Chapter 1 Discrete semiconductor devices

1.1 MOSFETs

Learning objectives:

→ State the terminals of a MOSFET.
→ Define the threshold voltage of a MOSFET.
→ Explain how the current through a MOSFET is controlled.

MOSFET characteristics

A metal-oxide semiconducting field-effect transistor (MOSFET) is a three-terminal device that can be used to amplify electronic signals or to switch other devices on and off.

The circuit symbol for the most commonly used MOSFET, a negative-channel or **n-channel enhancement MOSFET**, is shown in Figure 1. Its three terminals are called the source, the drain, and the gate.

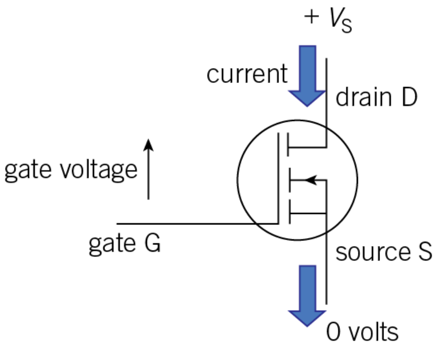


Figure 1 *The n-channel enhancement MOSFET*

When it is in use, the drain is at a positive potential relative to the source. When the gate is made sufficiently positive relative to the source, a conducting channel opens between the source and the drain, and so a current passes through the device from the drain to the source, as shown in Figure 1.

The **threshold voltage** V_{Th} of the transistor is the minimum potential difference (pd) needed to form the conducting channel between the drain and the source. The transistor is ON when the gate–source pd, V_{GS}, is greater than the threshold voltage, V_{Th}. When V_{GS} is less than V_{Th}, the transistor is OFF because there is no effective conducting channel between the drain and the source.

Figure 2 shows how the drain–source current, I_D, depends on the pd between the drain and the source, V_{DS}, for different values of V_{GS}.

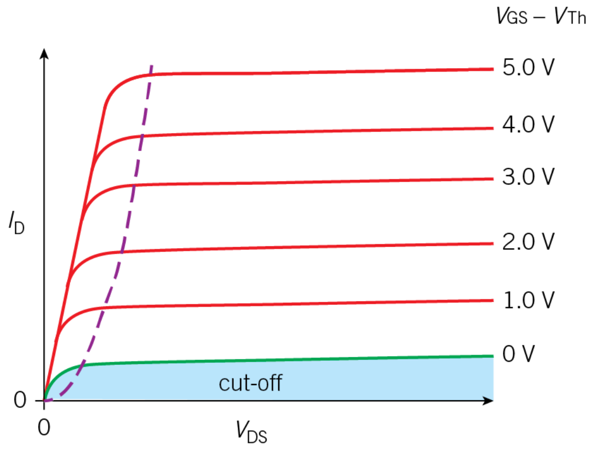


Figure 2 *Graph of* I_D *against* V_{DS}

The curves all have the same characteristic shape:

- For $V_{GS} - V_{Th} < 0$, the drain current is zero and the transistor is OFF. Because there is no conducting channel between the drain and the source, the drain is **cut off** from the source.
- For any constant value of $V_{GS} - V_{Th} > 0$, increasing the drain–source voltage, V_{DS}, from zero 'enhances' the conduction channel and therefore increases the drain current, I_D, linearly at first. Then the rate of increase decreases, and I_D becomes constant. The drain current is then said to be **saturated**. The dashed line in Figure 2 shows where saturation starts for each characteristic curve.
- For a constant drain–source pd, when the transistor is ON and the drain current is saturated, a change of the gate–source pd ΔV_{GS} causes the saturation current to change in proportion to ΔV_{GS}.
- The input resistance of a MOSFET (i.e., the resistance between the gate and the source) is very high. Therefore the current between the gate and the source is negligible.

MOSFETs in use

An n-channel MOSFET can be used to switch a relay on or off, as shown in Figure 3, or to switch a high-current device such as a motor on or off.

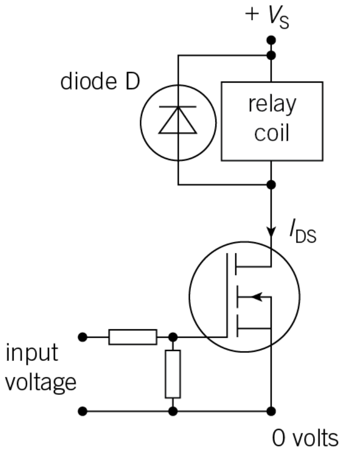


Figure 3 *A MOSFET in use*

The drain is connected in series with the relay coil which is connected to the positive terminal, $+V_S$, of the supply voltage, as shown in Figure 3.

- When the input voltage is zero, the transistor is OFF, so the drain current is zero. Therefore, the relay coil is not magnetised, and the relay is in its unactivated (i.e., normal) state.
- When the input voltage is made sufficiently positive, the transistor is switched ON, and the relay coil is magnetised. So the relay is activated, and its switch closes if it is a normally open (NO) relay, or its switch opens if it is a normally closed (NC) relay.

Notes

1 The resistance of the device in series with the drain must be small enough so that the current through the device is large enough to operate the device. However, if its resistance is too small, the current would be too high, and the heating effect of the current in the device would destroy the device or the transistor.

2 A reverse-biased diode must be connected across an electromagnetic device such as a motor or a relay. This is to short-circuit the emf induced in the device when the current is switched on or off, which would otherwise destroy the transistor.

3 The very high input resistance of a MOSFET means that very little current is drawn from any device connected to the MOSFET input. For this reason, a logic gate can be used to switch a MOSFET on and off. The logic gate is connected to the input via a potential divider consisting of two resistors, as shown in Figure 3. The potential divider ensures the MOSFET gate is at 0 V when the output of the logic gate is 0.

4 MOSFETS in combination are used to make logic gates in integrated circuits. For example, two MOSFETS in parallel connected to $+V_S$ via the same resistor can act as a simple NOR gate. A positive voltage as a logic 1 signal applied to either MOSFET gate causes the output pd to become logic 0. See Figure 4.

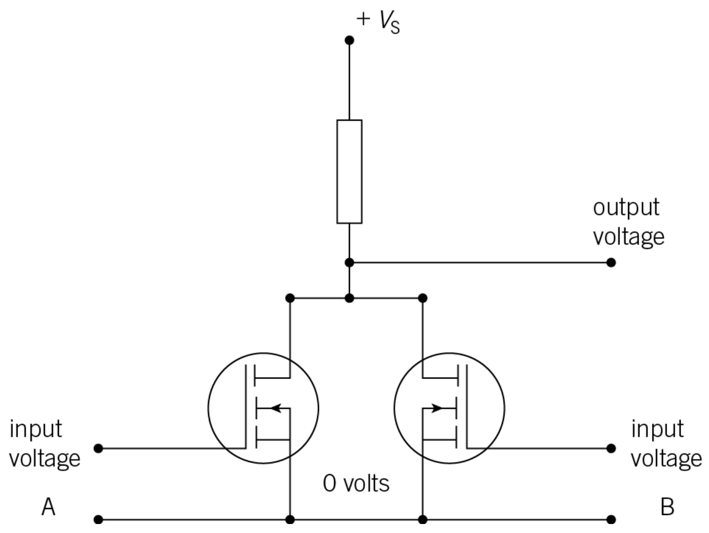


Figure 4 *A logic NOR gate*

MOSFET limits

The region between the source and the drain is not a perfect insulator. A very small leakage current of the order of microamperes passes through the transistor between the drain and the source. Although the leakage current increases when the temperature increases, this leakage is negligible compared with the current when V_{GS} is greater than V_{Th}. The leakage current when $V_{GS} = V_{Th}$ is called the **drain–source leakage current**, I_{DSS}.

The input gate resistance of a MOSFET is extremely high because the gate is separated from the semiconducting material of the MOSFET by an insulating layer of a dielectric or a metal oxide. When the metal gate is made positive relative to the source, the insulating layer prevents electrons from passing from the semiconducting material to the gate. However, if the gate voltage is greater than a particular value called the **breakdown voltage**, the strong electric field across the insulating layer destroys the insulating property of the layer.

Link

Semiconductors were looked at in Topic 12.1, Current and charge, in Year 1 of the *AQA Physics* student book.

How a MOSFET works

Figure 5 shows the design of an n-channel enhancement MOSFET. The gate is a metal tab separated by an insulating layer from a block of p-type semiconductor in which the drain and the source are embedded as smaller separate regions of n-type semiconductor. The p-type semiconductor is connected internally to the source.

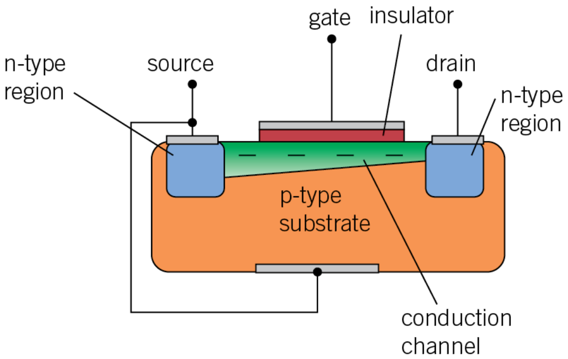


Figure 5 *Inside a MOSFET*

- When $V_{GS} = 0$, electrons from the n-type material near the p–n boundary move across the boundary, and ‘holes’ from the p-type material move across in the opposite direction. As a result, the n-type material loses electrons and becomes positive, whilst the p-type material loses holes and becomes negative. The transfer process stops when the p-type becomes sufficiently negative and the n-type sufficiently positive to stop further transfers. The material between the source and the drain is therefore unable to conduct when a pd is applied between them.

- When $V_{GS} > 0$, electrons are attracted into the p-type semiconductor region beneath the gate where they accumulate as the gate is made more positive. When $V_{GS} = V_{Th}$, then the electron density beneath the gate is equal to the hole density, and a conducting channel opens between the drain and the source.

QUESTION: Describe the ratio of conduction electrons to holes beneath the gate when $V_{GS} > V_{Th}$.

More about semiconductors

- n-type and p-type semiconductors are intrinsic semiconductors such as silicon with a small percentage of a different type of atom added.
- n-type semiconductors contain ‘donor’ atoms that have more electrons per atom than silicon atoms. When the donor atoms form bonds in the semiconductor, the surplus electrons become conduction electrons.
- p-type semiconductors contain ‘acceptor’ atoms that have fewer electrons per atom than silicon atoms. When the acceptor atoms form bonds in the semiconductor, they form bonds with ‘holes’ due to the missing electrons.

Summary questions

1 **a** An n-channel enhancement MOSFET has a threshold voltage of 0.85 V. Explain what is meant by threshold voltage in this context.

b Describe how the drain–source current in the MOSFET in **a** changes if the drain–source pd is increased from zero and the gate–source pd is such that the MOSFET conducts and is constant.

2 **a** Draw a circuit diagram including an n-channel MOSFET and a variable potential divider that may be used to switch on and control the speed of a low-voltage dc electric motor. Include a diode in your circuit across the motor.

b **i** When the motor is running, describe and explain how its speed may be reduced using your circuit.

ii Explain the purpose of the diode in the circuit.

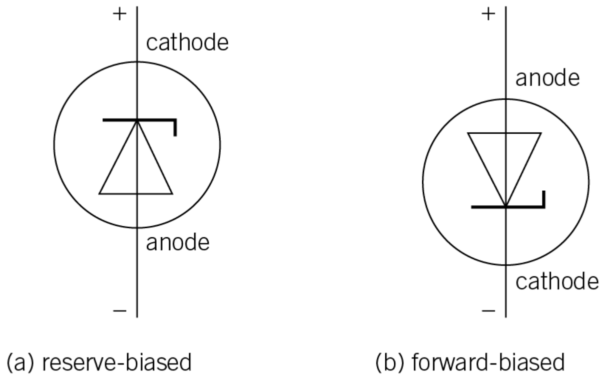
1.2 Zener diodes

Learning objectives:

→ Distinguish a zener diode from an ordinary diode.
→ Define the breakdown voltage of a zener diode.
→ Explain what is meant by a constant-voltage source.

Zener diode characteristics

When a semiconductor diode is reverse-biased and the pd across it is increased, its resistance is very large. The current through it is negligible until a particular voltage is reached, when it then breaks down and conducts. A zener diode is a semiconductor diode that is designed and manufactured so that when it is reverse-biased, it breaks down and conducts at a specific pd that depends on its design. The pd at which it breaks down is called its **breakdown voltage**. The pd across a diode changes very little when it conducts, as long as the diode current does not overheat and destroy the diode. Figure 1 shows the symbol for a zener diode. The terminal that is positive when it is forward-biased is called the **anode**; the other terminal is called the **cathode**.



(a) reserve-biased (b) forward-biased

Figure 1 *The zener diode*

In reverse-biased mode in series with a resistor, a zener diode can therefore be used in applications where a pd is required that stays constant when the current changes. Ordinary semiconductor diodes are not designed to break down, and their breakdown voltages are generally too high for most applications.

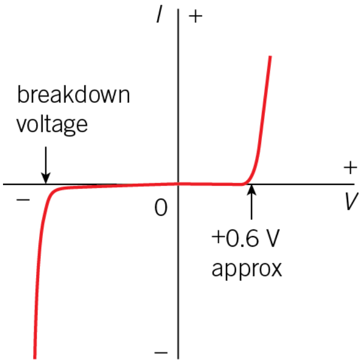


Figure 2 *Breakdown voltage*

Figure 2 shows how the current through a diode varies with pd when it is forward-biased and when it is reverse-biased. The current becomes very large when the pd reaches 0.6 V in the forward direction or when the breakdown voltage is reached in the reverse direction. In both cases, the line changes from horizontal to vertical over a small range of pd. In the reverse direction, the minimum current at the breakdown voltage is called the **minimum operating current** (typically 5–10 mA). All semiconductor diodes, including zener diodes, need to be connected in series with a resistor chosen to limit the current to a maximum value so that the heating effect of the current does not destroy the diode.

Zener diodes are specified in terms of their breakdown voltage and their **maximum power rating**. The current through a zener diode should therefore not exceed the maximum power rating divided by the breakdown voltage. For example, for a zener diode that has a breakdown voltage of 4.7 V and a maximum power rating of 100 mW, its current should not exceed 21 mA $\left(=\dfrac{100\,\text{mW}}{4.7\,\text{V}}\right)$.

Zener diodes in circuits

Zener diodes are used in constant-voltage sources where the source pd does not change when the current supplied changes. They can also be used to provide a **reference voltage** (i.e., a specified constant voltage), for example, a power supply with an output pd that does not change when the output current changes.

Figure 3 shows a 4.7 V, 100 mW zener diode in series with a resistor in a circuit that is used to supply a constant pd of 4.7 V to a car radio. The circuit is connected to a 12 V car battery.

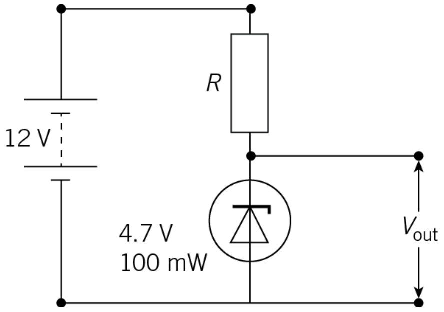


Figure 3 *Using a zener diode*

As explained above, the current through the zener diode must not exceed 21 mA. The resistance of *R* is chosen between two limits:

- low enough to allow enough current for the car radio to operate normally
- high enough to prevent the diode from overheating.

To calculate the *minimum* resistance of *R*, assume that the car radio is *not* connected to the circuit.

- The battery pd of 12 V is equal to the sum of the pd across resistor *R* and the pd across the zener diode. Therefore, the pd across *R* is equal to 7.3 V (= 12 V − 4.7 V).
- Because the current through the zener diode must not exceed 21 mA, the minimum resistance of *R* is equal to the pd across *R* divided by the maximum zener diode current = 350 Ω.

To calculate the *maximum* resistance of *R*, assume that the car radio *is* connected to the circuit.

- The zener diode pd stays at 4.7 V, and the pd across *R* stays at 7.3 V.
- Suppose that the car radio needs a current of 5 mA to operate normally and the minimum operating current for the diode is 8 mA. Therefore, the current in *R* needs to be at least 13 mA. So the maximum resistance of *R* must be 560 Ω $\left(=\dfrac{7.3\text{ V}}{13\text{ mA}}\right)$.

Taking the minimum and maximum together, then, the resistance of *R* must be between 350 Ω and 560 Ω.

Consider the effect of choosing different values of resistance *R* when the car radio is on:

R = 350 Ω, and pd across the car radio = the zener pd = 4.7 V:

- current through the resistor = 21 mA
- current through the car radio = 5 mA
- current through the zener diode = 21 mA − 5 mA = 16 mA.

R = 560 Ω, and pd across the car radio = the zener pd = 4.7 V:

- current through the resistor = 13 mA
- current through the car radio = 5 mA
- current through the zener diode = 13 mA − 5 mA = 8 mA.

For any value of resistance *R* between 350 Ω and 560 Ω, when the car radio is on, the current through the resistor is split between the zener diode and the car radio. Thus, the zener diode current is equal to the current through the resistor $\left(=\dfrac{7.3\text{ V}}{R}\right)$ minus 5.0 mA through the car radio.

Note

Suppose that the car radio was replaced by a new radio that needed a pd of 4.7 V and a current of 25 mA. The pd across the series resistor would still be 7.3 V, but the resistance range above would limit the resistor current to no more than 21 mA, so the radio would not operate normally. Using a series resistor with a resistance lower than 350 Ω would cause the zener diode to overheat when the car radio is off. The zener diode would therefore need to be replaced by one with a higher power rating.

Summary questions

1 **a** Draw a graph to show how the current through a diode varies with the pd across it.

b Explain what is meant by the breakdown voltage of a zener diode.

2 **a** For the zener diode in Figure 3, calculate the current in each component for a 4.7 V, 5.0 mA car radio for a 400 Ω series resistor. Assume the minimum operating current for the zener diode is 8 mA.

b Explain why the circuit would not function if a 2.0 kΩ series resistor was used.

c Design a circuit using a 6.0 V, 180 mW zener diode, a series resistor, and a 9.0 V battery to supply a pd of 6.0 V to a 20 mA, 6.0 V lamp. Assume the minimum operating current for the zener diode is 5 mA.

1.3 Photodiodes

Learning objectives:

→ Explain what is meant by the dark current of a photodiode.
→ Explain what is meant by the saturation current of a photodiode.
→ Describe how a photodiode is connected in a circuit when it is used to measure light intensity.

Photodiode characteristics

A photodiode is a diode that conducts when it is reverse-biased and light is incident on it. Its case has a window that allows light through to illuminate the semiconductor material inside. Figure 1 shows the symbol for a photodiode. Ordinary semiconductors do have a very small leakage current of the order of nanoamperes. The leakage current in a reverse-biased photodiode is of the order of milliamperes when light is incident on it. When it is reverse-biased, the photodiode is said to be in **photoconductive** mode, which means that a current passes through it when light is incident on it and it is reverse-biased.

Figure 1 *The symbol for a photodiode*

Figure 2 shows how the leakage current of a photodiode varies with the pd across the photodiode for different incident light intensities.

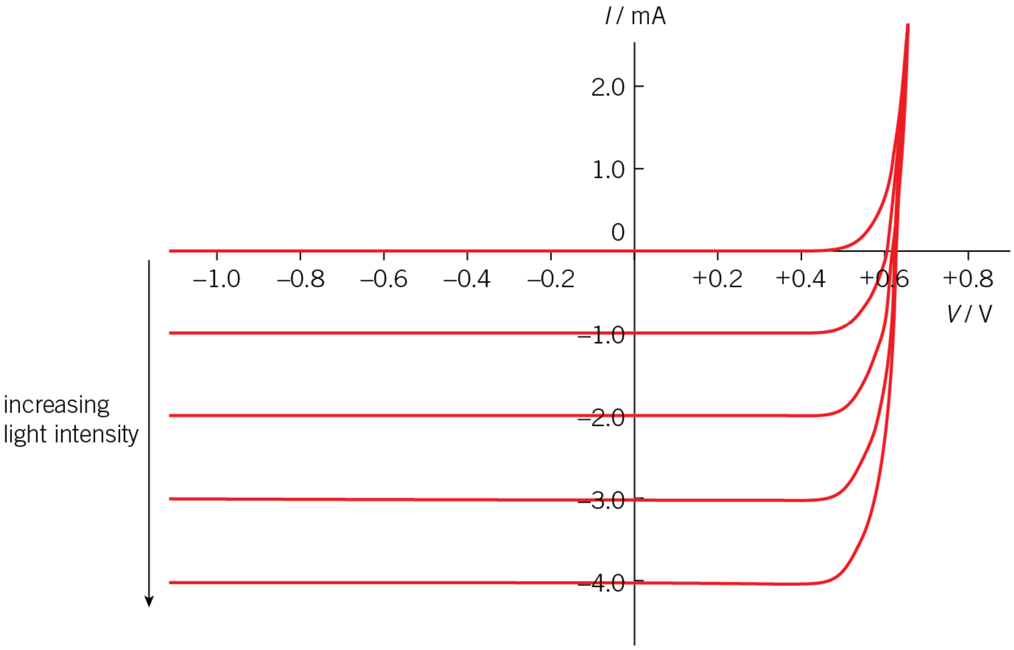


Figure 2 *I against V curves for a photodiode*

For zero light intensity:

- A photodiode behaves like a normal diode in that there is a leakage current of the order of microamperes or less when the diode is reverse-biased, so the pd across it is negative. The leakage current for zero light intensity is called the **dark current**.

For constant light intensity:

- When the pd is zero or negative so that the photodiode is reverse-biased, a leakage current of the order of milliamperes passes through the photodiode that does not change when the pd is made negative or more negative. The current in these conditions is called the **saturation current**.
- When the pd is increased from zero, the leakage current is the same as for zero pd until a pd of about 0.5 V is reached, when the leakage current decreases to zero at about 0.6 V. A relatively small increase of pd then causes the current to increase considerably in the forward direction.

The current versus pd graphs for various light intensities have the same characteristic shape with a saturation current that increases as the incident light intensity increases.

Photodiodes in use

Photodiodes respond to a wide range of wavelengths. For example, silicon photodiodes respond to wavelengths from 400 nm (blue light) to about 1000 nm (infrared radiation). Photodiodes are therefore used widely to detect pulses of infrared radiation sent along optical fibres.

The **spectral response** of a silicon photodiode to visible light and infrared light is shown in Figure 3. This is measured for different wavelengths at the same intensity as the reverse-biased photodiode current for each wavelength. Note that the response is greatest at a wavelength just less than 900 nm which is in the infrared part of the electromagnetic spectrum. In comparison, the response to visible light of wavelength 500 nm is about half.

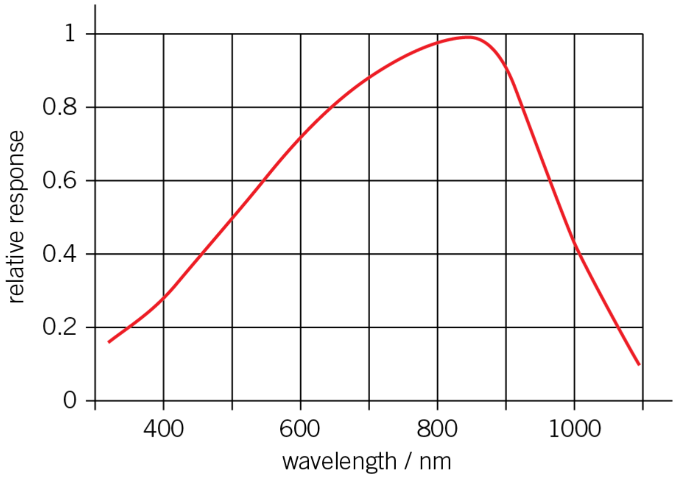


Figure 3 *The spectral response of a photodiode*

Figure 4 shows a reverse-biased photodiode in series with a resistor connected to a low-voltage power supply. The pd across the photodiode depends on the incident light intensity, so it gives an output pd V_{out} that depends on the incident light intensity on the photodiode.

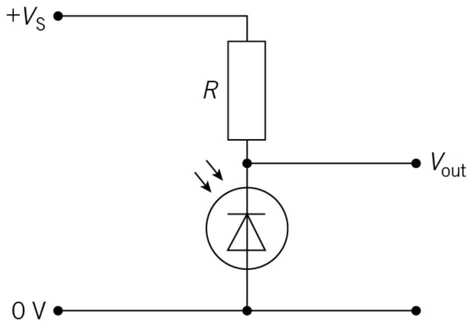


Figure 4 *A photodiode circuit*

- When the photodiode is in darkness, the current through the photodiode and the resistor is negligible, so there is no pd across the resistor. Hence, the pd across the photodiode is equal to V_S, where V_S is the supply pd. Therefore, the output pd, V_{out}, is equal to V_S.
- When light of constant intensity is directed at the photodiode, a constant current equal to the saturation current passes through the photodiode and the resistor. So the pd across the resistor is equal to IR, where R is the resistance of the resistor. Therefore, $V_{out} = V_S - IR$.

Figure 5 shows how the output pd varies with the saturation current for $V_S = 5.0$ V and $R = 2.5$ kΩ. The graph line, called the **load line**, shows that the pd decreases linearly as the saturation current increases.

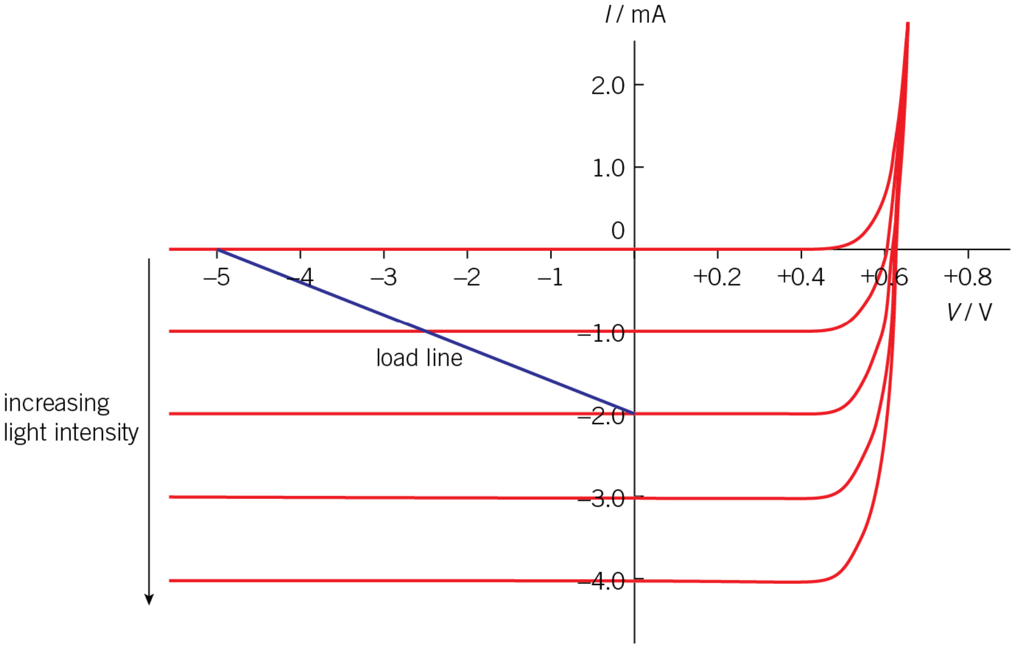


Figure 5 *A load line on the graph of I against V for a photodiode*

The light intensity characteristic graphs are also shown in Figure 5. The output pd for any given light intensity is the point where the load line intersects the relevant light intensity graph.

- The line intercepts the −*x*-axis where the current is zero, corresponding to $V_{out} = V_S$. Note that the *x*-intercept value is $-V_S$ because the photodiode is reverse-biased.

- The line intercepts the −*y*-axis where $V_{out} = 0$. The current at the *y*-intercept is therefore given by $V_{out} = V_S - IR$, so $I = \frac{V_S}{R}$.

The output voltage therefore changes from V_S to 0 as the light intensity increases from zero to a level where the saturation current is equal to $\frac{V_S}{R}$. The output voltage could be inverted using an inverting amplifier so that a change of light intensity from 0 to the maximum value (corresponding to a saturation current of $\frac{V_S}{R}$) causes the output voltage of the operational amplifier to increase from 0 to V_S.

Link

For more about the operational amplifier, see Topic 3.2, Ideal operational amplifiers, and Topic 3.3, Operational amplifier circuits.

Worked example

In a circuit similar to the one in Figure 3, a 5.0 V power supply and a 1.0 kΩ resistor *R* are used. The light intensity is adjusted until the output pd is 3.5 V. Calculate the photodiode current at this output pd.

Solution

The pd across the resistor = 5.0 − 3.5 V = 1.5 V

$$\text{Photodiode current} = \frac{\text{pd across resistor } R}{\text{resistance of } R} = \frac{1.5\ \text{V}}{1.0\ \text{k}\Omega} = 1.5\ \text{mA}$$

Photodiode applications

The photodiode in the circuit in Figure 4 can be used in a variety of applications:

Monitoring changes in light intensity

Constant illumination gives an output pd that is constant. But if the light intensity increases, the output pd decreases. If a voltmeter or a data logger is across the output, the photodiode can be used as a light meter to monitor changes in light intensity.

Converting light pulses into electronic pulses

A brief pulse of light directed at a photodiode causes the output pd to decrease briefly from V_S to a much lower voltage. Thus a photodiode can be used to convert light pulses into electronic pulses. Because photodiodes respond rapidly to changes in light intensity, they can be used to convert optical pulses carrying data into electronic pulses.

Detecting atomic particles when used with a scintillator

A scintillator is a substance that emits light when atoms within it de-excite after being excited by a high-energy atomic particle that enters it. A reverse-biased photodiode as in Figure 3 can be used to detect the light from individual scintillation events by fixing the scintillator to the window of the photodiode. The light from a scintillation event decays within microseconds or less, and because photodiodes have a fast response time, a scintillator–photodiode detector can be used to count individual high-energy atomic particles entering the scintillator. The Charged Particle Monitor (CPM) on ASTROSAT, a satellite carrying six different instruments, was launched by the Indian Space Research organisation in 2015. The CPM is a 1 × 1 × 1 cm caesium iodide crystal fixed to the window of a photodiode. In addition to its capacity for counting individual atomic particles at rates from slow to fast, the CPM has a relatively low mass (2 kg) and low power requirement (2 W).

Summary questions

1 **a** For a reverse-biased photodiode, explain what is meant by:
  i its dark current
  ii the saturation current.
  b Draw a circuit diagram of a photodiode circuit that can be used to monitor changing light intensity.
  c Describe the changes that take place in your circuit when the photodiode is illuminated after being in darkness.

2 In a circuit similar to the one in Figure 4, a 6.0 V power supply and a 3 kΩ resistor are used.
  a The light intensity is adjusted until the output pd is 4.5 V. Calculate the photodiode current at this output pd.
  b The light intensity is reduced by 75%. Calculate the output pd at this reduced light intensity.

www.oxfordsecondary.co.uk/acknowledgements

1.4 Hall effect sensors

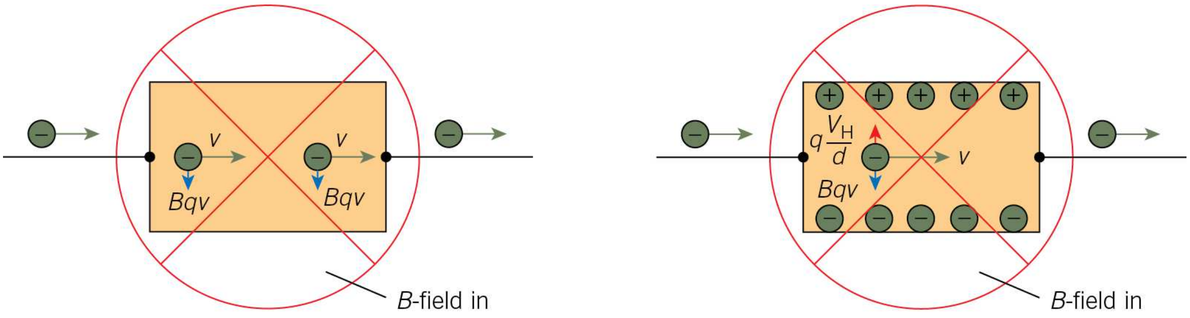
Learning objectives:

→ State the Hall voltage of a Hall effect sensor.
→ State how the Hall voltage depends on magnetic flux density.
→ Describe the uses of a Hall sensor.

The Hall effect

A Hall effect sensor is designed to detect, monitor, and measure magnetic fields. When a current is passed through a sample of a semiconductor or a conductor and a magnetic field is applied perpendicular to the direction of current, a pd is created between opposite sides of the sample. This is called the **Hall effect**. It occurs because the charge carriers moving through the sample are forced to one edge of the sample by the force of the magnetic field on them.

The pd created between the opposite sides of the sample is called the Hall voltage, V_H. It creates an electric field that prevents further charge carriers entering the sample from being forced down by the magnetic field. As shown in Figure 1(b), each further charge carrier entering the sample experiences an upward electric field force $q\frac{V_H}{d}$ that is equal and opposite to the magnetic force Bqv.



(a) initially

(b) finally

Figure 1 *The Hall voltage*

The pd across the sample, V_H, is proportional to:

- the current I through the sample, and
- the component of the magnetic flux density B perpendicular to the sample, as shown in Figure 1.

A Hall sensor consists of a rectangular sample of an n-type or p-type semiconductor in a circuit in which the current through the sample is constant. This ensures that the Hall voltage depends only on the magnetic flux density. When a magnetic field is applied perpendicular to the sample as in Figure 2, the Hall voltage created is amplified to provide the sensor output.

Link

Moving charges were looked at in Topic 24.2, Moving charges in a magnetic field, in Year 2 of the *AQA Physics* student book.

Hall effect uses

1 A Hall sensor may be used as **a proximity sensor** next to a rotating disc or wheel that has a magnet fixed near its perimeter, as shown in Figure 2. A voltage pulse is produced by the sensor each time the magnet moves past the sensor.

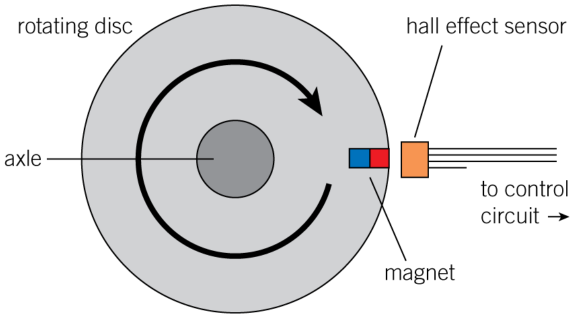


Figure 2 *A proximity sensor*

By recording:

- the total number of pulses, the number of turns made by the disc can be measured
- the number of pulses produced per second, the angular speed of the disc can be measured.

The electronic vehicle **tachometer** now in widespread use makes use of a Hall sensor fitted to an engine crankshaft. The sensor supplies voltage pulses to an on-board electronic system that records how far the vehicle travels and its speed at any time.

2 Hall sensors are also used to **monitor attitude** or orientation of objects in a magnetic field relative to the field direction. An attitude sensor is a device that monitors or measures its orientation relative to an external frame of reference. A Hall effect attitude device contains three Hall sensors in a single chip. The sensors are orientated in the *x*, *y*, and *z* directions relative to the chip, as shown in Figure 3. Each of the three sensors gives a voltage determined by the *x*, *y*, and *z* components respectively of the local magnetic flux density.

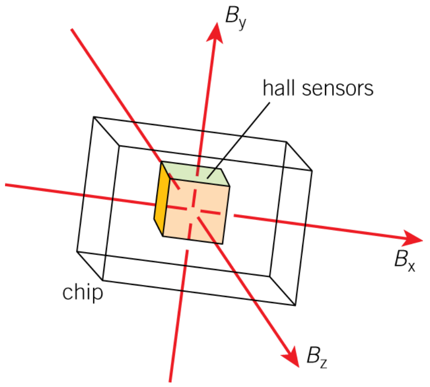


Figure 3 *An attitude sensor*

From these three components, B_x, B_y, and B_z:

- the magnitude of the local flux density B can be determined because $B^2 = B_x^2 + B_y^2 + B_z^2$ (from Pythagoras' theorem), and
- the orientation of the sensors relative to the local magnetic field can be determined by using appropriate trigonometrical rules.

Such devices fitted to satellites are used to determine the orientation of a satellite by comparing the orientation of the device relative to the local magnetic field with the known direction of the local magnetic field relative to the Earth.

Link

Magnetic flux density was looked at in Topic 24.1, Current-carrying conductors in a magnetic field, in Year 2 of the *AQA Physics Level* student book.

Summary questions

1 **a** Explain why the current in a Hall effect sensor must be constant.

b A Hall effect sensor is placed near a rotating disc that has two embedded magnets near its perimeter on opposite sides of the centre of the disc. When the disc is rotating at a constant frequency, the sensor produces 35 pulses per second. Calculate the time for one rotation of the disc.

2 **a** State the function of an attitude sensor.

b When a Hall effect attitude device on the ground is orientated so that its *z*-axis is vertical, it gives *x*, *y*, *z* readings of 10 μT, 15 μT, and 55 μT respectively.

i Calculate the magnitude of the magnetic flux density *B*.

ii The sensors were calibrated with a current of 100 ±2 μA in each sensor. Estimate the uncertainty in the magnitude of the magnetic flux density calculated in **a**.