REVISION HANDOUTS 复习讲义

AS-Level Handouts

Every AS topic across all eight subjects.

八门学科的全部 AS 单元。

Bilingual study handouts • 74 topics.

Contents 目录

AS-Level Physics 5

1 Physical quantities and units 5

2 Kinematics 18

3 Dynamics 30

4 Forces, density and pressure 42

5 Work, energy and power 52

6 Deformation of solids 61

7 Waves 70

8 Superposition 83

9 Electricity 94

10 D.C. circuits 105

11 Particle physics 115

AS-Level Chemistry 125

1 Atomic structure 125

2 Atoms, molecules and stoichiometry 138

3 Chemical bonding 147

4 States of matter 157

5 Chemical energetics 165

6 Electrochemistry 171

7 Equilibria 176

8 Reaction kinetics 185

9 The Periodic Table: chemical periodicity 190

10 Group 2 196

11 Group 17 201

12 Nitrogen and sulfur 207

13 An introduction to AS Level organic chemistry 213

14 Hydrocarbons 221

15 Halogen compounds 229

16 Hydroxy compounds 235

17 Carbonyl compounds 240

18 Carboxylic acids and derivatives 245

19 Nitrogen compounds 249

20 Polymerisation 252

21 Organic synthesis 256

22 Analytical techniques 262

AS-Level Biology 267

1 Cell structure 267

2 Biological molecules 276
3 Enzymes 287
4 Cell membranes and transport 297
5 The mitotic cell cycle 305
6 Nucleic acids and protein synthesis 312
7 Transport in plants 321
8 Transport in mammals 329
9 Gas exchange 338
10 Infectious diseases 343
11 Immunity 348

AS-Level Computer Science 359

1 Information representation 359
2 Communication 372
3 Hardware 386
4 Processor Fundamentals 405
5 System Software 418
6 Security, privacy and data integrity 428
7 Ethics and Ownership 436
8 Databases 443
9 Algorithm Design and Problem-solving 452
10 Data Types and Structures 461
11 Programming 472
12 Software Development 484

AS-Level Economics 494

1 Basic economic ideas and resource allocation 494

2 The price system and the microeconomy 507
3 Government microeconomic intervention 522
4 The macroeconomy 532
5 Government macroeconomic intervention 543
6 International economic issues 552

AS-Level Business 560

1 Business and its environment 560
2 Human resource management 573
3 Marketing 582
4 Operations management 594
5 Finance and accounting 602

AS-Level Mathematics 612

1 Pure Mathematics 1 612
2 Pure Mathematics 2 630
4 Mechanics 639
5 Probability & Statistics 1 647

AS-Level Further Mathematics 655

1 Further Pure Mathematics 1 655
3 Further Mechanics 662
4 Further Probability & Statistics 670

Physical quantities and units

A-Level Physics

Physical quantities

A **physical quantity** 物理量 has two parts: a number (its **magnitude** 大小) and a **unit** 单位. The number on its own tells you nothing. You must also say what is measured and in which unit.

Example: "the length is 1.5" is not complete. "The length is 1.5 m" is a physical quantity.

Making estimates

You should be able to **estimate** 估算 the size of the physical quantities in this syllabus. Learn these rough values:

- **mass** 质量 of an adult human: ~ 70 kg
- **weight** 重力 of an adult human: ~ 700 N
- height of an adult human: ~ 1.7 m
- mass of an apple: ~ 0.1 kg (so its weight is about 1 N)
- speed of sound in air: ~ 340 m s^{-1}
- speed of light in a **vacuum** 真空: 3.0×10^8 m s^{-1}
- **acceleration** 加速度 of **free fall** 自由落体: $g \approx 9.81$ m s^{-2}

A good estimate has the right **order of magnitude** 数量级 (the right power of ten). For a human, 70 kg is a good guess; 7 kg is not.

SI units 国际单位制

Base units

The SI system has seven **base quantities** 基本量; five of them are used in this topic. You must know these five and their units:

- **mass** — kilogram, kg
- **length** 长度 — metre, m
- time — second, s
- **current** 电流 — ampere 安培, A
- **temperature** 温度 — kelvin 开尔文, K

Every other unit in this syllabus is built from these five.

Derived units

A **derived unit** 导出单位 is made by multiplying or dividing base units. You should be able to write any quantity in this syllabus in base units.

Build a derived unit from the equation that defines it:

- **speed** 速率 = distance / time, so its unit is m s^{-1}
- **acceleration** = change in **velocity** 速度 / time, so its unit is m s^{-2}
- **force** 力 = mass × acceleration, so its unit is kg m s^{-2}. The **newton** 牛顿 is 1 N = 1 kg m s^{-2}.
- **work** 功 and **energy** 能量 = force × distance, so the unit is kg m^2 s^{-2}. The **joule** 焦耳 is 1 J = 1 kg m^2 s^{-2}.
- **power** 功率 = energy / time, so the unit is kg m^2 s^{-3}. The **watt** 瓦特 is 1 W = 1 kg m^2 s^{-3}.
- **pressure** 压强 and **stress** 应力 = force / area, so the unit is kg m^{-1} s^{-2}. The **pascal** 帕斯卡 is 1 Pa = 1 kg m^{-1} s^{-2}.

When a question asks for the SI base units of a quantity, replace each named unit with its base units, then simplify. Example: the SI base units of the watt are kg m^2 s^{-3}.

Checking that the units match

An equation is **homogeneous** 量纲一致 when both sides have the same base units. In plain words: the units on both sides match.

Write each side in base units and compare. Take the equation $v^2 = u^2 + 2as$:

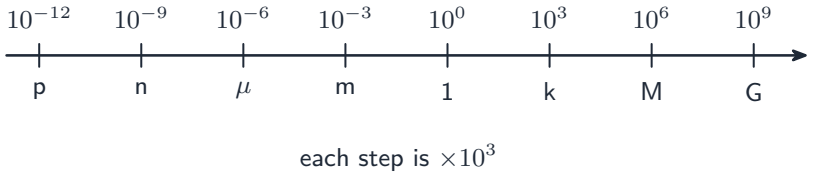
- left side: $(\text{m s}^{-1})^2 = \text{m}^2\ \text{s}^{-2}$
- right side, first term: $(\text{m s}^{-1})^2 = \text{m}^2\ \text{s}^{-2}$
- right side, second term: $\text{m s}^{-2} \cdot \text{m} = \text{m}^2\ \text{s}^{-2}$

Both sides give $\text{m}^2\ \text{s}^{-2}$, so the units match.

Be careful: matching units do **not** prove the whole equation is correct. It could still have a wrong number, or a missing factor of 2. But if the units do **not** match, the equation is wrong for sure.

Prefixes

A **prefix** 词头 is a letter put in front of a unit to make it bigger or smaller by powers of ten. You must know these:



The SI prefixes climb in steps of a thousand, from pico to giga

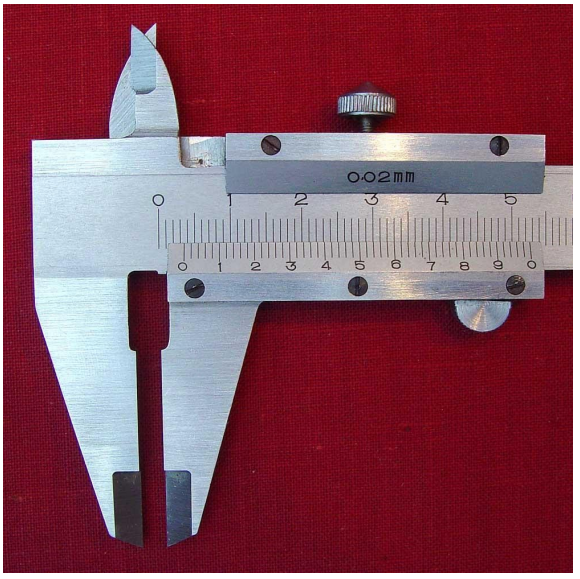
Prefix	Symbol	Factor
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}

To change a prefixed unit into base units, replace the prefix with its factor, then simplify. Example: change 0.25 kN mm^{-2} into N m^{-2}:

$$0.25\ \text{kN mm}^{-2} = 0.25 \times \frac{10^3\ \text{N}}{(10^{-3}\ \text{m})^2} = 0.25 \times \frac{10^3}{10^{-6}}\ \text{N m}^{-2} = 2.5 \times 10^8\ \text{N m}^{-2}.$$

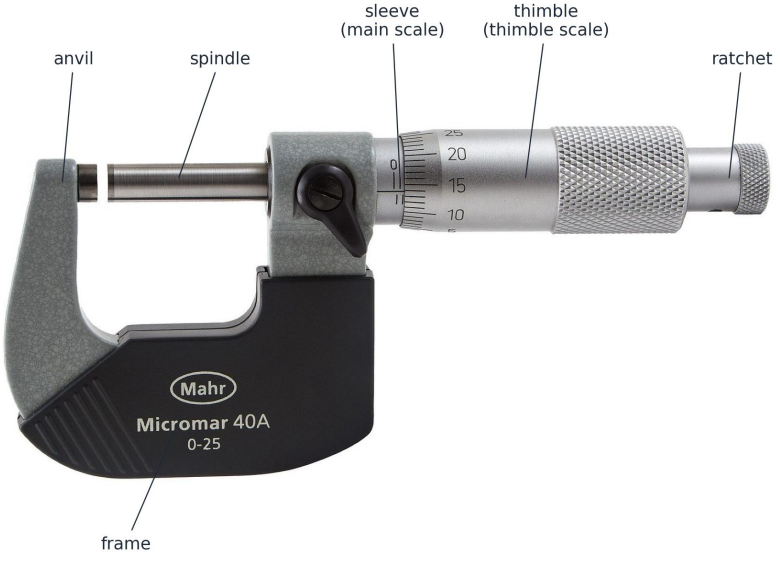
Take special care with squared units like mm^2: you must square the factor too.

Errors and uncertainties

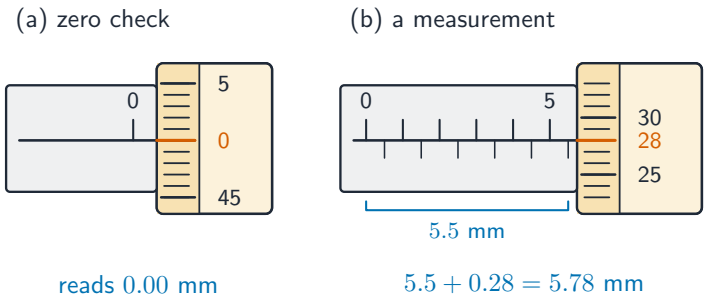


A vernier caliper measures length precisely, with a small uncertainty.

Every **measurement** 测量 has some **uncertainty** 不确定度 — we are never fully sure of the value. A good experimenter knows where the uncertainty comes from, makes a fair estimate of it, and carries it through to the final answer.



The main parts of a real micrometer screw gauge, which measures to the nearest 0.01 mm



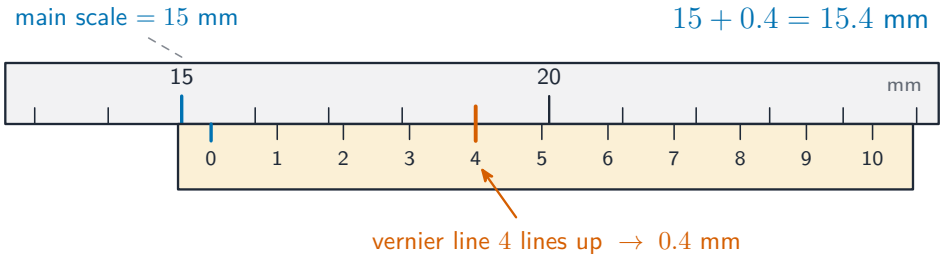
Reading a micrometer: add the main scale reading to the thimble reading



Reading a real micrometer: read the mm and half-mm on the sleeve, then add the thimble scale



Vernier calipers measure to the nearest 0.1 mm — the sliding scale gives the extra digit

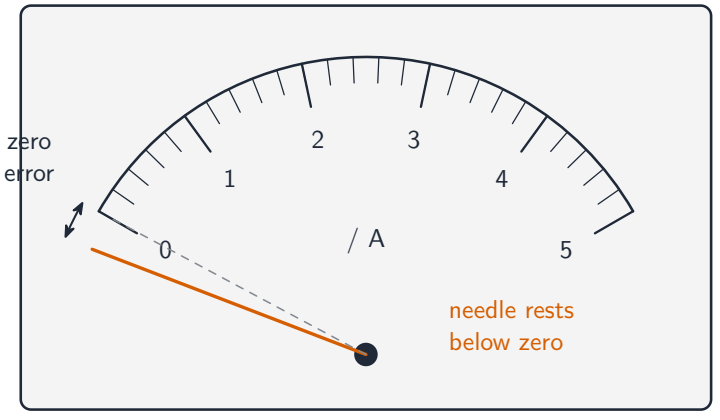


Reading a vernier caliper: whole millimetres from the main scale, plus the tenths from the vernier line that lines up

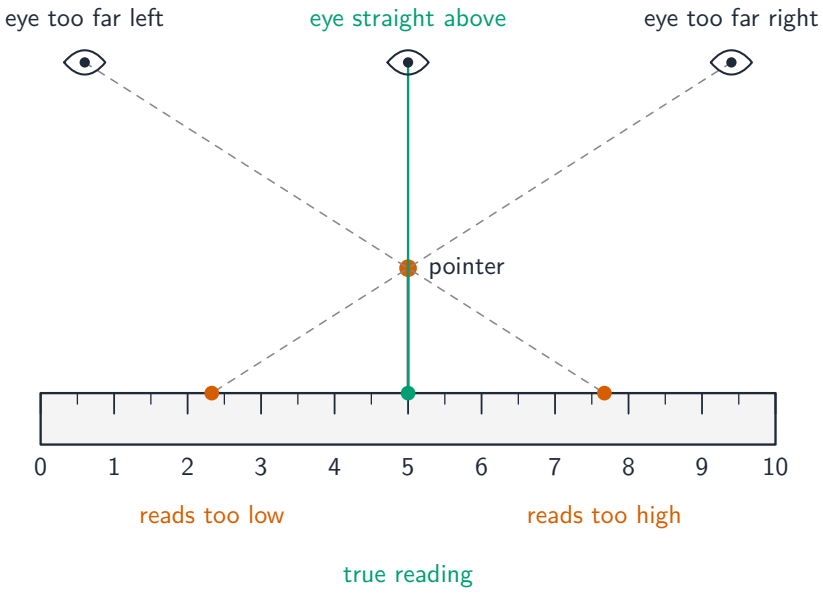
Systematic and random errors

A **systematic error** 系统误差 changes every reading by the same amount, in the same direction. You cannot find it by repeating the measurement. Common causes:

- a **zero error** 零点误差 (the scale does not read zero when the true value is zero)
- a **calibration** 校准 error (the scale itself is wrong)
- **parallax** 视差 (your eye is always to one side of the scale)



An ammeter with a zero error: the needle reads below zero before any current flows



Parallax error: different viewing angles give different scale readings

A systematic error makes the **accuracy** 准确度 worse, but it does not change the **precision** 精密度.

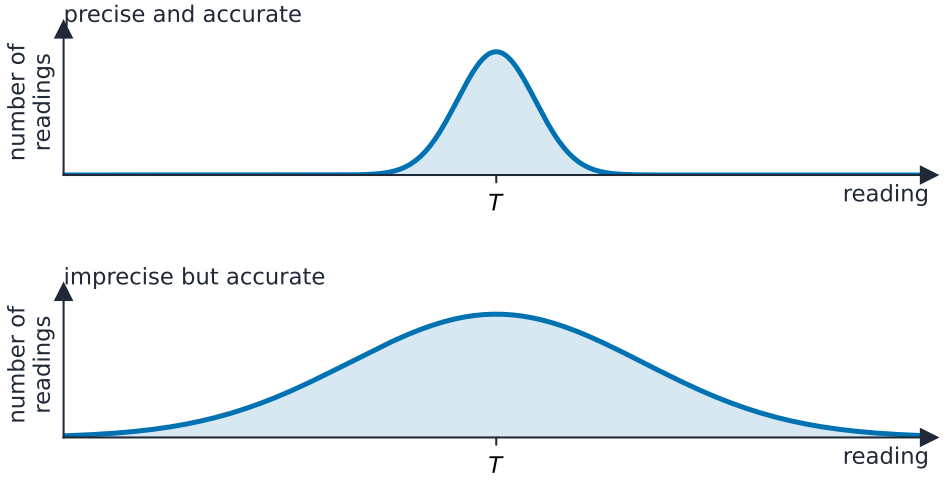
A **random error** 随机误差 makes readings jump above and below the true value, with no pattern. Causes include how carefully you read the scale, changing conditions, and the smallest step the **instrument** 仪器 can show. If you repeat the measurement many times and take the **mean** 平均值 (the average), random errors partly cancel out.

A random error makes the precision worse. But with enough repeats, the mean can still be accurate.

Precision and accuracy

Precision is how close repeated readings are to each other. Precise readings are grouped very close together.

Accuracy is how close a reading (or the mean of several readings) is to the true value.

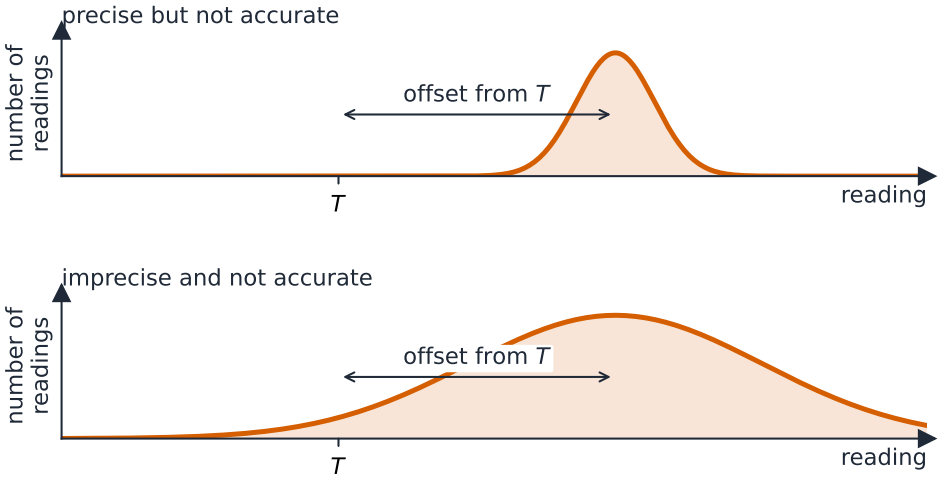


precision: how closely repeated readings agree (random error / spread)

Precision: how narrow the distribution is around the true value T

A set of readings can be:

- precise and accurate — close together and near the true value
- precise but not accurate — close together, but away from the true value (a systematic error)
- accurate but not precise — spread out, but the mean is near the true value
- neither — spread out and away from the true value



accuracy: how close readings are to the true value (systematic error reduces accuracy)

Accuracy: whether the peak of the distribution is centred on the true value T

When a question gives a table of repeated readings, look at the **spread** (precision) and the **mean** (accuracy) separately.

Uncertainty in a derived quantity

A measurement is often written as $x \pm \Delta x$. Here Δx is the **absolute uncertainty** 绝对不确定度. The **percentage uncertainty** 百分比不确定度 is

$$\text{percentage uncertainty in } x = \frac{\Delta x}{|x|} \times 100\%.$$

A **derived quantity** 导出量 is one you calculate from measured values. Its uncertainty is found by simple rules:

- **Adding or subtracting** — add the absolute uncertainties. If $y = a+b$ or $y = a-b$, then $\Delta y = \Delta a + \Delta b$.
- **Multiplying or dividing** — add the percentage uncertainties. If $y = \dfrac{a \cdot b}{c}$, then

$$\frac{\Delta y}{|y|} = \frac{\Delta a}{|a|} + \frac{\Delta b}{|b|} + \frac{\Delta c}{|c|}.$$

- **Powers** — multiply the percentage uncertainty by the power. If $y = a^n$, then $\dfrac{\Delta y}{|y|} = |n| \cdot \dfrac{\Delta a}{|a|}$.

Worked example. A ball's **diameter** 直径 is measured as $d = (5.26 \pm 0.02)$ cm. The **volume** 体积 of a sphere is $V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(d/2)^3$, so $V \propto d^3$ (V depends on d cubed).

The percentage uncertainty in d is

$$\frac{0.02}{5.26} \times 100\% \approx 0.38\%.$$

Because $V \propto d^3$, the percentage uncertainty in V is three times this, about 1.14%. The volume is $\frac{4}{3}\pi(2.63)^3 \approx 76.2$ cm^3. So the absolute uncertainty is $0.0114 \times 76.2 \approx 0.87$ cm^3. The final answer is $V = (76.2 \pm 0.9)$ cm^3.

Significant figures

When you write a calculated quantity, give it the **same number of significant figures** 有效数字 as the least precise measurement you used — usually two or three in this syllabus. Too many significant figures makes the answer look more exact than it really is. Too few loses useful information.

Scalars and vectors

A **scalar** 标量 has size only. A **vector** 矢量 has both size and direction.

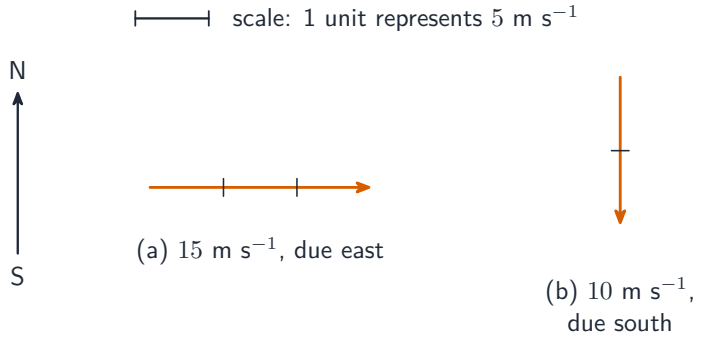
Examples from the syllabus:

- scalars: mass, time, temperature, energy, work, power, distance, speed, pressure, **density** 密度, **electric charge** 电荷
- vectors: **displacement** 位移, velocity, acceleration, force (including weight), **momentum** 动量

Quick test: if it makes sense to ask "in which direction?", the quantity is a vector. You cannot ask "in which direction is the temperature?", so temperature is a scalar. You can ask "in which direction is the velocity?", so velocity is a vector.

Adding and subtracting vectors

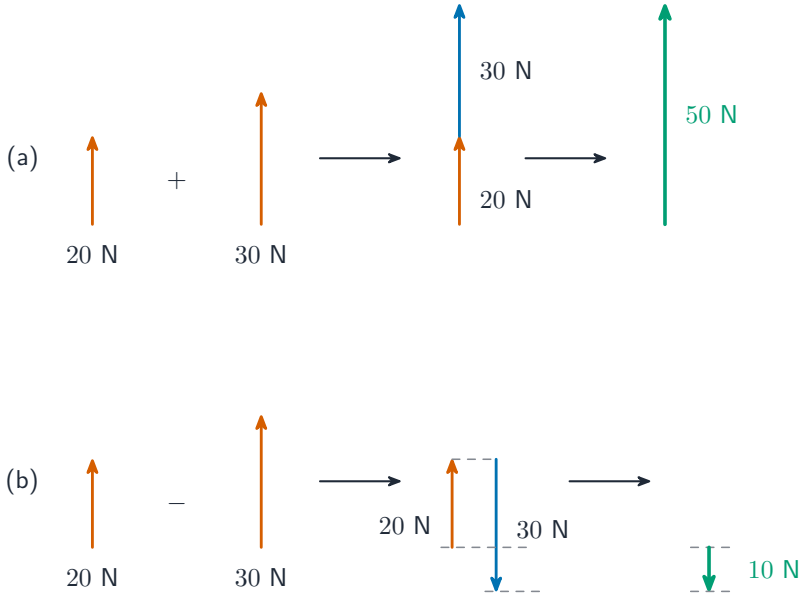
A vector is drawn as an arrow: the **direction** of the arrow gives the direction of the quantity, and the **length** of the arrow (drawn to scale) gives the magnitude.



Vectors represented as arrows drawn to scale

To add two **coplanar** 共面 vectors (vectors in the same flat plane), draw them **tip to tail**. The **resultant** 合矢量 goes from the tail of the first arrow to the tip of the second.

To find $\vec{X} - \vec{Y}$, add the reverse of $\vec{Y}$: $\vec{X} + (-\vec{Y})$. The reverse of $\vec{Y}$ has the same size as $\vec{Y}$ but points the opposite way.



Adding and subtracting parallel vectors

If the two vectors are at right angles (90°), the size of the resultant is

$$|\vec{R}| = \sqrt{X^2 + Y^2},$$

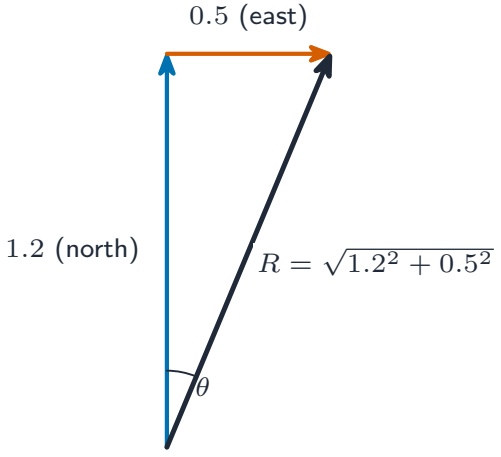
and its direction comes from $\tan\theta = Y/X$.

Worked example. A swimmer heads north at 1.2 m s^{-1} across a river that flows east at 0.5 m s^{-1}. Find the size and direction of the resultant velocity.

The two velocities are perpendicular, so

$$|\vec{R}| = \sqrt{1.2^2 + 0.5^2} = \sqrt{1.69} = 1.3 \text{ m s}^{-1},$$

at an angle $\tan\theta = 0.5/1.2$, giving $\theta \approx 23°$ east of north.



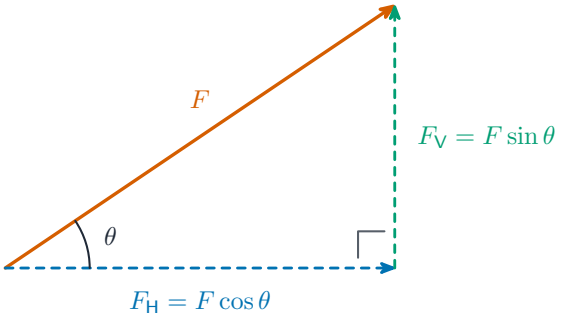
Two perpendicular vectors add to a resultant of size $\sqrt{X^2 + Y^2}$ at angle θ

If two vectors have the same size F with an angle 2α between them, the resultant has size $2F\cos\alpha$ and lies along the line that cuts the angle in half.

Splitting a vector into perpendicular parts

Any vector can be split into two **perpendicular** 垂直 (at right angles) **components** 分量. Usually these are **horizontal** 水平 and **vertical** 竖直, or along and across a surface. For a vector $\vec{v}$ at angle θ to the horizontal:

$$v_{\text{H}} = v\cos\theta, \qquad v_{\text{V}} = v\sin\theta.$$



Resolving a vector into horizontal and vertical components

Choose the directions that make the problem easiest. On a slope (an **inclined plane** 斜面), split the weight into one part along the slope and one part at right angles to it:

$$W_{\parallel} = W \sin\theta, \qquad W_{\perp} = W \cos\theta,$$

where θ is the angle of the slope to the horizontal.

You split a vector into components whenever you need to know how much of it acts in one direction. For example: the part of a force that acts along a slope, or the horizontal and vertical parts of a ball's velocity after it is thrown.

Exam tips

- Give every answer a **unit**, and check **homogeneity** — both sides of an equation must have the same base units.
- Distinguish **random error** (reduce by repeating and averaging) from **systematic error** (a zero or calibration error that repeats do not remove).
- Combine uncertainties: **add absolute** uncertainties when adding/subtracting, **add percentage** uncertainties when multiplying/dividing (and multiply the % by any power).
- Distinguish **precision** (small spread) from **accuracy** (close to the true value).
- Resolve a vector into perpendicular components ($F\cos\theta$, $F\sin\theta$); add vectors tip-to-tail or by components.

Kinematics

A-Level Physics

Key definitions



A speedometer shows speed: the distance travelled per unit time.

These five quantities come up in almost every **kinematics** 运动学 question. Learn the exact words — the examiner gives marks for precise wording.

- **distance** 距离 — the total length of the path travelled. A **scalar** 标量.
- **displacement** 位移 — the straight-line distance from the start to the end, with a direction. A **vector** 矢量.
- **speed** 速率 — the rate of change of distance with time. A scalar.
- **velocity** 速度 — the rate of change of displacement with time. A vector.
- **acceleration** 加速度 — the rate of change of velocity with time. A vector.

The unit of speed and velocity is m s^{-1}; the unit of acceleration is m s^{-2}.

A common mistake: **deceleration** 减速度 just means acceleration in the opposite direction to the velocity. It is not a separate quantity.

Motion graphs



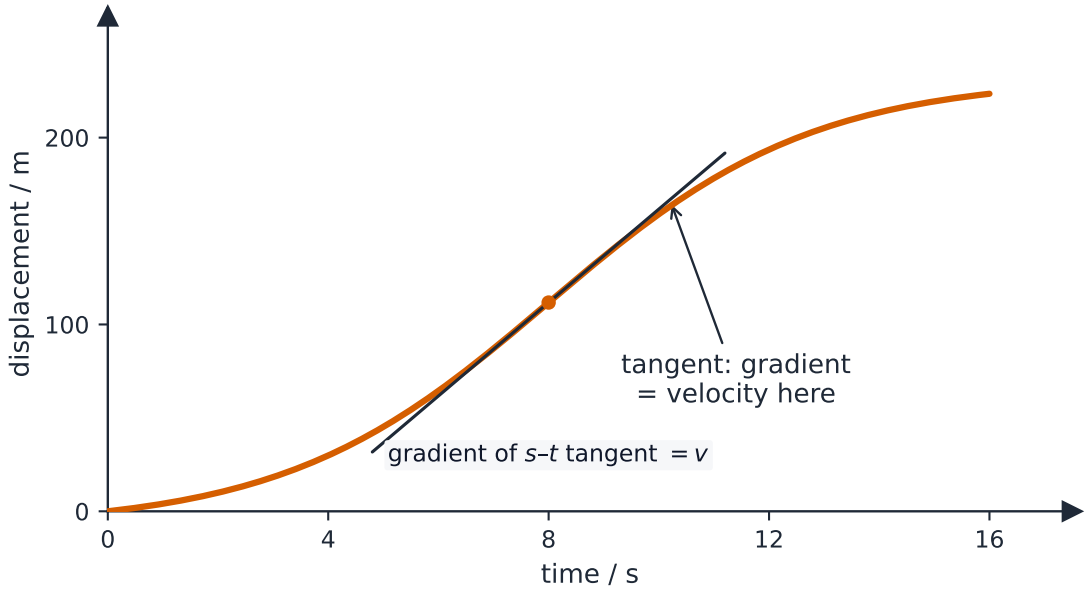
A high-speed train: its motion can be shown on a distance-time graph.

Many marks come from reading or drawing motion graphs. Two graphs matter.

Displacement–time graph

The **gradient** 斜率 (steepness) of a displacement–time graph at a point gives the velocity at that moment.

- flat line → the object is at rest.
- straight sloping line → constant velocity (gradient = velocity).
- curved line → changing velocity. Draw a **tangent** 切线 at the point and find its gradient.



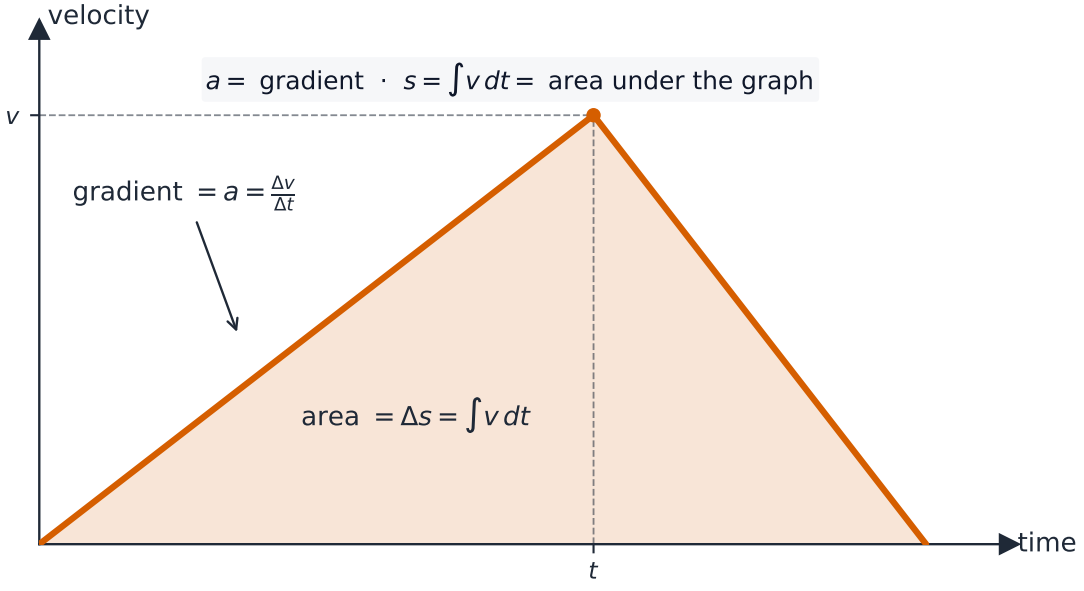
Displacement–time graph of a car on a test track

Velocity–time graph

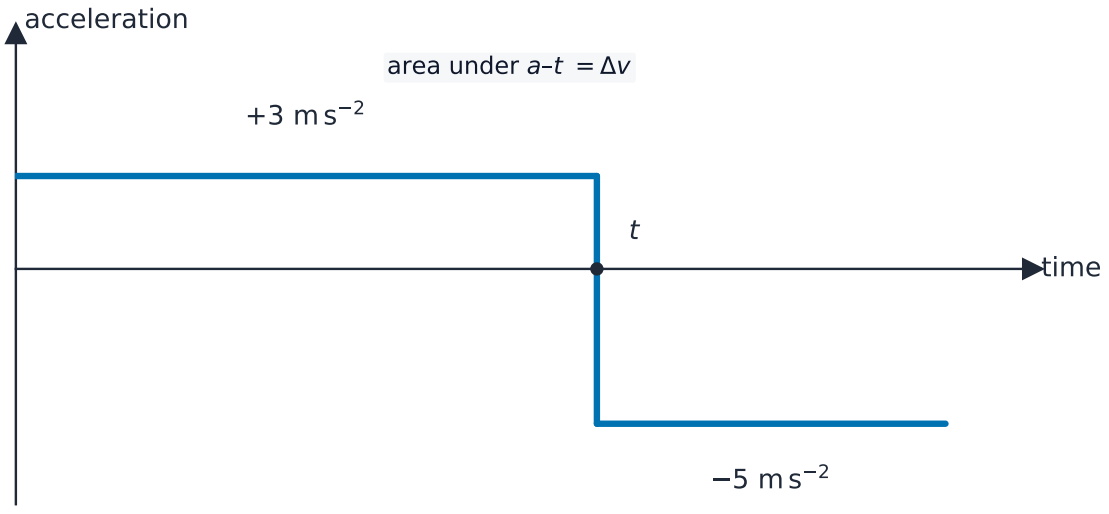
The gradient of a velocity–time graph gives the acceleration at that moment.

The **area** between the line and the time axis gives the displacement in that time.

- flat line → constant velocity (zero acceleration).
- straight sloping line → **uniform acceleration** 匀加速 (constant acceleration).
- curved line → changing acceleration.
- area above the time axis is positive displacement; area below is negative (the object moved backwards).



Velocity–time graph — gradient gives acceleration, area gives displacement



Acceleration–time graph derived from the same motion

To find the displacement, split the area into triangles and rectangles, or count grid squares. Area of a triangle is $\frac{1}{2} \times$ base $\times$ height; area of a rectangle is base $\times$ height.

The four SUVAT equations

For motion in a straight line with uniform acceleration, we use five symbols: starting velocity u, final velocity v, acceleration a, displacement s, and time t. Four equations link them:

$$v = u + at$$

$$s = ut + \tfrac{1}{2}at^2$$

$$s = \tfrac{1}{2}(u + v)t$$

$$v^2 = u^2 + 2as$$

Each equation uses four of the five symbols. To pick the right one: write down what you know and what you want, then choose the equation with exactly those four.

Worked example. A car accelerates uniformly from 8 m s^{-1} to 20 m s^{-1} over a distance of 56 m. Find its acceleration.

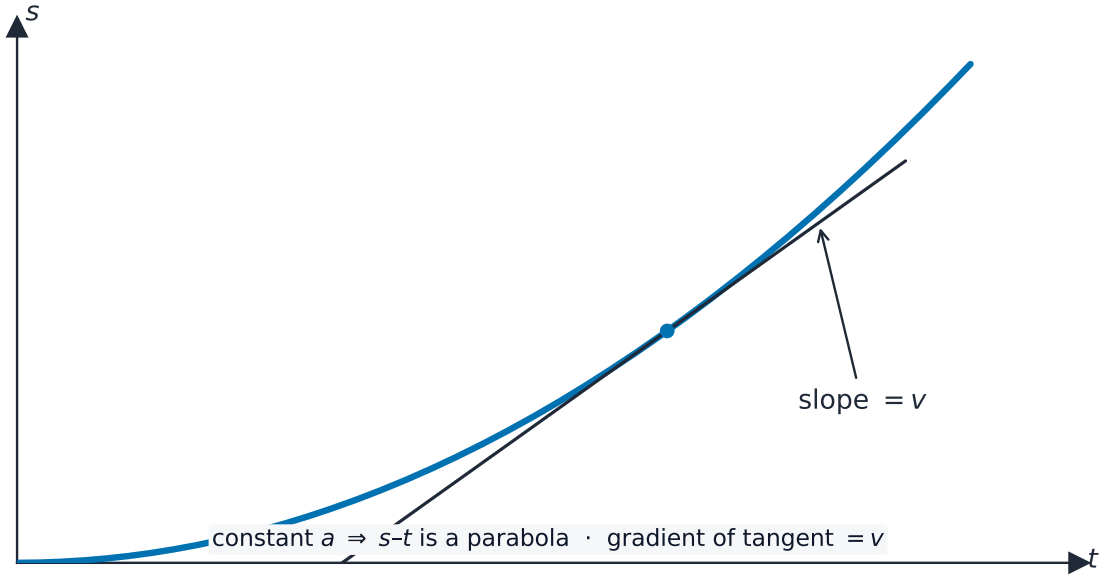
We know u, v and s and want a, so use $v^2 = u^2 + 2as$:

$$20^2 = 8^2 + 2a(56) \quad \Rightarrow \quad 336 = 112\,a \quad \Rightarrow \quad a = 3.0 \text{ m s}^{-2}.$$

Where the SUVAT equations come from

You should be able to get these from the definitions of velocity and acceleration:

- $v = u + at$ comes from $a = (v - u)/t$, the gradient of the line.
- $s = \frac{1}{2}(u + v)t$ is the area under the line — a **trapezium** 梯形 with parallel sides u and v and width t.
- $s = ut + \frac{1}{2}at^2$ comes from putting $v = u + at$ into the area.
- $v^2 = u^2 + 2as$ comes from removing t from the first two.



Displacement–time graph for uniform acceleration — the slope at any point equals the instantaneous velocity

If a question asks "which equation can be found using only the gradient of a velocity–time graph?", the answer is $v = u + at$ (the gradient is the acceleration).

Choosing a positive direction

Pick a positive direction at the start and keep it. Anything pointing the other way gets a minus sign. For a ball thrown straight up, if "up" is positive: u is positive, $a = -g$ (**gravity** 重力 pulls down), and at the highest point the displacement is positive but the velocity is zero.

Free fall under gravity

When **air resistance** 空气阻力 can be ignored, an object in **free fall** 自由落体 has a constant acceleration $g \approx 9.81$ m s^{-2} downwards. This is the same for every mass.

For a ball dropped from rest and falling a distance h:

$$h = \tfrac{1}{2}gt^2, \qquad v = gt, \qquad v^2 = 2gh.$$

For a ball thrown straight up with speed u:

- greatest height: put $v = 0$ in $v^2 = u^2 - 2gh$, giving $h = u^2/(2g)$.
- time to reach the top: put $v = 0$ in $v = u - gt$, giving $t = u/g$.
- total time to fall back to the start height: $2u/g$ (the motion is **symmetric** 对称).

Worked example. A ball is thrown straight up at 20 m s^{-1}. Find the greatest height it reaches (take $g = 9.81$ m s^{-2}).

At the highest point $v = 0$, so from $h = u^2/(2g)$:

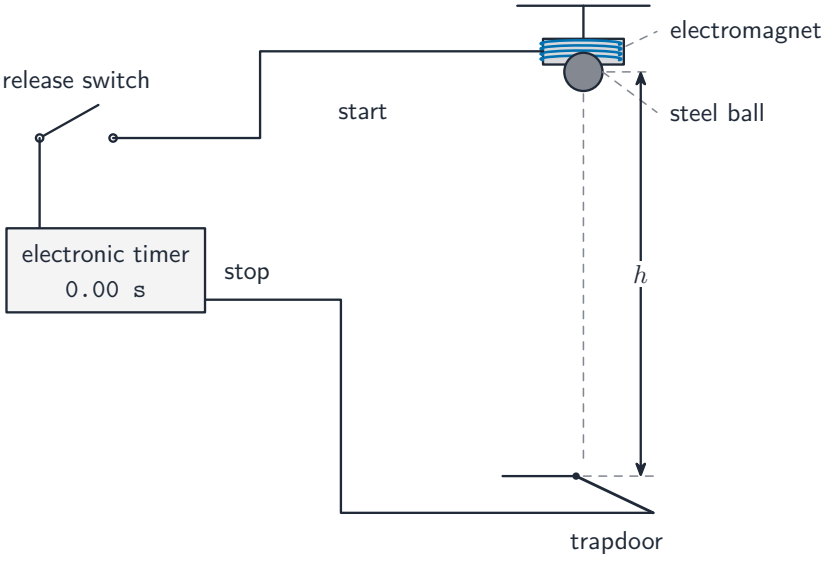
$$h = \frac{20^2}{2 \times 9.81} = \frac{400}{19.62} \approx 20.4 \text{ m}.$$

Experiment to find g

A common method: drop an object from rest, then measure the distance h it falls and the time t it takes. Then

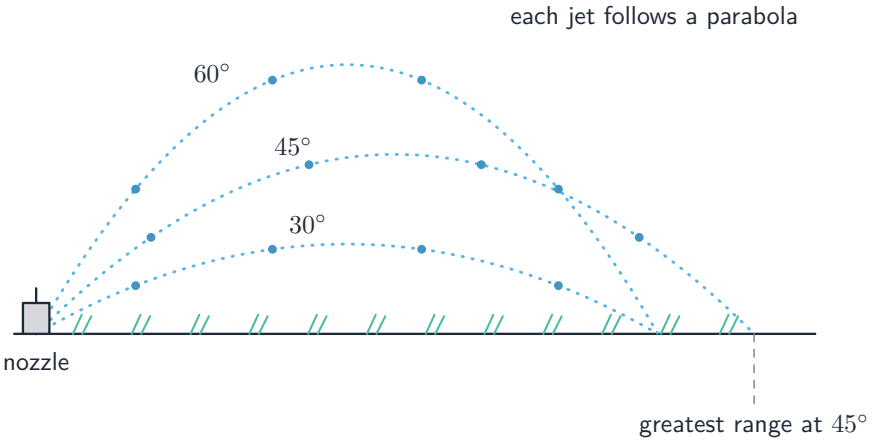
$$g = \frac{2h}{t^2}.$$

Repeat for several heights and plot h against t^2. The gradient of the best straight line is $g/2$, so g is twice the gradient. Repeating reduces **random error** 随机误差. An electronic timer — using **light gates** 光电门, or a switch the ball hits — removes **reaction-time** 反应时间 error.



Experimental set-up for measuring the acceleration due to free fall

Motion in two directions



Water jets from a sprinkler trace parabola paths — a real example of projectile motion

When an object moves at constant velocity in one direction (say **horizontal** 水平) and speeds up in a direction at right angles to it (say **vertical** 竖直, under gravity), the two motions do not affect each other. Treat each direction on its own, with its own SUVAT equation.

Horizontal throw

An object thrown horizontally with speed u_{H} from height h, with air resistance ignored:

- **horizontal**: constant velocity u_{H}. After time t, the horizontal distance is $x = u_{\mathrm{H}}t$.
- **vertical**: starts from rest and speeds up downwards at g. After time t, it has fallen $y = \frac{1}{2}gt^2$ and has vertical velocity $v_{\mathrm{V}} = gt$.

The time to reach the ground depends **only** on the height h, not on u_{H}. Solve $h = \frac{1}{2}gt^2$ for t; then the horizontal **range** 射程 is $u_{\mathrm{H}}t$.

Worked example. A ball is thrown horizontally at 15 m s^{-1} from the top of a cliff 20 m high. Find the time it takes to land and how far from the base it lands (take $g = 9.81$ m s^{-2}).

Vertical motion gives the time: from $h = \frac{1}{2}gt^2$,

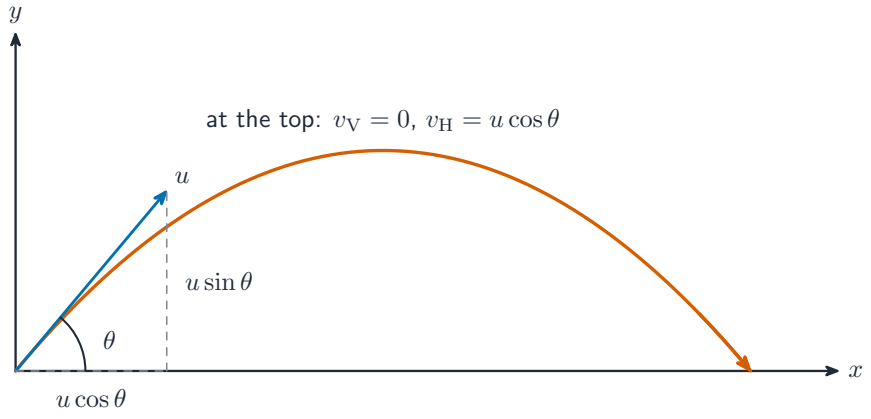
$$t = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2 \times 20}{9.81}} \approx 2.0 \text{ s}.$$

Horizontal motion then gives the range: $x = u_{\mathrm{H}}t = 15 \times 2.0 \approx 30$ m.

The horizontal-velocity graph is a flat line at u_{H}. The vertical-velocity graph is a straight line from the origin with gradient g.

Projectile at an angle

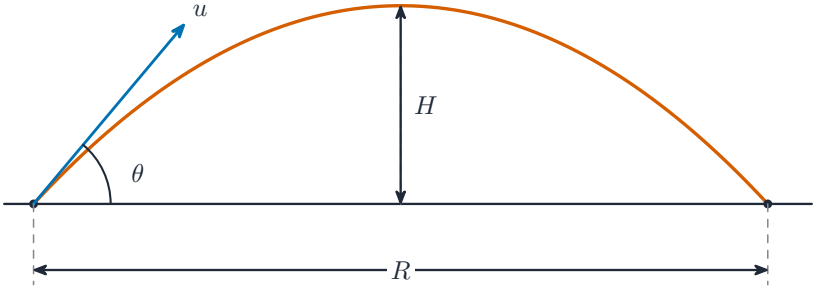
A **projectile** 抛体 thrown at speed u at angle θ above the horizontal:



Projectile launched at angle θ — horizontal and vertical motions are independent

- horizontal **component** 分量 of the starting velocity: $u_{\mathrm{H}} = u\cos\theta$ (stays constant during the flight).
- vertical component of the starting velocity: $u_{\mathrm{V}} = u\sin\theta$ (gets smaller, becomes zero at the top, then grows downwards).

At the highest point, $v_{\mathrm{V}} = 0$, but v_{H} is still $u\cos\theta$. The time to the top is $t_{\mathrm{up}} = u\sin\theta/g$; the total flight time (back to the start height) is $2t_{\mathrm{up}}$.



Range R of a projectile launched from and landing on level ground

Bouncing ball

When a ball bounces, its velocity–time graph is a set of straight sloping lines (constant g) with a sudden jump at each bounce (the velocity flips direction, and gets smaller if some **energy** 能量 is lost). Add up the times and the distances across the bounces.

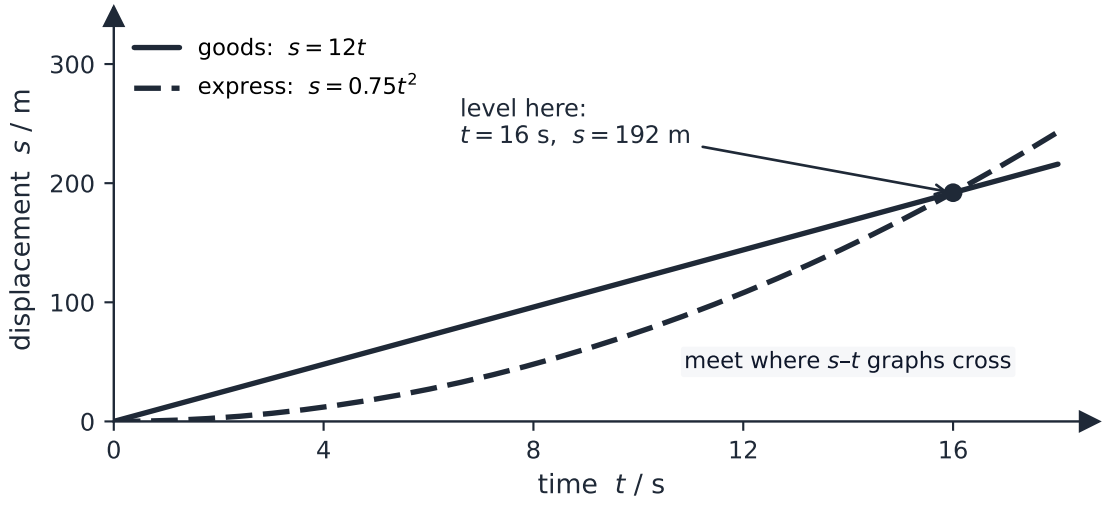
Two objects meeting

When two objects move along the same line in different ways, write a displacement equation for each. Use the same start time and the same positive direction. Then set the two displacements equal (or set their difference to a given gap).

For a goods train at constant velocity u_{G} and an express train starting from rest with acceleration a, both passing the same point at $t = 0$:

$$s_{\mathrm{G}} = u_{\mathrm{G}} t, \qquad s_{\mathrm{E}} = \tfrac{1}{2} a t^2.$$

They are level again when $s_{\mathrm{G}} = s_{\mathrm{E}}$, giving $t = 2u_{\mathrm{G}}/a$.



Meet where the displacement–time graphs cross (same s at same t)

Tips for solving problems

1. **Draw a diagram** and mark the positive direction.
2. **List the SUVAT symbols** with their known and unknown values, including signs.

3. **Choose the SUVAT equation** with exactly the four symbols you have, plus the one you want.
4. For projectile motion, **split into horizontal and vertical** SUVAT problems, linked only by the time t.
5. Always **check the units** of your answer, and that its size is sensible.

Exam tips

- Use the **SUVAT** equations only for **constant acceleration**; list s, u, v, a, t and pick the equation missing your unknown.
- Choose one direction as **positive** and keep signs consistent (usually $g = -9.81$ m s^{-2}).
- On a **velocity-time graph**, gradient = acceleration and area = displacement.
- For projectiles, treat **horizontal (constant velocity) and vertical ($a = g$) motion separately**, linked by the same time.

Dynamics

A-Level Physics

Mass, momentum and force

Mass

Mass 质量 tells you how hard it is to change an object's motion. The larger the mass, the larger the **force** 力 needed to give it a certain **acceleration** 加速度. Mass is measured in kilograms (kg) and is a **scalar** 标量.

Momentum

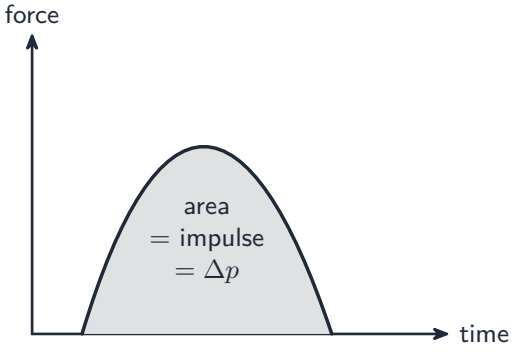
Linear momentum 动量 is the product of mass and velocity:

$$p = mv.$$

Momentum is a **vector** 矢量 — it points the same way as the **velocity** 速度. Its unit is kg m s^{-1}, which is the same as N s.

Force as the rate of change of momentum

Newton's second law, in its general form: the **resultant force** 合力 on an object equals the rate of change of its momentum.



Impulse is the area under a force-time graph, equal to the change in momentum

$$F = \frac{\Delta p}{\Delta t}.$$

When the mass is constant this becomes $F = ma$, because $\Delta p = m\,\Delta v$ and $\Delta v/\Delta t = a$. Cambridge questions often want you to use $F = \Delta p/\Delta t$ directly for a **collision** 碰撞 or an **impulse** 冲量: the average force equals the change in momentum divided by the contact time.

A ball hits a wall with momentum p_1 and bounces back with momentum p_2 in the opposite direction. The change in momentum is $\Delta p = p_2 - p_1$ (give each direction the correct sign). The average force is $\Delta p/\Delta t$, where Δt is the contact time.

Worked example. A 0.20 kg ball hits a wall at 8.0 m s^{-1} and bounces straight back at 6.0 m s^{-1}. The contact lasts 0.050 s. Find the average force on the ball.

Take the rebound direction as positive, so $u = -8.0$ m s^{-1} and $v = +6.0$ m s^{-1}:

$$\Delta p = m(v - u) = 0.20 \times \big(6.0 - (-8.0)\big) = 2.8 \text{ kg m s}^{-1},$$

$$F = \frac{\Delta p}{\Delta t} = \frac{2.8}{0.050} = 56 \text{ N}.$$

When you know the momentum but not the speed, the change in **kinetic energy** 动能 is

$$\Delta E_\text{k} = \frac{p_2^2 - p_1^2}{2m}.$$

This comes from $E_\text{k} = \tfrac{1}{2}mv^2 = p^2/(2m)$.

Newton's three laws of motion



A rocket pushes gas down; by Newton's third law the gas pushes the rocket up.

First law

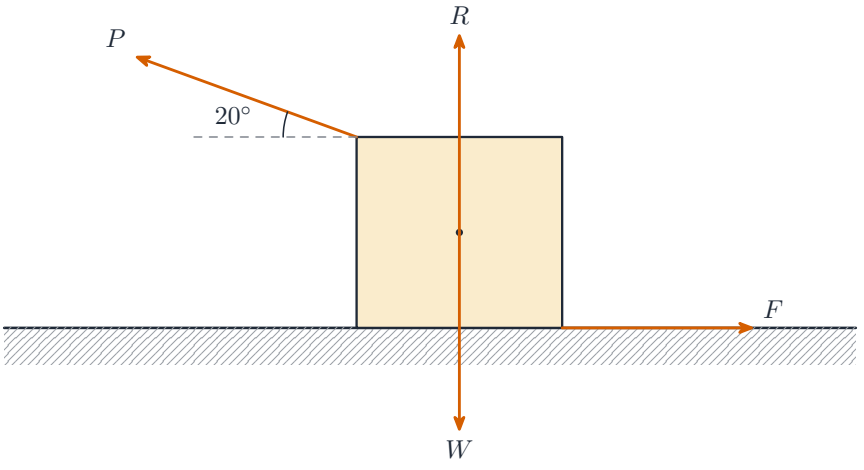
An object stays at rest, or keeps moving at constant velocity in a straight line, unless a resultant **external force** 外力 acts on it. In short: zero resultant force means zero acceleration.

Second law

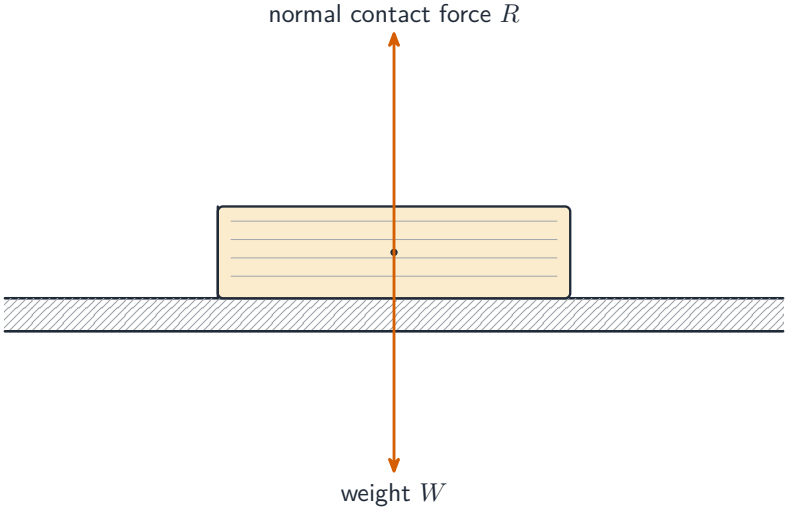
The resultant force on an object equals its rate of change of momentum, and acts in the same direction as that change. In SI units,

$$F = \frac{\Delta p}{\Delta t} = ma \quad \text{(for constant mass)}.$$

Acceleration and resultant force always point the **same way**.



Free-body diagram showing all forces on a block being pulled at an angle



Weight and normal contact force on a book resting on a table

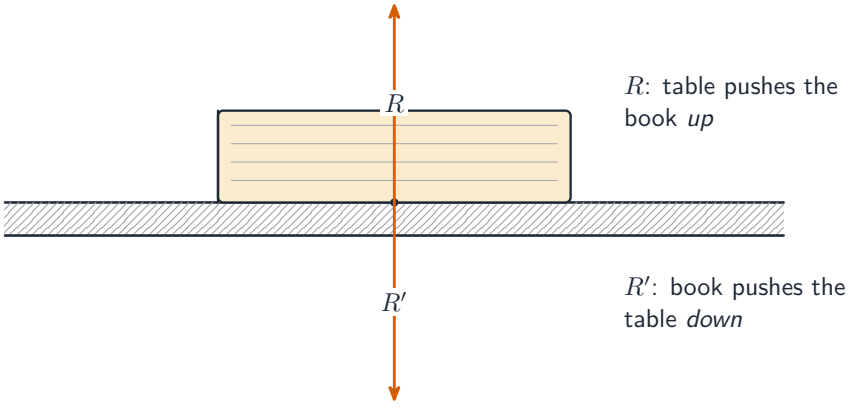
Third law

When body A pushes on body B, body B pushes back on body A with an **equal and opposite** force. The two forces:

- act on **different objects**,
- are of the **same type** (both gravitational, both contact, both **electrostatic** 静电, and so on),
- have the same size and opposite direction.

A common trap: the **weight** 重力 of a block on a table and the **normal contact force** 支持力 from the table are **not** a third-law pair (they act on the same object and are different types). The third-law partner of the block's weight is the pull the block makes on the Earth. The third-law partner of the table's contact force is the push the block makes on the table.

For a rocket: the **thrust** 推力 on the rocket and the force on the gases are a third-law pair (the engine pushes the gas down, the gas pushes the engine up). Weight and air resistance are not part of this pair.



Newton's third-law pair: R on the book (up) and R' on the table (down)

Weight

Weight is the force on an object from a **gravitational field** 重力场. Near the Earth's surface,

$$W = mg,$$

where $g \approx 9.81\ \text{m s}^{-2}$ is the acceleration of free fall. Weight is a vector that points towards the centre of the Earth. Do not mix it up with mass: mass is the same

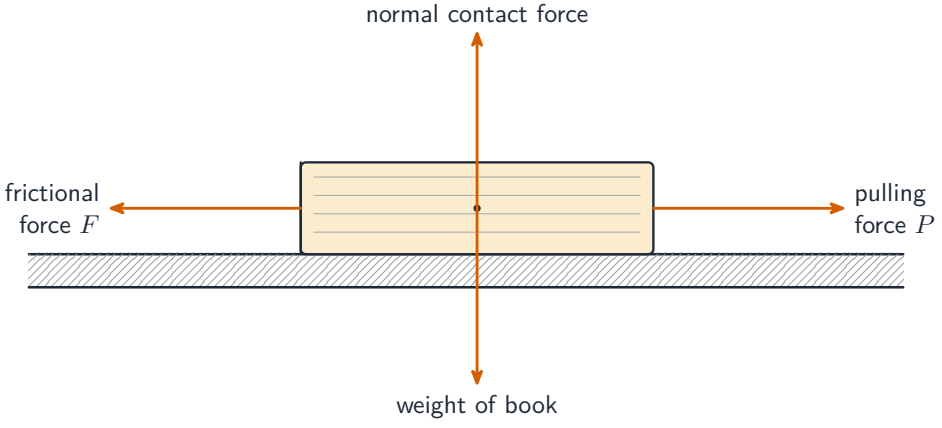
everywhere, but weight changes with place.

Non-uniform motion: friction, drag and terminal velocity

Friction and drag forces

A **friction** 摩擦力 force between two solid surfaces acts along the surface and opposes the sliding. A **viscous** 黏性 or **drag** 阻力 force is the resistive force from a **fluid** 流体 (a liquid or gas) on an object moving through it; air resistance is the case for air. You do not need to use any **coefficient** 系数 of friction or viscosity.

A simple model: the drag gets bigger as the speed gets bigger. At zero speed, the drag is zero.



Free-body diagram of a book being pulled on a table

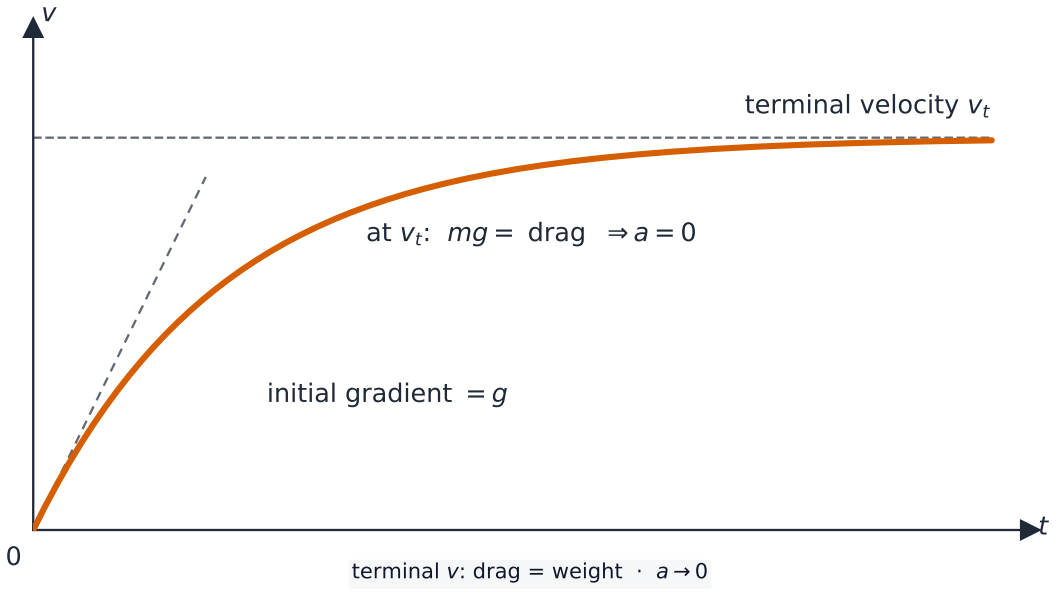
An object falling through air

For an object dropped from rest and falling through air:

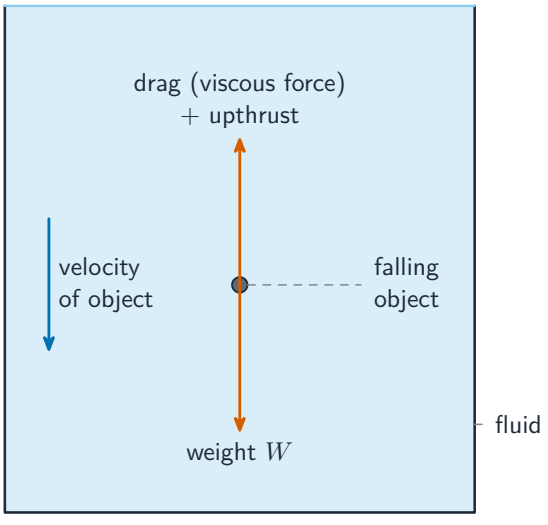
1. At first, only weight acts, so the object speeds up downwards at g.
2. As the speed grows, the upward drag grows. The resultant force gets smaller, so the acceleration gets smaller.
3. In the end, the drag equals the weight. The resultant force is zero, the acceleration is zero, and the speed stays constant — the **terminal velocity** 收尾速度.

On a velocity–time graph, the line starts straight with gradient g, then bends and flattens at the terminal velocity. This shape (fast start, then slowing acceleration,

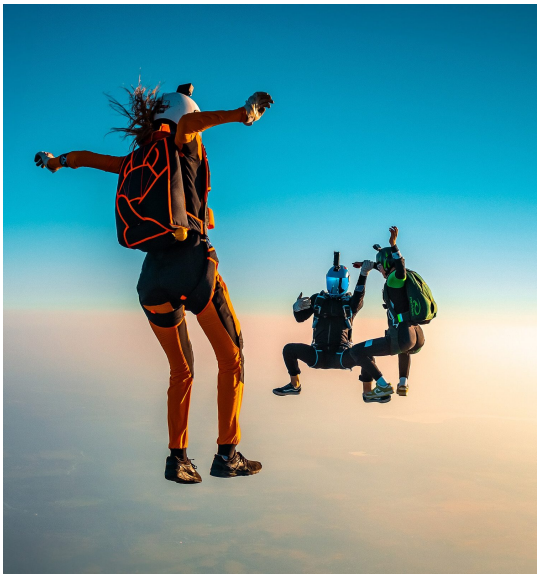
then constant speed) is how ”falling with air resistance” differs from ”free fall in a vacuum”.



Velocity–time graph for an object falling through air



Forces on a falling object in a fluid



Why the spread-eagle pose? Spreading out gives the largest area, so the most drag. The bigger the drag, the sooner drag balances weight — and the lower the steady terminal velocity they fall at

Energy during a terminal-velocity fall

At terminal velocity, a **parachutist** 跳伞者 has constant kinetic energy. But the **gravitational potential energy** 重力势能 keeps falling as they go down. Where does it go? Almost all of it turns into **thermal energy** 热能 of the air around them. It does **not** become kinetic energy of the parachutist — that stays constant.

Cyclist or car at constant speed

A vehicle at constant speed on a flat road has zero resultant force. The forward driving force is equal and opposite to the total resistive force (friction, air resistance, and rolling resistance). At higher speed the drag is larger, so the driving force must be larger too — and so the **power** 功率 must be larger.

Conservation of linear momentum



In a crash, a large force acts over a very short time to change momentum.

The principle

For a system with no resultant external force, the total momentum stays constant. This is **conservation of momentum** 动量守恒.

It always holds when there is no outside resultant force — in collisions, **explosions** 爆炸, and **recoil** 反冲. In two dimensions, momentum is conserved along each direction on its own.

Worked example. A 2.0 kg trolley and a 3.0 kg trolley are held together against a compressed spring, then released from rest. The 2.0 kg trolley flies off at 6.0 m s^{-1}. Find the speed of the other trolley.

The total momentum stays zero (it started at rest), so

$$0 = (2.0)(6.0) + (3.0)(-v) \quad \Rightarrow \quad v = \frac{12}{3.0} = 4.0 \text{ m s}^{-1}$$

in the opposite direction.



equal size, opposite direction
(forces act on different objects)

Newton's third law in an isolated two-particle system: equal and opposite forces

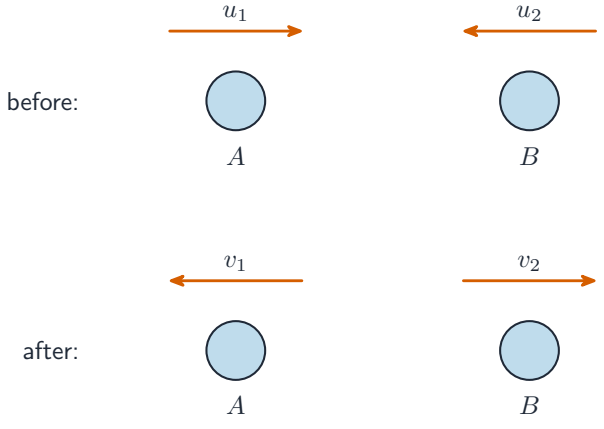
Elastic and inelastic collisions

In **every** collision, momentum is conserved (if there is no outside resultant force).

An **elastic collision** 弹性碰撞 is one where the total kinetic energy is **also** conserved. A quick test: in an elastic collision, the **relative speed** 相对速率 of approach equals the relative speed of separation.

In an **inelastic collision** 非弹性碰撞, momentum is conserved but kinetic energy goes down — some becomes thermal, sound, or **deformation** 形变 energy. If the two objects stick together, the collision is perfectly inelastic.

Solving collision problems (one dimension)



Head-on collision: velocities before and after

For two objects with masses m_1, m_2 and starting velocities u_1, u_2 that hit **head-on** 正面, write

$$m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2.$$

Use **signed** velocities (positive in one chosen direction). If the collision is elastic, add the relative-speed equation

$$u_1 - u_2 = -(v_1 - v_2),$$

or, the same thing, $\tfrac{1}{2}m_1u_1^2 + \tfrac{1}{2}m_2u_2^2 = \tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2$. That gives two equations for two unknowns.

Worked example. A 1500 kg car moving at 12 m s^{-1} runs into a stationary 1000 kg car and they lock together. Find their common velocity just after the collision.

Momentum is conserved (the cars stick, so $v_1 = v_2 = v$):

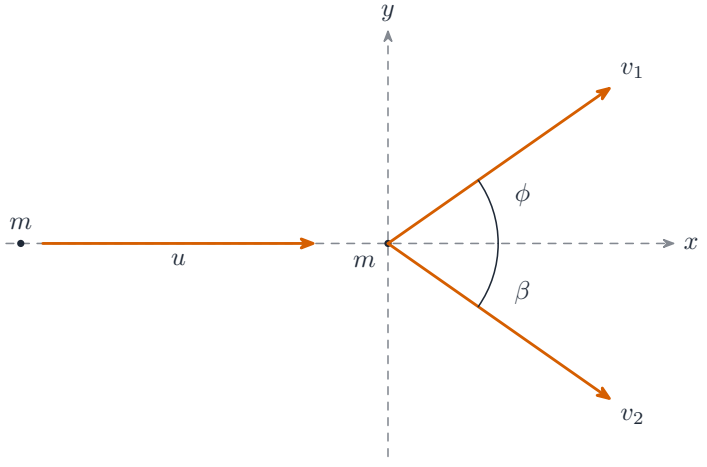
$$1500 \times 12 + 1000 \times 0 = (1500 + 1000)\,v \quad \Rightarrow \quad v = \frac{18\,000}{2500} = 7.2 \text{ m s}^{-1}.$$

A useful result for a head-on elastic collision of mass m with a **stationary** 静止 mass M:

$$v_m = \frac{m - M}{m + M}u, \qquad v_M = \frac{2m}{m + M}u.$$

Collisions in two dimensions

If the objects move in two dimensions, split the velocities into **perpendicular** 垂直 **components** 分量 and apply conservation of momentum along each direction on its own. For a collision where the objects hit at an angle, choose one axis along the first object's motion and one across it. The total momentum is conserved along each axis.



A glancing collision resolved along two perpendicular axes

Rocket / pushing out mass

A rocket pushes out gas at velocity u (relative to itself) at a **mass-flow rate** 质量流率 $\dot{m}$ (kg per second). It feels a thrust

$$F = \dot{m} \cdot u,$$

which comes from $F = \Delta p/\Delta t$. The momentum given to the gas each second equals the thrust on the rocket (Newton's third law: the rocket pushes the gas one way, the gas pushes the rocket the other way).

Exam tips

- **Newton's second law is** $F = \Delta p/\Delta t$ (rate of change of momentum); $F = ma$ is the special case for constant mass.
- Identify **third-law pairs** correctly: the same type of force, acting on **two different bodies** — not the balanced forces on one body.
- **Momentum is conserved** in every collision; kinetic energy is conserved only in an **elastic** collision.
- For **terminal velocity**, explain that drag rises with speed until drag = weight, so the acceleration becomes zero.

Forces, density and pressure

A-Level Physics

Turning effects of forces

Centre of gravity

The **weight** 重力 of a large object can be treated as acting at one single point, called the **centre of gravity** 重心 (the same as the **centre of mass** 质心 in a uniform gravitational field). For a uniform, regular shape — a rectangle, a sphere, a uniform rod — the centre of gravity is at the middle.

When you draw a **free-body diagram** 受力图, always put the weight arrow at the centre of gravity.

Moment of a force

The **moment** 力矩 of a force about a point is

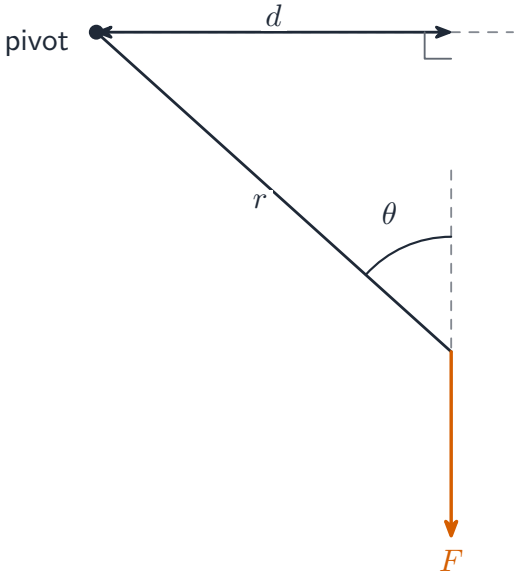
$$M = F \cdot d,$$

where F is the size of the **force** 力 and d is the **perpendicular distance** 垂直距离 from the point to the **line of action** 作用线 of the force. Unit: N m.

If the force acts at angle θ to a **lever arm** 力臂 of length r from the **pivot** 支点, then $d = r\sin\theta$, so $M = Fr\sin\theta$. Only the part of the force **perpendicular** 垂直 to the lever arm makes it turn.

Worked example. A force of 50 N is applied to the end of a spanner 0.40 m long, at 30° to the spanner. Find the moment of the force about the nut.

$$M = Fr\sin\theta = 50 \times 0.40 \times \sin 30° = 50 \times 0.40 \times 0.50 = 10 \text{ N m}.$$



The moment depends on the perpendicular distance d from the pivot to the line of action of the force

A moment is either **clockwise** 顺时针 or **anticlockwise** 逆时针 about the chosen point.

Couple and torque

A **couple** 力偶 is a pair of forces that are:

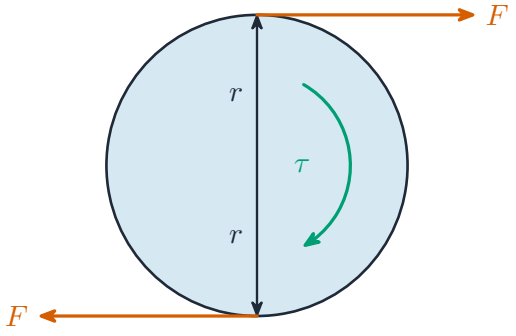
- **equal in size**,
- **opposite in direction**,
- with their lines of action a perpendicular distance apart.

A couple makes the body **turn only** — its resultant force is zero, so it gives no straight-line acceleration.

The **torque** 力偶矩 of a couple is the turning effect it makes:

$$\tau = F \cdot d,$$

where F is the size of one force and d is the perpendicular distance between the two lines of action. Unit: N m. The torque is the same about any point — a special property of couples.



A couple: two equal and opposite forces, a distance apart, producing a torque

A common multiple-choice trap: two equal forces in the **same** direction are not a couple (they have a resultant force and cause **translation** 平动). A couple needs equal size **and** opposite direction.

Equilibrium of forces

Conditions for equilibrium

A body is in **equilibrium** 平衡 when:

1. the **resultant force** 合力 is zero (no straight-line acceleration), AND

2. the resultant moment about any point is zero (no **angular acceleration** 角加速度).

Both must hold. A body with no resultant force can still be turning; a body with no resultant moment can still be moving in a straight line.

Principle of moments

For a body that is not turning, **the total clockwise moment about any point equals the total anticlockwise moment about the same point**. This is the **principle of moments** 力矩原理.

To solve a balance problem:

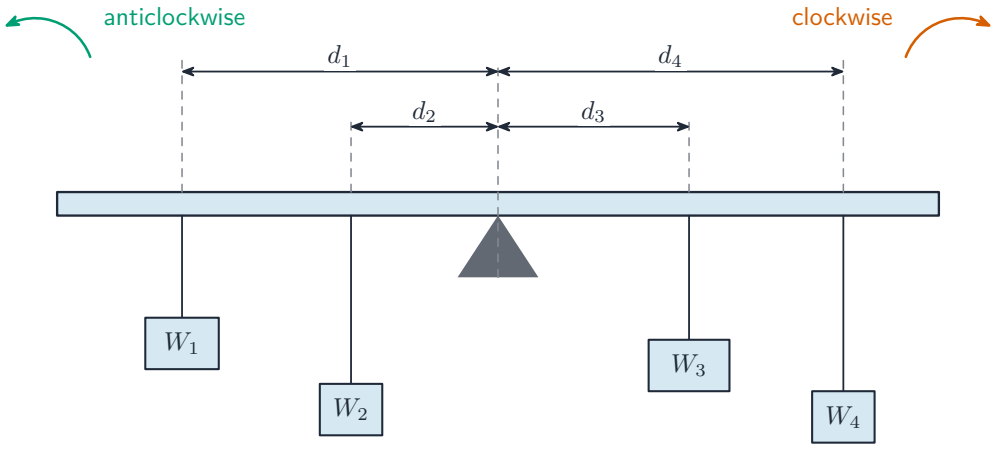
1. Choose a pivot — usually where an unknown force acts, so that force drops out (its distance is zero).
2. List every force and its perpendicular distance from the pivot.
3. Set $\sum M_{\text{clockwise}} = \sum M_{\text{anticlockwise}}$.
4. Use $\sum F = 0$ if you need a second equation.

A ruler balanced on a pivot with masses on each side is solved this way. For a heavy uniform rod, remember to include its weight acting at its centre of gravity.

Worked example. A uniform beam of length 4.0 m is pivoted at its centre. A 40 N weight hangs 1.5 m from the pivot on one side. How far from the pivot, on the other side, must a 30 N weight hang to balance the beam?

The beam's own weight acts at the centre (the pivot), so it has no moment. Setting clockwise = anticlockwise moments:

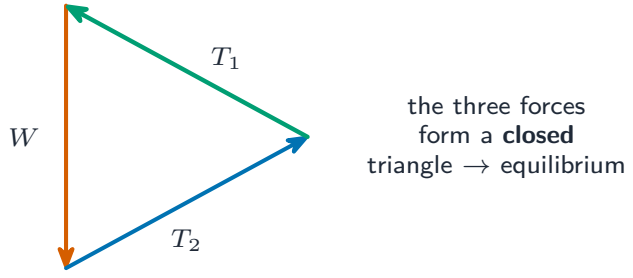
$$40 \times 1.5 = 30 \times d \quad \Rightarrow \quad d = \frac{60}{30} = 2.0 \text{ m}.$$



Weights on a rule balanced at a pivot — used to test the principle of moments

Vector triangle

Three forces in the same plane that are in equilibrium can be drawn as a **closed vector triangle** 矢量三角形 — drawn tip-to-tail, the three arrows come back to the start. This is a drawing method instead of splitting into **components** 分量.



Three forces in equilibrium form a closed vector triangle

Use the **sine rule** 正弦定理 or the **cosine rule** 余弦定理 on the triangle to find unknown sizes or directions, or draw the triangle to scale on graph paper.

You can also split each force into **horizontal** 水平 and **vertical** 竖直 components and set $\sum F_x = 0$ and $\sum F_y = 0$.

Density



An iceberg floats with most of its volume hidden: ice is slightly less dense than water.

Density 密度 is the mass per unit volume:

$$\rho = \frac{m}{V}.$$

Unit: kg m^{-3} (or g cm^{-3}; 1 g cm^{-3} = 1000 kg m^{-3}). Density is a **scalar** 标量.

Some useful densities to know:

- water: 1000 kg m^{-3}
- air at room conditions: $\sim$ 1.2 kg m^{-3}
- iron / steel: $\sim$ 7800 kg m^{-3}

Pressure

Pressure 压强 is the force per unit area, where the force acts at right angles to the surface:

$$p = \frac{F}{A}.$$

Unit: Pa = N m^{-2}. Pressure is a scalar.



A precision aneroid barometer measures atmospheric pressure

Hydrostatic pressure

Take a column of **fluid** 流体 with density ρ, **cross-sectional area** 横截面积 A and height Δh. Its weight is

$$W = mg = (\rho \cdot A \cdot \Delta h) \cdot g.$$

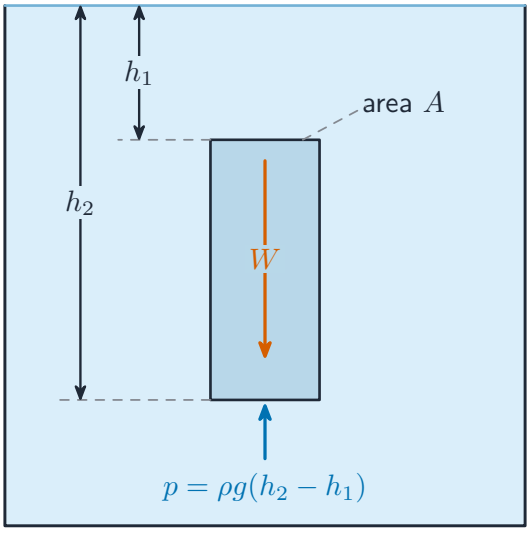
This weight presses down on the area A at the bottom, so the extra pressure at the bottom compared with the top is

$$\Delta p = \frac{W}{A} = \rho g \Delta h.$$

This is the **hydrostatic pressure** 流体静压强 equation. It depends only on the density of the fluid and the **depth** 深度 — the shape of the container does not matter.

Worked example. Find the extra pressure due to the water at the bottom of a swimming pool 2.5 m deep. (Water density 1000 kg m^{-3}, $g = 9.81$ m s^{-2}.)

$$\Delta p = \rho g \Delta h = 1000 \times 9.81 \times 2.5 \approx 2.5 \times 10^4 \text{ Pa}.$$

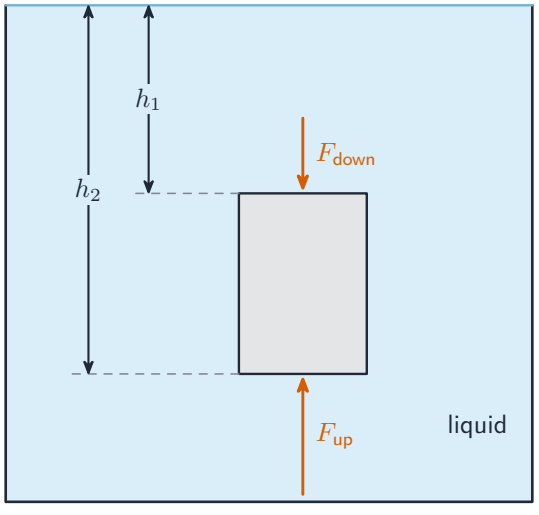


A column of liquid of area A: its weight sets the extra pressure at the depth below

For a submarine at depth h below the surface, the pressure from the water is $\rho_{\text{seawater}}\, g\, h$. For the total pressure, add the **atmospheric pressure** 大气压强 at the surface (about 1.0×10^5 Pa).

Upthrust and Archimedes' principle

When an object is **submerged** 浸没 in a fluid, the pressure at the bottom of the object is greater than the pressure at the top (by $\rho g \Delta h$, where Δh is the object's height). This difference gives a net upward force called the **upthrust** 浮力.



Upthrust arises because the pressure on the bottom of the object is greater than on the top

For an object of volume V (the volume of fluid **displaced** 排开), the upthrust is

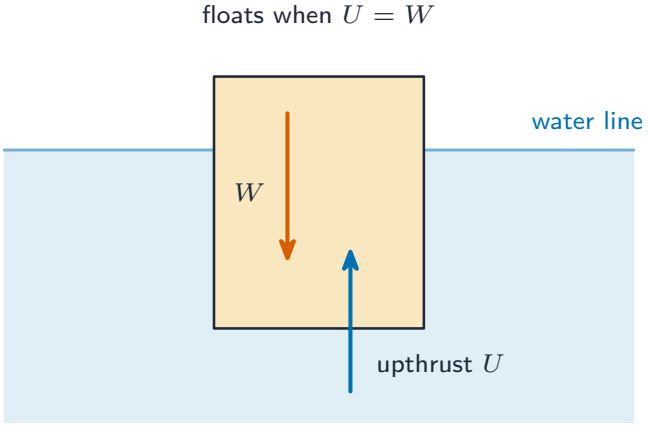
$$F_{\text{upthrust}} = \rho_{\text{fluid}}\, g\, V.$$

This is **Archimedes' principle** 阿基米德原理: the upthrust on a body in a fluid equals the weight of the fluid it pushes aside.

Worked example. A metal block of volume 2.0×10^{-3} m^3 is fully submerged in water. Find the upthrust on it. (Water density 1000 kg m^{-3}, $g = 9.81$ m s^{-2}.)

$$F_{\text{upthrust}} = \rho_{\text{fluid}}\, g\, V = 1000 \times 9.81 \times 2.0 \times 10^{-3} \approx 20 \text{ N}.$$

For a fully submerged object, V is its full volume. For a **floating** 漂浮 object, V is only the volume below the surface — the object floats when the upthrust on the part below the surface equals its weight.



A floating object sinks until the upthrust on the submerged part equals its weight

Force balance with upthrust

A block held under water by a string tied to the bottom of the container is in equilibrium under three vertical forces: weight (down), **tension** 张力 (down), upthrust (up). Set $F_{\text{upthrust}} = W + T$ to find the tension.

A submerged block hanging from a **newton meter** 弹簧测力计 reads less than its weight in air, because of the upthrust: reading $= W - F_{\text{upthrust}}$.

The upthrust depends on the fluid density and the displaced volume, not on the object's material or depth (for an **incompressible** 不可压缩 fluid). On a planet with smaller g, the upthrust is smaller in the same ratio as the weight, so a floating object still floats with the same fraction below the surface.

Exam tips

- Take **moments** about a chosen pivot; for equilibrium, total clockwise = total anticlockwise moments **and** the resultant force is zero.
- A body in equilibrium under three forces gives a **closed triangle** of forces.
- Fluid pressure = ρgh; **upthrust = weight of fluid displaced** (Archimedes).
- Distinguish mass, weight and density, and give the base unit each time.

Work, energy and power

A-Level Physics

Work, energy and power



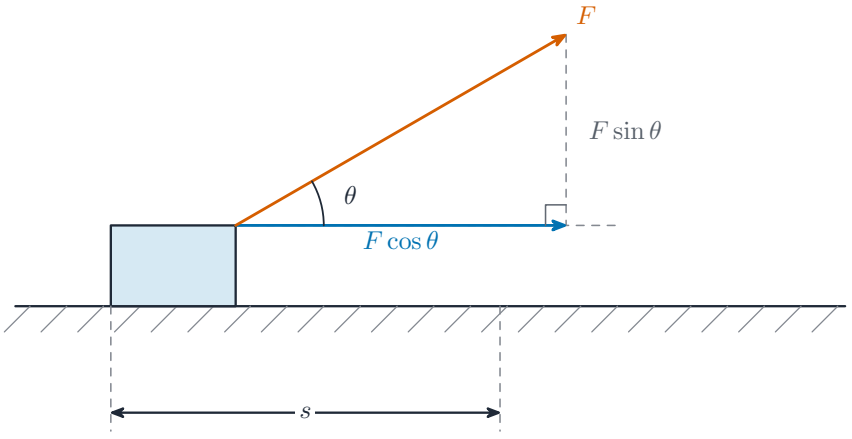
Wind turbines transfer the kinetic energy of the wind into electrical energy.

Work done by a force

Work 功 is done when a force moves its point of contact along the line of the force. The work done by a constant force F that causes a **displacement** 位移 s is

$$W = F \cdot s \cdot \cos\theta,$$

where θ is the angle between the force and the displacement. Only the **component** 分量 of the force along the displacement does work.



Only the component of the force along the displacement ($F\cos\theta$) does work

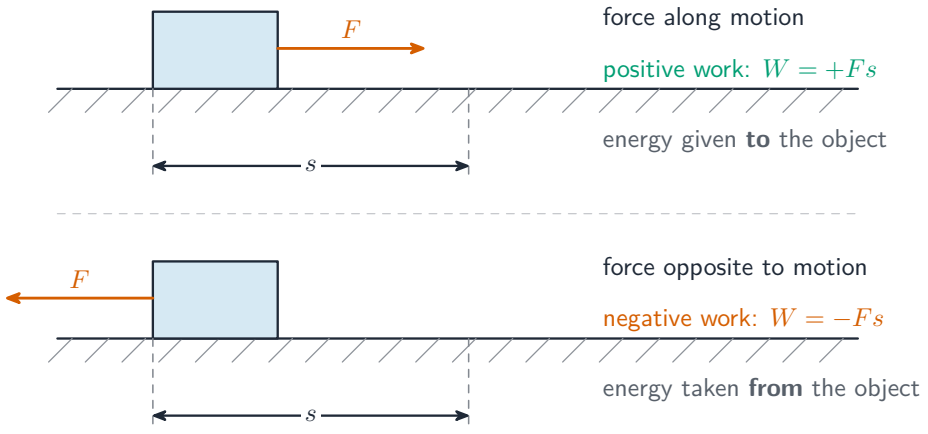
Worked example. A child pulls a sledge 5.0 m across the snow with a rope, using a force of 20 N at 60° to the ground. Find the work done by the rope.

$$W = Fs\cos\theta = 20 \times 5.0 \times \cos 60° = 20 \times 5.0 \times 0.50 = 50 \text{ J}.$$

Unit: J = N m. Work is a **scalar** 标量.

Special cases:

- force in the same direction as the motion ($\theta = 0$): $W = Fs$, positive work, **energy** 能量 given **to** the object.
- force at right angles to the motion ($\theta = 90°$): $W = 0$. The **normal contact force** 支持力 on a car on a flat road does no work.
- force opposite to the motion ($\theta = 180°$): $W = -Fs$, negative work, energy taken **from** the object (for example **friction** 摩擦力).



Positive work ($W = +Fs$): force along the motion (top). Negative work ($W = -Fs$): force opposite to the motion, e.g. friction (bottom)

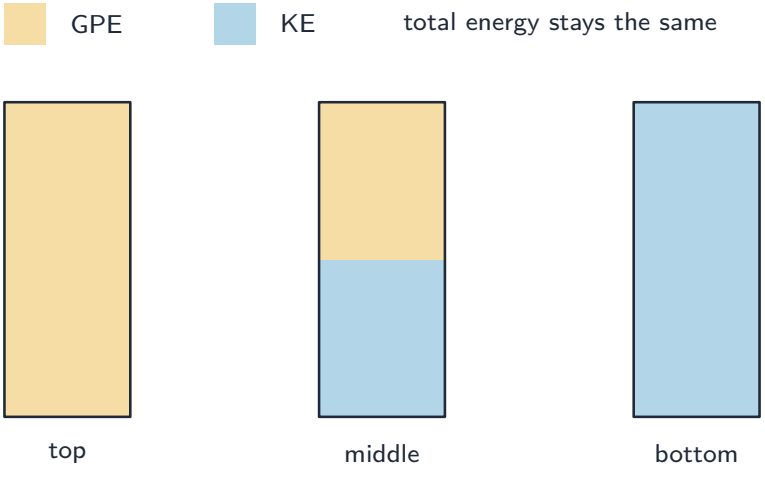
For an object moving up a slope at angle α to the **horizontal** 水平, the work done against gravity in rising a height h is mgh, while the work done by a horizontal push over the slope length L uses $\cos\alpha$.

Conservation of energy

Energy is **never made or destroyed** — it only changes from one form to another, or moves from one object to another. In a **closed system** 封闭系统, the total energy stays constant. This is **conservation of energy** 能量守恒.

When you write an energy equation, list every form the energy starts as and ends as. Common forms in this syllabus: kinetic, gravitational potential, **elastic potential energy** 弹性势能, **electrical** 电能, thermal, sound, **chemical energy** 化学能.

A ball rolling down a **frictionless** 无摩擦 **ramp** 斜坡 turns **gravitational potential energy** 重力势能 into **kinetic energy** 动能: $mgh = \frac{1}{2}mv^2$, so $v = \sqrt{2gh}$. With friction, some of this energy becomes **thermal energy** 热能 of the ramp and the air.



On a frictionless ramp, GPE turns into KE while the total energy stays constant

Worked example. A ball is released from rest at the top of a smooth ramp 1.2 m high. Find its speed at the bottom (take $g = 9.81$ m s^{-2}).

All the gravitational potential energy becomes kinetic energy, so $v = \sqrt{2gh}$ (the mass cancels):

$$v = \sqrt{2 \times 9.81 \times 1.2} \approx 4.9 \text{ m s}^{-1}.$$

Efficiency

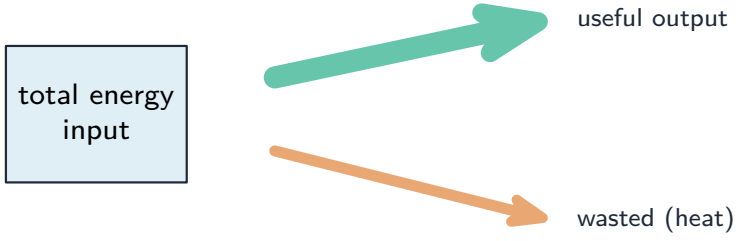
The **efficiency** 效率 of a system is

$$\text{efficiency} = \frac{\text{useful energy output}}{\text{total energy input}} \times 100\%.$$

The same idea with power:

$$\text{efficiency} = \frac{\text{useful power output}}{\text{total power input}} \times 100\%.$$

Efficiency is always less than 100% in a real system, because some input energy becomes "useless" forms — usually thermal energy.



$$\text{efficiency} = \frac{\text{useful output}}{\text{total input}} \times 100\%$$

Efficiency: only part of the energy input leaves as useful output; the rest is wasted, usually as heat

For an electric motor lifting a load with efficiency η at **voltage** 电压 V and **current** 电流 I, the useful output power is ηVI. From this you can find a force, a lifting speed, or a **tension** 张力.

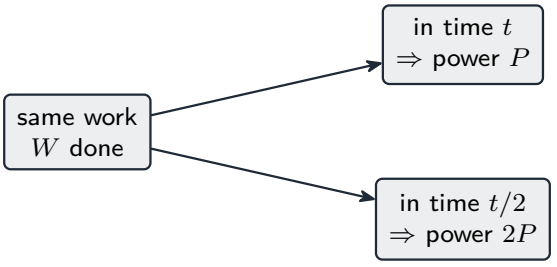
Worked example. An electric motor lifts a 50 kg load at a steady 0.40 m s^{-1} while drawing 250 W of electrical power. Find its efficiency (take $g = 9.81$ m s^{-2}).

The useful output power is $P = mgv = 50 \times 9.81 \times 0.40 = 196$ W, so

$$\text{efficiency} = \frac{196}{250} \times 100\% \approx 78\%.$$

Power

Power 功率 is the rate of doing work, or the rate of transferring energy:



$$P = \frac{W}{t} \text{ — less time, more power}$$

The same work done in less time means more power

$$P = \frac{W}{t} = \frac{\Delta E}{\Delta t}.$$

Unit: W = J s^{-1}. Power is a scalar.

Power, force and velocity

For an object moving at **velocity** 速度 v with a force F along the direction of motion, in a short time Δt the displacement is $v\,\Delta t$ and the work done is $Fv\,\Delta t$. Dividing by Δt:

$$P = Fv.$$

This is one of the most useful results in mechanics.

Worked example. A car travels at a steady 25 m s^{-1} against a total resistive force of 600 N. Find the output power of its engine.

At constant speed the driving force equals the resistive force, so

$$P = Fv = 600 \times 25 = 15\,000 \text{ W} = 15 \text{ kW}.$$

- For a car at constant velocity v on a flat road, the engine power must balance the total resistive force: $P = F_{\text{resist}} \cdot v$. If the **drag** 阻力 grows with v^2, doubling the speed roughly **quadruples** the power needed.

- For lifting a **weight** 重力 mg straight up at constant speed v, the useful output power is $P = mg \cdot v$.
- For an aircraft hovering at a fixed height, the lift force equals the weight, and a large power is needed because air must be pushed downwards all the time.

Gravitational potential energy

In a **uniform** gravitational field (close to a planet's surface), the change in gravitational potential energy of **mass** 质量 m rising or falling through a height Δh is

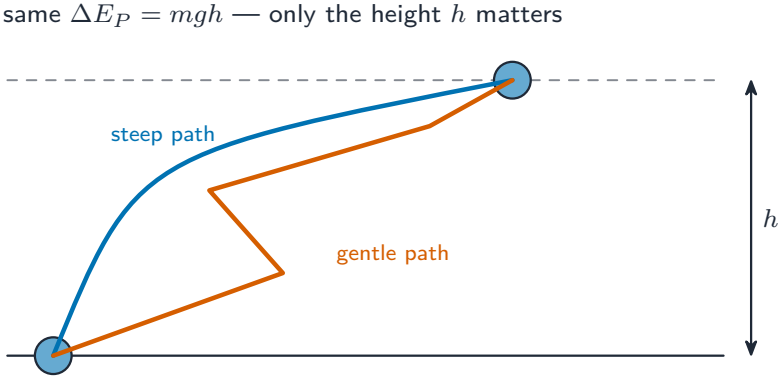
$$\Delta E_{\mathrm{P}} = mg\Delta h.$$

Where it comes from

The work done **against** gravity to raise a mass m slowly (no change in kinetic energy) through height Δh equals the gravitational potential energy gained:

- the gravitational force on the mass is mg downwards,
- the force needed to lift it slowly is mg upwards,
- the work done by this force is $W = F \cdot s = mg \cdot \Delta h$,
- this work becomes ΔE_{P}.

So $\Delta E_{\mathrm{P}} = mg\Delta h$. To use it you need m and Δh (and g). You do not need speed or time.



Gravitational PE depends only on the height risen, not the path taken

Kinetic energy

The kinetic energy of an object of mass m moving at speed v is

$$E_{\mathrm{K}} = \tfrac{1}{2}mv^2.$$

Where it comes from

Apply a resultant force F to a mass m that starts at rest. It speeds up evenly from 0 to v over a displacement s. From $v^2 = u^2 + 2as$ with $u = 0$,

$$s = \frac{v^2}{2a}.$$

The work done on the mass is

$$W = F \cdot s = ma \cdot \frac{v^2}{2a} = \tfrac{1}{2}mv^2.$$

All this work becomes kinetic energy, so $E_{\mathrm{K}} = \tfrac{1}{2}mv^2$.

Kinetic energy and momentum

Combining $p = mv$ and $E_{\mathrm{K}} = \tfrac{1}{2}mv^2$:

$$E_{\mathrm{K}} = \frac{p^2}{2m}.$$

This is handy when the **momentum** 动量 is given but not the velocity. For a momentum change from p_1 to p_2 at constant mass, the change in kinetic energy is $(p_2^2 - p_1^2)/(2m)$.

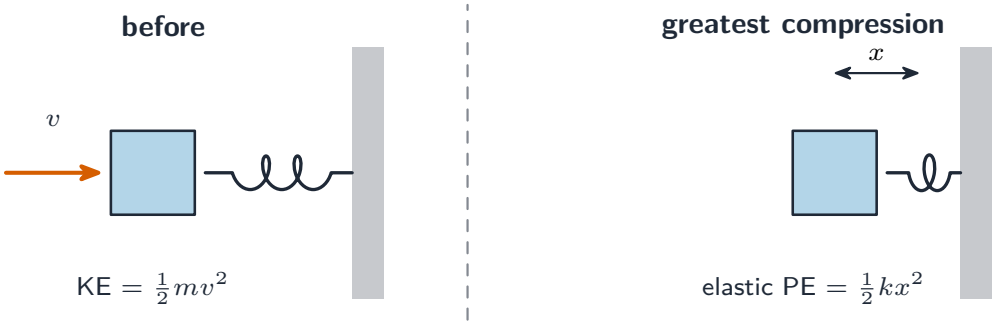
Using energy methods

A useful plan for problems that mix forces and energy:

1. **Find the start and end states.** Write the kinetic and potential energies in each.
2. **List any work done by outside forces** (friction, a push). Friction usually takes energy out; a push can add it.
3. **Conservation of energy**: $E_{\text{start}} + W_{\text{in}} = E_{\text{end}} + W_{\text{lost as heat etc.}}$.

Examples:

- A box pushed at **constant velocity** up a ramp of length L rising by h: E_K does not change, so the work done by the push goes into ΔE_P plus the work done against friction.
- A block sliding into a **spring** 弹簧 with kinetic energy E_K on a frictionless surface: at greatest **compression** 压缩 x, all the kinetic energy has become elastic potential energy $\frac{1}{2}kx^2$ (where k is the **spring constant** 劲度系数).
- A ball dropped from height h_1 that bounces to height h_2: the ratio h_2/h_1 is the fraction of mechanical energy kept, $h_2/h_1 = (v_{up}/v_{down})^2$.
- A **projectile** 抛体 thrown to the same height at different angles: the final speed is the same (only the height matters); use components to get its direction.



A block's kinetic energy becomes elastic potential energy as it compresses the spring

Exam tips

- **Work = force × distance moved in the direction of the force** — use $Fs\cos\theta$ when they are at an angle.
- Use **conservation of energy**: loss in GPE = gain in KE (+ work done against resistance).
- **Power = work / time** $= Fv$; efficiency = useful output ÷ total input.
- GPE uses the **vertical** height gained, not the distance along a slope.

Deformation of solids

A-Level Physics

Forces that cause deformation

When a force acts on a solid along its length, the object changes shape (**deformation** 形变). Two cases (treated as one-dimensional here):

- a **tensile** 拉伸 force stretches the object — it makes an **extension** 伸长量 x,
- a **compressive** force squeezes the object — it makes a **compression** 压缩, treated as a negative extension.

The applied force is the **load** 负载. The change from the natural length is the extension (or compression).

Hooke's law and the spring constant



A spring obeys Hooke's law: extension is proportional to the force applied.

For many materials at small extensions, **the extension is proportional to the load** — this is **Hooke's law** 胡克定律. The constant that links them is the **spring constant** 劲度系数 k:

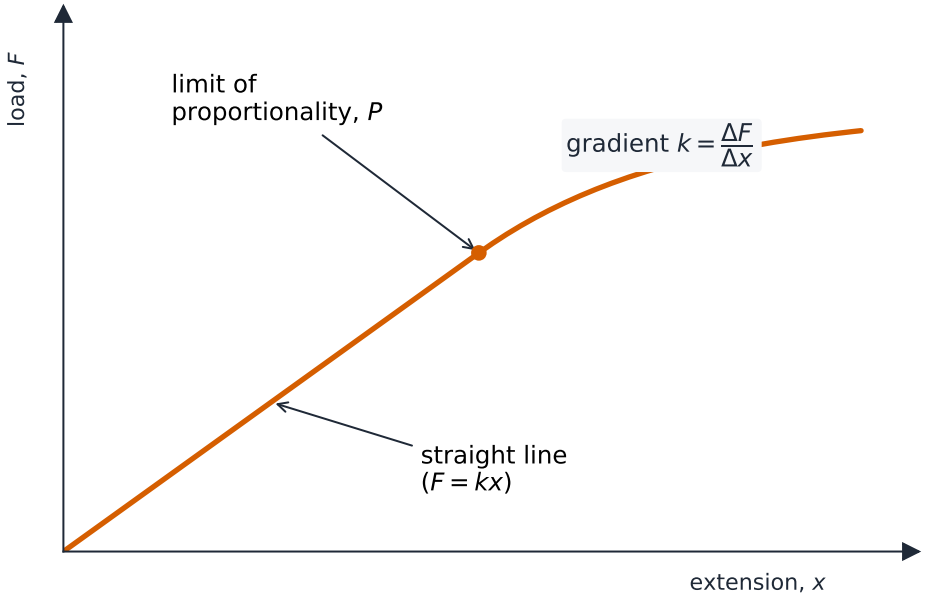
$$F = kx \qquad \Longleftrightarrow \qquad k = \frac{F}{x}.$$

Unit of k: N m^{-1}.

Worked example. A spring stretches by 4.0 cm when a 2.0 N load is hung from it. Find its spring constant.

Converting the extension to metres (4.0 cm = 0.040 m):

$$k = \frac{F}{x} = \frac{2.0}{0.040} = 50 \text{ N m}^{-1}.$$



A load–extension graph: straight up to the limit of proportionality P, then it curves

Reading a graph:

- A force–extension (F against x) graph has gradient k in the Hooke's-law region.
- An extension–force (x against F) graph has gradient $1/k$ in the Hooke's-law region.

A common trap: if a graph plots **length** L against force, you can still find the spring constant from the gradient (since $L = L_0 + F/k$, the gradient is $1/k$ — read it off carefully).

Limit of proportionality

Hooke's law only holds up to the **limit of proportionality** 比例极限. Past this point the F against x line curves and is no longer straight. The material may still be **elastic** 弹性 (it returns to its first length when you remove the load) a little further, then it becomes **plastic** 塑性.

Springs in series and parallel

You may need to combine spring constants:

- **Series** 串联 (one spring hangs from another): the same load passes through both, the total extension is the sum, so $\frac{1}{k_{\text{total}}} = \frac{1}{k_1} + \frac{1}{k_2}$.
- **Parallel** 并联 (two springs side by side holding the same load): each takes half the load (if they are identical), the extensions are equal, so $k_{\text{total}} = k_1 + k_2$.



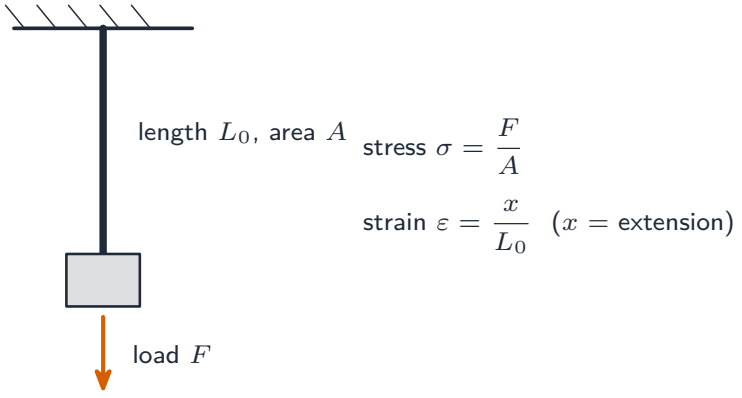
$\frac{1}{k} = \frac{1}{k_1} + \frac{1}{k_2}$ $\qquad$ $k = k_1 + k_2$

Combining spring constants: series gives a softer spring, parallel a stiffer one

Stress, strain and the Young modulus

For a wire of uniform cross-section under a tensile load:

- **Stress** 应力 $\sigma = \dfrac{F}{A}$, where F is the load and A is the **cross-sectional area** 横截面积. Unit: Pa.
- **Strain** 应变 $\varepsilon = \dfrac{x}{L_0}$, where x is the extension and L_0 is the original length. Strain has no unit (it is a ratio of lengths).



Stress is the load per cross-sectional area; strain is the extension per original length

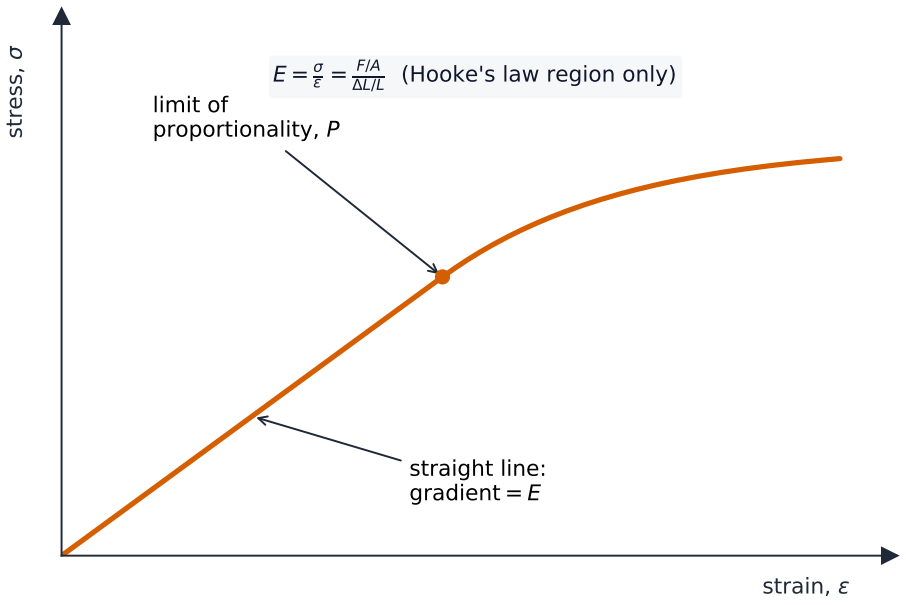
The **Young modulus** 杨氏模量 is the ratio of stress to strain in the Hooke's-law region:

$$E = \frac{\sigma}{\varepsilon} = \frac{F/A}{x/L_0} = \frac{FL_0}{Ax}.$$

Unit: Pa (about 10^{11} for metals; e.g. steel $\approx 2.0 \times 10^{11}$ Pa).

Worked example. A steel wire of length 2.0 m and cross-sectional area 1.5×10^{-7} m^2 stretches by 1.0 mm under a load of 15 N. Find the Young modulus.

$$E = \frac{FL_0}{Ax} = \frac{15 \times 2.0}{(1.5 \times 10^{-7})(1.0 \times 10^{-3})} = 2.0 \times 10^{11} \text{ Pa}.$$



A stress–strain graph, straight up to the limit of proportionality P

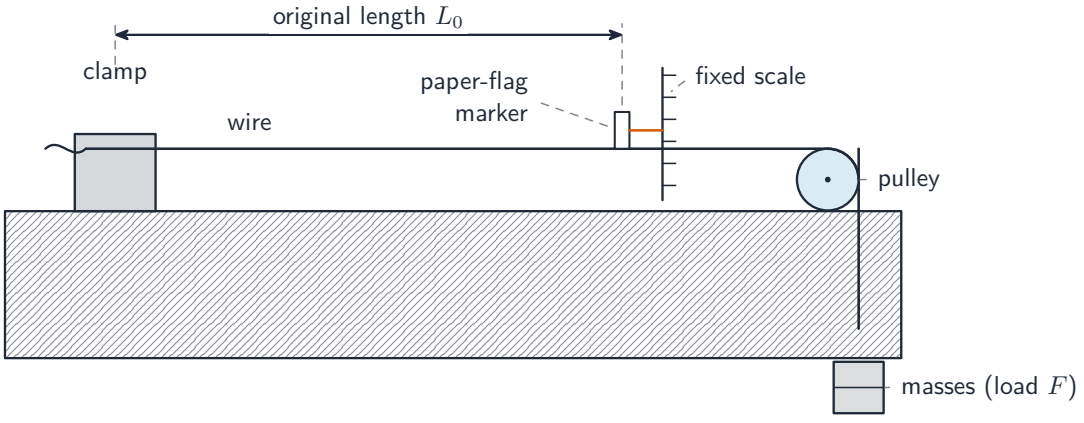
The Young modulus is a property of the **material** — it does not depend on the wire's shape or size. The spring constant k depends on both the material and the size: $k = EA/L_0$.

Experiment to find the Young modulus of a metal wire

A standard setup:

1. Clamp one end of a long, thin wire to a fixed support. Pass the wire over a **pulley** 滑轮 at the edge of the bench so it hangs straight down.
2. Measure the original length L_0 between the clamp and a marker near the pulley, using a metre rule.
3. Measure the **diameter** 直径 d of the wire at several places with a **micrometer** 螺旋测微器 and take the average. Work out $A = \pi d^2/4$.
4. Hang weights one at a time. Record the load F and the extension x (how far the marker moves against a fixed scale).
5. Plot F against x. In the straight region the gradient is EA/L_0, so $E = \text{gradient} \times L_0/A$.

Why a **long, thin** wire? To make the extension big enough to measure well. Why repeat readings and measure d at several places? To reduce **random error** 随机误差 and check the wire is uniform.



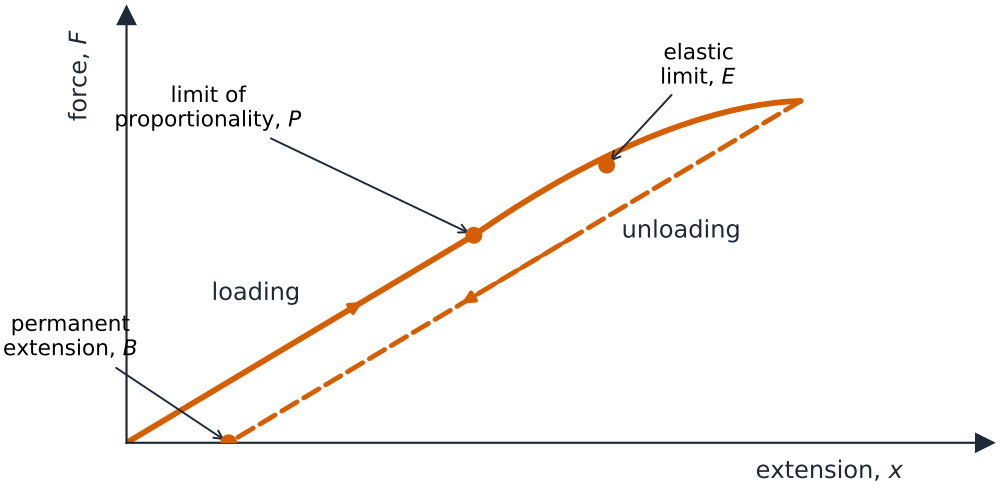
Apparatus for measuring the Young modulus of a wire

Elastic and plastic behaviour

As the load grows:

1. **Elastic and straight (Hooke obeyed)** — up to the limit of proportionality. Removing the load returns the object to its first length.
2. **Elastic but curved** — between the limit of proportionality and the **elastic limit** 弹性极限. The extension is no longer straight in the load, but on unloading the object still returns to its first length.
3. **Plastic** — past the elastic limit. On unloading, the object does **not** return to its first length; a permanent extension stays.

Hooke's law only holds in the straight, elastic region.



Hooke up to limit of proportionality · plastic beyond yield

Force–extension past the elastic limit: P and E marked, with a permanent extension B left after unloading

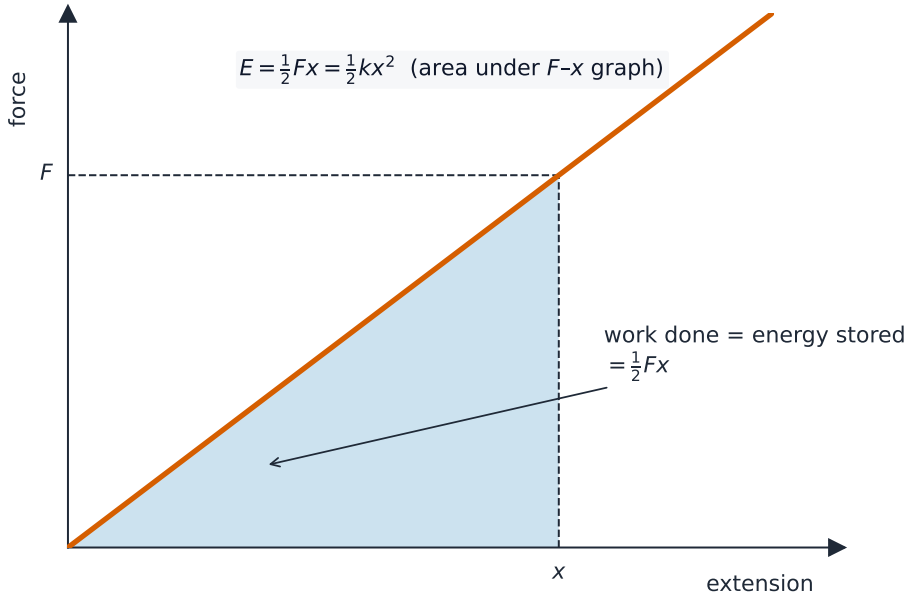


A modern universal (tensile) testing machine stretches a sample and records the force and extension

On a force–extension graph for a material taken into the plastic region and then unloaded, the loading line and the unloading line are different. The unloading line is parallel to the first Hooke line but shifted to the right (the permanent extension left when the load reaches zero). The area between the loading and unloading lines is the energy turned into **thermal energy** 热能 in the material.

Energy stored in a stretched material

The **work done** in stretching a material from 0 to extension x, as the load grows from 0 to F, is the **area under the force–extension graph**.



The work done stretching a material is the area under the force–extension graph

Hooke's-law material

When Hooke's law holds, the F against x graph is a straight line through the origin. The area under it from 0 to x is a triangle:

$$E_{\mathrm{P}} = \tfrac{1}{2}Fx = \tfrac{1}{2}kx^2.$$

This is the **elastic potential energy** 弹性势能 stored in a spring or wire stretched within its limit of proportionality. An equal form:

$$E_{\mathrm{P}} = \frac{F^2}{2k}.$$

Worked example. A spring of spring constant 50 N m^{-1} is stretched by 0.20 m, within its limit of proportionality. Find the elastic potential energy stored.

$$E_{\mathrm{P}} = \tfrac{1}{2}kx^2 = \tfrac{1}{2} \times 50 \times 0.20^2 = 1.0 \text{ J}.$$

Non-Hooke material

For a graph that is not a straight line (a stretched rubber band, or a spring past its limit of proportionality), find the area by counting grid squares or by using **trapezia** 梯形. The same idea holds: **the area under the force–extension graph is the work done on the material**.

Comparing stored energy

A common multiple-choice case: two materials are stretched by the same force, or by the same extension. Using $E_{\mathrm{P}} = \frac{1}{2}Fx$:

- same F, smaller k (less stiff) $\rightarrow$ larger x $\rightarrow$ more energy stored.
- same x, larger k (stiffer) $\rightarrow$ larger F $\rightarrow$ more energy stored.

When a stretched spring is released onto a mass, the elastic potential energy becomes **kinetic energy** 动能 (and gravitational potential energy if the mass rises). Set $\frac{1}{2}kx^2$ equal to $\frac{1}{2}mv^2$ (plus any mgh) to find the speed or height.

Exam tips

- **Hooke's law ($F = kx$)** holds only up to the limit of proportionality.
- **Stress** $= F/A$**, strain** $= x/L$**, Young modulus = stress/strain** (gradient of the straight part of the stress-strain graph) — watch the units (Pa).
- Energy stored = **area under the force-extension graph** $= \frac{1}{2}Fx$ in the elastic region.
- Distinguish **elastic** (returns to shape) from **plastic** (permanent) deformation.

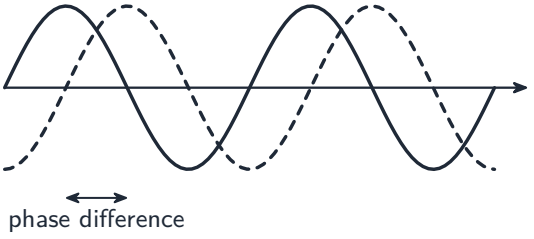
Waves

A-Level Physics

Progressive waves



Ripples spreading on water are progressive waves that carry energy outward.



Two waves of the same frequency, shifted by a phase difference

A **wave** 波 carries **energy** 能量 from one place to another without moving matter overall. The particles of the **medium** 介质 **oscillate** 振动 about fixed rest positions; only the disturbance (and its energy) **propagates** 传播. Examples: a **transverse wave** 横波 on a rope, a **longitudinal wave** 纵波 on a slinky spring, ripples on water, and sound in air. A wave that travels and carries energy is a **progressive wave** 行波.

Key terms

- **displacement** 位移 y — how far a particle has moved from its rest position at a moment. A **vector** 矢量.
- **amplitude** 振幅 A — the largest displacement from the rest position.
- **wavelength** 波长 λ — the shortest distance along the wave between two points that move **in phase** 同相 (for example, two next-door **crests** 波峰).
- **period** 周期 T — the time for one full oscillation of a particle.
- **frequency** 频率 f — the number of full oscillations per second; $f = 1/T$. Unit: **hertz** 赫兹, Hz.
- **speed** 速率 v — how fast a crest travels along the medium.
- **phase difference** 相位差 — the fraction of a cycle by which one oscillation leads or lags another. Given in **radians** 弧度 (a full cycle is 2π) or degrees (a full cycle is 360°).

Two points one wavelength apart are in phase (phase difference 0 or 2π). Two points half a wavelength apart are exactly out of phase (phase difference π).

The wave equation

In one period T, the wave moves forward by one wavelength λ. So speed = distance/time = $\lambda/T = \lambda f$:

$$v = f\lambda.$$

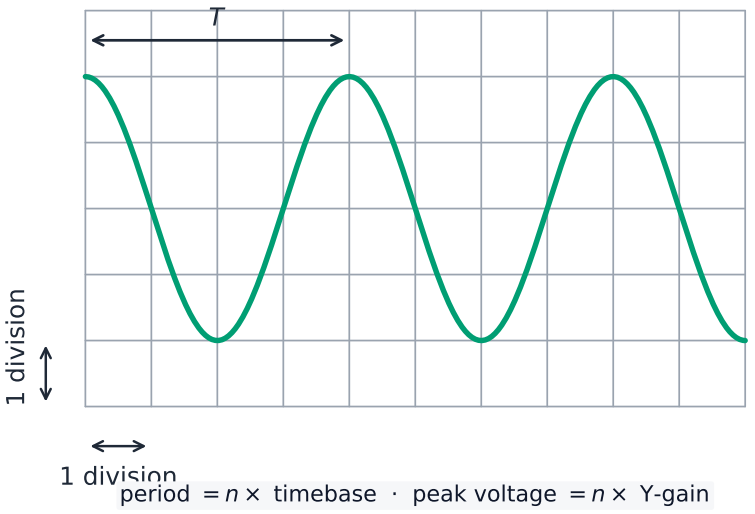
This comes straight from the definitions of speed, frequency and wavelength, and works for every progressive wave.

Reading a CRO trace

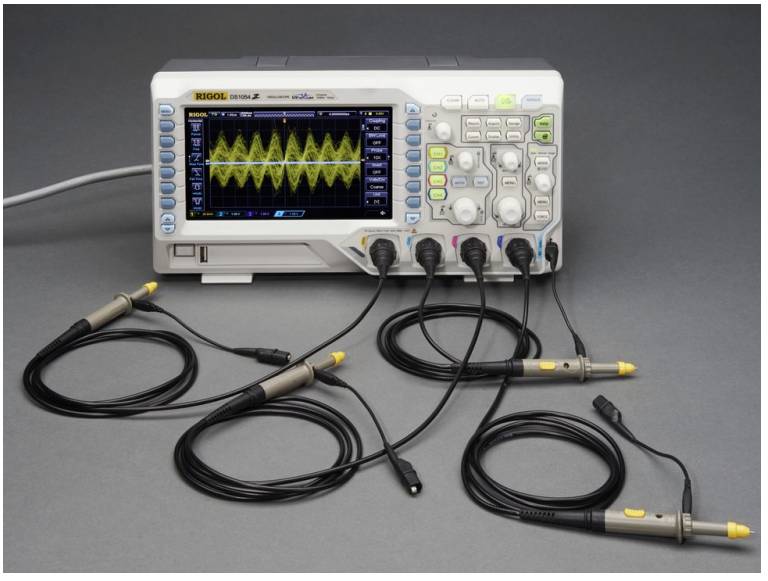
A **cathode-ray oscilloscope** 示波器 (CRO) draws a voltage signal — for sound, the output of a microphone — against time. Two controls matter:

- **time-base** 时基 (seconds per division across): turns horizontal distance on the screen into time. Read the period T as the distance between two next-door peaks, then $f = 1/T$.
- **y-gain** 垂直增益 (volts per division up): turns vertical distance into voltage. The amplitude in volts is the peak height from the centre line.

If the time-base is 5 ms/div and one full cycle takes 4 divisions, then $T = 4 \times 5$ ms = 20 ms and $f = 50$ Hz.



Reading the period T from a CRO trace using the grid and time-base



A real oscilloscope: the grid lets you read off the period and the amplitude

Intensity of a wave

A wave carries energy. The **intensity** 强度 at a point is the **power** 功率 passing through unit area at right angles to the direction of travel:

$$I = \frac{P}{A}.$$

Unit: W m^{-2}.

Intensity is proportional to the square of the amplitude:

$$I \propto A^2.$$

For a **point source** 点源 sending out energy equally in all directions, the **wavefronts** 波前 are spheres; the surface area at distance r is $4\pi r^2$, so

$$I = \frac{P}{4\pi r^2}, \qquad I \propto \frac{1}{r^2}.$$

Doubling the distance cuts the intensity to a quarter, which means the amplitude is halved (since $I \propto A^2$).

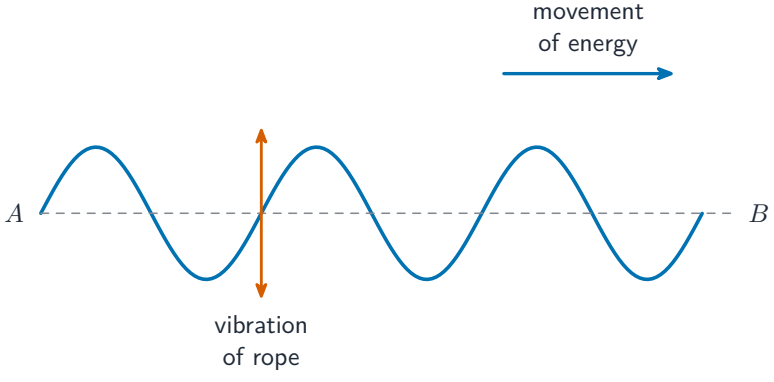
Worked example. A lamp emits 60 W of light equally in all directions. Find the intensity of the light 2.0 m away.

$$I = \frac{P}{4\pi r^2} = \frac{60}{4\pi (2.0)^2} \approx 1.2 \text{ W m}^{-2}.$$

Transverse and longitudinal waves

Transverse waves

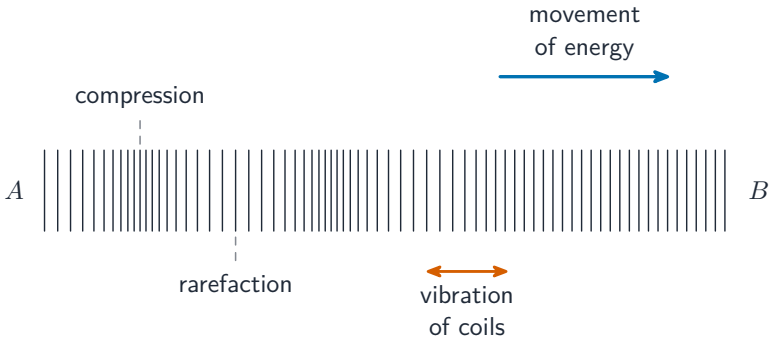
The particles oscillate **perpendicular** 垂直 to the direction the energy travels. A wave on a rope, all electromagnetic waves, and S-waves in the Earth are transverse.



Transverse wave on a rope

Longitudinal waves

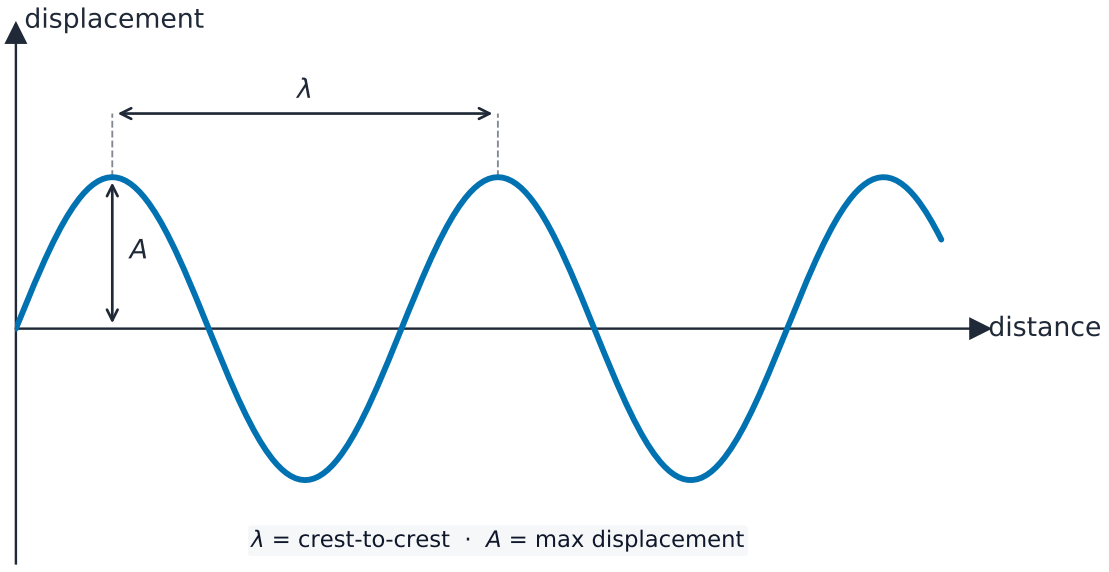
The particles oscillate **parallel** to the direction the energy travels. Sound in any medium, P-waves in the Earth, and the squashes on a slinky are longitudinal. The wave is made of **compressions** 压缩 (higher pressure, particles close together) and **rarefactions** 稀疏 (lower pressure, particles spread out).



Longitudinal wave on a slinky spring

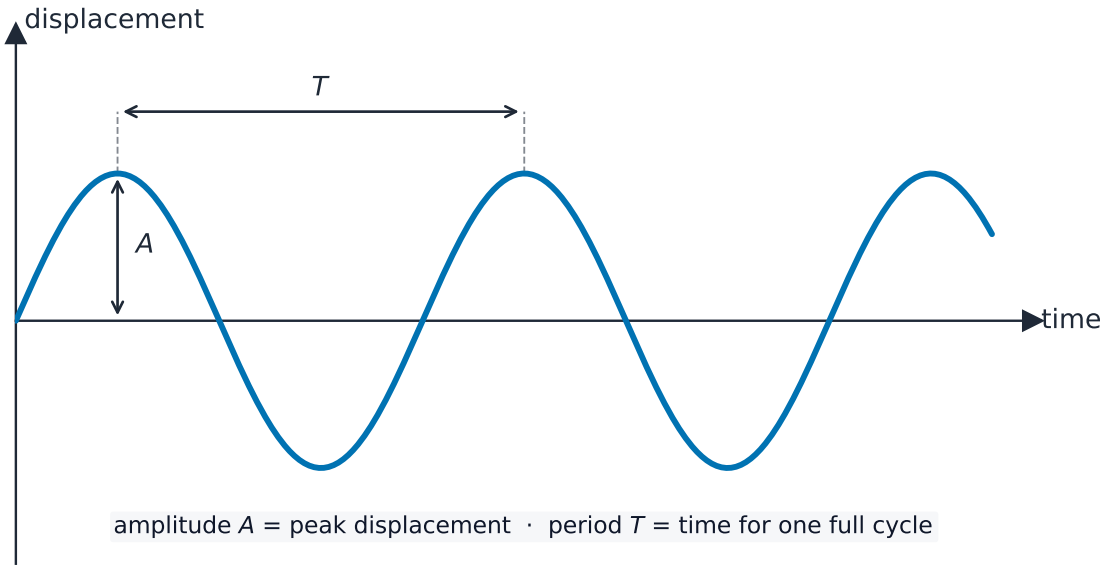
Graphs of waves

A graph of **particle displacement against position** at one moment looks like a **sine curve** 正弦曲线 for both kinds of wave. The difference: for a transverse wave the displacement axis is the real sideways displacement; for a longitudinal wave it is the small back-and-forth displacement along the direction of travel (positive one way, negative the other).



A displacement–distance graph shows the wave's amplitude and wavelength

A graph of **particle displacement against time** at one point in space is also a sine curve for both kinds. Read the period T from this graph.

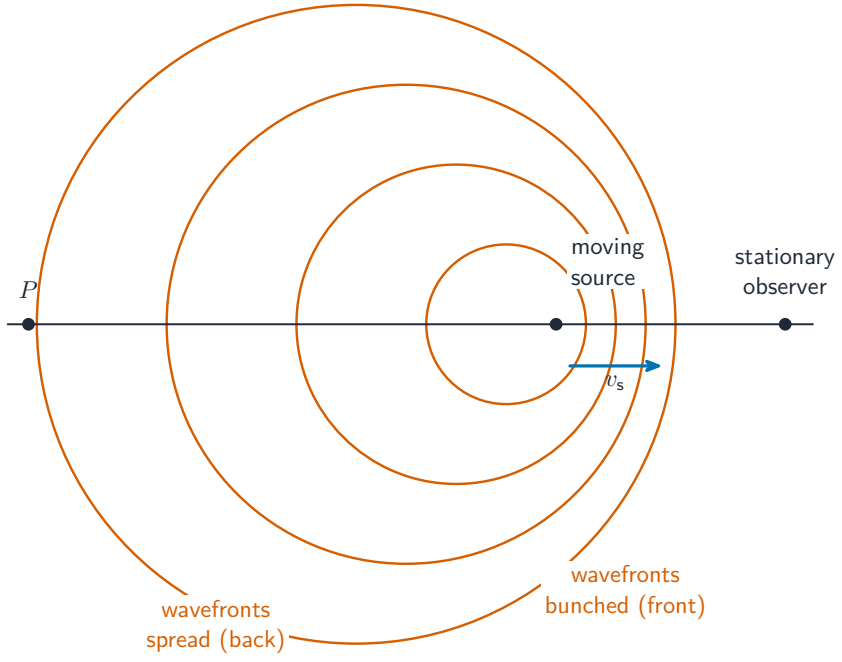


A displacement–time graph shows the wave's amplitude and period

Doppler effect (moving source, stationary observer)

When the **source** 波源 of a sound moves relative to a **stationary** 静止 **observer** 观察者, the heard frequency is different from the source frequency. This is the **Doppler effect** 多普勒效应.

- source moving towards the observer: the wavefronts in front are squashed, so the wavelength is shorter and the heard frequency is **higher**.
- source moving away from the observer: the wavefronts behind are spread out, so the wavelength is longer and the heard frequency is **lower**.



A moving source squashes the wavefronts ahead of it, raising the observed frequency

The formula (source moving at speed v_s along the line to the observer; wave speed v, source frequency f_s, heard frequency f_o):

$$f_o = \frac{v \cdot f_s}{v \pm v_s}.$$

Choose the sign to match the physics:

- minus sign on the bottom when the source moves **towards** the observer ($f_o > f_s$),
- plus sign when the source moves **away** ($f_o < f_s$).

You only need the case of a stationary observer.

Worked example. A car horn at $f_s = 800$ Hz moves at 30 m s^{-1} towards a still listener. Speed of sound $v = 340$ m s^{-1}:

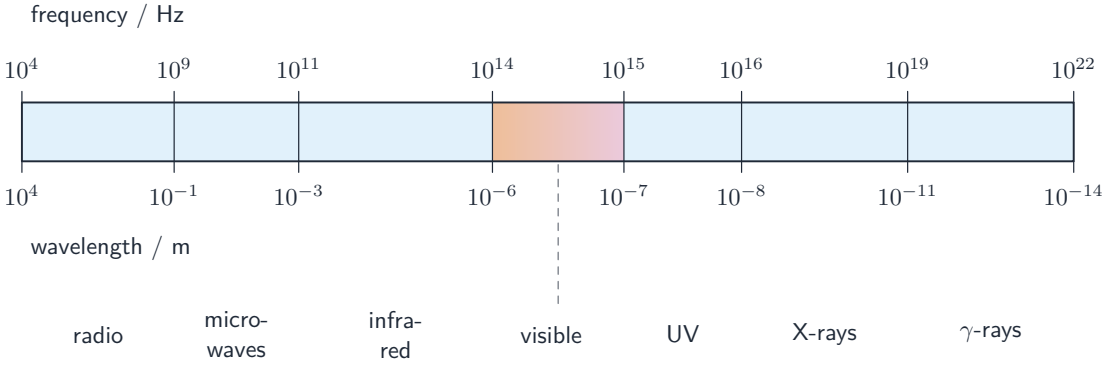
$$f_o = \frac{340 \times 800}{340 - 30} = \frac{272\,000}{310} \approx 877 \text{ Hz}.$$

Electromagnetic spectrum

All **electromagnetic waves** 电磁波 (EM waves) are **transverse** and travel in a **vacuum** 真空 at the same speed:

$$c = 3.00 \times 10^8 \text{ m s}^{-1}.$$

The **electromagnetic spectrum** 电磁波谱 includes radio waves, **microwaves** 微波, **infrared** 红外线, visible light, **ultraviolet** 紫外线, **X-rays** X 射线 and γ**-rays** γ 射线.



The electromagnetic spectrum

Approximate wavelength ranges in free space (learn the orders of magnitude):

- radio waves: $> 10^{-1}$ m (up to many km).
- microwaves: 10^{-3} m to 10^{-1} m.
- infrared: $\sim 7 \times 10^{-7}$ m to 10^{-3} m.
- visible light: 400 nm (violet) to 700 nm (red), i.e. 4×10^{-7} m to 7×10^{-7} m.
- ultraviolet: $\sim 10^{-8}$ m to 4×10^{-7} m.
- X-rays: $\sim 10^{-11}$ m to 10^{-8} m.
- γ-rays: $< 10^{-11}$ m.

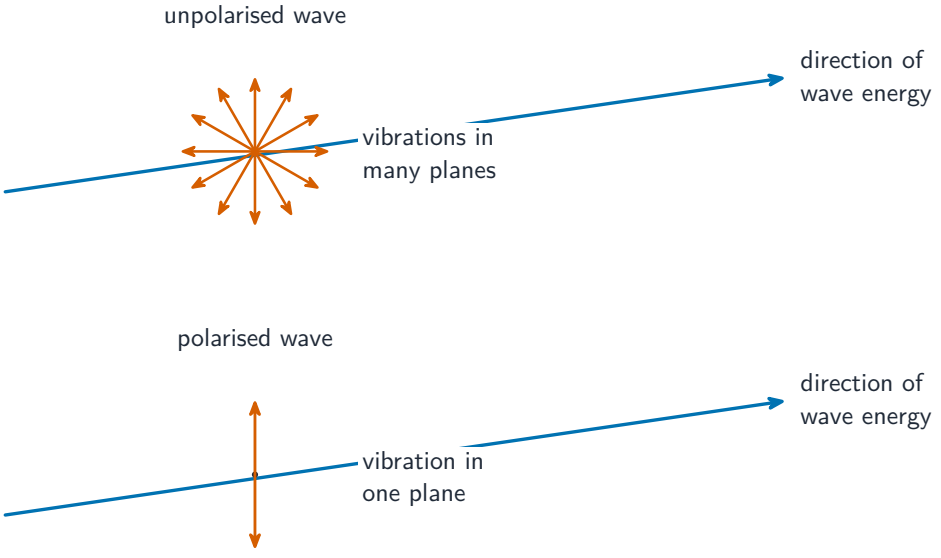
The boundaries between regions are not sharp. Use $c = f\lambda$ to change between wavelength and frequency. Only light with wavelengths 400–700 nm can be seen.

Polarisation

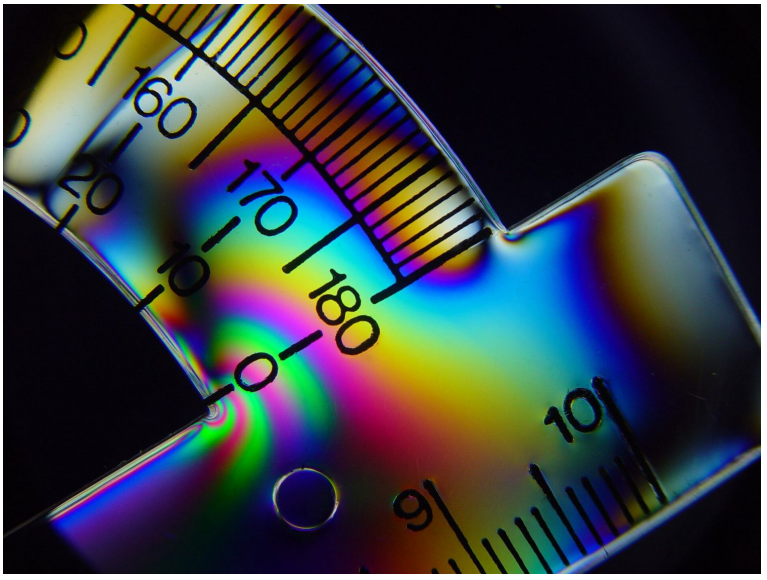
Polarisation 偏振 means making a transverse wave oscillate in one plane only.

- Only **transverse** waves can be polarised — the oscillation is perpendicular to the direction of travel, so different perpendicular planes are real choices.
- **Longitudinal** waves (sound) cannot be polarised — the oscillation is along the direction of travel, so there is no other plane.

So polarisation is a test: if a wave can be polarised, it must be transverse.



Unpolarised waves vibrate in many planes; a polarised wave vibrates in one plane



A see-through plastic protractor placed between two crossed polarising filters. Light only reaches your eye because the stressed plastic rotates its plane of polarisation — the colours map where the plastic is squeezed most. With ordinary light it would just look clear

Malus's law

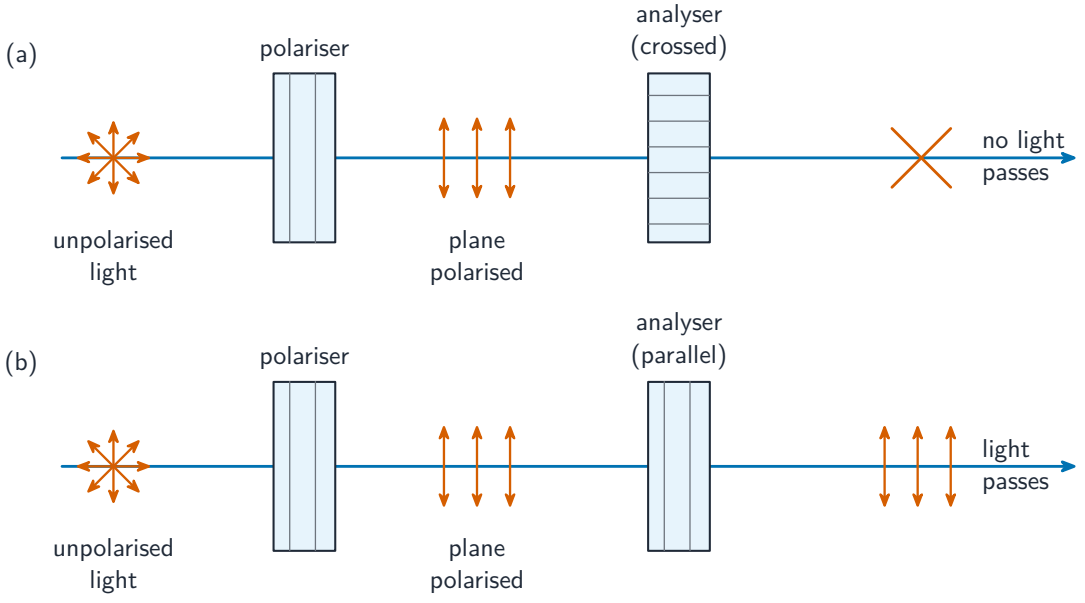
Plane-polarised 平面偏振 light of intensity I_0 passes through a **polarising filter** 偏振片 whose **transmission axis** 透光轴 is at angle θ to the plane of polarisation. The transmitted intensity is given by **Malus's law** 马吕斯定律:

$$I = I_0 \cos^2 \theta.$$

- $\theta = 0°$: filter lined up with the polarisation, $I = I_0$, all passes through.
- $\theta = 90°$: filter at right angles, $I = 0$, all blocked.
- $\theta = 60°$: $I = I_0 \cos^2 60° = I_0 \cdot 0.25 = I_0/4$.

Worked example. Plane-polarised light of intensity 12 W m^{-2} meets a polarising filter whose axis is at 30° to the plane of polarisation. Find the transmitted intensity.

$$I = I_0 \cos^2 \theta = 12 \times \cos^2 30° = 12 \times 0.75 = 9.0 \text{ W m}^{-2}.$$



Crossed filters (a) block the light; parallel filters (b) let it pass

For two filters in a row, use Malus's law twice with the angle between each pair. Be careful with the angle each time — after the first filter the polarisation is along that filter's axis, and the second filter's angle is measured from there.

(You do not need to work out the effect of a polarising filter on an unpolarised wave.)

Exam tips

- Define terms precisely (displacement, amplitude, wavelength, period, frequency) and use $v = f\lambda$.
- Distinguish **transverse** (vibration perpendicular to travel) from **longitudinal** (parallel); only transverse waves can be **polarised**.
- For the **Doppler effect**, the observed frequency **rises** as the source approaches and falls as it recedes.
- Learn the **electromagnetic spectrum** order; all its waves travel at c in a vacuum.

Superposition

A-Level Physics

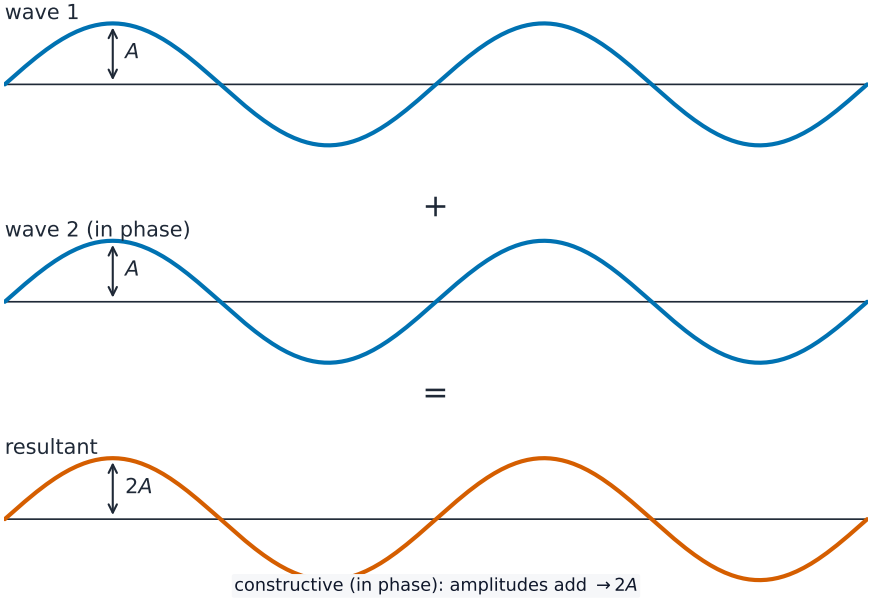
Principle of superposition

When two or more **waves** 波 overlap at a point, the **displacement** 位移 there is the **vector sum** 矢量和 of the displacements each wave would make on its own. This is the **principle of superposition** 叠加.

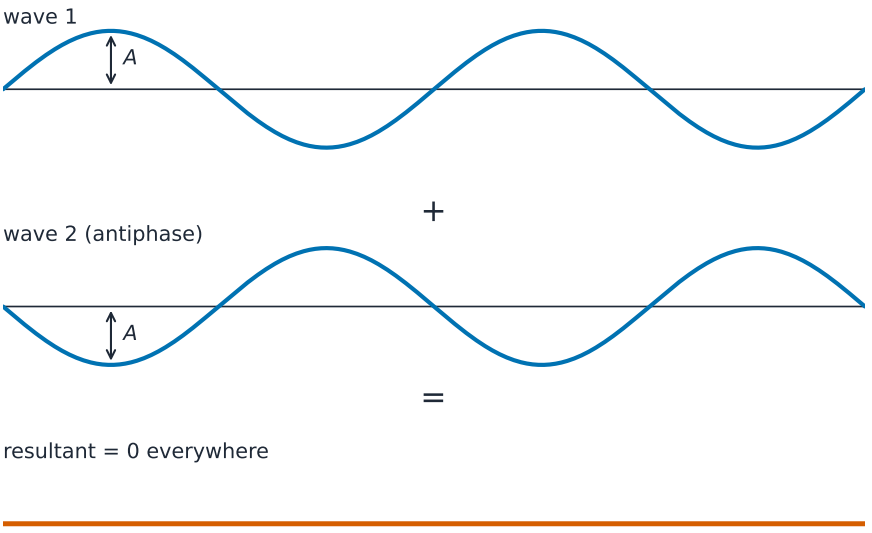
The waves pass through each other and come out unchanged. Superposition is the base of everything in this topic.

If two waves of **amplitude** 振幅 A_1 and A_2 meet:

- **in phase** 同相 (**crest** 波峰 meets crest): the amplitude is $A_1 + A_2$ (**constructive interference** 相长干涉).
- exactly out of phase (crest meets **trough** 波谷, **phase difference** 相位差 π): the amplitude is $|A_1 - A_2|$ (**destructive interference** 相消干涉).
- any other phase difference ϕ: the amplitude is somewhere between these two.



Two waves arriving in phase add to give double the amplitude (constructive)



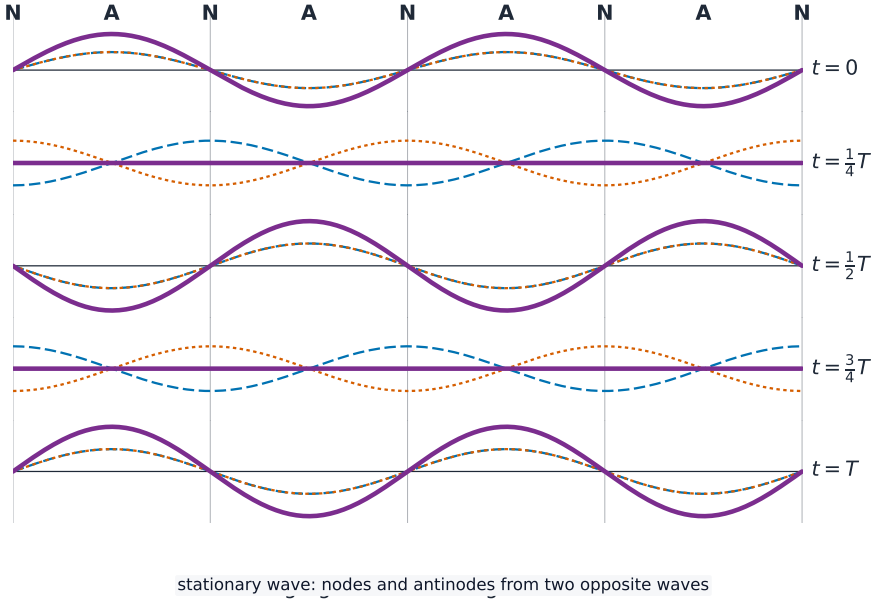
destructive (out of phase by λ/2): cancel if equal *A*

Two waves arriving exactly out of phase cancel to zero (destructive)

For **intensity** 强度, $I \propto A^2$. Two equal waves meeting in phase give intensity $(2A)^2 = 4A^2$ — **four times** the intensity of one wave alone.

Stationary (standing) waves

When two identical **progressive waves** 行波 travel in **opposite directions** and overlap, they make a **stationary wave** 驻波. Examples: a wave on a string **reflected** 反射 from a fixed end overlapping the incoming wave; sound in an air column reflected from a closed end; microwaves between an emitter and a metal sheet.



stationary wave: nodes and antinodes from two opposite waves

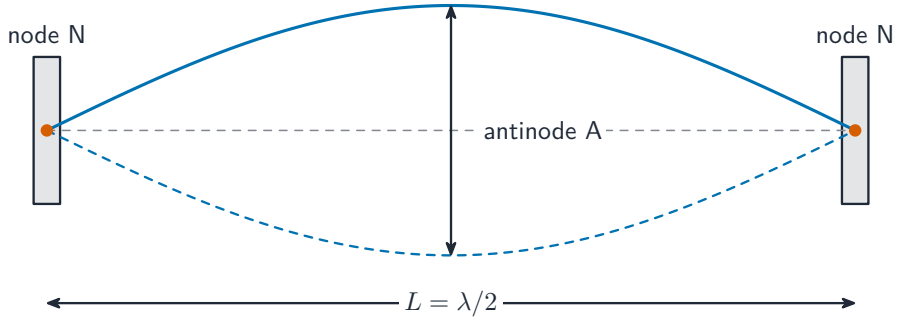
A stationary wave forms where two opposite waves overlap (N marks a node, A an antinode)

Nodes and antinodes

In a stationary wave:

- **node** 波节 — a point that is always at zero displacement (the two waves always cancel). The distance between next-door nodes is $\lambda/2$.
- **antinode** 波腹 — a point of largest amplitude (the two waves always add). The distance between next-door antinodes is $\lambda/2$.
- a node and the next antinode are $\lambda/4$ apart.

Particles between two nodes oscillate **in phase** with each other, but with different amplitudes (largest at the antinode, zero at the nodes). Particles on opposite sides of a node oscillate in **antiphase** 反相 (phase difference π).



Fundamental mode on a stretched string — one loop, with L equal to half a wavelength

Worked example. A string of length 0.80 m is fixed at both ends and vibrates in its fundamental mode, where the wave speed is 240 m s^{-1}. Find the fundamental frequency.

One loop fits the string, so $\lambda = 2L = 1.60$ m. Then

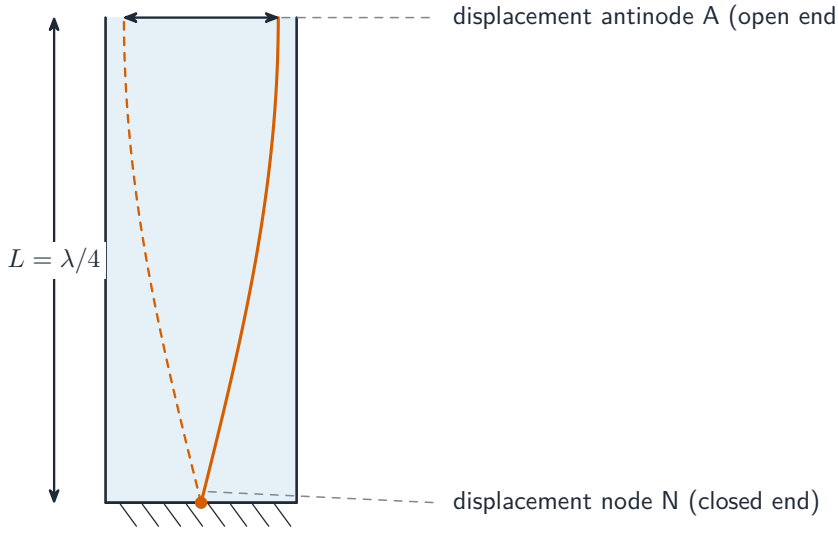
$$f = \frac{v}{\lambda} = \frac{240}{1.60} = 150 \text{ Hz}.$$

How a stationary wave differs from a progressive wave: it does not carry **energy** 能量 along its length, the pattern does not move along, and the nodes stay fixed; a progressive wave has the same amplitude everywhere and carries energy.

Measuring wavelength from node spacing

Drive a string with a vibrator at frequency f until a stationary pattern appears. Measure the distance between two well-separated nodes and divide by the number of half-**wavelengths** 波长 between them. Then λ is known, and $v = f\lambda$ gives the wave speed.

For a tube closed at one end and open at the other (a **resonance tube** 共鸣管), the closed end is a displacement node and the open end is a displacement antinode. The **fundamental** 基频 has $L = \lambda/4$; the next resonance is at $L = 3\lambda/4$; and so on. For a tube open at both ends, both ends are antinodes; the fundamental is at $L = \lambda/2$.



Fundamental mode in a closed pipe — a node at the closed end, an antinode at the open end

Diffraction

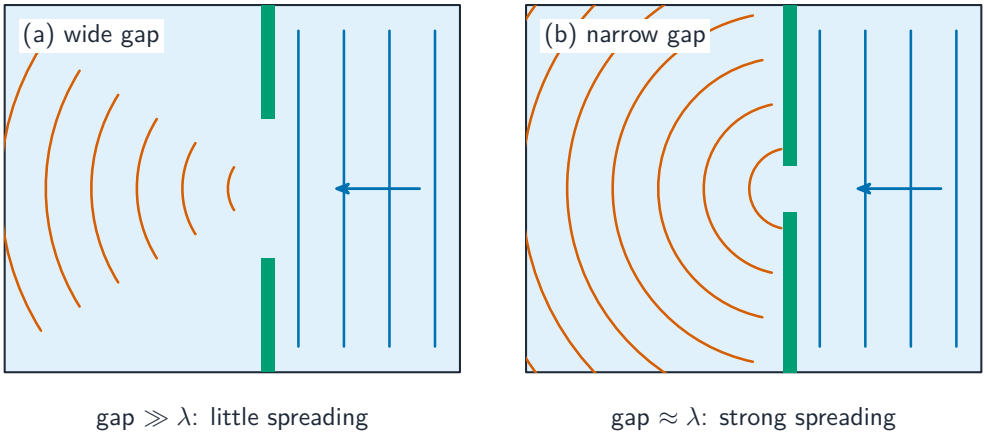
Diffraction 衍射 is the spreading of a wave after it passes through a gap or around an **obstacle** 障碍物. All waves diffract — water, sound, light, microwaves.

The amount of spreading depends on the ratio of **wavelength to gap width**:

- gap **much wider** than λ: very little spreading; the wave goes nearly straight through.
- gap **about the size of** λ: a lot of spreading; the wave fans out.
- gap **smaller than** λ: very strong spreading; the gap acts almost like a point source.

Show this with water waves in a **ripple tank** 水波槽: straight waves meet a barrier with a gap, and the waves curve more as the gap is made narrower. The same idea is

why you can hear someone around a corner (speech has λ near 1 m, close to the gap size) but cannot see them (visible light has $\lambda \sim 500$ nm, far smaller than the gap).



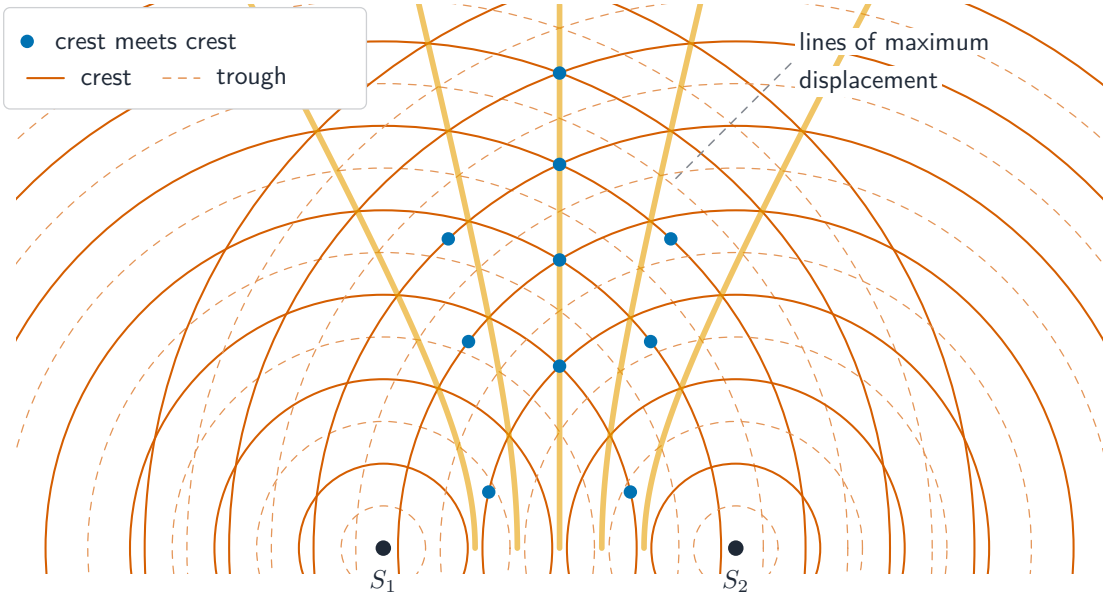
Diffraction in a ripple tank — a wide gap (a) spreads the waves little, a narrow gap (b) much more

Interference

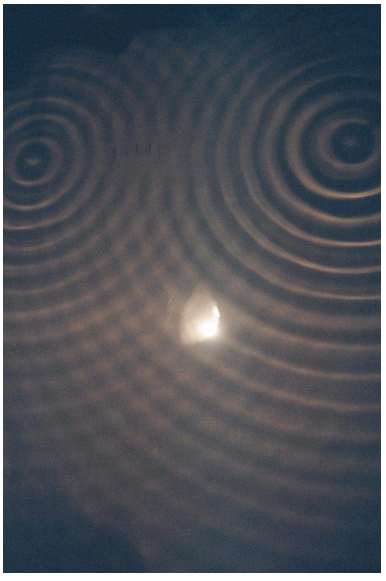


The shifting colours on a soap bubble come from the interference of light.

Interference 干涉 is the superposition of two **coherent** 相干 waves to give a steady pattern of high-amplitude regions (constructive) and low-amplitude regions (destructive).



Two coherent sources give lines of maximum displacement where crests meet crests



The same effect in a real ripple tank — two coherent sources give a steady interference pattern

Coherence

Two sources are **coherent** when they emit waves with a **constant phase difference** (which also needs the same frequency). Two separate lamps are not coherent — their phase changes randomly, so any pattern flickers too fast to see and you get only an average.

To make coherent light from one source, pass it through two **slits** 狭缝 in a **double-slit** 双缝 setup. Both slits are lit by the same wavefront, so the two beams keep a fixed phase relationship.

Conditions for a clear pattern

To see two-source fringes you need:

1. two **coherent** sources (constant phase difference).
2. roughly equal amplitudes (or the dark regions are not very dark).
3. the waves overlap where you look.
4. for light (a transverse wave), the same plane of **polarisation** 偏振.

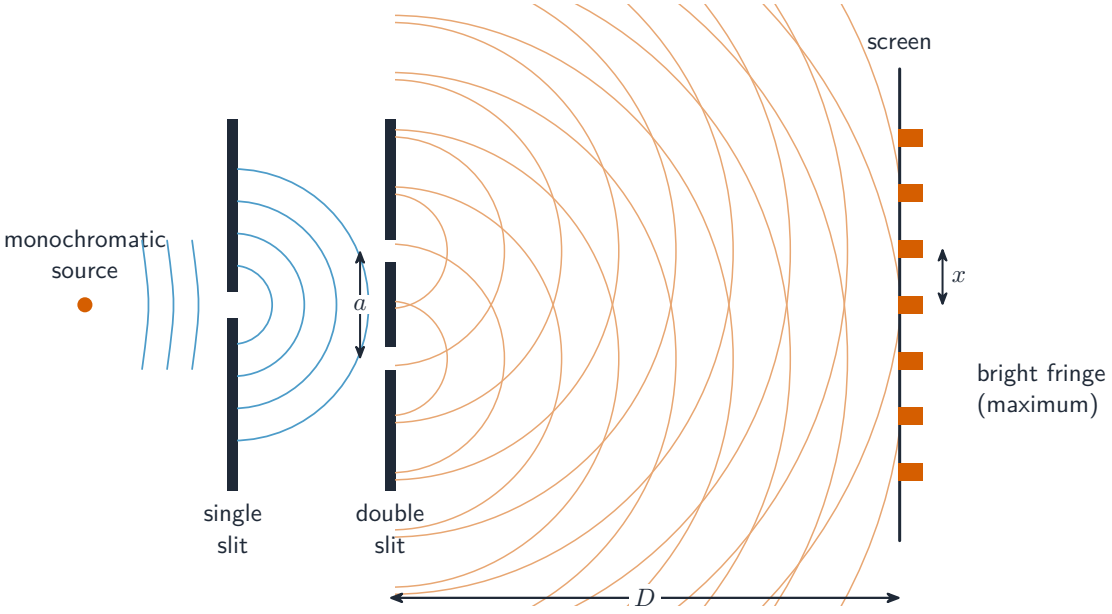
Path difference

For two coherent sources, what happens at a point depends on the **path difference** 路程差 Δx between the two waves arriving there:

- constructive: $\Delta x = n\lambda$ (for whole numbers $n = 0, 1, 2, \ldots$).
- destructive: $\Delta x = (n + \frac{1}{2})\lambda$.

Double-slit (Young's) experiment

For two slits a distance a apart, with a screen a distance D away (assume $D \gg a$), light of wavelength λ makes fringes on the screen.



Young's double-slit experiment — the single slit makes the two slits coherent sources

The **fringe spacing** 条纹间距 x (one **fringe** 条纹 to the next) is

$$\lambda = \frac{ax}{D}, \qquad x = \frac{\lambda D}{a}.$$

Bright fringes (**maximum** 极大) are where the path difference is a whole number of λ; dark fringes (**minimum** 极小) where it is $(n+\frac{1}{2})\lambda$. The fringes are equally spaced.

To make the fringe spacing smaller: increase a (slits further apart), reduce D (screen closer), or use a shorter λ (bluer light).

Worked example. In a double-slit experiment the slits are 0.50 mm apart and lit by light of wavelength 600 nm. The screen is 2.0 m away. Find the fringe spacing.

$$x = \frac{\lambda D}{a} = \frac{(600\times10^{-9})(2.0)}{0.50\times10^{-3}} = 2.4\times10^{-3}\ \text{m} = 2.4\ \text{mm}.$$

Diffraction grating

A **diffraction grating** 衍射光栅 has many equally spaced slits — often hundreds or thousands per millimetre. Each slit is a coherent source. A maximum is seen at angle θ from the **normal** 法线 to the grating when

$$d \sin\theta = n\lambda,$$

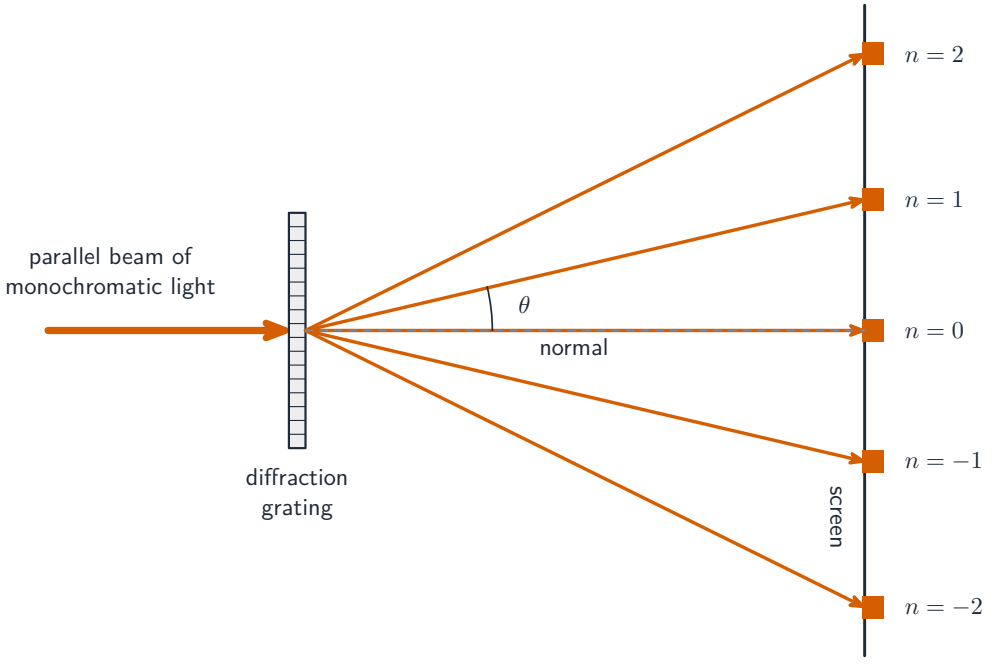
where d is the **slit spacing** 缝间距 (centre to centre), $n = 0, \pm 1, \pm 2, \ldots$ is the **order** 级次, and λ is the wavelength.

Worked example. A diffraction grating has 500 lines per mm. Light of wavelength 600 nm is shone normally on it. Find the angle of the first-order ($n = 1$) maximum.

The slit spacing is $d = \dfrac{1}{500}$ mm $= 2.0 \times 10^{-6}$ m, so

$$\sin\theta = \frac{n\lambda}{d} = \frac{600 \times 10^{-9}}{2.0 \times 10^{-6}} = 0.30 \quad \Rightarrow \quad \theta \approx 17°.$$

Compared with the double slit, a grating gives much **sharper** maxima, because more slits add together — every other direction is cancelled by many slits.



A diffraction grating splits monochromatic light into sharp maxima on a screen

Slit spacing from "lines per mm"

If a grating has N lines per millimetre, then $d = 1/N$ millimetres $= 10^{-3}/N$ metres. For 450 lines per mm, $d = 1/450$ mm ≈ 2.22 μm.

Highest order

For a given grating and wavelength, $\sin\theta = n\lambda/d$ cannot be more than 1, so the highest order seen is

$$n_{\max} = \left\lfloor \frac{d}{\lambda} \right\rfloor .$$

If $d/\lambda = 3.27$, orders up to $n = 3$ exist; $n = 4$ would need $\sin\theta > 1$ and is not seen.

Finding λ with a grating

Shine parallel light of unknown wavelength straight at the grating. Measure the angle θ_1 of the first-order maximum from the centre. Then $\lambda = d\sin\theta_1$. Repeating for higher orders and averaging reduces error.

Exam tips

- Two-source interference: **constructive** when path difference $= n\lambda$, **destructive** when $= (n + \frac{1}{2})\lambda$; the sources must be **coherent**.
- Double slit: $\lambda = ax/D$; diffraction grating: $d\sin\theta = n\lambda$ — know every symbol.
- On a **stationary wave** mark nodes and antinodes; adjacent nodes are $\lambda/2$ apart; it stores energy but does not transfer it.
- A stationary wave needs two waves of the **same frequency travelling in opposite directions**.

Electricity

A-Level Physics

Electric current

An **electric current** 电流 is a flow of **charge carriers** 载流子. In a metal the carriers are negative conduction **electrons** 电子; in an **electrolyte** 电解质 they are positive and negative **ions** 离子; in a **semiconductor** 半导体 they may be electrons or "**holes**" 空穴. The **conventional current** 常规电流 direction is the way **positive** charge would flow — opposite to the real flow of electrons in a wire.

Charge

Charge is **quantised** 量子化: the smallest free unit of charge is the **elementary charge** 基本电荷

$$e = 1.60 \times 10^{-19}\ \text{C}.$$

Every free charge in this syllabus is a whole-number multiple of e. The unit of charge is the **coulomb** 库仑, C.

Current as the rate of flow of charge

If charge Q passes a point in time t, the current is

$$I = \frac{Q}{t}, \qquad Q = It.$$

Unit of current: ampere, A ($= \text{C s}^{-1}$). For a changing current, the charge that has flowed in a time is the **area under** an I–t graph.

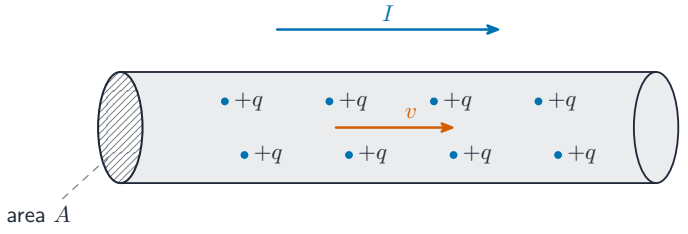
Worked example. A current of 0.50 A flows for 2.0 minutes. Find the charge that passes, and how many electrons this represents. ($e = 1.60 \times 10^{-19}$ C.)

$$Q = It = 0.50 \times 120 = 60\ \text{C}, \qquad N = \frac{Q}{e} = \frac{60}{1.60 \times 10^{-19}} \approx 3.8 \times 10^{20}.$$

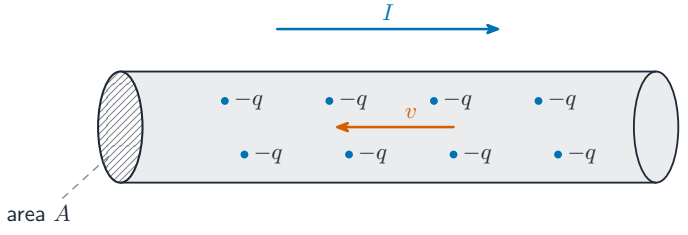
Drift velocity equation

For a uniform conductor of cross-section area A, with n charge carriers per unit volume (the **number density** 数密度), each carrying charge q, moving with average **drift velocity** 漂移速度 v:

$$I = Anvq.$$



(a) positive carriers drift with I



(b) electrons drift against I

Charge carriers drifting inside a conductor: positive carriers drift with I, electrons against it

Worked example. A copper wire of cross-sectional area 1.0×10^{-6} m^2 carries a current of 5.0 A. Copper has $n = 8.5 \times 10^{28}$ free electrons per m^3. Find the drift velocity. ($e = 1.60 \times 10^{-19}$ C.)

Rearranging $I = Anvq$ gives $v = \dfrac{I}{Anq}$:

$$v = \frac{5.0}{(1.0 \times 10^{-6})(8.5 \times 10^{28})(1.60 \times 10^{-19})} \approx 3.7 \times 10^{-4} \text{ m s}^{-1}.$$

The electrons drift very slowly — less than a millimetre per second.

Use this to compare currents:

- a thinner wire (smaller A) at the same I needs a faster drift v.

- a semiconductor has far fewer free carriers than a metal (smaller n), so for the same I the drift velocity is much larger.
- in **series** 串联 components, I is the same everywhere, so if A stays the same but the material changes, nv changes the other way.

Potential difference

The **potential difference** 电势差 (p.d.) across a component is the **energy** 能量 transferred per unit charge as that charge passes through it:

$$V = \frac{W}{Q}.$$

Unit: **volt** 伏特, V ($= \text{J C}^{-1}$).

If 1 J of electrical energy changes into other forms (thermal, light, kinetic, …) when 1 C of charge passes through a component, the p.d. across it is 1 V.

The **electromotive force** 电动势 (e.m.f.) of a source is the energy given **per unit charge by the source**. The formula is the same as for p.d.; the difference is direction: e.m.f. is energy given **to** the charge by the source; p.d. is energy given **up** by the charge to the component.

Electrical power



High-voltage power lines carry electrical energy across the country.

Combining $V = W/Q$ and $I = Q/t$:

$$P = \frac{W}{t} = VI.$$

Using Ohm's law $V = IR$:

$$P = VI = I^2R = \frac{V^2}{R}.$$

Pick the form with the quantities you know. Examples:

- two heaters of equal resistance — the one with the larger current gives more **power** 功率 ($P = I^2R$).
- two resistors in **parallel** 并联 across the same **voltage** 电压 — the one with smaller R gives more power ($P = V^2/R$).
- a kettle marked "2.4 kW, 240 V" draws $I = P/V = 10$ A and has resistance $R = V^2/P = 24\ \Omega$.

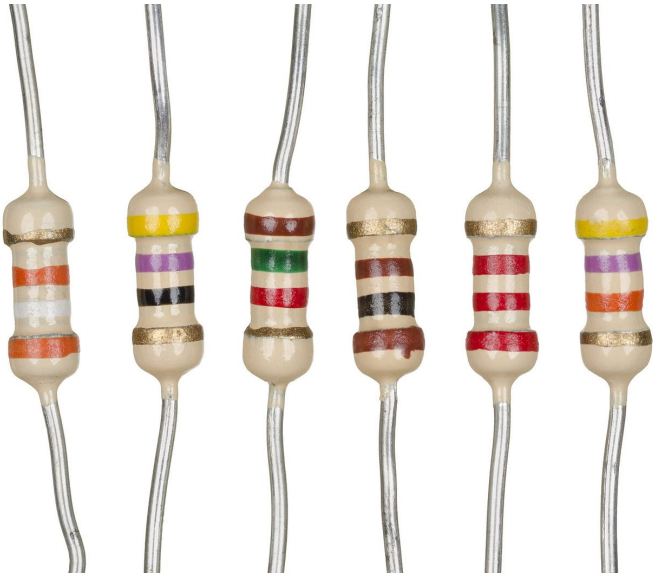
Energy transferred in time t is $E = Pt$.

Resistance and Ohm's law

The **resistance** 电阻 R of a component is

$$R = \frac{V}{I}.$$

Unit: **ohm** 欧姆, Ω ($= \text{V A}^{-1}$). Resistance depends on the conditions (such as temperature) when it is measured.



Real fixed resistors — the coloured bands code the resistance in ohms

Ohm's law

A conductor obeys **Ohm's law** 欧姆定律 when the current through it is **proportional to the p.d.** across it, as long as the conditions (especially temperature) stay constant. For such a conductor R is constant and the I–V graph is a straight line through the origin.

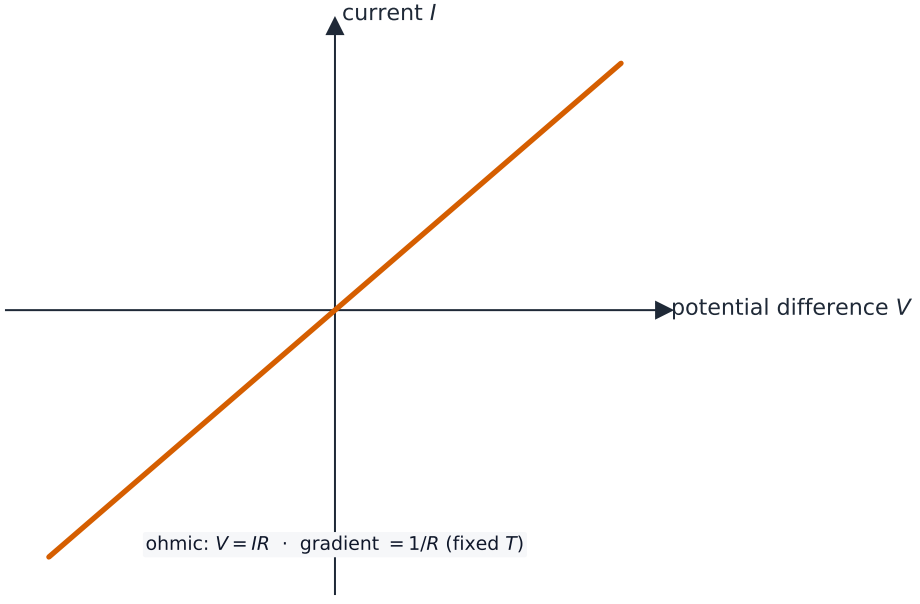
Ohm's law is an experimental result, not a definition. The definition $R = V/I$ works for **any** component; only ohmic ones have constant R.

I–V characteristics

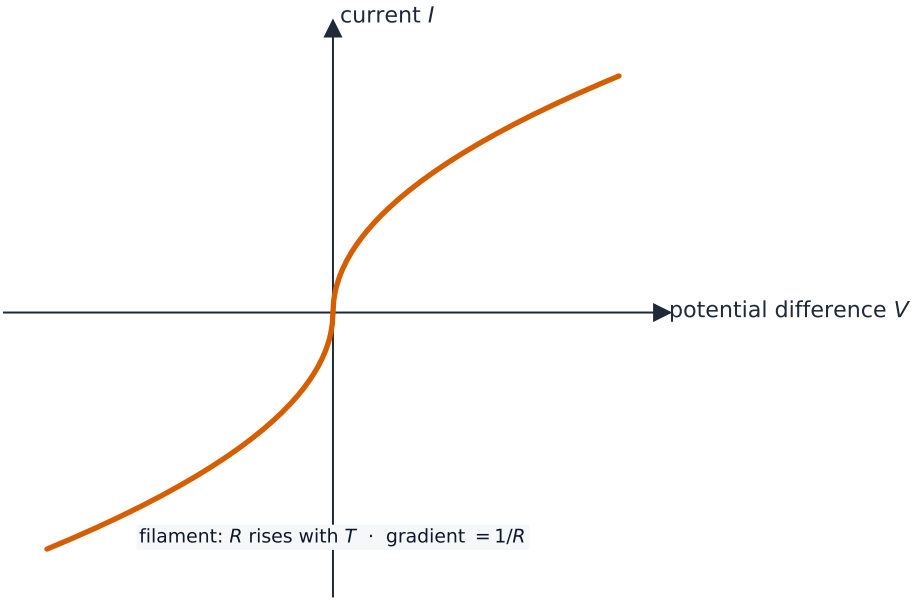
You should be able to sketch these:

- **metal wire at constant temperature** — a straight line through the origin (constant R). Reversing the p.d. drives the current the other way, giving a straight line in both directions.

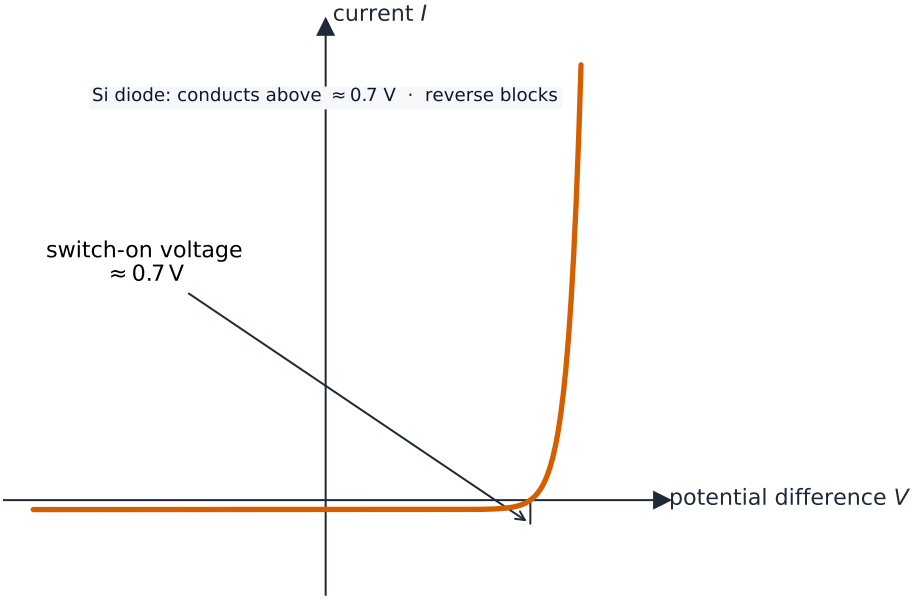
- **filament lamp** 灯丝灯泡 — through the origin, steep at first, then flatter as V (and I) grow. Reason: more current heats the filament, so its resistance rises and the gradient $1/R$ falls.
- **semiconductor diode** 二极管 — almost no current for negative V or small positive V. Above a "switch-on" voltage (about 0.7 V for silicon), the current rises sharply.



I–V characteristic of an ohmic conductor (metal wire at constant temperature)



I–V characteristic of a filament lamp



I–V characteristic of a semiconductor diode

Resistivity

For a uniform conductor of length L and cross-section area A,

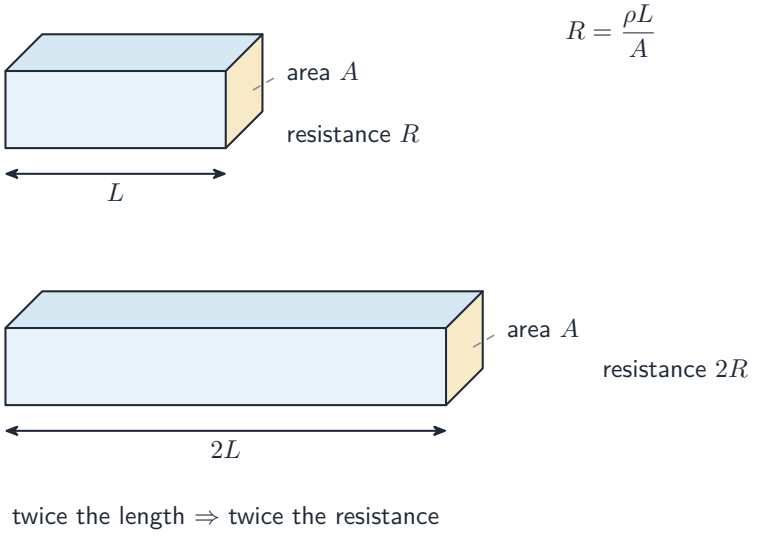
$$R = \frac{\rho L}{A}.$$

ρ is the **resistivity** 电阻率, a property of the material, with unit Ω m. Doubling the length doubles R; doubling the area halves it; halving the diameter quarters the area and so makes R four times bigger.

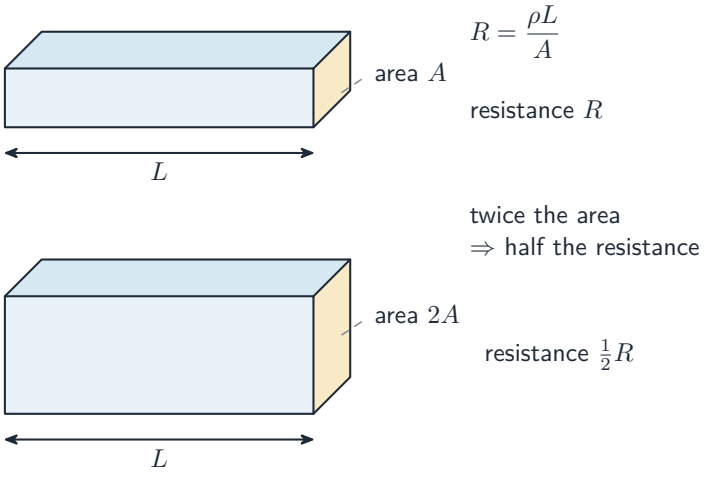
Worked example. A copper wire of length 2.0 m and cross-sectional area 1.7×10^{-7} m^2 has resistivity 1.7×10^{-8} Ω m. Find its resistance.

$$R = \frac{\rho L}{A} = \frac{(1.7 \times 10^{-8})(2.0)}{1.7 \times 10^{-7}} = 0.20\ \Omega.$$

Typical values: copper at room temperature $\rho \sim 1.7 \times 10^{-8}$ Ω m; an **insulator** 绝缘体 $\rho \sim 10^{15}$ Ω m or more.



A longer conductor has more resistance — doubling L doubles R

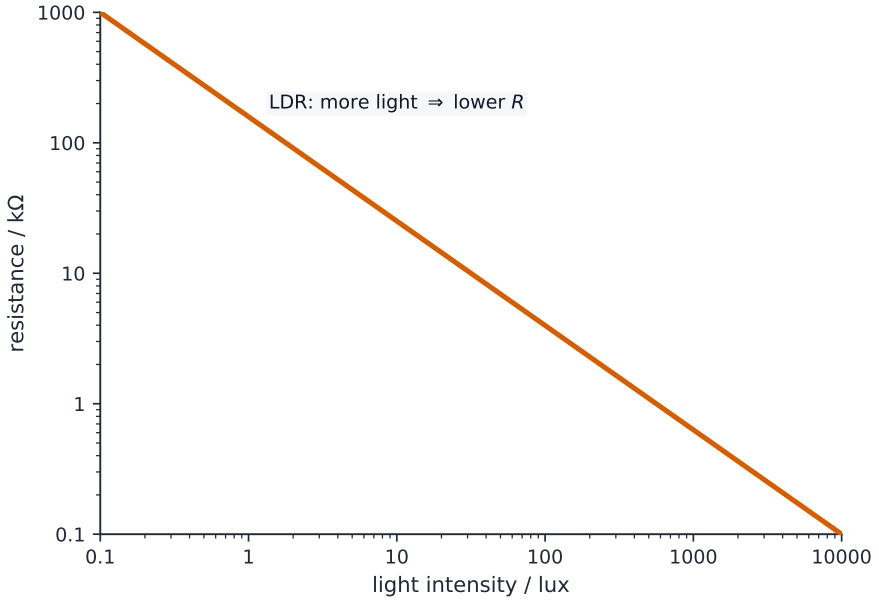


A wider conductor has less resistance — doubling A halves R

The resistivity of a metal **rises with temperature** (more **lattice vibration** 晶格振动 **scatters** 散射 the electrons), which is why the filament lamp's I–V line curves.

Light-dependent resistor (LDR)

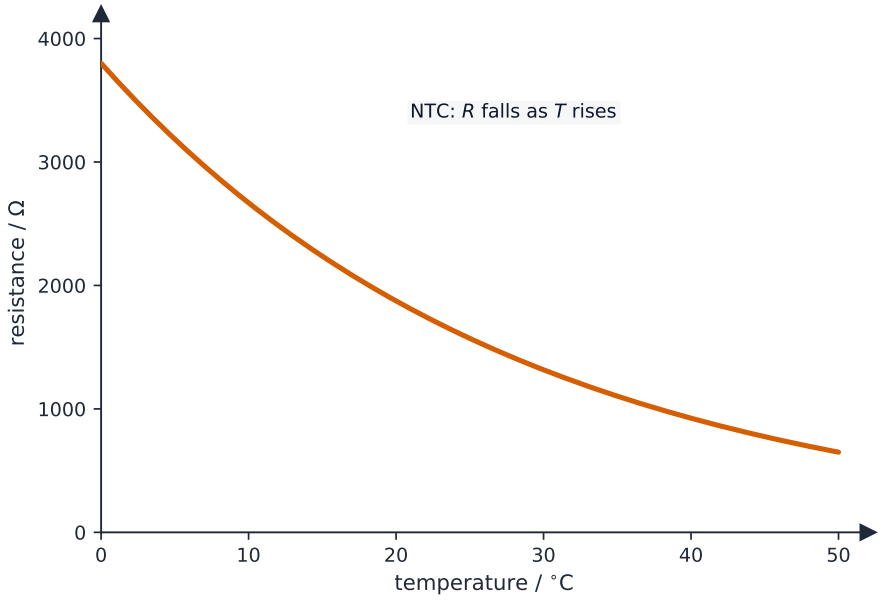
A **light-dependent resistor** 光敏电阻 (LDR) is a semiconductor whose **resistance falls as the light intensity rises**. In bright light R may be a few hundred Ω; in the dark it can be in the megaohms. LDRs are used in light-sensing circuits (street lamps, camera light meters). Here the **light intensity** 光强 controls the resistance.



Resistance of an LDR decreases as light intensity increases

Thermistor

In this syllabus a **thermistor** 热敏电阻 has a **negative temperature coefficient** 负温度系数: its **resistance falls as its temperature rises**. This is useful for sensing temperature — put it in a **potential divider** 分压器 and the output voltage changes with temperature.



Resistance of a thermistor falls as temperature rises

This is the opposite of a metal: in a semiconductor, more thermal energy frees more charge carriers, and this matters more than the extra scattering.

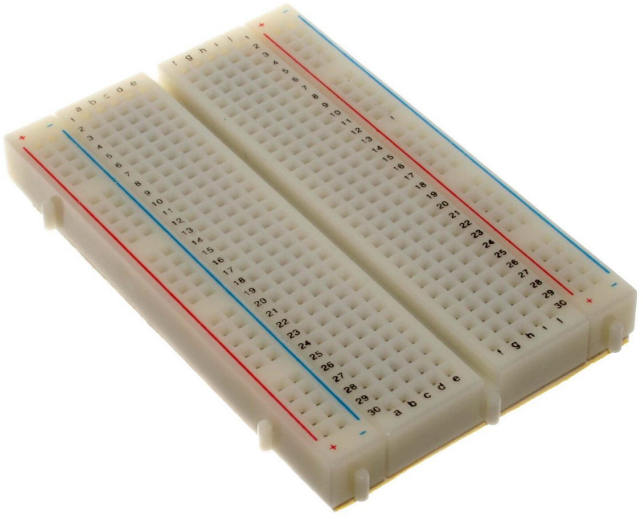
Exam tips

- Use $I = Q/t$, $V = W/Q$ (energy per unit charge) and $P = VI = I^2R = V^2/R$.
- **Ohm's law ($V = IR$)** applies only to an ohmic conductor at constant temperature — a filament lamp is non-ohmic.
- Sketch and interpret the I-V **characteristics** of a resistor, filament lamp and diode.
- An **LDR's** resistance falls with light; a **thermistor's** falls as temperature rises.

D.C. circuits

A-Level Physics

Practical circuits



A breadboard lets you build and test practical circuits without soldering.

e.m.f. and p.d.

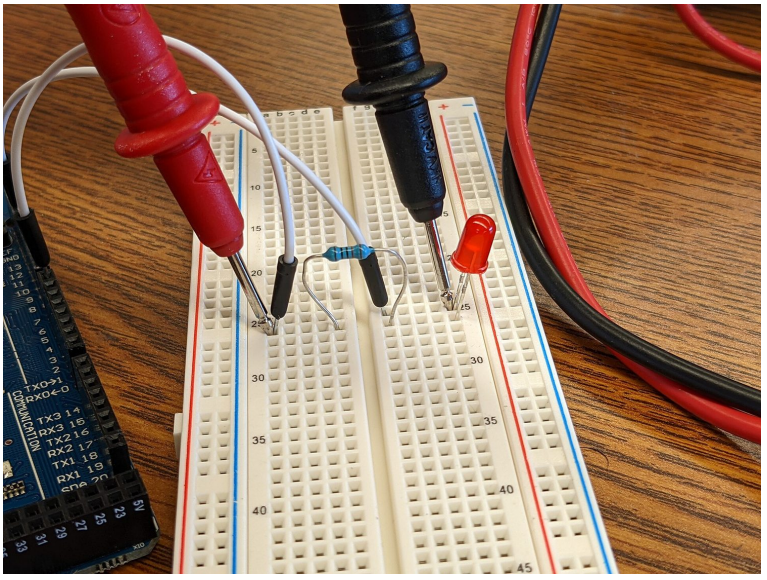
The **electromotive force** 电动势 (e.m.f.) ε of a source is the **energy** 能量 given to each unit of charge by the source as it drives the charge around a full circuit. Unit: volt.

The **potential difference** 电势差 (p.d.) across a component is the energy changed from electrical to other forms by each unit of charge as it passes through that component.

Both are in volts; they differ in energy direction:

- e.m.f. — energy put **into** the circuit by the source (chemical → electrical in a battery, mechanical → electrical in a generator).
- p.d. — energy taken **out** of the electrical form (electrical → thermal in a resistor, → light in a lamp, → kinetic in a motor).

In the lab you often build a circuit on a **breadboard** 面包板 (a board with rows of holes that connect components without soldering) and measure currents and p.d.s with a **multimeter** 万用表.



A real circuit on a breadboard, being measured with a multimeter

Internal resistance

A real source has some **internal resistance** 内阻 r — usually the resistance of the **electrolyte** 电解质 in a **cell** 电池, or the wire windings in a generator. When current I flows, an internal p.d. of Ir is "lost" inside the source, so the **terminal p.d.** 端电压 across the outside circuit is

$$V_{\text{terminal}} = \varepsilon - Ir.$$

So:

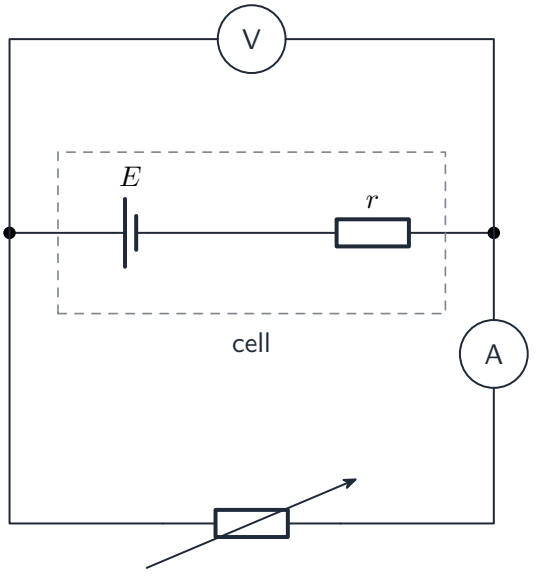
- **no current** (**open circuit** 开路, $I = 0$): the terminal p.d. equals the e.m.f.
- **larger current**: the terminal p.d. falls.
- **short circuit** 短路 ($R_{\text{external}} \to 0$): $I = \varepsilon/r$, a large current, with all the energy turned to heat inside the source.

To measure r, change the outside resistance and plot V_{terminal} against I: the line has y-intercept ε and gradient $-r$.

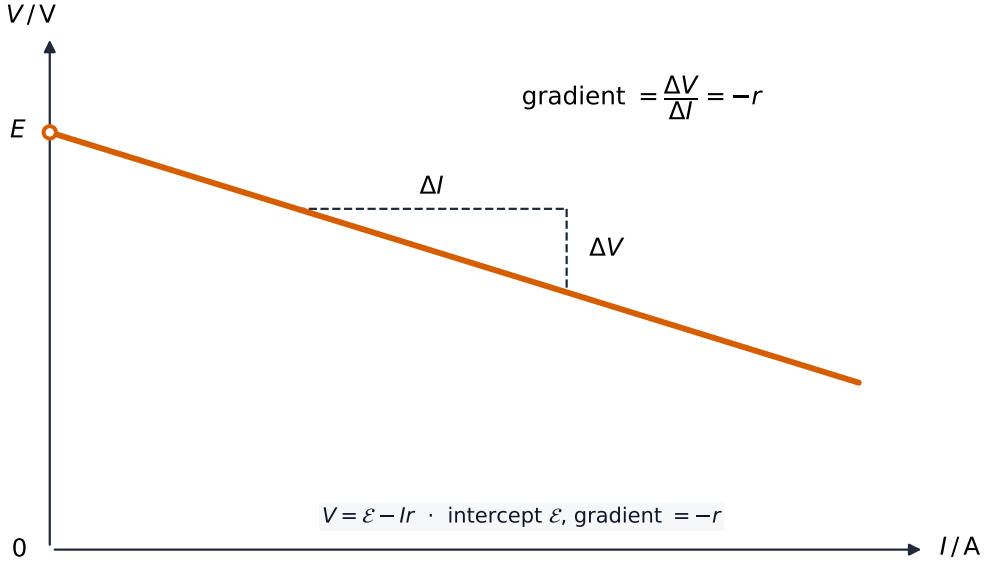
Worked example. A cell of e.m.f. 1.5 V and internal resistance 0.50 Ω is connected to a 2.5 Ω resistor. Find the current and the terminal p.d.

The e.m.f. drives the current through both resistances: $I = \dfrac{\varepsilon}{R + r} = \dfrac{1.5}{2.5 + 0.50} = 0.50$ A. Then

$$V_{\text{terminal}} = \varepsilon - Ir = 1.5 - 0.50 \times 0.50 = 1.25 \text{ V}.$$



Circuit for measuring the e.m.f. and internal resistance of a cell

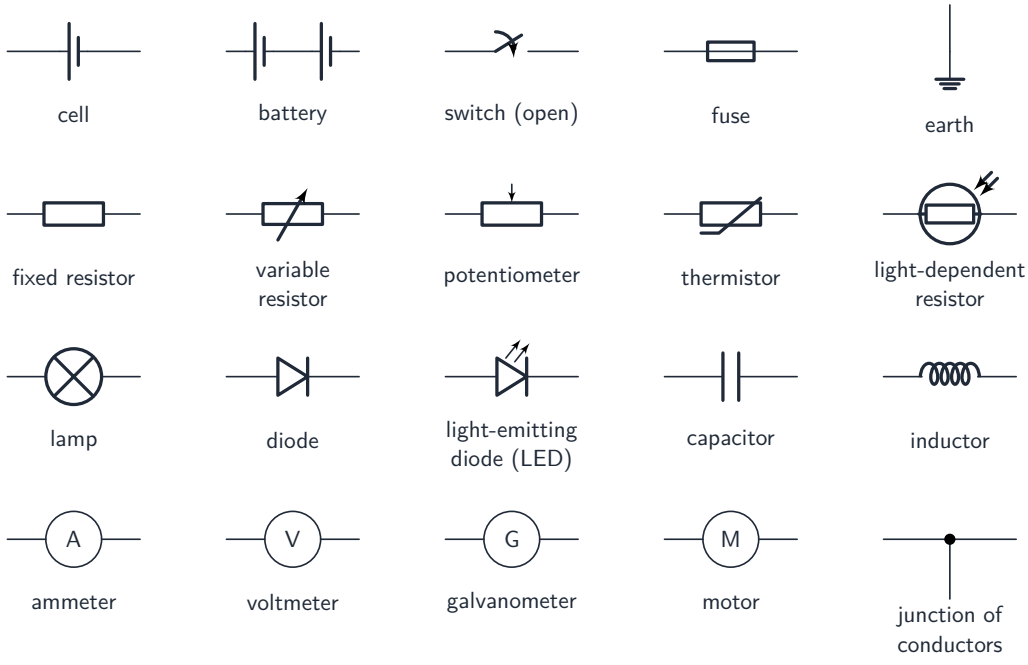


Terminal p.d. against current — the intercept is the e.m.f. and the gradient is minus the internal resistance

The **power** 功率 given to the outside load is $P_{\text{ext}} = (\varepsilon - Ir)I$; the power lost inside is $P_{\text{int}} = I^2 r$; the total power from the source is εI.

Circuit symbols

You must recognise and draw the standard symbols in the syllabus: cell, battery, switch, resistor, variable resistor, **ammeter** 电流表, **voltmeter** 电压表, lamp, **diode** (and **LED** 发光二极管), **capacitor** 电容器, inductor, thermistor, light-dependent resistor, **fuse** 保险丝, earth, junction. An ideal ammeter has zero **resistance** 电阻 and goes in **series** 串联. An ideal voltmeter has infinite resistance and goes in **parallel** 并联.



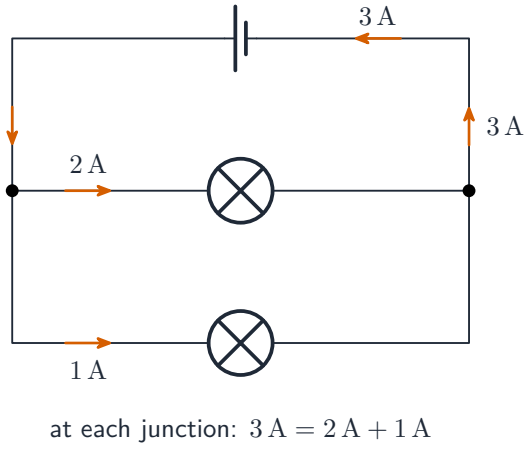
The standard circuit symbols you need to recognise and draw

Kirchhoff's laws

First law (junction rule)

At any **junction** 节点, **the total current flowing in equals the total current flowing out**. This follows from **conservation of charge** 电荷守恒 — charge cannot build up at a point in a steady circuit, so charge in per second equals charge out per second.

For a junction with three wires: $I_1 = I_2 + I_3$ if currents 2 and 3 flow out and current 1 flows in.



Current divides at a junction in a parallel circuit (3 A in equals 2 A plus 1 A)

Second law (loop rule)

Around **any closed loop** 回路, **the total e.m.f. equals the total p.d. across the components in that loop**. This follows from **conservation of energy** 能量守恒: as a unit of charge goes once round a loop, the energy it gains from sources equals the energy it gives up to components.

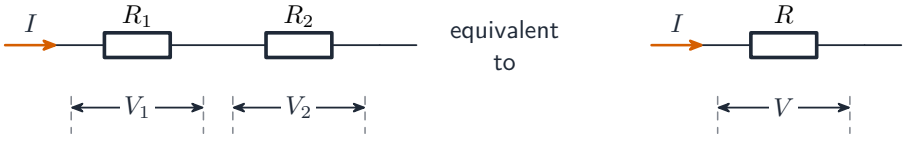
Pick a direction round the loop. Take an e.m.f. as positive when the loop direction goes from − to + of the source, and a p.d. as positive when the loop direction is the conventional current direction through the resistor.

Combining resistors

Resistors in series carry the same current; the total p.d. is the sum:

$$\varepsilon = IR_1 + IR_2 + \ldots = I(R_1 + R_2 + \ldots),$$

so $R_{\text{series}} = R_1 + R_2 + \ldots$.

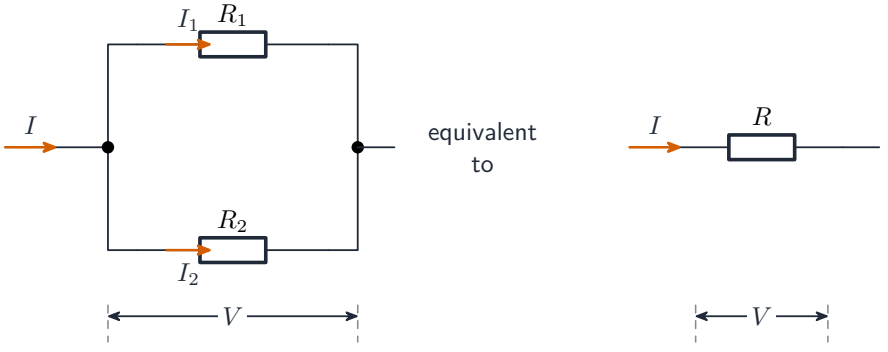


Two resistors in series and their single equivalent resistor

Resistors in parallel have the same p.d.; the total current is the sum:

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \ldots = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \ldots\right),$$

so $\dfrac{1}{R_{\text{parallel}}} = \dfrac{1}{R_1} + \dfrac{1}{R_2} + \ldots$.



Two resistors in parallel and their single equivalent resistor

Two equal resistors R in parallel give $R/2$; N equal ones give R/N. A parallel combination is **always smaller** than any of its resistors; a series combination is always larger.

Worked example. A 4.0 Ω resistor and a 12 Ω resistor are connected in parallel. Find their combined resistance.

$$\frac{1}{R} = \frac{1}{4.0} + \frac{1}{12} = \frac{3}{12} + \frac{1}{12} = \frac{4}{12} = \frac{1}{3} \quad \Rightarrow \quad R = 3.0\ \Omega.$$

Solving a circuit

1. **Label** every current with a symbol and a chosen direction.
2. **Use Kirchhoff's first law** 基尔霍夫第一定律 at each junction to link the currents.
3. **Use Kirchhoff's second law** 基尔霍夫第二定律 around each loop to get equations in the p.d.s.
4. **Use** $V = IR$ for each resistor.
5. Solve the equations together.

For symmetric resistor networks, use the symmetry to spot branches with equal currents — the branch with the most current gives the most power ($P = I^2R$).

Potential dividers

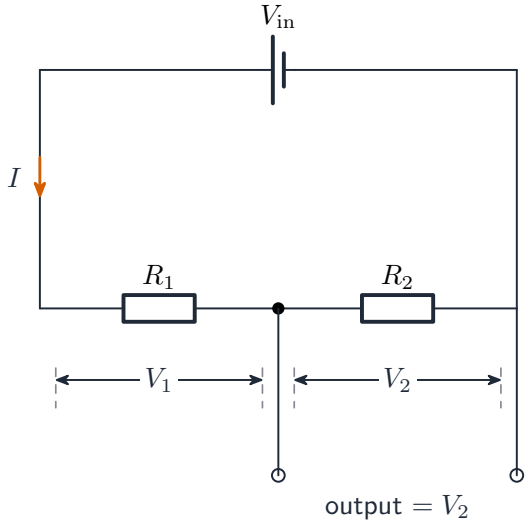
A **potential divider** 分压器 is two (or more) resistors in series across a source. The p.d. across each resistor is in **direct proportion** to its resistance:

$$V_1 = V_{\text{in}} \cdot \frac{R_1}{R_1 + R_2}, \qquad V_2 = V_{\text{in}} \cdot \frac{R_2}{R_1 + R_2}.$$

The output (tapped between R_1 and R_2) can be set to any **voltage** 电压 between 0 and V_{in} by choosing the resistances. A **rheostat** 变阻器 (a slider on a uniform-resistance wire) gives a smoothly variable divider.

Worked example. A 6.0 V supply is connected across a 2.0 kΩ resistor in series with a 4.0 kΩ resistor. Find the output voltage tapped across the 4.0 kΩ resistor.

$$V_2 = V_{\text{in}} \cdot \frac{R_2}{R_1 + R_2} = 6.0 \times \frac{4.0}{2.0 + 4.0} = 4.0 \text{ V}.$$



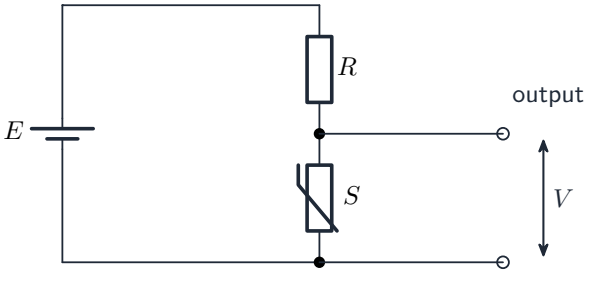
A potential divider — the p.d. splits between R1 and R2 in proportion to their resistances

Sensor circuits

Replace one fixed resistor with a **sensor** 传感器 whose resistance changes with a physical quantity:

- **thermistor** 热敏电阻 (NTC): R falls as temperature rises. In a divider, the output voltage changes with temperature in a fixed direction.
- **light-dependent resistor** 光敏电阻 (LDR): R falls as **light intensity** 光强 rises, giving a brightness-dependent output.

Connect the output to a **transistor** 晶体管 base or a **comparator** 比较器 to switch a load on or off when the temperature or light passes a **threshold** 阈值.



A thermistor in a potential divider gives an output voltage that changes with temperature

Potentiometer and the null method

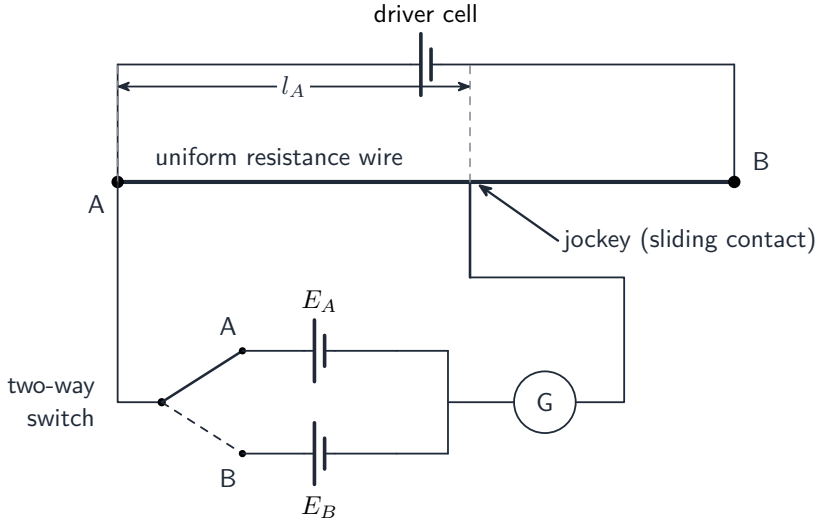
A **potentiometer** 电位差计 is a uniform resistance wire of length L_0 with a sliding contact (**jockey** 滑动触头). The resistance per unit length is uniform, so the p.d. from one end to the jockey is proportional to the length:

$$V_x = V_{\text{full}} \cdot \frac{x}{L_0}.$$

To **compare two e.m.f.s** (an unknown cell against a standard cell), connect each in turn with the jockey through a **galvanometer** 检流计. Slide the jockey until the galvanometer reads zero (a **null** — no current flows through the cell being measured, because the potentiometer's voltage there exactly opposes the cell's e.m.f.). The two balance lengths are in the ratio of the e.m.f.s:

$$\frac{\varepsilon_1}{\varepsilon_2} = \frac{l_1}{l_2}.$$

This is a **null method** 零点法: you find the balance (zero current) instead of measuring a current's value. Its advantage is that at balance the unknown cell gives no current, so its internal resistance does not affect the result.



A potentiometer comparing two cell e.m.f.s by the null method

Exam tips

- Apply **Kirchhoff's laws**: current into a junction = current out (charge conserved); $\sum$ e.m.f. = $\sum$ p.d. round a loop (energy conserved).
- Combine resistors: **series** $R = R_1 + R_2$; parallel $1/R = 1/R_1 + 1/R_2$.
- A **potential divider** splits voltage in the ratio of the resistances.
- Include **internal resistance**: e.m.f. $= I(R + r)$ — the "lost volts" are Ir.

Particle physics

A-Level Physics

The nuclear atom

Geiger–Marsden α-particle scattering

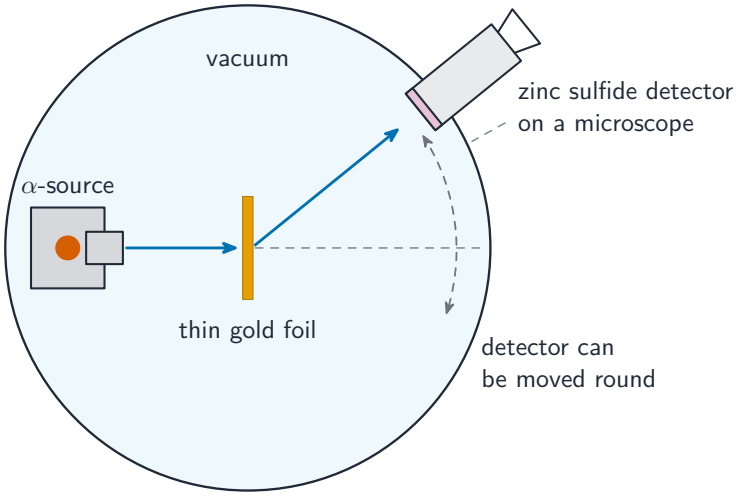
Alpha particles α 粒子 fired at a thin gold foil were seen to:

- mostly pass straight through, with very little **deflection** 偏转,
- sometimes deflect through small angles,
- rarely (about 1 in 8000) deflect through angles greater than 90°.

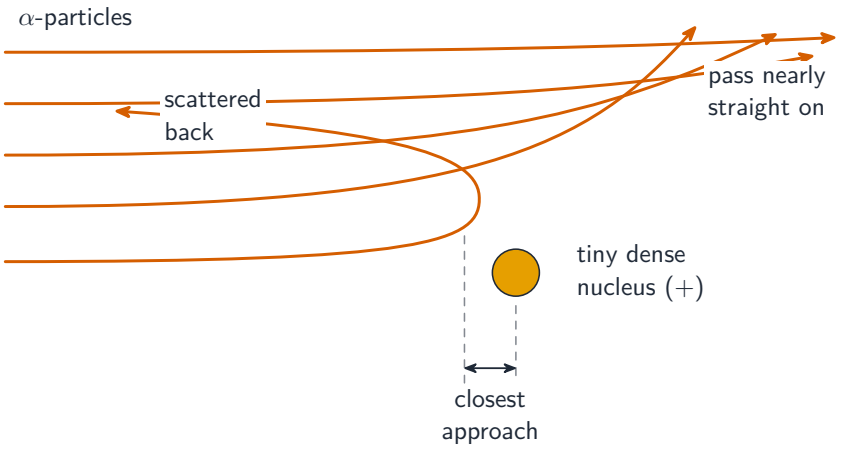
From this Rutherford worked out:

- the atom is **mostly empty space** (most α-particles pass straight through),
- there is a **tiny, dense, positively charged nucleus** 原子核 at the centre (the rare large deflections need a concentrated charge to push the α away),
- almost all of the atom's mass is in this nucleus.

Order of magnitude: atom diameter $\sim 10^{-10}$ m, nucleus diameter $\sim 10^{-15}$ m — the nucleus is about 10^5 times smaller than the atom.



The α-scattering experiment — α-particles strike a thin gold foil in a vacuum



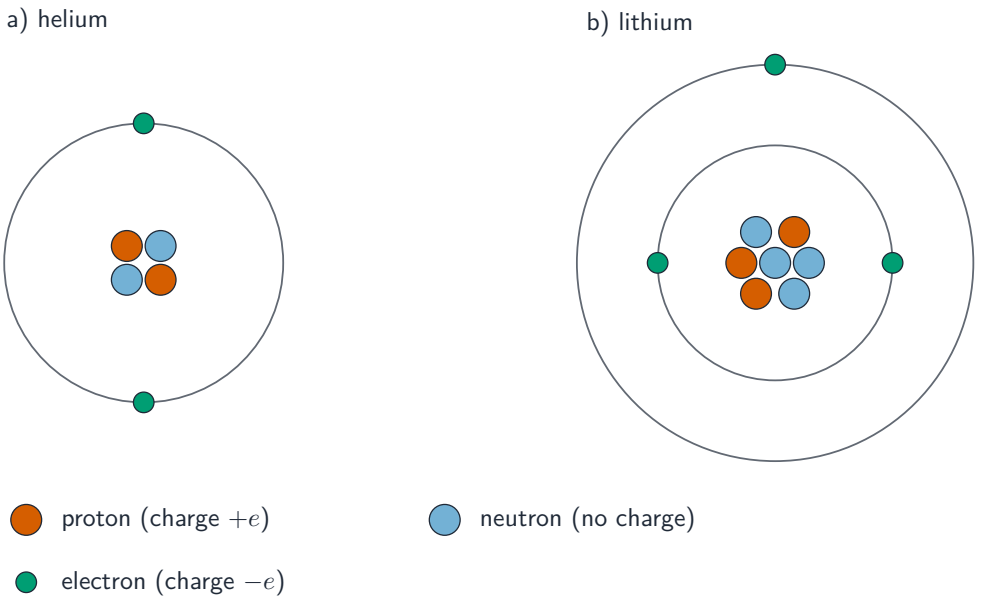
Most α-particles pass nearly straight through; a few are deflected sharply by the tiny nucleus

Simple nuclear model

An atom has:

- a central nucleus of **protons** 质子 (positive, charge $+e$) and **neutrons** 中子 (no charge),
- **electrons** 电子 (charge $-e$) around the nucleus.

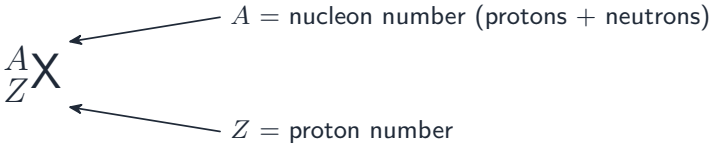
The proton and neutron have almost the same mass ($\approx$ 1 u); the electron is about $\frac{1}{1836}$ of the proton's mass.



Simple models of a helium atom and a lithium atom (not to scale)

Notation and key numbers

For a **nuclide** 核素 written ${}^{A}_{Z}\mathrm{X}$:



Nuclide notation: nucleon number on top, proton number below

- **proton number** 质子数 Z (also the atomic number): the number of protons. It fixes the element.
- **nucleon number** 核子数 A (also the mass number): the total number of **nucleons** 核子 (protons + neutrons).
- number of neutrons $N = A - Z$.

A neutral atom has the same number of electrons as protons.

Isotopes

Isotopes 同位素 are atoms of the same element (same Z) with different numbers of neutrons (different A). They behave the same chemically but differently in the nucleus. Example: ${}^{12}_{6}\mathrm{C}$ and ${}^{14}_{6}\mathrm{C}$ are isotopes of carbon.

Conservation laws in nuclear processes

In any nuclear process:

- nucleon number A is conserved (total A before = total A after),
- **charge** is conserved (this is **conservation of charge** 电荷守恒).

These two rules let you balance decay and reaction equations.

Unified atomic mass unit

The **unified atomic mass unit** 统一原子质量单位, symbol u, is set so that an atom of ${}^{12}_{6}\mathrm{C}$ has mass exactly 12 u. Numerically,

$$1\ \mathrm{u} = 1.661 \times 10^{-27}\ \mathrm{kg}.$$

A proton has mass ≈ 1.007 u; a neutron ≈ 1.009 u; an electron $\approx 5.5 \times 10^{-4}$ u.

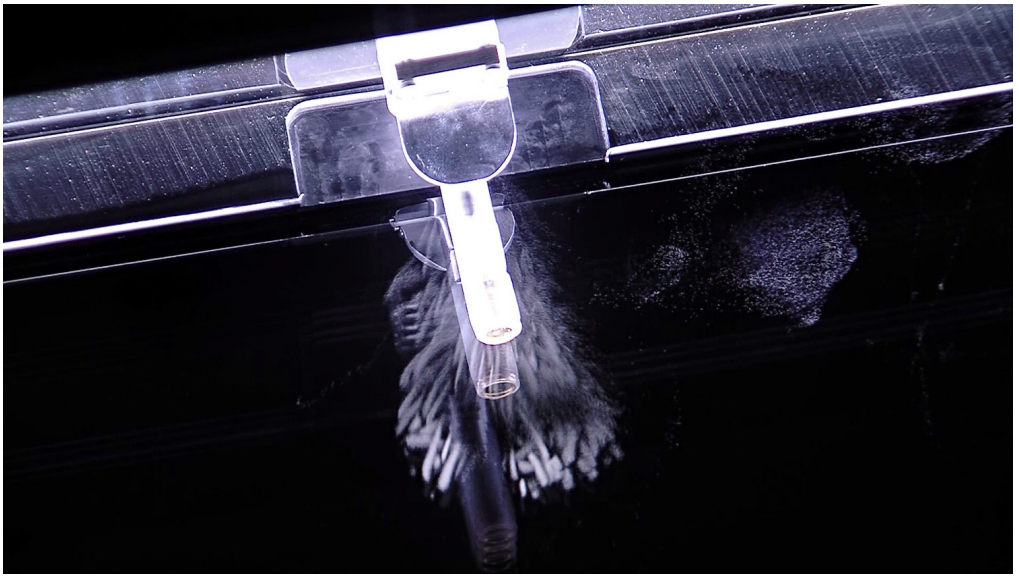
Radioactive emissions

An unstable nucleus rearranges itself and gives out one of three kinds of **radiation** 辐射. This is **radioactive** 放射性 decay. Each kind has its own properties.

α-radiation

- **Made of**: a helium-4 nucleus, ${}^{4}_{2}\alpha$ (two protons + two neutrons).
- **Mass**: ≈ 4 u.
- **Charge**: $+2e$.
- **Range in air**: a few cm. Stopped by a sheet of paper.
- **Ionising power**: strong — it is good at **ionising** 电离.
- **Energy spectrum**: **discrete** 分立 (one decay gives α-particles at one or a few sharp energies).

A **cloud chamber** 云室 makes the tracks visible: each α-particle leaves a short, straight, thick trail of tiny droplets as it ionises the air. The short equal lengths show the α-particles all carry about the same energy.



Alpha-particle tracks in a cloud chamber, fanning out from an americium-241 source

β-radiation

Two types of **beta particle** β 粒子:

- β^-: a fast electron, given out when a neutron turns into a proton.
- β^+: a **positron** 正电子 (the electron's antiparticle), given out when a proton in a proton-rich nucleus turns into a neutron.

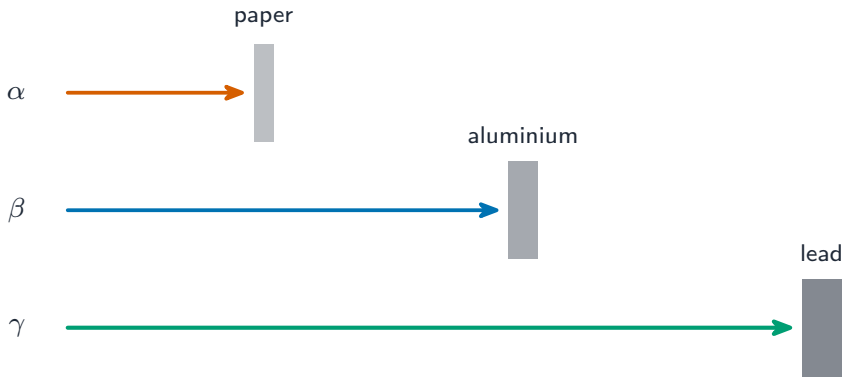
Properties (both types):

- **Mass**: $\approx 1/1836$ u (much less than α).
- **Charge**: $-e$ for β^-, $+e$ for β^+.
- **Range in air**: about 1 m. Stopped by a few mm of aluminium.
- **Energy spectrum**: **continuous** 连续 up to a maximum (see below).

γ-radiation

- **Made of**: a high-energy **photon** 光子 — part of the **electromagnetic spectrum** 电磁波谱.
- **Mass**: zero (rest mass).
- **Charge**: zero.
- **Range in air**: large (follows the inverse-square law). Strongly **attenuated** 衰减 by several cm of lead or about a metre of concrete.
- **Ionising power**: weakest.
- **Energy spectrum**: discrete (a **gamma ray** γ 射线 is given out as the nucleus drops between two nuclear energy levels).

A nucleus often gives out a γ-photon as a "tidy-up" step after an α or β decay leaves the **daughter nucleus** 子核 in an **excited state** 激发态.



Penetrating power: α is stopped by paper, β by aluminium, γ only attenuated by lead

Antiparticles, neutrinos and antineutrinos

Every particle has an **antiparticle** 反粒子 with the **same mass** but **opposite charge**. The positron is the antiparticle of the electron.

In β-decay, a third particle is always given out as well:

- β^- decay: an **antineutrino** 反中微子 $\bar{\nu}_e$.
- β^+ decay: a **neutrino** 中微子 ν_e.

Neutrinos and antineutrinos have zero charge, very small mass, and barely interact — they are very hard to detect, but they must be there to balance energy, **momentum** 动量 and other conserved quantities in β-decay.

Why β has a continuous spectrum (and α does not)

In α-decay the **energy** 能量 released is shared between just two particles (the daughter nucleus and the α). Conservation of momentum and energy then fixes the α's energy to one value (discrete).

In β-decay the energy is shared between **three** particles (the daughter nucleus, the β, and the (anti)neutrino). The β can take any share from zero up to a maximum, so its energy spectrum is continuous.

Writing decay equations

A general α-decay:

$$^{A}_{Z}\mathrm{X} \rightarrow {}^{A-4}_{Z-2}\mathrm{Y} + {}^{4}_{2}\alpha$$

(check: $A = (A-4)+4$, $Z = (Z-2)+2$.)

A general β^- decay:

$$^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z+1}\mathrm{Y} + {}^{0}_{-1}\beta + \bar{\nu}_\mathrm{e}.$$

A general β^+ decay:

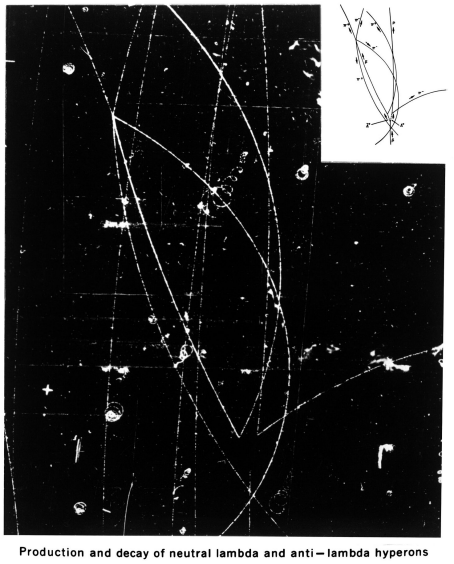
$$^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z-1}\mathrm{Y} + {}^{0}_{+1}\beta + \nu_\mathrm{e}.$$

Worked example. Uranium-238, $^{238}_{92}\mathrm{U}$, decays by α-emission; carbon-14, $^{14}_{6}\mathrm{C}$, decays by β^--emission. Find each daughter nuclide.

α-decay lowers A by 4 and Z by 2; β^--decay leaves A unchanged and raises Z by 1:

$$^{238}_{92}\mathrm{U} \rightarrow {}^{234}_{90}\mathrm{Th} + {}^{4}_{2}\alpha, \qquad {}^{14}_{6}\mathrm{C} \rightarrow {}^{14}_{7}\mathrm{N} + {}^{0}_{-1}\beta + \bar{\nu}_\mathrm{e}.$$

Fundamental particles



A bubble chamber reveals the curved tracks of charged particles.

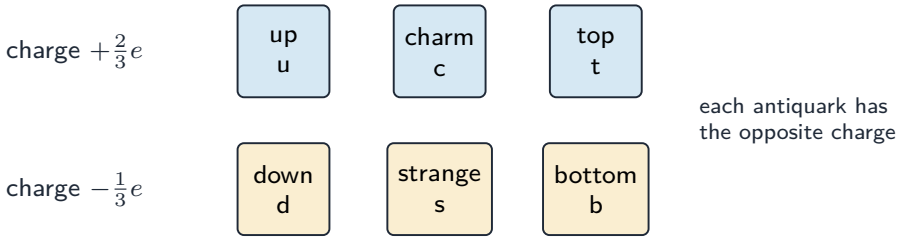
Some particles are **fundamental** 基本粒子 (point-like, with no smaller parts as far as we know); others are built from fundamental ones.

Quarks

A **quark** 夸克 is a fundamental particle. There are six **flavours** 味:

- **up** (u), charge $+\frac{2}{3}e$,
- **down** (d), charge $-\frac{1}{3}e$,
- **charm** (c), charge $+\frac{2}{3}e$,
- **strange** (s), charge $-\frac{1}{3}e$,
- **top** (t), charge $+\frac{2}{3}e$,
- **bottom** (b), charge $-\frac{1}{3}e$.

Each quark has an **antiquark** 反夸克 with the same size of charge but the opposite sign: $\bar{u}$ (charge $-\frac{2}{3}e$), $\bar{d}$ (charge $+\frac{1}{3}e$). No other quark property is tested.



The six quarks: up/charm/top carry $+\frac{2}{3}e$, down/strange/bottom carry $-\frac{1}{3}e$

Hadrons: baryons and mesons

Particles built from quarks are **hadrons** 强子. Two types:

- **baryons** 重子 — three quarks. Examples: proton (u u d), neutron (u d d). Charge check: $\frac{2}{3}+\frac{2}{3}-\frac{1}{3}=+1$ for the proton; $\frac{2}{3}-\frac{1}{3}-\frac{1}{3}=0$ for the neutron.
- **mesons** 介子 — one quark and one antiquark (for example π^+ is $u\bar{d}$).

Protons and neutrons are **not** fundamental — they are baryons made of quarks.

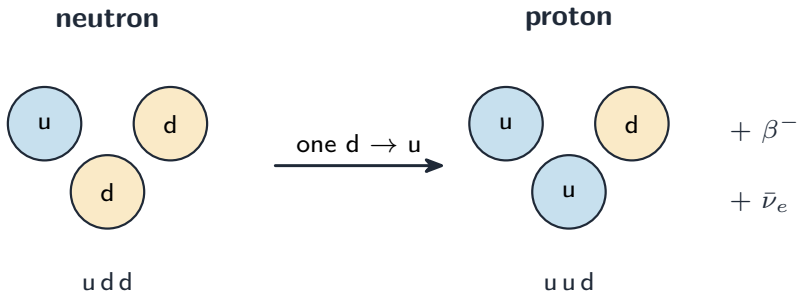
Quark changes in β-decay

In β^- decay a neutron turns into a proton; in quark terms, **one down quark turns into an up quark**:

$$\mathrm{d} \rightarrow \mathrm{u} + \beta^- + \bar{\nu}_\mathrm{e}.$$

In β^+ decay a proton turns into a neutron; **one up quark turns into a down quark**:

$$\mathrm{u} \rightarrow \mathrm{d} + \beta^+ + \nu_\mathrm{e}.$$



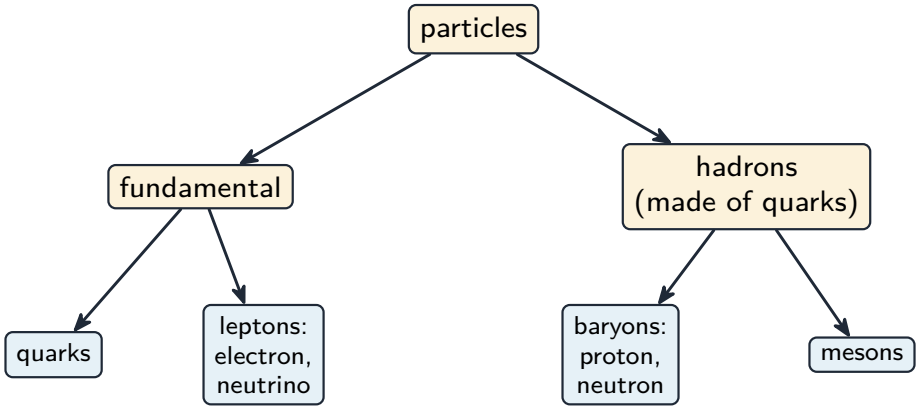
Beta-minus decay: one down quark becomes an up quark, turning a neutron into a proton

Leptons

Leptons 轻子 are fundamental particles that are not made of quarks. Electrons and neutrinos are leptons. (Heavier leptons — the muon and tau — are not needed for this syllabus.)

Classifying particles

To answer "which are fundamental?": **quarks** and **leptons** (electrons, positrons, neutrinos, antineutrinos) are fundamental; protons, neutrons, baryons, mesons and hadrons are **not** — they are built from quarks.



Fundamental particles (quarks, leptons) versus hadrons (baryons, mesons)

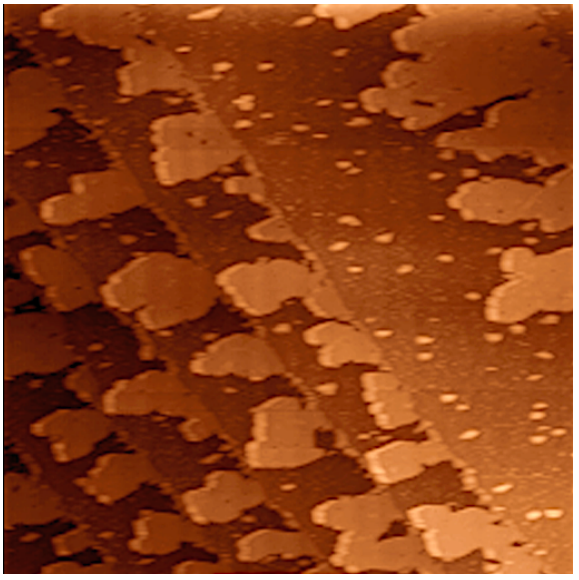
Exam tips

- Describe the **nuclear atom** (a small, dense, positive nucleus) using the alpha-scattering evidence.
- Compare α, β and γ by **charge, mass, ionising power and penetration**.
- Use quark composition (proton uud, neutron udd) and check that **charge and nucleon number balance** in an equation.
- In β^- decay a neutron becomes a proton, an electron and an antineutrino.

Atomic structure

A-Level Chemistry

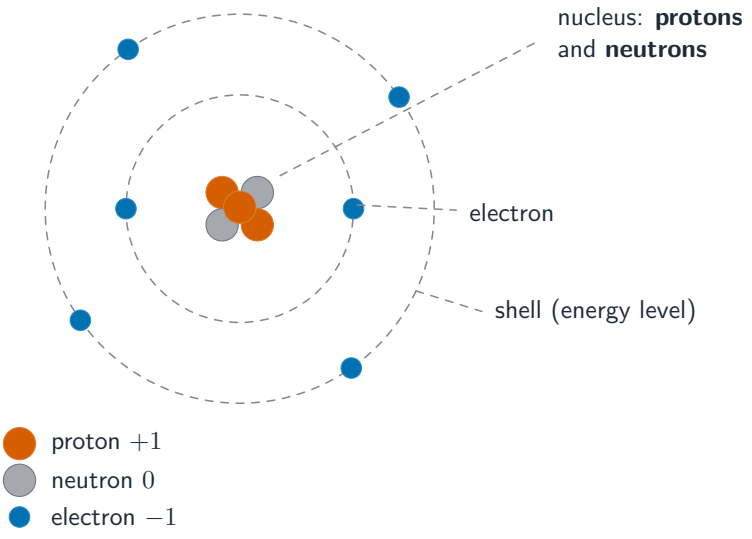
What an atom is made of



A scanning tunnelling microscope can image individual atoms.

Everything is made of **atoms** 原子. An atom is mostly empty space. At its centre is a tiny, heavy **nucleus** 原子核. The nucleus holds two kinds of particle: **protons** 质子 and **neutrons** 中子. Around the nucleus, in the empty space, move the **electrons** 电子. The electrons stay in **shells** 壳层 — layers at set distances from the nucleus.

The nucleus is very small but holds almost all the mass. The electrons take up almost all the space but have almost no mass.



An atom is mostly empty space: protons and neutrons sit in the tiny central nucleus, while electrons move in shells around it

Relative charge and relative mass

We compare the three particles using **relative charge** 相对电荷 and **relative mass** 相对质量. These are simple numbers, not real units.

Particle	Relative charge	Relative mass
proton	+1	1
neutron	0	1
electron	−1	$\frac{1}{1836}$ (about 0)

A proton and a neutron have almost the same mass. An electron is about 1836 times lighter. The proton is positive, the electron is negative, and the neutron has no charge — it is **neutral** 中性.

Proton number and nucleon number

Two numbers describe the nucleus:

- the **proton number** 质子数 (also called the **atomic number** 原子序数), symbol Z — the number of protons.
- the **nucleon number** 核子数 (also called the **mass number** 质量数), symbol A — the total number of protons and neutrons. Protons and neutrons are both **nucleons** 核子.

So the number of neutrons is $A - Z$.

Counting particles in an atom or ion

For a neutral atom, the number of electrons equals the number of protons, which equals Z.

An **ion** 离子 is an atom that has lost or gained electrons, so it has a charge:

- a positive ion has fewer electrons than protons.
- a negative ion has more electrons than protons.

Example: ${}^{27}_{13}\mathrm{Al}^{3+}$ has 13 protons, $27 - 13 = 14$ neutrons, and $13 - 3 = 10$ electrons (it lost 3 electrons to become 3+).

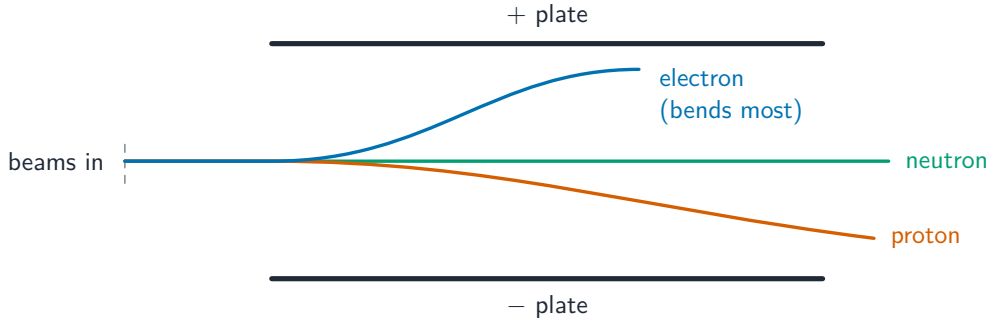
How mass and charge are spread out

Almost all the mass sits in the nucleus, because protons and neutrons are heavy and electrons are very light. All the positive charge is in the nucleus (the protons). The negative charge is spread out in the shells (the electrons).

Beams of particles in an electric field

Imagine beams of protons, neutrons and electrons moving at the same speed into an **electric field** 电场 between two charged plates:

- the proton beam bends towards the negative plate (protons are positive).
- the electron beam bends the other way, towards the positive plate. It bends much more, because the electron is far lighter — the same force gives a bigger **deflection** 偏转 to a smaller mass.
- the neutron beam goes straight through. It has no charge, so the field gives it no force.



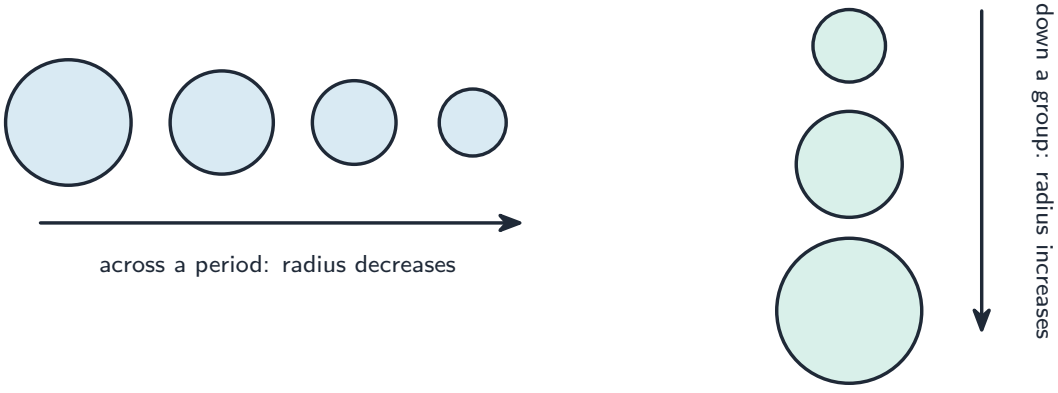
In an electric field the proton bends towards the − plate and the electron bends the opposite way and far more (it is much lighter); the neutron passes straight through

Atomic radius and ionic radius

The **atomic radius** 原子半径 is the size of an atom. The **ionic radius** 离子半径 is the size of an ion.

Across a **period** 周期 (left to right), the atomic radius gets **smaller**. The **nuclear charge** 核电荷 (the pull from the protons) rises, but the electrons go into the same outer shell, so the **shielding** 屏蔽 by inner shells stays about the same. The stronger pull draws the outer shell inwards.

Down a **group** 族 (top to bottom), the atomic radius gets **larger**. Each step down adds a new shell, so the outer electrons are further out and feel more shielding from the nucleus.



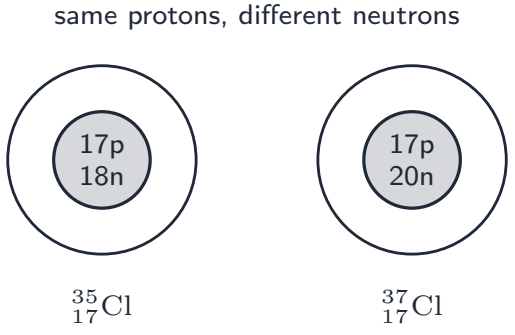
Across a period the atoms shrink (stronger nuclear pull on the same outer shell); down a group they grow (each step adds a shell)

For ions:

- a positive ion (**cation** 阳离子) is smaller than its atom. It has lost its outer shell, and the electrons that remain feel a stronger pull each.
- a negative ion (**anion** 阴离子) is larger than its atom. It has gained electrons, so there is more **repulsion** 排斥 between the electrons.
- among ions that have the same number of electrons, the one with more protons is smaller.

Isotopes

Isotopes 同位素 are atoms of the same element with the same number of protons but a different number of neutrons. So isotopes have the same proton number Z but a different nucleon number A.



Two chlorine isotopes: same protons, different neutrons

We write an isotope as ${}^{A}_{Z}\mathrm{X}$: the nucleon number A on top, the proton number Z below. For example, chlorine has two main isotopes, ${}^{35}_{17}\mathrm{Cl}$ and ${}^{37}_{17}\mathrm{Cl}$.

Same chemical properties

Chemical properties 化学性质 depend on the electrons, especially the outer electrons. Isotopes of one element have the same number of electrons arranged in the same way. So they react in exactly the same way — they have the same chemical properties.

Different physical properties

Some **physical properties** 物理性质 depend on mass, so they differ between isotopes. A heavier isotope has more neutrons, so more mass, and therefore a higher **density** 密度. (The syllabus limits this difference to mass and density.)

Electrons, energy levels and orbitals



Hydrogen emits light only at discrete, characteristic wavelengths — its line emission spectrum.

Electrons are arranged in shells, sub-shells and orbitals.

Shells and the principal quantum number

Each shell is labelled by the **principal quantum number** 主量子数 $n = 1, 2, 3, \ldots$ A larger n means a shell that is further from the nucleus and higher in energy.

Sub-shells and orbitals

Each shell is split into **sub-shells** 亚层, named s, p and d. Each sub-shell is built from **orbitals** 轨道. An orbital is a small region that can hold up to two electrons.

Sub-shell	Number of orbitals	Maximum electrons
s	1	2
p	3	6
d	5	10

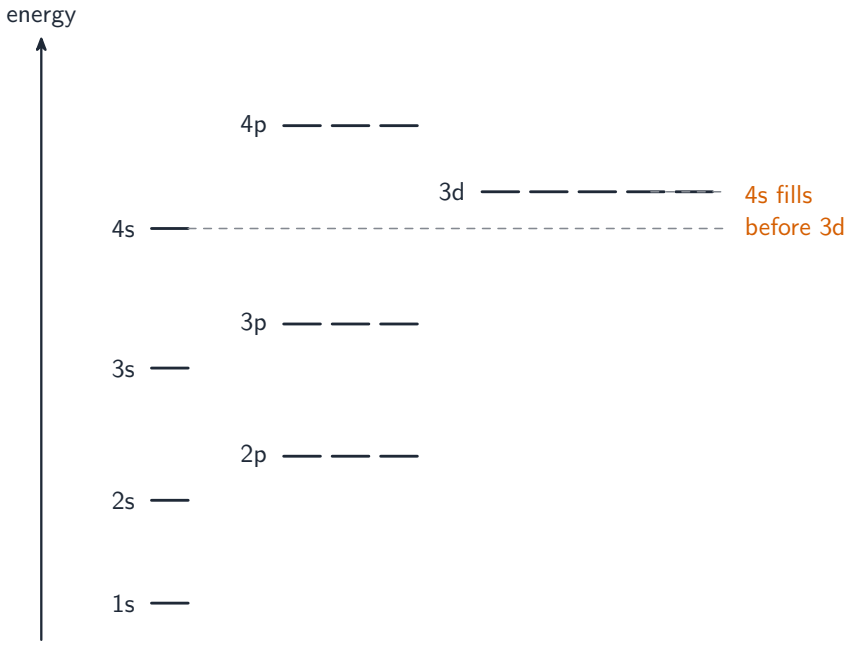
So an s sub-shell holds 2 electrons, a p sub-shell holds 6, and a d sub-shell holds 10.

Order of increasing energy

Electrons fill the lowest-energy sub-shell first. For the first three shells, plus 4s and 4p, the order of rising energy is:

$$1s < 2s < 2p < 3s < 3p < 4s < 3d < 4p$$

Notice the surprise: 4s is slightly lower in energy than 3d, so 4s fills first.



The sub-shells in order of increasing energy. 4s lies just below 3d, so 4s fills first

Electronic configuration

The **electronic configuration** 电子排布 lists how many electrons are in each sub-shell. The lowest-energy arrangement is the **ground state** 基态.

For iron (Fe, $Z = 26$):

$$1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,3d^6\,4s^2$$

You can write a shorthand using the nearest **noble gas** 稀有气体 in square brackets:

$$[\mathrm{Ar}]\,3d^6\,4s^2$$

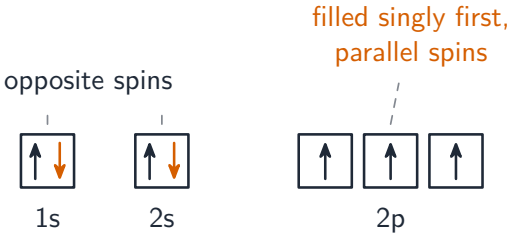
Here [Ar] stands for the full configuration of argon.

For ions, you add or remove electrons. One key rule: when a transition metal forms a positive ion, it loses its 4s electrons **before** its 3d electrons. So Fe^{3+} is $[Ar]\,3d^5$.

Electrons in boxes

The **electrons in boxes** notation draws each orbital as a box and each electron as an arrow. Two electrons in the same orbital must point opposite ways, because each electron has a property called **spin** 自旋, and a shared orbital needs opposite spins.

Within a sub-shell, electrons fill empty orbitals one at a time, with parallel arrows, before any orbital gets a second electron. Spreading out like this keeps the electrons apart and lowers the repulsion between them.



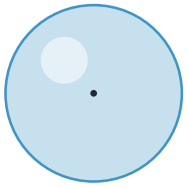
Electrons in boxes for nitrogen ($1s^2\,2s^2\,2p^3$): paired electrons point opposite ways, and the 2p orbitals fill singly with parallel spins

Why the configuration takes this shape

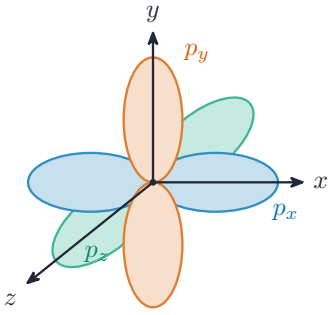
Electrons fill from low energy to high energy because that gives the most stable (lowest-energy) atom. Within a sub-shell they spread out singly first to reduce the repulsion between the negative electrons.

Shapes of s and p orbitals

- an s orbital is a **sphere** 球形 centred on the nucleus.
- a p orbital has two lobes, like a dumbbell, pointing along one axis. The three p orbitals point along three directions at right angles (the x, y and z axes).



s orbital
(sphere)



p orbitals
(two lobes each)

An s orbital is a sphere; each p orbital is a dumbbell, and the three p orbitals point along the x, y and z axes

Free radicals

A **free radical** 自由基 is a species with one or more **unpaired electrons** 未成对电子. Free radicals are very reactive.

Ionisation energy

First ionisation energy

The **first ionisation energy** 第一电离能 (IE) is the energy needed to remove one electron from each atom in one mole of gaseous atoms, forming one mole of gaseous +1 ions.

We use gaseous atoms so there are no forces between the particles. The unit is kJ mol^{-1}. As an equation, for an element X:

$$X(g) \rightarrow X^{+}(g) + e^{-}$$

The (g) shows the species is a gas.

Successive ionisation energies

After removing one electron, you can remove another. The second ionisation energy removes one electron from each +1 ion:

$$X^+(g) \rightarrow X^{2+}(g) + e^-$$

You can keep going. These are the **successive ionisation energies** 逐级电离能. Each is larger than the one before, because every electron is pulled away from a more positive ion.

What ionisation energy depends on

Ionisation energy comes from the attraction between the positive nucleus and the outer electron. Three main factors set how strong that attraction is:

- **nuclear charge**: more protons pull the electrons more strongly, so the ionisation energy is higher.
- **atomic radius**: the further the outer electron sits from the nucleus, the weaker the pull, so the ionisation energy is lower.
- **shielding**: inner shells block some of the pull on the outer electron. More inner shells mean more shielding and a lower ionisation energy.

There is a smaller effect too — **spin-pair repulsion** 自旋成对排斥. When two electrons share one orbital, they push each other a little, so one is easier to remove.

Trends in first ionisation energy

Across a period, the first ionisation energy generally **rises**. The nuclear charge grows while shielding stays about the same, so the outer electrons are held more tightly.

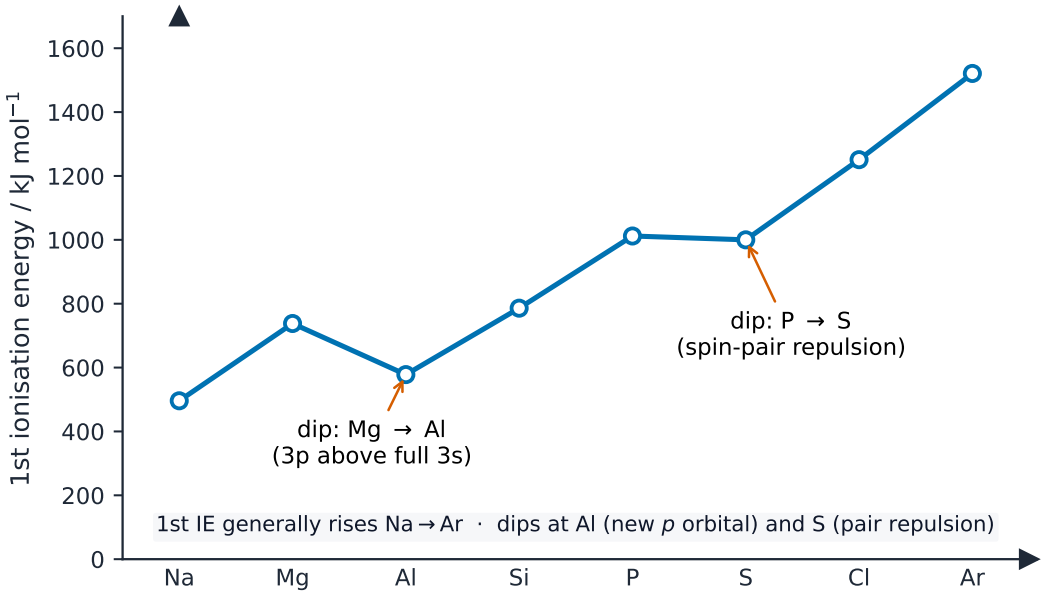
Down a group, the first ionisation energy **falls**. Lower elements have more shells, so more shielding and a larger radius, and the outer electron is easier to remove.

The dips are evidence for sub-shells

The rise across a period is not smooth. Two small dips appear, and you should be able to explain both:

- **Group 2 to Group 13** (for example Mg to Al): the electron removed from Al comes from a 3p sub-shell, which is higher in energy than the full 3s sub-shell in Mg. A 3p electron is easier to remove, so the value dips.
- **Group 15 to Group 16** (for example P to S): in S, one 3p orbital now holds a pair of electrons. Spin-pair repulsion makes one of them easier to remove, so the value dips.

These dips are evidence that sub-shells exist.

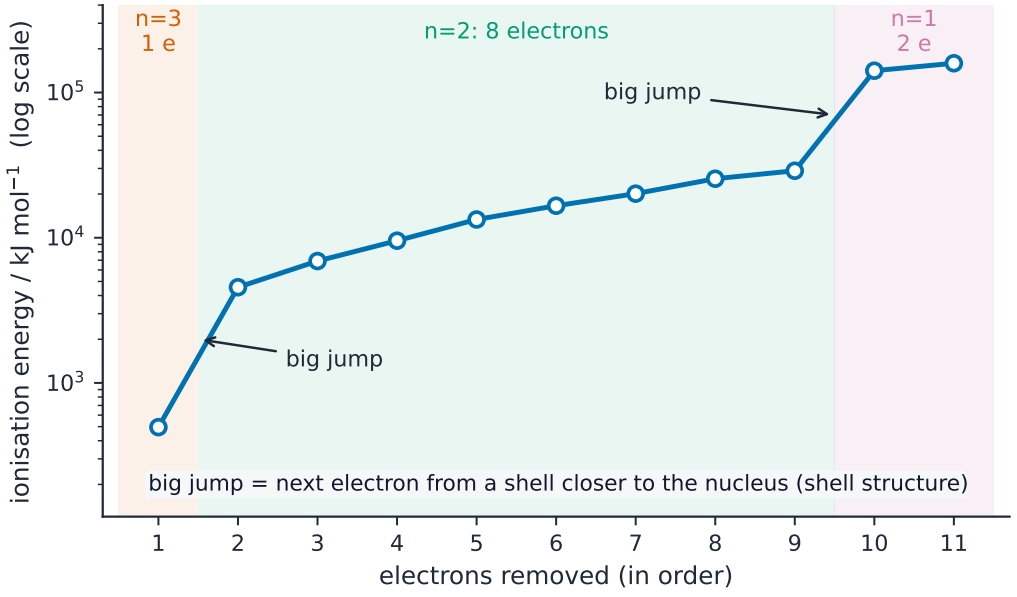


First ionisation energy rises across Period 3 but dips at Al and at S — evidence that sub-shells exist

Successive ionisation energies are evidence for shells

If you plot the successive ionisation energies of one element, the values rise, with **big jumps** at certain points. A big jump happens when the next electron must come from a shell closer to the nucleus.

Count how many electrons come off easily before the first big jump — that is the number of electrons in the outer shell, which tells you the **group** the element is in. You can also use the pattern to work out the electronic configuration and the position of the element in the Periodic Table.



Successive ionisation energies of sodium (log scale): the big jumps reveal the $2, 8, 1$ *shell structure*

Electrons removed before the first big jump	Group
1	Group 1
2	Group 2
3	Group 13

Worked example. The first five successive ionisation energies of an element are 590, 1150, 4940, 6480 and 8120 kJ mol^{-1}. Which group is it in? Look for the **big jump**, not the biggest number. From the 1st to the 2nd the value roughly doubles, which is a normal rise. From the 2nd (1150) to the 3rd (4940) it more than quadruples: that is the jump. So **two** electrons come off easily before it, the outer shell holds 2 electrons, and the element is in **Group 2**. Count the electrons removed **before** the jump, and explain the jump properly: the next electron is being pulled from a shell **closer to the nucleus**, not from a different element.

Exam tips

- Define **isotopes** in full: atoms of the same element with the same number of protons but a different number of neutrons — the mark scheme wants both halves.
- 4s fills **before** 3d, but electrons are **removed from 4s first** when forming ions, so Fe^{3+} is $[Ar]3d^5$ (not $[Ar]3d^3 4s^2$).
- In successive ionisation energies, a **big jump** marks the start of a new (inner) shell — use the jumps to place the element in its group.
- Explain every ionisation-energy trend with the same three factors: **nuclear charge, distance and shielding**, plus sub-shell effects for the small dips.
- Learn the exact reason first ionisation energy of oxygen is **below** nitrogen: oxygen's **paired 2p electrons repel**, so one is easier to remove.

Atoms, molecules and stoichiometry

A-Level Chemistry

Relative masses of atoms and molecules

Atoms 原子 are far too light to weigh in grams, so we compare every mass to one standard. The standard is the **unified atomic mass unit** 统一原子质量单位 (symbol u), defined as exactly one twelfth of the mass of one carbon-12 atom.

Using this unit, we state masses as simple numbers:

- the **relative atomic mass** 相对原子质量 A_r of an element is the average mass of its atoms compared with $\frac{1}{12}$ of a carbon-12 atom. It is an average over all the **isotopes** 同位素, weighted by how common each one is.
- the **relative isotopic mass** 相对同位素质量 is the mass of one atom of a single isotope, compared with $\frac{1}{12}$ of a carbon-12 atom.
- the **relative molecular mass** 相对分子质量 M_r of a **molecule** 分子 is the sum of the relative atomic masses of all its atoms.
- the **relative formula mass** 相对式量 is the same idea for a substance that is not made of molecules (such as an ionic compound). Add up the relative atomic masses shown in the formula.



$$A_r = \frac{(75 \times 35) + (25 \times 37)}{100} = 35.5$$

Relative atomic mass is a weighted average: chlorine's two isotopes (75% ^{35}Cl, 25% ^{37}Cl) average to $A_r = 35.5$

To find A_r from isotope data, multiply each isotope mass by its percentage, add these up, and divide by 100.

Worked example. Chlorine is 75% ^{35}Cl and 25% ^{37}Cl. Find its relative atomic mass.

$$A_r = \frac{(75 \times 35) + (25 \times 37)}{100} = \frac{2625 + 925}{100} = 35.5.$$

Worked example. A mass spectrometer shows copper is 69.2% ^{63}Cu and 30.8% ^{65}Cu. Find its relative atomic mass.

$$A_r = \frac{(69.2 \times 63) + (30.8 \times 65)}{100} = \frac{4359.6 + 2002.0}{100} = 63.6.$$

The mole and the Avogadro constant



A modern electronic balance measures mass — the basis of mole calculations.

Chemists count particles in groups called moles, just as we count eggs in dozens.

One **mole** 摩尔 (symbol mol) is the amount of substance that contains the same number of particles as there are atoms in exactly 12 g of carbon-12. That number is the **Avogadro constant** 阿伏伽德罗常量:

$$N_A = 6.02 \times 10^{23}\ \text{mol}^{-1}$$

So one mole of anything contains 6.02×10^{23} particles. The particles may be atoms, molecules or **ions** 离子 — always say which.

The mass of one mole in grams equals the relative mass (A_r or M_r). This is the **molar mass** 摩尔质量, with units g mol^{-1}. The key equation is:

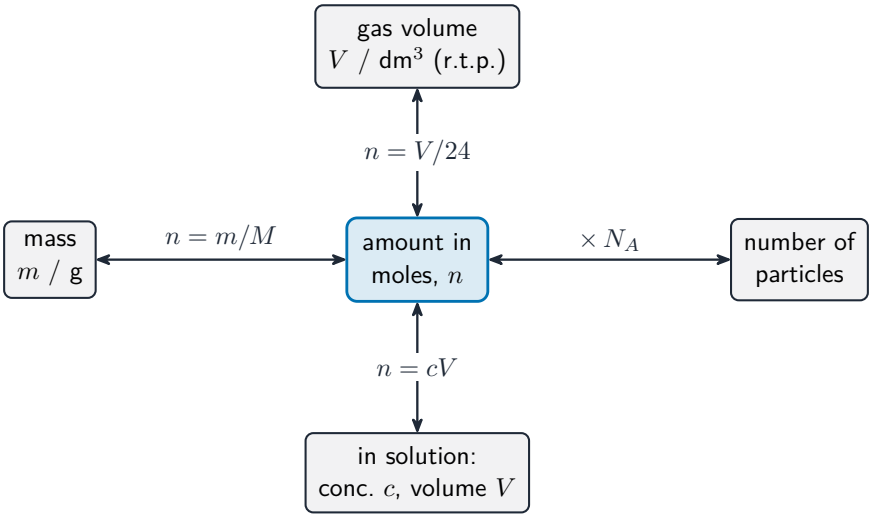
$$n = \frac{m}{M}$$

where n is the amount in moles, m is the mass in grams, and M is the molar mass.

Worked example. How many moles are in 8.0 g of methane, CH_4? (A_r: C = 12, H = 1.)

The molar mass is $M = 12 + 4(1) = 16$ g mol^{-1}, so

$$n = \frac{m}{M} = \frac{8.0}{16} = 0.50 \text{ mol}.$$



The mole is the hub of every amount calculation: convert to mass ($n = m/M$), particles ($\times N_A$), gas volume ($n = V/24$) or solution ($n = cV$)

Formulas

A **compound** 化合物 is a substance made of two or more elements chemically joined.

Charges and formulas of ionic compounds

In an **ionic compound** 离子化合物 the total positive charge balances the total negative charge, so the compound is neutral overall.

You can predict the charge of many ions from the element's position in the Periodic Table:

Group	**1**	**2**	**13**	**15**	**16**	**17**
Usual ion charge	+1	+2	+3	−3	−2	−1

Hydrogen forms H^+, and the Group 18 noble gases do not normally form ions.

Some ions you must know by name and formula:

Name	Formula
nitrate	NO_3^-
carbonate	CO_3^{2-}
sulfate	SO_4^{2-}
hydroxide	OH^-
ammonium	NH_4^+
zinc	Zn^{2+}
silver	Ag^+
hydrogencarbonate	HCO_3^-
phosphate	PO_4^{3-}

For a metal that can have more than one charge, a Roman numeral shows the **oxidation number** 氧化数. For example, iron(II) is Fe^{2+} and iron(III) is Fe^{3+}. To write a formula, balance the charges: iron(III) oxide is Fe_2O_3, because two Fe^{3+} balance three O^{2-}.

Equations and state symbols

A chemical equation must be **balanced** 配平 — the same number of each kind of atom on both sides. Add **state symbols** 状态符号 to show the state of each species: (s) solid, (l) liquid, (g) gas, and (aq) **aqueous** 水溶液 (dissolved in water).

An **ionic equation** 离子方程式 shows only the ions and molecules that actually change. The ions that do not change are **spectator ions** 旁观离子, and you leave them out. For example, the reaction that forms silver chloride is:

$$Ag^+(aq) + Cl^-(aq) \rightarrow AgCl(s)$$

Empirical and molecular formulas

The **empirical formula** 实验式 is the simplest whole-number ratio of the atoms of each element in a compound. The **molecular formula** 分子式 shows the actual number of atoms of each element in one molecule.

For example, ethane has empirical formula CH_3 but molecular formula C_2H_6.

Hydrated and anhydrous solids

Some solids hold water inside their crystals. This water is the **water of crystallisation** 结晶水. A solid that contains it is **hydrated** 水合的; the same solid with the water removed is **anhydrous** 无水的.

For example, hydrated copper(II) sulfate is $CuSO_4 \cdot 5H_2O$. Heating it drives off the water to leave anhydrous $CuSO_4$.

Calculating empirical and molecular formulas

To find the empirical formula from masses (or percentages by mass):

1. divide each element's mass by its A_r to get the moles.
2. divide all the mole values by the smallest one.
3. round to the nearest whole numbers — that ratio is the empirical formula.

To get the molecular formula, you also need M_r. Find how many times the empirical formula mass fits into M_r, then multiply the formula by that number.

Worked example. A compound is 40.0% carbon, 6.7% hydrogen and 53.3% oxygen by mass. Find its empirical formula. (A_r: C = 12, H = 1, O = 16.)

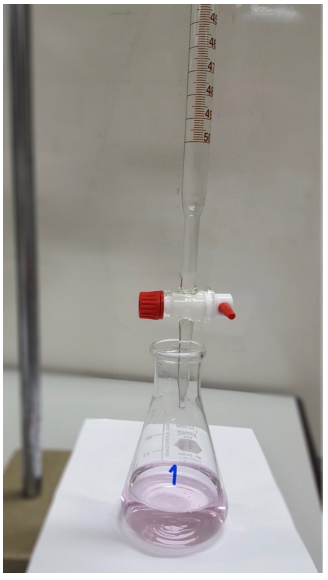
Take 100 g and divide each mass by its A_r to get moles: C = 40.0/12 = 3.33, H = 6.7/1 = 6.7, O = 53.3/16 = 3.33. Dividing through by the smallest (3.33) gives a ratio C : H : O = 1 : 2 : 1, so the empirical formula is CH_2O.

element	% by mass	$\div A_r \rightarrow$ moles	$\div$ smallest (3.33)	ratio
C	40.0	$40.0/12 = 3.33$	1.0	1
H	6.7	$6.7/1 = 6.7$	2.0	2
O	53.3	$53.3/16 = 3.33$	1.0	1

empirical formula CH_2O - if $M_r = 180$, it fits $180/30 = 6$ times: molecular formula $C_6H_{12}O_6$

The empirical-formula recipe: divide by Ar, divide by the smallest, read off the ratio

Reacting masses and volumes



A titration finds reacting volumes precisely.

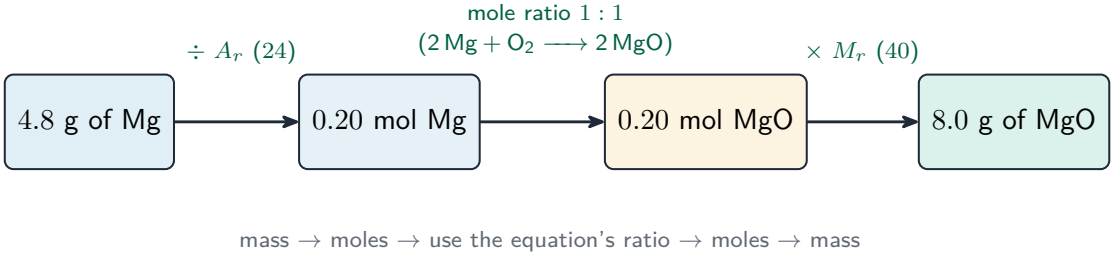
Reacting masses and percentage yield

The numbers in front of each species in a balanced equation give the mole ratio — this is the **stoichiometry** 化学计量. To find a reacting mass: change the known mass to moles, use the mole ratio to find the moles you want, then change back to mass.

Worked example. What mass of magnesium oxide forms when 4.8 g of magnesium burns completely? $2\mathrm{Mg} + \mathrm{O_2} \rightarrow 2\mathrm{MgO}$. ($A_r$: Mg = 24, O = 16.)

Moles of Mg = $4.8/24 = 0.20$ mol. The ratio Mg : MgO is 1 : 1, so 0.20 mol of MgO forms. Its molar mass is $24 + 16 = 40$ g mol^{-1}, so

$$m = nM = 0.20 \times 40 = 8.0 \text{ g}.$$



The mole bridge: mass to moles, ratio, then back to mass

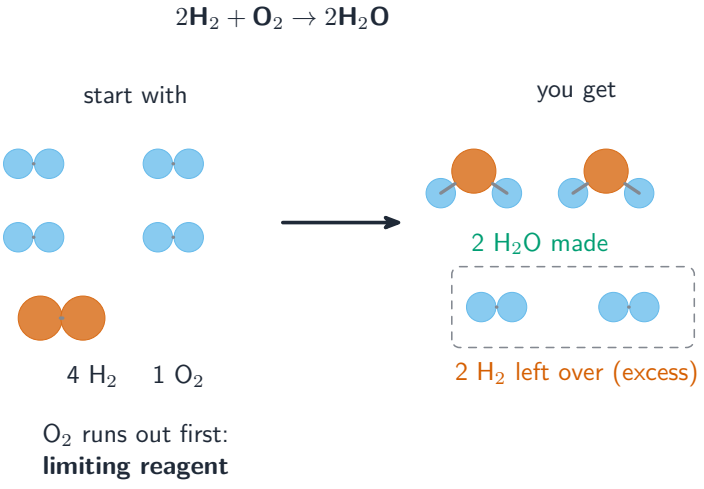
In real reactions you usually get less product than the maximum. The **percentage yield** 产率 compares the amount you actually made with the most you could make:

$$\text{percentage yield} = \frac{\text{actual amount of product}}{\text{maximum possible amount}} \times 100\%$$

Limiting and excess reagent

When two reactants are mixed, one usually runs out first. The **limiting reagent** 限量试剂 is the one that runs out — it decides how much product forms. The other is the **excess reagent** 过量试剂, because there is more than enough of it. Always base the calculation on the limiting reagent.

To find it: work out the moles of each reactant, divide each by its number in the equation, and the smallest result is the limiting reagent.



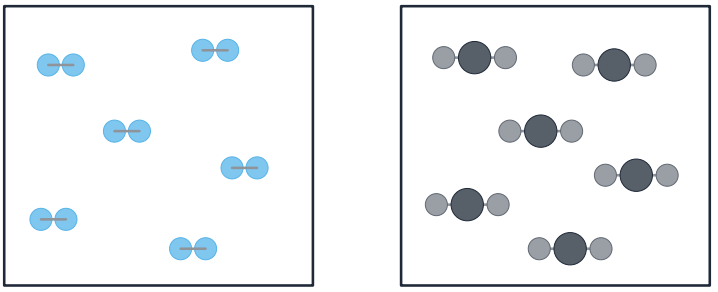
The limiting reagent runs out first and decides how much product forms; the leftover reactant is in excess

Volumes of gases

At the same temperature and pressure, equal volumes of any gases contain equal numbers of molecules. At room temperature and pressure (r.t.p.), one mole of any gas takes up 24.0 dm^3, so:

$$n = \frac{V}{24.0} \qquad (V \text{ in } \text{dm}^3 \text{ at r.t.p.})$$

same volume, same temperature and pressure



hydrogen, H_2 carbon dioxide, CO_2

the same **number** of molecules (6 each), even though CO_2 is bigger

Equal volumes of gases at the same temperature and pressure hold equal numbers of molecules, whatever the gas

This is used when burning **hydrocarbons** 碳氢化合物 (compounds of only carbon and hydrogen), where you compare gas volumes.

Volumes and concentrations of solutions

The **concentration** 浓度 of a solution is the amount of **solute** 溶质 in each cubic decimetre of **solution** 溶液, measured in mol dm^{-3}:

$$n = c \times V$$

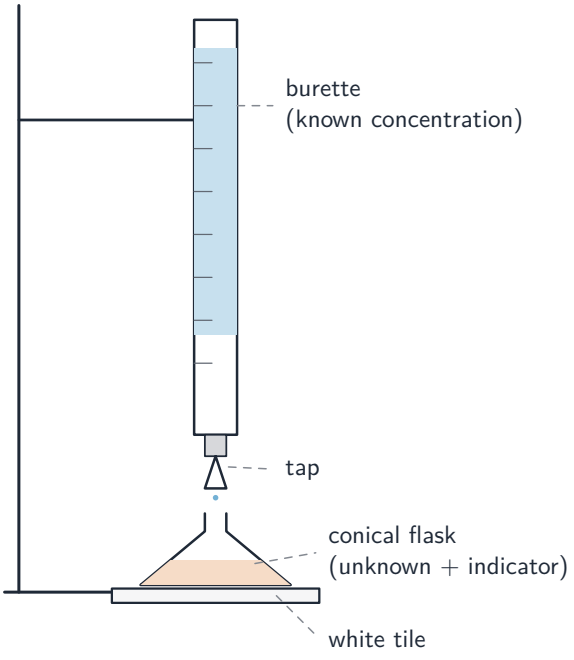
where c is the concentration and V is the volume in dm^3. Remember that 1000 cm^3 = 1 dm^3.

Worked example. In a titration, 25.0 cm^3 of sodium hydroxide solution is exactly neutralised by 20.0 cm^3 of 0.100 mol dm^{-3} hydrochloric acid: $NaOH + HCl \rightarrow NaCl + H_2O$. Find the concentration of the sodium hydroxide.

Moles of HCl $= cV = 0.100 \times \dfrac{20.0}{1000} = 2.00 \times 10^{-3}$ mol. The ratio is 1 : 1, so there are 2.00×10^{-3} mol of NaOH in 25.0 cm^3:

$$c = \frac{n}{V} = \frac{2.00 \times 10^{-3}}{25.0/1000} = 0.0800 \text{ mol dm}^{-3}.$$

This is the basis of a **titration** 滴定, where you find an unknown concentration by reacting it with a solution whose concentration you already know.



In a titration a burette adds a solution of known concentration to the unknown in the conical flask, until the indicator changes

Significant figures

Give your answer to a sensible number of **significant figures** 有效数字 — usually match the data in the question. Do not write more digits than the data supports, and do not round so early that you lose accuracy.

Exam tips

- Convert volumes to dm^3 **before** using concentration ($25.0\ cm^3 = 0.0250\ dm^3$); the missing $\div 1000$ is the most common titration error.
- Use the **balancing numbers** as the mole ratio between species — never the M_r values.
- Empirical formula: divide each mass/percentage by A_r, then by the **smallest**, then scale to whole numbers; use M_r to reach the molecular formula.
- For gases at r.t.p. use volume = moles $\times$ 24 dm^3; quote the equation you use every time.
- Give the answer to the same **significant figures** as the data (usually 3) and always include units.

Chemical bonding

A-Level Chemistry

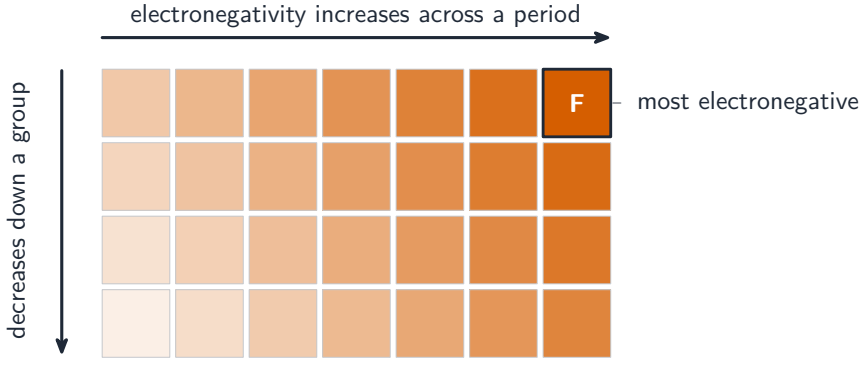
Electronegativity

Electronegativity 电负性 is the power of an atom to attract the **electrons** 电子 in a bond towards itself.

Three factors decide how electronegative an atom is:

- **nuclear charge** 核电荷: more protons pull the bonding electrons more strongly.
- **atomic radius** 原子半径: the closer the bond is to the nucleus, the stronger the pull.
- **shielding** 屏蔽 by inner shells and sub-shells: more inner electrons weaken the pull on the bonding electrons.

So electronegativity **rises** across a period (more nuclear charge, smaller radius) and **falls** down a group (larger radius, more shielding). Fluorine is the most electronegative element.



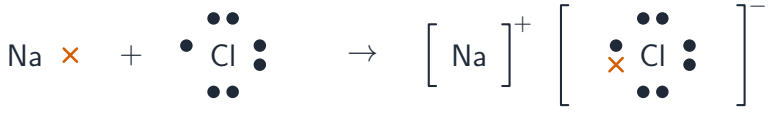
Electronegativity rises across a period and falls down a group, so fluorine is the most electronegative element

You can use the difference in **Pauling electronegativity** 鲍林电负性 values to predict the bond type. A large difference gives an **ionic** bond; a small difference gives a **covalent** bond.

Ionic bonding

Ionic bonding 离子键 is the **electrostatic attraction** 静电引力 between oppositely charged **ions** 离子 — positive **cations** 阳离子 and negative **anions** 阴离子.

It forms when a metal gives electrons to a non-metal. Good examples are sodium chloride (NaCl), magnesium oxide (MgO) and calcium fluoride (CaF_2). The ions pack into a regular giant **lattice** 晶格, held together by the attraction in every direction.



sodium gives its outer electron (×) to chlorine

Ionic bonding in NaCl: sodium transfers its single outer electron to chlorine, giving Na^+ and a full-octet Cl^-

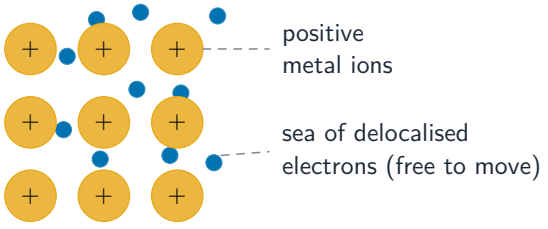


A real crystal of rock salt (halite, NaCl); the cubic shapes mirror the giant ionic lattice inside

Metallic bonding

Metallic bonding 金属键 is the electrostatic attraction between positive metal ions and a "sea" of **delocalised electrons** 离域电子.

The outer electrons are free to move through the whole metal. This explains why metals conduct electricity and are strong.



Metallic bonding: positive metal ions sit in a sea of delocalised electrons that are free to move

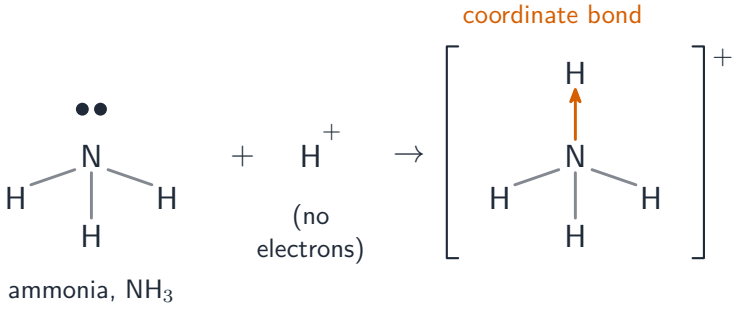
Covalent and coordinate bonding

Covalent bonding 共价键 is the electrostatic attraction between the nuclei of two atoms and a shared pair of electrons.

Simple molecules with covalent bonds include H_2, O_2, N_2, Cl_2, HCl, CO_2, NH_3, CH_4, C_2H_6 and C_2H_4. A double bond shares two pairs; a triple bond (as in N_2) shares three pairs.

Atoms in Period 3 and below can **expand the octet** 扩展八隅体 — hold more than eight electrons in their outer shell. Examples are SO_2, PCl_5 and SF_6.

A **coordinate bond** 配位键 (also called a dative covalent bond) is a covalent bond where both shared electrons come from the same atom. For example, when ammonia and hydrogen chloride gases meet, the lone pair on the nitrogen forms a coordinate bond to H^+, making the ammonium ion NH_4^+. Coordinate bonds also join the two halves of the Al_2Cl_6 molecule.



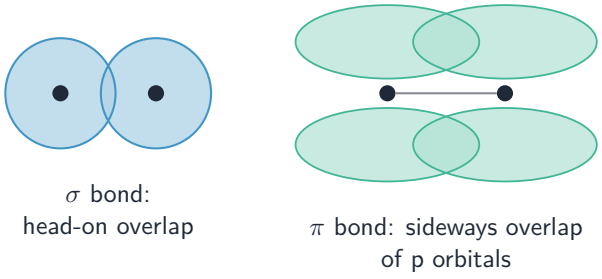
A coordinate (dative) bond: nitrogen's lone pair forms the fourth N–H bond, both electrons coming from N

Sigma and pi bonds

Covalent bonds form when **orbitals** 轨道 overlap:

- a **sigma bond** σ 键 forms by the direct, head-on **overlap** 重叠 of orbitals between the two atoms.
- a **pi bond** π 键 forms by the sideways overlap of two p orbitals, above and below the sigma bond.

A single bond is one sigma bond. A double bond (as in C_2H_4) is one sigma plus one pi bond. A triple bond (as in N_2 and HCN) is one sigma plus two pi bonds.



A σ bond forms by direct head-on overlap; a π bond forms by the sideways overlap of two p orbitals, above and below

Hybridisation

Hybridisation 杂化 mixes orbitals in the same shell to make new, equal orbitals for bonding:

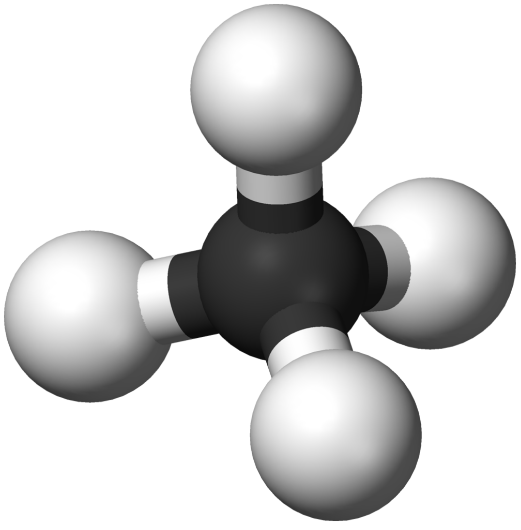
- sp: two equal orbitals, used in a linear molecule.
- sp^2: three equal orbitals, used in a flat molecule like C_2H_4.
- sp^3: four equal orbitals, used in CH_4.

Bond energy and bond length

- **bond energy** 键能 is the energy needed to break one mole of a particular covalent bond in the gas state.
- **bond length** 键长 is the distance between the centres of the two bonded atoms.

A shorter bond is usually stronger (higher bond energy). Triple bonds are shorter and stronger than double bonds, which are shorter and stronger than single bonds. Stronger bonds make a molecule harder to react.

Shapes of molecules



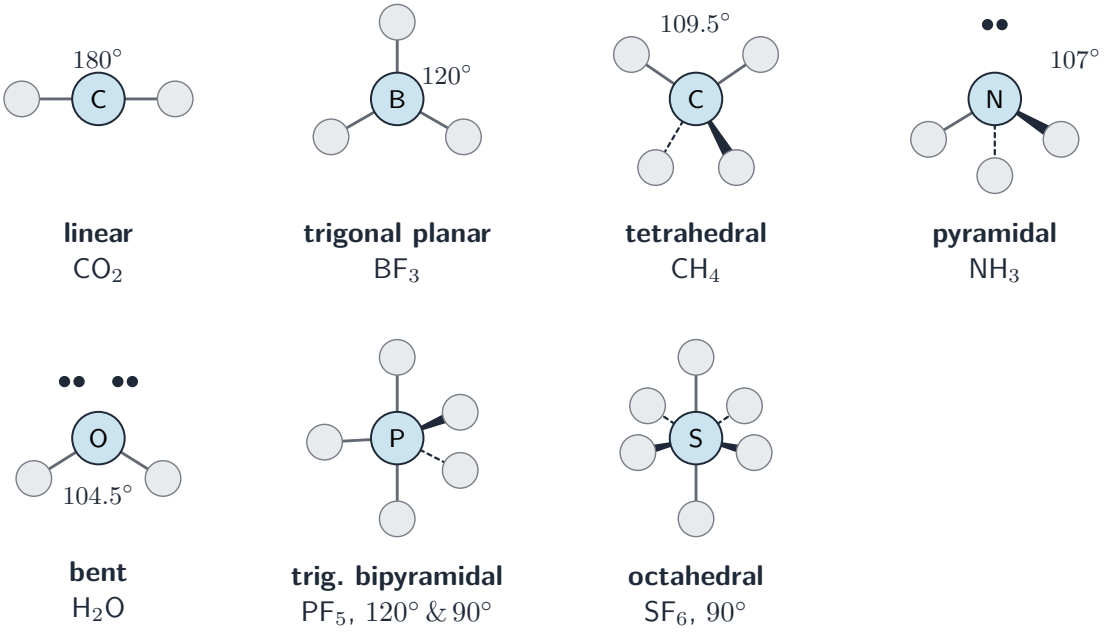
A molecular model shows the three-dimensional shape of a covalent molecule.

To work out a shape, use **VSEPR theory** 价层电子对互斥理论: the pairs of electrons around the central atom push apart as far as possible, because like charges repel.

A **lone pair** 孤对电子 (not in a bond) pushes more strongly than a **bonding pair** 成键电子对. Each lone pair squeezes the **bond angle** 键角 by about 2.5°.

Molecule	Shape	Bond angle
CO_2	**linear** 直线形	180°
BF_3	**trigonal planar** 平面三角形	120°
CH_4	**tetrahedral** 四面体形	109.5°
NH_3	**pyramidal** 三角锥形	107°
H_2O	**bent** 角形	104.5°
PF_5	**trigonal bipyramidal** 三角双锥形	120° and 90°
SF_6	**octahedral** 八面体形	90°

NH_3 has one lone pair and H_2O has two, which is why their angles drop below the 109.5° of CH_4. You can predict the shapes of similar molecules and ions in the same way.



The seven shapes from VSEPR theory; the lone pairs on NH_3 and H_2O push harder, squeezing the bond angle below 109.5°

Worked example. Predict the shape and bond angle of NH_3 and of H_2O. Nitrogen has 5 outer electrons and forms 3 bonds, leaving **3 bonding pairs and 1 lone pair** - four pairs in total, so they start from the tetrahedral 109.5°. The lone pair repels more strongly and squeezes the angle by about 2.5°, giving a **pyramidal** shape at about 107°. Oxygen forms 2 bonds and keeps **2 lone pairs**: still four pairs, but now two squeezes, so the shape is **bent** at about 104.5°. Count **all** the pairs to fix the basic geometry, subtract 2.5° for **each** lone pair, and name the shape from the **atoms** - NH_3 has four pairs but is pyramidal, not tetrahedral.

Intermolecular forces

Intermolecular forces 分子间作用力 are the forces *between* **molecules** 分子. They are much weaker than the ionic, covalent and metallic bonding *inside* substances.

Bond polarity and dipoles

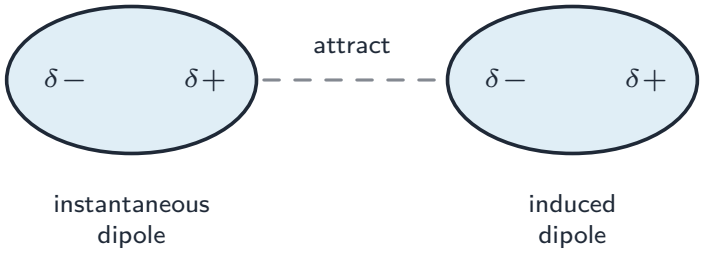
When two atoms with different electronegativity share a bond, the electrons sit closer to the more electronegative atom. The bond then has a **polarity** 极性: one end is slightly negative ($\delta-$) and the other slightly positive ($\delta+$). This separation of charge is a **dipole** 偶极.

If the dipoles in a molecule do not cancel, the whole molecule has a **dipole moment** 偶极矩 and is polar. If they cancel by symmetry (as in CO_2), the molecule is non-polar.

Van der Waals' forces

Van der Waals' forces 范德华力 is the general name for all intermolecular forces. There are two main types.

The first type is the instantaneous dipole–induced dipole force, also called the **London dispersion force** 伦敦色散力. Moving electrons make a brief **instantaneous dipole** 瞬时偶极, which then creates a matching **induced dipole** 诱导偶极 in a nearby molecule. These forces act between all molecules and get stronger when there are more electrons.

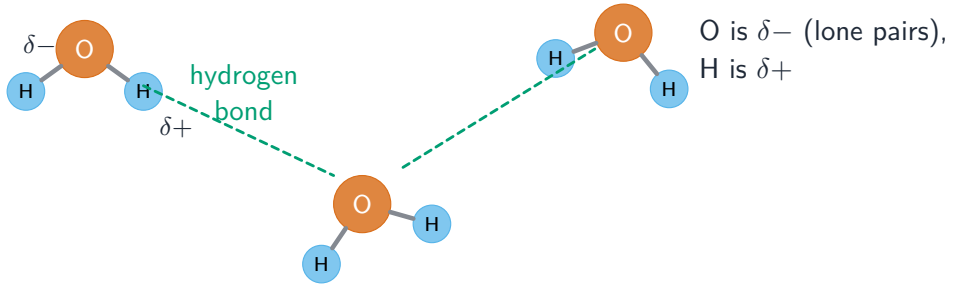


A London force: a momentary dipole in one molecule induces a dipole in its neighbour, so they attract — this acts between all molecules

The second type is the permanent dipole–permanent dipole force. It acts between molecules that are always polar, because each one has a **permanent dipole** 永久偶极.

Hydrogen bonding

Hydrogen bonding 氢键 is a strong, special case of permanent dipole forces. It forms when hydrogen is bonded to a very electronegative atom — nitrogen, oxygen or fluorine — and is attracted to a lone pair on an N, O or F atom in a neighbour. Look for N–H and O–H groups, as in ammonia and water.



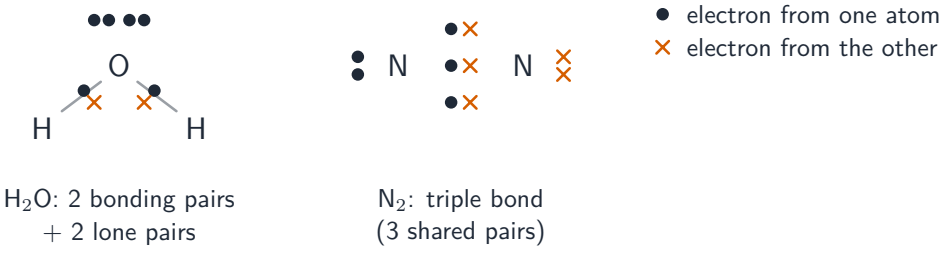
Hydrogen bonding in water: a δ+ hydrogen is attracted to a lone pair on the δ− oxygen of a neighbouring molecule

Hydrogen bonding explains the strange behaviour of water:

- its high melting and **boiling point** 沸点, because many hydrogen bonds must be broken.
- its high **surface tension** 表面张力.
- ice is **less dense** than liquid water, because hydrogen bonds hold the molecules in an open, spread-out structure, so ice floats.

Dot-and-cross diagrams

A **dot-and-cross diagram** 点叉图 shows the outer electrons of each atom, using dots for one atom and crosses for the other. This makes it clear where each bonding electron came from. You can draw them for ionic, covalent and coordinate bonding, including molecules with an expanded octet or an odd number of electrons.



Covalent dot-and-cross: each shared pair is one electron from each atom. Water has two bonding pairs and two lone pairs; nitrogen shares three pairs (a triple bond)

Exam tips

- For **shapes**, count bonding pairs and lone pairs, name the shape, then give the exact bond angle (e.g. NH_3: pyramidal, 107°) — each lone pair lowers the angle by about 2.5°.
- A **dative (coordinate) bond** has both electrons from one atom (e.g. NH_4^+, H_3O^+); draw the arrow from the lone pair.
- Name the intermolecular force precisely: **hydrogen bonding** needs H bonded to N, O or F; otherwise it is permanent-dipole or induced-dipole (van der Waals). Never call van der Waals forces "bonds".
- Explain a physical property by stating **which forces are broken**, not just "strong bonds".

States of matter

A-Level Chemistry

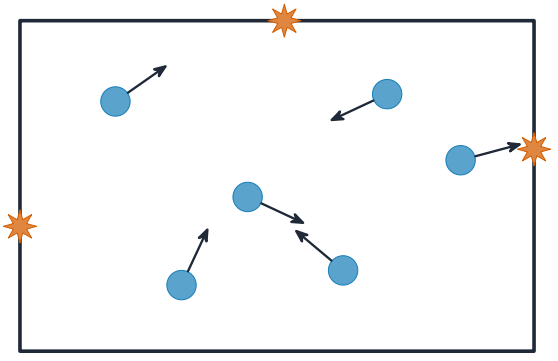
The gaseous state



Boiling turns liquid water into steam — a change between states of matter.

Where gas pressure comes from

Gas molecules move fast in all directions. They keep hitting — **colliding** 碰撞 with — the walls of their container. Each hit gives the wall a tiny push. The **pressure** 压强 of the gas is the overall result of these many collisions on the walls.



fast particles collide with the walls; the many pushes make the **pressure**

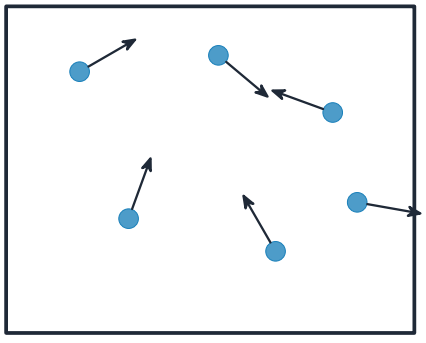
Gas pressure: fast molecules move in all directions and collide with the walls; the many tiny pushes add up to the pressure

Ideal gases

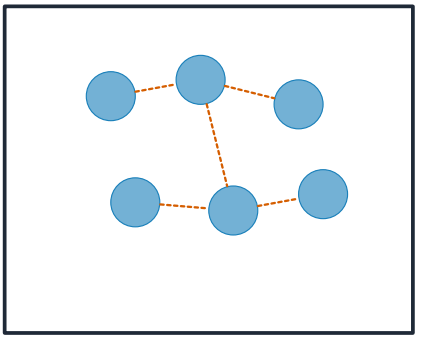
An **ideal gas** 理想气体 is a simple model. We assume two things:

- the particles themselves take up zero volume.
- there are no **intermolecular forces** 分子间作用力 of attraction between the particles.

A **real gas** 实际气体 follows this model closely at low pressure and high temperature. It behaves least like an ideal gas at high pressure and low temperature, when the particles are squeezed close together and the forces between them start to matter.



ideal gas (model)
point particles, *no* forces



real gas: high p, low T
real size + attractions (dashed)

An ideal gas is a model: point particles with no forces between them. A real gas behaves least like this at high pressure and low temperature, when the particles are crowded and their real size and attractions start to matter

The ideal gas equation

The **ideal gas equation** 理想气体方程 links pressure, volume, amount and temperature:

$$pV = nRT$$

where p is the pressure in Pa, V is the volume in m^3, n is the amount in moles, T is the temperature in **kelvin** 开尔文 (K), and R is the **gas constant** 气体常量 (8.31 J K^{-1} mol^{-1}).

Always change the units first: °C to K (add 273), and cm^3 or dm^3 to m^3.

Worked example. Find the volume of 0.50 mol of an ideal gas at 27 °C and 100 kPa. ($R = 8.31$ J K^{-1} mol^{-1}.)

Convert first: $T = 300$ K, $p = 1.00 \times 10^5$ Pa. Then

$$V = \frac{nRT}{p} = \frac{0.50 \times 8.31 \times 300}{1.00 \times 10^5} = 0.0125 \text{ m}^3 \ (= 12.5 \text{ dm}^3).$$

You can also use the equation to find a **molar mass** 摩尔质量. Since $n = m/M$:

$$pV = \frac{m}{M}RT \qquad \Rightarrow \qquad M = \frac{mRT}{pV}$$

This lets you work out M_r from the mass (or the density) of a gas.

Worked example. A flask holds 0.96 g of a gas in 600 cm^3 at 100 kPa and 27 °C. Find the molar mass of the gas. ($R = 8.31$ J K^{-1} mol^{-1}.)

Convert: $V = 6.00 \times 10^{-4}$ m^3, $T = 300$ K. Then

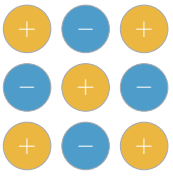
$$M = \frac{mRT}{pV} = \frac{0.96 \times 8.31 \times 300}{(1.00 \times 10^5)(6.00 \times 10^{-4})} \approx 40 \text{ g mol}^{-1}.$$

Bonding and structure

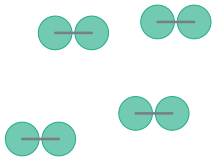


Quartz is a giant covalent structure of silicon and oxygen.

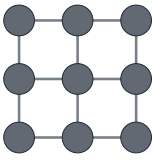
How a substance behaves depends on how its particles are joined. There are four main structures of a **crystalline solid** 晶体.



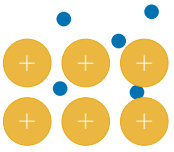
giant ionic
NaCl: high m.p.,
conducts when molten



simple molecular
I_2, ice: low m.p.,
weak forces between



giant covalent
diamond, SiO_2: very
high m.p., insulating



giant metallic
copper: high m.p.,
conducts (solid + molten)

The four structures of a crystalline solid — the structure decides the melting point, conductivity and solubility

Giant ionic

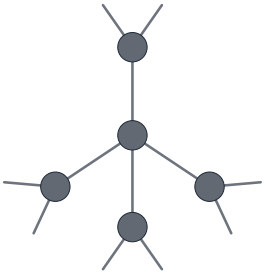
A **giant ionic** 离子晶体 structure is a huge regular **lattice** 晶格 of positive and negative **ions** 离子, held together by strong attraction in every direction. Examples are sodium chloride and magnesium oxide.

Simple molecular

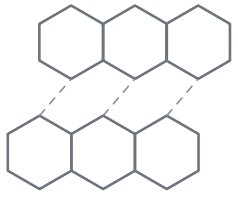
A **simple molecular** 分子晶体 structure is made of small **molecules** 分子. The bonds inside each molecule are strong, but the intermolecular forces between the molecules are weak. Examples are iodine (I_2), **fullerene** 富勒烯 (C_{60}) and ice.

Giant molecular

A **giant molecular** 原子晶体 structure (also called giant covalent) is a huge network of atoms joined by strong **covalent bonds** 共价键. Examples are **silicon(IV) oxide** 二氧化硅, **graphite** 石墨 and **diamond** 金刚石.



diamond
every C bonded to 4:
rigid 3D network, very hard



graphite
layers slide; spare
electrons let it conduct

Two giant covalent forms of carbon: diamond is a rigid 3D network (very hard); graphite has sliding layers and spare electrons that conduct



A real diamond in the rock it grew in. That whole crystal is one giant molecule — a single unbroken network of carbon atoms, each bonded to four others. Breaking it means breaking countless strong covalent bonds, which is why diamond is the hardest natural material

Giant metallic

A **giant metallic** 金属晶体 structure is a lattice of positive metal ions in a "sea" of **delocalised electrons** 离域电子. An example is copper.



A piece of pure copper metal. The shine, and the way metals conduct and bend, all come from that giant lattice of copper ions sitting in a shared sea of delocalised electrons

Physical properties

The structure decides the physical properties:

Structure	Melting/boiling point	Conducts electricity?	Solubility in water
giant ionic	high	only when molten or dissolved	usually soluble
simple molecular	low	no	usually low
giant molecular	very high	no (except graphite)	insoluble
giant metallic	high	yes (solid and molten)	insoluble

- **melting point** 熔点 and **boiling point** 沸点 are high when strong forces (ionic, covalent or metallic) must be broken, and low when only weak intermolecular forces break.
- **electrical conductivity** 导电性 needs charged particles that can move — ions that are free (when molten or dissolved) or delocalised electrons. Graphite conducts because some of its electrons are delocalised.
- **solubility** 溶解度 in water is usually high for ionic solids and low for molecular and giant covalent solids.

You can work backwards too: from the melting point, conductivity and solubility of an unknown substance, deduce the type of structure and bonding it has.

Exam tips

- Use **SI units** in $pV = nRT$: pressure in Pa, volume in m^3 ($cm^3 \times 10^{-6}$), temperature in K ($°C + 273$).
- State the **ideal-gas assumptions** (negligible molecular volume, no intermolecular forces) and when real gases deviate (high pressure, low temperature).
- Link each property to structure: giant ionic (high m.p., conducts molten), giant covalent (very high m.p.), simple molecular (low m.p.), giant metallic (conducts, malleable).
- Graphite conducts because each carbon has a **delocalised** electron; diamond does not — a favourite comparison.

Chemical energetics

A-Level Chemistry

Enthalpy change, ΔH



Burning fuel is exothermic, releasing energy to the surroundings.



An instant cold pack uses an endothermic reaction that takes in heat.

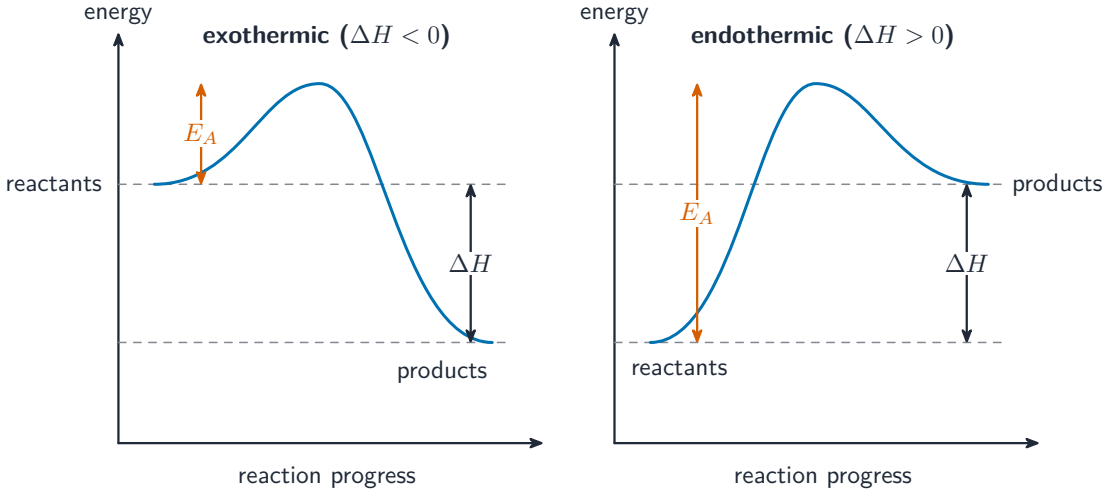
Every chemical reaction takes in or gives out energy. This energy change, measured at constant pressure, is the **enthalpy change** 焓变, with symbol ΔH.

- in an **exothermic** 放热 reaction the system gives out heat, so the products have less energy than the reactants and ΔH is **negative**.
- in an **endothermic** 吸热 reaction the system takes in heat, so the products have more energy than the reactants and ΔH is **positive**.

Reaction pathway diagrams

A **reaction pathway diagram** 反应路径图 shows the energy of the reactants and products, and the energy "hill" between them. The height of the hill is the **activation energy** 活化能 — the least energy the particles need before they can react.

- exothermic: products sit **lower** than reactants ($\Delta H < 0$).
- endothermic: products sit **higher** than reactants ($\Delta H > 0$).

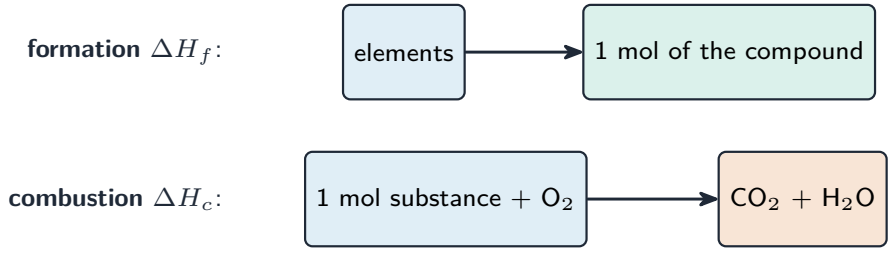


Exothermic reactions end lower than they start ($\Delta H < 0$); endothermic reactions end higher ($\Delta H > 0$)

Standard conditions and types of enthalpy change

Energy values are compared under **standard conditions** 标准条件: 298 K and 101 kPa, shown by the symbol $^{\ominus}$. Each substance is in its normal physical state at those conditions.

Symbol	Name	Definition (per mole, under standard conditions)
$\Delta H_r^{\ominus}$	**enthalpy change of reaction** 反应焓变	for the amounts shown in the equation
$\Delta H_f^{\ominus}$	**enthalpy change of formation** 生成焓变	one mole of a compound forms from its elements
$\Delta H_c^{\ominus}$	**enthalpy change of combustion** 燃烧焓变	one mole of a substance burns completely in oxygen
$\Delta H_{\text{neut}}^{\ominus}$	**enthalpy change of neutralisation** 中和焓变	one mole of water forms from an acid and an alkali



Don't confuse them: formation builds 1 mol of a compound from its elements; combustion burns 1 mol of a substance in oxygen

Energy from breaking and making bonds

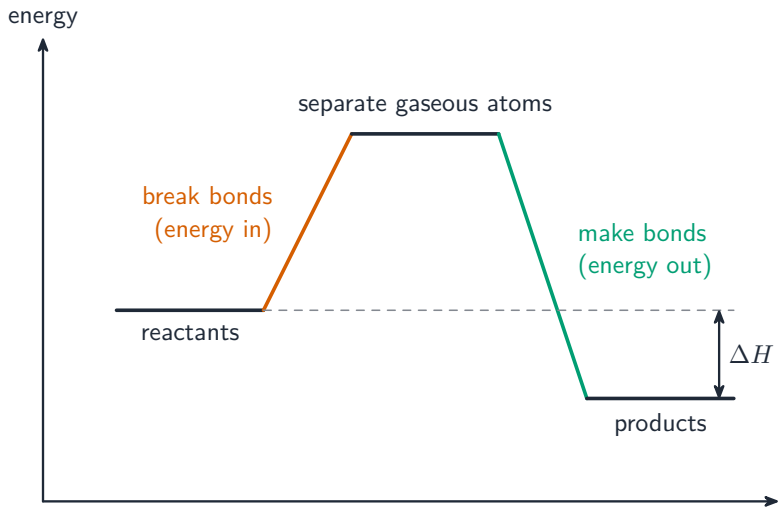
During a reaction, old bonds break and new bonds form. **Breaking** a bond needs energy (endothermic); **making** a bond releases energy (exothermic). The enthalpy change of the reaction is the difference between the two:

$$\Delta H_r = \sum(\text{bond energies broken}) - \sum(\text{bond energies made})$$

The **bond energy** 键能 is the energy needed to break one mole of a particular bond in the gas state, so it is always positive. Some bond energies are exact (for one specific molecule); others are averages taken over many different molecules, so calculations using them are only approximate.

Worked example. Use bond energies to find ΔH for $H_2 + Cl_2 \rightarrow 2HCl$. Bond energies (kJ mol^{-1}): H–H = 436, Cl–Cl = 242, H–Cl = 431.

$\Delta H = \sum(\text{broken}) - \sum(\text{made}) = (436 + 242) - (2 \times 431) = 678 - 862 = -184$ kJ mol^{-1}.



Breaking bonds takes energy in; making bonds gives energy out. ΔH is the difference between the two

Measuring enthalpy change in the lab

When a reaction heats up (or cools down) a known mass of water or solution, the heat transferred is:

$$q = mc\Delta T$$

where m is the mass, c is the **specific heat capacity** 比热容 (how much energy raises 1 g by 1 K), and ΔT is the temperature change. The enthalpy change per mole is then:

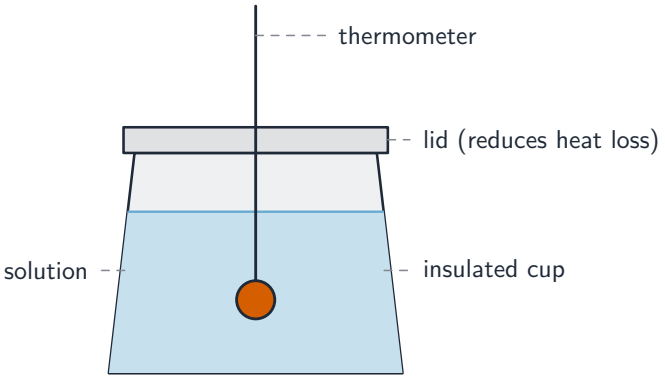
$$\Delta H = -\frac{mc\Delta T}{n}$$

The minus sign makes ΔH negative when the temperature rises (an exothermic reaction).

Worked example. Burning 0.50 g of methanol ($M_r = 32$) raises the temperature of 100 g of water by 18 °C. Find the enthalpy change of combustion per mole. ($c = 4.18$ J g^{-1} K^{-1}.)

The heat released is $q = mc\Delta T = 100 \times 4.18 \times 18 = 7520$ J. The amount burnt is $n = 0.50/32 = 0.0156$ mol, so

$$\Delta H_c = -\frac{q}{n} = -\frac{7520}{0.0156} \approx -480 \text{ kJ mol}^{-1}.$$

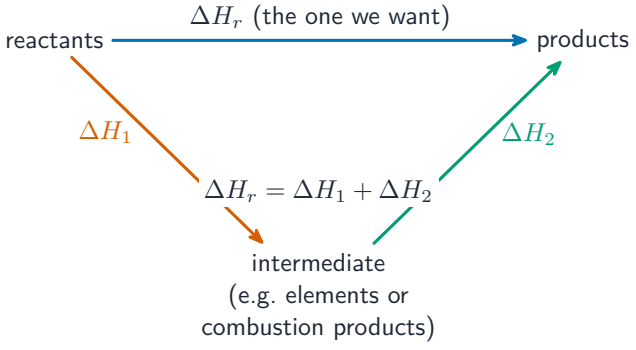


An insulated cup and a thermometer measure the temperature change of a known mass of solution

Hess's law

Hess's law 盖斯定律 says that the total enthalpy change for a reaction is the same, no matter which route you take from reactants to products. This is because energy is conserved.

This lets you draw an **energy cycle** 能量循环: link the reactants and products by a direct step and by an indirect route, then add the steps so that both routes give the same total.



The direct route equals the indirect route, so $\Delta H_r = \Delta H_1 + \Delta H_2$

Hess's law is useful in two ways:

- it lets you find an enthalpy change that you **cannot** measure directly (for example, the formation of a compound that forms slowly or with side reactions).
- it lets you calculate ΔH_r from bond energy data, or from formation or combustion data given in the question.

Exam tips

- Define each enthalpy change with its **exact standard conditions** (e.g. combustion = one mole burned completely in excess oxygen).
- Use $\Delta H = \sum$(bonds broken) $- \sum$(bonds made); breaking is endothermic (+), making is exothermic (−) — getting the sign the wrong way round is the classic error.
- In $q = mc\Delta T$ use the mass of the **water/solution**, then divide by moles and add the minus sign for an exothermic reaction.
- Draw **Hess cycles** with arrows the same way round, follow the alternative route, and always give ΔH a sign and units (kJ mol^{-1}).

Electrochemistry

A-Level Chemistry

Oxidation number

The **oxidation number** 氧化数 (also called the oxidation state) shows how many **electrons** 电子 an atom has lost or gained compared with the free element. You work it out using simple rules:

- an uncombined element has an oxidation number of 0.
- a simple ion has an oxidation number equal to its charge (so Mg^{2+} is +2).
- Group 1 is always +1, Group 2 is always +2.
- hydrogen is +1 (but −1 in metal hydrides).
- oxygen is −2 (but −1 in peroxides).
- fluorine is always −1.
- the oxidation numbers in a neutral compound add up to 0; in an ion they add up to the charge.

A Roman numeral shows the size of the oxidation number of an element, for example iron(II) means +2 and manganese(VII) in $KMnO_4$ means +7.

known	unknown	known
K	Mn	O_4
+1	+7	$4 \times (-2)$

they sum to 0: $(+1) + \text{Mn} + (-8) = 0$, so $\text{Mn} = +7$

Assigning an oxidation number: fix the known atoms (K = +1, O = −2 each), then use "they sum to zero" to find the unknown (Mn = +7)

Worked example. Find the oxidation number of sulfur in the sulfate ion, SO_4^{2-}.

Each oxygen is −2, so the four oxygens give $4 \times (-2) = -8$. In an ion the numbers add up to the charge, here −2. If sulfur is x:

$$x + (-8) = -2 \quad \Rightarrow \quad x = +6.$$

Worked example. Find the oxidation number of nitrogen in the nitrate ion, NO_3^-.

$$x + 3 \times (-2) = -1 \quad \Rightarrow \quad x - 6 = -1 \quad \Rightarrow \quad x = +5.$$

Redox in terms of electrons



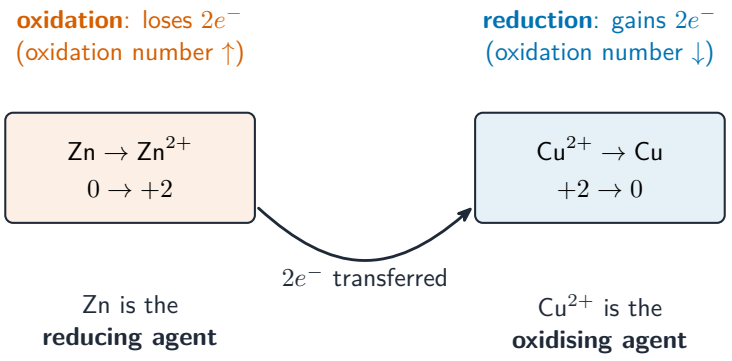
A battery uses redox reactions to push electrons round a circuit.

A **redox** 氧化还原 reaction is one where electrons move from one species to another.

- **oxidation** 氧化 is the **loss** of electrons. The oxidation number goes **up**.
- **reduction** 还原 is the **gain** of electrons. The oxidation number goes **down**.

A useful memory aid is **OIL RIG**: Oxidation Is Loss, Reduction Is Gain.

Oxidation and reduction always happen together, because the electrons lost by one species are gained by another. This **electron transfer** 电子转移 is why we call it a redox reaction.



Redox is electron transfer: the reducing agent loses electrons (oxidised, number up); the oxidising agent gains them (reduced, number down)

Using oxidation numbers to balance equations

Changes in oxidation number help you balance a redox equation:

1. find the element whose oxidation number **rises** (it is oxidised) and the one whose number **falls** (it is reduced).
2. the total rise must equal the total fall, because every electron lost is gained somewhere.
3. choose the ratio of the two species so the rise and fall match, then balance the rest of the equation.

$Fe^{2+} \rightarrow Fe^{3+}$ $\quad (+2 \rightarrow +3)$ — oxidised: rises 1 ($\times 5$)

$MnO_4^- \rightarrow Mn^{2+}$ $\quad (+7 \rightarrow +2)$ — reduced: falls 5 ($\times 1$)

5 electrons lost $=$ 5 gained $\Rightarrow$ ratio $\mathbf{5\,Fe^{2+}}$ **:** $\mathbf{1\,MnO_4^-}$

Balancing a redox equation: the total rise in oxidation number must equal the total fall, which fixes the ratio

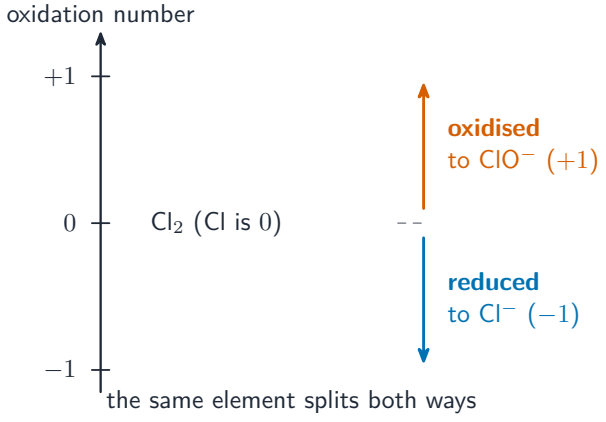
Worked example. Balance the reaction of manganate(VII) with iron(II) in acid: $MnO_4^- + Fe^{2+} + H^+ \rightarrow Mn^{2+} + Fe^{3+} + H_2O$.

Manganese falls from $+7$ to $+2$ (a fall of 5); iron rises from $+2$ to $+3$ (a rise of 1). To make the total fall equal the total rise, take 5 iron ions for every 1 manganate ion. Then balance oxygen with water ($4\,H_2O$) and hydrogen with H^+ ($8\,H^+$):

$$MnO_4^- + 5Fe^{2+} + 8H^+ \rightarrow Mn^{2+} + 5Fe^{3+} + 4H_2O.$$

Disproportionation

Disproportionation 歧化 is a special redox reaction in which the **same** element is both oxidised and reduced at the same time. For example, when chlorine reacts with cold water, some chlorine atoms are reduced (to Cl^-) and others are oxidised (to ClO^-).



Disproportionation: chlorine (0) is both oxidised to ClO^- (+1) and reduced to Cl^- (−1) at once

Oxidising and reducing agents



Rusting is the oxidation of iron by oxygen.

- an **oxidising agent** 氧化剂 takes electrons away from another species. In doing so, it is itself reduced.
- a **reducing agent** 还原剂 gives electrons to another species. In doing so, it is itself oxidised.

So in any redox reaction, the oxidising agent causes the oxidation of the other species, while the reducing agent causes the reduction.

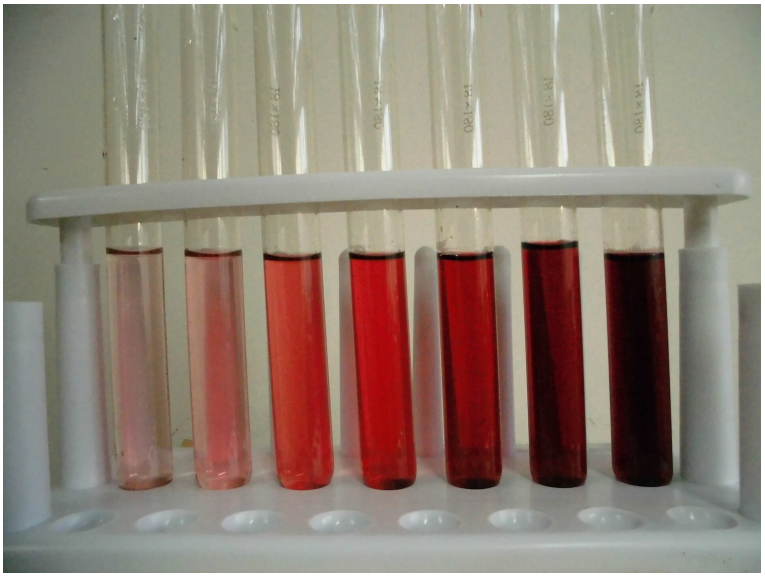
Exam tips

- Apply the oxidation-number rules in order: O is -2 **except** in peroxides (-1); H is $+1$ **except** in metal hydrides (-1).
- In a redox equation the **total increase equals the total decrease** in oxidation number — use this to fix the ratio, then balance O with H_2O and H with H^+.
- An **oxidising agent is itself reduced**; name both what is oxidised/reduced *and* the agent for full marks.
- **Disproportionation** is the same element both oxidised and reduced — show both oxidation-number changes.

Equilibria

A-Level Chemistry

Reversible reactions and dynamic equilibrium



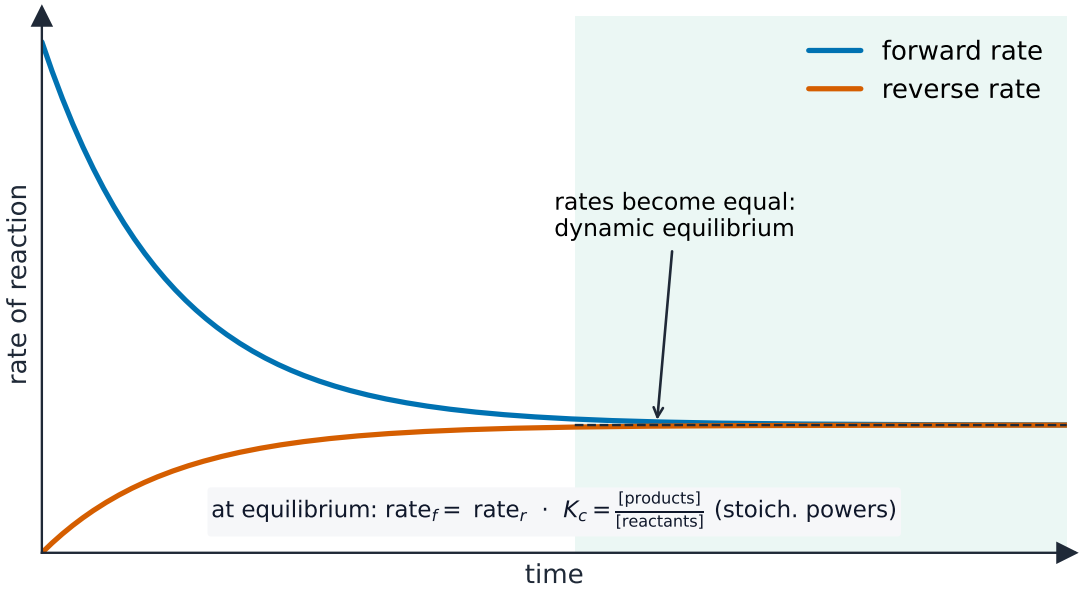
Cobalt chloride changes colour as its equilibrium shifts.

A **reversible reaction** 可逆反应 can go both ways. The **forward reaction** 正反应 makes products; the **reverse reaction** 逆反应 turns products back into reactants. We show this with the sign $\rightleftharpoons$.

In a **closed system** 封闭系统 (nothing enters or leaves), the reaction reaches a **dynamic equilibrium** 动态平衡 when:

- the rate of the forward reaction equals the rate of the reverse reaction.
- the concentrations of reactants and products stay constant.

It is called *dynamic* because both reactions are still happening — they just cancel out. A closed system is needed, or products would escape and equilibrium could never be reached.



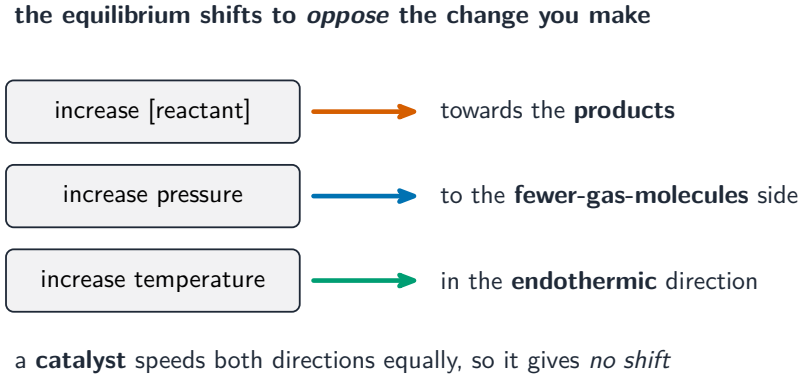
At dynamic equilibrium the forward and reverse rates have become equal, so the concentrations stay constant

Le Chatelier's principle

Le Chatelier's principle 勒夏特列原理 says: if you change a system at equilibrium, the position of **equilibrium** 平衡 moves to oppose (reduce) that change.

Change you make	Which way the equilibrium moves
increase concentration of a reactant	towards the products
increase pressure	towards the side with fewer gas molecules
increase temperature	towards the **endothermic** direction
add a **catalyst** 催化剂	no shift (it speeds up both ways equally)

A catalyst lets equilibrium be reached faster, but it does **not** change the position of equilibrium.



Le Chatelier's principle: the equilibrium always shifts to oppose the change you make

Equilibrium constants

For a reaction at equilibrium, the **equilibrium constant** 平衡常数 links the amounts of products and reactants. For the reaction $a\mathrm{A} + b\mathrm{B} \rightleftharpoons c\mathrm{C} + d\mathrm{D}$:

$$K_c = \frac{[\mathrm{C}]^c\,[\mathrm{D}]^d}{[\mathrm{A}]^a\,[\mathrm{B}]^b}$$

where the square brackets mean concentration in mol dm^{-3}.

Worked example. At equilibrium the mixture $\mathrm{H_2} + \mathrm{I_2} \rightleftharpoons 2\mathrm{HI}$ has $[\mathrm{H_2}] = 0.20$, $[\mathrm{I_2}] = 0.20$ and $[\mathrm{HI}] = 1.6$ mol dm^{-3}. Find K_c.

$$K_c = \frac{[\mathrm{HI}]^2}{[\mathrm{H_2}][\mathrm{I_2}]} = \frac{1.6^2}{0.20 \times 0.20} = 64$$

(no units here, because the concentration powers cancel top and bottom).

Worked example (ICE table). 1.00 mol each of H_2 and I_2 are sealed in a 1.00 dm^3 flask; at equilibrium 0.20 mol of H_2 remains. Set out an **ICE table** (Initial, Change, Equilibrium):

	$\mathbf{H_2}$	$\mathbf{I_2}$	$2\mathbf{HI}$
Initial	1.00	1.00	0
Change	-0.80	-0.80	$+1.60$
Equilibrium	0.20	0.20	1.60

0.80 mol of H_2 reacted, so $2 \times 0.80 = 1.60$ mol HI formed. In the 1.00 dm^3 flask the concentrations equal the moles, so $K_c = \dfrac{1.60^2}{0.20 \times 0.20} = 64$.

For reactions of gases, we use **partial pressure** 分压 instead of concentration. The partial pressure of a gas is the share of the total pressure that it provides. It is found from the **mole fraction** 摩尔分数 (the fraction of all the moles that are that gas):

$$\text{partial pressure} = \text{mole fraction} \times \text{total pressure}$$

The constant written with partial pressures is K_p.

Worked example. A gas mixture holds 2.0 mol of N_2 and 6.0 mol of H_2 at a total pressure of 200 kPa. Find the partial pressure of each gas.

There are 8.0 mol in total, so

$$p(N_2) = \frac{2.0}{8.0} \times 200 = 50 \text{ kPa}, \qquad p(H_2) = \frac{6.0}{8.0} \times 200 = 150 \text{ kPa}.$$

Worked example (a numeric K_p). For $N_2 + 3H_2 \rightleftharpoons 2NH_3$ the equilibrium partial pressures are $p(N_2) = 20$, $p(H_2) = 40$ and $p(NH_3) = 10$ kPa. Then

$$K_p = \frac{p(NH_3)^2}{p(N_2)\,p(H_2)^3} = \frac{10^2}{20 \times 40^3} = 7.8 \times 10^{-5} \text{ kPa}^{-2},$$

the units coming from $\dfrac{\text{kPa}^2}{\text{kPa} \times \text{kPa}^3} = \text{kPa}^{-2}$.

Only **temperature** changes the value of K_c or K_p. Changing concentration or pressure shifts the position of equilibrium but leaves the constant unchanged; adding a catalyst changes **neither** the constant nor the position — it only makes equilibrium arrive faster.

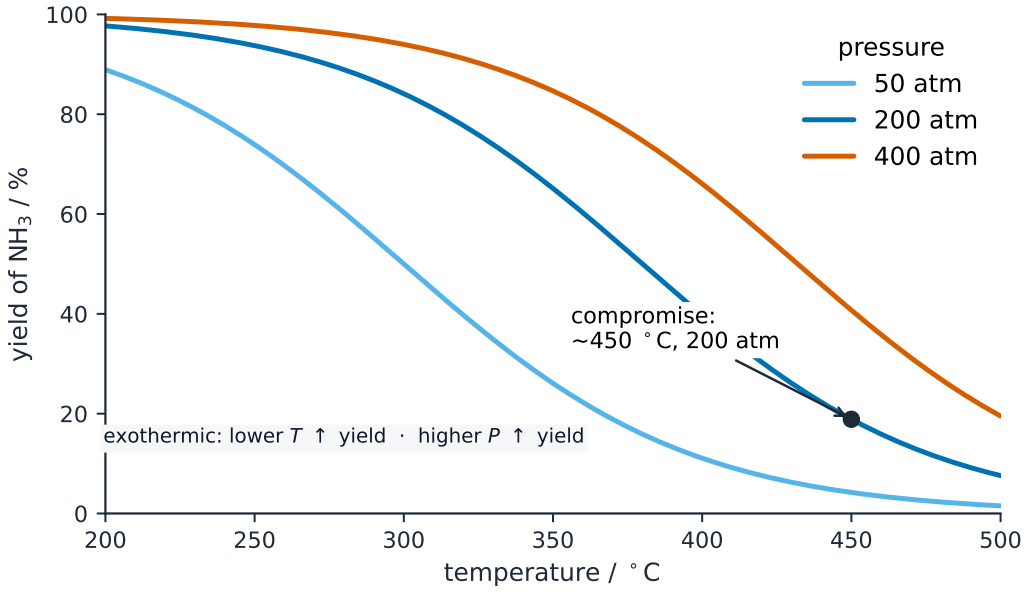
The Haber and Contact processes



Ammonia for fertiliser is made by the Haber process in large industrial plants like this one.

These two industrial processes are chosen by balancing yield, rate and cost using Le Chatelier's principle.

- the **Haber process** 哈伯法 makes ammonia: $N_2 + 3H_2 \rightleftharpoons 2NH_3$ (exothermic). It uses about 450 °C, 200 atm and an iron catalyst. A low temperature would give more ammonia but too slowly, so a moderate temperature is a compromise.
- the **Contact process** 接触法 makes sulfur trioxide: $2SO_2 + O_2 \rightleftharpoons 2SO_3$ (exothermic). It uses about 450 °C, near 1–2 atm and a vanadium(V) oxide catalyst.

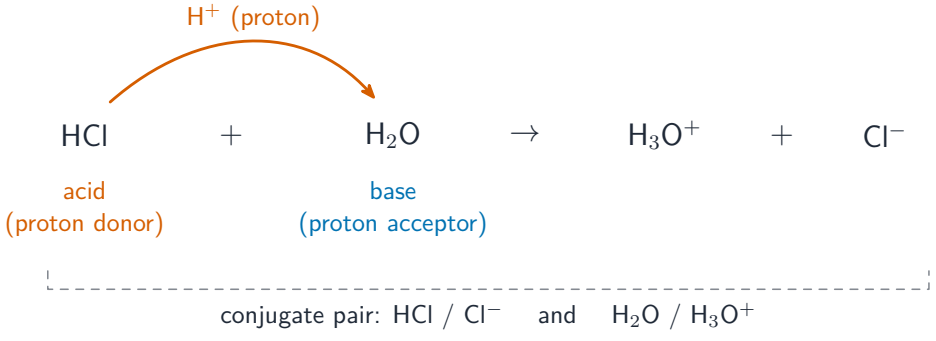


The Haber equilibrium gives more ammonia at lower temperature and higher pressure; about 450 °C and 200 atm is the working compromise

Acids and bases: the Brønsted–Lowry theory

A **proton** 质子 is simply an H^+ ion. The Brønsted–Lowry theory defines acids and bases by what they do with protons:

- an **acid** 酸 is a **proton donor** 质子供体 (it gives away H^+).
- a **base** 碱 is a **proton acceptor** 质子受体 (it takes H^+). A base that dissolves in water is called an alkali.



Brønsted–Lowry: the acid donates a proton (H^+) to the base, making two conjugate acid–base pairs

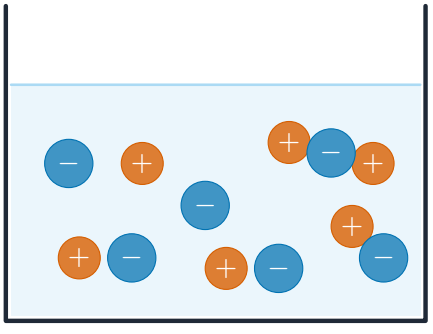
Common acids you must know: hydrochloric acid (HCl), sulfuric acid (H_2SO_4), nitric

acid (HNO_3) and ethanoic acid (CH_3COOH). Common alkalis: sodium hydroxide (NaOH), potassium hydroxide (KOH) and ammonia (NH_3).

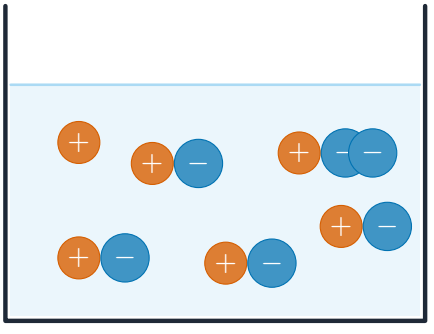
Strong and weak acids and bases

This is about how fully an acid or base splits up in water — not how concentrated it is.

- a **strong acid** 强酸 or **strong base** 强碱 is fully **dissociated** 解离 in water (almost every molecule splits into ions).
- a **weak acid** 弱酸 or **weak base** 弱碱 is only partly dissociated (most molecules stay whole).



strong acid
fully dissociated: many ions



weak acid
mostly molecules: few ions

Same concentration, different ionisation: a strong acid is fully dissociated into ions; a weak acid stays mostly as whole molecules

The **pH** scale measures how acidic a solution is: pure water is pH 7, acids are below 7, and alkalis are above 7. A strong acid has a lower pH than a weak acid of the same concentration.

You can tell strong and weak acids apart by:

- **reaction with a reactive metal**: a strong acid fizzes faster.
- **pH**: measured with a pH meter or **universal indicator** 通用指示剂.
- **electrical conductivity**: a strong acid conducts better, because it has more ions.

Neutralisation, salts and titration curves

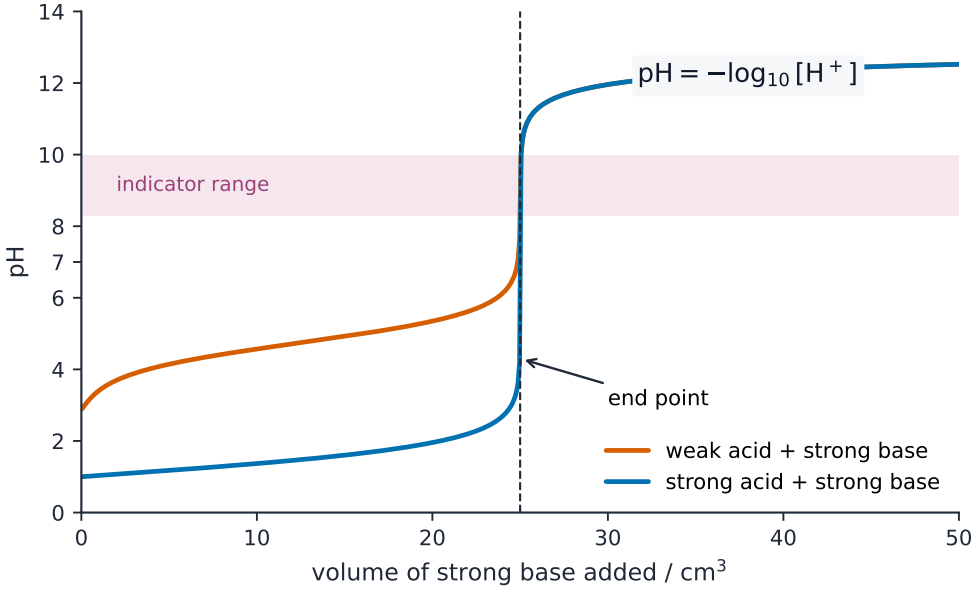
Neutralisation 中和 happens when the hydrogen ions from an acid react with the hydroxide ions from an alkali:

$$H^+(aq) + OH^-(aq) \rightarrow H_2O(l)$$

A **salt** 盐 is also formed, from the rest of the acid and base.

In a **titration** 滴定 you add one solution to another and follow the pH. The **titration curve** 滴定曲线 has a steep, almost vertical jump near the end point. The exact shape depends on whether each reactant is strong or weak.

To pick an **indicator** 指示剂, choose one whose colour change falls inside that steep jump. For a strong acid with a strong base most indicators work; for a weak acid with a strong base you need one that changes in the higher pH range.



Titration curves each have a steep pH jump at the end point. The weak-acid jump sits higher, so the indicator must change colour in that range

Exam tips

- Define **dynamic equilibrium** with all three points: forward and reverse rates equal, concentrations constant, closed system.
- Answer **Le Chatelier** questions by stating the change, the direction of shift, and the reason; a **catalyst does not shift** the position (it speeds both rates).

- Write K_c as **products over reactants**, each raised to its balancing number; include state symbols to decide what appears.
- Justify Haber/Contact conditions as a **compromise** between yield, rate and cost — do not just quote them.
- Brønsted: acid = **proton donor**, base = proton acceptor; conjugate pairs differ by one H^+.

Reaction kinetics

A-Level Chemistry

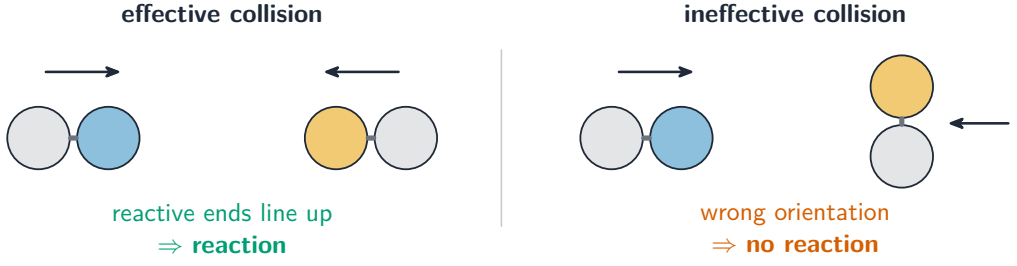
Rate of reaction

The **rate of reaction** 反应速率 is how fast reactants turn into products. We measure it as the change in concentration (or amount) in each unit of time.

To react, particles must **collide** 碰撞. The **collision frequency** 碰撞频率 is how often the particles hit each other. But not every collision leads to a reaction:

- an **effective collision** 有效碰撞 has enough energy *and* the correct direction, so a reaction happens.
- a **non-effective collision** 无效碰撞 does not have enough energy, or the particles hit at the wrong angle, so nothing happens.

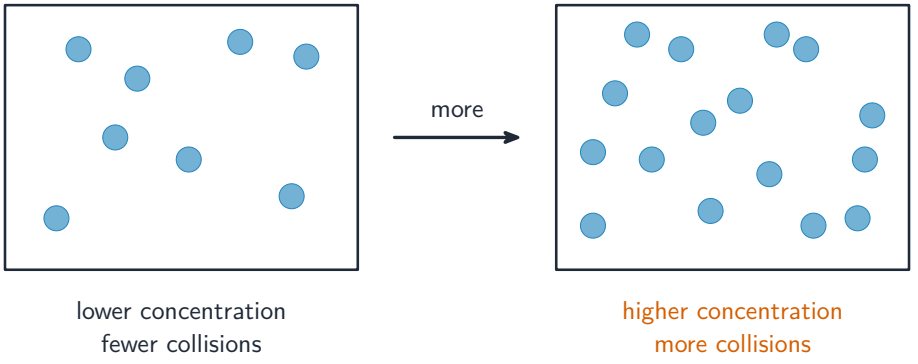
So the rate depends on the **frequency of effective collisions** — how many useful collisions happen each second.



A collision only reacts with the right orientation and enough energy (coloured ends = the reactive part)

Concentration and pressure

If you increase the concentration of a solution (or the pressure of a gas), the particles are packed closer together. They collide more often, so there are more effective collisions each second, and the rate goes up.



More particles in the same volume collide more often, so the rate rises

You can calculate a rate from experimental data — for example, the volume of gas made divided by the time taken.

Worked example. A reaction gives off carbon dioxide. In the first 30 s, 48 cm^3 of gas is collected. Find the average rate of reaction over this time.

$$\text{average rate} = \frac{\text{volume of gas}}{\text{time}} = \frac{48}{30} = 1.6\ \text{cm}^3\,\text{s}^{-1}.$$

The rate is fastest at the start (the graph is steepest there) because the reactants are most concentrated, then it slows as they are used up.

Temperature and activation energy

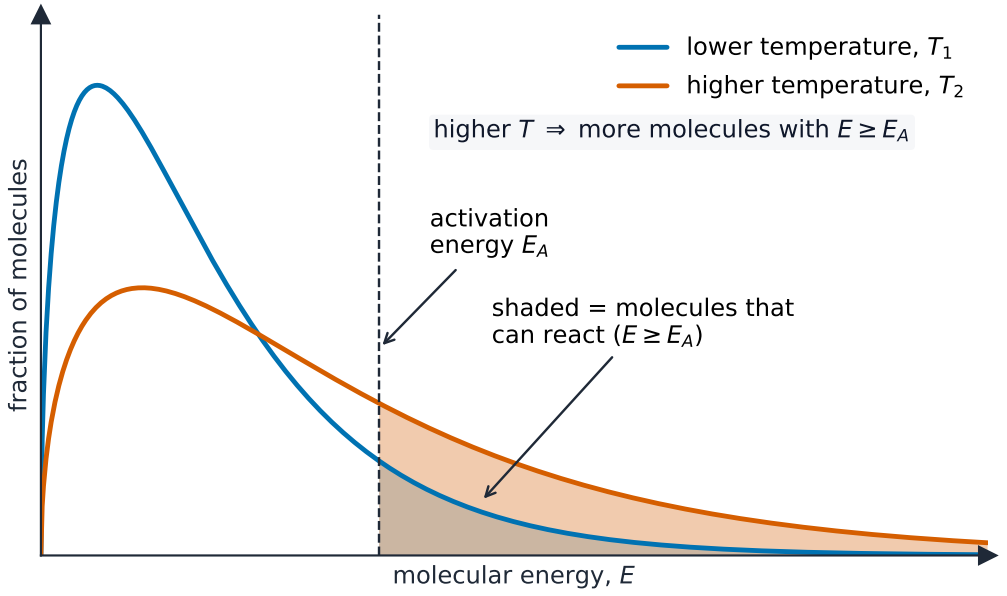
The **activation energy** 活化能 (E_A) is the minimum energy a collision needs in order to be effective.

The **Boltzmann distribution** 玻尔兹曼分布 is a graph showing how the energies of the molecules are spread out at one temperature. The curve starts at the origin, rises to a peak, then falls away in a long tail. The total area under the curve is the total number of molecules. Only the molecules to the right of E_A have enough energy to react.

When you raise the temperature:

- the curve flattens and spreads to the right, so a **much larger fraction** of molecules now have energy greater than E_A.
- the molecules also move faster and collide more often.

The first effect is the bigger one. This is why a small rise in temperature gives a large rise in rate.



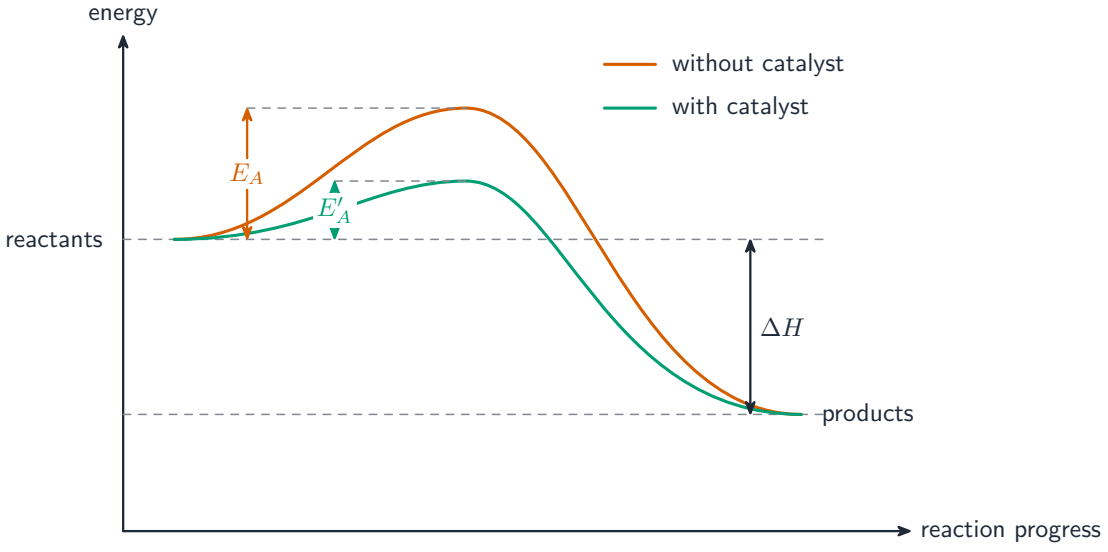
At higher temperature the curve spreads to the right, so a larger fraction of molecules can react

Catalysts

A **catalyst** 催化剂 speeds up a reaction but is not used up itself. **Catalysis** 催化作用 is the name for this action.

A catalyst works by giving the reaction a different **reaction mechanism** 反应机理 — a new route with a **lower** activation energy. On the Boltzmann distribution, lowering E_A moves the line to the left, so more molecules now have enough energy. This means more effective collisions each second, and a faster rate.

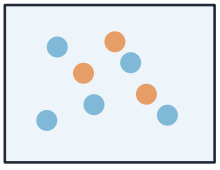
On a **reaction pathway diagram** 反应路径图, the catalysed route has a lower energy "hill". The enthalpy change of the reaction, ΔH, is **not** changed by the catalyst.



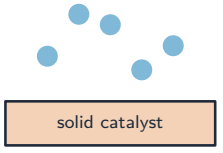
A catalyst gives a route with lower activation energy; the enthalpy change is unchanged

There are two types:

- a **homogeneous catalyst** 均相催化剂 is in the **same** physical state as the reactants — for example, an acid catalyst dissolved in a solution of liquids.
- a **heterogeneous catalyst** 多相催化剂 is in a **different** state from the reactants — for example, solid iron speeding up the reaction of gases in the Haber process.

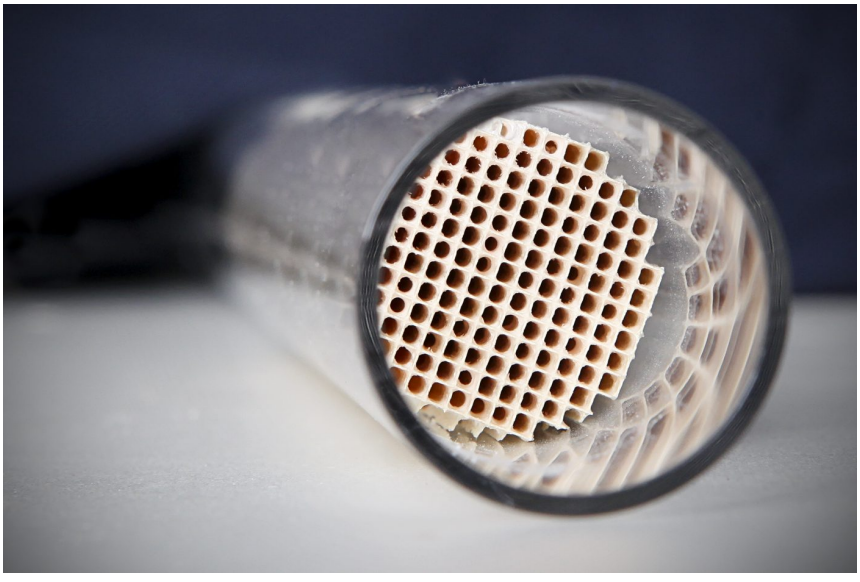


homogeneous
(same state, mixed in)



heterogeneous
(different state, a surface)

A homogeneous catalyst is mixed in with the reactants (same state); a heterogeneous catalyst is a separate surface (different state), where reactants meet and react



A car's catalytic converter is a heterogeneous catalyst; its honeycomb gives a huge surface area

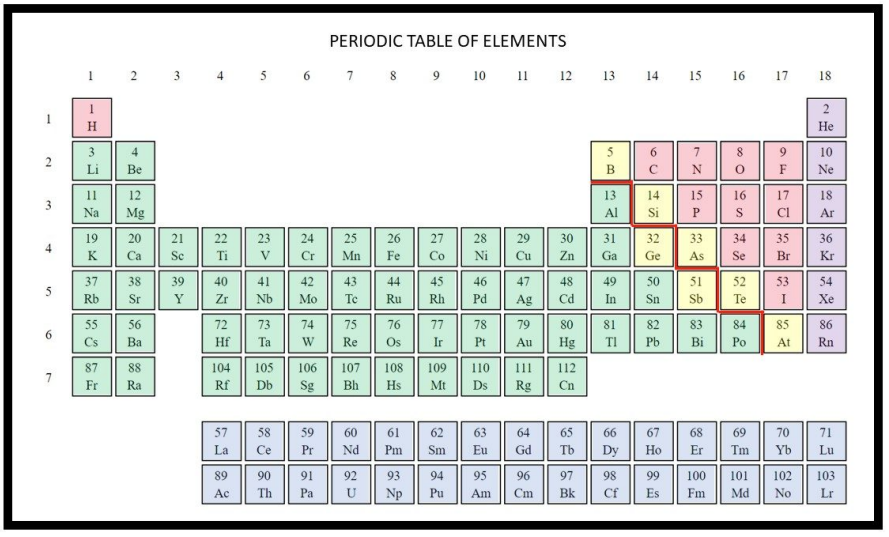
Exam tips

- Explain rate changes with **collision theory** — more frequent and/or more energetic **effective collisions**.
- On a **Boltzmann distribution** mark E_A, shade to its right, and show the curve **flatten and shift right** at higher temperature; it starts at the origin and never touches the axis.
- A **catalyst** gives an alternative route with **lower** E_A and does **not** change ΔH.
- Say why a small temperature rise gives a large rate rise: a **much greater proportion** of molecules now exceed E_A (the main effect).

The Periodic Table: chemical periodicity

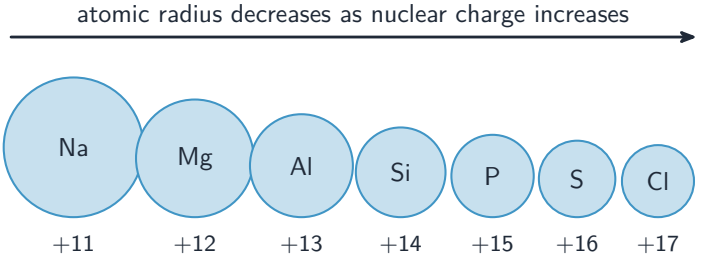
A-Level Chemistry

Physical properties across Period 3



Properties repeat in a regular pattern across each period of the table.

Periodicity 周期性 means that properties repeat in a regular pattern as you go across each **period** 周期 of the Periodic Table. Period 3 (Na to Ar) is the standard example.



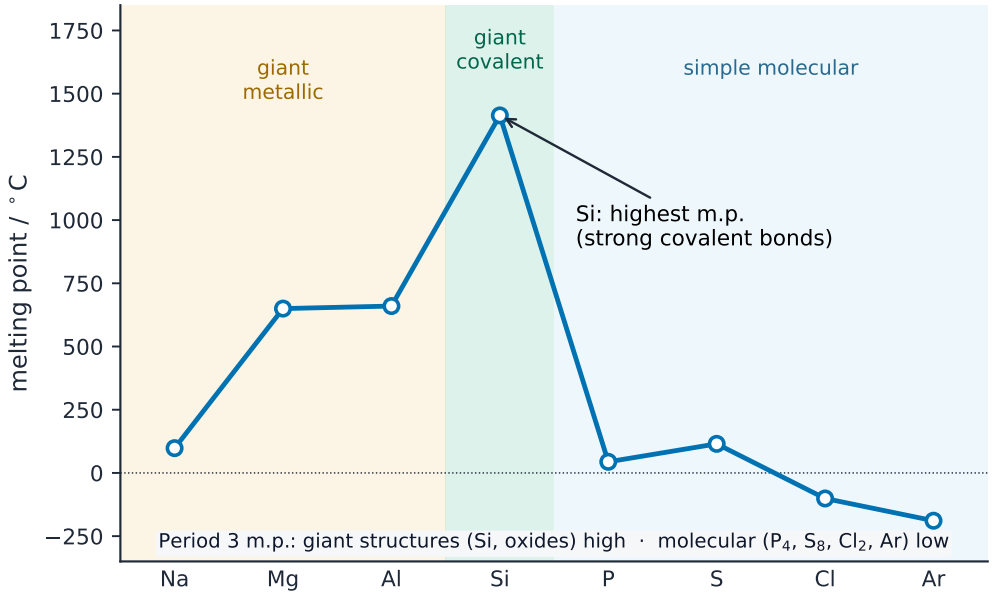
Atomic radius decreases across Period 3: the rising nuclear charge pulls the same outer shell inwards

Property	Trend across Period 3
atomic radius 原子半径	gets smaller (more nuclear charge pulls the same shell in)
ionic radius 离子半径	positive ions are small; from P^{3-} onwards the negative ions are larger
melting point 熔点	rises to a peak at silicon, then falls sharply
electrical conductivity 导电性	high for Na, Mg, Al; almost zero from Si onwards

The melting point and conductivity follow from the structure and bonding:

- **Na, Mg, Al** are **giant metallic** 金属晶体. Melting points rise (Na → Al) because each atom gives more delocalised electrons and the ions get smaller, so the bonding is stronger. They conduct well.
- **Si** is **giant molecular** 原子晶体 (giant covalent). It has the highest melting point, because strong covalent bonds must be broken. It barely conducts.
- **P, S, Cl, Ar** are **simple molecular** 分子晶体 (or single atoms). Their melting points are low, because the only forces that break are the weak **van der Waals' forces** 范德华力 **between** the molecules - the strong covalent bonds **inside** each molecule are never broken at all. They do not conduct.

Name **van der Waals' forces** when you explain that drop: it is the marking point, and "the covalent bonds are weak" is the error that loses it. Melting S_8 pulls whole molecules apart from each other; it does not touch a single S–S bond. Their sizes still follow the molecule: S_8 melts higher than P_4 because it is bigger, with more electrons and so stronger van der Waals' forces, while Ar (a single atom) is lowest of all.



Melting point across Period 3 peaks at silicon (giant covalent); it is high for the metals and low for the simple molecular elements

Chemical properties across Period 3

Reactions with oxygen, chlorine and water

With **oxygen**:

$$4Na + O_2 \rightarrow 2Na_2O \qquad 2Mg + O_2 \rightarrow 2MgO \qquad 4Al + 3O_2 \rightarrow 2Al_2O_3$$

$$P_4 + 5O_2 \rightarrow P_4O_{10} \qquad S + O_2 \rightarrow SO_2$$

With **chlorine**:

$$2Na + Cl_2 \rightarrow 2NaCl \qquad Mg + Cl_2 \rightarrow MgCl_2 \qquad 2Al + 3Cl_2 \rightarrow 2AlCl_3$$

$$Si + 2Cl_2 \rightarrow SiCl_4 \qquad 2P + 5Cl_2 \rightarrow 2PCl_5$$

With **water** (only Na and Mg react):

$$2Na + 2H_2O \rightarrow 2NaOH + H_2 \qquad Mg + 2H_2O \rightarrow Mg(OH)_2 + H_2$$

Sodium reacts fast; magnesium reacts only very slowly with cold water.

Oxidation number of the oxides and chlorides

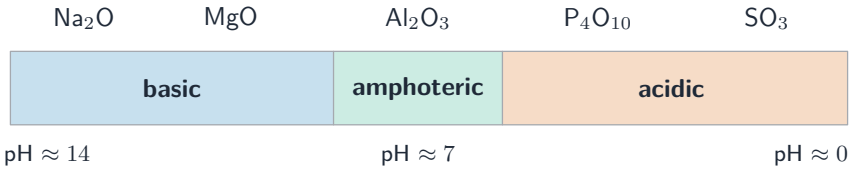
The **oxidation number** 氧化数 of the Period 3 element in its oxide or chloride rises across the period, because it equals the number of outer-shell (**valence shell** 价层) **electrons** 电子 the atom uses in bonding:

Oxides	**Na_2O**	**MgO**	**Al_2O_3**	**P_4O_{10}**	**SO_2 / SO_3**
oxidation number	+1	+2	+3	+5	+4 / +6

The chlorides NaCl, $MgCl_2$, $AlCl_3$, $SiCl_4$, PCl_5 show oxidation numbers +1 to +5 in the same way.

Oxides with water, and acid–base behaviour

Across the period the **oxides** 氧化物 change from basic to acidic:



metal oxides (left) are basic; non-metal oxides (right) are acidic

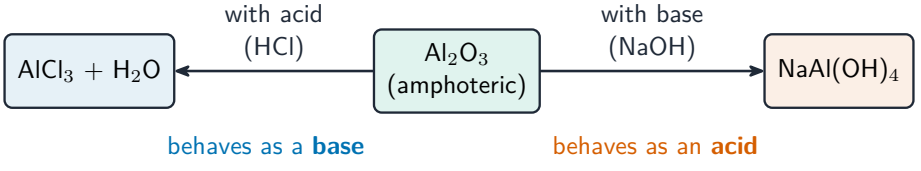
The Period 3 oxides change from basic (the metals) through amphoteric (Al_2O_3) to acidic (the non-metals)

Oxide	With water	Acid–base nature	Approximate pH
Na_2O	$Na_2O + H_2O \rightarrow 2NaOH$	basic	13–14
MgO	$MgO + H_2O \rightarrow Mg(OH)_2$	basic	9–10
Al_2O_3	insoluble	**amphoteric** 两性	7
SiO_2	insoluble	weakly acidic	7
P_4O_{10}	$P_4O_{10} + 6H_2O \rightarrow 4H_3PO_4$	acidic	1–2
SO_3	$SO_3 + H_2O \rightarrow H_2SO_4$	strongly acidic	0–1

Metal oxides (left) are basic; non-metal oxides (right) are acidic. Al_2O_3 and its **hydroxide** 氢氧化物 $Al(OH)_3$ are amphoteric — they react with **both** acids and bases:

$$Al_2O_3 + 6HCl \rightarrow 2AlCl_3 + 3H_2O$$

$$Al_2O_3 + 2NaOH + 3H_2O \rightarrow 2NaAl(OH)_4$$

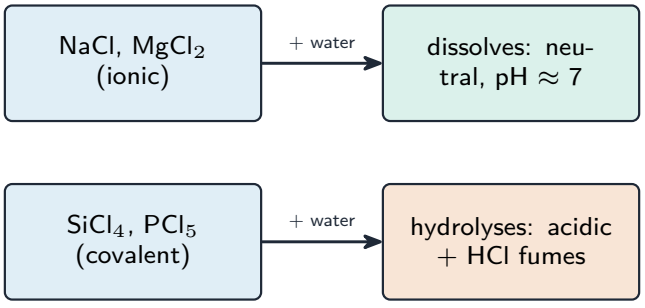


Aluminium oxide is amphoteric — it reacts with acids (behaving as a base) and with bases (behaving as an acid)

Chlorides with water

- $NaCl$ and $MgCl_2$ are ionic. They simply dissolve, giving a near-neutral solution.
- $SiCl_4$ and PCl_5 are covalent. They undergo **hydrolysis** 水解 (react with water) to make acidic solutions and fumes of HCl:

$$SiCl_4 + 2H_2O \rightarrow SiO_2 + 4HCl \qquad PCl_5 + 4H_2O \rightarrow H_3PO_4 + 5HCl$$



The ionic chlorides just dissolve (neutral); the covalent chlorides react with water (hydrolyse) to give an acidic solution and HCl fumes

Explaining the trends

These trends follow from the change in bonding and **electronegativity** 电负性. On the left, the elements are metals with low electronegativity, so their oxides and chlorides are ionic and basic (or neutral). On the right, the elements are non-metals with high electronegativity, so their oxides and chlorides are covalent and acidic. You can use a chloride's or oxide's properties (melting point, conductivity, effect on water) to suggest whether its bonding is ionic or covalent.

Worked example. Predict the pH when Na_2O, Al_2O_3 and SO_3 are each added to water. Acid-base character follows the **metal to non-metal** change across Period 3. Na_2O is ionic and basic: it dissolves to give NaOH, so the pH is about **13**. SO_3 is covalent and acidic: it reacts to give H_2SO_4, so the pH is about **1**. Al_2O_3 sits at the changeover - it is **amphoteric** and essentially **insoluble** in water, so the pH stays about **7**. The trap is that last one: amphoteric does **not** mean neutral, it means the oxide reacts with acids **and** with alkalis - it simply does not react with water.

Periodicity of other elements

The same idea works for any group. If you know the pattern down a group and across a period, you can:

- **predict** the properties of an element from its position (for example, a Group 1 element will be a reactive metal forming a +1 ion).
- **deduce** the likely position and identity of an unknown element from its physical and chemical properties.

Exam tips

- Always pair the **trend with its reason**: atomic radius falls across Period 3 (rising nuclear charge, similar shielding).
- Explain the **melting-point pattern** by structure: giant metallic (Na→Al), giant covalent (Si, highest), then simple molecular (P_4, S_8, Cl_2) and monatomic (Ar).
- Oxides go **basic → amphoteric (Al_2O_3) → acidic** across the period; link to ionic-to-covalent bonding.
- For reactions of oxides/chlorides with water, give the **products and the resulting pH**.

Group 2

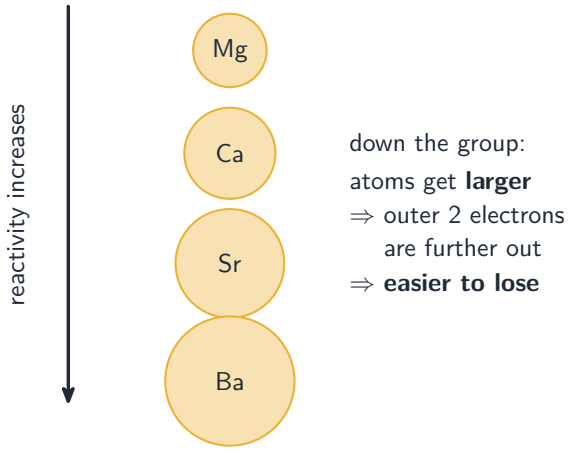
A-Level Chemistry

The Group 2 metals



Limestone is calcium carbonate — a compound of the Group 2 metal calcium.

Group 族 2 holds the metals magnesium, calcium, strontium and barium. They all have two outer electrons, which they lose to form 2+ ions. Going **down** the group, the atoms get larger and the outer electrons are easier to lose, so the metals get **more reactive** — their **reactivity** 反应活性 increases down the group.



Reactivity increases down Group 2: larger atoms hold their two outer electrons less tightly, so they are lost more easily

Reactions of the elements

With **oxygen**, they burn to form an oxide:

$$2\mathrm{Mg} + \mathrm{O_2} \rightarrow 2\mathrm{MgO}$$



Magnesium burns in air with a brilliant white flame, forming white magnesium oxide

With **water**, they form a hydroxide and hydrogen. The reaction gets faster down the group:

$$Ca + 2H_2O \rightarrow Ca(OH)_2 + H_2$$

Magnesium is slow with cold water but reacts fast with steam, giving MgO and H_2.

With **dilute hydrochloric or sulfuric acid**, they form a salt and hydrogen:

$$Mg + 2HCl \rightarrow MgCl_2 + H_2 \qquad Mg + H_2SO_4 \rightarrow MgSO_4 + H_2$$

With sulfuric acid the reaction slows down the group, because the **sulfates** 硫酸盐 formed (such as $BaSO_4$) are insoluble and coat the metal.

Reactions of the compounds

The **oxides** 氧化物 and **hydroxides** 氢氧化物 are basic. They react with water and with dilute acids:

$$MgO + H_2O \rightarrow Mg(OH)_2 \qquad Mg(OH)_2 + 2HCl \rightarrow MgCl_2 + 2H_2O$$

The **carbonates** 碳酸盐 react with dilute acids to give a salt, water and carbon dioxide:

$$CaCO_3 + 2HCl \rightarrow CaCl_2 + H_2O + CO_2$$

Thermal decomposition

Thermal decomposition 热分解 means breaking a compound apart by heating it.

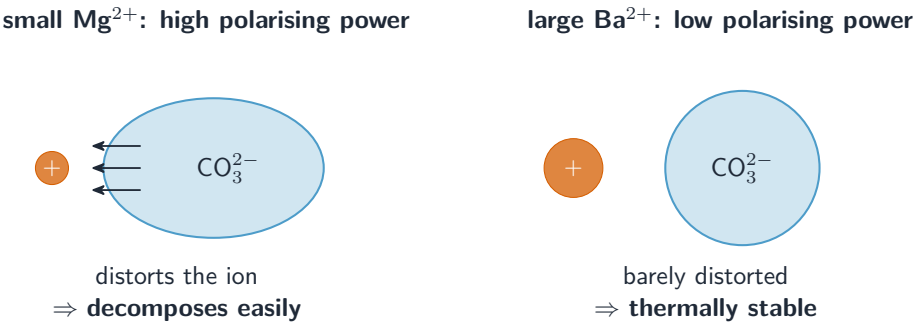
- the carbonates break into the oxide and carbon dioxide:

$$CaCO_3 \rightarrow CaO + CO_2$$

- the **nitrates** 硝酸盐 break into the oxide, brown nitrogen dioxide gas and oxygen:

$$2Ca(NO_3)_2 \rightarrow 2CaO + 4NO_2 + O_2$$

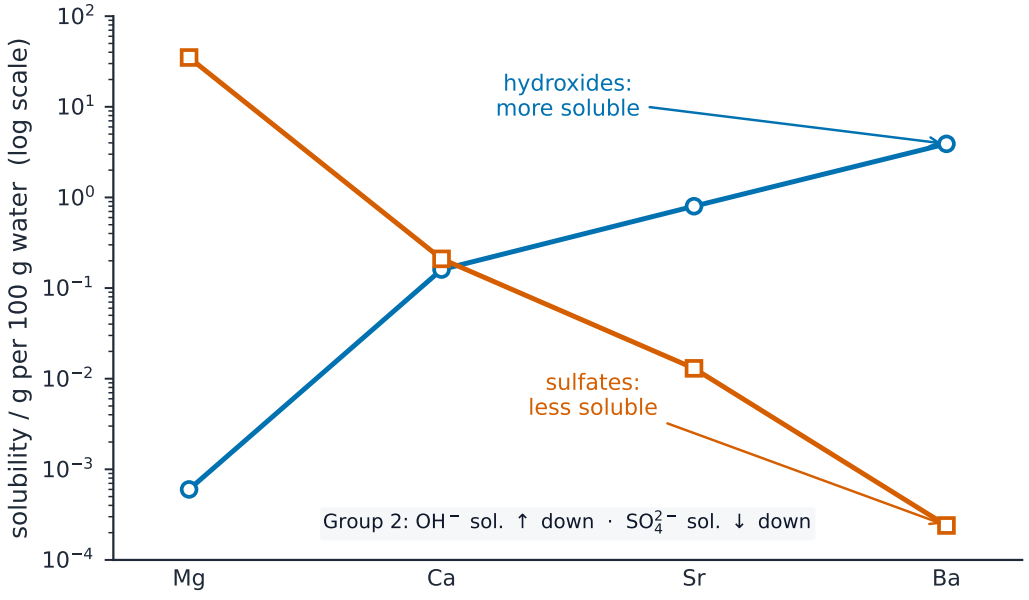
The **thermal stability** 热稳定性 of both the carbonates and the nitrates **increases** down the group. A larger metal ion pulls less on the carbonate or nitrate ion, so the compound is harder to break apart — it needs a higher temperature. So magnesium carbonate decomposes most easily, and barium carbonate is the hardest.



Thermal stability rises down the group: a small cation polarises (distorts) the carbonate ion more, weakening it so it decomposes more easily

Trends in solubility

Compound	Trend in solubility 溶解度 down the group
hydroxides	**increase** ($Mg(OH)_2$ almost insoluble; $Ba(OH)_2$ soluble)
sulfates	**decrease** ($MgSO_4$ soluble; $BaSO_4$ insoluble)



Down Group 2 the hydroxides become more soluble while the sulfates become less soluble (note the log scale)

From these trends you can predict the properties of the next element down, radium, and of its compounds.

Worked example. $MgCO_3$ decomposes at about 540 °C and $BaCO_3$ at about 1360 °C. Explain the trend and predict where $CaCO_3$ sits. Going **down** Group 2 the cation gets **larger**, so its charge is spread over a bigger surface and its **polarising power** falls. A less polarising cation distorts the carbonate ion less, so the C-O bond is weakened less and more heat is needed to break it. Thermal stability therefore **increases down the group**, and $CaCO_3$, lying between Mg and Ba, decomposes at an intermediate temperature of about 900 °C. Argue the whole chain - **cation size, then polarising power, then distortion of the anion**; "it is more reactive" explains nothing here.

Exam tips

- Reactivity **increases down** Group 2: ionisation energy falls as atoms get larger, so electrons are lost more easily.
- Reactions with water become **more vigorous** down the group; Mg reacts slowly with cold water but fast with steam ($\rightarrow$ MgO).
- Hydroxides get **more soluble** down the group; sulfates get **less soluble** ($BaSO_4$ insoluble — the basis of the sulfate test).
- Write equations with the correct +2 ion and state the observations (fizzing, dissolving).

Group 17

A-Level Chemistry

Physical properties of the halogens

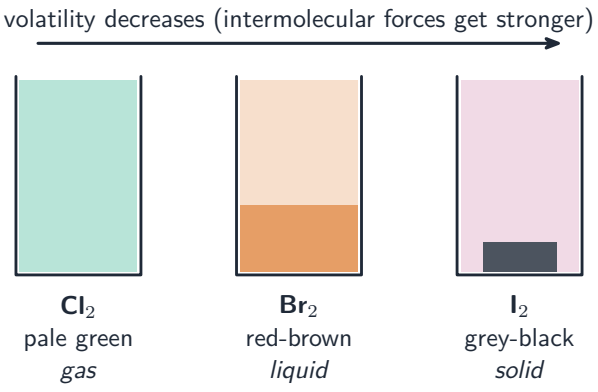


Bromine is a Group 17 halogen — a dark liquid that gives off an orange vapour.

The **halogens** 卤素 are the **Group** 族 17 elements. They exist as diatomic molecules (Cl_2, Br_2, I_2). Going down the group:

Element	Colour and state at room temperature
chlorine	pale green gas
bromine	red-brown liquid
iodine	grey-black solid (purple vapour)

The **volatility** 挥发性 (how easily a substance turns to vapour) **decreases** down the group: chlorine is a gas, but iodine is a solid. This is because the molecules get larger and have more electrons, so the **instantaneous dipole** 瞬时偶极 and **induced dipole** 诱导偶极 forces between them get stronger. Stronger forces are harder to break, so the boiling point rises and volatility falls.



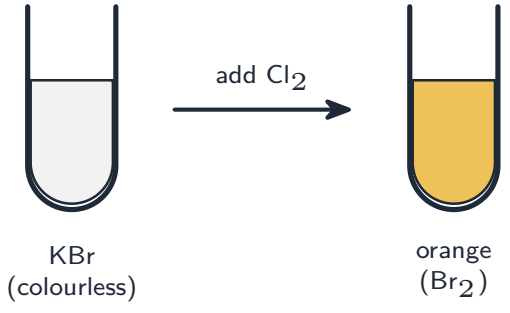
Down Group 17 the halogens change from a pale green gas to a red-brown liquid to a grey-black solid as volatility falls

The **bond energy** 键能 (bond strength) of the X–X molecules generally falls from Cl_2 to I_2, because the shared electrons are further from the nuclei in the larger atoms.

Chemical properties of the halogens and hydrogen halides

Halogens as oxidising agents

Each halogen reacts by gaining one electron to form a 1− ion, so it acts as an **oxidising agent** 氧化剂. This power **decreases** down the group, because the larger atoms attract an extra electron less strongly. A more reactive halogen can push out a less reactive one from its salt:



Chlorine displaces bromine: the solution turns orange

$$Cl_2 + 2KBr \rightarrow 2KCl + Br_2$$

Reactions with hydrogen

Each halogen reacts with hydrogen to form a **hydrogen halide** 卤化氢:

$$H_2 + Cl_2 \rightarrow 2HCl$$

The reaction gets less vigorous down the group: chlorine reacts explosively in light, bromine needs heat, and iodine reacts slowly and only partly.

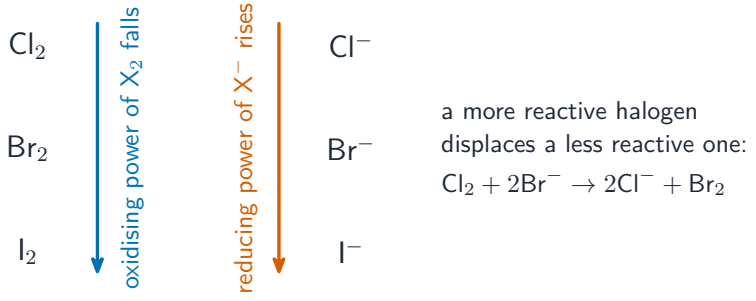
Thermal stability of the hydrogen halides

The **thermal stability** 热稳定性 of the hydrogen halides **decreases** down the group. The H–X bond gets weaker as the halogen atom gets larger, so HI breaks apart on gentle heating while HCl is very stable.

Reactions of the halide ions

Halide ions as reducing agents

A **halide ion** 卤离子 (such as Cl^-) can give away an electron, acting as a **reducing agent** 还原剂. This power **increases** down the group, because a larger ion holds its outer electron less tightly.

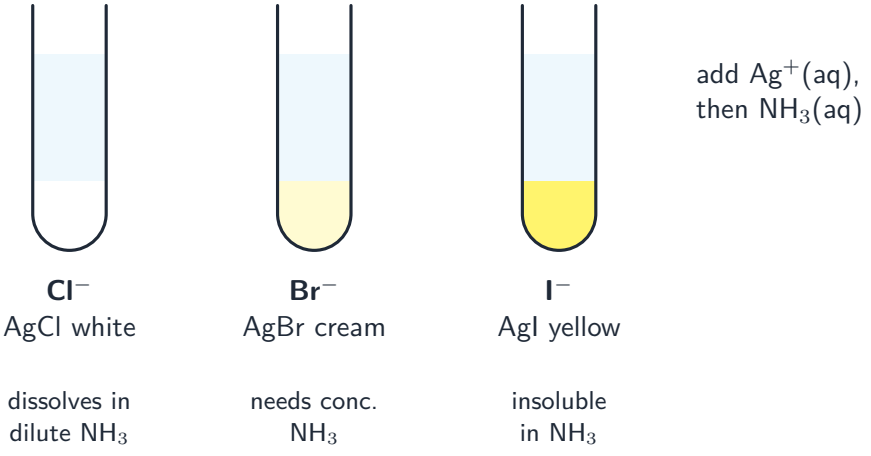


Two opposite trends: the oxidising power of the halogens falls down the group, while the reducing power of the halide ions rises

Reaction with aqueous silver ions

Add aqueous silver nitrate, then aqueous **ammonia** 氨, to identify the halide from the colour of the silver halide **precipitate** 沉淀:

Halide	Precipitate with Ag^+	Solubility in ammonia
Cl^-	white	dissolves in dilute ammonia
Br^-	cream	dissolves only in concentrated ammonia
I^-	yellow	insoluble in ammonia



Silver halide precipitates: AgCl is white, AgBr cream and AgI yellow — and their solubility in ammonia confirms which halide is present



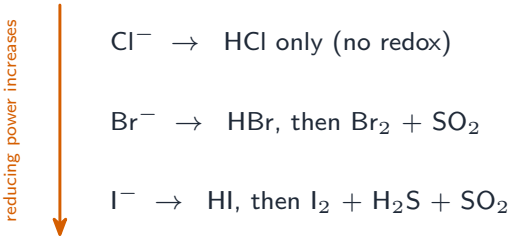
The silver halide test: AgCl is white, AgBr cream and AgI yellow

Reaction with concentrated sulfuric acid

All the halides first give the hydrogen halide. The lower halides are then oxidised by the acid, because they are stronger reducing agents:

$$NaCl + H_2SO_4 \rightarrow NaHSO_4 + HCl$$

- chloride gives only HCl (no redox).
- bromide also gives some brown Br_2 and SO_2.
- iodide gives I_2 and the smelly gases H_2S and SO_2, because I^- is the strongest reducing agent.



With concentrated sulfuric acid, chloride gives only HCl, but bromide and iodide (stronger reducing agents) are also oxidised to the halogen

Reactions of chlorine



Chlorine is added to pool water to kill microbes.

With sodium hydroxide

With **cold, dilute** sodium hydroxide, chlorine reacts to form chloride and chlorate(I):

$$Cl_2 + 2NaOH \rightarrow NaCl + NaClO + H_2O$$

With **hot, concentrated** sodium hydroxide, it forms chloride and chlorate(V):

$$3Cl_2 + 6NaOH \rightarrow 5NaCl + NaClO_3 + 3H_2O$$

In both, the **oxidation number** 氧化数 of chlorine goes both up and down (from 0), so both are **disproportionation** 歧化 reactions.

Chlorine in water purification

A little chlorine is added to water for **water purification** 水净化. It reacts with water:

$$Cl_2 + H_2O \rightleftharpoons HOCl + HCl$$

The active species HOCl and ClO^- kill **bacteria** 细菌, making the water safe to drink.

Worked example. Solid NaCl, NaBr and NaI are each warmed with concentrated H_2SO_4. Predict the products. **Reducing power increases down the group**, so how far each halide reduces the sulfuric acid differs. Cl^- is too weak to reduce it at all, so you get only steamy HCl - an acid-base reaction. Br^- reduces it a little: HBr plus brown Br_2 and SO_2. I^- is the strongest reducing agent: HI plus I_2, and it drives the sulfur all the way down to H_2S, with its bad-egg smell. **Every** halide gives the hydrogen halide first; the extra products appear only where the halide is a strong enough **reducing agent** to attack the sulfur.

Exam tips

- Halogens get **less reactive down** the group (harder to gain an electron); oxidising power decreases.
- **Displacement**: a more reactive halogen displaces a less reactive halide — state the colour change.
- Test halide ions with $AgNO_3$: white (Cl^-), cream (Br^-), yellow (I^-), then confirm with dilute/concentrated ammonia.
- Chlorine with water and with cold NaOH are **disproportionation** — show the oxidation-number changes.

Nitrogen and sulfur

A-Level Chemistry

Why nitrogen is unreactive

Nitrogen gas, N_2, makes up most of the air but reacts with very little. There are two reasons:

- the two nitrogen atoms are joined by a **triple bond** 三键, which has a very high **bond energy** 键能. A lot of energy is needed to break it.
- the molecule has no **polarity** 极性 — it is perfectly symmetrical, so nothing pulls other molecules towards it.



N≡N

strong triple bond: very high bond energy ⇒ hard to break

symmetrical molecule ⇒ no polarity ⇒ nothing pulls it into a reaction

Nitrogen is unreactive for two reasons: a very strong triple bond, and a symmetrical, non-polar molecule

Ammonia and the ammonium ion

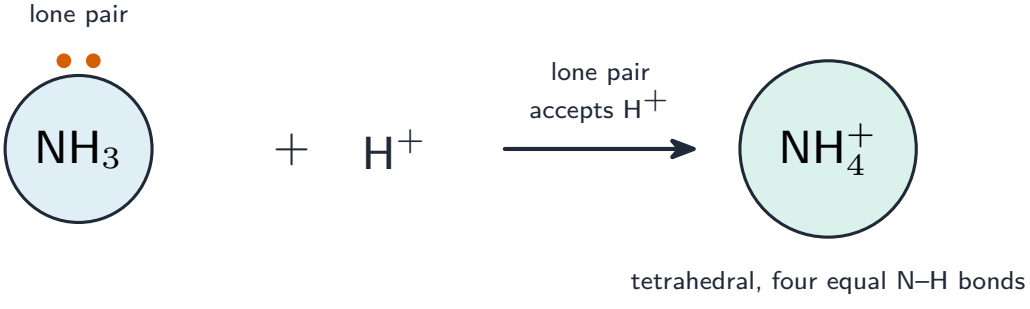


Ammonia is converted into nitrogen fertilisers in industrial plants on a huge scale.

Basicity of ammonia

The **basicity** 碱性 of ammonia (its ability to act as a base) comes from the **lone pair** 孤对电子 of electrons on the nitrogen atom. Using the Brønsted–Lowry theory, ammonia is a base because this lone pair can accept a **proton** 质子 (H^+):

$$NH_3 + H^+ \rightarrow NH_4^+$$



Ammonia acts as a base because the lone pair on its nitrogen accepts a proton, making the ammonium ion

The ammonium ion

When the lone pair forms a bond to H^+, it makes the **ammonium ion** 铵离子, NH_4^+. Because both shared electrons came from the nitrogen, this new bond is a **coordinate bond** 配位键. The ion has four identical N–H bonds and a tetrahedral shape.

Displacement of ammonia from its salts

If you warm an ammonium salt with a base (such as sodium hydroxide), you push out ammonia gas. This is an acid–base **displacement** 置换:

$$NH_4Cl + NaOH \rightarrow NaCl + NH_3 + H_2O$$

The sharp smell of ammonia, and damp red litmus turning blue, is a test for an ammonium salt.

Oxides of nitrogen and air pollution

Where they come from

Oxides of nitrogen (NO and NO_2, together called NO_x) come from two sources:

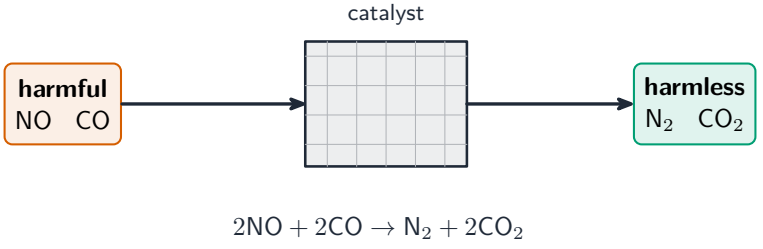
- **natural**: lightning gives enough energy for nitrogen and oxygen in the air to combine.
- **man-made**: the high temperature inside an **internal combustion engine** 内燃机 makes nitrogen and oxygen react:

$$N_2 + O_2 \rightarrow 2NO$$

Removing them from car exhaust

A **catalytic converter** 催化转化器 cleans the **exhaust gases** 尾气. It lets the harmful gases react together to form harmless ones:

$$2NO + 2CO \rightarrow N_2 + 2CO_2$$



In a catalytic converter the harmful gases NO and CO react over the catalyst to form harmless N_2 and CO_2

Photochemical smog

In sunlight, NO and NO_2 react with unburned **hydrocarbons** to form peroxyacetyl nitrate (PAN). PAN is a harmful part of **photochemical smog** 光化学烟雾, the brown haze seen over busy cities.

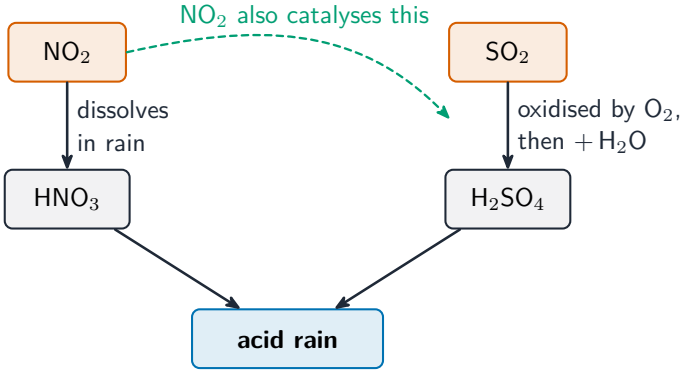


Photochemical smog over a big city. The brown layer is trapped low down, exactly where the traffic fumes are — sunlight turns those nitrogen oxides and unburned hydrocarbons into the haze you can see

Acid rain

The oxides of nitrogen also help make **acid rain** 酸雨 in two ways:

- directly: NO_2 dissolves in rain to form nitric acid.
- as a catalyst: NO_2 speeds up the oxidation of atmospheric **sulfur dioxide** 二氧化硫 (SO_2) into SO_3, which then forms sulfuric acid in the rain.



Nitrogen and sulfur oxides make acid rain: NO_2 forms nitric acid directly and also catalyses the oxidation of SO_2 to sulfuric acid

Worked example. A white solid is warmed with aqueous NaOH, and a gas is released that turns damp red litmus blue. Identify the gas and the ion in the solid, and explain the reaction. The only common gas that turns damp red litmus **blue** is **ammonia**, NH_3, so the solid contains the **ammonium ion**, NH_4^+. The hydroxide ion is the stronger base, so it takes the proton back from the ammonium ion:

$$NH_4^+ + OH^- \rightarrow NH_3 + H_2O$$

This is the standard test for NH_4^+. Always say the litmus is **damp**: the ammonia must dissolve in the water before it can show its basicity, so dry litmus would give no colour change at all.

Exam tips

- N_2 is unreactive because of its **strong triple bond** (very high bond energy) — state this exactly.
- **Ammonia** is a base and a ligand because of its **lone pair**; the ammonium ion forms by a dative bond.
- Explain how **oxides of nitrogen** form (high temperature in engines) and their link to acid rain and photochemical smog.

- Give balanced equations and correct observations for reactions of ammonia and the ammonium ion.

An introduction to AS Level organic chemistry

A-Level Chemistry

Formulas, functional groups and naming



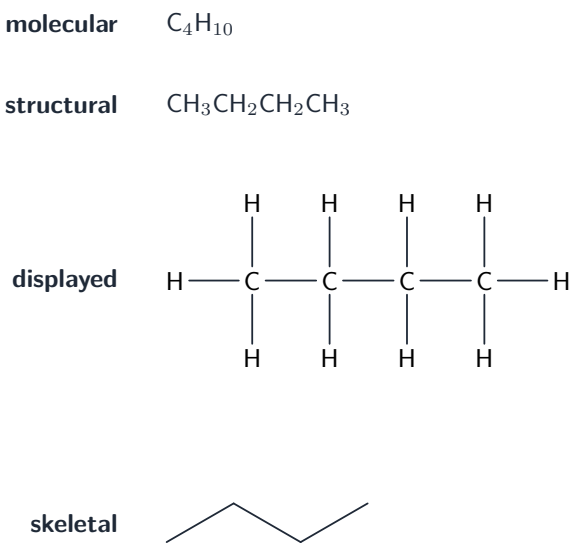
Crude oil is a complex mixture of hydrocarbons — the feedstock for organic chemistry.

A **hydrocarbon** 碳氢化合物 is a compound of only carbon and hydrogen. **Alkanes** 烷烃 are the simplest hydrocarbons and have no functional group.

A **functional group** 官能团 is the reactive part of a molecule. It decides the physical and chemical properties of the compound, so molecules with the same functional group behave alike (for example the –OH group in alcohols).

Types of formula

Formula	What it shows
general formula 通式	the pattern for a whole family, e.g. alkanes are C_nH_{2n+2}
molecular formula 分子式	the actual number of each atom, e.g. C_4H_{10}
structural formula 结构式	the groups in order, e.g. $CH_3CH_2CH_2CH_3$
displayed formula 展开式	every atom and every bond drawn out
skeletal formula 骨架式	lines for bonds; carbons at corners, hydrogens on carbon not shown



The same molecule (butane) shown four ways — molecular, structural, displayed and skeletal — each hiding more detail than the last

You can read off the **empirical formula** 实验式 (simplest ratio) from any of these.

Naming

Use systematic **nomenclature** 命名法: a stem for the number of carbons (meth-, eth-, prop-, but-, pent-, hex- for 1 to 6), an ending for the functional group, and numbers to show where groups are.

Characteristic organic reactions

Some key terms

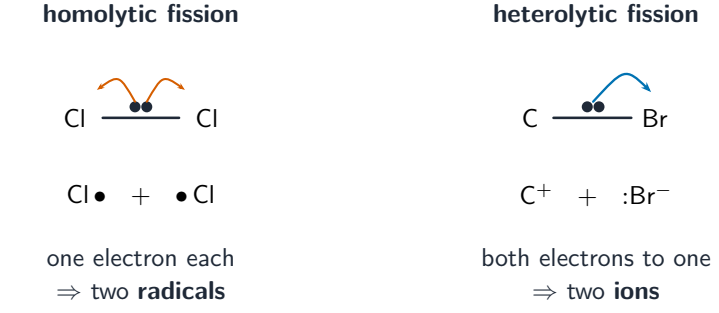
- a **homologous series** 同系列 is a family of compounds with the same functional group and general formula, where each member differs by CH_2.

- a **saturated** 饱和 compound has only single C–C bonds; an **unsaturated** 不饱和 compound has a C=C double bond (or a triple bond).

Breaking bonds

A covalent bond can break in two ways:

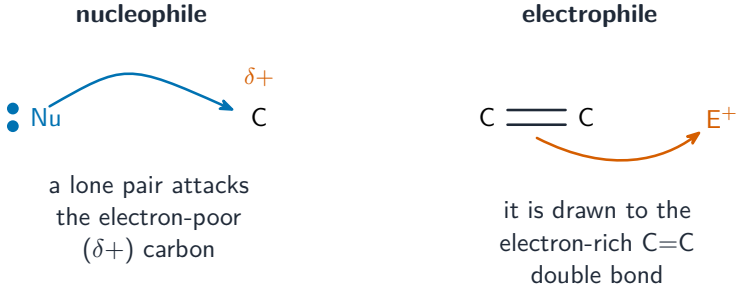
- **homolytic fission** 均裂: the bond splits evenly, one electron to each atom. This makes two **free radicals** 自由基 (species with an unpaired electron).
- **heterolytic fission** 异裂: the bond splits unevenly, both electrons going to one atom. This makes two ions.



Homolytic fission gives one electron to each atom (two radicals); heterolytic fission gives both electrons to one atom (two ions)

Attacking species

- a **nucleophile** 亲核试剂 is a species with a lone pair that is attracted to a positive (electron-poor) centre.
- an **electrophile** 亲电试剂 is a species attracted to a negative (electron-rich) centre, such as a C=C double bond.



A nucleophile uses its lone pair to attack an electron-poor centre; an electrophile is drawn to an electron-rich one such as a C=C bond

Types of reaction

Reaction	What happens
addition 加成	two molecules join to make one
substitution 取代	one atom or group is swapped for another
elimination 消去	a small molecule is removed, making a double bond
hydrolysis 水解	a molecule is split apart by water
condensation 缩合	two molecules join and a small molecule (such as water) is lost

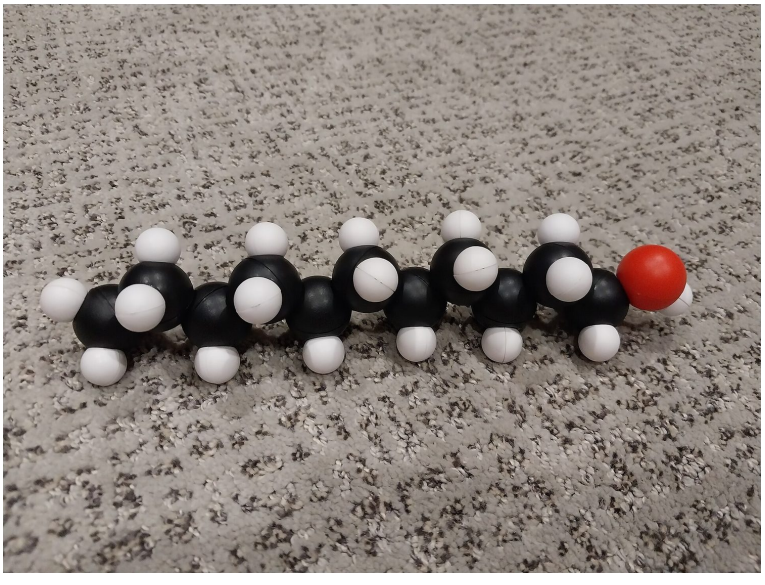
For organic redox, the symbol [O] stands for one oxygen atom from an oxidising agent, and [H] for one hydrogen atom from a reducing agent.

Types of mechanism

A free-radical reaction happens in three steps: **initiation** 引发 (radicals are made), **propagation** 增长 (radicals react and make new radicals), and **termination** 终止 (two radicals join and stop the chain).

The main mechanisms you meet are **free-radical substitution** 自由基取代 (alkanes), **electrophilic addition** 亲电加成 (alkenes), **nucleophilic substitution** 亲核取代 (halogenoalkanes) and **nucleophilic addition** 亲核加成 (carbonyls). In mechanisms, a **curly arrow** 弯箭头 shows a pair of electrons moving; it starts at a bond or a **lone pair** 孤对电子.

Shapes of organic molecules



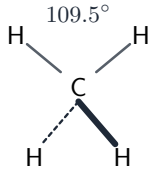
Organic molecules have definite three-dimensional shapes.

Organic molecules can be straight-chained, branched or **cyclic** 环状 (in a ring).

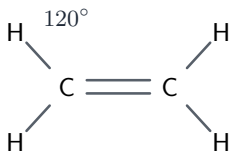
The shape around a carbon depends on its **hybridisation** 杂化:

Hybridisation	Bonds	Shape	Angle
sp^3	4 single	tetrahedral	109.5°
sp^2	1 double + 2 single	**planar** 平面 (flat)	120°
sp	1 triple (or 2 doubles)	linear	180°

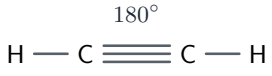
Every single bond is a **sigma bond** σ 键, made by direct overlap. A double bond is one sigma bond plus one **pi bond** π 键, made by sideways overlap of p orbitals. Ethene is planar because of its sp^2 carbons.



sp^3: 4 single bonds
tetrahedral



sp^2: 1 double + 2 single
planar (flat)



sp: 1 triple bond
linear

The shape at a carbon follows from its hybridisation: sp^3 is tetrahedral (109.5°), sp^2 is planar (120°), sp is linear (180°)

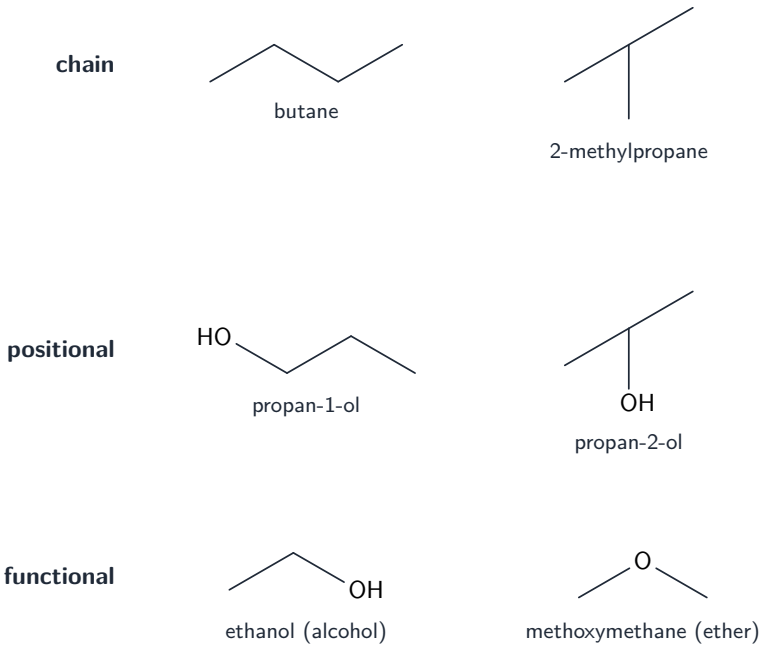
Isomerism

Isomers are compounds with the same molecular formula but a different arrangement of atoms. This is called **isomerism** 异构.

Structural isomerism

In **structural isomerism** 结构异构 the atoms are joined in a different order. There are three kinds:

- **chain isomerism** 链异构: the carbon chain is branched in different ways.
- **positional isomerism** 位置异构: the functional group is on a different carbon.
- **functional group isomerism** 官能团异构: the atoms form a different functional group (for example an alcohol and an ether).

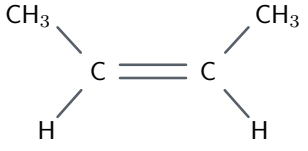


The three kinds of structural isomerism — chain, positional and functional-group — each pair sharing the same molecular formula

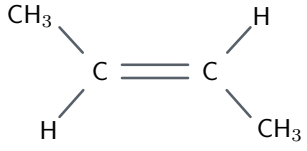
Stereoisomerism

In **stereoisomerism** 立体异构 the atoms are joined in the same order but point in different directions in space.

- **geometrical isomerism** 几何异构 (cis/trans) happens at a C=C double bond. The pi bond stops the two carbons rotating (**restricted rotation** 受限旋转), so groups are fixed on the same side (**cis** 顺式) or opposite sides (**trans** 反式).



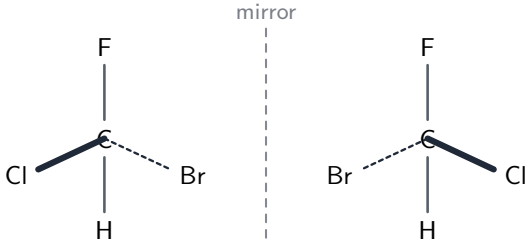
cis: both CH_3 on the same side



trans: CH_3 on opposite sides

Cis–trans isomerism at a C=C bond: the methyl groups are fixed on the same side (cis) or opposite sides (trans) because the bond cannot rotate

- **optical isomerism** 旋光异构 happens at a **chiral** 手性 carbon — a **chiral centre** 手性中心 is a carbon with four different groups attached. Such a carbon gives two mirror-image forms called **enantiomers** 对映体. A molecule may have more than one chiral centre.



enantiomers: mirror images that cannot be superimposed (the central carbon has 4 different groups)

Optical isomerism: a carbon with four different groups gives two mirror-image forms (enantiomers) that cannot be superimposed

From a molecular formula you can deduce the possible isomers by trying different chains, positions and functional groups.

Worked example. Explain why but-2-ene shows **cis-trans** (E/Z) isomerism but but-1-ene does not. Stereoisomerism at a C=C needs two things: restricted rotation about the double bond (both alkenes have that), **and** two **different** groups on **each**

of the two double-bond carbons. In but-2-ene, $CH_3CH{=}CHCH_3$, each double-bond carbon carries an H and a CH_3 - different, so the methyl groups can sit on the same side (**cis** / **Z**) or on opposite sides (**trans** / **E**). In but-1-ene, $CH_2{=}CHCH_2CH_3$, the first carbon carries **two hydrogens** - identical, so swapping them changes nothing and only one form exists. Test **each** double-bond carbon separately: two identical groups on **either** carbon kills the isomerism, however different the other carbon may be.

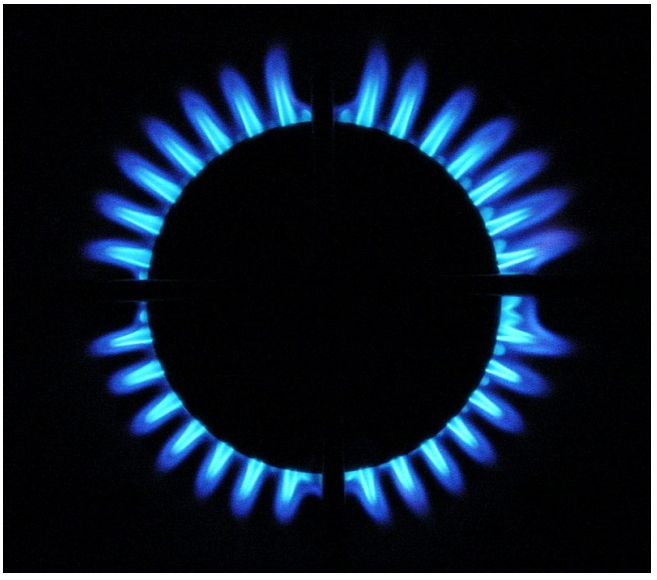
Exam tips

- Know the difference between **general, molecular, structural, displayed and skeletal** formulas — questions ask for a specific one.
- Name systematically: longest chain, lowest locants, substituents alphabetical; a wrong locant loses the mark.
- Classify each reaction by **type and reagent** (addition, substitution, elimination, oxidation).
- **E/Z isomerism** needs restricted rotation about a C = C **and** two different groups on each carbon.

Hydrocarbons

A-Level Chemistry

Alkanes



Natural gas — mainly methane, the simplest alkane — burns on a stove.

Alkanes 烷烃 are saturated hydrocarbons (general formula C_nH_{2n+2}).

Making alkanes

- **hydrogenation** 氢化: add hydrogen to an **alkene** 烯烃, using a Pt or Ni catalyst and heat.
- **cracking** 裂化: break a long-chain alkane into shorter ones by heating with Al_2O_3.

Combustion

In **combustion** 燃烧 an alkane burns in oxygen:

$$\text{complete: alkane} + O_2 \rightarrow CO_2 + H_2O$$

$$\text{incomplete: alkane} + \text{less } O_2 \rightarrow CO + \text{soot} + H_2O$$

Complete combustion gives CO2 and water; incomplete also gives CO and soot

- **complete combustion** 完全燃烧 (plenty of oxygen) gives carbon dioxide and water.
- **incomplete combustion** 不完全燃烧 (not enough oxygen) gives water plus toxic **carbon monoxide** 一氧化碳 (CO) and soot (carbon).

You are often asked to **write the balanced equation**. Balance it in a fixed order - **carbon first, then hydrogen, and oxygen last** - because oxygen is the only element left on just one side:

$$C_6H_{14} + 9\tfrac{1}{2}\,O_2 \rightarrow 6\,CO_2 + 7\,H_2O$$

The 6 carbons fix $6\,CO_2$; the 14 hydrogens fix $7\,H_2O$; counting the oxygens on the right gives $12+7=19$, so the left needs $19 \div 2 = 9\tfrac{1}{2}$. A **half** of O_2 is perfectly acceptable in this equation - and if the question asks for whole numbers, simply double everything ($2\,C_6H_{14} + 19\,O_2 \rightarrow 12\,CO_2 + 14\,H_2O$).

Free-radical substitution

Alkanes react with chlorine or bromine by **free-radical substitution** 自由基取代, in **ultraviolet light** 紫外线. For ethane and chlorine the mechanism has three steps:

- **initiation** 引发 — UV light splits the halogen into two **free radicals** 自由基: $Cl_2 \rightarrow 2\,Cl\cdot$ This kind of break is **homolytic fission** 均裂: the bond splits **evenly**, one electron going to each atom, which is what makes two radicals. (The opposite, **heterolytic fission** 异裂, sends **both** electrons to one atom and makes a pair of **ions** - that is what happens in the polar mechanisms such as electrophilic addition.) Examiners ask for this word by name, so use it.
- **propagation** 增长 — radicals react and make new radicals:

$$Cl\cdot + C_2H_6 \rightarrow C_2H_5\cdot + HCl \qquad C_2H_5\cdot + Cl_2 \rightarrow C_2H_5Cl + Cl\cdot$$

- **termination** 终止 — two radicals join, ending the chain: $Cl\cdot + C_2H_5\cdot \rightarrow C_2H_5Cl$

1. initiation

$Cl_2 \xrightarrow{UV} 2\,Cl^{\bullet}$

UV light splits Cl–Cl into two radicals

2. propagation

$Cl^{\bullet} + C_2H_6 \longrightarrow C_2H_5^{\bullet} + HCl$

$C_2H_5^{\bullet} + Cl_2 \longrightarrow C_2H_5Cl + Cl^{\bullet}$

each radical makes a new radical, so the chain continues

3. termination

$Cl^{\bullet} + C_2H_5^{\bullet} \longrightarrow C_2H_5Cl$

two radicals join, ending the chain

Free-radical substitution in three steps: initiation makes radicals, propagation carries the chain, and termination ends it

Why cracking is useful, and why alkanes are unreactive

Cracking turns heavy fractions of **crude oil** 原油 into more useful, lower-M_r alkanes and alkenes. (A **fraction** 馏分 is a group of molecules with a similar boiling-point range.)

Alkanes are generally unreactive, especially towards polar reagents. This is because the C–H and C–C bonds are strong and have little **polarity** 极性, so there is no charge to attract an attacking species.

Environmental effects

Burning alkanes in an internal combustion engine gives off carbon monoxide, oxides of nitrogen and unburnt hydrocarbons. A catalytic converter removes these by turning them into harmless gases.

Alkenes



Cracking heavy fractions at a refinery produces alkenes for plastics and fuels.

Alkenes have a C=C double bond (general formula C_nH_{2n}). The double bond is the functional group, so alkenes are reactive.

Making alkenes

- **elimination** 消去 of HX from a **halogenoalkane** 卤代烷, using NaOH dissolved in ethanol, with heat.
- **dehydration** 脱水 of an **alcohol** 醇, using a hot Al_2O_3 catalyst or concentrated sulfuric acid.
- cracking of a longer-chain alkane.

Reactions of alkenes

Most reactions are **electrophilic addition** 亲电加成 across the double bond:

Reagent and conditions	Product
H_2, Pt/Ni catalyst, heat	alkane
steam 水蒸气 (H_2O), H_3PO_4 catalyst	alcohol
a **hydrogen halide** 卤化氢 (HX), room temperature	halogenoalkane
a halogen X_2	a di-substituted alkane

There are also two oxidation reactions with acidified $KMnO_4$:

- **cold, dilute** $KMnO_4$ adds two –OH groups to give a **diol** 二醇.
- **hot, concentrated** $KMnO_4$ breaks the C=C bond right apart, and what each half turns into tells you exactly where the double bond used to be.

That second reaction is worth learning properly, because it is how the exam asks you to **locate a double bond in a large molecule**. Oxidation is written with [O], meaning "an oxygen atom from the oxidising agent". Cut the molecule at the C=C, then look at **what each of the two carbons was carrying**:

The C=C carbon carries	It becomes
two alkyl groups (=CR_2)	a **ketone** 酮, $R_2C{=}O$
one alkyl and one H (=CHR)	a **carboxylic acid** 羧酸, RCOOH
two hydrogens (=CH_2)	CO_2 and water

The pattern is simply how many hydrogens that carbon had: none stops at a ketone, one is pushed on to an acid, and two are oxidised all the way to CO_2.

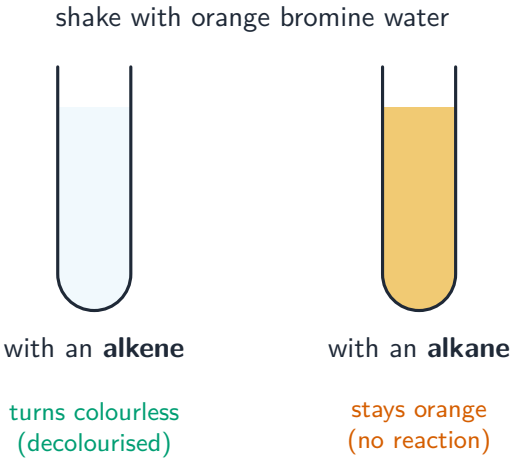
Worked example. Give the products when $(CH_3)_2C{=}CHCH_3$ is heated with hot concentrated acidified $KMnO_4$. Cut at the C=C and take each carbon separately. The **left** carbon carries **two methyl groups** and no hydrogen, so it stops at a **ketone**: $(CH_3)_2C{=}O$, which is propanone. The **right** carbon carries **one methyl and one H**, so it goes on to a **carboxylic acid**: CH_3COOH, ethanoic acid. So:

$$(CH_3)_2C{=}CHCH_3 + 3[O] \rightarrow CH_3COCH_3 + CH_3COOH$$

Now run it **backwards**, which is the way the question is usually set. Given the products, rebuild the alkene by putting the two carbonyl carbons back together as a C=C: a ketone means that carbon had **two alkyl groups**, an acid means **one alkyl and one H**, and CO_2 means the chain **ended** in $=CH_2$. Getting CO_2 is the strongest clue of all - it can only come from a **terminal** double bond.

Test for a C=C bond

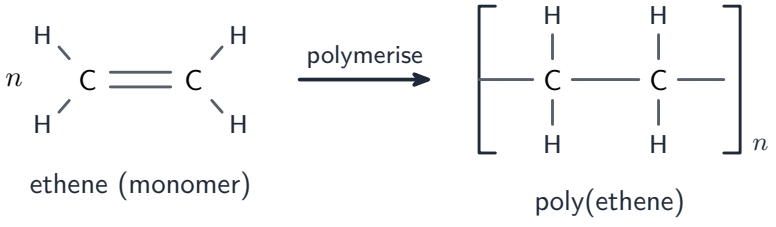
Shake the compound with orange **bromine water** 溴水. An alkene **decolourises** it (turns it colourless) by electrophilic addition. An alkane does not.



The bromine-water test: an alkene decolourises the orange bromine water, while an alkane leaves it orange

Addition polymerisation

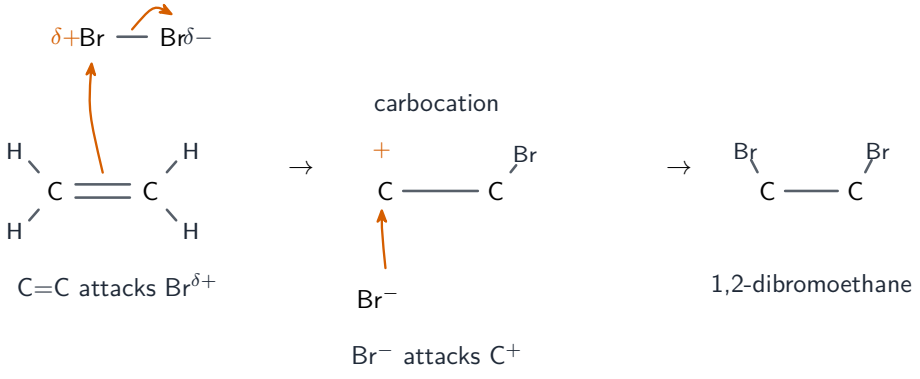
In **addition polymerisation** 加成聚合, many alkene molecules join into one long chain, with no other product. Ethene gives poly(ethene). The long-chain product is a **polymer** 聚合物.



Addition polymerisation: many ethene molecules open their double bonds and join into the long chain of poly(ethene)

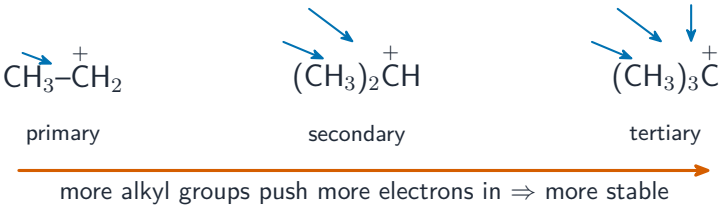
The mechanism and Markovnikov's rule

In electrophilic addition (for example bromine with ethene), the electron-rich C=C attracts the electrophile. This forms a positive intermediate called a **carbocation** 碳正离子, which the negative part then attacks.



Electrophilic addition of bromine to ethene: the C=C attacks $Br^{\delta+}$, a carbocation forms, then Br^- attacks it

Alkyl groups push electrons towards the positive carbon — this is the **inductive effect** 诱导效应. So a carbocation with more alkyl groups is more stable: tertiary is more stable than secondary, which is more stable than primary. When HBr adds to propene, the more stable carbocation forms, so hydrogen adds to the carbon that already has more hydrogens. This pattern is **Markovnikov's rule** 马氏规则.



Markovnikov: HBr adds so the *more stable* carbocation forms

Carbocation stability rises from primary to tertiary as more alkyl groups push electrons in — the basis of Markovnikov's rule

Worked example. Predict the major product when HBr adds to propene, $CH_3CH{=}CH_2$. The H^+ adds first, and it adds in whichever way makes the **more stable carbocation**. Adding the H to the **end** carbon puts the positive charge on the middle carbon, giving a **secondary** carbocation, which is stabilised by electron-releasing alkyl groups on **two** sides. Adding it to the middle carbon would leave a less stable **primary** carbocation. The bromide ion then attacks the secondary

carbocation, so the major product is **2-bromopropane**. That is Markovnikov's rule - but quote the **reason** (carbocation stability: tertiary > secondary > primary), because the rule on its own is not the explanation.

Exam tips

- **Alkanes**: free-radical substitution needs **UV light** — show initiation, propagation and termination with the radical dots.
- **Alkenes**: electrophilic addition via a **carbocation**; draw curly arrows from the C = C.
- **Markovnikov**: H adds to the carbon with more H's, via the **more stable carbocation** (alkyl groups are electron-releasing).
- Test for C = C: **bromine water decolourises** (orange → colourless) — state the observation.

Halogen compounds

A-Level Chemistry

Halogenoalkanes

A **halogenoalkane** 卤代烷 is an alkane with one or more halogen atoms in place of hydrogen (a C–X bond, where X is a halogen).



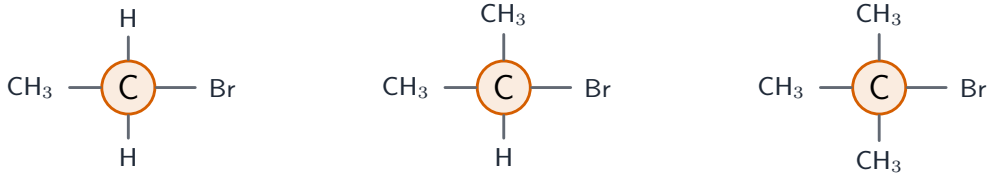
Volatile halogenoalkanes were once widely used as refrigerants and aerosol propellants

Making halogenoalkanes

- **free-radical substitution** 自由基取代 of an alkane with Cl_2 or Br_2 in ultraviolet light.
- **electrophilic addition** 亲电加成 of an alkene with a halogen X_2 or a hydrogen halide HX.
- substitution of an **alcohol** 醇, for example by HX, by PCl_5, by PCl_3 with heat, or by $SOCl_2$.

Three classes

A halogenoalkane is **primary** 伯, **secondary** 仲 or **tertiary** 叔, depending on how many carbon atoms are joined to the carbon that holds the halogen (one, two or three).



primary
1 C on the C–Br carbon

secondary
2 C

tertiary
3 C

Primary, secondary and tertiary halogenoalkanes, set by how many carbons are joined to the carbon bearing the halogen

Nucleophilic substitution

The C–X bond is polar, so the carbon is slightly positive. A **nucleophilic substitution** 亲核取代 happens when a **nucleophile** 亲核试剂 (a lone-pair species) attacks that carbon and replaces the halogen.



Many halogenoalkane reactions are carried out by heating the mixture under reflux

Reagent and conditions	Product
NaOH(aq), heat	an alcohol
KCN in ethanol, heat	a **nitrile** 腈 (adds one carbon to the chain)
NH_3 in ethanol, heated under pressure	an **amine** 胺

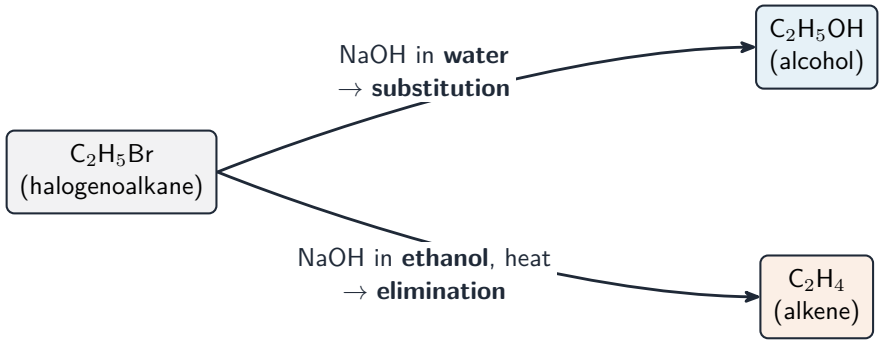
To **identify the halogen**, warm the halogenoalkane with **silver nitrate** 硝酸银 in ethanol. A silver halide **precipitate** 沉淀 forms, and its colour shows which halogen is present (white AgCl, cream AgBr, yellow AgI).

Elimination

The same halogenoalkane can instead undergo **elimination** 消去 to form an **alkene** 烯烃. The conditions decide which reaction wins:

- NaOH in **water** → nucleophilic substitution → an alcohol.
- NaOH in **ethanol**, heated → elimination → an alkene.

$$C_2H_5Br + NaOH \rightarrow C_2H_4 + NaBr + H_2O$$



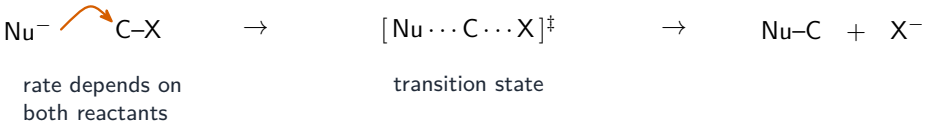
The same halogenoalkane: NaOH in water substitutes to an alcohol, while NaOH in ethanol (with heat) eliminates to an alkene

The S_N1 and S_N2 mechanisms

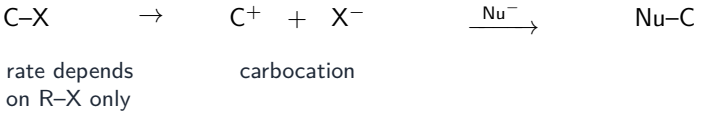
Nucleophilic substitution can follow two routes:

- **S_N2**: one step. The nucleophile attacks at the same time as the halogen leaves, passing through a crowded **transition state** 过渡态 where both are half-bonded. The rate depends on both the halogenoalkane and the nucleophile.
- **S_N1**: two steps. First the C–X bond breaks to give a **carbocation** 碳正离子; then the nucleophile attacks it. The rate depends only on the halogenoalkane.

S_N2: one step



S_N1: two steps



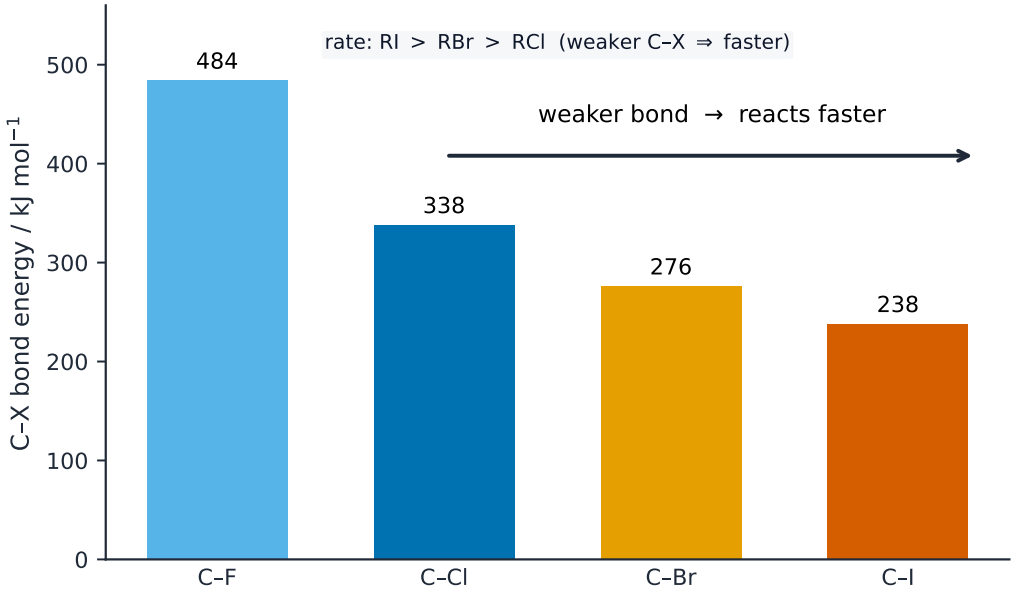
S_N2 is a single step through a crowded transition state; S_N1 is two steps via a carbocation

Alkyl groups push electrons towards the positive carbon (the **inductive effect** 诱导效应), so they stabilise the carbocation. This is why:

- **primary** halogenoalkanes mostly react by S_N2.
- **tertiary** halogenoalkanes mostly react by S_N1 (their carbocation is well stabilised).
- **secondary** halogenoalkanes use a mixture of the two.

Different reactivities

How fast a halogenoalkane reacts depends on the strength of the C–X bond, measured by its **bond energy** 键能. The C–I bond is the weakest, so iodoalkanes react fastest; the C–Cl bond is the strongest of the three, so chloroalkanes react slowest. So when tested with silver nitrate, an iodoalkane gives its precipitate first — its higher **reactivity** 反应活性 comes from the weaker C–X bond.



The weaker the C–X bond, the faster the halogenoalkane reacts — so iodoalkanes react fastest and chloroalkanes slowest

Worked example. Predict the mechanism for the hydrolysis of 1-bromobutane and of 2-bromo-2-methylpropane. Classify the halogenoalkane first. 1-bromobutane is **primary**: the carbon carrying the Br is barely shielded, so the nucleophile can attack the back of it and the mechanism is S_N2 - one step, with the rate depending on **both** the halogenoalkane and the nucleophile. 2-bromo-2-methylpropane is **tertiary**: three bulky methyl groups block that attack, but they also stabilise the carbocation formed once the Br leaves, so it goes S_N1 - two steps, with the rate depending on the halogenoalkane **only**. Decide from the **class** (primary → S_N2, tertiary → S_N1), and note that the tertiary one hydrolyses **faster** despite being the more crowded.

Exam tips

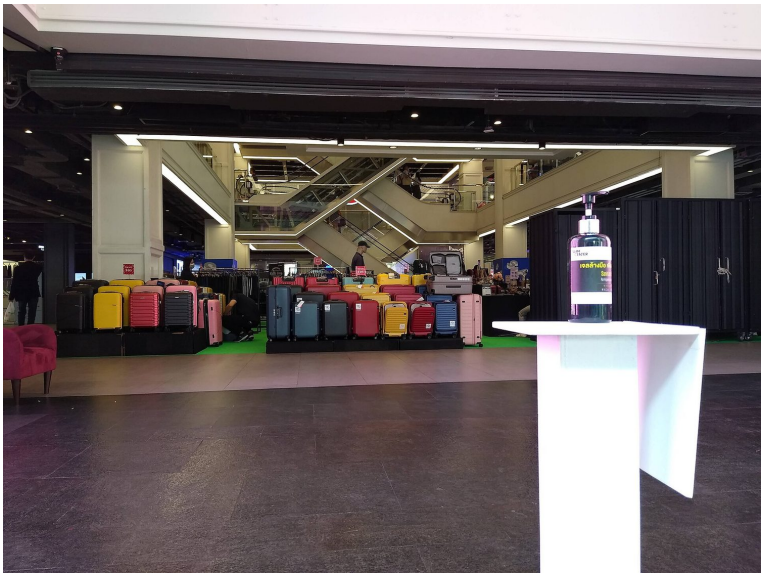
- Draw **nucleophilic substitution** with a curly arrow from the nucleophile's lone pair to the $\delta+$ carbon and one from the C-X bond.
- Match mechanism to class: $S_{\mathbf{N}}1$ (tertiary, carbocation, two steps) vs $S_{\mathbf{N}}2$ (primary, one step, transition state).
- Reactivity is set by **bond enthalpy**: C-I is weakest, so iodoalkanes react fastest (not electronegativity).
- **Elimination** (hot, ethanolic KOH) vs **substitution** (warm, aqueous KOH) — the conditions decide the product; state them.

Hydroxy compounds

A-Level Chemistry

Alcohols

An **alcohol** 醇 has the –OH (hydroxyl) functional group.



Ethanol, the alcohol in hand sanitiser, kills microbes

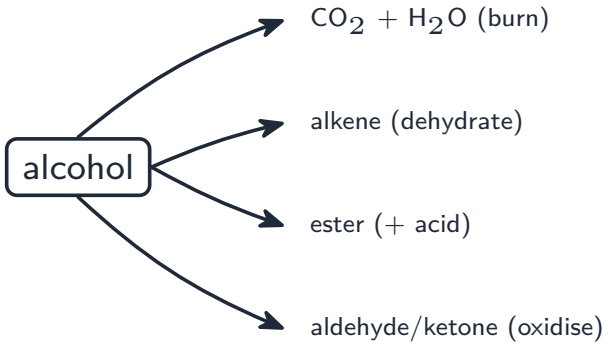
Making alcohols

Method	Reagents and conditions
addition of steam to an alkene	$H_2O(g)$, H_3PO_4 catalyst (**electrophilic addition** 亲电加成)
an **alkene** 烯烃 with cold dilute $KMnO_4$	gives a **diol** 二醇 (two –OH groups)
substitution of a **halogenoalkane** 卤代烷	NaOH(aq), heat
reduction 还原 of an **aldehyde** 醛 or **ketone** 酮	$NaBH_4$ or $LiAlH_4$
reduction of a **carboxylic acid** 羧酸	$LiAlH_4$
hydrolysis 水解 of an **ester** 酯	dilute acid or alkali, heat

Reactions of alcohols

- **combustion**: alcohols burn in oxygen to give carbon dioxide and water.
- **substitution** to a halogenoalkane, for example with HX, PCl_5, PCl_3 and heat, or $SOCl_2$.
- **with sodium**: alcohols react with sodium metal to give hydrogen and a sodium alkoxide — like water, but more slowly.

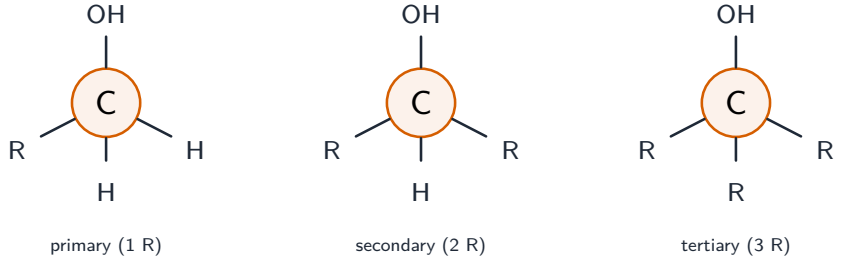
- **oxidation** 氧化 with acidified $K_2Cr_2O_7$ (or $KMnO_4$). The product depends on the class of alcohol (see below).
- **dehydration** 脱水 to an alkene, using a hot Al_2O_3 catalyst or concentrated acid.
- **ester formation**: an alcohol reacts with a carboxylic acid (with concentrated H_2SO_4 catalyst) to make an ester.



An alcohol can burn, dehydrate, esterify or be oxidised

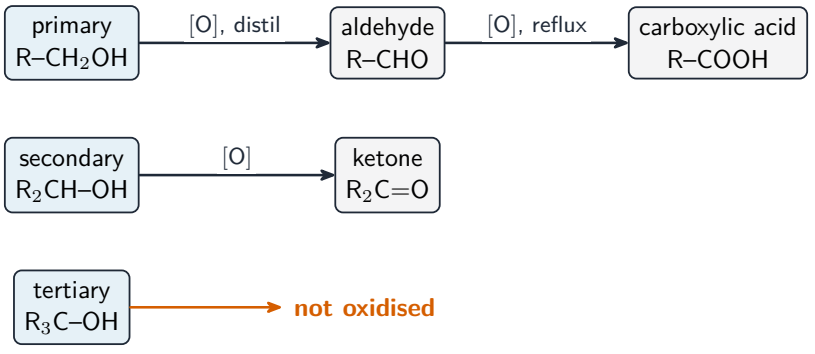
Three classes and how oxidation tells them apart

An alcohol is **primary** 伯, **secondary** 仲 or **tertiary** 叔, depending on how many carbons are joined to the carbon holding the –OH. Some molecules have more than one –OH group.



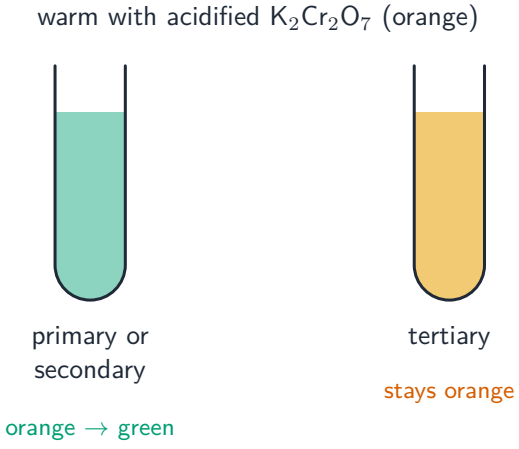
The class is set by the carbon holding the –OH: one R group makes it primary, two secondary, three tertiary — which decides how it oxidises

Class	Oxidation product
primary	**aldehyde** (by **distillation** 蒸馏), then **carboxylic acid** (by **reflux** 回流)
secondary	a **ketone**
tertiary	not oxidised



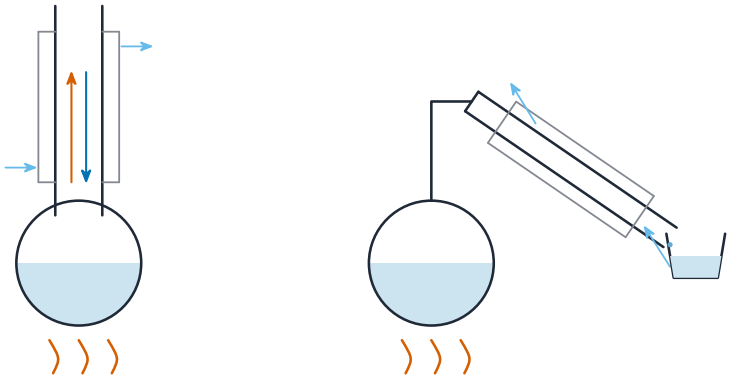
Oxidation by class: a primary alcohol gives an aldehyde then a carboxylic acid, a secondary gives a ketone, a tertiary is not oxidised

In a quick test, acidified $K_2Cr_2O_7$ turns from **orange to green** with a primary or secondary alcohol, but stays orange with a tertiary alcohol.



Acidified dichromate turns orange to green with a primary or secondary alcohol, but stays orange with a tertiary alcohol

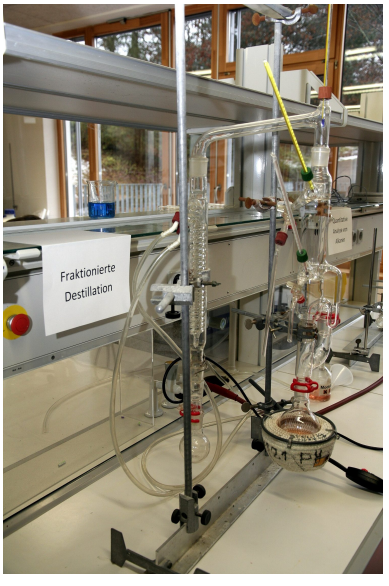
- **distillation** removes the aldehyde as it forms, before it can be oxidised further.
- **reflux** keeps boiling the mixture and returning the vapour, so the alcohol is fully oxidised to the carboxylic acid.



reflux
vapour returns: fully oxidised to the acid

distillation
product distils off as it forms (the aldehyde)

Distillation removes the aldehyde as it forms; reflux keeps boiling and returning the vapour, fully oxidising to the acid



A real distillation set-up: the vapour boils off, cools in the condenser and is collected — the same idea separates ethanol from a fermented mixture

The iodoform test

If you warm an alcohol that contains the $CH_3CH(OH)-$ group with alkaline aqueous iodine, you get a pale yellow precipitate of **tri-iodomethane** 三碘甲烷 (CHI_3) and the ion RCO_2^-. This is a useful test for that group.

Acidity of alcohols

The –OH group makes alcohols very weakly acidic: they can lose the H^+ to form an RO^- ion. But their **acidity** 酸性 is **lower** than that of water. This is because the alkyl group pushes electron density onto the oxygen, which makes the RO^- ion less stable, so the alcohol holds onto its H^+ more tightly.

Worked example. Three unlabelled bottles hold butan-1-ol, butan-2-ol and 2-methylpropan-2-ol. How does warming each with acidified $K_2Cr_2O_7$ tell them apart? What matters is how many hydrogens sit on the **carbon carrying the OH**. Butan-1-ol is **primary** (two such hydrogens): the orange dichromate turns **green**, and by distilling you collect an **aldehyde** (butanal), or by refluxing you get the **carboxylic acid** (butanoic acid). Butan-2-ol is **secondary** (one such hydrogen): also **green**, but the product is a **ketone** (butanone), which will not oxidise further. 2-methylpropan-2-ol is **tertiary** (**no** such hydrogen): there is nothing to remove, so the dichromate stays **orange**. The colour only separates the tertiary from the other two - to split primary from secondary you must identify the **product** (an aldehyde gives a silver mirror with Tollens', a ketone does not).

Exam tips

- Classify the alcohol as **primary, secondary or tertiary** first — it decides the oxidation product.
- **Oxidation** with acidified $K_2Cr_2O_7$ (orange → green): primary → aldehyde (distil) → acid (reflux); secondary → ketone; tertiary → no reaction.
- Distinguish the conditions for **aldehyde vs acid** from a primary alcohol (distillation vs reflux).
- Name esters correctly (the acid part comes second, ending "-oate").

Carbonyl compounds

A-Level Chemistry

Aldehydes and ketones

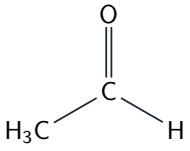
Aldehydes and ketones are **carbonyl compounds** 羰基化合物 — they contain the C=O **carbonyl** 羰基 group.

- in an **aldehyde** 醛 the carbonyl carbon is on the end of the chain (it also carries an H), written –CHO.
- in a **ketone** 酮 the carbonyl carbon is in the middle, between two other carbons.



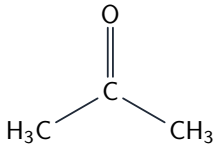
Propanone (acetone), the simplest ketone, is the solvent in nail-polish remover

aldehyde



C=O carbon on the end
(also carries an H): –CHO

ketone



C=O carbon in the middle,
between two carbons

Both have the C=O carbonyl group: in an aldehyde it is on the end of the chain (–CHO), in a ketone it is in the middle

Making aldehydes and ketones

Both are made by the **oxidation** 氧化 of an **alcohol** 醇 with acidified $K_2Cr_2O_7$ or $KMnO_4$:

- a **primary** alcohol, with **distillation** 蒸馏, gives an aldehyde.
- a **secondary** alcohol gives a ketone.

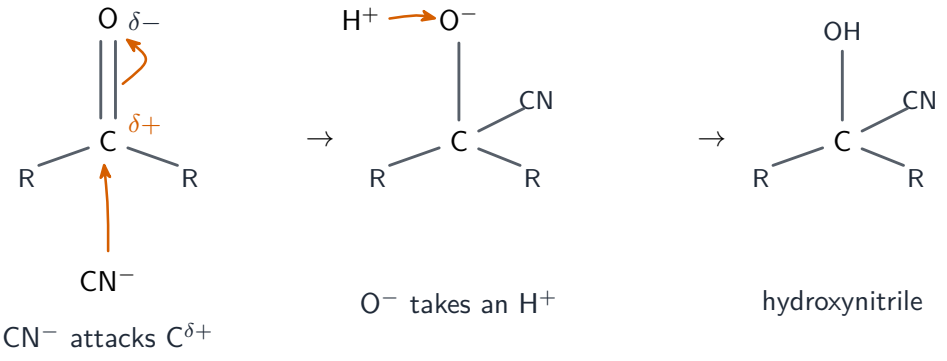
Reactions

- **reduction** 还原 with $NaBH_4$ or $LiAlH_4$ turns a carbonyl compound back into an alcohol (an aldehyde gives a primary alcohol; a ketone gives a secondary alcohol).
- reaction with **hydrogen cyanide** 氰化氢 (HCN), with KCN as catalyst and heat, adds H and CN across the C=O to make a **hydroxynitrile** 羟基腈. This adds one carbon to the chain.

The mechanism: nucleophilic addition

The reaction with HCN is a **nucleophilic addition** 亲核加成. The carbonyl carbon is slightly positive (oxygen pulls the electrons away). So:

1. the CN^- ion (a nucleophile) attacks the slightly positive carbon.
2. this breaks the C=O double bond, leaving a negative oxygen (O^-).
3. the O^- takes an H^+ (from HCN) to finish the hydroxynitrile.



Nucleophilic addition of HCN: CN^- attacks the $\delta+$ carbonyl carbon, the C=O breaks to O^-, then O^- takes an H^+

Tests for carbonyl compounds

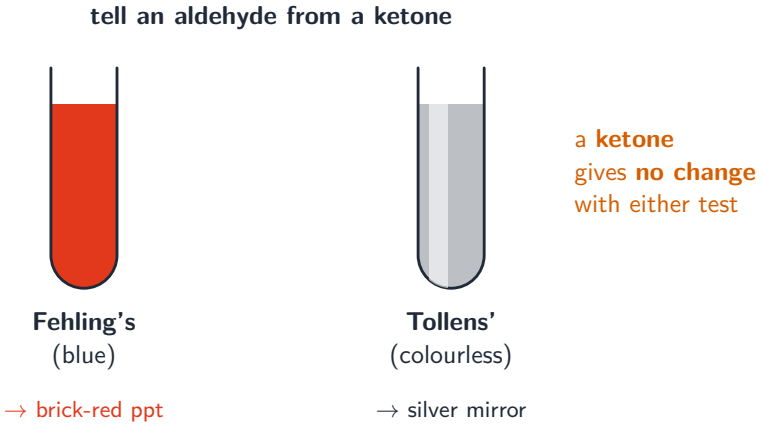
Detecting any carbonyl

Add **2,4-DNPH** reagent (2,4-dinitrophenylhydrazine). An orange precipitate confirms that the compound is an aldehyde or a ketone.

Telling an aldehyde from a ketone

Aldehydes are easily oxidised to carboxylic acids, but ketones are not. Two tests use this difference:

Test	Aldehyde	Ketone
Fehling's reagent 斐林试剂 (blue solution)	turns to a brick-red precipitate	no change
Tollens' reagent 托伦试剂 (colourless)	gives a silver mirror	no change



Telling an aldehyde from a ketone: an aldehyde gives a brick-red precipitate with Fehling's and a silver mirror with Tollens'; a ketone gives no change



A positive Tollens' test: an aldehyde coats the tube with a shiny silver mirror

The iodoform test

If the compound has the CH_3CO- group, warming it with alkaline aqueous iodine gives a pale yellow precipitate of **tri-iodomethane** 三碘甲烷 (CHI_3) and the ion RCO_2^-.

Worked example. A liquid gives an orange precipitate with 2,4-DNPH, gives **no** silver mirror with Tollens' reagent, and gives a yellow precipitate with alkaline aqueous iodine. Identify it. Take the tests one at a time, using each for exactly what it proves. The orange precipitate with **2,4-DNPH** proves a **carbonyl** group is present - an aldehyde or a ketone, nothing else. **No** silver mirror with Tollens' rules out an **aldehyde**, so it must be a **ketone**. The yellow precipitate in the **iodoform** test proves a CH_3CO- group next to the carbonyl. The simplest compound satisfying all three is **propanone**, CH_3COCH_3. Keep the roles straight: 2,4-DNPH finds **any** carbonyl, Tollens' separates aldehyde from ketone, and iodoform detects the **methyl group beside** the C=O.

Exam tips

- **2,4-DNPH** gives an orange precipitate with **any** carbonyl — it tests for C = O.
- **Tollens' (silver mirror)** and **Fehling's (brick-red)** are positive for **aldehydes only** — this is how you tell aldehydes from ketones.
- Nucleophilic addition of HCN adds **one carbon** ($\rightarrow$ hydroxynitrile); show the curly arrows.
- The **iodoform (tri-iodomethane)** test is positive for CH_3CO- groups.

Carboxylic acids and derivatives

A-Level Chemistry

Carboxylic acids

A **carboxylic acid** 羧酸 has the –COOH (carboxyl) functional group. It is a weak acid.



Vinegar is a dilute solution of ethanoic acid, a carboxylic acid

Making carboxylic acids

- **oxidation** 氧化 of a primary **alcohol** 醇 or an **aldehyde** 醛 with acidified $K_2Cr_2O_7$ or $KMnO_4$, with **reflux** 回流 (so it is fully oxidised).
- **hydrolysis** 水解 of a **nitrile** 腈 with dilute acid or alkali, then acidifying.
- hydrolysis of an **ester** 酯 with dilute acid or alkali and heat, then acidifying.

Reactions of carboxylic acids

These reactions all show that carboxylic acids are acids:

- **with a reactive metal** (a **redox** 氧化还原 reaction): gives a **salt** 盐 and hydrogen.

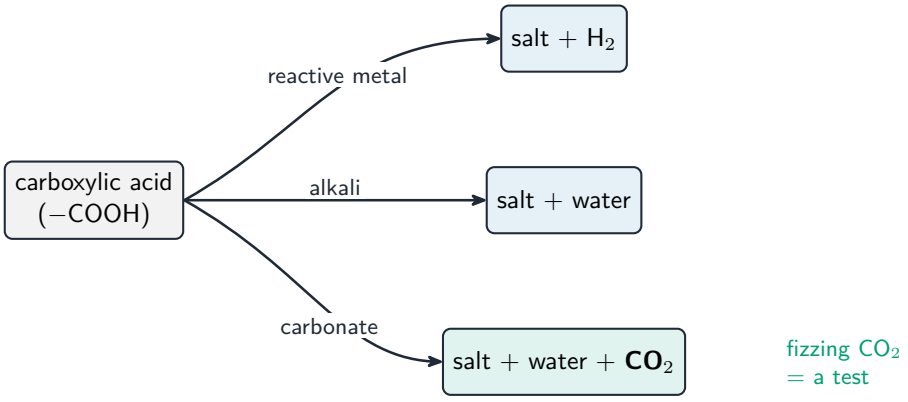
$$2CH_3COOH + Mg \rightarrow (CH_3COO)_2Mg + H_2$$

- **with an alkali** (a **neutralisation** 中和): gives a salt and water.

$$CH_3COOH + NaOH \rightarrow CH_3COONa + H_2O$$

- **with a carbonate** 碳酸盐: gives a salt, water and carbon dioxide. The fizzing of CO_2 is a test for a carboxylic acid.

$$2CH_3COOH + Na_2CO_3 \rightarrow 2CH_3COONa + H_2O + CO_2$$



Carboxylic acids behave as acids: a salt with a metal (+ H_2), with an alkali (+ water), or with a carbonate (+ water + CO_2 — the fizz is a test)

Two reactions change the functional group:

- **esterification** 酯化 with an alcohol (concentrated H_2SO_4 catalyst) gives an ester.
- **reduction** 还原 by $LiAlH_4$ gives a primary alcohol.

Esters

An ester has the –COO– group. It often smells sweet or fruity.



Esters give many fruits and perfumes their sweet smell

Making esters

An ester forms in a **condensation** 缩合 reaction between an alcohol and a carboxylic acid, with concentrated H_2SO_4 as catalyst. A water molecule is lost, and the reaction is reversible:

$$CH_3COOH + C_2H_5OH \rightleftharpoons CH_3COOC_2H_5 + H_2O$$

$$CH_3COOH \quad + \quad HOC_2H_5 \quad \rightleftharpoons \quad CH_3CO\ O\ C_2H_5 \quad + \quad H_2O$$

the –OH (acid) and –H (alcohol) leave as water; the C–O–C ester link forms

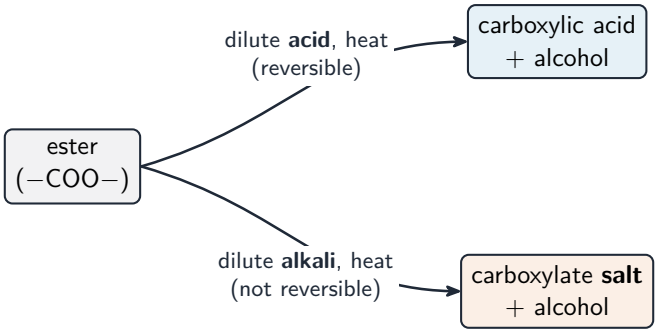
(concentrated H_2SO_4 catalyst; reversible)

Esterification is a condensation: the –OH from the acid and the –H from the alcohol leave as water, forming the C–O–C ester link

Hydrolysis of esters

Hydrolysis splits the ester back apart. The conditions change the products:

- **dilute acid** and heat: reversible. Gives back the carboxylic acid and the alcohol.
- **dilute alkali** and heat: not reversible. Gives the alcohol and the **salt** of the carboxylic acid (the carboxylate ion).



Hydrolysing an ester: dilute acid (reversible) gives the carboxylic acid and alcohol; dilute alkali (not reversible) gives the carboxylate salt and alcohol

Worked example. Name the ester made from ethanol and propanoic acid, and say which part comes from which. In an ester name the **alcohol** gives the first word (the alkyl part) and the **acid** gives the second (the -oate part). Ethanol supplies C_2H_5- and propanoic acid supplies CH_3CH_2COO-, so the ester is **ethyl propanoate**, $CH_3CH_2COOC_2H_5$. It is made by refluxing the two with a concentrated H_2SO_4 catalyst, and the reaction is **reversible**, so the yield is never complete. The trap is naming it backwards: **propyl ethanoate** is a completely different ester (made from propan-1-ol and ethanoic acid). Alcohol first, acid second.

Exam tips

- Carboxylic acids react with carbonates to give $\mathbf{CO_2}$ (fizzing) — this distinguishes them from phenols.
- Explain acid strength via the **delocalised carboxylate ion**; electron-withdrawing groups (e.g. Cl) increase strength.
- **Esterification** (with concentrated H_2SO_4) is reversible; name esters as *alkyl alkanoate*.
- Know the products of **acid vs alkaline** hydrolysis of an ester (they differ).

Nitrogen compounds

A-Level Chemistry

Primary amines

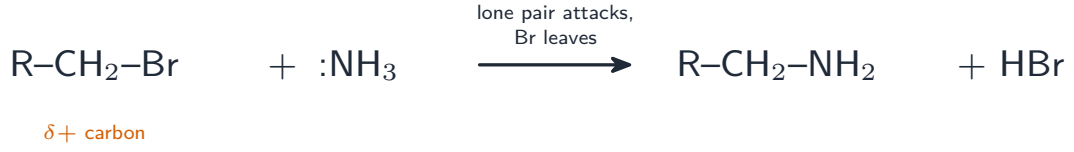
An **amine** 胺 has an $-NH_2$ group (the nitrogen replaces a hydrogen of ammonia).

Making a primary amine

Heat a **halogenoalkane** 卤代烷 with ammonia dissolved in ethanol, under pressure:

$$C_2H_5Br + NH_3 \rightarrow C_2H_5NH_2 + HBr$$

This is a **nucleophilic substitution** 亲核取代: the lone pair on the nitrogen of ammonia attacks the slightly positive carbon and pushes out the halogen. You use an excess of ammonia, or the amine made can react again.



Making a primary amine: ammonia's lone pair attacks the δ+ carbon and pushes out the halogen — a nucleophilic substitution

Nitriles and hydroxynitriles

Making a nitrile

Heat a halogenoalkane with potassium cyanide (KCN) in ethanol:

$$C_2H_5Br + KCN \rightarrow C_2H_5CN + KBr$$

This is also a nucleophilic substitution, with the CN^- ion as the nucleophile. It is useful because it **adds one carbon** to the chain. The product is a **nitrile** 腈.

Making a hydroxynitrile

Add HCN (with KCN as catalyst, and heat) to an **aldehyde** 醛 or **ketone** 酮. The H and CN add across the C=O bond to give a **hydroxynitrile** 羟基腈. The reagent is **hydrogen cyanide** 氰化氢, and the mechanism is **nucleophilic addition** 亲核加成.

$$R\text{–}CH_2\text{–}Br + CN^- \longrightarrow R\text{–}CH_2\text{–}CN$$

nucleophilic **substitution** (nitrile)

$$R_2C{=}O + HCN \longrightarrow R_2C(OH)(CN)$$

nucleophilic **addition** (hydroxynitrile)

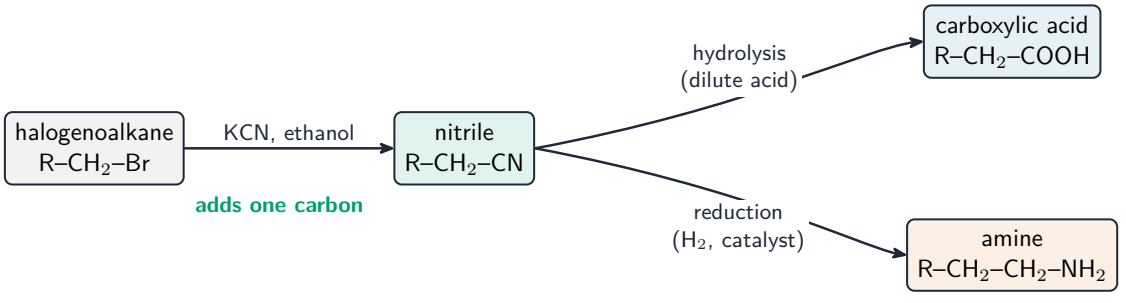
Two ways cyanide adds a carbon: KCN with a halogenoalkane substitutes to a nitrile; HCN with a carbonyl adds to a hydroxynitrile

Hydrolysis of nitriles

Warm a nitrile with dilute acid (or dilute alkali, then acidify). This **hydrolysis** 水解 turns the –CN group into a –COOH group, giving a **carboxylic acid** 羧酸:

$$CH_3CN + 2H_2O + HCl \rightarrow CH_3COOH + NH_4Cl$$

A nitrile can also be **reduced** by hydrogen and a catalyst to form an amine, which is the **reduction** 还原 route to a longer-chain amine.



Nitriles are a useful hub: KCN adds a carbon to make the nitrile, which then hydrolyses to a carboxylic acid or reduces to an amine



Reducing a nitrile gives an amine; amines and diacids link up to make polyamides such as nylon — the fibre first made famous in stockings

Worked example. Starting from bromoethane, make propanoic acid. Compare the carbons first: bromoethane has 2, propanoic acid has 3, so a carbon must be **added** - and the KCN step is the reaction that does it. Step 1: warm bromoethane with **ethanolic KCN**; nucleophilic substitution gives propanenitrile, CH_3CH_2CN, which now has 3 carbons because the CN carbon joins the chain. Step 2: reflux the nitrile with **dilute HCl**; hydrolysis gives propanoic acid. Count the carbons **before** you plan: whenever the target has exactly one more than the starting material, the nitrile route is almost always the intended answer, and remember the CN carbon is part of the new chain.

Exam tips

- **Amines are bases** (the N lone pair accepts H^+); aliphatic amines are stronger bases than ammonia.
- Making amines: **reduce a nitrile**, or react a halogenoalkane with excess ammonia.
- **KCN adds one carbon** (halogenoalkane → nitrile) — track the carbon count carefully in synthesis routes.
- State reagents and conditions precisely for each conversion.

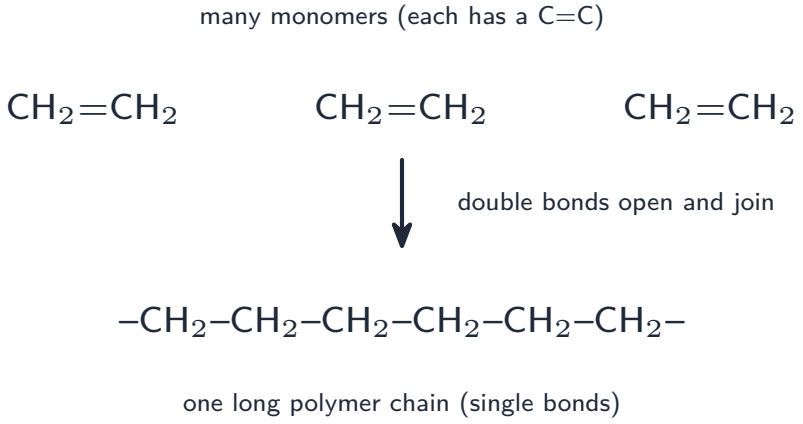
Polymerisation

A-Level Chemistry

Addition polymerisation

In **addition polymerisation** 加成聚合, many small molecules join into one very long chain, with no other product made.

Each small molecule is a **monomer** 单体. It must be unsaturated — it has a C=C double bond. The double bond opens up so that the monomers can link together. The long chain that forms is the **polymer** 聚合物.



Each monomer's double bond opens, and the monomers join end to end into one long chain — with no other product made



Poly(ethene), a common addition polymer, is supplied as tiny pellets a few millimetres across; these are later melted and moulded into bottles, bags and other products

Repeat units

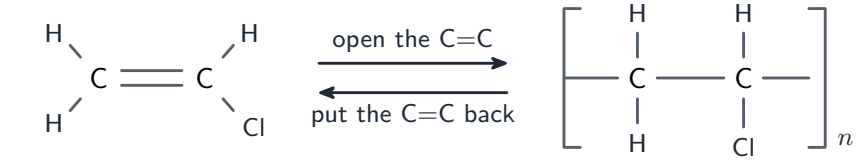
The **repeat unit** 重复单元 is the small part that is copied again and again along the chain. To find it, take the monomer, change the C=C to a single C–C, and draw bonds going out at each end.

- **poly(ethene)** 聚乙烯 is made from ethene:

$$n\,CH_2{=}CH_2 \rightarrow -(CH_2{-}CH_2)_n-$$

- **poly(chloroethene)** 聚氯乙烯 (PVC) is made from chloroethene:

$$n\,CH_2{=}CHCl \rightarrow -(CH_2{-}CHCl)_n-$$



chloroethene (monomer) poly(chloroethene), PVC

Finding the repeat unit: change the monomer's C=C to a single bond and draw bonds out at each end; reverse it to find the monomer

Finding the monomer

To go the other way, look at one repeat unit of the polymer, and put the C=C double bond back in. That gives you the monomer.

Worked example. A polymer has the repeat unit $-(CH_2-CH(CH_3))_n-$. Identify the monomer and name the polymer. Reverse the rule you used to build it: rub out the bonds sticking out at each end, and turn the single C-C **in the backbone** back into a **C=C**. That gives $CH_2=CH(CH_3)$, which is propene - so the polymer is **poly(propene)**. Two checks catch most errors: the monomer must have the **same molecular formula** as the repeat unit (addition polymerisation adds nothing and loses nothing), and the double bond goes back into the **backbone**, never into the side group.

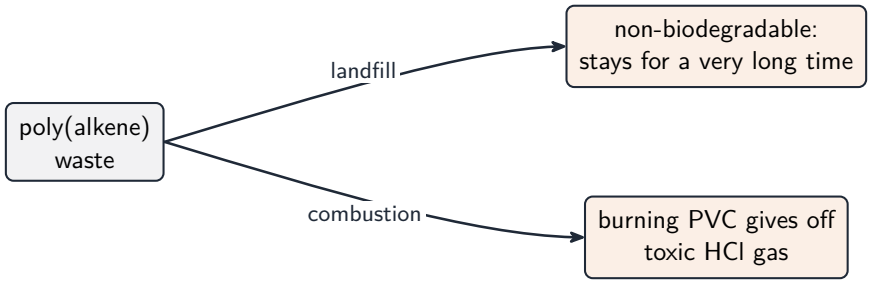
The problem of disposal

Poly(alkene)s are very hard to get rid of:

- they are **non-biodegradable** 不可生物降解 — microbes cannot break them down, so they stay in the ground for a very long time.
- their **combustion** 燃烧 (burning) can release harmful gases. For example, burning PVC gives off toxic hydrogen chloride.



Most addition polymers are non-biodegradable, so they build up as waste in the environment



Poly(alkene) waste is hard to dispose of: it is non-biodegradable, and burning PVC releases toxic hydrogen chloride

Exam tips

- Draw the **repeat unit** in brackets with the two bonds crossing them and n outside; keep the backbone carbons.
- Deduce the **monomer from a polymer** (and vice versa) — a very common question.
- Addition polymers are **inert and non-biodegradable** — link to landfill, recycling and incineration issues.
- Do not confuse with condensation (no small molecule is lost in addition polymerisation).

Organic synthesis

A-Level Chemistry

Organic synthesis

This topic does not add new reactions. Instead it asks you to **join up** the reactions you already know, so you can build a target molecule in several steps.



Medicines are built up from simple starting materials by multi-step organic synthesis

Identifying functional groups

A molecule may have more than one **functional group** 官能团. Use the test reactions from the syllabus to identify each one, and then predict how the molecule will behave. For example:

- decolourises bromine water → a C=C double bond (an **alkene** 烯烃).
- gives a precipitate with silver nitrate → a **halogenoalkane** 卤代烷.
- orange $K_2Cr_2O_7$ turns green → a primary or secondary **alcohol** 醇.
- orange precipitate with 2,4-DNPH → an **aldehyde** 醛 or **ketone** 酮.
- fizzes with a carbonate → a **carboxylic acid** 羧酸.

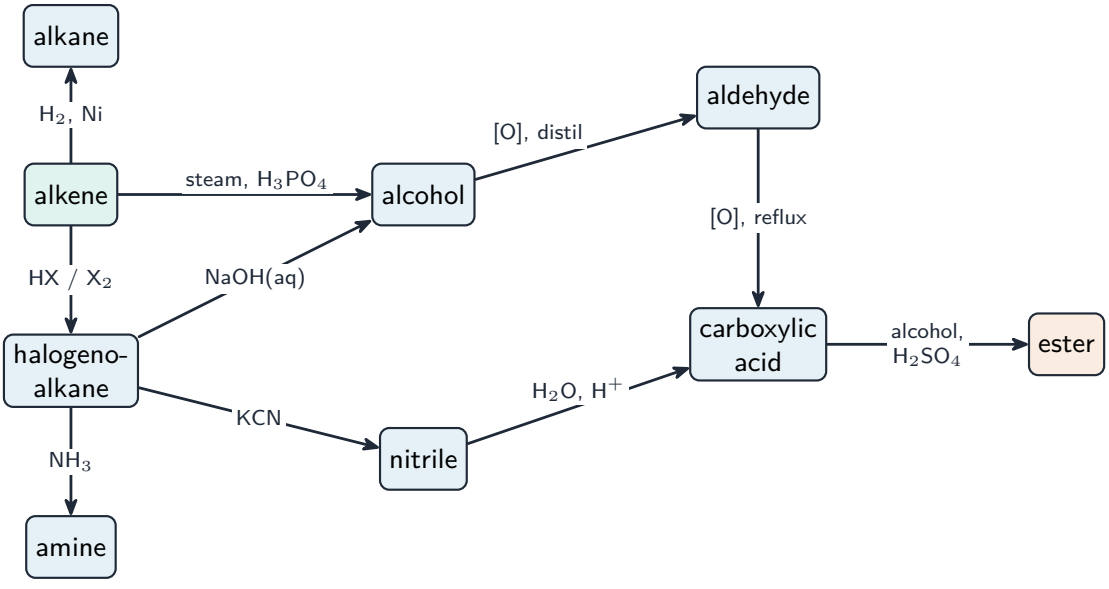
add this	you see	functional group
bromine water	orange → colourless	**C=C (alkene)**
silver nitrate	a precipitate forms	**halogenoalkane**
acidified $K_2Cr_2O_7$	orange → green	**1°/2° alcohol**
2,4-DNPH	orange precipitate	**aldehyde / ketone**
a carbonate	fizzes (CO_2 given off)	**carboxylic acid**

The standard identification tests: each reagent gives a characteristic observation that points to one functional group

A map of the AS reactions

Each row turns one functional group into another. Learn it as a map you can travel around:

Start	Reagent and conditions	Product
alkene	H_2, Ni	alkane
alkene	HX, or X_2	halogenoalkane
alkene	steam, H_3PO_4	alcohol
halogenoalkane	NaOH(aq), heat	alcohol
halogenoalkane	KCN in ethanol, heat	a **nitrile** 腈 (adds one carbon)
halogenoalkane	NH_3 in ethanol, pressure	an **amine** 胺
alcohol	$K_2Cr_2O_7$, distil / reflux	aldehyde / carboxylic acid
alcohol	concentrated acid, heat	alkene
aldehyde or ketone	$NaBH_4$	alcohol
aldehyde or ketone	HCN, KCN	hydroxynitrile
nitrile	dilute acid, heat	carboxylic acid
carboxylic acid + alcohol	concentrated H_2SO_4	an **ester** 酯



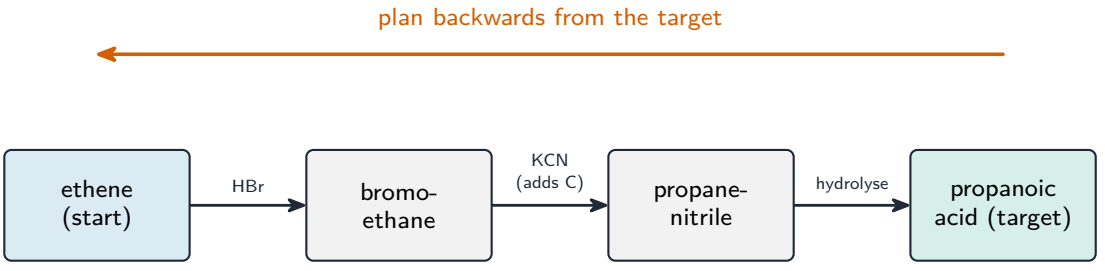
A map of the AS reactions: each arrow turns one functional group into another. Work backwards from your target to plan a route

Planning a multi-step route

To devise a **synthetic route** 合成路线:

1. compare the target with the starting material — what has changed (the functional group, the number of carbons)?
2. work **backwards** from the target: which single reaction could make it, and from what?
3. repeat until you reach the starting material.
4. write each step with its **reagent** 试剂 and conditions.

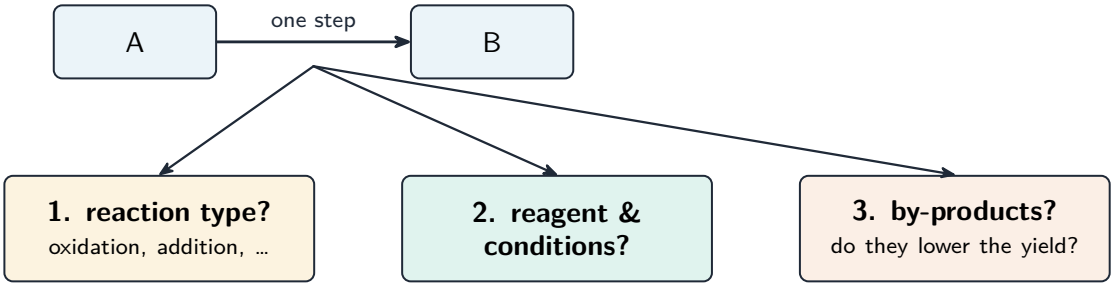
If you need to add a carbon, the KCN step is the key — it is the only AS reaction that lengthens the chain.



Plan a route by working backwards from the target, one reaction at a time, until you reach the starting material

Analysing a route

When you are given a route, for each step state the **type of reaction** (such as **oxidation** 氧化, **reduction** 还原, substitution, addition or elimination) and the reagent used. Also think about possible **by-products** 副产物 — for example, making an amine from a halogenoalkane also gives a mixture of further-substituted amines, so the yield of the simple amine is low.



e.g. a halogenoalkane → amine also makes further-substituted amines, so the simple-amine yield is low

For every step in a given route, ask its reaction type, reagent and by-products

name the reaction type at each step

oxidation	loses electrons / gains O or loses H (e.g. alcohol → acid)
reduction	gains electrons / loses O or gains H
substitution	one atom or group swaps for another
addition	two molecules join; a C=C double bond opens
elimination	a small molecule is removed; a C=C forms

also give the **reagent** for each step, and watch for **by-products** (e.g. amines over-substitute)

Name the reaction type at each step: oxidation, reduction, substitution, addition or elimination

Worked example. Devise a route from propene to propanone, CH_3COCH_3. Work **backwards** from the target. A ketone comes from oxidising a **secondary** alcohol, so the step before propanone is propan-2-ol with acidified $K_2Cr_2O_7$ under reflux. Propan-2-ol comes from propene by adding **steam** over an H_3PO_4 catalyst - and Markovnikov's rule conveniently puts the OH on the **middle** carbon, which is exactly the secondary alcohol needed. So the route is: propene, then steam with H_3PO_4, giving propan-2-ol; then acidified $K_2Cr_2O_7$ under reflux, giving propanone. Give a **reagent and**

its conditions on every arrow: a route with the right intermediates but no reagents scores very little.

Exam tips

- Learn the **reagents and conditions** for each conversion — that is exactly what synthesis questions test.
- Plan multi-step routes by functional group and watch the **carbon count** (KCN adds one carbon).
- Choose the **shortest valid route** and state every reagent and condition.

Analytical techniques

A-Level Chemistry

Infrared spectroscopy

Infrared spectroscopy 红外光谱 helps you find the **functional group** 官能团 in a molecule. Each kind of bond soaks up (absorbs) infrared radiation at its own range of frequencies. Where the bond shows strong **absorption** 吸收, the spectrum has a dip.

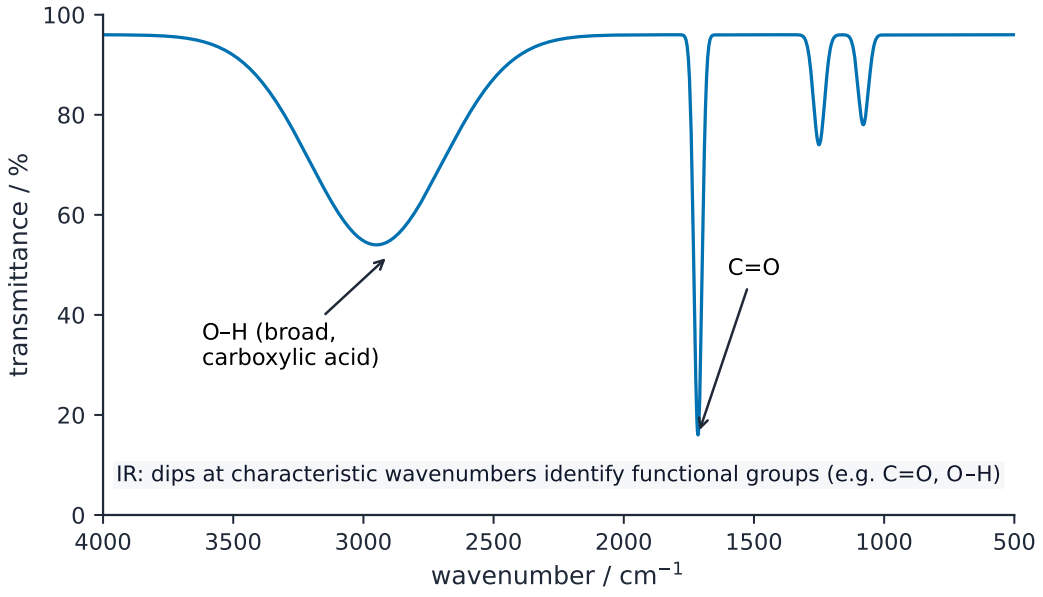


An infrared spectrometer shines infrared through a sample and records which wavenumbers its bonds absorb

The position is measured in **wavenumber** 波数 (in cm^{-1}). You are given a data table, so you do not memorise the numbers. You just match the dips to bonds:

- a broad dip around 3200–3650 cm^{-1} → an O–H bond in an alcohol.
- a dip around 1700 cm^{-1} → a C=O bond (aldehyde, ketone, acid or ester).
- a broad dip 2500–3000 cm^{-1} together with a C=O dip → a carboxylic acid.

This is useful for checking a reaction. For example, if propene has been turned into propan-2-ol, the C=C dip should be gone and an O–H dip should appear.



An infrared spectrum: each bond gives a dip at its own wavenumber. A broad O–H dip together with a C=O dip identifies a carboxylic acid

Mass spectrometry

In **mass spectrometry** 质谱, a molecule is turned into ions and sorted by its **mass-to-charge ratio** 质荷比 (m/e). The spectrum is a set of peaks at different m/e values.



A modern mass spectrometer: the sample is loaded at the front, then the machine ionises it and sorts the ions by their mass-to-charge ratio to give the spectrum

Relative atomic mass from isotopes

An element's **isotope** 同位素 mixture gives several peaks. From the **isotopic abundance** 同位素丰度 (how common each isotope is) you can find the **relative atomic mass** 相对原子质量 — a weighted average:

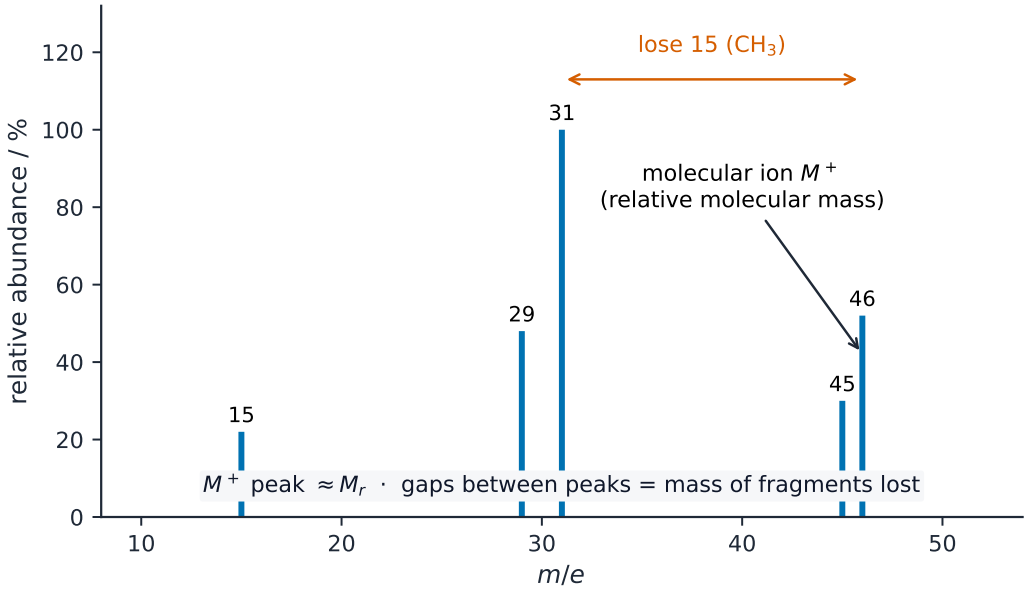
$$A_r = \frac{\sum(\text{isotope mass} \times \text{abundance})}{\sum \text{abundance}}$$

For example, chlorine is 75% ^{35}Cl and 25% ^{37}Cl, giving $A_r = \dfrac{35 \times 75 + 37 \times 25}{100} = 35.5$.

The molecular ion and fragmentation

The peak at the highest m/e (the **molecular ion** 分子离子 peak, or **molecular ion peak** 分子离子峰, M^+) gives the relative molecular mass of the whole molecule.

The molecule also breaks into smaller pieces — this is **fragmentation** 碎裂. The gap between two peaks tells you the mass of the lost piece, so you can suggest each **fragment** 碎片. For example, a loss of 15 means a CH_3 group was lost, and a loss of 29 means CHO or C_2H_5.



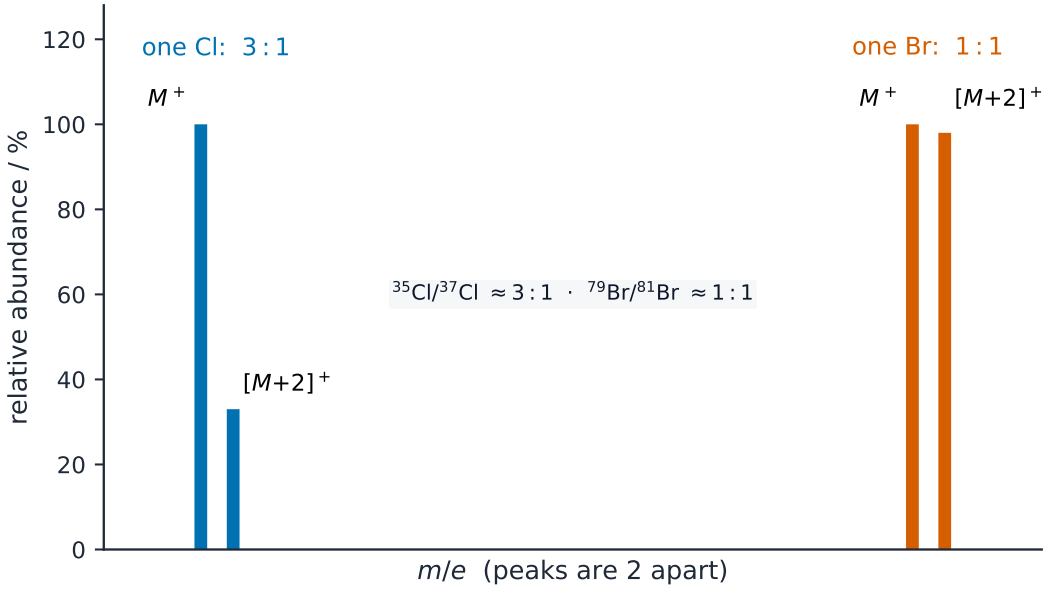
A mass spectrum: the highest-m/e peak is the molecular ion (M^+, the M_r); the gaps between peaks give the masses of the lost fragments

The [M + 1] and [M + 2] peaks

- a small **[M + 1]** peak comes from the ^{13}C isotope. The number of carbon atoms n is:

$$n = \frac{100 \times \text{abundance of } [M+1]^+}{1.1 \times \text{abundance of } M^+}$$

- an **[M + 2]** peak shows chlorine or bromine. One chlorine gives an $[M+2]$ peak about one third the height of M^+ (from ^{37}Cl); one bromine gives an $[M+2]$ peak about the **same** height as M^+ (from ^{81}Br).



An $[M+2]$ peak two mass units above M^+ shows a halogen: one chlorine gives a 3 : 1 ratio, one bromine a 1 : 1 ratio

Worked example. A compound shows a molecular ion at $m/e = 108$ and a peak of almost equal height at $m/e = 110$. Its infrared spectrum shows **no** broad absorption near 3300 cm^{-1}. Deduce its identity. Two peaks **two units apart with roughly equal heights** are the 1 : 1 signature of **one bromine** atom (from ^{79}Br and ^{81}Br); one chlorine would have given a 3 : 1 ratio instead. Take the bromine away from the molecular ion: $108 - 79 = 29$, which fits a C_2H_5 fragment. The absence of a broad peak near 3300 cm^{-1} rules out an O-H, so there is no alcohol group. The compound is **bromoethane**, C_2H_5Br. Read the $[M+2]$ **ratio** rather than merely noting the peak: 1 : 1 means bromine, 3 : 1 means chlorine.

Exam tips

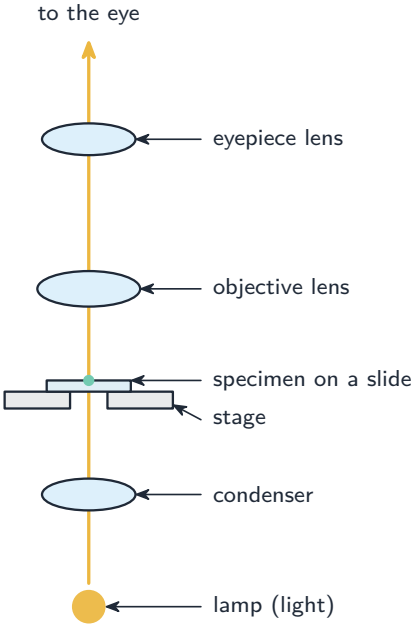
- On an **IR spectrum** quote the **wavenumber and the bond** (O-H broad $\sim 3200 - 3600$, C=O ~ 1700); use the data booklet.
- The peak at the **highest** m/z is the molecular ion M^+ and gives the M_r.
- Explain **fragmentation**: the gap between peaks is the lost fragment (loss of 15 = CH_3, 29 = C_2H_5 or CHO).
- Learn common **isotope patterns** (Cl gives M and M+2 in about 3:1; M+1 from ^{13}C).

Cell structure

A-Level Biology

How we study cells

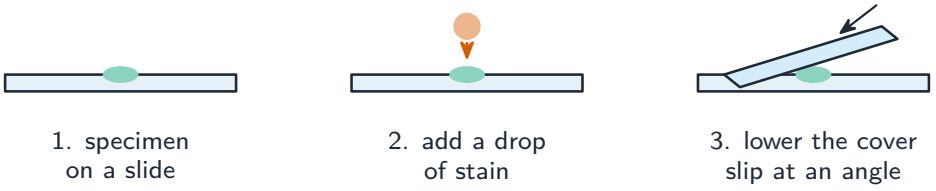
Cells 细胞 are very small, so you cannot see them with your eyes alone. You use a **microscope** 显微镜 to make a bigger picture of them. The first kind you meet is the **light microscope** 光学显微镜. It shines light through a thin **specimen** 标本 (the material you look at) and uses glass lenses to enlarge the view.



A light microscope: light passes up through the specimen, then two lenses (the objective and the eyepiece) magnify it

Making a slide and drawing what you see

To look at living material, you make a **temporary preparation** 临时装片. You put a small, thin piece of material on a glass **slide** 载玻片, add a drop of **stain** 染色剂 (a coloured liquid that makes parts easier to see), then lower a thin **cover slip** 盖玻片 on top to flatten it and keep out air.



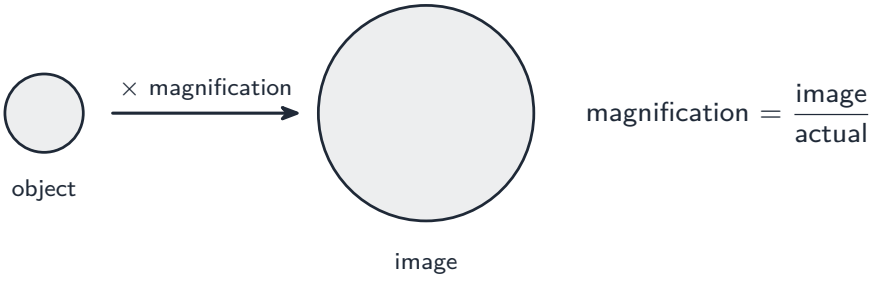
Making a temporary wet mount — lower the cover slip at an angle so no air bubbles are trapped

When you draw cells from a slide or a photograph, follow simple rules:

- use a sharp pencil and clear, single lines (no shading).
- draw only what you can really see, with the parts in the correct sizes.
- label the parts with straight lines that do not cross.

Magnification and actual size

Magnification 放大倍数 tells you how many times bigger the image is than the real object. It has no unit. You find it with one equation:



$$\textit{Magnification} = \textit{image size} \div \textit{actual size}$$

$$\text{magnification} = \frac{\text{size of image}}{\text{actual size of object}}$$

You can rearrange this to find any one value from the other two:

$$\text{actual size} = \frac{\text{size of image}}{\text{magnification}}$$

The top and the bottom of the fraction **must use the same unit**. Cells are tiny, so you work in small units:

- 1 mm = 1000 micrometre 微米 (µm)
- 1 µm = 1000 nanometre 纳米 (nm)
- so 1 mm = 1 000 000 nm.

Worked example. In a photomicrograph at magnification 5000, a **chloroplast** 叶绿体 measures 25 mm across. Its actual size is

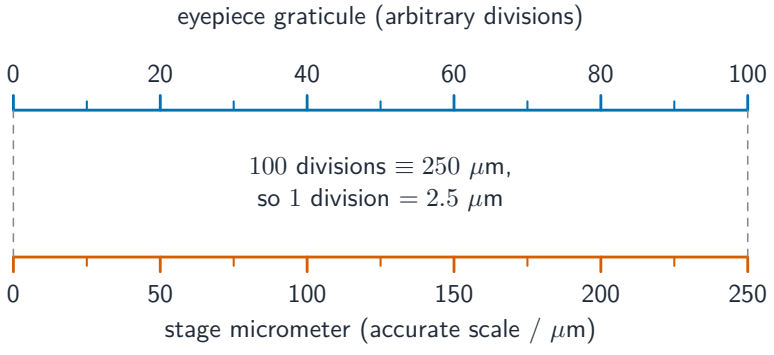
$$\frac{25 \text{ mm}}{5000} = 0.005 \text{ mm} = 5 \text{ µm}.$$

The same equation works for drawings, photomicrographs 显微照片 (photos taken through a light microscope) and **electron micrographs** 电子显微照片 (the most detailed photos, explained below). Always convert to the same unit first, then divide.

Eyepiece graticule and stage micrometer

To measure a real cell under the microscope, you use an **eyepiece graticule** 目镜测微尺 — a tiny scale inside the eyepiece. Its divisions have no fixed size, so first you must **calibrate** 校准 them (work out what one division is worth).

You calibrate using a **stage micrometer** 载物台测微尺 — a special slide with an accurate scale on it (often 1 mm split into 100 parts, so each part is 10 µm). You line up the two scales, count how many graticule divisions fit a known length, and divide. Once calibrated, you can swap in your specimen and measure it with the graticule.



Calibrate the graticule by lining it up with the stage micrometer's known scale

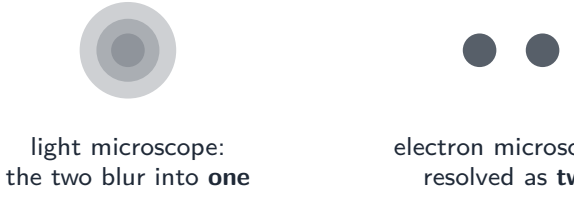
Resolution and magnification

These two words are easy to mix up. The examiner gives marks for the difference.

- **Magnification** is how many times bigger the image is than the object.
- **Resolution** 分辨率 is the smallest distance between two points that still lets you see them as **two** separate points.

Making an image bigger does not always show more detail. Past a certain point you just get a bigger, blurry image. Resolution sets the real limit on detail.

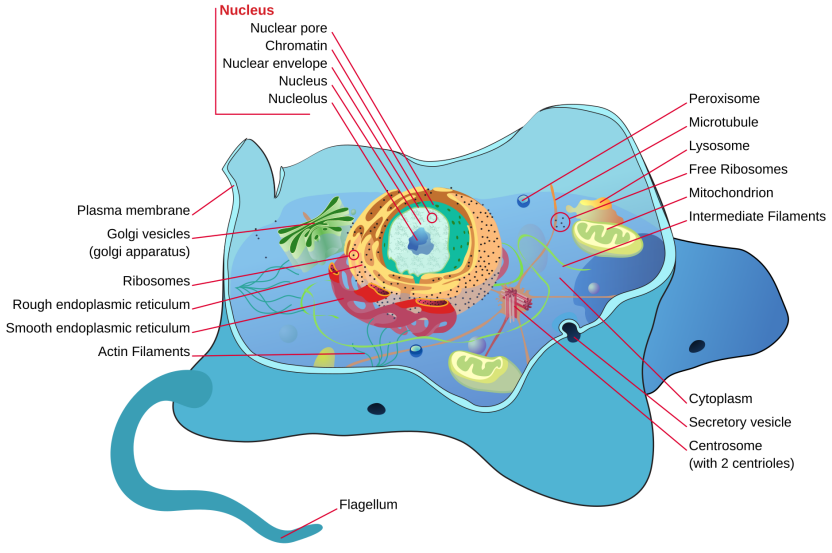
A light microscope has lower resolution because light has a fairly long **wavelength** 波长. An **electron microscope** 电子显微镜 uses beams of **electrons** 电子 instead of light. Electrons have a much shorter wavelength, so the resolution is far higher and you can see very small structures inside the cell. There are two kinds: **scanning** 扫描 (shows the surface in 3D) and **transmission** 透射 (passes electrons through a thin slice to show inside detail).



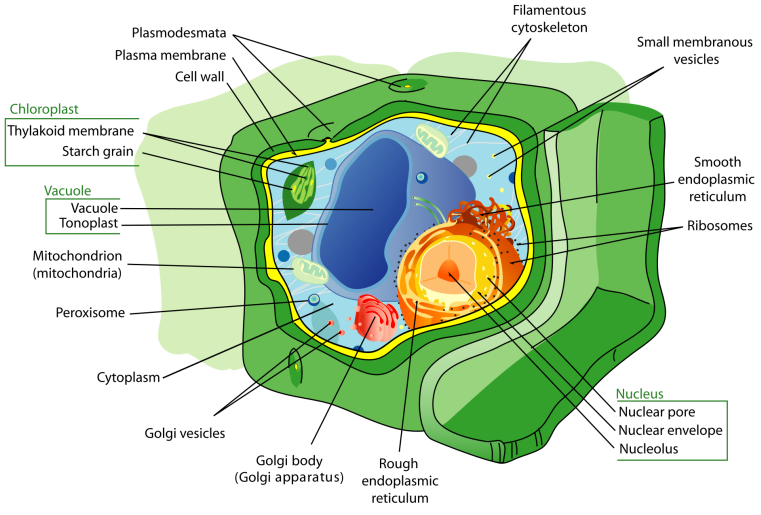
Resolution: a light microscope blurs two very close points into one; an electron microscope, with its shorter wavelength, resolves them as two

Eukaryotic cells and their organelles

Plant and animal cells are **eukaryotic cells** 真核细胞: their DNA is kept inside a **nucleus** 细胞核. Inside the cell are many small parts called **organelles** 细胞器, each with its own job. The jelly-like fluid around them is the **cytoplasm** 细胞质.



A generalised animal cell and its organelles



A plant cell also has a cell wall, chloroplasts and a large vacuole

In a photomicrograph or electron micrograph you identify organelles by their shape, size and position; in a drawing you show their outlines and label them.

Organelle	Structure	Function
cell surface membrane 细胞膜	thin layer around the cell	controls what enters and leaves the cell
nucleus	large, surrounded by a **nuclear envelope** 核膜 (a double membrane with holes); contains a **nucleolus** 核仁	holds the DNA; controls the cell; the nucleolus makes ribosomes
rough endoplasmic reticulum 粗面内质网 (rough ER)	sheets of membrane with ribosomes on the surface	makes and transports **proteins** 蛋白质 (for example **antibodies** 抗体)
smooth endoplasmic reticulum 滑面内质网 (smooth ER)	sheets of membrane, no ribosomes	makes **lipids** 脂质
Golgi body 高尔基体	stack of flat membrane sacs	changes and packs proteins and lipids into **vesicles** 囊泡 for **secretion** 分泌
mitochondria 线粒体	oval, with a folded inner membrane; has small circular DNA	site of **respiration** 呼吸作用 — releases energy 能量 as ATP
ribosomes 核糖体	very small; 80S in the cytoplasm, 70S in chloroplasts and mitochondria	join **amino acids** 氨基酸 to **synthesise** 合成 proteins
lysosomes 溶酶体	small sacs of **enzymes** 酶	break down old organelles and waste
centrioles 中心粒 and **microtubules** 微管	small tubes made of protein	help move chromosomes and form the cell's "skeleton"
cilia 纤毛	tiny hairs on the cell surface that beat	move fluid or move the cell
microvilli 微绒毛	tiny folds of the cell surface membrane	increase surface area for **absorption** 吸收
chloroplasts (plants)	green, with stacked membranes; has small circular DNA	site of **photosynthesis** 光合作用
cell wall 细胞壁 (plants)	strong outer layer of **cellulose** 纤维素	supports and protects the cell; stops it bursting
plasmodesmata 胞间连丝 (plants)	tiny channels through the cell walls	link the cytoplasm of neighbouring cells
large permanent **vacuole** 液泡 (plants)	big sac of watery fluid, with a membrane called the **tonoplast** 液泡膜	stores water and keeps the cell firm

Cells use **ATP** made in respiration as their energy supply for every job that needs energy, such as making proteins, moving things and dividing.

Comparing plant and animal cells

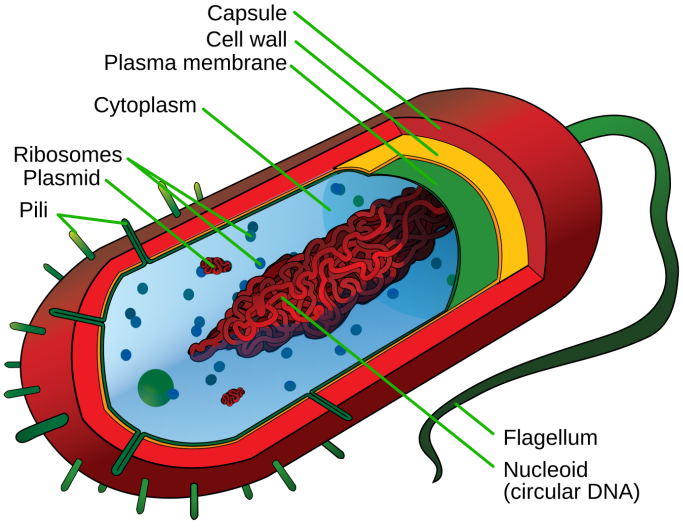
Feature	Plant cell	Animal cell
cell wall	present (cellulose)	absent
chloroplasts	present	absent
large permanent vacuole	present	absent (only small, temporary ones)
centrioles	absent in most	present
shape	fixed and regular	rounder and more flexible

Both have a cell surface membrane, cytoplasm, a nucleus, mitochondria, ribosomes, ER and a Golgi body.

Prokaryotic cells (bacteria)

A **prokaryotic cell** 原核细胞, such as a **bacterium** 细菌, is much smaller and simpler than a eukaryotic cell. Its key features are:

- **unicellular** 单细胞 — it is a single cell.
- generally 1–5 µm across.
- a cell wall made of **peptidoglycan** 肽聚糖 (not cellulose).
- circular DNA lying free in the cytoplasm — there is **no nucleus**.
- 70S ribosomes (smaller than the 80S ones in the cytoplasm of eukaryotes).
- **no organelles surrounded by a double membrane** — so no nucleus, no mitochondria and no chloroplasts.



A prokaryotic cell: the circular DNA (nucleoid) lies free, with no nucleus and no double-membrane organelles

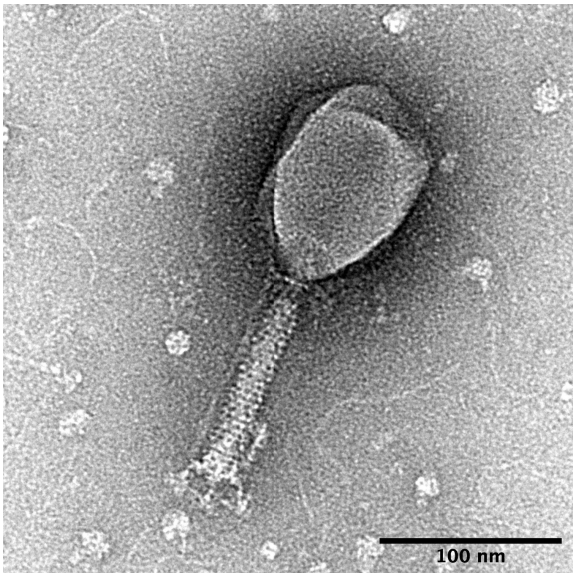
Comparing prokaryotic and eukaryotic cells

Feature	Prokaryotic cell	Eukaryotic cell
size	about 1–5 µm	about 10–100 µm
DNA	circular, free in cytoplasm	linear, inside a nucleus
nucleus	none	present
double-membrane organelles	none	mitochondria (and chloroplasts in plants)
ribosomes	70S	80S (with 70S inside mitochondria and chloroplasts)
cell wall	peptidoglycan	cellulose (plants) or none (animals)

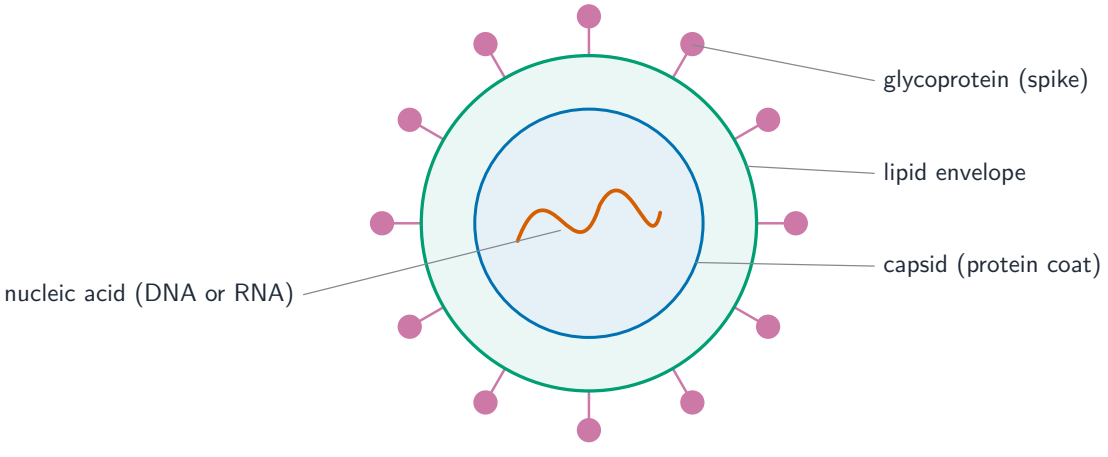
Viruses

All **viruses** 病毒 are **non-cellular** 非细胞 — they are not made of cells at all. Each virus is built from just two or three parts:

- a core of **nucleic acid** 核酸, which is **either DNA or RNA** (never both).
- a protein coat around the core called a **capsid** 衣壳.
- in some viruses, an outer **envelope** 包膜 made of **phospholipids** 磷脂.



A real virus, magnified hugely. The rounded head is the protein capsid wrapped around the nucleic acid core; the tail injects that nucleic acid into a bacterium. Note the scale bar — the whole virus is about 100 nm, far smaller than any cell



A generalised virus: nucleic acid inside a protein capsid, with a lipid envelope in some viruses

A virus has no cytoplasm, no organelles and no ribosomes. It cannot respire or make its own proteins. It can only copy itself inside a living **host** 宿主 cell, so it sits at the edge of what we call "living".

Exam tips

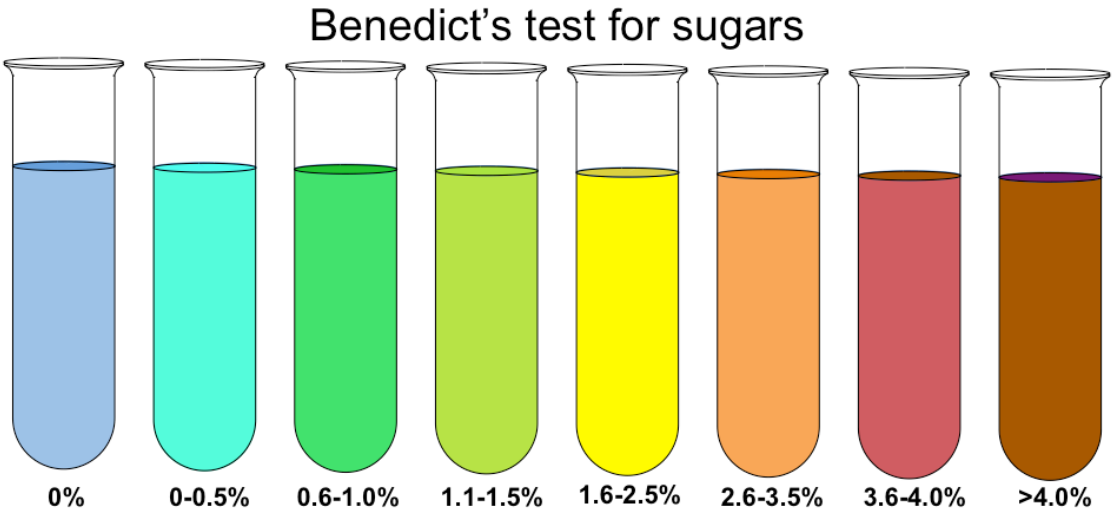
- Magnification = image size ÷ actual size — **convert units first** (mm ↔ μm ↔ nm), then rearrange for whichever is unknown.
- Distinguish **magnification** (how many times larger) from **resolution** (smallest distance still seen as two points); electron microscopes resolve more because electrons have a shorter wavelength.
- Give each organelle a **structure + function** pair (e.g. mitochondrion: folded inner membrane → aerobic respiration).
- State the **prokaryote vs eukaryote** differences exactly: no nucleus, smaller (70S) ribosomes, no membrane-bound organelles, circular DNA.
- Viruses are **non-living** — describe them only by capsid, genetic material and (sometimes) an envelope.

Biological molecules

A-Level Biology

Testing for biological molecules

Living things are built from four main kinds of large **molecule** 分子: **carbohydrates** 碳水化合物, **lipids** 脂质, **proteins** 蛋白质 and **nucleic acids** 核酸. You can use simple chemical tests to find out which kinds are present in a sample.



Benedict's test: blue turns green, then orange, then brick-red as more reducing sugar is present

Test	What it finds	Method	Positive result
Benedict's test	**reducing sugar** 还原糖	add Benedict's solution and heat in a water bath	blue changes to green, yellow, orange, then brick-red **precipitate** 沉淀
iodine test	**starch** 淀粉	add orange-brown **iodine** 碘 solution	colour changes to blue-black
emulsion test	**lipid**	mix sample with ethanol, then pour into water	a white, cloudy **emulsion** 乳浊液 forms
biuret 双缩脲 test	**protein**	add biuret solution at room temperature	blue changes to purple

Semi-quantitative Benedict's test

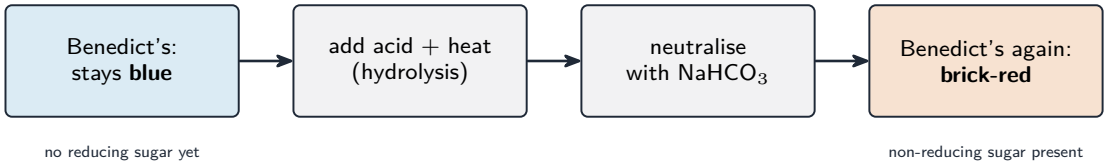
The normal Benedict's test only tells you "yes or no". A **semi-quantitative** 半定量 test gives a rough amount. First you **standardise** 标准化 the test: you run it on solutions of known **concentration** 浓度 and record the result for each. Then you can estimate an unknown by either:

- the **time to the first colour change** (more sugar changes colour faster), or
- comparing the final colour to your set of colour standards.

Testing for non-reducing sugars

Some sugars, such as sucrose, are **non-reducing sugar** 非还原糖: they give a negative Benedict's test. To detect them:

1. Do a normal Benedict's test first. It stays blue (no reducing sugar).
2. Take a fresh sample and add dilute **hydrochloric acid** 盐酸, then heat. This **acid hydrolysis** breaks the sugar into smaller reducing sugars.
3. Cool, then **neutralise** 中和 the acid with sodium hydrogencarbonate.
4. Now do the Benedict's test again. A brick-red colour shows a non-reducing sugar was present.



Testing for a non-reducing sugar: hydrolyse it with acid first, then the second Benedict's test turns red

Worked example. A solution gives a **negative** Benedict's test. It is then boiled with dilute hydrochloric acid, neutralised with sodium hydrogencarbonate, and re-tested with Benedict's - now it turns brick-red. What was present, and why is each step needed? The first negative result rules out a **reducing sugar**. Boiling with acid **hydrolyses** the glycosidic bond, splitting a non-reducing sugar such as **sucrose** into its reducing monosaccharides. The neutralising step is essential because Benedict's **only works in alkaline conditions** - skip it and the test fails even when sugar is present. The positive re-test therefore shows a **non-reducing sugar** was there all along. Quote the **first, negative** test as part of the answer: without it, the final red colour cannot tell a non-reducing sugar from a reducing one.

Carbohydrates

Monomers, polymers and macromolecules

- a **monomer** 单体 is a small molecule that is a single unit.
- a **polymer** 聚合物 is a long molecule made of many monomers joined together.
- a **macromolecule** 大分子 is any very large molecule.

Sugars come in three sizes:

- a **monosaccharide** 单糖 is a single sugar unit, such as **glucose** 葡萄糖 and

fructose.

- a **disaccharide** 二糖 is two units joined, such as maltose and sucrose.
- a **polysaccharide** 多糖 is many units joined into a polymer.

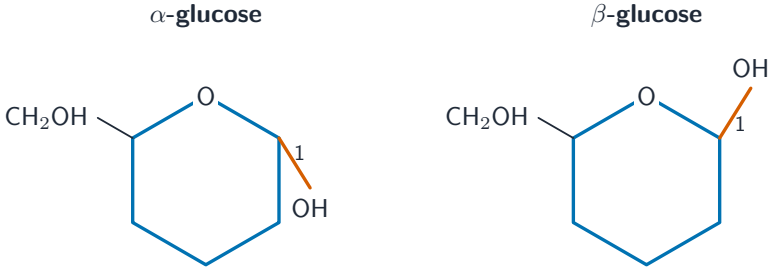
Glucose, **fructose** 果糖 and **maltose** 麦芽糖 are reducing sugars. **Sucrose** 蔗糖 is a non-reducing sugar.

Two ring forms of glucose

Glucose has six carbon atoms and forms a ring. There are two ring forms. They differ only at carbon 1:

- in **α-glucose**, the –OH group on carbon 1 points **down**, below the ring.
- in **β-glucose**, the –OH group on carbon 1 points **up**, above the ring.

This small difference decides which polysaccharide the glucose can build.



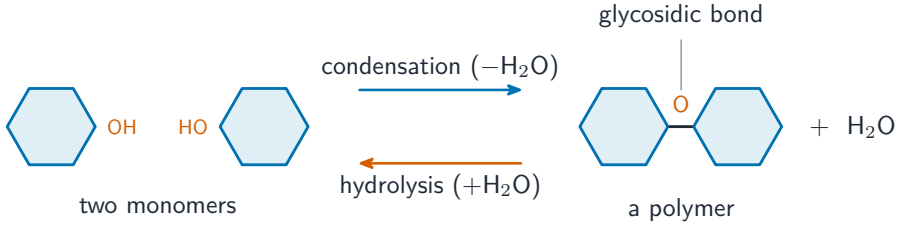
only carbon 1 differs: the OH points **down** in α, **up** in β

The two ring forms differ only at carbon 1: the –OH points down in α, up in β

Joining and breaking sugars

Monomers are joined by strong **covalent bonds** 共价键. When two sugars join, a **glycosidic bond** 糖苷键 forms between them. This happens by **condensation** 缩合: a molecule of water is removed each time a bond forms.

The reverse is **hydrolysis** 水解: a water molecule is added to break a glycosidic bond. This is why the non-reducing sugar test needs acid and heat — they hydrolyse sucrose into glucose and fructose.



Condensation removes water to join monomers; hydrolysis adds water to split them

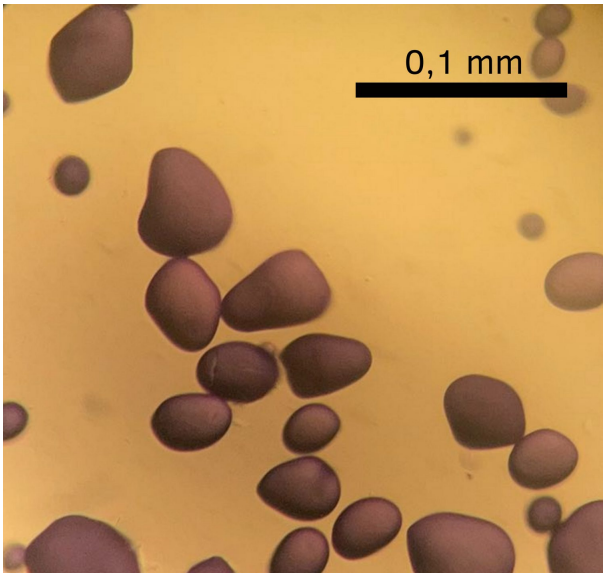
Storage polysaccharides: starch and glycogen

Starch is the energy 能量 store in plants. It is made of two polymers of α-glucose:

- **amylose** 直链淀粉 — a long, unbranched chain that coils into a spiral.
- **amylopectin** 支链淀粉 — a chain with many side branches.

Glycogen 糖原 is the energy store in animals. It is like amylopectin but has even more branches, so it can be broken down quickly when energy is needed.

These stores suit their job well: they are compact, they are **insoluble** 不溶 (so they do not leave the cell), and they do not change the **water potential** 水势 of the cell (so they do not pull water in by **osmosis** 渗透). The many branches give many ends, so glucose can be added or removed fast.



Real starch grains inside a potato, stained by iodine. Each grain is a dense, insoluble package of amylose and amylopectin — note the scale: the largest are only about 0.1 mm across

Cellulose

Cellulose 纤维素 is made of β-glucose. Because of the β form, every other glucose is flipped over, so the chains are long and straight. Many straight chains lie side by side and are held together by **hydrogen bonds** 氢键 into strong bundles called **microfibrils** 微纤丝. These give the plant **cell wall** 细胞壁 its strength and stop the cell bursting.

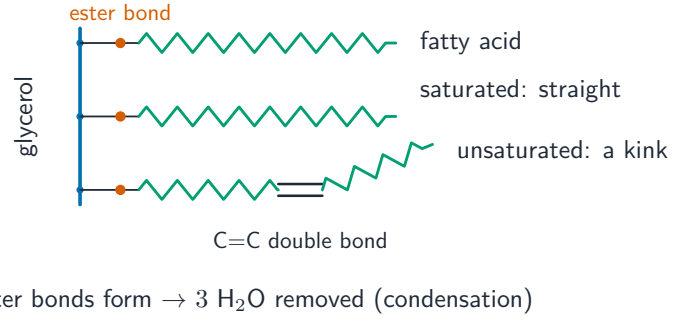


The shape fits the job: amylose coils and amylopectin/glycogen branch (for compact stores), while straight cellulose chains pack into strong fibres

Lipids

Triglycerides

A **triglyceride** 甘油三酯 is the main fat or oil. It is **non-polar** 非极性 and **hydrophobic** 疏水 (it does not mix with water). It is made from one **glycerol** 甘油 molecule joined to three **fatty acids** 脂肪酸 by **ester bonds** 酯键. Each ester bond forms by condensation, so three water molecules are removed.



One glycerol plus three fatty acid tails; a straight tail is saturated, a kinked one unsaturated

Fatty acids are of two kinds:

- **saturated** 饱和 — no carbon–carbon **double bonds** 双键; these fats are usually solid.
- **unsaturated** 不饱和 — one or more double bonds; these oils are usually liquid.

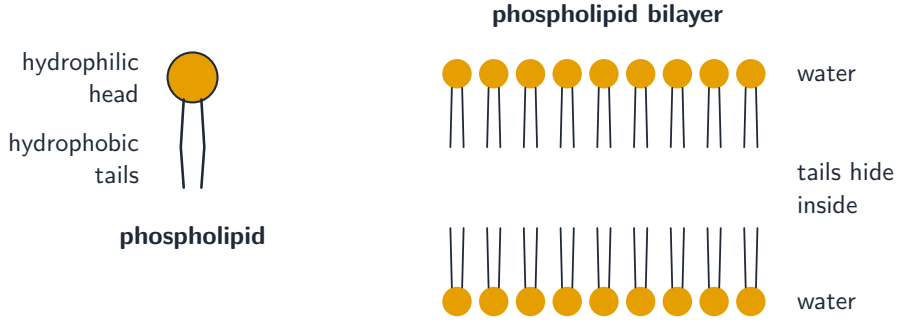
Triglycerides make a good long-term energy store: they release about twice as much energy per gram as carbohydrates, they are insoluble, and they store little extra mass because they hold no water. Under the skin they also give **insulation** 隔热 and protect the organs.

Phospholipids

A **phospholipid** 磷脂 is like a triglyceride, but one fatty acid is replaced by a **phosphate** 磷酸 group. This gives the molecule two ends with different behaviour:

- a **hydrophilic** 亲水 ("water-loving") **polar** 极性 phosphate head.
- two hydrophobic ("water-fearing") fatty acid tails.

This split personality is why phospholipids form the membranes around cells.



The hydrophilic heads face the water; the hydrophobic tails hide inside, forming a bilayer

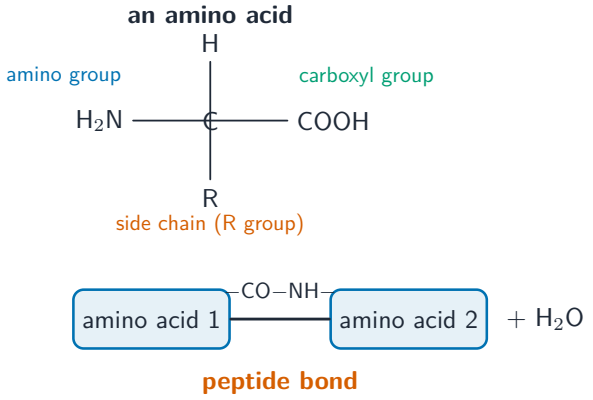
Proteins

Amino acids and the peptide bond

Proteins are polymers of **amino acids** 氨基酸. Every amino acid has the same general structure around a central carbon atom: an **amino group** 氨基 ($-NH_2$), a **carboxyl group** 羧基 (–COOH), a hydrogen atom, and a variable **side chain** 侧链 (the R group). The R group is different in each amino acid.

Two amino acids join by condensation. The bond formed between the amino group of one and the carboxyl group of the next is a **peptide bond** 肽键, and a water molecule

is removed. Many amino acids joined this way make a **polypeptide** 多肽. Adding water (hydrolysis) breaks a peptide bond.



Every amino acid has an amino group, a carboxyl group and an R group; two join by a peptide bond

Four levels of protein structure

Level	What it means
primary structure 一级结构	the order of amino acids in the chain
secondary structure 二级结构	local shapes — the **α-helix** 螺旋 and the **β-pleated sheet** 折叠片 — held by hydrogen bonds
tertiary structure 三级结构	the whole chain folded into a precise 3-D shape
quaternary structure 四级结构	two or more polypeptide chains joined into one protein



The four levels: primary → secondary → tertiary → quaternary structure

The folded shape is held together by four kinds of interaction between R groups:

- **hydrophobic interactions** 疏水作用 (non-polar R groups cluster away from water).
- hydrogen bonding.

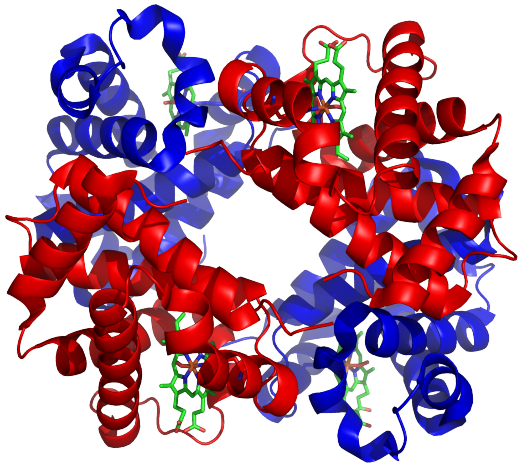
- **ionic bonds** 离子键 (between charged R groups).
- covalent bonding, including strong **disulfide bonds** 二硫键.

Globular and fibrous proteins

- **globular proteins** 球状蛋白质 fold into a rounded shape, are usually **soluble** 可溶, and do jobs in the body (for example enzymes and haemoglobin).
- **fibrous proteins** 纤维状蛋白质 form long strands, are usually insoluble, and give structure and support (for example collagen).

Haemoglobin — a globular protein

Haemoglobin 血红蛋白 carries **oxygen** 氧气 in red blood cells. It has a quaternary structure made of four polypeptide chains: two alpha (α-globin) chains and two beta (β-globin) chains. Each chain holds a **haem group** 血红素. At the centre of each haem group is an **iron** 铁 atom, and this is where one oxygen molecule binds. Four chains mean one haemoglobin molecule can carry four oxygen molecules.



A real haemoglobin molecule, worked out from X-ray data. Count them: two alpha chains (red), two beta chains (blue), and one green haem group held in each — so four oxygen molecules in total

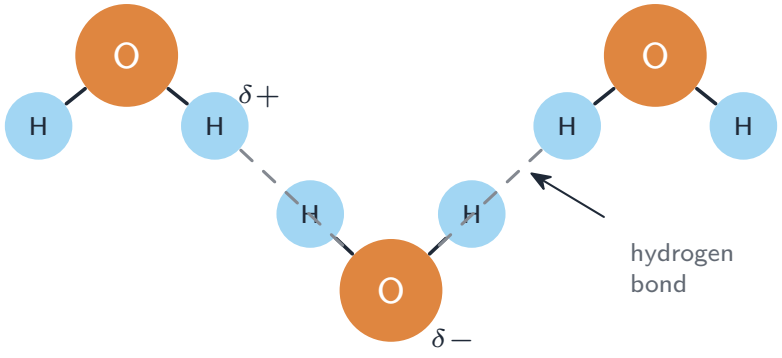
Collagen — a fibrous protein

Collagen 胶原蛋白 gives strength to skin, **tendons** 肌腱, bone and blood vessel walls. One collagen molecule is three polypeptide chains wound tightly around each other in a triple strand, held by hydrogen bonds. Many of these molecules lie side by side, slightly staggered, and are cross-linked into thick **fibres** 纤维. The staggered, cross-linked arrangement makes collagen very strong when pulled.

Water

Water is a small molecule, but its two O–H bonds are **polar**: the oxygen end is slightly negative and the hydrogen ends are slightly positive. So one water molecule attracts its neighbours, forming weak hydrogen bonds between them. These hydrogen bonds explain water's useful properties:

- **solvent action** — water is a good **solvent** 溶剂, so many substances dissolve in it. This lets reactions happen and lets substances be carried around the body.
- **high specific heat capacity** 比热容 — water needs a lot of energy to warm up, so its temperature stays steady. This protects living things from quick temperature changes.
- **latent heat of vaporisation** 汽化潜热 — water needs a lot of energy to **evaporate** 蒸发. So when water evaporates (for example as sweat dries), it carries away a lot of heat and cools the body.



Water is polar (δ− oxygen, δ+ hydrogens), so its molecules attract each other by hydrogen bonds — the reason for all the properties above

Exam tips

- Learn each **food test** as reagent + positive result + colour change (Benedict's → brick-red; iodine → blue-black; biuret → purple; emulsion → white).
- Benedict's detects **reducing** sugars; for a non-reducing sugar you must **hydrolyse with acid first**, then re-test.
- Name the bond precisely: **glycosidic** (carbohydrates), **ester** (lipids), **peptide** (proteins) — all made by **condensation** (water removed).
- Link structure to function: cellulose (straight chains, H-bonds → strong), glycogen/starch (branched/coiled → compact store).
- For proteins, name the bond at each level: primary (peptide), secondary (hydrogen), tertiary (R-group interactions), quaternary.

Enzymes

A-Level Biology

What enzymes are

Enzymes 酶 are **globular proteins** 球状蛋白质 — a type of **protein** 蛋白质 with a rounded, soluble shape. They are biological catalysts: they **catalyse** 催化 (speed up) the chemical reactions in living things, and they are not used up, so each enzyme works again and again.

Enzymes work in two places:

- **intracellular** 细胞内 enzymes work inside the **cell** 细胞 that made them. An example is **catalase** 过氧化氢酶, which breaks down harmful hydrogen peroxide.
- **extracellular** 细胞外 enzymes are **secreted** 分泌 (sent out) to work outside the cell. An example is **amylase** 淀粉酶, which is released into the gut to **digest** 消化 **starch** 淀粉.



Yeast enzymes ferment sugar, giving off bubbles of carbon dioxide

How enzymes work

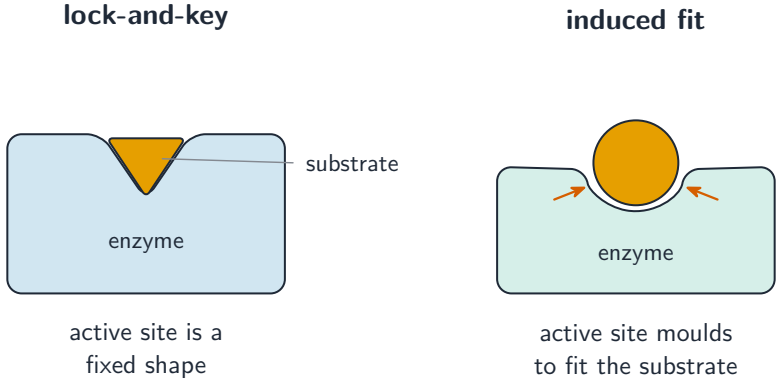
Each enzyme has a special pocket called the **active site** 活性位点. The molecule it acts on is its **substrate** 底物. The substrate fits into the active site to form an **enzyme–substrate complex** 酶底物复合物. The reaction then happens, and the **products** 产物 leave, freeing the active site for the next substrate.

Specificity

An enzyme is **specific**: it usually works on only one substrate. This is because the shape of the active site is **complementary** 互补 to (fits) the shape of that substrate and no other. We call this **specificity** 专一性.

Two ideas explain how the substrate fits:

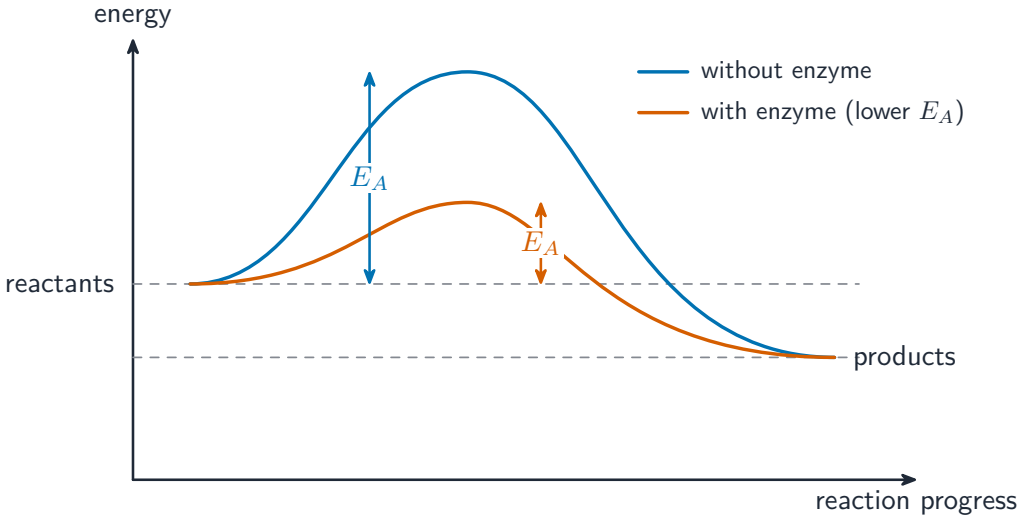
- the **lock-and-key hypothesis** 锁钥学说 — the active site is a fixed shape, and only a substrate with the matching shape fits, like a key in a lock.
- the **induced-fit hypothesis** 诱导契合学说 — the active site is not quite the right shape at first. When the substrate binds, the active site changes shape a little to wrap around it tightly. This idea fits the evidence better.



Lock-and-key: a fixed active site. Induced fit: the active site changes shape to grip the substrate

Lowering activation energy

Every reaction needs a small "push" of energy to start, called the **activation energy** 活化能. An enzyme lowers the activation energy. This lets the reaction go quickly at the cell's normal **temperature** 温度, instead of needing high heat.



The enzyme route has a lower activation energy (E_A), so more molecules can react

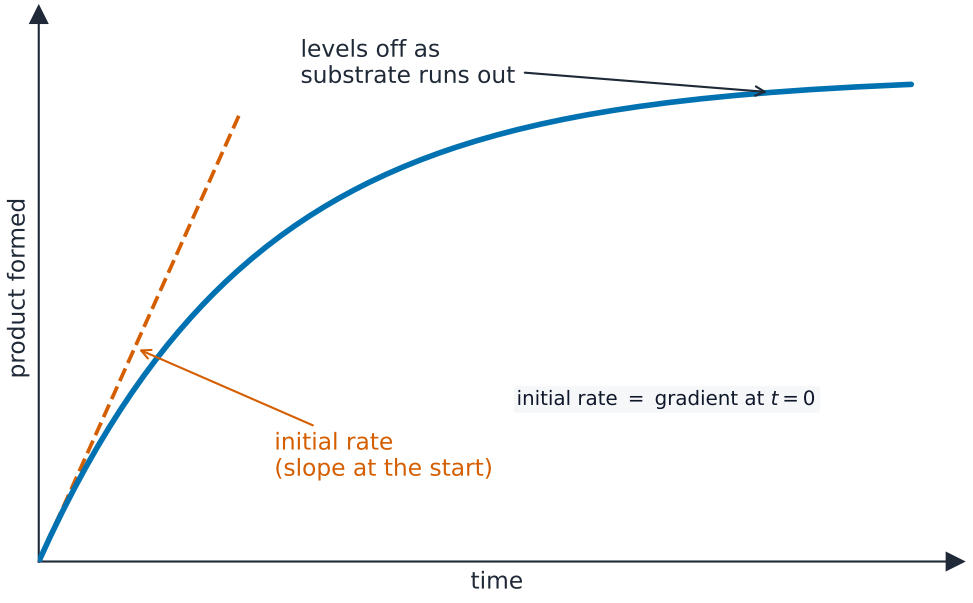
Measuring the rate of a reaction

You can follow an enzyme reaction in two ways:

- measure how fast **product** is made. With catalase, oxygen gas is a product, so you collect the gas and measure its volume over time.
- measure how fast substrate disappears. With amylase, you remove samples and use the iodine test; the blue-black colour fades as the starch is used up.

A **colorimeter** 比色计 makes this exact. It shines light through the tube and measures how much light is absorbed, so a colour change becomes a number you can plot.

The **rate of reaction** 反应速率 is steepest at the start (most substrate present), so the **initial rate** (the slope at time zero) is the fairest value to compare.



Product builds fastest at the start; the initial rate is the slope of the tangent at time zero — the fairest value to compare

Worked example. In the first 20 seconds of a catalase reaction, 16 cm^3 of oxygen is collected. Estimate the rate of reaction.

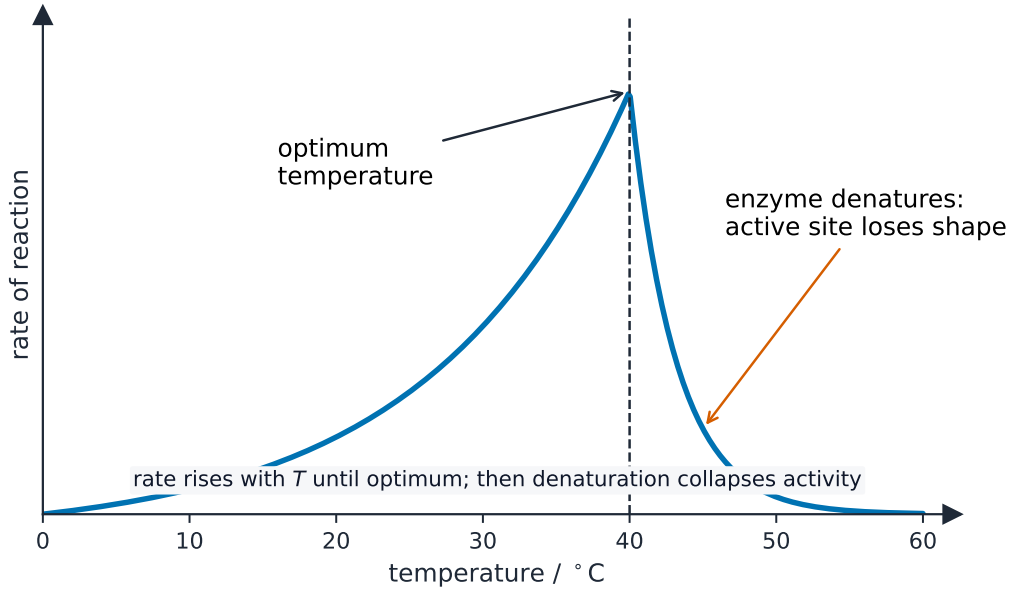
$$\text{rate} = \frac{\text{volume of product}}{\text{time}} = \frac{16}{20} = 0.8 \text{ cm}^3\,\text{s}^{-1}.$$

Because the reaction is fastest at the start, measuring over this short early interval gives a value close to the initial rate; averaging over a longer time would include the slower later stages and underestimate it.

Factors that affect enzyme activity

Temperature

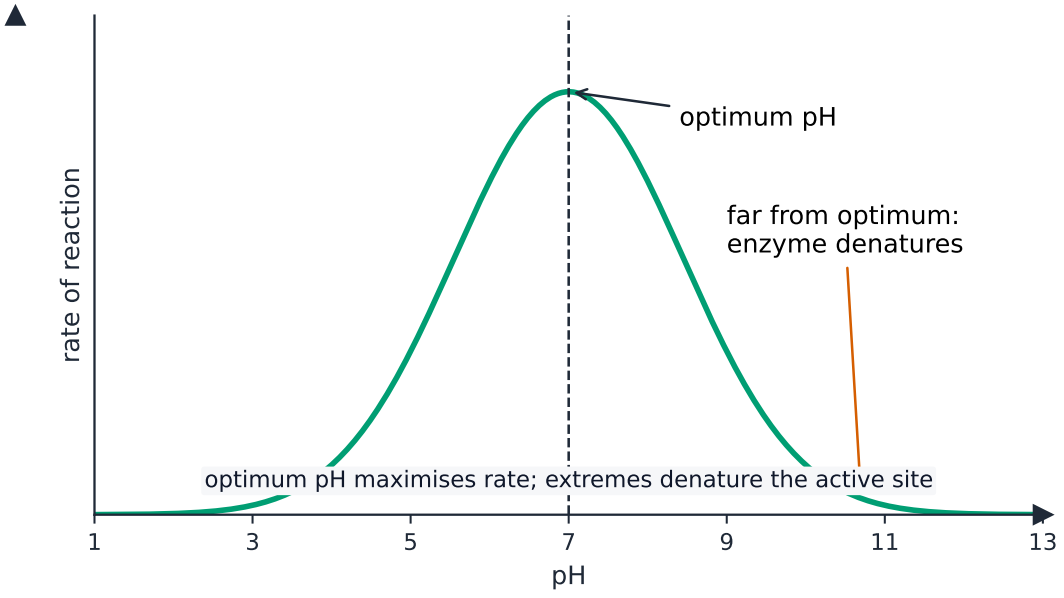
As temperature rises, molecules gain more **kinetic energy** 动能 and **collide** 碰撞 more often, so the rate rises. But above the **optimum temperature** 最适温度 the enzyme begins to **denature** 变性: the heat breaks the bonds holding its shape, so the active site changes and no longer fits the substrate. The rate then falls quickly.



Rate rises to the optimum, then falls fast as the enzyme denatures

pH

Each enzyme has an optimum pH. If the pH moves too far from it, the enzyme denatures and the rate drops. To study pH fairly, you keep it steady with a **buffer solution** 缓冲液.



The rate peaks at the optimum pH and falls away on either side

Enzyme concentration

With plenty of substrate, more enzyme means more active sites, so the rate goes up in proportion to enzyme concentration.

Substrate concentration

At first, adding more substrate speeds the reaction. But once every active site is busy, adding more makes no difference — the rate levels off at a maximum.

Inhibitor concentration

An **inhibitor** 抑制剂 is a molecule that slows an enzyme. The more inhibitor present, the lower the rate.

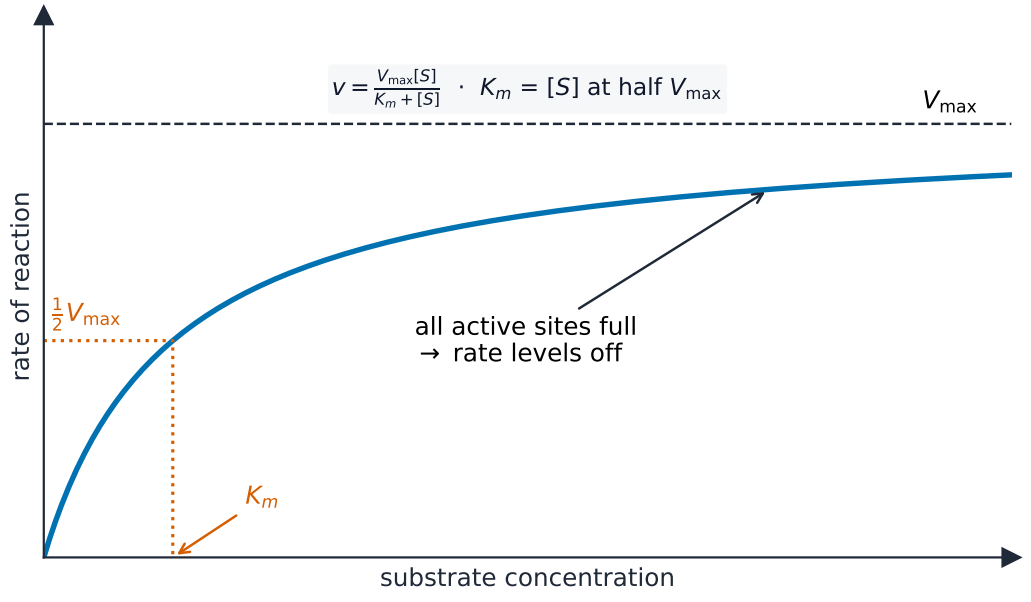
V_max and the Michaelis–Menten constant

The levelling-off rate, when all active sites are full, is the maximum rate, written $V_{\max}$.

The **Michaelis–Menten constant** 米氏常数 (K_{m}) is the substrate concentration that gives **half** of $V_{\max}$. It tells you about the enzyme's **affinity** 亲和力 (pulling power) for its substrate:

- a **low** K_{m} means the enzyme reaches half-speed at a low substrate concentration, so it has a **high** affinity.
- a **high** K_{m} means a **low** affinity.

So K_{m} lets you compare how strongly different enzymes hold their substrates.

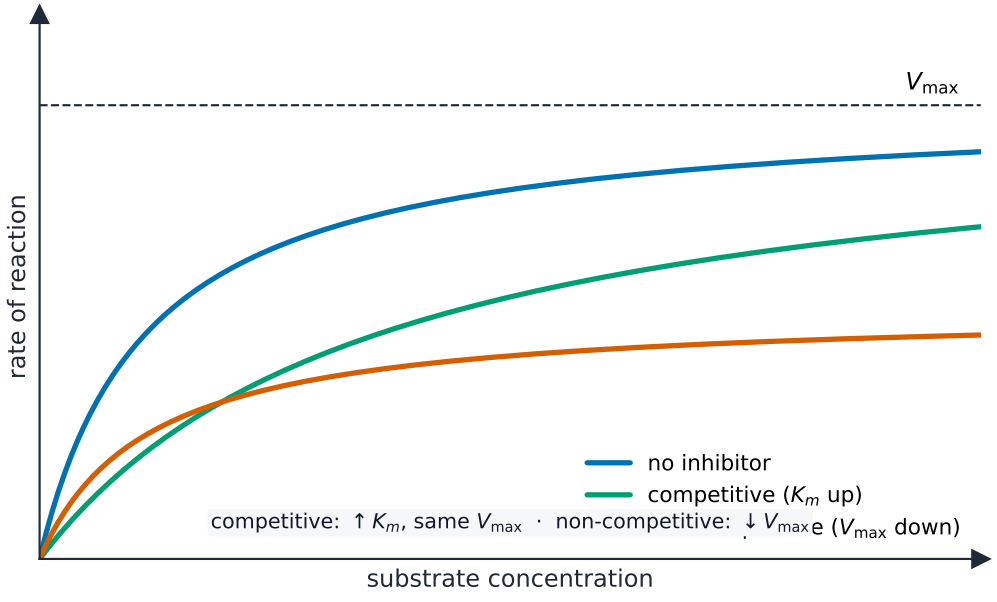


Rate climbs to V_{max} when all active sites are full; K_m is the substrate concentration giving half V_{max}

Reversible inhibitors

Some inhibitors are **reversible** 可逆: they can leave the enzyme again. There are two types.

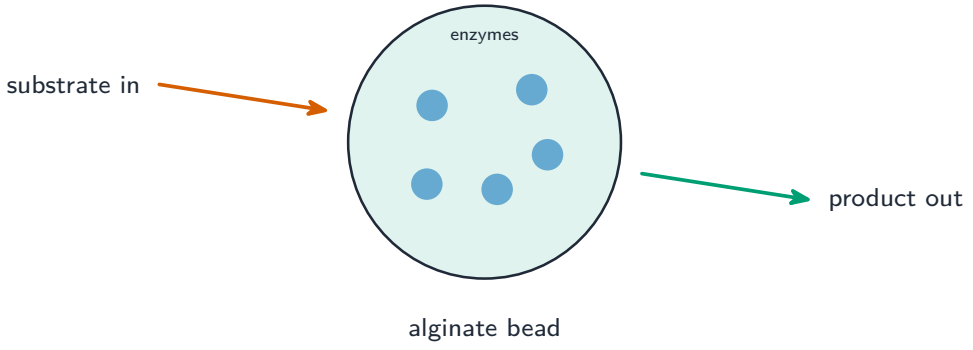
Type	Where it binds	Effect of adding more substrate	Effect on V_{max} and K_m
competitive inhibitor 竞争性抑制剂	in the active site (it has a similar shape to the substrate)	more substrate out-competes it, so its effect is reduced	V_{max} unchanged; K_m rises
non-competitive inhibitor 非竞争性抑制剂	at another site, changing the active site's shape	adding more substrate does not help	V_{max} falls; K_m unchanged



A competitive inhibitor raises K_m (more substrate overcomes it); a non-competitive one lowers V_{max}

Immobilised enzymes

An **immobilised enzyme** 固定化酶 is fixed in place — for example, trapped inside small beads of **alginate** 海藻酸盐 — instead of floating free in solution. The substrate solution flows past the beads.



The enzymes stay trapped in the bead while substrate flows in and product flows out — so the enzyme is never washed away

A free enzyme usually works a little faster, because the substrate can reach it easily. But immobilised enzymes have big practical advantages:

- the enzyme is not washed away, so it can be used again and again.
- the product is pure — it is not mixed with enzyme.
- the enzyme is more stable, so it survives changes in temperature and pH better.
- the process can run continuously, with substrate flowing in and product flowing out.



Biological washing powders contain enzymes that digest food and blood stains at low temperatures

Exam tips

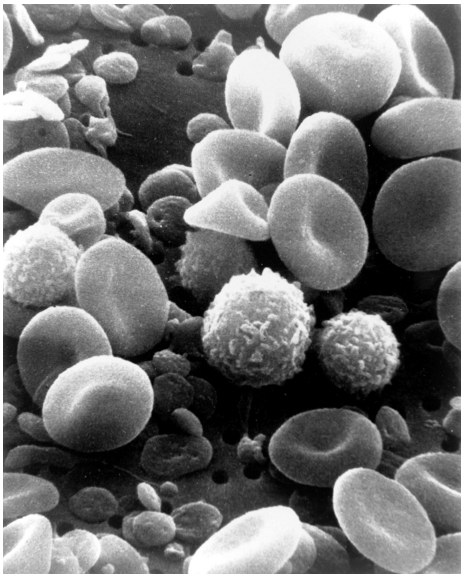
- Explain enzyme action with the **induced-fit** model (the active site moulds around the substrate) — now preferred to lock-and-key.
- Above the optimum the enzyme **denatures** (hydrogen bonds and tertiary structure break) — write "denatures", never "dies" or "is killed".
- Distinguish inhibitors: **competitive** binds the active site (overcome by more substrate, same V_{max}); **non-competitive** binds elsewhere (lower V_{max}).
- Compare rates using the **initial rate** (tangent at time zero) — the fairest measure, before substrate becomes limiting.

Cell membranes and transport

A-Level Biology

The cell surface membrane

Every cell is wrapped in a **cell surface membrane** 细胞膜. We describe its structure with the **fluid mosaic model** 流动镶嵌模型.



Red blood cells under a scanning electron microscope — each is wrapped in a cell surface membrane

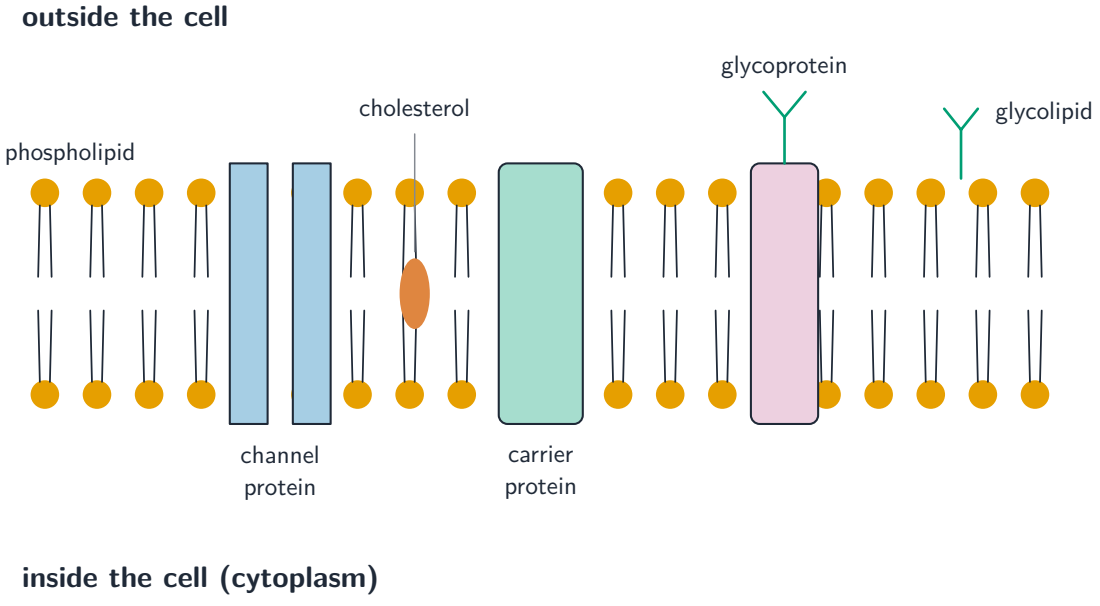
The phospholipid bilayer

The membrane is built mainly from **phospholipids** 磷脂. Each phospholipid has a **hydrophilic** 亲水 ("water-loving") head and two **hydrophobic** 疏水 ("water-fearing") tails. There is water on both sides of the membrane, so the phospholipids line up in two layers — a **bilayer** 双层 — with the heads facing the water outside and inside, and the tails hidden in the middle, away from water. This arrangement forms by itself because of those hydrophilic and hydrophobic interactions.

The model is called "fluid" because the phospholipids are not fixed: they slide past each other, so the membrane can move and bend. It is called a "mosaic" because many **proteins** 蛋白质 are dotted through it, like tiles in a picture.

What floats in the membrane

Part	Where it sits	Main roles
phospholipids	the two layers	form the basic barrier
proteins	through the membrane or on its surface	transport, support and signalling
carrier proteins 载体蛋白	span the membrane	carry specific molecules across
channel proteins 通道蛋白	span the membrane	form water-filled pores for ions to pass
cholesterol 胆固醇	between the phospholipid tails	controls **fluidity** 流动性 and adds strength
glycolipids 糖脂 and **glycoproteins** 糖蛋白	carbohydrate chains on the outer surface	cell recognition; some act as **antigens** 抗原



The fluid mosaic model: proteins, cholesterol and carbohydrate chains sit in a fluid phospholipid bilayer

So the membrane molecules together give the membrane its **stability** 稳定性, its fluidity, its **permeability** 通透性 (control over what gets through), its transport jobs, its signalling jobs, and its cell recognition.

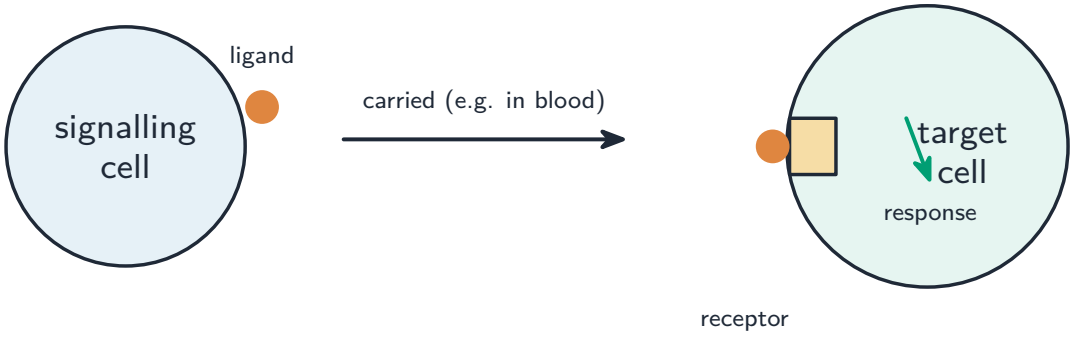
The membrane is **partially permeable** 半透膜: it lets some substances through easily but blocks others.

Cell signalling

Cells talk to each other by **cell signalling** 细胞信号传递. The main stages are:

1. a cell **secretes** a signal chemical called a **ligand** 配体 (for example a **hormone** 激素).
2. the ligand is carried (often in the blood) to a **target cell** 靶细胞.

3. the ligand binds to a specific **receptor** 受体 on the target cell's surface membrane. The shape of the receptor matches that ligand only. Binding then triggers a particular response inside the target cell.



The ligand fits one receptor shape only, so only cells with that receptor respond to the signal

Moving substances across the membrane

There are six processes. Some are **passive** 被动 (they need no energy 能量), and some are **active** (they use energy from ATP).

Simple diffusion

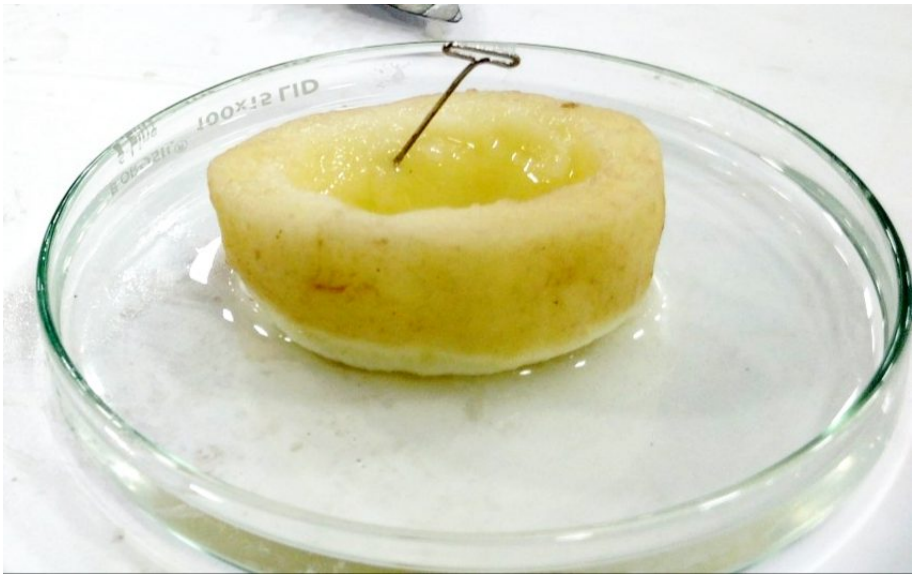
Diffusion 扩散 is the net movement of particles from where they are at a high **concentration** 浓度 to where they are at a low concentration, until they are spread evenly. **Simple diffusion** 简单扩散 is when particles pass straight through the bilayer, down the **concentration gradient** 浓度梯度. Only small or non-polar molecules can do this — such as **oxygen** 氧气 and **carbon dioxide** 二氧化碳. It is passive.

Facilitated diffusion

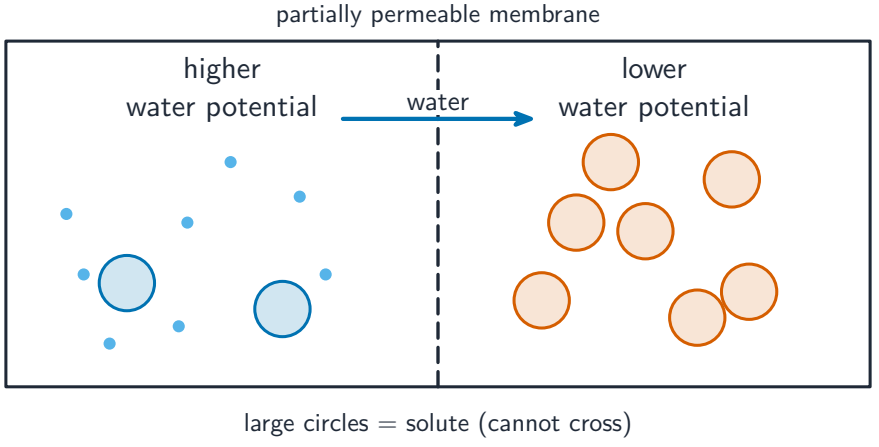
Charged **ions** 离子 and large polar molecules (such as **glucose** 葡萄糖) cannot cross the oily bilayer by themselves. In **facilitated diffusion** 易化扩散 they cross through channel proteins or carrier proteins, still moving down the concentration gradient. It is also passive.

Osmosis

Osmosis 渗透 is the diffusion of water across a partially permeable membrane, from a higher **water potential** 水势 to a lower water potential. It is passive.



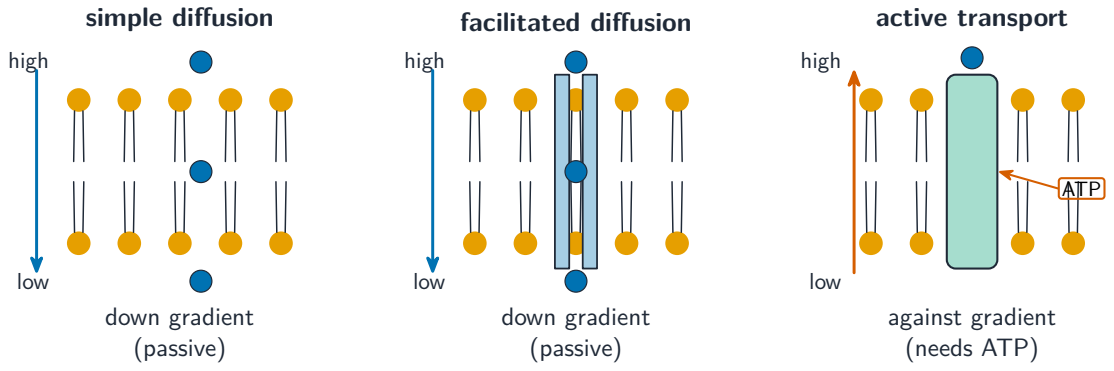
A potato osmometer: the sugar solution rises up the tube as water enters the potato by osmosis



Water crosses to the lower water potential; the solute is too big to cross the partially permeable membrane

Active transport

Active transport 主动运输 moves a substance **against** its concentration gradient — from low to high concentration. This needs carrier proteins and energy from ATP.

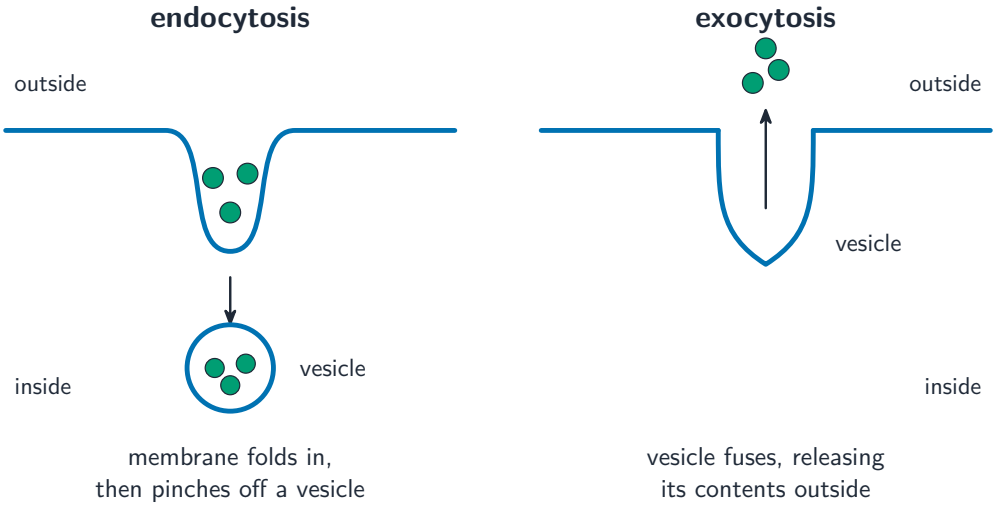


Diffusion and facilitated diffusion are passive (down the gradient); active transport goes against it and needs ATP

Endocytosis and exocytosis

These move large amounts of material in bulk, using ATP.

- in **endocytosis** 胞吞作用, the membrane folds inwards around material and pinches off a **vesicle** 囊泡 to bring it into the cell.
- in **exocytosis** 胞吐作用, a vesicle fuses with the membrane and releases its contents outside the cell.



Endocytosis brings material in by forming a vesicle; exocytosis fuses a vesicle to release its contents

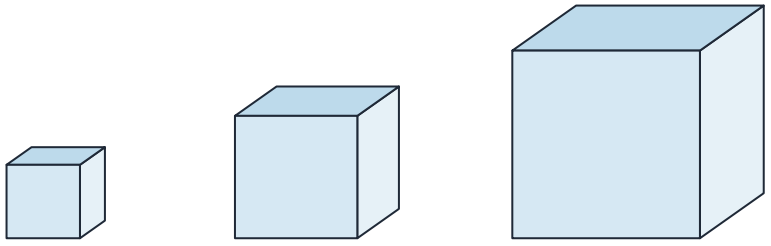
Surface area to volume ratio

A cell takes in and removes substances across its surface. As an object gets bigger, its **volume** 体积 grows faster than its **surface area** 表面积. So the surface area to volume ratio gets **smaller** as size increases.

For a cube of side L:

$$\text{surface area} = 6L^2, \qquad \text{volume} = L^3, \qquad \text{ratio} = \frac{6}{L}.$$

A large L gives a small ratio. This is why small cells (and thin, flat shapes) exchange materials quickly, while large cells cannot rely on diffusion alone.



side $L = 1$
SA : V $= 6 : 1$

side $L = 2$
SA : V $= 3 : 1$

side $L = 3$
SA : V $= 2 : 1$

a bigger cube (or cell) has a **smaller** surface area : volume ratio

As a cube (or cell) grows, its surface area : volume ratio gets smaller

Worked example. Compare the surface area : volume ratio of a cube-shaped cell of side 4 with one of side 10.

For side 4: surface area $= 6 \times 4^2 = 96$ and volume $= 4^3 = 64$, so the ratio is $96 : 64 = 1.5 : 1$. For side 10: surface area $= 6 \times 10^2 = 600$ and volume $= 10^3 = 1000$, so the ratio is $600 : 1000 = 0.6 : 1$. The larger cell has the smaller ratio, so it exchanges materials across its surface more slowly for its size — which is why large organisms need specialised exchange surfaces such as lungs and gills.

You can show this with **agar** 琼脂 blocks of different sizes soaked in dye or acid: the smallest block, with the largest surface area to volume ratio, changes colour all the way through fastest. Diffusion across non-living materials can also be studied with **dialysis tubing** 透析袋 (Visking tubing).

Water potential and living cells

Water potential measures how likely water is to leave a solution. Pure water has the highest water potential. Adding a **solute** 溶质 (a dissolved substance) lowers it. Water always moves by osmosis from a higher to a lower water potential.

To estimate the water potential of plant tissue, you place pieces in sucrose solutions of different water potentials. The solution that causes **no change** in mass or length has about the same water potential as the tissue.

Effect on plant cells

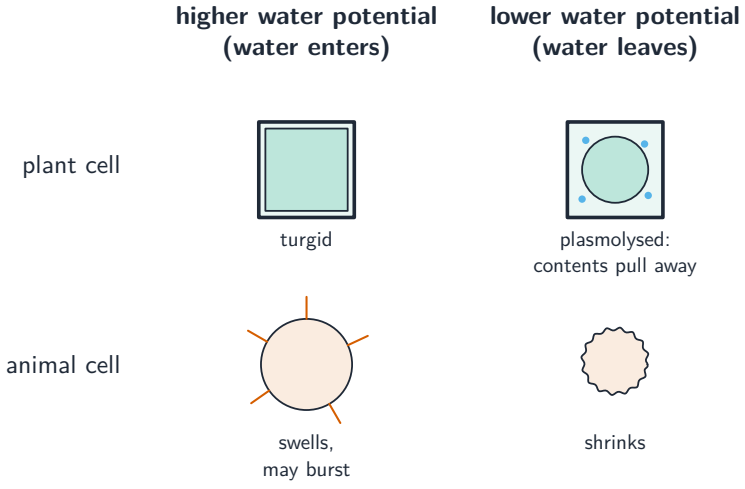
- in a solution of **higher** water potential (for example **distilled water** 蒸馏水), water enters the cell. The cell swells and becomes **turgid** 膨胀, but the strong cell wall stops it bursting.

- in a solution of **lower** water potential, water leaves. The cell contents shrink and the membrane pulls away from the cell wall — this is **plasmolysis** 质壁分离.

Effect on animal cells

Animal cells have no cell wall to protect them.

- in a solution of higher water potential, water enters and the cell may burst. In a red blood cell this bursting is called **haemolysis** 溶血.
- in a solution of lower water potential, water leaves and the cell shrinks.



A plant cell becomes turgid or plasmolysed; an animal cell may burst (haemolysis) or shrink

Exam tips

- Describe the membrane as a **fluid mosaic**: a phospholipid bilayer with proteins, cholesterol and glycoproteins.
- Sort each process: diffusion, facilitated diffusion and osmosis are **passive** (down a gradient); active transport and endo/exocytosis need **ATP** and can go against it.
- For osmosis always use **water potential** (Ψ): water moves from **high (less negative) to low (more negative)** Ψ; pure water is 0, the highest.
- State the outcome by cell type: plant cell turgid/plasmolysed; animal cell lyses/crenates — link it to the water-potential gradient.

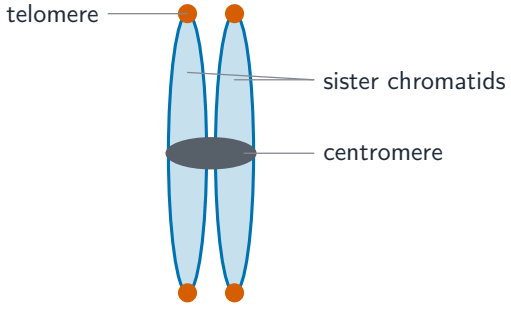
The mitotic cell cycle

A-Level Biology

The structure of a chromosome

A **chromosome** 染色体 is one very long molecule of DNA wound tightly around special **proteins** 蛋白质 called **histones** 组蛋白. Winding the DNA like this lets a huge length fit inside the nucleus and keeps it tidy.

Before a cell divides, its DNA is copied (this copying is called **replication** 复制). After copying, each chromosome is made of two identical copies joined together. These two copies are the **sister chromatids** 姐妹染色单体, and they are held together at a point called the **centromere** 着丝粒. The tips of each chromosome are capped by **telomeres** 端粒, which protect the ends.



one chromosome = two identical copies after DNA replication

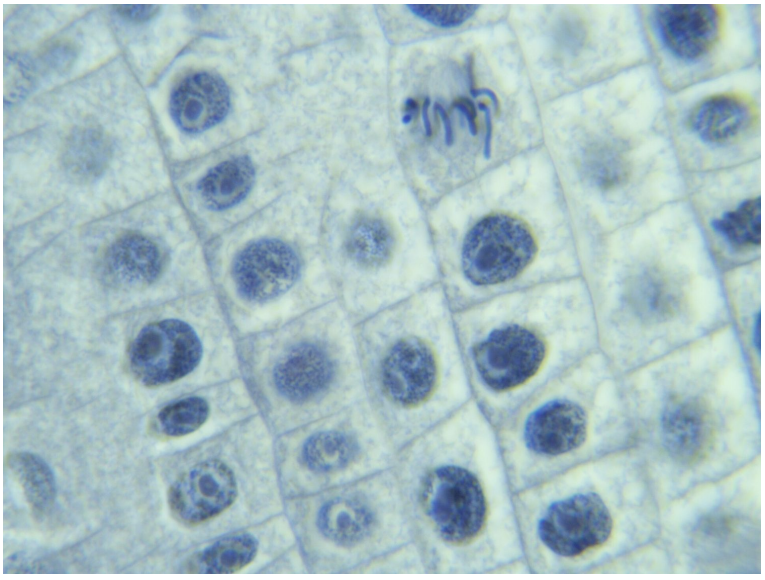
After replication a chromosome is two sister chromatids joined at the centromere, with telomeres at the tips

Why mitosis matters

Mitosis 有丝分裂 is a type of nuclear division that makes two **daughter cells** 子细胞 that are **genetically identical** — they carry exactly the same **genes** 基因 as the parent cell and as each other.

This matters for:

- **growth** of **multicellular** 多细胞 **organisms** 生物体 (making more cells).
- **replacement** of damaged or dead cells.
- **repair** of **tissues** 组织 by making new cells.
- **asexual reproduction** 无性生殖 (one parent makes identical offspring).

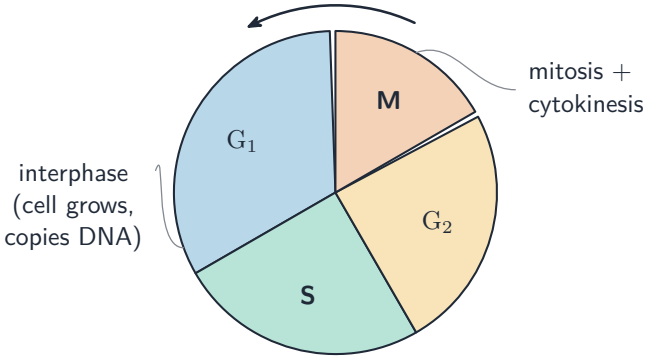


Onion root tip at metaphase: mitosis produces identical body cells for growth and repair

The mitotic cell cycle

The **cell cycle** 细胞周期 is the full life of a cell from one division to the next. It has three parts:

1. **interphase** 间期 — the longest part. The cell grows in the $\mathbf{G_1}$ phase, copies its DNA in the **S** phase (replication), and grows again and prepares to divide in the $\mathbf{G_2}$ phase.
2. **mitosis** — the nucleus divides into two identical nuclei.
3. **cytokinesis** 胞质分裂 — the rest of the cell splits, giving two separate daughter cells.



Most of the cycle is interphase (G_1, S, G_2); mitosis (M) and cytokinesis are a short part

The mitotic index

The mitotic index is the fraction of cells in a sample that are in mitosis. You find it by counting cells under a microscope:

$$\text{mitotic index} = \frac{\text{number of cells in mitosis}}{\text{total number of cells}}.$$

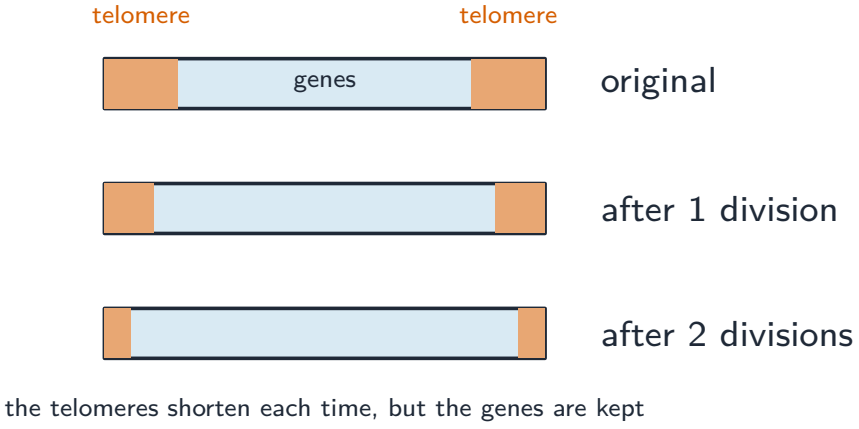
Worked example. In a root-tip sample, 8 of the 50 cells seen are in mitosis. Find the mitotic index.

$$\text{mitotic index} = \frac{8}{50} = 0.16.$$

A high mitotic index means many cells are dividing quickly — normal in a growing root tip, but in adult animal tissue it can be a warning sign of uncontrolled growth.

Telomeres and the ends of chromosomes

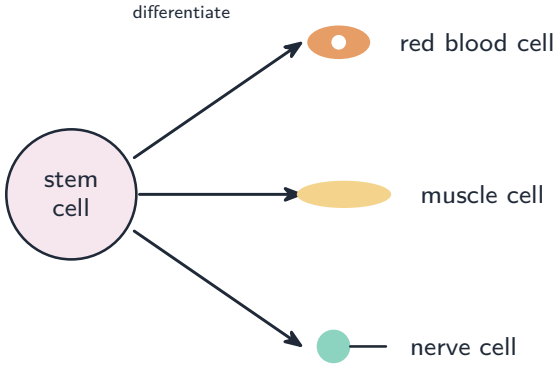
When DNA is replicated, the copying cannot reach the very end of the molecule, so a little is lost each time. Telomeres are short, repeated lengths of DNA at the ends that carry **no genes**. Because the telomeres are shortened instead, no important genes are lost during replication.



The telomeres (which carry no genes) shorten a little at each division, so the genes in the middle are never lost

Stem cells

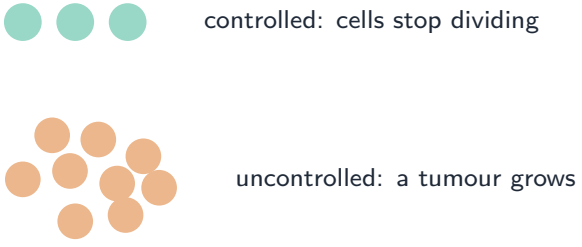
A **stem cell** 干细胞 is an unspecialised cell that can keep dividing by mitosis and can **differentiate** 分化 (change) into different specialised cell types. Stem cells are the source of new cells for replacing lost cells and repairing tissues.



One unspecialised stem cell can differentiate into many specialised cell types

Uncontrolled division and tumours

The cell cycle is normally tightly controlled, so cells divide only when needed. If this control is lost, a cell may divide again and again without stopping. This uncontrolled division produces a lump of cells called a **tumour** 肿瘤.



Normally cells stop dividing when they should; if that control is lost, they keep dividing into a tumour

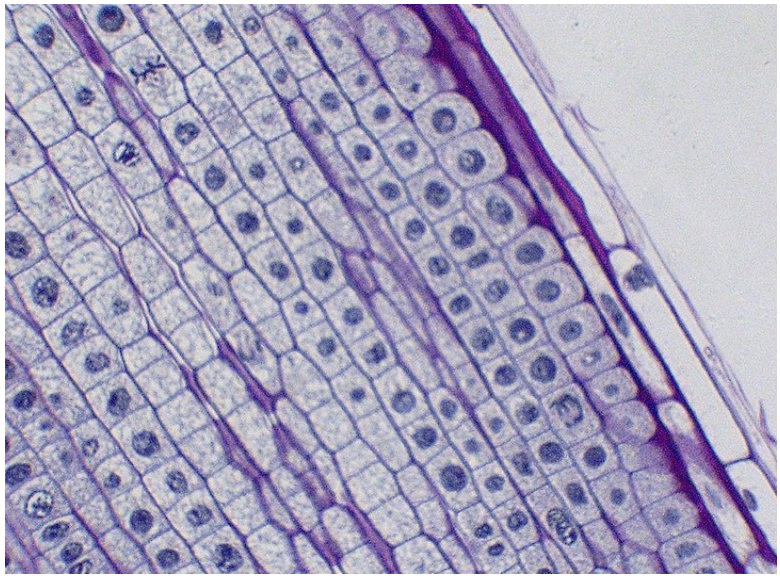
The stages of mitosis

Mitosis runs through four stages. You should be able to recognise them in photomicrographs and slides.

Stage	What happens
prophase 前期	chromosomes coil up and become visible as two sister chromatids; the **nuclear envelope** 核膜 breaks down; a **spindle** 纺锤体 of fibres forms across the cell
metaphase 中期	chromosomes line up along the middle (the **equator** 赤道); spindle fibres attach to each centromere
anaphase 后期	the centromeres split; the sister chromatids are pulled to opposite ends (poles) of the cell
telophase 末期	a set of chromosomes reaches each pole; a new nuclear envelope forms around each set, making two nuclei



The four stages: prophase, metaphase, anaphase, telophase



A real onion root tip: in a growing tip many cells are caught dividing

Cytokinesis then follows. In an animal cell the cell surface membrane pinches inwards to split the cell; in a plant cell a new wall forms across the middle. The result is two genetically identical daughter cells.

Exam tips

- Learn the stages **in order** (prophase, metaphase, anaphase, telophase) with one key event each; interphase (G1, S, G2) is **not** part of mitosis.
- Mitosis gives **two genetically identical diploid** cells — for growth, repair and asexual reproduction.
- **Mitotic index** = cells in mitosis ÷ total cells; a high value means rapid division.
- Link **uncontrolled mitosis** to tumours and telomere shortening to the limit on the number of divisions.

Nucleic acids and protein synthesis

A-Level Biology

Nucleotides — the building blocks

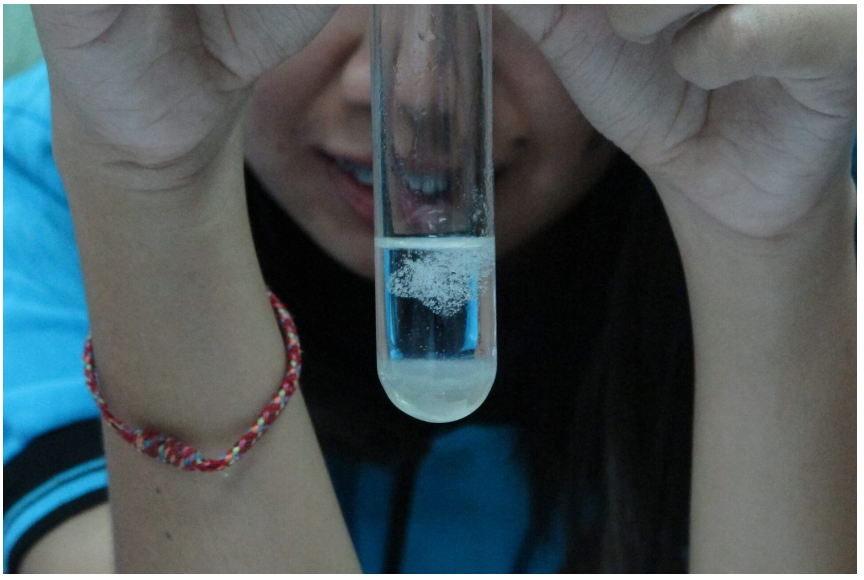
Nucleic acids 核酸 (DNA and RNA) are polymers of small units called **nucleotides** 核苷酸. Each nucleotide is made of three parts joined together:

- a **phosphate** 磷酸 group,
- a sugar (a 5-carbon sugar),
- a nitrogen-containing **base** 碱基.

ATP is a special nucleotide. It has the base **adenine** 腺嘌呤, the sugar **ribose** 核糖, and **three** phosphate groups. Breaking off the last phosphate releases **energy** 能量 for the cell.



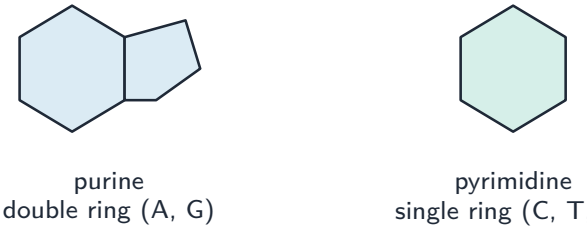
A nucleotide is a phosphate, a sugar and a base; ATP is a nucleotide with three phosphates



DNA extracted from cells appears as pale, stringy strands of these nucleotide polymers

There are five bases, in two groups:

- **purines** 嘌呤 have a **double ring** (two rings): adenine and **guanine** 鸟嘌呤.
- **pyrimidines** 嘧啶 have a **single ring** (one ring): **cytosine** 胞嘧啶, **thymine** 胸腺嘧啶 and **uracil** 尿嘧啶.



The two groups differ in shape: a purine is a double ring, a pyrimidine a single ring — which is why a purine always pairs with a pyrimidine

The structure of DNA

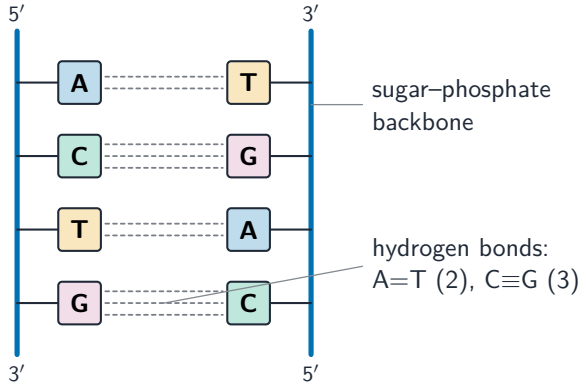
A DNA molecule is two strands twisted together into a **double helix** 双螺旋.

Each **strand** 链 has a backbone of alternating sugar (here the sugar is **deoxyribose** 脱氧核糖) and phosphate. The sugar of one nucleotide is joined to the phosphate of the next by a **phosphodiester bond** 磷酸二酯键.

The two strands are held together by their bases, which meet in the middle. The pairing is exact — this is **complementary base pairing** 碱基互补配对:

- A always pairs with T, held by **two hydrogen bonds** 氢键.
- C always pairs with G, held by **three** hydrogen bonds (so a C–G **base pair** 碱基对 is harder to separate).

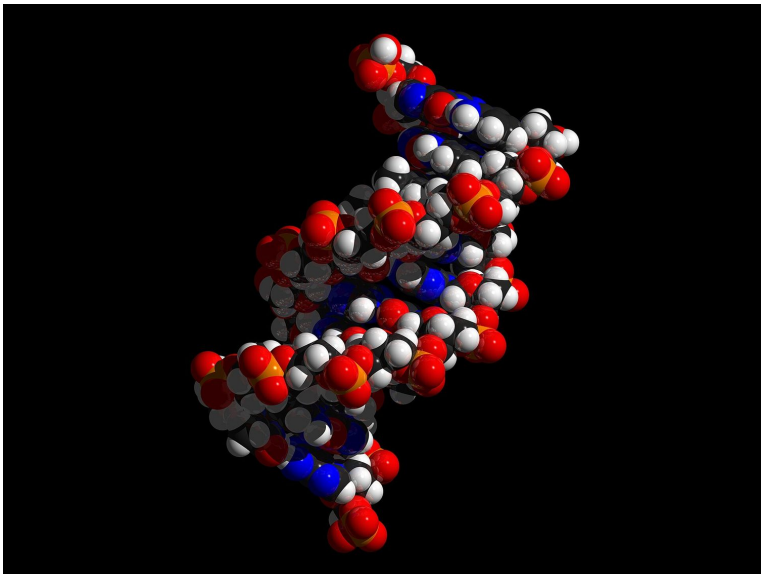
The two strands run in opposite directions: one goes 5′ to 3′ while the other goes 3′ to 5′. We say they are **antiparallel** 反平行.



the two strands run in opposite directions (antiparallel)

Complementary base pairing: A pairs with T (2 hydrogen bonds), C with G (3); the strands are antiparallel

The flat ladder above is twisted into a spiral. This space-filling model, where every atom is a ball, shows the real shape of the **double helix**:



A space-filling model of DNA: the two strands twist around each other into the ***double helix*** *— the ladder of the diagram, coiled up*

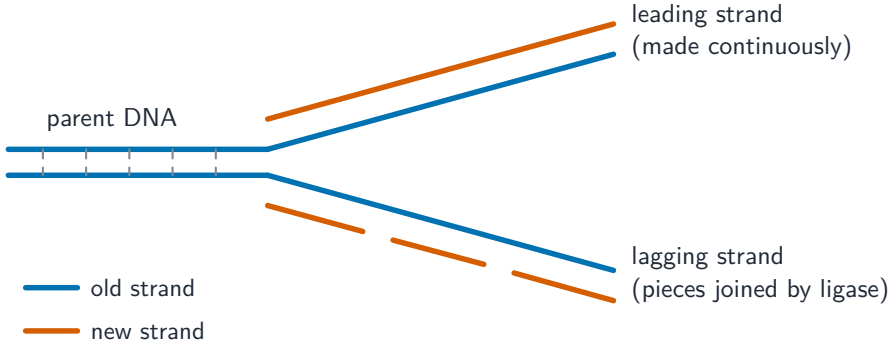
DNA replication

DNA **replication** 复制 (copying) happens during the S phase of the cell cycle. It is **semi-conservative** 半保留复制: each new molecule keeps one old strand and one new strand. The steps are:

1. the double helix unwinds and the hydrogen bonds break, so the two strands separate.
2. each old strand acts as a template. Free nucleotides pair with the exposed bases by complementary base pairing.
3. the **enzyme** 酶 **DNA polymerase** 聚合酶 joins the new nucleotides into a strand. It can only add nucleotides in the 5′ to 3′ direction.

Because of that 5′ to 3′ rule, the two new strands are made differently:

- the **leading strand** 前导链 is built continuously, following the unwinding.
- the **lagging strand** 后随链 is built in short pieces, working away from the unwinding point. The enzyme **DNA ligase** 连接酶 then joins these pieces together.

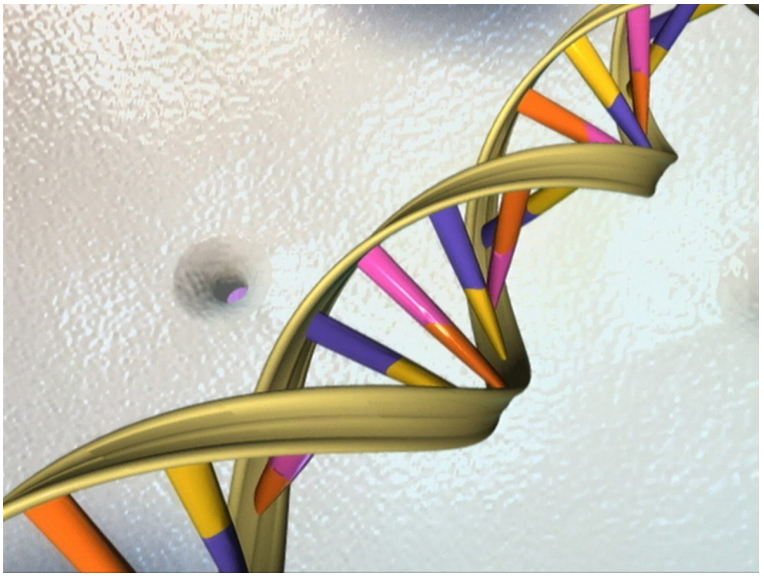


each new molecule = one old strand + one new strand (semi-conservative)

Replication is semi-conservative: each new molecule keeps one old strand (blue) and one new strand (orange)

RNA

RNA is also made of nucleotides, but it is a single strand, its sugar is ribose, and it uses uracil in place of thymine. The most important type here is **messenger RNA (mRNA)**, which carries a copy of a gene's instructions out of the nucleus to be used.



DNA double helix: RNA is a related nucleic acid that carries genetic information into protein synthesis

The genetic code

A **gene** 基因 is a sequence of DNA nucleotides that codes for one **polypeptide** 多肽.

The code is read in **triplets** 三联体 — groups of three bases. Each triplet either codes for one specific **amino acid** 氨基酸, or acts as a start or stop signal. The code is **universal**: nearly all living things use the same triplets for the same amino acids.

There are 64 possible triplets but only 20 common amino acids, so most amino acids are coded by more than one triplet.

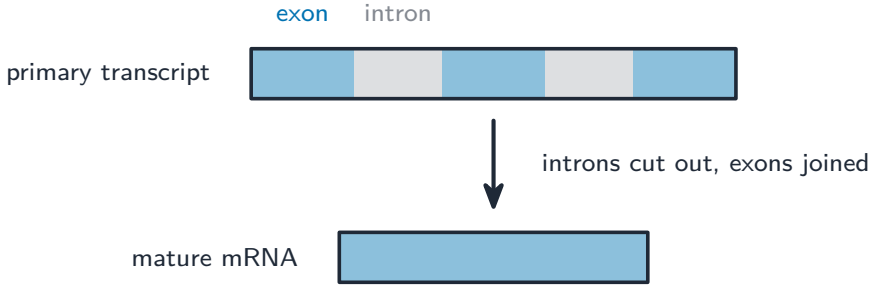
Protein synthesis: transcription and translation

Transcription (in the nucleus)

The DNA gene is copied into mRNA. This is **transcription** 转录.

- the strand of DNA that is copied is the **template strand** 模板链; the partner strand is the **non-transcribed strand** 非转录链.

- the enzyme **RNA polymerase** joins RNA nucleotides that pair with the template bases (with uracil pairing to adenine).
- in eukaryotes the first RNA made (the **primary transcript** 初级转录本) contains coding parts called **exons** 外显子 and non-coding parts called **introns** 内含子. The introns are cut out and the exons joined together to form the finished mRNA.

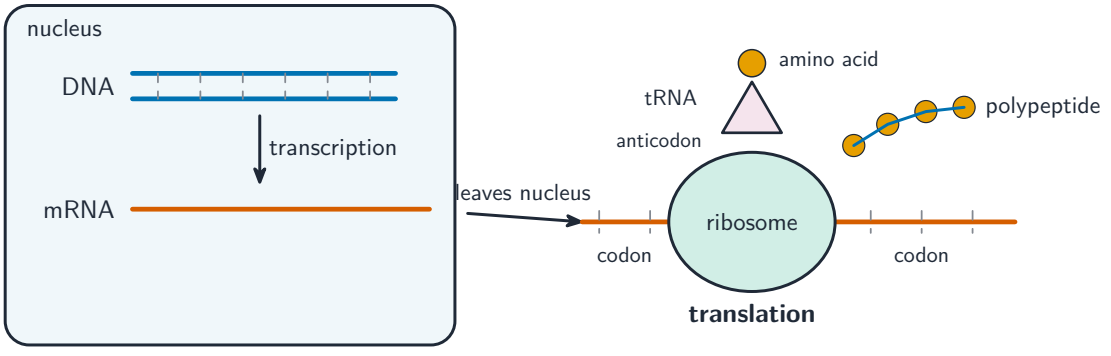


Splicing: the non-coding introns are cut out and the coding exons joined to make the finished mRNA

Translation (at the ribosome)

The mRNA leaves the nucleus and attaches to a **ribosome** 核糖体. Building the polypeptide from the mRNA code is **translation** 翻译.

- the mRNA is read in **codons** 密码子 (each codon is one triplet of mRNA bases).
- molecules of **transfer RNA (tRNA)** bring amino acids to the ribosome. Each tRNA has an **anticodon** 反密码子 that pairs with a matching codon.
- as the codons are read in order, the ribosome joins the amino acids with **peptide bonds** 肽键, building the polypeptide.



Transcription copies DNA into mRNA in the nucleus; translation at the ribosome builds the polypeptide

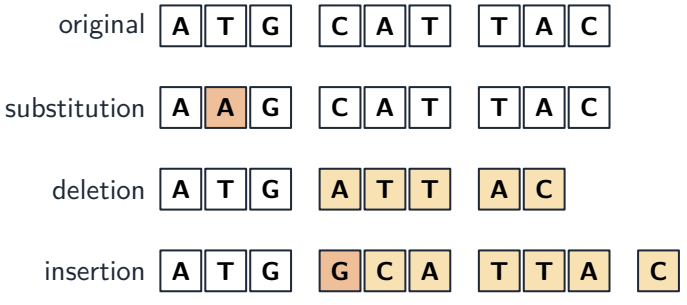
Worked example. A DNA template strand reads `TAC GGA CTT`. Give the mRNA codons, and say how long the polypeptide is. Transcribe by **complementary base pairing**, remembering that RNA uses **uracil** in place of thymine: `TAC` gives `AUG`, `GGA` gives `CCU`, `CTT` gives `GAA`. So the mRNA reads `AUG CCU GAA`. `AUG` is the **start** codon (methionine), so the polypeptide is **three** amino acids long - two once the start methionine is removed. Two errors cost most of the marks here: pairing A with **T** instead of **U** when writing mRNA, and transcribing the **coding** strand instead of the **template** strand. The mRNA is complementary to the template, and identical to the coding strand apart from U replacing T.

Gene mutations

A gene **mutation** 突变 is a change in the base sequence of a DNA molecule. It may change the polypeptide made. There are three types:

- **substitution** 替换 — one base is swapped for a different base. This changes at most one amino acid, and sometimes none (because most amino acids have more than one triplet).
- **deletion** 缺失 — a base is removed.
- **insertion** 插入 — an extra base is added.

A deletion or insertion shifts how every later triplet is read, so it usually changes many amino acids after that point and has a large effect on the polypeptide.



A substitution changes one triplet; a deletion or insertion shifts every later triplet (a frameshift)

Exam tips

- DNA is **antiparallel**, held by hydrogen bonds (A=T two, C≡G three); replication is **semi-conservative** (each product keeps one old strand).
- State **where** and **what**: transcription (nucleus → mRNA), translation (ribosome, mRNA + tRNA → polypeptide).
- Describe the code with all four terms: **triplet, non-overlapping, degenerate, universal**.
- For a **gene mutation** (substitution, insertion, deletion) explain the effect on the protein — and why a substitution can be silent (degeneracy).

Transport in plants

A-Level Biology

The two transport tissues

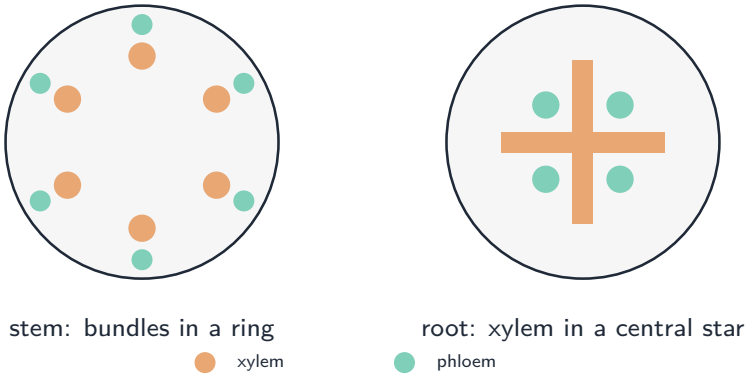
Plants move substances through two transport **tissues** 组织:

- **xylem** 木质部 carries water and dissolved **mineral ions** 矿物离子 up from the roots.
- **phloem** 韧皮部 carries dissolved foods (called **assimilates** 同化物, mainly sugars) to wherever they are needed.

In a **transverse section** 横切面 (a cut straight across) of a **dicotyledonous plant** 双子叶植物:

- in the **stem**, xylem and phloem sit together in bundles near the outside, with xylem on the inside of each bundle.
- in the **root**, the xylem is in the centre, often in a star shape, with phloem between the arms.
- in the **leaf**, both are found in the veins.

When you draw a plan diagram, you draw only the outlines of the tissues, not the single cells.



Where the tissues sit: stem bundles form a ring near the edge (xylem inside); in the root the xylem makes a central star



The wood of a tree trunk is xylem; each ring is one year of growth

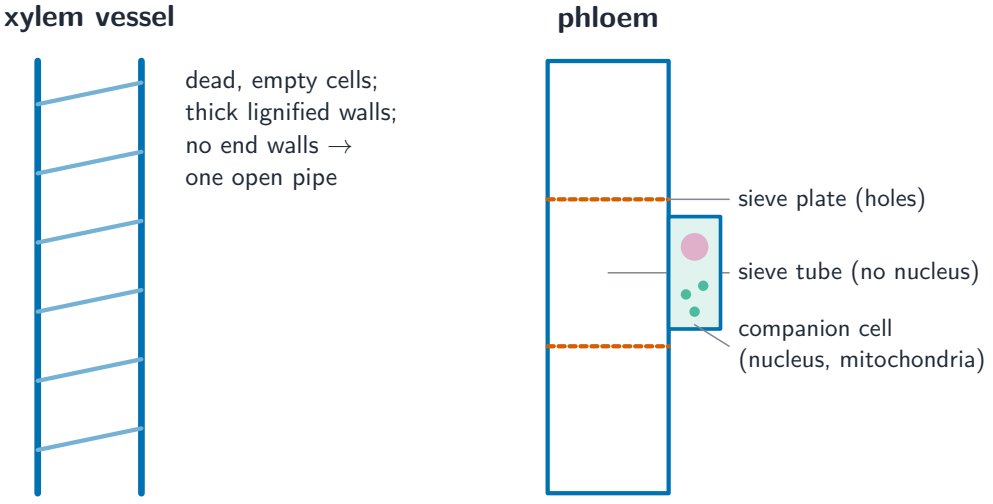
Xylem vessels

Xylem water-carrying tubes are called **vessels** 导管. They are made of dead, empty cells joined end to end, with the end walls gone, so they form one long open pipe. Their walls are thickened and waterproofed with **lignin** 木质素. So the structure suits the job: the hollow, open tube with no contents lets water flow fast, and the lignin gives strength and support.

Phloem sieve tubes and companion cells

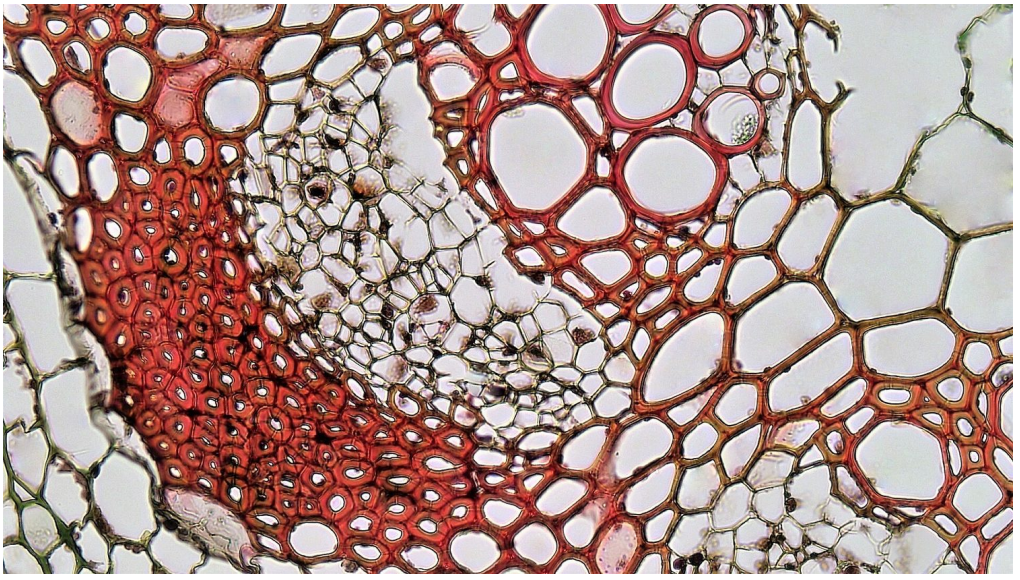
Phloem food-carrying tubes are called **sieve tubes** 筛管. They are living cells joined end to end, but their end walls are not gone — they become **sieve plates** 筛板 with many holes that sap flows through. To leave room for flow, a sieve tube cell loses most of its contents and has no nucleus.

Beside each sieve tube is a **companion cell** 伴胞. It keeps its **nucleus** 细胞核 and has many **mitochondria** 线粒体. It does the living work for the sieve tube and loads sugars into it.



Xylem is a dead, open pipe; phloem is living sieve tubes with sieve plates and companion cells

This is what a real **vascular bundle** 维管束 looks like under the microscope, in a stained section of a young sunflower stem:



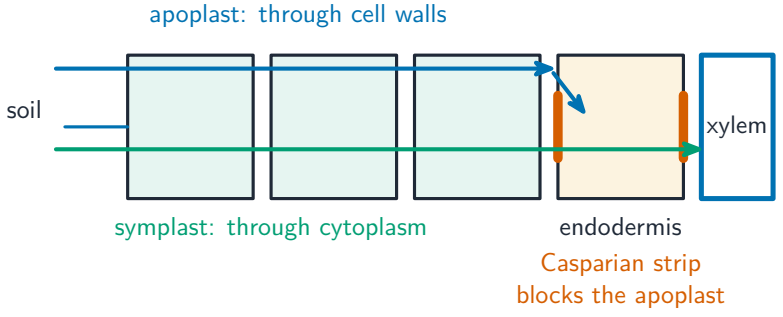
*A real **vascular bundle** in section: the big red cells are xylem vessels (thick lignified walls); the smaller cells above are phloem*

Water from the soil to the xylem

Water enters a **root hair cell** 根毛细胞 by **osmosis** 渗透, because the root hair has a lower water potential than the soil water. Water then crosses the root to the xylem by two pathways:

- the **apoplast pathway** 质外体途径 — water moves through the **cell walls** 细胞壁 (made of **cellulose** 纤维素) and the spaces between cells, without entering the cytoplasm. This is fast.
- the **symplast pathway** 共质体途径 — water moves through the cytoplasm of cells, passing from cell to cell through the **plasmodesmata** 胞间连丝.

At a ring of cells called the **endodermis** 内皮层, the apoplast pathway is blocked by the **Casparian strip** 凯氏带, a waterproof band of **suberin** 木栓质. This forces all the water through the cell membranes, which lets the plant control what enters the xylem.



The apoplast goes through the walls, the symplast through the cytoplasm; the Casparian strip blocks the apoplast at the endodermis

Transpiration and the movement of water up the xylem

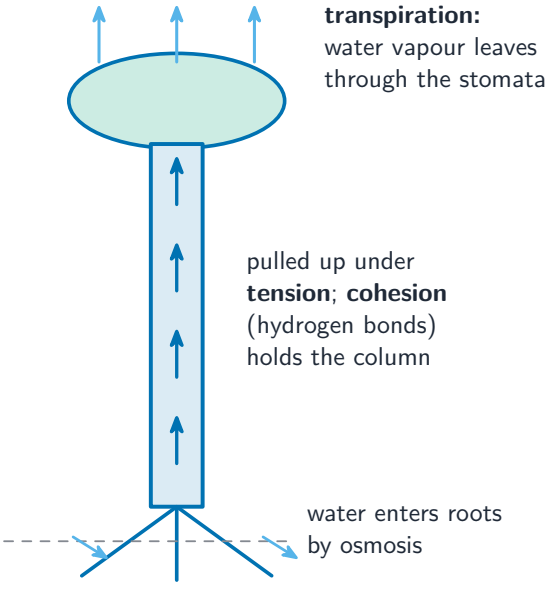
Transpiration 蒸腾作用 is the loss of water vapour from a plant. Water **evaporates** (turns to vapour) from the wet cell surfaces inside the leaf — this is **evaporation** 蒸发. The **water vapour** 水蒸气 then **diffuses** 扩散 out through the **stomata** 气孔 into the **atmosphere** 大气.

This loss at the top pulls water up the xylem in a continuous column. It works because of hydrogen bonding between water molecules:

- water molecules attract each other through **hydrogen bonds** 氢键, so they stick together. This sticking is **cohesion** 内聚力, and it lets the whole column be pulled

up under **tension** 张力 (the cohesion–tension idea).

- water molecules also stick to the cellulose of the cell walls. This is **adhesion** 附着力, which helps hold the column in place.

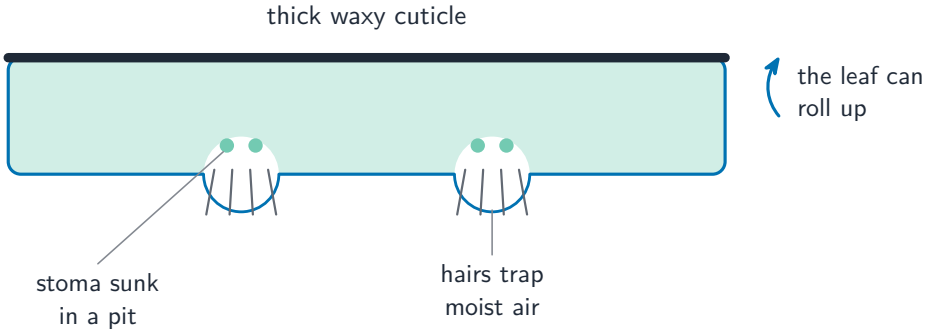


Transpiration at the leaf pulls the whole water column up the xylem; cohesion (hydrogen bonds) keeps it together

Worked example. A plant cell with a water potential of −800 kPa is placed in a solution of water potential −400 kPa. Which way does water move, and what happens to the cell? Water always moves **down** a water potential gradient - from the **less negative** (higher) value to the **more negative** (lower) one. The solution at −400 is **higher** than the cell at −800, so water moves **into** the cell. The cell swells, the protoplast presses on the cell wall, the **pressure potential** rises, and the cell becomes **turgid**. Two traps: −400 is **greater** than −800, which is the sign error that reverses half of all answers; and pure water is the maximum at **0**, so every solution is negative and a water potential can never rise above zero.

Xerophytes

A **xerophyte** 旱生植物 is a plant **adapted** 适应 to live where water is scarce. Its leaves reduce water loss by transpiration in several ways: a thick waxy **cuticle** 角质层, **stomata** sunk in pits, hairs that trap moist air, and leaves that can roll up. You should be able to draw a labelled leaf section showing these features.



Xerophyte leaves cut water loss: a thick cuticle, sunken stomata, trapped moist air and rolling up

Translocation: moving assimilates in the phloem

Assimilates such as **sucrose** 蔗糖 and **amino acids** 氨基酸 are carried in the phloem from a **source** 源 to a **sink** 库.

- a **source** is where the assimilate is made or released (for example a photosynthesising leaf).
- a **sink** is where it is used or stored (for example a growing root).

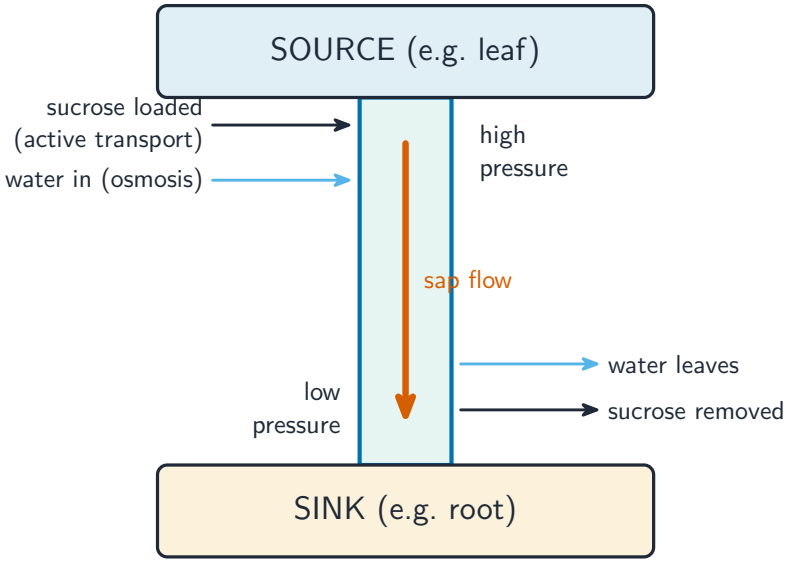
Loading at the source

Companion cells load sucrose into the sieve tubes against its concentration gradient. They use **proton pumps** 质子泵 to pump hydrogen ions out, then **cotransporter proteins** 协同运输蛋白 bring sucrose back in together with those ions. This is a form of **active transport** 主动运输.

Loading sucrose lowers the **water potential** 水势 inside the sieve tube, so water follows by osmosis. This raises the **hydrostatic pressure** 静水压 there.

Mass flow

At the sink, sucrose is removed, so the water potential rises, water leaves, and the pressure falls. The result is a pressure difference between source (high) and sink (low). Sap flows from high to low pressure down this gradient. This pressure-driven flow is called **mass flow** 集流.



Loading sucrose at the source raises the pressure; sap then flows by mass flow to the sink

Exam tips

- State the tissue for each direction: **xylem** carries water up (dead, lignified), **phloem** carries assimilates both ways (living sieve tubes + companion cells).
- Explain water movement by **cohesion-tension**: transpiration pulls a continuous water column held by cohesion (H-bonds) and adhesion.
- List the factors affecting **transpiration rate** (light, temperature, humidity, air movement) and how a potometer measures uptake.
- For **xerophytes**, link each adaptation (thick cuticle, sunken stomata, rolled leaves, hairs) to reduced water loss.

Transport in mammals

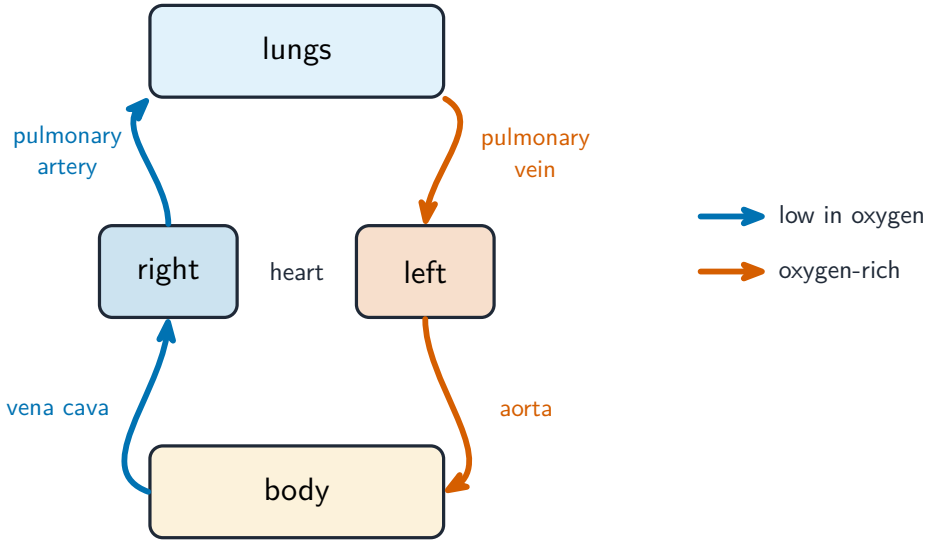
A-Level Biology

The circulatory system

Mammals have a **closed double circulation**. "Closed" means the blood stays inside **blood vessels** 血管 the whole time. "Double" means the blood passes through the **heart** 心脏 twice for each full trip around the body. This gives two linked loops, so we call it a **double circulation** 双循环 and the whole thing a **circulatory system** 循环系统:

- the **pulmonary circulation** 肺循环 carries blood from the heart to the lungs and back.
- the **systemic circulation** 体循环 carries blood from the heart to the rest of the body and back.

Blood travels in this order: **arteries** 动脉 → **arterioles** 小动脉 → **capillaries** 毛细血管 → **venules** 小静脉 → **veins** 静脉.



Double circulation: the pulmonary loop goes to the lungs, the systemic loop to the body; blood passes through the heart twice

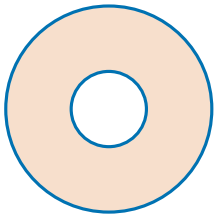
The main vessels are:

- the **pulmonary artery** 肺动脉 — carries blood low in **oxygen** 氧气 from the heart to the lungs.

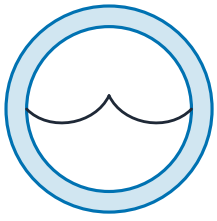
- the **pulmonary vein** 肺静脉 — carries oxygen-rich blood from the lungs back to the heart.
- the **aorta** 主动脉 — the big artery that carries oxygen-rich blood from the heart to the body.
- the **vena cava** 腔静脉 — the big vein that returns oxygen-poor blood from the body to the heart.

How the vessels suit their jobs

Vessel	Structure	Function
artery	thick wall of **muscle** 肌肉 and **elastic** 弹性 fibres; narrow **lumen** 管腔 (the space inside)	carries blood at high pressure away from the heart; elastic walls stretch and recoil to smooth the flow; muscle controls the flow
capillary	wall just one cell thick; very narrow	short distance for exchange of substances between blood and cells
vein	thin wall; wide lumen; has **valves** 瓣膜	returns blood at low pressure to the heart; valves stop blood flowing backwards



artery
thick muscular, elastic wall; narrow lumen (high pressure)



vein
thin wall; wide lumen; valves (low pressure)

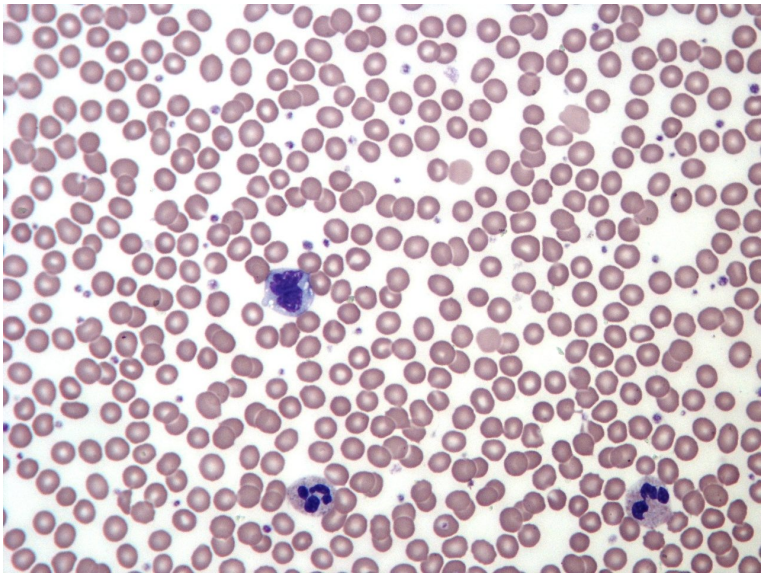


capillary
wall one cell thick; very narrow (exchange)

An artery has a thick wall and narrow lumen; a vein a thin wall, wide lumen and valves; a capillary is one cell thick

Blood cells

You should recognise: **red blood cells** 红细胞 (which carry oxygen), and three white blood cells — **monocytes** 单核细胞, **neutrophils** 中性粒细胞 and **lymphocytes** 淋巴细胞.

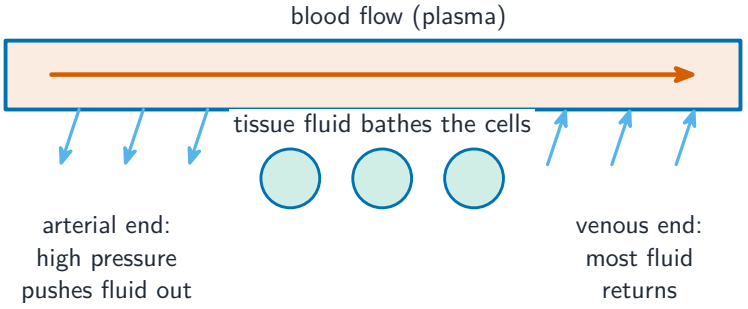


*A stained **blood smear**: many small red cells, plus a lymphocyte (centre) and neutrophils (lobed nucleus) — the white cells are larger and have a nucleus*

Water, plasma and tissue fluid

Water is the main part of blood. It is a good **solvent** 溶剂, so it carries dissolved substances, and it has a high **specific heat capacity** 比热容, so the blood's temperature stays steady.

At the start of a capillary, the high blood pressure pushes liquid (but not the cells or large proteins) out of the **plasma** 血浆 and through the capillary wall. This liquid around the cells is **tissue fluid** 组织液. It supplies the cells with oxygen and glucose and carries waste away. Most of it returns to the capillary at the far end, where the pressure is lower.



High pressure at the arterial end pushes fluid out to form tissue fluid; most returns at the venous end

Transport of oxygen and carbon dioxide

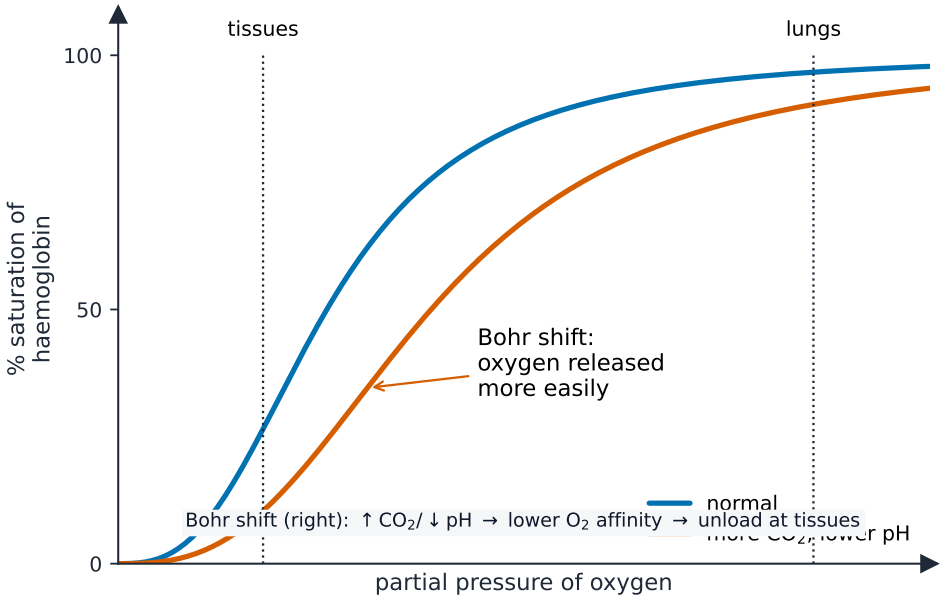
Carrying oxygen

Oxygen is carried by **haemoglobin** 血红蛋白 in the red blood cells. We show how well haemoglobin holds oxygen with the **oxygen dissociation curve** 氧解离曲线. This S-shaped graph plots the **saturation** 饱和度 (how full of oxygen the haemoglobin is) against the **partial pressure** 分压 of oxygen:

- where the partial pressure of oxygen is **high** (in the lungs), haemoglobin loads up and becomes almost fully saturated.
- where it is **low** (in respiring tissues), haemoglobin unloads its oxygen for the cells to use.

The Bohr shift

When tissues are very active, they release more **carbon dioxide** 二氧化碳, which lowers the pH. This makes haemoglobin release oxygen more easily, so the curve moves to the right. This helpful change is the **Bohr shift** 波尔位移: oxygen is given up exactly where it is most needed.



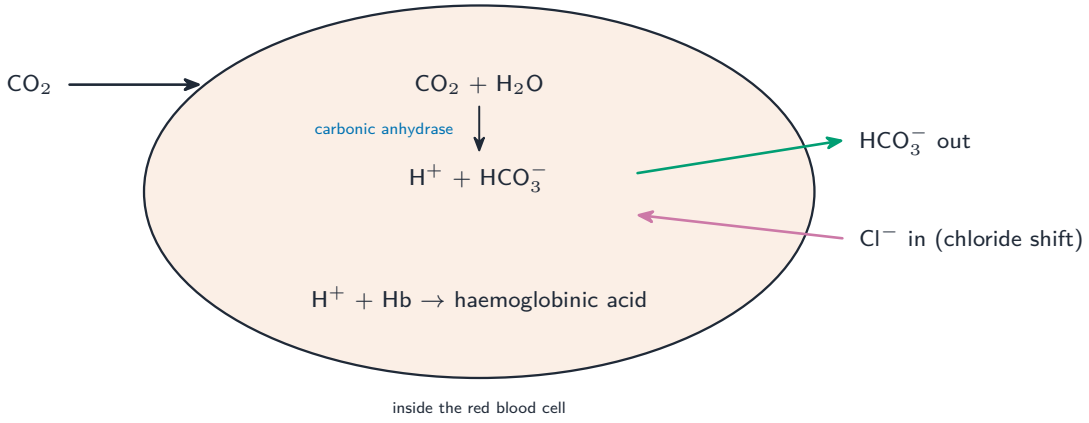
Haemoglobin loads oxygen in the lungs and unloads it in the tissues; the Bohr shift moves the curve right so more is released

Carrying carbon dioxide

A little carbon dioxide dissolves straight into the plasma, but most is carried after a reaction inside the red blood cells:

1. the **enzyme** 酶 **carbonic anhydrase** 碳酸酐酶 speeds up the reaction of carbon dioxide with water to make carbonic acid.
2. the carbonic acid splits into hydrogen ions and **hydrogencarbonate ions** 碳酸氢根离子.
3. the hydrogencarbonate ions move out into the plasma. This is the main way carbon dioxide is carried.
4. to keep the charge balanced, **chloride ions** 氯离子 move into the red blood cells. This movement is the **chloride shift** 氯转移.
5. the hydrogen ions join haemoglobin to form **haemoglobinic acid** 血红蛋白酸; this mops up the hydrogen ions and keeps the pH steady.

Some carbon dioxide also joins haemoglobin directly to form **carbamino-haemoglobin** 氨甲酰血红蛋白.



Most CO_2 travels as hydrogencarbonate ions; the chloride shift keeps the charge balanced and haemoglobin mops up the H^+

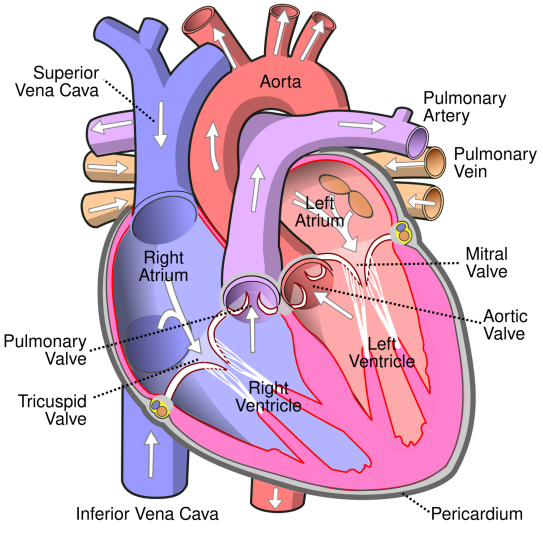
Worked example. At the lungs the partial pressure of oxygen is about 12 kPa and haemoglobin is about 98% saturated; in an exercising muscle it is about 3 kPa and saturation falls to about 40%. How much oxygen is unloaded, and why is the curve S-shaped? Subtract the saturations: $98 - 40 = \mathbf{58\%}$ of the haemoglobin's oxygen is released in the muscle. The **S shape** comes from **cooperative binding** - the first oxygen to bind changes haemoglobin's shape so the next ones bind more easily, which is why the middle of the curve is so steep. That steepness is the point: a **small** fall in partial pressure in a respiring tissue causes a **large** release of oxygen. The extra carbon dioxide in an exercising muscle lowers the pH and shifts the curve **right**, so even more is unloaded at the same partial pressure - that is the **Bohr shift** doing its job.

The heart

Structure

The heart has four chambers. The two upper chambers are the **atria** 心房 (singular: atrium); they have thin walls because they only push blood down into the chambers below. The two lower chambers are the **ventricles** 心室; they have thick muscular walls because they pump blood out of the heart.

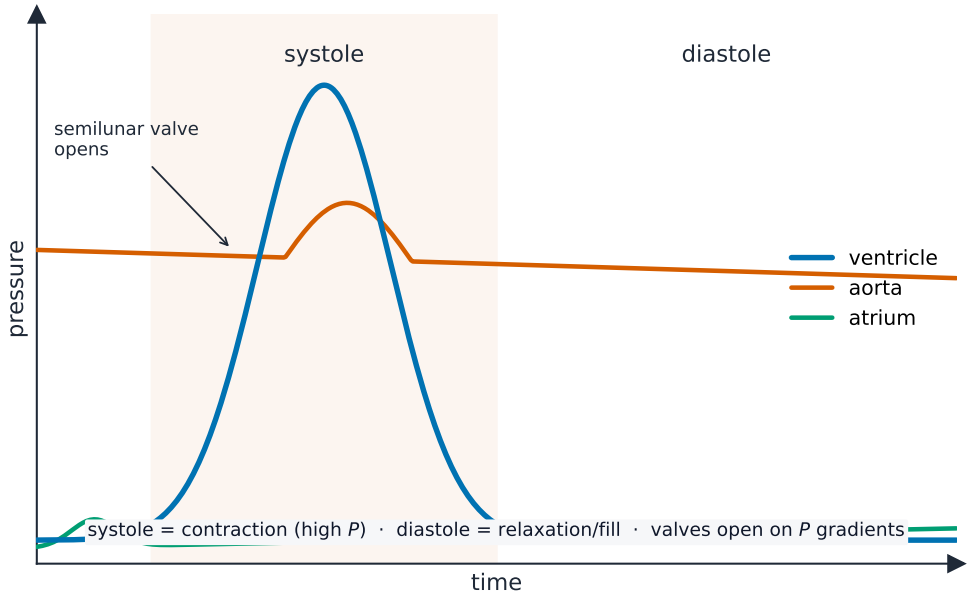
The left ventricle wall is thicker than the right ventricle wall, because the left side must pump blood all the way round the body, while the right side only pumps to the nearby lungs.



The four chambers, the valves and the main vessels; the left ventricle wall is the thickest

The cardiac cycle

One heartbeat is the **cardiac cycle** 心动周期. It has two parts: **systole** 收缩期 (when heart muscle contracts) and **diastole** 舒张期 (when it relaxes and fills). When a chamber contracts, the pressure inside rises; this pressure change opens and closes the valves so that blood flows one way only.

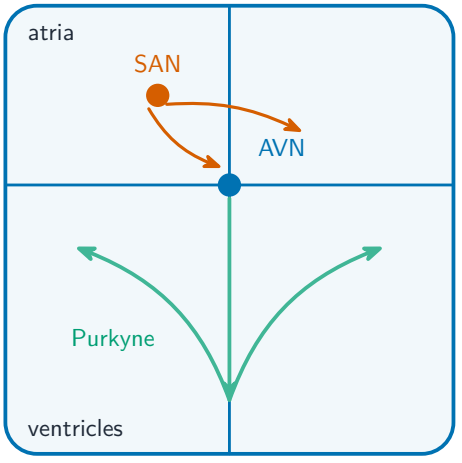


The ventricle pressure spikes in systole and pushes blood into the aorta; a valve opens or closes whenever two pressure curves cross

Controlling the heartbeat

The heart sets its own rhythm:

- the **sinoatrial node** 窦房结 in the right atrium is the pacemaker. It sends out a wave of electrical excitation that spreads across the atria and makes them contract.
- the **atrioventricular node** 房室结 picks up the wave, holds it back for a moment (so the atria empty first), then passes it on.
- the **Purkyne tissue** 浦肯野组织 carries the wave down and through the ventricle walls, so the ventricles contract from the bottom upwards and push blood out.



SAN fires → atria contract → AVN delays → Purkyne → ventricles contract from the bottom up

The SAN sets the rhythm; the wave passes to the AVN, then the Purkyne tissue makes the ventricles contract bottom-up

Exam tips

- Trace the **double circulation** and name the chambers, valves and vessels; the left ventricle wall is thicker (pumps further at higher pressure).
- Read the **cardiac-cycle** pressure graph: a valve opens/closes when the pressures across it cross over.
- Explain the **oxygen dissociation curve** (S-shaped) and the **Bohr shift** (more CO2 shifts it right, unloading more O2 to active tissue).
- Most CO2 is carried as **hydrogencarbonate** (via the chloride shift).

Gas exchange

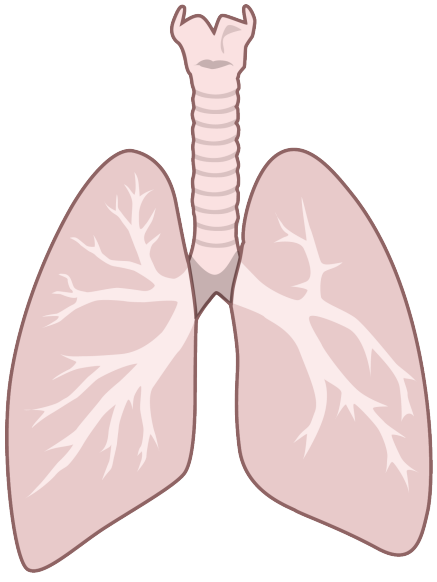
A-Level Biology

The gas exchange system

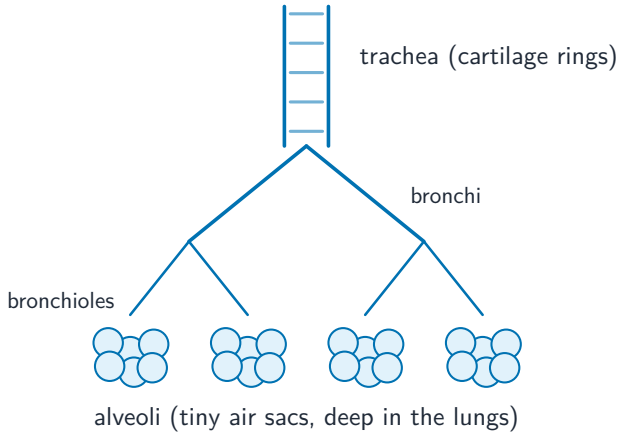
Your body needs to take in oxygen and get rid of carbon dioxide. This swap happens in the **gas exchange** 气体交换 system. Air follows this path into the body:

- down the **trachea** 气管 (the windpipe),
- into two **bronchi** 支气管 (one to each lung),
- into many smaller **bronchioles** 细支气管,
- and finally into tiny air sacs called **alveoli** 肺泡, deep in the **lungs** 肺.

Each alveolus is wrapped in a network of **capillaries** 毛细血管, so air and blood are brought very close together.



The human lungs, where gas exchange takes place across millions of alveoli

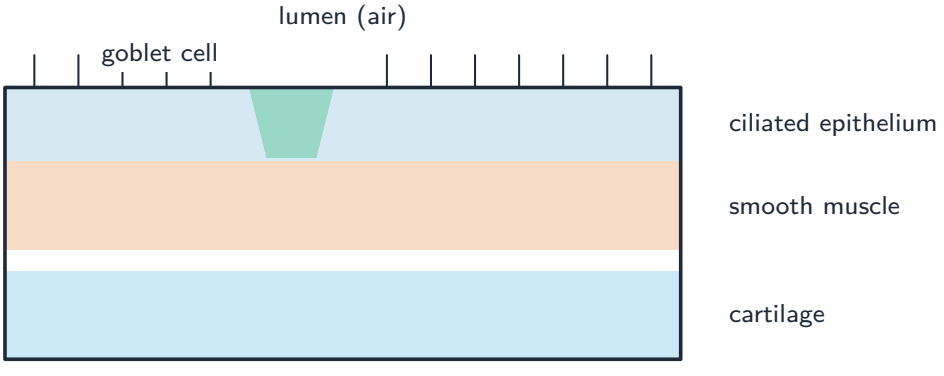


Air passes down the trachea, into the bronchi and bronchioles, to the alveoli

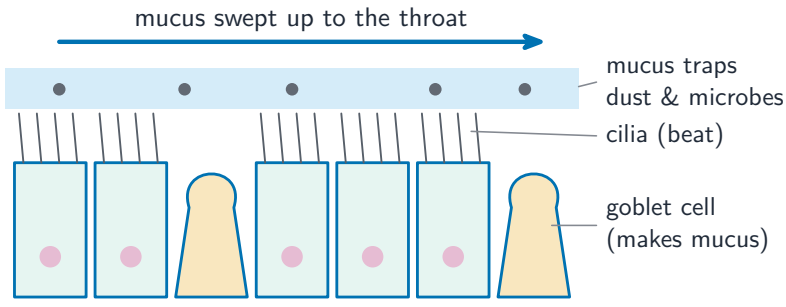
The tissues of the airways and what they do

Tissue	Where it is	Function
cartilage 软骨	C-shaped rings in the trachea and bronchi	holds the airway open so it cannot collapse when you breathe in
ciliated epithelium 纤毛上皮	lining the trachea and bronchi	tiny hairs called **cilia** 纤毛 beat to sweep **mucus** 黏液 up towards the throat, away from the lungs
goblet cells 杯状细胞 and **mucous glands** 黏液腺	in the lining of the airways	make the mucus, which traps dust and **microbes** 微生物 that you breathe in
smooth muscle 平滑肌	in the walls of bronchi and bronchioles	contracts to make the airway narrower
elastic fibres 弹性纤维	in the airway and alveolus walls	stretch when you breathe in, then spring back to help push air out
squamous epithelium 扁平上皮	the very thin, flat lining of the alveoli	gives a very short distance for gases to cross

The cilia, goblet cells and mucous glands work together to keep the lungs clean and healthy: the mucus traps dirt and microbes, and the cilia carry it away to be swallowed.



A section through the airway wall: cilia and goblet cells line the lumen, with smooth muscle and supporting cartilage behind



Goblet cells make mucus that traps dust and microbes; the cilia sweep it up to the throat

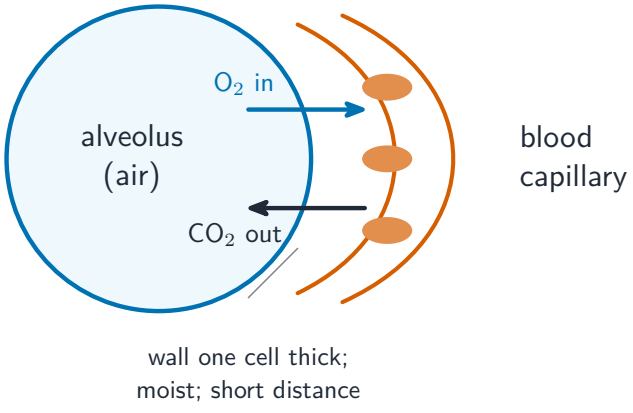
Gas exchange in the alveoli

The alveoli are excellent surfaces for exchanging gases, because they have:

- a very large total surface area (millions of tiny sacs),
- very thin walls — the squamous epithelium of the alveolus and the capillary wall are each only one cell thick, so the distance to cross is tiny,
- a rich blood supply from the capillary network,
- a moist lining, so gases dissolve before crossing.

Gases move by **diffusion** 扩散 down their **concentration gradients** 浓度梯度:

- **oxygen** 氧气 is at a high concentration in the alveolar air and a low concentration in the blood, so it diffuses from the air into the blood.
- **carbon dioxide** 二氧化碳 is at a high concentration in the blood and a low concentration in the alveolar air, so it diffuses from the blood into the air to be breathed out.



Across the thin, moist wall, oxygen diffuses into the blood and carbon dioxide diffuses out

Under the microscope, real lung tissue looks like a fine pink lace. The many open spaces are the alveoli, and the thin pink lines between them are the walls where gas exchange happens:



Real lung tissue stained for the microscope: the open spaces are alveoli and the thin pink walls are where gases cross; a small airway sits in the centre

Breathing keeps fresh air in the alveoli, and the flowing blood keeps carrying gases away. Both of these keep the concentration gradients steep, so gas exchange stays fast.

Worked example. Use Fick's law to explain why the alveoli allow such rapid gas exchange. Fick's law makes the rate of diffusion proportional to

$$\frac{\text{surface area} \times \text{concentration difference}}{\text{diffusion distance}}$$

so a fast rate needs all three terms working for it. **Surface area**: millions of alveoli give a huge total area, roughly 70 m^2. **Diffusion distance**: the alveolar epithelium and the capillary endothelium are each one flattened cell thick, so oxygen crosses in under 1 μm. **Concentration difference**: ventilation constantly refreshes the air while the circulation constantly carries oxygenated blood away, so a **steep gradient is maintained**. Tie each adaptation to the **term of the equation** it serves - listing "big surface area, thin walls, good blood supply" without linking them to Fick's law is the weaker answer.

Exam tips

- Link each **alveolar** feature to fast diffusion: large surface area, thin (one-cell) walls, moist surface, good blood supply — a "surface area, short distance, steep gradient" answer.
- Match airway tissue to function: cartilage (holds airways open), ciliated + goblet cells (trap and sweep mucus), smooth muscle and elastic fibres.
- Frame answers with **Fick's law** ideas: rate $\propto$ surface area $\times$ concentration difference $\div$ distance.

Infectious diseases

A-Level Biology

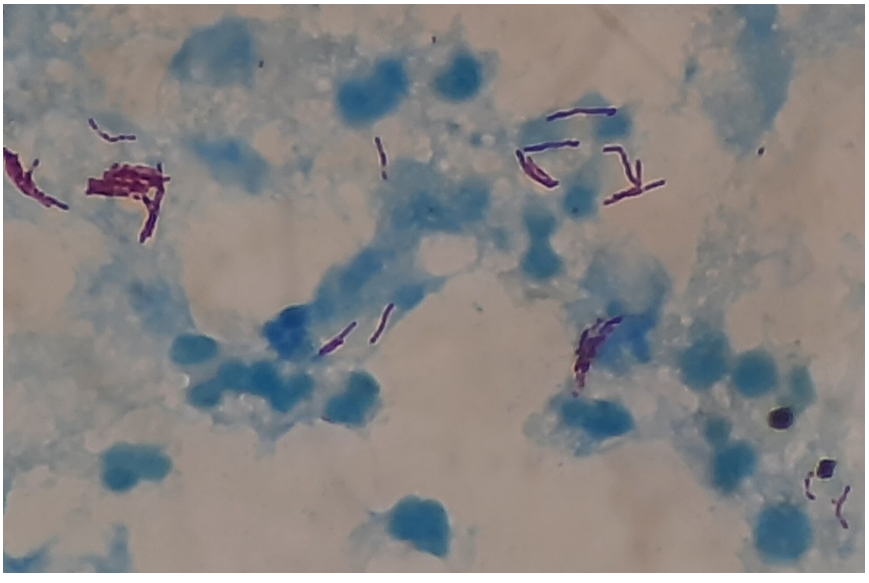
What causes infectious disease

An **infectious disease** 传染病 is caused by a **pathogen** 病原体 — an organism that lives in or on a host and causes harm. Infectious diseases are **transmissible**: the pathogen can be **transmitted** 传播 (passed) from one person to another.

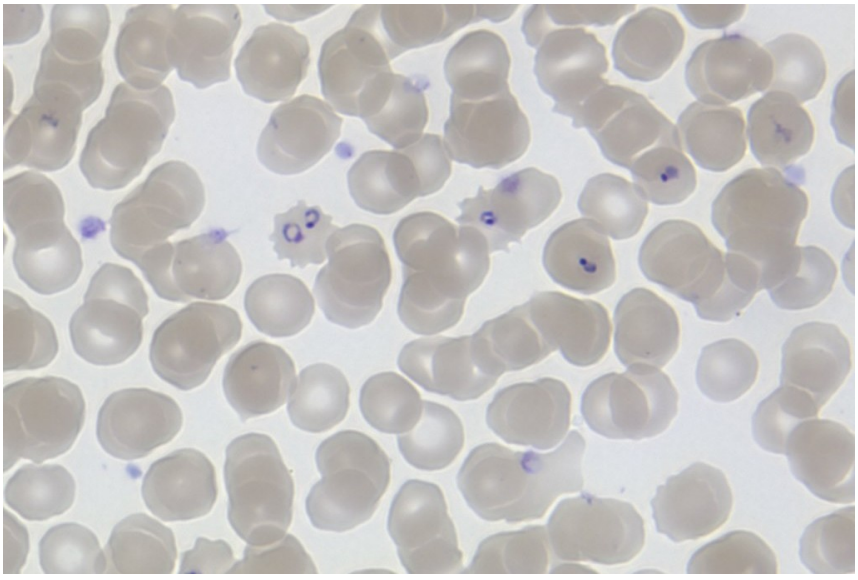
You need to know four diseases, the pathogen that causes each, and how each spreads.

Disease	Pathogen and type	How it spreads
cholera 霍乱	the **bacterium** 细菌 *Vibrio cholerae*	drinking water or food **contaminated** 污染 with **faeces** 粪便 (human waste)
malaria 疟疾	the **protoctist** 原生生物 *Plasmodium*	the bite of an infected **mosquito** 蚊子, which acts as a **vector** 媒介 (a carrier of the pathogen); also through infected blood
tuberculosis (TB) 结核病	the bacterium *Mycobacterium*	tiny airborne **droplets** 飞沫 from coughs and sneezes; spreads fast where people are crowded
HIV/AIDS	the **virus** 病毒 HIV (which leads to **AIDS** 艾滋病)	unprotected sex, infected blood (for example shared needles), and from mother to baby

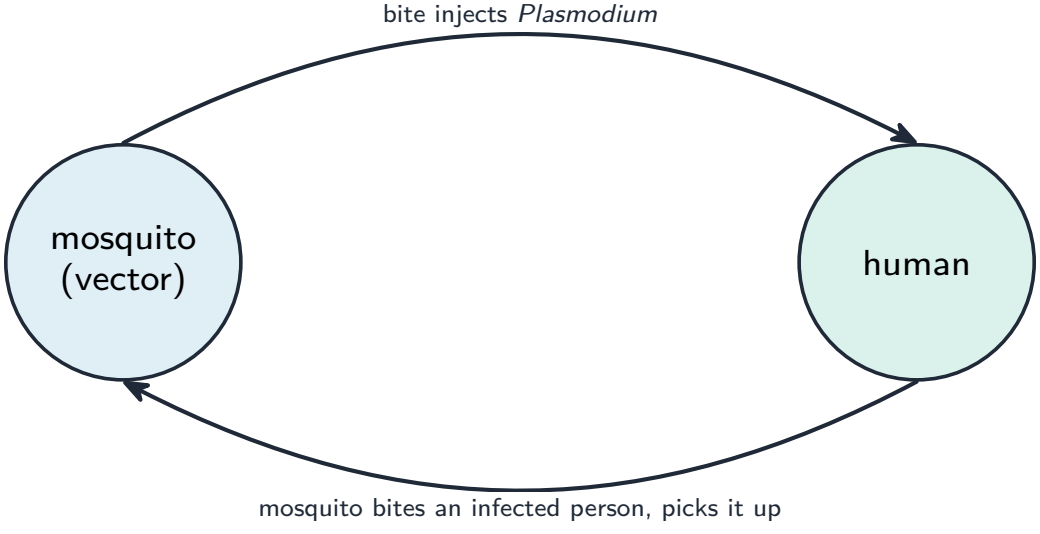
HIV infects and destroys certain white blood cells, so it slowly weakens the body's **immune system** 免疫系统.



The bacterium Mycobacterium (red rods) that causes tuberculosis, stained in a sputum sample



The protoctist Plasmodium (small dark rings) that causes malaria, living inside red blood cells



Malaria needs a mosquito vector: the parasite passes from human to mosquito to human with each bite — which is why control targets the mosquito

Preventing and controlling these diseases

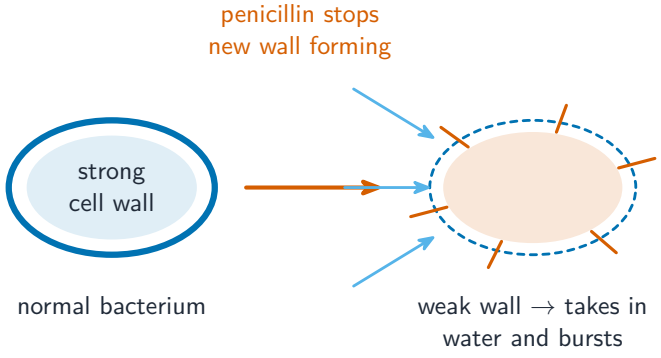
Control has **biological**, **social** and **economic** sides — the science of the pathogen, people's behaviour and education, and the money and resources available. Examples:

- **cholera**: provide clean water and proper sewage treatment; good hygiene; **vaccines** 疫苗 in some areas.
- **malaria**: sleep under nets; remove pools of still water where mosquitoes breed; spray **insecticides** 杀虫剂; take anti-malarial drugs.
- **TB**: find and treat infected people with a long course of antibiotics; give the BCG vaccine; reduce overcrowding; trace contacts of patients.
- **HIV**: use condoms; use clean needles; test donated blood; educate people. There is no cure and no vaccine yet, but drugs can slow the virus down.

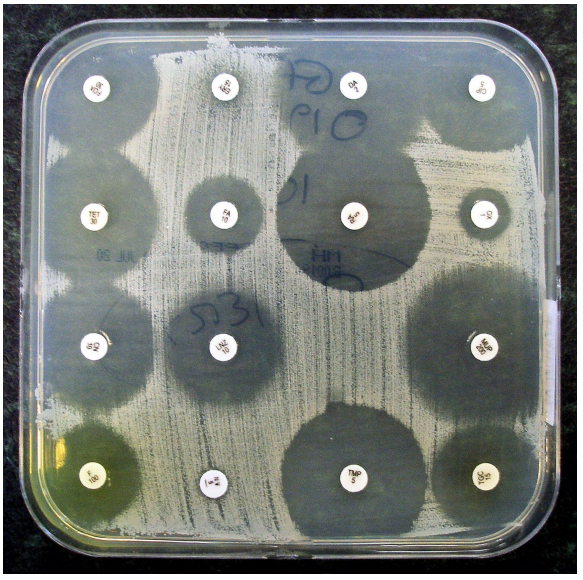
In every case, cost (economic), people's willingness to change behaviour (social) and the supply of drugs or vaccines (biological) all affect how well a disease can be controlled.

Antibiotics

An **antibiotic** 抗生素 is a drug that kills bacteria or stops them growing. For example, **penicillin** 青霉素 stops bacteria from building their **cell walls** 细胞壁. As the bacterium grows, its weak wall cannot hold it, so the cell takes in water and bursts.



Penicillin stops new cell wall forming, so the bacterium takes in water and bursts



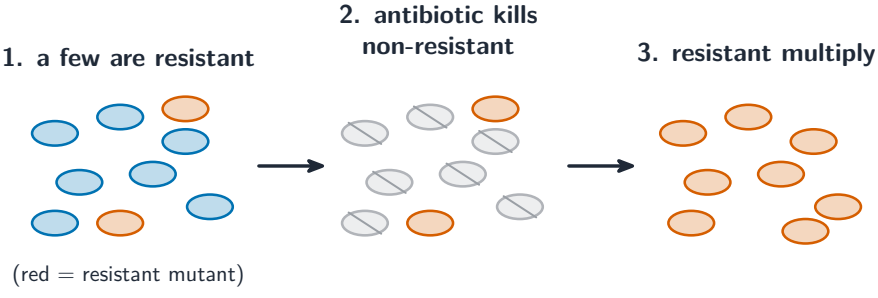
How we test which antibiotic works. Each paper disc holds a different antibiotic. A clear ring means that antibiotic killed the bacteria around it — the bigger the ring, the more effective the drug. Discs with no ring are ones the bacteria resist

Antibiotics do **not** work against **viruses**. A virus has no cell wall and no chemical reactions of its own to attack — it simply uses the machinery of the host cell. So there is no antibiotic target in a virus.

Antibiotic resistance

Sometimes a **mutation** 突变 makes a bacterium **resistant** to an antibiotic, which means the antibiotic no longer kills it. This **resistance** 耐药性 is a serious problem:

- when an antibiotic is used, the non-resistant bacteria die, but any resistant ones survive and multiply. Over time, more and more bacteria carry the resistance.
- some infections then become very hard, or impossible, to treat.



Antibiotic resistance spreads by natural selection: the antibiotic kills the rest, so the resistant survivors take over

Steps to slow resistance down:

- only use antibiotics when they are really needed (not for viral illnesses such as colds).
- always finish the full course, so no bacteria are left alive.
- use the correct antibiotic for the infection.
- reduce the heavy use of antibiotics in farming.

Worked example. A patient with influenza is prescribed an antibiotic. Explain why it will not help, and why it is actively harmful. Antibiotics work by attacking structures or processes that **bacteria** have and human cells do not - penicillin, for instance, blocks **cell wall** synthesis, so the growing bacterium bursts under osmotic pressure. A **virus** has no cell wall, no ribosomes of its own and no metabolism: it replicates **inside** the host's cells using the host's machinery, so there is simply no bacterial target for the drug to attack. Taking it anyway exposes the patient's harmless resident bacteria to the antibiotic, killing the susceptible ones and **selecting for resistant** ones - which is how resistance spreads through a population. Name the **target that is missing**: "viruses are not alive" is not the reason and earns nothing.

Exam tips

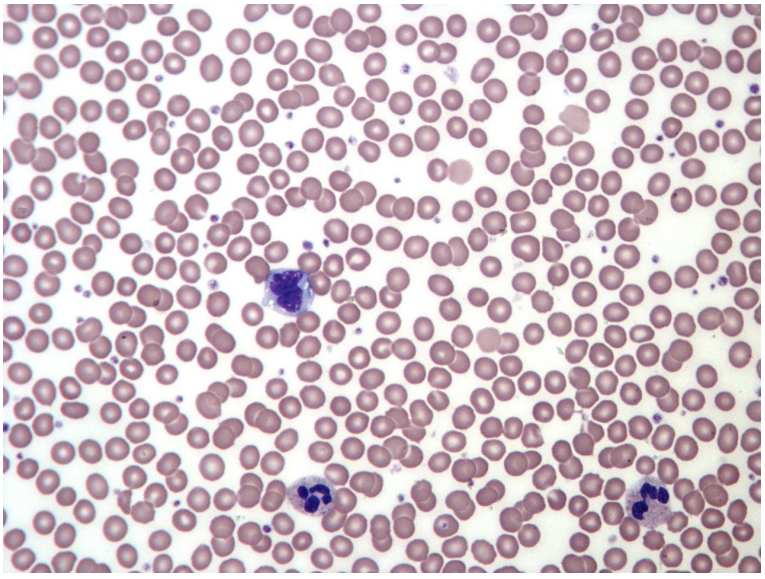
- For each named disease give the **pathogen and route** (cholera – bacterium, water; malaria – Plasmodium, mosquito vector; also TB, HIV/AIDS).
- Explain why **antibiotics** work on bacteria only (they target cell walls/enzymes, not viruses).
- Explain **antibiotic resistance** as natural selection: a resistant mutant survives and reproduces; reduce it by finishing courses and avoiding overuse.

Immunity

A-Level Biology

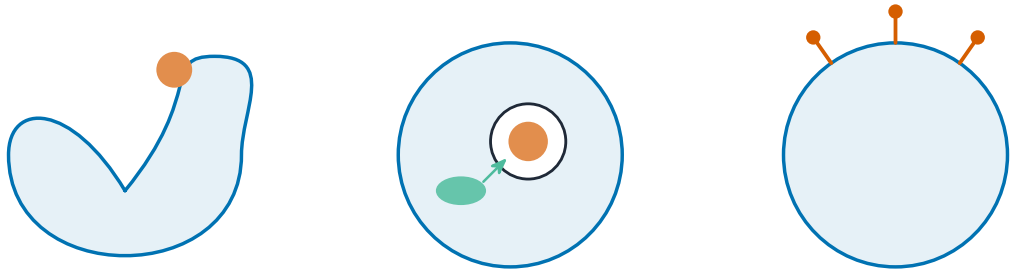
Phagocytes — the first defence

Your **immune system** 免疫系统 is the set of cells that defends the body against **pathogens** 病原体. The first cells to act are **phagocytes** 吞噬细胞, a group of white blood cells that includes **macrophages** 巨噬细胞 and **neutrophils** 中性粒细胞.



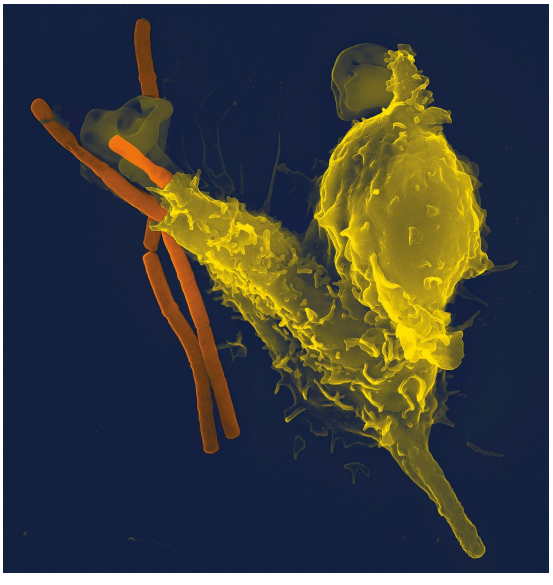
A blood smear: the large stained cells are white blood cells, scattered among the many red cells

A phagocyte destroys a pathogen by **phagocytosis** 吞噬作用: it surrounds the pathogen, takes it inside in a vesicle, and digests it with enzymes. After this, a macrophage displays parts of the pathogen on its own surface, ready to alert other immune cells.



Phagocytosis: the phagocyte engulfs the pathogen, digests it, then displays its antigen

This false-coloured electron micrograph captures the moment in real life: a phagocyte reaching out to grab and engulf rod-shaped bacteria.



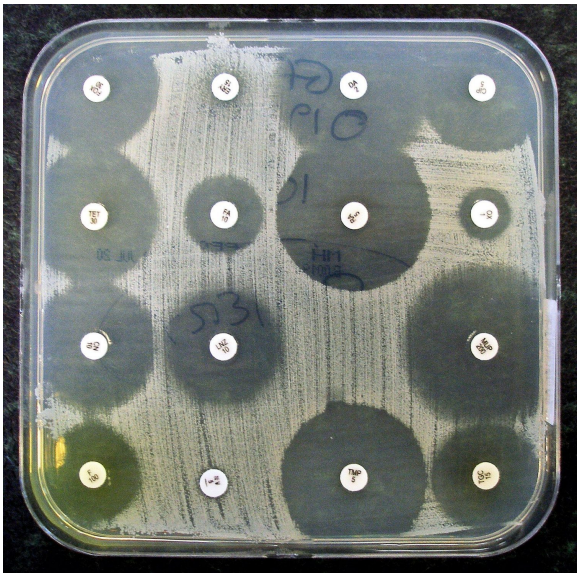
*A real **neutrophil** (yellow), one kind of phagocyte, reaching out to engulf rod-shaped bacteria (orange) by phagocytosis*

Antigens: self and non-self

An **antigen** 抗原 is a molecule (usually a protein) on a cell surface that the immune system can recognise.

- **self antigens** 自身抗原 are the body's own markers. The immune system learns to ignore them.

- **non-self antigens** 非自身抗原 are foreign, for example the antigens on a pathogen. These trigger an immune response.

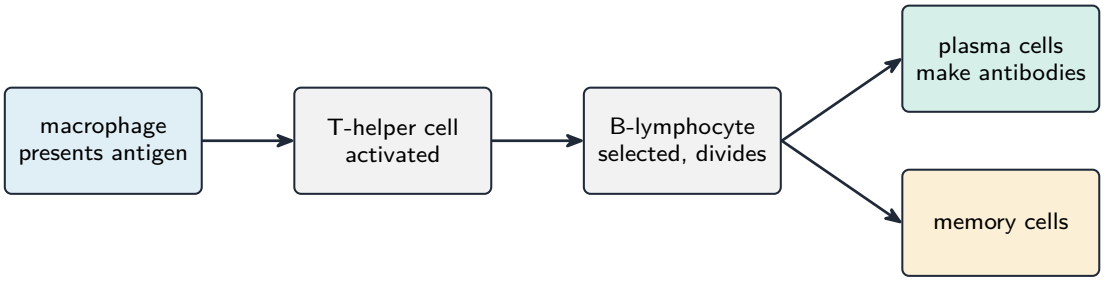


Clear zones on a culture plate: the immune system recognises non-self antigens and attacks foreign cells

The primary immune response

The first time a new pathogen enters, the body makes a slow **immune response** 免疫反应. The main steps are:

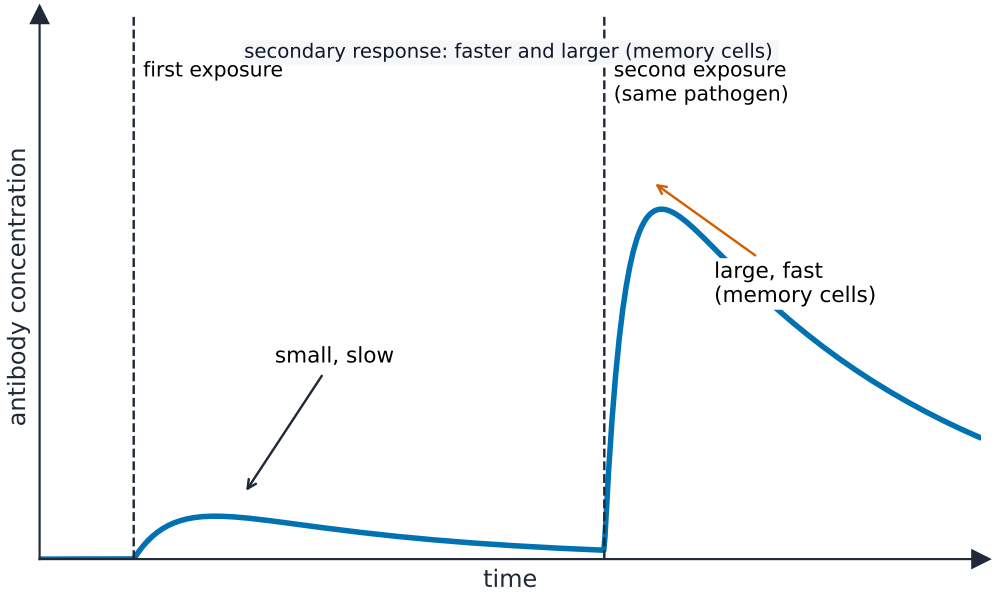
1. a macrophage engulfs the pathogen and displays its antigen.
2. **T-helper cells** 辅助性 T 细胞 recognise that antigen and become active. They release chemicals that switch on other cells.
3. **B-lymphocytes** 淋巴细胞 with a matching shape are selected. They divide to form **plasma cells** 浆细胞, which pour out **antibodies** 抗体, and **memory cells** 记忆细胞.
4. **T-killer cells** 杀伤性 T 细胞 destroy the body's own cells that have been infected.



Only the B-lymphocyte whose shape matches the antigen is selected and cloned — into antibody-making plasma cells and long-lived memory cells

Memory cells and long-term immunity

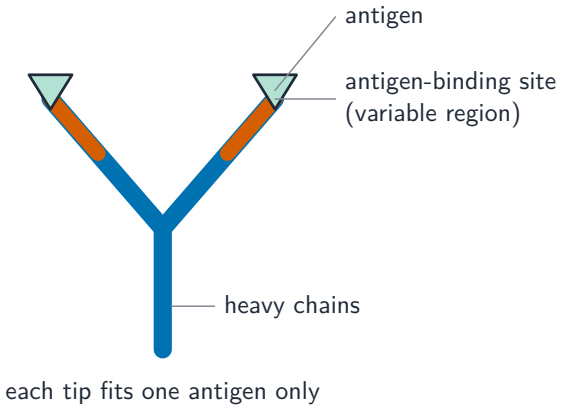
Memory cells stay in the body for years after the infection is over. If the **same** pathogen enters again, the memory cells start a **secondary** immune response that is much faster and larger than the first. The pathogen is destroyed before it can make you ill. This is what we mean by long-term immunity.



The secondary response is faster and larger, because memory cells are ready

Antibodies

An antibody is a Y-shaped protein made by plasma cells. The two tips of the Y are **antigen-binding sites** 抗原结合位点. Each site has a special shape, called the **variable region** 可变区, that fits one antigen only, like a lock and key.



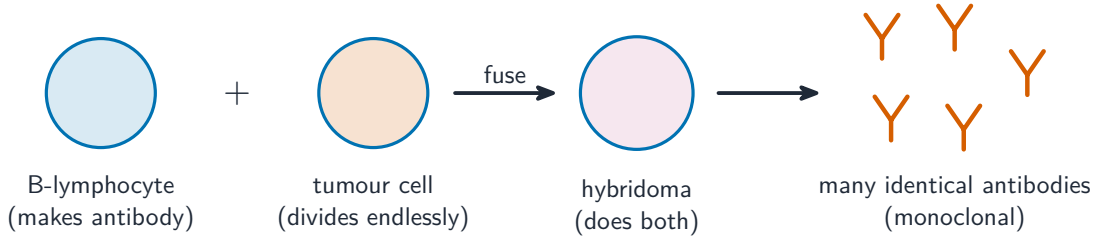
An antibody is Y-shaped; the two tips are the variable regions that bind one specific antigen

Antibodies help in several ways: they stick to antigens, clump pathogens together so they are easier to deal with, mark pathogens so phagocytes find them, and block harmful toxins.

Monoclonal antibodies

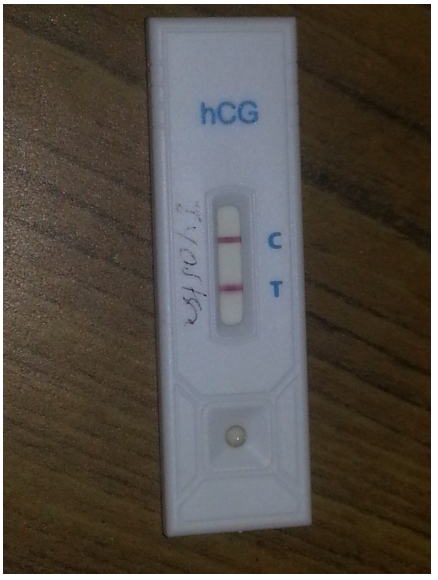
A **monoclonal antibody** 单克隆抗体 is a single type of antibody, all identical. They are made by the **hybridoma** 杂交瘤 method:

1. an animal is given an antigen, so it makes B-lymphocytes that produce the wanted antibody.
2. these B-lymphocytes are fused with tumour cells, which divide endlessly.
3. the fused cell (the hybridoma) both makes the antibody **and** divides without stopping, producing large amounts of one identical antibody.



Fusing a B-lymphocyte with a tumour cell makes a hybridoma that pours out one identical (monoclonal) antibody

Monoclonal antibodies are used in the **diagnosis** 诊断 of disease (to detect a specific molecule, as in a pregnancy test) and in treatment (to carry drugs to specific target cells, such as cancer cells).

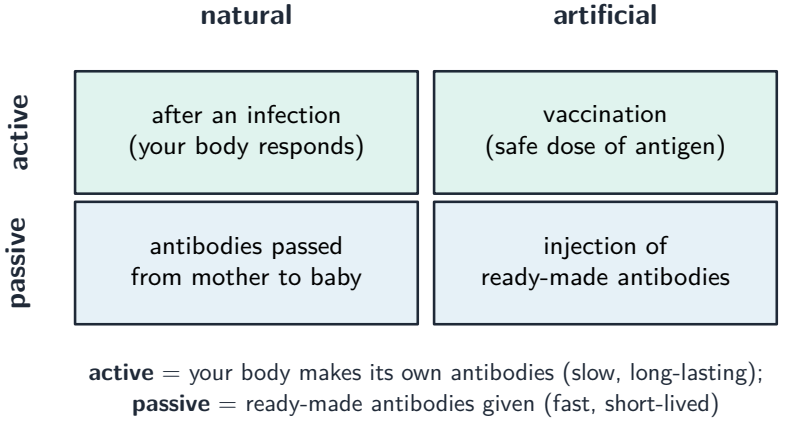


A pregnancy test is monoclonal antibodies at work. The antibodies are fixed at the T line and grab only one molecule — the pregnancy hormone hCG. The T line turns red only if hCG is in the urine; the C line always appears, to show the test ran properly

Types of immunity

Immunity can be active or passive, and natural or artificial.

- **active immunity** 主动免疫 — your own body meets an antigen and makes its own antibodies and memory cells. It is slow to start but long-lasting.
- **passive immunity** 被动免疫 — ready-made antibodies are given to you from outside. It works at once but does not last, because there are no memory cells.
- **natural immunity** 天然免疫 — gained in a natural way (active: after an infection; passive: antibodies passed from mother to baby).
- **artificial immunity** 人工免疫 — gained on purpose (active: your body is made to respond to a safe dose of antigen; passive: you are injected with ready-made antibodies).



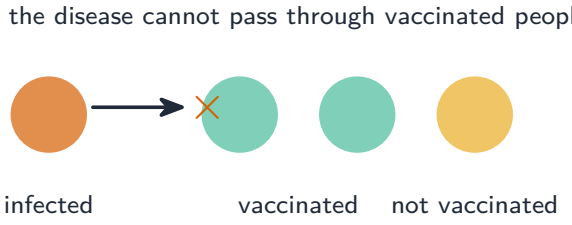
active = your body makes its own antibodies (slow, long-lasting);
passive = ready-made antibodies given (fast, short-lived)

Active immunity (your body responds) lasts; passive immunity (ready-made antibodies) is fast but short

Vaccination

A **vaccine** 疫苗 contains antigens — often a dead or weakened pathogen, or part of one. The antigens trigger a primary immune response and make memory cells, so you gain long-term immunity **without** becoming ill.

Vaccination 疫苗接种 programmes can control the spread of a disease across a population. If enough people are vaccinated, the pathogen cannot pass easily from person to person. This protects even the people who are not vaccinated, an effect called **herd immunity** 群体免疫.



Herd immunity: when enough people are vaccinated, the pathogen cannot reach the few who are not

Worked example. Classify each as active or passive, natural or artificial: (a) a baby receives antibodies in breast milk; (b) a child is vaccinated against measles; (c) someone recovers from chickenpox; (d) a patient bitten by a snake is given antivenom. Ask two questions each time. **Did the person make the antibodies themselves?** If yes it is **active**; if they were handed ready-made ones, it is **passive**. **Did it happen by chance or deliberately?** Naturally, or artificially. So (a) is **natural passive** - ready-made antibodies by a natural route; (b) is **artificial active** - the vaccine's antigens make the child produce their own; (c) is **natural active** - the infection made them produce their own; (d) is **artificial passive** - ready-made antibodies given deliberately. Only **active** immunity makes **memory cells**, which is exactly why passive immunity acts immediately but is **short-lived**.

Exam tips

- Distinguish **phagocytes** (engulf, non-specific) from lymphocytes: B cells → antibodies (humoral), T cells (helper/killer, cell-mediated).
- Label an **antibody** (variable region, antigen-binding site) and explain agglutination and neutralisation.
- Compare the **primary and secondary response** on a graph — memory cells make the secondary faster and larger.

- Distinguish **active vs passive** and **natural vs artificial** immunity with an example of each; explain vaccination and herd immunity.

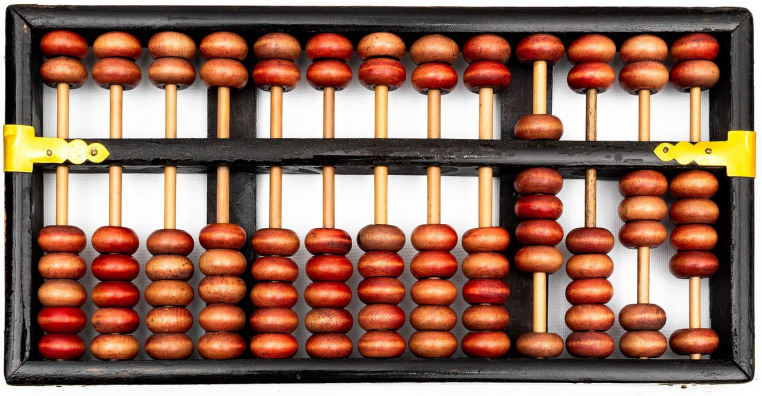
Information representation

A-Level Computer Science

Number systems

The three **number systems** 数制 you must use:

- **denary** 十进制 (decimal, base 10) — uses digits 0–9. Place values are powers of ten.
- **binary** 二进制 (base 2) — uses 0 and 1. Place values are powers of two. Every **byte** 字节 is 8 **bits** 位.
- **hexadecimal** 十六进制 (base 16) — uses 0–9 then A–F for 10–15. Each hex **digit** 数位 stands for exactly 4 bits.



An abacus represents numbers by place value — the same idea behind decimal, binary and hexadecimal

Conversions

Denary → binary: keep dividing by 2 and record the remainders, read bottom-up. Or subtract the largest **place value** 位值 (power of 2) that fits.

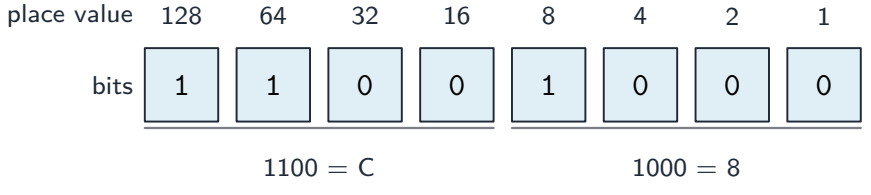
Example: 558_{10}: $558 = 512 + 32 + 8 + 4 + 2 = 2^9 + 2^5 + 2^3 + 2^2 + 2^1$. In 12 bits: `0010 0010 1110`.

Binary → hex: group the bits into **nibbles** 半字节 (4 bits) from the right and convert each. `0010 0010 1110` → `2 2 E` → `22E`.

Hex → binary: replace each hex digit with its 4-bit pattern. **Hex → denary**: multiply each digit by its place value. `22E` $= 2 \times 256 + 2 \times 16 + 14 = 558$.

Worked example. Convert denary 200 to 8-bit binary, then to hexadecimal.

$200 = 128 + 64 + 8$, so the binary is `11001000`. In nibbles, `1100 1000` = 12 and 8, i.e. C and 8, so the hexadecimal is `C8`.



$$200 = 128 + 64 + 8 = \texttt{11001000}_2 = \texttt{C8}_{16}$$

Reading 200 from its place values, then grouping the bits into nibbles to get hex C8

Binary vs decimal prefixes

Two prefix families look similar but differ — decimal (powers of 10) and binary (powers of 2):

Decimal (SI)	Binary (memory)
kilo $= 10^3$	kibi (Ki) $= 2^{10} = 1024$
mega $= 10^6$	mebi (Mi) $= 2^{20}$
giga $= 10^9$	gibi (Gi) $= 2^{30}$
tera $= 10^{12}$	tebi (Ti) $= 2^{40}$

So a tebibyte (TiB) is slightly more than a terabyte (TB). A "1 TB" drive holds 10^{12} bytes, but an operating system that reports in TiB shows a smaller number.

Binary arithmetic

Binary addition

Add column by column from the right, carrying as in denary:

Bit A	Bit B	Carry in	Sum bit	Carry out
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	1	0	0	1
1	1	1	1	1

Overflow 溢出 happens when the result needs more bits than the **register** 寄存器 can hold — the carry-out of the leftmost column is the overflow bit.

Binary subtraction

The usual way is **two's complement** 补码 addition: to do $A - B$, form the two's complement of B (invert every bit and add 1), then add, and discard any final carry-out.

To subtract 00011110 from 01100100 (unsigned 8-bit):

- two's complement of 00011110: invert $\rightarrow$ 11100001, add 1 $\rightarrow$ 11100010.
- add to 01100100: result 1 01000110 (9 bits) — discard the leading 1 $\rightarrow$ 01000110 = 70_{10}. Check: $100 - 30 = 70$. ✓

Two's complement signed integers

In an n-bit two's-complement number:

- the **most significant bit** 最高有效位 (MSB) is the **sign bit** 符号位: 0 = positive, 1 = negative.
- to read a negative number: invert every bit, add 1, then negate.

So 11100010 is negative; invert $\rightarrow$ 00011101, add 1 $\rightarrow$ 00011110 = 30, so it is -30. This is a **signed integer** 有符号整数 (unlike an **unsigned** 无符号 one). The range for n bits is -2^{n-1} to $+2^{n-1} - 1$; for 8 bits, -128 (10000000) to $+127$ (01111111).

The same bits mean different numbers depending on the agreed reading. As an **unsigned** integer every bit is a place value, so 8 bits run from 0 to 255; as a **signed** two's-complement integer the top bit is the sign, so the same 8 bits run from -128 to

+127. The pattern 11111111 is 255 read one way and −1 read the other — nothing in the bits themselves says which.

8 bits	as unsigned every bit a place value	as signed top bit is the sign
00000000	0	0
01111111	127	+127
10000000	128	−128
11111111	255	−1
range for n bits	$0 \dots 2^n - 1$	$-2^{n-1} \dots 2^{n-1} - 1$

The same byte read as unsigned and as signed: only the agreed interpretation tells them apart  *8-bit two's complement: the sign bit splits the range into negative (−128 to −1) and positive (0 to 127)*

Worked example. What denary value does the 8-bit two's-complement number 10110100 represent?

The MSB is 1, so it is negative. Invert → 01001011, add 1 → 01001100 = 76, so the value is −76. Check with place values: $-128 + 32 + 16 + 4 = -76$.

Overflow in signed arithmetic happens when the true result falls outside this range — spotted when the sign bit flips wrongly (two positives giving a negative, or two negatives giving a positive).

One's complement

Before two's complement, an older scheme called **one's complement** 反码 represented a negative number by simply **inverting every bit** of the positive — there is no "add 1" step.

- +30 = 00011110, so in one's complement −30 = 11100001 (just the inverse).
- Drawback: it has **two zeros** — 00000000 (+0) and 11111111 (−0) — which wastes a bit pattern and makes arithmetic awkward.

Two's complement (invert **and** add 1) removes the negative zero: it has a single zero and lets addition and subtraction use the same circuit. That is why modern computers store signed integers in two's complement, not one's complement.

Binary Coded Decimal (BCD)

In **BCD** 二进码十进数, each denary digit is written as its own 4-bit pattern. The number 93 is `1001 0011` in BCD — not binary 93 (01011101). Each nibble uses only 0–9; patterns 1010–1111 are invalid.

BCD reading: `0010 0111 0101` → 2, 7, 5 → 275.

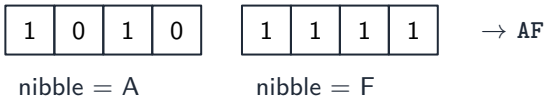
Use: calculators, digital clocks, and devices that show denary digits — each digit drives a **7-segment display** 七段显示器. Currency code often uses BCD to avoid the rounding errors of converting fractions like 0.1 to binary.



A seven-segment display shows one denary digit, often driven by BCD

Hexadecimal — practical uses

Hex is a compact way to write binary (1 hex digit = 4 bits):



A byte is two nibbles; each nibble is one hex digit

- **memory addresses** 内存地址 in low-level programming — `0x7FFE`.
- **colour values** in HTML/CSS — `#FF8800`.
- **MAC addresses** — `AC:DE:48:00:11:22`.

Hex does not change the stored data — it just makes binary easier for humans.

Character codes

Computers store text as numbers; each character has a numeric **code point** 码点 set by a **character set** 字符集.

ASCII

- **ASCII** uses 7 bits — 128 code points. Basic Latin letters, digits, punctuation, and control codes.
- **Extended ASCII** uses 8 bits — 256 code points; the lower 128 match ASCII, the upper 128 vary by region.

character	code (denary)	binary
A	65	01000001
a	97	01100001
0	48	00110000
(space)	32	00100000

Each character is stored as a number — a few ASCII code points in denary and binary

Unicode

- **Unicode** is a universal character set covering almost every script, plus symbols and emoji.
- common **encodings** 编码: **UTF-8** (1–4 bytes, ASCII-compatible), **UTF-16** (2 or 4 bytes), **UTF-32** (fixed 4 bytes).

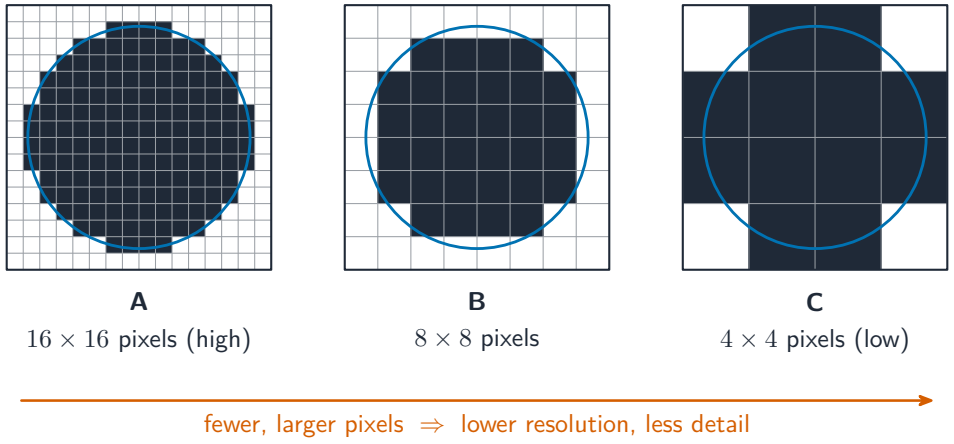
Why Unicode beats ASCII

- it represents **far more characters** (every script, emoji); ASCII covers only basic English.
- files are **portable** with no code-page confusion, and allow **multilingual** text in one document.
- trade-off: Unicode files are usually **larger** for English-only text.

Bitmap images

A **bitmap** 位图 image (also called a **bitmapped image**) stores the colour of every **pixel** 像素 in a grid. At the start of the file a **file header** 文件头 records the image's metadata — its width, height and colour depth — so software knows how to read the pixel data that follows.

- **image resolution** 图像分辨率: the bitmap's own size, width × height in pixels (e.g. 1920 × 1080).
- **screen resolution** 屏幕分辨率: the width × height the **display** can show. If an image's resolution is larger than the screen it is scaled down to fit; a low-resolution image looks blocky when stretched onto a higher-resolution screen.
- **colour depth** 颜色深度 (**bit depth** 位深度): bits per pixel. 1 bit → black/white; 8 bits → 256 colours; 24 bits → 16.7 million ("true colour").



The same image stored at three resolutions, from high (A) to low (C): fewer, larger pixels mean less detail

File size

$$\text{size in bits} = \text{width} \times \text{height} \times \text{bit depth}.$$

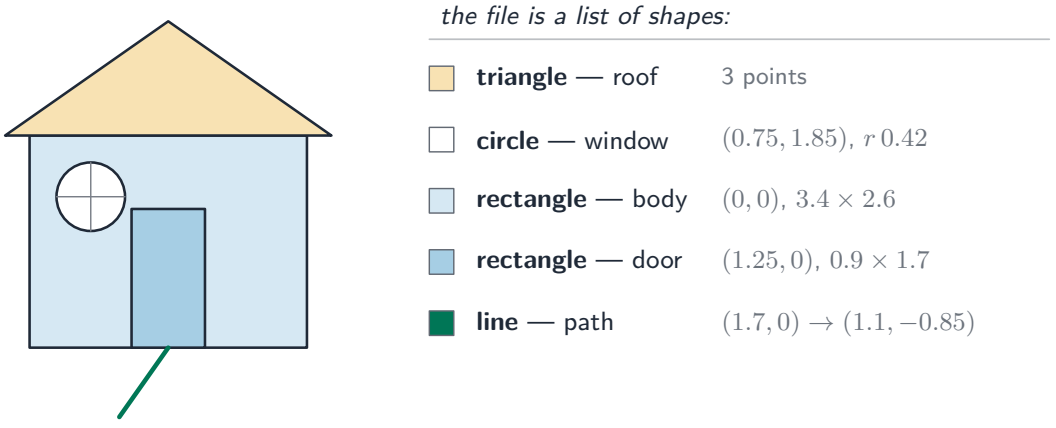
Divide by 8 for bytes, by 1024 for KiB, etc. Example: a 3000 × 2000 image at 24 bpp is $3000 \times 2000 \times 24 = 1.44 \times 10^8$ bits ≈ 17.2 MiB.

Changing settings

- **lower resolution** → smaller file, less detail (looks blocky when enlarged).
- **lower colour depth** → smaller file, but smooth shades show banding.
- **higher** of either → larger file, better quality.

Vector graphics

A **vector graphic** 矢量图形 stores the **instructions** to draw the image as a **drawing list** 绘图列表 — an ordered list of **drawing objects** 绘图对象 (geometric **primitives** 图元: lines, curves, polygons, circles). Each drawing object has **properties** 属性 such as colour, fill, line width and position (coordinates). To show it, the program **renders** 渲染 the drawing list at any resolution needed.

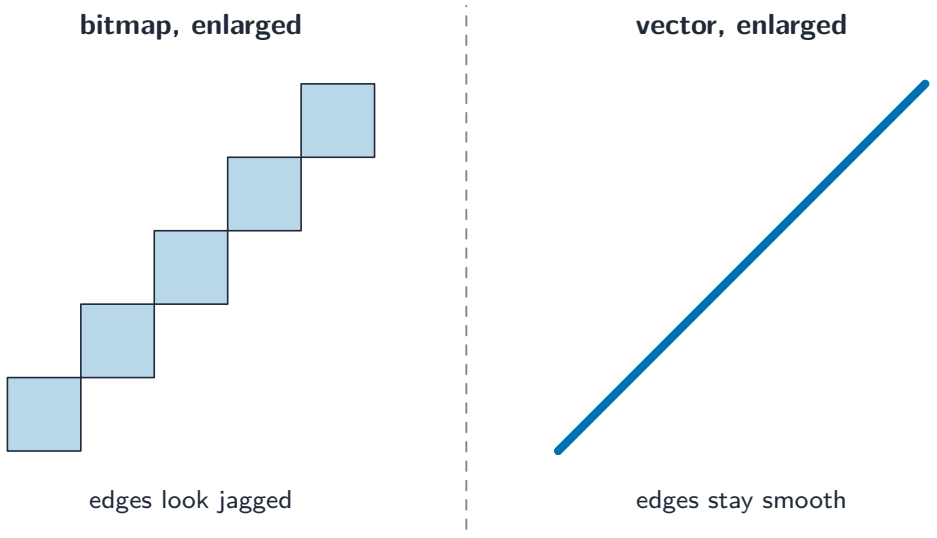


A vector image is built from labelled geometric shapes, each with attributes

Bitmap vs vector

Task	Better choice	Why
Photograph	Bitmap	Complex pixel-level detail can't be described as shapes.
Logo, icon, sign	Vector	Sharp edges; scales to any size without blur.
Engineering drawing	Vector	Precise geometry and scaling.
Painting, texture	Bitmap	Smooth tonal detail per area.

Vector advantage: it **scales without losing quality** — a vector logo stays sharp at any size, while a bitmap blurs when enlarged. **Vector disadvantage**: it cannot describe arbitrary pixel detail (photographs).

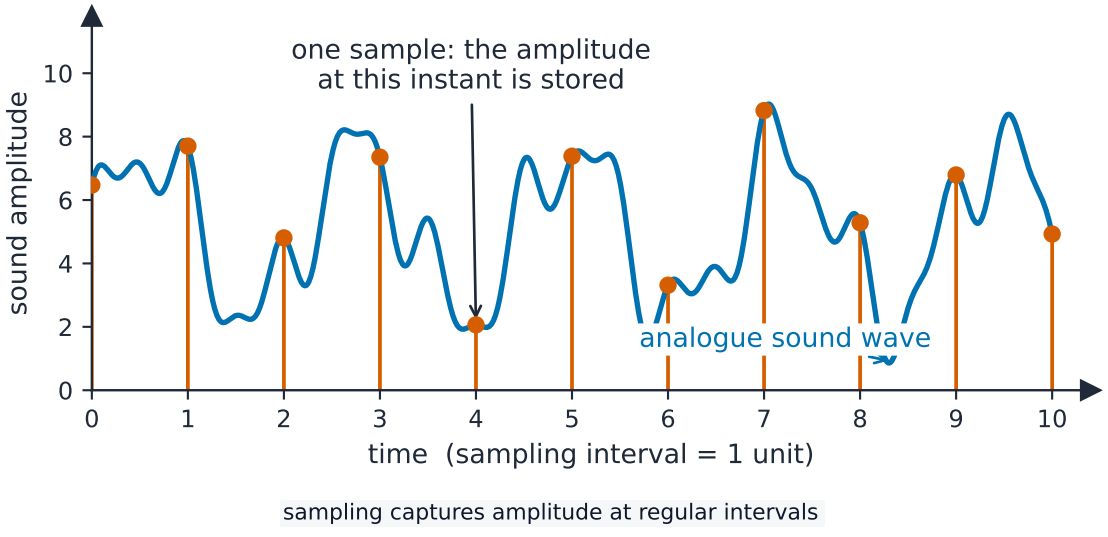


Enlarged, a bitmap's pixels turn jagged; a vector stays smooth at any size

Sound

A continuous wave of **analogue data** 模拟数据 (the sound) is converted into **digital data** 数字数据 by **sampling** 采样:

- **sampling rate** 采样率 — samples per second (Hz). CD quality is 44.1 kHz.
- **sampling resolution** 采样分辨率 (bit depth) — bits per sample's **amplitude** 振幅. CD quality is 16 bits.



Sampling a sound wave: its amplitude is read at each time interval

File size

$$\text{size in bits} = \text{sampling rate} \times \text{resolution} \times \text{duration} \times \text{channels}.$$

A 10-second stereo CD clip: $44100 \times 16 \times 10 \times 2 = 14\,112\,000$ bits ≈ 1.68 MiB.

Changing settings

- **higher sampling rate** → captures higher pitches, larger file.
- **higher sample resolution** → finer amplitude steps, less **quantisation** 量化 noise, larger file.
- **lower** of either → smaller file, clear quality loss.

(The sampling rate must be at least twice the highest frequency you want to keep.)

Compression

Compression 压缩 reduces file size, saving storage and transmission **bandwidth** 带宽. Two kinds:

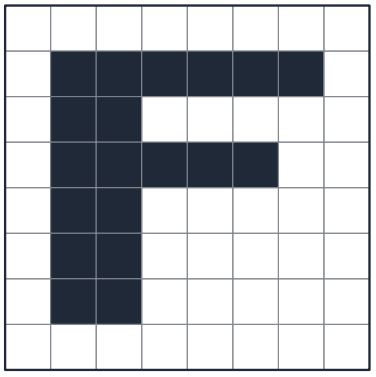
- **lossless** 无损 — the original data is recovered exactly (text, programs, ZIP/PNG).
- **lossy** 有损 — some detail is dropped for much smaller files (JPEG, MP3, video).

When to use which

- **lossless** for documents, source code, medical images — anything needing exact data.
- **lossy** for streaming media. **Real-time video streaming** uses lossy compression because it must send huge amounts of data in real time over limited bandwidth; lossless would not shrink it enough. Raw HD video is gigabytes per minute, so without compression the picture would keep freezing.

Lossless methods

- **run-length encoding** 行程编码 (RLE): store "the next n values are x" instead of repeating x. Great for flat areas; useless for noisy data.
- **dictionary methods** 字典编码 (ZIP, PNG): replace repeated byte sequences with a short reference. Good for text and code.
- **Huffman coding** 霍夫曼编码: give short codes to common symbols and long codes to rare ones, bringing the average code length near the data's **entropy** 熵.

8×8 bitmap	binary	RLE code
	00000000	8W
	01111110	1W 6B 1W
	01100000	1W 2B 5W
	01111100	1W 5B 2W
	01100000	1W 2B 5W
	01100000	1W 2B 5W
	01100000	1W 2B 5W
	00000000	8W

Key: 1 = black (B), 0 = white (W). RLE stores *count* + *colour*: e.g. 01111110 → 1W 6B 1W

Run-length encoding of the letter F in an 8 × 8 black-and-white grid

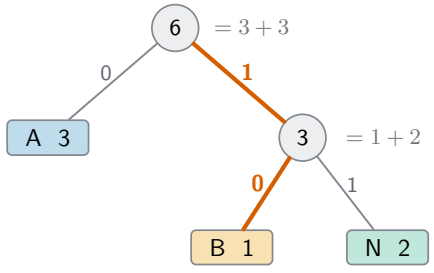
source ABC ABC ABC XYZ 12 characters

dictionary 1 → ABC 2 → XYZ stored once each

encoded 1 1 1 2 4 tokens + dictionary

The decoder rebuilds the dictionary as it reads, so the original returns exactly — **lossless** (ZIP, PNG).

Dictionary coding: each repeated sequence is stored once, and every occurrence becomes a short index



Walk **down from the root**, writing a bit at each branch: left = 0, right = 1.

A = 0

B = 10 right, then left

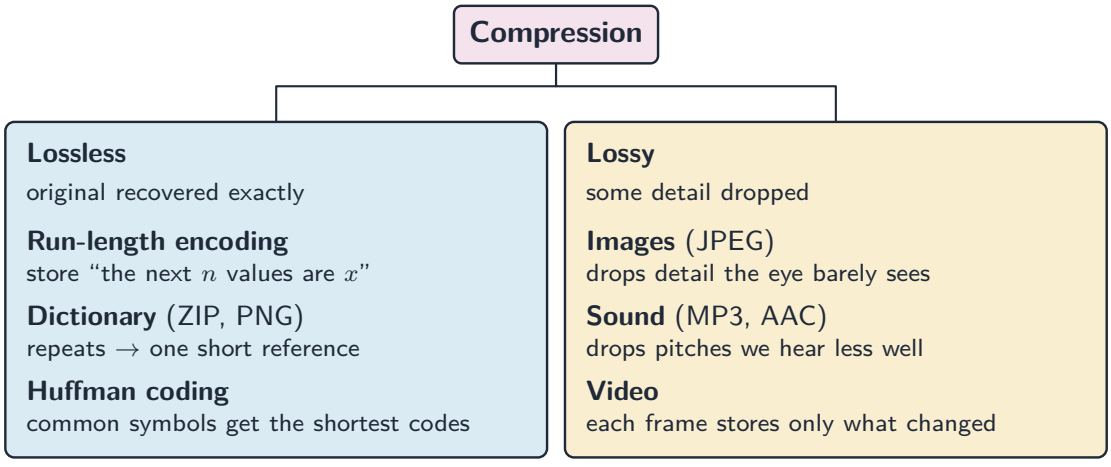
N = 11

BANANA becomes 10 0 11 0 11 0 — **10 bits**, against 12 for fixed 2-bit codes: the commonest letter is nearest the root.

Huffman coding: the commonest symbol gets the shortest code, so BANANA needs 10 bits instead of 12

Lossy methods

- **images** (JPEG): drop fine detail and colour differences the eye barely sees.
- **sound** (MP3, AAC): drop pitches we hear less well, and quiet sounds hidden by louder ones.
- **video** combines **spatial** 空间 compression (within each frame, like JPEG) with **temporal** 时间 compression (most frames store only the differences from the previous frame).



Compression methods: lossless versus lossy, with common examples

Exam tips

- Show working for base conversions: denary → binary by place values, binary → hexadecimal in **nibbles** (groups of 4 bits).
- For **two's complement** the MSB is negative; to negate, **invert and add 1**; watch for **overflow** when the sign bit flips wrongly.
- Distinguish **bitmap** (pixels; file size = width × height × colour depth) from **vector** (drawing commands; scales without loss).
- Sound file size depends on **sample rate** × **bit depth** × **time** — more of each means better quality but a bigger file.
- Compare **lossless vs lossy** compression and give a use for each.

Communication

A-Level Computer Science

Networks: purpose and benefits

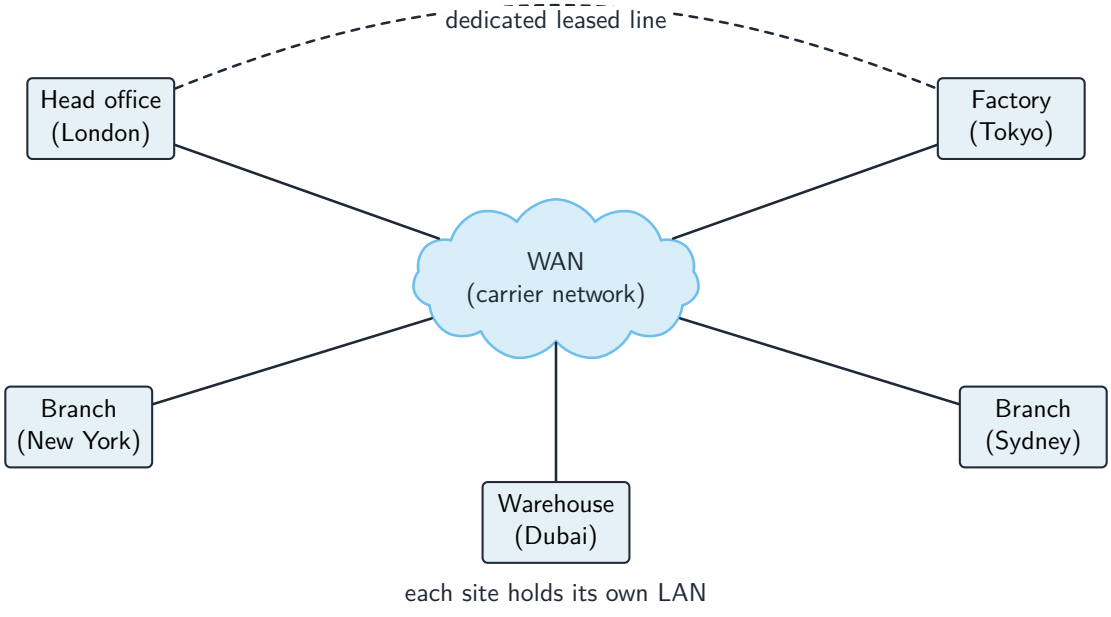
A **network** 网络 is a set of computing devices connected so they can communicate and share resources. Benefits:

- **sharing resources** (printers, file servers, internet) — cheaper than equipping each computer.
- **sharing data** — many users access the same files.
- **central management** — install software, manage users and back up once on a server.
- **communication** — email, video calls, messaging.
- **remote access** — work from anywhere.

LAN vs WAN

A **local area network** 局域网 (LAN) covers a **small area** — a home, office or school, usually owned by the organisation, with high data rates and low **latency** 延迟.

A **wide area network** 广域网 (WAN) covers a **large area** — a city, country, or the world (the internet is the largest WAN). It uses telecom-company infrastructure — often the **Public Switched Telephone Network** 公共交换电话网 (PSTN), leased lines or fibre — with lower data rates and higher latency. A WAN connects LANs together.

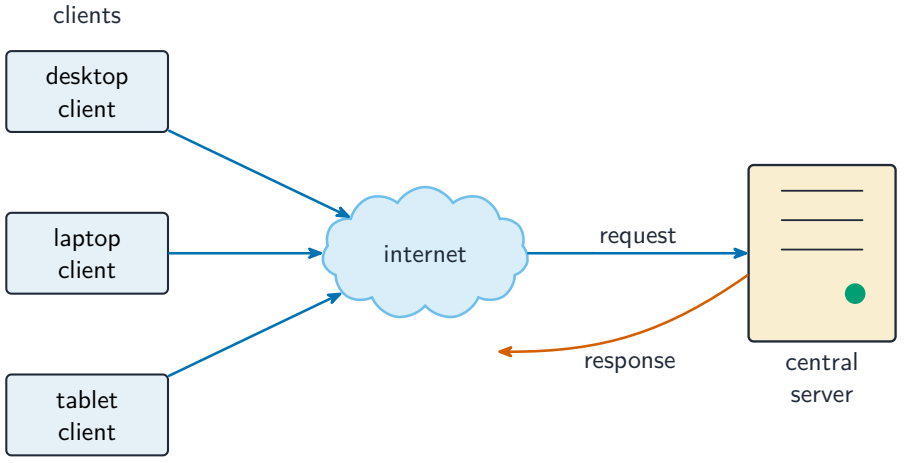


A wide-area network links many systems across a large area

Client-server and peer-to-peer

Client-server

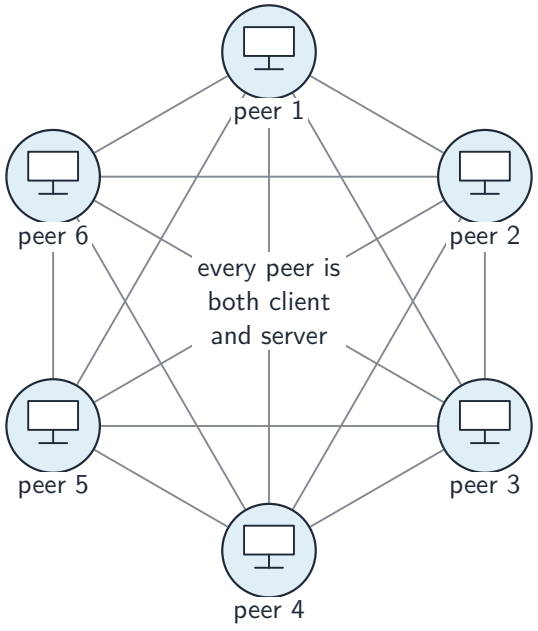
- powerful machines act as **servers** 服务器, providing services (files, web pages, email).
- other machines are **clients** 客户端 that request services.
- central and easy to manage, but the server is a single point of failure unless backed up.



In a client-server network, clients request services from a central server

Peer-to-peer (P2P)

- all machines are equal **peers**; each can be both client and server (**peer-to-peer** 对等网络).
- resources are spread across the peers — no central server. Robust to one failure, but harder to keep secure and consistent.



In a peer-to-peer network, every node is both client and server

Thin and thick clients

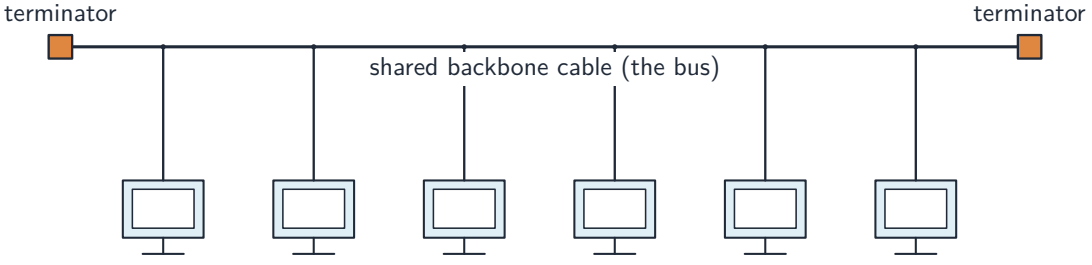
A **thin client** 瘦客户端 does **little processing locally** and relies on a powerful server (web terminals, remote desktops). A **thick client** 胖客户端 has **strong local processing and storage** and runs full applications itself (a normal desktop PC).

Feature	Thin client	Thick client
Local processing	minimal	substantial
Local storage	minimal	substantial
Reliance on network	high	lower
Server load	high	lower

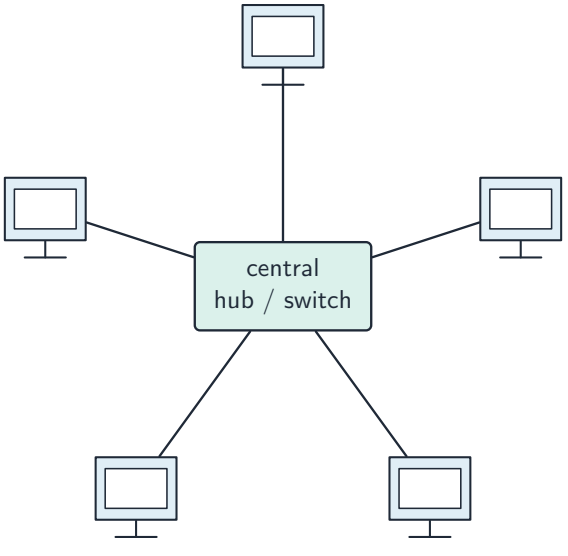
Network topologies

The **topology** 拓扑 is how the nodes and links are arranged.

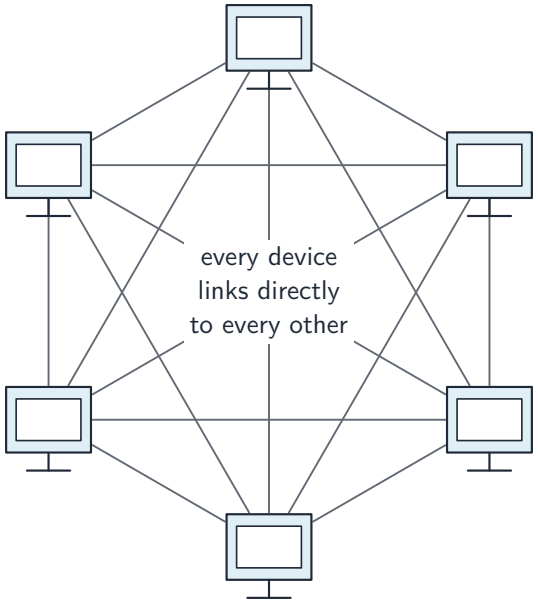
- **bus** 总线 — all devices on one shared cable. Cheap; the whole LAN fails if the bus fails; performance drops as more devices share the **bandwidth** 带宽.
- **star** 星形 — every device connects to a central switch. One device failing does not affect others; the switch failing brings all down. Most common today.
- **mesh** 网状 — every device links directly to others, with many paths. Very **fault-tolerant** 容错 (traffic reroutes) but needs lots of cabling.
- **hybrid** — a mix (a star in each office, mesh links between offices).



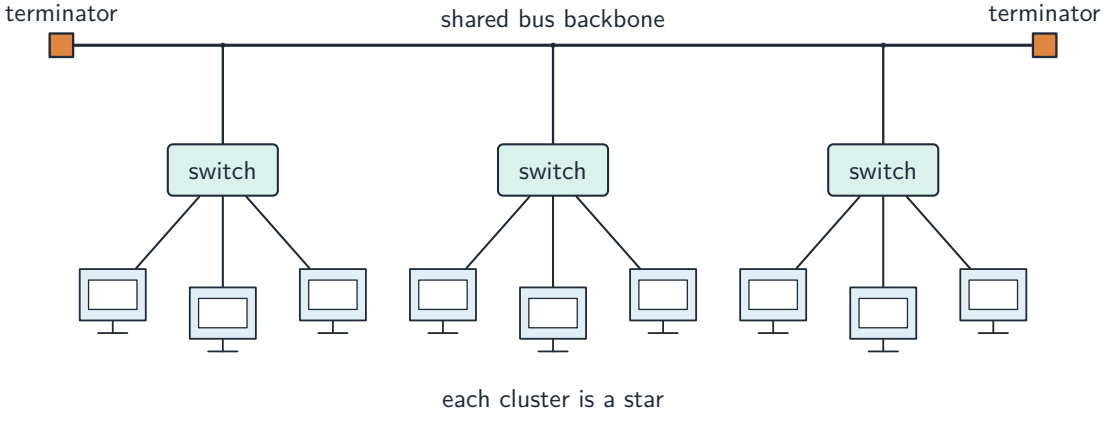
Bus topology: all devices share one cable with a terminator at each end



Star topology: every device connects to a central hub or switch



Mesh topology: every device links directly to the others



Hybrid topology: star clusters joined by a central bus

Cloud computing

Cloud computing 云计算 delivers computing services (servers, storage, software) **over the internet**, hosted by a third party. Benefits: **scalability** 可扩展性 (pay for what you need), lower cost, access from anywhere, and reliable redundant data centres. Drawbacks: needs internet, your data is held by a third party, and possible vendor lock-in.

Wired vs wireless

- **wired** (**Ethernet** 以太网 over **twisted-pair** 双绞线 or **fibre-optic** 光纤): higher speed, lower latency, fewer errors, more secure.
- **wireless** (Wi-Fi, Bluetooth, cellular): no cables, devices can move, but slower, prone to interference and eavesdropping.

For the same generation, wired wins on speed and reliability; wireless wins on convenience.

LAN hardware

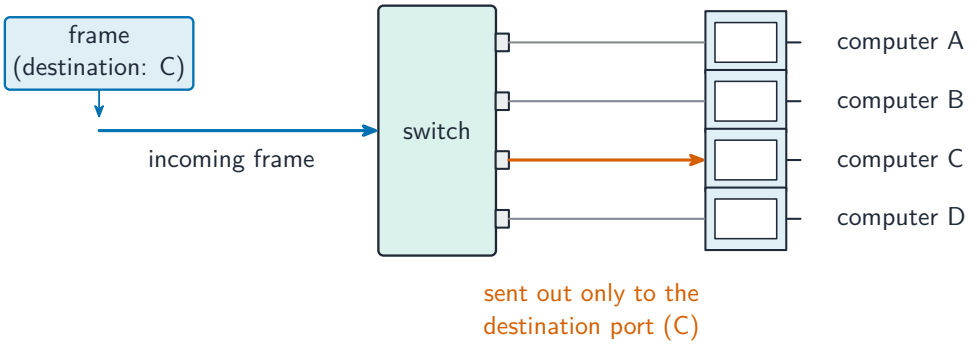
- **network interface card** 网络接口卡 (NIC) — lets a device send and receive on the network; has a unique **MAC address** MAC 地址 (a 48-bit hardware address). A wireless device uses a **wireless network interface card** 无线网络接口卡 (WNIC).
- **switch** 交换机 — forwards Ethernet frames only to the port for the destination MAC address.
- **hub** 集线器 — a simpler device that copies traffic to all ports (now obsolete).
- **wireless access point** 无线接入点 (WAP) — lets wireless clients join a wired LAN.
- cabling — twisted-pair for short runs; fibre-optic for longer, faster runs.



A network switch: each device's cable plugs into one of its ports



An RJ-45 plug on a twisted-pair Ethernet cable

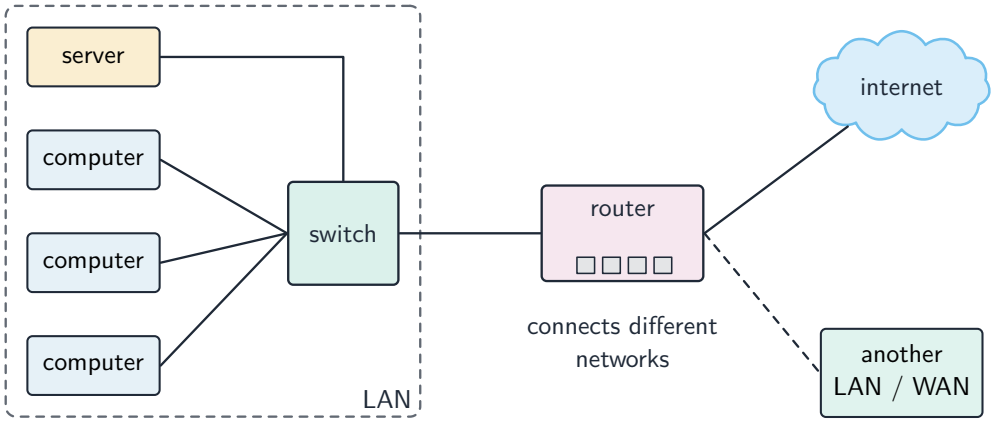


A switch sends each frame only to the port for its destination

Routers

A **router** 路由器 connects **different networks** and forwards data between them — usually at the boundary of a LAN and the internet. It does:

- **forwarding** — reads each **packet** 数据包's destination **IP address** IP 地址 and sends it out the right port, using a **routing table** 路由表.
- **network address translation** 网络地址转换 (NAT) — lets many private LAN addresses share one public IP.
- **DHCP** 动态主机配置协议 — hands out private IP addresses to LAN devices.
- **firewall** 防火墙 — blocks unwanted incoming traffic.



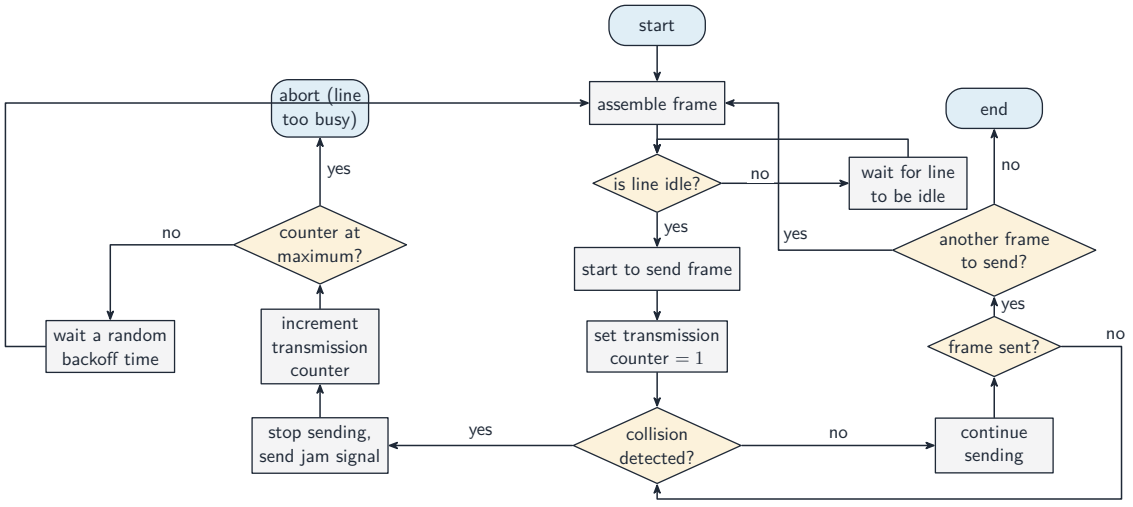
A router connects a LAN to the internet or another network

Ethernet and CSMA/CD

Ethernet is the main wired LAN technology. On shared media a **collision** 冲突 can happen when two devices send at once. The protocol is **CSMA/CD** 载波侦听多路访问/冲突检测 (Carrier Sense Multiple Access with Collision Detection):

1. **carrier sense** — listen before sending; wait if the cable is busy.
2. **multiple access** — many devices share the medium.
3. **collision detection** — keep listening while sending; a clash is a collision.
4. on a collision, both stop, send a brief "jam" signal, then wait a **random backoff** time before retrying.

Modern switched Ethernet uses **full-duplex** 全双工 point-to-point links, so collisions no longer happen.



The CSMA/CD process for handling collisions on shared media

Bit streaming

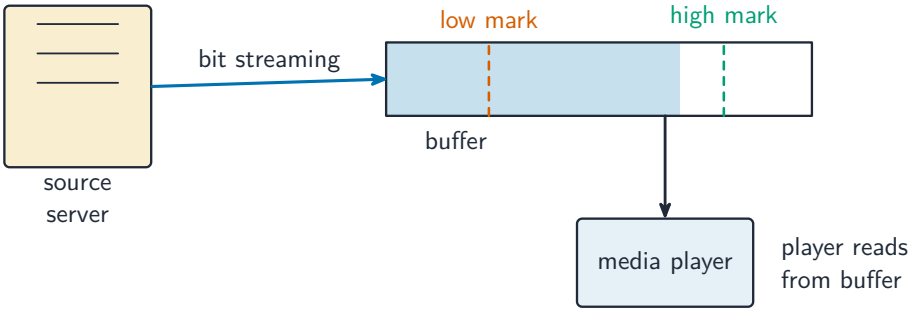
Bit streaming 流式传输 sends multimedia as a continuous stream that the receiver plays as it arrives, instead of downloading the whole file first.

- **real-time** (live): captured and streamed as it happens (live sport, video calls). You cannot rewind; low latency is vital.
- **on-demand**: pre-recorded on a server (YouTube, Netflix). You can pause and rewind; the server can **buffer** 缓冲 ahead.

Real-time streaming works as a short pipeline:

1. **capture and sample** the source (a camera or microphone).
2. **encode** it, using **compression** 压缩 to shrink the data.
3. **send** it across the network as packets.
4. the receiver buffers a little, then plays it live — **dropping any packet that arrives late**, because a live stream cannot wait for it.

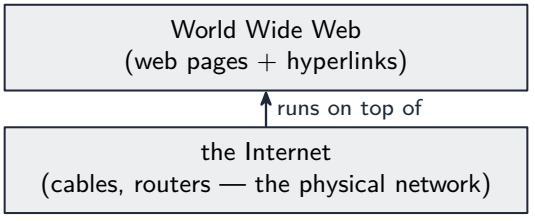
Lossy 有损 compression is used here: moving pictures hide small losses, and the stream must be small enough to fit the bandwidth.



Data streams from the server into a buffer before the media player reads it

The internet and the World Wide Web

The **internet** 互联网 is a global **network of networks** using a common **protocol** 协议 suite (TCP/IP). The **World Wide Web** 万维网 (WWW) is a **service** that runs over it: hyperlinked documents identified by URLs, viewed in browsers via HTTP/HTTPS. Email and file transfer are other internet services that are not part of the WWW.



The Web is one service running on top of the Internet

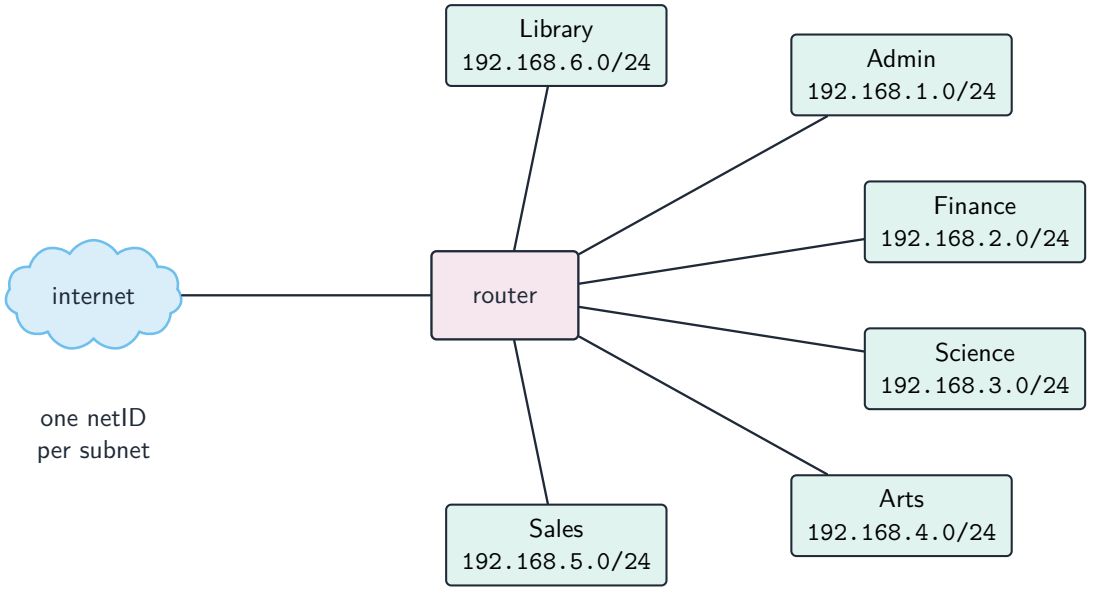
IP addresses

An IP address uniquely identifies a device.

- **IPv4** — 32-bit, four denary numbers 0–255 (`192.168.1.10`); about 4.3×10^9 addresses (now exhausted).
- **IPv6** — 128-bit, eight groups of four hex digits; about 3.4×10^{38} addresses.

Subnetting

A network can be split into **subnets** 子网. The IP address splits into a network part and a host part, given by a **subnet mask** 子网掩码 (e.g. `255.255.255.0` = first 24 bits are network). Subnetting improves management, cuts broadcast traffic, and improves security.



Splitting a network into subnets, one netID per department

Public vs private addresses

- **private** addresses are used within a LAN and are not routable on the internet (e.g. `192.168.0.0/16`).
- a **public IP address** is globally unique and routable, assigned by an **ISP** 互联网服务提供商.

Devices behind NAT with private addresses are not directly reachable from the internet, giving some protection.

Static vs dynamic

- a **static IP address** is fixed; used for servers that must be found at a known address.
- a **dynamic IP address** is assigned by DHCP and may change; easier for client devices and uses a limited address pool efficiently.

Worked example. A host has IP address `192.168.10.130` with subnet mask `255.255.255.192`. Which network is it on, and is `192.168.10.200` on the same one? The mask's last octet, `192`, is `11000000` in binary, so the first **26** bits are the network part and the last 6 bits address the host. That makes the subnets step in blocks of $256 - 192 = 64$: `.0`, `.64`, `.128`, `.192`. The address `130` falls in the block starting at `.128`, so the host is on network **`192.168.10.128/26`**, whose usable hosts run `.129` to `.190` (`.191` is the broadcast address). `200` falls in the **next** block (`.192`), so it is on a **different** subnet and traffic between the two must pass through a **router**. Get the block size from the mask first (256 minus the mask octet) - guessing from the first three octets is what makes these go wrong.

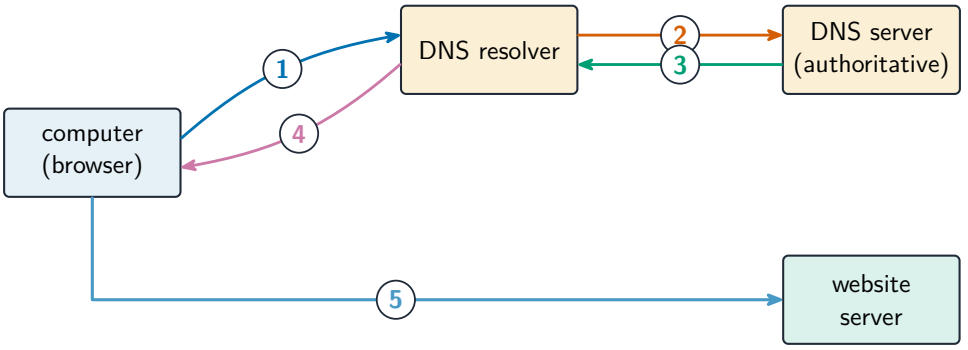
URL and DNS

A **URL** 统一资源定位符 (Uniform Resource Locator) locates a resource on the WWW:

```
https://www.example.com/about/contact.html
protocol     domain name        path
```

- **protocol**: `http`, `https`, etc.
- **domain name** 域名: a readable server address.
- **path**: the resource on that server.

The **Domain Name System** 域名系统 (DNS, also called the Domain Name Service) is a distributed set of servers that turns **domain names into IP addresses**. When you type a URL, the browser asks a DNS resolver for the IP, which queries DNS servers (root → top-level → authoritative) until it finds it; the browser then connects to that IP and requests the path. DNS saves humans from memorising IP addresses and lets a site change server without changing its name.



1 computer asks the resolver for the IP **2** resolver queries a DNS server
3 server returns the IP address **4** resolver passes the IP back to the computer
5 browser connects to that IP and requests the page

How DNS finds a website's IP address before the browser connects

Exam tips

- Distinguish **LAN vs WAN** and **client-server vs peer-to-peer** by who stores and controls the resources.
- Match each **topology** (bus, star, mesh) to its advantages and drawbacks (cost, reliability, collisions).
- Know the job of each device: a **switch** directs within a LAN by MAC address, a **router** routes between networks by IP.
- Explain **bit streaming** and why buffering is needed (data arrives at a different rate from playback).
- Distinguish **IPv4 vs IPv6** and public vs private addresses; DNS turns a URL into an IP address.

Hardware

A-Level Computer Science

Computers and their components

A general-purpose computer has four building blocks:

- **input devices** 输入设备 — get data in (keyboard, mouse, microphone, scanner, sensors).
- **output devices** 输出设备 — give results out (monitor, speakers, printer, actuators).
- **primary memory** 主存储器 — fast memory the **processor** 处理器 (CPU) reaches directly (RAM and ROM). Holds the running program and its data.
- **secondary storage** 辅助存储器 — slower, larger, keeps programs and data when not in use (hard disk, SSD, optical disc, USB stick).



A keyboard: a common input device for typing text and commands



A mouse: a pointing input device



A flatbed scanner: an input device that turns a paper page into a digital image



A monitor: a common output device that displays the screen image

Embedded systems

An **embedded system** 嵌入式系统 is a computer built **into another device** to do one fixed job (washing machine, microwave, car engine unit, thermostat).

- **benefits**: optimised for one task (low power, small, cheap); reliable; fast to start; cheap in volume.
- **drawbacks**: limited to its one task; hard to update (its **firmware** 固件 may need special tools); often not repairable; sometimes weak security.

Principal hardware devices

Laser printer

A **laser printer** 激光打印机 scans the page image onto a charged photosensitive **drum** 感光鼓. **Toner** 墨粉 sticks to the charged areas, transfers to the paper, and is melted on by a fuser. Fast, sharp, high-volume.



A laser printer: fast, sharp printing using a charged drum and toner

3D printer

A **3D printer** 3D 打印机 builds an object **layer by layer**: FDM melts plastic filament through a nozzle; stereolithography cures liquid resin with a UV laser. Used for prototypes and custom medical parts.



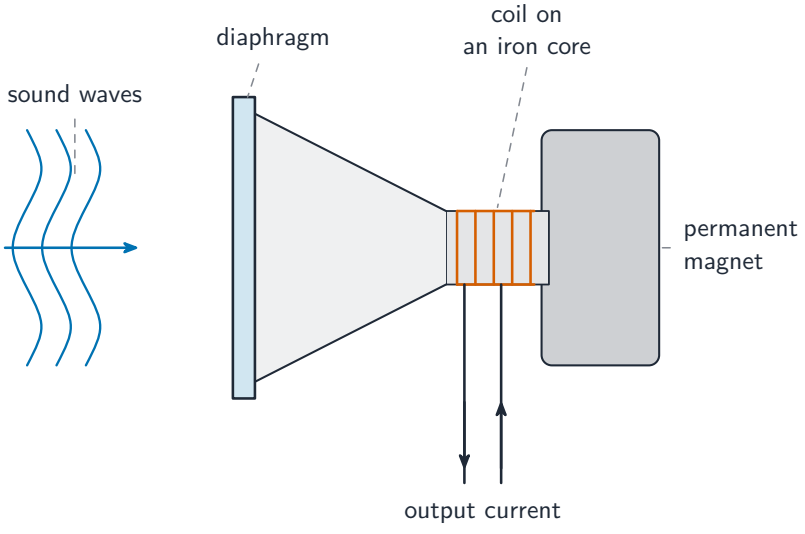
An FDM 3D printer builds an object layer by layer by melting plastic filament

Microphone and speakers

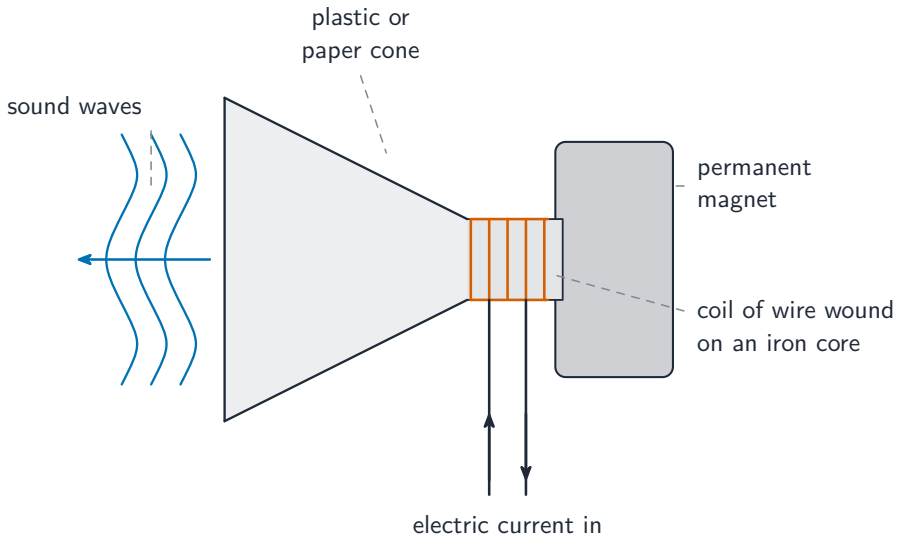
A **microphone** 麦克风 turns **sound into an electrical signal** (a diaphragm vibrates, changing **capacitor** 电容器 charge or coil position); the signal is digitised by an **analogue-to-digital converter** 模数转换器 (ADC). A speaker does the reverse — a varying signal drives a coil in a magnetic field, moving a cone to make sound.



A microphone turns sound into an electrical signal



Inside a microphone: sound vibrates the diaphragm and coil to produce a current



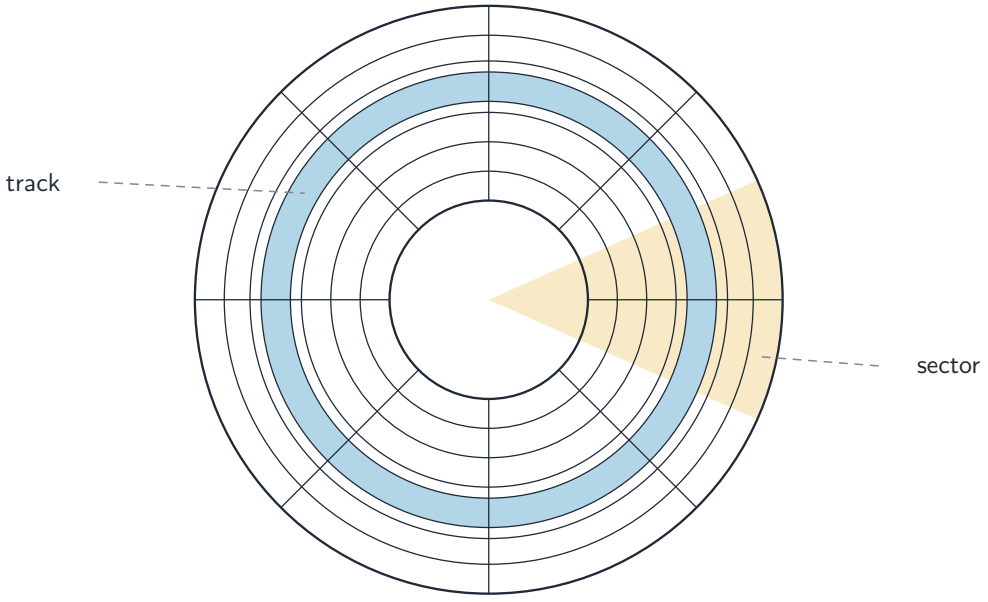
Inside a loudspeaker: a varying current in the coil moves the cone to make sound

Magnetic hard disk (HDD)

A **hard disk** 硬盘 stores data on spinning platters coated with magnetic material. Each platter has **tracks** 磁道 divided into **sectors** 扇区. A **read/write head** 读写头 floats just above and magnetises tiny regions (write) or senses them (read). Cheap per gigabyte, but slower than SSDs and has moving parts.



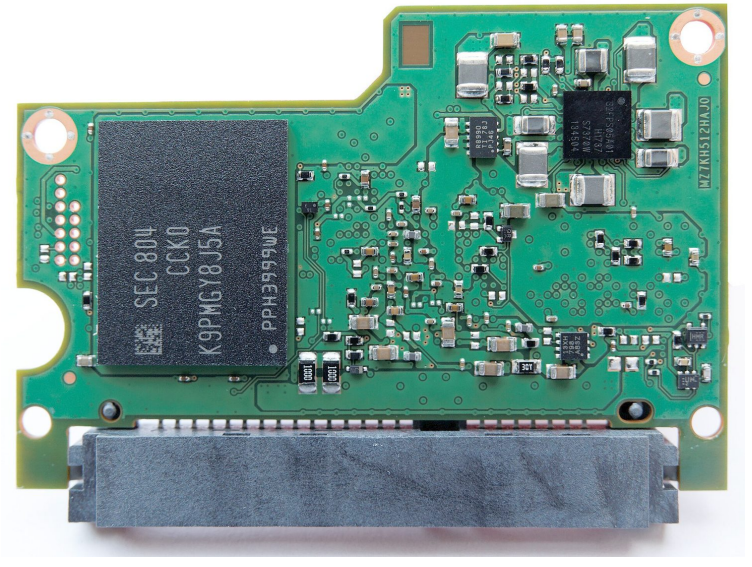
An opened hard disk: the actuator arm carries the read/write head over a platter



Tracks and sectors on a hard disk platter

Solid-state (flash) memory

A **solid-state drive** 固态硬盘 stores data as charge in **transistors** 晶体管, with no moving parts. Faster random access than HDDs, tougher, lower power, but dearer per gigabyte; each cell wears out after many writes.



Inside an SSD: data is stored in flash memory chips, with no moving parts (compare the hard disk above)

Optical disc

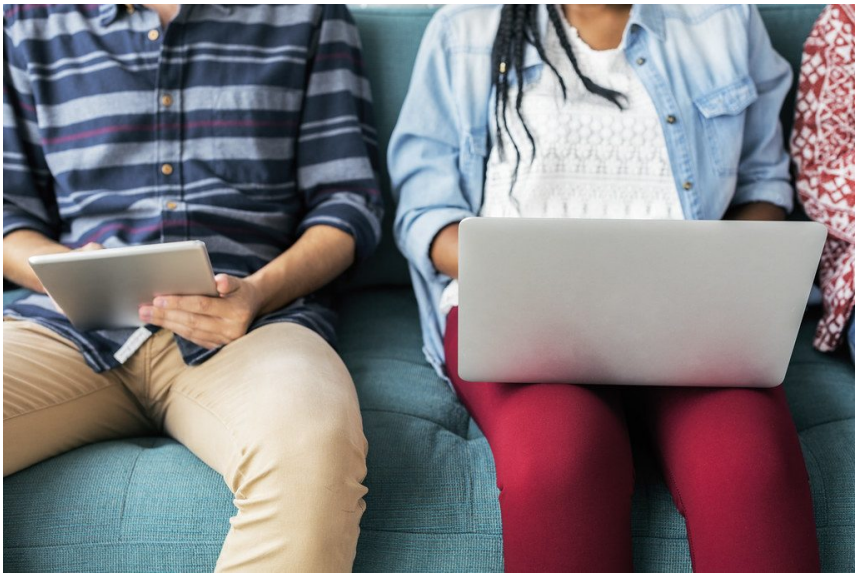
A laser detects reflections from tiny pits on an **optical disc** 光盘 (CD, DVD, Blu-ray). The drive is an **optical disc reader/writer**: writing uses a stronger laser to change the surface's reflectivity.



An optical disc drive: a laser reads tiny pits on a CD, DVD or Blu-ray disc

Touchscreen

A **touchscreen** 触摸屏 senses contact. **Resistive** 电阻式: two conductive layers pressed together; works with anything but is less accurate. **Capacitive** 电容式: a finger disturbs a charge field; accurate, multi-touch, used in phones.



A touchscreen senses where a finger touches the glass

Virtual reality headset

A **virtual reality** 虚拟现实 (VR) headset has two small displays (one per eye) and motion sensors (**accelerometer** 加速度计, **gyroscope** 陀螺仪) that track head movement so the scene shifts as you look around.



A virtual reality headset: two small displays and motion sensors track the head

Buffers

A **buffer** 缓冲 is memory that **holds data temporarily** while it moves between devices of different speeds. Example: the CPU writes a document to a printer buffer quickly, then is free to do other work while the printer prints from the buffer at its own pace. Buffers stop the fast device waiting for the slow one (also used in streaming, the keyboard, and disk access).

RAM and ROM

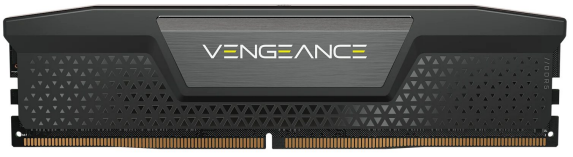
- **RAM** 随机存取存储器 (Random Access Memory) — **volatile** 易失性 (loses data without power). Holds the OS, running programs and their data; read and written constantly.
- **ROM** 只读存储器 (Read-Only Memory) — **non-volatile** 非易失性 (keeps data without power). Usually written once; holds firmware needed at start-up (the BIOS / boot loader).

RAM
volatile — lost at power off
read *and* write
holds running programs & data

ROM
non-volatile — kept at power off
read-only
holds start-up instructions

RAM is volatile and read/write; ROM is non-volatile and read-only

ROM starts the system; RAM then holds the active work.



A RAM module (DIMM) plugs into the motherboard as the computer's fast main memory

SRAM vs DRAM

- **SRAM** 静态 RAM (Static RAM) stores each bit in a **flip-flop** 触发器 of several transistors. Fast, but expensive and not dense. Used for CPU **cache** 高速缓存.

- **DRAM** 动态 RAM (Dynamic RAM) stores each bit as charge on a tiny capacitor. Cheaper and denser but slower, and must be **refreshed** 刷新 (rewritten) thousands of times a second. Used for main memory.

Use SRAM for small fast memory (cache); DRAM for large main memory.

PROM, EPROM and EEPROM

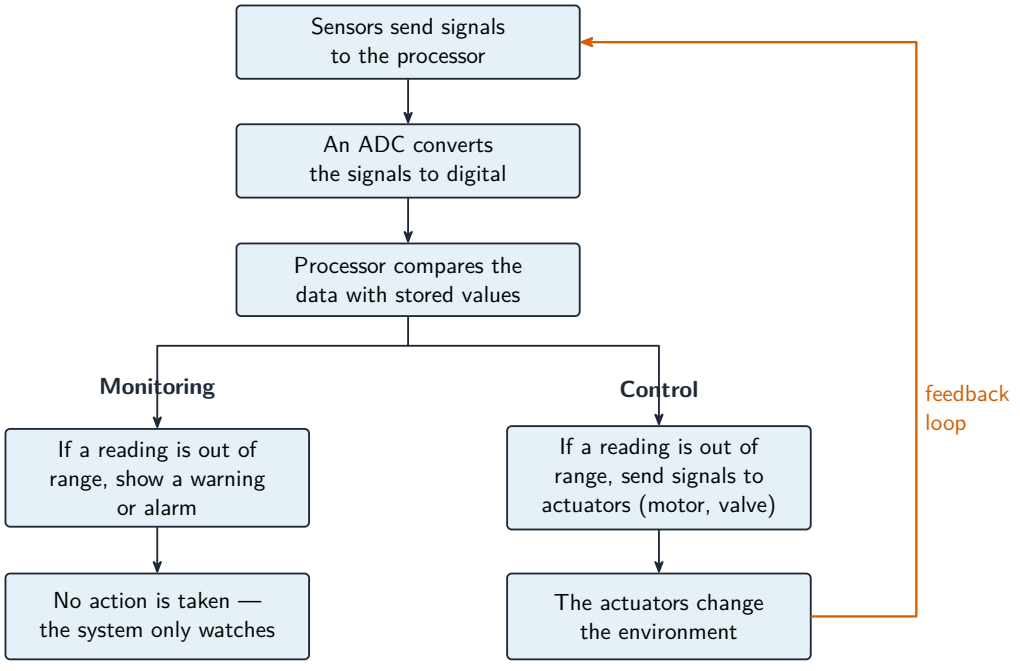
ROM variants you can program after manufacture:

- **PROM** (Programmable ROM) — written **once** (fuses burned by a programmer); cannot be changed.
- **EPROM** (Erasable Programmable ROM) — erased by **strong UV light** through a window, then rewritten (whole chip at once).
- **EEPROM** (Electrically Erasable Programmable ROM) — erased and rewritten **electrically**, a byte at a time, in circuit. Flash memory is a derivative optimised for block erase.

Monitoring and control systems

Both read **sensors**; the difference is what they do next.

- **monitoring** 监控 — collects and reports data but **takes no action** (a weather station logging readings).
- **control system** 控制系统 — uses sensor data to **decide and act** through actuators, usually in a feedback loop (a thermostat turning a boiler on/off).



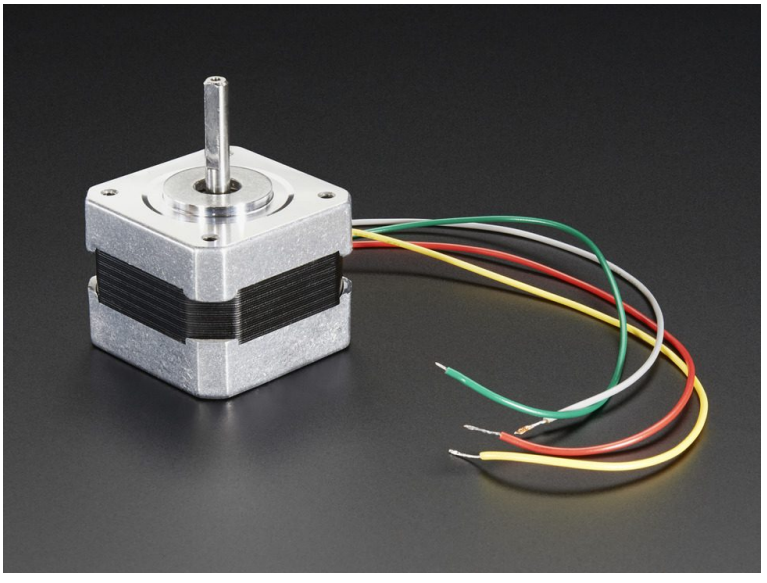
Monitoring reports data; a control system acts through a feedback loop

Sensors and actuators

A **sensor** 传感器 turns a physical quantity into a signal: temperature (a **thermistor** 热敏电阻 or thermocouple), pressure (strain gauge), infra-red, sound. Analogue signals need an ADC first. An **actuator** 执行器 does the reverse — turns a signal into an action (a motor, valve, heater, buzzer).



A thermistor: a temperature sensor whose resistance changes with heat



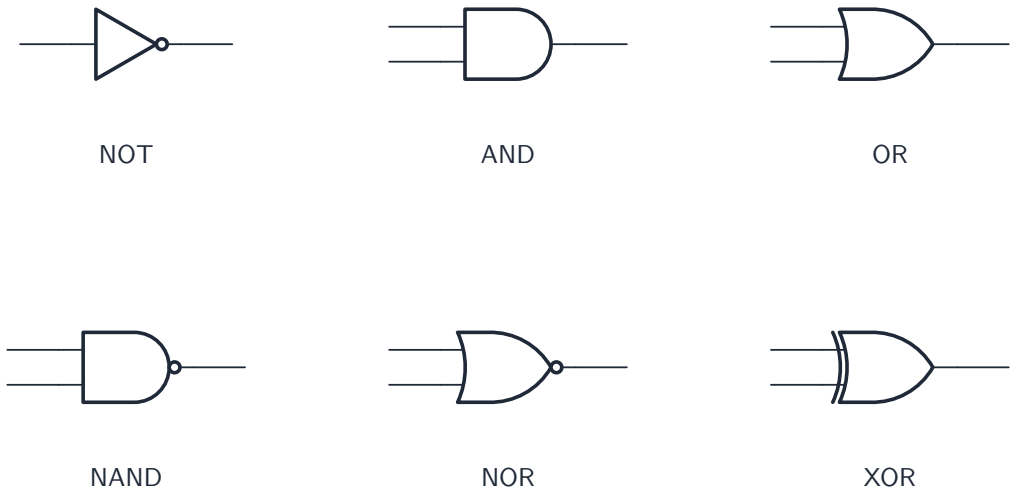
A small electric motor: an actuator that turns a signal into movement

Feedback

In a control system the actuator changes the environment, which the sensors then re-measure — a **feedback** 反馈 loop. Without feedback the system cannot correct itself or know when to stop (a thermostat with no temperature feedback would heat forever).

Logic gates

A **logic gate** 逻辑门 is a small circuit that does one **Boolean** 布尔 operation. Inputs and outputs are 0 (false, low) or 1 (true, high). Know the symbol, function and **truth table** 真值表 for each gate.



The symbols for the six logic gates

NOT (inverter)

A	NOT A
0	1
1	0

AND — output 1 only if all inputs are 1

A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

OR — output 1 if at least one input is 1

A	B	A OR B
0	0	0
0	1	1
1	0	1
1	1	1

NAND (NOT AND) — output 0 only when all inputs are 1

A	B	A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

NOR (NOT OR) — output 1 only when all inputs are 0

A	B	A NOR B
0	0	1
0	1	0
1	0	0
1	1	0

XOR (Exclusive OR, also called EOR) — output 1 if the inputs are different

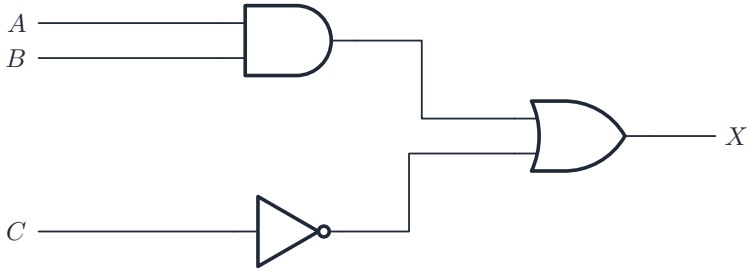
A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Logic circuits

A **logic circuit** 逻辑电路 is a network of gates that carries out a Boolean expression. You should be able to move between a problem statement, a logic expression, a truth table, and a circuit diagram.

From expression to circuit

Draw one gate per operator and wire them up. For $X = (A \text{ AND } B) \text{ OR } (\text{NOT } C)$: a NOT gate on C, an AND gate on A and B, then an OR gate on the two results.



Gates wired together to carry out a Boolean expression

From circuit to expression

Work forwards from the inputs, labelling each gate's output, until you reach the final output.

From circuit to truth table

For n inputs there are 2^n rows. List every input combination; for each, work out the internal gates then the output.

From truth table to expression (sum of products)

For each row that outputs 1, write an AND of the inputs (with NOT on any input that is 0 in that row); OR these together. Example: a table that is 1 only on $(A = 0, B = 1)$ and $(A = 1, B = 0)$ gives $\overline{A}B + A\overline{B}$, which is A XOR B.

From a problem statement

Turn the English into a Boolean expression first: "A and B" → A AND B; "A or B or both" → A OR B; "exactly one of A and B" → A XOR B; "neither A nor B" → A NOR B; "not both" → A NAND B.

Worked example. A machine's alarm X sounds when the guard is open ($A = 1$) **and** either the motor is running ($B = 1$) or the temperature is high ($C = 1$). Write the Boolean expression, and give the rows where $X = 1$. Turn the English into logic one clause at a time: "either B or C" is $B+C$, and "A **and** that" is $X = A \cdot (B+C)$. For the rows, $X = 1$ needs $A = 1$ **and** at least one of B, C equal to 1 - so $(A, B, C) = (1, 0, 1)$, $(1, 1, 0)$ and $(1, 1, 1)$, three rows out of eight. Notice $A = 0$ can never sound the alarm, whatever B and C do. **Bracket the OR** before ANDing it: $X = A \cdot B + C$ is a different circuit altogether, one that would sound the alarm on a high temperature even with the guard **closed**.

Exam tips

- Distinguish **RAM (volatile, read/write)** from **ROM (non-volatile, holds the bootstrap)**; SRAM (cache, faster) from DRAM (main memory, needs refreshing).
- For a **logic circuit**, build the Boolean expression gate by gate, then a truth table covering **every** input combination.
- Learn the symbol, expression and truth table for each gate (AND, OR, NOT, NAND, NOR, XOR).
- Explain a **buffer** (a temporary store bridging two different speeds) and the role of an interrupt.

Processor Fundamentals

A-Level Computer Science

Von Neumann architecture

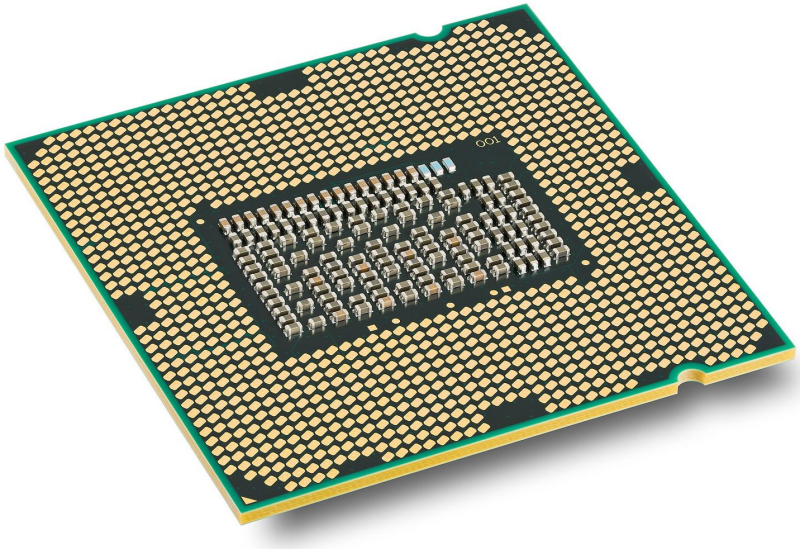
The **Von Neumann architecture** 冯 · 诺依曼体系结构 underlies almost every general-purpose computer:

- a **single memory** — the **Immediate Access Store** 立即存取存储器 (IAS) — holds both **program instructions** and **data** (the **stored program** 存储程序 concept).
- a **processor** 处理器 (CPU) fetches instructions from memory and runs them one at a time.
- instructions run **in order** unless a branch changes the flow.

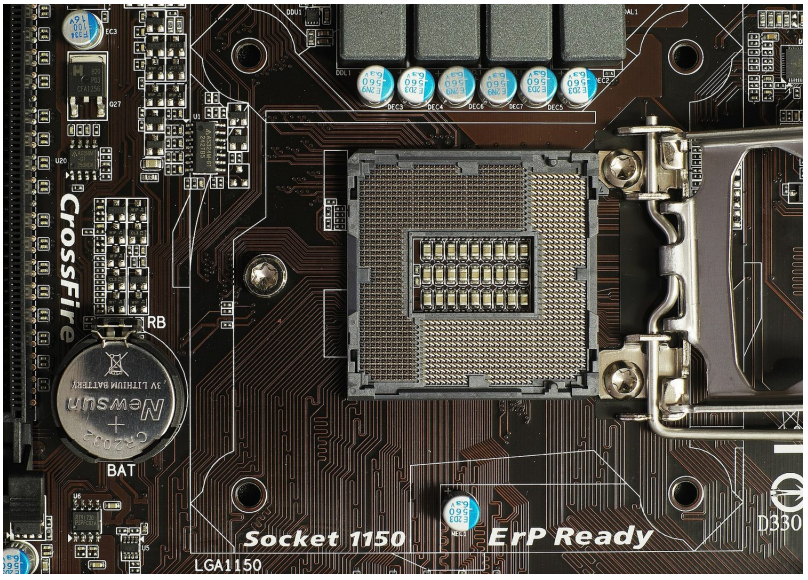
The stored-program idea is what makes a computer flexible: change the program and you change what it does, with no rewiring.

The CPU's main parts

All of these parts sit inside one small chip. The diagram later in this section shows how they connect; the photo below shows the real thing.



A modern CPU: the whole processor is one small chip (here seen from below, showing the contacts)



The matching CPU socket on the motherboard: the chip's contacts press onto these pins

Arithmetic and Logic Unit (ALU)

The **ALU** 算术逻辑单元 does the arithmetic (add, subtract, …) and logic (AND, OR, comparisons). It takes operands from **registers** 寄存器 and puts results back in a register.

Control Unit (CU)

The **control unit** 控制单元 **decodes** each instruction and sends the **control signals** to carry it out — opening data paths, telling the ALU what to do, and controlling memory reads and writes.

System clock

The clock sends a steady stream of pulses that keep the CPU in step. Each instruction takes a fixed number of cycles, and the **clock speed** 时钟频率 (e.g. 3.8 GHz) is one factor in performance.

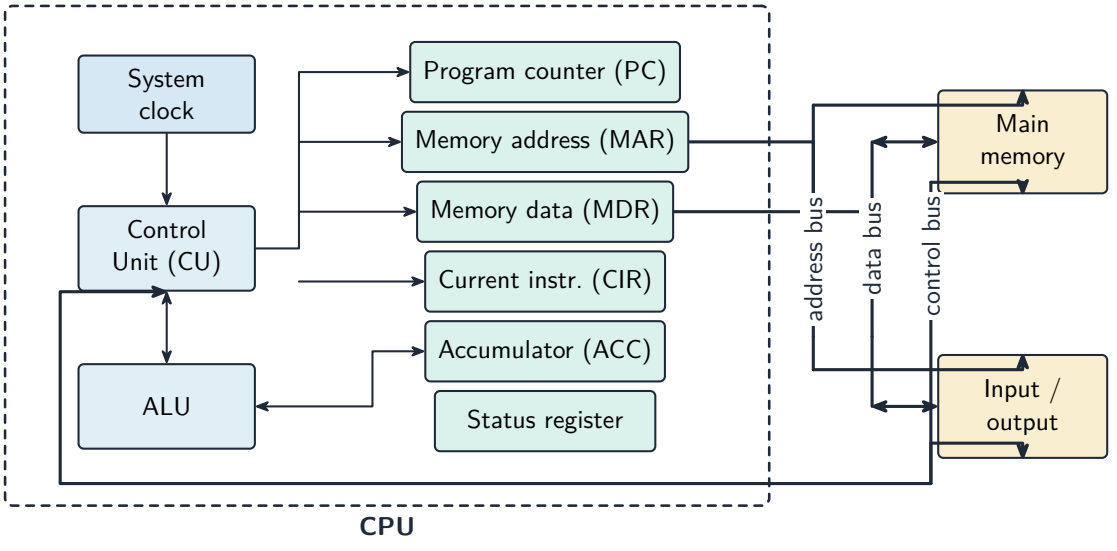
Registers

Registers are tiny, very fast stores inside the CPU. The **special purpose registers** 专用寄存器 each have a fixed job in the cycle:

- **Program Counter** 程序计数器 (PC) — the address of the **next** instruction.

- **Memory Address Register** 内存地址寄存器 (MAR) — the address being read or written.
- **Memory Data Register** 内存数据寄存器 (MDR) — the data going to or from memory.
- **Current Instruction Register** 当前指令寄存器 (CIR) — the instruction being decoded.
- **Accumulator** 累加器 (ACC) — the value the ALU is working on.
- **Status Register** 状态寄存器 — holds **flags** 标志 (carry, zero, negative, overflow) used by branches.
- **Index Register** 变址寄存器 — an offset used in indexed addressing.

General-purpose registers 通用寄存器 are used by the programmer for temporary values during a calculation. Movements of data between registers and memory are written in **register transfer** 寄存器传送 notation — e.g. `MAR ← [PC]` ("copy the contents of PC into MAR").



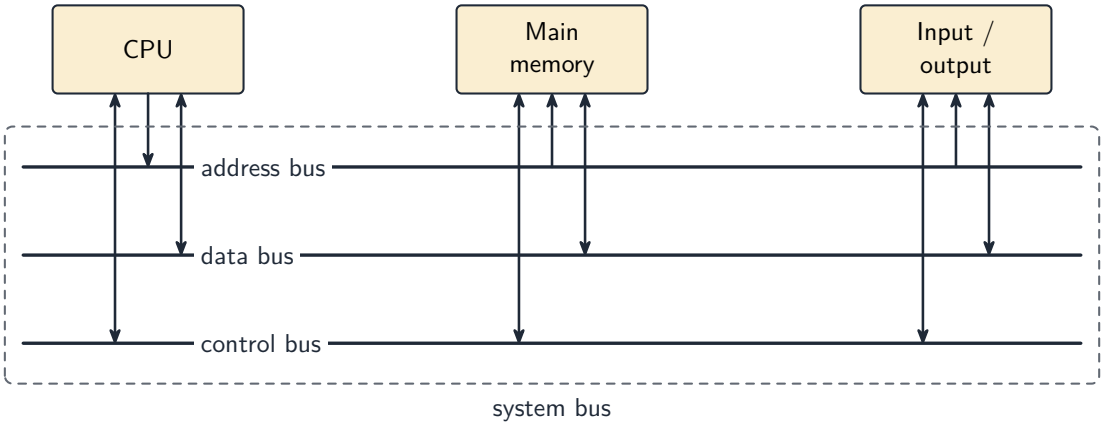
The Von Neumann CPU: registers, control unit and ALU linked by buses

Buses

Three internal **buses** 总线 (sets of parallel wires) connect the parts:

- **address bus** 地址总线 — carries the memory address. **One-way** (CPU → memory).
- **data bus** 数据总线 — carries the data. **Two-way**.
- **control bus** 控制总线 — carries control signals (read, write, interrupt). **Two-way**.

An n-bit address bus can reach 2^n memory locations. The data-bus width sets how many bits move per access (often the word size).



The three system buses connecting the CPU, memory and input/output

What affects performance

- **clock speed** — more cycles per second.
- **number of cores** 核心 — a multi-core CPU runs several threads at once.
- **word size** 字长 — a 64-bit CPU handles 64-bit chunks per cycle and can address far more memory than a 32-bit one.
- **amount of RAM** 随机存取存储器 — more RAM holds more of the working set; too little forces the OS to **page** 页 to disk.
- **cache memory** 高速缓存 size — more cache cuts average memory access time.
- **secondary storage** 辅助存储器 type — an SSD loads programs far faster than an HDD.
- **bus width and speed** — wider/faster buses move data more quickly.

Match the specs to the workload: a quad-core beats a dual-core on parallel work, but

higher per-core speed wins on single-threaded work.

Ports

A **port** 端口 is a physical socket for connecting a **peripheral** 外围设备:

- **USB** (Universal Serial Bus) — general-purpose (keyboards, drives, phones).
- **HDMI** (High Definition Multimedia Interface) — digital video and audio to a screen.
- **VGA** (Video Graphics Array) — older analogue video output to a monitor.
- **Ethernet (RJ-45)** — wired LAN. Audio jacks — headphones/microphone.

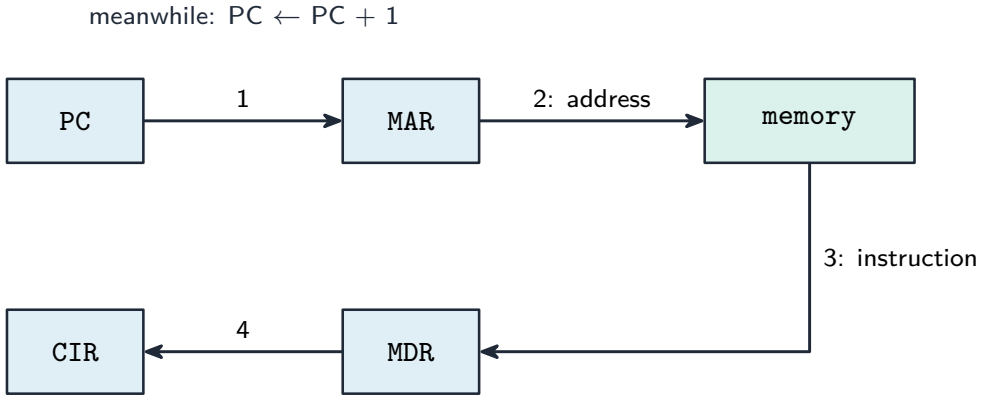
Different ports use different signals, so an HDMI cable will not fit a USB socket. USB-C is unusual in carrying video, data and power.

Fetch-Execute cycle

The CPU repeats the **fetch-execute cycle** 取指-执行周期, one run per machine instruction.

Fetch

1. the PC's address is copied to the MAR.
2. the PC is **incremented** to point to the next instruction.
3. a **read** signal goes over the control bus.
4. memory puts the instruction on the data bus.
5. it is copied into the MDR, then into the CIR.



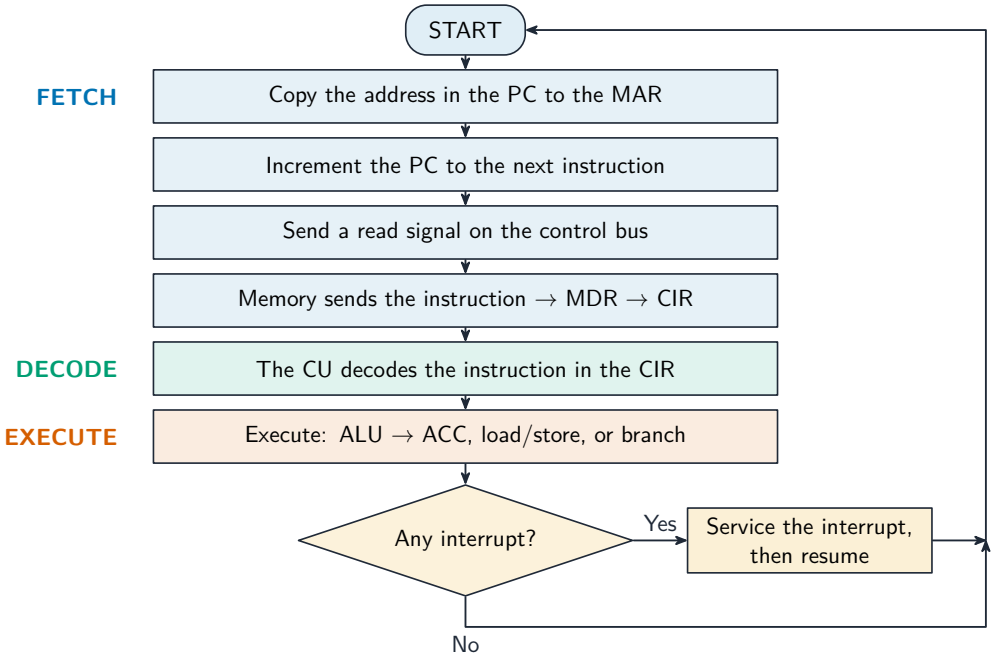
The register transfers in a fetch: PC → MAR → memory → MDR → CIR, with the PC incremented

Decode

The CU decodes the instruction in the CIR — what operation, and which operands or addresses.

Execute

The CU carries it out: arithmetic/logic goes to the ALU (result to the ACC); a load/store moves data between memory and a register; a branch changes the PC. Then the cycle repeats.



The fetch-execute cycle, with a check for interrupts each time

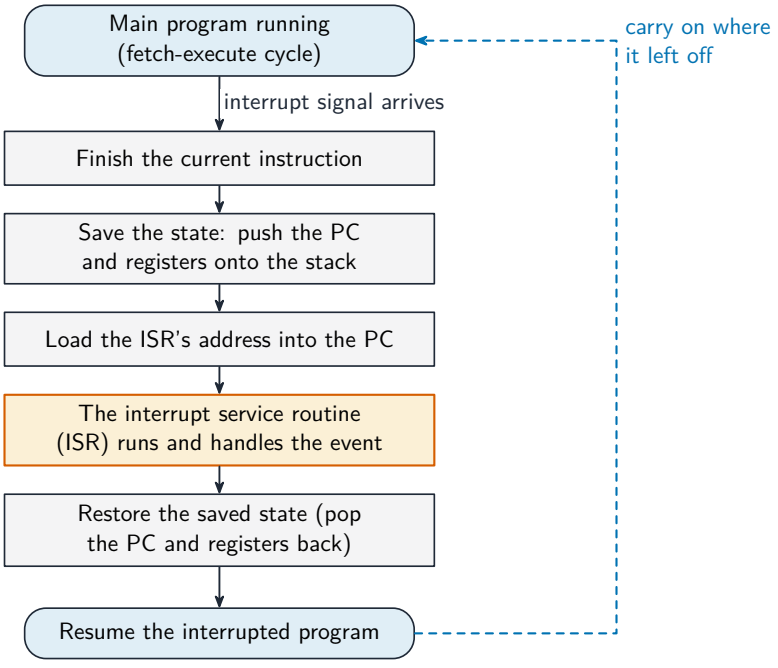
Interrupts

An **interrupt** 中断 is a signal that **pauses** the normal cycle so the CPU can handle an urgent event (a key press, a packet arriving, a hardware fault, division by zero, the OS timer).

Handling one:

1. finish the current instruction.
2. **save the state** (PC and registers).
3. load the address of the **interrupt service routine** 中断服务程序 (ISR) into the PC and run it.
4. the ISR handles the event.
5. **restore** the saved state and carry on.

Interrupts let the system respond promptly without the CPU constantly checking devices, and are how the OS multitasks.



How an interrupt fits into the fetch-execute cycle

Assembly language and machine code

The CPU actually runs **machine code** 机器码 — bit patterns, specific to one architecture. **Assembly language** 汇编语言 is a readable form, with one instruction per machine instruction, written using **mnemonics** 助记符 like `LDD`, `ADD`, `JMP`. An **assembler** 汇编器 translates it to machine code.



An assembler turns mnemonics into machine-code bit patterns

Two-pass assembler

A two-pass assembler reads the source twice:

- **pass 1** builds a **symbol table** 符号表: each time a **label** 标签 (like `LOOP:`) appears, record its address; no code yet.
- **pass 2** generates code: translate each instruction, and when one refers to a label (like `JMP LOOP`), look up its address in the symbol table.

Two passes handle **forward references** 前向引用 (a jump to a label defined later).

Example instruction set

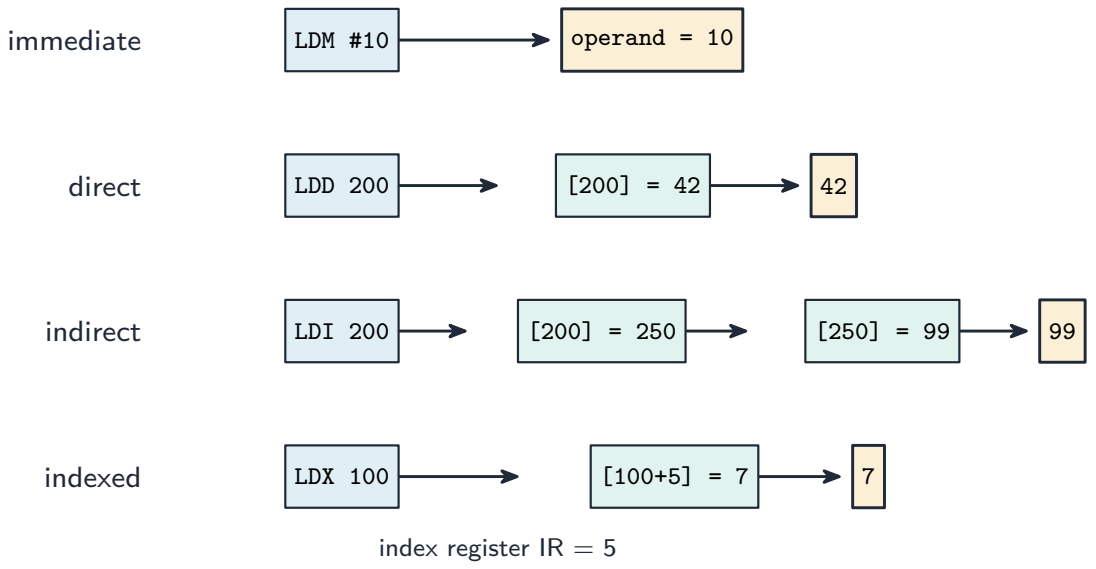
Cambridge uses a small generic set: data movement (`LDD`, `LDM`, `LDI`, `LDX`, `STO`, `MOV`), arithmetic (`ADD`, `SUB`, `INC`, `DEC`), logic/bit (`AND`, `OR`, `XOR`, `LSL`, `LSR`), compare and branch (`CMP`, `JMP`, `JPE`, `JPN`), I/O (`IN`, `OUT`), and `END`. The exact mnemonics are given in the paper's reference table.

Addressing modes

The **addressing mode** 寻址方式 (the **modes of addressing**) says how the CPU finds the operand:

- **immediate addressing** 立即寻址 — the operand is the value in the instruction. `LDM #10` loads 10.
- **direct addressing** 直接寻址 — the instruction holds an address; the operand is the value there. `LDD 200`.
- **indirect addressing** 间接寻址 — the instruction holds an address that holds **another** address, which is the data. `LDI 200`.
- **indexed addressing** 变址寻址 — effective address is `address + index register`; used for arrays. `LDX 100` with IR = 5 reads address 105.

(**Relative addressing** 相对寻址 gives the address as an offset from the PC — used for jumps.)



How each addressing mode reaches its operand — immediate, direct, indirect and indexed

Worked example. Memory holds: location 200 = 250, location 250 = 99, location 105 = 7. The index register holds 5. What is in the accumulator after each of `LDM #200`, `LDD 200`, `LDI 200` and `LDX 100`? Follow how far each mode has to look. `LDM #200` is **immediate** - the operand *is* the number written in the instruction, so the accumulator holds **200**. `LDD 200` is **direct** - go to location 200 and take what is there: **250**. `LDI 200` is **indirect** - location 200 holds 250, which is *another address*, so go on to location 250: **99**. `LDX 100` is **indexed** - add the index register to the address, $100 + 5 = 105$, and read location 105: **7**. Count the hops to keep them apart: immediate 0, direct 1, indirect 2, indexed 1 (once the index has been added).

Tracing an assembly program

To **trace** it: make a table with columns for the PC, ACC, index register, each variable and any flags. Step through the instructions, updating the table after each; follow branches when they change the PC; stop at `END`. A common pattern is a loop over an array using indexed addressing.

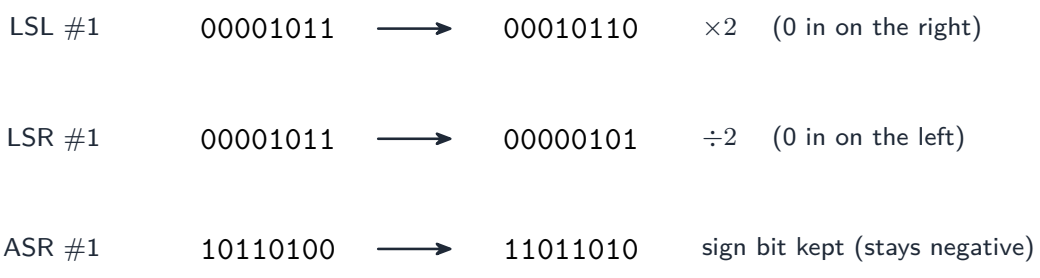
Binary shifts

A **logical shift** 逻辑移位 moves all the bits left or right by some places, filling new positions with 0.

- **left shift by 1** (`LSL #1`) — bits move left, a 0 enters on the right; for an unsigned number this is $\times$ **2**.
- **right shift by 1** (`LSR #1`) — bits move right, a 0 enters on the left; for an unsigned number this is **integer $\div$ 2**.

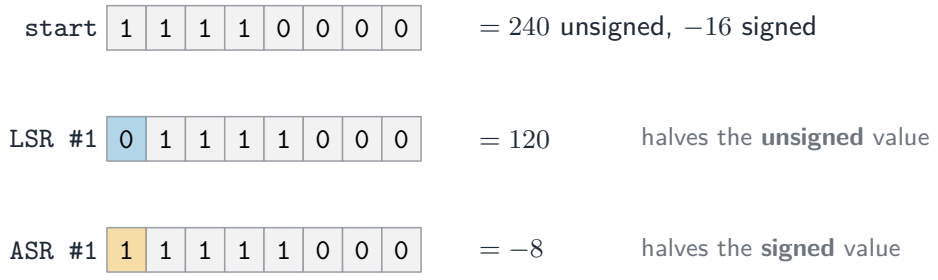
Shifting by n places multiplies or divides by 2^n. Example: `00001011` (11) `LSL #1` $\rightarrow$ `00010110` (22).

An **arithmetic right shift** keeps the sign bit so a negative signed number stays negative. A **cyclic shift** 循环移位 (rotate) feeds the bit that drops off one end back in at the other end, so no bits are lost.



Logical left ($\times 2$), logical right ($\div 2$) and arithmetic right (keeps the sign bit)

The difference between the two right shifts is a single bit. Take `11110000`, which is 240 read as unsigned and -16 read as signed. `LSR #1` brings in a 0 and gives `01111000` = 120, which is the correct half of 240. `ASR #1` copies the sign bit instead and gives `11111000` = -8, which is the correct half of -16. Neither is wrong — each halves the value under one reading.



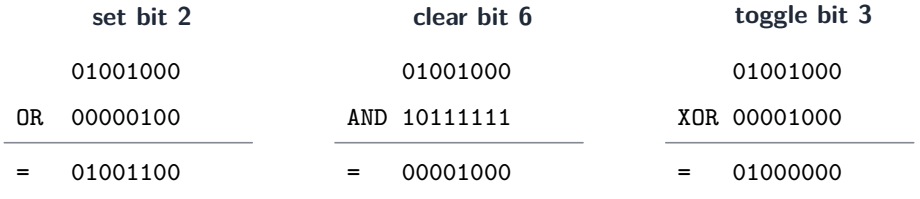
The two results differ in **one bit** — the one that entered on the left. **LSR** always brings in a 0; **ASR** copies the sign bit, so a negative number stays negative and $-16 \div 2$ comes out right.

Logical and arithmetic right shift on the same byte: only the bit that enters on the left differs

Bit manipulation for monitoring/control

Embedded devices often use one **bit** 位 of a register per signal (e.g. bit n = LED n). Using a **mask** 掩码 — **bit masking** — you can:

- **set** bit n: `R = R OR` a mask with bit n set.
- **clear** bit n: `R = R AND` a mask with bit n clear and the rest set.
- **toggle** bit n: `R = R XOR` a mask with bit n set.
- **test** bit n: `R AND` the mask, then check if the result is non-zero.



Set a bit with OR, clear it with AND, toggle it with XOR — each using a mask

Bit manipulation is fast, uses little memory, and lets one byte hold up to 8 on/off states.

Exam tips

- Learn the **fetch-execute cycle** in register-transfer terms (PC, MAR, MDR, CIR, ACC) and what increments the PC.

- Name each register's job; the **address bus is one-way**, the data bus is two-way.
- Distinguish the **addressing modes** (immediate, direct, indirect, indexed) — a frequent question.
- Explain how clock speed, number of cores, cache size and word length affect **performance**.
- For a **binary shift**, state whether it is logical or arithmetic; a left shift multiplies by 2, a right shift divides by 2.

System Software

A-Level Computer Science

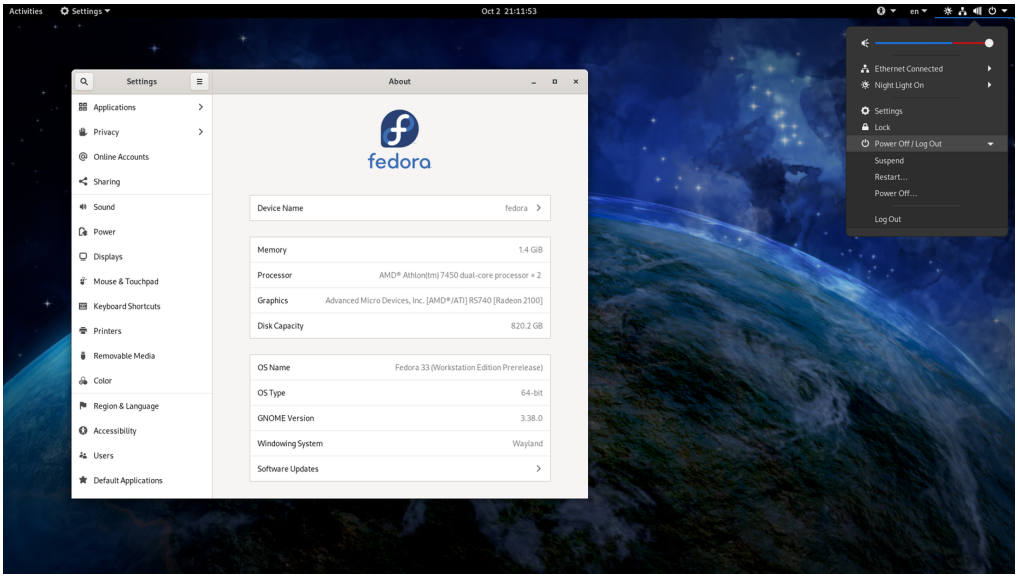
Operating systems

Why a computer needs an OS

Hardware on its own can only fetch and run instructions — it knows nothing about files, programs, networks or users. The **operating system** 操作系统 (OS) is the **software layer** that:

- **manages the hardware** (**processor** 处理器, memory, I/O, storage) for the running programs.
- **provides services** (file system, network, user accounts) through a clear interface, so programs need not talk to the hardware directly.
- **provides a user interface** (command line, GUI, touch).
- **lets several programs share the hardware safely** — each gets fair CPU time and is kept out of the others' memory.

Without an OS, every program would need its own drivers, and only one program could safely run at a time.



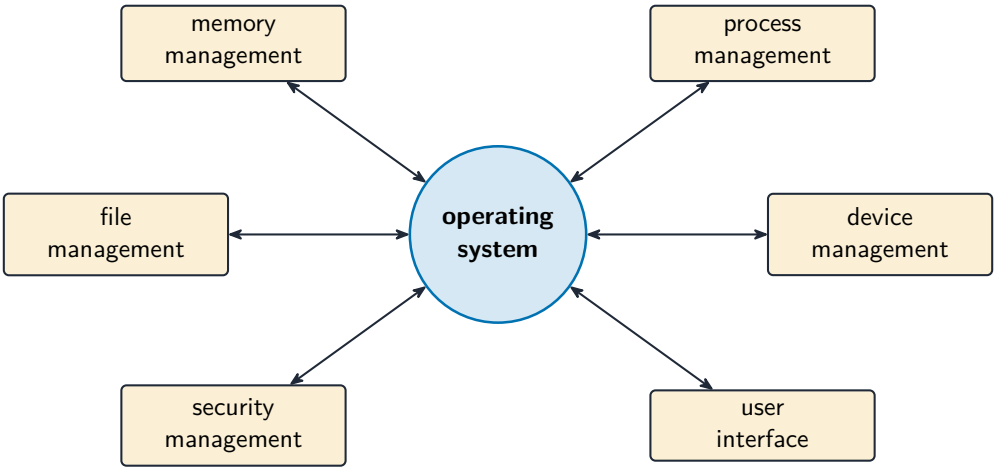
A desktop operating system manages the screen, files and programs for the user



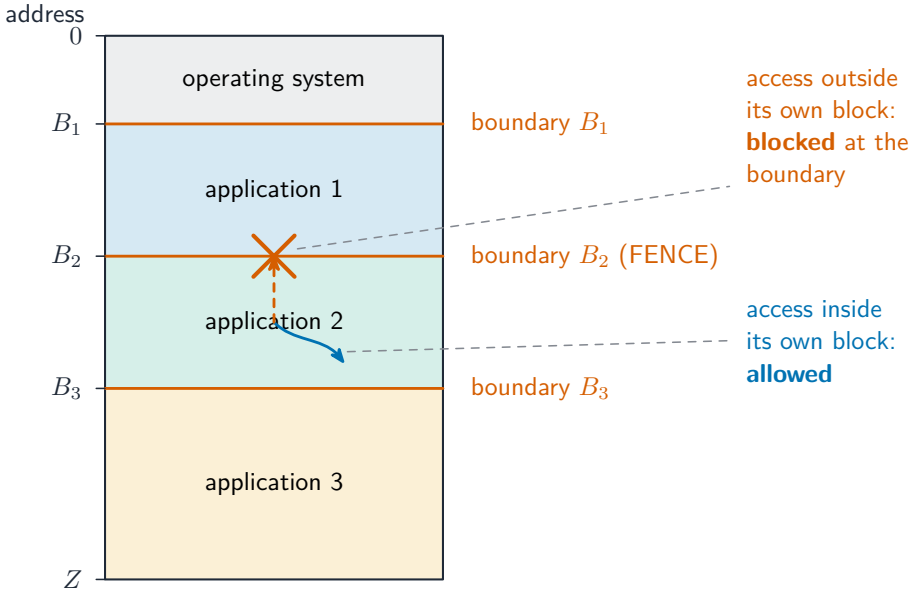
A smartphone runs a mobile operating system such as Android

Key management tasks

- **process management** 进程管理 — load programs, schedule them on the CPU, switch between them, and kill misbehaving ones. (A running program is a **process** 进程.)
- **memory management** 内存管理 — give memory to processes, keep them apart, and use **virtual memory** 虚拟内存 / **paging** 分页 so the working set can exceed physical **RAM** 随机存取存储器.
- **file management** — organise files and folders on **secondary storage** 辅助存储器, control permissions, prevent write corruption.
- **device management** (hardware management) — handle I/O and peripherals through **device drivers** 设备驱动, buffer data, manage interrupts, and give a uniform interface.
- **security management** — accounts, permissions, firewall, encryption.
- **user interface** and **networking**.



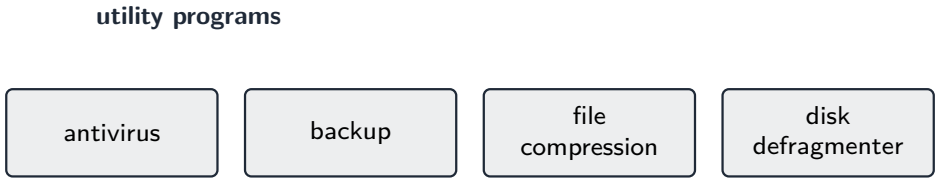
The main jobs the operating system manages



Memory protection keeps each application in its own block of memory

Utility software

Most OSes include **utility programs** 实用程序 to maintain the system:



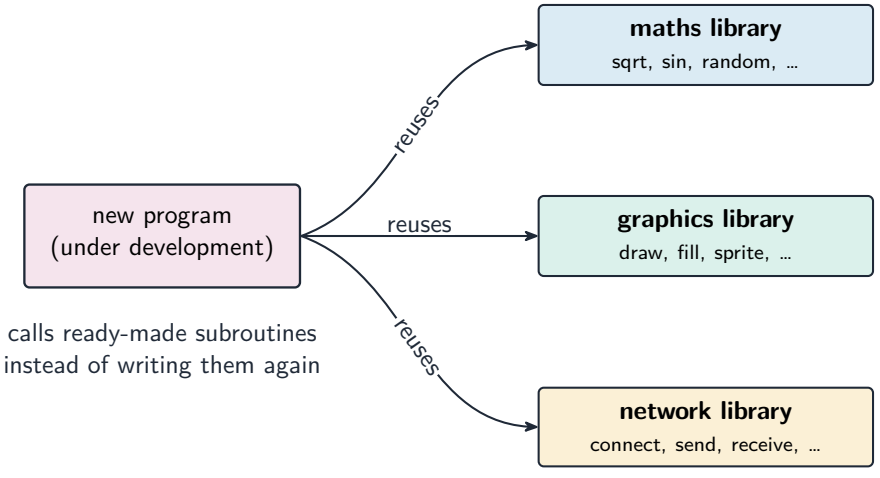
Utility programs: antivirus, backup, compression and defragmenter

- **disk formatter** — a **file / disk management** tool that prepares a disk (sets up its file system); related tools copy, move and delete files.
- **defragmentation software** (**disk defragmenter** 碎片整理) — moves the pieces of fragmented files together on a hard disk to cut seek time (not useful on SSDs).
- **disk contents analysis / disk repair software** — check disk integrity and fix file-system errors and bad sectors.
- **back-up software** (**backup** 备份) — copies user data elsewhere so it can be recovered.
- **virus checker** (**antivirus** 杀毒软件) — scans for malware and quarantines threats.
- **firewall** 防火墙 — filters network traffic by rules.
- **file compression** (**compression** 压缩) / archiving; **system monitor**; **updates**.

Bundling these with the OS saves the user installing each one.

Program libraries

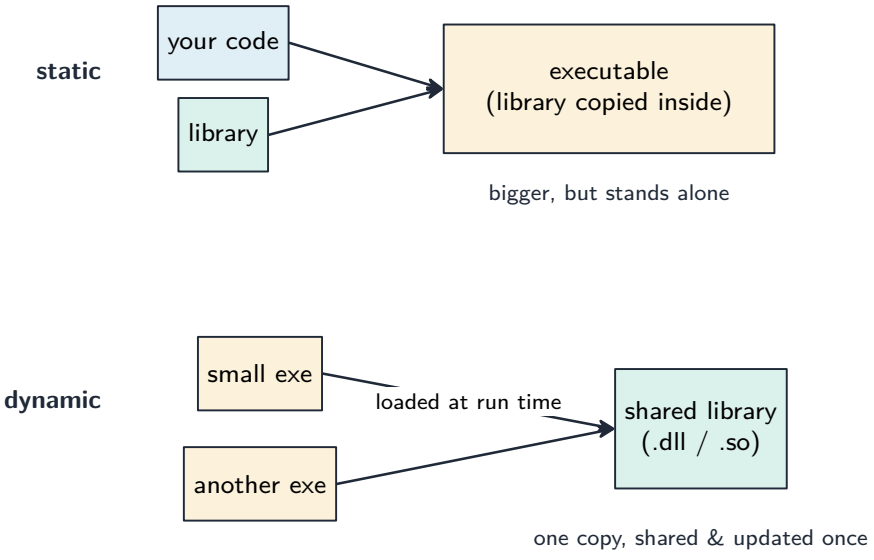
A **program library** 程序库 is pre-written code (**subroutines** 子程序, classes, modules) that programs reuse instead of writing it themselves — e.g. a maths library, a network library, a graphics library.



A new program reusing ready-made routines from libraries

Benefits: **saves time** (off-the-shelf code), **reliable** (well-tested, widely used), and **standardised** (consistent behaviour).

- a **static library** 静态库 is copied into the executable at compile time (stands alone, but larger and needs rebuilding to update).
- a **dynamic library** 动态库 (DLL, Dynamic Link Library; `.so`) is loaded at run time (smaller executables, shared by many programs, updated once for all).



Static: the library is copied into the executable. Dynamic: a shared library file is loaded at run time

Language translators

You write source code; the computer runs **machine code** 机器码. A **translator** 翻译器 converts between them.

Assembler

An **assembler** 汇编器 translates **assembly language** 汇编语言 into machine code, one instruction per instruction. Used for low-level code (embedded systems, drivers).

Compiler

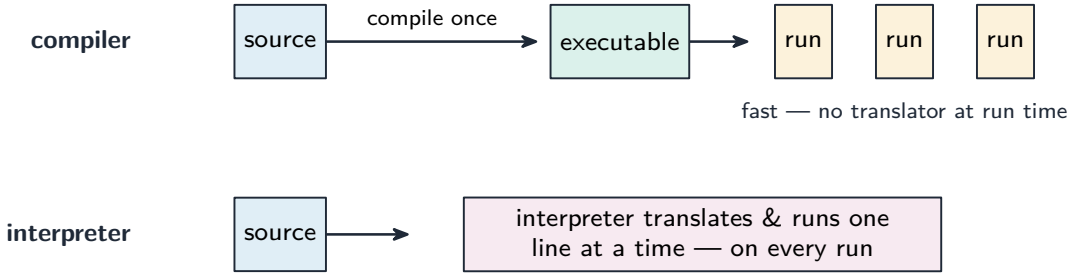
A **compiler** 编译器 translates a high-level program into machine code **once, before it runs**.

- it reports **all** errors at compile time; once clean, it produces a stand-alone **executable** 可执行文件 that runs **without the compiler installed** and can be run many times.
- generally **faster at run time** (no translation while running), but tied to one CPU/OS — recompile for each platform.

Interpreter

An **interpreter** 解释器 translates and runs a high-level program **one line at a time**, producing no executable.

- it reports an error **when it reaches that line**, then stops; you can fix it and continue — good for development.
- the **interpreter must be installed** to run the program; generally **slower** (each run re-translates), but easy to **port** across platforms.



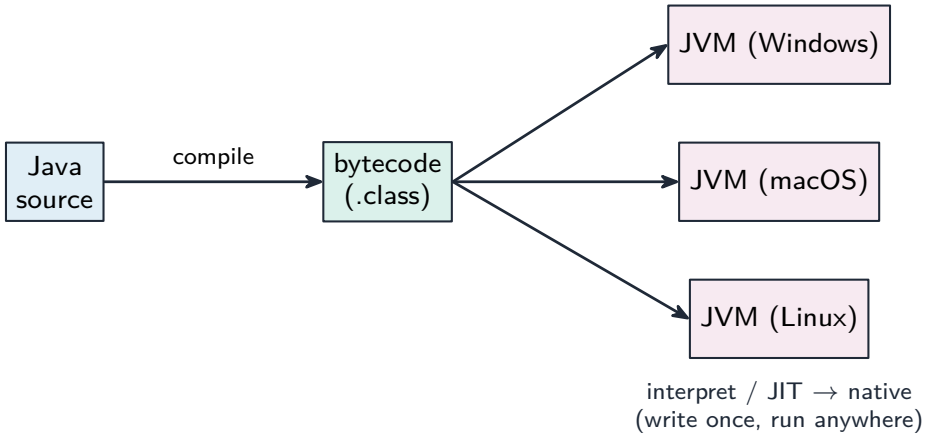
A compiler translates once into a standalone program; an interpreter translates line by line, every run

Choosing between them

Use a compiler when:	Use an interpreter when:
run-time speed matters	you want fast edit–run cycles
distributing to users without dev tools	writing cross-platform scripts
the program runs many times	the program is small or run once
	teaching beginners

Hybrid: Java

Java is compiled into **bytecode** 字节码 (a platform-independent intermediate form), which a **virtual machine** 虚拟机 (the JVM) then interprets — or uses **just-in-time compilation** 即时编译 to turn hot parts into native code. So errors are caught early, the bytecode runs anywhere with a JVM ("write once, run anywhere"), and long-running programs reach near-native speed. C# and Python use similar designs.

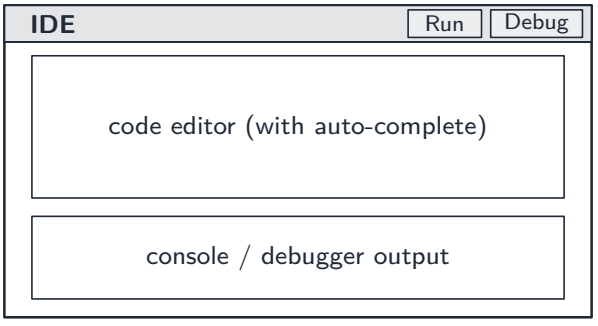


Java compiles to portable bytecode that any JVM runs — write once, run anywhere

Worked example. Java source is compiled to **bytecode**, which a JVM then interprets. Why use both, instead of compiling straight to machine code? A compiler produces machine code for **one** processor and operating system, so a program compiled on one machine will not run on another. Java's compiler instead targets a **virtual** machine, so the bytecode it produces is identical everywhere; each platform then supplies its own **JVM** to interpret that bytecode into its own native instructions. One compiled file therefore runs anywhere a JVM exists - "write once, run anywhere". The price is speed: interpreting bytecode is slower than running native code, which is why a real JVM also uses **JIT** compilation to turn frequently-run bytecode into native code while the program runs. Name **both** sides - the marks are for portability **bought at the cost of speed**.

Integrated Development Environment (IDE)

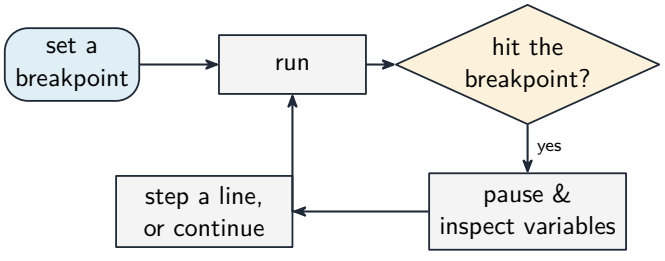
An **integrated development environment** 集成开发环境 (IDE) brings the tools to write, test and debug code into one application:



An IDE bundles the editor, a Run button and a debugger

- **source code editor** with **syntax highlighting** 语法高亮 (keywords, strings, comments in different colours), auto-indent and bracket matching.
- **auto-complete** 自动补全 — suggests names and shows function parameters as you type.
- **translator integration** — compile/run with one keystroke; errors shown inline.
- **debugger** 调试器 — set **breakpoints** 断点 to pause, **step through** line by line, and **inspect variables**.
- **version control** 版本控制 integration (git), **project management**, a help system, **refactoring** 重构 tools (safe renaming), and **unit test** 单元测试 integration.

An IDE speeds development by putting writing → running → debugging → fixing behind one interface. Common IDEs: Visual Studio, PyCharm, Eclipse, VS Code.



A debugger: set a breakpoint, run, then pause to inspect variables and step through the code

Exam tips

- List the OS's jobs (memory, process, file, device and security management) — "manages resources" alone is too vague.
- Compare **compiler vs interpreter vs assembler**: what each translates and when errors are reported.
- Explain what an **IDE** provides (editor, debugger, error diagnostics, auto-complete).

Security, privacy and data integrity

A-Level Computer Science

Security, privacy and integrity — three different ideas

These sound alike but mean different things:

- **security** 安全 — protecting data from **unauthorised** 未授权 access, change or destruction.
- **privacy** 隐私 — an individual's right to **control who sees their personal data**, with consent and a clear purpose.
- **integrity** 完整性 — the data being **accurate and complete** — not corrupted or accidentally changed.

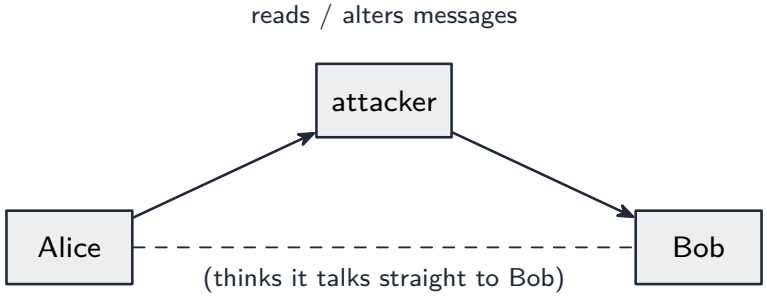
A file can be secure (only the right people can open it) but lack integrity (a typo corrupted it); or accurate but not private (anyone can read it). All three are needed.

Why security matters

Two things to protect: the **data** itself (keep it confidential, intact and available) and the **computer system** (a compromised system can attack others, steal credentials, or be held to ransom).

Threats from networks and the internet

Threats fall into three groups.



A man-in-the-middle attacker sits between the two parties

1. **Malware** 恶意软件 (malicious software) — harmful programs:

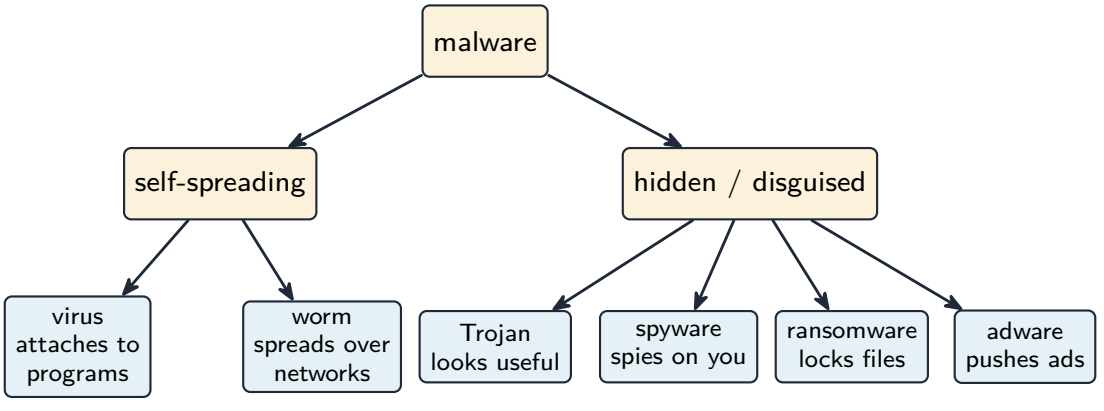
- **virus** 病毒 — self-copying code that attaches to other programs and spreads when they run.
- **worm** 蠕虫 — self-copying code that spreads over **networks** 网络 with no user action.
- **Trojan horse** 木马 — looks useful but hides malicious code.
- **spyware** 间谍软件 — secretly collects information (keystrokes, passwords).
- **ransomware** 勒索软件 — encrypts your files and demands payment.
- **adware** 广告软件 — pushes unwanted adverts.

2. **Tricking people** (social attacks):

- **phishing** 网络钓鱼 — fake emails/sites that trick users into giving credentials.
- **pharming** 域名欺骗 — redirects a user to a fake site even when they type the correct address.
- **social engineering** 社会工程 — tricking people into giving up information.

3. **Attacks on the network:**

- **hacking** 黑客入侵 by **hackers** 黑客 — unauthorised access, often via weak passwords or software flaws.
- **denial of service** 拒绝服务 (DoS/DDoS) — floods a server so real users cannot reach it.
- **eavesdropping** 窃听 — capturing data in transit (a risk on open Wi-Fi).
- **man-in-the-middle** 中间人攻击 — an attacker secretly relays or alters messages between two parties.



Malware by behaviour: self-spreading (virus, worm) versus hidden/disguised (Trojan, spyware, ransomware, adware)

Security measures

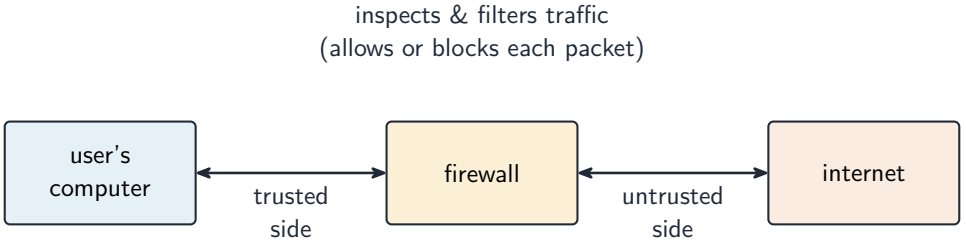
Measures protect both the **security of data** (against loss, theft or corruption) and the **security of the computer system** (its hardware, software and network).

A standalone PC

- a strong password; **antivirus** kept up to date; prompt **software updates**; **backup** 备份 to separate media; full-disk **encryption** 加密; a locked screen.

A networked PC

All the above, plus a **firewall** 防火墙, per-user **permissions** (admin rights only for admins), central management of **user accounts** 用户账户, and **audit logs** 审计日志 (who logged in, what they touched).



A firewall sits between the user's computer and the internet

Across the internet

- **VPN** 虚拟专用网 — encrypts traffic between the user and the corporate gateway.
- **HTTPS / TLS** — encrypt web traffic.
- **digital signatures** 数字签名 — prove who sent a message and that it was not altered in transit.
- intrusion detection — watches traffic for known attack patterns.

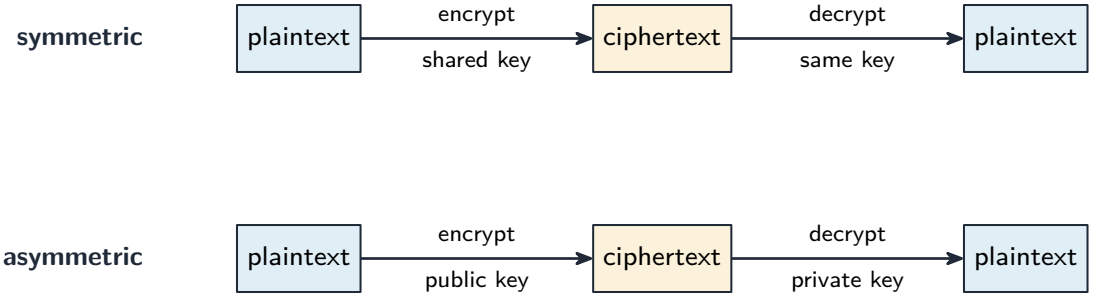
Matching measures to threats

- **interception in transit** → **encrypt** the data (HTTPS, VPN). Intercepted ciphertext is useless without the key.
- **unauthorised access** → strong **authentication** 身份验证 (long passwords; **two-factor authentication** 双因素认证 with a phone code or key); user **authorisation** 授权; lock-out after failed logins.

- **malware** → **anti-virus software** and **anti-spyware** 反间谍软件 with real-time scanning; patching; avoid untrusted downloads.
- **phishing** → user training; email filtering; check the URL before entering credentials.
- **internal threats** → the **least-privilege** 最小权限 principle (give each user only what they need); auditing.
- **DDoS** → rate limiting and traffic filtering.

Protecting the data itself

- **encryption** — turn **plaintext** 明文 into **ciphertext** 密文 with a key. **Symmetric encryption** 对称加密 (AES) uses one shared key; **asymmetric encryption** 非对称加密 (RSA) uses a **public key** 公钥 and a **private key** 私钥. Protects data at rest and in transit.
- **access control** 访问控制 — file permissions (read/write/execute) and **access rights** 访问权限, enforced by the OS.
- **authentication** — **authentication techniques** verify the user: something you know (password), have (token, phone), or are (**biometrics** 生物识别 — fingerprint, face, iris); strongest combined.
- **backups** — keep copies (some **off-site**) so loss or corruption is recoverable.
- **physical security** — locked server rooms, cable locks.



Symmetric uses one shared key; asymmetric uses a public key to encrypt and a private key to decrypt



A security token shows a changing code for two-factor authentication ("something you have")



A fingerprint reader checks "something you are" — a feature of the person, not a password

Data integrity

Data has integrity when it is accurate and complete. Two techniques: **data validation** (catch bad data before storing) and **data verification** (confirm data was entered or transferred correctly).

Validation — does the data make sense?

Validation 验证 checks data against sensible rules, automatically:

- **range check** — within limits (a month is 1–12).
- **limit check** — on the correct side of a single limit (e.g. age $\geq$ 18).
- **existence check** — the referenced item exists (e.g. a product code is in the table).
- **length check** — the right number of characters.
- **type / character check** — the right kind of data (a phone field allows only digits).
- **format check** — matches a pattern (an email must contain `@`).
- **presence check** — required fields are not empty.
- **check digit** 校验位 — an extra digit computed from the others (ISBN, card numbers) that spots transcription errors.
- **lookup check** and **consistency check** (e.g. delivery date $\geq$ order date).

Validation catches data that is wrongly **formatted**, but not data that is the right format yet factually wrong ("Bob" for "Bib").

Verification — was the data entered or transferred correctly?

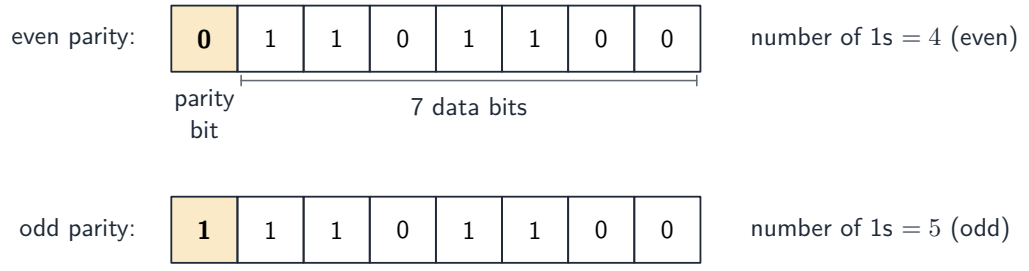
Verification 核对 checks the data was not changed in moving from one place to another.

During entry: **double entry** (type it twice and compare, as for a new password) or **visual** check.

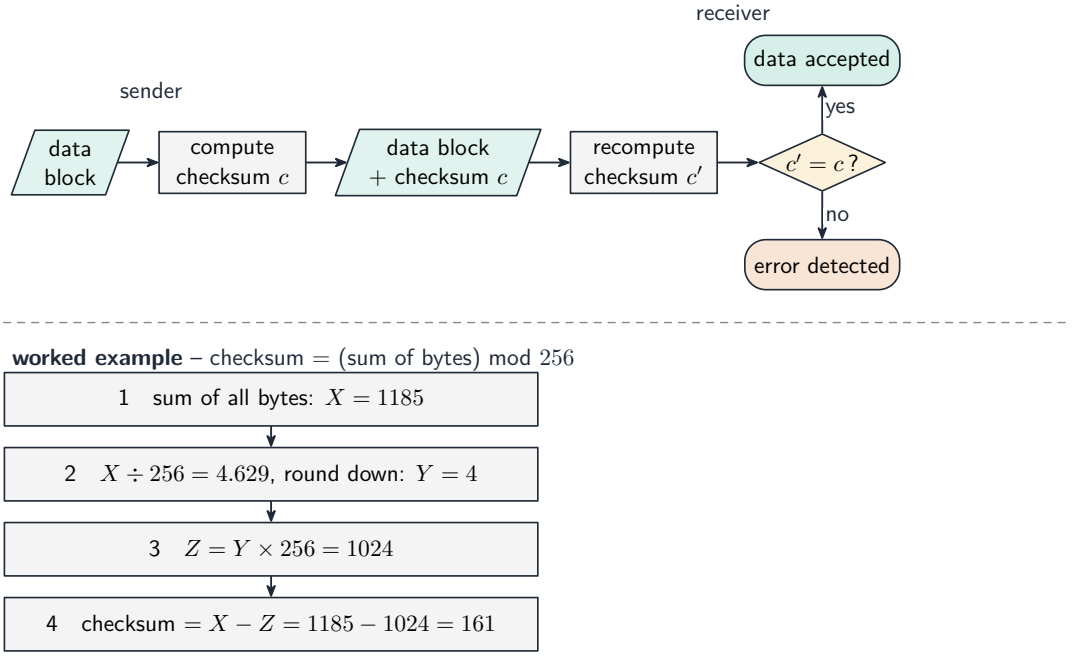
During transfer (bits can flip):

- **parity check** 奇偶校验 — an extra bit makes the number of 1s even (even parity) or odd. The receiver re-counts. Catches single-bit errors.
- **checksum** 校验和 — the sender sends a summary value of the data; the receiver recomputes it and compares.
- **cyclic redundancy check** 循环冗余校验 (CRC) — a stronger checksum using polynomial division, catching many more error types.

A **parity block check** 奇偶块校验 goes further and *locates* the error. Arrange the bytes in a grid: give each byte a **row** parity bit, then compute one extra **parity byte** whose bits are the **column** parity of the bytes above. A single flipped bit now fails **one row and one column** – their intersection pinpoints exactly which bit changed, so it can even be corrected.



The parity bit is set to make the number of 1s even or odd



worked example – checksum = (sum of bytes) mod 256

1 sum of all bytes: $X = 1185$

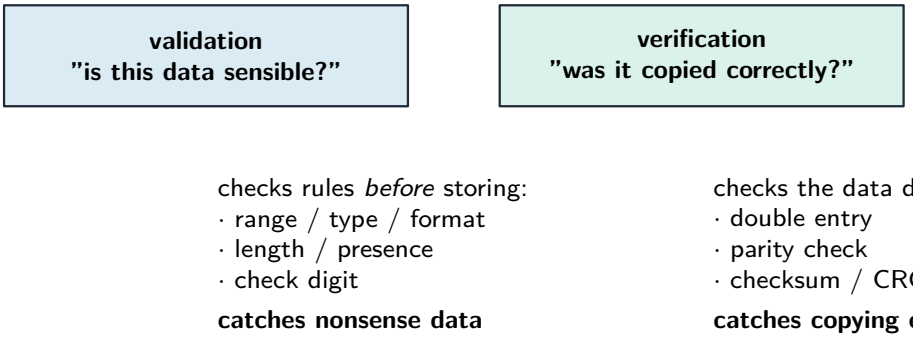
2 $X \div 256 = 4.629$, round down: $Y = 4$

3 $Z = Y \times 256 = 1024$

4 checksum $= X - Z = 1185 - 1024 = 161$

Working out a checksum for a block of data

Verification only proves what arrived matches what was sent — not that the data is correct, and not against deliberate tampering. **Validation** asks "is this sensible?"; **verification** asks "was this copied correctly?" — use both.



Validation checks the data makes sense; verification checks it was copied without change

Worked example. A user types their date of birth as `31/02/2009`, and types their email address twice. Which check catches which error, and what is the difference? **Validation** asks "is this data **sensible**?" - the computer tests it against a rule, and a format or range check rejects `31/02/2009` because February never has 31 days. **Verification** asks "was this data **entered correctly**?" - typing the email twice is **double entry**, and comparing the two copies catches a typing slip. The limit is what makes this a favourite question: validation can never tell you the data is **right**, only that it is **possible** - `01/02/2009` passes every validation rule even if the user was actually born on a different day. Say what each check **can** and **cannot** catch.

Exam tips

- Keep the three ideas separate: **security** (keeping data safe), **privacy** (who may see it), **integrity** (keeping it correct).
- Match each **threat** (malware, hacking, phishing, interception) to a **measure** (firewall, encryption, authentication, access rights).
- Encryption protects **confidentiality, not integrity** — use a checksum, parity or check digit for integrity.
- Distinguish a **virus, worm and Trojan** and how each spreads.

Ethics and Ownership

A-Level Computer Science

Ethics for computing professionals

A **computing professional** is someone whose work — software, systems, networks, data — affects other people. Because the work is technical, others often cannot judge whether it was done well or honestly. So the profession follows shared **ethics** 伦理 (principles for good behaviour).

Why ethics matters

- **trust** — users and employers trust professionals to act in their interest. Without that trust, software loses credibility.
- **impact** — software runs medical devices, banking, vehicles. Careless or dishonest work can hurt people.

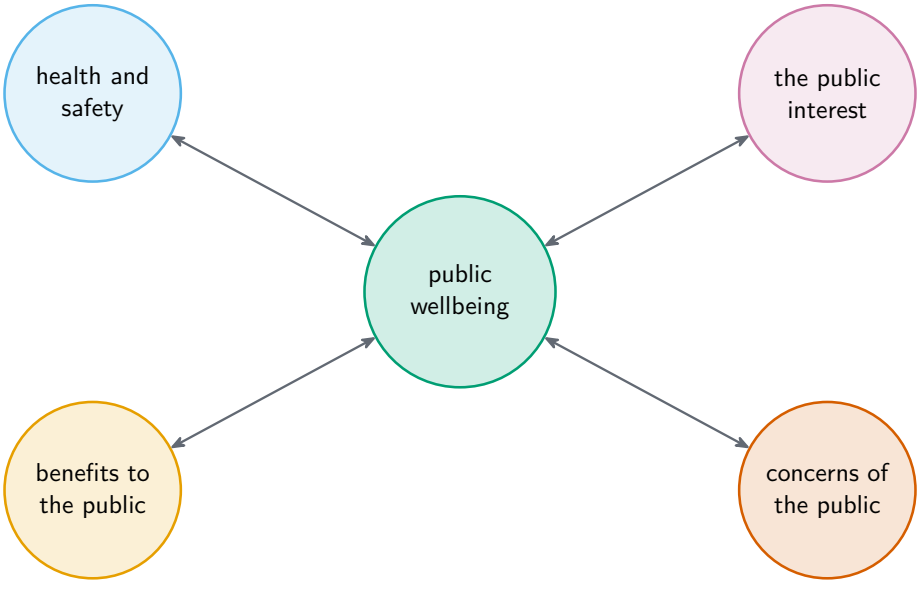
Professional bodies (BCS, ACM, IEEE) publish codes of ethics for members.



CCTV raises privacy concerns — one of the ethical issues a computing professional must weigh



Discarded electronics (e-waste) are a growing environmental cost of computing



Software development affects the public's wellbeing in several ways

Typical principles

- **public interest first** — protect the safety and welfare of those affected.
- **honesty and competence** — be honest about your skills; don't claim expertise you lack.
- **confidentiality** 保密性 — protect clients' and employers' private information.

- **avoid conflicts of interest** 利益冲突 — don't take work where your interest clashes with the client's.
- keep your skills current; respect **intellectual property** 知识产权 and **privacy** 隐私; treat colleagues fairly.

Acting ethically vs unethically

Acting **ethically** protects users, strengthens reputation, reduces legal risk, and builds trust. Acting **unethically** (skipping testing, hiding bugs, misusing data) can harm real users, lead to dismissal or legal action, damage reputation, and erode trust in technology generally.

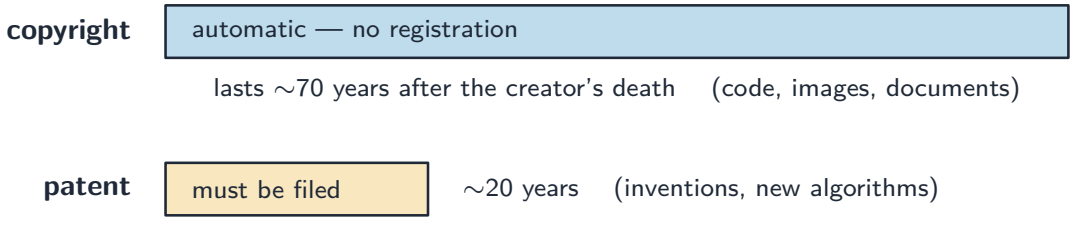
When you face a borderline decision: identify whose interests are affected, check the code of ethics and the law, weigh the consequences, ask a trusted senior, and choose the option that protects users above short-term convenience.

Worked example. Your team's new AI hiring tool sorts CVs ten times faster, but you notice it rejects more older applicants. Shipping it pleases your manager, but it treats one group unfairly. The ethical choice is to **hold it back until the bias is fixed** — public interest and fairness come before short-term convenience.

Copyright

Copyright 版权 is the legal right of the creator of an original work to control how it is **copied, distributed, modified and performed**. It applies automatically (no registration) to source code, software, documents, images, audio and video.

Without copyright, anyone could copy software freely, the developer would not be paid, and plagiarism would be legal. With copyright, developers can earn from their work (encouraging more software), users know who made it, and re-use happens on the developer's terms through licensing. Copyright lasts a long time (often 70 years after the creator's death). General ideas and algorithms are not covered by copyright but may be covered by a **patent** 专利.



Copyright is automatic and long-lasting; a patent must be filed and lasts about 20 years

Software licences

A **software licence** 软件许可证 is a contract granting permission to use software on the owner's terms; choosing and applying one is called **software licencing**.

Commercial (proprietary)

- **commercial software** is sold: you **buy a licence**; the software is used only within its terms.
- the source code is not given (a **proprietary** 专有 product); you cannot modify or redistribute it.
- examples: Microsoft Office, Adobe Photoshop, most games.

Used when the developer wants **revenue** per user and to **keep control** of the code.

Open-source

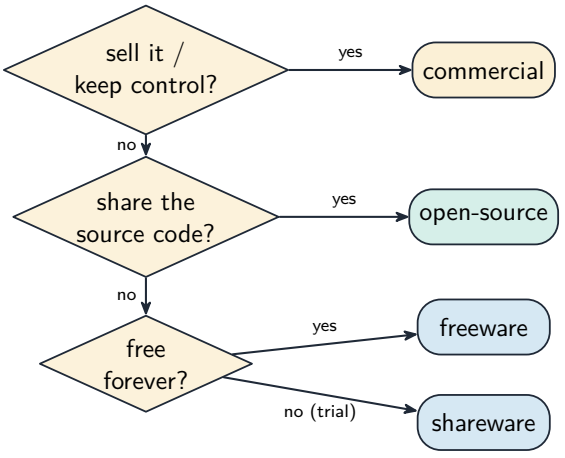
- the source code is **public**; users can read, modify and redistribute it (**open-source** 开源).
- **permissive** licences (MIT, BSD) allow almost any use; **copyleft** 著佐权 licences (GPL) require that modified versions are released under the same licence ("share-alike").
- the **Free Software Foundation** (FSF) and the **Open Source Initiative** (OSI) promote and approve open-source licences.
- examples: Linux, Python, Apache.

Used when the developer wants the software **widely used and improved by the community**.

Freeware and shareware

- **freeware** 免费软件 — free of charge, no source code, may be redistributed but not modified (Acrobat Reader, WhatsApp).
- **shareware** 共享软件 — free for a trial period, then you pay to keep using it; no source code.

Type	Cost	Source	Redistribute	Modify
Commercial	Paid	No	No	No
Open-source	Free	Yes	Yes	Often, with conditions
Freeware	Free	No	Yes	No
Shareware	Free trial, then paid	No	Sometimes	No



Choosing a licence from the developer's goal

To justify a licence choice, link it to the developer's goal (revenue, reach, community), the user's needs (cost, customising), and the use case.

Artificial Intelligence (AI)

Artificial intelligence 人工智能 builds systems that do tasks once thought to need **human intelligence** — recognising speech and images, translating, playing games, driving.

Most modern AI uses **machine learning** 机器学习 — algorithms that improve at a task by learning patterns from large amounts of data, instead of being programmed step by step. **Deep learning** 深度学习, using **neural networks** 神经网络 with many layers, is the leading approach today.

Everyday examples

AI tasks split into two kinds — **understanding input**, and **producing output or decisions**.

Understanding input:

- **speech recognition** 语音识别 — spoken words to text (voice assistants).
- **image recognition** 图像识别 — finding objects, faces or text in images.

Producing output or decisions:

- **machine translation** 机器翻译 — automatic translation between languages.
- **recommendation systems** 推荐系统 — suggesting products, videos or music.
- **autonomous vehicles** 自动驾驶汽车 and robots.

A common exam scenario: a program reads a label with a camera, translates it, and reads it aloud — using **optical character recognition** 光学字符识别 to find the words, machine translation to convert them, and **text-to-speech** 文本转语音 for the audio.



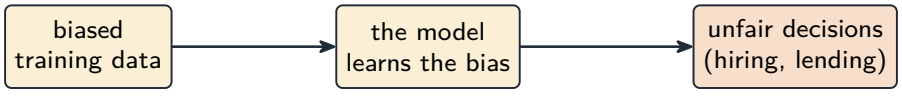
A common scenario: OCR → machine translation → text-to-speech reads a foreign label aloud

Benefits

- **accessibility** — speech/image AI helps users with impairments; translation helps non-native speakers.
- **productivity** — automating repetitive tasks frees people for creative work.
- **decision support** — AI spots patterns in huge datasets (medical diagnosis, fraud detection).
- **always available**, and **personalised** to each user.

Concerns

- **bias** 偏见 — unfair patterns in the training data become unfair AI decisions (hiring, lending).
- **job displacement** — AI may replace some roles.
- **privacy** — training often uses large amounts of personal data.
- **transparency** — large models are "black boxes", hard to explain.
- **accountability** — when AI is wrong, who is responsible: developer, user, or operator?
- **misuse** — deepfakes, misinformation, surveillance.



How bias gets into AI: biased data → a biased model → unfair decisions

Professionals must **understand the limits** of the AI they build, **inform users**, and **reduce harm**.

Exam tips

- Answer ethics questions against a **professional code of conduct** (public interest, competence, honesty), not personal opinion.
- Distinguish **copyright** (protects the expression) from a **patent** (protects an invention).
- Compare software licences: proprietary, open-source, freeware, shareware and FOSS.

Databases

A-Level Computer Science

File-based storage and its limits

Before databases, programs stored data in **flat files** 平面文件 — usually one file per program. This is fine for small data but breaks down at scale.



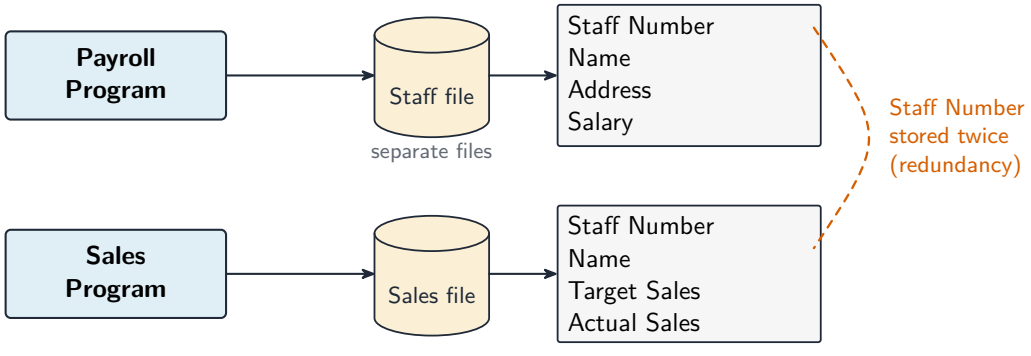
File-based storage keeps data in separate files, like papers in a filing cabinet — hard to search and easy to duplicate

Limitations

- **data redundancy** 数据冗余 — the same data (a customer's address) is held in several files.
- **data inconsistency** 数据不一致 — redundant copies updated separately get out of sync.
- **data dependence** — programs are tied to the file format; change the format and every program must be rewritten.
- hard to enforce **integrity** 完整性, hard to share safely, weak querying, and weak per-field security.

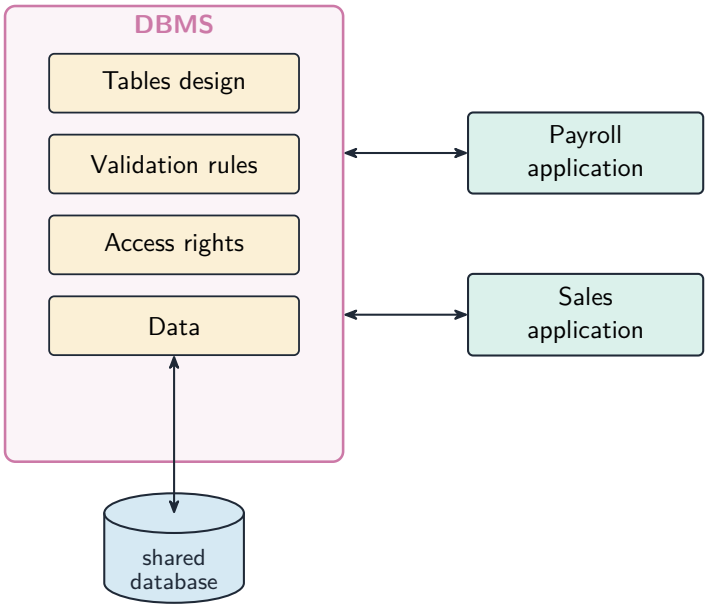


The files themselves are stored on devices such as hard disk drives



The file-based approach: each program keeps its own files

A **relational database** 关系数据库 fixes these by storing data in **tables** managed by one piece of software (the DBMS) that all programs use.



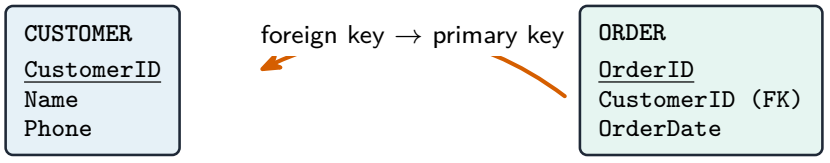
The database approach: one DBMS serves all the programs

Relational model — terms

- **table** 表 (relation) — a grid of rows and columns; one table per type of **entity** 实体 (e.g. `CUSTOMER`).
- **record** 记录 (row, also called a **tuple** 元组) — one row; one instance of the entity.
- **field** 字段 (column, also called an **attribute** 属性) — one column; one piece of information about each record.
- **primary key** 主键 — a field (or fields) that **uniquely identifies** each record; never null or duplicated.
- **foreign key** 外键 — a field whose value matches the primary key of another table, linking the two.
- **composite key** 复合键 — a primary key made of two or more fields together.
- **candidate key** 候选键 — any field(s) that could be the primary key.
- **secondary key** 次键 — a non-primary field that is indexed for fast searching.
- **indexing** 索引 — building an index on a field so look-ups and joins run faster.
- **referential integrity** 参照完整性 — every foreign-key value must match an existing primary key (no orphan records).

A table is written in shorthand with the primary key underlined and foreign keys noted:

```
CUSTOMER(CustomerID, Name, Phone)
ORDER(OrderID, CustomerID, OrderDate)   -- CustomerID is FK → CUSTOMER
```

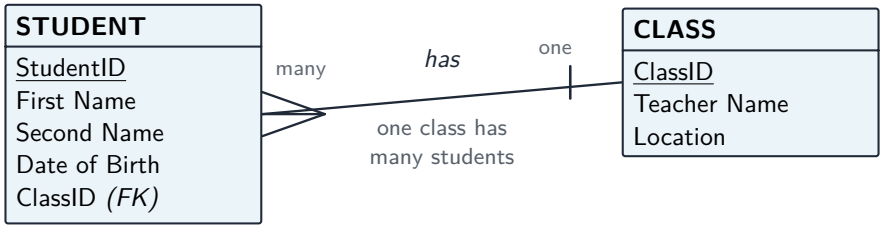


A foreign key links two tables: ORDER.CustomerID matches the primary key CUSTOMER.CustomerID

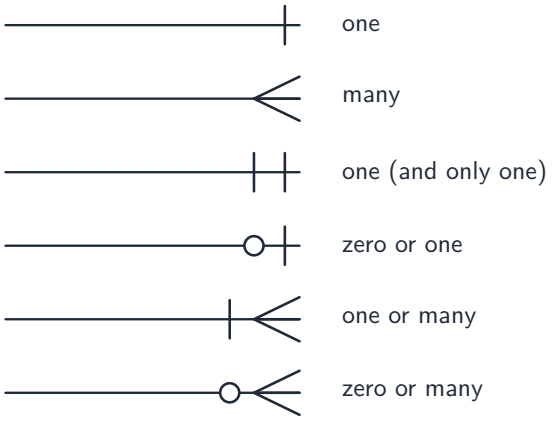
Entity-relationship (E-R) diagrams

An **entity-relationship diagram** 实体关系图 shows the structure: each entity is a rectangle, each relationship a line, with the **cardinality** 基数 marked at each end:

- **one-to-one** (1:1).
- **one-to-many** 一对多 (1:M) — each Customer has many Orders; each Order has one Customer.
- **many-to-many** (M:N) — Students take many Courses, and Courses have many Students.



An E-R diagram: one class has many students



Crow's-foot symbols for the cardinality of a relationship

A many-to-many relationship cannot be stored directly. Break it into two one-to-many relationships through a **link table** 连接表 holding the two foreign keys:

```
ENROLMENT(StudentID, CourseID, EnrolmentDate)
```



a link table turns many-to-many into two one-to-many relationships

A link table resolves a many-to-many relationship into two one-to-many relationships

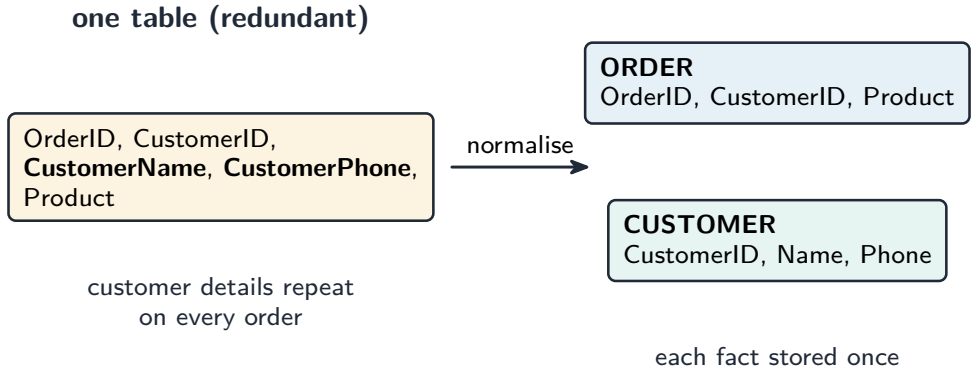
Normalisation

Normalisation 规范化 organises tables to **cut redundancy and inconsistency**, going through **normal forms** 范式 in order.

- **First normal form (1NF)** — every field holds a single (**atomic** 原子) value, with no repeating groups, and a primary key.
- **Second normal form (2NF)** — in 1NF, and every non-key field depends on the **whole** primary key (only matters for a composite key).
- **Third normal form (3NF)** — in 2NF, and every non-key field depends **only** on the primary key, not on another non-key field (no **transitive dependency** 传递依赖).

A 3NF design stores each fact once, so insert/update/delete anomalies disappear. The trade-off is more tables and more joins. Aim for 3NF.

To produce a 3NF design: find the entities and their attributes; choose a primary key for each; split repeating/non-atomic fields (1NF); split fields depending on part of a composite key (2NF); split fields depending transitively on the key (3NF); add foreign keys for the relationships.



Normalisation removes redundancy by splitting repeated data into its own table

Worked example. The table `ORDER(OrderID, CustomerID, CustomerName, ProductID, Quantity)` has the composite primary key `(OrderID, ProductID)`. Normalise it to 3NF. Test each non-key field against the key. `Quantity` depends on **both** `OrderID` and `ProductID`, which is fine. But `CustomerID` depends on **`OrderID` alone** - only *part* of the composite key. That is a **partial dependency**, so the table is not in **2NF**. Split it into `ORDER_LINE(OrderID, ProductID, Quantity)` and `ORDER(OrderID, CustomerID, CustomerName)`. Now test 3NF: in that new `ORDER` table, `CustomerName` depends on `CustomerID`, which is **not** the key - a **transitive dependency**. Split again: `ORDER(OrderID, CustomerID)` and `CUSTOMER(CustomerID, CustomerName)`. Name the **dependency** that breaks each form (partial breaks 2NF, transitive breaks 3NF); "it has repeated data" describes the symptom and earns nothing.

Database Management System (DBMS)

A **DBMS** 数据库管理系统 manages the database centrally. Features that fix the file-based limits:

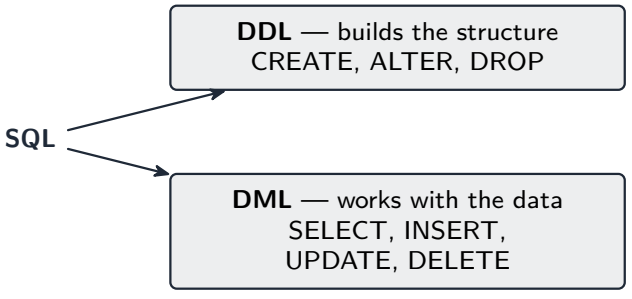
- **data dictionary** 数据字典 — a description of every table, field, type and key; programs query it instead of hard-coding the structure.
- redundancy/consistency control — each fact stored once.

- **concurrent access** 并发访问 control — locks and transactions let many users work at once.
- **backup** 备份 and recovery; security and per-user permissions.
- **integrity rules** — keys, unique and range constraints, enforced centrally.
- **transactions** 事务 — a group of operations that all succeed or all fail.
- **views** 视图 — virtual tables that show each user "their" slice of the data.
- **data management** 数据管理 and **data modelling** 数据建模 — control how data is stored and define its structure as a **logical schema** 逻辑模式 (the logical design, independent of physical storage).
- **data integrity** 数据完整性 and **data security** 数据安全 — enforce correctness and control access centrally.
- a **query processor** 查询处理器 runs queries; a **developer interface** 开发者接口 gives tools and APIs for building applications.

Its tools include a data-dictionary editor, a query builder, a forms builder, a report generator, user management, and an SQL editor.

DDL and DML

SQL 结构化查询语言 (Structured Query Language) has two halves:



DDL builds the database structure; DML works with the data

- **Data Definition Language** 数据定义语言 (DDL) — creates or changes the **structure** (tables, keys, constraints).
- **Data Manipulation Language** 数据操纵语言 (DML) — works with the **data** (insert, update, delete, **query** 查询).

DDL basics

```
CREATE TABLE CUSTOMER (
  CustomerID INTEGER PRIMARY KEY,
  Name VARCHAR(50) NOT NULL,
  Phone VARCHAR(20)
);
```

Add a foreign key:

```
CREATE TABLE ORDER (
  OrderID INTEGER PRIMARY KEY,
  CustomerID INTEGER,
  OrderDate DATE,
  FOREIGN KEY (CustomerID) REFERENCES CUSTOMER(CustomerID)
);
```

Modify and drop:

```
ALTER TABLE CUSTOMER ADD Email VARCHAR(100);
DROP TABLE CUSTOMER;
```

Common types: `INTEGER`, `REAL`, `VARCHAR(n)`, `CHAR(n)` (also `CHARACTER(n)`), `DATE`, `TIME`, `BOOLEAN`, `DECIMAL(p, s)`.

DML basics

Query with `SELECT`:



A SELECT query returns only the rows that match its condition

```
SELECT Name, Phone
FROM CUSTOMER
WHERE City = 'London'
ORDER BY Name ASC;
```

`SELECT` lists fields, `FROM` names the table, `WHERE` filters rows, `ORDER BY` sorts.

A **join** 连接 combines two tables using a foreign-key relationship:

```
SELECT C.Name, O.OrderDate
FROM CUSTOMER C INNER JOIN ORDER O
  ON C.CustomerID = O.CustomerID
WHERE O.OrderDate >= '2024-01-01';
```

Aggregate functions 聚合函数 (`COUNT`, `SUM`, `AVG`, `MIN`, `MAX`) are often used with `GROUP BY`:

```
SELECT CustomerID, COUNT(*) AS NumOrders
FROM ORDER
GROUP BY CustomerID;
```

Insert, update, delete:

```
INSERT INTO CUSTOMER (CustomerID, Name, Phone)
VALUES (101, 'Ada Lovelace', '020-1234-5678');

UPDATE CUSTOMER SET Phone = '020-9999-0000' WHERE CustomerID = 101;

DELETE FROM CUSTOMER WHERE CustomerID = 101;
```

Always put a `WHERE` clause on `UPDATE` and `DELETE`, or the change hits **every** row.

Tips for exam SQL

- use the exact table and field names from the question.
- quote strings with single quotes (`'Smith'`); don't quote numbers.
- comparisons: `=`, `<`, `>`, `<=`, `>=`, `<>`.
- `LIKE 'A%'` matches anything starting with A (`%` = any string, `_` = one character); `IN (1,2,3)`; `BETWEEN 10 AND 20`.
- combine conditions with `AND` / `OR` / `NOT`, and end each statement with a semicolon.

Exam tips

- Define the terms exactly: entity, attribute, **primary key, foreign key**, and the relationship types (1:1, 1:many, many:many).
- Give a reason at each normal form: **1NF** (no repeating groups), **2NF** (no partial dependency), **3NF** (no non-key dependency).
- Explain what a **DBMS** provides (data independence, security, integrity, concurrent access).
- Distinguish **DDL** (define the structure) from **DML** (query and change the data).

Algorithm Design and Problem-solving

A-Level Computer Science

Computational thinking

Computational thinking 计算思维 is the set of mental tools for **analysing a problem and designing a solution a computer can run**. Two key ones are abstraction and decomposition.



Computational thinking breaks a big problem into smaller, easier parts — like solving a jigsaw

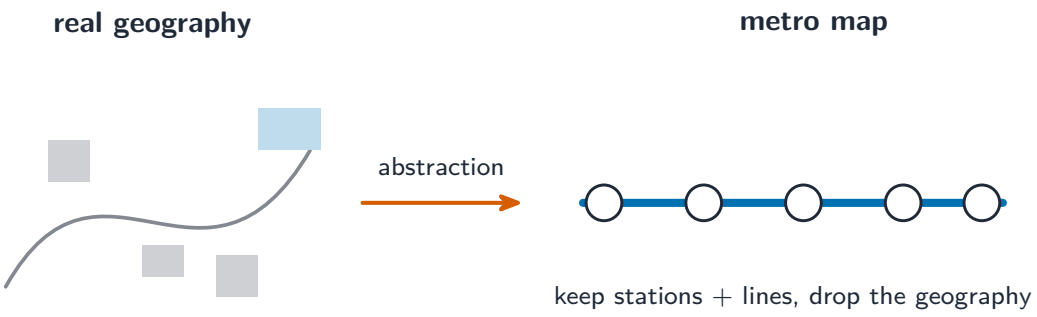
Abstraction

Abstraction 抽象 means **keeping the essential features** of a problem and **ignoring the irrelevant detail**, giving a simpler model.

Examples:

- a **train-network map** keeps the stations and lines but drops the geography.
- a **class** in object-oriented programming keeps only the attributes and methods the system needs.
- a **function** hides a piece of work behind a name.

A full model of any real problem would be too big to reason about, so abstraction is essential.



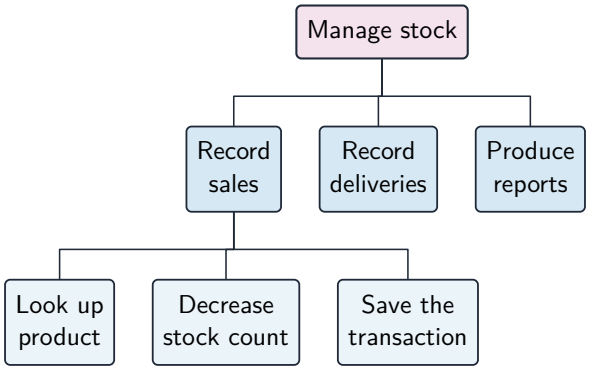
Abstraction keeps the essentials (stations and lines) and drops irrelevant detail (the geography)

Decomposition

Decomposition 分解 means **breaking a large problem into smaller sub-problems**, each easier to solve and tackled one at a time.

1. find the main parts of the task.
2. break each into smaller sub-tasks.
3. continue until each is small enough to design directly.
4. solve the small tasks and combine them.

For stock control: "manage stock" → "record sales", "record deliveries", "produce reports" → ("record sales") "look up product", "decrease stock count", "save the transaction". Decomposition makes big problems manageable, lets a team divide the work, and gives modular code — each module becomes a **procedure** 过程 or function.



Decomposing a program into modules and sub-modules

Algorithms

An **algorithm** 算法 is a **solution expressed as a sequence of defined steps**. Each step is **unambiguous** 无歧义 (one meaning), **deterministic** 确定性 (same input → same output), **finite** (the steps end), and **effective** (each can be done). An algorithm says *what to do*, independent of the programming language used to implement it.

Identifier table

When you start an algorithm, list every piece of data in an **identifier table** 标识符表 — its **variable** 变量 name, **data type** 数据类型, and description:

Variable name	Data type	Description
`Category`	`STRING`	the product category
`SaleDate`	`DATE`	when the item was sold
`ItemCost`	`REAL`	cost of the item
`InStock`	`BOOLEAN`	`TRUE` if in stock

Use **descriptive** names (`ItemCost`, not `x`); common types are `INTEGER`, `REAL`, `STRING`, `CHAR`, `BOOLEAN`, `DATE`, plus arrays. The table forces you to name every piece of data before writing code.

list every piece of data first - name, type, description

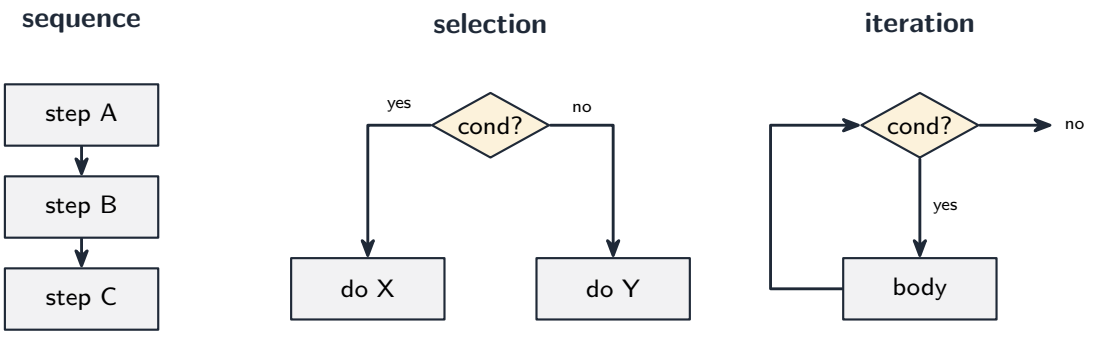
name	data type	description
`ItemCost`	`REAL`	cost of the item
`Category`	`STRING`	the product category
`SaleDate`	`DATE`	when the item was sold
`InStock`	`BOOLEAN`	TRUE if in stock

use **descriptive** names (`ItemCost`, not `x`) - the table forces you to plan the data

An identifier table names every piece of data before you write code

Pseudocode — the three basic constructs

Pseudocode 伪代码 is a structured, language-neutral way to describe algorithms.



The three building blocks of any algorithm: sequence, selection and iteration

1. Sequence

Steps run one after another (**sequence** 顺序):

```
INPUT Name
INPUT Age
OUTPUT "Hello", Name
```

2. Selection

A choice of which steps run, based on a condition (**selection** 选择):

```
IF Age >= 18 THEN
    OUTPUT "Adult"
ELSE
    OUTPUT "Minor"
ENDIF
```

For more options, use `CASE OF ... ENDCASE`.

3. Iteration

Repeating a block (**iteration** 迭代, a **loop** 循环):

```
FOR i ← 1 TO 10
    OUTPUT i
```

```
NEXT i
```

A **WHILE** loop tests the condition **before** each pass (may run zero times); a **REPEAT...UNTIL** loop tests **after** each pass (always runs at least once).

Common operations

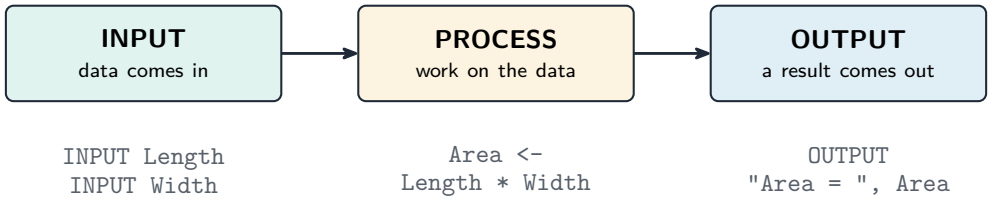
- **assignment** 赋值: `x ← 5` (an arrow; `=` is for comparison).
- input/output: `INPUT variable`, `OUTPUT expression`.
- comparisons `=`, `<>`, `<`, `>`, `<=`, `>=`; logic `AND`, `OR`, `NOT`.
- arithmetic `+ - * /`, plus `DIV` (integer division) and `MOD` (remainder).
- strings: `LENGTH`, `LEFT`, `RIGHT`, `MID`, and `&` for **concatenation** 拼接 (joining).

Input → Process → Output

Every program follows this shape:

```
INPUT Length
INPUT Width
Area ← Length * Width
OUTPUT "Area = ", Area
```

Listing the inputs and outputs first makes the algorithm cleaner.



Every program follows the Input, Process, Output shape

Three notations

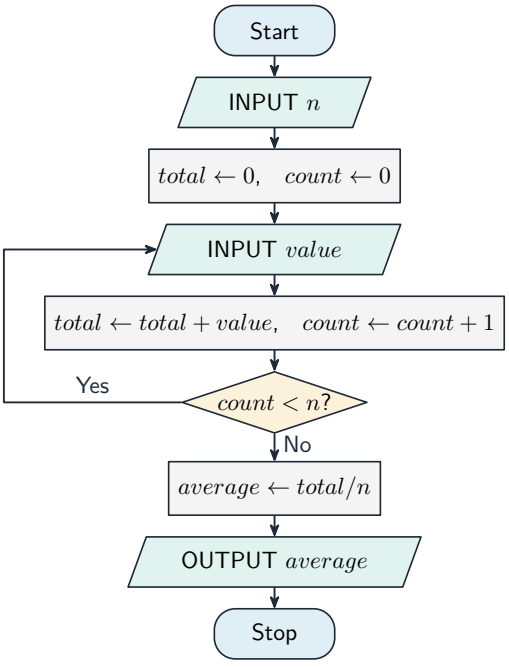
The same algorithm can be written three ways.

- **structured English** 结构化英语 — natural language with indentation and fixed keywords; good for a high-level description.
- **flowchart** 流程图 — a diagram with standard shapes:

Shape	Meaning
Rounded rectangle	Start / Stop
Parallelogram	Input / Output
Rectangle	Process
Diamond	Decision
Arrow	Flow of control

- **pseudocode** — the keyword notation above; closest to code.

You should be able to convert between any pair: each `IF` is a decision diamond, each loop is a back-arrow, and a sequence is stacked rectangles.



A flowchart for averaging a list of numbers, using the standard shapes

Stepwise refinement

Stepwise refinement 逐步求精 starts with a high-level outline and **expands each step** until it is small enough to code. For an average of n numbers:

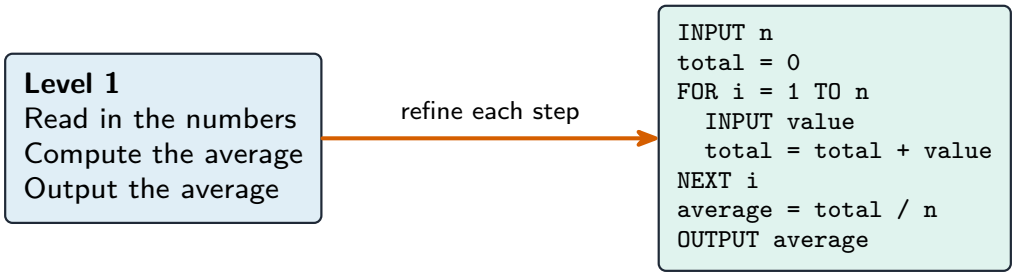
Level 1:

```
Read in the numbers
Compute the average
Output the average
```

Level 2:

```
INPUT n
total ← 0
FOR i ← 1 TO n
    INPUT value
    total ← total + value
NEXT i
average ← total / n
OUTPUT average
```

Each refinement keeps the previous structure and adds detail.



Stepwise refinement: expand each high-level step into detailed pseudocode

Logic statements

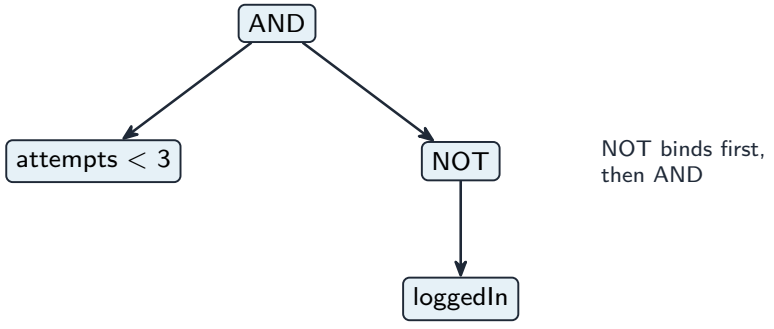
A **logic statement** 逻辑语句 is a **Boolean** 布尔 condition that controls branching, built from comparisons (`x > 10`), connectives (`AND`, `OR`, `NOT`) and brackets. Use it as the condition of `IF`, `WHILE` or `REPEAT...UNTIL`:

```
WHILE attempts < 3 AND NOT loggedIn DO
    INPUT password
    IF password = correctPassword THEN
        loggedIn ← TRUE
    ELSE
        attempts ← attempts + 1
    ENDIF
ENDWHILE
```

Precedence 优先级 (highest to lowest): `NOT`, then `AND`, then `OR`. Use brackets when unsure. Common mistakes:

- `a = 1 OR 2` is wrong — write `a = 1 OR a = 2`.
- `NOT a > 5` means `NOT (a > 5)`, i.e. `a <= 5`.

- `NOT (A AND B)` is the same as `(NOT A) OR (NOT B)` (**De Morgan's law** 德摩根定律) — handy for simplifying conditions.



Precedence: NOT binds to loggedIn first, then AND combines the two sides

Worked example. Write an identifier table and pseudocode to read 10 numbers and output the largest. The **identifier table** names each variable with its data type and purpose: `Count : INTEGER` (loop counter), `Num : REAL` (the number just read), `Max : REAL` (largest so far).

```
Max ← -999999
FOR Count ← 1 TO 10
    INPUT Num
    IF Num > Max THEN
        Max ← Num
    ENDIF
NEXT Count
OUTPUT Max
```

The design decision carrying the marks is **initialising `Max`**: it must start **lower than any possible input** - or, safer still, be set to the **first** number read. Initialise it to `0` and the algorithm wrongly returns 0 for a list of negative numbers, a bug your trace only exposes if the test data include a negative.

Exam tips

- Define an **algorithm** as an unambiguous, finite, deterministic sequence of steps, independent of language.
- Use the three constructs correctly — **sequence, selection, iteration** — and keep an identifier table with data types.
- Break a problem down by **decomposition and abstraction**, then stepwise refinement.

- Write pseudocode that would actually run: declare variables and follow the exam's pseudocode style.

Data Types and Structures

A-Level Computer Science

Choosing data types

Every variable needs a **data type** 数据类型 — the kind of value it holds and the operations allowed:

- `INTEGER` — a whole number (`42`, `-7`). For counts, indexes, IDs.
- `REAL` — a number with a fractional part (`3.14`). For money, measurements.
- `STRING` — characters in quotes (`"Hello"`). For text.
- `CHAR` — a single character (`'A'`).
- `BOOLEAN` — `TRUE` or `FALSE`. For flags.
- `DATE` — a calendar date.

Pick the smallest precise type that fits: `INTEGER` for whole counts, `BOOLEAN` for flags (not the strings `"yes"`/`"no"`).

Records

A **record** 记录 (a **record structure** 记录结构) holds **several fields of different types** under one name — useful when several values describe one thing.

```
TYPE TStockItem
    DECLARE ItemID : INTEGER
    DECLARE Category : STRING
    DECLARE ItemCost : REAL
    DECLARE InStock : BOOLEAN
ENDTYPE
```

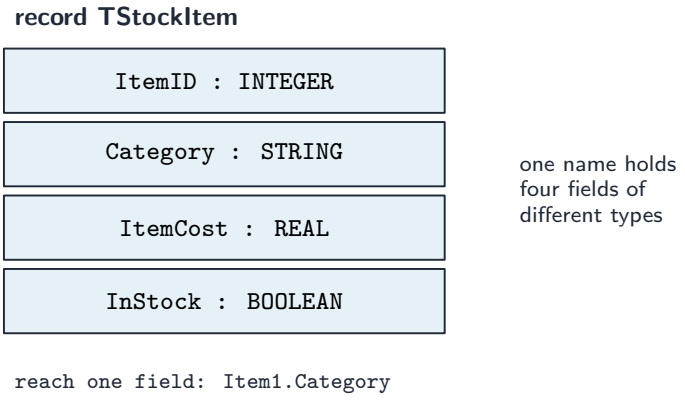
This defines the **type** `TStockItem`; declare variables of it:

```
DECLARE Item1 : TStockItem
DECLARE Items : ARRAY[1:100] OF TStockItem
```

Use dot notation to reach each **field** 字段:

```
Item1.Category ← "Fruit"
OUTPUT Item1.Category, " costs ", Item1.ItemCost
```

Use a record when values **always belong together** (a customer, a stock item); use separate variables for unrelated values.



A record holds several fields of different types under one name

Arrays

An **array** 数组 is an ordered collection of items of the **same type**, under one name, reached by an **index** 索引.

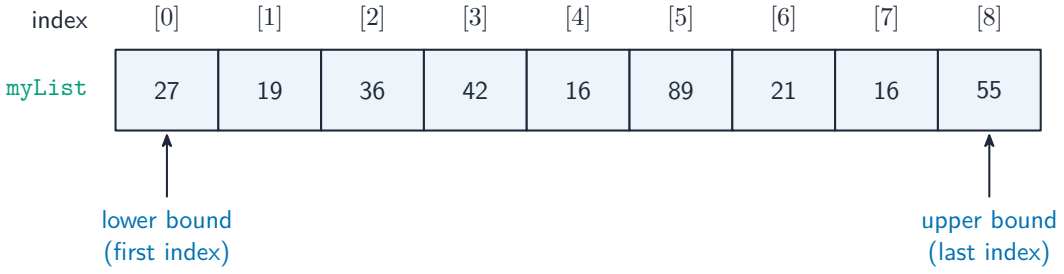
- **element** 元素 — one item in the array.
- **bounds** 边界 — the lowest and highest valid indices.
- **dimension** 维度 — 1-D (a list), 2-D (a table), etc.

1-D arrays

```
DECLARE Names : ARRAY[1:5] OF STRING
Names[3] ← "Cara"
OUTPUT Names[3]
```

Process every element with a `FOR` loop:

```
FOR i ← 1 TO 5
    OUTPUT Names[i]
NEXT i
```

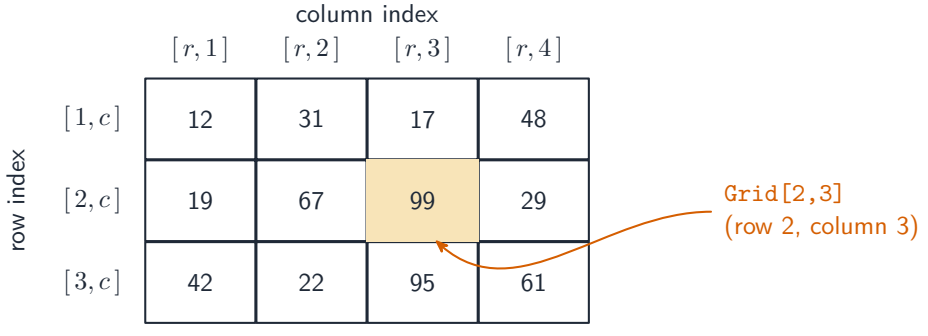


A 1-D array (a list) with indices and bounds

2-D arrays (2D array)

```
DECLARE Grid : ARRAY[1:3, 1:4] OF INTEGER
Grid[2, 3] ← 99
```

The first index is the row, the second the column. Use nested loops to visit every cell. Use **1-D** for a single sequence, **2-D** for two natural dimensions (a grid, rows × columns).



A 2-D array (a table) with row and column indices

Common operations

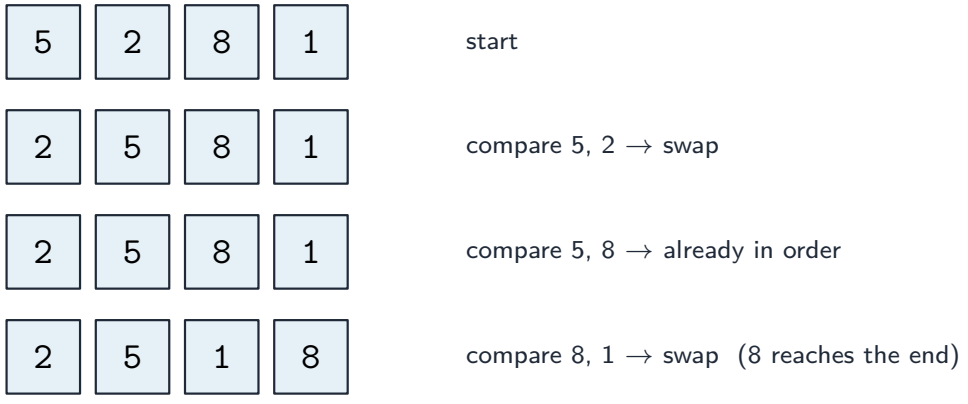
A **linear search** 线性查找 checks each element until found:

```
FOR i ← 1 TO n
    IF A[i] = Target THEN
        OUTPUT "Found at ", i
    ENDIF
NEXT i
```

To find a sum, count, maximum or minimum, set a running variable then sweep through:

```
Max ← A[1]
FOR i ← 2 TO n
    IF A[i] > Max THEN Max ← A[i]
NEXT i
```

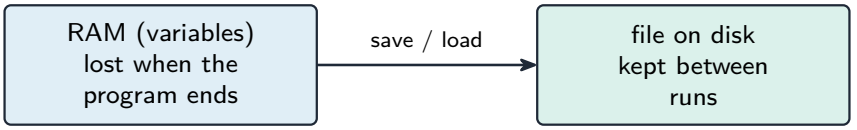
A **bubble sort** 冒泡排序 puts an array in order: pass through it comparing each adjacent pair and **swapping** any that are out of order; repeat the passes until one pass makes no swaps.



One pass of a bubble sort: adjacent pairs are compared and swapped, bubbling the largest value to the end

Files

A **file** 文件 is data stored on **secondary storage** 辅助存储器, kept between program runs. Variables in RAM disappear when the program ends, so to save data permanently (high scores, records, settings) the program writes to a file. Files also let programs share data and restart from a saved state.



Variables in RAM vanish when the program ends; a file on disk persists between runs

A **text file** 文本文件 holds one or more lines of readable characters; programs read and write **text files** line by line. Open a file before use and **close** it after:

```
OPENFILE "data.txt" FOR READ      // or FOR WRITE, FOR APPEND
WHILE NOT EOF("data.txt") DO
    READFILE "data.txt", LineString
    OUTPUT LineString
ENDWHILE
CLOSEFILE "data.txt"
```

`EOF` tests the **end of file** 文件结束 before reading. To write:

```
OPENFILE "log.txt" FOR WRITE
FOR i ← 1 TO 100
    WRITEFILE "log.txt", "Event " & i
NEXT i
CLOSEFILE "log.txt"
```

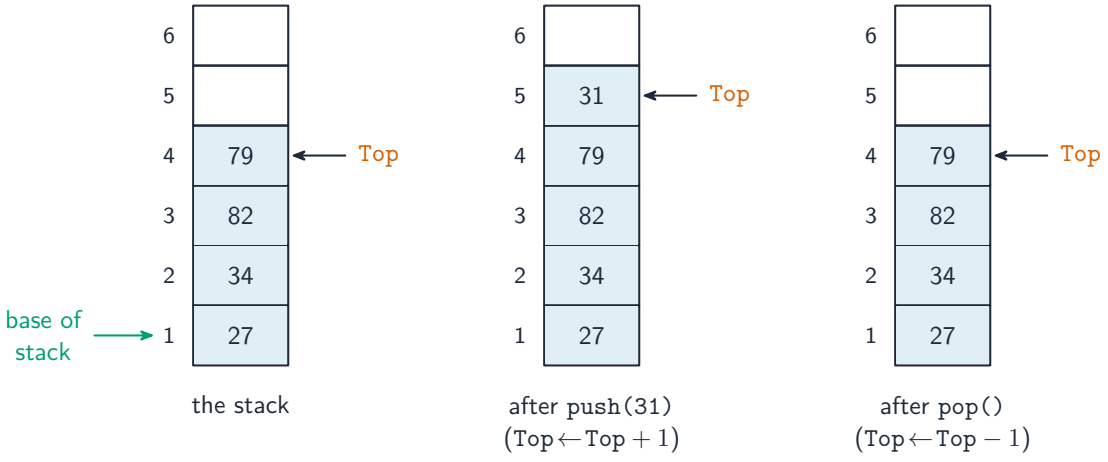
Always **close** every file — otherwise buffered writes may be lost and other programs may be locked out.

Abstract Data Types (ADTs)

An **Abstract Data Type** 抽象数据类型 (ADT) is a **collection of data plus operations on it**, defined by **what** it does, not how it is stored. The user works only through the operations; the implementation is hidden, so it can change without affecting code that uses the ADT. Know three: stack, queue, linked list.

Stack

A **stack** 栈 works in **LIFO** 后进先出 order (Last In, First Out). Operations: **push** 入栈 (add to the top), **pop** 出栈 (remove from the top), peek (look at the top), and tests for empty/full. Uses: undo history, function-call return addresses, expression parsing, backtracking.



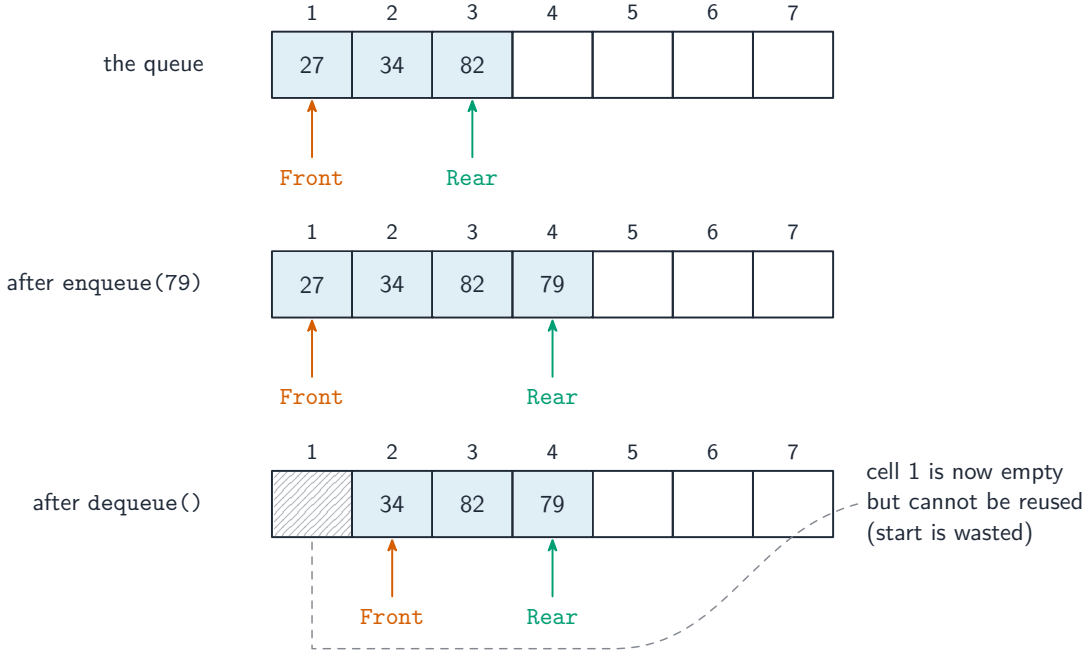
Push and pop change the top pointer; the base pointer stays put



*A pile of books is a stack you can see. You can only add or take a book from the **top**, so the last one you put on is the first one you take off — that is exactly LIFO*

Queue

A **queue** 队列 works in **FIFO** 先进先出 order (First In, First Out). Operations: **enqueue** 入队 (add to the rear), **dequeue** 出队 (remove from the front), and tests for empty/full. Uses: print spooling, scheduling, breadth-first search, buffering.



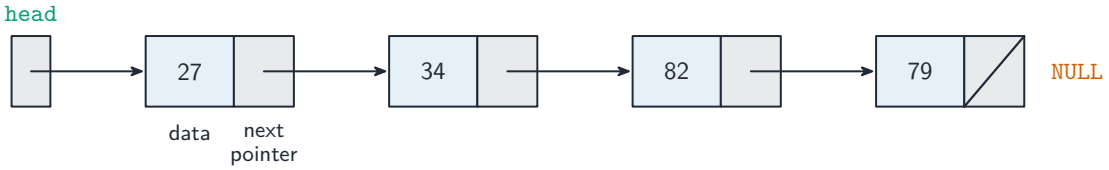
Enqueue adds at the rear; dequeue removes from the front



*A line of people is a queue you can see. You join at the **back** and are served from the **front**, so whoever waited longest is served first — that is exactly FIFO*

Linked list

A **linked list** 链表 stores data as a sequence of **nodes** 节点. Each node holds a value and a **pointer** 指针 to the next node; a head pointer marks the start, and the last node's pointer is a sentinel (e.g. `NULL`). Operations: insert, delete, search, and **traverse** 遍历 (visit each node in order). Its advantage over an array is cheap insertion/deletion (just adjust pointers); its disadvantage is slow random access (you must follow pointers from the head).



A linked list: each node points to the next

Implementing ADTs using arrays

Stack using an array

Hold items in `Stack[1:MaxSize]` with an integer `Top` (0 when empty).

- `Push(x)`: if `Top = MaxSize` the stack is full (**overflow** 溢出); else `Top ← Top + 1`; `Stack[Top] ← x`.
- `Pop()`: if `Top = 0` the stack is empty (**underflow** 下溢); else return `Stack[Top]` and `Top ← Top - 1`.

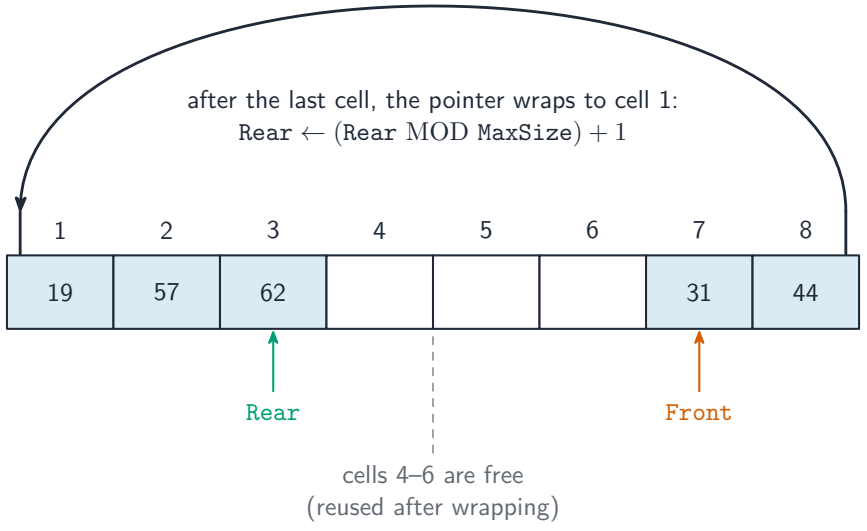
Queue using a circular array

A simple queue lets `Front` and `Rear` march off the end, wasting the start. The fix is a **circular array** 循环数组 — when a pointer reaches `MaxSize` it wraps back to 1:

- `Enqueue(x)`: check full; else `Rear ← (Rear MOD MaxSize) + 1`; `Queue[Rear] ← x`.
- `Dequeue()`: check empty; else return `Queue[Front]` and `Front ← (Front MOD MaxSize) + 1`.

Track a separate count to tell empty from full.

For example, with `MaxSize = 6`: if `Rear = 5`, then `(5 MOD 6) + 1 = 6`, so the next item goes in cell 6; if `Rear = 6`, then `(6 MOD 6) + 1 = 1`, so the pointer **wraps back** to cell 1.



A circular queue wraps the pointers back to the start of the array

Linked list using an array

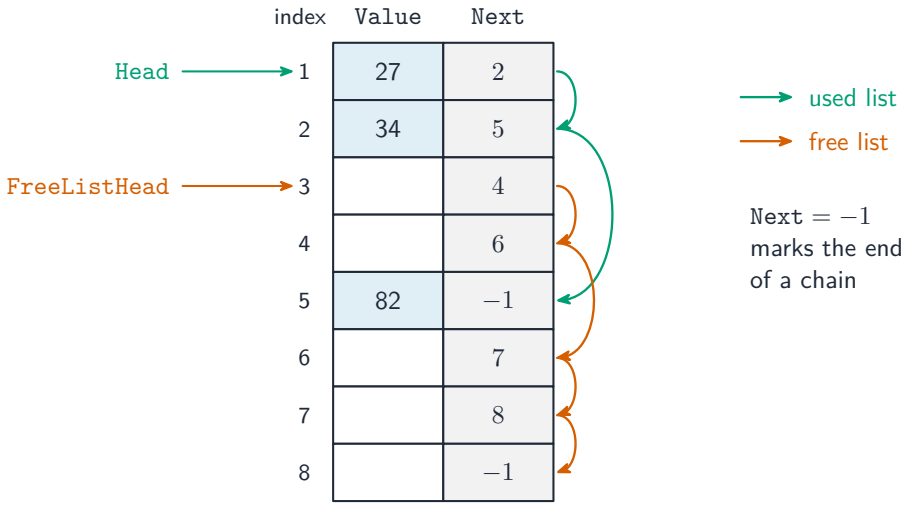
Use an array of records, each with a `Next` index:

```
TYPE TNode
    DECLARE Value : INTEGER
    DECLARE Next : INTEGER     // index of the next node, or -1 for end
ENDTYPE

DECLARE Nodes : ARRAY[1:MaxSize] OF TNode
```

```
DECLARE Head : INTEGER         // index of first node, -1 if empty
DECLARE FreeListHead : INTEGER // first available free node
```

A **free list** 空闲列表 chains the unused slots, just as the data list chains its used ones. To insert: take a slot from `FreeListHead`, set the new node's value and `Next`, and update the previous node's `Next` (or `Head`). To delete: unlink the node and return its slot to the free list. This gives the flexibility of a linked structure with the static allocation of an array.



A linked list stored in an array: a data array and a pointer array

Worked example. A circular queue is held in an array of size 5 (indices 0 to 4) with `Front = 3`, `Rear = 3` and one item stored. Two items are added, then two are removed. Where are the pointers, and why use a *circular* queue at all? Every move uses `(pointer + 1) MOD size`, so the pointers **wrap**. Adding twice moves `Rear`: $3 \to 4$, then $4 \to 0$ (because $(4+1) \bmod 5 = 0$), so `Rear = 0` and three items are stored. Removing twice moves `Front` the same way: $3 \to 4$, then $4 \to 0$, leaving `Front = 0` and one item. The wrap is the whole point: in a **linear** array queue the pointers march to the end and the freed space at the front is wasted even when the queue is empty. Remember a queue removes at the **Front** and adds at the **Rear** - a stack uses one pointer for both.

Exam tips

- Choose the right **data structure** and justify it (a record for mixed fields, a 2-D array for a grid).
- Know how to implement a **stack, queue and linked list** with an array and pointers (top; front/rear; next).

- Distinguish an **ADT** (its behaviour) from its implementation (array plus pointers).

Programming

A-Level Computer Science

Programming basics

```
1 usage
public Hotel (Integer capacity) { this.capacity = capacity; }
1 usage
public synchronized boolean checkIn(AccommodationRequest request) {
    if (guests.size() < capacity) {
        guests.add(request);
        return true;
    }
    else {
        return false;
    }
}
```

Programming turns a design into instructions written as code

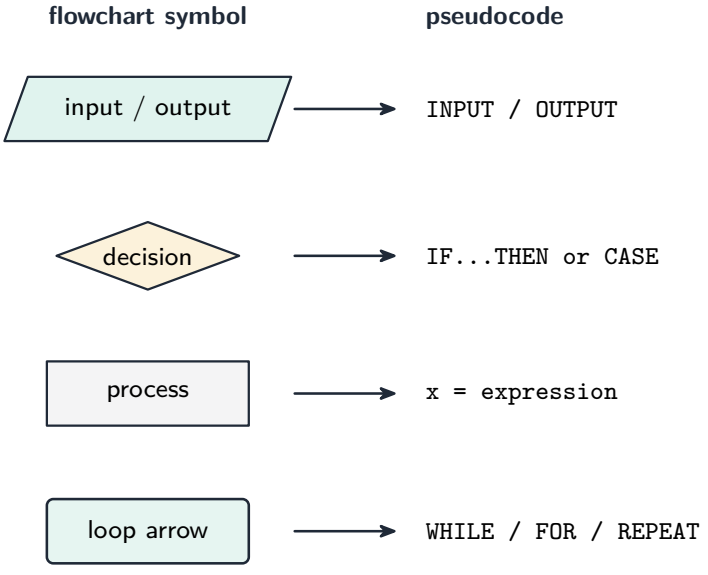


A programmer writes the code and tests it as they go

From design to code

You should be able to turn a design — a **flowchart** 流程图 (**program flowchart**) or **structured English** 结构化英语 — into **pseudocode** 伪代码, and then into a real language:

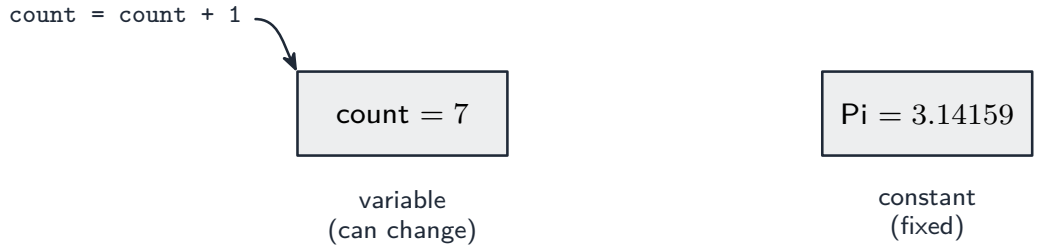
1. find the **variables** 变量 and their **data types** 数据类型.
2. turn input/output boxes into `INPUT` / `OUTPUT`.
3. turn decision diamonds into `IF...ELSE...ENDIF` (or `CASE`).
4. turn loop arrows into `WHILE`, `REPEAT...UNTIL`, or `FOR`.
5. turn process boxes into assignments or calculations.
6. check by tracing a small input.



Each flowchart symbol becomes a pseudocode keyword

Constants and variables

A **constant** 常量 holds a value that **never changes**; a variable holds one that may change. Declare them with a type:



A variable's value can change; a constant stays fixed

```
CONSTANT Pi ← 3.14159
DECLARE Radius : REAL
DECLARE Area : REAL

Radius ← 5
Area ← Pi * Radius * Radius
```

Use constants for fixed values that recur (`Pi`, `MaxScore`); they make code clearer and easy to change in one place.

Assignment and expressions

Use ← for **assignment** 赋值:

```
Total ← Total + 1
Average ← Sum / Count
```

Expressions use **operators** 运算符:

- arithmetic `+ - * /`, plus `DIV` (integer division) and `MOD` (remainder): `7 DIV 2 = 3`; `7 MOD 2 = 1`.
- comparisons `=`, `<>`, `<`, `>`, `<=`, `>=`.
- logic `AND`, `OR`, `NOT`.

Precedence 优先级 (highest to lowest): `NOT` → `* / DIV MOD` → `+ -` → comparisons → `AND` → `OR`. Use brackets when unsure.

Input and output

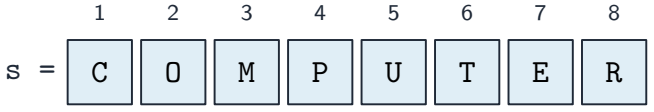
```
OUTPUT "Enter your name:"
INPUT Name
OUTPUT "Hello, ", Name
```

Built-in functions and library routines

Many tasks have ready-made **library routines** 库例程, so you need not write them:

- string: `LENGTH(s)`, `LEFT(s, n)`, `RIGHT(s, n)`, `MID(s, start, len)`, `UCASE(s)`, `LCASE(s)`.
- numeric: `INT(x)`, `ROUND(x)`, `ABS(x)`, `MOD(a, b)`, `RANDOM()`.
- conversion: `STR(x)` (number → string), `VAL(s)` (string → number).

Use the exact names from the question paper's reference list.



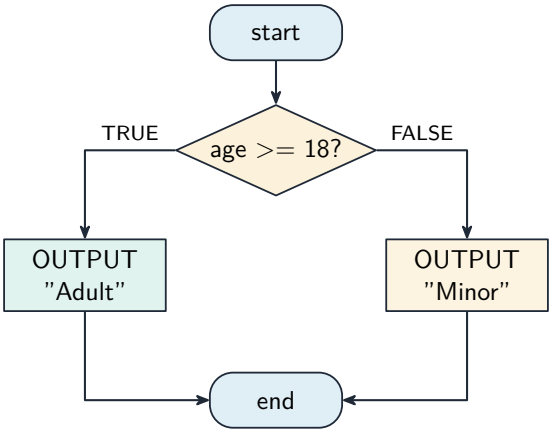

```
LENGTH(s) = 8              RIGHT(s, 2) = "ER"
LEFT(s, 3) = "COM"         UCASE(s) = "COMPUTER"
MID(s, 4, 3) = "PUT"       LCASE(s) = "computer"
```

The common string routines acting on `s = "COMPUTER"` (positions 1–8)

Selection

Selection 选择 chooses which steps run.

```
IF age >= 18 THEN
    OUTPUT "Adult"
ELSE
    OUTPUT "Minor"
ENDIF
```

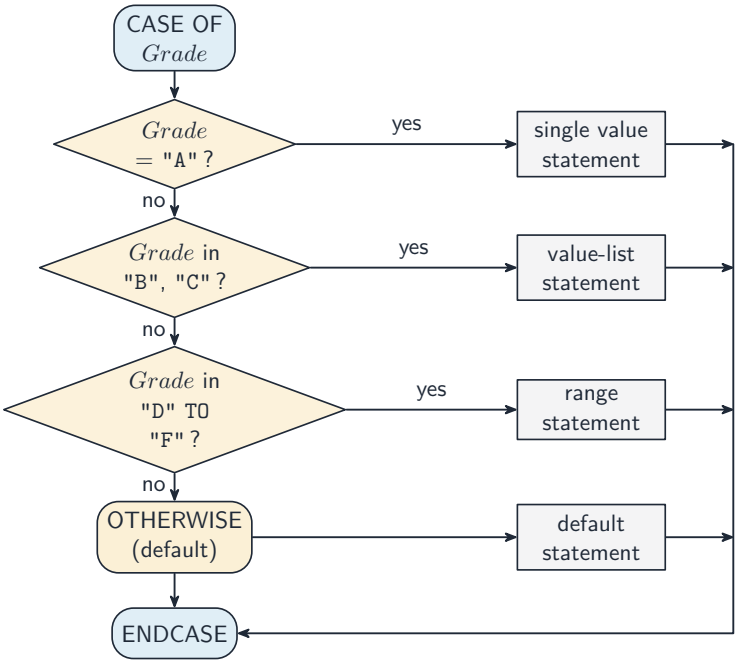


An IF…ELSE tests the condition once, then runs exactly one branch

For more than two cases you can use a **nested** 嵌套 IF, but deep nesting is hard to read — a `CASE` is cleaner when testing one value against several options:

```
CASE OF Grade
    "A": OUTPUT "Excellent"
    "B": OUTPUT "Good"
    OTHERWISE: OUTPUT "Try again"
ENDCASE
```

Cambridge `CASE` allows single values, value lists (`1, 2, 3:`), and ranges (`1 TO 5:`).



A CASE statement runs the branch that matches the value

Iteration

Iteration 迭代 repeats a block. Three loops differ in how many times the body runs.

Count-controlled (FOR) loop

A **count-controlled loop** 计数循环 — use it when you know how many times to repeat:

```
FOR i ← 1 TO 10
    OUTPUT i
NEXT i
```

A `STEP` can change the count (e.g. `FOR i ← 10 TO 1 STEP -1`). Best for a fixed number of repeats or processing each element of an **array** 数组.

Pre-condition (WHILE) loop

A **pre-condition loop** 前测循环 tests the condition **before** each pass, so it may run **zero** times:

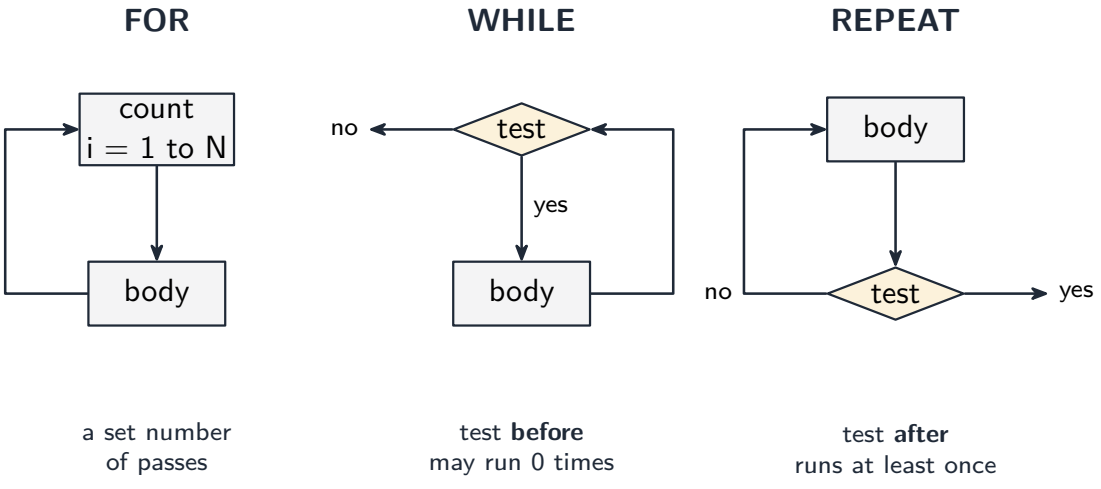
```
WHILE total < 100 DO
    INPUT n
    total ← total + n
ENDWHILE
```

Post-condition (REPEAT...UNTIL) loop

A **post-condition loop** 后测循环 tests the condition **after** each pass, so it always runs **at least once**:

```
REPEAT
    INPUT password
UNTIL password = correctPassword
```

Choosing the right loop



The three loops differ in where the condition is tested — before the body (WHILE), after it (REPEAT), or a set number of times (FOR)

- count known up front → **FOR**.
- may need zero passes → **WHILE**.
- always at least one pass → **REPEAT...UNTIL**.

Justify your choice by whether the count is known and whether the body must run at least once. A typical question gives a scenario ("ask for a password until correct, but always ask at least once") and asks which loop fits.

Procedures and functions

Structured programming 结构化编程 builds a program from small named **subroutines** 子程序, each with one job.

Procedure

A **procedure** 过程 is a named block that **does an action**; it may take **parameters** 参数 but does **not return a value**.

```
PROCEDURE Greet(name : STRING)
    OUTPUT "Hello, ", name
ENDPROCEDURE

CALL Greet("Ada")
```

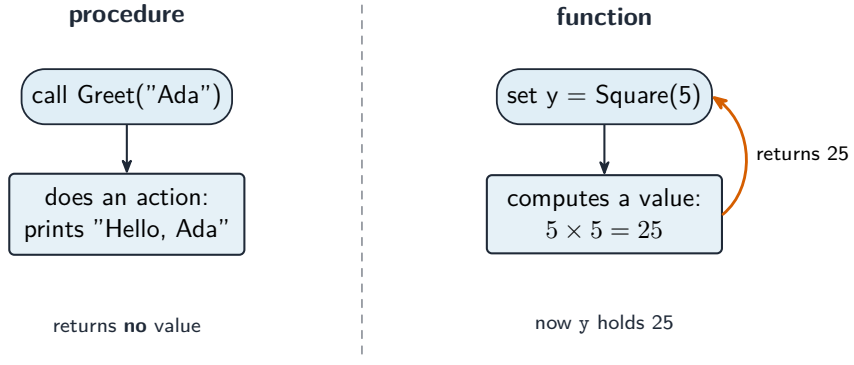
Function

A **function** 函数 is like a procedure but it **returns a value** that becomes part of an expression.

```
FUNCTION Square(x : INTEGER) RETURNS INTEGER
    RETURN x * x
ENDFUNCTION

result ← Square(5) + 1     // result = 26
```

Use a **procedure** when the subroutine performs an action; use a **function** when it computes a value for the caller.

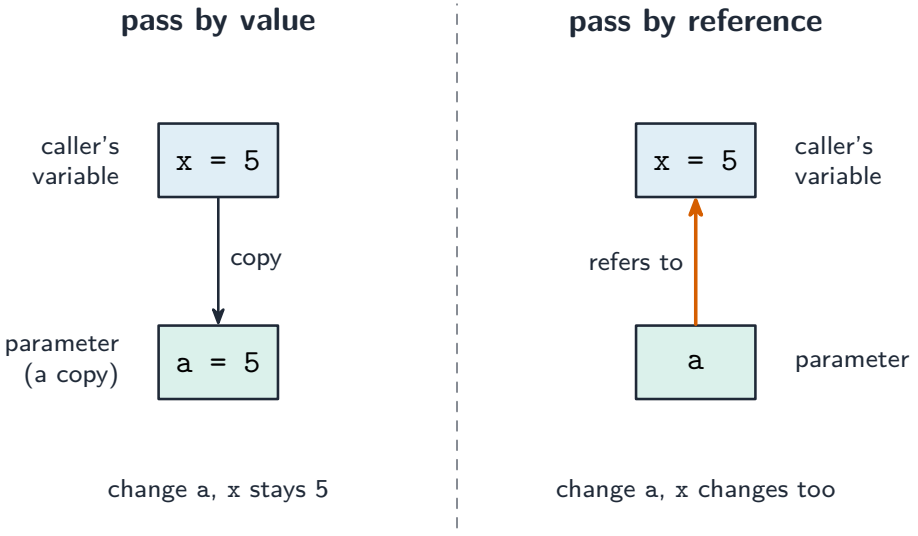


A procedure does an action and returns nothing; a function returns a value you use in an expression

Parameters

A **parameter** is a variable a subroutine declares to receive input; the values the caller supplies are **arguments** 实参. Two ways to pass them:

- **pass by value** 传值 — the routine gets a **copy**; changes inside it do not affect the caller. Use for inputs it only reads.
- **pass by reference** 传引用 — the routine gets a **reference** to the caller's variable; changes **do** affect the caller. Use when it must update a parameter.

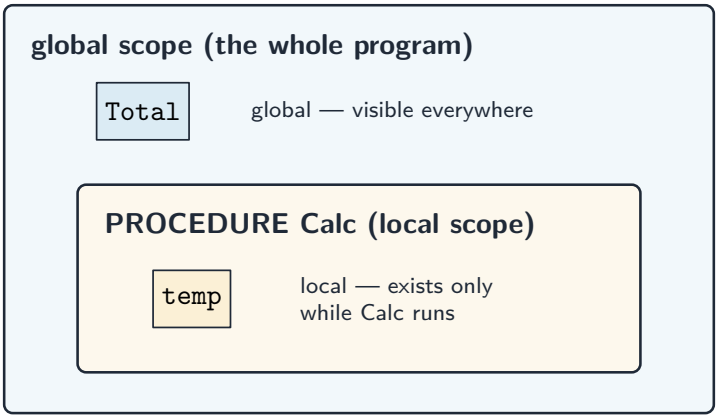


Pass by value copies the value into a new box; pass by reference lets the routine change the caller's own variable

```
PROCEDURE Swap(BYREF a : INTEGER, BYREF b : INTEGER)
    DECLARE temp : INTEGER
    temp ← a
    a ← b
    b ← temp
ENDPROCEDURE
```

Local vs global variables

A **local variable** 局部变量 is declared inside a subroutine and exists only while it runs. A **global variable** 全局变量 is declared outside and is visible everywhere. Prefer locals and parameters — heavy use of globals makes code hard to follow and test. (The region where a name is visible is its **scope** 作用域.)



A global variable is visible everywhere; a local variable exists only inside its own procedure

When to use a subroutine

Use a subroutine when:

- the same logic appears in **more than one place** — write it once, call it many times.
- a block has a **clear named purpose** — the name documents what it does.
- the program is **complex** — break it into parts (**decomposition** 分解).
- you want to **test a piece in isolation**.

Don't make them so tiny that the call costs more than the work inside.

Terminology

- **definition** — the `PROCEDURE ... ENDPROCEDURE` (or function) block.
- **call** — where it is invoked. **argument** — a value passed in. **parameter** — the variable that receives it.
- **return value** — what a function passes back.
- **procedure/function header** — the first line giving the name and parameters (`PROCEDURE Name(params)` or `FUNCTION Name(params) RETURNS type`).

- **procedure/function interface** / **signature** 签名 — name + parameters + return type: what a caller must know to use it.

Worked example. Which loop suits each task? (a) print the 12 times table; (b) keep reading numbers until the user enters 0; (c) ask for a password until it is correct. Choose by asking **how many times** the body runs and **when the test happens**. (a) The count is known in advance (12), so use a **FOR** loop. (b) The count is unknown, and the very first input might already be 0 - so the test must come **before** the body: a **WHILE** loop, which runs **zero** or more times. (c) The count is unknown, but you must always ask **at least once** before there is anything to test - so the test comes **after** the body: a **REPEAT...UNTIL**, which runs **one** or more times. The deciding question is whether the body must run at least once: WHILE may run zero times, REPEAT always runs once.

Writing efficient pseudocode

- **move invariants out of loops** — if a value (an **invariant** 不变量) does not change with the loop counter, compute it once before the loop.
- **exit a loop early** when the answer is found (stop a **linear search** 线性查找 as soon as the target appears).
- **avoid redundant work** — store a result and reuse it instead of recomputing.
- **choose the right data structure** — an array beats many separate variables when the items belong together.
- **replace deep nested IFs with CASE** when testing one value against many.
- **comment the intent**, not the mechanics (`// validate the postcode`, not `// loop 6 times`).
- **use meaningful names** (`numberOfPupils`, not `n`) and **initialise variables** before use.

slower

```
FOR i = 1 TO n
 x = slow() // every pass
 use(x, i)
NEXT i
```

faster

```
x = slow() // once
FOR i = 1 TO n
 use(x, i)
NEXT i
```

Move unchanging work out of the loop so it runs once

Exam tips

- Distinguish a **procedure** (no return value) from a **function** (returns a value); know **pass by value vs by reference**.
- Choose the right loop: **count-controlled (FOR)** when the number of repeats is known, **condition-controlled (WHILE/REPEAT)** otherwise.
- Distinguish **local vs global** variables and scope; prefer local variables in reusable modules.

Software Development

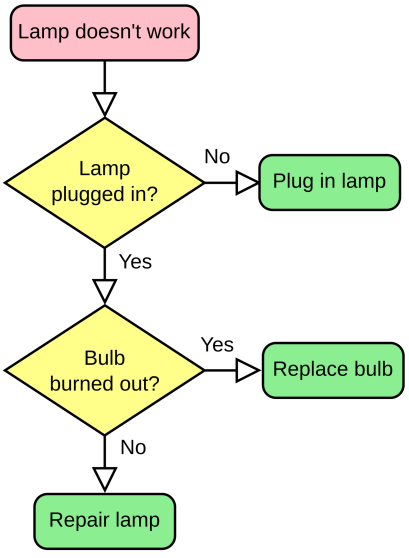
A-Level Computer Science

Program development life cycle

A **development life cycle** 开发生命周期 is the set of stages from idea to finished, maintained software. It exists to **plan, manage and control** a project — to build the right product, on time, with good quality.



Software is built by teams who follow a development life cycle to stay coordinated



A flowchart plans a program's logic during the design stage of the cycle

Why a life cycle is needed

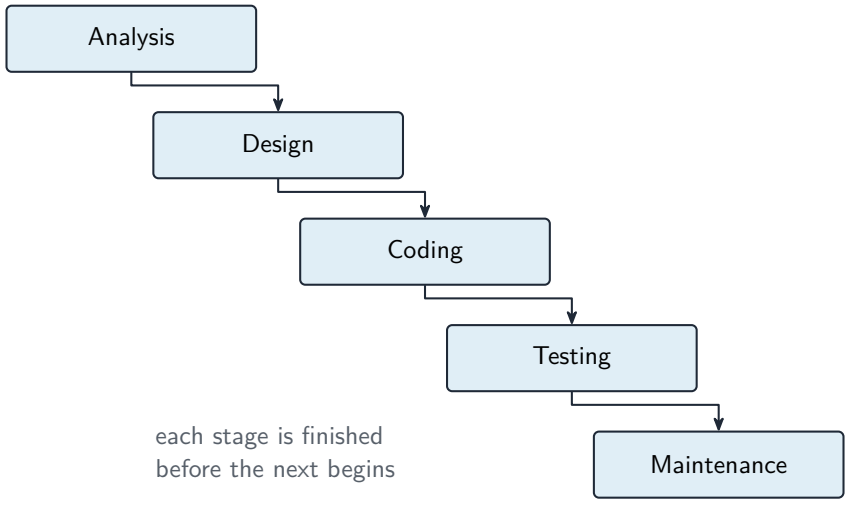
It manages complexity (break a big program into phases), coordinates teams, tracks progress with milestones, builds in testing, records design decisions for later, and manages risk.

Why there are different ones

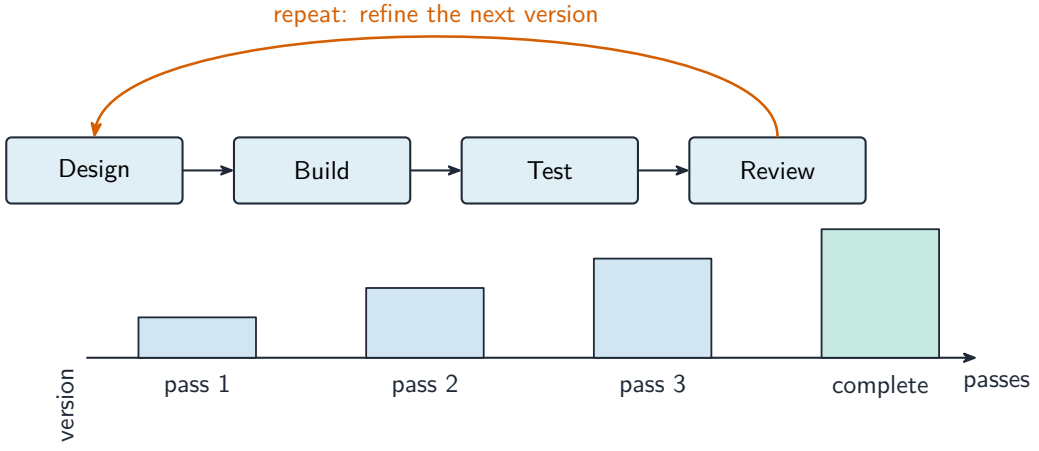
No single life cycle fits every project, so several **development life cycles** exist. The choice depends on the size and complexity, how clear the **requirements** 需求 are at the start, how much change is expected, the risk level, the team, and the deadline.

Common models

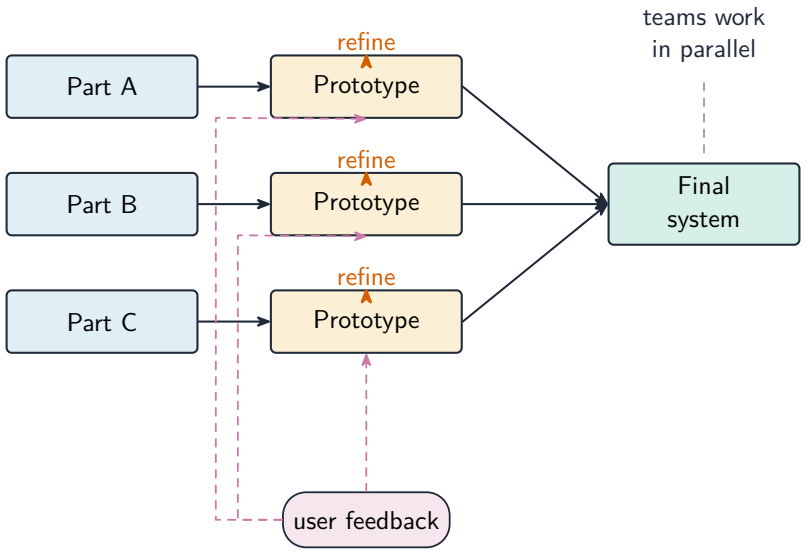
- **Waterfall** 瀑布模型 — a linear sequence (Analysis → Design → Coding → Testing → Maintenance), each stage finished before the next. Clear and well-documented; good for **stable requirements**, but poor at coping with mid-project change, and the customer sees nothing working until the end.
- **Iterative model** 迭代模型 — repeated passes, each producing a partial version that is reviewed and refined. Catches problems earlier; good when requirements are **discovered over time**, but harder to estimate.
- **Rapid Application Development** 快速应用开发 (RAD) — heavy use of a **prototype** 原型 and user feedback. Very fast first delivery; good for **changing requirements**, but depends on user availability and suits smaller systems.
- **Agile** 敏捷 — short iterations ("sprints"), constant collaboration and testing. Flexible and adaptive, but needs a committed customer and a skilled team.



The waterfall model: each stage is finished before the next begins



The iterative model: repeated passes refine the program



Rapid application development: teams work on parts in parallel

The standard stages

- **analysis** — find **what** the program must do; gather and document requirements.
- **design** — decide **how**: data structures, algorithms, modules, interface, file layouts.
- **coding (implementation** 实现 **)** — write the source code following the design.
- **testing** — run against test data and fix bugs.
- **maintenance** 维护 — after release, keep it working and useful.

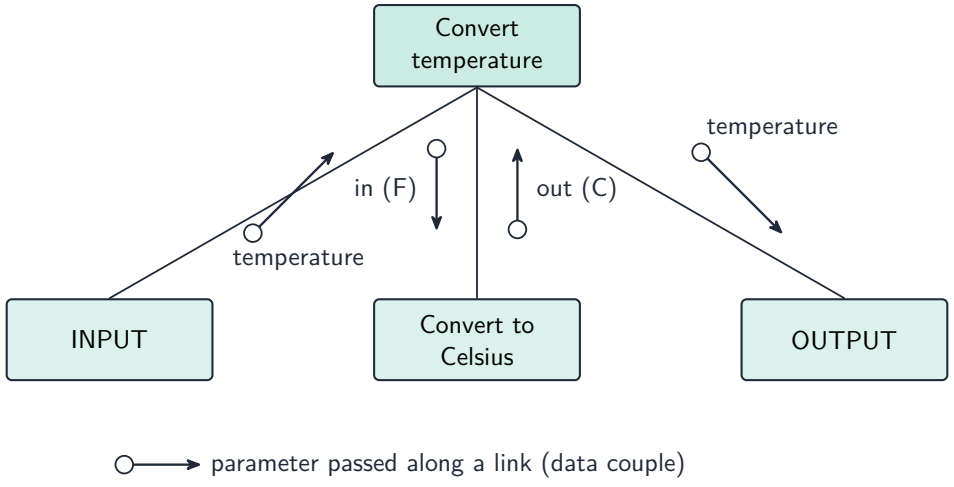
Program design tools

Structure chart

A **structure chart** 结构图 shows the **hierarchical decomposition** 分解 of a program into modules (**subroutines** 子程序) and the **parameters** 参数 passed between them. Each module is a rectangle; lines link caller (above) to callee (below); small arrows show data going down and results coming back up. The design can then be turned into equivalent **pseudocode** 伪代码.

```
              CalculatePay
          /        |         \
GetEmployee  CalculateBonus  CalculateTax
Returns:     Takes: sales    Takes: gross
employeeID   Returns: bonus  Returns: tax
```

It is a design-stage tool, and you can read the procedure signatures off it.



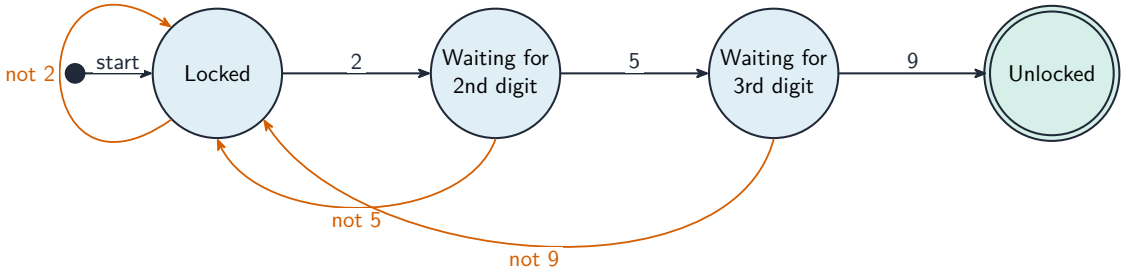
A structure chart: modules with the parameters passed between them

State-transition diagram

A **state-transition diagram** 状态转换图 shows the **states** 状态 a system can be in and the **events** that move it between them — good for vending machines, traffic lights, user interfaces. **State-transition diagrams** are used to document the behaviour of an algorithm or system. Each state is a circle; each transition is an arrow labelled with the event.

```
   coin inserted                item selected
[Idle] ---------------→ [Awaiting selection] -----------→ [Dispensing]
```

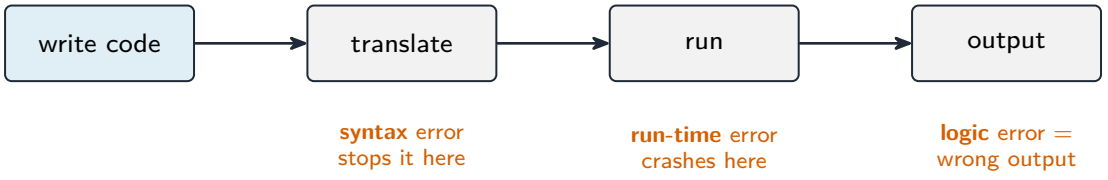
It makes missing transitions easy to spot ("what if a second coin is inserted while awaiting selection?").



A state-transition diagram for a door lock with code 259

Errors

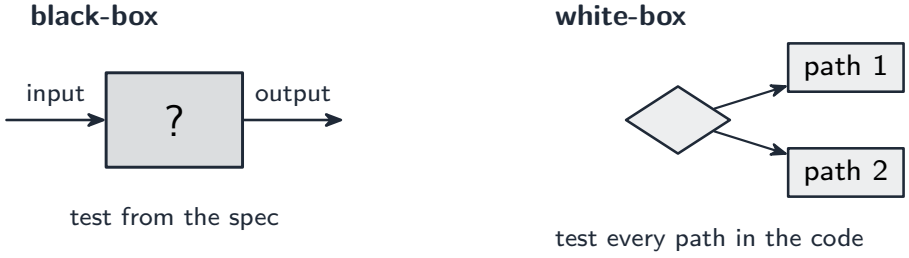
- **syntax error** 语法错误 — breaks the language's grammar (missing bracket, misspelled keyword). Caught at translation time; the program won't run until fixed.
- **run-time error** 运行时错误 — happens while running (divide by zero, file not found, array index out of range). The program crashes or raises an exception; fix by adding checks.
- **logic error** 逻辑错误 — the program runs but gives **wrong results** (using `+` for `-`, an off-by-one loop, conditions in the wrong order). The hardest to find; the only sign is wrong output, so use careful testing and tracing.



When each error shows up: syntax at translation, run-time during the run, logic in the output

Testing methods

- **dry run** 手工跟踪 — trace the code on paper, writing each variable's value in a table.
- **walkthrough** 走查 — a team review of the code.
- **white-box testing** 白盒测试 — designed from the code's internal structure, covering every statement, branch and loop.
- **black-box testing** 黑盒测试 — designed from the specification only: feed inputs, check outputs.
- **integration testing** 集成测试 — combine modules and test the interfaces between them.
- **alpha testing** α 测试 — by the developers/in-house before release; **beta testing** β 测试 — by a limited group of real users in their own environment.
- **acceptance testing** 验收测试 — by the customer, to decide if the product is fit for purpose.
- **stub** 桩 — a placeholder for a module that does not exist yet, so the structure can be tested top-down.



Black-box tests the specification; white-box tests the code paths

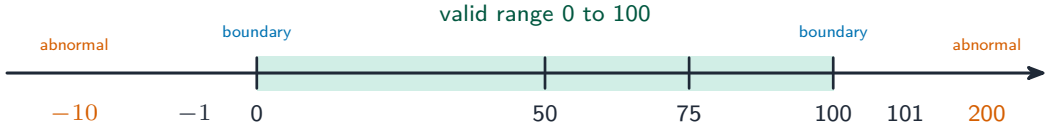
Test strategy and test plan

A **test strategy** 测试策略 is the **high-level approach** — which kinds of testing, who does them, when, and the criteria to move on. A **test plan** 测试计划 is the **detailed list of tests** — each with input data, expected output, and a column for the actual output.

Choosing test data

For each field or condition, include three kinds:

- **normal data** 正常数据 — typical values inside the valid range (for marks 0–100: `50`, `75`).
- **abnormal data** 异常数据 — values that should be rejected (`-10`, `200`, `"abc"`).
- **extreme data** 极端数据 — the largest and smallest values still **accepted** (`0` and `100`).
- **boundary data** 边界数据 — values at the edges, where off-by-one errors hide (each accepted extreme and the rejected value just outside it: `0`/`-1`, `100`/`101`).



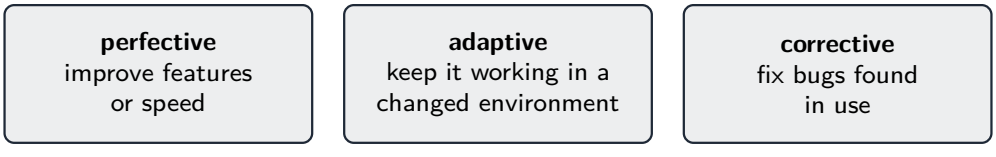
normal (50, 75) inside; extremes (0, 100) the accepted edges; abnormal (−1, 101, −10, 200) rejected

Test data for a 0–100 field: normal inside, extremes at the boundaries, abnormal outside

Worked example. A field accepts an exam mark from 0 to 100. Give test data of each kind with its expected result. **Normal**: `50` - accepted, a typical value inside the range. **Abnormal**: `-10`, `200`, `"abc"` - all rejected, being out of range or the wrong data type. **Extreme**: `0` and `100` - the largest and smallest values that are still **accepted**. **Boundary**: the pairs straddling each edge - `-1` rejected alongside `0` accepted, and `100` accepted alongside `101` rejected. Every value must carry its **expected result**, or the test plan proves nothing. Extreme and boundary are the pair most often confused: an extreme value sits **inside** and is accepted, while a boundary test is always a **pair** either side of the edge - which is exactly where off-by-one errors hide.

Maintenance

Most of a program's lifetime cost is in maintenance. Three kinds:



Three kinds of maintenance: perfective, adaptive and corrective

- **perfective maintenance** 完善性维护 — improving performance or features even though it works (a faster query, a new option).
- **adaptive maintenance** 适应性维护 — keeping it working in a changing environment (a new OS, a new API, a legal change).
- **corrective maintenance** 纠正性维护 — fixing bugs found in use.

A program may need all three throughout its life.

Amending an existing program

When asked to add a feature or fix a bug:

1. **read the existing code** until you understand the algorithm and data flow.
2. **find where the change goes** — which subroutine, which lines.
3. **make the change as small as possible** — don't rewrite working code.
4. **update related parts** — every caller of a changed parameter list, every routine using a changed data structure.
5. **test** the new behaviour **and** the old (**regression testing** 回归测试 — check you broke nothing).
6. **document** the change.

Clear comments, meaningful names, decomposed subroutines and a structure chart make a program much easier to amend — which is why the design tools matter even after the first release.

Exam tips

- Compare development models (waterfall, iterative, RAD) and know the stages of the **program development life cycle**.
- Distinguish **syntax, logic and run-time** errors and how each is found.
- Choose **test data** of three kinds — normal, boundary and erroneous — and give an example of each for the stated range.
- Distinguish the types of **maintenance** (corrective, adaptive, perfective).

Basic economic ideas and resource allocation

A-Level Economics

The fundamental economic problem

Economics starts with one big idea: **scarcity** 稀缺性. We never have enough to give everyone everything they want.

- **resources** 资源 are the things we use to make goods and services. They are **finite** 有限的 — there is only a limited amount, so they are **scarce** 稀缺.
- **goods** 物品 (things you can touch, like food) and **services** 服务 (work done for you, like a haircut) are what we make from resources.
- human **wants** 欲望 are unlimited. People always want more.

Because wants are bigger than resources, this is called the **economic problem** 经济问题. Every society must answer three questions:

1. **What** to produce?
2. **How** to produce it?
3. **For whom** to produce it?

Choice and opportunity cost

Because resources are scarce, you must choose. Every choice means you give up something else.

opportunity cost 机会成本 — the next best **alternative** 替代选择 you give up when you make a decision.

Opportunity cost is faced by everyone:

- **consumers** 消费者 (people who buy and use goods): buy a phone, and you give up the holiday that money could have paid for.
- **producers** 生产者 and **firms** 企业 (the businesses that make goods): use a factory to make cars, and you give up the trucks it could have made.
- **governments** 政府: spend tax money on hospitals, and you give up the schools that money could have built.

Thinking like an economist

Economics as a social science

Economics is a **social science** 社会科学. It studies how people behave when they make choices. Like other sciences, it builds **models** 模型 (simple pictures of the real world) and tests them with data.

The real world is complex, so economists change one thing at a time and keep everything else fixed. This is the idea of **ceteris paribus** 其他条件不变 — a Latin phrase meaning "all other things stay the same". For example: "if the price of coffee rises, people buy less coffee, *ceteris paribus*".

Positive and normative statements

- a **positive statement** 实证表述 is about fact. It can be tested as true or false. Example: "a higher price of petrol reduces the amount people buy."
- a **normative statement** 规范表述 is about opinion. It says what *should* happen, and it cannot be tested as true or false. Example: "the government should make petrol cheaper."

Normative statements rest on a **value judgement** 价值判断 — a personal view about what is good or fair. Two people can agree on the facts and still disagree about what *should* be done.



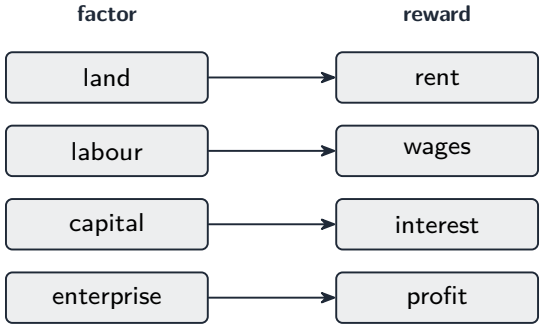
A positive statement can be tested; a normative statement is a value judgement

Microeconomics and macroeconomics

- **microeconomics** 微观经济学 studies single parts of the economy: one market, one firm, one consumer.
- **macroeconomics** 宏观经济学 studies the whole economy together: total **output** 产出, jobs, and rising prices.

Factors of production

The **factors of production** 生产要素 are the four kinds of resource used to make goods and services. Each one earns a **reward** 报酬.



The four factors of production and the reward each earns

Factor	What it is	Its reward
land 土地	natural resources (fields, water, oil, minerals)	**rent** 地租
labour 劳动	human work, both physical and mental	**wages** 工资
capital 资本	man-made things used to produce, like machines, tools and factories	**interest** 利息
enterprise 企业家才能	the skill of bringing the other three together and taking the risk	**profit** 利润

Land includes some **non-renewable resources** 不可再生资源 (like oil and coal) that can run out. Enterprise is supplied by the **entrepreneur** 企业家, the person who has the business idea and risks their own money.



An open-pit copper mine: land includes the natural resources (minerals, oil, water) that firms extract.

Specialisation and division of labour

specialisation 专业化 is when a worker, firm or country makes only the things it is best at, then trades for the rest.

Inside a workplace, specialisation leads to the **division of labour** 劳动分工: a big job is split into small tasks, and each **worker** 工人 does one task again and again.

Advantages	Disadvantages
workers get faster and better at one task	doing one task is boring, so workers may quit
less time wasted moving between jobs	if one worker is absent, the whole line can stop
output per worker rises, so costs fall	workers learn only one narrow skill



A car assembly line: capital (machines and robots) plus the division of labour, each worker repeating one task.

Mobility of factors

mobility 流动性 means how easily a factor can move to a new use.

- **occupational mobility** 职业流动性 — how easily a factor moves from one *job or use* to another. A teacher who retrains as a nurse is occupationally mobile.
- **geographical mobility** 地理流动性 — how easily a factor moves from one *place* to another. Land cannot move at all, so it has no geographical mobility.

Labour is often not very mobile. Family ties, the cost of housing, and missing skills all make it hard for workers to move.

Economic systems

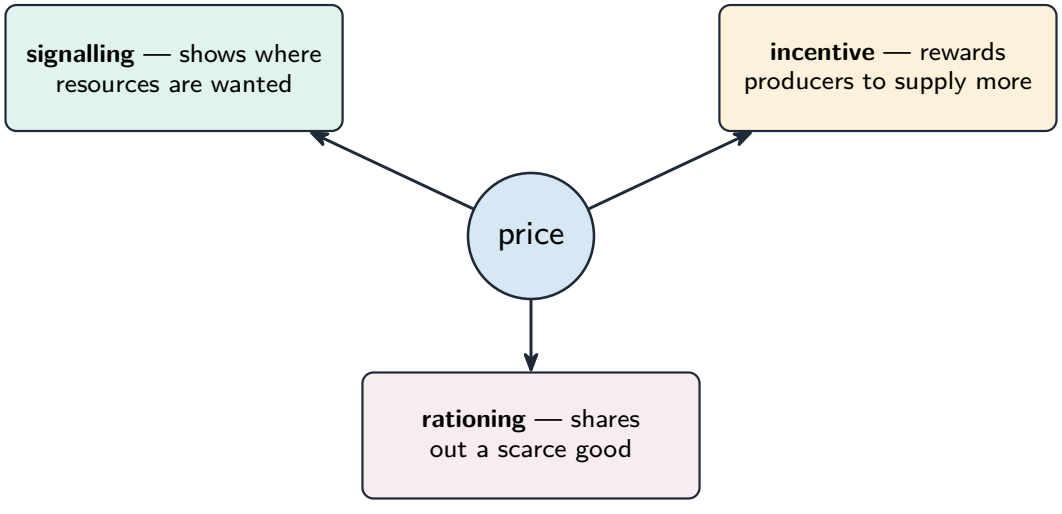
An economic system is the way a country decides how to **allocate** 配置 (share out) its scarce resources. There are three types.

- a **market economy** 市场经济 — firms and people own resources, and prices decide what is made. The government does little.
- a **planned economy** 计划经济 (also called a **command economy** 指令经济) — the government owns resources and decides what is made, how, and for whom.
- a **mixed economy** 混合经济 — a blend of the two. The **private sector** 私营部门 (firms and households) and the **public sector** 公共部门 (the government) both make decisions. Almost every real country is mixed.

The price mechanism

In a market economy, resources are guided by the **price mechanism** 价格机制. Prices do three jobs:

- **signalling** 信号传递 — a price tells buyers and sellers where resources are wanted. A rising price says "this is wanted — make more".
- **incentive** 激励 — a higher price rewards producers, so they want to supply more of that good.
- **rationing** 配给 — when a good is scarce, a higher price shares it out to those who are willing and able to pay.

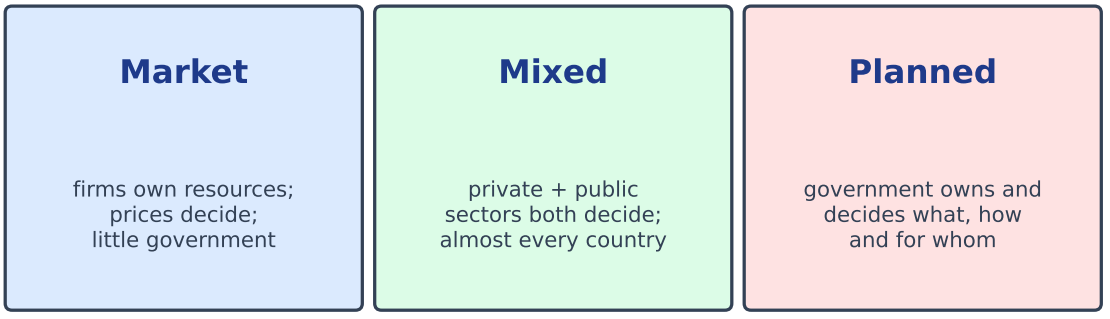


In a market the price does three jobs at once — it signals, it gives an incentive, and it rations the scarce good

Comparing the systems

System	Strengths	Weaknesses
market	choice, low prices, strong incentive to work and invent	large gap between rich and poor; some goods not supplied
planned	smaller gap between rich and poor; key goods provided for all	weak incentive; shortages and surpluses; slow to change
mixed	tries to keep the good points of both	the right balance is hard to find

In practice no country sits at a pure extreme. China's economy, for instance, is officially a *socialist market economy*, where markets set most prices while the government guides key sectors and owns major enterprises. Every real system mixes the two and sits somewhere on the spectrum below.



Exam: real economies are mixed — the question is how far market vs government decides

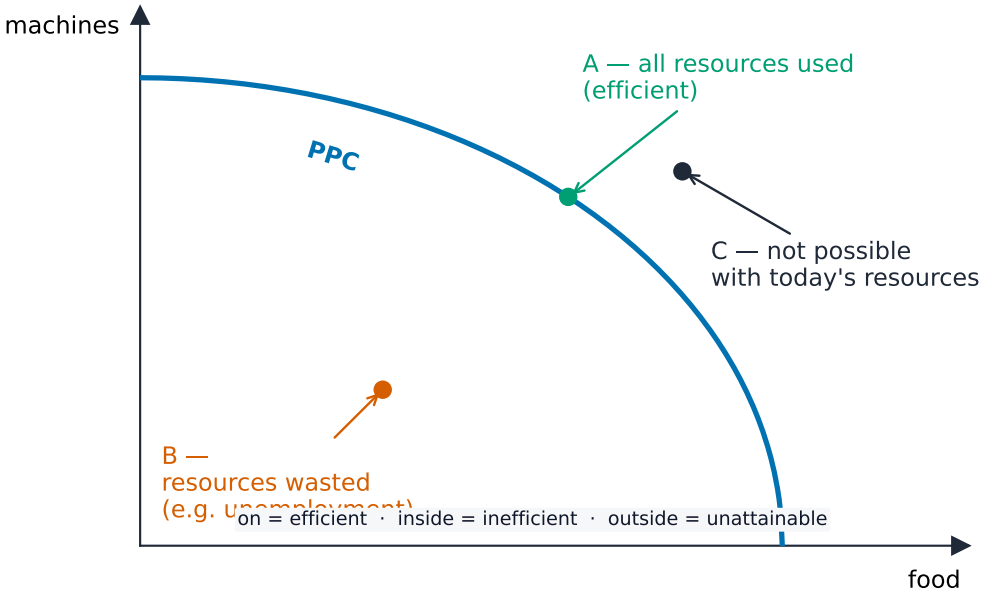
Economic systems form a spectrum: a pure market at one end, full government planning at the other, mixed economies in between.

Production possibility curves

A **production possibility curve** 生产可能性曲线 (PPC) is a line showing the largest combinations of two goods an economy can make when all its resources are fully and well used.

Imagine an economy that makes only food and machines. The PPC is drawn with food on one axis and machines on the other.

- a point **on** the curve: all resources are used with full **efficiency** 效率.
- a point **inside** the curve: resources are wasted (for example, **unemployment** 失业 of workers). The economy could do better.
- a point **outside** the curve: impossible today, because there are not enough resources. This shows scarcity.

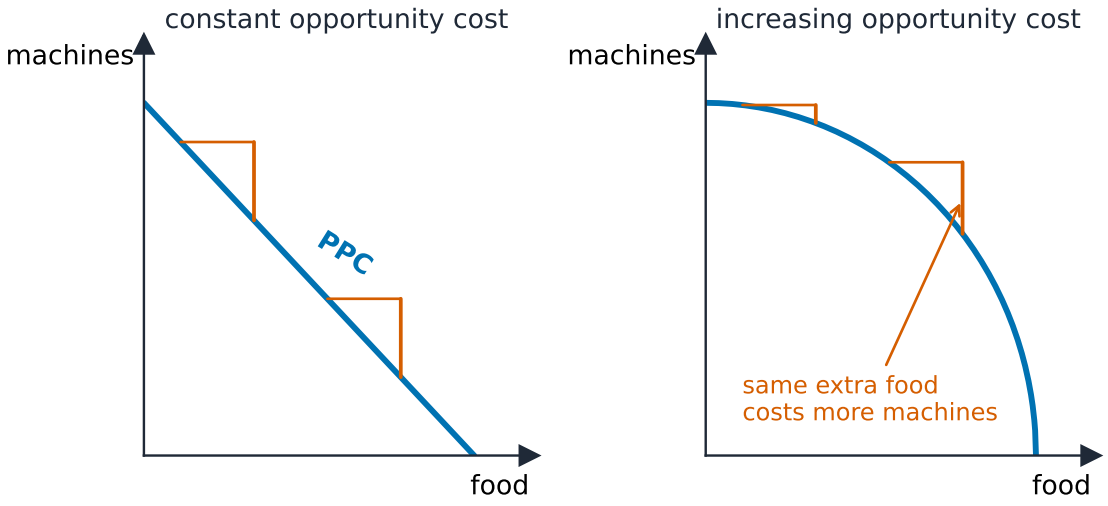


A point on the curve uses all resources (efficient); inside the curve means waste, e.g. unemployment; outside is not possible yet.

Moving along the curve shows opportunity cost: to make more food, you must move resources away from machines, so you make fewer machines. This is the **trade-off** 取舍 between the two goods.

The *shape* of the curve shows how opportunity cost behaves:

- a straight-line PPC means **constant** opportunity cost — each extra unit of food costs the same number of machines.
- a curve that bows outwards means **increasing** opportunity cost — each extra unit of food costs more and more machines, because resources are not equally good at making both goods.

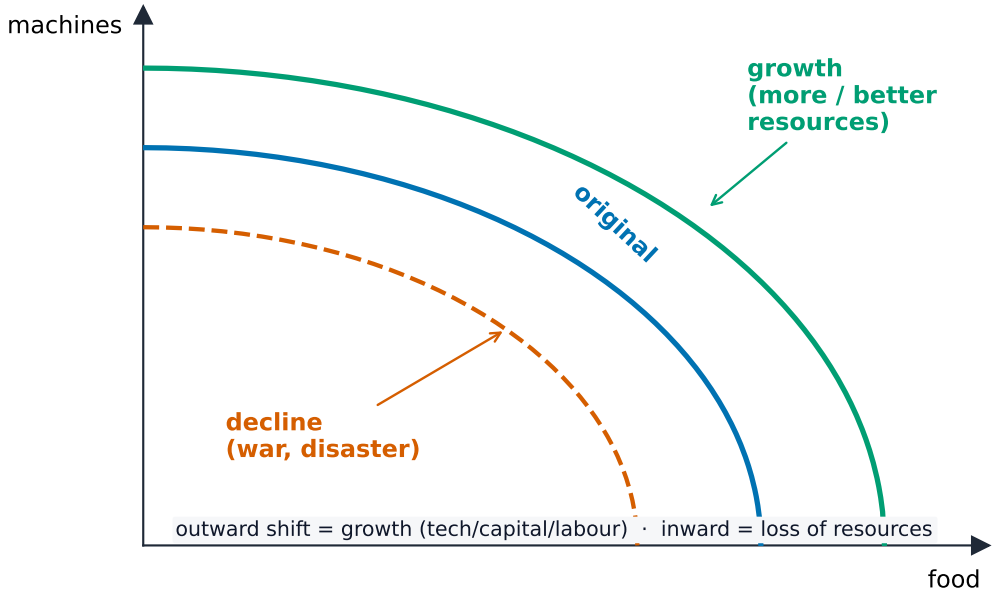


opportunity cost = goods forgone · bowed-out PPC ⇒ increasing opportunity cost

A straight PPC means each extra unit of food costs the same; a bowed PPC means each extra unit costs more and more machines.

Shifts and growth

When the PPC moves outward (to the right), the economy can make more of both goods. This is **economic growth** 经济增长. It happens when there are more resources, or better ones (new **technology** 技术, a bigger or more skilled workforce). War or disaster can move the PPC inward — this is **economic decline** 经济衰退.

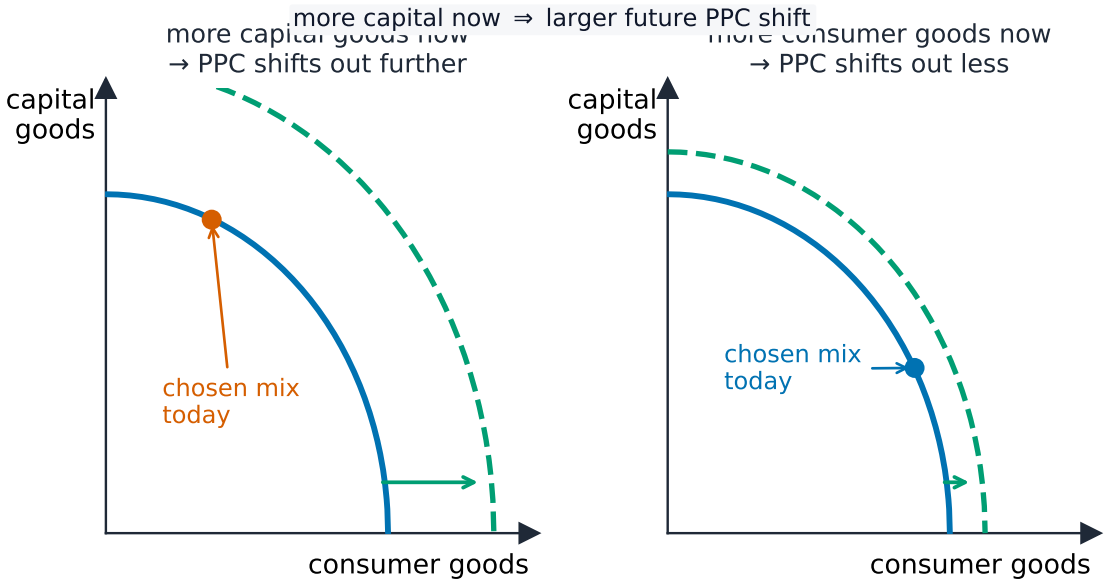


More or better resources shift the whole PPC outward (growth); war or disaster shifts it inward (decline).

A country must also choose between two kinds of good now:

- **capital goods** 资本品 — goods used to make other goods (machines, factories).
- **consumer goods** 消费品 — goods people use directly (food, clothes).

Making more capital goods today means fewer consumer goods today. But more capital goods help the economy grow faster, so the PPC shifts out by more in the future.



Choosing more capital goods today means fewer consumer goods now, but the future PPC shifts out further.

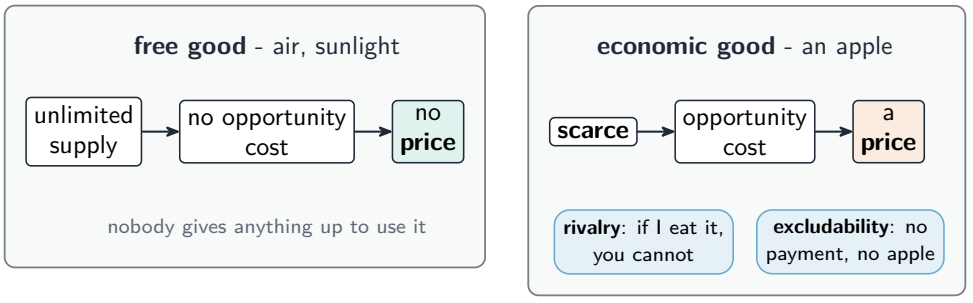
Worked example. Why is a production possibility curve usually drawn **concave** (bowed outwards) rather than as a straight line? A straight line would mean the opportunity cost of each extra unit stays **constant**. The curve bows out because factors of production are **not perfectly substitutable** - they are specialised. Starting from all-food production, the first factors moved into making cars are the ones **least suited** to farming, so very little food is given up for a lot of cars. Keep going and you must move factors that are **increasingly good at farming and increasingly bad at making cars**, so each extra car costs **more and more** food. That rising cost is **increasing opportunity cost**, and the bowed shape is simply its picture. A straight-line PPC is not wrong - it just assumes factors are perfectly substitutable, and naming that assumption is what earns the mark.

Types of goods and services

Free goods and economic goods

- a **free good** 免费物品 has no opportunity cost. Its supply is unlimited, so it has no price (for example, air or sunlight).
- an **economic good** 经济物品 (also called a **private good** 私人物品) is scarce. Making it has an opportunity cost, so it has a price.

A private good has two features. It has **rivalry** 竞争性 (if I eat the apple, you cannot) and **excludability** 排他性 (the seller can stop you using it unless you pay).



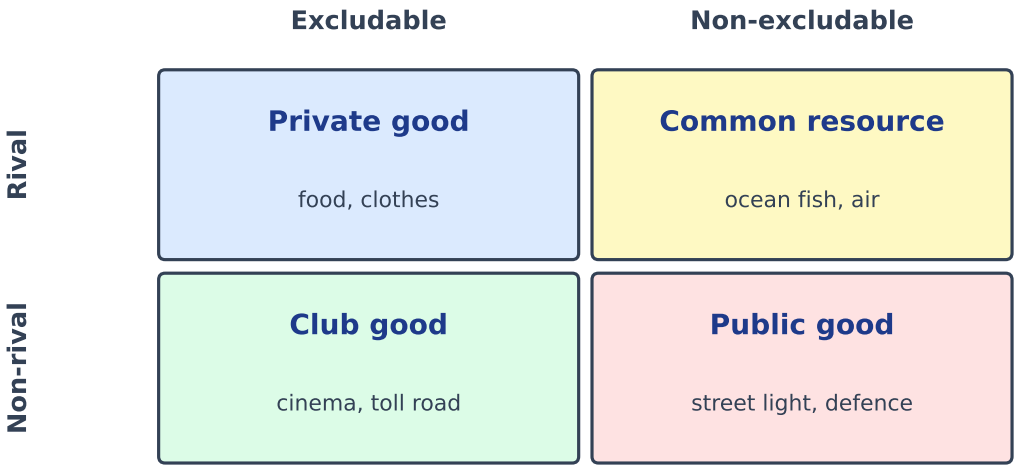
A free good has no opportunity cost and no price; an economic good is scarce, rival and excludable

Public goods

A **public good** 公共物品 is the opposite of a private good. It has two features:

- **non-rivalry** 非竞争性 — one person using it does not reduce the amount left for others.
- **non-excludability** 非排他性 — you cannot stop people who do not pay from using it.

Because of non-excludability, public goods suffer from the **free-rider problem** 搭便车问题: people enjoy the good without paying for it. So private firms cannot make a profit and will not supply it. The government must provide public goods instead — for example, street lighting and **national defence** 国防.



Exam: classify by rivalry × excludability — free-rider problem only for non-excludable goods

Rivalry and excludability give four types of good: private, club, common resource, and public.

Merit and demerit goods

- a **merit good** 优值品 is better for you (and for society) than you realise, so people buy too little of it. Examples: education and health care.
- a **demerit good** 劣值品 is worse for you than you realise, so people buy too much of it. Examples: cigarettes and alcohol.



Education is a classic merit good. Its full value — better jobs, and a more skilled society — is easy to underestimate, so left to themselves people buy less of it than is good for them or the country. That is why governments provide and subsidise it

People misjudge these goods because of **information failure** 信息失灵 — they do not have, or do not use, the full facts about the benefit or harm.

Deciding what counts as a merit or demerit good is partly a value judgement, so people may disagree about which goods belong on each list.

Exam tips

- Frame answers around **scarcity and choice**: every choice has an **opportunity cost** — the next best alternative given up.
- Read a **PPC**: a point inside means unused resources, on the curve is efficient, outside is unattainable; the slope shows opportunity cost.
- Distinguish **positive** (a testable fact) from **normative** ("should", a value judgement) statements.
- Classify goods: free versus economic, and **public / merit / demerit** goods (non-excludable, non-rival, and so on).

The price system and the microeconomy

A-Level Economics

Demand

demand 需求 is the quantity of a good that buyers are willing and able to buy at each **price** 价格 in a period of time. It is not just a wish — buyers must also be able to pay.

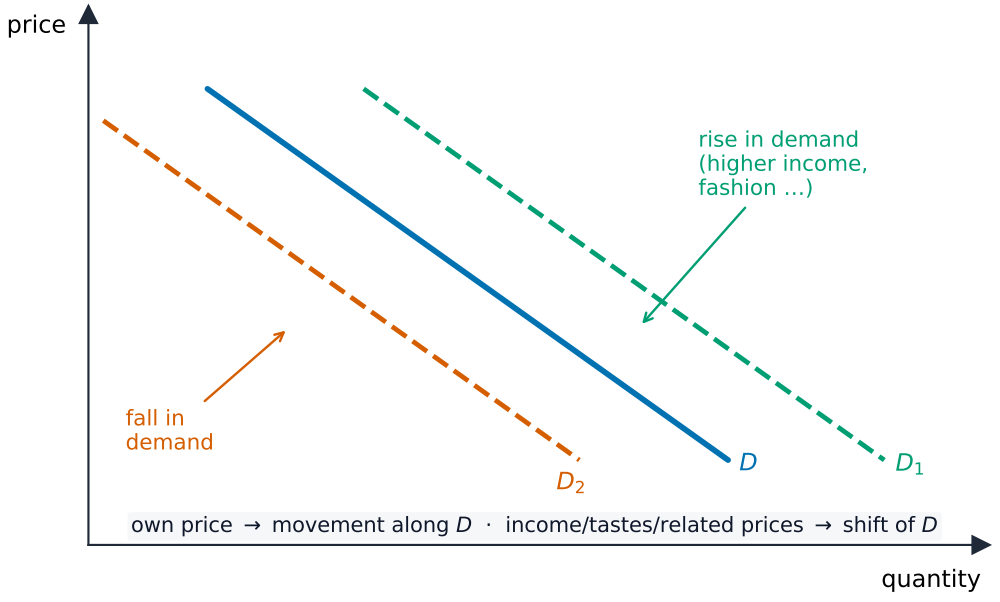
The **law of demand** says: when the price rises, the **quantity demanded** 需求量 falls; when the price falls, the quantity demanded rises. Price and quantity move in opposite directions.

So the **demand curve** 需求曲线 slopes downward, from top-left to bottom-right.

What shifts the demand curve

A change in the good's **own price** is shown as a *movement along* the demand curve. A change in any other cause shifts the whole curve to a new position — this is a **shift** 移动 of demand. The main causes (the **determinants** 决定因素) are:

- **income** 收入 — for most goods, higher income raises demand.
- the price of a **substitute** 替代品 — a good you can use instead (tea for coffee). If the substitute's price rises, demand for this good rises.
- the price of a **complement** 互补品 — a good used together with this one (cars and petrol). If the complement's price rises, demand for this good falls.
- tastes and fashion, the size of the population, and advertising.



A change in income, tastes or related prices shifts the whole curve (D_1 right, D_2 left); a change in the good's own price is a movement along it.

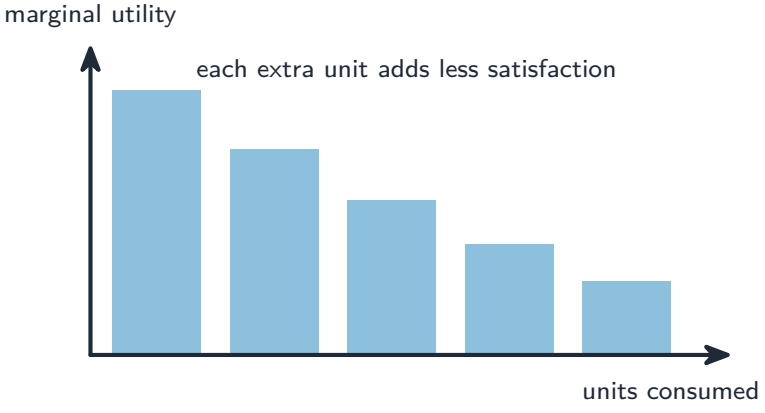
Individual and market demand

One **consumer** 消费者 has an individual demand. The **market demand** 市场需求 is found by adding up the quantity that all consumers want at each price (adding the curves sideways).

Marginal utility and demand

utility 效用 is the satisfaction you get from consuming a good. **marginal utility** 边际效用 is the extra utility from one more unit.

The law of **diminishing marginal utility** 边际效用递减 says each extra unit gives less extra satisfaction than the one before. Because later units are worth less to you, you will only buy them at a lower price. This is why the demand curve slopes downward.



Each extra unit adds less satisfaction than the last — so buyers will only pay less for later units, which is why demand slopes downward

Supply

supply 供给 is the quantity of a good that **producers** 生产者 are willing and able to sell at each price in a period of time.

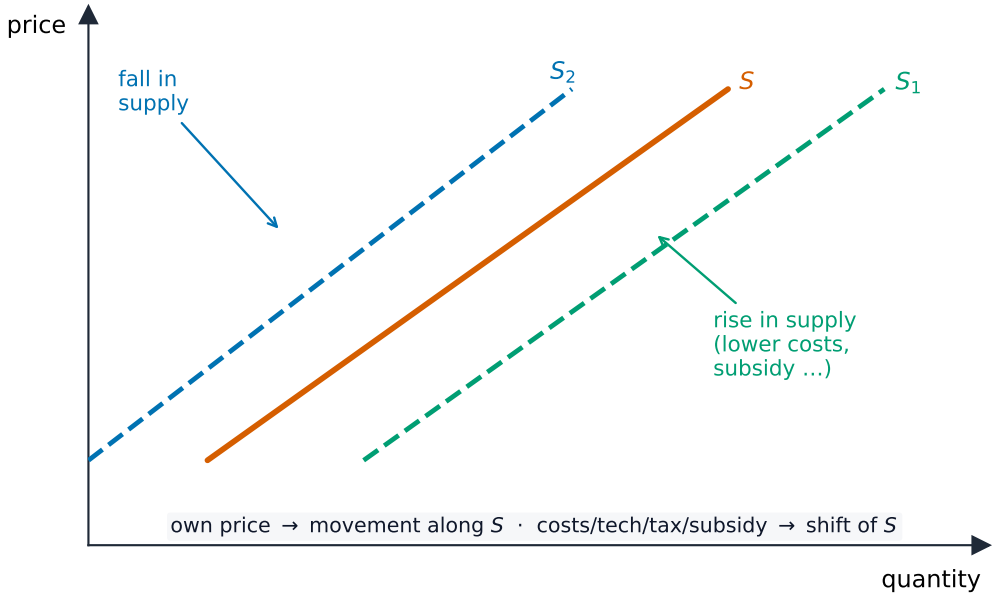
The **law of supply** says: when the price rises, the **quantity supplied** 供给量 rises. Higher prices mean more **profit** 利润, so firms make more. So the **supply curve** 供给曲线 slopes upward.

What shifts the supply curve

A change in the good's own price is a movement along the supply curve. Other causes shift the whole curve:

- **costs of production** 生产成本 — if wages, raw materials or rent get dearer, supply falls.
- technology — better machines raise supply.

- a **subsidy** 补贴 (government payment to producers) raises supply; an indirect tax lowers it.
- the number of **firms** 企业 in the market, and weather (for farm goods).



Lower costs, better technology or a subsidy shift supply right (S_1); higher costs or a tax shift it left (S_2).

Individual and market supply

The **market supply** 市场供给 is found by adding up the quantity all firms will sell at each price.

Price elasticity of demand

price elasticity of demand 需求价格弹性 (PED) measures how much the quantity demanded responds when the price changes.

$$\text{PED} = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in price}}$$

PED is normally negative (because of the law of demand), but we usually look at the size and drop the minus sign.

Worked example. The price of a good rises by 10% and the quantity demanded falls by 25%. Find the PED and classify the demand.

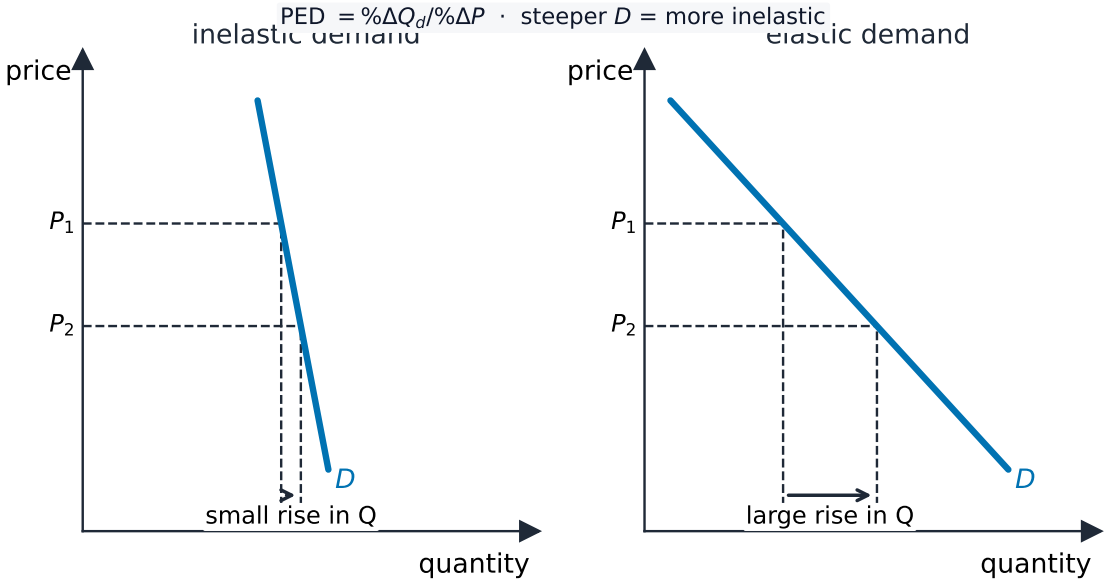
$$\text{PED} = \frac{-25}{10} = -2.5$$

Ignoring the minus sign, $2.5 > 1$, so demand is elastic — quantity responded much more than the price did.



Petrol has price-inelastic demand: drivers must buy it, so a higher price cuts the quantity bought only a little.

Value	Name	Meaning
more than 1	**elastic** 富有弹性	quantity responds a lot
less than 1	**inelastic** 缺乏弹性	quantity responds a little
equal to 1	**unit elastic** 单位弹性	quantity changes by the same % as price
0	**perfectly inelastic** 完全无弹性	quantity does not change at all
infinity	**perfectly elastic** 完全有弹性	any price rise sends demand to zero



When demand is inelastic (steep), a price change barely moves quantity; when it is elastic (shallow), the same price change moves quantity a lot.

What makes demand elastic or inelastic

- the number and closeness of substitutes (more substitutes → more elastic).
- whether the good is a **necessity** 必需品 (inelastic) or a **luxury** 奢侈品 (elastic).
- the share of income spent on the good (a large share → more elastic).
- time — demand is more elastic in the long run, when buyers can find other options.
- habit or addiction makes demand inelastic.

PED and total revenue

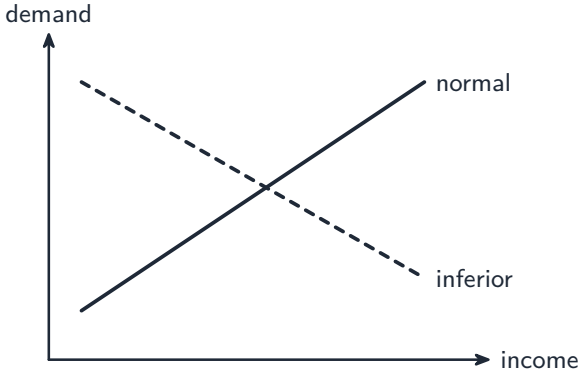
total revenue 总收益 is the money a firm earns: TR = price × quantity. PED tells a firm what happens to revenue when it changes the price:

If demand is...	price rises	price falls
inelastic	total revenue rises	total revenue falls
elastic	total revenue falls	total revenue rises
unit elastic	total revenue stays the same	total revenue stays the same

This is useful to firms (how to price for more revenue) and to **governments** 政府 (a tax on an inelastic good, like petrol, raises a lot of revenue without cutting sales much).

Income elasticity of demand

income elasticity of demand 需求收入弹性 (YED) measures how much demand responds when income changes.



As income rises, demand for a normal good rises but an inferior good falls

$$\text{YED} = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in income}}$$

- a **normal good** 正常品 has positive YED: demand rises as income rises. Necessities have YED between 0 and 1; luxuries have YED above 1.
- an **inferior good** 低档品 has negative YED: demand falls as income rises (people switch to better goods). An example is cheap instant noodles.

Worked example. A person's income rises by 20% and their demand for restaurant meals rises by 30%. Find the YED and classify the good.

$$\text{YED} = \frac{+30}{+20} = +1.5$$

YED is positive and greater than 1, so restaurant meals are a normal good — and a luxury.

This helps firms plan: as incomes grow, demand for luxuries grows fastest.

Cross elasticity of demand

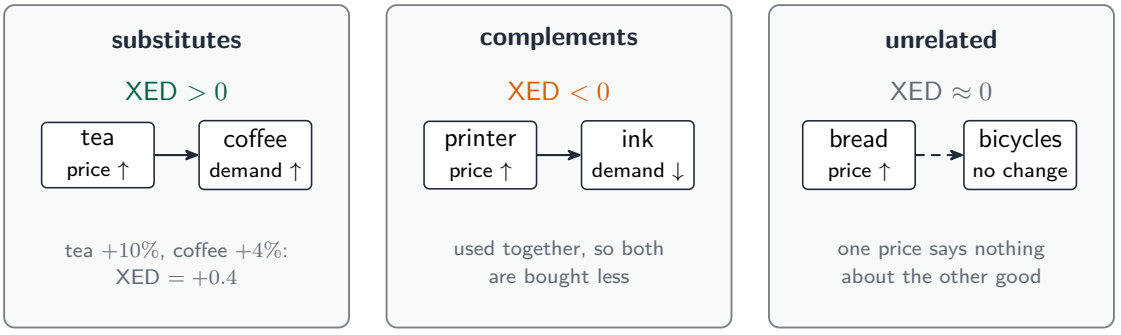
cross elasticity of demand 需求交叉弹性 (XED) measures how much demand for one good responds when the price of *another* good changes.

$$\text{XED} = \frac{\%\text{ change in quantity demanded of A}}{\%\text{ change in price of B}}$$

- for **substitutes**, XED is positive (the price of B rises, so people buy more A).
- for **complements**, XED is negative (the price of B rises, so people buy less A and less B).
- for goods that are not related, XED is about zero.

Worked example. When the price of tea rises by 10%, the quantity of coffee demanded rises by 4%. Find the XED and say how the two goods are related.

$$\text{XED} = \frac{+4}{+10} = +0.4$$



The sign of XED: positive for substitutes, negative for complements, about zero for unrelated goods

XED is positive, so tea and coffee are substitutes.

This helps a firm watch its rivals: if XED with a rival's good is high and positive, a price cut by the rival will hurt its sales.

Price elasticity of supply

price elasticity of supply 供给价格弹性 (PES) measures how much the quantity supplied responds when the price changes.

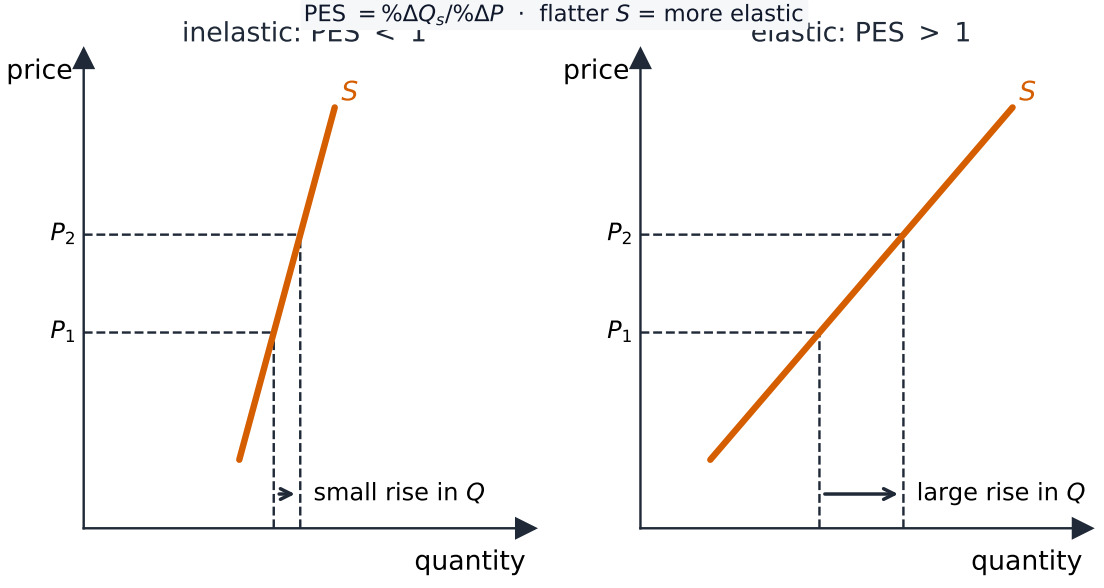
$$\text{PES} = \frac{\%\text{ change in quantity supplied}}{\%\text{ change in price}}$$

PES is positive. Supply is elastic if it responds a lot, inelastic if it responds little.

Worked example. The price of a good rises by 8% and the quantity supplied rises by 4%. Find the PES.

$$\text{PES} = \frac{4}{8} = 0.5$$

PES is less than 1, so supply is inelastic.



The same price rise: a small supply response when PES is below 1, a large one when PES is above 1

What makes supply more elastic:

- **spare capacity** 闲置产能 — unused machines and workers, so firms can quickly make more.
- **stocks** 库存 — goods stored in a warehouse that can be sold fast.
- how easily factors of production can move into the industry.

- time. Economists split time into three periods:
  - the **momentary** 瞬时 period — supply is fixed; PES is zero.
  - the **short run** 短期 — at least one factor is fixed, so supply can rise only a little.
  - the **long run** 长期 — all factors can change, so supply is most elastic.

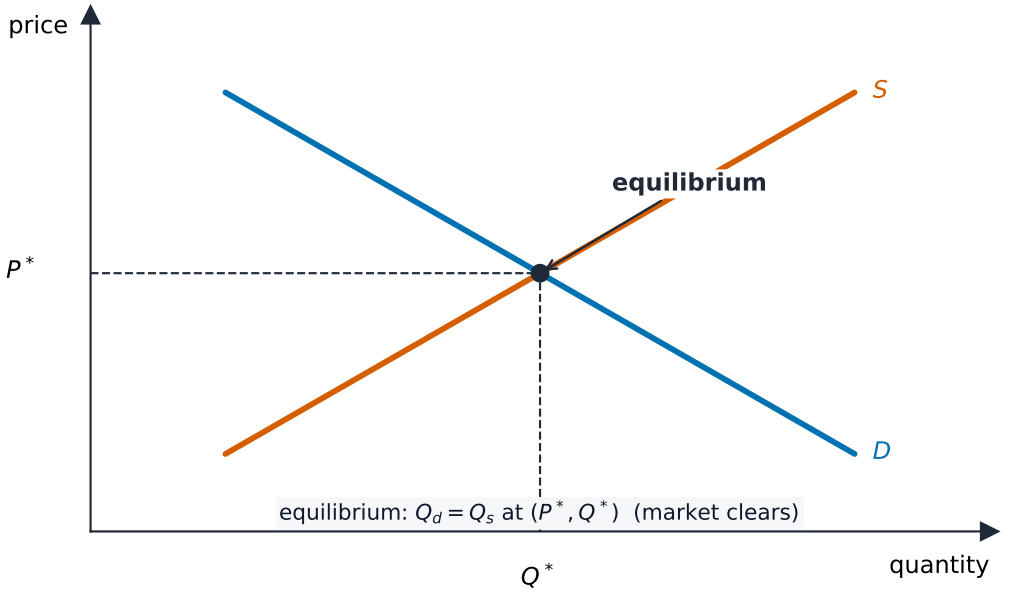
The interaction of demand and supply



In a market, buyers and sellers meet and the price settles where demand equals supply

Equilibrium

The market is in **equilibrium** 均衡 where the demand curve crosses the supply curve. Here demand equals supply. This gives the **equilibrium price** 均衡价格 and the **equilibrium quantity** 均衡数量. There is no pressure for the price to change.



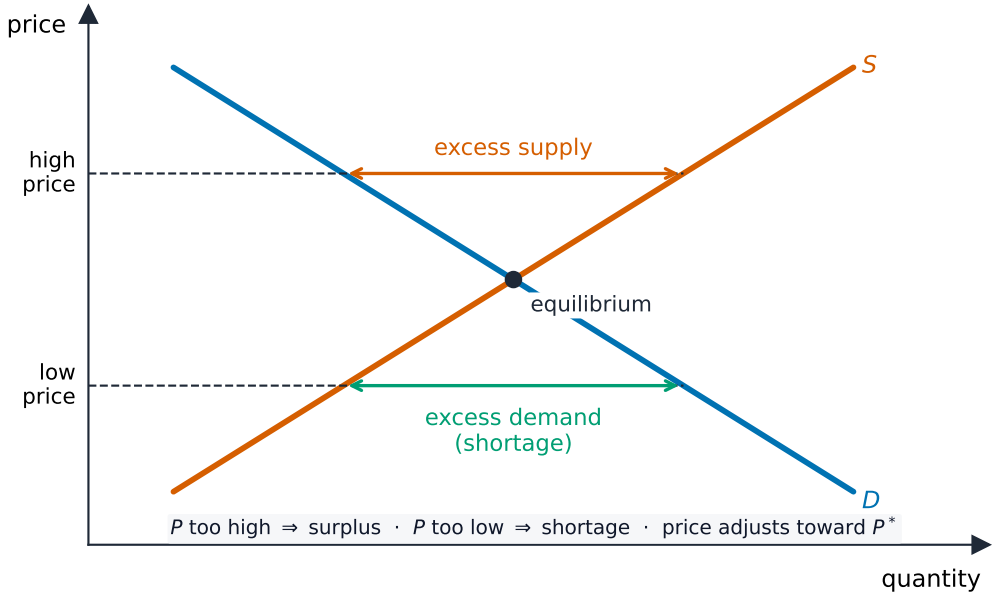
At the equilibrium price P^* and quantity Q^* the curves cross, so quantity demanded equals quantity supplied.

Disequilibrium

When the price is not at equilibrium, the market is in **disequilibrium** 非均衡:

- if the price is **too low**, there is **excess demand** 超额需求 (a **shortage** 短缺). Buyers compete, so the price is pushed up.
- if the price is **too high**, there is **excess supply** 超额供给. Sellers cannot sell everything, so the price is pushed down.

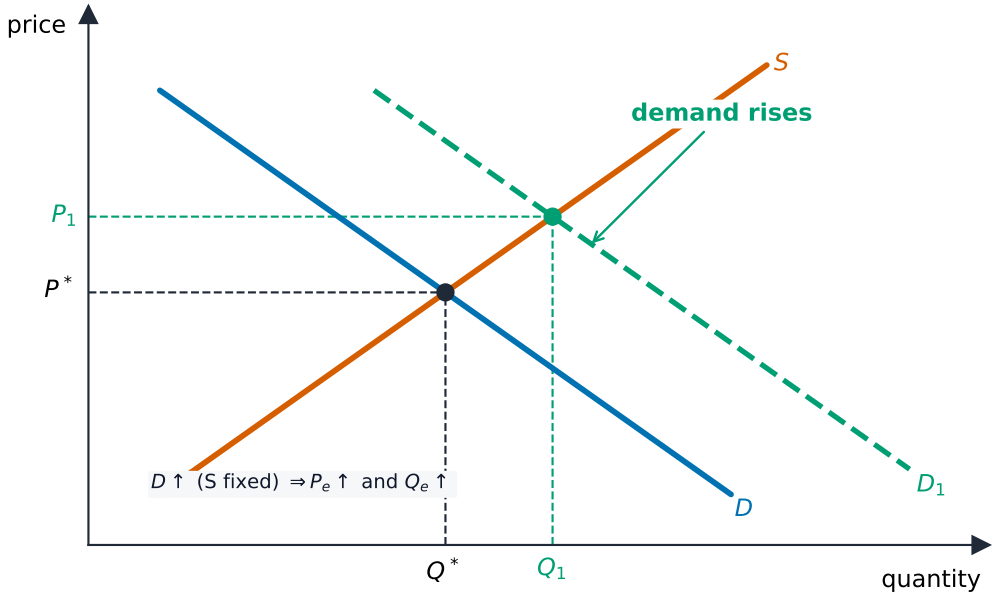
In both cases the price moves back to equilibrium. This shows the price mechanism doing its jobs of signalling, rationing and giving an incentive.



Above the equilibrium price there is excess supply; below it there is excess demand (a shortage). Either way the price is pushed back to equilibrium.

How shifts change the equilibrium

Change	Effect on price	Effect on quantity
demand rises (shifts right)	rises	rises
demand falls (shifts left)	falls	falls
supply rises (shifts right)	falls	rises
supply falls (shifts left)	rises	falls



With supply fixed, a rise in demand ($D \rightarrow D_1$) raises both the equilibrium price and the equilibrium quantity.

Demand and supply in different markets

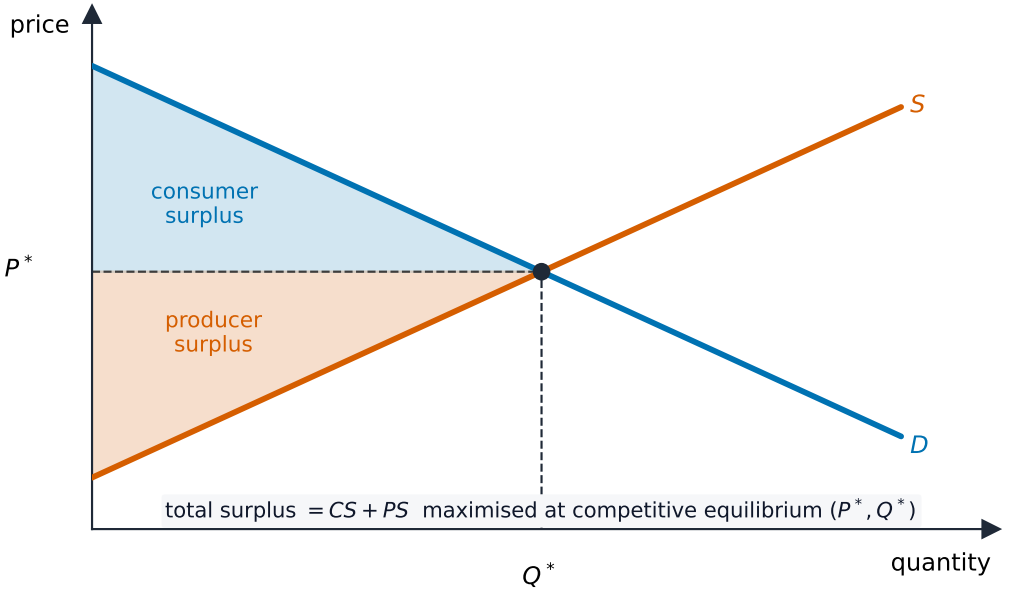
The same analysis works in many markets:

- in **product markets** 产品市场, demand and supply set the price of goods and services.
- in **factor markets** 要素市场, they set the price of factors — for example, the demand for and supply of labour set the **wage** 工资.
- in the **foreign exchange market** 外汇市场, the demand for and supply of a currency set the **exchange rate** 汇率 (the price of one currency in terms of another).

Consumer and producer surplus

consumer surplus 消费者剩余 is the gain to buyers. It is the difference between the most a consumer was willing to pay and the price actually paid. On a diagram it is the area below the demand curve and above the price line.

producer surplus 生产者剩余 is the gain to sellers. It is the difference between the price a producer receives and the lowest price the producer would have accepted. On a diagram it is the area above the supply curve and below the price line.



Consumer surplus is the area below demand and above the price; producer surplus is the area above supply and below the price.

How they change:

- if demand rises, the price and quantity rise, so producer surplus rises and consumer surplus usually rises too.
- if supply rises, the price falls and quantity rises, so consumer surplus rises.
- a higher price (with the curves fixed) raises producer surplus but lowers consumer surplus.

Together, consumer surplus plus producer surplus is the total **welfare** 福利 (gain to society) from the market. It is largest at the free-market equilibrium.

Exam tips

- Quote the **elasticity formula** and interpret the value: PED inelastic (< 1), elastic (> 1); link PED to total revenue.
- Distinguish a **movement along** (a price change) from a **shift** (a non-price factor) of demand or supply.
- Use the **sign** of XED (substitutes $+$, complements $-$) and YED (normal $+$, inferior $-$).
- On a diagram, label **consumer and producer surplus** and show how a price change alters each.

Government microeconomic intervention

A-Level Economics

Why markets fail

A free market often gives a good result, but not always. **market failure** 市场失灵 is when the market does not **allocate** 配置 (share out) resources efficiently — too much or too little of a good is made. Then the government may step in.

There are several reasons for market failure.

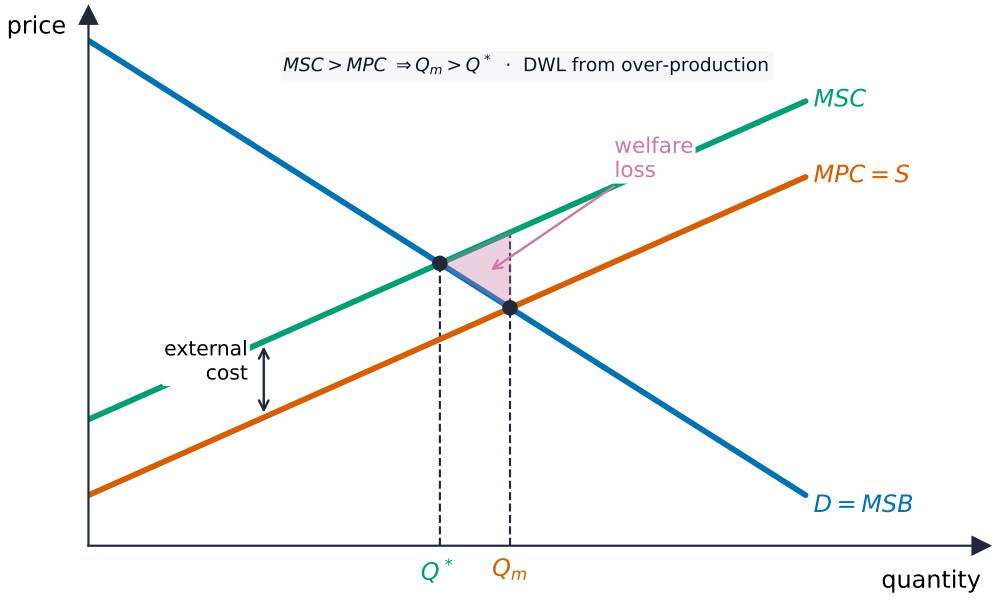


A coal power station: producing electricity can impose a negative externality such as pollution on people who are not part of the deal.

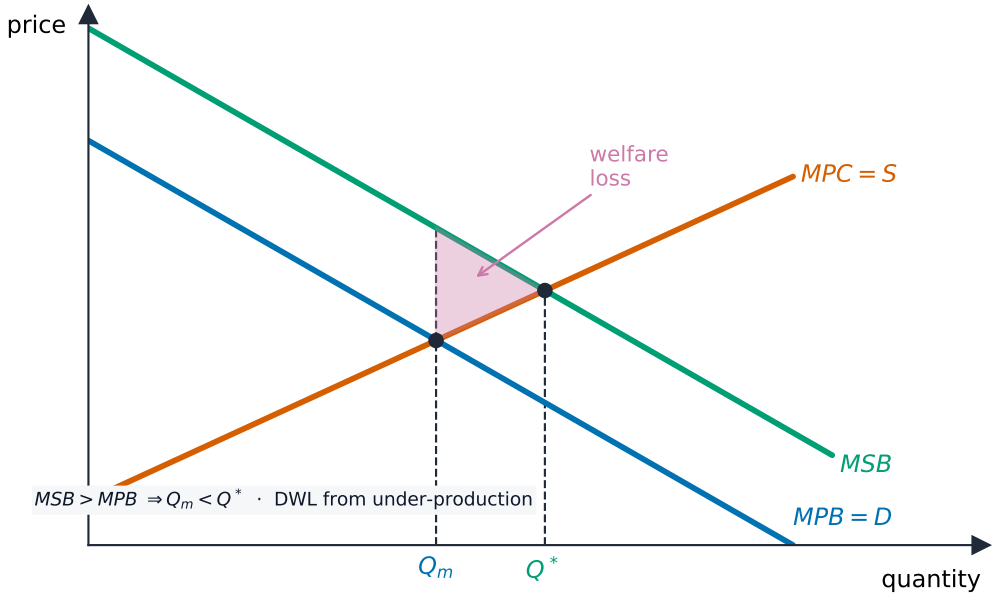
Externalities

An **externality** 外部性 is a cost or a benefit that falls on a **third party** 第三方 — someone who is not the buyer or the seller.

- a **negative externality** 负外部性 is an **external cost** 外部成本 paid by others, like the pollution from a factory. The **social cost** 社会成本 (the cost to all of society) is bigger than the **private cost** 私人成本 (the cost to the producer). When people ignore the external cost, the market makes **too much** of the good.
- a **positive externality** 正外部性 is an **external benefit** 外部收益 enjoyed by others, like the protection that vaccination gives to everyone. Here the market makes **too little**, because buyers ignore the benefit to others.



With a negative externality, social cost (MSC) is above private cost (MPC), so the market over-produces at Q_m instead of the social optimum Q^*; the shaded triangle is the welfare loss.



With a positive externality, social benefit (MSB) is above private benefit (MPB), so the market under-produces at Q_m instead of the social optimum Q^*.

Public goods

A **public good** 公共物品 (like street lighting or national defence) has **non-rivalry** 非竞争性 and **non-excludability** 非排他性. Because non-payers cannot be shut out, there is a **free-rider problem** 搭便车问题, so private firms make no profit and supply none. The market fails completely.

Merit and demerit goods

A **merit good** 优值品 (like education) is under-consumed; a **demerit good** 劣值品 (like cigarettes) is over-consumed. People misjudge them because of **information failure** 信息失灵 — they do not know, or do not use, the true benefit or harm.

Monopoly power

monopoly 垄断 means one firm dominates a market. A firm with **monopoly power** 垄断势力 can cut output and raise the price above the competitive level. This wastes resources and harms consumers.

Inequality and asymmetric information

- the market rewards people who own resources, so it can produce great **inequality** 不平等 — some people are left with too little to live on.
- **asymmetric information** 信息不对称 is when one side of a deal knows more than the other (a used-car seller knows the faults; the buyer does not). This leads to poor decisions and unfair deals.

Methods of government intervention

Indirect taxes

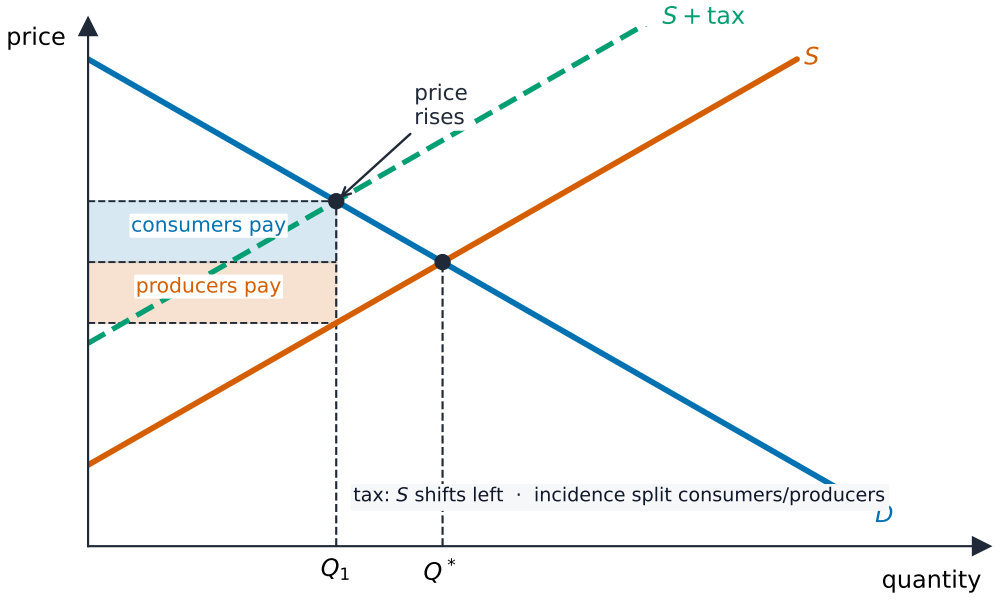
An **indirect tax** 间接税 is a tax on spending. It raises a firm's costs, shifts the supply curve to the left, and so raises the price and cuts the quantity. Governments tax demerit goods (like tobacco) and goods with negative externalities to make people use less.



Cigarettes are a demerit good — overconsumed because people misjudge the harm — so governments tax them heavily.

The **incidence of a tax** 税收归宿 is who really pays it. This depends on elasticity:

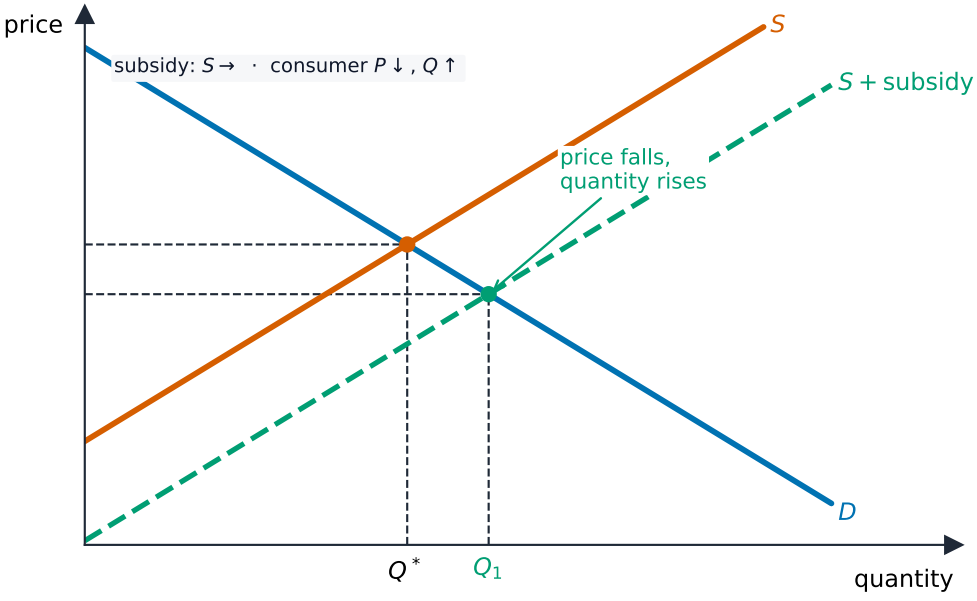
- if demand is **inelastic** 缺乏弹性, consumers pay most of the tax (the price rises by nearly the full tax).
- if demand is **elastic** 富有弹性, producers pay most of the tax (they cannot pass it on).



An indirect tax shifts supply up to $S + tax$, raising the price and cutting the quantity; the burden is shared between consumers and producers.

Subsidies

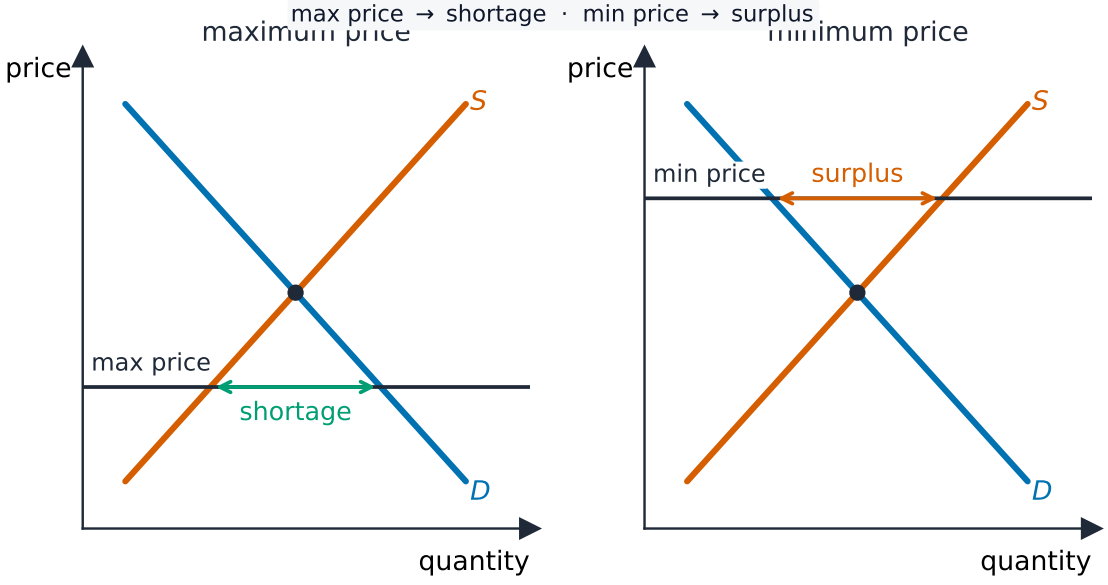
A **subsidy** 补贴 is a government payment to producers. It lowers their costs, shifts the supply curve to the right, and so lowers the price and raises the quantity. Subsidies are used for merit goods and goods with positive externalities.



A subsidy shifts supply down to S + subsidy, lowering the price and raising the quantity traded.

Maximum and minimum prices

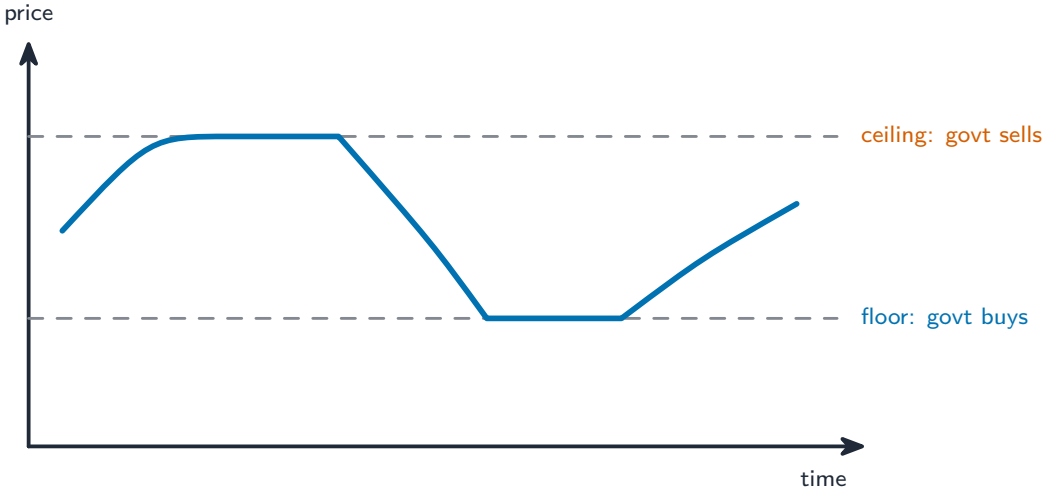
- a **maximum price** 最高限价 is a legal top price, set **below** equilibrium (for example, rent control). Because the price is held down, quantity demanded is greater than quantity supplied, so there is a **shortage** 短缺.
- a **minimum price** 最低限价 is a legal bottom price, set **above** equilibrium (for example, a minimum wage, or a floor for farm prices). Because the price is held up, quantity supplied is greater than quantity demanded, so there is **excess supply** 超额供给.



A maximum price set below equilibrium leaves a shortage; a minimum price set above equilibrium leaves a surplus.

Buffer stocks

A **buffer stock** 缓冲库存 is used to steady the price of farm goods. The government buys and stores the good when the price is low, and sells from the store when the price is high. This keeps the price inside a target band.



A buffer stock holds the price in a band: the government buys when it hits the floor and sells when it hits the ceiling

Regulation, provision and permits

- **regulation** 管制 means rules and laws — bans, age limits, safety standards, or pollution limits.
- **state provision** 政府提供 means the **government** 政府 supplies a good directly, often free at the point of use (public goods and merit goods like schools and hospitals).
- **tradable permits** 可交易许可证 set a total cap on pollution and give firms permits to pollute. Firms that pollute less can sell spare permits to firms that pollute more. This caps total pollution at the lowest cost.

Government failure

Intervention does not always help. **government failure** 政府失灵 is when the government's action makes the use of resources **worse**, not better. Common causes:

- poor information, so the government sets the wrong tax or subsidy.
- high cost of running the policy.

- unintended results — for example, a maximum price can create an illegal **black market** 黑市.
- decisions made for political reasons rather than for efficiency.

Worked example. A government puts a £2 per-packet tax on cigarettes, whose demand is very **price-inelastic**. Who really pays it, and what does that mean for the policy's two aims? The **incidence** falls mainly on whichever side is **less able to respond** - the more inelastic one. Because smokers keep buying almost the same quantity as the price rises, producers can pass most of the tax on, so **consumers bear most of the incidence**: the price might rise by about £1.80, with producers absorbing only about £0.20. That creates the tension the question is really testing. As a **revenue raiser** the tax succeeds: quantity barely falls, so revenue is large and steady. As a way of **cutting consumption** it largely fails - for exactly the same reason. And because poorer smokers spend a larger share of their income on it, the tax is **regressive**. Argue from **elasticity** to incidence, then judge each aim separately.

Income and wealth inequality

It helps to separate two ideas:

- **income** 收入 is a **flow** 流量 — money received over a period, such as wages, rent and interest.
- **wealth** 财富 is a **stock** 存量 — the value of assets owned at a point in time, such as houses, shares and savings.

Wealth produces income (savings earn interest), and income can be saved to build wealth, so the two are linked. Inequality has many causes: differences in skills and wages, the ownership of wealth, inheritance, and unemployment.

Policies to reduce inequality

- a **progressive tax** 累进税 takes a larger **percentage** 百分比 from higher incomes than from lower incomes. This narrows the gap after tax.
- **benefits** 福利金 (also called **transfer payments** 转移支付) are money paid by the government to the poor, the old, and the unemployed.
- **state provision** of free services, like education and health care, raises the real living standards of poorer people.

Together these policies bring about **redistribution** 再分配 — moving income and wealth from richer to poorer people. But there is a **trade-off** 取舍: very high taxes and large benefits may weaken the **incentive** 激励 to work and to invest.



Wealth and poverty side by side: markets alone may leave large income inequality unaddressed

Exam tips

- Define **market failure** as a misallocation of resources, then name the type (externality, public good, information gap).
- On an **externality** diagram, mark the divergence of private and social cost/benefit and shade the welfare loss.
- Evaluate each intervention (tax, subsidy, regulation, tradable permits) with one advantage and one drawback.
- Show that **tax incidence** depends on elasticity — the more inelastic side bears more of the tax.

The macroeconomy

A-Level Economics

Macroeconomics 宏观经济学 looks at the whole economy at once: its total output, its jobs, and its prices. This handout covers how we measure the economy and the main problems it can face.

National income statistics

To judge how an economy is doing, we measure its **national income** 国民收入 — the total value of everything it produces in a year.



National income (GDP) is the value of all the goods and services a country produces in a year.

- **gross domestic product** 国内生产总值 (GDP) is the value of all goods and services made **inside** a country in a year.
- **gross national income** 国民总收入 (GNI) is GDP plus the income a country's people and firms earn **abroad**, minus income foreigners earn inside it.

Two important differences:

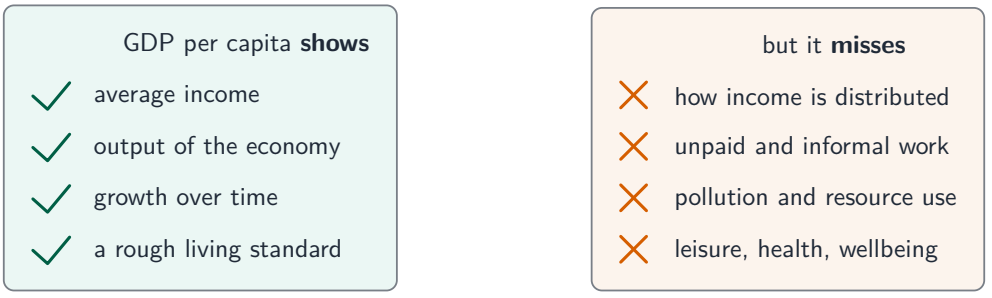
- **nominal** 名义 values are measured at current prices. **real** 实际 values are adjusted to remove the effect of rising prices, so they show the true change in output. Always compare real values over time.
- **total** values cover the whole country. **per capita** 人均 values are divided by the population, so they show the average per person.

Uses and limitations

National income is used to measure the **standard of living** 生活水平 and to compare countries. But it has limits:

- it ignores the **distribution** 分配 of income — a high average can hide great inequality.
- it leaves out unpaid work (housework) and the **informal economy** 非正规经济 (hidden, untaxed activity).
- it says nothing about pollution, leisure time, or health.

So a higher GDP per capita usually means a better life, but not always.



a higher GDP per capita usually means a better life - but not always

GDP per capita shows average income and output, but misses distribution, unpaid work and pollution

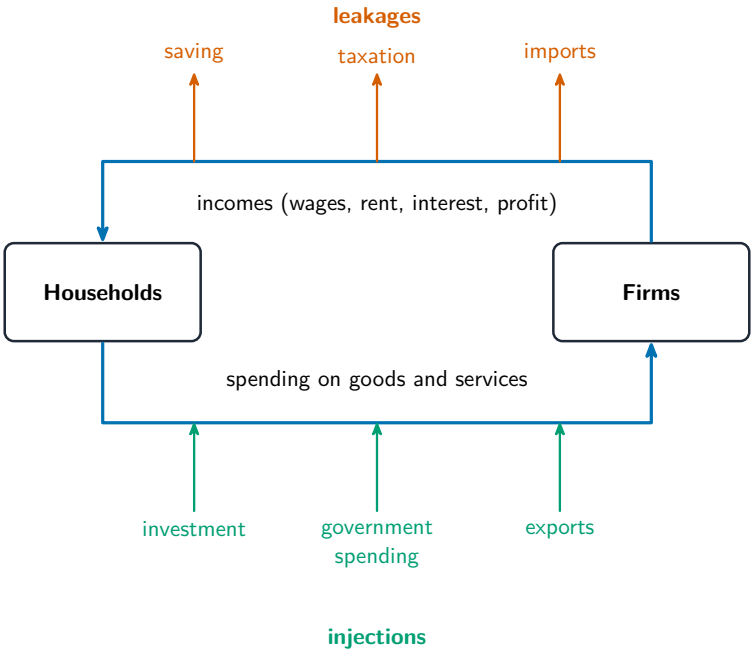
The circular flow of income

The **circular flow of income** 收入循环流 shows money moving around the economy. **Households** 家庭 supply factors of production to **firms** 企业 and receive income (wages, rent, interest, profit). They spend that income on goods and services, and the money flows back to firms.

Money also leaves and enters this flow:

- **leakages** 漏出 take money **out** of the flow: **saving** 储蓄, **taxation** 税收, and spending on **imports** 进口.
- **injections** 注入 put money **into** the flow: **investment** 投资, **government spending** 政府支出, and **exports** 出口.

The flow is in **equilibrium** 均衡 when total injections equal total leakages. If injections are greater, the economy grows; if leakages are greater, it shrinks.



Money loops between households and firms; leakages (saving, taxation, imports) drain it, while injections (investment, government spending, exports) add to it.

A **closed economy** 封闭经济 has no trade with other countries. An **open economy** 开放经济 trades, so it has exports and imports.

Aggregate demand and aggregate supply

Aggregate demand

aggregate demand 总需求 (AD) is the total spending on a country's goods and services at each price level. It has four parts:

$$AD = C + I + G + (X - M)$$

- **C** is **consumption** 消费 — spending by households.
- **I** is **investment** — spending by firms on capital.
- **G** is **government spending**.
- **(X − M)** is **net exports** 净出口 — exports minus imports.



$AD = C + I + G + (X - M)$: total spending on a country's goods and services

Aggregate demand is total spending: consumption, investment, government spending and net exports

The AD curve slopes downward: a lower price level means higher real spending, so a greater real output is demanded. AD shifts when any of its four parts changes (for example, a tax cut raises C).

Aggregate supply

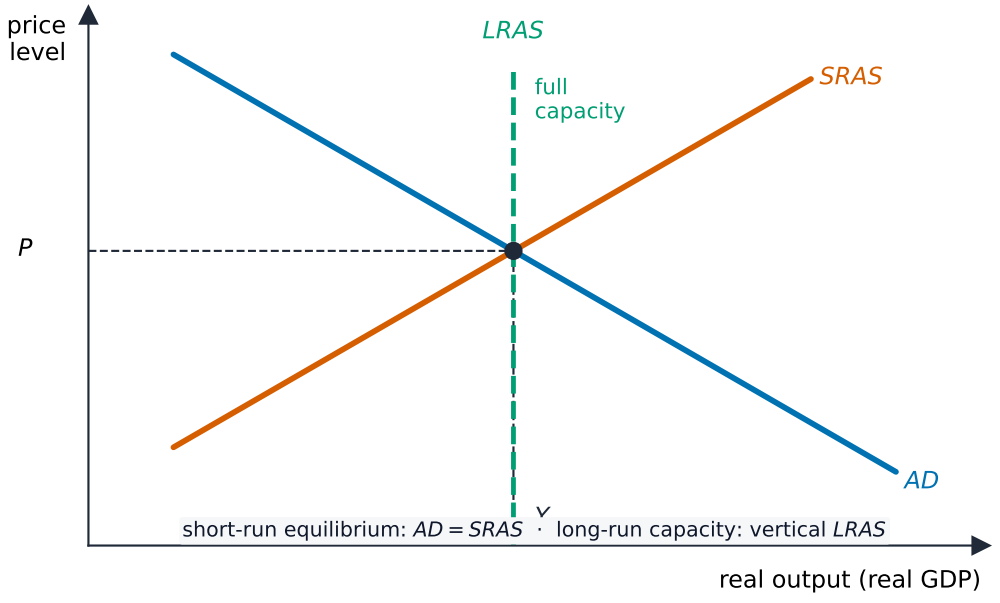
aggregate supply 总供给 (AS) is the total output firms are willing to make at each price level.

- **short-run aggregate supply** 短期总供给 (SRAS) slopes upward. In the short run, higher prices make production more profitable, so firms supply more. SRAS shifts when costs change (wages, raw materials, taxes).

- **long-run aggregate supply** 长期总供给 (LRAS) shows the economy's full **productive capacity** 生产能力. It shifts only when the quantity or quality of factors changes.

Macroeconomic equilibrium

macroeconomic equilibrium 宏观经济均衡 is where AD crosses AS. It sets the **price level** 价格水平 and the **real output** 实际产出 of the economy. A shift in AD or AS changes both.



Equilibrium sets the price level P and real output Y where AD meets short-run AS; vertical LRAS marks full capacity.

Economic growth

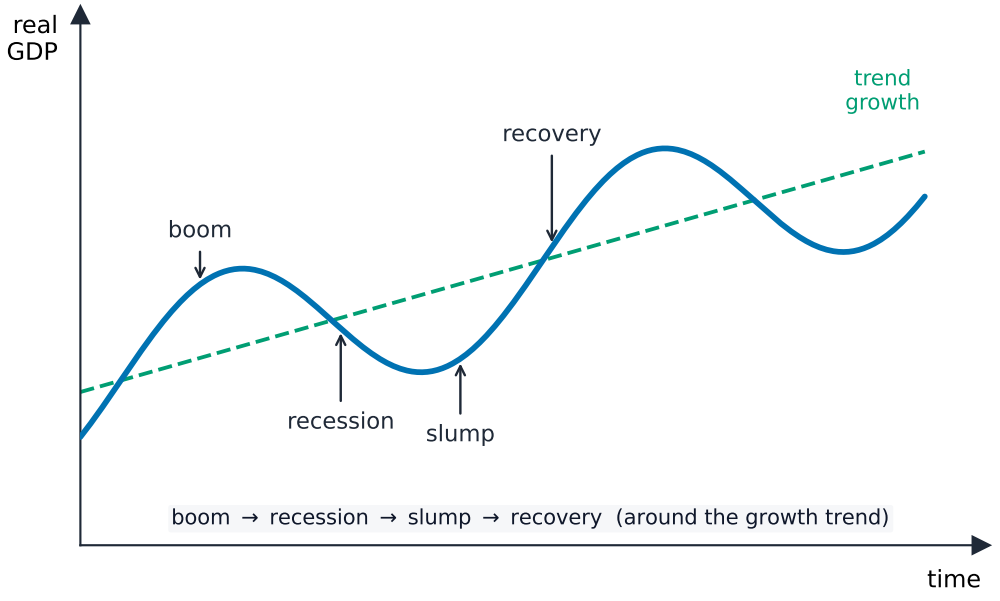
economic growth 经济增长 is a rise in a country's real output (real GDP) over time.

- **actual growth** 实际增长 is the rise in output that really happens (a movement to a point further out, or from inside the LRAS towards it).
- **potential growth** 潜在增长 is the rise in the economy's capacity (the LRAS shifting outward).

Growth is caused by more or better factors of production: a bigger or more skilled workforce, more investment in capital, new technology, and higher **productivity** 生产率 (output per worker).

The business cycle

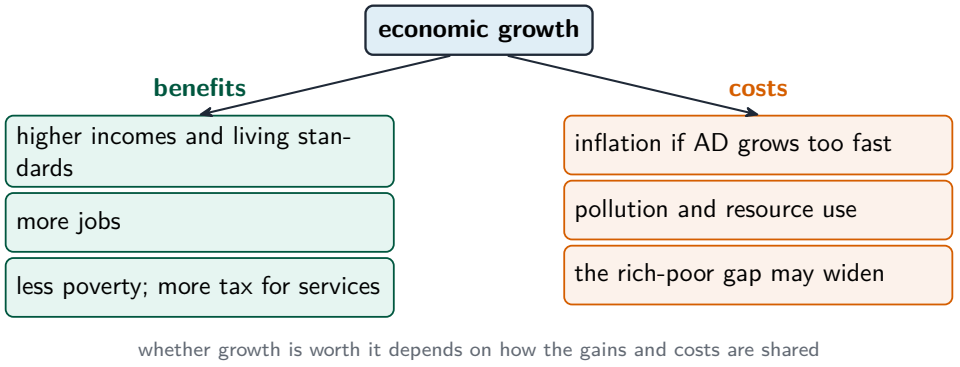
Real output does not grow smoothly. The **business cycle** 经济周期 (also called the trade cycle) is the up-and-down pattern of output over the years. It has four phases: **boom** 繁荣 (fast growth), **recession** 衰退 (falling output), **slump** 萧条 (the low point), and **recovery** 复苏.



Real GDP swings through boom, recession, slump and recovery around its rising long-run trend.

Costs and benefits of growth

Benefits	Costs
higher incomes and living standards	can cause inflation if AD grows too fast
more jobs	pollution and use of resources
less poverty; more tax for public services	the gap between rich and poor may widen



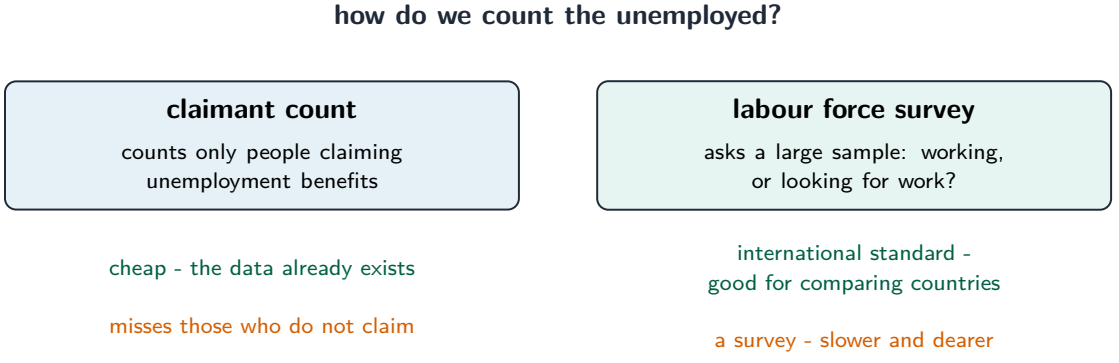
Economic growth brings higher incomes and jobs, but can bring inflation, pollution and inequality

Unemployment

unemployment 失业 means people who are able and willing to work, and are looking for a job, but cannot find one.

Measuring unemployment

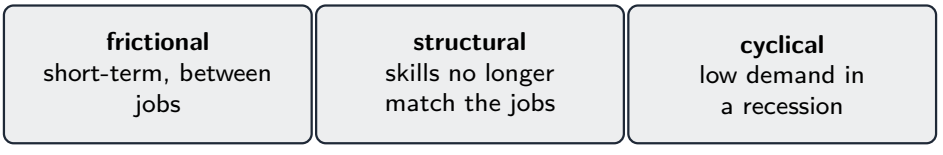
- the **claimant count** 申领救济人数 counts only people claiming unemployment benefits. It is cheap but misses those who do not claim.
- the **labour force survey** 劳动力调查 asks a large sample whether they are working or seeking work. It follows an international standard, so it is better for comparing countries. The **labour force** 劳动力 is everyone working or looking for work.



Two ways to count the unemployed - and what each one misses

Types and causes

- **frictional unemployment** 摩擦性失业 — short-term, while people move between jobs.
- **structural unemployment** 结构性失业 — when an industry declines and workers' skills no longer match the jobs available.
- **cyclical unemployment** 周期性失业 — caused by low AD in a recession. This is usually the largest type.
- **seasonal unemployment** 季节性失业 — when work is only available at certain times of the year (such as farming or tourism).



Three types of unemployment: frictional, structural and cyclical

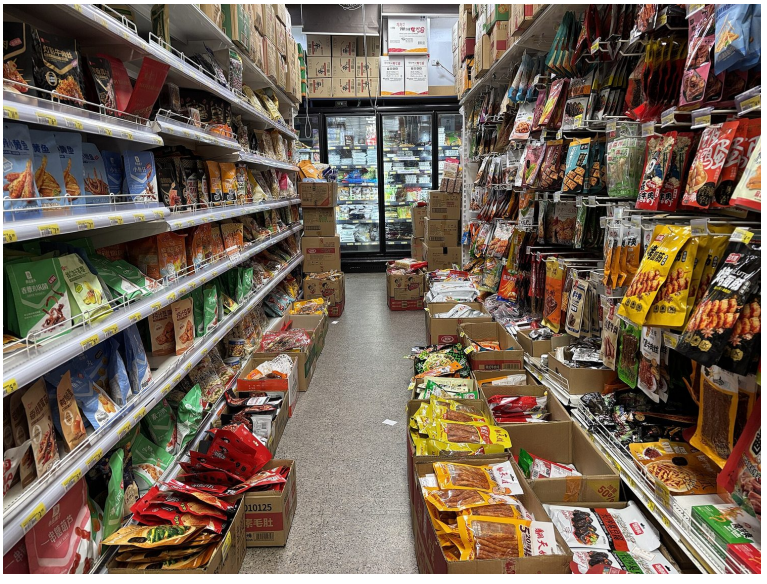
Consequences

Unemployment wastes output (the economy produces inside its PPC), lowers incomes, and costs the government (more benefits, less tax). It also harms families and can raise crime and ill health.

Price stability

Inflation, deflation and disinflation

- **inflation** 通货膨胀 is a sustained rise in the general price level. Money then buys less, so its **purchasing power** 购买力 falls.
- **deflation** 通货紧缩 is a sustained fall in the general price level.
- **disinflation** 反通货膨胀 is when inflation is still positive but is slowing down (prices rise, but more slowly).



Inflation is measured from the price of a basket of everyday goods like these.

The price level is measured with the **consumer price index** 消费者价格指数 (CPI). A typical household's spending is used to build a "basket" of goods. Each good is given a **weight** 权重 for how much is spent on it, and the change in the basket's cost from year to year is the inflation rate:

$$\text{inflation rate} = \frac{\text{change in index}}{\text{starting index}} \times 100\%.$$

Worked example. A country's CPI rises from 100 to 104 over one year, then to 107 the year after. Find the inflation rate in each year.

Year 1: $\dfrac{104-100}{100} \times 100\% = 4\%$. Year 2: $\dfrac{107-104}{104} \times 100\% = 2.9\%$.

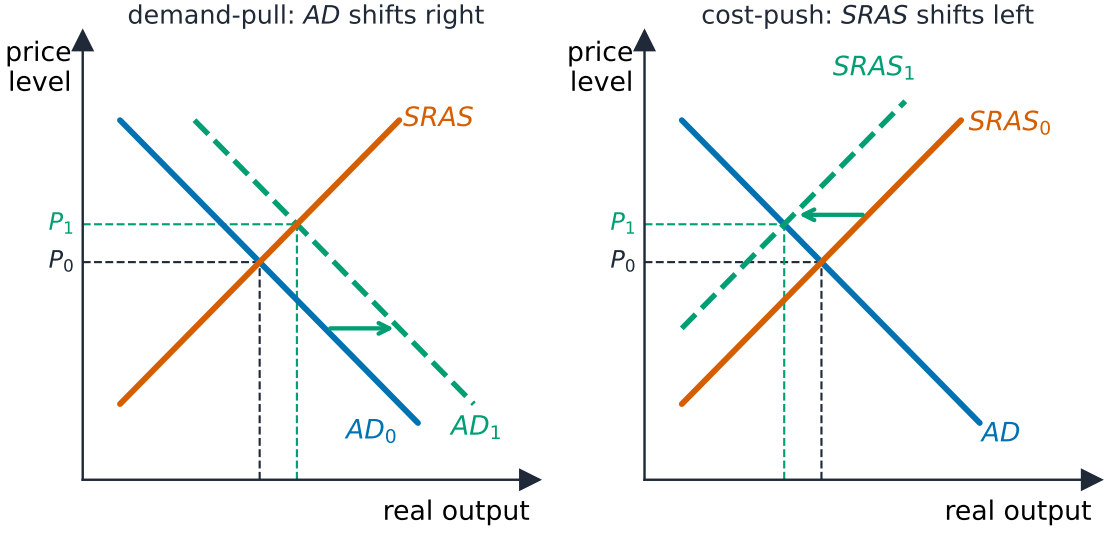
Notice that prices are still **rising** in year 2, just more slowly — this fall in the inflation rate (from 4% to 2.9%) is disinflation, not deflation.

Worked example. A worker's wage rises from £500 to £520 a week (a 4% rise) while the CPI rises by 4%. What happens to the **real wage**?

Money (nominal) pay is up 4%, but prices are also up 4%, so the real wage — what the pay can actually buy — is unchanged. A pay rise only raises living standards if it beats inflation.

Causes of inflation

- **demand-pull inflation** 需求拉动型通货膨胀 — AD rises faster than AS, so "too much money chases too few goods".
- **cost-push inflation** 成本推动型通货膨胀 — firms' costs rise (wages or raw materials), so they raise prices, shifting SRAS to the left.



Demand-pull inflation comes from AD shifting right; cost-push inflation comes from SRAS shifting left. Both raise the price level.

Consequences

Inflation lowers purchasing power, especially for people on fixed incomes. It creates uncertainty, can make a country's exports dearer, and helps borrowers but harms savers. Deflation sounds good but can be harmful: people delay buying (waiting for lower prices), so AD falls, firms cut output, and unemployment rises.

Exam tips

- Distinguish **real vs nominal** (real removes inflation) and calculate an **inflation rate** from a price index.
- Explain **AD** $= C + I + G + (X - M)$ and what shifts each part; distinguish an AD shift from an AS shift.
- Distinguish the **types of unemployment** (frictional, structural, cyclical) and how each is tackled.
- Distinguish **demand-pull vs cost-push** inflation on an AD/AS diagram.

Government macroeconomic intervention

A-Level Economics

Governments try to steer the whole economy. This handout covers what they aim for and the three sets of tools they use.

Macroeconomic policy objectives

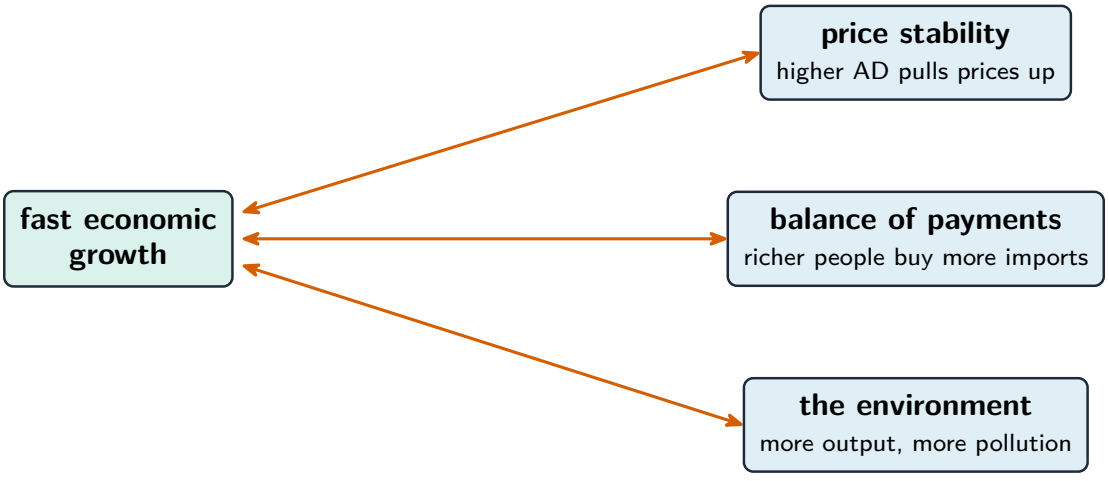
A government usually has five main aims:

- **price stability** 物价稳定 — low and steady **inflation** 通货膨胀 (often a target of about 2% a year).
- **full employment** 充分就业 — jobs for almost everyone who wants to work.
- **economic growth** 经济增长 — a rising real output over time.
- **balance of payments** 国际收支 stability — not importing far more than the country exports for years on end.
- **redistribution of income** 收入再分配 — narrowing the gap between rich and poor.

Conflicts between objectives

The aims can **conflict** 冲突 — reaching one can hurt another:

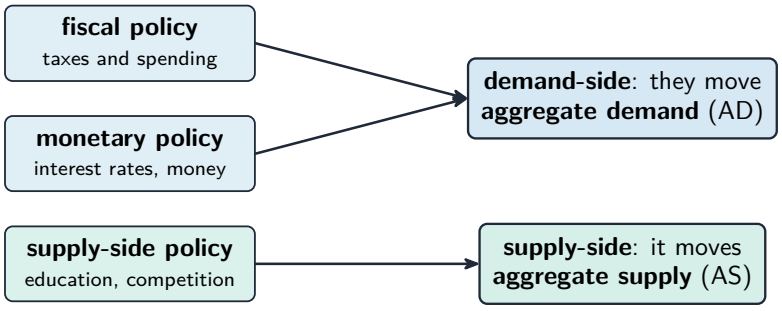
- **growth versus inflation**: fast growth raises **aggregate demand** 总需求, which can pull prices up.
- **growth versus the balance of payments**: as incomes rise, people buy more imports, worsening the trade balance.
- **unemployment versus inflation**: cutting unemployment by raising demand can cause more inflation.
- **growth versus the environment**: more output often means more pollution.



the same tension links **full employment** and **price stability**:
cutting unemployment by raising demand can cause more inflation

Chasing fast growth can pull against price stability, the balance of payments and the environment

The government has three sets of tools. Fiscal and monetary policy are **demand-side policies** 需求侧政策 (they work on AD). Supply-side policy works on **aggregate supply** 总供给.



Fiscal and monetary policy work on aggregate demand; supply-side policy works on aggregate supply

Fiscal policy

fiscal policy 财政政策 is the use of government **spending** and **taxation** to influence the economy.



Governments use fiscal policy — government spending and taxation, set through the legislature — to steer the economy.

The **budget** 预算 compares the two:

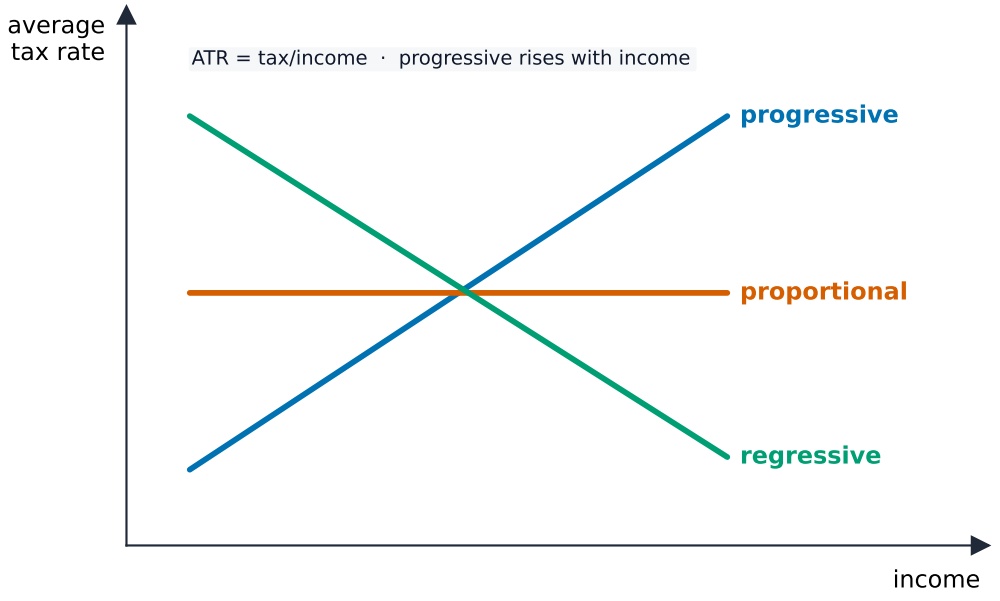
- a **budget deficit** 预算赤字 is when spending is greater than tax revenue. The government must borrow, adding to the **national debt** 国债.
- a **budget surplus** 预算盈余 is when tax revenue is greater than spending.

Types of tax

- a **direct tax** 直接税 is paid straight to the government on income or profit (income tax, corporation tax).
- an **indirect tax** 间接税 is a tax on spending (added to the price of goods).

Taxes are also grouped by how the rate changes with income:

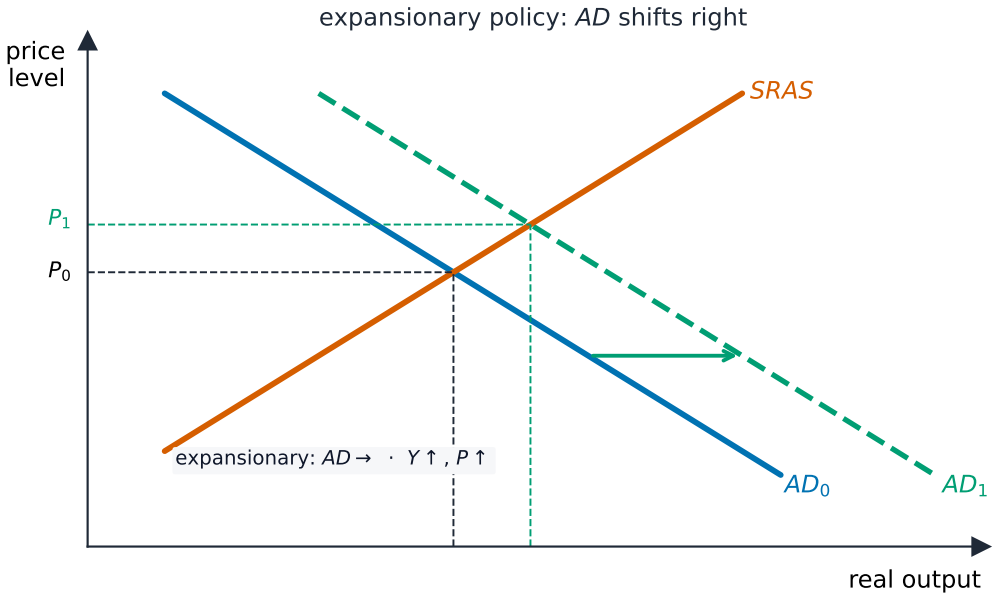
Type	How it works
progressive tax 累进税	the rate **rises** as income rises (the rich pay a larger share)
proportional tax 比例税	the rate **stays the same** at every income
regressive tax 累退税	the rate **falls** as income rises (the poor pay a larger share)



Under a progressive tax the average rate rises with income; under a proportional tax it stays flat; under a regressive tax it falls.

Using fiscal policy

- **expansionary** 扩张性 fiscal policy raises government spending or cuts taxes. This raises AD, to fight a recession and unemployment.
- **contractionary** 紧缩性 fiscal policy cuts spending or raises taxes. This lowers AD, to fight inflation.



Expansionary fiscal or monetary policy shifts AD right, raising output and the price level; contractionary policy does the reverse.

Effectiveness. Fiscal policy can act fast on AD, but it has problems: time lags (a budget is set once a year), large deficits and debt, and **crowding out** 挤出效应 — when heavy government borrowing pushes up interest rates and reduces private investment.

Monetary policy

monetary policy 货币政策 is run by the **central bank** 中央银行. Its main tool is the **interest rate** 利率, but it can also change the **money supply** 货币供给 or influence the **exchange rate** 汇率.



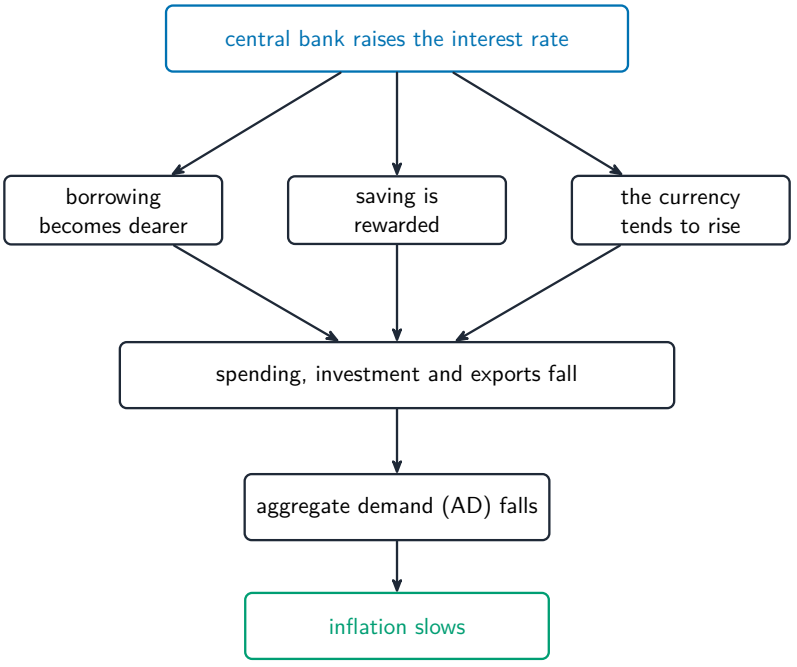
A central bank, like the Bank of England, sets the interest rate — the main tool of monetary policy.

The transmission mechanism

The **transmission mechanism** 传导机制 is how a change in the interest rate reaches the wider economy. If the central bank **raises** the interest rate:

- borrowing becomes dearer, so households and firms spend and invest less.
- saving is rewarded, so people spend less.
- the currency tends to rise, so exports fall.

All of these lower AD, which slows inflation. A **cut** in the interest rate does the opposite, raising AD.

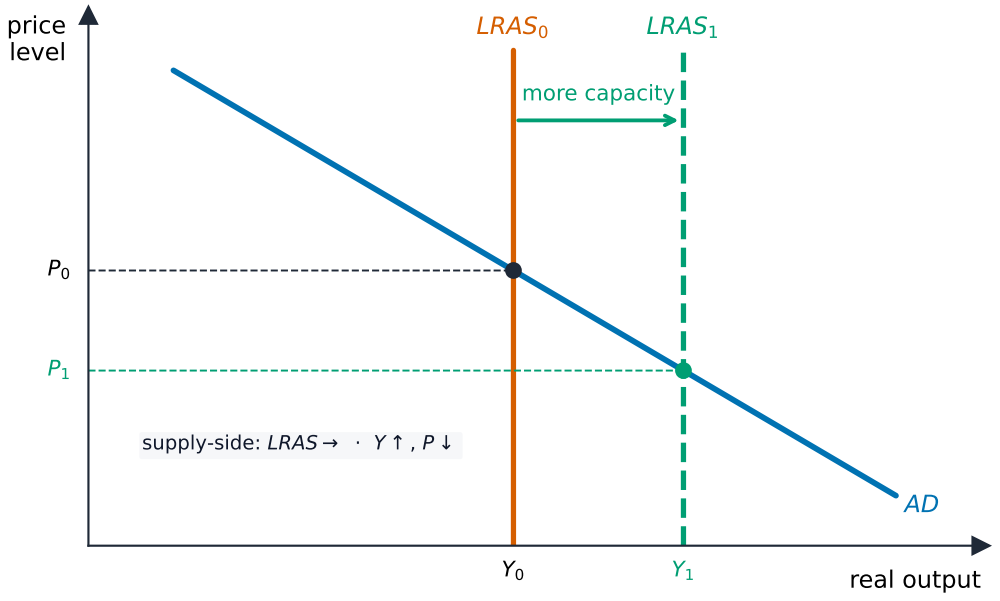


A higher interest rate works through three channels to cut spending, investment and exports, lowering AD and slowing inflation.

Effectiveness. Monetary policy is flexible (the rate can change at any time), but it works with a time lag of a year or more. In a deep slump, even very low rates may not make worried firms and households borrow.

Supply-side policy

supply-side policy 供给侧政策 tries to raise the economy's **productive capacity** 生产能力 — to shift long-run aggregate supply to the right. It aims to make markets work better and raise **productivity** 生产率.



Supply-side policy raises capacity, shifting LRAS right: real output rises and the price level falls — growth with lower inflation.

There are two kinds.

Market-based policies let markets work more freely:

- cutting income tax and benefits to strengthen the **incentive** 激励 to work.
- **privatisation** 私有化 — selling state firms to private owners, who run them for profit.
- **deregulation** 放松管制 — removing rules that block competition.

Interventionist policies use the government directly:

- spending on education and **training** 培训 to raise workers' skills.
- spending on **infrastructure** 基础设施 — roads, ports and broadband.
- support for research and new technology.

Effectiveness. Supply-side policy is powerful because it can raise growth **and** lower inflation at the same time, with no trade-off. But it is slow (education takes years to pay off), costly, and some measures, like cutting benefits, can widen inequality.

Worked example. An economy is in a deep recession with a large output gap, and interest rates are already near zero. Which policy should the government use, and why is the other one weak here? **Monetary policy is weak**: with the base rate already near zero there is almost no room to cut further, and even cheap credit will not make firms invest when they expect no demand. **Fiscal policy is stronger**: government spending raises aggregate demand **directly**, without waiting for anyone to choose to borrow, and the **multiplier** means each pound spent raises national income by more than a pound as it is re-spent. Weigh the costs against it - a bigger **budget deficit** and rising debt, plus **time lags** on large projects. Match the policy to the **constraint named in the question**: here it is the zero lower bound, and that is precisely what makes fiscal policy the answer.

Exam tips

- Match a **policy to an objective**: fiscal and monetary policy act on demand; supply-side policy raises long-run capacity.
- Explain how a change in **interest rates** works through AD, and fiscal policy through G and T.
- Evaluate policy with **time lags, side effects and conflicts** between objectives.
- **Supply-side** measures (education, tax incentives) are long-run and slow to work.

International economic issues

A-Level Economics

Why countries trade

international trade 国际贸易 lets countries buy and sell with each other. Trade lets each country **specialise** 专业化 in what it makes best, so the world makes more in total.



A container port: international trade moves goods between countries in huge volumes

Absolute and comparative advantage

- a country has an **absolute advantage** 绝对优势 in a good if it can make **more** of it with the same resources than another country.
- a country has a **comparative advantage** 比较优势 in a good if it can make it at a **lower opportunity cost** 机会成本 than another country.

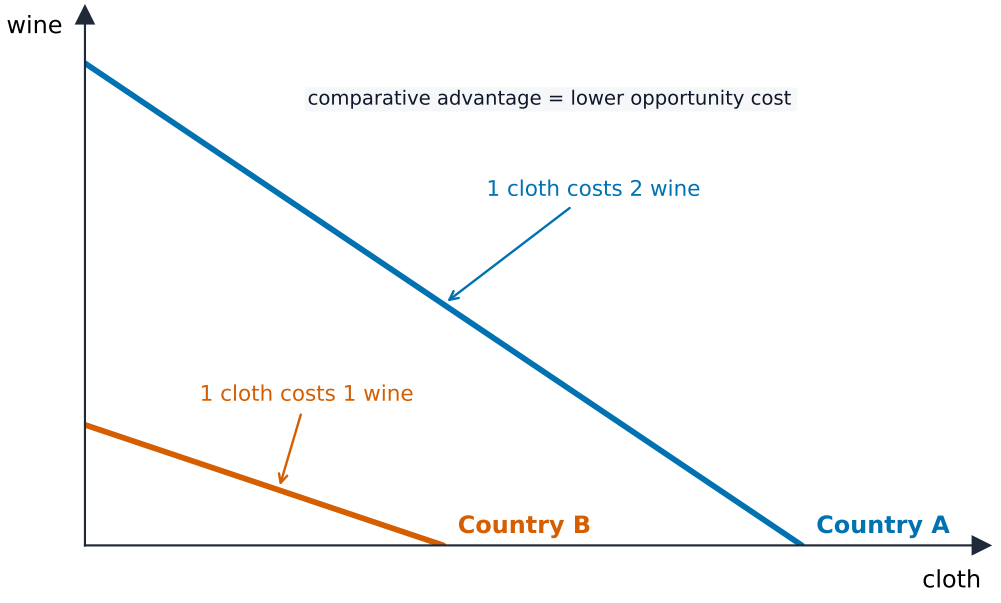
Comparative advantage is the key idea: a country should make the goods it gives up least to produce, and trade for the rest. Look at this example, where each country uses the same resources:

Country	Wine (units)	Cloth (units)
Country A	20	10
Country B	5	5

Country A is better at both (absolute advantage in both). But think about opportunity cost:

- in A, making 1 cloth means giving up 2 wine.
- in B, making 1 cloth means giving up 1 wine.

So B gives up less to make cloth — B has the comparative advantage in cloth. A has it in wine. If A makes only wine and B makes only cloth, and they trade, **both** can end up with more than before.



Country A can make more of both goods, but its PPC is steeper: it gives up 2 wine per cloth, while B gives up only 1, so B has the comparative advantage in cloth.

Terms of trade

The **terms of trade** 贸易条件 measure the price of a country's exports compared with the price of its imports. It is worked out as an index:

$$\text{terms of trade} = \frac{\text{index of export prices}}{\text{index of import prices}} \times 100.$$

If export prices rise faster than import prices, the terms of trade "improve" — the country can buy more imports for the same exports.

Worked example. Over a year, a country's export price index rises from 100 to 110, while its import price index rises from 100 to 105. Find the new terms of trade.

$$\text{terms of trade} = \frac{110}{105} \times 100 = 104.8.$$

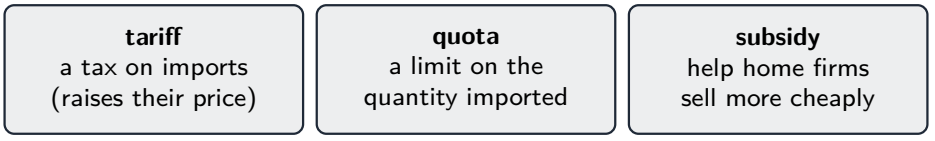
The figure is above 100, so the terms of trade have **improved**: each unit of exports now buys about 4.8% more imports than before.

Limits of the theory

The gains are real, but the simple theory assumes no transport costs, no trade barriers, and that costs do not change with output. In the real world these do not fully hold, and specialising in one good is risky if its price falls.

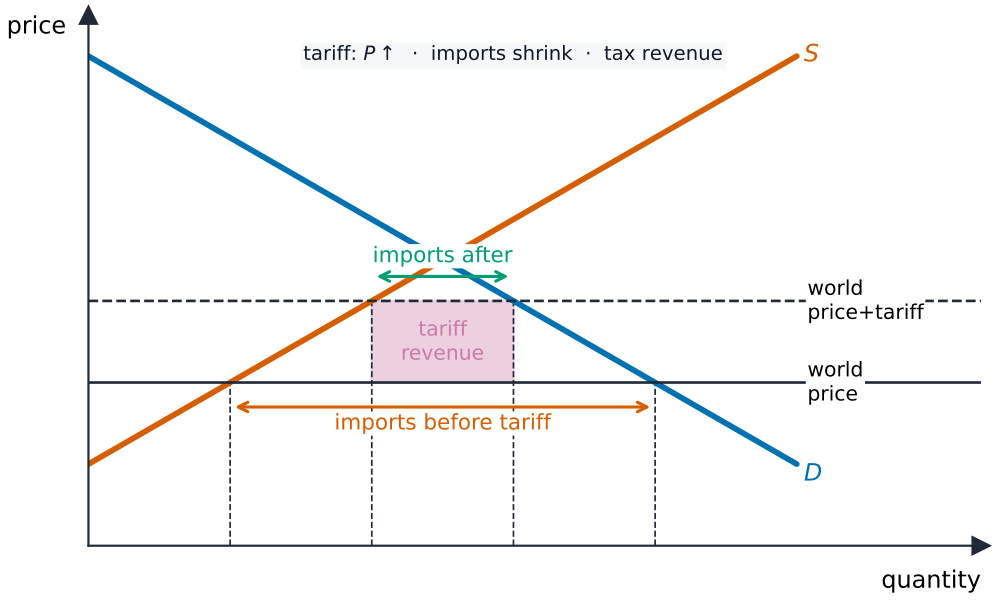
Protectionism

protectionism 贸易保护主义 means a government protecting home producers from foreign competition. The main tools:



Three tools of protection: tariffs, quotas and subsidies

- a **tariff** 关税 — a tax on imports, which raises their price.
- a **quota** 配额 — a legal limit on the quantity of a good that may be imported.
- a **subsidy** 补贴 to home producers, so they can sell more cheaply than foreign rivals.
- an **embargo** 禁运 — a complete ban on trade with a country or in a good.
- **administrative barriers** 行政壁垒 — extra paperwork and tough standards that make importing hard.



A tariff raises the price from the world price to "world price + tariff": domestic supply rises, demand falls, and imports shrink; the shaded box is the government's tariff revenue.

For and against

Arguments **for** protection include protecting an **infant industry** 幼稚产业 (a young industry that needs time to grow), saving jobs, and stopping **dumping** 倾销 (foreign firms selling below cost to destroy local rivals).

Arguments **against** are strong: protection raises prices for consumers, protects weak firms so they stay inefficient, and can lead to **retaliation** 报复 — other countries hitting back with their own barriers.

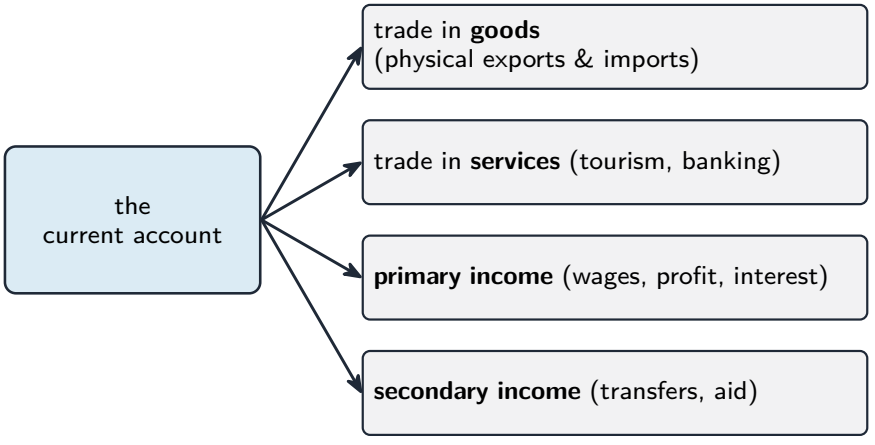
Effects on stakeholders

A tariff affects each **stakeholder** 利益相关者 differently: home producers gain, the government gains tariff revenue, but consumers lose (higher prices, less choice) and foreign producers lose sales.

The balance of payments

The **balance of payments** 国际收支 is a record of all money flows between one country and the rest of the world. Its most-studied part is the **current account** 经常账户, which has four parts:

- **trade in goods** 货物贸易 — exports and imports of physical goods.
- **trade in services** 服务贸易 — exports and imports of services (tourism, banking).
- **primary income** 初次收入 — income from work and investment abroad (wages, profit, interest).
- **secondary income** 二次收入 — transfers with nothing given back, like foreign aid.



The current account adds up four flows: trade in goods and services, plus primary income (from work and investment) and secondary income (transfers)

A **current account deficit** 经常账户赤字 means a country buys more from abroad than it sells. A **current account surplus** 经常账户盈余 is the opposite. A long deficit can mean rising debt to other countries; a large surplus can mean the country saves a lot but consumes little.

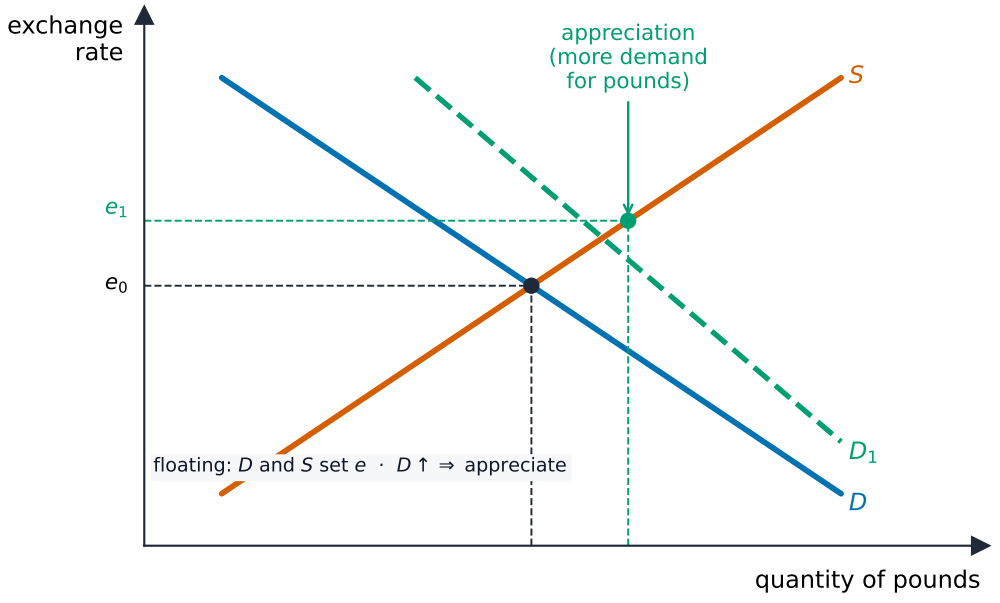
Exchange rates

An **exchange rate** 汇率 is the price of one currency in terms of another. There are three main systems.



A currency exchange: the exchange rate is the price of one currency in terms of another.

- a **floating exchange rate** 浮动汇率 is set freely by the demand for and supply of the currency. If more people want a currency (to buy its exports), its value rises.
- a **fixed exchange rate** 固定汇率 is held at a set value by the central bank, which buys and sells the currency to keep it there.
- a **managed exchange rate** 管理浮动汇率 mostly floats, but the central bank steps in now and then to steady it.



A floating exchange rate is set where demand for the currency meets supply. More demand for pounds ($D \to D_1$) raises the rate — an appreciation.

Appreciation and depreciation

Under floating rates:

- **appreciation** 升值 is a rise in the currency's value. Exports become dearer for foreigners, and imports become cheaper.
- **depreciation** 贬值 is a fall in the currency's value. Exports become cheaper, and imports become dearer.

A depreciation can therefore raise export sales and cut imports, helping the current account — but it can also raise inflation, because imports cost more.

Worked example. Suppose the exchange rate is £1 = \$1.40.

A UK exporter's £200 machine costs an American buyer £200 × 1.40 = \$280. If the pound then appreciates to £1 = \$1.60, the same machine costs £200 × 1.60 = \$320 — dearer for the American, so fewer are sold. Going the other way, a \$700 invoice from a US supplier costs a UK firm \$700 ÷ 1.40 = £500.

Correcting a current account deficit

A government can use three kinds of policy to cut a deficit:

- **expenditure-reducing** 支出减少 policies lower total demand (higher taxes or interest rates), so people buy fewer imports. But they also slow growth and raise unemployment.
- **expenditure-switching** 支出转换 policies move spending away from imports towards home goods (a tariff, or a lower exchange rate). But they can cause retaliation or inflation.
- **supply-side** policies raise the **competitiveness** 竞争力 of home firms (better skills and technology), so exports rise. These work best in the long run but are slow.



Trade flows at a container port: a current account deficit means imports of goods and services exceed exports

Exam tips

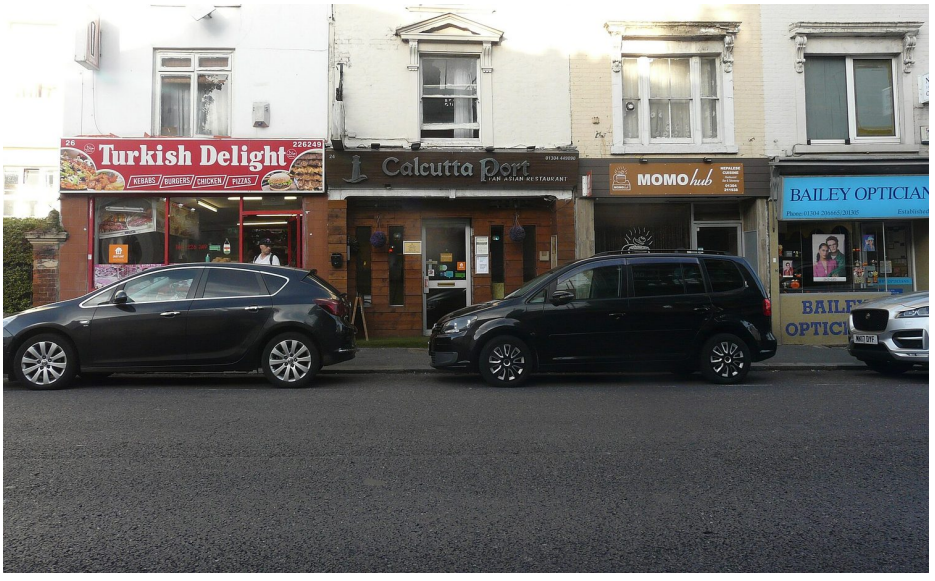
- Explain **comparative advantage** through opportunity cost (specialise where you give up least) and state its assumptions and limits.
- Evaluate **protectionism** (tariffs, quotas): it can protect jobs but raises prices and risks retaliation.
- Read the **balance of payments** (the current account covers trade, income and transfers).
- A **depreciation** makes exports cheaper and imports dearer, helping the current account only if demand is elastic enough (Marshall-Lerner).

Business and its environment

A-Level Business

Why businesses exist

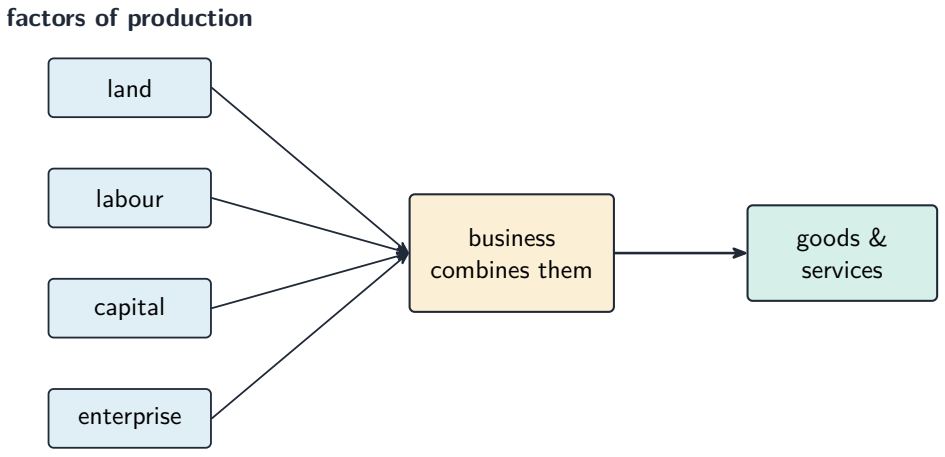
People have **needs** 需要 (things you must have, like food and shelter) and **wants** 欲望 (things you would like, like a phone or a holiday). A business exists to meet these needs and wants by making **goods** 商品 and **services** 服务.



Shops competing on a high street — the market a business operates in.

To do this, a business uses **resources** 资源. We sort resources into four **factors of production** 生产要素:

- **land** 土地 — natural things, such as soil, water and minerals.
- **labour** 劳动力 — the work done by people.
- **capital** 资本 — the machines, tools, buildings and money used to produce.
- **enterprise** 创业精神 — the skill of bringing the other three together and taking the **risk** 风险 to start a business.



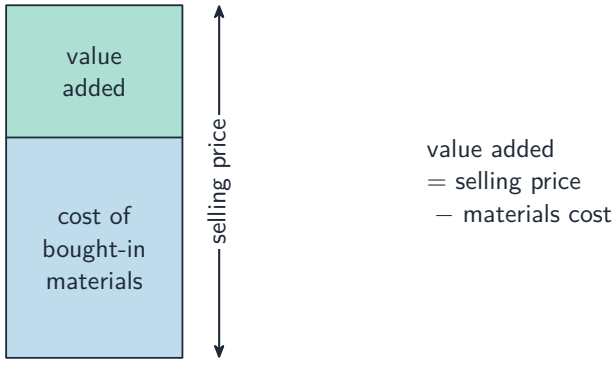
Land, labour, capital and enterprise are the four factors a business combines to make goods and services

Resources are **scarce** 稀缺 — there are not enough to meet everyone's wants. So every business must choose how to use them well.

Adding value

Adding value 增值 means the selling price of a product is higher than the cost of the **inputs** 投入 (the bought-in materials) used to make it.

$$\text{value added} = \text{selling price} - \text{cost of bought-in materials}$$



Value added is the selling price minus the cost of bought-in materials

A business can add value by **branding** 品牌, better quality, good design, fast service, or convenience. More value added can mean more **profit** 利润, but only if costs stay under control.

Entrepreneurs and intrapreneurs

An **entrepreneur** 企业家 is a person who takes the risk to set up and run a business. They have the idea, organise the factors of production, and make the decisions. If the business fails, they can lose their money.

An **intrapreneur** 内部创业者 is an employee inside a larger business who acts like an entrepreneur. They build new ideas, products or ways of working, using the company's resources.



A food truck is enterprise you can see. The owner spotted a gap, spent their own savings on the van and stock, and now takes the daily risk of running it — the essence of being an entrepreneur

Characteristics of successful entrepreneurs

Successful entrepreneurs are usually:

- **innovative** 有创新精神的 — full of new ideas.
- willing to take risks, but in a careful way.
- hard-working and **determined** 有决心的 — they do not give up easily.
- **self-confident** 自信的 and good at making decisions.
- good leaders who can plan ahead and manage money.

Why enterprise matters to a country

When people start new businesses (**business start-ups** 初创企业), the whole country can gain:

- new jobs are created, which lowers **unemployment** 失业.
- more goods and services are made, which raises **output** 产出 (and the country's **GDP** 国内生产总值).
- more **competition** 竞争, giving customers lower prices and more choice.
- some firms grow and **export** 出口, bringing money into the country.
- **economic growth** 经济增长 and a higher standard of living.

The business plan

A **business plan** 商业计划书 is a written document that describes the business idea and how it will work. It usually contains:

Part of the plan	What it covers
The idea	the product, and what makes it different
The market	target customers, market size, competitors
Marketing	price, promotion and place
Operations	how the product is made or delivered
Finance	start-up costs, a **cash-flow forecast** 现金流量预测, expected profit
People	the owners and key staff

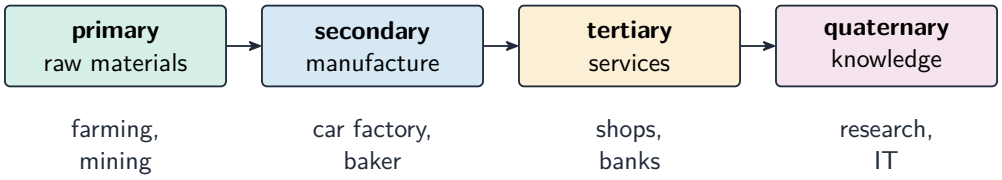
A business plan is useful because it helps the owner:

- get **finance** 融资 from a bank or an **investor** 投资者.
- set clear goals and lower the risk of failure.
- guide decisions as the business grows.

Sectors of business activity

Every business works in one of four sectors:

Sector	What it does	Example
primary sector 第一产业	takes raw materials from nature	farming, mining, fishing
secondary sector 第二产业	makes (manufactures) goods	car factory, baker
tertiary sector 第三产业	provides services	shops, banks, transport
quaternary sector 第四产业	knowledge and information services	research, IT, consulting



Business activity runs from the primary sector through to the quaternary sector

Here **raw materials** 原材料 are the natural inputs a business starts with, and to **manufacture** 制造 means to make goods in large amounts. As a country develops, it usually moves from mostly primary work towards more tertiary and quaternary work.

Private and public sectors

- The **private sector** 私营部门 is made up of businesses owned by private people. Their main aim is usually profit.
- The **public sector** 公共部门 is made up of organisations owned by the government, such as state schools and public hospitals. Their aim is usually to provide a service, not to make profit.

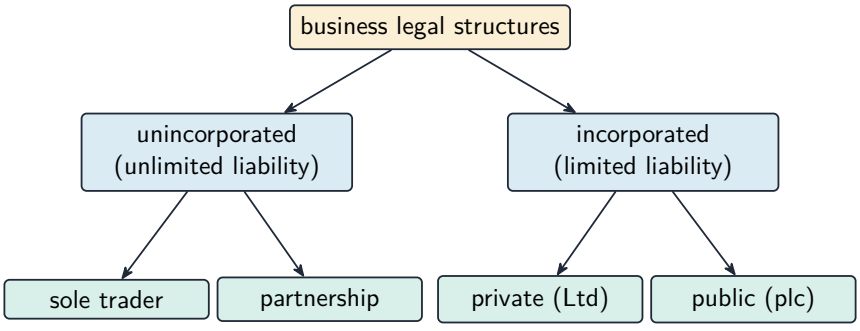
Legal structures of business



A company headquarters: businesses range from sole traders to giant public limited companies.

A business must choose a legal form. The main ones are below.

- A **sole trader** 个体经营者 is owned by one person. It is easy to set up and the owner keeps all the profit, but raising money is hard.
- A **partnership** 合伙企业 is owned by two or more partners who share the work and profit. They often sign a **deed of partnership** 合伙协议 that sets out each partner's share.
- A **private limited company** 私人有限公司 (Ltd) sells **shares** 股份 to a small group, such as family and friends. Its owners are **shareholders** 股东. Its shares are not sold to the public.
- A **public limited company** 上市公司 (plc) sells its shares to the public on the **stock exchange** 证券交易所. It can raise large amounts of money, but it must follow more rules and faces the risk of a **takeover** 收购.
- A **franchise** 特许经营 lets one firm use another firm's brand and business model. The **franchisor** 特许方 owns the brand; the **franchisee** 加盟方 pays a fee and a **royalty** 提成 to run a branch.
- A **joint venture** 合资企业 is when two businesses join to run one project and share the costs and risks.
- A **social enterprise** 社会企业 trades like a normal business but aims to do social or **environmental** 环境 good, and reinvests most of its profit.



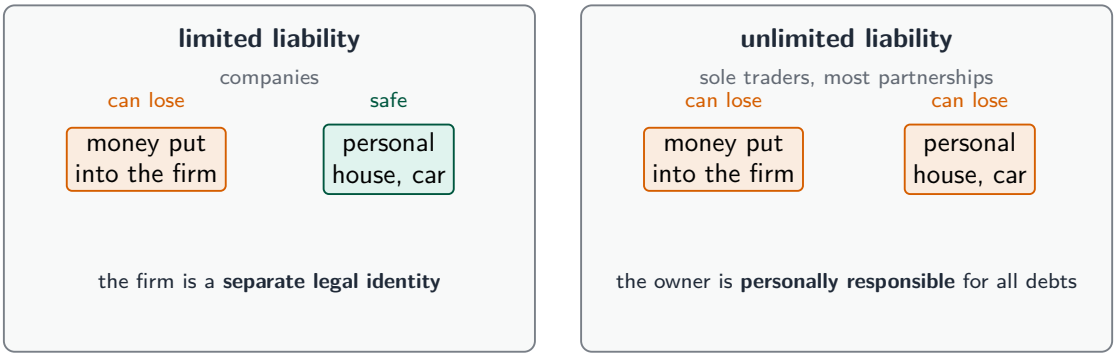
Legal forms split by liability: unincorporated (unlimited) and incorporated (limited)

Limited and unlimited liability

Liability 责任 means who must pay the business's **debts** 债务 if it cannot pay them itself.

- **limited liability** 有限责任 — the owners can only lose the money they put in. Their personal things, like a house or car, are safe. Companies have limited liability.
- **unlimited liability** 无限责任 — the owners are personally responsible for all the debts. They could lose personal things. Sole traders and most partnerships have unlimited liability.

A company is a **separate legal identity** 独立法人 — it can own things, owe money, and be taken to court in its own name, apart from its owners.



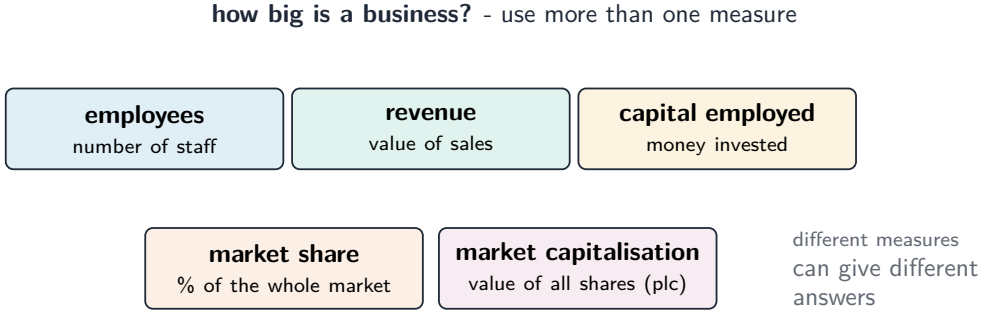
Limited liability risks only the money put in; unlimited liability risks the owner's personal things too

The size of a business

There is no single best way to measure how big a business is. Common measures are:

- the number of **employees** 员工.
- **revenue** 营业收入 (also called sales turnover) — the value of what it sells.
- **capital employed** 所用资本 — the total money invested in the business.
- **market share** 市场份额 — the firm's sales as a percentage of the whole market.
- **market capitalisation** 市值 — the total value of a company's shares (for a plc).

Different measures can give different answers, so it is best to use more than one.



Five ways to measure how big a business is; use more than one

Why small businesses matter

Small businesses create jobs, serve local **communities** 社区, give personal service, react quickly to change, bring new ideas, and supply parts and services to large firms.

Business growth

A business can grow in two ways.

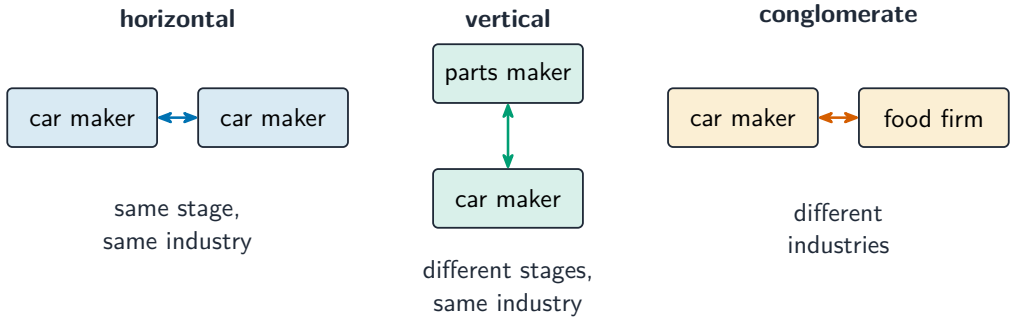
Internal growth 内部增长 (organic growth) — the business grows by itself: selling more, opening new branches, or adding new products. It is slower but lower risk.

External growth 外部增长 — the business joins with another firm. This is called **integration** 整合, and it happens by:

- a **merger** 合并 — two firms agree to join together as one.
- a **takeover** — one firm buys control of another.

There are three types of integration:

- **horizontal integration** 横向整合 — joining a firm at the same stage in the same industry (e.g. two car makers).
- **vertical integration** 纵向整合 — joining a firm at a different stage in the same industry. **Forward integration** 前向整合 is towards the customer (e.g. buying a shop); **backward integration** 后向整合 is towards the supplier (e.g. buying a parts maker).
- **conglomerate integration** 混合整合 — joining a firm in a completely different industry. This spreads risk, which is called **diversification** 多元化.



The three types of integration when one firm joins with another

Firms grow to lower their costs, win a bigger market share and more **market power** 市场支配力, raise profit, and spread risk. But growth also brings problems: higher average costs, harder management and communication, different company **cultures** 企业文化 that may clash, and a need for a lot of finance.

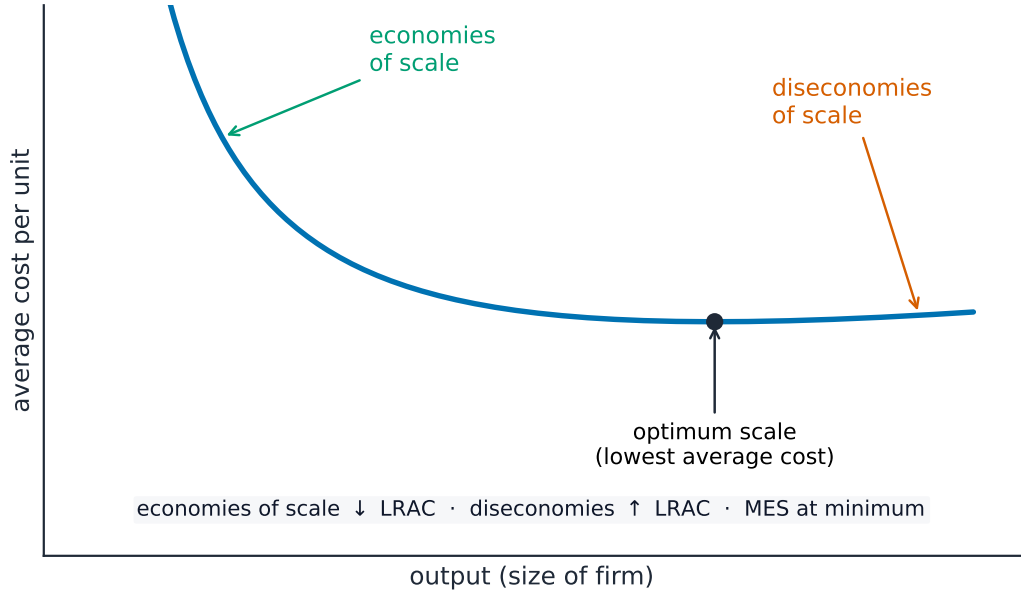
Economies and diseconomies of scale

Economies of scale 规模经济 are the cost savings a business gets as it grows larger. The **average cost** 平均成本 of each unit falls as output rises. The main types are:

Type	How it lowers cost
purchasing	buying in **bulk** 大批量 earns discounts
technical	large machines are more efficient
financial	large firms borrow at lower **interest** 利息
managerial	the firm can employ expert specialist managers
marketing	advertising cost is spread over more units

Diseconomies of scale 规模不经济 happen when a firm grows too big and the average cost starts to rise again. The usual causes are:

- **communication** 沟通 becomes slow and unclear.
- the **coordination** 协调 of many departments is hard.
- workers may feel unimportant and lose **motivation** 激励.



Average cost per unit falls (economies of scale) then rises (diseconomies of scale) as a firm grows

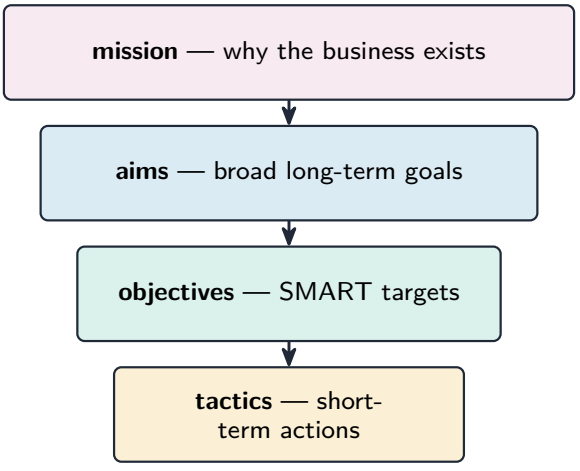
Worked example. A factory making 10,000 units has fixed costs of £200,000 and variable costs of £6 per unit. It expands to 50,000 units, and bulk buying cuts the variable cost to £5. Find the unit cost at each output and explain the fall. Unit cost = total cost ÷ output. At 10,000: total = 200,000 + (10,000 × 6) = £260,000, so the unit cost is 260,000 ÷ 10,000 = **£26**. At 50,000: total = 200,000 + (50,000 × 5)

= £450,000, so the unit cost is 450,000 ÷ 50,000 = **£9**. Two economies of scale are at work: the **fixed** costs are spread over far more units (£20 down to £4 each), and **purchasing** economies cut the variable cost. Note that the fixed cost **total does not change** - only the fixed cost **per unit** falls. Writing "fixed costs fall" is the error that loses the mark.

Aims, objectives and mission

A **mission statement** 使命宣言 is a short sentence that says why the business exists and what it wants to be. It guides everyone in the firm.

From the mission come **aims** 目的 (broad, long-term goals), and then **objectives** 目标 (the specific targets that help reach the aims).



Goals get narrower and more specific down the chain: a broad mission becomes aims, then SMART objectives, then day-to-day tactics

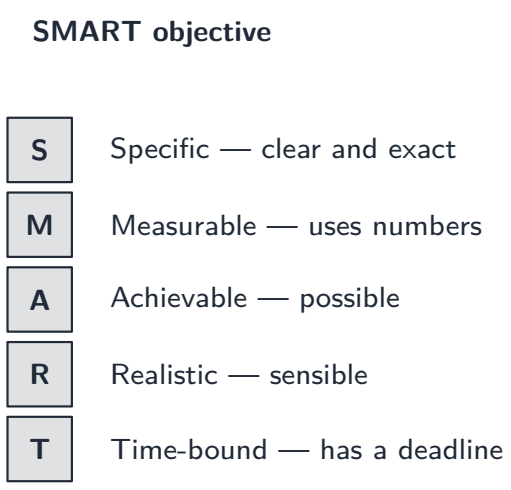
Good objectives are **SMART** — the first letters of:

- **Specific** 具体的 — clear and exact.
- **Measurable** 可衡量的 — you can check it with numbers.
- **Achievable** 可实现的 — it is possible to reach.
- **Realistic** 现实的 — sensible for the resources you have.
- **Time-bound** 有时限的 — it has a deadline.

A **strategy** 战略 is the long-term plan to meet objectives. **Tactics** 策略 are the short-term actions that carry out the strategy. Objectives are shaped by the size and age of the firm, who owns it, the company culture, and conditions in the market.

Types of business objective

Objective	What it means
profit maximisation 利润最大化	making the largest possible profit
growth	getting bigger — more sales, more branches
survival 生存	staying in business, vital for new firms and in hard times
market share	winning a larger share of the market
social and **ethical** 道德 objectives	doing what is morally right, such as fair pay



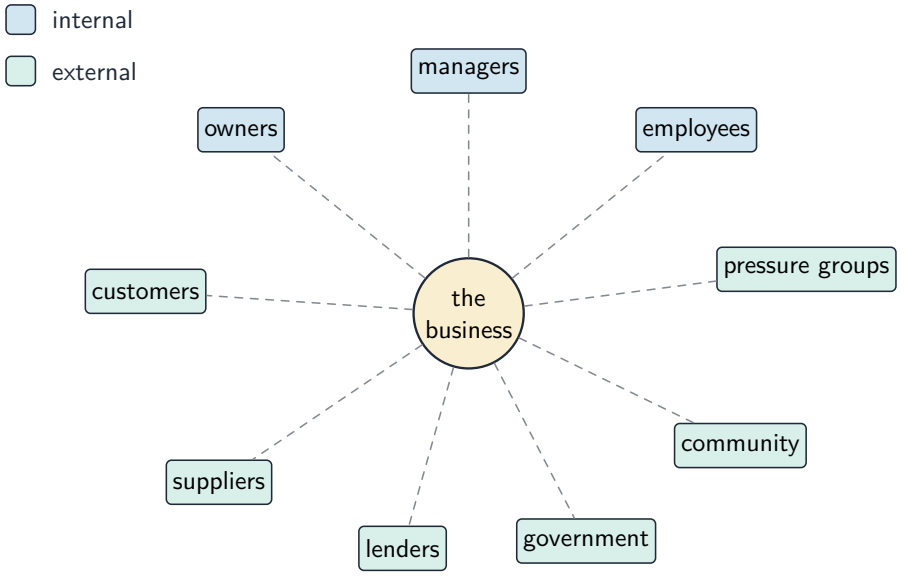
SMART objectives: Specific, Measurable, Achievable, Realistic, Time-bound

Corporate social responsibility 企业社会责任 (CSR) is the idea that a business should act in the best interests of society and the environment, not just make profit. Examples are treating workers fairly, cutting **pollution** 污染, and helping the local community.

A firm's objectives can change over time. A new firm may aim to survive; later it may aim to grow; a large, safe firm may focus on profit or CSR.

Stakeholders

A **stakeholder** 利益相关者 is any person or group with an interest in a business and how well it does.



Internal and external stakeholders all have an interest in the business

Internal stakeholders 内部利益相关者 are inside the business:

- **owners** and shareholders — want profit and a good return.
- **managers** 管理者 — want growth, job security and bonuses.
- employees — want fair pay, safe work and job security.

External stakeholders 外部利益相关者 are outside the business:

Stakeholder	Main objective
customers 顾客	good quality at a fair price
suppliers 供应商	regular orders, paid on time
lenders 贷款方 (banks)	to be repaid with interest
the government	tax, jobs, and that the firm obeys the law
local community	jobs, but little pollution or noise
pressure groups 压力团体	to protect a cause, such as the environment

Why stakeholder objectives conflict

Stakeholders often want different things, so their objectives can **conflict** 冲突 (clash). For example:

- owners want higher profit, but employees want higher pay — and higher pay lowers profit.
- customers want low prices, but owners want a high profit margin.
- the local community wants less pollution, but cutting pollution can raise costs.

How conflict can be managed

- keep good communication with every group.
- **compromise** 妥协 — give each group part of what it wants.
- **negotiation** 谈判 — talk until the groups reach an agreement.
- decide which stakeholders matter most for each decision, and act on that.

Exam tips

- Define **added value** = selling price − cost of bought-in inputs, and give a way to raise it (branding, quality, design).
- Distinguish **unlimited vs limited liability** and link it to sole trader/partnership versus Ltd/plc.
- Measure business **size** in more than one way (employees, revenue, capital employed, market share) — note each has limits.
- Distinguish **objectives** (SMART, specific) from a **mission** (overall purpose).
- In evaluation, use **stakeholder conflicts** (owners vs employees vs community) to argue both sides.

Human resource management

A-Level Business

What human resource management does

Human resource management 人力资源管理 (HRM) is the part of a business that looks after its people. Good HRM makes sure the firm has enough staff, with the right skills, who are well trained and motivated.



Managers plan and organise the workforce — the core of human resource management.

The main roles of HRM are:

- **workforce planning** 人力规划 — working out how many staff, and what skills, the business will need in the future.
- **recruitment** 招聘 and **selection** 甄选 — finding people to apply for jobs, then choosing the best.
- **training** 培训 and **development** 培养 — improving staff skills now and helping them grow for the future.

Workforce planning, recruitment and selection

Workforce planning looks at the firm's plans, then at the staff it already has, to find the gap. The gap may be filled by hiring, training, or moving staff.



Recruitment: describe the job, advertise, then select the best applicant

To recruit, the firm first writes:

- a **job description** 职位描述 — the title, duties and tasks of the job.
- a **person specification** 人员规格 — the skills, **qualifications** 资格 and qualities the right person should have.

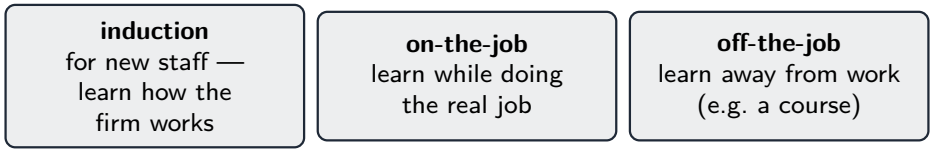
The firm can recruit from inside or outside:

- **internal recruitment** 内部招聘 — filling the job with someone who already works there. It is cheaper and faster, and the person is known, but no new ideas come in.
- **external recruitment** 外部招聘 — hiring someone from outside. It brings new skills and ideas, but costs more and is slower.

Selection methods include reading application forms and CVs, **interviews** 面试, tests, and group tasks.

Training and development

Training improves the skills staff need to do their jobs well. Common types are:



Three kinds of training: induction, on-the-job and off-the-job

- **induction training** 入职培训 — given to new staff so they learn how the firm works.
- **on-the-job training** 在岗培训 — learning while doing the real job.

- **off-the-job training** 脱产培训 — learning away from the workplace, such as a course.

Training costs time and money, but it raises quality, **productivity** 生产率 and motivation, and helps keep staff.

Ending employment

Sometimes a worker has to leave.

- **redundancy** 裁员 happens when the job itself is no longer needed — for example, when a machine replaces it. It is not the worker's fault.
- **dismissal** 解雇 happens when a worker is made to leave because of poor work or bad behaviour.

Rights and managing change

Both sides have rights and duties. The **employer** 雇主 must follow the law on pay, safety, working hours and unfair dismissal. The **employee** 雇员 must work as agreed and follow safe, fair rules. These **rights** 权利 protect both sides.

Businesses change often — new technology, new markets, or **restructuring** 重组. The **management of change** 变革管理 in HR means planning the change, explaining it clearly, training staff for new roles, and listening to worries, so staff accept the change instead of resisting it.

Measuring the workforce

Two simple measures help managers judge how the workforce is doing.

Labour turnover 员工流失率 shows the percentage of staff who leave in a year:

$$\text{labour turnover} = \frac{\text{number of staff leaving}}{\text{average number employed}} \times 100\%$$

High labour turnover can mean low pay, poor motivation or weak management. It raises the cost of recruiting and training.

Worked example. A firm employs 200 people on average, and 30 leave during the year. Find its labour turnover.

$$\text{labour turnover} = \frac{30}{200} \times 100\% = 15\%.$$

Labour productivity 劳动生产率 shows the output made by each worker:

$$\text{labour productivity} = \frac{\text{total output}}{\text{number of employees}}$$

Higher labour productivity lowers the cost of each unit, so the firm can compete better.

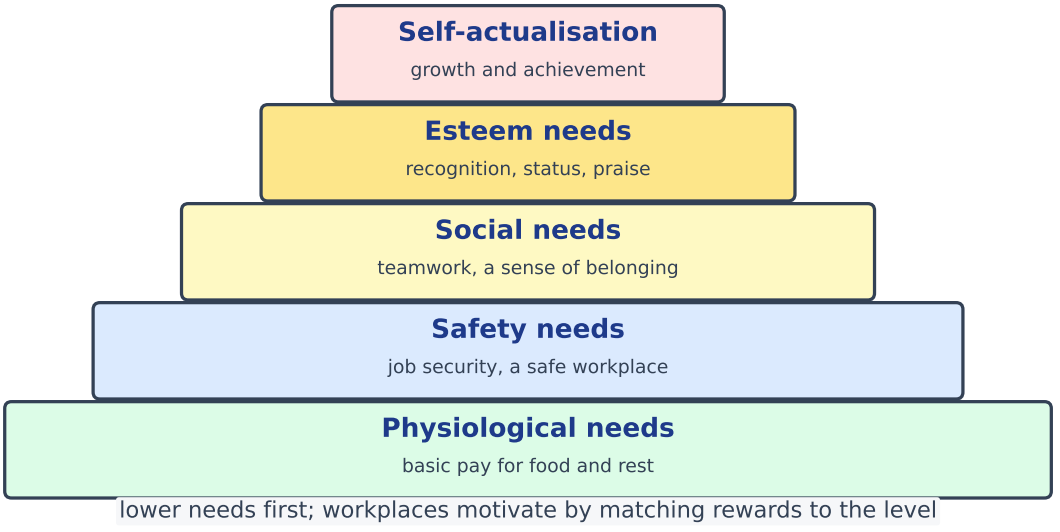
Worked example. A team of 8 workers makes 2400 units in a week. Find the labour productivity.

$$\text{labour productivity} = \frac{2400}{8} = 300 \text{ units per worker per week.}$$

Why motivation matters

Motivation 激励 is the desire to work hard and do a good job. Motivated staff produce more and better work, take fewer days off, stay longer, and give better service. So motivation lowers costs and raises quality.

Theories of motivation



Maslow's hierarchy of needs, with workplace examples at each level.

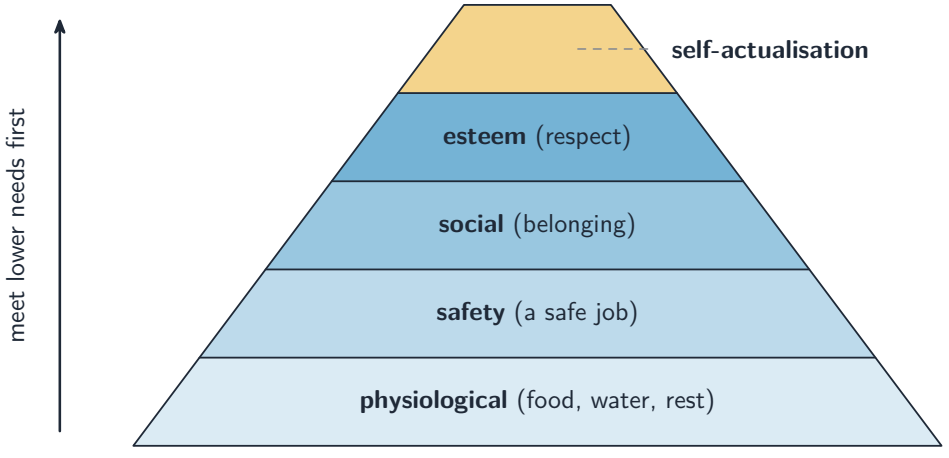
Several thinkers tried to explain what motivates people. You should know the main idea of each.

Thinker	Main idea
Taylor	**scientific management** 科学管理 — workers are mainly motivated by money. Pay them per unit made (**piece rate** 计件工资) and they work harder.
Mayo	**human relations** 人际关系 — workers are also motivated by attention, teamwork and feeling part of a group.
Maslow	**hierarchy of needs** 需求层次 — people meet needs in order, from low to high (see below).
Herzberg	**two-factor theory** 双因素理论 — some things cause unhappiness if missing, others truly motivate (see below).
McClelland	people are driven by three needs in different amounts: **achievement** 成就, power and affiliation (belonging).
Vroom	**expectancy theory** 期望理论 — people work hard only if they believe the effort will lead to a reward they want.

Maslow's hierarchy of needs

Maslow said people try to meet lower needs first, then move up:

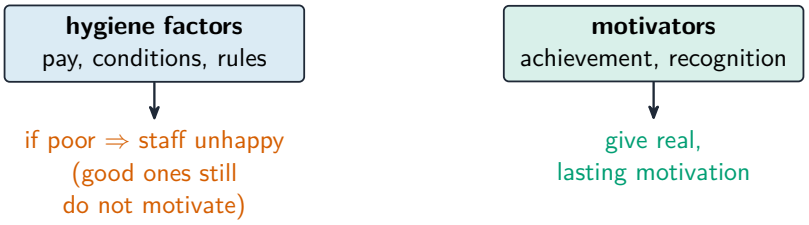
1. **physiological needs** 生理需求 — food, water, rest (met by pay).
2. safety needs — a safe job and workplace.
3. social needs — friendship and being part of a team.
4. **esteem needs** 尊重需求 — respect and **recognition** 认可.
5. **self-actualisation** 自我实现 — reaching your full potential.



Maslow's hierarchy: people meet lower needs before higher ones

Herzberg's two factors

- **hygiene factors** 保健因素 — things like pay, conditions and rules. If they are poor, staff are unhappy; but fixing them does not truly motivate.
- **motivators** 激励因素 — things like achievement, responsibility and recognition. These give real, lasting motivation.



Hygiene factors stop unhappiness; only motivators give real, lasting motivation

Financial motivators

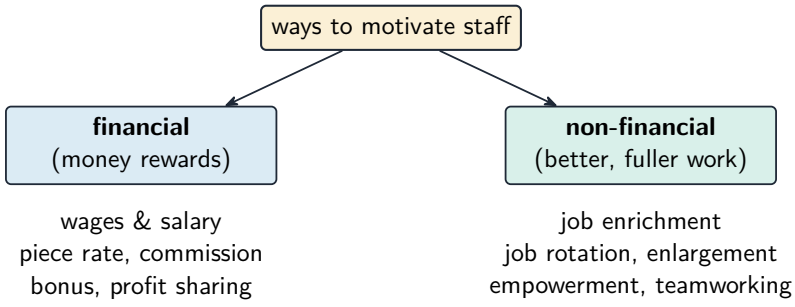
Financial motivators 经济激励 are rewards paid in money:

- **wages** 工资 — pay by the hour or week, often for **manual work** 体力劳动.
- **salary** 薪水 — a fixed yearly amount, paid monthly.
- piece rate — pay for each unit made.
- **commission** 佣金 — pay based on how much you sell.
- **bonus** 奖金 — an extra payment for good results.
- **profit sharing** 利润分享 — staff get a share of the firm's profit.
- **performance-related pay** 绩效工资 — extra pay for those who meet targets.
- **fringe benefits** 附加福利 — non-cash extras, such as a company car or health care.

Non-financial motivators

Non-financial motivators 非经济激励 reward staff without extra money, by making work more interesting and giving more control:

- **job enrichment** 工作丰富化 — giving more challenging tasks and responsibility.
- **job rotation** 工作轮换 — moving between different tasks to reduce boredom.
- **job enlargement** 工作扩大化 — adding more tasks at the same level.
- **empowerment** 授权 — letting staff make their own decisions about their work.
- **teamworking** 团队合作 — organising staff into teams that share goals.



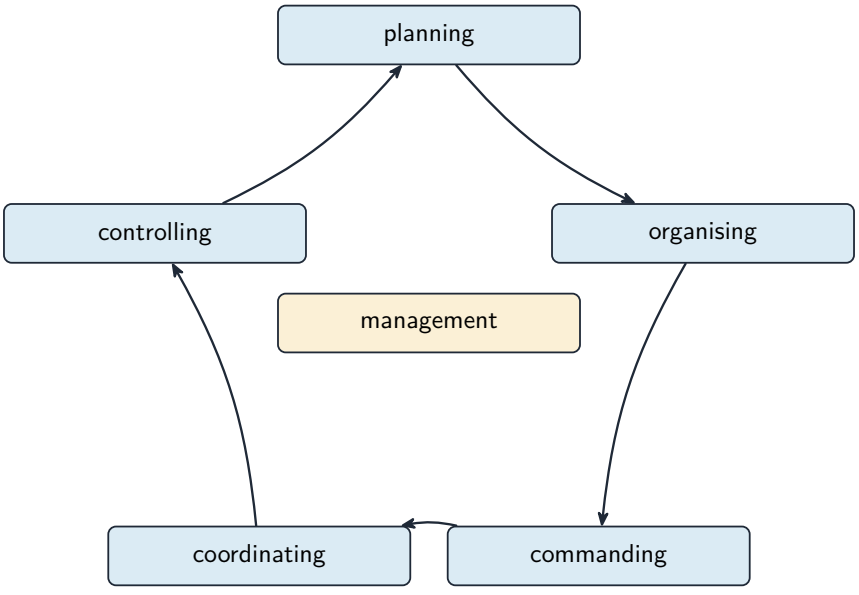
Motivators are either financial (money) or non-financial (better, fuller work)

Management and its functions

Management 管理 means getting work done through other people to reach the firm's objectives.

Fayol said managers carry out five functions:

Function	What the manager does
planning 计划	set objectives and decide how to reach them
organising 组织	arrange people and resources
commanding 指挥	give clear instructions and guidance
coordinating 协调	make sure all parts work together
controlling 控制	check results against the plan and correct problems



Fayol's five functions of management form a repeating cycle

Mintzberg said managers also play ten **roles** 角色 in three groups: interpersonal (working with people), informational (sharing information), and decisional (making decisions). This shows that real management is busy and varied, not just the five neat functions.

Leadership versus management

Leadership 领导 and management are not the same.

- a manager plans, organises and controls the day-to-day work.
- a leader sets a **vision** 愿景, inspires people, and guides change.



A manager runs the day-to-day work; a leader sets direction and inspires — the best people do both

The best managers are also good leaders. Effective management matters because it raises productivity, keeps good staff, helps the firm adapt to change, and turns plans into real results.

Exam tips

- Calculate and interpret **labour turnover** and **labour productivity**; high turnover raises recruitment and training costs.
- Link a **motivation theory** to a method: Maslow (needs), Herzberg (hygiene vs motivators), Taylor (pay).
- Distinguish **financial** (piece rate, bonus) from **non-financial** (job enrichment, empowerment) motivators.
- Distinguish **leadership from management** and match a style (autocratic, democratic, laissez-faire) to the situation.

Marketing

A-Level Business

The role of marketing

Marketing 市场营销 is the job of finding out what customers want, then making and selling products that meet those wants at a profit. Good marketing links the business to its **market** 市场 (all the buyers and sellers of a product).

A business can be:

- **product-orientated** 产品导向 — it focuses on the product first and hopes customers will buy it. This suits new or technical products.
- **market-orientated** 市场导向 — it finds out what customers want first, then makes it. This lowers the risk of failure.

Customer relationship marketing 客户关系营销 (CRM) means building long-term links with customers so they come back again. Keeping a customer is cheaper than finding a new one.

Market size, growth and share

- **market size** 市场规模 — the total sales of all firms in a market, by value or by number of units.
- **market growth** 市场增长 — how fast the market is getting bigger, shown as a percentage.
- **market share** 市场份额 — one firm's sales as a percentage of the total market:

$$\text{market share} = \frac{\text{firm's sales}}{\text{total market sales}} \times 100\%$$

A rising market share is a sign of strong marketing.

Worked example. A firm sells £4 million of goods in a market worth £50 million in total. Find its market share.

$$\text{market share} = \frac{4}{50} \times 100\% = 8\%.$$

If the whole market grows but the firm's sales stay at £4 million, its market share **falls** — so a firm can sell the same amount yet lose ground to faster-growing rivals.

Mass and niche markets

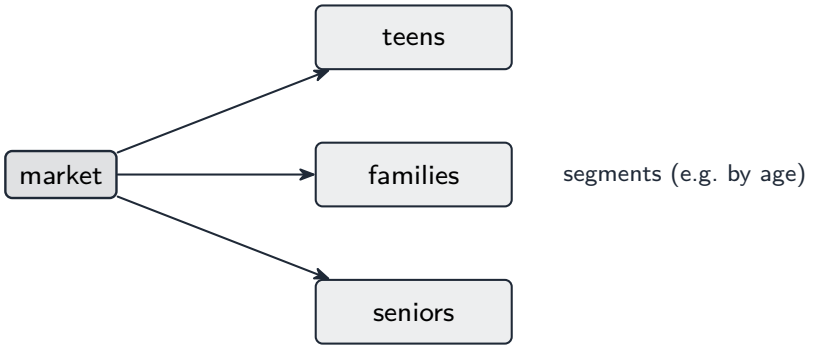
- a **mass market** 大众市场 is large, with products aimed at most buyers (e.g. soft drinks). Sales are high, but competition is strong.
- a **niche market** 利基市场 is a small part of a market with special needs (e.g. left-handed tools). Competition is weaker, but sales are smaller.



A mass market targets everyone; a niche serves a small specialist group

Market segmentation

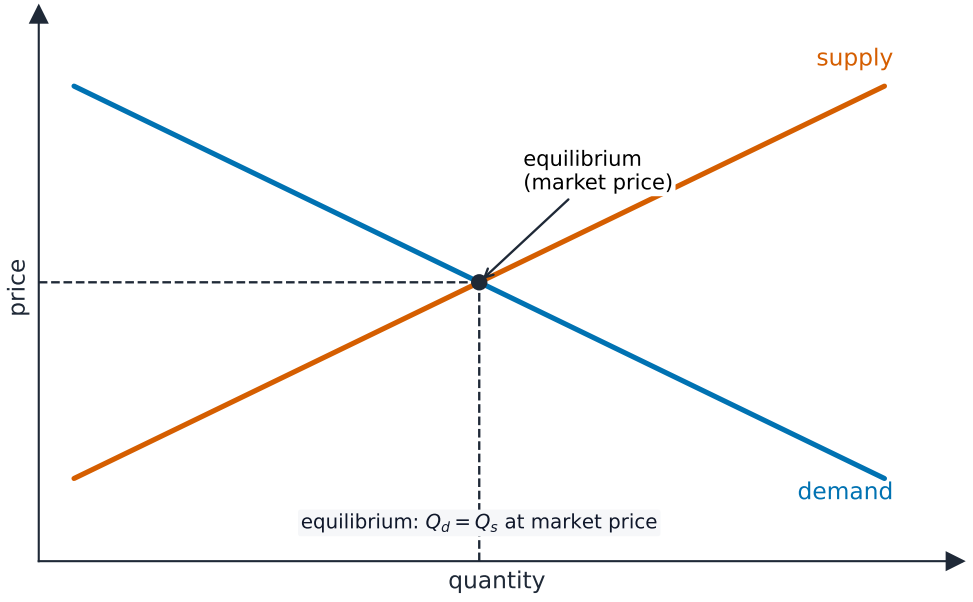
Market segmentation 市场细分 means splitting a market into groups (**segments** 细分市场) of buyers who are alike. Common ways to segment are by age, income, gender, location, and lifestyle. Segmentation lets a firm aim the right product and message at each group, which raises sales and cuts waste.



Splitting a market into segments by age, income and lifestyle

Demand, supply and elasticity

Demand 需求 is how much of a product buyers will buy at each price. **Supply** 供给 is how much sellers will offer at each price. Usually, when price falls, demand rises.



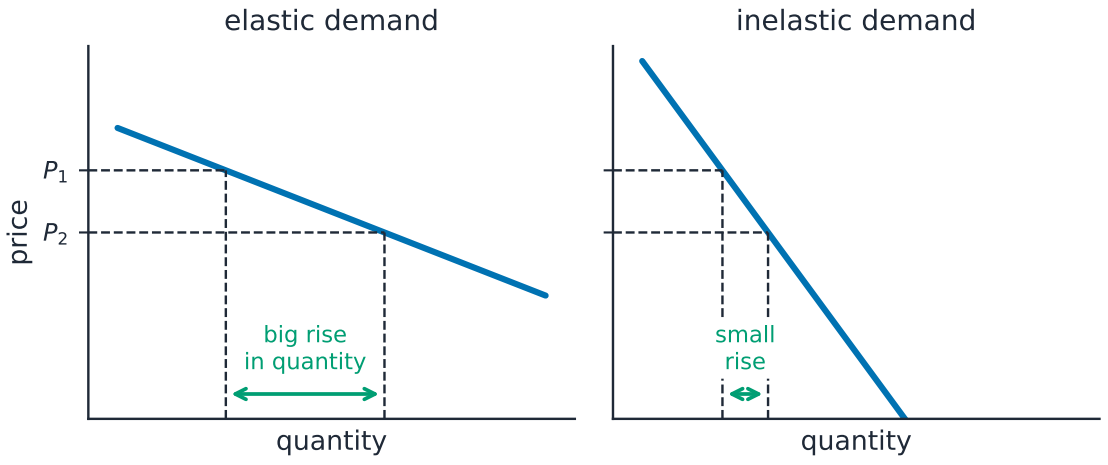
Supply and demand meet at the equilibrium market price

Elasticity 弹性 measures how strongly demand reacts to a change. You should know three types.

Price elasticity of demand 需求价格弹性 (PED) measures how demand reacts to a price change:

$$\text{PED} = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in price}}$$

- if PED is greater than 1, demand is **price elastic** 富有弹性 — a small price cut raises sales a lot, so lowering price can raise revenue.
- if PED is less than 1, demand is **price inelastic** 缺乏弹性 — demand changes little, so raising price can raise revenue.



$$\text{PED} = \frac{\%\,\Delta Q_d}{\%\,\Delta P} \quad \cdot \quad |PED| > 1 \text{ elastic} \quad \cdot \quad |PED| < 1 \text{ inelastic}$$

The same price fall raises quantity a lot when demand is elastic, only a little when inelastic

Income elasticity of demand 需求收入弹性 (YED) measures how demand reacts to a change in customers' income. Luxury goods react strongly; basic goods react weakly.

Cross elasticity of demand 需求交叉弹性 (XED) measures how the demand for one product reacts to a price change of another. It is positive for **substitutes** 替代品 (e.g. tea and coffee) and negative for **complements** 互补品 (e.g. cars and fuel).

Elasticity is useful because it helps a firm predict how a price change will affect sales and revenue.

Market research

Market research 市场调研 is collecting and studying information about customers, **competitors** 竞争对手 and the market.

- **primary research** 一手调研 — new information you collect yourself, by surveys, interviews or observation. It fits your exact needs, but costs more and takes time.
- **secondary research** 二手调研 — information that already exists, such as reports, websites and sales records. It is cheap and quick, but may be old or not exactly right.

Data can be:

- **quantitative data** 定量数据 — facts in numbers (e.g. "60% buy weekly"). Easy to compare.

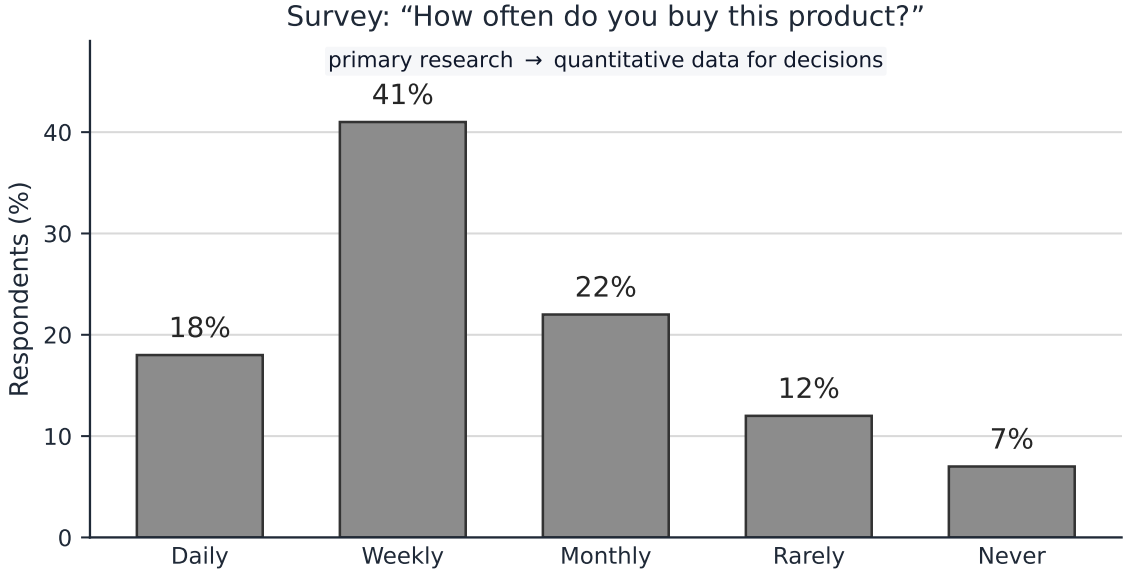
- **qualitative data** 定性数据 — opinions and reasons (e.g. "why people like the brand"). Gives deeper understanding.

Sampling and reliability

A **sample** 样本 is the small group of people asked, chosen to stand for the whole market. **Sampling** 抽样 methods include:

- **random sampling** 随机抽样 — everyone has an equal chance of being picked.
- **quota sampling** 配额抽样 — set numbers from each group are asked.
- **stratified sampling** 分层抽样 — the sample copies the make-up of the whole market.

A larger, well-chosen sample gives more **reliable** 可靠的 results. Poor sampling can cause **bias** 偏差, which makes the findings wrong. Results are often shown in tables and charts, then read carefully to guide decisions.



Primary research → quantitative data that guides marketing decisions

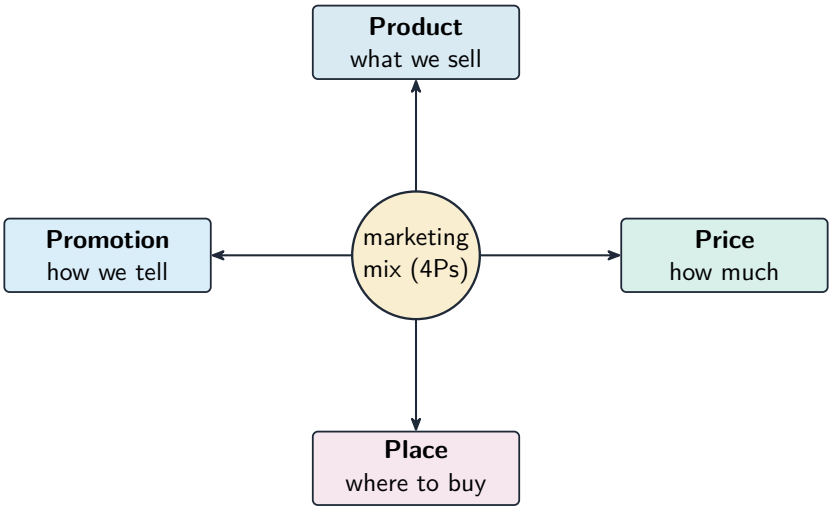
The marketing mix



A branded shop window — branding and place help sell the product.

The **marketing mix** 营销组合 is the set of choices a firm makes to sell a product. It is often called the **4Ps**:

P	Question it answers
product 产品	what are we selling?
price 价格	how much do we charge?
place 渠道	where and how do customers buy it?
promotion 促销	how do we tell customers about it?



The marketing mix combines the four Ps

The four parts must work together. For example, a high price needs high quality and a good image. Making the parts fit each other is called **integration** 整合.

Product

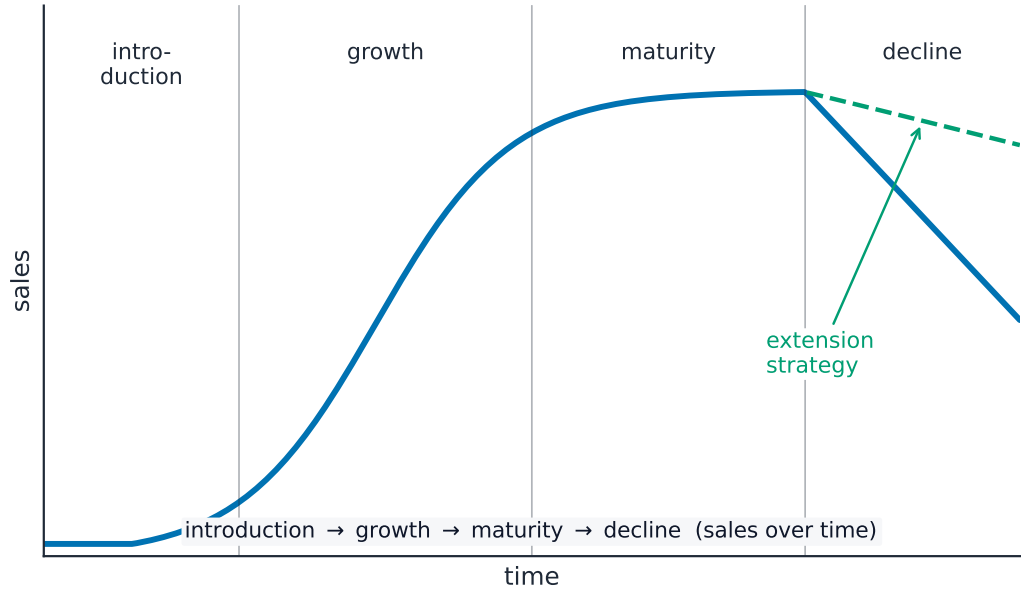
A good product is the heart of the mix.

Product development 产品开发 is the work of designing and improving products to meet customer needs and beat rivals.

The product life cycle

The **product life cycle** 产品生命周期 shows the sales of a product over time, in four stages:

1. **introduction** 引入期 — launch; low sales, high promotion cost.
2. **growth** 成长期 — sales rise quickly.
3. **maturity** 成熟期 — sales reach their peak and level off.
4. **decline** 衰退期 — sales fall as tastes change.



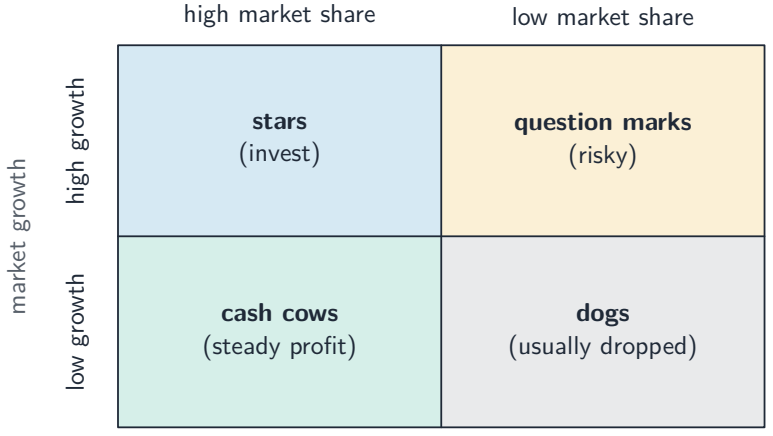
Sales of a product over its four life-cycle stages, with an extension strategy in maturity

To keep sales up in maturity, firms use **extension strategies** 延长策略, such as new packaging, new uses, or selling in new markets.

The product portfolio (Boston Matrix)

The **product portfolio** 产品组合 is the full range of products a firm sells. The **Boston Matrix** 波士顿矩阵 sorts products by market share and market growth:

	High market growth	**Low market growth**
High share	**stars** 明星产品	**cash cows** 现金牛
Low share	**question marks** 问题产品	**dogs** 瘦狗产品



The Boston Matrix sorts products by market share and market growth

Stars need investment but may become cash cows. Cash cows bring steady profit. Question marks are risky. Dogs are usually dropped.

Branding

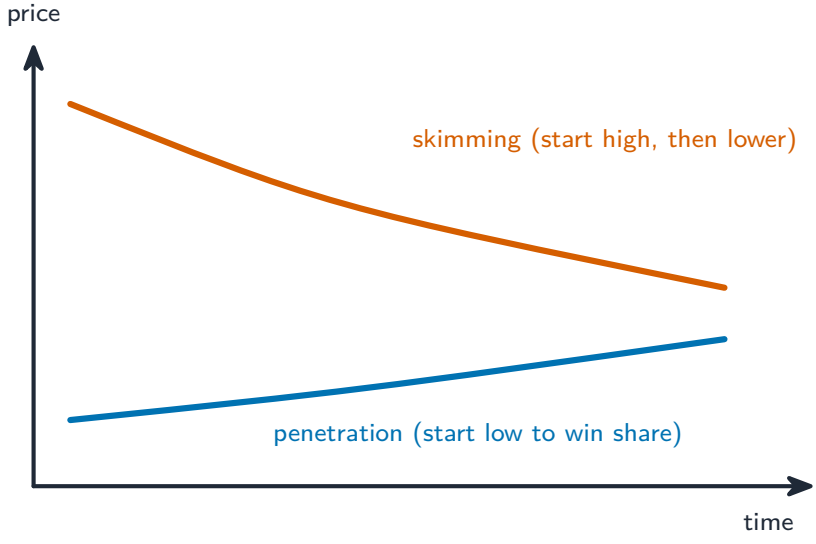
Branding 品牌 gives a product a name, design and image that customers recognise. A strong brand builds **loyalty** 忠诚度, lets the firm charge more, and makes new products easier to launch.

Price

Choosing a price is a key decision. The main methods and strategies are:

Method / strategy	How it works
cost-plus pricing 成本加成定价	add a fixed profit margin to the unit cost
penetration pricing 渗透定价	set a low price to enter a market and win share
price skimming 撇脂定价	set a high price first for a new, special product, then lower it
competitive pricing 竞争定价	set a price close to rivals' prices
price discrimination 价格歧视	charge different prices to different groups (e.g. peak vs off-peak)
dynamic pricing 动态定价	change the price often as demand changes
psychological pricing 心理定价	use prices like \$9.99 that feel lower

The best price depends on costs, competitors, the product's life-cycle stage, and how price elastic demand is.



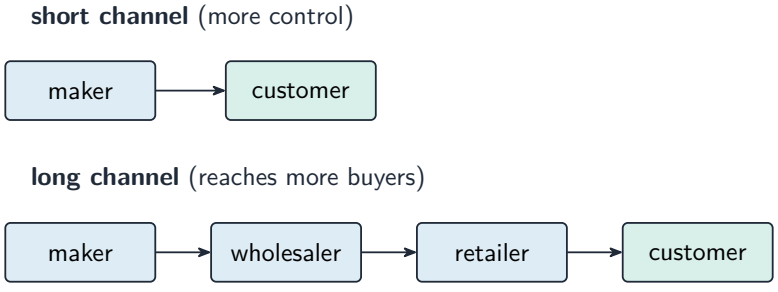
Two launch strategies over time: penetration sets a low price to win share; skimming sets a high price first, then lowers it

Place and promotion



Advertising on giant screens — promotion is part of the marketing mix.

Place is how the product reaches the customer, through a **distribution channel** 分销渠道. A short channel (maker → customer) gives more control; a longer channel (maker → **wholesaler** 批发商 → shop → customer) reaches more buyers.



A short channel gives more control; a longer channel reaches more buyers

Promotion tells customers about a product and persuades them to buy. It is split into:

- **above-the-line promotion** 大众媒体广告 — paid **advertising** 广告 in mass media, such as TV and online ads, aimed at a wide audience.
- **below-the-line promotion** 直接促销 — targeted methods the firm controls, such as discounts, free samples and direct mail.

Digital marketing 数字营销 and **e-commerce** 电子商务 (buying and selling online)

now play a huge role. They let small firms reach world markets cheaply, target adverts to the right people, and sell at any time of day.

Exam tips

- Calculate **market share** and **market growth** as percentages; a firm can grow yet lose share in a faster-growing market.
- Use **price elasticity of demand** to decide pricing: raise price when demand is inelastic (revenue rises), cut it when elastic.
- Distinguish **primary vs secondary** and **quantitative vs qualitative** research, and comment on sampling reliability.
- Treat the **marketing mix (4Ps)** as a coordinated set — each P should support the others.

Operations management

A-Level Business

What operations management does

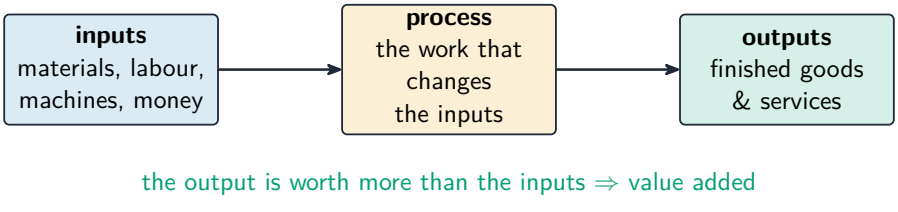
Operations 运营 is the part of a business that makes the product. **Operations management** 运营管理 is the job of turning **inputs** 投入 into **outputs** 产出 well.



A distribution centre: operations management turns inputs into finished goods ready for customers.

This is the **transformation process** 转化过程:

- **inputs** — raw materials, labour, machines and money.
- **process** — the work that changes the inputs.
- **outputs** — finished goods and services for the customer.



Operations turn inputs into outputs, adding value along the way

Good operations **add value** 增值: the output is worth more than the inputs used to

make it. Using **raw materials** 原材料 (the basic materials a business starts with) well is a big part of this.

Productivity and efficiency

Productivity 生产率 measures how much output you get from your inputs:

$$\text{productivity} = \frac{\text{output}}{\text{input used}}$$

Efficiency 效率 means making products with as little waste of time, materials and money as possible. Higher productivity and efficiency lower the cost of each unit, so the firm can charge less or earn more.

Worked example. A factory of 20 workers makes 800 units in a day. Find the output per worker.

$$\text{productivity} = \frac{800}{20} = 40 \text{ units per worker per day.}$$

If training raised output to 1000 units with the same 20 workers, productivity would rise to 50 units each — the same staff cost now spread over more units, so each unit is cheaper to make.



Assembly line: productivity is output per worker (or per hour) — rising productivity cuts unit costs

Capital- and labour-intensive operations

- a **capital-intensive** 资本密集型 operation uses mostly machines (e.g. a car factory). It has high set-up costs but low cost per unit, and is good for large amounts.
- a **labour-intensive** 劳动密集型 operation uses mostly people (e.g. a hair salon). It is flexible and good for personal service, but quality can vary.



Factory robots: capital-intensive production uses more machines relative to labour

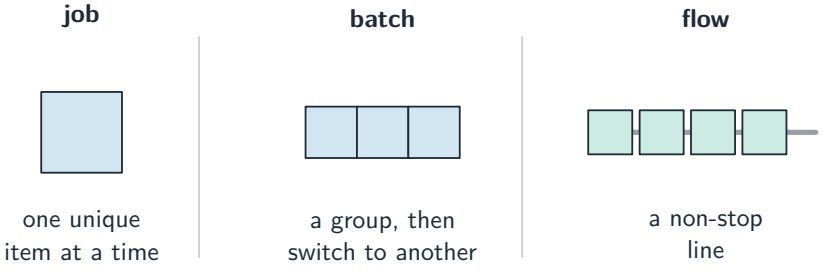
Methods of production

A firm chooses how to organise production to suit its product and demand.



A car assembly line: flow production of identical units, one stage after another

Method	What it is	Best for
job production 单件生产	making one unique item at a time	a wedding cake, a bridge
batch production 批量生产	making a group of the same item, then switching	bread in flavours
flow production 流水线生产	making items non-stop on a line	bottled drinks, cars
mass customisation 大规模定制	a flow line that still lets customers choose options	cars with chosen colours



The three methods of production

Job production is flexible but costly per unit. Flow production has a low unit cost but needs high, steady demand.

Sustainability and technology

Sustainability 可持续性 means meeting today's needs without harming the future or the planet. Sustainable operations cut waste, use less energy, recycle, and use materials that can be replaced.

Technology 技术 is widely used in operations. **Automation** 自动化 (machines and robots doing tasks) raises output, improves quality and lowers cost, but it has a high set-up cost and can replace jobs.

Inventory and why it matters

Inventory 库存 (stock) is the goods a business holds: raw materials, part-finished goods, and finished goods waiting to be sold.

A firm holds inventory so it can keep producing and meet customer orders quickly. But holding inventory has costs:

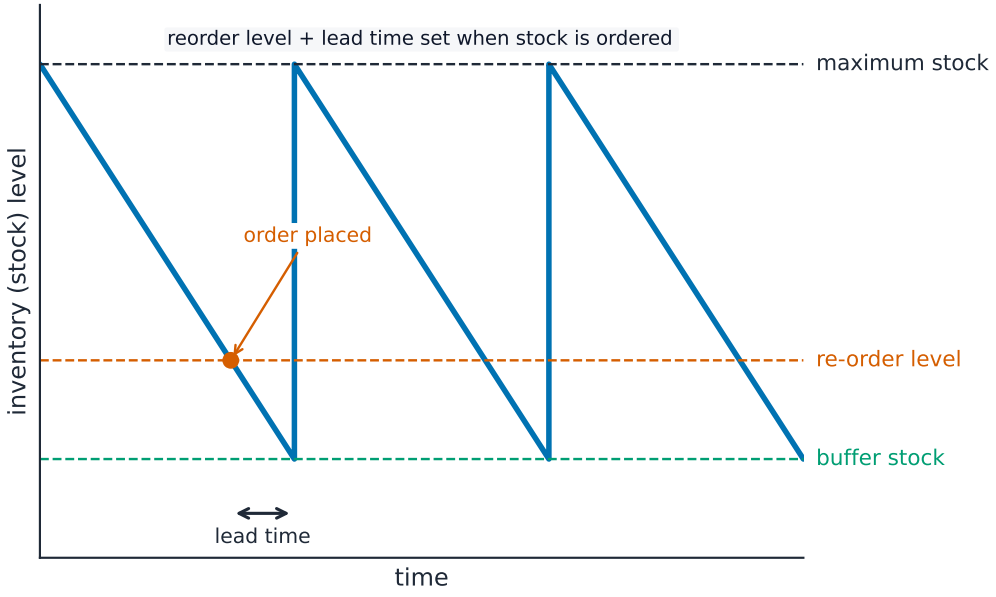
- **holding costs** 持有成本 — storage, insurance, and money tied up in stock.
- the risk that goods go out of date or are damaged.

Holding too little stock is also risky: the firm may run out and lose sales.

Inventory control charts

An **inventory control chart** 库存控制图 shows how stock rises when an order arrives and falls as it is used. Key terms:

- **re-order level** 再订货水平 — the stock level at which a new order is placed.
- **lead time** 交货周期 — the time between placing an order and receiving it.
- **buffer inventory** 缓冲库存 — a minimum "safety" stock kept in case of delays or extra demand.
- **maximum inventory** 最大库存 — the most stock the firm wants to hold.

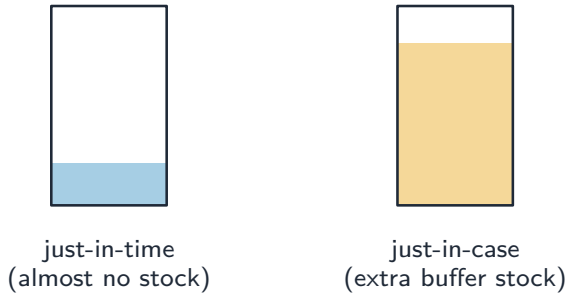


Stock is used up, then replenished when a new order arrives after the lead time

Just-in-time and just-in-case

There are two main ways to manage stock.

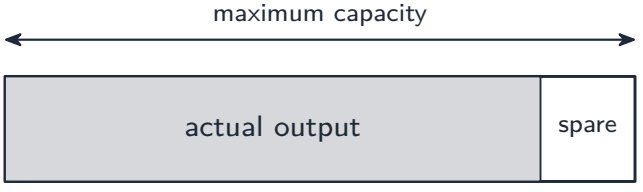
- **just-in-time** 准时制 (JIT) — stock arrives only as it is needed, so almost no inventory is held. This cuts holding costs and waste, but a late delivery can stop production.
- **just-in-case** 备货制 — extra stock is held in case of problems. This is safer but costs more to store.



Just-in-time holds almost no stock (low cost, but risky if a delivery is late); just-in-case keeps a buffer (safer, but more to store)

Capacity utilisation

Capacity 产能 is the most a business can produce with its current resources. **Capacity utilisation** 产能利用率 shows how much of that capacity is being used:



Capacity utilisation: how much of the maximum output is used

$$\text{capacity utilisation} = \frac{\text{actual output}}{\text{maximum possible output}} \times 100\%$$

- **under-utilisation** 产能利用不足 (a low figure) means resources sit idle, so the fixed cost spread over each unit is high.

- **over-utilisation** 产能过度利用 (very close to 100%) leaves no time for repairs, can tire staff, and may lower quality.

Most firms aim for a high but not full level, around 85–90%.

Worked example. A factory can make 5000 units a week but currently makes 4000. Find its capacity utilisation.

$$\text{capacity utilisation} = \frac{4000}{5000} \times 100\% = 80\%.$$

This leaves 20% spare. Winning an order for another 500 units a week would lift utilisation to 90%, spreading the fixed costs over more units without needing new machines.

Outsourcing

Outsourcing 外包 means paying another firm (a **supplier** 供应商) to do work the business used to do itself, such as cleaning, IT or making parts.

Benefits: it can be cheaper, lets the firm focus on what it does best, and adds flexibility. Drawbacks: less control over quality, possible delays, and the risk of sharing secrets with another firm.

Exam tips

- Calculate **capacity utilisation** and **productivity**; high utilisation spreads fixed costs but leaves no slack for breakdowns or new orders.
- Read an **inventory control chart** (re-order level, lead time, buffer stock, maximum stock).
- Compare **JIT vs JIC** (low holding cost but delivery risk vs security of supply but higher cost).
- Match the **production method** (job, batch, flow) to the product and the size of the order.

Finance and accounting

A-Level Business

Why a business needs finance

Business finance 企业资金 is the money a firm needs to start, run and grow. A new firm needs money to buy **premises** 经营场所 and equipment; a growing firm needs money for new projects; and every firm needs money to pay day-to-day bills.



Banks are a key external source of finance, through loans and overdrafts.

Capital and revenue expenditure

Spending falls into two types:

capital expenditure	revenue expenditure
fixed assets that last (machines, vehicles, buildings)	day-to-day running (wages, rent, materials)

Capital spending buys fixed assets; revenue spending covers running costs

- **capital expenditure** 资本性支出 — money spent on **fixed assets** 固定资产 that last a long time, such as machines, vehicles and buildings.

- **revenue expenditure** 收益性支出 — money spent on day-to-day running, such as wages, rent and raw materials.

Working capital

Working capital 营运资本 is the money a firm has for its daily needs. It is found from the **balance sheet** 资产负债表:

$$\text{working capital} = \text{current assets} - \text{current liabilities}$$

Here **current assets** 流动资产 are things that become cash within a year (cash, inventory, money owed by customers), and **current liabilities** 流动负债 are debts due within a year. Too little working capital is dangerous — the firm may not be able to pay its bills.



Cash is working capital: firms need enough liquid assets to pay day-to-day bills

Cash is not the same as profit

This is a key idea.

- **profit** 利润 is what is left when total costs are taken from total revenue over a period.
- **cash** 现金 is the money the firm actually has right now to spend.

A firm can be **profitable** but still run out of cash — for example, if it sells goods but customers have not paid yet. Many firms fail not from low profit, but from running out of cash.

a firm can be profitable but still run out of cash

profit (on paper)
revenue − costs = positive

cash (in the bank)
low — customers
haven't paid yet

Profit and cash are different: a firm can look profitable on paper yet have no cash to pay its bills — which is why many fail

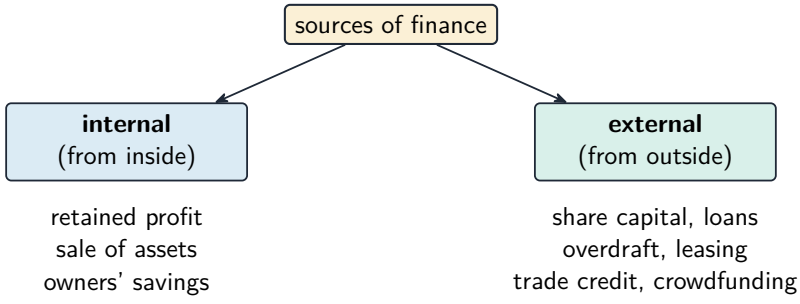
Sources of finance



A stock exchange: public limited companies raise finance by selling shares.

Finance can come from **inside** the business (**internal sources** 内部来源) or from **outside** (**external sources** 外部来源). It can also be **short-term** 短期 (paid back within a year) or **long-term** 长期.

Source	Type	What it is
retained profit 留存利润	internal, long	profit kept in the business instead of paid out
sale of assets 出售资产	internal	selling items the firm no longer needs
share capital 股本	external, long	money raised by selling shares
loan 贷款	external, long	borrowed money repaid with interest
debenture 公司债券	external, long	a long-term loan certificate sold by a company
overdraft 透支	external, short	the bank lets the account go below zero for a short time
leasing 租赁	external	renting an asset instead of buying it
hire purchase 分期付款	external	buying an asset by paying in instalments
trade credit 商业信用	external, short	paying a supplier later (e.g. after 30 days)
venture capital 风险资本	external, long	money from investors into risky young firms
crowdfunding 众筹	external	small amounts raised from many people online
microfinance 小额信贷	external	small loans for people with little access to banks

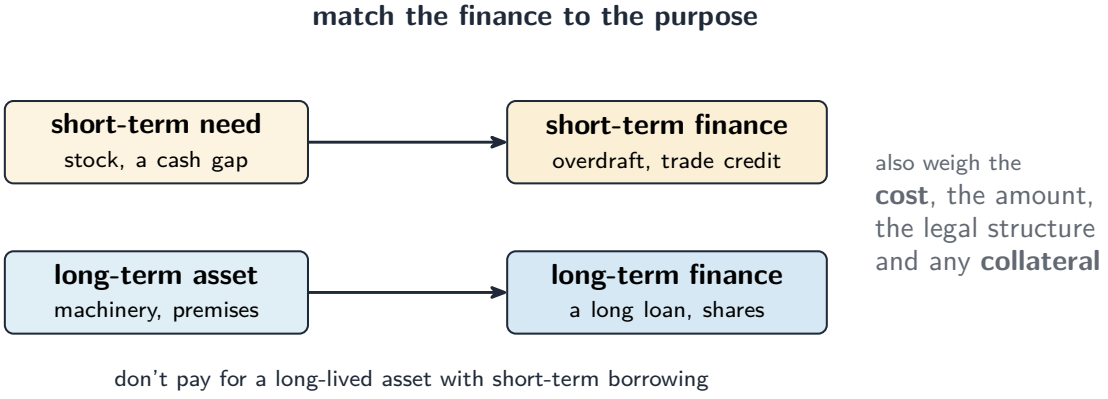


Sources of finance come from inside the business (internal) or outside it (external)

Choosing a source of finance

There is no single best source. The choice depends on:

- the **amount** needed — large amounts may need shares or a long loan.
- the **purpose** — long-term assets should use long-term finance.
- the **cost** — interest and fees differ between sources.
- the **legal structure** — only a company can sell shares.
- how much the firm already owes, and whether it can offer **collateral** 抵押品 (an asset the lender can take if the loan is not repaid).



Match the finance to the purpose: short-term needs use short-term finance, long-term assets use long-term finance

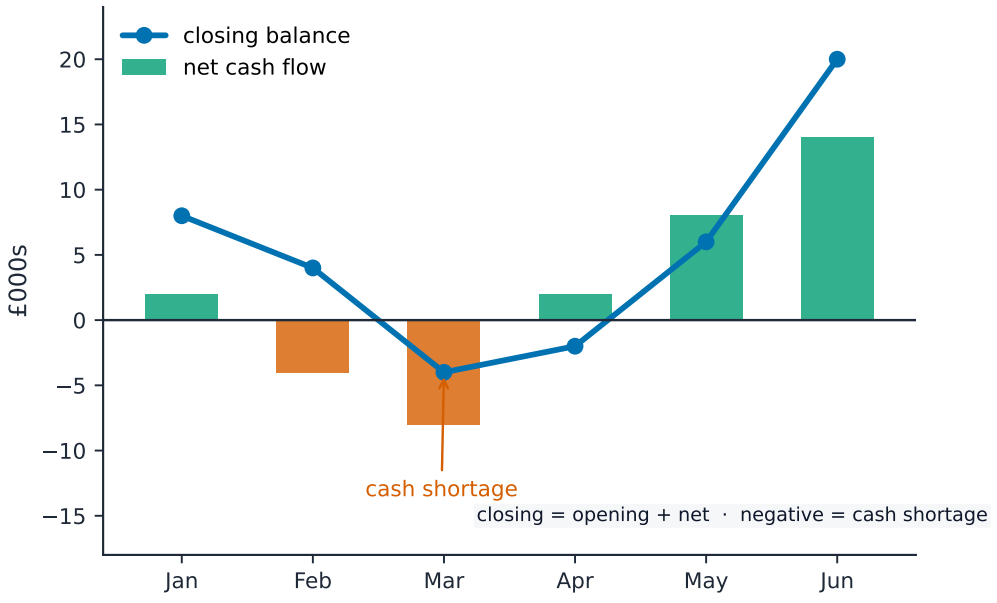
Cash-flow forecasts

A **cash-flow forecast** 现金流量预测 predicts the cash coming in and going out each month. It helps a firm spot a shortage early.

- **cash inflows** 现金流入 — money coming in, mostly from sales.
- **cash outflows** 现金流出 — money going out, such as wages and rent.
- **net cash flow** 净现金流 — inflows minus outflows for the period.

$$\text{net cash flow} = \text{cash inflows} - \text{cash outflows}$$

The **opening balance** 期初余额 is the cash at the start of the month. Adding net cash flow gives the **closing balance** 期末余额, which becomes next month's opening balance.



A cash-flow forecast tracks the closing balance each month, warning of a shortage early

Cash-flow problems and how to fix them

Common causes of cash-flow problems are: too many sales on credit, holding too much stock, buying too many assets at once, and **overtrading** 过度交易 (growing too fast for the cash available).

Ways to improve cash flow:

- bring cash in sooner — ask customers to pay faster, or offer discounts for quick payment.
- delay cash going out — agree longer trade credit with suppliers.
- arrange an overdraft or short loan to cover a gap.
- hold less stock to free up cash.

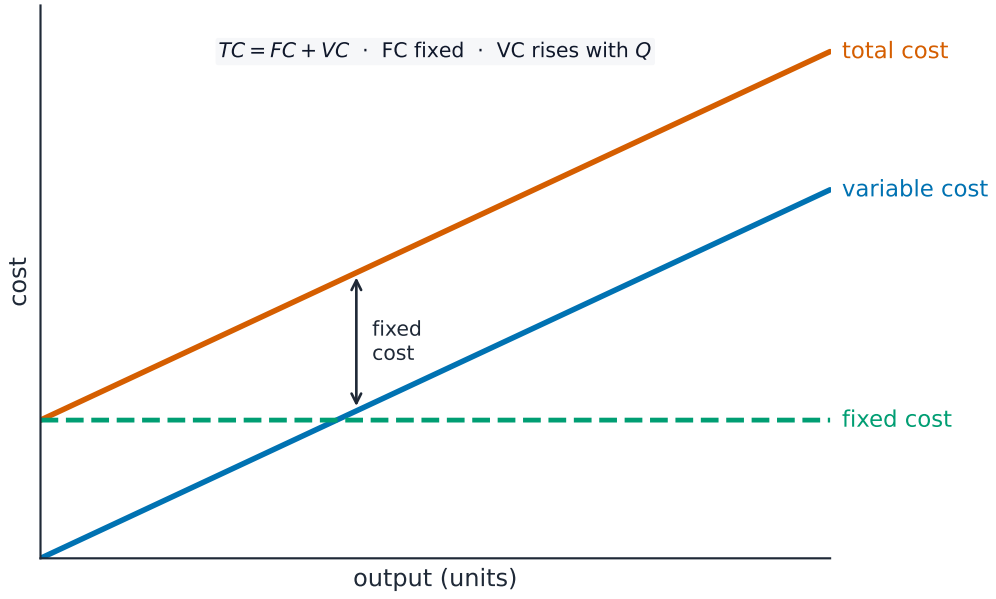
Good **working capital management** keeps enough cash to stay safe without holding so much that money sits idle.

Costs and how we classify them

A **cost** 成本 is money the business spends to make its product. We sort costs in several ways:

- **fixed costs** 固定成本 — costs that do not change with output, such as rent.

- **variable costs** 可变成本 — costs that rise and fall with output, such as raw materials.
- **direct costs** 直接成本 — costs clearly linked to one product, such as its materials.
- **indirect costs** 间接成本 (overheads) — costs not linked to one product, such as manager salaries.
- **marginal cost** 边际成本 — the cost of making one more unit.
- **average cost** 平均成本 — the cost per unit, found by total cost ÷ number of units.



Total cost = fixed cost + variable cost; only variable cost rises with output

Cost information helps managers set prices, choose what to make, and control spending.

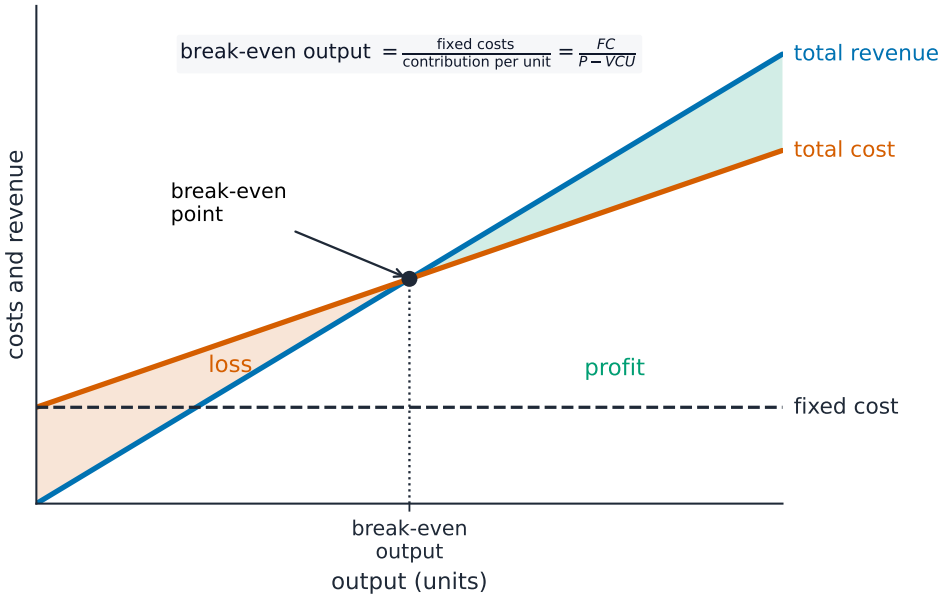
Break-even analysis

Contribution 贡献 is how much each sale adds towards paying the fixed costs:

$$\text{contribution per unit} = \text{selling price} - \text{variable cost per unit}$$

Break-even 盈亏平衡 is the point where total revenue equals total costs, so profit is zero. The **break-even point** 盈亏平衡点 in units is:

$$\text{break-even point} = \frac{\text{fixed costs}}{\text{contribution per unit}}$$



The break-even point is where total revenue equals total cost — left of it is loss, right of it is profit

The **margin of safety** 安全边际 is how far current sales are above the break-even point:

$$\text{margin of safety} = \text{actual output} - \text{break-even output}$$

Worked example. A product sells for £20 with a variable cost of £12 per unit. Fixed costs are £40,000, and the firm makes 6,000 units. Find the contribution per unit, the break-even output and the margin of safety.

$$\text{contribution per unit} = £20 - £12 = £8$$

$$\text{break-even point} = \frac{£40\,000}{£8} = 5\,000 \text{ units}$$

$$\text{margin of safety} = 6\,000 - 5\,000 = 1\,000 \text{ units}$$

A large margin of safety means sales can fall a lot before the firm makes a loss. Break-even analysis is quick and useful, but it has **limitations** 局限性: it assumes the selling price and costs stay the same, and that everything made is sold.

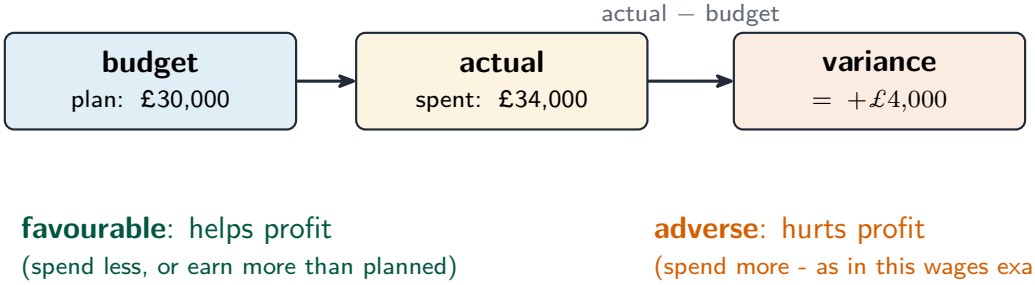
Budgets and variances

A **budget** 预算 is a financial plan for the future — a target for income or spending. Budgets help a firm plan, control money, and check performance. Two types are:

- **incremental budget** 增量预算 — last year's budget changed by a small amount.
- **zero-based budget** 零基预算 — every cost must be explained from zero each year.

A **variance** 差异 is the difference between the budgeted figure and the actual figure:

$$\text{variance} = \text{actual figure} - \text{budgeted figure}$$



a **budget** is a financial plan; the **variance** shows how far the actual result missed it

A variance is actual minus budget; favourable helps profit, adverse hurts it

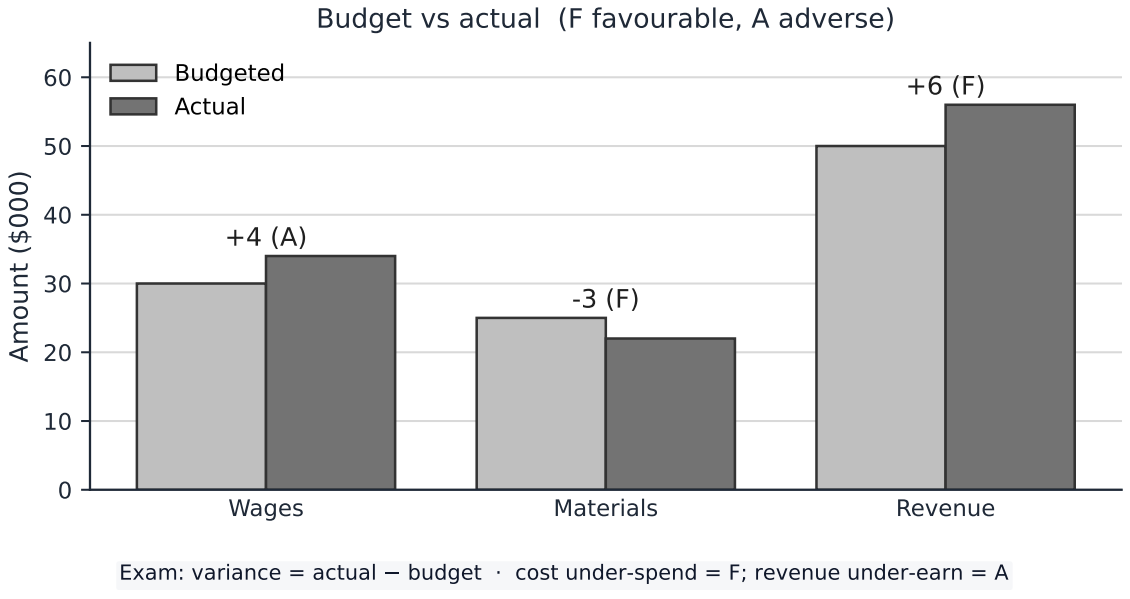
Worked example. A firm budgeted £30,000 for wages but actually spent £34,000. Find the variance and say whether it is favourable or adverse.

$$\text{variance} = £34\,000 - £30\,000 = £4\,000$$

Costs came in higher than planned, so this is an adverse variance.

- a **favourable variance** 有利差异 is better for profit (costs lower, or revenue higher, than planned).
- an **adverse variance** 不利差异 is worse for profit (costs higher, or revenue lower, than planned).

Studying variances shows managers where the plan went wrong, so they can act.



Variance = actual − budget: cost overspend is adverse; revenue over-earn is favourable

Exam tips

- Show that **cash is not profit** — a profitable firm can still fail from a cash-flow shortage.
- For **break-even**: contribution per unit = price − variable cost; break-even = fixed costs ÷ contribution; margin of safety = actual − break-even output.
- Match a **source of finance** to the need (short-term for cash flow, long-term or equity for assets) and to the type of business.
- Calculate a **variance** and label it **favourable or adverse** (actual against budget).

Pure Mathematics 1

A-Level Mathematics

This handout covers Topic 1: **Pure Mathematics** 纯数学 1. It is the algebra 代数 and calculus 微积分 core of the course. Each `##` section is one syllabus subtopic.

Quadratics



A suspension bridge: the main cable hangs in a parabola.

A **quadratic** 二次式 is an expression of the form $ax^2 + bx + c$, where $a \neq 0$. The letters a, b, c are the **coefficients** 系数 (the fixed numbers). Much of this section is about solving the **equation** $ax^2 + bx + c = 0$.

Completing the square

To **complete the square** 配方 means to write the quadratic in the form

$$a(x + p)^2 + q.$$

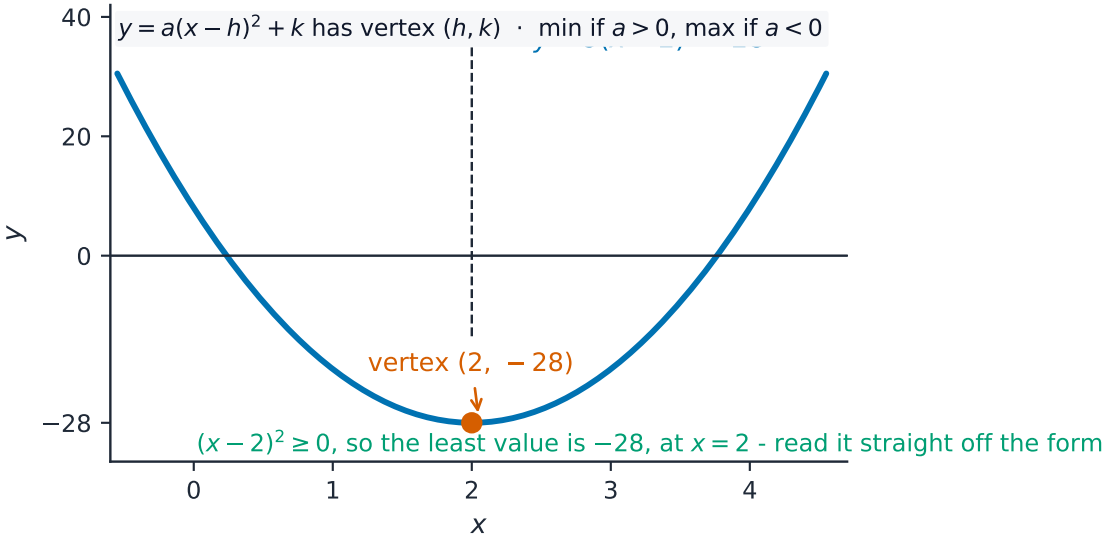
This form is useful: it shows the **vertex** 顶点 (turning point) of the curve at $(-p,\ q)$, and it gives a quick way to solve the equation.

Worked example. Write $9x^2 - 36x + 8$ in the form $p(x + q)^2 + r$.

Take the factor 9 out of the first two terms, then complete the square inside:

$$\begin{aligned} 9x^2 - 36x + 8 &= 9\left(x^2 - 4x\right) + 8 \\ &= 9\left((x-2)^2 - 4\right) + 8 \\ &= 9(x-2)^2 - 36 + 8 = 9(x-2)^2 - 28. \end{aligned}$$

So $p = 9$, $q = -2$, $r = -28$.



The completed square hands you the vertex: least value -28 at x equals 2

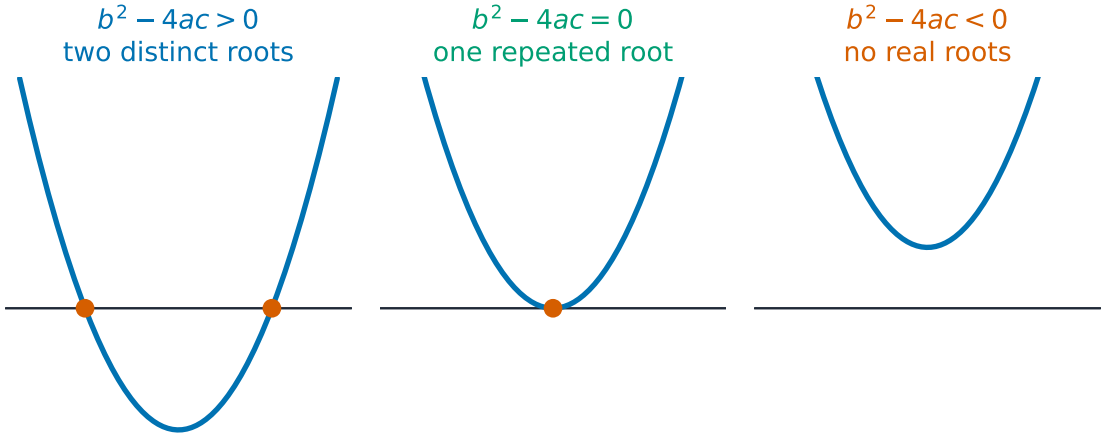
The discriminant

The **discriminant** 判别式 of $ax^2 + bx + c$ is

$$\Delta = b^2 - 4ac.$$

It tells you how many **real roots** 实根 (real solutions) the equation $ax^2 + bx + c = 0$ has:

Discriminant	Roots
$b^2 - 4ac > 0$	two **distinct** 相异 real roots
$b^2 - 4ac = 0$	one repeated real root
$b^2 - 4ac < 0$	no real roots



$$ax^2 + bx + c = 0 \;\Rightarrow\; x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (\text{real roots need } b^2 - 4ac \geq 0)$$

The sign of $b^2 - 4ac$ decides how many times the parabola meets the x-axis.

Worked example. Find the values of the constant k for which $3kx^2 + (k+8)x + 3 = 0$ has two distinct real roots.

Here $a = 3k$, $b = k + 8$, $c = 3$. For two distinct real roots you need $b^2 - 4ac > 0$:

$$(k+8)^2 - 4(3k)(3) > 0 \;\Rightarrow\; k^2 + 16k + 64 - 36k > 0 \;\Rightarrow\; k^2 - 20k + 64 > 0.$$

Factorise: $(k - 4)(k - 16) > 0$, so $k < 4$ or $k > 16$. You also need $a \neq 0$, so $k \neq 0$.

Quadratic equations and inequalities

To solve a **quadratic equation** 二次方程, factorise, complete the square, or use the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

To solve a **quadratic inequality** 二次不等式 such as $(k-4)(k-16) > 0$, find the two roots, then decide which side of each root makes the statement true. A sketch of the parabola 抛物线 helps: the curve is above the x-axis (positive) outside the roots and below it (negative) between them.

Simultaneous equations

To solve a pair of **simultaneous equations** 联立方程 where one is linear and one is quadratic, use **substitution** 代入: rearrange the linear equation for one letter, then put that into the quadratic. This gives a single quadratic to solve.

Equations that are quadratic in disguise

Some equations are quadratic in **some function of** x. For example $x^4 - 5x^2 + 4 = 0$ is quadratic in x^2: let $u = x^2$, solve $u^2 - 5u + 4 = 0$, then go back to x. You will use this idea again in trigonometry.

Functions

A **function** 函数 is a rule that sends each input to exactly one output. Write it as $f(x)$. The **composition of functions** 复合函数 $fg(x)$ means apply g first, then apply f to the result.

- The **domain** 定义域 is the set of allowed inputs x.
- The **range** 值域 is the set of outputs the function actually produces.

A function is **one-one** 一一对应 if different inputs always give different outputs. (No output is repeated.) Only a one-one function has an **inverse function** 反函数 f^{-1}, which reverses the rule.

The **composition** 复合 of two functions means doing one after the other. $fg(x)$ means "do g first, then f": $fg(x) = f(g(x))$. The composite function fg exists only when the range of g lies inside the domain of f.

Finding an inverse

To find f^{-1}: write $y = f(x)$, make x the subject, then swap letters.

Worked example. The function $f(x) = (x+3)^2 - 12$ is defined for $x \geqslant 0$. Find $f^{-1}(x)$.

Write $y = (x+3)^2 - 12$ and solve for x:

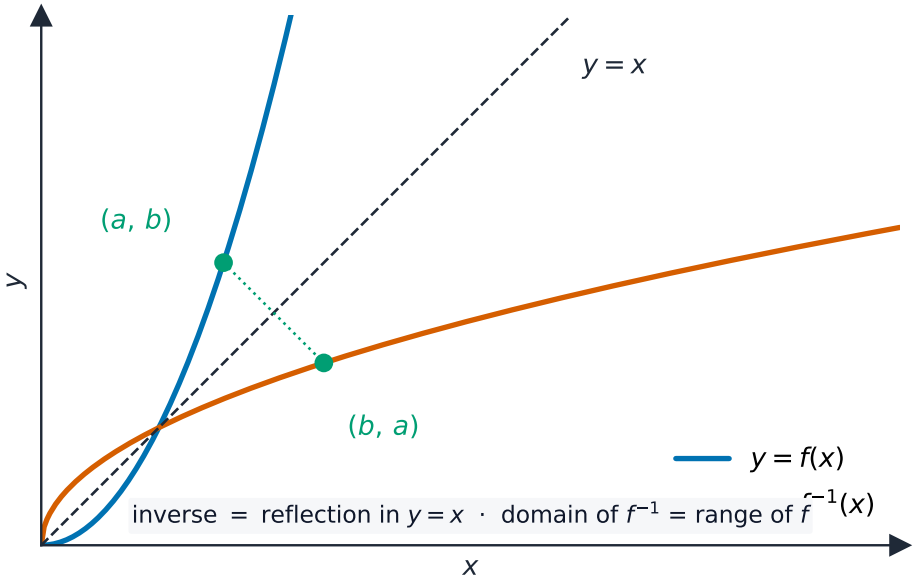
$$(x+3)^2 = y + 12 \;\Rightarrow\; x + 3 = \sqrt{y+12} \;\Rightarrow\; x = \sqrt{y+12} - 3.$$

You take the **positive** square root because $x \geqslant 0$ means $x + 3 \geqslant 3 > 0$. So

$$f^{-1}(x) = \sqrt{x+12} - 3.$$

Graphs of inverses and transformations

The graph of $y = f^{-1}(x)$ is the **reflection** 反射 of $y = f(x)$ in the line $y = x$.

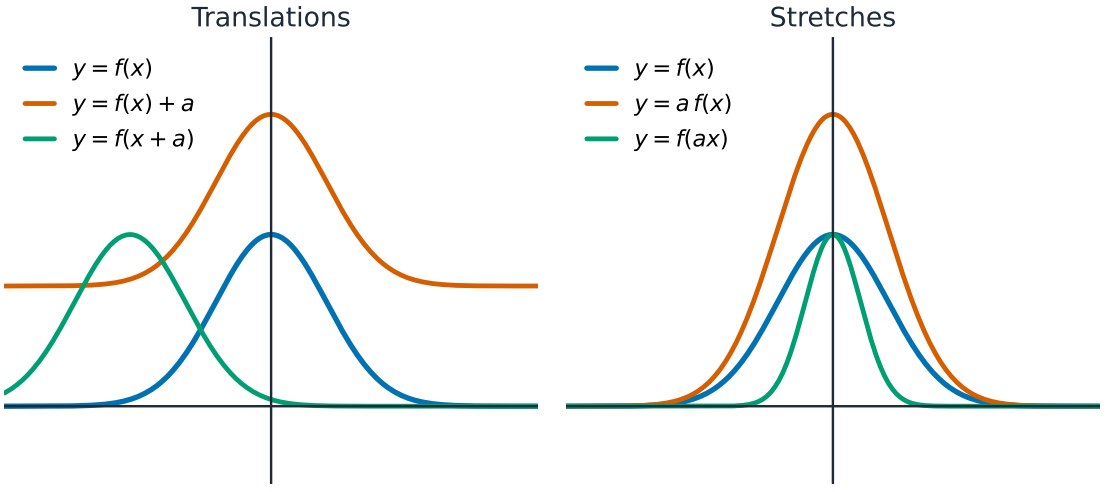


Reflecting $y = f(x)$ in the line $y = x$ gives its inverse; a point (a, b) becomes (b, a).

You should know these **transformations** 变换 of $y = f(x)$:

New equation	Effect on the graph
$y = f(x) + a$	**translation** 平移 up by a
$y = f(x + a)$	translation left by a
$y = a\,f(x)$	**stretch** 伸缩 in the y-direction, scale factor a
$y = f(ax)$	stretch in the x-direction, scale factor $\frac{1}{a}$

When two transformations are combined, the order can matter. State each one fully (type, direction, and amount).



$f(x)+a$ moves up · $f(x+a)$ moves left · $af(x)$ stretch y · $f(ax)$ compress x if $a>1$

Adding to the output or input slides the curve; a multiplier stretches it.

Coordinate geometry

Coordinate geometry 坐标几何 studies lines and circles using their equations.

Straight lines

The **gradient** 斜率 (steepness) of the line joining (x_1, y_1) and (x_2, y_2) is

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

You can write the **equation of a straight line** 直线方程 in any of these forms:

$$y = mx + c, \qquad y - y_1 = m(x - x_1), \qquad ax + by + c = 0.$$

Two lines are **parallel** 平行 when their gradients are equal, and **perpendicular** 垂直 (at right angles) when the product of their gradients is -1.

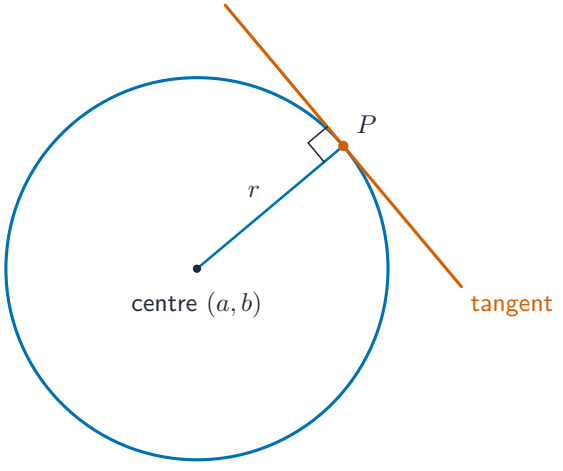
Circles

The **circle** 圆 with **centre** 圆心 (a, b) and **radius** 半径 r has equation

$$(x-a)^2 + (y-b)^2 = r^2.$$

An expanded form like $x^2 + y^2 - 6x + 10y - 27 = 0$ is the same circle: complete the square in x and in y to find the centre and radius.

A **tangent** 切线 to a circle touches it at one point and is perpendicular to the radius at that point. This right-angle fact solves most circle problems.



A tangent touches the circle once and meets the radius at a right angle.

Worked example. The points $P(1, 1)$ and $Q(7, 11)$ are the ends of a **diameter** 直径 of a circle. Find the equation of the circle.

The centre is the midpoint of PQ:

$$\left(\frac{1+7}{2}, \frac{1+11}{2}\right) = (4, 6).$$

The radius is half the length of PQ:

$$r = \tfrac{1}{2}\sqrt{(7-1)^2 + (11-1)^2} = \tfrac{1}{2}\sqrt{36+100} = \tfrac{1}{2}\sqrt{136} = \sqrt{34}.$$

So the circle is $(x-4)^2 + (y-6)^2 = 34$.

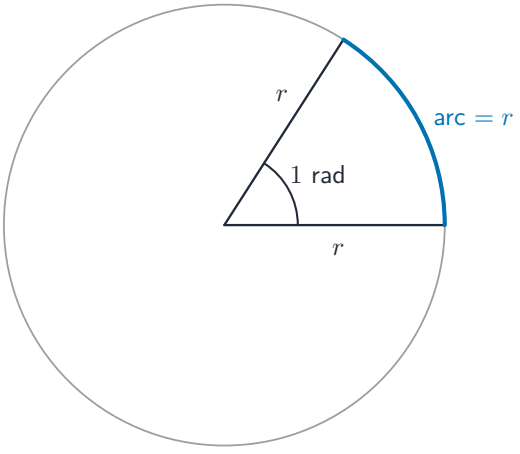
Circular measure

Radians

A **radian** 弧度 is another way to measure angles. One radian is the angle at the centre of a circle that cuts off an **arc** 弧 equal in length to the radius. The link between radians and **degrees** 度 is

$$\pi \text{ radians} = 180^\circ.$$

So to change degrees to radians, multiply by $\dfrac{\pi}{180}$; to change radians to degrees, multiply by $\dfrac{180}{\pi}$.



one **radian**: the angle whose arc is exactly one radius long

π rad $= 180^\circ$, so 1 rad $\approx 57.3^\circ$

degrees $\to$ radians: $\times \dfrac{\pi}{180}$ radians $\to$ degrees: $\times \dfrac{180}{\pi}$

One radian: the angle whose arc equals the radius

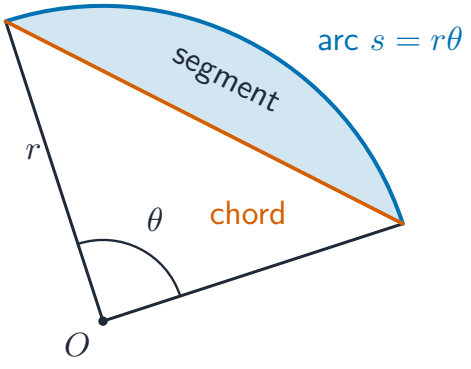
Arc length and sector area

For a **sector** 扇形 with radius r and angle θ **in radians**:

$$\text{arc length} = s = r\theta, \qquad \text{sector area} = A = \tfrac{1}{2}r^2\theta.$$

A **chord** 弦 cuts the sector into a triangle and a **segment** 弓形. The segment area is the sector minus the triangle:

$$\text{segment} = \tfrac{1}{2}r^2\theta - \tfrac{1}{2}r^2\sin\theta = \tfrac{1}{2}r^2(\theta - \sin\theta).$$



The shaded segment is the part of the sector between the chord and the arc.

Worked example. A sector has centre O and the angle at O is $\frac{2}{3}\pi$ radians. Show that the segment cut off by the chord has area about $0.614r^2$.

$$\text{segment} = \tfrac{1}{2}r^2\left(\tfrac{2}{3}\pi - \sin\tfrac{2}{3}\pi\right) = \tfrac{1}{2}r^2(2.0944 - 0.8660) = \tfrac{1}{2}r^2(1.2284) \approx 0.614r^2.$$

Trigonometry

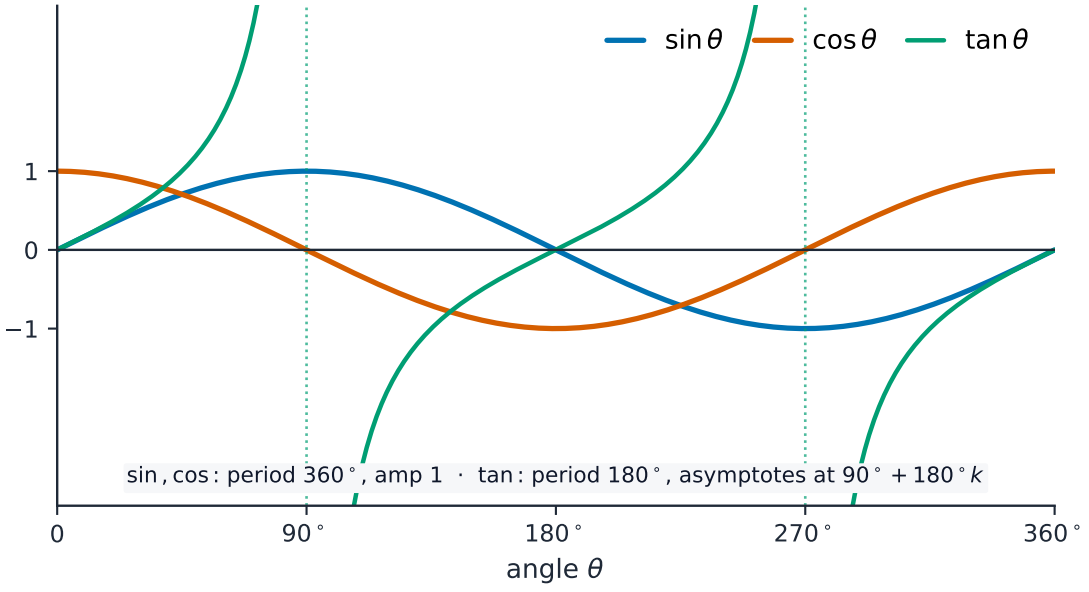


A Ferris wheel: a point on the rim rises and falls like a sine curve.

Graphs and exact values

You must know the shape of the graphs of the **sine** 正弦, **cosine** 余弦 and **tangent function** 正切 (written sin, cos, tan). The sine and cosine graphs wave between -1 and 1 and repeat every $360°$ (2π). Learn these exact values:

θ	$30°$	$45°$	$60°$
$\sin\theta$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$
$\cos\theta$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$
$\tan\theta$	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$



Over one turn sin *and* cos *stay between* -1 *and* 1*;* tan *shoots off at* $90°$ *and* $270°$*.*

The notations $\sin^{-1}x$, $\cos^{-1}x$, $\tan^{-1}x$ mean the **inverse** angle (the **principal value** 主值).

Identities

An **identity** 恒等式 is true for every value of the angle. The two you must know are

$$\tan\theta \equiv \frac{\sin\theta}{\cos\theta}, \qquad \sin^2\theta + \cos^2\theta \equiv 1.$$

Use them to rewrite an equation so it contains only one trig function.

Solving trigonometric equations

To solve a **trigonometric equation** 三角方程, first reduce it to one function, then find every solution in the given interval.

Worked example. Solve $6\sin\theta = 1 + \dfrac{2}{\sin\theta}$ for $-180° < \theta < 180°$.

Multiply through by $\sin\theta$ to clear the fraction. This makes a **quadratic in** $\sin\theta$:

$$6\sin^2\theta - \sin\theta - 2 = 0 \;\Rightarrow\; (3\sin\theta - 2)(2\sin\theta + 1) = 0.$$

So $\sin\theta = \frac{2}{3}$ or $\sin\theta = -\frac{1}{2}$.

- $\sin\theta = \frac{2}{3}$: $\theta = 41.8°$ or $\theta = 180° - 41.8° = 138.2°$.
- $\sin\theta = -\frac{1}{2}$: $\theta = -30°$ or $\theta = -150°$.

The four solutions are $\theta = -150°,\ -30°,\ 41.8°,\ 138.2°$.

Series

The binomial expansion

For a positive integer n, the **binomial expansion** 二项展开式 is

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + b^n,$$

where $\binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$ is a **binomial coefficient** 二项式系数.

Worked example. Find the first three terms, in ascending powers of x, of $(2-px)^5$.

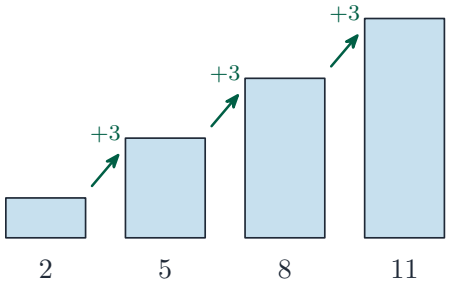
$$(2-px)^5 = 2^5 + \binom{5}{1}2^4(-px) + \binom{5}{2}2^3(-px)^2 + \cdots = 32 - 80px + 80p^2x^2 + \cdots$$

Arithmetic and geometric progressions

A **progression** 数列 (sequence) is a list of terms following a rule.

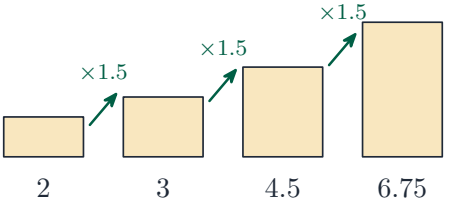
- An **arithmetic progression** 等差数列 (AP) adds a fixed **common difference** 公差 d each step. The nth **term** 项 is $u_n = a + (n-1)d$, and the sum of the first n terms is $S_n = \frac{n}{2}\big(2a + (n-1)d\big)$.
- A **geometric progression** 等比数列 (GP) multiplies by a fixed **common ratio** 公比 r each step. The nth term is $u_n = ar^{n-1}$, and $S_n = \dfrac{a(1-r^n)}{1-r}$.

arithmetic: add d each step



$u_n = a + (n-1)d$: equal steps

geometric: multiply by r each step



$u_n = ar^{n-1}$: growing steps

An AP climbs in equal steps; a GP's steps grow by the same ratio

A GP is **convergent** — it **converges** 收敛 (settles to a limit) — when $|r| < 1$. Then it has a **sum to infinity** 无穷和

$$S_\infty = \frac{a}{1-r}.$$

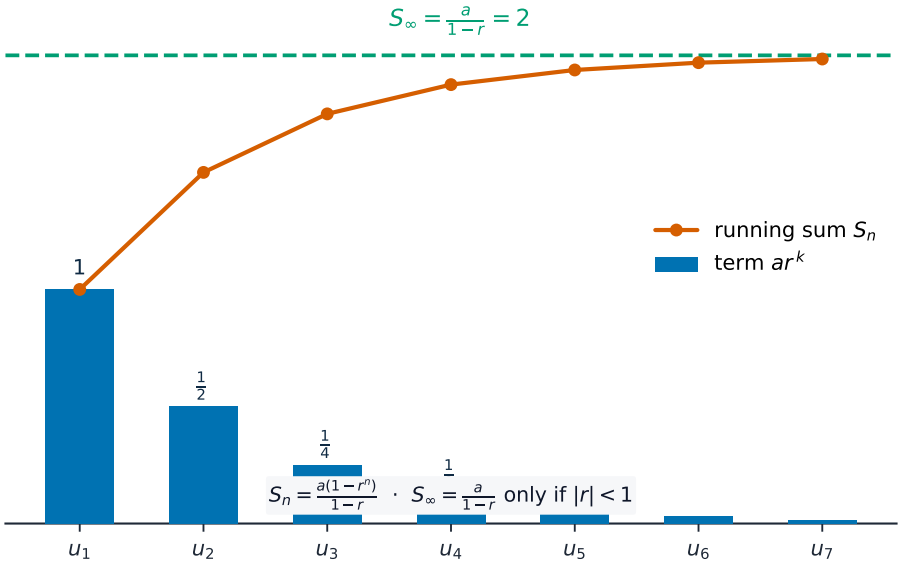
Worked example. The third term of a GP is 18 and the sum of the first three terms is 26. The common ratio is negative. Find the sum to infinity.

From $ar^2 = 18$ you get $a = \dfrac{18}{r^2}$. Put this into $a(1 + r + r^2) = 26$:

$$18(1 + r + r^2) = 26r^2 \;\Rightarrow\; 8r^2 - 18r - 18 = 0 \;\Rightarrow\; (4r + 3)(r - 3) = 0.$$

The ratio is negative, so $r = -\frac{3}{4}$ and $a = \dfrac{18}{(3/4)^2} = 32$. Then

$$S_\infty = \frac{32}{1 - (-\frac{3}{4})} = \frac{32}{\frac{7}{4}} = \frac{128}{7}.$$



Convergent GP: bars shrink toward zero while the running sum approaches S infinity = a/(1-r)

Differentiation

Differentiation 微分 finds the **gradient of a curve** 曲线斜率 at each point. The gradient is the **limit** 极限 of the gradients of shorter and shorter **chords**, called the **derivative** 导数.

The rules

Write the derivative as $f'(x)$ or $\dfrac{dy}{dx}$. The basic rule is

$$\frac{d}{dx}\left(x^n\right) = nx^{n-1} \quad \text{for any rational } n.$$

Differentiate sums term by term, and use the **chain rule** 链式法则 for a function inside a function:

$$\frac{d}{dx} f(g(x)) = f'(g(x)) \cdot g'(x).$$

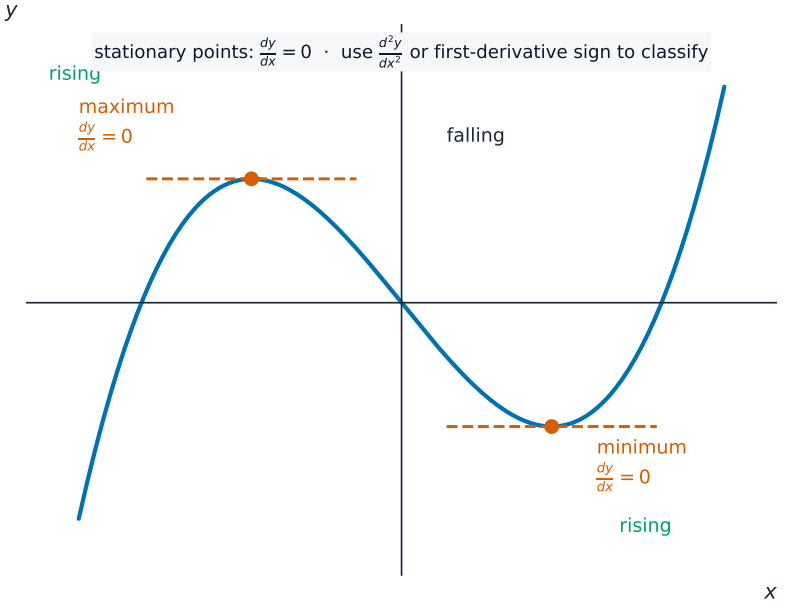
Differentiating again gives the **second derivative** 二阶导数 $f''(x)$ or $\dfrac{d^2y}{dx^2}$.

Using the derivative

- **Tangent and normal.** The gradient of the curve at a point is the gradient of the **tangent** there. The **normal** 法线 is perpendicular to the tangent, so its gradient is $-\dfrac{1}{(\text{tangent gradient})}$.
- **Increasing or decreasing.** The function is an **increasing function** 增函数 where $\dfrac{dy}{dx} > 0$, and a **decreasing function** 减函数 where $\dfrac{dy}{dx} < 0$.
- **Rate of change.** A derivative is a **rate of change** 变化率. Linked rates use the chain rule, e.g. $\dfrac{dy}{dt} = \dfrac{dy}{dx} \cdot \dfrac{dx}{dt}$.

Stationary points

A **stationary point** 驻点 is where $\dfrac{dy}{dx} = 0$. Test its nature with the second derivative: $f''(x) > 0$ gives a **minimum point** 极小值点, and $f''(x) < 0$ gives a **maximum point** 极大值点. When $f''(x) = 0$ the test is inconclusive: the point may be a **point of inflexion** 拐点, where the curve changes concavity (the way it bends) – check the sign of $\dfrac{dy}{dx}$ just before and after to decide.



At a maximum or a minimum the tangent is flat, so $\frac{dy}{dx} = 0$.

Worked example. The curve $y = 4x^{1/2} - x$ has a maximum point at $x = a$. Find a.

$$\frac{dy}{dx} = 2x^{-1/2} - 1 = 0 \;\Rightarrow\; \frac{2}{\sqrt{x}} = 1 \;\Rightarrow\; \sqrt{x} = 2 \;\Rightarrow\; x = 4.$$

So $a = 4$.

Integration

Integration 积分 is the reverse of differentiation. Reversing the power rule gives

$$\int (ax+b)^n \, dx = \frac{(ax+b)^{n+1}}{a(n+1)} + C \quad (n \neq -1).$$

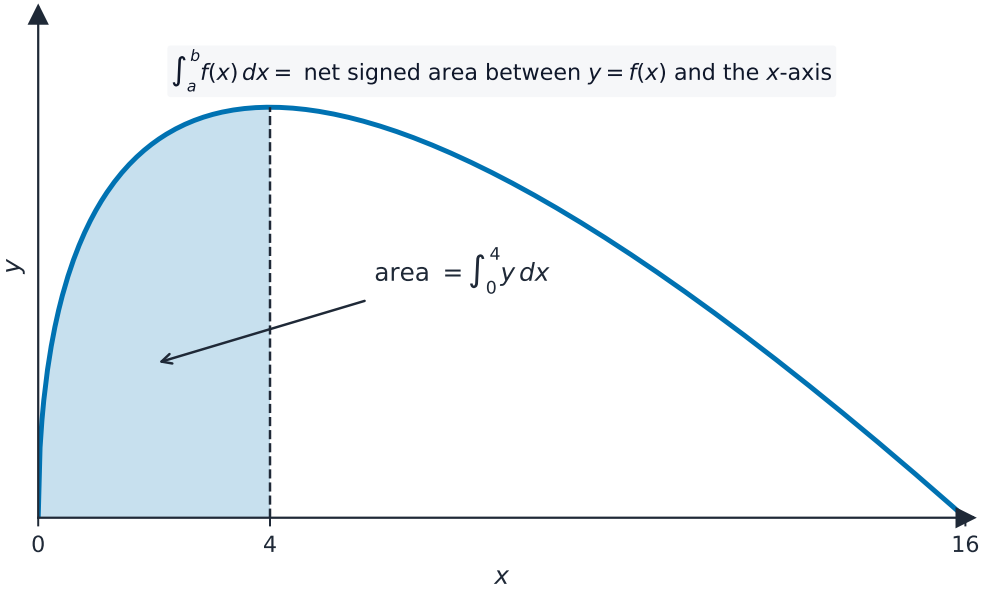
The $+C$ is the **constant of integration** 积分常数. If you know one point on the curve, substitute it to find C.

Definite integrals and area

A **definite integral** 定积分 has limits and gives a number:

$$\int_p^q f(x) \, dx = \big[F(x)\big]_p^q = F(q) - F(p).$$

The area of the **region** 区域 between a curve and the x-axis, from $x = p$ to $x = q$, is $\int_p^q y \, dx$. For the area between two curves, integrate (top curve − bottom curve).



The definite integral $\int_0^4 y \, dx$ is the shaded area under the curve.

Worked example. The curve $y = 4x^{1/2} - x$ meets the x-axis again at $x = 16$. Find

the area between the curve and the x-axis from $x = 0$ to $x = 4$.

$$\int_0^4 \left(4x^{1/2} - x\right) dx = \left[\frac{8}{3}x^{3/2} - \frac{x^2}{2}\right]_0^4 = \frac{8}{3}(8) - \frac{16}{2} = \frac{64}{3} - 8 = \frac{40}{3}.$$

An **improper integral** 广义积分 has an infinite limit or an endpoint where the integrand is undefined; evaluate it as a **limit**. For example,

$$\int_1^\infty x^{-2}\, dx = \lim_{b\to\infty}\left[-\frac{1}{x}\right]_1^b = \lim_{b\to\infty}\left(1 - \frac{1}{b}\right) = 1,$$

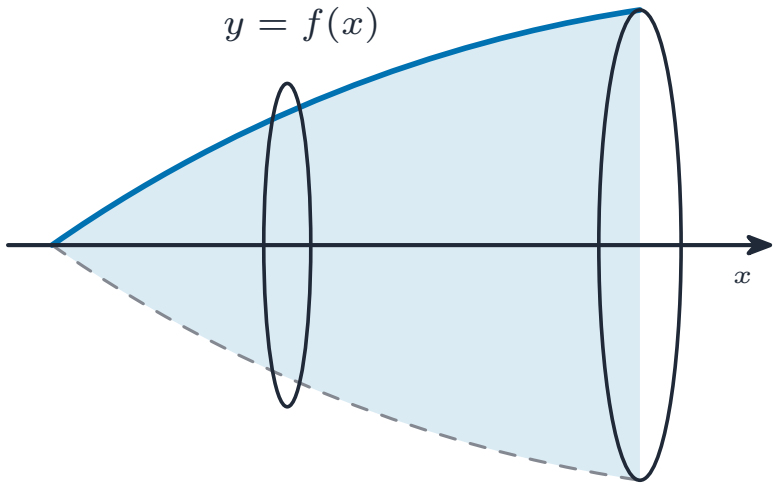
and $\displaystyle\int_0^1 x^{-1/2}\, dx = \left[2x^{1/2}\right]_0^1 = 2$, which is finite even though the integrand blows up at $x = 0$.

Volume of revolution

When a region is turned all the way around an axis it sweeps out a solid. The **volume of revolution** 旋转体体积 about the x-axis is

$$V = \pi \int_p^q y^2 \, dx,$$

and about the y-axis it is $V = \pi \int x^2 \, dy$. For example, the region under $y = \sqrt{x}$ from $x = 0$ to $x = 4$, turned about the x-axis, has volume $\pi \int_0^4 x \, dx = \pi \left[\frac{x^2}{2}\right]_0^4 = 8\pi$.



rotate the region around the x-axis to sweep a solid

Rotating the region under a curve around the x-axis sweeps out a solid; each thin slice is a disc of area πy^2, so $V = \pi \int y^2 \, dx$

Exam tips

- Show **every line of algebra** — method marks are lost by jumping steps; use the **discriminant** $b^2 - 4ac$ to decide the number of real roots.
- Work in **radians** for circular measure (arc $= r\theta$, sector area $= \frac{1}{2}r^2\theta$) and for calculus of trig functions.
- For **differentiation**, set $\frac{dy}{dx} = 0$ for stationary points and use the second derivative to classify them.
- For **integration**, add $+c$ to an indefinite integral and use limits for area; an area below the axis gives a negative integral.

Pure Mathematics 2

A-Level Mathematics

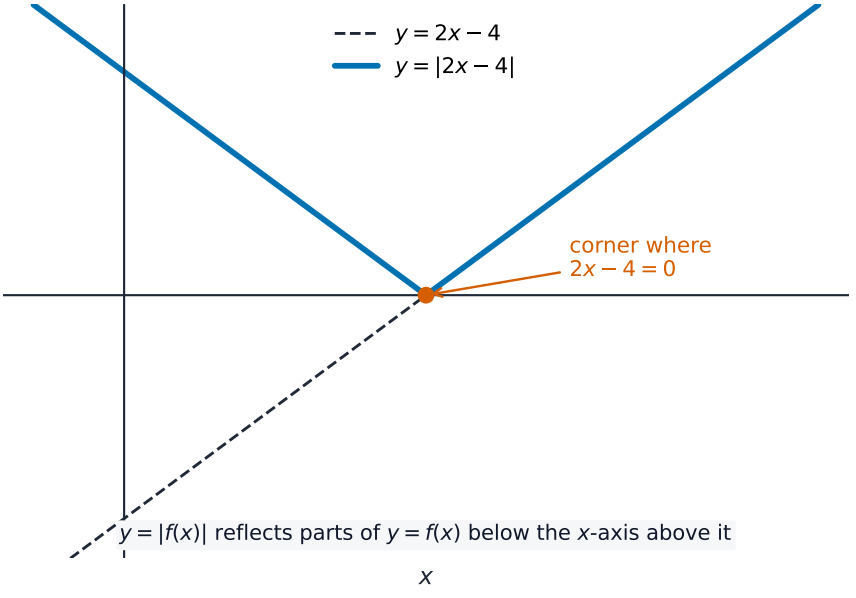
This handout covers Topic 2: **Pure Mathematics** 纯数学 2. It adds the modulus and polynomial algebra, **logarithms** 对数 and the **exponential function** 指数函数, more trigonometry, and new ways to differentiate and integrate.

Algebra

The modulus

The **modulus** 绝对值 $|x|$ is the size of a number with its sign removed, so $|x| \geqslant 0$ always. The graph of $y = |ax + b|$ is a "V" shape that bounces off the x-axis. Two useful rules for solving equations and inequalities are

$$|a| = |b| \;\Leftrightarrow\; a^2 = b^2, \qquad |x - a| < b \;\Leftrightarrow\; a - b < x < a + b.$$



Taking the modulus folds the part of the line below the axis upward into a V.

Worked example. Solve $|3x + 8| < 9$.

Using the second rule with the inequality written as $-9 < 3x + 8 < 9$:

$$-9 < 3x + 8 < 9 \;\Rightarrow\; -17 < 3x < 1 \;\Rightarrow\; -\tfrac{17}{3} < x < \tfrac{1}{3}.$$

Polynomial division and the factor and remainder theorems

A **polynomial** 多项式 is a sum of powers of x, such as $2x^4+3x^2-5$. Its **degree** 次数 is the highest power. When you divide one polynomial by another you get a **quotient** 商 and a **remainder** 余数.

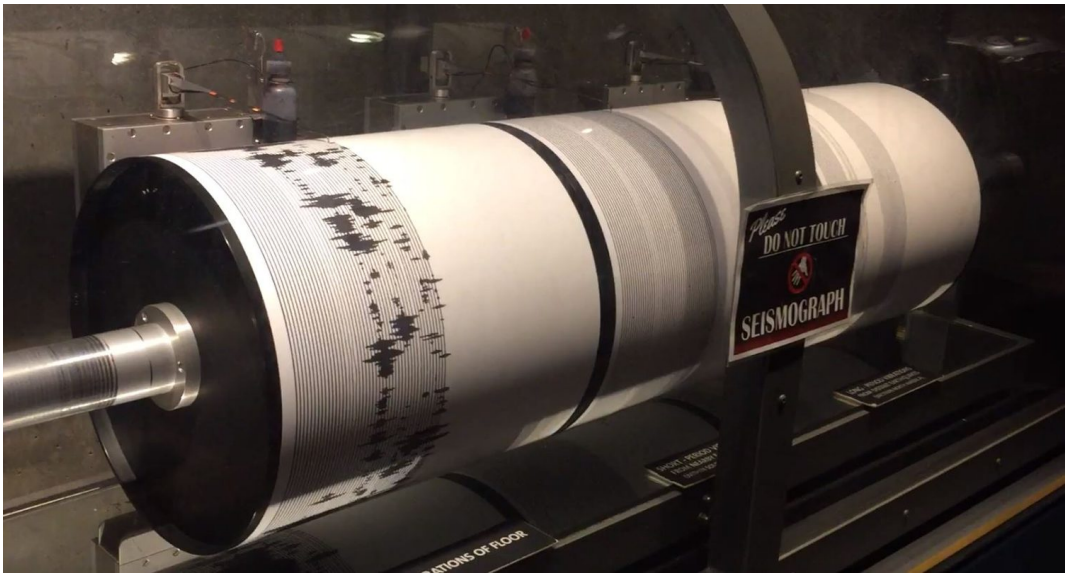
- **Remainder theorem** 余数定理: the remainder when $p(x)$ is divided by $(x-a)$ is $p(a)$.
- **Factor theorem** 因式定理: $(x-a)$ is a factor of $p(x)$ exactly when $p(a)=0$.

Worked example. The polynomial $p(x) = 2x^4 + kx^3 + kx^2 + 17x + 18$ has factor $(x+2)$. Find k.

By the factor theorem $p(-2)=0$:

$$2(16) + k(-8) + k(4) + 17(-2) + 18 = 0 \;\Rightarrow\; 16 - 4k = 0 \;\Rightarrow\; k = 4.$$

Logarithmic and exponential functions



Earthquake strength is measured on the logarithmic Richter scale.

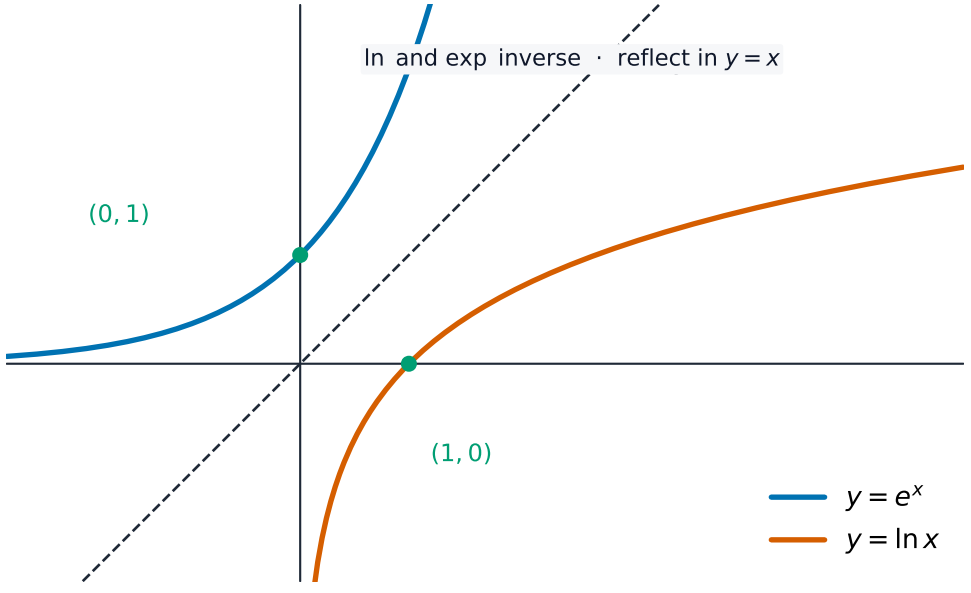


Populations can grow exponentially when resources are plentiful.

A **logarithm** answers the question "what power?". If $a^x = y$ then $x = \log_a y$. Logarithms and **indices** 指数 (powers) are reverse ideas. The **laws of logarithms** 对数定律 are

$$\log(mn) = \log m + \log n, \qquad \log\frac{m}{n} = \log m - \log n, \qquad \log(m^k) = k \log m.$$

The exponential function e^x and the **natural logarithm** 自然对数 $\ln x$ are inverse functions, so $\ln(e^x) = x$ and $e^{\ln x} = x$. When the unknown is in the power, take logs of both sides.



e^x and $\ln x$ undo each other, so each is the other reflected in $y = x$.

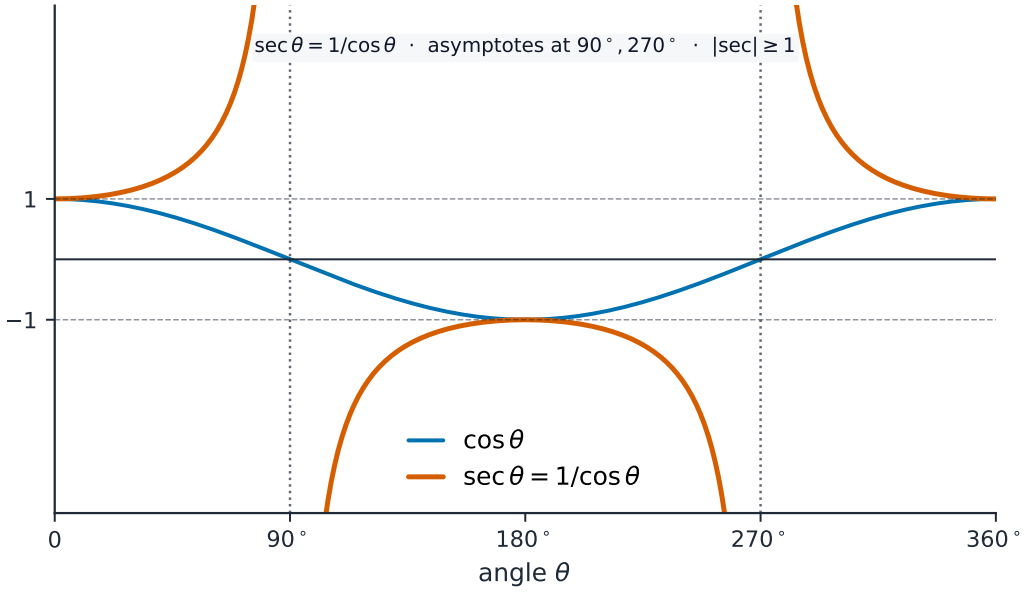
Worked example. Solve $4^x < 0.05$.

$$4^x < 0.05 \;\Rightarrow\; x \ln 4 < \ln 0.05 \;\Rightarrow\; x < \frac{\ln 0.05}{\ln 4} = -2.16 \text{ (3 s.f.)}.$$

Linear form 线性形式: a relationship like $y = Ax^n$ becomes a straight line if you take logs: $\ln y = \ln A + n \ln x$. Plotting $\ln y$ against $\ln x$ gives a line with gradient n and intercept $\ln A$, so you can find the unknown constants.

Trigonometry

There are three more functions, each the reciprocal of one you know: the **secant** 正割 $\sec\theta = \dfrac{1}{\cos\theta}$, the **cosecant** 余割 $\csc\theta = \dfrac{1}{\sin\theta}$, and the **cotangent** 余切 $\cot\theta = \dfrac{1}{\tan\theta}$.



$\sec\theta = 1/\cos\theta$ *rises to an asymptote wherever* $\cos\theta = 0$.

You must know these **trigonometric identities** 三角恒等式 and choose the right one for each problem:

$$\sec^2\theta \equiv 1 + \tan^2\theta, \qquad \csc^2\theta \equiv 1 + \cot^2\theta.$$

You also use the **compound angle** 复合角 formulae for $\sin(A \pm B)$, $\cos(A \pm B)$, $\tan(A \pm B)$, the **double angle** 二倍角 formulae

$$\sin 2A = 2\sin A\cos A, \quad \cos 2A = 2\cos^2 A - 1, \quad \tan 2A = \frac{2\tan A}{1 - \tan^2 A},$$

and the **R-formula** 辅助角公式 $a\sin\theta + b\cos\theta = R\sin(\theta + \alpha)$, where $R = \sqrt{a^2 + b^2}$ and $\tan\alpha = \dfrac{b}{a}$.

Worked example. Solve $2\tan^2\theta + 3\sec\theta = 18$ for $-180° < \theta < 180°$.

Replace $\tan^2\theta$ with $\sec^2\theta - 1$ to get one function:

$$2(\sec^2\theta - 1) + 3\sec\theta = 18 \;\Rightarrow\; 2\sec^2\theta + 3\sec\theta - 20 = 0 \;\Rightarrow\; (2\sec\theta - 5)(\sec\theta + 4) = 0.$$

So $\sec\theta = \frac{5}{2}$ or $\sec\theta = -4$, giving $\cos\theta = \frac{2}{5}$ or $\cos\theta = -\frac{1}{4}$. The solutions are $\theta = \pm 66.4°$ and $\theta = \pm 104.5°$.

Differentiation

Learn these standard derivatives:

$$\frac{d}{dx}e^x = e^x, \quad \frac{d}{dx}\ln x = \frac{1}{x}, \quad \frac{d}{dx}\sin x = \cos x, \quad \frac{d}{dx}\cos x = -\sin x, \quad \frac{d}{dx}\tan x = \sec^2 x.$$

For a product or a quotient of two functions, use:

- the **product rule** 乘积法则: $(uv)' = u'v + uv'$;
- the **quotient rule** 商法则: $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$.

When a curve is given by **parametric equations** 参数方程 $x = x(t)$, $y = y(t)$, the gradient is $\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$. When y is defined **implicitly** 隐式 (not made the subject), differentiate every term with respect to x, using the chain rule on the y terms, then solve for $\frac{dy}{dx}$.

Worked example. Given $y = 6x\cos(x^2+1)$, find $\frac{dy}{dx}$.

Use the product rule with $u = 6x$ and $v = \cos(x^2+1)$ (and the chain rule for v):

$$\frac{dy}{dx} = 6\cos(x^2+1) + 6x\cdot\big(-2x\sin(x^2+1)\big) = 6\cos(x^2+1) - 12x^2\sin(x^2+1).$$

Integration

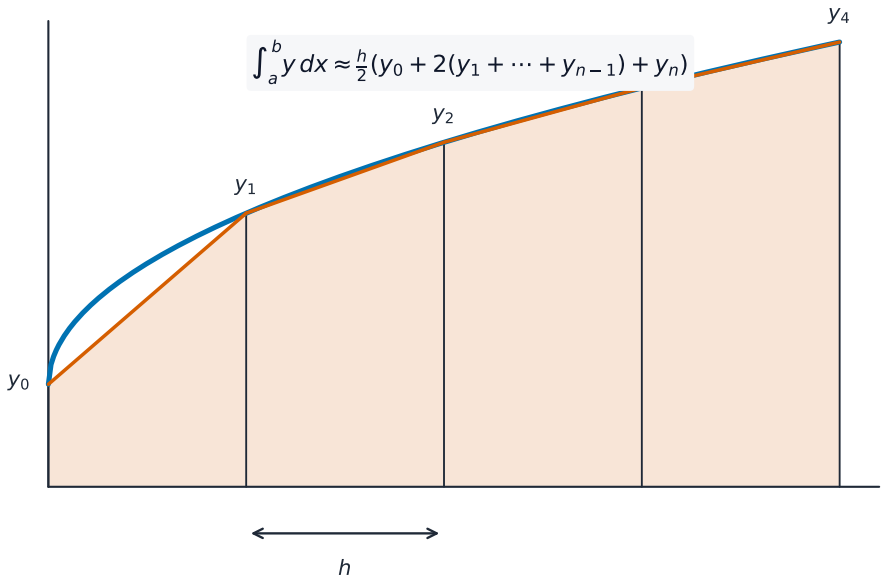
Integration is **reverse differentiation** — reverse each new derivative; harder integrals may need **integration by substitution** 换元积分 (developed in Pure 3). For a linear inside function $(ax+b)$:

$$\int e^{ax+b}\,dx = \frac{1}{a}e^{ax+b} + C, \qquad \int \frac{1}{ax+b}\,dx = \frac{1}{a}\ln|ax+b| + C,$$

$$\int \sin(ax+b)\,dx = -\frac{1}{a}\cos(ax+b) + C, \qquad \int \cos(ax+b)\,dx = \frac{1}{a}\sin(ax+b) + C, \qquad \int \sec^2(ax+b)\,dx = \frac{1}{a}\tan(ax+b) + C.$$

To integrate a power of sin or cos, first use an identity to remove the power. When you cannot integrate exactly, the **trapezium rule** 梯形法则 estimates a definite integral:

$$\int_a^b y\,dx \approx \tfrac{h}{2}\big[y_0 + y_n + 2(y_1 + y_2 + \cdots + y_{n-1})\big].$$



Each strip of width h is a trapezium; their areas add up to estimate the integral.

Worked example. Find $\displaystyle\int 6\sin^2 x\,dx$.

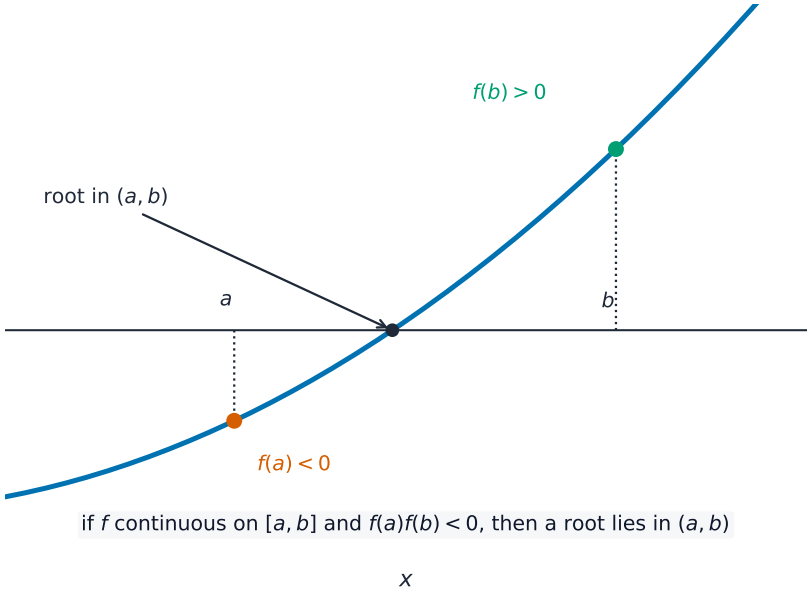
Use $\sin^2 x = \frac{1}{2}(1 - \cos 2x)$:

$$\int 6\sin^2 x\,dx = \int (3 - 3\cos 2x)\,dx = 3x - \tfrac{3}{2}\sin 2x + C.$$

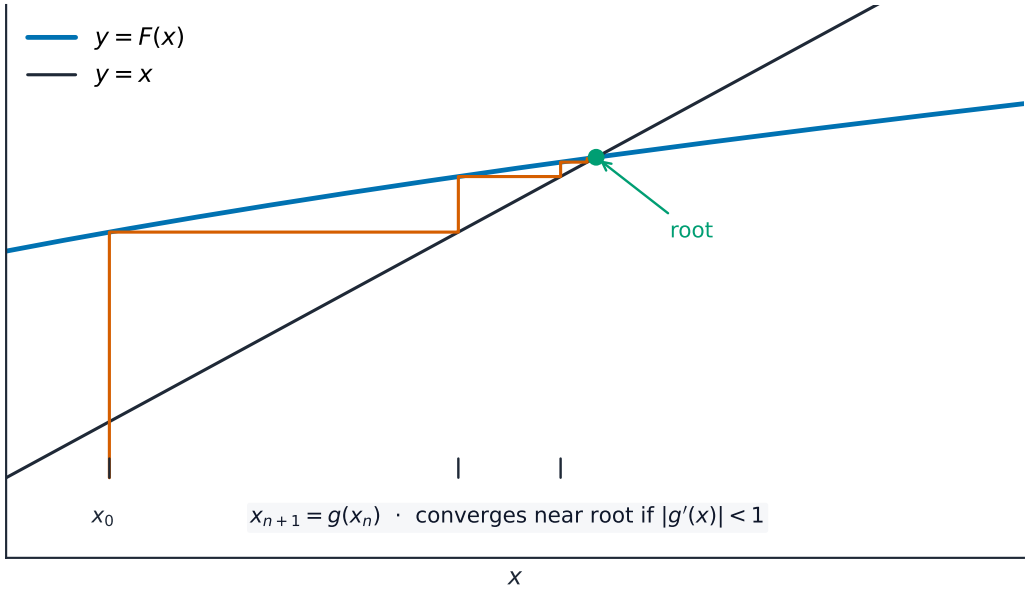
Numerical solution of equations

Many equations cannot be solved exactly. Two ideas help you find a **root** 根 (a solution).

- **Sign change** 变号: if $f(a)$ and $f(b)$ have opposite signs (and the graph has no break between them), a root lies between a and b.
- **Iteration** 迭代: rearrange the equation into the form $x = F(x)$, then use the **iterative formula** 迭代公式 $x_{n+1} = F(x_n)$. Start from a first guess x_0 and repeat. If the values are **convergent** they settle down and **converge** 收敛 to a root. Keep going until the answer is steady to the accuracy asked for.



$f(a)$ and $f(b)$ have opposite signs, so a root is trapped between a and b.



Each step goes up to $y = F(x)$ then across to $y = x$; the staircase closes in on the root.

Worked example. A root β of an equation satisfies $x = \sqrt[3]{-2x - 4.5}$, and $-1.4 < \beta < -1.0$. Use the iteration $x_{n+1} = \sqrt[3]{-2x_n - 4.5}$ with $x_0 = -1.2$.

$$x_1 = \sqrt[3]{-2(-1.2) - 4.5} = \sqrt[3]{-2.1} = -1.281, \qquad x_2 = \sqrt[3]{-2(-1.281) - 4.5} = -1.247, \quad \ldots$$

The values settle near -1.26, so $\beta = -1.26$ (3 s.f.).

Exam tips

- Use the **laws of logarithms** to solve equations; remember ln and e^x are inverses.
- For **numerical methods**, show a **sign change** to locate a root, set out the iteration clearly, and give the answer to the stated accuracy.
- Learn the **chain, product and quotient rules** and identify which the function needs.
- Solve modulus equations $|f(x)| = g(x)$ by considering both the positive and negative cases, and sketch to check.

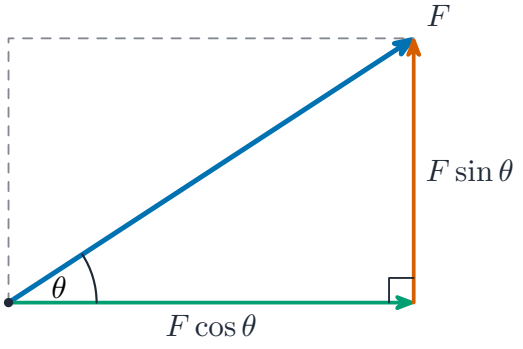
Mechanics

A-Level Mathematics

This handout covers Topic 4: **Mechanics** 力学. It studies how forces make objects move. Throughout this topic, take the acceleration of free fall as $g = 10$ m s^{-2}.

Forces and equilibrium

A **force** 力 is a push or a pull. It is a vector, so it has size and direction. Because it is a vector, you can split a force into **components** 分量 (usually horizontal and vertical), and you can add several forces into one **resultant** 合力.



A force at angle θ has a horizontal part $F\cos\theta$ and a vertical part $F\sin\theta$.

A particle is in **equilibrium** 平衡 when the forces are balanced: the vector sum of the forces is zero. In practice this means the components in any direction add to zero.



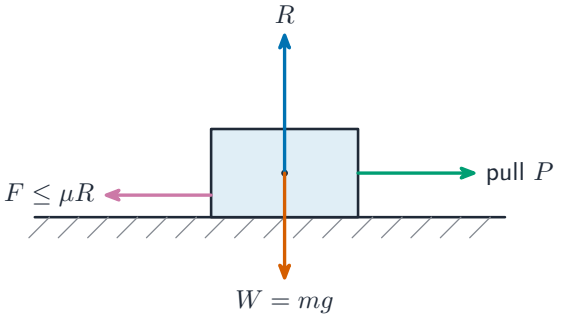
Every mechanics problem is a picture like this. Three forces act on the climber — weight straight down, tension along the rope, and a push from the rock — and because they balance, the climber hangs still in equilibrium. Resolving each into horizontal and vertical parts turns the picture into equations

Friction

When two surfaces touch, the **contact force** 接触力 between them has two parts: the **normal reaction** 法向反作用力 R, at right angles to the surface, and the **friction** 摩擦力 F, along the surface, which opposes sliding. A "smooth" surface is a model with no friction.

Friction can only grow up to a maximum. At that maximum the body is in **limiting equilibrium** 极限平衡, about to slip, and the friction is **limiting friction** 最大静摩擦力. The maximum is set by the **coefficient of friction** 摩擦系数 μ:

$$F \leqslant \mu R, \qquad \text{with } F = \mu R \text{ at the point of slipping.}$$



On a rough surface the contact force splits into the normal reaction R and friction F (at most μR).

By **Newton's third law** 牛顿第三定律, the two surfaces push on each other with equal and opposite forces.

Worked example. A block of weight 20 N rests on a rough horizontal table with coefficient of friction $\mu = 0.4$. Find the largest horizontal force that can be applied before the block slides.

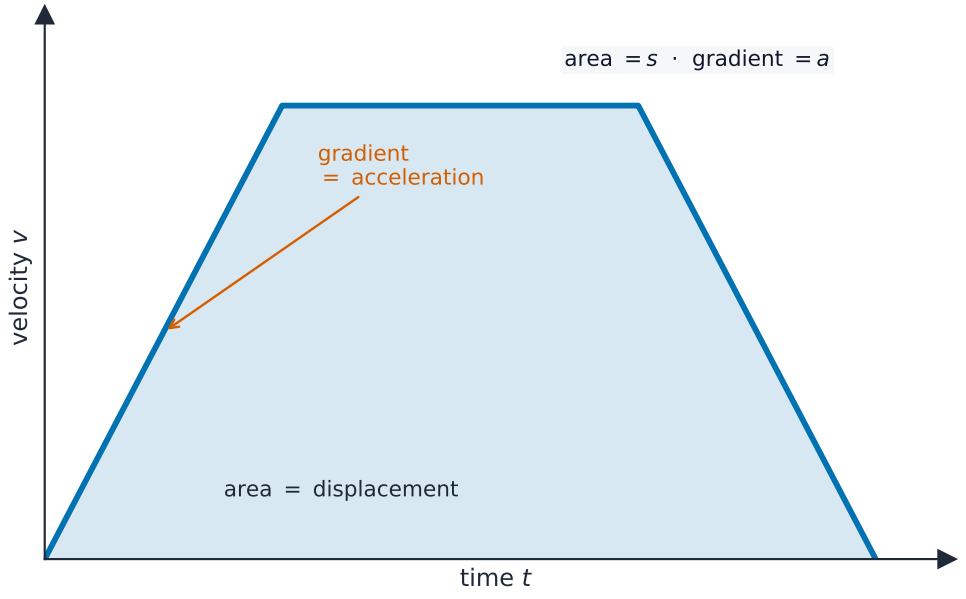
The table's normal reaction balances the weight, so $R = 20$ N. The block is on the point of slipping when friction reaches its maximum $F = \mu R = 0.4 \times 20 = 8$ N. In equilibrium the applied force equals the friction, so the largest force it can resist is 8 N.

Kinematics of motion in a straight line

Distance 距离 and **speed** 速率 are scalars (size only). **Displacement** 位移, **velocity** 速度 and **acceleration** 加速度 are vectors (size and direction).

On a **velocity-time graph** 速度时间图, the area under the graph is the displacement and the gradient is the acceleration (a negative acceleration is a **deceleration** 减速

度). On a displacement-time graph, the gradient is the velocity. More generally, differentiate with respect to time to go from displacement to velocity to acceleration, and integrate to go back.



The shaded area gives the distance travelled; the slope of the line gives the acceleration.

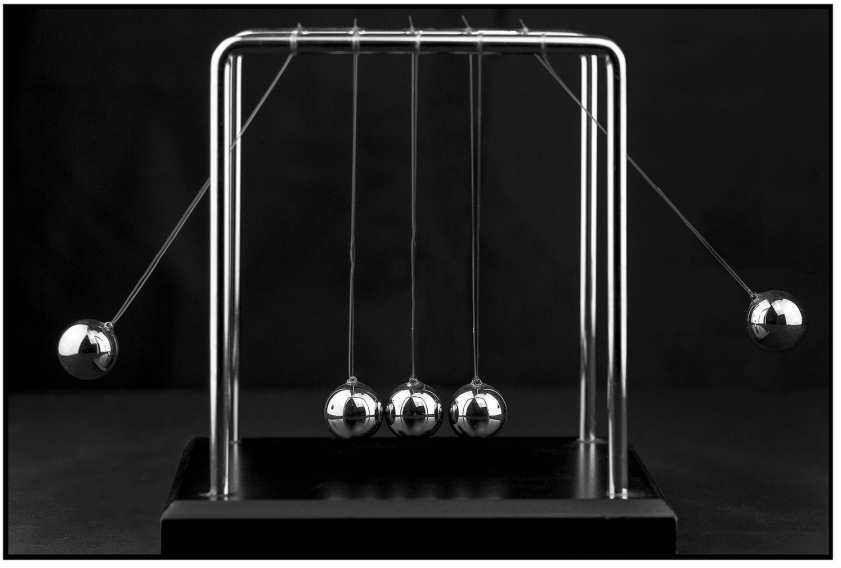
For motion with **constant acceleration** 匀加速, use these formulae (the "suvat" equations):

$$v = u + at, \qquad s = ut + \tfrac{1}{2}at^2, \qquad v^2 = u^2 + 2as, \qquad s = \tfrac{1}{2}(u + v)t.$$

Worked example. A car starts from rest and accelerates at 2.5 m s^{-2} for 4 s. Find its speed and the distance travelled.

$$v = 0 + 2.5 \times 4 = 10 \text{ m s}^{-1}, \qquad s = 0 + \tfrac{1}{2}(2.5)(4^2) = 20 \text{ m}.$$

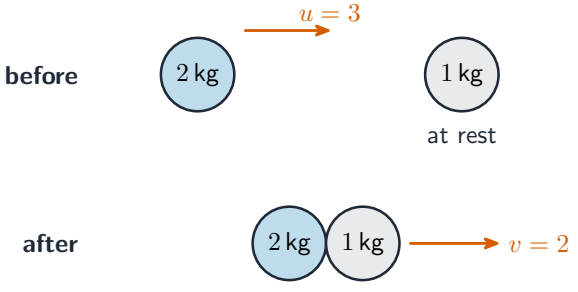
Momentum



A Newton's cradle demonstrates conservation of momentum in collisions.

The **linear momentum** 动量 of a body is mass × velocity. It is a vector. In a direct collision of two bodies, the total momentum is unchanged. This is the **conservation of linear momentum** 动量守恒:

$$m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2.$$



$$2(3) + 1(0) = (2 + 1)\,v$$

The total momentum is the same before and after the collision.

Worked example. A body of mass 2 kg moving at 3 m s^{-1} hits a stationary body of mass 1 kg, and they stick together. Find their common speed afterwards.

$$2(3) + 1(0) = (2 + 1)v \;\Rightarrow\; v = \frac{6}{3} = 2 \text{ m s}^{-1}.$$

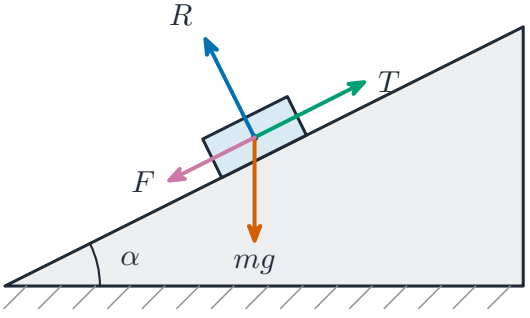
Newton's laws of motion

Newton's laws of motion 牛顿运动定律 connect force and acceleration. The key one is: resultant force = **mass** 质量 × acceleration,

$$F = ma.$$

The **weight** 重力 of a body is the force of gravity on it: $W = mg$. Forces in a problem may include weight, friction, **tension** 张力 in a string, the **thrust** 推力 (push) in a rod, and **air resistance** 空气阻力.

For motion on an **inclined plane** 斜面, split each force into a part along the slope and a part at right angles to it, then apply $F = ma$ along the slope. For **connected particles** 连接质点 (joined by a string), apply $F = ma$ to each body, or to the whole system.



On a slope, resolve the forces along the slope and at right angles to it.

Worked example. A block of mass 12 kg is pulled up a rough plane by a rope parallel to the slope. The plane is at 20° to the horizontal, the coefficient of friction is 0.4, and the acceleration is 2 m s^{-2}. Find the tension in the rope.

The normal reaction is $R = mg\cos 20° = 120\cos 20° = 112.8$ N, so the friction is $F = \mu R = 0.4 \times 112.8 = 45.1$ N. Along the slope, $T - mg\sin 20° - F = ma$:

$$T = ma + mg\sin 20° + F = 12(2) + 120\sin 20° + 45.1 = 24 + 41.0 + 45.1 = 110 \text{ N (3 s.f.)}.$$

Energy, work and power



A roller coaster trades potential energy for kinetic energy as it rises and falls.

The **work done** 功 by a constant force is the **scalar product** of force and displacement — the force times the distance moved in the direction of the force: $W = Fd\cos\theta$, where θ is the angle between the force and the motion. Work is measured in joules (J).

Energy comes in forms you can calculate:

- **Kinetic energy** 动能 (energy of movement): $\text{KE} = \frac{1}{2}mv^2$.
- **Gravitational potential energy** 重力势能 (energy of height): $\text{PE} = mgh$.

The work done by the outside forces equals the change in the total energy. When no friction acts, the total energy stays the same — the **conservation of energy** 能量守恒.

Power 功率 is the rate of doing work. For a force pulling in the direction of motion, $P = Fv$ (power = force × velocity). Power is measured in watts (W).

Worked example. A car engine works at 12 kW while the car moves at 20 m s^{-1} on a level road. Find the driving force.

$$P = Fv \;\Rightarrow\; F = \frac{P}{v} = \frac{12000}{20} = 600 \text{ N}.$$

Exam tips

- Draw a clear **force diagram** and resolve into perpendicular components; for equilibrium, each direction sums to zero.
- Use **SUVAT** only for constant acceleration and keep a consistent positive direction.
- Apply $F = ma$ along the direction of motion, including friction ($F = \mu R$) on a rough surface.
- State your **assumptions** (light inextensible string, smooth pulley, particle) — they are often worth a mark.

Probability & Statistics 1

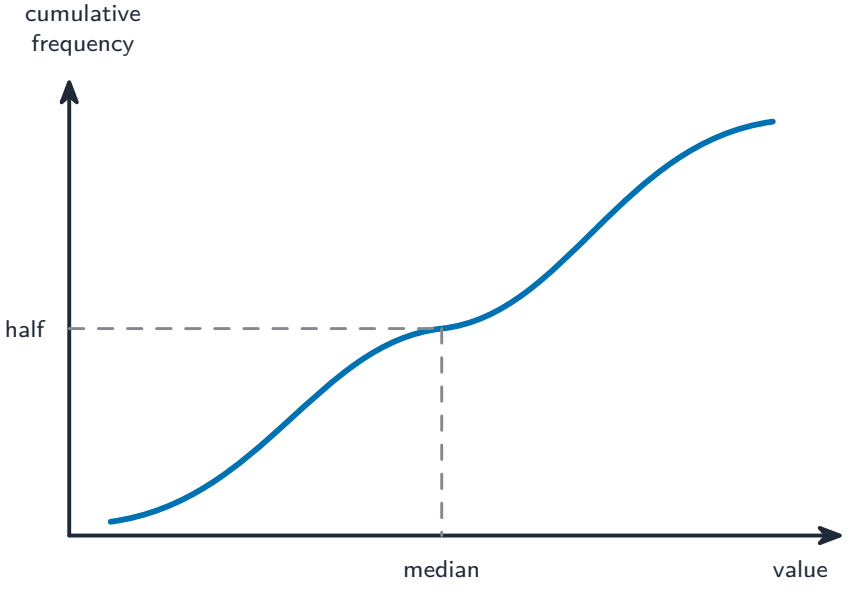
A-Level Mathematics

This handout covers Topic 5: **Probability & Statistics** 概率统计 1. It is about describing data, counting choices, and working out the chance of events.

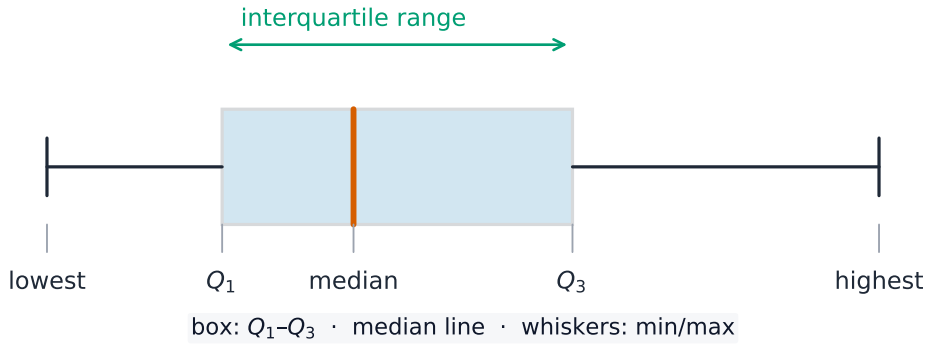
Representation of data

Choose a diagram that suits the data. You should be able to draw and read:

- a **stem-and-leaf diagram** 茎叶图 (keeps the original values and shows the shape); two data sets are compared with a **back-to-back** 背靠背 version – a shared central stem, one set's leaves increasing to the left and the other's to the right, so you can compare their medians and spreads at a glance;
- a **box-and-whisker plot** 箱线图 (shows the lowest value, the three quartiles, and the highest value);
- a **histogram** 直方图 (for grouped data, where the **area** of each bar shows the frequency);
- a **cumulative frequency** 累积频数 graph (running totals, used to estimate the median and quartiles).



A cumulative frequency curve is S-shaped; read the median across from half the total frequency, and the quartiles from one and three quarters



The box spans the quartiles Q_1 to Q_3; the whiskers reach the lowest and highest values.

Averages and spread

A **measure of central tendency** 集中趋势 is a single "middle" value:

- the **mean** 平均数 $\bar{x} = \dfrac{\sum x}{n}$ (the average);
- the **median** 中位数 (the middle value when the data is in order);
- the **mode** 众数 (the most common value).

A measure of **variation** 离散程度 shows how spread out the data is:

- the **range (of data)** 极差 (highest − lowest);
- the **interquartile range** 四分位距 (upper quartile − lower quartile);
- the **standard deviation** 标准差 $\sigma = \sqrt{\dfrac{\sum x^2}{n} - \bar{x}^2}$.

You often work from the totals $\sum x$ and $\sum x^2$. The square of the standard deviation is the **variance** 方差.

Worked example. For 10 values, $\sum x = 50$ and $\sum x^2 = 300$. Find the mean and standard deviation.

$$\bar{x} = \frac{50}{10} = 5, \qquad \sigma = \sqrt{\frac{300}{10} - 5^2} = \sqrt{30 - 25} = \sqrt{5} = 2.24.$$

Coding 编码 makes big numbers easier. Replace each value by $t = x - a$ for a convenient **assumed mean** 假定平均数 a. Then $\bar{x} = a + \bar{t}$, while the standard deviation is **unchanged** (shifting every value does not spread the data). So from the

coded totals $\sum(x-a)$ and $\sum(x-a)^2$ you get $\bar{x}$ and σ directly, and two data sets can be compared through their coded totals.

Permutations and combinations

A **permutation** 排列 is an arrangement where order matters; a **combination** 组合 is a selection where order does not matter. The numbers are

$$ {}^{n}P_{r} = \frac{n!}{(n-r)!}, \qquad {}^{n}C_{r} = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}. $$

To arrange objects in a line when some are repeated, divide by the factorial of each repeat count.

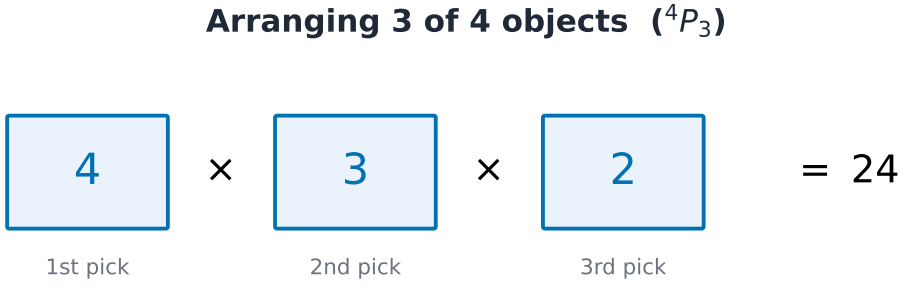


Order matters for a permutation, but not for a combination

Worked example. How many different arrangements are there of the letters of the word NEEDLESS?

There are 8 letters, with E repeated 3 times and S repeated 2 times:

$$ \frac{8!}{3!\,2!} = \frac{40320}{6 \times 2} = 3360. $$



nPr = n!/(n−r)!; nCr = nPr/r!

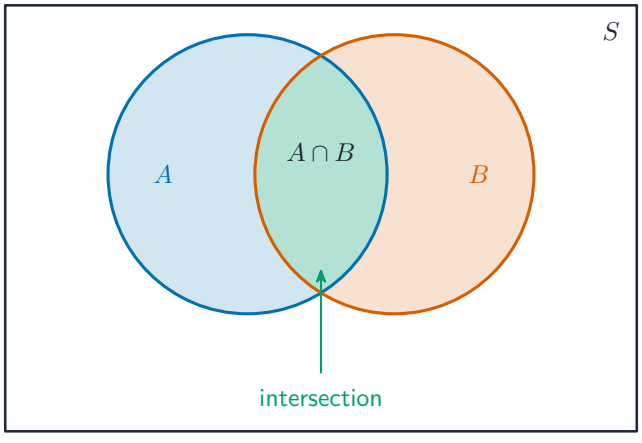
Probability



Dice: a familiar starting point for probability.

Find a probability by counting equally likely outcomes, or by using permutations and combinations. Combine probabilities with these rules:

- **addition** for "or": $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- **multiplication** for "and" when events are independent: $P(A \cap B) = P(A)\,P(B)$.



The overlap of the two circles is $A \cap B$; the addition rule subtracts it once so it is not counted twice.

Two events are **mutually exclusive events** 互斥事件 if they cannot both happen, and **independent events** 独立事件 if one happening does not change the chance

of the other. To test independence, check whether $P(A \cap B) = P(A) \times P(B)$. A **conditional probability** 条件概率 is the chance of A given that B has happened: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$.

Worked example. Events have $P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cap B) = 0.2$. Are A and B independent?

Test: $P(A) \times P(B) = 0.5 \times 0.4 = 0.2 = P(A \cap B)$. The values are equal, so A and B are independent.

Discrete random variables

A **discrete random variable** 离散型随机变量 X takes separate values, each with a probability. List them in a **probability distribution table** 概率分布表; the probabilities must add to 1. Then the **expectation** 期望 (mean) and **variance** are

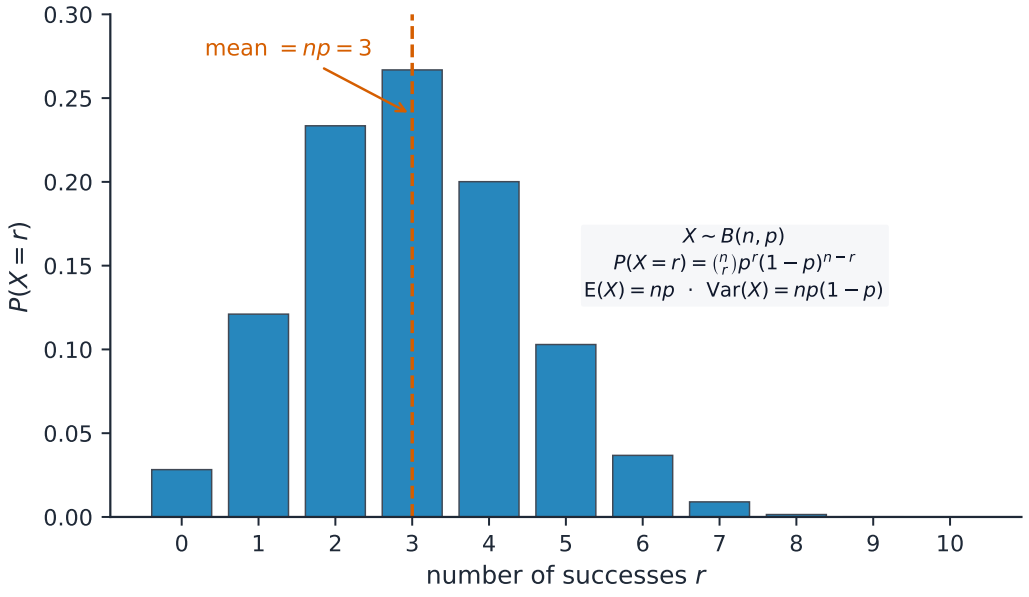
$$E(X) = \sum x\,P(X = x), \qquad \mathrm{Var}(X) = \sum x^2\,P(X = x) - \big(E(X)\big)^2.$$

Two special models:

- the **binomial distribution** 二项分布 $X \sim B(n, p)$, for the number of successes in n independent trials: $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$, with $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$;
- the **geometric distribution** 几何分布, for the trial on which the first success happens: $P(X = r) = (1-p)^{r-1}p$, with $E(X) = \dfrac{1}{p}$.

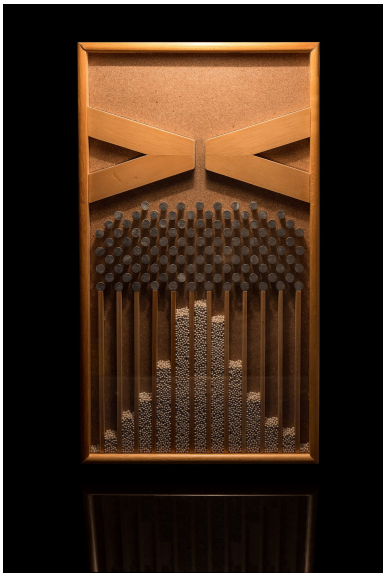
Worked example. $X \sim B(10, 0.3)$. Find $P(X = 2)$ and $E(X)$.

$$P(X = 2) = \binom{10}{2}(0.3)^2(0.7)^8 = 45 \times 0.09 \times 0.05765 = 0.233, \qquad E(X) = 10 \times 0.3 = 3.$$



The distribution $B(10, 0.3)$: each bar is $P(X = r)$, clustered around the mean $np = 3$.

The normal distribution



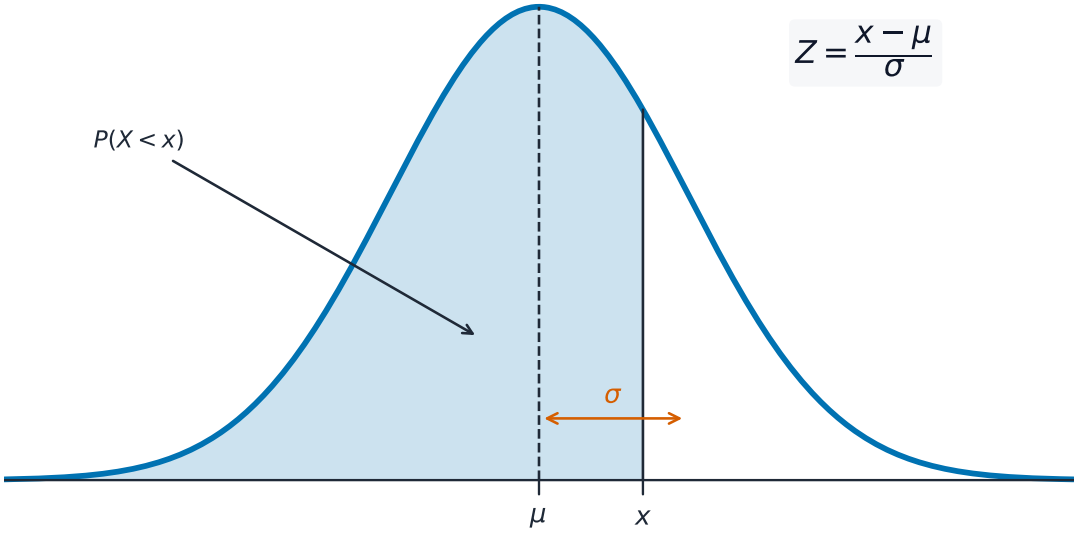
A Galton board: balls falling through pins pile up into the bell-shaped normal distribution.

The **normal distribution** 正态分布 models a **continuous random variable** 连续型随机变量 with a symmetric bell shape. Write $X \sim N(\mu, \sigma^2)$, where μ is the mean and σ is the standard deviation. To use the tables, by **standardisation** 标准

化 convert to the variable $Z \sim N(0,1)$:

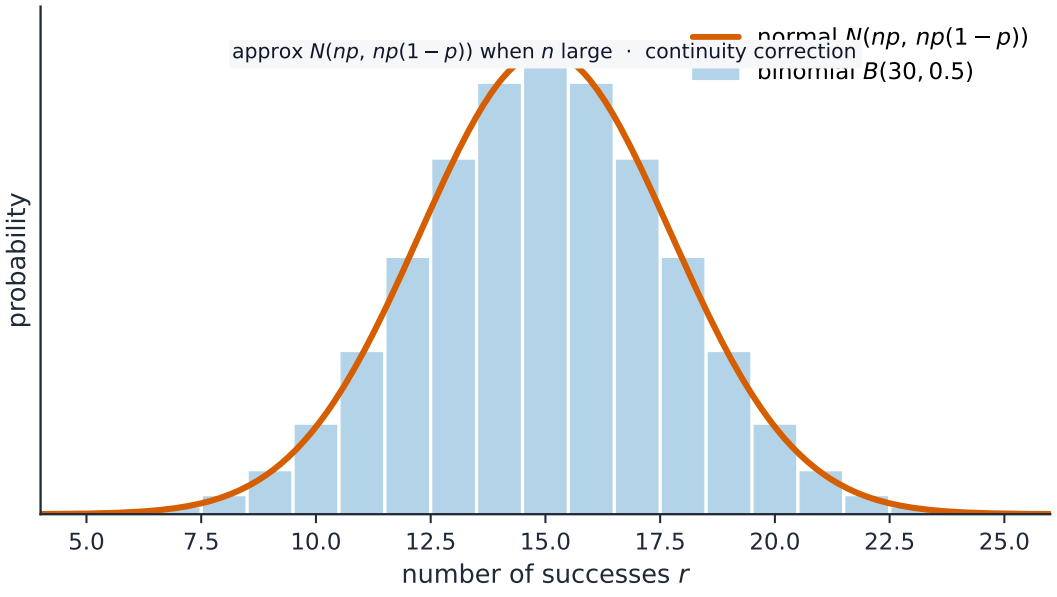
$$Z = \frac{X - \mu}{\sigma}.$$

Then $P(X < x) = P\left(Z < \frac{x - \mu}{\sigma}\right)$, which you read from the normal table Φ.



A normal probability is an area under the bell curve; standardizing rescales it to $Z \sim N(0,1)$.

The normal distribution is also a good **approximation** 近似 to the binomial when n is large. Because you replace a discrete variable by a continuous one, apply a **continuity correction** 连续性校正 (adjust by 0.5).



When n is large the binomial bars follow a normal curve of the same mean and variance.

Worked example. Bags of rice have mass $X \sim N(\mu, 0.14^2)$. Given that $P(X < 1.48) = 0.22$, find μ.

From the table, $P(Z < z) = 0.22$ gives $z = -0.772$. So

$$\frac{1.48 - \mu}{0.14} = -0.772 \;\Rightarrow\; \mu = 1.48 + 0.772 \times 0.14 = 1.59 \text{ kg (3 s.f.)}.$$

Exam tips

- Decide whether order matters: **permutations** (${}^nP_r$) when it does, **combinations** (${}^nC_r$) when it does not.
- For a **discrete random variable**, check the probabilities sum to 1 and use $E(X) = \sum x\,P(X = x)$.
- For the **normal distribution**, standardise with $z = (x-\mu)/\sigma$, **sketch and shade**, then read the table.
- Apply a **continuity correction** when approximating a discrete variable by the normal.

Further Pure Mathematics 1

A-Level Further Mathematics

This handout covers Topic 1: **Further Pure Mathematics** 进阶纯数学 1. It adds new algebra, matrices, polar coordinates, more vectors, and proof by induction.

Roots of polynomial equations

For a polynomial equation, the **roots** 根 are linked to the **coefficients** 系数. For $ax^2 + bx + c = 0$ with roots α, β:

$$\alpha + \beta = -\frac{b}{a}, \qquad \alpha\beta = \frac{c}{a}.$$

For a cubic $ax^3 + bx^2 + cx + d = 0$ with roots α, β, γ:

$$\sum \alpha = -\frac{b}{a}, \qquad \sum \alpha\beta = \frac{c}{a}, \qquad \alpha\beta\gamma = -\frac{d}{a}.$$

Sums like $\sum \alpha$ and $\sum \alpha\beta$ are **symmetric functions** 对称函数 of the roots (unchanged if the roots are swapped). To find an equation whose roots are changed in a simple way, use a **substitution** 代换 (for example, put $w = \alpha + 1$).

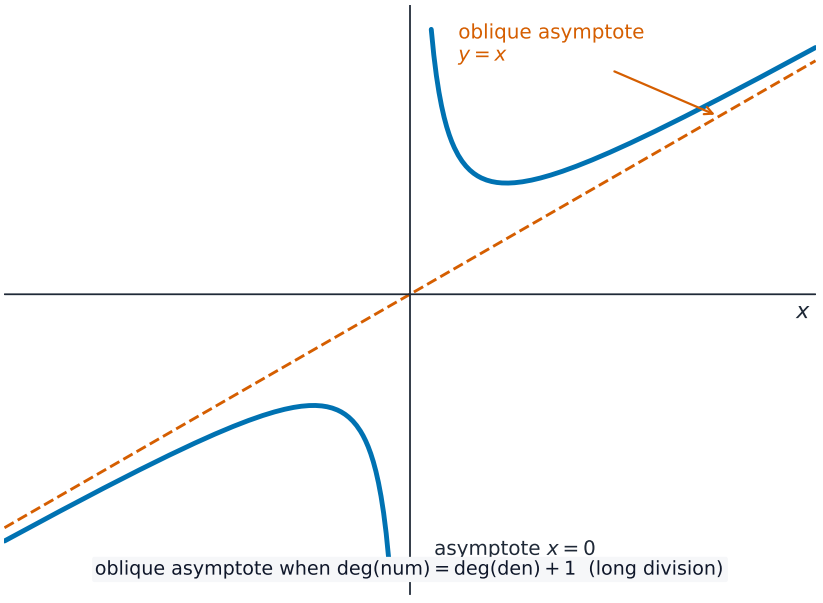
Worked example. The equation $x^2 - 5x + 6 = 0$ has roots α, β. Find the equation with roots $\alpha + 1, \beta + 1$.

Here $\alpha + \beta = 5$ and $\alpha\beta = 6$. The new sum is $(\alpha + 1) + (\beta + 1) = 7$, and the new product is $(\alpha + 1)(\beta + 1) = \alpha\beta + \alpha + \beta + 1 = 6 + 5 + 1 = 12$. So the new equation is

$$x^2 - 7x + 12 = 0.$$

Rational functions and graphs

A **rational function** 有理函数 is a fraction of two polynomials. When the top has degree exactly one higher than the bottom, the graph has an **oblique asymptote** 斜渐近线 (a slanted line the curve approaches); find it by dividing out. You should also be able to relate the graph of $y = f(x)$ to those of $y^2 = f(x)$, $y = \dfrac{1}{f(x)}$, $y = |f(x)|$ and $y = f(|x|)$. Find the set of values the function can take using the **discriminant** 判别式, and locate its **turning points** 驻点.



Far from the origin the curve hugs the slant line $y = x$; near $x = 0$ it runs off along the vertical asymptote.

Worked example. Find the oblique asymptote of $y = \dfrac{x^2 + x + 1}{x + 1}$.

Divide out the fraction: $\dfrac{x^2 + x + 1}{x + 1} = x + \dfrac{1}{x + 1}$. As $x \to \pm\infty$ the remainder $\dfrac{1}{x+1} \to 0$, so the curve approaches the line $y = x$ — the oblique asymptote.

Summation of series

Learn the standard results:

$$\sum_{r=1}^{n} r = \tfrac{1}{2}n(n+1), \qquad \sum_{r=1}^{n} r^2 = \tfrac{1}{6}n(n+1)(2n+1), \qquad \sum_{r=1}^{n} r^3 = \tfrac{1}{4}n^2(n+1)^2.$$

The **method of differences** 差分法 sums a series by cancelling middle terms: if **partial fractions** 部分分式 write each term as $f(r) - f(r+1)$, almost everything cancels. From the sum to n terms you can see whether a series is **convergent** 收敛 and, if so, find its **sum to infinity** 无穷和.

Worked example. Find $\displaystyle\sum_{r=1}^{n}(2r+1)$.

$$\sum_{r=1}^{n}(2r+1) = 2\sum_{r=1}^{n} r + \sum_{r=1}^{n} 1 = 2 \cdot \tfrac{1}{2}n(n+1) + n = n(n+1) + n = n(n+2).$$

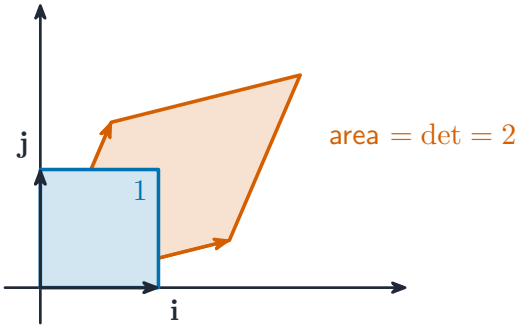
Matrices

A **matrix** 矩阵 is a rectangular block of numbers. **Matrix addition** adds entries in matching positions; you can also subtract and multiply matrices (multiplication is row × column). The **zero matrix** 零矩阵 has every entry 0, and the **identity matrix** 单位矩阵 (or **unit matrix**) I leaves any matrix unchanged under multiplication.

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the **determinant** 行列式 is $\det A = ad - bc$. If $\det A \neq 0$ the matrix is non-singular (otherwise it is a **singular matrix** 奇异矩阵), and the **inverse matrix** 逆矩阵 is

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

For a product, $(AB)^{-1} = B^{-1}A^{-1}$. A 2×2 matrix can represent a **geometric transformation** 几何变换 of the plane — a **rotation** 旋转, **reflection** 反射, **enlargement** 放大, **stretch** 拉伸 or **shear** 切变: the determinant gives the area scale factor, and a product AB means "do B, then A". Points or lines that do not move are called **invariant points** 不变点 and **invariant lines** 不变直线.



The matrix sends the unit square to a parallelogram whose area is the determinant.

Worked example. Find the inverse of $A = \begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix}$.

$\det A = 3 \times 4 - 1 \times 2 = 10$, so

$$A^{-1} = \frac{1}{10} \begin{pmatrix} 4 & -1 \\ -2 & 3 \end{pmatrix}.$$

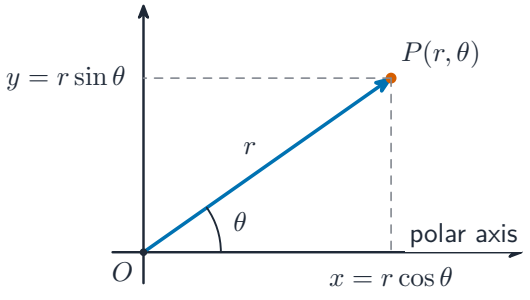
Polar coordinates



A spiral staircase: spirals are described naturally using polar coordinates.

Polar coordinates 极坐标 give a point by its distance r from the origin and its angle θ. They link to **Cartesian coordinates** 直角坐标 by

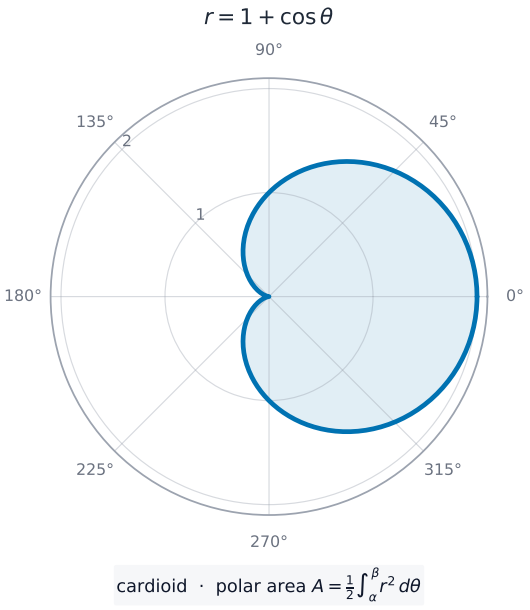
$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad r^2 = x^2 + y^2.$$



Polar coordinates (r,θ) convert to Cartesian by $x = r\cos\theta$ and $y = r\sin\theta$.

You should sketch simple **polar curves** 极坐标曲线, and find the area of a sector with

$$\text{area} = \tfrac{1}{2}\int r^2\,d\theta.$$



The cardioid $r = 1 + \cos\theta$, plotted straight from r as a function of θ.

Worked example. Convert the polar equation $r = 4\cos\theta$ to Cartesian form.

Multiply both sides by r: $r^2 = 4r\cos\theta$, so $x^2 + y^2 = 4x$. Completing the square gives $(x-2)^2 + y^2 = 4$, a circle of radius 2 centred at $(2, 0)$.

Vectors

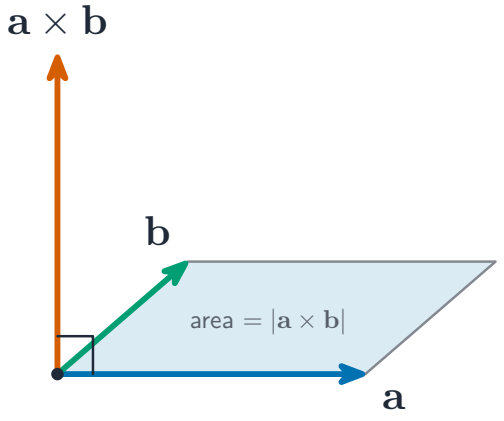


Forces like wind and water are vectors — they have both size and direction.

In three dimensions a **plane** 平面 can be written as $ax + by + cz = d$, or $\mathbf{r} \cdot \mathbf{n} = p$, or $\mathbf{r} = \mathbf{a} + \lambda\mathbf{b} + \mu\mathbf{c}$. Besides the scalar product, there is the **vector product** 向量积 of two **vectors** 向量:

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \sin\theta\, \hat{\mathbf{n}},$$

which gives a vector at right angles to both. With scalar and vector products you can find distances, angles, and where lines and planes meet — including the shortest distance between **skew lines** 异面直线.



$\mathbf{a} \times \mathbf{b}$ *is perpendicular to both vectors; its length equals the area of the parallelogram they span.*

Worked example. Find $\mathbf{a} \times \mathbf{b}$ for $\mathbf{a} = \mathbf{i} + \mathbf{j}$ and $\mathbf{b} = \mathbf{j} + \mathbf{k}$.

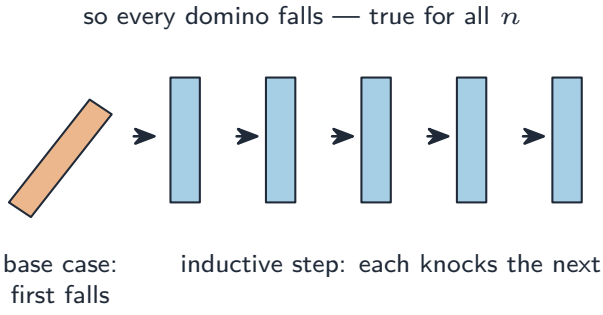
$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = \mathbf{i}(1) - \mathbf{j}(1) + \mathbf{k}(1) = \mathbf{i} - \mathbf{j} + \mathbf{k}.$$

Proof by induction

Mathematical induction 数学归纳法 proves a result for every positive integer n in two steps:

1. **Base case**: show the result is true for $n = 1$.
2. **Inductive step**: assume it is true for $n = k$, then prove it for $n = k + 1$.

If both steps work, the result is true for all n. Often you first make a **conjecture** 猜想 (a sensible guess) from a few cases, then confirm it by **inductive proof** 归纳证明.



Induction is like dominoes: the base case topples the first, and the inductive step makes each one knock over the next — so the result holds for every n

Worked example. Prove that $\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$.

Base case $n = 1$: the left side is 1 and the right side is $\frac{1}{2}(1)(2) = 1$. True. **Inductive step**: assume $\sum_{r=1}^{k} r = \frac{1}{2}k(k+1)$. Then

$$\sum_{r=1}^{k+1} r = \tfrac{1}{2}k(k+1) + (k+1) = (k+1)\left(\tfrac{k}{2}+1\right) = \tfrac{1}{2}(k+1)(k+2).$$

This is the formula with $n = k+1$, so by induction it holds for all n.

Exam tips

- Use the **relationships between roots and coefficients** (sum, sum of pairs, product) instead of solving the polynomial.
- For **proof by induction**, set out the base case, assume the result for $n = k$, prove it for $n = k+1$, and write a clear closing statement.
- Sketch **rational-function graphs** with their asymptotes and turning points; use the **method of differences** for a summation.
- For **matrices**, know the determinant, inverse and invariant lines, and interpret each transformation geometrically.

Further Mechanics

A-Level Further Mathematics

This handout covers Topic 3: **Further Mechanics** 进阶力学. It extends mechanics to projectiles, rigid bodies, circular motion, elastic strings, variable forces and collisions. Take $g = 10\ \text{m s}^{-2}$.

Motion of a projectile

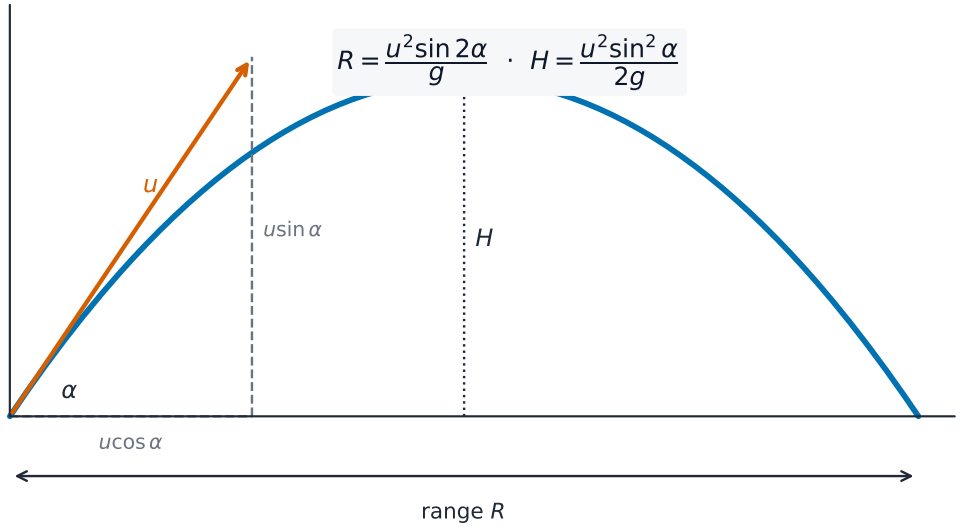


Water jets follow parabolic paths — the classic projectile motion under gravity.

A **projectile** 抛射体 moves freely under gravity, so it has **constant acceleration** 匀加速 g downwards and no horizontal acceleration. Treat the horizontal and vertical motions separately. If it is launched at speed u and angle α:

$$\text{horizontal: } x = u\cos\alpha \cdot t, \qquad \text{vertical: } y = u\sin\alpha \cdot t - \tfrac{1}{2}gt^2.$$

Eliminating t gives the **Cartesian equation** 直角坐标方程 of the **trajectory** 轨迹 (the path), which is a parabola. For a fixed launch speed, all the trajectories lie under a **bounding parabola** 包络抛物线 (the envelope of reachable points). The range on level ground is $\dfrac{u^2\sin 2\alpha}{g}$ and the greatest height is $\dfrac{u^2\sin^2\alpha}{2g}$.



The path is a parabola; the launch speed u splits into a steady horizontal part and a vertical part slowed by gravity.

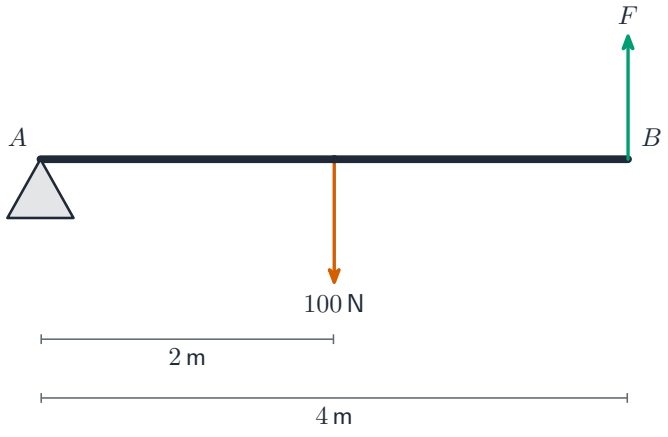
Worked example. A ball is thrown at $u = 20$ m s^{-1} at 30° to the horizontal. Find the range and greatest height.

$$\text{range} = \frac{20^2 \sin 60^\circ}{10} = 40 \times 0.866 = 34.6 \text{ m}, \qquad \text{height} = \frac{20^2 \sin^2 30^\circ}{2 \times 10} = \frac{400 \times 0.25}{20} = 5 \text{ m}.$$

Equilibrium of a rigid body

The **moment** 力矩 of a force about a point is force × perpendicular distance; it measures turning effect. The weight of a body acts at its **centre of mass** 质心, which you can find using **symmetry** 对称 for a uniform flat shape (a **lamina** 薄片), or by treating a composite body as a set of particles.

A rigid body under **coplanar forces** 共面力 is in **equilibrium** 平衡 when two conditions both hold: the vector sum of the forces is zero, and the sum of the moments about any point is zero. A body may also be on the edge of **toppling** 翻倒 (turning over) or **sliding** 滑动 (slipping).



Taking moments about the pivot A balances the turning effects: $F \times 4 = 100 \times 2$.

Worked example. A uniform beam AB of length 4 m and weight 100 N rests on a pivot at A. A vertical force F at B keeps it horizontal. Find F.

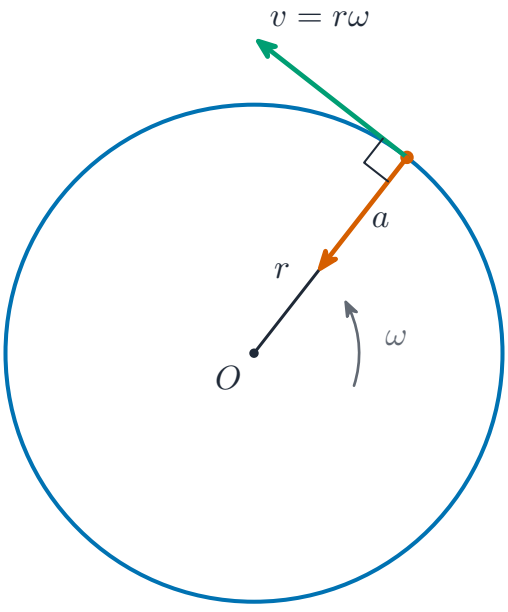
Take moments about A (the weight acts at the centre, 2 m from A):

$$F \times 4 = 100 \times 2 \;\Rightarrow\; F = 50 \text{ N}.$$

Circular motion

For a particle moving in a circle of radius r, the **angular speed** 角速度 ω links to the speed by $v = r\omega$. The acceleration points towards the centre — the **centripetal acceleration** 向心加速度 — with size

$$a = r\omega^2 = \frac{v^2}{r}.$$



The velocity points along the tangent; the acceleration points inward to the centre.



The hammer thrower is a moving diagram of this page. The taut wire pulls the ball toward the centre — that inward pull is the centripetal force — while the ball's velocity points along the tangent. Let go, and with no inward force left the ball flies straight off that tangent

In a **horizontal circle** 水平圆 the speed is constant — as in a **conical pendulum** 圆锥摆 (a mass swung on a string). In a **vertical circle** 竖直圆 use energy conservation, because the speed changes with height; the **normal contact force** 法向接触力 or

string tension provides the centripetal force.

Worked example. A particle moves in a horizontal circle of radius 2 m with angular speed 3 rad s^{-1}. Find its speed and acceleration.

$$v = r\omega = 2 \times 3 = 6 \text{ m s}^{-1}, \qquad a = r\omega^2 = 2 \times 3^2 = 18 \text{ m s}^{-2}.$$

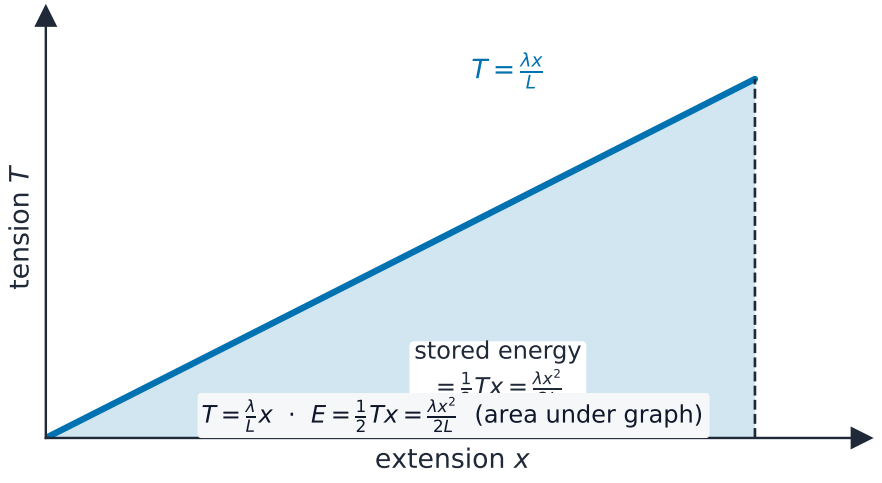
Hooke's law

Hooke's law 胡克定律 says the tension in an elastic string or spring is proportional to its extension x:

$$T = \frac{\lambda x}{L},$$

where L is the natural length and λ is the **modulus of elasticity** 弹性模量. Stretching stores **elastic potential energy** 弹性势能:

$$E = \frac{\lambda x^2}{2L}.$$



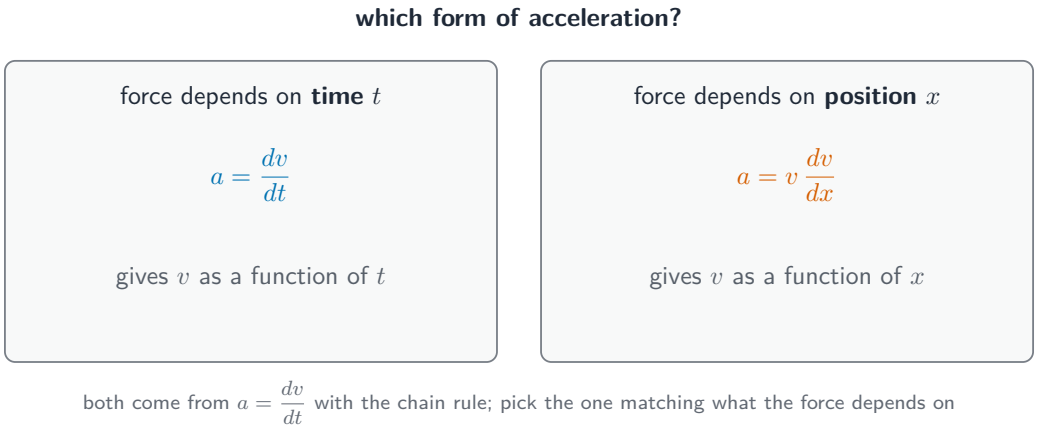
Tension rises in proportion to extension; the shaded triangle is the stored elastic energy.

Worked example. An elastic string of natural length 2 m and modulus 50 N is stretched by 0.5 m. Find the tension and the stored energy.

$$T = \frac{50 \times 0.5}{2} = 12.5 \text{ N}, \qquad E = \frac{50 \times 0.5^2}{2 \times 2} = 3.125 \text{ J}.$$

Linear motion under a variable force

When the force depends on position x, use acceleration in the form $a = v\dfrac{dv}{dx}$, which turns Newton's law into a **differential equation** 微分方程 relating v and x.



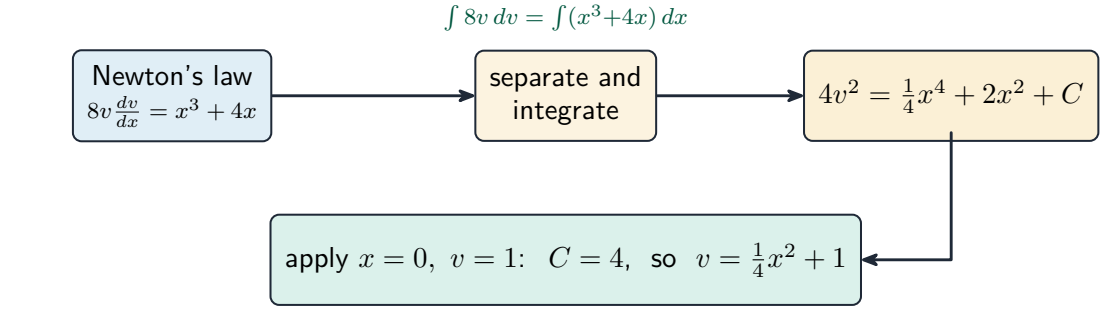
Use dv/dt when the force depends on time, v dv/dx when it depends on position

Worked example. A particle of mass 8 kg moves along a line under a **variable force** 变力 of size $(x^3 + 4x)$ N acting in the direction of motion. When $x = 0$, $v = 1$. Find v in terms of x.

Newton's law gives $8v\dfrac{dv}{dx} = x^3 + 4x$. Separating and integrating:

$$\int 8v\,dv = \int (x^3 + 4x)\,dx \;\Rightarrow\; 4v^2 = \tfrac{1}{4}x^4 + 2x^2 + C.$$

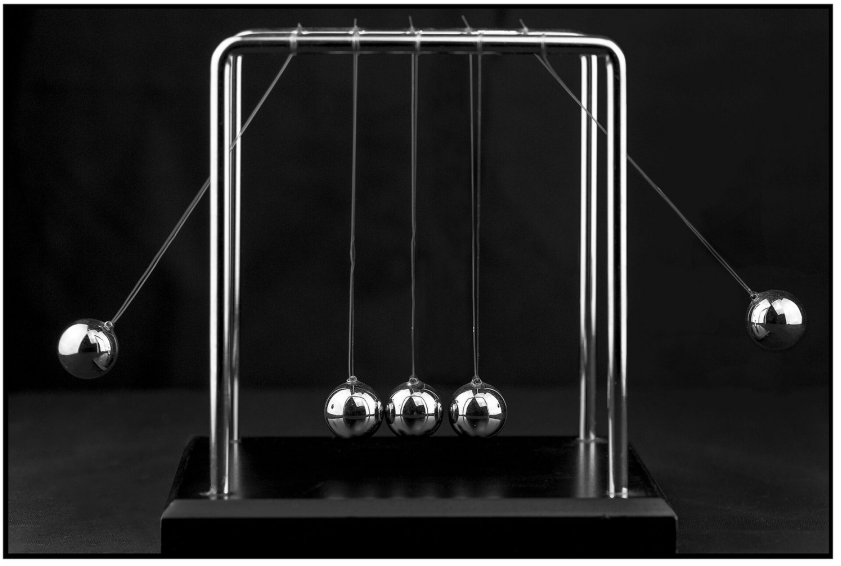
At $x = 0$, $v = 1$ gives $C = 4$, so $4v^2 = \frac{1}{4}x^4 + 2x^2 + 4 = 4\left(\frac{x^2}{4} + 1\right)^2$. Hence $v = \frac{1}{4}x^2 + 1$.



a position-dependent force becomes a **differential equation** in v and x

A position-dependent force becomes a differential equation in v and x

Momentum

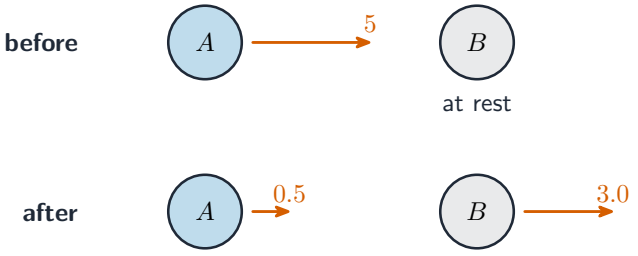


A Newton's cradle demonstrates conservation of momentum in collisions.

In a collision, total **linear momentum** is conserved — the **conservation of linear momentum** 动量守恒. The bounciness is measured by **Newton's experimental law** 牛顿实验定律, which defines the **coefficient of restitution** 恢复系数 e:

$$e = \frac{\text{speed of separation}}{\text{speed of approach}}, \qquad 0 \leqslant e \leqslant 1.$$

Here $e = 1$ is perfectly elastic (no energy lost) and $e = 0$ is inelastic (the bodies stay together). In an **oblique impact** 斜碰撞, resolve the velocities along and perpendicular to the line of impact, applying restitution along it.



$$e = \frac{\text{speed of separation}}{\text{speed of approach}} = \frac{3.0 - 0.5}{5} = 0.5$$

The coefficient of restitution e compares how fast the bodies separate with how fast they approached.

Worked example. A sphere A of mass 2 kg moving at 5 m s^{-1} hits a stationary sphere B of mass 3 kg, with $e = 0.5$. Find the speeds afterwards.

Momentum: $2(5) = 2v_A + 3v_B$, so $2v_A + 3v_B = 10$. Restitution: $v_B - v_A = 0.5(5) = 2.5$. Solving together gives $v_A = 0.5$ m s^{-1} and $v_B = 3.0$ m s^{-1}.

Exam tips

- For **projectiles**, resolve into horizontal (constant velocity) and vertical ($a = g$) motion, linked by the same time.
- For a **rigid body in equilibrium**, take moments about a point that removes an unknown force.
- For **circular motion**, use $F = mv^2/r = m\omega^2 r$ towards the centre; in a vertical circle check the minimum speed at the top.
- With a **variable force**, use $a = v\,\frac{dv}{dx}$ and integrate; elastic PE $= \frac{\lambda x^2}{2L}$.

Further Probability & Statistics

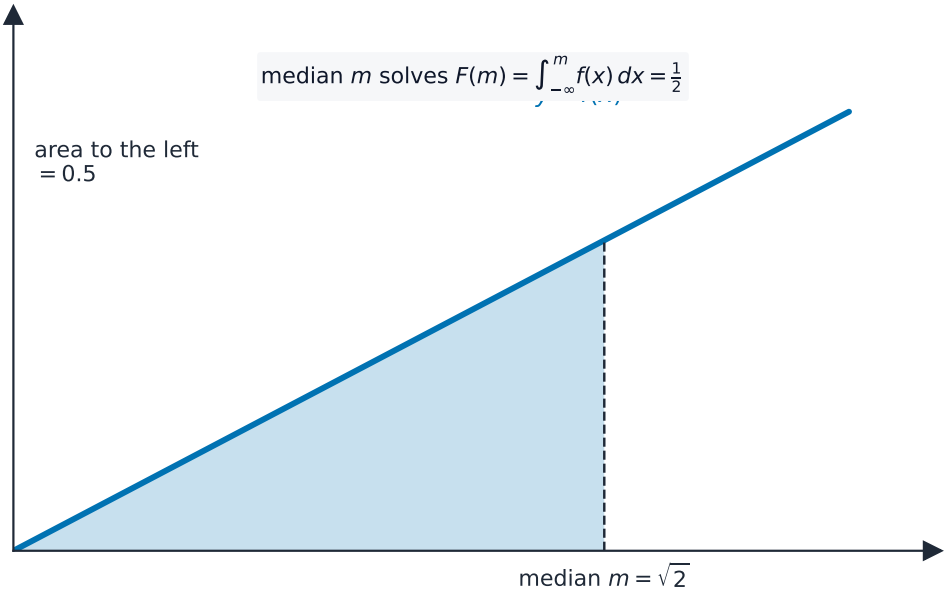
A-Level Further Mathematics

This handout covers Topic 4: **Further Probability & Statistics** 进阶概率统计. It adds continuous distributions, small-sample inference, the chi-squared and non-parametric tests, and probability generating functions.

Continuous random variables

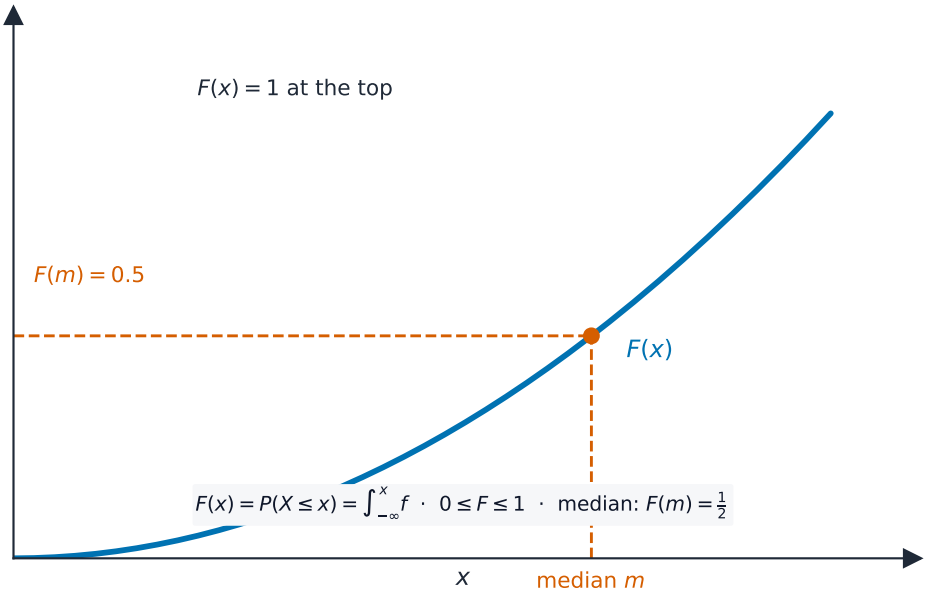
A continuous variable X is described by a **probability density function** 概率密度函数 $f(x)$, which may be defined piecewise. The probability over a range is the area under f, and the mean of any function of X is

$$E(g(X)) = \int g(x)\, f(x)\, dx.$$



The median sits where the area under $f(x)$ to its left is exactly 0.5.

The **cumulative distribution function** 累积分布函数 $F(x) = P(X \leqslant x)$ is the running total: $F(x) = \int_{-\infty}^{x} f(t)\, dt$, and $f(x) = F'(x)$. Use F to find probabilities and **percentiles** 百分位数 (for example, the **median** 中位数 solves $F(x) = 0.5$).



The cumulative graph $F(x)$ rises from 0 to 1; the height 0.5 is reached at the median.

Worked example. A variable has $f(x) = \frac{1}{2}x$ for $0 \leqslant x \leqslant 2$. Find the median.

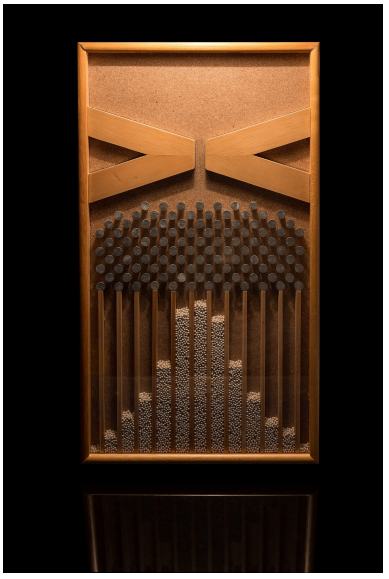
The cumulative distribution function is $F(x) = \displaystyle\int_0^x \tfrac{1}{2}t\,dt = \tfrac{1}{4}x^2$. Set $F(m) = 0.5$:

$$\tfrac{1}{4}m^2 = 0.5 \;\Rightarrow\; m^2 = 2 \;\Rightarrow\; m = \sqrt{2} = 1.41.$$

The distribution of a related variable. If $Y = g(X)$, find Y's distribution through its cumulative function. For $Y = X^2$: $F_Y(y) = P(X^2 \leqslant y) = P(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$, then differentiate for $f_Y = F_Y'$. When g is monotonic increasing there is a shortcut, $F_Y(y) = F_X\big(g^{-1}(y)\big)$.

Worked example. With $f_X(x) = \frac{1}{2}x$ on $[0, 2]$ (so $F_X(x) = \frac{1}{4}x^2$), let $Y = X^2$. For $0 \leqslant y \leqslant 4$, $F_Y(y) = F_X(\sqrt{y}) = \frac{1}{4}y$, so $f_Y(y) = F_Y'(y) = \frac{1}{4}$ – that is, Y is uniform on $[0, 4]$.

Inference using normal and t-distributions

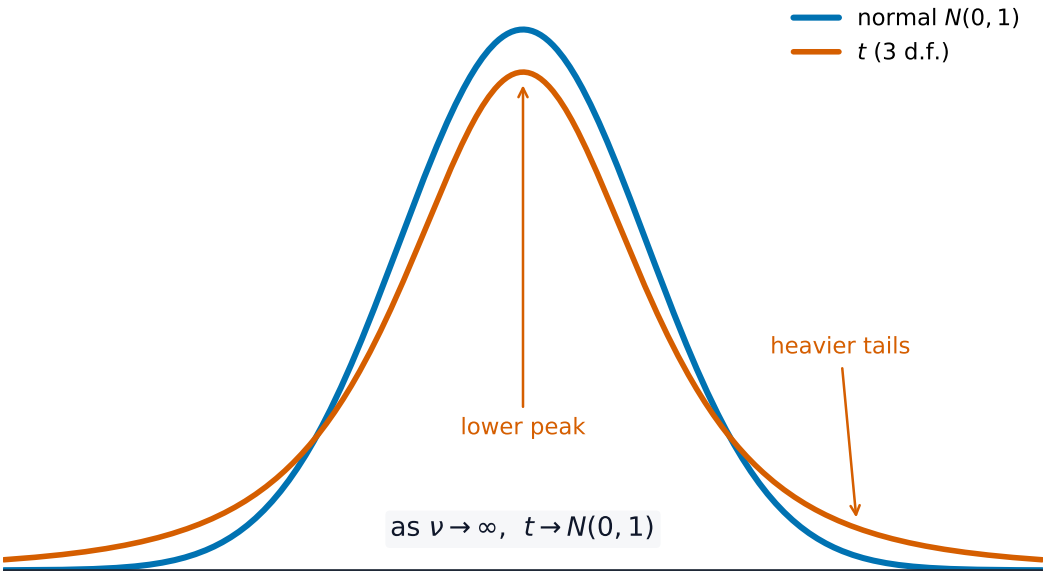


A Galton board: balls falling through pins pile up into the bell-shaped normal distribution.

When a sample is small and the population variance is unknown, base your **hypothesis test** 假设检验 on the t-distribution instead of the normal. The same idea gives a **confidence interval** 置信区间 for the mean:

$$\bar{x} \pm t \frac{s}{\sqrt{n}},$$

where t comes from the t-tables with $n - 1$ degrees of freedom. To compare two populations, use a **two-sample** (2-sample) or **paired-sample** t-test, after finding a **pooled estimate** 合并估计 of the shared variance when appropriate.



For a small sample the t-distribution is flatter with heavier tails, so its critical values are larger.

Worked example. A sample of $n = 10$ has mean $\bar{x} = 50$ and standard deviation $s = 4$. Find a 95% confidence interval for the mean (use $t = 2.262$ for 9 degrees of freedom).

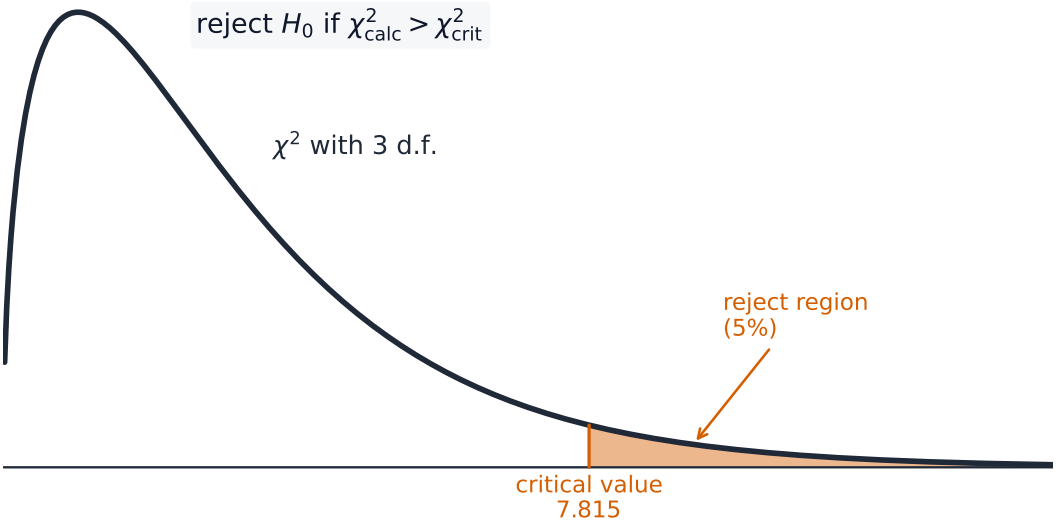
$$50 \pm 2.262 \times \frac{4}{\sqrt{10}} = 50 \pm 2.86 \;\Rightarrow\; (47.1,\ 52.9).$$

Chi-squared tests

A χ^2-test (**chi-squared test** 卡方检验) compares observed counts O with expected counts E from a **theoretical distribution** 理论分布:

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

Compare this with a table value for the right number of **degrees of freedom** 自由度. Two uses: a **goodness of fit** 拟合优度 test (does the data follow the proposed model?), and a test for **independence** 独立性 of two variables in a **contingency table** 列联表.



The test rejects the model when χ^2 exceeds the critical value, landing in the shaded 5% tail.

Worked example. Four equally likely categories give observed counts $20, 30, 25, 25$ (so each expected count is 25). Test the fit at the 5% level.

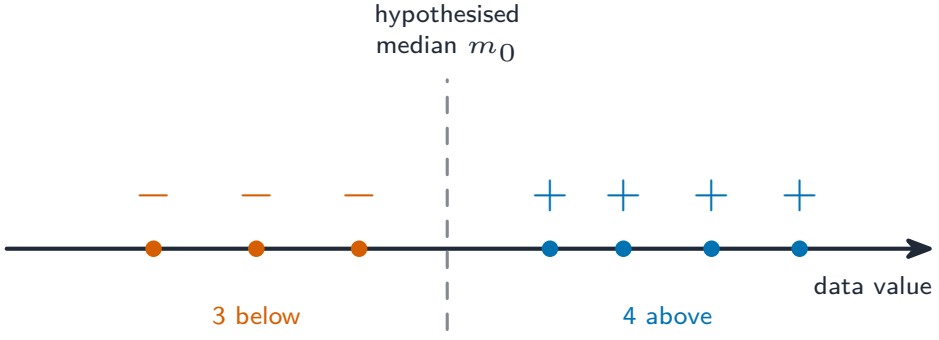
$$\chi^2 = \frac{(20-25)^2 + (30-25)^2 + 0 + 0}{25} = \frac{25+25}{25} = 2.$$

With $4 - 1 = 3$ degrees of freedom the table value is 7.815. Since $2 < 7.815$, do not reject the model.

Non-parametric tests

A **non-parametric test** 非参数检验 makes no assumption that the data is normal, so it is useful when that assumption fails. The basic ones are:

- the **sign test** 符号检验: count how many values fall above and below a proposed median, and test those counts with a binomial model;
- the **Wilcoxon signed-rank test** 威尔科克森符号秩检验 (the **matched-pairs** test for paired data), which also uses the sizes of the differences, not just their signs;
- the **Wilcoxon rank-sum test** 威尔科克森秩和检验, for comparing two separate samples.



The sign test counts how many values fall above and below the proposed median, then tests those counts with a binomial model

Worked example. Test whether a median is 5. In a sample of 10 values (none equal to 5), 9 lie **above** 5 and 1 lies below. Test at the 5% level (two-tailed).

Under H_0 (median $= 5$) the number above follows $B(10, 0.5)$. The observed result (9 above) is extreme, so find $P(X \geq 9) = \binom{10}{9}(0.5)^{10} + (0.5)^{10} = \dfrac{11}{1024} = 0.0107$. For a two-tailed test compare with $\frac{1}{2}(5\%) = 0.025$. Since $0.0107 < 0.025$, **reject** H_0: there is evidence the median is not 5.

Probability generating functions



Dice: a starting point for probability and discrete random variables.

The **probability generating function** 概率母函数 of a discrete variable X is

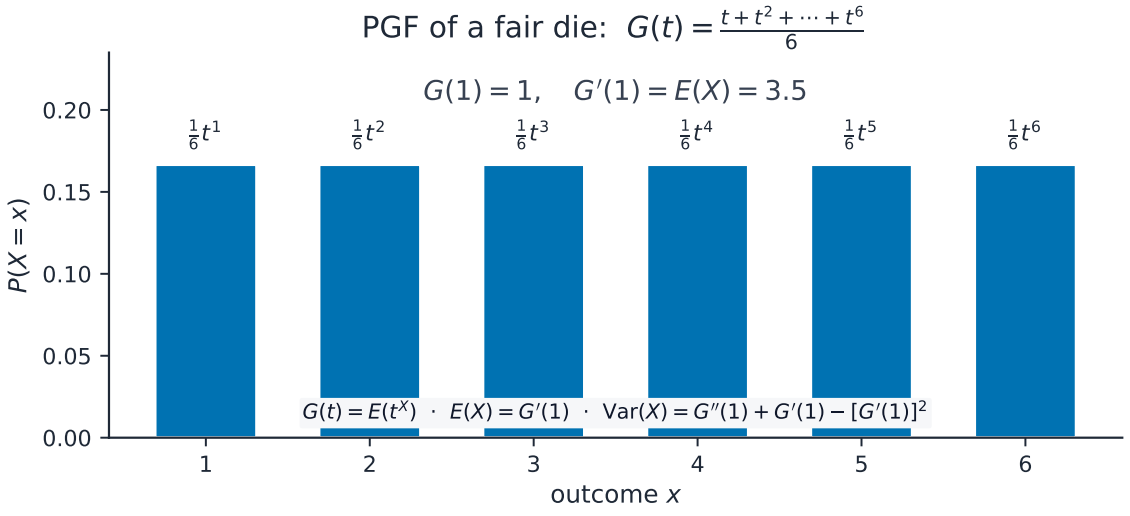
$$G(t) = E(t^X) = \sum_x P(X = x)\, t^x.$$

It packs the whole distribution into one function. The mean and variance come from its derivatives at $t = 1$: $E(X) = G'(1)$ and $\mathrm{Var}(X) = G''(1) + G'(1) - \big(G'(1)\big)^2$. Also, the PGF of a sum of independent variables is the product of their PGFs.

Worked example. X has $P(X = 0) = 0.5$, $P(X = 1) = 0.3$, $P(X = 2) = 0.2$. Find $E(X)$ using the PGF.

Here $G(t) = 0.5 + 0.3t + 0.2t^2$, so $G'(t) = 0.3 + 0.4t$ and

$$E(X) = G'(1) = 0.3 + 0.4 = 0.7.$$



PGF of a fair die packs equal masses at 1 through 6 into G(t)

Exam tips

- For a **continuous random variable**, the pdf integrates to 1 over its range, and $E(X) = \int x f(x)\,dx$.
- Use the **t-distribution** when the sample is small and the population variance is unknown; state the degrees of freedom.
- For a **chi-squared** test compute $\sum (O - E)^2 / E$, compare with the critical value at the right degrees of freedom, and **combine classes with** $E < 5$.
- State H_0 and H_1 and give the conclusion **in context** for every test.