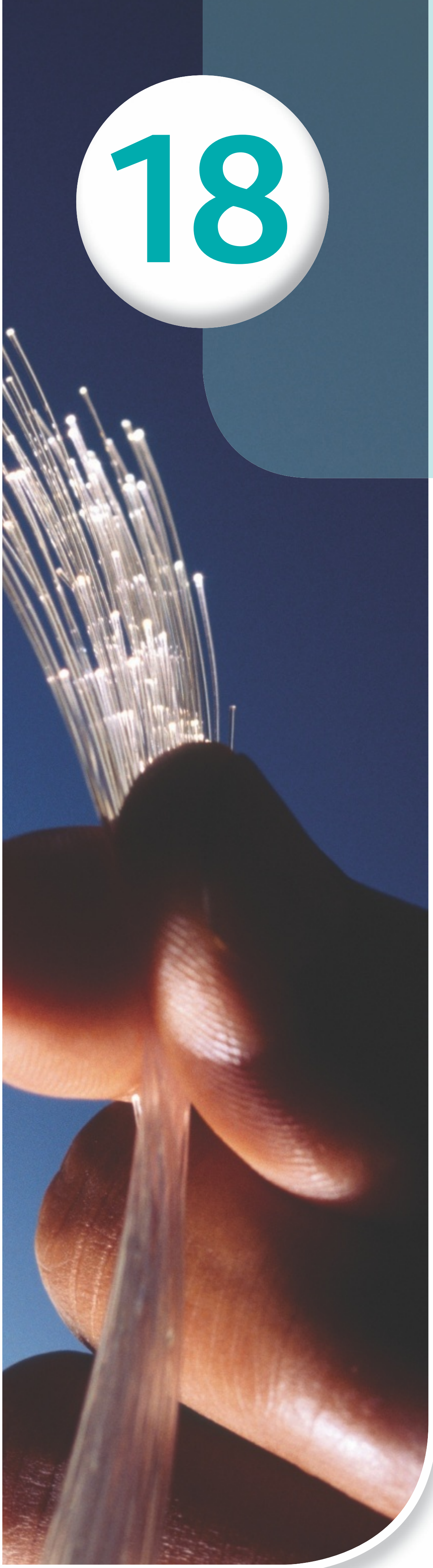


18 Electronics

PRIOR KNOWLEDGE

Before you start, make sure that you are confident in your knowledge and understanding of the following points:

Discrete electronic devices

- A potential divider consists of two resistors where the potential difference across one resistor depends on the ratio of that resistance to the whole resistance: $V_2 = V_{in}\left(\frac{R_2}{R_1 + R_2}\right)$
- A potentiometer can be used as a potential divider to give a continuous range of potential differences.
- Pure semiconductor materials, for example silicon, have poor conductivity.
- A charge, Q, in an electric field, E, experiences a force, $F = EQ$.
- A potential difference across a conductor will produce a potential gradient, $\frac{\Delta V}{\Delta x} = E$, in which charges will experience a force.
- The symbol for a magnetic field is B. A magnetic field or magnetic flux density is measured in tesla (T).

Analogue signal processing

- A capacitor is a component that stores charge and consists of a pair of parallel plates separated by a dielectric.
- The value of a capacitor, C, is given in farad and $C = \frac{Q}{V}$ where Q is the charge stored for a given potential difference, V.
- A coil of wire wound on a magnetisable core will create magnetic flux, ϕ through and around the coil when a current flows through the coil.
- When a magnetic flux linked to a conductor, or coil, changes, an emf is induced in the conductor in a direction so as to oppose the change.
- The emf induced in a conductor of N turns is given by $\varepsilon = -\frac{\Delta N\phi}{\Delta t}$ where $N\phi$ is the flux linkage and $\frac{\Delta N\phi}{\Delta t}$ is the rate of change of flux linkage.
- $\phi = BA$ where B is the flux density and A is the area.
- Simple harmonic motion occurs in a mechanical system when the restoring force, F, on a displaced object is directly proportional to the displacement, x, of the object and in the opposite direction.
  - $a = -\omega^2 x$ where ω^2 is a constant and a is the acceleration.
  - $F = -m\omega^2 x$ where m is the mass of the object.
  - $\omega = 2\pi f$ where f is the frequency of the simple harmonic oscillation.
- In a simple harmonic system, energy is constantly exchanged between one form and another; e.g. in a mass–spring system between kinetic energy and potential energy.

- A mass, m, on a spring of spring constant, k, will oscillate at a particular frequency given by $f_0 = \dfrac{1}{2\pi\sqrt{\dfrac{m}{k}}}$ if it is displaced. The amplitude of the oscillations will die away exponentially at a rate that is determined by the amount of damping in the system.
- If a mass on a spring is hung on a vibrator and driven at various frequencies, f_d, the amplitude of oscillation will be greatest when $f_0 = f_d$.
- The width of the resonance curve will depend on the amount of damping present.

Sequential logic

- Frequency, f, is given by $f = \dfrac{1}{T}$ where T is the periodic time or time for one cycle of an oscillation.
- A capacitor, C, will charge through a resistor, R, with a time constant RC. At time, t, the potential difference across the capacitor will be V.
- $V = V_0\left(e^{-\frac{t}{RC}}\right)$ where V_0 is the maximum potential difference.
- If $t = RC$, V will reach 37% of the starting potential difference, V_0.

Data communication systems

- An audible sound will contain frequencies between about 20 Hz and 20 kHz.
- For all waves, $v = f \times \lambda$ where v is the speed of the wave in m s^{-1}, f is the frequency in Hz and λ is the wavelength in m.
- Total internal reflection occurs when a light wave is incident on a boundary to a transparent medium at an angle greater than the critical angle.
- Diffraction of a wave occurs when the wave encounters an object or a gap. Significant diffraction happens when the gap or object is of similar size to the wavelength of the wave.
- A polarised wave is one that only vibrates in one direction. Any transverse wave can be polarised, so all electromagnetic waves can be polarised.

Modulation and multiplexing

- All periodic analogue signals are a combination of a series of sinusoidal waves of different frequencies.

TEST YOURSELF ON PRIOR KNOWLEDGE

1 A resistor of value 2.2 kΩ is in series with another resistor of value 6.8 kΩ. The combination is connected across a 12 V power supply. Calculate:
  a) the potential difference across the 2.2 kΩ resistor
  b) the current through both resistors
  c) the power dissipated in the two resistors.

2 An electron is in an electric field of field strength 1000 N C^{-1}. Calculate the force on the electron. The charge on an electron is -1.6×10^{-19} C.

3 A potential difference of 5 V is applied across a piece of semiconductor of length 0.5 μm. Calculate the force on an electron in the electric field.

4 A positive charge carrier is in a semiconductor region where the electric field strength is 1.0×10^{6} V m^{-1}. The positive charge is $+1.6\times10^{-19}$ C.
  a) Calculate the force on the charge carrier.
  b) Explain the direction of the force.

5 A capacitor is charged to a potential difference of 5.0 V and is found to have stored 10 μC of charge. Calculate the value of the capacitor in μF.

6 A coil of wire with 1000 turns and area 1.0 cm^2 is moved into a magnetic field of flux density 200 mT in a time of 2.5 s.
Calculate the emf induced in the coil.

7 A mass and spring system oscillates at 3.0 Hz. Calculate the restoring force on the mass of 0.5 kg when the displacement is 5.0×10^{-2} m.

8 A mass of 100 g is hung on a spring of spring constant 4.0×10^{2} N m^{-1}. The mass is displaced a little.
Calculate the frequency with which it will oscillate.

9 One complete cycle of an oscillation takes 0.5 ms. Calculate the frequency, *f*.

10 A capacitor of value 0.22 μF is in series with a 470 kΩ resistor and connected to a 12 V power supply. Calculate the potential difference across the capacitor after 9.0 s.

11 A 1 μF capacitor is discharged through a 100 kΩ resistor.
Calculate the time taken for the potential difference across the capacitor to fall to 37% of its starting value.

12 A radio wave has a frequency of 101.1 MHz. The speed of light is 3.0×10^{8} m s^{-1}.
Calculate the wavelength of the wave.

13 A light ray is incident at an angle just greater than the critical angle on the boundary between a glass block and air.

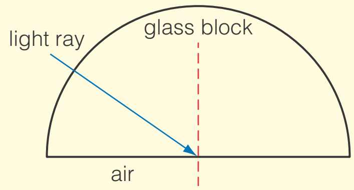


Figure 18.1

  a) On a copy of Figure 18.1, label the angle of incidence.
  b) Complete the path of the light ray.

14 A wave is incident on a gap as shown in Figure 18.2.

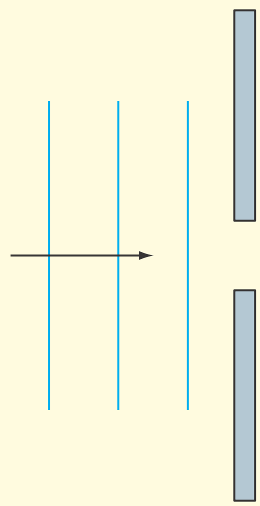


Figure 18.2

On a copy of the diagram, add the next three wavefronts.

15 A microwave transmitter is set up pointing at a receiver.
As the transmitter is rotated through 90° around the axis shown in Figure 18.3, the signal from the receiver falls from maximum to zero. Explain this observation.

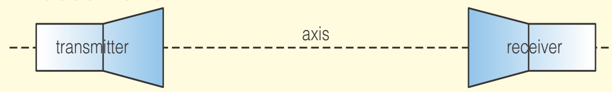


Figure 18.3

Doping is when a small amount of another element is intentionally added to the pure semiconductor material to change its properties.

An n-type semiconductor is material that has been doped so that it conducts with the majority of charge carriers being electrons.

A p-type semiconductor is material that has been doped so that it conducts with the majority of charge carriers being vacancies (holes) where electrons could be.

A discrete semiconductor device is a semiconductor component that performs a single specific function.

TIP

Electronics is essentially practical and you will learn much by trying out the different circuits for yourself. Where possible, enough detail is given in order for you to try out the circuits described.

Discrete semiconductor devices

Pure semiconductor materials, for example silicon, have poor conductivity but their conductivity can be improved and controlled by **doping** the material with an impurity.

If doped with a 5-valent element, for example phosphorus, silicon becomes an **n-type** semiconductor where the majority charge carriers are electrons.

If doped with a 3-valent element, for example boron, silicon becomes a **p-type** semiconductor where the majority charge carriers are holes.

In a doped material, the minority charge carriers are the opposite charge to the majority charge carriers; electrons in the case of p-type material and holes in the case of n-type material. The minority carriers take a very small part in conduction.

There are many different **discrete semiconductor devices**, for example transistors, diodes, light-emitting diodes and sensors for specific tasks. This section looks in detail at a few of these discrete devices.

In many semiconductor devices, there is a junction between p-type material and n-type material. In practice this must be made by doping different regions of a single semiconductor crystal to produce n-type and p-type regions.

The p-n junction

A diode is an example of a semiconductor p-n junction. The construction is shown in Figure 18.4.

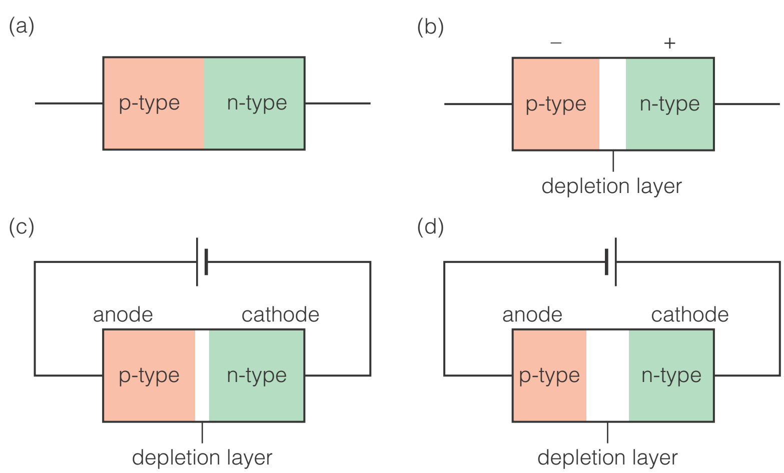


Figure 18.4 (a) A p–n junction. (b) An unbiased p-n junction. Electrons from the n-type region combine with holes from the p-type region to form a depletion layer. (c) Forward bias. The anode is made positive with respect to the cathode. (d) Reverse bias. The cathode is made positive with respect to the anode.

Figure 18.4(a) shows a junction that has just formed. However, immediately holes from the p-type region combine with electrons from the n-type region forming a region with almost no charge carriers called the depletion layer, shown in Figure 18.4(b).

The movement of charge carriers causes a potential difference between the two regions, which prevents any further movement of charge carriers.

A **forward-biased** junction is one where the p-type region is made positive with respect to the n-type region.

Figure 18.4(c) shows the junction with a potential difference applied that makes the p-type region, the anode, positive compared to the n-type region, the cathode. This reduces the width of the depletion layer until, at a certain potential difference, electrons can flow from n-type to p-type and holes from p-type to n-type and a current flows. The junction is said to be **forward biased**. The potential difference at which current starts to flow depends on the type of semiconductor but is typically about 0.7 V for silicon.

In Figure 18.4(d) the potential difference is reversed, making the anode negative compared to the cathode. The depletion layer widens, making it less likely that current can flow. It should be noted that, when reverse biased, conditions are right for the minority charge carriers (electrons in the p-type material and holes in the n-type material) to flow across the junction and so cause a very small current, usually negligible.

Figure 18.5 shows a typical characteristic for a silicon diode.

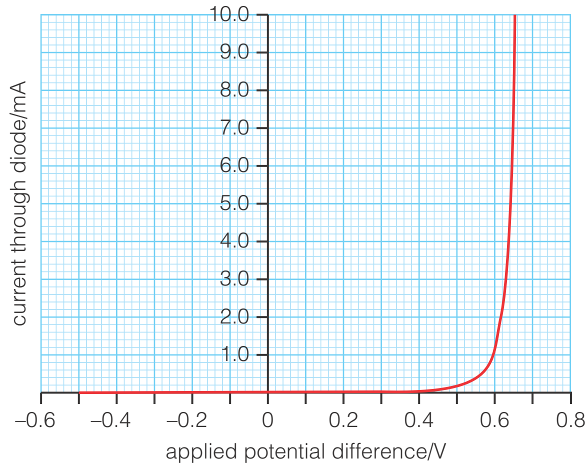


Figure 18.5 Typical characteristic for a silicon diode.

Figure 18.5 shows that the diode will conduct in the forward direction, once a certain applied potential difference is reached. The forward potential difference at which conduction starts will depend on the semiconductor material used: for silicon it is about 0.7 V, for germanium it is about 0.3 V. LEDs, which are composed of a variety of different materials, can have forward potentials that vary from about 1.2 V for an infrared (IR) LED to about 4 V for a white LED.

In the reverse direction, the current is carried by the minority carriers only and is usually negligible.

MOSFET (metal-oxide semiconducting field-effect transistor)

William Shockley, John Bardeen and Walter Brattain are generally credited with the invention of the transistor in 1947, for which they were awarded the Nobel prize in 1956. However, an Austrian-Hungarian physicist, Julius Edgar Lilienfeld, filed the first patent for a field-effect transistor in 1925, although there is no evidence that he actually made one.

There are a number of different types of transistor in common use now. All transistors perform a similar function: a signal on one terminal, usually

The **gate** in a MOSFET is the terminal that is connected to a signal and controls the flow of electrons from the **source** (of electrons) to the **drain** (for electrons). Hence the source is connected to the negative supply.

called the **gate** or the base, controls a current through the device between the **drain** and the **source** or the collector and emitter. Bipolar transistors have regions of p- and n-type semiconductor and a current between the base and the emitter controls a larger current between the collector and emitter.

Field-effect transistors are different in that a potential difference between the gate and the source controls a current between the drain and the source. There are two main families of field-effect transistors: junction field-effect transistors, or JFET; and the insulated gate field-effect transistor, IGFET. The IGFET is more commonly known by a name that describes its construction – the metal-oxide semiconductor field-effect transistor, or MOSFET.

It is this latter type of transistor that we will describe and use in this chapter. In particular we will be describing the construction and use of an n-channel enhancement mode MOSFET. A similar device can be constructed using a p-channel.

The circuit symbol for an n-channel MOSFET is shown in Figure 18.6.

Figure 18.7 illustrates how the transistor is constructed.

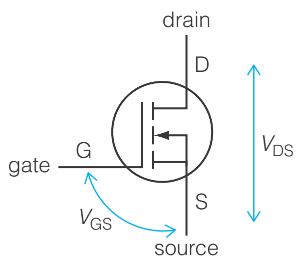


Figure 18.6 Circuit symbol for an n-channel MOSFET.

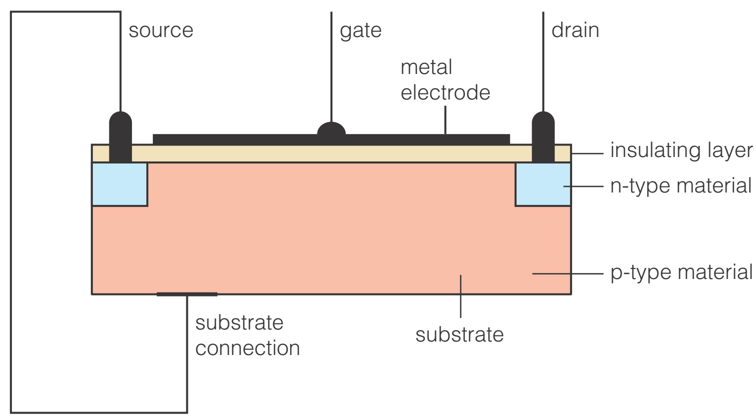


Figure 18.7 Construction of an n-channel enhancement mode MOSFET.

The drain and source are n-type regions doped onto a p-type substrate. Metal connections are made to the two regions.

On the top of the substrate, a layer of metal oxide insulator is deposited. On top of the metal oxide is a metal electrode with the gate connection.

The substrate is connected to the source as shown. If you compare the diagram in Figure 18.7 with the circuit symbol in Figure 18.6, you can see how the circuit symbol represents the construction and connections.

In use, a potential difference is connected between the drain and the source with the drain positive with respect to the source. A small positive potential difference is applied to the gate, as shown in Figure 18.8.

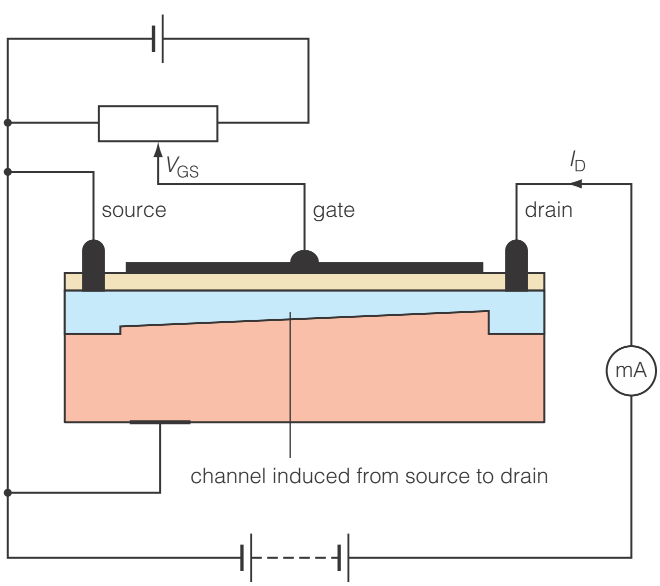


Figure 18.8 A MOSFET with connections.

With no gate voltage, there is a reverse biased p-n junction between the n-type drain and the p-type substrate. This means that almost no current can flow between the drain and the source.

When a potential difference, V_{GS}, is applied between the gate and the source and substrate, the electric field between the gate electrode and the substrate repels the majority charge carriers in the substrate, which are holes, and attracts minority charge carriers, electrons. Thus the region immediately below the insulating layer becomes n-type. This n-type region connects the drain and the source and a current, I_D, can flow through it. Because the channel is n-type in a p-type region, it is called an induced channel.

As there is a potential difference between the drain and the source, the electric field caused by the gate voltage is greater towards the source and so the channel is not uniform between the two electrodes.

The greater the potential difference between the gate and the source, the wider the channel and the greater the drain current that can flow. Thus the MOSFET can act as a switch or, by carefully selecting the gate voltage range, as a variable resistor controlled by the gate voltage.

MOSFET characteristics

The characteristics of an enhancement MOSFET are shown in Figure 18.9.

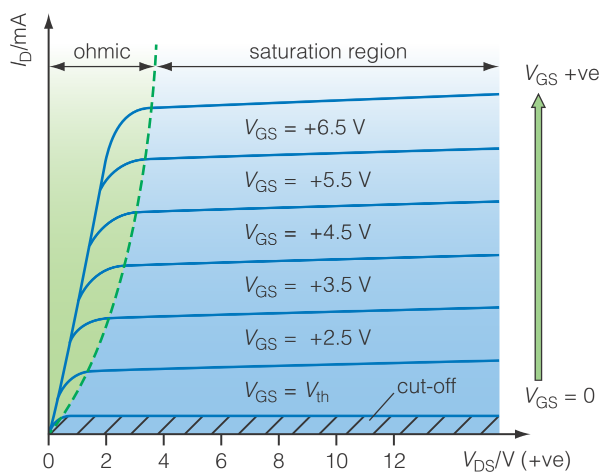


Figure 18.9 Typical characteristics of an enhancement MOSFET.

Notice in Figure 18.9 that there is almost no drain current up to a gate-to-source potential difference, V_{GS}, labelled V_{th}. This is the threshold voltage, below which there is no complete channel, and is typically about 1.5 V depending on the particular MOSFET.

There are three distinct regions:

- $V_{GS} < V_{th}$ – there is no current between drain and source and the transistor acts as an open switch.
- $V_{GS} > V_{th}$ – there is a region where the MOSFET acts like a variable resistor and the drain current, I_D, is dependent on the gate voltage, V_{GS}. This is shown as the ohmic region in Figure 18.9.
- $V_{GS} > V_{th}$ and above a certain value of V_{DS} – the MOSFET is fully on, acting as a closed switch. This is the saturated region, where the MOSFET acts as a constant current source with the drain current depending only on the value of V_{GS}.

The relationship between V_{GS} and I_D is generally shown as a transfer characteristic, as shown in Figure 18.10.

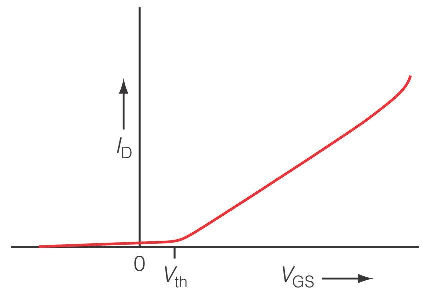


Figure 18.10 Typical transfer characteristic of a MOSFET.

The gain of any device is generally given by the ratio of the change in output parameter divided by the corresponding change in input parameter.

For a MOSFET, the output parameter is the drain current, I_D, and the input parameter is the gate-to-source potential difference, V_{GS}. The gain ratio is therefore $\frac{\Delta I_D}{\Delta V_{GS}}$. It has the unit of A V^{-1} or siemens (S), which is the unit of conductance, G.

This ratio is referred to as the forward transconductance, g_m.

Thus for a MOSFET:

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

Generally, ΔI_D is in milliamps and ΔV_{GS} is in volts, so the unit of g_m is mS.

Typically a MOSFET may have a forward transconductance of about 200 mS.

The very small drain current when the gate-to-source potential difference is zero is caused by conduction by the minority charge carriers across the reverse biased p-n junction between drain and substrate. It is referred to as the zero gate voltage drain current, I_{DSS}, and would be typically a few μA for a small MOSFET. Clearly if this current is too large, the MOSFET becomes less useful as a switch.

Because there is no connection to the gate other than any input, the input resistance of the MOSFET is very high indeed. This means that it can be used, for example, as the input stage for a very high resistance voltmeter or electrometer, as no current will flow from gate to source.

This very high input resistance does have one disadvantage – the gate is susceptible to electrostatic charges from handling or materials in contact with it and a large charge can destroy the transistor. This is why they are stored in conducting plastic bags until used in a circuit.

ACTIVITY

Measuring the characteristics of a MOSFET

A suitable, cheap and readily available transistor to use for this activity is the 2N7000.

The connections to the 2N7000 are shown in Figure 18.11. The case style is known as a TO-92 case.

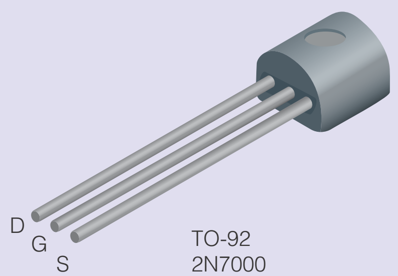


Figure 18.11 A small MOSFET with connections.

On a piece of prototype board or 'breadboard' assemble the circuit in Figure 18.12.

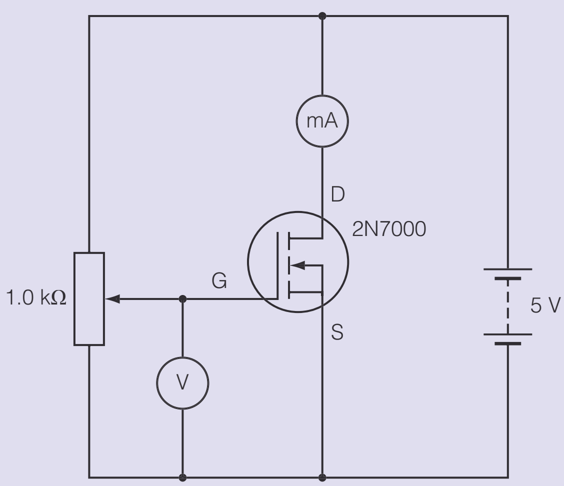


Figure 18.12 Circuit to measure the characteristics of a MOSFET.

Be very careful that before you switch on, the potentiometer is set to give zero voltage at the gate. A higher voltage may cause too large a drain current, which will overheat and destroy the transistor. If in doubt, disconnect the milliammeter before connecting the power supply so that no drain current can flow until you have checked that the gate voltage is zero.

With zero gate voltage, measure the zero gate voltage drain current, I_{DSS}. You will need a microammeter for this, so it is a good idea to use a multimeter as your ammeter so that you can switch ranges easily.

Very slowly increase the gate voltage, V_{GS}, noting the drain current, I_D.

Increase V_{GS} in steps of 0.2 V. At the same time, ensure that the drain current does not go above 200 mA, otherwise the transistor will overheat and is likely to be destroyed. In tests, this limit was exceeded with a gate voltage between 2.4 V and 2.5 V, but values for individual transistors can vary a little.

By plotting I_{DS} against V_{GS} you will obtain the transfer characteristic (Figure 18.13). The gradient of the nearly vertical part of the line will give you a value for g_m.

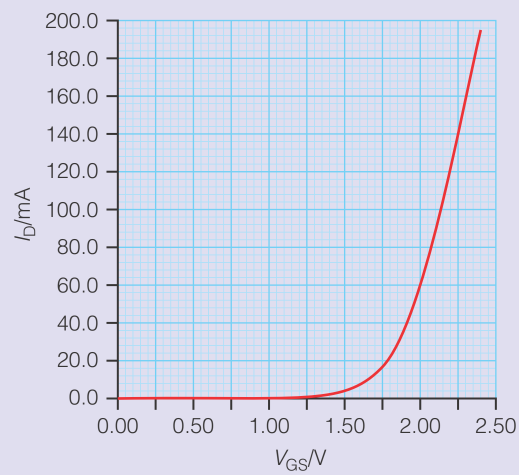


Figure 18.13 A typical transfer characteristic for a 2N7000 MOSFET.

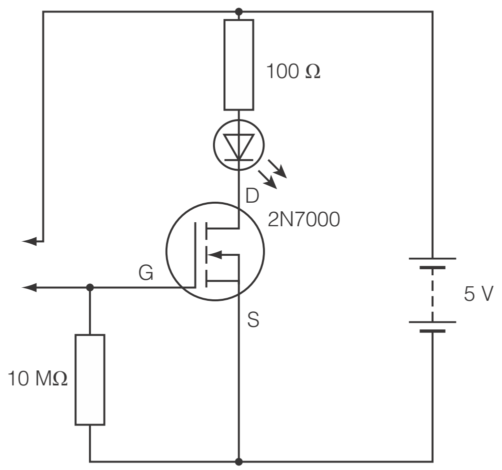


Figure 18.14 A circuit to show a MOSFET as a switch.

Using a MOSFET as a switch

The circuit of Figure 18.12 can easily be adapted to show the MOSFET being used as a switch. Figure 18.14 shows a circuit to switch on an LED when a contact is touched.

The 10 MΩ resistor ensures that V_{GS} normally stays below V_{th}, so keeping the MOSFET switched off. However, when the two contacts shown by the arrowheads are touched with a finger, the finger resistance is likely to be less than 10 MΩ and with the 10 MΩ resistor acts as a potential divider so that $V_{GS} > V_{th}$ and the transistor switches on, causing the LED to light. It may only be necessary to touch the single gate contact to light the LED.

Because I_{DSS} is only about 2 μA, this circuit could be left permanently connected to a battery and it would only draw a current when the contact to the gate is touched.

TEST YOURSELF

1 Describe how doping silicon with phosphorus makes the material n-type with negative charge carriers.

2 Describe how doping silicon with boron makes the material p-type with positive charge carriers.

3 Explain why the construction of a MOSFET means that it has a very high input resistance.

4 Use the graph in Figure 18.13 to show that the transfer characteristic, or g_m, of the MOSFET is about 350 mS.

5 The circuit in Figure 18.15 is for a battery-operated rain alarm. The connections shown by the two arrows are connected to two copper strips which are placed outside. When it rains, a drop of water connects the two copper strips.

a) Explain why this circuit is suitable to be battery operated.

b) The threshold voltage, V_{th}, for this MOSFET is 1.35 V. Calculate the minimum resistance of the rainwater when the alarm just operates.

c) In use, as drawn, the circuit could be damaged if the copper strips were to be shorted out. Explain why damage could occur in this case.

d) Suggest a way in which the circuit could be improved to avoid damage if the copper strips were to be shorted out.

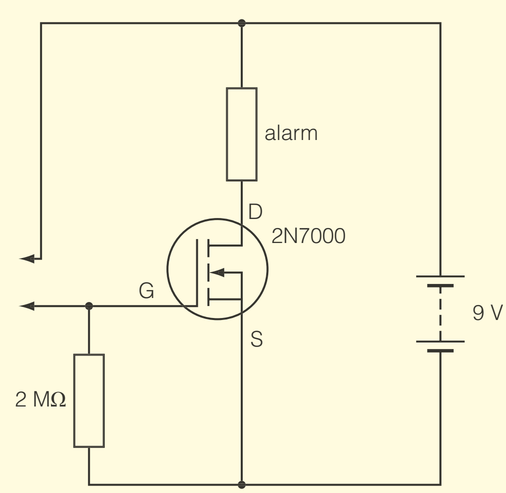


Figure 18.15 A rain alarm. When it rains the MOSFET conducts and the alarm sounds.

Zener diode

A reverse biased p-n junction will normally allow only a very small current to flow, carried by the minority charge carriers. However, if the reverse bias voltage is increased there comes a point when the junction suddenly starts to conduct. This is because the electric field across the junction causes the minority carriers to gain sufficient energy to liberate more electron–hole pairs by their collisions. These, in turn, cause further electron–hole pairs to be liberated and avalanche breakdown occurs, which will often result in a high current causing destructively excessive heating.

However, special diodes are manufactured that have quite low reverse breakdown voltages of 3 V or more. These are called Zener diodes after Carl Zener, who suggested the breakdown mechanism in 1934. The reverse breakdown voltage can be very precisely defined and a wide range of diodes with different breakdown voltages can be purchased. Figure 18.16 shows the characteristics of a Zener diode.

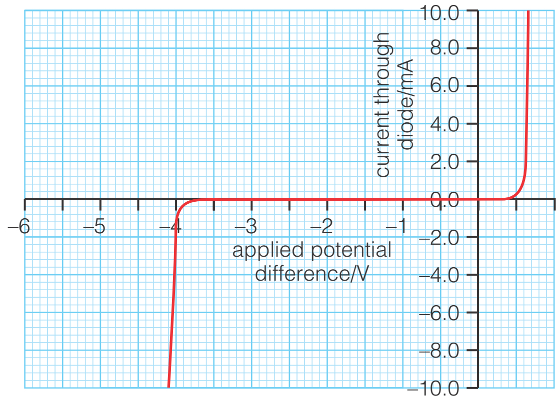


Figure 18.16 Characteristics of a 4 V Zener diode.

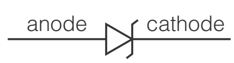


Figure 18.17 Circuit symbol for a Zener diode with the cathode and anode identified.

The circuit symbol for a Zener diode is shown in Figure 18.17.

The Zener diode can be used with a resistor as a simple constant voltage source. A circuit for this is shown in Figure 18.18.

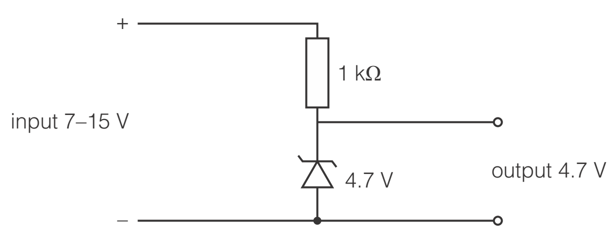


Figure 18.18 A simple constant voltage source.

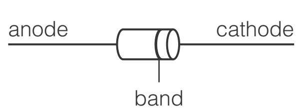


Figure 18.19 Identifying the anode and cathode on a diode.

The anode and cathode of the Zener diode are identified as shown in Figure 18.19.

Notice that in Figure 18.18, the anode is connected to the negative. The value of the resistor is calculated so that the current flowing through the diode is not so large as to cause overheating.

A typical Zener diode might be rated 4.7 V and 100 mW. This means that the maximum current, I, that can flow is given by:

$$P = V \times I \rightarrow \frac{100 \times 10^{-3}}{4.7} \text{A} = 21 \text{ mA}$$

If the maximum input voltage is 15 V then the resistor must have a pd of about 10 V across it when a current of 21 mA flows.

This means that the minimum resistance, R, is given by:

$$R = \frac{V}{I} = \frac{10}{21 \times 10^{-3}} = 470 \ \Omega$$

In practice, one would use a value larger than 470 Ω to ensure that the diode is kept within its operating conditions and does not overheat.

However, the Zener diode used in this way will only give a constant voltage output if the current drawn from the output is kept small so that the potential difference across the resistor is not greater than (input voltage – 4.7) V.

The Zener diode is used to provide a reference voltage for power supply circuits so that the output stays at the voltage defined by the Zener diode.

TEST YOURSELF

6 Draw a circuit to provide a 5 V reference voltage from a 12 V source.

7 A circuit uses a 5 V Zener diode rated at 500 mW.
Calculate the minimum resistance that should be used in the circuit.

8 A sinusoidal wave is fed into the circuit in Figure 18.20.
Sketch the output signal.

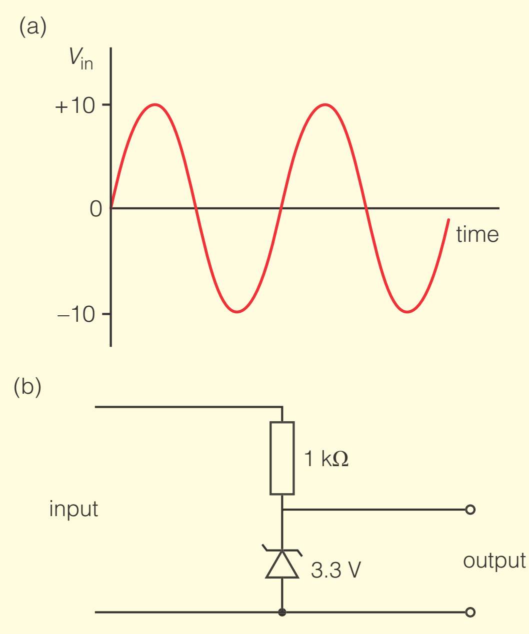


Figure 18.20

Photodiode

A diode that is reverse biased will conduct a very small current unless, as in a Zener diode, energy can be supplied to the lattice in order to release more electron–hole pairs. A photodiode exploits this effect where the energy to release more electron–hole pairs is provided by incident photons.

The construction of a photodiode is different from that of a normal diode in that the junction must be able to be illuminated by incident light (see Figure 18.21). If kept dark, the diode behaves in the same way as any diode, allowing a current to flow when forward biased above a threshold voltage and allowing only a very small dark current, I_D, typically about 1 nA, to flow when reverse biased.

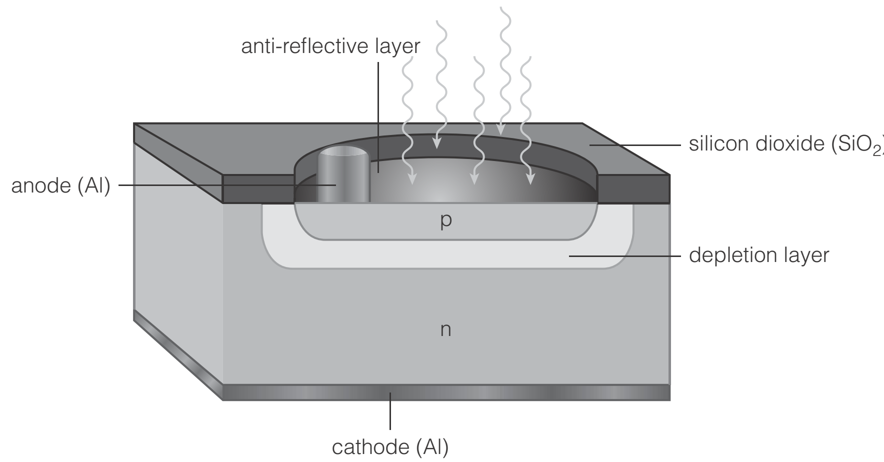


Figure 18.21 Simplified diagram of the construction of a photodiode.

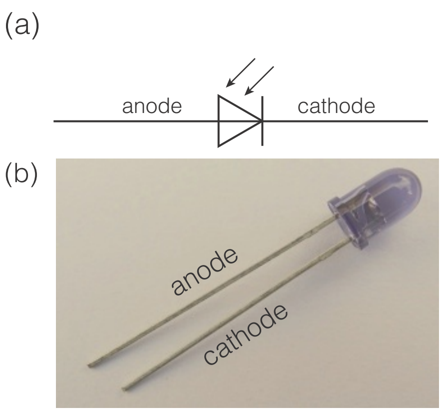


Figure 18.22 (a) Circuit symbol for a photodiode. The symbol is sometimes drawn with a circle around it. (b) A typical package is also shown. The cathode is the shorter lead.

When the reverse biased junction is exposed to light, the current rises very nearly linearly with incident light power to a few milliamps, depending on the reverse bias applied and the light level. This is known as the photo-conductive mode of operation.

The diode is responsive to wavelengths of light from about 400 nm (near IR) to 1100 nm (near UV) with a peak response about 850 nm. However, some photodiodes are manufactured with an integrated daylight filter so that they do not respond to visible light.

Photodiodes specifically designed to receive the signals from a TV remote control, which uses near IR, generally include a daylight filter.

The circuit symbol for a photodiode is shown in Figure 18.22.

Use of the photodiode as an optical detector

The operation of a photodiode as an optical detector can be demonstrated with the circuit in Figure 18.23. Checking the output with a voltmeter will show a change of voltage as the light level changes.

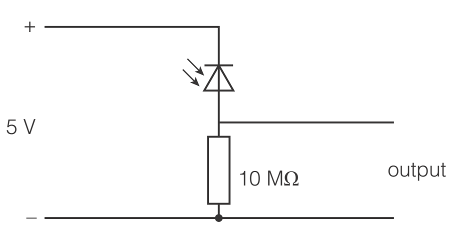


Figure 18.23 Photodiode sensor using a SFH 203 photodiode. The construction of this photodiode is designed to allow it to switch very quickly, which is essential for fibre optic signalling applications.

The output can be connected to a cathode-ray oscilloscope (CRO) and a TV remote control pointed at the photodiode. The CRO will show a series of pulses that correspond to the signal produced by the remote control.

One use of the photodiode is for fibre optic communications where the IR signal pulses are detected by the photodiode. Photodiodes are used in cameras to detect light level and control the focus and the flash. Medical uses include pulse detectors. Some smoke detectors also use a photodiode.

Reversing the positions of the resistor and photodiode in Figure 18.23 will result in an output that falls to a low voltage when light is detected.

Using this output as an input to the gate connection in Figure 18.14 will cause the LED to come on in the dark.

Using a photodiode to detect atomic particles

A number of materials will scintillate, that is, emit a flash of visible light, when bombarded with an atomic particle or high-energy photon, for example X-rays. A number of different scintillation materials are used but generally the intensity of the flash of light is linearly related to the energy of the particle.

By attaching a photodiode to a piece of scintillation material the activity and energy of the radiation can be detected and measured. Some specialist photodiodes can detect single photons.

Generally, photodiodes are used in the reverse biased, or photoconductive mode, although for this application, they are sometimes used in photovoltaic mode.

In photovoltaic mode, the diode is neither forward nor reverse biased. When photons or other atomic particles are absorbed in the depletion layer, a small current will flow in an externally connected circuit.

In both cases, a suitable amplifier can be used to produce a pulse in response to the arrival of the particles.

Hall effect sensor

The Hall effect is the production of a potential difference across a conductor, at right angles to the current, when the conductor is placed in a magnetic field so that the direction of the current is perpendicular to the applied magnetic flux density, *B*.

The potential difference is called the Hall voltage, V_H (Figure 18.24). The Hall voltage is directly proportional to the applied magnetic field.

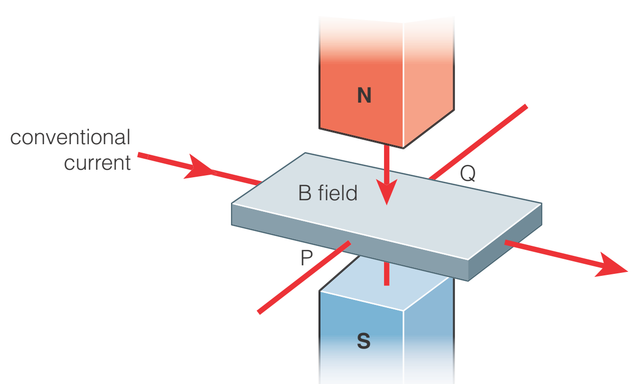


Figure 18.24 When a conductor carries a current in a magnetic field, *B*, as shown, V_H is measured between P and Q. With current and field in the directions shown, Q is negative with respect to P.

Because the Hall voltage is also proportional to $\frac{1}{n \times d}$ where n is the number of charge carriers per unit volume in the conductor and d is the thickness of the conductor, a thin slice of semiconductor with a relatively small number of charge carriers per unit volume gives a high Hall voltage for a given B field.

Suitable slices of semiconductor are manufactured into chips to make Hall effect sensors that give an output that is directly proportional to the applied magnetic field.

Applications of Hall effect sensors

Most mobile phones will have a Hall effect sensor in them that allows them to detect their orientation in the Earth's magnetic field. If you have such a phone, you can test the sensor by downloading an app that gives you access to the data for the magnetic flux detected, usually in μT. Changing the orientation of the phone will show different values for the magnetic field detected. Since the direction of the Earth's magnetic field is known, this allows the attitude of the phone to be detected.

Moving a magnet near the phone will also show the magnetic flux varying.

A Hall effect sensor is commonly used as part of a tachometer, for example in a car, to measure engine revolutions per minute (rpm) and road speed (Figure 18.25). They are also used for determining the position of the crankshaft in the engine for timing purposes.

A sensor designed for these applications will have additional circuitry included on the chip to provide an output that is either a high voltage or a low voltage; it will not be proportional to the magnetic flux density.

TIP

You may have used a Hall effect sensor in your Physics A-level course to measure magnetic flux density.

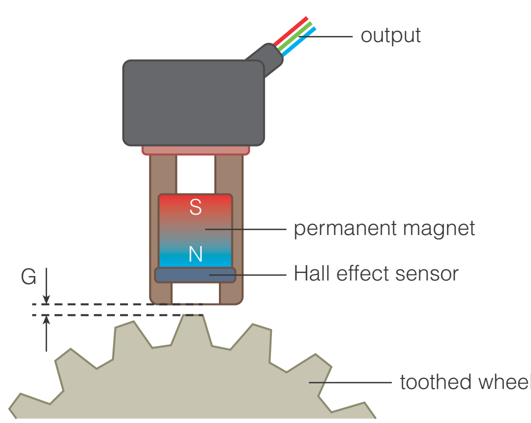


Figure 18.25 Hall effect sensor used in a tachometer. The toothed wheel is magnetic and there is a small gap, G, between the sensor and the toothed wheel. The permanent magnet induces a magnetic effect in the toothed wheel and the variation of this as the wheel turns is detected by the Hall effect sensor, which outputs a pulse each time a tooth passes the sensor.

TEST YOURSELF

9 Explain why a photodiode is able to conduct when exposed to light.

10 Explain why the photodiodes used for TV remote controls need to include a daylight filter.

11 A Hall effect sensor is used to measure the rpm of a motor using an arrangement as shown in Figure 18.25. The toothed wheel has 32 teeth. The output from the sensor is measured and found to be 1600 Hz.
Calculate the rpm of the motor.

12 A Hall effect sensor has an output of 2.5 V with no magnetic field present. The sensitivity of the sensor is 25×10^3 V T^{-1}.
When placed in the Earth's magnetic field at a particular point, the output changes to 3.3 V. Calculate the flux density of the Earth's magnetic field at that point.

Analogue and digital signals

An analogue signal is one that varies with time and can take any value, usually between +V and –V, where V is any potential difference determined by the circuit being used. In an analogue signal, information is represented as a function of voltage and time, e.g. a signal from a microphone that represents a sound (Figure 18.26).

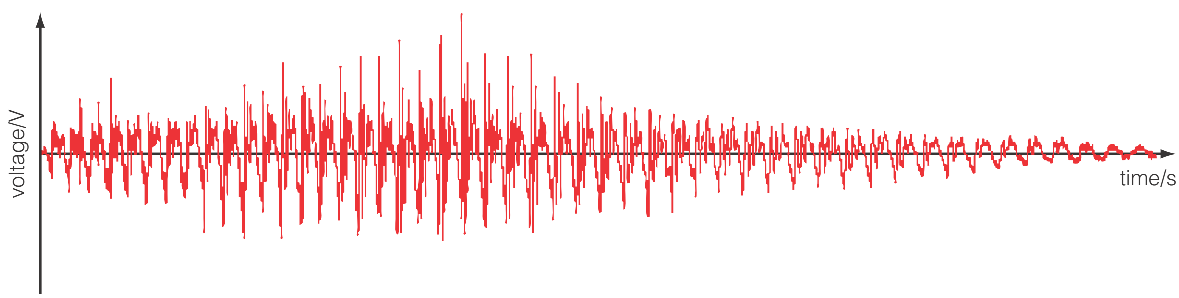


Figure 18.26 The signal from a microphone, in this case the word 'no', is an analogue signal that varies continuously.

A digital signal can take only one of two values: +V and 0 V. In an ideal digital signal, the transition between one level and the other is instantaneous although in practice this is not the case. A value of +V (often +5 V) is generally referred to as 1 or 'high' and a value of 0 V is referred to as 0 or 'low'. In a digital signal, information is represented by a sequence of 1s and 0s (Figure 18.27).

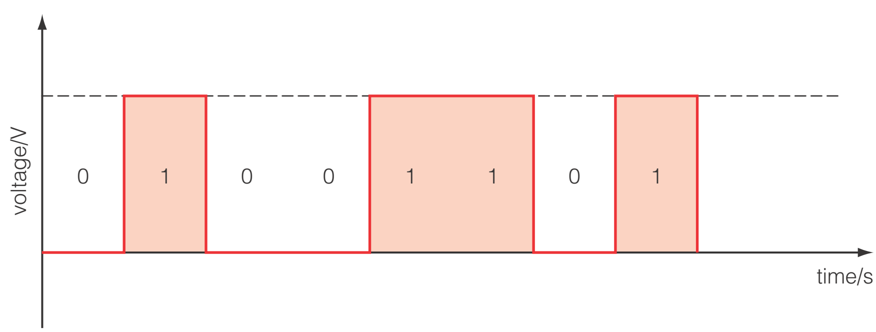


Figure 18.27 A digital signal consists of 1s and 0s only.

Binary counting is a system of counting that uses 1 and 0 only and any decimal number can be represented in binary format.

Each binary digit (0 or 1) is referred to as a 'bit' (**bi**nary digi**t**). In the binary system, the position of a bit determines the power of 2 it represents, in the same way as in the decimal counting system we usually use, the position of a digit determines the power of 10 it represents.

Table 18.1 gives examples of numbers represented in the decimal and binary systems. The decimal number 3456_{10} means 6 units, plus 5 tens, plus 4 hundreds, plus 3 thousands – making 3456 in total. The binary number 1101_2 means 1 unit, plus 0 twos, plus 1 four, plus 1 eight. This makes 1101 in binary, or 13 in decimal.

Table 18.1 Comparison of place value in the decimal and binary systems.

	Decimal					Binary					Decimal equivalent
power	10^3	10^2	10^1	10^0		2^3	2^2	2^1	2^0		
value	1000	100	10	1		8	4	2	1		
example	3	4	5	6	3456_{10}	1	1	0	1	1101_2	13

The bit that represents the units is often referred to as the least significant bit (LSB) and the bit that represents the highest power of 2 in the number is referred to as the most significant bit (MSB).

Table 18.2 shows the equivalent numbers in binary and decimal.

Table 18.2

Binary	Decimal
0000	0
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
1000	8
1001	9
1010	10
1011	11
1100	12
1101	13
1110	14
1111	15

TIP
It is possible to perform all the usual arithmetic with binary as is done in decimal, although it is not required in this unit.

Counting in binary up to decimal 15 requires a 4-bit binary number, which gives 16 possible combinations of digits. It represents the decimal numbers from 0 to 15. In binary, unlike the normal practice in decimal, 0000 is regarded as a piece of information, or the first count.

If b is the number of bits then the total number of possible combinations, n, is given by $n = 2^b$. Thus a 16-bit number can represent $2^{16} = 65\,536$ different combinations. The decimal number represented by the binary number 1111 1111 1111 1111 is 65 535.

A binary number of 8 bits is known as a byte (b). A byte can represent $2^8 = 256$ different combinations. 1111 1111 is 255 in decimal.

A kilobyte (kb) is 1024 bytes. 1024 is used rather than 1000 because 1024 uses all the possible combinations for a 10-bit number. A megabyte (Mb) is $1024 \times 1024 = 1\,048\,576$ bytes.

Converting analogue signals to digital signals

Most sensors initially produce an analogue signal. A large variety of sensors are available and such sensors might detect temperature, pressure, light, magnetic field, sound, potential difference, position or angle.

Data from a computer and other electronic devices is usually digital. Computers process information in a digital form and therefore analogue signals usually need to be converted to digital signals before being processed, stored or transmitted.

Most telephone conversations are transmitted in digital format over long distances and music is stored on a CD or in an mp3 file in digital form.

Mobile phones use digital code to transmit voices and now all television is transmitted in this form. DAB, digital audio broadcasting, is also being introduced across the UK for radio transmissions.

Digital code is almost immune to interference and noise since the receiver only has to distinguish 0 or 1 and there are complicated error correction codes included with all signals. It is also easier to include security measures into digital signals.

Sampling analogue signals

The process of converting an analogue signal to a digital signal is carried out by an electronic circuit called an analogue-to-digital converter (ADC).

The method of converting an analogue signal, particularly a sound signal, is by using pulse code modulation (PCM). In PCM, the analogue signal is sampled at regular time intervals. The instantaneous value of the signal at the sampling time is given the nearest available binary code. An 8-bit or 1-byte code can be used, but 16-bit codes are also used. Thus 00000000 would be the code for the lowest voltage possible and 11111111 would be the code for the highest voltage possible.

Figure 18.28 shows a simplified analogue-to-digital conversion using just 4 bits. No information about the analogue signal is recorded between sample times.

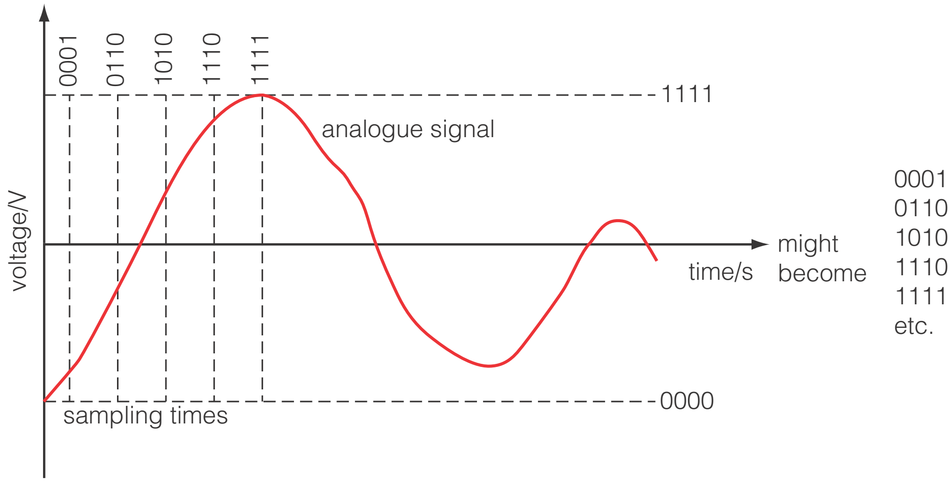


Figure 18.28 An analogue signal is sampled at regular time intervals and the value of the analogue signal at that instant is converted to a digital value by an ADC circuit.

If the analogue signal needs to be reproduced, a circuit called a digital-to-analogue converter (DAC) converts the digital signal to a varying voltage using the digital values recorded. The circuit fills in the missing information by producing a smooth transition of output voltage from one point to the next.

Quantisation

One problem with digital sampling is that of quantisation.

When an analogue signal is sampled, the digital value given to the signal can only be the nearest digital value, as illustrated in Figure 18.29. This means that there is an error in some of the values, called the quantisation error. The more digital levels there are, the smaller this error will be, but it will always be present to an extent.

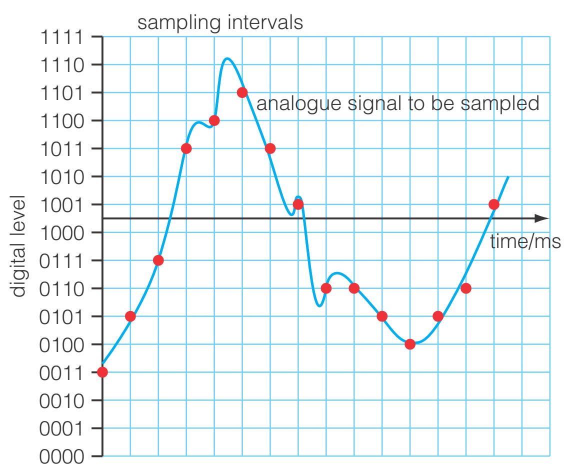


Figure 18.29 The analogue signal can only be given the nearest digital value each time it is sampled.

Noise is an unwanted signal that is added to the original signal to alter it in some way. All signals, whether analogue or digital, will gather noise when they are stored or transmitted.

Figure 18.30 shows how the reproduced analogue signal might look. It is different from the original and, in particular, some of the high-frequency detail has been lost. This effect introduces **noise** into the reproduced signal. The amount of noise introduced will depend on the number of bits used for the sampling levels.

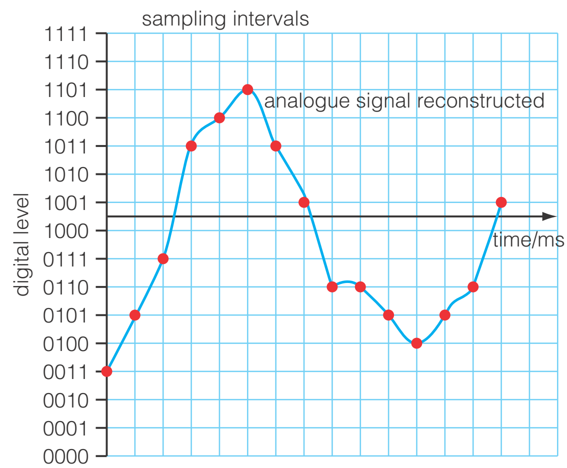


Figure 18.30. The reproduced analogue signal is different from the original because of the quantisation errors. In particular, some detail has disappeared.

Sampling rate

The frequency at which samples are taken also makes a significant difference to the quality of the reproduced signal. Figure 18.31 shows an original signal and the signal reproduced from the digital samples.

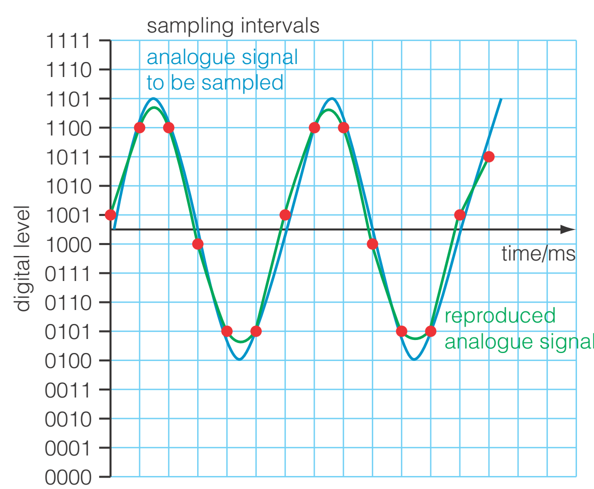


Figure 18.31 An analogue signal is sampled and then reproduced. There is a small error in the reproduced signal but it is quite close to the original.

If the sampling frequency is too small, as in Figure 18.32, then a very large error can occur. This effect is known as aliasing. In a sound signal, this will produce unwanted low-frequency sounds and stops the higher frequency sound being reproduced at all.

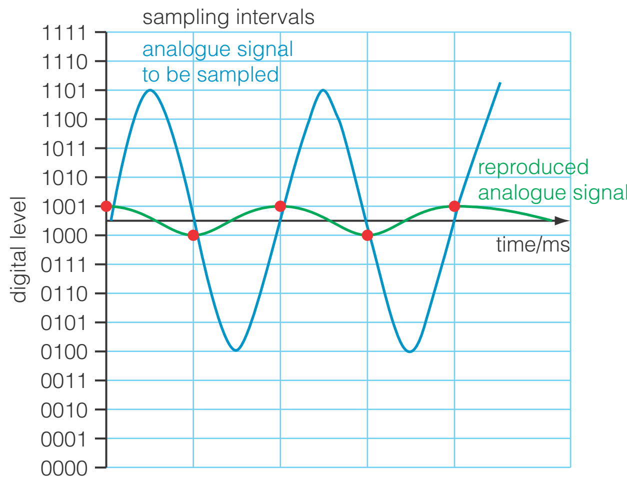


Figure 18.32 The reproduced signal is of lower frequency than the original – in this case the frequency is half the original frequency.

The Nyquist sampling theorem is that the sampling frequency should be at least twice the maximum frequency of the highest analogue frequency to be sampled. Human hearing can detect sounds between about 20 Hz and 20 kHz, so to reproduce the full range of frequencies that can be detected by the human ear, the sampling frequency must be at least 40 kHz. The music for a CD, for example, is sampled at 44.1 kHz using 16 bits for each sample. This gives 65 536 levels for each sample. The sound for a DVD may be sampled at 96 kHz using 24 bits per sample. Clearly this generates a lot more digital data per second of recorded sound but results in better quality reproduction.

Phones are based on an 8 kHz sampling rate using 8 bits per sample, which is why telephone speech lacks high frequencies and why attempting to listen to music through a telephone can be a painful experience! However, it is not quite as simple as this because phones, especially mobile phones, will adapt their sampling rate to an extent depending on the sounds detected. This allows for them to have to transmit less data, for example, during pauses or times of silence.

Advantages and disadvantages of a digital signal

A digital signal that results from the sampling of an analogue signal will always lose some information from the original. The amount of data lost will depend on the sampling frequency and the number of bits per sample and in many cases is unnoticeable. However, it is a disadvantage and some people prefer to retain the original analogue signal, as in vinyl discs or magnetic tape.

Digital signals nevertheless have many advantages over analogue signals when stored or transmitted.

All signals, digital or analogue, will become noisy when transmitted and will also be reduced in amplitude. The signal can be amplified at the receiving end but this also amplifies the noise. Figure 18.33 shows a noisy analogue signal. There is little that can be done to remove the noise other than using an analogue filter to remove high frequencies. The use of filters is covered in the next section.

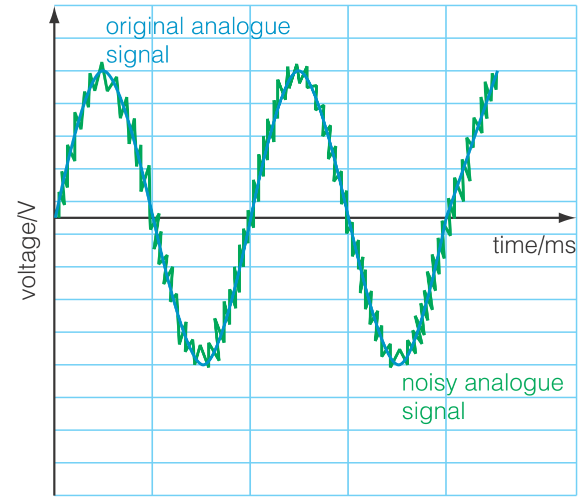


Figure 18.33 A signal will always have some noise associated with it.

When sampling a signal, the sampling levels do not need to be closer together than the noise in the signal. The number of 'levels' available for b bits is 2^b, but there is no point in having a smaller gap between levels than the noise signal present, as this will not improve the quality of the resulting signal. If we can measure the average total signal voltage amplitude, $V_{\text{signal + noise}}$, and the average amplitude of the noise signal alone, V_{noise}, then the maximum number of bits that it is worth using to sample the signal is given by:

$$b = \log_2\left(\frac{V_{\text{signal+noise}}}{V_{\text{noise}}}\right)$$

Clearly this means that the smaller the noise signal, the more bits it is worth using for sampling the signal.

EXAMPLE

A sound signal has an average total amplitude of 100 mV. A noise signal of 0.2 mV is measured. Calculate the maximum number of bits that should be used to sample the signal.

Answer

$$b = \log_2\left(\frac{V_{\text{signal + noise}}}{V_{\text{noise}}}\right) = \log_2\left(\frac{100.2}{0.2}\right) = \log_2(501) \approx 9$$

$2^9 = 512$, so with 9 bits there would be 512 levels for each sample. This would mean that most of the noise would not be encoded in the sampling process.

When a digital signal is transmitted, it will become noisy (Figure 18.34). It will also become less 'square', as transmitting a true square wave requires a very large range of frequencies to be transmitted, or a very high 'bandwidth'. This will be covered in the section on Q factor.

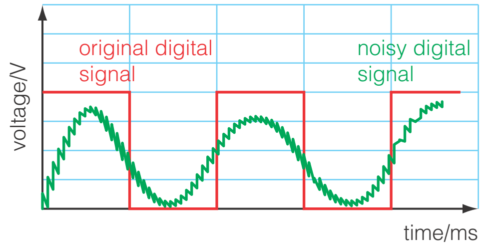


Figure 18.34 A transmitted digital signal is likely to become noisy and be reduced in amplitude.

However, since all that is required is the original 0 and 1, it is quite straightforward to process the noisy digital signal to restore the original signal exactly as it was. If there is a particularly noisy part, the digital signal may be corrupted slightly but the system of transmission will usually also include error correction codes that allow the receiving equipment to reconstruct the original signal, or else request a re-send.

TEST YOURSELF

13 **a)** Write down the digital values of the analogue signal shown in Figure 18.35 at each of the sampling intervals.

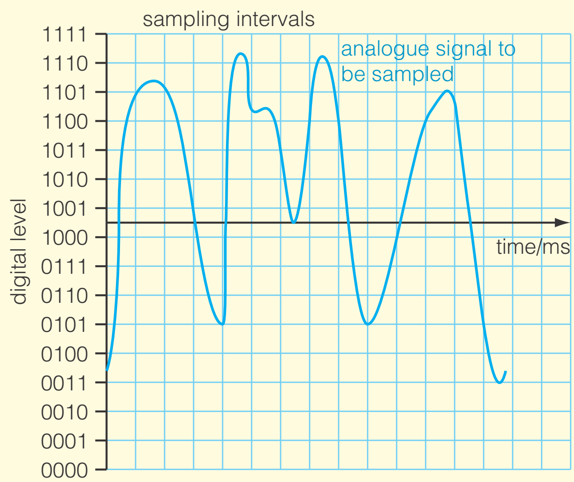


Figure 18.35

b) Reconstruct an analogue signal from your digital data on graph or squared paper and comment on the result.

14 Explain what is meant by quantisation.

15 Explain how a signal becomes noisy.

16 An analogue signal has a maximum peak-to-peak voltage of ±1 V and a maximum noise peak-to-peak signal of ±1 mV. It is to be sampled to produce a digital signal.
Calculate the maximum number of bits to be used to encode the signal.

17 Explain why a digital signal is better than an analogue signal for reducing the effects of noise.

Analogue signal processing

As noted in the previous section, it is sometimes necessary to process analogue signals. Generally this is to remove frequencies that are not required but which are present in the signal. This may be a noise signal, or there might be particular frequencies in a signal that are not wanted – for example in a telephone signal, where any frequencies above 4 kHz need to be removed so that when the signal is sampled at 8 kHz, frequencies higher than 4 kHz are not present, which would cause aliasing.

LC resonance filters

One method of processing analogue signals uses filters to remove unwanted frequencies. The LC resonance filter makes use of two components, the inductor and the capacitor.

Ferrite is a ceramic material that contains iron oxide and the oxides of other metals. It is ferromagnetic but non-conducting.

The **henry** is defined by the emf induced across an inductor when the current changes. When the current through an inductor is changing at a constant rate of one ampere per second ($1\ \mathrm{A\ s^{-1}}$), an inductance of 1 H results in an emf of one volt (1 V) across the inductor. Thus $1\ \mathrm{H} = 1\ \mathrm{V\ A^{-1}\ s}$.

Reactance is the ac equivalent of resistance in a dc circuit. It is measured in ohms.

Inductors

An inductor is a coil of wire, wound on a magnetisable core, either iron or, more usually, **ferrite**. The effect of an inductor in an alternating current (ac) circuit, the inductance, L, is measured in **henry** (H). An inductor has little effect on direct current (dc) in a circuit but has a significant effect on ac in a circuit. The effect depends on the frequency of the ac. A quantity, similar to dc resistance, is defined which is called the **reactance**, X_L, of the inductor. $X_L = 2\pi fL$ so is directly proportional to the frequency, f, of the ac. Like resistance, X_L is measured in ohms.

Capacitors

A capacitor is open circuit to dc but allows an ac current to pass. The reactance of a capacitor, X_C, also depends on the frequency of the ac and is given by $X_C = \dfrac{1}{2\pi fC}$, also measured in ohms. As can be seen, this means that the reactance of a capacitor is inversely proportional to the frequency of the ac.

Figure 18.36 shows how the reactance of a capacitor and an inductor change with frequency.

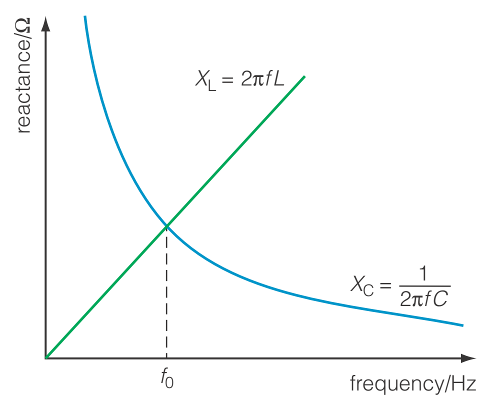


Figure 18.36 The reactances of an inductor and capacitor depend on frequency. At the frequency f_0 their reactances are equal.

Parallel LC resonance circuit

When an inductor and a capacitor are connected in parallel in an LC circuit, the total reactance of the parallel combination is a maximum at f_0. This means that the current through the parallel combination is a minimum and the pd across the combination is a maximum. The frequency, f_0, is known as the resonant frequency and is given by $f_0 = \frac{1}{2\pi\sqrt{LC}}$.

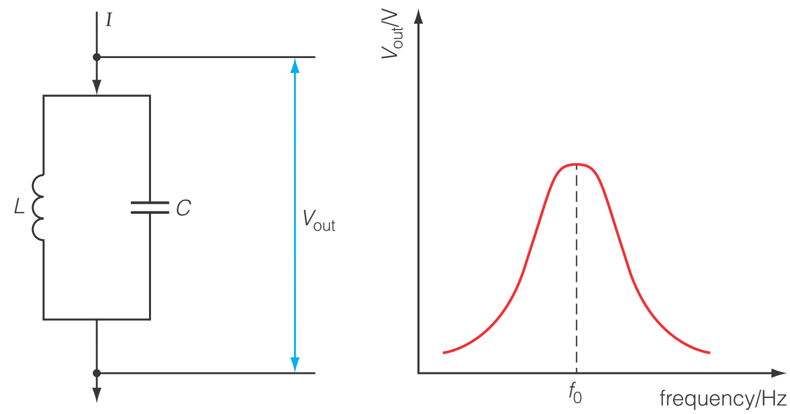


Figure 18.37 An inductor and capacitor in parallel gives a maximum output pd at $f_0 = \frac{1}{2\pi\sqrt{LC}}$.

Mechanical analogy

A mechanical analogy to this parallel LC circuit is a mass on a spring system. In this analogy, the spring acts like the capacitor with the capacitance, C, analogous to the reciprocal of the spring constant, $\frac{1}{k}$. The mass acts like the inductor with the inductance, L, analogous to the mass, m.

The energy stored in a capacitor is given by $E = \frac{1}{2}\frac{Q^2}{C}$, where Q is the charge stored. The energy stored in a spring is given by $E = \frac{1}{2}kx^2$, where k is the spring constant and x is the extension. Comparing the equations, C is analogous to $\frac{1}{k}$.

The energy stored in a spring is transferred to the mass as kinetic energy, $E = \frac{1}{2}mv^2$. The energy stored in the magnetic field in an inductor is given by $E = \frac{1}{2}LI^2$. Here L is analogous to m.

A mass on a spring has a resonant frequency $f_0 = \frac{1}{2\pi\sqrt{m/k}}$ so, using the analogy, the resonant frequency of the capacitor and inductor system is $f_0 = \frac{1}{2\pi\sqrt{LC}}$.

Displacing a mass on a spring without friction will cause it to oscillate forever at its resonant frequency but, of course, it will be damped by friction and its amplitude will reduce exponentially. The rate of reduction will depend on the amount of damping. The amount of damping affects the width of the resonance curve when the mass on a spring is driven by a driving oscillator.

If a brief pulse is applied to the LC circuit, it will oscillate at the resonant frequency. In an ideal circuit with zero resistance the oscillations would continue indefinitely, but in practice they will die out exponentially as energy is lost heating the resistance. Thus resistance in the circuit causes damping. The width of the resonance curve therefore depends on the resistance in the circuit, particularly the dc resistance of the inductor. The lower the resistance, the higher and narrower is the resonance curve.

Q factor

The circuit in Figure 18.37 acts as a filter for a signal. The output voltage at low frequencies is small and is also small for high frequencies. It is sometimes known as a band pass filter. It means that a range of frequencies can be selected from a much wider range. This circuit is used as a radio tuner to select just the frequency required for a particular radio station. It could also be used, for example, to limit the range of frequencies in an audio signal to 4 kHz before sampling for telephone transmission.

The Q factor or quality factor is similar to that for a spring system and defines the width of the resonance curve. It describes how selective the circuit is for a particular range of frequencies. A wide resonance curve has a low Q factor while a sharp resonance curve has a high Q factor.

The bandwidth of the signal, f_b, is the range of frequencies over which the output power level, P_{out}, is greater than 50% of the maximum output at f_0 (Figure 18.38).

Since electrical power is given by $P = \dfrac{V^2}{R}$, then to halve the power requires a reduction in output voltage, V_{out}, to 0.707 or 70.7% of the maximum value since 0.707 × 0.707 = 0.500.

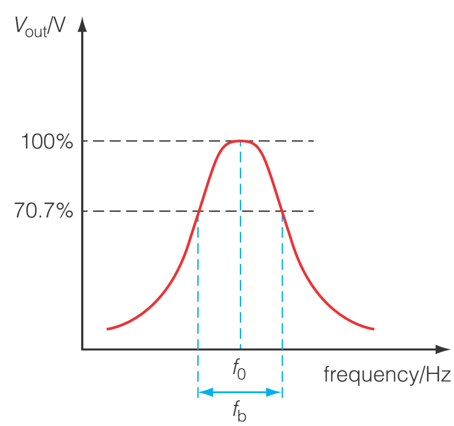


Figure 18.38 The bandwidth, f_b, of the filter is the range of frequencies over which the output power is greater than 50% of the maximum at f_0.

This means that the output voltage must be greater than 70.7% of the maximum at f_0.

The Q factor is defined as:

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

A radio tuner will have a high Q factor, typically as high as 400, whereas an audio band pass filter may have quite a small Q factor.

EXAMPLE

A filter circuit has a resonant frequency, f_0 = 2.5 kHz and a bandwidth of 3.0 kHz. What is the Q factor for this circuit?

Answer

The Q factor for this circuit is given by:

$$Q = \frac{f_0}{f_B} = \frac{2.5}{3.0} = 0.8$$

As the Q factor is a frequency divided by a frequency, it has no unit.

TEST YOURSELF

18 What is the reactance of an inductor of value 1.0 mH at the following frequencies?
- **a)** 50 Hz
- **b)** 50 kHz

19 What is the reactance of a 1.0 μF capacitor at the following frequencies?
- **a)** 50 Hz
- **b)** 50 kHz

20 The 1.0 mH inductor and 1.0 μF capacitor are connected in parallel. Show that the resonant frequency of the circuit, f_0, is about 5.0 kHz.

21 The bandwidth of the combination in Question 20 is found to be 2.0 kHz. Calculate the Q factor for the circuit.

22 Sketch a graph to show the response curve for the circuit in Question 20.

The operational amplifier

An ideal operational amplifier

A **differential amplifier** is one that amplifies the difference in potential difference between two inputs, denoted – and + or inverting and non-inverting. If the –ve input is positive compared to the +ve input, the output swings negative. If the +ve input is positive compared to the –ve input, the output swings positive.

The operational amplifier (op amp) is a fundamental building block in electronic devices and circuits. It is a **differential amplifier** that has, ideally, infinite input resistance, infinitely large gain at all frequencies and very low output resistance. In practice none of these ideal characteristics are entirely possible but all general purpose operational amplifiers will have a very large input resistance, typically 10^{12} Ω, a very large gain, perhaps 100 000 or more, a low output resistance of a few hundred ohms, and operate uniformly over a wide range of frequencies, perhaps 3 MHz. These characteristics are essential for reliable and predictable operation.

Op amps are produced as integrated circuits and generally come in standard packages, as shown in Figure 18.39.

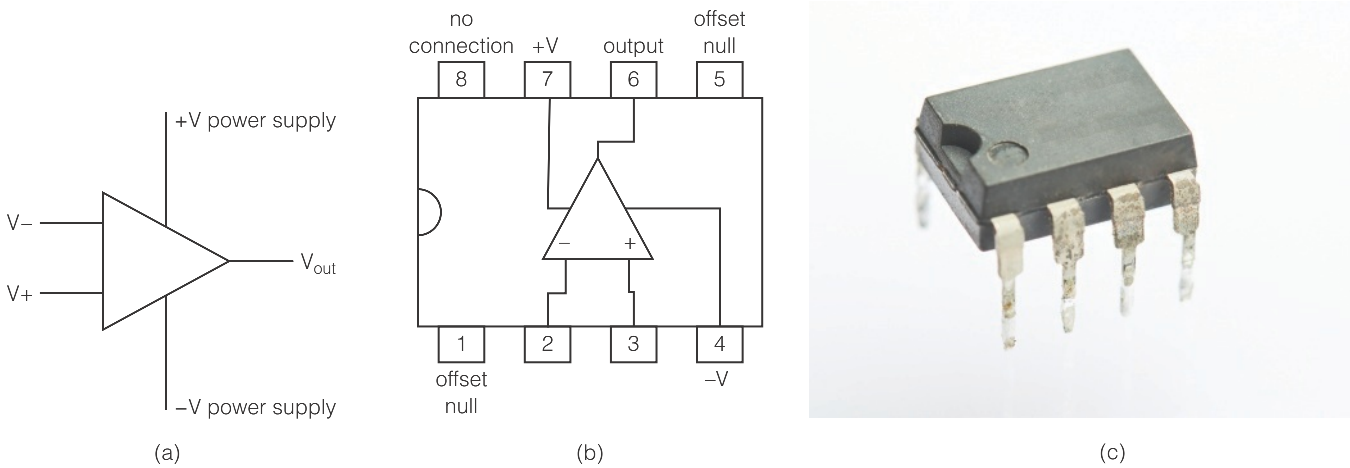


Figure 18.39 (a) The connections to an op amp showing the differential inputs and power supply connections. (b) Typically an op amp comes in an 8-pin DIP (dual inline package) and most follow the same pin configuration. (c) An integrated circuit op amp. The indented dot signifies pin 1.

Op amps generally require a split power supply, that is, +V, 0 V and −V. Typically this might be +12 V, 0 V and −12 V. This is because the output can swing positive or negative, depending on the state of the inputs.

The infinite input resistance means that no current is drawn from the source so adding the op amp to a circuit does not affect the output voltage from the previous stage.

The open-loop gain, A_{OL}, is the amount by which the op amp amplifies the difference between the two inputs. Although ideally infinite, the open-loop transfer characteristic of a real op amp is given by $V_{out} = A_{OL}(V_+ - V_-)$, where $(V_+ - V_-)$ is the difference in voltage between the non-inverting and the inverting inputs (Figure 18.40).

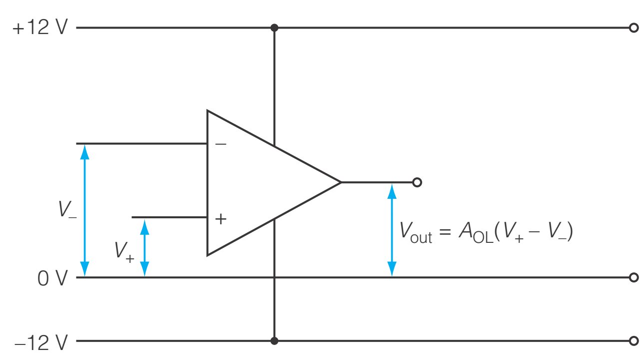


Figure 18.40 An op amp used in open-loop mode. The output relative to 0 V will be given by $V_{out} = A_{OL}\left(V_+ - V_-\right)$. Because of the split power supply, the output can be positive or negative with respect to 0 V.

Saturated means that the output is as far positive or negative as it can be and is no longer affected by changes to the input.

Because of the very high gain, the smallest difference between V_+ and V_- will result in an output that is maximum positive or maximum negative. The output is then said to be **saturated**. This maximum output will be close to the supply rail voltage, depending on the exact op amp

A **comparator** is a circuit that compares two inputs and outputs a signal that depends on the difference.

used. For the TL081 op amp shown in Figure 18.39, the output will be approximately 1 V away from the supply voltage, e.g. ± 11 V in the circuit in Figure 18.40.

An operational amplifier is rarely used in its open-loop mode but one circuit where it is used in this way is as a **comparator**. A comparator compares an input voltage with a fixed reference, for example from a Zener diode, and outputs a positive voltage if the output is below the reference and a low voltage if the output is above the reference.

ACTIVITY

The circuit in Figure 18.41 is a comparator circuit. The 100 kΩ potentiometer will provide a voltage between +12 V and −12 V to the inverting input. If this is adjusted so that it is below the 4.7 V reference voltage from the Zener diode, the output of the op amp will be high and the green LED will light. If the voltage at the inverting input is above the 4.7 V reference voltage, the output of the op amp will be low and the red LED will light.

Swapping the connections to the + and − inputs will reverse the effect.

Comparators are used as part of a number of different devices and circuits; voltage regulators and analogue-to-digital encoders are examples.

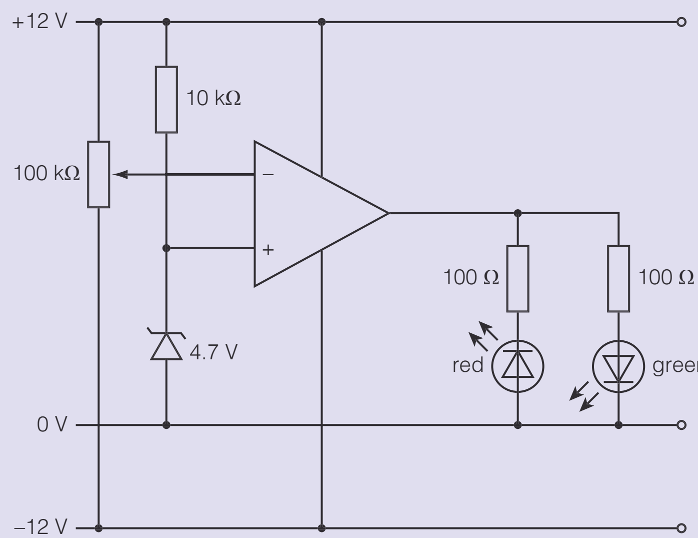


Figure 18.41 A circuit that shows how any op amp can be used as a comparator. The LEDs will light depending on the voltage at the inverting input.

TEST YOURSELF

23 In the circuit in Figure 18.42, TH1 is a thermistor, a temperature-dependent resistor, whose resistance falls with increasing temperature.

a) Explain the purpose of the 100 kΩ potentiometer.

b) Explain which LED will light for a low temperature.

24 An op amp has an open-loop gain of 30×10^6 and a power supply of ±12 V. What difference in input voltage between the + and the − inputs will make the output swing to maximum positive?

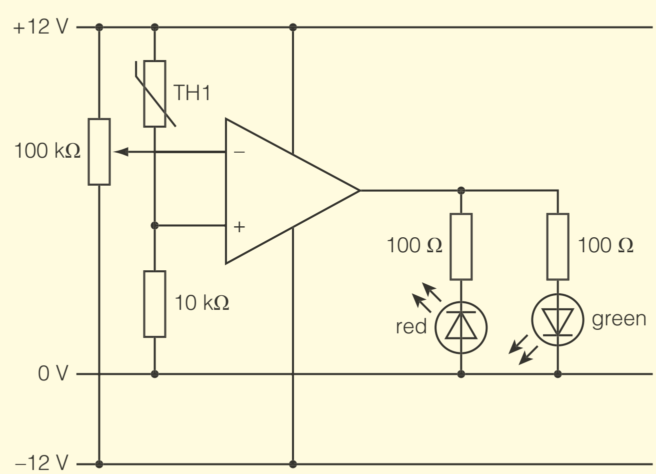


Figure 18.42

Inverting amplifier configuration

An operational amplifier is more commonly used in a configuration where the gain can be controlled so that the output does not become saturated. In order to achieve this, negative feedback is used.

Negative feedback occurs when a little bit of the output signal is fed back to the inverting input. This little bit fed back goes to reduce the signal that actually reaches the amplifier and so reduces the gain.

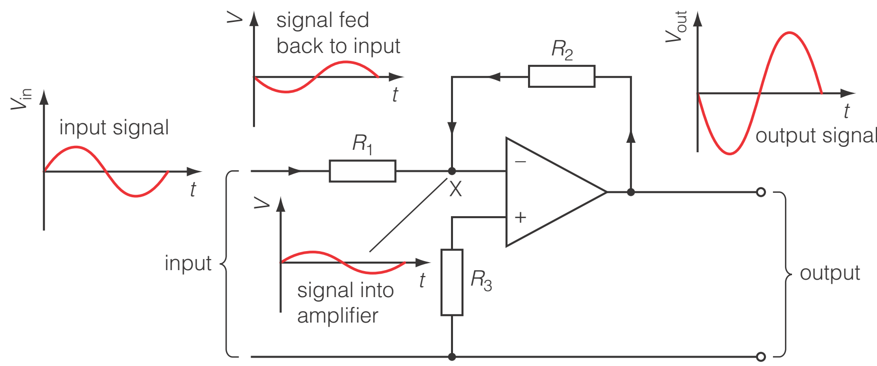


Figure 18.43 An op amp with negative feedback. It is common practice when drawing op amp circuits to omit the power supply for clarity. There must, however, be power supply connections.

In Figure 18.43, an input signal, V_{in}, is amplified to give an inverted copy. The input may be ac, as shown in Figure 18.43, or it may be dc.

A little of the output is fed back to the input via resistor R_f. This fed-back signal mixes with the input signal at point X so making the actual input to the amplifier smaller than it would be otherwise. Thus the gain is reduced.

Virtual earth is a point that, although not connected to earth (or 0 V), is always at very nearly the same potential as earth.

Point X is known as a **virtual earth**. This is because the non-inverting input is connected to earth or 0 V via resistor R_X and since the amplifier has infinite (or very high) open-loop gain, then for the output not to be saturated there must be a negligible potential difference between the inverting and the non-inverting input. Since the non-inverting input is at 0 V, point X must be virtually at 0 V; as far as the signal is concerned, the inverting input looks virtually the same as the non-inverting input which is at 0 V.

Given that X is a virtual earth, the gain of the circuit in Figure 18.43 can be derived:

The input current, I_{in}, is given by $I_{in} = \frac{V_{in}}{R_{in}}$.

The feedback current, I_f, is given by $I_f = \frac{-V_{out}}{R_f}$.

Since the input resistance of the op amp is infinite, no current flows into the input and therefore:

$$I_{in} = I_f$$

Thus $= \frac{V_{in}}{R_{in}} = \frac{-V_{out}}{R_f}$ and since the closed loop gain, A_{VCL}, is equal to $\frac{V_{out}}{V_{in}}$ the gain is given by:

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

The negative sign simply indicates that the output is inverted. In most cases this does not matter at all.

EXAMPLE

For an inverting amplifier, R_f = 560 kΩ and R_{in} = 1.0 kΩ. Calculate the gain.

Answer

The gain is $-\frac{560\ \text{k}\Omega}{1.0\ \text{k}\Omega} = -560$

The larger the value of R_f, the larger the gain.

TEST YOURSELF

25 Explain why point X in Figure 18.43 is called a virtual earth.

26 An inverting amplifier has R_f = 100 kΩ and R_{in} = 1.0 kΩ. Calculate the gain.

Non-inverting amplifier configuration

Sometimes it is necessary for the operational amplifier to be used in a non-inverting configuration.

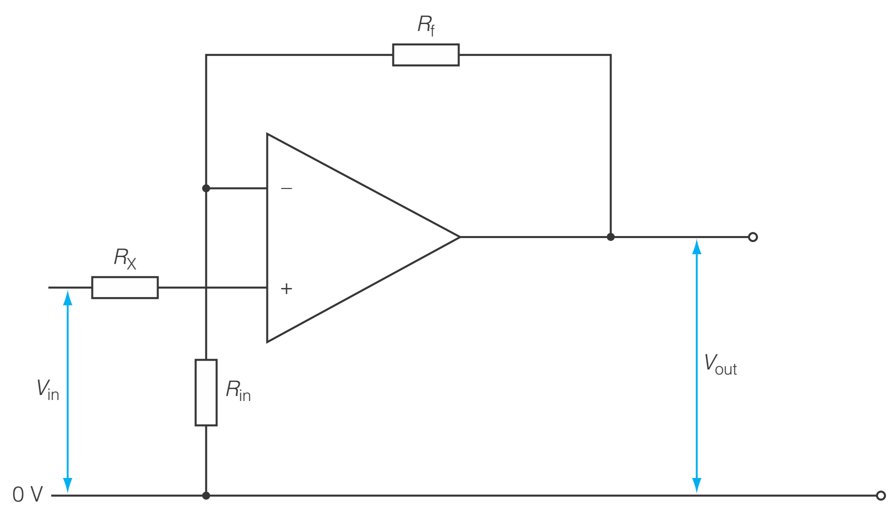


Figure 18.44 A non-inverting amplifier configuration. The input now goes to the non-inverting input.

A non-inverting amplifier circuit is shown in Figure 18.44. Once again, the gain is controlled by the negative feedback to the inverting input.

It can be shown that the closed loop gain, A_{VCL} in this case is given by:

$$A_{VCL} = \frac{V_{out}}{V_{in}} = 1 + \frac{R_f}{R_{in}}$$

R_X has no effect on the gain because of the effectively infinite input resistance of the op amp.

A particular use of this circuit is if R_{in} is removed completely (made = ∞) and R_f is made zero. R_X would also be made zero. In this case, the gain becomes 1 and the output exactly follows the input. It is known as a voltage follower.

This is used for applications where a signal that cannot supply a current, e.g. a photodiode sensor, is required to drive an output that does require a current, for example a meter or an LED.

TEST YOURSELF

27 In the circuit in Figure 18.45, R_f is 470 kΩ and R_{in} is 10 kΩ. Calculate the gain of the circuit.

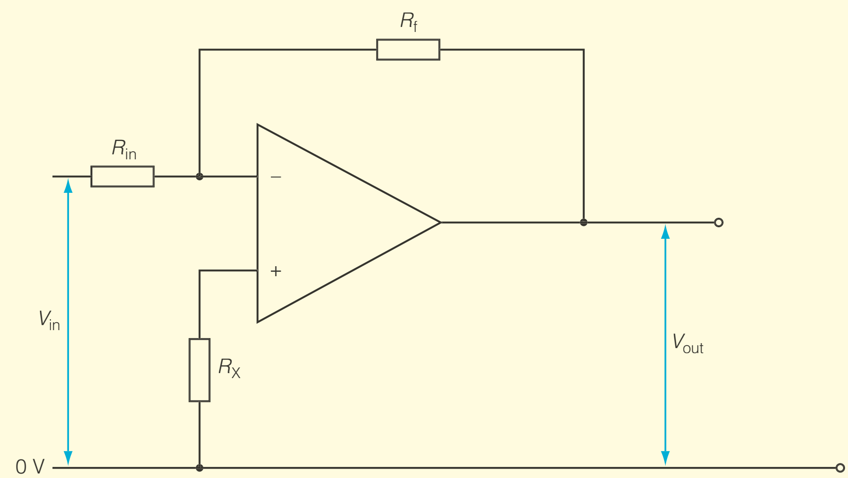


Figure 18.45

28 In the circuit in Figure 18.44, R_f is 39 kΩ and R_{in} is 1.0 kΩ. Calculate the gain of the circuit.

29 Draw circuits based on op-amps that have the following gains:

a) −1

b) +48

c) −82

d) +1

30 Explain why the value of the resistor R_X in Figure 18.44 has no effect on the gain.

Summing and difference amplifiers

A summing amplifier

The inverting amplifier can be used to add potential differences. The circuit of Figure 18.43 is used but with additional input resistors.

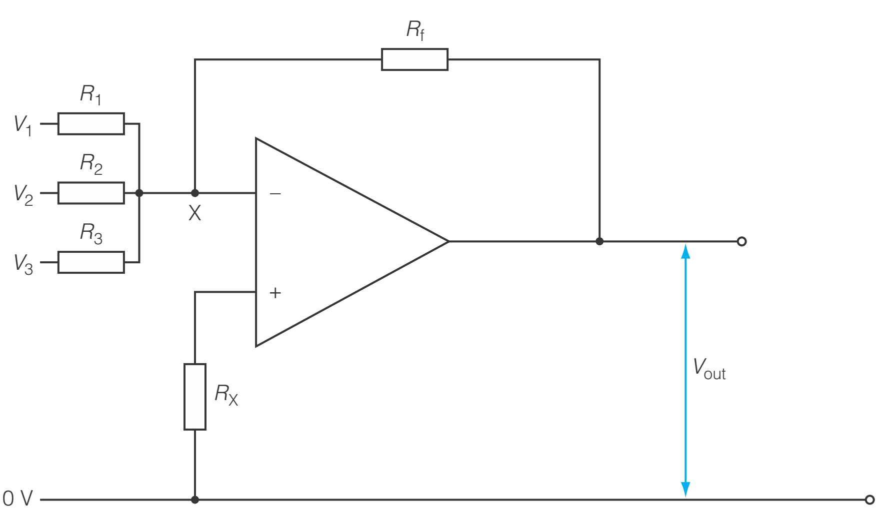


Figure 18.46 A summing amplifier.

In the circuit in Figure 18.46, because of the virtual earth at point X, the output, V_{out}, is given by:

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

Thus, if $R_1 = R_2 = R_3 = \ldots$ the output is simply given by $V_{out} = -\frac{R_f}{R_1}(V_1+V_2+V_3+\ldots)$, although different proportions of each input can be added by varying the input resistors.

A summing amplifier is used to add voltages, ac or dc, for example as the basis for a simple microphone mixing amplifier.

A difference amplifier

The difference amplifier is a special case of the differential amplifier. It uses both inverting and non-inverting inputs, usually with a gain of one, to produce an output equal to the difference between the inputs (Figure 18.47).

The output voltage is given by $V_{out} = (V_+ - V_-)\frac{R_f}{R_{in}}$ and if $R_f = R_{in}$ then the gain is 1 and $V_{out} = (V_+ - V_-)$.

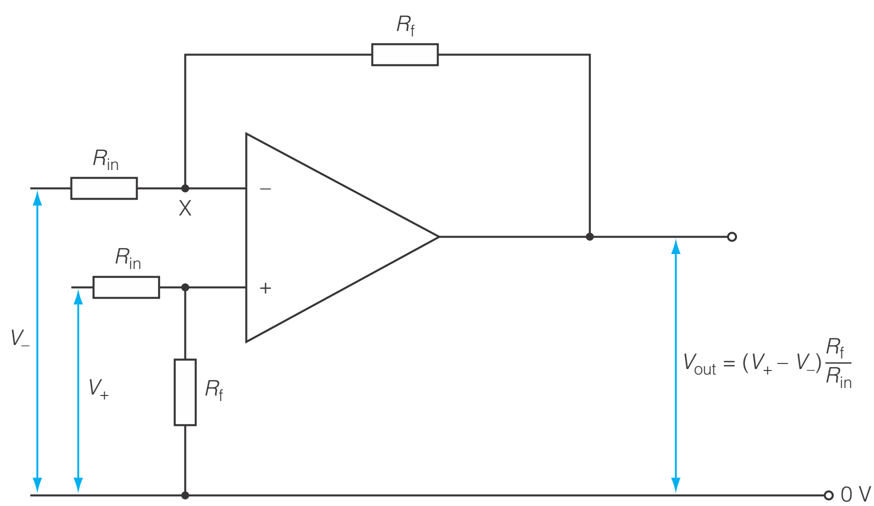


Figure 18.47 A difference amplifier is a special case of the differential amplifier where the gain is determined by the input and feedback resistors.

Difference amplifiers are used in noise reduction circuits where one input is connected to the wanted signal plus noise and the other input is connected to the same noise source, for example a transmission line, but without the wanted signal. The output from the difference amplifier will consist of only the wanted signal.

Another use is for amplifying the output from a strain gauge where the difference in pd across a pair of platinum resistors carrying a current depends on the strain applied.

TEST YOURSELF

31 In Figure 18.46, R_f = 10 kΩ and R_1, R_2 and R_3 are all 2.0 kΩ. V_1 = 0.1 V, V_2 = 0.05 V, V_3 = 0.08 V. Calculate V_{out}.

32 In Figure 18.48, the input voltage on each of A, B, C and D is 1 V. Calculate the output voltage, V_{out}.

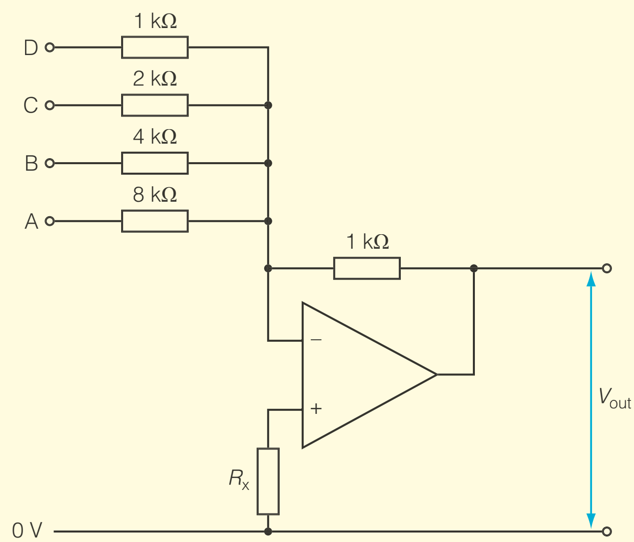


Figure 18.48 A simple digital-to-analogue converter with input D as the MSB.

33 In Figure 18.47, R_f = 10 kΩ and R_{in} = 1.0 kΩ.

a) Calculate the output voltage, V_{out}, if V_- = 1.0 V and V_+ = 1.1 V.

b) Calculate the output voltage, V_{out}, if V_- = 1.1 V and V_+ = 1.0 V.

Real operational amplifiers

An ideal operational amplifier has a constant gain over all frequencies, but this would make the op amp unstable and so real op amps have their gain limited by an internal capacitor. This ensures that the feedback can never be positive at all frequencies where the open-loop gain, A_{OL}, is greater than one. Positive feedback would be like the 'howling' that can occur in an audio system when the microphone picks up sound from the loudspeaker – it generally renders the system unusable unless it is controlled.

The bandwidth is the range of frequencies over which the gain is constant for a particular combination of resistors.

The op amp is designed so that gain × bandwidth = constant for any given device. For example, one popular and common op amp has gain × bandwidth = 3 MHz.

The graph in Figure 18.49 shows how this affects the actual gain of the amplifier with various values for desired gain. For example, for a gain of 100, the bandwidth, as shown by the red line, is about 3×10^4 Hz. This is consistent with:

$$\text{gain} \times \text{bandwidth} = 100 \times 3 \times 10^4 \text{ Hz} = 3 \text{ MHz}$$

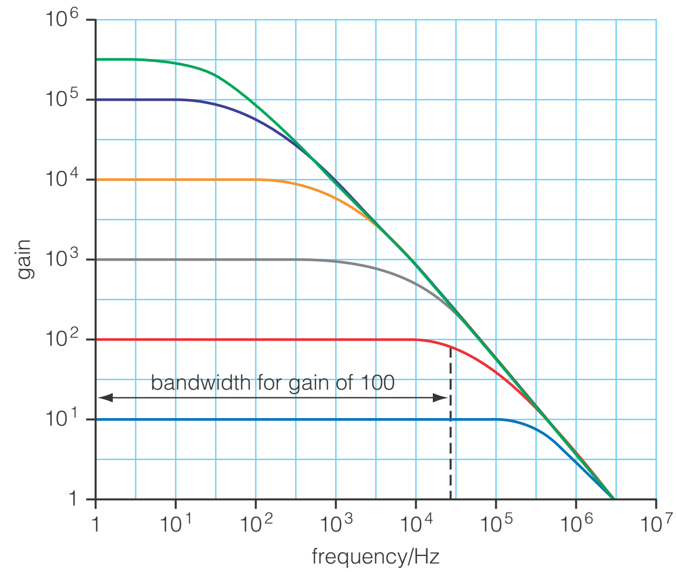


Figure 18.49 The frequency response of an op amp at various values of closed loop gain.

Different op amps will have different characteristics. This op amp would be fine for an audio application if the gain is set at less than 100 but not if the gain is set to 1000 since, at this gain, the gain is reduced at the higher range of audio frequencies.

For radio frequencies, which may be greater than 3 MHz, this particular op amp is unlikely to be suitable, but there will be other ones available that are suitable.

TEST YOURSELF

34 Explain why the open-loop gain of real op amps drops off at high frequencies.

35 An op amp has gain × bandwidth = 4 MHz.

a) Calculate the bandwidth of an amplifier constructed using this op amp with a gain of 10 000.

b) Comment on whether this would be suitable for use as an audio amplifier.

Digital signal processing

Combinational logic

Two switches in series in a circuit can be described as an AND circuit because both switches have to be closed in order for a current to flow.

Two switches in parallel in a circuit can be described as an OR circuit because either one or both switches have to be closed for a current to flow.

The function of these circuits can be described in truth tables (Tables 18.3 and 18.4).

Table 18.3 Truth table for an AND circuit.

Switch 1	Switch 2	Circuit on/off
open	open	off
open	closed	off
closed	open	off
closed	closed	on

Table 18.4 Truth table for an OR circuit.

Switch 1	Switch 2	Circuit on/off
open	open	off
open	closed	on
closed	open	on
closed	closed	on

Digital signal processing is the fundamental function in all digital processors and computers. In digital signals we are only concerned with two levels, high and low, generally referred to as 1 and 0.

A **logic gate** is an electronic circuit that performs a logical function on one or more digital inputs where the output is always exactly defined.

Digital circuits are based on **logic gates**, which perform logical functions on one or more digital signals to provide an output that depends on the inputs. These logical functions are described in terms of Boolean logic and truth tables are used to show the Boolean functions.

Logical functions and truth tables

NOT

The simplest logical function is to give the inverse of a digital value; an input 1 becomes an output 0 and an input 0 becomes an output 1.

If A is the original signal, the inverse is written as $\overline{\mathrm{A}}$, where the line above the A shows the inverse function. Thus we might write:

$$Q = \overline{\mathrm{A}}$$

This means that the value of Q is always the inverse of A or Q is NOT A. We can also show this in a truth table (Table 18.5).

Table 18.5 Truth table for a NOT gate.

A	Q
A	$\overline{\mathrm{A}}$
0	1
1	0

In drawing logic circuits and truth tables, the convention is to use A, B, C, etc. as inputs and Q or Q_0, Q_1, Q_2, etc. as outputs.

A circuit to perform the NOT function is quite simple but it is only shown as a circuit building block. These circuit building blocks are known as logic gates.

Figure 18.50 shows the circuit symbol for a NOT gate. The small circle at the output indicates inversion.

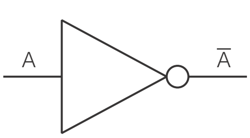


Figure 18.50 Circuit symbol for a NOT gate.

You will note that there are no power supply connections shown. This is always the case when drawing circuits using logic gates but it does not mean that a logic gate does not need a power supply – it does, but logic gates are manufactured in integrated circuits, perhaps with six NOT gates on a single chip, and the connections to the power supply, usually 5 V, are assumed. Including power supply connections would make the circuits unnecessarily complicated.

AND

The next simplest logic function is when the output of a circuit is 1 if both of two inputs are 1. This is analogous to having two switches in series in a circuit – both have to be closed for a current to flow.

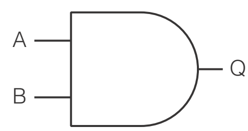
For two inputs A and B, we write:

$$Q = A \bullet B = A \text{ AND } B$$

Boolean algebra is a branch of mathematics that is able to simplify working out truth tables for a complicated logic system. It has specific rules that apply to logic functions.

where Q is the output and the dot between the A and B is a shorthand for the AND function. This is the operator for **Boolean algebra**.

The circuit symbol for the AND gate is shown in Figure 18.51, together with the truth table.



AND

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

Figure 18.51 Circuit symbol and truth table for an AND gate.

Note that, by convention, the inputs are shown in order of binary counting. There can be more than two inputs, but the same rule applies for as many inputs as are required.

OR

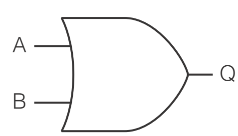
An OR function has the output 1 if either or both of the inputs are 1. This is analogous to two switches in parallel in a circuit.

For two inputs, A and B, we write:

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

where Q is the output and the + sign between the A and B is the shorthand for the OR function.

The circuit symbol and truth table for the OR gate are shown in Figure 18.52.



OR

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

Figure 18.52 Circuit symbol and truth table for an OR gate.

NAND

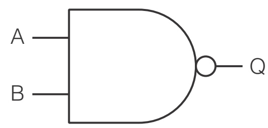
Perhaps surprisingly, one of the most common logic gates used in constructing circuits is the NAND gate. One reason for this is that all other logic functions can be created from a combination of NAND gates. NAND is short for NOT AND.

For two inputs, A and B, we write:

$$Q = \overline{A \bullet B} = A \text{ NAND } B$$

where Q is the output.

The circuit symbol and truth table for the NAND gate are shown in Figure 18.53. The small circle on the output indicates an inversion.



NAND

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

Figure 18.53 Circuit symbol and truth table for a NAND gate.

NOR

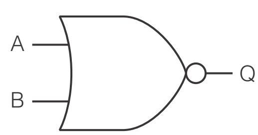
Another common logic gate used in constructing circuits is the NOR. NOR is short for NOT OR.

For two inputs, A and B, we write:

$$Q = \overline{(A + B)} = A \text{ NOR } B$$

where Q is the output. The brackets are important, in the same way as when using brackets in arithmetic or algebra, as anything inside the bracket is performed first. In this case they show that the result of A OR B is then inverted.

The circuit symbol for the NOR gate is shown in Figure 18.54, together with its truth table. Again, the small circle at the output indicates an inversion.



NOR

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

Figure 18.54 Circuit symbol and truth table for a NOR gate.

EOR – exclusive OR

The final logic function we will use is the exclusive OR function. In this the output is 1 if either input A is high or input B is high, but not both. An exclusive OR gate is used as part of a circuit to add binary numbers.

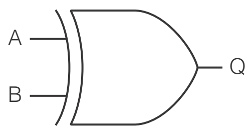
There is no universally agreed Boolean symbol for an EOR gate, although $\oplus$ is often used.

For two inputs, A and B, we write:

$$Q = ((A \bullet \overline{B}) + (\overline{A} \bullet B)) = A \text{ EOR } B$$

where Q is the output. We shall see later why this is the case.

The circuit symbol for the EOR gate and its truth table are shown in Figure 18.55.



EOR

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

Figure 18.55 Circuit symbol and truth table for an EOR gate.

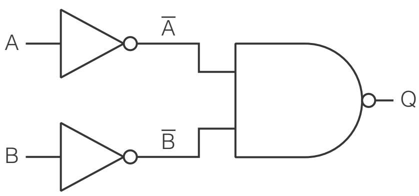


Figure 18.56

Boolean logic and combinations of gates

The compact way of describing logic functions is using Boolean logic but truth tables are also very useful.

Consider the combination of logic gates in the circuit in Figure 18.56.

The circuit of Figure 18.56 can be represented by the truth table given in Table 18.6.

Table 18.6 Truth table for the circuit in Figure 18.56.

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

If you look at the inputs A and B and output Q, you will see that this is actually the truth table for an OR gate.

Another way of analysing this is to use De Morgan's theorem. This states that to invert an expression, replace AND with OR and each variable with its inverse, that is:

$$\overline{A \cdot B} = \overline{A} + \overline{B}$$

$$\overline{A + B} = \overline{A} \cdot \overline{B}$$

In the truth table shown in Table 18.6, (NOT A) NAND (NOT B) = Q, or in Boolean algebra:

$$Q = \overline{\overline{A} \cdot \overline{B}}$$

Using De Morgan's theorem on this expression, we can find the inverse. We know that $\overline{A \cdot B} = \overline{A} + \overline{B}$, so:

$$\overline{\overline{A} \cdot \overline{B}} = A + B$$

This gives Q = A + B, which is OR.

TEST YOURSELF

36 a) Draw a truth table for the circuit in Figure 18.57 and write down the Boolean expression for Q.

b) Use De Morgan's theorem to determine Q.

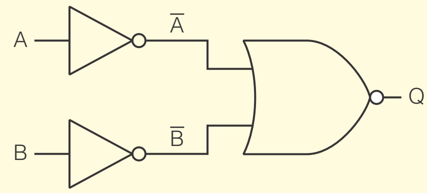


Figure 18.57

37 Use $Q = ((A \bullet \overline{B}) + (\overline{A} \bullet B))$ to devise a circuit of logic gates using only NOT and NAND to produce an EOR gate. You may find it helps to copy and complete the full truth table first.

Table 18.7 Truth table for Question 37

A	B	$\overline{A}$	$\overline{B}$	$(A \cdot \overline{B})$	$(\overline{A} \cdot B)$	Q
0	0					
0	1					
1	0					
1	1					

You should find that it is possible to simplify your circuit to use just two NOT gates and three NAND gates since $\overline{\overline{A}} = A$.

38 Draw a truth table for the arrangement of gates in Figure 18.58 and use it to determine which logic gate the circuit produces.

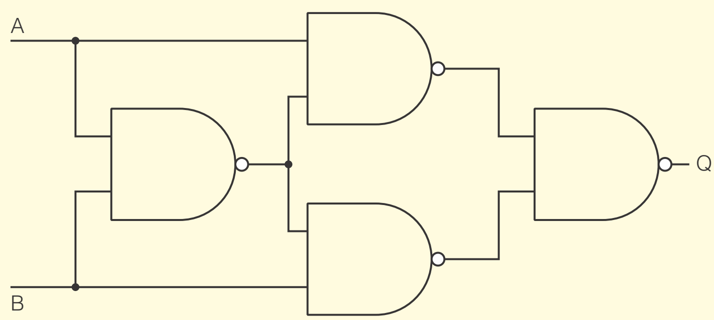


Figure 18.58

Sequential logic

Astable circuits

Many digital circuits involve a series of processes, for example counting, sampling, switching and timing. All these processes need to be kept going by means of a steady stream of pulses, usually referred to as a clock (Figure 18.59).

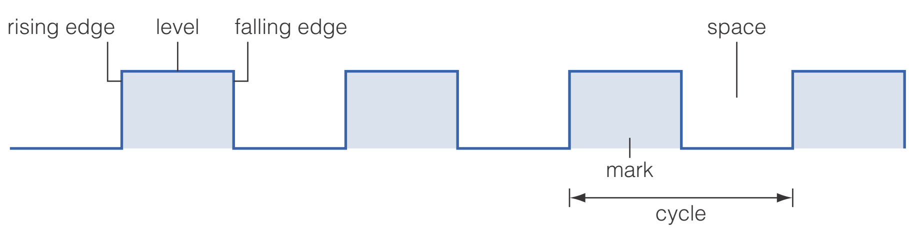


Figure 18.59 Clock pulses (CK).

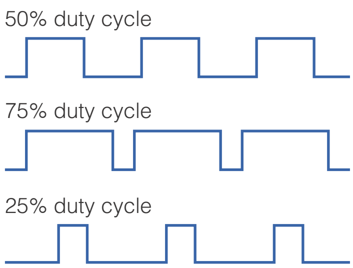


Figure 18.60 A 50% duty cycle signal has a mark-to-space ratio of 1 : 1. A 75% duty cycle has a mark-to-space ratio of 3 : 1.

Clock pulses are provided by a circuit called an astable.

The pulse rate or frequency is the number of cycles (a mark and a space) per second. The frequency is given by $f = \frac{1}{T}$ where T is the periodic time for 1 cycle.

The clock shown in Figure 18.59 has a mark-to-space ratio of 1:1 or a duty cycle of 50% but this does not need to be the case. A mark-to-space ratio of 2:1 would have the mark twice as long as the space; the duty cycle for such a pulse would be 67%. Different duty cycles are shown in Figure 18.60.

Sometimes the duty cycle or the mark-to-space ratio is changed according to a separate, often an analogue, input. When this is done, it is known as pulse width modulation (PWM). This is often used to control the speed of a motor or the brightness of a lamp or LED. A duty cycle of 100% would mean a continuous output of 1 or full power to the lamp or motor. Reducing the duty cycle for a motor can reduce the speed while maintaining the turning effect or torque.

There are many different astable circuits, including some commonly available dedicated integrated circuits such as the 555 timing chip. An example of a simple astable circuit is shown in Figure 18.61.

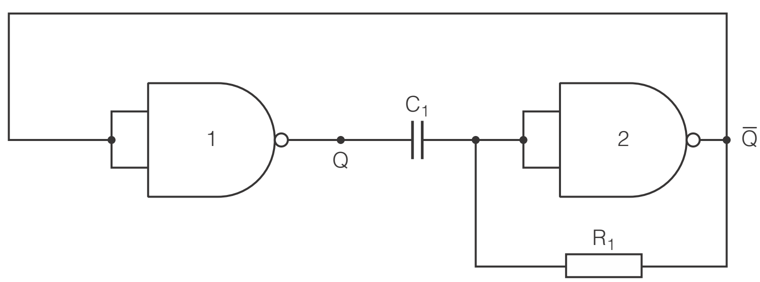


Figure 18.61 A NAND gate with the two inputs connected together acts like a NOT. NOT gates can equally be used.

Suppose when switched on, the output $\overline{Q}$ is high (it doesn't matter what starting condition you assume). This means that the input to NOT gate 1 is high so Q is low.

The capacitor, C_1, will charge towards V_0 and after a time $0.9R_1C_1$ will reach just above 50% of V_0. At about this point, the input to NOT gate 2 is high enough to switch it and make the output $\overline{Q}$ low. This, in turn, makes the input to NOT gate 1 low, which makes Q high. The capacitor now starts to discharge through the resistor R_1 and after a time $0.9R_1C_1$ will reach just below 50% of V_0. This is low enough for NOT gate 2 to switch, making the output $\overline{Q}$ high, and the cycle repeats itself.

The frequency of this astable using CMOS NAND gates, is given by $f = \frac{1}{1.8R_1C_1}$. This simple circuit will give a duty cycle of close to 50% for certain frequencies. A more sophisticated circuit or dedicated integrated circuit is needed for a 50% duty cycle at all frequencies.

EXAMPLE

In the circuit of Figure 18.61, the value of C_1 is 0.1 μF and R_1 = 4.7 MΩ. Calculate the clock frequency produced.

Answer

$$f = \frac{1}{1.8R_1C_1} = \frac{1}{1.8 \times 4.7 \times 10^6 \times 0.1 \times 10^{-6}} = 1.2\ \text{Hz}$$

Other astable circuits, such as ones based on a 555 timer integrated circuit have a slightly different formula for frequency but all depend on a *RC* timing circuit similar to the one used above. Unlike the simple circuit of Figure 18.61, a 555 allows the duty cycle to be adjusted from 0 to 100%.

If a more accurate clock pulse is required, a timing crystal is incorporated in the circuit.

TEST YOURSELF

39 The circuit in Figure 18.62 is for an astable constructed using the popular LM555 timer chip.

R_1 = 4.7 kΩ, R_2 = 47 kΩ and C = 0.01 μF

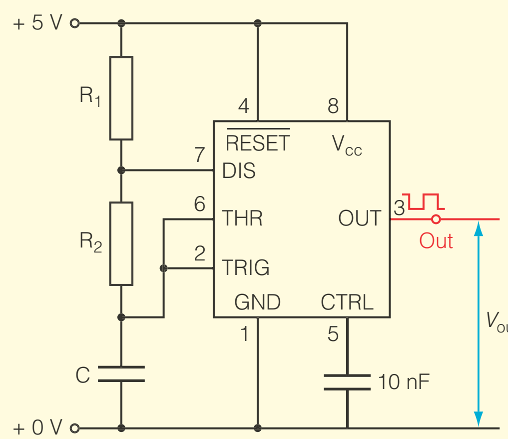


Figure 18.62

The output clock frequency is given by $f = \frac{1.44}{(R_1 + 2R_2) \times C}$ and the duty cycle is $\frac{R_2}{(R_1 + 2R_2)} \times 100\%$.

a) Calculate the frequency of the output clock pulse.
b) Calculate the duty cycle.
c) Suggest how to make the duty cycle as close as possible to:
  i) 50%
  ii) 33%.

Counters

The use of a clock pulse from an astable circuit allows logic functions to be carried out one after another – sequentially. A common application of sequential logic is in counting.

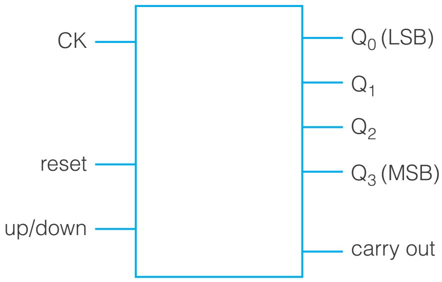


Figure 18.63 A circuit block for a binary counter.

Binary counter

In many digital applications it is required that a series of pulses (0 or 1) are output as a binary count. For this a binary counter integrated circuit is used. The counter is constructed on the chip with a combination of logic gates but we shall use the counter simply as another circuit building block.

A 4-bit binary counter will output 4 bits that will count in binary sequence from 0000 to 1111. It can be reset at any point and most integrated binary counters also include a control that allows them to count up or count down.

There are several integrated circuit binary counters available and they may have other inputs or outputs but Figure 18.63 shows the basic functions.

Q_0 is the least significant bit and Q_3 the most significant bit. Provided the up/down input is kept high, on the rising edge of the first clock pulse, Q_0 will go high. On the next rising edge of the clock pulse, Q_0 will go low and Q_1 will go high. The bits, Q_0 to Q_3 will go high or low on each subsequent clock pulse until the 17th clock pulse, when they all go low again and counting starts again.

The function of a counter is most easily shown in a timing diagram (Figure 18.64).

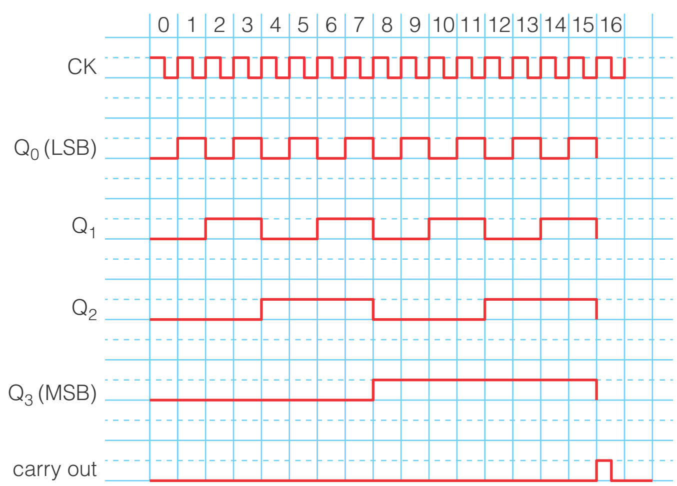


Figure 18.64 The timing diagram for a binary counter.

The carry out bit can be used to connect to another counter that will count up by one every time the counter completes a count up to 1111. The next counter will count 1 on the rising edge of the carry out bit.

Although the description has referred to changes happening on the rising edge of the clock pulse, it is equally possible to have the changes occur on the falling edge, either by using a different design of integrated circuit or by inverting the clock pulse using a NOT gate. These counters, however, will always operate on either the rising edge or the falling edge.

Binary coded decimal (BCD) counter

The BCD counter is very similar to the binary counter but instead of counting to 1111, it counts to 1001 and then resets to 0000. Although dedicated integrated BCD counters are manufactured, one can easily be made from a binary counter. If you look at Figure 18.64, you will see that on clock pulse 10, Q_1 and Q_3 are high for the first time and so an AND gate connected to Q_1 and Q_3 with the output to the reset input will reset the counter on the 10th clock pulse.

Often the output from a BCD counter is input to a decoder circuit for a 7-segment display so that a decimal figure is displayed. If a binary counter and gate is used, the carry out bit is no longer of any use, but the output from the AND that is connected to the reset can also be used as the clock pulse for another BCD counter in order to count 10s.

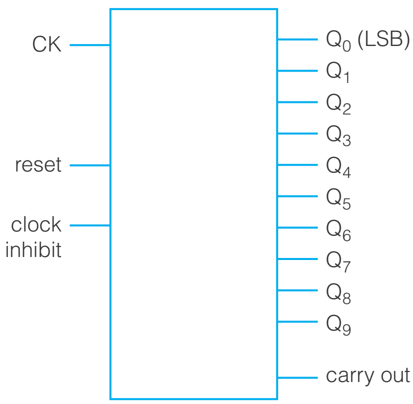


Figure 18.65 A block diagram for a 5-stage Johnson counter.

Johnson counter

A Johnson counter consists of a number of circuit elements, called shift registers, which are connected one after the other. Any number can be used but we are going to look at the case of a 5-stage Johnson counter (Figure 18.65).

The counter will have a clock input and a number of outputs, in this case ten, Q_0 to Q_9. When the counter is reset, all the outputs Q_0 to Q_9 are low.

On the rising edge of the first clock pulse after reset, Q_0 goes high. On the rising edge of the next clock pulse, Q_0 goes low and Q_1 goes high, on the rising edge of the next clock pulse, Q_1 goes low and Q_2 goes high, etc. On the 11th clock pulse, Q_9 goes low and Q_0 goes high as the cycle repeats.

The term shift register can be understood in that the 1 is shifted along the sequence of registers on each clock pulse.

As before, the best way of showing the operation of the counter is by means of a timing diagram (Figure 18.66).

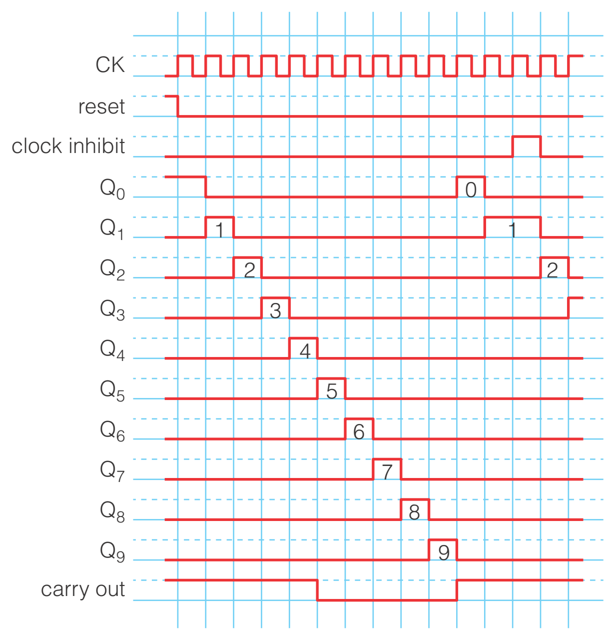


Figure 18.66 The timing diagram for a 5-stage Johnson counter. The carry out pin provides a single pulse every time the counter completes a cycle. The rising edge occurs at the end of the tenth clock pulse. In this particular design, a high on the clock inhibit pin stops the counter advancing while the pin is kept high.

Modulo-n counter

By connecting any of the outputs to the reset pin, the counter can be made to reset after a given number of clock pulses and can therefore be used as a modulo-*n* counter or frequency divider.

For example, connecting Q_5 to the reset pin will result in one output pulse from any of Q_0 to Q_4 for every five input clock pulses. This would be a modulo-5 counter.

It is not necessary to use a Johnson counter for this as a binary or BCD counter can be used in the same way, but the Johnson counter is easy to use and it is easy to change *n*.

More than one Johnson counter can be used one after another, for example dividing a clock pulse by 60 would be done by using a modulo-10 counter followed by a modulo-6 counter.

This can be used in applications such as clocks where a high-frequency clock pulse needs to be divided down to produce an output in seconds, minutes and hours. The use of a high-frequency clock reduces timing errors.

An astable for a clock might run at 2.097152 MHz. This can be divided down to a 1 Hz clock using a 21-bit counter. If the clock runs at ±5 Hz, this is an error of 2.4×10^{-4} %, which is probably adequate for most clocks.

TEST YOURSELF

40 Use Figure 18.64 to show that the output from the binary counter on clock pulse 10 is 1010.

41 Outline the difference between a binary counter and a BCD counter.

42 Figure 18.67 shows an oscilloscope trace for a modulo-*n* counter.

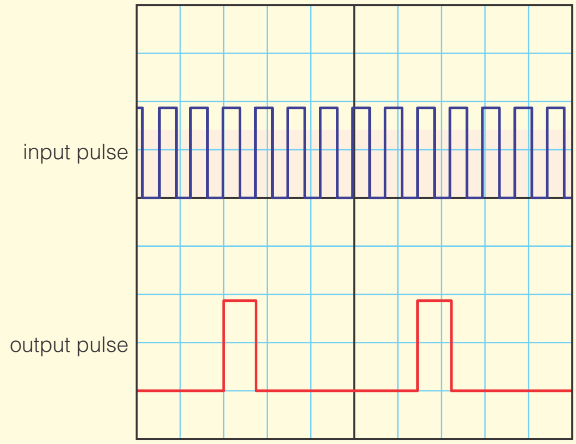


Figure 18.67

a) Use the trace to determine *n*.

b) Suggest which of the outputs shown in Figure 18.65 would be connected to the reset pin in order to make *n* the value you have determined in part (a).

43 Suggest why a high-frequency clock is used for accurate timing applications.

Data communication systems

Communication systems building blocks

All communication systems have common elements; these elements, which are the basic building blocks of any system are described in Table 18.8.

Table 18.8

Input transducer	This takes the signal from the environment, for example, a sound is picked up by a microphone, which converts it to an electrical signal.
Encoder	This takes the signal or message and converts it into a code that is suitable for transmission. For example an analogue signal may be sampled and converted into a digital signal for transmission.
Modulator	The coded signal has to be placed onto the carrier. This is done by the modulator. For example a coded message might be used to switch on and off a light beam; in this case, the carrier is the light wave and the modulator is the switch. Carriers can generally be amplitude modulated (AM) or frequency modulated (FM).
Amplifier	Often the modulated signal will need to be amplified before it is transmitted. This is because all transmission paths attenuate (reduce in strength) the signal and also introduce some noise. It is better to amplify the signal before it becomes noisy than to have to amplify it more when it is received, which will amplify both the desired signal and the noise.
Transmitter	The modulated carrier has to be transmitted from one place to another. For example a modulated radio wave is transmitted from an aerial. The power of the transmitter determines the distance from the transmitter that the carrier can be detected.
Transmission path	This is the route by which the modulated signal gets from the transmitter to the receiver. For example it could be radio waves, microwaves, light or even ultrasound. Sometimes, for example in fibre optic transmission, the signal may need to be amplified or boosted a few times over the entire length of the fibre optic cable.
Receiver	The transmitted signal will be detected by the receiver. A radio will receive radio waves via an aerial.
Amplifier	The received signal will usually need to be amplified because it will have been attenuated in the transmission path.
Demodulator	The message or signal has to be separated from the carrier so that the original message can be read. For example, a light signal down a fibre optic will be amplitude modulated according to the 1s and 0s of the digital signal. The 1s and 0s need to be recovered from the light signal using a photodiode.
Decoder	The signal may need to be decoded, for example a digital signal from a sampled analogue signal will need to be reconstructed as an analogue signal.
Output transducer	This is a device that changes the signal or message from one type to another. For example, a loudspeaker changes an electrical signal into a sound.

A typical communication system will be made up of all or most of the building blocks listed above. It is easier to see how these all relate to each other by using a system diagram (Figure 18.68).

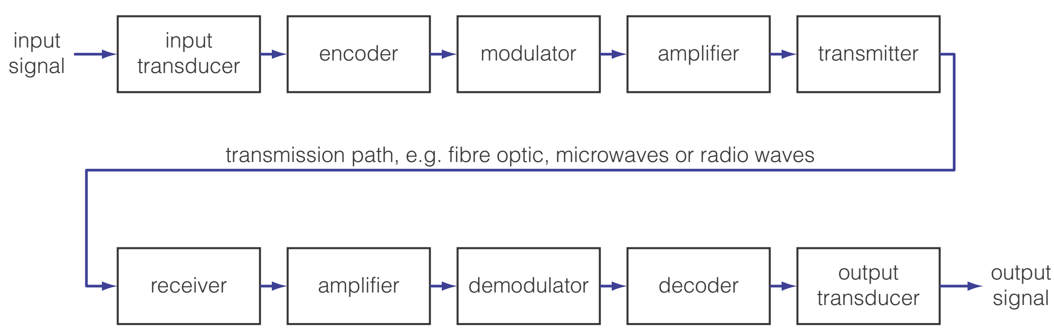


Figure 18.68 A typical communications system block diagram. Not every block will be required in every system.

TEST YOURSELF

44 Outline the purpose of each of the following in a transmission system:

a) carrier
b) modulator
c) receiver
d) amplifier.

Figure 18.69 A filter used for a copper wire telephone line to separate the broadband signal from the telephone signal.

Transmission media

The transmission path can be formed from a number of different transmission media. Telephone lines from most houses are copper wires and the voice signal to and from the telephone exchange is analogue. The digital signal used for broadband internet connection is achieved by using an amplitude modulated high-frequency carrier wave which is separated from the voice signal by means of filters at either end (Figure 18.69). Upstream and downstream signals are transmitted on different carrier frequencies.

Fibre optic cables

Fibre optic cables are used extensively to transmit digital signals. This might be between a domestic TV or tuner and an amplifier, between telephone exchanges, around the country for internet, TV and telephone or for communications between countries. Fibre optic cable is also used around the sites of schools, hospitals and other organisations to carry telephone signals and fast computer links.

In a fibre optic cable, light, usually infrared, is passed along a very thin glass fibre by total internal reflection.

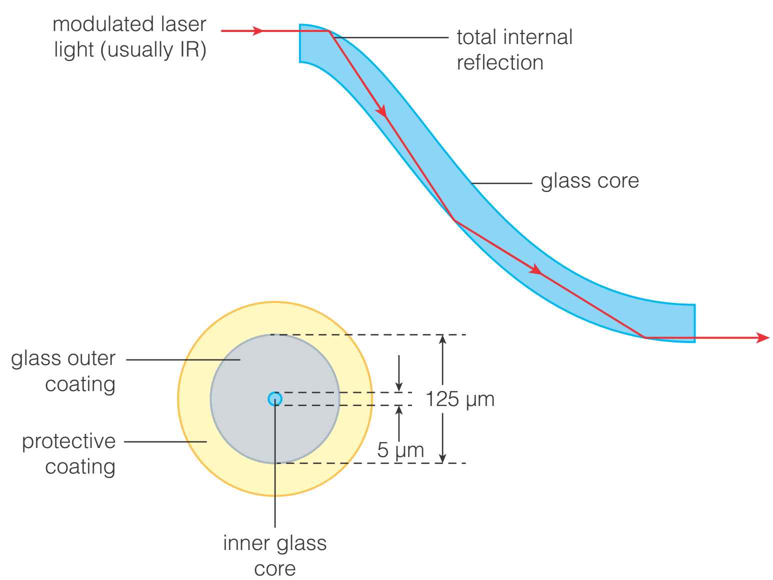


Figure 18.70 Fibre optic cable uses total internal reflection to guide light along a glass core.

Figure 18.70 shows part of an optical fibre with a ray of light passing along the fibre. In practice, some of the light will escape from a fibre like this and so the core is surrounded by more pure glass called cladding. This cladding ensures that almost all the light is reflected back into the fibre.

Optical fibres can be as long as 200 km although there is a limit on the distance between boosters. This is for two main reasons; one is that despite the careful manufacture of the glass fibre, there is still some attenuation of the signal because of absorption of the light and loss. The second is that the light can take slightly different paths down the fibre, which results in what started as a sharp digital pulse becoming broader, a process called pulse broadening. This will eventually cause the pulses to overlap and become indistinguishable.

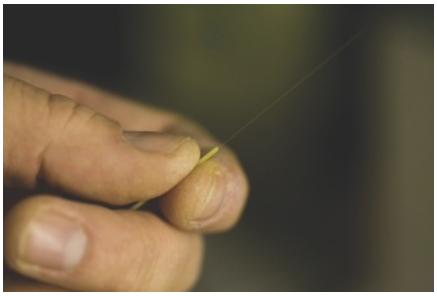


Figure 18.71 A fibre optic cable that can carry many thousands of telephone calls is much smaller than the equivalent copper cables.

A single optical fibre is used to transmit telephone, TV and computer data. As many as 10 000 telephone conversations can be transmitted down a single fibre at the same time using a system of multiplexing, which will be covered later. Figure 18.71 shows a typical fibre optic cable.

Advantages of fibre optic cables

- They use a high-frequency carrier wave; light, and the high frequency of light means that they can transmit data very much faster than copper wires, which will use a much lower carrier frequency.
- They do not pick up electrical 'noise' in the same way as copper cables, which means that the signal coming out of a fibre can be amplified without amplifying the noise.
- Glass is also much cheaper and has a lower density than copper and a single fibre can transmit as much data as many copper wires.
- They are more secure. It is possible to intercept the signal in a copper wire without damaging or cutting the wire simply by using electromagnetic induction. As far as a telephone line is concerned, this is very straightforward and simply requires a coil of wire and an amplifier. It is not at all straightforward to intercept the signal in a fibre optic cable directly.

Disadvantages of fibre optic cables

- If damaged, they cannot be re-joined simply as can copper wires.
- Special units are required at each end to process the optical signal.
- There is a limit on how sharply cables can be curved.

Radio waves and microwaves

Radio waves are part of the electromagnetic spectrum. Like all waves, radio waves can be reflected, refracted, diffracted and show interference effects. All these wave effects are particularly important for the reception of radio signals. The speed of all radio waves in air is the same: 3×10^8 m s^{-1}. Since for all waves, $v = f\lambda$, a long wavelength means a low frequency and a short wavelength means a high frequency.

How radio waves travel

Radio waves travel in straight lines. The curvature of the Earth means that a radio wave travelling a large distance in a straight line goes off into space, but it soon became clear from the experiments of Marconi and others that long distance radio communication is possible. This is due to the ionosphere. The ionosphere is a layer of reflecting ions that starts about 50 km above the surface of the Earth. This layer bounces radio waves back down to the surface of the Earth. Radio waves are also reflected from the surface of the Earth and under the right conditions can propagate all around the world by successive reflections (Figure 18.72).

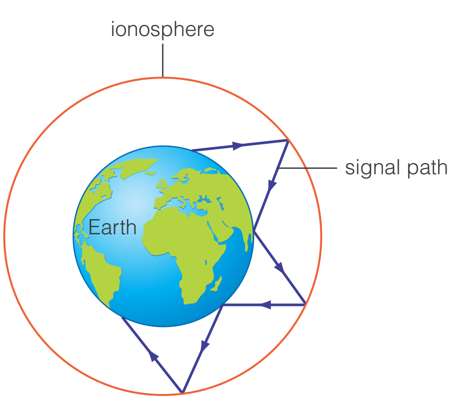


Figure 18.72 Radio waves can propagate around the Earth because of the ionosphere.

Although reflection from the ionosphere plays an important part in the propagation of radio waves, it is not the only way that radio waves can travel. The method of propagation depends on the frequency of the carrier wave. Figure 18.73 shows the paths taken by waves of different frequency ranges.

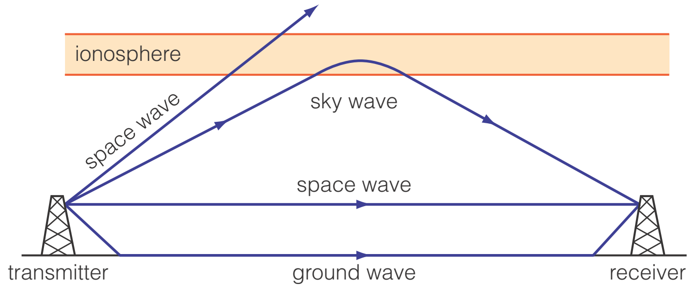


Figure 18.73 The propagation of radio waves around the Earth: ground waves, sky waves and space waves.

- Ground waves – waves up to about 2 MHz (wavelength about 150 m) follow the surface of the Earth due to diffraction. These are known as ground waves and their range is of the order of 1000 km.
- Sky waves – waves between 3 MHz and 30 MHz (wavelength 100 m to 10 m) travel around the Earth by reflecting off the surface of the Earth and the ionosphere. Sky waves are used for worldwide radio communications.
- Space waves – for frequencies above 30 MHz, the waves pass straight through the ionosphere. These are used for short range communications and 'line of sight' communication, e.g. microwaves. Space waves are ideal for communicating with aircraft, satellites and spacecraft, as well as for local radio and mobile phones.

Features on the ground can absorb radio waves, reflect or diffract them. The short radio waves associated with FM broadcasts are in the range 88 MHz to 108 MHz so are space waves and obey the 'line of sight' rule. In built-up areas, reflections off buildings and hills can allow reception without a direct 'line of sight' to the transmitter. However, for television reception, these reflections can be a nuisance and can result in additional noise or a weak signal, which can result in 'blocking' or a temporary picture freeze visible on the screen.

DAB radio is broadcast in the UK on frequencies between 218 MHz and 229 MHz and each channel can carry several radio streams by multiplexing. Multiplexing is covered later.

Long-wave radio transmissions, e.g. Radio 4, 198 kHz (wavelength 1500 m), diffract sufficiently to follow the curvature of the Earth. In the shadow of a hill or a large building, long-wave reception is usually much better than short-wave reception because the longer waves are diffracted more by the obstacles (Figure 18.74). This helps to explain why they have a greater area of coverage than FM broadcasts. Only one long-wave transmitter is required for the whole of England; it is located at Droitwich. Two lower power transmitters cover Scotland.

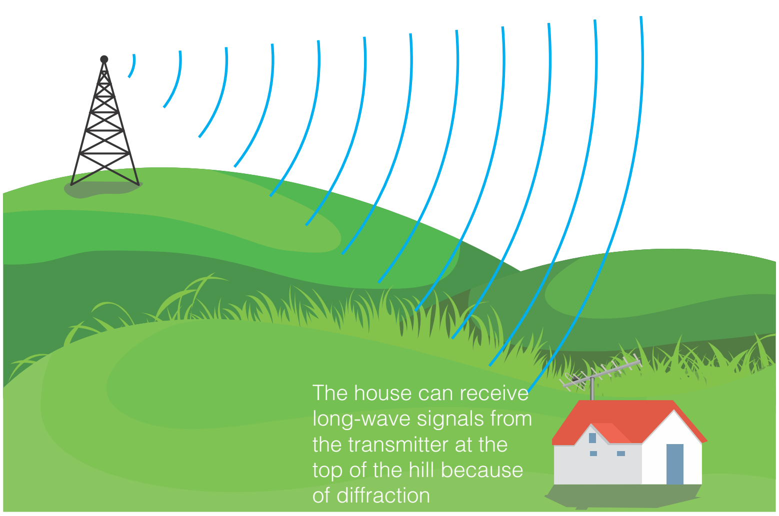


Figure 18.74 Long waves will diffract and so are less affected by obstacles.

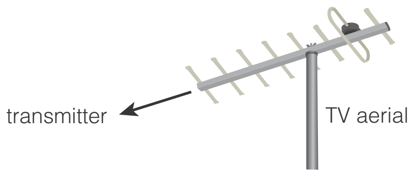


Figure 18.75 A TV aerial is designed to be highly directional and to only pick up signals with one direction of polarisation.

Interference effects are a constant problem for broadcasters; if two radio or television transmitters, transmitting on the same frequency, overlap their coverage area, some receivers could be in positions of destructive interference and receive no signal at all. In order to reduce this problem, national FM radio transmitters have to broadcast their signals with different carrier frequencies in different areas, which is why a car radio, set to, for example, BBC Radio 1, has to be re-tuned as the car travels to different parts of the country (most now do this automatically). The same is true of television broadcasts but the reception problems are helped by having aerials that are highly directional and will only receive waves with one direction of polarisation. The aerial is pointed directly at the transmitter (Figure 18.75).

Satellites

The early communications satellites worked in the passive mode. This means that they simply reflected signals back down to Earth. The signals could be received over a very wide area but were weak, needing a large receiving dish. The transmitters had to be very powerful.

All satellites are now active devices. The signals are beamed up to the satellite on a high-frequency carrier. The received signal is then processed within the satellite and returned to Earth in a narrow beam and at a lower frequency. It is necessary to have different frequencies for the up-link and the down-link in order to avoid de-sensing. De-sensing is the effect of a strong signal from a transmitter interfering with the detection of a weak signal by a receiver when operating at the same or similar frequencies at the same time.

There are four main frequency bands used by satellites:

- C band: up-link 5.9–6.4 GHz; down-link 3.7–4.2 GHz.
- X band: up-link 7.9–8.4 GHz, down-link 7.3–7.8 GHz.
- Ku band: up-link 14.0 GHz; down-link 10.9–12.8 GHz.
- Ka band: up-link 26.5–40.0 GHz; down-link 18–20 GHz.

Telstar, launched in 1962, was designed to relay signals between the United States, Europe and Japan. It was put into a low elliptical orbit and could only relay signals during short periods of each orbit. This is clearly a disadvantage and in 1967, INTELSAT 2B, the first geostationary communications satellite, was launched. Geostationary satellites have the advantage that the transmitter and receiving dish aerials can be permanently pointed at one place and the satellite can relay signals continuously.

A large number of other satellites have been launched. Some are geostationary and provide communications and satellite TV, others have remote sensing equipment and monitor atmospheric and surface conditions on Earth, others are equipped to look out into space and provide information on the nature of the Universe that would be difficult or impossible to gather from the Earth's surface.

Geostationary satellites

A geostationary satellite is one that stays above the same place on the surface of the Earth. It must be above the equator (Figure 18.76).

Figure 18.76 Geostationary or geosynchronous satellites are positioned above the equator.

Figure 18.77 Polar orbiting satellites are close to the Earth.

Polar orbits

A polar orbit allows a satellite to be closer to the surface of the Earth. As the satellite orbits, the Earth rotates below (Figure 18.77). The result is that the satellite covers the surface of the Earth in a long, continuous strip. Spy satellites, GPS and remote sensing satellites orbit in this way so that they can cover the whole surface of the Earth.

TEST YOURSELF

45 Outline the advantages and disadvantages of fibre optic and copper wires as a transmission medium.

46 BBC Radio 5 live transmits at frequencies of 693 kHz and 909 kHz in the UK.

a) Calculate the wavelength for each of these broadcasts.

b) Describe the characteristics of these waves and explain why this means that only a few transmitters are required to cover the UK.

47 Satellite communication uses radio frequencies in the range 5–40 GHz.

a) Explain one reason why these high frequencies are used.

b) Why is it necessary for satellites to use different frequencies for their up-link and down-link?

48 The BBC long-wave transmissions from Droitwich can be received in most of the UK and much of northern Europe. The FM broadcasts of the same material can only be received locally.
Explain why there is this difference in reception areas.

49 Describe the advantage of using a geostationary satellite for communications over a polar orbiting satellite.

Modulation and multiplexing

Modulation

In order for information, digital or analogue, to be transmitted over a distance, a carrier has to be used. This is for two reasons: the low-frequency speech or music would not travel very far in space and only one signal could be transmitted and received at a time in any one place. A carrier is a higher-frequency wave that can be modulated in order to transmit the information. There are two basic forms of modulation: frequency modulation (FM) and amplitude modulation (AM).

Both of these forms of modulation are used in radio broadcasting. FM is used in the higher-frequency ranges, as it is more reliable and gives better reproduction than AM, but it uses a lot of the available frequency range. If FM were used on the medium waveband, there would only be room for three or four radio stations. AM is used for transmission down optical fibres but when used in optical fibres, the information is first made into a digital code.

Amplitude modulation (AM)

Amplitude modulation is the simplest modulation system to understand. Medium- and long-wave radio, TV signals and optical fibres all use forms of AM.

In its simplest form, the original signal (analogue or digital) is made to modulate or change the amplitude of a carrier (radio waves or light).

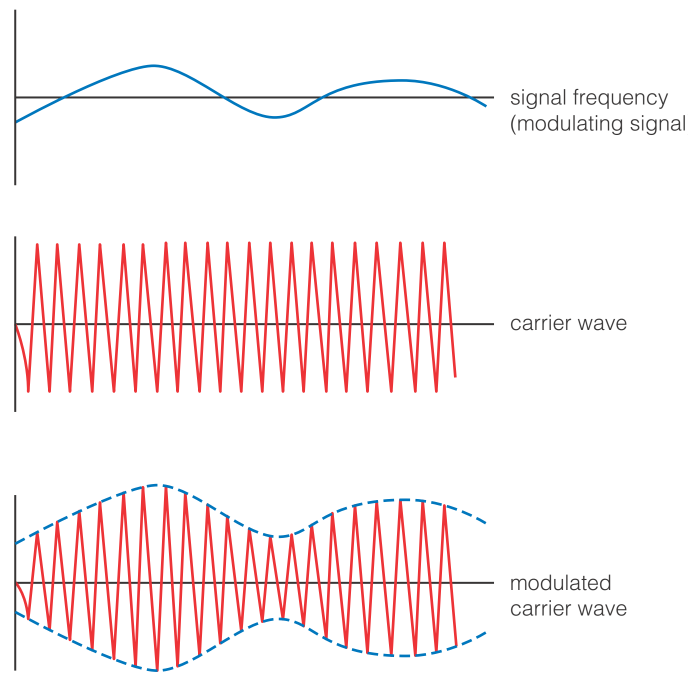


Figure 18.78 Amplitude modulation is when the amplitude of a high-frequency carrier wave is modified by the signal to be transmitted.

Figure 18.78 shows that the amplitude of the final signal follows the original signal. This signal could have just as easily been a digital signal.

To decode this signal in the receiver the bottom half of the signal is removed and the 'humps' of the carrier are smoothed out (Figure 18.79).

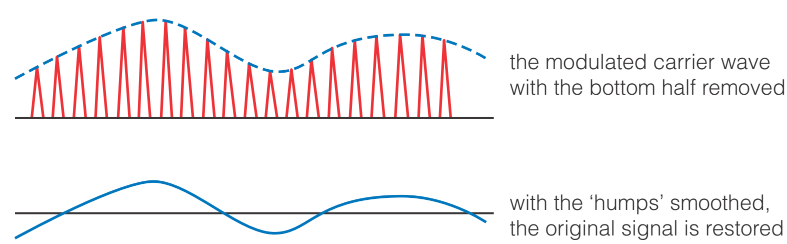


Figure 18.79 The amplitude-modulated signal is decoded by removing the negative part of the signal and smoothing the carrier wave.

It might be thought that only the frequency of the carrier need be transmitted but this is not the case. The bandwidth of the required signal for an amplitude modulated wave is the range of frequencies, centred on the carrier frequency, that need to be transmitted and received in order for the original signal to be recovered.

When a carrier wave of frequency f is amplitude modulated by a signal of frequency f_m, signals of $(f + f_m)$ and $(f - f_m)$ are also generated. These additional frequencies, called sidebands, must be received in order for information from the modulating signal to be recovered. The bandwidth of an AM signal is therefore $2f_m$, where f_m is the maximum frequency that is to be transmitted (Figure 18.80).

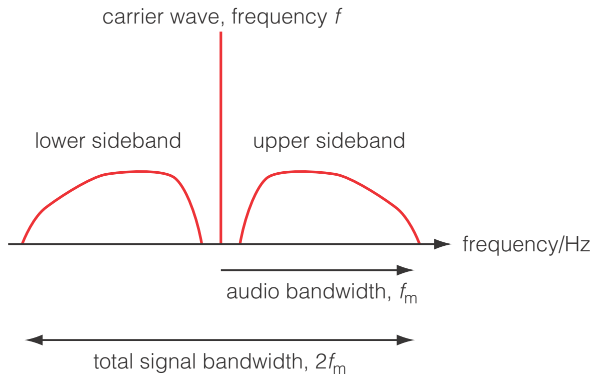


Figure 18.80 Modulating the carrier wave using AM generates sidebands which must also be transmitted.

AM broadcasts on the medium waveband in the UK and Europe limit the frequency range for the audio signals to between 40 Hz and 5 kHz. This is why the audio quality of medium-wave radio is limited and is mainly suited to talk radio.

The reason why the bandwidth is restricted is because the medium waveband lies between 526.5 kHz and 1606.5 kHz and this allows only about 100 separate stations to broadcast in a wide geographical area.

Frequency modulation (FM)

FM is used to give higher quality audio than is possible with AM. FM has several advantages over AM. The main two advantages are:

- It is resilient to noise: as most noise comes from a varying amplitude, this can be removed so that only frequency variations appear.
- It is resilient to signal strength variations: in the same way that varying amplitude noise can be removed, so too can any signal variations. This means that frequency modulation does not suffer from audio amplitude variations as the signal level varies, and it makes FM ideal for use in mobile applications where signal levels constantly vary.

Two disadvantages of FM are:

- It requires a more complicated demodulator: the demodulator is more complicated than for AM because the demodulator has to track the variations of frequency of the carrier rather than just the variations of amplitude.

- It requires a greater bandwidth than AM: the sidebands for an FM transmission theoretically extend out to infinity. To limit the bandwidth of the transmission, filters are used and these cause some distortion of the signal.

Frequency modulation is achieved by making the modulating signal change the frequency of the carrier wave (Figure 18.81).

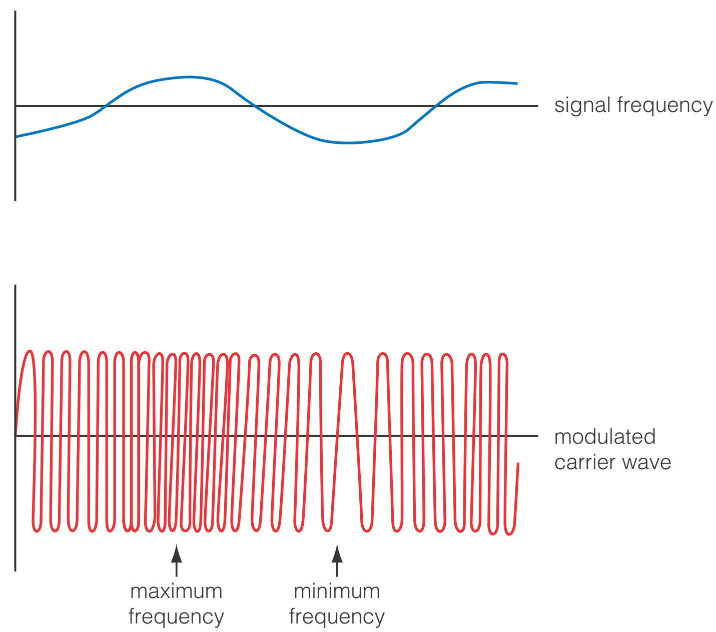


Figure 18.81 With frequency modulation the original signal changes the frequency of the carrier wave according to the amplitude and frequency of the signal.

When the frequency of a carrier wave is changed two sinusoidal functions are being multiplied and this produces an infinite series as a result. Thus if the carrier frequency is f which we modulate at a frequency f_m, then there are sidebands at $\pm f_m$, $\pm 2f_m$, $\pm 3f_m$ … etc. This is shown in Figure 18.82.

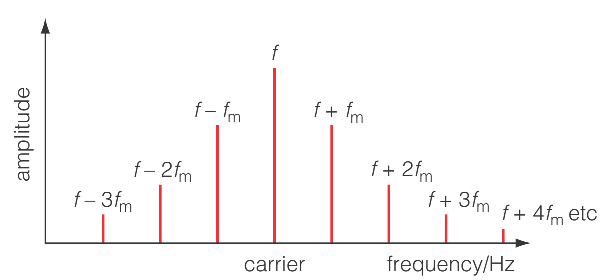


Figure 18.82 The sidebands for FM modulation are at multiples of the modulating frequency, f_m.

However, a rule of thumb, known as Carson's rule, is that 98% of the signal is transmitted if the bandwidth is given by:

$$\text{bandwidth} = 2(\Delta f + f_m)$$

where Δf is the maximum deviation of the frequency from the carrier frequency and is determined by the instantaneous maximum amplitude of the modulating signal, so a signal with a large maximum amplitude (regardless of its frequency) will require a large Δf.

Some FM radio stations will compress the dynamic range of their output in order to reduce Δf.

EXAMPLE

FM radio broadcasts have an allowed maximum deviation, Δf, of 75 kHz and a maximum modulating frequency of 15 kHz. Calculate the bandwidth required.

Answer

bandwidth = $2(\Delta f + f_m)$ = 2(75 kHz + 15 kHz) = 180 kHz

The requirement for a larger bandwidth is why FM broadcasts are not used on the medium waveband but use the VHF band, which operates at frequencies between 87.5 MHz and 108.0 MHz. This is more so because, unlike the simple example given, a stereo signal has to be encoded which requires a greater bandwidth.

In the UK, the carrier frequency of each broadcasting station, broadcasting a stereo signal, is separated by 400 kHz. This gives space for about 50 radio channels but since VHF frequencies fall into the category of space waves, they are generally 'line of sight' and quite short range so that carrier frequencies can be re-used in different parts of the country.

Data capacity of a transmission channel

In 1949 Claude Shannon worked out how many bits per second are required to send a digitised signal.

To digitise a signal, it needs to be sampled at least twice as frequently as the maximum frequency required, f. Therefore:

$$\text{sampling rate} = 2f$$

The number of bits per second, n, is equal to the number of bits per sample, b, times the sampling rate:

$$n = 2fb$$

The worst-case situation is to have to transmit 0101010101… A '01' pair counts as one cycle of a wave so the maximum frequency of transmission (signal bandwidth) is $\frac{n}{2}$ where n is the number of bits per second to be transmitted.

Therefore, the signal bandwidth required to transmit a signal of maximum frequency f with b bits per sample is given by:

$$\text{signal bandwidth} = \frac{n}{2} = \frac{2fb}{2} = f \times b$$

EXAMPLE

A mobile phone signal is to be sampled using 8 bits per sample so that the maximum frequency transmitted is 4 kHz. Calculate the signal bandwidth required.

Answer

signal bandwidth = $f \times b$ = 4000 × 8 = 32 kHz

The value calculated in the example is quite a large bandwidth when a number of mobile phones are being used in the same area, but fortunately various data compression techniques are used to reduce the amount of data that is actually required to be transmitted.

If CD-quality music is to be transmitted, the required bandwidth is about 700 kHz and this would cause problems for most transmission paths. Data compression techniques, e.g. mp3, are used to reduce the bandwidth, but these inevitably reduce the range of frequencies transmitted and thus the quality.

Multiplexing

Multiplexing is when several different signals are sent down the same transmission path at the same time. The two main forms of multiplexing are:

- frequency division multiplexing – carriers of several different frequencies are used in the same transmission path, each modulated using AM or FM
- time-division multiplexing – signals are chopped up into small segments and single segments from each signal are sent sequentially down the transmission path.

Some transmission paths use both forms of multiplexing at the same time.

Digital television (DVB) and digital audio broadcasting (DAB) use multiplexing to code several digital signals into one channel. A form of frequency division multiplexing is used to do this.

Time-division multiplexing (TDM)

Time-division multiplexing is mainly used for voice and data channels.

The capacity of most transmission paths is considerably greater than is required for one single phone call or other data transfer, so that several data sources or voice channels can be sent down the same transmission path. TDM is the way this is achieved.

Suppose we have five telephone calls we want to route down the same optical fibre transmission path.

First, the analogue signals are sampled to give a digital signal. Because the analogue signal is sampled, there is plenty of time between samples to send samples from other signals down the transmission path.

The digital signals are then broken up into small segments, which may each be one sample, and to each segment is added some information that identifies the call from which it came and the destination.

In synchronous TDM each call is allocated a particular time slot and in the allocated time slot a segment from the first call is sent down the transmission path.

This is immediately followed by a segment from the next call in the second time slot, then the next, and so on until a segment from each of the five calls has been sent. The set of five segments is called a frame. The process of combining the segments into one digital stream is called multiplexing.

The next segment from the first call is then sent, and so on, to produce the next frame.

At the other end, the segments for each call from each frame are routed separately, reconstructed and the digital signal processed as needed. This process is, unsurprisingly, called demultiplexing.

In practice, up to 30 voice calls can be assembled into a single frame and a single transmission path can be used to send many frames at the same time. This is illustrated in Figure 18.83.

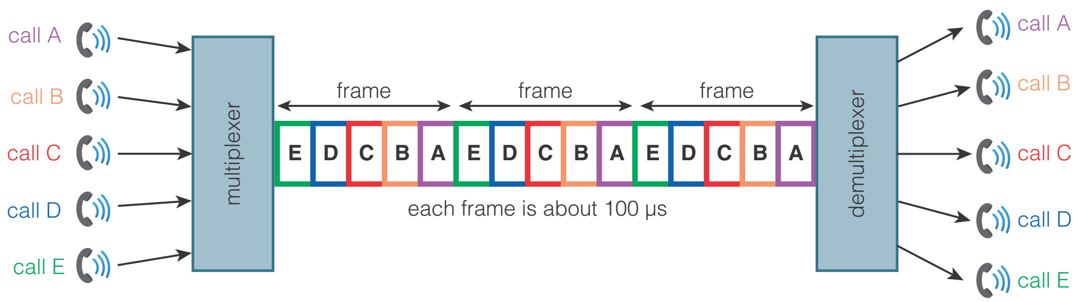


Figure 18.83 Segments of each call are sent sequentially down the transmission path. Each call is sampled about 10 000 times a second, or one sample every 100 μs.

Ethernet – a means of communication between computers – uses a form of TDM. Packets of data are assembled into frames that have error checking codes and destination addresses included before being transmitted. The destination address allows the data to be routed, often by several different routers, to the correct destination. Not all the frames will necessarily follow the same route so the packets have to be re-assembled in the correct order at the final destination.

TEST YOURSELF

50 An audio signal has a frequency range between 18 Hz and 18 kHz. It is to be used to amplitude modulate a carrier wave of frequency 600 kHz.

a) Calculate the bandwidth of the modulated signal.

The medium waveband in Europe extends from 526.5 kHz to 1606.5 kHz.

b) Calculate the maximum number of these audio signals that could be fitted into the medium waveband.

c) Explain why the frequency range for audio broadcasting is limited to about 5 kHz.

51 Outline the advantages and disadvantages of FM compared to AM.

52 An FM broadcast is allocated a bandwidth of 150 kHz. The maximum deviation of the frequency, Δf, is 60 kHz.

Calculate the maximum modulating frequency.

53 A digital recording is to be made of some music using 16 bits per sample. The frequency range is from 20 Hz to 20 kHz.

a) Calculate the required sampling rate.

b) Calculate the number of bits per second.

Since the recording is in stereo, there are two separate sets of samples.

c) Calculate the total number of bits per second for both channels.

d) Calculate the signal bandwidth required to transmit the stereo signal.

e) Suggest why music streaming uses significant data compression techniques.

54 A remote telephone exchange has many telephone users in the local area but only one fibre optic cable connecting it to the next exchange.

a) Explain how the signals from a number of users are multiplexed onto the fibre optic.

b) What is the advantage of a fibre optic connection over a copper wire connection?

Exam practice questions

1 A NAND gate requires an LED as its output. The logic gate cannot give sufficient current to drive the LED so an n-channel enhancement mode MOSFET is to be used as shown.

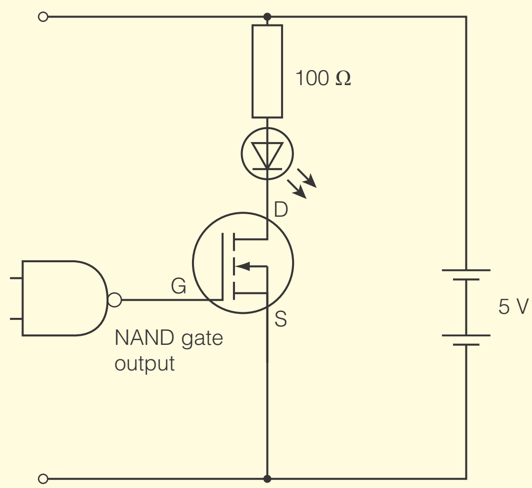


Figure 18.84

a) When the output of the NAND gate is high, will the LED be on or off? Explain your answer. *(2)*

b) Explain why the characteristics of the MOSFET make it suitable for use in this application. *(2)*

2 A Zener diode is to be used to provide a steady 3.3 V from a 12 V supply. The Zener diode to be used is rated at 3.3 V 100 mW.

a) Draw the circuit you would use for this application using the Zener diode and a resistor. *(2)*

b) Calculate the minimum value for the resistor. *(2)*

3 The ideal op amp circuit is designed to give an output voltage, V_{out}, that can be varied by means of a 47 kΩ potentiometer.

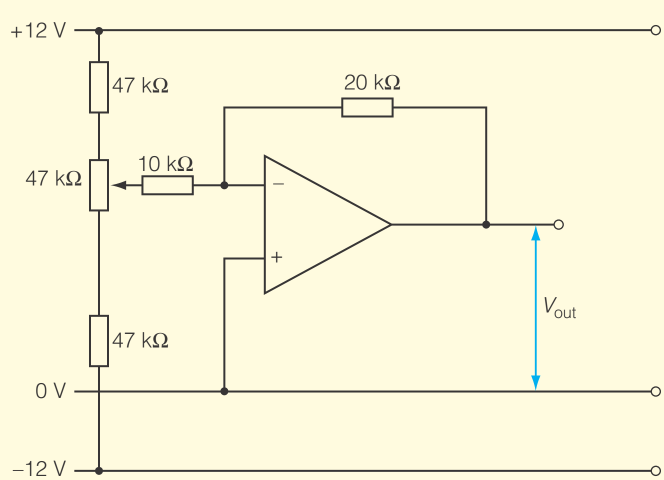


Figure 18.85

a) Calculate the maximum and minimum input voltages obtainable from the 47 kΩ potentiometer. *(2)*

b) Calculate the range of V_{out} from the operational amplifier as the potentiometer is adjusted from maximum to minimum. *(3)*

4 A filter for a telephone is constructed using an inductor and a capacitor, as shown in Figure 18.86.

It is required to have a resonant frequency of 2.5 kHz and a bandwidth of 4.0 kHz.

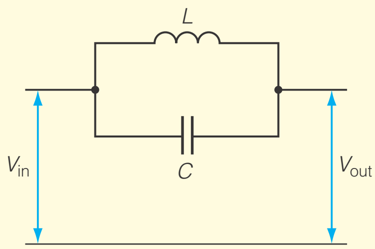


Figure 18.86

If the inductor has a value of 1.0 mH, calculate:

a) the value for the capacitor, *C* *(2)*

b) the *Q* factor. *(1)*

A *LC* system like this is analogous to a mass–spring system.

c) Explain how the two systems are analogous and what feature in the electrical circuit corresponds to damping of the mass–spring system. *(4)*

5 An arrangement of logic gates is set up as shown.

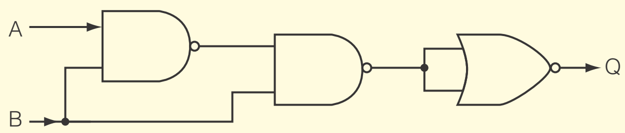


Figure 18.87

a) Draw a truth table for this system. *(3)*

b) Using your truth table, or otherwise, determine the Boolean expression for the output Q. *(2)*

6 A clock pulse is generated by an astable circuit.

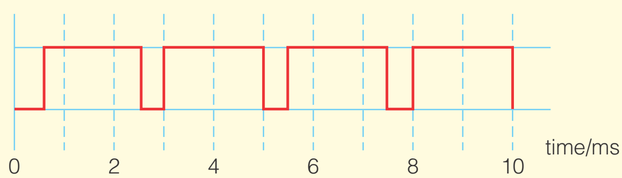


Figure 18.88

a) Use the scale to show that the frequency of the clock is 400 Hz. *(1)*

b) Give the mark-to-space ratio and the duty cycle for this set of pulses. *(2)*

A modulo-n counter is to be used to produce a 10 Hz clock pulse from this high-frequency clock, using two 10-stage Johnson counters.

c) Explain how a 10 Hz clock pulse can be produced. *(2)*

The 400 Hz frequency is accurate to within ±2 Hz.

d) Calculate the accuracy of the 10 Hz clock pulse. *(1)*

7 Compare the advantages and disadvantages of using AM and FM for radio transmission. *(4)*

8 A Hall effect sensor is used as part of a tachometer. A sensor is positioned next to a magnetic toothed wheel that has seven teeth. The sensor outputs a pulse every time a tooth passes it.

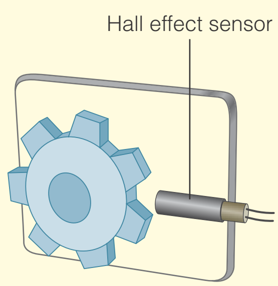


Figure 18.89

a) Explain how a Hall effect sensor can detect the teeth on the wheel. (2)

After the sensor, it is required that there is a single pulse output for every complete turn of the wheel.

b) Describe a logic system that will produce the required output. (2)

9 The circuit shown below was constructed to be an automatic night-light. However, it does not operate correctly as the LED is on in the light and off in the dark.

When it is dark, the voltage at the gate of the MOSFET is below V_{th}.

V_{th} for this MOSFET is 1.5 V.

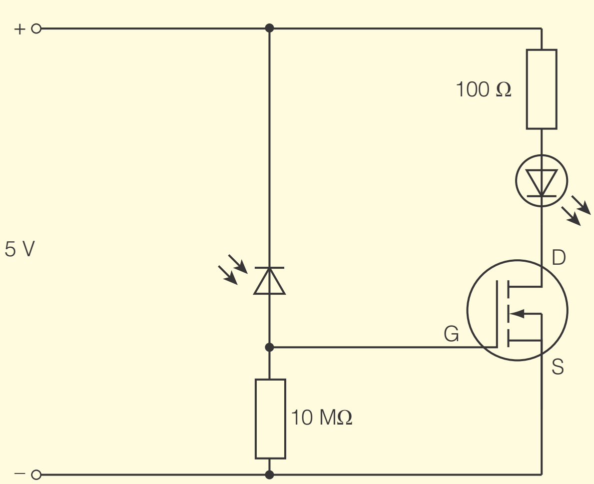


Figure 18.90

a) Calculate the maximum dark current through the photodiode. *(2)*

b) Suggest a modification that will make the LED switch on in the dark. *(1)*

10 The noisy analogue signal shown is sampled every 0.5 ms.

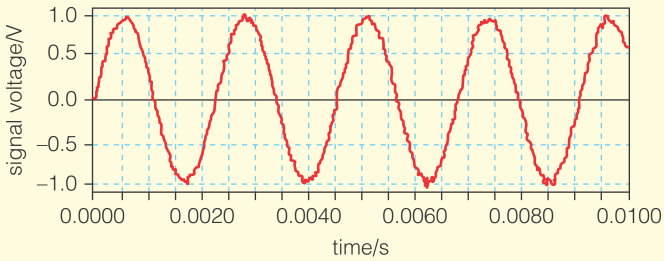


Figure 18.91

a) Explain whether this sampling rate is suitable for this signal. *(2)*

b) The maximum signal is 1.8 V peak-to-peak. From the diagram, estimate the maximum peak-to-peak noise signal. *(1)*

c) Calculate the maximum number of bits it is appropriate to use to encode this signal. *(1)*

d) Sampled signals can suffer from *quantisation* errors. Explain the term quantisation. *(2)*

11 An ideal operational amplifier is used in an amplifier circuit.

The power supply is ±12 V.

R_f = 820 kΩ and R_{in} = 4.7 kΩ.

For this amplifier, gain × bandwidth = 3 MHz.

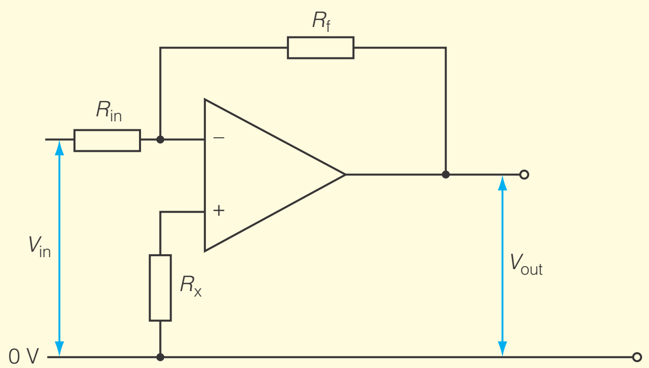


Figure 18.92

a) Calculate the gain of the amplifier. *(1)*

b) Calculate the bandwidth of this amplifier and discuss whether this is adequate for an amplifier for audio signals. *(3)*

c) What is the maximum amplitude of input signal that can be used? *(1)*

12 The circuit shown is set up with an AND gate connected to the Q_1 and Q_3 outputs of a binary counter.

When there is 1 on the reset pin, the outputs, Q_0 to Q_3 are all reset to 0.

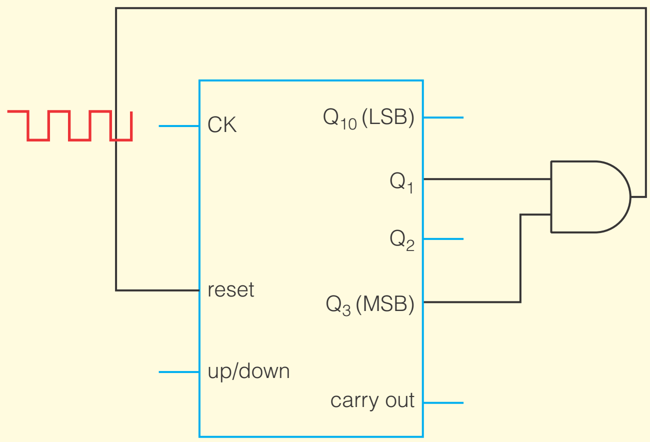


Figure 18.93

a) Give the conditions under which the output of an AND gate is 1. *(1)*

b) Explain the effect of adding the AND gate to the binary counter. *(3)*

c) The inputs to the AND gate are now taken from outputs Q_1 and Q_2. Describe the effect on the counter. *(1)*

13 Many telephone calls can be sent down the same transmission path on a single carrier using time-division multiplexing (TDM).

a) Explain the term multiplexing. *(1)*

b) Briefly outline the different transmission media that might be used for the transmission path. *(3)*

c) Describe how TDM allows many calls to be sent along the same path at the same time. *(4)*

14 A single long-wave radio transmitter provides coverage for the whole country but an FM transmitter will only provide coverage in its local area.

Explain why this is the case. *(4)*

Answers

Answers to Test yourself on prior knowledge

1 a) $V_2 = V_{in}\left(\frac{R_2}{(R_1+R_2)}\right) = 12\times\left(\frac{2.2\times10^3}{6.8\times10^3+2.2\times10^3}\right) = 2.9\ \text{V}$

b) $I = \frac{V}{R} = \frac{12}{9.0\times10^3} = 1.3\times10^{-3}\ \text{A} = 1.3\ \text{mA}$.

c) $P = V\times I = 12\times1.3\times10^{-3} = 16\times10^{-3}\ \text{W} = 16\,\text{mW}$.

2 $F = EQ = 1000\times -1.6\times10^{-19} = -1.6\times10^{-16}\,\text{N}$

3 $F = EQ = \frac{V}{d}\times Q = \frac{5}{0.5\times10^{-6}}\times1.6\times10^{-19} = 1.6\times10^{-12}\,\text{N}$

4 a) $F = EQ = 1.0\times10^6\times1.6\times10^{-19} = 1.6\times10^{-13}\,\text{N}$

b) The direction of the force is in the same direction as the electric field, that is, from the positive to the negative potential.

5 $C = \frac{Q}{V} = \frac{10\times10^{-6}}{5.0} = 2.0\times10^{-6}\,\text{F} = 2.0\,\mu\text{F}$

6 $\varepsilon = -\frac{\Delta N\phi}{\Delta t} = -\frac{\Delta NBA}{\Delta t} = -\frac{1000\times200\times10^{-3}\times1.0\times10^{-4}}{2.5} = 8.0\times10^{-3}\,\text{V}$

7 $F = -m\omega^2 x = 4m\pi^2 f^2 x = 4\times0.5\times\pi^2\times3.0^2\times5.0\times10^{-2} = 8.9\,\text{N}$

8 $f_0 = \frac{1}{2\pi\sqrt{\frac{m}{k}}} = \frac{1}{2\pi\sqrt{\frac{0.10}{4.0\times10^2}}} = 10\,\text{Hz}$

9 $f = \frac{1}{T} = \frac{1}{0.5\times10^{-3}} = 2000\ \text{Hz}$

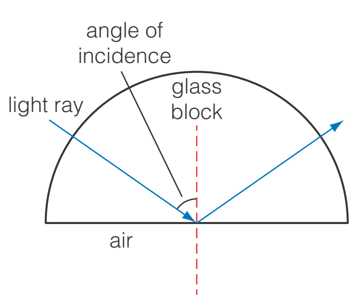
10 $V = V_0\left(1-e^{-\frac{t}{RC}}\right) = 12\times\left(1-e^{-\frac{9.0}{470\times10^3\times22\times10^{-6}}}\right) = 7.0\ \text{V}$

11 Time to fall to 37% of its starting value is RC, the time constant.

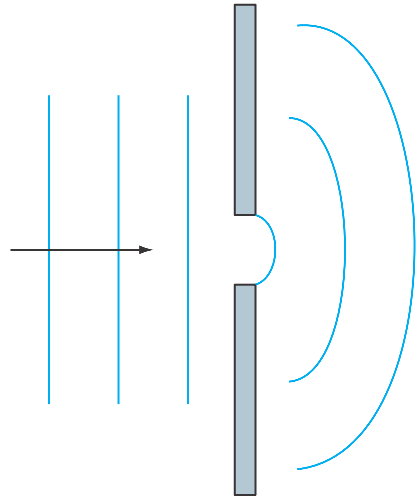
$RC = 100\times10^3\times1.0\times10^{-6} = 0.1\ \text{s}$

12 $v = f\times\lambda \rightarrow \lambda = \frac{3.0\times10^8}{101.1\times10^6} = 3.0\ \text{m}$

13 a) and **b)**



14



15 The microwaves from the transmitter are polarised. When the received signal is maximum, the planes of polarisation of the transmitter and receiver are the same. When either the transmitter or receiver are rotated through 90°, the planes of polarisation are crossed and so no signal is received.

Answers to Test yourself questions

1 Phosphorus is 5-valent, which means it has five electrons in its outer shell that can take part in bonding. When phosphorus is added to the silicon lattice, four of the electrons bond with the neighbouring silicon atoms to complete their outer shells, leaving one electron spare. This can become free and take part in conduction. Under the influence of an electric field the electrons will drift in a direction opposite to the direction of the electric field.

2 Boron is 3-valent, which means it has three electrons in its outer shell that can take part in bonding. When boron is added to the silicon lattice, these three electrons are shared with the neighbouring silicon atoms, leaving one space or 'hole' where there is a vacancy for an electron that is needed to complete the outer shell. Electrons from nearby atoms can move from their bound states into the vacancy so that the hole moves around the lattice. Thus the vacancy is effectively a positive charge carrier that moves under the influence of an electric field.

3 Because the gate contact is to an electrode that is separated from the substrate by an insulating layer, there is no direct connection between the gate and source or gate and drain. This ensures that the input resistance is very high.

4 The gradient of the nearly vertical part of Figure 18.13 is g_m, the gradient

is $\frac{(2.35-1.9)}{(180-40)}$ S = 320 mS

5 a) When there is no rain, there is zero gate voltage and with a zero gate voltage, V_{GS}, the very small drain current, the *zero gate voltage drain current,* I_{DSS}, caused by conduction by the minority charge carriers across the reverse biased p-n junction between drain and substrate is typically a few µA so a battery would last a long time.

b) The water forms part of a potential divider circuit with the 2 MΩ resistor. The output voltage from this potential divider circuit must be about 1.35 V.

$$\text{Therefore: } 1.35 = 9.0\left(\frac{2.0\times10^6}{R_1 + 2.0\times10^6}\right)$$

c) The circuit could be damaged because if the copper strips are shorted the gate will be connected directly to the positive of the power supply so V_{GS} is 9 V, allowing a very large I_{DS} which may overheat the transistor.

d) To prevent damage a resistor of at least 2 MΩ should be put in the lead from the positive the power supply.

6 See Figure 18.18.

7 The maximum current, I, that can flow is given.

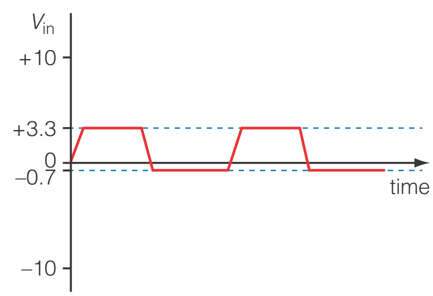
$$P = V\ I \rightarrow \frac{500\times10^{-3}}{5}\text{ A} = 100\text{ mA}$$

With an input voltage of 12 V then the resistor must have a pd of 7 V across it when a current of 100 mA flows.

This means that the minimum resistance, R, is given by:

$$R = \frac{V}{I} = \frac{7}{100\times10^{-3}} = 70\ \Omega$$

8



9 The reverse biased p-n junction is exposed to light and the energy of the photons is transferred to the lattice and is sufficient to release more electron–hole pairs in the depletion layer. These additional charge carriers can take part in conduction. The number of electron–hole pairs, and hence the photocurrent, will be proportional to the amount of light incident on the junction.

10 Without a daylight filter, bright light in a room would cause a large unwanted signal to be detected, making the receiver give a permanently high or low output and so it would not be sensitive to the required signal. It would only work correctly and reliably in a dark room.

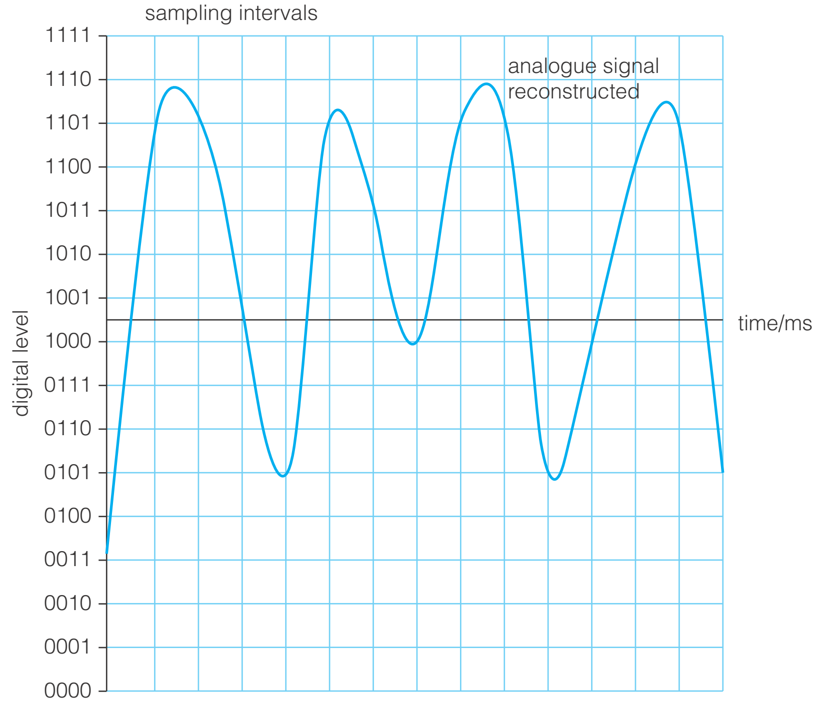
11 $\frac{1600}{32} = 50$ Hz

12 $\Delta V = 3.3 - 2.5 = 0.8$ V

$$\frac{0.8}{25\times10^3} = 3.2\times10^{-5} = 32\ \mu\text{T}$$

13 a) 0011, 1101, 1101, 1001, 0101, 1101, 1011, 1000, 1101, 1101, 0101, 1000, 1100, 1101, 0101.

b) The resultant signal has lost some of the detail although it is roughly the same shape.



14 Quantisation is when the actual level of a sampled analogue signal is approximated to the nearest digital level at the sampling time.

15 Signals become noisy because of electrical noise introduced from thermal movement of charge carriers in electronic components and from electromagnetic induction in wires. Electrical connections can also introduce noise.

16 $b = \log_2\left(\frac{V_{\text{signal + noise}}}{V_{\text{noise}}}\right) = \log_2\left(\frac{1.001}{0.001}\right) = 10$. Therefore the maximum number of bits is 10.

17 When analogue signals become noisy, it is difficult to separate the signal from the noise and any amplification after the signal has been attenuated in transmission amplifies the noise as well.

A digital signal will also become noisy but at the receiving end, it can be cleaned up as all that is required is for the difference between a high and a low to be distinguishable. As long as the noise is not so great as to make the two levels indistinguishable, the original digital signal can be recovered.

18 a) $X_L = 2\pi fL = 2 \times \pi \times 50 \times 1.0 \times 10^{-3} = 0.31\ \Omega$

b) $X_L = 2\pi fL = 2 \times \pi \times 50 \times 10^3 \times 1.0 \times 10^{-3} = 310\ \Omega$

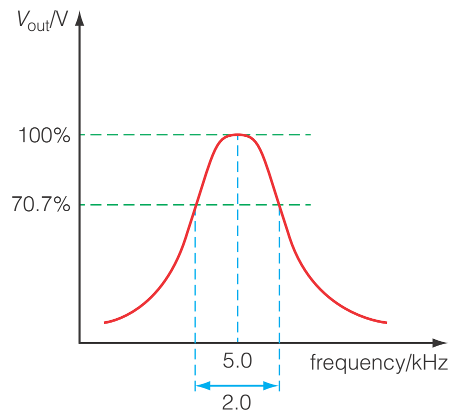
19 a) $X_C = \frac{1}{2\pi fC} = \frac{1}{2 \times \pi \times 50 \times 1.0 \times 10^{-6}} = 3.2 \times 10^3\ \Omega$

b) $X_C = \frac{1}{2\pi fC} = \frac{1}{2 \times \pi \times 50 \times 10^3 \times 1.0 \times 10^{-6}} = 3.2\ \Omega$

20 $f_0 = \dfrac{1}{2\pi\sqrt{LC}} = \dfrac{1}{2\times\pi\times\sqrt{1.0\times10^{-3}\times1.0\times10^{-6}}} = 5032\text{ Hz} \approx 5.0\text{ kHz}$

21 $Q = \dfrac{f_0}{f_b} = \dfrac{5.0\times10^3}{2.0\times10^3} = 2.5$

22



23 **a)** The 100 kΩ potentiometer will allow the point at which the comparator switches to indicate high or low temperature to be adjusted.

b) At low temperature, the resistance of TH1 is high. This means that the pd at the + input to the comparator is low because TH1 and the 10 kΩ resistor form a potential divider.

If the + input is below that of the – input the output will go low (–12 V) and the red LED will light.

24 Assuming that the op amp is ideal, the maximum positive output is 12 V.

$V_{out} = A_{OL}(V_+ - V_-)$ therefore $\dfrac{12}{30\times10^6} = (V_+ - V_-) = 4.0\times10^{-7}$ V

25 Point X is not actually connected to 0 V or earth but because of the very large open-loop gain of the op amp, the difference between the +ve input (which is at 0 V) and the –ve input must be virtually zero.

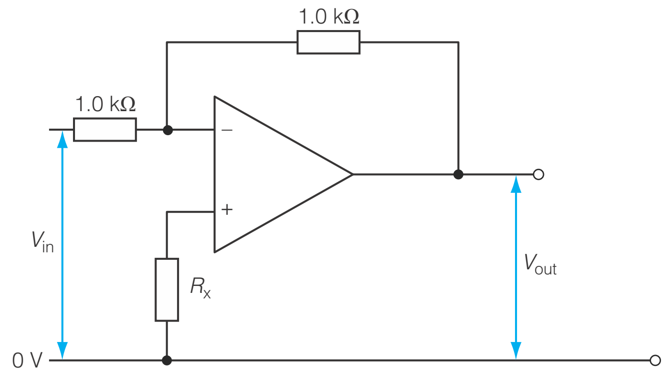
26 The gain is $-\dfrac{100\text{ k}\Omega}{1.0\text{ k}\Omega} = -100$

27 Gain $= -\dfrac{R_f}{R_{in}} = -\left(\dfrac{470\times10^3}{10\times10^3}\right) = -47$

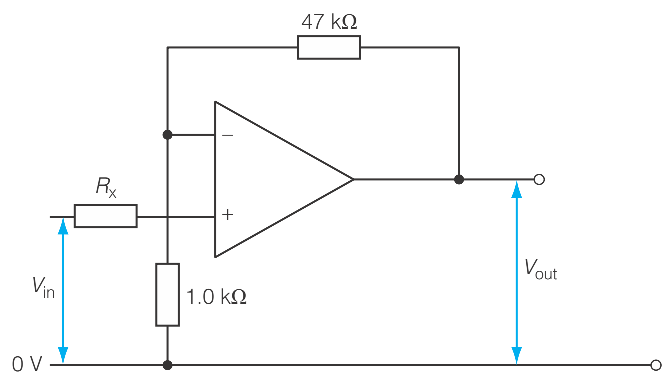
28 Gain $= 1 + \dfrac{R_f}{R_{in}} = 1 + \dfrac{39\times10^3}{1.0\times10^3} = 1 + 39 = 40$

29 Example circuits are shown.

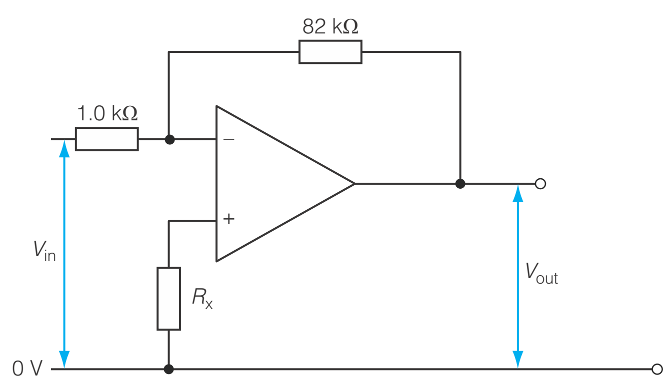
a)



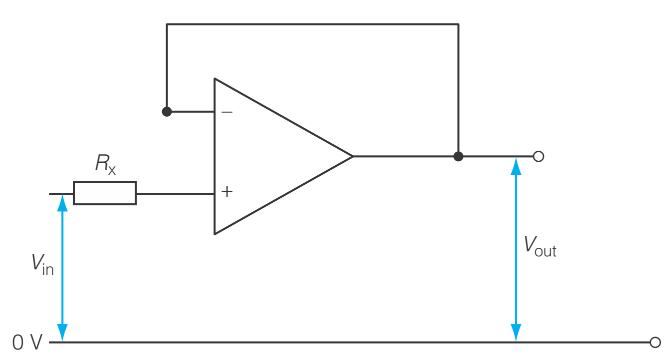
b)



c)



d)



30 Resistor R_X is not part of the feedback loop. The input is through R_X but since the ideal op amp has infinite input resistance, there is no current through R_X and so no pd across it and the input at the non-inverting input is equal to V_{in}.

31 $$V_{out} = -R_f\left(\frac{V_1}{R_1}+\frac{V_2}{R_2}+\frac{V_3}{R_3}\right)$$

$$= -10 \times 10^3\left(\frac{0.1}{2.0 \times 10^3}+\frac{0.05}{2.0 \times 10^3}+\frac{0.08}{2.0 \times 10^3}\right) = -1.15\ \text{V}$$

32 $$V_{out} = -R_f\left(\frac{V_1}{R_1}+\frac{V_2}{R_2}+\frac{V_3}{R_3}+\frac{V_4}{4}\right)$$

$$= -1.0 \times 10^3\left(\frac{1.0}{1.0\ \ 10^3}+\frac{1.0}{2.0\ \ 10^3}+\frac{1.0}{4.0\ \ 10^3}+\frac{1.0}{8.0\ \ 10^3}\right)$$

$$= -1.875\ \text{V}$$

33 a) $$V_{out} = (V_+ - V_-)\frac{R_f}{R_{in}} = (1.1-1.0) \times \frac{10 \times 10^3}{1.0 \times 10^3} = 1.0\ \text{V}$$

b) $$V_{out} = (V_+ - V_-)\frac{R_f}{R_{in}} = (1.0-1.1) \times \frac{10 \times 10^3}{1.0 \times 10^3} = -1.0\ \text{V}$$

34 If the open-loop gain was high at very high frequencies, positive feedback could occur which would cause the output to be unstable and render the op amp unusable. The gain is therefore limited by an internal capacitor so that this cannot occur at any frequency when the open-loop gain is more than one.

35 a) $$\text{Gain} \times \text{bandwidth} = \text{constant} \rightarrow \text{bandwidth} = \frac{\text{constant}}{\text{gain}}$$

$$\therefore \text{bandwidth} = \frac{4 \times 10^6}{10\,000} = 400\ \text{Hz}$$

b) This would not be suitable as an audio amplifier, which would need the bandwidth to be about 20 kHz. This would require a maximum

$$\text{gain} = \frac{4 \times 10^6}{20 \times 10^3} = 200.$$

36 a)

A	B	$\bar{A}$	$\bar{B}$	Q
0	0	1	1	0
0	1	1	0	1
1	0	0	1	1
1	1	0	0	1

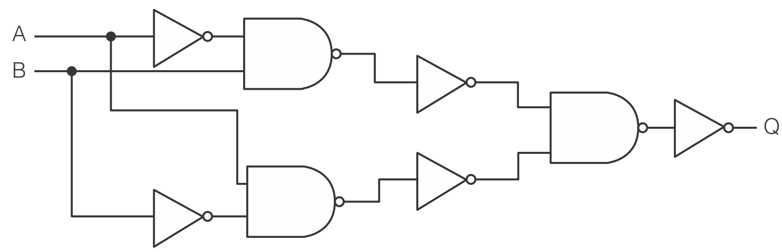
b) The function is: $Q = \overline{(\bar{A} \bullet \bar{B})}$

Replace AND with OR and invert each variable.

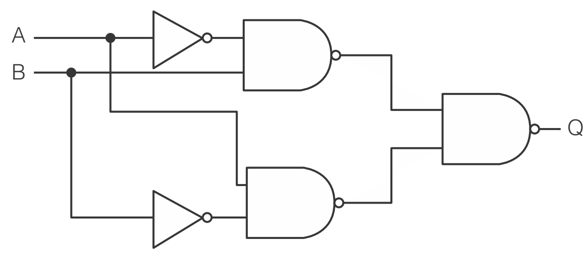
$Q = \overline{(\bar{A} \bullet \bar{B})} = (A + B)$ so this is actually an OR function, which can be checked from the truth table in the answer to part (a).

37

A	B	$\bar{A}$	$\bar{B}$	$(A \cdot \bar{B})$	$(\bar{A} \cdot B)$	Q
0	0	1	1	0	0	0
0	1	1	0	0	0	1
1	0	0	1	1	1	1
1	1	0	0	0	0	0



This simplifies to:



38 The truth table is below; this is the truth table for an EOR gate.

A	B	$\overline{A \cdot B}$	$\overline{(\overline{A \cdot B}) \cdot A}$	$\overline{(\overline{A \cdot B}) \cdot B}$	Q
0	0	1	1	1	0
0	1	1	1	0	1
1	0	1	0	1	1
1	1	0	1	1	0

39 **a)** $f = \dfrac{1.44}{(R_1 + 2R_2) \times C} = \dfrac{1.44}{(4.7 \times 10^3 + 2 \times 47 \times 10^3) \times 0.01 \times 10^{-6}} = 1460\text{ Hz}$

b) Duty cycle $= \dfrac{47 \times 10^3}{(4.7 \times 10^3 + 2 \times 47 \times 10^3)} \times 100\% = 48\%$

c) i) To get the duty cycle as near to 50% as possible, would need $R_2 \gg R_1$. For example, R_2 = 100 kΩ and R_1 = 1.0 kΩ would give a duty cycle of 49.8%.

ii) To get the duty cycle to 33%, make $R_2 = R_1$. For example, $R_2 = R_1$ = 100 kΩ will give a duty cycle of 33%.

40 In Figure 18.64, Q_1 and Q_3 are high on the tenth pulse while Q_0 and Q_2 are low. Since Q_3 is the MSB, this gives 1010 (which is decimal 10).

41 A binary counter will count from 0000 to 1111, if it is a 4-bit binary counter, before it resets and starts at 0000 again. A BCD counter will reset on the tenth pulse so will count from 0000 to 1001 and then reset to 0000. This allows the output to be decoded for a seven-segment digit.

42 **a)** This is a modulo-7 counter. The output goes high every 7th clock pulse.

b) Q_6 should be connected to the reset pin in order to make this a modulo-7 counter.

43 Any clock will have an uncertainty of ± a few clock pulses per second. Using a high frequency clock pulse and dividing down, the actual uncertainty will be reduced.

44 **a)** A carrier is a high-frequency wave that is modulated by the required signal. Examples of a carrier would be light, radio waves or microwaves. Ultrasound can be used as a carrier.

b) The coded signal has to be placed onto the carrier. This is done by the modulator. For example a coded message might be used to switch on and off a light beam; in this case, the carrier is the light wave and the modulator is the switch. Carriers can generally be amplitude modulated (AM) or frequency modulated (FM).

c) The receiver detects the modulated carrier. A receiver could be an aerial for radio waves or a photodiode for light.

d) All signals will be attenuated in transmission. An amplifier is needed to increase the attenuated signal so that it can be processed or output by a transducer.

45 Optical fibre uses a high-frequency carrier wave; light, and the high frequency of light means that it can transmit data very much faster than copper wires, which will use a much lower carrier frequency.

Optical fibres do not pick up electrical 'noise' in the same way as copper cables which means that the signal coming out of a fibre can be amplified without amplifying the noise.

Glass is much cheaper and has a lower density than copper and a single fibre can transmit as much data as many copper wires.

Optical fibres are more secure. It is possible to intercept the signal in a copper wire without damaging or cutting the wire simply by using electromagnetic induction.

A disadvantage is that, if damaged, they cannot be re-joined simply, as can copper wires.

Generally, more complicated units are required at each end to process the optical signal because an optical signal has to be converted to an electrical signal.

Optical fibres cannot be bent and curved in the same way as copper cables.

46 a) $v = f \times \lambda \rightarrow \lambda = \dfrac{3.0 \times 10^8}{693 \times 10^3} = 433\text{ m}$

$v = f \times \lambda \rightarrow \lambda = \dfrac{3.0 \times 10^8}{909 \times 10^3} = 330\text{ m}$

b) These waves are ground waves and so follow the surface of the Earth by diffraction. This means that they have a larger range than sky waves or space waves. Because their range can be up to 1000 km (depending on the transmitter power), only a few transmitters are required to cover the UK.

47 a) 5–40 GHz are space waves and pass straight through the ionosphere so that they can reach a satellite.

b) It is necessary to have different frequencies for the up-link and the down-link in order to avoid de-sensing. De-sensing is the effect of a strong signal from a transmitter interfering with the detection of a weak signal by a receiver when operating at the same frequency at the same time.

48 Long waves are ground waves and so are diffracted around the Earth. A powerful transmitter enables these waves to have a very large local coverage. Long-wave transmissions are amplitude modulated.

FM can only be used for higher frequencies and shorter wavelengths, and these are space waves which can only be used for 'line of sight' communications. They will not diffract significantly around the surface of the Earth and so can only be received in an area local to the transmitter.

49 A geostationary satellite is positioned above the same place on the Earth's surface and so it can be seen all the time from a point on the surface of the Earth. This means that transmitting and receiving aerials can be permanently pointed at the satellite and will be in communication continuously.

A polar orbiting satellite is only visible for a short period of time from any point on Earth and is moving when it is visible. This makes continuous communication impossible. However, because polar orbiting satellites are closer to the Earth they are better for remote sensing, imaging and GPS signals which need to be picked up by relatively simple equipment on the Earth's surface.

50 a) Bandwidth = $2f_m \approx 36$ kHz

b) $\dfrac{(1606.5\text{ kHz} - 526.5\text{ kHz})}{36\text{ kHz}} = 30$

c) 30 separate radio stations on the medium waveband is not very many. By limiting the frequency range to 5 kHz, the bandwidth is 10 kHz and this means that about 100 separate radio stations can be fitted into the medium waveband.

51 FM is resilient to noise, as most noise comes from a varying amplitude, this can be removed so that only frequency variations appear. It is also resilient to signal strength variations and in the same way that varying amplitude noise can be removed, so too can any signal strength variations. This means that one of the advantages of frequency modulation is that it does not suffer audio amplitude variations as the signal level varies, and it makes FM ideal for use in mobile applications where signal levels constantly vary.

FM, however, does have disadvantages: it requires a more complicated demodulator than AM because the demodulator has to track the variations of frequency of the carrier rather than just the variations of amplitude.

FM requires a greater bandwidth than AM since to transmit most of the modulating signal requires a bandwidth $= 2\left(\Delta f + f_m\right)$.

52 Bandwidth $= 2\left(\Delta f + f_m\right) \rightarrow f_m = \left(\frac{\text{bandwidth}}{2}\right) - \Delta f$

$$= \left(\frac{150 \times 10^3}{2}\right) - 60 \times 10^3 = 15 \times 10^3 = 15 \text{ kHz}$$

53 **a)** Sampling rate = 2 × the maximum frequency, = 40 kHz

b) $40 \times 10^3 \times 16 = 640 \times 10^3$ bits per second.

c) 1.28×10^6 bits per second for both channels.

d) Signal bandwidth $= \frac{n}{2} = 640 \times 10^3$ Hz

e) This is a large signal bandwidth and would take up a lot of capacity on a transmission path. Therefore data compression is generally used for music streaming so that the signal bandwidth can be significantly reduced.

54 **a)** The exchanges will use time-division multiplexing (TDM) to transmit the telephone calls.

The analogue signals will be sampled and each sample from each call will be put together into a frame and sent down the transmission path. Each frame will consist of a sample from each telephone call.

At the receiving end, the samples will be sorted and routed to the separate receivers and the analogue signal reconstructed.

b) Because the fibre optic cable has a higher carrier frequency, the available bandwidth is greater and so more telephone calls can be routed down the same cable than is possible with copper wires.

Answers to Exam practice questions

1 **a)** The LED will be on. *(1)*

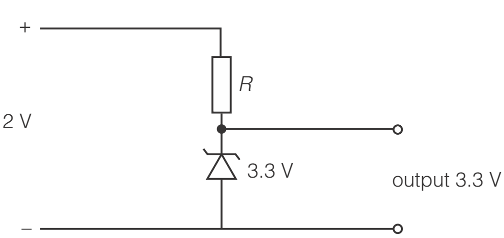
The high output from the NAND gives a high potential difference between the gate and source of the MOSFET with allows a current to flow from drain to source. *(1)*

b) The input resistance of the MOSFET is very high and so the gate will not load the output of the NAND gate. *(1)*

The drain-to-source resistance is low when the gate is high so sufficient current can flow for the LED to light. *(1)*

2 **a)** Correct circuit. *(1)*

Diode correct way around. *(1)*



b) Potential difference across Zener diode = 8.7 V *(1)*

$R = \frac{V^2}{P} = \frac{8.7^2}{100 \times 10^{-3}} = 760\ \Omega$. This is the minimum resistance required. *(1)*

3 a) Potential difference across each 47 kΩ resistor is $\frac{24\ \text{V}}{3} = 8\ \text{V}$. *(1)*

Therefore the minimum input voltage is –4 V and the maximum is +4 V. *(1)*

b) The gain of the op amp circuit is $-\frac{20 \times 10^3}{10 \times 10^3} = -2$. *(1)*

Therefore the maximum is +8 V (when input is –4 V). *(1)*

The minimum is –8 V (when input is +4 V). *(1)*

4 a) $f_0 = \frac{1}{2\pi\sqrt{LC}} \rightarrow C = \frac{1}{4\pi^2 f^2 L} = \frac{1}{4 \times \pi^2 \times 2.5 \times 10^3 \times 1.0 \times 10^{-3}}$

$= 4.1 \times 10^{-6}\ \text{F} = 4.1\ \mu\text{F}$ *(2)*

b) $Q = \frac{f_0}{f_b} = \frac{2.5 \times 10^3}{4.0 \times 10^3} = 0.625$ *(1)*

c) A mass and spring system continuously transfers energy between kinetic and potential while the capacitor inductor system continuously transfers energy between an electric field and a magnetic field. *(1)*

Both systems have a resonant frequency that depends on the system quantities. *(1)*

The mass is acting like the inductor with the inductance, *L*, analogous to the mass, *m*. *(1)*

C is analogous to $\frac{1}{k}$, where *k* is the spring constant. *(1)*

5 a) *(3)*

A	B	$\overline{A \bullet B}$	$\overline{(A \bullet B)} \bullet A$	Q
0	0	1	1	0
0	1	1	0	1
1	0	1	1	0
1	1	0	1	0

b) From the truth table, $Q = \overline{\overline{(A \bullet B)} \bullet B}$. *(1)*

$\therefore Q = \overline{(A \bullet B)} \bullet B$ *(1)*

6 a) 4 complete pulses in 10 ms is 2.5 ms per pulse.

$f = \frac{1}{2.5 \times 10^{-3}} = 400\ \text{Hz}$ *(1)*

b) Mark-to-space ratio is 4 : 1 *(1)*

The duty cycle is therefore $\frac{4}{5} \times 100\% = 80\%$. *(1)*

c) To obtain a 10 Hz clock pulse from a 400 Hz clock requires dividing by 40. *(1)*

This can be done with one 10-stage (modulo-10) counter followed by a modulo-4 counter. *(1)*

d) The 400 Hz pulse is accurate to ±0.5%.

This makes the 10 Hz clock pulse accurate to the same percentage or ±0.05 Hz. *(1)*

7 AM easily picks up noise because noise is generated by varying amplitude of the carrier. Electrical interference, reflections and attenuation will cause these variations. *(1)*

FM is resilient to noise, as most noise comes from a varying amplitude, this can be removed so that only frequency variations appear. *(1)*

A disadvantage of FM is that it requires a more complicated demodulator than AM because the demodulator has to track the variations of frequency of the carrier rather than just the variations of amplitude. *(1)*

FM requires a greater bandwidth than AM since to transmit most of the modulating signal requires an FM bandwidth $= 2\left(\Delta f + f_m\right)$. AM requires a bandwidth = $2f_m$. *(1)*

8 a) As each tooth passes near the sensor, the change in magnetic effect is detected by the Hall effect sensor. *(1)*

This outputs a pulse each time a tooth passes it. *(1)*

b) The wheel has 7 teeth and therefore the output pulses from the Hall effect sensor need to be divided by 7. *(1)*

This can be achieved using a modulo-7 counter. *(1)*

9 a) If V_{th} for the MOSFET is 1.5 V, the pd across the 10 MΩ resistor must be < 1.5 V. *(1)*

$$I_D = \frac{V}{R} = \frac{1.5}{10\times10^6} = 0.15\times10^{-6}\ \text{A} = 0.15\ \mu\text{A} \quad (1)$$

b) Swap the positions of the photodiode and the 10 MΩ resistor. *(1)*

10 a) 4 cycles occur in 0.0091 s therefore f = 440 Hz. *(1)*

Sampling frequency is $\frac{1}{0.5\times10^{-3}} = 2000$ Hz. This is more than twice the frequency being sampled so is adequate. *(1)*

b) Peak-to-peak noise signal is about 0.1 V. *(1)*

c) $b = \log_2\left(\frac{V_{signal+noise}}{V_{noise}}\right) = \log_2\left(\frac{1.9}{0.1}\right) = 4\ (.2)$ bits *(1)*

d) When an analogue signal is sampled, the digital value given to the signal can only be the nearest digital value. *(1)*

This means that there is an error in some of the values and this is called the quantisation error. The more digital levels there are, the smaller this error will be but it will always be present to an extent. *(1)*

11 a) Gain $= \frac{-R_f}{R_{in}} = -\frac{820 \times 10^3}{4.7 \times 10^3} = 174$ *(1)*

b) Bandwidth $= \frac{3.0 \times 10^6}{174} = 17 \times 10^3$ Hz *(1)*

This is just about adequate for an audio amplifier. *(1)*

Because a bandwidth of about 20 kHz is needed although 17 kHz will probably not be noticeable. *(1)*

c) The power supply is ±12 V and an ideal op amp will allow the output to swing between the power supply voltages.

The maximum input is $\frac{12}{174}$ = 0.07 V. *(1)*

12 a) The output of an AND gate is 1 if both of the inputs are 1. *(1)*

b) When Q_1 and Q_3 are first both 1, i.e. at the count 1010 (10), the output of the AND gate will go high. *(1)*

This will put a 1 on the reset pin. *(1)*

So the output will immediately go to 0000 and it will be a BCD counter. *(1)*

c) Q_1 and Q_2 are first high on count 0110 (6).

This means the counter will reset and will count only from 0000 to 0101 and the counter will now be a modulo-6 counter. *(1)*

13 a) Multiplexing is when several different signals are sent down the same transmission path at the same time. *(1)*

b) The different transmission media used in a transmission path could be:

- optical fibre using light as the carrier wave
- radio waves and microwaves
- copper wires. *(3)*

c) Time-division multiplexing is when signals are chopped up into small segments and single segments from each signal are sent sequentially down the transmission path. *(1)*

Analogue signals will be sampled and each sample from each call will be put together into a frame and sent down the transmission path. *(1)*

Digital signals will be split into segments and the segments from different sources put together in a frame. *(1)*

At the receiving end, the samples will be sorted and routed to the separate receivers and the analogue signal reconstructed. *(1)*

14 Long waves are ground waves and so are diffracted around the Earth. *(1)*

A powerful transmitter enables these waves to have a very large local coverage. *(1)*

Long-wave transmissions are AM but FM can only be used for higher frequencies and shorter wavelengths. *(1)*

As FM transmissions are space waves they can only be used for 'line of sight' communications and so can only be received in an area local to the transmitter. *(1)*

Photo credits

p.1 © Tony Cordoza/Alamy Stock Photo; **p.28 right** © PixieMe/Shutterstock.com **p.47** © Iain McGillivray/Shutterstock.com; **p.48** © Mike Goldwater/Alamy Stock Photo