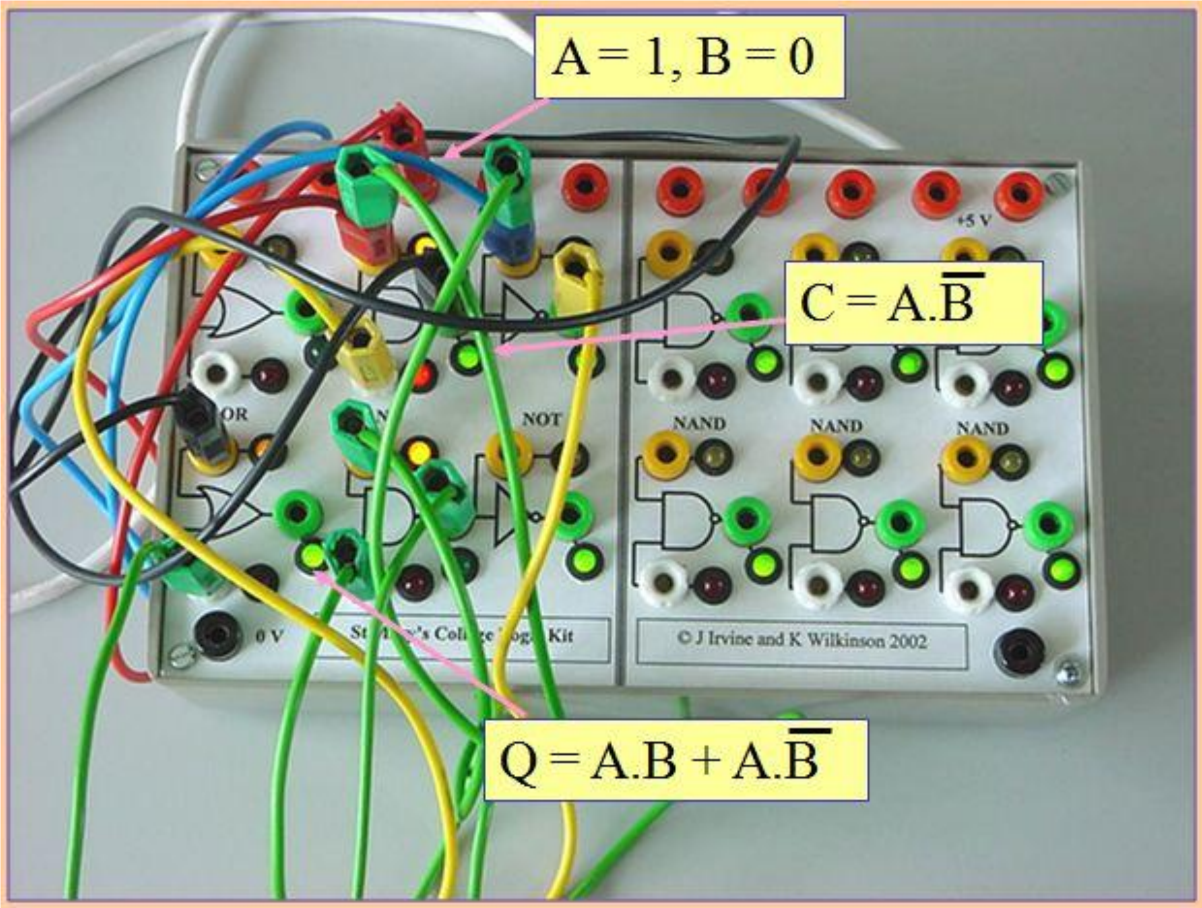


Figure 143 Circuit on a logic tutor board

A reminder of the rules:

Simple Logic Gates

- The **NOT gate** gives a 1 when input A is 0.
- The **AND gate** gives a 1 when both A AND B are 1.
- The **OR gate** gives a 1 when A OR B are 1 or both.
- The **Exclusive OR gate** gives a 1 when A or B are 1 but not both.
- The **NOR gate** gives a 1 when both A and B are 0.
- The **NAND gate** gives a 1 when either A or B are 0.
- The outputs can be summed up in truth tables.

Combining Logic Gates

- The output of one logic gate can be connected to the input of another.
- Truth tables can be made for simple combinations.

Boolean Algebra

- States what the inputs must be for an output to be 1.
- Can simplify the design of circuits.
- Has rules to simplify long expressions.

Questions

Tutorial 14E.09

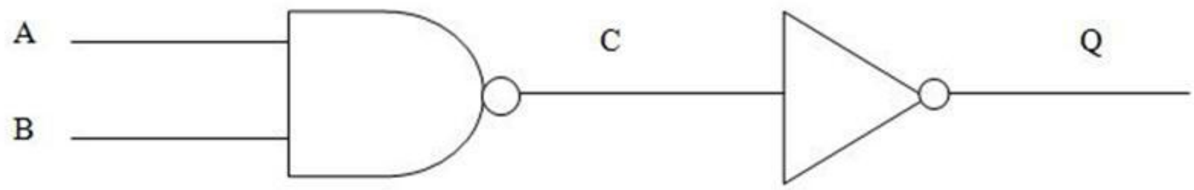
14E.09.1

Match the statement to the gate:

Gives a 1 if both Input A and Input B are 1	OR
Gives a 1 if either Input A or Input B is a 1	AND
Gives a 1 if either Input A or Input B is a 1, but not both.	X-OR
Gives a 0 when both Input A and Input B are a 1.	NAND

14E.09.2

Complete the following table for this circuit:



A	B	C	Q
0	0		
0	1		
1	0		
1	1		

14E.09.3

Refer to *Figure 132* (Page 144) to complete the truth table:

A	B	C	D	Q
0	0			
0	1			
1	0			
1	1			

14E.09.4

Refer to *Figure 135* (Page 145) to complete this truth table:

A	B	C	D	Q
0	0			
0	1			
1	0			
1	1			

14E.06.5 to 14E.09.8 Refer to *Figure 137* on Page 147.

14E.09.5

The output Q is given by the Boolean Expression:

A: Q = C

B: Q = D

C: Q = C.D

D: Q = C + D

14E.09.6

What is the Boolean expression for C?

A: $C = \bar{B}$

B: $C = \bar{A}$

C: $C = A.B$

D: $C = A + B$

14E.09.7

What is D in terms of A and B?

A: $D = \bar{B}$

B: $D = \bar{A}$

C: $D = A + B$

D: $D = A.B$

14E.09.8

What is the output Q in terms of A and B?

A: Q = B.(A + B)

B: Q = NOT A.(A + B)

C: Q = A + B

D: Q = NOT B.(A+B)

14E.09.9

Simplify the last answer using the laws of Boolean Algebra (Page 148).

Tutorial 14 E.10 Sequential Logic Systems

AQA Syllabus

Contents

14E.101 What is a sequential circuit?

We have looked at combinational logic systems in which the output was determined by the combinations of one or more inputs. The output state at any time is dependent on the state of the inputs. In a **sequential circuit**, the output is dependent on:

- The current input to the circuit.
- The previous inputs to the circuit.

In effect the circuit has a **memory.**

Sequential circuits are the basic building blocks of:

- Counters.
- Shift registers.
- Memories.

In **synchronous** sequential circuits, changes in output do not occur immediately there is a change in input, but the next time there is a clock pulse. In **asynchronous** circuits the next stage is triggered by the completion of the previous stage without reference to a clock pulse.

Clock pulses are **square wave** oscillations that are produced by a **pulse generator** that can be based on two kinds of circuit:

- **Astable** generators produce trains of **square waves**.
- **Monostables** produce single pulses.

Circuits are triggered in one of two ways:

- **Level triggering** in which the changes occur when the level of the pulse is either at 0 or 1.
- **Edge triggering** in which the change occurs as the clock pulse rises from 0 to 1(rising or positive edge) or from 1 to 0 (falling or negative edge).

The diagram shows the idea (*Figure 144*).

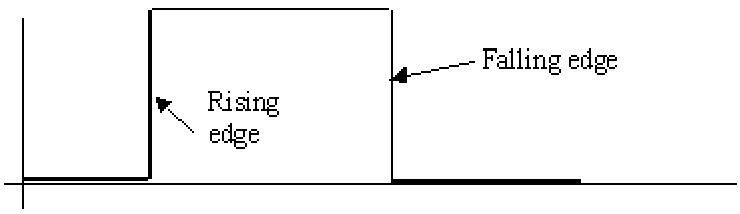


Figure 144 Edge triggering

Level sensitive devices are often referred to as **latches**, while edge triggered devices are called **flip-flops**. Pulses can be provided manually with switches, but they can provide spurious pulses due to **bounce**. They can be de-bounced using a **Schmitt trigger**, which also can be used to clean up noisy signals.

Here some important definitions to learn.

Term	Definition
Combinational circuit	The output is determined by the combinations of one or more inputs.
Sequential circuit	Output is dependent on: • The current input to the circuit. • The previous inputs to the circuit.
Synchronous	Changes in output do not occur immediately there is a change in input, but the next time there is a clock pulse
Asynchronous	The next stage is triggered by the completion of the previous stage without reference to a clock pulse
Triggering	Level triggering in which the changes occur when the level of the pulse is either at 0 or 1. Edge triggering in which the change occurs as the clock pulse rises from 0 to 1(rising or positive edge) or from 1 to 0 (falling or negative edge).
Flip-Flop	Edge triggered devices are called flip-flops.
Latch	Level sensitive devices are often referred to as latches

14E.102 Bistable Latches

Bistables have two stable states; one output remains high while the other remains low. These are **complementary states**. The situation remains until an external input signal such as a clock pulse switches the complimentary states over.

We can use two NAND gates to produce an **S-bar - R-bar latch**. The circuit diagram is shown below (*Figure 145*).

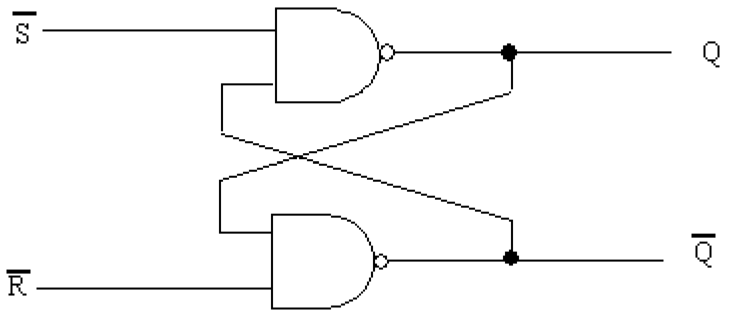


Figure 145 S-bar - R-bar latch

There are two inputs to the latch, the set, **S-bar**, and the reset, **R-bar**. There are two output states, **Q** and **Q-bar** which are complementary to each other. This means that when Q = 0, Q-bar = 1, and *vice versa*. The common symbol for the latch is not the circuit diagram above (*Figure 145*), but either of the alternatives shown (*Figure 146*).

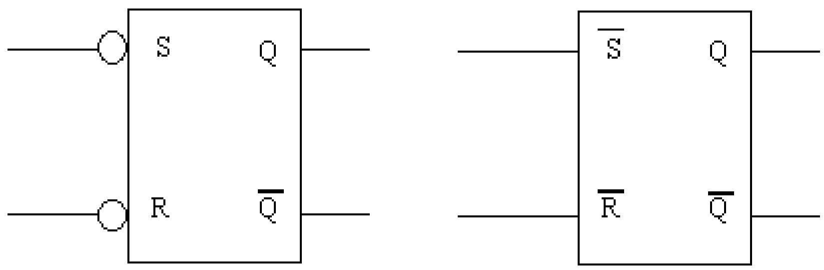


Figure 146 Bistable latch

This bistable is the industry standard, made from NAND gates. We say that its inputs are **active-low** which means that the state changes when the inputs go low.

We can draw up a truth table, often called a **transition table** for the circuit. We can also show what is happening in a **timing diagram**, which is three voltage-time graphs stacked one on top of the other (*Figure 147*).

S-bar	R-bar	Q	Q-bar	Notes
0	1	1	0	S-bar = 0 sets Q = 1 (SET)
1	1	1	0	Outputs remain in previous states
1	0	0	1	R-bar = 0 sets Q-bar = 1 and Q = 0 (RESET)
1	1	0	1	Outputs remain in previous states
0	0	0	0	Indeterminate state (not allowed)

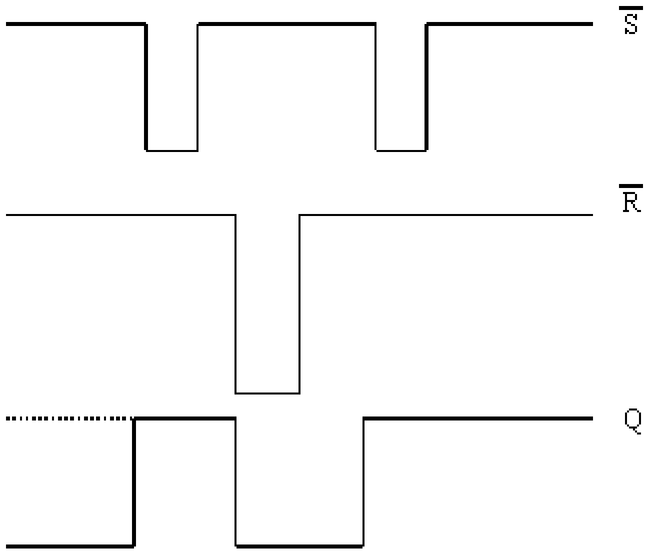


Figure 147 Timing diagram

When S-bar falls from 1 to 0, there is no effect until the R-bar falls from 1 to 0. Then the output Q changes from 1 to 0. Then R-bar goes to 1, but there is no change in Q until S-bar goes to 1.

This latch is a circuit that can be used to de-bounce a switch, cleaning its action to get rid of unwanted pulses. The layout is shown in the diagram (*Figure 148*):

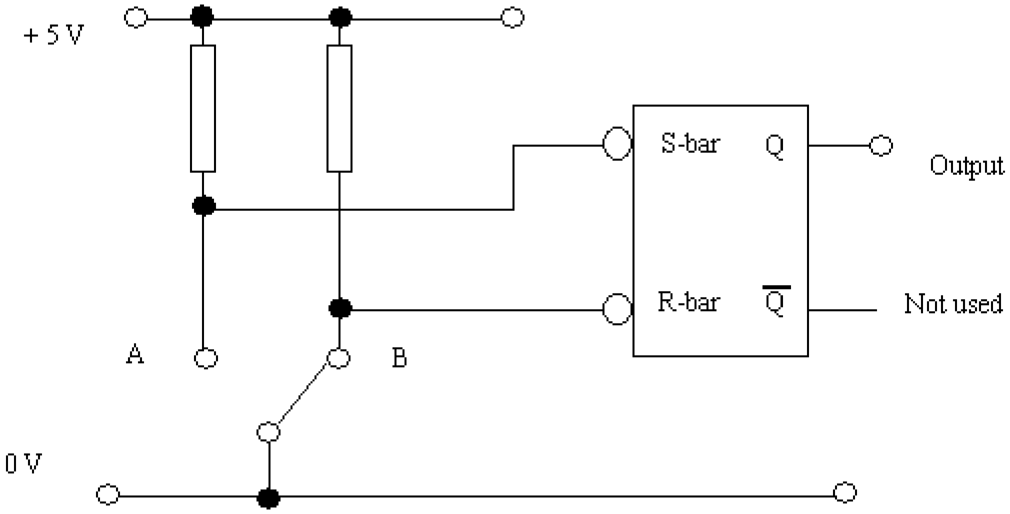
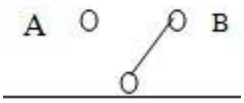
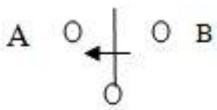
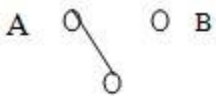
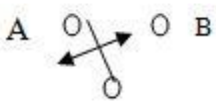
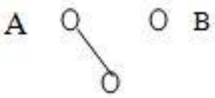


Figure 148 Debouncing a switch

Let us analyse the circuit as the switch is moved from B to A.

Switch Position	*S-bar*	*R-bar*	*Q*
	1	0	0
	1	1	0
	0	1	1
	1	1	1
	0	1	1

Notice how the output does not change as the switch bounces.

There is one disadvantage about the S – R bistable circuit, and that is what happens when both the inputs are 0. This is an **indeterminate** state, and the output is not predictable. We cannot say if the bistable will return to the SET or the RESET state. We can avoid this by ensuring that the inputs are changed alternately.

14E.103 The D-type Flop-Flip

Latches can be used to act as **memories** but have a major problem. They can be what is called **transparent**. If one of the inputs is high, and the other is connected to a clock impulse, the output will change as the clock pulses pass through. Let us think about this more using the S – R latch (made from NOR gates). Data are put in through the S input while the R input is connected to a square wave pulse generator as below (*Figure 149*).

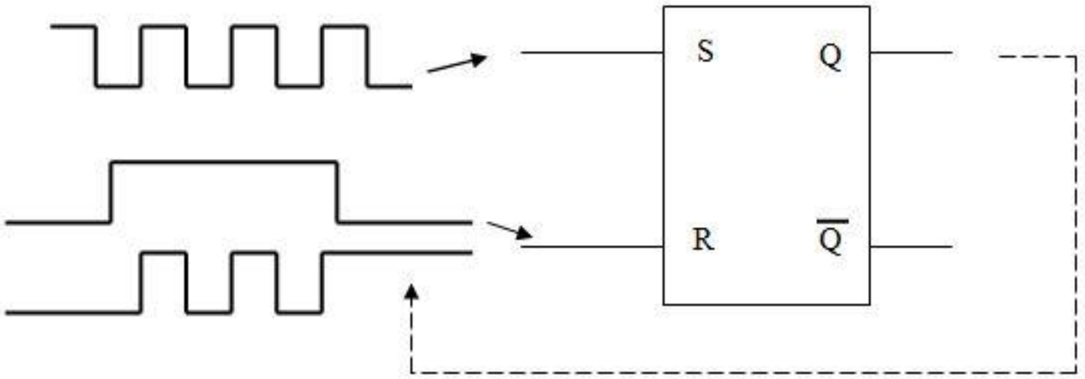


Figure 149 Transparency in the SR latch

This latch is active high, which means that it changes state when S or R goes high. We start off with R low and S changing from low to high and back again. The output Q is low. Then we change R to high and this should give us a high at the output Q. We get that initially, but when the input S goes low, the output goes low. When it goes high, the output goes high as well. This situation lasts as long as R stays high. So, we get the clock pulse passing through the latch, which is why we call it **transparent**. This can be a nuisance in computers and other systems where data changes rapidly.

We can overcome this problem by using **edge-triggering**. Bistables that use edge triggering are called **flip-flops**. Flip-flops do not have this problem with being transparent. The **D-type flip-flop** is the basic design unit for **sequential** circuits, which are circuits whose outputs change with time. The symbol with the D-type flop-flip is shown (*Figure 150*).

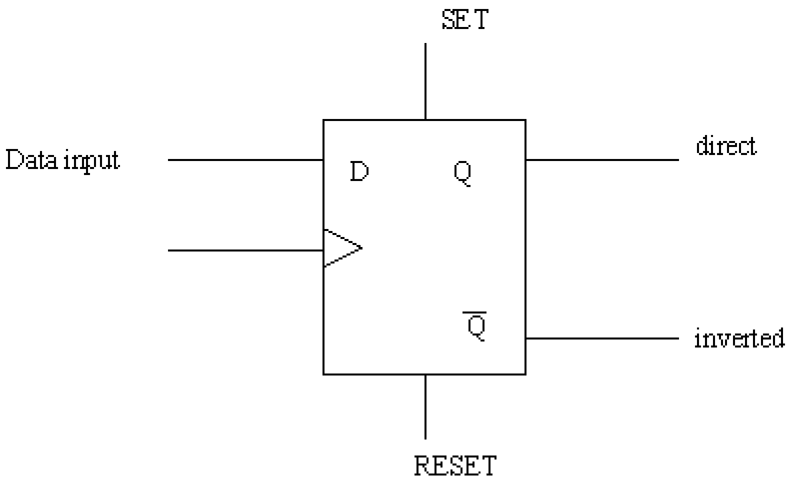


Figure 150 D-type flip-flop

We should note the following about the flip-flop:

- The outputs are complimentary. When Q is 1, Q-bar is 0 and *vice versa*.
- Terminal S and R are there to set and reset the flip-flop. Signals at these inputs take priority over the other two inputs
- The **data input** takes in the data, while the **clock input** takes in the clock pulses. The triangle indicates edge triggering. An upward pointing arrow indicates positive edge triggering, while a downward arrow shows negative edge triggering.
- CMOS flip-flops are active high, while many TTL flip-flops are active low.

The behaviour of the flip-flop is shown (*Figure 151*):

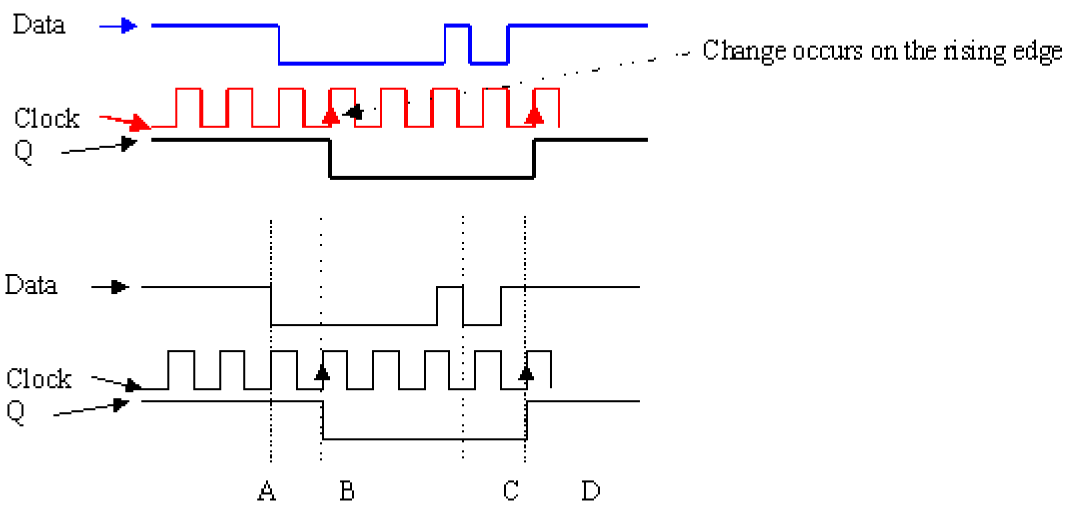


Figure 151 Timing diagram for a D-type flip-flop

When we send a pulse down the SET or RESET lines, the results can be shown in the truth table:

S	R	Q	Q-bar
0	0	1/0	0/1
0	1	0	1
1	0	1	0
1	1	1	1

The circuit shows how the D-type flip-flop can be made using NAND gates (*Figure 152*):

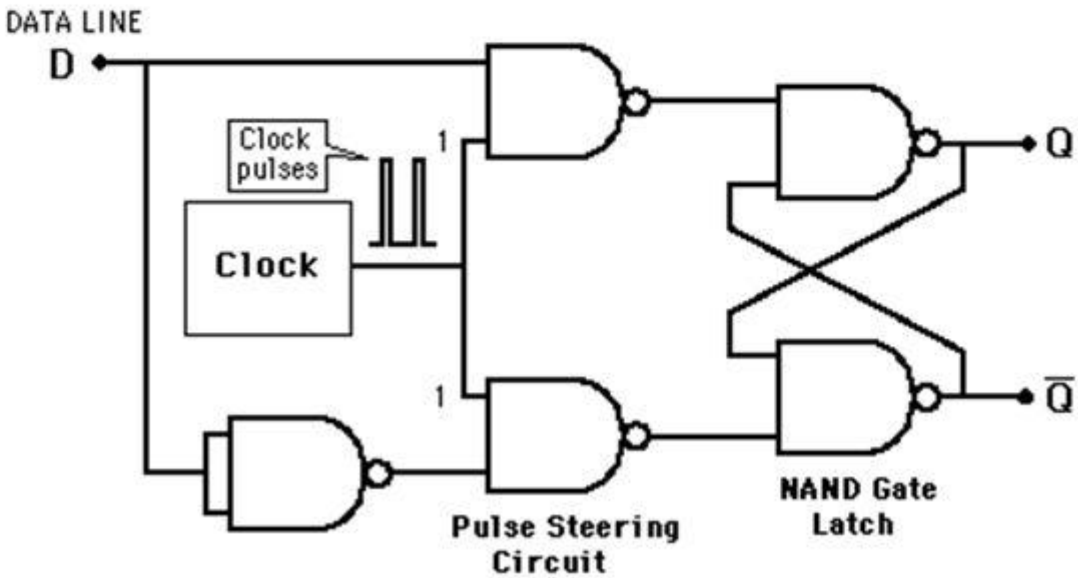


Figure 152 D-type flip-flop from NAND gates

14E.104 Binary and Hexadecimal Counting

Computers work on **binary** numbers, which means numbers to the **base 2**. We normally count in tens, because we have ten digits on our front and hind paws. This is **base 10** or **decimal**. We can count in any base we like. In the UK the currency was run on a duodecimal system, counting in base twelve. 1 shilling = 12 old pence; 1 pound = 20 shillings. This lasted until 1971.

When we express a number, we start off with the **most significant bit** at the left hand side, and the **least significant** at the right. Consider the number 245:

	Hundreds	*Tens*	*Units*
Powers of Ten	10^2	10^1	10^0
Number	2	4	5
	2 × 100	4 × 10	5 × 1

So, 245 is the sum of 200 + 40 + 5

We do a similar thing in **binary**. The least significant bit is 2^0 (= 1), followed by 2^1 (= 2), followed by 2^2 (= 4), etc. A **bit** is a **b**inary dig**it**. We will look at a **four bit** number:

	Eight	*Four*	*Two*	*Units*
Powers of Two	2^3	2^2	2^1	2^0
Number	1	0	1	1
	1 × 8	0 × 4	1 × 2	1 × 1

So, 1011 = 8 + 0 + 2 + 1 = decimal 11.

Computer memories are designed to act rather like a set of pigeonholes, or lockers, in which data is posted. Each location has a unique address, which is given a number in base 16 (= 2^4), called a **hexadecimal** or **HEX** code.

The first 9 hexadecimal numbers are like the first 9 decimal numbers. The character 10 represents decimal 16. So, there have to be alternative characters for decimal 10, 11, 12, 13, 14, and 15. These are A, B, C, D, E, and F respectively. The table shows decimal numbers 0 to 16 with their four bit binary and hexadecimal codes:

Decimal	***Four-bit binary***	***Hexa-decimal***
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F
16	0000	10

The binary code for decimal 16 or hexadecimal 10 is 10000, which is a five bit number.

Address codes of four figures give 16^4 (= 65536) combinations. In real computers the addresses can be 16-digit (16-bit) codes or even 32-digit, which give many more combinations. Modern computers now have 64-bit codes.

14E.105 Counters

Latches can act as a memory for binary numbers that have been put into them. Eight flip-flops can act as a memory for an eight-bit **word**, or a single **byte**. In computers, which work with bytes, each eight-bit number stands for something, be it a letter, a number, or a character, determined by the **ASCII code**, universal for all computers.

Counters are special memories that store a word that represents the number of pulses that have passed into the circuit. The D-type flip-flop is the simplest counter for 1 bit (*Figure 153*).

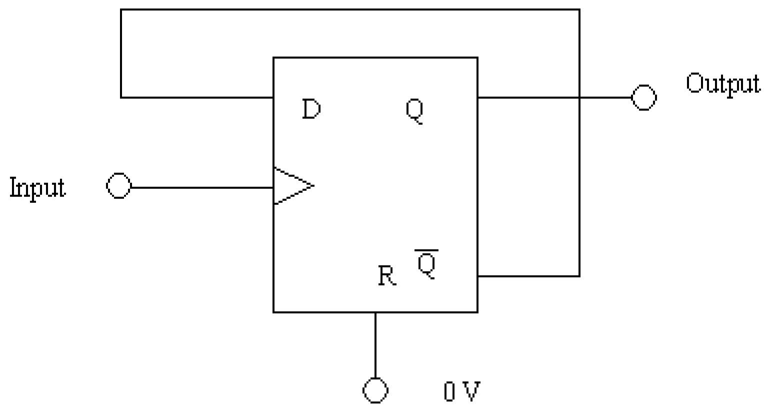


Figure 153 D-type flip-flop as a counter

The timing diagram (*Figure 154*) shows the behaviour of the circuit:

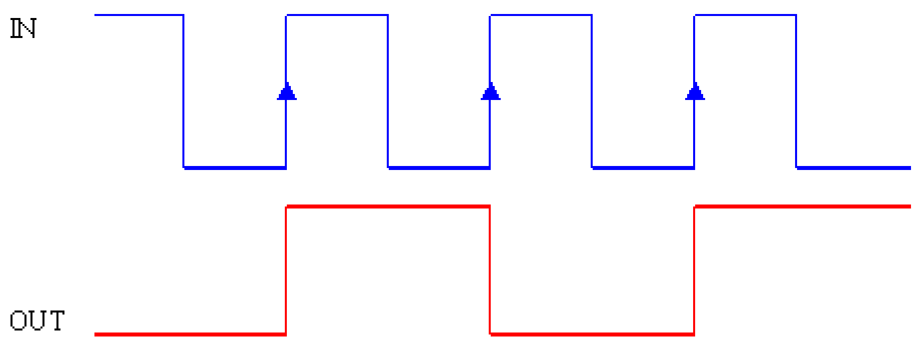


Figure 154 Timing diagram for a D-type flip-flop as a counter

This circuit (*Figure 155*) can count two bits:

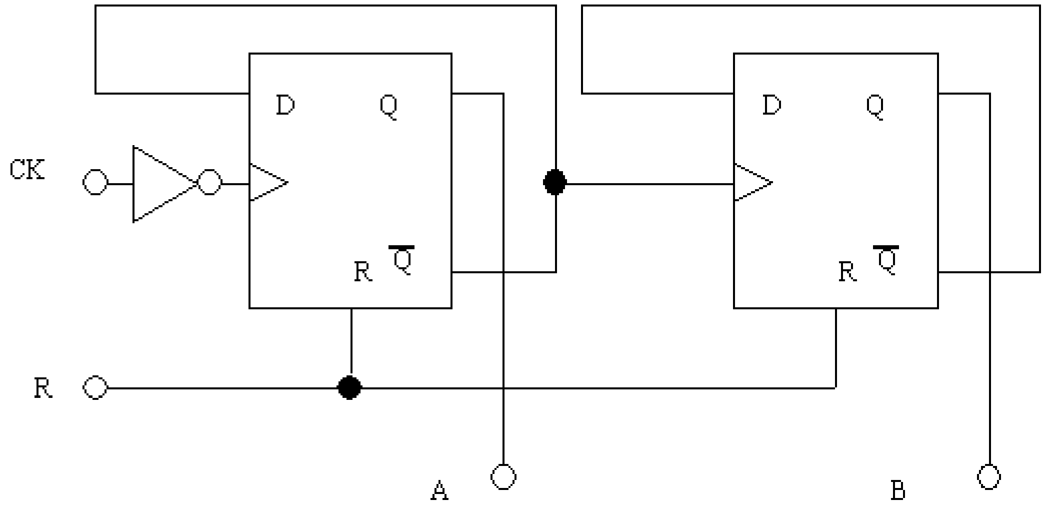


Figure 155 2 bit counter

When the reset is made high, the outputs Q go to give a two bit word BA 00. The circuit then changes state on the falling edge of the clock pulse, according to the following table:

Pulse	B	A
0	0	0
1	0	1
2	1	0
3	1	1
4	0	0

The word BA tells us the number of pulses that have arrived. B is the **most significant** bit worth 2^1 (= 2), while A is the **least significant** bit worth 2^0 (= 1). So, BA = 11 represents 2 x 1 + 1 x 1 = 3. The timing diagram is like this (*Figure 156*):

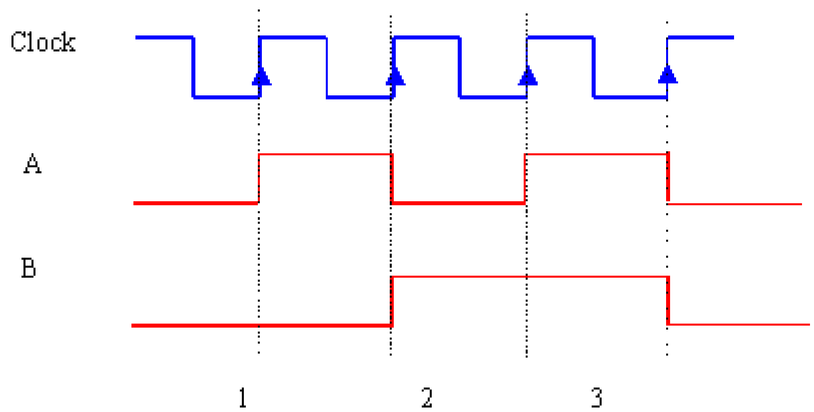


Figure 156 Timing diagram for a two bit counter

14E.106 Divide by 2 Circuit

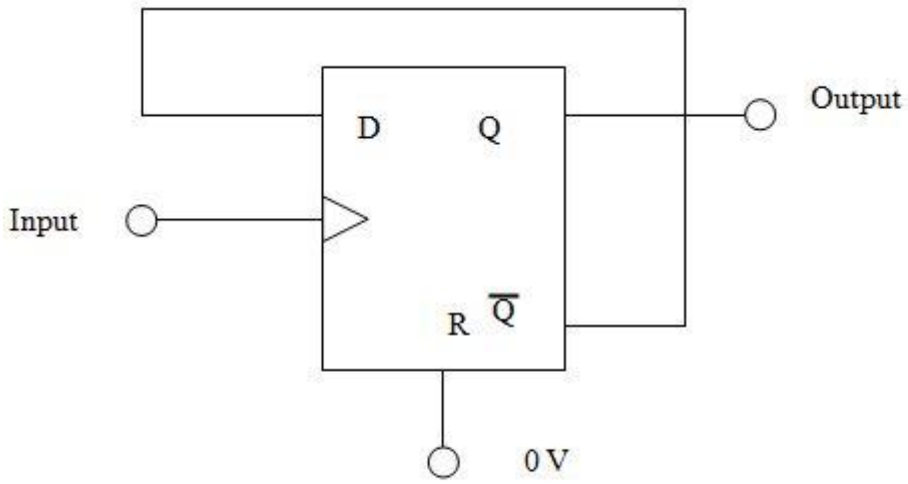


Figure 157 Divide by 2 circuit

As well as acting as a single bit counter, the circuit above, which is a D-type flip-flop with feedback, acts as a **divide by two circuit**. If we look at the timing diagram, we can see that the number of output pulses is half the number of clock pulses (*Figure 158*).

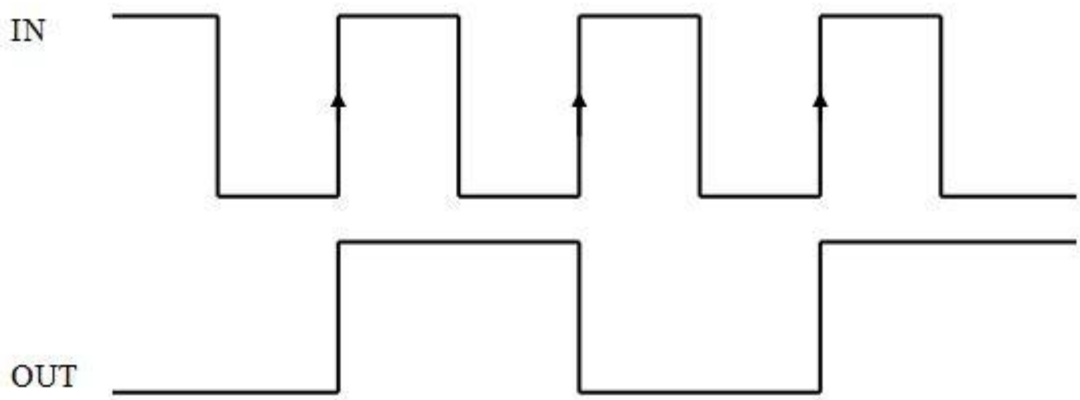


Figure 158 Timing diagram for a divide by 2 circuit

The circuit works like this:

- The output Q-bar is 1 when Q is 0.
- On the first rising edge, Q changes to 1.
- And stays there as the clock pulse falls to 0.
- Then on the next clock pulse, the rising edge causes the output to change to 0.
- The output changes every other clock pulse.

14E.107 4-bit Counters

We can cascade flip-flops so that we can have as many bits as we want. The next diagram shows a four bit counter (*Figure 159*):

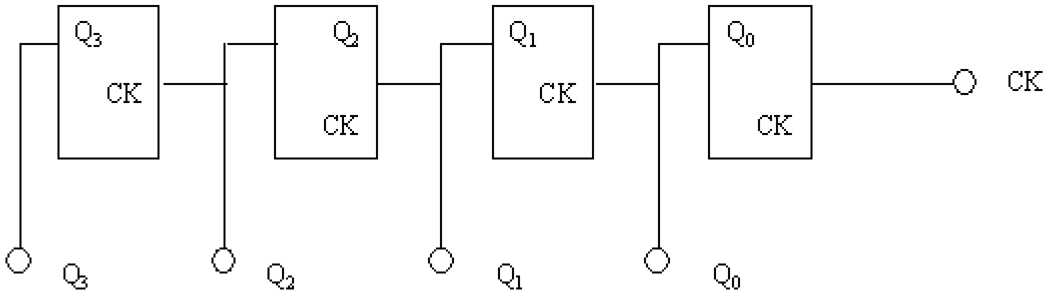


Figure 159 4-bit counter

This circuit is rising edge triggered, and each flip-flop has its Q-bar output fed back to the data input. The timing diagram shows the idea:

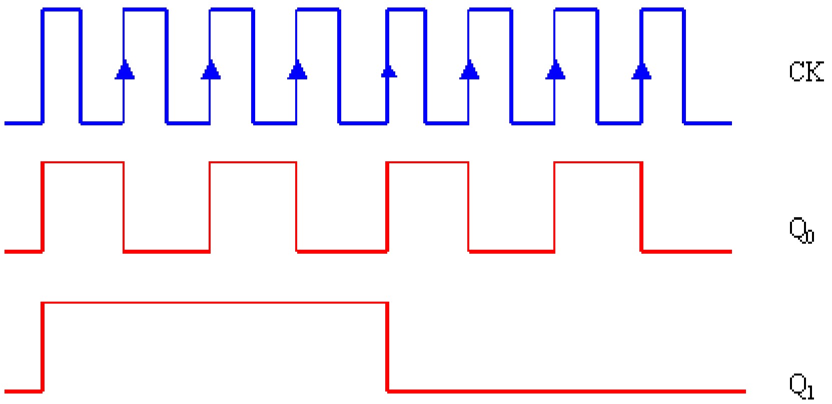


Figure 160 How the output changes for the least significant bits for a 4-bit counter.

Notice that the unit counter goes through a change every two clock pulses, and the twos counter every four pulses. The fours would be every eight pulses, and the eights every sixteen pulses. If we look at the output of the counter, we will see it increase by 1 every two clock pulses. This is an **up-counter**.

To make a **down-counter**, we connect the Q-bar output to the CK input of the next flip-flop, while the Q output is connected to the data input (*Figure 161*).

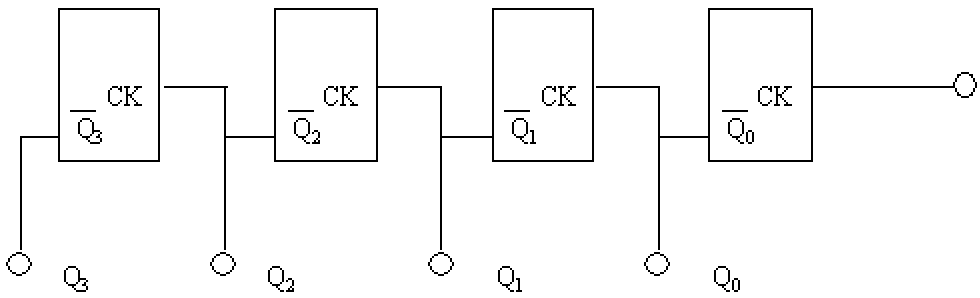


Figure 161 Down-counter

14E.108 4-bit Ring Counter

A 4-bit ring counter (*Figure 162*) is a serial input parallel output shift register (SIPO) that has the output of the last counter which has a line connecting the final output with the input.

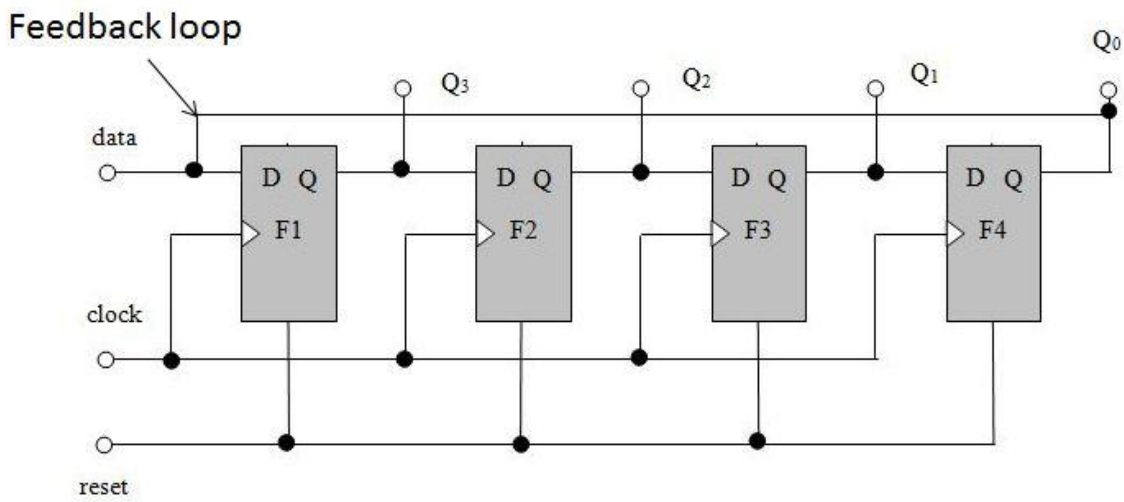


Figure 162 4-bit ring counter

It works like this:

1. All the flip-flops are set to 0 by sending a pulse along the reset line.
2. The input is set to 1 before the clock is set going. This makes Q_3 to be 1. Q_0 is 1 as well because it's directly connected to the input.
3. On the next clock pulse the 1 in Q_3 is passed to Q_2. The 0 in Q_1 is passed to Q_0. So the input becomes 0.
4. Then the 1 is passed from Q_2 to Q_1. Q_0 is 0 and the input is 0. Q_3 and Q_2 are 0 as well.
5. Finally, the 1 gets passed to Q_0 and is fed back to the input.

We can sum this up in this table:

Clock	*Q_3*	*Q_2*	*Q_1*	*Q_0*
Start	1	0	0	1
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1
4	1	0	0	0

So, the 1 gets cycled around. The idea is shown in the diagram (*Figure 163*):

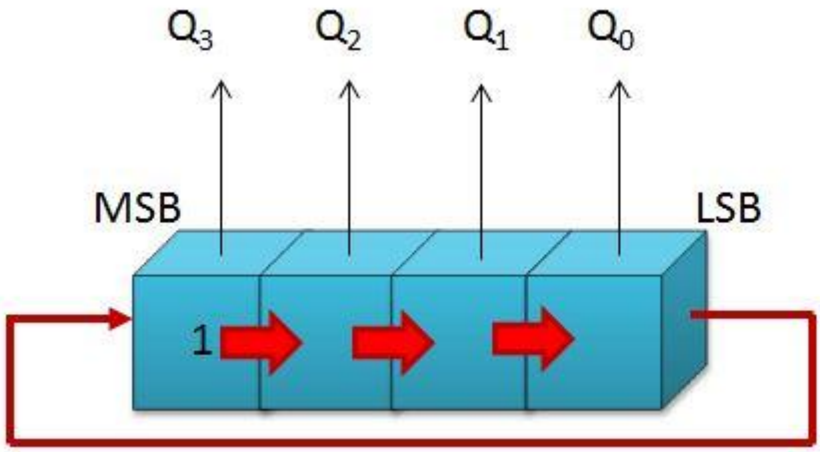


Figure 163 Cycling a 1 through a 4-bit ring counter

This is often called a **modulo-4 counter** as it has 4 states before it repeats itself. If you wanted 8 states, you would have 8 flip-flops to make a modulo-8 counter. Similarly, you would need 16 flip-flops to make a modulo-16 counter. This can be rather inefficient.

14E.109 Johnson Counter

This is a very similar counter, except that the feedback is taken from the Q-bar output. Alternatively, you can put a NOT-gate into the feedback, as has been done here (*Figure 164*):

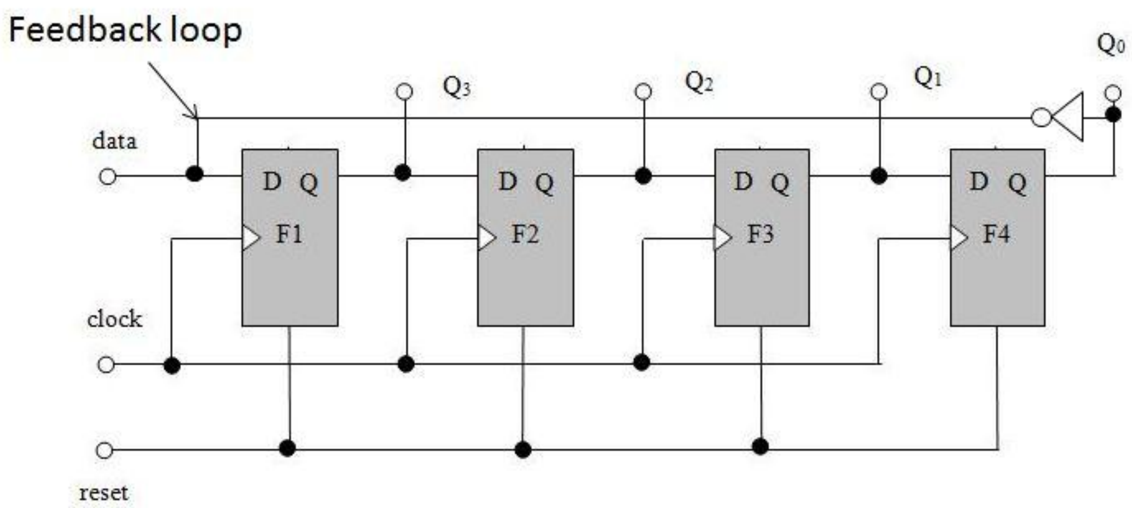


Figure 164 Johnson counter

The inversion causes the counting to be done in a different way, shown in the table below:

Clock	Q_3	Q_2	Q_1	Q_0
Start	0	0	0	0
1	1	0	0	0
2	1	1	0	0
3	1	1	1	0
4	1	1	1	1
5	0	1	1	1
6	0	0	1	1
7	0	0	0	1
8	0	0	0	0

While the normal 4-bit counter can circulate 4 different combinations, the Johnson counter can circulate 8 different combinations. Its modulus is twice the number of the flip-flops. The Johnson counter is often called a **Modulo-2n** counter.

A 3-stage Johnson counter can be used as a 3-phase square wave generator, with each phase being 120° apart. A 5-stage Johnson counter can be used as a **decade counter**.

14E.1010 Modulo N Counters

The 4-bit counters we have seen above count from 0 to 15 (decimal) before resetting to zero. We say that it is a **modulo 16** counter. The **modulo** refers to the number of states that a counter goes through until it resets to zero.

A counter with *n* flip-flops will go through 2^n states before it resets to zero.

If we want to reset to zero before that 2^nth state is reached, we need to add an AND gate to the circuit and feed the output of the AND gate to the reset line. The diagram shows the four bit counter with the AND gate feeding the reset line (*Figure 165*):

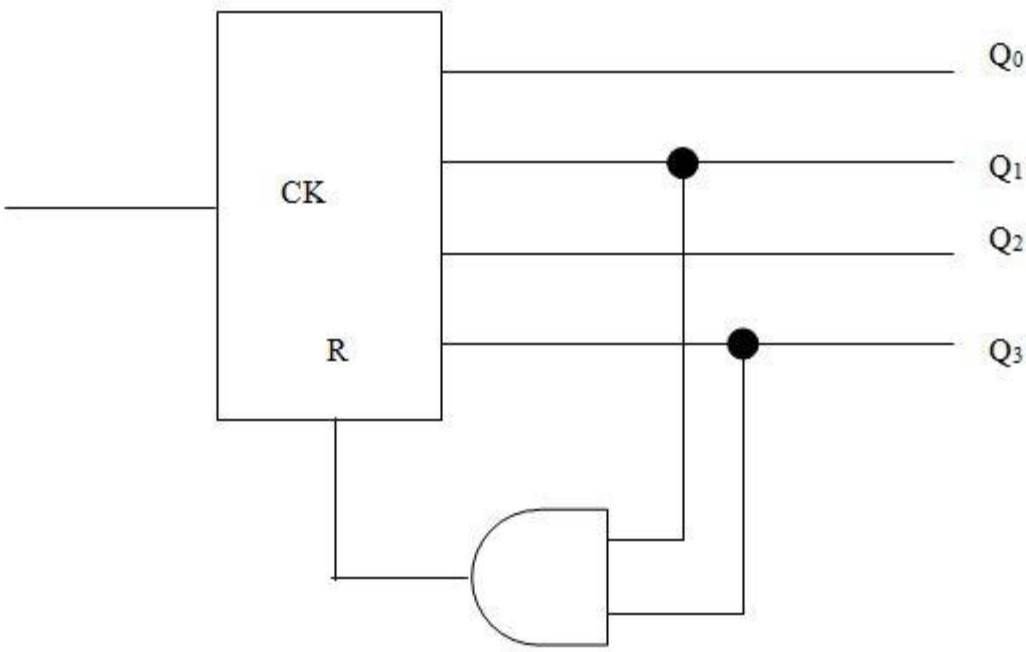


Figure 165 4-bit counter with an AND gate for resetting (Modulo 10)

The counter counts up to binary 1010 (decimal 10). Since Q_1 and Q_3 are 1, the output of the AND gate is 1 and that makes the reset line 1, knocking the counter back to zero. This circuit is called a **binary coded decimal** (BCD) counter. We could choose any of the lines. The AND gate between Q2 and Q3 would give us a modulo 12 counter (*Figure 166*).

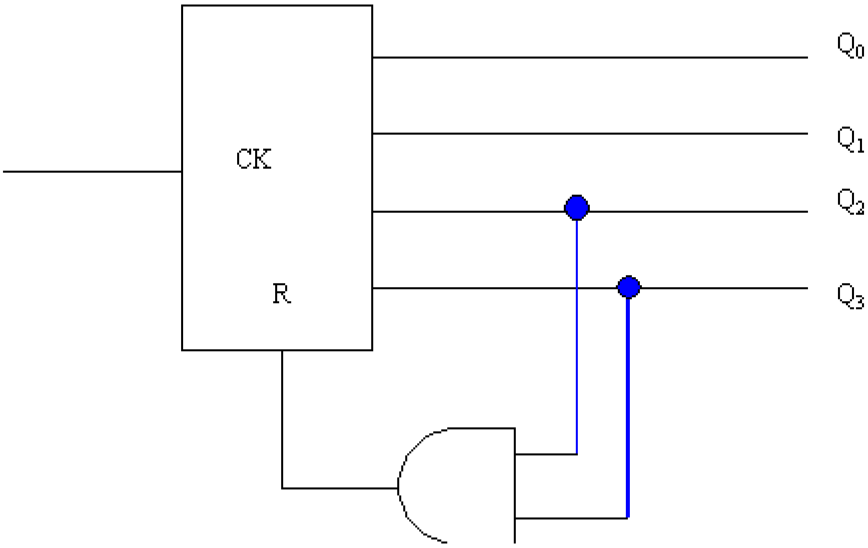


Figure 166 Modulo 12 counter

Questions

Tutorial 14E.10

14E.10.1

What is the difference between pulses produced by a monostable and an astable?

14E.10.2

What is the behaviour you would expect from a synchronous falling edge triggered flip-flop?

14E.10.3

What is the problem with there being an indeterminate state?

14E.10.4

What is meant by a circuit being transparent?

14E.10.5

Look at *Figure 154*. What do you notice about the output compared to the input?

14E.10.6

Look at *Figure 156*. What do you notice about the output trace B?

14E.10.7

Look at *Figure 158*. When does the output trace change state?

14E.10.8

Look at *Figure 159* and *Figure160*. What is the maximum decimal number that this counter can count to?

14E.10.9

Look at *Figure 166*. Why is the circuit a Modulo-12 counter?

Tutorial 14 E.11 Monostables and Astables	
AQA Syllabus	
Contents	

14E.111 Monostables

Monostables are NOT on the AQA syllabus. It is worth mentioning them to provide some information as a context for astables. You will come across them in university level electronics. If you want to miss out this section, go to 14E.112.

A monostable gives a pulse that lasts for a fixed period of time. It does it once and once only. We will look at monostables that are based on the 555-timer and the NAND gate. You are NOT expected to recall details of these circuits, but it's worth being able to recognise them.

<u>555-Timer Monostable</u>

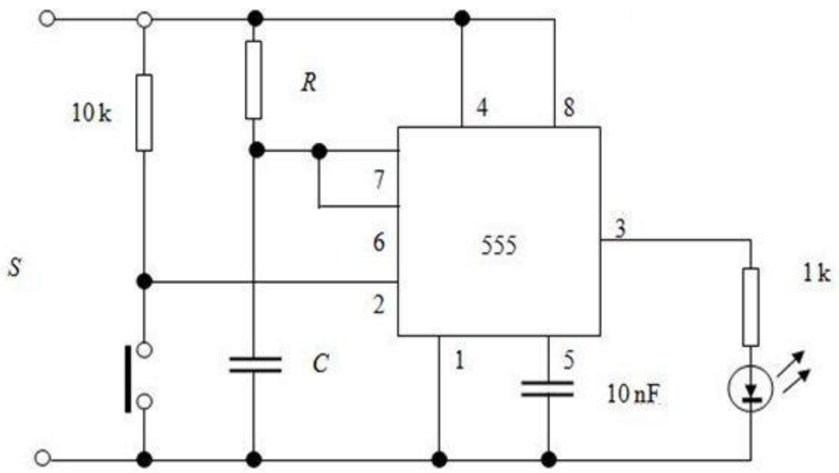


Figure 167 Monostable based on a 555-timer

The circuit will give a single output pulse like this (*Figure 168*):

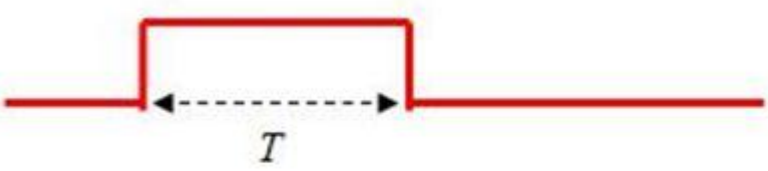


Figure 168 Output of a monostable

The components R and C determine the time period T of the output pulse.

- When the push switch *S* is closed and released, the voltage at pin 2 goes from high to low to high again.
- This triggers the output to go to high.
- When the voltage across *C* gets to about 2/3 of the supply voltage, the output goes low.
- The period of the pulse is given by a simple relationship:

$$T \approx 1.1\ RC$$ ………… *Equation 85*

- Once triggered, the circuit cannot be re-triggered to extend the period T.

Monostable Circuits based on NAND gates

Here is a monostable circuit based on a NAND gate (*Figure 169*):

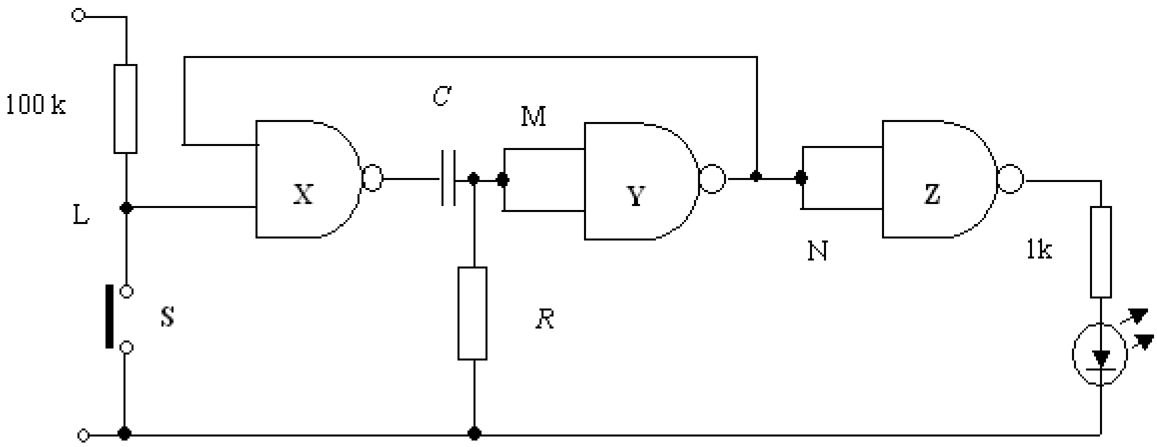


Figure 169 NAND gate monostable

How does the circuit work? At first the circuit is stable.

- When switch **S** is open, the point L is at 1.
- The capacitor C is not charged and the voltage at M is 0.
- Therefore, the output of **Y** is 1, and both inputs of **X** are 1.
- The output of **X** is 0.
- Since the inputs of **Z** are 1, and the output of **Z** is 0.

Now let us momentarily close the switch **S**.

- T becomes 0.
- Therefore, the input L into gate **X** becomes 0.
- The output of **X** becomes 1.
- The flow of charge onto the capacitor *C* causes there to be a current through the resistor *R*. This makes the voltage at M high.
- Therefore, gate **Y** gives out a 0.
- Since the voltage at N is a zero, the output of gate **Z** is 1, and the LED lights up.

This lasts a short time, determined by the RC time constant.

- The current through *R* decays exponentially as the capacitor charges up. So the voltage at M decays exponentially as well.
- As the voltage at M drops, it passes below the threshold at which **Y** is triggered to change state.
- Gate **Y** gives out a 0 until the voltage passes below the threshold.
- Now gate **Y** changes to 1.
- Since L is at 1, this makes the output of **X** low again.

If the switch is held closed, the output of **X** remains 1, but no current flows onto the plates of the capacitor, so the voltage at M remains low. Therefore, holding the switch has no effect on the behaviour of the circuit. We can show the behaviour of the circuit with timing diagrams for each of the points (*Figure 170*).

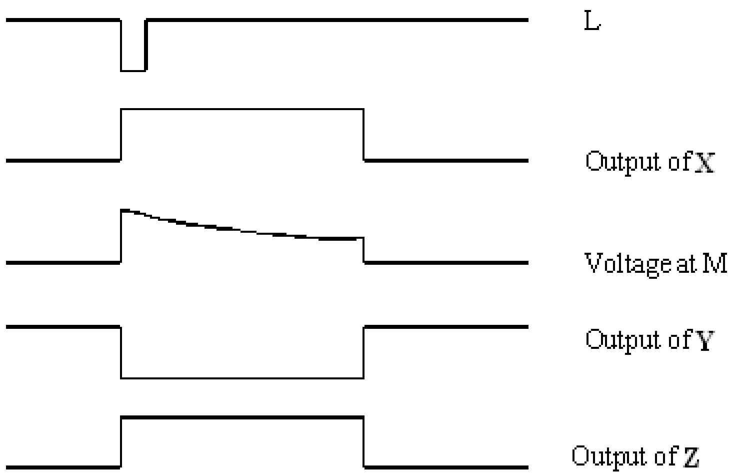


Figure 170 Timing diagram for a NAND gate monostable

Depending on the threshold at which the gate triggers, it can be shown that the time period T at for which the output of **Z** is high is approximately RC.

$$T \approx RC$$ ………. *Equation 86*

If the gate triggers at $0.5\ V_s$, then T **is approximately** $0.7\ RC$.

$$T \approx 0.7\ RC$$ ………… *Equation 87*

14E.112 - Astables

Monostable circuits give out a single pulse. These are not on the syllabus, but if you want to find out more about them, look at **14E.111**.

Astable circuits do not have a stable state. They are constantly changing from 0 to 1 and back. They emit a train of pulses which are high (1) or low (0). These pulses can be square waves or can be configured to be rectangular.

A common device to produce these pulses is the 555-timer. You are NOT expected to describe the 555 timer circuit, but it's worth being able to recognise the circuit.

555-timer in Astable Mode

The 555 timer can be wired up to produce a train of pulses by ensuring that the circuit is **astable**, which means that it is not in a stable state. We can make astable circuits from other components, but the 555 timer gives a train of digital pulses. The diagram shows the circuit (*Figure 171*).

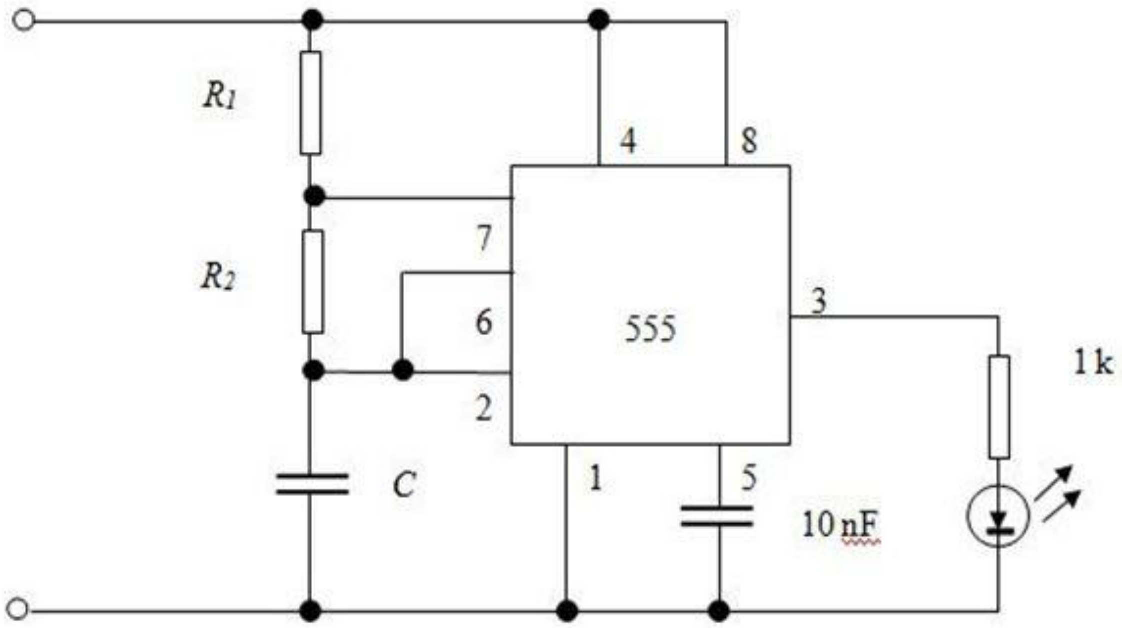


Figure 171 555-timer in astable mode

The output of the circuit is a **square wave**, as shown (*Figure 172*).

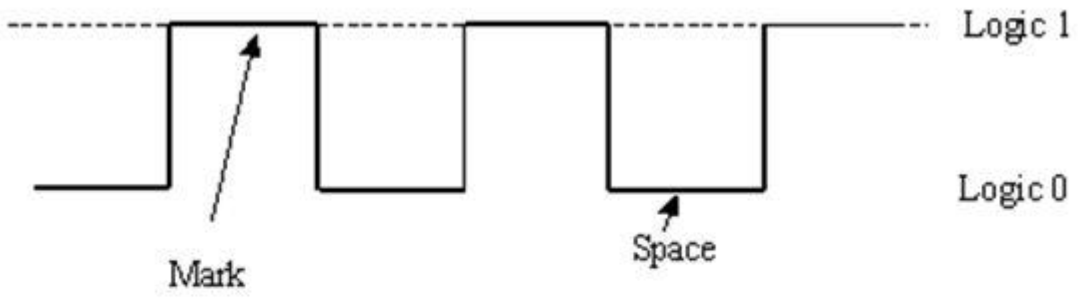


Figure 172 Output of a 555-timer astable

We need to consider some definitions:

- The **mark time** [$t_{(H)}$] is the time at which the output is a 1.

$$t_H = 0.7(R_A + R_B)C \text{ .... Equation 88}$$

- The **space time** [$t_{(L)}$] is the time at which the output is a 0.

$$t_L = 0.7\, R_B C \text{ .... Equation 89}$$

- The **mark to space ratio** = mark time ÷ space time.

$$\text{Ratio} = t_H \div t_L \text{ .... Equation 90}$$

- The **astable period** T is the time taken for one complete cycle, the mark and the space times added together.

$$T = \text{mark} + \text{space} = t_L + t_H \text{ .... Equation 91}$$

- The **frequency** = 1 ÷ period.

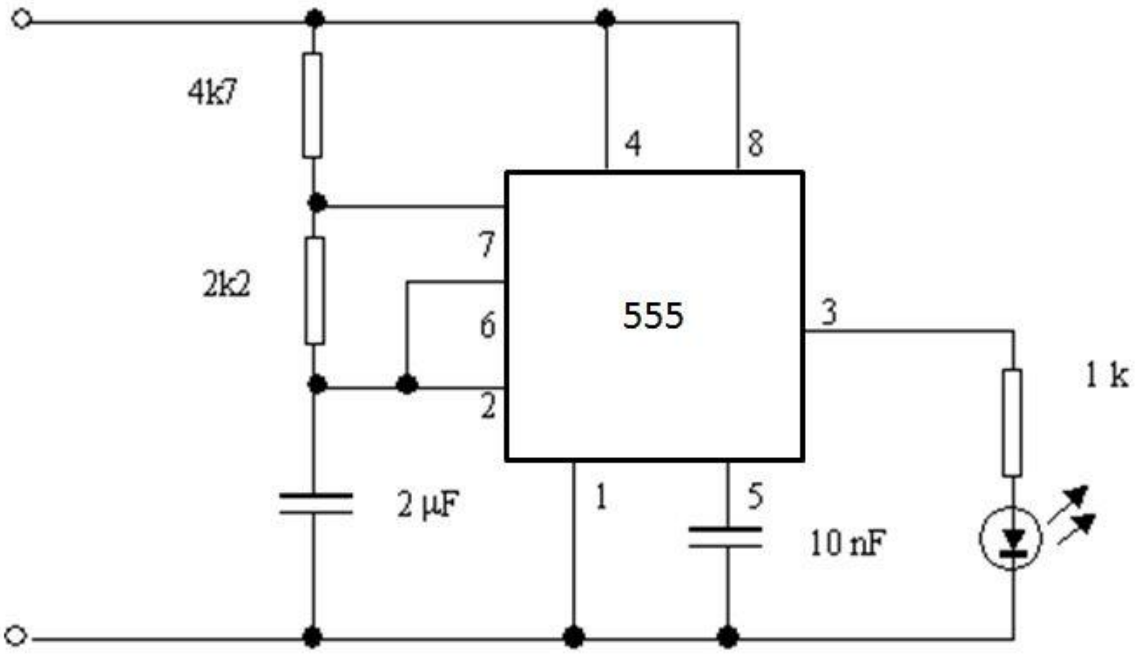
$$f = \frac{1.4}{(R_1 + 2R_2)C}$$ .... *Equation 92*

The time t_H will be longer than t_L, unless R_1 is very small compared to R_2. If this is the case, then t_H will be approximately equal to t_L, but not quite equal. We can say to a first approximation that the mark to space ratio is 1. This will result in a square wave output.

The **duty cycle** = mark time ÷ period.

Worked Example

What is the frequency of the square wave output from the circuit?



Use the formula

$$f = \frac{1.4}{(R_1 + 2R_2)C}$$

Answer

From the diagram we can see that:

- $R_1 = 4700\ \Omega$
- $R_2 = 2200\ \Omega$
- $C = 2.0 \times 10^{-6}$ F

We need to substitute these values into the formula:

$$f = \frac{1.4}{(R_1 + 2R_2)C} = \frac{1.4}{[4700\ \Omega + (2 \times 2200\ \Omega)] \times 2.0 \times 10^{-6}\ \text{F}}$$

$$= \frac{1.4}{9100 \times 2.0 \times 10^{-6}} = \mathbf{77\ Hz}$$

This is a 555-timer in astable mode. It is NOT a tidy looking circuit (*Figure 173*):

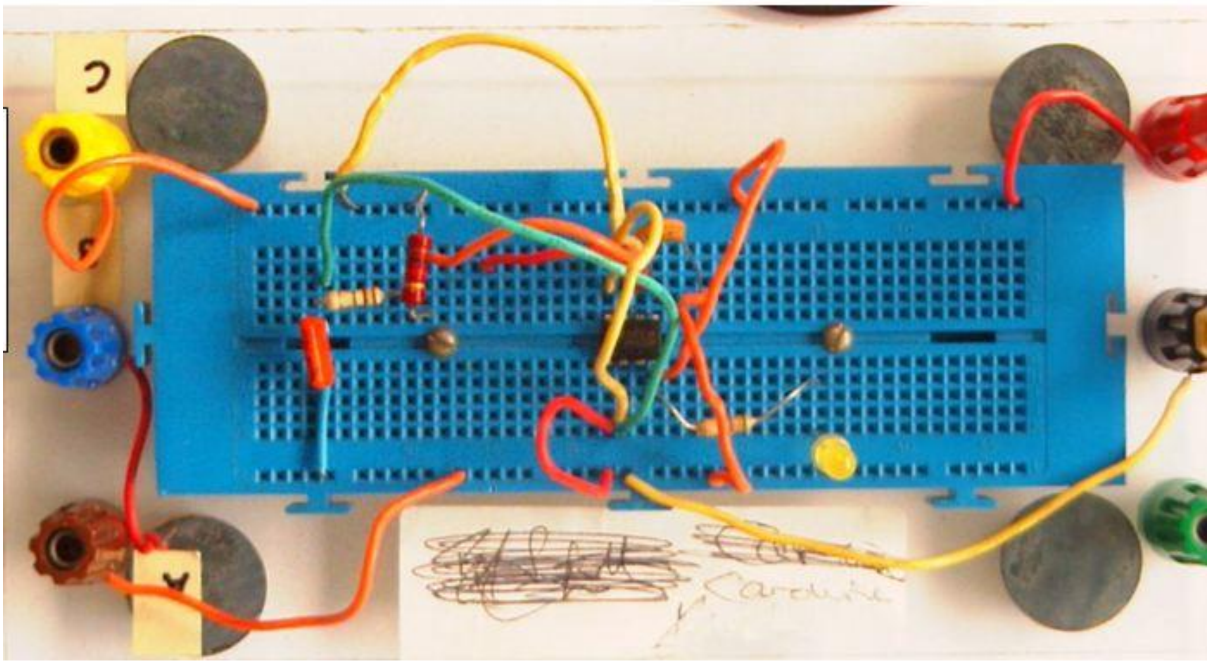


Figure 173 555-timer as an astable

But it does give out a good rectangular wave (*Figure 174*):

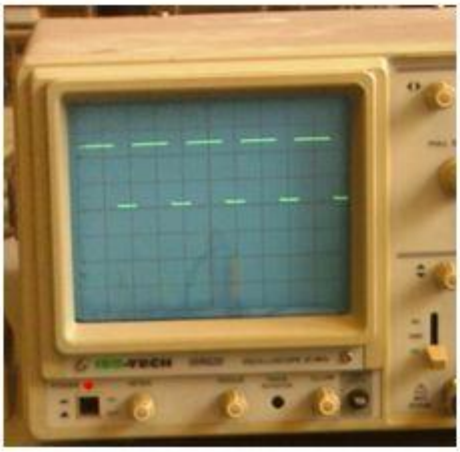


Figure 174 Square wave output of the circuit in Figure 173

The mark time is longer than the space time.

<u>NAND gate Astable</u>

We can make an astable circuit the output of which oscillates at a frequency determined by the value of the time constant of a capacitor and a resistor (*Figure 175*).

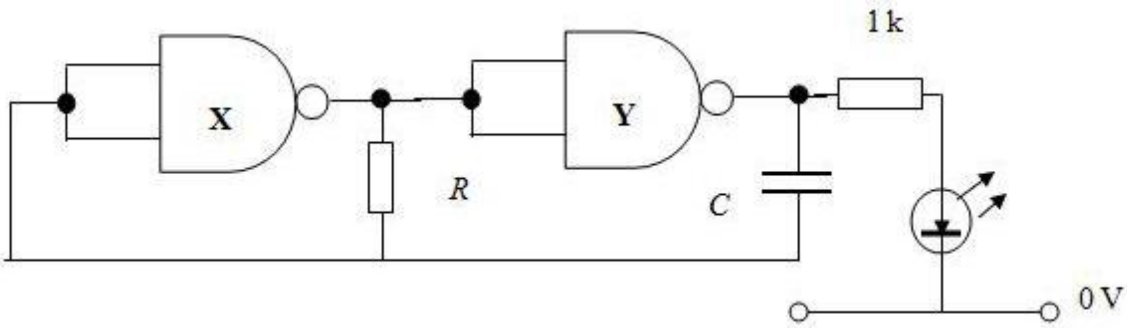


Figure 175 NAND gate astable

If you look carefully at the arrangements of the NAND gates, it does not take a genius to see that the two NAND gates are wired as NOT gates, so this set up is also called a **NOT gate astable**.

Let's have a look at how the circuit works:

- Suppose the output of **Y** is high.
- This means that the input to **Y** is low.
- The capacitor will charge up.
- A current flows through the resistor *R* which means there will be a voltage across it.
- This raises the input to **Y** to high and it will trigger to the output being low.
- Since **X** is connected to the feedback loop, its output will be low.
- The low output of **X** will cause the capacitor to discharge.
- This makes the input to **Y** low, hence the output to go to high.
- And so on...

We can summarise this in the timing diagram (*Figure 176*):

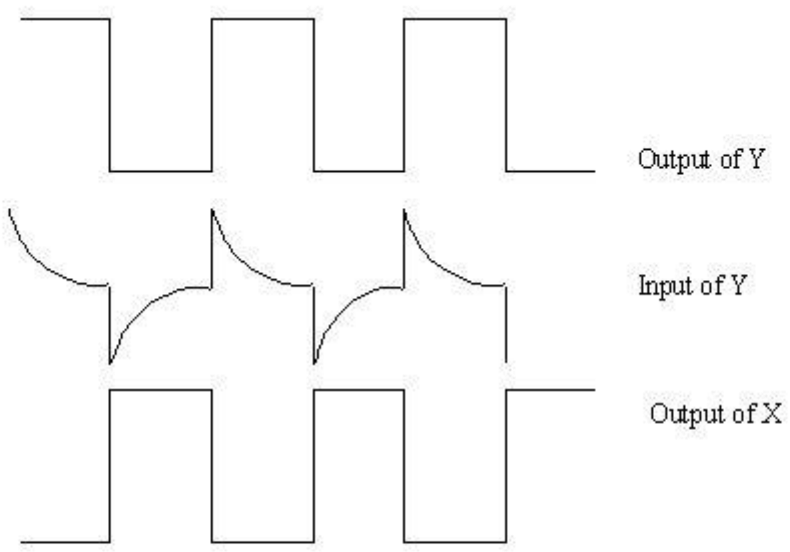


Figure 176 Timing diagram of NAND gate astable

We can show that the **mark time** is given by the relationship:

$$t_{\mathrm{H}} = 1.1\ RC$$ ……….. *Equation 93*

Similarly, the **space time** is given by:

$$t_{\mathrm{L}} \approx 1.1\ RC$$ ………… *Equation 94*

Therefore, the period:

$$T = t_H + t_L = 2.2\ RC$$ .... *Equation 95*

So, the frequency:

$$f = 1/T = 1/2.2\ RC$$ .... *Equation 96*

Questions

Tutorial 14E.11

14E.11.1

An astable 555 timer has the following external component values: R_1 = 100 kilohms; R_2 = 47 kilohms; C_1 = 10 microfarads.

What is the mark and space?

14E.11.2

An astable has a 2.2 μF capacitor, R_1 value of 10 k, and R_2 value 20 k.
(a) Calculate the astable frequency.
(b) Work out the mark time, the space time, and the mark to space ratio.
(c) Is the output a square wave?

14E.11.3

A NAND gate astable has a capacitor of capacitance 20 μF with a resistor of resistance 150 kΩ. What is the period of the astable?

Tutorial 14 E.12 Principles of Communication and Transmission Media	
AQA Syllabus	
Contents	
14E.121 Practical Telecommunication	14E.122 General Principles
14E.123 Transmission Media	14E.124 Electrical Wires
14E.125 Optical Fibres	14E.126 The Decibel Scale
14E.127 Radio Transmission	14E.128 Satellite Transmission
14E.129 Satellite Communication Frequencies	14E.1210 Advantages and Disadvantages of Transmission Media

14E.121 Practical Telecommunication

Practical **telecommunication** systems have been around for about two hundred years. The earliest **electromechanical telegraph** was invented in the UK by Francis Ronalds (1788 - 1873). It was William Cooke (1806 - 1879) and Charles Wheatstone (1802 - 1875) who put the electric telegraph into commercial use. The electric telegraph and code were further developed by **Samuel Morse** (1791 - 1872) in America. The code is still recognised internationally. Aircraft navigation beacons transmit their identifiers in Morse Code. Morse himself was also an artist, with a good number of portraits and other artworks to his name. He also invented a marble-cutting machine that could cut in three dimensions. One could argue that Morse code was an early digital transmission medium.

Telephony arrived in the late nineteenth century when of capturing sound wave and converting them to electrical waves were applied. The invention of the telephone was generally credited to Alexander Bell (1847 - 1922), although others have claimed its invention as their own.

Wireless telegraphy started to be developed after the work of James Clerk-Maxwell (1831 - 1879) and Heinrich Hertz (1857 - 1894). The idea was developed commercially by Guglielmo Marconi (1874 - 1937) who was the first to transmit a transatlantic message in 1901. In the early Twentieth Century the development of speech radio spread widely, and we are going to study the principles of radio and other communication technology.

14E.122 General Principles

Most modern communication systems follow a general schematic like this (*Figure 177*):

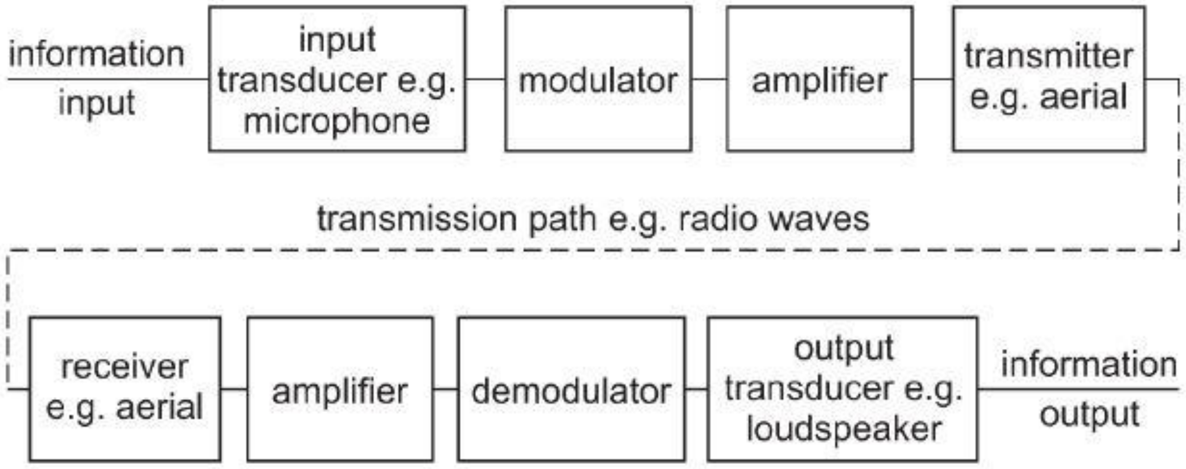


Figure 177 Schematic of data transmission (Credit AQA syllabus)

- The **input transducer** converts an input signal, like the sound waves of speech, into electrical waves.
- The **modulator** mixes the signals from the microphone with a **carrier wave** which **oscillates** at a particular frequency.
- The carrier wave with the signal is then **amplified**.
- The amplified signal is then sent to the **transmitter**, which acts as the primary of a transformer.
- The radio waves are **electromagnetic waves** that **propagate** spherically from the transmitter. In most text books, you will see the wavefronts propagating radially from the point source
- The aerial on the receiver acts as a **secondary** to a very inefficient transformer. The voltage is very small, but the phenomenon of **electrical resonance** boosts it. Electrical resonance also selects the right signal. You change the **capacitance** of a **variable capacitor** when you tune into a radio station. You change the **inductance** of an **inductor** when you change waveband.
- The signal is further **amplified**.
- The **demodulator** separates the broadcast information from the carrier wave.
- The **analogue** signal is further boosted by an amplifier. This may be separate from the **tuner section** or contained within the **receiver**.
- The signal then passes to a **loudspeaker** so that we can hear the output.

Important Note
If you are intending to do experiments with radio apparatus, you MUST have a licence. It is illegal to do otherwise, and you could end up getting a hefty fine, especially if your apparatus causes interference with other radio users.

The picture below shows a radio tuner that is separate from the audio amplifier (*Figure 178*):

Figure 178 A hi-fi radio tuner (bottom left)

In the Exam

Although these numerical calculations have not been explicitly shown in the syllabus for the Electronics option, the relationships are elsewhere on the Physics syllabus. It is quite likely that you would be asked to do similar calculations in the exam, which is why I have included these examples.

You also may get new situations in which you need to apply the physics that you know.

14E.123 Transmission Media

Radio wave are **electromagnetic waves**. They travel at 3.0×10^8 m s^{-1} in a vacuum, and very close to that speed in air. The wave equation applies:

$$c = f\lambda$$

.... *Equation 97*

All the properties of electromagnetic waves are true, and they reflect and refract in the same way as light does. The picture shows a receiver disc for signals from a relay satellite (*Figure 179*):

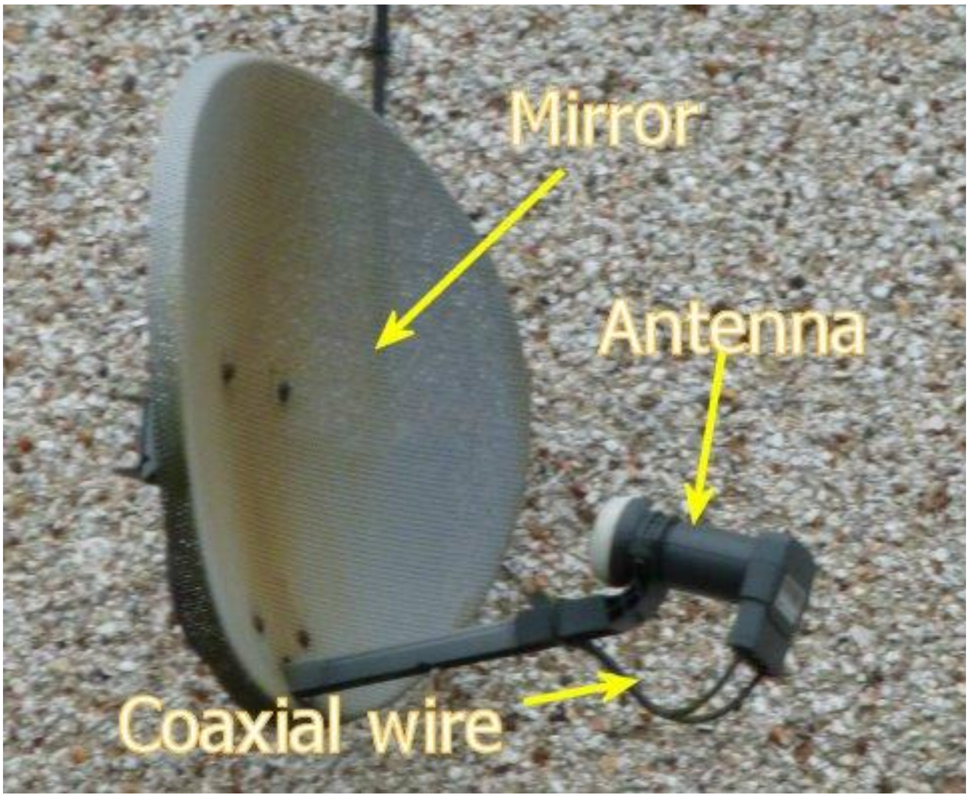


Figure 179 Satellite dish mounted on a house wall

Notice that the reflector is not solid but is made of a perforated material. This does not affect its performance as a mirror at all.

Radio waves, like all electromagnetic waves are attenuated according to the inverse square law. You have looked at the inverse square law for gamma rays and saw the equation:

$$\frac{I_1}{I_2} = \left(\frac{x_2}{x_1}\right)^2$$

.... *Equation 98*

The intensity is the power per unit area. The equation assumes that the source is a point source which is infinitely small. The distance x_1 is the radius of the aerial, which is a distance of x_1 metres from the point source. The distance x_2 is the distance between the transmitter and the receiver. The idea is shown in the diagram (*Figure 180*).

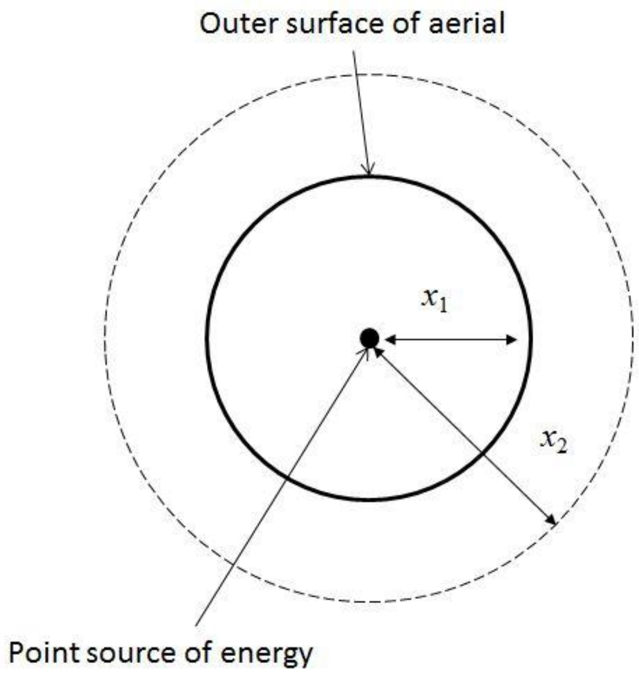


Figure 180 Distance between the aerial and the point source, and the point source and the receiver

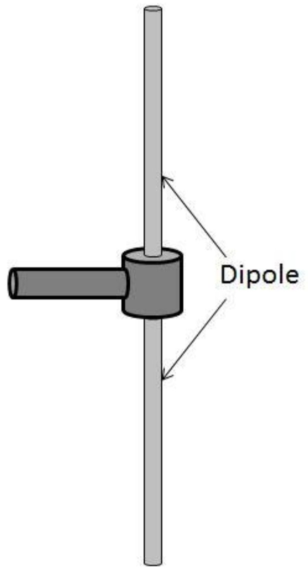


Figure 181 A dipole aerial

The right hand picture (*Figure 181*) shows a **dipole aerial** that is commonly used for radio transmission. The total length of the aerial is **half a wavelength**, as an **electrical standing wave** is set up in the dipole. The whole thing is mounted on a bracket that is bolted to the wall of a building.

From the answer to 14E.12.3 you can see that the intensity of the radiation is very low indeed. Radio transmission feeds very little energy per second to the receivers. However, you can have thousands of receivers and the total energy per second they pick up is a small fraction of that from the source.

14E.124 Electrical Wires

Radio waves seem to be a remarkably inefficient and wasteful way of getting information about. Most of the energy used to do so is dissipated.

Data can be transmitted using **electric wires**. This may seem a logical way to connect a radio station with a remote transmitter, but wires present their own problems. When we do circuits, we tend to regard the wires as perfect conductors. They are not. They have resistance, according to the resistivity equation:

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

.... *Equation 99*

Electric currents **always** have magnetic fields associated with them. With continuously flowing direct current, there is little effect. However, with alternating currents (and direct currents that flow in pulses) there is a reverse emf due to Faraday's Law and Lenz's Law:

$$\mathcal{E} = -N\frac{\Delta \Phi}{\Delta t}$$

.... *Equation 100*

You can see the effect of this if you have an **unshielded** wire connected to a CRO in the physics lab. You get a fuzzy wave along the screen which has a period of 0.02 s, even if there is no signal on the wire. While this may seem a curiosity, it has important implications in the design of audio amplifiers. The wave will end up being amplified as **mains hum** which can be intrusive. Other spurious signals end up as **noise**. Unshielded wires are often called **twisted pairs**.

Figure 182 A mess of mains wires

The mess of wires here (*Figure 182*) is very likely to pick up mains hum and other noise.

Often there are cables made of hundreds of individual conductors. It is possible for the wires in such **multi-core** cables to pick up interference from signals in other conductors (*Figure 183*).

Figure 183 Cable made of many individual conductors

The wire also acts as an **inductor**. The higher the frequency, the higher the **reactance**.

There is a way of preventing these problems. The wire can be shielded easily; we simply surround the wire with an insulating layer and then copper braiding to shield the wire. We have a **coaxial** cable. Voltages induced in the braiding by changing magnetic fields are connected to earth (*Figure 184*).

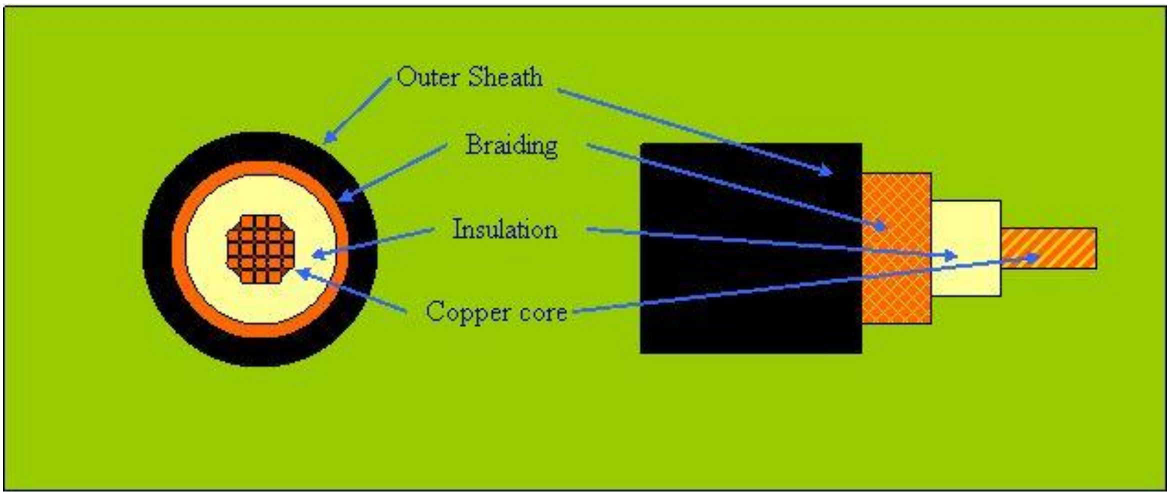


Figure 184 Coaxial cable

The problem with coaxial cable is that:

- It is heavy.
- It is bulky. Lots of coaxial cables take up a lot of space.

There is another problem. The copper core is separated from the braiding by an insulating layer. The coaxial therefore makes a perfectly good **cylindrical capacitor**. Its capacitance is not very high, but it can have a significant effect at high frequencies. The capacitance per unit length varies from 10 to 100 pF per metre.

If an alternating signal is being conducted, there will be a reactance from the capacitor given by:

$$X_C = \frac{1}{2\pi f C}$$

.... *Equation 101*

The higher the frequency, the lower the reactance.

There is a number of equations that deal with the capacitance of the coaxial cable. These are NOT on the syllabus.

Another problem with a wire is that while a direct current is carried evenly across the conductor, this is not the case with alternating current. High frequency alternating currents are carried on the outer layers of the conductor. This is called the **skin effect**.

14E.125 Optical Fibres

The most up-to-date data transmission systems use **optical fibres**. These can carry large amounts of data with a very thin fibre. They use the principle of **total internal reflection**, which you have studied in the section on waves. See *Figure 185*.

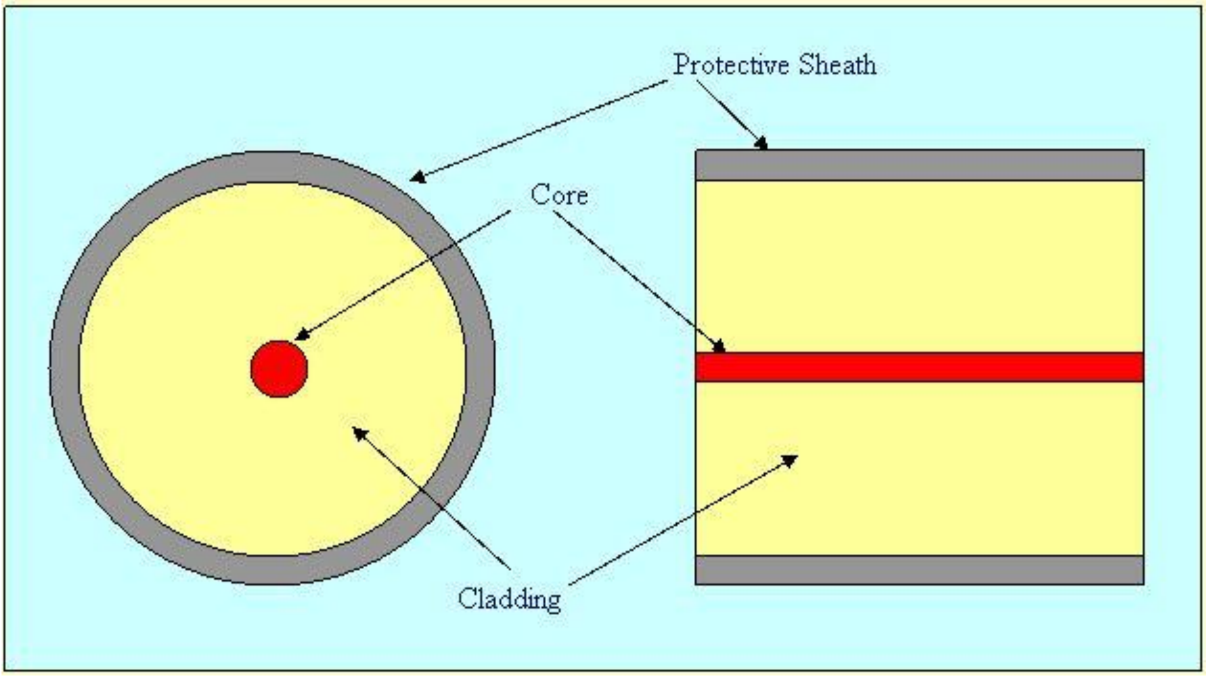


Figure 185 Optical fibre

The light ray makes a certain angle of incidence when it hits the boundary of an optically dense material (like glass) and an optically less dense material (like air). If this angle is greater than the critical angle, the ray is totally internally reflected. The critical angle, q_c, is determined by the formula:

$$n = \frac{1}{\sin \theta_c}$$ .... *Equation 102*

The rays of light should travel like this (*Figure 186*):

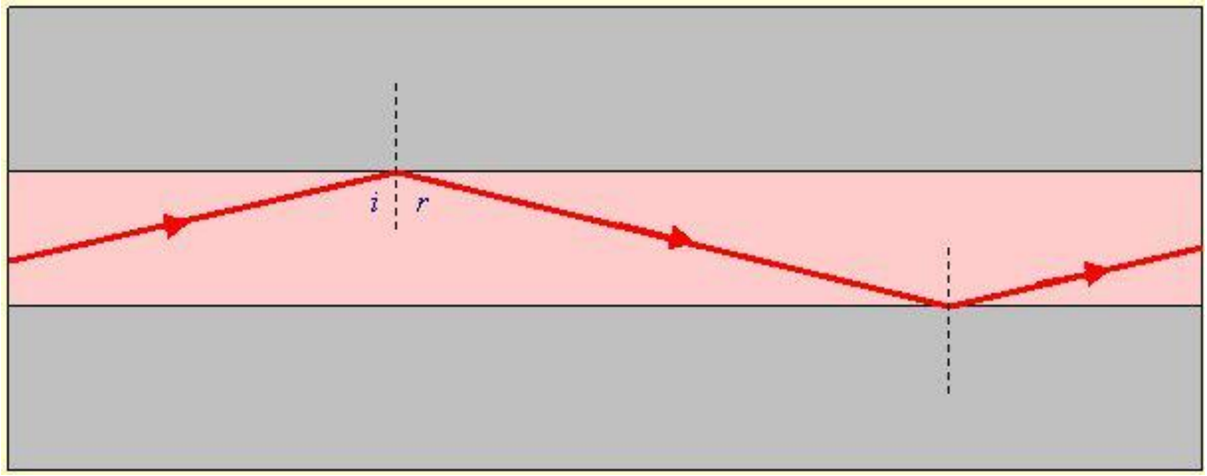


Figure 186 Light travelling in an ideal path

But instead, light rays can travel several paths (*Figure 187*):

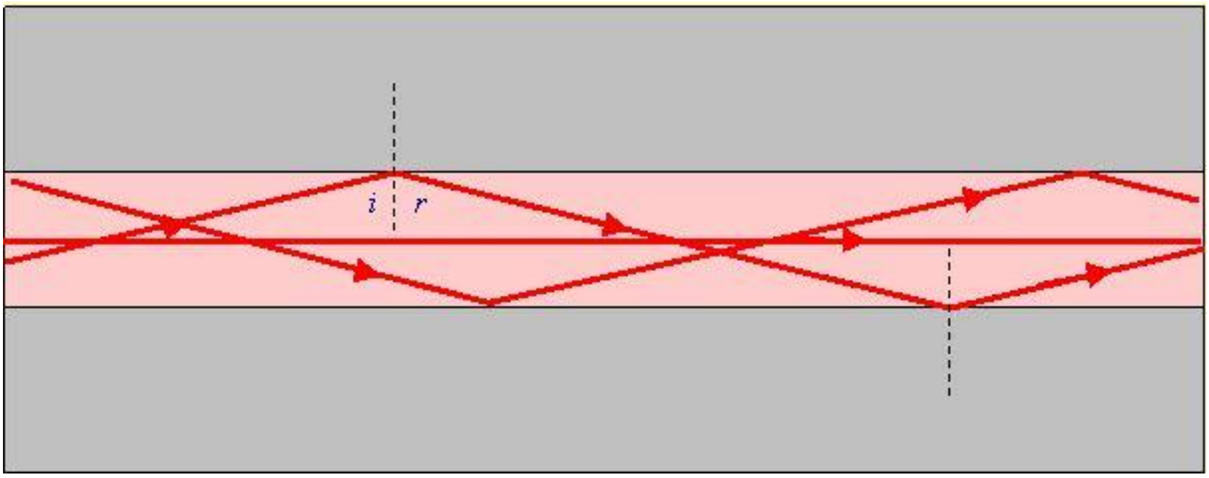


Figure 187 In reality, light travels on many different paths

This means that the light rays can arrive at different times, resulting in **dispersion** or **smearing**. The signal that was sharp when it left the transmitter is smeared (*Figure 188*).

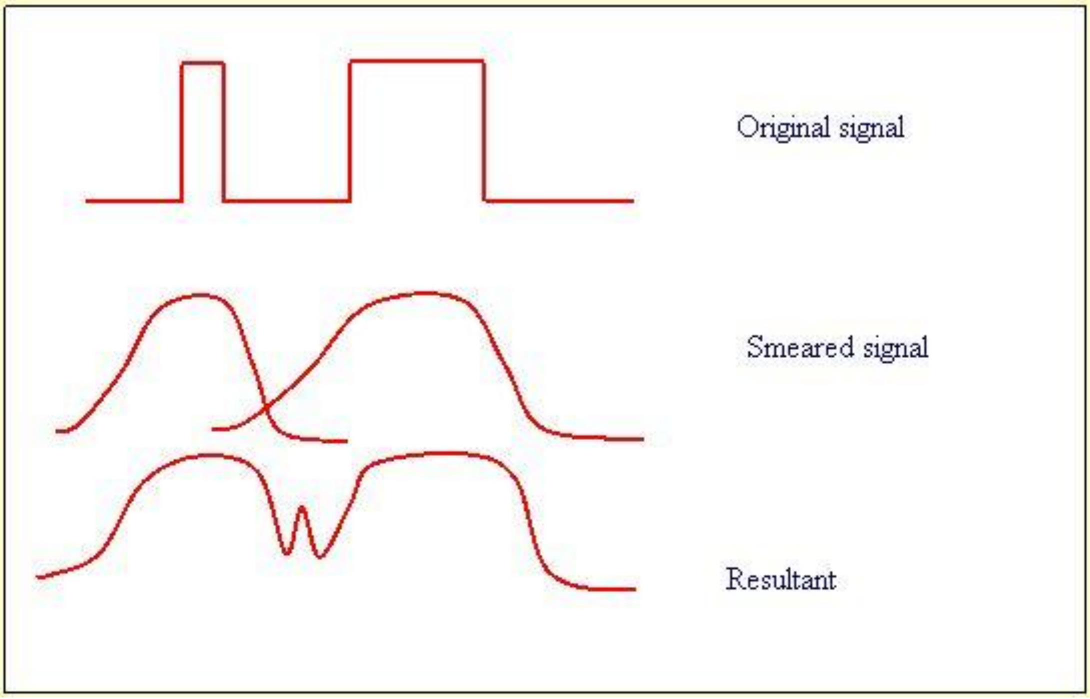


Figure 188 Smearing

The picture shows us how the signal can be unacceptably distorted and even produce **spurious** signals that were not there.

The problem can be resolved by **cladding** the core with a material of slightly lower refractive index. For example, the core might have a refractive index of 1.6, while the cladding has a refractive index of 1.4.

Dispersion can be reduced further by use of a **graded index** or **multimode** fibre. Some light is passes down the middle, which has a higher refractive index, therefore slower rate of travel. With clever manipulation of the refractive indices, the ray travelling down the middle can be made to arrive at the same time as the ray that goes from side to side. They can meet with a time difference of about 1 ns km^{-1}.

Monomode fibres are designed such that the rays pass only down the middle. If the light were perfectly **monochromatic**, i.e. of one wavelength only, the rays would all arrive at the same time. However, even the best lasers produce a slight spread, and since refractive index varies with wavelength, there can be slight differences in arrival times, leading to smearing.

Optical fibres experience **attenuation**.

As light travels down an optical fibre, it loses **intensity**. This attenuation is caused by very slight impurities that you get even in the purest of glass. Also there will be defects and anomalies in the crystal structure caused by the manufacturing process. Light may even "leak out" of the fibre. Whatever the cause the light at the receiver will be dimmer than the light at the transmitter.

If the light intensity travels through 1 km of optical fibre, and its intensity is reduced to 50 % of its original, we can expect that after 2 km, then intensity is 25 % of the original, and after 3 km, it's 12.5 % (1/8) of the original. The change is **exponential**.

So the relationship between the intensity and the distance is going to be governed by an exponential function:

$$I = I_0 e^{-\mu x}$$

.... *Equation 103*

Where:

- I is the intensity (W m^{-2})
- I_0 is the intensity of the source (W m^{-2})
- m is the attenuation coefficient (m^{-1})
- x is the length of the optical fibre (m)
- e is the exponential number 2.718...

If we have 1000 metres of optical fibre with an attenuation coefficient of 0.002 m^{-1}, the graph is like this (*Figure 189*):

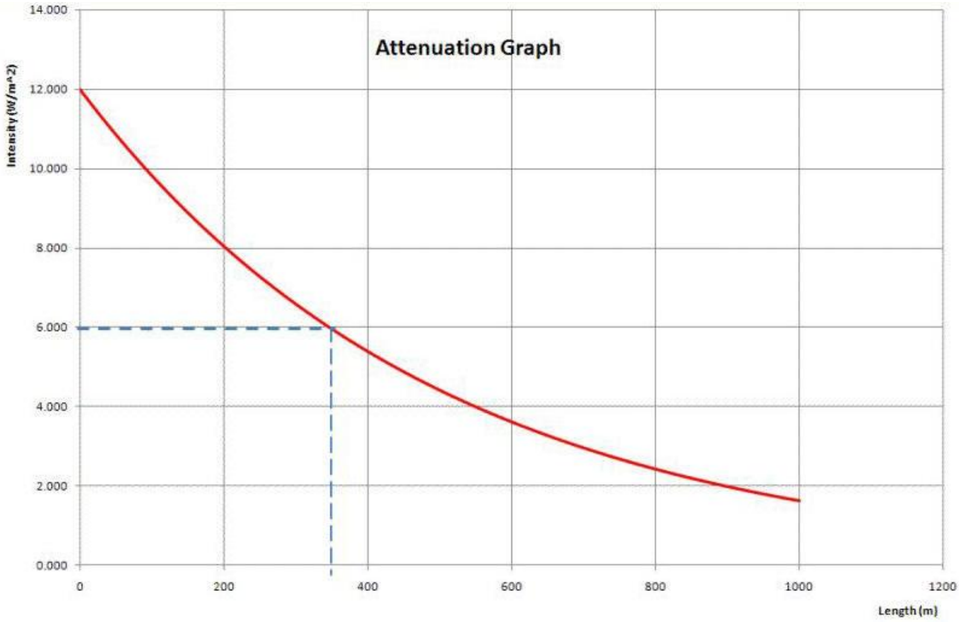


Figure 189 Attenuation graph for an optical fibre

We can see that the intensity has dropped to 50 % of its value after 350 m.

If we plot the **natural logarithm** of the intensity against the distance, we get a straight line:

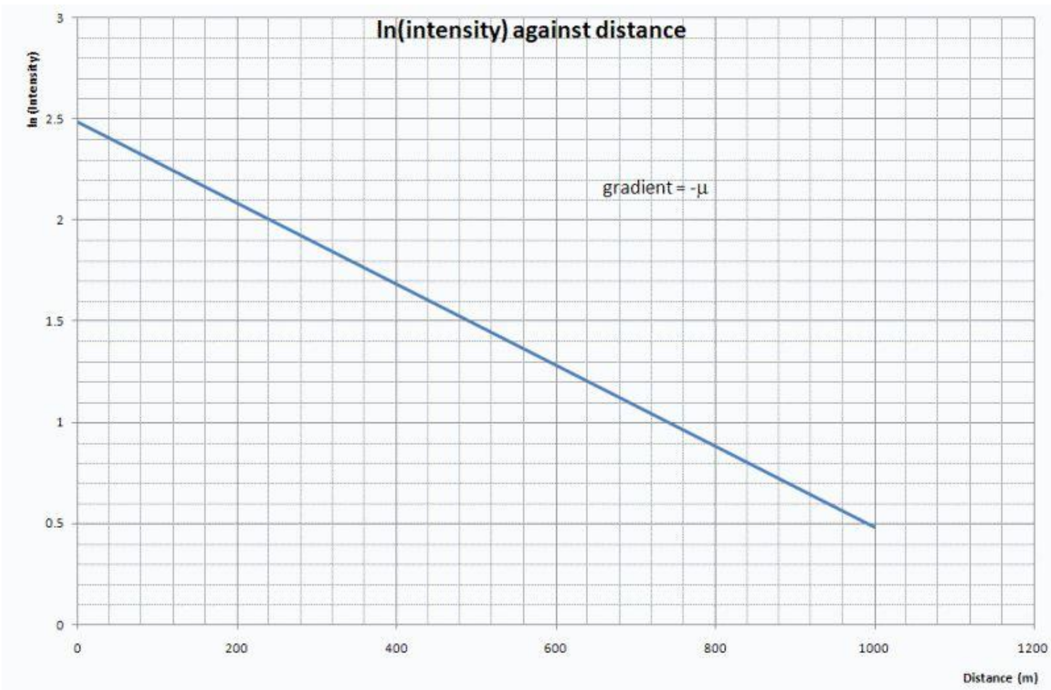


Figure 190 Semi-log plot for an attenuation graph

We would get a similar shape if we used logs to the base 10.

Typically attenuation coefficients are in the order of 10^{-5} m^{-1}. For long distance communication, it can be significant. Also, the attenuation coefficient is different for different wavelengths in the same optical fibre.

14E.126 The Decibel Scale

In communication systems, the sound intensity that we hear is result of sound energy from loudspeakers. The rate at which this energy is given out is the power. The human ear can just detect a doubling in power. See Topic 14B to see how the ear behaves. We measure the intensities of sounds, represented by **powers** using logarithmic scales to compress the graph into something more manageable. In the decibel scale we use a reference power, P_1, a reference point to which we compare a second power P_2. We can write an expression for the change in power ΔP.

1 Bel is the power change from 10^{-6} to 10^{-4} W. The Bel (B) is rather a big unit, and we use the **decibel** (dB) instead.

$$\Delta P = 10 \log_{10} \frac{P_2}{P_1}$$

.... *Equation 104*

A **doubling** of power gives a **3 dB** increase. The 3 dB increase is because if we double the power from 10 W to 20 W:

$$\Delta P = 10 \log_{10} (20 \text{ W} \div 10 \text{ W})$$

$$\Delta P = 10 \log_{10} (2) = 10 \times 0.3010 = 3.01 \text{ dB}$$

We often describe an amplifier as having a **gain** of *X* decibels. This means that the amplifier boosts signals by *X* decibels.

Worked example
A signal of 25 mW is boosted by 18 dB using a booster amplifier. What is the output power of the amplifier?
Answer ΔP = 18 dB = 10 $\lg_{10}$ (P_2 ÷ 25 × 10^{-3} W) $1.8 = \lg_{10}$ (P_2 ÷ 25 × 10^{-3} W) $10^{1.8}$ = P_2 ÷ 25 × 10^{-3} W P_2 = $10^{1.8}$ × 25 × 10^{-3} W = 63.1 × 25 × 10^{-3} W = 1.58 W = **1.6 W (2 s.f.)**

The **dBA scale** is used to take into account the frequency dependence. Remember that the ear is most sensitive at frequencies between 100 to 10 000 Hz. The table shows the levels of certain noises:

Level (dBA)	***Noise***	***Effect***
0	Threshold of hearing	
20	Blood pulsing	
30	ticking watch	
40	Quiet conversation	
50	Quiet street	
70	Hoover in a room	
90	Road drill at 7 m	Prolonged exposure can lead to hearing damage
100	Noisy factory	
120	Loud discothèque	Threshold of discomfort
140	Aircraft at 25 m	Threshold of pain
160	Rifle close to ear	Ear drum ruptured

14E.127 Radio Transmission

Most radio transmissions travel directly from the transmitter to the receiver. This is known as **line-of sight transmission**.

Often the transmitter may not be in direct line of sight, but the waves can diffract around objects in the way and diffract around the curvature of the Earth. The waves are known as **ground waves** or **surface waves**.

While the ground is opaque to light waves, it can transmit radio waves to a limited extent, especially those of a frequency lower than 3 MHz. There is a value for refractive index for the ground, while the refractive index for air for radio waves is very close to 1.0. Therefore, the waves can be guided along the **interface** between the two media.

This has nothing to do with total internal reflection.

Long distance transmission of **short wave** radio waves uses sky-waves which are reflected off the ionosphere. Short wave radio waves have a frequency between 1.6 to 30 MHz (187 m down to 10 m wavelength). See *Figure 191*.

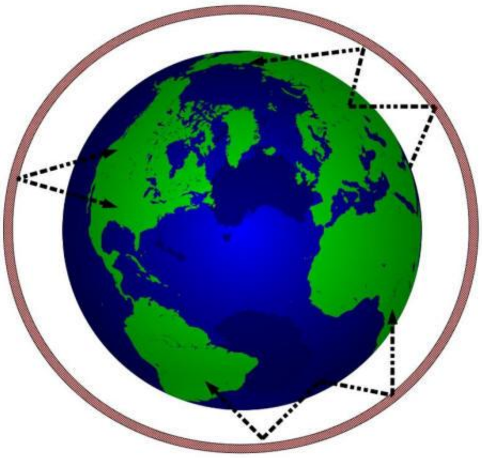


Figure 191 Reflection of radio waves by the ionosphere (Image from Wikimedia Commons. Authors: Kf4yfd, Noldoaran, Augiasstallputzer)

The **ionosphere** is a layer of the upper atmosphere from about 80 km to 1000 km above the Earth's surface. Neutral air molecules are ionised by cosmic rays and high energy photons from the Sun. The physics process is called **back-scattering**. If the ionisation is strong enough, the radio waves are reflected back to the ground. When they strike the ground, they are diffusely reflected back to the ionosphere. So, the waves bounce off the ionosphere to the ground, enabling the waves to be received much further than would be expected. Transmissions of just a few watts can be picked up 3500 km away.

The ionosphere is not a stable layer. It undulates like a sea, leading to **periodic variations in signal strength**, which can be observed in the phenomenon of **fading**. Fading can make broadcasts unpleasant to listen to, if not unintelligible. (Ask your parents or grandparents who listened to *Radio Luxemburg* during the evening on 208 m of the Medium Wave Band when they were teens.) Fading can be explained by considering the behaviour of **standing waves.** Similar fading can be heard when listening to FM radio while driving through towns. Waves reflected by buildings destructively interfere with the incoming radio wave. The resultant wave has a much lower amplitude.

For waves below 10 MHz, sky-wave propagation is most efficient at night. For frequencies above 10 MHz, the propagation is most efficient by day.

This process does not happen with high frequency waves (100 MHz), except in very unusual atmospheric conditions.

14E.128 Satellite Transmission

Communication satellites are widely used in broadcasting nowadays, enabling **live events** to be broadcast throughout the world in **real time**. Before the satellites were invented, events were recorded on film, then flown around the world. A process that take days then is now done in a fraction of a second. The satellites have a receiver which they pick up waves from a ground station. Then they retransmit the signals back to the ground, where they are picked up using a satellite dish like the one above. The idea is like this (*Figure 192*):

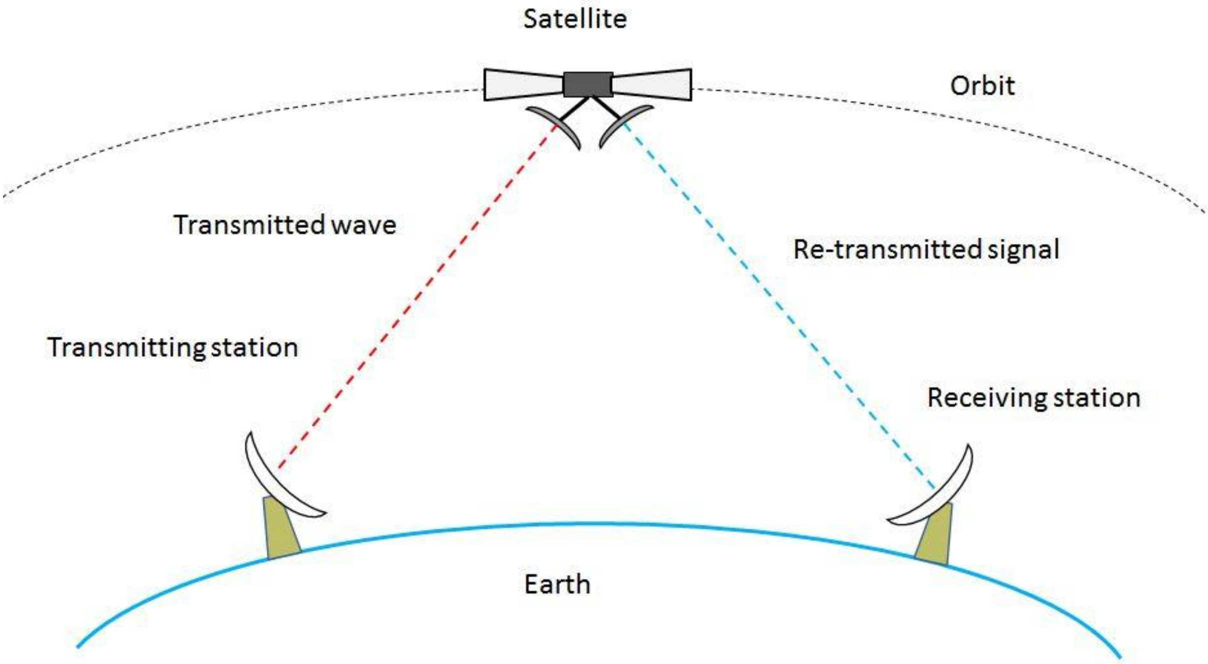


Figure 192 Using satellites to transmit data

The physics of satellite orbits is covered in the section of work on gravity fields. **Kepler III** is the rule that links orbital height and periodicity. There are three types of orbit:

- **Low earth orbit** - Height: 160 - 2000 km. It is less expensive to send a satellite into low earth orbit. Also, the radio signals are stronger. However, the satellite is only in contact for a short while, as it makes an orbit every 90 minutes. To get continuity of connection, several satellites are needed. This is known as a **satellite constellation**. The *Iridium* system provides satellite phone services to remote areas. Another use is for a satellite to pick up data in passing one transmitting station and transmitting them to a second station as it passes over it at a later time.
- **Medium Earth Orbit** - Height 2000 - 35 000 km. The functionality is similar to the Low Earth Orbit, but the period of connectivity and coverage are improved.
- **Geostationary Orbit** - Height 35,786 km. At this height, the orbit period is exactly 1 day (86 400 s), so the satellite remains above a particular region of the Earth.

The angle of the beam leaving the parabolic dish on the satellite determines the **coverage** on the ground (*Figure 193*).

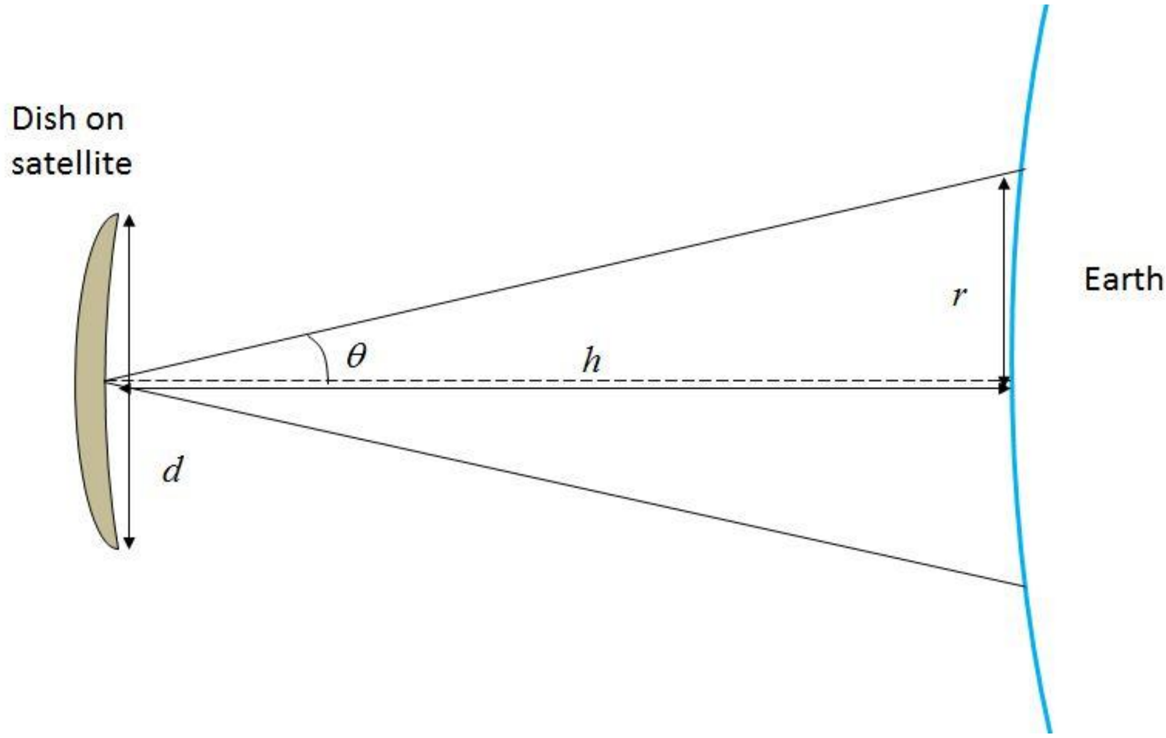


Figure 193 Beam coverage

Consider a satellite that has a dish that is d m in diameter. The satellite is h m above the ground and transmits with a wavelength λ m. The **angle of aperture** is given by the equation:

$$\sin\theta = 1.22\frac{\lambda}{d}$$

.... *Equation 105*

We can then work out the radius r by using:

$$r = h\tan\theta$$

.... *Equation 106*

The UHF waves from a satellite pass straight through the ionosphere.

14E.129 Satellite Communication Frequencies

These are in the range 1 GHz to 40 GHz (wavelength 0.3 m to 7.5 × 10^{-3} m). They are in internationally agreed bands which are set out in the table below:

Band	*Frequency /GHz*	*Use*
L	1 - 2	Global Positioning System; satellite telephony; maritime and aeronautical communications
S	2 - 4	Weather radar; ship surface radar; some communications
C	4 - 8	Satellite communications, TV relays
X	8 - 12	Military radar, weather monitoring, air traffic control, vehicle tracking in law enforcement.
Ku	12 - 18	Satellite broadcasting downlink, e.g. Astra Satellite.
Ka	26 - 40	Satellite broadcasting uplink. high resolution radar.

High frequencies tend to be affected by very heavy rainfall in tropical areas (known as **rain-fade**). So, satellites used in these areas tend to use lower frequencies.

The radio signals sent from the earth are referred to as the **uplink**. The signals sent from the satellite back to the Earth are called the **downlink**. These two links have different frequencies. The uplink has a higher frequency. The reason for this is that high frequencies are needed to get radio signals through the atmosphere without being reflected. However, the high frequency signals are attenuated more by the atmosphere, so need to have a source with higher power. This is easily achieved on a ground station; you simply use a big enough power supply.

Where there are several satellites in a **network** (sometimes called a **constellation**), there will be channels of communication between the satellites. These are called **cross-links**. The idea is shown below (*Figure 194*):

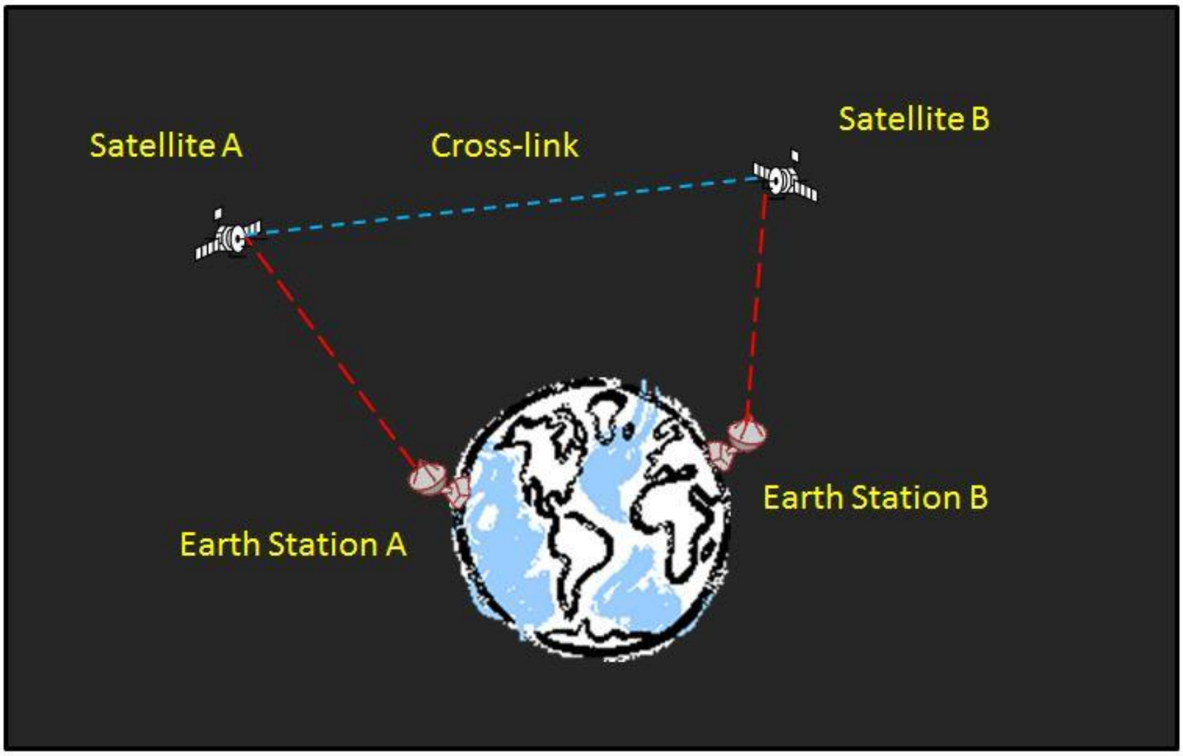


Figure 194 Cross links between satellites

The signal from **Earth Station A** will not reach **Satellite B**, for it to **relay** the communication to **Earth Station B**. So, Earth Station A has to communicate to Satellite A. This acts as a repeater station, sending the information on a **cross-link** to Satellite B, which in turn repeats the information to Earth Station B.

Satellites need to be lightweight as they are hugely expensive to launch. The power supplies (from solar panels with back-up batteries) are limited, so a lower downlink frequency is used. In the C-band the uplink frequency is 6 GHz, while the downlink is 4 GHz. Another reason for this is to prevent interference that would occur if the uplink and downlink signals had the same frequency. Another advantage of a lower frequency is less tendency for rain fade.

De-sensing is a problem encountered frequently by radio engineers. It is the effect of a strong transmitted signal on the weak received signal. It happens if the reception and transmission frequencies are close and results in **audio distortion**, and **loss of range**. At worst it can lead to the **complete loss of the signal**. The solutions to these problems are well-known to radio-engineers, e.g. antenna design and separation. Frequency separation is the easiest to achieve with a satellite.

14E.1210 **Advantages and Disadvantages of Transmission Media**

These are summarised in the table:

Medium	*Advantage*	*Disadvantage*	*Security*	*Cost*
Radio	No wires needed. Can be spread over a wide area	Energy inefficient. Interference can make signals unintelligible. Poor security	Poor. Signals can be easily intercepted.	Low
Wire	Less energy needed. Transmission and reception equipment are simpler, with no need for modulation.	Only a limited number of receivers are possible. Degradation of quality over distance due to attenuation and noise. High cost of infrastructure.	Better than radio. Physical connection needed to tap wire. Electromagnetic radiation can be detected and used to pick up signals (eavesdropping).	High
Fibre-optic	High rates of data transmission are possible. Less susceptible to noise and interference.	Optical fibre is digital only. Therefore, conversion from analogue and back are needed. Cracks in glass can interfere with transmission.	Better than electrical wires, but still possible to eavesdrop using sophisticated equipment.	High

Security is an issue with all data transmission. With radio, it is possible for anyone with appropriate equipment to intercept radio messages. If the data are sensitive, then **encryption** methods will enhance security. During the Second World War, the German *Enigma* machines provided sophisticated levels of encryption, but the capture of several machines enabled allied cryptographers to crack the code. This was helped by an early day computer called *Colossus*. In an experiment conducted just a few years ago, a replica Colossus cracked the Enigma code faster than a modern-day PC.

Security of modern **Wi-Fi** systems, which are radio-based is certainly a concern. A wired network is more secure. An optical fibre network is even more secure, but not completely so. Individuals and organisations that hold sensitive data on their computer systems must ensure that these data are kept secure from hackers. The computer is a more likely target than the transmission system, so measures need to be taken to:

- ensure the physical security of the equipment, e.g. ensuring only trusted employees have access to servers.
- keep anti-virus and network security up to date.

- back up data to ensure that they are not lost in the case of a drive failure.

Some embassies that handle information that is sensitive towards national security have started to go back to a more low-tech approach. They use **typewriters and paper**.

Questions

Tutorial 14E.12

14E.12.1

The frequency from a satellite is 600 MHz. What is the wavelength of the waves?

14E.12.2

A keen DIY enthusiast wants to make his satellite dish smaller, so that it's less obvious on the house. So, he makes it 20 cm in diameter. Will it work? Explain your answer using the physics you know.

14E.12.3

A transmitter has a dipole aerial which consists of two rods of total length 1.50 m and diameter 1.0 cm. The power from the dipole is 3.5 W.

(a)

Calculate the wavelength and frequency of this transmitter.

(b)

Calculate the power intensity at the surface.

(c)

Calculate the power intensity at the aerial of a receiver 2.5 km away.   Give your answer to an appropriate number of significant figures.

14E.12.4

A remote transmitter is 10 km from its base station and is connected by a land line that consists of a coaxial cable which has a centre conductor of diameter 1.0 mm. It has a thick outer sheath that has a resistance that is much lower than the centre conductor.

Calculate the resistance of the central conductor.

Resistivity of copper = $1.68 \times 10^{-8}\ \Omega$ m

14E.12.5

Data are transmitted along a wire at a rate of 100 megabits per second. When the current that forms a '1' flows, a magnetic field of flux density 2.5 nT is measured. Calculate the EMF.

14E.12.6

A coaxial cable is 25 metres long and has a capacitance of 56 pF m^{-1}. It is carrying a signal of 105 MHz.

(a) Calculate the capacitance of the cable.

(b) Calculate the reactance at this frequency. Give the correct unit.

14E.12.7

An optical fibre has an attenuation coefficient of 1.2×10^{-5} m^{-1}. At one end is an infra-red laser of power 250 mW. The light passes a rectangular window that is 6.0 mm wide and 2.0 mm high.

(a)

Show that the intensity of the radiation is about 21×10^{6} W m^{-2}.

(b)

Calculate the intensity of the radiation at the receiver, if the optical fibre is 50 km long.

(c)

What is the efficiency of the transmission?

14E.12.8

A geostationary satellite is 36000 km above the Earth. It has a transmitting antenna of diameter 2.0 m. If the frequency of the waves transmitted is 7.5 GHz, calculate the radius of coverage.

Tutorial 14 E.13 Time Division Multiplexing	
AQA Syllabus	
Contents	

14E.131 Remote Controlling

In the earliest days of steam power, the engines had to be operated where the engine stood. There were controls on the engine itself and the boiler to provide the steam was close by, if not in the same room. Orders to increase (or decrease) the output of the engine had to be sent to the engine house by a runner. This was not an efficient process. This was particularly the case in steam ships. One way of improving that was to have a speaking tube, in which the captain of the ship would speak into a tube to give his orders to the engine room crew (*Figure 195*). It was simple but not foolproof.

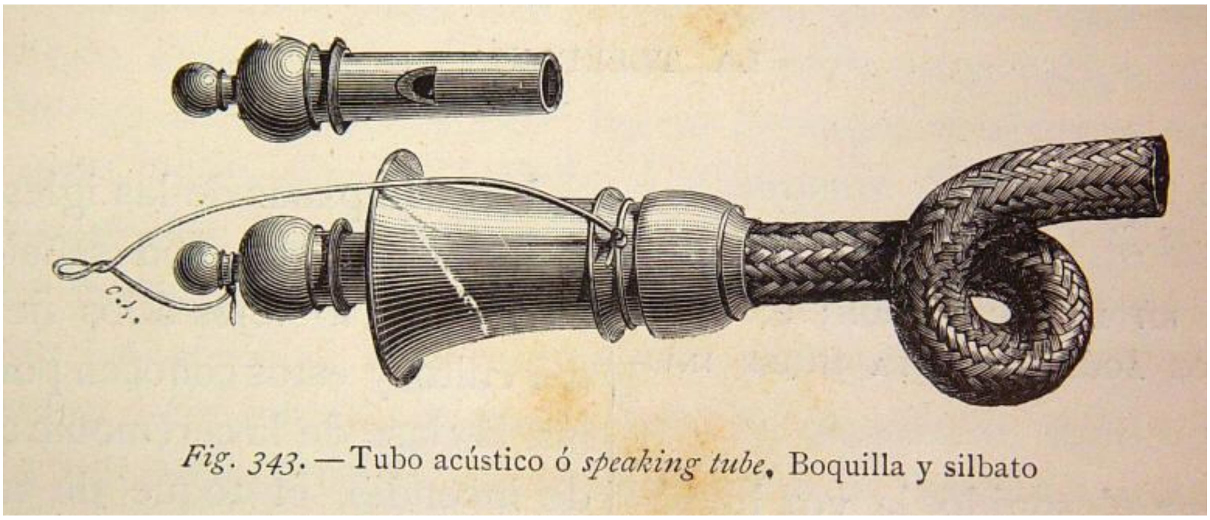


Figure 195 Speaking Tube (Wikimedia Commons: Biblioteca de la Facultad de Derecho y Ciencias del Trabajo Universidad de Sevilla)

Such devices are still in use today.

A development of this was the ship's telegraph as shown here (*Figure 196*):

Figure 196 A ship's telegraph Photo from Pinterest (Ruby Lane))

The captain would pull the lever to, say, *Half Speed Ahead*. A repeater in the engine room would relay that message to the operators of the engine.

On small branch railway lines, a small engine would push and pull a couple of coaches up and down the branch line. This would save the need for the engine to run around the train at the far end. The driver would stand in a compartment at the far end of the train and would issue bell codes to the fireman to open up the regulator (throttle) and change the valve gear.

When electric multiple unit trains were first introduced, they offered the advantage that they could easily be controlled from either end. However, they often had a motorised carriage (power car) in the middle of the train that needed to be controlled remotely. If a pair of wires was needed for each control, this could end up needing a lot of wiring, which is expensive and bulky. The wires needed to be able to be split for maintenance purposes, which required multi-way plugs. These could be a source of trouble.

A single cable can be used to control remotely a locomotive using the system of **Time Division Multiplexing** (TDM).

14E.132 Basic Ideas of Multiplexing

Multiplexing (sometimes abbreviated to **muxing** - horrible word) is about loading multiple signals onto a single connecting channel. It is not a new technology. It was introduced in 1910 by an American engineer, George Owen Squier (1865 - 1934). Both analogue and digital signals can be multiplexed. The multiple channels are combined by a multiplexer (MUX). The signals are sent down a single channel (wire, optical fibre, or radio link). They are then sorted out into the same number of channels by a demultiplexer (DMX). This is shown by the diagram below (*Figure 197*):

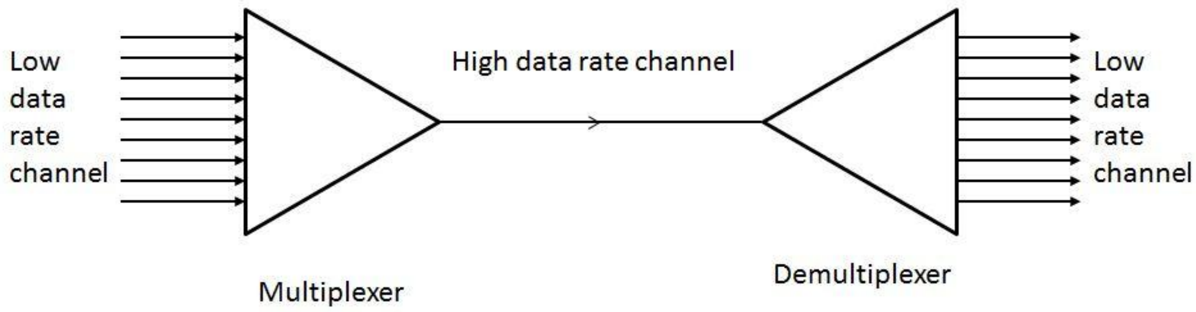


Figure 197 Multiplexing

In this example there are 9 low data rate input channels. The high data rate channel must be able to carry data at a rate of 9 times each of the input channels.

There are several ways that data can be multiplexed. We will consider the ideas of **time division multiplexing** (TDM) which is the one on the syllabus. The diagram (*Figure 198*) shows the idea of time division multiplexing:

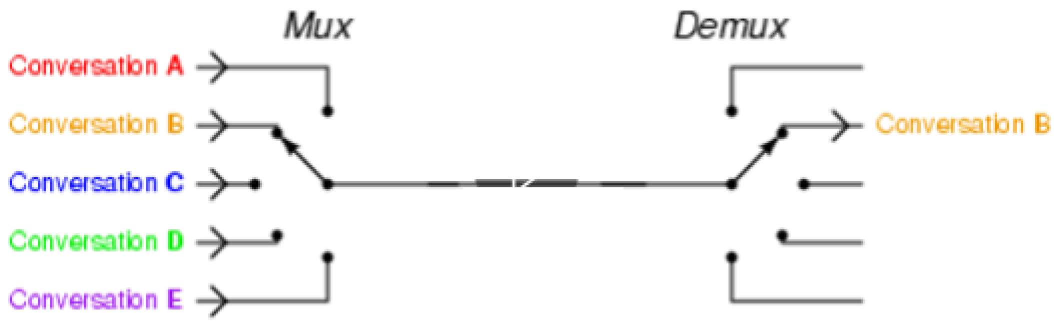


Figure 198 Time division multiplexing (Image from Wikimedia Commons Tony R. Kuphaldt - Socratic Electronics website)

The mux and demux spend a certain period on Conversation A, then move to Conversation B, then to C, then D, and finally E, before starting over again. Each cycle occurs many times a second.

In this tutorial, we will consider how multiplexing principles can be used with a number of different applications. We will use TDM to explain what is going on in these systems. However, in reality, other techniques can be used. We will mention these at the end of the tutorial.

14E.133 Aircraft Systems

Most aeroplanes have control surfaces that are operated by the pilot through a system of rods and wires with levers. With larger aircraft, simple cable and rod systems can be heavy to use, so power assistance is included. Such systems are widely employed but are rather bulky. In a machine like the huge aeroplane in the picture below, the controls would be operated by hydraulic systems. In other older aeroplanes compressed air would be used.

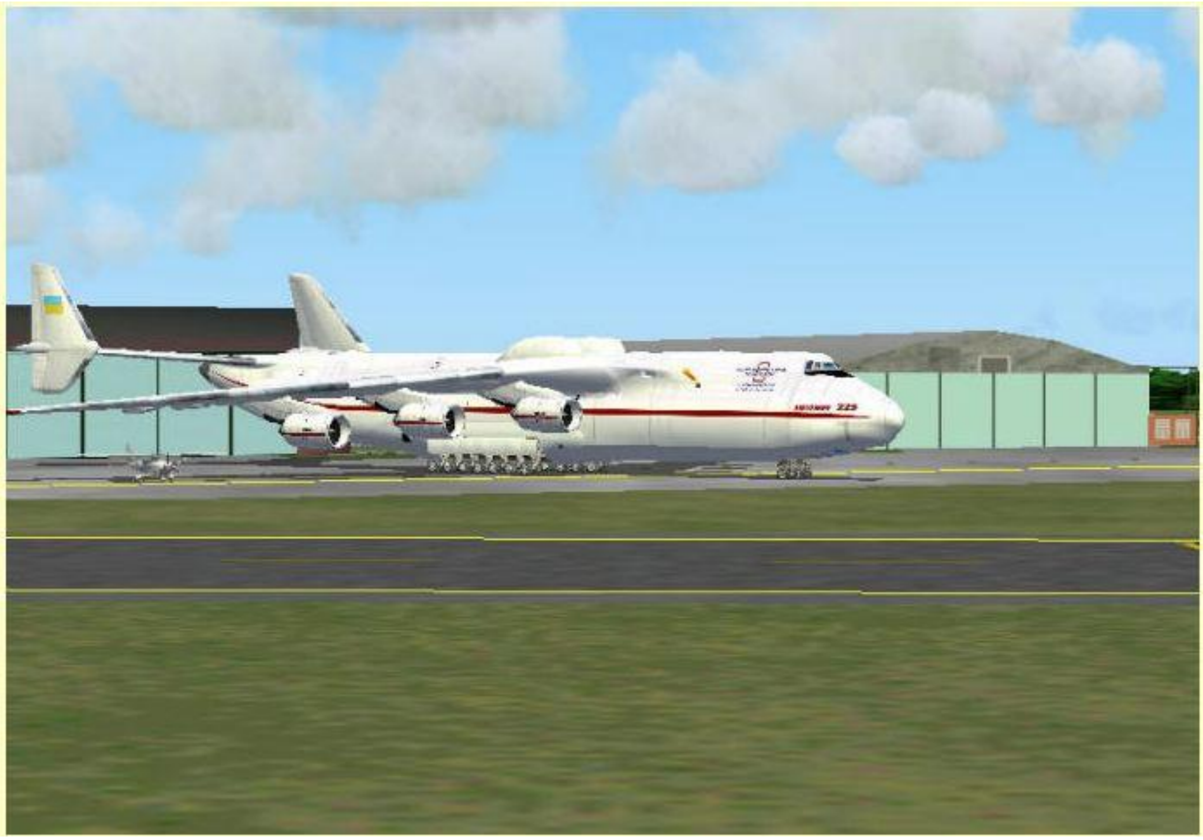


Figure 199 A very large aeroplane has control surfaces operated by hydraulic systems

Author's Note: This giant aeroplane is the Antonov An – 225. This was the world's largest aeroplane until it was wrecked during the initial Russian attacks against Ukraine in February 2022. This squalid war of attrition still continues at the time of writing, March 2026. It is intended to rebuild the AN – 225 once the war is concluded.

Unnecessary weight wastes fuel, which is expensive. So instead of using rods, cables, levers, and power assistance, we could use a system of motors and actuators connected to pilot's controls. Let's look at how it could be done (*Figure 200*):

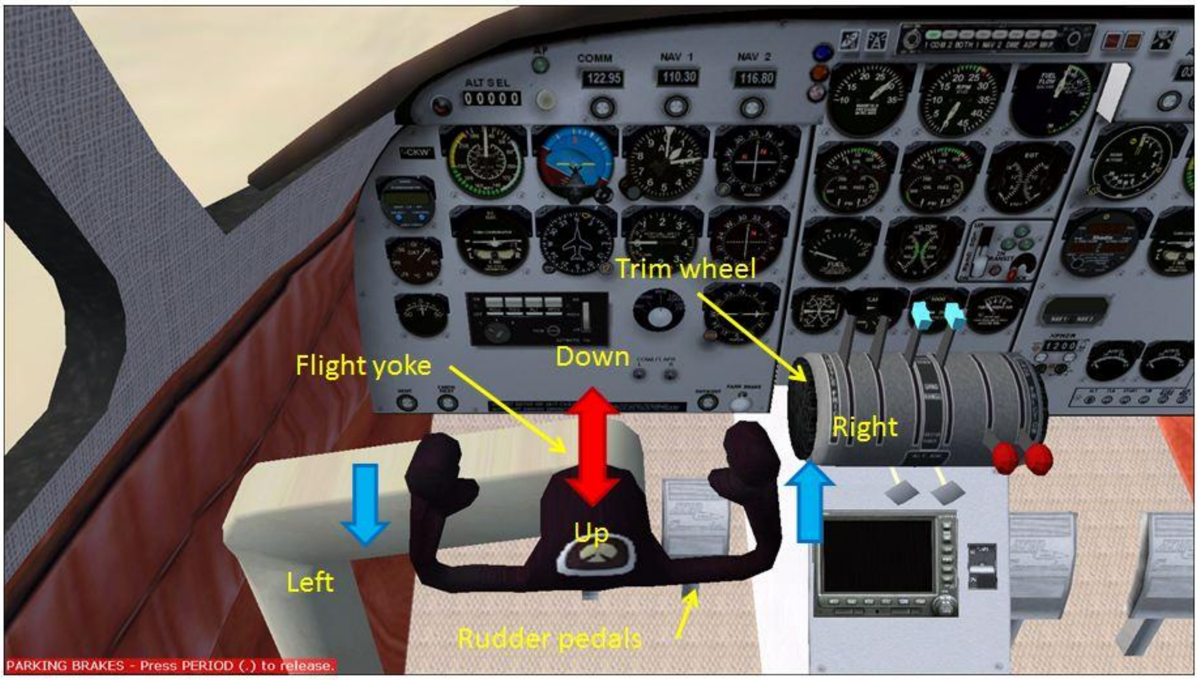


Figure 200 Pilot's controls in an aeroplane.

The **flight yoke** is the primary control. It activates various **control surfaces**. It can steer left and right. If the pilot pulls it towards her, the aeroplane will go up. If she pushes it away from her, the aeroplane will go down. This is, of course, a simplistic summary, but this is a physics tutorial, not a flight school! The pilot also has a trim-wheel that enables her to fly the aeroplane level without having to push or pull on the yoke to keep the aeroplane level. With an electric (or hydraulic) control system, the control surfaces are moved by **motors**.

The systems are shown in a schematic form here (*Figure 201*):

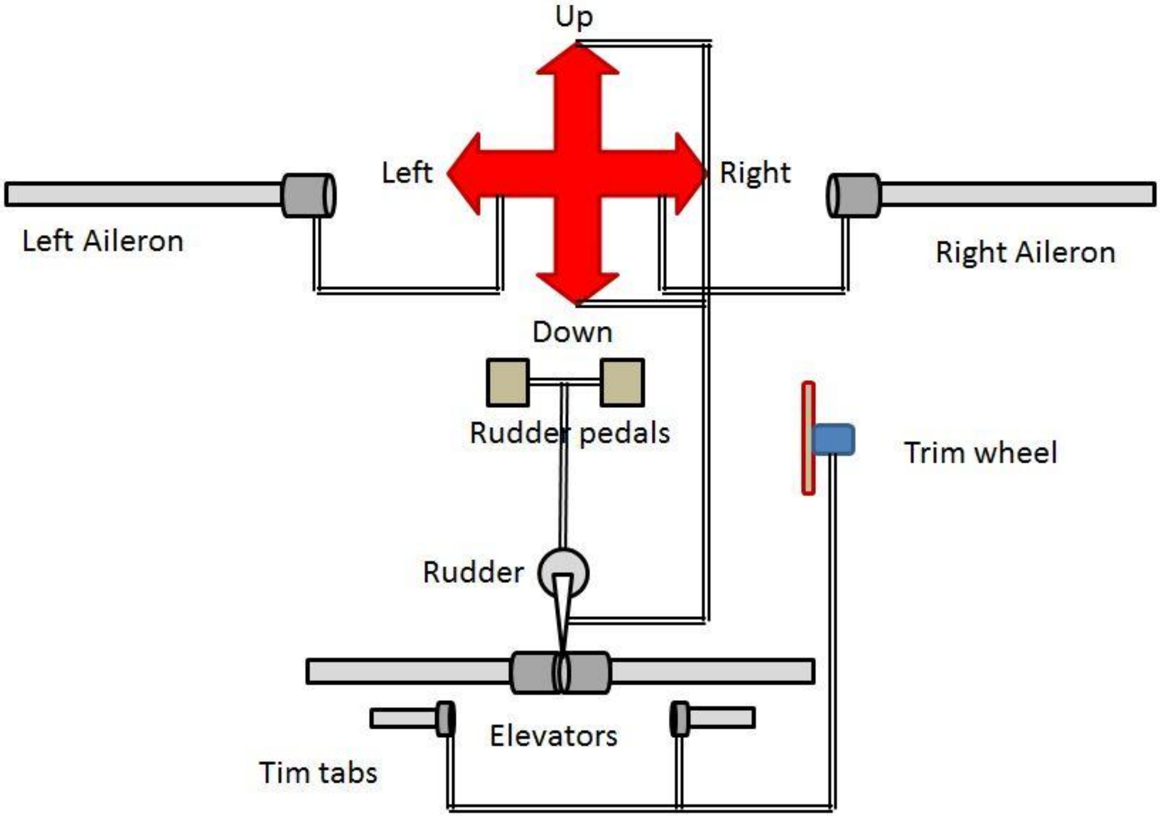


Figure 201 Aeroplane control surfaces operated by hydraulic/compressed air/electric motors

You can see that it's starting to get complicated. The more circuits there are, the more likely it is that there are going to be problems. Murphy's Law (*If it can go wrong, it will*) was coined by an aircraft engineer. If something goes wrong when you are up in the air, you can't land on a cloud, park up, and sort it out.

14E.134 Multiplexing

In aeroplanes the saving of weight is critical. If we can have one wire feeding several sensors, that saves a lot of weight. We use a technique called **multiplexing** to splice the output of each sensor before it is sent down the wire to the computer. In the picture below the control surface systems have been multiplexed. The control inputs from the pilot are carried to a multiplexer and computer (*Figure 202*).

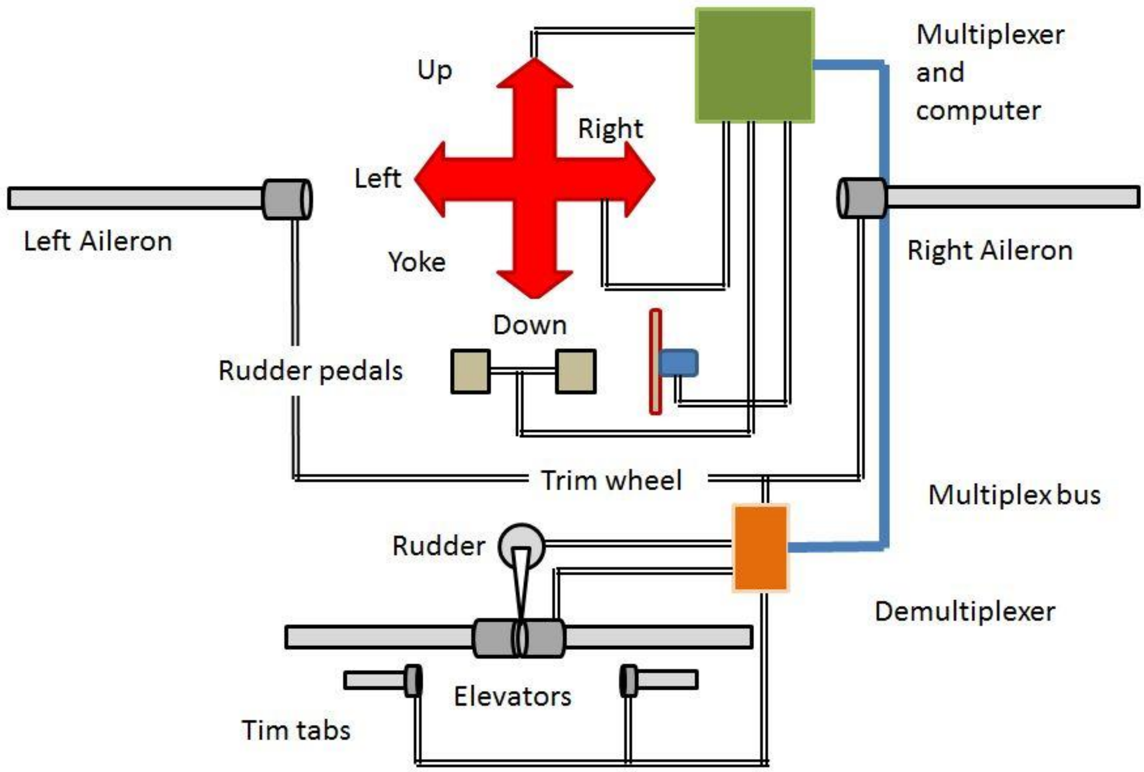


Figure 202 Multiplexing to operate aircraft control surfaces

Then the signals are sent to the ailerons and tail plane through **multiplex buses**. (A bus is a common conductor.) Then the command signals are **de-multiplexed** so that the relevant control surfaces are activated.

In **time division multiplexing** (TDM) the data from each sensor are split up into **equal sections of time**. They are then sent down the wire to a demultiplexer, where they are reassembled, giving as many outputs as there were inputs. TDM is often used in trains where a single wire carries the control signals from the remote driving cab to the locomotive. In earlier systems, the data were **analogue**, but modern systems are **digital**.

So, for our aeroplane the inputs from the pilot are as follows:

- The **yoke** controls *up and down*, called the **pitch**.
- The yoke controls the **roll** (or bank) to *left or right*.
- The rudder pedals control the **yaw** to the *left or the right*; (Yaw is a horizontal rotation to the left or right. It is used to co-ordinate turns to prevent the aeroplane from losing height.)
- The trim wheel controls the **trim** *up and down*, the adjustment of which keeps the aeroplane in level flight.

So, in the diagram (*Figure 203*), Y1 refers to Yaw command 1, while R1 refers to Roll command 1, etc.

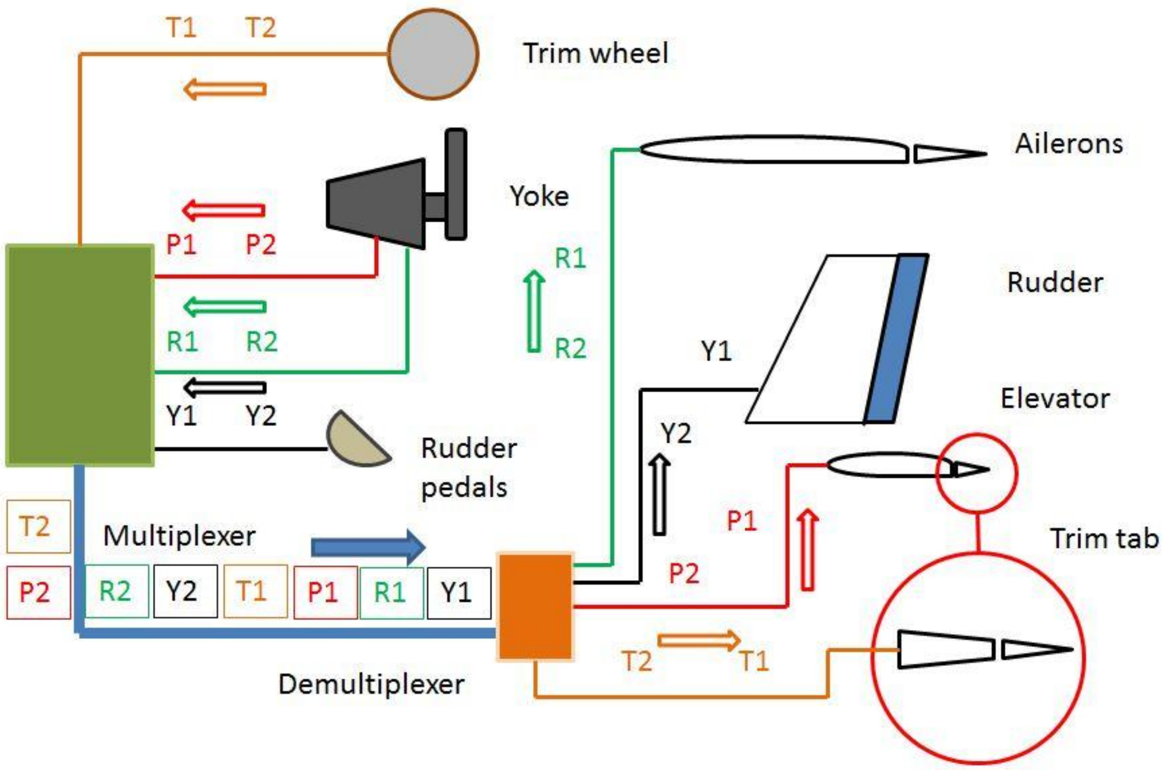


Figure 203 Multiplexing control inputs in an aeroplane

So, the **multiplexer** sends command Y1 in the first block of time, then R1 in the next block of time, then P1, and then T1. Each command is sent **one by one** in its own **block of time**. Each system shares the time equally. The **demultiplexer** then distributes the relevant commands to the different surfaces. Each command will have an identifier, so that it is sent to the correct destination. If this didn't happen, the aeroplane would do things it shouldn't do. This could rapidly make it un-flyable.

Let us suppose that the system uses 8-bit words (i.e. from 00000000 to 11111111). Let's say that the yoke is at 00000000 at fully nose-up and 11111111 at full nose-down. Suppose the **pitch** of the yoke were **neutral** (neither up nor down). It would give out a command 01000000. If two successive commands were the same, the demultiplexer would interpret the situation as **no change**. Therefore, the elevator actuators would do nothing, and the aeroplane would maintain the same pitch.

Suppose we allocated each system an identification number using an 8-bit word:

- System 1 is the **pitch** system using the **elevators** (which make the plane go up or down). Its identifier is 00000001.
- System 2 is the **roll** system using the **ailerons** (which makes the plane bank left or right). Its identifier is 00000010.
- System 3 is the **rudder** system that makes the plane **yaw** from left to right. Its identifier 00000011.
- System 4 is the **trim** system that maintains the plane in a particular attitude. Its identifier is 00000100.

The commands have values between 00000000 (full up, full left) and 11111111 (full down, full right).

The system needs to know three things:

- When a message is being sent.
- What system is being addressed.
- What the message is.

Let's think about the message in a block of time. The multiplexer and demultiplexer are co-ordinated with a **clock** that sends out pulses. This system of multiplexing is called **synchronous** multiplexing. The message in each block of time consists of **two bytes** (two 8-bit words). First is the system to be addressed. Second is the value that the system should be at. Consider this message:

00000001 01000000

Answer question 14E.13.3.

So, let's look at the sequence of commands that occurs through the multiplex wire:

Time Block	*Command*	*First Byte*	*Second Byte*	*Change*	*Effect on the plane*
1	Y1	00000011	01000000		Rudder neutral
2	R1	00000010	01000000		Ailerons neutral
3	P1	00000001	01000000		Pitch neutral
4	T1	00000100	01000010		Trim 2 places down from neutral, so nose slightly down
5	Y2	00000011	01000000	No change	Rudder neutral
6	R2	00000010	01000000	No change	Ailerons neutral
7	P2	00000001	01000000	No change	Pitch neutral
8	T2	00000100	01000110	Change	Trim 6 places down from neutral, so nose slightly further down.

This is a very simplified version of what really happens. The real situation is more complex with other systems being addressed and controlled. There are also sensors that feed the positions of control surfaces to the flight computer.

In real systems the clocks run at many thousands of times per second. They also carry much more information than suggested by this model.

Aircraft that use this system are called **fly-by-wire** and are controlled with **flight computers**. The aircraft manufactured by *Airbus Industrie SA* use this system. The aeroplane below (*Figure 204*) is one such example.

Figure 204 Airbus A-400 M

There are back up computers in case the main computer goes wrong. The systems are very expensive and have back-up systems which take over should the main one go wrong. The aircraft can fly but will be grounded until it is repaired. While the computers do much of the flying, they are supervised by pilots who can take over to fly the aeroplane if needed. However, if all the computers fail completely, there is no way of controlling the plane and it becomes unflyable. In one case, a pilot put his aeroplane into a **stall** during bad weather. The other pilot panicked and pulled back on his control stick, while the first pilot pushed forward... The aeroplane fell into the sea.

(A stall is where the aeroplane goes too slowly to fly. You should push the yoke forward in a stall to make the aeroplane drop and gain speed. It will then resume flight. All novice pilots are taught how to recover from a stall.)

The *Eurofighter* aircraft uses the system to such effect that an aeroplane, which is very unstable and in theory impossible for a human to fly, is as easy to fly as a light sports aircraft. The aircraft are very expensive, so only highly trained pilots are allowed anywhere near them.

14E.135 Computer Networks

Time Division Multiplexing is used widely in **computer networking**, especially where a company has two sites that are many kilometres apart. If each site had 100 computers, each of which needed to exchange data with the others, there would be thousands of kilometres of wire needed. With multiplexing, one wire can carry all the data. Consider this set of computers (*Figure 205*):

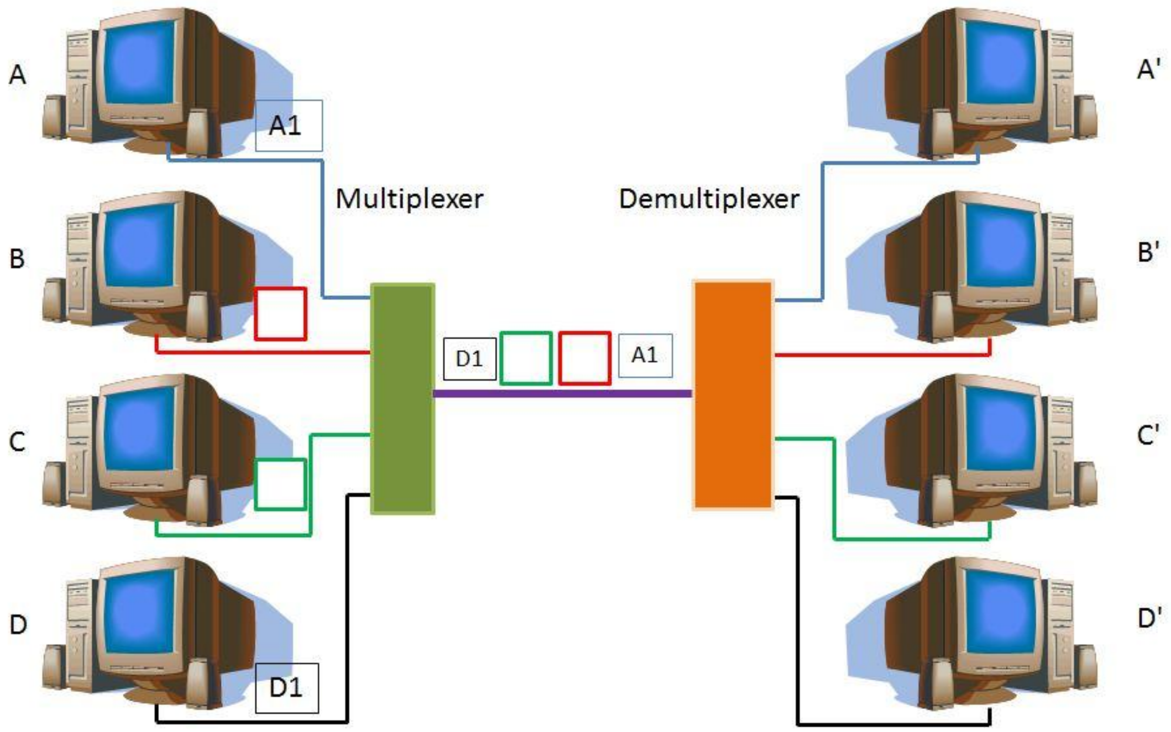


Figure 205 Multiplexing for a computer network

(Yes, I know that they all have CRT monitors, not TFT, but that was the clipart I had.)

Each computer can send messages to any of the others. So, A can send messages to A' or B' and so on. The demultiplexer in this case would have a **router**, and the data would have an address to which to send the data item. This is different to the aeroplane multiplexing in which one control input would only affect one control system. Notice that computers B and C are sending null or blank messages. They may be off-line or even switched off. That is no problem to the multiplexing system.

There are two drawbacks:

- The speed of the multiplex bus must be greater than the speeds of all the data inputs added together. So, if there are four inputs that each have a speed of 100 Mb s^{-1}, the multiplex data link must be able to carry at least 400 Mb s^{-1}.
- With computers B and C off-line, only half the capacity is being used.

14E.136 Asynchronous TDM

There is another method of multiplexing that addresses these two drawbacks. This is **asynchronous** TDM. The multiplexer reacts to the inputs according to the demands of the machines. It uses a **statistical algorithm** (computer-speak for say how often each computer needs to send messages).

Consider the four computers (complete with their ancient CRT monitors) multiplexing using asynchronous TDM (*Figure 206*).

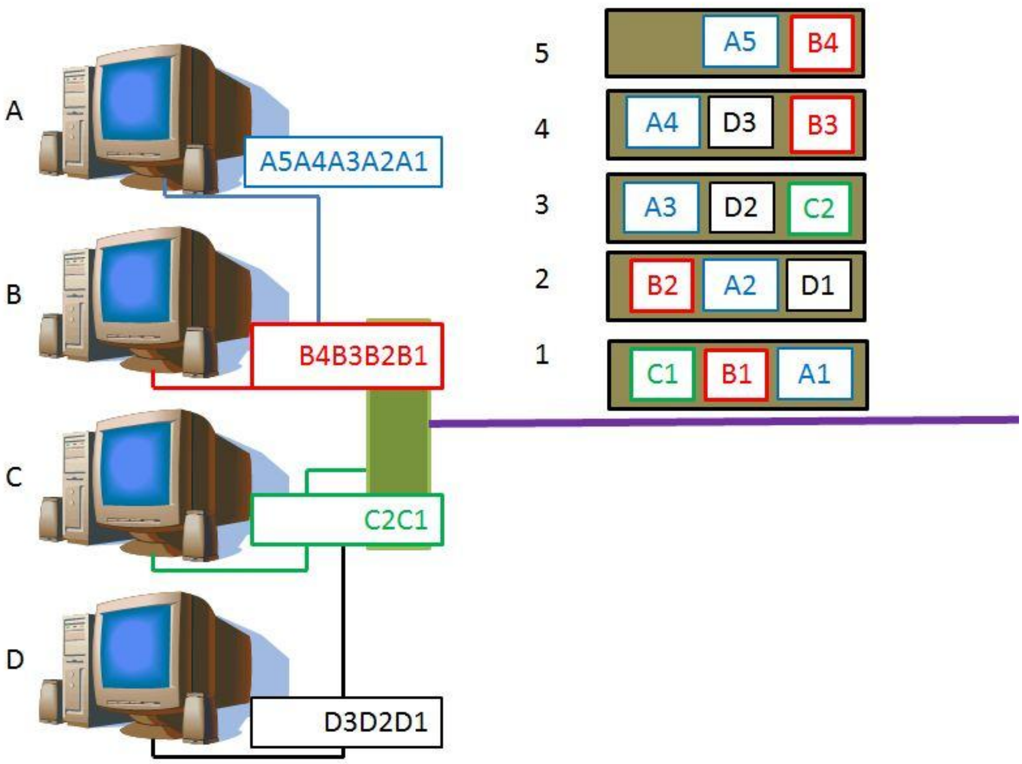


Figure 206 Asynchronous TDM

The data from each computer are put into blocks of three. So, the first block is A1, B1, and C1. Then the second block is D1, A2, B2. And so on. The fifth block is B4 and A5. The third space is empty. The multiplex system can cope with that.

We can think of each block of three as a wagon onto which three containers are placed.

There are fourteen blocks of data that are being carried in fifteen spaces.

The advantages of this system are:

- The whole bandwidth can be used for each computer.
- The TDM equipment is less complex (no synchronisation signal is needed).
- There is less crosstalk (picking up of signals from adjacent cables).

The disadvantage is that the synchronisation is lost. Therefore, checking signals have to be sent to verify the data to ensure that they are not corrupted.

14E.137 Multiplex Transmission

We have assumed that the multiplexed data are being transmitted using a wire. Remember that wires have problems of their own (see Tutorial 12), especially with high frequency signals. We can use optical fibres, in which the problems are much fewer. Optical fibre landlines are much more common nowadays.

Where it's difficult or expensive to lay landlines, radio signals are used.

Multiplex radio signals are essential for mobile telephones. Your mobile number does not have a dedicated frequency of its own. Instead, the frequency is determined by the service provider. For example, Virgin Media 4 G has frequencies of either 1800 MHz or 2600 MHz. O2 4G is on 800 MHz. (Other mobile phone providers are available.)

The calls to your phone are multiplexed with many other calls in a similar manner to that described above. You will be sharing your connection with several thousand other users. Your phone will only pick up the blocks of data that are addressed to your number. The blocks of data are transmitted so rapidly that you would not detect the gaps between each block.

14E.138 Digital Audio Broadcasting

Radios using **digital audio broadcasting** (DAB) have been around for about thirty years. DAB for the BBC is transmitted at a frequency of about 226 MHz. You do not have to retune the radio to a different frequency to change station. Instead, signals are multiplexed, and when you change stations, you are changing the demultiplexer to recognise a different identifier. The picture (*Figure* 207) below shows a stereo DAB radio tuner for a Hi-Fi system.

Figure 207 A DAB Radio Tuner

Consider four commercial stations that are multiplexed to a digital audio broadcast. *Junko Punko Radio* has identifier JP. *Brain Dead FM* has the identifier BD. *Drivel FM* has the identifier DV. *Ear-Buster Radio* has the identifier EB. Each is transmitting data as shown in the picture below (*Figure 208*).

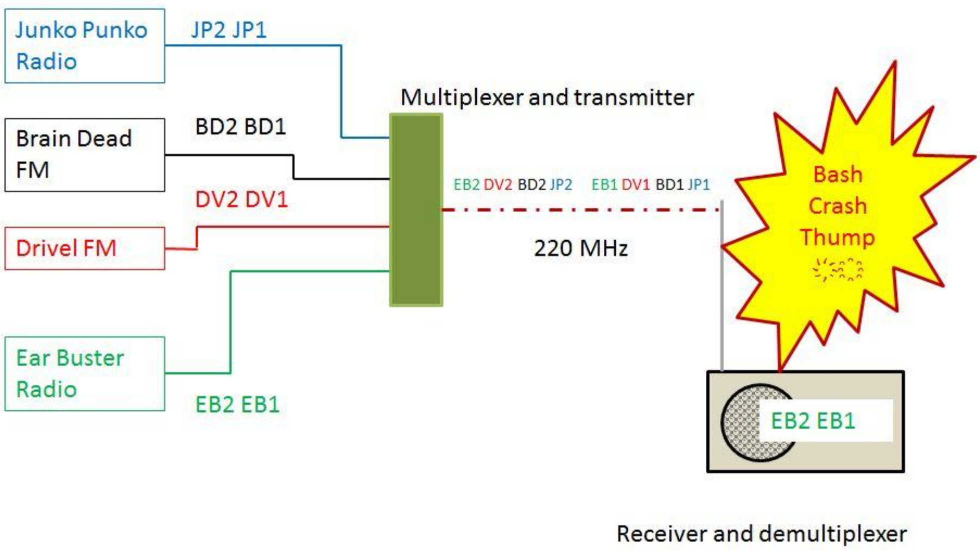


Figure 208 Multiplexing for DAB radio

The multiplexer is transmitting one block of data at a time, JP1, BD1, etc. The receiver is picking up all the blocks of data. Since it is set to receive *Ear-Buster Radio*, the demultiplexer only responds to the code with the identifier EB. The rest are ignored. If the listener wants to listen to *Brain Dead FM*, then the multiplexer then is changed to detect BD. Then only the data with the identifier BD are processed.

As well as the data that make the music, the receiver will display other information about the music being played (*Figure 209*):

Figure 209 A DAB radio display for a commercial radio station

Advantages

- There is coverage for much of the nation. If you want to listen to a Sheffield based station and you live in Bristol, you can do so.
- Extra data can be added (see above).
- Lots of extra stations are available.
- There is less interference.
- The broadcasts can be rewound, if you want to hear a bit of a programme again. You are really interested in what that politician said, so you can wind back.

Disadvantages

- There are **dead spots** where DAB cannot be picked up. In these islands dead spots include much of Wales, parts of the Pennines, a lot of Scotland, and quite a bit of East Anglia. In Germany, the coverage is over 90 %.
- Weak signals can sound terrible, as if the programme is "blowing bubbles". It sounds as if the radio is under water.

- DAB can be of variable quality when driving around in the car. In some places it is fine, but a few kilometres up the road, it is unpleasant to listen to. Also, the signals can reflect off buildings in towns, causing **multipath interference**. The same happens with FM radio.
- Not all cars have DAB radios. Retrofitting is expensive.
- Some people say that FM radio sounds better. It's like the debate between CDs and Vinyl LP records.

Figure 210 A typical small DAB radio receiver

Digital audio broadcasting uses **statistical time division multiplexing**. The details are beyond the scope of this discussion.

14E.139 Other Ways of Multiplexing

We have considered time division multiplexing as it's the easiest way of describing the concept. However, there are other ways that data can be transmitted down a channel. You are not expected to know these for the exam, but it is useful to be aware of them.

- **Frequency division multiplexing** (FDM). This involves sending signals of a number of different frequency ranges across a single channel. It is mostly used for **analogue** signals and receivers need to be tuned to a particular frequency to receive a particular signal.
- **Wavelength division multiplexing** (WDM). This is a variant of FDM. Various different wavelengths of light are used in optical fibres. The wavelengths used are all in the infra-red region, from 1270 nm to 1610 nm. Digital data are transmitted in this way.
- **Polarisation division multiplexing** (PDM). This was used in radio, with one channel being vertically polarised, and a second channel being horizontally polarised.
- **Code division multiplexing** (CDM). Each channel has its own code for transmission, and the data are demultiplexed asynchronously.

It is also possible to have **inverse multiplexing** whereby a single data stream is broken up into several parallel data streams and is transmitted over several data transmission lines. Then the data streams are reassembled into a single data stream. This is useful for increasing the data transfer rate if the transmission lines have a low data transfer rate.

Questions

Tutorial 14E.13

14E.13.1

What do you think the weakness is of such practices?

14E.13.2

Refer to Pages 220 to 221.

(a) How many possible positions would there be with 8-bit words?

(b) Why is the neutral position given by 01000000 rather than 00001111?

14E.13.3

Refer to Page 222. What is this message (00000001 01000000) saying?

14E.13.4

Later on in the flight, the multiplex wire carries these commands. Refer to Page 223.

Time Block	*Command*	*First Byte*	*Second Byte*
1001	Y1	00000011	01000011
1002	R1	00000010	01000011
1003	P1	00000001	01000000
1004	T1	00000100	01000010
1005	Y2	00000011	01000000
1006	R2	00000010	01000000
1007	P2	00000001	00111111
1008	T2	00000100	01000010

How does the plane respond?

14E.13.5

(a) What is the transmission efficiency of the asynchronous multiplexing?

(b) Suppose the data are transmitted using synchronous multiplexing. What is the transmission efficiency in this case?

Tutorial 14 E.14 AM & FM Techniques

AQA Syllabus

Contents

14E.141 What is Modulation?

When a radio signal is transmitted, a high frequency electronic circuit makes electrons **oscillate** in an **antenna**. This results in an **electromagnetic** wave propagating from the transmitter. It is detected by the antenna of a receiver, and the voltage amplitude is magnified by the phenomenon of **electrical resonance**.

A simple sine wave on its own is no good for transmitting information, whether it's analogue or digital. The remote reception of such a wave is merely a physics curiosity. However, if we can get information to ride piggy-back on the transmitted wave, things become very much more useful. We can do this by superposing the information that we want to broadcast to the transmitted wave.

The transmitted wave at the station frequency is the **carrier** wave. The information is put onto the carrier wave using a technique called **modulation**. In modulation the **audio signal** is added to the carrier wave.

14E.142 Amplitude Modulation

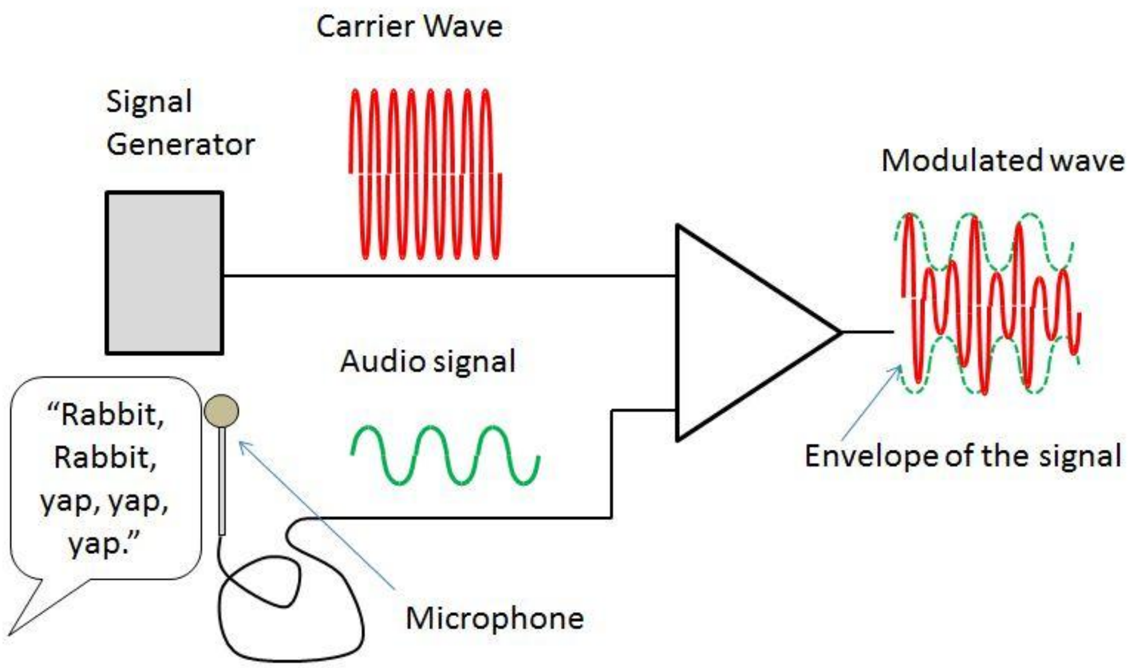


Figure 211 Amplitude Modulation

In **Amplitude Modulated** (AM) broadcasting the amplitude of the carrier wave varies so that the envelope traces the audio signal around the modulated wave. In the diagram above (*Figure 211*) the audio signal is shown as a sine wave. The diagram above shows the idea of amplitude modulation. In reality the audio signal of "Rabbit, Rabbit, yap, yap, yap" would be much more complex. The antenna of the receiver detects the changing voltage due to the variations in the amplitude. These are **demodulated** and sent to the amplifier and loudspeaker.

The separation of stations is about 10 kHz.

The graph below (*Figure 212*) shows the interaction between the signal wave and the carrier wave. In this case, the carrier wave is at a slightly higher frequency than the signal wave. In reality, the carrier wave is at a much higher frequency than the signal wave.

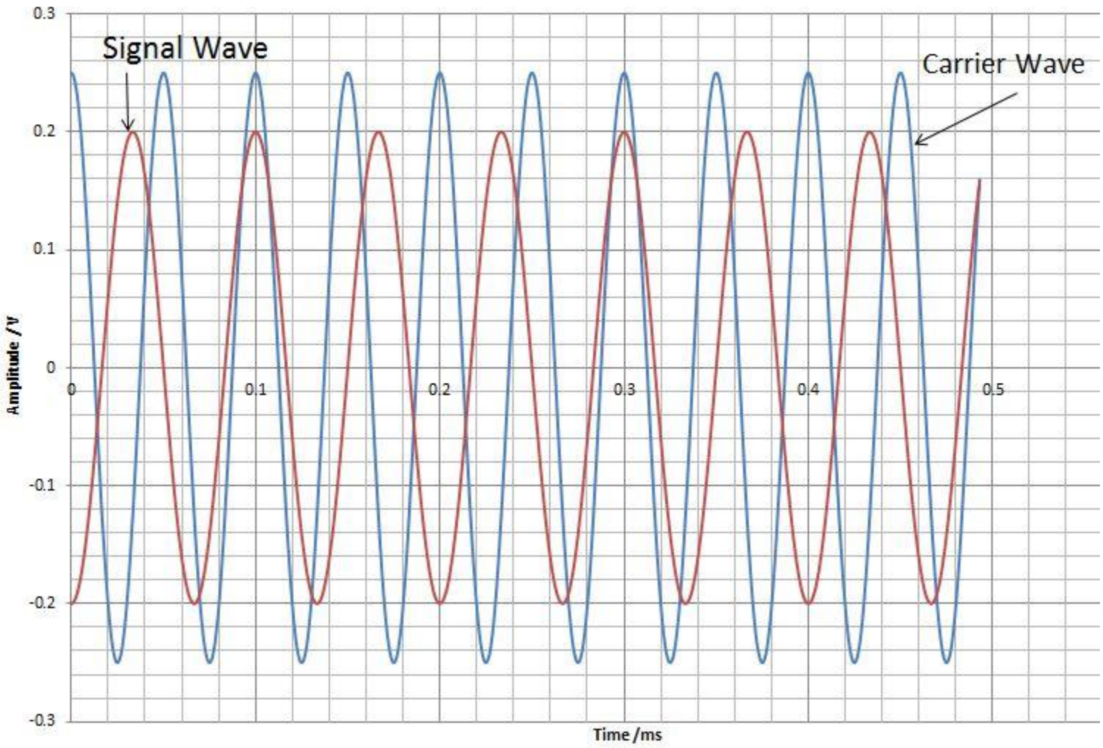


Figure 212 Signal wave and carrier wave

The resultant waveform looks like this (*Figure 213*):

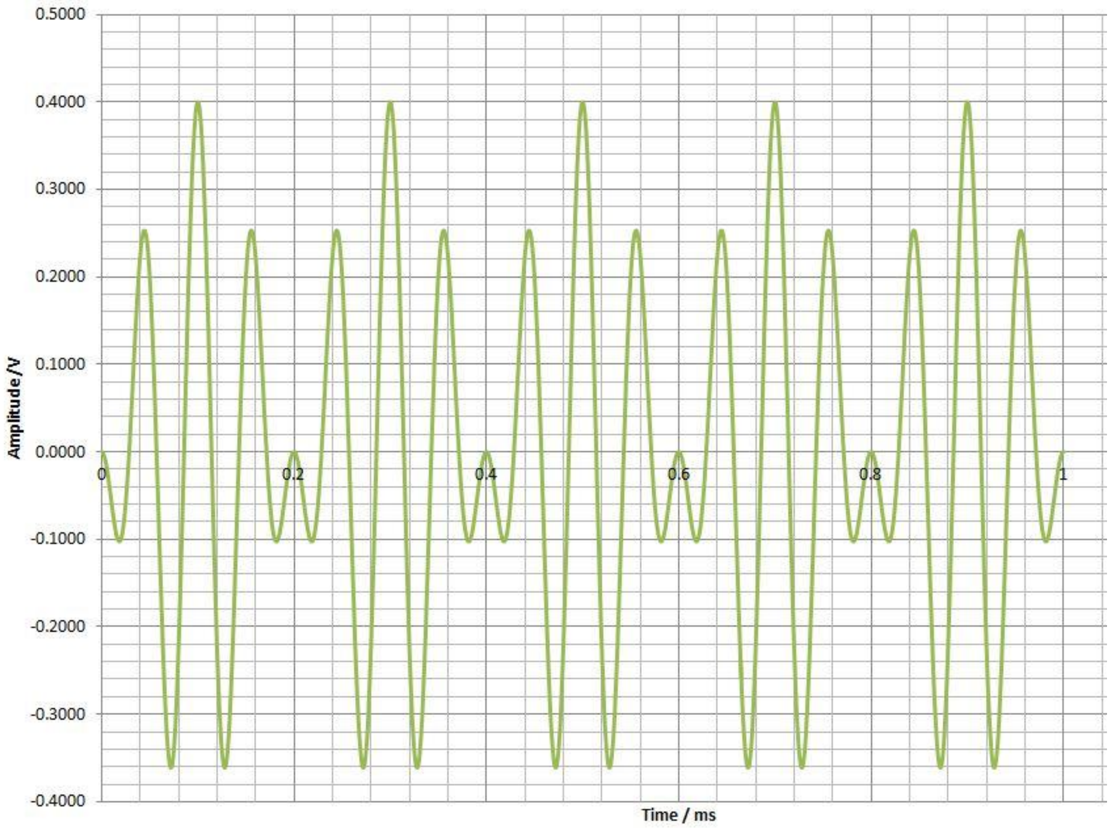


Figure 213 Resultant waveform of the signal wave superposed on the carrier wave.

14E.143 Frequency Modulation

The diagram below (*Figure 214*) shows an audio signal being combined with a carrier wave. In this case, the frequency of the carrier wave is changed by the frequency of the audio signal. Hence, it's called **frequency modulation** (FM)

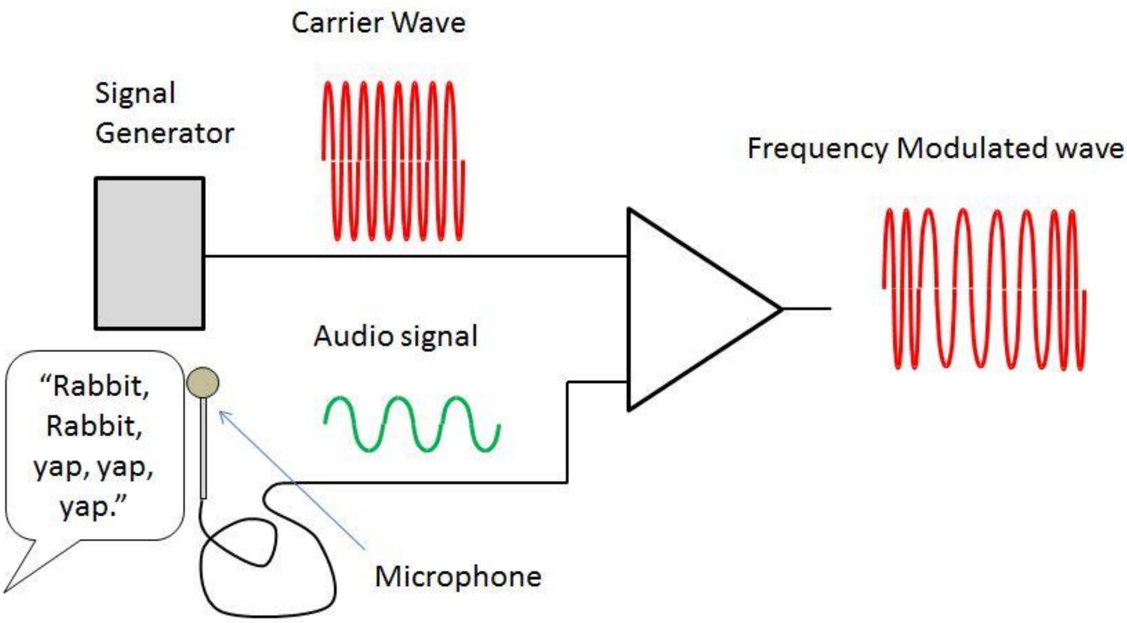


Figure 214 Frequency Modulation

Each station has a frequency that is separated by 200 kHz. Each frequency quoted is the middle value of the range of frequencies.

In reality the frequency variation is 53 kHz above and below the middle frequency. Above and below this is a 25 kHz cushion (**guard band**) to prevent interference between adjacent radio stations. The wide range of frequency allows for a high fidelity signal to be broadcast. The audio range is limited to 15 kHz. Although 53 kHz is much higher than this, it allows for other information such as the carrier signal (19 kHz) for stereo reception.

In the Exam

You are NOT expected to know details of any modulation or demodulation circuit.

You are expected to know about AM and FM. However, you are NOT expected to know about any other form of modulation. Nor are you expected to do a mathematical treatment of the modulation.

You are not expected to know about stereo reception.

14 E.144 Wave Bands

This radio receiver (*Figure* 215) can pick up stations from all the radio bands that are available for broadcasting.

Figure 215 SW LW MW FM radio receiver

The **medium wave band** has stations transmitting on AM and the frequencies are from 535 kHz to 1605 kHz. The **long wave band** has frequencies from 150 to 300 kHz. They too transmit using AM. In the UK Radio 4 still broadcasts on long wave at 198 kHz.

Short wave radio transmits at frequencies 1.6 MHz to 30 MHz. Mostly such stations use AM, but there are other forms of modulation available, including:

- Single sideband.
- Vestigial sideband.
- Continuous wave.
- Digital Radio Mondiale.
- Narrowband Frequency modulation for frequencies above 20 MHz.

14E.145 Bandwidth

Bandwidth is a term that describes a range of frequencies.

In electronic circuit theory the bandwidth of an amplifier is the **range of frequencies in which the power is greater than half the maximum power**. This is shown on the graph (*Figure 216*):

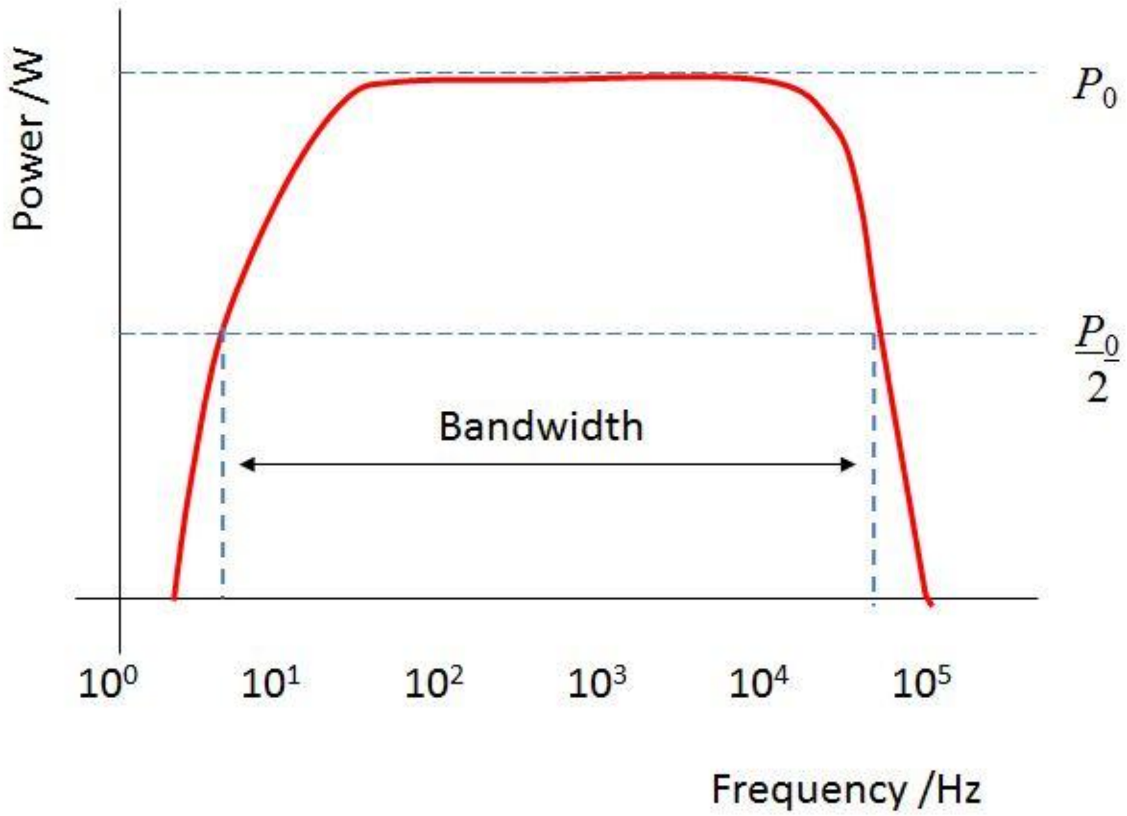


Figure 216 Bandwidth

If we consider the voltage amplitude, the bandwidth is the range of frequencies in which the voltage is greater than 71 % (1/√2) of the maximum voltage.

14E.146 AM Bandwidth

In radio communication the term **bandwidth** is used slightly differently. When we quote the frequency of a radio station, we are stating the carrier frequency. The signals that are super-imposed on the carrier wave form **sidebands**. The idea is shown in the diagram below (*Figure 217*).

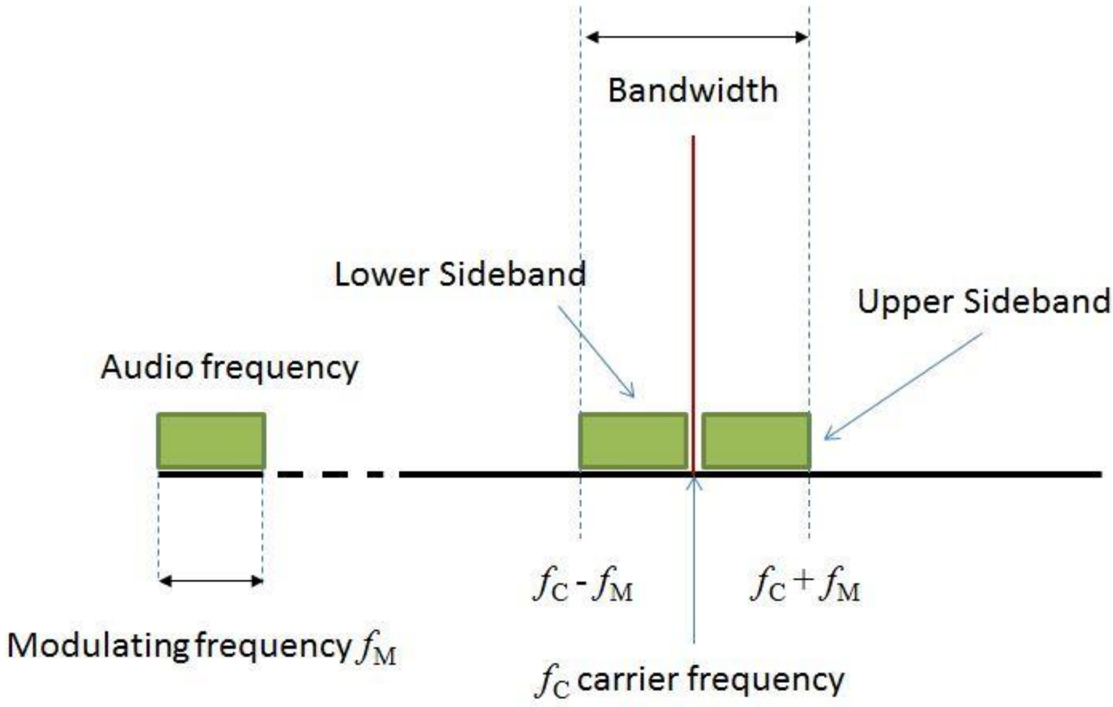


Figure 217 Showing sidebands

The range of audio frequencies (shown in the green box) is called the **modulating frequency**, f_M. The **carrier frequenc**y is given the code f_C. The **bandwidth** is simply the difference between the lowest frequency and the highest frequency. In AM radio the bandwidth is limited. Channel spacing in Europe is 9 kHz. To avoid interference, the lower and upper side bands need to be about 4 kHz each.

The bandwidth for AM is summed up in this simple equation:

$$\text{Bandwidth} = 2f_M$$

.... *Equation 107*

The narrow bandwidth explains why AM radio does not give high fidelity reception. There is no technical reason why AM should not have a higher bandwidth, but that would restrict the number of radio stations that would be available. Stations on the short-wave bands have even more restricted bandwidth.

A way round restricted bandwidth is to prevent stations with adjacent frequencies being anywhere near each other geographically.

14E.147 FM Bandwidth

The bandwidth for FM transmission is not so easy to understand. In theory the frequency range in an FM signal can be infinite. For narrow bandwidth FM, the situation is similar to AM, with a central carrier with two sidebands. However, the lower sideband is 180° out of phase with the upper sideband.

For wide bandwidth FM, the situation is even more complex. There are a lot of sidebands. It is the usual practice to consider the bandwidth as that which contains **98 % of the powe**r. The detailed analysis of FM sideband spectra is not on the syllabus, so we will not worry about it here.

The difference between the **maximum modulation frequency** of the signal and the **carrier frequency** of the FM transmission is called the **deviation frequenc**y, which has the physics code Δf. This quantity is used in **Carson Bandwidth Rule**, devised by John Renshaw Carson (1886 - 1940). This is given in the equation:

$$\text{Bandwidth} = 2\Delta f + f_{\text{M}}$$

.... *Equation 108*

Most FM radio transmissions use a deviation frequency of 75 kHz. The audio frequency is set at 15000 Hz.

The spacing is set at 200 kHz on the FM band (87.5 MHz to 108 MHz). This allows for a cushion of 25 kHz above and below to prevent interference.

Aircraft radios operate on the **Aircraft band** frequencies, from 108 MHz to 137 MHz. From 108 MHz to 118 MHz, the stations are set at 50 kHz intervals transmitting using narrowband FM techniques. There are 200 channels. These are used for navigation beacons. The picture (*Figure 218*) shows an aeroplane on a flight simulator approaching the aerodrome at Aix-les-Bains (LFLB).

Figure 218 On approach to an aerodrome

The navigation radio is tuned to 115.40 MHz (Chambéry VOR, CBY) and the instrument just below the aircraft registration (F-JLCO) is the navigation display linked with this. Since this is a physics tutorial, not a flight school, the meaning of the display is not relevant. The VOR (VHF Omnidirectional Radio Ranger) transmits on narrow band the Morse code identifier, in this case CBY, or "-.-. -... -.--". The pilot can hear this through the headphones. It is repeated every fifteen seconds and gets rather monotonous if left on. Knowledge of Morse code is expected of all pilots.

The communications radio is set to 121.20 MHz, which is the radio frequency for the tower at Aix-les-Bains.

In the last few years aircraft radios in Europe have to use a separation of 8.33 kHz with a frequency deviation of 2 kHz. It used to be 25 kHz. This has not pleased pilots who have had to replace their radio equipment (not cheap).

Your answer to Question 14E.14.4 will show that air-to-ground radio communication does not exactly result in High-Fidelity sound quality.

14E.148 Data Capacity of a Channel

Digital data are transmitted as a series of pulses of value 1 (ON) and value 0 (OFF). When digital signals are sent by radio waves, they are carried like this (*Figure 219*):

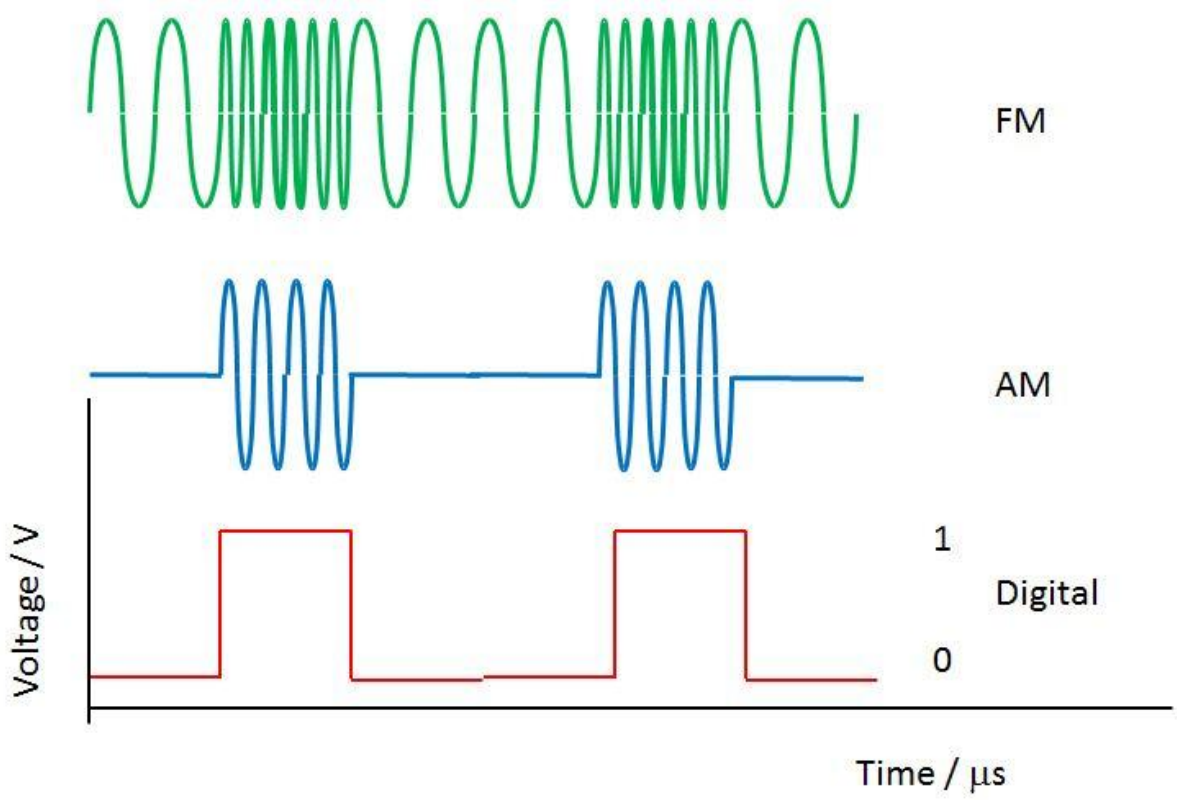


Figure 219 Transmission of data by radio waves

The higher the frequency of transmission, the more data are carried every second. Also, the bandwidth can be increased, so that even more data can be carried. So, let's have a look at some definitions:

- **Signal Bandwidth** is the range of frequencies in the transmitted signal.
- **Channel bandwidth** is the range of frequencies allowed that do not lose a significant amount of energy.
- **Channel Capacity** (or **Maximum data rate**) is the maximum rate in bits per second (bps) at which data are able to be transferred across the channel.

A channel of bandwidth B Hz can be used to carry $2B$ symbols (or level changes) every second. The channel capacity for binary digital signals is given by this equation:

$$\text{Capacity} = 2 \times B \times \log_2 M$$

.... *Equation 109*

B is the bandwidth in Hz, and M is the number of signalling values. For a **binary** channel, $M = 2$, so $\log_2(2) = 1$. Therefore, the capacity is twice the bandwidth. This relationship is called the **Nyquist Equation**. It is NOT on the syllabus, but it's the sort of thing that might come in an extension question.

For noisy data channels, the **Shannon data capacity relationship** applies, which takes into account the **signal-to-noise ratio**. A weak signal is noisy and this can blot out the data.

The limiting factors for the channel's data carrying capacity are:

- The bandwidth - the bigger the bandwidth, the more data that can be carried.
- The bandwidth of the transmitter and receiver.
- The signal to noise ratio.

14E.149 Bandwidths of Different Kinds of Data Channels

High speed internet communication is, nowadays, reckoned to be as much of vital service as electricity, gas, water, sewerage, and telephone. The **world wide web** was invented in 1989 to allow university computer systems to exchange data, using a language, **html**, devised by the Physicist Tim Berners-Lee in three hours one Christmas Day (doesn't say much for his Christmas, does it?). The **Internet** was a network of military computers in the USA set up in the 1970s, so that a nuclear strike on one site would not destroy the capability of the others. Nowadays the terms are interchangeable.

At the start of this century, more people and institutions were connected to the web using **modems** connected to their telephone lines. These were used to dial-up a number given by the internet service provider. The maximum channel carrying capacity was 56 kilobits per second. This page would take 5 minutes and 3 seconds to load, sufficient for you to make a cup of coffee. The bandwidth of a telephone line is 4 kHz.

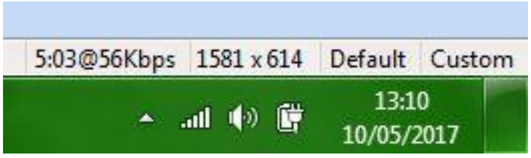


Figure 220 Old-fashioned dial-up internet

This speed would be considered primitive today. Also, your phone was engaged while you were on-line. You could neither make nor receive calls.

ADSL **Broadband** internet of 1 megabit per second came in about ten to fifteen years ago. It was considered (rightly) a considerable advance on dial-up. You were on-line all the time, and your phone-line was available for use. The photo shows an old ADSL modem.

Figure 221 Old-fashioned ADSL modem

Today, a broadband of 1 Mbps is considered unacceptable, although there are places in these islands where this is the speed. One village in Kent has had speeds as slow as 540 kbps. The village of Miserden in Gloucestershire has the dubious honour of being the slowest broadband in the UK, at 1.2 Mbps. And it's not just in the most isolated villages, either. Some parts of London have surprisingly low broadband speed. Of the major cities, Kingston-Upon-Hull has an average broadband speed of 12 Mbps, which is considered poor. Where I live, 38 Mbps is the quoted speed. (Since I wrote this, things may have changed. It seems that Canterbury in Kent has the slowest broadband,)

Gigabit Ethernet has a capacity of 1000 Mbps. You can also get ethernet adapters that use the house wiring to transmit data from the modem and router to a second computer.

Figure 222 Ethernet Adapter

This one operates at 200 Mbps.

USB 3.0 can transmit data at 5 Gbps. Some wireless networks can operate at 54 Gbps.

Questions

Tutorial 14E.14

14E.14.1

Classic FM has a frequency of 101.1 MHz. Calculate the minimum and maximum frequencies that are used for Classic FM.

14E.14.2

The bandwidth for a radio station broadcasting on a wavelength of 247 m is 8200 Hz.

(a) Calculate the audio bandwidth.

(b) Calculate the carrier frequency to this station.

(b) Calculate the minimum and maximum frequencies of this station.

14E.14.3

(a) What is the bandwidth of a radio station in the FM band?

(b) A station is transmitting on 93.5 MHz. What is the minimum and maximum frequency that it uses?

14E.14.4

In the last few years aircraft radios in Europe have to use a separation of 8.33 kHz with a frequency deviation of 2 kHz.

(a) How many channels does this allow?

(b) What is the bandwidth, assuming that the frequency deviation is 2 kHz and that the modulation frequency is 500 Hz?

14E.14.5

A transmission has a deviation frequency of 100 kHz and a modulation frequency of 50 kHz.

(a) What's the bandwidth?

(b) What is the data capacity in bytes per second?

14E.14.6

A student downloads a film which is a 1.5 Gigabyte file. It takes 10 hours to download. What is the rate of data transfer?

Answers to Questions

Tutorial 14E.01

14E.01.1

Use:

$$\text{Gain} = I_{ce} \div I_{be}$$

Gain = 495 mA ÷ 5 mA

Gain = **99**

14E.01.2

G = gate.

D = drain.

S = source

14E.01.3

A: The gate-source voltage is less than the threshold voltage. The depletion layer prevents any current from flowing

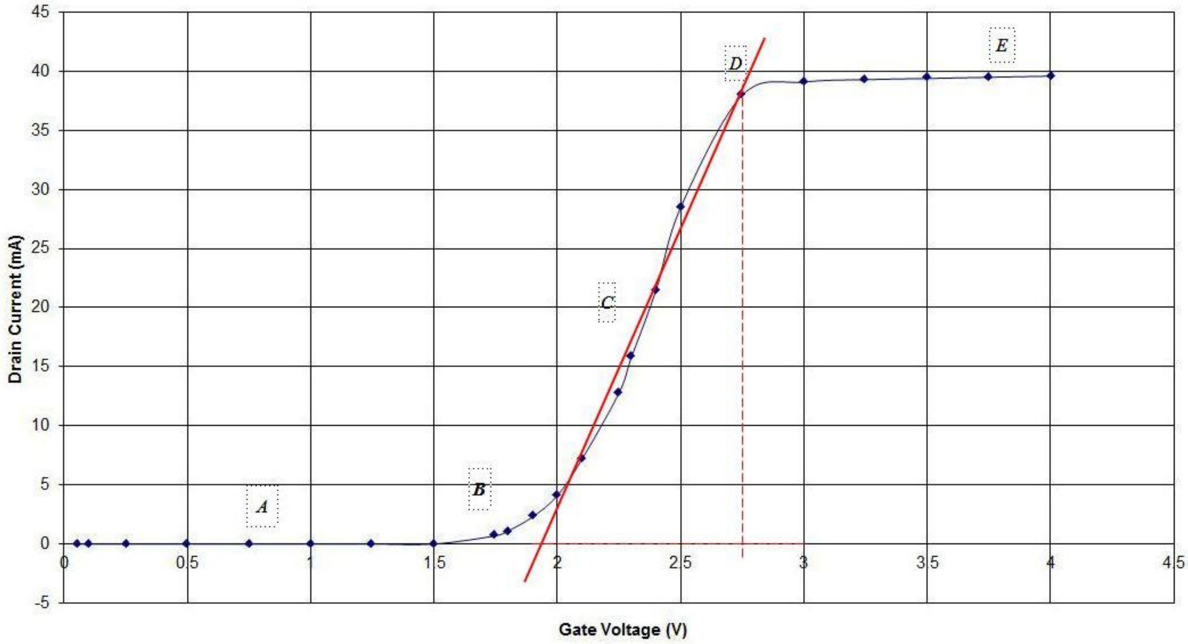
B: The threshold voltage has been reached. The conductive path in the n-channel has just opened to allow electrons to flow.

C: The gate source voltage is above the threshold. The n-channel path is now wide enough to allow a current of about 20 mA to flow.

D: The gate-source voltage is high enough for the n-channel to be fully open, so that the maximum current is flowing.

E: The MOSFET is carrying the maximum current possible, given the number of free electrons. It is saturated.

14E.01.4



The transconductance is the gradient of the red line:

Gradient = (38 mA - 0 mA) ÷ (2.75 V - 1.9 V) = 38 × 10^{-3} A ÷ 0.85 V

= **0.045 S** (2 s.f. are appropriate here)

This is equivalent to a resistance of 22 ohms.

14E.01.5

The MOSFET had a resistance of 22 ohms. The current limiting resistor was 150 ohms

Total resistance = 22 Ω + 150 Ω = 172 Ω

I = 38 mA

V = 38 × 10^{-3} A × 172 Ω = **6.5 V**

14E.01.6

When S1is open, the gate voltage is 0 V, as it's connected to the 0 V line by the 1 MΩ resistor.

When S2 is closed, the gate voltage rises to +Vs, which fully opens the n-channel by reducing and pushing back the depletion layer. No current flows through the gate.

The current limited by the 1 kΩ resistor flows through the MOSFET and the LED, making it light up

14E.01.7

Voltage from drain to source = IR_{ds} = 0.50 A × 22 Ω = 11 V

Voltage of the motor = 6.5 Ω ÷ 0.50 A = 13 V

The supply voltage = 11 V + 13 V = **<u>24 V</u>**

14E.01.8

(a)

It is a thermistor.

(b) Equation:

$$V_{\text{out}} = V_{\text{in}}\left(\frac{R_2}{R_1 + R_2}\right)$$

Put the numbers in:

2.0 V = 24 V × [R_2 ÷ (150 000 Ω + R_2)]

2.0 V × (150 000 Ω + R_2) = 24 V × R_2

300 000 V Ω + 2.0 V × R_2 = 24 V × R_2

22 V × R_2 = 300 000 V Ω

R_2 = 13600 Ω = **<u>14 kΩ to 2 s.f.</u>**

Tutorial 14E.02

14E.02.1

The voltage drop across the resistor = 12 V - 5.6 V = 6.4 V

Resistance = 6.4 V ÷ 10 × 10^{-3} A = **640 Ω**

14E.02.2

Current through the 1000 Ω resistor = 5.6 V ÷ 1000 Ω = 5.6 × 10^{-3} A

Current through the diode = 10 mA - 5.6 mA = **4.4 mA**

14.02.3

The current through the diode is 0 mA, so the 10 mA must go through the load resistor.

R = 5.6 V ÷ 10 × 10^{-3} A = **560 Ω**

14.02.4

Use the voltage divider equation:

$$V_{\text{out}} = V_{\text{in}} \left(\frac{R_2}{R_1 + R_2} \right)$$

V_{load} = 12 V × [300 Ω ÷ (640 Ω + 300 Ω)] = **3.8 V**

Tutorial 14E.03

14E.03.1

(a)

Photocurrent = 60 μA - 1.5 μA = **58.5 μA**

(b)

Area = π × (5.0 × 10^{-3} m)^2 = 7.85 × 10^{-5} m^2

Power = 10 W m^{-2} × 7.85 × 10^{-5} m^2 = **7.85 × 10^{-4} W**

(c)

Responsivity = 58.5 × 10^{-6} A ÷ 7.85 × 10^{-4} W = **0.074 A W^{-1} (2 s.f.)**

14E.03.2

Power = 7.85 × 10^{-4} W

Equation for Photon energy:

$$E = \frac{hc}{\lambda}$$

Photon energy = (6.63 × 10^{-34} J s × 3.00 × 10^{8} m s^{-1}) ÷ 510 × 10^{-9} m = 3.9 × 10^{-19} J

Photons per second = *P/E* = 7.85 × 10^{-4} W ÷ 3.9 × 10^{-19} J = **2.01 × 10^{15} s^{-1}**

Photocurrent = 58.5 × 10^{-6} A

Electron-hole pairs per second = 58.5 × 10^{-6} A ÷ 1.6 × 10^{-19} C = 3.65 × 10^{14} s^{-1}.

Quantum efficiency, η = (3.65 × 10^{14} s^{-1} ÷ 2.01 × 10^{15} s^{-1}) × 100 %

= 18.2 % = **18 % (2 s.f.)**

(Good revision, what?)

14 E.03.3

Total time for the signal = 2.5 ns + 1.0 ns + 2.5 ns = 6 ns

Frequency = 1 ÷ 6.0 × 10^{-9} s = **<u>1.67 × 10^{8} Hz</u>**

14E.03.4

From the graph:

The relative sensitivity at 1000 nm is 90 %

Quantum efficiency = 90 % × 0.7 = **<u>63 %</u>**

14E.03.5

Intensity = 2500 lux ÷ 683 lux m^{2} W^{-1} = **<u>3.66 W m^{-2}</u>**

14E.03.6

Read from the graph at Intensity = 10 W m^{-2}. Remember that it's a logarithmic scale. It's about 0.3 above the 100 mA level. Log^{-1} (0.3) = 2.0.

Therefore, the current is 200 mA = 200 × 10^{-6} A.

Voltage across the resistor = 3.0 V - 0.50 V = 2.5 V

Resistance = 2.5 V ÷ 200 × 10^{-6} A = 12500 Ω = **<u>13 kΩ (to 2 s.f.)</u>**

Tutorial 14E.04

14E.04.1

The circuit is unlikely to work, because the Hall Voltage is very much less than the 2 V that is needed at the gate to turn it on.

Tutorial 14E.05

14E.05.1

10110010 = **178**

14E.05.2

We can have 4 levels: 00, 01, 10, 11. (0, 1, 2, 3)

14E.05.3

128 + 64 + 32 + 16 + 8 + 4 + 2 + 1 = 255

14E.05.4

2^{32} = 4294967296 colours = **4.294×10^9**

14E.05.5

6: binary 0110, decimal 6.

C: binary 1100, decimal 12.

13: binary 100101; decimal 19.

3F: binary 111111; decimal 63.

14E.05.6

1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0

14E.05.7

1000 1011 1011 1010 1011 1101 1010 1111 1010 1011 1000 0101
0101 0100 0101 0100 0011 0101 0001 0101 0101 0111

8 11 11 10 11 13 10 15 10 11 8 5 5 4 5 4 3
5 1 5 5 7

14E.05.8

T = 1 ÷ 512 Hz = 1.95 × 10^{-3} s.

There are 22 samples. Therefore, the sampling time = 1.95 × 10^{-3} s ÷ 22 = 8.88 × 10^{-5} s

Sampling frequency = 1 ÷ 8.88 × 10^{-5} s = 11264 Hz = **11 kHz (2.s.f.)**

Tutorial 14E.06

14E.06.1

The voltage across the inductor leads the current.

The voltage across the capacitor lags the current.

14E.06.2

The inductor has no resistance (in theory).

The capacitor has infinite resistance.

14E.06.3

a.

The voltage as it's not part of the formula.

b. Formula:

$$f = \frac{1}{2\pi\sqrt{(LC)}}$$

$f^2 = (4 \times \pi^2 \times 50 \times 10^{-3}\text{ H} \times 330 \times 10^{-6}\text{ F})^{-1} = 1540\text{ s}^{-2}$

f = **39 Hz**

14E.06.4

Formula:

$$f^2 = \frac{1}{4\pi^2 LC}$$

Rearranging:

$$C = \frac{1}{4\pi^2 L f^2}$$

$C = [4 \times \pi^2 \times 0.141 \times 10^{-3}\text{ H} \times (25000\text{ Hz})^2]^{-1} = 2.87 \times 10^{-7}\text{ F}$ = **0.29 μF**

14E.06.5

Frequency range is 1850 Hz to 1950 Hz.

Bandwidth = 1950 Hz - 1850 Hz = 100 Hz

f_0 = 1900 Hz

$$Q = \frac{f_0}{f_B}$$

Q = 1900 Hz ÷ 100 Hz = **19**

Tutorial 14E.07

14E.07.1

$$V_{out} = \text{gain} \times V_{in}$$

$$V_{out} = 200 \times 75 \times 10^{-3}\text{ V} = \mathbf{\underline{1.5\text{ V}}}$$

14E.07.2

$$\Delta V_{in} = V_{out} \div \text{gain}$$

$$\Delta V\text{in} = \pm\ 13.5\text{ V} \div 100\ 000 = \pm\ \mathbf{\underline{13.5 \times 10^{-4}\text{ V}}}$$

14E.07.3

The ideal op-amp:

- Infinite open loop gain.
- Infinite input impedance so that no current is drawn.
- Zero output impedance so that maximum current can be transferred to the load.
- Very wide bandwidth.

The real op-amp:

- Gain is about 200 000 in open loop.
- Current drawn is very small but is measurable.
- Output impedance is about 150 ohms.
- Bandwidth is very narrow in open loop but improved with lower gain.

14E.07.4

(a)

The voltage at both X and Y is 4.5 V

(b)

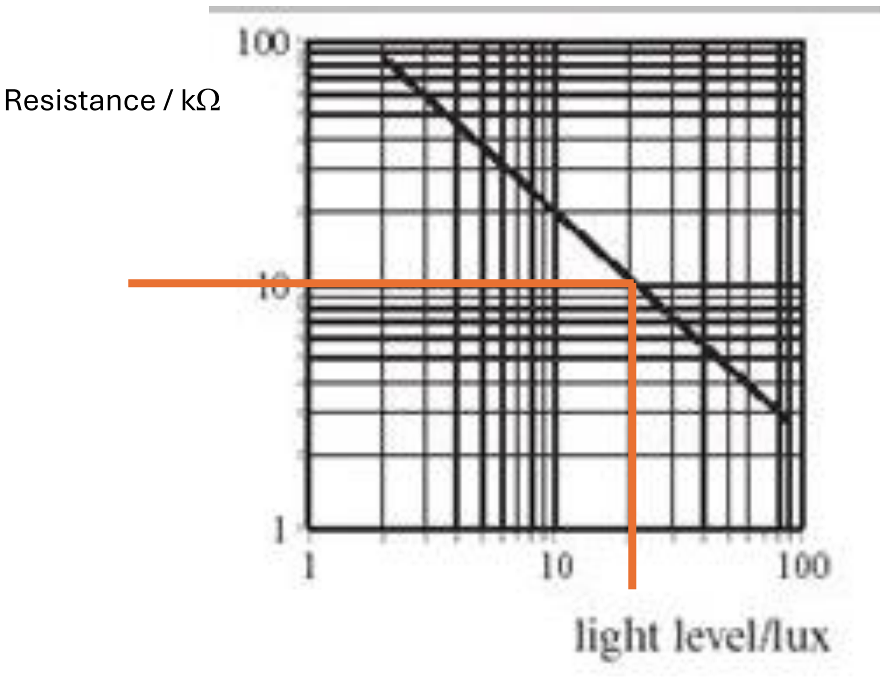
The voltage across the 10 k resistor is 4.5 V.

Since the current is the same throughout (I = 4.5 ÷ 10 000 = 0.45 mA),

the resistance of the LDR must be 10 k.

(c)

It's not very clear, but the light level is about 12 lux.



Tutorial 14E.08

14E.08.1

If the non-inverting input is at zero,

the inverting input is almost at zero,

because the difference is only 60 μV

14E.08.2

Gain = $\frac{-100000}{1000}$ = **-100** (No units for gain)

14E.08.3

The voltage difference
between the inverting and non-inverting input is very small.
Point P is connected by a wire to the – input, so no voltage drop here.

14E.08.4

V_{out} / V_{in} = gain.

Gain = 1 + 10000/100 = 1 + 100 = **101**

14E.08.5

The circuit is like an inverting amplifier.

14E.08.6

The currents in each branch add up to the total current.

14E.08.7

$V = IR$ (provided the temperature is constant).

You knew that, didn't you?

14E.08.8

The R terms cancel out

The negative sign tells us we are going up the potential hill.

14E.08.9

$$V_{out} = -R_f\left(\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}\right)$$

$$V_{out} = -2000\ \Omega \times \left(\frac{4}{1000} + \frac{5}{5000}\right) = -2000(0.004 + 0.001)$$

$$\Rightarrow V_{out} = -2000 \times 0.005 = \mathbf{-10\ V}$$

14E.08.10

$$V_{out} = -1000\left(\frac{5}{10000} + \frac{5}{20000} + \frac{0}{40000} + \frac{5}{80000}\right)$$

$$= -\left(\frac{5}{10} + \frac{5}{20} + \frac{0}{40} + \frac{5}{80}\right)$$

$$= -(0.5 + 0.25 + 0 + 0.0625)$$

$$= \mathbf{-0.8125\ V}$$

14E.08.11

$$V_{out} = (4 - 2) \times (100 \div 50) = 2 \times 2 = \mathbf{4\ V}$$

14E.08.12

$$V_{out} = (V_2 - V_1) \times (R_1/R_1) = (V_2 - V_1)$$

Tutorial 14E.09

14E.09.1

Gives a 1 if both Input A and Input B are 1	AND
Gives a 1 if either Input A or Input B is a 1	OR
Gives a 1 if either Input A or Input B is a 1, but not both.	X-OR
Gives a 0 when both Input A and Input B are a 1.	NAND

14E.09.2

A	B	C	Q
0	0	1	0
0	1	0	1
1	0	0	1
1	1	0	1

14E.09.3

A	B	C	D	Q
0	0	0	1	0
0	1	1	0	0
1	0	1	1	1
1	1	1	0	0

14E.09.4

A	B	C	D	Q
0	0	1	1	0
0	1	1	0	1
1	0	0	1	1
1	1	0	0	1

14E.09.5

C: Q = C.D

14E.09.6

D: $C = A + B$

14E.09.7

A: $D = \bar{B}$

14E.09.8

D: Q = NOT B.(A+B)

14E.09.9

Q = NOT B.A

Tutorial 14E.10

14E.10.1

A monostable produces a single pulse.
An astable produces a continuous train of pulses.

14E.10.2

When the clock pulse goes from 1 to 0, the bistable changes state.

14E.10.3

The output is not predictable.

This could cause undesirable effects in subsequent circuits.

14E.10.4

The output changes with the one of the changing S input
while The R input is high.

14E.10.5

The number of output pulses is half the number of input pulses.

Each output pulse has a period that is double the input pulse period.

14E.10.6

The output trace B has a period that is four times longer than the input pulse.

14E.10.7

The output trace changes state on the rising edge. It is rising edge triggered.

14E.10.8

The maximum number is 15 (binary 1111).

On the next pulse the counter will reset to 0000.

14E.10.9

The counter counts up to binary 1100 (decimal 12).
Since Q1 and Q3 are 1, the output of the AND gate is 1
and that makes the reset line 1,
knocking the counter back to zero.

Tutorial 14E.11

14E.11.1

Equations:

$$t_H = 0.7(R_A + R_B)C$$

$$t_L = 0.7\,R_B C$$

t_H = 0.7 × (100 000 Ω + 47 000 Ω) × 10 × 10^{-6} F = **1.03 s**

t_L = 0.7 × 47000 Ω × 10 × 10^{-6} F = **0.329 s**

14E.11.2

Equations:

$$t_H = 0.7(R_A + R_B)C$$

$$t_L = 0.7\,R_B C$$

$$f = \frac{1.4}{(R_1 + 2R_2)C}$$

(a)

f = 1.4 ÷ [10 000 Ω + (2 × 20 000 Ω)] × 2.2 × 10^{-6} F = 12.7 Hz = **13 Hz (2 s.f.)**

(b)

t_H = 0.7 × (10 000 Ω + 20 000 Ω) × 2.2 × 10^{-6} F = **0.0462 s**

t_L = 0.7 × 20000 Ω × 2.2 × 10^{-6} F = **0.0308 s**

mark to space ratio = 0.0462 s ÷ 0.0308 s = **1.5**

(c)

It is not a square wave has t_H is 1.5 times longer than t_L.

14E.11.3

Equation:

$$f = 1/T = 1/2.2\ RC$$

T = 2.2 × 150 × 10^3 Ω × 20 × 10^{-6} F = **6.6 s**

Tutorial 14E.12

14E.12.1

$\lambda = c/f = 3.0 \times 10^8$ m s^{-1} ÷ 600 × 10^6 Hz = **0.50 m**

14E.12.2

$\lambda = c/f = 3.0 \times 10^8$ m s^{-1} ÷ 600 × 10^6 Hz = 0.50 m

It will not work. The reflector must be at least 1 wavelength in diameter.

Waves will not diffract if they pass an aperture that is less than 1 wavelength.

So, such a reflector will not cause any deviation of the waves.

14E.12.3

(a)

$\lambda = 1.50$ m × 2 = 3.0 m

$f = 3.0 \times 10^8$ m s^{-1} ÷ 3.0 m = 1.0×10^8 Hz = **100 MHz**.

(b)

Area = $\pi \times d \times l = \pi \times 1.0 \times 10^{-2}$ m × 1.50 m = 0.0471 m^2

$I = 3.5$ W ÷ 0.0471 m^2 = **74.3 W m^{-2}**

(c)

Formula:

$$\frac{I_1}{I_2} = \left(\frac{x_2}{x_1}\right)^2$$

Remember that x_1 is the **radius**, 0.5×10^{-2} m.

$I_2 = 74.272$ W m^{-2} × $(0.5 \times 10^{-2}$ m ÷ 2500 m$)^2$ = 2.97×10^{-10} W m^{-3}

= **3.0×10^{-10} W m^{-3}** (2 s.f.)

14E.12.4

Formula:

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

A = π × (0.5 × 10^{-3} m)^2 = 7.854 × 10^{-7} m^2

R = (1.68 × 10^{-8} Ω m × 10 000 m) ÷ 7.854 × 10^{-7} m^2 = **214 W**

14E.12.5

Formula:

$$\mathcal{E} = -N\frac{\Delta\Phi}{\Delta t}$$

Δt = (100 × 10^6 s^{-1}) = 1.0 × 10^{-8} s

E = -1 × (2.5 × 10^{-9} T ÷ 1.0 × 10^{-8} s) = **- 0.25 V**

(Remember the minus sign!)

14E.12.6

(a)

Capacitance = 25 m × 56 × 10^{-12} F m^{-1} = **1.4 × 10^{-9} F**

(b) Formula:

$$X_C = \frac{1}{2\pi f C}$$

X_C = 1 ÷ (2 × π × 105 × 10^6 Hz × 1.4 × 10^{-9} F) = **1.1 Ω** (Unit for reactance is ohm)

14E.12.7

(a)

$A = 6.0 \times 10^{-6}$ m $\times 2.0 \times 10^{-6}$ m $= 12 \times 10^{-12}$ m^2

$I_0 = 250 \times 10^{-6}$ W $\div 12 \times 10^{-12}$ m^2 = **<u>20.8×10^6 W m^{-2}.</u>**

(b)

Formula:

$$I = I_0 e^{-\mu x}$$

$I = 20.8 \times 10^6$ W m^{-2} $\times$ e$^{-1.2\text{^}-5 \text{ m}\text{^}-1 \times 50000 \text{ m}}$

$I = 20.8 \times 10^6$ W m^{-2} $\times$ e$^{-0.6}$ = **<u>11.4×10^6 W m^{-2}</u>**

(c)

Efficiency = (11.4×10^6 W m^{-2} $\div$ 20.8×10^6 W m^{-2}) $\times$ 100 % = **<u>55 %</u>**

14E.12.8

Formula:

$$\sin\theta = 1.22\frac{\lambda}{d}$$

$\lambda = 3.0 \times 10^8$ m s^{-1} $\div$ 7.5×10^9 Hz = 0.04 m

$\sin\theta = 1.22 \times (0.04 \text{ m} \div 2.0) = 0.0244$

$\theta = 1.40^\circ$

$r = 36 \times 10^6$ m $\times$ tan 1.40° = 880×10^3 m = **<u>880 km</u>**

Tutorial 14E.13

14E.13.1

People.

They might not be attentive,

or they might do the wrong thing, especially if they are not fully competent.

14E.13.2

(a)

2^8 = 256 different states.

(b)

01000000 represents 128 (which is half way between 0 and 255),

while 00001111 represents 15.

It would not make much sense to have just 15 positions for nose up

and 240 positions for nose down.

14E.13.3

The pitch system is at neutral.

14E.13.4

The aeroplane is banking to the right.

The rudder is yawing to the right.

Then the ailerons then go back to neutral, and the rudder goes to neutral.

The pitch goes from neutral to nose up.

The aeroplane starts to climb.

The trim has not been changed.

14E.13.5

(a)

Transmission Efficiency = (14 ÷ 15) × 100 % = 93 %

(b) Look at the TDM:

Block 1: A1; B1; C1; D1.

Block 2: A2; B2; C2; D2.

Block 3: A3; B3: Blank; D3.

Block 4: A4; B4; Blank; Blank.

Block 5: A5; Blank; Blank; Blank.

In this case, the transmission efficiency = (14 ÷ 20) × 100 % = 70 %.

Tutorial 14E.14

14E.14.1

Lowest frequency = 101.1×10^6 Hz - 100×10^3 Hz = **101.0×10^6 Hz**

Highest frequency = 101.1×10^6 Hz + 100×10^3 Hz = **101.2×10^6 Hz**

14E.14.2

(a)

Audio bandwidth = 8200 Hz ÷ 2 = **4100 Hz**

(b)

Carrier frequency = 3×10^8 m s^{-1} ÷ 247 m = **1.215×10^6 Hz**

(c)

Minimum frequency = 1.215×10^6 Hz - 4100 Hz = **1.211×10^6 Hz**

Maximum frequency = 1.215×10^6 Hz + 4100 Hz = **1.219×10^6 Hz**

14E.14.3

(a) Formula:

$$\text{Bandwidth} = 2\Delta f + f_M$$

Bandwidth = (2 × 75 000 Hz) + 15 000 Hz = 150 000 Hz + 15000 Hz = **165 000 Hz**

(= 165 kHz = 0.165 MHz)

(b)

Minimum frequency = 93.5×10^6 Hz - 82500 Hz = **93.42×10^6 Hz**

Maximum frequency = 93.5×10^6 Hz + 82500 Hz = **93.58×10^6 Hz**

14E.14.4

(a)

The frequency range for communication radios is 137 - 118 = 19 MHz

No of channels = 19×10^6 Hz ÷ 8.33×10^3 = **2280**

(2280.9, but you can't have 0.9 of a radio channel)

(b)

$$\text{Bandwidth} = 2\Delta f + f_M$$

Bandwidth = (2 × 2000 Hz) + 500 Hz = **4500 Hz**

14E.14.5

(a)

$$\text{Bandwidth} = 2\Delta f + f_M$$

Bandwidth = (2 × 100 000 Hz) + 50 000 Hz = **250 000 Hz**

(b)

Capacity = 2 × 250 000 Hz = 500 000 bits per second.

This is **62500** bytes per second (= 62.5 kB s^{-1})

14E.14.6

1.5 GB = 12×10^9 bits

10 h = 36000 s

Rate of data transfer = **330 kb s^{-1}**

(Not very fast)