Collins

PHYSICS

Electronics
AQA A-level
Year 2

Chris Bishop

William Collins' dream of knowledge for all began with the publication of his first book in 1819.

A self-educated mill worker, he not only enriched millions of lives, but also founded a flourishing publishing house. Today, staying true to this spirit, Collins books are packed with inspiration, innovation and practical expertise. They place you at the centre of a world of possibility and give you exactly what you need to explore it.

Collins. Freedom to teach

HarperCollins*Publishers*

The News Building,

1 London Bridge Street

London

SE1 9GF

Browse the complete Collins catalogue at
www.collins.co.uk

This optional topic is part of the Collins AQA A-Level Physics Year 2 Student Book.

10 9 8 7 6 5 4 3 2 1

ISBN 978-0-00-759764-2

A catalogue record for this book is available from the British Library

Authored by Chris Bishop
Commissioned by Emily Pither
Development by Jane Roth
Editorial management by Mike Appleton and Kate Ellis
Edited by Geoff Amor
Proofread by Mitch Fitton and Sue Glover
Artwork and typesetting by Jouve
Cover design by We are Laura
Printed by Grafica Veneta S.p.A.

The publisher would like to thank Sue Glover and Peter Robinson.

Approval Message from AQA

This textbook has been approved by AQA for use with our qualification. This means that we have checked that it broadly covers the specification and we are satisfied with the overall quality. Full details for our approval process can be found on our website.

We approve textbooks because we know how important it is for teachers and students to have the right resources to support their teaching and learning. However, the publisher is ultimately responsible for the editorial control and quality of this book.

Please note that when teaching the A-level Physics course, you must refer to AQA's specification as your definitive source of information. While this book has been written to match the specification, it cannot provide complete coverage of every aspect of the course.

A wide range of other useful resources can be found on the relevant subject pages of our website: www.aqa.org.uk

CONTENTS

ELECTRONICS

The *Tianhe-2* (whose name in English means 'Milky Way 2') is currently the most powerful computer in the world (see figure). It is capable of 33.86 'petaflops' per second, which means it can perform 33 860 trillion calculations each second, and it has a memory capacity of 1 PB (1 PB = 1 petabyte = 1000 terabytes (TB) or 1 000 000 gigabytes (GB)). It uses 80 000 separate microprocessors linked together in 162 storage cabinets – over 700 m^2 of floor space are required to house it. *Tianhe-2* is the first step in building computers capable of processing on the 'exaflop' level – a further thousand-fold increase in power.

The Tianhe-2 supercomputer developed and built by China's National University of Defence Technology

The first supercomputers were developed in the 1960s but these used only a few processors. During the 1990s, as advances in the manufacture of large-scale integrated circuits progressed and their costs decreased, machines with thousands of processors began to appear. Now supercomputers with thousands of 'off-the-shelf' processors, not unlike those found in your PC, are routinely connected together in massively parallel combinations to form extremely powerful computing architectures that can solve complex mathematical problems.

The supercomputer *Deep Blue* was in 1997 the first machine to defeat a world reigning chess champion (Gary Kasparov). Since then, supercomputers have increasingly been used in the field of artificial intelligence, in attempts to mimic the way humans think. In 2013, one such computer was hailed as passing the Turing test, which was developed by computer science pioneer and Second World War codebreaker Alan Turing. The machine convinced more than 30% of a panel of judges, with whom it was holding a series of keyboard conversations, that it was a 13-year-old boy.

Supercomputers play important roles in our lives, being used for a wide range of computationally intensive tasks, such as weather forecasting, climate research, oil and gas exploration, molecular modelling, physical simulations and data analysis. Despite their complexity, supercomputers are built up from simpler digital electronic subsystems containing discrete electronic devices that you will learn about in this option.

1 DISCRETE SEMICONDUCTOR DEVICES

PRIOR KNOWLEDGE

You will need to have a good understanding of potential difference (pd), current and resistance. You will need to remember how a potential divider works (*see Chapter 14 of Year 1 Student Book*). It will help if you refresh your memory of the behaviour of a semiconductor diode, outlined in *Chapter 13 of Year 1 Student Book*. You should also be familiar with photons (*see Chapter 8 of Year 1 Student Book*), and with magnetic fields and magnetic flux density (Chapter 7 in this book).

LEARNING OBJECTIVES

In this chapter you will learn about discrete semiconductor devices such as the MOSFET, the Zener diode, the photodiode and the Hall effect sensor, and how they can be used in electronic circuits for amplification, switching, voltage reference and sensing.

(Specification 3.13.1.1–3.13.1.4)

1.1 SEMICONDUCTING MATERIALS

To understand how semiconductor devices work, we need to consider the motion of charges in the materials they are constructed from – **semiconductors** (*see section 14.3 in Chapter 14 of Year 1 Student Book*). Examples of common semiconductor materials are silicon, germanium, gallium arsenide and selenium. Semiconductors allow the movement of charge carriers – that is, electric current – but in a limited way. They have an electrical conductivity somewhere between that of a conductor and that of an insulator.

On an atomic level, the important difference between conductors, semiconductors and insulators lies in the number of free electrons present in the material. A single isolated atom has a well-defined set of energy levels dependent on the electron configuration, and electrons can move between the different levels. In a solid, the atoms are linked together by inter-atomic bonds, which cause the discrete energy levels to broaden to form **energy bands** (Figure 1).

In a conductor, the outermost electrons or **valence electrons** can move freely through the material. In an insulator, the valence electrons are tightly bound

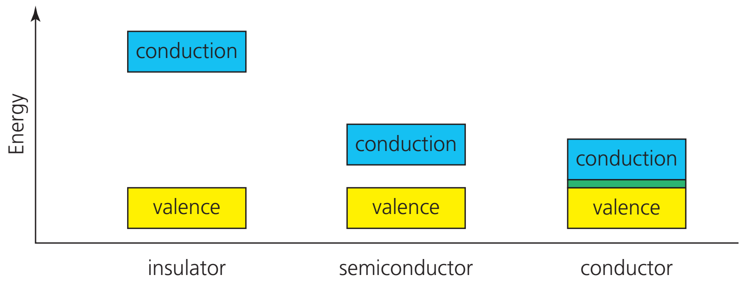


Figure 1 *The energy band model of a solid. The gap between the conduction and valance bands is the energy gap that determines the conductivity of the material.*

to the atom and cannot move freely. When atoms in a solid material form energy bands, a **valence band** and a **conduction band** are formed. In the conduction band, electrons can move freely. If the material is given enough energy, then electrons from the valence band can move to higher energies in the conduction band and move through the material.

The size of the **energy gap** between the valence band and the conduction band determines the conductivity of the material. In an insulator, the energy gap is large and electrons need a large amount of energy to jump into the conduction band. In a conductor, the valence and conduction bands overlap (there is no energy gap between them) and hence valence electrons can easily move into the conduction band. In a semiconductor, the energy gap is somewhere between that of a conductor and insulator. If the material is given a small amount of energy (for example, by an applied potential difference, by heat or by light), then electrons are able to bridge the gap into the conduction band.

The structure of a semiconductor material such as silicon is in the form of a crystalline lattice. If an electron in the lattice is dislodged from its position by thermal energy, it will move through the lattice, leaving behind an atom with a missing electron, called a **hole**. A hole is the absence of an electron in a particular place in an atom. Although it is not a physical particle in the same sense as an electron, a hole can be passed from atom to atom in a semiconductor material. A hole forms when an atomic electron moves from the valence band into the conduction band. The hole moves when it is filled with an electron coming in from the conduction band – in Figure 2, an electron from the left. However, this electron has now created a hole to the left – so the hole has effectively moved to the left.

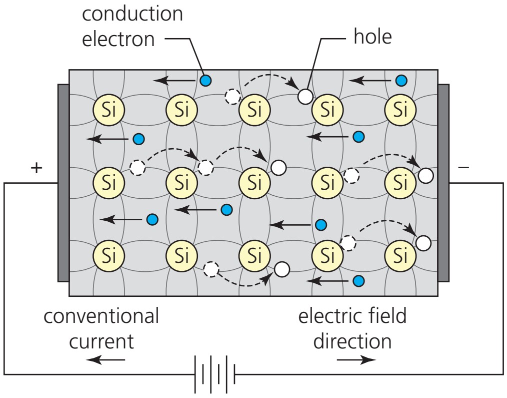


Figure 2 *Charge carriers in the crystalline lattice of a semiconductor are made up of holes in the valence band and electrons in the conduction band. Holes move in the direction of conventional current, that is, in the direction of the electric field applied.*

Holes behave as particles with a positive charge, equal in magnitude to the electronic charge. They are always in the valence band. Holes and electrons are the two types of **charge carriers** responsible for current in semiconductor materials. If a potential difference is applied across a semiconductor lattice, as shown in Figure 2, then we can think of the current as the flow of charge carriers made up of conduction electrons in one direction *plus* holes in the other direction.

Intrinsic and extrinsic semiconductors

The electrical properties of semiconductors can be controlled by a process called **doping**, in which small amounts of impurities are introduced into the semiconductor material. An **intrinsic semiconductor** is an undoped semiconductor. No impurity atoms have been added to the material. Any holes in the valence band are vacancies created by electrons that have been thermally excited into the conduction band as an intrinsic property of the material itself. An **extrinsic semiconductor** is one that has been doped with impurity atoms that modify the material's intrinsic electrical conductivity by adding charge carriers in a precise way. This makes it suitable for the manufacture of semiconductor devices such as diodes and transistors. Extrinsic semiconductors can be 'n-type' or 'p-type'.

- An **n-type** semiconductor is an extrinsic semiconductor where the dopant atoms are 'donors' – they provide extra conduction electrons to the semiconductor material. An example is the doping of silicon with phosphorus atoms. This increases the conductivity by creating an excess of negative (n-type) charge carriers. Electrons are the majority charge carriers.
- A **p-type** semiconductor is one that has been doped with 'acceptor' atoms, such as boron, that create an excess of holes as the majority charge carriers.

The p–n junction

This difference in doping becomes important when the two types of extrinsic semiconductor materials are joined together to form a **p–n junction**, as in a **diode** (*see section 13.5 in Chapter 13 of Year 1 Student Book*). A p–n junction diode is the simplest semiconductor device. It allows current to flow only in one direction. At the p–n junction, some of the extra electrons in the n-type material will diffuse across the junction and occupy the holes in the p-type material, forming a region that is depleted of charge carriers, called the

depletion region. As a consequence, there will be a small negative charge on the left of the depletion layer and a small positive charge on the right (Figure 3). The depletion layer is effectively an electrically insulating barrier between the n-type and p-type materials, so there can be no further net charge flow across the junction.

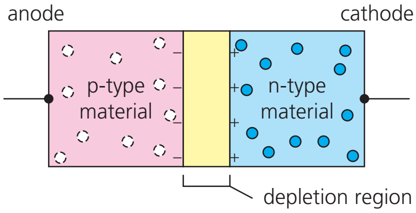


Figure 3 *Depletion region in a p–n junction. The combining of electrons and holes depletes the holes in the p-type region and depletes the electrons in the n-type region near the junction.*

If the p-type side of the diode, the **anode**, is connected to the positive terminal of a source of pd, and the n-type side, the **cathode**, is connected to the negative terminal, the free electrons in the n-type material are drawn to the positive potential and the holes in the p-type material move the other way. If the pd across the diode is high enough, the valence electrons in the depletion zone have enough energy to move freely again. The depletion zone disappears, and current moves across the diode. In this state, the diode is **forward-biased** (Figure 4a).

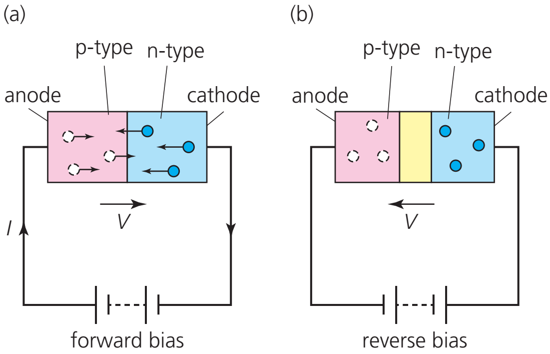


Figure 4 *(a) Forward-biased and (b) reverse-biased p–n junction diode*

If the diode is connected the other way round, with the p-type side (anode) connected to the negative potential and the n-type side (cathode) to the positive potential, diffusion of electrons and holes at the junction causes the depletion region to increase, so current will not flow. In this state, the diode is **reverse-biased** (Figure 4b).

If the reverse-bias voltage is high enough, however, at a **breakdown voltage** the depletion region breaks down and a very high reverse **avalanche current** flows. The breakdown is permanent and the diode is damaged.

We will look at some special types of p–n junction diode in Electronics sections 1.3 and 1.4.

1.2 MOSFET (METAL OXIDE SEMICONDUCTING FIELD-EFFECT TRANSISTOR)

A **transistor** ('transfer-resistor') is a three-terminal semiconductor device that can regulate current and voltage and can also act as a switch. It was invented at Bell Laboratories in the USA in 1947 as a replacement (in radio and telephony technology) for bulky vacuum tubes or 'valves', which required a lot of power to run and were frequently unreliable.

The **MOSFET** (metal oxide semiconducting field-effect transistor) is a type of transistor that works by varying the width of a conducting **channel** along which charge carriers flow. It has three terminals, called the **gate**, the **source** and the **drain**. In an n-channel MOSFET (Figure 5), which is the type we will be considering, the charge carriers are electrons, and these enter the channel at the source and exit via the drain. The conventional current is from drain to source.

Figure 5a shows the construction of an n-channel MOSFET. An underlying layer or **substrate** of p-type material (see Electronics section 1.1) is used, with two n-type regions diffused into one surface. A very thin layer, about 10^{-4} mm thick, of metal oxide forms an insulating layer on the outside of that same surface. Aluminium contacts, or **electrodes**, make electrical contacts to different parts of the MOSFET, as shown, to form the gate, source and drain.

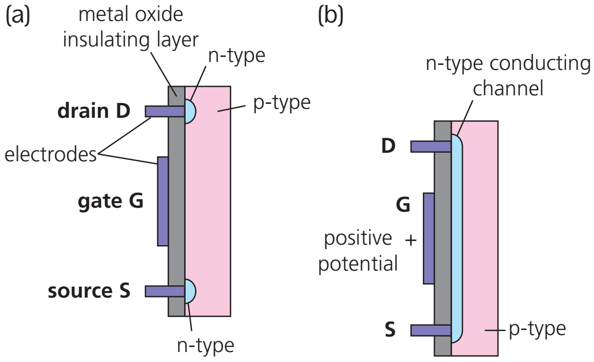


Figure 5 *An n-channel MOSFET: (a) construction; (b) when a positive gate potential is applied*

When the gate electrode has a voltage applied to it that is positive with respect to the source electrode, holes in the p-type substrate are repelled away from the metal oxide layer, into the substrate. Free electrons in the p-type material (the minority carriers) are attracted towards the metal oxide layer. These create an n-type conductive channel between the source and drain (Figure 5b). Increasing the gate voltage increases the number of electrons in the channel; decreasing the gate voltage decreases the number of electrons in the channel. If the gate voltage is reduced to zero, the conductive channel collapses and the region near the gate reverts to p-type again, as in Figure 5a.

If a potential difference is applied between the drain and the source while there is a conductive channel, then a current will flow through the MOSFET. As the gate voltage is increased, the size of the conducting channel, and hence the size of the current, between the source and the drain grows or is 'enhanced', and the MOSFET is said to be operating in **enhancement mode**.

A MOSFET may be a discrete device (Figure 6) or it can be fabricated on an integrated circuit (IC) chip. On an IC chip, it might be only of the order of nanometres from source to drain. A single IC could contain many thousands of MOSFETs, along with other components, such as resistors, capacitors and diodes in very high densities.

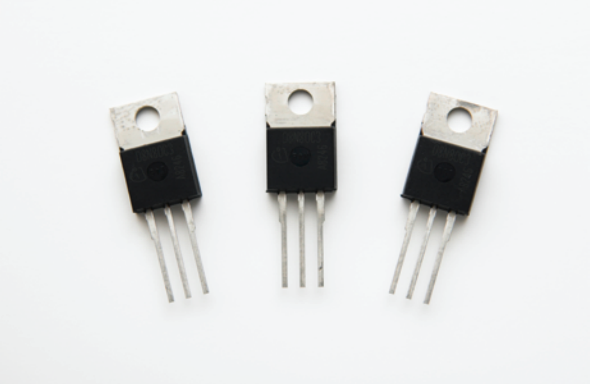


Figure 6 *MOSFET, showing the three terminals*

The circuit symbol for an n-channel MOSFET is shown in Figure 7. If an **input** signal voltage is applied to the gate, this controls the drain–source current and hence the **output** in the external circuit. The insulating metal oxide layer gives the MOSFET an extremely high **input resistance** (about 10^{13}–$10^{14}\,\Omega$), which is so high that the MOSFET draws essentially no current from the input signal. This means that it will draw very little power from the input signal when operating as an amplifier.

However, the metal oxide layer is extremely thin, and a MOSFET is susceptible to destruction by electrostatic charges accumulating on the oxide layer between the gate and the source (which behaves like a capacitor). So the gate should not be left unconnected – a path to ground is needed to allow the charge to flow off.

QUESTIONS

1. What is meant by an *n-type semiconductor* and a *p-type semiconductor*?
2. What property does the insulating metal oxide layer give to a MOSFET?
3. Why must MOSFETs be handled carefully?

MOSFET measurement parameters and characteristic curves

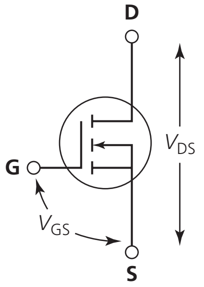


Figure 7 *Circuit symbol for an n-channel MOSFET and the voltages* V_{DS} *and* V_{GS}

A MOSFET (Figure 7) is specified by a number of measurement parameters:

- V_{DS} is the voltage between the drain and the source
- V_{GS} is the voltage between the gate and the source
- V_{th} is the gate-to-source **threshold voltage** – the minimum value of V_{GS} required to form a conducting channel between the drain and the source
- I_{DS} is the current between the drain and the source
- I_{DSS} is the small leakage current between the drain and the source when the gate voltage V_{GS} is 0 V.

The drain–source current and drain–source voltage characteristic is shown in Figure 8.

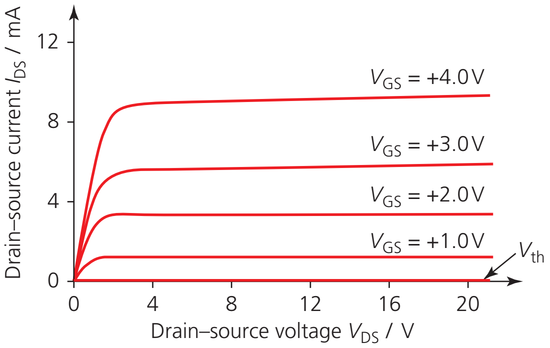


Figure 8 *The I_{DS} versus V_{DS} characteristic for a MOSFET*

When V_{GS} is less than V_{th}, the MOSFET is OFF because there is no conducting channel. However, a small leakage current I_{DSS} flows, which is of the order of a few nanoamps (nA).

Above V_{th}, a channel starts to form and the MOSFET turns ON. The I_{DS} versus V_{DS} characteristic curves have almost vertical and almost horizontal parts. The linear, almost vertical parts of the curves correspond to the ohmic region, where the MOSFET channel acts like a resistor. The linear, almost horizontal part corresponds to the constant-current region, where there is almost no increase in drain current for increasing V_{DS}. This is the **saturation region**. The drain–source current is then controlled by the value of V_{GS}.

Figure 9 shows a typical I_{DS} versus V_{GS} **transfer characteristic** (V_{GS} input for I_{DS} output) for the saturation region. Above V_{th}, the drain current I_{DS} increases slowly at first with an increase in V_{GS}, and then much more rapidly. In practice, V_{th} is the value when the current in the channel attains the manufacturer's minimum operating value, which may typically be 10 μA. The value of V_{th} varies for different devices, from less than 1 V to more than 5 V.

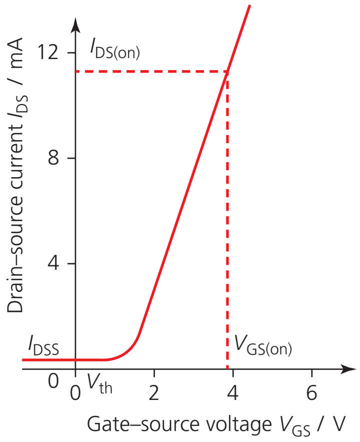


Figure 9 *The I_{DS} versus V_{GS} transfer characteristic for a MOSFET*

The operation of an enhancement-mode MOSFET can be summarised in terms of three operating regions.

- *Cut-off region*: with $V_{GS} < V_{th}$, the gate–source voltage is lower than the threshold voltage, so the MOSFET is switched fully OFF and $I_{DS} = 0$; the MOSFET acts as if it was an open circuit.
- *Ohm's law region*: with $V_{GS} > V_{th}$, the MOSFET acts like a variable resistor whose value is determined by the gate voltage V_{GS}, up to the point where it becomes saturated.
- *Saturation region*: with $V_{GS} > V_{th}$ and a high enough value of V_{DS}, the MOSFET is in its constant-current region and is switched fully ON or saturated; the current I_{DS} = maximum and depends of the value of V_{GS}.

Beyond the saturation region is the *breakdown region*, where, at a certain value of V_{DS}, called the **breakdown voltage**, the drain–source path of the MOSFET breaks downs internally and a large current will flow, destroying the transistor.

A MOSFET has a performance parameter called the **transconductance**, g_m, which is the change in drain current ΔI_{DS} caused by the change in the voltage between the gate and the source, ΔV_{GS}:

$$g_m = \frac{\Delta I_{DS}}{\Delta V_{GS}}$$

It can be found by working out the gradient of the graph in Figure 9, and has the unit mA V^{-1}.

The MOSFET as a switch

MOSFETs can be used as switches. Figure 10 shows a circuit using a MOSFET to turn on a lamp from a signal input that produces a square voltage pulse.

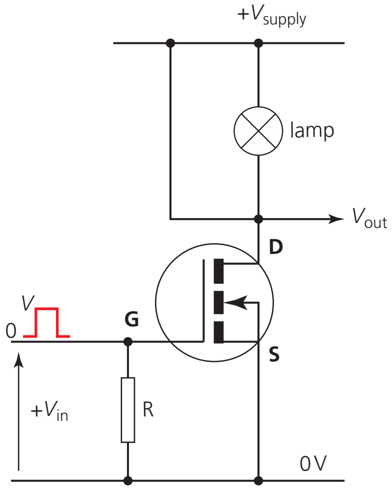


Figure 10 *MOSFET as a simple switch*

When the voltage pulse is at 0, the input voltage $V_{in} = 0$:

- the voltage at the gate of the MOSFET is 0 V, so less than the threshold voltage V_{th}
- the MOSFET is fully OFF and no drain–source current flows.

When the voltage pulse is at V, then $V_{in} = V$:

- if $V > V_{th}$, the MOSFET is now fully ON (saturation region) and the maximum drain–source current flows.

So the lamp is turned on and off by the level of voltage on the gate. The lamp in this circuit could be replaced by a light-emitting diode (LED) or any other load that needs to be switched.

The resistor R ensures that the gate is not left unconnected and so not damaged by static electricity.

An important advantage of a MOSFET switch is that it needs very little current to switch devices on and off – it will draw hardly any current from a signal source. This makes it ideal for battery-operated devices where current consumption needs to be kept to a minimum.

The MOSFET has a much faster switching time than other types of transistor. However, it can get hot when acting as a switch, and wastes power as a result.

Table 1 summarises the main characteristics of a MOSFET.

Output current	Controlled by gate voltage
Input resistance	Extremely high $> 10^{12}\,\Omega$
Transconductance	Given by $g_m = \frac{\Delta I_{DS}}{\Delta V_{GS}}$
Static-sensitive	Yes
Switching time	Fast
Input voltage	An n-channel MOSFET requires voltages greater than V_{th} (a few volts) to turn on with virtually zero current

Table 1 *MOSFET characteristics*

Worked example

The circuit in Figure 11 shows a touch sensor that uses a MOSFET to light a lamp. The MOSFET has $V_{th} = 4.0$ V and the saturation current is 50 mA. A 6 V, 60 mA lamp is connected to the MOSFET as shown.

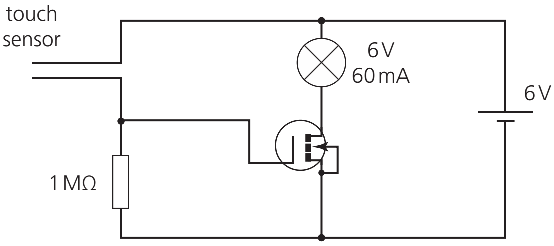


Figure 11

a. If the resistance of a human finger is about 20 kΩ, what voltage will appear between the gate and source of the MOSFET when the sensor is touched?

b. Why will the MOSFET turn on?

c. When the MOSFET turns on it becomes fully saturated. What will the current through the lamp be?

d. Will the lamp glow at maximum brightness?

e. What is the power dissipated by the MOSFET if the channel resistance between the drain and source R_{DS} is 0.1 Ω and the MOSFET is fully saturated when on?

a. V_{GS} is found from the potential divider relation, so

$$V_{GS} = \left(\frac{10^6}{10^6 + 20 \times 10^3}\right) \times 6 = 5.9\,\text{V}$$

b. The MOSFET will turn on because V_{GS} is greater than V_{th}.

c. If it is in saturation, the maximum current through the lamp will be 50 mA.

d. No. The maximum current the lamp can take is 60 mA. Only 50 mA flows because the MOSFET is saturated.

e. Power $P = I^2R$ (*see section 13.7 in Chapter 13 of Year 1 Student Book*). With a channel resistance R_{DS} of 0.1 Ω, the power dissipated in the MOSFET will be

$$P = I_{DS}^2 \times R_{DS} = (50 \times 10^{-3})^2 \times 0.1 = 2.5 \times 10^{-4}\,\text{W}$$
$$= 0.25\,\text{mW}$$

QUESTIONS

4. What property of MOSFETs makes them suitable for low-power devices?

5. Why are MOSFETs commonly used in integrated circuits?

6. What controls the amount of current flowing in the MOSFET channel?

7. In the circuit of Figure 12, a touch sensor is used to turn on an LED. What is the purpose of

   a. the resistor R1

   b. the resistor R2?

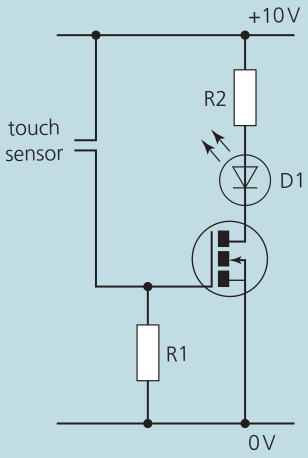


Figure 12

KEY IDEAS

- A MOSFET is a transistor that has three terminals called the gate, source and drain.
- In enhancement mode, the current flowing through a MOSFET (source to drain) I_{DS} varies linearly with the voltage between drain and source V_{DS} up to the MOSFET's saturation value.
- In the saturation region, the current I_{DS} is no longer dependent on V_{DS}. Its value is controlled by the width of the MOSFET's conducting channel, which is determined by the input voltage between the gate and the source, V_{GS}.
- When there is no voltage on the gate, there is no conducting channel. The MOSFET does not conduct until the gate voltage is greater than a threshold voltage V_{th}.
- A MOSFET may be used as a switch to turn a device on and off.
- A MOSFET has an extremely high input resistance, which means that it takes virtually no current from the input source, and this makes it useful in applications that require low current consumption.

1.3 ZENER DIODES

A **Zener diode** is a special type of semiconductor junction diode that allows current to flow in the forward direction, but also allows it to flow in the reverse direction when the reverse-bias voltage across the diode is above a certain value. An ordinary junction diode (see Electronics section 1.1), when reverse-biased, will not conduct significantly until the reverse-bias voltage reaches the breakdown voltage, and then the avalanche current will permanently damage the diode.

When a Zener diode is connected in reverse bias and the applied voltage is increased, the large reverse current at breakdown does not damage the diode because of the Zener diode's special construction. Zener diodes are manufactured to have their reverse breakdown occur at a specific, well-defined voltage, and are designed so that they can be operated continuously in that **breakdown mode**.

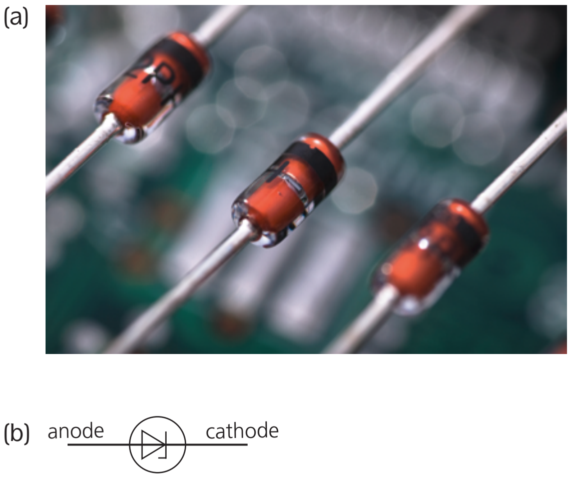


Figure 13 *A Zener diode has the position of the cathode indicated by a stripe on the diode body and by a crooked bar line on the circuit symbol. In forward bias the cathode is connected to the negative source terminal; in reverse bias it is connected to the positive source terminal.*

The Zener diode was named after Clarence Zener, who described theoretically how an insulator could break down when a voltage was applied across it, and whose work led to the development of such diodes by Bell Laboratories in the USA. Figure 13 shows a Zener diode and its circuit symbol. Note that sometimes the circle in the symbol is omitted.

Figure 14 shows the current–voltage characteristic of a Zener diode. In forward bias it operates like other types of diode (*see section 13.5 in Chapter 13 of Year 1 Student Book*), with a normal 'turn on' forward voltage V_F of about 0.7 V. However, a Zener diode is normally used in reverse bias, and at a certain voltage called the **Zener breakdown voltage**, or simply **Zener voltage**, V_Z, the diode breaks down internally in a controlled way and an avalanche current flows. It can be seen from the reverse-bias characteristic in Figure 14 that, at V_Z, the voltage across the diode then remains almost constant regardless of the current flowing through it. The diode requires a **minimum operating current** to ensure that it is working in breakdown mode. It will also have a maximum avalanche current, above which it will be permanently damaged by overheating.

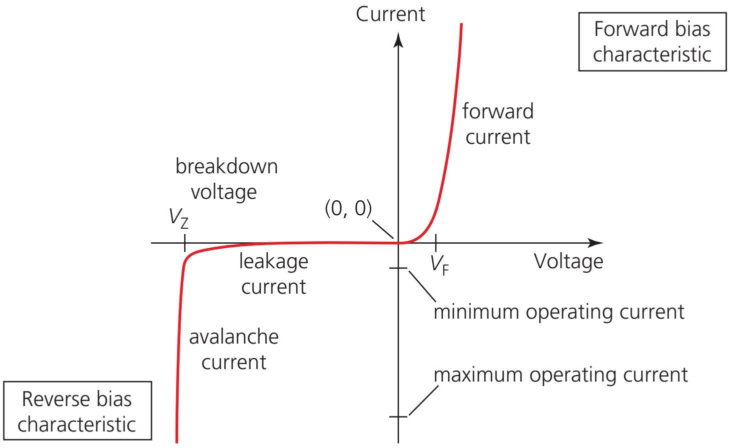


Figure 14 *Forward and reverse characteristics of a Zener diode*

Zener diode as constant-voltage source and reference voltage

The property of Zener diodes to maintain almost constant voltage while in the breakdown region makes them useful as constant-voltage sources and in circuits where a reference voltage is needed for comparison purposes. An unstable input voltage that is varying may be steadied to a constant reference voltage given by the Zener voltage V_Z if a Zener diode is connected, reverse-biased, as shown in Figure 15.

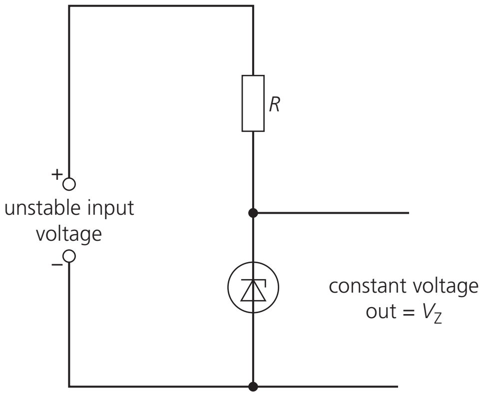


Figure 15 *Zener diode as a constant-voltage source. If the input varies, then the output remains constant at* V_Z.

The input voltage has to be higher than the Zener voltage V_Z, and the resistor value R must be chosen such that there is always current flowing through the Zener diode between its minimum and maximum operating currents in the breakdown mode. Zener diodes are available with V_Z anywhere from 1.8 to 200 V.

Worked example

A portable CD player requires a maximum current of 150 mA and a voltage of 6.0 V. It is connected to a car battery via the cigarette lighter socket. Depending on how well the car battery is charged, the battery voltage can vary from as low as 9.6 V to as high as 13.2 V.

a. Draw a circuit diagram with a Zener diode and a single resistor R that could provide a constant voltage to the CD player regardless of the battery voltage. What should the Zener breakdown voltage of the diode be?

b. What would be the maximum and minimum voltages across the resistor depending on the state of the car battery?

c. The Zener diode needs a minimum current of 5 mA flowing through it to operate, under all input conditions. What current flows through the resistor when the output current is 150 mA?

d. The CD player is drawing its full current and the battery voltage is 9.6 V. What is the value of *R*?

a. The circuit would be as in Figure 16, with the car battery voltage as the input, and the CD player connected to the output from the Zener diode. The Zener breakdown voltage V_Z should be 6.0 V.

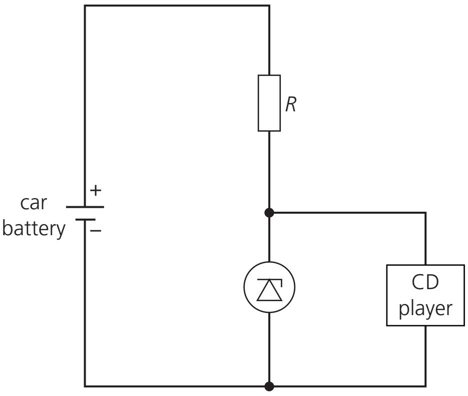


Figure 16

b. Maximum voltage across *R* is 13.2 − 6.0 = 7.2 V; minimum voltage is 9.6 − 6.0 = 3.6 V.

c. Current is 5 mA + 150 mA = 155 mA
= 0.155 A

d. Resistance is $R = \frac{V}{I} = \frac{3.6}{0.155} = 23.2\,\Omega$

QUESTIONS

8. a. In which sense is a Zener diode normally connected in a circuit when used as a voltage reference?

b. Why should you not connect a Zener diode directly across a varying voltage source (without the use of a resistor) to provide a voltage reference?

9. The circuit in Figure 17 shows a Zener diode connected as a voltage reference in series with a resistor *R*.

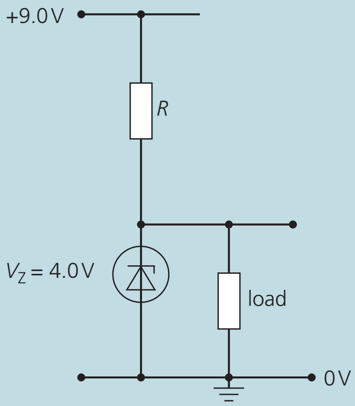


Figure 17

a. What is the voltage across the load if the Zener diode is operating between its minimum and maximum currents?

b. Assume the load takes a maximum current of 95 mA, and for normal operation the Zener diode must take a current of at least 5 mA. What will be the maximum current flowing through the resistor?

c. Calculate the value of *R*.

d. If the load is disconnected, how much current will flow through the Zener diode?

KEY IDEAS

- A Zener diode is a semiconductor p–n junction that permits current to flow in the forward direction but also allows it to flow in the reverse direction when the voltage is above a certain value called the Zener breakdown voltage or Zener voltage V_Z.
- In the breakdown region, the voltage across the Zener diode remains constant even when the current through it is changing.
- Zener diodes can be used as constant-voltage sources and references when a steady voltage is needed in a circuit.
- The Zener diode must be connected in series with a resistor that keeps the current through the diode between its minimum operating current and its maximum avalanche current.

1.4 PHOTODIODES

A **photodiode** is a semiconductor diode that conducts only when light is incident on it. It consists of a p–n junction and a transparent window through which light can enter. Figure 18 shows a photodiode and its symbol.

Figure 18 *Photodiode and its symbol. Note that sometimes the circle in the symbol is omitted.*

A photodiode has a p–n junction just like any other diode (see Electronics section 1.1), except that the junction is exposed to light. When it is connected in reverse bias, as shown in Figure 19, it is said to be working in **photoconductive mode**.

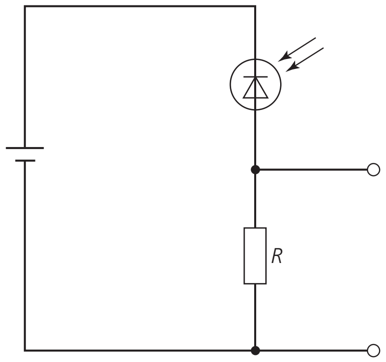


Figure 19 *A photodiode connected in reverse bias with a resistor in series to limit the current*

The depletion region at the junction has an electric potential across it when reverse-biased (see Figure 4b in Electronics section 1.1). When photons of light are incident on the depletion region, electron–hole pairs are produced. The electric potential sweeps the holes towards the anode and the electrons towards the cathode, and a **reverse photocurrent** flows. The greater the light intensity, the greater the number of electron–hole pairs created and the greater the photocurrent (Figure 20).

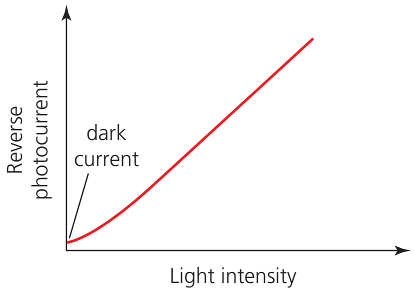


Figure 20 *The reverse current in a photodiode has a linear relationship with the intensity of light, except that when there is no incident light a small 'dark current' flows.*

Spectral response

A photodiode has different sensitivity to different wavelengths of light, which is described by its **spectral response**. Photodiodes can be manufactured with spectral responses in different parts of the electromagnetic spectrum. In Figure 21 the green curve represents a photodiode whose spectral response peaks at 550 nm – it is most sensitive to visible light at that wavelength. The red and purple curves are for photodiodes with spectral responses in the infrared and the ultraviolet.

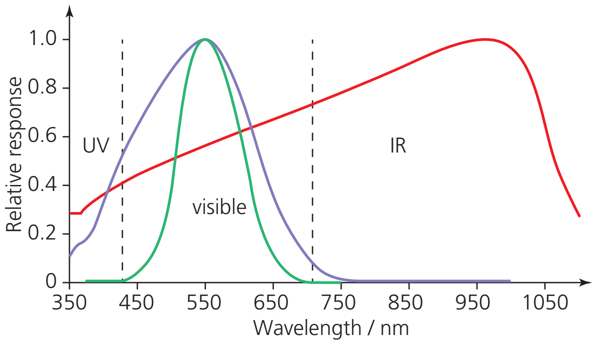


Figure 21 *Spectral response curves for three different photodiodes*

The response of a photodiode can be expressed as its **photosensitivity**. This is defined as

$$\text{photosensitivity} = \frac{\text{photocurrent generated}}{\text{power incident}}$$

It is measured in $\mathrm{A\,W^{-1}}$.

Worked example

Figure 22 shows the spectral response curve of a silicon photodiode.

a. At what wavelength does the photodiode produce the greatest photocurrent, for a given incident light intensity?

b. If 0.01 mW of monochromatic light, of wavelength that was calculated in part **a**, is incident on the photodiode receiving area, what size of photocurrent is generated?

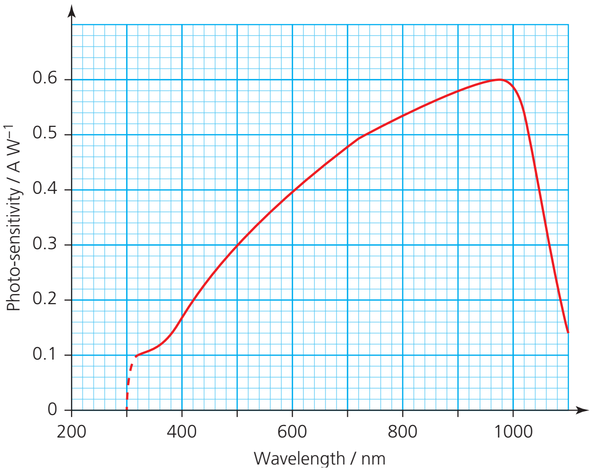


Figure 22

a. From the graph, the peak photosensitivity is at about 960–980 nm, say 970 nm.

b. Photosensitivity at the peak wavelength is 0.6, so the photocurrent is 0. 01 × 10^{-3} × 0.6 = 6.0 μA.

QUESTIONS

10. What is meant by a photodiode working in photoconductive mode?

11. What is the difference between the spectral response and the photosensitivity of a photodiode?

12. A photodiode has an active light-receiving area of diameter 0.4 mm. Light of wavelength 700 nm and intensity 0.1 mW cm^{-2} is incident on the receiving area.

- **a.** What power falls on the photodiode receiving area?
- **b.** If the photodiode generates a photocurrent of 56.6 nA, what is the photosensitivity, in A W^{-1}, of the photodiode at this wavelength?

Photodiodes in optical systems

Photodiodes in photoconducting mode are used in many different types of circuits and applications, for example, light meters, smoke detectors, position sensors, photocopiers and light detectors for optical fibre communication systems.

In one type of smoke detector, a pulsing infrared LED is located in a chamber. The chamber is designed to exclude light from any external source. A photodiode is positioned in the chamber so that normally it does not intercept the beam from the LED (Figure 23). In the event of smoke from a fire entering the chamber, the light from the LED will be scattered and some will be directed to the photodiode. This will generate a photocurrent that can be amplified to sound an alarm.

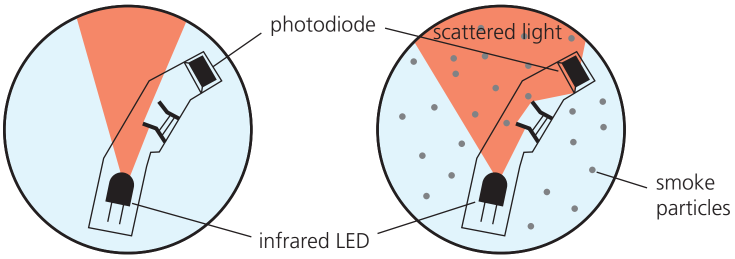


Figure 23 A *photodiode used in a smoke detector*

Scintillator and photodiodes

A **scintillator** is a material that produces a flash of light when a particle such as an ion, electron, alpha particle or high-energy photon passes through it. If coupled to a photodiode (Figure 24), then the number of light pulses can be detected and amplified, and the energy of the particle passing through the scintillator measured.

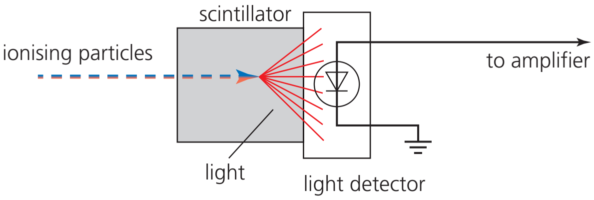


Figure 24 *Scintillator and photodiode used to detect ionising particles*

Scintillator materials can be liquid or solid, but a common material used is a crystalline solid, often caesium iodide (CsI) or sodium iodide (NaI). Figure 25 shows a CsI scintillator bonded to a photodiode.

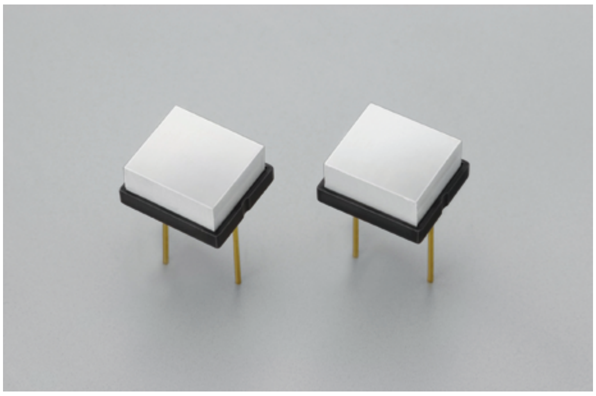


Figure 25 *Scintillation photodiode. The white cube is the scintillator material.*

Only a fraction of the original particle energy deposited in the scintillator is available to produce photons of light that can detected by the photodiode. This fraction is expressed as the **scintillator efficiency**:

$$\text{scintillator efficiency} = \frac{\text{total energy of light photons produced in scintillator}}{\text{energy deposited in scintillator by particles or high-energy photons}} \times 100\%$$

The scintillator efficiency is quite low, of the order of 3% to 15%. However, a single particle event depositing energy in the scintillator may typically produce a few thousand photons. This is known as the **light yield** and is measured as the number of photons per MeV. A photon produced in the scintillator typically has an energy of 3–4 eV, which is sufficient to create an electron–hole pair in the depletion region of the photodiode. For maximum detection efficiency, the spectral response of the photodiode needs to be matched as closely as possible to the wavelength produced by the scintillator. The electrons and holes are collected respectively at the anode and the cathode of the diode. The charge collected is amplified and a pulse produced that can be counted.

Worked example

A beta particle of energy 0.4 MeV is incident on a CsI scintillator coupled to a photodiode. The light yield of the scintillator is 55×10^3 photons/MeV.

a. How many photons are produced in the scintillator by a single beta particle of 0.4 MeV?

b. If each of the photons produces an electron–hole pair, what is the total electronic charge produced?

c. Each photon has an energy of 2.5 eV. What is the wavelength of the light detected by the photodiode?

a. Number of photons $= 55 \times 10^3 \times 0.4$
$= 22 \times 10^3$

b. Total electronic charge $= 22 \times 10^3 \times (-1.6 \times 10^{-19}) = -3.5 \times 10^{-15}\,\text{C}$

c. Photon energy (*see section 8.3 in Chapter 8 of Year 1 Student Book*) is $E = hc/\lambda$ so

$$\lambda = \frac{hc}{E} = \frac{6.6 \times 10^{-34} \times 3.00 \times 10^{8}}{2.5 \times 1.6 \times 10^{-19}}$$
$$= 4.95 \times 10^{-7} = 495\,\text{nm}$$

Photodiode scintillation detectors are used where a low-power supply is needed or desired. Also, because they are relatively unaffected by magnetic fields, they are used in medical scanners and in high-energy physics experiments where high magnetic fields may be present. They are also useful for radiation detection in small spaces or where a rugged non-liquid detector is required.

QUESTIONS

13. **a.** Explain what is meant by a *scintillator*.
    b. An atomic particle of energy 2 MeV is incident on a scintillation photodiode. The scintillator material has a light yield of 23 000 photons/MeV. If the photons detected by the photodiode have a wavelength of 447 nm, what is the scintillation efficiency of the scintillator material?

KEY IDEAS

- A photodiode is a p–n junction diode that converts the energy of incident light into the electrical energy of a current.
- A photodiode is in photoconductive mode (able to produce a photocurrent) when it is reverse-biased.
- Photodiodes have a spectral response, which means that they generate maximum photocurrent for a specific wavelength range of incident light.
- Applications of photodiodes are in optical detectors, where light needs to be converted into an electrical signal.
- Photodiodes can be used with scintillators to detect atomic particles.

1.5 HALL EFFECT SENSOR

When a current-carrying semiconductor (or conductor, to a lesser extent) is placed in a magnetic field, a small potential difference will arise across it in a direction perpendicular to the field. This effect is called the **Hall effect**, discovered in 1879 by Edwin Hall. The voltage produced is called the **Hall voltage** or **Hall pd** V_H and is proportional to the magnetic flux density B in which the semiconductor lies.

The Hall effect arises from a magnetic force on, and hence movement of, charge carriers across the semiconductor material, so creating an electric field across it (Figure 26). The direction of the Hall pd shows (from Fleming's right hand rule) whether the majority carriers are negative or positive. If the majority are positive, side X of the semiconducting material becomes positive with respect to side Y; if the majority are negative, side X becomes negative with respect to side Y.

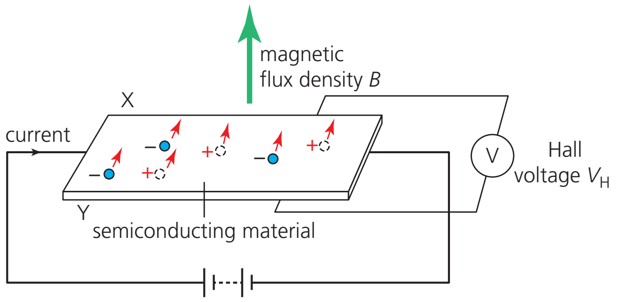


Figure 26 *The Hall effect in a semiconductor*

Since V_H is proportional to B, then it is possible to use the Hall effect in a transducer that varies its output voltage in response to a varying magnetic field. This is called a **Hall sensor** (Figure 27). It allows the direction of the magnetic field and the magnitude of the flux density to be determined.

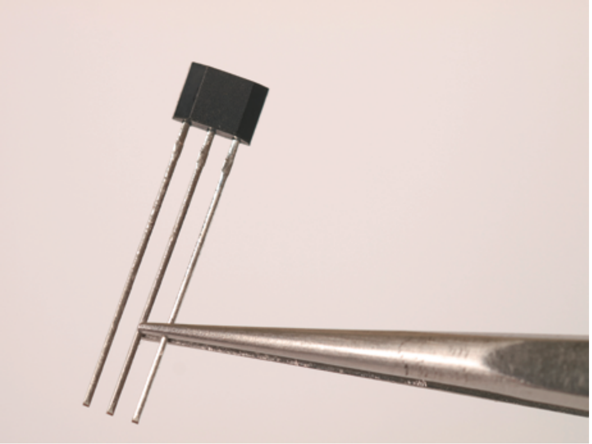


Figure 27 *A Hall sensor is a small device with three terminals.*

Hall sensor characteristic curve

The characteristic curve, of output (Hall) voltage against magnetic flux density, of a Hall sensor is shown in Figure 28. At large values of flux density, the Hall sensor saturates – the output no longer increases with the field. When there is no significant magnetic field present, then V_H is half-way between the saturation points.

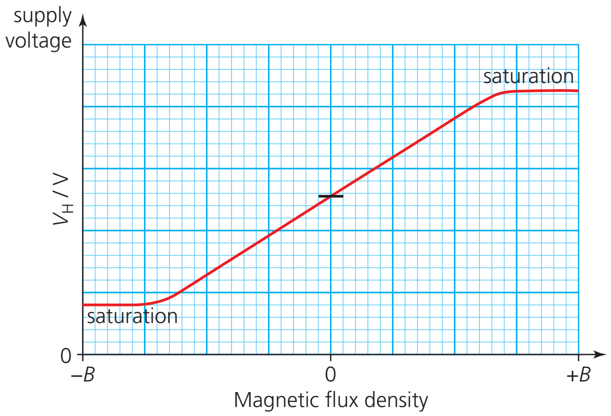


Figure 28 *Hall sensor characteristic curve*

Hall sensor attitude sensing

A Hall sensor can be used to detect the position and attitude of an object in three dimensions. This requires attaching a magnet to the object.

In Figure 26, the magnetic field detected is the field normal to the plane of the semiconducting material. If the magnet is at an angle, then this changes the component of B normal to the Hall sensor (Figure 29), which changes the output Hall voltage. In this way the attitude or angle change can be detected.

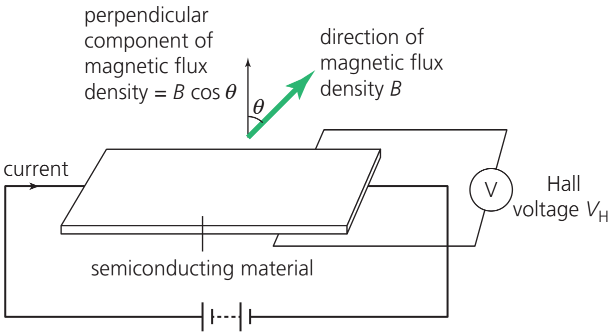


Figure 29 *The principle of attitude sensing with a Hall sensor. The value of the Hall voltage for a non-normal magnetic field is proportional to B cos θ.*

When a magnet moves across the face of a Hall sensor in a sideways motion with a fixed gap (Figure 30), V_H will vary proportionally with the perpendicular component of the flux density of the magnetic field through its face, which can be related to the position of the magnet.

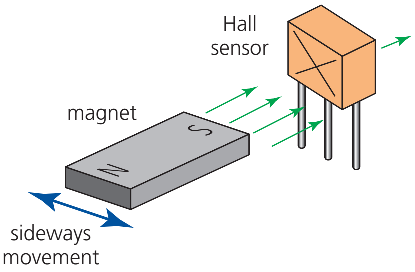


Figure 30 *Sideways motion detected by a Hall sensor*

This directional movement detection works for vertical components of the motion as well as horizontal, and for rotation of the magnet. Hence, from the output voltage, the attitude of the magnet relative to the sensor can be determined. In a joystick used in computer games (Figure 31), the movement of the stick, which has a magnet in its base, gives rise to a varying voltage that is dependent on the orientation of the stick.

Figure 31 *A joystick used in computer games makes use of the Hall effect.*

Hall sensor tachometers

Hall sensors can be used in **tachometers** to measure the rotation speed of wheels or rotating machinery. Figure 32 shows two configurations.

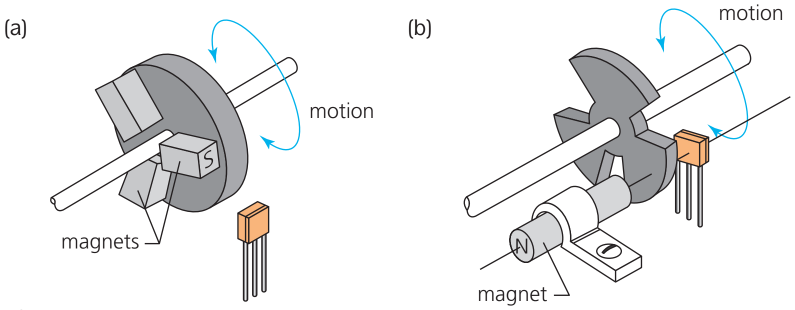


Figure 32 *Hall effect tachometers*

In Figure 32a, magnets rotate with the rotating shaft, and the stationary Hall sensor produces a voltage output pulse with each pass of a magnet pole.

In Figure 32b, the shaft has a vane attached made out of ferrous material. In this configuration, both the Hall sensor and the magnet are stationary and the vane alters the strength of the magnetic field in a periodic fashion as it passes through the magnetic field, resulting in a corresponding variation of V_H as the ferrous vane rotates.

The output from the Hall sensor can be amplified to produce a series of pulses, the frequency of which allows the number of times the shaft rotates per second to be determined.

Hall effect tachometers can work at frequencies of 100 kHz or more, which means they can measure very fast rotational speeds. They are also quite small, so they can be put in confined spaces, and they do not make physical contact with the shaft, so there are no frictional forces.

QUESTIONS

14. Explain how a Hall sensor can be used to sense attitude (angle).

15. Describe *two* advantages that a Hall effect tachometer has in measuring rotation speeds.

KEY IDEAS

- The Hall effect is the production of a voltage across a current-carrying semiconductor or conductor when in a magnetic field.
- The Hall voltage V_H is proportional to the flux density B of the magnetic field.
- Hall sensors are semiconductor devices that use the Hall effect to measure magnetic fields.
- Hall sensors can be used to measure the position and attitude of objects, and can be used in tachometers to measure very fast rotational speeds.

PRACTICE QUESTIONS

1. a. MOSFETs are commonly used in circuits where low power consumption is important to extend battery life. State and explain the property of MOSFET devices that makes them useful in these circuits.

Figure Q1 shows an n-channel enhancement mode MOSFET, being used as part of a circuit for the water level alarm in a garden pond. When the gap between the copper strips is filled with water, the MOSFET turns on and the alarm sounds.

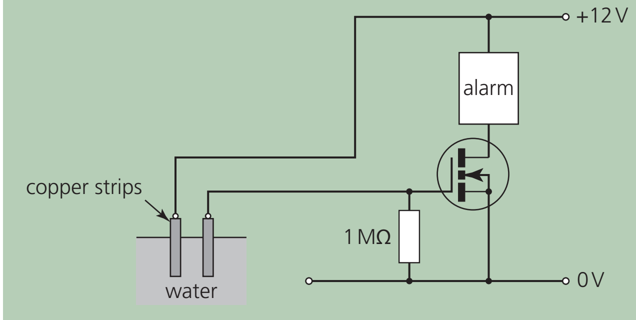


Figure Q1

b. Explain the reason for the 1 MΩ resistor in this application.

c. The circuit is tested by immersing the copper strips in the water, and bringing them closer together until the alarm sounds. V_{th} for the MOSFET in Figure Q1 is 2.4 V.

Determine the resistance of the water between the copper strips when the alarm sounds.

AQA Paper 3BE Specimen 2014 Q1

2. A student designs a circuit to obtain a USB-compatible output voltage from a car power socket.

a. The car's power socket voltage can vary from 10.5 V to 14.7 V, and the USB voltage needs to be steady. What would be the problem of using a simple potential divider with two resistors to supply a steady voltage to the USB device?

b. The student decides to use the circuit shown in Figure Q2, using a Zener diode of 4.7 V to give a constant voltage to the USB device. The operating current of the Zener diode is 15 mA.

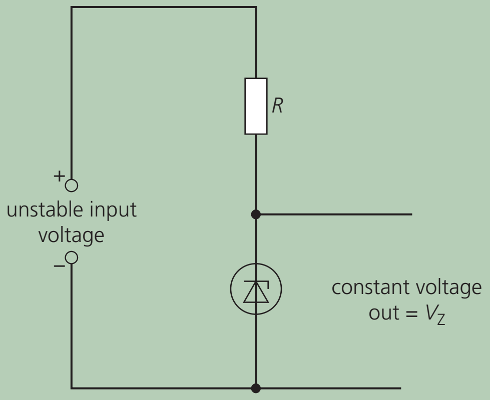


Figure Q2

Calculate the current through this resistor when the USB device is drawing its maximum current of 80 mA and the power socket is at minimum voltage.

c. When revving the car engine, the power socket voltage increases to its maximum of 14.7 V.

- **i.** What is the new voltage across *R*?
- **ii.** What is the current through the resistor under these conditions?
- **iii.** What power is dissipated through the resistor when the socket voltage is a maximum?

3. The spectral response of a photodiode is shown in Figure Q3.

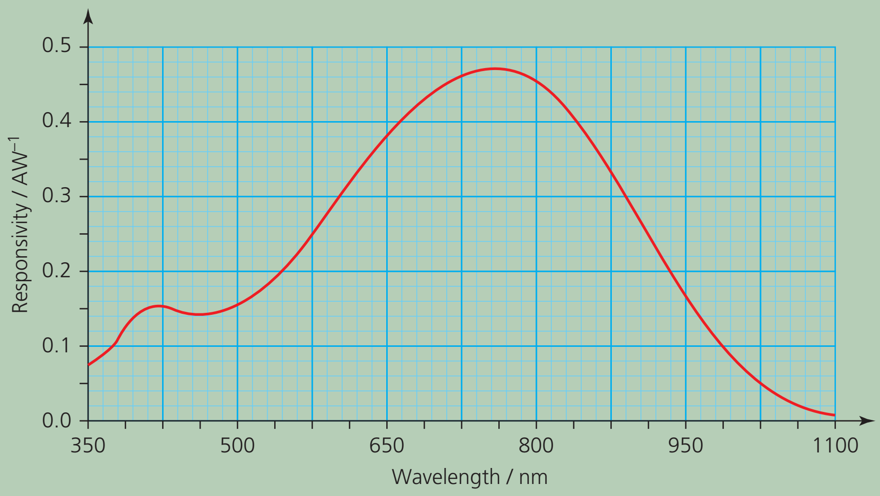


Figure Q3

a. i. At what wavelength does the photodiode produce the greatest photocurrent, for a given incident light intensity?

ii. Light of power 10 μW and wavelength 650 nm is incident on the photodiode. What photocurrent does it generate?

b. A photodiode scintillator has a light yield of 44 × 10^3 photons/MeV. Each photon has an energy of 3.3 eV. An alpha particle of energy 6.8 MeV is incident on the scintillator.

- **i.** What is the efficiency of the scintillator?
- **ii.** A photodiode coupled to the scintillator detects light photons of wavelength 458 nm. What is the energy of the photons produced in the scintillator?

2 ANALOGUE AND DIGITAL SIGNALS

PRIOR KNOWLEDGE

You will have a good understanding of the basic ideas of wave motion, including the meanings of the terms amplitude, period and frequency. You should remember the physics of the motion of masses oscillating on springs and of resonance in a driven oscillator (Chapter 2). You will also need a knowledge of electromagnetic induction (Chapter 8) and capacitance (Chapter 6).

LEARNING OBJECTIVES

In this chapter you will learn about the two kinds of electronic signal (analogue and digital), how analogue information can be converted into a digital form that represents that information, and about the advantages and disadvantages of each. Then you will find out about *LC* resonant filters, which are circuits that respond selectively to particular frequencies, much like a harmonic oscillator.

(Specification 3.13.2.1, 3.13.3.1)

2.1 SIGNALS

Electronic communication signals are either analogue or digital. A signal in electronic terms is an electric current that represents information, and this is derived from a varying voltage.

An **analogue signal** is continuous: it can have an infinite number of values, within limits. The simplest analogue signal waveform is a sine wave (Figure 1), but analogue waveforms usually have more complex shapes.

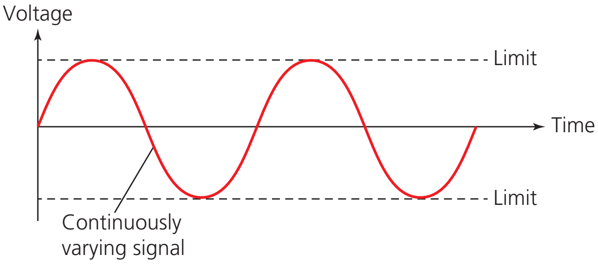


Figure 1 *An analogue signal. A voltage that varies as a sine wave is a simple analogue waveform.*

A **digital signal** is not continuous: it has just two states, and switches straight from one to the other, as shown in Figure 2. These states are 'high' (non-zero voltage, or ON) and 'low' (low or zero voltage, or OFF), and they can be represented by 1 and 0. The actual voltage value of the ON state does not matter as long as the two states can be distinguished. Common values are 5 V for a high state and 0 V for a low state. The time for which the signal is in an ON state and an OFF state can be the same or different, and are typically in the range from nanoseconds to milliseconds, but can be longer.

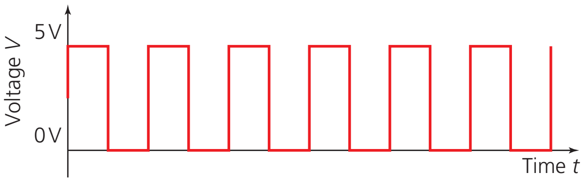


Figure 2 *A digital signal. The high state here is 5 V and the low state is 0 V.*

Bits and bytes of digital data

A **bit** is the basic unit of information both in computer systems and in digital communications. A bit can have just two values: 1 or 0.

A **byte** (B) is a unit of information that is eight bits long. Multiples of bytes are kilobytes (kB), megabytes (MB) and gigabytes (GB). We need to be careful about our use of unit multiples here. In data transmission, the usual meaning of the prefixes applies: kilo = 10^3, mega = 10^6, giga = 10^9. But in relation to memory storage capacity in computers, the prefixes have different meanings: 1 kilobyte is not 1000 bytes but 1024 (2^{10}) bytes; 1 megabyte is 1 048 576 (2^{20}) bytes or 1024 kilobytes; 1 gigabyte is 1 073 741 824 (2^{30}) bytes or 1024 megabytes and 1 terabyte is 1 099 511 627 776 (2^{40}) bytes or 1024 gigabytes.

The reason for this difference is that binary numbers are used to represent digital signals numerically. Unlike the decimal system, which uses powers of 10 to represent numbers with 10 different symbols (0 to 9), the binary system uses powers of 2 (Table 1).

128s	64s	32s	16s	8s	4s	2s	1s
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	0	0	0	0	1	1	1

Table 1 *The decimal number 7 is represented by 111 in binary.*

Table 1 shows the decimal number 7 represented by $2^0 + 2^1 + 2^2 = 1 + 2 + 4 = 7$ (decimal) or 111 (binary). A combination of just two states, 0 and 1, can represent any numerical value (provided we have enough powers of 2), which makes binary ideal for digital signals. Notice that three bits are sufficient in binary to represent the decimal number 7. To represent higher numbers, we need more bits. Table 2 shows the decimal numbers 0 to 10 and their binary equivalents.

Decimal number	Binary number
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
10	1010

Table 2 *The decimal numbers 0 to 10 and their binary equivalents.*

A digital signal consisting of a pattern of low–high states can therefore represent a binary number. So, for example, the pattern of states made up of 11 bits shown in Figure 3 represents the binary number

$$\begin{aligned} \mathrm{L\,H\,L\,L\,H\,H\,L\,L\,L\,H\,H} &= 01001100011 \text{ (binary)} \\ &= 611 \text{ (decimal)} \end{aligned}$$

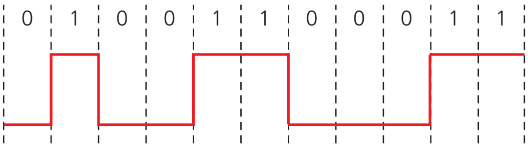


Figure 3 *An 11-bit digital signal*

QUESTIONS

1. Sketch the digital signal that represents each of the following decimal values.
   - **a.** 3
   - **b.** 5
   - **c.** 10
2. How many bits are there in
   - **a.** a computer memory hard drive of 500 gigabytes
   - **b.** a data transmission system that transfers 20 megabytes per second?
3. The Short Message Service (SMS) on a mobile phone can send a maximum of 160 characters per message. Each character consists of seven bits. How many bits are in the following message?

   GOOD MORNING
4. What is the highest decimal number that can be represented by four bits?

Stretch and challenge

5. What are the binary equivalents of each of the following decimal numbers?
   - **a.** 16
   - **b.** 32
   - **c.** 64

KEY IDEAS

- An analogue signal consists of a continuous waveform that can take any value, within limits.
- A digital signal consists of two states: high and low.

- Digital signals can be represented as bits: a series of 1s and 0s, and so can be described numerically using binary.
- One byte of information consists of eight bits.

2.2 SIGNALS AND NOISE

Humans communicate using analogue information. Analogue audio signals intelligible to humans, such as speech and music, are continuous streams of varying frequency and amplitude. To receive the signal correctly, the exact value at every moment must be detected and interpreted. So the information carried by an analogue signal can be badly affected by any slight distortion to the wave.

Figure 4 shows an analogue audio waveform, which consists of many waves of different frequencies superimposed on one another. The signal will be a complex mix of different frequencies and will include **noise**. Noise in a speech waveform may be unwanted background sounds. Noise in a voltage–time signal, for example that produced from the speech signal by a microphone, consists of random voltage fluctuations at different frequencies embedded in the signal. These will be due to **interference** from the electronic system that processes the signal as well as from external sources.

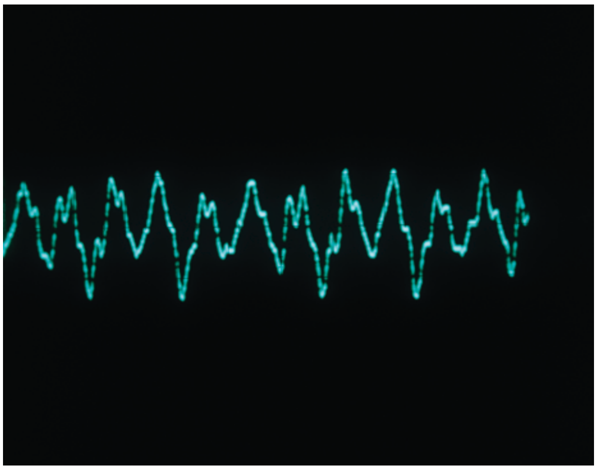


Figure 4 *An analogue audio waveform is often complex.*

Usually, the noise level is much smaller than the signal level (Figure 5), but noise can become noticeable as hiss and crackle in a received audio signal. It adds extra random information to the signal.

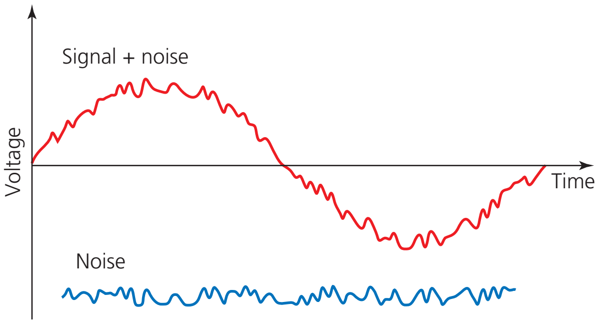


Figure 5 *Random noise is superimposed on an analogue signal.*

The processing of analogue signals requires amplification, and each time the signal is amplified, the noise is also amplified. At each stage, the signal becomes less and less like the original signal. Eventually, it may be impossible, for example, to make out the music in a radio broadcast from the background noise. The noise on an analogue signal is randomly mixed with the signal itself, so is difficult to clean up.

The comparison of signal level and noise level is expressed as a ratio called the **signal-to-noise ratio** (**SNR**). It may be expressed in a number of ways, such as the ratio of the signal amplitude to the average noise amplitude, or the ratio of signal power P_{signal} to noise power P_{noise}. A logarithmic measure of SNR is more often used, expressed in **decibel** (**dB**) as (*see Assignment 3 in Chapter 7 of Year 1 Student Book*)

$$\text{SNR}_{\text{dB}} = 10\log\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right)$$

Using the properties of logarithms this can be written as

$$\begin{aligned}\text{SNR}_{\text{dB}} &= 10\log(P_{\text{signal}}) - 10\log(P_{\text{noise}})\\ &= P_{\text{signal,dB}} - P_{\text{noise,dB}}\end{aligned}$$

Noise is also present on a digital waveform. However, the information is contained in the pattern of states (1s and 0s), and not in the shape of the signal, as it is with analogue waveforms. Therefore, the information pattern is usually still recognisable even if the shapes of the digital states are distorted and less square (Figure 6). This means that digital signals have inherently high noise immunity. Additionally, since the noise on a digital signal is usually much lower in amplitude than the amplitude of the ON state, the electronic transmission system can filter out the random noise and the original signal can be recovered.

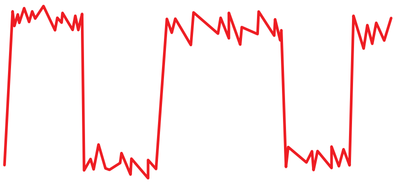


Figure 6 *A noisy digital signal. Despite the noise, the pattern of high–low states can still be recognised.*

One reason for converting data from analogue to digital form (see Electronics section 2.3) is to prevent deterioration by interference during transmission and processing. The ability of digital signals to maintain their quality despite the presence of noise is a major reason why broadcasters are using digital signals more and more for their transmissions (Figure 7).

Figure 7 *DAB (Digital Audio Broadcasting) is a digital radio technology used for broadcasting radio content using digital signals. This exploits the ability of digital signals having very low noise to give a pure sound on reception.*

QUESTIONS

6. Explain what is meant by digital signals having 'higher noise immunity' than analogue signals.

Stretch and challenge

7. Using the properties of logarithms, and the relationship between power and voltage in a resistive load R, show that the signal-to-noise ratio in dB for the signal voltage and noise voltage may be expressed as

$$\mathrm{SNR}_{\mathrm{dB}} = 20\log\left(\frac{V_{\mathrm{signal}}}{V_{\mathrm{noise}}}\right)$$

2.3 CONVERTING ANALOGUE SIGNALS TO DIGITAL SIGNALS

Analogue signals from sensors

Information from the environment is in analogue form. Sensors are devices that detect and respond to a certain type of analogue input, which could be sound, temperature, light, heat or pressure, for example. The sensor is a **transducer** – as well as *detecting* a physical input, it *converts* it into an electrical property, such as resistance or capacitance, from which an analogue voltage output can be transmitted from the sensor location for reading or further processing. Table 3 shows different physical sensor inputs and their associated transducers.

Sensor input	Transducer
Temperature	Thermistor
Light	Photodiode/light-dependent resistor
Sound	Microphone
Mechanical strain	Strain sensor
Magnetic fields	Hall sensor
Pressure	Flow sensor

Table 3 *Sensor inputs and their transducers. The transducer converts the input into an analogue signal that can be amplified for further processing.*

Sensors that produce analogue outputs are characterised by their **sensitivity, resolution** and **response time**.

- The sensitivity is the amount of change in output quantity with unit change in input quantity. Often, this is in terms of the amount of voltage output. But, for some sensors, it can be in terms of current. So a photodiode, for example, has a sensitivity measured in amps per watt ($\mathrm{A\,W^{-1}}$).
- The resolution is the smallest change in the physical property that can be measured by the sensor as a ratio of the value measured, and is dimensionless. So a temperature sensor may be able to measure down to 0.1 °C. If it measures a temperature of 30 °C, then its resolution would be $0.1/30 = 0.003$.
- The response time of a sensor is the amount of time it takes to completely respond to a change in input, often stated as its time to respond from zero input to some specified value.

A voltage output from any sensor that varies with time can, like the audio output from a microphone

(Figure 4), be considered as a complex combination of frequencies. Any such analogue signal has a **bandwidth** that is defined as

$$\text{bandwidth} = \text{highest frequency in signal} - \text{lowest frequency in signal}$$

QUESTIONS

8. **a.** A Hall sensor (see Electronics section 1.5) is stated to have a sensitivity of 2.5 mV/gauss. The gauss is an old unit of flux density, and 1 gauss = 10^{-4} T (tesla). When a magnetic field of 40 mT is detected by the sensor, what is the voltage output?

   b. The minimum magnetic field detectable by the sensor is ± 2 mT. What is the resolution when making the measurement in part **a**?

In many cases, the sensor output needs to be processed by a computer, so at some stage it needs to be converted into digital form by an **analogue-to-digital converter** (**ADC**). An ADC is an integrated circuit that takes an analogue input and converts it into a digital output (Figure 8).

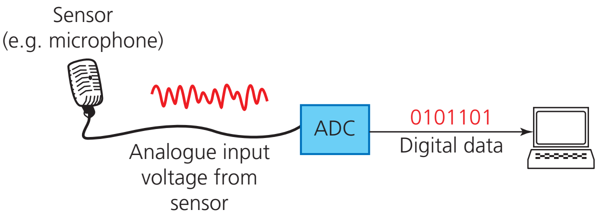


Figure 8 *An analogue audio signal from a microphone is fed via an ADC into a computer for digital processing.*

Sampling an analogue signal

An ADC converts an analogue signal into digital form using a technique called **sampling**. The analogue signal is passed into a **sampling gate**, which samples the value of the signal over an interval determined by an external clock pulse. Samples taken at clock intervals are produced at the output (Figure 9).

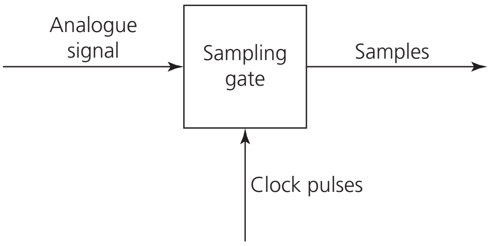


Figure 9 *A sampling gate*

Figure 10 shows how sampling generates a digital output. An analogue signal is sampled at equal time intervals and the voltage level is read off. Figure 10 shows eight sampling times (at intervals of t), with their corresponding voltage values. Each decimal value is converted to its binary equivalent, and a corresponding digital pattern of states results. In this case, only three bits are used to represent each voltage level in steps of 1 V from 0 V to 7 V.

Conversion quality

A sampling scheme such as that in Figure 10 will give a very poor copy of the analogue signal because we have only sampled at eight points and have missed values in between. In order to produce a more accurate replica of the signal, we need to take more samples (Figure 11).

The rate at which samples are taken in practice is controlled by a sampling theory first defined by Harry Nyquist. This states that, in order to obtain a reconstructed signal that is recognisable:

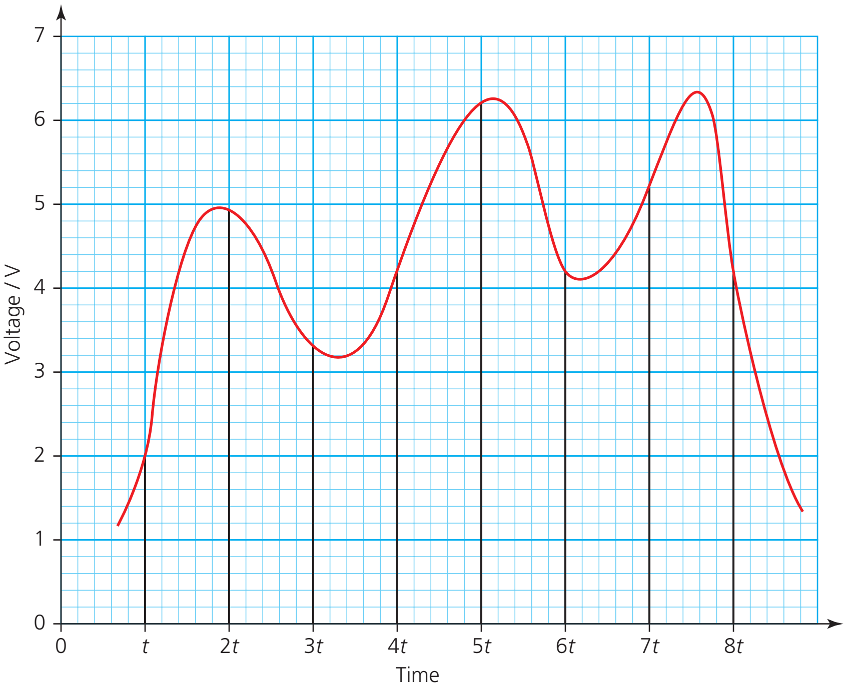
> The **sampling rate** must be at least twice (2 ×) the highest information frequency in the signal bandwidth.

For example, if an analogue speech signal with a bandwidth that ranges from 300 Hz to 3000 Hz is to be sampled and converted into a recognisable digital signal, the sampling rate must be at least $2 \times 3000\,\text{Hz} = 6000\,\text{Hz}$.

A second factor in the analogue-to-digital conversion process that affects quality is **quantisation** – the number of available voltage values. This is determined by the number of bits (0s and 1s) used for each sample. In Figure 10 the number of bits used was three. The number of voltage values possible with three-bit sampling is eight (0, 1, 2, 3, 4, 5, 6, 7). When a sample is taken, the instantaneous snapshot of its analogue value has to be rounded or quantised to the nearest available digital value. The difference between the analogue value and the digital value is called the approximation error or quantising error.

In Figure 10 the resulting digital signal cannot represent a fraction, so the voltage is read to the nearest integer and the fraction ignored; for example, 6.3 V is read as 6 V. This value of 6 is converted into, or **encoded** as, the binary number 110, and a binary voltage signal is generated from it.

If 0.3 V is too large a quantising error, the accuracy needs to be increased with smaller voltage intervals, for example, millivolts. The voltage 6.3 V now reads as 6300 mV. This requires more binary



Decimal voltage value / V	2	5	3	4	6	4	5	4
Transmitted binary value	010	101	011	100	110	100	101	100
Signal pulses								

Figure 10 *Sampling an analogue signal. The sampling rate is 1/t.*

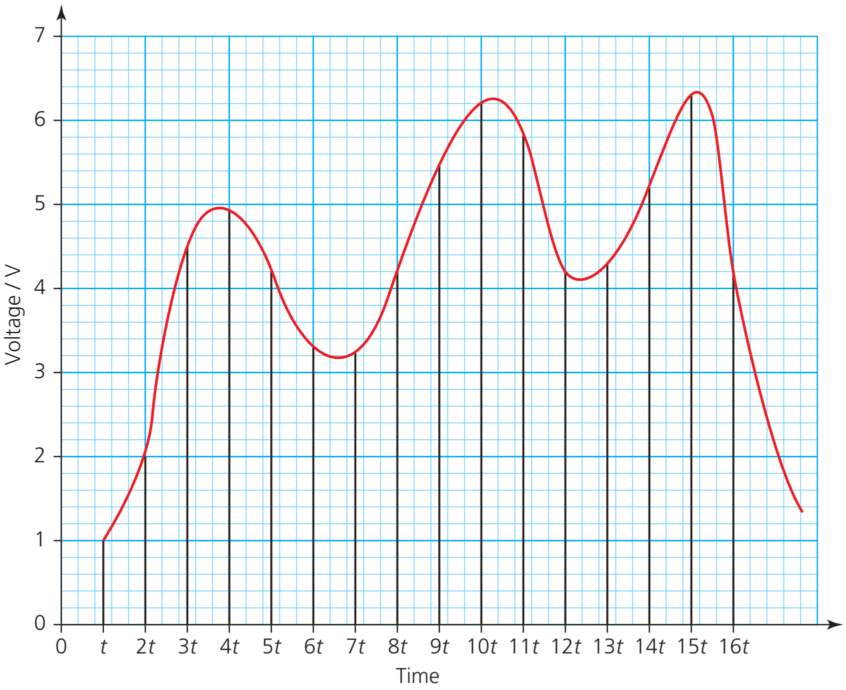


Figure 11 *Here the sampling rate has been doubled compared with that in Figure 10.*

digits for its encoding, since 6300 is equivalent to 1100010011100 in binary.

So if we want to increase the number of values that we can read, then we need to increase the number of bits that represent each value:

> The number of values possible with N bits is 2^N.

So eight-bit sampling gives $2^8 = 256$ possible values that could be read, and 16-bit sampling gives $2^{16} = 65\,536$ values.

If the digital signal becomes distorted in shape, with it edges sloping out due to noise or interference, then, provided the sampling rate of the original signal is known, the ADC can sample at the same rate and correct for the sloping edges, so maintaining the integrity of the information.

Pulse code modulation

There are three stages in the analogue-to-digital conversion process of an analogue signal:

- Sampling is the process of reading the values of the analogue signal at equal time intervals at a constant sampling rate.
- Quantisation is the process of assigning a discrete value from a range of possible values to each sample obtained.
- Encoding is the process of representing the sampled values as a binary number with a number of bits in the range 0 to N. The value of N is chosen as a power of 2, depending on the accuracy required.

The three-stage process, which is called **pulse code modulation**, is illustrated in Figure 12. The resulting stream of binary numbers is called a pulse code modulated (PCM) signal. The **bit rate** of a PCM signal is the number of bits generated per second during sampling. This is given by

$$\text{bit rate} = \text{number of bits used in quantisation} \times \text{sampling frequency}$$

The unit is bit s^{-1} but more usually it will be quoted in kbit s^{-1} or Mbit s^{-1}. These are sometimes presented as kbps and Mbps, respectively.

Worked example

An analogue signal undergoes pulse code modulation. What is the bit rate of the resulting digital signal if the sampling rate is 20 kHz and a quantisation of 16 bits is used?

$$\begin{aligned}\text{Bit rate} &= 20\,000 \times 16 = 320\,000\ \text{bit s}^{-1}\\ &= 320\ \text{Mbit s}^{-1}\end{aligned}$$

QUESTIONS

9. A rock band in a recording studio produces a single analogue music signal containing frequencies ranging from 50 Hz to 16 500 Hz when mixed by a recording engineer. It is sampled using a 32-bit sound card in the mixing desk to produce PCM output.

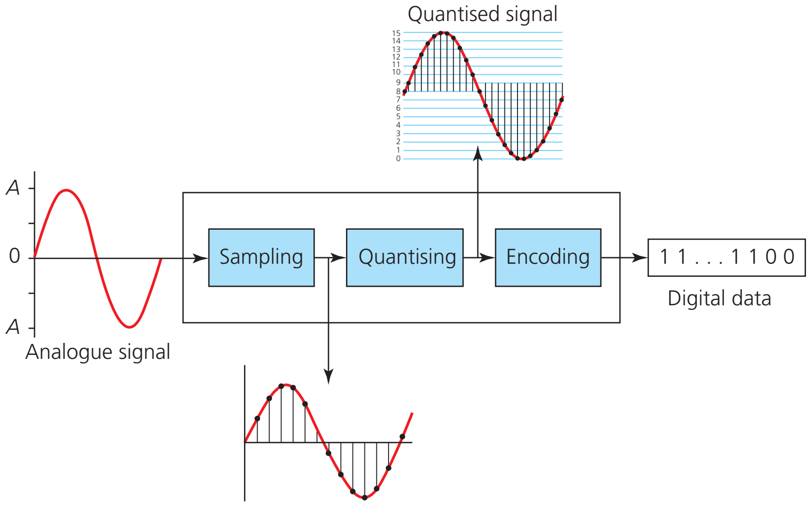


Figure 12 *Pulse code modulation*

a. What is the bandwidth of the signal that has been mixed?
b. What minimum sampling rate must be used to produce a recognisable copy of the music signal?
c. What is the maximum range of values that can be assigned to the different voltage levels in the analogue music signal?
d. What is the bit rate?

KEY IDEAS

- The information carried by digital signals is much less affected by noise than analogue information.
- A sensor is a transducer that senses the environment and produces an analogue voltage output.
- An analogue signal can be converted into a digital signal for transmission using an analogue-to-digital converter.
- The conversion is achieved by sampling, quantisation and encoding. This three-stage process is called pulse code modulation.
- To obtain a reconstructed digital signal that is recognisable, the sampling rate must be at least twice (2 ×) the highest information frequency in the bandwidth of the analogue signal.
- The higher the sampling rate and the higher the number of quantised voltage levels, the better the quality of the digital signal.
- A large number of quantisation levels requires encoding with a large number of bits.
- The bit rate of a digital transmission is

  bit rate = number of bits used in quantisation × sampling frequency

2.4 ADVANTAGES AND DISADVANTAGES OF DIGITAL SIGNALS

Advantages of digital signals over analogue signals:

- The high immunity to noise of a digital signal means that it maintains its quality over long distances during transmission.
- When a digital signal is affected by a high level of noise, the original signal can be recovered.
- More information can be sent by digital signals than analogue signals in the same time, using the same method of transmission (for example, cables, radio waves).
- The transmitted signals can be interfaced readily with other digital systems such as computers.
- Digital signals enable optical transmission through optical fibres with added advantages of security, freedom from interference and lighter cabling.

Disadvantages of digital signals compared with analogue signals:

- The information is not as exact as that of analogue signals, because discrete values are used to form the signal.
- The systems for transmission and processing can be more complex to build than for analogue systems.

2.5 *LC* RESONANCE FILTERS

A **filter** is a circuit that is capable of selectively filtering one frequency or a range of frequencies out of a mix of different ac frequencies in a circuit. One example is a filter circuit that filters audio signals to high-performance loudspeakers in a stereo system so that low frequencies go to a bass speaker called a ‘woofer’ and higher frequencies to a second loudspeaker called a ‘tweeter’.

Another example of a filter circuit is that used in radio and television reception. Both digital and analogue broadcasting stations transmit over a wide range of frequency bands, and it is necessary for the receiver to accept one frequency, corresponding to a particular broadcasting station. The filter circuit allows the receiver to ‘tune in’ to a specific signal frequency by filtering out others (Figure 13). In a digital radio, the tuning component is analogue, but the frequency is digitally selected by the user.

One way of constructing a filter is the use of an ***LC* circuit** (or ‘*LC* resonant circuit’ or ‘inductor–capacitor circuit’). This is an electric circuit made up of an inductor *L* and a capacitor *C* connected together in parallel, as shown in Figure 14, used to filter out frequencies when processing analogue signals.

Figure 13 *Tuning in to a radio station involves using a filter circuit, which filters out all frequencies except that of the broadcasting station frequency.*

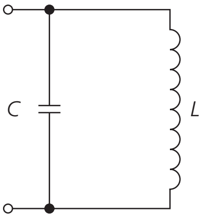


Figure 14 *A parallel LC circuit*

To explain how an *LC* resonant circuit works, we need to understand the electrical properties of an **inductor**. An inductor is a passive electronic component that stores energy in the form of a magnetic field. It consists of a coil of wire, sometimes wound around an iron core. An inductor and its circuit symbol are shown in Figure 15.

The action of an inductor can be explained using Faraday's laws of electromagnetic induction (Chapter 8). When a varying current starts to flow through an inductor from a source of emf, a varying magnetic field is created. This varying magnetic field induces an emf in the coil itself, generating an induced current, which opposes the changing primary current, according to Lenz's law.

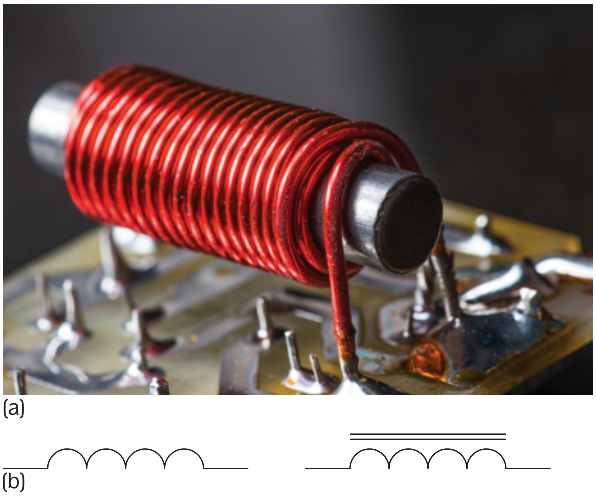


Figure 15 *An inductor and its circuit symbol. The bars on the second symbol indicate that the inductor has an iron core.*

The **inductance** *L* of a coil is a measure of its 'resistance' to the change of the current flowing through the circuit; it has the unit **henry** (H). A coil has an inductance of 1 H when a current flowing through it that changes at a rate of 1 A per second induces a voltage of 1 V across it. Reduced to SI base units, the henry has the equivalence $1\,\text{H} = 1\,\text{kg}\,\text{m}^2\,\text{s}^{-2}\,\text{A}^{-2}$.

Operation of an *LC* circuit

The operation of an *LC* circuit is shown in Figure 16 and described below.

1. When the switch S is in position 'a', the capacitor is connected to a source of emf, there will be a pd across the plates of the capacitor and the capacitor will charge up.
2. When S is disconnected, the capacitor stores electrostatic potential energy because it has an electric field across it due to the accumulation of charge on its plates.

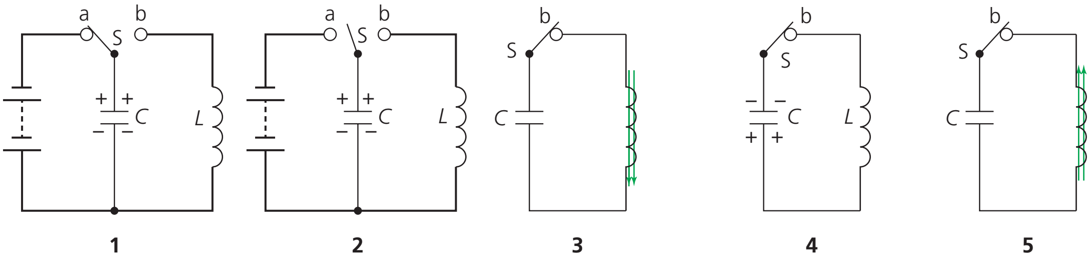


Figure 16 *Operation of an LC circuit*

3. When S is in position 'b', the current from the capacitor will start to flow through the inductor and a magnetic field will build up. As the capacitor discharges, the electrostatic potential energy stored in the capacitor is transferred as magnetic field energy in the inductor.

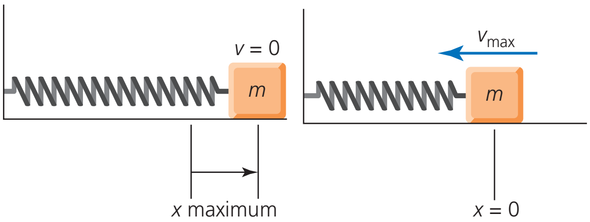
   When the capacitor is completely discharged and no more current flows, the magnetic field in the inductor starts to collapse and an emf is induced, which, according to Lenz's law, produces a current in the opposite direction.

4. The capacitor will now charge up, but with opposite polarity. The energy stored in the magnetic field is converted back into electrostatic potential energy in the capacitor, which charges to the voltage it had to start with.
5. When the capacitor is fully charged, the cycle repeats, but the capacitor now begins to discharge current through the inductor in the opposite direction.

The sequence of charging and discharging and conversion from electrostatic to magnetic field energy goes back and forth, and the *LC* circuit is said to oscillate.

An oscillating *LC* circuit can be compared to a mass oscillating on a spring (Figure 17).

- The *LC* circuit in part 4 of Figure 16, with energy stored only in the capacitor as electrostatic potential energy, is analogous to a stretched spring at its maximum displacement (Figure 17a). Energy is stored here as elastic potential energy in the spring, and the mass is stationary.
- The *LC* circuit in part 5 of Figure 16, with energy stored entirely in the magnetic field of the inductor, is analogous to the spring being at its equilibrium length (Figure 17b). The mass moves with maximum velocity and all the energy is in the form of its kinetic energy.



(a) (b)

Figure 17 *An LC circuit is analogous to the behaviour of a mass–spring system. (a) Energy is all stored as elastic potential in the spring, and the mass is stationary with zero kinetic energy. (b) Elastic potential energy is entirely converted into kinetic energy of the mass, which has maximum velocity at zero displacement.*

QUESTIONS

10. An *LC* circuit may be regarded in an analogous way to a mass–spring system. What physical quantities correspond to

- **a.** the mass
- **b.** the spring?

Frequency of oscillation of a parallel *LC* circuit

Any oscillating system will have a **natural frequency**, which is the frequency at which a system tends to oscillate in the absence of any driving force. But for the oscillations to continue, a driving force is needed, supplying energy, otherwise the amplitude of the oscillations will decrease as energy is lost. In a mass–spring system, for example, energy is lost due to frictional forces while in an *LC* circuit, energy is dissipated due to electrical resistance. The driving force should be periodic, matching the natural period of oscillation. A playground swing, for example, will be most sensitive to a pushing force that comes regularly with the natural frequency of the to-and-fro motion of the swing. A pushing force that comes at a different frequency will not build up as large an amplitude because it will sometimes oppose the motion of the swing. If the driving force is applied with a frequency close to or equal to the natural frequency of an oscillating system, then large-amplitude oscillations will build up in the system, and we call this **resonance** (see section 2.4 of Chapter 2).

In the same way, oscillations in a parallel *LC* circuit can be driven by an alternating frequency source of emf that supplies it with electrical energy (Figure 18). The charge oscillates backwards and forwards between the inductor and the capacitor *at the frequency of the source*. If the frequency of the source of emf is the same as the natural frequency of the parallel *LC* circuit, then resonance will occur, and there will be an increase in current at that frequency. The natural frequency or **resonant frequency** f_0 of an *LC* circuit is given by

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

If L is in henry (H) and C is in farad (F), then f is in hertz (Hz, s^{-1}).

Worked example 1

What is the resonant frequency of the parallel *LC* circuit with $C = 10\,\mu\text{F}$ and $L = 100\,\text{mH}$ in Figure 18?

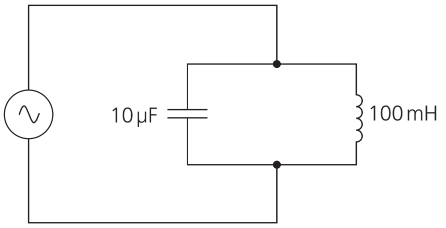


Figure 18 *A parallel LC circuit driven by an alternating current*

$$f_0 = \frac{1}{2\pi \times \sqrt{(100 \times 10^{-3}) \times (10 \times 10^{-6})}} = 159\,\text{Hz}$$

Energy response curve in an *LC* circuit and *Q* factor

Figure 19 shows the energy response curve of a parallel *LC* circuit with the frequency of the driving source of emf. At the resonant frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$, the energy stored in the circuit is a maximum, and the voltage across the *LC* circuit is at its resonance peak. The voltage response curve is the same shape – the voltage also has a maximum at the resonance peak at f_0.

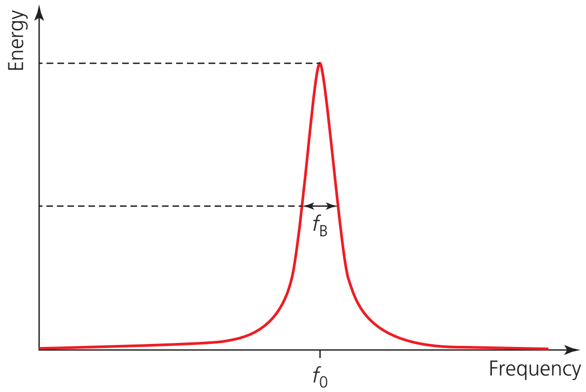


Figure 19 *The energy response curve of a parallel LC circuit*

The resonance peak has a certain **bandwidth** f_B, which measures its 'sharpness'. This is the frequency range between the 50% maximum energy (or voltage) points on the response curve. (This is not to be confused with a signal bandwidth – see Electronics section 2.3.)

A dimensionless quantity called the 'quality' or ***Q* factor** of the resonant circuit is defined as

$$Q = \frac{f_0}{f_B}$$

High-*Q* *LC* resonators have very narrow bandwidths, whereas low-*Q* resonators have wide bandwidths.

The analogy between the driven *LC* circuit and a driven mass on a spring is again apparent. The amplitude–frequency curve for forced oscillations of a mass on a spring is the same shape as the *LC* circuit response curve – see Figure 26 in section 2.4 of Chapter 2. The effect of damping on the sharpness of the resonance in the mechanical system has its analogy in the electrical resistance of the *LC* circuit.

Worked example 2

A parallel resonant *LC* circuit has a capacitance of 0.1 μF and an inductance of 2.0 μH. What is the *Q* value if the bandwidth at the 50% points is 5.0 kHz?

Resonant frequency is

$$f_0 = \frac{1}{2\pi \times \sqrt{(2.0 \times 10^{-6}) \times (0.1 \times 10^{-6})}} = 356\,\text{kHz}$$

The *Q* value is $Q = \frac{f_0}{f_B} = \frac{356 \times 10^3}{5.0 \times 10^3} = 71.2$

LC resonant circuit as a tuning filter

As we have mentioned, an *LC* resonant circuit may be used as a tuning filter to pick up a radio broadcast. A radio wave is an electromagnetic wave that has a transverse electric field oscillating at a particular frequency. When this is picked up by an aerial, an alternating voltage is produced that drives a current through the receiving circuit, which includes a parallel *LC* resonator (Figure 20).

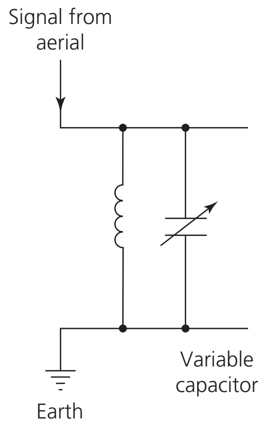


Figure 20 *An LC resonant circuit in a radio aerial*

By replacing the fixed capacitor in the parallel *LC* circuit with a variable capacitor, the *LC* circuit can be 'tuned', so that its resonant frequency changes and it only responds to the frequency of the radio wave required to be picked up. When you tune a radio, you are adjusting the value of a variable capacitor so that the *LC* circuit in the radio will resonate at a particular transmission frequency. Then a peak voltage response is produced across the circuit. Other transmission frequencies are filtered out. The resonant voltage response can be amplified and further processed by the receiver to give an output.

QUESTIONS

11. a. Compare the *Q* factors of two *LC* resonant circuits, both of which have the same resonant frequency of 1 MHz, and bandwidths of **i** 1 kHz and **ii** 10 kHz at the 50% energy point.

b. Which has the sharper resonance peak?

12. As well as in receiving aerial circuits, *LC* circuits can be used in radio-frequency oscillators to produce a radio wave. If a radio frequency of 20 MHz is to be produced, using a 0.1 μH inductor, what value of capacitor is required?

KEY IDEAS

- An *LC* parallel resonant circuit is an inductor and a capacitor connected in parallel.
- An inductor has inductance *L* measured in the unit henry (H).
- In an *LC* circuit, energy can be exchanged between the electric field of the capacitor and the magnetic field of the inductor, causing the current to oscillate to and fro.
- An input of alternating emf at the resonant frequency of the *LC* circuit causes the *LC* circuit to resonate. The energy of the circuit reaches a maximum.
- The resonant frequency is given by

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

- The *Q* factor is a measure of the sharpness of the resonance peak and is a dimensionless quantity:

$$Q = \frac{f_0}{f_B}$$

where f_B is the bandwidth of the circuit at the 50% voltage (or energy) points.

PRACTICE QUESTIONS

1. a. In computer networking, each computer is assigned an Internet Protocol (IP) address. An IP address contains 32 bits, which are grouped into bytes.

i. Explain the difference between a *bit* and a *byte*.

ii. How many bytes are there in an IP address?

b. i. Explain the difference between an *analogue signal* and a *digital signal*.

ii. Why do digital signals have greater noise immunity?

c. i. What is meant by the *signal-to-noise ratio* of a signal processing system?

ii. A digital signal passing through a noisy communications system has a 'high value' of 5 V. If the average noise level on the digital waveform is 200 mV, what is the signal-to-noise ratio of the system in dB?

2. a. Explain what is meant by *pulse code modulation*.

b. Figure Q1 shows part of a pulse code modulation system.

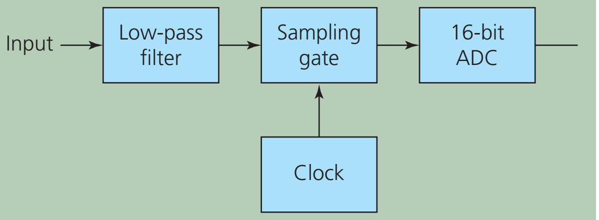


Figure Q1

The analogue signal is filtered by the low-pass filter so it contains frequencies in the audio range between 20 Hz and 20 kHz.

i. What is the signal bandwidth of the filtered signal?

ii. The output of the low-pass filter and the output of the clock are shown in Figure Q2a and b. Draw a graph of the output of the sampling gate against time, using the same scales as in Figure Q2a.

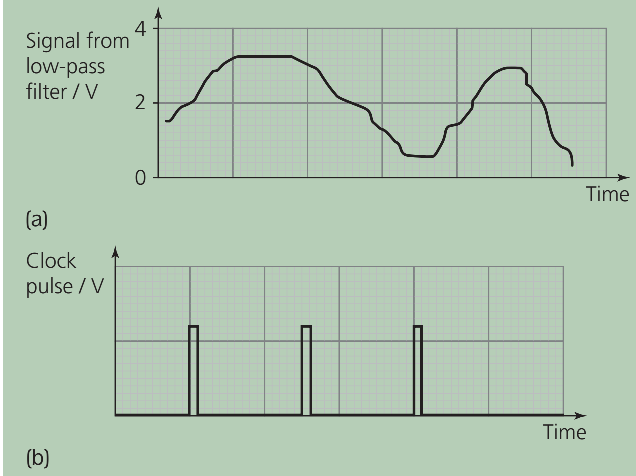


Figure Q2

iii. What is the minimum sampling frequency needed for the system to reproduce a recognisable copy of the filtered signal?

iv. The 16-bit analogue-to-digital converter (ADC) has an input voltage range of 0 to 10 V. What is the quantisation and minimum voltage resolution of the system?

v. How could the digital reproduction of the analogue signal be made more accurate?

3. An engineer uses copper cable to connect an intercom system between her office and workshop. The signals have to travel a long distance and she finds that interference (hum) from the mains supply is a problem. She reduces the interference using a filter tuned to the frequency of the mains supply. The mains frequency is 50 Hz.

Figure Q3 shows her solution which is based on a parallel *LC* resonant circuit.

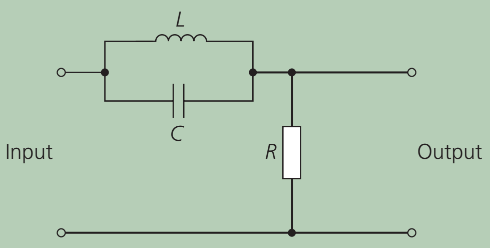


Figure Q3

a. The engineer uses a 2.0 H inductor. Calculate the required value for *C* for the filter to operate at 50 Hz.

Figure Q4 is the response curve for the inductor–capacitor circuit which shows how the pd *V* across the inductor–capacitor circuit varies with frequency.

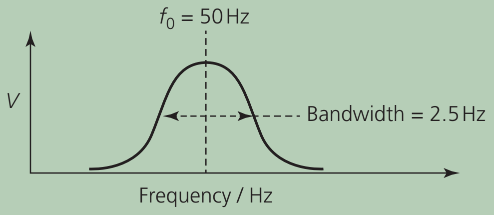


Figure Q4

b. Calculate, from the graph, the *Q* factor of the inductor–capacitor circuit.

c. The inductor is replaced by one that has an inductance of 8.0 H and a lower resistance than that of the original inductor. The capacitor is not changed. Describe how this change affects the response curve of the inductor–capacitor circuit.

AQA Paper 3BE Specimen 2014 Q4

3 OPERATIONAL AMPLIFIERS

PRIOR KNOWLEDGE

You may need to review your understanding of potential divider circuits (*see Chapter 14 of Year 1 Student Book*). You will need to be familiar with the Zener diode, which was discussed in Electronics Chapter 1.

LEARNING OBJECTIVES

In this chapter you will learn that the operational amplifier (op-amp) is an integrated circuit amplifier and how it can be used in a variety of ways to perform many useful operations in electronic systems. You will learn about the characteristics of an op-amp, such as its voltage gain, in different circuit configurations, how its gain can be controlled by negative feedback, and how a real op-amp differs from an 'ideal' one.

(Specification 3.13.3.2, 3.13.4.1–3.13.4.4)

3.1 THE IDEAL OPERATIONAL AMPLIFIER

Amplifiers are devices that take a relatively weak analogue signal as an input and produce a much stronger analogue signal as an output. An **operational amplifier** (**op-amp**) is an integrated circuit amplifier that can be configured in a number of ways to provide amplification of analogue signals for different tasks. In electronic systems it is considered as a 'system block' to which external components can be connected in various ways. We are concerned here not with what is inside an op-amp but with what tasks it performs in a system.

The symbol for an op-amp is shown in Figure 1. For the 'non-inverting input' V_+, the output changes in the same direction as the input applied at this terminal. For the 'inverting input' V_-, the output changes in the opposite direction. Note that, in circuit diagrams, the non-inverting input can be at the top and the inverting input at the bottom, or vice versa.

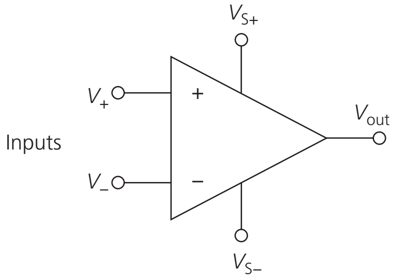


Figure 1 *The symbol for an operational amplifier (op-amp)*

The power supplies to an op-amp (V_{S+} and V_{S-}) may have the same magnitude but opposite values, or one (usually the V_{S-} supply) may be connected to zero volts (ground), depending on the circuit application. Also, note that there are also two other connections to the op-amp (not shown in Figure 1) called 'offset null', which are there to cancel out small internal voltages at the output generated by the op-amp itself.

An important characteristic of any amplifier is its voltage gain, or simply **gain**, A, which is defined as

$$A = \frac{\text{output voltage}}{\text{input voltage}}$$

In an op-amp, the gain can be expressed in two ways:

- **open-loop gain** A_{OL}, where no part of the amplifier's output is fed back to the input
- **closed-loop gain** A_{CL}, where a fraction of the amplifier's output is fed or 'looped back ' to its input (see Electronics section 3.3).

We contrast the characteristics of an *ideal* op-amp with those of a typical real op-amp in Table 1.

	Ideal op-amp	Typical real op-amp
Open-loop gain A_{OL}	Infinite	In excess of 10^5 or more
Input resistance (between the $V+$ and $V\_$ terminals)	Infinite	10^7–10^{12} or more
Output resistance (this limits the output current in the same way as the internal resistance of a cell)	Zero	Depends on the actual device, but can be in the range from one to a few kΩ, and typically 70 Ω
Maximum output voltage	V_+	Between 1 and 2 V less than V_{S+}
Minimum input voltage	V_-	Between 1 and 2 V more than V_{S-}

Table 1 *Characteristics of ideal and real operational amplifiers*

An op-amp has an open-loop output function called a **transfer function** such that

$$V_{out} = A_{OL}(V_+ - V_-)$$

In other words, the op-amp amplifies the *difference* between its non-inverting and inverting terminals.

Figure 2 shows a plot of the output voltage V_{out} against $(V_+ - V_-)$. The output function is linear. However, since the open-loop gain of the op-amp is so high, the output is soon driven into **saturation** at the supply voltages V_{S+} and V_{S-}. There is no further increase in V_{out} if the difference between V_+ and V_- becomes greater, as shown at points A and B. Since A_{OL} for an op-amp is very high, the transfer function has a very steep slope, equal to A_{OL}.

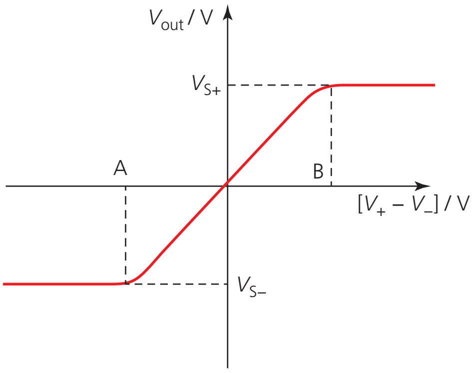


Figure 2 *Open-loop transfer function of an operational amplifier, showing saturation*

3.2 THE OPERATIONAL AMPLIFIER AS A COMPARATOR

The property of amplifying the difference between its input terminals makes the op-amp useful when comparing voltages. The circuit in Figure 3 shows an op-amp being used as a **comparator**.

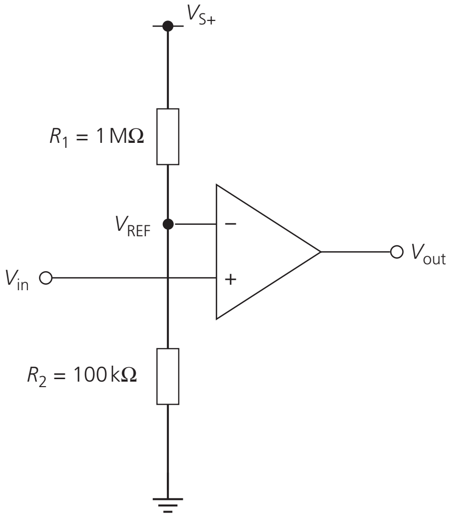


Figure 3 *The operational amplifier as a simple comparator*

Resistors R_1 and R_2 form a potential divider network (*see section 14.5 in Chapter 14 of Year 1 Student Book*), with a reference voltage V_{REF} presented to the inverting input. At the non-inverting input, V_{in} is a varying input that is to be compared with V_{REF}. Because of the large open-loop gain of the amplifier, the output will be saturated except when the value of V_{in} is the same as V_{REF} and the difference between the amplifier's input terminals is zero. When this happens, the transfer function of the op-amp tells us that the output will be zero, so it can be used to detect when V_{in} is equal to V_{REF}.

Worked example

In Figure 3, the value of V_{S+} is 10 V.

a. What is the value of V_{REF}?

b. What value will V_{in} need to be for the output of the comparator to go to zero?

a. The op-amp has infinite input resistance, so it takes no current. Therefore

$$V_{REF} = \left(\frac{10^5}{10^6 + 10^5}\right) \times 10 = 0.9\,\text{V}$$

b. The same as V_{REF}, that is, 0.9 V.

QUESTIONS

1. A practical op-amp has an open-loop gain of 10^5. If the voltage at the non-inverting input is 60 µV and that at the inverting input is 10 µV, what is the value of the output voltage?
2. The circuit in Figure 4 shows an op-amp with the inverting output connected to ground and an alternating voltage applied to the non-inverting input. Sketch the graph of V_{out} against time that you would expect to see, and explain the operation of the circuit.

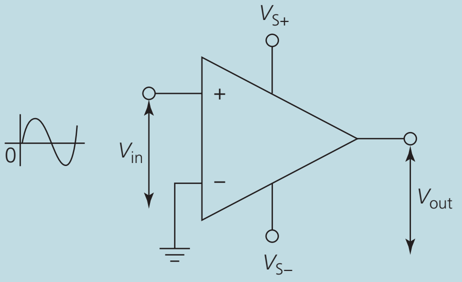


Figure 4

3. Figure 5 shows an op-amp connected as a voltage comparator. D is a Zener diode with $V_Z = 6.2\,V$. The positive power supply to the op-amp is connected to 15 V and the negative supply to zero volts.

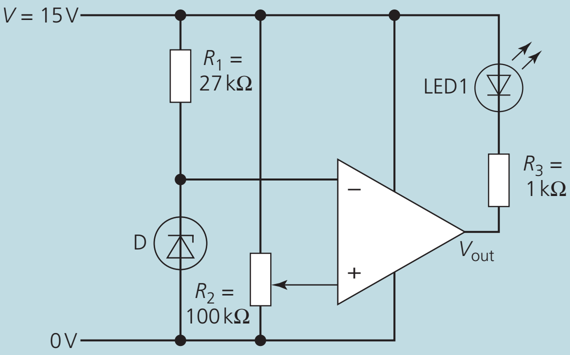


Figure 5

   a. What is the purpose of the Zener diode in this circuit?
   b. Explain what will happen when the slider on variable resistor R_2 is moved from its bottom position to its top position.
   c. What is the purpose of R_1 and R_3?

KEY IDEAS

- An operational amplifier, or op-amp, is an integrated circuit amplifier that is used as a system block in electronic systems to amplify voltages.
- An ideal op-amp has an infinite open-loop gain A_{OL} and infinite input resistance.
- An op-amp has two input terminals, a non-inverting input and an inverting input.
- An op-amp amplifies the difference between the non-inverting and inverting inputs:

$$V_{out} = A_{OL}(V_+ - V_-)$$

- An op-amp can be used as a comparator to compare two voltages. When they are equal, the output is zero.

3.3 THE OPERATIONAL AMPLIFIER AS AN INVERTING AND A NON-INVERTING AMPLIFIER

Feedback

The gain of an op-amp, which on open loop is essentially infinite, can be controlled with the use of **feedback**. If some of the output voltage is fed back to the inverting input, then the voltage gain can be altered to a much lower value.

Inverting amplifier configuration

Consider the circuit shown in Figure 6. Part of the output of the amplifier is fed back via a feedback resistor R_f to the inverting input. In this configuration, the op-amp is called an **inverting amplifier**.

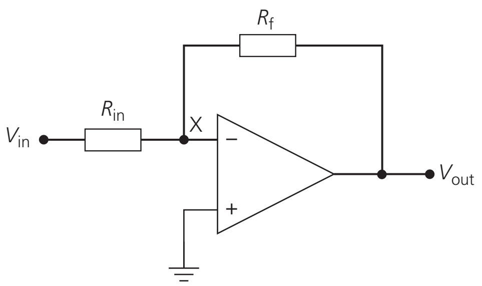


Figure 6 An operational amplifier in the inverting amplifier configuration

We can understand the inverting amplifier by using a **virtual earth analysis**. We assume that V_{out} is at some reasonable value, that is, it is somewhere between the values of the negative and positive supply voltages. The non-inverting input is connected to ground, so its value is zero. So, from the transfer characteristic, we have

$$V_{out} = A_{OL}(0 - V_-) = -A_{OL}V_-$$

Therefore, the voltage at the inverting input is $V_- = -V_{out}/A_{OL}$. Since A_{OL} is a very large value, approximately 10^5, the voltage V_- at the inverting input can be regarded as being zero. We say that the input to the inverting input X is a **virtual earth** because, although it is not directly connected to 0 V (earth), its value is so small as to be considered virtually at earth.

The input to the op-amp itself draws no current, as it has infinite resistance. This means that the current flowing through R_{in} to X must be same as that flowing through R_f. Therefore, from $I = V/R$, we obtain

$$\frac{V_{out}}{R_f} = -\frac{V_{in}}{R_{in}}$$

The negative sign is because the input is the inverting input, so the output will be the opposite value.

Hence the **closed-loop gain** of the circuit is

$$A_{CL} = \frac{V_{out}}{V_{in}} = -\frac{R_f}{R_{in}}$$

and the output is a negative or *inverted* voltage.

Note that this is called the closed-loop gain because there is now a feedback loop from the output to the input of the amplifier. The open-loop gain A_{OL} is the gain quoted when there is no feedback. In this case we are applying **negative feedback** to the op-amp because we are reducing its gain.

Worked example 1

In the inverting amplifier in Figure 7, what is the closed-loop gain?

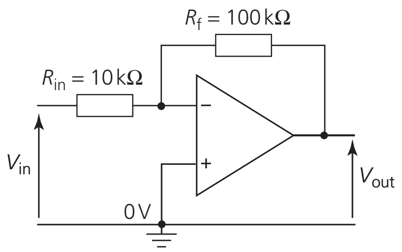


Figure 7

The closed-loop gain (as we are working out a ratio, we can keep the resistance values in kΩ) is

$$A_{CL} = -\frac{100}{10} = -10$$

Non-inverting amplifier configuration

It is possible to configure the op-amp to give a non-inverted output, that is, one that gives a positive voltage gain. This is shown in the circuit in Figure 8. Part of the output of the amplifier is fed back via a simple potential divider made up of R_f and R_1 to the inverting input.

The closed-loop gain in this **non-inverting amplifier** configuration is given by

$$A_{CL} = \frac{V_{out}}{V_{in}} = 1 + \frac{R_f}{R_1}$$

This is again negative feedback.

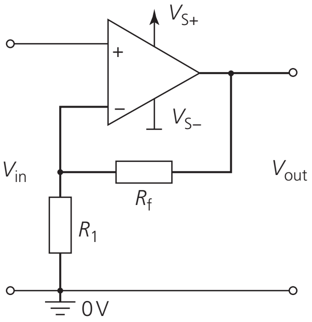


Figure 8 *The operational amplifier in the non-inverting configuration*

Worked example 2

Calculate the closed-loop gain of a non-inverting amplifier, as in Figure 8, where R_f = 100 kΩ and R_1 = 10 kΩ.

The closed-loop gain (again, as we are working with a ratio, we can keep the resistance values in kΩ) is

$$\frac{V_{out}}{V_{in}} = 1 + \frac{100}{10} = 11$$

QUESTIONS

4. Explain the difference between open-loop gain and closed-loop gain for an op-amp.
5. It is required to design an inverting amplifier with a gain of –50. The input resistor has a value of 11.2 kΩ. What should be the value of the feedback resistor?
6. Figure 9 shows an op-amp in the non-inverting configuration, with $R_1 = 4.7\,\text{k}\Omega$ and $R_f = 470\,\text{k}\Omega$. What is the value of the closed-loop gain?

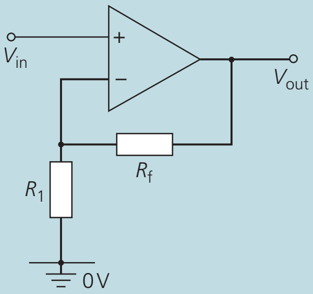


Figure 9

KEY IDEAS

- The gain of an operational amplifier can be controlled by the use of negative feedback.
- The inverting amplifier makes use of negative feedback to amplify signals by a factor equal to the ratio of the feedback resistance divided by the resistance of the input resistor at the inverting input. The output is inverted:

$$A_{CL} = \frac{V_{out}}{V_{in}} = -\frac{R_f}{R_{in}}$$

- For a non-inverting amplifier, the closed-loop gain is given by

$$A_{CL} = \frac{V_{out}}{V_{in}} = 1 + \frac{R_f}{R_1}$$

where R_f and R_1 are resistors that form a potential divider network in negative feedback.

3.4 THE OPERATIONAL AMPLIFIER AS A SUMMING AND A DIFFERENCE AMPLIFIER

Summing amplifier configuration

The op-amp can be used to sum voltages together as a **summing amplifier** using the inverting input. In the circuit in Figure 10, we have three inputs, V_1, V_2 and V_3, to the inverting input, with input resistors R_1, R_2 and R_3, and a feedback resistor R_f.

The output is given by

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

The gain of each input is set by the ratios $-R_f/R_1$, $-R_f/R_2$ and $-R_f/R_3$, so the voltages are effectively summed at the inverting input.

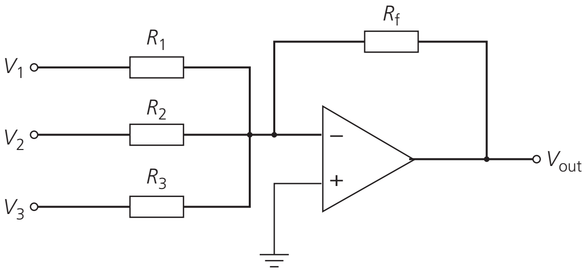


Figure 10 *The operational amplifier as a summing amplifier*

Difference amplifier configuration

An op-amp may also be used to *subtract* two voltages from one another. The circuit shown in Figure 11 is a **difference amplifier**. If the feedback resistor R_f is made the same value as the resistor from the non-inverting input to ground, then the output V_{out} is given by

$$V_{out} = (V_+ - V_-)\frac{R_f}{R_1}$$

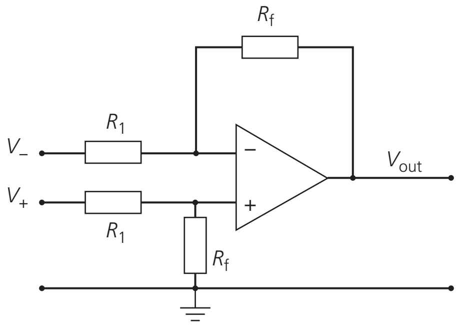


Figure 11 *The operational amplifier as a difference amplifier*

QUESTIONS

7. In a summing amplifier, what value do the input resistors have to be so that all the inputs are amplified with a gain of 1?

8. A difference amplifier has a non-inverting input of 50 mV and an inverting input of 8 mV. The feedback resistor is 10 kΩ. If the input resistors are both 2.2 kΩ, what is the value of the output voltage?

9. Figure 12 shows an op-amp connected as a summing amplifier.

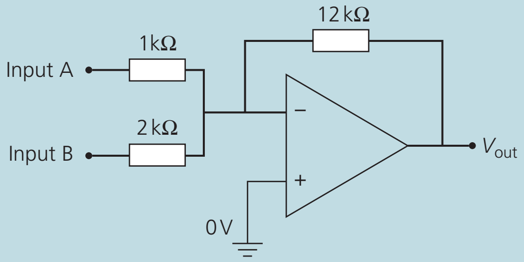


Figure 12

a. If input A is 3 mV and input B is 6 mV, what closed-loop voltage gains do these inputs have?

b. What is the value of V_{out}?

10. Figure 13 shows an op-amp acting as a difference amplifier.

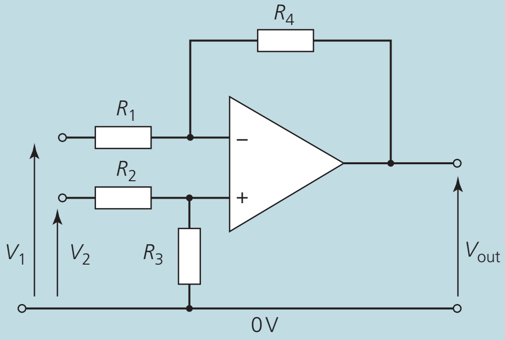


Figure 13

a. What values do R_1, R_2 and R_4 have to be to ensure that inputs to the difference amplifier have a closed-loop gain of 1?

b. If $V_1 = 2$ mV and $V_2 = 5$ mV, and R_1, R_2 and R_4 are all 10 kΩ, what is the value of V_{out}?

KEY IDEAS

- Operational amplifiers can be configured to add two or more voltages and to subtract one voltage from another voltage.
- In summing amplifier configuration:

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

- In difference amplifier configuration:

$$V_{out} = (V_+ - V_-)\frac{R_f}{R_1}$$

where R_1 is the input resistance to the inverting input.

ASSIGNMENT 1: MEASURING THE MECHANICAL STRENGTH OF A MATERIAL USING AN OPERATIONAL AMPLIFIER

(PS 1.1, PS 1.2, PS 3.2, MS 0.2, MS 0.3, MS 2.2, MS 2.3)

In this assignment you will see how an operational amplifier (op-amp) combined with a strain sensor can be used to measure the deformation of materials when under stress.

Stress is the force per unit area exerted on an object. Strain is the deformation an object undergoes while under stress, and is equal to the extension Δl as a fraction of the original length l of the object (*see section 12.1 in Chapter 12 of Year 1 Student Book*).

For mechanical structures, it is important to know by how much they can be stressed and deform without breaking. A sensor called a 'strain gauge' can be bonded to the structure (Figure A1a), so that when the material deforms the gauge deforms in the same way. Strain gauges are sensitive enough to identify minute deformations in the material that the eye cannot see, deformations that could lead to cracks being formed and subsequent failure of the structure.

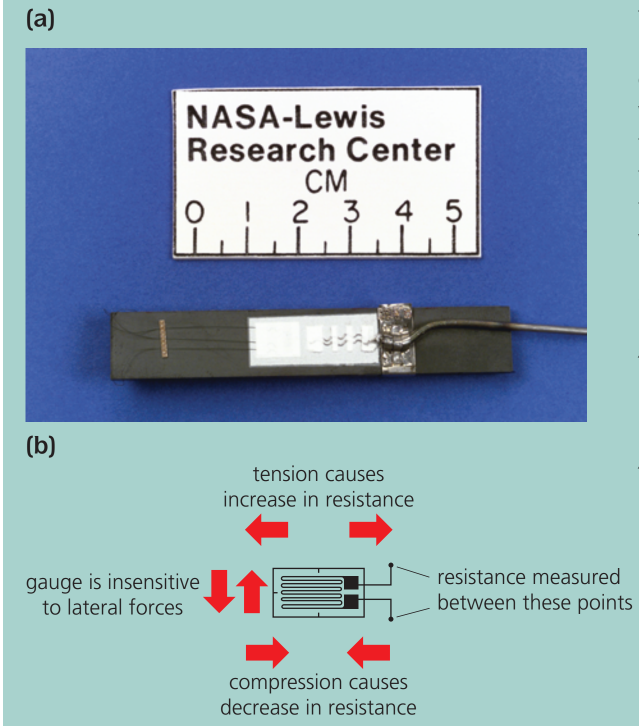


Figure A1 *(a) A strain gauge bonded to a sample of reinforced carbon–carbon composite that was used for the heat-resistant tiles on the space shuttles. (b) How a strain gauge works.*

A strain gauge consists of a thin wire filament or strip of metal foil in a parallel arrangement (Figure A1b). The resistance of a conducting material is directly proportional to the length l and inversely proportional to the cross-sectional area A:

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

where ρ is the resistivity of the material (*see section 13.8 in Chapter 13 of Year 1 Student Book*), which remains constant. Thus, as the wire is stretched or compressed, the resistance of the gauge changes.

The sensitivity of a strain gauge of resistance R is defined by the gauge factor GF, expressed as the fractional change ΔR in electrical resistance divided by the strain:

$$\text{GF} = \frac{\Delta R \,/\, R}{\text{strain}}$$

This is a dimensionless quantity.

Strain gauge resistances when unstressed can range from 30 Ω to 3 kΩ. The change in resistance in use may be only a fraction of a per cent for the range of stresses applied (limited because high stresses would permanently deform both the material and the strain gauge bonded to it). So, in order to use the strain gauge as a practical sensor, there needs to be a way to measure small changes in resistance with high accuracy.

Questions

A1 What will happen to the resistance if a wire or conducting strip is **a** stretched and **b** compressed?

A2 A strain gauge of resistance 120 Ω and GF = 2.0 is attached to the surface of a material. The material is lightly stressed, so that it experiences a strain of 500×10^{-6}. What is the change in resistance?

The Wheatstone bridge

One way to measure a small change in resistance is to use a ‘Wheatstone bridge’ circuit, as shown in Figure A2. This consists of two potential dividers connected to a common voltage V_{in}.

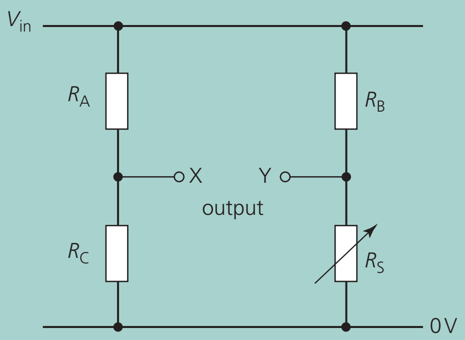


Figure A2 *A Wheatstone bridge including a variable resistor* R_S.

Questions

A3 a. If $R_A = R_B$ and $R_C = R_S$, what can you say about the voltages at X and Y?

b. If a centre-zero voltmeter was connected between X and Y, what value would it read?

c. If the value of R_S is adjusted so that it no longer equals that of R_B, what would you see on the voltmeter?

Using an operational amplifier

Figure A3 shows a Wheatstone bridge connected to an op-amp. If the difference in potential between X and Y is applied across the inverting and non-inverting terminals of the op-amp, this difference, and hence the change in resistance of R_S, can be amplified.

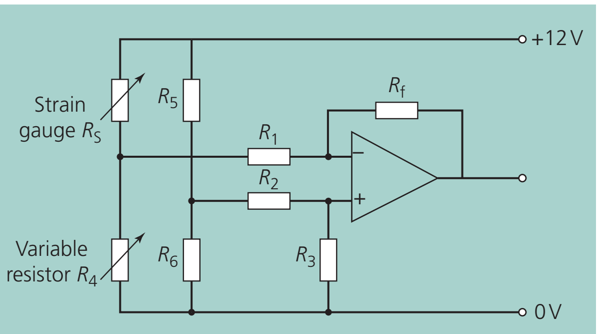


Figure A3 *A strain gauge in a Wheatstone bridge network connected to an operational amplifier.*

Questions

A4 a. In what type of configuration is the op-amp in the circuit of Figure A3?

b. If $R_f = 500\,\text{k}\Omega$ and $R_1 = 10\,\text{k}\Omega$, what is the voltage gain of the circuit?

c. Resistors R_5 and R_6 are both $150\,\Omega$. The strain gauge (R_S) has gauge factor GF = 3.0. It has an unstressed resistance of $150\,\Omega$. The variable resistor (R_4) is adjusted so that the voltages to the inverting and non-inverting inputs of the op-amp are the same.

i. What will be the voltage at the output of the op-amp?

ii. The strain gauge is bonded to a material and stressed so that it experiences a strain of 750×10^{-6}. What is the change in resistance of the strain gauge?

iii. What is the resulting voltage difference at the inputs to the op-amp?

iv. What will be the voltage at the output of the op-amp?

v. If the change in resistance of the strain gauge doubles, what will be the voltage output? Comment on the nature of the amplification.

3.5 REAL OPERATIONAL AMPLIFIERS

There is no such thing as an ideal op-amp. Real op-amps do not have infinite open-loop gain nor infinite input resistance. For high-performance devices, the highest open-loop gain is about 10^8 and the highest input resistance is $10^{13}\,\Omega$.

Also, even when the inputs to the non-inverting and inverting terminals are the same or zero (both grounded), there is still a small output voltage, called the **offset voltage**, which needs to be cancelled or 'nulled out' to ensure proper operation.

Figure 14 shows a common op-amp, the '741', in an eight-pin integrated circuit package, and a diagram showing its pin connections. Pins 1 and 5 are connections to cancel or 'null out' this offset voltage by using an external voltage divider circuit. Pins 2 and 3 are the inverting and non-inverting inputs, and pin 6 is the output. Pins 7 and 4 are the supply voltage connections; pin 8 (NC = 'not connected') is unused.

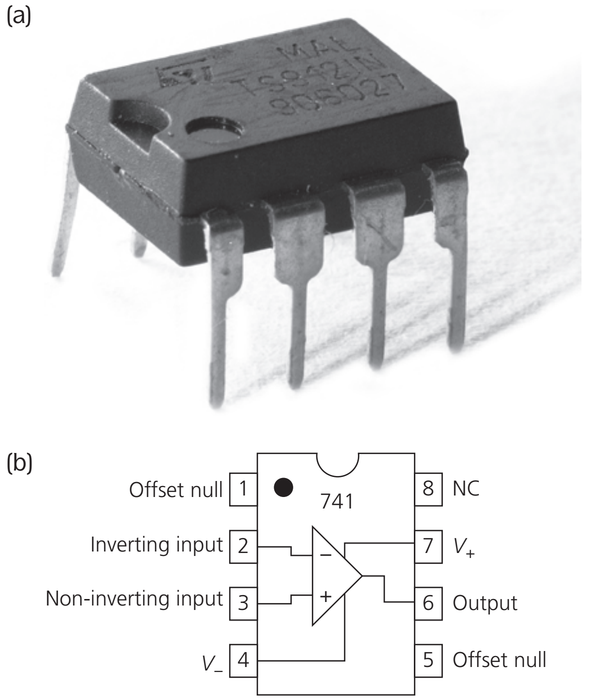


Figure 14 *(a) The 741 operational amplifier and (b) its pin connections*

Although the amplifier has a high input resistance, the input current is not exactly zero. Real op-amps will take a small current of the order of a few nanoamps (nA) and this will be amplified and will generate a small offset voltage. The output resistance is low, but it is not zero. Real op-amps are also affected by ambient temperature and their performance characteristics are usually quoted at a specific temperature value.

Gain and frequency response of an operational amplifier

A common application for op-amps is to amplify analogue signals. An analogue signal can vary in both amplitude and frequency, so an amplifier must be able to maintain its gain over the full range of frequencies in the signal, that is, across the whole signal bandwidth. An amplifier is thus specified not only by its voltage gain but also by the bandwidth over which it can amplify the range of frequencies uniformly.

Ideal op-amps have infinite open-loop gain at all frequencies. Real op-amps, however, have frequency response curves that look like that in Figure 15a, with a fall-off in gain as the frequency of an input signal increases. The voltage gain of a typical op-amp will start to fall linearly at a very low frequency when operated in an open-loop configuration. Thus, when the op-amp is on open loop, there is a high gain only for a small bandwidth. But if the op-amp is operated in a closed-loop configuration by applying feedback (see Electronics section 3.3), it is possible to exchange some of the very high open-loop gain for increase in signal bandwidth (Figure 15b).

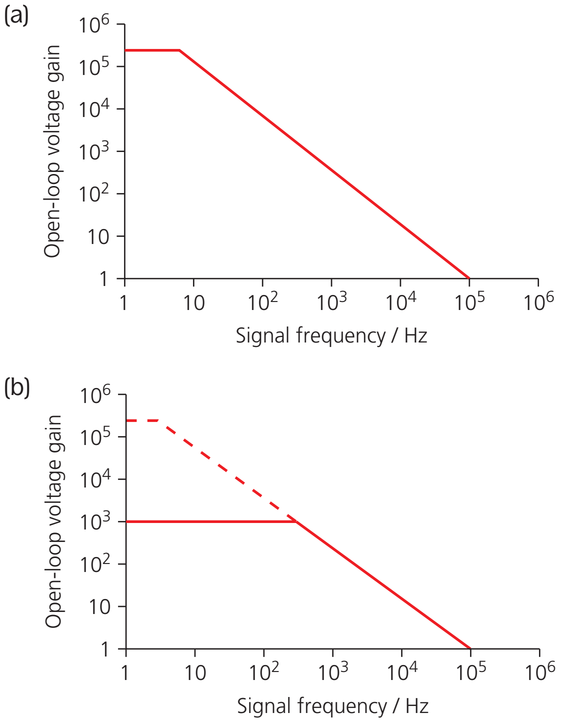


Figure 15 *(a) Open-loop frequency response for an operational amplifier. (b) In closed loop, the operational amplifier has a larger bandwidth for constant gain.*

We define a quantity called the **gain–bandwidth product** as

$$\text{gain} \times \text{bandwidth} = \text{constant}$$

where 'bandwidth' here means the range of frequencies for which the gain can be expected to be uniform. This is a constant quoted value for a real op-amp. Its unit is Hz (or, more realistically, kHz or MHz).

Worked example

The 741 amplifier has a gain–bandwidth product of 1 MHz. It is to be used to amplify analogue signals with a bandwidth of 10 kHz. What is the maximum gain that the amplifier will be able to provide for all frequencies at that bandwidth?

We have

$$\text{gain} \times \text{bandwidth} = 1\,\text{MHz} = 10^6\,\text{Hz}$$

Therefore, for a bandwidth of 10 kHz (10^4 Hz),

$$\text{gain} = 10^6/10^4 = 100$$

QUESTIONS

11. An op-amp has a gain–bandwidth product of 6 MHz. Copy and complete Table 2 for this op-amp.

Gain required	Useful bandwidth
1	
	20 kHz
	600 Hz
30	

Table 2

PRACTICE QUESTIONS

1. a. Figure Q1 shows an operational amplifier in the inverting configuration.

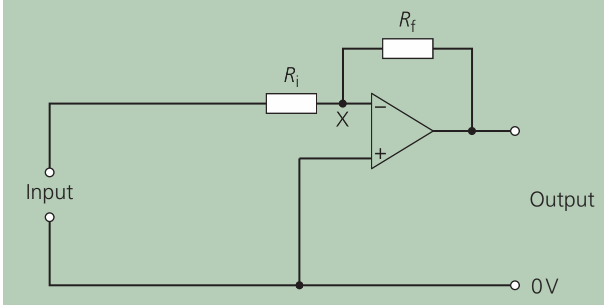


Figure Q1

i. The input resistor R_i is 4.7 kΩ and the feedback resistor R_f is 470 kΩ. What is the voltage gain of the circuit?

ii. Point X is a *virtual earth*. Explain what is meant by this term in the context of the circuit in Figure Q1.

b. Figure Q2 shows an op-amp connected as a temperature sensor.

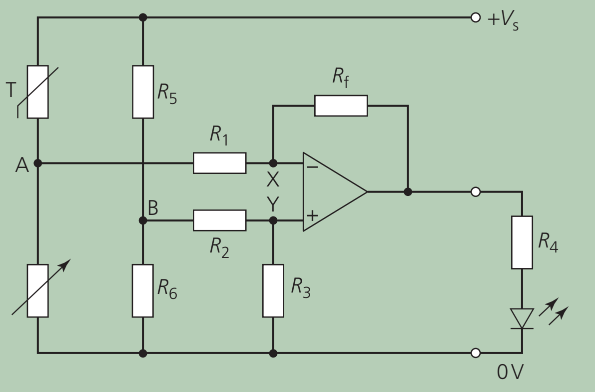


Figure Q2

i. In what configuration is the op-amp connected?

ii. Name the circuit component represented by T.

iii. Explain how the circuit works as a temperature sensor.

iv. What is the purpose of the resistor R_4 in series with the LED?

2. Stereo music recordings are made by having two separate microphones, one to the left and one to the right of the musicians. For these signals to be transmitted by radio, they have to be processed so that a listener with a mono radio receives all of the information, while a listener with a stereo receiver can receive both the left and right channel signals separately.

   a. For the mono radio listener, the left and right signals are added together and transmitted normally. Draw the circuit diagram for an op-amp circuit that can add together two audio signals.

   b. The magnitude of the voltage gain of the summing circuit is 1. On your diagram for part **a**, mark suitable resistor values.

   c. So that the two separate channels can be obtained for the stereo listener, the left and right signals are subtracted from each other, and this information is also transmitted, but in a way that cannot be detected by the mono listener. Draw the circuit diagram for an op-amp circuit that can subtract one signal from the other.

   d. The magnitude of the voltage gain of the subtraction circuit is 1. Mark on your diagram for part **c** suitable resistor values.

   e. The stereo radio receives two signals, L + R and L – R. Explain how the left and right signals can be extracted from these combined signals.

   AQA Electronics Unit 2 June 2011 Q4

3. A circuit is being designed to pre-amplify the audio signal from a microphone before being fed into an audio mixer for recording in a recording studio. The pre-amplifier (pre-amp) needs to amplify the signal by a factor of 50. The resistance of the microphone is 2.7 kΩ and the pre-amp input resistance must be 2.7 kΩ for optimum power transfer.

   a. It is decided to use an op-amp to construct the pre-amp. Draw two possible circuit diagrams for the pre-amp design using an op-amp

      i. that would amplify by a factor of 50 and give an inverting output

      ii. that would amplify by a factor of 50 and give a non-inverting output.

   b. The op-amp would be powered by a ±15 V supply. Assuming an ideal op-amp is used, what is the maximum amplitude of the input signal from the microphone before the output becomes saturated?

   c. Explain what is meant by the *gain–bandwidth product* for an op-amp.

   d. The pre-amp has a gain–bandwidth product of 1 MHz. Draw a graph using axes as in Figure Q3 to show how the open-loop voltage gain (which can be assumed to be infinite) of the op-amp varies with frequency.

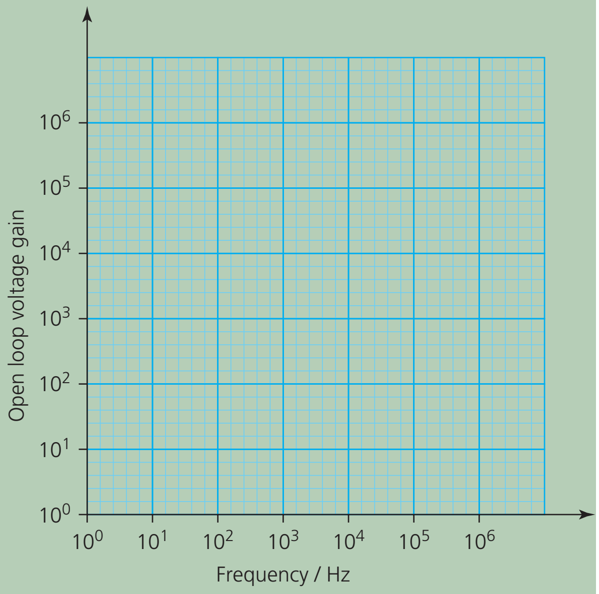


Figure Q3

   e. Above what frequency does the voltage gain become less than 50?

   f. The range of audio signal frequencies is 20 to 20 000 Hz. Comment on the suitability of the amplifier for maintaining a constant gain for audio signals.

4. A temperature sensor input subsystem is shown in Figure Q4.

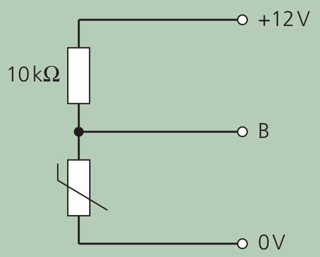


Figure Q4

a. The thermistor in Figure Q4 has a resistance of 45 kΩ at 0 °C, 20 kΩ at 25 °C and 1 kΩ at 100 °C. Calculate the output voltage at B at a temperature of 25 °C.

The temperature sensor input subsystem is connected to the comparator circuit as shown in Figure Q5.

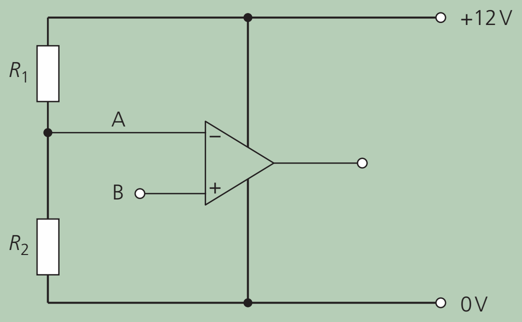


Figure Q5

b. Calculate and choose values of resistors, in the 1 kΩ to 10 kΩ range, for the circuit that will make the comparator switch at 25 °C.

c. What voltage would you expect from the output of this circuit when

i. the temperature is 20 °C

ii. the temperature rises to 30 °C?

AQA Electronics Unit 1 June 2010 Q6

4 DIGITAL SIGNAL PROCESSING

PRIOR KNOWLEDGE

You should know, from Electronics Chapter 2, about the nature of digital signals and how they may be represented by binary numbers, as patterns of 1s and 0s. You will be familiar with the concepts of frequency and period, and have a knowledge of the electrical properties of resistors and capacitors.

LEARNING OBJECTIVES

In this chapter you will learn that logic gates are fundamental building blocks in computers and other digital devices, and that they give an output depending on the state of the inputs. You will learn how logic gates can be combined to make logic circuits that perform designed logical functions in the processing of digital signals. You will find out how circuits are made that can count in binary and shift data through circuit components in logical blocks in a sequential manner. You will also learn about timing circuits and how they can be used to provide pulses to control digital circuits.

(Specification 3.13.5.1 to 3.13.5.3)

4.1 COMBINATIONAL LOGIC

Logic gates, Boolean algebra and truth tables

Most modern electronic devices used in communications, manufacturing, transport and scientific research contain **digital electronics** to carry out operations where decisions need to made from one or more inputs. Electronic switching circuits called **logic gates** give outputs that depend on the states of their inputs, or **logic states**, which are represented by pulses of electricity being present or absent. If a current is present, then this is represented by a 'high' state or '1'. If current is not present, then the state is represented by a 'low' state or '0'. The logic gate gives a logical output (0 or 1) depending on the present logic states (0 or 1) of its inputs.

A logic gate can be created in a number of ways. The most common method uses transistors as switches, connected in such a way that the required logical output is generated based on the logical inputs. There are different types of logic gate depending on the relationship required between input and output. We are not concerned here with what is inside a logic gate, but simply treat it as a building block that performs logical functions.

A very common example of an electronic system that makes use of logic gates is a mobile phone. When you speak into the phone, your voice is converted into digital signals and processed using logic gates as a series of 1s and 0s and then transmitted to the receiver phone. Digital circuits are used for mobile telephony because they are less prone to noise (see Electronics section 2.2) and easier to transmit over long distances. Computers also rely on logic gates for their operation. A typical microprocessor found in the heart of a personal computer may contain millions of logic gates formed from transistors, all working together to process digital information and allow the computer to perform a multitude of different tasks (Figure 1).

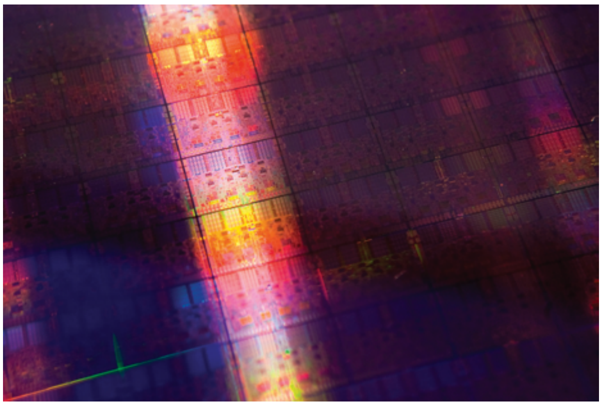
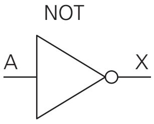


Figure 1 *The internal layout of the Core i7 microprocessor from Intel. This chip contains more than 1 400 000 000 transistors that form logic gates to process digital signals.*

In digital electronics, **combinational logic** means the use of connecting logic gates to give the required outputs from a particular combination of the input states. In order to determine the output states, **Boolean algebra** is used. This a mathematical method of manipulating logic states according to a set of rules in which a ‘1’ represents the concept of *true* and a ‘0’ represents the concept of *false*. Boolean algebra was invented by George Boole, an English mathematician and philosopher, in the 19th century. In 1854, he published a book called *An Investigation of the Laws of Thought*, which outlined the fundamental principles of the algebraic method. It was subsequently developed for use in digital electronics.

The simplest type of logic gate is a **NOT gate** (Figure 2). If the input state at its single input A is represented by A, then the output at X is the opposite state, denoted by $\bar{A}$ (meaning ‘NOT A’) in Boolean algebra. So, if the input at A is 0, then the output at X is 1; and if the input at A is 1, then the output at X is 0. A NOT gate is sometimes called an **inverter** because it inverts the input.



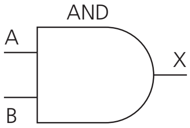
A	X
0	1
1	0

$X = \bar{A} = \text{NOT } A$

Figure 2 *A NOT gate symbol and its truth table*

A **truth table** is a list of all possible input states of a logic gate, with their corresponding logical outputs. This enables circuit designers to see at a glance how a gate works. We will show them for each type of logic gate, beside the circuit symbol for the gate.

An **AND gate** is shown in Figure 3. An AND gate gives an output of 1 at X if and only if the inputs at A *and* B are both 1; otherwise it gives a 0 output at X. In Boolean algebra, this is written in an equation as $X = A.B$, where the dot represents the logical AND operation.

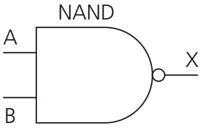


A	B	X
0	0	0
0	1	0
1	0	0
1	1	1

$X = A.B = A \text{ AND } B$

Figure 3 *An AND gate symbol and its truth table*

A **NAND gate** (Figure 4) is a logic gate that gives an inverted output from an AND gate, that is, $X = \bar{A}.\bar{B}$. It can be thought of as an AND gate with a NOT gate on its output.

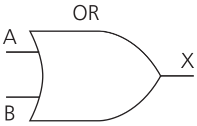


A	B	X
0	0	1
0	1	1
1	0	1
1	1	0

$X = \bar{A}.\bar{B} = \bar{A} \text{ AND } \bar{B}$

Figure 4 *A NAND gate symbol and its truth table*

An **OR gate** (Figure 5) gives a 1 output when one or the other of its inputs is 1 *or* both are 1. In Boolean algebra, this is written as $X = A + B$, where the + sign denotes the logical OR operation.

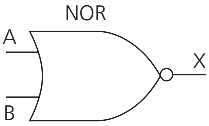


A	B	X
0	0	0
0	1	1
1	0	1
1	1	1

$X = A + B = A \text{ OR } B$

Figure 5 *An OR gate symbol and its truth table*

A **NOR gate** (Figure 6) is a logic gate that gives an inverted output from an OR gate. It can be thought of as an OR gate with a NOT gate on its output.

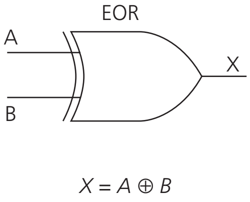


A	B	X
0	0	1
0	1	0
1	0	0
1	1	0

$X = \bar{A} + \bar{B} = \bar{A} \text{ OR } \bar{B}$

Figure 6 *A NOR gate symbol and its truth table*

One other type of gate used in digital electronics, called an **exclusive-OR gate** (**EOR**), is shown in Figure 7. The output is high if *either one* of the inputs is 1, but not if both are 1 or both are zero. The output is written as $X = A \oplus B$.



A	B	X
0	0	0
0	1	1
1	0	1
1	1	0

Figure 7 *An EOR gate symbol and its truth table*

Logic circuits

Logic gates can be combined together to form **logic circuits** that can also be described using Boolean algebra. Consider the Boolean expression $A + A.B$. This contains an AND function $A.B$ and an OR function $A + A.B$, so, to get the result expressed by this algebra, both an AND gate and an OR gate must be present in the circuit one after the other. The logic circuit is shown in Figure 8.

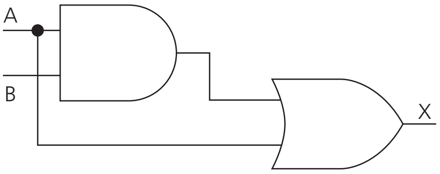


Figure 8 *Logic circuit for $X = A + A.B$*

QUESTIONS

1. For the logic circuit in Figure 8, state, in Boolean algebra,
   - **a.** the output of the AND gate
   - **b.** the two inputs to the OR gate.
   - **c.** The (incomplete) truth table for the logic circuit in Figure 8 is shown below. Copy and complete the table.

A	B	$A.B$	$X = A + A.B$
0	0	0	0
0	1	0	
1	0	0	
1	1	1	

2. **a.** Draw the arrangement of logic gates for the Boolean expression $A.(A + B)$.
   - **b.** Construct the truth table for your arrangement in part **a**.
   - **c.** What do you notice about this compared with your answer to question **1** part **c**? What can you conclude?

Given a truth table, it is possible to construct the corresponding digital logic circuit.

Worked example 1

The Boolean equation for a particular logic circuit with inputs A and B and output X is given by

$$X = (\bar{A}.B) + (A.\bar{B})$$

Draw up a truth table for the circuit and complete it for all possible inputs. Then draw the logic circuit, labelling the inputs and outputs at each stage.

Table 1 and Figure 9 show the solutions.

A	B	$\bar{A}$	$\bar{B}$	$\bar{A}.B$	$A.\bar{B}$	$X = (\bar{A}.B) + (A.\bar{B})$
0	0	1	1	0	0	0
0	1	1	0	1	0	1
1	0	0	1	0	1	1
1	1	0	0	0	0	0

Table 1

When drawing the logic circuit, it is necessary to identify what logical functions are involved and how many of each. In this case, for $(\bar{A}.B) + (A.\bar{B})$, NOT, AND and OR are involved. There are

- two NOT operations, (NOT A) and (NOT B)
- two AND operations, (NOT A AND B) and (A AND NOT B)
- one OR operation, (…) + (…).

So there must be two NOT gates, two AND gates and a single OR gate in the circuit (Figure 9).

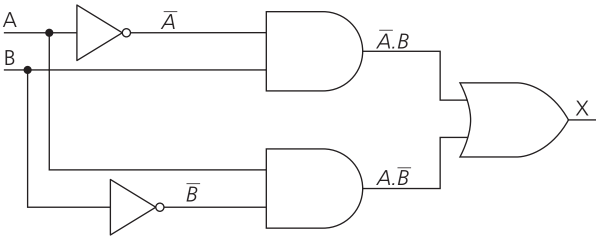


Figure 9

QUESTIONS

3. **a.** Draw a NAND gate with its inputs connected together.
   b. Draw the truth table for the NAND gate with its inputs connected together.
   c. Which logic gate is this equivalent to?

Stretch and challenge

4. **a.** Draw the logic circuit for the Boolean equation $X = (A + B).(\bar{A} + \bar{B})$.
   b. Compile its truth table.
   c. Which logic gate is this digital circuit equivalent to?

Questions 3 and 4 show that the same logical function can result from using different combinations of logic gates. A digital logic designer will try to use the most efficient set-up with the minimum number of gates available to reduce power consumption and cost and, when possible, use one gate to give the same output as another by connecting inputs together to reduce the number of individual inputs.

KEY IDEAS

- A logic gate is an electronic switch that performs a logical function. It gives an output of 0 or 1 depending on the logic states (0 or 1) of its inputs.
- NOT, AND, NAND, OR, NOR and EOR are different types of gate.
- A truth table lists all possible input states of a logic gate with their corresponding logical outputs.
- Boolean algebra is a mathematical way of describing a logic circuit using logical rules. It enables the truth table of a combination of logic gates to be determined.
- The following notation is used:

$\bar{A}$ = NOT A

$A.B = A$ AND B

$A + B = A$ OR B

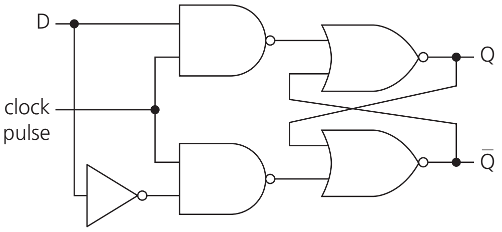
4.2 SEQUENTIAL LOGIC

Sequential logic circuits are logic circuits whose outputs depend, in part, on the sequence of *previous* input states. This is in contrast to combinational logic, whose outputs only depend on the state of the *present* input.

Sequential logic is important in **counting circuits**. A counting circuit is an electronic device that records the number of times an event has taken place in relation to a **clock pulse**, which is a regular sequence of high and low states. The clock pulse coordinates the action of the counting circuit. An example of the use of a counting circuit is to count events such as the number of cars entering a car park or the number of people walking through a turnstile. Electronic counters are used also in timing circuits – when it is necessary to count up or count down before an event is triggered. Counters can be manufactured as separate stand-alone integrated circuits or incorporated into much larger integrated circuits such as microprocessors.

The flip flop

A counting circuit is a logic circuit, based on a combination of logic gates called a **flip flop**, that is able to store logic state information. The flip flop has two logic states, high or low, and can 'settle' into one or other in a stable condition. Figure 10 shows a **D-type flip flop**, which is made out of two NOR gates, two NAND gates and a NOT gate, and its truth table.



D	Q	$\bar{Q}_{(next)}$
0	0	0
1	0	1
0	1	0
1	1	1

Figure 10 *A D-type flip flop and its truth table*

- The first column of the truth table in Figure 10 is the value of D, the *data input*: either 0 or 1. This could be the output of another digital circuit or a switch.
- The second column is the *output state*: the value output at Q.
- The third column, $Q_{(next)}$, is the *next output state* – that is, the value of Q at the next positive edge of the clock pulse.

The D-type flip flop has a data input D, a **clock pulse** or clock input, and two outputs Q and $\bar{Q}$. The flip flop is triggered to *flip* from one output state to the other by the *positive edge* of a series of clock pulses (see Figure 11) applied to the clock input. (Clock pulses and their generation are explained in detail in Electronics section 4.3.)

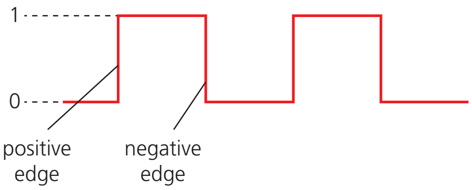


Figure 11 *A clock pulse has a positive (rising) edge and a negative (falling) edge.*

When the clock pulse is *not* at a positive edge, the flip flop ignores the data input D and remains in its current state. When the clock pulse is at its next positive edge, the flip flop is triggered to read the value of D and, based on this, it updates the value of Q. So we can say that:

> The logic state at the D input is transferred to the Q output only on the rising edge of the clock signal.

The flip flop output thus tracks the input, making transitions that match those of the input data, and meanwhile stores the value of the data input (Figure 12). In this sense, the D-type flip flop can be thought of as a basic memory unit.

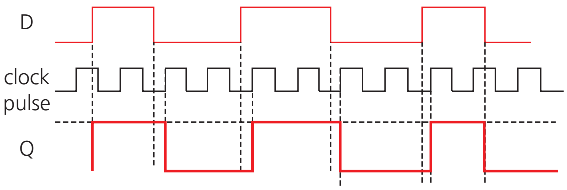


Figure 12 *The relationship between D, the clock pulse and Q for the D-type flip flop. Output Q tries to follow the input at D and then keeps this value until the clock input changes from low to high.*

D-type flip flops are used as system blocks to construct counting circuits. The circuit symbol for a D-type flip flop and its connections are shown in Figure 13.

The output of the D-type flip flop can be forced to a high or low state by means of the SET and CLEAR/RESET inputs.

- **SET** input: this is normally held low. When it is made high, the outputs of the D-type are forced immediately to the SET state, Q = 1, $\bar{Q}$ = 0.

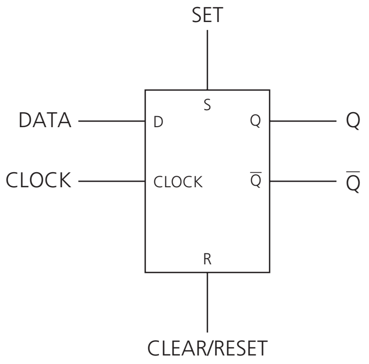


Figure 13 *Circuit symbol for a D-type flip flop*

- **CLEAR/RESET** input: this is normally held low. When it is made high, the outputs of the D-type are forced immediately to the CLEAR/RESET state, Q = 0, $\bar{Q}$ = 1.

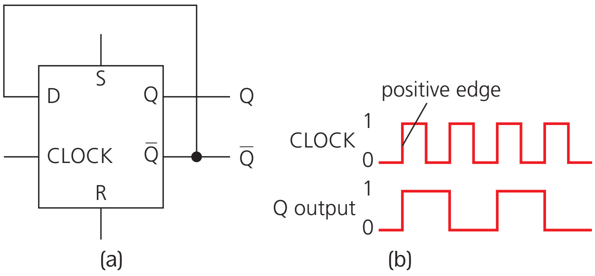
These inputs are used when it is necessary to return the outputs of the D-type to some initial logic state value.

Binary counter

D-type flip flops can be connected together to make a counter that counts in binary.

> For *n*-bit binary counting, *n* flip flops are needed.

First consider a single D-type flip flop connected as in Figure 14a, with the data input D connected to output $\bar{Q}$. At the start, both the CLOCK and Q are low. Since Q and $\bar{Q}$ are opposite states, $\bar{Q}$ must be high, and because D is connected to $\bar{Q}$, this must also be high. When the first positive edge of the clock pulse occurs, the input at D is read and transferred to Q, which goes high in turn. When Q goes high, $\bar{Q}$ goes low, and since it is connected to D, this goes low as well. When the second positive edge of the clock pulse occurs, D is now low, so Q becomes low, making $\bar{Q}$ high and D high again, ready for the next positive edge. The number of pulses at the output of the D-type is divided by 2 compared with the number of clock pulses. This configuration of a D-type flip flop is a one-bit **divide-by-two counter** (Figure 14b).

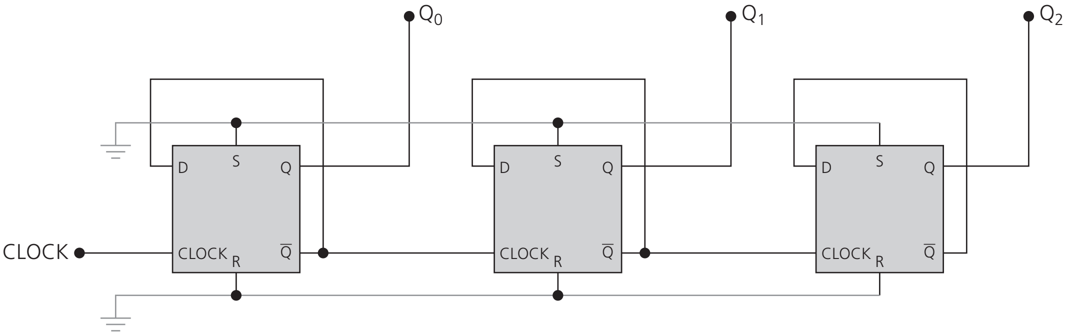


(a) (b)

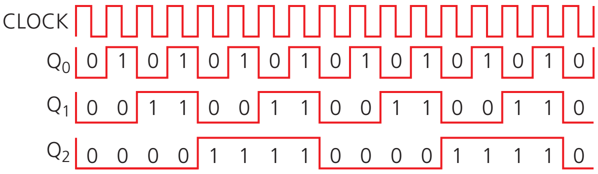
Figure 14 *In a single D-type flip flop with $\bar{Q}$ connected to D, the flip flop is a divide-by-two counter.*

Figure 15 shows a three-bit binary counter made with three D-type flip flops. Using three bits, it can count in binary from 0 to decimal number 7. The flip flops are connected with their D inputs connected to $\bar{Q}$ *and* to the clock input of the next flip flop (Figure 15a). The digital outputs Q_0, Q_1 and Q_2 are shown in Figure 15b. The binary numbers corresponding to these outputs are shown in the table in Figure 15c. The output from Q_0 corresponds to the least significant bit (LSB) and the output from Q_2 to the most significant bit (MSB). The rows of the table are a time sequence going forwards in time with each clock pulse as the flip flops count up to (decimal) 7.

Notice from Figure 15b that the binary counter is effectively a **frequency divider**. Output Q_1 has half the frequency of Q_0, and Q_2 has half the frequency of Q_1. This is a useful function in circuits where it is necessary to reduce the frequency of a signal. For example, inside a quartz clock or watch, the pd from the battery causes the quartz crystal to oscillate at a very precise and stable frequency, producing pulses at exactly 32768 Hz. These can be ‘divided down’ using a series of binary counters to lower frequencies to produce pulses for the hours, minutes and seconds clock display (see Assignment 1).



(a)



(b)

Q_2	Q_1	Q_0	
Binary			
MSB		LSB	
2^2	2^1	2^0	Decimal
0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	3
1	0	0	4
1	0	1	5
1	1	0	6
1	1	1	7

(c)

Figure 15 *A three-bit binary counter and its outputs*

While the Q outputs of the three flip flops generate an *up*-counting sequence, as shown in Figure 15, if you look at the $\bar{Q}$ outputs in Figure 16 you can see that they generate a *down*-counting sequence, so the binary counter can be used to count down as well as up. It is then sometimes called a **binary down counter**.

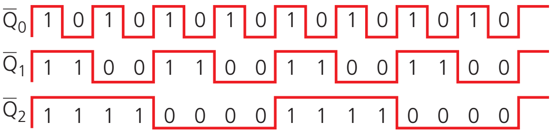


Figure 16 *The $\bar{Q}$ outputs of the three-bit binary counter generate a down-counting sequence: 111, 110, 101, 100, … in decimal is 7, 6, 5, 4, ….*

QUESTIONS

5. Figure 17 shows the signals applied to the D and Clock inputs of a D-type flip flop that is triggered on the rising edge. Copy these signals and below them, to the same time scale, complete the outputs for Q and $\bar{Q}$.

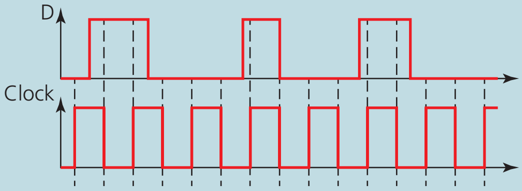


Figure 17

6. How could the counter in Figure 15 be modified to allow counting to more than 7?

Modulo-*n* counter

A **modulo-*n* counter** is a counter that counts up to a chosen number. It does this by detecting the unique logic state of the number using external logic gates and sending an output to reset the count.

Figure 18 shows a modulo-6 counter using three D-type flip flops. The binary for decimal 6 is 110. The output of Q_1 and Q_2 is fed into an AND gate, which, when both Q_1 and Q_2 outputs are 1, sends a 'high' to the Reset terminals that resets the counter to zero. This state with Q_1 and Q_2 both 1 will occur only once in the sequence.

BCD (binary coded decimal) counter

Four-bit binary numbers are common in digital electronics, as they can represent any decimal number from 0 to 9. A **binary coded decimal** (**BCD**) counter is a special case of the modulo-*n* counter that counts from 0 to 9 and then resets. It therefore goes through a sequence of ten states when it is clocked and returns to 0 on the binary count of 1001 (decimal 9). Four D-type flips flops are connected, as shown in Figure 19.

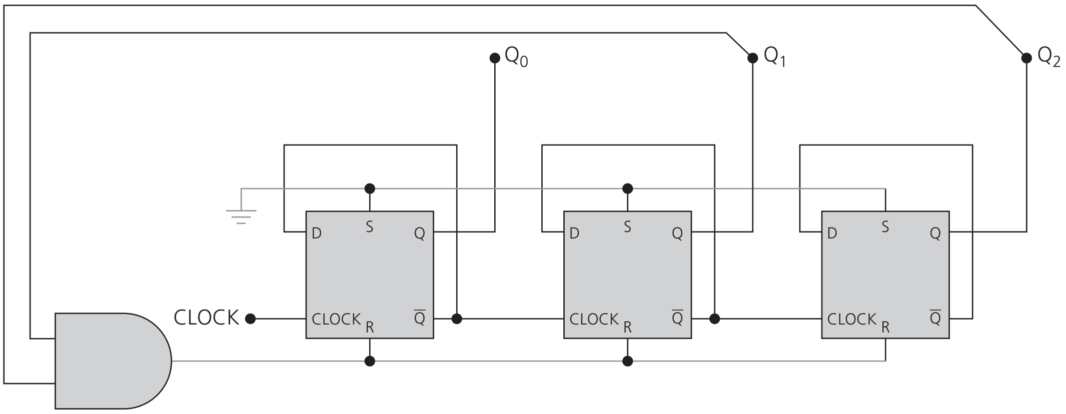


Figure 18 *A modulo-6 counter*

Table 2 shows the Q outputs of a four-bit binary counter. Its maximum count is at (decimal) 15, but at the count of 9 the states of the Q outputs are 1001. The outputs of Q_0 and Q_3 are fed to an AND gate, and when they are both 1 the AND gate resets the counter back to zero. This only occurs once in the sequence.

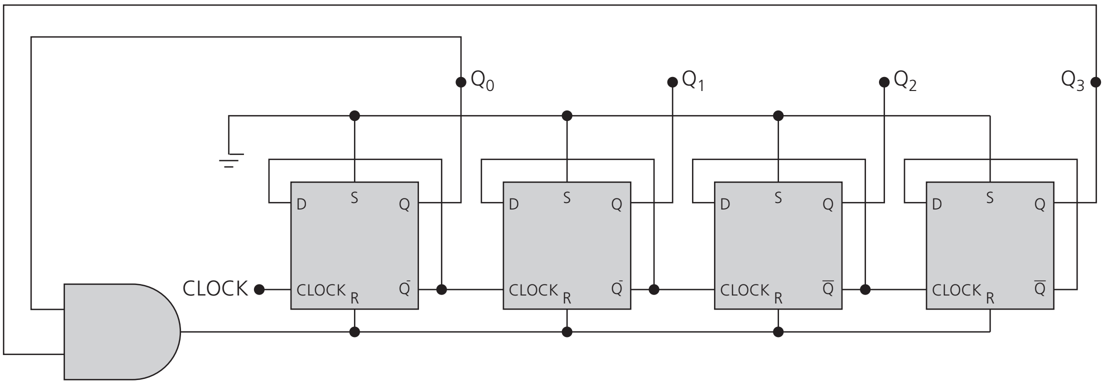


Figure 19 *A BCD counter*

Count	Q_3	Q_2	Q_1	Q_0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1

Table 2 *The count of 9 has state 1001*

Johnson counter

If you connect together D-type flip flops so that the Q output of one forms the data input of the next, and if you clock them all at the *same time*, then data is 'shifted' through the flip flops sequentially every clock cycle. This creates what is known as a **shift register** (Figure 20).

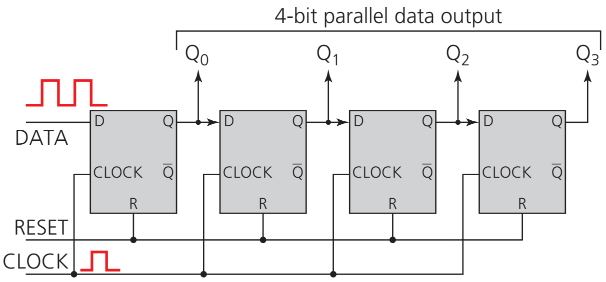


Figure 20 *A four-bit shift register*

Table 3 shows the movement of a single '1' bit fed to the data input of the first flip flop through a four-bit shift register. The bit is 'held' in the shift register and moved along to the next flip flop until it reaches the last one.

Clock pulse	Q_0	Q_1	Q_2	Q_3
0	0	0	0	0
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1

Table 3 *The movement of data through the shift register*

In a **Johnson counter**, the $\bar{Q}$ output of the last D-type flip flop is fed back to the data input D of the first (Figure 21). The data pattern contained within the register will recirculate as long as clock pulses are applied. Figure 22 show the outputs of a four-bit Johnson counter.

Suppose that the data inputs to all the D-type flip flops are initially zero. The first positive edge of the clock pulse shifts three 0s through the flip flops, that is, the 0s in Q_0, Q_1 and Q_2 go to Q_1, Q_2 and Q_3. The output of $\overline{Q_3}$, which is '1', is shifted back to Q_0 and then 1s shift through the flip flops, replacing the 0 states. Thus the counter circulates the pattern shown in the register repeatedly. An *n*-stage Johnson counter will give a count sequence of length 2*n* before the pattern is repeated.

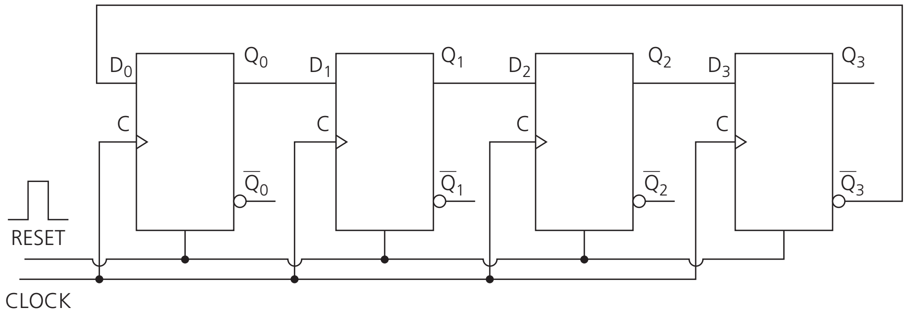


Figure 21 *A four-bit Johnson counter. Note that $\overline{Q}_3$ is fed back to D_0.*

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

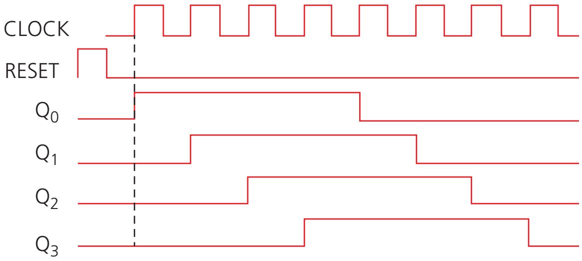


Figure 22 *Johnson counter outputs after eight clock pulses*

QUESTIONS

7. **a.** Compile a table of all 10 states of a BCD counter.

b. Explain, referring to Figure 19 and your table in part **a**, how outputs Q_0 and Q_3 can be used to reset the counter to zero with the AND gate.

8. **a.** A D-type flip flop is shown in Figure 23. Describe its operation, with reference to its connections.

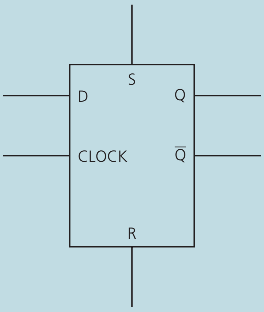


Figure 23

b. Draw the connections between four such D-type flip flops to form a four-bit shift register. Add external logic gates to the register so that, when all the Q outputs are 0, the logic state at the data input is 1.

9. A Johnson counter is constructed with five D-type flip flops. What is the length of its count sequence?

KEY IDEAS

- Sequential logic circuits have outputs that depend in part on the sequence of previous states of the inputs.
- The D-type flip flop is a basic building block of sequential logic circuits. It transfers the value of its data input to its output on the positive edge of a clock pulse. Otherwise its output does not change.
- D-type flip flops can be connected together to form counting circuits that count in binary. The counter can either count up using the Q outputs or down using the $\overline{Q}$ outputs.
- A Johnson counter is a counter made of *n* D-type flip flops, where the $\overline{Q}$ output of the last flip flop is fed back to the input of first flip flop. This causes a pattern of bits to be shifted along with each clock pulse and to be repeated after a certain count. An *n*-stage Johnson counter will give a count sequence of length 2*n* before the pattern is repeated.
- A modulo-*n* counter is a counter that counts up to a chosen number. It does this by detecting the unique logic state of the number using external logic gates and sending an output to reset the count.
- A BCD counter is a special case of a modulo-*n* counter with $n = 10$. It counts in binary from 0 to decimal 9 and then resets.

ASSIGNMENT 1: KEEPING TIME


(a) (b)

Figure A1 *A quartz watch uses a quartz crystal as a super-accurate time-keeping element.*

The need to keep time has been important throughout history. Some of the oldest time-keeping devices were developed by Neolithic people as long ago as 3100 BC. The changing position of the sunrise throughout the year was determined at the time when the Sun rose and was aligned with a certain stone. Stonehenge in Wiltshire can be thought of as a ‘clock’ that records the time of the summer and winter solstices and provides a reference when to plant and harvest crops. Other types of simple clock were developed, including sundials and hourglasses. During the centuries of major seafaring, the need arose to make more accurate clocks for navigation purposes. These were based on moving mechanical assemblies that produced regular oscillations. With the advent of electronics, we can now construct clocks that are extremely accurate and contain no moving parts at all.

In this assignment we will see how modern digital clocks and watches use a quartz crystal as a time-keeping element that can be used to generate pulses to drive digital logic circuits to produce a highly accurate timepiece (Figure A1).

A quartz crystal is a ‘piezoelectric‘ material, which means that it will mechanically oscillate when there is a varying potential difference across it and will resonate when the frequency of the pd is near to its natural frequency. It will then *generate* an emf at this exact natural frequency. This can be put through an external circuit that produces amplified stable square wave pulses of this accurate frequency. A common natural frequency for quartz crystal oscillators is 32768 Hz.

Figure A2 shows the original circuit diagram (redrawn) from the 1975 patent for the first digital watch. It used an oscillator amplifier based on a quartz crystal and sequential logic circuits acting as frequency dividers.

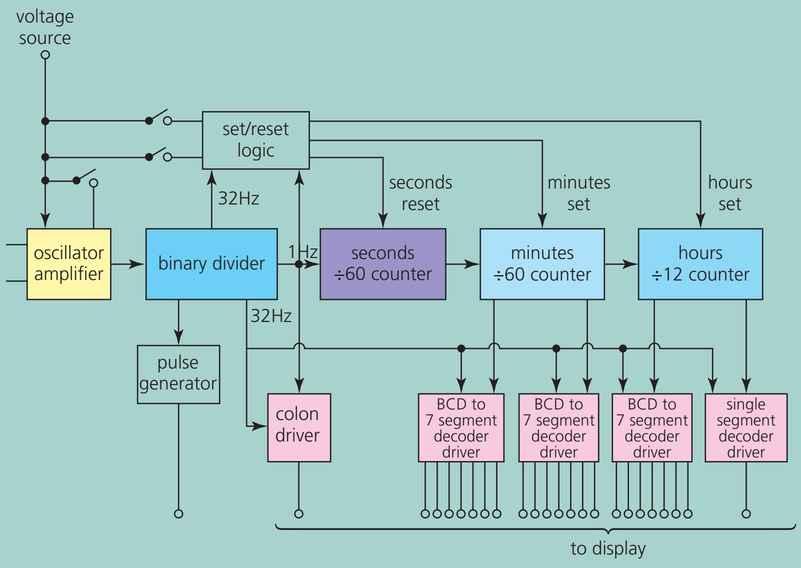


Figure A2 *Circuit for the first digital watch (redrawn from the US Patent application)*

Questions

A1 Explain what is meant by *resonate*. (You may want to look back at section 2.4 of Chapter 2.)

A2 Why is a quartz oscillator of frequency 32768 Hz used?

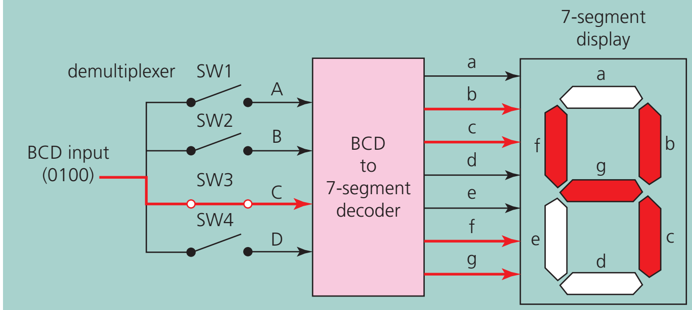
A3 What is the purpose of the binary divider in Figure A2?

A4 For an oscillator frequency of 32768 Hz, how many divide-by-two flip flops are needed in the binary divider to produce an output of 1 Hz?

A5 Digital watches use modulo-*n* counters to produce the hours and minutes display. What is a modulo-*n* counter?

Stretch and challenge

A digital watch face has a seven-segment LED display that allows the numbers 0 to 9 to be shown. The outputs from the counters are fed to a 'BCD to seven-segment decoder'. This looks at the binary number coming in and turns on the appropriate bars in the seven-segment LED to display that number. Figure A3 shows a block diagram of the decoder and the segment combinations that have to be active to display a particular number.



Digit shown	Illuminated segment (1 = illumination)						
	a	**b**	**c**	**d**	**e**	**f**	**g**
0	1	1	1	1	1	1	0
1	0	1	0	0	0	0	0
2	1	1	0	1	1	0	1
3	1	1	1	1	0	0	1
4	0	1	1	0	0	1	1
5	1	0	1	1	0	1	1
6	1	0	1	1	1	1	1
7	1	1	1	0	0	0	0
8	1	1	1	1	1	1	1
9	1	1	1	1	0	1	1

Figure A3 *A seven-segment decoder*

Questions

A6 The binary input to the decoder is in BCD. What is meant by BCD?

A7 What BCD value will display the number 7?

A8 A seven-segment decoder allows a four-bit binary number (half a byte) to be used to display the numbers from 0 to 9. But the numbers on a watch need to exceed the value 9. How could numbers from 00 to 99 be displayed using seven-segment decoders?

4.3 ASTABLES

Clock pulses

An **astable** is an oscillator circuit that produces a continuous output of regular ON–OFF pulses that can be used as clock pulses. Sequential logic circuits (see Electronics section 4.2) use clock pulses for synchronisation – they depend on the frequency and width of an input clock pulse to activate their switching action. Clock pulses are continuous and are rectangular in shape, going from a low (OFF), normally 0 V, to a high (ON), usually 5 V, with constant period, as shown in Figure 24.

Clock pulses have a rising or positive edge and a falling or negative edge. Sequential circuits can be triggered on either the rising or falling edge of a clock pulse, or both. The **pulse width** is the duration of the pulse when it is high or ON. A **clock cycle** is the **period**, that is, the time between the start of one pulse and the start of the next.

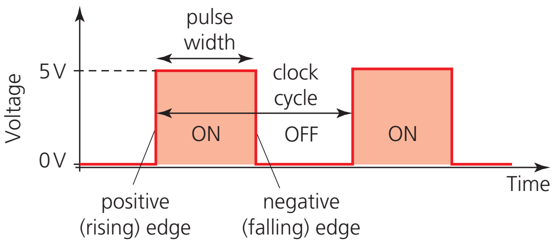


Figure 24 *Clock pulses*

The **clock rate** or **pulse frequency** in Hz is the reciprocal of the period:

clock rate or pulse frequency

$$f = \frac{1}{\text{period}} = \frac{1}{\text{ON time} + \text{OFF time}}$$

The time for which the pulse is high or ON – the pulse width – is sometimes called the 'mark', and the time for which the pulse is low or OFF is then called the 'space'. In Figure 24 the pulse is ON for the same time as it is OFF, but in general the mark and the space do not need to be of equal duration (see Figure 25). We define the **mark-to-space ratio** as

$$\text{mark-to-space ratio} = \frac{\text{ON time}}{\text{OFF time}}$$

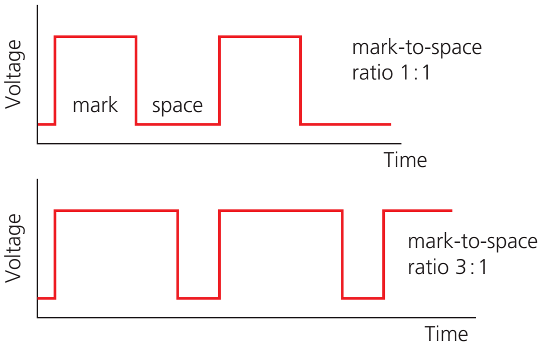


Figure 25 *Clock pulses with different mark-to-space ratios*

The **duty cycle** is the percentage of the time that the pulse is in the ON state:

$$\text{duty cycle} = \frac{\text{ON time}}{\text{period}} \times 100\% = \frac{\text{mark}}{\text{period}} \times 100\%$$

So, for the clock pulses in Figure 25, when the mark-to-space ratio is 1 : 1, the duty cycle is 50%, and when the mark-to-space ratio is 3 : 1, the duty cycle is $\frac{3}{4} \times 100\% = 75\%$.

QUESTIONS

10. a. Draw a clock pulse with a duty cycle of 25%.

b. What is the mark-to-space ratio for the pulse in part **a**?

c. If the ON pulse lasts 0.02 s, what is the clock rate for this pulse?

An astable in an *RC* network

There are many kind of circuits and devices that can produce clock pulses. One of the most common is the 555 astable integrated circuit. This can be connected with external resistors and a capacitor in an ***RC* network** (resistor–capacitor network) to produce clock pulses (Figure 26). The 0.01 µF capacitor is not involved in the timing function and is there to prevent electrical noise on the power supply from affecting the timer operation. The frequency of the clock pulses and the duty cycle are set by the values of R_1, R_2 and C_1 in the *RC* network. (See the Worked example 2.)

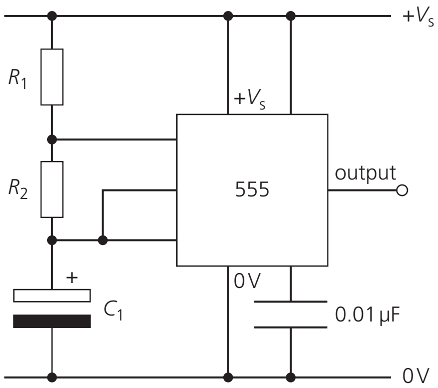


Figure 26 *The 555 astable clock pulse generator*

Worked example 2

A 555 astable is used to produce clock pulses with an external capacitor and two resistors, in a circuit like that in Figure 26. The pulse frequency and the duty cycle are as given in the previous expressions. The value of C_1 is 10 µF, R_1 is 100 kΩ and R_2 is 47 kΩ.

The frequency of the clock pulses and the duty cycle are given as follows.

$$\text{pulse frequency} = f = \frac{1.4}{(R_1 + 2R_2) \times C_1}$$

$$\text{duty cycle} = \frac{R_1 + R_2}{R_1 + 2R_2} \times 100\%$$

a. Calculate the frequency of the clock pulses produced by the astable.

b. Calculate the duty cycle of the clock pulses.

c. Determine the mark-to-space ratio.

d. How could the *RC* network be modified to produce an astable with a variable clock frequency?

a. Using the expression given:

$$\text{pulse frequency} = \frac{1.4}{(10^5 + 2 \times 47 \times 10^3) \times 10 \times 10^{-6}}$$
$$= \frac{1.4}{19.4} = 0.72\,\text{Hz}$$

b. Using the expression given:

$$\text{duty cycle} = \frac{10^5 + 47 \times 10^3}{10^5 + 2 \times 47 \times 10^3} \times 100\%$$
$$= \frac{1.47 \times 10^5}{1.94 \times 10^5} = 76\%$$

c. From the definitions in the main text, and the given expressions for pulse frequency and duty cycle, it can be shown that

$$\frac{1}{\text{period}} = \frac{1.4}{(R_1 + 2R_2) \times C_1}$$

$$\frac{\text{mark}}{\text{period}} = \frac{R_1 + R_2}{R_1 + 2R_2}$$

Then with some algebraic substitutions:

$$\text{mark} = \frac{R_1 + R_2}{R_1 + 2R_2} \times \text{period}$$
$$= \frac{R_1 + R_2}{R_1 + 2R_2} \times \frac{(R_1 + 2R_2)C_1}{1.4} = \frac{(R_1 + R_2)C_1}{1.4}$$

and

$$\text{space} = \text{period} - \text{mark}$$
$$= \frac{(R_1 + 2R_2)C_1}{1.4} - \frac{(R_1 + R_2)C_1}{1.4} = \frac{R_2C_1}{1.4}$$

Therefore

$$\frac{\text{mark}}{\text{space}} = \frac{R_1 + R_2}{R_2} = \frac{(100 + 47) \times 10^3}{47 \times 10^3} = \frac{147}{47} = 3.1$$

and so the mark-to-space ratio is 3.1:1.

d. One of the resistors in the *RC* network could be replaced with a variable resistor.

QUESTIONS

Use the equations given in Worked example 2 for these questions.

11.a. What are the frequency and period of clock pulses from a 555 astable in an *RC* network with $R_1 = 10\,\text{k}\Omega$, $R_2 = 20\,\text{k}\Omega$ and $C_1 = 220\,\text{nF}$?

b. What is the duty cycle of the clock pulses? Explain the meaning of this.

c. How long does each ON pulse last?

12. A 555 astable generates pulses that have an ON time of 5.0 μs and an OFF time of 2.5 μs.

a. What is the pulse frequency of the pulses?

b. If $R_1 = 5.6\,\text{k}\Omega$ and R_2 is $56\,\text{k}\Omega$, what is the value of C_1?

c. What is the duty cycle?

d. What is the mark-to-space ratio?

NAND gate astable

Another type of astable circuit that produces a clock signal has two NAND gates connected with an *RC* network, as shown in Figure 27.

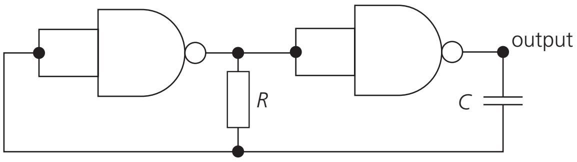


Figure 27 *NAND gate astable*

The frequency of the pulse from such a NAND gate astable is given by

$$f = \frac{1}{2.2RC}$$

The circuit gives an output clock pulse with a mark-to-space ratio of approximately 1:1.

QUESTIONS

13. What is the pulse frequency of a NAND gate astable with an external *RC* network of $R = 15\,\text{k}\Omega$ and $C = 0.1\,\mu\text{F}$?

14. How could a NAND gate astable provide a variable clock frequency?

15. What is the disadvantage of this type of astable, compared with the 555 astable circuit?

KEY IDEAS

- Astables are oscillator circuits, producing clock pulses for counter circuits.
- A clock pulse is a continuous ON–OFF signal with a constant period.
- The clock rate or pulse frequency is the reciprocal of the period.
- The pulse width or 'mark' is the time for which a pulse is ON.
- The duty cycle is the percentage of the time that the pulse is ON:

$$\text{duty cycle} = \frac{\text{ON time}}{\text{period}} \times 100\%$$

- The mark-to-space ratio is:

$$\text{mark-to-space ratio} = \frac{\text{ON time}}{\text{OFF time}}$$

- The pulse frequency of an astable is determined by an external *RC* network.

PRACTICE QUESTIONS

1. The Boolean equation for a particular logic circuit with inputs A and B and output Q is

$$Q = (A.B) + (\overline{A}.\overline{B})$$

Table Q1 shows intermediate logic signals for the circuit with inputs A and B and output Q for all combinations of the inputs A and B.

A	B	$\overline{A}$	$\overline{B}$	A.B	$\overline{A}.\overline{B}$	Q
0	0	1	1	0	1	
0	1	1	0	0	0	0
1	0	0	1	0		0
1	1	0	0	1	0	1

Table Q1

a. Copy the truth table and complete the missing two entries.

b. Copy and complete Figure Q1 to show the logic circuit that has the same function as the Boolean equation given above. Your circuit should only contain *two* AND gates, *two* NOT gates and *one* OR gate.

Figure Q1

AQA Paper 3BE Specimen 2014 Q6

2. As part of his project, a student constructs the logic circuit in Figure Q2.

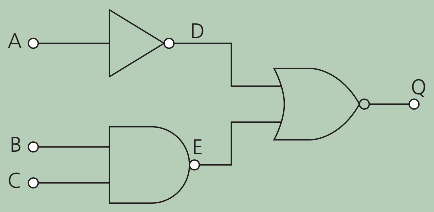


Figure Q2

a. Write down the Boolean expressions for:

i. D

ii. E

b. Write down the Boolean expression for Q in terms of D and E.

c. Copy and complete the truth table in Table Q2 for the logic circuit in Figure Q2.

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

Table Q2

AQA Electronics Unit 1 June 2013 Q2 (part)

3. **a.** Figure Q3 shows a combination of logic gates.

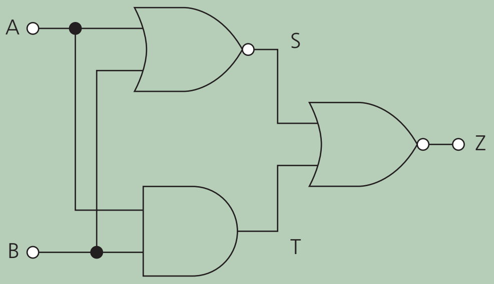


Figure Q3

Copy and complete the truth table (Table Q3) that shows the output at S, T and Z.

A	B	S	T	Z
0	0			
1	0			
0	1			
1	1			

Table Q3

b. Figure Q4 shows an EOR gate.

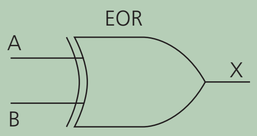


Figure Q4

i. Write down the truth table for the output of an EOR gate.

ii. What other gate would you need to put in front of the EOR gate to get the same output as the logic circuit in part **a**?

4. Copy and complete the circuit diagram in Figure Q5 for a four-bit shift register. Label the data input and clock input to the shift register.

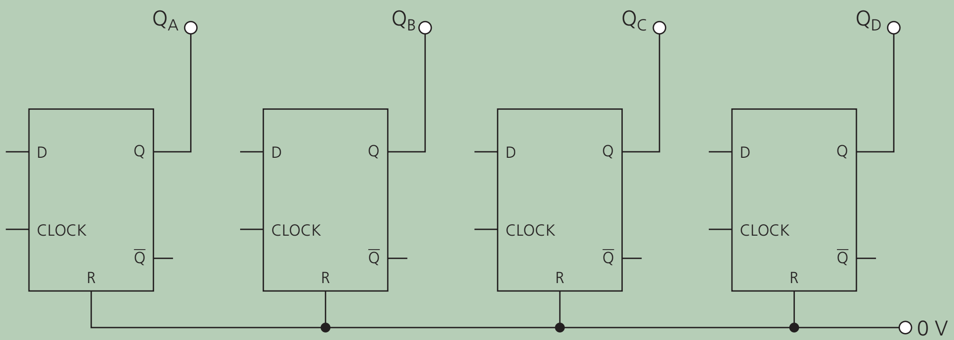


Figure Q5

AQA Electronics Unit 2 June 2012 Q1 (part)

5. **a. i.** Draw a circuit diagram to show how a D-type flip flop (Figure Q6) can be connected to divide a clock frequency by 2, and label the input and output.

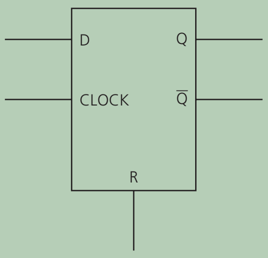


Figure Q6

ii. Copy and complete the timing diagram in Figure Q7 to show the relationship between the input signal and Q.

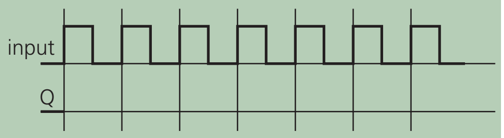


Figure Q7

b. Bats navigate using ultrasonic waves between 30 kHz and 60 kHz. A student designs the system in Figure Q8 to investigate these sounds.

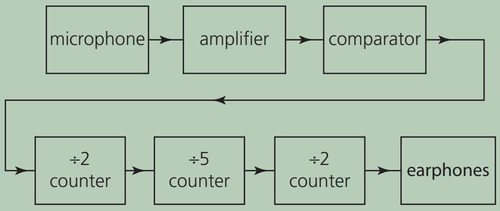


Figure Q8

i. If the system receives ultrasonic waves at a frequency of 48 kHz, calculate the frequency of the signal at the earphones.

ii. Copy and complete the circuit diagram in Figure Q9 for a divide-by-5 counter.

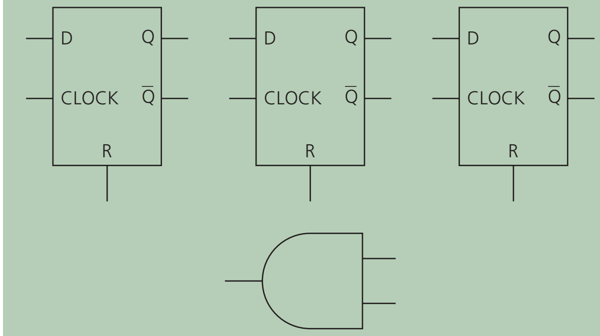


Figure Q9

AQA Electronics Unit 2 June 2013 Q2

6. The circuit in Figure Q10 shows a 555 timer configured as an astable, with $R_1 = 1\,\text{k}\Omega$, $R_2 = 2\,\text{k}\Omega$ and $C = 10\,\mu\text{F}$.

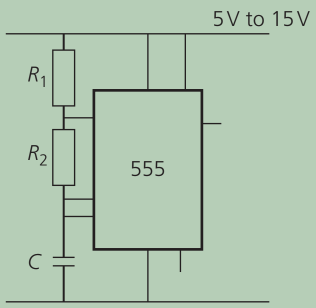


Figure Q10

a. i. Describe the output signal from this astable circuit.

ii. What might this output be used for?

b. The pulse frequency and the duty cycle for the circuit are given as follows:

$$\text{pulse frequency} = \frac{1.4}{(R_1 + 2R_2) \times C_1}$$

$$\text{duty cycle} = \frac{R_1 + R_2}{R_1 + 2R_2} \times 100\%$$

Calculate

i. the duty cycle

ii. the frequency

iii. the period

iv. the mark-to-space ratio.

5 DATA COMMUNICATION SYSTEMS

PRIOR KNOWLEDGE

You should have a knowledge of wave characteristics, in particular, electromagnetic waves and their properties of reflection, refraction, total internal reflection, diffraction and superposition (*Chapters 5, 6 and 7 of Year 1 Student Book*). You will be familiar with the theory and application of optical fibres (*Chapter 7 of Year 1 Student Book*). You should also have a good understanding of the difference between analogue and digital signals (Electronics Chapter 2).

LEARNING OBJECTIVES

In this chapter you will learn how data can be sent from one place to another via different transmission media. You will learn what constitutes the main components of a communications system and how they work together. You will see how information can be impressed on – or 'carried' using – a wave, so that it can be transmitted over long distances, and how this can be done efficiently. You will also learn what determines the maximum amount of information that can be transmitted.

(Specification 3.13.6.1 to 3.13.6.4)

5.1 PRINCIPLES OF COMMUNICATION SYSTEMS

A **communication system** is way of communicating information from a source to a receiver. The transfer of information can occur between two humans, between a human and a machine, or between two machines, and it can be in analogue or digital form. The information is propagated or 'carried' along a communication channel or transmission path from source to receiver by a **carrier**, which can be electromagnetic waves, sound, liquid or some other transmission medium that is able to carry energy.

The communication channel in a communication system may be:

- **simplex**, where the information flows in one direction only
- **duplex**, where communications between participants can flow between source and receiver and back again simultaneously
- **half-duplex**, where two-way communication is possible, but only one participant can communicate at a time
- **multiplex**, where multiple analogue digital data streams are combined into one signal over a common transmission path.

A communication system can be considered as a number of distinct system blocks through which the information passes on its way to the receiver (Figure 1). We will look at each component in turn.

Information input

Information that needs to be sent elsewhere comes in many forms. It might be speech, images, digital information, or analogue information from a sensor such as temperature or positional data, for example.

Input transducer

Whatever the type of information, it needs to be converted into a form (usually electrical) that is suitable for further transmission through the system. A **transducer** is a device, such as a microphone, that converts one form of energy into another – see Electronics section 2.3.

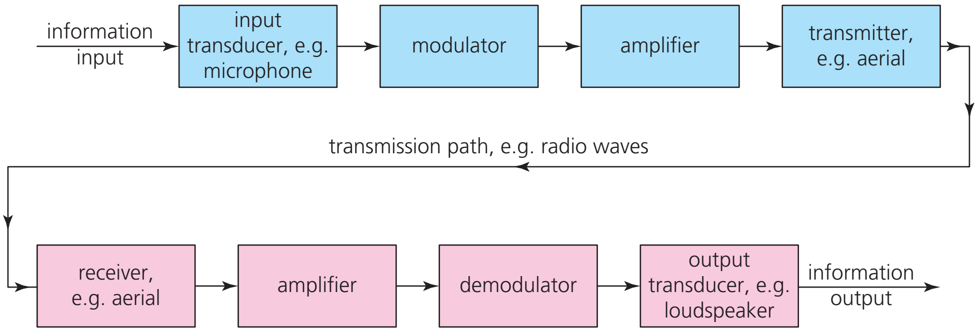


Figure 1 *The basic components of a communication system*

Modulator

Modulation is the addition of the information that needs to be transmitted onto a signal **carrier**. A **modulator** is a device that impresses the information onto the carrier and blends it in a way that is suitable for transmission. One can think of waving a blanket over a smoky fire to send smoke signals. The blanket can be said to modulate the steady stream of smoke (the carrier) with a signal. Another example is Morse code, invented for wireless telegraphy, which uses a two-state digital code to modulate a light or radio signal with information (Figure 2).

Figure 2 *A signalman modulating a light signal with Morse code*

Amplifier

The purpose of the amplifier is to increase the voltage, current or power of the modulated signal so that it is well above any noise (see Electronics section 2.2) and is boosted to a level that will increase its range when transmitted (since all signals are weakened, or suffer **attenuation**, with distance). Amplifiers can generate a lot of power. The output power of a TV transmitter is typically a few hundred kilowatts and can broadcast over ranges of hundreds of miles.

Transmitter

A transmitter is a transducer that can project the signal for transmission. In the case of radio transmission, it is an **aerial** or **antenna** that converts an electrical signal into radio waves. Another example is the conversion of electrical signals to light signals by laser diodes in optical communications, and also by infrared light-emitting diodes (LEDs) in TV remote controls (Figure 3).

Figure 3 *A TV remote control contains an infrared LED that transmits a modulated infrared signal to communicate with the TV.*

Transmission path

This is the physical medium in the data communication system over which a signal propagates. It may be copper cable, optical fibre, or an electromagnetic wave in air. The nature of the transmission path can affect the quality of the transmission, as we will see in Electronics section 5.2.

Receiver

The purpose of the receiver is to collect the signal energy that has been sent. Often the signal strength is very weak at the receiving end, so the receiver has to be large to collect enough energy. Figure 4 shows the receiver dish aerial for the NASA Deep Space

Tracking Network at Goldstone in California. The dish is 70 m across and focuses weak radio signals with its parabolic reflector to a detector at the focus.

Figure 4 *The Goldstone 70 m receiver dish receives weak signals from interplanetary probes far out in space that are exploring the solar system.*

Amplifier

As on the transmission side of a communications system, an amplifier is necessary on the receiver side so that weak signals can be identified and read easily above any background noise.

Demodulator

The purpose of the **demodulator** is to extract the original information signal from the modulated carrier signal that was used to send it.

Output transducer

Once the information signal has been demodulated and recovered from the carrier, it is passed into an output transducer, which converts the electrical signal to the required output for reading. Examples of output transducers are loudspeakers (electrical energy to sound), printers (electrical energy to mechanical energy) and projectors (electrical energy to light).

QUESTIONS

1. State whether the following communication channels are duplex, half-duplex or simplex:
    a. listening to your iPod
    b. using Skype
    c. using a walkie-talkie
    d. internet surfing
    e. talking to someone on your mobile phone
    f. using an automatic garage door opener.

KEY IDEAS

- A communication system is made up of a number of system blocks, each of which has a specific purpose in transferring an information signal from a source to a receiver.
- Modulation is the process of impressing information onto a signal carrier.
- The transmission path is the physical medium over which the information is transmitted.
- Demodulation is the process of extracting the original information content from the signal carrier.
- Communication channels can be simplex (one-way), duplex (simultaneously two-way), half-duplex (two-way, but one at a time), or multiplex.

5.2 TRANSMISSION-PATH MEDIA

In electronic communication systems, the most common transmission-path media are metal wires, optical fibres and electromagnetic waves in the radio and microwave range.

Metal wires

Metal wire, such as that in copper cable, uses the varying flow of electric current to carry information. It provides a **guided transmission** path to the destination. The cable is either **coaxial cable** or **twisted-pair cable**.

A coaxial cable (Figure 5) consists of a central core of copper wire (a single copper strand or several braided strands), which transports the data, surrounded

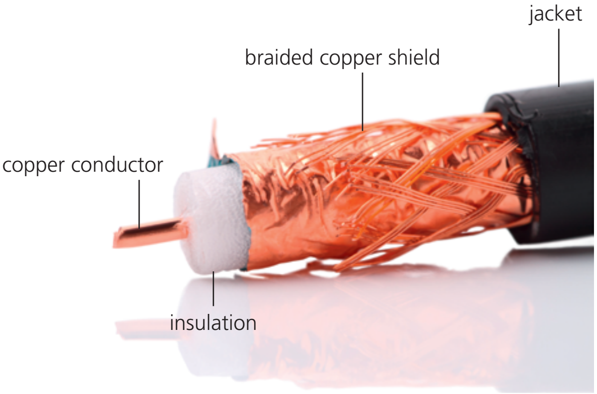


Figure 5 *The structure of coaxial cable*

by an insulator to isolate it from the braided metal **shield** that protects the transmitted data from electrical interference. The cable is covered by a tough protective jacket made out of PVC or Teflon to protect it from the environment.

In a twisted-pair cable (Figure 6), pairs of insulated copper strands are woven together and a bundle of such pairs is covered with outer plastic insulation. The twisting of the individual pairs helps eliminate electrical interference from adjacent pairs of wires or external sources of electrical noise. Twisted pairs are suitable for local-area computer networks (LANs), where high-density connections are made. They are colour-coded so that the connections can be followed through.

Figure 6 *A twisted-pair LAN cable. The pairs are colour-coded for identification.*

Both coaxial cable and twisted-pair cable suffer from loss of signal energy or **attenuation**. These can be due to resistive losses in the wires or losses when they are joined together. Being made of metal, they are also subject to corrosion and oxidisation, which can degrade their performance.

Optical fibres

Optical fibres are another form of guided path transmission medium. They use the total internal reflection of light to guide light pulses down a glass fibre (*see sections 7.4 and 7.5 in Chapter 7 of Year 1 Student Book*).

Optical fibres are long, thin and highly transparent strands of glass used to transmit pulses of light (usually infrared) from a semiconductor laser diode or an LED over long distances with very little energy loss. A typical optical fibre cable (Figure 7) consists of a very pure glass core with a high refractive index and a diameter of a few micrometres, surrounded by glass cladding of lower refractive index. This is surrounded by a strengthener and a protective jacket to give the optical fibre cable durability. Being glass, they do not corrode, so overall they are robust in harsh environments.

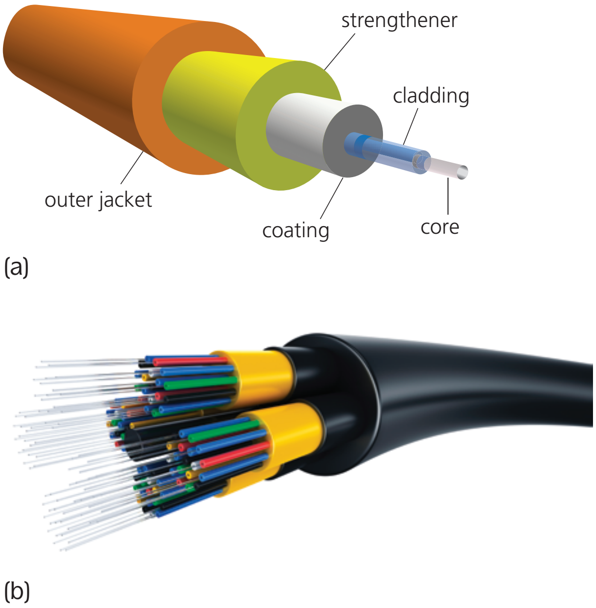


Figure 7 *(a) The structure of an optical fibre cable, simplified to show just one fibre. (b) Optical fibre communication cable contains many individual fibres bundled together.*

Optical cable can transmit much more information at one time than coaxial cable or radio waves. The digitally coded pulses of light are produced at very high frequencies and pass along the optical fibres, being totally internally reflected from the cladding several thousand times per metre. While attenuation is not as great as in metal cable, it can however be significant over long distances. The signal may also be distorted by pulse broadening or ‘smearing’, because of modal dispersion and material dispersion (*see section 7.6 in Chapter 7 of Year 1 Student Book*). These types of dispersion are effectively reduced, respectively, by using monomode fibres and monochromatic laser light.

Optical fibres are used extensively in communication systems. **Fibre broadband** is the use of optical fibre cable to achieve very high bit rates (see Electronics section 2.3) and hence high download speeds on the internet. Most landline phone networks now use optical fibre cable over long distances, and cable TV companies rely on the ability of optical fibres to transmit digital video signals at high speeds.

QUESTIONS

2. Suggest which are the two most important technical advantages of an optical fibre communications system over a metal wired system.

Electromagnetic waves

Electromagnetic waves in the form of radio and microwaves are a form of **unguided transmission** medium. There is no physical link between the source and the receiver. The electromagnetic waves are spread throughout the air, and are received by antennas tuned to the carrier frequency.

Radio waves and microwaves are part of the electromagnetic spectrum, with wavelengths longer than those of infrared. These are easily generated and can penetrate through walls. They can have wavelengths from 1 mm (microwave) to 10^5 km (long-wave) and frequencies ranging from a few tens of hertz (extremely low frequency, ELF) to 300 GHz (microwave frequencies, or extremely high frequency, EHF). They propagate in different ways, depending on their frequency.

- Lower-frequency radio waves can bypass obstacles such as hills and large structures, because they diffract around them. These propagate as **ground waves**.
- Higher frequencies are diffracted less and travel in straight lines. Some may be refracted by layers of different density in the atmosphere – these are **sky waves** – and also may be reflected from the Earth's surface.
- The greater the frequency of radio waves, the more they are attenuated (suffer energy loss) as they pass through the atmosphere, so very high-frequency waves cannot propagate as sky waves. They propagate along the line of sight and are called **space waves**.

Table 1 shows the frequency bands for radio communications, including their mode of propagation.

Ground waves propagate close to the ground up to hundreds of kilometres. They diffract as they encounter obstacles and closely follow the curvature of the surface of the Earth (Figure 8), enabling coverage to be achieved beyond the horizon. They are mainly transmitted in the VLF, LF and MF bands (see Table 1) and require high transmission power.

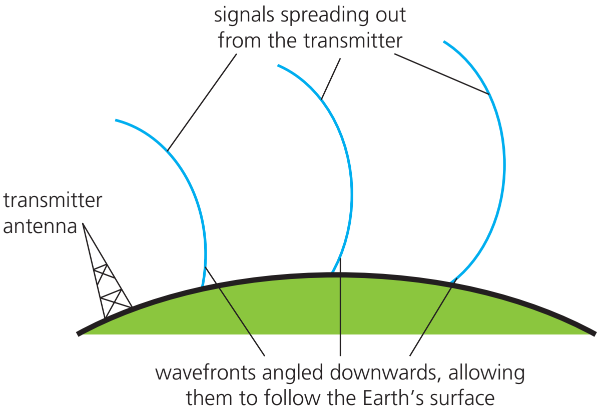


Figure 8 *Ground waves diffract as they propagate and follow the Earth's curvature.*

Radio waves between about 3 and 30 MHz can pass through and be *refracted* in the atmosphere as sky waves, and they can be *reflected* by an

	Band	Frequency	Wavelength	Mode of wave propagation
ELF	Extremely low frequency	3–30 Hz	10 000–100 000 km	Ground
SLF	Super low frequency	30–300 Hz	10 000–1000 km	Ground
ULF	Ultra low frequency	0.3–3 kHz	1000–100 km	Ground
VLF	Very low frequency	3–30 kHz	100–10 km	Ground
LF	Low frequency	30–300 kHz	10–1 km	Ground
MF	Medium frequency	300–3000 kHz	1000–100 m	Ground/sky
HF	High frequency	3–30 MHz	100–10 m	Sky
VHF	Very high frequency	30–300 MHz	10–1 m	Space
UHF	Ultra high frequency	300–3000 MHz	100–10 cm	Space
SHF	Super high frequency	3–30 GHz	10–1 cm	Space
EHF	Extremely high frequency	30–300 GHz	10–1 mm	Space

Table 1 *Frequency bands for radio communications*

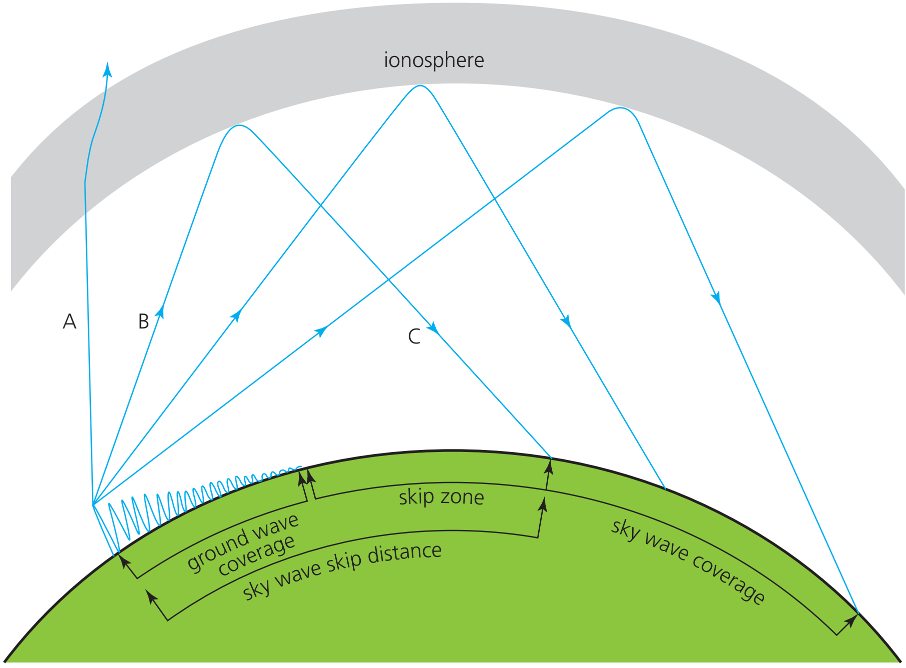


Figure 9 *Sky-wave propagation by reflection from the ionosphere. A is a sky wave that has been refracted by the ionosphere but not totally internally reflected, and escapes into space. B is a sky wave that is totally internally reflected by the ionosphere. C is its return to Earth at a different location, well beyond the range of a ground wave.*

electrically charged layer of the atmosphere called the **ionosphere**. The ionosphere is in the upper atmosphere at an altitude of 80–1000 km where neutral air atoms are ionised by solar radiation and cosmic rays. As the sky waves reach the ionosphere, they are totally internally reflected and come back down to Earth at very large distances from the transmitter (Figure 9). As a result, global communications are possible across continents.

Sky-wave coverage varies according to frequency, latitude and time of day, and also depends on the angle at which the sky wave enters the ionosphere. Sky waves are used by commercial AM radio stations, in military communications and also by amateur ‘radio hams’ to communicate with each other around the world.

Microwave communication links are space waves and propagate by line of sight. They effectively travel in straight lines and cannot pass around obstacles. They range in frequency from about 1 to 300 GHz. Because of their small wavelengths, parabolic transmitter antennas (*see Figure 15 in section 6.2 in Chapter 6 of Year 1 Student Book*) can focus the microwaves into narrow beams for point-to-point communications. They are used in mobile phone networks, in satellite communications, in satellite TV and also in satellite navigation systems (satnavs). They have a large information-carrying capacity. Mobile networks such as 3G phones use microwave links that enable mobile internet download speeds in excess of 6 Mbit s^{-1}, and 4G enables speeds of over 15 Mbit s^{-1}.

Satellite communication uses satellites in orbit around the Earth as relay stations to send communications to any point on the Earth’s surface (Figure 10). The information signal is first transmitted from the ground to the satellite (the **uplink**), where the signal is re-amplified and sent back down to a point on the Earth’s surface, where it is received (the **downlink**).

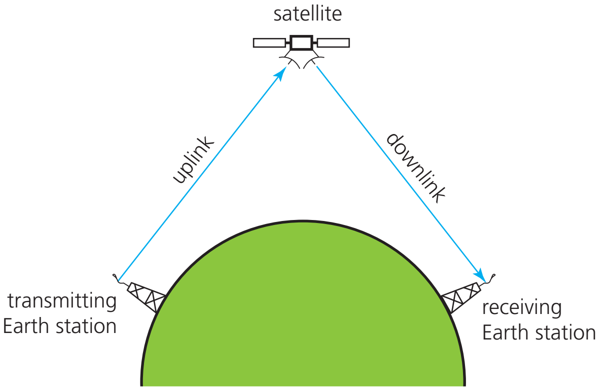


Figure 10 *Satellite transmission*

Microwave frequencies are used for the uplink and downlink. Table 2 shows the frequency bands commonly used for satellite communications and their applications.

Band	Frequency range	Applications
L	1–2 GHz	Mobile telephony and data transmission
S	2–3 GHz	Mobile telephony and data transmission
C	3.4–7 GHz	Fixed telephone services, radio broadcast services, business networks
X	7–8.4 GHz	Government or military communications, encrypted for security reasons
Ku	10.7–18.1 GHz	High-data-rate transmission, television, videoconferencing, business networks
Ka	18.1–31 GHz	High-data-rate transmission, television, videoconferencing, business networks

Table 2 *Satellite communication bands*

Most communication satellites are in **geostationary orbits** (see Chapter 4). This means that the orbital period of the satellite is the same as the rotation period of the Earth. Then, from the ground, the satellite's position in space appears fixed, so the transmitting and receiving antennas do not have to move to track it.

A satellite has a specific frequency for an uplink and a different frequency for the downlink. Uplinks are stronger signals because ground stations have more powerful transmitters than those on satellites. If the downlink band overlapped with the uplink band, then some of the uplink channels might 'swamp' the weaker downlink channels or **de-sense** the satellite transmissions.

The downlink frequency is lower than the uplink frequency. This helps to reduce the attenuation of the relatively weak downlink, since the higher the frequency, the greater the attenuation of a radio wave in the atmosphere.

Security of transmission-path media

Radio and satellite communications are unguided transmission-path media, which can be picked up by any receiver in range. This makes the information insecure. Transmission by guided media is more secure, as the information cannot be easily read unless a physical 'tap' is made into the cable. Optical fibres are particularly secure in this respect, as it is difficult to cut into a fibre without it being self-evident. Wi-Fi links, which are wireless communications between a computer and a network, are potentially insecure – information downloaded from the network onto a computer or other Wi-Fi-enabled device can be read by a user on another device that is using the same link (Figure 11). To stop this happening, Wi-Fi links have a Pre Shared Key (PSK), which is an encrypted code that only the network and your computer knows in order to be able to communicate.

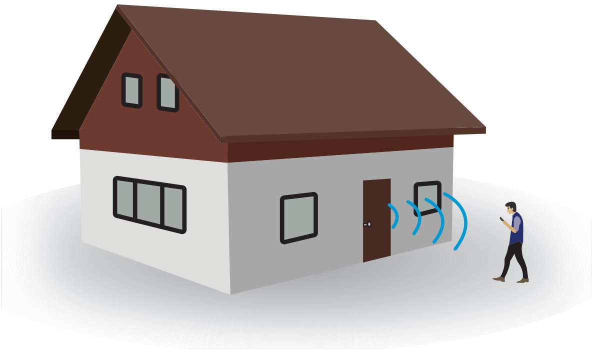


Figure 11 *Radio waves from a wireless network can be intercepted by anyone nearby, unless encrypted.*

Advantages and disadvantages of different transmission media

We can compare the relative merits of different types of transmission media in terms of their nominal data transmission rate (bits per second), frequency range, transmission range, security and cost (Table 3).

Metal cable is the least expensive transmission medium and is easier to install and connect than optical fibres, although its maximum data rate is lower. Optical fibre is capable of very high data rates, but is more expensive to install. It also requires precise interfacing and alignment where fibres are joined together. Radio communications for long-range and microwave links require expensive transmitters and receivers, and their output is unguided and so not secure, although their range is much larger. Satellite communications, while having worldwide coverage, are very expensive owing to the need to launch them into space on rockets.

The effective range of any transmission medium needs to take into account the attenuation of the signal as it is propagated. In metal wire, resistive losses

Transmission media	Data rate	Frequency range	Range	Security	Cost
Metal cable (twisted pair)	1 Gbit s^{-1}	Up to 100 MHz	According to length of cable	Good	Less expensive than optical fibres but more expensive per amount of data carried
Metal cable (coaxial)	10 Mbit s^{-1}	Up to 500 MHz			
Optical fibre	100 Tbit s^{-1}	180–370 THz	According to length of cable	Very good	Costly but more cost-effective than metal cable in terms of data carried
Radio (ground waves)	~ 20 kbit s^{-1}	Up to 1.5 MHz	A few hundred kilometres	Poor	High
Radio (sky waves)	~ 20 kbit s^{-1}	1.5–30 MHz	Global (via reflection of the ionosphere)	Poor	High
Radio (microwave)	Up to 275 Mbit s^{-1}	300 MHz–300 GHz	Line of sight, typically 50–70 km	Poor	High
Satellite communications	Up to 50 Mbit s^{-1}	1–31 GHz	Global	Poor	Very high

Table 3 *Advantages and disadvantages of different transmission media*

dissipate signal energy. In optical fibres, attenuation of the light by absorption in the glass has the effect of reducing data rates with length. So the signal has to be periodically boosted by an amplifier called a **repeater** to maintain signal strength. Radio communications are affected by spreading, absorption and scattering in the atmosphere, which affects the signal strength and data transmission rates.

QUESTIONS

3. State what is meant by *guided* and *unguided* transmission-path media and explain the relative benefits.

4. The ASTRA satellite is a telecommunications satellite that provides satellite TV to millions of homes across Europe. Explain why the dish aerial on a home with a satellite TV installation is in a fixed position, pointing in the same direction.

5. Why do satellite systems use different uplink and downlink frequencies?

6. The TV mast on Emley Moor in West Yorkshire is 330 m tall and transmits TV signals at frequencies between 563 and 754 MHz. Explain why TV masts like Emley Moor have to be so high.

KEY IDEAS

- Transmission-path media can be metal wire, optical fibre or radio waves.
- Metal wire can be in the form of coaxial cable or twisted-pair cable.
- Optical fibres are very long thin glass strands that carry modulated light signals by total internal reflection and are capable of very high data rates.
- Ground waves are radio waves that diffract over the curvature of the Earth over long distances.
- Sky waves are radio waves that are reflected off the ionosphere by total internal reflection and can travel hundreds of kilometres.
- Microwave links are line-of-sight transmission paths to transmit and receive information at microwave frequencies.
- Satellite communications use geostationary satellites and ground stations to receive and transmit information around the Earth. They transmit and receive on different frequencies to avoid de-sensing.
- Guided transmission paths are more secure than unguided ones.

ASSIGNMENT 1: DATA SECURITY

Stretch and challenge

The Government, banks, the Health Service, businesses and the Armed Forces rely on data contained on their computer systems. If the data is damaged, lost, stolen or read by unauthorised users, it could have catastrophic consequences. It is always important to keep data backed up in case of computer crashes or memory corruption. But how can we ensure that data, when it is sent, is not intercepted and read by unauthorised users? When you use the internet to pay for things using a debit or credit card, you will notice that the start of the URL in the address bar of your internet browser changes from *http* to *https*. This shows that a secure connection between your computer and the bank or business has been established. Your card details are **encrypted** to ensure that they can only be read by the bank or business and no one else.

Digital signals lend themselves readily to encryption. There are many ways of encrypting digital data, but we will look at just one method here (Figure A1). The information to be encrypted is the 'plain text'. This is paired with a randomly generated binary number called the 'key' and combined into an encrypted signal using modulo-2 addition (see Electronics section 4.2) based on the EOR operation (Electronics section 4.1) – two binary numbers are added together to give an output the same as the EOR function. The encrypted signal is the 'cypher text'. The key is generated by a Keymat, which provides a succession of randomly generated binary numbers called a 'keystream' from a Pseudo Random Number Generator (PRANG) for each data string transmitted. The randomly generated binary number is only used once. The receiver of the encrypted signal has the same key and applies the same modulo-2 addition operation to de-encrypt the data.

Questions

A1 **a.** For two binary inputs A and B, write down the truth table for the EOR logical operation.

b. Using modulo-2 addition based on the EOR operation, add together:

- **i.** 0 + 0
- **ii.** 0 + 1
- **iii.** 1 + 0
- **iv.** 1 + 1.

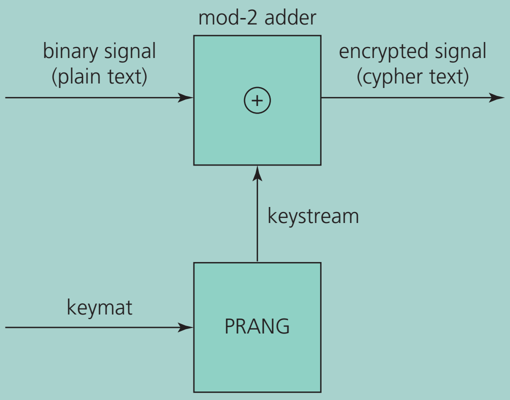


Figure A1 *Encrypting data*

A2 **a.** The characters on a computer keyboard are encoded into eight-bit binary using a convention called the American Standard Code for Information Interchange (ASCII). You can see what the characters are and their binary equivalents by going to the website www.rapidtables.com. Assume we want to transmit the letter 'a'. Find the binary equivalent of the letter 'a' by typing 'a' in the ASCII-to-binary conversion calculator.

b. By adding one digit of each at a time, starting at the right with the least significant bit, add the eight-bit random binary number key 11010001 to the plain-text binary code in part **a** using the EOR operation, to generate an encrypted version of the letter 'a', ready for transmission.

c. The encrypted data is transmitted and received and the same key added at the destination. Using the EOR operation add the encrypted number to the key. What do you notice?

d. If there are *n* bits in a binary number, then the number of different values that can be represented is given by 2^n. How many different keys could be generated using eight-bit binary numbers?

A3 Suggest ways in which this encryption method could be made more secure.

5.3 MODULATION OF ANALOGUE SIGNALS

In Electronics section 5.1 we saw that the purpose of a modulator is to impress the information onto a carrier and blend it in a way that is suitable for transmission. Two methods of modulating an analogue carrier wave such as a radio wave are **amplitude modulation** (**AM**) and **frequency modulation** (**FM**).

Amplitude modulation (AM)

Consider the case of a music recording that is to be transmitted by radio. Radio waves have frequencies of the order of megahertz (MHz) or gigahertz (GHz). The audio signal will have frequencies in the range hertz (Hz) to kilohertz (kHz). In order to transmit this signal, the *amplitude* of the higher-frequency carrier wave can be modulated so that it varies with the variation of the lower-frequency signal. This is illustrated in Figure 12.

Recall from Electronics section 2.3 that the **bandwidth** of an analogue signal is defined as the range between the highest frequency in the signal and the lowest frequency in the signal. AM radio broadcasts are in the frequency range 535–1605 kHz and carrier frequencies of 540–1600 kHz are assigned at 10 kHz intervals. The modulation process will produce a band of frequencies higher and lower either side of the carrier frequency called **sidebands**. Amplitude modulation of a carrier wave normally results in two mirror-image sidebands, and means that the highest audio frequencies in the audio bandwidth will be furthest away from the carrier on either side. Figure 13 shows the sidebands either side of the carrier. On the extreme left of the diagram is the audio frequency spectrum, which is the range of frequencies in the audio signal, with f_L being the lowest and f_H the highest.

The amplitude-modulated (AM) signal thus consists of a carrier with frequency f_C with two frequency sidebands: an **upper sideband** of $f_C + f_H$ where f_H is the *highest* frequency of the modulating signal; and a **lower sideband** of $f_C - f_H$ that extends either side from the main carrier frequency.

The overall bandwidth of the AM signal can be seen to be twice that of the highest audio frequency transmitted: $(f_C + f_H) - (f_C - f_H) = 2f_H$.

The transmission medium must be able to carry at least that frequency. We define the **AM bandwidth** as

$$\text{AM bandwidth} = 2f_H$$

Worked example 1

A carrier wave is amplitude-modulated by a single-frequency 3 kHz audio tone. What is the bandwidth of the transmission?

Two frequencies of 3 kHz will appear either side of the carrier, so the modulation bandwidth will be $2 \times 3\,\text{kHz} = 6\,\text{kHz}$.

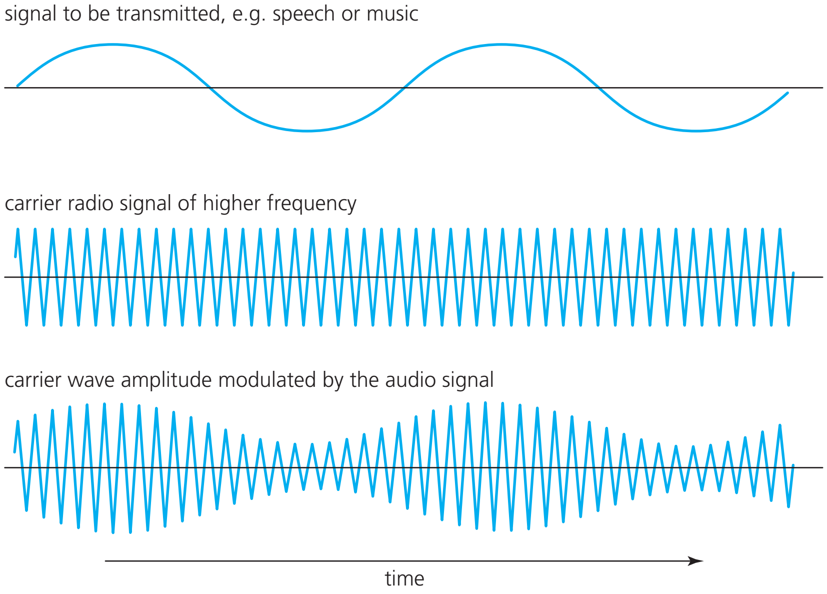


Figure 12 *Amplitude modulation of a carrier wave by an audio signal*

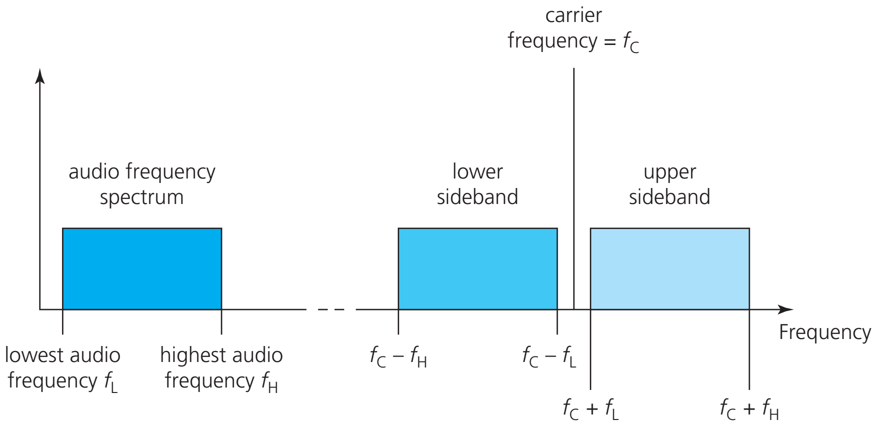


Figure 13 *AM sidebands. The highest audio frequency in the modulating signal will determine the frequency width either side of the carrier frequency.*

QUESTIONS

7. a. An amplitude-modulated signal has a maximum information frequency of 6.5 kHz. What is the bandwidth of the information signal?

b. The signal is to be transmitted on the medium waveband, which extends from 522 kHz to 1710 kHz in the radio band electromagnetic spectrum. How many channels are available for AM transmissions?

Frequency modulation (FM)

Frequency modulation is widely used for many radio communication applications and provides extremely high-quality analogue audio signals. Unlike in amplitude modulation, the amplitude of the carrier signal is kept constant. Instead, the *frequency* of the carrier is modulated in sympathy with the modulation signal. Figure 14 shows how this works.

The transmitted FM signal now has a range of frequencies. The maximum instantaneous difference between an FM modulated frequency and the carrier frequency is called the **frequency deviation** Δf.

We define the **FM bandwidth** as

$$\text{FM bandwidth} = 2 \times (\Delta f + f_M)$$

where f_M is the peak frequency of the modulated signal.

For FM radio stations, frequency deviation is of particular importance, because less deviation means that more channels can be allocated to the frequency space in the FM transmission spectrum used by broadcasters, which is 87.5–108 MHz. In FM broadcasting, the channel spacing (the separation between transmission channels) is 200 kHz, and the maximum frequency deviation of 75 kHz means that there is a 25 kHz ‘buffer’ above the highest and below the lowest frequencies to reduce interaction with other channels. By comparison, AM broadcasting has channel spacing of only 10 kHz, but because the signal is amplitude-modulated, frequency deviation is not relevant.

FM broadcasts are resistant to noise since the amplitude of the carrier remains constant and does not reach the small levels that would be comparable with the amplitude of the noise. The audio quality of an FM signal increases with increasing frequency deviation, which is why FM modulation can be **wide-band** or **narrow-band**. Wide-band FM has a larger frequency deviation and is used by FM broadcasting stations for high-quality music transmissions that require a wider range of frequencies, whereas narrow-band FM is used for speech and data (Figure 15).

Worked example 2

What is the FM bandwidth of an FM signal with a 5 kHz frequency deviation and a maximum audio frequency of 4 kHz?

$$\begin{aligned}\text{FM bandwidth} &= 2 \times (\Delta f + f_M) = 2 \times (5 + 4)\\ &= 9\,\text{kHz}\end{aligned}$$

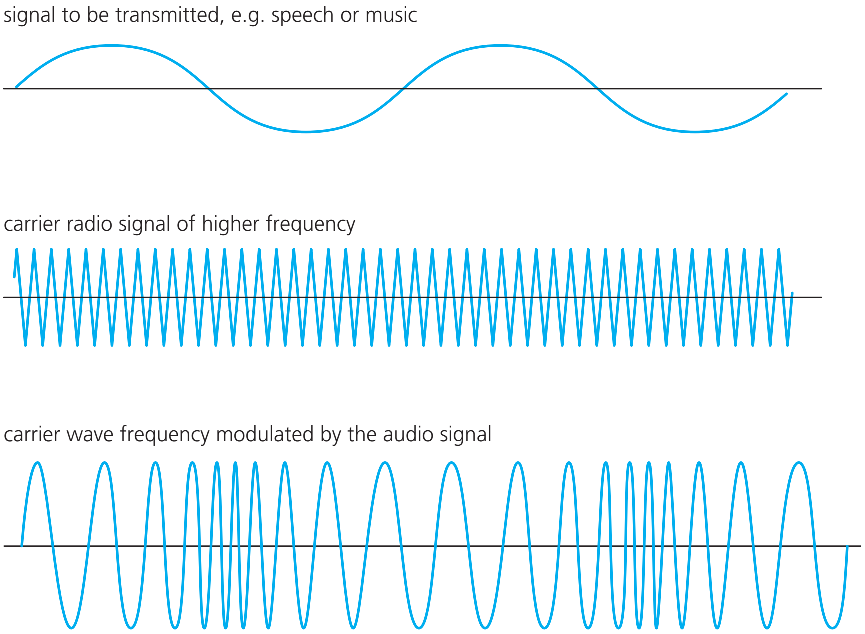
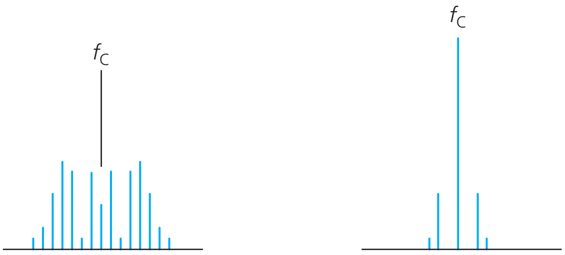


Figure 14 *Frequency modulation of a carrier wave by an audio signal*



wide-band FM

narrow-band FM

Figure 15 *Wide-band and narrow-band FM*

QUESTIONS

8. An audio signal transmitted on an FM music station has a maximum frequency of 18 kHz. If the maximum frequency deviation of the carrier is ± 75 kHz, what is the bandwidth of the FM signal?

9. FM is transmitted between 88 and 108 MHz in the VHF band of the electromagnetic spectrum. If a channel spacing of 200 kHz is allowed for maximum frequency deviation, how many FM radio channels are available in this VHF band?

KEY IDEAS

- Audio signal information can modulate a higher-frequency (radio) carrier wave for transmission over long distances.
- In amplitude modulation (AM), the amplitude of the carrier waveform is varied, corresponding to variations in the audio signal:

  $$\text{AM bandwidth} = 2f_H$$

  where f_H is the highest frequency in the audio signal.
- In frequency modulation (FM), the amplitude of the carrier wave is constant but its frequency is varied corresponding to variations in frequency of the audio signal. The maximum variation in frequency is called the frequency deviation, Δf:

  $$\text{FM bandwidth} = 2 \times (\Delta f + f_M)$$

  where f_M is the peak frequency of the modulated signal.

- The frequency deviation of FM is important because less deviation means that more channels can fit into the same amount of frequency spectrum. For high-quality FM broadcasts, wide-band FM with a larger frequency deviation is used to accommodate a greater range of frequencies.
- FM signals have higher noise immunity than AM signals.

5.4 COMPARISON OF BANDWIDTHS

In digital communications the bandwidth is defined differently.

The minimum bandwidth, in Hz, of a digital communications channel needs to be equal to the number of bits per second that are to be transmitted.

The maximum possible data rate or **data capacity** of a channel is equal to the 2 × the maximum available bandwidth of the communications channel.

(This is to do with the Nyquist relation in Electronics section 2.3 – an analogue signal of bandwidth B can be completely re-created from its sampled form provided it is sampled at a rate equal to at least twice its bandwidth. So the data capacity of the digital communications channel must be able to accommodate the rate at which the original analogue signal was sampled.)

Typical available bandwidths for twisted-pair cable, coaxial cable and optical fibre are shown in Table 4.

Transmission media	Range of frequencies / Hz	Bandwidth / Hz
Twisted-pair cable	0 to 10^9	10^9
Coaxial cable	0 to 10^{10}	10^{10}
Optical fibre	10^{14} to 10^{16}	9.9×10^{15}

Table 4 *Bandwidth availability of twisted-pair cable, coaxial cable and optical fibre*

QUESTIONS

10. a. A ‘superfast’ broadband internet connection has a bit rate of 152 Mbit s^{-1} when downloading information from the server. What is the minimum bandwidth required to carry this bit rate?

b. The internet connection uploads data to its server at a lower bandwidth of 150 kHz. What is the bit rate when uploading information?

c. How much faster is the download speed compared to the upload speed?

d. How long does it take an 8 Mbyte image from a digital camera to be uploaded to a website using the internet connection?

KEY IDEAS

- The data capacity is the maximum rate at which information can be sent over a communications channel in bits per second and depends on the available bandwidth of the medium.
- The data capacity is equal to twice the bandwidth of the transmission medium.
- Optical fibres have the highest bandwidth of transmission media.

5.5 TIME-DIVISION MULTIPLEXING

Transmissions in metal cables and by radio waves can be analogue or digital signals. (Optical cables can only transmit digital signals.) In Electronics section 2.4 we compared the advantages and disadvantages of digital and analogue signals. One advantage mentioned was that more information can be sent by digital signals than by analogue signals, for the same transmission medium. Multiple digital signals can be transmitted almost simultaneously in one cable, for example, by using a technique called **time-division multiplexing** (**TDM**).

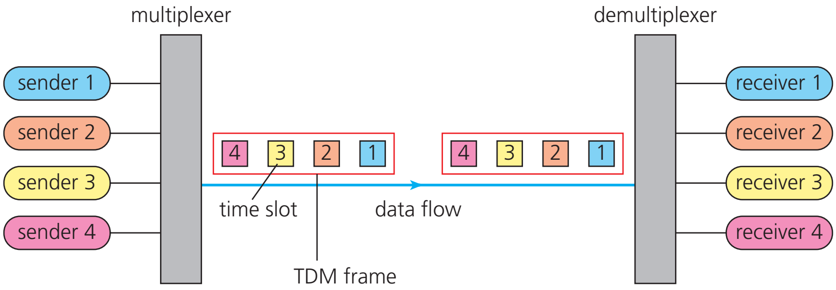


Figure 16 *Time-division multiplexing (TDM)*

The separate digital input signals are combined or **multiplexed** into a single continuous signal. The multiplexed signal is transmitted and then **demultiplexed** at the receiving end, that is, separated out into the original signals. Figure 16 shows the principle of this.

Suppose there are *N* senders (Figure 16 shows four for simplicity) who want to send a signal input over a common transmission path. Each input is assigned a fixed-duration **time slot** in the order of microseconds for transmission. The total transmission time is divided between the *N* users and the time slots are grouped into **TDM frames**. For *N* sending devices, there will be *N* slots in each frame, that is, one slot for each sender. Frames are sent one after another. Each sender takes control of the transmission-path bandwidth turn by turn when its time slot arrives, for a fixed amount of time during which its data is transmitted. Upon reception, the slots of each frame are separated out or demultiplexed and sent to their respective receivers.

In TDM, the data rate of the transmission medium needs to be greater than the data rate of the sending or receiving devices. A user gets the full digital bandwidth of the transmission medium and a single frequency can be used for transmitting the whole time frame, making efficient use of the available frequency transmission bands.

TDM can also be used for analogue signals, but is more suitable for digital signals.

One disadvantage of TDM is that a time slot is still formed even if one of the senders has no data to send, so this is wasteful of transmission-path bandwidth, as one of the slots in the frame will be empty.

QUESTIONS

11. a. Describe *two* advantages of TDM.

b. If there are five senders in a TDM system, and each is allocated a time slot of 125 μs to transmit their data, what will be the total duration of the TDM frame?

c. Describe *two* disadvantages of TDM.

KEY IDEAS

- Time-division multiplexing (TDM) is a method of communication where several users can send signals over a common transmission path and frequency.
- Each user is assigned a time slot for transmission, and a collection of time slots containing different users' data is called a TDM time frame.
- Upon reception, the time frame is demultiplexed and the users' signals sent to their respective receivers.
- In TDM, the data rate of the transmission medium needs to be greater than the data rate of the sending or receiving devices.

PRACTICE QUESTIONS

1. Broadband internet services can make use of the wired telephone network to transmit digital computer data. The system uses a technique called ADSL, where the frequencies used for the data are above the normal range for speech, which is about 300 Hz–4 kHz. Data received by your computer from the internet (i.e. downloaded) are called downstream data, and data sent from your computer to the internet (i.e. uploaded) are called upstream data.

   This is shown in Figure Q1.

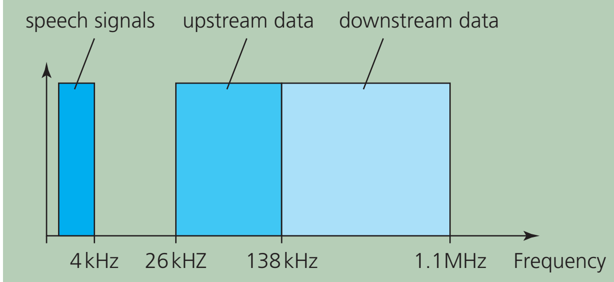


Figure Q1

   a. Calculate the bandwidth allocated to the computer downstream data signals.

   b. The upstream and downstream data bands are each divided into many sections, called sub-channels, each of which is approximately 4.31 kHz wide. Calculate the maximum number of downstream sub-channels possible.

   c. Each one of the 4.31 kHz wide sub-channels is used for data and many sub-channels are used simultaneously for this purpose. This is why ADSL can achieve such high speeds. Because of the complex way in which data is encoded into the sub-channels, each sub-channel can carry data at a maximum rate of about 56 kbps. Show that the maximum downstream data rate for ADSL is about 12 Mbps.

   d. In practice a maximum rate of 8 Mbps is more commonly achieved. Assume that each byte of data has a stop bit, a start bit and a parity bit added to it before sending. Calculate how long it would take at this rate of 8 Mbps to download a movie, with a size of 3 GBytes.

   e. Explain why the maximum possible uploading speed is lower than the maximum possible downloading speed.

   f. State why it is reasonable, given common usage, that the uploading speed is less than the downloading speed.

AQA Electronics Unit 5 June 2012 Q4

2. **a.** An information signal and a carrier wave are shown on the axes in Figure Q2. Make sketches to show how these can be combined to form

   i. an AM signal

   ii. an FM signal.

   b. The information signal has a maximum frequency of 3 kHz. Calculate the bandwidth of the resulting AM signal.

   c. The maximum frequency deviation of the FM carrier is ± 5 kHz. Calculate the practical bandwidth of the resulting FM signal.

AQA Electronics Unit 5 June 2011 Q2

3. PMR446 (Private Mobile Radio, 446 MHz) is a part of the UHF radio frequency range that is available for business and personal use in most of the EU. PMR446 is used in consumer-grade walkie-talkies where analogue FM is used.

   a. i. State the meaning of the term *UHF*.

   ii. State the meaning of the term *FM*.

   iii. Explain how analogue information is carried on FM.

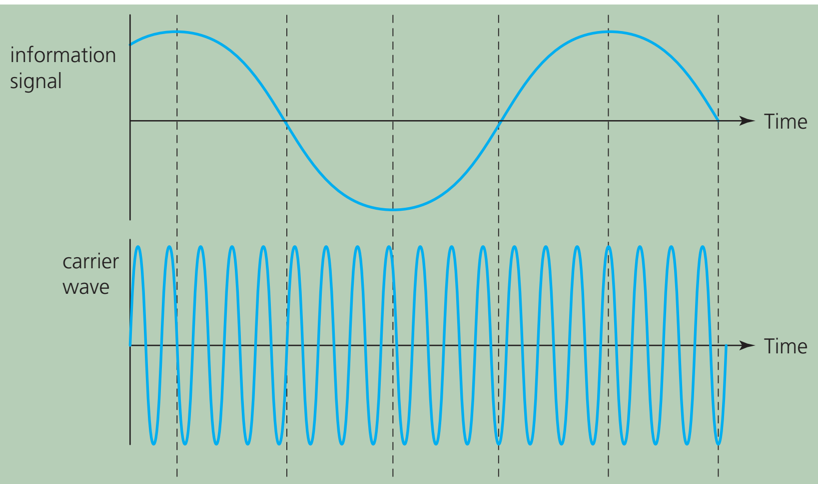


Figure Q2

b. Table Q1 shows the channels available for use.

PMR channel	Frequency / MHz
1	446.006 25
2	446.018 75
3	446.031 25
4	446.043 75
5	446.056 25
6	446.068 75
7	446.081 25
8	446.093 75

Table Q1

- **i.** Calculate the wavelength of the radio signal generated on channel 8.
- **ii.** Calculate the spacing of PMR446 channels.
- **iii.** Show by calculation that a maximum information frequency of 3 kHz and a deviation of ± 2.5 kHz can be supported in this system.
- **iv.** Explain the effect of using ± 5 kHz deviation.

c. All these channels are used for half-duplex communication. Explain the meaning of the term *half-duplex* and how each channel can be used for two-way communication.

AQA Electronics Unit 5 June 2014 Q3 (part)

4. a. Explain what is meant by a *guided* and an *unguided* transmission medium.

b. Guided transmission paths can use metal wire or optical fibre to transmit information. When transmitting information, state the advantages and disadvantages of metal wire and optical fibre in terms of their

- **i.** physical properties
- **ii.** external interference
- **iii.** signal-carrying properties.

5. a. Describe the main difference between amplitude modulation (AM) and frequency modulation (FM).

b. A radio wave has an unmodulated carrier frequency of 150 kHz. An input from a transducer amplitude-modulates the signal by 1.8 kHz. What is the bandwidth of the amplitude-modulated wave?

c. FM radio stations transmit in the VHF radio band. Explain why they do not transmit on medium- or long-wave frequencies.

d. Explain *one* advantage of transmitting digital signals using frequency modulation (FM) rather than amplitude modulation (AM).

ANSWERS TO IN-TEXT QUESTIONS

1 DISCRETE SEMICONDUCTOR DEVICES

1. An n-type semiconductor is one in which the majority charge carriers are electrons. A p-type semiconductor is one in which the majority charge carriers are holes.
2. An extremely high input resistance.
3. They can be damaged by static electricity, as the oxide layer behaves like a capacitor and can store charge.
4. They have an extremely high input resistance, which means that they take essentially no current from a signal source when operating, thus reducing device power consumption.
5. They can be made very small and packed onto integrated circuits in high densities.
6. The voltage applied to the gate, V_{GS}. (With a high enough value of V_{DS}, the MOSFET is in saturation – the current no longer depends on V_{DS}.)
7. **a.** R1 is to ensure that the gate is not left unconnected, when it could pick up static electricity and damage the MOSFET. It also forms a potential divider network, so that the gate voltage is high enough for the MOSFET to turn ON when the sensor is touched.

   b. R2 is a current-limiting resistor to ensure that the current through the LED when the MOSFET channel is formed does not exceed its maximum value before it is damaged.
8. **a.** The Zener diode is connected in reverse bias.

   b. Connecting a Zener diode directly across a varying voltage source could cause a very large current to flow in the diode, destroying it.
9. **a.** Voltage will be equal to the Zener voltage $V_Z = 4.0\,\text{V}$

   b. $95\,\text{mA} + 5\,\text{mA} = 100\,\text{mA} = 0.1\,\text{A}$

   c. $R = \frac{V}{I} = \frac{9.0 - 4.0}{0.1} = 50\,\Omega$

   d. 100 mA
10. Photoconductive mode is when the photodiode is connected in reverse bias, so it only conducts when light is incident on it (apart from a negligible 'dark current').
11. The spectral response is the photocurrent produced by the photodiode at a specific wavelength. The photosensitivity is the photocurrent per watt of power incident on the photodiode and has the unit of $\text{A}\,\text{W}^{-1}$.
12. **a.** Receiving area, $\pi r^2 = \pi \times (0.2 \times 10^{-3})^2$
   $= 4\pi \times 10^{-8}\,\text{m}^2$

   Power at receiving area =

   $$\frac{0.1 \times 10^{-3}}{10^{-4}} \times 4\pi \times 10^{-8} = 1.3 \times 10^{-7}\,\text{W}$$

   b. Photosensitivity $= \frac{56.6 \times 10^{-9}}{1.3 \times 10^{-7}} = 0.4\,\text{A}\,\text{W}^{-1}$
13. **a.** A scintillator is a material that produces a flash of light when a charged particle or high-energy photon passes through it.

   b. Energy of each photon is

   $$E = \frac{hc}{\lambda} = \frac{6.6 \times 10^{-34} \times 3.00 \times 10^{8}}{447 \times 10^{-9}} = 4.4 \times 10^{-19}\,\text{J}$$

   A 2 MeV particle will produce $2 \times 23\,000 = 46\,000$ photons in the scintillator material.

   Total photon energy per incident particle = $46\,000 \times 4.4 \times 10^{-19} = 2.0 \times 10^{-14}\,\text{J}$

The scintillation efficiency is therefore

$$\frac{2.0\times10^{-14}}{2.0\times10^{6}\times1.6\times10^{-19}}\times100\% = 6.3\%$$

14. The Hall voltage V_H generated by a Hall sensor is proportional to $B\cos\theta$, the resolved B component perpendicular to the plane of the semiconducting material. This variation of V_H with angle enables the position and attitude of an object (with a magnet attached) relative to the sensor to be detected.

15. They are non-contact devices. They can measure high rotational speeds.

2 ANALOGUE AND DIGITAL SIGNALS

1. See Figure 1.

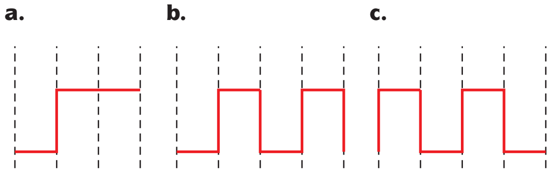


Figure 1

2. **a.** $500\times(2^{30})\times8 = 4.3\times10^{12}$ bits

b. $20\times10^{6}\times8 = 1.6\times10^{8}$ bits per second

3. There are 12 characters in the message (11 letters and one space), each represented by seven bits. Therefore, the number of bits = 12 × 7 = 84 bits.

4. Binary 1111, which is 15 in decimal.

5. **a.** 10000

b. 100000

c. 1000000

6. The information in a digital signal is in the pattern of ON and OFF states. The exact waveform is not crucial provided that the ON and OFF states can be distinguished.

7. Substituting $P = V^2/R$ in the expression for SNR in the text gives

$$\begin{aligned}\mathrm{SNR_{dB}} &= 10\log\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right)\\ &= 10\log\left(\frac{V^2_{\mathrm{signal}}/R}{V^2_{\mathrm{noise}}/R}\right)\\ &= 10\log\left(\frac{V_{\mathrm{signal}}}{V_{\mathrm{noise}}}\right)^2\\ &= 20\log\left(\frac{V_{\mathrm{signal}}}{V_{\mathrm{noise}}}\right)\end{aligned}$$

8. **a.** 40 mT = $(40\times10^{-3})/10^{-4}$ = 400 gauss. Therefore, voltage output = 400 × 2.5 mV = 1 V

b. Resolution = 2/40 = 0.05

9. **a.** Bandwidth = (16 500 − 50) Hz = 16 450 Hz

b. Maximum frequency in the music signal is 16 500 Hz

Therefore sampling rate = 2 × 16 500 = 33 kHz

c. 32-bit sampling means there are $2^{32} = 4.3\times10^{9}$ values

d. Bit rate = sampling rate × number of bits $= 33\times10^{3}\times32 = 1.05\times10^{6}\,\mathrm{bit\,s^{-1}}$

10. **a.** Inductance

b. Capacitance

11. **a.** **i.** $Q = 10^6/10^3 = 1000$

ii. $Q = 10^6/10^4 = 100$

b. The LC circuit **i** with a bandwidth of 1 kHz and $Q = 1000$ has the sharper peak.

12. Rearranging the equation $f_0 = \dfrac{1}{2\pi\sqrt{LC}}$ gives

$$\begin{aligned}C &= \frac{1}{4\pi^2 f_0^2 L}\\ &= \frac{1}{4\pi^2\times(20\times10^{6})^2\times(0.1\times10^{-6})}\\ &= 6.33\times10^{-10}\,\mathrm{F} = 633\,\mathrm{pF}\end{aligned}$$

3 OPERATIONAL AMPLIFIERS

1. $V_{\mathrm{out}} = 10^5\times(60-10)\times10^{-6} = 5\,\mathrm{V}$

2.

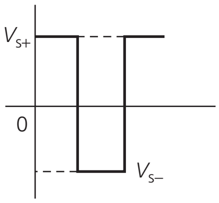


Figure 2

The inverting input V_- is held at 0 V, so $V_{\mathrm{out}} = A_{\mathrm{OL}}V_+$. When the input signal crosses the 0 V reference at the inverting input, the output of the circuit is saturated at its positive supply voltage V_{S+}. When the input goes below 0 V, the output voltage of the circuit immediately switches to its negative saturation supply voltage V_{S-}. Every time the input signal crosses the zero voltage level, the output switches between one saturation level and the other. Regardless of the shape of the input wave, the output is always a rectangular wave. (This op-amp circuit is commonly known as a *zero-crossing detector*, as it switches when the input goes past zero volts.)

3. **a.** The Zener diode provides a steady voltage reference (of 6.2 V) to the inverting input of the op-amp.

   b. The output V_{out} is given by $V_{out} = A_{OL}(V_+ - V_-)$, where A_{OL} is the open-loop gain of the op-amp. The input V_- is held at 6.2 V by the Zener diode. When the slider on R_2 is at the bottom position, the input to V_+ is zero, so the output V_{out} will be saturated to the 0 V supply because of the high gain. The LED will light because it will be forward-biased.

   As the slider on R_2 is moved to the top position, there will come a point where the potential difference at V_+ exceeds that of the Zener voltage at V_-. The output V_{out} then swings to the positive supply voltage of 15 V. There will no longer be a potential difference across the LED and it will not light.

   c. Resistors R_1 and R_3 are current-limiting resistors to ensure that the Zener diode and the LED, respectively, operate within their safe limits.

4. Open-loop gain is the gain of the amplifier without negative feedback, which is very high. Closed-loop gain is the gain that results when negative feedback is applied from the output to the input. The effect of this is to control the open-loop gain and reduce it to a value that is useful for a practical circuit.

5. From $A_{CL} = -R_f/R_{in}$ we get $-50 = -R_f/11.2$, so $R_f = 11.2 \times 50 = 560\,k\Omega$

6. From $A_{CL} = V_{out}/V_{in} = 1 + R_f/R_1$ we get $A_{CL} = 1 + (470/4.7) = 101$

7. The input resistors have to be the same value as the feedback resistor.

8. $V_{out} = (V_+ - V_-) \times R_f/R_1 = (0.050 - 0.008) \times 10/2.2 = 0.190\,V = 190\,mV$

9. **a.** Input A will have $V_{CL} = -12/1 = -12$.
   Input B will have $V_{CL} = -12/2 = -6$.

   b. In millivolts (mV) $V_{out} = (-12) \times 3 + (-6) \times 6 = -72\,mV$

10. **a.** They have to be the same value.

    b. The closed-loop gain is $R_4/R_1 = 10/10 = 1$, so $V_{out} = (5 - 2) \times 1 = 3\,mV$

11. Gain × bandwidth = 6 MHz = 6×10^6 Hz

Gain required	Useful bandwidth
1	6 MHz
300	20 kHz
10 000	600 Hz
30	200 kHz

4 DIGITAL SIGNAL PROCESSING

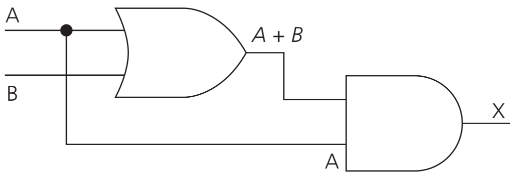
1. **a.** $A.B$

   b. $A.B$ and A

   c.

A	B	$A.B$	$X = A + A.B$
0	0	0	0
0	1	0	0
1	0	0	1
1	1	1	1

2. **a.** See Figure 3.



$X = A.(A + B)$

Figure 3

   b.

A	B	$A + B$	$X = A.(A + B)$
0	0	0	0
0	1	1	0
1	0	1	1
1	1	1	1

   c. The truth table is the same as in question **1** part **c**. So $A + A.B = A.(A + B)$. This means that it does not matter whether the AND gate is followed by the OR gate, or vice versa.

3. **a.** See Figure 4.

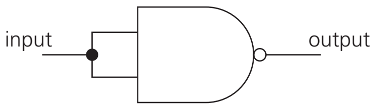


Figure 4

   b. The inputs are equal, so the NAND gate will effectively have a single input, A, of 0 or 1. The output, X, is the inverse of A. The truth table is

A	X
0	1
1	0

   c. This is equivalent to a NOT gate.

4. a. See Figure 5.

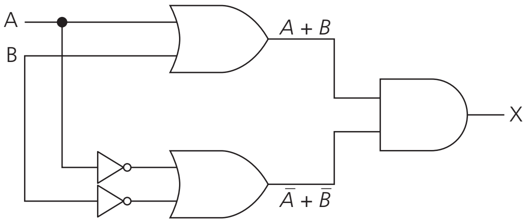


Figure 5

b.

A	B	$\overline{A}$	$\overline{B}$	$A + B$	$\overline{A} + \overline{B}$	X
0	0	1	1	0	1	0
0	1	1	0	1	1	1
1	0	0	1	1	1	1
1	1	0	0	1	0	0

c. The circuit is equivalent to an EOR gate. (It is also equivalent to $X = (A.B) + (A.\overline{B})\overline{\ }$ in the worked example.)

5. See Figure 6.

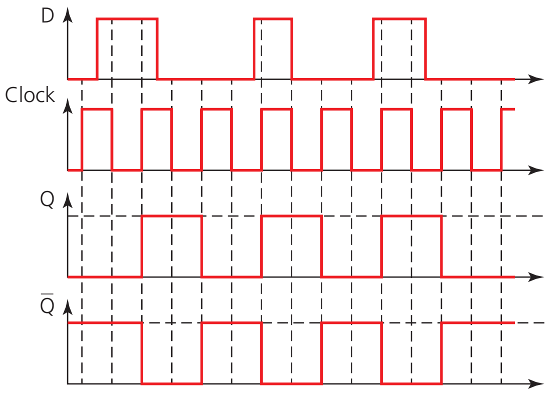


Figure 6

6. Add a fourth D-type flip flop.

7. a.

Count	Q_3	Q_2	Q_1	Q_0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	**1**	**0**	**0**	**1**

b. The outputs of Q_0 and Q_3 are both 1 at (decimal) count 9, so the output from the AND is 1 and the flip flops are reset to 0. This occurs only once in the counting sequence.

8. a. On the positive edge of a clock pulse at the CLOCK input, whatever is the value of the data input at D (0 or 1), this value is transferred to Q. The output of $\overline{Q}$ is always the opposite value to Q. When the clock is not at a positive edge, then the output values are held, that is, they do not change. The outputs can be forced to a high or low state by means of the S (SET) or R (CLEAR/RESET) inputs.

b. See Figure 7.

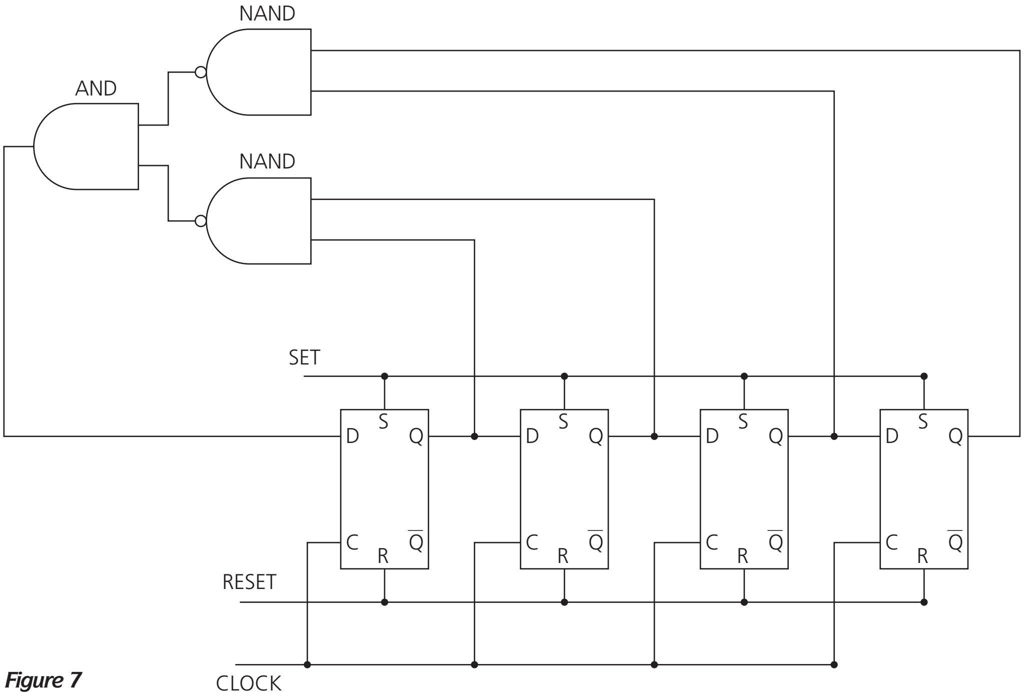


Figure 7

9. Count = $2n = 2 \times 5 = 10$

10. **a.** See Figure 8.

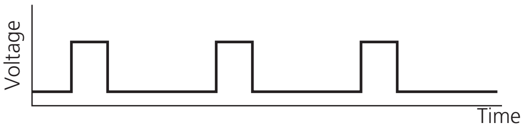


Figure 8

b. 1 : 3

c. Period = ON time + OFF time = 0.02 + (3 × 0.02) = 0.02 + 0.06 = 0.08 s

$$\text{Pulse frequency or clock rate} = \frac{1}{\text{period}} = \frac{1}{0.08} = 12.5\,\text{Hz}$$

11. **a.** From the expression given in the text

$$f = \frac{1.4}{(R_1 + 2R_2) \times C_1}$$

$$= \frac{1.4}{(10 \times 10^3 + 2 \times 20 \times 10^3) \times 220 \times 10^{-9}} = 127\,\text{Hz}$$

$$\text{period} = \frac{1}{f} = \frac{1}{127} = 7.9\,\text{ms}$$

b. From the expression given in the text

$$\text{duty cycle} = \frac{R_1 + R_2}{R_1 + 2R_2} \times 100\%$$

$$= \frac{(10 + 20) \times 10^3}{[10 + (2 \times 20)] \times 10^3} \times 100\% = \frac{30}{50} = 60\%$$

The duty cycle is the proportion of the time that the pulses are ON.

c. Length of ON pulse = 0.6 × 7.9 = 4.7 ms

12. **a.** Period = 5.0 + 2.5 = 7.5 μs

$$\text{Frequency } f = \frac{1}{\text{period}} = \frac{1}{7.5 \times 10^{-6}} = 1.33 \times 10^5\,\text{Hz}$$

b. Rearranging the expression for frequency in the text to make C_1 the subject

$$C_1 = \frac{1.4}{(R_1 + 2R_2) \times f}$$

$$= \frac{1.4}{(5.6 \times 10^3 + 2 \times 56 \times 10^3) \times 1.33 \times 10^5}$$

$$= 90\,\text{pF}$$

c. $$\text{Duty cycle} = \frac{\text{ON time}}{\text{period}} \times 100\% = \frac{5}{7.5} \times 100\% = 67\%$$

d. $$\text{Mark-to-space ratio} = \frac{\text{ON time}}{\text{OFF time}} = \frac{5}{2.5} = 2 : 1$$

13. Pulse frequency is

$$f = \frac{1}{2.2RC} = \frac{1}{2.2 \times 15 \times 10^3 \times 0.1 \times 10^{-6}} = 303\,\text{Hz}$$

14. The resistor R could be a variable resistor or C could be varied.

15. The mark-to-space ratio cannot be varied.

5 DATA COMMUNICATION SYSTEMS

1. **a.** Simplex

b. Duplex – you can both talk to each other (and see each other) at the same time.

c. Half-duplex

d. Half-duplex – a user issues a request for a web document, then that document is downloaded and displayed before the user issues another request.

e. Duplex – you can both talk to each other at the same time.

f. Simplex

2. They can carry more signals at once/greater possible bit rates. They are long-lasting (resistant to corrosion).

3. Guided transmission-path media provide a physical path for the information to travel from the input to the output. Examples are metal wire and optical fibre. Unguided media is where there is no physical path from the input to the output. Radio transmissions are examples of unguided transmission-path media. Unguided media need no wiring, connections, etc., but the data transmitted is less secure.

4. The ASTRA satellite is in a geostationary orbit so it appears to be in a fixed position in the sky above the equator. Therefore, receiving aerials do not have to track it.

5. To avoid being de-sensed where a strong 'up' signal stops the detection of a weak 'down' signal.

6. TV signals at these frequencies are space waves and travel by line of sight. The mast has to be high so that there can be good reception at large distances.

7. **a.** Bandwidth = 2 × 6.5 = 13 kHz

b. From 522 kHz to 1710 kHz is 1188 kHz. With bandwidth of 13 kHz, number of channels is 1188/13 = 91 channels

8. FM bandwidth = $2 \times (\Delta f + f_M) = 2 \times (75 + 18)$ = 186 kHz

9. From 88 MHz to 108 MHz is 20 MHz. With channel spacing of 200 kHz, number of channels is $(20 \times 10^6)/(200 \times 10^3)$ = 100 channels

10. **a.** Data capacity is 2 × available bandwidth B. So data capacity = 152 Mbit s^{-1} = $2B$. Therefore bandwidth $B = (152 \times 10^6)/2$ = 76 MHz

b. Bit rate = $2B = 2 \times 150 \times 10^3$ = 300 kbit s^{-1}

c. $(152 \times 10^6)/(300 \times 10^3) = 507$. The download speed is about 500 times faster than the upload speed.

d. 8 Mbytes $= 8 \times 10^6 \times 8 = 64 \times 10^6$ bits. Therefore upload time $= (64 \times 10^6)/(3 \times 10^5) = 213$ s (about 3½ minutes).

11. **a.** Several senders can use a single transmission path and the user gets the full bandwidth of the transmission medium in their time slot. All users share the same transmission frequency, so it is efficient use of the available frequency transmission bands.

b. TDM frame duration $= 5 \times 125\,\mu\text{s} = 625\,\mu\text{s}$

c. Information from the senders cannot be sent simultaneously. They have to wait their turn. Some time slots in a time frame remain unused if a sender is idle and not sending information, meaning that bandwidth is wasted.

GLOSSARY

Aerial (or antenna) A device that converts an electrical signal into radio waves, or radio waves into an electrical signal.

AM bandwidth Twice the values of the highest frequency of the modulating signal, f_H.

Amplitude modulation (AM) The superposition of a signal wave with a carrier wave to create a transmitted wave which varies in amplitude.

Analogue signal A continuous signal which can have an infinite number of values within limits.

Analogue-to-digital converter (ADC) An ADC is an integrated circuit that takes an analogue input and converts it into a digital output.

AND gate A two input logic gate where both inputs must be 1 for the output to be 1.

Anode A positive electrode.

Astable An oscillator circuit, which produces a continuous output of regular ON–OFF pulses.

Attenuation The reduction of intensity of a signal or wave as it travels through a material.

Avalanche current A high current that flows in a reverse biased diode causing permanent damage.

BANDWIDTH The difference between the highest and the lowest frequency of an analogue signal.

Binary coded decimal (BCD) counter A counter which counts from 0 to 9 and then resets.

Binary down counter A counter consisting of two or more flip-flops which counts down in binary numbers.

Bit (is the basic unit of information both in Computer systems and in digital Communications) The basic unit of information both in computer systems and in digital communications. A bit can have just two values: 1 or 0.

Bit rate The number of bits generated per second during sampling of a pulse code modulated (PCM) signal.

Boolean algebra A mathematical method of manipulating logic states according to a set of rules in which a 1 represents the concept of true and a 0 represents the concept of false.

Breakdown mode The mode of operation of a Zener diode where they operate continuously in revers bias.

Breakdown voltage The voltage causing an avalanche current to flow in a reverse-biased diode.

Byte A unit of information that is eight bits long.

Carrier A wave or transmission medium which carries a signal from source to receiver.

Cathode A negative electrode.

Channel A transmission path along which a signal is carried.

Charge carriers In general terms, a particle which carries an electric charge. In terms of semiconductors, a mobile electron or hole responsible for carrying current in semiconductor materials.

CLEAR/RESET Inputs to a flip-flop which can force a high or low output state.

Clock cycle The period between the start of one pulse and the start of the next in a clock pulse.

Clock pulse A regular sequence of high and low states in a counting circuit.

Clock rate The reciprocal of the period in a clock cycle, also called *pulse frequency*.

Closed-loop gain The gain where a fraction of an operational amplifier's output is fed or 'looped back ' to its input: $A_{CL} = V_{out}/V_{in} = -R_f/R_{in}$.

Coaxial cable A cable that consists of a central core of copper wire which transports the data, surrounded by an insulator to isolate it from a braided metal shield that protects the transmitted data from electrical interference.

Combinational logic circuit A logic circuit whose output depends only on the present *input* states. (Compare with *sequential logic circuits*.)

Communication system A way of communicating information from a source to a receiver via a carrier.

Comparator An operational amplifier used to compare two input voltages.

Conduction band An energy band within a conductor where electrons can move freely.

Counting circuits An electronic device that records the number of times an event has taken place in relation to a clock pulse.

Data capacity The maximum possible data rate of a communication channel which is equal to twice the maximum available bandwidth.

Decibel (dB) Logarithmic scale used to measure intensity of sound.

Demodulator The part of a communication system which extracts the original information signal from the modulated carrier signal that was used to send it.

Demultiplexed Reseparation of multiplex (mixed) signals in a communication system.

Depletion region A region within a semiconductor material that has fewer electrons or 'holes'.

De-sense An effect in satellite communications where a strong upward signal to a satellite can swam a weaker returning signal.

Difference amplifier An operational amplifier used to subtract two voltages from one another.

Digital electronics Electronic systems built from logic gates which operate with digital signals.

Digital signal A signal which can only have two states: *on* or *off* which can be represented in Boolean algebra as a *1* or *0*.

Diode An electric circuit device made of semiconducting material that has low resistance for current in one direction (when forward biased) but very high resistance for current in the opposite direction (when reverse biased).

Divide-by-two counter A flip flop configured to halve the frequency of a clock pulse.

Doping A method of controlling the electrical properties of a semiconductor by introducing small amounts of impurity.

Downlink The signal returned by a satellite in a communication system.

Drain ~ (on the MOSFET) One of the three terminals on a MOSFET through which electrons leave.

D-type flip flop A logic circuit used in counting circuits consisting of two NOR gates, two NAND gates and a NOT gate.

Duplex A communication channel which can carry a signal in both directions simultaneously.

Duty cycle The percentage of the time that the pulse of a clock pulse is in the ON state.

Electrodes Positive and negative connections to a device through which electricity can enter or leave.

Encoded The conversion of a signal from one form into another, such as a keyboard character into a binary number.

Encrypted The recoding of a signal for transmission to increase security of the information.

Energy bands Broad energy states for electrons which exist within a conductor; see conduction *band* and *valence band*.

Energy gap The gap between the conduction band and the valence band which determines whether a material is a conductor, semiconductor or insulator.

Enhancement mode When the current flowing through a MOSFET (source to drain) I_{DS} varies linearly with the voltage between drain and source V_{DS} up to the MOSFET's saturation value.

Exclusive-OR gate (EOR) A logic gate where the output is high if either one of the inputs is 1, but not if both are 1 or both are zero.

Extrinsic semiconductor A semiconductor which has been doped with impurity atoms that modify the material's natural electrical conductivity by adding charge carriers in a precise way. (*See intrinsic semiconductor*.)

Feedback Connecting the output voltage of an operational amplifier to an input to reduce the voltage gain.

Fibre broadband Use of optical fibre cable to achieve very high bit rates in communication systems.

Filter In the context of electronic signals, the removal of unwanted random 'noise' which distorts the original signal.

Flip flop A logic circuit based on a combination of logic gates, which is able to store logic state information.

FM bandwidth FM bandwidth is defined as $(\Delta f + f_m)$ where Δf is the frequency deviation and f_m is the peak frequency of the modulated signal: FM bandwidth $2 \times (\Delta f + f_m)$

Forward-biased A diode connected so that the potential difference across it allows it to conduct.

Frequency deviation The maximum instantaneous difference between an FM modulated frequency and the carrier frequency.

Frequency divider A combination of two D-type flip flops which halves the frequency of a clock pulse.

FREQUENCY MODULATION (FM) A radio communication signal where the amplitude remains constant but the frequency of the signal varies.

Gain The voltage gain, A of an amplifier defined as: A = output voltage/input voltage.

Gain-bandwidth product For an operational amplifier, **gain bandwidth = constant**.

Gate An electronic switching circuit used in digital electronics also called a *logic* gate.

Geostationary orbit An orbit in which a satellite moves in the same direction as the Earth's rotation, above the equator with an orbital period of 24 hours.

Ground waves Radio waves that propagate close to the ground for large distances They diffract as they encounter obstacles and closely follow the curvature of the surface of the Earth.

Guided transmission The process of communicating a signal along a wire or fibre optic cable.

Half-duplex A communication channel which can carry a signal in both directions but only one participant can communicate at a time.

Hall effect The perpendicular deflection of electrons in a semiconductor material carrying an electric current when placed in a magnetic field.

Hall voltage The potential difference produced across a semiconducting material carrying an electric current when placed in a magnetic field.

Hall sensor A sensor used to measure the flux density of a magnetic field by using the Hall effect

Henry (H) The unit of inductance ('resistance' to the change of current in a coil).

Intrinsic semiconductor An undoped semiconductor (*see extrinsic semiconductor*).

Inductance The 'resistance' to the change of current in a coil with the henry (H) as its unit.

Inductor A passive electronic component that stores energy in the form of a magnetic field. It consists of a coil of wire sometimes wound around an iron core.

Input The voltage into an electronic component.

Input resistance The resistance between the V+ and V– terminals of an operational amplifier.

Interference The interaction of two (or more) waves resulting in areas of maximum and minimum intensity.

Inverter A single input logic gate that inverts the input signal also called a NOT gate.

Inverting amplifier A configuration of an operational amplifier where part of the output is fed back to the inverting input.

Ionosphere A region of the upper atmosphere at an altitude of between 80 – 1000 km where neutral air atoms are ionised by solar radiation and cosmic rays. It can be used to reflect radio waves giving them a range of thousands of kilometres.

Johnson counter A counter made form D-type flop flops that has a recirculating data pattern.

Closed-loop gain The gain where a fraction of an operational amplifier's output is fed or 'looped back ' to its input: $A_{CL} = V_{out}/V_{in} = -R_f/R_{in}$.

Light yield The efficiency of a scintillator, measured in the number of photons per MeV.

Logic circuits Circuits made from a combination of logic gates.

Logic gates Electronic switching circuits which give outputs that depend on the status of their inputs, or logic states, and which are represented by pulses of electricity being present or absent.

Logic states An input or output from a logic gate which can be 0 or 1.

Lower sideband The frequency band lower than the carrier frequency of a AM signal created by the modulation process.

Mark-to-space ratio For a clock pulse, the mark-to-space ratio is the ratio of ON time to OFF time:

$$\text{mark-to-space ratio} = \frac{\textit{ON time}}{\textit{OFF time}}$$

Microwave communication A type of space waves used for communication propagated by line of sight. They effectively travel in straight lines and cannot pass around obstacles ranging in frequency from 1 to 300 GHz.

Minimum operation current The smallest value of the current required to ensure a Zener diode operates in breakdown mode.

Modulation The addition of the information that needs to be transmitted onto a signal carrier.

Modulator A device that impresses the information that needs to be transmitted on to the signal carrier and blends it in a way that is suitable for transmission.

Modulo-n counter A counter that counts up to a chosen number by detecting the unique logic state of the number using external logic gates and sending an output to reset the count.

MOSFET (metal oxide semiconducting field-effect transistor) A type of transistor that works by varying the width of a conducting channel along which charge carriers flow.

Multiplex Where multiple analogue or digital data streams are combined into one signal over a common transmission path.

Multiplexed The combination of multiple analogue or digital data streams to form on signal.

NAND gate A two input logic gate that give an inverted output from an AND gate.

Narrow-band FM transmissions with a narrow frequency deviation used for speech and data only (*see wide-band*).

Natural frequency For a vibrating system this is the frequency at which the system oscillates when disturbed.

Negative feedback Connecting the output voltage of an operational amplifier to the inverting input to reduce the voltage gain.

Noise Random voltage fluctuations at different frequencies embedded in the signal caused by interference from the electronic system that processes the signal as well as from external sources.

Non-inverting amplifier A negative feedback configuration of an operational amplifier that gives a positive voltage gain

NOR gate A two input logic gate that gives an inverted output from an OR gate.

NOT gate A single input logic gate that inverts the input signal also called an inverter.

N-type An extrinsic semiconductor where the dopant atoms are 'donors' which provide extra conduction electrons to the semiconductor material.

Offset voltage A small output voltage from an operational amplifier which needs to cancelled or 'nulled out' to ensure proper operation when the inputs to the non-inverting and inverting terminals are the same or zero.

Open-loop gain (symbol) The gain of an operational amplifier where no part of the amplifier's output is fed back to the input: A_{OL}.

Operational amplifier (op-amp) An integrated circuit amplifier that that can be configured in a number of ways to provide amplification of analogue signals for different tasks.

Output The voltage out of an electronic component.

Period The time taken, T, for one complete oscillation or wave; the reciprocal of the frequency, f : $= 1/f$.

Photoconductive mode A photodiode operating in reverse bias.

Photodiode A semiconductor diode that conducts only when light is incident on it.

Photosensitivity The sensitivity of a photodiode to different wavelengths of light or regions of the electromagnetic spectrum.

P-n junction The junction between p-type and n-type semiconductor materials in a semiconductor device where extra electrons in the n-type material will diffuse across the junction and occupy the holes in the p-type material.

P-type A semiconductor that has been doped with 'acceptor' atoms, such as boron, that create an excess of holes as the majority charge carriers.

Pulse frequency The reciprocal of the period in a clock cycle, also called clock rate.

Pulse width The duration of a clock pulse when it is high or ON.

Q factor The Q factor of a resonance peak for an *LC* circuit is defined as $\frac{f_0}{f_B}$ where f_0 is the resonant frequency of the circuit and f_B is the half height bandwidth of the resonance curve:

$$Q = \frac{f_0}{f_B}$$

Quantisation In terms of analogue and digital signals, quantisation refers to the number of available voltage values in the analogue-to-digital conversion process.

***RC* network** A combination of an astable integrated circuit connected with external resistors and capacitors, used to produce clock pulses.

Repeater An amplifier used to boost periodically a communications signal to maintain signal strength.

RESOLUTUION The resolution of a measuring device is the smallest increment in the measured quantity that can be shown on the device. The resolution of an optical instrument is its ability to distinguish very close objects as separate.

Resonance This occurs when a system accepts energy from a driving source at its natural frequency - the amplitude increases greatly.

Resonant frequency The natural frequency of a mechanical or electrical oscillating system.

RESPONSE TIME The response time of a sensor is the amount of time it takes to completely respond to a change in input, often stated as its time to respond from zero input to some specified value.

Reverse photocurrent An electrical current produced when light is incident on a reverse bias photodiode.

Reverse-biased A diode connected so that the potential difference across it does not allow it to conduct.

Sampling The selection of values of an analogue signal for conversion to a digital signal by a sampling gate. The sampling interval is determined by an external clock pulse.

Sampling gate The device which samples values of an analogue signal for conversion to a digital signal determined by an external clock pulse.

Satellite communication Communication which uses satellites in orbit around the Earth as relay stations to send communications to any point on the Earth's surface.

Saturation region The constant current region on a graph of drain current I_{DS} against drain-source voltage V_{DS} for a MOSFET where there is almost no increase in drain current for increasing V_{DS}.

Scintillator A material that produces a flash of light when a particle such as an ion, electron, alpha particle or high energy photon passes through it.

Scintillator efficiency Scintillator efficiency = total energy of light photons produced in scintillator / energy deposited in scintillator by particles or high-energy photons × 100%

Semiconductors A type of material with electrical conductivity between that of metals and insulators; it has a limited number of mobile charges.

Sensitivity The amount of change in output quantity with unit change in input quantity for a sensor. Often this is in terms of the amount of voltage output but for some sensors it can be in terms of current.

Sequential logic circuits A logic circuit whose output depends in part on the sequence of previous input states. (Compare with *combinational logic circuits*).

Shield In a coaxial cable, the braided copper around the signal wire which protects the transmitted data from electrical interference.

Shift register A cascade of flip flops, sharing the same clock, in which the output of each flip-flop is connected to the "data" input of the next flip-flop in the chain, resulting in a circuit that shifts by one position the data stored in it.

Sidebands Frequency bands higher and lower than the carrier frequency of a AM signal created by the modulation process.

Signal-to-noise ratio (SNR) The comparison of signal level and noise level for a waveform carrying information. $SNR_{dB} = 10\log\left(\frac{P_{signal}}{P_{noise}}\right)$

P_{signal} is the power of the signal and P_{noise} is the power of the noise. SNR_{dB} is expressed in decibels.

Simplex A communication channel which can carry a signal in only one direction.

Sky waves Radio waves between 3 and 30 MHz that are refracted in the atmosphere and can be reflected by the ionosphere.

Source The origin of a pd or communication signal.

Space waves Waves such as microwaves used for communication which are propagated by line of sight. They effectively travel in straight lines and cannot pass around obstacles ranging in frequency from 1 to 300 GHz.

Spectral response The relative sensitivity of a photodiode to different wavelengths which can be shown as a graph.

Substrate An underlying layer of a substance such as p-type semiconductor.

Summing amplifier A configuration of an operational amplifier where different voltages are fed into the inverting input and summed.

Tachometer An instrument which measures the working speed of rotating wheels or machinery, typically in revolutions per minute.

TDM frames A series of N timeslots for sending a multiplex signal comprising of N data streams

Threshold voltage V_{th}, the minimum value of the gate voltage, V_{GS} required to form a conducting channel between the drain and the source of a MOSFET.

Time-division multiplexing (TDM) A method of combining several analogue or digital data streams into one signal over a common transmission path such that each data stream is sent within a discrete timeslot within a frame.

TRANSDUCER A device that converts variations in a physical quantity such as pressure, temperature or flux density, into an electrical signal or vice versa.

Transfer characteristic The relationship between output and input of an electronic component or system especially as depicted graphically.

Transfer function The relationship between output and input of an electronic component or system especially as depicted as an equation.

Transistor A semiconductor device with three connections, capable of amplification.

Transconductance A performance parameter for a MOSFET, g_m, which is the change in drain current ΔI_{DS} caused by the change in the voltage between the gate and the source ΔV_{GS}.

Truth table A list of all possible input states of a logic gate, with their corresponding logical outputs.

Twisted-pair cable A cable with pairs of insulated copper strands woven together, bundled and covered with outer plastic insulation. The twisting of the individual pairs helps eliminate electrical interference from adjacent pairs of wires or external sources of electrical noise.

Unguided transmission Transmission of a signal where there is no physical link between the source and receiver such as radio and microwaves.

Uplink In satellite communications, the information signal transmitted from the ground to the satellite.

Upper sideband The frequency band higher than the carrier frequency of a AM signal created by the modulation process.

Valence electrons Electrons which lie within the valence band which cannot move freely.

Valence band An energy band within a conductor where electrons cannot move freely.

Virtual earth analysis A method of understanding the behaviour of an operational amplifier by considering an input to have such a low value, that it is virtually zero (earthed).

Wide-band FM transmissions with a wide frequency deviation used for high quality music transmissions that require a wider range of frequencies than speech (*see narrow band*).

Zener breakdown voltage The voltage, V_z, causing a Zener diode to conduct in reverse-biased.

Zener diode A special type of semiconductor junction diode that allows current to flow in the forward direction, but also allows it to flow in the reverse direction when the reverse-bias voltage across the diode is above a certain value.

INDEX

ACKNOWLEDGEMENTS

The publishers wish to thank the following for permission to reproduce photographs. Every effort has been made to trace copyright holders and to obtain their permission for the use of copyright materials. The publishers will gladly receive any information enabling them to rectify any error or omission at the first opportunity.

Chapter 1

p1: Richard Bizley/Science Photo p1: Scorpp/Shutterstock; p1, left: Xinhua/Alamy; p5, Fig 6: GIPhotostock/Science Photo Library; p8, Fig 13: Ziga Cetrtic/Shutterstock; p11, Fig 18: Heintje Joseph T. Lee/Shutterstock; p13, Fig 25: used with kind permission from Hamamatsu Photonics; p14, Fig 27: Kirill Volkov/Shutterstock; p15, Fig 31: Jakub Krechowicz/Shutterstock

Chapter 2

p20, Fig 4: Andrew Lambert Photography/Science Photo Library; p21, Fig 7: David J. Green/Alamy; p26, Fig 13: T.Dallas/Shutterstock; p26, Fig 15: Zigzag Mountain Art/Shutterstock

Chapter 3

p37, Fig A1: NASA/Science Photo Library; p39, Fig 14: Antoine Beyeler/Shutterstock

Chapter 4

p44, Fig 1: Bloomberg/Getty Images; p52, Fig A1a: John Kasawa/Shutterstock; p52, Fig A1b: Sheila Terry/Science Photo Library

Chapter 5

p60, Fig 2: Jonathan Eastland/Alamy; p60, Fig 3: Concept Photo/Shutterstock; p61, Fig 4: Paulo Afonso/Shutterstock; p61, Fig 5: ra3rn/Shutterstock; p62, Fig 6: yurazaga/Shutterstock; p62, Fig 7: zentilia/Shutterstock