Preparation Course

Physics

Disclaimer

According to Belgian Law ("Wetboek Economisch Recht"), Olivier Loose is the rightful author of this work, whereby all rights are reserved. No part of this preparation course may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without prior permission in writing from the author.

The use of this work is furthermore restricted to educational and informational purposes only and any commercial aspect with respect to its use is a violation of the Law.

All the images in this preparation course are created by the author himself with the assistance of Microsoft PowerPoint and no external images (from the public domain or otherwise) were used. Any numerical examples provided are equally original creations of the author. The preparation course itself has been written with the help of the freely available text editing software LaTeX.

Copyright © Olivier Loose 2022

Contents

1 Electricity

1.1 Electrostatics

The field of electrostatics describes the behaviour of **electric charges at rest**—moving charges is what constitutes electricity and is the topic of section 1.2. Being at rest means that the charges, which can be positive, neutral, or negative, have reached a *static equilibrium*. An example of a negatively charged particle is an electron, while the proton and the neutron, which both sit at the atom's nucleus, are an example of a positively and neutrally charged particle, respectively.

The notion of **charging** usually refers to the gain or loss of electrons. That is, the surface of an object that is initially neutral in charge can become negatively charged when it accumulates electrons, or positively charged (i.e., less negative) when it loses electrons. Such a temporary transfer of electrons can be accomplished by simple contact or friction (rubbing) between objects. After the objects have been separated again, one will be negatively charged and the other positively charged.

When a material does not allow the electrons to circulate freely across its surface and instead holds the charges together at a specific location, the object is said to be an **insulator**, i.e., a poor conductor of electrons. In other words, the buildup of charge on the surface of an insulator is locally concentrated rather than uniformly distributed.

Examples of insulators include plastic, glass, wool, paper and rubber. Some materials are more inclined to attract electrons than others, so that when placed into contact with other materials, they are quicker to steal electrons from the other object. This means that after the period of contact has ended, their surface will be negatively charged.

In table 1.1, the most common insulators are listed, together with the strength of their tendency to either give up electrons (and become positive in charge) or attract electrons (and become negative in charge).

Table 1.1: Insulating Materials

Material	Charge after contact	Material	Charge after contact
Air	+++	Silicone Rubber	− − −
Dry Skin	+++	Plastic	− − −
Rabbit Fur	++	Polyester	− −
Glass	++	Rubber Balloons	−
Human Hair	++	Hard Rubber	−
Wool	+	Wood	−
Paper	+		

For instance, when combing our hair, the comb made out of hard rubber or plastic will be negatively charged while our hair will be positively charged. Another example is when we have dry skin and wearing polyester clothes: touching our clothes results in having a positively charged skin and negatively charged clothes. After a charge is present, we will feel a spark (i.e., the *electrostatic discharge* or rebalancing of the charges) when touching our clothes again. A final example refers to the situation where we rub a plastic rod with rabbit fur or wool (the rod gets negatively charged) or a glass rod with wool (the rod gets positively charged, as it has a greater tendency (++) to get rid of electrons than wool (+)).

The electric charges exert a force, called the **Coulomb force**, upon one another: equal charges repel each other while opposite charges attract (see Fig. 1.1). In other words, our hair will be attracted to the hard rubber comb, the plastic rod will be able to attract small pieces of paper, and our strands of hair will try to remain as far as possible away from each other after combing, especially if it is dry (giving us "flyaway hair").

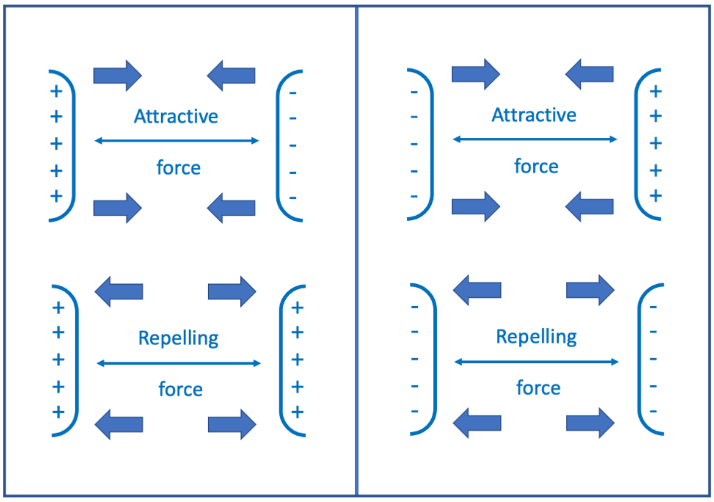


Figure 1.1: The Coulomb force between electric charges

In the examples stated above, the amount of accumulated charge is minimal, so that the energy associated with the electrostatic discharge is also minimal. The **electrostatic discharge** refers to the event between two *differently* charged objects whereby, upon contact, the excess of electrons in the negatively charged object flows towards the positively charged object.

However, when that charge difference becomes large and the two objects are put close enough together, the spontaneous discharge that follows can cause lots of damage—this difference in charge is called the voltage or the electric potential difference (see section 1.2). Bear in mind that what is essential here is the *difference* in charge; that is, objects with equal charges cannot produce any electrostatic discharge.

In many industries, these sparks of electricity constitute a safety hazard, as they might ignite flammable or explosive materials, triggering fires or explosions in the process. Even by walking around people build up electric charges (shoes have rubber soles), so that coming into contact with machines could lead to a spark, possibly setting ablaze inflammable materials.

One of the most important methods to minimize the buildup of electric charges is the phenomenon of **earthing or grounding**. This safety measure entails that an electrically charged object is connected via cables, wires, or clips to a charge-conducting rod stuck in the ground, so that the excess of electrons can flow into the Earth, keeping in this way the difference in electrostatic charge between objects to a minimum. It is hereby important to ensure that the resistance of the path towards the ground is minimal.

1.2 Electric Circuits

Electricity basically implies the movement of electrons (negatively charged particles). The unit in which this flow of electricity (the electric current) is expressed, i.e., ampere (A), is one of the 7 fundamental SI (Système International d'unités) units—the others being kilogram, meter, second, Kelvin, candela, and mole. The **electric current I** is defined as the amount of charge Q that flows in a certain period of time t:

$$I = \frac{Q}{t} \tag{1.1}$$

with I expressed in ampere (A), Q in coulomb (C), and t in seconds. An elementary charge Q is measured as 1.6×10^{-19}C.

Electric current is the result of electrons (which orbit around an atom's nucleus) crossing from one atom of a conducting material to the next and so on. That is, a **conductor** allows the electrons to move freely between its atoms, so that any charge present is distributed uniformly across its surface. Examples include metals (e.g., silver, copper, aluminum, etc.), aqueous solutions of salts, water, and the human body (depending on humidity levels).

In contrast, an **insulator** prohibits the free movement of electrons, whereby any charge present will be situated locally on its surface instead of spreading out. Examples include dry air, glass, rubber, etc. (see section 1.1).

An electric current only flows in a *closed* electric circuit, which consists fundamentally of three main parts:

1. An **energy source**. This is usually a chemical battery, which is a device that converts chemical energy into electric energy by means of an electrochemical cell—a battery can contain one or more of these cells. The chemistry involved within a cell is a redox reaction, whereby the produced electrons at the anode (−) side of the cell (oxidation) flow via an external path towards the cathode (+) side (reduction). The anode and cathode are called electrodes.

2. A **path** that is made of a conducting material to sustain the flow of electrons.

3. An **appliance** that harvests the electricity (e.g., a light bulb). A light bulb is a kind of resistor, which is a component in an electric circuit that restricts the flow of electricity. Due to the very thin wire (the filament) that is housed within the light bulb, the electrons experience more friction (resistance) as they bump into each other more frequently, moving therefore slower, so that the energy of their movements (the kinetic energy) is transformed into heat energy, which is in turn responsible for the glowing of the light bulb.

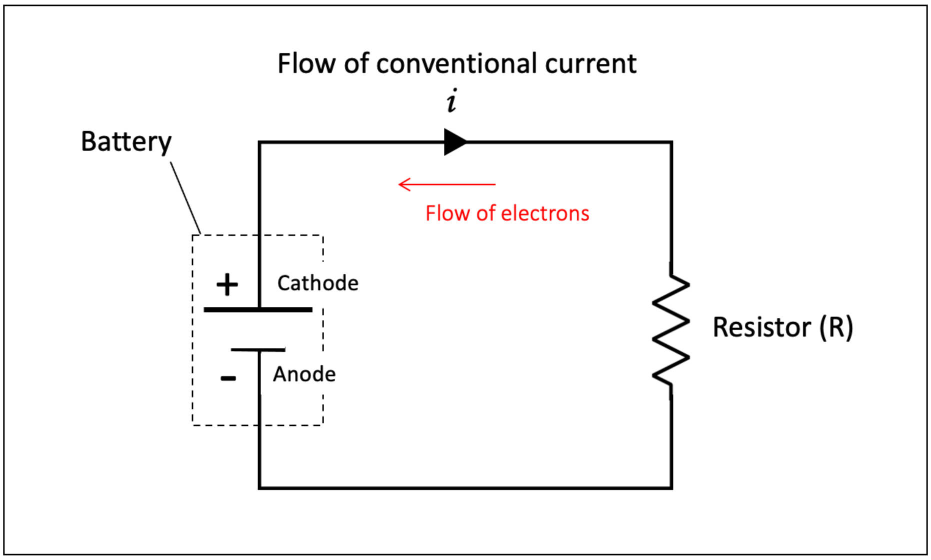


Figure 1.2: The Electric Circuit and its main components

Fig. 1.2 shows that, as stated previously, the electrons flow from the anode (−) to the cathode (+) via the circuit. However, given that electric charge can take on a negative as well as a positive value, the general convention is to identify in a diagram of an electric circuit the **direction of the current *i*** as the direction in which the *positive* charges flow, i.e., from the cathode to the anode (we have to thank Ben Franklin for this unfortunate convention).

A battery also comes with a characteristic referred to as **voltage**, which is the difference between the electric potential of the two electrodes. The electric potential is the amount of electric potential energy per unit charge, whereby we can think of the electric potential energy as a kind of stored energy in an electrochemical reaction, or, if you will, the strength of that reaction to produce or attract electrons. Ideally, the electric potential of the anode must be much lower (more negative, as it needs to produce electrons) than the electric potential of the cathode (which needs strength to attract the electrons). The electric potential and the voltage can thus be defined as follows:

$$\begin{cases} V = \frac{E}{Q} \\ \\ \Delta V = \frac{\Delta E}{Q} \end{cases} \tag{1.2}$$

with V the electric potential expressed in volt (V), ΔV the voltage (in V), E (ΔE) the (difference in) electric potential energy (in Joule J), and Q the electric charge (in coulomb C). By convention, the common symbols that we may encounter to describe voltage include ΔV, V, or sometimes U.

We can increase the voltage of a battery by either stacking more electrochemical cells on top of each other within a battery or changing the type of material from which the electrodes are fashioned. Also note that recharging the battery reverses the role of the electrodes in an electrochemical cell: the anode becomes positive and the cathode negative.

The electric current furthermore comes in two types: **direct current (DC) and alternating current (AC)**. The former implies a *unidirectional* flow of electric charges, meaning that the current can only flow in one direction. As a result, the polarity of the voltage remains unchanged. DC is primarily used in batteries, solar cells, and some railway systems.

Quite the opposite behaviour is observed with AC: the alternating electric current reverses its direction in a *periodical* way whereby its magnitude continuously evolves with time, causing thus the voltage to also change accordingly. This means that the waveform of AC when plotted in a diagram is a sine wave (more about waves in section 6). It is this type of current that comes out of the wall sockets in our houses. Remark that AC cannot be stored (unlike DC) and is less dangerous than DC (with DC there is a constant current, while with AC it periodically measures 0V).

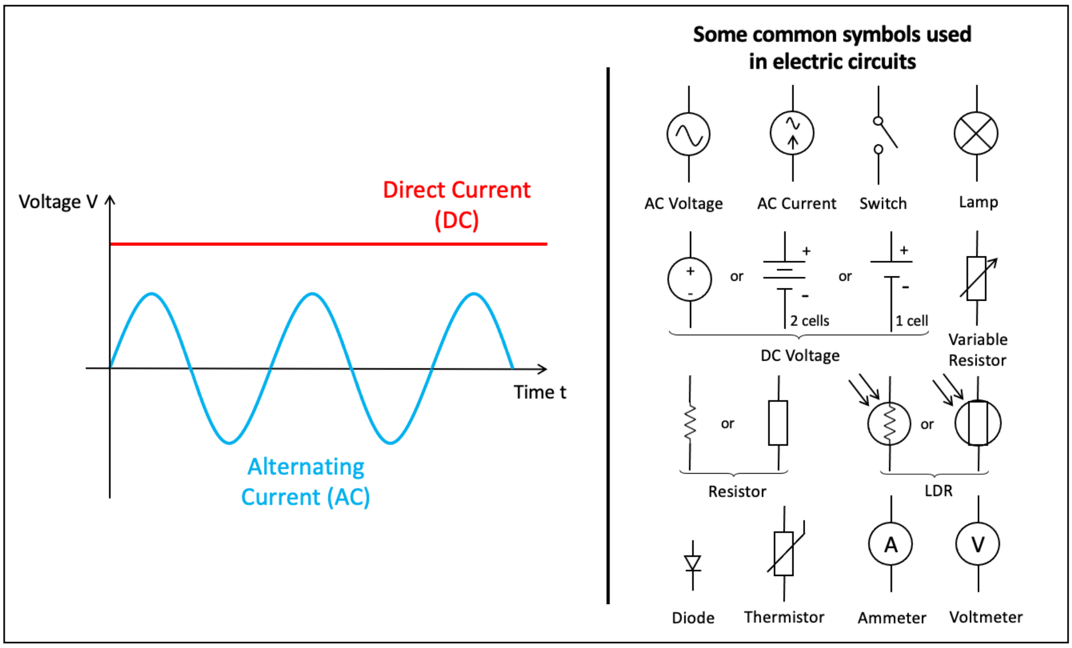


Figure 1.3: AC versus DC and some common circuit symbols

Apart from the basic elements (as depicted in Fig. 1.2), an electric circuit can contain additional components (see Fig. 1.3). A **switch** determines whether electricity flows through the circuit and can be opened or closed. While a common resistor has a fixed value, a **variable resistor** allows its value to be adjusted. Another type of resister is called a **light-dependent resistor (LDR)**, a.k.a. a photocell or photoconductor, which is commonly used in street lights. When daylight reaches the resistor, the conductivity of the electrons increases, thereby reducing the resistance (this prohibits the filament of the street light to become heated and emit light). At night, the opposite scenario takes place: the lack of light enhances resistance, switching on the lights. An **ideal diode** allows a current to flow *only* when a positive voltage is applied across the electrodes; if we reverse the voltage, the diode blocks the current. A resistor whose functioning is heavily temperature-dependent is known as a **thermistor**; the resistance of a *negative temperature coefficient (NTC) thermistor* weakens with rising temperatures, since more heat means greater thermal agitation of the atom's electron shells, thereby freeing up more traveling electrons (and thus manifesting less resistance). A *positive temperature coefficient (PTC) thermistor* works just the other way around. That is, the resistance builds up with increasing temperatures. Finally, a **voltmeter** measures the voltage (the difference in electric potential) between two points in an electric circuit, whereas an **ammeter** indicates the current. The way in which they are connected to the circuit differs: the voltmeter is hooked up in parallel, i.e., next to or parallel to the circuit, whereas the ammeter in series, i.e., in line with the circuit.

An important relationship between the current, the voltage, and the resistance is reflected by **Ohm's law**, which states that for a given resistance R, the current I is directly proportional to the voltage V:

$$I = \frac{V}{R} \tag{1.3}$$

with I expressed in ampere (A), V in volt (V), and R in ohm (Ω). It is instructive to remark at this point that even though a light bulb exhibits similar characteristics to a resistor, it does not always follow Ohm's law given that its resistance varies with temperature and voltage. More precisely, a filament lamp only follows Ohm's law for small values of V and I, while at large values, V and I push the temperature up, which in turn boosts the resistance R (and this behaviour is no longer described by Ohm's law). So, if we want to maintain a constant resistance in our circuit, we are advised to implement a resistor instead of a light bulb.

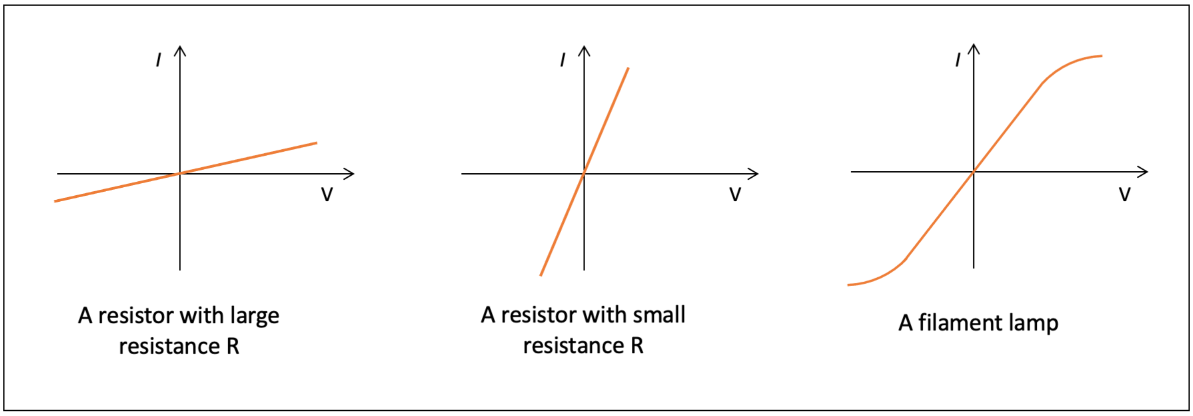


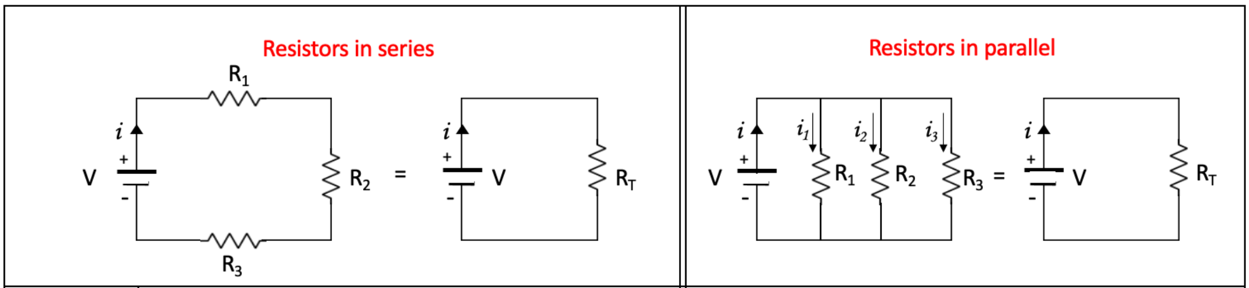
Figure 1.4: Ohm's Law in the context of a resistor and a filament lamp

Ohm's law is a useful tool to calculate the different values of the current, the voltage, and the resistance, when we consider electric circuits with multiple resistors, placed either **in series or in parallel**. Below in Fig. 1.5, an overview is provided of the rules to identify the current, voltage, and resistance in both series and parallel electric circuits.

Let us consider an example whereby V = 12.0V, $R_1 = 1.00\Omega$, $R_2 = 6.00\Omega$, and $R_3 = 13.0\Omega$. For the resistors positioned in series, we obtain the following values: $R_T = 20.0\Omega$, $i = 0.60A$, $V_1 = 0.60V$, $V_2 = 3.60V$, and $V_3 = 7.80V$.

For the resistors in parallel, we get these values: $R_T = 0.80\Omega$, i=14.9A, i_1=12.0A, i_2=2.00A, and i_3=0.92A.

From the rules for a *parallel* electric circuit, it follows that the calculated total resistance R_T is always lower than that of any other individual resistor in a parallel-connected circuit. Therefore, an easy trick to ensure that our calculations are on the right track is to make sure that R_T is lower than the smallest resistance value among all individual resistors. In the case of our example, we find indeed that $R_T = 0.80\Omega < R_1 = 1.00\Omega$.



Current	The current i is the same through every resistor	$i = i_1 + i_2 + i_3 = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$ $= V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) = \frac{V}{R_T}$
Voltage	$V = V_1 + V_2 + V_3$ $= iR_1 + iR_2 + iR_3$ $= i(R_1 + R_2 + R_3)$ $= iR_T$	The voltage V is the same across every resistor
Resistance	$R_T = R_1 + R_2 + R_3$	$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$

Figure 1.5: The rules for current, voltage, and resistance in series and parallel circuits

Both the voltage and the current are crucial to establish the rate at which a battery can perform some work. That rate is defined as the battery's **power**. Therefore, work can be interpreted as the amount of **energy E** that the battery, endowed with a certain power P, can deliver or "transfer" for a certain period of time t. We then obtain the following equations:

$$P = IV = I^2R \tag{1.4}$$

$$\Delta E = Pt = VIt \tag{1.5}$$

with P the power expressed in Watt (W) and ΔE the energy transferred in Joule (J). The second step in Equation 1.4 makes use of Ohm's law (see Equation 1.3), and the second step in Equation 1.5 relies on Equation 1.4.

2 Magnetism

2.1 Properties of Magnets

Any object that possesses an intrinsic **magnetic field** is referred to as a magnet, such as a horseshoe magnet, a bar magnet, and the Earth. We can think of a field as an imaginary sheet stretched infinitely across space whereby each point of the field provides us with information about the magnitude of the field, the direction in which the field is moving, or both. In the case of a magnetic field, every point in space gives us information about the strength of the field as well as the direction in which it flows.

We visualize the magnetic field around a magnet by drawing **field lines**, which *always* run from the magnet's north pole to its south pole. In other words, magnetic field lines are *always* closed loops—even if in a picture there appear lines which seem headed for infinity, they will always arrive at their south pole at some point. What is more, the field lines never cross each other, and the number of lines in a certain region represents the strength of the magnetic field, whereas an arrow signals its direction.

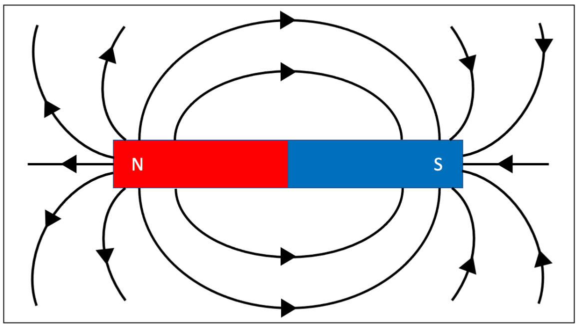


Figure 2.1: The field lines of a bar magnet, flowing from the north to the south pole

As with electric charges, the magnetic poles equally experience **attractive and repelling forces**. In magnetism, we find exactly the same rule: equal poles repel each other while opposite poles attract. More precisely, bringing two north (south) poles closer together results in the manifestation of an increasingly repelling force. In contrast, when a north and a south pole approach each other, we observe an ever attracting force.

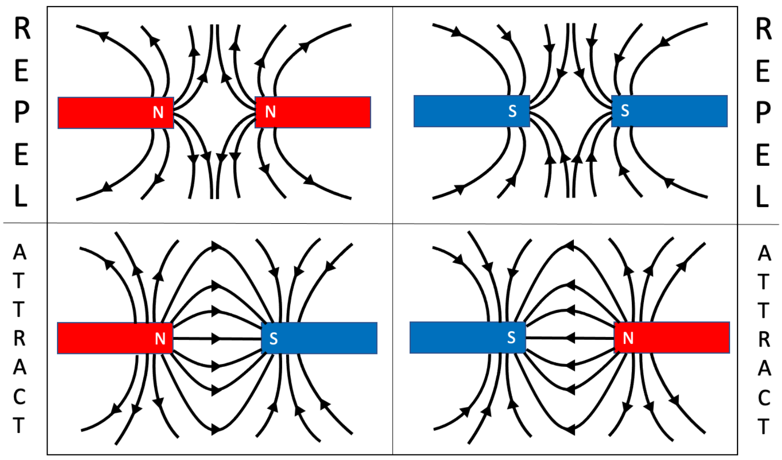


Figure 2.2: The repulsive and attractive characteristics of a bar magnet

A magnet is made by exposing a material to a magnetic field. How easy it is to (de)magnetize certain materials defines the distinction between hard and soft magnetic materials. **Hard**

magnets retain their magnetic properties once the magnetic field is removed and are called permanent magnets; it is hard to get rid of their magnetic properties once magnetized. Steel is usually the base material used to make this type of magnet. Examples include refrigerator magnets, and magnets used in disk drives and motors.

If on the other hand (de)magnetizing a material requires little effort, and it loses its magnetic properties upon deactivation of the magnetic field, the material qualifies as a **soft magnet**. These magnets are typically forged out of iron or nickel. An example of a soft magnet is the electromagnet (see section 2.2).

The act of exposing a magnetic material to a magnetic field is called **induced magnetism**. For instance, if one brings, let's say, the south pole of a permanent magnet close to one side of a magnetic material, that material will become magnetized with the exposed side turning into a north pole. Also note that a pole of a magnet cannot exist on its own. So, if we were to break a bar magnet (with the north pole on the left side) in two pieces, then both the left and right piece would develop an additional pole, i.e., a south and a north pole, respectively.

2.2 Magnetic Field Due to an Electric Current

Moving electric charges create a magnetic field B. To study the properties of this field, i.e., its direction and its magnitude, we consider three types of current-carrying wires: **a straight wire, a coil, and a solenoid**. Regardless of the shape of the wire, the magnetic field lines will take the form of circular loops positioned around the wire. The **direction** in which they circulate can be found as follows: if the thumb of your right hand points into the direction in which the current flows, then curling the other fingers of your right hand gives you the direction of the magnetic field lines.

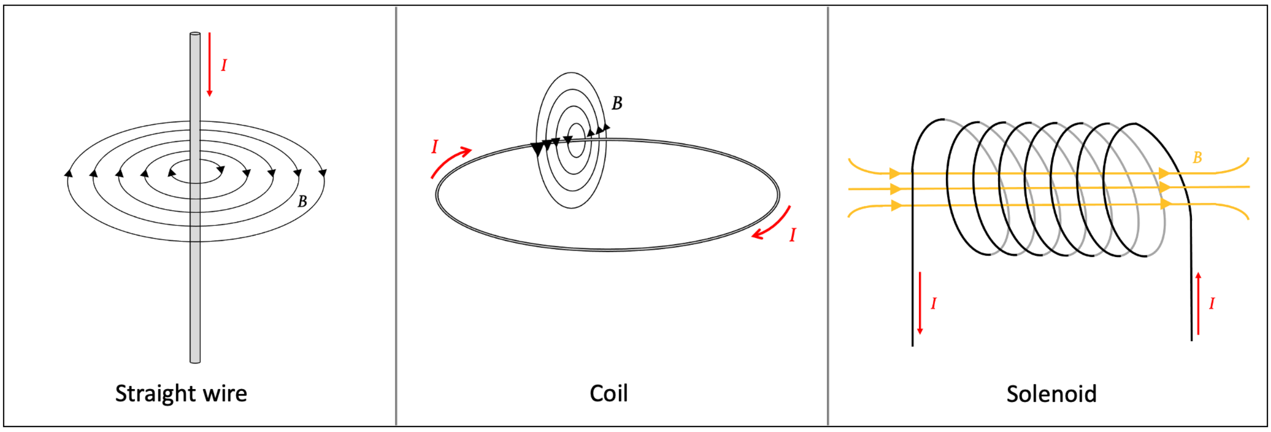


Figure 2.3: The direction of the magnetic field for a straight wire, a coil, and a solenoid

The direction of the magnetic field B in the case of a solenoid is perhaps less obvious. The magnetic lines still take the form of a circle though, if we zoom out and imagine the loops of the solenoid being packed very closely together. The lines that come out of the right-hand side of the solenoid will circle back via the outside of the solenoid to the left-hand side and flow back into the solenoid. Zooming out in this way and flipping the solenoid 90° to the right, we obtain approximately the same situation as that described for the coil (see the middle image of Fig. 2.3). Zooming back in and considering that the field lines *inside* the solenoid run parallel to its axis, we can in fact envisage the entry of the solenoid (the left-hand side) as the south pole of a magnet and the exit (the right-hand side) as the north pole, given the direction of the flow of the magnetic field lines.

With regard to **the magnitude or the strength** of the magnetic field B, we can say that for the cases of the straight wire and the coil the field strength reduces when farther away from the wire. The following equations summarize the various factors that affect the magnetic field's strength:

$$B_{Straightwire} = \mu_0 \frac{I}{2\pi r} \tag{2.1}$$

$$B_{Coil} = \mu_0 \frac{I}{2r} \tag{2.2}$$

$$B_{Solenoid} = \mu_0 \frac{NI}{L} \tag{2.3}$$

with B the strength of the magnetic field expressed in Tesla (T), $\mu_0 = 4\pi \times 10^{-7}$ T·m·A^{-1} (μ_0 is a constant and is called the permeability of free space), I the current (in A), r the distance between the wire and the point of the magnetic field B that we are measuring (in m), N the number of loops of the solenoid, and L the length of the solenoid (in m).

Equations 2.1 and 2.2 tell us that for a straight wire and a coil the two factors that impact the strength of the magnetic field B are the current I and the distance r from the wire. So, the strength of the field B not only increases when being closer to the wire, but also when cranking up the amount of current running through the wire. As a side note, B_{Coil} is being measured at the centre of the coil, so the distance r in Equation 2.2 refers not to any random distance but to the radius of the coil—coils with a larger radius will experience a weaker field at their centre.

For the solenoid, we have three factors affecting the magnetic field: the current I, the number of loops (N), and the length L of the solenoid. A longer solenoid for the same number of loops results in a weaker magnetic field, while a larger number of loops for the same length enhances its magnitude. Finally, as is also the case for a straight wire and a coil, a higher current reinforces the magnetic field.

2.3 The Motor Effect

Placing a current-carrying wire within an external magnetic field B causes the wire to experience a **force F** (see section 2.1), which results from an interaction between the magnetic field produced by the wire (see section 2.2) and the field B. Both the direction in which the current I flows and the orientation of the magnetic field B can impact the direction in which the force exerts its influence (see Fig. 2.4 for the case of a straight wire).

To identify the **force's direction**, there are two equivalent tricks we can apply: one involves our right hand, the other our left hand. In the case of the right-hand rule, when aligning our thumb with the direction of the current I and the rest of our fingers with the direction of the field B, then the force points away from the palm of our hand. Alternatively, if the middle finger of our left hand indicates the direction of the current I and the index finger marks the direction of the magnetic field B, then our thumb reveals the orientation of the force—assuming only angles of 90°, the thumb has to be perpendicular to both the index and middle finger.

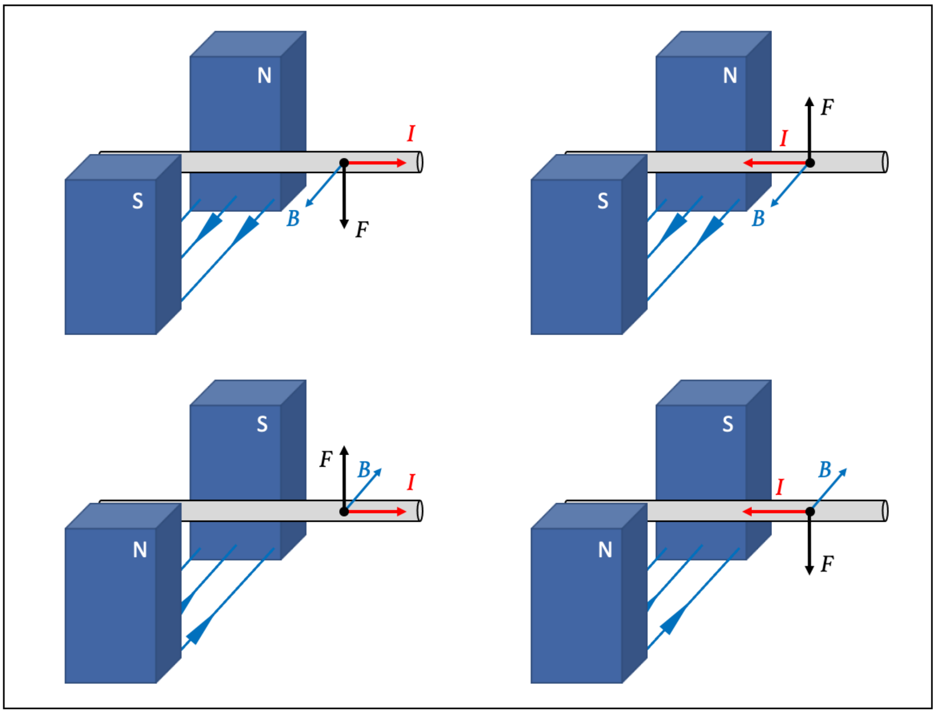


Figure 2.4: The direction of the force in function of the current I and the magnetic field B

The **magnitude of this force F** can be impacted by three factors: the magnetic field's strength B, the amount of current I, and the length L of the straight wire. If we only consider 90° angles between the wire and the applied magnetic field B, then the magnitude of the force can be calculated as follows:

$$F = BIL \tag{2.4}$$

with F the force expressed in Newton (N), B the strength of the magnetic field in Tesla (T), I the current in ampere (A), and L the length of the wire in meter (m). Some indirect effects that augment the force are a greater diameter of the wire (which elevates the current I) and more powerful magnets (which boosts the magnetic field's magnitude B).

The above-described notion of a force is the driving principle behind a **DC motor**, which is a mechanical device that converts direct current (DC) electrical energy into mechanical energy, producing a rotary movement.

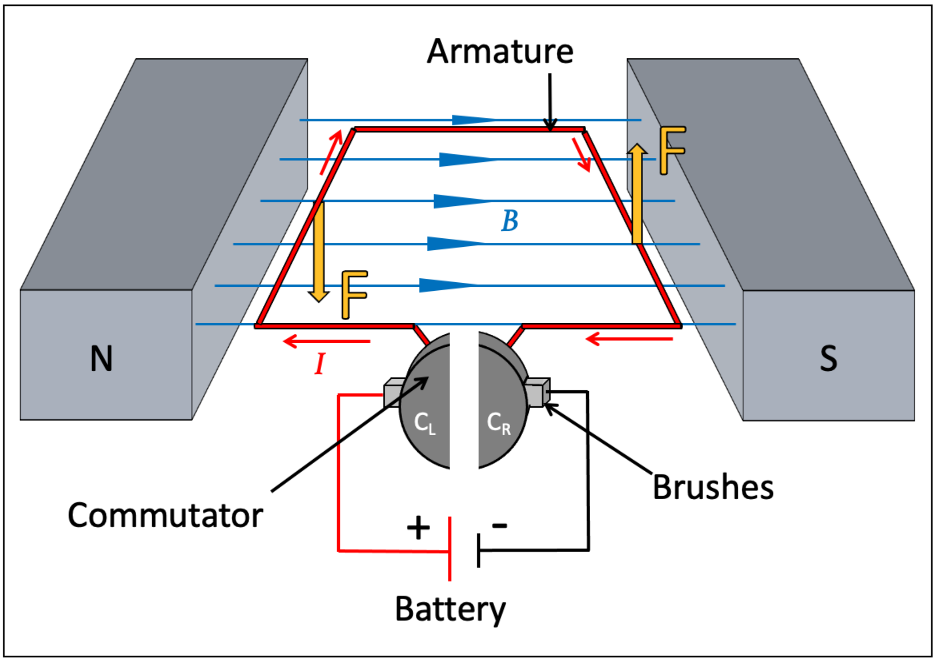


Figure 2.5: The main concept of a DC motor

In Fig. 2.5, the only parts that rotate are the armature and the commutator. The current runs from the positive side of the battery to the left brush, which is in contact with the left segment of the commutator (C_L), which in turn is connected to the armature, so that the current eventually arrives at the right segment of the commutator (C_R), which touches the right brush and thus allows the current to enter the negative side of the battery, ready to start another loop.

At this point, C_L is in contact with the left brush. This means that the left side of the current-carrying armature experiences a downward force, whereas an upward force is exerted upon the right side. The aggregated effect is that both the armature and the commutator start rotating *anti-clockwise*.

When the right (left) side of the armature reaches its highest (lowest) point, the segment C_R (C_L) is no longer touching the right (left) brush, but instead makes contact with the left (right) brush. As a consequence, C_R (C_L), which up till this point experienced an incoming (outgoing) current, now experiences an outgoing (incoming) current. That is, the current across the entire armature has effectively reversed its direction, and it will continue to do so after every half turn, i.e., a rotation of 180°.

If we want the DC motor to rotate clockwise, we can either swap the position of the bar magnets (not very practical) or reverse the polarity of the applied voltage V. The magnitude of the force can be amplified in various ways (as per Equation 2.4), including adding coils to the armature so that the magnetic field B grows stronger and increasing the current I or the voltage V.

Smaller versions of DC motors are used in many types of tools and appliances, such as robot arms, DVD disk drives in laptops, and toys, while larger DC motors are responsible for the functioning of electric vehicles and bikes, elevators, and hoists.

2.4 Electromagnetic Induction

Electromagnetic induction refers to the situation whereby a change in the magnetic field B elicits an electric current I in a closed electrical conductor, which is placed within this field, together with an accompanying force—called the electromotive force (emf)—in such a way that the emf attempts to counteract the initial change of the magnetic field. This induced electrical action or emf is expressed in volts and is defined as the amount of energy required for one unit of charge to engage in this opposing behaviour. Equation 1.2 explains why emf is expressed in volts, and emf is therefore also referred to as the **induced voltage**.

If we consider the surface which is enclosed by a closed electrical conductor (e.g., a current-carrying ring) and through which the magnetic field lines flow, a change in the magnetic field B can then be taken to mean three things: a variation in the magnetic field's strength, a modified surface area of the conductor, or a different orientation of the conductor relative to the magnetic field. Bear in mind, however, that no voltage is induced if a change is carried out *parallel to* the magnetic field lines—in other words, in the case of a constant magnetic field, an emf is only obtained if a change implies that the field lines are being cut.

Identifying the induced voltage comes down to *knowing in which direction the current I will flow* when an electrical conductor is exposed to an altered magnetic field B. The three images in Fig. 2.6 correspond to the three types of change that we pointed out in the last paragraph.

The left image represents an **increase of the magnetic strength** experienced within the enclosed area of the coil as the magnet is being lowered towards the coil. The direction of the induced voltage must be such that the increase of the field's magnitude is opposed. That is, we have to identify the direction of the induced current in a way whereby we create a magnetic field that points upwards at the centre of the coil. When we curl the fingers of our right hand accordingly, then we find that the current flows anti-clockwise.

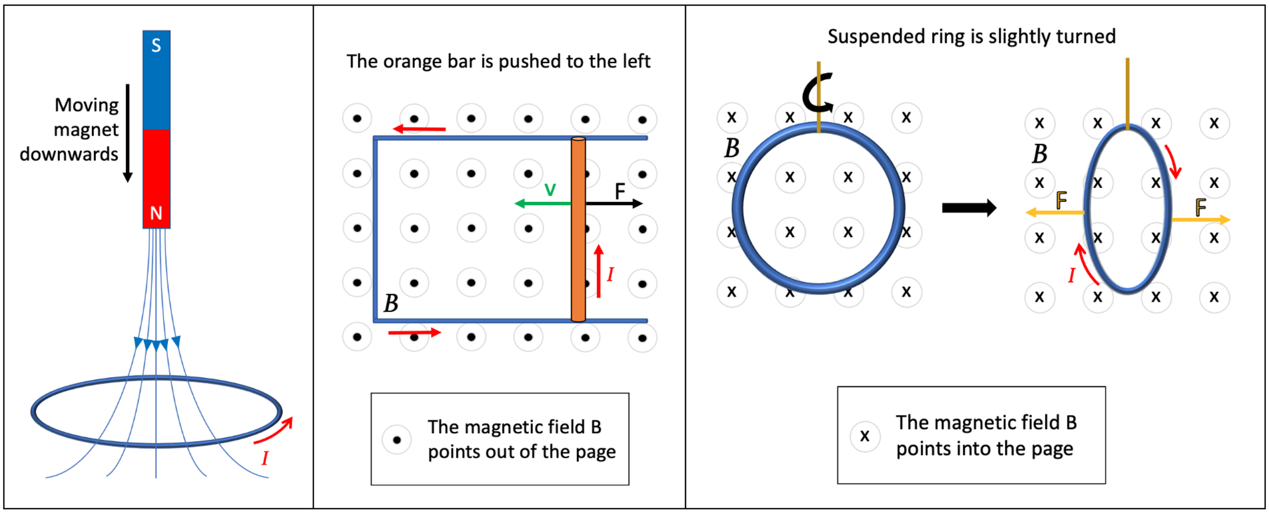


Figure 2.6: The concept of electromagnetic induction

The middle image is an example of a **modified surface area of the conductor**. The orange-coloured bar can be moved to the right or the left and forms together with the blue-coloured frame a closed electrical conductor. When we slide the bar to the left, we are basically diminishing the surface area of the conductor. Therefore, the induced voltage has to be established in a way so that the bar experiences a force that is orientated towards the opposite direction of its movement. Knowing that the magnetic field points out of the page and the force to the right, we can use the left- or right-hand rule to observe a current that must run anti-clockwise.

The right image shows how a **change in orientation of the conductor relative to the magnetic field's direction** leads to an induced voltage. When the suspended coil rotates as indicated by the black arrow, the angle that the surface area of the conductor makes with the magnetic field lines changes along with the rotation. The direction of the subsequently induced voltage must counter this action; given that a force pointing to the opposite direction of the ring's rotational movement has to be exerted upon the sides of suspended coil—and again relying on either the left- or right-hand rule—we then know that the induced current must flow clockwise.

The **magnitude of the induced voltage** is affected by the three components of change as just discussed above: the greater the change in the magnetic field's strength, in the enclosed surface area of the conductor, or in the conductor's orientation with respect to the direction of the field, the higher the induced voltage. Moreover, the magnitude can also be manipulated by the number of turns or loops out of which an electrical conductor is constructed—more loops increases the induced voltage.

The **direction of the induced voltage** can be reversed by switching the direction of motion in the three images of Fig. 2.6—that is, moving the magnet upwards instead of downwards, pushing the slide bar to the right instead of left, and rotating the suspended ring in the other direction—or by interchanging the direction of the magnetic field B.

Remark that the rotating armature of the DC motor as discussed in section 2.3 is an example of the third type of change, i.e., a change in orientation of the conductor in relation to the magnetic field (see the right image of Fig. 2.6). As a matter of fact, the induced voltage—in the context of the DC motor this induced voltage is referred to as "back emf"—tries to oppose the direction of the rotary movement, which is sustained by the applied battery voltage V. This means that the voltage V must constantly coerce the current I to run through the armature as it struggles against the back emf (which is always smaller than the voltage V). As a result, it is this electric work invested in maintaining an "upstream" current that is being transformed into mechanical energy within the armature, giving rise to the rotary movement.

The principle of electromagnetic induction lies at the basis of the workings of an **alternating current (AC) generator**. While a DC motor uses electric energy to produce mechanical energy, the AC generator applies mechanical energy to convert it into electricity.

In Fig. 2.7, the only parts that rotate are the connection to the turbine, the armature, and the slip rings. The brushes are in contact with the slip rings, yet remain stationary. We refer to the first (second) slip ring as SR_1 (SR_2), and it is accompanied by a fixed brush, which we will indicate as Br_1 (Br_2). The enclosed surface of the armature is denoted as the rectangular 1234. As a starting position of our *clockwise* rotation, imagine that the side 12 is located *at the top* of the circular motion carried out by the armature.

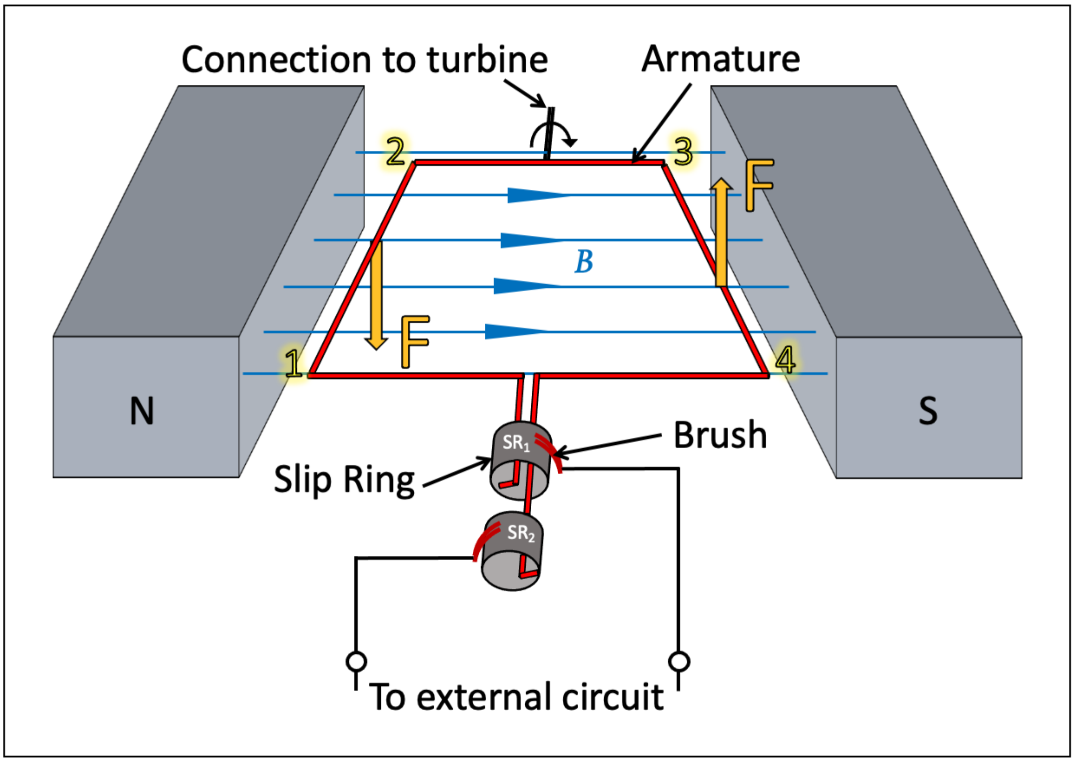


Figure 2.7: The main concept of an AC generator

When we start rotating, the orientation of the closed surface with respect to the magnetic field changes constantly, so that the armature develops an induced voltage. During the first 90° of rotation, the side 12 experiences an upwards force so that the induced current runs from point 2 to 1. The same upwards force is present during the next 90° of rotation. The side 12 is now located at the lowest point of the circular movement of the armature.

However, during the next half turn of the armature the side 12 registers a downwards force, whereby the induced current flows from point 1 to 2, which is the opposite direction relative to the first 180° of rotation. Stated differently, as it rotates along with the armature, SR_1 receives an incoming flow of current during the first half turn and experiences an outgoing flow during the second half turn.

Since the side 34 sits at the opposite end of the armature with regard to the side 12, the side 34 follows the exact opposite evolution: at the starting position, it is located at the lowest point of the rotary motion, and during the first (second) half turn it encounters a downward (upward) force. Accordingly, SR_2 is first submitted to an outgoing current which is followed by an incoming current during the second half turn.

Putting it all together, the continuous switch of the direction of the current every 180° of rotation causes the external circuit to undergo a flow of current that alternates with every half turn. Hence the name AC generator.

The **orientation of the output voltage** can be affected by reversing the direction of rotation or swapping the magnets. The **magnitude of the induced voltage** can be enhanced by increasing the number of coils in the armature, by strengthening the magnetic field B, by enlarging the closed surface area of the conductor (the rectangular 1234), or by speeding up the rate of rotation.

The angle between the direction of the surface area and that of the magnetic field B impacts both the magnitude and the orientation of the induced voltage. What is more, if we plot how the induced voltage evolves with the *angle of rotation*, then we observe a **sine wave**, which is typical for AC as already alluded to in section 1.2 (see Fig. 1.3)—the below wave is technically a cosine function, which can be converted into as sine function through the relation $\cos(\alpha) = \sin(\alpha + \frac{\pi}{2})$.

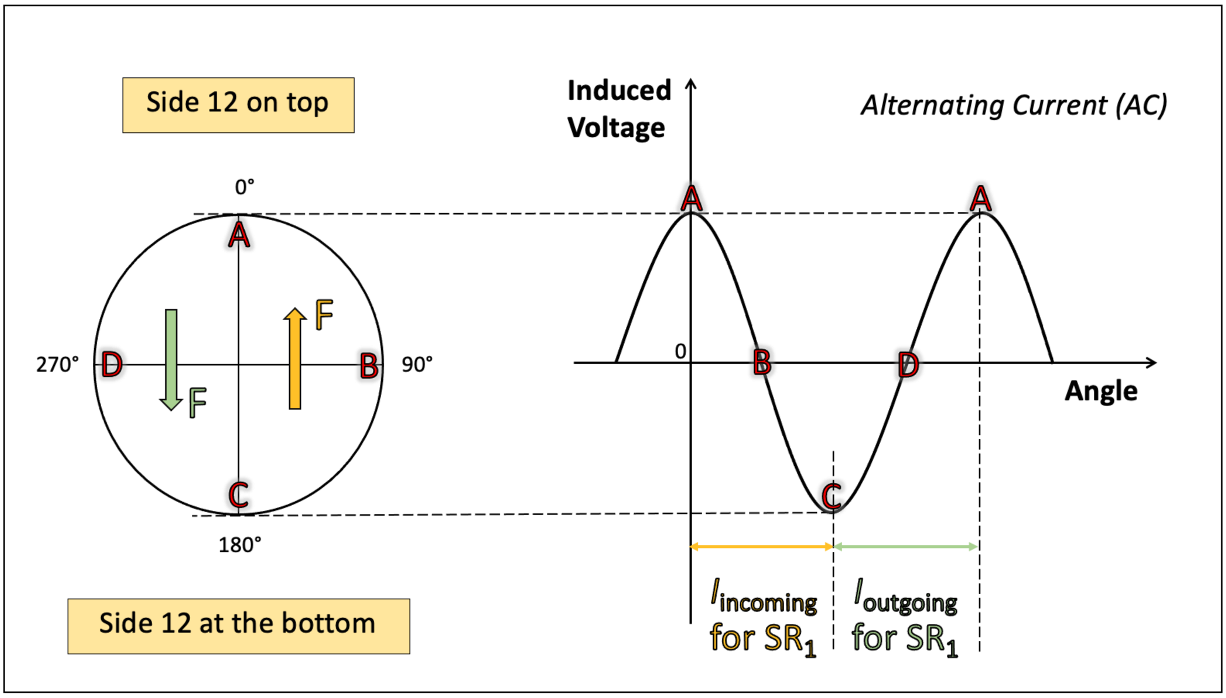


Figure 2.8: The output voltage of an AC generator expressed in terms of the angle of rotation

Point A in Fig. 2.8 corresponds to our agreed upon starting position of the armature, and this is the moment when the induced voltage reaches a maximum value. At point B the plane of the closed surface area is parallel to the magnetic field lines, therefore no induced voltage is present. When we arrive at point C, we again obtain a maximum output voltage, but this time with an opposite polarity; from this moment onwards the current runs in the other direction. Another 90° rotation brings us to point D, whereby the armature is once more orientated parallel to the magnetic field, thus not producing any current. Finally, completing one full turn takes us back to our starting position.

One example where an AC generator is put to use is a hydroelectric dam whereby the waterwheels of the turbines inside the power station are connected to the rotating axis of an AC generator, transforming in this way the mechanical energy delivered by falling water into electricity. In thermal or nuclear power stations, the turbines are driven by steam instead.

Besides the AC generator, other examples that rely on the principle of electromagnetic induction include microphones, induction stoves, safety circuit breakers, and transformers, which we will discuss next.

2.5 Transformers

A **transformer** is a stationary framework that passes on electrical energy (or, if you will, voltage or electrical power) from one circuit to another based on the principle of electromagnetic induction—there is no physical transfer of electricity (no contact) between the circuits.

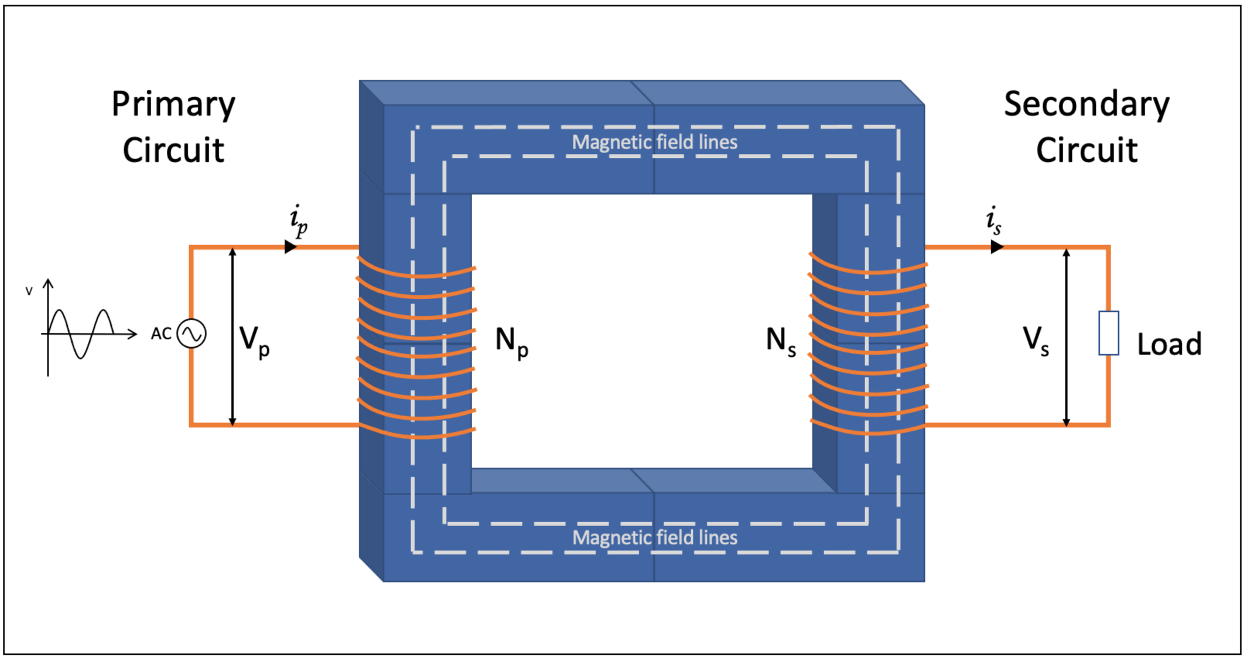


Figure 2.9: The basic setup of a transformer

The primary circuit is the circuit that is connected to an AC source, while the secondary circuit is the one that contains the transformer load, which is an electrical device coupled to the output of the transformer. The *main idea* of a transformer is that the alternating current of the primary circuit generates a constantly changing magnetic field that is transported via the static frame (called the magnetic core) to the secondary circuit in which it develops an induced current as a result of electromagnetic induction.

As the fluctuating magnetic field also gives rise to an induced voltage in the primary circuit, it is said that both circuits are in communication with each other through mutual induction. In fact, if the number of windings (turns) of both circuits are equal ($N_p = N_s$), then the same amount of voltage is transmitted to the load. However, if the windings of the secondary circuit are greater in number compared to the primary circuit ($N_p < N_s$), then the output voltage increases, and the transformer is referred to as a **step-up transformer**. In the opposite scenario ($N_p > N_s$), the transformer is known as a **step-down transformer**. This is summarized by the following relationship between the voltage and the turn ratio:

$$\frac{V_p}{V_s} = \frac{N_p}{N_s} \tag{2.5}$$

In case of an ideal transformer, which carries over the entire amount of electrical energy, we can assume that the electrical power of the primary circuit and the load are equal to one another. Based on the definition of electrical power (see Equation 1.4) and using Equation 2.5, we can then write the following equivalent relations:

$$P_p = P_s \iff V_p I_p = V_s I_s \iff \frac{N_p}{N_s} = \frac{I_s}{I_p} = \frac{V_p}{V_s} \tag{2.6}$$

Transformers play a key role in **power transmission**. Roughly, the process goes as follows. A power station generates electricity, whose output voltage of up to 30,000 V is boosted to values in the range between 110,000 V and 765,000 V by means of step-up transformers, which are located at an electrical station nearby the power plant. The electric charges are subsequently transported via high-voltage transmission lines to eventually have their voltage reduced via step-down transformers at local substations to a value between approximately 3,000 V and 66,000 V. Lower-voltage transmission lines then transfer the electricity close to our homes and buildings, where its voltage is further decreased to 240 V by step-down transformers attached to electricity poles. The alternating current is now ready to power our household appliances—depending on where you live in the world, the voltage can still be further brought down to 110 V, 115 V, 120 V, 127 V, 220 V, or 230 V.

The transmission of power cannot occur without some **losses of electrical energy**, because the flowing of electrical charges causes the production of heat due to friction (resistance). In order to minimize the incurred losses, Equation 2.6 tells us that it is beneficial to transport electricity over high-voltage transmission lines, since a higher voltage V_s implies a lower electrical current I_s and thus a minimum of resistance and resistive losses. For instance, if the voltage V_s is increased by a factor 20, then the amount of current I_s decreases by that same factor. Therefore, the power losses are minimized by a factor of 400 (remember that according to Equation 1.4, $P = I^2R$).

3 Mechanics

3.1 Kinematics

The term kinematics is used to denote the subfield within classical mechanics that studies the motion of objects without taking into consideration its underlying causes, i.e., the forces acting on them—forces are discussed in section 3.2.

In section 2.1, we briefly touched upon the fact that fields contain information which may tell us something about the field's magnitude, direction, or both, at every location in space. If the only information available about a certain physical quantity is its magnitude, we refer to that field or quantity as a **scalar field or scalar quantity**. On the other hand, **a vector field or vector quantity** is defined as a field or physical quantity that reveals information about the quantity's magnitude as well as its direction in space. Note that vector quantities carry an arrow on top of their respective symbols, while scalar quantities do not.

To illustrate the difference with an example, temperature is a scalar quantity (or scalar field), as it makes no sense to talk about the direction of temperature, whereas velocity is a vector quantity (or vector field), since at every point in space we can figure out both an object's speed and the direction in which it is headed. Other scalar quantities include time, energy, distance, and speed, and other examples of vector quantities are position, displacement, acceleration, force, momentum, and weight.

The notion of **displacement $\Delta\vec{\mathbf{x}}$ or $\vec{\mathbf{s}}$** (a vector quantity) refers to the difference between an object's initial ($\vec{x}_0$) and final position ($\vec{x}_f$) in space—in other words, displacement represents a change in position—whereas the **distance x** (a scalar quantity) indicates the total length of the path trodden to arrive at that final position $\vec{x}_f$. For instance, after completing one lap at a running track of 800 m (this is the distance x) we are back at the point of departure, so that the total displacement Δx (or s) measures 0 m—Δx is the accompanying scalar (magnitude) of the vector $\Delta\vec{x}$ and carries no arrow.

The same distinction can be made for speed v versus velocity $\vec{v}$. The concept of **velocity $\vec{\mathbf{v}}$** (a vector quantity) implies a direction in space, e.g., we are driving at 75 km·h^{-1} to the northeast, while the **speed v** (a scalar quantity) is the magnitude of the velocity vector, e.g., we are driving at 75 km·h^{-1}. So, a change in velocity ($\Delta\vec{v}$) can either mean a change in speed or a change in the direction of motion. For example, an object moving in a circle at a constant speed is continuously changing its velocity, as its direction is changing from moment to moment.

Another vector quantity is **acceleration ($\vec{\mathbf{a}}$)**, by which is meant the change of velocity over time. To come back to the example of the circular movement, due to the continuous change in direction, the object is continually undergoing a constant, non-zero acceleration. The definitions of speed, velocity, and acceleration can be summarized by the following equations:

- $$\text{speed} = \frac{\text{distance}}{\text{time}} \iff v = \frac{x}{t} \tag{3.1}$$
- $$\text{velocity} = \frac{\text{change in displacement}}{\text{time}} \iff \vec{v} = \frac{\Delta\vec{x}}{\Delta t} = \frac{(\vec{x}_f - \vec{x}_0)}{\Delta t} = \frac{\vec{s}}{\Delta t} \tag{3.2}$$

- acceleration $= \dfrac{\text{change in velocity}}{\text{time}} \iff \vec{a} = \dfrac{\Delta\vec{v}}{\Delta t} = \dfrac{(\vec{v}_f - \vec{v}_0)}{\Delta t}$ (3.3)

whereby t or Δt is the time (in seconds (s)), x is the distance (in meters (m)), v is the speed (in m·s^{-1}), $\Delta\vec{x} = \vec{s}$ is the displacement (in m), $\vec{x}_0$ ($\vec{x}_f$) is the initial (final) position (in m), $\vec{v}$ is the velocity (in m·s^{-1}), $\vec{v}_0$ ($\vec{v}_f$) is the initial (final) velocity (in m·s^{-1}), and $\vec{a}$ is the acceleration (in m·s^{-2}). If we wish to calculate the **average value** for any of these quantities for the total amount of elapsed time Δt, then we replace distance and displacement by the total distance covered and the total displacement, respectively.

In classical mechanics, the **three equations of motion** describe how physical objects move through space in function of time. From these equations, we can determine an object's position (or displacement) as well as its velocity. For our convenience, we will consider only one spatial dimension (so, we can omit the vector notation) and also write t instead of Δt. Equation 3.3 provides the first equation of motion:

The first equation of motion: $v_f = v_0 + at$ (3.4)

From Equation 3.2, we see that $v_{average} = \frac{s}{t}$ (with s being the total displacement). We also know that we can obtain the average velocity by adding the initial and final velocity and dividing that sum by two ($v_{average} = \frac{v_0+v_f}{2}$). Matching these two definitions of the average velocity and replacing v_f with the description given by Equation 3.4, we obtain the second equation of motion:

The second equation of motion: $s = v_0 t + \dfrac{at^2}{2} \iff x_f = x_0 + v_0 t + \dfrac{at^2}{2}$ (3.5)

Finally, if we insert $t = \frac{v_f - v_0}{a}$ (which we obtain from Equation 3.4) into Equation 3.5, the third equation of motion appears:

The third equation of motion: $v_f^2 - v_0^2 = 2as \iff v_f^2 - v_0^2 = 2a(x_f - x_0)$ (3.6)

To further highlight the difference between distance and displacement as well as between speed and velocity, let us plot these quantities with an example, whereby we can also put into practice the equations of motion.

Imagine that we start running on a straight line with an initial velocity of $v_0 = 2$ m·s^{-1} for a distance of $x = 200$ m. After that distance, we pick up the pace and accelerate ($a = 0.1$ m·s^{-2}) for a distance of $x = 100$ m. Next, we gradually slow down until we come to a full stop. The magnitude of the rate at which we wind down our velocity (this rate is called the deceleration) is the same as that of our acceleration, but it points in the other direction ($a = -0.1$ m·s^{-2}). In a last step, we casually run back at a speed of $v = 2$ m·s^{-1} to the location where we previously initiated our deceleration. Schematically, our physical exercise looks like this:

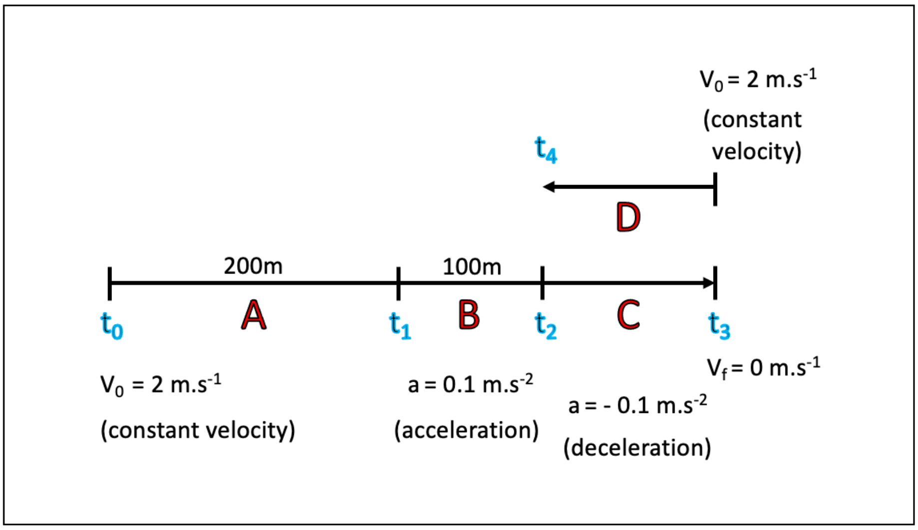


Figure 3.1: The schematic representation of the example in one spatial dimension

Before we can draw some plots, we need to calculate the following missing values: t_1, t_2, t_3, t_4, the final velocity v_f of phase A and B, and the displacement of phase C and D. Equation 3.1 reveals that the time it takes to run the first 200 m equals $t_1 = 100$s. With the help of Equation 3.6, we learn that after accelerating for 100m during phase B our final velocity measures $v_f = 4.9$ m·s^{-1}—which is also the initial velocity v_0 of phase C—and Equation 3.4 tells us that we pull off this sprint in $\Delta t_B = (t_2 - t_1) = 29$s. This means that so far we have been running for $t_2 = 129$s.

In order to come to a halt at the end of phase C, we have to bridge a displacement of $\Delta x = s = 120$ m, according to Equation 3.6, and it would require $\Delta t_C = (t_3 - t_2) = 49$s, as per Equation 3.4. At this point, our total elapsed time amounts to $t_3 = 178$s, and we now start to run back to the position where we initiated the deceleration phase, i.e., the beginning of phase C. As we already calculated that $\Delta x = s = 120$ m for phase C, it then follows that we obtain the same amount of displacement for phase D. Finally, Equation 3.1 indicates that our last stretch lasts for $\Delta t_D = (t_4 - t_3) = 60$s, so that the total elapsed time for our entire run equals $t_4 = 238$s, during which we have run a total distance of 540 m.

Using these calculated values, we can now create the following four graphs: a distance-time graph, a displacement-time graph, a speed-time graph, and a velocity-time graph.

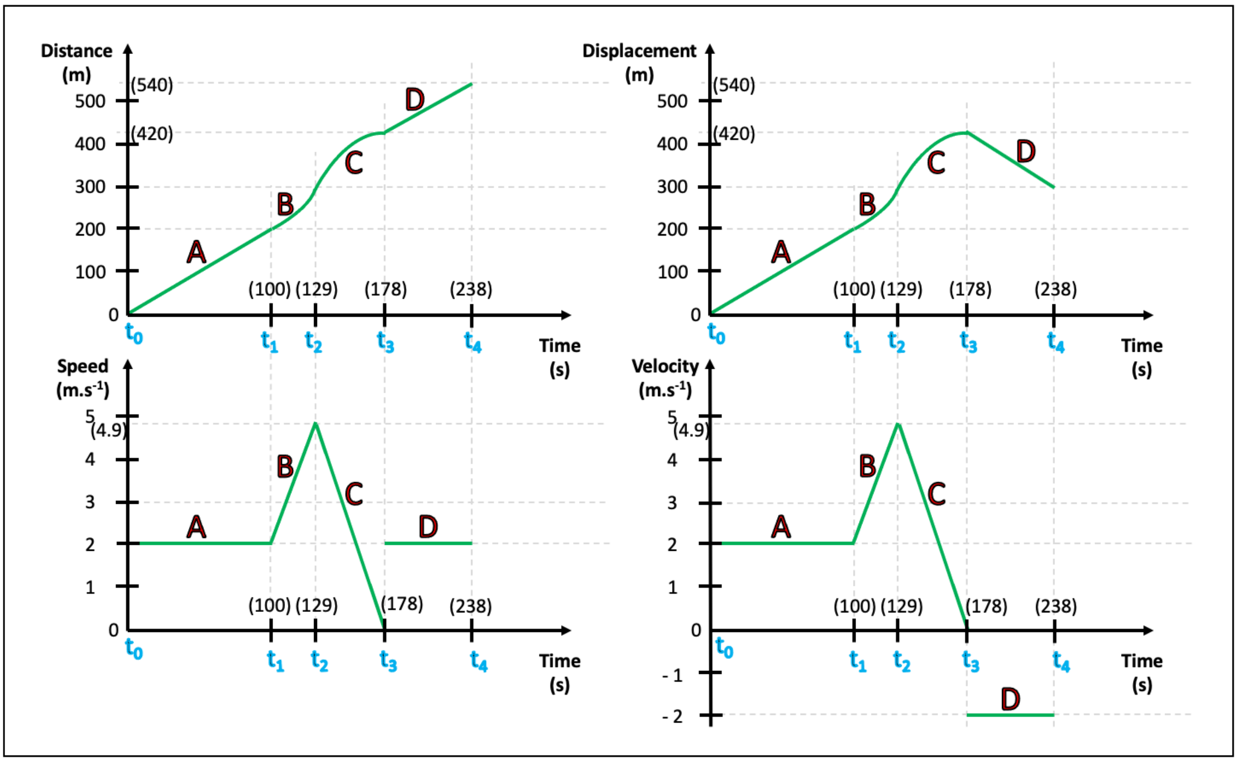


Figure 3.2: Our example: distance versus displacement and speed versus velocity

There are two relationships between the distance-time (displacement-time) and the speed-time (velocity-time) graphs: **the slope (gradient) of a curve and the area under a graph**. Concerning the first relation, Fig. 3.2 shows that the slope of a *straight line* in the distance-time (displacement-time) plot equals the speed (velocity) of the corresponding phase in the speed-time (velocity-time) plot. In our example, this applies to both phase A and D. Remember that the slope for a straight line between two points (x_1, y_1) and (x_2, y_2) is equal to $\frac{y_2-y_1}{x_2-x_1}$. For instance, looking at phase A in the displacement-time plot, we can identify two points on the straight line, i.e., (0, 0) and (100, 200), so that we find a slope of $\frac{200-0}{100-0} = 2$. If we rewrite this fraction, expressing it in the appropriate physical quantities in line with the displacement-time plot, we see that the slope equals $\frac{\Delta x}{\Delta t}$, which is nothing but the definition of velocity (see Equation 3.2). Indeed, the slope of phase A that we just calculated matches the velocity of 2 m·s^{-1} of phase A in the corresponding velocity-time graph.

In the case of phase B and C, however, we are dealing with *curved lines*. If we apply the above formula for a slope, we basically approximate these curves by straight lines, and the values that we obtain are *average* values. For example, the slope of phase C in the distance-time plot is calculated as $\frac{420-300}{178-129} = 2.45$. This corresponds to the average speed that we infer as follows from the speed-time graph: $\frac{v_0+v_f}{2} = \frac{4.9+0}{2} = 2.45$ m·s^{-1}.

Also the slope of the curves in both the speed-time and the velocity-time graphs carries a meaning: it coincides with the value of acceleration (or deceleration). For instance, the slope of B in the velocity-time plot amounts to $\frac{4.9-2}{129-100} = 0.1$, which agrees with the acceleration of $a = 0.1$ m·s^{-2}.

The second relationship implies that the area under a graph in the speed-time (velocity-time) plot equals the amount of distance (displacement) covered. For instance, in the case of phase D in the velocity-time graph, the area between the graph and the time axis is equal to $(2 \times (238 - 178)) = 120$, which is precisely the amount of displacement that we can read off from the displacement-time graph for phase D, i.e., 420 m − 300 m = 120 m.

As a final illustration to demonstrate the difference between distance and displacement and between speed and velocity, let us calculate the total average speed and the total average velocity for the run we performed in our example. The total average speed equals $\frac{\text{total distance}}{\text{total elapsed time}} = \frac{540}{238} = 2.3\ \text{m·s}^{-1}$, while the total average velocity is measured as $\frac{\text{total displacement}}{\text{total elapsed time}} = \frac{300}{238} = 1.3\ \text{m·s}^{-1}$.

3.2 Forces

The behaviour of an object can vary over time: it can go faster, it can slow down, it can alter its direction of movement, it can resist a motion, it can be deformed, etc. The question that section 3.1 does not address is what causes motion in the first place. In classical mechanics, it is the presence of a **force** $\vec{\boldsymbol{F}}$ (expressed in SI units of newton (N)) which is ultimately responsible for these changes.

Forces can be divided up in two broad categories: contact forces and non-contact forces. Forces that come into play when objects are in direct contact with one another are appropriately known as **contact forces**, whereas forces that can wield their influence without the need for physical contact are called **non-contact forces or field forces**—that is, the mere existence of other objects gives rise to an interaction between them, and force is the concept that quantitatively describes their connection.

With respect to contact forces, we can distinguish various types, such as **tensional forces** (e.g., pulling a rope at both ends), the force related to **springs** (see section 3.3), the **normal force** (i.e., the reactive force when an object experiences a force), **friction** (i.e., the force that arises when an object moves across another object, e.g., sliding a book across the table), **drag forces** (i.e., forces that counter the direction of motion of an object in a fluid, such as air resistance; in physics, fluids comprise liquids and gases), **weight** (i.e., the force that acts on objects with a *mass* as a result of gravity (which in itself is a non-contact force); this force always points to the centre of the Earth), **lift** (i.e., the force that a fluid exerts upon an object, such as buoyancy (also called upthrust), and it usually points to the opposite direction of gravity; for instance, releasing a submerged ice cube in a glass of water results in the ice cube rising to the surface), and **thrust** (i.e., the force that points to the opposite direction in which a system is expelling (accelerating) mass, such as the forward thrust generated when a rocket engine ejects burnt fuel).

In terms of non-contact forces, there is the **gravitational force or gravity** (i.e., the force that objects exert upon each other because of the fact that they have mass), the **electrostatic force** (i.e., the force that describes the interaction between electrical charges; see section 1.1), the **magnetic force** (i.e., the force that exists between magnets as well as the force that arises as a result of electrical charges moving through a magnetic field; see section 2), the **strong nuclear force** (i.e., the force that holds the nucleus of an atom together), and the **weak nuclear force** (i.e., the force that, for instance, transforms a neutron into a proton within the nucleus of an atom; neutrons and protons are the composite particles that make up the nucleus, and they consist of smaller particles called quarks).

Since force $\vec{F}$ is a vector, it has both a magnitude and a direction, and a **force diagram** helps to visualize how forces act on a system. Before considering a couple of examples, bear in mind that, in classical mechanics, the behaviour of macroscopic objects can be adequately described without taking into account the nuclear forces, which act on very small distance scales. Therefore, in the below examples, which discuss macroscopic objects, the nuclear forces do not appear.

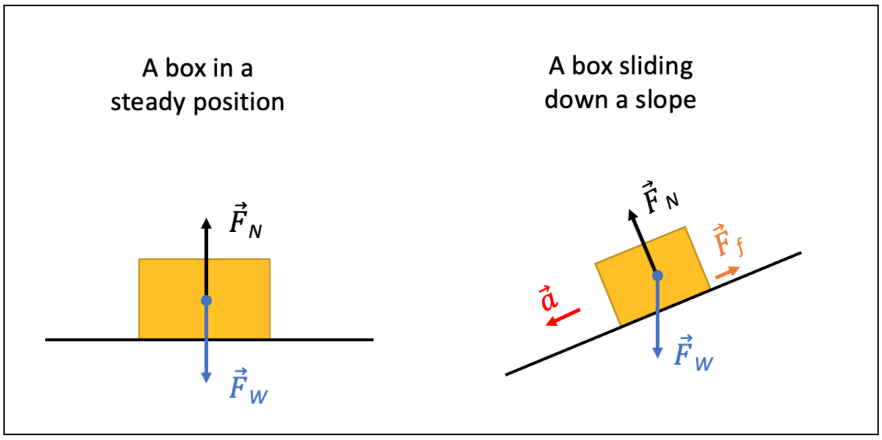


Figure 3.3: Force Diagrams: A box in balance and a box moving down a slope

In the image to the left in Fig. 3.3, a box is sitting on, let's say, a table top and is *not moving*. In this particular physical system, the only two forces present are the weight force $\vec{F}_W$ and the normal force $\vec{F}_N$. The weight of the box points towards the centre of the Earth and acts on the table top, which reacts to the weight force by manifesting a normal force, which acts on the box. Since the box is in perfect balance, both forces are equal in magnitude. The heavier the mass of the box, the greater the magnitude of both forces.

If we tilt the table top enough so that the box *starts sliding*, we arrive at the right image in Fig. 3.3. The weight force never changes its direction, yet the normal force is always directed perpendicular to the surface with which the object makes contact. The direction of motion is to the left, so that the frictional force $\vec{F}_f$ points to the opposite direction. Given that the box is moving, it means that the magnitude of $\vec{F}_f$ is not large enough to prevent the sliding from happening. In fact, larger friction could be established by increasing the mass of the object, decreasing the tilted angle of the table top, or replacing the surface with a material that possesses a higher coefficient of friction.

The left image of Fig. 3.4 represents an ice cube *floating* in a glass of water. The buoyant force $\vec{F}_L$, which is an example of a lifting force, acts on the part of the ice cube that is still immersed in the water. As the ice cube is balanced, $\vec{F}_L$ perfectly counteracts the weight force $\vec{F}_W$. That is, they are equal in magnitude, but opposite in direction. If the ice cube would be placed in the middle of a full glass of water, $\vec{F}_L$ would exhibit a larger magnitude than $\vec{F}_W$, and the ice cube would rise to the surface. The buoyant force grows stronger if the density of the fluid is larger than that of the object (as is the case for the ice cube; density is defined as $\frac{mass}{volume}$, see section 5.4) and if the volume of the object becomes bigger.

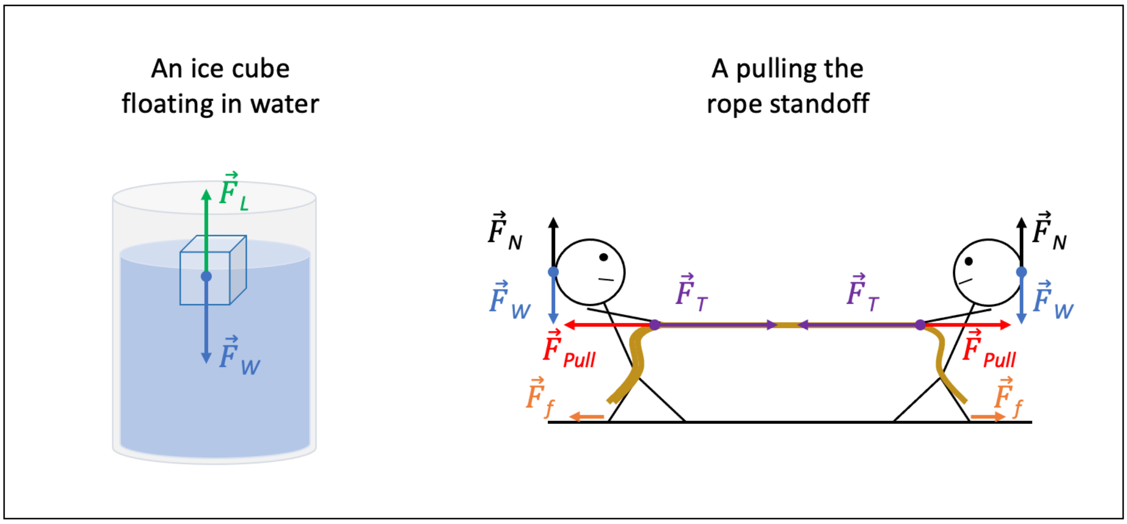


Figure 3.4: Force Diagrams: An ice cube in water and pulling the rope

To the right in Fig. 3.4 we see two individuals engaged in a pulling the rope contest. As it concerns a *standoff*, none of the contestants are moving, as they show overall equal strength. Apart from the usual weight and normal force, each individual experiences a frictional force $\vec{F}_f$ at their feet, preventing them from sliding towards the other contestant. Moreover, the force $\vec{F}_{Pull}$ with which they pull the rope triggers a tensional force $\vec{F}_T$ within the rope, acting in the opposite direction of the pull. However, the forces $\vec{F}_T$ and $\vec{F}_{Pull}$ are not equal in magnitude. Nonetheless, since the entire situation is in perfect equilibrium, all forces must cancel each other out. In other words, $\vec{F}_T = \vec{F}_{Pull} + \vec{F}_f$.

The car in Fig. 3.5 is *accelerating to the right*, which means that the force applied to drive the car ($\vec{F}_{Drive}$) exceeds the sum of the drag force $\vec{F}_D$ and the friction $\vec{F}_f$ (this is represented in the image by choosing appropriate lengths for the respective vectors). If we would consider the scenario whereby $\vec{F}_{Drive} = \vec{F}_D + 4\vec{F}_f$, then the car would be cruising at a constant speed. It is useful to note that the drag force can be reduced by optimizing the aerodynamic design of the car.

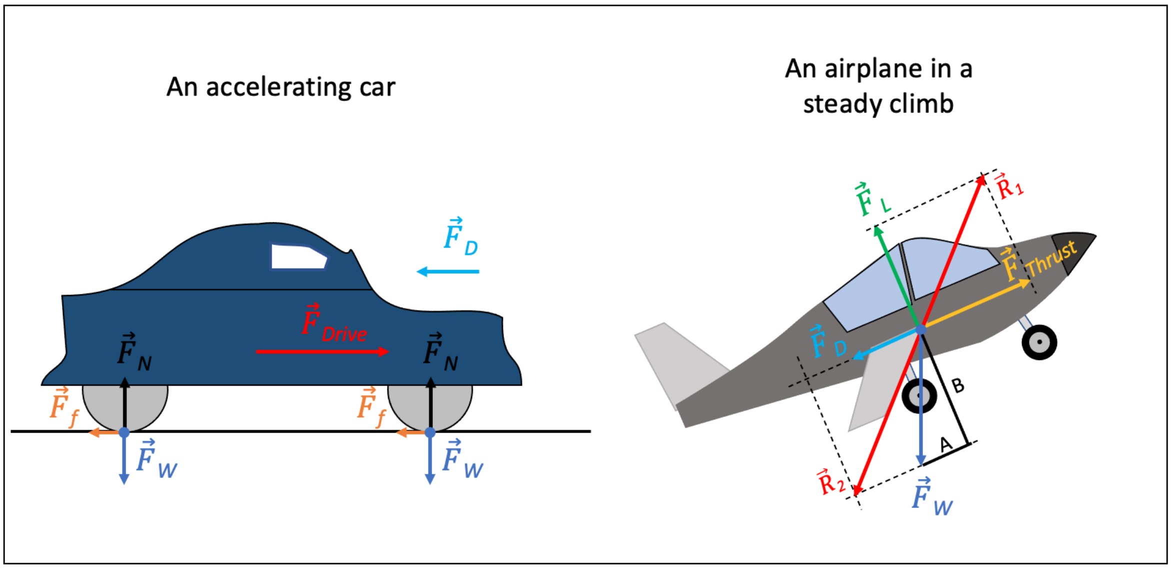


Figure 3.5: Force Diagrams: An accelerating car and a climbing airplane

The airplane in Fig. 3.5 is propelled forwards by means of fuel-based engines, which provide the airplane with a thrust $\vec{F}_{Thrust}$. Generally, an airplane experiences four forces: thrust $\vec{F}_{Thrust}$, drag $\vec{F}_D$, lift $\vec{F}_L$, and weight $\vec{F}_W$. In this example, the aircraft is engaged in a *steady climb*, meaning that the four forces are in balance; in other words, *the airplane is not accelerating*. This is best understood by looking at the two red-coloured vectors $\vec{R}_1$ and $\vec{R}_2$, which are equal in magnitude and opposite in direction (they are the result of vector addition, which will be shortly discussed below).

So, even though the length of the vector $\vec{F}_{Thrust}$ is longer than that of the drag force $\vec{F}_D$, there is no acceleration. The reason why is because, due to the inclination of the airplane, the thrust is also carrying part of the weight vector $\vec{F}_W$, i.e., the part that is parallel to the drag force and indicated in Fig. 3.5 by the black line A, so that $\vec{F}_{Thrust} = \vec{F}_D + \text{A}$. What is more, as a result of the ascending motion, the lift vector $\vec{F}_L$ is growing smaller than $\vec{F}_W$ and equals the part of $\vec{F}_W$ that is parallel to $\vec{F}_L$, i.e., the black line B.

To come back to the concept of **vector addition**, consider Fig. 3.6, which provides three examples in one and two spatial dimensions (only numerical calculations are given for the one-dimensional case). The resultant force $\vec{R}$ signals the direction in which the system *as a whole* will move, whereby the direction of this final vector is determined by the orientation of the individual forces. Regardless of the number of dimensions, vector $\vec{R}$ can be found as follows: by placing the starting point of the second vector (the green dot of vector $\vec{F}_2$) at the end of the first vector (the black arrow of $\vec{F}_1$), the vector $\vec{R}$ is found by connecting the starting point of vector $\vec{F}_1$ (the black dot) to the end of vector $\vec{F}_2$ (the green arrow).

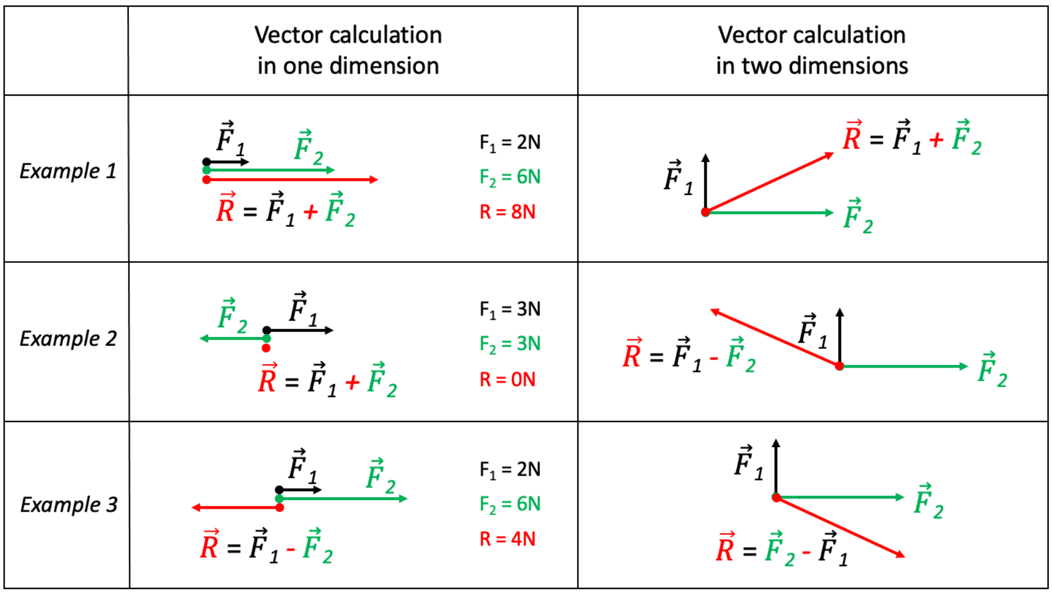


Figure 3.6: The concept of vector addition in one and two dimensions

In example 1 in one dimension, as both individual forces point towards the same direction, the vector $\vec{R}$ has a length that is the sum of the lengths of the individual forces. In this case, the entire system is accelerating to the right. In example 2 in one dimension, as the vector $\vec{F}_2$ brings us back to the starting point of vector $\vec{F}_1$, i.e., the black dot, there is no resultant force $\vec{R}$ ($R = 0$N). This means that this physical system is in equilibrium, i.e., it is not accelerating. It also means that the system is either not moving or is moving in some direction at a *constant* velocity. In example 3 in one dimension, the vector $\vec{R}$ requires the subtraction of $\vec{F}_2$, which is the same as adding the vector with the opposite direction, i.e., $\vec{R} = \vec{F}_1 - \vec{F}_2 = \vec{F}_1 + (-\vec{F}_2)$. In other words, when placing the green dot of $\vec{F}_2$ at the position of the black arrow of $\vec{F}_1$, the vector $\vec{F}_2$ points *to the left*. Connecting the black dot of $\vec{F}_1$ with the green arrow of $\vec{F}_2$ gives the resultant vector $\vec{R}$, which points to the left, as the magnitude of the force $\vec{F}_2$ is larger than that of the force $\vec{F}_1$. Note that it is the absolute value of the magnitude of $\vec{R}$ that is expressed, i.e., $R = 4N$ and not $R = -4N$.

3.3 Force and Extension

An object that is submitted to a force can become compressed or extended. The amount by which it is compressed or extended is called the **extension**. When the applied force is removed and the object returns to its original shape, the extension is designated as **elastic**. Conversely, if the object is deformed and cannot recover its original shape, it is said to exhibit **inelastic** behaviour. The point at which deformation is irreversible for a particular type of material is referred to as the **elastic limit**. Fig. 3.7 shows the relationship between the force and the extension, whereby point B marks the elastic limit.

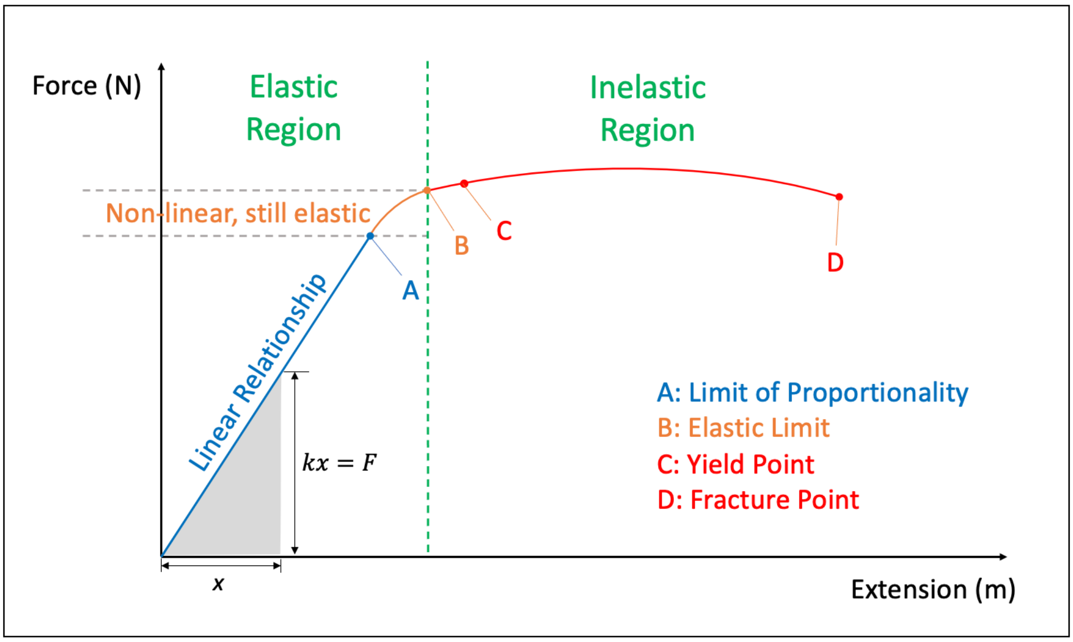


Figure 3.7: The force-extension graph

To the right of point B, we enter the inelastic region (also known as the plastic region), where from point C (the yield point) onwards relatively little additional force is needed to further deform the object. Eventually, the material starts falling apart at point D, i.e., the breaking point or fracture point.

Going back to the elastic region, there is another point of interest (point A), called **the limit of proportionality**. Up to this limit (the blue-coloured straight line), the extension of the object occurs in a linear fashion. That is, the amount of applied force is directly proportional to the amount of extension. This linear relationship is described by **Hooke's**

law:

$$\vec{F} = k\vec{x} \tag{3.7}$$

with $\vec{F}$ the force expressed in newton (N), k the proportionality constant that reflects how rigid or stiff a material is (k is expressed in N·m^{-1}), and $\vec{x}$ the extension expressed in meter (m). The constant k is the slope of the blue-coloured straight line in Fig. 3.7; the steeper the slope, the higher the constant k, the stronger and also the more brittle the material, such as steel or glass—in these cases, there is no or little inelastic behaviour, so that it immediately breaks when moving beyond point B.

There is furthermore a region between point A (the limit of proportionality) and B (the elastic limit) where the object does not behave linearly anymore, but still exhibits elastic behaviour. In other words, even though the extension is non-linear (and thus Hooke's law does not apply anymore), there is no permanent deformation of the material. An example of a type of material that demonstrates such type of behaviour is rubber.

Hooke's law is applicable to the behaviour of springs, as long as the extension does not cause any deformations in the spring. Very often, the limit of proportionality for strings is taken to mean the same thing as their elastic limit, i.e., point A = point B. Be that as it may, the most important point to remember is that a spring abides by Hooke's law, unless it is deformed. If $\vec{F}$ in Equation 3.7 is thought of as the restoring force rather than the applied force, then Hooke's law becomes $\vec{F} = -k\vec{x}$, as the restoring force is directed opposite to the direction of compression or extension.

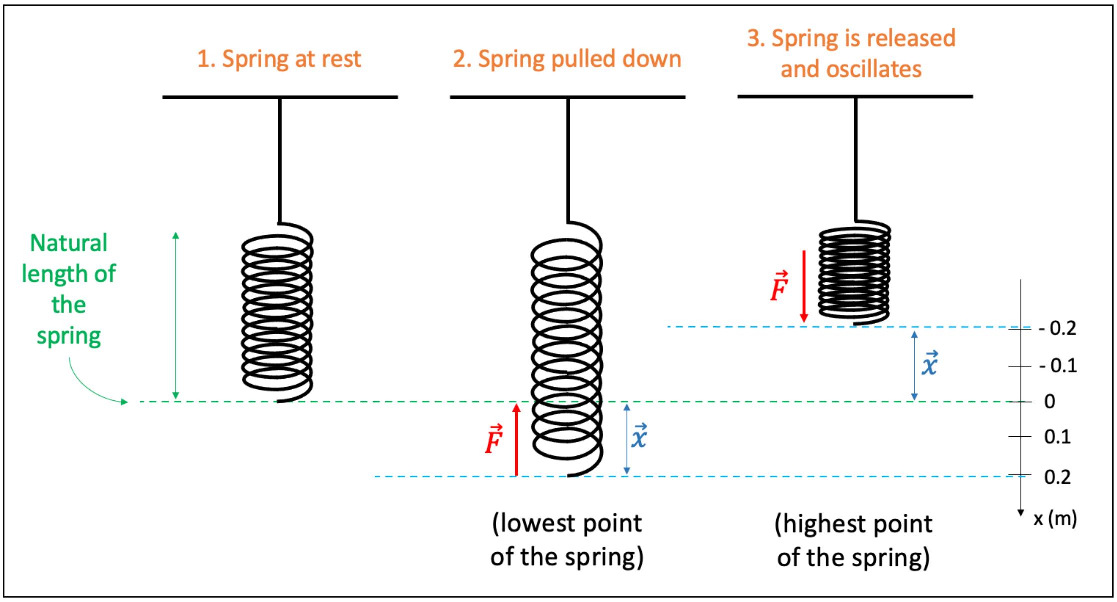


Figure 3.8: The oscillatory motion of a suspended spring

In Fig. 3.8, a spring suspended from the ceiling is in a first instance at rest, then pulled down, and finally released. After being released, the spring engages in a vertical oscillatory movement. During the oscillation, the spring experiences an upward restoring force $\vec{F}$ when it falls below the green line, as it wishes to return to its equilibrium position (the green line). The direction of the force $\vec{F}$ is reversed (i.e., it points downwards) when the spring is moving in the region above the green line.

As a result of the oscillatory motion, the spring has stored energy within its system which is referred to as the *elastic potential energy* E_p. The total amount of energy in the system also includes another form of energy called the kinetic energy E_k, which is related to the motion of the spring (more on this in section 3.7). The potential energy E_p reaches its maximum when the spring is located either at its highest or at its lowest point, since at these two moments during the oscillation the spring briefly stops moving ($E_k = 0$) and all the energy of the system is contained within the potential energy E_p. Conversely, at the moment when the spring crosses the green line ($x = 0$), it reaches its highest speed and all the energy is absorbed by the kinetic energy E_k ($E_p = 0$). The amount of potential energy E_p equals the area under the force-extension graph (see the grey area in Fig. 3.7) and can be calculated as follows:

$$E_p = \frac{1}{2}(F \cdot x) = \frac{kx^2}{2} \tag{3.8}$$

3.4 Newton's Laws

Section 3.1 discusses how objects move over time, while section 3.2 explains the types of forces that can act on objects. The three laws of motion by Sir Isaac Newton make the bridge between these two sections and explain how the forces to which the object is submitted are responsible for its motion.

Newton's first law of motion is known as the principle of inertia, which is defined as the resistance of an object to alter its velocity due to its mass. Stated in words, the first law is summarized in the following way:

<u>Newton's First Law</u> : "a body will remain at rest or in a state of uniform motion in a straight line unless acted on by a resultant external force" (3.9)

In other words, if all the forces within a physical system cancel each other out (i.e., the resultant force $\vec{R}$ has a zero magnitude), the system will not experience any *change in velocity.* Since velocity implies both a speed and a direction, Newton's first law says two things: an object at rest ($v = 0$) will not obtain a higher speed by itself, and an object that moves with a constant speed will not change its direction of motion. Expressed in mathematics, the first law reads (if the mass of the object remains unchanged):

$$\vec{R} = \vec{F}_{net} = \sum \vec{F} = \vec{0} \iff \frac{\Delta \vec{v}}{\Delta t} = \vec{0} \tag{3.10}$$

Newton's second law of motion expresses the fact that any resultant force (also called the net force) acting on an object is proportional to the acceleration of the object:

$$\underline{\textbf{\textit{Newton's Second Law}}} : \qquad \vec{F}_{net} = m \cdot \frac{\Delta \vec{v}}{\Delta t} = m \cdot \vec{a} \tag{3.11}$$

Again, the above equation is valid only if the mass of the object remains constant. Once we define the momentum of a system (see section 3.6), we will see that the first and second law can be expressed more generally in terms of momentum.

Newton's third law of motion is known as the action-reaction law and basically states that every force is always associated with another force, which is equal in magnitude but opposite in direction. If we consider two objects, then we can rephrase the third law as follows:

Newton's Third Law : "if object 1 exerts a force on object 2 then object 2 exerts an equal and opposite force on object 1" (3.12)

We can rewrite the third law as $\vec{F}_{12} = -\vec{F}_{21}$, whereby $\vec{F}_{12}$ ($\vec{F}_{21}$) represents the force that object 2 (object 1) exerts upon object 1 (object 2).

3.5 Mass and Weight

As pointed out in the previous section 3.4, the notion of **mass** is closely related to inertia; mass is described as a property of an object with the characteristic that it resists as much as possible any change in motion. Because of the existence of mass and the presence of gravity, an object undergoes a force called **weight** that always points to the centre of the Earth.

According to Newton, the **universal law of gravitation** posits that the force of gravity $\vec{F}_G$ between two objects is directly proportional to the product of their masses but inversely related to the square of the distance that sits between them. The direction of $\vec{F}_G$ is given by the straight line that connects both objects, and its magnitude equals:

$$F_G = G\frac{m_1 m_2}{r^2} \tag{3.13}$$

with G the gravitational constant ($G = 6.67 \times 10^{-11}\ m^3 \cdot kg^{-1} \cdot s^{-2}$), m_1 (m_2) the mass of object 1 (object 2) expressed in kg, and r the distance between the objects expressed in meter (m).

Invoking Newton's second law of motion (see Equation 3.11), we can see that:

$$F_G = G\frac{m_1 m_2}{r^2} = m_1(\frac{Gm_2}{r^2}) = m_1 a_g \iff a_g = \frac{Gm_2}{r^2} \tag{3.14}$$

whereby F_G can be thought of as the magnitude of the gravitational force between an object (with mass m_1) on the surface of the Earth and the Earth itself (with mass m_2). Then a_g is nothing else but the acceleration that the object experiences if it were entirely unencumbered by external forces (including air resistance), except by gravity. In such a state, it is said that the object is **free-falling** towards the Earth—according to Newton's third law, the Earth is also free-falling towards the object, although that effect is extremely small.

Plugging in the numbers for the mass m_2 of the Earth and the Earth's radius r (i.e., the distance between the Earth's centre and its surface, where the object is positioned), the value of the acceleration a_g is calculated as $a_g = 9.81$ m·s^{-2} (or N·kg^{-1}), or rounded to $a_g = 10$ N·kg^{-1}. In the context of free-falling, a_g is called the **gravitational field strength**, which is indicated by the letter g. Only when we are talking about the gravitational force *that the Earth* exerts upon an object does it make sense to say that the gravitational force equals the **weight** of an object:

$$F_G = m_1 a_g = m_1 g = F_W \tag{3.15}$$

If we were to roam other planets, we would equally refer to our *weight* in view of the gravitational force between our body and the respective planet. However, it is more accurate to speak of the gravitational force between planets or the gravitational force that the Earth exerts upon the moon, instead of the weight of the moon.

Free-falling furthermore implies that all objects, *regardless of their mass*, fall at the same rate, i.e., the gravitational field strength g of the Earth. For instance, if you would simultaneously drop a cannon ball and a feather side by side without any initial push in a chamber where the air is sucked out of it, they would touch the ground *at exactly the same moment*, since they are both submitted to only one force, i.e., gravity. As a matter of fact, even though the Earth is pulling harder on the heavier cannon ball (Equation 3.15), it is also more difficult to get the cannon ball to accelerate—remember that a greater mass means a greater inertia, i.e., the resistance of an object to change its velocity (see Newton's first law of motion, Equation 3.9). As a result, these two effects cancel out and the only force left to have an impact on the objects is gravity.

Performing this experiment in the open air close to the surface of the Earth, we know that it would take longer for the feather to reach the ground. These observations lead us to conclude that on Earth, it is ***air resistance***, *not weight*, that dictates how fast an object falls.

The air resistance, which is a type of drag force, grows *stronger* if the surface of the object is *larger*, if the object travels at *greater* velocity, and if the density of the object is *low* (in other words, *lighter* objects encounter more air resistance). As long as the downward gravitational force during a fall exceeds the drag force issued by the air, the object continues to accelerate, since the system experiences a net force. As the velocity keeps increasing during acceleration, the air resistance builds up accordingly, up to a point where the drag force eventually catches up with the gravitational force. At that moment, the acceleration drops to zero (there is no net force anymore), and the object now falls with a constant velocity, which is designated as the **terminal velocity**. Given that the Earth pulls less intensely on lighter objects and that lighter objects experience more drag, it then follows that lighter objects reach their terminal velocity quicker than their heavier counterparts.

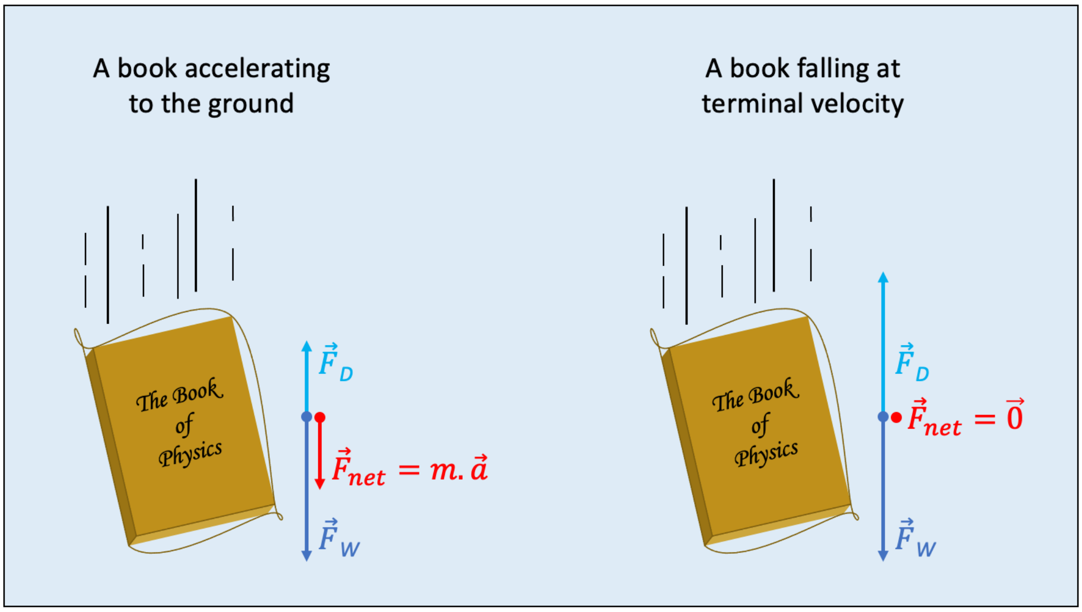


Figure 3.9: The force diagram of a falling book experiencing air resistance

3.6 Momentum

The **momentum** of an object can be described as the amount of motion it possesses, and it is defined as follows:

$$\vec{p} = m \cdot \vec{v} \tag{3.16}$$

with m the mass (in kg), $\vec{v}$ the velocity of the object (in m$\cdot s^{-1}$), and $\vec{p}$ the momentum (in kg$\cdot$m$\cdot s^{-1}$). The more mass an object has, the more momentum it possesses at a certain velocity, and the harder it is to make the object change its velocity (remember that this can refer to either its speed or its course of direction), due to inertia.

As mentioned already in section 3.4, Newton's first and second law of motion are stated for the case of an object whose mass remains constant. A counterexample whereby the mass of an object varies over time includes a rocket engine, which loses mass as it burns fuel. To rewrite these two laws in a more general fashion, we rely on the concept of momentum:

$$\underline{\textbf{\textit{Newton's First Law}}}: \quad \frac{\Delta\vec{p}}{\Delta t} = \vec{0}$$

$$\underline{\textbf{\textit{Newton's Second Law}}}: \quad \vec{F}_{net} = \frac{\Delta\vec{p}}{\Delta t} \tag{3.17}$$

In other words, saying that a force equals mass times acceleration means exactly the same thing as saying that a force acting on an object can be defined as the rate by which the momentum of the object changes over time. The following calculations show how the Equations 3.17 are equivalent to Equations 3.10 and 3.11 when the mass is assumed constant (whereby Equation 3.16 and 3.3 are used during the first and third step, respectively):

$$\frac{\Delta\vec{p}}{\Delta t} = \frac{\Delta(m\vec{v})}{\Delta t} = m \cdot \frac{\Delta\vec{v}}{\Delta t} = m \cdot \vec{a} = \vec{F}_{net} \tag{3.18}$$

If $\frac{\Delta\vec{p}}{\Delta t} = \vec{0}$, it follows then that both $\frac{\Delta\vec{v}}{\Delta t} = \vec{0}$ and $\vec{F}_{net} = \vec{0}$, as per Newton's first law (see Equation 3.10).

For an isolated physical system, i.e., a system that is not under the influence of an external net force, the **law of conservation of momentum** says that the change in the total momentum of that system is always zero. That is to say, although this law still allows for changes in the momentum within this isolated system, the initial momentum and the final momentum of the *total* system measured over a certain period of time must remain the same.

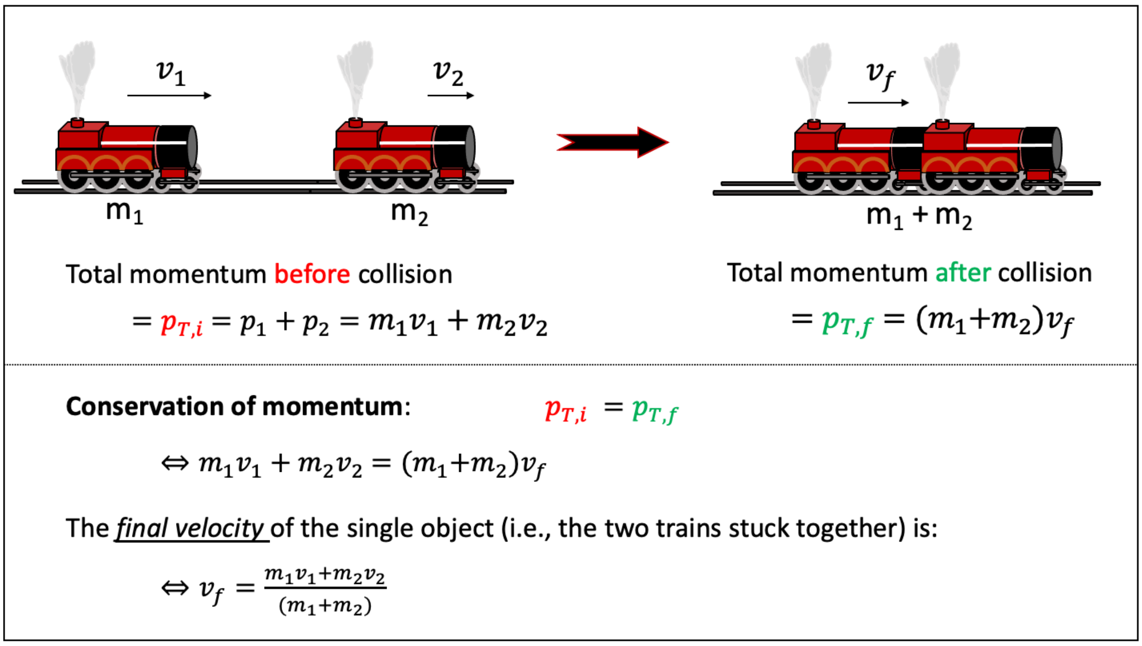


Figure 3.10: The law of conservation of momentum applied to an inelastic collision

As an example, consider the isolated system of two miniature trains on a straight railway track in Fig. 3.10. They both move in the same direction (i.e., to the right), but the first train possesses a higher speed, so that the trains eventually collide. After the collision, they stick together (through magnets) and now move (still to the right) as a single object along the track—this type of collision is called an *inelastic collision*. The law of conservation of momentum states that the total momentum before the collision must be equal to the total momentum after the collision, and the law can be used to calculate the one unknown variable in this scenario: the final velocity of the merged object after the collision.

3.7 Energy

The concept of **work** in physics points to the idea that **energy is being transferred** from one system to another. If you push your car and manage to get it moving in the same direction that you are pushing, the force applied to your car delivers energy to the system in the form of kinetic energy (i.e., energy of motion), as your car is now moving. Put differently, the force exerted upon the car does *positive work* on the car, since it transfers energy *towards* the car. Given that Newton's third law states that to every action there is an equal and opposite reaction (see Equation 3.12), the normal force that the car exerts upon you is doing *negative work*, as it transfers energy *away* from you.

In order for a system to perform work, three criteria must be satisfied: there has to be a force $\vec{F}$, there has to be a displacement $\Delta\vec{x}$ that is *caused by the force*, and the direction of the applied force cannot be perpendicular to the direction of the displacement. In the language of mathematics, work is defined as:

$$W = F \cdot \Delta x \cdot \cos\theta \tag{3.19}$$

with W the work done (expressed in units of energy, i.e., Joule (J)), F the magnitude of the applied force (expressed in N), Δx the magnitude of the displacement (expressed in m), and θ the angle between the force $\vec{F}$ and the displacement $\Delta\vec{x}$. Note that if the angle θ is less (more) than 90°, positive (negative) work is being done by the force.

An example that implies the *absence* of work done is the situation whereby you attach an object to one end of a rope and perform a continual circular motion by holding the other end of the rope. Even though there is a force present pointing from the object towards the centre of the imaginary circle, i.e., the centripetal force (supplied by the tensional force in the rope), as well as a displacement (the object is moving along the tangent at each point of the circle), the work done by the centripetal force is zero, since the angle between the direction of motion and the force equals 90° at all times.

As a matter of fact, every time a horizontal motion occurs the force of gravity never does any work on the object, given that it points downwards and thus always makes an angle of 90° with the direction of motion.

Consider a couple of additional examples. Carrying a heavy bag of garden waste to a container 100 m further away is an example of *not* doing any work, as the upward force required to hold the bag does not cause the horizontal movement. When a coffee mug falls off a table top, the gravitational force (pointing downwards) is doing *positive* work during the fall, as both the force and the displacement vectors point in the same direction (or, alternatively, the angle θ between them is $0°$). Preventing a sofa from sliding down the staircase during a relocation by letting it lean against your back (while you are waiting for your partner to show up) is a *non*-example of work, since there is no displacement. Finally, the air resistance is performing *negative* work on your car while driving at a constant velocity, given that the drag force and the direction of motion point in the opposite direction (or, alternatively, the angle that they make is equal to $180°$).

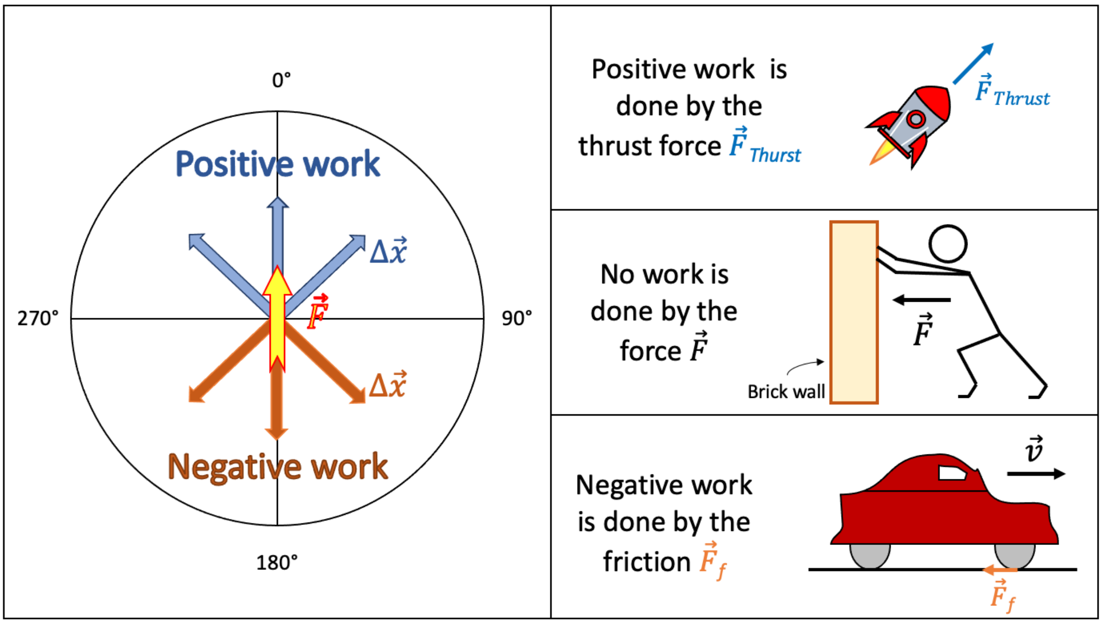


Figure 3.11: The concept of work in physics

A concept related to the gravitational force is the **gravitational potential energy**. We already touched upon the notion of potential energy in the context of electrical circuits (see section 1.2) and springs (see section 3.3) and described it as a form of stored energy in a physical system. The gravitational potential energy *at a certain position* x above the surface of the Earth can then be understood as an amount of energy stored within the gravitational field, whereas the work done *during a certain displacement* along the vertical direction by an external force, i.e., a force other than gravity, is equal to the *difference* between the gravitational potential energy at the final and initial position of the motion:

$$\begin{cases} E_{p,G} = mgx \\ \\ W_{ext} = E_{p,G}(x_f) - E_{p,G}(x_i) = mg(x_f - x_i) = mgh \end{cases} \tag{3.20}$$

with $E_{p,G}$ the gravitational potential energy (expressed in Joule (J)) of an object that is positioned at point x along the vertical direction, m the mass of the object (expressed in kg), g the gravitational field strength ($g = 9.81$ N$\cdot$kg^{-1}), W_{ext} the work done by an external force (i.e., a force other than gravity) in the vertical direction, $E_{p,G}(x_i)$ ($E_{p,G}(x_f)$) the gravitational potential energy at the initial (final) position x_i (x_f), and $h = x_f - x_i$ the vertical displacement or difference in height (expressed in meter (m)).

This change in gravitational potential energy manifested by the external force can equally be defined as the negative of the work done by the gravitational force:

$$E_{p,G}(x_f) - E_{p,G}(x_i) = -W_G = W_{ext} = mg(x_f - x_i) = mgh \tag{3.21}$$

Lifting a box filled with course notes above your head means that the (external) force you exert upon the box is doing positive work. What is more, the energy that your force is transferring to the box (that is, the work performed on the box) is stored as gravitational potential energy in the gravitational field. That is, the gravitational potential energy of the field is *increasing*—in Equation 3.20, this means that point x_f is located at a higher altitude than point x_i. In the meantime, the gravitational force is doing negative work, as the direction of motion of the box is opposite to it (as in agreement with Equation 3.21). Fig. 3.12 provides the opposite example of lowering a box towards the floor.

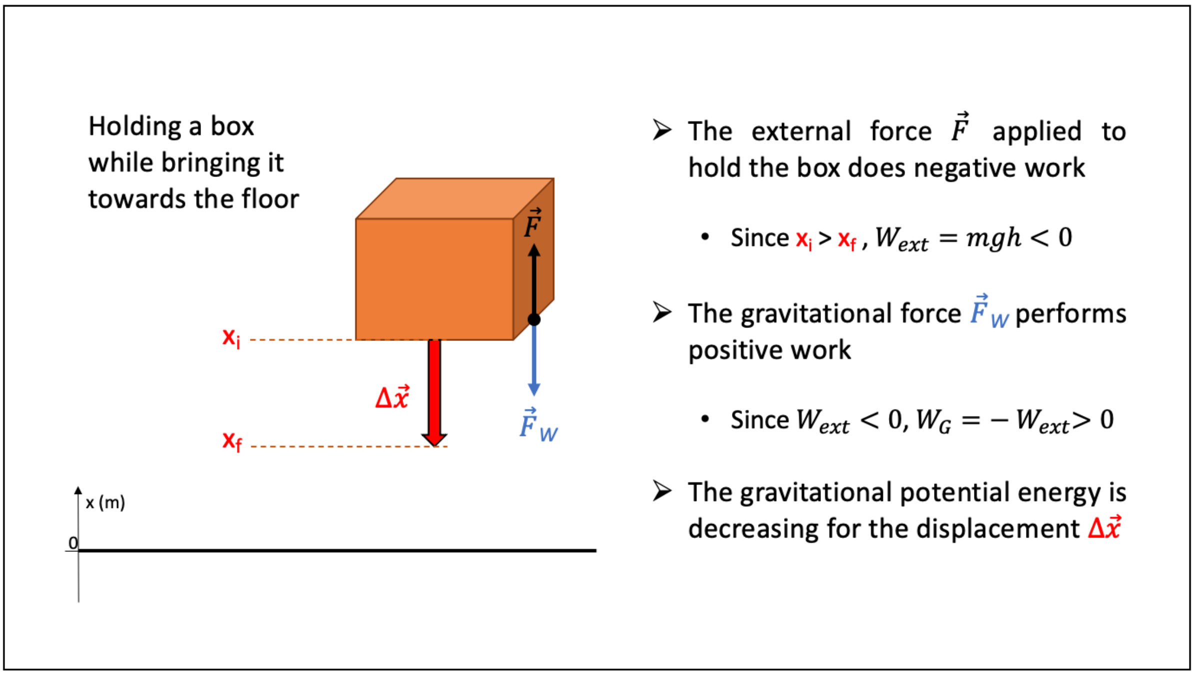


Figure 3.12: An example of a decreasing gravitational potential energy

Imagine you are holding a toothbrush in your hand and in the next moment you drop it on the floor—in Equation 3.20, this means that point x_i is now located at a higher altitude than point x_f. While the gravitational force is doing positive work on the toothbrush as it falls, the amount of gravitational potential energy in the field is *decreasing* as its stored energy is being released and converted into energy of motion (kinetic energy). In other words, the gravitational potential energy of the toothbrush in your hand is higher than that of the toothbrush lying on the floor. This time, it is the air resistance that is performing negative work.

As already alluded to in section 3.3 on springs, the **total amount of mechanical energy** of a system consists of both the potential energy E_p and the energy of motion, i.e., **kinetic energy $\boldsymbol{E_k}$**, and are defined as:

$$\begin{cases} E_k = \dfrac{mv^2}{2} \\ \\ E_{Tot} = E_k + E_p = \dfrac{mv^2}{2} + mgh \end{cases} \tag{3.22}$$

with the kinetic energy E_k, the potential energy E_p, and the total mechanical energy E_{Tot} expressed in Joule (J).

If the total amount of energy in a system equals the total amount of *mechanical* energy, then the system is referred to as a *conservative* system. That is, the only forces present are **conservative forces**, which are characterized by the feature that their work done on an object only depends on the initial and final position of the displacement and not on the path taken. Examples are the gravitational force, the electrostatic force, the magnetic force, and spring forces. For instance, for an asteroid approaching Earth, the change in gravitational potential energy (i.e., the work done by gravity) between two different altitudes remains the same whether it crosses that difference following a straight line or whether it traverses it under a certain angle (which means a longer trajectory), as it is only the difference in altitude relative to Earth that matters to the work done by gravity.

This is not the case for **non-conservative forces**, since the path taken *does affect* the amount of work done by these forces. The forces that fall under this category include applied forces, friction, drag forces (air resistance), tensional forces, and the normal force. For instance, pushing a box across the floor from point A to B via a longer, curved path instead of following a straight line increases the work, as the force of friction has a greater time window (and more distance to cover) during which it can act upon the box.

Another difference between these two types of forces is that conservative forces are *always* associated with a potential energy. Moreover, for conservative systems the ***total amount of mechanical energy is conserved***, meaning that, regardless of the individual values for E_k and E_p, the change in the total energy always equals zero:

$$\begin{aligned} &\Longrightarrow E_{k,i} + E_{p,i} = E_{k,f} + E_{p,f} \\ &\Longleftrightarrow E_{k,f} - E_{k,i} = -(E_{p,f} - E_{p,i}) \\ &\Longleftrightarrow \Delta E_k = -\Delta E_p \\ &\Longleftrightarrow \Delta E_k + \Delta E_p = 0 \\ &\Longleftrightarrow \Delta E_{Tot} = 0 \end{aligned} \tag{3.23}$$

with $E_{k,i}$ ($E_{p,i}$) and $E_{k,f}$ ($E_{p,f}$) referring to the kinetic (potential) energy of the initial and final state of the system, respectively. If any non-conservative force would act on the system, the total mechanical energy would not be conserved.

To rephrase the law of energy conservation: energy is exchanged exclusively between the potential and the kinetic energy—remember that to establish an exchange (transfer) of energy a force must actually perform work. In other words, there are no other types of energy present, such as thermal energy (heat) produced by the non-conservative force of friction, which would *dissipate mechanical energy from the system.*

In fact, the work done by any non-conservative force either adds mechanical energy to a system or removes some amount from it. This means that in the previous example of the box being pushed across the floor via the longer path, the force of friction is able to dissipate a greater amount of mechanical energy, which not only clarifies that the amount of work done depends on the path taken, but it also explains why the total amount of mechanical energy cannot be conserved when non-conservative forces are present.

Starting from Newton's second law of motion (see Equation 3.11), an interesting relationship arises—known as the **work-energy theorem**—between the notion of work W and the change in kinetic energy ΔE_k (making use of Equations 3.2, 3.3, 3.6, 3.19, and 3.22):

$$\begin{aligned}
&\Longrightarrow F = m \cdot a \\
&\Longleftrightarrow F = m \cdot \frac{\Delta v}{\Delta t} = m \cdot \frac{\Delta x}{\Delta t^2} \\
&\Longleftrightarrow F = \frac{m}{\Delta t^2} \cdot \left(\frac{v^2 - v_0^2}{2a}\right) \\
&\Longleftrightarrow F = m \cdot \frac{(v^2 - v_0^2)}{2\Delta x} \\
&\Longleftrightarrow F \cdot \Delta x = \frac{mv^2}{2} - \frac{mv_0^2}{2} \\
&\Longleftrightarrow F \cdot \Delta x = E_{k,f} - E_{k,i} \\
&\Longleftrightarrow W = \Delta E_k
\end{aligned} \tag{3.24}$$

From the conservation of mechanical energy (see Equation 3.23) we then obtain that $W = -\Delta E_p$ (given that $\Delta E_k = -\Delta E_p$), which is exactly what Equation 3.21 describes: an increase (decrease) in gravitational potential energy is the result of gravity doing negative (positive) work.

Apart from understanding the amount of work a system is delivering, it may equally be helpful to gain insight into the question of how much work the system can do within a specific timeframe, i.e., the rate of work. The concept of **power** provides an answer to that question—we mentioned power earlier in section 1.2 in the context of batteries and electricity. With the assistance of the definition of both work (see Equation 3.19, whereby we assume that $\vec{F}$ and $\Delta\vec{x}$ are parallel to each other) and velocity (see Equation 3.2), the notion of power can be formulated as follows:

$$P = \frac{W}{\Delta t} = \frac{F \cdot \Delta x}{\Delta t} = F \cdot v \tag{3.25}$$

with the power P expressed in Watt (W). Not all of the energy that a system generates is beneficial for the intended purposes. The amount of energy produced that is valuable for further processes is called **useful energy**, while the part that is less desirable is referred to as **wasted energy**, which usually takes the form of heat—heat is the amount of energy *transferred* between systems—and in some cases also sound or light.

For instance, even though the objective of an incandescent light bulb is to generate light (useful energy), it also gives off a substantial amount of heat (wasted energy). Food processors and lawn mowers convert electrical energy into useful mechanical energy, but produce noise (sound) as an unwanted form of energy. A television set provides entertainment because of the energy of light and sound (useful energy), but the accompanying heat needs to kept at a minimum. Finally, a hairdryer is designed to deliver heat (useful energy), despite the noise that comes with it.

During any real-world physical process, there is *always* wasted energy, and a measure used to quantify the proportion of useful energy with respect to the total input of energy is known as the **energy efficiency**:

$$\begin{aligned} \Longrightarrow \textit{ energy efficiency (\%)} &= \frac{\textit{useful energy output}}{\textit{total energy input}} \times 100 \\ &= \frac{\textit{useful power output}}{\textit{total power input}} \times 100 \end{aligned} \tag{3.26}$$

4 Thermal Physics

In section 3.7, we referred to **heat** as the amount of energy transferred between systems. In the following sections, we discuss the three modes by which heat can be transmitted: conduction, convection, and radiation.

4.1 Conduction

Conduction is the transfer of thermal energy (heat) in solid materials as a result of temperature differences, whereby the free electrons help mediate the dispersion of energy. The heat transfer takes place either within one and the same object or after *physical contact* between different objects, and always runs in the direction from warmer to colder regions.

The following mechanism explains how the energy is transferred across the material by means of conduction. More heat means that an atom, which is held in a *fixed position* in solid materials, is vibrating at a higher rate—it has effectively gained a higher kinetic energy. Due to its increased vibrational motion, the atom not only transmits these vibrations throughout the material structure, but it also sheds its attached electrons more easily. As these free electrons move around, they bump into the tightly bound electrons of neighbouring atoms, which absorb the kinetic energy and enhance their own vibrations. These neighbouring atoms, in turn, release their electrons, which pass along their kinetic energy to their neighbours, and so on. This exchange of energy between the atoms continues to develop until they all vibrate at the same rate. That is, until the thermal energy is on average evenly distributed across the material so that the net flow of energy is zero, i.e., the surface is in a state of thermal equilibrium and heat will not flow. In this view, **temperature** can be defined as a measure of the average kinetic energy of all the atoms combined within a material.

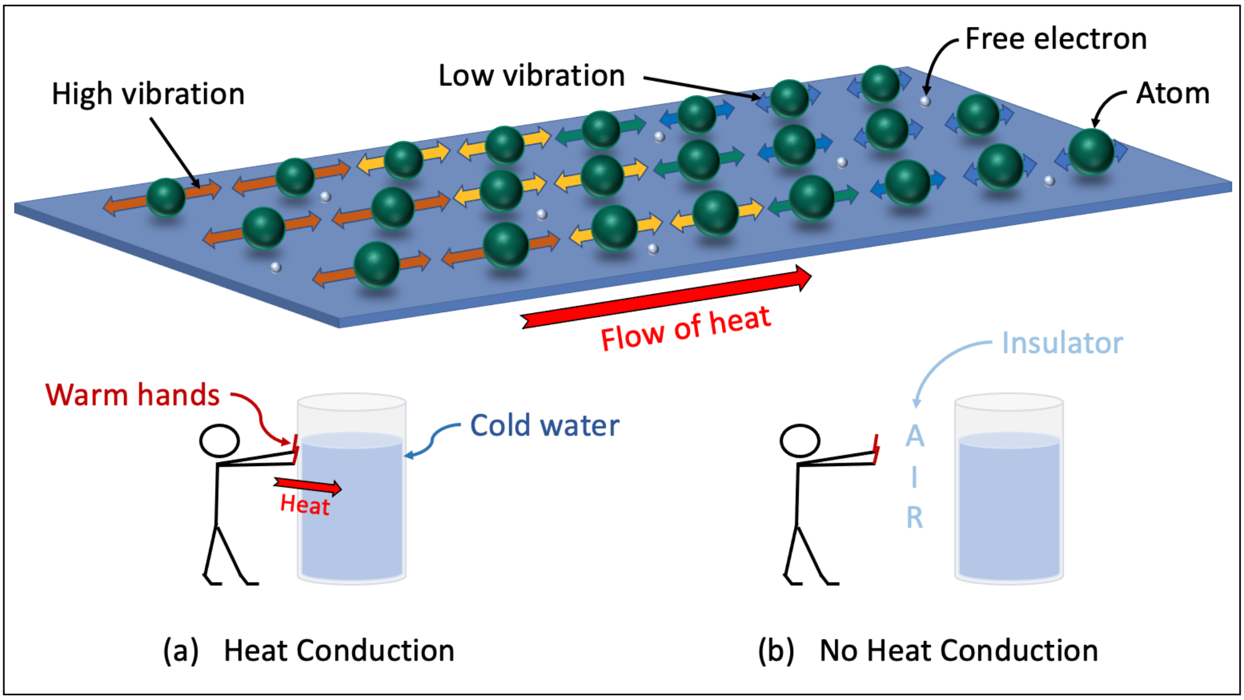


Figure 4.1: The principle of conductive heat transfer

Metals are materials that already contain lots of free electrons to start with, as the atoms of these metals have a greater tendency to give away their electrons. In other words, metals possess the property that electrons can easily move across their surface. As such, they greatly facilitate the transfer of kinetic energy—and thus heat—among atoms. Materials that exhibit these features are called **thermal conductors**. Examples include the materials diamond, silver, copper, gold, aluminum, iron, and steel.

By contrast, the electrons in *non-metals* are firmly attached to the atoms, so that the process of heat transfer takes place only at a very slow rate. These materials hardly contain any free electrons, which is a key characteristic of **thermal insulators**; they want to prevent the flow of heat through their surface as much as possible. In fact, the energy in non-metals is mainly passed on through vibrations of the atoms, rather than collisions with free electrons. Examples are mineral wool, fiberglass, air, cellulose, polystyrene, fur, polyurethane foam, water, and rubber (see also section 1.1).

A number of factors impact the **rate of conduction**. How the property of materials influences conduction is reflected by the **constant k of thermal conductivity**; solid materials tend to have a *larger value for k* than fluids, given that the atoms are much more packed together in solids (which enables heat transfer), which *enhances* conduction. A *greater* **temperature gradient**, which signals the direction and the rate of the change in temperature, means a *larger difference in temperature*, which implies a *faster* flow of heat from warmer to colder regions. Finally, a *larger* **contact area** between objects *increases* the rate of heat flow, whereas a *longer* **path length** (i.e., a greater thickness of the material) *lowers* the rate of conduction. In mathematical language, the rate of conduction becomes:

$$q = \frac{\Delta Q}{\Delta t} = k \cdot A \cdot \frac{(T_{Warm} - T_{Cold})}{\Delta x} \tag{4.1}$$

with q the rate of conduction (expressed in $J{\cdot}s^{-1}$ or Watt W), ΔQ the amount of heat flow (in joule J), Δt the duration of the heat flow (in seconds), k the constant of thermal conductivity (in $W{\cdot}m^{-1}{\cdot}K^{-1}$), A the contact area (in m^2), T_{Warm} (T_{Cold}) the temperature of the warmer (colder) region (in Kelvin K), and Δx the path length or thickness of the material (in m).

When speaking of the flow of electrons, it is reminiscent of the concept of electricity (see section 1.2), and, indeed, poor (good) conductors of heat are *usually* also poor (good) conductors of electricity. Bear in mind, however, that the effect of temperature makes this relationship more subtle: an increase in temperature generally reduces electrical conductivity for metallic conductors, while the thermal conductivity is boosted for gases as well as non-metals and decreased for liquids and metallic conductors.

4.2 Convection

Whereas in conduction heat transfer occurs *without* the movement of atoms (i.e., mass), in **convection** the flow of heat is manifested through the motion of mass. Moreover, while conduction is the key mode of energy transfer for solid materials, convection is the dominant process for heat transport in *fluids* (i.e., liquids and gasses). And contrary to conduction, which is sustained by temperature differences, heat exchange in convection is driven by differences in fluid density.

In the example of a fluid contained inside a box made of a conductive material that is being heated at the bottom, the mechanism for convective heat transport works as follows. The constitutive elements of a fluid called molecules (i.e., specific groupings of atoms) bump into a surface of higher temperature and receive additional kinetic energy from the atoms embedded in that surface (which is basically conductive heating). That particular region of the fluid undergoes a subsequent increase in temperature, and as the atoms within the molecules now vibrate stronger, they make the molecules grow a bit in volume, effectively making them *less dense*. As discussed in section 3.2, an object with a lower density that sits in a fluid with a higher density experiences an upward buoyant force—in addition, that force grows stronger when the object expands in volume—so that it starts moving upwards. As a result, the molecules with lesser density *start rising* in the fluid. At the same time, the molecules within the colder regions now move downwards, since they have become more dense *relative to* the rising particles. When they then bump into that surface of a higher temperature, they undergo the same process as just described for the first batch of particles. What is more, during the upwards motion of less dense molecules, these more energetic molecules come into contact with less energetic molecules of colder regions, thereby donating part of their kinetic energy to these slower-moving molecules, which are on their way down. This constant circulation of molecules throughout the fluid as a result of heat exchange is referred to as a *convection current*.

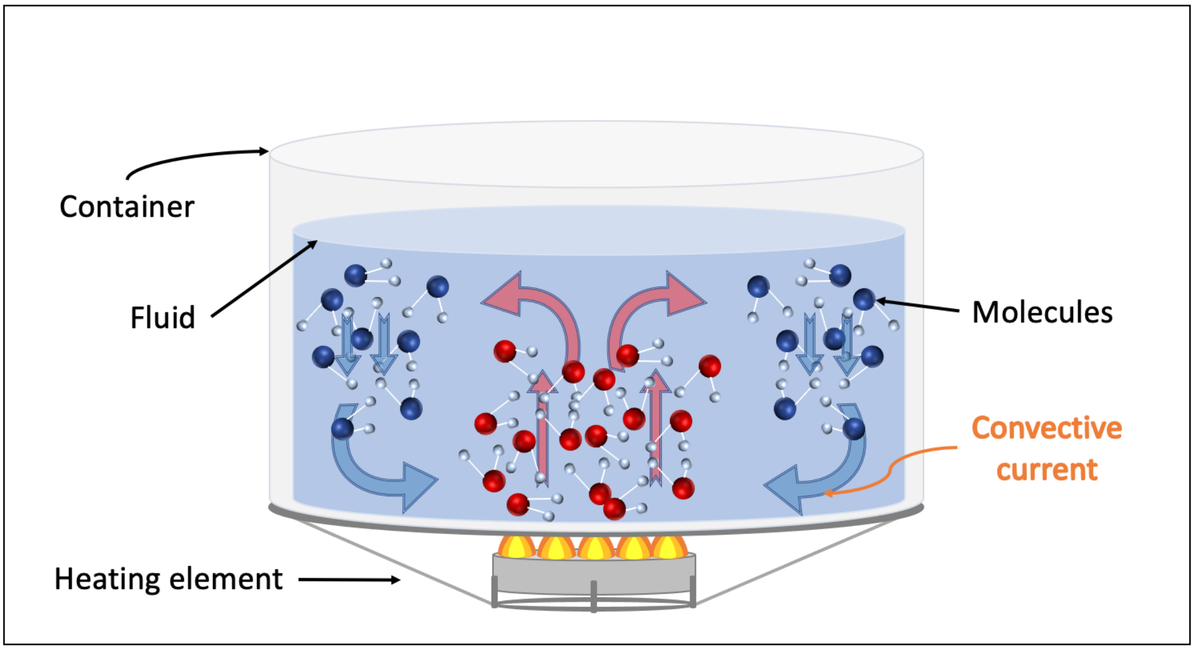


Figure 4.2: The principle of convective heat transfer

The above description is an example of **natural convection**, because the motion of the fluid emerges spontaneously as a consequence of differences in fluid density. Another example is letting a freshly baked cake cool down at room temperature; as the highly energetic food molecules pass on their kinetic energy to the surrounding air molecules, the latter molecules start rising (due to their lower density), thereby making room for colder molecules, and naturally generating a convective current. Other examples include land and sea breezes, conventional ovens, hot air balloons, and defrosting frozen food in the open air.

In contrast, during **forced convection**, external factors, such as pumps, fans, utensils, and sucking devices, influence the motion of the fluid so that the heat flow can be used optimally. Applications of this type of convection include hair dryers, air conditioning, convection ovens, central heating, stirring in a cooking pot, refrigerators, car radiators, blood circulation, and steam turbines.

A couple of factors influence the **rate of convection**. A *larger* **contact area** between the fluid and the surface of the warmer object as well as a *greater difference* between the surface and the air **temperature** *enhance* the rate of heat transfer. Also, a *larger* **convective coefficient $\mathbf{h_c}$**, which depends on the physical properties of the fluid and on the physical situation, pushes up the rate of convection. In equation form, the rate of convection is expressed as follows:

$$q = \frac{\Delta Q}{\Delta t} = h_c \cdot A \cdot (T_{Surface} - T_{Fluid}) \tag{4.2}$$

with q the rate of convection (in Watt W), ΔQ the amount of heat flow (in joule J), Δt the duration of the heat flow (in seconds), h_c the convective heat transfer coefficient (in W·m^{-2}·K^{-1}), A the heat transfer area of the surface (in m^2), and $T_{Surface}$ (T_{Fluid}) the temperature of the surface (the fluid) (in Kelvin K).

4.3 Thermal Radiation

The atoms of any object with a temperature above absolute zero (i.e., 0 Kelvin or -273.15° Celsius) vibrate, since the temperature is defined as the average kinetic energy of the atoms within that object. Due to the vibrational motion of the atom, the electrically charged electron attached to the atom recurrently changes its velocity, leading to the production of *electromagnetic waves (a.k.a. electromagnetic radiation)*, which are oscillating electric and magnetic fields coupled to one another and *carry energy with them.*

This electromagnetic radiation spontaneously emitted by matter (including the human body) is referred to as **thermal radiation**, and falls largely within the range of infrared radiation, but can also extend into the region of visible light and ultraviolet radiation (see section 6.5), *depending on the temperature of the object.*

Radiation heat transfer is then defined as the exchange of energy (i.e., heat) between objects by way of thermal radiation, which arises as a result of the objects possessing a non-zero temperature (in Kelvin). The higher the temperature of the object, the greater the amount of energy transferred—ultraviolet radiation contains more energy than visible light, which harbours more energy than infrared radiation. Put differently, the intensity of radiation of a warmer object is greater than that of a colder object.

Thermal radiation does furthermore *not require any medium* to carry out the process of heat transfer, unlike conduction and convection, which rely on solid materials and fluids, respectively. That is, heat can be transported between objects even when they are separated by empty space (i.e., vacuum); this is for instance how the Sun conveys its radiative energy to the Earth. What is more, electromagnetic radiation travels very fast, i.e., at the speed of light (299,792,458 m$\cdot s^{-1}$), and is emitted in all directions.

The electron in an atom is able to hold on to an additional amount of energy for a short period of time. When electromagnetic radiation is delivering this extra bit of energy to the electron, the process of energy accumulation is called **absorption of radiation**. When the electron then gets rid of this energy surplus, it sends out electromagnetic radiation, which is referred to as **emission of radiation**, with roughly the same amount of energy that was initially delivered to the electron. That is, when an electron absorbs thermal radiation, it will eventually emit thermal radiation. Moreover, absorption and emission only occurs when the amount of energy corresponds to a very specific quantity, i.e., a photon.

The rate of absorption and emission depends on the temperature of the objects, the nature of the surface of the material, and the emitting object's surface area. With regard to the **temperature**, a warmer object emits energy at a greater rate than that it absorbs, thereby decreasing its temperature. Conversely, a colder object absorbs faster, so that its temperature rises. This means that the exchange of heat still flows from warmer to colder objects—note that, to calculate the *net* radiation heat transfer of an object, one has to subtract the incoming energy (absorption) from the outgoing energy (emission). Also, if objects maintain equal rates of absorption and emission, the temperature will remain constant.

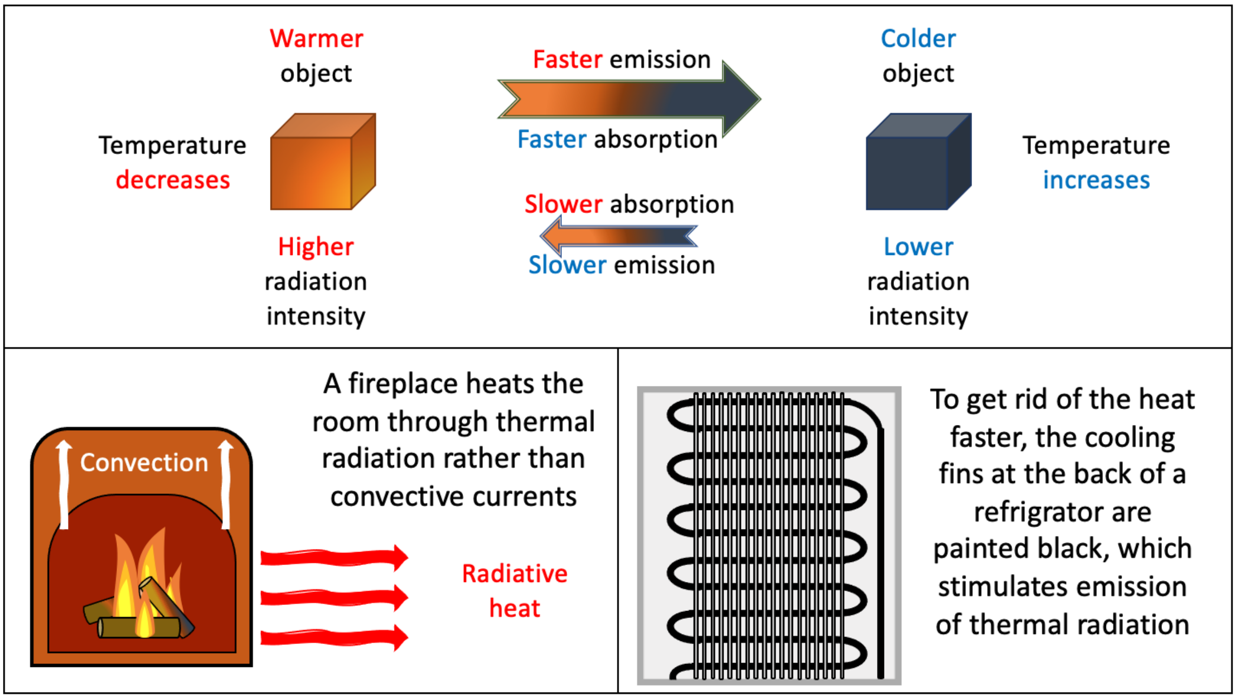


Figure 4.3: The principle of radiative heat transfer

When it comes to **the type of surface**, shiny silvery, mirror-like metallic (polished) surfaces are the poorest absorbers and emitters of thermal radiation, while the opposite usually holds for darkened, weathered, and matte surfaces. In general, poor (good) absorbers make poor (good) emitters as well as *good (poor)* reflectors. Examples of good absorbers and emitters include ice, water, glass, bricks, asphalt, wood, paper (including white paper), and snow. An exception to this rule is for instance white paint, which is a poor absorber but a very efficient emitter of thermal radiation.

The rate of radiative heat transfer furthermore increases with a larger **surface area** of the warmer object. Converting these three factors into mathematics, we obtain this equation for the *net* rate of radiation heat transfer:

$$q = \frac{\Delta Q}{\Delta t} = \epsilon \cdot \sigma \cdot A_{Warm} \cdot (T_{Warm}^4 - T_{Cold}^4) \tag{4.3}$$

with q the net rate of radiative heat transfer (in Watt W), ΔQ the amount of heat flow (in joule J), Δt the duration of the heat flow (in seconds), ϵ the emissivity coefficient of the warmer object with values between 0 and 1 (a higher ϵ equals a better emitter), $\sigma = 5.67 \cdot 10^{-8}$ $\text{W} \cdot \text{m}^{-2} \cdot \text{K}^{-4}$ the Stefan-Boltzmann constant, A_{Warm} the surface area of the warmer object (in m^2), and T_{Warm} (T_{Cold}) the temperature of the warmer (colder) object (in Kelvin K).

4.4 Heat Capacity

The specific heat capacity c is a way to quantize the capacity of a material to store energy due to the transfer of heat. It is defined as the amount of energy that is required to change the temperature of a specific material substance of 1 kg of mass by a difference in temperature of 1 Kelvin (or 1°C):

$$c = \frac{1}{m} \cdot \frac{\Delta Q}{\Delta T} \tag{4.4}$$

with c the specific heat capacity (expressed in $\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$), m the mass of the substance (in kg), ΔQ the required energy (in joule J), and ΔT the change in temperature (in K). If the temperature difference is expressed in °C, then the units of c change to $\text{J} \cdot \text{kg}^{-1} \cdot {}^\circ\text{C}^{-1}$.

The specific heat capacity varies according to the state in which the substance of matter finds itself, i.e., solid, liquid, or gas (see section 5.1), the type of material, and the temperature of the object. For example, for steam that measures 110°C, we find $c = 2,010$ $\text{J} \cdot \text{kg}^{-1} \cdot {}^\circ\text{C}^{-1}$, while for liquid water at 20°C, $c = 4,182$ $\text{J} \cdot \text{kg}^{-1} \cdot {}^\circ\text{C}^{-1}$. This means that if we had to increase the same mass of steam and liquid water by the same amount of temperature, it would take much longer for water than for steam to reach that new temperature. Conversely, it would take much longer for water to cool down as well. Another example is the difference between aluminum and lead. The specific heat capacity of aluminum is higher ($c = 897$ $\text{J} \cdot \text{kg}^{-1} \cdot {}^\circ\text{C}^{-1}$) than that of lead ($c = 129$ $\text{J} \cdot \text{kg}^{-1} \cdot {}^\circ\text{C}^{-1}$), meaning that aluminum would radiate heat longer after it has been taken out of an oven.

In exercises, the specific heat capacity c is usually provided (see Table 4.1), and one is then asked to find the final temperature, or the heat transferred. For instance, the amount of heat needed to bring a pot of 1 liter of water at 20°C to a temperature of 100°C equals

$\Delta Q = c{\cdot}m{\cdot}\Delta T = 4,182 \times 1 \times (100-20) = 334,560$ J (note that 1 L of water is equal to 1 kg of water). Another example might consist of finding the final temperature of a brick of 4 kg at 18°C when exposed to 500,000 J of heat. The answer is $T_2 = T_1 + \frac{\Delta Q}{mc} = 18 + \frac{500{,}000}{4\times 840} = 167$°C.

Table 4.1: The specific heat capacity c for different materials

Substance	Specific Heat c (in $J{\cdot}kg^{-1}{\cdot}°C^{-1}$)	Substance	Specific Heat c (in $J{\cdot}kg^{-1}{\cdot}°C^{-1}$)
Aluminum	897	Olive Oil	1,790
Brick	840	Paper	1,336
Clay (sandy)	1,381	Polypropylene	1,920
Copper	385	Porcelain	1,085
Helium	5,193	Silver	235
Hydrogen	14,304	Steel	490
Ice (-5°C)	2,090	Water (20°C)	4,182
Lead	129	Zinc	388

5 Matter

5.1 States of Matter

In general, the matter that we observe around us comes in three states: solid, liquid, and gas. At the microscopic level, matter is made of particles, which can exist either in a fundamental form, such as electrons, or in a more aggregate form, e.g., atoms, molecules, and ions (which are charged atoms or molecules; see section 7.1). The specific *arrangement and behaviour* of particles in matter determine whether matter is classified as a solid, liquid, or gas.

In **solids**, particles are closely packed together and remain in a locked position—an exception are individual electrons in metals, which can easily move around (see section 4.1). Typically, this gives rise to a material with a microscopically regular pattern; very orderly solids are said to be crystalline. Although they stay put in their location, there is still room though for the particles to vibrate, which they do as a result of their thermal energy. Because of the overall rigid structure, not only can a solid maintain its own shape and volume when it is placed in a container, but it is also strong, hard to press together, and difficult to break apart.

In fact, when it comes to the binding structure of matter, there are broadly speaking three types of chemical bonds that hold atoms together *within* molecules or compounds: covalent bonds, ionic bonds, and metallic bonds. *Covalent bonds* are bonds among non-metals whereby pairs of electrons are shared, as in the case of diamond or graphite. The other two types of bonds rely on electrostatic forces due to the difference in polarity (see section 1.1). That is, while *ionic bonds* are formed between positively charged metal ions and negatively charged non-metal ions (e.g., the ions Na^+(metal) and F^-(non-metal) combine to create the compound sodium fluoride NaF), *metallic bonds* arise from the attraction between positively charged metal ions within a metallic substance and the free-roaming negatively charged bonding electrons (which is the case, for instance, for copper and tungsten).

With respect to interactions *between* molecules, ions, or compounds, there is a fourth type of chemical bonding. The so-called *weak inter-molecular bonds* are moulded between chemical substances following an asymmetry within their internal charge distribution, which leads to the formation of electrostatic forces (for instance, in paraffin wax and caffeine).

Covalent bonds are generally considered to be the strongest type of chemical bonding, followed by ionic and metallic bonds on roughly equal footing, and the weak intermolecular bonds taking up the last position in the ranking.

In **liquids**, the particles are less tightly packed together than in solids and move around rather easily, which leads to a random microscopic arrangement. In other words, physically breaking up a liquid is done effortlessly (as you can tell from holding your hands under a tap of running water). Another difference with solids is that when a liquid of a certain volume is introduced in a container, it takes on the shape of the container. Yet, it retains its volume—much like solids, liquids have a fixed volume.

The molecules in matter are held together by weak intermolecular bonds, but these attractive intermolecular forces have lost a bit of their strength in liquids with respect to solids since the particles are no longer located in fixed positions. However, they are still strong enough to keep the ensemble of particles from disintegrating. Nonetheless, both liquids and solids are difficult to compress, given that the spacing between particles is minimal.

In **gases**, the distance between particles significantly increases, which enables them to move around with much higher velocities. This also means that a gas can be easily compressed and that its microscopic arrangement is completely random. The attractive intermolecular forces in gases are furthermore no longer strong enough to hold the gas together, so that a gas naturally expands. As a result, and contrary to solids and liquids, when a gas is inserted in a container, it does not maintain its volume and is dispersed across the entire container until its volume matches that of the container.

In hindsight, the final state of matter essentially depends on the specific interplay between the kinetic energy of the particles, which drives them apart, and the intermolecular forces involved, which draw the particles closer together. In gases, the kinetic energy gains the upper hand, while solids are firmly disciplined by the intermolecular forces, with liquids somewhere stuck in the middle.

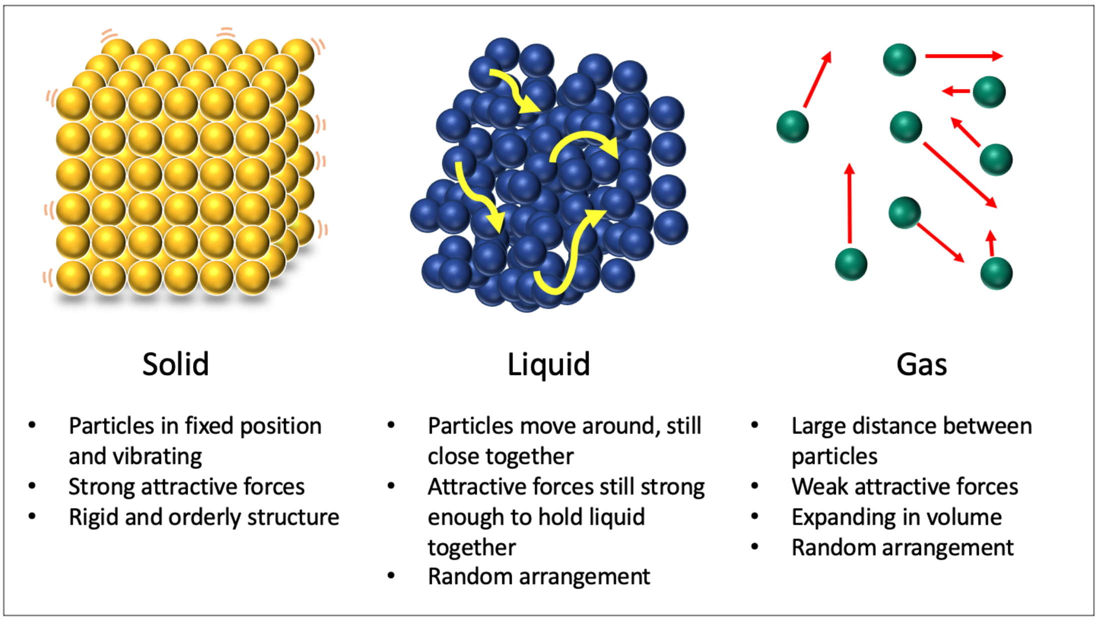


Figure 5.1: The different states of matter

5.2 Ideal Gases

As stated in section 5.1, the precise state of matter is contingent upon the balance between the kinetic energy of the particles, which is driven by temperature, and the intermolecular forces, which can be manipulated by increasing or reducing the pressure exerted upon the substance (see section 5.5). That is, introducing an external force on, let's say, a liquid, pushes the particles closer together, whereby the greater proximity translates into larger intermolecular forces. Moreover, keeping the temperature low implies slower-moving particles and thus a higher likelihood of particle interaction upon collision, which results in stronger intermolecular forces.

In the toy model of **ideal gases**, it is assumed that interactions between particles are *unaffected by intermolecular forces* and that the collisions occur in a *perfectly elastic fashion* (i.e., particles do not stick together when colliding). The law under this model that describes the behaviour of ideal gases takes into account three main variables, i.e., the pressure p, the volume V, and the temperature T, and is formulated as follows:

$$pV = nRT \tag{5.1}$$

with p the pressure that the gas carries out on the walls of the container in which it is stored (expressed in pascal (Pa)), V the volume (in cubic meters m^3), n the number of moles (in mol), $R = 8.314$ Pa·m^3·K^{-1}·mol^{-1} the gas constant, and T the temperature (in Kelvin K).

Let us examine one specific case whereby an ideal gas is contained in a closed system and both the temperature (T) and the amount of gas (n) are kept constant—this is known as **Boyle's Law**. The relationship between the pressure p and the volume V can then be formulated in the following way:

$$pV = \text{constant} \iff p_i V_i = p_f V_f \tag{5.2}$$

with p_i (p_f) the initial (final) pressure (expressed in Pa or N·m^{-2}) and V_i (V_f) the initial (final) volume (expressed in m^3 or liter L). Figure 5.2 graphically represents Boyle's Law for three different temperatures.

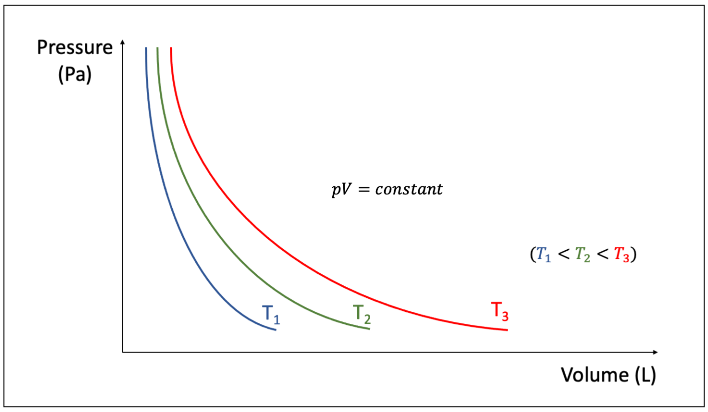


Figure 5.2: The relationship between pressure and volume for an ideal gas (Boyle's Law)

In other words, Equation 5.2 says that the pressure p of an ideal gas and the volume V of the container in which the gas is stored are inversely proportional to one another. For a given temperature of a certain gas, for instance T_2 in Fig. 5.2, if the volume V of the container is being reduced (by means of a moving piston), the particles are left with an increasingly lesser amount of space to move in, so that the particles hit the walls of the container more often, which results in a buildup of pressure. Keep in mind that the average velocity of the particles (i.e., the temperature) is maintained constant during this process.

Consider the example of a container of 5.2 liters (L) filled with helium gas that experiences a pressure of 4.32 atm (whereby atm is the standard atmospheric pressure and is equal to 101,325 Pa). If the pressure is upgraded to 6.12 atm, then we know that the container, according to Boyle's law, must decrease in size, if the temperature is kept unchanged. Equation 5.2 says that $p_i V_i = p_f V_f \Leftrightarrow 4.32 \text{ atm} \times 5.2 \text{ L} = 6.12 \text{ atm} \times V_f$ L, so that $V_f = 3.67$ L, which indeed refers to a smaller volume.

5.3 State Changes

A **state change or phase change** refers to the moment of transition between two states of matter. Table 5.1 sums up the six different state changes that can occur for the three states of matter that we discussed in section 5.1, i.e., solids, liquids, and gases.

Table 5.1: The six different state changes

State Change	Name	Example
Solid to liquid	Melting	Ice to water
Solid to gas	Sublimation	Dry ice at room temperature
Liquid to solid	Freezing	Water to ice
Liquid to gas	Vaporization	Water to steam
Gas to solid	Deposition	Water vapour to ice
Gas to liquid	Condensation	Water vapour to dew

The **melting point** of a material substance is defined as the temperature at which the substance transitions from its solid state into its liquid state at atmospheric pressure (i.e., the pressure of the Earth's atmosphere at at sea level). What happens at microscopical level is that the particles in the solid state gain additional kinetic energy as the temperature rises. When the average kinetic energy grows to the point where it is *able to overcome the attractive strength of the intermolecular forces*, the particles start leaving their locked positions, which signals the *onset* of the melting stage—bear in mind that the average kinetic energy *during* the transition from a solid to a liquid state does not change, as we shall see below.

Table 5.2 below provides the melting points for a handful of different materials. Generally, the melting point *rises with the strength of the bonds*. That is, networks of covalent bonds result in very high melting points (e.g., carbon), while simple covalent compounds (e.g., ammonia) give the lowest melting points, as they are held together by weak inter-molecular forces. Bear in mind that melting is about overcoming the intermolecular forces and does *not* imply that the chemical bonds *within* molecules are being broken.

Table 5.2: The melting points for different materials at atmospheric pressure

Substance	Melting Point (in °C)	Substance	Melting Point (in °C)
Aluminum	658	Lead	327
Ammonia	-77.7	Lithium	180.5
Bronze	910	Mercury	-38.8
Carbon	3,550	Nickel	1,455
Chloroform	-63.4	Nitrogen	-210
Copper	1,083	Palmitic acid	62.9
Glucose	146	Phosphorus	44.1
Gold	1,065	Polystyrene	240
Ice	0	Sand	1,550
Iron	1,538	Titanium	1,668

Although the melting point usually equals the freezing point, these two temperatures do not necessarily have to match. As a case in point, it is found that very pure water can reach temperatures well below 0°C and still remain liquid, even though ice melts at 0°C.

The **boiling point** of a substance refers to the temperature of a substance, measured at atmospheric pressure, whereby the average kinetic energy has reached the stage at which the particles become unimpeded by the intermolecular forces and detach from the liquid to roam freely throughout the space around them. The corresponding state change is known as vaporization, which marks the transformation from a liquid into a gaseous state; as in the case of melting or freezing, *during* this transition, the average kinetic energy does not further increase.

Again, the *stronger* the intermolecular forces, the *more elevated* the boiling point for a certain substance (bear in mind that other factors also have an influence, such as the shape of the molecule and its molecular weight). Since external pressure can influence the strength of the intermolecular forces, a liquid on top of the Kilimanjaro mountain, where the atmospheric pressure is *lower* compared to the sea level—it measures around 0.39 atm—boils at a *lower* temperature. If you would like to cook your food while enjoying the scenic view at an altitude of 5,896 m, you will in fact have to wait longer for your meal to be ready, given that the water boils at a lower temperature. Similarly, a pressure cooker operates at roughly 1.9 atm, which ensures that water boils at a higher temperature, reducing the time needed for food to cook, if that water were to be used for cooking. Table 5.3 supplies the boiling points of the substances mentioned in Table 5.2.

Table 5.3: The boiling points for different materials at atmospheric pressure

Substance	Boiling Point (in °C)	Substance	Boiling Point (in °C)
Aluminum	2,327	Lead	1,740
Ammonia	-33.3	Lithium	1,336
Bronze	2,300	Mercury	357
Carbon	4,827	Nickel	2,730
Chloroform	61.1	Nitrogen	-196
Copper	2,595	Palmitic acid	351
Glucose	527.1	Phosphorus	279.7
Gold	2,700	Polystyrene	430
Water	100	Sand	2,230
Iron	2,861	Titanium	3,287

In section 4.4, the concept of specific heat capacity is used to identify the amount of energy required for a substance to achieve a certain increase in temperature—or, alternatively, the amount of energy released during a cooling period. However, at the moment of a phase change, even though the temperature stays *at a constant value*, additional energy is required (or released) to complete the transition between different states. This energy is called *latent heat*, which is *not* kinetic energy (otherwise the temperature would rise) but a potential energy needed to *overcome the intermolecular forces of the bonds between the neighbouring particles*.

The latent heat related to melting and freezing is called the **latent heat of fusion**, whereas the **latent heat of vaporization** indicates the required or released phase energy during either vaporization or condensation. The *higher* the strength of the intermolecular forces, the *higher* the amount of energy required for the particles to break away either from their fixed locations within a solid or from the relatively loose positions within a liquid.

For any material, the latent heat of vaporization is always *larger* than the latent heat of fusion, since it requires much more energy to break *entirely* free from intermolecular forces (which is the case for the phase of vaporization) rather than *weakening* the intermolecular bond strength (which is the case for the phase of melting). Table 5.4 provides the latent heat of both fusion and vaporization for the substances mentioned in Table 5.3.

Table 5.4: The latent heat of fusion and vaporization for different materials (at 1 atm)

Substance	Heat of Fusion (in kJ·kg^{-1})	Heat of Vaporization (in kJ·kg^{-1})	Substance	Heat of Fusion (in kJ·kg^{-1})	Heat of Vaporization (in kJ·kg^{-1})
Aluminum	399.9	10,874	Lead	23.2	857.6
Ammonia	332.3	1,372	Lithium	432.2	21,023
Bronze	/	/	Mercury	11.4	295.3
Carbon	9,741	29,624	Nickel	297.6	6,310.8
Chloroform	73.7	263	Nitrogen	25.7	199.4
Copper	205.4	4,725.7	Palmitic acid	167.2	291.1
Glucose	110.6	652.3	Phosphorus	21.2	391.6
Gold	63.7	1,697.75	Polystyrene	132.1	348
Water	333	2,258	Sand	/	/
Iron	247.3	6,260.2	Titanium	322.8	8,795

Let us illustrate these concepts with the case of the chemical element Au (gold), which possesses a latent heat of fusion of 63.7 kJ·kg^{-1} (with kJ equal to kilojoule or 1,000 joule) at its melting point of 1,065°C (see table 5.2) and a latent heat of vaporization of 1,697.75 kJ·kg^{-1} at its boiling point of 2,700°C (see table 5.3). The energies required to bring 1 kg of solid gold at a temperature of 0°C to a state of vaporized gold above 2,700°C are graphically represented by Fig. 5.3.

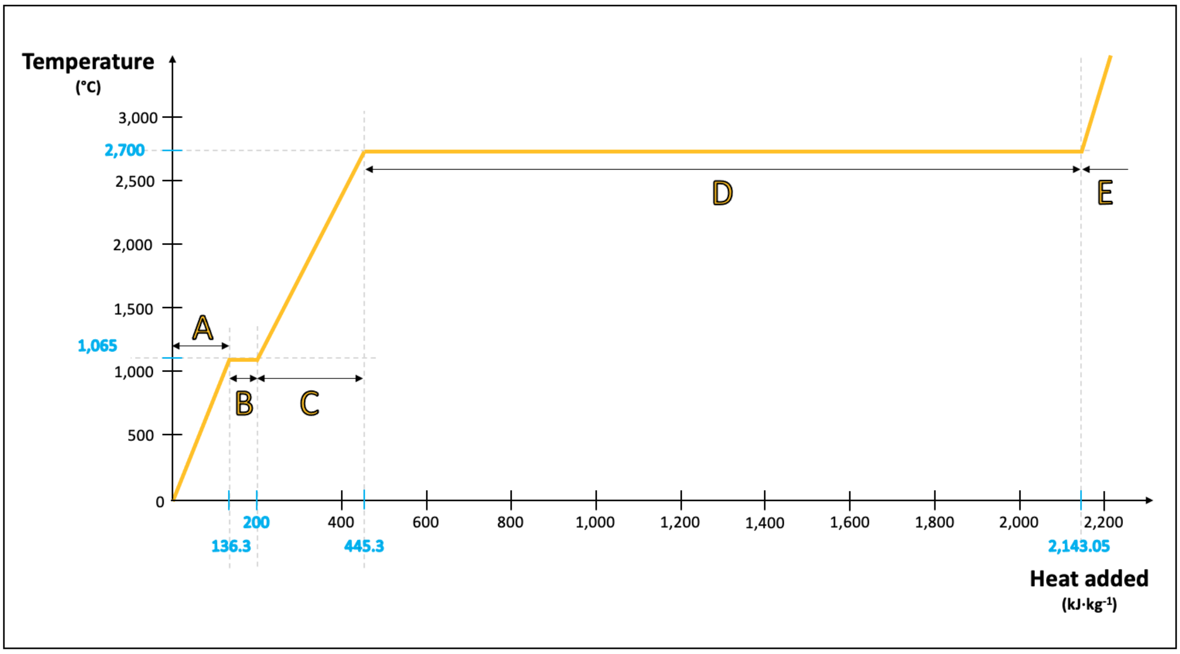


Figure 5.3: The latent heat of fusion and vaporization in the case of 1 kg of gold

Given a specific heat capacity of 128 J·kg^{-1}·°C^{-1} for solid gold, Equation 4.4 reveals that it requires $\Delta Q = 0.128 \times 1 \times (1,065 - 0) = 136.3$ kJ to raise the temperature from 0°C to 1,065°C—this corresponds to the region A in Fig. 5.3, where gold is in a solid state. At the melting point of 1,065°C, as a latent heat of fusion of 63.7 kJ is gradually being added to the process, gold finds itself in a mixed state between solid and liquid, i.e., region B. Upon completion, gold has entirely melted and can be further heated.

During region C, an additional energy of $\Delta Q = 0.150 \times 1 \times (2,700 - 1,065) = 245.3$ kJ (see Equation 4.4, with a specific heat capacity of 150 J·kg^{-1}·°C^{-1} for liquid gold) brings the gold to its boiling point of 2,700°C. Throughout region D, gold particles are incrementally detaching themselves from the liquid gold while 1,697.75 kJ of energy is progressively supplied to the process. At the beginning of stage E, the gold has completely vaporized, which is now ready to be further heated up.

The total amount of energy required to arrive at this point is equal to 2,143.05 kJ—as a point of reference, to convert 1 kg of ice of 0°C into steam (assuming a specific heat capacity of 4,217 J·kg^{-1}·°C^{-1} for water at 0°C) requires an energy of 3,012.7 kJ, which is higher due to the strong intermolecular forces in water.

5.4 Density

The **density** of a substance is defined as:

$$\rho = \frac{m}{V} \tag{5.3}$$

with ρ the density (expressed in kg·m^{-3} or often in g·mL^{-1}, with mL = milliliter = 10^{-3} L), m the mass of the substance (in kg or g), and V the volume (in m^3 or mL; keep in mind that 1 kg·m^{-3} is equal to 0.001 g·mL^{-1}). A commonly used density is that of water, whereby $\rho_{water} = 1,000$ kg·m^{-3} or $\rho_{water} = 1$ kg·L^{-1}, so that 1 kg of water corresponds to 1 L of water.

Both pressure and temperature can affect the density of a substance. While a *larger, uniform pressure* on the material tends to reduce its volume and therefore *increase* the substance's density, a *rise in temperature* gives a boost to the average kinetic energy of the particles, effectively expanding the volume of the substance, which leads to a *lower* density. A noteworthy exception is water, which reaches its maximum density at 4°C. This means that not only cooling down water from 4°C to 0°C but also reducing the temperature of ice causes a *decline* in density. Moreover, since ice has a more expanded molecular structure than water, the density of ice falls below that of water (below room temperature). Water beyond 4°C follows again the standard pattern, i.e., a higher (lower) temperature decreases (increases) its density.

As pointed out in section 5.1, solids and liquids are hard to compress and the distances between their particles remain rather minute, so that any variation in pressure and temperature has an overall minimal effect on the substance's density. In contrast, gases are much more susceptible to the impact of both temperature and pressure, due to the greater amount of space available between the particles. Using the ideal gas law (Equation 5.1) and the fact that the number of moles n equals the mass m of the substance divided by its molecular weight M ($n = \frac{m}{M}$), the density of a substance can be reformulated in this way:

$$\rho = \frac{pM}{RT} \tag{5.4}$$

with ρ the density (in kg·m^{-3}), p the pressure (in Pa), M the molecular weight (in kg·mol^{-1}), $R = 8.314$ Pa·m^3·K^{-1}·mol^{-1} the gas constant, and T the temperature (in Kelvin K). As expected, Equation 5.4 reveals that increasing the pressure of a gas augments the density, whereas rising the temperature diminishes its density.

Experimentally, the density of a solid is determined by establishing the mass of the substance with a scale and measuring the geometric features of the substance (which gives the volume). Alternatively, the object's density can be identified through *hydrostatic weighing*. This technique calls upon Archimedes' principle, which postulates that the buoyant force acting on an object submerged in water corresponds to the weight of the fluid displaced by the object. For instance, if an underwater scale indicates a mass of 3 kg, whereas the substance's mass is measured as 4 kg in a laboratory on land, then the amount of displaced fluid comes down to 1 kg. Given that 1 kg of water equals 1 L of water, the density can be calculated as $\rho = \frac{4}{1} = 4$ kg·L^{-1} (or kg·dm^{-3}, with dm = decimeter, given that 1 L = 1 dm^{-3}).

The density of liquids and gases can be measured by means of a hydrometer and a dasymeter, respectively, which both rely on Archimedes' principle. A *hydrometer*, which is marked with a scale of relative densities, is lowered into a container filled with a liquid. For a heavier (lighter) liquid, the hydrometer rises (sinks) and indicates a higher (lower) relative density on its scale. A *dasymeter* uses the difference in weight of a thin glass sphere when placed in vacuum and when submerged in the gas in order to determine the density of the gas.

5.5 Pressure

The **pressure** is defined as the force applied by a substance, i.e., a solid, a liquid, or a gas, on a certain surface area, whereby the direction of the force is perpendicular to the surface:

$$p = \frac{F}{A} \tag{5.5}$$

with p the pressure (expressed in Pa), F the magnitude of the exerted force (in N), and A the surface area (in m^2). Other units of pressure include the *atmospheric pressure* (1 atm = $101,325$ Pa), the *torr* (1 torr = $\frac{1}{760}$ atm = 133.32 Pa), the *bar* (1 bar = $100,000$ Pa), the *psi* (the pound-force per square inch, whereby 1 psi = $6,895$ Pa; note that 1 pound (lb) = 0.4536 kg and 1 inch = 0.0254 m), and the *millimeter of mercury* (1 mmHg = 133.32 Pa, which is basically identical to the torr; note that Hg is the notation of the chemical element mercury).

In the context of fluids, i.e., liquids and gases, the concept of **hydrostatic pressure** refers to the pressure that a point within a fluid (in equilibrium) experiences due to the weight of the fluid above that point. In other words, the force of gravity is responsible for the hydrostatic pressure. With the assistance of Equations 3.15, 5.3, and 5.5, and given that the volume V of an object is defined as the area A times the height h, the mathematical expression for hydrostatic pressure can be derived as follows:

$$\begin{aligned}
&\Longrightarrow p = \frac{F}{A} \\
&\Longleftrightarrow p = \frac{mg}{A} \\
&\Longleftrightarrow p = \frac{\rho V g}{A} \\
&\Longleftrightarrow p = \rho g h
\end{aligned} \tag{5.6}$$

with p the hydrostatic pressure (expressed in Pa), ρ the density of the fluid (in kg·m^{-3}), g the gravitational field strength ($g = 9.81$ m·s^{-2}), and h the vertical distance of the point in the fluid to the surface. Bear in mind that the *total* pressure for a point within a fluid is the hydrostatic pressure *plus* the atmospheric pressure (1 atm) on the surface of the fluid. Fig. 5.4 provides a visual representation together with three case studies.

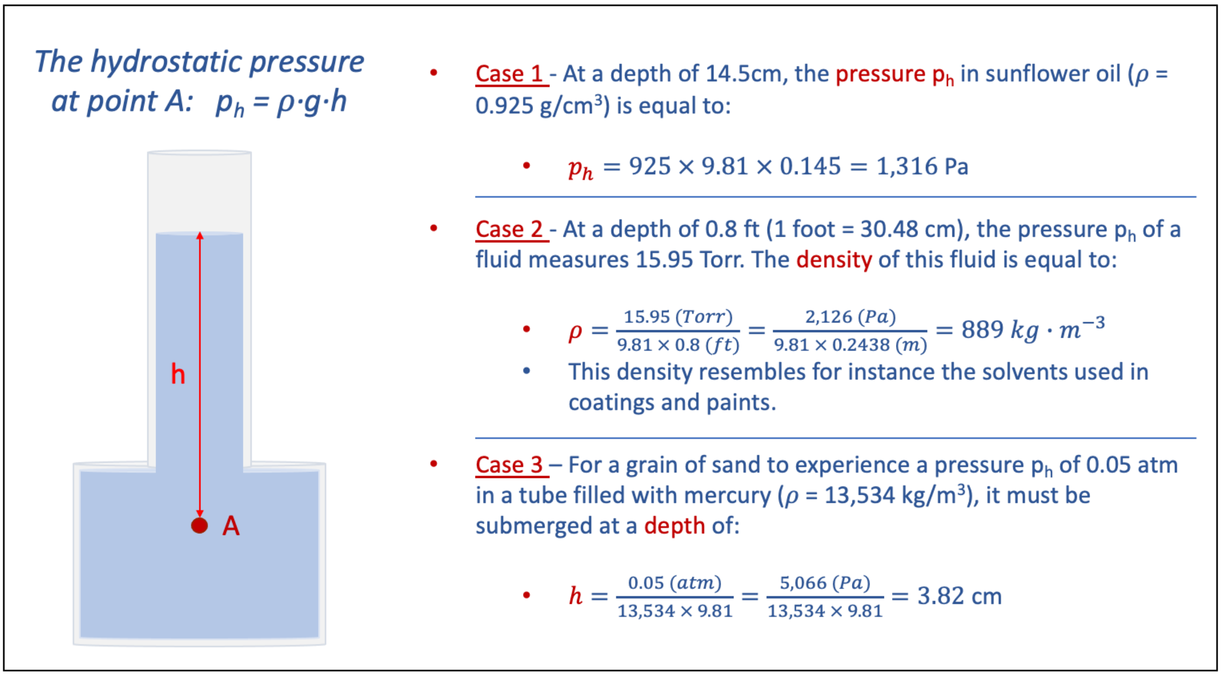


Figure 5.4: The hydrostatic pressure and three case studies

6 Waves

6.1 Wave Properties

As already briefly touched upon in section 4.3 on thermal radiation, **electromagnetic waves** consist of oscillating electric and magnetic fields, which are basically the fundamental constituents of what *light* is made of. Broadly speaking, electromagnetic waves can be subdivided into seven different categories: radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays. What is more, they carry energy even when propagating through empty space, e.g., beyond the Earth's atmosphere in outer space (more on electromagnetic waves in section 6.5).

Apart from electromagnetic waves, there is another type of wave, called **mechanical waves**, which differ from electromagnetic waves in at least one way: mechanical waves can only travel through space with the assistance of a *physical medium*, i.e., any solid, liquid, or gas—electromagnetic waves do not need such a medium, as they can travel in vacuum where there is no medium present. The transfer of energy is then enabled by particles in the medium that *oscillate around a fixed position*, i.e., they do not move along with the wave. Examples of mechanical waves include sound waves (see section 6.4), water waves, seismic waves, waves associated with springs, waves produced by vibrating strings between two fixed ends, and waves generated by the oscillatory motion of the surface of solids.

As a matter of fact, there are two other types of waves, i.e., matter waves and gravitational waves. Matter waves or de Broglie waves are waves related to matter particles themselves, e.g., an electron, whereas gravitational waves are ripples of spacetime itself, which are the result of high-energy events in space (e.g., the collision of two black holes). These two types of waves are not further explored here.

Generally, a distinction is made between transverse and longitudinal waves. While **transverse waves (a.k.a. shear waves**) are characterized by a medium that oscillates in a direction *perpendicular* to the wave's direction of motion, in the case of **longitudinal waves (a.k.a. pressure waves)** the direction of displacement of the medium is *parallel* to the direction in which the wave travels. Basically, it is the oscillations, which make up the wave, that travel through space. One key characteristic that is shared among all types of waves is that they transport energy *without* taking any matter particles with them.

Examples of transverse waves are electromagnetic waves, waves that result from moving one end of a rope up and down while the other end is attached to a wall, and waves produced by the skin of a tambourine when tapping it, whereas sound waves traveling through the air, spring waves, and certain seismic waves called P-waves are brought under the group of longitudinal waves. What is more, waves can adopt a mixed form of shear and pressure waves, such as water waves, sound waves that move through solids, and general seismic waves.

Regarding transverse waves, if we consider a flat line as a wave at rest, then the maximum vertical displacement from this equilibrium position is called the **amplitude** $\boldsymbol{A_w}$, whereby the **peak** indicates the highest point of the amplitude and the **trough** the lowest. In longitudinal waves, a peak (trough) is referred to as a **compression (rarefaction)**, which signals the region where the spacing between the particles of the medium is minimal (maximal). The amplitude is related to the degree by which the density of the particles within the medium during compressions (or rarefactions) deviate from the density at equilibrium position; a greater relative difference in particle density refers to a higher amplitude.

The distance between two peaks or troughs (or two rarefactions or compressions in the case of longitudinal waves) is referred to as the **wavelength $\boldsymbol{\lambda}$**. The number of oscillations carried out by the wave in the timespan of one second is defined as the **frequency f**, whereas the **period T** is equal to the time it takes the wave to traverse the distance of one wavelength λ. Put differently, the period T is the duration to complete one wave cycle. This implies that the frequency f is the inverse of the period T, and vice versa.

Moreover, we can also formulate a **wave speed $\mathbf{v_w}$**, which marks the speed with which the wave propagates through space and is entirely determined by a certain period T and wavelength λ. Keep in mind that the wave speed v_w is *independent* of the amplitude A_w and that the **energy of the wave** is directly proportional to the square of the amplitude A_w. In mathematical form, the above definitions become:

- $$f = \frac{1}{T} \iff T = \frac{1}{f} \tag{6.1}$$
- $$v_w = \frac{\lambda}{T} = \lambda f \tag{6.2}$$
- $$E_{wave} \propto A_w^2 \tag{6.3}$$

with f the frequency (expressed in Hertz (Hz) or s^{-1}), T the period (in s), λ the wavelength (in m), v_w the wave speed (in m·s^{-1}), A_w the amplitude (its units are equal to those of the oscillating variable, e.g., displacement), and E_{wave} the energy of the wave (in joule J).

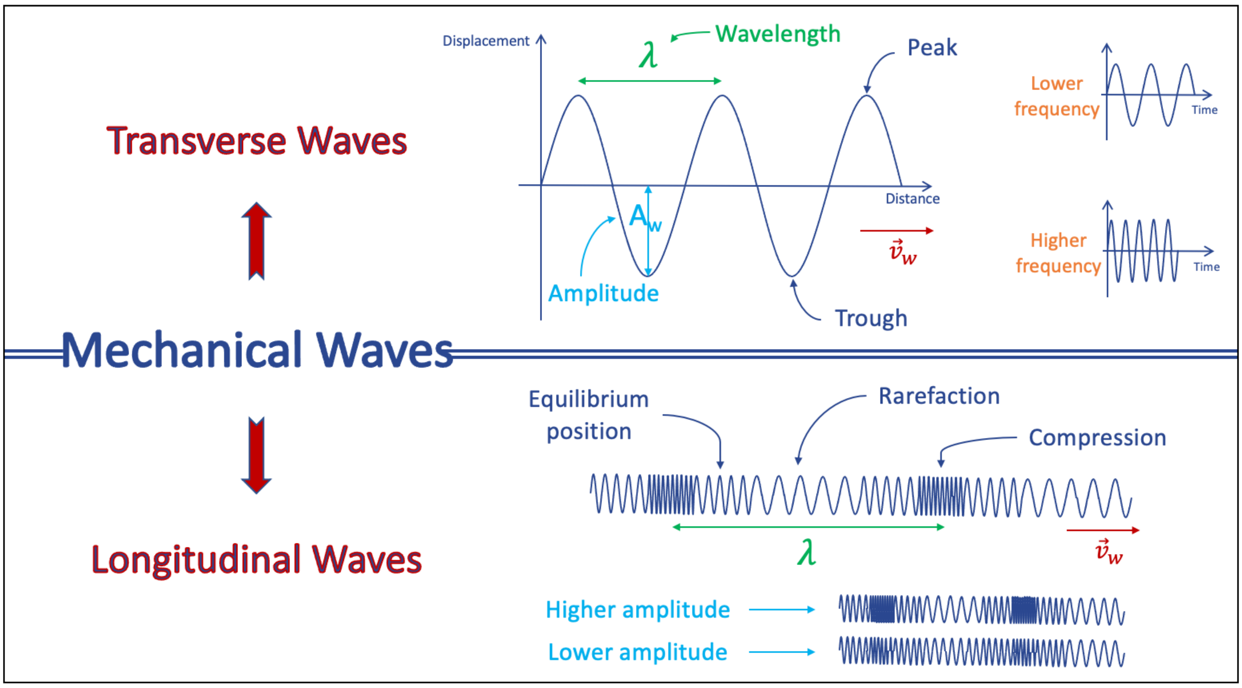


Figure 6.1: Transverse and longitudinal mechanical waves

6.2 Wave Behaviour

When waves travel through space, they can encounter a boundary, which may either take the form of an impenetrable obstacle or imply a change of (the density of) the physical medium. At the boundary, at least some part of the wave bounces back off it, which is a behaviour called **reflection**. That is, the **incident wave**, which is the name given to the wave before it meets the boundary, splits up in a **reflected wave** and a **transmitted wave**, i.e., the part that is allowed to pass through the boundary.

The angle at which the incident wave hits the boundary equals the angle under which the reflected wave travels away from it (under the assumption of a boundary with a smooth surface). The former angle is referred to as the **angle of incidence**, which is defined as the angle between the incident wave's direction of motion and the normal of the boundary—the normal is the vector that is always oriented perpendicular to a surface or boundary—while the latter angle is called the **angle of reflection**.

In the event of an entirely impermeable boundary or if the wave reaches the end of the medium, the incident wave is *completely* reflected. Consider an example of the second scenario whereby a transverse incident wave travels with an *upward* displacement along a rope, which is attached to a wall—the rope is the medium and ends at the wall where it is *fixed* (see upper part of Fig. 6.2)—and whereby the rope makes an angle of 90° with the wall (that is, the angle of incidence is 0°, given that the wave's direction of motion is parallel to the normal of the wall). Because of the downward restoring forces of the propagating wave, the end of the rope experiences an upward force, to which the wall has to respond by exerting a downward force on the rope (due to Newton's third law; see Equation 3.12). This means that the *reflected* wave travels with a *downward* displacement along the rope at an angle of reflection of 0°—in other words, the wave has effectively reversed its direction of displacement.

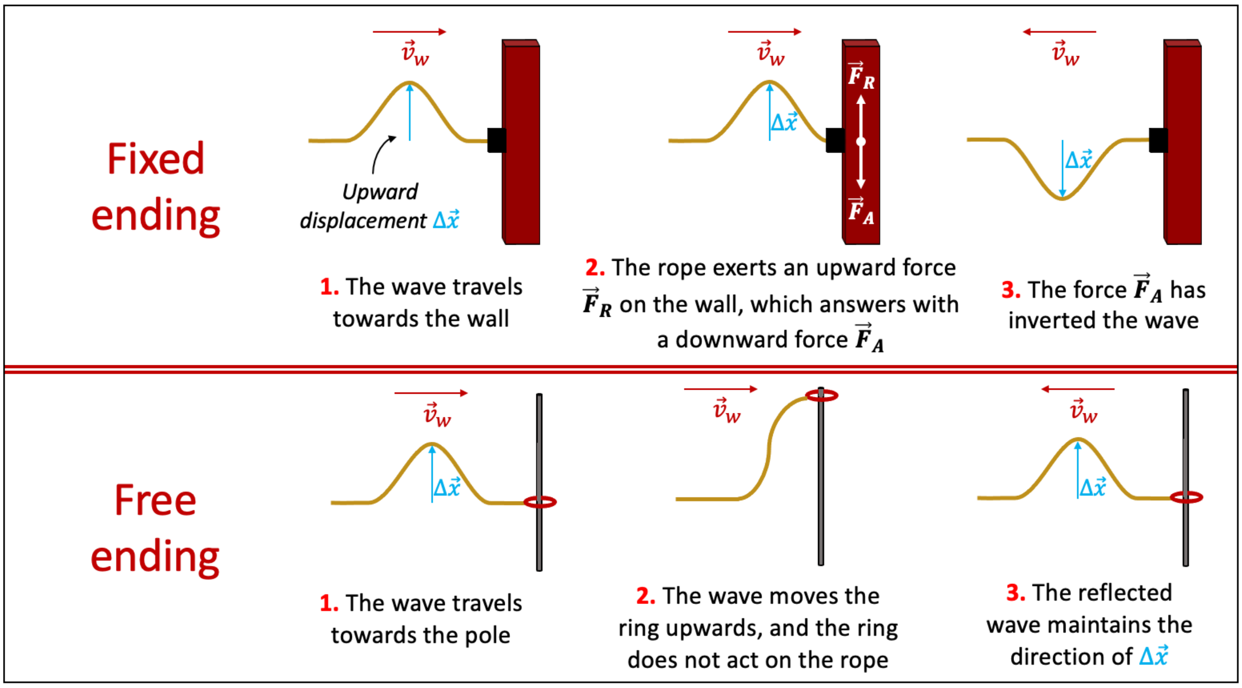


Figure 6.2: Transverse mechanical waves propagating along a stretched rope

In contrast, if the rope is attached to a *moveable (free) ending*, e.g., a massless ring which can slide up and down a pole, the end of the rope moves upwards with the ring (see lower part of Fig. 6.2). This time, there is no downward force acting on the rope, so that the wave is reflected at the *same* side (the upper side) of the rope; that is, the direction of displacement is still upwards.

In both cases (i.e., an attached and a free ending), the wave speed v_w as well as the amplitude A_w of the reflected wave remain unchanged *due to the conservation of momentum and energy.*

Regardless of the type of mechanical wave, the wave speed v_w depends on the nature of the physical medium, and the *more dense* the medium, the *lower* the wave speed v_w. On top of that, in the case of transverse waves, a stronger tensional (elastic) force applied to the medium enhances the wave speed v_w, whereas either a more pronounced resistance to compression by fluids or a greater stiffness in solid materials boosts the wave speed v_w for longitudinal waves. Translated into mathematical equations, this gives:

- *Transverse waves*: $$v_w = \sqrt{\frac{F_T}{\mu}} \tag{6.4}$$

- *Longitudinal waves*: $$v_{w,solids} = \sqrt{\frac{E}{\rho}} \tag{6.5}$$

$$v_{w,fluids} = \sqrt{\frac{B}{\rho}} \tag{6.6}$$

with v_w the wave speed (expressed in m·s^{-1}), F_T the tensional force (in N), $\mu = \frac{m}{\delta x}$ the linear density (in kg·m^{-1}), E the elastic modulus (the stiffness of a material; expressed in Pa), B the bulk modulus (resistance to compressibility; expressed in Pa), and ρ the density (in kg·m^{-3}).

Take the example of a thick rope that is attached to another rope of a lower density, i.e., a thinner rope. Consider first the scenario whereby a wave propagates *from the higher-density to the lower-density region* (see left side of Fig. 6.3). When an incident wave reaches the boundary, which is represented by this change in density, it gives rise to a reflected wave and a transmitted wave, whereby the reflected wave is not inverted, because it regards the end of its medium as a free ending. Furthermore, as the reflected wave remains on the higher-density side, it retains the velocity of the incident wave. Since the medium particles at both sides of the boundary are connected with one another, the frequency of both the reflected and the transmitted wave are identical. It follows then from the definition of the wave speed (see Equation 6.2) that the wavelength of the reflected wave does not change. However, given that part of the energy of the incident wave is conveyed to the transmitted wave, the amplitude of the reflected wave is lower than that of the incident wave.

As the transmitted wave travels in a medium of lower (linear) density, its wave speed goes up, resulting in a larger wavelength due to a constant frequency. Moreover, the transmitted wave always retains the direction of displacement of the incident wave. Keep also in mind that it propagates with a lower amplitude, as the energy of the incident wave is shared between the reflected and transmitted wave.

In the other scenario whereby a wave moves *from the lower-density region into the higher-density region* (see right side of Fig. 6.3), the reflected wave is inverted, as it treats the end of its medium as a fixed ending. Both the reflected and the transmitted wave exhibit a lower amplitude, with the reflected (transmitted) wave propagating at equal (lower) speed with respect to the incident wave. The wavelength of the *transmitted* wave is thus reduced.

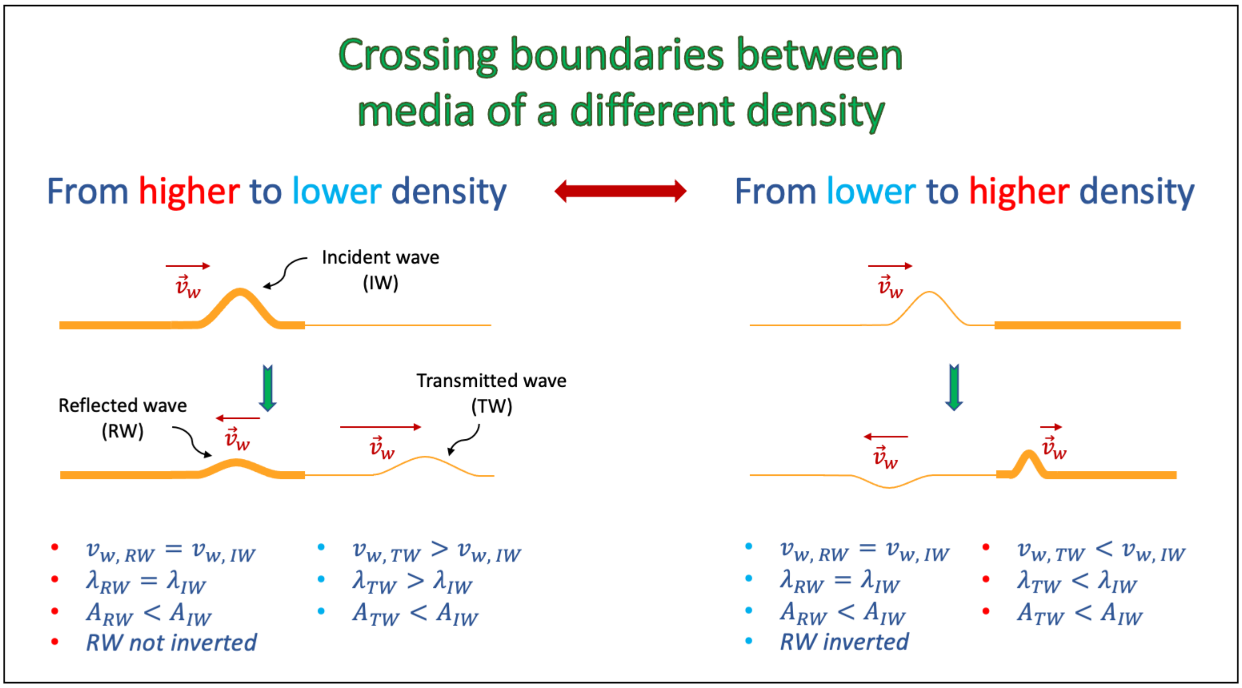


Figure 6.3: Transverse mechanical waves crossing boundaries

In the event that the incident wave reaches a boundary at an angle of incidence *different* from 0°, then the direction of traveling of the *transmitted wave* is slightly altered relative to that of the incident wave. This kind of wave behaviour is known as **refraction**, and the angle between the normal of the boundary and the direction of motion of the transmitted wave is called the **angle of refraction**.

If the density of the medium through which the transmitted wave is moving is *higher (lower)* compared to the situation prior to the boundary crossing, then the angle of refraction is *smaller (larger)*, the wave speed *lowered (boosted)*, and the wavelength *reduced (increased)*. In other words, if a mechanical wave changes its speed, it alters its direction. Similar to the case where the angle of incidence equals 0°, the frequency of the wave does not vary in the scenario of refraction, and the amplitude of the transmitted wave is also diminished.

Regarding **water waves**, which are a mixture of transverse and longitudinal waves (as indicated in section 6.1), their behaviour differs with respect to the distance from the origin of disturbance. That is, close to the source, the waves spread out in a circular (two-dimensional) or spherical (three-dimensional) fashion, whereas at larger distances they mainly move along a straight line (one-dimensional). The peaks of the propagating wave are called *wave fronts*, and the wavelength λ is equal to the distance between two consecutive wave fronts. So, in close proximity to the source, the wave fronts are curved, while farther away they approximate straight lines, which are then appropriately named *plane waves*.

The direction of motion of wave fronts or plane waves is indicated by a straight line, called a *ray*, that runs perpendicular to the wave front or plane wave. Fig. 6.4 depicts the behaviour of reflection and refraction for water waves. In this respect, the ray's behaviour is analogous to that of a ray of light, i.e., electromagnetic waves, which is further explored in section 6.3.

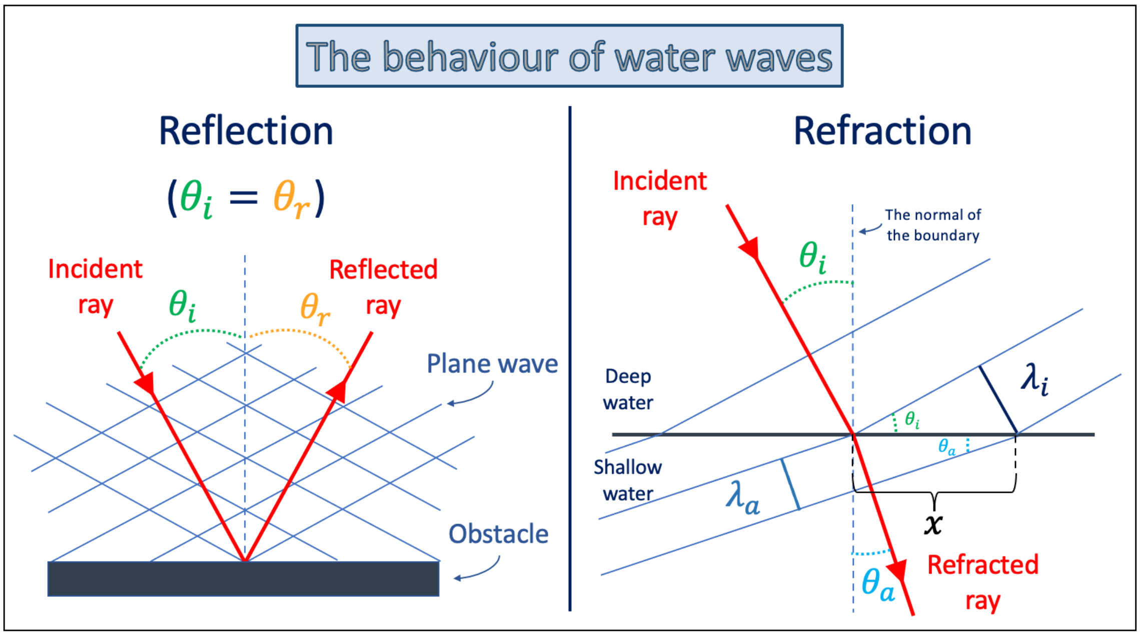


Figure 6.4: Reflective and refractive behaviour of water waves

When incident water waves transition from deep waters into shallows waters, they lose speed largely due to frictional forces exerted by the seabed (and not a difference in density). In other words, they refract. As a result, transmitted shallow-water waves exhibit a reduced wavelength λ and a smaller angle of refraction. With the assistance of trigonometry and Equation 3.2, we can deduce the **law of refraction** from Fig. 6.4:

$$\begin{aligned}
&\Longrightarrow \sin\theta_i = \frac{\lambda_i}{x} = \frac{v_{w,i}T}{x} \qquad \textit{and} \qquad \sin\theta_a = \frac{\lambda_a}{x} = \frac{v_{w,a}T}{x} \\
&\Longleftrightarrow \frac{v_{w,i}T}{\sin\theta_i} = x = \frac{v_{w,a}T}{\sin\theta_a} \\
&\Longleftrightarrow \frac{\sin\theta_a}{\sin\theta_i} = \frac{v_{w,a}}{v_{w,i}} = \frac{\lambda_a}{\lambda_i}
\end{aligned} \tag{6.7}$$

with θ_i (θ_a) the angle of incidence (refraction), λ_i (λ_a) the wavelength of the incident (transmitted) wave, x the shared hypotenuse of the two adjacent triangles, T the period of the incident and the transmitted wave (the period does not change, because the frequency does not change at boundary crossings; they are related by Equation 6.1), and $v_{w,i}$ ($v_{w,a}$) the wave speed of the incident (transmitted) wave.

A final behaviour of waves that is explored in this section is what is known as the **Doppler effect**, which is defined as the *observed* change in frequency of a wave as the result of the relative motion between the emitting source of the wave and the measuring device, i.e., the observer. The Doppler effect is a feature of both mechanical and electromagnetic waves.

If the emitting source moves, let's say, towards a stationary observer (or, alternatively, the observer approaches a stationary emitting source), the Doppler effect implies that consecutive peaks (or compressions) of the wave reach the observer at a greater rate (i.e., a higher frequency). In the case of sound waves, this translates into the observer experiencing the wave fronts as more squeezed together. Conversely, moving away from the emitting source means that this rate goes down (i.e., a lower frequency) and that the wave fronts of the sound waves are experienced as being farther apart from one another.

The Doppler effect of sound waves is experienced, for example, when listening to the siren of an ambulance while it rushes past us—as the ambulance approaches we hear an increasingly higher pitch, whereas we observe a gradually declining pitch when it recedes from us. In the context of electromagnetic waves, if we measure, for instance, the visible light of a nearby star that is moving at high velocity towards Earth, the observed frequency is higher (or, similarly, the wavelength is shortened) relative to the proper frequency of the light—put differently, the electromagnetic wave is observed to have been shifted towards the blue spectrum (see section 6.5).

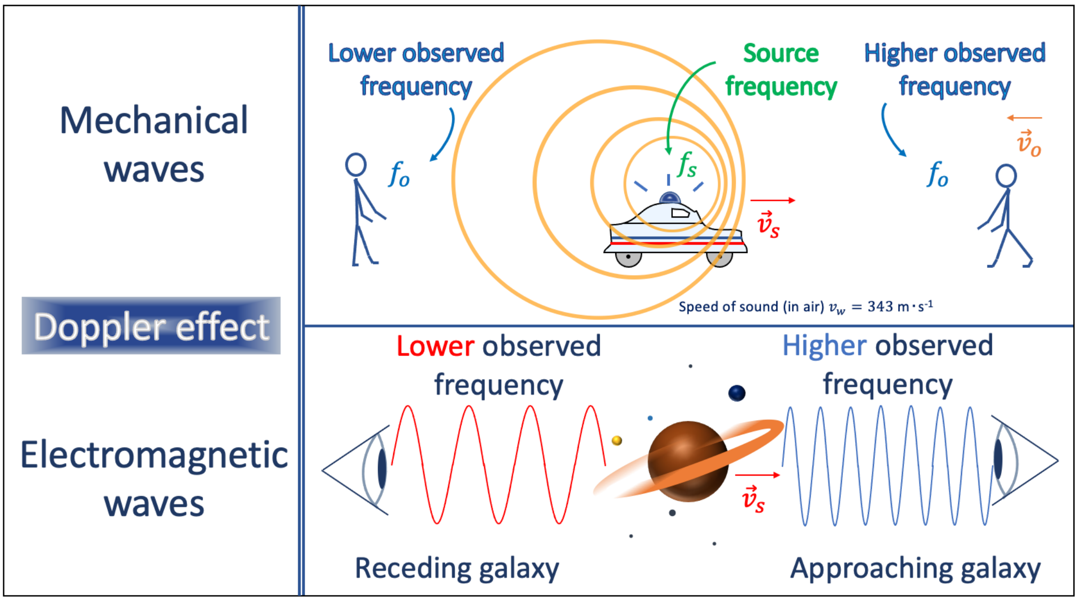


Figure 6.5: A schematic view of the Doppler effect

The mathematical formula for the Doppler effect (for mechanical waves) is as follows:

$$f_o = \frac{(v_w \pm v_o)}{(v_w \pm v_s)} f_s \qquad (6.8)$$

with f_o (in Hz) the frequency as experienced by the observer, f_s (in Hz) the frequency of the emitting source, v_w (in m·s^{-1}) the wave speed, and v_o (v_s) the speed (in m·s^{-1}) at which the observer (the emitting source) moves. The choice for the plus or minus sign depends on the direction in which the observer and/or the source is moving. Bear in mind that the frequency increases (decreases) when the observer approaches (moves away from) the source or when the source moves closer to (farther from) the observer.

Consider the following two numerical examples. While blowing its whistle with a frequency of $f_s = 380$ Hz, a train moves at $v_s = 20$ m·s^{-1} towards an individual sitting on a bench (i.e., $v_o = 0$ m·s^{-1}) next to the railway tracks. Given a speed of sound of $v_w = 343$ m·s^{-1}, the frequency observed by the individual equals $f_o = \frac{(343 \pm 0)}{(343-20)} \times 380 = 404$ Hz. While you are riding a bike at $v_o = 10$ m·s^{-1}, you measure the frequency of the siren ($f_o = 450$ Hz) of a police car that is driving away from you. Given that you know that the actual frequency measures $f_s = 495$ Hz and assuming a speed of sound of $v_w = 343$ m·s^{-1}, you come to the conclusion that the police car is driving at a speed of $v_s = f_s \cdot \frac{(v_w + v_o)}{f_o} - v_w = 495 \cdot \frac{(343+10)}{450} - 343 = 45$ m·s^{-1}.

6.3 Optics

Rays of **light** consist of electromagnetic waves, which travel at a constant speed, i.e., the speed of light c, which is equal to 299,792,458 m·s^{-1} (in vacuum), or often rounded off to 3×10^8 m·s^{-1}. They follow the behaviour of waves as explained in section 6.2: light is reflected, transmitted, and refracted. Similar to mechanical waves, when light travels across different media, its speed varies according to the nature of the material it passes through. The reason for this is that the electrons of the atoms within the medium are able to absorb and emit light (see section 4.3 on thermal radiation), which slightly slows down the speed of light. This fact about electromagnetic waves is represented by the **index of refraction n**, which is defined as:

$$n = \frac{c}{v_w} \tag{6.9}$$

with n the index of refraction (dimensionless), $c = 299,792,458$ m·s^{-1} as measured in empty space, and v_w the wave speed of light in the respective medium. The *larger* the index, the *greater* the reduction of the speed of light. Keep in mind that the index can never fall below the value of 1. Table 6.1 provides the refraction index for a couple of different media.

Table 6.1: The index of refraction for different materials

Material	n	Type of Material	Material	n	Type of Material
Air (at 15°C)	1.00028	gas	Benzene	1.50	liquid
Ammonia	1.00038	gas	Carbon disulfide	1.63	liquid
Methane	1.00044	gas	Ice	1.31	solid
Chlorine	1.00077	gas	Polystyrene	1.6	solid
Water (at 20°C)	1.33	liquid	Flint glass	1.75	solid
Olive oil	1.46	liquid	Diamond	2.42	solid

The fact that there is much more distance between the molecules and atoms in gases compared to liquids and solids (see section 5.1) is clearly demonstrated by Table 6.1. Given that light interacts on average less frequently with the atoms in gases, the index is overall lower for gases, which results in a higher speed when light propagates through these gaseous media.

A relationship is experimentally established between the angle of incidence, the angle of refraction, and the index of refraction, which is known as **Snell's law**:

$$n_i \sin\theta_i = n_a \sin\theta_a \tag{6.10}$$

with θ_i (θ_a) the angle of incidence (refraction), and n_i (n_a) the index of refraction of the medium through which the incident (refracted) light ray is traveling. By using the definition of the refraction index (Equation 6.9), Snell's law can be connected to the law of refraction (see Equation 6.7 of section 6.2) in the following way:

$$\frac{n_i}{n_a} = \frac{\sin\theta_a}{\sin\theta_i} = \frac{v_{w,a}}{v_{w,i}} = \frac{\lambda_a}{\lambda_i} \tag{6.11}$$

Equation 6.11 reveals that a light ray traveling from a region with a *lower* refraction index to a region with a *higher* index results in a *smaller* angle of refraction relative to the angle of incidence, whereby the light ray *loses* a fraction of its initial speed and propagates with a *reduced* wavelength.

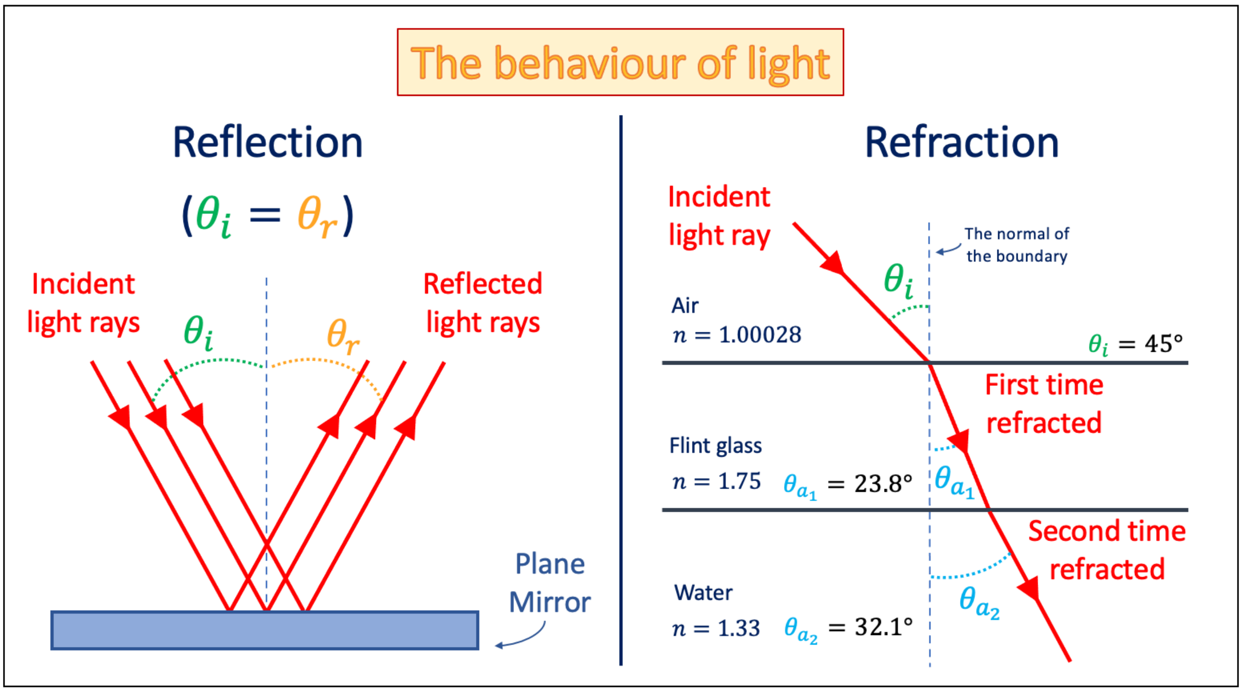


Figure 6.6: Reflective and refractive behaviour of light

6.4 Sound Waves

As pointed out in section 6.1 on the characteristics of waves, both transverse and longitudinal mechanical waves require a medium in order to travel, whereas electromagnetic waves propagate without a medium and are thus able to travel in empty space (i.e., vacuum). Longitudinal waves are known as pressure waves, since they propagate by way of periodically compressing and decompressing the particles within the medium. In order to set in motion a pressure wave, a source must be present that produces vibrations (e.g., a person who sings). These tiny distortions of the medium then travel longitudinally until they reach an observer or until they fade away.

Sound waves are a class of longitudinal waves and can even behave as transverse waves in solids. As they qualify as waves, they **reflect** when meeting boundaries, which explains the origin of **echoes**, whereby the angle of incidence (θ_i) equals the angle of reflection (θ_r)—this applies generally to both mechanical and electromagnetic waves.

Sound waves furthermore **refract** when transitioning into a medium of *different density*. For instance, when entering a more dense region as they propagate through the air, sound waves slow down (as per Equation 6.6), which results in a smaller angle of refraction (as per Equation 6.7). However, their refractive behaviour is more gradual and also varies with the temperature of the medium (although this temperature dependence is mainly relevant for gases). That is, *sound waves move faster at higher temperatures*, since the increased vibrations, caused by a higher average kinetic energy of the particles, help to facilitate the longitudinal motion of sound.

For example, when you are having a conversation in the garden late at night, the layer of air close to the ground surface is colder relative to the warmer air higher up, which slows down the bottom part of the sound wave, and thus refracts the sound downwards towards the ground—this is why you can better overhear conversations at longer distances at night than during daytime, given that by day the ground is warmer than the air, which causes the sound to gradually refract upwards.

As a side note, this temperature dependence of the **speed of sound waves** can be seen mathematically when incorporating the ideal gas law (Equation 5.4) into the definition for the speed of longitudinal waves in fluids (Equation 6.6), after replacing the bulk modulus B (which is defined as $B = -V\frac{\Delta p}{\Delta V}$) with the pressure p under the assumption of Boyle's law (Equation 5.2). In other words:

$$\begin{cases} pV = \text{constant} \iff \Delta p \cdot V + p \cdot \Delta V = 0 \iff p = -V\dfrac{\Delta p}{\Delta V} = B \\ \\ v_{w,fluids} = \sqrt{\dfrac{B}{\rho}} = \sqrt{\dfrac{p}{\rho}} = \sqrt{\dfrac{RT}{M}} \end{cases} \tag{6.12}$$

We can also approach this result from the perspective of hydrostatic pressure in fluids (see section 5.5). That is, the speed of sound *decreases with altitude*. The reason is that, as we rise to higher altitudes, the atmospheric pressure p drops, since we decrease the distance h between the point where we find ourselves and the surface of the fluid, i.e., the boundary between the Earth's atmosphere and space, and Equation 6.12 subsequently reveals that a lower pressure translates into a reduced speed of sound. Put differently, as it is colder higher up in the atmosphere (the ideal gas law states that for a constant volume a lower pressure leads to a lower temperature), Equation 6.12 predicts a lower wave speed of sound. With the help of the definition of the hydrostatic pressure in fluids (Equation 5.6), the speed of sound waves becomes:

$$v_{w,fluids} = \sqrt{\frac{p}{\rho}} = \sqrt{\frac{\rho g h}{\rho}} = \sqrt{gh} \tag{6.13}$$

Apart from the density of the medium and temperature, there is another aspect that impacts the speed of sound waves: the *elasticity* of materials. As solids are better able to hold together their form under pressure with respect to liquids and gases (i.e., the elasticity modulus E is typically larger than the bulk modulus B; see Equations 6.5 and 6.6) and given that the particles in solids are much closer to one another, solids assist the flow of sound *to a much greater extent* than liquids, which in turn do a far better job than gases, as is evident from Table 6.2 in which the wave speed of sound for a number of different materials is listed.

Table 6.2: The speed of sound for different materials (at 1 atm and 25°C)

Material	Speed of sound (in $m{\cdot}s^{-1}$)	Type of Material	Material	Speed of sound (in $m{\cdot}s^{-1}$)	Type of Material
Sulfur Dioxide (20°C)	201	gas	Sea water	1,533	liquid
Nitrous oxide	268	gas	Glycerol	1,904	liquid
Air (20°C)	343	gas	Gold	3,240	solid
Methane (20°C)	446	gas	Aluminum	5,100	solid
Ethyl alcohol	1,207	liquid	Pyrex glass	5,640	solid
Mercury	1,450	liquid	Diamond	12,000	solid

Two other characteristics of sound waves are the pitch and loudness. The *frequency* of a sound wave determines the **pitch** of a sound, which refers to how high or low a sound is. For instance, the lowest key on a standard piano rings at 27.5 Hz (the note C), while the highest note has a frequency of 4,186 Hz (the note A). The *audible range of hearing* for human beings goes approximately from 20 Hz to 20,000 Hz, while that of animals is much more extended than ours. For example, the upper frequency boundary for dogs, cats, bats, and dolphins can go as high as 45,000 Hz, 85,000 Hz, 120,000 Hz, and 200,000 Hz, respectively, whereas blue whales, ferrets, and elephants are able to reach minimum levels of 14 Hz, 16 Hz and 17 Hz, respectively.

Ultrasound (Infrasound) is a sound wave that is ultrasonic (infrasonic), i.e., its frequency exceeds (falls below) the threshold of 20,000 Hz (20 Hz). The *pulse-echo technique* in the field of medical imaging is one example that relies on ultrasound, whereby sound pulses of frequencies in the order of 10^6 Hz are directed towards the body. What is measured are the reflections off the boundaries between internal organs and structures, and the results are visually represented in an echography, such as the ultrasound image of a fetus.

The other feature of sound waves, i.e., **loudness**, is related to the amount of energy that the wave is carrying, which in turn is proportional to the *amplitude squared* (see Equation 6.3). In fact, loudness is represented by the physical quantity of *intensity* I, which is defined as the amount of energy affiliated with the wave that flows through a surface of a given area over a certain timespan. In other words, the intensity of a sound wave explains how much power is being transferred through a certain surface area. Based on Equation 3.25 of section 3.7, the intensity of sound can be written as follows:

$$I = \frac{P}{A} = \frac{\Delta E}{(\Delta t) \cdot A} \propto A_w^2 \tag{6.14}$$

with I the intensity (expressed in W·m^{-2}), P the power (in Watt W), A the surface area (in m^2), t the time (in s), and A_w the amplitude (in m). Note that although the Equation 3.25 defines power as the amount of *work* per unit time, given that work equals the energy transferred in a certain time frame, the power can be formulated as $P = \frac{\Delta E}{\Delta t}$, as is done in Equation 6.14.

The intensity of sound is then used to establish the *sound level* β (expressed in decibel (dB)), which is defined as:

$$\beta = 10 \cdot log\left(\frac{I}{I_0}\right) \tag{6.15}$$

with β the sound level (expressed in dB), I the intensity (in W·m^{-2}), $I_0 = 1.0 \times 10^{-12}$ W·m^{-2} the minimum intensity to which a human ear is responsive, and log the logarithm to the base of 10 ($\log_{10}$). Equation 6.15 shows that the operation of doubling (halving) the intensity of sound I leads approximately to a 3 dB increase (decline) in sound level β, whereas the sound level β rises (decreases) with exact steps of 10 dB for every expansion (reduction) of the intensity I by a factor of 10—remember that $log_a(x) = y \Leftrightarrow a^y = x$.

For example, doubling a sound energy of $I = 4.5 \times 10^{-8}$ W·m^{-2} three times results in an increase of sound level of $\beta_2 - \beta_1 = 55.6 - 46.5 = 9.1$ dB, whereby $\beta_1 = 10 \cdot log(\frac{4.5\times10^{-8}}{1.0\times10^{-12}}) = 46.5$ dB and $\beta_2 = 10 \cdot log(\frac{36\times10^{-8}}{1.0\times10^{-12}}) = 55.6$ dB. Table 6.3 supplies the sound level and intensity for various sounds.

Table 6.3: The sound level β and intensity I of different sounds

Source of sound	Sound level β (in dB)	Intensity (in W·m^{-2})	Source of sound	Sound level β (in dB)	Intensity (in W·m^{-2})
Threshold of hearing	0	1×10^{-12}	Loud radio	80	1×10^{-4}
Rustle of leaves	10	1×10^{-11}	Truck traffic	90	1×10^{-3}
Whispering	20	1×10^{-10}	Noisy factory	100	1×10^{-2}
Quiet home	30	1×10^{-9}	Siren within 30m	110	1×10^{-1}
Average home	40	1×10^{-8}	Loud rock concert	120	1
Soft music	50	1×10^{-7}	Threshold of pain	120	1
Normal conversation	60	1×10^{-6}	Jet airplane at 30m	140	1×10^{2}
Busy traffic	70	1×10^{-5}	Bursting of eardrums	160	1×10^{4}

6.5 Electromagnetic Spectrum

As already alluded to in section 4.3 (on thermal radiation), section 6.1 (on wave properties), and section 6.3 (on optics), electromagnetic waves—they are also referred to as electromagnetic radiation—consist of oscillating electric and magnetic fields which transport energy, even in empty space, i.e., the vacuum. As such, they do not need any medium to be able to move around, even though their speed depends on the medium through which they travel. Electromagnetic waves are *transverse* waves and propagate in vacuum at a constant speed of $c = 299,792,458$ m·s^{-1}, which is known as the speed of light, given that electromagnetic waves are what light is made of.

The **electromagnetic spectrum** is defined as the continuous range of frequencies and wavelengths of electromagnetic waves. The relationship between the wavelength and frequency of electromagnetic radiation is based on Equation 6.2, with v_w replaced by the speed of light c, and the amount of energy that the waves carry is determined by the frequency (or wavelength):

$$\begin{cases} c = \lambda f \\ E = hf = \dfrac{hc}{\lambda} \end{cases} \tag{6.16}$$

with $c = 299,792,458$ m·s^{-1} (or rounded off to $c = 3 \times 10^8$ m·s^{-1}) the speed of light, λ the wavelength (in m), f the frequency (in Hz), E the energy of the electromagnetic radiation (in J) per photon (which is a discrete packet of energy, i.e., a particle, and considered the fundamental constituent of light, whereby the photon can behave both as a wave and a particle), and $h = 6.626 \times 10^{-34}$ J·s Planck's constant. For instance, electromagnetic radiation with a frequency of $f = 4.84 \times 10^{14}$ Hz (or 484 THz, with T the symbol for tera, which indicates an order of magnitude of 12 (10^{12})) corresponds to a wavelength of $\lambda = 6.19 \times 10^{-7}$ m (or 619 nm, with n the symbol for nano, which implies an order of magnitude of -9 (10^{-9})), and an energy of 3.2×10^{-19} J (or 2 eV, with 1 eV $= 1.602 \times 10^{-19}$ J, whereby eV is called the electron-volt).

As the electromagnetic spectrum deals with very large and very small numbers, it is useful to get acquainted with several commonly used unit prefixes and their related order of magnitude. Table 6.4 presents an overview.

Table 6.4: An overview of some commonly used unit prefixes

Prefix		Power of	Prefix		Power of
Name	**Symbol**	**ten**	**Name**	**Symbol**	**ten**
Zetta	Z	10^{21}	Deci	d	10^{-1}
Exa	E	10^{18}	Centi	c	10^{-2}
Peta	P	10^{15}	Milli	m	10^{-3}
Tera	T	10^{12}	Micro	μ	10^{-6}
Giga	G	10^{9}	Nano	n	10^{-9}
Mega	M	10^{6}	Pico	p	10^{-12}
Kilo	k	10^{3}	Femto	f	10^{-15}
Hecto	h	10^{2}	Atto	a	10^{-18}
Deca	da	10^{1}	Zepto	z	10^{-21}

Equation 6.16 discloses that electromagnetic waves with *higher (lower)* frequencies possess *shorter (longer)* wavelengths and are carrying a *higher (lower)* amount of energy with them.

Although the electromagnetic spectrum is a *continuous* range of electromagnetic radiation, it is typically categorized in **various bands of frequency**. Ordered from lower to higher frequency (or, alternatively, from longer to shorter wavelengths) the different types of radiation are: radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays. The reason why these bands are established in the first place is due to their distinct characteristics. That is to say, the various types of electromagnetic radiation differ from one another in terms of their origin, the particular way in which they interact with matter (which depends on their wavelength and energy levels), and their practical applications.

Table 6.5: The different regions of electromagnetic radiation

Name	Frequency band (in Hz)	Wavelength range	Energy (in eV, per photon)
Radio waves	$< 3 \times 10^9$	> 10 cm	$< 10^{-5}$
Microwaves	$3 \times 10^9 - 3 \times 10^{12}$	10 cm – 100 μm	$10^{-5} - 10^{-2}$
Infrared (IR)	$3 \times 10^{12} - 4 \times 10^{14}$	100 μm – 750 nm	$10^{-2} - 1.65$
Visible light	$4 \times 10^{14} - 7.5 \times 10^{14}$	750 nm – 400 nm	$1.65 - 3.1$
Ultraviolet (UV)	$7.5 \times 10^{14} - 3 \times 10^{17}$	400 nm – 1 nm	$3.1 - 10^3$
X-rays	$3 \times 10^{17} - 3 \times 10^{20}$	1 nm – 1 pm	$10^3 - 10^6$
Gamma rays	$> 3 \times 10^{20}$	< 1 pm	$> 10^6$

Due to the continuous characteristic of the electromagnetic spectrum, the boundaries of the various regions indicated in Table 6.5 ought to be interpreted as *soft* boundaries, in that the numbers can vary. For instance, there is usually some overlap when defining radio waves and microwaves; some sources mention that microwaves start from a wavelength of 1 mm or even 1 m. Notwithstanding the discussion on the fluidity of the width of the different bands, what is typical for radiation in the proximity of these borders is that they exhibit features of *both* the upper and the lower band. Let us explore now the distinct behaviour of the respective parts of the electromagnetic spectrum.

Radio waves possess the *longest* wavelengths of all the electromagnetic waves. The radiation at this low-frequency end of the electromagnetic spectrum can be generated electronically by oscillating electric and magnetic fields, as in the case of alternating currents in electrical circuits and electromagnetic induction (see section 2), and can be emitted and received by means of radio antennas. However, also stars, planets, as well as gas and dust clouds throughout the Universe, including the Earth and the Sun, produce radio waves (together with most of the other types of radiation) due to accelerating charges and fluctuating magnetic lines. Radio waves travel mostly undisturbed through the Earth's atmosphere, except in the uppermost regions where ions circulate, i.e., the ionosphere, which reflects the radio waves back to the Earth's surface. As a result, radio waves can bounce their way across the planet, which is why we use this type of electromagnetic radiation mainly for *long-distance telecommunication* purposes, e.g., AM and FM radio broadcasts, television, maritime and aviation navigation, mobile phones, and military communication.

Microwaves are defined as radio waves with shorter wavelengths and can be artificially created not only by smaller dish antennas—in contrast to radio antennas, the emitting and receiving microwave antennas must be positioned in line of sight—but also with the assistance of a device called a *magnetron*, which is installed in our typical microwave ovens. Inside the magnetron, accelerating electrons produce microwaves, which constantly reflect off the metal walls of the microwave oven. Since these waves resonate with the water and fat molecules within the food, they can deposit their energy in them, which enhances the average kinetic energy of the molecules, leading to a rise in temperature, and heating up the food. Besides cooking food, microwaves are also useful for radar technology, radio astronomy, and communication via satellite (e.g., the Global Positioning System (GPS), personal communication services, and digital satellite TV and radio), as they propagate almost unrestrictedly through the Earth's ionosphere. Due to their strength in point-to-point communication, microwaves also come in handy for short-distance wireless communication services (e.g., WLAN via Wi-Fi network, Bluetooth, and Wi-Fi Direct).

Infrared (IR) radiation is emitted and absorbed by all objects that possess a temperature above absolute zero, since the frequency of IR radiation resonates with the molecular vibrations of gas, liquid, and solid particles (see section 4.3). It is therefore commonly referred to as *thermal radiation or heat*. These electromagnetic waves are responsible for the (rising) temperatures on Earth, because instead of granting them free passage towards space, gas particles in the atmosphere absorb (and re-emit) the reflected infrared waves from the Earth's surface, so that thermal radiation dwells for longer periods of time within the atmosphere. Examples of their practical applications include electrical heaters, infrared cameras, and infrared spectroscopy, i.e., a technique that is used to identify chemical compounds. IR radiation also serves a purpose in astronomy, as it can probe regions that are hidden from sight by gas and dust clouds—due to their smaller wavelength, IR waves are able to penetrate deeper into these clouds and reveal astrophysical objects and phenomena that visible light could not detect.

Visible light is the electromagnetic radiation that the cones and rods at the back of the human eye detect, transform into electrical signals, and transmit via the optical nerve to the brain, where it is processed by the interplay of the parietal, temporal, and occipital lobes to form a visual image. This idiosyncratic characteristic sets visible light apart from the rest of the electromagnetic spectrum, in that it is the only frequency band that human beings are able to detect visually—interestingly, some animals, such as goldfish, pet vipers, mosquitos, and salmon, can also see infrared radiation. Although the primary reservoir of visible light is the Sun—in fact, among all types of radiation that the Sun emits, visible light makes up the *largest* portion—other origins include artificial sources (e.g., fluorescent glass tubes, LEDs, incandescent light bulbs, and lasers), stars, the Northern Lights, comets, lightning, volcanoes, fires, and bioluminescence (whereby marine organisms, e.g., algae, worms, and sea stars, emit visible light). The range of visible light runs from the colour red, which sits with its longest wavelength (i.e., 750 nm) at the lower frequency end, to the colour violet at the highest end, which displays the shortest wavelength (i.e., 400 nm). Visible light is furthermore absorbed by atoms to a greater extent than IR radiation, as it possesses the right amount of energy not only to excite atoms but also to *push electrons to higher energy levels* within the atom. Apart from vision, other applications of visible light are the process of photosynthesis by plants, cyanobacteria, and phytoplankton, as well as information transmission via optic fibers made of glass.

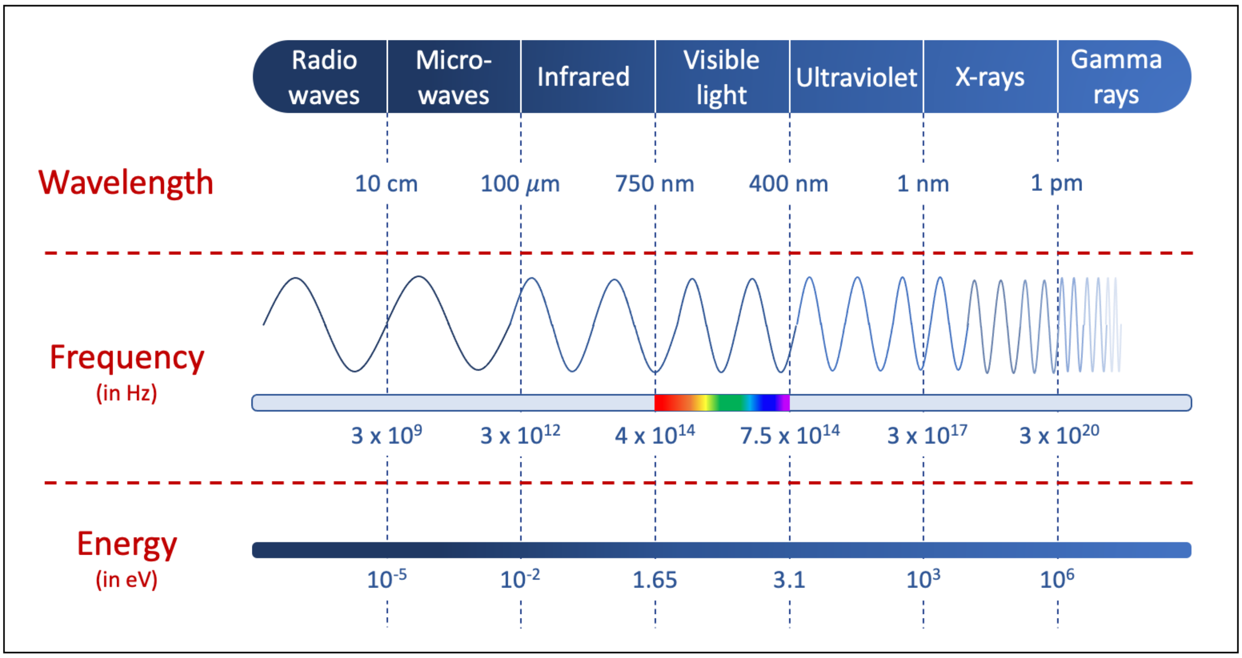


Figure 6.7: The electromagnetic spectrum

Ultraviolet (UV) radiation consists of electromagnetic waves that contain energy to push electrons to still higher levels within the atom (relative to visible light) and even enough energy to *severe chemical bonds and knock electrons off atoms*—this creates charged particles called ions—which sets in motion chemical reactions. The UV waves higher up in the UV frequency band together with X-rays and gamma rays are jointly designated as *ionizing radiation* (see section 7.3), which, besides impairing the skin and other tissue, causes irreversible damage to DNA molecules. The major source of UV light is the Sun, whereby the UV waves with the shortest wavelengths (called UV-B and UV-C) are almost entirely stopped in their tracks by the upper layers of the Earth's atmosphere—they are the reason why the ionosphere exists, as they carry sufficient energy to ionize the gas particles. Longer, unprotected exposure to lower-frequency UV radiation (called UV-A) and UV-B can lead to skin cancer, snow blindness, or cataracts. UV light can also be made artificially by lasers, blue-emitting LEDs, UV lamps (which have the potential to disinfect air, water, and surfaces of solids), transilluminators (which are devices that make DNA visible in gels), and crosslinkers (which are irradiation systems that allow DNA to bond with certain proteins). Other applications of UV radiation are fluorescence-based energy-efficient lamps and treatments of psoriasis (which is a chronic autoimmune disease, characterized by the appearance of red patches on the skin).

X-rays are electromagnetic waves that are created when electrons *closest* to the atom's nucleus are kicked out of the atom. Although the main source of X-rays is the Sun—more precisely, it is the upper layer of the Sun's atmosphere, i.e., the corona—very dense stars, such as neutron stars and black holes, as well as explosions of stars (supernovae) and other high-energy events in space also give rise to X-rays. Other than natural sources, these high-frequency waves can be artificially produced in X-ray tubes, whereby high-speed electrons are hitting a fixed (tungsten) target, producing X-rays in the process. Towards the higher frequency end of the X-ray band (the so-called *hard X-rays*), the waves are absorbed or scattered (reflected) to a lesser degree due to their smaller wavelength, which renders them very suitable for diagnostic medical imaging (radiography) or airport security scanners—which is also why this type of radiation is often referred to as *Röntgen radiation*. Even though

X-rays qualify as ionizing radiation and therefore pose a substantial health hazard, they can be turned into an advantage for instance in the case of cancer treatment.

Gamma rays are electromagnetic waves that are emitted by atoms when changes occur *within* the atom's nucleus, i.e., radioactive decay (see section 7.2). As such, they are a form of nuclear radiation, in contrast to X-rays, which are produced *outside* the nucleus. A couple of examples of atoms that emit gamma rays include cobalt-60, cesium-137, iridium-192, and radium-226. Of all radiation types, gamma rays possess the *shortest* wavelengths, which explains, for instance, why they travel the farthest in water. Besides during collisions between electron beams and laser pulses, gamma rays can also be created during nuclear explosions, lightning, and high-energy events in space (e.g., gamma-ray bursts, neutron stars, and supernova explosions)—note that the Sun usually does *not* emit gamma rays, except during the occasional solar flares. Despite the fact that gamma rays are a form of ionizing radiation due to their high-energy content, they are utilized, among other practical applications, in the production process of nuclear energy, in radiotherapy, in food irradiation (gamma rays help to reduce the presence of germs and diseases as well as enhance food preservation; note that irradiated food does not become radioactive), and in radionuclide scanning (this is an medical imaging technique whereby gamma rays, which are emitted by minuscule radioactive tracers inserted into the body, make infections, tumours, or other internal traumas light up more brightly on the produced scan).

7 Radioactivity

7.1 Atomic Structure

Although the structure and behaviour of an atom is usually described by the nuclear model, bear in mind that this description mostly applies to the lightest of elements (i.e., hydrogen or helium). For the heavier elements, the valence shell model, which is an extension of the nuclear model and not further elaborated here, is more accurate in explaining their behaviour.

In the **nuclear model** (a.k.a. the Bohr model after physicist Niels Bohr), an atom consists of both a nucleus, which harbours positively charged **protons** $\mathbf{p^+}$ and neutrally charged **neutrons n**, and negatively charged **electrons** $\mathbf{e^-}$, which circle around the nucleus in stable orbits. While the protons and electrons are attracted to each other electrostatically by the Coulomb force (see section 1.1), the electrons remain in an *orbit at fixed distances from the nucleus* due to the fact that energy comes in *discrete packets called photons* or quanta of energy. In other words, the electrons can only orbit the nucleus at certain discrete energy levels and cannot exist in between two energy states.

Within their respective stable orbits, electrons find themselves in the lowest possible energy state and it *requires* energy for an electron to reside temporarily in another orbit that lies *farther* away from the nucleus. Put differently, electrons in the outer orbits contain more energy than those populating the inner orbits. That necessary energy to jump to a more distant orbit is procured at the moment when an atom is excited by (i.e., *absorbs*) an incoming photon that possesses an energy which matches exactly the difference in energy between these two orbits—this energy is defined by Equation 6.16. Conversely, an electron that momentarily dwells in a higher-energy orbit can fall down to a lower-energy orbit by *sending out* a photon with an energy that corresponds to the energy difference of the two orbits—this only occurs when there is an available slot for the electron in the lower-energy orbit.

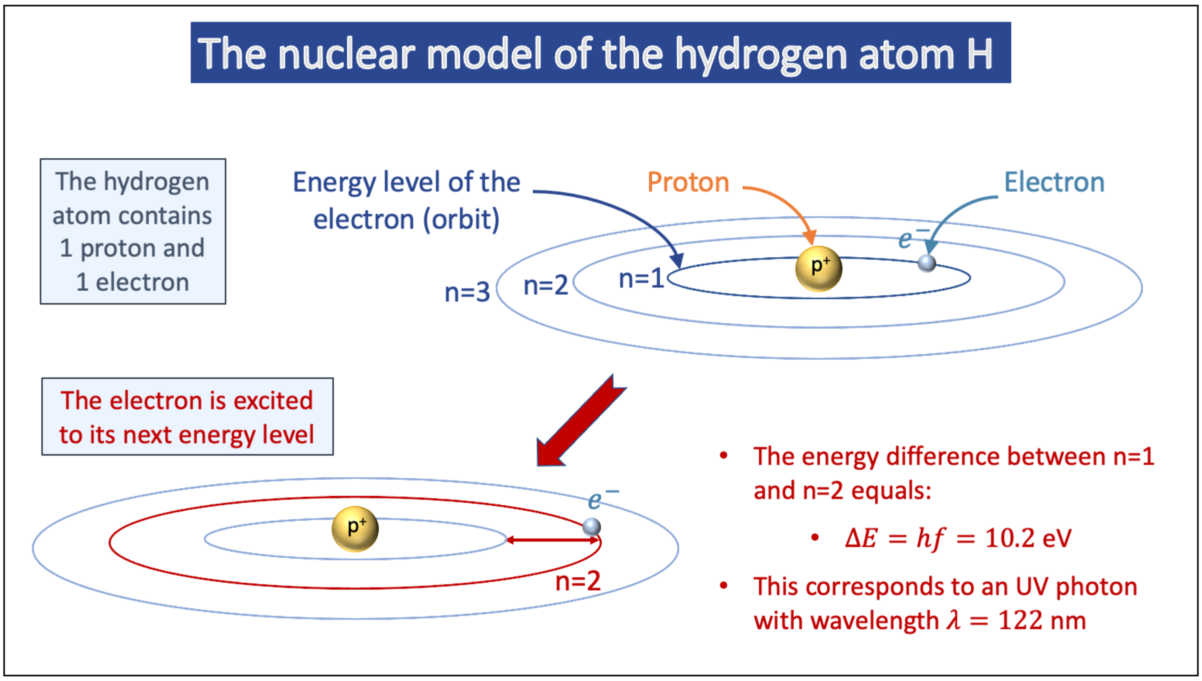


Figure 7.1: The nuclear model of the hydrogen atom

Making an atom vibrate more does not necessarily mean that the electrons become detached from the atoms—remember that an atom is electrically neutral when it holds an equal number of protons and electrons and is said to be **ionized** when either losing or gaining electrons. As a matter of fact, the *closer* to the nucleus electrons flow, the *more* energy it takes to remove them from their orbits, since the electrostatic force acts more intensely on them. This especially applies to heavier atoms, which house more protons within their nucleus and unleash a larger Coulomb force between the protons and the electrons. This is why it takes X-ray photons to knock out electrons from the innermost regions in heavier atoms (see section 6.5).

An atom of a particular chemical element is distinguished from an atom of another element by the number of protons it accommodates within its nucleus. This number is referred to as the **atomic number Z** and is typically written at the bottom left of the symbol of the respective chemical element. When it comes to the mass of atoms, it is predominately determined by the nucleus, i.e., protons and neutrons. That is to say, if we assign a mass of 1 to the neutron, then the mass of the proton amounts to $0.9986 \approx 1$, whereas that of the electron is equal to 0.00054. As a result, the **mass number A** is defined as the total number of protons and neutrons in the nucleus of an atom, and it closely resembles the actual mass of the atom, called the atomic mass—the mass number is usually positioned at the top left of the chemical symbol. The number of neutrons in an atom can then be found as follows:

$$\begin{cases} \textit{Number of neutrons} & = & \textit{mass number A} & - & \textit{atomic number Z} \\ \#(n) & = & \#(n+p) & - & \#(p) \end{cases} \tag{7.1}$$

What is more, the number of neutrons within the nucleus of atoms of the *same* chemical element can vary. That is, one atom can take on several stable forms with each time a different number of neutrons. These various forms are designated as **isotopes**, whereby the mass number A changes per isotope while the atomic number Z remains the same. One *specific* isotope is then called a **nuclide**, which is indicated by the **symbol** $\mathbf{{}^{A}_{Z}E}$ (with E a particular chemical element or atom). In other words, for a nuclide, *both* the atomic number Z and the mass number A must be known. For instance, one of the isotopes of the element gold (${}^{197}_{79}\mathrm{Au}$) is the nuclide ${}^{195}_{79}\mathrm{Au}$, which contains $195 - 79 = 116$ neutrons instead of the usual $197 - 79 = 118$ for the standard gold atom. As a counterexample, an isotope with mass number 15 is not a nuclide, given that it is not specified or unique, because it can refer both to the isotope 15 of oxygen (${}^{15}_{8}\mathrm{O}$) and to the (rare) isotope 15 of nitrogen (${}^{15}_{7}\mathrm{N}$)—these two particular isotopes are indeed nuclides, as their respective atomic number is now identified.

The **relative atomic mass** of a chemical element is defined as the weighted average of the atomic masses of all the stable isotopes of that element per atomic mass unit (amu). As stated previously, the atomic mass is usually approximated by the mass number A. The weights applied are the relative abundances of the isotopes, i.e., the relative extent to which the different isotopes are naturally present on Earth (given in percentages). Finally, 1 amu is measured as $\frac{1}{12}$ of the mass of the atom ${}^{12}_{6}\mathrm{C}$. Given that the atomic masses of the elements are expressed in amu, the relative atomic mass is unitless and is formulated as:

$$\textit{Relative atomic mass} = \sum_{i=1}^{n} \left(\frac{\textit{natural abundance (\%)}}{100} \right)_i \times \textit{(mass number A)}_i \qquad (7.2)$$

whereby i runs from 1 to n, with n the total number of isotopes. For example, silver has two stable isotopes, i.e., ${}^{107}_{47}\mathrm{Ag}$ and ${}^{109}_{47}\mathrm{Ag}$, with natural abundances of 51.84% and 48.16%, respectively. Its relative atomic mass is then calculated as $(0.5184 \times 107) + (0.4816 \times 109) = 107.96$. Another example involves the relative atomic mass of 39.13 for potassium. If the relative abundance of the first of the two stable isotopes ${}^{39}_{19}\mathrm{K}$ and ${}^{41}_{19}\mathrm{K}$ is equal to 93.26%, then the natural abundance of the second isotope is calculated as $(\frac{39.13-0.9326\times 39}{41}) \times 100 = 6.73\%$.

7.2 Radioactive Decay

Within the nucleus of an atom, there is a continuous give and take between, on the one hand, repelling electrostatic forces due to the presence of positively charged protons and, on the other hand, attractive forces—called the strong nuclear force—which act on the interaction between protons and neutrons and hold the nucleus together. As a result, the (in)stability of a nucleus is determined by the exact ratio between the number of neutrons and protons. Nevertheless, *all the nuclei* with an atomic number Z equal to or higher than 84—the element with $Z = 84$ is called polonium—are **radioactive**, i.e., all of their isotopes are unstable.

In the case of an *unstable* nucleus (regardless of the atomic number), nuclear changes take place whereby the parent nucleus (${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{E}_{\mathrm{p}}$) either transforms into different daughter nuclei (${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{E}_{\mathrm{d}}$) or lowers its energy levels. This is the process defined as **radioactive decay**, which occurs in a *spontaneous and random fashion* and is accompanied by the *emission of ionizing radiation.* The process of nuclear decay only grinds to a halt when the resultant daughter nucleus is stable.

The accompanying radiation takes the form of either *high-energy radiation*, i.e., gamma rays, or *emitted particles.* Even so, gamma rays are usually always associated with nuclear changes, regardless of the type of radioactive decay. Three modes of nuclear decay that are discussed in this section are: alpha decay, beta decay, and gamma emission—note that the type of beta decay that is mentioned here is what is known as beta-minus decay.

Alpha decay is the type whereby an alpha particle α is emitted, which is equal to the nucleus of a helium atom, i.e. ${}^{4}_{2}\mathrm{He}$. This means that the parent nucleus loses 4 units of mass number A and 2 units of atomic number Z. Put differently, in the case of alpha decay, the daughter nucleus is always the atom which is positioned *two places to the left* of the parent nucleus in the periodic table of chemical elements.

Beta decay refers to the type of radioactive decay that allows one neutron of the parent nucleus to decay into one proton and a beta particle β, which consists of an electron ${}^{0}_{-1}\mathrm{e}$. Losing one neutron and gaining one proton means that the daughter nucleus retains the original mass number A and gains 1 unit of atomic number Z. In other words, in the event of beta decay, the daughter nucleus is equal to the chemical element *one place to the right* of the parent nucleus in the periodic table.

Gamma emission is understood as nuclear decay during which both the mass number A and the atomic number Z remain *constant*—the chemical element does not transform into another element—and during which the energy level of the parent nucleus *diminishes*. Similar to the situation whereby electrons emit photons when falling to a lower-energy level orbit, the nucleus emits gamma radiation ${}^{0}_{0}\gamma$ when reducing its energy level, with the difference that this nuclear radiation contains a higher amount of energy with respect to that of the emitting electron—gamma rays are indeed the type of electromagnetic radiation that contains the most energy (see section 6.5). Note that the *excited* parent nucleus is recognized by an asterisk at the top right of the symbol (${}^{A}_{Z}\mathrm{E}^{*}_{p}$).

Table 7.1: Three different types of radioactive decay

Name	Symbol	Emitted radiation	Change in A	Change in Z	General Equation
Alpha decay	α	${}^{4}_{2}\alpha$ (${}^{4}_{2}\mathrm{He}$)	-4	-2	${}^{A}_{Z}\mathrm{E}_{p} \longrightarrow {}^{A-4}_{Z-2}\mathrm{E}_{d} + {}^{4}_{2}\alpha$
Beta decay	β	${}^{0}_{-1}\beta$ (${}^{0}_{-1}\mathrm{e}$)	0	+1	${}^{A}_{Z}\mathrm{E}_{p} \longrightarrow {}_{Z+1}^{\;\;A}\mathrm{E}_{d} + {}^{0}_{-1}\beta$
Gamma emission	γ	${}^{0}_{0}\gamma$	0	0	${}^{A}_{Z}\mathrm{E}^{*}_{p} \longrightarrow {}^{A}_{Z}\mathrm{E}_{p} + {}^{0}_{0}\gamma$

Consider the following set of examples of the three types of nuclear decay:

$$\begin{cases} \textit{Alpha decay}: & {}^{223}_{87}\mathrm{Fr} \longrightarrow {}^{219}_{85}\mathrm{At} + {}^{4}_{2}\alpha \\ \textit{Beta decay}: & {}^{211}_{83}\mathrm{Bi} \longrightarrow {}^{211}_{84}\mathrm{Po} + {}^{0}_{-1}\beta \\ \textit{Gamma emission}: & {}^{234}_{91}\mathrm{Pa}^{*} \longrightarrow {}^{234}_{91}\mathrm{Pa} + {}^{0}_{0}\gamma \end{cases} \tag{7.3}$$

The examples of alpha and beta decay in Equation 7.3 form part of a nuclear decay chain called the Actinium Cascade or the Actinium Series. With regard to the α-decay, the parent nucleus francium (${}^{223}_{87}\mathrm{Fr}$) decays into the daughter nucleus astatine (${}^{219}_{85}\mathrm{At}$) whereby the mass number and the atomic number decrease from $A_{Fr} = 223$ to $A_{At} = 219$ (with a change equal to $\Delta A = -4$) and from $Z_{Fr} = 87$ to $Z_{At} = 85$ (with a change equal to $\Delta Z = -2$), respectively. The difference in both numbers is compensated by the emitted α-particle ${}^{4}_{2}\alpha$.

In the event of the β-decay, the parent nucleus bismuth (${}^{211}_{83}\mathrm{Bi}$) decays into the daughter nucleus polonium (${}^{211}_{84}\mathrm{Po}$), whereby the mass number does not change ($\Delta A = 0$) and the atomic number increases with one unit ($\Delta Z = +1$), which is offset by the β-particle ${}^{0}_{-1}\beta$.

In the example of the γ-emission, the element proactinium does not modify the composition of its nucleus, i.e., $\Delta A = 0$ and $\Delta Z = 0$, but lowers its energy level from an excited state (${}^{234}_{91}\mathrm{Pa}^{*}$) to a reduced energy level (${}^{234}_{91}\mathrm{Pa}$) by emitting a high-energy photon (${}^{0}_{0}\gamma$).

It becomes clear from the above examples that the *total* number of protons and neutrons before and after the radioactive decay remains *unaltered*. In other words, the mass number A is conserved. Other physical quantities that are also conserved include energy, momentum, and electric charge. Finally, Fig. 7.2 provides an example of a full radioactive decay chain.

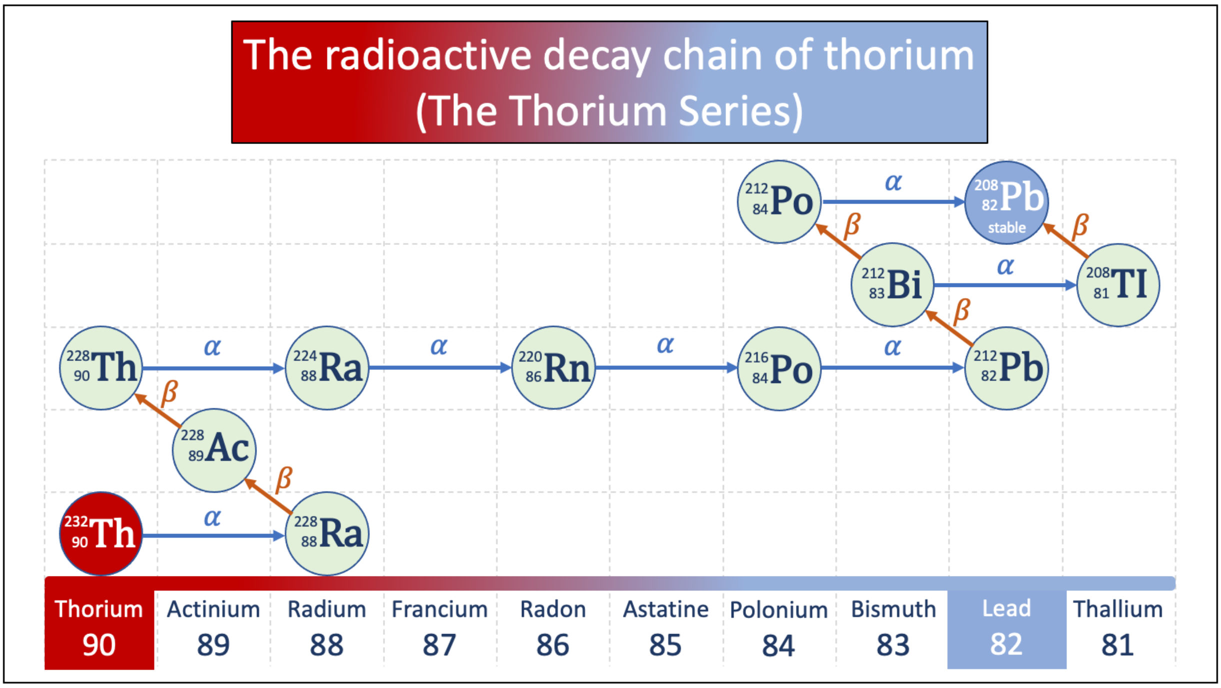


Figure 7.2: The Thorium Series of radioactive decay

7.3 Ionizing radiation

The three forms of emitted radiation, i.e., α-, β-, and γ-radiation, that are coupled to the respective type of nuclear decay differ from each other in at least three ways: by their penetrating abilities, by their ionizing power, and by their deflection in electric and magnetic fields.

The **penetration power** of the ionizing radiation describes to what extent nuclear radiation is able to infiltrate matter. Relative to the β-particle, the α-particle is *larger in size* (by a factor of roughly 10,000), *slower* (it is larger in mass) and has *more charge* (2 protons versus 1 electron), which implies that the α-particle has the *lowest* ability to pierce through matter, since it quickly collides and interacts with the surrounding molecules. Since electromagnetic radiation has no mass or charge and propagates close to the speed of light in matter (see section 6.3), γ-rays possess the *strongest* penetration power. Whereas a layer of dead skin or a sheet of paper is sufficient to block α-radiation, it requires a couple of millimeters of aluminum and a number of centimeters of lead to impede the passage of β- and γ-radiation, respectively.

To provide another example of their relative penetration power, γ-rays can move through the air without much obstruction for several hundreds and even thousands of meters, whereas β-particles are put to a stop already after a couple of meters. In an even sharper contrast, α-radiation does not get much farther in its air travels than an average of 3.7cm.

The **ionizing power** refers to the likelihood of nuclear radiation to strip away electrons from the atoms and molecules of the material through which it travels. In view of the fact that α-

particles move slower, are larger in size, and possess more charge with respect to β-particles, it follows that α-particles come with an *increased* chance of engaging in interactions. That is, α-particles have a *larger* potential to ionize atoms and molecules—by getting into contact with molecules, the α-particles, which have a positive electric charge of 2, attract two electrons from nearby molecules and transform into a neutral, harmless helium atom, leaving the molecules ionized. In contrast, γ-rays *hardly* interact with matter, relatively speaking, and can easily pass through a human body without ionizing any atoms or molecules. However, due to their high penetration power and high-energy content, they can cause considerable damage to the deeper regions within living matter, such as the nucleus of a cell where DNA is stored.

Table 7.2: The relative penetration and ionizing power of nuclear radiation

Radiation	Relative penetration power	Relative ionizing power	Type of shielding
α	weakest	strongest	A sheet of paper
β	average	average	Aluminum plate (2-4 mm)
γ	strongest	weakest	Lead (1-4 cm)

Considering that ions are the workhorses of living cells, the reason why ionizing radiation can be problematic for living tissue is because too much of radiation modifies the structure and functioning of molecules, so that cells are no longer able to carry out their activities in the way that they should. What is more, exposure to radiation is almost always accompanied by the destruction of cell walls, which additionally hampers the cell's functioning. Although α-particles penetrate poorly, they can become significantly harmful to the human body in the case that the particles are inhaled, ingested, or incorporated within the bloodstream through other means—this is called internal exposure—since they possess a substantial ionizing power.

The third way by which these three types of nuclear radiation can be distinguished from one another is their **deflective behaviour within electric and magnetic fields**. As pointed out in section 1.1, particles of equal charge repel each other, while those with an opposite charge attract. Imagine the setup of two vertically positioned, electrically charged plates whereby the plate on the left (right) is negatively (positively) charged. The electric field lines are then flowing perpendicular to the plates, from the right to the left plate. If a *charged* particle is passing through the opening between the plates, the *positively* charged α-particle will then bend to the left towards the *negatively* charged plate, whereas the *negatively* charged β-particle will curve to the right in the direction of the *positively* charged plate. Given a *zero* net charge, γ-rays are unhindered by the electric field and do therefore *not* deflect. The left panel in Fig. 7.3 describes the situation whereby the polarity of the plates in the above-mentioned example are reversed.

Consider next the presence of a magnetic field pointing straight upwards through which ionizing radiation is moving—it is furthermore assumed that the radiation can propagate at any angle with the magnetic field lines, but not parallel to them. Since charged particles undergo a force when subject to a magnetic field (see section 2.3), the α-particle (β-particle) that traverses the magnetic field in the direction of the geographical *north* is deflected towards the *east (west)*—keep in mind that no deflection would occur if you were doing this experiment on the South Pole, as the direction of motion would be parallel to the magnetic field lines. As with electric fields, γ-rays are unaffected by the magnetic field and do *not* deflect. The right panel of Fig. 7.3 provides another example of radiation deflecting in magnetic fields.

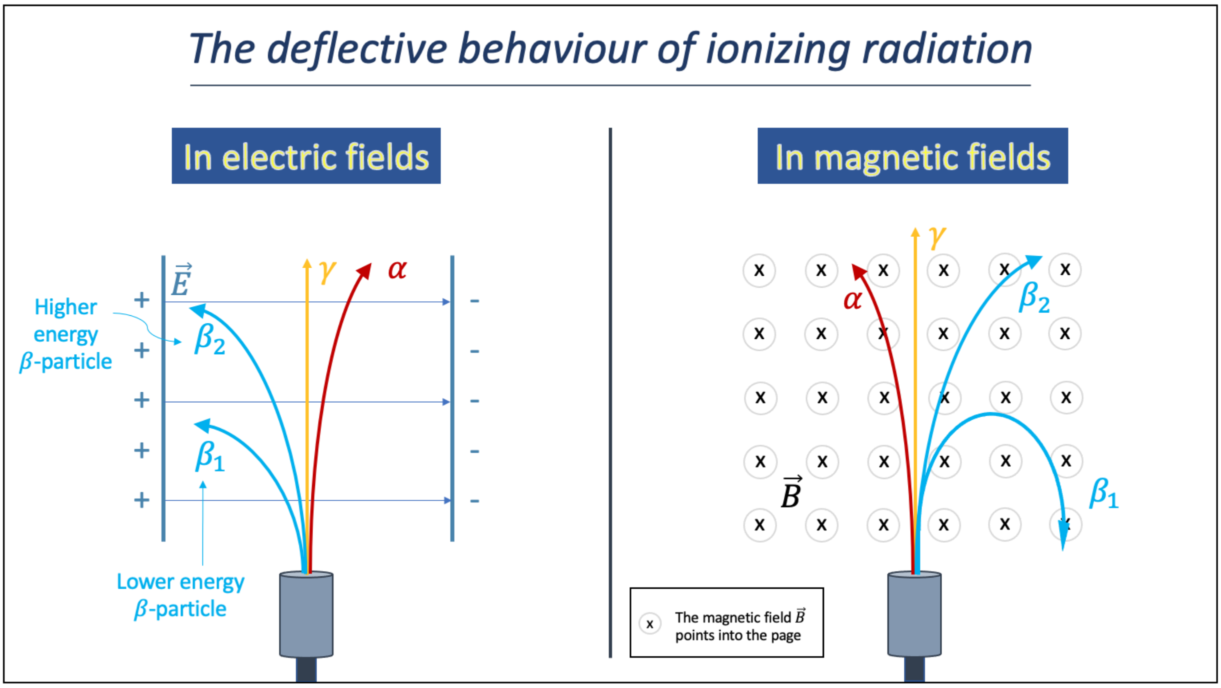


Figure 7.3: The deflective behaviour of nuclear radiation in electric and magnetic fields

With regard to both electric and magnetic fields, the amount by which deflection takes place depends on the charge to mass ratio as well as the energy of the incoming radiation particle. The *faster* a particle is moving, the *larger* its momentum, the *more pronounced* its inertia, and the *greater* the amount of kinetic energy that the particle is carrying. In other words, it shows a *greater* resistance to modify its direction of motion and thus exhibits *less* deflection (see Newton's first law (Equation 3.9)). Moreover, as α-particles are much heavier than β-particles—and thus exhibit a larger momentum—they generally deflect to a lesser extent.

Despite the existing health hazards of ionizing radiation, there are many **applications** that turn these risks into opportunities, as already alluded to in section 6.5. For instance, ionizing radiation plays a role in cancer treatments (radiotherapy), in the preservation of food, in medical imaging (radiography, e.g., via PET scans), in the generation of electricity (via the process of nuclear fission in nuclear power stations), and in smoke detectors (the alarm goes off when a drop in current is registered, which is the result of smoke particles *not* being ionized—and thus preventing a current from flowing—by alpha particles (which are emitted, for instance, by the isotope americium-241); this is considered a harmless application, since alpha particles do not travel very far through the air).

The amount of exposure to human-produced ionizing radiation might be a source of concern to many. However, to put things into perspective, human activities only account for approximately 15% of the total dose of radiation to which we are annually exposed. The remaining 85% comes from **natural background radiation**, which reaches us by *inhalation of radon gas (52%)*, especially when it accumulates in indoor and enclosed spaces, by *external terrestrial exposure (20%)* through building materials and radioactive rocks and soil within the Earth's crust, by *cosmic radiation (16%)* from space, which increases when we spend more time at higher altitudes, and by *ingestion (12%)* of mainly the radioactive isotopes potassium $^{40}_{19}$K, thorium $^{232}_{90}$Th, and uranium $^{238}_{92}$U, including their decay products, which are present in foods and drinking water.

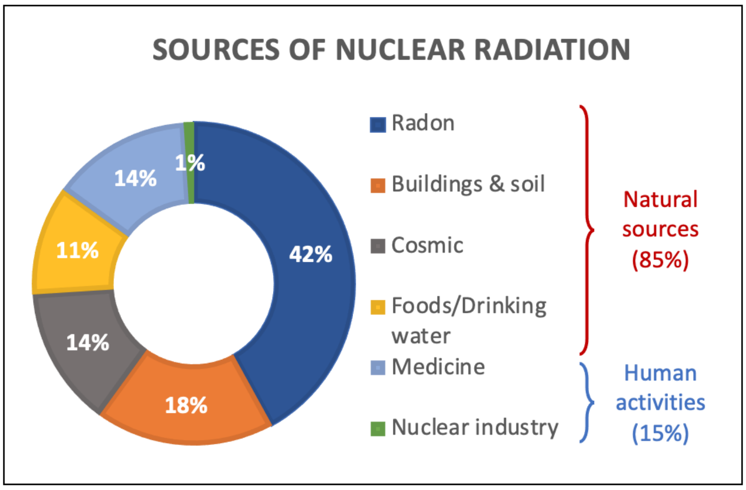


Figure 7.4: The sources of nuclear radiation

Radon gas is a naturally occurring decay product of the uranium and thorium radioactive decay series (see Fig. 7.2), which are together with potassium the *main source* of natural background radiation. What is more, radon is often spewed out by volcanoes, given that potassium, uranium, and thorium subsist within the Earth's core ever since the birth of the Earth some 4.5 billion years ago—in fact, the geothermal heat within the Earth's innermost region, which is responsible for bringing the temperature of the Earth's interior to roughly 6,000°C, is mostly (90%) supplied by the nuclear decay activities of these three elements.

7.4 Half-life

The reason why the radionuclides potassium $^{40}_{19}$K, thorium $^{232}_{90}$Th, and uranium $^{238}_{92}$U have been able to deliver heat energy already for such a long time is that they possess a half-life of many billions of years. The **half-life** of an ensemble of radioactive nuclei is the amount of time it takes that sample to reduce its radioactivity *by half* through nuclear decay. In other words, the half-life is the characteristic time span that is required to slow down the disintegration rate, i.e., the number of disintegrations per second, of a specific radioactive element by 50%.

Although the nuclear decay of *one nucleus* is spontaneous and random—that is, it is not possible to tell exactly when it will decay—it is nevertheless feasible to infer a pattern, i.e., the half-life, of a *large sample* of individual nuclei—it is equivalent to not knowing the exact behaviour of every molecule in a fluid, but being able to understand the behaviour of the fluid as a whole (the underlying reason for why this is possible is embedded within the laws of mechanical statistics). Put differently, for a large sample of radioactive nuclei, it is possible to figure out the **number of atoms** that will decay over a certain time period:

$$N = N_0 \cdot e^{-\lambda t} \tag{7.4}$$

with N the number of remaining parent nuclei after a time t, N_0 the initial number of parent nuclei (at time $t = 0$ s), and λ the decay constant (in s^{-1}). The **decay rate R, a.k.a. the (radio)activity A**, of a sample of nuclei is then defined as:

$$R = \left|\frac{dN}{dt}\right| = |(-\lambda N_0)| \cdot e^{-\lambda t} \quad \Longleftrightarrow \quad A = A_0 \cdot e^{-\lambda t} \tag{7.5}$$

with R the remaining activity of the sample after a time t (in s), which is often denoted by the letter A, whereby $A_0 = |(-\lambda N_0)|$ refers to the absolute value of the initial activity of the sample (at time $t = 0$ s). From Equation 7.5 (or, equally, from Equation 7.4), we can then deduce the definition of the **half-life $\boldsymbol{T_{1/2}}$** by setting A equal to $\frac{A_0}{2}$ (after all, we are interested in finding the time it takes for the radioactive sample to reduce its activity by one-half):

$$\begin{aligned} &\Longrightarrow A = \frac{A_0}{2} = A_0 \cdot e^{-\lambda t} \\ &\Longleftrightarrow e^{\lambda t} = 2 \\ &\Longleftrightarrow \lambda t = \ln 2 \\ &\Longleftrightarrow t = T_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda} \end{aligned} \tag{7.6}$$

whereby ln indicates the natural logarithm of a function and $\ln(e^x) = x$. The half-life $T_{1/2}$ is unique to every radioactive element. For instance, for the radioisotope carbon-14 ($^{14}_{6}C$) the half-life is equal to $T_{1/2} = 5,730$ years, for gallium-68 ($^{68}_{31}C$) it is $T_{1/2} = 68$ minutes, and for uranium-238 ($^{238}_{92}U$) it measures $T_{1/2} = 4.468$ billion years.

Let us consider two numerical examples. The initial radioactivity A_0 of a $^{14}_{6}C$-sample that contains 3.2×10^{15} nuclei is equal to $A_0 = \lambda N_0 = \frac{0.693}{(5{,}730 \times 31{,}536{,}000)} \cdot (3.2 \times 10^{15}) = 12,272$ decays per second (remember to convert the half-life in years into seconds, as the decay rate is expressed in seconds). For the radioisotope $^{13}_{7}N$, which has a half-life of 10 min, an initial radioactivity of $A_0 = 5.83 \times 10^{13}$ disintegrations per second will have slowed down after 4 hours to an activity of $A = A_0 \cdot e^{-\lambda t} = (5.83 \times 10^{13}) \cdot e^{-(\frac{0.693}{10 \times 60})(4 \times 3{,}600)} = 3.5$ million decays per second.

In conclusion, the below Fig. 7.5 shows the exponential behaviour of the radioactive decay rate for the example of copper-67 ($^{67}_{29}Cu$), which has a half-life of 78.26 hours. Remark that radioisotopes with a shorter (longer) half-life, e.g., rubidium-81 ($^{81}_{37}Rb$) (xenon-133 ($^{133}_{54}Xe$)), have a larger (smaller) decay constant λ, which is graphically reflected by a more (less) pronounced declining exponential curve.

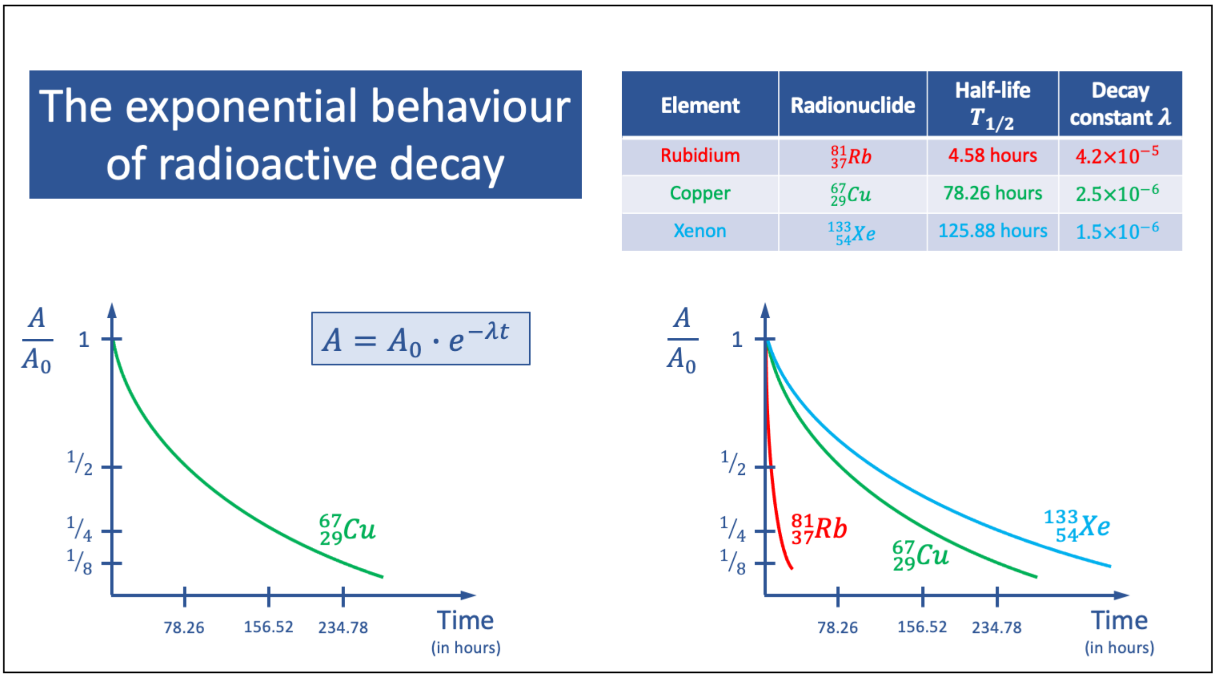


Figure 7.5: The exponential behaviour of radioactivity