Introduction to electronics

Unit 5

Contents

Section	Learning competencies
5.1 Vacuum tube devices (page 156)	• Define the term electronics. • State the importance of electronics in your daily life. • State what is meant by thermionic emission. • Describe the behaviour of vacuum tubes. • Describe the function of a cathode ray tube. • Describe the uses of a cathode ray tube. • Represent both d.c. and a.c. on current–time or voltage–time graphs. • Use the current–time or voltage–time graphs to find the period and frequency of alternating currents or voltages.
5.2 Conductors, semiconductors and insulators (page 163)	• Distinguish between conductors, semiconductors and insulators. • Give examples of semiconductor elements. • Distinguish between instrinsic and extrinsic semiconductors. • Describe a semiconductor in terms of charge carriers and resistance.
5.3 Semiconductors (impurities, doping) (page 166)	• Explain doping to produce the two types of semiconductors. • Indentify semiconductors as p-type and n-type. • Describe the mode of conduction by the majority and minority carriers. • Define the term diode and show its circuit symbol. • Draw a current versus voltage characteristics (graph) to show the behaviour of p-n junction. • Describe how a semiconductor diode can be used in a half-wave rectification. • Sketch voltage time graphs to compute the variation of voltage with time before and after rectification. • Distinguish between direct current from batteries and rectified alternating current by consideration of their voltage–time graphs. • Show the circuit symbols of semiconductor devices such as thermistor, LED, LDR and transistors.
5.4 Transistors (p-n-p, n-p-n) (page 176)	• Distinguish between p-n-p and n-p-n transistors. • Identify the base, emitter and collector of a transistor. • Use the following terms correctly: forward biased and reverse biased. • Describe the behaviour of semiconductor devices such as thermistor, LED, LDR, photodiode and transistors. • Use the circuit symbols for the gates. • Draw the truth tables for the different logic gates and for a combination of logic gates. • Explain the action of logic gates: NOT, OR, AND, NOR, NAND.

Electronics is the study and design of systems that use the flow of electrons through such components as semiconductors, resistors and capacitors. Many of the concepts which we met in Unit 3 (Current Electricity) are also relevant to electronic circuits, but several of the components used in electronic circuits have very specialised characteristics.

Activity 5.1: Observing objects

Observe the objects around you – at home, at school, when travelling – and note down how many are dependent on electronics.

At school, display the objects you have considered on a concept map and discuss them in the class.

KEY WORDS

thermionic emission *the escape of electrons from a heated metal surface*

cathode ray oscilloscope *electronic test equipment that provides visual images of electrical signals and oscillations*

5.1 Vacuum tube devices

By the end of this section you should be able to:

- Define the term electronics.
- State the importance of electronics in your daily life.
- State what is meant by thermionic emission.
- Describe the behaviour of vacuum tubes.
- Describe the function of a cathode ray tube.
- Describe the uses of a cathode ray tube.
- Represent both d.c. and a.c. on current–time or voltage–time graphs.
- Use the current–time or voltage–time graphs to find the period and frequency of alternating currents or voltages.

Thermionic emission

In Unit 3 we learned about electrons (conduction electrons) that are free to move around within a metal at random, even without the application of a voltage to cause a general drift in one direction. If the metal is heated up, they move faster, and some of the more energetic electrons can then escape from the surface of the metal in a way very similar to evaporation of the faster molecules from the surface of a liquid.

The effect is known as **thermionic emission**. Its rate is negligible at ordinary temperatures, and does not become significant until high temperatures are reached.

Thermionic emission provides a controllable supply of electrons in a vacuum. This is the basis of the **cathode ray oscilloscope**, a piece of equipment that has important uses in studying and displaying electrical signals or oscillations.

Vacuum tubes

The thermionic diode

The thermionic diode, a vacuum tube, is now almost obsolete, its job is now being performed by the semiconductor diode, which we shall study later in this unit. Nevertheless it is still useful to take a quick look at how it worked. The name **diode** refers to the fact that the device has two electrodes – an anode and a cathode. A metal plate (the cathode of Figure 5.1) was heated so as to emit electrons. This was done electrically by placing the cathode in front of what amounted to an electric heater, as shown. This heater was a tungsten wire like the filament of a light bulb, being run at red heat. It operated off a few volts, either a.c or d.c. – the 6 V a.c. in the drawing was typical.

KEY WORDS

diode *an electrical component with two electrodes, used for rectification*

rectification *converting alternating current to direct current*

electron gun *an electrical component producing a beam of electrons moving through a vacuum at high speed*

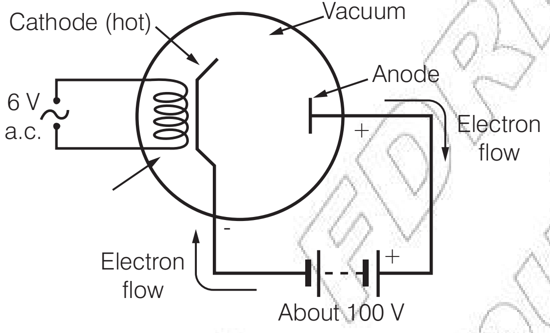


Figure 5.1 *Thermionic diode.*

The electrons 'boiled off' from the hot cathode and were attracted to the anode – a cold metal plate which was commonly at about +100 V with respect to the cathode. In this way the circuit was completed, and current flowed.

The main use of the device was for **rectification** – to obtain a d.c. current from an alternating voltage. It behaved like a valve in a water pipe and permitted one-way flow only. Current could pass across it in the direction shown in Figure 5.1 but if the polarity of the main battery was reversed, the cold anode would not emit electrons into the vacuum and so current would not flow. A diode valve was wasteful of power because of the heater.

Cathode rays

Cathode rays are a beam of electrons moving through a vacuum at a high speed. They are produced by an **electron gun**, which is a vacuum tube device. A simple example of one is shown in Figure 5.2. Practical examples may include arrangements for controlling the number of electrons in the beam and for focusing the beam to counteract the natural tendency of the particles to spread out. Nevertheless the drawing shows the principles.

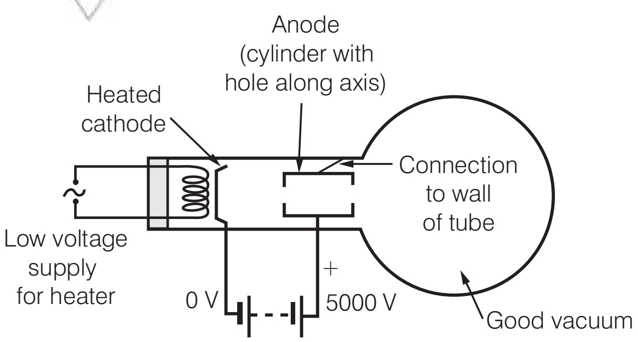


Figure 5.2 *Electron gun.*

As in the thermionic diode, electrons are released from the heated cathode by thermionic emission, and are attracted towards the anode.

Unlike the diode, the voltage on the anode may be as high as 5000 V relative to the cathode. The electrons are pulled towards the centre of the anode, accelerating to a very high speed – of the order of a tenth that of light – shooting straight through the hole along the axis of the anode.

X-ray tube

The X-ray machine (Figure 5.3) is also a vacuum tube device. Electrons are released from the cathode by thermionic emission and are then accelerated through a p.d. of the order of 100 kV so as to hit the anode at an extremely high speed. When such fast-moving cathode rays are suddenly stopped, X-rays are produced.

Most of the beam's energy is released as heat rather than X-rays, so the anode gets very hot. To minimise this it is made of a large block of copper to conduct the heat away – the other end being equipped with cooling fins or having cold liquid pumped round it. Even so, the working tip of the anode is made from tungsten because of its high melting point.

In the early days

Early workers in this area of science were puzzled by this strange radiation which could pass through wood and flesh and cause a photographic plate to blacken; they were unaware of its dangers, and did not realise that it was a previously undetected part of the family of electromagnetic waves (see Unit 6). The name they chose – X-rays – indicates clearly the extent of their knowledge at the time!

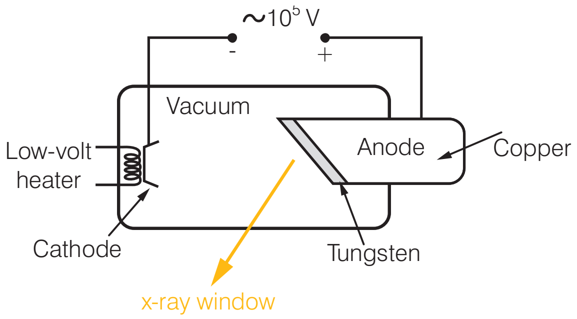


__Figure 5.3__ X-ray tube.

Cathode ray oscilloscope (CRO)

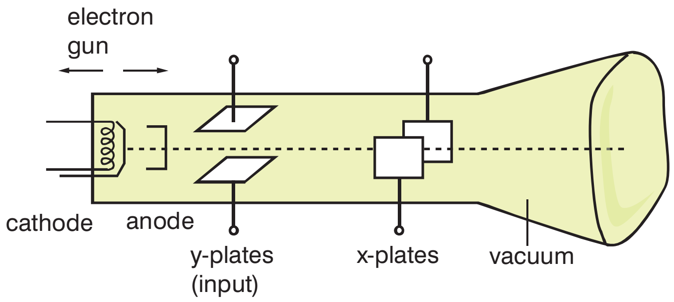


__Figure 5.4__ The cathode ray oscilloscope.

Another very useful vacuum tube device is the cathode ray oscilloscope (Figure 5.4).

In a cathode ray oscilloscope the beam of electrons produces a spot on a fluorescent screen at the end of the tube. Before reaching the screen it passes between two sets of deflecting plates – one pair to deflect it in the y-direction and the other in the x-direction.

Any rapidly varying voltage you wish to examine is connected across the y-plates. This may be the output from a microphone, for instance, or from an a.c. power supply. This voltage will cause the spot to move quickly up and down the screen (the deflection being proportional to the voltage), so a vertical line is visible. Because of the very low mass of the electrons which make up the beam, they can respond to very rapidly changing voltages.

KEY WORDS

direct current (d.c.) *an electric current that flows in a constant direction*

gain control *a device adjusting the amount of beam deflection in a cathode ray oscilloscope*

Some uses of the CRO

Direct current

The CRO can be used as a voltmeter and will represent the voltage of a source of **direct current** as a stationary spot of light on the screen. It can be calibrated by using a known voltage across the y plates so that the value of an unknown voltage can be measured.

Sensitivity

The sensitivity – the size of the deflection caused by the voltage applied across the y plates of a CRO – can be adjusted using the **gain control**. For example, if the sensitivity is set to 3 V per cm and the spot is deflected by 2 cm by an unknown p.d., the value of the p.d. must be 2 × 3 = 6 volts.

Worked example 5.1

A 1.5 V cell is connected to the y-plates of a CRO and the gain control adjusted so that the trace is 1 cm above the zero line (Figure 5.5a). The cell is removed, an unknown p.d. applied to the y-plates and a new trace is seen on the screen of the CRO (Figure 5.5b). What is the size of the unknown p.d.?

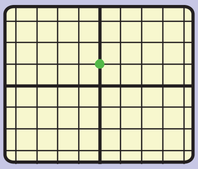


Figure 5.5a

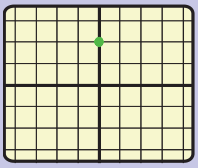


Figure 5.5b

A 1.5 V cell is represented by a deflection of 1 cm. The unknown p.d. is represented in Figure 5.4b by a 2 cm deflection. The unknown p.d. is therefore 3 V.

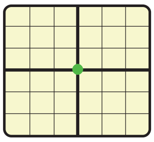


***Figure 5.6a** Time base off.*

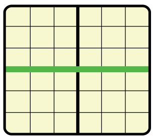


***Figure 5.6b** Time base on.*

Time base

To discover more about a signal being studied a voltage called a time base is applied to the x-plates. This pulls the spot across the screen from left to right at a steady rate; it then flies back to the left-hand side and repeats its movement. If the time base is applied to a signal made by a direct current, the trace changes from a dot (Figure 5.6a) to a line (Figure 5.6b).

The time base control can be calibrated (in milliseconds per cm) to show how long the spot takes to cross each centimetre of the screen. For example, if it is set to 1 ms/cm the spot takes 1 ms (1 millisecond) to move 1 cm to the right. This can be very useful in measuring distances if, for example, traces of a sonar signal and its echo are studied.

Provided the gain control and the time base are synchronised (and the instrument does this for you), what appears on the screen is an apparently static graph of the input voltage plotted against time.

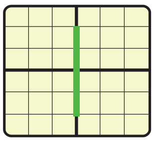


***Figure 5.7a** Time base off.*

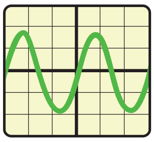


***Figure 5.7b** Time base on.*

Alternating current

The voltage of an alternating current varies between a (positive) maximum to a (negative) minimum. If the time base is switched off, this is represented as a vertical line on the screen of a CRO (Figure 5.7a). If the time base is switched on, this becomes a curve (Figure 5.7b) whose shape is known as a **sine wave**.

KEY WORDS

sine wave *a mathematical function that describes a smooth repetitive oscillation*

Finding the period and frequency of alternating currents or voltages using the CRO

In the same way that an unknown d.c. voltage can be measured using a CRO by comparing its trace with that made by a known p.d., the CRO can measure frequency by comparing a wave of unknown frequency with one of known frequency.

The known signal is applied to the CRO and the time base adjusted so that one complete wave appears on the screen (Figure 5.8a). The unknown signal is then applied in place of the known signal, without altering any of the CRO controls, and the trace on the screen is studied (Figure 5.8b).

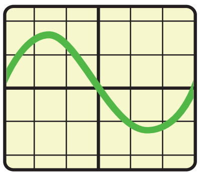


***Figure 5.8a** (50 Hz).*

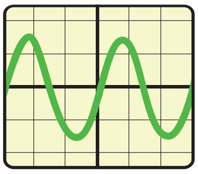


***Figure 5.8b** (unknown).*

As Figure 5.8b shows two complete waves it follows that the unknown frequency is twice that of the known signal, i.e. 2 × 50 = 100 Hz.

Worked example 5.2

Figure 5.9a shows the trace on a CRO made by a 100 Hz signal. What is the frequency of the signal in Figure 5.9b (which was made without altering any of the CRO controls after measuring the 100 Hz signal)?

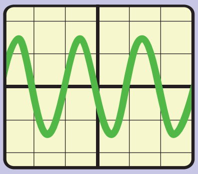


Figure 5.9a

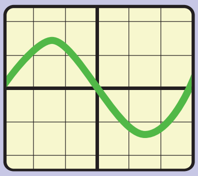


Figure 5.9b

The 100 Hz signal in Figure 5.8a shows three complete waves while the signal of unknown frequency in Figure 5.8b shows one wave. The unknown frequency in this example is therefore one-third that of the known signal

The unknown frequency = $100 \times \frac{1}{3} = 33.3$ Hz

Worked example 5.3

The sensitivity of a CRO is set as 5 V/cm, and the time base setting is 1 ms/cm. Find the peak voltage and frequency of the a.c. signal shown in Figure 5.10.

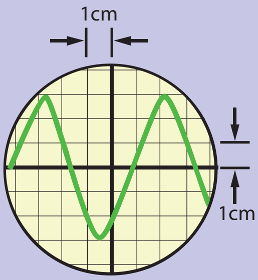


Figure 5.10

The maximum vertical displacement is 3 cm. Each centimetre of vertical displacement corresponds to 5 V.

The peak voltage of the signal shown in Figure 5.10 is therefore $5 \times 3 = 15$ V.

The time base setting is 1 ms /cm. The spot therefore takes 1 ms to move horizontally by one centimetre. One complete cycle of the signal shown in Figure 5.10 takes 5 cm. The time for one cycle is therefore 5 ms = 5×10^{-3} and there must therefore be 200 cycles each second.

The frequency of the a.c. signal is therefore 200 Hz.

TV picture tube

The older sort of television, with the big heavy tube, is similar to the CRO. The receiver sends currents through coils that are mounted just outside the tube; their magnetic fields deflect the spot of light on the screen so it moves rapidly to trace an ordered path all over the screen.

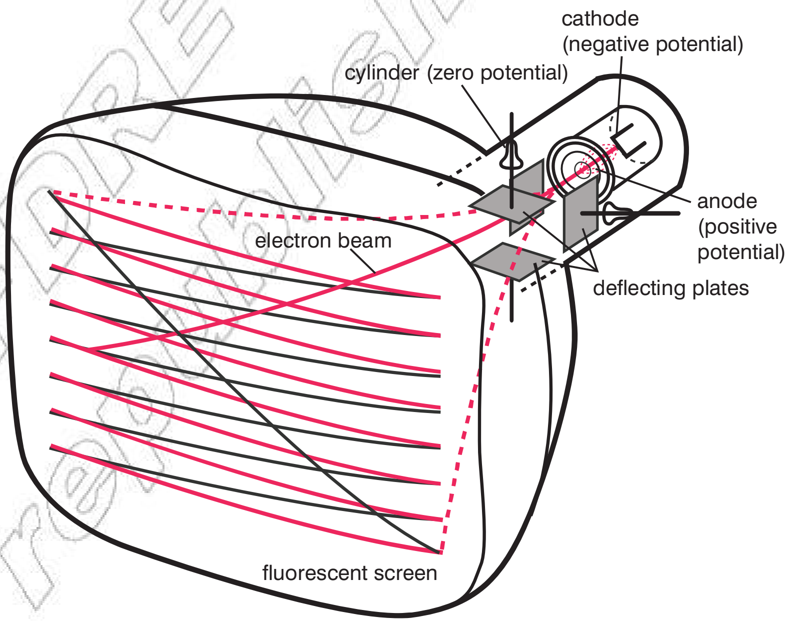


Figure 5.11a *Television tube.*

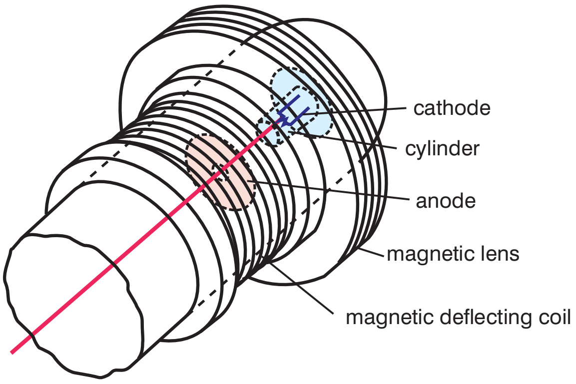


Figure 5.11b *Coils forming a magnetic lens.*

Summary

- Thermionic emission is the escape of conduction electrons from a hot metal surface.
- Thermionic electrons may be accelerated through a high voltage to produce a beam of cathode rays.
- These cathode rays convey negative charge, and may be deflected accordingly by magnetic and electric fields.

Review questions

1. In an oscilloscope tube what is the purpose of:
   a) the heater
   b) the cathode
   c) the anode
   d) the x- and y-plates?
2. Explain what it means if the time base of an oscilloscope is set at 2 ms/cm (1 ms = 0.001 s.)
3. Figure 5.12**a** shows a reading of 1.5 V. Give the values displayed in **b**, **c** and **d**.

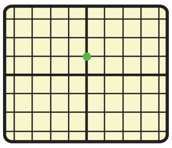


Figure 5.12a

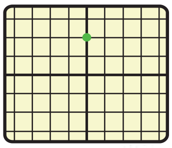


Figure 5.12b

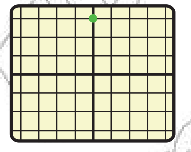


Figure 5.12c

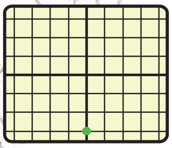


Figure 5.12d

4. Explain carefully why a cathode ray oscilloscope:
   a) acts as a voltmeter
   b) responds to changes in that voltage within less than a millionth of a second
   c) draws virtually no current from the voltage source.
5. Why is the anode of the X-ray tube drawn in Figure 5.3 made of copper with a tungsten tip?
6. Describe three different uses for X-rays. Say what precautions should be taken while using them, and explain why.

5.2 Conductors, semiconductors and insulators

By the end of this section you should be able to:

- Distinguish between conductors, semiconductors and insulators.
- Give examples of semiconductor elements.
- Distinguish between instrinsic and extrinsic semiconductors.
- Describe a semiconductor in terms of charge carriers and resistance.

Conductors, semiconductors and insulators

We saw in Unit 3 that materials can be divided into three classes:

- Insulators, such as glass and plastic, which do not conduct electricity because every electron in them is tightly bound to its parent atom.

KEY WORDS

intrinsic semiconductor *a pure semiconductor not containing any dopant*

hole *the lack of an electron at a position where one could exist in an atomic lattice*

intrinsic conduction *electrons and holes in a semiconductor moving in opposite directions when an e.m.f. is applied*

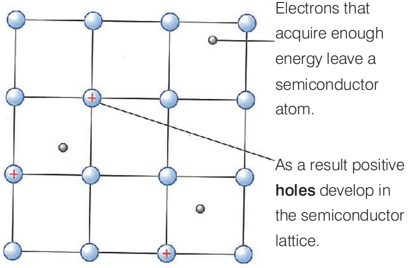


Figure 5.13a

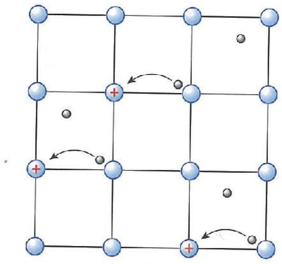


Figure 5.13b

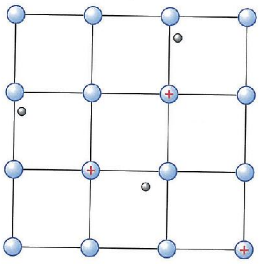


Figure 5.13c

- Conductors such as metals. All the electrons in the inner shells are still tightly bound to their atoms. However, those electrons in the outermost shell of every atom are free to move within the metal. We describe them as being in the conduction band.
- Semiconductors. Just a few of the outermost electrons have enough energy to be in the conduction band (that is, to break free from their parent atom), but this number rises as the material becomes hotter. Silicon, germanium, lead sulphide, selenium and gallium arsenide are all semiconductors.

Intrinsic semiconductors

Conduction in a pure (intrinsic) semiconductor

Substances such as silicon and germanium have resistivities between those of insulators and those of conductors. These substances are known as semiconductors, and they form the basis of many of the devices that we take for granted in a technological society. Pure semiconductors are usually referred to as **intrinsic semiconductors**, since their conductivity is not affected by any external factors. As we shall see in Section 5.3, trace impurities can greatly alter the conductivity of semiconductors.

In an intrinsic semiconductor, electric current is carried by moving electrons, as it is in metals, although the number of charge carriers in silicon is perhaps a billion times fewer than in copper. However, in addition to electrons, intrinsic semiconductors can also be considered to contain moving positive charges that carry current. This can be explained as follows.

Lattice structure of the atoms in an intrinsic semiconductor

Some of the electrons in an atom of an intrinsic semiconductor are held less tightly than others. This means that in a piece of intrinsic semiconductor material at room temperature there will always be a few free electrons that have been 'shaken free' of their atoms by thermal excitation (when the material has absorbed energy from the surroundings). When an electron leaves an atom in this way, the atom becomes positively charged (Figure 5.13a). The effect of an electron leaving an atom is therefore to create a positive charge in the semiconductor lattice. This positive charge is called a **hole**.

When an electric field is applied to the semiconductor (that is, when it is connected to a source of e.m.f.) the electrons and holes move in opposite directions, and the semiconductor exhibits **intrinsic conduction**.

This happens because, under the influence of this electric field, electrons still bound to atoms in the lattice are able to move through the lattice from an atom to a nearby hole (Figure 5.13b), thus causing the hole to appear to move through the lattice (Figure 5.13c). This motion happens in the opposite direction to the motion of the electrons.

The current in a pure semiconductor consists of free electrons moving through the semiconductor lattice in one direction, with an equal number of positively charged holes moving in the other direction (see Figure 5.14).

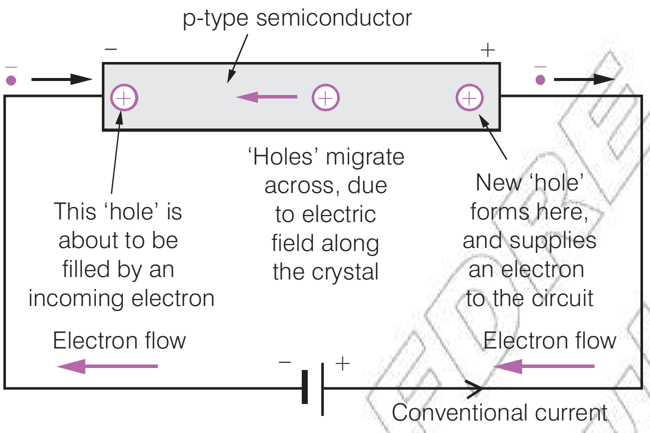


Figure 5.14

Charge carriers and resistance in a semiconductor

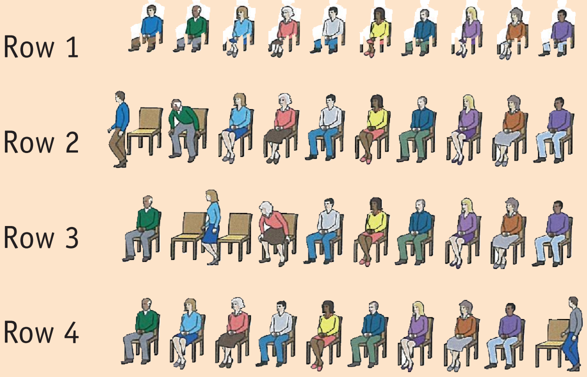
While the resistance of metallic conductors rises as they warm up, with semiconductors the resistance falls greatly as their temperature goes up.

This is because when the temperature of a semiconductor is raised, more electrons (charge carriers) have enough energy to break free. As the number of charge carriers increases, the resistance of the semiconductor material decreases and the material conducts better.

Activity 5.2: Mysteriously moving chair

Place 10 chairs in a row and ask 10 people to sit on the chairs (Figure 5.15 row 1). Ask the person at one end of the row to stand up and move away (Figure 5.15 row 2). Ask the person sitting next to the empty chair to move into that chair and then ask the remaining people to move to the empty chair as their neighbour moves up (row 3). Eventually the empty chair will be at the end of the row (a new person could take the empty chair! (row 4)).

Figure 5.15



The motion of holes through the lattice is like the motion of the empty chair in this line. As the people move from right to left, the empty chair moves from left to right.

Summary

- In a semiconductor just a few electrons have enough energy to break free from their atoms. These enable it to carry a current.
- As the electrons move in one direction, positive 'holes' appear to move in the other direction.

Review questions

1. Explain why a semiconductor can conduct at all.
2. Explain why a semiconductor conducts better when it is hot.
3. Explain what we mean by a positive 'hole' in a silicon crystal.

5.3 Semiconductors (impurities, doping)

By the end of this section you should be able to:

- Explain doping to produce the two types of semiconductors.
- Indentify semiconductors as p-type and n-type.
- Describe the mode of conduction by the majority and minority carriers.
- Define the term diode and show its circuit symbol.
- Draw a current versus voltage characteristics (graph) to show the behaviour of p-n junction.
- Describe how a semiconductor diode can be used in a half-wave rectification.
- Sketch voltage–time graphs to compute the variation of voltage with time before and after rectification.
- Distinguish between direct current from batteries and rectified alternating current by consideration of their voltage–time graphs.
- Show the circuit symbols of semiconductor devices such as thermistor, LED, LDR and transistors.

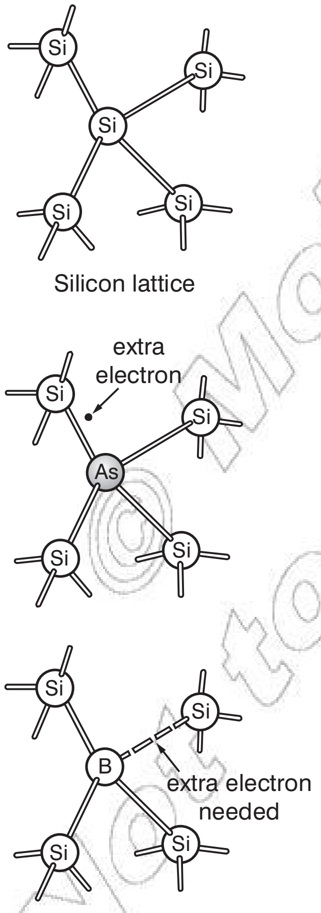


Figure 5.16 Dopants in a silicon lattice.

Doping

The variety of uses to which semiconductor materials such as silicon can be put in devices such as transistors and diodes depends on introducing minute quantities of impurities into their atomic lattice structure in a process called **doping**. This process introduces extra charge carriers by replacing atoms in the semiconductor lattice with atoms of an impurity of similar size (this is important so as to minimise the distortion of the semiconductor lattice). This alters

KEY WORDS

doping *deliberately introducing impurities into a semiconductor to change its electrical properties*

extrinsic semiconductor *a semiconductor that has been doped* the semiconductor's conducting properties by introducing extra large carriers to the semiconductor lattice, forming what is called an **extrinsic semiconductor**.

Majority and minority carriers

Silicon is in group 4 of the Periodic Table, which means the outer shell of the neutral silicon atom contains four electrons.

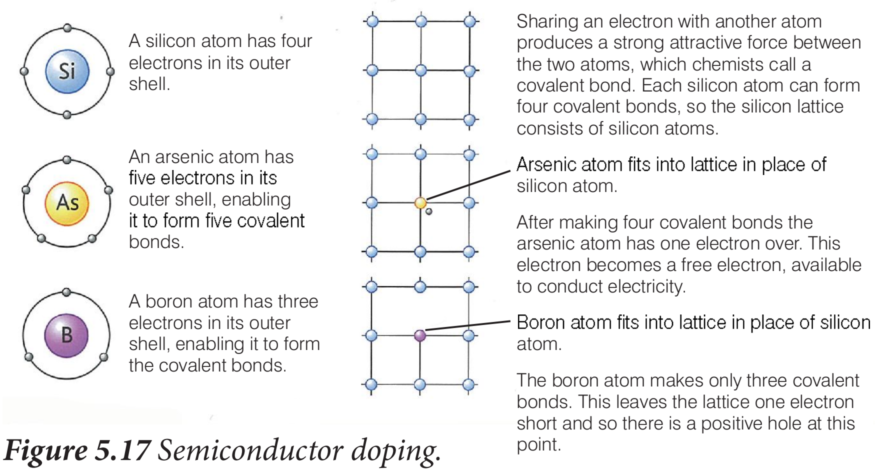
Suppose that atoms of an impurity from group 5 of the Periodic Table (neutral atoms in group 5 have five electrons in their outer shell) such as arsenic are added to the silicon lattice. The crystal as a whole remains uncharged, but these impurities help to provide 'spare' free electrons to the crystal, causing much improved conduction. Since the **majority carriers** in this type of semiconductor material are negative electrons, we describe this as an **n-type semiconductor**. (The **minority carriers** in n-type semiconductor material are holes.) Arsenic is described as a **donor** impurity, because it releases free electrons into the lattice.

Alternatively, if silicon is doped with impurity atoms from group 3 of the Periodic Table (neutral atoms in group 3 have three electrons in their outer shell) such as boron, there is again no overall charge. However, the apparent electron deficiency can shift from one atom to another and so behave like a kind of positive 'hole' which can move through the crystal and thereby carry a current through it. These 'holes' are the majority carriers in this type of semiconductor – called **p-type semiconductor** – and the minority carriers are electrons. Boron is an example of an **acceptor** impurity which traps electrons when introduced into the lattice, resulting in an increase in the number of positive holes.

Conduction in a doped semiconductor

The number of free electrons and holes can be altered dramatically by doping. For example, the addition of only one arsenic atom per million silicon atoms increases the conductivity 100 000 times.

Figure 5.17 shows how these processes occur .



A silicon atom has four electrons in its outer shell.

Sharing an electron with another atom produces a strong attractive force between the two atoms, which chemists call a covalent bond. Each silicon atom can form four covalent bonds, so the silicon lattice consists of silicon atoms.

An arsenic atom has five electrons in its outer shell, enabling it to form five covalent bonds.

Arsenic atom fits into lattice in place of silicon atom.

After making four covalent bonds the arsenic atom has one electron over. This electron becomes a free electron, available to conduct electricity.

A boron atom has three electrons in its outer shell, enabling it to form the covalent bonds.

Boron atom fits into lattice in place of silicon atom.

The boron atom makes only three covalent bonds. This leaves the lattice one electron short and so there is a positive hole at this point.

***Figure 5.17** Semiconductor doping.*

KEY WORDS

majority carriers *the type of carrier, electron or hole, that constitutes more than half the carriers in a semiconductor*

n-type semiconductor *a semiconductor in which the majority carriers are electrons, due to doping*

minority carriers *the type of carrier, electron or hole, that constitutes less than half the carriers in a semiconductor*

donor *impurity atoms added to a semiconductor which release free electrons*

p-type semiconductor *a semiconductor in which the majority carriers are holes, due to doping*

acceptor *impurity atoms added to a semiconductor which trap electrons*

junction *the region where two types of semiconducting materials touch*

Activity 5.3: Human wire

In this activity a group of students will model a doped p-type semiconductor lattice – a silicon lattice with boron impurities.

In a group of nine, each student will a model silicon atom – each person will have four rocks to represent four electrons.

One student needs to represent a boron atom – s/he will have three rocks to represent electrons and one basket to represent a hole.

Students should move to show how the electron travels in one direction and the hole in the other.

Activity 5.4: Modelling the lattice of a semiconductor

Figure 5.18 shows the three-dimensional lattice structure of a semiconductor such as silicon. Each silicon atom has four electrons available for making bonds with other atoms

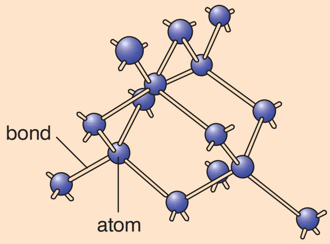


Figure 5.18 *Lattice structure.*

Using gummy sweets to represent silicon atoms and toothpicks to represent the bonds made by a silicon atom, build a lattice like that shown in Figure 5.18, where each silicon atom is joined to four others.

Change your lattice to show how doping affects the structure.

To represent an n-type semiconductor, stick an extra toothpick into some of the gummy sweets. A sweet with five toothpicks represents a donor impurity (such as arsenic) because the fifth electron is free to move into the lattice.

To represent a p-type semiconductor, remove one of the gummy sweets in the lattice and replace it with a marshmallow. The marshmallow represents an acceptor impurity (such as boron) – think of the squashy marshmallow as representing a 'hole' into which electrons would be attracted.

The p–n junction diode

Suppose a p-type semiconductor (with 'spare' positive holes) is in contact with an n-type one (with 'spare' mobile electrons). It is important to realise that before they were brought into contact both pieces of semiconductor were electrically neutral overall. At the junction where they meet, some of the n-type's electrons move, or 'fall' into the p-type's holes (Figure 5.19). This movement is known as idffusion current.

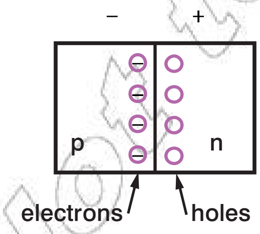


Figure 5.19 *p–n junction.*

This diffusion current causes the p-type to become slightly negative while the n-type is left equally positive, leaving a 'depletion zone' for a small distance each side of the boundary (of the order of 1 µm) – a shortage of 'holes' on one side and of free electrons the other (Figure 5.20).

This adverse voltage (of about 0.6 V for silicon) between the p-type and n-type semiconductors will prevent any more electrons from crossing the boundary, so the diode will not conduct in that direction. In the depletion zone there are no more 'holes' in the p-type and no free electrons in the n-type, so it forms a non-conducting strip which blocks all current.

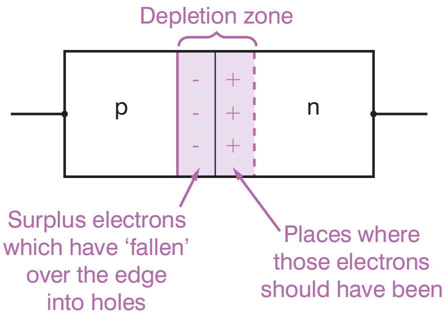


Figure 5.20 *A p–n junction.*

Forward and reverse bias

Imagine a cell connected as shown in Figure 5.21. Remember that it tries to move electrons from its negative terminal round the circuit to its positive terminal, which will only serve to make the situation at the boundary worse. After this momentary transfer, the voltage at the boundary (which is opposing that of the cell) becomes as large as that of the cell. No current will then flow. We say the junction is **reverse biased**.

If instead you apply a voltage across the diode to make the p-type positive, and make it more than 0.6 V so as to overcome the depletion-zone voltage across the junction, then the junction is said to be **forward biased** (Figure 5.22). In this direction the cell is helping to keep the transfer going. Excess electrons are being removed from the p-type, and are being fed back into the *n*-type to fill the extra 'holes'. That could go on forever – in this direction the junction will conduct.

Electrons will exit from the left-hand lead, continually leaving fresh positive 'holes' behind them. The voltage across the crystal will cause these holes to move to the right. At the junction, each hole meets a free electron and so ceases to exist. All the time, replacement electrons are flowing round the circuit and into the n-type via the right-hand lead and so the whole process is maintained.

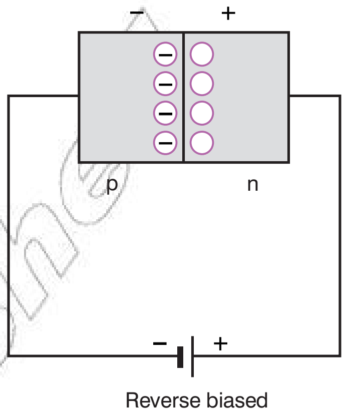


Figure 5.21 *Reverse biased, no current.*

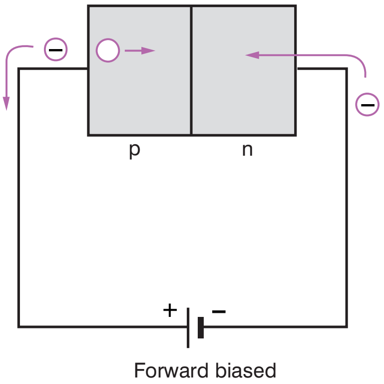


Figure 5.22 *Forward biased, current flows.*

Current–voltage characteristics of the semiconductor diode

The p–n junction as described acts as a diode: in one direction it will conduct, in the other direction it will not. The behaviour of such a diode may be illustrated by a current–voltage graph like the one shown in Figure 5.23.

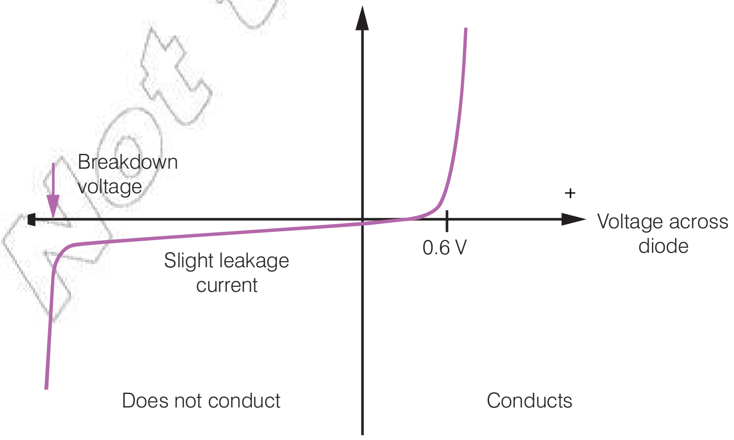


Figure 5.23

In the forward direction, silicon requires about 0.6 V before conduction will start, but after that the current is usually limited by the resistance of the rest of the circuit. In the other direction, notice there will be a tiny leakage current. If the reverse p.d. becomes too great for the device, the barrier at the junction breaks and permits a large current to flow.

KEY WORDS

reverse biased *connecting the positive terminal of a cell to the n-type region of a diode, and the negative terminal to the p-type region, preventing conduction*

forward biased *connecting the positive terminal of a cell to the p-type region of a diode, and the negative terminal to the n-type region, allowing conduction*

Demonstration: Testing the conduction of a diode

If you have the equipment, you could test the conduction of a diode by connecting a battery, a variable resistor, an ammeter and a diode in a circuit. If possible, use an LED as your diode as this should give you useful visual clues about how the experiment is progressing. You might also need a small resistor in the circuit to protect the diode. Make sure that you know in which direction the current is flowing in the circuit so that you can connect the anode and cathode of the diode correctly in the circuit. Connect a voltmeter across the diode. Use the variable resistor to change the current in the circuit and observe the voltage across the diode and the current in the circuit. You should see a series of values like those displayed in Figure 5.23.

Some semiconductor devices

Diode

A diode is an electronic component with two electrodes – an anode and a cathode – which will only allow electric current to pass through it in one direction (Figure 5.24).

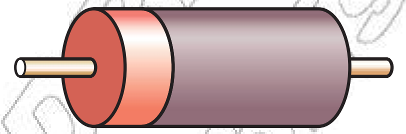


Figure 5.24a *An example of a diode (ringed end shows cathode).*

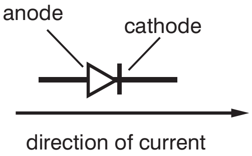


Figure 5.24b *Symbol for a diode.*

The semiconductor diode – formed from a layer of *p*-type semiconductor joined to a layer of *n*-type semiconductor material as seen above – is a very important electronic component.

LDR

A light-dependent resistor (LDR for short) conducts electricity, but in the dark it has a very high resistance. Shining light on it appears to 'unjam' it, because its resistance falls. The brighter the light, the better it conducts. The symbol for an LDR is shown in Figure 5.25b.

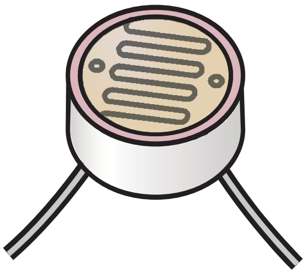


Figure 5.25a *LDR.*

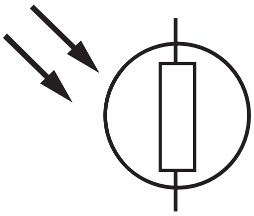


Figure 5.25b *Symbol for LDR.*

Thermistor

The thermistor shown in Figure 5.26a is a piece of semiconductor material that has a high resistance in the cold. Its resistance drops as it becomes warmer.

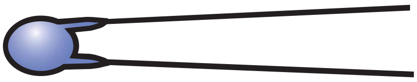


Figure 5.26a *Thermistor.*

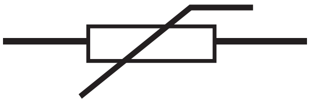


Figure 5.26b *Symbol for thermistor.*

Variable resistor

The variable resistor (Figure 5.27) is a very useful component in electronic circuits, particularly in circuits containing **transistors**.

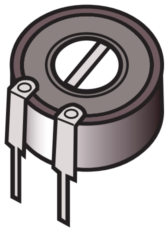


Figure 5.27a *Variable resistor.*

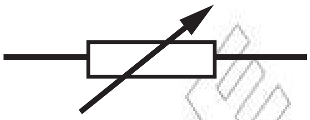


Figure 5.27b *Symbol for variable resistor.*

LED

The light emitting diode, whose symbol is shown in Figure 5.28b, can be seen in a multitude of devices. When a current is passed in the forward direction, an LED emits light. The LED is a very useful component – if there is one in a circuit, it is possible to see immediately if current is flowing. LEDs are now available in a range of colours – red, green, blue, white. White LEDs are increasingly being used in lighting; they produce light very efficiently (using relatively little energy).

Figure 5.28a *LED.*

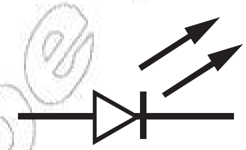


Figure 5.28b *Symbol for LED.*

Transistor

The transistor (Figure 5.29) is a very significant semiconductor component which we shall learn more of in Section 5.4.

Figure 5.29a *Transistor.*

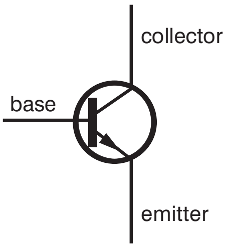


Figure 5.29b *Symbol for transistor.*

A **bipolar junction transistor** is made of three layers of doped semiconductor and it has three terminals – the base is connected to the central layer, the other two (the collector and the emitter) are each connected to one of the outer layers. Figure 5.29 shows

KEY WORDS

transistor *a semiconductor device used to amplify or switch electronic signals*

bipolar junction transistor *a device in which the current flow between two terminals, the collector and the emitter is controlled by the amount of current that flows through a third terminal, the base* an n–p–n transistor, which has a layer of p-type semiconductor sandwiched between two layers of n-type.

Photodiode

The photodiode (Figure 5.30) is a light-sensitive diode used to detect light or to measure its intensity. Photodiodes are reverse-biased so they do not conduct. Light incident on the photodiode frees a few more electrons and the device starts to conduct.

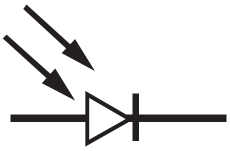


Figure 5.30 Symbol for photodiode.

Photovoltaic cell

The photovoltaic cell (Figure 5.31) is a form of photodiode.

The base layer of a photovoltaic solar cell is made of p-type semiconductor material. This is covered with a layer of n-type semiconductor material. When light strikes the junction between n- and p-types of semiconductor, electrons flow through the structure of the cell.

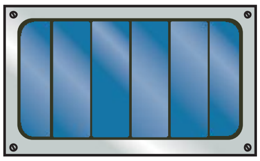


Figure 5.31a *Photovoltaic cell .*

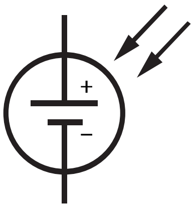


Figure 5.31b *Symbol for photovoltaic cell .*

KEY WORDS

photovoltaic cell *a cell that converts solar energy into electrical energy*

Activity 5.5: Light into power

In this experiment we study the way in which photovoltaic solar cells convert light energy to electrical energy.

When light strikes the junction between n- and p-types (see Figure 5.32), electrons flow through the crystal structure and round the circuit connected between them. This electric current can be used directly or stored in a rechargeable cell.

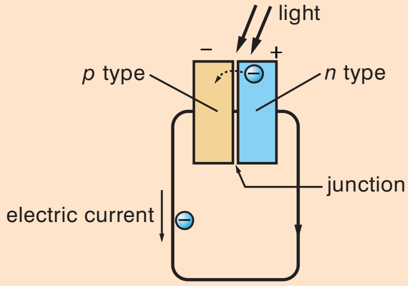


Figure 5.32

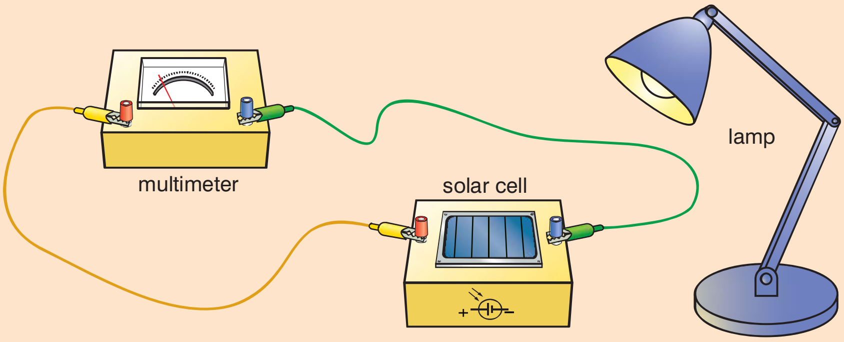


Figure 5.33 *Testing a photovoltaic cell in different light conditions.*

Connect one end of a crocodile clip lead to the positive terminal on the back of a small photovoltaic cell (0.4 V 100 mA) and the other end of this lead to the positive probe of a multimeter (see Figure 5.33). Connect the negative terminal on the back of the cell to the negative probe of the multimeter in the same way.

Switch the multimeter to read voltage – in hundredths of volts.

Cover the photovoltaic cell and record the multimeter reading in the table below (in the row 'no light, covered').

Uncover the photovoltaic cell and record the multimeter reading (observe the speed with which the multimeter reading changes when you uncover and cover the solar cell).

Move the photovoltaic cell into the sunlight and repeat the reading for sunlight (note on the table if the sun is falling directly on the solar cell).

Light source	Voltage reading
No light (covered)	
Room illumination	
Sunlight	

You will notice that the voltage is higher when more light falls on the photovoltaic cell.

You will probably see that the photovoltaic cell responds rapidly to the presence and absence of light (the voltage reading on the multimeter rises and falls abruptly).

Photovoltaic cells tend to be black because black objects are more efficient at absorbing radiation.

In order to produce a power supply for a particular appliance, a number of cells can be connected to produce the required voltage and current.

Rectification using the p–n junction diode

Using one diode

Direct current can be obtained from an alternating current generator by putting a diode in the circuit (Figure 5.34).

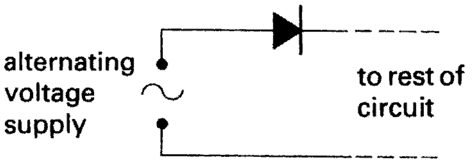


Figure 5.34 *A diode in a circuit with a supply of alternating current.*

KEY WORDS

capacitor *a device for storing electric charge, consisting of two conductors separated by a dielectric*

The diode allows the current to flow one way, but on the other half of the cycle the current cannot flow back again through the diode. The resulting current is shown in graph form in Figure 5.35. We call this half-wave rectification.

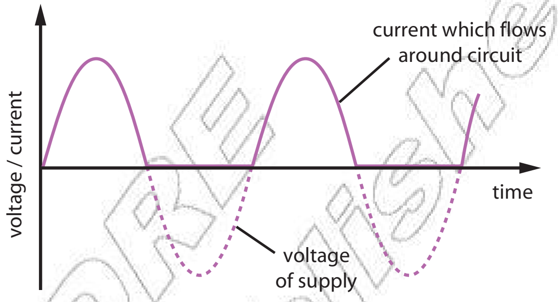


Figure 5.35 *Graph showing current flowing round circuit.*

It is direct current, because the charges always flow in the same direction and so will make progress round the circuit. However, it is not a steady smooth flow of direct current such as that obtained from a battery. Instead the charges move forward in a series of 'spurts'.

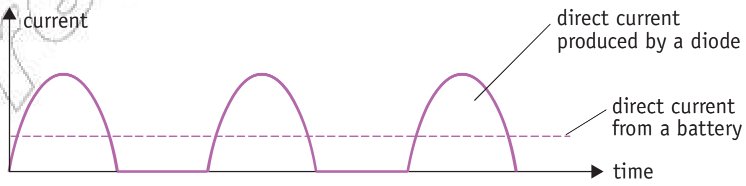


Figure 5.36 *Irregular direct current produced by diode compared with current from battery.*

Using a diode and a capacitor

It is possible to use a **capacitor** to help to smooth the fluctuations in this current. As you saw in Unit 2, a capacitor stores charge, and can release it later. The capacitor is connected across the terminals inside the casing of the power supply (Figure 5.37).

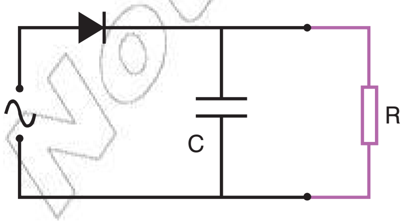


Figure 5.37

The capacitor is filled up to the peak voltage of the supply, then as the power supply's voltage drops to zero we can think of the capacitor as feeding the outside circuit. It has to keep doing this until the supply voltage next peaks and the capacitor is once more filled up (Figure 5.38).

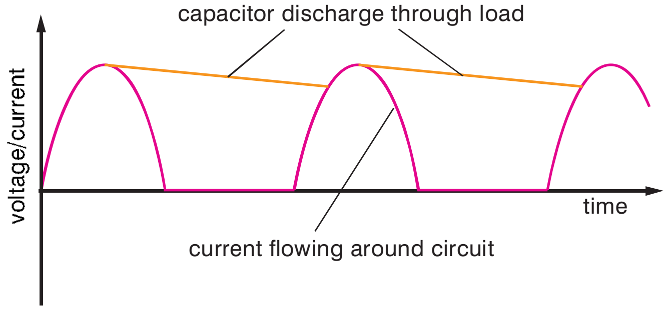


Figure 5.38

If the supply is a 50 Hz one, this will happen every $\frac{1}{50}$ s (0.02 s). The capacitor will be effective only if the time constant CR is large compared to 0.02 s, which will mean that it has emptied very little over that length of time. We often use an electrolytic capacitor as the smoothing capacitor, sometimes 1000 μF or even larger. An electrolytic capacitor is rather leaky, but since it gets refilled 50 times a second that hardly matters.

Full wave rectification

An arrangement made from four diodes is known as a **bridge rectifier**. To understand how it works needs some thought from you, so you are invited to work through it yourself. In Figure 5.39, the dotted line is the outside casing of a d.c. power pack. Inside are an a.c. power supply marked AB (obtained from the mains via a transformer) and four diodes labelled C, D, E and F. G and H are the terminals that connect it to an outside circuit, represented by resistor R.

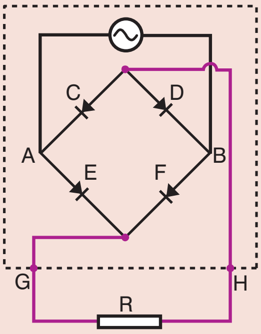


Figure 5.39

Remember that a current can flow only if a continuous circuit is provided from one terminal of power supply AB right round to the other. Check that at a moment when A is positive and B negative, conventional current has such a route through AEGRHDB. Now find the complete circuit half a cycle later, when B is the positive terminal and the conventional current starts from there. What do you notice about the direction in which the current passes through R on each half of the alternating cycle? This is known as full-wave rectification, and has the form shown in Figure 5.40.

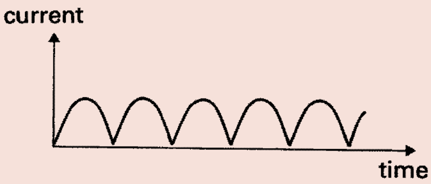


Figure 5.40

Summary

- An extrinsic conductor has its crystal doped with just a few other atoms, which can help it to conduct still better.
- A p–n junction acts as a diode, and will conduct in one direction only.
- A semiconductor diode can produce half-wave rectification from an alternating supply.

KEY WORDS

bridge rectifier *an arrangement of four diodes which produce full-wave rectification of an alternating current*

Review questions

1. An arsenic atom is about the same size as a silicon atom. Silicon has four electrons in its outer shell, arsenic has five.
    a) Explain how these two facts enable us to dope a silicon crystal with small numbers of arsenic atoms, and this makes it conduct much better.
    b) Why do we call it an n-type semiconductor?

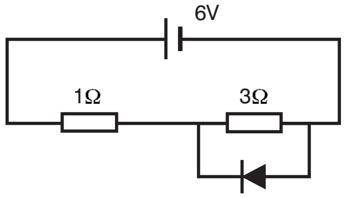


Figure 5.41

2. a) What size current will be drawn from the battery in Figure 5.41? Give your reasoning.

   b) If the battery is turned round, what size current will it now supply? Explain.

   (Assume the diode is an ideal one, so it requires zero voltage in the forward direction and possesses infinite resistance in the reverse direction.)

3. Explain in your own words why a diode valve acts as a rectifier.

4. A p-type material is brought into contact with an n-type material. Explain in your own words:

   a) Why both materials are initially uncharged.

   b) Why some free electrons from the n-type 'fall' over the boundary.

   c) Why a voltage appears across the junction, the p-type being negative.

   d) What we mean by a depletion zone, and why this will not conduct.

5. For your own notes make a table with four columns. They should be headed diode, light emitting diode (LED), light-dependent resistor (LDR) and thermistor. Under each heading give its circuit symbol and then give a brief description of how the component behaves when it is part of an electrical circuit.

5.4 Transistors (p–n–p, n–p–n)

By the end of this section you should be able to:

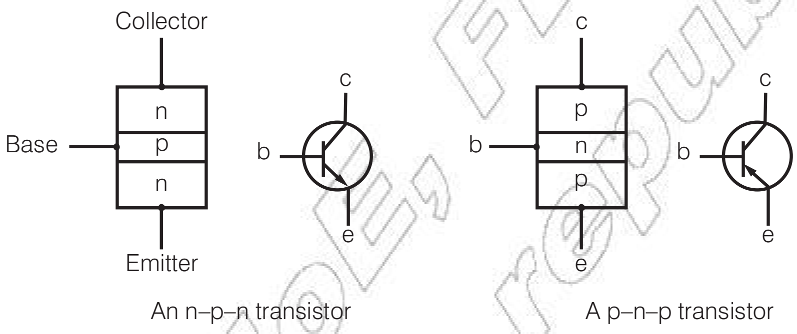
- Distinguish between p–n–p and n–p–n transistors.
- Identify the base, emitter and collector of a transistor.
- Use the following terms correctly: forward biased and reverse biased.
- Describe the behaviour of semiconductor devices such as thermistor, LED, LDR, photodiode and transistors.
- Use the circuit symbols for the gates.
- Draw the truth tables for the different logic gates and for a combination of logic gates.
- Explain the action of logic gates: NOT, OR, AND, NOR, NAND.

The bipolar junction transistor

Transistors use the input of relatively small signals to control circuits carrying large currents. This makes them very important as switches and amplifiers.

As we saw in Section 5.3, the bipolar junction transistor is a three-layer semiconductor device. The outer layers can be n-type materials with a p-type layer in the centre (in which case it is described as an n–p–n transistor), or it may be a p–n–p transistor with the layers the other way round (Figure 5.42).

A transistor of this type has three terminals – one to each semiconductor layer. The connection to the central layer is known as the **base**. The outer two are called the **collector** and the **emitter**.



An n–p–n transistor

A p–n–p transistor

***Figure 5.42** Transistor layouts.*

The arrow on the circuit symbol indicates the way conventional current should flow through it. The term ‘bipolar’ means that both negative electrons and positive ‘holes’ play a part in conduction through the transistor, but it should be remembered that in the connecting leads the current will actually be a flow of electrons in the opposite direction to the arrow.

Such an arrangement is sometimes described as two diodes back to back. We have already seen that the diode (a single layer of n-type semiconductor joined to a layer of p-type semiconductor) only allows current to flow in one direction, when the p-type layer is connected to a positive voltage. It might therefore seem that two diodes back to back would never allow current flow, but the important thing about the transistor is that, as we shall see below, by applying a small voltage to the base connection, the transistor can be ‘unblocked’ and current will flow through the it. This is the feature which makes the transistor so valuable as an electrical component.

Two essential features of the transistor are:

- The base layer has to be extremely thin.
- The collector must be arranged so as to be in physical contact with and surround as much of the base as possible.

KEY WORDS

base *one of three regions forming a bipolar junction transistor. This layer separates the emitter and collector layers. If a voltage greater than around 0.6 V is applied to the base terminal, current will flow through the transistor from base to emitter*

collector *one of the three regions forming a bipolar junction transistor. When used in a circuit, a positive voltage is applied to the collector terminal.*

emitter *one of the three regions forming a bipolar junction transistor. Current will only flow from the emitter terminal if a voltage greater than around 0.6 V is applied to the base terminal.*

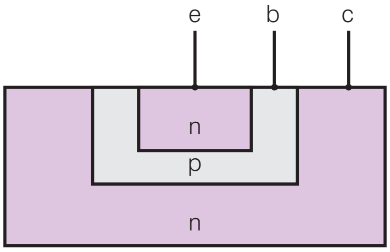


***Figure 5.43** n–p–n transistor – e (emitter), b (base) c (collector).*

Transistor biasing

‘Biasing’ refers to placing voltages across the terminals of a device. A diode is said to be ‘forward biased’ if a voltage is applied to it which enables it to conduct (from Figure 5.44 that will be when the positive side of the battery is connected to the p-type material). If the voltage is applied the other way round, the diode is said to be ‘reverse biased’ and it will not conduct.

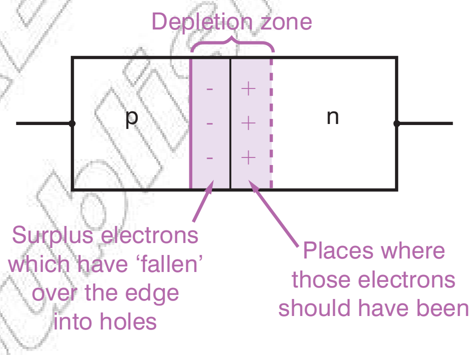


Figure 5.44 *A p–n junction.*

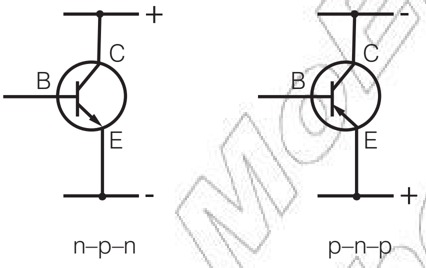


Figure 5.45 *Transistor biasing.*

The arrow on the emitter of a transistor tells you the direction of the conventional current it will allow through. Considering the emitter–collector as the through route, this means that the polarity of the battery across the device must be as in Figure 5.45.

Despite this there should be no current flowing from C because the first junction, the one from collector to base, is reverse biased.

While this is true, it turns out that a small current via the base terminal somehow seems to ‘unblock’ it, and in those circumstances current will flow from C to E.

Current sent into the base has only one possible route out. The junction towards the collector is reverse biased, so it all has to escape via the emitter (Figure 5.46).

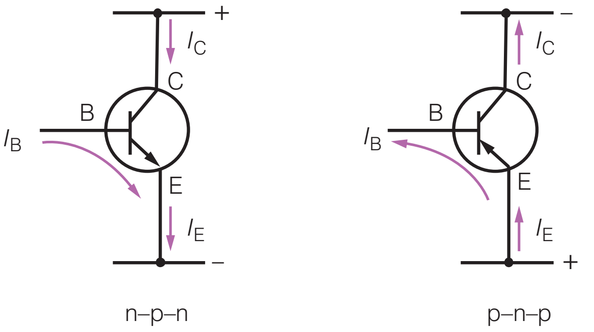


Figure 5.46

By Kirchhoff’s current law, (the total current entering any junction in a circuit equals the current leaving it) $I_E = I_C + I_B$.

In terms of voltages, the base has to be at a voltage intermediate between that of the emitter and the collector. Provided V_{BE} is greater than about 0.6 V, the base–collector junction will conduct, so a base current will flow.

Transistor characteristics

We will consider just the n–p–n transistor because that is the more common one. By its characteristics we mean how one variable affects a second one, all other variables staying constant.

In most applications, the signal at the base is used to control what is received at the collector. Two leads are needed to feed an input into a transistor, however, and two leads will be needed for the output as well. A transistor has only three terminals, so one of them will inevitably be common to both. It can be any of the three, so the circuitry can be classified as common emitter, common base or common collector (Figure 5.47).

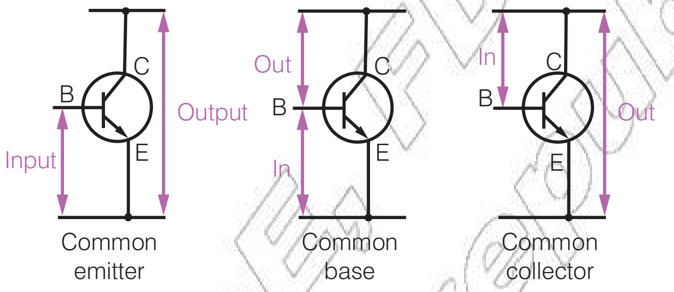


Figure 5.47 *Transistor circuitry.*

These diagrmas do not depict actual circuits – although the input and output are essentially currents, a resistor in the lead between the positive line at the top and the collector will enable the changing voltage across it to be used as an output. Such a common-emitter arrangement is shown in Figure 5.48.

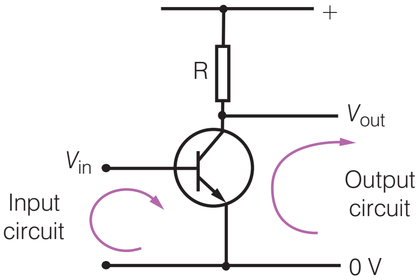


Figure 5.48 *A common emitter.*

Demonstration: Transistor as amplifier

To show how the transistor can be used as an amplifier, connect a microphone to the base of a transistor and a speaker in the collector–emitter part of a circuit (Figure 5.49).

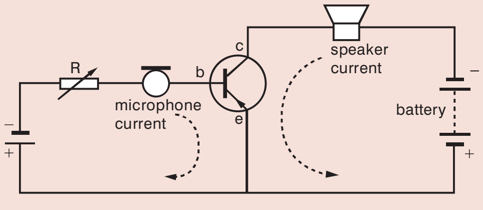


Figure 5.49 *Amplification.*

Such a circuit will cause the very small signal produced by the microphone to become loud enough to be heard through the speaker. In other words, the transistor amplifies the changes in the base current. A very small change in the base current (of the order of μA) causes a much larger change in the collector current (mA).

Figure 5.50a Input to base (microphone).

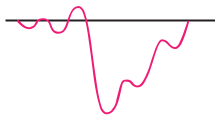


Figure 5.50b Collector current (speaker).

If a CRO trace of the signals in the two circuits can be made, you will be able to observe that the variations in the speaker signal are similar in shape but larger than those of the microphone. Figure 5.50a gives an example of variations of the input current (microphone) while Figure 5.50b shows the variations in the output current (speaker). You can see that the signal has been amplified and an estimate of the current gain can be made.

$$\text{Current gain } (h_{fe}) = \frac{\text{transistor output (collector current)}}{\text{transistor input (base current)}} = \frac{I_C}{I_B}$$

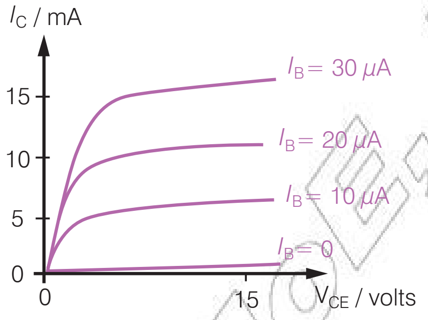


Figure 5.51 Current through semiconductor plotted against voltage.

Variation in collector current (I_C) with increasing voltage (V_{CE})

One important characteristic is how the collector current (I_C) varies as an increasing voltage V_{CE} is placed across the transistor from emitter to collector. The base current I_B must be held constant, but the graph is usually plotted for more than one base current (Figure 5.51).

You can see from this that, so long as there is a sufficient voltage between the collector and the emitter, the dominant influence on the collector current is what is being fed into the base. Notice that the base current is small – with the transistor represented by the graph, a change of 10 μA in the base current would produce a change of around 5 mA in the collector current.

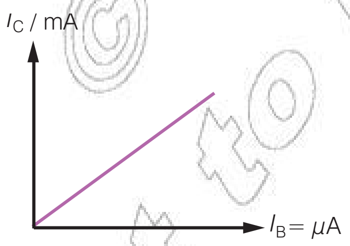


Figure 5.52 Collector current against base current.

Response of collector current to changes in base current

With a transistor which amplifies well, the graph will be nearly a straight line.

Transistor as switch and as amplifier

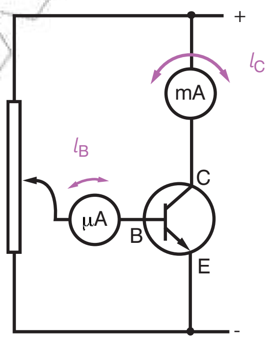


Figure 5.53 Transistor circuit.

The circuit of Figure 5.53 can be used to demonstrate the behaviour of a transistor. A potential divider is used to raise or lower the voltage at the base to control the base current. If there is no base current, the milliammeter records no collector current either. Move the slider on the potential divider up and down – the base current rises and falls, and the movement on the milliammeter seems to track whatever the base current is doing but with much larger variations.

This demonstration also illustrates two of the important functions of a transistor – it can act as an amplifier, and it can serve as a switch. All transistors can do both jobs, but they are usually designed for one purpose or the other. For amplifiers a linear response is all-important, whereas for a switch what matters is that it does it very quickly.

To act as a switch a small current into the base will switch on a larger current at the collector. This is very much what a **relay** might also do. A transistor may not be able to switch on and off really large currents, but otherwise it has the advantages of no moving parts (and hence reliability) and rapid action.

A transistor might be able to switch off in a time of around 10^{-8} s, while a relay may require something nearer a second. You may feel a second is not too long to wait, but to perform a task in a calculation may call for over a million switching operations, one after another. A second for each is too long then!

KEY WORDS

relay *an electrically-operated switch*

A simple amplifier

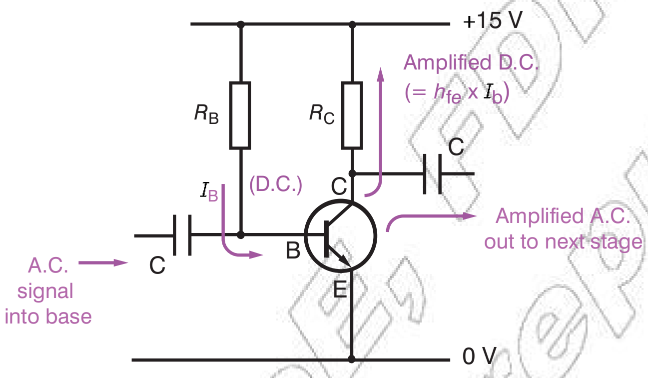


Figure 5.54 *A simple amplifier.*

In Figure 5.54, notice first that the resistor R_B supplies the quiescent base current (that is, a steady d.c. before any input signal is applied). The base–emitter junction is a forward-biased diode, so we may assume that when it is conducting, the voltage across it will be about 0.6 V. If we wish for a base current of, say, 10 µA we can calculate from $R = \frac{V}{I}$ what base resistor is needed. A voltage across it which will be 15–0.6 V has to send 1×10^{-5} A through it – check that R_B will have to be 1.44 MΩ.

Then notice how the input signal is fed in via a capacitor. None of the direct current from R_B can go that way, but the a.c. signal can pass through – and then faced with a choice of the route through to the emitter of the transistor (a low-resistance forward biased diode) or through the 1.44 MΩ, virtually all the signal enters the base.

Now look at the resistor R_C. The collector has to be joined to the positive line to provide a return route for the quiescent current. The value of R_C must be big enough but not too big. It has to be sufficiently large to cause most of the amplified a.c. signal to go instead through the capacitor which links it to the next stage – the larger the better in that respect. The d.c. component of the collector current has to pass through R_C, however, and for that to happen there must be a voltage drop given by $I_C R_C$ across it. The top of R_C is at the +15 V of the supply, so V_{CE} across the transistor must fall by that amount – and we know that if R_C is too big this will make V_{CE} so low that the transistor will stop behaving in a linear fashion or even cease to conduct altogether.

Figure 5.55 is the graph of how I_C varies with V_{CE} for the transistor considered in Figure 5.54. We can see that over its working range a change of 10 μA in the base current will cause a change of around 5 mA in the collector current. The gain (h_{fe}) of the transistor is roughly linear and about 5×10^{-3} A/10×10^{-6} A, which works out to be 500.

We have already chosen a quiescent base current of 10 μA, so the amplitude of the a.c. signal which is to be added must not exceed that value (if it did, the base current would go negative at one of the peaks which would cause the transistor to turn off at that time thereby distorting the output). We are therefore expecting total base currents which will never exceed 20 μA at the most.

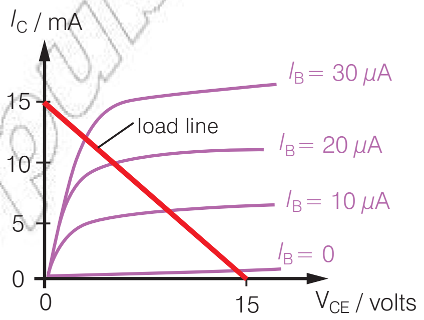


Figure 5.55

The graph now has what we call a load line added. This line shows what the actual value of V_{CE} across the transistor will be – for each value of I_C it is the 15 V of the supply less $I_C R_C$ at that moment. Only when $I_C = 0$ does the full 15 V appear across the transistor.

The straight line must therefore start at (15 V, 0 mA) and is positioned by eye. What is required is that it cuts the I_B = 20 μA line still on an acceptably linear part of its curve. It does not matter that when I_B = 30 μA the voltage has dropped so low that the characteristic is off its linear working range, because we do not intend I_B to get as large as that.

Using the load line drawn, therefore, let us see what is the largest workable value for R_C. Looking at the other intercept, at 15 mA the voltage across the transistor would drop to zero (though it would of course turn off a little before then). A resistor with 15 mA through it and 15 V across it gives a value for R_C of $V/I = 15/15 \times 10^{-3}$ = 1000 Ω. Any value greater than that would mean that in order to handle the expected collector current, the voltage at the collector would drop too low so the output would distort or switch off.

A simple amplifier with negative feedback

Feedback means taking some of the output signal and feeding it back to the input. Negative feedback means that as the output signal rises, a portion is fed back in such a way as to make the input go down a bit and therefore cause the output to drop (Figure 5.56).

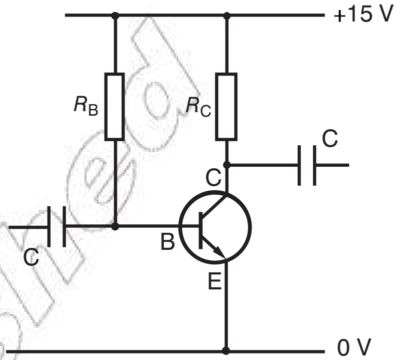


***Figure 5.56** A simple amplifier with negative feedback*

At first sight this might appear undesirable, since it is bound to limit the amplification produced. It does, however, provide useful stability to the transistor. Semiconductors conduct better as they warm up, but this poses a risk: if they conduct better they might pass a larger current, which causes them to heat up further, which makes them conduct more. A good circuit design guards against what could lead to destruction of the semiconductor.

The easiest way to do this is to add a resistor R_E in the emitter line.

To understand how this feedback resistor R_E does its job, suppose the transistor is suffering a surge in the output collector current I_C. Consider the following sequence of events.

1. The current I_E into the emitter must rise too (since $I_E = I_B + I_C$ and the base current is very small).
2. The p.d. across R_E must rise according to $V = IR$.
3. The voltage at emitter E must therefore rise further above 0 V.
4. When the transistor is conducting, like any forward-biased diode the voltage at B will be about 0.6 V higher than that at E, so as the voltage at E rises so must that at B.
5. A higher voltage at B means a smaller p.d. across R_B, so the base current falls.
6. Therefore the collector current I_C drops.

A voltage-operated switch

This circuit is essentially that of an amplifier, but with one difference – the input voltage is either zero or it is a pre-set value (6 V, perhaps), but never anything in between. This means that the transistor is either switched off or it is turned on (Figure 5.57).

The purpose of R_B is to limit the base current to a suitable level, since when 6 V is at the input the only other resistance present is the forward-biased junction.

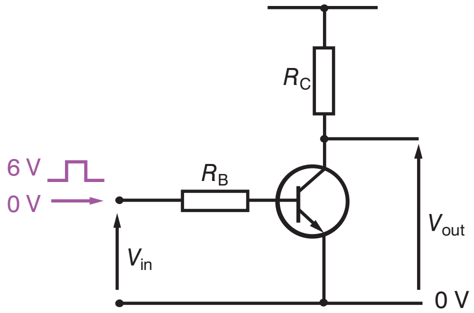


***Figure 5.57** A voltage-operated switch.*

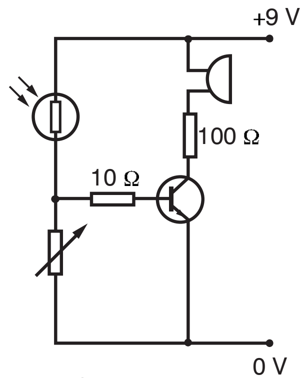


Figure 5.58

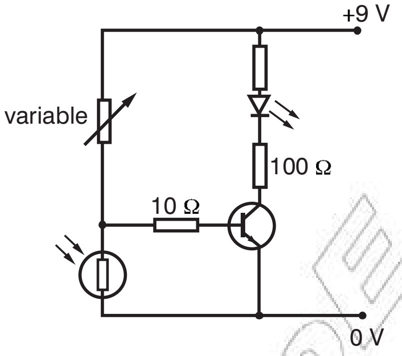


Figure 5.59

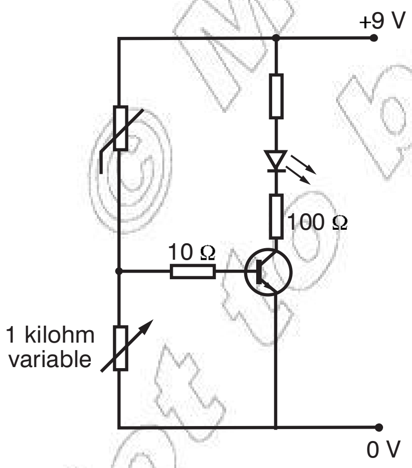


Figure 5.60

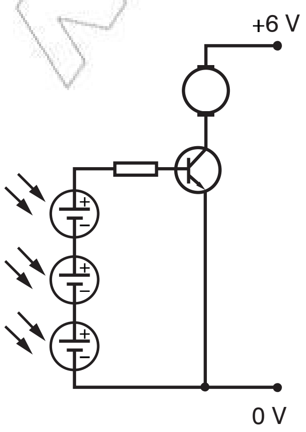


Figure 5.61

This idea is employed in logic circuits, which are discussed later. We say that the input is either a logic '0' or a logic '1', according to whether the voltage is either absent or present.

Using LDR, LED, thermistor, photovoltaic cell and transistor

LDR, LED and transistor

The LDR has a very high resistance when light levels are low, preventing current flowing in the circuit in which it is connected. When light levels rise, the resistance of the LDR drop, allowing current to flow to the base of the transistor.

In the circuit shown in Figure 5.58, the transistor allows current to flow in the circuit containing the buzzer when the LDR is illuminated. This circuit could be the basis for a burglar alarm – if a light shone on the sensor when a room was supposed to be dark, an alarm would sound.

In the circuit shown in Figure 5.59, the transistor allows current to flow in the circuit containing the LED when the LDR receives no light. This circuit could, for example, switch lights on as it gets dark in the evening.

Thermistor, LED and transistor

The thermistor has a very high resistance at low temperatures so, in the circuit shown in Figure 5.60, no current flows to the base of the transistor at low temperatures, and no current is able to flow in the circuit containing the LED. The resistance of the thermistor decreases as the temperature increases allowing current to flow to the base of the transistor. This allows current to flow through the LED.

This circuit could give a visual (or audible if a buzzer were used in place of the LED) warning of overheating.

Photovoltaic cells and transistor

If no light falls on the photovoltaic cells in Figure 5.61, no current is generated in that circuit, and so no current flows to the base of the transistor and no current is therefore able to flow in the circuit containing the motor.

When light shines on the photovoltaic cell, a current flows to the base of the transistor and allows current to flow in the circuit in which a motor is connected.

The motor in such a circuit could be used to operate a blind to cover a window if the sun shone onto the sensor.

Logic gates

Logic gates are tiny silicon chips on which are etched combinations of transistors and resistors. They typically have two inputs and one output. What happens at the output is determined by the situation at those inputs. An example is the AND gate, which we shall consider first.

The AND gate

The power supply needed will be a steady smooth d.c. of around 5 V. Each gate must be connected to this power supply by two wires, which are not normally indicated on the diagrams: one at +5 V, and the other at 0 V (Figure 5.62).

Each input may either be joined to the 0 V line, in which case we say the input is a ‘0’, or it may be joined to the +5 V line, in which case the input is a ‘1’. In Figure 5.63 input A is a ‘1’ and B is a ‘0’.

The output may be a ‘0’ (so it is at the same voltage as the 0 V line) or it may be a ‘1’ (and so acts as if it was the +5 V line). The output of a logic gate is shown by its truth table which is a list of all possible input combinations, showing what you get each time at the output.

Figure 5.64 shows the truth table for the AND gate. As its name suggests, the only way to get a ‘1’ at the output is to have a ‘1’ at input A and a ‘1’ at input B. In Figure 5.63, you can see that input A is ‘1’, input B is ‘0’. From the truth table you can see that the output for these input values is ‘0’.

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

Figure 5.64 *Truth table for AND gate.*

You can think of an AND gate as a bit like two switches in series (Figure 5.65). Both switches must be placed in the ‘1’ position to make the output ‘live’ (that is, at +5 V).

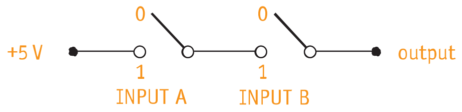


Figure 5.65 *You can think of an AND gate as two switches in series.*

As the inputs are voltages and not currents, all you have to do is touch the input contact onto the ‘live’ (+5 V) line for a ‘1’, or touch it on the earth line for a ‘0’. The input resistance is huge, so in effect no current flows into the gate.

The OR gate

The OR gate may not be quite what you would expect from its name. If the AND gate could be pictured as two switches in series, the OR gate behaves like the same two switches in parallel (Figure 5.66b).

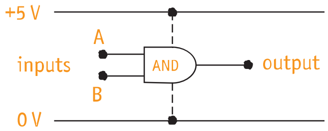


Figure 5.62 *AND gate – the dotted lines are not usually shown.*

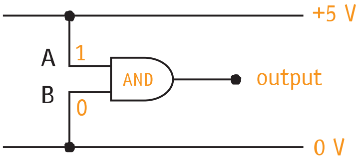


Figure 5.63 *Input A is ‘1’ (one) and input B is ‘0’ (zero).*

KEY WORDS

logic gate *an electronic device that performs a logical operation on two inputs and produces a single logic output*

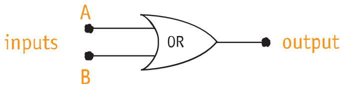


Figure 5.66a *The symbol for an OR gate.*

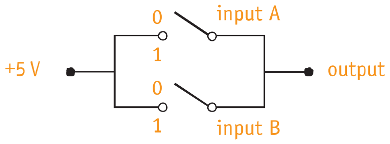


Figure 5.66b *You can think of an OR gate as two switches in parallel.*

The truth table for the OR gate is shown in Figure 5.67.

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

Figure 5.67 *Truth table for OR gate.*

You might not have expected the last line in the table from what we normally understand by the word 'or', but if you think of those two switches in parallel (Figure 5.66b), you should soon be able to work out the correct truth table. The output will always be at +5 V except when both switches are at 'zero'.

The NOT gate

Yet another gate is the NOT gate. This has a single input, and its output is always the opposite. A '0' at the input means a '1' at the output, and vice versa. Figure 5.68 shows the symbol for a NOT gate.

If you are puzzled as to how a '1' can come out when nothing is going in, you are forgetting that the gate also has connections to the power supply which are not shown. If the input is connected to earth (0), therefore, the +5 V of the 'live' line appears at the output. The truth table for the NOT gate is shown in Figure 5.69.

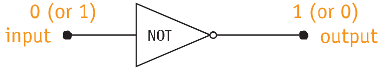


Figure 5.68 *The symbol for a NOT gate.*

INPUT	OUTPUT
0	1
1	0

Figure 5.69 *Truth table for NOT gate.*

The NAND gate and the NOR gate

There are two final gates to consider – the NAND gate and the NOR gate. These are just the AND gate and the OR gate respectively, but with the output inverted. Instead of a '0' there is a '1', and instead of a '1' a '0'. Their symbols are shown in Figure 5.70 and the little circle at the output indicates that each value is reversed.

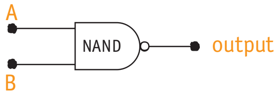


Figure 5.70a *Symbol for a NAND gate.*

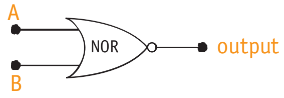


Figure 5.70b *Symbol for a NOR gate.*

The four truth tables are summarised in Figure 5.71.

			Output		
INPUT A	INPUT B	AND	NAND	OR	NOR
0	0	0	1	0	1
0	1	0	1	1	0
1	0	0	1	1	0
1	1	1	0	1	0

Figure 5.71 *Truth table for NAND and NOR gates.*

Nobody would expect you to sit and learn that table by heart. You should be able to work out how the AND and OR gates behave, and then obtain the other two simply by reversing their outputs.

Combinations of logic gates

More than one logic gate may be combined to increase the range of control tasks that can be performed.

As an example, consider the arrangement shown in Figure 5.72. How does it behave?

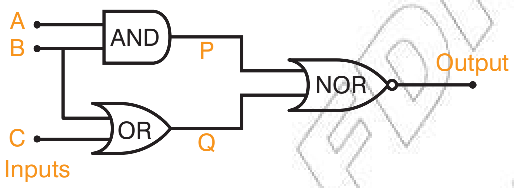


***Figure 5.72** Combining different logic gates.*

Start by preparing a table with columns for all possible combinations of inputs for the corresponding logic states at intermediate points P and Q, and finally for the output.

There are three inputs this time, which will give $2^3 = 8$ different combinations of 0s and 1s. If you know your binary notation, the neatest way is to write them down is in sequence by listing in turn the binary equivalents of 0 to 7. Figure 5.73 shows the truth table filled in with these figures.

INPUT A	INPUT B	INPUT C	POINT P	POINT Q	OUTPUT
0	0	0			
0	0	1			
0	1	0			
0	1	1			
1	0	0			
1	0	1			
1	1	0			
1	1	1			

***Figure 5.73** Initial truth table for logic gate combination shown in Figure 5.72.*

Once the inputs are entered, work out the states of points P (from inputs A and B) and Q (from inputs B and C) and record them in the table (Figure 5.73).

Then use P and Q as inputs to the final gate in order to find the output.

Check that in this case the result is as shown in Figure 5.74.

INPUT A	INPUT B	INPUT C	POINT P	POINT Q	OUTPUT
0	0	0	0	0	1
0	0	1	0	1	0
0	1	0	0	1	0
0	1	1	0	1	0
1	0	0	0	0	1
1	0	1	0	1	0
1	1	0	1	1	0
1	1	1	1	1	0

Figure 5.74 *Final truth table for logic gate combination shown in Figure 5.72.*

Worked example 5.4

Consider the arrangement shown in Figure 5.75. How does it behave?

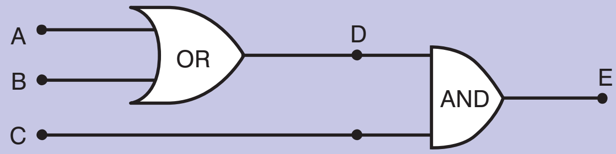


Figure 5.75

Start by preparing a table with columns for all possible combinations of inputs (A, B and C). There are three inputs this time, which will give $2^3 = 8$ different combinations of 0s and 1s. As seen above, the neatest way is to write them down is in sequence by listing in turn the binary equivalents of 0 to 7. Figure 5.76 shows the table filled in like this.

INPUT A	INPUT B	INPUT C	POINT D	POINT E
0	0	0		
0	0	1		
0	1	0		
0	1	1		
1	0	0		
1	0	1		
1	1	1		

Figure 5.76 *Intermediate truth table for logic gate combination shown in Figure 5.75.*

Now fill in the column for the logic state at intermediate point D (remember to use columns for Input A and Input B to work out the results for point D), and finally for the output, E (using the columns for Input C and point D). Figure 5.77 shows the completed table.

INPUT A	INPUT B	INPUT C	POINT D	POINT E
0	0	0	0	0
0	0	1	0	0
0	1	0	1	0
0	1	1	1	1
1	0	0	1	0
1	0	1	1	1
1	1	1	1	1

Figure 5.77 *Final truth table for logic gate combination shown in Figure 5.75.*

The action of logic gates

In this section we will give some circuits involving logic gates and say what they are meant to do. Your task each time is to try to explain how they work.

A simple burglar alarm

In the circuit drawn in Figure 5.78, two doors and two windows are equipped with micro-switches such that if they are opened the switch closes and the signal from it changes from a 0 to a 1.

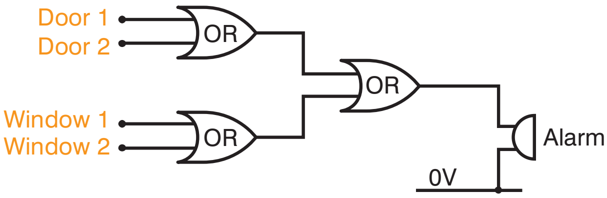


Figure 5.78 *A circuit for a simple burglar alarm.*

The switches are connected to one of a pair of two-input OR-gates, so if any of the inputs becomes a 1 the alarm will sound.

A thermostat for a hot water tank

In the circuit drawn in Figure 5.79,

(a) two contacts are fixed near the top of the tank and when submerged (i.e. when the water is deep enough) the water completes a circuit between the contacts, signalling that the water is deep enough,

(b) a thermistor (set to the required temperature) is immersed in the water in the tank; it signals if the temperature of the water is too cold.

The water heater will only be switched on if both (a) the tank is full of water AND (b) the water is too cold.

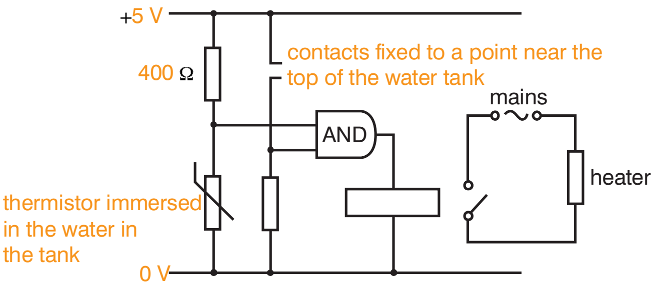


Figure 5.79 *A thermostat for a hot water tank.*

What the output of logic gates can do – using a relay

When there is an '0' at the output, it acts as if it was the 0 V line. When there is a '1', it behaves like part of the +5 V line and is capable (in principle, at least) of lighting a bulb (Figure 5.80).

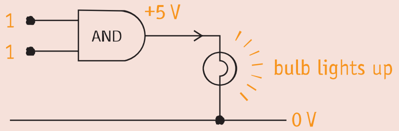


Figure 5.80 *When the output is a '1' a current can flow.*

The only snag is that the tiny silicon chip cannot really handle the sort of current needed to make the filament of a bulb white-hot, but it can cope with the much smaller current needed to make a light emitting diode glow.

There is one other thing that the small current which a gate will supply is capable of doing – it can operate a relay. The output of the gate feeds current to an electromagnet. An entirely separate circuit contains a switch made of iron. The electromagnet will attract the iron switch towards it against the pull of a spring, and this action either turns the separate circuit on or switches it off (Figure 5.81).

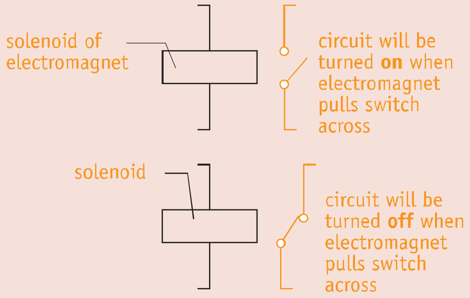


Figure 5.81 *How a relay works.*

The circuit symbol for each type of relay is shown in Figure 5.81. Figure 5.82 shows an AND gate whose inputs are both '1', so it will turn on a powerful electric motor.

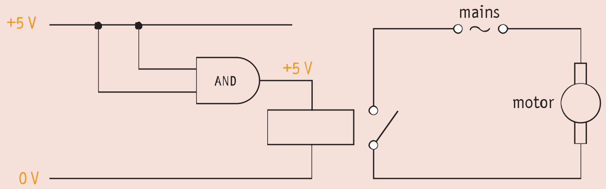


Figure 5.82 *How an AND gate can operate a relay to switch a motor on and off.*

A security lock to get 'behind the counter' at a bank

In the circuit drawn in Figure 5.83, button A is on the bank manager's desk and button B is by the door of the bank, on the outside.

The solenoid which pulls the bolt in, unlocking the door, is only activated if both button A AND button B are pressed and held down.

The warning buzzer sounds if button B is pressed AND button A is NOT pressed.

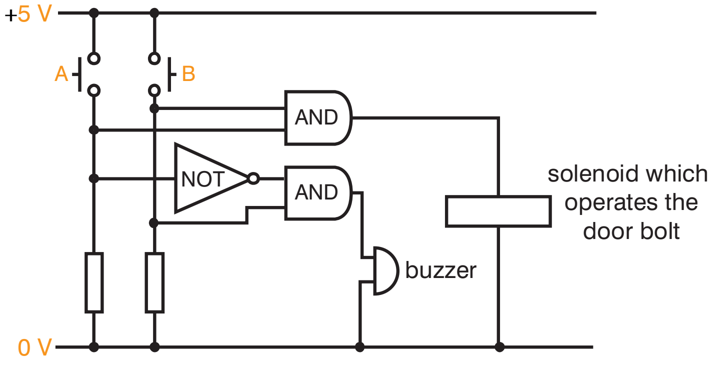


Figure 5.83 *A security lock circuit.*

An automatic plant waterer

In the circuit drawn in Figure 5.84:

(a) Contacts sensing nmoisture content are placed in the soil. These generate a signal when the soil is damp. As the system is required to switch the pump on when the soil is dry, the signal passes through a NOT gate.

(b) A light sensor, consisting of a photovoltaic cell, generates a voltage when light falls on it, but as the system is required to switch on when it is dark, the signal passes through a NOT gate.

(c) Switch S is a manual override. Pressing this switch will switch on the pump whatever the conditions.

The solenoid switching on the pump to water the plants will come on if the soil is dry AND it is dark OR if the manual override switch is pressed.

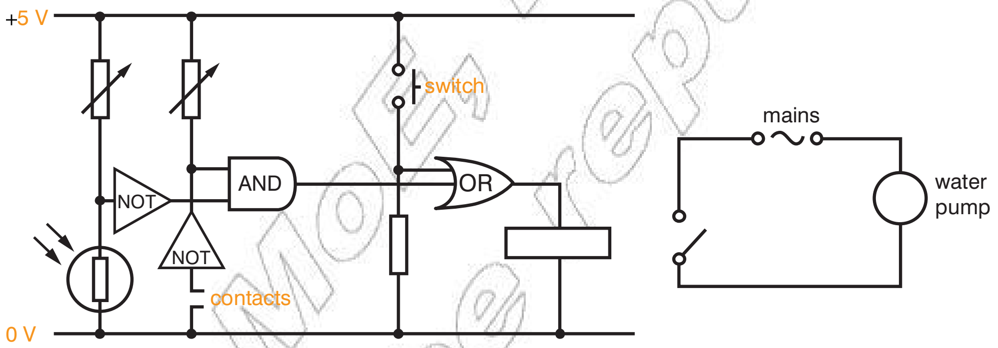


Figure 5.84 *An automatic plant watering system.*

Integrated circuits

Logic gates are grouped together in circuits – such as the gate shown in Figure 5.85a.The tiny silicon chip is encased in resin and is connected to the outside world via its 14 metal 'legs' (Figure 5.85b). Note that the gates are identified just by their symbols. It is a 'quad 2-input NAND' meaning that it contains four separate NAND-gates, each with an input A and an input B. A single +5 V connection and a single 0 V one serve all four gates inside.

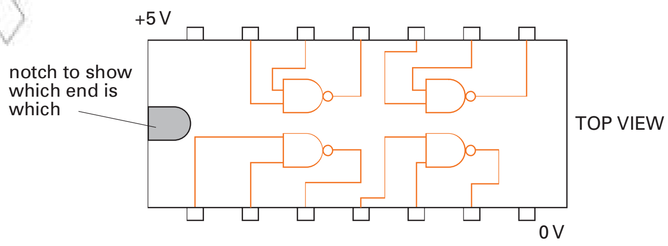


Figure 5.85a *Layout.*

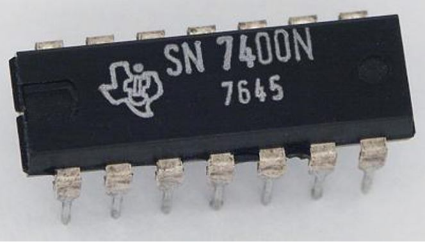


Figure 5.85b *Photo of a quad 2-input NAND.*

Summary

- With an n–p–n transistor, there is a low input resistance at the base, and the base current flows out of the emitter via the forward-biased junction.
- The gain h_{fe} of the transistor is defined to be $\frac{\Delta I_C}{\Delta I_B}$.
- The collector current is almost as big as the emitter current: $I_E = I_C + I_B$.
- A transistor can be used as an amplifier or as a switch.
- Logic gates work in terms of 1s (a voltage is present) and 0s (the voltage is not present).
- The truth table for a logic gate is simply a list of all the possible input readings with their corresponding output.
- Logic gates can be combined to perform all sorts of tasks relating to control and alarms.

Review questions

1. A transistor has a gain h_{fe} of 200. If a current of 3.0 μA is sent into the base, what size collector current would you expect?
2. Compile your own notes on types of logic gate.
   a) Start with the NOT gate. Give its symbol (just its outline shape this time, without writing the word 'NOT' on it) and its truth table, which for this gate will have only two columns and two lines to it.
   b) Now do the same for the AND gate. When you come to the truth table, see if you can get all four lines right before you check – if it helps, add a note comparing it to two switches in series (Figure 5.65).
   c) Do the NAND gate next. Again, working from the AND gate, try to write down the truth table before looking.
   d) Next do the OR-gate in the same way. This time, if it helps, add a note comparing it to two switches in parallel (Figure 5.66b).
   e) Finally repeat for the NOR gate.
3. a) In Figure 5.86, does the output of the logic gate need to be a '0' or a '1' in order for the tiny bulb to light? Explain.
   b) What inputs will be needed at A and B to achieve this?
4. Complete the truth table for the arrangement of gates shown in Figure 5.87.

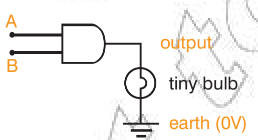


Figure 5.86

A	B	C	D	E
0	0	1		
1	0	0		
0	1	1		
1	1	0		

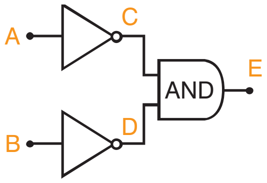


Figure 5.87

End of unit questions

1. A battery is connected to the y terminals of an oscilloscope and the spot deflects 3 cm vertically. The sensitivity of the oscilloscope is set at 1 V/cm. What is the p.d. across the battery?
2. Boron has only three electrons in its outer shell. Explain why doping a silicon crystal with a small amount of boron makes it a p-type semiconductor.
3. What is the difference between d.c. obtained from a battery and that obtained from an alternating supply when you add a diode in series?
4. What do you understand by n-type and p-type semiconductors?
5. The input of a logic gate may be either a '0' or a '1'. Explain what that means.
6. a) Name the type of logic gate shown in a Figure 5.88a. Write down its truth table.

   b) Show that when its two inputs are joined together as in Figure 5.88b it will act as a NOT gate.

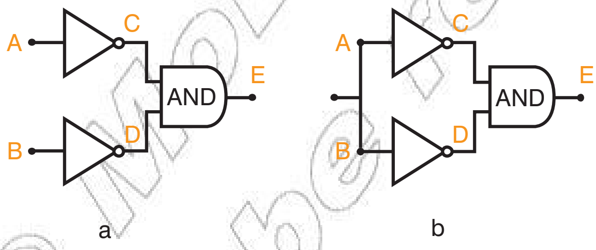


Figure 5.88