# AS & A-Level Physics Lecture Notes

*(Draft)*

BY

*Yuhao Yang*

Shanghai Easyday Education

April 27, 2022

# Practical Issues

## Last Update: April 27, 2022

## Contacts

**Email**: colin-young@live.com

The latest update can be found via: https://github.com/yuhao-yang-cy/a2physics

## About the Lecture Notes

This is a set of very concise lecture notes written for CIE A-level Physics (syllabus code 9702). Presumably the target audience of the notes are students studying the relevant course.

I have been teaching A-Level courses for many years, and I have been planning to typeset my collection of handwritten notes with LaTeX not long after my teaching career started. The project never appeared to get any close to completion for many years, so I found joy when I eventually finished this project.

These notes are supposed to be self-contained. I believe I have done my best to make the lecture notes reflect the spirit of the syllabus set by the Cambridge International Examination Board. Apart from the essential derivations and explanations, I also included a handful of worked examples and problem sets, so that you might get some rough idea about the styles of questions that you might encounter in the exams. If you are a student studying this course, I believe these notes could help you get well-prepared for the exams.

Throughout the notes, key concepts are marked red, key definitions and important formulas are boxed. But to be honest, the main reason that I wrote up these notes was not to serve any of my students, but just to give myself a goal. Since physics is such a rich and interesting subject, I cannot help sharing a small part of topics beyond the syllabus that I personally find interesting. If you see anything beyond the syllabus, they usually show up in the footnotes and are labelled with an asterisk sign ($*$).

Note that there will be an update for the syllabus since 2022. I am still working on the updates so that the notes are tailored for the new syllabus. I am glad to see some *introductory astronomy and cosmology* being added into the new syllabus, so you will see one new chapter coming in.

The contents of *electronics* and *telecommunication* have been removed from the new syllabus. As those future engineers might want to look at these topics, so I kept the two chapters in the new edition nonetheless.

Also very importantly, I am certain that there are tons of typos in the notes. If you spot any errors, please let me know.

## Literature

I borrow heavily from the following resources:

- Cambridge International AS and A Level Physics Coursebook, by *David Sang*, *Graham Jones*, *Richard Woodside* and *Gurinder Chadha*, Cambridge University Press
- International A Level Physics Revision Guide, by *Richard Woodside*, Hodder Education
- Longman Advanced Level Physics, by *Kwok Wai Loo*, Pearson Education South Asia
- Past Papers of Cambridge International A-Level Physics Examinations
- HyperPhysics Website: `http://hyperphysics.phy-astr.gsu.edu/hbase/index.html`
- Wikipedia Website: `https://en.wikipedia.org`

## Copyright

# Contents


**CHAPTER 1** **Circular Motion** — PAGE 1

1.1 angular quantities 1

angular displacement - 1 / angular velocity - 1 / uniform circular motion - 2 / centripetal acceleration - 3

1.2 centripetal force 4

**CHAPTER 2** **Gravitational Fields** — PAGE 8

2.1 gravitational forces 8

Newton's law of gravitation - 8 / planetary motion - 9 / apparent weight - 10

2.2 gravitational fields 12

2.3 gravitational field strength 12

gravitational field strength - 12 / acceleration of free fall - 13 / gravitational field lines - 14

2.4 gravitational potential & potential energy 14

potential energy - 14 / gravitational potential energy - 15 / gravitational potential - 18

**CHAPTER 3** **Oscillation** — PAGE 20

3.1 oscillatory motion 20

amplitude, period, frequency - 20 / phase angle - 20 / acceleration & restoring force - 21

3.2 simple harmonic oscillation 22

equation of motion - 22 / examples of simple harmonics - 23 / velocity & acceleration - 25 / vibrational energy - 28

3.3 damped oscillations 29

light damping - 29 / heavy damping - 30 / critical damping - 30

3.4 forced oscillations 31

free & forced oscillation - 31 / resonance - 31 / damping & resonance - 32

CHAPTER 4 **Ideal Gases** PAGE 33

4.1 gas molecules 33

motion of gas particles - 33 / amount of molecules - 33 / pressure (qualitative view) - 34

4.2 ideal gas 35

ideal gas equation - 35 / empirical laws - 37

4.3 kinetic theory of ideal gases 39

assumptions of ideal gas model - 39 / pressure (quantitative view) - 40 / kinetic energy - 41

CHAPTER 5 **Thermodynamics** PAGE 44

5.1 thermal physics basics 44

temperature scales - 44 / kinetic theory of matter - 44 / specific latent heat - 45 / specific heat capacity - 46

5.2 internal energy 47

kinetic energy - 48 / potential energy - 48 / internal energy of ideal gas - 48 / change of states - 49 / first law of thermodynamics - 51

5.3 temperature 54

temperature & thermal energy - 54 / absolute zero - 54 / thermometer - 55

CHAPTER 6 **Electrostatics** PAGE 57

6.1 electric forces 57

Coulomb's law - 57 / electric fields - 58

6.2 electric field strength 58

electric field strength - 58 / electric field lines - 60

6.3 potential & potential energy 61

electric potential energy - 61 / electric potential - 63 / equipotential lines - 65

6.4 further discussions on electric fields 67

comparison with gravitational fields - 67 / electric field inside conductors - 68 / field strength & potential - 69

CHAPTER 7 **Capacitors** PAGE 74

7.1 capacitors: an introduction 74

capacitors - 74 / mutual capacitance - 75 / self-capacitance - 75

7.2 energy stored in a capacitor 77

7.3 capacitor networks 78

capacitors in parallel - 78 / capacitors in series - 79 / capacitor networks - 80 / capacitors & resistors - 81

7.4 charging & discharging capacitors 82

charging phase - 82 / discharging phase - 82 / time constant - 83

CHAPTER 8 **Magnetic Fields** PAGE 84

8.1 magnetism 84

magnets - 84 / magnetic field due to currents - 85 / solenoids & electromagnets - 86

8.2 magnetic force on current-carrying conductor 87

magnetic force on current-carrying conductor - 87 / magnetic flux density - 88

8.3 magnetic force on charged particles 91

magnetic force on charged particles - 91 / charged particle in uniform magnetic fields - 93 / mass spectrometer - 94 / cyclotron - 95 / velocity selector - 97

8.4 Hall effect 98

CHAPTER 9 **Electromagnetic Induction** PAGE 100

9.1 magnetic flux 100

9.2 laws of electromagnetic induction 101

Faraday's law - 101 / Lenz's law - 101

9.3 applications of electromagnetic induction 104

magnetic braking - 104 / the generator - 106 / electromagnetic gun - 107

CHAPTER 10 **Alternating Currents** PAGE 108

10.1 sinusoidal a.c. 108

sinusoidal waveform - 108 / power - 109 / r.m.s. current & r.m.s. voltage - 110 / measurement of a.c. with an oscilloscope - 111

10.2 power supply systems 112

high-voltage transmission - 112 / transformers - 113 / rectifiers - 114 / smoothing - 116

CHAPTER 11 **Quantum Physics Basics** PAGE 118

11.1 classical theories 118

particle models - 118 / wave models - 119 / history of light - 119

11.2 photon theory 120

photoelectric effect - 120 / photon theory - 122 / photoelectric effect explained - 123 / photon momentum - 125

11.3 wave-particle duality 127

matter waves - 127 / electron diffraction - 128 / wave-particle duality - 129

11.4 quantisation of electron energy levels 129

hydrogen atom (∗) - 130 / electron energy levels - 131 / atomic spectrum - 131

11.5 band theory (∗) 136

energy bands - 136 / band theory & electrical conductivity - 137

CHAPTER 12 **Nuclear Physics** PAGE 140

12.1 nuclear energy 140

mass-energy equivalence - 140 / binding energy - 141 / nuclear stability - 142 / fusion - 143 / fission - 144 / energy release from nuclear reactions - 145

12.2 radioactive decay 147

radioactive decay - 147 / decay equations - 148 / half-life - 149 / measurement of radioactive decay - 151

CHAPTER 13 **Medical Imaging** PAGE 154

13.1 X-ray imaging 154

nature of X-rays - 154 / production of X-rays - 155 / X-ray spectrum - 156 / X-ray attenuation - 158 / X-ray imaging techniques - 160 / CT scans - 161

13.2 ultrasonic scans 164

nature of ultrasound - 164 / production of ultrasounds - 165 / attenuation of ultrasounds - 166 / reflected & transmitted intensity at boundaries of structures - 166 / ultrasound imaging - 168

13.3 MRI (∗) 169

nuclear magnetic resonance - 169 / magnetic resonance imaging - 172

13.4 PET 173

positron - 173 / electron–positron annihilation - 173 / positron emission tomography (PET) - 174

CHAPTER 14 **Astrophysics** PAGE 176

14.1 astronomical distances 176

unit of length in astronomy - 176 / luminosity - 176 / radiant flux intensity - 177 / standard candles - 177

14.2 spectrum of stars 179

black body radiation - 179 / surface temperature - 180 / stellar radii - 181

14.3 cosmology 182

quick review: absorption spectrum - 182 / quick review: Doppler effect - 182 / redshift - 183 / recession of galaxies - 183 / Big Bang theory - 184 / fate of the universe (∗) - 187

CHAPTER 15 **Electronics (∗)** PAGE 188

15.1 sensors 188

thermistor - 188 / LDR - 189 / strain gauge - 189 / piezo-electric transducer - 190

15.2 processors 191

potential divider - 191 / op-amp: basics - 192 / voltage comparator - 194 / negative feedback - 195 / inverting amplifier - 197 / non-inverting amplifier - 198

15.3 output devices 202

LED - 202 / relay - 202 / calibrated meters - 203

15.4 practical electronic circuits 204

automatic illumination system - 204 / multi-range voltmeter - 205 / digital signal regenerator - 206 / summing amplifier - 207

CHAPTER 16 **Telecommunication (∗)** PAGE 209

16.1 modulation 209

amplitude modulation - 211 / frequency modulation - 214 / comparison between AM and FM - 215

16.2 digital transmission 215

analogue & digital signals - 215 / regeneration - 216 / binary numbers - 217 / analogue-to-digital conversion - 218

16.3 communication channels 220

wire pairs - 220 / coaxial cables - 220 / optic fibres - 221 / radio waves - 221 / microwaves - 223

16.4 the decibel scale 224

# Chapter 1

# Circular Motion

## 1.1 angular quantities

movement or rotation of an object along a circular path is called **circular motion**

to describe a circular motion, we can use *angular quantities*, which turn out to be more useful than linear displacement , linear velocity , etc.

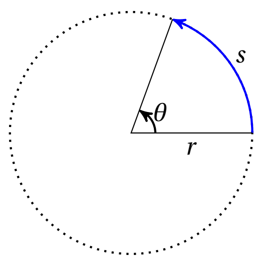
### 1.1.1 angular displacement

**angular displacement** is angle swiped out by object moving along circular

- unit: $[\theta] =$ rad (natural unit of measurement for angles)

  conversion rule: $2\pi \text{ rad} = 360°$
- if two radii form an angle of $\theta$, then length of arc: $s = r\theta$

  two radii subtending an arc of same length as radius form an angle of one **radian**




### 1.1.2 angular velocity

angular velocity describes how fast an object moves along a circular path

**angular velocity** is defined as angular displacement swiped out per unit time: $\boxed{\omega = \frac{\Delta\theta}{\Delta t}}$

- unit of: $[\omega] =$ rad $s^{-1}$, also in radian measures
- angular velocity is a *vector* quantity

  this vector points in a direction normal to the plane of circular motion but in A-level course, we treat angular velocity as if it is a scalar

angular velocity and angular speed may be considered to be the same idea

➢ relation with linear velocity

in interval $\Delta t$, distance moved along arc $\Delta s = v\Delta t = r\Delta\theta \quad\Rightarrow\quad \omega = \dfrac{\Delta\theta}{\Delta t} = \dfrac{v}{r} \quad\Rightarrow\quad \boxed{v = \omega r}$

this relation between linear speed and angular speed holds at any instant

### 1.1.3 uniform circular motion

when studying linear motion, we started from motion with constant velocity $v$

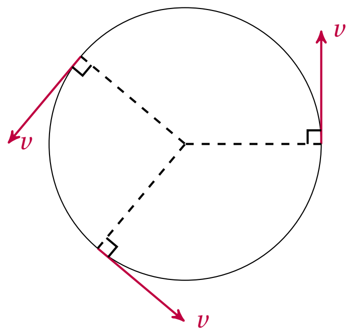
consider the simplest possible circular motion $\longrightarrow$ circular motion with constant $\omega$

analogy with linear motion with constant $v$

uniform linear motion: $s = vt$

displacement $s \leftrightarrow \theta$, velocity $v \leftrightarrow \omega$

for uniform circular motion, one has: $\boxed{\theta = \omega t}$




➢ time taken for one complete revolution is called **period** $T$

in one $T$, angle swiped is $2\pi$, so $\boxed{\omega = \dfrac{2\pi}{T}}$

➢ uniform circular motion is still *accelerated* motion

speed is unchanged, but *velocity* is changing

direction of velocity always *tangential* to its path, so direction of velocity keeps changing

in general, any object moving along circular path is accelerating

**Example 1.1** An object undergoes a uniform motion around a circular track of radius 2.5 m in 40 s, what is its angular speed and linear speed?

✍ $$\omega = \frac{2\pi}{T} = \frac{2\pi}{40} \approx 0.157 \text{ rad s}^{-1} \qquad v = \omega r = 0.157 \times 2.5 \approx 0.39 \text{ m s}^{-1}$$ □

**Question 1.1** What is the angular velocity of the minute hand of a clock?

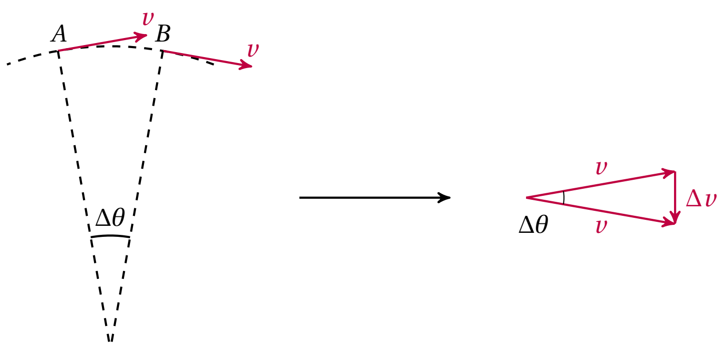
**Question 1.2** A spacecraft moves around the earth in a circular orbit. The spacecraft has a speed of 7200 m s$^{-1}$ at a height of 1300 km above the surface of the earth. Given that the radius of the earth is 6400 km. (a) What is the angular speed of this spacecraft? (b) What is its period?

#### 1.1.4 centripetal acceleration

> **centripetal acceleration** is the acceleration due to the change in direction of velocity vector, it points toward the centre of circular path

consider motion along a circular path from $A$ to $B$ with constant speed $v$

under small (infinitesimal) duration of time $\Delta t$[1]




change in velocity: $\Delta v = 2v\sin\dfrac{\Delta\theta}{2} \approx v\Delta\theta \quad$ (as $\Delta\theta \to 0$, $\sin\Delta\theta \approx \Delta\theta$)

acceleration: $a = \dfrac{\Delta v}{\Delta t} \approx v\dfrac{\Delta\theta}{\Delta t} = v\omega \quad$ (as $\omega = \dfrac{\Delta\theta}{\Delta t}$)

recall relation $v = \omega r$, we find centripetal acceleration: $\boxed{a_c = \dfrac{v^2}{r} = \omega^2 r}$

- direction of centripetal acceleration: always towards centre of circular path
- centripetal acceleration is only responsible for the change in *direction* of velocity

  change in *magnitude* of velocity will give rise to *tangential acceleration*

  this is related to *angular acceleration*[2] , which is beyond the syllabus

**Question 1.3** A racing car makes a 180° turn in 2.0 s. Assume the path is a semi-circle with a radius of 30 m and the car maintains a constant speed during the turn. (a) What is the angular velocity of the car? (b) What is the centripetal acceleration?

[1] A more rigorous derivation can be given by using differentiation techniques

[2] Angular acceleration is analogous to linear acceleration $\alpha$, defined as rate of change of angular velocity: $\alpha = \dfrac{d\omega}{dt} = \dfrac{d^2\theta}{dt^2}$ $(*)$. Similar to $v = \omega r = \dfrac{ds}{dt}$, the relation $a = \alpha r = \dfrac{dv}{dt}$ also holds.

## 1.2 centripetal force

circular motion must involve change in velocity, so object is not in equilibrium

there must be a *net force* on an object performing circular motion

> **centripetal force** ($F_c$) is the resultant force acting on an object moving along a circular path, and it is always directed towards centre of the circle

➢ centripetal force causes centripetal acceleration

using Newton's 2nd law: $\boxed{F_c = m\frac{v^2}{r} = m\omega^2 r}$

➢ $F_c$ is not a new force by nature, it can have a variety of origins

$F_c$ is a resultant of forces you learned before (weight, tension, contact force, friction, etc.)

➢ $F_c$ acts at right angle to direction of velocity

or equivalently, if $F_{\text{net}} \perp v$ and $F_{\text{net}}$ is of constant magnitude

then this net force provides centripetal force for circular motion

➢ effect of $F_c$: change *direction* of motion, or maintain circular orbits

to change *magnitude* of velocity, there requires a *tangential* component for the net force

again the idea of tangential force is beyond the syllabus

**Example 1.2** A rock is able to orbit around the earth near the earth's surface. Let's ignore air resistance for this question, so the rock is acted by weight only. Given that radius of the earth $R = 6400$ km. (a) What is the orbital speed of the rock? (b) What is the orbital period?

✍ weight of object provides centripetal force: $mg = \frac{mv^2}{R}$

orbital speed: $v = \sqrt{gR} = \sqrt{9.81 \times 6.4 \times 10^6} \approx 7.9 \times 10^3 \text{ m s}^{-1}$

period: $T = \frac{2\pi R}{v} = \frac{2\pi \times 6.4 \times 10^6}{7.9 \times 10^3} \approx 5.1 \times 10^3 \text{ s} \approx 85 \text{ min}$ □

**Example 1.3** A turntable can rotate freely about a vertical axis through its centre. A small object is placed on the turntable at distance $d = 40$ cm from the centre. The turntable is then set to rotate, and the angular speed of rotation is slowly increased. The coefficient of friction between the object and the turntable is $\mu = 0.30$. If the object does not slide off the turntable, find the maximum number of revolutions per minute.

✍ if object stays on turntable, friction provides the centripetal force required: $f = m\omega^2 d$

increasing $\omega$ requires greater friction to provide centripetal force

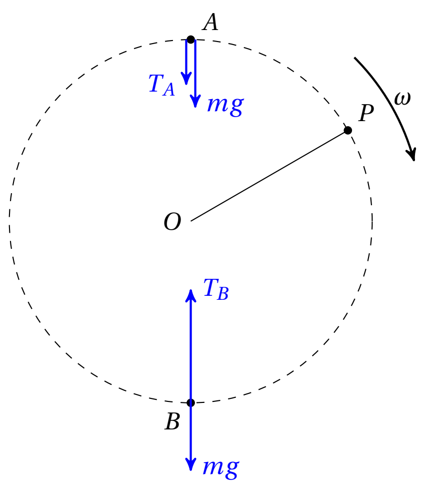
but maximum limiting friction possible is: $f_{\text{lim}} = \mu N = \mu mg$, therefore

$$f \le f_{\text{lim}} \quad \Rightarrow \quad m\omega^2 d \le \mu mg \quad \Rightarrow \quad \omega^2 \le \frac{\mu g}{d} \quad \Rightarrow \quad \omega_{\max} = \sqrt{\frac{0.30 \times 9.81}{0.40}} \approx 2.71 \text{ rad s}^{-1}$$

period of revolution: $T_{\min} = \dfrac{2\pi}{\omega_{\max}} = \dfrac{2\pi}{2.71} \approx 2.32$ s

number of revolutions in one minute: $n_{\max} = \dfrac{t}{T_{\min}} = \dfrac{60}{2.32} \approx 25.9$ □

**Example 1.4** Particle $P$ of mass $m = 0.40$ kg is attached to one end of a light inextensible string of length $r = 0.80$ m. The particle is whirled at a constant angular speed $\omega$ in a vertical plane. (a) Given that the string never becomes slack, find the minimum value of $\omega$. (b) Given instead that the string will break if the tension is greater than 20 N, find the maximum value of $\omega$.




✍ at top of circle (point $A$): $F_c = T_A + mg = m\omega^2 r \quad \Rightarrow \quad T_A = m\omega^2 r - mg$

at bottom of circle (point $B$): $F_c = T_B - mg = m\omega^2 r \quad \Rightarrow \quad T_B = m\omega^2 r + mg$

tension is minimum at $A$, but string being taut requires $T \ge 0$ at any point, so $T_A \ge 0$

$$m\omega^2 r - mg \ge 0 \quad \Rightarrow \quad \omega^2 \ge \frac{g}{r}$$

$$\omega_{\min} = \sqrt{\frac{g}{r}} = \sqrt{\frac{9.81}{0.80}} \approx 3.5 \text{ rad s}^{-1}$$

tension is maximum at $B$, but string does not break requires $T \le T_{\max}$, so $T_B \le T_{\max}$

$$m\omega^2 r + mg \le T_{\max} \quad \Rightarrow \quad \omega^2 \le \frac{T_{\max}}{m} - \frac{g}{r}$$

$$\omega_{\max} = \sqrt{\frac{T_{\max}}{m} - \frac{g}{r}} = \sqrt{\frac{20}{0.40} - \frac{9.81}{0.80}} \approx 6.1 \text{ rad s}^{-1}$$ □

**Example 1.5** A pendulum bob of mass 120 g moves at constant speed and traces out a circle or radius $r = 10$ cm in a horizontal plane. The string makes an angle $\theta = 25^\circ$ to the vertical. (a) What is the tension in the string? (b) At what speed is the bob moving?

✍ vertical component of tension $T_y$ equals weight

$$T_y = mg \quad \Rightarrow \quad T\cos\theta = mg$$

$$T = \frac{mg}{\cos\theta} = \frac{0.12 \times 9.81}{\cos 25^\circ} \approx 1.3 \text{ N}$$

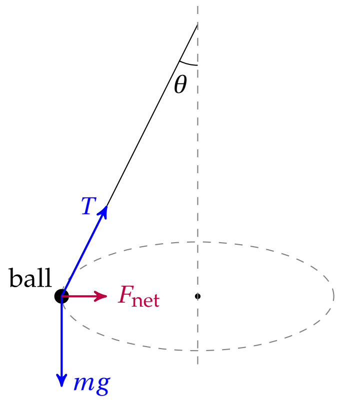
net force equals horizontal component of tension $T_x$

so component $T_x$ provides centripetal force

$$F_c = T_x \quad \Rightarrow \quad T\sin\theta = \frac{mv^2}{r}$$

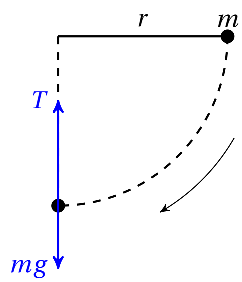
by eliminating $T$ and $m$, one can find

$$v^2 = \frac{r\tan\theta}{g} = \frac{0.10 \times \tan 25^\circ}{9.81} \quad \Rightarrow \quad v \approx 0.069 \text{ m s}^{-1} \quad \square$$




**Example 1.6** A small ball of mass $m$ is attached to an inextensible string of length $l$. The ball is held with the string taut and horizontal and is then released from rest.

When the ball reaches lowest point, find its speed and the tension in the string in terms of $m$ and $l$.




✍ energy conservation: G.P.E. loss = K.E. gain

$$mgr = \frac{1}{2}mv^2 \quad \Rightarrow \quad v = \sqrt{2gr}$$

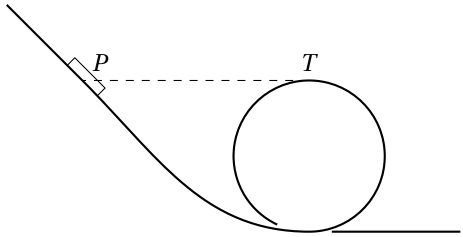
at lowest point: $F_c = T - mg = m\dfrac{v^2}{r}$

$$T = mg + m\frac{v^2}{r} = mg + m\frac{2gr}{r} = 3mg \qquad \square$$

**Question 1.4** Suggest what provides centripetal force in the following cases. (a) An athlete running on a curved track. (b) An aeroplane banking at a constant altitude. (c) A satellite moving around the earth.

**Question 1.5** A turntable that can rotate freely in a horizontal plane is covered by dry mud. When the angular speed of rotation is gradually increased, state and explain whether the mud near edge of the plate or near the mud will first leave the plate?

**Question 1.6** A bucket of water is swung at a constant speed and the motion describes a circle of radius $r = 1.0$m in the vertical plane. If the water does not pour down from the bucket even when it is at the highest position, how fast do you need to swing the bucket?



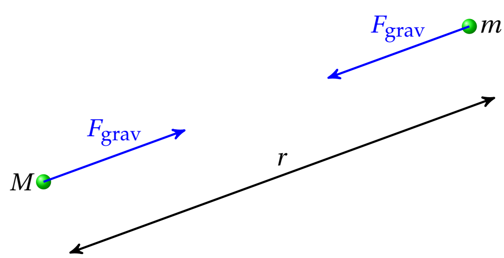

**Question 1.7** This question is about the design of a roller-coaster. We consider a slider that starts from rest from a point $P$ and slides along a frictionless circular track as sketched below. $P$ is at the same height as the top of the track $T$. (a) Show that the slider cannot get to $T$. (b) As a designer for a roller-coaster, you have to make sure the slider can reach point $T$ and continue to slide along the track, what is the minimum height for the point of release?

# Chapter 2
# Gravitational Fields

## 2.1 gravitational forces

### 2.1.1 Newton's law of gravitation

any object attracts any other object through the gravitational force




gravitational attraction between $M$ and $m$

**Newton's law of gravitation** states that gravitational force between two *point* masses is proportional to the product of their masses and inversely proportional to the square of their distance $\left(F_{\text{grav}} \propto \dfrac{Mm}{r^2}\right)$

this law was formulated in *Issac Newton*'s work 'The Principia', or 'Mathematical Principles of Natural Philosophy', first published in 1687

mathematically, gravitational force takes the form: $\boxed{F_{\text{grav}} = \dfrac{GMm}{r^2}}$

$G = 6.67 \times 10^{-11}\ \text{N m}^2\ \text{kg}^{-2}$ is the *gravitational constant*

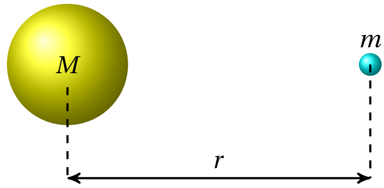
- gravitational force is always *attractive*
- gravity is *universal*, i.e., gravitational attraction acts between *any* two masses
- Newton's law of gravitation refers to *point masses*

  i.e., particles with no size, therefore distance $r$ can be easily defined

➢ a sphere with uniform mass distribution (e.g., stars, planets) can be treated as a *point model*

distance $r$ is taken between centres of the spheres [3]

(see Example 2.8, field lines around a planet *seem* to point towards centre of planet)

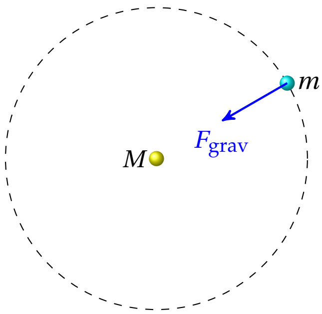



**Example 2.1** The Earth can be thought as a uniform sphere of radius $R = 6.4 \times 10^6$ m and mass $M = 6.0 \times 10^{24}$ kg. Estimate the gravitational force on a man of 60 kg at sea level.

✍ $$F = \frac{GMm}{R^2} = \frac{6.67 \times 10^{-11} \times 6.0 \times 10^{24} \times 60}{(6.4 \times 10^6)^2} \approx 586 \text{ N}$$ □

**Question 2.1** Estimate the gravitational force between you and your deskmate.

### 2.1.2 planetary motion




a planet/satellite orbiting around a star/earth

a planet/satellite can move around a star/earth in circular orbit

circular motion requires centripetal force

for these objects, gravitational force provides centripetal force

$$F_{\text{grav}} = F_c \quad \Rightarrow \quad \boxed{\frac{GMm}{r^2} = \frac{mv^2}{r}} \quad \text{or} \quad \boxed{\frac{GMm}{r^2} = m\omega^2 r}$$

**Example 2.2** GPS (Global Positioning System) satellites move in a circular orbits at about 20000 km above the earth's surface. The Earth has a radius $R = 6.4 \times 10^6$ m and mass $M = 6.0 \times 10^{24}$ kg. (a) Find the speed of GPS satellites. (b) Find its orbital period.

[3]This is known as *shell theorem*: a spherically symmetric shell (i.e., a hollow ball) affects external objects gravitationally as though all of its mass were concentrated at its centre, and it exerts no net gravitational force on any object inside, regardless of the object's location within the shell. (∗)

✍ $$\frac{GMm}{r^2}=\frac{mv^2}{r} \quad\Rightarrow\quad v=\sqrt{\frac{GM}{r}}=\sqrt{\frac{6.67\times10^{-11}\times6.0\times10^{24}}{6.4\times10^6+2.0\times10^7}}\approx3.9\times10^3\ \mathrm{m\ s^{-1}}$$
$$v=\frac{2\pi r}{T} \quad\Rightarrow\quad T=\frac{2\pi r}{v}=\frac{2\pi\times(6.4\times10^6+2.0\times10^7)}{3.9\times10^3}\approx4.3\times10^4\ \mathrm{s}\approx11.8\ \text{hours}$$ □

**Example 2.3** A **geostationary satellite** moves in a circular orbit that appears motionless to ground observers. The satellite follows the Earth's rotation, so the satellite rotates from west to east above equator with an orbital period of 24 hours. Find the radius of this orbit.

✍ $$\frac{GMm}{r^2}=m\omega^2 r \quad\Rightarrow\quad \frac{GMm}{r^2}=m\left(\frac{2\pi}{T}\right)^2 r \quad\Rightarrow\quad r^3=\frac{GMT^2}{4\pi^2}$$
$$r=\left(\frac{GMT^2}{4\pi^2}\right)^{1/3}=\left(\frac{6.67\times10^{-11}\times6.0\times10^{24}\times(24\times3600)^2}{4\pi^2}\right)^{1/3}\approx4.23\times10^7\ \mathrm{m}$$ □

**Example 2.4** Assuming the planets in the solar system all move around the sun in circular orbits, show that the square of orbital period is proportional to the cube of orbital radius. [4]

✍ $$\frac{GMm}{r^2}=m\omega^2 r \quad\Rightarrow\quad \frac{GMm}{r^2}=m\left(\frac{2\pi}{T}\right)^2 r \quad\Rightarrow\quad T^2=\frac{4\pi^2}{GM}\cdot r^3$$

$G$ is gravitational constant, $M$ is mass of the sun, so $\frac{4\pi^2}{GM}$ is a constant, so $T^2\propto r^3$ □

**Question 2.2** Given that it takes about 8.0 minutes for light to travel from the sun to the earth. (a) What is the mass of the sun? (b) At what speed does the earth move around the sun?

### 2.1.3 apparent weight

an object's *actual weight* is the gravitational attraction exerted by the earth's gravity

an object's *apparent weight* is the upward force (e.g., normal contact force exerted by ground, tension in a spring balance, etc.) that opposes gravity and prevents the object from falling

apparent weight can be different from actual weight due to vertical acceleration or buoyancy

but if we consider rotation of the earth, this also causes apparent weight to be lessened

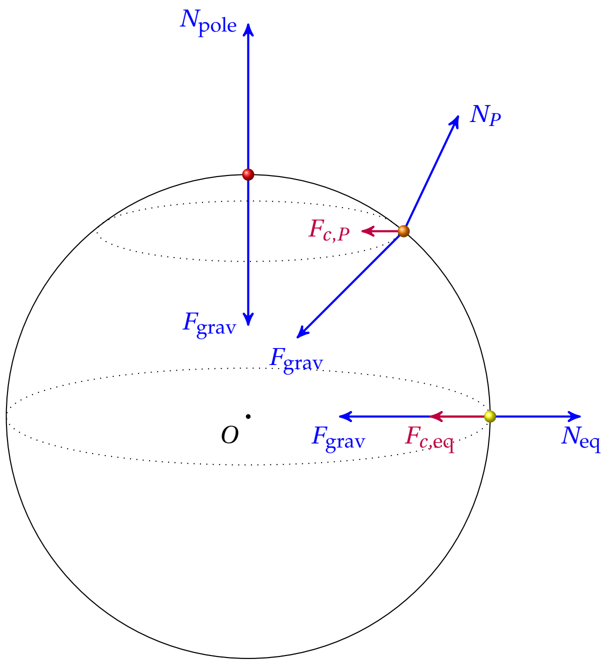
object resting on ground is actually rotating together with earth

resultant of gravitational force and contact force should provide centripetal force

for object on equator: $F_{c,\mathrm{eq}}=m\omega^2R \quad\Rightarrow\quad F_{\mathrm{grav}}-N_{\mathrm{eq}}=m\omega^2R \quad\Rightarrow\quad N_{\mathrm{eq}}=\frac{GMm}{R^2}-m\omega^2R$

[4] This is known as *Kepler's 3rd law* for planetary motions. In the early 17th century, German astronomer Johannes Kepler discovered three scientific laws which describes how planets move around the sun. This $T^2\propto r^3$ relation not only holds for circular orbits but are also correct for elliptical orbits.

Isaac Newton proved that Kepler's laws are consequences of his own law of universal gravitation, and therefore explained why the planets move in this way.




apparent weight at various positions near earth's surface (not to scale)

for object at poles: $F_{c,\text{pole}} = 0 \quad\Rightarrow\quad F_{\text{grav}} - N_{\text{pole}} = 0 \quad\Rightarrow\quad N_{\text{pole}} = \dfrac{GMm}{R^2}$

at lower latitudes, object describe larger circles, hence requires greater centripetal force

this offsets the balancing normal force, so apparent weight decreases near the equator

**Example 2.5** A stone of mass 5.0 kg is hung from a newton-meter near the equator. The Earth can be considered to be a uniform sphere of radius $R = 6370$ km and mass $M = 5.97 \times 10^{24}$ kg. (a) What is the gravitational force on the stone? (b) What is the reading on the meter?

✍ gravitational force: $F_{\text{grav}} = \dfrac{GMm}{R^2} = \dfrac{6.67 \times 10^{-11} \times 5.97 \times 10^{24} \times 5.0}{(6.37 \times 10^6)^2} \approx 49.07$ N

centripetal force required: $F_c = m\omega^2 R = m\left(\dfrac{2\pi}{T}\right)^2 R = 5.0 \times \dfrac{4\pi^2}{(24 \times 3600)^2} \times 6.37 \times 10^6 \approx 0.17$ N

apparent weight, or reading on meter: $N = F_{\text{grav}} - F_c = 49.07 - 0.17 \approx 48.90$ N □

**Question 2.3** Why astronauts in space stations are said to be *weightless*?

**Question 2.4** How do you find the apparent weight of an object at an arbitrary latitude $P$? Does the apparent weight act vertically downwards? Give your reasons.

## 2.2 gravitational fields

to explain how objects exert gravitational attraction upon one another at a distance, we introduce the concept of *force fields*

> **gravitational field** is a region of space where a mass is acted by a force

any mass $M$ (or several masses) can produce a gravitational field around it

a test mass $m$ within this field will experience a gravitational force

to describe the effect on a small mass $m$ in the field, we will further introduce

- *gravitational field strength*, to help us compute gravitational force on objects
- *gravitational potential*, to help us compute gravitational potential energy between objects

## 2.3 gravitational field strength

### 2.3.1 gravitational field strength

> **gravitational field strength** is defined as gravitational force per unit mass: $\boxed{g = \frac{F_{\text{grav}}}{m}}$

➢ unit of $g$: $[g] = \text{N kg}^{-1} = \text{ m s}^{-2}$, same unit as acceleration

➢ field strength due to an isolated source of mass $M$

at distance $r$ from the source, a test mass $m$ is acted by a force: $F_{\text{grav}} = \frac{GMm}{r^2}$

field strength at this position: $g = \frac{F_{\text{grav}}}{m} = \quad \Rightarrow \quad \boxed{g = \frac{GM}{r^2}}$

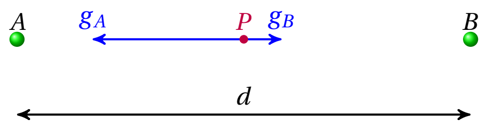
note that the field is produced by the source $M$, so field strength $g$ depends on $M$, not $m$

➢ field strength $g$ is a *vector* quantity, it has a direction

gravitation is *attractive*, so $g$ points towards source mass

to compute combined field strength due to several sources, should perform *vector sum* of contributions from each individual

**Example 2.6** Star $A$ of mass $6.0 \times 10^{30}$ kg and star $B$ of mass $1.5 \times 10^{30}$ kg are separated by a distance of $2.0 \times 10^{12}$ m. (a) What is the field strength at the mid-point $P$ of the two stars? (b) If a comet of mass $4.0 \times 10^6$ kg is at the mid-point, what force does it experience?




✍ $g_A$ acts towards $A$, $g_B$ acts towards $B$, they are in opposite directions

$$g_P = g_A - g_B = \frac{GM_A}{r_A^2} - \frac{GM_B}{r_B^2} = 6.67 \times 10^{-11} \times \left[ \frac{6.0 \times 10^{30}}{(1.0 \times 10^{12})^2} - \frac{1.5 \times 10^{30}}{(1.0 \times 10^{12})^2} \right] \approx 3.0 \times 10^{-4} \text{ N kg}^{-1}$$

force on comet: $F = mg = 4.0 \times 10^6 \times 3.0 \times 10^{-4} \approx 1.2 \times 10^3$ N □

### 2.3.2 acceleration of free fall

if field strength $g$ is known, gravitational force on an object of mass $m$ is: $F_{\text{grav}} = mg$

if the object is acted by gravity only, then $F_{\text{net}} = F_{\text{grav}} \quad \Rightarrow \quad ma = mg \quad \Rightarrow \quad \boxed{a = g}$ [5]

this shows gravitational field strength gives the acceleration of free fall!

**Example 2.7** The earth has a radius of 6370 km. (a) Find the mass of the earth. [6] (b) Find the acceleration of free fall at the top of Mount Everest. (height of Mount Everest $H \approx 8.8$ km)

✍ consider acceleration of free fall near surface of earth:

$$g_s = \frac{GM}{R^2} \quad \Rightarrow \quad 9.81 = \frac{6.67 \times 10^{-11} \times M}{(6.37 \times 10^6)^2} \quad \Rightarrow \quad M \approx 5.97 \times 10^{24} \text{ kg}$$

[5] Rigorously speaking, the two $m$'s are different concepts. There is the *inertia* mass, decribing how much an object resists the change of state of motion. There is also the *gravitational* mass, describing the effect produced and experienced by the object in gravitational fields. Yet no experiment has ever demonstrated any significant difference between the two. The reason why the two masses are identical is very profound. We have shown here acceleration of free fall equals gravitational field strength, but Albert Einstein's *equivalence principle* suggests that it is actually impossible to distinguish between a uniform acceleration and a uniform gravitational field. This idea lies at the heart of the *general theory of relativity*, where I should probably stop going further.

[6] British scientist Henry Cavendish devised an experiment in 1798 to measure the gravitational force between masses in his laboratory. He was the first man to yield accurate values for the gravitational constant $G$. Then he was able to carry out this calculation, referred by himself as 'weighing the world'.

at top of Mount Everest:

$$g_{\mathrm{ME}} = \frac{GM}{(R+H)^2} = \frac{6.67\times10^{-11}\times5.97\times10^{24}}{(6.37\times10^{6}+8.8\times10^{3})^2} \approx 9.78\ \mathrm{N\,kg^{-1}} \quad \Rightarrow \quad a_{\mathrm{ME}} \approx 9.78\ \mathrm{m\,s^{-2}} \qquad \square$$

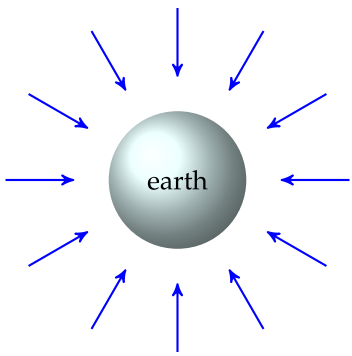
### 2.3.3 gravitational field lines

**gravitational field lines** are drawn to graphically represent the pattern of field strength

- *direction* of field lines show the *direction* of field strength in the field
- *spacing* between field lines indicates the *strength* of the gravitational field
- gravitational field lines always end up at a mass

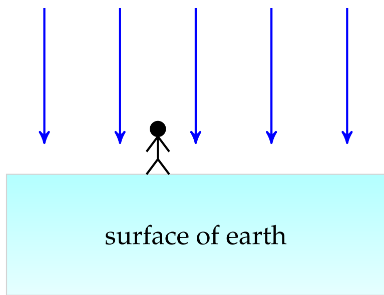
this arises from the attractive nature of gravitation

**Example 2.8** field around the earth




*radial* field (field lines normal to surface)

**Example 2.9** field near earth's surface




almost a *uniform* field

(field lines are parallel and equally spaced)

## 2.4 gravitational potential & potential energy

### 2.4.1 potential energy

*potential energy* is the energy possessed by an object due to its position in a force field

work done *by* force field decreases P.E., and work done *against* a force field increases P.E.

let $W$ be work by the force field, then we have: $\boxed{W = -\Delta E_p}$

to define potential energy of an object at a specific point $X$, we can

(1) choose a position where potential energy is defined to be zero

(2) find work done by force field to bring the object from zero P.E. point to $X$

(3) consider change in P.E.: $\Delta E_p = E_{p,X} - E_{p,\text{initial}} = E_{p,X} - 0 = E_{p,X}$

but $\Delta E_p = -W$, so P.E. at point $X$ is found: $E_{p,X} = -W$

so potential energy is equal to (negative) work done to move the object to a specific position

**gravitational potential energy near earth's surface**

we may choose a zero G.P.E. point, for example, $E_p(0) = 0$ at sea level

if mass $m$ is moved up for a height $h$, work done by gravity is $W = -mgh$[7]

this causes a change in gravitational potential energy $\Delta E_p = -W = mgh$

then at altitude $h$, G.P.E. can be given by $E_p(h) = mgh$

### 2.4.2 gravitational potential energy

we are now ready to derive an expression for G.P.E. between two masses $M$ and $m$

we define $E_p = 0$ at $r = \infty$ (choice of zero potential energy, no force so no G.P.E.), then

**gravitation potential energy** is equal to the work done by gravitational force to bring a mass to a specific position from *infinity*

consider a mass $m$ at infinity with zero energy and a source mass $M$ at origin

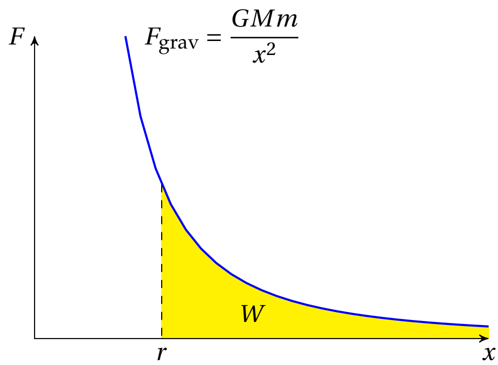
let's find out how much work is done by gravitational force to pull $m$ towards the origin


but $F_{\text{grav}}$ varies as inverse square of separation $x$

so here we need to evaluate work done by a non-constant force

we can plot a $F$-$x$ graph, then magnitude of work done equals area under the graph

[7]Negative sign because this is work against gravity.




integrate[8] to evaluate the area: $W = \int_r^\infty \frac{GMm}{x^2}\mathrm{d}x = -\frac{GMm}{x}\Big|_r^\infty = \frac{GMm}{r}$

$\Delta E_p = -W \quad \Rightarrow \quad E_p(r) - E_p(\infty) = -\frac{GMm}{r}$

but we have defined $E_p(\infty) = 0$, therefore: $\boxed{E_p(r) = -\frac{GMm}{r}}$

$E_p(r)$ gives G.P.E between masses $M$ and $m$ when they are at distance $r$ from each other

- as $r \to \infty$, $E_p \to 0$, this agrees with our choice of zero G.P.E. point
- potential energy is a *scalar* quantity (negative sign cannot be dropped)
- G.P.E. is always *negative*, this is due to *attractive* nature of gravity

  to separate masses, work must be done to overcome attraction

  so G.P.E. increases with separation $r$, i.e., G.P.E. is maximum at infinity, which is zero

  G.P.E. between masses at finite separation must be less than zero
- $E_p = mgh$ is only applicable near earth's surface where field is almost *uniform*

  $E_p = -\frac{GMm}{r}$ is a more *general* formula for gravitational potential energy [9]

**Example 2.10** A meteor is travelling towards a planet of mass $M$. When it is at a distance of $r_1$ from centre of $M$, it moves at speed $v_1$. When it is $r_2$ from $M$, it moves at speed $v_2$. Assume

---

[8] In general, work done by a non-constant force over large distance is $W = \int_{\text{initial}}^{\text{final}} F\mathrm{d}x$.

For our case, $x$ is the displacement away from the source, but gravitational force tends to pull the mass towards the source. $F$ and $x$ are in opposite directions, a negative sign is needed for $F$. Therefore the work done by gravity to bring mass $m$ from infinity is: $W = \int_\infty^r F\mathrm{d}x = \int_\infty^r \left(-\frac{GMm}{x^2}\right)\mathrm{d}x = +\frac{GMm}{x}\Big|_\infty^r = \frac{GMm}{r}$.

[9] One can recover $\Delta E_p = mg\Delta h$ from $E_p = -\frac{GMm}{r}$. Near the earth's surface, if $r_1 \approx r_2 \approx R$, and $r_1 > r_2$, then we have: $\Delta E_p = E_p(r_1) - E_p(r_2) = -GMm\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = GMm\frac{r_1 - r_2}{r_1 r_2} \approx m\frac{GM}{R^2}\Delta r \xlongequal{g=GM/R^2} mg\Delta h$.

only gravitational force applies, establish a relationship between these quantities.

✍ energy conservation: K.E. + G.P.E. = const $\Rightarrow \frac{1}{2}mv_1^2 + \left(-\frac{GMm}{r_1}\right) = \frac{1}{2}mv_2^2 + \left(-\frac{GMm}{r_2}\right)$ □

**Example 2.11** If an object is thrown from the surface of a planet at sufficiently high speed, it might escape from the influence of the planet's gravitational field. The minimum speed required is called the *escape velocity*. Using the data from previous examples, find the escape velocity from the surface of earth.

✍ assuming no energy loss to air resistance, then total energy is conserved

$$\text{K.E.} + \text{G.P.E. at surface of planet} = \text{K.E.} + \text{G.P.E. at infinity}$$

$$\frac{1}{2}mu^2 + \left(-\frac{GMm}{R}\right) = \frac{1}{2}mv^2 + 0 \quad \xRightarrow{v\geq 0} \quad u^2 \geq \frac{2GM}{R} \quad \Rightarrow \quad u_{\text{min}} = \sqrt{\frac{2GM}{R}}$$

for earth, escape velocity $u_{\text{min}} = \sqrt{\frac{2\times 6.67\times 10^{-11}\times 6.0\times 10^{24}}{6.4\times 10^{6}}} \approx 1.12\times 10^4 \text{ m s}^{-1}$ □

**Question 2.5** A planet of uniform density distribution is of radius $R$ and mass $M$. A rock falls from a height of $3R$ above the surface of the planet. Assume the planet has no atmosphere, show that the speed of the rock when it hits the ground is $v = \sqrt{\frac{3GM}{4R}}$.

**Question 2.6** A space probe is travelling around a planet of mass $M$ in a circular orbit of radius $r$. (a) Show that the total mechanical energy (sum of kinetic energy and gravitational energy) of the space probe is $E_{\text{total}} = -\frac{2GMm}{r}$. (b) If the space probe is subject to small resistive forces, state the change to its orbital radius and its orbiting speed.

**Question 2.7** A *black hole* is a region of spacetime where gravitation is so strong that even light can escape from it. For a star of mass $M$ to collapse and form a black hole, it has to be compressed below a certain radius. (a) Show that this radius is given by $R_S = \frac{2GM}{c^2}$, known as the *Schwarzschild radius*[10]. (b) Show that the Schwarzschild radius for the sun is about 3 km.

[10]When you deal with very strong gravitational fields, Newton's law of gravitation breaks down and effects of Einstein's *general theory of relativity* come into play. The radius of a *Newtonian* black hole being equal to the radius of a Schwarzschild black hole is a mere coincidence.

### 2.4.3 gravitational potential

it is useful to introduce a quantity called *potential* at a specific point in a gravitational field

gravitational potential can be considered as the potential energy per unit mass: $\varphi = \dfrac{E_p}{m}$

**gravitational potential** at a point is defined as the work done to bring *unit* mass from *infinity* to that point

- unit: $[\varphi] = \mathrm{J\,kg^{-1}}$
- gravitational potential due to an isolated source $M$

$$\varphi = \frac{E_p}{m} = \frac{-\frac{GMm}{r}}{m} \quad \Rightarrow \quad \boxed{\varphi = -\frac{GM}{r}}$$

- potential at infinity is zero: $\varphi_\infty = 0$

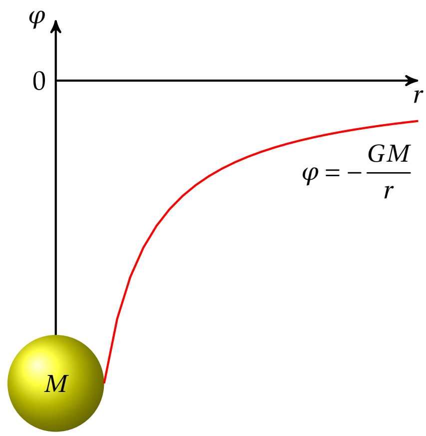
  this is our choice of zero potential point
- gravitational potential is a *scalar*

  combined potential due to several masses equals scalar sum of potential of each individual
- gravitational potential is always *negative*

  again this arises from attractive nature of gravity

  work is done to pull unit mass away from source

  farther from source means higher potential




**Example 2.12** A star $A$ of mass $M_A = 1.5 \times 10^{30}$ kg and a planet $B$ of mass $M_B = 6.0 \times 10^{26}$ kg form an isolated astronomical system. Point $P$ is between $A$ and $B$, and is at distance $r_A = 2.0 \times 10^{12}$ m from $A$, and distance $r_B = 8.0 \times 10^{10}$ m from $B$. (a) Find the gravitational potential at $P$. (b) A meteor is initially at very large distance from the system with negligible speed. It then travels towards point $P$ due to the gravitational attraction. Find its speed when it reaches $P$.

✍ gravitational potential at $P$: $\varphi_P = \varphi_A + \varphi_B = \left(-\dfrac{GM_A}{r_A}\right) + \left(-\dfrac{GM_B}{r_B}\right)$

$$\varphi_P = -6.67 \times 10^{-11} \times \left(\frac{1.5 \times 10^{30}}{2.0 \times 10^{12}} + \frac{6.0 \times 10^{26}}{8.0 \times 10^{10}}\right) \approx -5.05 \times 10^7 \mathrm{\ J\,kg^{-1}}$$

gain in K.E. = loss in G.P.E.: $\dfrac{1}{2}mv^2 = m\Delta\varphi \quad \Rightarrow \quad v^2 = 2(\varphi_\infty - \varphi_P) = -2\varphi_P$

$$v = \sqrt{-2 \times (-5.05 \times 10^7)} \approx 1.01 \times 10^4 \mathrm{\ m\,s^{-1}}$$

□

**Example 2.13** The Moon may be considered to be an isolated sphere of radius $R = 1.74 \times 10^3$ km. The gravitational potential at the surface of the moon is about $-2.82 \times 10^6$ J kg$^{-1}$. (a) Find the mass of the moon. (b) A stone travels towards the moon such that its distance from the centre of the moon changes from $3R$ to $2R$. Determine the change in gravitational potential. (c) If the stone starts from rest, find its final speed.

✍ at surface: $\varphi(R) = -\dfrac{GM}{R} \quad \Rightarrow \quad -2.82 \times 10^6 = -\dfrac{6.67 \times 10^{-11} \times M}{1.74 \times 10^6} \quad \Rightarrow \quad M = 7.36 \times 10^{22}$ kg

from $3R$ to $2R$: $\Delta\varphi = \varphi_{(3R)} - \varphi_{(2R)} = \left(-\dfrac{GM}{3R}\right) - \left(-\dfrac{GM}{2R}\right) = \dfrac{GM}{6R} = \dfrac{2.82 \times 10^6}{6} \approx 4.70 \times 10^5$ J kg$^{-1}$

note this change is a *decrease* in gravitational potential

gain in K.E. = loss in G.P.E.: $\dfrac{1}{2}mv^2 = m\Delta\varphi \quad \Rightarrow \quad v = \sqrt{2\Delta\varphi} = \sqrt{2 \times 4.70 \times 10^5} \approx 970$ m s$^{-1}$ □

**Question 2.8** Given that the moon is of radius 1700 km and mass $7.4 \times 10^{22}$ kg. (a) Find the change in gravitational potential when an object is moved from moon's surface to 800 km above the surface. (b) If a rock is projected vertically upwards with an initial speed of 1800 m s$^{-1}$ from surface, find the rock's speed when it reaches a height of 800 km. (c) Suggest whether the rock can escape from the moon's gravitational field completely.

# Chapter 3

# Oscillation

## 3.1 oscillatory motion

**oscillation** refers to a repetitive back and forth motion about its *equilibrium position*

the equilibrium position is a point where all forces on oscillator are balanced

release an object from its equilibrium position from rest, it will stay at rest

examples of oscillation includes pendulum of a clock, vibrating string, swing, etc.

### 3.1.1 amplitude, period, frequency

to describe motion of an oscillator, we define the following quantities:

- **displacement** ($x$): distance from the equilibrium position
- **amplitude** ($x_0$): maximum displacement from the equilibrium position
- **period** ($T$): time for one complete oscillation
- **frequency** ($f$): number of oscillations per unit time

frequency is related to period as: $\boxed{f = \frac{1}{T}}$

displacement $x$ varies with time $t$ repetitively, for which we can plot an $x$-$t$ graph

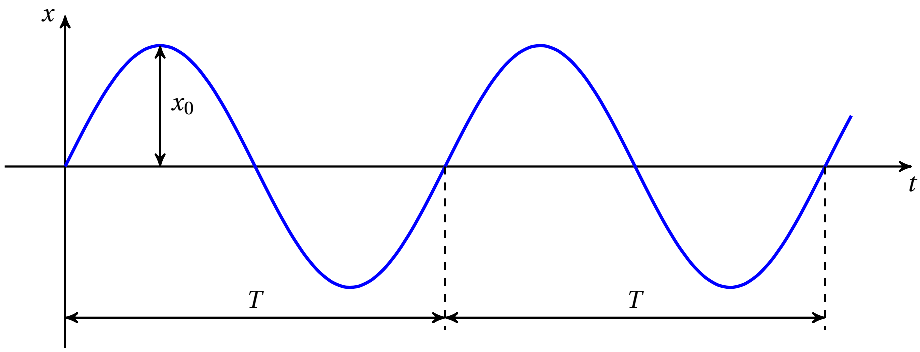
amplitude $x_0$ and period $T$ are labelled on the graph

### 3.1.2 phase angle

the point that an oscillator has reached within a complete cycle is called **phase angle** ($\phi$)

- unit of phase angle: $[\phi] = \text{rad}$

it looks like an angle, but better think of it as a number telling fraction of a complete cycle

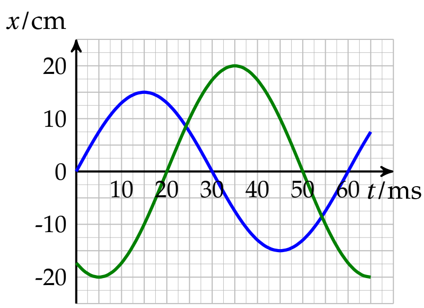



displacement-time graph for a typical oscillator

➢ we use **phase difference** $\Delta\phi$ to compare how much one oscillator is ahead of another

$\Delta\phi$ is found in terms of fraction of an oscillation: $\Delta\phi = \frac{\Delta t}{T} \times 2\pi$ (also measured in radians)

**Example 3.1** Compare the two oscillations from the $x$-$t$ graph below.




✍ both have period $T = 60$ ms

frequency $f = \frac{1}{60 \times 10^{-3}} \approx 16.7$ Hz

they are of different amplitudes

one has $x_0 = 15$ cm, the other has $x_0 = 20$ cm

time difference: $\Delta t = 20$ ms

phase difference: $\Delta\phi = \frac{\Delta t}{T} \times 2\pi = \frac{20}{60} \times 2\pi = \frac{2\pi}{3}$ rad

### 3.1.3 acceleration & restoring force

for any oscillatory motion, consider its velocity and acceleration at various positions

its acceleration must be always pointing towards the equilibrium position

resultant force always acts in the direction to restore the system back to its equilibrium point, this net force is known as the **restoring force**

if at equilibrium position, then no acceleration or restoring force

## 3.2 simple harmonic oscillation

if an oscillator has an acceleration always proportional to its displacement from the equilibrium position, and acceleration is in opposite direction to displacement, then the oscillator is performing **simple harmonic motion**

many phenomena can be approximated by simple harmonics

examples are motion of a pendulum, molecular vibrations, etc.

complicated motions can be decomposed into a set of simple harmonics

simple harmonic motion provides a basis for the study of many complicated motions [11]

### 3.2.1 equation of motion

defining equation for simple harmonics can be written as $\boxed{a = -\omega^2 x}$

$\omega$ is some constant, so $a$ is proportional to $x$

the minus sign shows $a$ and $x$ are in opposite directions

general solution to this this equation of motion[12] takes the form: $\boxed{x = x_0 \sin(\omega t + \phi)}$

$x_0$ represents the amplitude, $\omega$ is called the angular frequency, $\phi$ is the phase angle

#### angular frequency

- **angular frequency** satisfies the relation: $\boxed{\omega = \dfrac{2\pi}{T} = 2\pi f}$
- unit of angular frequency: $[\omega] = \mathrm{rad} \cdot \mathrm{s}^{-1}$
- angular frequency $\omega$ is determined by the system's *physical constants* only

[11]This can be done through a mathematical technique known as *Fourier analysis*. For example, a uniform circular motion can be considered as the combination of two simple harmonic motion in $x$- and $y$-directions.

[12]You probably know that acceleration can be written as the second derivative of displacement: $a = \dfrac{\mathrm{d}^2 x}{\mathrm{d}t^2}$, so $a = -\omega^2 x$ is equivalent to $\dfrac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \omega^2 x = 0$, which a *second-order differential equation*. If you do not know how to solve it, you may have the chance to study this in an advanced calculus course.

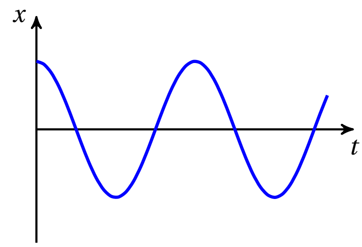
if an object is set to oscillate *freely* with no external force, its period will always be the same

frequency of an free oscillatory system is called the **natural frequency**

#### phase angle

- phase angle $\phi$ is dependent on *initial conditions* (e.g. initial position and initial speed at $t = 0$?)
- in many cases, phase angle term can be avoided if a suitable trigonometric function is chosen

**Example 3.2** A simple harmonic oscillator is displaced by 6.0 cm from its rest position and let go at $t = 0$. Given that the period of this system is 0.80 s, state an equation for its displacement-time relation.



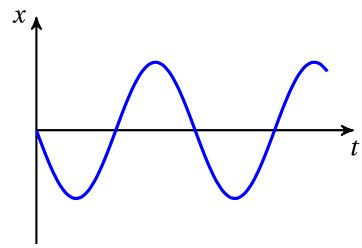

✍ angular frequency: $\omega = \frac{2\pi}{T} = \frac{2\pi}{0.80} = \frac{5\pi}{2}\pi$ rad s$^{-1}$

initial displacement $x(0) = +x_0 = 6.0$ cm

for displacement-time relation, we use cosine function

$$x(t) = x_0 \cos \omega t \quad \Rightarrow \quad x = 6.0 \cos\left(\frac{5\pi}{2} t\right) \qquad \square$$

**Example 3.3** A simple harmonic oscillator is initially at rest. At $t = 0$, it is given an initial speed in the negative direction. Given that the frequency is 1.5 Hz and the amplitude is 5.0 cm, state an equation for its displacement-time relation.




✍ angular frequency: $\omega = 2\pi f = 2\pi \times 1.5 = 3\pi$ rad s$^{-1}$

initial displacement $x(0) = 0$

for displacement-time relation, we use sine function

$$x(t) = -x_0 \cos \omega t \quad \Rightarrow \quad x = -5.0 \sin(3\pi t) \qquad \square$$

### 3.2.2 examples of simple harmonics

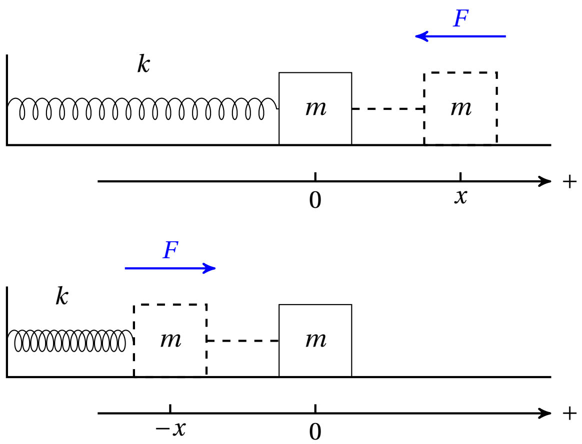
#### mass-spring oscillator

a mass-spring oscillator system consists of a block of mass $m$ and an ideal spring

when a spring is stretched or compressed by a mass, the spring develops a restoring force

magnitude of this force obeys *Hooke's law*: $F = kx$

direction of this force is in opposite direction to displacement $x$




restoring force acting on the ideal mass-spring oscillator

take vector nature of force into account, we find

$$F_{\text{net}} = ma \quad \Rightarrow \quad -kx = ma \quad \Rightarrow \quad a = -\frac{k}{m}x$$

spring constant $k$ and mass $m$ are constants, so $a \propto x$

negative sign shows $a$ and $x$ are in opposite directions

so mass-spring oscillator executes simple harmonic motion

compare with $a = -\omega^2 x \quad \Rightarrow \quad \omega^2 = \frac{k}{m} \quad \Rightarrow \quad \boxed{\omega = \sqrt{\frac{k}{m}}}$

period of mass-spring oscillator: $T = \frac{2\pi}{\omega} \quad \Rightarrow \quad T = 2\pi\sqrt{\frac{m}{k}}$

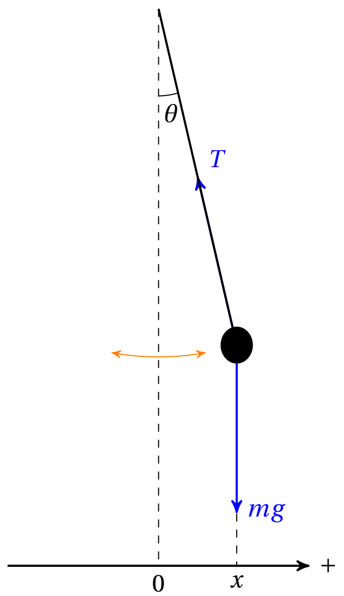
- period and frequency are solely determined by mass of oscillator $m$ and spring constant $k$

  identical mass-spring systems will oscillate at same frequency no matter what amplitude
- $m \nearrow \Rightarrow T \nearrow$, greater mass means greater inertia, oscillation becomes slower
- $k \nearrow \Rightarrow T \searrow$, greater $k$ means stiffer spring, greater restoring force makes oscillation go faster

**simple pendulum**

a simple pendulum is set up by hanging a bob on a light cord from a fixed point

displace the bob by some angle and release from rest, it can swing freely




one can show this performs simple harmonic motion for *small-angle* oscillation

if angular displacement $\theta$ is small, then the pendulum has almost no vertical displacement, the motion can be considered to be purely horizontal

vertically: $T\cos\theta \approx mg \xRightarrow{\cos\theta\approx 1 \text{ as } \theta\to 0} T \approx mg$

horizontally: $-T\sin\theta = ma \xRightarrow{\sin\theta = x/L} a \approx -\dfrac{g}{L}x$

this shows simple pendulum undergoes simple harmonics

compare with defining equation for simple harmonics:

$$a = -\omega^2 x \quad \Rightarrow \quad \omega = \sqrt{\frac{g}{L}}$$

period for a simple pendulum: $\boxed{T = 2\pi\sqrt{\frac{L}{g}}}$

- ➢ period and frequency of a pendulum are determined by length of the string $L$ only

  as long as displacement remains small, frequency does not depend on amplitude
- ➢ $L\nearrow \Rightarrow T\nearrow$, longer pendulums oscillate more slowly
- ➢ $g\searrow \Rightarrow T\nearrow$, if there is no gravity ($g=0$), then the bob will not move at all ($T\to\infty$)

**Question 3.1** A cylindrical tube of total mass $m$ and cross sectional area $A$ floats upright in a liquid of density $\rho$. When the tube is given a small vertical displacement and released, the magnitude of the resultant force acting on the tube is related to its vertical displacement $y$ by the expression: $F_{\text{net}} = \rho g A y$. (a) Show that the tube executes simple harmonic motion. (b) Find an expression for the frequency of the oscillation.

**Question 3.2** A small glider moves along a horizontal air track and bounces off the buffers at the ends of the track. Assume the track is frictionless and the buffers are perfectly elastic, state and explain whether the glider describes simple harmonic motion.

### 3.2.3 velocity & acceleration

displacement of simple harmonic oscillator varies with time as: $x = x_0\sin(\omega t + \phi)$

from this displacement-time relation, we can find velocity and acceleration relations

**velocity**

to find velocity-time relation, let's recall that velocity $v$ is rate of change of displacement $x$

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} x_0 \sin(\omega t + \phi) \quad \Rightarrow \quad \boxed{v(t) = \omega x_0 \cos(\omega t + \phi)}$$

by taking $v^2 + \omega^2 x^2$, the sine and cosine terms can be eliminated, we find:

$$v^2 + \omega^2 x^2 = \omega^2 x_0^2 \cos^2(\cdots) + \omega^2 x_0^2 \sin^2(\cdots) = \omega^2 x_0^2$$

this gives velocity-displacement relation: $\boxed{v(x) = \pm\omega\sqrt{x_0^2 - x^2}}$

➢ at equilibrium position $x = 0$, speed is maximum: $\boxed{v_{\max} = \omega x_0}$

➢ when $x = \pm x_0$, oscillator is momentarily at rest: $v = 0$

**acceleration**

acceleration-time relation is found by further taking rate of change of velocity $v$

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \omega x_0 \cos(\omega t + \phi) \quad \Rightarrow \quad \boxed{a(t) = -\omega^2 x_0 \sin(\omega t + \phi)}$$

this is actually unnecessary, if we compare this with $x(t) = x_0 \sin(\omega t + \phi)$, we have: $a = -\omega^2 x$

we have recovered the definition for simple harmonics

(if $a \propto x$ and in opposite directions to $x$, then simple harmonic motion)

so acceleration-displacement relation is given by the defining equation explicitly $\boxed{a(x) = -\omega^2 x}$

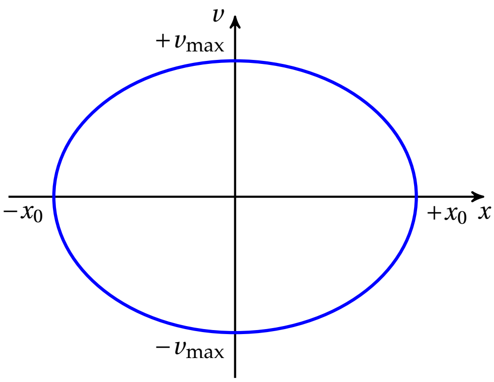
➢ at equilibrium position $x = 0$, zero acceleration

➢ when $x = \pm x_0$, acceleration is greatest: $\boxed{a_{\max} = \omega^2 x_0}$

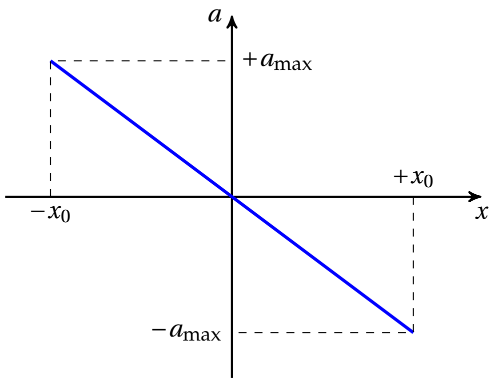
let's take $x = x_0 \sin\omega t$ as example, changes of $x$, $v$, $a$ over time are listed below

| time $t$ | 0 | $\frac{1}{4}T$ | $\frac{1}{2}T$ | $\frac{3}{4}T$ | $T$ |
|---|---|---|---|---|---|
| displacement: $x = x_0 \sin\omega t$ | 0 | +max | 0 | −max | 0 |
| velocity: $v = \omega x_0 \cos\omega t$ | +max | 0 | −max | 0 | +max |
| acceleration: $a = -\omega^2 x = -\omega^2 x_0 \sin\omega t$ | 0 | −max | 0 | +max | 0 |

**Example 3.4** The motion of a simple pendulum is approximately simple harmonic. As the pendulum swings from one side to the other end, it moves through a distance of 6.0 cm and




velocity-displacement graph




acceleration-displacement graph

the time taken is 1.0 s. (a) State the period and amplitude. (b) Find the greatest speed during the oscillation. (c) Find its speed when displacement $x = 1.2$ cm.

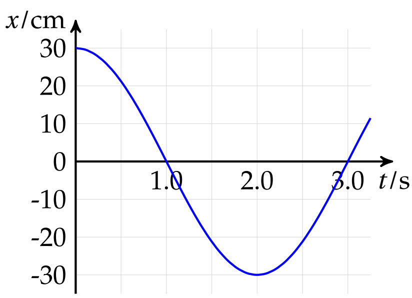
✍ period: $T = 2 \times 1.0 = 2.0$ s, and amplitude: $x_0 = \frac{1}{2} \times 6.0 = 3.0$ cm

angular frequency: $\omega = \frac{2\pi}{T} = \frac{2\pi}{2.0} = \pi$ rad s$^{-1}$

greatest speed: $v_{max} = \omega x_0 = \pi \times 3.0 \approx 9.4$ cm s$^{-1}$

speed at 1.2 cm: $v = \omega\sqrt{x_0^2 - x^2} = \pi \times \sqrt{3.0^2 - 1.2^2} \approx 8.6$ cm s$^{-1}$ □

**Example 3.5** Given the $x$-$t$ graph of a simple harmonic oscillator. (a) Find its speed at $t = 0$. (b) Find its greatest speed. (b) Find its acceleration at $t = 1.0$ s.




✍ at $t = 0$, $x = +x_0 \quad \Rightarrow \quad v = 0$ (zero gradient)

from graph: amplitude $x_0 = 30$ cm, period $T = 4.0$ s

angular frequency: $\omega = \frac{2\pi}{T} = \frac{2\pi}{4} = \frac{\pi}{2}$ rad s$^{-1}$

greatest speed: $v_{max} = \omega x_0 = \frac{\pi}{2} \times 30 \approx 47$ cm s$^{-1}$

at $t = 1.0$ s, $x = 0 \quad \Rightarrow \quad a = 0$

(equilibrium position so no acceleration) □

**Question 3.3** Assume the motion of a car engine piston is simple harmonic. The piston completes 3000 oscillations per minute. The amplitude of the oscillation is 4.0 cm. (a) Find the greatest speed. (b) Find the greatest acceleration.

### 3.2.4 vibrational energy

consider the *ideal* mass-spring oscillator, its vibrational energy consists of two parts:

- kinetic energy of the mass: $E_k = \frac{1}{2}mv^2 = \frac{1}{2}m\omega^2 x_0^2 \cos^2\omega t \xlongequal{v=\pm\omega\sqrt{x_0^2-x^2}} \frac{1}{2}m\omega^2(x_0^2 - x^2)$
- (elastic) potential energy in the spring: $E_p = \frac{1}{2}kx^2 \xlongequal{\omega=\sqrt{\frac{k}{m}}} \frac{1}{2}m\omega^2 x^2$

total energy of the oscillator: $E = E_k + E_p \quad \Rightarrow \quad \boxed{E = \frac{1}{2}m\omega^2 x_0^2}$

➢ although this formula is derived from the mass-spring model

$E = \frac{1}{2}m\omega^2 x_0^2$ can be used to compute vibrational energy of all simple harmonic oscillators

➢ for an ideal system, total energy remains constant

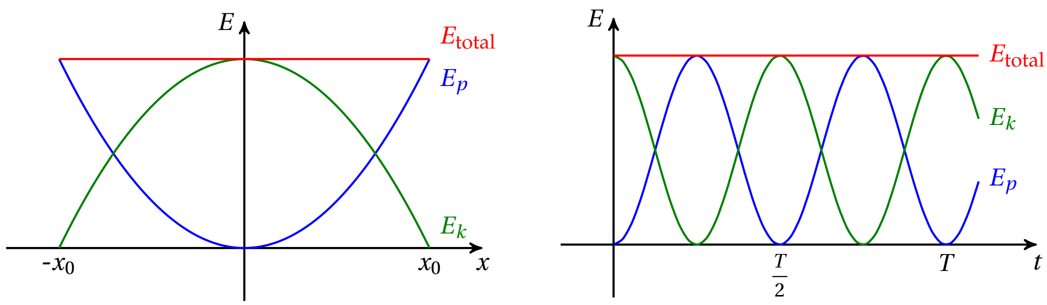
$E_k$ and $E_p$ keep changing, one transfers into another, but total energy is *conserved*

➢ when $x = 0$, $E_k = \max$, $E_p = 0$, vibrational energy is purely kinetic

$$E = E_{k,\max} = \frac{1}{2}mv_{\max}^2 \xlongequal{v_{\max}=\omega x_0} \frac{1}{2}m\omega^2 x_0^2$$

➢ when $x = \pm x_0$, $E_k = 0$, $E_p = \max$, vibrational energy is purely potential

$$E = E_{p,\max} = \frac{1}{2}kx_0^2 \xlongequal{\omega=\sqrt{\frac{k}{m}}} \frac{1}{2}m\omega^2 x_0^2$$




vibrational energy of a mass-spring oscillator

**Example 3.6** A block of mass 150 g at the end of a spring oscillates with a period of 0.80 s. The maximum displacement from its rest position is 12 cm. Find the energy of the vibration.

✍ $$E = \frac{1}{2}m\omega^2 x_0^2 = \frac{1}{2}m\left(\frac{2\pi}{T}\right)^2 x_0^2 = \frac{1}{2} \times 0.15 \times \frac{4\pi^2}{0.80^2} \times 0.12^2 \approx 6.7 \times 10^{-2}\ \text{J}$$ □

**Question 3.4** An oscillator is given an energy of 20 mJ and starts to oscillate, it reaches an amplitude of 8.0 cm. If we want to double the amplitude, find the vibrational energy required.

## 3.3 damped oscillations

total vibrational energy stays constant for an ideal system

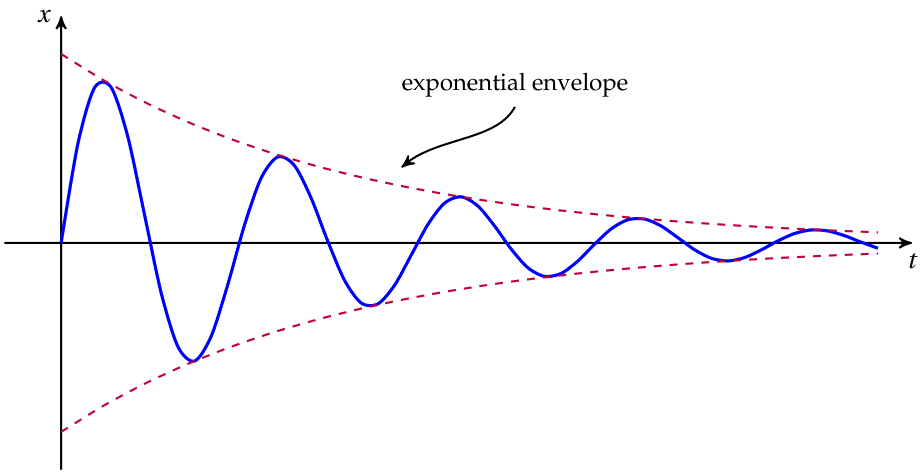
but in reality, there are friction, resistance and viscous forces that oppose motion

amplitude of an oscillator decreases due to energy loss to friction, this is called **damping**

### 3.3.1 light damping

for a **lightly-damped oscillator**, amplitude decreases *gradually*

oscillator will not stop moving back and forth after quite a few oscillations




- ➢ decrease in amplitude is *non-linear* in time (exponential decay in many cases)
- ➢ frequency and period are (almost) unchanged

**Example 3.7** An oscillator is composed of a block of mass $m = 250$ g and a spring of $k = 1.6$ N/cm. It is displaced by 5.0 cm from its rest position and set free. (a) What is its angular frequency? (b) what is the initial vibrational energy? (c) After a few oscillations, 40% of its energy is lost due to damping. What is its new amplitude?

✍ angular frequency: $\omega = \sqrt{\dfrac{k}{m}} = \sqrt{\dfrac{160}{0.25}} \approx 25.3 \text{ rad s}^{-1}$

energy of oscillator: $E = \dfrac{1}{2}m\omega^2 x_0^2 = \dfrac{1}{2} \times 0.25 \times 25.3^2 \times 0.050^2 = 0.20$ J[13]

[13]An easier approach: $E = \dfrac{1}{2}kx_0^2 = \dfrac{1}{2} \times 160 \times 0.050^2 = 0.20$ J.

since $E \propto x_0^2$, so: $\frac{E'}{E} = \frac{x_0'^2}{x_0^2} \quad \Rightarrow \quad 60\% = \frac{x_0'^2}{x_0^2} \quad \Rightarrow \quad x_0' = \sqrt{0.6}x_0 = \sqrt{0.6} \times 5.0 \approx 3.9$ cm □

**Question 3.5** A small toy boat of mass 360 g floats on surface of water. It is gently pushed down and then released. During the fist four complete cycles of its oscillation, its amplitude decreased from 5.0 cm to 2.0 cm in a time of 6.0 s. Find the energy loss.

#### 3.3.2 heavy damping

if resistive forces are too strong, there will be no oscillatory motion

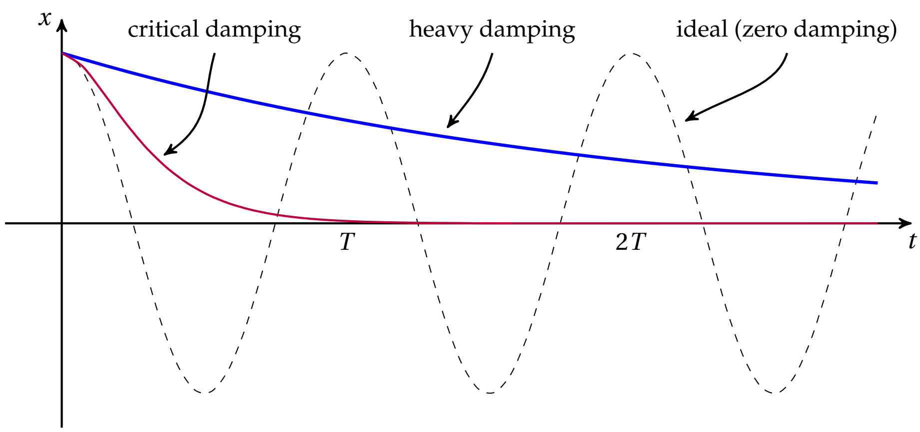
the system will return to the equilibrium position very slowly

this system is said to be **heavily damped**

#### 3.3.3 critical damping

**critical damping** is the border between light damping and heavy damping

it occurs when system returns to equilibrium in *shortest* time without any oscillation




➢ critical damping is desirable in many engineering designs [14]

examples include door-closing mechanism, shock absorbers in vehicles and artillery, etc.

[14]When a damped oscillator is required, critically-damped system provides the quickest approach to equilibrium without overshooting, while lightly-damped system reaches the zero position quickly but continues to oscillate, and heavily-damped system reaches zero position in very long time.

## 3.4 forced oscillations

### 3.4.1 free & forced oscillation

an oscillator moving on its own with no gain or loss of energy is called **free oscillation**

amplitude of the oscillation is constant, its frequency called **natural frequency**

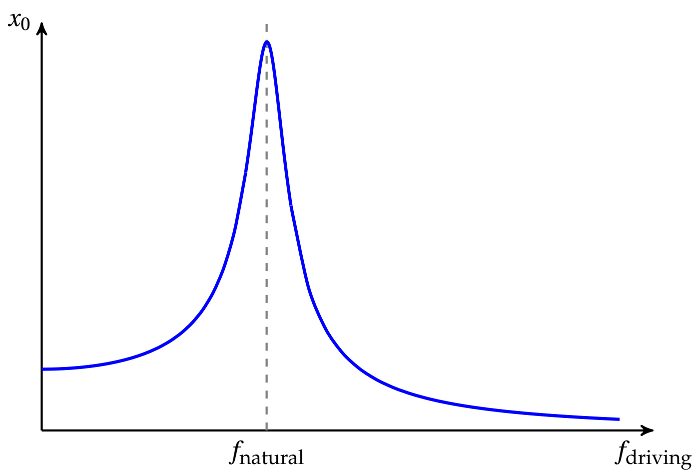
an oscillator may also move under an external driving force, it is **forced oscillation**

frequency of forced oscillator tends to driving frequency after sufficiently long time

### 3.4.2 resonance

for a forced oscillation system, when frequency of driving force $f_{\text{driving}}$ is close to natural frequency $f_{\text{natural}}$, amplitude of oscillator increases rapidly

> when driving frequency of external force equals natural frequency of the system, amplitude of the system becomes maximum, this phenomenon is called **resonance**




resonance is achieved when $f_{\text{driving}} = f_{\text{natural}}$

(amplitude tends to infinity if no damping)

➢ practical application of resonance

- microwave oven – water molecules resonate at microwave frequency and vibrate greatly

- MRI (magnetic resonance imaging) — precession of nuclei resonate at radio frequency, signals are processed to image nuclei of atoms inside a human body in detail
- radio/TV — *RLC* tuning circuits resonate at frequency of signals being received

➢ possible problems caused by resonance

- buildings during earthquake – resonate at frequency of shockwaves and collapse
- car suspension system – going over bumps may give large amplitude vibrations
- bridges and skyscrapers – resonance due to wind conditions

### 3.4.3 damping & resonance

an oscillation system can be subject to both driving force and resistive force

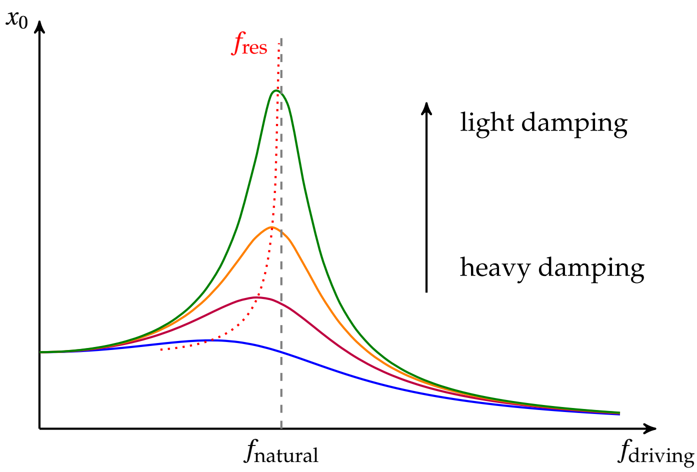
resonance behaviour will be changed by damping effects

➢ damping decreases amplitude of oscillation at all frequencies

greater damping causes resonance peak to become *flatter*

engineering systems are often deliberately damped to minimise resonance effect

➢ damping also shifts resonance frequency (slightly reduced for light damping)




resonance effect for various damping conditions

# Chapter 4

# Ideal Gases

## 4.1 gas molecules

### 4.1.1 motion of gas particles

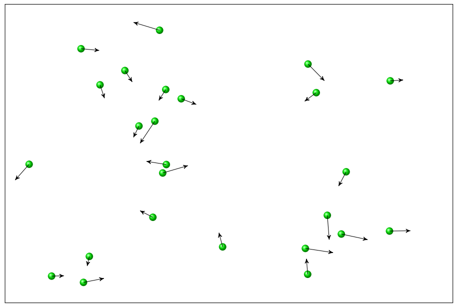
gas consists of a large number of molecules

gas molecules move *randomly* at high speeds

➢ randomness results from *collisions* of fast-moving molecules in the gas

for an individual molecule, its velocity changes constantly as it collides with other molecules

for the gas at any instant, there is a range of velocities for molecules



motion of gas molecules in a container

➢ experimental evidence of random motion: **Brownian motion**

dust or smoke particles in air undergo jerky random motion (viewed through microscope)

this is due to collisions with gas molecules that move randomly

➢ speed of gas molecules depend on temperature

molecules move faster at higher temperature[15]

### 4.1.2 amount of molecules

there are a huge number of molecules in a gas

we introduce **amount of substance** to measure the size of a collection of particles

[15]We will prove this statement later in this chapter.

➢ unit of amount of substance: $[n] = \mathrm{mol}$

one **mole** is defined as the amount carbon-12 atoms in a sample of 12 grams

➢ 1 mole of substance contains $6.02\times 10^{23}$ particles

this number is called **Avogadro constant**: $N_A = 6.02 \times 10^{23}\ \mathrm{mol}^{-1}$ [16]

conversion between number of molecules and amount of substance: $\boxed{N = nN_A}$

➢ it is useful to introduce the notion of molar mass $M$

**molar mass** of a substance is defined as the mass of a given sample divided by the amount of substance: $M = \dfrac{m}{n}$

– amount of substance $= \dfrac{\text{mass of sample}}{\text{molar mass}}$, or $n = \dfrac{m}{M}$

– mass of single molecule $= \dfrac{\text{molar mass}}{\text{Avogadro constant}}$, or $m_0 = \dfrac{M}{N_A}$

**Example 4.1** Find the number of molecules in 160 grams of argon-40 gas.

✍ amount of gas: $n = \dfrac{m}{M} = \dfrac{160\ \mathrm{g}}{40\ \mathrm{g\ mol^{-1}}} = 4.0\ \mathrm{mol}$

number of gas molecules: $N = nN_A = 4.0\ \mathrm{mol} \times 6.02 \times 10^{23}\ \mathrm{mol}^{-1} \approx 2.41 \times 10^{24}$ □

**Question 4.1** Find the mass of a sample of uranium-235 that contains $6.0 \times 10^{20}$ atoms.

### 4.1.3 pressure (qualitative view)

when gas molecules collide with walls of container and rebound, they are acted by a force

by Newton's third law, gas molecules must exert a reaction force on container in return

contributions from many molecules give rise to a pressure

**Example 4.2** If a gas is heated with its volume fixed, how does the pressure change?

✍ at higher temperature, gas molecules move faster

they will collide *harder* and produce a greater force upon each collision

[16] In 2018, IUPAC suggested a new definition of the mole, which is defined to contain exactly $6.02\times 10^{23}$ particles. This new definition fixed numerical value of the Avogadro constant, and emphasized that the quantity 'amount of substance' is concerned with counting number of particles rather than measuring the mass of a sample.

they will also collide more *frequently* with the container

so pressure of the gas will increase □

**Question 4.2** If you pump gas into a bicycle tyre, state and explain how the pressure changes.

**Question 4.3** A fixed amount of gas is allowed to expand at constant temperature, state and explain how the pressure changes.

## 4.2 ideal gas

### 4.2.1 ideal gas equation

a gas that satisfies the equation $\boxed{pV = nRT}$ or $\boxed{pV = NkT}$ at any pressure $p$, any volume $V$, and thermodynamic temperature $T$ is called an **ideal gas**

**molar gas constant**: $R = 8.31\ \mathrm{J\ mol^{-1}\ K^{-1}}$

**Boltzmann constant**: $k = 1.38 \times 10^{-23}\ \mathrm{J\ K^{-1}}$

values of $R$ and $k$ apply for any ideal gas, i.e., they are *universal* constants

➢ recall conversion between number of molecules and amount of substance: $\boxed{N = nN_A}$

we have relation between the constants: $R = kN_A$, or $k = \dfrac{R}{N_A}$

➢ one must use *thermodynamic temperature* in the equation

thermodynamic temperature is measured in kelvins (K), so it is also called the *Kelvin scale*[17]

conversion between Kelvin temperature and Celsius temperature: $\boxed{T_K(\mathrm{K}) \underset{+273}{\overset{-273}{\rightleftharpoons}} T_C(^\circ\mathrm{C})}$

#### real gases

real gas behaves ideally at sufficiently high temperature and low pressure

- at very low temperatures, real gas will condense into liquid or solid
- at very high pressures, intermolecular forces become important

however, under normal conditions (room temperature $T \approx 300$ K and standard atmospheric pressure $p \approx 1.0 \times 10^5$ Pa), there is no significant difference between a real gas and an ideal gas

[17] We will discuss in details about Kelvin scale in §5.1.1 and §5.3.2.

so ideal gas approximation can be used with good accuracy for most of our applications

**Example 4.3** A sealed cylinder of volume of 0.050 $\mathrm{m}^3$ contains 75 g of air. The molar mass of air is 29 g $\mathrm{mol}^{-1}$. (a) Find the air pressure when its temperature is 30°C. (b) The gas is allowed to expand with its pressure fixed. Find the temperature of the gas when the volume doubles.

✍ amount of gas: $n = \dfrac{m}{M} = \dfrac{75}{29} \approx 2.59$ mol

pressure at 30°C: $p = \dfrac{nRT_1}{V_1} = \dfrac{2.59 \times 8.31 \times (30+273)}{0.050} \approx 1.30 \times 10^5$ Pa

pressure fixed, so $V \propto T \quad \Rightarrow \quad \dfrac{T_2}{T_1} = \dfrac{V_2}{V_1} = 2 \quad \Rightarrow \quad T_2 = 2 \times (30+273) = 606\ \mathrm{K} = 333°\mathrm{C}$ □

**Example 4.4** A gas cylinder holding 5000 $\mathrm{cm}^3$ of air at a temperature of 27 °C and a pressure of $6.0 \times 10^5$ Pa is used to fill balloons. Each balloon contains 1000 $\mathrm{cm}^3$ of air at 27 °C and $1.0 \times 10^5$ Pa when filled. (a) Find the initial amount of gas in the cylinder. (b) Find the number of balloons that can be filled.

✍ initial amount of gas in cylinder: $n_0 = \dfrac{p_0 V}{RT} = \dfrac{9.0 \times 10^5 \times 5000 \times 10^{-6}}{8.31 \times (27+273)} \approx 1.203$ mol

final amount of gas in cylinder: $n_{\text{remain}} = \dfrac{pV}{RT} = \dfrac{1.0 \times 10^5 \times 5000 \times 10^{-6}}{8.31 \times (27+273)} \approx 0.201$ mol[18]

amount of gas in each balloon: $n_b = \dfrac{pV_b}{RT} = \dfrac{1.0 \times 10^5 \times 1000 \times 10^{-6}}{8.31 \times (27+273)} \approx 0.040$ mol

number of balloons: $N = \dfrac{n_0 - n_{\text{remain}}}{n_b} = \dfrac{1.203 - 0.201}{0.040} \approx 25$ □

**Example 4.5** A storage cylinder has a volume of $5.0 \times 10^{-4}\ \mathrm{m}^3$. The gas is at a temperature of 300 K and a pressure of $4.0 \times 10^6$ Pa. (a) Find the number of molecules in the cylinder. (b) The gas molecules slowly leak from the cylinder at a rate of $1.6 \times 10^{16}\ \mathrm{s}^{-1}$. Find the time, in days, after which the pressure will reduce by 5.0%.

✍ initial number of molecules: $N_0 = \dfrac{p_0 V}{kT} = \dfrac{4.0 \times 10^6 \times 5.0 \times 10^{-4}}{1.38 \times 10^{-23} \times 300} \approx 4.83 \times 10^{23}$

volume fixed, so $N \propto p \quad \Rightarrow \quad \dfrac{\Delta N}{N_0} = \dfrac{\Delta p}{p_0} = 5.0\%$

number of molecules escaped: $\Delta N = 0.05 \times 4.83 \times 10^{23} \approx 2.42 \times 10^{22}$

time needed: $t = \dfrac{2.42 \times 10^{22}}{1.6 \times 10^{16}} \approx 1.51 \times 10^6\ \mathrm{s} \approx 17.4$ days □

[18] Air will leave the cylinder to fill balloons only if pressure inside the cylinder is higher than pressure of the balloon. When the two pressures become equal, no more balloons can be filled, there will be some air remain in cylinder.

**Question 4.4** Containers $A$ has a volume of $2.5\times10^{-2}$ m$^3$ contains a gas at a temperature of 17°C and pressure of $1.3\times10^5$ Pa and . Another container $B$ of same size holds a gas at same temperature and a pressure of $1.9\times10^5$ Pa. The two containers are initially isolated from each another. (a) Find the total amount of molecules. (b) The two containers are now connected through a tube of negligible volume. Assume the temperature stays unchanged, find the final pressure of the gas.

**Question 4.5** The air in a car tyre can be assumed to have a constant volume of $3.0\times10^{-2}$ m$^3$ . The pressure of this air is $2.8\times10^5$ Pa at a temperature of 25°C. The pressure is to be increased using a pump. On each stroke 0.015 mol of air is forced into the tyre. If gas has a final pressure of $3.6\times10^5$ Pa and final temperature of 28°C. Find the number of strokes of the pump required.

### 4.2.2 empirical laws

historically, the ideal gas law was first stated by *Émile Clapeyron* in 1834:

for a fixed amount of gas, $\boxed{\dfrac{PV}{T}=\text{const}}$

his work was based on the empirical Boyle's law, Charles's law, and Gay-Lussac's law

we will next recover these laws from the ideal gas equation

#### Boyle's law

Boyle's law was discovered by *Robert Boyle* in 1662, based on experimental observations

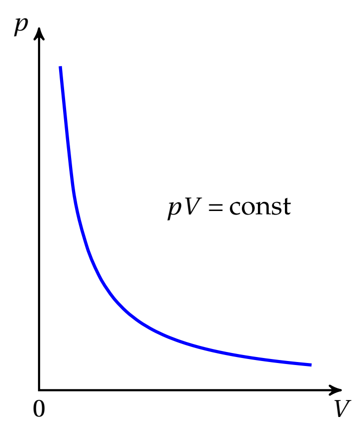
if temperature $T$ remains constant, then

$$\boxed{pV=\text{const}}, \text{ or } \boxed{p\propto\frac{1}{V}}$$

pressure $p$ of gas is inversely proportional to volume $V$

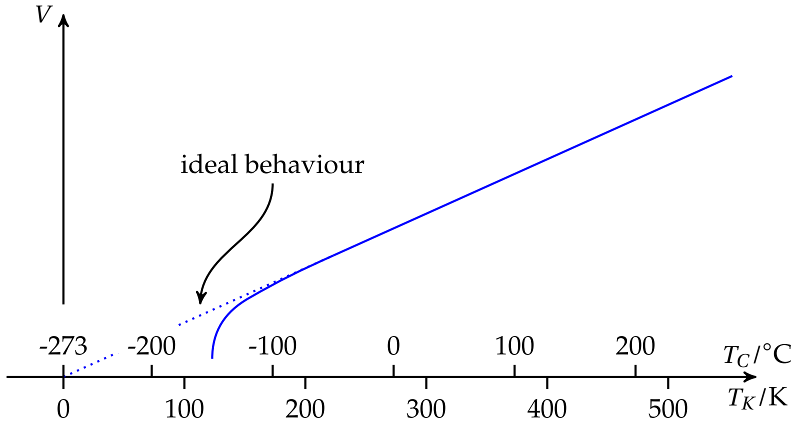
- for a gas with fixed temperature: $p_1V_1=p_2V_2$
- a thermodynamic process for which temperature is kept constant is called an *isothermal* process

$p$-$V$ relation for an isothermal process is shown



**Charles's law**

Charles's law was discovered by *Jacques Charles* in 1787, based on experimental observations




if pressure $p$ remains constant, then: $\boxed{\frac{V}{T} = \text{const}}$, or $\boxed{V \propto T}$

i.e., volume $V$ of gas is directly proportional to its temperature $T$

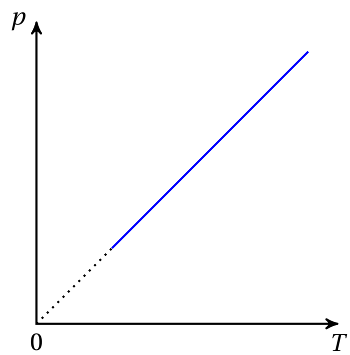
- ➢ proportionality relation only applies if Kelvin scale is used
- ➢ a thermodynamic process for which pressure is kept constant is called an *isobaric* process

  $V$-$T$ relation for an isobaric process is shown
- ➢ Charles's law implies that volume of gas tends to zero at a certain temperature

  historically this is how the idea of *absolute zero* first arose
- ➢ as $T \to 0$, a real gas condenses into solid

  there will be deviation from ideal behaviour (dotted line)

**Gay-Lussac's law**




Gay-Lussac's law was discovered by *Joseph Louis Gay-Lussac* between 1800 and 1802

if volume $V$ remains constant, then

$$\boxed{\frac{p}{T} = \text{const}}, \text{ or } \boxed{p \propto T}$$

i.e., pressure $p$ is directly proportional to temperature $T$

➢ a thermodynamic process for which volume is kept constant

is called an *isochoric* process, or *isometric* process

$p$-$T$ relation for an isochoric process is shown

➢ behaviour of real gas again deviates from ideal behaviour (dotted line) as $T \to 0$

## 4.3 kinetic theory of ideal gases

**kinetic model of gases**: a theory based on microscopic motion of molecules of a gas that explains its macroscopic properties

### 4.3.1 assumptions of ideal gas model

kinetic theory of the ideal gas model is based on the following assumptions:

- gas molecules are in constant *random* motion
- *intermolecular separation* is much greater than size of molecules

  volume of molecules is negligible compared to volume occupied by gas
- *intermolecular forces* are negligible
- collisions between molecules are perfectly *elastic*, i.e., no kinetic energy lost
- molecules travel in straight line between collisions

**Example 4.6** A mass of 20 g helium-4 at a temperature of 37°C has a pressure of $1.2 \times 10^5$ Pa. Each helium-4 atom has a diameter of 280 pm. (a) Find the volume occupied by the gas and the volume of atoms in this gas. (b) Compare the two volumes, suggest whether this gas can be considered as an ideal gas.

✍ number of helium molecules: $N = nN_A = \frac{m}{M} \times N_A = \frac{20}{4.0} \times 6.02 \times 10^{23} \approx 3.01 \times 10^{24}$

volume of gas: $V_{\text{gas}} = \frac{NkT}{p} = \frac{3.01 \times 10^{24} \times 1.38 \times 10^{-23} \times (37 + 273)}{1.2 \times 10^5} \approx 0.107 \text{ m}^3$

volume of one atom: $V_{\text{atom}} = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi \times (140 \times 10^{-12})^3 \approx 1.15 \times 10^{-29} \text{ m}^3$

volume of all atoms: $V_{\text{atoms}} = NV_{\text{atom}} = 3.01 \times 10^{24} \times 1.15 \times 10^{-29} \text{ m}^3 \approx 3.46 \times 10^{-5} \text{ m}^3$

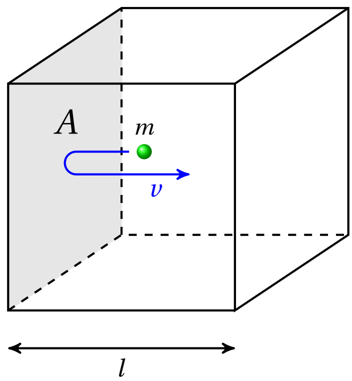
$V_{\text{gas}} \gg V_{\text{atoms}}$, so this gas can approximate to an ideal gas □

### 4.3.2 pressure (quantitative view)

we are ready to derive a formula for pressure due to ideal gas

pressure of gas is due to collision of gas molecules with container

let's first consider the effect of one single molecule moving in one dimension only, and then generalise the result to a gas containing $N$ molecules moving in all three dimensions




one gas molecule moving in 1-D

let's assume this single molecule only moves in $x$-direction (see figure)

change in momentum when colliding with wall: $\Delta P_x = mv_x - (-mv_x) = 2mv_x$[19]

time interval between collisions: $\Delta t = \frac{2l}{v_x}$

average force acting: $F_x = \frac{\Delta P_x}{\Delta t} = \frac{2mv_x}{\frac{2l}{v_x}} = \frac{mv_x^2}{l}$

average pressure: $p_x = \frac{F}{A} = \frac{mv_x^2}{lA} = \quad \Rightarrow \quad p_x = \frac{mv_x^2}{V}$

generalisation to $N$ molecules moving in 3-D

- $N$ molecules so $N$ times the contributions to pressure

  but there is a *distribution* of speeds for $N$ molecules, so should take average of $v^2$
- in three-dimensional space, we have: $v^2 = v_x^2 + v_y^2 + v_z^2$

  but molecules have no preference in any specific direction, so: $\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle = \frac{\langle v^2 \rangle}{3}$

  pressure should be shared equally among three dimensions: $p = p_x = p_y = p_z$

therefore we find the pressure of an ideal gas is given by: $\boxed{p = \frac{Nm\langle v^2 \rangle}{3V}}$

[19]In this section we use $P$ for momentum of a particle and $p$ for pressure of a gas to avoid confusion.

➢ $\langle v^2 \rangle$ is the *mean square velocity* of gas molecules

we can further define r.m.s. (root mean square) velocity: $v_{\text{rms}} = \sqrt{\langle v^2 \rangle}$

gas molecules in random motion so there exists a range of velocities

we cannot tell exact velocity of a specific molecule, but can only tell mean values

➢ $N$ is number of molecules, $m$ is mass of one molecule

then $Nm$ gives total mass of the gas, and $\dfrac{Nm}{V}$ gives gas density $\rho$

we can rewrite the pressure formula as: $\boxed{p = \dfrac{1}{3}\rho \langle v^2 \rangle}$

(pressure depends only on density and mean square speed of molecules)

➢ physical interpretation of the formula

- $N \nearrow \Rightarrow$ more molecules, more collisions $\Rightarrow p \nearrow$
- $m \nearrow \Rightarrow$ greater mass, greater force upon collision $\Rightarrow p \nearrow$
- $v \nearrow \Rightarrow$ strike container harder, also more often $\Rightarrow p \nearrow$
- $V \nearrow \Rightarrow$ spend more time in gas, less frequent collision with container $\Rightarrow p \searrow$

### 4.3.3 kinetic energy

we now have two equations for ideal gases:

$$\begin{cases} pV = nRT\,,\ \text{or}\ \ pV = NkT & \text{ideal gas law} \\ p = \dfrac{Nm\langle v^2 \rangle}{3V} & \text{pressure law} \end{cases}$$

compare the two equations: $pV = \dfrac{1}{3}Nm\langle v^2 \rangle = NkT \quad \Rightarrow \quad m\langle v^2 \rangle = 3kT$

*mean kinetic energy* of a single molecule in a gas is: $\boxed{\langle E_k \rangle = \dfrac{1}{2}m\langle v^2 \rangle = \dfrac{3}{2}kT}$

mean K.E. of ideal gas molecules is *proportional* to its thermodynamic temperature

➢ useful relation for molecular speeds: $\boxed{v_{\text{rms}}^2 \propto T}$

recall our statement in §4.1.1, higher temperature means higher speed for molecules

➢ we only talk about *translational* K.E. here

molecules have this energy because they are moving through space

total kinetic energy may also include *rotational* K.E. and *vibrational* K.E. [20]

[20] There is an important result in classical thermal physics, known as the *equipartition of energy theorem*. It states that the average energy per molecule is $\frac{1}{2}kT$ for each independent *degree of freedom*. A molecule

➢ $\langle E_k \rangle = \dfrac{3}{2}kT$ gives the *mean*, or *average* K.E. per molecule

gas molecules exchange energies with each other upon collisions

for an individual molecule, its K.E. is not a constant

but mean K.E. is constant, which depends on temperature $T$ only

➢ in a mixture of several gases, K.E. is shared *equally* among its components

this is because of repeated collisions between particles

though all molecules have same K.E., heavier molecules will move more slowly

**Example 4.7** Air consists of oxygen ($O_2$, molar mass 32 g mol$^{-1}$) and nitrogen ($N_2$, molar mass 28 g mol$^{-1}$). (a) Calculate the mean translational kinetic energy of these molecules at 300 K. (b) Estimate the typical speed for each type of the molecule.

✍ mean K.E. of single molecule: $\langle E_k \rangle = \dfrac{3}{2}kT = \dfrac{3}{2} \times 1.38 \times 10^{-23} \times 300 \approx 6.21 \times 10^{-21}$ J

$$\langle E_k \rangle = \frac{1}{2}m\langle v^2 \rangle = \frac{3}{2}kT \quad \Rightarrow \quad \frac{1}{2}\frac{M}{N_A}\langle v^2 \rangle = \frac{3}{2}kT \quad \Rightarrow \quad \langle v^2 \rangle = \frac{3kN_AT}{M} = \frac{3RT}{M}$$

for oxygen molecule: $v_{O_2} \approx \sqrt{\dfrac{3 \times 8.31 \times 300}{0.032}} \approx 483$ m s$^{-1}$

for nitrogen molecule: $v_{N_2} \approx \sqrt{\dfrac{3 \times 8.31 \times 300}{0.028}} \approx 517$ m s$^{-1}$ □

**Example 4.8** A cylinder container initially holds a gas of helium-4 at a temperature of 54°C. (a) Find the mean square speed of these helium atoms. (b) If the temperature is raised to 540°C, find the r.m.s. speed of the atoms.

✍ mass of one helium-4 atom: $m = 4\text{u} = 4 \times 1.66 \times 10^{-27} \approx 6.64 \times 10^{-27}$ kg

at 54°C: $\dfrac{1}{2}m\langle v^2 \rangle = \dfrac{3}{2}kT \quad \Rightarrow \quad \langle v^2 \rangle = \dfrac{3kT}{m} = \dfrac{3 \times 1.38 \times 10^{-23} \times (54+273)}{6.64 \times 10^{-27}} \approx 2.04 \times 10^6$ m$^2$ s$^{-2}$

note relation between $v$ and $T$: $\langle v^2 \rangle \propto T \quad \Rightarrow \quad \dfrac{\langle v'^2 \rangle}{\langle v^2 \rangle} = \dfrac{T'}{T} \quad \Rightarrow \quad v'_{\text{rms}} = \sqrt{\dfrac{T'}{T}} \times v_{\text{rms}}$

at 540°C: $v'_{\text{rms}} = \sqrt{\dfrac{540+273}{54+273}} \times \sqrt{2.04 \times 10^6} \approx 2.25 \times 10^3$ m s$^{-1}$ □

**Question 4.6** A fixed mass of gas expands to twice its volume at constant temperature. (a)

can move in three directions, corresponding to three translational degrees of freedom, thus its mean translational kinetic energy is $\frac{3}{2}kT$. For a polyatomic gas (each molecule consists of several atoms), apart from translational motion , it has additional rotational degrees of freedom and different vibrational modes, so its average energy can be calculated by counting the total number of degrees of freedom.

How does its pressure change? (b) How does mean kinetic energy change?

**Question 4.7** In order for a molecule to escape from the gravitational field of the earth, it must have a speed of $1.1 \times 10^6$ m s$^{-1}$ at the top of the atmosphere. (a) Estimate the temperature at which helium-4 atoms could have this speed. (b) Helium atom actually escape from top of the atmosphere at much lower temperatures, explain how this is possible.

# Chapter 5

# Thermodynamics

## 5.1 thermal physics basics

### 5.1.1 temperature scales

- Celsius scale (unit: °C)

  0°C defined as temperature of ice-water mixture

  100°C defined as temperature of boiling water
- Kelvin scale (unit: K)

  0 K (*absolute zero*) is lowest temperature possible
- conversion rule: $T_K(\mathrm{K}) \xrightleftharpoons[+273]{-273} T_C(^\circ\mathrm{C})$
- change of 1°C equals change of 1 K

| $T_C$(°C) | $T_K$(K) | |
|---|---|---|
| 100°C | 373 K | boilng water |
| 25°C | 298 K | room temperature |
| 0°C | 273 K | ice-water mixture |
| -196°C | 77 K | liquid nitrogen |
| -273°C | 0 K | absolute zero |

### 5.1.2 kinetic theory of matter

there are three common states of matter: solid, liquid and gas

they have very different physical properties (density, compressibility, fluidity, etc.)

but deep down, they are all composed of a large number of small molecules

in the **kinetic theory of matter**, we look at microscopic behaviour at molecular level (arrangement, motion, intermolecular forces, separation, etc.)

*microscopic* behaviour of molecules cause differences in *macroscopic* properties of matter

- solid: molecules close together, tightly bonded, vibrate about their positions
- liquid: molecules quite close together, vibrate but has some freedom to move about
- gas: molecules widely separated, free from neighbours, move rapidly

### 5.1.3 specific latent heat

it requires heat energy to *melt* a solid or *boil* a liquid

melting and boiling usually occur at a fixed temperature

thermal energy to cause the change of state at a constant temperature is called *latent heat*

amount of latent heat needed depends on mass of substance: $\boxed{Q = Lm}$

> we define **specific latent heat** ($L$) as the thermal energy required to change the state of *unit* mass of substance with no change in temperature is called

- unit of specific latent heat: $[L] = \mathrm{J \cdot kg^{-1}}$
- specific latent heat is an *intensive* property

  i.e., $L$ does not depend on size or shape of sample, $L$ depends on type of substance only
- for melting, $L$ is called *specific latent heat of fusion*

  for boiling, $L$ is called *specific latent heat of vaporisation*
- latent heat is related to breaking bonds and increasing intermolecular separation

  vaporisation requires larger increase in particle separation than fusion

  for a given substance, $L_{\text{vapour}} > L_{\text{fuse}}$

**Example 5.1** A 3.0 kW electric kettle contains 0.5 kg of water already at its boiling point. Neglecting heat losses, determine how long it takes to boil dry. ($L_{\text{water}} = 2.26 \times 10^6\ \mathrm{J\ kg^{-1}}$)

✍ heat required: $Q = mL = 0.50 \times 2.26 \times 10^6 = 1.13 \times 10^6$ J

time needed: $t = \dfrac{Q}{P} = \dfrac{1.13 \times 10^6}{3.0 \times 10^3} \approx 380\ \mathrm{s} \approx 6.3\ \mathrm{min}$ □

**Example 5.2** A student measures specific latent heat of fusion for ice. He uses an electric heater to melt ice but the insulation is not perfect. The experiment is carried out twice, with the heater operating at different powers. Use the data table to calculate specific latent heat of fusion.

| | Power (W) | time interval (min) | mass of ice melted (g) |
|---|---|---|---|
| test 1 | 60 | 3.0 | 40.4 |
| test 2 | 90 | 3.0 | 56.6 |

✍ there exists heat gain from surroundings, so effective power $P_{\text{eff}} = P_{\text{heater}} + P_{\text{sur}}$

heat energy to melt ice: $Q = mL = (P_{\text{heater}} + P_{\text{sur}})\,t$

$$\begin{cases} 40.4 \times L = (60 + P_{\text{sur}}) \times 3.0 \times 60 \\ 56.6 \times L = (90 + P_{\text{sur}}) \times 3.0 \times 60 \end{cases} \Rightarrow \begin{cases} L \approx 333\ \text{J g}^{-1} \\ P_{\text{sur}} \approx 14.8\ \text{W} \end{cases} \qquad \square$$

**Question 5.1** A student designs an experiment to determine the specific latent heat of fusion $L$ of ice. Some ice at 0°C is heated with an electric heater. The experiment is carried out twice and the following data are obtained.

| | energy supply from heater (J) | time interval (min) | mass of ice melted (g) |
|---|---|---|---|
| heater off | 0 | 10.0 | 14.3 |
| heater on | 21000 | 5.0 | 70.0 |

(a) Suggest why two sets of readings are taken. (b) Find specific latent heat of fusion for ice.

### 5.1.4 specific heat capacity

heating a substance could cause an increase in its temperature

heat required is proportional to its mass $m$ and temperature change $\Delta T$: $\boxed{Q = cm\Delta T}$

we define **specific heat capacity** ($c$) as the thermal energy required per unit mass of substance to cause an increase of one unit in its temperature

- unit of specific heat capacity: $[c] = \text{J kg}^{-1}\ \text{K}^{-1}$ or $\text{J kg}^{-1}\ {}^\circ\text{C}^{-1}$
- $c$ is also an *intensive* property, i.e., independent of size or shape of the sample

**Example 5.3** A block of 30 g ice at −20°C is added to a large cup of 270 g water at 80°C. Assume there is no energy lost, what is the final temperature of the mixture? (data: specific heat capacity of water is 4200 J kg$^{-1}$ K$^{-1}$, specific heat capacity of ice is 2100 J kg$^{-1}$ K$^{-1}$, specific latent heat of ice is $3.3 \times 10^5$ J kg$^{-1}$.

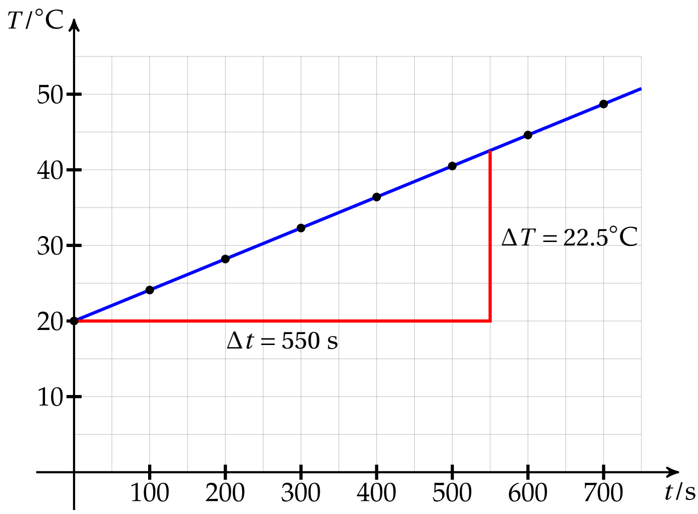
✍ energy lost by hot water = energy gain by ice cube

$$\underbrace{4200 \times 0.27 \times (80 - T)}_{95\,^\circ\text{C water} \to T\,^\circ\text{C water}} = \underbrace{2100 \times 0.030 \times [0 - (-20)]}_{-20\,^\circ\text{C ice} \to 0\,^\circ\text{C ice}} + \underbrace{3.3 \times 10^5 \times 0.030}_{0\,^\circ\text{C ice} \to 0\,^\circ\text{C water}} + \underbrace{4200 \times 0.030 \times (T - 0)}_{0\,^\circ\text{C water} \to T\,^\circ\text{C water}}$$

$$90720 - 1134T = 1260 + 9900 + 126T$$

$$T = \frac{90720 - 1260 - 9900}{1134 + 126} \approx 63^\circ\text{C} \qquad \square$$

**Example 5.4** A 1.00 kg aluminium block is heated using an electrical heater. The current in the heater is 4.2 A and the p.d. across is 12 V. Measurements of the rising temperature are represented by the graph. Determine specific heat capacity of aluminium.




✍ energy supplied: $Q = cm\Delta T \quad \Rightarrow \quad IV\Delta t = cm\Delta T \quad \Rightarrow \quad c = \dfrac{IV}{m\frac{\Delta T}{\Delta t}}$

$\dfrac{\Delta T}{\Delta t}$ is gradient of fitting line: $\dfrac{\Delta T}{\Delta t} = \dfrac{22.5}{550} \approx 4.09 \times 10^{-2}\ {}^\circ\mathrm{C\ s^{-1}}$

specific heat capacity: $c = \dfrac{4.2 \times 12}{1.00 \times 4.09 \times 10^{-2}} \approx 1230\ \mathrm{J\ kg^{-1}\ {}^\circ C^{-1}}$ □

**Question 5.2** A mixture contains 5% silver and 95% of gold by weight. Some gold is melted and the correct weight of silver is added. The initial temperature of silver is 20°C. Use the data to calculate the initial temperature of gold so that the final mixture is at melting point of gold.

| | silver | gold |
|---|---|---|
| melting point (K) | 1240 | 1340 |
| specific heat capacity (solid or liquid) (J kg$^{-1}$ K$^{-1}$) | 235 | 129 |
| specific latent heat of fusion (kJ kg$^{-1}$) | 105 | 628 |

## 5.2 internal energy

we now consider the total energy within a thermodynamic system

molecules in a system undergo random motion, so they have kinetic energy there are potential energy between molecules due to intermolecular interaction

**internal energy** is defined as the sum of random kinetic energy of molecules and potential energy between molecules: $\boxed{U = E_k + E_p}$

- internal energy is a *state function* of the system

  it only depends on current state of system, not on process to arrive at this state

### 5.2.1 kinetic energy

- internal energy counts K.E. due to random motion at molecular level

  K.E. of macroscopic motion of the system as a whole is not included
- mean K.E. of molecules is directly proportional to temperature: $\boxed{E_k \propto T}$

  K.E. of molecules depends on temperature only

  higher temperature means molecules move faster, vibrate more intensively, etc.

### 5.2.2 potential energy

- internal energy counts P.E. due to force fields *within* the system

  P.E. of the system as a whole due to *external* force fields is not included
- P.E. between molecules depends on intermolecular separation and chemical bonding

  in general, greater intermolecular separation means greater P.E.[21]: $\boxed{r \nearrow \Leftarrow E_p \nearrow}$

  breaking intermolecular bonds also causes an increase in P.E.

  mean P.E. of gas > mean P.E. of liquid > mean P.E. of liquid solid

### 5.2.3 internal energy of ideal gas

for ideal gas, mean K.E. of one molecule: $E_k = \frac{3}{2}kT$

[21]Intermolecular separation does not necessarily increase during melting processes. A typical counter example is melting of ice into water, for which intermolecular separation actually decreases (density of water > density of ice), but potential energy of the system will still increase because hydrogen bonds between $H_2O$ molecules are broken.

there is no intermolecular force, so P.E. of ideal gas is defined to be zero: $E_p = 0$

internal energy per molecule: $U = E_k + E_p = \frac{3}{2}kT$

hence internal energy of ideal gas is purely kinetic and directly proportional to temperature

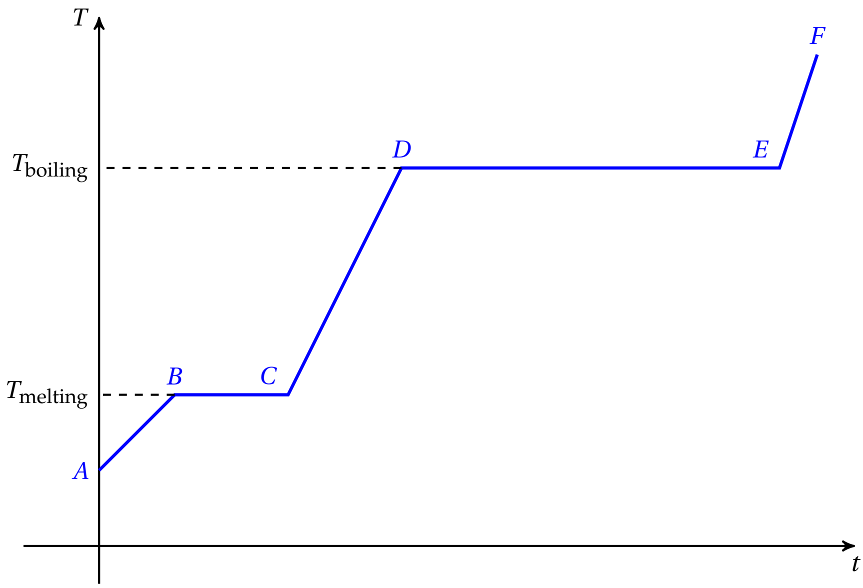
total internal energy of the gas: $U_{\text{gas}} = NU \quad \Rightarrow \quad \boxed{U_{\text{gas}} = \frac{3}{2}NkT}$

### 5.2.4 change of states

consider a substance being heated from its solid state

it will melt into a liquid and further vaporise into its gaseous state

we now look into the changes of internal energy during each stage




- $AB$ (solid state): $T \nearrow \Rightarrow E_k \nearrow$, greater vibration for solid particles

  but no (significant) change in mean separation[22] $\Rightarrow$ no change in P.E.
- $BC$ (*melting*): latent heat goes into breaking intermolecular bonds, $r \nearrow \Rightarrow E_p \nearrow$

  but melting occurs at constant temperature[23] $\Rightarrow$ no change in K.E.

[22]A typical solid material expands when it is heated, so intermolecular separation will increase slightly.

[23]Here we talk about *pure substance*, which changes from solid into liquid at a particular temperature, called the *melting point*. But for a *mixture* of substances, melting may occur over a *range* of temperatures. It is also possible for a substance to decompose before they change states.

- *CD* (liquid state): $T \nearrow \Rightarrow E_k \nearrow$, greater vibration and free motion

  but no (significant) change in mean separation $\Rightarrow$ no change in P.E.
- *DE* (*boiling*): molecules break free, $r \nearrow \Rightarrow E_p \nearrow$

  boiling occurs at constant temperature[24] $\Rightarrow$ no change in K.E.
- *EF* (gas state): $T \nearrow \Rightarrow E_k \nearrow$ particles move even faster

  particles completely separated, no intermolecular force, so constant $E_p = 0$

**Question 5.3** For a particular substance, why is the specific latent heat of vaporisation much greater than the specific latent heat of fusion?

**evaporation**

liquid changes into gas without boiling $\longrightarrow$ **evaporation**

particles move randomly, i.e., they move at various speeds

some molecules move fast enough to break free

➢ *cooling effect*: evaporation causes a decrease in temperature of the liquid

most energetic molecules escaped, those remain in the liquid have less energy, $E_k \searrow \Rightarrow T \searrow$

➢ rate of evaporation increases with temperature, surface area of liquid

➢ different between boiling and evaporation

| | boiling | evaporation |
|---|---|---|
| occurrence | throughout the liquid | at surface only |
| temperature | occur at boiling point | occur at any temperature |
| bubble formation | bubbles formed | no bubbles |
| rate of process | fast | slow |

[24] Again we only concern pure substances. Mixtures that boil over a range of temperatures or substance decompose before phase transition are not considered here.

#### 5.2.5 first law of thermodynamics

internal energy of a system changes upon heat transfer or doing work

**first law of thermodynamics** states that the increase in internal energy equals sum of heat supply to the system and work done on the system: $\boxed{\Delta U = Q + W}$

- first law of thermodynamics is an extension of the law of conservation of energy
- sign conventions for $Q$ and $W$
  - $Q > 0$ if heat supplied to system
  - $Q < 0$ if heat released by system to surroundings
  - $W > 0$ if work done *on* system by external

    i.e., if system is compressed and volume decreases, then $W > 0$
  - $W < 0$ if system does work *against* surroundings

    i.e., system expands and volume increases, then $W < 0$
- amount of heat energy: $Q = \begin{cases} cm\Delta T & \text{(if no change of state)} \\ Lm & \text{(during change of state)} \end{cases}$
- amount of work is related to pressure and change of volume

  if volume changed at *constant* pressure, then $W = F\Delta s = pA\Delta s \Rightarrow \boxed{W = p\Delta V}$[25]

  if no change of volume, then no work is done

**Example 5.5** A gas is heated by supplying it with 25 kJ of energy. The gas expands so that the volume increases by 0.10 $m^3$. Assume the gas has a fixed pressure of 150 kPa during the process. Calculate the change in internal energy.

✍ amount of work done: $W = p\Delta V = 150 \times 0.10 = 15$ kJ

but gas expands means work is done against surroundings, so this is negative work

change in internal energy: $\Delta U = Q + W = (+25) + (-15) = +10$ kJ □

**Example 5.6** Use the idea of internal energy and the first law of thermodynamics, explain why boiling water requires heat supply.

[25] If pressure changes with volume during a thermodynamics process, then work done $W = \int p\mathrm{d}V$. Alternatively, we can evaluate the area under a $p$-$V$ graph to find the work done.

✍ boiling occurs at constant temperature, so $\Delta E_k = 0$

but separation between molecules increased, so $\Delta E_p > 0$

by definition, internal energy $U = E_k + E_p$, so $\Delta U > 0$

during boiling, there is an increase in volume, so work against surroundings, $W < 0$

recall first law of thermodynamics $\Delta U = Q + W$, must have $Q > 0$

this means heat must be supplied for boiling processes □

**Example 5.7** When you pump up a bicycle tyre, the temperature of air inside the tyre will go up. Explain why this happens using the first law of thermodynamics.

✍ pumping up tyre involves compressing gas, so positive work is done: $W > 0$

for each stroke, there is little time for heat transfer, so $Q \approx 0$

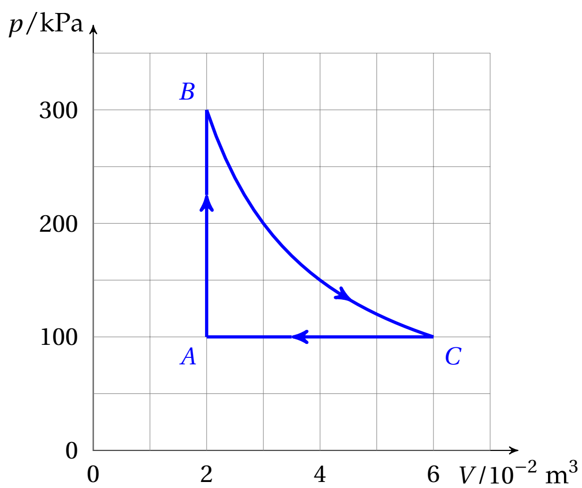
according to first law of thermodynamics $\Delta U = Q + W \quad \Rightarrow \quad \Delta U > 0$

by definition, internal energy $U = E_k + E_p \quad \Rightarrow \quad \Delta U = \Delta E_k + \Delta E_p > 0$

but for gas, there is negligible intermolecular force, so $\Delta E_p = 0$, then must have $\Delta E_k > 0$

K.E. of molecules is proportional to temperature, higher K.E. so higher temperature □

**Example 5.8** An ideal gas of 0.080 mol is initially at state $A$ and then undergoes a cycle $ABCA$. The variation of its pressure $p$ with its volume $V$ is shown on the graph.




Temperature of state $A$ is 300 K. The magnitude of work on gas from state $B$ to $C$ is 6570 J.

For each stage $A \to B$, $B \to C$ and $C \to A$ during the cycle, determine work done and heat supply to the gas, and also find the change in internal energy.

✍ work done depends on change in volume

$A \to B$: no change in volume, so $W_{AB} = 0$

$B \to C$: $|W_{BC}| = 6570$ J, but expansion implies $W < 0$, so $W_{BC} = -6570$ J

$C \to A$: $|W_{CA}| = p\Delta V_{CA} = 1 \times 10^5 \times (6-2) \times 10^{-2} \quad \Rightarrow \quad W_{CA} = +4000$ J (compression so $W > 0$)

change in internal energy of ideal gas depends on change in temperature

$A \to B$: same $V$ but $p_B = 3p_A$, so $T_B = 3T_A = 900$ K

$$\Delta U_{AB} = \frac{3}{2}Nk\Delta T_{AB} = \frac{3}{2} \times 0.80 \times 6.02 \times 10^{23} \times 1.38 \times 10^{-23} \times (900-300) \approx +5980 \text{ J}$$

$B \to C$: note that $p_B V_B = p_C V_C$, so $T_B = T_C$, no change in temperature, so $\Delta U_{BC} = 0$

$C \to A$: $\Delta U_{CA} = \frac{3}{2}Nk\Delta T_{CA} = \frac{3}{2} \times 0.80 \times 6.02 \times 10^{23} \times 1.38 \times 10^{-23} \times (300-900) \approx -5980$ J

for cycle $ABCA$, same initial and final state, so total change in internal energy must be zero

one can check that $\Delta U_{\text{cycle}} = \Delta U_{AB} + \Delta U_{BC} + \Delta U_{CA} = 0$

to find supply of thermal energy, we apply first law of thermodynamics: $\Delta U = Q + W$

$A \to B$: $+5980 = Q_{AB} + 0 \quad \Rightarrow \quad Q_{AB} = +5980$ J

$B \to C$: $0 = Q_{BC} + (-6570) \quad \Rightarrow \quad Q_{BC} = +6570$ J

$C \to A$: $-5980 = Q_{CA} + (+4000) \quad \Rightarrow \quad Q_{CA} = -9980$ J

the table below summarises all energy changes during the cycle $ABCA$

| change | $W$/J | $Q$/J | $\Delta U$/J |
|---|---|---|---|
| $A \to B$ | 0 | +5980 | +5980 |
| $B \to C$ | -6570 | +6570 | 0 |
| $C \to A$ | +4000 | -9980 | -5980 |

□

**Question 5.4** Show that when $n$ mol of gas is heated at a fixed volume, thermal energy required to raise the temperature by 1.0 K is $nR$.

**Question 5.5** Two identical balloons $A$ and $B$ hold the same amount of gas at the same initial temperature. They are given the same amount of heat. Suppose volume of $A$ is fixed, while $B$ is allowed to expand, compare the final temperatures of the gases in the two balloons.

## 5.3 temperature

### 5.3.1 temperature & thermal energy

➢ temperature can be considered as a *relative* measure of thermal energy

temperature can tell the *direction* of thermal energy flow

heat always (spontaneously) flows from high temperature regions to colder regions[26]

➢ if two objects in contact have the same temperature, then there is no net heat transfer

the two objects are said to be in **thermal equilibrium**

➢ if two systems $A$ and $B$ are each in thermal equilibrium with a third system $C$, $A$ and $B$ are also in thermal equilibrium, this is called the **zeroth law of thermodynamics** [27]

**Question 5.6** A student thinks that temperature measures the amount of heat in an object. Suggest why this statement is incorrect with examples.

### 5.3.2 absolute zero

mean K.E. of molecules is a microscopic description of temperature $T$

minimum K.E. occurs if molecules do not move at all (completely frozen)[28]

this corresponds to the lowest possible temperature, called **absolute zero**

➢ **Kelvin scale of thermodynamic temperature** is defined based on absolute zero as 0 K[29]

[26]This is the consequence of the *second law of thermodynamics*, which is concerned with the direction of natural processes. The law states that the total *entropy*, a quantity that counts the number of microstates of a system, of an isolated system can never decrease over time. You will learn more about entropy if you study A-Level chemistry.

[27]This law is important for the formulation of thermal physics. The physical meaning of the law was expressed by Maxwell: "All heat is of the same kind." The zeroth law allows us to give the mathematical definition of temperature.

[28]In the *classical* description, there is no reason not allowing a molecule to cease motion. However, due to *quantum mechanical effects*, kinetic energy of a system cannot be zero even at absolute zero.

[29]More precisely, 0 K for absolute zero and 273.16 K for water triple point.

- conversion rule between Celsius scale and Kelvin scale: $\boxed{T_K(\text{K}) \underset{+273.15}{\overset{-273.15}{\rightleftharpoons}} T_C(^\circ\text{C})}$ [30]
- Kelvin scale is said to be an *absolute scale*

  zero of Kelvin scale does not depend on property of a specific substance

  in contrast, zero of Celsius scale is based on properties of water
- it is impossible to remove any more energy from a system at 0 K (or -273.15 °C)

  but there is no practicable means to bring a physical system to exactly 0 K [31]

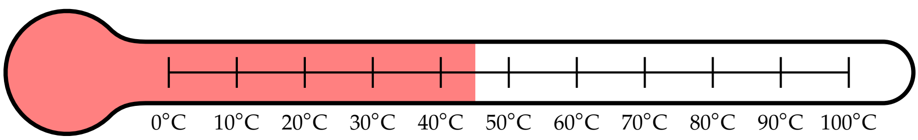
### 5.3.3 thermometer

a **thermometer** is a device which can be used to measure temperature

**liquid-in-glass thermometer**

basic principle: liquid expands in volume at higher temperature

examples include alcohol thermometer, mercury-in-glass thermometer, etc.




**resistance temperature detectors (RTD)**

basic principle: resistance of electronic element changes with temperature

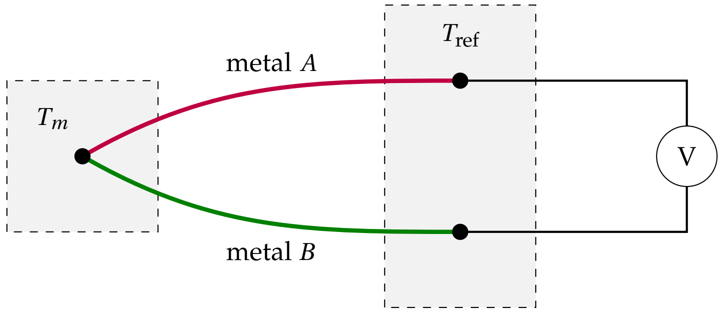
metal wires and thermistor are both used in RTD elements

**thermocouple**

basic principle: difference in temperature can produce a *thermoelectric voltage* across junctions, thermocouple measures temperatures by means of this voltage

[30] The numerical value of 273.15 will only be quoted in this section. For everywhere else in the notes, we will use the less precise value 273 for simplicity.

[31] This is known as the *third law of thermodynamics*, which states that it is impossible, no matter how idealized, to reduce the temperature of any closed system to absolute zero in a finite number of operations.




typical configuration of a thermocouple unit

two pieces of different metal wires are joined at their ends

if there exists a temperature difference between the ends, a *thermoelectric voltage* is developed

this voltage depends on temperature difference, captured by some characteristic function

in practice, we place the *measurement* junction in an environment of unknown temperature

the other end, or the *reference* junction, is at a known temperature

temperature difference is deduced from the voltage reading

hence the desired temperature can be determined

➢ features of a thermometer:

- *range*: whether the thermometer can measure very low or very high temperatures
- *sensitivity*: whether a small change in temperature can be detected
- *response time*: whether changes in temperature can be immediately measured
- *linearity*: whether changes in temperature are proportional to changes in output

| | liquid-in-glass | RTD wire | thermistor | thermocouple |
|---|---|---|---|---|
| valid range | narrow | wide | narrow | very wide |
| sensitivity | low | fair | high | high |
| response time | slow | fast | fast | very fast |
| linearity | good | good | limited | non-linear |

# Chapter 6

# Electrostatics

## 6.1 electric forces

### 6.1.1 Coulomb's law

charged objects will attract or repel one another through the electric force

**Coulomb's law** states that the electric force between two electrically charged particles is proportional to their charges and inversely proportional to the square of their separation: $\boxed{F = \dfrac{Qq}{4\pi\epsilon_0 r^2}}$

$\epsilon_0 = 8.85 \times 10^{-12}\ \mathrm{C^2\ N^{-1}\ m^{-2}}$ is the *permittivity of free space*

$k = \dfrac{1}{4\pi\epsilon_0} = 8.99 \times 10^9\ \mathrm{N\ m^2\ C^{-2}}$ is a useful constant for calculations

this law was first published by French physicist *Charles Augustin de Coulomb* in 1785

- charges $Q$, $q$ in Coulomb's law are *point charges*
- for uniformly charged spheres, they can be thought as point charges

  separation $r$ is taken to be centre-to-centre distance
- symbolically, sign of $F$ can tell direction of the electric force

  for like charges (both positive or both negative), $Q_1Q_2 > 0 \Rightarrow F > 0 \Rightarrow$ *repulsion*

  for opposite charges (one positive and one negative), $Q_1Q_2 < 0 \Rightarrow F < 0 \Rightarrow$ *attraction*

**Example 6.1** The hydrogen atom has a radius of about 53 pm. Estimate the electric force between the proton and the orbiting electron.

✍ $$F = \frac{Q_pQ_e}{4\pi\epsilon_0 r^2} = 8.99 \times 10^9 \times \frac{(1.60 \times 10^{-19})^2}{(53 \times 10^{-12})^2} \approx 8.2 \times 10^{-8}\ \mathrm{N}$$ □

**Question 6.1** Two protons are separated by a distance $r$. Find the ratio of the electric force to the gravitational force between them.

### 6.1.2 electric fields

to explain how charges affect each other at a distance, we introduce notion of *electric fields*

**electric field** is a region of space where a charged object is acted by a force

any charge $Q$ (or several charges) can produce an electric field

any test charge $q$ within the field will experience an electric force

Next, we will introduce the concepts of *electric field strength* and *electric potential*, and see how they are related to the force acting on a charged object and the potential energy it possesses.

You might have noticed that Coulomb's law for electrostatic forces and Newton's law of gravitation are both *inverse square laws*, it turns out that electric fields are very similar to gravitational fields in various aspects.

## 6.2 electric field strength

### 6.2.1 electric field strength

**electric field strength** is defined as electric force per unit positive charge: $\boxed{E = \dfrac{F}{q}}$

- unit of $E$: $[E] = \mathrm{N\,C^{-1}} = \mathrm{V\,m^{-1}}$[32]
- field strength due to an isolated source of *point* charge $Q$

  a small test charge $q$ at distance $r$ is acted by a force: $F = \dfrac{Qq}{4\pi\epsilon_0 r^2}$

  field strength at this point: $E = \dfrac{F}{q} \quad \Rightarrow \quad \boxed{E = \dfrac{Q}{4\pi\epsilon_0 r^2}}$

  the field is produced by $Q$, so field strength only depends on the source $Q$
- if the source is a charged *sphere* of radius $R$ with uniform charge distribution

  viewed from *outside* the sphere, it acts like a point charge concentrated at the centre [33]

[32] You will later find in §6.4.3 the deeper reason why $\mathrm{V\,m^{-1}}$ is also a reasonable unit for field strength.
[33] A brief explanation is given in Example 6.4.

therefore, $E = \dfrac{Q}{4\pi\epsilon_0 r^2}$ also holds for field due to charged sphere at $r > R$[34]

where $r$ is the distance from the point of interest to *centre* of the sphere

➢ field strength $E$ is a *vector* quantity, it has a direction

to compute combined field strength due to several sources, should perform *vector sum* of contributions from each individual

➢ direction of field strength depends on the source charge $Q$

for positive source ($Q > 0$): field points away from the source

for negative source ($Q < 0$): field points towards the source

➢ electric force on a charge $q$ can be found if field strength $E$ is known

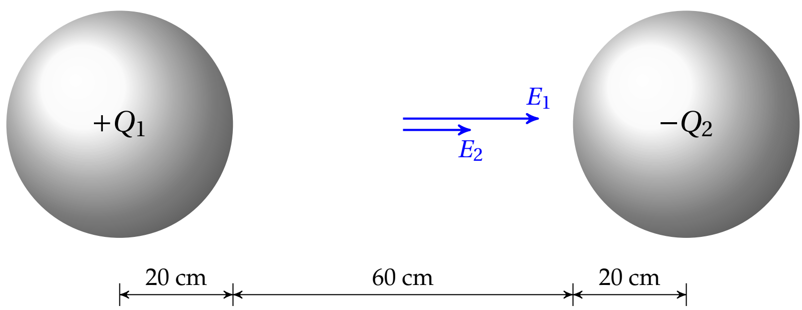
magnitude of electric force: $F = Eq$

direction of force: same direction as $E$ if $q > 0$, but opposite to $E$ if $q < 0$

**Example 6.2** A *Van de Graaff generator* produces sparks when its surface electric field strength $4.0 \times 10^4$ V cm$^{-1}$. If the diameter of the sphere is 40 cm, what is the charge on it?

✍ $$E = \frac{Q}{4\pi\epsilon_0 r^2} \quad\Rightarrow\quad Q = 4\pi\epsilon_0 E r^2 = 4\pi \times 8.85 \times 10^{-12} \times 4.0 \times 10^6 \times 0.20^2 \approx 1.8 \times 10^{-5}\ \text{C}$$ □

**Example 6.3** Two identical metal spheres of radius 20 cm carry charges $+2.0\ \mu$C and $-1.0\ \mu$C respectively. There is a 60 cm gap between them. (a) Find the electric field strength midway along the line joining their centres. (b) A dust particle carrying a charge of $-1.3 \times 10^{-8}$ C is at this position. Find the electric force it experiences.




✍ field strengths due to the two spheres are in same direction

$$E = E_1 + E_2 = \frac{1}{4\pi\epsilon_0}\left(\frac{Q_1}{r_1^2} + \frac{Q_2}{r_2^2}\right) = 8.99 \times 10^9 \times \left(\frac{2.0 \times 10^{-6}}{0.25^2} + \frac{1.0 \times 10^{-6}}{0.25^2}\right) \approx 4.32 \times 10^5\ \text{N C}^{-1}$$

[34]For electric field strength *inside* a conducting sphere, detailed discussions are given in §6.4.2.

field strength points to the right

force on dust particle: $F = Eq = 4.32 \times 10^5 \times 1.4 \times 10^{-8} \approx 5.6 \times 10^{-3}$ N

dust particle is negatively-charged means force is opposite to field strength

so force on dust particle acts to the left □

**Question 6.2** When the charge on the Van de Graaff generator is $4.0 \times 10^{-7}$ C, the electric field strength at the sphere's surface is $2.4 \times 10^6$ V m$^{-1}$. Determine the additional charge added to the sphere if the field strength at the surface becomes $3.0 \times 10^6$ V m$^{-1}$.

**Question 6.3** Two positively charged particles $A$ and $B$ are situated in a vacuum. Point P lies on the line joining the centres of the two spheres and is a distance $x$ from $A$. Sketch the variation with $x$ of electric field strength $E$ due to the two particles.

### 6.2.2 electric field lines

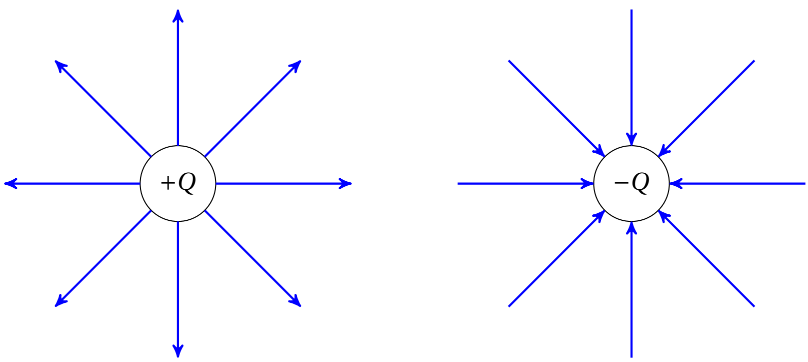
we can use **electric field lines** to visualize an electric field

- *arrows* of field lines show direction of the field

field lines always tend to leave positive charge, and end up at negative charges

- *density* or spacing of lines show strength of the field

**Example 6.4** Sketch the electric field around a positively-charged sphere or a negatively charged sphere, and explain why they can be considered as point charges.

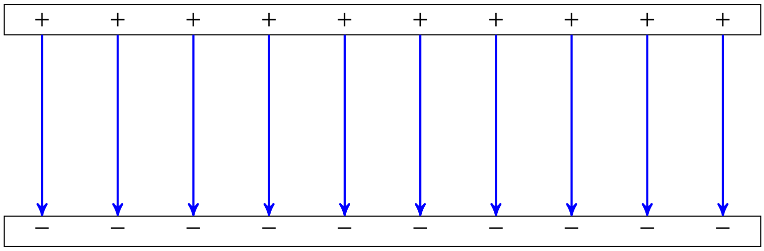



field lines of either case are *radial*, i.e., perpendicular to surface

field lines appear to start from or converge towards centre of the sphere

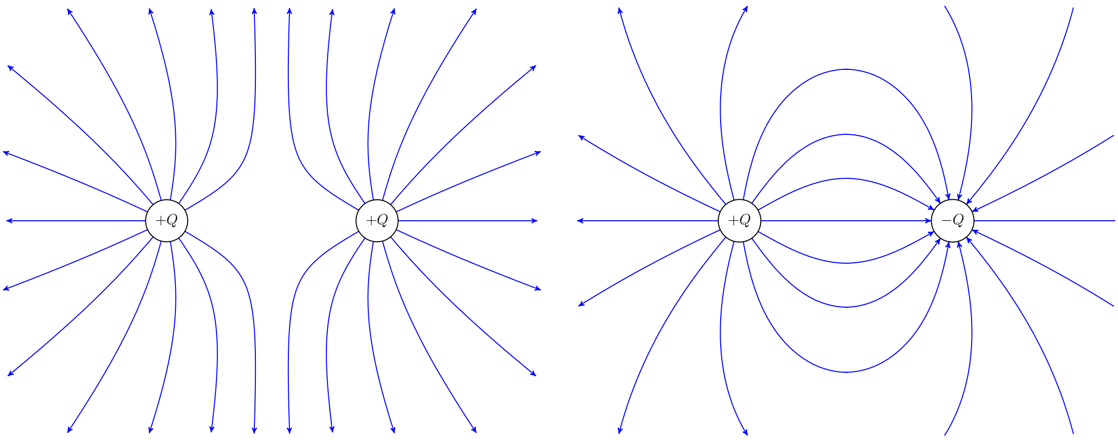
so charged spheres act like point charges □

**Example 6.5** Field lines between two oppositely-charged large metal plates.



field lines are *parallel* and equally spaced, so this is a *uniform* electric field □

**Example 6.6** Field pattern due to two charges of equal magnitude.




two positive charges two opposite charges □

## 6.3 potential & potential energy

### 6.3.1 electric potential energy

gain/loss in **electric potential energy** is defined as work done against/by electric force

(compare everything in this section with what you have learned about gravitational P.E.!)

let's start to derive the electrical P.E. between two charges $Q$ and $q$ separated by $r$

again we define $E_p = 0$ at $r = \infty$ (choice of zero potential energy, no force so no P.E.), then

> **electric potential energy** is equal to the work done by electric force to bring a charge to a specific position from *infinity*

moving a test charge $q$ from $r = \infty$ to a distance of $r$ from $Q$


work done by electric force: $W = \int_\infty^r F\,dr = \int_\infty^r \frac{Qq}{4\pi\epsilon_0}\frac{\mathrm{d}x}{x^2} = -\frac{Qq}{4\pi\epsilon_0 x}\Big|_\infty^r = -\frac{Qq}{4\pi\epsilon_0 r}$

since $\Delta E_p = -W$, we find: $E_p(r) - E_p(\infty) = \frac{Qq}{4\pi\epsilon_0 r}$

but $E_p(\infty) = 0$, so electric P.E. between two charges $Q$ and $q$ is $\boxed{E_p(r) = \frac{Qq}{4\pi\epsilon_0 r}}$

➢ as $r \to \infty$, $E_p \to 0$, this agrees with our definition for zero P.E. point

➢ for like charges, $Qq > 0$, so $E_p > 0$

to bring like charges closer, work must be done to overcome their *repulsion*, P.E. increases

minimum P.E. $E_p(\infty) = 0$ at infinity, so positive P.E. at finite $r$

➢ for opposite charges, $Qq < 0$, so $E_p < 0$

to pull opposite charges apart, work must be done to overcome their *attraction*, P.E. increases

maximum P.E. $E_p(\infty) = 0$ at infinity, so negative P.E. at finite $r$

➢ electric P.E. is a *scalar quantity*, sign is important

repulsion implies positive P.E., and attraction implies negative P.E.

sign of P.E. is hidden in polarities of charges

**Example 6.7** To gain information about the gold nucleus ($^{197}_{79}$Au), we fire $\alpha$-particles ($^4_2\alpha$) towards a thin gold foil. The size of a typical nucleus is about $10^{-14}$m, what is the minimum initial speed for $\alpha$-particles so that radius of gold nucleus can be determined?

✍ as $\alpha$-particle approaches the nucleus, it slows down due to the repulsive interaction

kinetic energy decreases and electric potential energy increases

if it gets close enough to the nucleus before coming to a stop, nuclear radius can be estimated

$$\text{K.E. loss} = \text{P.E. gain} \quad \Rightarrow \quad \frac{1}{2}mu^2 - \underbrace{\frac{1}{2}mv^2}_{0} = E_p(r) - \underbrace{E_p(\infty)}_{0} \quad \Rightarrow \quad \frac{1}{2}mu^2 = \frac{Qq}{4\pi\epsilon_0 r}$$

$$\frac{1}{2} \times 4 \times 1.66 \times 10^{-27} \times u^2 = \frac{79 \times 1.60 \times 10^{-19} \times 2 \times 1.60 \times 10^{-19}}{4\pi \times 8.85 \times 10^{-12} \times 10^{-14}} \quad \Rightarrow \quad u \approx 3.3 \times 10^7 \text{ m s}^{-1} \qquad \square$$

**Question 6.4** A metal sphere of radius 20 cm carries a charge of $5.0 \times 10^{-7}$ C. A proton is sent towards the sphere at a speed of $1.8 \times 10^6$ m s$^{-1}$. Can the proton reach the surface of the sphere?

### 6.3.2 electric potential

it is also useful to define the *electric potential* for any specific point in a field

electric potential can be thought as the electric potential energy per unit charge: $V = \dfrac{E_p}{q}$

> **electric potential** is the work needed to bring a unit positive charge from infinity

- unit: $[V] = \text{J C}^{-1} = \text{V}$
- electric potential due to an isolated source $Q$: $V = \dfrac{E_p}{q} = \dfrac{\frac{Qq}{4\pi\epsilon_0 r}}{q} \quad \Rightarrow \quad \boxed{V = \dfrac{Q}{4\pi\epsilon_0 r}}$
- potential at infinity vanishes: $V_\infty = 0$
- electric potential can take both signs

  the sign depends on whether unit positive charge is repelled or attracted by the source

  for positively-charged sources $V > 0$, while for negatively-charged sources $V < 0$
- electric potential is a *scalar* quantity

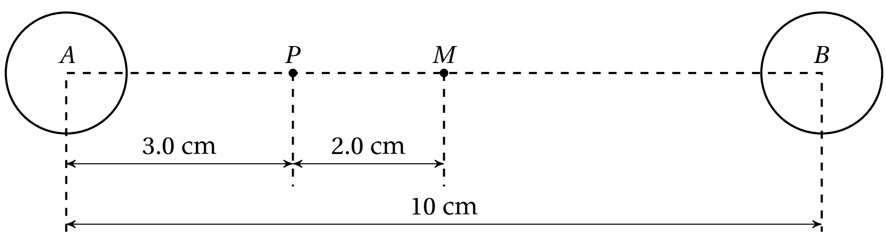
  to find combined potential due to multiple charges, add up contributions of each charge

**Example 6.8** An electron is accelerated from rest through a potential difference of 600 V. Find the final speed of the electron.

✍ gain in K.E. = change in electric P.E. $\quad \Rightarrow \quad \dfrac{1}{2}mv^2 = q\Delta V$

$$v = \sqrt{\frac{2q\Delta V}{m}} = \sqrt{\frac{2.60 \times 10^{-19} \times 600}{9.11 \times 10^{-31}}} \approx 1.45 \times 10^7 \text{ m s}^{-1}$$ □

**Example 6.9** Two small metal spheres $A$ and $B$ are in a vacuum. Sphere $A$ has charge +20 pC and sphere $B$ has charge +84 pC. The arrangement is shown below.




(a) Find the electric potential at point $P$ and point $M$ respectively.

(b) Find the work done to move an $\alpha$-particle from $P$ to $M$.

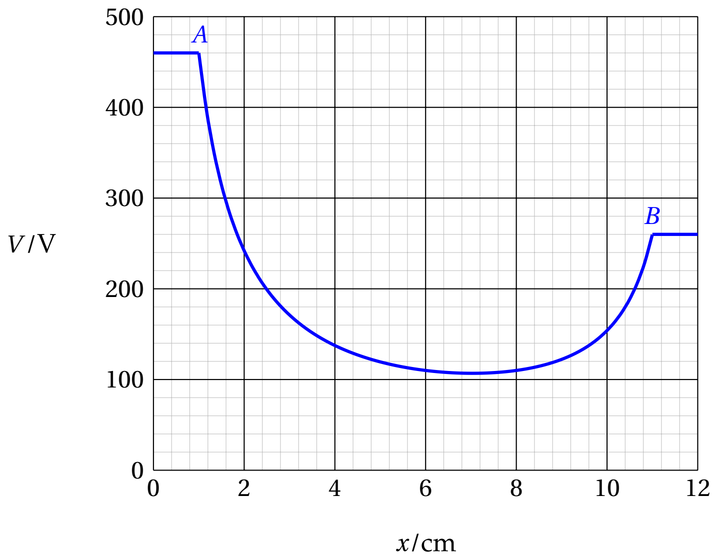
✍ combined electric potential: $V = V_A + V_B = \dfrac{1}{4\pi\epsilon_0}\left(\dfrac{Q_A}{r_A} + \dfrac{Q_B}{r_B}\right)$

at $P$: $V_P = 8.99 \times 10^9 \times \left(\dfrac{+20 \times 10^{-12}}{0.030} + \dfrac{+84 \times 10^{-12}}{0.070}\right) \approx 16.8$ V

at $M$: $V_M = 8.99 \times 10^9 \times \left(\dfrac{+20 \times 10^{-12}}{0.050} + \dfrac{+84 \times 10^{-12}}{0.050}\right) \approx 18.7$ V

from $P$ to $M$: $W = \Delta E_p = q\Delta V = q(V_M - V_P) = 2 \times 1.60 \times 10^{-19} \times (18.7 - 16.8) \approx 6.2 \times 10^{-19}$ J □

**Example 6.10** $A$ and $B$ are two positively-charged spheres of radius 1.0 cm. A proton $P$ initially at rest on the surface of $A$ moves along the line joining the centres of the two spheres. The variation with distance $x$ from the centre of $A$ of electric potential $V$ at point $P$ is given.




(a) Find the maximum speed as the proton moves from $A$ to $B$.

(b) Find the speed when the proton reaches surface of $B$.

✍ increase in K.E. = loss in P.E., so: $\dfrac{1}{2}mv^2 - 0 = q\Delta V \quad \Rightarrow \quad \dfrac{1}{2}mv^2 = q(V_A - V_P)$

maximum speed when $\Delta V$ is maximum, or $V_P = 107$ V becomes minimum (at $x = 7.2$ cm)

$$\frac{1}{2} \times 1.67 \times 10^{-27} \times v_{\max}^2 = 1.60 \times 10^{-19} \times (460 - 107) \quad \Rightarrow \quad v_{\max} \approx 2.60 \times 10^5 \text{ m s}^{-1}$$

at surface of $B$, $V_P = 260$ V (at $x = 11.0$ cm)

$$\frac{1}{2} \times 1.67 \times 10^{-27} \times v_B^2 = 1.60 \times 10^{-19} \times (460 - 260) \quad \Rightarrow \quad v_B \approx 1.96 \times 10^5 \text{ m s}^{-1}$$ □

**Question 6.5** Electrical breakdown occurs when electric field strength at surface of a metal sphere exceeds $5.0 \times 10^6$ N C$^{-1}$. Given that the radius of the sphere is 16 cm. What is the electric potential at the surface when electrical breakdown occurs?

**Question 6.6** Two charged particles $A$ and $B$ are separated by 20 cm. $P$ is a point on the line $AB$. Given that particle $A$ carries charge +7.2 $\mu$C, and electric potential is zero where $AP = 5.0$ cm. Find the electric charge of $B$.

**Question 6.7** A particle with specific charge (ratio of its electric charge to its mass) $+9.58 \times 10^7$ C kg$^{-1}$ is moving towards a fixed metal sphere. The sphere has a potential of +500 V. The initial speed of the particle is $3.0 \times 10^5$ m s$^{-1}$ when it is a large distance from the sphere. Determine whether the particle can reach the surface of the sphere.

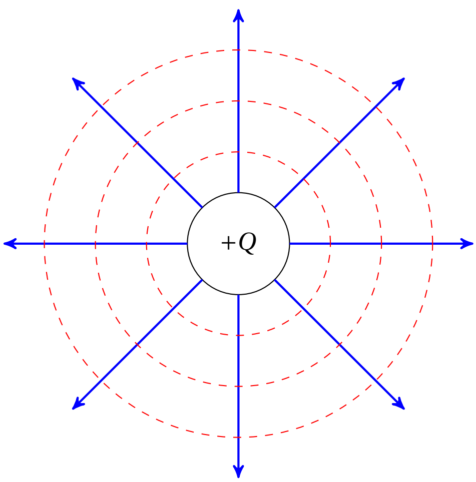
### 6.3.3 equipotential lines

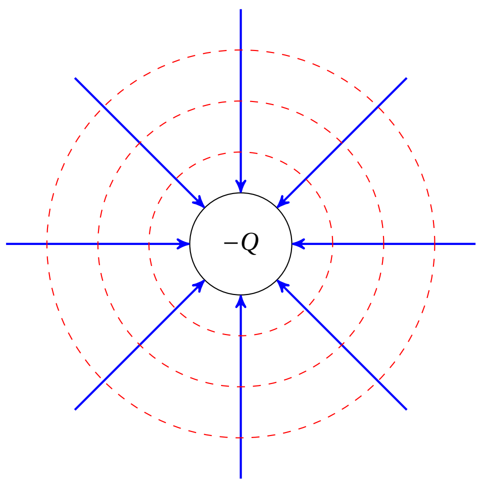
to show potential distributions, we draw **equipotential lines** [35]

points on same equipotential line have constant electric potential

i.e., equipotential lines are *contour lines* of equal electric potential

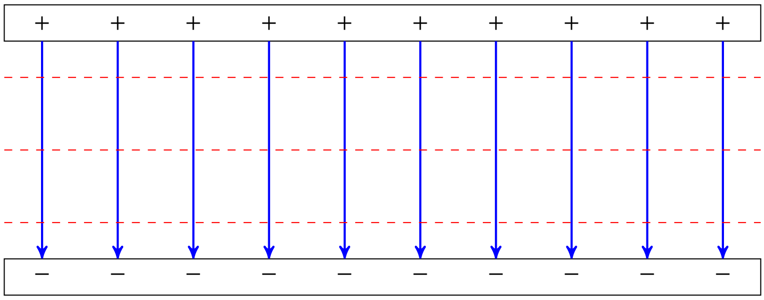
- for a field near a point charge, equipotential lines are a set of *concentric* circles




[35] In three dimensions, these lines form equipotential *surfaces*.

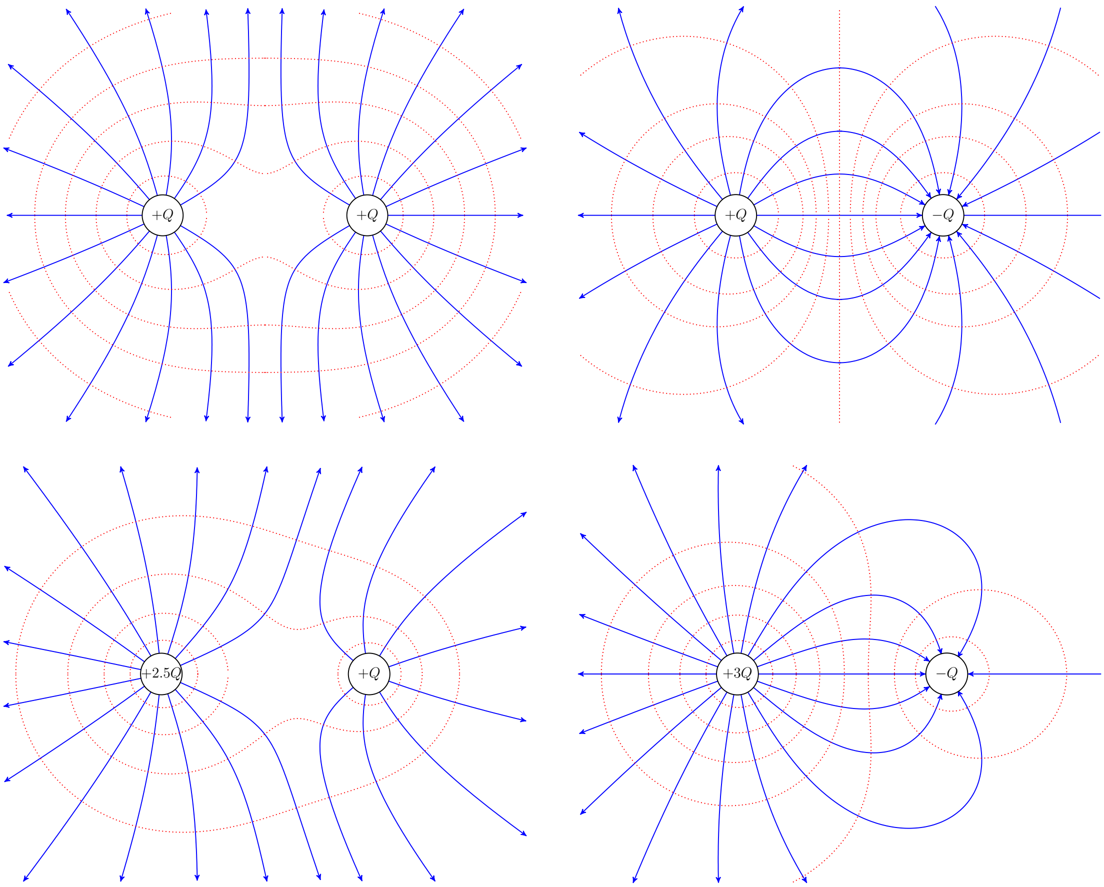
➢ for uniform fields, equipotential lines are a set of *parallel* straight lines

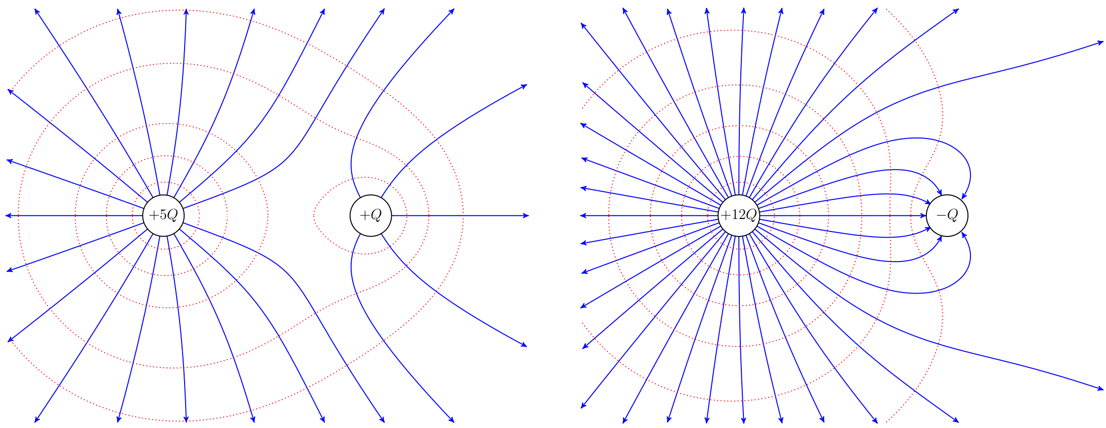



➢ equipotential lines are always perpendicular to the electric field lines

moving along an equipotential line requires no work done

**Example 6.11** Field lines and equipotential lines due to two charges of various magnitudes






## 6.4 further discussions on electric fields

### 6.4.1 comparison with gravitational fields

both gravitational and electric force are described by inverse square law, so it follows that the mathematical language for both theories are very similar

➢ physical quantities that describe gravitational/electric fields

| | vector description | scalar description |
|---|---|---|
| interaction between two masses/charges | force $F$ | potential energy $E_p$ |
| effect of source mass/charge | field $g/E$ | potential $\varphi/V$ |

➢ comparing gravitational field with electric field [36]

| | gravitational field | electric field | meaning |
|---|---|---|---|
| force | $F = (-)G\dfrac{Mm}{r^2}$ | $F = \dfrac{1}{4\pi\epsilon_0}\dfrac{Qq}{r^2}$ | force between masses/charges |
| field strength | $g = \dfrac{F}{m} = (-)G\dfrac{M}{r^2}$ | $E = \dfrac{F}{q} = \dfrac{1}{4\pi\epsilon_0}\dfrac{Q}{r^2}$ | force per unit mass/charge |
| potential energy | $E_p = -G\dfrac{Mm}{r}$ | $E_p = \dfrac{1}{4\pi\epsilon_0}\dfrac{Qq}{r}$ | related to work done by force |
| potential | $\varphi = \dfrac{E_p}{m} = -G\dfrac{M}{r}$ | $V = \dfrac{E_p}{q} = \dfrac{1}{4\pi\epsilon_0}\dfrac{Q}{r}$ | energy per unit mass/charge |

[36]Note that there is no negative mass, gravitational force always interacts attractively. This is the fundamental difference between gravitational fields and electric fields.

➢ similarities between gravitational field and electric field

- force and field strength both obey *inverse square laws*
- potential energy and potential is inversely proportional to separation
- no potential energy and no potential at infinite separation

➢ differences between gravitational field and electric field

mass (source of gravity) is always positive, but electric charges can be positive or negative

this fact leads to many fundamental differences between the two force fields

- electric force can be repulsive or attractive, but gravitational force is always attractive
- electric potential can take both signs, but gravitational potential is always negative

### 6.4.2 electric field inside conductors

consider electric field *inside* a metal conductor carrying charge $Q$

conductor means there are free charge carriers that can move around

but charge distribution should be stable for a charged conductor (no circulating currents)

so charge carriers must experience no force, i.e., field strength inside conductor is zero

put it the other way round, if there is an excess field, it will push free charge carriers to move around, until they are distributed so that the field inside the conductor becomes zero

moreover, there shall be no potential difference between any two points inside the conductor, otherwise charge carriers would flow, so electric potential must be constant

electric field strength is everywhere zero inside a conductor: $E = 0$

electric potential is everywhere constant inside a conductor: $V = \text{const}$

**Example 6.12** Consider the electric field due to a metal sphere of radius $R$ carrying charge $Q$. Plot the variation with the distance $r$ from sphere's centre of the field strength, and the variation with $r$ of the electric potential.

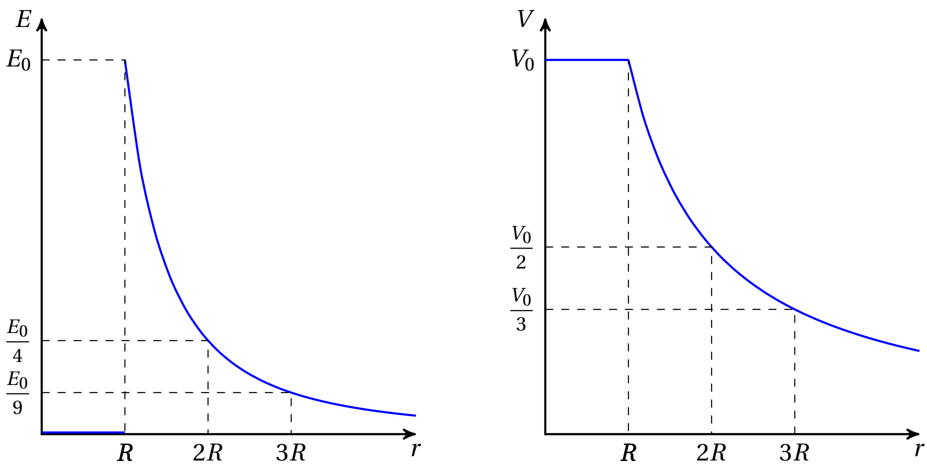
✍ charge $Q$ is uniformly spread out on *surface* of sphere

viewed from *outside*, the sphere appears to have all of its charge concentrated at the centre

so it can be modelled as a point charge due to its symmetric distribution of charges electric field strength at distance $r$ from sphere's centre is: $E = \dfrac{Q}{4\pi\epsilon_0 r^2}$ for $r > R$

electric potential at distance $r$ from sphere's centre is: $V = \dfrac{Q}{4\pi\epsilon_0 r}$ for $r > R$

*inside* the sphere, i.e., for $r < R$, we have $E = 0$, and $V = \dfrac{Q}{4\pi\epsilon_0 R} = \text{const}$ □




Example 6.12: field strength and potential due to a charged metal sphere

**Question 6.8** State whether the two spheres in Example 6.10 are conductors. State what feature of the potential graph supports your answer.

**Question 6.9** You might have the experience that your mobile phone signal gets much weaker when you get into an elevator. Explain why this happens.

### 6.4.3 field strength & potential

for a small displacement $\Delta r$ in an electric field, change in potential $\Delta V$ is

$$\Delta V = \frac{\Delta E_p}{q} \xlongequal{\Delta E_p = -W} -\frac{\Delta W}{q} = -\frac{F\Delta r}{q} \xlongequal{F=Eq} -E\Delta r \quad \Rightarrow \quad E = -\frac{\Delta V}{\Delta r}$$

as change in displacement $\Delta r \to 0$, we have $\boxed{E = -\dfrac{\mathrm{d}V}{\mathrm{d}r}}$

therefore we have the following theorem:

field strength is negative gradient of potential with respect to displacement can also consider change in potential $\Delta V$ for large distance due to work done in a field

$$\Delta V = \int \mathrm{d}V \xlongequal{E=-\mathrm{d}V/\mathrm{d}r} \int (-)E\mathrm{d}r = -\int E\mathrm{d}r$$ [37]

this gives the inverse relation: $\boxed{\Delta V = -\int E\mathrm{d}r}$ [38]

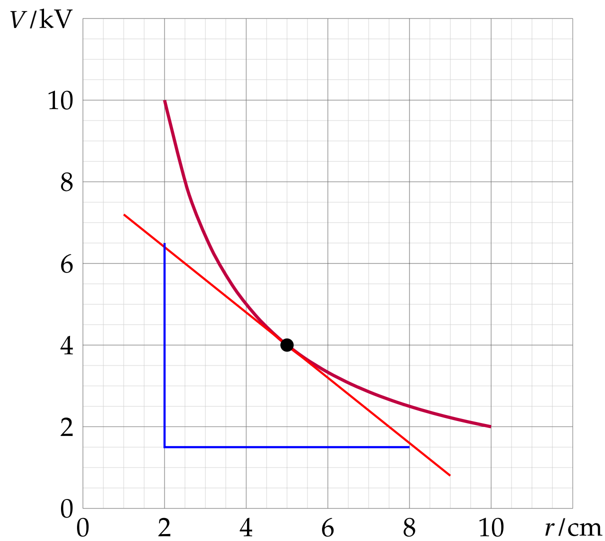
➢ given a $V$-$r$ graph, gradient of curve gives field strength

conversely, given a $E$-$r$ graph, area under curve gives change in potential

➢ can also write $F = -\dfrac{\mathrm{d}U}{\mathrm{d}r}$ and $\Delta U = -\int F\mathrm{d}r$

force always acts in a direction to lower the potential energy of an object [39]

**Example 6.13** The variation of electric potential near a charged object is shown on the graph. Calculate the electric field strength at 5.0 cm from the centre of the object.




✍ draw tangent to the graph at $r = 5.0$ cm (red line), gradient of tangent gives field strength:

$$\text{gradient} = \frac{\Delta V}{\Delta r} = \frac{(1.5 - 6.5) \times 10^3}{(8.0 - 2.0) \times 10^{-2}} \approx -8.3 \times 10^4 \text{ V m}^{-1} \quad \Rightarrow \quad E = -\frac{\Delta V}{\Delta r} = 8.3 \times 10^4 \text{ V m}^{-1}$$ □

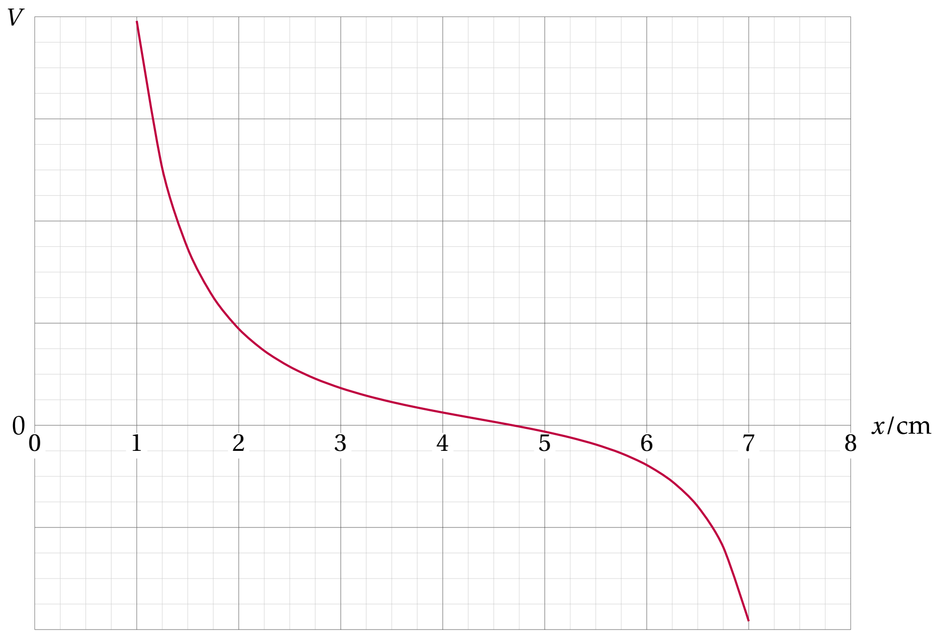
[37] The expression is implicitly integrated from initial position to final position.

[38] These equations are correct if the charge is moving in the parallel direction to the field, i.e., the motion is along the field lines. But an object can move in all directions in the field. More rigorously, if we take the vector nature of electric field into account, we should write $\Delta V = -\int \mathbf{E}\cdot\mathrm{d}\mathbf{r}$, and $\mathbf{E} = -\dfrac{\partial V}{\partial \mathbf{r}}$. (∗)

[39] This result can be generalised to a very important principle of physical laws called the *least action principle*. It states that any motion of a system tends to minimise the action, a physical quantity related to the energy of the system. This fundamental law plays a crucial role in the study of theoretical physics. (∗)

**Question 6.10** Show that the charged object in Example 6.13 behaves like a point charge. Determine the charge it carries, and hence calculate the field strength at $r = 5.0$ cm.

**Example 6.14** The variation of electrical potential along a certain line is shown. State and explain where in the field an electron will experience the greatest force.




✍ greatest force means greatest field strength, which means maximum potential gradient

largest gradient of $V$-$x$ curve at $x = 1$ cm, so greatest force at $x = 1$ cm □

**Example 6.15** electric field due to an isolated point charge

we have learned that the electric potential due to a point charge is: $V = \dfrac{Q}{4\pi\epsilon_0 r}$

using $E = -\dfrac{\mathrm{d}V}{\mathrm{d}r}$, we have $E = -\dfrac{\mathrm{d}}{\mathrm{d}r}\left(\dfrac{Q}{4\pi\epsilon_0 r}\right) = -\dfrac{Q}{4\pi\epsilon_0}\dfrac{\mathrm{d}}{\mathrm{d}r}\left(\dfrac{1}{r}\right) = \dfrac{Q}{4\pi\epsilon_0 r^2}$

this agrees with the expression for field strength due to an isolated charge □

**Question 6.11** For the electric field due to a charged metal sphere (see Example 6.12), convince yourself that the field strength equals negative gradient of potential at any point.

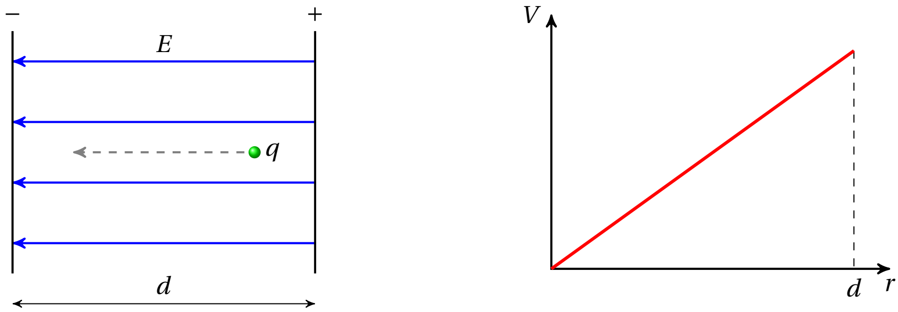
**Question 6.12** We have seen the statement field strength equals negative potential gradient holds for electric fields. Does it also hold for gravitational fields?

**uniform fields revisited**

given two oppositely-charged metal plates separated by a distance of $d$

if p.d. between the plates is $V$, then electric field strength between is given by $E = \dfrac{V}{d}$[40]

we will derive this result using the theorem introduced in the last section




moving a test charge in a uniform electric field

moving a test charge $q$ in a uniform field, work done by electric force: $W = Fd = Eqd$

change in P.E.: $\Delta E_p = -W = -Eqd$

change in potential: $\Delta V = \dfrac{\Delta E_p}{q} = -Ed$, or $E = -\dfrac{\Delta V}{d}$

plotting $V$-$r$ graph, gradient of line $= -E$

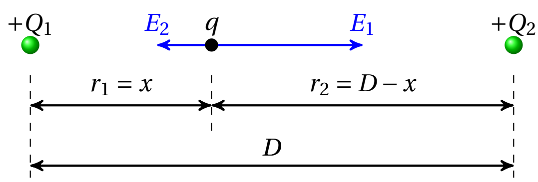
the minus sign means field strength points in the direction such that potential decreases

i.e., electric field acts from high potential to low potential

**electric field due to two positive point charges**

two point charges $+Q_1$, $+Q_2$ are separated by a distance of $D$

let's look into the electric field along the segment joining the two charges




[40] You should have learned this in AS-level physics.

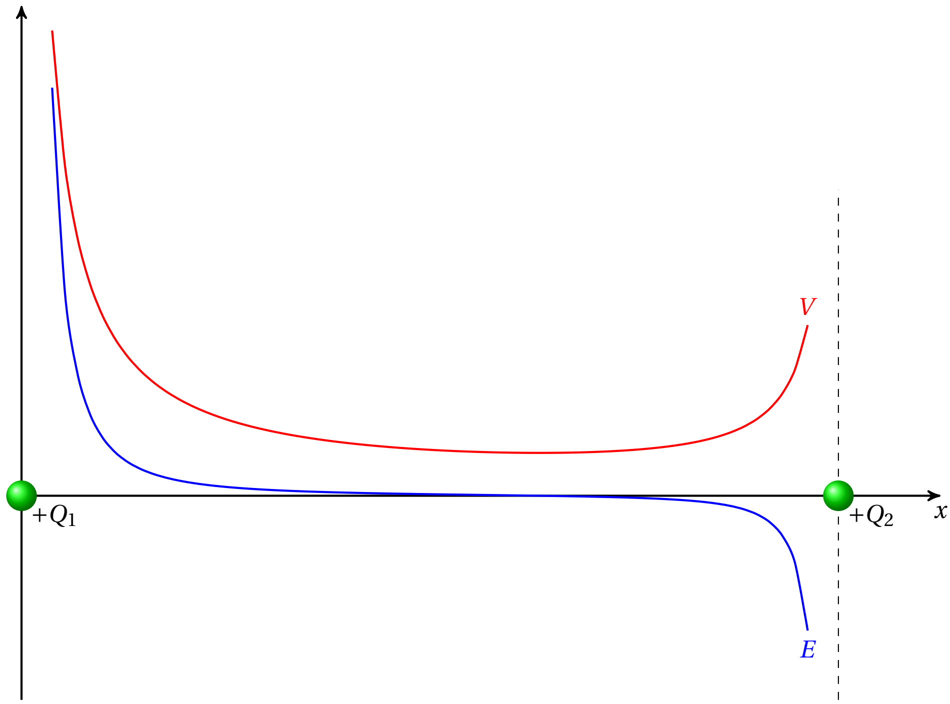
combined potential: $V = V_1 + V_2 = \frac{Q_1}{4\pi\epsilon_0 r_1} + \frac{Q_2}{4\pi\epsilon_0 r_2} = \frac{1}{4\pi\epsilon_0}\left(\frac{Q_1}{x} + \frac{Q_2}{D-x}\right)$

combined field strength: $E = E_1 - E_2 = \frac{Q_1}{4\pi\epsilon_0 r_1^2} - \frac{Q_2}{4\pi\epsilon_0 r_2^2} = \frac{1}{4\pi\epsilon_0}\left(\frac{Q_1}{x^2} - \frac{Q_2}{(D-x)^2}\right)$

notice that when computing $V$, we carry out *scalar sum*

but for $E$, we carry out *vector sum*, i.e., directions of $E_1$ and $E_2$ become important

$V$-$x$ graph and $E$-$x$ graph for the case where $Q_1 = 3Q_2$ are sketched




**Question 6.13** Convince yourself that field strength is indeed given by negative potential gradient. You may interpret it either graphically (think about gradient of tangent along the curve) or algebraically (think about the derivative of $V$).

**Question 6.14** We have looked into the electric field between two positively-charged particles. Discuss the cases where (a) both particles are negatively charged, (b) the two particles carry opposite charges.

# CHAPTER 7
# Capacitors

## 7.1 capacitors: an introduction

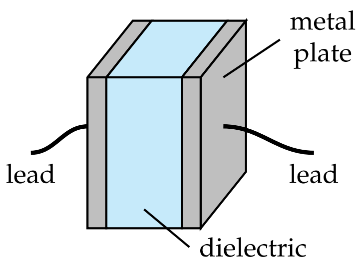
### 7.1.1 capacitors

**capacitors** are elementary electrical units widely used in electrical and electronic engineering

a typical capacitor has two conductive plates

between the plates there is usually an insulating material called *dielectric*

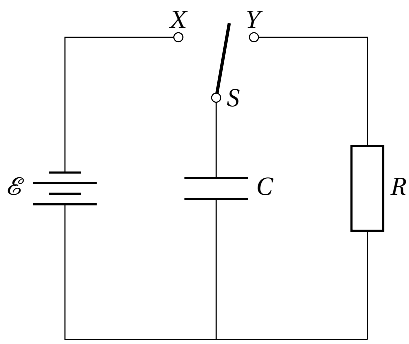
circuit symbol for a capacitor is ⊣⊢




if we construct an electric circuit as shown

when contact $S$ is moved to $X$, capacitor is connected to a voltage supply and becomes charged

positive and negative charges are separated onto two plates, and they will stay where they are even if we disconnect the capacitor from the voltage supply




if we then move $S$ to $Y$, the charged capacitor discharges and drives a current through resistor $R$, i.e., it can act as a temporary power source[41]

so capacitors can be used to store and release energy[42]

[41]Details on charging and discharging processes will be gone through in §7.4.

[42]Other important functions of capacitors in electronic circuits include smoothing output voltage of power supplies, blocking direct current while allowing alternating current to pass, etc.

### 7.1.2 mutual capacitance

to describe ability of a capacitor to store charges, we define the notion of capacitance

> **(mutual) capacitance** of a parallel-plate capacitor is defined as the ratio of the charge stored on one plate to the potential difference across the two plates.

in a word equation: mutual capacitance $C = \dfrac{\text{charge on one plate } Q}{\text{p.d. } V \text{ across the plates}} \quad \Rightarrow \quad \boxed{C = \dfrac{Q}{V}}$

- unit of capacitance: **farad** [43] : $[C] = \mathrm{F}$

  farad a derived unit: $1\ \mathrm{F} = 1\ \mathrm{C \cdot V^{-1}}$

  farad is a large unit, more common subunits of capacitance in use are sub-multiples of farad:

$$1\ \mu\mathrm{F} = 10^{-6}\ \mathrm{F}, \quad 1\ \mathrm{nF} = 10^{-9}\ \mathrm{F}, \quad 1\ \mathrm{pF} = 10^{-12}\ \mathrm{F}$$

- for a parallel-plate capacitor, charges on the two plates are equal but opposite

  net charge on the capacitor: $Q_{\text{net}} = (+Q) + (-Q) = 0$

  so we should emphasise on the notion of charge on *one* plate in the definition

- capacitance depends on *geometry* of the device and permittivity of the dielectric material

  capacitance does not depend on electric field or potential[44]

  for example, capacitance between two metal plates is: $C = \dfrac{\epsilon_0 A}{d}$ [45]

  $A$ is area of plate, $d$ is distance between plates, both are geometrical quantities

### 7.1.3 self-capacitance

there are two closely related notions of capacitance: *mutual* capacitance and *self* capacitance

[43]The unit is named after Michael Faraday, a British physicist who developed the concept of capacitance. Faraday's other main discoveries include electromagnetic induction and electrolysis. He established the basis for the concept of the electromagnetic field in physics.

[44]Recall the resistance of an electrical component. Resistance is defined as the ratio of p.d. to current, but the value of resistance is essentially dependent on the length, cross-sectional area and material of the component, instead of the p.d. applied or the current flowing through it.

[45]If there is *dielectric* in between, the formula should be rewritten as $C = \dfrac{\epsilon A}{d}$, where $\epsilon$ is permittivity of dielectric. These formulae are not examinable by the syllabus.

the definition for capacitance given in the previous section, is actually *mutual* capacitance

on the other hand, all bodies are able to store electrical charge

any object that can be electrically charged exhibits capacitance

we define **self-capacitance** of an object as the amount of charge that must be added to increase per unit electrical potential

in a word equation, self capacitance $C = \dfrac{\text{charge of object } Q}{\text{electric potential of object } V} \Rightarrow \boxed{C = \dfrac{Q}{V}}$ [46]

**Example 7.1** Self-capacitance of a charged metal sphere in a vacuum

consider a metal sphere of radius $R$ and carries an electric charge of $Q$

its electric potential: $V = \dfrac{Q}{4\pi\epsilon_0 R}$

self-capacitance of the sphere: $C_{\text{sphere}} = \dfrac{Q}{V} = Q \times \dfrac{4\pi\epsilon_0 R}{Q} \Rightarrow \boxed{C_{\text{sphere}} = 4\pi\epsilon_0 R}$

note that capacitance is only dependent on its geometrical property (radius $R$) □

**Example 7.2** A conducting sphere of radius 1.0 m is situated in free space. (a) Find its capacitance. (b) In order to raise its potential to 5000 V, find the amount of charge needed.

✍ capacitance of the sphere: $C = 4\pi\epsilon_0 R = 4\pi \times 8.85 \times 10^{-12} \times 1.0 \approx 1.11 \times 10^{-10}$ F

(here you can see farad being an impractically huge unit)

charge on sphere: $Q = CV = 1.11 \times 10^{-10} \times 5000 \approx 5.56 \times 10^{-7}$ C □

### analogy with ideal gases

an interesting analogy can be made between capacitors and ideal gases

recall an ideal gas is described by equation $pV = nRT$ [47]

$$\text{compare} \begin{cases} \text{amount of charge:} & Q = CV \\ \text{amount of substance:} & n = \left(\dfrac{V}{RT}\right) p \end{cases}$$

volume $V$ of a container has a certain space *capacity*, at fixed $T$, pumping more gas (increase $n$) into system, pressure $p$ increases

[46] For either mutual capacitance or self capacitance, defining equation $C = \dfrac{Q}{V}$ takes the same form, but you should keep in mind that $Q$ and $V$ represent different things in different contexts.

[47] Don't confuse voltage $V$ with volume $V$!

similarly, a capacitor has a certain charge capacity, adding more charge $Q$ increases p.d. $V$, so the quantity $C$ is naturally called *capacitance*

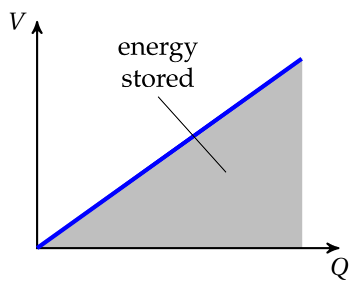
also for a container, there exists a maximum pressure which it can withstand

for a capacitor, there exists a *breakdown voltage*, or *withstand voltage*, beyond which there could be sparkling across the capacitor

## 7.2 energy stored in a capacitor

to charge a capacitor, need to push electrons off one plate and onto the other

separation of positive and negative charges requires work done ⇒ energy is stored




since charge $Q$ varies with p.d. $V$, we shall use the $V$-$Q$ graph to find work done $W$

area under $V$-$Q$ graph is equal to work done $W$

$$W = \frac{1}{2}QV$$

substitute $Q = CV$, energy stored in capacitor is[48]:

$$\boxed{W = \frac{1}{2}CV^2 = \frac{Q^2}{2C}}$$

**Example 7.3** When the p.d. across a capacitor of $1.8 \times 10^{-4}$ F is increased from 10 V to 20 V, how much additional energy is stored?

✍ energy change: $\Delta W = W_f - W_i = \frac{1}{2}CV_f^2 - \frac{1}{2}CV_i^2 = \frac{1}{2} \times 1.8 \times 10^{-4} \times (20^2 - 10^2) = 2.7 \times 10^{-2}$ J □

**Question 7.1** A capacitor of 2500 $\mu$F is charged to a working voltage of 18 V. (a) What is the magnitude of positive charge on its plate? (b) What is the energy stored?

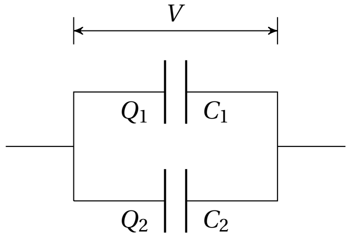
**Question 7.2** A capacitor initially charged to a potential difference of 16 V discharges and loses 40% of its energy. What is its new p.d.?

**Question 7.3** For an isolated metal sphere of radius 30 cm situated in vacuum, what is the electric potential energy stored when charged to a potential of 120 kV?

[48] Rigorously speaking, this electrical potential energy is stored within the *electric fields* between metal plates of capacitor.

## 7.3 capacitor networks

### 7.3.1 capacitors in parallel




consider two capacitors connected in parallel

same p.d. $V$ across the network: $V = V_1 = V_2$

but charge $Q$ is shared: $Q_{\text{total}} = Q_1 + Q_2$

$$\frac{Q_{\text{total}}}{V} = \frac{Q_1}{V} + \frac{Q_2}{V}$$

$$C_{\text{total}} = C_1 + C_2$$

if three of more capacitors in parallel: $\boxed{C_{\text{total}} = C_1 + C_2 + C_3 + \cdots}$ $\boxed{Q_{\text{total}} = Q_1 + Q_2 + Q_3 + \cdots}$

➢ adding extra capacitor in parallel to a network, total capacitance will increase

explanation: when several capacitors connected in parallel, equivalent to a single capacitor with larger plates, so more charge on the plates $\Rightarrow C \nearrow$

**Example 7.4** A capacitor with capacitance $C_0$ is charged to a p.d. $V_0$. It is disconnected from the power supply, and then connected across an identical capacitor. Discuss the change in p.d., and change in energy stored in the system.

✍ initial charge $Q = C_0 V_0$, initial energy stored $W_0 = \frac{1}{2} C V_0^2$

combined capacitance: $C = C_0 + C_0 = 2C_0$

charge is conserved, so final p.d. across: $V = \frac{Q}{C} = \frac{C_0 V_0}{2C_0} = \frac{1}{2} V_0$

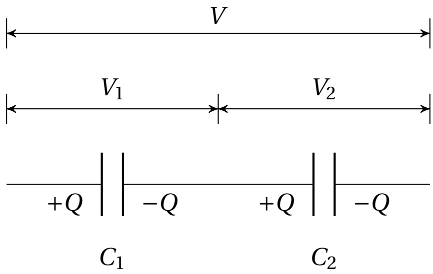
charge shared between capacitors, the first capacitor loses half its charge and p.d.

final energy stored in system: $W = \frac{1}{2} C V^2 = \frac{1}{2} \times 2C_0 \times (\frac{1}{2} V_0)^2 = \frac{1}{4} C_0 V_0^2 \Rightarrow W = \frac{1}{2} W_0$

half of stored energy is lost as heat when electrons flow between two capacitors □

**Question 7.4** A capacitor $A$ of capacitance $C$ and a second capacitor $B$ of capacitance $3C$ are connected in parallel. If a voltage is applied across the network, what is the ratio of energy stored in $A$ to that in $B$? What about the ratio of electric charge?

### 7.3.2 capacitors in series




consider next two series capacitors

same charge $Q$ on each plate: $Q = Q_1 = Q_2$[49]

p.d. is shared: $V_{\text{total}} = V_1 + V_2$

$$\frac{V_{\text{total}}}{Q} = \frac{V_1}{Q} + \frac{V_2}{Q}$$

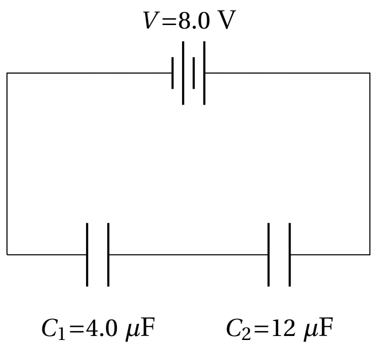
$$\frac{1}{C_{\text{total}}} = \frac{1}{C_1} + \frac{1}{C_2}$$

if three of more capacitors in series: $\boxed{\frac{1}{C_{\text{total}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots}$ $\boxed{V_{\text{total}} = V_1 + V_2 + V_3 + \cdots}$

➢ adding extra capacitor in series to a network, total capacitance will decrease

explanation: when several capacitors connected in series, equivalent to a parallel-plate capacitor with greater separation, so more charge on the plates $\Rightarrow C \searrow$

**Example 7.5** For the circuit shown below, find the p.d. across each of the capacitor.




✍ using result for combined capacitance, we have: $C_{\text{total}} = \left(\frac{1}{C_1} + \frac{1}{C_2}\right)^{-1} = \left(\frac{1}{4.0} + \frac{1}{12}\right)^{-1} = 3.0\ \mu\text{F}$

charge for the network: $Q = C_{\text{total}} V_{\text{total}} = 3.0 \times 8.0 = 24\ \mu\text{C}$

series network so all capacitors have same $Q$, so: $Q_1 = Q_2 = 24\ \mu\text{C}$

p.d. across each individual capacitor: $V_1 = \frac{Q_1}{C_1} = \frac{24}{4.0} = 6.0\ \text{V}, \quad V_2 = \frac{Q_2}{C_2} = \frac{24}{12} = 2.0\ \text{V}$

alternatively, we can use properties of series network: $Q_1 = Q_2$ and $V_{\text{total}} = V_1 + V_2$

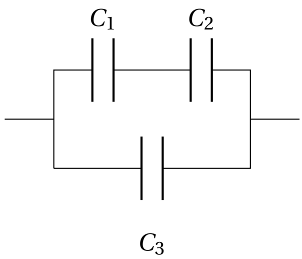
we can solve simultaneous equations: $\begin{cases} 4.0V_1 = 12V_2 \\ V_1 + V_2 = 8.0 \end{cases} \Rightarrow \begin{cases} V_1 = 6.0\ \text{V} \\ V_2 = 2.0\ \text{V} \end{cases}$ □

[49]Initially, the H-shaped isolated section between capacitors is uncharged. Since no charge can enter or leave this section, its net charge must remain zero, so same charge $Q$ on each plate of series capacitors.

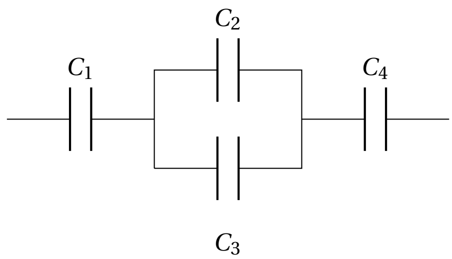
### 7.3.3 capacitor networks

more complicated capacitor networks can be considered as a combination of some smaller networks with capacitors in parallel or in series

**Example 7.6** $C_1 = C_2 = C_3 = C_4 = 10\ \mu\text{F}$, calculate the capacitance of the network (a) and (b).




(a)




(b)

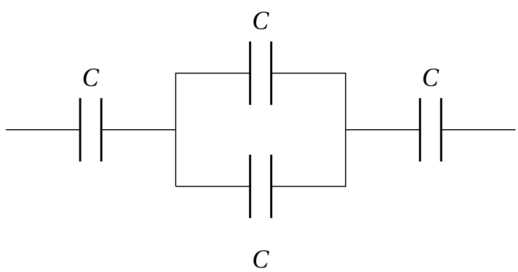
(a) $C_{12} = \left(\dfrac{1}{C_1} + \dfrac{1}{C_2}\right)^{-1} = \left(\dfrac{1}{10} + \dfrac{1}{10}\right)^{-1} = 5.0\ \mu\text{F}$

$C_{\text{total}} = C_{12} + C_3 = 5 + 10 = 15\ \mu\text{F}$

(b) $C_{23} = C_2 + C_3 = 10 + 10 = 20\ \mu\text{F}$

$C_{\text{total}} = \left(\dfrac{1}{C_1} + \dfrac{1}{C_{23}} + \dfrac{1}{C_4}\right)^{-1} = \left(\dfrac{1}{10} + \dfrac{1}{20} + \dfrac{1}{10}\right)^{-1} = 4.0\ \mu\text{F}$ □

**Example 7.7** Four identical capacitors are arranged as shown. Each capacitor can withstand a maximum p.d. of 12 V, what is the maximum safe p.d. to be applied between the terminals?




suppose charge on the leftmost capacitor is $Q$, then charge of rightmost capacitor is also $Q$

but for the two capacitors in parallel, charge is shared, so each has charge $\dfrac{Q}{2}$

p.d. across the parallel network is half of the p.d. across the two capacitors near the ends

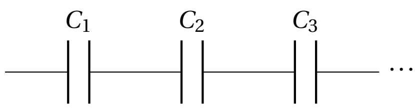
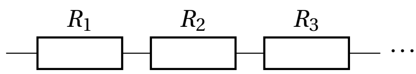
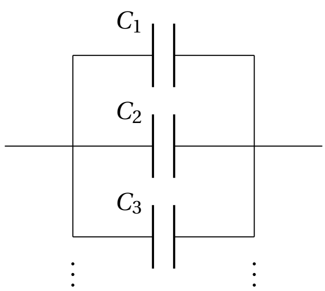
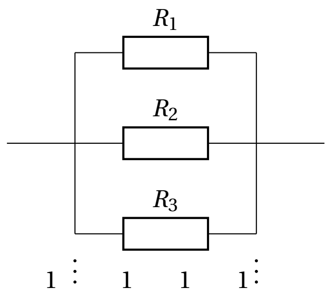
so maximum p.d. across the terminals: $V_{\text{max}} = 12 + 6 + 12 = 30\ \text{V}$ □

**Question 7.5** Using at most four capacitors of $24\ \mu\text{F}$, design a network that has a combined capacitance of (a) $72\ \mu\text{F}$, (b) $8\ \mu\text{F}$, (c) $36\ \mu\text{F}$, and (d) $18\ \mu\text{F}$.

**Question 7.6** If the two networks in Example 7.6 are both connected to a supply voltage of 15 V, determine the p.d. across each individual capacitor.

### 7.3.4 capacitors & resistors

as a quick review, we compare capacitor networks with resistor networks

| | capacitors | resistors |
|---|---|---|
| in series | same charge   $\frac{1}{C_{\text{total}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots$ | same current   $R_{\text{total}} = R_1 + R_2 + R_3 + \cdots$ |
| in parallel | same p.d. across   $C_{\text{total}} = C_1 + C_2 + C_3 + \cdots$ | same p.d. across   $\frac{1}{R_{\text{total}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots$ |

## 7.4 charging & discharging capacitors

in this section, we will investigate how the p.d. across a capacitor changes with time when it is being charged, and we will also look into discharging processes

### 7.4.1 charging phase

initial state: no charge in capacitor: $Q(0) = 0, V_C(0) = 0$

at any instant: $\mathrm{d}Q = C\mathrm{d}V_C$

current in circuit: $I = \dfrac{V_R}{R} = \dfrac{\mathscr{E} - V_C}{R}$

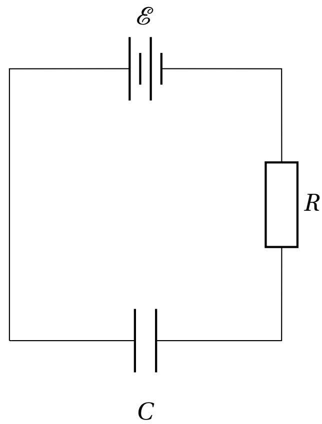
change of charge: $\mathrm{d}Q = I\mathrm{d}t = \dfrac{\mathscr{E} - V_C}{R}\mathrm{d}t = C\mathrm{d}V_C$

$$\frac{\mathrm{d}t}{RC} = \frac{\mathrm{d}V_C}{\mathscr{E} - V_C}$$

$$\int_0^t \frac{\mathrm{d}t}{RC} = \int_0^{V_C} \frac{\mathrm{d}V_C}{\mathscr{E} - V_C}$$

$$\frac{t}{RC}\bigg|_0^t = -\ln(\mathscr{E} - V_C)\bigg|_0^{V_C}$$

simplify everything, we get: $\boxed{V_C = \mathscr{E}\left(1 - \mathrm{e}^{-t/RC}\right)}$




- p.d. of capacitor increases at a decreasing rate when it is being charged

as electric charges are separated onto the two plates, pushing more $+Q/-Q$ onto +ve/−ve plate requires more work done to overcome the repulsion ⇒ increase in p.d. slows down

- p.d of capacitor eventually tends to the battery e.m.f.

charge will continue to flow if there exists a potential difference

when $V_C = \mathscr{E}$, no charge flow, hence charging current gradually drops to zero

### 7.4.2 discharging phase

capacitor initially charged with $Q(0) = Q, V_C(0) = V_0$

at any instant, $\mathrm{d}Q = -C\mathrm{d}V_C$ (minus sign because charge decreases during discharging) also $V_C = V_R$ because $C$ and $R$ are in parallel

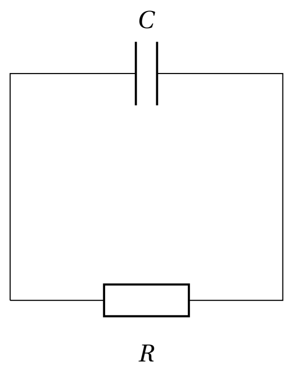
charge change: $\mathrm{d}Q = I\mathrm{d}t = \frac{V_C}{R}\mathrm{d}t = -C\mathrm{d}V_C$

$$-\frac{\mathrm{d}t}{RC} = \frac{\mathrm{d}V_C}{V_C}$$

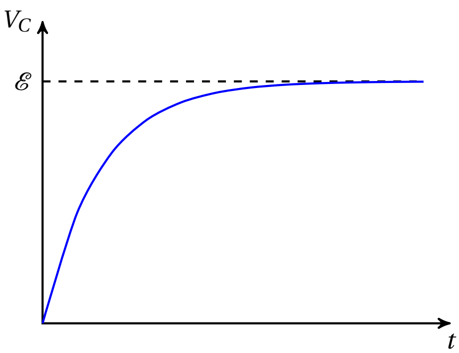
$$-\int_0^t \frac{\mathrm{d}t}{RC} = \int_{V_0}^{V_C} \frac{\mathrm{d}V_C}{V_C}$$

$$-\frac{t}{RC}\Big|_0^t = \ln(V_C)\Big|_{V_0}^{V_C}$$

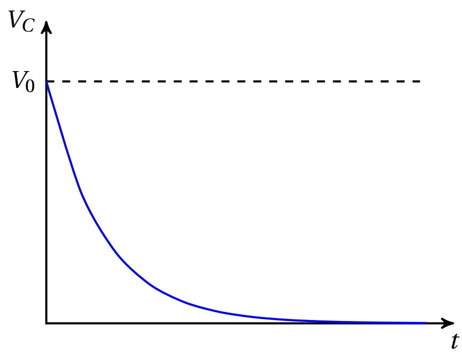
simplify everything, we get: $\boxed{V_C = V_0 \mathrm{e}^{-t/RC}}$




- p.d. of capacitor gradually drops to zero during discharging
- discharging current also gradually approaches zero




charging phase of capacitor




discharging phase of capacitor

### 7.4.3 time constant

$\tau \equiv RC$ is called **time constant**, which determines charging and discharging rate of a capacitor

$R \nearrow \Rightarrow$ smaller charging/discharging current $\Rightarrow$ takes longer to charge/discharge

$C \nearrow \Rightarrow$ more charge to be charged/discharged $\Rightarrow$ takes longer

charging or discharging of capacitors is not instantaneous, there is always a certain delay

as a rule of thumb, after a time $t = 3 \sim 5\tau$, charging or discharging is almost complete

# Chapter 8
# Magnetic Fields

## 8.1 magnetism

### 8.1.1 magnets

magnetic effects are commonly seen in *magnets*

a magnet creates a magnetic field that attracts or repels other magnets

➢ polarity of magnets

a magnet has two poles, the *north pole* and the *south pole*

freely suspend a magnet, the north pole points towards earth's geographic north pole

when two magnets are brought near each other, like poles repel, and opposite poles attract

➢ we use **magnetic field lines** to graphically represent how magnetic field permeate space

by convention, fields lines emerge from north pole and go into south pole

density of field lines shows strength of magnetic field

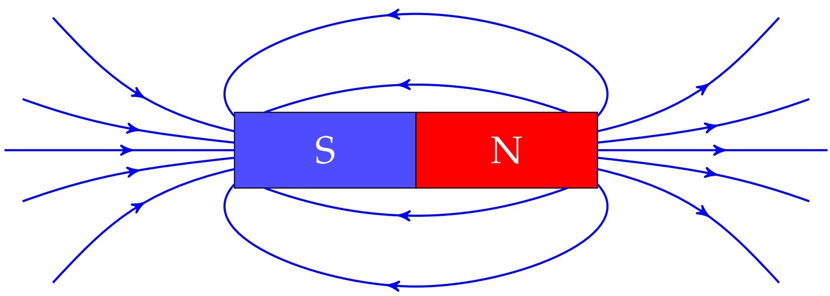
direction of field line tells how the north pole of a small compass will line up at that point

➢ strength of the field is described by a quantity called **magnetic flux density**, denoted by $B$

when we draw field lines, we are actually drawing the pattern of flux density $B$

the notion of flux density will be defined later in details in §8.2.2

**Example 8.1** magnetic field around a bar magnet



**Question 8.1** For two identical bar magnets placed side by side as shown, what does the magnetic field look like? Try sketching the magnetic field lines.


(a) two attracting bar magnets


(b) two repelling bar magnets

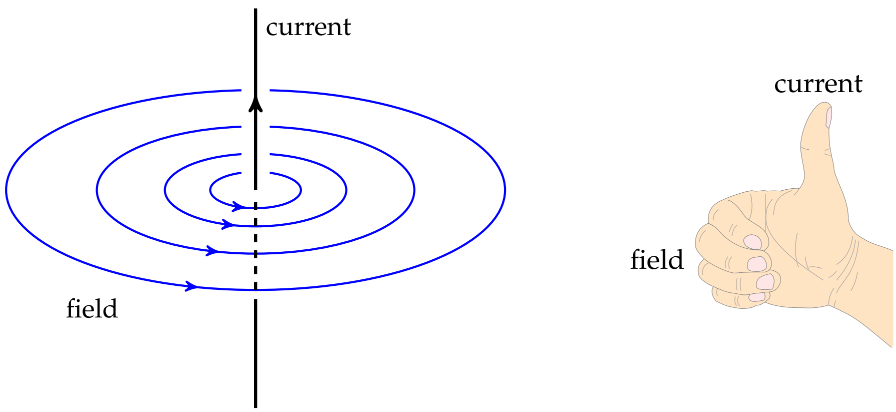
### 8.1.2 magnetic field due to currents

an electric current also induces a magnetic field around it

this was first discovered by Danish physicist *Hans Christian Orsted* in 1820, when he noticed the turning of a compass needle placed next to a wire carrying current.

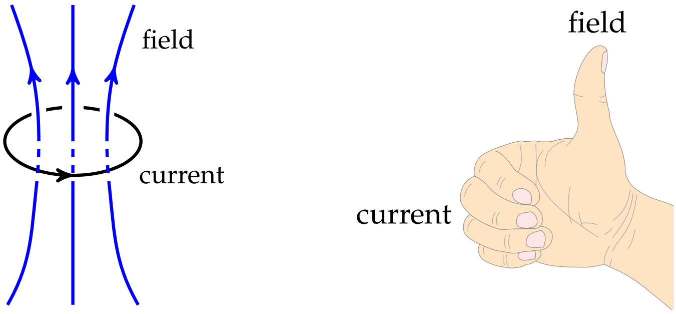
we will look into what happens when a current flows through a straight wire or a coil

➢ pattern of the field can be determined using **right-hand (grip) rule**[50]




field due to a long straight current-carrying wire

[50]The awesome right-hand rule illustrations below are copied from the PSTricks web site: http://tug.org/PSTricks/main.cgi?file=examples. The credits for these figures are attributed to CTAN community member *Thomas Söll*.




field due to a circulating current

➢ strength of the field is proportional to the current: $B \propto I$

for both straight wires [51] and coils[52] , greater current means stronger field

➢ strength of magnetic field can be increased with *soft iron*

this is because *ferromagnetic materials* (iron, cobalt, nickel) can attract magnetic field lines[53]

### 8.1.3 solenoids & electromagnets

strength and polarity of the field due to a coil can be changed easily by tuning currents, so coils are widely used to create magnetic fields where needed

a current-carrying coil is also called a **solenoid**

➢ a solenoid generates a magnetic field similar to that of a bar magnet

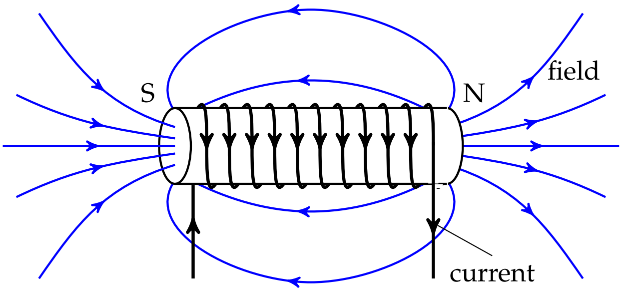
useful to talk about the north and south poles of a solenoid

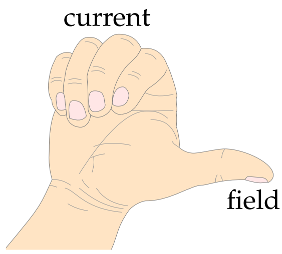
➢ direction of magnetic field in a solenoid is given by the **right-hand (grip) rule**

[51]The magnetic flux density generated by a current flowing in an infinitely long wire in free space is given by: $\boxed{B = \frac{\mu_0 I}{2\pi r}}$, $I$ is the electric current, $r$ is the perpendicular distance from the current, and $\mu_0 = 4\pi \times 10^{-7}$ T m A$^{-1}$ is a fundamental physical constant called the *vacuum permeability*. (∗)

[52]The magnetic flux density at the centre of a current-carrying coil in free space is given by: $\boxed{B = \frac{\mu_0 N I}{L}}$, where $N$ is the number of turns, $I$ is the electric current flowing through the coil, $L$ is the length of coil, and $\mu_0$ is the vacuum permeability. (∗)

[53]If the straight wire is immersed in a material with *relative permeability* $\mu_r$, then the field becomes: $B = \frac{\mu_0 \mu_r I}{2\pi r}$. Similarly, if a material with relative permeability $\mu_r$ is present, then the magnetic flux density inside a coil becomes: $B = \frac{\mu_0 \mu_r N I}{L}$. A good magnetic material (high permeability material), such as iron, has large $\mu_r$, and therefore can greatly intensify the magnetic field. (∗)




magnetic field around a solenoid

- magnetic field produced by a solenoid can be controlled

  current $I \nearrow \Rightarrow$ stronger field, also number of turns $N \nearrow \Rightarrow$ stronger field

- inserting an *iron core* inside greatly strengthens the field, this makes an *electromagnet*

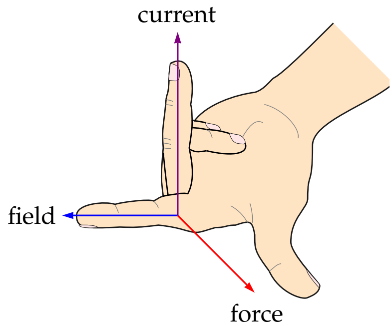
**Question 8.2** Describe the magnetic field due to an alternating current through a solenoid.

## 8.2 magnetic force on current-carrying conductor

### 8.2.1 magnetic force on current-carrying conductor

a current-carrying wire produces its own magnetic field, when it is surrounded by an external magnetic field, the two fields would interact $\Rightarrow$ a **magnetic force** on the conductor

- direction of magnetic force on current can be worked out with **Fleming's left-hand rule**[54]




Fleming's left-hand rule

- magnitude of magnetic force: $\boxed{F = BIL\sin\theta}$

  $B$ is the magnetic flux density to be specified later

  $I$ is the current, $\theta$ is the angle between $B$ and $I$

- when $B \perp I$, magnetic force $F = BIL$

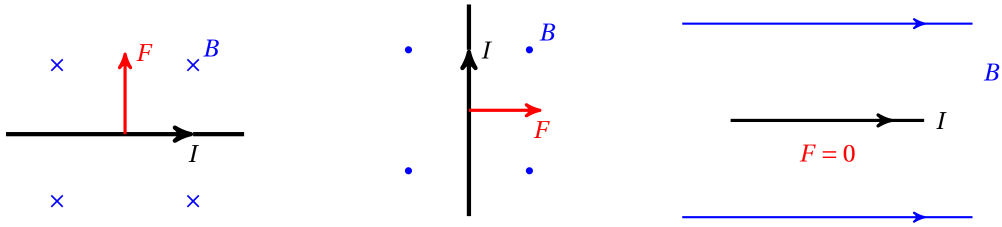
  when $B \parallel I$, there is no magnetic force

  when $B$ forms angle $\theta$ with $I$, only the *perpendicular component* contributes to the force, giving rise to the $\sin\theta$ factor in the formula

- magnetic force is perpendicular to both $B$ and $I$

[54] Figure credit again goes to *Thomas Söll* from the CTAN community.

**Example 8.2** Determine the direction of the magnetic force acting. Check yourself.




(a) (b) (c)

### 8.2.2 magnetic flux density

rewrite $F = BIL\sin\theta$ as $B = \dfrac{F}{IL\sin\theta}$,we can now give a formal definition for flux density $B$:

> **magnetic flux density** $B$ at a point is defined as the force acted per unit length on a conductor carrying a unit current at right angle to the magnetic field

- flux density describes the strength of a magnetic field[55]
- $B$ is measured in **tesla** (T)[56]: $[B] = \mathrm{T}$, $1\ \mathrm{T} = 1\ \mathrm{N \cdot A^{-1} \cdot m^{-1}}$

> if a wire of 1 m normal to the magnetic field that carries a current of 1 A experiences a force of 1 N, then the magnetic flux density is 1 T

- $B$ is a *vector* quantity, with both magnitude and direction

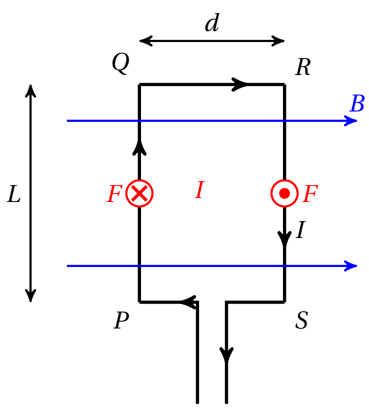
**Example 8.3** A wire of 0.80 m carrying a current of 5.0 A is lying at right angles to a magnetic field, it experiences a force of 0.60 N, what is the flux density?

✍ $B = \dfrac{F}{IL} = \dfrac{0.60}{5.0 \times 0.80} = 0.15\ \mathrm{T}$ (a field of over 0.1 T is actually quite strong!) □

[55]Magnetic field strength is a different quantity, defined by $H = \dfrac{B}{\mu}$, with $\mu = \mu_0\mu_r$ being the *magnetic permeability* of material. The naming of field strength $H$ and flux density $B$ are due to historical reasons.

[56]Telsa is a very large unit. For example, the strength of a typical refrigerator magnet is of about $10^{-3}$ T, even the very strong superconducting coils used in MRI are of around $1 \sim 3$ T.

**Example 8.4** Torque on a rectangular metal frame in uniform magnetic field




a rectangular frame lies in parallel with the field

$QR, PS \parallel B$, so no force acting on these two sides

$PQ, RS \perp B$, so there is magnetic force $F = BIL$

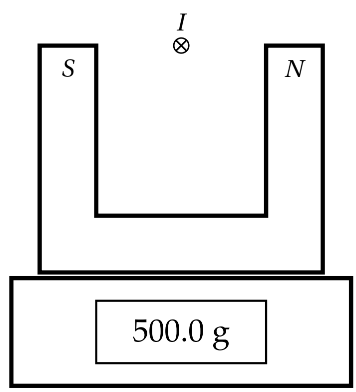
using left-hand rule, we find $F_{PQ}$ acts into the paper, $F_{RS}$ acts out of the paper

this is a pair of equal but opposite forces

they produce a torque about central axis of frame: $\tau = Fd = BILd$, causing the frame to rotate[57] □

**Question 8.3** If the plane of the rectangular frame is at right angle to the magnetic field, describe the magnetic forces acting on each side and hence state what happens to the frame.

**Example 8.5** A U-shaped magnet is placed on a balance with a wire suspended above it as shown. Magnetic flux density between the poles is about 0.50 T. The part of the wire that is in the field is of length 8.0 cm. The balance initially shows a reading of 500.0 g. When a current of 10 A flowing into the paper is switched on, what is the new reading on the balance?




✍ use left-hand rule, force on wire acts upwards

from *Newton's third law*, reaction force on magnet acts downwards

so there will be an increase in the balance reading

magnitude of the force: $F = BIl = 0.50 \times 10 \times 8.0 \times 10^{-2} = 0.40$ N

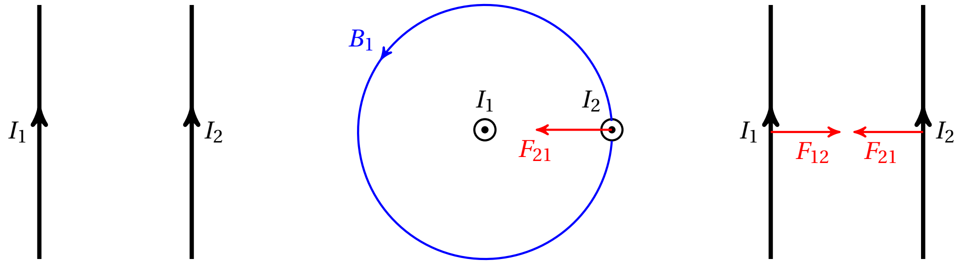
change in the mass reading: $\Delta m = \frac{F}{g} = \frac{0.40}{9.81} \approx 0.0407 \text{ kg} \approx 40.7 \text{ g}$

new reading on balance: $m_{\text{new}} = 500.0 + 40.7 = 540.7$ g □

**Question 8.4** If the current in Example 8.5 is replaced by a current of 6.0 A flowing out of the plane of the paper. What is the reading on the balance?

[57]This effect allows engineers to build *electric motors*, such as those in electric vehicles, washing machines, blenders, etc.

**force between long straight current-carrying wires**



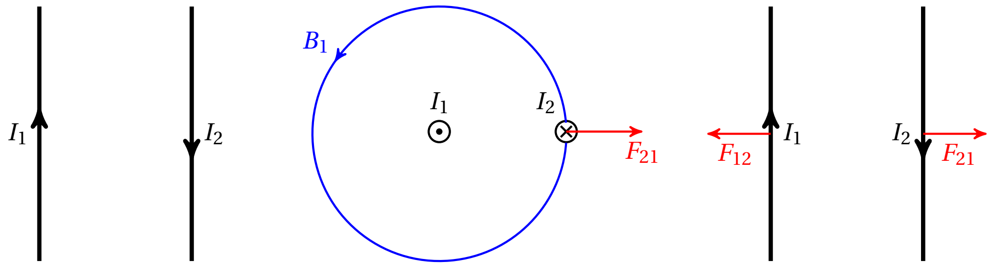

if we have two straight parallel wires carrying currents in the same direction

force acting on $I_2$ is due to magnetic field $B_1$ generated by $I_1$

from top view, $I_1$ creates a counter-clockwise $B_1$, so $I_2$ experiences an upward field

using left-hand rule, force acting on $I_2$ by $I_1$, denoted $F_{21}$[58], points to the left

since $F_{21}$ and $F_{12}$ are a pair of action-reaction, so $F_{12}$ points to the right




similar discussions apply for $I_1$ and $I_2$ flowing in opposite directions (left as exercise)

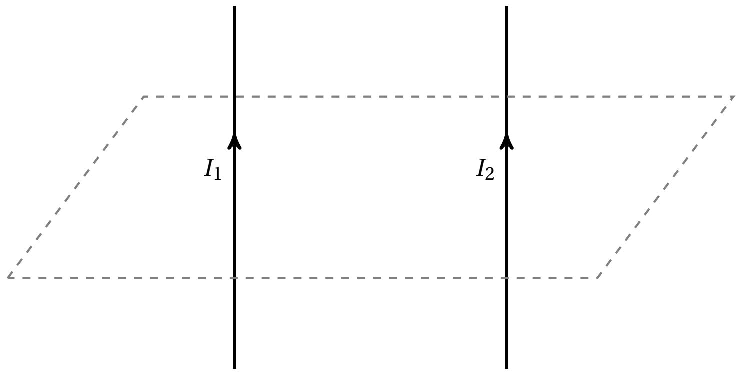
hence we come to an conclusion: *parallel currents attract, and anti-parallel currents repel*[59]

**Question 8.5** A light coil of wire of several loops is suspended from a fixed point. When an electric current is switched on in the coil, state and explain the change in the separation between the loops.

**Question 8.6** Two long parallel wires carry currents of $I_1$ =3.0 A and $I_2$ =5.0 A in the same direction as shown. The flux density at a point at perpendicular distance $r$ from a straight long wire carrying current $I$ is given by $B = \dfrac{\mu_0 I}{2\pi r}$.

[58] $F_{AB}$ denotes the force acted on object $A$ by object $B$.

[59] The S.I. unit for current, ampere, is defined using two long wires parallel to each other carrying the same current in the same direction. If the wires are separated by 1 m and the magnetic force experienced per metre is $2.0 \times 10^{-7}$ N, then the current is of 1 A.




(a) Draw at least three lines to show the magnetic field due to $I_1$.

(b) State and explain the direction of force acting on $I_2$.

(c) Given that $I_1$ and $I_2$ are separated by 4.0 cm, what is flux density due to $I_1$ at $I_2$?

(d) What is the magnetic force per unit length acting on $I_2$?

(e) What is the magnitude and the direction of the force per unit length acting on $I_1$?

**Question 8.7** If the current in one of the two wires in Question 8.6 is replaced by an alternating current, then the two wires should begin to vibrate. However, the wires are not observed to move. Make some reasonable estimates and explain why the vibration is not observed.

## 8.3 magnetic force on charged particles

electric currents are formed by moving charges, since current-carrying wires in a magnetic field experience force, charged particle moving in magnetic field should also experience force

### 8.3.1 magnetic force on charged particles

starting from $F = BIl\sin\theta$, let's substitute $I = nAqv \quad \Rightarrow \quad F = B(nAqv)l\sin\theta$

notice $n$ is the number density of charged particles, so $nAl$ together gives the total number

if we look at the force acting on *one* charged particle, then $nAl = 1$, so: $F = Bqv\sin\theta$

➢ magnitude of magnetic force on charged particle: $\boxed{F = Bqv\sin\theta}$

$B$ is magnetic flux density, $q$ is electric charge of the particle, $v$ is its velocity

$\theta$ refers to the angle between $B$ and $v$

➢ direction of magnetic force can be determined using *Fleming's left-hand rule*

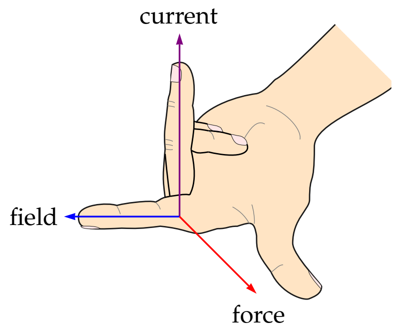
for $+q$, current in same direction as $v$

for $-q$, current in opposite direction as $v$

➢ force depends on *perpendicular component* of $B$

if $v \perp B$, magnetic force $F = Bqv$

if $v \parallel B$, there is no magnetic force

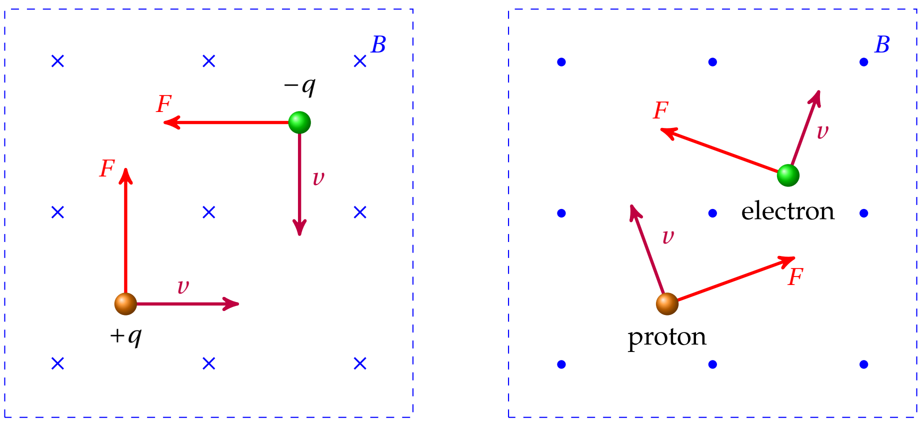



Fleming's left-hand rule

when particle moves at angle $\theta$ to $B$, contribution to the force only comes from the *perpendicular component*, giving rise to the $\sin\theta$ factor

➢ magnetic force always perpendicular to both velocity $v$ and magnetic field $B$[60]

**Example 8.6** Use the left-hand rule to find the direction of the magnetic forces acting on the following moving charges. Check yourself.




(a) (b)

**charged particles in electric & magnetic fields**

in either electric field or magnetic field, charged particle may experience a force[61]

in this section, we will compare the difference between electric field and magnetic field

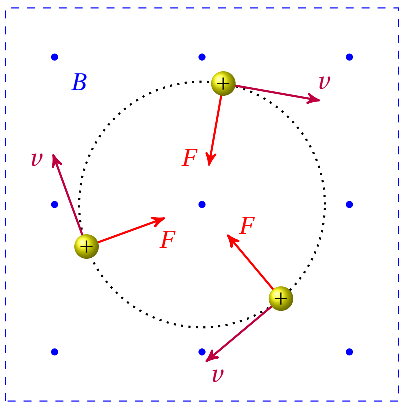
[60] Vector form of the formula: $\vec{F} = q\vec{v} \times \vec{B}$. $(*)$

[61] The combination of electric and magnetic force on the charge due to electromagnetic fields is called the *Lorentz force*: $\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$. If you have a good knowledge in maths, everything we cover in this section can be recovered from this vector equation. $(*)$

| | electric field | magnetic field |
|---|---|---|
| strength of the field | electric field strength $E$ | magnetic flux density $B$ |
| magnitude of force | $F_E = Eq$ | $F_B = Bqv\sin\theta$ |
| whether force depends on velocity | no dependence, acts equally on stationary or moving charges | depend on perpendicular component of velocity |
| direction of force | $F_E \parallel E$<br>same as/opposite to $E$ for $+q/-q$ | $F_B \perp B$ and $F_B \perp v$<br>use left-hand rule |
| effect of force on motion of particle | can change both magnitude and direction of velocity | only changes direction of velocity, cannot change magnitude of velocity |
| work done by force | work can be done, can define notions of potential and P.E. | magnetic force does no work for charged particles |

### 8.3.2 charged particle in uniform magnetic fields

suppose a charged particle is moving at right angle to a uniform magnetic field




circular motion of a charged particle in uniform magnetic field

magnetic force $F_B$ always at right angle to motion, so $F_B$ keeps changing direction of velocity

uniform field means a constant force, so $F_B$ deflects the particle at the same rate

the particle should describe a *circular* path!

for a charged particle moving in a uniform magnetic field, *magnetic force provides centripetal force for circular motion*: $\boxed{Bqv = \frac{mv^2}{r}}$, or $\boxed{Bqv = m\omega^2 r}$

➢ can solve for radius of the orbiting particles: $r = \frac{mv}{Bq}$

- $v \nearrow \Rightarrow r \nearrow$, faster particles take larger circles
- $B \nearrow \Rightarrow r \searrow$, stronger magnetic field, larger centripetal force, smaller circles
- $m \nearrow \Rightarrow r \nearrow$, larger mass, larger inertia, so larger circles

➢ radius of curvature relates to charge-to-mass ratio (also called *specific charge*) of the particle

rearrange the equation we have $\frac{q}{m} = \frac{v}{Br}$, which can be computed using experimental data

**Example 8.7** An $\alpha$-particle travelling at $2.5 \times 10^4$ m s$^{-1}$ enters a region of uniform magnetic field. The field has flux density of 5.4 mT and is normal to direction of particle's velocity. What is the radius of $\alpha$-particle's path?

✍ radius of circular arc: $r = \frac{mv}{Bq} = \frac{4 \times 1.66 \times 10^{-27} \times 2.5 \times 10^4}{5.4 \times 10^{-3} \times 2 \times 1.60 \times 10^{-19}} \approx 0.096 \text{ m} \approx 9.6 \text{ cm}$ □

**Question 8.8** For an $\alpha$-particle and a $\beta$-particle entering the same uniform magnetic field at a same speed, compare the radius of their paths.

**Question 8.9** For a charged particle undergoing circular motion in a uniform magnetic field, show that its angular velocity is independent of the radius of its path.

**Question 8.10** If a charged particle enters a uniform magnetic field at an angle $\theta \neq 90°$, state and explain the path of this particle. (Hint: think about components of its velocity.)

### 8.3.3 mass spectrometer

a **mass spectrometer** is a device to measure the charge-to-mass ratio of charged particles

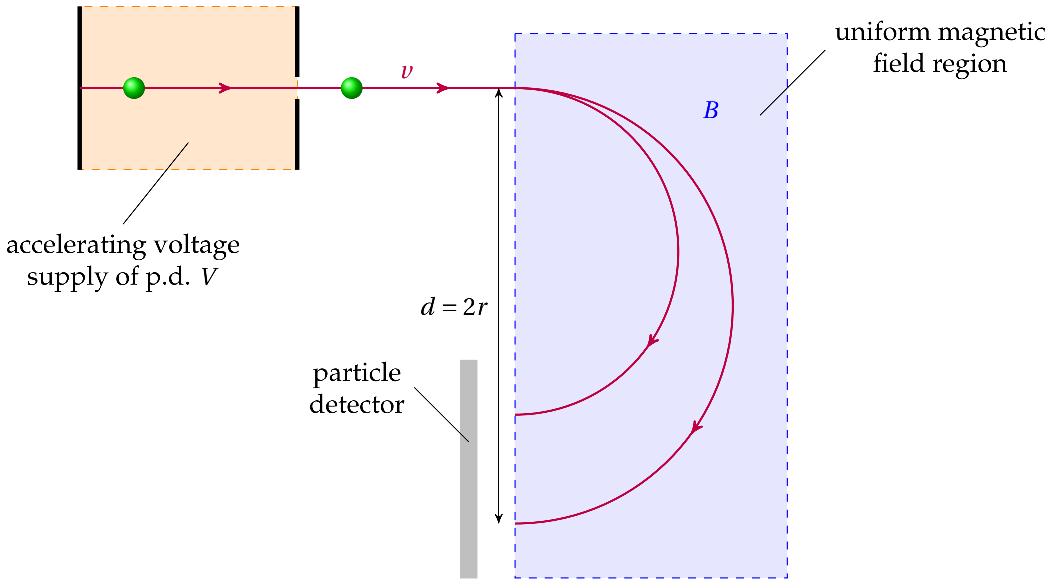
charged particles accelerated through an electric field: $\frac{1}{2}mv^2 = qV$

they then enter a uniform magnetic field: $Bqv = \frac{mv^2}{r} \Rightarrow v = \frac{Bqr}{m}$

eliminating $v$: $v^2 = \frac{2qV}{m} = \frac{B^2q^2r^2}{m^2} \Rightarrow \frac{q}{m} = \frac{2V}{B^2r^2}$

we can measure $V, B, r$ in practice, so the charge-to-mass ratio $\frac{q}{m}$ is worked out

different particles have different values of $\frac{q}{m}$, unknown particles can be identified




deflection of two charged particles in a mass spectrometer

**Question 8.11** In the figure above, two paths of deflected particles are shown. Give reasons why the radius of the circular path can be different.

**Question 8.12** If the particles sent into the mass spectrometer are positively-charged, in which direction should the magnetic field be applied?

**Question 8.13** A particle carrying a charge of $+e$ enters a uniform magnetic field of 8.8 mT at right angles with an initial speed of $1.4 \times 10^5$ m s$^{-1}$. It describes a semi-circle with diameter of 66 cm. (a) Find the mass of the particle. (b) Suggest possible composition of this particle.

### 8.3.4 cyclotron

a **cyclotron** is a type of *particle accelerator*

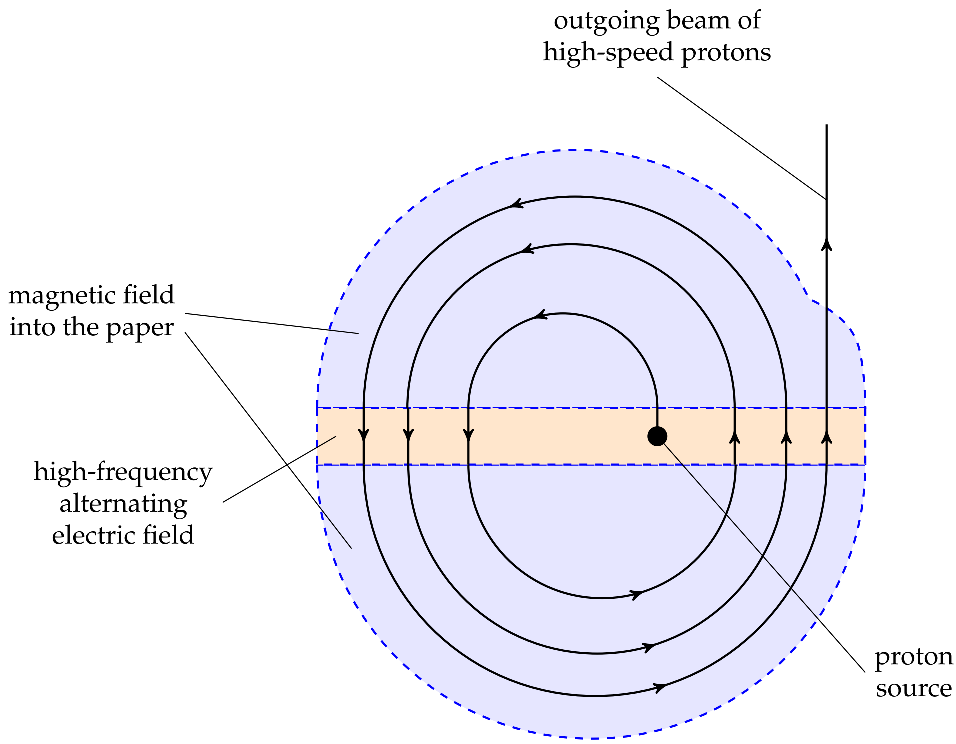
the idea is to make use of a magnetic field to guide moving charges into a spiral path between accelerations by an electric field

as the particle enters and leaves the region of electric field, it gains extra energy of $qV$

it then follows a semi-circular path under magnetic force and re-enters the electric field

polarity of the electric field is reversed so the particle continues to accelerate across the gap

energy of particle increases by $qV$, it then moves in a larger semi-circle in magnetic field




protons being accelerated in a cyclotron

repeat this process, the particle leaves the exit port with very high speed

➢ cyclotron frequency

for charged particles circulating in a uniform magnetic field: $Bqv = m\omega^2 r$

time to complete one full turn is: $T = \frac{2\pi r}{v} = \frac{2\pi}{v} \times \frac{mv}{Bq} \Rightarrow T = \frac{2\pi m}{Bq}$

this shows the period is independent of radius of the circular path!

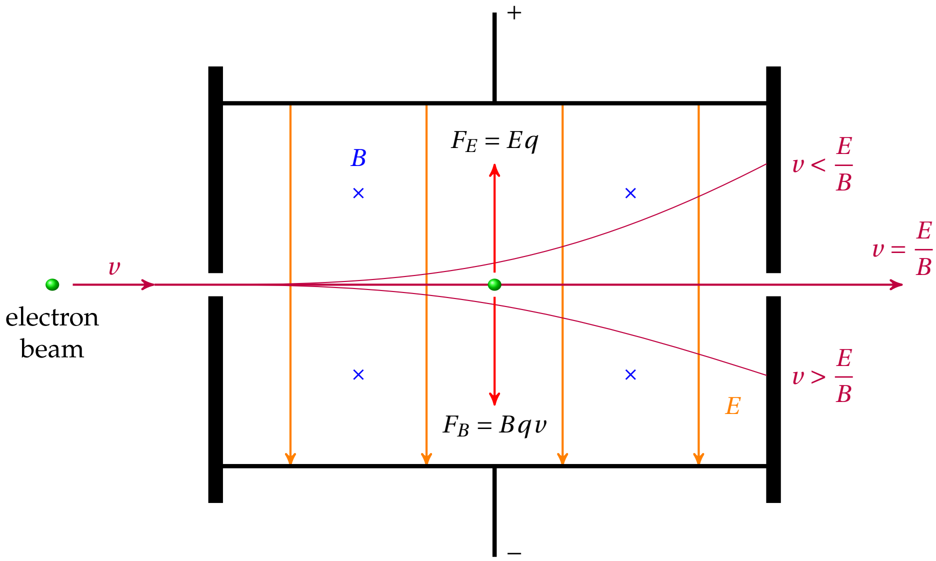
if an alternating voltage is applied at frequency $f = \frac{Bq}{2\pi m}$, particles can be accelerated continuously at just the right time when they cross the gap

this frequency is known as the *cyclotron frequency*

**Question 8.14** Protons are accelerated in a cyclotron where the magnetic field region has a uniform flux density of 1.5 T and the voltage between the gap is 20 kV. (a) Find the frequency of the voltage supply needed. (b) Find the number of circles the protons must make in order to gain an energy of 10 MeV.

### 8.3.5 velocity selector

**velocity selector** is a device to produce a beam of charged particles all with same speed $v$




velocity selector

let's consider a beam of electrons passing through a region where both uniform electric field and magnetic field are applied, electrons at desired speed $v$ are *undeflected*

no net force acting on these electrons, so equilibrium between electric and magnetic force

$$F_E = F_B \quad \Rightarrow \quad Eq = Bqv \quad \Rightarrow \quad \boxed{v = \frac{E}{B}}$$

for particles entering the same region with higher speed, $F_B > F_E$, they deflect downwards

for particles entering the same region with lower speed, $F_B < F_E$, they deflect upwards

**Question 8.15** If electrons at speed $v$ are undeflected when they pass through the velocity selector, what about a beam of $\alpha$-particles entering the same region at the same speed $v$?

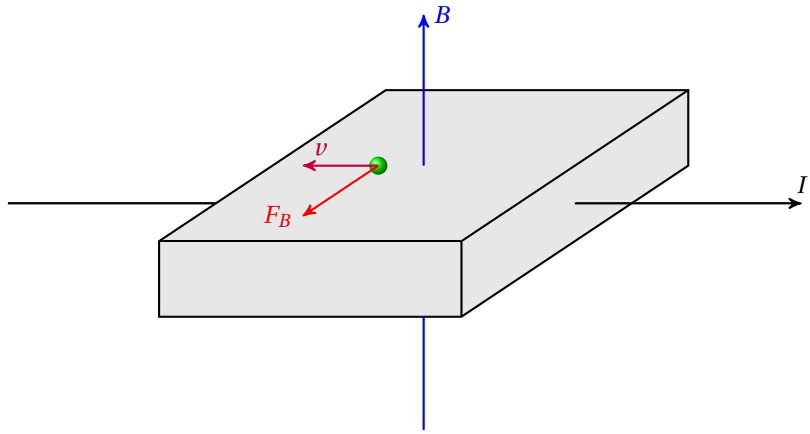
**Question 8.16** A uniform magnetic field with flux density $6.0 \times 10^{-2}$ T is applied out of the plane of the paper. A beam of protons are travelling into this region at a speed of $3.5 \times 10^4 \text{m s}^{-1}$. (a) In which direction do the protons deflect? (b) A uniform electric field is now applied in the same region so that the protons become undeviated. What is the strength and the direction of this electric field?

## 8.4 Hall effect

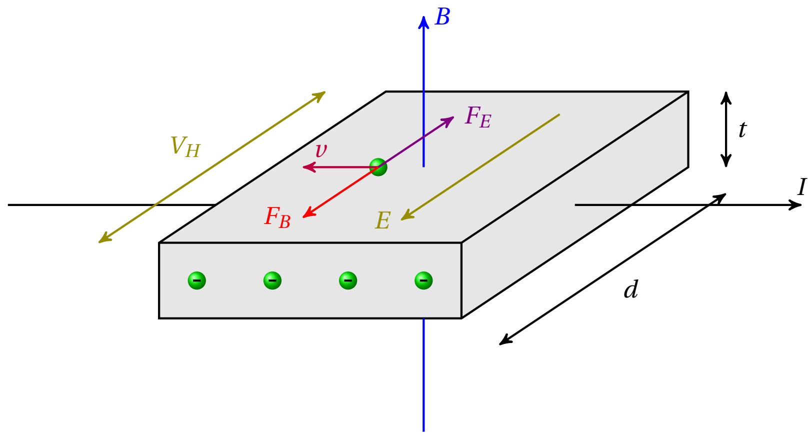
when an electric current passes through a conductor surrounded by an external magnetic field perpendicular to the current, a transverse potential difference is produced

this phenomenon is known as the **Hall effect** (discovered by American physicist *Edwin Hall* in 1879), the voltage difference is called the **Hall voltage**

suppose magnetic field goes upward, current flows to the right (see figure), then charge carriers (assumed to be electrons for now) experience a magnetic force pointing out of paper




this magnetic force causes a build-up of extra charge on front side of conductor, which produces an internal electric field $E$ inside conductor, i.e., a potential difference $V_H$ is formed



as charge carriers accumulate on one side, Hall voltage opposes further migration of charges

steady $V_H$ is established when magnetic force and electric force are balanced: $Eq = Bqv$

- strength of internal electric field is related to Hall voltage: $E = \dfrac{V_H}{d}$

  where $d$ is width of the conductor

- drift velocity $v$ of particles is related to current $I$: $I = nAqv$

  where $A$ is cross section of conductor and $n$ is number density of charge carriers

we now have: $\dfrac{V_H}{d} q = Bq \dfrac{I}{nAq} \Rightarrow V_H = \dfrac{BId}{nAq}$

note that cross section $A = dt$, where $t$ is thickness of conductor as shown

we therefore obtain an useful expression for Hall voltage: $V_H = \dfrac{BId}{n(td)q} \Rightarrow \boxed{V_H = \dfrac{BI}{ntq}}$

➢ to produce noticeable $V_H$, we want smaller $n$ and smaller $t$

- $n \downarrow \Rightarrow$ few free charge carriers $\Rightarrow$ *semi-conductors* are preferable
- $t \downarrow \Rightarrow$ small thickness $\Rightarrow$ *thin slice* of component is preferable

➢ polarity of $V_H$ is determined by nature of charge carries

charge carriers can be *free electrons* (negatively charged) or *holes* (positively charged)

then Hall voltage induced would have opposite polarity

➢ if current $I$ forms angle $\theta$ with flux density $B$, again only perpendicular component matters

expression for Hall voltage becomes: $V_H = \dfrac{BI}{ntq} \sin\theta$

➢ apply a fixed current in a conductor, $V_H$ is proportional to $B$

so magnetic flux density can be calculated once we find $V_H$

one can make use of Hall effect to build a *Hall probe*, a device used to measure flux density $B$

when using a Hall probe, one should rotate and record the greatest voltage reading to ensure the current applied is at right angle to the external magnetic field

**Question 8.17** Why is it difficult to detect Hall voltage in a thin slice of copper?

**Question 8.18** A Hall probe is placed near to one end of a strong magnet. State and explain the variation in the Hall voltage as the probe is rotated for one complete revolution.

# Chapter 9

# Electromagnetic Induction

## 9.1 magnetic flux

**magnetic flux** is defined as the product of a closed area and the magnetic flux density going through it at right angles: $\boxed{\phi = BA\cos\theta}$

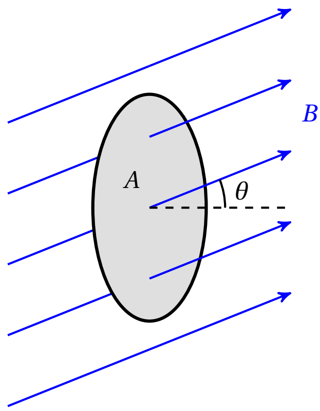
- unit for magnetic flux: $[\phi] = \mathrm{Wb}$, $1\ \mathrm{WB} = 1\ \mathrm{T \cdot m^2}$
- if magnetic field is perpendicular to the area

  magnetic flux simply becomes: $\boxed{\phi = BA}$
- for a coil with $N$ turns, total magnetic flux is

  $$\boxed{\Phi = N\phi = NBA}$$

  this is called **magnetic flux linkage**
- magnetic flux $\phi$ can be graphically thought as the number of field lines through area $A$

  cutting of field lines means a change in flux
- change in flux can be caused by many processes, for example
  - moving a magnet towards/away from a coil
  - varying current in a solenoid/electromagnet
  - inserting an iron core into a solenoid/electromagnet
  - pushing a straight conductor through a magnetic field
  - rotating a coil in a magnetic field
  - …




**Question 9.1** Explain why the processes mentioned above would give rise to a change in magnetic flux.

## 9.2 laws of electromagnetic induction

electromagnetic induction is the phenomena that magnetism produces electricity

the laws of electromagnetic induction were discovered by *Michael Faraday* and *Heinrich Lenz* in the 1830s, and later described mathematically by *James Clerk Maxwell*

we will study in what conditions voltages and currents could be induced, and how to find their magnitudes and polarities

### 9.2.1 Faraday's law

**Faraday's law** states that induced e.m.f./voltage is proportional to rate of change in magnetic flux (linkage): $\boxed{\mathscr{E} \propto \frac{\Delta\Phi}{\Delta t}}$

➢ Faraday's law gives the *magnitude* of the induced e.m.f.

the key here is the *change* in flux: as long as flux changes, e.m.f. will be induced

➢ if $\mathscr{E}$, $\Phi$, $t$ are all given in SI units, this proportional relation becomes an identity: $\boxed{\mathscr{E} = \frac{\Delta\Phi}{\Delta t}}$

**Example 9.1** A coil of 80 turns is wound tightly around a solenoid with a cross-sectional area of 35 cm$^2$. The flux density at the centre of the solenoid is 75 mT. (a) What is the flux linkage in the coil? (b) The current in the solenoid is *reversed* in direction in a time of 0.40 s, what is the average e.m.f. induced?

✍ flux linkage: $\Phi = NBA = 80 \times 75 \times 10^{-2} \times 35 \times 10^{-4} = 0.21$ Wb

average e.m.f. induced: $\mathscr{E} = \frac{\Delta\Phi}{\Delta t} = \frac{(+\Phi)-(-\Phi)}{\Delta t} = \frac{2 \times 0.21}{0.40} = 1.05$ V □

### 9.2.2 Lenz's law

**Lenz's law** states that induced e.m.f. or current is always in the direction to produced effects that *oppose* the change in flux that produced it

➢ Lenz's law gives the *polarity* of the induced current

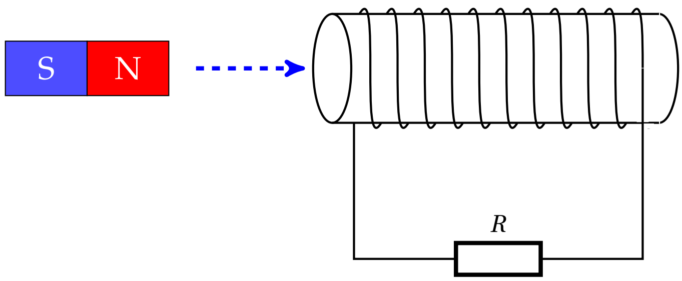
the key here is the nature dislikes any change in flux: effects of induced current always *oppose* the cause of its production

➢ Lenz's law is a consequence of the conservation of energy

since induced current dissipates electrical energy as heat, it must cause loss of original forms of energy possessed by the system

if the change in flux is caused by motion, then magnetic force on/by the induced current must *resist* this motion

**Example 9.2** A coil of 200 turns is connected to a resistor $R = 4.0\ \Omega$. The coil has a diameter 10 cm. Initially a bar magnet is at great distance from the coil. The magnet is then inserted into the coil and field inside coil becomes 0.40 T. The process occurs within a duration of 2.0 s.




(a) What is average induced e.m.f. in the coil? (b) What is average induced current through resistor $R$? (c) In which direction does this current flow?

✍ change in flux linkage: $\Delta\Phi = NBA - 0 = 200 \times 0.40 \times \pi \times 0.050^2 \approx 0.628$ Wb

average e.m.f. induced: $\mathscr{E} = \frac{\Delta\Phi}{\Delta t} = \frac{0.628}{2.0} \approx 0.314$ V

average current induced: $I = \frac{\mathscr{E}}{R} = \frac{0.314}{4.0} \approx 0.0785$ A

as magnet approaches, coil experiences an increasing flux to the right

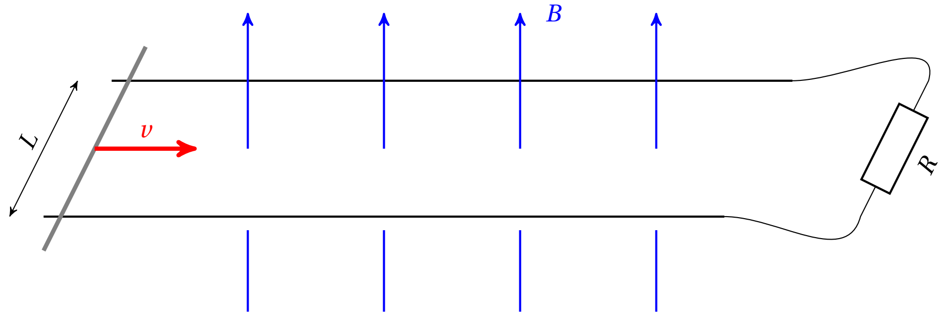
according to Lenz's law, field due to induced current acts to the left to oppose the change

alternatively, left side of the coil should behave as a north pole to oppose motion of magnet

use right-hand rule, we find induced current flows through $R$ to the right

**Question 9.2** If the magnet is now pulled away from the coil, state ad explain the direction of the induced current.

**Example 9.3** Two parallel metal tracks are separated by a distance of $L = 45$ cm and placed in a uniform magnetic field of 0.80 T. The tracks are connected to a resistor $R = 6.0\ \Omega$. A long metal rod is being pushed under an external force at $v = 3.0$ m s$^{-1}$ along the tracks as shown.




(a) What is the induced current through resistor? (b) In which direction does this current flow? (c) For the rod to travel at constant speed, what is magnitude of external force required?

✍ change of flux in time interval $\Delta t$ is: $\Delta\phi = \Delta(BA) = B\Delta A = BLv\Delta t$

e.m.f. induced: $\mathscr{E} = \dfrac{\Delta\phi}{\Delta t} \Rightarrow \boxed{\mathscr{E} = BLv}$

induced current: $I = \dfrac{\mathscr{E}}{R} = \dfrac{BLv}{R} = \dfrac{0.80 \times 0.45 \times 3.0}{6.0} = 0.18A$

according to Lenz's law, magnetic force on induced current opposes the rod's motion of cutting field lines, so magnetic force acts to the left

using Fleming's left-hand rule, we find induce current flows in anti-clockwise direction

if rod travels at constant speed, then equilibrium between external push and magnetic force

external force required: $F_{\text{ext}} = F_B = BIL = 0.80 \times 0.18 \times 0.45 \approx 0.065$ N □

**Question 9.3** In Example 9.3, can you find alternative arguments to determine the direction of the current induced as the rod cuts through the magnetic field?

### Hall voltage & induced voltage in a coil

in this section, we compare readings on voltmeter connected to a Hall probe and a coil

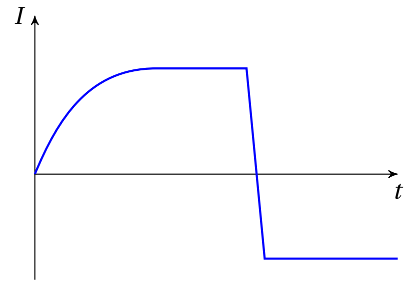
Hall probe picks up voltage proportional to flux density: $V_H \propto B$

coil picks up induced e.m.f. proportional to *rate of change* in flux: $\mathscr{E} \propto \dfrac{\mathrm{d}\Phi}{\mathrm{d}t}$

**Question 9.4** If a Hall probe is placed near a permanent magnet, what voltage do you measure?

If a small coil is placed in the same field, state and explain whether you can measure a non-zero voltage in the coil. If not, state three different ways in which you can produce a voltage.

**Question 9.5** A Hall probe is placed near one end of a solenoid that carries a varying current as shown in the graph. (a) Sketch the variation of Hall voltage with time. (b) If Hall probe is replaced by a small coil parallel to the solenoid's end, sketch the variation of the voltage induced in the coil.




## 9.3 applications of electromagnetic induction

### 9.3.1 magnetic braking

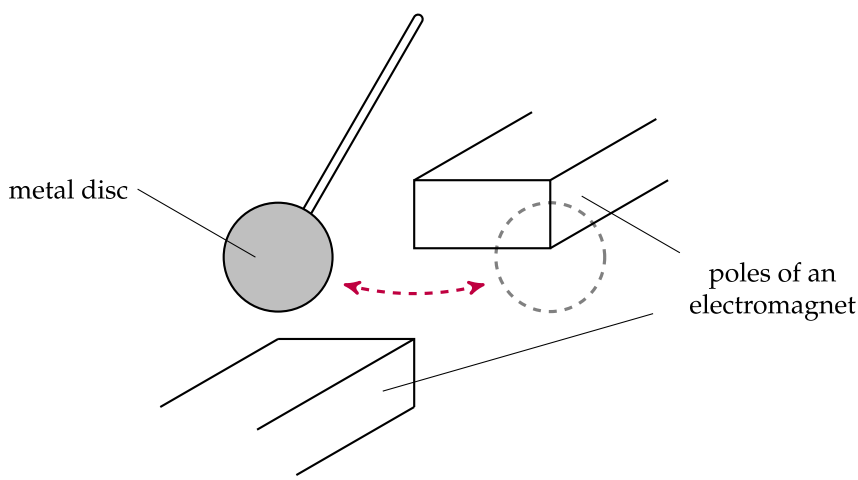
induction brakes make use of induced current to dissipate kinetic energy of a moving object

in this section, we will look at two simple demonstrations, but the principle can be used in braking system of high-speed trains, roller-coasters, etc.

#### the damped pendulum

a metal disc can swings freely between the poles of an electromagnet

when the electromagnet is switched on, the disc comes to rest very quickly



as disc moves in and out of field, change in flux gives rise to e.m.f induced

note that different parts of the disc experience different rates of change in flux, different e.m.f. is induced in different parts for the disc

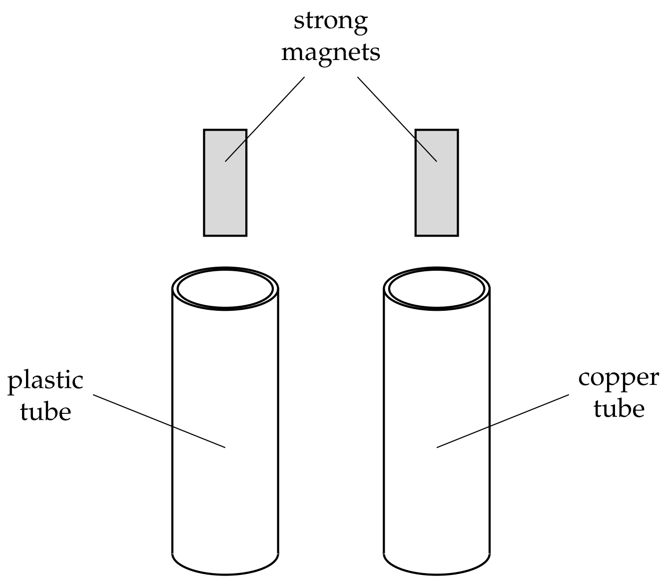
this causes circulating currents flowing in the disc, called **eddy currents**

- vibrational energy is lost as heat due to the eddy currents induced
- magnetic force on induced current causes damping

so amplitude of oscillation decreases quickly

**falling magnet**

suppose we drop two strong magnets down a plastic tube and a copper tube




as magnet falls, tube experiences change in flux, so e.m.f is induced

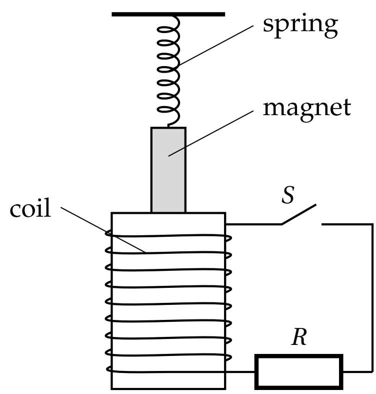
since plastic is an insulator, so no current flows in plastic tube, magnet undergoes free fall

but in the conducting copper tube, eddy current is induced in the tube

- energy is lost as heat due to induced current, not all G.P.E. converted into K.E.
- induced current exerts magnetic force on magnet to oppose its motion, acceleration $a < g$

so magnet falls more slowly through the copper tube

**Question 9.6** A magnet is suspended from the free end of a spring. When displaced vertically and released, the magnet can oscillate in and out of a coil (see diagram). The switch $S$ is initially open, there is negligible change in the amplitude. However, when the switch is closed, the amplitude is seen to decrease quickly. Explain the reasons.

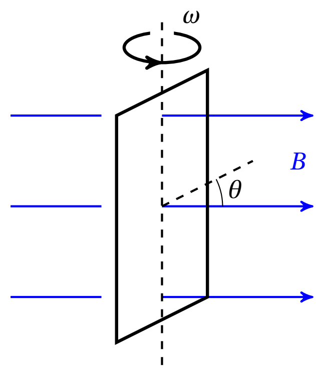



### 9.3.2 the generator

imagine a coil rotates with constant angular speed $\omega$ in a uniform magnetic field $B$

let's assume that the coil initially lies in parallel to the field, i.e., $\theta = 0$ at $t = 0$

at time $t$, coil forms an angle $\theta = \omega t$ with the magnetic field (see diagram)




magnetic flux linkage through coil is:

$$\Phi = NBA\sin\theta = NBA\sin\omega t$$

magnitude of induced e.m.f. is:

$$\mathscr{E} = \frac{\mathrm{d}\Phi}{\mathrm{d}t} = \omega NBA\cos\omega t$$

this is a sinusoidal voltage with maximum value: $\mathscr{E}_{\max} = \omega NBA$

for a coil rotating in a magnetic field like this, an *alternating current* is generated

this is basically how a *generator* works

in practice, it is a strong electromagnet rotating inside a large coil to generate electricity

**Question 9.7** State and explain the effect on the voltage generated if the coil rotates faster.

**Question 9.8** For the coil rotating at a uniform angular speed in a uniform magnetic field, is the magnetic flux in phase with the e.m.f. generated? If not, what is the phase difference?

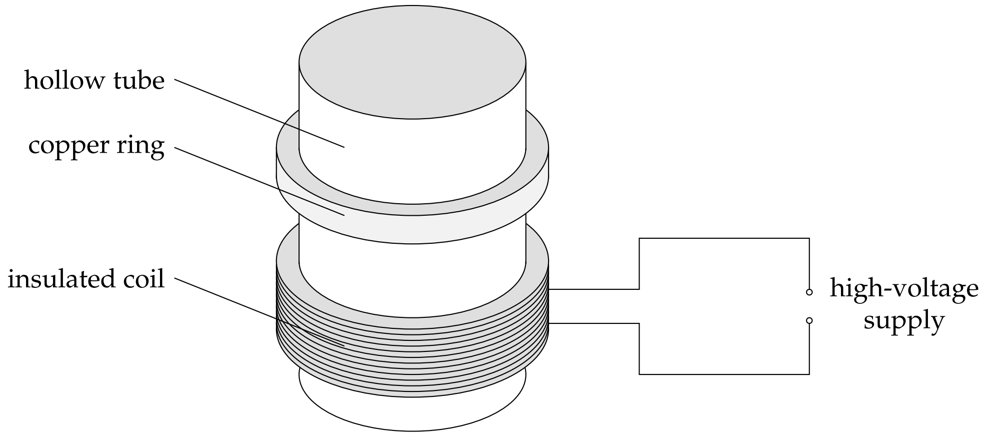
**Question 9.9** Does the generator output the maximum voltage when the rotating coil is in parallel to the magnetic field or when the coil is at right angle to the field?

### 9.3.3 electromagnetic gun

a coil of wire of many turns is wound on a hollow tube

a light copper ring that can move freely along the tube is placed on the coil

let's find out what happens to the ring when a high-voltage supply is switched on




when supply is switched on, flux density in coil greatly increases

sudden change in magnetic flux could induce a huge e.m.f. in copper ring

induced current in ring would experience a repulsive magnetic force

this repulsive force could be way larger that ring's weight, ring would jump out from tube

if the ring is replaced by a small metal sphere placed inside the tube, the same strong magnetic force arising from a sudden change in flux could fire the sphere at very high speed

**Question 9.10** The coil is now connected to a stable d.c. voltage supply. If we quickly insert an iron core into the tube, what might happen to the light copper ring?

#### other applications

as a final remark, electromagnetic induction is widely used in many other areas as well

for those who are interested, you may research on the principles behind wireless charging, induction cooking, contactless payment technology, smart pencils for tablets or computers, etc.

# Chapter 10
# Alternating Currents

a **direct current** (d.c.) flows in one direction only

an **alternating current** (a.c.) reverses its flow direction from time to time

a.c. has certain advantages than d.c., as you will see in this chapter

we will study the mathematics of a.c. and the transmission process for a.c.

## 10.1 sinusoidal a.c.

as we have seen in §9.3.2, currents produced from generators are naturally sinusoidal

sinusoidal a.c. is one of the most important types of a.c. waveform in electrical engineering.

for most cases in this course, we focus on a.c. that varies like a sine wave

### 10.1.1 sinusoidal waveform

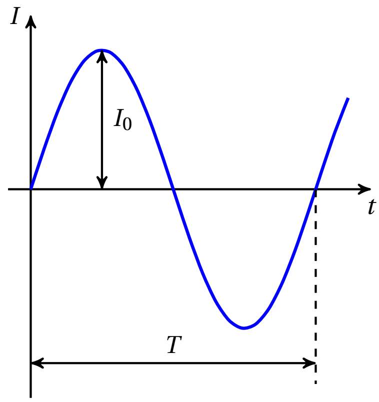
current: $\boxed{I = I_0 \sin \omega t}$ / voltage: $\boxed{V = V_0 \sin \omega t}$

➢ $I_0$, $V_0$ are called *peak current* and *peak voltage* (amplitudes of the a.c. signal)

➢ $\omega$ is **angular frequency**, which describes how fast a.c. signal oscillates (same idea as for simple harmonic motion, see §3)

frequency and period of the signal are given by:

$$\boxed{\omega = 2\pi f}, \quad \boxed{T = \frac{1}{f} = \frac{2\pi}{\omega}}$$




➢ mean current: $\langle I \rangle = 0$, mean voltage: $\langle V \rangle = 0$

this is because a.c. signal fluctuates with time, positive and negative bits cancel out

**Question 10.1** An alternating voltage has a peak value of 16 V and period of 0.10 s. Write down a mathematical equation that describes the variation of this voltage.

**Question 10.2** An alternating voltage is produced from a simple generator. If the rotating speed of the coil in the generator doubles, describe quantitatively the change in the peak value and frequency of the alternating voltage.

**Question 10.3** A student argues that when an alternating current is driven through a resistor, the mean current is zero, so an alternating current does not produce heating power on the resistor. State and explain whether this is correct.

### 10.1.2 power

electrical power dissipated in a resistor: $P = I^2 R$, or, $P = \dfrac{V^2}{R}$

since $I^2, V^2 \geq 0$, $P$ can never be negative, so an a.c. can produce effective power in a resistor

note that for an a.c., $P$ keeps changing with time, as $I$ and $V$ are both varying with time

in everyday life, we are more concerned about the *mean power*

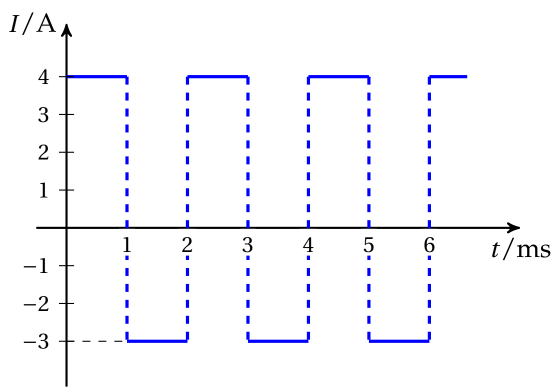
mean power output for an a.c. is: $\langle P\rangle = \langle I^2\rangle R = \dfrac{\langle V^2\rangle}{R}$

so we see the necessity to introduce mean square values $\langle I^2\rangle$ and $\langle V^2\rangle$

let's further introduce root mean square (r.m.s.) values: $I_{\text{rms}} = \sqrt{\langle I^2\rangle}$, and $V_{\text{rms}} = \sqrt{\langle V^2\rangle}$

we can now write the mean power for a.c. as: $\boxed{\langle P\rangle = I_{\text{rms}}^2 R = \dfrac{V_{\text{rms}}^2}{R}}$

**Example 10.1** The variation with time of an alternating current in a resistor of 120Ω is shown. (a) What is the value of its r.m.s. current? (b) What is the mean power dissipated by the resistor?




✍ $\langle I^2\rangle = \dfrac{4^2 + (-3)^2}{2} = 12.5\ \text{A}^2$

$I_{\text{rms}} = \sqrt{\langle I^2\rangle} = \sqrt{12.5} \approx 3.54$ A

$\langle P\rangle = I_{\text{rms}}^2 R = 12.5 \times 120 = 1500$ W □

### 10.1.3 r.m.s. current & r.m.s. voltage

in last section, we studied the mathematical aspect of the r.m.s. value

but we still need a definition for r.m.s. current and r.m.s. voltage from a physical viewpoint

physically, r.m.s. value of an alternating current is defined as follows

> **r.m.s. current/voltage** of an a.c. equals a steady d.c. current/voltage that delivers same average power to a resistive load

➢ for *sine waves*, r.m.s values are related to peak values by: $\boxed{I_{\text{r.m.s}} = \frac{1}{\sqrt{2}} I_0}$ and $\boxed{V_{\text{r.m.s}} = \frac{1}{\sqrt{2}} V_0}$

proof (∗): total energy dissipation in one period is: $W_T = \int_0^T P \mathrm{d}t$

so mean power can be given by: $\langle P \rangle = \frac{W_T}{T} = \frac{1}{T}\int_0^T P \mathrm{d}t$

substitute $P = I^2 R \xlongequal{I = I_0 \sin\omega t} I_0^2 R \sin^2 \omega t$, we have: $\langle P \rangle = \frac{I_0^2 R}{T} \int_0^T \sin^2 \omega t \mathrm{d}t$

the integral is carried out: $\int_0^T \sin^2 \omega t \mathrm{d}t = \frac{1}{2}\int_0^T (1 - \cos 2\omega t)\mathrm{d}t = \frac{1}{2}\left(t - \frac{\sin \omega t}{2\omega}\right)\Big|_0^T = \frac{1}{2}$

now we have: $\langle P \rangle = I_{\text{rms}}^2 R = \frac{1}{2} I_0^2 R \quad \Rightarrow \quad I_{\text{rms}} = \frac{1}{\sqrt{2}} I_0$

a similar calculation for voltage would show: $V_{\text{rms}} = \frac{1}{\sqrt{2}} V_0$ □

➢ it is worth pointing out that the $\frac{1}{\sqrt{2}}$-relation for r.m.s. values only holds for sine waves

for other waveforms, e.g., square waves or triangle waves, numerical constant is different

➢ value of voltages stated for mains electricity supply usually refers to the r.m.s. value[62]

**Example 10.2** An a.c. power supply produces a sinusoidal output across a resistor of 30 Ω. The maximum voltage is found to be 75 V. Find energy dissipated in the resistor in 2.0 minutes.

✍ r.m.s voltage: $V_{\text{rms}} = \frac{1}{\sqrt{2}} V_0 = \frac{75}{\sqrt{2}} \approx 53.0$ V

mean power output: $\langle P \rangle = \frac{V_{\text{rms}}^2}{R} = \frac{53.0^2}{30} \approx 93.8$ W

energy dissipation: $W = \langle P \rangle t = 93.8 \times 2.0 \times 60 \approx 1.13 \times 10^4$ J □

[62] Different countries have different standards. China mains electricity supplies a r.m.s. voltage of 220 V. UK uses a r.m.s. 230 V distribution system. USA has national standard of a r.m.s. 110 V voltage.

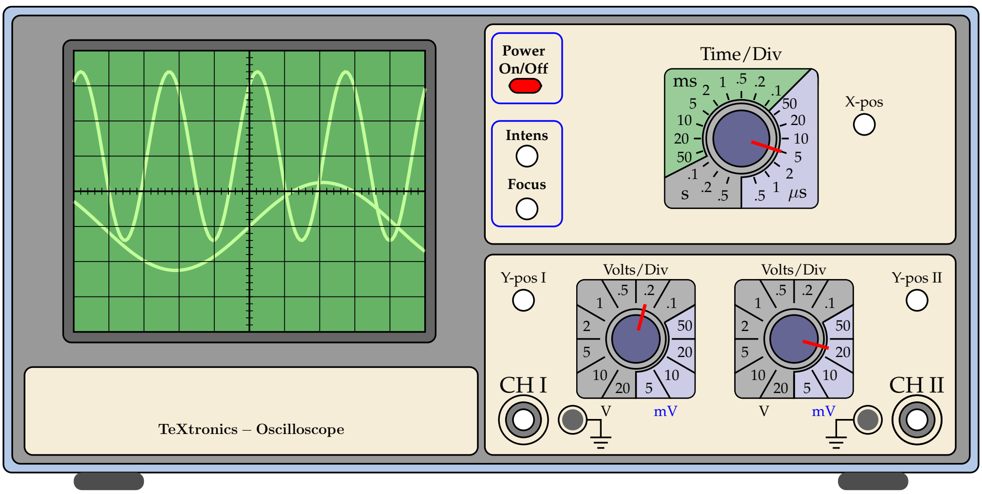
**Question 10.4** An alternating voltage $V$ is represented by the equation: $V = 310\sin(100\pi t)$, where $V$ and $t$ are measured in SI units. For this voltage, find (a) the peak voltage, (b) the r.m.s voltage, (c) the frequency, (d) the mean power when it is applied across a resistor of 250 Ω.

**Question 10.5** What is the average power dissipated in a resistor when the alternating supply has a peak current of 5.0 A and a peak voltage of 8.0 V?

**Question 10.6** The peak value of a sinusoidal alternating current is equal to a steady direct current. When they are applied to the same load, what is the ratio of power dissipation $\frac{P_{\text{d.c.}}}{P_{\text{a.c.}}}$?

### 10.1.4 measurement of a.c. with an oscilloscope

to measure a varying voltage, we can use a **cathode-ray oscilloscope** (c.r.o.) (see figure[63])




an oscilloscope is basically an electron deflection tube that uses a beam of electrons to trace the input voltage as a function of time.

electron beam is controlled by two sets of parallel plates

horizontal plates control how fast beam moves across screen, giving a horizontal time axis

a.c. voltage is applied across vertical plates, causing beam to bend upwards or downwards

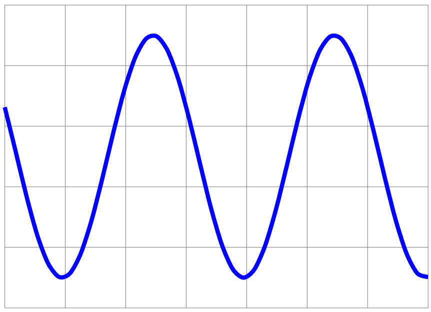
when beam hits fluorescent screen, entire trace of the beam can be seen across the screen

[63]The illustration of the oscilloscope was created by *Hugues Vermeiren*. The source code is downloaded from TeXample: `http://www.texample.net/tikz/examples/textronics-oscilloscope/`.

➢ to take readings from oscilloscope, remember it displays a voltage against time graph

- **voltage gain**, or **Y-gain**, tells the number of volts per vertical division
- **time base** gives the time unit per horizontal division

**Example 10.3** An oscilloscope displays an a.c. voltage signal as shown.



suppose the time base setting is 10 ms/div, and the voltage gain is 5 V/div.

4 vertical divisions from highest to lowest, so

peak voltage: $V_0 = \frac{1}{2} \times 4 \times 5 = 10$ V

3 horizontal divisions between peak to peak, so

period: $T = 3 \times 10 = 30$ ms

frequency $f = \frac{1}{T} = \frac{1}{30 \times 10^{-3}} \approx 33.3$ Hz □

**Question 10.7** In Example 10.3, if the time-base setting is 5 ms/div and the voltage gain is 2 V/div, write down an equation that represents this alternating voltage.

## 10.2 power supply systems

### 10.2.1 high-voltage transmission

electricity is sent from power stations to consumers around the country

for long-distance transmission, we need minimise energy losses

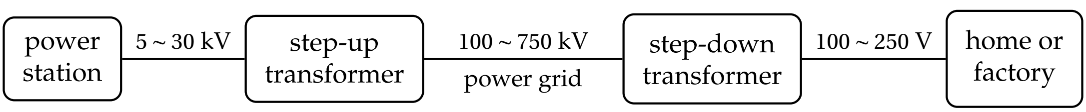
power dissipated due to resistance in cables is $I^2R$, should transmit at low currents

since output power of a power station is fixed, low current means high voltage

so transmission at high voltages minimises energy loss in power grids

but for reasons of safety and efficiency, desirable to have low voltages at both generating end (power station) and receiving end (home or factory)

this requires converting a.c. into higher or lower voltages ⇒ need for *transformers*

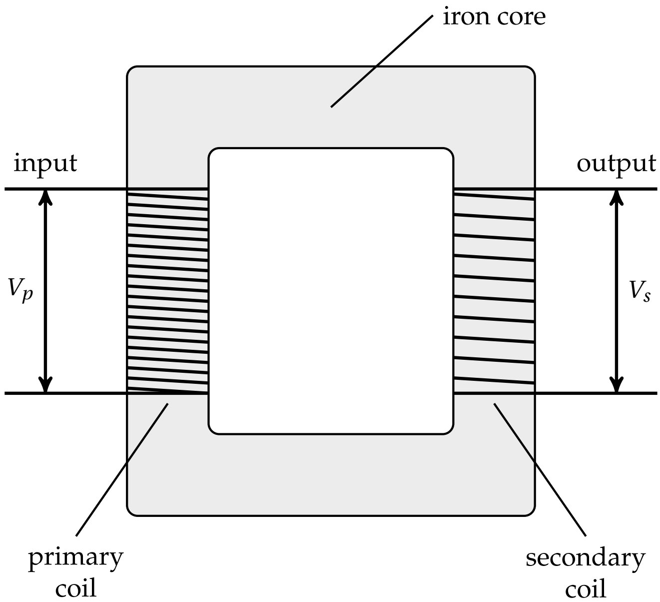


#### 10.2.2 transformers

**transformer** is a device that can change values of voltage for an alternating current

transformers are an important part of the national power grid

transformers are essential for transmission, distribution, and utilization of electrical energy




structure of a typical transformer

➢ principle of transformers

a.c. flowing in **primary coil** (input) produces a changing magnetic field, i.e. a changing flux

this flux is linked with **secondary coil** through iron core

an a.c. voltage with same frequency is then induced in secondary coil (output)

➢ **turns-ratio equation** for a transformer

assume no loss in magnetic flux, i.e., all field inside transformer's iron core

from Faraday's law: $V_p = N_p \frac{\mathrm{d}\phi}{\mathrm{d}t}$, $V_s = N_s \frac{\mathrm{d}\phi}{\mathrm{d}t}$, where $N_p$, $N_s$ are number of turns of coils

cancel out $\frac{\mathrm{d}\phi}{\mathrm{d}t}$, we find: $\boxed{\frac{V_s}{V_p} = \frac{N_s}{N_p}}$

if $N_p < N_s$, output voltage is increased → *step-up transformers*

if $N_p > N_s$, output voltage is decreased → *step-down transformers*

➢ if transformer is 100% efficient, then input power equals output power

for ideal transformers: $\boxed{I_p V_p = I_s V_s}$

➢ in practice, there always exists losses of energy from the transformer

causes of energy loss from the transformer include

- heat produced by *eddy currents* induced in iron core

  this is reduced by *laminating* the core with insulate layers

- heat generated in coils due to resistance

  can use thick copper wire to minimise resistance

- leakage of magnetic flux into surroundings

  transformer's core is made of a continuous loop of iron to minimise this effect

**Example 10.4** An ideal transformer has 200 turns on the primary coil and 5000 turns on the secondary coil. The r.m.s. input voltage to the primary coil is 8.0 V. What is the *peak* voltage across a resistor connected to the secondary coil?

✍ $$\frac{V_{s,0}}{V_{p,0}} = \frac{N_s}{N_p} \quad \Rightarrow \quad \frac{V_{s,0}}{\sqrt{2}V_{p,\text{rms}}} = \frac{N_s}{N_p} \quad \Rightarrow \quad \frac{V_{s,0}}{8.0 \times \sqrt{2}} = \frac{5000}{200} \quad \Rightarrow \quad V_{s,0} \approx 2830 \text{ V}$$ □

**Question 10.8** An ideal transformer has 6000 turns on its primary coil. It converts a mains supply of 220 V r.m.s. to an a.c. voltage with a peak value of 12.0 V. Find the number of turns on the secondary coil.

**Question 10.9** If a *steady* d.c. voltage is applied to the input of a simple transformer, what is the output voltage produced?

**Question 10.10** State and explain whether the current in the primary coil of a transformer is in *phase* with the voltage induced in the secondary coil.

### 10.2.3 rectifiers

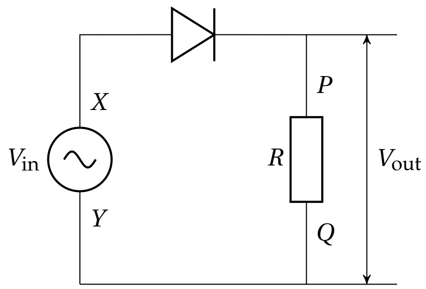
some electronic equipments (e.g., your smartphone, laptop, etc.) must work with d.c.

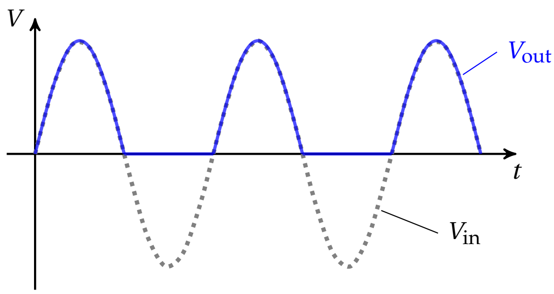
for these appliances require **rectification**, a process that converts an a.c. into a d.c.

rectification uses *diodes*, electronic components that only allow current in one direction

**half-wave rectification**

half-wave rectification uses a single diode




when input terminal $X$ is positive, current can flow through diode, terminal $P$ is positive with respect to $Q$ for load resistor $R$

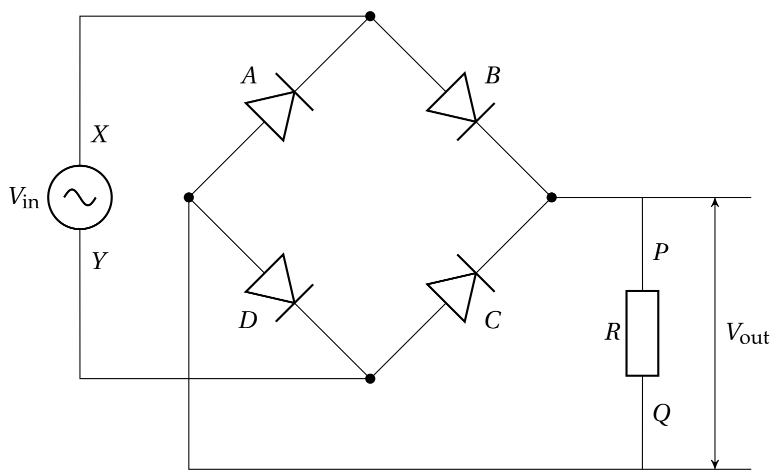
when input terminal $Y$ is negative, flow of current is blocked, so zero output voltage on $R$

$V_{out}$ across load $R$ is in one direction only, i.e., it becomes a d.c.

but power available from half-wave rectifier is only half of supply power

**full-wave rectification**

full-wave rectification requires using a combination of four-diode bridge structure



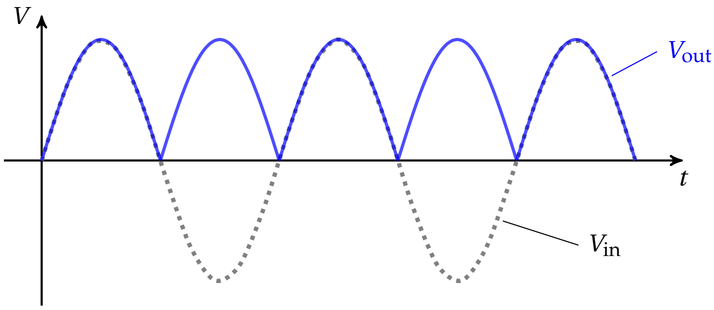

when input terminal $X$ is positive, current flow: $X \to B \to R \to D \to Y$

when input terminal $Y$ is positive, current flow: $Y \to C \to R \to A \to X$

either case, resulting output current in load resistor $R$ always flows in same direction

terminal $P$ is always positive with respect to $Q$ for $R$ no matter what polarity for $V_{\text{in}}$

output voltage across $R$ is now full-wave rectified



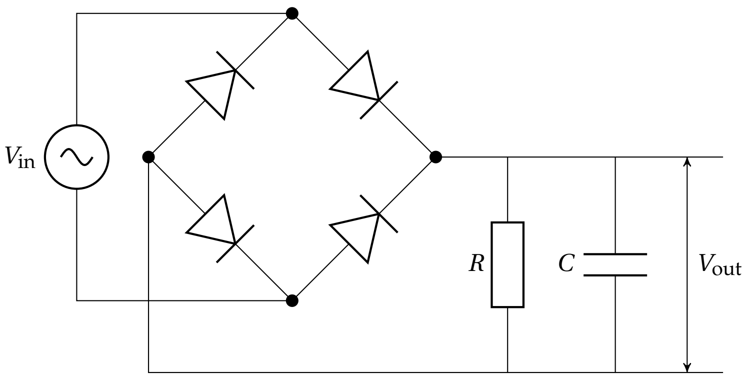

**Question 10.11** If we want to have terminal $Q$ to be positive with respect to $P$ for load resistor $R$, how should we rebuild the bridge rectifier with the same four diodes?

### 10.2.4 smoothing

note that d.c. resulting from rectification still varies with time

to produce steady d.c. from fluctuating d.c., a process called **smoothing** is carried out

smoothing uses *capacitors*, which are connected in *parallel* with the load resistor (see figure)

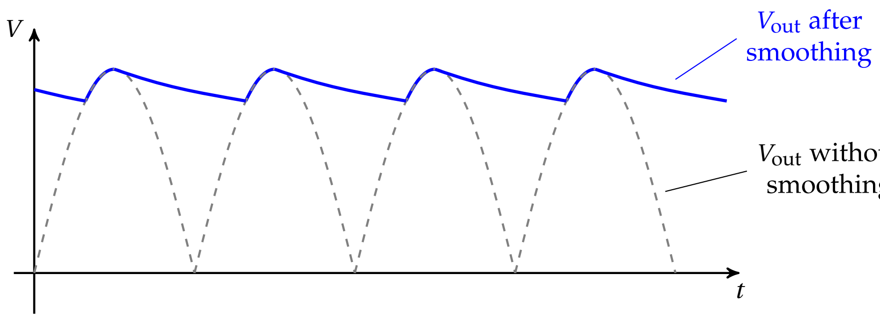



capacitor can store electrical energy as voltage output from rectifier rises

as voltage from rectifier drops, capacitor slowly discharges and feeds energy to the load

output voltage $V_{\text{out}}$ across load will have less ripples over the cycles

less fluctuation in $V_{\text{out}}$ so we say $V_{\text{out}}$ is now smoothed




➢ effect of smoothing is controlled by choice of load resistor $R$ and smoothing capacitor $C$

rate of charging and discharging depend on *time constant* $RC$ (see §7.4)

greater $RC$, smoothing capacitor discharges more slowly, giving less ripples

a larger capacitor for a fixed resistor usually gives better smoothing effect

but if $RC$ is too large, charging would become very slow, capacitor might not be completely charged after each cycle, this could also result into undesirable effects

**Question 10.12** If a second capacitor is added in *series* with the first smoothing capacitor, use sketches to show the changes you would expect for the output voltage.

# CHAPTER 11
# Quantum Physics Basics

In this chapter, we will study the laws of nature at very small scales.

We start by reviewing classical concepts like particles or waves, and then see how they break down when being applied to the microscopic world. For those observations that could not be explained with classical physics, we need *quantum physics*. The theory of quantum physics is very deep and profound, so surely way beyond the scope of our course. What we will study in this A-Level course are merely four tips of an enormous iceberg, namely

- photon theory / photoelectric effect
- matter waves / electron diffraction
- electron energy levels / atomic spectra
- band theory / electrical conductivity of solids

## 11.1 classical theories

in classical physics, both particle models and the wave models have been very useful

though both being successful, the two model are distinct in many aspects

to understand a phenomenon, we take either the particle picture or the wave picture

it seems that there is no way particle models would reconcile with wave models, or is it not?

### 11.1.1 particle models

in particle model, any system is considered to consist of particles governed by *Newtonian mechanics*, physical properties of this system are predicted by studying behaviour of particles

➢ areas of science where particle models are used to interpret and make predictions include:

- *electricity*: electric current formed by motion of charge carriers

- *ideal gas*: pressure caused by collisions of gas molecules
- *solid*: elasticity due to interaction between solid atoms
- *radioactivity*: decay interpreted as the emission of $\alpha$-/$\beta$-particles from the nucleus
- *chemistry*: chemical reactions due to exchange of electrons between atoms and molecules
- …

### 11.1.2 wave models

in the wave picture, energy is transferred via the vibration of medium or force fields

➢ phenomena that can be explained in terms of wave models include:

- *water waves*: variation in the vertical displacement of water surface
- *propagation of sound*: variation in the pressure and density of medium
- *light*: variation of electric and magnetic fields
- …

➢ a key feature that makes waves distinct from particles is that waves can *superpose*

when two or more waves meet, they add up or cancel out to give a resultant wave

this gives rise to the characteristic properties of waves:

- *interference*: waves superpose to form a resultant wave of greater or lower amplitude
- *diffraction*: bending of waves around obstacles, or spreading out of waves through slits

### 11.1.3 history of light

efforts to understand the nature of light can be traced all the way back to ancient Greeks

we are not going to examine the ideas of early thinkers well over 3,000 years ago

let's skip a couple of years and jump to the ideas developed since the Scientific Revolution

➢ particle theory of light (*Issac Newton*, 1671)

key idea: light rays is comprised of a stream of massless particles called *corpuscles*

– explains straight-line propagation
– explains reflection and refraction
– explains colours of light seen in dispersion in prism (corpuscles have different colours)

the problems with the particle model include:

- does not agree with observations on refraction
- cannot predict the interference and diffraction of light

➢ wave theory of light (*Christian Huygens*, 1678)

key idea: light is a wave that transfers energy within a medium known as *aether*

- follows laws of reflection and refraction
- explains colours of light (light have different wavelengths)
- explains interference (Thomas Young's double slit experiment)
- explains diffraction (Poisson spot experiment)

only problem is that aether, the medium in which light lives, was not experimentally found

➢ electromagnetic theory (*James Maxwell*, 1865)

key idea: light is an *electromagnetic wave* [64]

- electric and magnetic fields travel through space in the form of waves at speed of light
- propagation of electromagnetic wave does not require medium, hence no need for aether

all behaviour of light known at that time could be explained with Maxwell's theory

so scientists were convinced that light travelled through space as an electromagnetic wave

## 11.2 photon theory

### 11.2.1 photoelectric effect

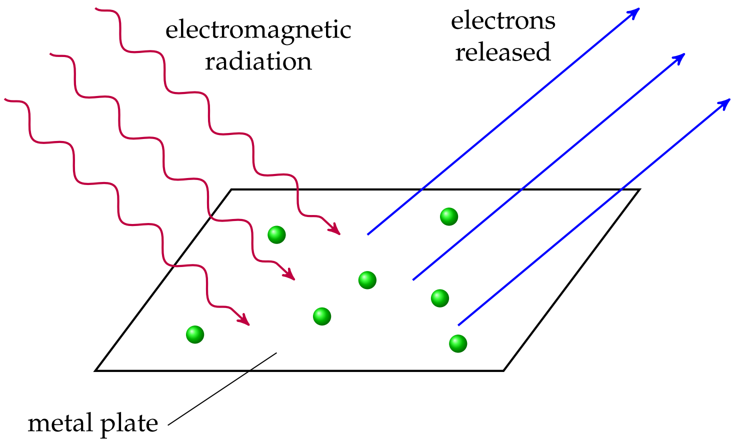
electromagnetic radiation incident upon a metal can cause emission of electrons from the metal surface, this is called **photoelectric effect**, [65] emitted electrons are called **photoelectrons**

[64] *Maxwell's equations* fully describe the behaviour of electric and magnetic fields. Based on the four equations that Maxwell established, he found the vibration of electric and magnetic field can propagate in space as a classical wave. The speed of electromagnetic wave is given by

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = \frac{1}{\sqrt{8.85 \times 10^{-12} \times 4\pi \times 10^{-7}}} = 3.00 \times 10^8 \text{ m s}^{-1},$$

which is the same as speed of light. This gave Maxwell the intuition that light is an electromagnetic wave.

[65] Photoelectric effect was first observed in 1887 by *Heinrich Hertz*, who found electrodes illuminated with ultraviolet light create electric sparks. In school labs, one can shine ultraviolet radiation from a




simply put, conduction electrons in metal gain additional energy by absorbing the incoming radiation, if this energy is sufficient for electrons to overcome the electrostatic attraction from the positive metal ions, they break free from the metal surface

➢ some detailed experimental observations on the photoelectric effect are:
  - there exists a minimum **threshold frequency** $f_0$ for incident radiation
    when $f < f_0$, no electrons released from metal surface
  - emission of electrons is immediate when radiation is incident as long as $f > f_0$
    even low-intensity light is effective
  - increasing intensity has no effect on energies of electrons
  - increasing intensity of incident light causes number of photoelectrons emitted to increase
  - increasing radiation frequency increases electron energies

➢ wave theory of light fails to explain any of these properties, according to wave model:
  - electrons could gradually build up energies by absorbing wave energy over time
    so radiation at any frequency should all work
  - need very intense light to have immediate effect
  - greater intensity mean higher energy, electrons released should have greater K.E.
  - varying frequency of radiation should have no effect on energy of electrons released

photoelectric effect sees the breakdown of the wave model, some new ideas are needed!

---

mercury lamp onto a zinc plate to cause photo-emission.

### 11.2.2 photon theory

photoelectric effect was explained by *Albert Einstein* in 1905 [66]

Einstein's revolutionary idea: when radiation delivers energy to matter, the transfer of energy is not continuous but carried in *discrete* packets, called *photons*

> a **photon** is a packet (*quanta*) of electromagnetic energy

- energy of one photon is given by: $\boxed{E = hf}$, where $h = 6.63 \times 10^{-34}$ J s is the **Planck constant**

  since $E \propto f$, higher/lower radiation frequency means greater/smaller photon energies
- wave equation $c = \lambda f$ relates frequency $f$ of an electromagnetic wave to its wavelength $\lambda$

  so energy of one photon is also given by: $\boxed{E = \frac{hc}{\lambda}}$

  since $E \propto \frac{1}{\lambda}$, longer/shorter wavelength means lower/greater photon energies
- the equation $E = hf$, or $E = \frac{hc}{\lambda}$, relates a particle property with a wave property

  $E$ is the energy of one photon, treated like a single particle

  but $f$ and $\lambda$ are both introduced to describe a wave, not a particle
- intensity of a beam of radiation is: $I = \frac{P}{A}$

  total energy incident on a given area per unit time determines the radiation intensity

  so intensity depends on the product of the number of photons arriving per unit time and energy of each photon: $\boxed{I \propto nhf}$, or more precisely, $I = \frac{nhf}{A}$
- when considering energy of a photon, **electronvolt** is a useful energy unit

  one electronvolt (1 eV) is work needed to make an electron travel through a p.d. of 1 V

  conversion between electronvolt and joule is: $\boxed{1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}}$

**Example 11.1** A laser emits red light of 650 nm at a power rating of 2.0 mW. (a) What is the energy carried by one photon? (b) How many photons are emitted per second?

✍ energy of one photon: $E = \frac{hc}{\lambda} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^{8}}{650 \times 10^{-9}} \approx 3.06 \times 10^{-19}$ J

number of photons: $N = \frac{\text{total energy output from laser}}{\text{energy of one photon}} = \frac{2.0 \times 10^{-3}}{3.06 \times 10^{-19}} \approx 6.54 \times 10^{15}$ □

[66] Albert Einstein was awarded the Nobel Prize in physics in 1921 for 'his discovery of the law of the photoelectric effect'. This discovery led to the quantum revolution of modern physics.

**Question 11.1** Calculate the energy, in eV, of a photon of light of wavelength 440 nm.

**Question 11.2** If the light source in Example 11.1 is replaced by a green laser with the same power output, how does the number of emitted photons per unit time change?

### 11.2.3 photoelectric effect explained

from the viewpoint of photon theory, photoelectric effect can be explained easily

to release an electron from metal, it requires a minimum energy $\Phi$, called the **work function**, for the electron to overcome attraction due to metal ions to escape from metal surface

as radiation shines upon metal, photon energies are absorbed by electrons

since photon energies are *discrete*, or *quantised*, the absorption is sort of all or nothing

if photon energy is greater than $\Phi$, electrons break free from metal

excess energy, if any, would become kinetic energy of the free electron

this is summarised in **Einstein's photoelectric equation**: $\boxed{hf = \Phi + E_{k,\max}}$

➢ note that we are talking about *maximum* K.E. of emitted electrons

electron emitted from the *surface* would have greatest K.E.

for electrons to be released from *below* the surface, they require more energy than work function, so less K.E. than maximum value

➢ at critical condition, incoming photon has just enough energy to release electron

so at threshold frequency $f_0$, photon energy equals work function: $\boxed{hf_0 = \Phi}$

➢ experimental observations mentioned in §11.2.1 can now be understood

- below threshold frequency $f_0$, not enough photon energy available to electron to overcome work function, so no effect for $f < f_0$
- interaction between photon and electron is *one-to-one*, so no time delay
- greater radiation intensity means more photons per unit time, so more electrons released
- greater frequency means higher photon energy, so greater K.E. for emitted electron

**Example 11.2** Given that work function energy of gold is 4.9 eV. Find the longest wavelength of electromagnetic wave that could release electrons from gold.

✍ threshold frequency: $f_0 = \frac{\Phi}{h} = \frac{4.9 \times 1.60 \times 10^{-19}}{6.63 \times 10^{-34}} \approx 1.18 \times 10^{18}$ Hz threshold wavelength: $\lambda_0 = \dfrac{c}{f_0} = \dfrac{3.00 \times 10^8}{1.18 \times 10^{15}} \approx 2.54 \times 10^{-7}$ m (ultraviolet light) □

**Question 11.3** Sodium has a work function of $3.8 \times 10^{-19}$ J. If a light of 500 nm is incident on sodium, determine whether electrons can be emitted from the surface.

**Question 11.4** When electromagnetic radiation of wavelength 1200 nm is incident on a metal surface, the maximum kinetic energy of the electrons released is found to be $5.4 \times 10^{-20}$ J. What is the work function of this metal?

**Question 11.5** When a beam of light of a particular frequency and intensity is shone onto a metal surface, electrons are released. If another beam of light of same intensity but higher frequency is used, what is the effect on the rate of emission of electrons from this surface?

**measurement of the Planck constant and work function energy**

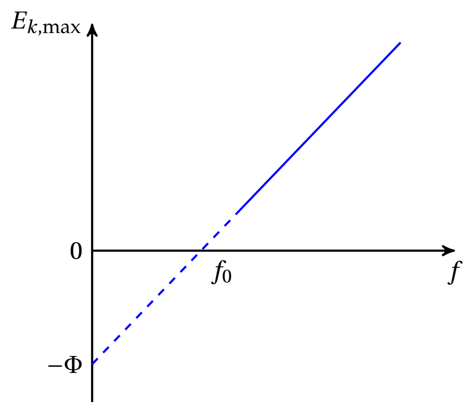
when radiation with different frequencies is incident onto a metal, we measure maximum K.E. of the electrons emitted from the surface, a set of readings $(f, E_{k,\max})$ can be found

note that the photoelectric equation can be rearranged as: $E_{k,\max} = hf - \Phi$

if we plot a graph of $E_{k,\max}$ against $f$, data points shall fall on a straight line

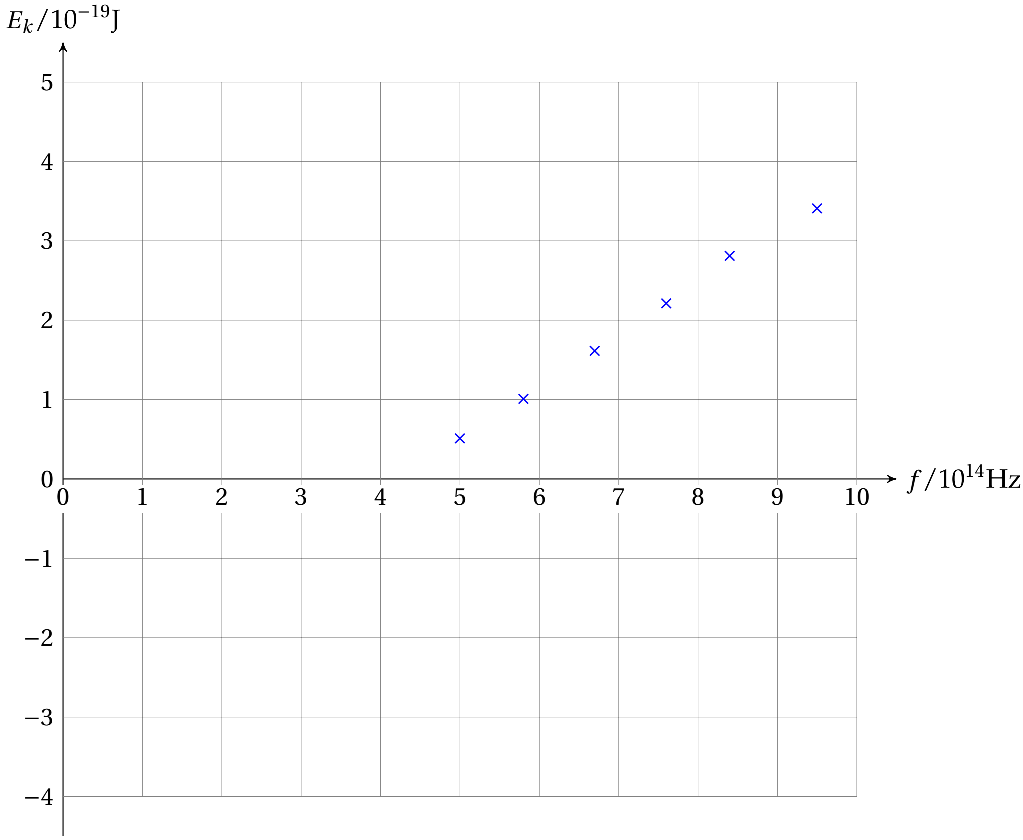
information about the Planck constant $h$, threshold frequency $f_0$, work function $\Phi$ can all be computed with the best-fit line

- gradient $= h$
- $y$-intercept $= -\Phi$
- $x$-intercept $= \dfrac{\Phi}{h} = f_0$




**Question 11.6** If a different metal with a greater work function energy is used, describe the change for the line that shows the variation with $f$ of $E_{k,\max}$ for this metal.

**Question 11.7** Electromagnetic adiation is incident upon a metal plate. The graph shows how maximum kinetic energy $E_k$ of emitted electrons varies with frequency $f$ of the radiation. Use the graph to find (a) the threshold frequency, (b) a value of Planck constant.




### 11.2.4 photon momentum

in photoelectric effect, photons can knock electrons out of a metal, this suggests that photons could have *momentum*, even though they do not have mass

earliest experimental evidence of photon momentum was came from *Arthur Compton* in 1923, who studied the scattering of X-ray photons by electrons in substances[67]

it can be shown[68] that photon momentum is given by: $\boxed{p = \dfrac{h}{\lambda}}$

[67]Arthur Compton was awarded the Nobel Physics Prize in 1929 for the discovery of this scattering effect, now known as *Compton scattering*.

[68]This derives from Einstein's theory of *special relativity*, which states that energy and momentum are related by the equation: $E^2 = m_0^2c^4 + p^2c^2$. Photons have zero rest mass, i.e., $m_0 = 0$. This relativistic relation becomes $E = pc$ for photons. Now recall photon energy is given by $E = hf$. Rearrange the terms, we can show: $p = \dfrac{hf}{c} = \dfrac{h}{\lambda}$.

➢ photon momentum a *relativistic* momentum, as photons move at speed of light

definition for classical momentum $p = mv$ does not apply for photons

➢ due to exchange of momentum, electromagnetic wave can exert *radiation pressure*

forces generated by radiation pressure are negligible under everyday circumstances, but they could have noticeable effects on spacecraft in outer space and comet tails

**Example 11.3** A beam of light has wavelength 600 nm, cross-sectional area 0.16 cm$^2$ and power 5.0 mW. The beam is normally incident onto a surface and is completely absorbed. Calculate, for a time of 1.0 s, (a) the number of photons incident onto the surface, (b) the change of total momentum of the photons, (c) the light pressure on the surface.

✍ energy of one photon: $E = \dfrac{hc}{\lambda} = \dfrac{6.63 \times 10^{-34} \times 3.0 \times 10^{8}}{600 \times 10^{-9}} \approx 3.32 \times 10^{-19}$ J 1.5083

number of photons arriving in 1.0 s: $N = \dfrac{5.0 \times 10^{-3} \times 1.0}{3.32 \times 10^{-19}} \approx 1.51 \times 10^{16}$

momentum for one photon: $p = \dfrac{h}{\lambda} = \dfrac{6.63 \times 10^{-34}}{600 \times 10^{-9}} \approx 1.11 \times 10^{-27}$ kg m s$^{-1}$

change of total momentum: $\Delta P = Np = 1.51 \times 10^{16} \times 1.11 \times 10^{-27} \approx 1.67 \times 10^{-11}$ kg m s$^{-1}$

average force due to these photons: $F = \dfrac{\Delta P}{\Delta t} = \dfrac{1.67 \times 10^{-11}}{1.0} \approx 1.67 \times 10^{-11}$ N

light pressure on surface: $p = \dfrac{F}{A} = \dfrac{1.67 \times 10^{-11}}{0.16 \times 10^{-4}} \approx 1.04 \times 10^{-6}$ Pa □

**Question 11.8** A laser of power $P$ is incident normally on a spot of area $A$, show that the pressure caused by the beam can be given by: $p = \dfrac{P}{cA}$.

**Question 11.9** When an electron and a positron meet together, they will annihilate and produce two $\gamma$-photons: ${}^{0}_{-1}e + {}^{0}_{+1}e \longrightarrow \gamma + \gamma$. Assume the electron and the positron have negligible kinetic energy before the interaction, explain why the two photons produced must move off in opposite directions with equal energies.

## 11.3 wave-particle duality

### 11.3.1 matter waves

inspired by photon theory, which shows electromagnetic waves have a particulate nature, *Louis de Broglie* suggested in his 1924 PhD thesis that all matter has a wave-like nature [69]

wave characteristic of a particle can be represented by a wavelength

> **de Broglie wavelength** of a matter particle is given by: $\boxed{\lambda = \dfrac{h}{p}}$, where $p = mv$ is the particle's momentum

**Example 11.4** What is the wavelength of a human of 70 kg walking at around 2.0 m s$^{-1}$?

✍ $\lambda = \dfrac{h}{mv} = \dfrac{6.63 \times 10^{-34}}{75 \times 2} \approx 4.7 \times 10^{-36}$ m

this wavelength is too small compared with any obstacle we encounter in everyday lives

so human bodies do not exhibit noticeable wave behaviour □

**Example 11.5** An electron is accelerated from rest through a potential difference of 50 V. What is the de Broglie wavelength of this electron?

K.E. of electron equals change in electric P.E., so

$$\frac{1}{2}mv^2 = qV \quad \Rightarrow \quad v = \sqrt{\frac{2qV}{m}} = \sqrt{\frac{2 \times 1.60 \times 10^{-19} \times 50}{9.11 \times 10^{-31}}} \approx 4.19 \times 10^6 \text{ m s}^{-1}$$

wavelength of electron: $\lambda = \dfrac{h}{mv} = \dfrac{6.63 \times 10^{-34}}{9.11 \times 10^{-31} \times 4.19 \times 10^6} \approx 1.74 \times 10^{-10}$ m

this wavelength is comparable to scale of atomic spacing (also around $10^{-10}$ m)

so these electron can be *diffracted* by solid crystals □

**Question 11.10** An $\alpha$-particle is moving with a kinetic energy of $2.4 \times 10^{-15}$ J. Find its speed, and hence find its de Broglie wavelength.

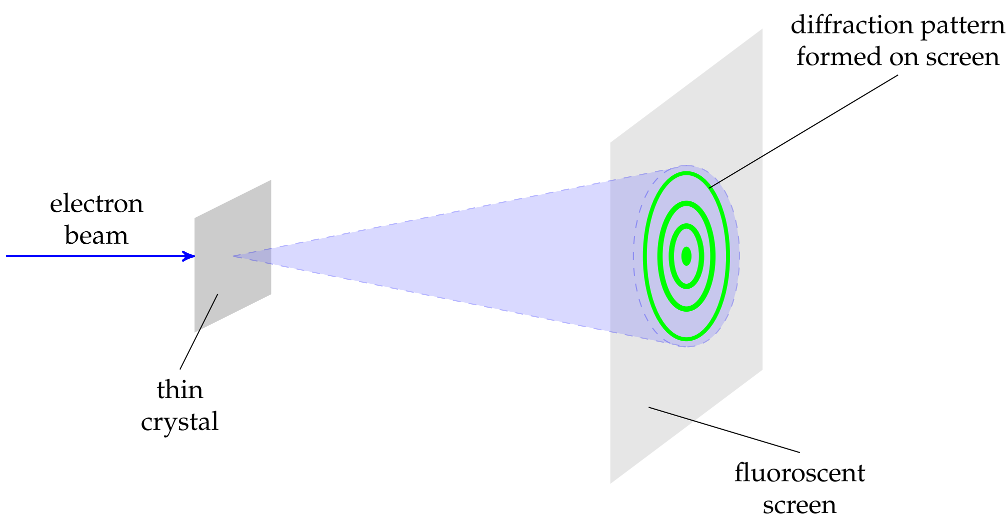
**Question 11.11** If a proton and an electron are accelerated through the same voltage, (a) how would their energy compare? (b) how would their wavelength compare?

[69] Louis de Broglie was awarded the 1929 Nobel Physics Prize 'for his discovery of the wave nature of electrons'. *Schröedinger's equation* and *Bohr's atomic model* was heavily influenced by ideas of de Broglie.

### 11.3.2 electron diffraction

wave property of electron was confirmed by *Clinton Davisson* and *George Thomson* in 1927

they showed experimentally electrons could be diffracted by metal crystals [70]




electron diffraction experiment

➢ electron diffraction experiment shows that particles do have wave-like properties

if electrons behaved like particles, we would see a round spot of *uniform* distribution

but the actual pattern formed on *fluorescent screen* is a set of *concentric rings*

this is a typical diffraction pattern, hence proves wave properties of electrons

➢ each metal has a different lattice structure, so each produces a different pattern

this allows investigation of structure of matter (explore arrangements of atoms, structures of complex molecules, structure of atomic nuclei, etc.) using electron diffraction

➢ since wavelength of an electron is much shorter than visible light

this allows *electron microscope* to have much higher resolving power than optical microscopes

**Question 11.12** If a higher p.d. is applied to accelerate the electron beam in the electron diffraction experiment, how would the pattern change?

[70] The Nobel Prize in Physics 1937 was awarded jointly to Clinton Davisson and George Thomson 'for their experimental discovery of the diffraction of electrons by crystals'

### 11.3.3 wave-particle duality

we have seen both particle-like and wave-like behaviour in light and electrons

| | electromagnetic radiation | electron |
|---|---|---|
| particle-like behaviour | photoelectric effect, radiation pressure | deflection in electric/magnetic fields, decay, scattering |
| wave-like behaviour | interference, diffraction, Doppler effect, reflection, refraction | electron diffraction |

when light/electron move through space, they behave like a wave

when light/electron interact with each other, they behave like particles

they show a two-sided nature, or a *dual* nature of being both a wave and a particle, described by either particle model or wave model under different circumstances

in fact, all matter (protons, neutrons, atoms, cells, basketballs, human body, earth, etc.) has this universal dual nature, called **wave-particle duality**

wave-particle duality addresses breakdown of classical concepts like particle or wave

to fully describe the behaviour of microscopic objects, we need *quantum mechanics* [71]

## 11.4 quantisation of electron energy levels

in a simple atomic model, electrons move around the nucleus in circular orbits

but now we understand electrons have wave properties, as an electron moves in its orbit as a wave, it can superpose with itself

only orbits in which the electron can superpose constructively with itself are preferable

so electrons are only allowed to move in certain orbits in an atom

this means can only take certain values of energy, called *energy levels*

[71]Being a central concept of quantum theory, wave-particle duality is deeply embedded into the foundations of quantum physics. In non-relativistic quantum mechanics, all information about a particle is encoded in its *wave function*, whose evolution with time is described by the famous *Schrödinger's equation*.

### 11.4.1 hydrogen atom (∗)

theoretical explanation for electron energy levels was developed in 1913 by Danish physicist *Niels Bohr* in his theory of hydrogen atom [72]

let's look at the hydrogen atom – the simplest possible atom in nature

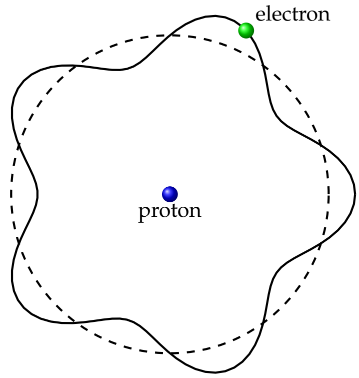
electrostatic attraction by proton provides centripetal force for electron to move in circles

$$\frac{e^2}{4\pi\epsilon_0 r^2} = \frac{m_e v^2}{r} \quad \Rightarrow \quad v^2 = \frac{e^2}{4\pi\epsilon_0 m_e r}$$

for electron to *constructively* superpose with itself, the perimeter of its orbit must be an integer multiple of its de Broglie wavelength

$$2\pi r = n\lambda \quad \Rightarrow \quad 2\pi r = \frac{nh}{m_e v} \quad \Rightarrow \quad v = \frac{nh}{2\pi m_e r}$$

where $n$ is positive integer $1, 2, 3, \cdots$




the hydrogen atom

compare the two equations, we can eliminate $v$

$$\frac{e^2}{4\pi\epsilon_0 m_e r} = \frac{n^2 h^2}{4\pi^2 m_e^2 r^2}$$

we hence find radius of allowed orbits satisfies:

$$\boxed{r_n = n^2 a_0} \quad \text{where } a_0 = \frac{h^2\epsilon_0}{\pi m_e e^2} \approx 5.29 \times 10^{-11}\ \text{m}$$

so we see all allowed orbits must have a radius equal to integer multiple of $a_0$

total energy possessed by the electron consist of K.E. and electric P.E.:

$$E = E_k + E_p = \frac{1}{2} m_e v^2 - \frac{e^2}{4\pi\epsilon_0 r} = \frac{1}{2} m_e \frac{e^2}{4\pi\epsilon_0 m_e r} - \frac{e^2}{4\pi\epsilon_0 r} \quad \Rightarrow \quad E = -\frac{e^2}{8\pi\epsilon_0 r}$$

substitute the allowed orbital radius, we find an expression for electron energy:

$$\boxed{E_n = -\frac{R_E}{n^2}} \quad \text{where } R_E = \frac{m_e e^4}{8h^2\epsilon_0^2} \approx 2.18 \times 10^{-18}\ \text{J} \approx 13.6\ \text{eV}$$

this shows electron can only have specific energies in the hydrogen atom

[72] Niels Bohr was surely one of the greatest physicists of the 20th century. He made foundational contributions to understanding atomic structure and quantum mechanics, for which he received the Nobel Physics Prize in 1922. He was also the founder of the Institute of Theoretical Physics at the University of Copenhagen, which soon became the centre of pioneering researches on quantum theory in the world.

### 11.4.2 electron energy levels

it can be shown that for any atom, electrons can only have certain fixed values of energy

we say energy of electrons in an atom is *discrete*, or **quantised**

these allowed values of energies are called **energy levels**[73]

➢ *quantisation* means an electron can have only specific values of energy in an atom

no intermediate value between levels is allowed

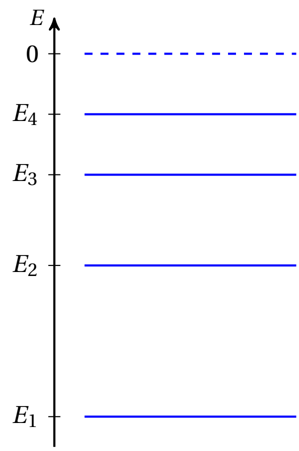
➢ the lowest energy level is called the **ground state**

any level higher than the ground state is called an **excited state**

in our illustration, $E_1$ is the ground state, while $E_2$, $E_3$, $E_4$ are all excited states

➢ electron energy levels are all *negative*

to pull an electron away from the nucleus, work must be done to overcome electrostatic attraction




electron energy levels of some atom

an electron at infinity would have greatest P.E., which is defined to be zero

so orbiting electrons have energies less than zero

an electron that has zero energy would become a *free* electron

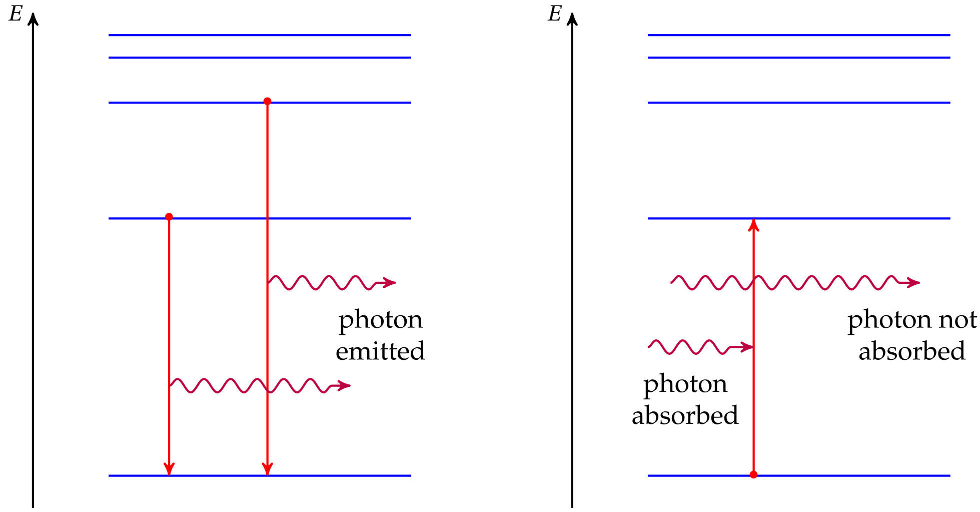
➢ electrons can jump, or *transit*, between energy levels

electron transition is associated with emission or absorption of photon with the right energy

### 11.4.3 atomic spectrum

passing light through a *prism* or a *diffraction grating*, different wavelengths are separated, a pheonomenon known as **dispersion**, this produces a **spectrum** which shows distribution of energy emitted from the source in order of wavelengths

[73] If you study chemistry, you shall recall this idea of discrete energy levels is related to the concepts of energy shells and sub-shells of atoms.




emission and absorption of photons as a result of electron transitions

for example, white light consists of a range of wavelengths from around 400 nm to 700 nm, so it gives a *continuous spectrum* with rainbow of colours

continuous spectrum of white light

electron transition between different levels causes emission or absorption of photons

this leads to the emission spectrum and absorption spectrum of atoms

**emission spectrum**

*hot gases* of an element can release photons, giving an **emission spectrum**

this happens when an electron transits from a high energy level to a lower level

energy of the photon emitted is equal to energy difference between the two levels

since electron energy levels are discrete, only specific changes of energies are possible, so only photons with specific energies can be emitted

this means photons emitted only have specific wavelengths, or specific frequencies

so a collection of sharp and bright lines are seen in emission spectrum

emission spectrum of the hydrogen atom

- emission spectrum is a *discrete* spectrum, also called a *line* spectrum
- atoms of each element has a unique electron energy level structure

  so each element produces a unique emission spectrum, leading to quite a few applications:

  - explains the colour of flames when a particular chemical element is present
  - allows the identification of elements in an unknown substance in chemical analysis
  - explains varied colours of electric signs lighted by gas-discharge tubes

**absorption spectrum**

pass white light through a *cool gas*, photons with the right energies can be absorbed, giving rise to an **absorption line spectrum**

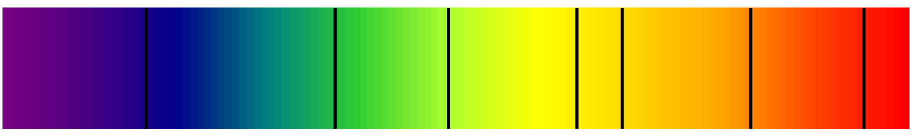
by absorbing photon, electron can transit from a low energy level to a higher level

energy of the photon absorbed must equal energy difference between the two levels

since electron energy levels are discrete, so only photons with specific energies can be absorbed, while other photons are unaffected as they pass through the gas

wavelengths of these absorbed photons will be missing in the emergent spectrum

so we would observe a set of *dark lines* appearing in a background of continuous spectrum



absorption spectrum for some element

- absorption spectrum is also a *discrete* spectrum, or a *line* spectrum
- as photons with the right energies are absorbed, they can be re-emitted through *de-excitation*

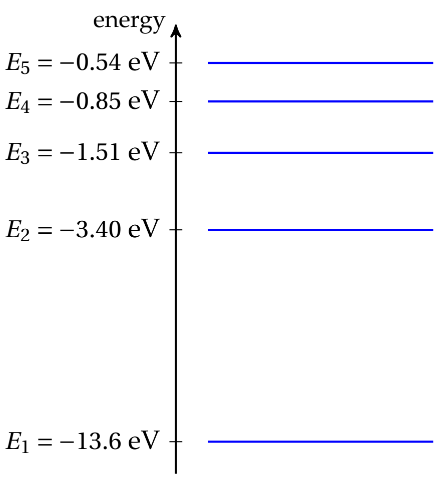
  but these photons are re-emitted in *random* directions

  so they would still appear dark compared with those unaffected wavelengths

➢ absorption spectrum is widely used in many areas

- determine the composition of a particular substance in analytical chemistry
- determine chemical compositions of stars in astronomical spectroscopy
- explain colour of chemicals in terms of the complementary colour of photons absorbed

**Example 11.6** Some electron energy levels of the hydrogen atom is shown. Find the longest and the shortest wavelength produced by electron transitions between the energy levels given.




energy levels of the hydrogen atom

✍ energy of photon emitted equals change of electron energy level: $\frac{hc}{\lambda} = \Delta E$, or $\lambda = \frac{hc}{\Delta E}$

transition with least/greatest energy change gives rise to longest/shortest wavelength, so

$$E_5 \to E_4 \Rightarrow \lambda_{\max} = \frac{hc}{E_5 - E_4} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^8}{(0.85 - 0.54) \times 1.60 \times 10^{-19}} \approx 4.0 \times 10^{-6} \text{ m} \quad \text{(infra-red)}$$

$$E_5 \to E_1 \Rightarrow \lambda_{\min} = \frac{hc}{E_5 - E_1} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^8}{(13.6 - 0.54) \times 1.60 \times 10^{-19}} \approx 9.5 \times 10^{-8} \text{ m} \quad \text{(ultraviolet)} \qquad \square$$

**Example 11.7** The emission spectrum of the hydrogen atom consists of a number of wavelengths in the visible spectrum. Given that the visible spectrum consists of light of wavelengths within the range from 380 nm to 740 nm, also use data in Example 11.6, find the energies of visible photons that can be produced by transitions between the energy levels shown.

✍ let's first find the range of energies, in eV, for visible photons

$$E_{\text{max}} = \frac{hc}{\lambda_{\text{min}}} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^{8}}{380 \times 10^{-9}} \approx 5.23 \times 10^{-19}\ \text{J} = \frac{5.23 \times 10^{-19}}{1.60 \times 10^{-19}}\ \text{eV} \approx 3.27\ \text{eV}$$

$$E_{\text{min}} = \frac{hc}{\lambda_{\text{max}}} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^{8}}{740 \times 10^{-9}} \approx 2.69 \times 10^{-19}\ \text{J} = \frac{2.69 \times 10^{-19}}{1.60 \times 10^{-19}}\ \text{eV} \approx 1.68\ \text{eV}$$

we shall find the combinations of electron energy levels such that their difference fall within the range between 1.68 eV and 3.27 eV, by trial and error, we find three possible combinations:

(1) $E_3 \to E_2 \Rightarrow 3.40 - 1.51 = 1.89$ eV

(2) $E_4 \to E_2 \Rightarrow 3.40 - 0.85 = 2.55$ eV

(3) $E_5 \to E_2 \Rightarrow 3.40 - 0.85 = 2.85$ eV

these emisssion lines are called $H_\alpha$ (656 nm), $H_\beta$ (486 nm) and $H_\gamma$ (435 nm) lines[74]

they are three of the four the hydrogen emission lines that are visible to human eyes[75] □

**Example 11.8** A white light is incident on a cloud of cool hydrogen gas. In the emergent spectrum, a dark line is observed at a wavelength of 435 nm. Determine the energy change that gives rise to this dark line.

✍ photon absorbed: $E = \frac{hc}{\lambda} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^{8}}{435 \times 10^{-9}} \approx 4.57 \times 10^{-19}\ \text{J} = \frac{4.57 \times 10^{-19}}{1.60 \times 10^{-19}}\ \text{eV} \approx 2.86\ \text{eV}$

note that $3.40 - 0.54 = 2.86$ eV, so this dark line is due to the electron transition: $E_2 \to E_5$ □

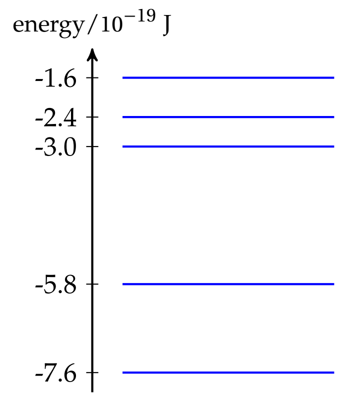
**Question 11.13** If we only consider the electron energy levels given in Example 11.6, how many different wavelengths can be detected in the emission spectrum of the hydrogen atom ?

**Question 11.14** Three lines are observed at wavelengths 486 nm , 656 nm, and 1880 nm in the emission spectrum of hydrogen atoms. (a) Calculate the photon energies for these wavelengths. (b) Draw a diagram with *three* labelled energy levels, and show the energy changes for the three wavelengths produced with arrows in your diagram.

[74]The $H_\alpha$ line is the brightest hydrogen line in the visible range. It plays an important role in astronomy, as it can be used to study a star's surface temperature, the velocity of a distant stellar object, etc.

[75]These lines belong to a family of the spectral lines of the hydrogen atom, known as the *Balmer lines*, named after Johann Balmer. Balmer discovered an empirical equation to predict the series in 1885, but the reason why the equation worked was eventually clarified by Neils Bohr with the atomic model which now bears his name. We now understand the Balmer lines correspond to emissions of photons by electrons jumping to the second level from higher energy states.

**Question 11.15** The relative cool atmosphere of the sun could give rise to dark lines in the spectrum of sunlight. One particular dark spectral line has a wavelength of 590 nm. By reference to the energy levels of the helium atom, suggest how this dark line provides evidence of the presence of helium in sun's atmosphere. You may draw an arrow to show the possible electron transition that gives rise to this dark line.




energy levels of the helium atom

## 11.5 band theory (∗)

### 11.5.1 energy bands

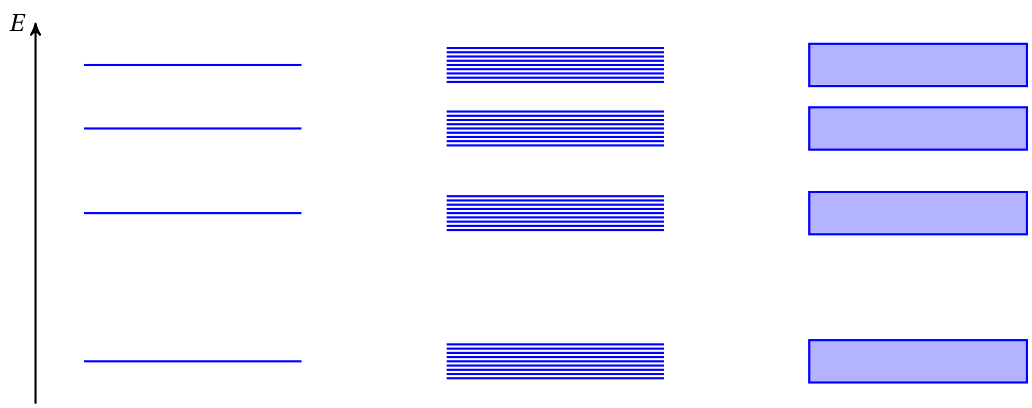
in an isolated atom, electrons have discrete energy levels

in solids, interaction between neighbouring atoms causes change of energies

original energy levels split into a band with many sub-levels

number of sub-levels in each band equals number of atoms in solid in general

due to very large number of atoms, sub-levels (seem to) form *continuous* **energy bands**




dicrete energy levels in isolated atom

split energy sub-levels due to interactions

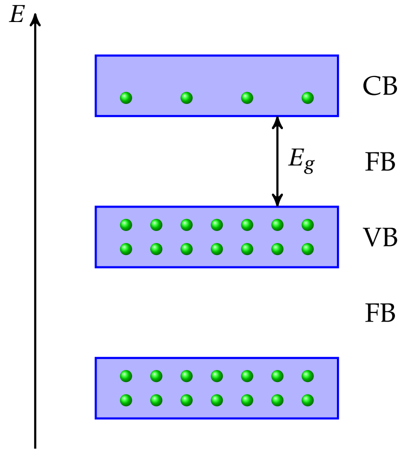
energy bands formed in many-particle system in solids, electrons fill up from lowest energy bands available (see figure for a rough idea)

**forbidden bands** (FB) represent energies that electrons are not allowed to take

highest fully filled band is called **valence band** (VB)

next partially filled or empty band is called **conduction band** (CB)

separation between valence band and conduction band defines **band gap** ($E_g$) of material




### 11.5.2 band theory & electrical conductivity

electrical conductivity of material can be explained by its band structure

low-energy bands below VB are usually fully filled at all times, all states are occupied, electrons cannot move freely to form electric currents, so we are not interested in low-energy bands

conductivity of material will depend on whether there are free *charge carriers* in CB and VB

**metallic conductors**

for typical metals, CB *overlaps* with VB, i.e., there is no band gap

there exist vacant states for conduction electrons to occupy at ease

this means conduction electrons are free to move around to form currents

as temperature rises, there is no significant change in number of free electrons

but *lattice vibration* of atoms increases, electrons become more likely to collide with vibrating atoms, so resistance of metal increases with temperature

➢ metals are *opaque* to visible light or other low-frequency radiation

width of CB for a typical metal is about 2 ~ 3 eV

low-energy photons can be absorbed by conduction electrons

visible light, infra-red, microwave therefore cannot pass through metal

➢ metals are *transparent* to high-frequency radiation such as X-rays if X-ray photons were absorbed, electrons would take energies in forbidden band

but this is not allowed, so high-frequency photons penetrate through metal

**insulators**

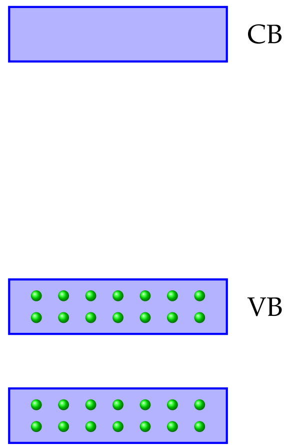
examples of insulators include glass, diamond, etc.

in normal conditions, CB of insulator is empty, energy bands up to VB are fully filled, so no movement of electron is possible, material has poor conductivity

insulators also have huge band gap[76][77], so thermal excitations cannot easily make VB electrons jump into CB

insulators remain poor conductors as temperature rises

**Question 11.16** Diamond has a band gap $E_g \approx 6.0$ eV. By reference to its electronic band structure, explain why diamond appears *transparent* to visible light?



band structure of insulator

**intrinsic semiconductors**

examples of an intrinsic semiconductor are silicon and germanium

band structure of an intrinsic semiconductor is similar to that of insulators

at low temperature, no free electron from the empty CB and completely occupied VB, so semi-conductors are not conductive in normal conditions

however, semiconductors have much narrower band gaps

as temperature rises, VB electrons gain thermal energy to cross band gap and enter CB

the electrons jumped into CB surely can move around to form electric currents

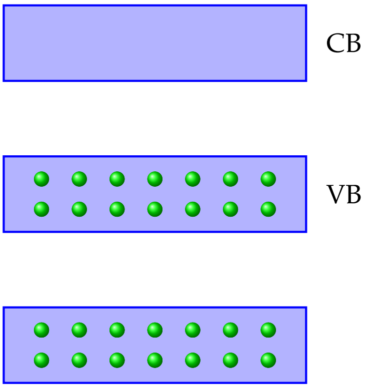
at the same time, VB is no longer completely filled, *holes* are formed

[76]To be a little more precise, materials with a band gap $E_g \approx 5$ eV or greater are regarded as insulators.

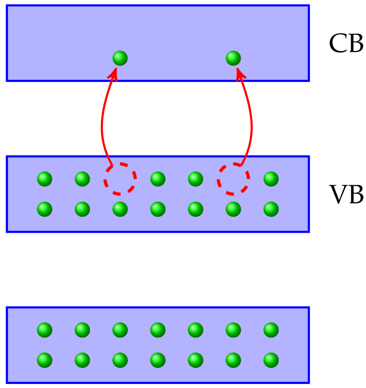
[77]There exist a class of insulators called *Mott insulators* which do not have large band gaps. Their poor conductivity at low temperatures is due to electron-electron interactions, which are not considered under conventional band theories.

a hole is a site where an electron is missing, neighbouring electrons can move to fill the hole and leave a new hole, but this can be thought of as the hole moving about

there are now more *charge carriers* (negatively-charge electrons and positively-charged holes) available, so semiconductor has better conductivity, resistance of material would decrease




band structure of an intrinsic semiconductor at low temperature




at higher temperature VB electrons cross band gap and enter CB

➢ *lattice vibration* could also affect resistance of semiconductors

as temperature rises, vibration of atoms increases, charge carrier are more likely to get scattered, causing a decrease in conductivity

but the effect of more charge carriers is far greater than effect due to lattice vibration

so resistance of semiconductor decreases at higher temperature

**Example 11.9** Silicon has a band gap $E_g \approx 1.1$ eV. State and explain whether it becomes conducting when exposed to red light of 600 nm.

✍ energy of radiation: $E = \frac{hc}{\lambda} = \frac{6.63 \times 10^{-34} \times 3.0 \times 10^{8}}{600 \times 10^{-9}} \approx 3.32 \times 10^{-19}\ \mathrm{J} \approx 2.1\ \mathrm{eV} > E_g$

this energy is sufficient to make valence electrons in silicon to cross band gap

silicon now possesses conducting electrons and holes therefore becomes conducting □

**Question 11.17** Use band theory to explain why an LDR (light-dependent resistor) has a very high resistance in the dark, but its resistance drops dramatically when being exposed to light.

# CHAPTER 12

# Nuclear Physics

## 12.1 nuclear energy

### 12.1.1 mass-energy equivalence

mass-energy equivalence principle states that mass and energy are equivalent to one another

this idea is represented mathematically by **Einstein's mass-energy relation**: $\boxed{E = mc^2}$ [78]

$m$ is **rest mass** of the object, $c = 3.00 \times 10^8$ m s$^{-1}$ is speed of light in vacuum

➢ mass-energy equivalence implies mass can be converted into pure energy

conversely, mass can be created out of energy

➢ $\Delta E = \Delta mc^2$ applies to *all* energy changes

if energy is supplied to a system, then mass of this system increases

for example, an object will have a greater mass when it is set to motion or is heated

mass is not a converved quantity, it is *mass-energy* that is conserved in all processes

➢ if an object is stationary, its mass is called the *rest mass* $m_0$

mass-energy equivalence implies an object at rest still has an intrinsic *rest energy*: $E_0 = m_0 c^2$

➢ energy transformations such as chemical reactions can cause a system to lose some mass content, but this change in mass is usually negligible

[78]This might be the most famous equation in all physical sciences. When Albert Einstein wrote down the theory of *special relativity* in 1905, he found the mass $m$ of an object with rest mass $m_0$ is related to the speed $v$ at which it is moving by: $m = m_0\left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}}$. Total energy is given by $E = mc^2 = m_0 c^2\left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}}$. At low speeds $v \ll c$, $E \approx m_0 c^2\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = m_0 c^2 + \frac{1}{2}m_0 v^2$. The second piece clearly gives the kinetic energy term, so the first piece can be thought as the rest energy stored within the mass. This is how the idea of mass-energy relation came to the great mind of Einstein.

**Example 12.1** When an antiproton (antiparticle of a proton) collides with a proton, they are annihilated and two photons of equal energy are formed. (a) What is the energy of each photon? (b) What is the photon frequency?

✍ rest energy of proton becomes photon energy:

$$E_\gamma = m_p c^2 = 1.67 \times 10^{-27} \times (3.0 \times 10^8)^2 \approx 1.50 \times 10^{-10}\ \text{J}$$

photon frequency: $f = \dfrac{E_\gamma}{h} = \dfrac{1.50 \times 10^{-10}}{6.63 \times 10^{-34}} \approx 2.27 \times 10^{23}$ Hz (high-frequency $\gamma$-photon) □

**Question 12.1** The combustion of one mole of solid carbon to form carbon dioxide ($CO_2$) at standard condition is 394 kJ. Find the change in mass for this amount of energy, and hence compare this mass change with the mass of carbon before combustion.

**Question 12.2** Given that the specific heat capacity of copper is 380 J kg$^{-1}$ K$^{-1}$. When a copper block is heated from 300 K to 1000 K, what is the additional mass as a fraction of its rest mass?

### 12.1.2 binding energy

*nucleons* (protons and neutrons) in a nucleus bind together through the *strong nuclear force*

to pull nucleons apart, work must be done to overcome the attraction

so free nucleons at infinity have greater potential energy than a single nucleus

since energy is equivalent to mass, free nucleons would appear more massive than when they are held together in a nucleus

this statement is supported by experimental data

useful notions in nuclear physics related to this idea can now be introduced

> difference between mass of a nucleus and total mass of its constituent nucleons when separated to infinity is called the **mass defect**

> energy needed to separate the nucleons in a nucleus to infinity, or equivalently, energy released when free individual nucleons combine to form a nucleus, is called the **nuclear binding energy** ($E_B$)

➢ for a nucleus $Z$ with proton number $Z$ and nucleon number $A$ (represented by ${}^A_Z X$)

its mass defect is given by: $\Delta m = Zm_\text{p} + (A-Z)m_\text{n} - m_X$

where $m_\text{p} = 1.673 \times 10^{-27}$ kg is mass of proton, $m_\text{n} = 1.675 \times 10^{-27}$ kg is mass of neutron

➢ by definition, binding energy is equivalent to mass defect: $\boxed{E_B = \Delta mc^2}$

➢ nucleons in a nucleus have *negative* P.E., while free nucleons have zero P.E.

$E_B$ is the energy required to fill this gap in order to pull nucleons apart

i.e., $E_B$ of a nucleus equals *loss* of potential energy during its formation

➢ nuclear mass is often measured in *unified atomic mass* u, where $\boxed{1\text{ u} = 1.66 \times 10^{-27}\text{ kg}}$

nuclide ${}^A_Z X$ has a nuclear mass of about $A$u

➢ binding energy is often measured in MeV, where $\boxed{1\text{ MeV} = 10^6\text{ eV} = 1.60 \times 10^{-13}\text{ J}}$

**Example 12.2** An iron-56 nucleus (${}^{56}_{26}$Fe) has a mass of $9.288 \times 10^{-26}$ kg. Calculate the nuclear binding energy per nucleon, in MeV, for ${}^{56}_{26}$Fe.

✍ mass defect: $\Delta m = 26m_\text{p} + (56-26)m_\text{n} - m_\text{Fe} = 26 \times 1.673 \times 10^{-27} + 30 \times 1.675 \times 10^{-27} - 9.288 \times 10^{-26}$

so we find $\Delta m = 8.68 \times 10^{-28}$ kg

binding energy: $E_B = \Delta mc^2 = 8.68 \times 10^{-28} \times (3.00 \times 10^8)^2 \approx 7.812 \times 10^{-11}$ J

binding energy per nucleon: $\epsilon_b = \dfrac{E_b}{A} = \dfrac{7.812 \times 10^{-11}}{56} \approx 1.395 \times 10^{-12}$ J

convert into MeV: $\epsilon_b = \dfrac{1.395 \times 10^{-12}}{1.60 \times 10^{-13}}$ MeV $\approx 8.72$ MeV □

**Question 12.3** Show that the energy equivalent of 1.0 u is 934 MeV.

**Question 12.4** Given that mass of proton is 1.007 u, mass of neutron is 1.009 u, and mass of uranium-235 nucleus is 234.992 u. Find the binding energy per nucleon of nuclide ${}^{235}_{92}$U.

**Question 12.5** What is the binding energy for the hydrogen nucleus ${}^1_1$H?

**Question 12.6** Why is rest mass of proton slightly larger than the unified atomic mass unit?

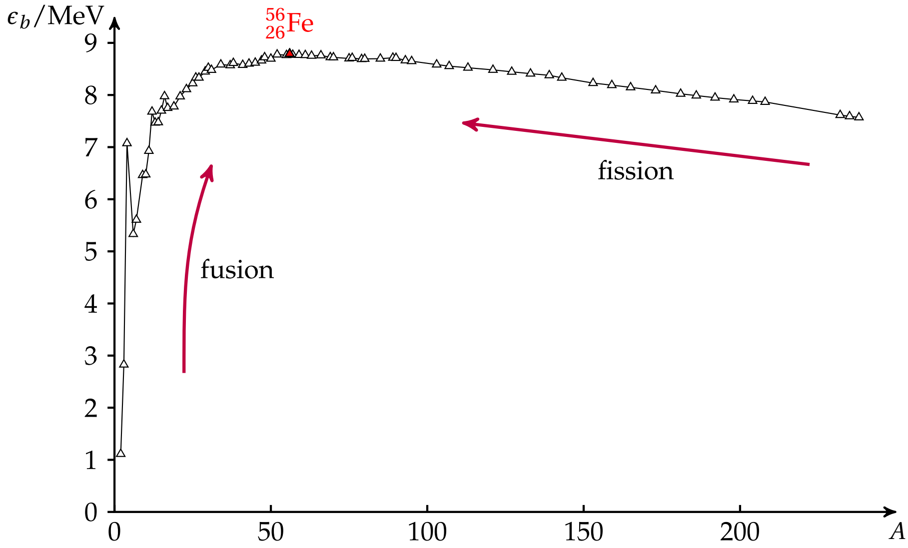
### 12.1.3 nuclear stability

binding energy per nucleon $\epsilon_b$ is closely related to nuclear stability

binding energy per nucleon gives average energy needed to remove a nucleon from nucleus

higher $\epsilon_b$ means more difficult to pull nucleons away, so nucleus has higher stability

a graph of $\epsilon_b$ against nucleon number $A$ can be plotted based on experimental data




binding energy per nucleon $\epsilon_b$ against nucleon number $A$

- $^{56}_{26}$Fe has greatest value of $\epsilon_b$ for all nucleus $\Rightarrow$ $^{56}_{26}$Fe is the most stable nuclide in nature
- any physical system would evolve such that it lowers total energy

  if there could be an increase in total binding energy, nuclear reactions could occur

#### 12.1.4 fusion

for nuclei smaller than $^{56}_{26}$Fe, they could combine together and release energy

> **fusion** is the process where two light nuclei join together to form a heavier nucleus

- fusion occurs only at very *high temperatures*

  nuclei are all positively charged, for two nuclei to come close enough to fuse together, high initial K.E. is needed for them to overcome their *electrostatic repusion*
- vast amount of energy of the sun and other *stars* are created through fusion

  extreme high temperature and high density at core of stars make fusion possible [79]

[79] Even for the simplest fusion reaction, fusing hydrogen into helium, at least $1.5 \times 10^7$ K is required. This is basically how the sun shines and nurtures lives on the earth, and also many other stars in the universe power energies. For a sufficiently massive star, when it burns out of its fuel, its core collapses and becomes hot enough to start to fuse heavier elements. If a star is not able to fuse again, nuclear

### 12.1.5 fission

for a nucleus larger than $^{56}_{26}$Fe, it can break into two or more parts and release energy

> **fission** is the process where a massive nucleus splits into two smaller nuclei of about the same size (and is usually associated with release of several neutrons)

➢ note that fission requires the two product nuclei are of simliar size

$\alpha$-decay (emission of helium necleus from an unstable nucleus) is not a fission process

➢ fission reactions can occur *spontaneously*, i.e. nuclei can undergo fission by themselves

fission can also be *induced* by bombarding the nucleus with an incident neutron

induced fission is used by humans to generate nuclear power or to build nuclear weapons

**nuclear reactors (∗)**

nuclear reactors make use of energy released from induced fission reactions

thermal energy produced from fission is further converted to electrical or mechanical forms

uranium-235 ($^{235}_{92}$U) is the most widely used fuel for fuel nuclear reactors

one of the many fission reactions of uranium-235 is: $^{235}_{92}\text{U} + ^{1}_{0}\text{n} \longrightarrow ^{141}_{56}\text{Ba} + ^{92}_{36}\text{Kr} + 3^{1}_{0}\text{n}$

➢ each reaction releases more than one neutrons

they can trigger further fission reactions, making *chain reaction* possible

hence huge amount of energy can be released in a very short time

➢ rate of fission should be controlled, a runaway reaction could lead to disastrous explosion

*control rods* (e.g., boron) are used to absorb neutrons: $^{10}_{5}\text{B} + ^{1}_{0}\text{n} \longrightarrow ^{7}_{3}\text{Li} + ^{4}_{2}\text{He}$

control rods are adjusted so that one neutron per reaction goes on to produce further fission

in emergency, release of boron rods shuts down reactor

➢ reaction is expected to continue at steady rate

---

reaction ceases, the star collapses and becomes a *white dwarf*. For a giant star, nuclear reactions can go all the way up to $^{56}_{26}$Fe. At this point, no more energy can be produced through fusion. The star collapses extremely rapidly and then explodes, creating a *supernova*, during which all elements heavier than $^{56}_{26}$Fe are formed. (∗)

need sufficient amount of neutrons to maintain the chain reaction

this requires a *critical mass* for the amount of $^{235}_{92}\mathrm{U}$ fuel

- only low energy neutrons (*thermal neutrons*) can be *captured* by $^{235}_{92}\mathrm{U}$

  but neutrons released through fission are very energetic

  a *moderator* (e.g., water) is needed to slow down neutrons
- heat produced in fission is removed by *coolant* (e.g., water)

  this heat can be used to power generators to produce electricity
- reactor is surrounded by a *shield* (e.g., a thick concrete) to prevent radiation from escaping

### 12.1.6 energy release from nuclear reactions

in this section, we consider only those nuclear reactions that release energy

let's write a generic nuclear reaction as: original particles $\longrightarrow$ product particles $+ Q$

two approaches to compute $Q$, the amount of energy release, will be introduced

**method 1**: use change in *mass* during the reaction

there must be a decrease in total mass to be converted into energy release

using mass-energy relation: $Q = \Delta mc^2 = (m_{\text{org}} - m_{\text{prod}})c^2$

**method 2**: use change of *total binding energy*

energy is released means product particles are more stable

so total binding energy of product particles is higher than that of original particles

energy released during reaction: $Q = E_{B,\text{prod}} - E_{B,\text{org}}$

**Example 12.3** The sun is a huge nuclear reactor. It fuses hydrogen nuclei into helium nuclei to produce large amounts of energy. The reaction can be expressed by the formula:

$$4\,^1_1\mathrm{H} \longrightarrow\, ^4_2\mathrm{He} + 2\,^0_{+1}\mathrm{e} \quad (^0_{+1}\mathrm{e} \text{ is a } \textit{positron}\text{, the anti-particle of electron})$$

(a) Evaluate the energy released in one reaction. (b) Given that the power output of the sun is roughly $4.0 \times 10^{26}$ W, estimate how much hydrogen the sun burns every second. (Data for this question: $m_{\text{p}} = 1.672622 \times 10^{-27}$ kg, $m_{\text{He}} = 6.644657 \times 10^{-27}$ kg, $m_{\text{e}} = 9.11 \times 10^{-31}$ kg)

✍ energy released in one reaction:

$$E_0 = \Delta mc^2 = (4m_{\text{H}} - m_{\text{He}} - 2m_{\text{e}})c^2 \approx 4.40 \times 10^{-29} \times (3.0 \times 10^8)^2 \approx 3.96 \times 10^{-12}\ \text{J}$$

in interval $\Delta t$, sun outputs a total energy of $P\Delta t$ via a number of $\Delta N$ fusion reactions

we write $P\Delta t = \Delta N E_0$, so number of fusion reactions every second is:

$$\frac{\Delta N}{\Delta t} = \frac{P}{E_0} = \frac{4.0 \times 10^{26}}{3.96 \times 10^{-12}} \approx 1.0 \times 10^{38}\ \mathrm{s}^{-1}$$

each reaction takes four hydrogen nuclei, so rate of hydrogen consumption:

$$\frac{\Delta M_{\mathrm{H}}}{\Delta t} = \frac{4\Delta N}{\Delta t} \times m_{\mathrm{H}} \approx 4 \times 1.0 \times 10^{38} \times 1.67 \times 10^{-27} \approx 6.7 \times 10^{11}\ \mathrm{kg\ s}^{-1}$$

the sun burns over 600 billion kilograms of hydrogen every second, just think about it! □

**Example 12.4** Uranium-235 nuclei ($^{235}_{92}\mathrm{U}$) bombarded with slow neutrons can undergo nuclear reaction:

$$^1_0\mathrm{n} + ^{235}_{92}\mathrm{U} \longrightarrow ^{141}_{56}\mathrm{Ba} + ^{92}_{36}\mathrm{Kr} + x^1_0\mathrm{n}$$

(a) Determine the number of $x$. (b) Using the data in the table, find the energy released for one reaction. (c) What is the change in mass, if any, before and after the reaction?

| | binding energy per nucleon |
|---|---|
| $^{235}_{92}\mathrm{U}$ | 7.591 MeV |
| $^{141}_{56}\mathrm{Ba}$ | 8.326 MeV |
| $^{92}_{36}\mathrm{Kr}$ | 8.513 MeV |

✍ from conservation of mass number: $1 + 235 = 141 + 92 + x \Rightarrow x = 3$

energy release during reaction equals change in total binding energy:

$$Q = E_{B,\mathrm{Ba}} + E_{B,\mathrm{Kr}} - E_{B,\mathrm{U}} = 141 \times 8.326 + 92 \times 8.513 - 235 \times 7.591 = 173.277\ \mathrm{MeV}$$

reduction in total mass: $\Delta m = \dfrac{Q}{c^2} = \dfrac{173.277 \times 1.60 \times 10^{-13}}{(3.0 \times 10^8)^2} \approx 3.08^{-28}$ kg □

**Question 12.7** Uranium-235 nuclei can undergo another fission process when being bombarded with neutrons. This is represented by:

$$^1_0\mathrm{n} + ^{235}_{92}\mathrm{U} \longrightarrow ^{144}_{55}\mathrm{Cs} + ^{90}_{37}\mathrm{Rb} + 2^1_0\mathrm{n}$$

Calculate the energy released in this reaction.

| | binding energy per nucleon |
|---|---|
| $^{235}_{92}\mathrm{U}$ | 7.591 MeV |
| $^{144}_{55}\mathrm{Cs}$ | 8.212 MeV |
| $^{92}_{36}\mathrm{Kr}$ | 8.631 MeV |

**Question 12.8** The *protonproton chain reaction* is a set of nuclear reactions through which stars fuse hydrogen into helium. One intermediate reaction may be summarised as:

$$^1_1\mathrm{H} + ^2_1\mathrm{D} \longrightarrow ^3_2\mathrm{He} + \text{energy}$$

Calculate the energy released in this reaction.

| | mass/u |
|---|---|
| $^1_1\mathrm{H}$ | 1.00728 |
| $^2_1\mathrm{D}$ | 2.01356 |
| $^3_2\mathrm{He}$ | 3.01605 |

**Question 12.9** Another fusion reaction in the proton-proton chain is represented by:

$$ {}_2^3\mathrm{He} + {}_2^3\mathrm{He} \longrightarrow {}_2^4\mathrm{He} + 2{}_1^1\mathrm{p} + \text{energy} $$

where energy released through this reaction is 12.86 MeV. Given that binding energy per nucleon of ${}_2^4$He is 7.074 MeV, calculate the binding energy per nucleon for ${}_2^3$He.

**Question 12.10** If the total mass of the product particles is greater than that of the original particles in a nuclear reaction, what can you say about the energy of the original particles?

## 12.2 radioactive decay

### 12.2.1 radioactive decay

the process of *random* and *spontaneous* emission of $\alpha$-, $\beta$-, and $\gamma$-radiation from unstable nuclei is called **radioactive decay**

- **random** means decay events are not predictable

  we cannot tell precisely when a particular nucleus is about to decay

  we can only tell the *probability* of decay within a certain time

- **spontaneous** means rate of decay does not depend on external conditions

  temperature, pressure, or has no effect on rate of radioactive decays

suppose initially a sample contains $N$ radioactive nuclei that are about to decay

after a time interval $\Delta t$, $\Delta N$ nuclei undergo decay processes

probability of decay for nuclei in the sample during this time interval is $\dfrac{\Delta N}{N}$

divide by $\Delta t$, probability of decay for each nucleus per unit time can be found

since nuclear decay is spontaneous, this number is a constant, called the decay constant

**decay constant** is the probability for one nucleus to decay per unit time: $\boxed{\lambda = \dfrac{\Delta N}{N\Delta t}}$

- to describe the rate of decay, we introduce the notion of **activity**

  activity is defined as the number of nuclei that undergo decay per unit time: $\boxed{A = \dfrac{\Delta N}{\Delta t}}$

  this can be given in a differential form: $A = -\dfrac{\mathrm{d}N}{\mathrm{d}t}$

a minus sign is included because number of nuclei $N$ decreases with time

➢ compare with the expression for decay constant, we find: $\boxed{A = \lambda N}$

this equation shows that if there are more nuclei present, the sample has greater activity

➢ units for decay constant and activity: $[\lambda] = [A] = \mathrm{s}^{-1}$

for activity, one decay event per unit time is defined as one *becquerel*: $1\ \mathrm{Bq} = 1\ \mathrm{s}^{-1}$

### 12.2.2 decay equations

variation of number of undecayed nuclei with time can be derived from the equation: $A = \lambda N$

recall that $A = -\frac{\mathrm{d}N}{\mathrm{d}t}$, so this equation is a differential equation in disguise: $-\frac{\mathrm{d}N}{\mathrm{d}t} = \lambda N$

this can be solved by separating variables and integrating

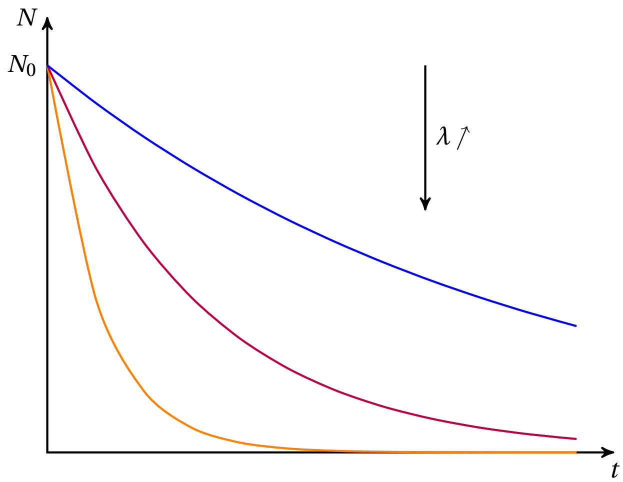
$$-\lambda \mathrm{d}t = \frac{\mathrm{d}N}{N} \Rightarrow -\lambda \int_0^t \mathrm{d}t = \int_{N_0}^N \frac{\mathrm{d}N}{N} \Rightarrow -\lambda t\Big|_0^t = \ln N\Big|_{N_0}^N \Rightarrow -\lambda t = \ln \frac{N}{N_0} \Rightarrow N = N_0 \mathrm{e}^{-\lambda t}$$

this shows that decay events obey *exponential decay laws*

➢ number of undecayed nuclei at time $t$ is given by: $\boxed{N(t) = N_0 \mathrm{e}^{-\lambda t}}$

$N_0$ is initial number of nuclei in the sample at $t = 0$

➢ $\lambda \nearrow \Rightarrow$ higher probability of decay, or greater rate of decay




exponential decay law for number of nuclei: $N(t) = N_0 \mathrm{e}^{-\lambda t}$

➢ since activity $A = \lambda N$, so activity varies with time as: $\boxed{A(t) = A_0 \mathrm{e}^{-\lambda t}}$

**Example 12.5** A sample containing $6.0 \times 10^7$ iodine-131 nuclei has an activity of 60 Bq. (a) What is the decay constant of iodine-131? (b) How many nuclei remain undecayed after 20 days?

✍ decay constant: $\lambda = \dfrac{A}{N} = \dfrac{60}{6.0\times 10^{7}} \approx 1.0\times 10^{-6}\ \mathrm{s^{-1}}$

number of remaining nuclei: $N = N_0 e^{-\lambda t} = 6.0\times 10^{7}\times e^{-1.0\times 10^{-6}\times 20\times 3600} \approx 1.1\times 10^{7}$ □

**Example 12.6** $^{24}_{11}$Na, an isotope of sodium, has a decay constant of $1.28\times 10^{-5}\ \mathrm{s^{-1}}$. Suppose a sample initially contains a mass of 9.0 $\mu$g of $^{24}_{11}$Na. Find (a) initial number of $^{24}_{11}$Na nuclei, (b) initial activity of this sample, (c) the activity after 48 hours.

✍ initial number of nuclei: $N_0 = \dfrac{9.0\ \mu\mathrm{g}}{24\mathrm{u}} = \dfrac{9.0\times 10^{-9}}{24\times 1.66\times 10^{-27}} \approx 2.26\times 10^{17}$

initial acitivy: $A_0 = \lambda N_0 = 1.28\times 10^{-5}\times 2.26\times 10^{17} \approx 2.89\times 10^{12}\ \mathrm{Bq}$

activity after 20 hours: $A = A_0 e^{-\lambda t} = 2.89\times 10^{12}\times e^{-1.28\times 10^{-5}\times 48\times 3600} \approx 3.17\times 10^{11}\ \mathrm{Bq}$ □

**Question 12.11** A sample of polonium-205 has an activity of $6.7\times 10^{15}$ Bq. The decay constant of polonium-205 is known to be $1.16\times 10^{-4}\ \mathrm{s^{-1}}$. (a) Find the number of nuclei needed to give this activity. (b) Find the mass of polonium-205 in this sample. (c) Calculate the time needed for the activity reduce to 1% of its initial value.

**Question 12.12** Technetium-99m is widely used as a radioactive tracer for medical diagnostic procedures. If some of this isotope, with an activity of 900 MBq was injected into a patient. The activity is found to reduce to 56.4 MBq after 24 hours. (a) What is the decay constant of technetium-99m? (b) How many technetium-99m nuclei are stil left in the patient's body?

### 12.2.3 half-life

it is more convenient to define a time quantity to describe how fast radioactive nuclei decay

it is useful to define half-life of a radioactive sample

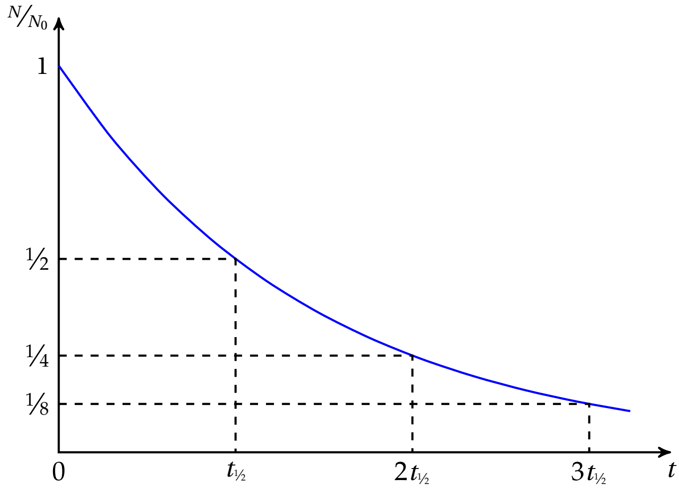
> **half-life** ($t_{1/2}$) is the mean time taken for number of radioactive nuclei in the sample, or activity of the sample, to reduce to half of its initial value

- half-life $t_{1/2}$ is closely related to decay constant $\lambda$

  by definition, at $t = t_{1/2}$, $N = \frac{1}{2}N_0$, $A = \frac{1}{2}A_0 \quad\Rightarrow\quad e^{-\lambda t_{1/2}} = \frac{1}{2} \quad\Rightarrow\quad e^{\lambda t_{1/2}} = 2 \quad\Rightarrow\quad \boxed{\lambda t_{1/2} = \ln 2}$

- $\lambda$ is a constant, so $t_{1/2}$ is also a constant over the lifetime of nuclear decay

  half-life of a sample of isotope does not depend on initial number of nuclei or initial activity




half-life of radioactive decay

**Example 12.7** Radium-224 has a half-life of 3.63 days. (a) What is the activity from 6.0 mg of pure radium-224? (b) How many radium-224 nuclei have undergone decay after 10 days?

✍ decay constant: $\lambda = \dfrac{\ln 2}{t_{½}} = \dfrac{\ln 2}{3.63 \times 24 \times 3600} \approx 2.21 \times 10^{-6}\ \mathrm{s}^{-1}$

initial number of Ra-224 nuclei: $N_0 = \dfrac{6.0\ \mathrm{mg}}{224\ \mathrm{u}} = \dfrac{6.0 \times 10^{-6}}{224 \times 1.66 \times 10^{-27}} \approx 1.61 \times 10^{19}$

initial acitivy: $A_0 = \lambda N_0 = 2.21 \times 10^{-6} \times 1.61 \times 10^{19} \approx 3.57 \times 10^{13}$ Bq

number of undecayed nuclei: $N = N_0 \mathrm{e}^{-\lambda t} = 1.61 \times 10^{19} \times \mathrm{e}^{-2.21 \times 10^{-6} \times 10 \times 24 \times 3600} \approx 2.39 \times 10^{18}$

number of nuclei that have decayed: $\Delta N = N_0 - N = 1.61 \times 10^{19} - 2.39 \times 10^{18} \approx 1.37 \times 10^{19}$ □

**Example 12.8** Living trees contain a certain percentage of ${}^{14}_{6}$C, an radioactive isotope of carbon that has a half-life of 5570 years. A sample of dead wood is found to have an activity of 0.42 Bq, while an equal mass of living wood has an activity of 1.60 Bq. Find the age of the dead wood.

✍ $A = A_0 \mathrm{e}^{-\lambda t} = A_0 \mathrm{e}^{-\frac{\ln 2}{t_{½}} t} \Rightarrow 0.42 = 1.60 \mathrm{e}^{-\frac{\ln 2}{5570} t} \Rightarrow -\dfrac{\ln 2}{5570} t = \ln\left(\dfrac{0.42}{1.60}\right) \Rightarrow t \approx 10700$ years □

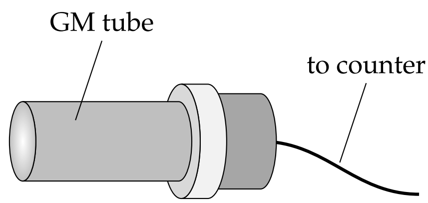
**Question 12.13** The number of uranium-238 nuclei in a rock sample is believed to have decreased from $3.50 \times 10^{17}$ to $3.27 \times 10^{17}$ in 480 million years. Estimate the half-life of uranium-238.

**Question 12.14** Plutonium-238 is a powerful alpha emitter with a half-life of 87.7 years. One decay of plutonium-238 releases an energy of about $9.0 \times 10{-}13$ J. A nuclear battery containing a sealed plutonium source is implanted into patient's body to power heart pacemakers. The battery has an initial activity of $6.0 \times 10^{10}$ Bq. (a) Calculate the initial power released by the source. (b) Find the mass of plutonium required to produce this power. (c) It is required that power output to the pacemaker is at least 60% of the initial power. Calculate the time, in years, for which the battery provides sufficient power.

**Question 12.15** Show that the variation of the number of undecayed nuclei with time $t$ can be given by: $N(t) = N_0 \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}}$, where $N_0$ is the initial number of nuclei.

### 12.2.4 measurement of radioactive decay

a GeigerMüller tube, or a **GM tube**, is a device used to detect ionizing radiation




GM tube measures the number of $\alpha$-particles, $\beta$-particles and $\gamma$-photons arriving per unit time

the number of decays recorded per unit time by GM tube is called the **count rate** $R$

since GM tube only picks up emissions to one particular direction, but a radioactive source emits radiation in *all* directions, so count rate is a fraction of the activity of the sample

activity obeys exponential decay, so we would expect count rate to satisfy: $\boxed{R = R_0 e^{-\lambda t}}$

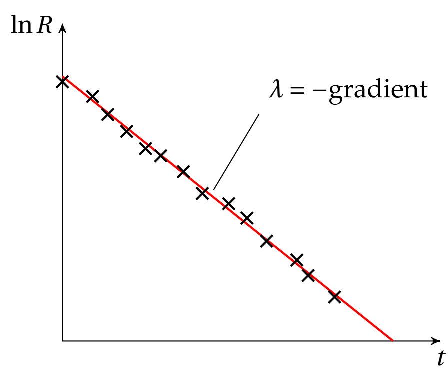
#### verifying the exponential decay law

to verify the exponential decay law, we first rearrange the equation as: $\ln R = -\lambda t + \ln R_0$

if a set of measurements of $R$ at time $t$ are obtained, a graph of $\ln R$ against $t$ can be plotted

if trend curve shows a straight line, the relation $R = R_0 e^{-\lambda t}$ is then verified

decay constant $\lambda$ is given by negative gradient of the best-fit line




$\ln R = -\lambda t + \ln R_0$

**issues with count rate**

for now, we take for granted that the count rate measured is 100% accurate

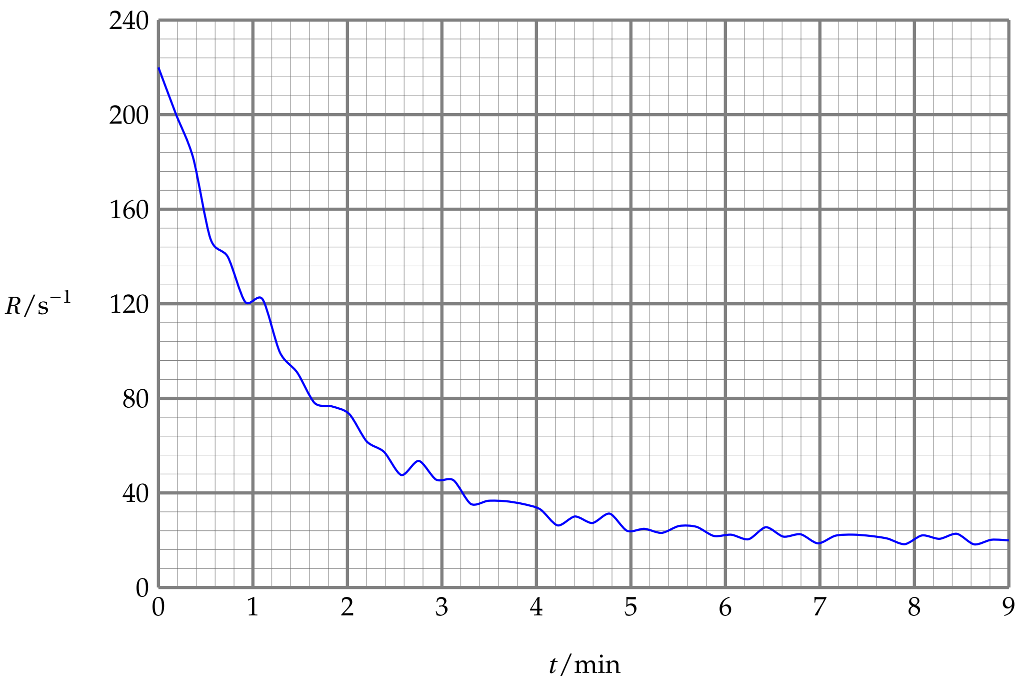
but in practice, there might be a number of factors leading to measurement errors

- there is always *background radiation* from earth minerals, cosmic rays, food and water, etc.

  background reading must be subtracted to give a corrected count rate
- $\alpha$- and $\beta$-particles emitted could be absorbed by sample itself
- product nuclei could also be radioactive, contributing to additional counts
- GM tube has *dead time*, or *resolving time*

  when a count is recorded, it takes GM tube a certain time to reset for the next count

**Example 12.9** A GM tube is placed close to a source of radioactive isotope. The variation with time $t$ of the measured count rate $R$ is shown. Determine the half-life of this isotope.




✍ there is evidence of *background radiation* as the count rate decreases and tends to a non-zero value of 20 $s^{-1}$, so we need subtract this number to obtain the true count rate of sample

there are also *fluctuations* in count rate due to *random* nature of decay processes, so we need to make several calculations and take average to obtain an accurate value for half-life

let's call true count rate $R_T$, and call measured count rate $R$

as $R_T = 200\ \mathrm{s}^{-1} \to 100\ \mathrm{s}^{-1}$, $R = 220\ \mathrm{s}^{-1} \to 120\ \mathrm{s}^{-1}$, so $t_{½} \approx 1.0 - 0.0 \approx 1.0$ min

as $R_T = 160\ \mathrm{s}^{-1} \to 80\ \mathrm{s}^{-1}$, $R = 180\ \mathrm{s}^{-1} \to 100\ \mathrm{s}^{-1}$, so $t_{½} \approx 1.3 - 0.4 \approx 0.9$ min

as $R_T = 120\ \mathrm{s}^{-1} \to 60\ \mathrm{s}^{-1}$, $R = 140\ \mathrm{s}^{-1} \to 80\ \mathrm{s}^{-1}$, so $t_{½} \approx 1.6 - 0.7 \approx 0.9$ min

as $R_T = 80\ \mathrm{s}^{-1} \to 40\ \mathrm{s}^{-1}$, $R = 100\ \mathrm{s}^{-1} \to 60\ \mathrm{s}^{-1}$, so $t_{½} \approx 2.3 - 1.3 \approx 1.0$ min

take average for these results, we find $t_{½} \approx 0.95$ min □

# CHAPTER 13
# Medical Imaging

symptoms of disease are to be diagnosed without cutting patient open, or inserting surgical instruments into patient's body, we need **non-invasive** techniques, which include:

- X-ray/CT imaging
- ultrasonic scans
- MRI (magnetic resonance imaging)
- PET (positron emission tomography) scans]

## 13.1 X-ray imaging

### 13.1.1 nature of X-rays

X-ray is electromagnetic radiation with short wavelength $\lambda$, or high frequency $f$

radiation with a wavelength in the range $10^{-12} \sim 10^{-8}$ m are categorised as X-rays

➢ X-rays are further categorised into *hard* and *soft* X-rays

**hardness** of X-ray refers to its penetrating ability

harder X-ray photons are more energetic, so more penetrating

hard X-rays have higher $f$ and shorter $\lambda$ ($10^{-13} \sim 10^{-11}$m)

soft X-rays have slightly lower $f$ and longer $\lambda$ ($10^{-11} \sim 10^{-9}$m)

➢ soft-X-rays are easily absorbed by patient's body, hence not contributing to imaging

in medical imaging, metal sheet (e.g., aluminium) is used to filter out soft X-rays

hard X-rays can penetrate through patient's body and are more useful to form images

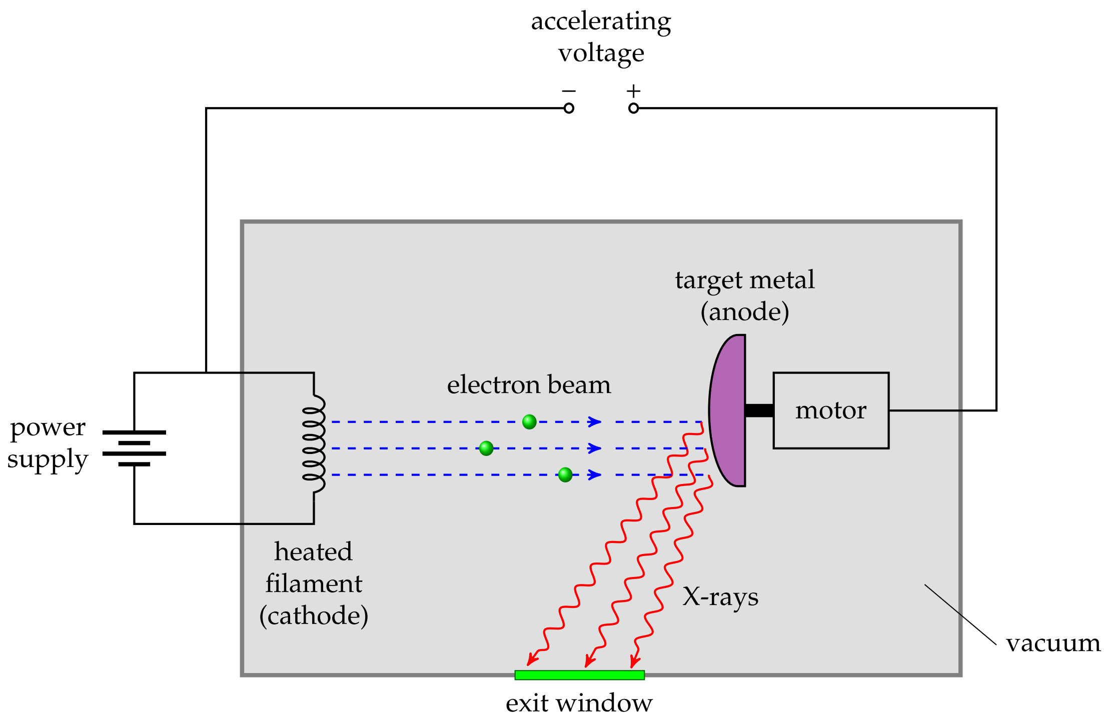
➢ there is no specific critical $f/\lambda$ to distinguish X-ray from $\gamma$-ray in electromagnetic spectrum

radiation emitted by *electrons* is called X-ray, radiation emitted from *nuclei* is called $\gamma$-ray

### 13.1.2 production of X-rays

X-ray can be produced by smashing electrons onto a target metal

the diagram below illustrates the production of X-ray in a *X-ray tube*




X-ray production tube

under a strong electric field, electrons are pulled out from a heated cathode

these electrons are accelerated through a high p.d. towards the anode

anode is made of target metal (tungsten, cobalt, etc.)

electrons smash into anode, their K.E. are converted into the form of X-ray radiation

beams of X-rays leave the X-ray tube through an exit window

➢ accelerating voltage of X-ray tube $V_{ac}$ controls energy of X-ray photons emitted

at higher $V_{ac}$, electrons have greater K.E., X-ray photons of higher energies are produced

➢ varying cathode current alters number of electrons flowing through filament per unit time

for a larger current, more electrons are pulled towards anode, producing more X-ray photons

so increasing current in X-ray tube increases intensity of X-ray beam

- in X-ray tube, only a tiny fraction of electron K.E. is converted into X-rays

  most energy is converted into thermal energy in anode metal

  to avoid overheating, anode is rotated by a motor, and water can be used as coolant

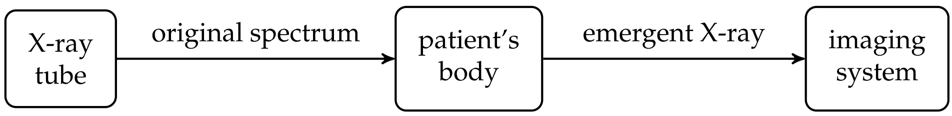
### 13.1.3 X-ray spectrum

to image a patient's body using X-rays, X-rays produced are sent towards the patient

X-ray emergent from patient's body can be measured and compared with original spectrum

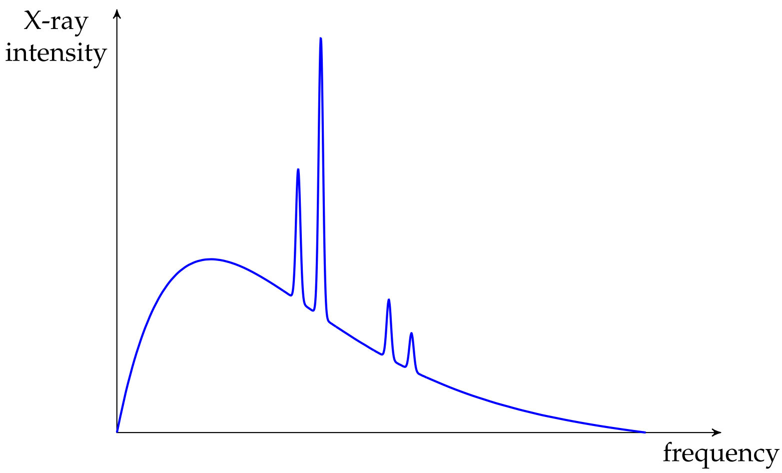
from this we can learn about absorption process of X-ray in patient's body

so information about patient's body can be obtained




so we need to study properties of *X-ray spectrum* as it is produced in X-ray tubes

the graph below shows a typical frequency spectrum of X-ray emitted from an X-ray tube




frequency spectrum of X-ray beam produced in an X-ray tube

this spectrum consists of two parts: a *continuous background*, and a few *discrete spikes*

these are due to two different mechanisms, about which we are about to discuss

**braking radiation**

continuous background in X-ray spectrum is due to **braking radiation**

high-speed electrons decelerate when they strike into the target metal

loss of K.E. of electrons are converted into electromagnetic energy, i.e. X-ray photons

no constraint on final K.E. of electrons, so there could be a *distribution* of K.E. loss

X-ray photons produced therefore are allowed to take a range of energies

so a range of frequencies of the X-ray photons, giving a broad continuous spectrum

➢ there exists a maximum frequency for emitted X-rays

X-ray photon of highest energy is produced if an electron loses all its K.E. in one collision

photon energy is proportional to frequency of radiation, so sharp cut-off at high $f$

**Example 13.1** An X-ray tube uses accelerating voltage of 5000 V, what is the highest frequency of the X-ray produced?

✍ electrical P.E. converts into K.E. of electron, then converts into X-ray photon

$$qV_{\text{ac}} = hf_{\text{max}} \quad \Rightarrow \quad f_{\text{max}} = \frac{qV_{\text{ac}}}{h} = \frac{1.60 \times 10^{-19} \times 5000}{6.63 \times 10^{-34}} \approx 1.2 \times 10^{18} \text{ Hz} \qquad \square$$

**Question 13.1** Explain why the spectrum of X-ray would have a minimum wavelength.

**characteristic radiation**

discrete peaks in the X-ray spectrum arise from a process called **characteristic radiation**

incident electrons can knock out orbital electrons from inner shells in target metal

orbital electrons from higher energy levels drop to fill these vacancies

*de-excitation* of the electrons gives out energy as X-ray photons

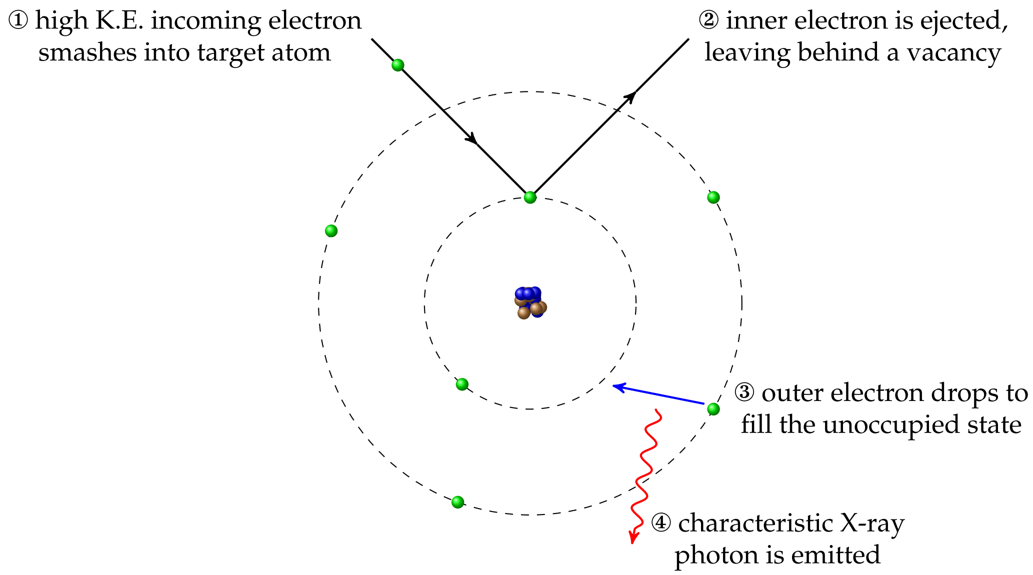
electron energy levels in atoms are discrete, so only certain X-ray frequencies are emitted

➢ positions of characteristic lines are determined by target metal

each metal has its unique electron energy level, producing unique characteristic lines

**Question 13.2** If accelerating voltage in the X-ray tube is increased, suggest and explain the changes to the X-ray spectrum.

**Question 13.3** Summarise the principles of production of X-rays.




production of characteristic radiation

### 13.1.4 X-ray attenuation

as an X-ray beam penetrates through a substance, its intensity decreases

gradual decrease in X-ray intensity is called **attenuation**

change in X-ray intensity can be measured to extract information about patient's body

the law of X-ray attenuation is given by $\boxed{I = I_0 e^{-\mu x}}$

this shows a beam of initial intensity $I_0$ decreases exponentially with distance $x$ travelled

$\mu$ is **linear attenuation constant**, or **absorption constant** of the substance

➢ the equation $I = I_0 e^{-\mu x}$ applies to *parallel* beam of X-ray only

intensity of a *divergent* beam will decrease even without absorption

➢ recall that *intensity* of radiation is defined as power transmitted per unit time: $I = \frac{P}{A}$

this is related to number of photons in the beam per unit time, given by: $\frac{\Delta N}{\Delta t} \times hf$

as each photon passes through some distance, it has certain probability to be absorbed

we may think of $\mu$ as probability for one single photon to be absorbed per unit distance

make analogy with nuclear decays, X-ray intensity should also decrease exponentially [80]

[80]You may have noticed the similarity between the law of X-ray attenuation ($I = I_0 e^{-\mu x}$) and nuclear

➢ substance with high value of $\mu$ is a good absorber of X-ray

bones are much better absorbers than soft tissues ($\mu_{\text{bone}} > \mu_{\text{tissue}}$)

with X-ray imaging techniques, bones are easily distinguished from soft tissues

➢ attenuation constant $\mu$ depends on hardness of an X-ray beam

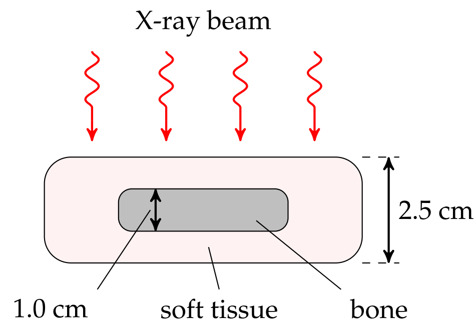
high-frequency X-ray is more penetrating, so a material has smaller $\mu$ at higher frequency

➢ can define *half-value thickness* $x_{½}$ (recall half-life $t_{½}$ for nuclear decays)

this is thickness of material at which intensity of X-ray becomes halved

it follows that: $\frac{1}{2}I_0 = I_0 e^{-\mu x_{½}} \Rightarrow \boxed{x_{½} = \frac{\ln 2}{\mu}}$ [81]

**Example 13.2** A parallel beam of 30 keV X-ray of initial intensity $I_0$ is incident on a specimen of soft tissue and bone as shown. Find the emergent intensity from the specimen if (a) the beam passes through soft tissue alone, (b) the beam passes through the bone and the soft tissue.




attenuation constants for 30 keV X-ray

| | $\mu/\text{cm}^{-1}$ |
|---|---|
| tissue | 0.40 |
| bone | 2.55 |

✍ through soft tissue only:

$$I = I_0 e^{-\mu x} = I_0 e^{-0.40\times 2.5} \approx 0.37 I_0$$

through soft tissue and bone:

$$I = I_0 e^{-\mu_m x_m} e^{-\mu_b x_b}$$

$$I_0 e^{-0.40\times 1.5} \times e^{-2.55\times 1.0} \approx 0.043 I_0$$

emergent intensity will be very different depending on whether it travels through bone

so there will be clear difference between bone and soft tissue in the image □

**Question 13.4** Copper has a linear attenuation coefficient 0.41 mm$^{-1}$ for X-ray radiation of energy 100 keV. If a copper filter is able to reduce the intensity of a parallel beam of X-ray by 90%, what is the thickness required?

decay laws ($N = N_0 e^{-\lambda t}$). The analogy is clear: intensity $I$ of radiation is analogous to number of undecayed nuclei $N$, distance $x$ is analogous to time $t$, and probability constant are $\mu$ and $\lambda$ respectively, where $\lambda$, the decay constant, tells the probability of decay for a radioactive nucleus per unit time.

[81] Compare with half-life of nuclear decay: $t_{½} = \frac{\ln 2}{\lambda}$.

#### 13.1.5 X-ray imaging techniques

basic principle of conventional X-ray imaging: expose patient to a beam of X-ray, a fraction of the beam is absorbed by patient's body, a film or a detector captures attenuated X-rays, resulting in a black-and-white shadow image

high-quality imaging requires good sharpness, good contrast and minimal radiation dosage

**image sharpness**

**sharpness** of image means whether details on edges of structures can be seen with ease

sharpness of image is mainly determined by *width* of X-ray beam

so a narrow beam of *parallel* X-ray is desired

➢ parallel X-ray beam can be realized by

- reducing size of target metal at anode of X-ray tube
- reducing size of exit window of X-ray tube
- using a set of lead slits to make the beam more parallel, known as *collimation*
- using *anti-scatter screen* to reduce influence of scattering of radiation in patient's body

➢ good sharpness also requires patient to keep still when image is taken

**image contrast**

**contrast** means whether different structures show up with different degrees of blackening

good contrast requires a great difference in X-ray intensity

➢ for substances with large difference in $\mu$, they naturally produce good contrast

➢ most soft tissues are poor absorbers of X-rays, not easy to distinguish them on same image

artificial **contrast media** such as iodine (I) injections , barium (Ba) meals[82] can be used to turn a tissue into a better absorber, therefore greatly improve contrast of image

[82]Contrast media usually contain elements of large atomic numbers, which means each atom accommodates a large number of electrons, so it is more likely for them to interact with X-ray photons, making these atoms good absorbers of radiation.

**Question 13.5** The table below shows the linear attenuation coefficient $\mu$ for X-ray beams of different energies in cortical bone and muscle. To produce an X-ray image where the bone is clearly distinguished from muscle, why is it more desirable to use low-energy X-rays?

| maximum X-ray energy/keV | bone: $\mu/\text{cm}^{-1}$ | muscle: $\mu/\text{cm}^{-1}$ |
|---|---|---|
| 30 | 2.55 | 0.40 |
| 80 | 0.43 | 0.19 |
| 200 | 0.25 | 0.14 |
| 800 | 0.14 | 0.082 |

#### dosage of radiation

X-ray radiation is ionising, hence harmful to human body

image should be developed quickly, so exposure time is minimised to reduce possible harm

to reduce exposure, or **dosage** of radiation, efficient imaging systems are needed

➢ **image intensifiers** are used to reduce dosage of X-ray radiation

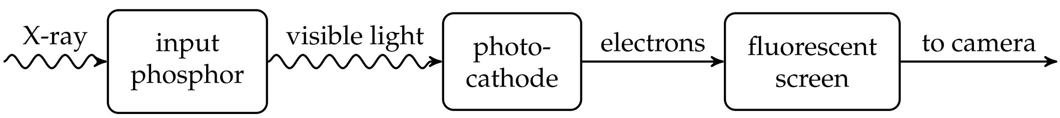



illustration of how the image is developed efficiently with an image intensifier

### 13.1.6 CT scans

conventional X-ray imaging only produces a 2D *shadow image*

information about *depth* of structures cannot be displayed

computerised axial tomography scan, or **CT scan**, is developed to produce 3D images

#### principles of CT imaging

patient's body can be thought to consist of many 2D slices each slice can be divided into a grid of many units, called **voxels**

to obtain image of one slice, computer-programmed X-rays are exposed from all directions

detectors opposite to X-ray tube pick up transmitted signals and feed them to computer

computer uses emergent intensities from all directions to deduce information about each voxel, therefore a 2D slice image can be built up

this process is repeated, so images for many 2D slices are obtained

2D slice images are combined to get a comprehensive 3D image

the 3D image can be rotated and viewed from different angles

➢ advantages of CT imaging

- precise position, shape, size of structures can be shown
- tissues with similar attenuation coefficients can be distinguished

➢ disadvantages of CT imaging

- more expensive than traditional 2D radiography
- patient is exposed to higher radiation dosage

**building up images (∗)**

we now discuss how information of each voxel is recovered from transmitted X-rays

we consider a toy model, a $2 \times 2$ grid with 4 voxels

each voxel makes its contribution to attenuation, described by numbers $a$, $b$, $c$ and $d$

grid is exposed to a parallel beam of X-ray, readings from detectors are assigned to associated voxel, the voxels are reconstructed and recorded in a memory gird

this process is repeated for four different directions

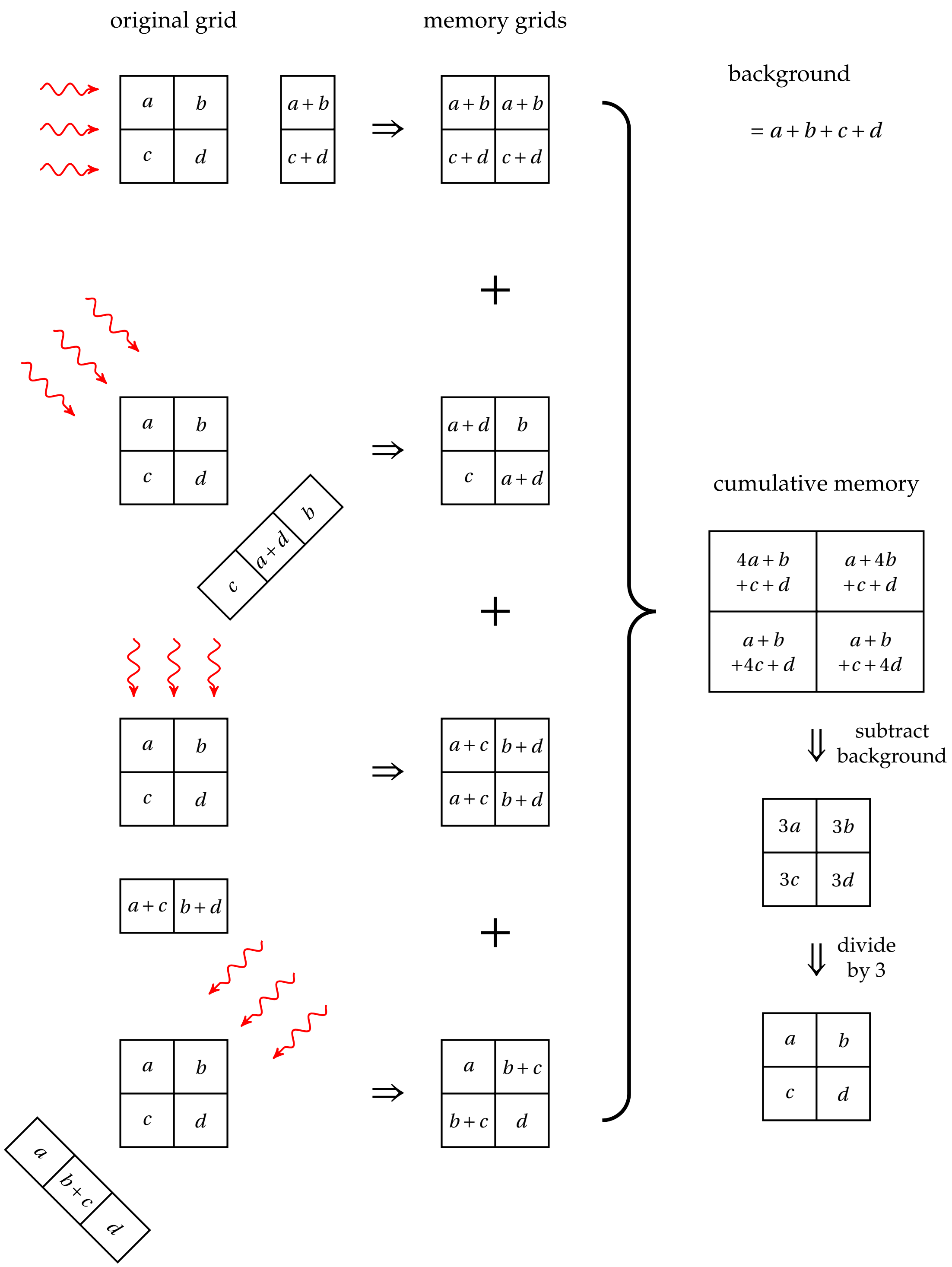
all the memory grids are then added to construct a cumulative grid

knowing background reading, which is $a+b+c+d$ from all voxels, and cumulative reading, information of the original voxels can be obtained within two simple steps

step 1: subtract background reading from cumulative grid

step 2: divide the result by three

the whole process is shown in the figure, which should be self-explanatory

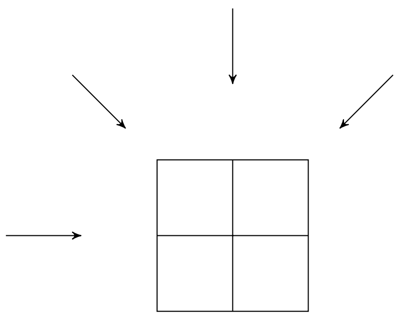



recovering information of the original $2 \times 2$ voxels using

detector readings obtained from exposure at different angles in practice, high-quality 2D slice image is built up from many voxels, attenuated intensities from many different directions are measured by an array of many detecting units [83]

huge amount of data is to be collected and processed, this requires use of power computers

**Question 13.6** An X-ray image of a 2-by-2 pixel section is obtained by exposing X-rays from four different angles. Detector readings from four consecutive directions are summed to give the cumulative pattern shown. For any one direction, total of detector readings is 19. (a) Find the value of each pixel and mark them on the graph. (b) For each of the directions shown, state the readings measured by detectors.



| | |
|---|---|
| 34 | 28 |
| 40 | 31 |

cumulative reading

**Question 13.7** By reference to the principles of CT imaging, suggest why the radiation dosage from a CT scan is greater than a simple X-ray image.

## 13.2 ultrasonic scans

### 13.2.1 nature of ultrasound

sound wave is a *longitudinal mechanical* wave

human hearing range is around 20 Hz (low pitch) ~ 20,000 Hz (high pitch)

sound waves of frequencies higher than 20,000 Hz is **ultrasound**

➢ meaning of longitudinal waves and mechanical waves

propagation of sound waves require material medium (air, water, tissue, etc.)

longitudinal wave travels in a parallel direction to compression and rarefaction of medium

[83] For our simple toy model, each voxel is exposed for four times and superimposed with each of the other voxels once, this is why the subtract-background-and-divide-by-three trick works.

➢ rationale for using ultrasounds in medical imaging

for a given medium, speed of sound is constant

$v = \lambda f \Rightarrow$ greater frequency means shorter wavelength, which means better *resolution*

so fine details of small structures can be detected

**Question 13.8** Sound waves travel at a speed of around 1600 m s$^{-1}$ in soft tissues. Compare the wavelength of sound waves if they have a frequency of (a) 500 Hz, (b) 10 MHz.

#### 13.2.2 production of ultrasounds

sound can be produced by a *piezo-electric transducer*[84]

if apply a high-frequency alternating voltage across a piezo-electric crystal, the crystal is made to vibrate, ultrasound wave is produced

➢ frequency of ultrasound is determined by frequency of alternating voltage: $f_{\text{sound}} = f_{\text{a.c.}}$

to generate ultrasounds, we need $f_{\text{a.c.}} > 20,000$ Hz

➢ there is an optimum thickness for crystal to give out ultrasound of greatest intensity

when thickness of crystal $d = \frac{1}{2}\lambda$, amplitude of vibration is maximum [85]

➢ in ultrasonic scanning, reflected sound waves are detected and processed to build images

the same piezo-electric transducer is also used to act as a sound detector

for each pulse emitted, vibration of crystal must stop quickly before reflected echo arrives

a damping mechanism is required to ensure vibration would stop in very short time

➢ a typical piezo-electric transducer used in ultrasonic scanning consists of several parts:

- a piece of piezo-electric crystal (quartz, PZT, PVDF, etc.)

  this is the heart of the transducer, responsible for generating and detecting sound waves

- an acoustic window (made of good sound transmitter)

  volume of sound generated is boosted with the acoustic window

[84] Piezo-electricity and its use as sensing devices will be discussed in §15.1.4.

[85] Resonance occurs when a nice *stationary wave* pattern is formed within the piezo-electric crystal. It can be shown that this occurs if the thickness of the crystal satisfies: $d = \left(n + \frac{1}{2}\right)\lambda$, where $n = 0, 1, 2, \cdots$. However, additional thickness would cause greater attenuation, so the optimum thickness is $d = \frac{1}{2}\lambda$.

– a damping block (usually made of epoxy-resin)

this stops vibration of crystal quickly before reflected wave returns

– electrodes/connecting wires

signals can be sent to transducers from control units to generate ultrasounds

signals detected by transducer can also be sent back to processing units to build images

### 13.2.3 attenuation of ultrasounds

just like X-rays, intensity of ultrasound also decreases as it propagates through a substance

intensity $I$ decreases *exponentially* with depth $x$: $\boxed{I = I_0 \mathrm{e}^{-\alpha x}}$

$I_0$ is original intensity, $\alpha$ is called **absorption coefficient** of the medium

➢ ultrasonic scans mostly rely on detection of reflected waves from various boundaries

so attenuation is not a big problem, weak echoes from deep tissues can be *amplified*

### 13.2.4 reflected & transmitted intensity at boundaries of structures

as ultrasound travels from one medium to another, it undergoes *reflection* and *transmission*

intensities of incident, reflected and transmitted ultrasounds satisfy the obvious relation:

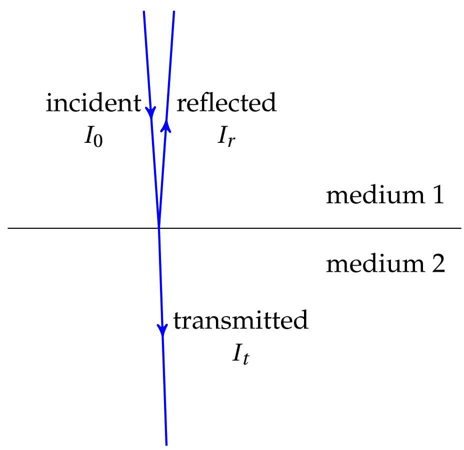
$$I_0 = I_r + I_t$$

for *normal incidence*, fraction of reflected and transmitted ultrasounds are further given by:

reflection coefficient: $\boxed{\dfrac{I_r}{I_0} = \left(\dfrac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2}$

transmission coefficient: $\boxed{\dfrac{I_t}{I_0} = 1 - \left(\dfrac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2}$ [86]

where $Z$ is the medium's *acoustic impedance*




ultrasound at boundary of two medium

[86] Proof of this relation involves solving second-order partial differential wave equations under certain boundary conditions, in which the idea of acoustic impedance is formally defined when one describes how much sound pressure is produced by vibration of acoustic medium at a given frequency.

➢ **acoustic impedance**, also called specific acoustic impedance, is defined as the product of density of medium and speed of sound wave travelling through it: $\boxed{Z = \rho c}$

**Example 13.3** Data for air and soft tissue is given in the table below. When an ultrasound is sent from air into human body at right angles, suggest what happens.

| | density/kg m$^{-3}$ | speed of sound/m s$^{-1}$ |
|---|---|---|
| air | 1.29 | 330 |
| soft tissue | 1060 | 1540 |

✍ let's first find the acoustic impedance for air and soft tissue:

$$Z_{\text{air}} = \rho_{\text{air}} c_{\text{air}} = 1.29 \times 330 = 4.26 \times 10^2 \text{ kg m}^2\text{s}^{-1}$$

$$Z_{\text{tissue}} = \rho_{\text{tissue}} c_{\text{tissue}} = 1060 \times 1540 = 1.63 \times 10^6 \text{ kg m}^2\text{s}^{-1}$$

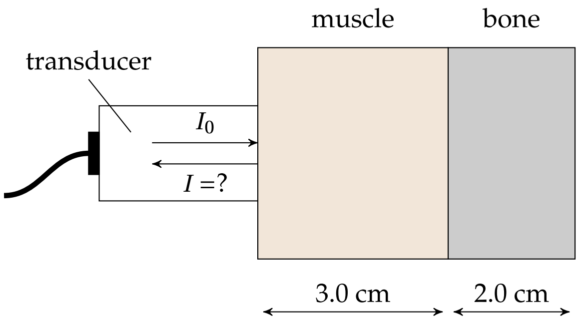
at air-tissue boundary, reflection coefficient is: $\dfrac{I_r}{I_i} = \left(\dfrac{4.26 \times 10^2 - 1.63 \times 10^6}{4.26 \times 10^2 + 1.63 \times 10^6}\right)^2 \approx 99.9\%$

almost complete reflection, so almost no ultrasound can enter human body

to overcome this problem, transduce is coupled to patient's skin using a *gel*, which is a transparent watery substance that has a similar acoustic impedance to soft tissue,

with *impedance matching*, ultrasound can get into patient's body □

**Example 13.4** An ultrasonic pulse of intensity $I_0$ is sent into a layer of muscle, it then arrives at a muscle-bone boundary (see figure). The speed of sound $c$, acoustic impedance $Z$ and absorption coefficient $\mu$ for muscle and bone are given in the table. (a) Find the intensity of the wave reflected from the muscle-bone boundary and received back at the surface of the muscle. (b) Find the arrival time of this signal after the pulse is transmitted into patient's body.




| | muscle | bone |
|---|---|---|
| $c$/m s−1 | 1600 | 4000 |
| $Z$/kg m$^2$s$^{-1}$ | $1.7 \times 10^6$ | $6.4 \times 10^6$ |
| $\mu$/m$^{-1}$ | 23 | 130 |

✍ intensity received at muscle surface: $I = I_0 \times \underbrace{e^{-\mu_m x_m}}_{\text{attenuation in muscle}} \times \underbrace{\left(\frac{Z_m - Z_b}{Z_m + Z_b}\right)^2}_{\text{reflection at boundary}}$

$$I = I_0 \times e^{-23\times2\times0.030} \times \left(\frac{6.4\times10^6 - 1.7\times10^6}{6.4\times10^6 + 1.7\times10^6}\right)^2 \approx I_0 \times 0.252 \times 0.337 \approx 0.085 I_0$$

note that ultrasound travels from surface of muscle to the muscle-bone boundary, and then travels from the boundary back to the muscle surface, so it is attenuated through a total distance that equals twice the thickness of the muscle

time taken for this return trip: $\Delta t = \frac{2x_m}{c_m} = \frac{2\times0.030}{1600} \approx 3.75\times10^{-5}$ s

in practice, arrival time for echo waves are measured so depth of structures are found □

**Question 13.9** Given that a gel has an acoustic impedance of $1.65\times10^6$ kg m$^2$s$^{-1}$. Use data in Example 13.3, calculate the reflection coefficient when ultrasound is normally incident from a gel to the human body, and hence explain why a gel is put on the patient's skin during medical diagnosis using ultrasounds.

**Question 13.10** It is a good idea to use ultrasound to image the lungs? Give your reasons.

**Question 13.11** In Example 13.4, if an ultrasound detector is couple to the surface of the bone at the far right, what intensity would be measured?

### 13.2.5 ultrasound imaging

the principles of ultrasonic imaging can be summarised as the following:

ultrasonic *pulses* produced from *piezo-electric transducer* are sent into patients body

pulses are reflected from *boundaries* of different media

*reflected waves*, or echoes, are received and detected by the same transducer

these signals are processed to construct the image

*arrival times* of the reflected waves can give information about *thickness* of structures

*intensity* of the reflected signals depend on acoustic impedance and attenuation coefficient of media, so can give information about *nature* of structures

➢ advantages of ultrasound imaging

- safe, no harmful radiation involved (can be used in diagnosis of fetus during pregnancy)
- provides images in real-time, almost no processing delay

  - low cost, apparatuses of ultrasound scans are not expensive
  - portable, apparatuses can be easily brought to a patient's bedside
- disadvantages of ultrasound imaging
  - difficult to image structures behind bones or air
  - require a highly-skilled operator to read images
- ultrasounds are also used in other areas of medical diagnosis
  - image flow of blood in vessels with Doppler ultrasonography
  - detect problems in liver, heart, kidney, abdomen, etc.
  - break up kidney stones by high-frequency resonance
  - clean teeth, medical devices, etc.

## 13.3 MRI (∗)

### 13.3.1 nuclear magnetic resonance

nuclei of some atoms have an intrinsic property called **spin**

at this stage, you can think of a nucleus as if it is spinning around [87]

since all nuclei are positively-charged, spinning gives rise to a circulating current, magnetic fields are created around any nucleus, so spin makes nuclei behave like tiny magnets

**step 1. split of energy levels**

in absence of magnetic field, nuclei align randomly, any orientation shares the same energy

apply a strong magnetic field $B_0$, nuclei would rearrange themselves

[87]This picture might be helpful to the understanding of NMR, but is not correct in the context of quantum mechanics. Spin is a *quantum* concept that has no analogy in classical physics. Spin is an *intrinsic* property of particles. For a proton, an electron, or any elementary particle, they have a fixed magnitude of a spin number, which only depends on the type of particle. This spin number has nothing to do with rotation. Spin also gives information about directions, which is even more peculiar. The spin component along any axis is *quantized*, i.e., only certain discrete values are allowed.

most nuclei line up parallel to $B_0 \longrightarrow$ stable, low-energy states

a few line up anti-parallel to $B_0 \longrightarrow$ unstable high-energy states

energy level of nuclei splits up when external magnetic field $B_0$ is switched on

note that nucleus energy levels in presence of $B_0$ are discrete

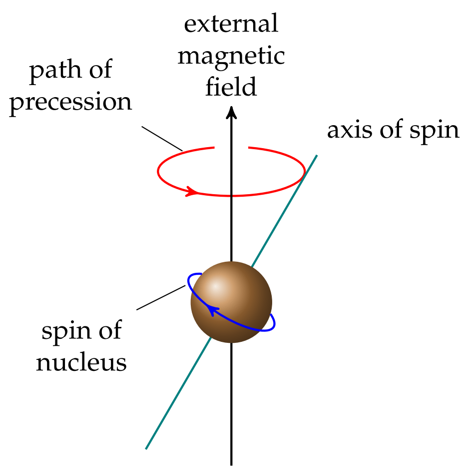
**step 2. precession of nuclei**

nuclei do not line up exactly along external magnetic field $B_0$, axis of spin is slightly tilted

axis of spin rotates about an axis parallel to $B_0$, this rotational motion is called **precession**

precession occurs at an angular frequency known as **Larmor frequency**, which is proportional to flux density $B_0$ of external field: $\boxed{\omega_0 = \gamma B_0}$

$\gamma$ is the gyromagnetic ratio of the nucleus, which depends of the type of the nucleus




**step 3. nuclear magnetic resonance (NMR)**

radio pulses $B_{RF}$ at same Larmor frequency are applied

low-energy nuclei absorb RF photons and flip up into higher states

this process is **nuclear magnetic resonance** [88]

**Example 13.5** The gyromagnetic ratio for proton is $2.68 \times 10^8$ rad s$^{-1}$ T$^{-1}$. What is the required frequency of the radiation pulse to cause magnetic resonance of hydrogen nuclei in a strong magnetic field of 1.8 T?

✍ Larmor frequency of precession: $\omega_0 = \gamma B_0 = 2.68 \times 10^8 \times 1.8 \approx 4.82 \times 10^8$ rad s$^{-1}$

resonant frequency matches Larmor frequency, so: $f = \dfrac{\omega_0}{2\pi} = \dfrac{4.82 \times 10^8}{2\pi} \approx 7.7 \times 10^7$ Hz

this frequency is within radio frequency, so *radio pulses* should be applied □

[88]NMR was first described and measured by *Isidor Rabi* in 1938. Rabi was awarded the 1944 Nobel Physics Prize 'for his resonance method for recording the magnetic properties of atomic nuclei'

**step 4. relaxation**

once RF pulses $B_{\mathrm{RF}}$ are switched off, nuclei flip back to lower energy states

during the process of **relaxation**, or **de-excitation**, RF photons of certain energies are released

time taken for nuclei to relax depends on the environment of the nuclei

relaxation time can give information about nature of materials

➢ there are two processes through with nuclei could relax

- *spinlattice relaxation*, characterized by $T_1$ relaxation

  nuclei flip so that spin orientations return to equilibrium orientation in external field $B_0$

  energy of spinning nuclei is transferred to lattice of surrounding atoms

- *spinspin relaxation*, characterized by $T_2$ relaxation

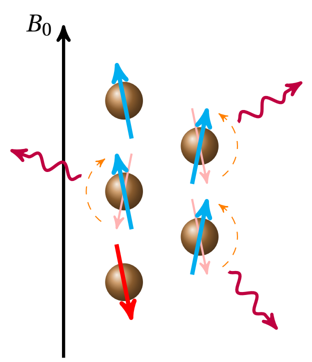
  component of spin transverse to $B_0$ would decay exponentially to its equilibrium state

  energy of spinning nuclei is transferred to other spinning nuclei

different tissues (e.g., water and fat) have distinct $T_1$ and $T_2$ relaxation times

hence various tissues can be identified based on combined information of $T_1$ and $T_2$

the whole idea of nuclear magnetic resonance and relaxation process is illustrated below




realignment of nuclei under strong external field $B_0$

magnetic resonance when radio pulses $B_{\mathrm{RF}}$ are applied

RF photons are re-emitted during de-excitation of nuclei

### 13.3.2 magnetic resonance imaging

procedures of magnetic resonance imaging (MRI) are summarised as the following:

large superconducting electromagnets are applied to produce *strong uniform field* $B_0$

nuclei then align parallel or anti-parallel to the field and *precess* at *Larmor frequency*

a set of *gradient coils* generate a *non-uniform* field that superimposes with $B_0$

this creates small variation in $B_0$ at different parts of the patient's body, so nuclei at different positions have slightly different precession frequencies

*RF coils* transmit *radio pulses* $B_{RF}$ at Larmor frequency to excite nuclei

this causes *magnetic resonance* as nuclei absorb energy from $B_{RF}$

nuclei then *de-excite*, re-emitted signals are detected by another set of RF coils

information is sent to a powerful computer, which processes data and builds up the image

*nature* of the substance is determined based on *duration of relaxation*

*location* of nuclei is be determined precisely based on the differences in *resonant frequencies*

- advantages of MRI
  - uses radiation in radio frequency, so no known biological hazards
  - good contrast between different soft tissues
  - produces 3D images, 2D slice image of any cross section can be obtained easily
- disadvantages of MRI
  - very expensive (superconducting coils must work at extremely low temperatures)
  - not suitable for patients with metallic implants, prostheses, heart pacemakers, etc.
  - patient must lie very still during MRI procedures
  - bones do not show up clearly on an MRI image (can combine MRI scan with CT)

**Question 13.12** State the main components of an MRI scanner and describe their functions.

## 13.4 PET

### 13.4.1 positron

**positron**, also called *anti-electron*, is the *anti-particle* of electron

➢ mass of a positron is same as mass of an electron ($m_e = 9.11 \times 10^{-31}$ kg)

➢ charge of a positron is opposite to charge of an electron ($+e = 1.60 \times 10^{-19}$ C)

➢ positrons can be produced using radioactive $\beta^+$-emitters

$$ {}^{A}_{Z}X \longrightarrow {}^{A}_{Z-1}Y + {}^{0}_{+1}\beta^+ + {}^{0}_{0}\nu $$

an isotope that is commonly used is fluorine-18: ${}^{18}_{9}\mathrm{F} \longrightarrow {}^{18}_{8}\mathrm{O} + {}^{0}_{+1}\beta^+ + {}^{0}_{0}\nu$

fluorine-18 is prepared by bombarding oxygen-18 with protons[89]: ${}^{18}_{8}\mathrm{O} + {}^{1}_{1}\mathrm{p} \longrightarrow {}^{18}_{9}\mathrm{F} + {}^{1}_{0}\mathrm{n}$

**Question 13.13** Another isotope of oxygen, oxygen-15 (${}^{15}_{8}\mathrm{O}$), can undergo decay via positron emission. Write down the equation representing this decay process.

### 13.4.2 electron–positron annihilation

when a positron meets up with an electron, they *annihilate* and produce two $\gamma$-photons[90]

$$ e^+ + e^- \longrightarrow \gamma + \gamma $$

➢ mass-energy is conserved during annihilation process

mass-energy of electron-positron pair transforms into electromagnetic energy of $\gamma$-photons

so energy of each $\gamma$-photon: $E_\gamma = m_e c^2 = 9.11 \times 10^{-31} \times (3.00 \times 10^8)^2 \approx 8.2 \times 10^{-14}$ J

using $E = hf$, we find frequency of each $\gamma$-photon: $f = \frac{E}{h} = \frac{8.2 \times 10^{-14}}{6.63 \times 10^{-34}} \approx 1.2 \times 10^{20}$ Hz

➢ total momentum is also conserved during annihilation

K.E. of electron-positron pair is usually negligible, so zero initial momentum

by conservation of momentum, total final momentum of the two $\gamma$-photons must be zero

so the two $\gamma$-photons carry equal but opposite momentum

- equal momentum means the two $\gamma$-photons have same energy

  so they have same frequency and same wavelength

[89] To send protons towards the target nuclei, we can accelerate them using a *cyclotron* (see §8.3.4).

[90] In case you have forgotten what is meant by a photon, go back and review §11.2.2 and §11.2.4.

– opposite momentum means the two $\gamma$-photons travel in opposite directions
so they move off from site of annihilation at 180° to each other

**Question 13.14** The annihilation of a proton and an antiproton has been observed by particle physicists. Find the energy released during the process of protonantiproton annihilation.

**Question 13.15** When an electron and a positron annihilate, (a) what is the wavelength of the $\gamma$-photon produced? (b) What is the momentum of the $\gamma$-photon produced?

### 13.4.3 positron emission tomography (PET)

for a PET scan, a small amount of tracer is injected into patient's body

**tracers** are molecules labelled with $\beta^+$-emitting nuclides (e.g., fluorine-18[91])

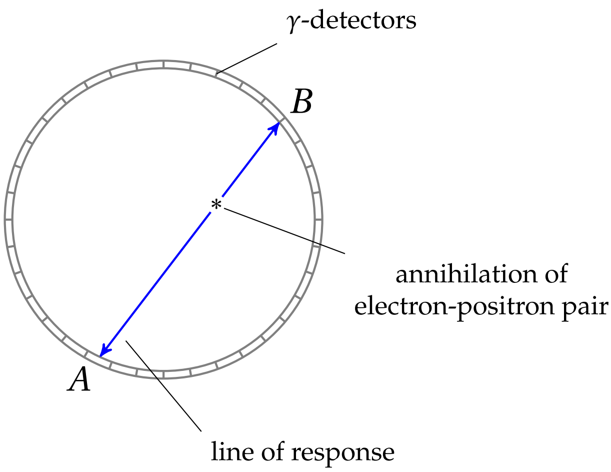
tracers travel around human body and accumulate at certain organs and tissues[92]

positrons are produced through $\beta^+$-decay of tracers

these positrons immediately annihilate with nearby electrons

two $\gamma$-photons are produced and move in opposite directions

the $\gamma$-photon pair are captured by a ring of detectors




**line of response** can be established

*difference in arrival time* can be used to find exact *location* of the annihilation event

*number of events per unit time* detected at one place can further tell tracer *concentration*

[91]The fluorine-18 nuclide can bind to a specific glucose based molecule, forming fluorodeoxyglucose (FDG, chemical formula $C_6H_{11}{}^{18}FO5$), which is one of the most common tracers.

[92]In particular, the tracers tend to accumulate in tumors. Cancer cells are more metabolically active than normal cells. They absorb the tracers at a higher rate, and therefore emit positrons at a higher rate. This effect can be seen on PET scans, allowing the doctor to detect caner at early stages or examine the effects of cancer therapy.

**Example 13.6** Two $\gamma$-photons are produced from the annihilation of an electron-positron pair, and they hit the detectors $A$ and $B$. Given that the two detectors are separated by 80 cm, and the arrival of the $\gamma$-photon at detector $A$ comes 1.2 ns later than the signal picked up at detector $B$. On which point along the line joining $A$ and $B$ does the annihilation event occur?

✍ the two $\gamma$-photons cover a total distance:

$$d = d_A + d_B = 80 \text{ cm}$$

but one photon covers a greater distance than the other:

$$d_A - d_B = c\Delta t = 3.00 \times 10^8 \times 1.2 \times 10^{-9} = 0.36 \text{ m} = 36 \text{ cm}$$

solving the two equations, we find: $d_A = 58$ cm and $d_B = 22$ cm □

# Chapter 14
# Astrophysics

## 14.1 astronomical distances

### 14.1.1 unit of length in astronomy

distances between stars and galaxies in the universe are very large

so it is more convenient to introduce alternative units of measurement for lengths

➢ astronomical unit (AU)

an astronomical unit is the average distance between the Earth and the Sun

$$1\ \text{AU} \approx 1.50 \times 10^{11}\ \text{m}$$

➢ light-year (ly)

a light-year is the distance that light can travel in one year

$$1\ \text{ly} = 3.00 \times 10^{8}\ \text{m s}^{-1} \times 365 \times 24 \times 3600\ \text{s} \approx 9.46 \times 10^{15}\ \text{m}$$

➢ parsec (pc) (∗)

one parsec is the distance at which 1 AU subtends an angle of 1 arcsec or $\dfrac{1^\circ}{3600}$

$$1\ \text{pc} = \frac{1.50 \times 10^{11}\ \text{m}}{\frac{1^\circ}{3600} \times \frac{2\pi}{360^\circ}} \approx 3.09 \times 10^{16}\ \text{m}$$

### 14.1.2 luminosity

**luminosity** (L) of an object is the total power emitted by this object

➢ unit for luminosity: $[L] = \text{W} = \text{J s}^{-1}$

➢ luminosity of a star depends on its surface temperature and its size

- temperature $T \nearrow \Rightarrow L \nearrow$
- stellar radius $R \nearrow \Rightarrow L \nearrow$

### 14.1.3 radiant flux intensity

**radiant flux intensity** ($F$) is defined as the radiant power passing through a surface per unit area at right angles

- unit for radiant flux intensity: $[F] = \mathrm{W\,m^{-2}}$
- $F$ measures *observed* intensity, i.e., $F$ gives *apparent brightness* of the object
- $F$ depends on distance from the emitting source

  if radiation through space is uniform with no absorption, then: $\boxed{F = \dfrac{L}{4\pi d^2}}$

**Example 14.1** The luminosity of the sun is approxiamately $3.86 \times 10^{26}$ W. Given that the distance between the earth and the sun is about $1.5 \times 10^{11}$ m, what is the theoretical power of a space probe that has a solar pane of an area of 40 m$^2$?

✍ radiant flux intensity: $F = \dfrac{L}{4\pi d^2} = \dfrac{3.86 \times 10^{26}}{4\pi \times (1.5 \times 10^{11})^2} \approx 1370\ \mathrm{W\,m^{-2}}$

power input: $P = FA = 1370 \times 40 \approx 5.46 \times 10^4$ W □

**Example 14.2** The radiant flux intensity due to the sun at the position of the earth is about 2.3 times greater than that at the position of the Mars. Given that the distance between the earth and the sun is about $1.5 \times 10^{11}$ m, how far is the Mars from the sun?

✍ $F \propto \dfrac{1}{d^2} \quad \Rightarrow \quad \dfrac{F_E}{F_M} = \dfrac{d_M^2}{d_E^2} \quad \Rightarrow \quad \dfrac{d_M^2}{d_E^2} = 2.3 \quad \Rightarrow \quad d_M = \sqrt{2.3} \times 1.5 \times 10^{11} \approx 2.3 \times 10^{11}$ m □

### 14.1.4 standard candles

astronomical objects of known luminosity are called **standard candles**

for a particular star, say $X$, we can look for a standard candle in the same galaxy

radiant flux intensity $F$ of the standard candle can be measured, and its luminosity $L$ is also known, so its distance can be found using: $d = \sqrt{\dfrac{L}{4\pi F}}$

distance of star $X$ is about same as distance of standard candle, as they are in same galaxy

this is how we determine the distance of stars by the use of *standard candles* among the many types of standard candles you need to know two in this course:

- cepheid variables
- supernovae

**cepheid variables**

cepheid variables are stellar objects whose brightness varies periodically[93]

once we know period of variation in brightness, luminosity of the star can be predicted

**Example 14.3** One cepheid variable has a radiuant flux intensity of $1.4 \times 10^{-16}$ W m$^{-2}$. It is known to have a luminosity of $1.0 \times 10^{30}$ W. Find the distance of this star in light-years.

✍ $F = \dfrac{L}{4\pi d^2} \quad \Rightarrow \quad d = \sqrt{\dfrac{L}{4\pi F}} = \sqrt{\dfrac{1.0 \times 10^{30}}{4\pi \times 1.4 \times 10^{-16}}} \approx 2.4 \times 10^{22}\ \text{m} = \dfrac{2.4 \times 10^{22}}{9.46 \times 10^{15}}\ \text{ly} \approx 2.5 \times 10^{6}\ \text{ly}$ □

**supernovae**

luminosity of supernovae is known to be constant[94]

➢ supernovae are extremely luminous, so suitable for measuring very large distances

[93]The variation in the stellar brightness is related to the mass of the cepheid variable. When the star contracts, its core temperature increases, this boosts the rate of nuclear reaction so radiation pressure increases to push the stellar matter outwards against the gravity. But as the star expands, its temperature drops so its radiation pressure lowers, the gravitational attraction would tend to pull the stellar matter inwards. The period of this feedback loop between the gravitational pull and the nuclear reaction taking place in the star depends on the mass of the star, and the mass of the star determines a number of the star's features including its luminosity. Therefore, if we know the period of the brightness fluctuations, we can know about the star's mass and hence its luminosity.

[94]To be more precise, Type Ia supernovae explosions are found to have a consistent luminosity. These events are believed to occur when a white dwarf accretes matter until its mass exceeds a critical value, known as the Chadellaskar limit. By then its gravitational attraction can no longer be overcome by the degeneracy pressure, the star collapses and explodes. There is also Type II supernovae explosions, these occur when a star of very large mass runs out of its nuclear fuel. Radiation pressure cannot be maintained, so the star collapses rapidly towards its centre under gravity.

**Question 14.1** (a) Type 1a supernovae have luminosities at the order of $10^{36}$ W (approximately 5 billions of the luminosity of the sun). If the sun undergoes a supernova explosion, find the radiant flux intensity measured on earth. (b) It is estimated that the Hiroshima bomb and the Nagasaki bomb (the two nuclear bombs dropped on Japan during the World War II) released a total energy of around 100 TJ. Suppose a similar nuclear bomb completely explodes at a distance of 50 m from your home in a period of one second, what radiant flux intensity do you measure? (c) Compare the two results, which event would appear brighter?

## 14.2 spectrum of stars

any object at a particular temperature $T$ emits electromagnetic radiation

radiation emitted by the object spreads into a continuous range of wavelengths $\lambda$

total intensity and typical wavelength emitted depend influentially on temperature $T$

for example, human body ($\sim 300$ K) emits infra-red, cooking fire ($\sim 1000$ K) emits visible red

stars also emit radiation with a range of wavelengths, giving rise to a characteristic spectrum

we can learn a lot about the properties of the star by looking at its spectrum

### 14.2.1 black body radiation

theoretically, stars can be modelled as black body emitters, whose emission spectrum at a particular temperature $T$ can be precisely predicted

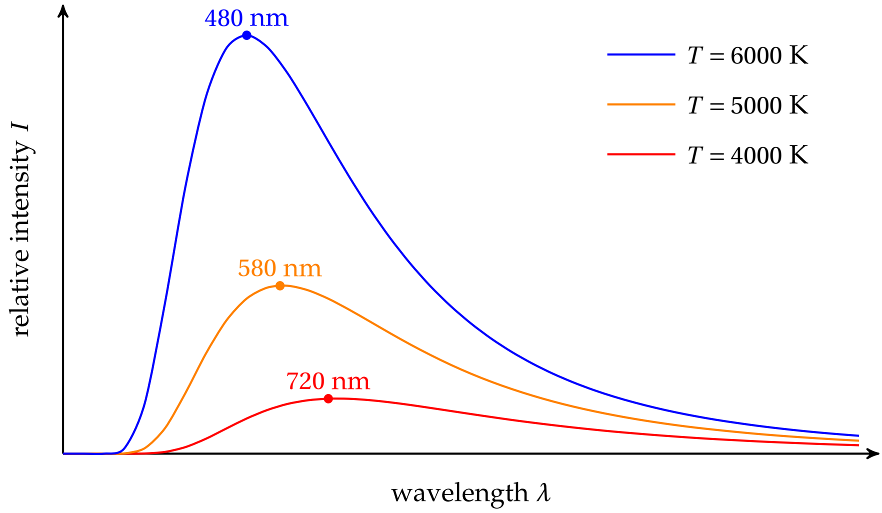
> a **black body** is an idealised object that absorbs all incident radiation

➢ intensity spectrum of black body radiation has the following features:

- for given temperature, intensity tends to zero at very short or very long wavelengths
- as temperature increases, peak intensity occurs at shorter wavelength ($T \nearrow \Rightarrow \lambda_{\max} \searrow$)
- for any $\lambda$, intensity increases with temperature ($T \nearrow \Rightarrow I \nearrow$)

➢ **Wien's law** states that wavelength of greatest intensity and surface temperature satisfies:

$$\boxed{\lambda_{\max} T = b} \qquad \text{where } \textit{Wien's displacement constant } b \approx 2.898 \times 10^{-3} \text{ m K}$$




radiation spectrum of an ideal black body

➢ **Stefan-Boltzmann law** states that total power emitted from a black body of surface area $A$ and surface temperature $T$ is given by

$\boxed{L = \sigma A T^4}$ where *Stefan-Boltzmann constant* $\sigma \approx 5.67 \times 10^{-8}$ W m$^{-2}$ K$^{-4}$

for a star of radius $R$, its surface area is $A = 4\pi R^2$

then total radiant power, i.e., luminosity of a star is given by: $\boxed{L = 4\pi\sigma R^2 T^4}$

**Question 14.2** Use the values labelled on the intensity-wavelength curves on top of this page, verify that the wavelength at which the maximum radiation intensity occur satisfies Wien's law at any one of the three given temperatures.

### 14.2.2 surface temperature

from radiation spectrum of a star, we can identify wavelength at peak intensity

then surface temperature of the star can be determined using Wien's law

**Example 14.4** The wavelength for which the maximum rate of emission from the redgiant Betelgeuse is 878 nm. What is the surface temperature of the star?

✍ $$T = \frac{b}{\lambda_{\text{max}}} = \frac{2.898 \times 10^{-3}}{878 \times 10^{-9}} \approx 3300 \text{ K}$$ □

**Example 14.5** The surface temperature of the sun is about 5800 K. What is the wavelength of the peak radiation ouput in the solar spectrum?

✍ $$\lambda_{\max} = \frac{b}{T} = \frac{2.898 \times 10^{-3}}{5800} \approx 5.0 \times 10^{-7}\ \text{m} \approx 500\ \text{nm}$$

this wavelength lies in the range of the visible spectrum (380 ~ 740 nm)

our calculation indicates a large portion of solar radiation is visible light □

**Question 14.3** It is difficult to see a human body or animals with naked eyes in a very dark environment, but night vision is possible using infra-red cameras. Explain why.

### 14.2.3 stellar radii

take a standard candle in a distant galaxy, we can calculate distance $d$ of the galaxy

from $d$ and radiant flux intensity $F$, we can find luminosity of the star: $L = 4\pi d^2 F$

from a star's spectrum, we can find surface temperature $T$ of the star using: $\lambda_{\max} T = b$

once we find $L$ and $T$, we can calculate the stellar radius $R$ using: $L = 4\pi\sigma R^2 T^4$

**Example 14.6** Aldebaran, the brightest star in the constellation Taurus, has a luminosity of $1.8 \times 10^{29}$ W. The wavelength at the peak radiation intensity of its spectrum is found to be 740 nm. What is the radius of the star?

✍ by Wien's law $\lambda_{\max} T = b$, we find surface temperature to be

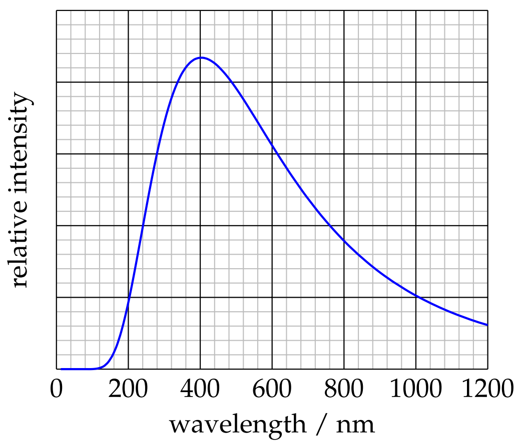
$$T = \frac{b}{\lambda_{\max}} = \frac{2.898 \times 10^{-3}}{740 \times 10^{-9}} \approx 3920\ \text{K}$$

by Stefan-Boltzmann law $L = 4\pi\sigma R^2 T^4$, so

$$1.8 \times 10^{29} = 4\pi \times 5.67 \times 10^{-8} \times 3920^4 \times R^2 \quad \Rightarrow \quad R \approx 3.3 \times 10^{10}\ \text{m}$$

Aldebaran is about 50 times larger than the solar radius (around $7.0 \times 10^8$ m), and its appearance is red (spectrum peaked at near infra-red), this makes it a typical *red giant* star □

**Question 14.4** Altair, the brightest star in the constellation Aquila, is at a distance of 16.7 light years from the earth. The radiant flux intensity due to Altair measured on the earth is $1.3 \times 10^{-8}\ \text{W m}^{-2}$. The intensity-wavelength curve of Altair is shown. (a) Find the luminosity of Altair. (b) Find the surface temperature of Altair. (c) Find the radius of the star.



## 14.3 cosmology

### 14.3.1 quick review: absorption spectrum

hot interior of star emits a continuous range of wavelengths

as light passes cooler outer layers of the star, photons of right energies can be absorbed

the absorption depends on the elements present in the star's outer layers

this gives rise to a characteristic *absorption spectrum* for the star[95]

the spectrum would consist of dark lines at specific wavelengths or frequencies[96]

### 14.3.2 quick review: Doppler effect

but the spectral lines observed by us will be different from where they are supposed to be

observed frequency/wavelength changes due to relative motion between star and observer

this phenomena is known as the *Doppler effect*

➢ changes in spectral lines can tell whether a star/galaxy is *receding* or *approaching*

- if wave source moves away, observed wavelength increases / frequency decreases
- if wave source moves closer, observed wavelength decreases / frequency increases

➢ changes in spectral lines can also tell how fast the star is moving

you might recall observed frequency is given by: $f = f_0 \frac{c}{c \mp v}$

this formula would not be particularly useful in this part of the course

we will derive another formula for Doppler shift in the following section

➢ Doppler shift can only tell *radial* velocity of a star/galaxy

Doppler method cannot tell about motion in a dirction that is normal to line of sight

[95]The chemical composition of most stars is mainly hydrogen, plus a small fraction of helium, and trace amounts of heavier elements such as carbon, oxygen, etc. The presence of each element gives rise to a unique set of absorption lines.

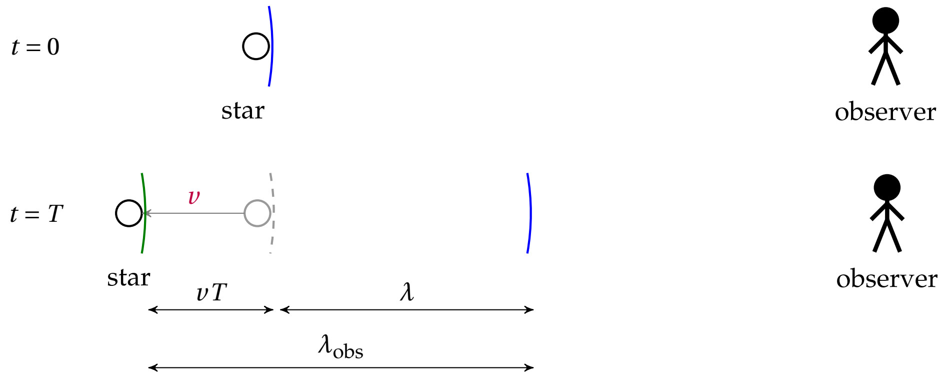
[96]More detailed discussions on the absorption spectrum has been given in §11.4.3.

### 14.3.3 redshift

data show that spectral lines of most galaxies shift in direction of increasing wavelength

for visible spectrum, the lines shift towards red end, so this is called **redshift**

redshift of line spectrum means most galaxies are moving away




for a star/galaxy recedes at velocity $v$, it travels a distance of $vT$ in one period

this causes an increase in observed wavelength: $\Delta\lambda = \lambda_{\text{obs}} - \lambda = vT$

so fractional change in wavelength is: $\dfrac{\Delta\lambda}{\lambda} = \dfrac{vT}{\lambda} = \dfrac{vT}{cT} \quad \Rightarrow \quad \boxed{z \equiv \dfrac{\Delta\lambda}{\lambda} = \dfrac{v}{c}}$

the parameter $z$ is usually called the *redshift factor*

### 14.3.4 recession of galaxies

*Edwin Hubble* discovered in 1929 that almost all galaxies are moving away

further analysis showed recession speed of a galaxy is roughly proportional to its distance

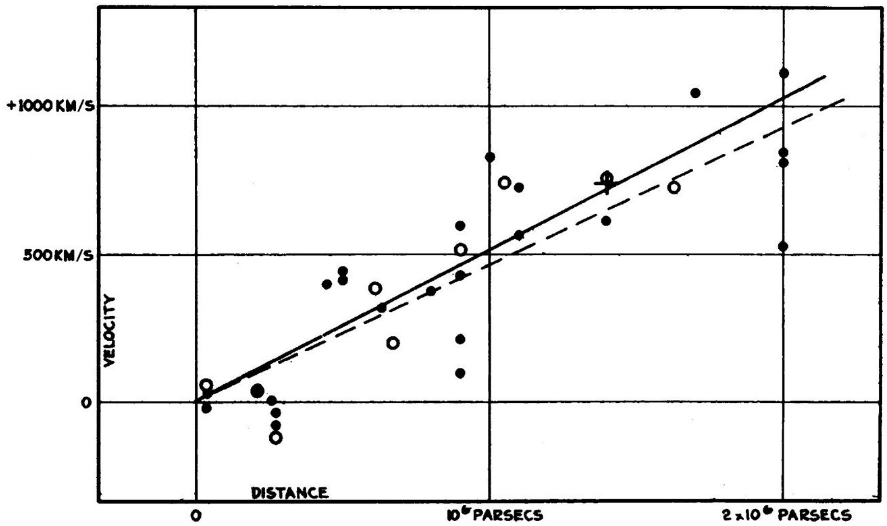
this is known as **Hubble's law**[97]

$$\boxed{v = H_0 d} \qquad \text{Hubble constant: } H_0 \approx 2.4 \times 10^{-18}\ \text{s}^{-1}$$

➢ exact value of the Hubble constant is not accurately known

$H_0$ is believed to be between $1.5 \times 10^{-18}\ \text{s}^{-1} \sim 3.5 \times 10^{-18}\ \text{s}^{-1}$ (uncertainty around ±40%!)

[97]Edwin Hubble made this conclusion based on his study of 46 galaxies. Later observations further supported the proportional relationship between the recession speed of galaxies and the distance, but the constant computed by Edwin Hubble himself was found to be greatly overestimated. Hubble's original paper can be seen here: https://www.pnas.org/doi/full/10.1073/pnas.15.3.168




Edwin Hubble's plot of the velocity-distance relation among galaxies in his original paper

**Example 14.7** The K-line of inoised calcium from the absorption spectrum of the galaxy NGC 4889 is measured to be 401.8 nm. The same K-line measured in a laboratory setting is known to be 393.3 nm. (a) What is the speed of the NGC 4889? (b) How long does it take for light to reach us from NGC 4889?

✍ recession speed: $v = \dfrac{\Delta\lambda}{\lambda} \times c = \dfrac{401.8 - 393.3}{393.3} \times 3.00 \times 10^8 \approx 6.48 \times 10^6\ \text{m s}^{-1}$

distance of galaxy: $d = \dfrac{v}{H_0} \approx \dfrac{6.48 \times 10^6}{2.4 \times 10^{-18}} \approx 2.7 \times 10^{24}\ \text{m}$

time for light to propagate: $t = \dfrac{d}{c} = \dfrac{2.7 \times 10^{24}}{3.00 \times 10^8} \approx 9.0 \times 10^{15}\ \text{s} \approx 2.9 \times 10^8$ years

### 14.3.5 Big Bang theory

Hubble's law implies the universe is *expanding*

go back in time, all galaxies must be close together

this suggests the universe must begin from a highly dense state

it is now generally accepted that the universe is created from a dramatic explosion

this starting point of the universe and everything is known as the **Big Bang**[98]

[98]The Big Bang Theory was proposed in 1927 by Belgian astronomer and cosmologist *Georges Lemaitre*. From a theoretical view, Lemaitre deduced that the natural consequence of Einstein's theory of general relativity must be an expanding universe. In 1929, *Edwin Hubble*, who was unaware of Lemaitre's work,

**age of the universe**

age of the universe can be deduced from Hubble's law

consider two galaxies started off close together during the Big Bang

assume galaxies have been moving away from one another at constant speed $v$ for time $t$

distance between the two galaxies today would be $d = vt$

but this $t$ is essentially all the time since beginning of the universe

so age of the universe: $t = \dfrac{d}{v} = \dfrac{d}{H_0 d} \quad \Rightarrow \quad \boxed{t = \dfrac{1}{H_0}}$

substitute $H_0 \approx 2.4 \times 10^{-18}\ \text{s}^{-1}$, one finds $t \approx 4.5 \times 10^{18}$ s, which is about 14 billion years (Gy)

➢ there is great uncertainty in the universe's age (varying from 11 ~ 20 Gy)

two main reasons for this uncertainty are:

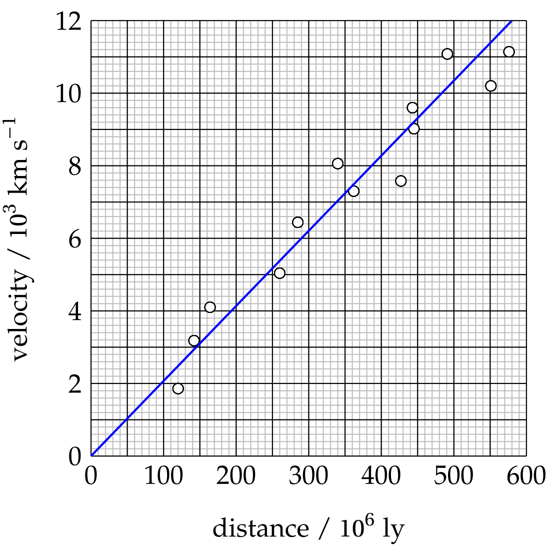
- our calculation is based on the assumption of constant recession speed

  there are cosmological models that suggest otherwise, so different ages are predicted

- our calculation also depends on the value of $H_0$

  but exact value of $H_0$ is in doubt due to large uncertainty in distance measurements

**Question 14.5** The plot shows some data on the velocities and distances for a set of galaxies. A fit line has also been drawn. (a) Compute a value for the Hubble constant $H_0$ based on the data given. (b) Estimate the age of the universe using your calculated value for $H_0$, and state any assumption that you make. (c) For a galaxy at a distance of $2 \times 10^9$ ly, what is the fractional change in the wavelength of its spectrum?




published his famous relation between distance and radial velocity among extra-galactic nebulae. Nevertheless, Hubble's discovery became a strong supporting evidence for the Big Bang theory.

**cosmic microwave background radiation**

another key evidence for the Big Bang model is the *cosmic microwave background*

after Big Bang, universe gradually cools down due to expansion

the universe today should be filled with radiation due to remnant heat from the Big Bang

theoretical calculation predicts that the universe should now have a temperature of 2.7 K

this suggests a background radiation peaked at around 1 mm (microwave band)

this radiation is therefore called **cosmic microwave background** (CMB)[99]

CMB radiation have been observed by space telescopes from all directions of cosmos[100]

CMB corresponds to a temperature of around 3 K, which is pretty close to the prediction

**Question 14.6** Use Wien's law to show that the peak intensity of the electromagnetic radiation given off by a black body at a temperature of 3 K is around 1 mm.

[99] *Ralph Alpher*, *George Gamow* and *Robert Herman* suggested the idea of CMB in theory in the 1940s. In 1965, *Robert Dicke* took up the problem and led a team of physicists at Princeton University to hunt for the CMB signal. In the same year, *Arno Penzias* and *Robert Wilson* at the Bell Laboratories built a large radio antenna for a communication satellite in New Jersey, while they failed in all attempts to get rid of the 'background noise' that came from all directions.

When Princeton's team heard about the Bell Lab's result, they realized at once that the CMB had already been found. As a result, two important articles were published: one by Penzias and Wilson about the 'noise', the other by Dicke's team explaining its true nature. Penzias and Wilson did not interpret what they had found in their article, nor did they have any intention to look for CMB in the first place. Nevertheless, they received the Nobel Prize in Physics in 1978, for their accidental discovery of CMB. What about Dicke and his team? Quoting *Bill Bryson's A Short History of Nearly Everything*, the Princeton researchers got only sympathy.

[100] Three major CMB missions include: (1) the Cosmic Background Explorer (COBE), launched by NASA in 1989, (2) the Wilkinson Microwave Anisotropy Probe (WMAP), launched by NASA in 2001 to study fluctuations in CMB in greater detail, (3) the Planck satellite, launched by ESA in 2009 to study CMB fluctuations with even greater sensitivity. These studies would allow scientists to trace back to what happened immediately after the Big Bang, and the understanding of the early universe is crucial for the understanding of the large-scale cosmic structures that we see today.

#### 14.3.6 fate of the universe (∗)

whether the universe will expand forever depends on strength of gravity

fate of universe narrows down to how density $\rho$ of universe compares to critical density $\rho_c$

- if $\rho > \rho_c$, force of gravity dominates as there is sufficient matter

  expansion will slow down, then contract back inwards

  this leads to a *closed* universe, eventually ending in a *Big Crunch*
- if $\rho < \rho_c$, there is not enough gravity to stop expansion

  expansion continues forever, leading to an *open* universe

but the big problem is, it is difficult to determine average density of our universe

➢ issue of *dark matter*

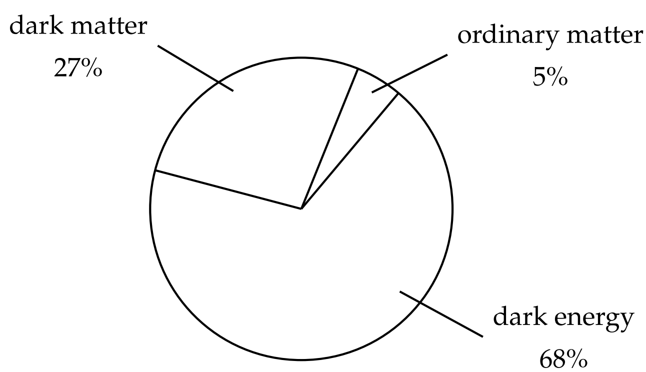
mass of galaxies suggested by luminosity calculation cannot provide the centripetal force needed to keep galaxies rotating

galaxies must contain mass that does not give off light, known as **dark matter**

➢ issue of *dark energy*

observation shows expansion of universe is not slowing down

it is suggested that cosmic acceleration is caused by **dark energy**




➢ it turns out ordinary matter only contributes to a very small part of the universe

great majority of the total mass-energy content of cosmos is dark matter and dark energy

but nature of dark matter and dark energy are not yet understood

# Chapter 15
# Electronics (∗)

electronic sensors basically consist of three parts

change in some physical property (temperature, pressure, etc.) is picked up by a **sensor**

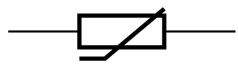
electrical signal is then fed to **processing units**

processor produces an output voltage to drive **output devices** to indicate this change

## 15.1 sensors

### 15.1.1 thermistor

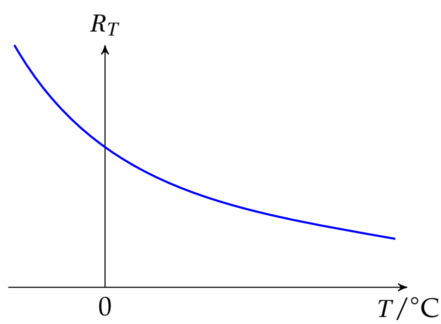
**thermistor** is an electrical component with a resistance that varies with temperature




electric symbol for thermistor

thermistors usually have a *negative temperature coefficient*, or NTC in short, which means their resistance decreases as temperature rises

in A-Levels, we consider NTC thermistors only




➢ resistance-temperature relation is *non-linear*

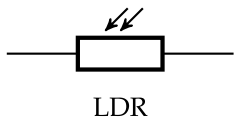
if a graph is sketched, the curve looks like an exponential decay

➢ behaviour of NTC thermistors can be explained in terms of *band theory* (see §11.5.2)

at low temperature, electrons are bound to atoms and cannot move freely; at high temperature, electrons gain energy and break free from atoms, so more free electrons available to conduct electricity

### 15.1.2 LDR

**light-dependent resistor**, or **LDR**, is an electrical component with a resistance that can change with intensity of light

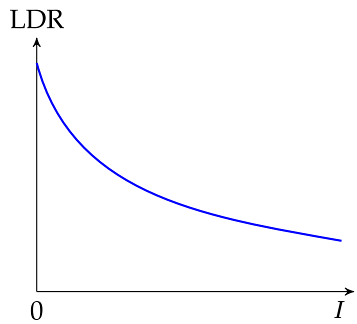



electric symbol for LDR

resistance of LDR decreases as light level increases

- resistance of LDR varies *non-linearly* with intensity
- behaviour of LDR is also explained with *band theory*

bound electrons acquire energy from radiation and become conducting, greater light intensity means more free electrons available



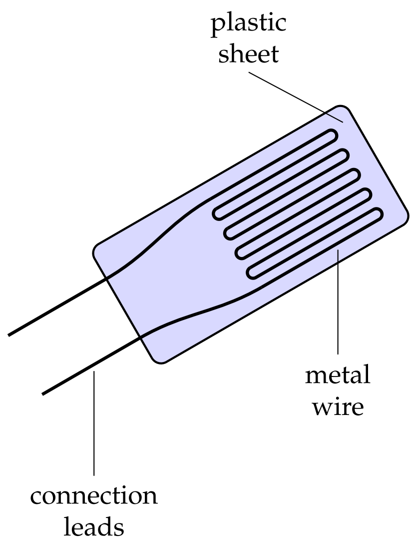

- in dark, typical resistance of LDR $\approx 10^5 \sim 10^7\ \Omega$

in bright, typical resistance of LDR $\approx 10^2 \sim 10^3\ \Omega$

### 15.1.3 strain gauge

to measure width of a narrow crack, a device called **strain gauge** is widely used in industry

a metal-wire strain gauge consists of a folded long wire sealed on a thin plastic box




structure of a strain gauge

as crack widens, strain gauge gets stretched, length of wire increases, causing a change in its resistance

resistance of a metal wire is given by: $R = \dfrac{\rho l}{A}$

increase in resistance is proportional to increase in length of wire (for small increments)

- if change in cross-sectional area of wire is negligible

increase in resistance: $\Delta R = \dfrac{\rho \Delta l}{A} \Rightarrow \boxed{\dfrac{\Delta R}{R} = \dfrac{\Delta l}{l}}$

- if cross-sectional area of wire changes with its length

total volume of wire $V = Al$ should remain constant it can be shown that [101] the percentage increase in resistance is: $\boxed{\frac{\Delta R}{R} = \frac{2\Delta l}{l}}$

notice the differences, change in length causes increase in resistance, reduction in cross section makes the same contribution, hence an extra factor of two.

### 15.1.4 piezo-electric transducer

**piezo-electric transducer** transforms mechanical energy into electrical energy, or vice versa

when a piezo-electric material[102] is compressed, a p.d. is produced across it

conversely, p.d applied across piezo-electric material can cause change in shape

since its discovery, piezo-electricity was exploited in many useful applications

- piezo-electric transducer can be used to generate sound waves

  apply an a.c. voltage, material is forced to vibrated at same frequency as a.c.

  this causes compression and rarefaction of air, producing sound waves

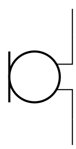
  so piezo-electric transducer is the key component in *loudspeakers* and *buzzers*

- piezo-electric transducer can also be used to detect sound waves

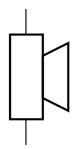
  exposed to sounds, variation in air pressure produces an a.c. voltage across the device

  this signal can be measured and processed to represent original sound wave

  so piezo-electric transducer is also found in *microphones*



symbol for a microphone



symbol for a loudspeaker

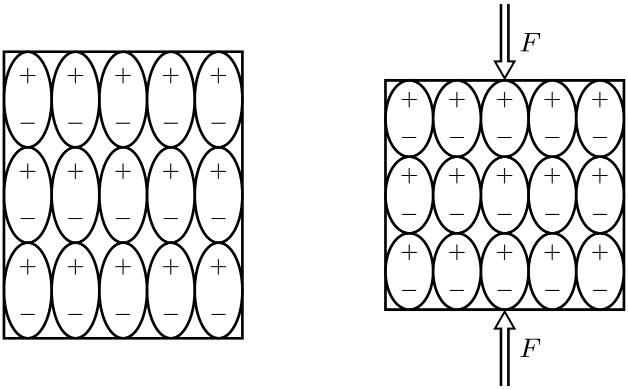
- mechanism of piezo-electricity is explained on molecular level

[101]To keep $V = Al$ constant, for small change in $l$, one has $\frac{\Delta A}{A} \approx \frac{\Delta l}{l}$. Resistance of the stretched wire is: $R + \Delta R = \frac{\rho(l+\Delta l)}{A - \Delta A} = \frac{\rho l}{A}\left(1+\frac{\Delta l}{l}\right)\left(1-\frac{\Delta A}{A}\right)^{-1} \approx R\left(1+\frac{\Delta l}{l}\right)\left(1-\frac{\Delta l}{l}\right)^{-1}$. For $x \ll 1$, we can use binomial expansion $(1+x)^n = 1 + nx + \cdots$. Drop higher order terms, we find to the first-order approximation that: $R + \Delta R \approx R\left(1+\frac{\Delta l}{l}\right)\left(1+\frac{\Delta l}{l}\right) \approx R\left(1+\frac{2\Delta l}{l}\right)$, so $\Delta R = 2R\frac{\Delta l}{l}$, it then follows that $\frac{\Delta R}{R} = \frac{2\Delta l}{l}$.

[102]Materials that exhibit piezo-electricity include quartz, certain ceramics, various proteins, etc.

piezo-electric materials consist of *polarised* molecules, also called *dipoles*

when being compressed, dipoles rearrange themselves, causing variation of charge density, so small voltage is generated (see illustration)



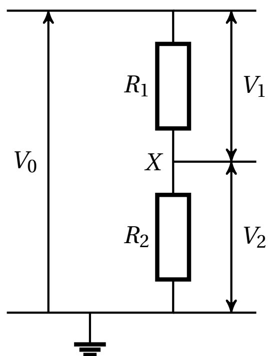

## 15.2 processors

### 15.2.1 potential divider

sensors such as thermistor and LDR only produce changes in resistance, but output devices are driven by voltages ⟶ **potential divider** circuit can be used

a potential divider consists of two resistors in series connected to a fixed voltage supply

voltage is shared between the resistors: $V_0 = V_1 + V_2$




p.d. across each resistor: $V_1 = IR_1$, $V_2 = IR_2$

but same current $I$ flows through $R_1$ and $R_2$, so $\boxed{\dfrac{V_1}{V_2} = \dfrac{R_1}{R_2}}$

this shows voltage is shared in proportion to resistance

for each resistor: $\boxed{V_1 = \dfrac{R_1}{R_1 + R_2} V_0}$, $\boxed{V_2 = \dfrac{R_2}{R_1 + R_2} V_0}$

➢ if $R_1$ increases, while $R_2$ is fixed, then $V_1 \nearrow$, $V_2 \searrow$

if $R_1$ decreases, while $R_2$ is fixed, then $V_1 \searrow$, $V_2 \nearrow$

➢ sometimes we need specify *electric potential* at a given point

potential at any point is measured with respect to a reference point

this point is called **earth** or **ground**, with a potential defined to be zero

for example, potential at $X$ is essentially equal to voltage $V_2$

**Example 15.1** For the sensing circuit shown, suggest how the output voltage changes as light level increases.

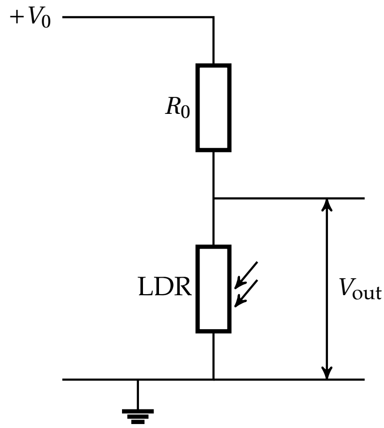
✍ intensity of light increases, resistance of LDR decreases

current in circuit increases, p.d. across $R_0$ increases

so p.d. across LDR becomes smaller, $V_{out}$ would decrease

from $V_{out} = \frac{LDR}{LDR + R_0} V_0$, one can also see $V_{out}$ decreases when resistance of LDR becomes lower

so $V_{out}$ decreases as light intensity increases □




**Question 15.1** In Example 15.1, if the fixed resistor has $R_0 = 8.0$ kΩ, and the LDR is in an environment such that it has a resistance of 12 kΩ. The voltage supply $V_0 = +15$ V. Find $V_{out}$.

**Question 15.2** Design a sensing circuit that produces a greater output voltage as the temperature of the sensor rises. Explain the operation of your circuit.

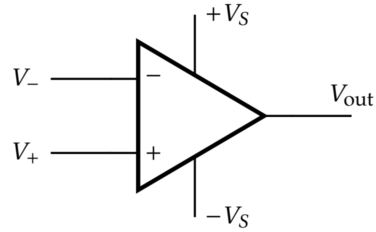
### 15.2.2 op-amp: basics

to amplify small input into large output signals, we need **operational amplifiers (op-amp)**[103]

op-amp is building block for many electronic devices

➢ an op-amp has several terminals:

- $V_+$: non-inverting input
- $V_-$: inverting input
- $\pm V_S$: positive/negative power supply voltages
- $V_{out}$: output voltage




circuit symbol for op-amp

➢ output voltage of op-amp is given by: $\boxed{V_{out} = G_0(V_+ - V_-)}$

$G_0$ is value of amplification of an open-loop[104] amplifier, called **open-loop gain**

[103] We are not concerned with what is inside an operational amplifier. An op-amp contains several *transistors*, and you are not required to know how they work in the A-Level syllabus. You only need to know what output voltage an op-amp would produce to drive output devices. If you want to learn more details on the functions of op-amps, you need to study a higher level course on *Electronics*.

[104] It is called an open loop in the sense that there is no loop linking $V_{out}$ to the input of the op-amp.

➢ value of $G_0$ is not well controlled in manufacturing processes

$G_0$ may also depend on temperature of op-amp, frequency of input signals, etc.

➢ output voltage of op-amp cannot exceed supply voltages

when $V_{\text{out}}$ calculated is greater than $V_S$, the op-amp becomes **saturated**, true $V_{\text{out}}$ of op-amp is close to $+V_S$ or $-V_S$ depending on polarity of $V_{\text{out}}$

➢ all voltages in amplifier circuits are measured with respect to a reference level set at *earth*

**properties of ideal op-amps**

an ideal op-amp is considered to have the following properties

- *infinite open-loop gain* ($G_0 \to \infty$)

  want large $V_{\text{out}}$ for small $V_{\text{in}}$, value of amplification should be as large as possible[105]

- *infinite input resistance/impedance*[106]

  want $V_{\text{in}}$ to op-amp to be as large as possible

  so large resistance between input terminals of op-amp[107]

  a consequence is that (almost) no current flows into or leaves the input terminals

- *zero output resistance/impedance*

  output impedance of op-amp acts like internal resistance of batteries

  want largest possible $V_{\text{out}}$ to load, so small resistance for output terminal[108]

- *infinite bandwidth*

  **bandwidth** refers to range of frequencies of input voltages that are amplified by same amount

  want same gain for all input frequencies, so infinite bandwidth is desired

- *infinite slew rate*: $V_{\text{out}}$ changes immediately with $V_{\text{in}}$, response is instantaneous

- *zero noise*: $V_{\text{out}}$ does not change with disturbance due to external conditions

[105]Open-loop gain of an actual op-amp: $G_0 \approx 10^5 \sim 10^7$.

[106]Impedance is an extension of the concept of resistance to a.c. circuits. You might just think of impedance of an electrical component as its ability to oppose currents.

[107]Typical input impedance of an actual op-amp: $10^6\ \Omega$ or higher.

[108]Typical output impedance of an actual op-amp: $10^1 \sim 10^2\ \Omega$.

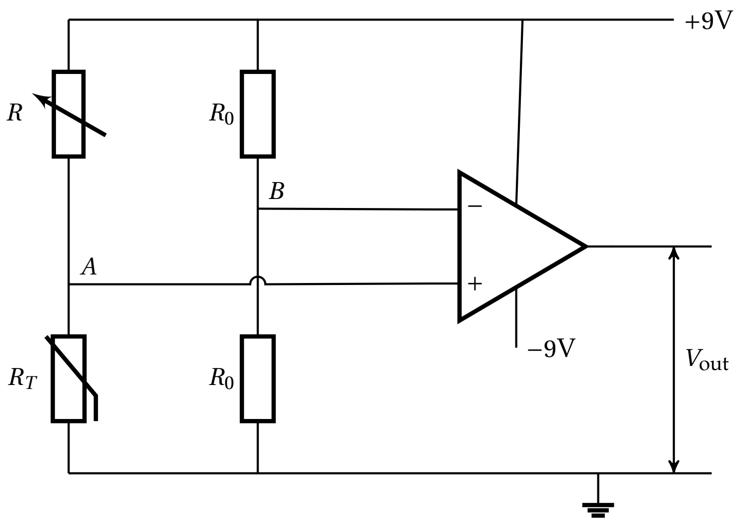
### 15.2.3 voltage comparator

for an open-loop op-amp, since gain $G_0$ is very large, so as long as there is a small difference in $V_+$ and $V_-$, $V_{\text{out}} = G_0(V_+ - V_-) \to \pm\infty$, this causes $V_{\text{out}}$ to saturate

so $V_{\text{out}} = \pm V_S$ depending on which one of $V_+$ and $V_-$ is larger, this makes a comparator

a **voltage comparator** compares two input voltages $V_+$ and $V_-$, and produces a $V_{\text{out}}$ to indicate which one is greater: if $V_+ > V_-$, then $V_{\text{out}} = +V_S$; if $V_+ < V_-$, then $V_{\text{out}} = -V_S$

**Example 15.2** A circuit that incorporates an ideal op-amp is shown. Given that $R_0 = 10$ kΩ, variable resistor is set to $R = 20$ kΩ, and the thermistor has a resistance of $R_T = 40$ kΩ at 20°C, and $R_T = 10$ kΩ at 50°C. Describe the change in $V_{\text{out}}$ as temperature rises from 20°C to 50°C.




✍ this is a comparator whose input voltages are determined from two potential divider circuits

resistor $R_0$ sets a constant potential for reference: $V_- = V_B = \dfrac{10}{10+10} \times (+9) = +4.5$ V

p.d. across $R_T$ determines $V_A$, which is fed into $V_+$ to be compared with $V_-$

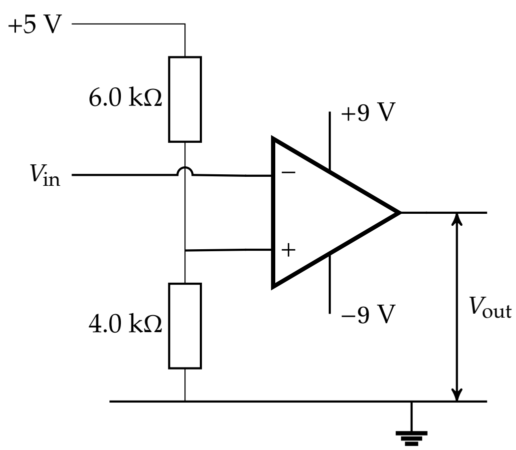
at 20°C: $V_+ = V_A = \dfrac{40}{40+20} \times (+9) = +6.0$ V, so $V_+ > V_- \Rightarrow V_{\text{out}} = +9$ V

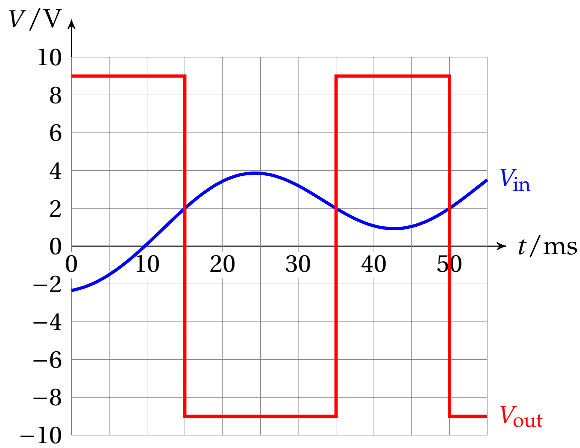
at 50°C, $V_+ = V_A = \dfrac{10}{10+20} \times (+9) = +3.0$ V, so $V_+ < V_- \Rightarrow V_{\text{out}} = -9$ V

as temperature goes from 20°C to 50°C, $V_{\text{out}}$ initially maintains at +9 V

it switches over to −9 V at some critical temperature then stays at −9 V until 50°C □

**Example 15.3** The diagram shows a circuit incorporating an ideal op-amp. The variation with time for input voltage $V_{in}$ is shown. Sketch the variation of the output voltage.




✍ potential at non-inverting input: $V_+ = \dfrac{4.0}{4.0 + 6.0} \times (+5) = +2$ V

potential at inverting input is equal to input voltage: $V_- = V_{in}$

for $V_{in}$ greater than +2 V, $V_- > V_+$, then $V_{out} = -9$ V

for $V_{in}$ less than +2 V, $V_- < V_+$, then $V_{out} = +9$ V

variation of $V_{out}$ can then be plotted, which shows the form of a square wave □

### 15.2.4 negative feedback

in practical uses, we may desire $V_{out}$ to reflect small changes in $V_{in}$, instead of stuck at $\pm V_S$

there are a few problems with open-loop amplifiers that we need to deal with:

- gain of amplifier is too high

  $V_{out}$ almost always saturates, so $V_{out}$ cannot change continuously with $V_{in}$

- open-loop gain $G_0$ of practical op-amp is affected by temperature

  different $V_{out}$ for the same $V_{in}$ at different temperatures is not convenient

- open-loop gain $G_0$ is a function of input frequency

  different input frequencies are not always amplified by same amount

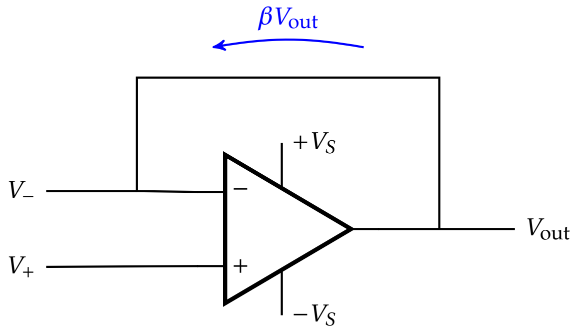
  this would cause distortion of original signal, bandwidth would be limited

one way to improve properties of amplifier circuits is to introduce negative feedback

feedback means sending a portion of output voltage back to the input terminals

a feedback network is also called a *closed loop*

if a fraction of output $V_{\text{out}}$ is sent back to inverting input $V_-$ such that it opposes changes of input, this is called **negative feedback**




negative feedback mechanism

let's look into the effect of introducing negative feedback into the amplifier circuit[109]

assume a fraction of output $\beta V_{\text{out}}$ $(0 < \beta \le 1)$ is fed back to inverting input of op-amp, then combined voltage to this terminal would be the sum of original $V_-$ and the feedback $\beta V_{\text{out}}$, the equation $V_{\text{out}} = G_0(V_+ - V_-)$ for open-loop circuit shall be rewritten, hence we have

$$V_{\text{out}} = G_0\Big[(V_+ - (V_- + \beta V_{\text{out}})\Big] \quad \Rightarrow \quad (1 + G_0\beta)V_{\text{out}} = G_0(V_+ - V_-) \quad \Rightarrow \quad V_{\text{out}} = \frac{G_0}{1 + G_0\beta}(V_+ - V_-)$$

overall gain of negative feedback amplifier is: $G = \dfrac{G_0}{1 + G_0\beta}$

this gain is less than open-loop gain $G_0$, so gain is reduced

further notice that $G_0\beta \gg 1$, we find $G \approx \dfrac{G_0}{\beta G_0} = \dfrac{1}{\beta}$

this shows for negative-feedback amplifier, as long as open-loop gain $G_0$ is sufficiently large, overall gain $G$ depends on $\beta$ only, i.e., only depends on how much output is sent back to $V_-$

$\beta$ can be controlled by a set of fixed resistors, which are not sensitive to change of frequency or temperature, so $G$ remains constant for a wider range of frequencies and temperatures

[109] Discussions in this section are not required by the syllabus. If you do not have the appetite, you may just skip to the conclusions.

we conclude that introducing negative feedback has the following benefits:

- reduced gain
- more stable gain
- increased bandwidth
- less distortion

**the golden rule**

with negative feedback, we expect that $V_{\text{out}}$ is able to change correspondingly with $V_{\text{in}}$

for any op-amp, $V_{\text{out}} = G_0(V_+ - V_-)$, but op-amp has very large open-loop gain $G_0$

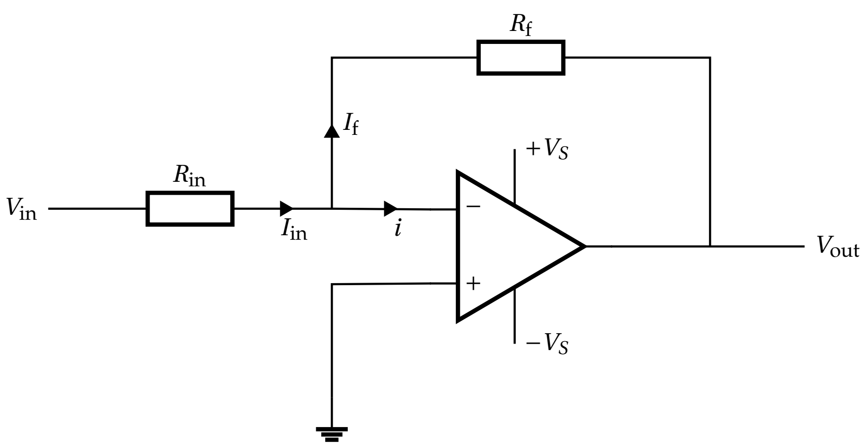
if $V_{\text{out}}$ does not saturate, one must have $V_+ - V_- \approx 0$

this can be summarised as the *golden rule*: output of op-amp attempts to do whatever is necessary to make the difference between the two input voltages zero

note that op-amp can do this only if there is negative feedback

### 15.2.5 inverting amplifier

one type of widely-used amplifier circuit is called **inverting amplifier**




inverting amplifier

$V_{\text{out}}$ is related to $V_{\text{in}}$ by $\boxed{V_{\text{out}} = \left(-\frac{R_{\text{f}}}{R_{\text{in}}}\right) V_{\text{in}}}$, where gain of the amplifier is $G = \frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{R_{\text{f}}}{R_{\text{in}}}$

to show this relation, we need two approximations

➢ virtual earth approximation: if $V_{\text{out}}$ does not saturate, must have $V_+ \approx V_-$, but $V_+$ is earthed, i.e., $V_+ = 0$, so $V_- \approx 0$, referred to as a **virtual earth**

➢ small current approximation: because of the very large input impedance of op-amp, current flowing into inverting input is negligible, i.e., $i \approx 0$

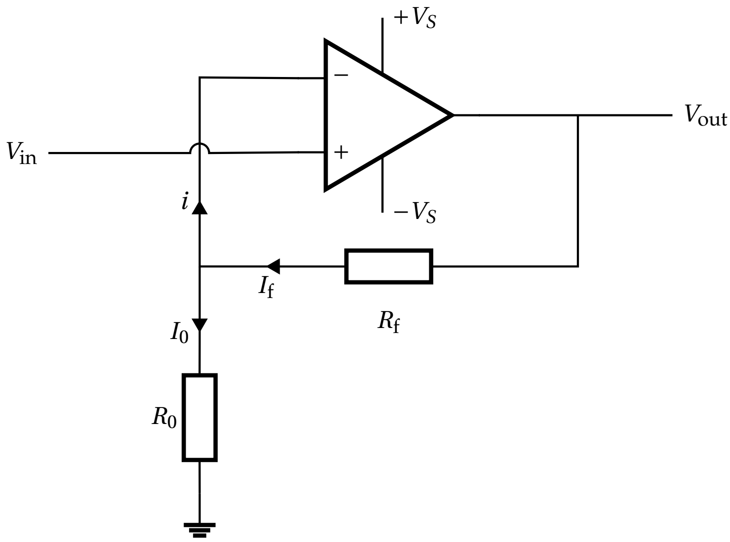
therefore currents through $R_{\text{in}}$ and $R_{\text{f}}$ are approximately the same: $I_{\text{in}} = I_{\text{f}}$

from this equation, we can relate $V_{\text{out}}$ to $V_{\text{in}}$:

$$\frac{V_{\text{in}} - V_-}{R_{\text{in}}} = \frac{V_- - V_{\text{out}}}{R_{\text{f}}} \quad \Rightarrow \quad \frac{V_{\text{in}} - 0}{R_{\text{in}}} = \frac{0 - V_{\text{out}}}{R_{\text{f}}} \quad \Rightarrow \quad V_{\text{out}} = \left(-\frac{R_{\text{f}}}{R_{\text{in}}}\right) V_{\text{in}}$$

### 15.2.6 non-inverting amplifier

another type of common amplifier circuit is called **non-inverting amplifier**




non-inverting amplifier

for this circuit, $V_{\text{out}}$ is given by: $\boxed{V_{\text{out}} = \left(1 + \frac{R_{\text{f}}}{R_0}\right) V_{\text{in}}}$, where overall gain $G = \frac{V_{\text{out}}}{V_{\text{in}}} = 1 + \frac{R_{\text{f}}}{R_0}$

to show this, recall current into input of op-amp $i \approx 0$ due to its large input impedance, so

$$I_0 = I_{\text{f}} \quad \Rightarrow \quad \frac{V_- - 0}{R_0} = \frac{V_{\text{out}} - V_- -}{R_{\text{f}}}$$

again, if $V_{\text{out}}$ is not saturated, $V_- \approx V_+ = V_{\text{in}}$, then we find

$$\frac{V_{\text{in}}}{R_0} = \frac{V_{\text{out}} - V_{\text{in}}}{R_{\text{f}}} \quad \Rightarrow \quad R_0 V_{\text{out}} - R_0 V_{\text{in}} = R_{\text{f}} V_{\text{in}} \quad \Rightarrow \quad V_{\text{out}} = \left(1 + \frac{R_{\text{f}}}{R_0}\right) V_{\text{in}}$$

**remarks on inverting & non-inverting amplifiers**

inverting and non-inverting amplifiers are starting point for more complex circuits

before going through detailed examples, let's briefly comment on the results we have found

$$\text{for inverting amplifier: } G = \frac{-R_f}{R_{in}} \qquad \text{for non-inverting amplifer: } G = 1 + \frac{R_f}{R_0}$$

➢ circuit's overall gain for both amplifiers only depend on choice of resistors

so gain can now be easily tuned, circuit design becomes more convenient

➢ no parameter of op-amp shows up in final expressions for overall gain

gain is determined by feedback network, rather than by op-amp's characteristics

since value of resistors does not change significantly with temperature or frequency

this again means gain becomes now more stable with negative feedback

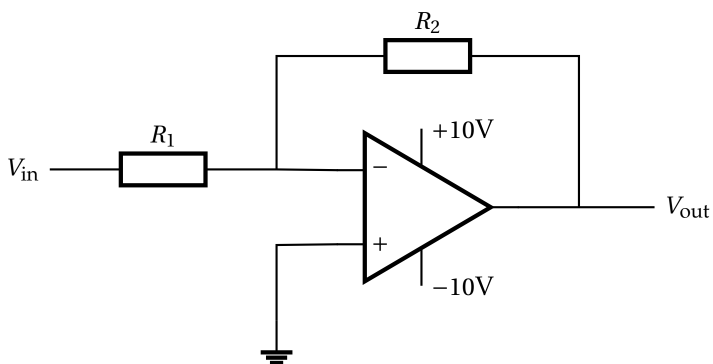
➢ for inverting amplifiers, $G < 0$, $V_{in}$ and $V_{out}$ always out of phase (opposite signs)

for non-inverting amplifiers, $G > 0$, $V_{in}$ and $V_{out}$ always in phase (same sign)

➢ the proportional relation between $V_{in}$ and $V_{out}$ only holds if op-amp is not saturated

if we find $V_{out}$ greater than supply voltage from a calculation, true $V_{out}$ equals $\pm V_S$

**Example 15.4** In the inverting amplifier circuit below, $R_1 = 10\ \text{k}\Omega$, $R_2 = 50\ \text{k}\Omega$. (a) Find $V_{out}$ when $V_{in} = +1.0$ V. (b) Find $V_{out}$ when $V_{in} = -4.0$ V.

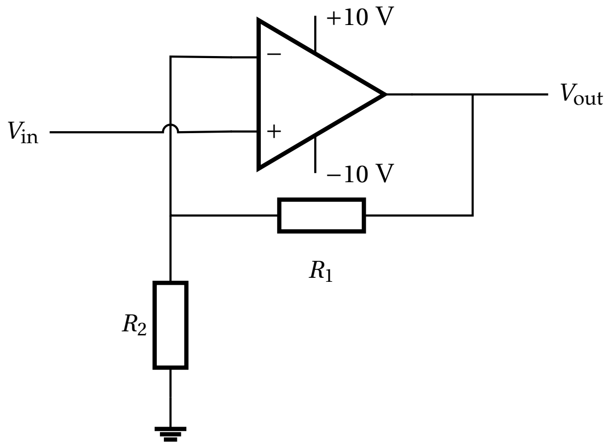



✍ overall gain of the amplifier: $G = -\frac{R_2}{R_1} = -\frac{50\text{k}}{10\text{k}} = -5.0$

for $V_{in} = +1.0$ V: $V_{out} = (-5.0) \times (+1.0) = -5.0$ V

for $V_{in} = -4.0$ V: $V_{out} = (-5.0) \times (-4.0) = +20$ V, output becomes saturated, so $V_{out} = +10$ V □

**Example 15.5** In the non-inverting amplifier circuit below, $R_1 = 48\ \mathrm{k\Omega}$, $R_2 = 6.0\ \mathrm{k\Omega}$. (a) Find $V_{\mathrm{out}}$ when $V_{\mathrm{in}} = +0.5$ V. (b) Find $V_{\mathrm{out}}$ when $V_{\mathrm{in}} = -3.0$ V.

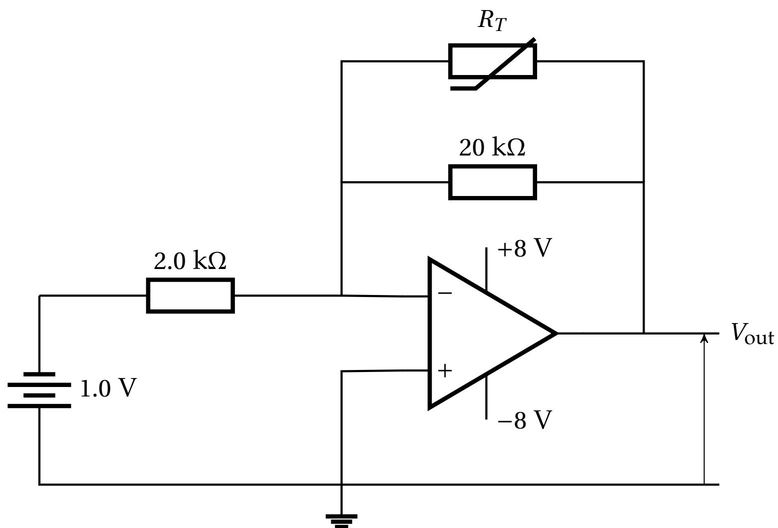



✍ overall gain of the amplifier: $G = 1 + \frac{R_1}{R_2} = 1 + \frac{48\mathrm{k}}{6.0\mathrm{k}} = 9.0$

for $V_{\mathrm{in}} = +0.5$ V: $V_{\mathrm{out}} = 9.0 \times (+0.5) = +4.5$ V

for $V_{\mathrm{in}} = -3.0$ V: $V_{\mathrm{out}} = 9.0 \times (-3.0) = -27$ V, output becomes saturated, so $V_{\mathrm{out}} = -10$ V □

**Example 15.6** An amplifier circuit incorporating an ideal op-amp is shown.




Initially, the thermistor has a resistance of 20 kΩ. (a) Find the gain of this circuit. (b) Find the output voltage. (c) State and explain how $V_{\mathrm{out}}$ changes as temperature rises. (d) The thermistor is then removed, find the new output voltage.

✍ (a) this is an inverting amplifier (there is negative feedback loop and $V_{out}$ is fed into $V_-$)

resistance of feedback loop: $R_f = \left(\frac{1}{20k} + \frac{1}{20k}\right)^{-1} = 10\,k\Omega$

gain of amplifer: $G = -\frac{R_f}{R_{in}} = -\frac{10k}{2.0k} \quad \Rightarrow \quad G = -5.0$

(b) $V_{in}$ is at negative terminal of the battery, which is 1.0 V lower than the positive terminal

but positive terminal is earthed, so $V_{in} = -1.0$ V

output voltage: $V_{out} = GV_{in} = (-5.0) \times (-1.0) \quad \Rightarrow \quad V_{out} = +5.0$ V

(c) as temperature increases, resistance of thermistor decreases

total resistance of feedback loop becomes smaller

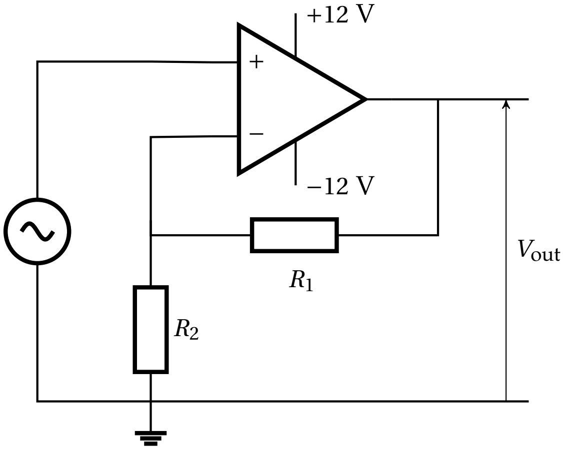
gain of the circuit then becomes less negative, so $V_{out}$ will decrease

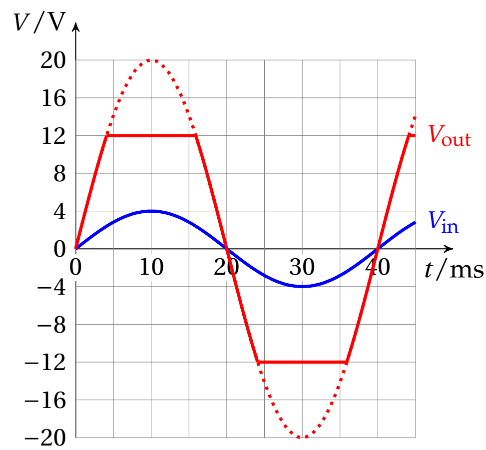
(d) gain of amplifier after removal of thermistor: $G' = -\frac{R'_f}{R_{in}} = -\frac{20k}{2.0k} \quad \Rightarrow \quad G' = -10$

new output voltage can be computed: $V'_{out} = G'V_{in} = (-10) \times (-1.0) = +10$ V

but voltage supply to op-amp is ±8 V, op-amp saturates, so $V'_{out} = +8$ V □

**Example 15.7** A circuit incorporating an ideal op-amp is shown below. The two fixed resistors have $R_1 = 36$ kΩ and $R_2 = 9.0$ kΩ. Variation of $V_{in}$ from a voltage source is shown with the blue curve. On the same graph, plot how $V_{out}$ varies with time.




✍ this is a non-inverting amplifier, whose voltage gain is: $G = 1 + \frac{R_1}{R_2} = 1 + \frac{36k}{9.0k} = +5.0$

so $V_{out} = +5.0 \times V_{in}$, with which we attempt to plot $V_{out}$ as the red dotted curve

but $V_{out}$ will saturate at voltage of power supply, which is ±12 V

taking this into account, we can plot the true $V_{out}$ as the red solid curve □

**Question 15.3** An amplifier circuit is used to monitor light intensity. The magnitude of its gain is required to increase as light intensity increases. Suggest how this can be done.

**Question 15.4** Show that the output voltage from an inverting amplifier is $V_{\text{out}} = \left(-\dfrac{R_{\text{f}}}{R_{\text{in}}}\right) V_{\text{in}}$.

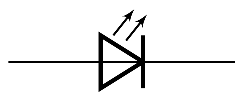
**Question 15.5** Show that the output voltage from a non-inverting amplifier is $V_{\text{out}} = \left(1 + \dfrac{R_{\text{f}}}{R_0}\right) V_{\text{in}}$.

## 15.3 output devices

### 15.3.1 LED

a **diode** is a device that only allows current to flow in one direction

**LED**, or light-emitting diode, is able to give off light when current passes through it

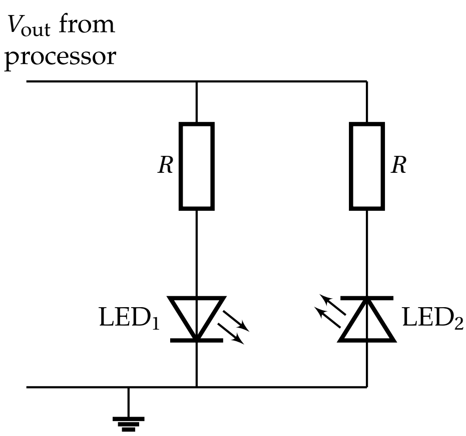


electric symbol for an LED

LEDs are efficient light emitters and they come in many different colours

LEDs can therefore be used as indicators to show the *polarity* of $V_{\text{out}}$ from processing units

a partial circuit incorporating two LEDs used as output devices is shown




when $V_{\text{out}} > 0$, LED 1 emits light, LED 2 is off

when $V_{\text{out}} < 0$, LED 2 emits light, LED 1 is off

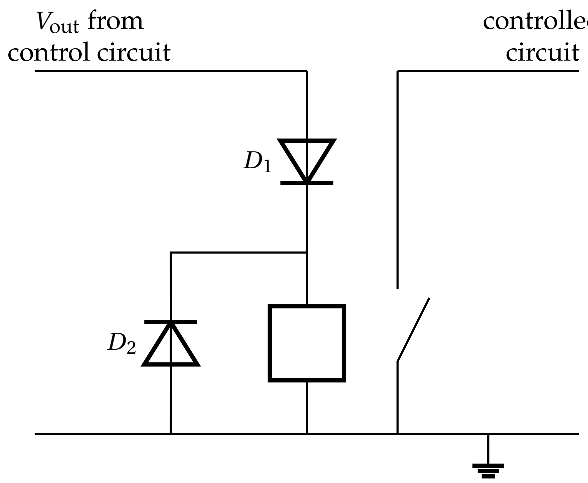
protector resistors $R$ are included to ensure LED units not to be broken by large $V_{\text{out}}$

### 15.3.2 relay

**relay** is an electromagnetic switch which enables the use of a current in one circuit (control circuit) to switch on another current in a second circuit (controlled circuit)

a relay consists of an electromagnet, a moveable arm and switch contacts current flowing in the coils of the electromagnet induces a magnetic field to attract the moveable arm, which then connects the contacts and completes the second circuit

➢ relay is used for many purposes, some of which are listed below
  - switching large voltage/current by means of a small voltage/current
  - high voltage isolation
  - remote switching




output from processing units connected to a relay

➢ in practice, a relay is usually connected with two *diodes* for the following reasons:
  - diode $D_1$ ensures relay is triggered only for positive $V_{\text{out}}$
    without $D_1$, relay would operate for currents flowing in either direction
  - diode $D_2$ protects sensitive processing units from large currents induced from relay coil
    when control circuit is turned off, large currents could be induced from relay coil
    this current might damage an op-amp or other sensitive components
    $D_2$ opens a path for induced current to flow (a short circuit across coil)

#### 15.3.3 calibrated meters

we can use sensing devices to monitor a specific physical quantity, say $x$

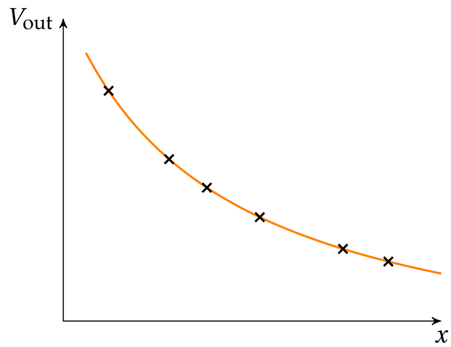
$x$ could be temperature, light level, sound intensity, etc.

change in $x$ would cause change in output voltage from the processing units for convenience, we would hope value of $x$ can be directly read off from a voltmeter, so we need associate a value of $V_{out}$ with a value for $x$

$V_{out}$ from processors is usually *non-linear* in $x$, so one must **calibrate** scale of meter in terms of $x$,

calibration can be done according to a curve of measured data, called *calibration curve*

scale will be non-uniform, but this is doable




example of a calibration curve

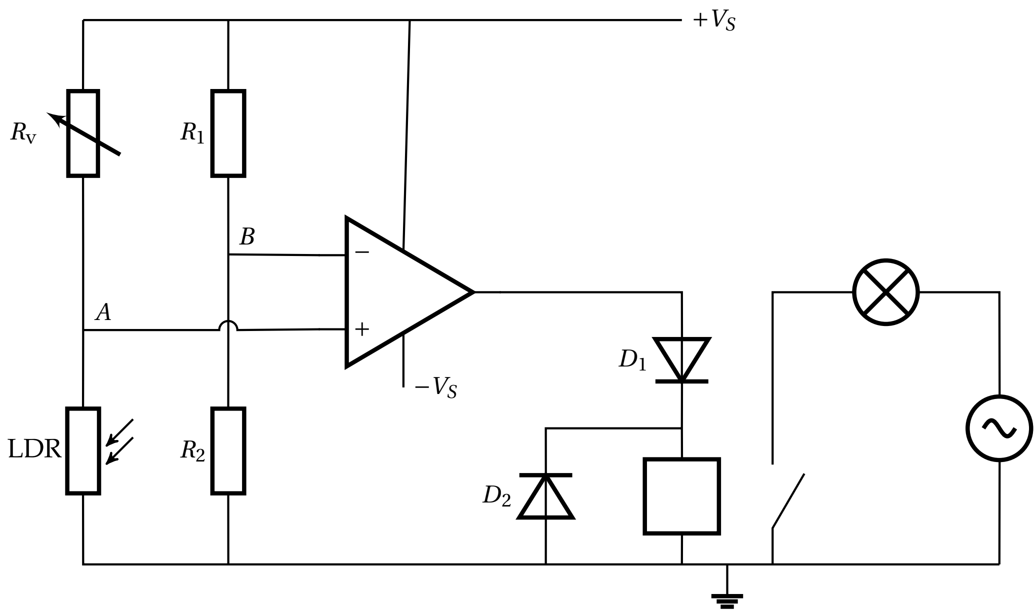
## 15.4 practical electronic circuits

the many things we learned so far can be put together to build functional electronic circuits

in last part of this section, we will see a few examples of practical circuits

### 15.4.1 automatic illumination system

the circuit below enables automatic switching of lamp when it gets dark




this circuit consists of a comparator, whose inputs are provided from two potential dividers output from the comparator could trigger a relay that controls an illumination circuit

fixed resistors $R_1$ and $R_2$ provide a constant potential to $V_-$ for reference

magnitude of $V_+$ depends on the voltage share on LDR, which depends on light conditions

when in bright, LDR has low resistance, $V_+$ is low, so $V_+ < V_-$, $V_{\text{out}}$ from op-amp is negative

current through relay is blocked by diode $D_1$, lamp circuit is off

as light level falls, resistance of LDR increases, $V_+$ then rises

at a certain light level, $V_+$ becomes greater than $V_-$, $V_{\text{out}}$ from op-amp becomes positive

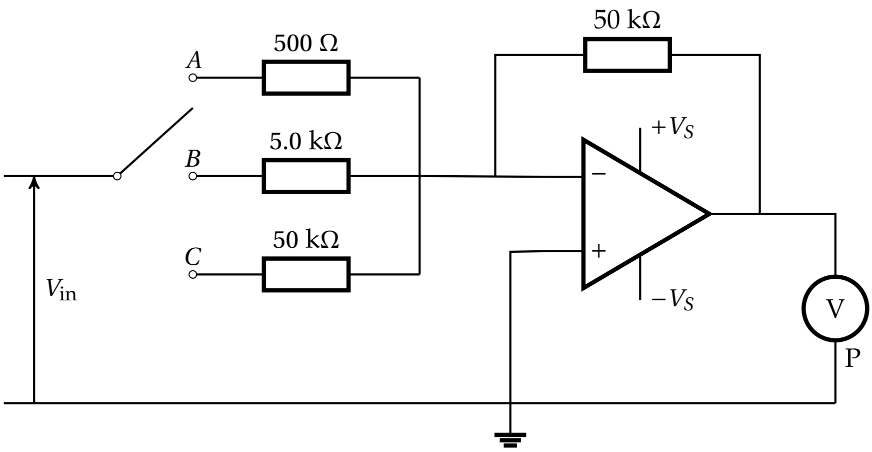
current in relay triggers the switch to close, the lamp is switched on

variable resistor $R_V$ can be set to determine the critical light level of switch-over

### 15.4.2 multi-range voltmeter

the inverting amplifier can be adapted to make a multi-range voltmeter

positive connection of voltmeter is at $P$ so that positive reading is obtained for $+V_{\text{in}}$




in this example, $V_{\text{in}}$ can be sent into the amplifier through three different channels

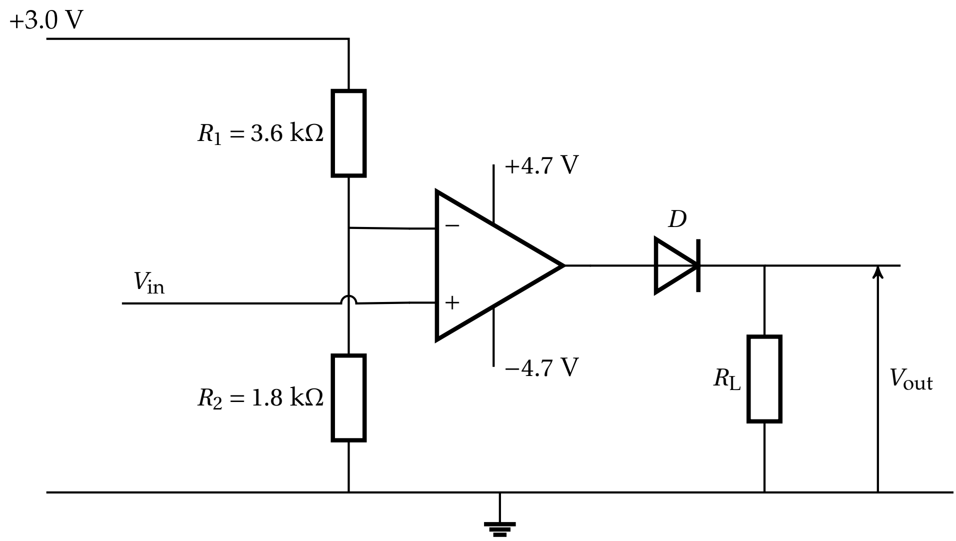
you can check that each channel has a different value of gain: $G_A = -100$, $G_B = -10$, $G_C = -1.0$

if the voltmeter connected to the output of op-amp has a range of $0 \sim 10$ V, then for $V_{\text{in}}$ less than 0.1 V, or in the range $0.1 \sim 1.0$ V, or in the range $1.0 \sim 10$ V, they can be sent through channel $A$, $B$ or $C$ respectively to get a noticeable deflection on the voltmeter

### 15.4.3 digital signal regenerator

the circuit given below can be used to regenerate a digital signal [110]

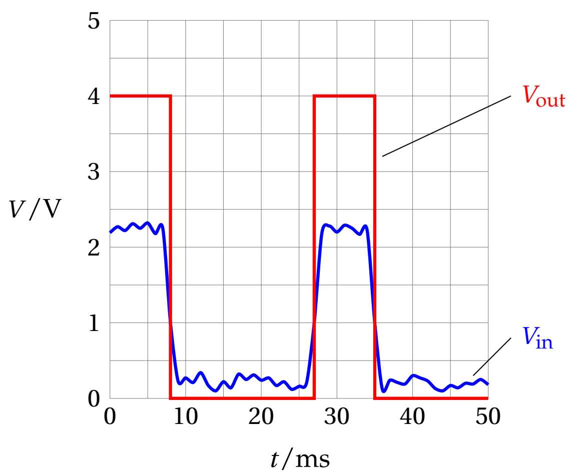
suppose $D$ is a silicon diode which develops a p.d. drop of 0.7 V when it is forward biased




$V_- = \dfrac{1.8}{1.8+3.6} \times (+3.0) = +1.0$ V gives a constant potential to be compared with $V_+$

when $V_{\text{in}} > 1.0$ V, i.e., $V_+ > V_-$, output from op-amp is +4.7 V, diode $D$ conducts and gets its voltage share of 0.7 V, the other 4.0 V is dropped across the load $R_L$, so $V_{\text{out}} = +4.0$ V

when $V_{\text{in}} < 1.0$ V, i.e., $V_+ < V_-$, output from op-amp is −4.7 V, diode $D$ does not conduct, no current flows through $R_L$, then $V_{\text{out}} = 0$




if a *digital* signal is transmitted over a long distance, it may suffer from loss in power and noise, we would want to amplify the signal and remove the noise from it

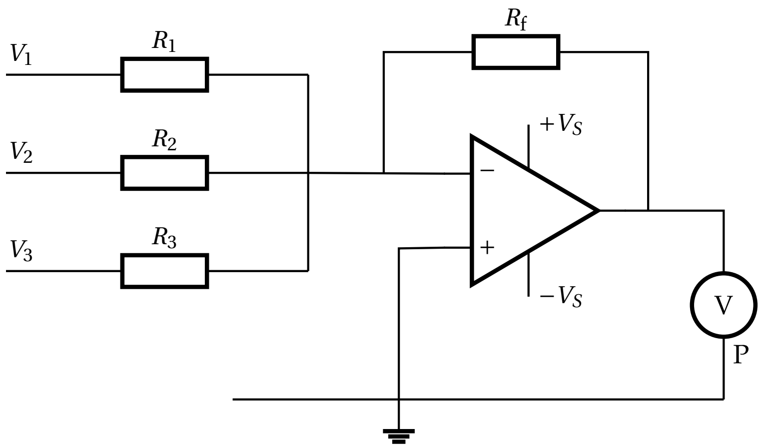
this circuit can take in this signal and produce an output that removes the fluctuations due to noise and hence regenerates the original digital signal with only highs and lows

[110]The use of digital signals in communication systems will be discussed in detail in §16.2.

### 15.4.4 summing amplifier

circuits incorporating op-amp can carry out many mathematical operations

the circuit we are going to introduce here, called the *summing amplifier*, is able to produce a voltage output as a *weighted sum* of several input signals [111]




this is basically a variation of the inverting amplifier

it can be shown that the output voltage from op-amp is given by: $V_{out} = -R_f\left(\dfrac{V_1}{R_1} + \dfrac{V_2}{R_2} + \dfrac{V_3}{R_3}\right)$

positive connection of voltmeter is at $P$ so positive reading is obtained for $V_1, V_2, V_3 > 0$

so reading on voltmeter is: $V = R_f\left(\dfrac{V_1}{R_1} + \dfrac{V_2}{R_2} + \dfrac{V_3}{R_3}\right)$, a weighted sum of inputs $V_1$, $V_2$, and $V_3$

- in the case where $R_1 = R_2 = R_3 = R_f$, the reading simply becomes: $V = V_1 + V_2 + V_3$

  this circuit produces the algebraic sum of the inputs, so it becomes a *voltage adder*

- in the case where $R_f = 4R_1 = 2R_2 = R_3$, the reading becomes: $V = 4V_1 + 2V_2 + V_3$

  if $V_1$, $V_2$, $V_3$ are only allowed to be 0 or 1, they can represent a three-bit binary number

  this circuit can convert this binary number into its decimal value

  so what we have here is a *binary-to-decimal converter*

[111] In a university-level electronic engineering course, you will see that op-amp can be used to build difference amplifiers, differentiator and integrator amplifiers (by including a capacitor), logarithmic amplifiers (by including a diode) and many other useful applications. Those who have interest in engineering science are encouraged to research on these topics.

**Question 15.6** For the summing amplifier introduced, if $R_f = R_1 = 10$ kΩ, $R_1 = 1.0$ kΩ, and $R_3 = 100$ Ω, find an expression for the voltage reading displayed on the voltmeter, and hence suggest the function of this circuit.

**Question 15.7** Design a non-inverting summing amplifier where the inputs are sent into the non-inverting input of the op-amp. (∗)

**Question 15.8** Design a circuit for a house such that a fire alarm can be automatically triggered when the temperature in the house becomes too high. (∗)

# Chapter 16

# Telecommunication (∗)

telecommunication concerns transmitting information by electromagnetic means

engineers focus on solving the problem of transmitting large volumes of information over long distances with minimal loss of signal strength and distortion due to noise

in this chapter, we will study three of the key components in a communication system [112]

- *modulator* and *demodulator*

  variation of an information signal is represented by a carrier wave through modulation

  original information can be extracted through the reverse process called demodulation

- *analogue-to-digital* and *digital-to-analogue converters*

  analogue signals are digitized to archive more reliable and efficient transmission

  digital signals are then converted back to analogue form reproduce original signal

- *communication channel*

  communication channel is a medium through which signals are transmitted

  each type of communication channel has its relative merits and disadvantages

## 16.1 modulation

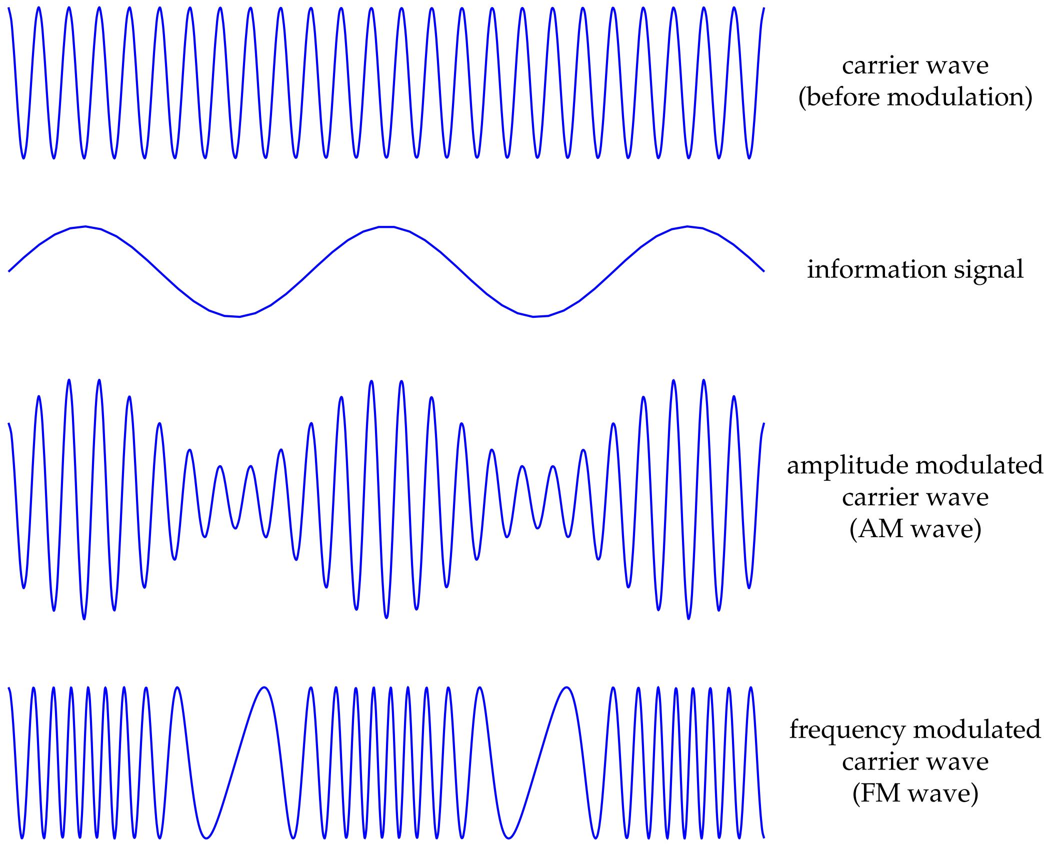
audio signals can be converted into a radio signal for transmission

rather than being transmitting directly as a radio wave, information signal is usually en-

[112] There are other important processes in telecommunication. For example, digital signals are usually passed through an *encoder* so that redundant information is removed and extra data is introduced to check for errors . Modulated signals are usually *multiplexed*, so multiple users can share the same communication channel. Common multiplexing schemes include FDMA, TDMA, CDMA, etc.

coded into a carrier wave for better transmission

> **modulation** means varying the amplitude or the frequency of a carrier wave in synchrony with the displacement of an information signal




modulation of a carrier wave in accordance with an information signal

- reasons for modulation
  - carrier wave has shorter wavelength, so shorter aerial is possible
  - each signal can be transmitted with a different carrier frequency

    so multiple signals can be transmitted at the same time without interfering each other
  - modulated carrier wave is less affected by noise
  - transmission over long distance is possible

### 16.1.1 amplitude modulation

for amplitude modulated carrier wave, frequency is unchanged

amplitude varies with time according to displacement of information signal

waveform of information signal determines the *envelope* of the modulated wave

#### spectrum of AM wave

amplitude modulated carrier wave is found to contain more than one frequencies

mathematically, each AM wave can be considered as the sum of several simple sine waves

proof (∗): suppose the original carrier wave takes the form: $x_c(t) = A_c \sin\omega_c t$

it is to be amplitude modulated by an information signal: $x_m(t)$

modulated wave has a new amplitude that varies with time: $A'_c(t) = A_c + x_m(t)$

for an audio signal of single frequency, say $x_m(t) = A_m \cos\omega_m t$, the AM wave takes the form:

$$x'_c(t) = (A_c + A_m \cos\omega_m t)\sin\omega_c t = A_c \sin\omega_c t + A_m \sin\omega_c t \cos\omega_m t$$

using trigonometric identity: $\sin\alpha\sin\beta = \frac{1}{2}\left[\sin(\alpha+\beta) + \sin(\alpha-\beta)\right]$, we find:

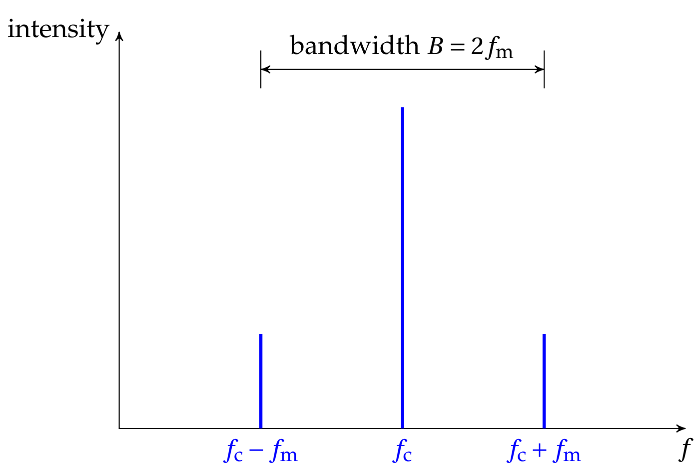
$$x'_c(t) = A_c \sin\omega_c t + \frac{1}{2} A_m \sin(\omega_c + \omega_m)t + + \frac{1}{2} A_m \sin(\omega_c - \omega_m)t$$

AM wave is a combination of three components of frequencies $f_c$, $f_c + f_m$ and $f_c - f_m$ □

➢ for an information signal of *single* frequency, we can draw the spectrum for the AM wave

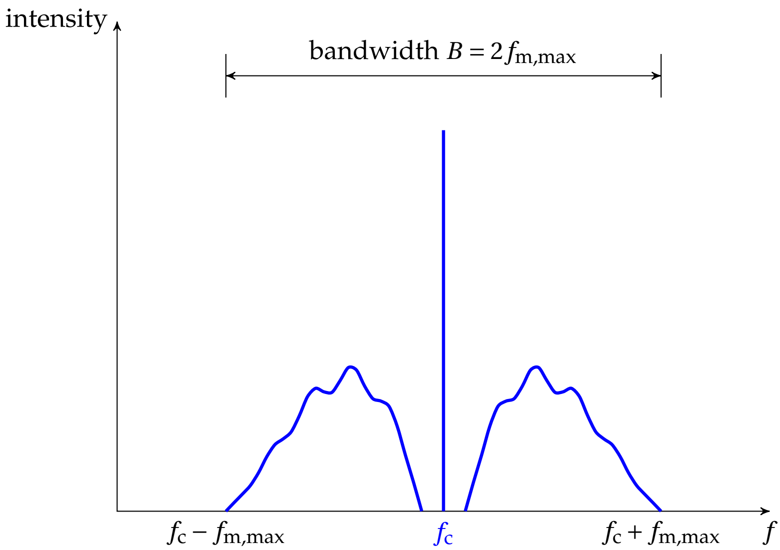
apart from original carrier frequency, there are two more frequencies $f_c \pm f_m$, called *sidebands*

positions of sideband is determined by frequency of the information signal $f_m$



➢ information signal could contain a *range* of frequencies (e.g., music signal)

the two sidebands would spread into two symmetrical humps




➢ we define **bandwidth** as the range of frequencies occupied by the modulated signal

- for a single-frequency information signal, bandwidth of AM wave is:

$$B = (f_c + f_m) - (f_c - f_m) \quad \Rightarrow \quad \boxed{B = 2f_m}$$

- for a multi-frequency information signal, bandwidth of AM wave is:

$$B = (f_c + f_{m,max}) - (f_c - f_{m,max}) \quad \Rightarrow \quad \boxed{B = 2f_{m,max}}$$

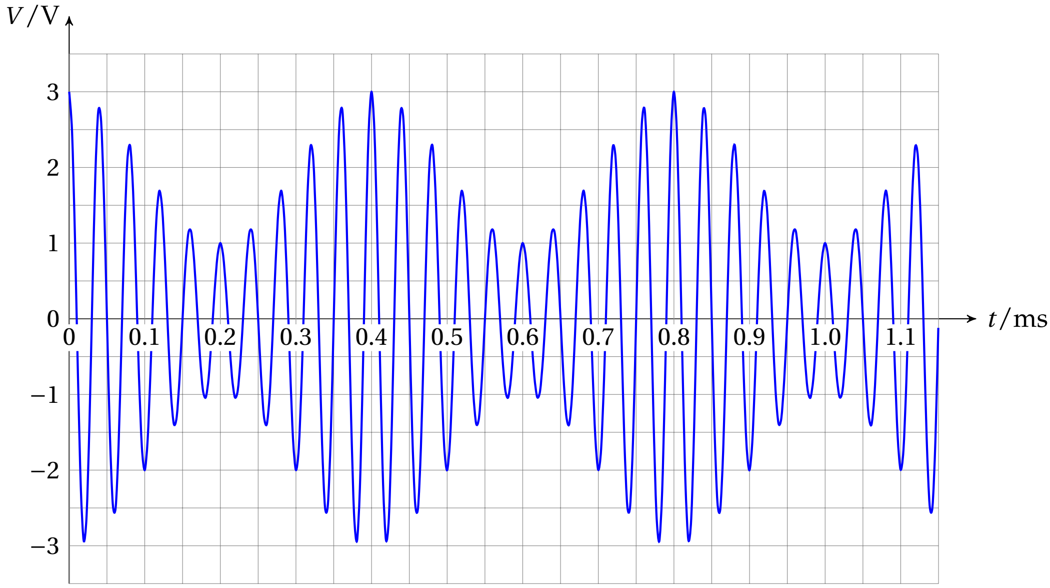
for sound signals, $f_m \approx 10^3$ Hz, so bandwidth is around a few kHz

compared with frequency of radio waves, bandwidth of AM is quite small

➢ size of bandwidth determines number of channels available in a given frequency range

smaller bandwidth for AM means more channels are possible

**Example 16.1** A radio station is broadcasting several audio signals with same bandwidth. Wavelengths of carrier waves available for this radio station ranges from $3.0 \times 10^3$ m to $1.5 \times 10^4$ m. The variation of voltage of a modulated signal is shown. (a) State the frequency of unmodulated carrier wave and the frequency of the signal wave. (b) Find the bandwidth of this signal. (c) Find the maximum number of radio channels that could be broadcast at the same time for this station.




✍ frequency of carrier wave: $f_c = \dfrac{1}{T_c} = \dfrac{1}{0.04 \text{ ms}} = 25 \text{ kHz}$

frequency of audio signal: $f_m = \dfrac{1}{T_m} = \dfrac{1}{0.40 \text{ ms}} = 2.5 \text{ kHz}$

bandwidth of the signal: $B = 2f_m = 2 \times 2.5 = 5.0 \text{ kHz}$

range of available carrier frequencies is found using $f = \dfrac{c}{\lambda}$:

maximum carrier frequency: $f_{max} = \dfrac{c}{\lambda_{min}} = \dfrac{3.0 \times 10^8}{3.0 \times 10^3} = 1.0 \times 10^5 \text{ Hz} = 100 \text{ kHz}$

minimum carrier frequency: $f_{min} = \dfrac{c}{\lambda_{max}} = \dfrac{3.0 \times 10^8}{1.5 \times 10^4} = 2.0 \times 10^4 \text{ Hz} = 20 \text{ kHz}$

number of radio channels: $n = \dfrac{f_{max} - f_{min}}{B} = \dfrac{100 - 20}{5.0} = 16$ □

**Question 16.1** An amplitude modulated radio wave carries an audio signal of a single frequency. In the power spectrum, it is found to contain three frequencies: 45 kHz, 50 kHz, and 55 kHz.(a) What is the frequency of the carrier wave? (b) What is the frequency of the audio signal? (c) What is the bandwidth for this AM wave?

**Question 16.2** A sinusoidal carrier wave has a frequency of 200 kHz. It is amplitude modulated for the broadcast of music that has frequencies ranging between 30 Hz and 4000 Hz. Sketch the power spectrum for the transmitted signal, hence or otherwise, find the bandwidth for this signal. You should label the frequency values on your graph.

### 16.1.2 frequency modulation

for frequency modulated carrier wave, amplitude is unaltered

frequency varies with time according to displacement of information signal

➢ shift in frequency of carrier wave is determined by a parameter called *frequency deviation*

frequency deviation is usually given in kHz $V^{-1}$

this means if displacement of information wave changes by one volt, frequency deviation gives the amount of change in frequency for the carrier wave

➢ FM wave give very complicated spectrum [113] [114]

as a consequence, each FM signal occupies very large bandwidth

**Example 16.2** A carrier wave of amplitude 5.0 V and frequency 80 kHz is modulated in frequency by a sinusoidal information wave of amplitude 3.0 V and frequency 4000 Hz. The frequency deviation of the carrier wave is 2.0 kHz $V^{-1}$. Describe the variation of this frequency modulated carrier wave.

✍ amplitude of modulated wave remains unchanged, so it is still 5.0 V

as displacement of information wave varies between +3.0 V to −3.0 V, maximum frequency shift for the carrier wave is: $\pm 3.0 \times 2.0 = \pm 6.0$ kHz

maximum frequency of FM wave: $f_{\max} = 80 + 6.0 = 86$ kHz

minimum frequency of FM wave: $f_{\min} = 80 - 6.0 = 74$ kHz

so frequency of modulated wave changes from a maximum value of 86 kHz to a minimum value of 74 kHz and then back to the maximum value repeatedly at a rate of 4000 $s^{-1}$ □

**Question 16.3** A carrier wave has a frequency of 900 kHz. It is frequency modulated by an

[113]FM spectrum is so complicated that you are only supposed to know it is complicated. You will not be asked to sketch a power spectrum for FM wave in A-Levels.

[114]If a carrier wave $x_c = A_c \sin\omega_c t$ is to be frequency modulated by an information signal $x_m$, the FM wave takes the form: $x_c(t) = A_c \sin\left[\omega_c t + f_\Delta \int_0^t x_m t(t) \mathrm{d}t\right]$, where $f_\Delta$ is the frequency deviation factor. In particular, for an information signal of single frequency $x_m = A_m \cos\omega_m t$, it can be shown that the FM wave becomes $x_c(t) = A_c \sin\left[\omega_c t + \frac{A_m f_\Delta}{\omega_m} \sin\omega_m t\right]$. Judging from the mathematical descriptions, FM wave is obviously much more complicated than the AM wave.

audio signal of frequency 6.0 kHz and amplitude 2.5 V. The frequency deviation for modulation is 30 kHz $V^{-1}$. (a) Find the maximum and minimum frequency of the modulated wave. (b) Describe how the frequency of the modulated wave changes.

#### 16.1.3 comparison between AM and FM

- relative advantages of AM transmission
  - electronics for AM transmission is less complicated, so cheaper radio sets
  - narrower bandwidth for AM channel, so more channels are available in a given range
  - AM wave has lower frequencies, hence longer wavelengths

    this means AM waves can better diffract around barriers (e.g., buildings, mountains)

    each AM transmitter covers a greater area, so fewer AM transmitters required than FM
- relative advantages of FM transmission
  - FM transmission is less affected by noise

    noise would superimpose with transmitted signal, so amplitude is altered by noise

    but for FM, variation of information signal is represented by variation in frequency
  - greater bandwidth for FM, so better reproduction quality of sound
  - FM transmission is more energy efficient

    for AM waves, information is carried in the sidebands

    but the sidebands only take a small fraction of total power

**Question 16.4** Suggest two reasons why FM broadcasting is more costly than AM.

## 16.2 digital transmission

#### 16.2.1 analogue & digital signals

an **analogue signal** a continuous signal that can take any value

a **digital signal** is a discrete signal that can only take 0's and 1's (or highs and lows), no intermediate value is allowed

### 16.2.2 regeneration

there are two major problems during the transmission of a signal: attenuation and noise

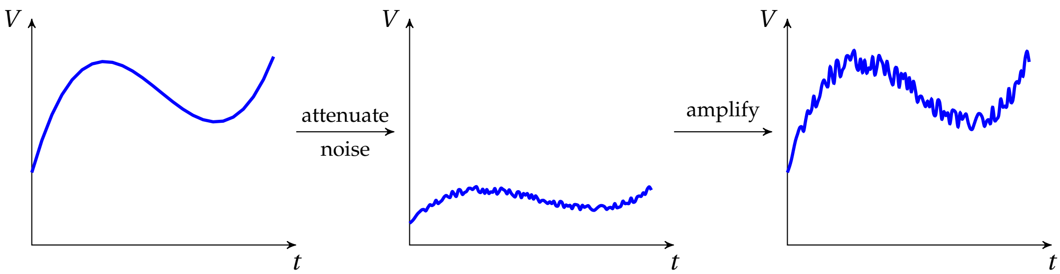
> **attenuation** means gradual loss of signal power/intensity during transmission

> **noise** refers to the unwanted random interference that distorts the signal

at the receiver end, we would want to remove noise and amplify the attenuated signal, so original signal is recovered, this process is called **regeneration**

- for an analogue signal, noise can not be distinguished

  so noise will be amplified together with the weakened signal

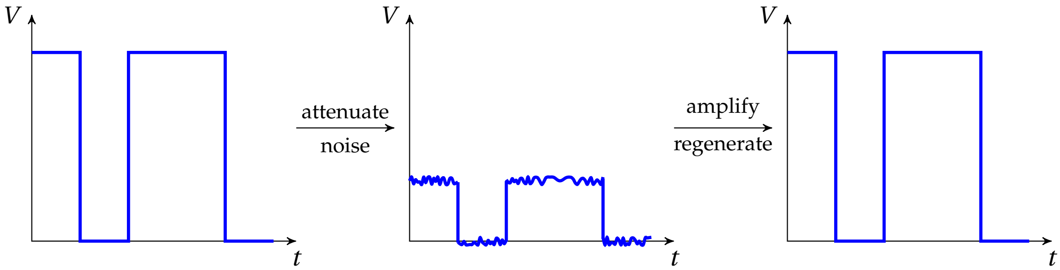



noise in an analogue system is amplified along with the weakened signal

- for a digital signal, it is either 0 or 1, any fluctuation must be due to noise

  so noise can easily be removed from a digital signal, signal is regenerated

  circuit for a simple regenerator amplifier is introduced in §15.4.3




noise in a digital system can be removed through regeneration

**advantage of digital transmission**

transmission of data in digital form brings many advantages:

- noise can be removed through regeneration
- only two states are involved, so more reliable and easier electronics
- digital signals can be encrypted, so better security
- extra data can be added, so errors can be checked
- bandwidth is small, so rate of data transmission is higher
- data in digital forms are stored and processed more easily

### 16.2.3 binary numbers

most signals that make up speech, music, radio, and television are analogue signals

for more reliable and more efficient transmission of data, it is necessary to convert analogue signals into digital forms

this requires the use of 0's and 1's to represent different voltage levels

we next study the binary number system in which a number is expressed in 0's and 1's

➢ each digit in a binary number is called a *bit*

➢ for a binary number, each bit represents an increasing power of 2

the bit on left-hand/right-hand side of a binary number has highest/lowest value, called the *most/least significant bit*, or MSB/LSB in short

**Example 16.3** Convert the binary numbers into decimal numbers: (a) 11001 (b) 10110.

✍ $11011_2 = 1\times 2^4 + 1\times 2^3 + 0\times 2^2 + 0\times 2^1 + 1\times 2^0 = 2^4 + 2^3 + 2^0 = 16 + 8 + 1 = 25$

$110110_2 = 1\times 2^5 + 1\times 2^4 + 0\times 2^3 + 1\times 2^2 + 1\times 2^1 + 0\times 2^0 = 2^5 + 2^4 + 2^2 + 2^1 = 32 + 16 + 4 + 2 = 54$ □

**Example 16.4** Convert the decimal numbers into binary numbers: (a) 13 (b) 28.

✍ $13 = 8 + 4 + 1 = 2^3 + 2^2 + 2^0 = 1\times 2^3 + 1\times 2^2 + 0\times 2^1 + 1\times 2^0 = 1101_2$

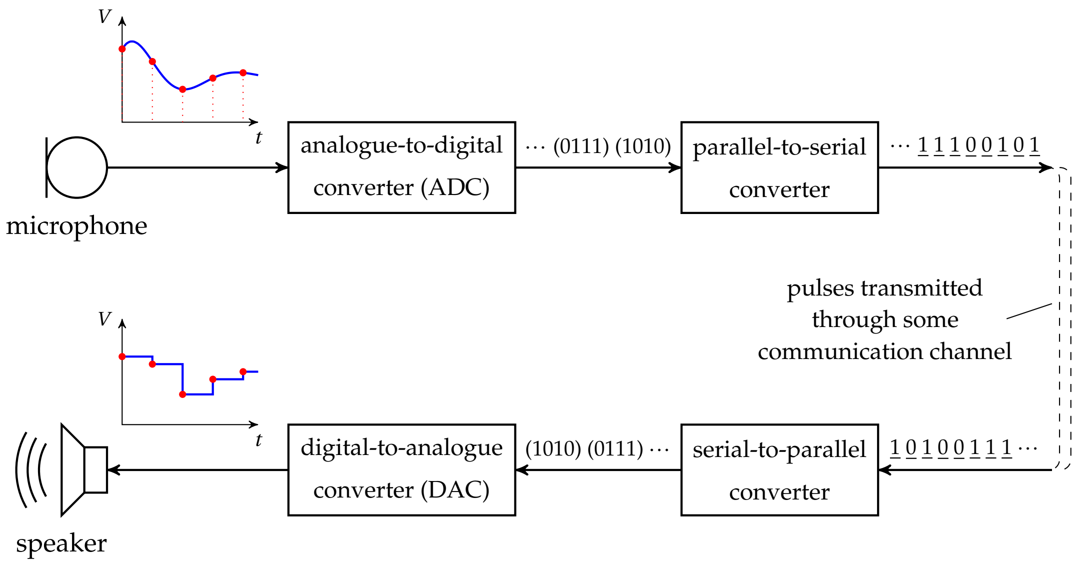
$28 = 16 + 8 + 4 = 2^4 + 2^3 + 2^2 = 1\times 2^4 + 1\times 2^3 + 1\times 2^2 + 0\times 2^1 + 0\times 2^0 = 11100_2$ □

➢ a $n$-bit binary system can represent $2^n$ different numbers

increasing the number of bits improves the precision of representation of voltage levels

### 16.2.4 analogue-to-digital conversion

digital transmission of an audio signal is represented by the diagram below:




transmission and reproduction of an audio signal in digital forms

- audio signal collected by a *microphone* is an analogue signal

  it must be converted into a digital signal for more reliable transmission
- **analogue-to-digital converter** (ADC) takes samples of the signal at regular time intervals

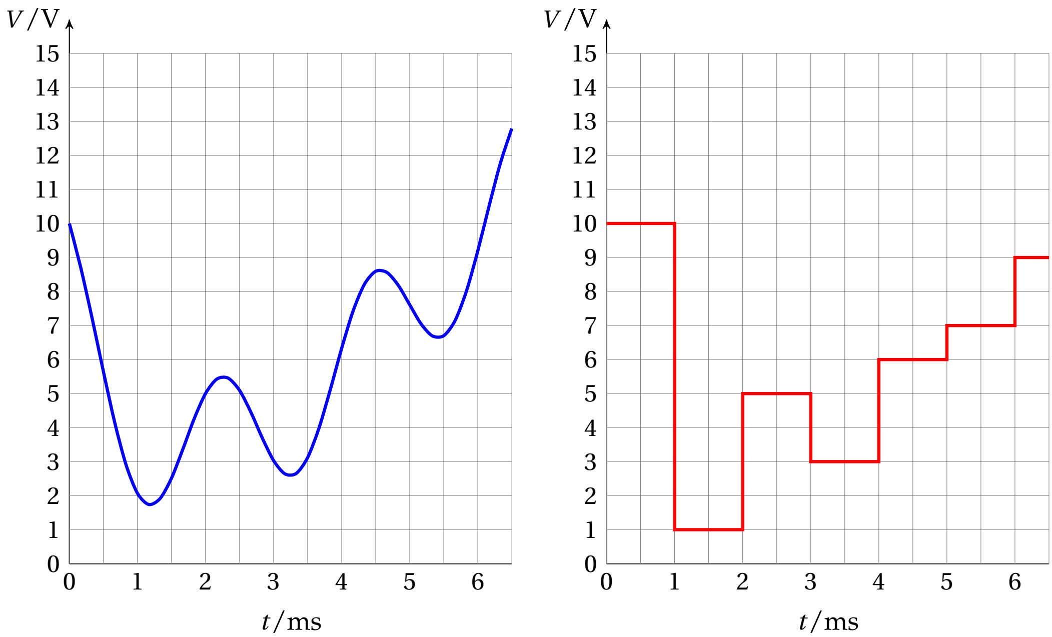
  each sample value is then converted into a binary number according to voltage levels
- **parallel-to-serial converter** takes all bits of a binary number produced by ADC

  each bit is then transmitted one after another
- at receiver end, the process is reversed to reproduce original audio signal
- quality of reproduction can be improved by two means:
  - increasing sampling frequency, so step width is reduced

    high-frequency components in original signal can be detected
  - increasing number of bits for the binary system, so step height is reduced

    smaller variations in voltages can be represented more precisely

**Example 16.5** An analogue signal is passed through a analogue-to-digital converter for digital transmission. Samples are taken at a frequency of 1.0 kHz. Each sample is then converted into a 4-bit binary number. (a) State the values of all samples taken for this signal. (c) Draw the pattern of the reproduced signal.




original analogue signal reproduced signal

✍ sampling frequency is 1.0 kHz, so samples are taken every 1.0 ms

four-bit binary system is implemented, so $2^4 = 16$ voltage steps are represented

these voltage levels are $0, 1, 2, \cdots, 15$ V

for a sample value between two voltage levels, it is rounded down

all samples taken for this signal can be summarised in the following table:

| $t$/ms | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| sample value/V | 10.0 | 2.0 | 5.0 | 3.0 | 6.3 | 7.6 | 9.2 |
| voltage level/V | 10 | 2 | 5 | 3 | 6 | 7 | 9 |
| digital number | 1010 | 0010 | 0101 | 0011 | 0110 | 0111 | 1001 |

these digital numbers are used to reproduce the analogue signal as shown □

## 16.3 communication channels

signals can be transmitted through various communication channels

we next look at relative advantages and disadvantages of each channel

### 16.3.1 wire pairs

a **wire pair** consists of two insulated wires twisted around each other[115]

- advantages of wire pairs: low cost, easy to install
- disadvantages of wire pairs
  - affected by noise (unwanted signals are easily picked up)
  - exist cross-linking (signal in one circuit easily induces a copy in a nearby circuit)
  - low security (easy to tap into a wire pair)
  - high attenuation (wire pair radiates electromagnetic waves when current in wire varies)
- applications: telephone to socket, amplifier to loudspeaker, etc.

### 16.3.2 coaxial cables

coaxial cable is a development of the wire pair

coaxial cable consists of a central core shielded by an outer conductor which is *earthed*

central core serves as the transmitting wire

outer conductor acts as signal return path and prevents interference getting into the core

- advantages of coaxial cables:
  - less noise and cross-linking (central core is shielded by outer conductor)
  - less attenuation (radiation of electromagnetic waves is prevented)
  - greater bandwidth, so higher rate of data transmission

[115]The idea behind the wire pair is that, by putting the two wires very close together, they are exposed to noise equally on average. So noise can be cancelled out at receiver by taking the difference signal. Compared with a single wire, twisting reduces interference, but still, wire pair is more affected by noise compared with other channels of communication.

  - higher security (more difficult to tap in than a wire pair)

➢ applications: television to aerial, computer to projector, landline to local exchange, etc.

### 16.3.3 optic fibres

**optic fibre** includes a core surrounded by a transparent cladding material

the core has a higher refraction index than the cladding, so this allows light to be kept in the core by *total internal reflection*

optic fibre can hence be used as a waveguide to transmit light signals

➢ advantages of optic fibres
  - very low attenuation (light is totally internally reflected, so it stays in core)
  - very little noise (light is immune to electromagnetic radiation)
  - very large bandwidth (light has much higher than frequencies than electrical signals)
  - ideal for digital transmission (light pulses can be switched on and off rapidly)
  - low weight and low cost (made of glass or plastic fibres)

➢ applications: internet broadband, transcontinental communication, etc.

### 16.3.4 radio waves

radio waves are part of electromagnetic spectrum with wavelengths $\lambda \approx 10$ cm ~ 10 km

radio waves can be used to transmit signals through different paths

➢ surface wave ($\lambda \approx 0.1 \sim 10$ km, $f \approx 30$ kHz ~ 3 MHz)

surface waves travel close to surface of earth

long wavelength means they can diffract around obstacles

➢ sky wave ($\lambda \approx 10 \sim 100$ m, $f \approx 3 \sim 30$ MHz)

sky waves can be reflected by the *ionosphere* (a layer of charged particles in atmosphere)

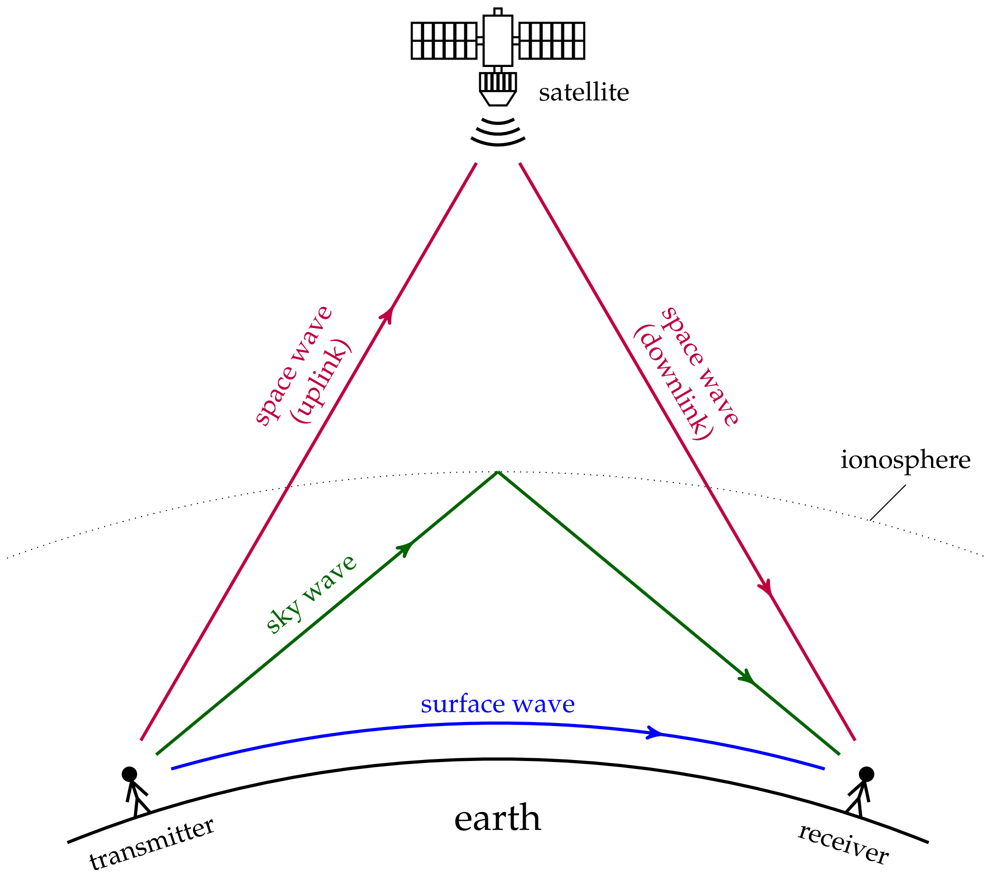
sky waves travel over large distances via multiple reflections by ionosphere and ground

sky wave depends on atmosphere conditions so reception is not very reliable

➢ space wave ($\lambda \approx 10$ cm ~ 10 m, $f \approx 30$ MHz ~ 3 GHz)

short-wave radio waves can penetrate atmosphere and reach satellites satellite can regenerate the signal and re-transmit them back to receiver

space waves are also used on surface of earth, but their short wavelength means they cannot diffract sufficiently, so transmission to receiver has to be *line-of-sight*




use of radio waves in telecommunication

**communication satellites**

➢ communication by satellite has several advantages compared with sky waves
  - satellite amplifies the signal for its return to surface, reception signal is stronger
  - space wave uses higher frequencies, this allows for greater bandwidth
    so more information can be transmitted per unit time
  - sky wave lacks stability, but communication by satellite is consistent

➢ uplink and downlink for satellite communication must use different frequencies

since uplink is greatly attenuated, downlink must be greatly amplified

to avoid uplink being swamped by downlink signals, different carrier frequencies are used

➢ one type of useful satellites are the *geostationary satellites*

geostationary satellite always stays in same position with respect to a ground observer

it has a period of 24 hours, moves from west to east in an equatorial orbit

- advantages

  relative at rest, so no need to track, satellite dish can be fixed

- disadvantages

  orbit is very high (radius of orbit is over 6 times of radius of earth), so long time delay

  orbit is above equator, so cannot cover areas of high latitudes

➢ another type of useful satellites are the *polar satellites*

polar satellites pass over the two poles of the earth at around 1000 km above ground

- advantages

  lower orbit, so shorter time delay

  give better details for surface observation (e.g., street map, weather forecast, spy, etc.)

- disadvantages

  not synchronous with ground observers, so satellite needs to be tracked

  channel must be swapped between satellites for continuous communication

#### 16.3.5 microwaves

microwaves have higher frequencies ($f \approx 3 \sim 300$ GHz) than radio waves [116]

high frequency means higher bandwidth, greater rate of data transmission is possible

but their wavelengths are shorter ($\lambda \approx 0.1 \sim 10$ cm), so not very able to diffract around barriers, transmission must be line-of-sight

➢ application: Wi-Fi, bluetooth connections, mobile phone networks, satellite links, etc.

[116]There is no clear boundary of frequency (or wavelength) between radio waves and microwaves. Usually we take a few GHz (or a few cm) as the borderline.

**causes of attenuation in communication channels**

each communication channel suffers some degree of power loss

- wire pair/coaxial cable: heat loss due to resistance and radiation of energy
- optic fibre: scattering due to impurities and defects in optic fibre[117]
- radio waves/microwaves: absorption by atmosphere/air molecules

## 16.4 the decibel scale

in telecommunication, engineers often want to evaluate the following:

- how much signal power is lost during transmission
- how the received power compares with noise level
- by what extent a signal is intensified if it is sent into an amplifier

these all involve computing ratios of powers of some kind, which are usually very large

engineers introduce a logarithmic scale to express power ratios, known as the **decibel scale**

with the decibel scale, power ratio $\frac{P_1}{P_2}$ is now expressed as $\boxed{10\log\frac{P_1}{P_2}}$[118]

in electrical engineering, signal attenuation (degree of power decrease), signal-to-noise ratio and gain of amplifer (value of amplification) are usually given in decibels:

- $\boxed{\text{attentuation} = 10\log\frac{P_{\text{in}}}{P_{\text{out}}}}$
- $\boxed{\text{signal-to-noise ratio} = 10\log\frac{P_{\text{out}}}{P_{\text{noise}}}}$
- $\boxed{\text{gain of amplifier} = 10\log\frac{P_{\text{out}}}{P_{\text{in}}}}$

➢ unit for the decibel scale: dB

ratio in dB is actually unit free

dB is simply a reminder that we are using a logarithmic scale to express a ratio

[117]In practice, infra-red is usually used in optic fibres rather than visible light. Infra-red has longer wavelength, so they have better diffraction ability. Less scattering effect means attenuation is lower.

[118]This is the logarithm to base 10, which can also be written as $\log_{10}$ or lg.

- rationales for using the decibel scale
  - ratios given in dB give more manageable numbers
  - ratios given in dB can be added or subtracted [119]

**Example 16.6** A signal of power 0.040 mW is sent into an amplifier. The gain of the amplifier is 36 db. What is the power for the output signal?

✍ $$10\log\frac{P_{\text{out}}}{P_{\text{out}}}=36 \quad\Rightarrow\quad 10\log\frac{P_{\text{out}}}{0.040}=3.6 \quad\Rightarrow\quad P_{\text{out}}=10^{3.6}\times 0.040\approx 159\text{ mW}$$ □

**Example 16.7** An optic fibre is known to have an attenuation of 0.12 dB km$^{-1}$. The transmitted signal sent into the optic fibre has a power of 50 mW. The noise level at receiver end is 8.0 $\mu$W. If the signal-to-noise ratio at receiver is at least 25 dB, what is the maximum length of the fibre?

✍ let's first find minimum reception power from the signal-to-noise ratio:

$$10\log\frac{P_{\text{out}}}{P_{\text{noise}}}=25 \quad\Rightarrow\quad \log\frac{P_{\text{out}}}{8.0\times 10^{-6}}=2.5 \quad\Rightarrow\quad P_{\text{out}}=10^{2.5}\times 8.0\times 10^{-6}\approx 2.53\times 10^{-3}\text{ W}$$

maximum attenuation in channel $=10\log\dfrac{P_{\text{in}}}{P_{\text{out}}}=10\log\dfrac{50\times 10^{-3}}{2.53\times 10^{-3}}\approx 130$ dB

maximum length of optic fibre: $L=\dfrac{130}{0.12}\approx 108$ km □

**Example 16.8** An electrical signal of 60 mW is transmitted along a cable of length 94 km. The cable has an attenuation of 3.2 dB km$^{-1}$. Repeater amplifiers of gain of 27 dB are placed every 8.0 km along the cable. What is the power of the received signal?

✍ total attenuation in cable: $94\times 3.2=300.8$ dB

number of amplifiers: $\dfrac{94}{8.0}=11.75\Rightarrow 11$ amplifiers are placed

total gain due to amplifiers: $11\times 27=297$ dB

so net attenuation during transmission: $300.8-297=3.8$ dB

$$10\log\frac{P_{\text{in}}}{P_{\text{out}}}=3.8 \quad\Rightarrow\quad \log\frac{60}{P_{\text{out}}}=0.38 \quad\Rightarrow\quad P_{\text{out}}=\frac{60}{10^{0.38}}\approx 25\text{ mW}$$ □

**Question 16.5** A signal is transmitted through a cable of total length of 5.0 km,. The signal attenuation per unit length is this cable is 12 dB km$^{-1}$. If the reception power is $2.9\times 10^{-7}$ W,

[119]This follows from the property of the logarithmic function: $\log x+\log y=\log(xy)$.

Note that: $10\log\dfrac{P_1}{P_2}+10\log\dfrac{P_2}{P_3}=10\log\left(\dfrac{P_1}{P_2}\times\dfrac{P_2}{P_3}\right)=10\log\dfrac{P_1}{P_3}$. This means if we know the power ratio of $P_1$ to $P_2$ in dB and the power ratio of $P_2$ to $P_3$ in dB, we can immediately add the two numbers together to get the the power ratio of $P_1$ to $P_3$ in dB.

find the power of the input signal.

**Question 16.6** An input signal of power 28 mW is passed along two cables followed by an amplifier at the receiver end for output. If the attenuation in the cables are 29 dB and 38 db respectively, and the amplifier has a gain of 43 dB. What is the output power from the amplifier?

**Question 16.7** Input signal of power 12 mW is sent along an optic fibre. The noise level at the receiver end of the optic fibre is 0.75 $\mu$W. It is required that the signal-to-noise ratio in the receiver must not fall below 35 dB. For this transmission, the maximum uninterrupted length of the optic fibre is 48 km. (a) What is the minimum acceptable power of the signal at the receiver? (b) What is the attenuation per unit length of the fibre?

# Index

absolute zero, 54

acoustic impedance, 167

activity, 147

amount of substance, 33

amplifier

- comparator, 194
- inverting amplifier, 197
- negative feedback, 196
- non-inverting amplifier, 198
- op-amp, 192

amplitude, 20

angular displacement, 1

angular frequency, 22, 108

angular velocity, 1

annihilation, 173

attenuation, 158, 216

Avogadro constant, 34

bandwidth, 212

Big Bang, 184

binding energy, 141

black body radiation, 179

Boyle's law, 37

braking radiation, 157

Brownian motion, 33

capacitance

- mutual capacitance, 75
- self-capacitance, 76

capacitor, 74, 116

centripetal acceleration, 3

centripetal force, 4

characteristic radiation, 157

Charles's law, 38

coaxial cable, 220

comsmic microwave background, 186

Coulomb's law, 57

count rate, 151

CT imaging, 161

cyclotron, 95

damped oscillation, 29

- critical damping, 30
- heavy damping, 30
- light damping, 29

dark energy, 187

dark matter, 187

de Broglie wavelength, 127

decay constant, 147

decibel scale, 224

digital signal, 215

diode, 114, 202
dispersion, 131
Doppler effect, 182

electric field
- electric field strength, 58
- electric potential, 63
- electric potential energy, 61

electromagnetic induction, 101
electron diffraction, 128
energy band, 136
- band gap, 137
- conduction band, 137
- forbidden band, 137
- valence band, 137

energy level, 131
equipotential lines, 65
escape velocity, 17

farad, 75
Faraday's law, 101
field line
- electric field line, 60
- gravitaional field line, 14
- magnetic field line, 84

fission, 144
forced oscillation, 31
fusion, 143

Gay-Lussac's law, 38
generator, 106
geostationary orbit, 10
GM tube, 151
gravitational field
- gravitational field strength, 12
- gravitational potential, 18

half-life, 149
Hall effect, 98
Hubble's law, 183

ideal gas, 35
internal energy, 48

kinetic model of gases, 39

Larmor frequency, 170
law of thermodynamics
- first law of thermodynamics, 51
- second law of thermodynamics (∗), 54
- third law of thermodynamics (∗), 55
- zeroth law of thermodynamics, 54

LDR, 189
LED, 202
left-hand rule, 87, 92
Lenz's law, 101
Lorentz force, 92
luminosity, 176

magnetic field
- magnetic flux, 100

magnetic flux density, 88
magnetic force, 87, 91
magnets, 84
mass defect, 141
mass spectrometer, 94
mass-energy relation, 140
modulation, 210
amplitude modulation, 211
frequency modulation, 214
molar mass, 34

negative feedback, 196
Newton's law of gravitation, 8
noise, 216
nuclear magnetic resonance, 169

optic fibre, 221
oscilloscope, 111

PET imaging, 173
phase angle, 20
photoelectric effect, 120
photoelectric equation, 123
threshold frequency, 121, 123
work function, 123
photon, 122
piezo-electric transducer, 190
Planck constant, 122
positron, 173
potential divider, 191

r.m.s. current, 110
r.m.s. voltage, 110
radian, 1
radiant flux intensity, 177
radio wave, 221
radioactive decay, 147
rectification, 114
full-wave rectification, 115
half-wave rectification, 115
redshift, 183
regeneration, 216
relaxation, 171
relay, 202
resonance, 31
restoring force, 21
right-hand rule, 85, 86

simple harmonic motion, 22
smoothing, 116
specific charge, 94
specific heat capacity, 46
specific latent heat, 45
spectrum, 131
absorption spectrum, 133
emission spectrum, 132
X-ray spectrum, 156
standard candle, 177
cepheid variable, 178

    supernova, 178

Stefan-Boltzmann law, 180

strain gauge, 189

tesla, 88

thermal equilibrium, 54

thermistor, 188

thermocouple, 55

thermometer, 55

transformer, 113

velocity selector, 97

voltage comparator, 194

wave-particle duality, 129

Wien's law, 179

X-ray

    spectrum, 156

    X-ray tube, 155