

Physics II:

Introduction to physics



PROF. BEN KILMINSTER

INTRODUCTION TO ELECTRICITY, MAGNETISM, ELECTROMAGNETISM, AND THERMODYNAMICS

This script is the second part of an undergraduate course in introductory physics. It is typically taught in the second semester, with part 1 often taught in the previous semester. The level of material is appropriate for physics majors as well as those in the life sciences. The latter may not be expected to learn the full level of detail that would be expected of physics majors, but may still benefit from the additional material in order to understand some concepts in more depth. It is recommended that students should be already familiar with geometry and also take or have taken a class in mathematics that covers vectors and calculus (derivatives and integrals).

Last edited on September 19, 2023
Typeset & figures by Izaak Neutelings

Contents

I	Electromagnetism	9
1	Electric Charges & Fields	11
1.1	Charges	11
1.1.1	Elementary charges	11
1.1.2	Conductors & insulators	14
1.2	The Coulomb force	14
1.3	Electric fields	16
1.3.1	Electric field lines	16
1.4	Electric dipoles	18
1.5	Charge distributions	20
1.5.1	Electric field of a charged rod	22
1.5.2	Electric field of a charged ring	24
1.5.3	Electric field of a charged sheet	24
2	Gauss's Law	27
2.1	Electric flux	27
2.1.1	Interlude: Surface integral and solid angle	28
2.2	Gauss's law	29
2.2.1	Trivial example: Single point charge	30
2.2.2	Electric field of a charged sphere	30
2.2.3	Electric field of a charged rod, revisited	31
2.2.4	Electric field of a charged sheet, revisited	32
2.2.5	Electric field of a charged slab	33
2.3	Interlude: Divergence theorem	33
2.4	Differential form of Gauss's law	34
2.5	Charge distribution in conductors	35
2.5.1	Charged conductors	36
2.5.2	Conductors with a cavity	37
2.5.3	Faraday's cage	38
3	Electric Potential	39
3.1	Electric potential energy	39
3.2	Electric potential	40
3.2.1	Electric potential of a uniform field	42
3.2.2	Electric potential of a point charge	42
3.2.3	Electric potential of a charged, conducting sphere	43
3.3	Equipotential surfaces	43
3.4	Discharge	43
3.5	Non-spherical conductor	45
3.6	Poisson's equation	46
3.7	Extra: Uniqueness theorem	48
3.8	Extra: Method of images	48
3.8.1	A point charge in front of an infinite conductor plane	48
3.8.2	A point charge in front of a spherical conductor	49
4	Capacitance & Dielectrics	51
4.1	Capacitors	51
4.1.1	Capacitance	52
4.1.2	Spherical capacitor	52
4.2	Dielectrics	53
4.2.1	Polarization & displacement	55
4.3	Storage of energy in capacitors	56
4.3.1	Forces in capacitors	56
4.3.2	Energy stored in electric fields	57
4.4	Combining capacitors	57
4.4.1	In parallel	57
4.4.2	In series	58
5	Current	59
5.1	Drift velocity	59
5.2	Current density vector	60
5.3	Continuity equation	61
5.4	Resistance	61
5.5	Resistivity and conductivity	62
5.6	Energy loss in a conductor	63
6	Circuits	65
6.1	Resistors in circuits	65
6.2	Resistors in series	66
6.3	Kirchhoff's rules	67
6.4	Resistors in parallel	68
6.5	Measuring circuits	69
6.6	RC circuits	69
6.6.1	RC discharge	69
6.6.2	Charging a RC circuit	70
6.7	Application: Tubular capacitors and Geiger counters	71
7	Magnetism	73
7.1	Gauss's law for magnetic fields	74
7.2	Magnetic force on an electric charge	75
7.2.1	Lorentz force	76
7.2.2	Application: Velocity selector	76
7.2.3	Application: Mass spectrometer	77
7.2.4	Application: Cathode-ray tube	78
7.3	Currents in magnetic fields	78
7.3.1	Force on a current loop in a magnetic field	79
7.3.2	Torque on a current loop in a magnetic field	80
7.3.3	Magnetic moment	80

7.3.4	Non-uniform field	81	12 Maxwell's Equations	119	
7.4	Magnetization of matter	82	12.1	Differential form of Ampère's law & Faraday's law	119
7.5	Application: Hall effect	83	12.1.1	Interlude: Kelvin-Stokes theorem	119
8	Laws of Magnetism	85	12.2	Ampère's law	120
8.1	Biot-Savart's law	85	12.3	Faraday's law	120
8.1.1	Single, moving charge	85	12.4	Maxwell's equations	121
8.1.2	Current	86	12.4.1	Displacement current	122
8.2	Ampère's law	87	13 Electromagnetic Waves	125	
8.2.1	Magnetic field of a cylinder	87	13.1	Electromagnetic spectrum	126
8.2.2	Magnetic field of a loop	88	13.2	Polarization	127
8.2.3	Magnetic field of a solenoid	88	13.3	Speed of light	127
8.2.4	Permeability	89	13.3.1	Measurement of speed of light	128
8.2.5	Magnetic force between conducting wires	89	13.3.2	Special relativity	128
8.3	Magnetic flux	90	13.3.3	Light year	129
8.3.1	Magnetic flux through a solenoid	91	13.4	Intensity & energy density	129
8.4	Faraday's law	91	13.5	Momentum	130
8.5	Lenz's law	91	13.5.1	Radiation pressure	130
8.5.1	Lenz's law in a changing loop	92	13.6	Creating & detecting electromagnetic waves	131
8.5.2	Moving conductor in a magnetic field	93	13.6.1	Dipoles & antennas	131
8.5.3	Application: Eddy currents	94	13.6.2	EMF of an electromagnetic wave	131
9	Inductance & RLC circuits	95	14 Optics	133	
9.1	Inductance	95	14.1	Huygens' principle	133
9.2	RL circuits	96	14.2	Reflection and refraction of light	134
9.2.1	Energy stored in an RL circuit	97	14.2.1	Total internal reflection	135
9.3	LC circuit	98	14.2.2	Mirage	136
9.3.1	RLC circuit	99	14.2.3	Refraction wavelength	136
10	AC Circuits & Phasors	101	14.2.4	Dispersion & prisms	137
10.1	Simple resistor	101	14.2.5	Rainbow	137
10.2	Capacitive reactance	102	14.3	Scattering	138
10.3	Inductive reactance	103	14.4	Polarization	138
10.4	Phasors	103	14.4.1	Polarization by absorption	139
10.4.1	RLC circuit in series	104	14.4.2	Polarization by reflection	140
10.4.2	RLC circuit in parallel	108	14.4.3	Polarization by scattering	141
10.5	Complex impedance	109	14.4.4	Circular polarization	141
10.5.1	RC circuit	111	14.5	Interference	142
10.5.2	RLC circuit in series	112	14.5.1	Interference pattern	142
10.5.3	RLC circuit in parallel	112	14.5.2	Double-slit interference pattern	144
11	Magnetization	113	14.6	Diffraction	146
11.1	Magnetic moment	113	14.6.1	Application: Crystallography	148
11.2	Quantization of magnetic moment	113	14.7	Iridescence	148
11.3	Magnetization	114	II Thermodynamics	149	
11.3.1	Magnetization in an external field	115	15 Temperature & The Zeroth Law	151	
11.3.2	Paramagnetism	116	15.1	Thermometers and temperature scales	152
11.3.3	Ferromagnetism	117	16 Ideal Gas Law	155	
11.3.4	Diamagnetism	118			
11.3.5	Superconductor	118			

17 Statistical Mechanics	157		
17.1 Equipartition theorem	157		
17.2 Interlude: Random walks	159		
17.3 Interlude: Gaussian distributions	160		
17.4 Maxwell-Boltzmann distribution	161		
17.5 Van der Waals equation	162		
18 Heat & The First Law	165		
18.1 Calorimeters	166		
18.2 Heat transfer	166		
18.3 1st law of thermodynamics	167		
18.4 Heat expansion of solids	168		
18.4.1 Thermal expansion at atomic level	169		
18.4.2 Strain & Young's Modulus	169		
19 Thermal Radiation	171		
19.1 Black body	171		
19.2 Rayleigh-Jeans law	172		
19.3 Plank's law	173		
19.3.1 Quantum mechanical model	174		
20 Phase Changes	175		
20.1 Isotherms, critical temperatures & va- por pressures	175		
20.2 Phase changes & latent heat	176		
		20.2.1 Phase diagrams	177
21 Thermodynamic Processes	179		
21.1 PV diagrams	179		
21.2 Relation between heat capacities	180		
21.3 Adiabatic expansion	182		
21.3.1 Adiabatic free expansion & en- thalpy	184		
22 Engines	187		
22.1 A simple engine	187		
22.2 Steam engine	187		
22.3 Internal combustion engine	188		
22.4 Alpha-type Stirling engine	189		
22.5 Beta-type Stirling engine	189		
23 Entropy & The Second Law	191		
23.1 Efficiency of engines	191		
23.2 Refrigerators	192		
23.3 2nd law of thermodynamics for a re- versible process	193		
23.3.1 Heat pump	195		
23.4 Entropy, reversibility and disorder	195		
23.4.1 Entropy for a reversible cycle	196		
23.4.2 Entropy for an ideal gas	196		
23.5 Entropy and probability	199		

Introduction

These lecture notes cover the basics of electromagnetism and thermodynamics for first-year physics students.

A good reference textbook to follow along with the lectures, which also has many practice exercises, is Tipler, *Physics for Scientists and Engineers*. The textbook *Electrodynamics and Optics* by Dementroeder, which has less explanation, but is sometimes higher level.

Also, *Fundamentals of Physics* by David Halliday, Robert Resnick & Jearl Walker contains many more examples and exercises.

In case you find any errors or typos, please send an e-mail to `izaak.neutelings@uzh.ch` and `ben.kilminster@physik.uzh.ch`.

Part I

Electromagnetism

Chapter 1

Electric Charges & Fields

1.1 Charges

Experiment 1.1 (Charging): In class, we rubbed a glass rod with a silk cloth, like in Fig. 1.1 to make it charged. By touching a metal ball, we transferred the charge. We repeated this experiment with a plastic rod and fur, and made several observations two charged balls: while the balls do nothing if both are uncharged, two charged metal balls *attract* or *repel* each other.

To explain the observations of our experiment, we construct the concept of charge:

1. Charges come in pairs, negative and positive.
2. Same-sign charges repel.
3. Opposite-sign charges attract.

Benjamin Franklin chose the sign convention for charge, where glass will be positively charged, while the silk cloth will be negatively charged.

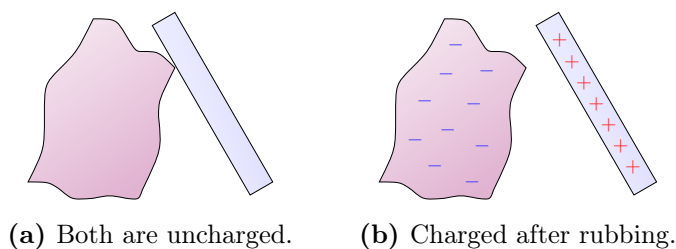


Figure 1.1: Charging a glass rod with a silk cloth.

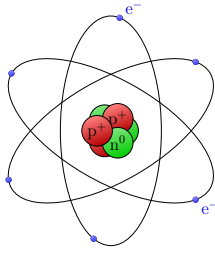
1.1.1 Elementary charges

It turns out that charge is quantized at the atomic-level: There is a smallest unit of charge. This means that any object should have a charge

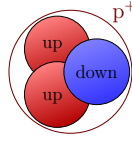
$$Q = ne, \tag{1.1}$$

where n is some integer (0, 1, 2, ...), and e is the *fundamental*, or *elementary unit of charge*,

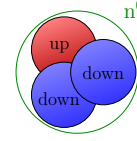
$$e = 1.6 \times 10^{-19} \text{ C}, \tag{1.2}$$



(a) The atom consists of a nucleus and electrons.

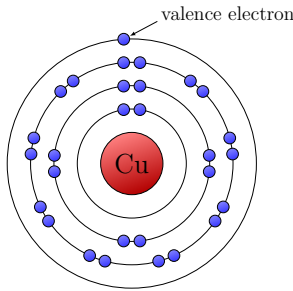


(b) The proton is composed of two up quarks and one down quark.

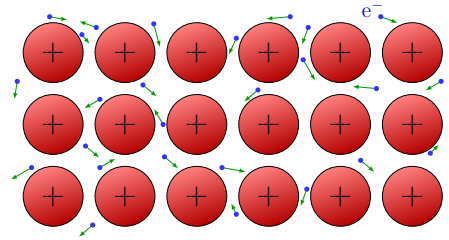


(c) The neutron is composed of one up quark and two down quarks.

Figure 1.2: Elementary charges: electrons ($-1e$), protons ($+1e$), up quarks ($+\frac{2}{3}e$) and down quarks ($-\frac{1}{3}e$).



(a) Copper atoms can share one valence electrons with other atoms.



(b) Sea model of conduction: Valence electrons move freely around the lattice of metal ions.

Figure 1.3: Metal atom and lattice.

where Coulomb (C) is the unit of charge.

We know that matter consists of atoms, which are in turn made out of a nucleus with protons and neutrons, and electron orbiting the nucleus. Protons have a positive charge $Q(p) = +e$, and electrons have a negative charge $Q(e) = -e$, while the neutron is neutral. But protons and neutrons turn out to be composite as well, as is shown in Fig. 1.2; they consist of smaller particles called *quarks*. The proton is actually three quarks bound together: two *up quarks*, one *down quark*. The up quark has $Q(u) = +\frac{2}{3}e$, while the down quark has $Q(d) = -\frac{1}{3}e$. Therefore, the proton will have an overall $+e$ charge. The neutron, on the other hand, is made out of one up and two down quarks, and is therefore neutral. However, one will never observe a “free quark” in nature, because they are only stable in combinations of integer charge. The only way to study them is in high-energy experiments, like in the particle colliders at CERN.

Like energy or momentum conservation, charge conservation is a law of nature. Electrons can move around, but do not spontaneously disappear or appear out of nothingness. In particle physics, however, particles can be created and destroyed, as long as the charge is conserved in the process. For example, it describes photon-electron interactions like

$$\gamma \rightarrow e^- e^+, \quad (1.3)$$

where γ is a photon, e^- is an electron, and e^+ is the anti-particle of an electron, the positron. In this interaction the charge before and after is the same:

$$Q(\gamma) = 0 = Q(e^-) + Q(e^+). \quad (1.4)$$

One can create *static electricity* by rubbing materials, like we saw in the beginning.

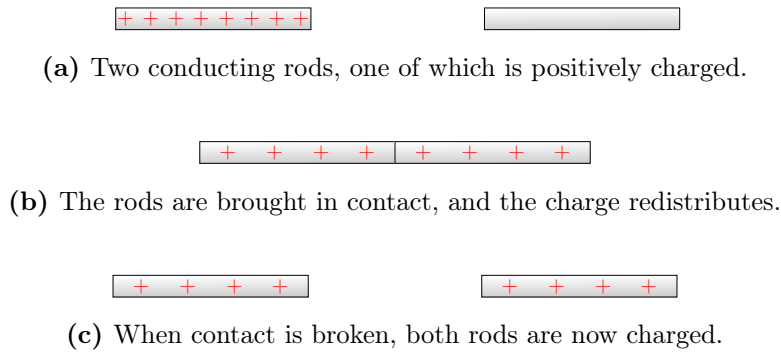


Figure 1.4: Charging of conducting rods by conduction.

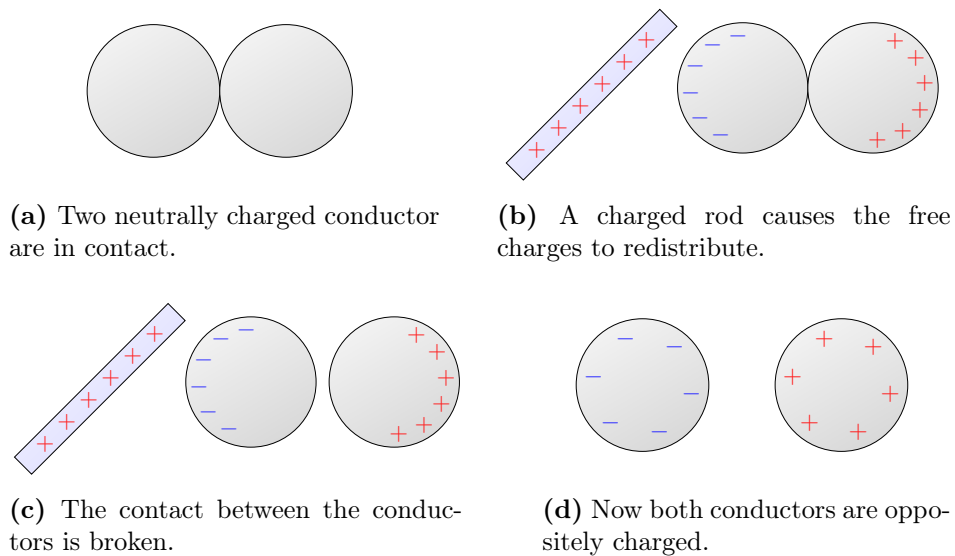


Figure 1.5: Charging of conductors by induction with a charged rod.

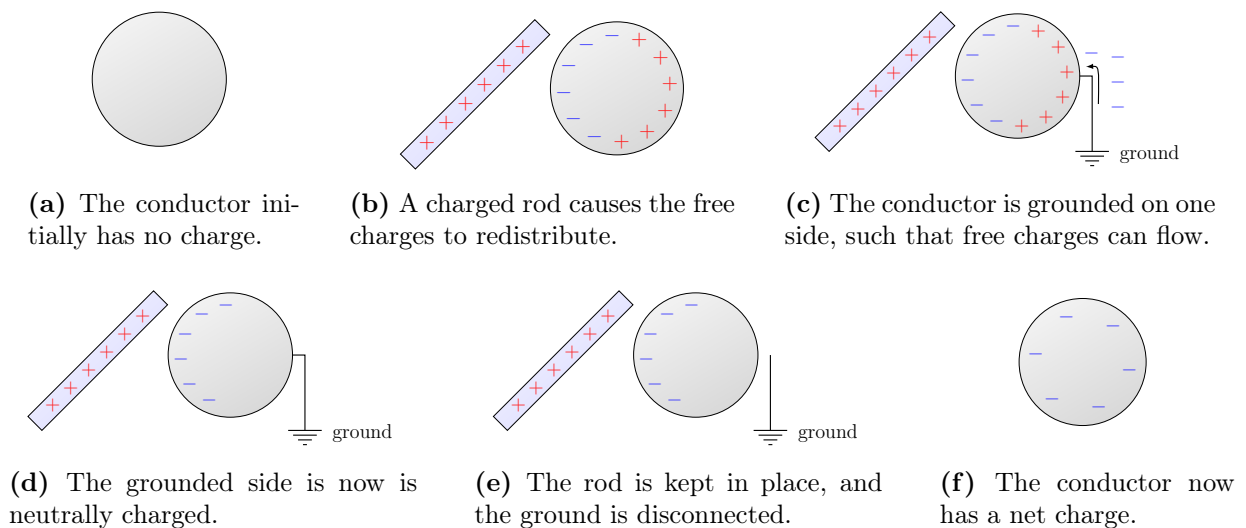


Figure 1.6: Charging of a conductor by induction with a charged rod and grounding.

This happens because electrons are built up. To create 10^{-8} C of charge, you would need

$$\#(\text{electrons}) = \frac{10 \times 10^{-9} \text{ C}}{1.6 \times 10^{-19} \text{ C}} \approx 6 \times 10^{10}. \quad (1.5)$$

1.1.2 Conductors & insulators

Conductors are materials wherein electrons can move freely such that the material can conduct electricity. As opposed to an ionic bond, where electrons are transferred between atoms, or a covalent bond, where electrons are shared between two atoms, in a metal electrons are shared by all atoms, as shown in Fig. 1.3b. The valence electrons of each metal atom can freely bounce around the metal, almost like a gas of free particles. In general, conductor are neutral, but they can be charged by *conduction*, like in Fig. 1.4, or by *induction*, like in Fig. 1.5 and Fig. 1.6.

Insulators, on the other hand, are materials wherein electrons are strongly bound to the atoms, such that they cannot flow freely.

1.2 The Coulomb force

The force between two point charges q_1 and q_2 , separated by a position vector \mathbf{r}_{21} like in Fig. 1.7, is given by *Coulomb force*

Coulomb force.

$$\mathbf{F}_{21} = \frac{kq_1q_2}{r_{21}^2} \hat{\mathbf{r}}_{21}, \quad (1.6)$$

with Coulomb's constant $k = 8.99 \times 10^9 \text{ N m}^2/\text{C}^2$, distance $r_{21} = |\mathbf{r}_{21}|$, unit vector $\hat{\mathbf{r}}_{21} = \mathbf{r}_{21}/|\mathbf{r}_{21}|$ in the direction \mathbf{r}_{21} . Here, \mathbf{F}_{21} is the force on q_1 exerted by q_2 . Thanks to Newton's 3rd law, we know that the force exerted on q_2 by q_1 is

$$\mathbf{F}_{12} = -\mathbf{F}_{21} \quad (1.7)$$

If the charges have the same sign, the forces point away from each other, if they have opposite signs, the forces point toward each other.

Notice that the Coulomb force is proportional to both charges. So if we double charge q_1 , the new Coulomb force is twice as large as well. This is illustrated in Fig. 1.7c. The Coulomb force is also inversely proportional to the distance squared, so if the distance r is doubled, the force decreases by a factor four. This is shown in Fig. 1.7d.

Note the similarity of the Coulomb force with Newton's law of gravity between two masses:

$$\mathbf{F}_{g,21} = \frac{Gm_1m_2}{r_{21}^2} \hat{\mathbf{r}}_{21}, \quad (1.8)$$

where the force is always attractive because $G, m_1, m_2 > 0$. There is a remarkable difference of many orders of magnitude in strength:

$$\frac{\mathbf{F}_{g,12}}{\mathbf{F}_{12}} = \frac{\frac{Gm_1m_2}{r_{21}^2} \hat{\mathbf{r}}_{12}}{\frac{kq_1q_2}{r_{21}^2} \hat{\mathbf{r}}_{21}}, \quad (1.9)$$

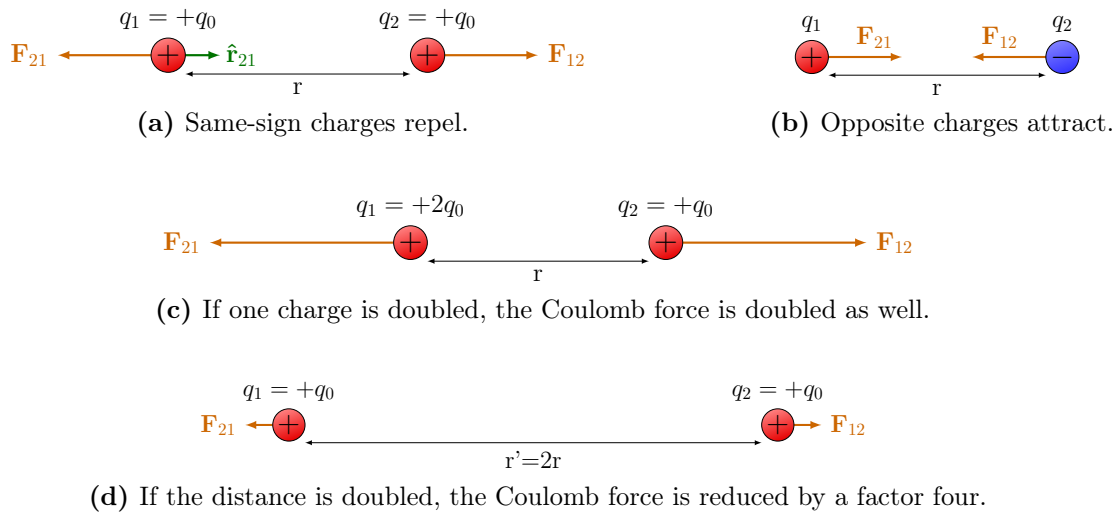


Figure 1.7: The Coulomb force on point charges is proportional to the charges and inversely proportional to the distance squared.

which for two protons is

$$\frac{\mathbf{F}_{g,21}}{\mathbf{F}_{21}} = 10^{-36}. \quad (1.10)$$

Example 1.1 (Multiple charges): Consider the configuration of three charges shown in Fig. 1.8. What is the total force \mathbf{F}_{tot} on q_3 due to q_1 and q_2 ? There are two forces, \mathbf{F}_{13} and \mathbf{F}_{23} . Notice that in two dimensions, we can split each force vector into its x and y components:

$$\mathbf{F}_{13} = \frac{kq_1q_3}{r_{13}^2} \hat{\mathbf{r}}_{13} = F_{x,13} \hat{\mathbf{x}} + F_{y,13} \hat{\mathbf{y}}$$

$$\mathbf{F}_{23} = \frac{kq_2q_3}{r_{23}^2} \hat{\mathbf{r}}_{23} = F_{x,23} \hat{\mathbf{x}} + F_{y,23} \hat{\mathbf{y}}.$$

The total force is then given by

$$\begin{aligned} \mathbf{F}_{\text{tot}} &= \mathbf{F}_{13} + \mathbf{F}_{23} \\ &= (F_{x,13} + F_{x,23}) \hat{\mathbf{x}} + (F_{y,13} + F_{y,23}) \hat{\mathbf{y}}. \end{aligned}$$

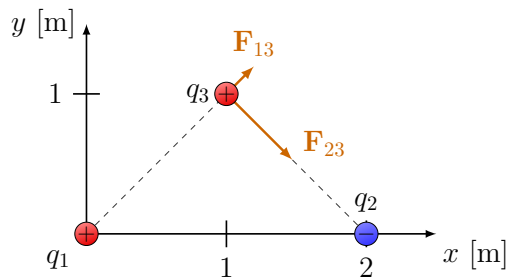
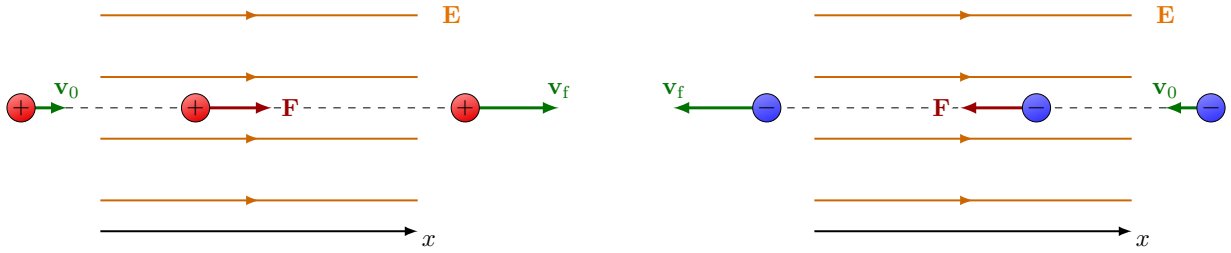
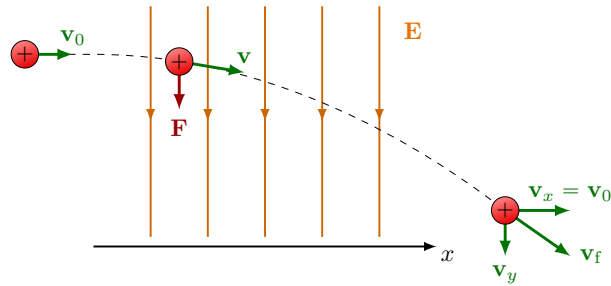


Figure 1.8: Configuration of multiple charges in a 2D plane.



(a) A positive charge moving along a uniform electric field.

(b) A negative charge moving along a uniform electric field.



(c) A positive charge crossing a uniform electric field.

Figure 1.9: Accelerated motion of a point charge in an electric field.

1.3 Electric fields

An electric charge produces an electric field around it. It is the field that exerts a force on other charges nearby. The force on a point charge q_0 in an electric field \mathbf{E} is given by

$$\mathbf{F}_E = q_0 \mathbf{E}. \quad (1.11)$$

The charge q_0 is often called a *test charge*. As per Newton's 2nd law, there will be a movement due to this force:

$$\sum \mathbf{F} = q_0 \mathbf{E} = m \mathbf{a}, \quad (1.12)$$

or,

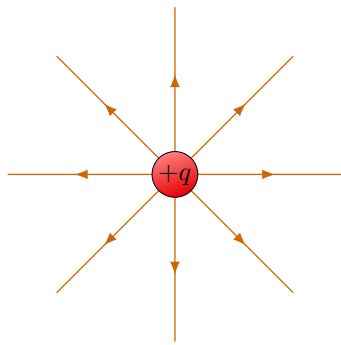
$$\mathbf{a} = \frac{q_0 \mathbf{E}}{m}. \quad (1.13)$$

Therefore, the acceleration is in the same direction as the electric field, as is shown in Fig. 1.9. Note, that the electric field points in the direction of acceleration of a positive charge $q_0 > 0$. For a negative charge, $q_0 < 0$, \mathbf{a} will be pointing in the opposite direction of \mathbf{E} . The normal laws of motion apply, as they would for a gravitational force. If the velocity gets too big, i.e. the velocity is close to the speed of light $c = 3 \times 10^8$ m/s, $v \gtrsim 0.1c$, then relativistic equations are needed.

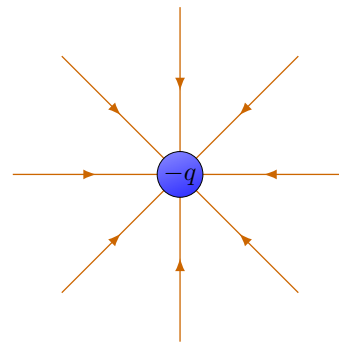
1.3.1 Electric field lines

Electric field lines indicate magnitude and direction of an electric field. Figures 1.10 and 1.11 show a examples for several different configurations of point charges. The density of electric field lines represent the strength of the electric field. In fact, the density is proportional the number of field lines in an area.

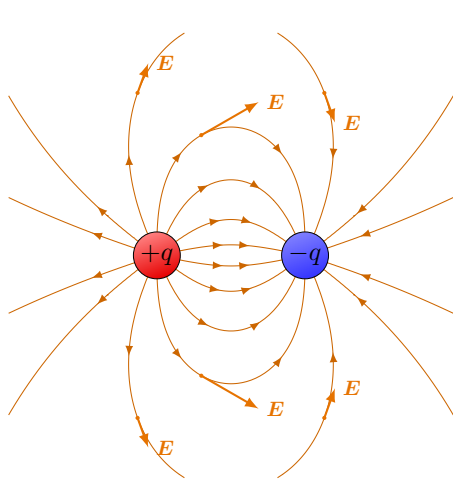
How do we draw electric field lines? There are several rules:



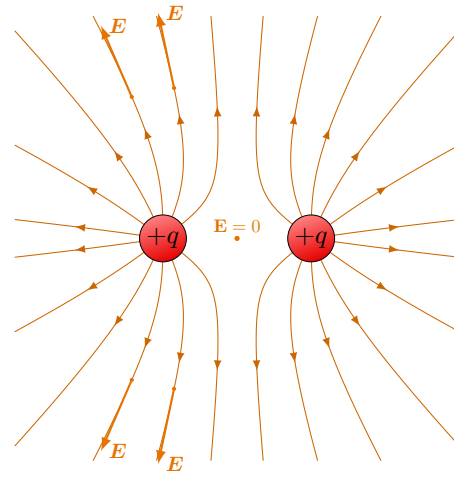
(a) Positive charge.



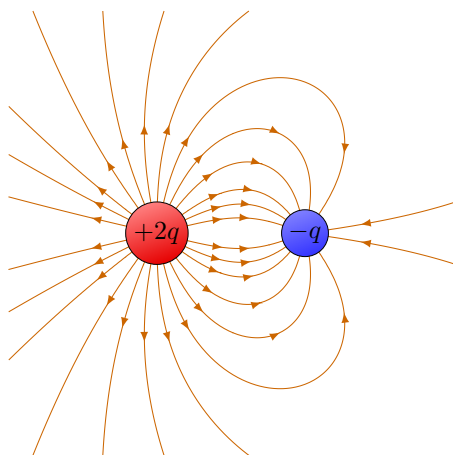
(b) Negative charge.

Figure 1.10: Point charges have radial electric field lines.

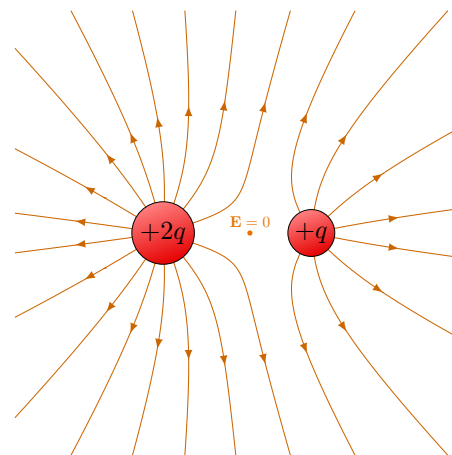
(a) Opposite point charges.



(b) Same-sign point charges.



(c) Opposite point charges, with one charge double the other.



(d) Same-sign point charges, with one charge double the other.

Figure 1.11: Electric field lines of two point charges. The total electric field is the vectorial sum of each charge's field. The electric field vectors \mathbf{E} are tangential to points on the field lines, and the higher the density of field lines, the stronger the field. At very large distance scales, the electric fields of (c) and (d) look like that of the single point charge in Fig. 1.10a.

- Lines go from + to – charges.
- Lines are symmetric when entering and leaving a point charge.
- The number of lines is proportional to the charge.
- The density of lines is proportional to the magnitude.
- At large distances, it appears as a single charge of the total charge.
- Two lines cannot cross.

1.4 Electric dipoles

A system of two opposite charges are separated by a distance. Then, the *electric dipole moment* is defined as

$$\mathbf{p} = q\mathbf{L}, \quad (1.14)$$

where \mathbf{L} is the separation vector from the negative to the positive charge, as in Fig. 1.12. What is the electric field produced by this dipole at some point x on the x axis? Because each charge creates its own electric field, we can simply add them vectorially as

$$\mathbf{E} = \frac{kq}{(x-a)^2}\hat{\mathbf{x}} + \frac{-kq}{(x+a)^2}\hat{\mathbf{x}}. \quad (1.15)$$

We do a bit of algebra to make the expression simpler:

$$\begin{aligned} \mathbf{E} &= kq \frac{(x+a)^2 - (x-a)^2}{(x-a)^2(x+a)^2} \hat{\mathbf{x}} \\ &= \frac{4akqx}{(x-a)^2(x+a)^2} \hat{\mathbf{x}}. \end{aligned}$$

Now, because $\mathbf{L} = 2a\hat{\mathbf{x}}$, we can use the definition of the dipole moment \mathbf{p} to write

$$\mathbf{E} = \frac{2kx\mathbf{p}}{(x-a)^2(x+a)^2}. \quad (1.16)$$

If we are far away, $x \gg a$, then these terms are small. The electric field of a dipole on its own axis and for distances much larger than the dipole size is

Electric field of a dipole (on its axis).

$$\mathbf{E} = \frac{2k\mathbf{p}}{x^3}. \quad (1.17)$$

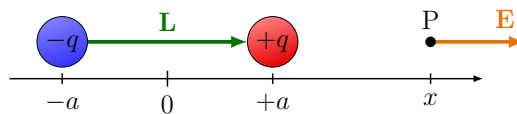


Figure 1.12: Electric dipole due to two point charges.

A practical example of a dipole is the water molecule. As is well known, water is H_2O , where each hydrogen shares an electron with the oxygen atom. As oxygen has eight protons pulls more strongly on the shared electrons, resulting in a net charge difference, often denoted with δ^\pm as in Fig. 1.13.

If a dipole is placed in an electric field, it will align as is shown in Fig. 1.14a. In a non-uniform field, the dipole will feel a force. Take for example a positive charge like in Fig. 1.14c. The electric field is stronger closer to the positive charge of the dipole than on the negative one, so there will be a net force.

Water is a permanent dipole, but dipole moments can also be induced with external electric fields. Examples are conductors in which electrons can freely redistribute in an electric field, such that negative charges (electrons) accumulate on one side, and positive ones on the other (protons without electrons to cancel the net charge). Another example are macroscopic objects that contain water. With a strong enough field, all the water dipoles will align, and the object effectively becomes a dipole.

Microwaves work on the fact that water is a dipole, and can be heated up by oscillating it in a changing electric field (i.e. the microwaves). If the dipole is not aligned with the external electric field, each charge will feel a force in the opposite direction. This means there will be a torque

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \quad (1.18)$$

with respect that rotates the dipole. Here we take \mathbf{r} from the negative to the positive charge. Using $\mathbf{F} = q\mathbf{E}$ and the fact that $\mathbf{r} = \mathbf{p}/q$,

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}. \quad (1.19)$$

The size of the torque is

$$\tau = pE \sin \theta, \quad (1.20)$$

where θ is the angle between the dipole and the electric field, like in Fig. 1.14b. The electric field rotated the dipole, so the field must have done work. The work is given by the potential energy

$$dU = -dW. \quad (1.21)$$

The work done to rotate a dipole by an angle $d\theta$ is given by

$$dW = -\tau d\theta, \quad (1.22)$$

$$= -pE \sin \theta d\theta. \quad (1.23)$$

Notice the sign is negative, because the torque makes the angle θ smaller. We find the potential energy U by integrating over the angle,

$$U = \int_0^\theta pE \sin \theta' d\theta' \quad (1.24)$$

$$= U_0 - pE \cos \theta. \quad (1.25)$$

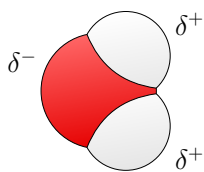
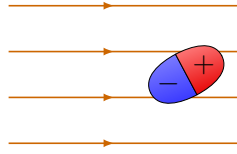
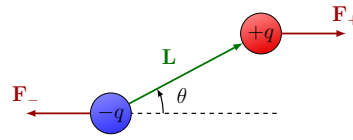


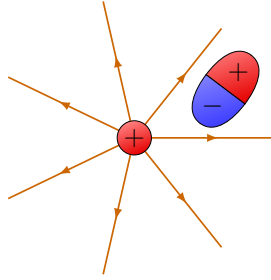
Figure 1.13: Water molecules (H_2O) are electric dipoles due to the spatial imbalance of electrons.



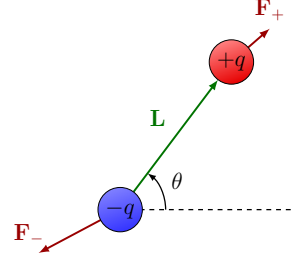
(a) Dipoles tend to align with uniform electric fields.



(b) Forces on either end of the dipole create a torque.



(c) A dipole in a non-uniform field tends to experience a net force.



(d) Different forces pull on either end of the dipole, creating a net force.

Figure 1.14: Dipoles in an electric field experience a force.

U_0 here is the potential energy for $\theta = 0$, which we can choose to set to zero. Recognizing the scalar product, we rewrite (1.25) to find

Potential energy of a dipole in an external electric field.

$$U = -\mathbf{p} \cdot \mathbf{E}. \quad (1.26)$$

1.5 Charge distributions

Charge can be distributed continuously over a volume V with charge density

$$\rho = \frac{\Delta q}{\Delta V} \quad (1.27)$$

In general, the charge distribution depends on the position, $\rho = \rho(\mathbf{r})$, and each small volume $\Delta V(\mathbf{r})$ has a charge $\Delta q(\mathbf{r})$. Taking the infinitesimal limit $\Delta V \rightarrow 0$,

$$\rho = \frac{dq}{dV} \left[\frac{\text{C}}{\text{m}^3} \right]. \quad (1.28)$$

Similarly, a surface with area A can have a charge density

$$\sigma = \frac{dq}{dA} \left[\frac{\text{C}}{\text{m}^2} \right]. \quad (1.29)$$

or, a length L has a linear density

$$\lambda = \frac{dq}{dL} \left[\frac{\text{C}}{\text{m}} \right]. \quad (1.30)$$

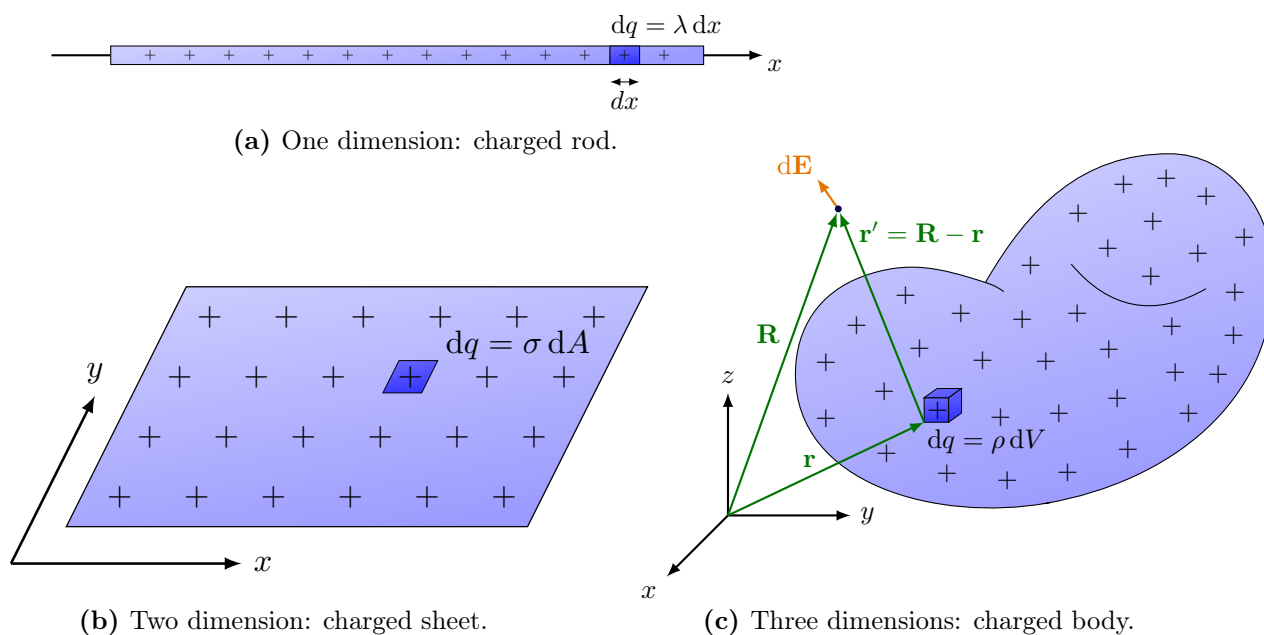


Figure 1.15: Charge distributions in different dimensions.

Figure 1.15 illustrates dq in different number of dimension. Say you have some three-dimensional body that has some charge distribution. Then the total charge is given by

$$Q = \int_V dq, \quad (1.31)$$

where the integrals runs over the infinitesimal charge dq of the whole volume V of the body. If the charge is distributed uniformly in three dimensions, $\rho(\mathbf{r})$ is constant in space, namely

$$\rho = \frac{Q}{V}. \quad (1.32)$$

The electric field at some point due to a continuous charge distribution in a volume V is

$$\mathbf{E} = \int_V \frac{k dq}{r^2} \hat{\mathbf{r}}. \quad (1.33)$$

Say you want to know the electric field due to a charged body in some point \mathbf{R} from the origin, like in Fig. 1.15c. Then one has to integrate over dq at a position \mathbf{r} from the origin. The distance between dq and the point is $\mathbf{r}' = \mathbf{R} - \mathbf{r}$, and therefore we can write

$$d\mathbf{E} = \frac{k dq}{r'^2} \hat{\mathbf{r}}', \quad (1.34)$$

where the denominator give the magnitude of the distance vector, and the unit vector

$$\begin{aligned} \hat{\mathbf{r}}' &= \frac{\mathbf{r}'}{|\mathbf{r}'|} = \frac{\mathbf{R} - \mathbf{r}}{|\mathbf{R} - \mathbf{r}|} \\ &= \frac{\mathbf{r}'}{\sqrt{\mathbf{r}'^2}} \end{aligned}$$

The integral becomes

$$\mathbf{E} = \int_V k \frac{\mathbf{r}'}{|\mathbf{r}'|^3} dq. \quad (1.35)$$

Often, we take advantage of symmetries to simplify the integrals, by choosing the right coordinate system. We will see some examples in the next sections with rods and sheets.

1.5.1 Electric field of a charged rod

Electric field of a charged rod along the x axis.

Let's consider the electric field for a uniformly charged rod (or equivalently, a charged wire, or a line of charge). We choose the x axis along the rod, with the origin at the left end of the rod, as in Fig. 1.16a. What is the electric field in a point x_0 on the x axis? First we notice that the electric field has no y component on the x axis, $\mathbf{E} = \mathbf{E}_x$. We need to integrate

$$\begin{aligned}\mathbf{E} &= k \int_0^L \frac{\lambda dx}{(x_0 - x)^2} \hat{\mathbf{x}} \\ &= k\lambda \left[\frac{1}{x_0 - L} - \frac{1}{x_0} \right] \hat{\mathbf{x}}.\end{aligned}$$

Using the fact that the total charge $Q = \lambda L$, gives us the result

$$\mathbf{E} = \frac{k\lambda L}{x_0(x_0 - L)} \hat{\mathbf{x}}. \quad (1.36)$$

So at large distances $x \gg L$ (or for very small L), the electric field looks like that of a point charge,

$$\mathbf{E}(x_0) = \frac{kQ}{x_0^2} \hat{\mathbf{x}}. \quad (1.37)$$

This is as we would expect, because when we zoom out, we cannot distinguish the length of the rod any more, and we only see the total charge.

Electric field of a charged rod along the y axis.

Similarly, what is the electric field in a point y_0 on the y axis, just above the middle of the rod? For this calculation, it's easier to choose the origin in the middle, as in Fig. 1.16b. Here, we simplify a lot of the calculations by considering the symmetries: Charges at $+x$ produce an electric field dE_x that is equal and opposite to that of the charges at $-x$. Therefore,

$$\int dE_x = 0, \quad (1.38)$$

And we see that $\mathbf{E} = \mathbf{E}_y$. In the y direction, we have to integrate over all the contributions

$$E = E_y = \int_{-\frac{L}{2}}^{\frac{L}{2}} dE_y \quad (1.39)$$

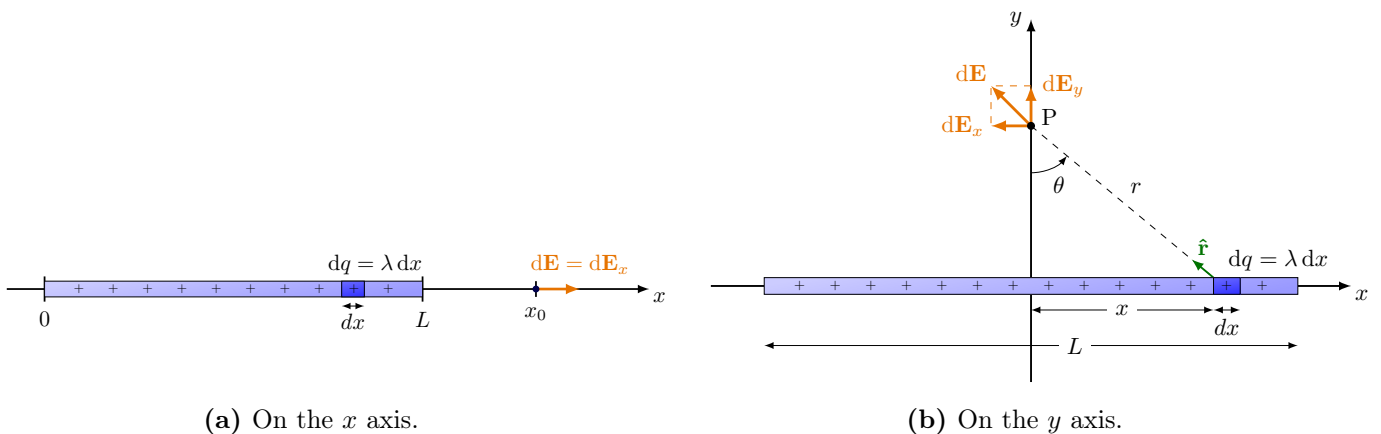


Figure 1.16: Calculation of the electric field due to a uniformly charged rod.

The contributions from $-x$ charges add up to those from $+x$ charges, so by symmetry,

$$E = 2 \int_0^{\frac{L}{2}} dE_y. \quad (1.40)$$

Because $dq = \lambda dx$ and $\hat{\mathbf{r}} = \sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{y}}$,

$$E = 2 \int_0^{\frac{L}{2}} \frac{k\lambda}{r^2} \cos \theta dx. \quad (1.41)$$

Note that θ and r vary with x , so we need some trigonometry. From the triangle in Fig. 1.16b:

$$x = y \tan \theta. \quad (1.42)$$

Differentiating this with respect to θ yields

$$\frac{dx}{d\theta} = y \sec^2 \theta = y \left(\frac{r}{y} \right)^2. \quad (1.43)$$

(Remember secant is the inverse of cosine.) So,

$$dx = \frac{r^2}{y} d\theta. \quad (1.44)$$

Substituting in Eq. (1.41):

$$E = \left[\frac{2k\lambda}{y} \sin \theta \right]_0^{\theta_{\max}}, \quad (1.45)$$

where the integral boundary $\theta = 0$ corresponds to $x = 0$, and $\theta = \theta_{\max}$ to $x = L/2$. Looking at the triangle in Fig. 1.16b again, we see that

$$\sin \theta_{\max} = \frac{L/2}{\sqrt{(L/2)^2 + y^2}} \quad (1.46)$$

In terms of L , Eq. (1.45) becomes

$$E(y) = \frac{2k\lambda}{y} \frac{L/2}{\sqrt{(L/2)^2 + y^2}}, \quad (1.47)$$

or as a vector,

$$\mathbf{E} = \frac{2k\lambda}{y} \frac{L/2}{\sqrt{(L/2)^2 + y^2}} \hat{\mathbf{y}}. \quad (1.48)$$

Close to the rod, $y \ll L$, and we approximate:

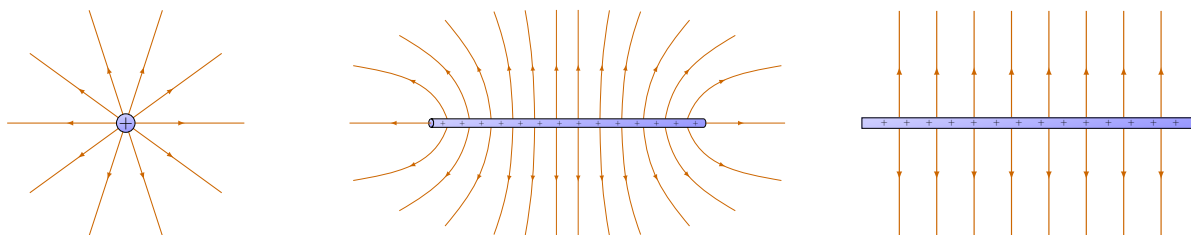
$$\mathbf{E} = \frac{2k\lambda}{y} \hat{\mathbf{y}}. \quad (1.49)$$

We see therefore, that $E_y \sim 1/y$, which is different from the the inverse quadratic behavior we saw before. Also, we see that it depends on the charge density rather than the total charge Q or length L , individually. This equation holds only in the middle of the rod: As you move closer to the edges, an x component appears, as is illustrated in Fig. 1.17b. On the other hand, if the rod is infinitely long, $L \rightarrow \infty$, Eq. (1.49) holds everywhere along the rod, like in Fig. 1.17c.

Instead, if we were far from the charged rod, $y \gg L$, then the approximation instead becomes:

$$\mathbf{E} = \frac{kQ}{y^2} \hat{\mathbf{y}} \quad (1.50)$$

Here, we see that the electric field now depends on the total charge Q , and $E \sim 1/y^2$, as is the case for a point charge.

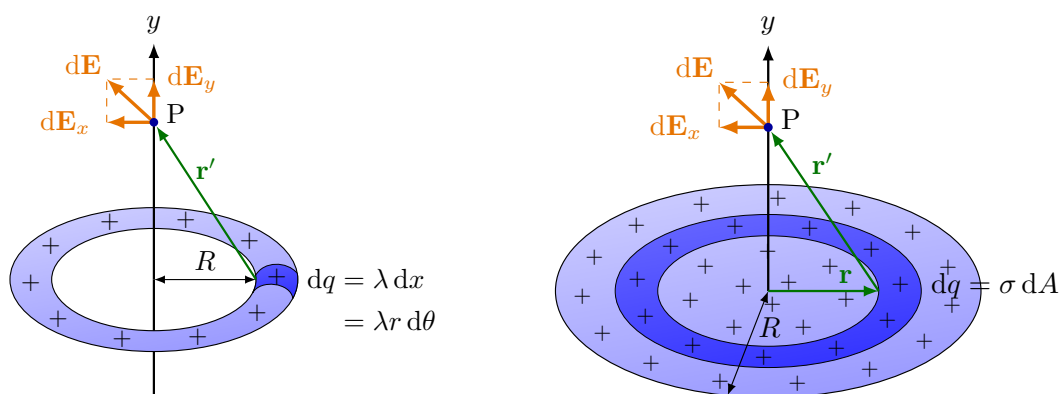


(a) Cross-sectional. Notice the field is radially symmetric in the transverse plane.

(b) Sideview for finite L . Equation (1.49) holds only in the middle and far from the rod ($y \gg L$).

(c) Sideview for an infinitely long rod ($L \rightarrow \infty$). Equation (1.49) holds everywhere.

Figure 1.17: Field lines of a charged rod with length L .



(a) On the axis of a uniformly charged ring.

(b) On the axis of a uniformly charged sheet.

Figure 1.18: Calculation of the electric fields by integration over dq .

1.5.2 Electric field of a charged ring

Consider a uniformly charged, one-dimensional ring with radius R that has a total charge Q , like in Fig. 1.18a. What is \mathbf{E} along its axis? By symmetry, only the y component survives, and the solution is

$$\mathbf{E} = \frac{kQy}{(y^2 + R^2)^{\frac{3}{2}}} \hat{\mathbf{y}}, \quad (1.51)$$

which is left as a homework assignment.

1.5.3 Electric field of a charged sheet

Now we consider the electric field due to a large, round sheet that has radius R and is uniformly charged with total charge Q . The charge density is

$$\sigma = \frac{Q}{\pi R^2}. \quad (1.52)$$

Because the sheet is round, it's easier to integrate over infinitesimal rings with radius r and width dr , like in Fig. 1.18b. One such ring has an area

$$dA = 2\pi r dr, \quad (1.53)$$

We can now use our previous result Eq. (1.51) to obtain the field element

$$\begin{aligned} dE &= \frac{kydq}{(y^2 + r^2)^{\frac{3}{2}}} \\ &= \frac{2\pi ky\sigma r dr}{(y^2 + r^2)^{\frac{3}{2}}}. \end{aligned}$$

After integrating over the radius r , we find

$$E = 2\pi k\sigma \left(1 - \frac{y}{\sqrt{y^2 + R^2}} \right). \quad (1.54)$$

For large distances $y \gg R$, we can approximate this using a binomial approximation

$$(1 - \epsilon)^n \approx 1 + n\epsilon \quad (1.55)$$

for $\epsilon \rightarrow 0$. We find

$$E \sim \frac{kQ}{y^2}, \quad (1.56)$$

which looks like a point charge. In other words, from very far away, one cannot see the sheet's actual shape, and it only appears a single charge. On the other hand, for $y \ll R$,

$$E = 2\pi k\sigma, \quad (1.57)$$

which is constant in both the y direction. This is exactly the field of an infinite plane with charge density σ , which also has no x component ($E_x = 0$).

Chapter 2

Gauss's Law

2.1 Electric flux

The calculations in the previous section involved a lot of integrals. There is an easier way to solve all these problems.

Let's imagine a spherical surface around a point charge as in Fig. 2.1a. We use $\hat{\mathbf{n}}$ to indicate the *normal vector* that is the unit vector perpendicular to the surface S , and points radially outward. Due to spherical symmetry of the field, we note that $\mathbf{E} \parallel \hat{\mathbf{n}}$ everywhere.

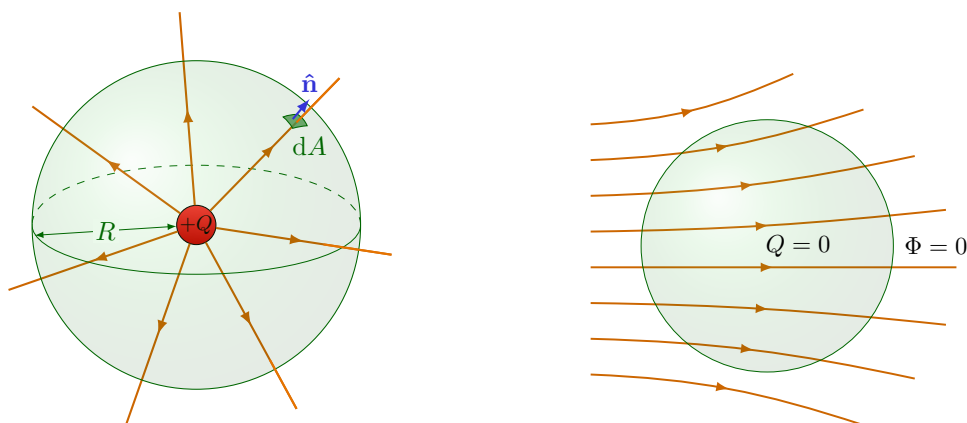
For a simple surface, we define the *electric flux* as simply

$$\Phi = EA. \quad (2.1)$$

This is a measure of how many field lines pass through a surface. The previous equation assumes the electric field is normal to the surface everywhere. If the electric field makes some angle with the surface, this becomes

$$\Phi = \mathbf{E} \cdot \hat{\mathbf{n}}A = EA \cos \theta, \quad (2.2)$$

where $\hat{\mathbf{n}}$ is the normal unit vector to the surface A , and θ is constant everywhere, like in Fig. 2.2. Even more generally, if the electric field is not exactly parallel to $\hat{\mathbf{n}}$ everywhere



(a) Spherical surface around a point charge. The normal vector $\hat{\mathbf{n}}$ stands perpendicular to the small surface segment dA .

(b) The total flux Φ_{net} is zero if the same number of field lines enter as leave the surface.

Figure 2.1: Flux of electric field lines.

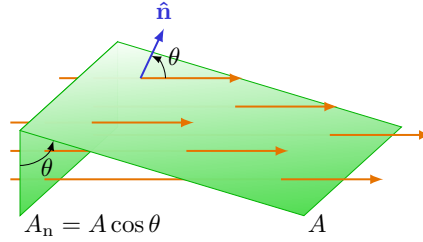


Figure 2.2: Flux of electric field lines through planes at different angles.

on a surface S , then we have to integrate over the whole surface

$$\Phi = \int_S \mathbf{E} \cdot \hat{\mathbf{n}} \, dA. \quad (2.3)$$

Some text books use the notation $d\mathbf{A} = \hat{\mathbf{n}} \, dA$, such that

$$\Phi = \int_S \mathbf{E} \cdot d\mathbf{A}. \quad (2.4)$$

Projecting the electric field \mathbf{E} onto $\hat{\mathbf{n}}$, we write $E_n = \mathbf{E} \cdot \hat{\mathbf{n}} = E \cos \theta$ with the angle θ between \mathbf{E} and $\hat{\mathbf{n}}$, and Eq. (2.3) becomes

$$\Phi = \int_S E_n \, dA. \quad (2.5)$$

Like in the case above, if $\mathbf{E} \parallel \hat{\mathbf{n}}$, then $\cos \theta = 1$. If E_n is also constant of the whole surface, the last equation reduces to

$$\Phi = E_n \int_S dA. \quad (2.6)$$

For a closed surface S , we use the notation

$$\Phi_{\text{net}} = \oint_S \mathbf{E} \cdot \hat{\mathbf{n}} \, dA \quad (2.7)$$

If more lines leave than enter the closed surface S , the total, or “net” flux Φ_{net} is positive, like in Fig. 2.1a. Similarly, if more field lines enter S , Φ_{net} is negative, and if the same number of line enter as leave S , the Φ_{net} is zero like in Fig. 2.1b.

2.1.1 Interlude: Surface integral and solid angle

The simple surface integral

$$A = \oint_S dA \quad (2.8)$$

is the area of the closed surface S . Often it is most convenient to make use of the spherical coordinate system shown in Fig. 2.3. At a fixed radius r , the integration is done over an infinitesimal surface element

$$dA = r^2 \sin \theta \, d\theta \, d\phi \quad (2.9)$$

Here, ϕ is the azimuthal angle between 0 and 2π from the x axis in the xy plane, polar angle θ between 0 and π from the z axis. Often, this is written in terms of the *solid angle* $d\Omega = \sin \theta \, d\theta \, d\phi$:

$$dA = r^2 \, d\Omega. \quad (2.10)$$

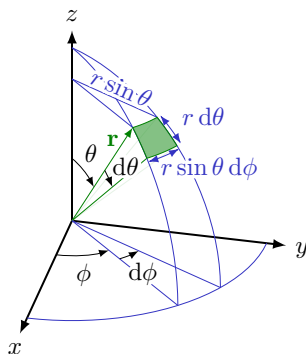


Figure 2.3: Integration over a spherical area with spherical coordinates.

Note that $d\Omega$ has units of radians squared, also called steradians. The integral over the whole ϕ and θ range yields

$$\Omega = \int d\Omega = \int_0^{2\pi} \int_0^\pi \sin \theta \, d\theta \, d\phi = 4\pi. \quad (2.11)$$

We will see later that 4π appears in several places. For a sphere with radius R then,

$$A = \oint_S dA = R^2 \int_0^{2\pi} \int_0^\pi \sin \theta \, d\theta \, d\phi = 4\pi R^2 \quad (2.12)$$

2.2 Gauss's law

For a point charge Q we can now find the flux using Eq. (2.6) and Eq. (2.12):

$$\begin{aligned} \Phi &= \frac{kQ}{R^2} \oint_S dA \\ &= \frac{kQ}{R^2} 4\pi R^2 \\ &= 4\pi kQ. \end{aligned} \quad (2.13)$$

The derivation of this result was simple enough, but it turns out that this relation between the flux and charge inside the surface holds for more general surfaces, and is given by

Gauss's law. *The electric flux through a closed surface S is given by the total charge Q_{in} inside S :*

$$\Phi = \oint_S \mathbf{E} \cdot \hat{\mathbf{n}} \, dA = 4\pi k Q_{in} = \frac{Q_{in}}{\epsilon_0}. \quad (2.14)$$

Here, ϵ_0 is the *vacuum electric permittivity*, which is often used instead of Coulomb's constant k ,

$$k = \frac{1}{4\pi\epsilon_0}. \quad (2.15)$$

The closed surface is sometimes called a *Gaussian surface*. In principle, Gauss's law will hold for any arbitrary, closed surface, whether it be a sphere, cylinder, cube, or any fantastic shape you can think of. However, to simplify calculations, one has to choose it intelligently by exploiting symmetries. We will see some examples below.

Notice that if the charge inside the surface, Q_{in} , is zero, the flux is zero as well, just like in Fig. 2.1b.

2.2.1 Trivial example: Single point charge

Let's look at a simple example of how to use Gauss's law. Say you have a test charge q_0 in the electric field a single point charge Q . It experiences a radial force $\mathbf{F} = q_0\mathbf{E}$ due to Q . Given the force $F = |\mathbf{F}|$ measured on q_0 , what is Q ? We consider again a spherical surface of radius R around Q as in Fig. 2.1, and find using Eq. (2.14),

$$\Phi = E(4\pi R^2) = 4\pi kQ, \quad (2.16)$$

because $\mathbf{E} \cdot \hat{\mathbf{n}} = E$ is constant over the whole sphere as in Eq. (2.6). Therefore,

$$Q = \frac{ER^2}{k} = \frac{FR^2}{q_0k}. \quad (2.17)$$

2.2.2 Electric field of a charged sphere

Another simple example is a solid sphere with uniform charge density $\rho = Q/V$ and radius R , like in Fig. 2.4. Consider a spherical gaussian surface of radius $r > R$ around the charge. Because of spherical symmetry, again the electric field is parallel to the normal vector and constant everywhere on this sphere. Therefore,

$$\Phi_{\text{net}} = E \oint_S dA = E(4\pi r^2), \quad (2.18)$$

and the electric field according to Gauss's law is

$$E = \frac{kQ}{r^2} \quad (2.19)$$

in the radial direction.

This was much faster and easier than finding a convenient coordinate system and integrating over the whole charge distribution of sphere! All we needed was an argument from symmetry and Gauss's law.

Notice that this result is the same as for a point charge. Also notice that we could have made the exact same argument for the electric field outside a hollow spherical surface with area A and charge density $\sigma = Q/A$, like in Fig. 2.5, and we would have found the exact same result (2.19).

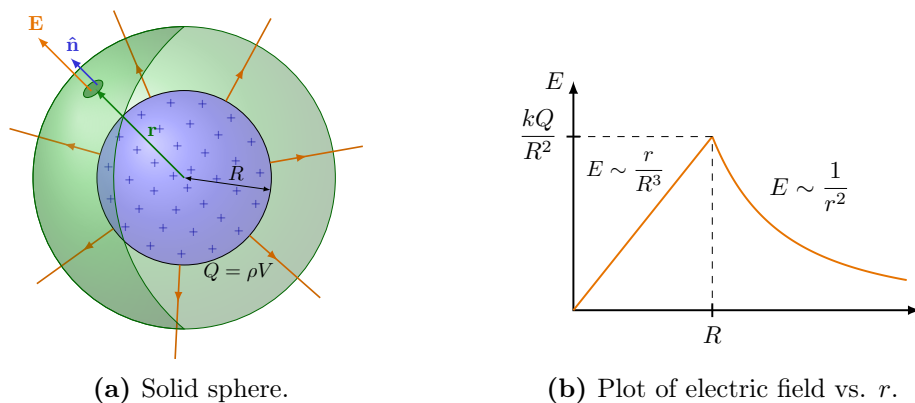


Figure 2.4: Applying Gauss's law to an uniformly charged, solid sphere with radius R and charge density $\rho = Q/V$.

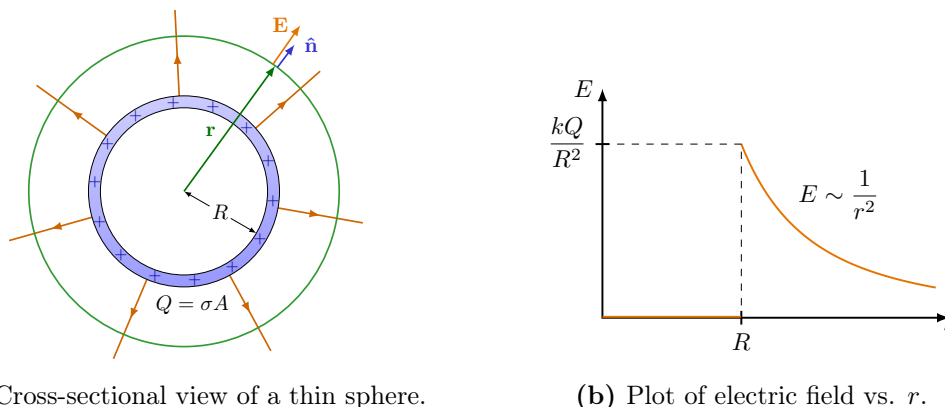


Figure 2.5: Applying Gauss's law to an uniformly charged, thin, hollow sphere with radius R and charge density $\sigma = Q/A$.

What about inside the solid sphere? If $r < R$, the volume of the gaussian sphere is $V_{\text{in}} = 4\pi r^3/3$, and therefore the enclosed charge is

$$Q_{\text{in}} = \rho V_{\text{in}} = Q \frac{V_{\text{in}}}{V} = Q \frac{r^3}{R^3}. \quad (2.20)$$

Using Eq. (2.18) and Gauss's law again,

$$E = \frac{kQr}{R^3}. \quad (2.21)$$

in the radial direction and for $r < R$. This time, the field is proportional to the distance r from the center. The total picture is plotted in Fig. 2.4b.

What is the electric field inside a hollow sphere?

2.2.3 Electric field of a charged rod, revisited

Let's consider the electric field of an infinitely long charged rod again, as we did in Section 1.5.1. Instead of using integrals and complicated coordinates, let's put Gauss's law to the test. This will turn out to be much easier. The trick is to exploit the symmetry of the rod, which is cylindrical. Therefore, we choose a cylindrical surface with length L and the rod as its axis, like in Fig. 2.6a.

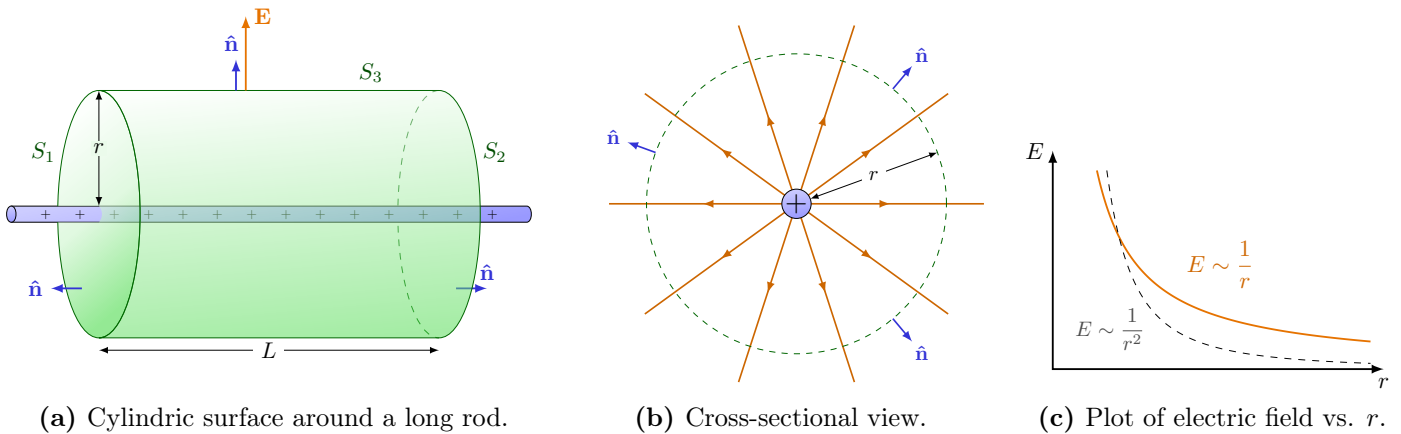
$$\Phi_{\text{net}} = \oint_S \mathbf{E} \cdot \hat{\mathbf{n}} \, dA = 4\pi k Q_{\text{in}}. \quad (2.22)$$

The closed cylindrical surface $S = S_1 + S_2 + S_3$ consist of three pieces: two "endcaps" S_1 and S_2 and the side S_3 . Therefore the surface integral can be split into the sum

$$\Phi_{\text{net}} = \int_{S_1} \mathbf{E} \cdot \hat{\mathbf{n}} \, dA + \int_{S_2} \mathbf{E} \cdot \hat{\mathbf{n}} \, dA + \int_{S_3} \mathbf{E} \cdot \hat{\mathbf{n}} \, dA \quad (2.23)$$

The electric field points radially outward from the rod and is therefore parallel to the endcaps S_2 and S_2 ($\mathbf{E} \perp \hat{\mathbf{n}}$), therefore the flux through S_1 and S_2 is zero. Meanwhile, the electric field is always perpendicular to the the side S_3 ($\mathbf{E} \parallel \hat{\mathbf{n}}$), and $E = |\mathbf{E}|$ is constant everywhere on S_3 . Therefore, the integral becomes

$$\Phi_{\text{net}} = 0 + 0 + E \int_{S_3} dA. \quad (2.24)$$

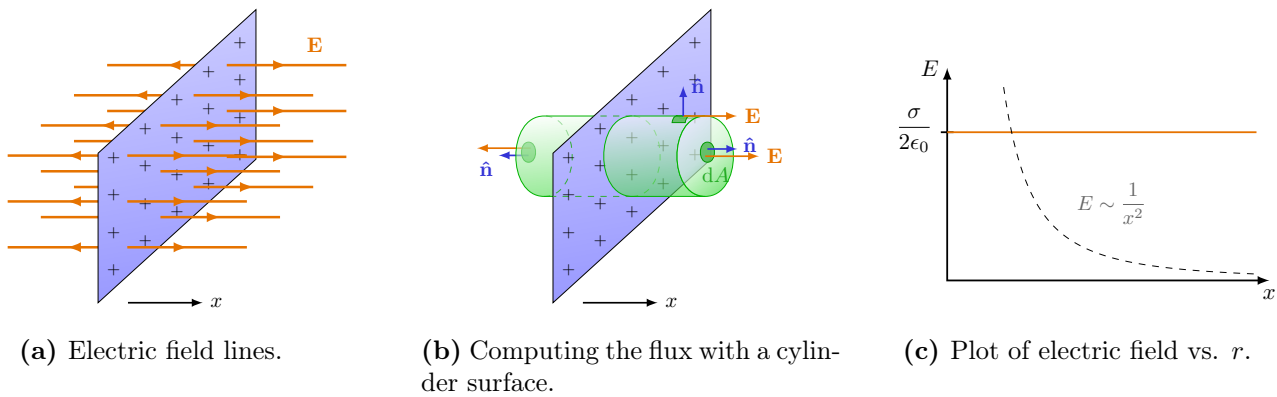


(a) Cylindric surface around a long rod.

(b) Cross-sectional view.

(c) Plot of electric field vs. r .

Figure 2.6: Applying Gauss's law to an infinitely long charged rod, using a cylinder as a Gaussian surface.



(a) Electric field lines.

(b) Computing the flux with a cylinder surface.

(c) Plot of electric field vs. r .

Figure 2.7: Applying Gauss's law to an infinitely large charged sheet, using a cylinder as a Gaussian surface to compute the flux.

The remaining integral is simply the total area of S_3 , so we find:

$$\Phi_{\text{net}} = E(2\pi rL) = 4\pi kQ_{\text{in}}. \quad (2.25)$$

And we retrieve our previous result, Eq. (1.49) for a charge density $\lambda = Q/L$:

$$\mathbf{E} = \frac{2k\lambda}{r} \hat{\mathbf{r}}. \quad (2.26)$$

This derivation was a lot simpler than the one in Section 1.5.1. We only had to know the field's angle to our intelligently chosen surface, exploiting the problem's symmetry, and we could then apply Gauss's law.

Note, however, this works because we assume L is large compared to r . By “large” here, we often mean “infinite”. In real-life, a rod will typically have a more complicated field configuration at the edges where the field is not perpendicular anymore to our simple cylindrical surface, like in Fig. 1.17b. Still, our results will be a good approximation if the rod is infinitely long, or if we stay close to the middle, and not too far removed from the rod.

2.2.4 Electric field of a charged sheet, revisited

Now let us revisit the electric field near a large plane with uniform charge, as in Section 1.5.3. Again, let's use a cylindrical surface. We draw it through and perpendicular

to the plane as in Fig. 2.7b. The surface is again made out of three surfaces: two “endcaps” S_1 and S_2 and the side S_3 . This time, the electric field is parallel to the side, but perpendicular to both endcaps. So:

$$\begin{aligned}\Phi_{\text{net}} &= \int_{S_1} \mathbf{E} \cdot \hat{\mathbf{n}} \, dA + \int_{S_2} \mathbf{E} \cdot \hat{\mathbf{n}} \, dA + \int_{S_3} \mathbf{E} \cdot \hat{\mathbf{n}} \, dA \\ &= EA + EA + 0,\end{aligned}\tag{2.27}$$

where A is the area of S_1 and also S_2 . Gauss’s law implies

$$2EA = 4\pi k Q_{\text{in}},\tag{2.28}$$

or, using the sheet’s charge density $\sigma = Q_{\text{in}}/A$,

$$\mathbf{E} = 2\pi k\sigma = \frac{\sigma}{2\epsilon_0} \hat{\mathbf{x}}.\tag{2.29}$$

Notice that thanks to the use of the charge density, any trace (like the area A) of the our arbitrarily chosen surface has vanished. This is what we want, because the result should not depend on the choice of the surface. We used a cylinder, but a cubic surface would have worked conveniently as well in this case, as the flux through its sides is zero.

2.2.5 Electric field of a charged slab

Now let’s look at a block, or slab, of uniform charge. This is the case where our sheet in Fig. 2.7 has some thickness t . By the exact same argument as in the previous section, we can use a cylindrical surface with $\Phi = 2EA$. This time, the charge inside the cylinder is given by charge volume-density and volume V enclosed by the cylinder:

$$Q_{\text{in}} = \rho V.\tag{2.30}$$

Because the volume is actually $V = At$, we can write the charge inside, $\rho = Q_{\text{in}}/At$. At the same time, the uniform charge volume-density can be expressed by the surface density as $\rho = \sigma/t$. Therefore:

$$Q_{\text{in}} = \left(\frac{\sigma}{t}\right)(At) = \sigma A\tag{2.31}$$

Applying Gauss’s law, we find

$$\mathbf{E} = \frac{\sigma}{2\epsilon_0} \hat{\mathbf{x}},\tag{2.32}$$

on the left side, the exact same as for a charged sheet, Eq. (1.56).

2.3 Interlude: Divergence theorem

Remember that the *gradient* vector is defined by

$$\nabla = \frac{\partial}{\partial x} \hat{\mathbf{x}} + \frac{\partial}{\partial y} \hat{\mathbf{y}} + \frac{\partial}{\partial z} \hat{\mathbf{z}}.\tag{2.33}$$

This is the sometimes called the *del operator*, or also the *nabla operator*. It is applied to vector fields, i.e. vectors that are a function of space like the electric field $\mathbf{E}(\mathbf{r})$. Take for example, the vector field

$$\mathbf{v} = v_x \hat{\mathbf{x}} + v_y \hat{\mathbf{y}} + v_z \hat{\mathbf{z}}.\tag{2.34}$$

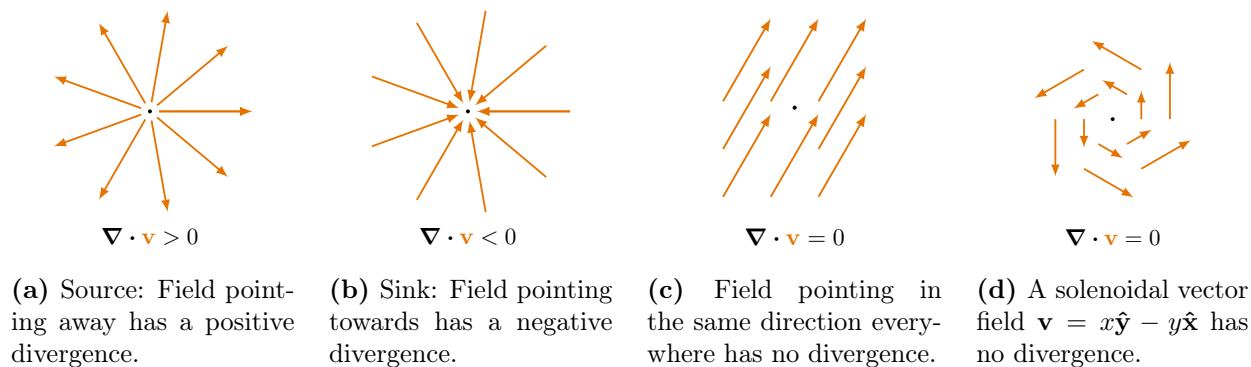


Figure 2.8: Intuitive examples of divergences in a point of a simple vector fields \mathbf{v} .

Here, \mathbf{v} is a function of x , y and z , and has components v_x , v_y and v_z , also functions of x , y and z . The *divergence* of \mathbf{v} then, is the dot product of del with \mathbf{v}

$$\nabla \cdot \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}. \quad (2.35)$$

The intuitive meaning of the divergence is shown in Fig. 2.8. If the divergence is positive, the vector field tends to point away, while if it is negative, the vector field points toward it. A point with positive divergence is sometimes called a *source*, while points with a negative one is called a *sink*.

In electromagnetism, the following theorem from calculus is very useful:

Fundamental theorem of divergence (Gauss's theorem). *The flux of a vector field \mathbf{v} through a closed surface S is related to the divergence of the field in the enclosed volume V as*

$$\int_V (\nabla \cdot \mathbf{v}) dV = \oint_S \mathbf{v} \cdot d\mathbf{A}. \quad (2.36)$$

This topic is covered in more detail in another course.

2.4 Differential form of Gauss's law

The form of the closed-surface integral on the right hand side of divergence theorem Eq. (2.36) looks similar to the one in Gauss's law Eq. (2.14). All we have to do is identify $\mathbf{v} = \mathbf{E}$, and we can write

$$\int_V (\nabla \cdot \mathbf{E}) dV = \oint_S \mathbf{E} \cdot \hat{\mathbf{n}} dA. \quad (2.37)$$

Gauss's law implies that

$$\int_V (\nabla \cdot \mathbf{E}) dV = \frac{Q}{\epsilon_0}, \quad (2.38)$$

where Q is the enclosed charge. Remember the definition of charge volume-density ρ in Eq. (1.28), which implies

$$Q = \rho \int_V dV. \quad (2.39)$$

We have derived a very nice differential formulation of Gauss's law:

Gauss's law (differential form).

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \quad (2.40)$$

This is the first of *Maxwell's four equations*, which describe all of electromagnetism. We will see more of them later.

Notice that if the charge density is positive ($\rho > 0$), the divergence is as well, meaning the electric field points away from the positive charge, like in Fig. 2.8a. Similarly for negative charge ($\rho < 0$), the field has a negative divergence, and points towards the charge, like in Fig. 2.8b. If there is no charge ($\rho = 0$), the divergence of the electric field is also zero,

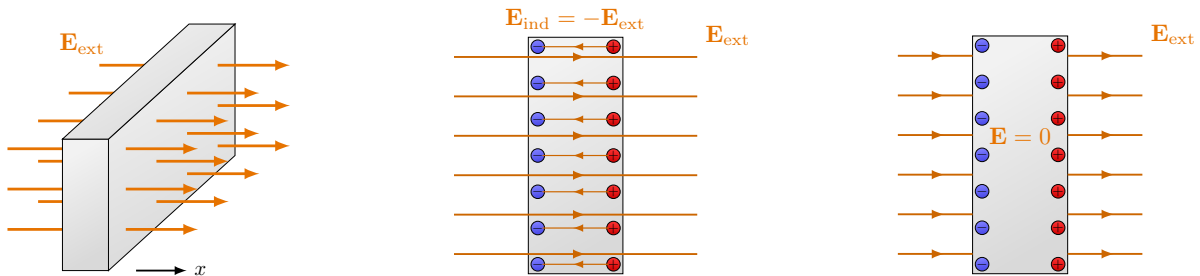
$$\nabla \cdot \mathbf{E} = 0. \quad (2.41)$$

2.5 Charge distribution in conductors

In a conductor, some electrons are free to move. Given an external field \mathbf{E} , they will feel a force $\mathbf{F} = q\mathbf{E}$. The electrons will move under this force until it is balanced. The negatively charged electrons in a conductor that move away, leave positively charged ions behind. They tend to redistribute in the conductor, so they are maximally separated on opposite sides of the conductor, like in Fig. 2.9. The new charge distribution will cause an induced field in the conductor that opposes the external field. At equilibrium, there are no forces on electrons, so the induced field exactly opposes the external field:

$$\mathbf{E}_{\text{ind}} = -\mathbf{E}_{\text{ext}}. \quad (2.42)$$

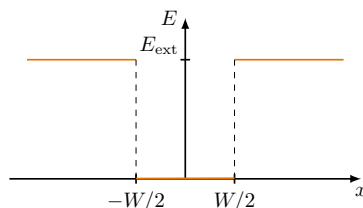
This implies that the electric field *inside* a conductor at equilibrium is always zero! This redistribution of charge and cancellation of fields happens almost instantaneously.



(a) An external field \mathbf{E}_{ext} is applied to a slab of conducting metal.

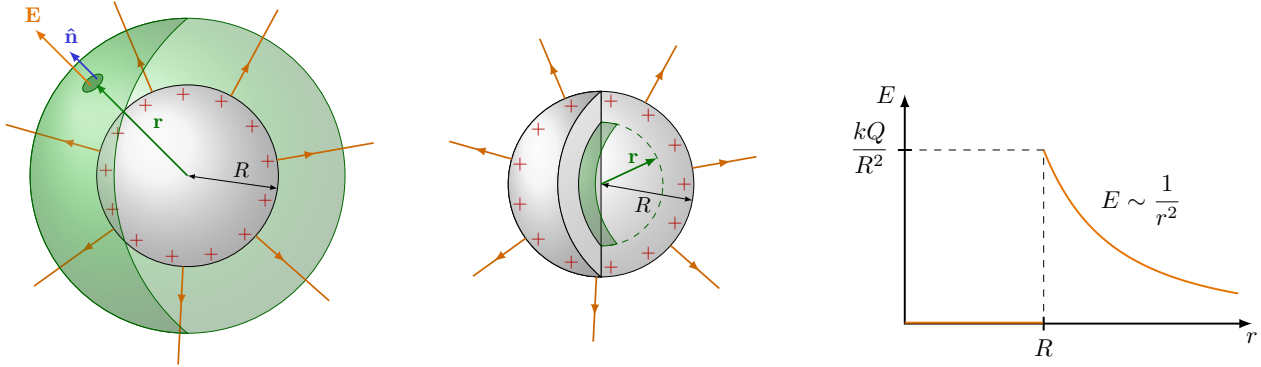
(b) The free charges redistribute and create an internal field \mathbf{E}_{ind} .

(c) Once the charge has settled, the internal and external fields cancel.



(d) Plot of electric field vs. r .

Figure 2.9: Conductor in an external field \mathbf{E}_{ext} . The free charges in the conductor (typically electrons) instantly redistribute in \mathbf{E}_{ext} to create their own induced field \mathbf{E}_{ind} that opposes \mathbf{E}_{ext} .



(a) Computing the flux outside. (b) Computing the flux inside. (c) Plot of electric field vs. r .

Figure 2.10: Applying Gauss's law to a solid and conducting ball with excess charge, using a cylinder as a Gaussian surface to compute the flux.

2.5.1 Charged conductors

In a conductor with excess charge, the charge will be pushed to the conductors' surface until *equilibrium equilibrium*, where there is no electric field inside anymore to move more charges. We can use Gauss's law to prove that all the charge has to accumulate at the surface. Draw a Gaussian surface just below the charged conductor's surface like in Fig. 2.10b. We argued that the electric field is zero inside, otherwise the charges would redistribute. Because the electric field, and therefore the flux ($\Phi_{\text{net}} = 0$), is zero everywhere inside, by Gauss's law, there can be no excess charge ($Q_{\text{in}} = 0$) inside this Gaussian surface. Therefore, we conclude that at equilibrium, all the charge have to be on the surface, and $E = 0$ inside.

Spherical conductor

Let's make this more precise: Consider for simplicity a spherical conductor with radius R and some net charge Q , like in Fig. 2.10. The electric field $\mathbf{E}(r) = 0$ inside, $r < R$. Outside, $r > R$,

$$\Phi = \oint_S E_n dA = 4\pi k Q_{\text{in}}. \quad (2.43)$$

Because this setup is spherically symmetric, because $\mathbf{E} \parallel \hat{\mathbf{n}}$ and $E_n = \mathbf{E} \cdot \hat{\mathbf{n}}$ only depends on the distance r . The integral is reduced to

$$E_n(4\pi r^2) = 4\pi k Q_{\text{in}}. \quad (2.44)$$

Therefore,

$$E_n = \frac{k Q_{\text{in}}}{r^2}. \quad (2.45)$$

This result is the same as for a uniformly charged, hollow sphere (see Section 2.2.2), which is no surprise, given the charge distribution inside the conductor.

Conducting slab

Now consider a conducting slab with charge density σ , as in Fig. 2.11. We have again a cylinder with one outer (S_1) and one inner endcap (S_2), and a side S_3 . The electric field at the surface of a charged conductor is constant and perpendicular to the outer endcap S_1 . Since there is no field inside, and the field is parallel to the side S_3 , we are left with

$$\Phi = EA + 0 + 0. \quad (2.46)$$

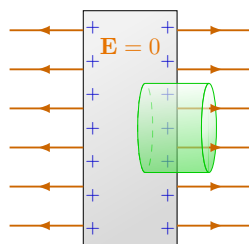


Figure 2.11: Computing the flux of a conducting slab with excess charge.

Using Gauss's law again, we find

$$\mathbf{E} = 4\pi k\sigma = \frac{\sigma}{\epsilon_0} \hat{\mathbf{x}}. \quad (2.47)$$

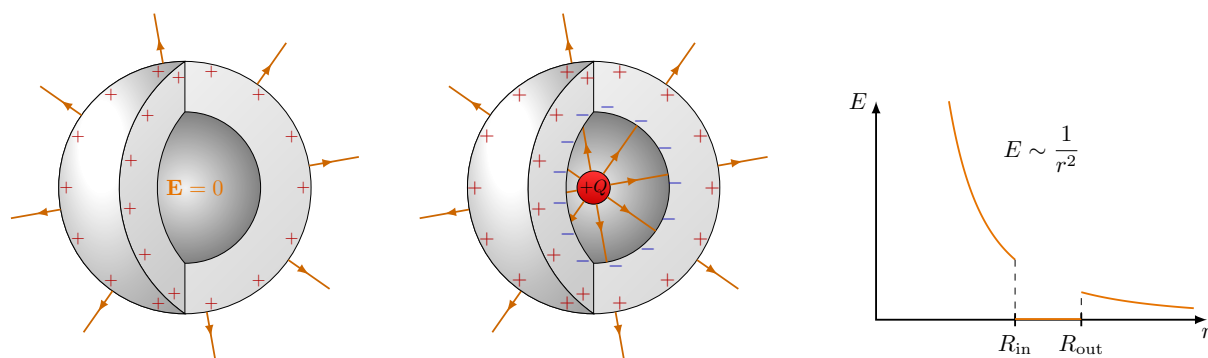
on the left side. This is the same result as before for a large, uniformly charged sheet, Eq. (1.56), or slab, Eq. (2.32), but this time there is a factor two difference. This can be explained by the fact that now the charge density is defined as the charge per area on the (disconnected) two sides of the conductor, whereas for the uniformly charged sheet or slab, the density was defined for the whole object.

For a more general conductor of any shape, Eq. (2.47) will hold close to the surface, and the electric field will always be perpendicular to the surface, no matter the curvature.

2.5.2 Conductors with a cavity

Now consider a hollow conductor with some excess charge as in Fig. 2.12a. What is the electric field inside the cavity? Because all the charge sits on the external surface, we conclude by Gauss's law that the field is zero.

Now imagine we suspend some charge Q inside the cavity of the conductor, without it touching the conductor, as in Fig. 2.12b. Suppose the conductor has no net charge while $+Q > 0$. What will happen with the electric field E_1 inside the cavity, the field inside the conductor, E_2 , and the field outside the conductor, E_3 ? Charge Q will bring about an electric field, causing the redistribution of the conductor's free charges. The negative charges (electrons) will be attracted to the positive charge Q inside, leaving positive charges on the outer surface. This will mean that E_1 stays the same as the original electric field of charge Q . On the other hand, $E_2 = 0$ because of the charge distribution, and outside, $E_3 \neq 0$, will be a field due to the net charge on the conductor's external surface. From close by, it will have a field $E_3 = \sigma/\epsilon_0$, where $\sigma = Q/A$ is the density over the outer surface



(a) Excess charge moves to the outer surface. Inside the charge and electric field are zero.

(b) A suspended charge in the center of a neutral conductor will cause a charge redistribution.

(c) Plot of electric field vs. r for a spherical conductor with a charge inside a cavity, like in (b).

Figure 2.12: Spherical conductor of radius R_{out} with cavity of radius R_{in} inside.

with area A . Notice that charge on the outer surface will be exactly Q if the conductor had no net charge in the beginning.

2.5.3 Faraday's cage

A Faraday's cage works exactly on this principle of redistributing charges in conductors. A Faraday's cage is a metal cage or box that will protect anything on the inside from electric fields and charge. This is why you are more safe from lightning inside a car than in a house (without a lightning rod). This principle also works with electromagnetic waves: If you put a radio or your mobile phone inside a metal cage like an elevator, you will not be able to pick up a signal.

Chapter 3

Electric Potential

3.1 Electric potential energy

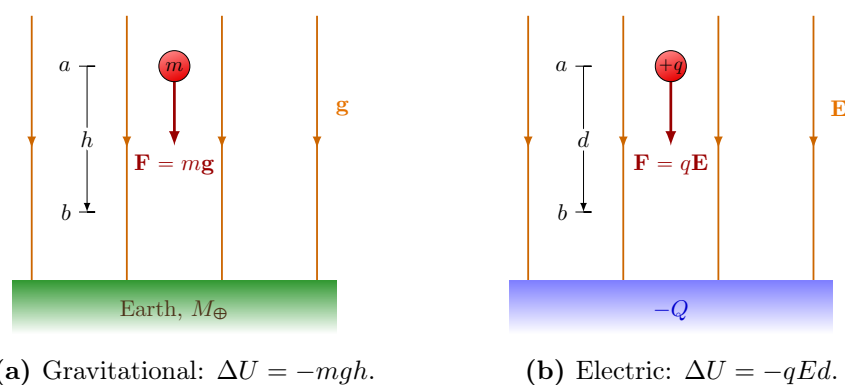


Figure 3.1: Comparison of potential energy difference $\Delta U = U_b - U_a$ in a force field.

Let's review the potential energy due to gravitational field, and compare it to that of an electric field. Imagine you drop a mass m from rest in a gravitational field \mathbf{g} , starting from a point a and falling some height h to a point b , like in Fig. 3.1a. The initial gravitational potential energy U_a decreases as we go from a higher to a lower point, $U_a > U_b$. The difference is given by

$$\Delta U = U_b - U_a = -mgh. \quad (3.1)$$

This potential difference is independent from the path the mass takes, because the gravitational force is conservative. The work done by a gravitational force is

$$W = -\Delta U = mgh. \quad (3.2)$$

More generally, one has to integrate the force over some path

$$W = \int_a^b \mathbf{F} \cdot d\boldsymbol{\ell}, \quad (3.3)$$

where $\boldsymbol{\ell}$ is a line segment of the the path along which the mass m travels. If the gravitational field is uniform (like close to Earth's surface), we can write $\mathbf{g} = -g\hat{\mathbf{y}}$, and $\mathbf{F} = -mg\hat{\mathbf{y}}$. Then only the path changes in the y direction matter, as the scalar product $\mathbf{F} \cdot d\boldsymbol{\ell}$ projects out the y component of the path.

Now compare this to an electric field. We hold a charge q from some distance of a charged infinitely large plane. Assume the charge is opposite so the force is attractive. We let q loose, and it freely moves towards the charged surface. As it moves closer, say over an distance d between points a and b like in Fig. 3.1b, the potential energy due to the electric force decreases, $U_a > U_b$, and

$$\Delta U = U_b - U_a = -qEd, \quad (3.4)$$

The minus here reflects that the force is attractive and the free charge is moving closer, just like for a gravitational force on a mass falling toward the Earth. The work done by the electric field, then, is

$$W = -\Delta U = qEd. \quad (3.5)$$

This was assuming a straight path along an uniform field. For a more general path in a more general electric field \mathbf{E} as in Fig. 3.2,

$$\begin{aligned} W &= \int_a^b \mathbf{F} \cdot d\boldsymbol{\ell} \\ &= \int_a^b q\mathbf{E} \cdot d\boldsymbol{\ell}, \end{aligned} \quad (3.6)$$

because $\mathbf{F} = q\mathbf{E}$. Notice that whether the work done along a line segment $d\boldsymbol{\ell}$ is positive or negative depends on not just whether $d\boldsymbol{\ell}$ is in the same ($\mathbf{E} \cdot d\boldsymbol{\ell} > 0$) or opposite direction ($\mathbf{E} \cdot d\boldsymbol{\ell} < 0$) as the field, but also on the charge sign of q , see Fig. 3.3.

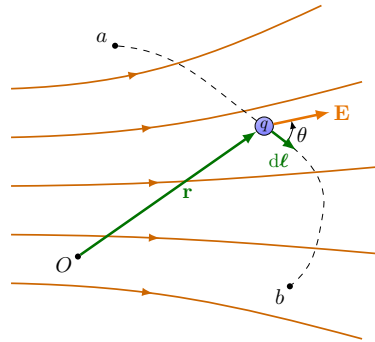


Figure 3.2: Path of a charge in an electric field.

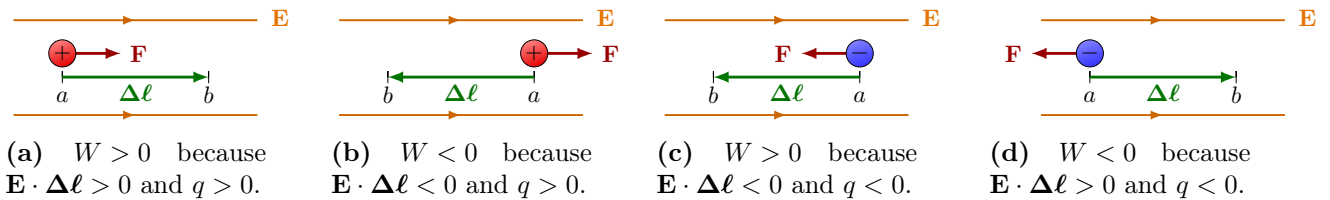


Figure 3.3: Work $W = \mathbf{F} \cdot \Delta\boldsymbol{\ell} = q\mathbf{E} \cdot \Delta\boldsymbol{\ell}$ done by the electric field on a charge q depends on the direction of motion in the field and the sign of q .

3.2 Electric potential

Instead of using the electric potential energy U , it is more common to use the *electric potential* V . Like energy, this is also a scalar quantity (as opposed to a vector quantity).

In fact, the important quantity is the *potential difference* ΔV , or *voltage*:

$$\Delta V = \frac{\Delta U}{q}, \quad (3.7)$$

where $\Delta V = V_b - V_a$ and $\Delta U = U_b - U_a$. Often just “ V ” for the voltage is used instead of “ ΔV ”. Because only the difference is important, we can often choose a reference potential, either V_a or V_b . Just like in the definition of the electric field, we divided out the test charge. The units of electric potential V and voltage ΔV are Volts (V). Clearly, they have the dimension of energy over charge,

$$1 \text{ V} = 1 \frac{\text{J}}{\text{C}} = 1 \frac{\text{Nm}}{\text{C}}. \quad (3.8)$$

Because the units of potential energy is Nm, we can now write the units of electric fields as

$$1 \frac{\text{N}}{\text{C}} = 1 \frac{\text{V}}{\text{m}} \quad (3.9)$$

Just like potential energy, only differences or changes ΔV in the electric potential V are important.

Conversely, we see that energy can be expressed in units of charge times voltage. In fact, a convenient unit of energy is electronvolts eV, where “ e ” indicates the elementary charge e :

$$\Delta U = 1 \text{ eV} = (1.6 \times 10^{-19} \text{ C})(1 \text{ V}) = 1.6 \times 10^{-19} \text{ J}. \quad (3.10)$$

This unit is used very often in particle physics, because you can interpret this as the energy that an electron gains when accelerated by a voltage of 1 V. For example, in the LHC at CERN, protons are accelerated up to energies corresponding to 6.5 TeV.

For a general path, we notice that

$$dV = dU/q = -dW/q, \quad (3.11)$$

and use Eq. (3.6) to find

$$dV = -\mathbf{E} \cdot d\boldsymbol{\ell}. \quad (3.12)$$

The negative sign here indicates that the potential V *decreases* if the movement $d\boldsymbol{\ell}$ is *along* the electric field \mathbf{E} , like in Fig. 3.4. The electric field therefore, goes from high to low potential. The voltage can therefore be computed with

$$\begin{aligned} V &= \int_a^b dV \\ &= - \int_a^b \mathbf{E} \cdot d\boldsymbol{\ell}. \end{aligned} \quad (3.13)$$

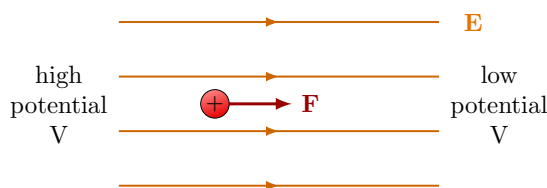


Figure 3.4: Electric field lines move from high to low electric potential.

3.2.1 Electric potential of a uniform field

Consider a uniform electric field $\mathbf{E} = E\hat{\mathbf{x}}$ like in Fig. 3.4. The electric field lines moves from a high potential to a low potential. What then, is the potential as a function of x along the field? Now we calculate the difference, from a reference point $x = 0$ to point x :

$$V(x) - V(0) = - \int_0^x E \, d\ell = -Ex. \quad (3.14)$$

At $x = 0$ we set the reference potential to $V(0) = 0$, such that

$$V(x) = -Ex. \quad (3.15)$$

The minus sign is expected, since the voltage moves from high to low values along the electric field (the x direction in this case).

Example 3.1: To get a feeling of numbers, suppose a proton is put at rest in an electric field $\mathbf{E} = 5.0\text{N/m}\hat{\mathbf{x}}$. What is its potential V after moving a distance of 1 meter in this field in the x direction? Using energy conservation:

$$\begin{aligned} 0 &= \Delta K + \Delta U \\ &= \frac{1}{2}mv^2 + q\Delta V. \end{aligned}$$

The electric potential over 1 meter is $\Delta V = -Ex = -5.0\text{V}$, so the final velocity is

$$v = \sqrt{\frac{2qEx}{m}} = \sqrt{\frac{2q\Delta V}{m}} = 31.0 \frac{\text{km}}{\text{s}}, \quad (3.16)$$

where we used the proton mass $m_p = 1.67 \times 10^{-27}\text{kg}$ and charge $q = e = 1.60 \times 10^{-19}\text{C}$.

3.2.2 Electric potential of a point charge

The electric field of a point charge is

$$\mathbf{E} = \frac{kq}{r^2}\hat{\mathbf{r}}, \quad (3.17)$$

and so, using Eq. (3.12),

$$\begin{aligned} dV &= -\frac{kq}{r^2} \cdot dr \hat{\mathbf{r}} \\ &= -\frac{kq}{r^2} dr. \end{aligned} \quad (3.18)$$

We can integrate this from a reference point r_0 at lower potential to r at higher potential:

$$\begin{aligned} V &= \int_{r_0}^r dV \\ &= \left[\frac{kq}{r} \right]_{r_0}^r. \end{aligned} \quad (3.19)$$

Only the difference in potential is important, so we can choose an arbitrary reference point r_0 to define our potential. For convenience, we typically set r_0 at infinity such that $V(r) = 0$ at $r \rightarrow \infty$. Then the last equation becomes

$$V(r) = \frac{kq}{r}. \quad (3.20)$$

If we have several charges, we simply sum up the potentials to get the total potential at a point P :

$$V = \sum_i \frac{kq_i}{r_i}, \quad (3.21)$$

where the sum runs over each charge i , and r_i is the distance between P and charge i . For a continuous body of charge, like in Fig. 1.15c, this generalizes into the integral

$$V(\mathbf{r}) = \int \frac{k dq}{|\mathbf{R} - \mathbf{r}|}. \quad (3.22)$$

Instead of performing this integral, it is often easier to first find the electric field for some charge configuration by applying Gauss's law, and then integrate over $-\mathbf{E} \cdot d\boldsymbol{\ell}$ to find the potential V . An example is shown in the next section.

3.2.3 Electric potential of a charged, conducting sphere

What is the potential of a conducting sphere with a net charge Q as a function of r ? Outside the sphere, $r > R$,

$$\mathbf{E} = \frac{kQ}{r^2} \hat{\mathbf{r}}. \quad (3.23)$$

To find $V(r)$, we integrate,

$$V(r) = \int_{\infty}^r \frac{kQ}{r'^2} dr' = \frac{kQ}{r}. \quad (3.24)$$

Inside the conductor, $r < R$, we already saw in Section 2.5.1 that $E = 0$. It follows that the sphere has some constant potential everywhere under its surface. No work will be done on a moving test charge. For $r = R$, we have

$$E = \frac{kQ}{R^2} \hat{\mathbf{r}}, \quad (3.25)$$

and consistently with Eq. (3.24),

$$V = \frac{kQ}{R}. \quad (3.26)$$

So this is the potential everywhere inside and on the conductor, which is constant. A plot of this electric potential field is shown in Fig. 3.5.

3.3 Equipotential surfaces

Equipotential surfaces are surfaces of constant voltage, or in other words, each point on the surface has the same potential. If a charge moves along an equipotential line, its electrical potential energy stays the same.

Some examples are shown in Fig. 3.6. One important property is that the equipotential surfaces are always perpendicular to the electric field.

3.4 Discharge

In reality, we are limited in how much potential V we can build up of charge Q . Namely, in high electric fields, air becomes ionized. This happens around $E_{\max} \approx 3 \times 10^6 \text{ V/m}$. Because electrons are freed from their molecular bounds, air becomes conductive. This is known as *dielectric break down*. Say E_{\max} is the dielectric strength of a material in

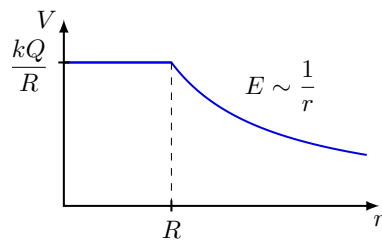
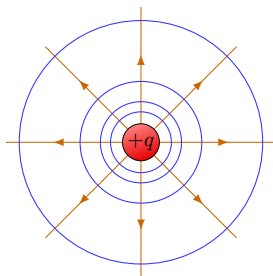
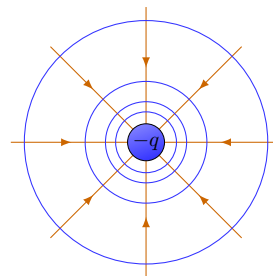


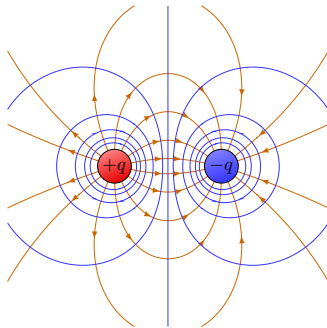
Figure 3.5: Electric plot V vs. r for a charged, spherical conductor. Compare to Fig. 2.10c.



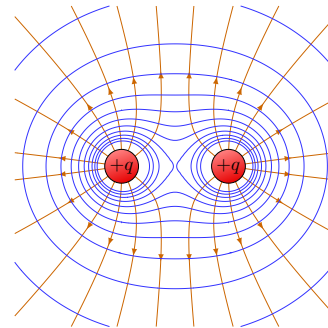
(a) Positive charge.



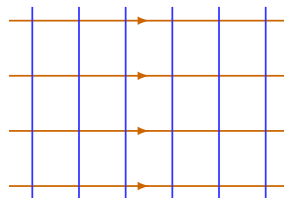
(b) Negative charge.



(c) Opposite point charges.



(d) Same-sign point charges.



(e) Uniform field (like for a infinite sheet of charge).

Figure 3.6: Equipotential surfaces (blue) of electric field lines (orange) for different configurations of point charges. All the points on the same equipotential have the same electric potential. The equipotential are equidistant to each other: Two neighbouring equipotentials differ by a fixed voltage ΔV .

some shape. At E_{\max} , charge will be able to move through the air. This is called *arc discharge*. Examples are lightning and static discharge (when you get zapped by built-up static electricity).

Example 3.2: For a spherical conductor of radius R in air, what is the maximum charge Q_{\max} and maximum V_{\max} in terms of E_{\max} ? Zoom into a small spot on the surface, such that we can approximate the field with $E_{\max} = \sigma_{\max}/\epsilon_0$. Where for a sphere

$$\sigma_{\max} = \frac{Q_{\max}}{4\pi R^2}. \quad (3.27)$$

So we find that

$$E_{\max} = \frac{Q_{\max}}{4\pi\epsilon_0 R^2}. \quad (3.28)$$

From our previous result, Eq. (3.24), we know that

$$V_{\max} = \frac{Q_{\max}}{4\pi\epsilon_0 R} = E_{\max} R. \quad (3.29)$$

So clearly, for a given E_{\max} (as a property of air), the voltage under which spherical conductor will discharge is higher for a larger radius, while conductors with smaller radii, reach can V_{\max} easier.

3.5 Non-spherical conductor

Consider a non-spherical conductor with potential V . For sake of argument, suppose we can approximate each end of this body by a sphere, one with radius r_a , and one with r_b , just like in Fig. 3.7. We have already seen in Section 3.2.3 that the potential is constant everywhere on the surface. But if $r_a < r_b$, what are the charge densities at points a and b ? Rewriting our result from the previous section, the charge density for a given potential V clearly depends on the radius:

$$\sigma = \frac{V\epsilon_0}{r}. \quad (3.30)$$

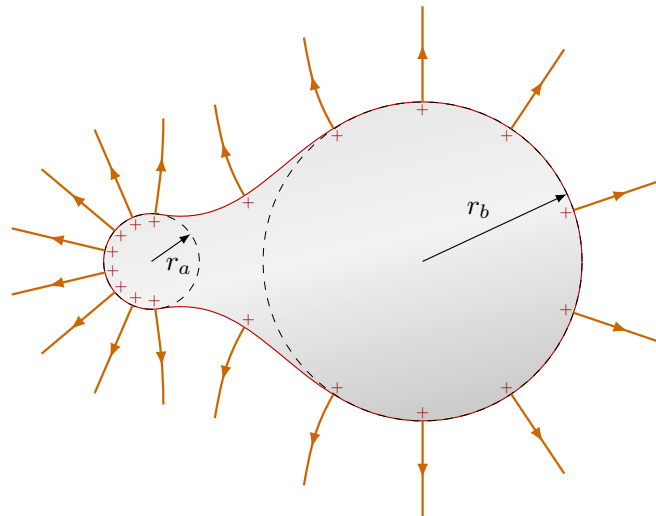


Figure 3.7: Conductor with varying curvatures. The charge density is larger and the electric field is strong at the surface with smaller curvature.

Even though everywhere on the conductor's surface the potential is the same, there is more charge where the surface's curvature has a smaller radius.

What about the electric field? Because $E = \sigma/\epsilon$, the electric field must be higher at the end with a smaller radius.

3.6 Poisson's equation

At the beginning of this chapter, we wrote the differential of potential as

$$dV = -\mathbf{E} \cdot d\boldsymbol{\ell}. \quad (3.31)$$

If we assume the simple case where the path is completely along the field, then

$$dV = -E d\ell, \quad (3.32)$$

and we see that

$$E = -\frac{dV}{d\ell}. \quad (3.33)$$

More generally, we can write the electric field as the *gradient* of the potential,

$$\mathbf{E} = -\nabla V. \quad (3.34)$$

Here, the gradient operator discussed in Section 2.3 shows up again. This time however, the operator acts on a scalar function $V(\mathbf{r})$, instead of on a vector function $\mathbf{E}(\mathbf{r})$. In a simple Cartesian xyz -coordinate system,

$$\nabla V = \frac{\partial V}{\partial x} \hat{\mathbf{x}} + \frac{\partial V}{\partial y} \hat{\mathbf{y}} + \frac{\partial V}{\partial z} \hat{\mathbf{z}}. \quad (3.35)$$

For case with spherical symmetric, one can also write

$$\nabla V = \frac{\partial V}{\partial r} \hat{\mathbf{r}}. \quad (3.36)$$

Relation 3.34 is a consequence of another theorem from calculus,

Fundamental theorem of line integrals (Gradient theorem). *The line integral over the gradient of a scalar field V between points a and b is*

$$\int_a^b \nabla V d\boldsymbol{\ell} = V(b) - V(a). \quad (3.37)$$

The gradient of a scalar field is a vector field. It can loosely be interpreted as a derivative that points in the directions of largest change.

Let's check the consistency with the potential in Eq. (3.20) that we computed for a point charge:

$$\begin{aligned} \mathbf{E} &= -\nabla V = -\frac{d}{dr} \left(\frac{kq}{r} \right) \hat{\mathbf{r}} \\ &= \frac{kq}{r^2} \hat{\mathbf{r}}, \end{aligned}$$

As expected, we retrieve the electric field of a point charge, which is proportional to $1/r^2$, while the potential goes as $1/r$.

Remember that Gauss's law can be written without an integral as

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \quad (3.38)$$

Taking the divergence of each side, we find

$$\nabla \cdot \mathbf{E} = -\nabla \cdot (\nabla V) = -\nabla^2 V. \quad (3.39)$$

This leads us to

Poisson's equation for electrostatics.

$$\nabla^2 V = -\frac{\rho}{\epsilon_0}. \quad (3.40)$$

Here, ∇^2 is called the *Laplacian operator*, and acts on a scalar field V as

$$\nabla^2 V = \frac{\partial^2}{\partial x^2} V + \frac{\partial^2}{\partial y^2} V + \frac{\partial^2}{\partial z^2} V. \quad (3.41)$$

The importance of Poisson's equation is that if you know $\rho(r)$, you can calculate $V(r)$. If $\rho(r) = 0$, then

Laplace's equation.

$$\nabla^2 V = 0. \quad (3.42)$$

This is called *Laplace's equation*. Its solutions are harmonic functions.

If the Laplacian of a scalar is zero at some point in space, that point is neither a minimum nor a maximum for that scalar. Therefore, if the Laplacian of the potential is zero everywhere inside a volume, the minimum and maximum of potential for that volume should be on the boundaries.

For example, let's assume that we have a constant electric field parallel to the x axis. According to Eq. (3.32) and for a constant electric field,

$$\Delta V = -E\Delta x \quad (3.43)$$

which means that the electric potential decreases by increasing x and given an arbitrary volume, the highest potential for that volume is at one or a set of points with the lowest x coordinate and therefore on the boundary of the volume. The minimum of electric potential is also at the point(s) with the highest x coordinate and on the boundary.

An important remark is that if the Laplacian of a scalar like electric potential is zero everywhere inside a volume and therefore the minimum and maximum of electric potential exist on the boundaries, then if the electric potential is the same anywhere on the boundary, we infer that the minimum and maximum of electric potential for the whole volume are the same which means we have a constant potential for the whole volume.

Sections 3.7 and 3.8 are written by Arash Jofrehei. They are supplemental and not covered on the exam.

3.7 Extra: Uniqueness theorem

The uniqueness theorem states that given the Laplacian of a scalar for each point of a volume and the value of scalar on the boundaries, the Laplace equation has only one unique solution.

To add more familiarity, let's assume this scalar is the electric potential and take an arbitrary volume for which we know that the potential on the boundary is $V_b(x, y, z)$. And we know the Laplacian of potential anywhere inside the volume or equivalently the charge density $\rho(x, y, z)$ according to Eq. (3.40). Assume that there are two solutions satisfying these requirements and let's call them U_1 and U_2 . So we would have:

$$\nabla^2 U_1 = \nabla^2 U_2 = \frac{\rho}{\epsilon_0} \quad (3.44)$$

$$U_1^{\text{boundary}} = U_2^{\text{boundary}} = V_b \quad (3.45)$$

These two equations lead to:

$$\nabla^2(U_1 - U_2) = 0 \quad (3.46)$$

$$(U_1 - U_2)^{\text{boundary}} = 0 \quad (3.47)$$

As mentioned earlier, if the Laplacian of a scalar function is zero everywhere inside a volume and is constant on the boundary, it is constant also everywhere inside. In this case the constant is zero and therefore $U_1 - U_2$ is zero and the two solutions are the same, hence the uniqueness theorem is proved.

3.8 Extra: Method of images

The uniqueness theorem provides a useful handle to solve some electrostatic problems. If one is able to simply guess a potential or equivalently electric field or charge distribution as a solution that satisfies the Poisson's equation and the boundary conditions, this solution would be uniquely correct. This theorem serves as the foundation of the method of images. Following, this method is introduced and utilized through a few examples.

3.8.1 A point charge in front of an infinite conductor plane

Assume that we put a point charge of q at distant h from an infinitely large conductor plane as in Fig. 3.8 and want to find the electric potential at any arbitrary point A at the same side of the plane ($z > 0$). The plane itself is connected to the ground so it has a constant electric potential of zero.

The classic way of solving this problem would be to solve the second order Poisson's differential equation for $z > 0$. The charge and therefore the Laplacian of potential is zero everywhere but at the point charge. The boundary condition is that the potential is constantly zero on all the boundaries. Solving this second order differential equation is difficult but if we were able to just guess a solution, which is sometimes easy for problems with a lot of symmetries, we would have our unique solution for the electric potential in $z > 0$.

The general idea behind the method of images is to replace the elements of the given problem, normally a conductor, with imaginary charge distributions for which it is easy to calculate the potential. It is important to remember that any proposed charge distribution should be outside the volume under study. Otherwise, the Poisson's equation for the volume would not stay the same.

Here, the volume under study is all the points with $z > 0$. So, we are allowed to replace the infinite conductor plane with any charge distribution in $z < 0$ as long as the new distribution gives the same electric potential on the boundaries. As seen in Fig. 3.8, the proposed imaginary charge distribution in this case can be simply a point charge of $-q$ at distant h from the plane, mirror to the main charge. Due to this mirroring, the imaginary charge is sometimes referred to as the image charge or basically the image of the main charge on the plane.

The imaginary charge is outside the volume under study so the Poisson's equation is unaffected and we just have to check the boundary conditions. At any point B with $z = 0$ and distance r from the main and imaginary charges, the electric potential is

$$V_B = \frac{q}{4\pi\epsilon_0 r} + \frac{-q}{4\pi\epsilon_0 r} = 0. \quad (3.48)$$

So the electric potential is still the same on the $z = 0$ plane and as the electric potential of point charges is also zero at infinitely far points, the boundary conditions are untouched after replacing the conductor with a point charge below it.

Having this quite simplified charge distribution, we can find the electric potential at any arbitrary point A at the distance of d_1 from q and d_2 from $-q$ to be

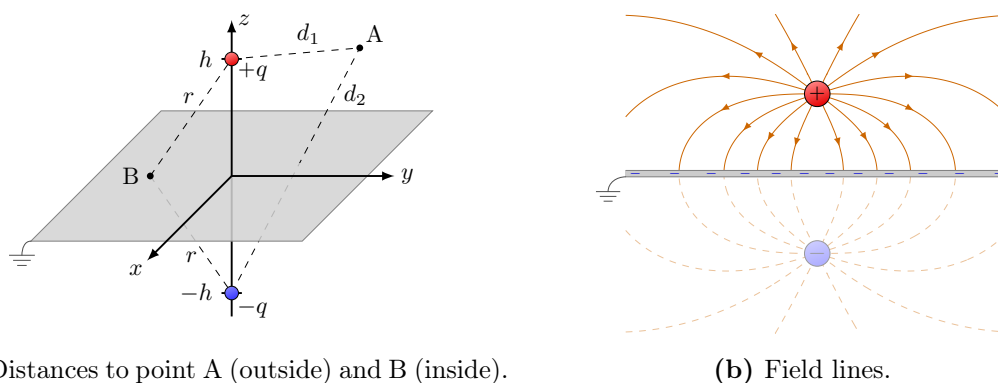
$$V_A = \frac{q}{4\pi\epsilon_0 d_1} + \frac{-q}{4\pi\epsilon_0 d_2}. \quad (3.49)$$

Having the electric potential, one can find the electric field using Eq. (3.34). One can even find the charge density at any point on the plane using the electric field.

3.8.2 A point charge in front of a spherical conductor

As another example, as depicted in Fig. 3.9, assume that we put a point charge of q in front of a spherical conductor with radius R at the distance of d from its center, and require the electric potential at any arbitrary point A outside the sphere. The sphere is grounded so the electric potential is zero on the boundaries of the volume under study ($r > R$). Similar to the previous problem, the aim is to propose an imaginary charge or charge distribution to replace the conductor and still satisfy the Poisson's equation in the volume under study and hold the same boundary condition.

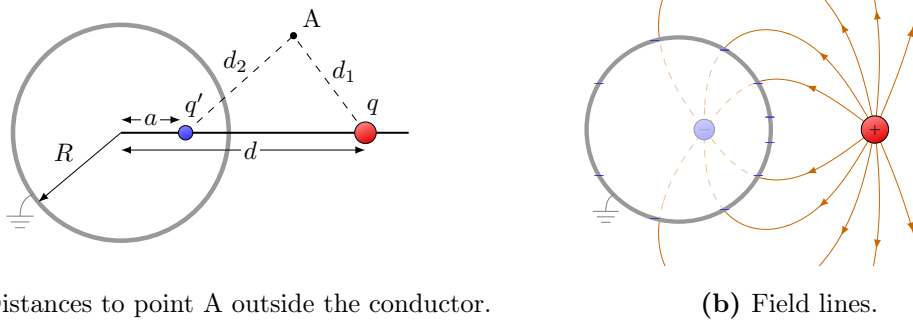
The proposed image charge in this case is a single point charge q' at radius a from the center of the sphere and on the same axis as the main charge. One can check that if we



(a) Distances to point A (outside) and B (inside).

(b) Field lines.

Figure 3.8: A point charge q in front of a grounded infinite conductor plane induces a charge distribution. To find the electric field, one can introduce an imaginary charge $q' = -q$ that mirrors q on the other side as trick.



(a) Distances to point A outside the conductor.

(b) Field lines.

Figure 3.9: A point charge q in front of a grounded spherical conductor induces a charge distribution. To find the electric field, one can introduce an imaginary charge $q' = -qR/d$ inside the sphere as trick.

take q' and a to be

$$q' = -q \times \frac{R}{d},$$

$$a = \frac{R^2}{d},$$

the electric potential will be zero at any point at radius R so the boundary condition is satisfied. Since the image charge is outside the volume under study, the Poisson's equation also stays the same in the volume. Thus, we have found a simplified charge distribution leading to a unique solution for electric potential inside this volume. The electric charge at any arbitrary point A outside the sphere at the distance d_1 from q and d_2 from q' will be

$$V_A = \frac{q}{4\pi\epsilon_0 d_1} + \frac{-qR}{4\pi\epsilon_0 d d_2}. \quad (3.50)$$

Given this electric potential, one can find the electric field anywhere outside the sphere and the charge density at any point on the sphere.

Chapter 4

Capacitance & Dielectrics

4.1 Capacitors

Any two oppositely charged conductors form a *capacitor*.

The simplest capacitor consists of two parallel plates with the same charge $\pm Q$, but opposite, like in Fig. 4.1. In Section 2.2.4), we previously found

$$E_1 = \frac{\sigma}{\epsilon_0} = \frac{Q}{2\epsilon_0 A} \quad (4.1)$$

for a single sheet, or plate, of charge. Now we have two plates of opposite charge, so in between them, the fields add up to

$$E = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A}, \quad (4.2)$$

while outside, they cancel out each other ($E = 0$). (In reality, of course, the field becomes less and less uniform as you approach the edges, as is shown in Fig. 4.1c, there is still non-zero field outside the plates, albeit much weaker than inside.) If the electric field is uniform between the plates, the voltage between them is

$$V = Ed = \frac{Qd}{\epsilon_0 A}. \quad (4.3)$$

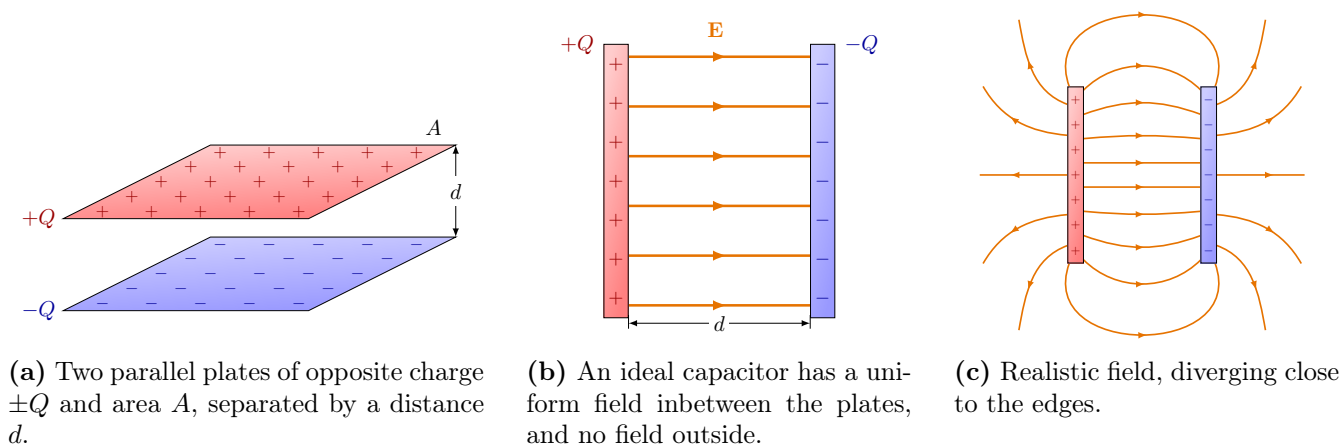


Figure 4.1: Parallel plate capacitors.

4.1.1 Capacitance

The *capacitance* measures the amount of charge Q that can be stored under a voltage V , and is defined as

$$C = \frac{Q}{V}. \quad (4.4)$$

The units of capacitance are Faraday, or just F, where

$$1 \text{ F} = 1 \frac{\text{C}}{\text{V}}. \quad (4.5)$$

For parallel planes, the capacitance is simply

$$C = \frac{\epsilon_0 A}{d}. \quad (4.6)$$

The charge completely canceled out, and it now only depends on the area A and distance d , so clearly, the capacitance is a function of the geometry of the capacitor, rather than the charge density.

In a electronic circuit diagram, the symbol is two parallel plates, like in Fig. 4.2.

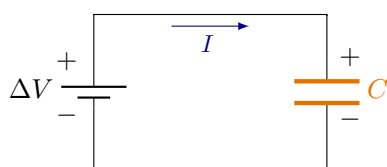
4.1.2 Spherical capacitor

Suppose you have a spherical capacitor made of two spherical conducting shells. The voltage holding a charge Q is

$$\begin{aligned} \Delta V &= V_a - V_b \\ &= \frac{Q}{4\pi\epsilon_0 r_a} - \frac{Q}{4\pi\epsilon_0 r_b} \\ &= \frac{Q}{4\pi\epsilon_0} \left(\frac{1}{r_a} - \frac{1}{r_b} \right). \end{aligned}$$



(a) Simple circuit with voltage ΔV and capacitor C .



(b) Simple circuit with battery and capacitor C .



(c) Alternative symbol for a (polar) capacitor.

Figure 4.2: Diagram of an electric circuit with a battery and a capacitor C (orange).

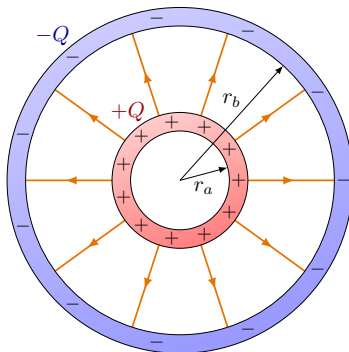


Figure 4.3: Cross section of a spherical capacitor with two shells holding charge $\pm Q$. The electric field is zero inside the inner and outside the outer sphere.

The capacitance then, is

$$\begin{aligned}
 C &= \frac{Q}{\Delta V} \\
 &= \frac{4\pi\epsilon_0}{\frac{1}{r_a} - \frac{1}{r_b}}.
 \end{aligned}
 \tag{4.7}$$

Up until now, capacitor was defined as *two* oppositely charge capacitors. But what if there were only *one* shell with some charge $+Q$? What would the capacitance be? In this case we can take $r_b \rightarrow \infty$, and therefore Eq. (4.7) reduces to

$$C = 4\pi\epsilon_0 r_a. \tag{4.8}$$

It is called the *self capacitance*, and any object that can be charged has a self capacitance.

4.2 Dielectrics

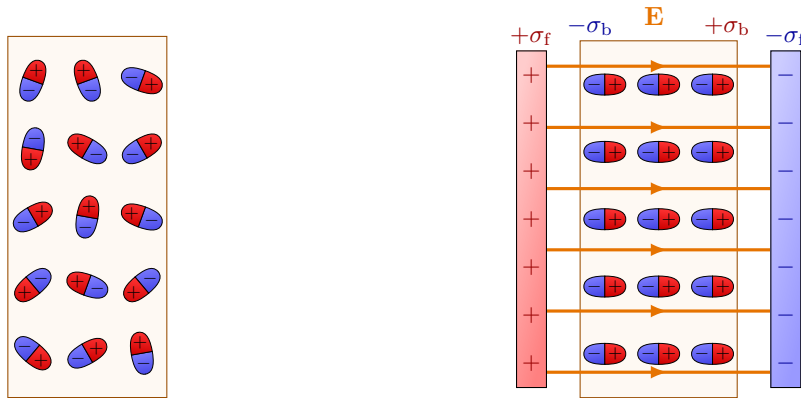
In a non-conducting materials, molecular dipoles tend to align oppositely to the electric field. The negative side of the dipole is attracted towards the positive side of the electric field. But remember that a dipole has a dipole moment $\mathbf{p} = q\mathbf{L}$ and will cause its own electric field, now in the opposite direction of the external electric field. If you have many small dipoles in an external electric field E_0 , it will therefore sum up and weaken the net field:

$$E_0 > E_0 - E_{\text{dielectric}}. \tag{4.9}$$

We can have a material composed of dipoles, called a *dielectric*. Each material has its own *dielectric constant*, $\kappa > 1$, that weakens the external electric field E_0 , such that

$$E = \frac{E_0}{\kappa}. \tag{4.10}$$

is the new field with the dielectric inserted. This is all illustrated in Fig. 4.4. Note that κ should always be larger than one, because in principle, a dielectric should only decrease the field, not increase it.



(a) The dipoles in a dielectric without an external field are not aligned.

(b) In an external field, the dipoles of the dielectric align and oppose the field.

Figure 4.4: Dielectric medium.

Table 4.1: Dielectric constant κ .

Material	κ
Air	1.00059
Paraffin	2
Plexiglass	3.2
Water	80

We know that $V = Ed$ for a parallel plate capacitor, so just like the external field electric field, the voltage reduces to

$$V = \frac{V_0}{\kappa}. \quad (4.11)$$

Consider the parallel plane capacitor with a slab of dielectric material with constant κ inserted, like in Fig. 4.4b. Because the electric field and voltage are weaker, the capacitor will also store less charge. So how does the initial capacitance C_0 under V_0 change? We find that the new capacitance C under the new voltage V can be expressed as

$$C = \frac{Q}{V} = \frac{Q\kappa}{V_0}. \quad (4.12)$$

For parallel planes then,

$$C = \frac{\kappa\epsilon_0 A}{d} = \frac{\epsilon A}{d}, \quad (4.13)$$

where we defined the *permittivity of the dielectric*

$$\epsilon = \kappa\epsilon_0. \quad (4.14)$$

In light of this, ϵ_0 can actually be thought of as the permittivity of the vacuum.

The charge density of the conductor on each plate is

$$\sigma_f = \frac{Q}{A}, \quad (4.15)$$

where the “f” label refers to the the charge that can move freely. Because the dipoles align, there will also be a buildup of charge on the either side of the dielectric’s surface, like in Fig. 4.4b. Its charge density is given by

$$\sigma_b = \frac{Q}{A}, \quad (4.16)$$

with the “b” label standing for bound charge. Remember that outside a charged surface, $E = \sigma/\epsilon_0$. Therefore, the electric field due to the plates is

$$E_0 = \frac{\sigma_f}{\epsilon_0}, \quad (4.17)$$

and the one due to the charge on the dielectric’s surface is

$$E_b = \frac{\sigma_b}{\epsilon_0}. \quad (4.18)$$

The net field becomes

$$\mathbf{E} = \mathbf{E}_0 - \mathbf{E}_b = \frac{\sigma_f}{\epsilon_0} - \frac{\sigma_b}{\epsilon_0}, \quad (4.19)$$

so we find that

$$\begin{aligned} E_b &= E_0 \left(1 - \frac{1}{\kappa}\right) \\ &= \frac{\kappa - 1}{\kappa} E_0. \end{aligned}$$

In terms of charge density,

$$\sigma_b = \frac{\kappa - 1}{\kappa} \sigma_f. \quad (4.20)$$

Because κ is always larger than 1, σ_b is smaller than σ_f .

4.2.1 Polarization & displacement

Let's look closer at the dielectric, and each individual charge due to the aligned dipoles on the surface. If there are N dipoles with some charge q and dipole moment $\mathbf{p} = q\mathbf{L}$, the total charge on the surface is

$$Q_b = \sigma_b A = NqLA, \quad (4.21)$$

or equivalently,

$$\sigma_b = \frac{Q}{A} = NqL. \quad (4.22)$$

This is also called the *polarization*

$$\mathbf{P} = Nq\mathbf{L} \quad (4.23)$$

of the dielectric. The electric field inside the dielectric becomes

$$\mathbf{E} = \mathbf{E}_0 - \frac{\mathbf{P}}{\epsilon_0}. \quad (4.24)$$

This allows us to define the *electric displacement*:

Electric displacement.

$$\mathbf{D} = \epsilon_0 \mathbf{E}_0 = \epsilon_0 \mathbf{E} + \mathbf{P}. \quad (4.25)$$

With this modification of the electric field \mathbf{E}_0 in vacuum, we can reformulate Gauss's law,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (4.26)$$

where now the charge volume-density is

$$\rho = \rho_f + \rho_b. \quad (4.27)$$

From the definition of \mathbf{D} , we see find

Gauss's law for dielectrics.

$$\nabla \cdot \mathbf{D} = \rho_f, \quad (4.28)$$

or in integral form,

$$\oint_S \mathbf{D} \cdot d\mathbf{A} = Q_f. \quad (4.29)$$

This is a useful form, because it references free charge Q_f , which we can directly measure.

Some useful references on the electric displacement \mathbf{D} and Gauss's law in a dielectric can be found in Section 1.7 of *Electrodynamics and Optics* by Wolfgang Demtröder, or Section 4.3 in *Introduction to Electrodynamics* by David Griffiths.

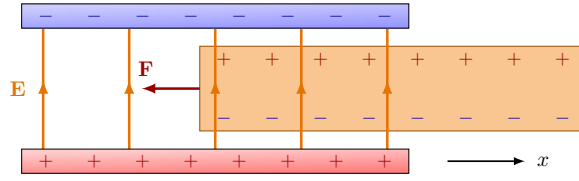


Figure 4.5: Force by the capacitor, pulling inside a dielectric .

4.3 Storage of energy in capacitors

In Section 3.2, we saw that the energy stored between a voltage V and a charge Q is $U = VQ$. Using the definition of capacitance, we find that

$$dU = VdQ = \frac{Q}{C}dQ. \quad (4.30)$$

When we charge up a capacitor, we continuously add charge from 0 to Q :

$$\begin{aligned} \int dU &= \int_0^Q \frac{Q}{C}dQ \\ &= \frac{Q^2}{2C}. \end{aligned}$$

Therefore, the energy stored in a capacitor will be

Energy stored in a capacitor.

$$U = \frac{Q^2}{2C} = \frac{QV}{2} = \frac{1}{2}CV^2. \quad (4.31)$$

4.3.1 Forces in capacitors

The parallel plates are attracting each other, so there is a force on either plates:

$$\begin{aligned} F &= QE \\ &= Q\frac{V}{d} \\ &= V^2\left(\frac{C}{d}\right). \end{aligned} \quad (4.32)$$

Note that the force $F \sim V^2$, and that C/d here acts as a geometrical factor of the capacitor.

A capacitor with a dielectric that is partially inside like in Fig. 4.5, will feel a force pulling it inside. To see this, consider that there will be a differential in capacitance, because in empty space, the capacitance is $C_0 = \epsilon_0 A/d$ and while inside the dielectric, $C = \epsilon A/d$. From dynamics (see PHY111), we already know that the relation between force and potential energy is

$$F = -\frac{\partial U}{\partial x}, \quad (4.33)$$

where x is along the length of the dielectric. So substituting Eq. (4.31),

$$|F_x| = \frac{V^2}{2} \frac{\partial C}{\partial x}. \quad (4.34)$$

Meanwhile, the forces on the dielectric along the y directions, i.e. towards the plates, cancel out.

4.3.2 Energy stored in electric fields

We can now think of the energy stored between two conductors (i.e. capacitor), as the energy stored in any electric field. For a parallel plate capacitor,

$$\begin{aligned} U &= \frac{1}{2}QV \\ &= \frac{1}{2}(E\epsilon A)(Ed), \end{aligned}$$

and therefore, the energy stored between the two plates is

$$U = \frac{1}{2}E^2 dA\epsilon. \quad (4.35)$$

Notice that dA is the volume between the plates, and therefore, we can express the energy density as

Energy density of an electrostatic field.

$$\eta_E = \frac{\text{energy}}{\text{volume}} = \frac{1}{2}E^2\epsilon \quad (4.36)$$

4.4 Combining capacitors

4.4.1 In parallel

What happens when we add capacitors in parallel? Consider two parallel plate capacitors that are connected by a wire on either side, as shown in Fig. 4.6. Because they are in parallel, the voltage V between each capacitor will be the same. Suppose that each capacitor i has a capacitance C_i , and a charge $Q_i = C_i V$. For two capacitors, the total charge is therefore

$$\begin{aligned} Q &= Q_1 + Q_2 \\ &= C_1 V + C_2 V \\ &= (C_1 + C_2)V. \end{aligned} \quad (4.37)$$

For any number of capacitors C_i that are in parallel, the equivalent capacitance is

Capacitance in parallel.

$$C_{\text{eq}} = \frac{Q}{V} = C_1 + C_2 + \dots \quad (4.38)$$

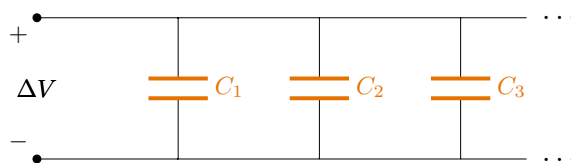


Figure 4.6: Combining in parallel capacitors C_1, C_2, C_3, \dots

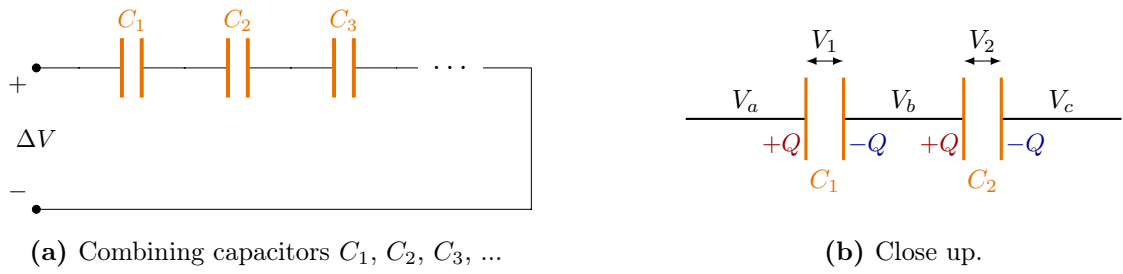


Figure 4.7: Capacitors in series.

4.4.2 In series

Now consider what happens if we put capacitors in series, as in Fig. 4.7a. Say each capacitor now has its own voltage V_i . Consider two parallel plate capacitors C_1 and C_2 as in Fig. 4.7b. The first plate of C_1 has some potential V_a , while the second one has V_b . The first plate of C_2 then, also has V_b , because it is directly connected to the second plate of C_1 . Finally the second plate of C_2 has a potential V_c . It is clear now that C_1 has a potential

$$V_1 = V_b - V_a = \frac{Q}{C_1} \quad (4.39)$$

and C_2 has

$$V_2 = V_c - V_b = \frac{Q}{C_2}. \quad (4.40)$$

The total potential difference across both capacitors is therefore

$$V = V_c - V_a. \quad (4.41)$$

We can write this out as

$$\begin{aligned} V &= (V_c - V_b) + (V_b - V_a) \\ &= \frac{Q}{C_1} + \frac{Q}{C_2} \\ &= Q \left(\frac{1}{C_1} + \frac{1}{C_2} \right) \end{aligned} \quad (4.42)$$

We see therefore, that in general for capacitors in series,

Capacitance in series.

$$\frac{1}{C_{\text{eq}}} = \frac{1}{C_1} + \frac{1}{C_2} + \dots \quad (4.43)$$

Notice that the equivalent capacitance gets weaker, each time you add a capacitor in series. So with the same capacitors, one can store more energy if you put them in parallel, than in series.

Example 4.1: If $C_1 = C_2 = 2 \mu\text{F}$, the equivalent capacitance is $C_{\text{eq}} = 4 \mu\text{F}$ if they are in parallel, $C_{\text{eq}} = 1 \mu\text{F}$ if they are in series.

Chapter 5

Current

The study of *electrodynamics* deals with moving charges. The *current* is the rate of flow of charges, expressed as

$$I = \frac{dQ}{dt}. \quad (5.1)$$

It has units of Ampere (A),

$$1 \text{ A} = 1 \frac{\text{C}}{\text{s}}. \quad (5.2)$$

5.1 Drift velocity

Consider a metal wire at a microscopic level. Because it is conducting, electrons are free to move around, and if there is no electric field, they move around randomly due to thermal energy. The thermal velocity is typically something like 10^6 m/s, but the direction is random. Once an external electric field is applied, the electrons will move in the opposite directions, as they are negatively charged. Suppose the electric field is applied along the wire, which has a cross sectional area A . The electrons do not move with a uniform speed, because they will constantly collide into atoms. Say the electrons are moving with an average velocity, or *drift velocity*, v_d against the field direction. If there are n freely moving charge carriers per unit of volume, we can write the total charge in the wire as.

$$\begin{aligned} \Delta Q &= \left(\frac{\# \text{charge carriers}}{\text{volume}} \right) (\text{volume})(\text{carrier charge}) \\ &= nVe. \end{aligned}$$

Because the volume can be expressed as $V = xA$, where the charges move a length of $x = v_d t$ after some time t ,

$$\Delta Q = n(v_d t A)e. \quad (5.3)$$

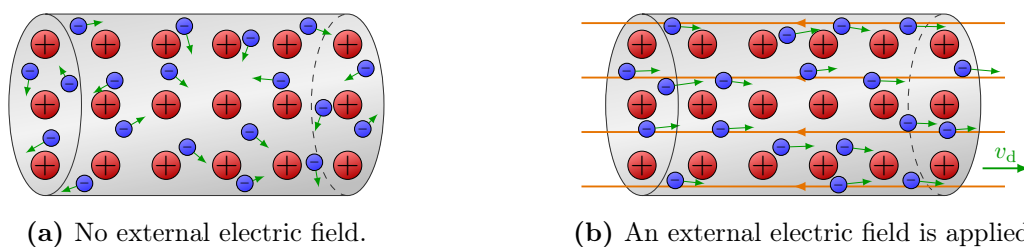


Figure 5.1: Current in a wire. Valence electrons are free to move around in a metal. When an external field is applied, all electrons move in the opposite direction of the field.

The current then, becomes

$$I = neAv_d. \quad (5.4)$$

Or equivalently, the drift velocity can be expressed as

$$v_d = \frac{I}{neA}. \quad (5.5)$$

Let's make this concrete. Assume a current of 1 A, a charge density similar to Avogadro's number, $n \sim N_A = 6.02 \times 10^{23}$ atoms/mol and a copper wire of radius $r = 1$ mm. Because copper has mass density $\rho = 8.93$ g/cm³, molar mass $M = 63.5$ g/mol, and one free electron per atom, the charge density per unit of volume is

$$n = \frac{\rho N_A}{M} \approx 8.5 \times 10^{28} \frac{\text{atoms}}{\text{m}^3}. \quad (5.6)$$

Now we can compute the drift velocity in a wire with area $A = 2\pi r$:

$$v_d \approx 0.02 \frac{\text{mm}}{\text{s}}. \quad (5.7)$$

This is surprisingly slow! It takes an electron on average more than 8 minutes to cross only 1 cm. But why then is electricity so fast when you flip a light switch? One way to think about it is like the electrons form an incompressible fluid: When you turn on the faucet to a long hose, the water flows immediately out of the other end (if there was still water in there). More importantly, the electric field propagates at the speed of light, so charges all along a conducting wire start moving instantly and simultaneously.

5.2 Current density vector

We can also write the current as a vector

$$\mathbf{I} = neA\mathbf{v}_d. \quad (5.8)$$

We sometimes express the current with the charge density $\rho = ne$. So

$$\mathbf{I} = \rho A\mathbf{v}_d, \quad (5.9)$$

which allows us to define the *current density vector*, which is the current per unit area:

$$\mathbf{j} = \frac{\mathbf{I}}{A} = \rho\mathbf{v}_d, \quad (5.10)$$

such that the current through a general surface is

$$I = \int_A \mathbf{j} \cdot d\mathbf{A}. \quad (5.11)$$

In general, current can be due to positive or negative charge carriers, or both at the same time. So the effective charge density is simply the sum of the positive and negative contributions:

$$\begin{aligned} \rho &= \rho^+ + \rho^- \\ &= n^+q^+ + n^-q^-, \end{aligned}$$

where n^\pm are the charge densities, and q^\pm are the charges. In the case of electrons, $q^- = -e$. So we can express the charge density current as

$$\mathbf{j} = \rho^+\mathbf{v}_d^+ + \rho^-\mathbf{v}_d^-, \quad (5.12)$$

where the positive and negative charges can have different drift velocities. Note that the convention of \mathbf{j} and also \mathbf{I} , is the direction a positive charge would go. This means that in most daily life applications, electrons actually move oppositely to the defined current.

5.3 Continuity equation

Using Gauss's law, the total current through a closed surface can be written as

$$I = \oint_S \mathbf{j} \cdot d\mathbf{A}. \quad (5.13)$$

We can use the fundamental theorem of divergences Eq. (2.36) to rewrite the first integral as

$$\oint_S \mathbf{j} \cdot d\mathbf{A} = \oint_V \nabla \cdot \mathbf{j} dV. \quad (5.14)$$

Meanwhile the current can also be written as

$$I = \frac{dQ}{dt} = \frac{d}{dt} \int \rho dV, \quad (5.15)$$

and therefore,

$$\oint_V \nabla \cdot \mathbf{j} dV = \frac{d}{dt} \int \rho dV. \quad (5.16)$$

Or in a differential form,

Continuity equation (charge conservation).

$$\nabla \cdot \mathbf{j} = -\frac{\partial \rho}{\partial t}. \quad (5.17)$$

This is called the *continuity equation*, and can be interpreted as a statement of charge conservation in our universe. Any current flowing through some point means charge changing over time in that point. The negative sign means that an outward flow of charge (out of a volume) causes the remaining charge (in a volume) to decrease. No charge can be created or destroyed in our universe, as this equation does not contain a so-called *source* or *sink* term.

5.4 Resistance

Consider again the conducting wire with an electric field along its length, like in Fig. 5.2. Because there is an electric field E , there also has to be a potential difference between any two points, say a and b , along the wire:

$$V = V_a - V_b = EL. \quad (5.18)$$

(Remember $V_a - V_b > 0$, because the electric field points from high to low potential.) For most materials, the current is proportional to the potential difference, $I \propto V$. These materials are called *ohmic*. The constant of proportionality is $1/R$, where R is the *resistance* of the material. Put together:

Ohm's law.

$$I = \frac{V}{R}, \quad (5.19)$$

or,

$$V = IR. \quad (5.20)$$

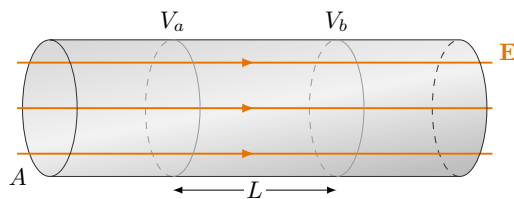


Figure 5.2: A wire with cross sectional area and a potential difference $V = V_a - V_b$ over a length L .

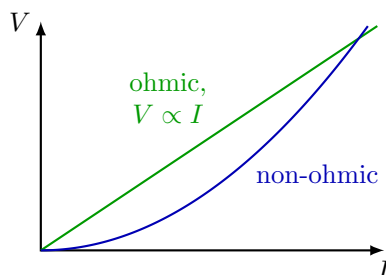


Figure 5.3: IV curve for Ohmic and non-ohmic resistance.

Resistance has units of Ohm (Ω), or

$$1 \Omega = 1 \frac{\text{V}}{\text{A}}. \quad (5.21)$$

Figure 5.3 shows the so-called *current-voltage characteristic*, or *IV curve*, of an ohmic resistor and an example of a non-ohmic resistor.

5.5 Resistivity and conductivity

The resistance can also be expressed in terms of the wire's length L and cross-sectional area A as

$$R = \frac{\rho L}{A}, \quad (5.22)$$

where ρ is the *resistivity*. (Not to be confused with the charge volume-density ρ .) The resistivity can be temperature dependent:

$$\rho = \rho_0 [1 + \alpha(T - T_0)], \quad (5.23)$$

where ρ_0 is the resistivity at T_0 , e.g. room temperature 20°C , and α is the temperature coefficient. This is plotted in Fig. 5.4. The $\alpha(T - T_0)$ term is a temperature-dependent correction factor to ρ_0 .

Table 5.1: Resistivity ρ and temperature coefficient α .

material	ρ [Ωm]	α [K^{-1}]
copper	1.7×10^{-8}	3.9×10^{-3}
aluminium	2.8×10^{-8}	3.9×10^{-3}
kanthal (FeCrAl)	1.5×10^{-6}	
wood	$\times 10^8 - \times 10^{14}$	
glass	$\times 10^{10} - \times 10^{14}$	

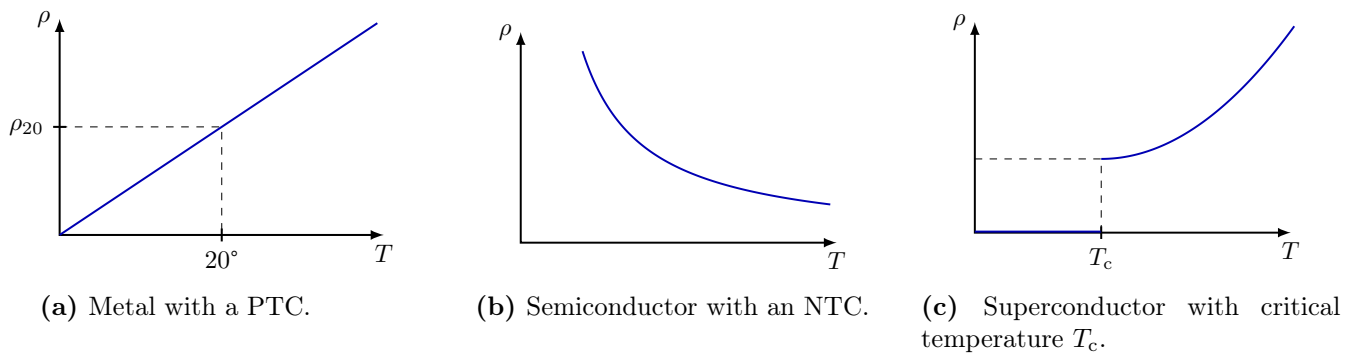


Figure 5.4: Temperature dependence of resistivity for different type of materials.

The *conductivity* is defined as the inverse of the resistivity:

$$\sigma = \frac{1}{\rho}. \quad (5.24)$$

If R , and therefore ρ , is small, then I is big. For example, for any give voltage, there will be more current in copper than in wood.

Resistivity can have many different temperature dependencies. Metals have a *positive temperature coefficient* (PTC), α , while semi-conductors can have *negative temperature coefficient* (NTC). For a NTC, the resistance decreases at higher temperatures, as electrons are held more loosely and can start to move. A special case is so-called *superconductors*. The superconductor's resistivity drops suddenly to zero below some critical temperature T_c . These materials with different types of resistivity are plotted in Fig. 5.4.

The current \mathbf{j} is along the electric field \mathbf{E} , and the potential difference is $V = EL$ over a length L in a wire. One can therefore find an interesting symmetry between the scalar and vector expressions:

$$V = IR \quad \longleftrightarrow \quad \mathbf{E} = \mathbf{j}\rho, \quad (5.25)$$

with the electric field \mathbf{E} .

5.6 Energy loss in a conductor

Energy will be lost in a conductor as electrical energy is converted to thermal energy, due to random collision of the moving charges. This loss of energy creates a potential decrease $V = V_a - V_b$. The loss of energy can therefore be written as

$$\Delta U = \Delta Q(V_a - V_b). \quad (5.26)$$

In some time Δt , the power can be expressed as

$$P = \frac{\Delta U}{\Delta t} = \frac{\Delta Q}{\Delta t} V \quad (5.27)$$

The power here, carries the usual units of Watts (W). We recognize the current on the left side,

$$P = IV. \quad (5.28)$$

So Watts can also be expressed as

$$1 \text{ W} = 1 \text{ A} \cdot \text{V} \quad (5.29)$$

For an ohmic resistance, $V = IR$, so P can also be written in terms of only I and R ,

$$P = I^2 R, \quad (5.30)$$

or only V and R :

$$P = \frac{V^2}{R}. \quad (5.31)$$

So the energy loss depends on the resistance R : For a fixed current I , a higher resistance R generates more heat. But for a fixed voltage V , a higher resistance R means that less heat is generated. We'll see this in the next chapter.

Chapter 6

Circuits

If a conductor has a non-zero electric field inside, $\mathbf{E} \neq 0$, it must be part of a complete loop, or *circuit*, where something is causing an difference in potential energy. If we just induce electric charge, the field will eventually become $\mathbf{E} = 0$ (*electrostatic*) due to the charge re-distributions, as discussed in Section 2.5. But if we add a source of energy, then we can create a current flow. This is called an *electromotive force* (EMF). It creates a potential difference \mathcal{E} , between either ends of the circuit to cause a non-zero field and charge current.

Intuitively, you can think of such an electric circuit as a pipe with water moving from a higher place to a lower one, i.e. from a larger to a lower gravitational potential. The EMF here, acts as a pump.

6.1 Resistors in circuits

Consider a circuit that contains a resistor R , and an EMF \mathcal{E} as in Fig. 6.1a. There will be a current I flowing along the circuit, and a potential difference V over the resistor. This is however an ideal case. In practice, the source of EMF has an internal resistance r that is non-zero, like in Fig. 6.1b. Therefore, the real potential difference V over resistor R is slightly smaller:

$$V = \mathcal{E} - Ir. \tag{6.1}$$

The voltage along the circuit is plotted in Fig. 6.2. The EMF increases the voltage, while the resistance lowers it.

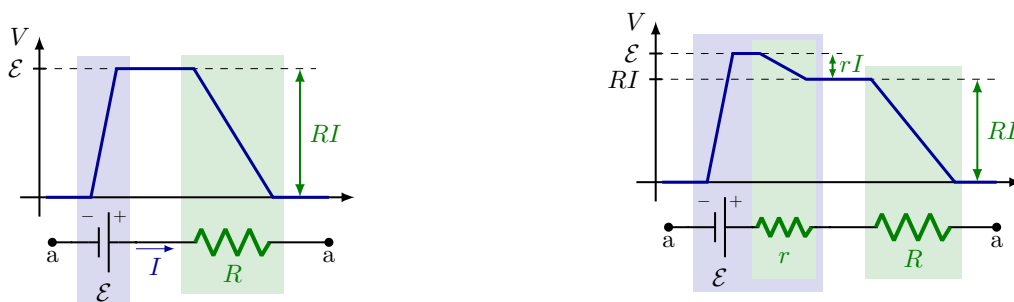
Example 6.1 (9V battery): Consider a 9 V battery with an internal resistance of 1Ω . If it



(a) Simple circuit with an EMF creating a potential \mathcal{E} and a resistor R .

(b) Circuit with an EMF and its internal resistance r , and a resistor R and voltmeter in parallel.

Figure 6.1: Simple circuit.



(a) Simple circuit with EMF and single resistor.

(b) EMF with internal resistance r .**Figure 6.2:** Plot of voltage in an electric circuit.

were connected to a resistor with $R = 11\ \Omega$, what would V be? We have

$$\mathcal{E} - Ir = V = IR.$$

Solving for V , we find that

$$V = \frac{\mathcal{E}R}{r + R} = 8.25\ \text{V}. \quad (6.2)$$

We see that, the larger the resistor R is, the closer V to 9 V becomes.

6.2 Resistors in series

Consider two resistors, R_1 and R_2 , that are connected in series on one wire (Fig. 6.3b). If the electric potential before R_1 is V_a , then the potential after R_1 , but before R_2 , is

$$V_b = V_a - I_1 R_1. \quad (6.3)$$

After R_2 then,

$$V_c = V_b - I_2 R_2 = V_a - I_1 R_1 - I_2 R_2. \quad (6.4)$$

This means that the equivalent resistance is

$$R_{\text{eq}} = R_1 + R_2, \quad (6.5)$$

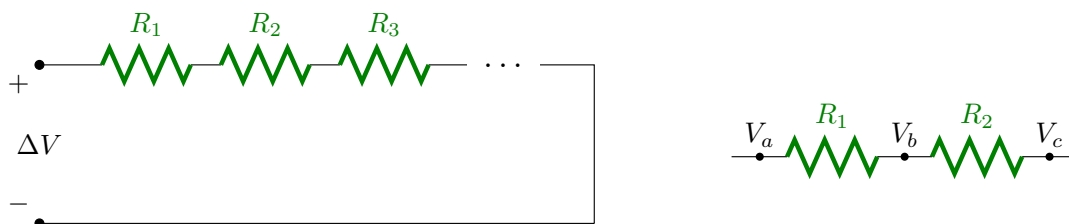
and the current stays constant $I = I_1 = I_2$ over each resistance, such that

$$V_c = V_a - R_{\text{eq}} I. \quad (6.6)$$

So generally,

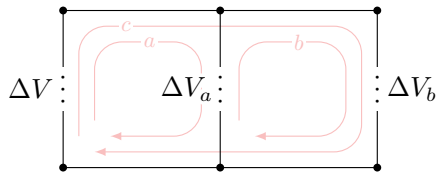
Resistors in series.

$$R_{\text{eq}} = R_1 + R_2 + \dots \quad (6.7)$$

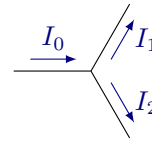
(a) Combining in series resistors R_1, R_2, R_3, \dots

(b) Close-up of two resistors.

Figure 6.3: Resistors in series.



(a) Rule 1: Total potential in a loop is zero.



(b) Rule 2: Current in equals current out.

Figure 6.4: Kirchhoff's rules.

6.3 Kirchhoff's rules

Before considering resistors in parallel, let's look at the Kirchhoff's rules.

Kirchhoff's rules. *Any simple circuit should follow these rules:*

1. Any complete loop has a total potential change of zero:

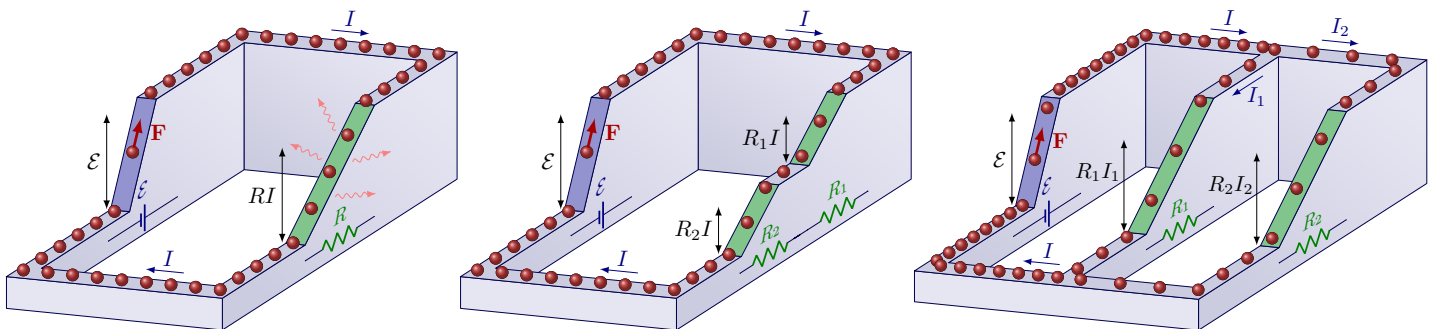
$$\sum \Delta V = 0. \quad (6.8)$$

2. The sum of currents into a junction must equal the sum of current going out:

$$\sum I_{in} = \sum I_{out}. \quad (6.9)$$

Rule 1 is illustrated in Fig. 6.4a, where there are three loops, a , b and c . Therefore, $\Delta V + \Delta V_a = 0$, $\Delta V + \Delta V_b = 0$, and $\Delta V_a + \Delta V_b = 0$. Rule 2 is illustrated in Fig. 6.4b, where $I_0 = I_1 + I_2$.

An intuitive way of thinking of the voltage and current in an electric circuit is shown in Fig. 6.5. The voltage is represented by a height difference, resulting in a difference in gravitational potential energy. The balls are like (positive) charge carriers flowing through the wire. The EMF raises them, while resistors lower them to extract the electric energy in the form of heat or light for example. The current I , on the other hand, is represented by the density of charge carriers flowing along the wires. In this analogy, the sum of the heights of two resistors must equal that of the EMF (Fig. 6.5b). In case of the two resistors in parallel in Fig. 6.5c, the sum of the currents I_1 and I_2 over each resistor must equal the total current I before the branching.



(a) Simple circuit with a single resistor, compare to Fig. 6.2.

(b) Circuit with a two resistors in series.

(c) Circuit with a two resistors in parallel.

Figure 6.5: An intuitive analogue for voltage and current in an electric circuit, using height for difference in gravitational potential energy. The EMF raises the (positive) charge carriers to increase their energy, while resistors convert the energy to heat or light.

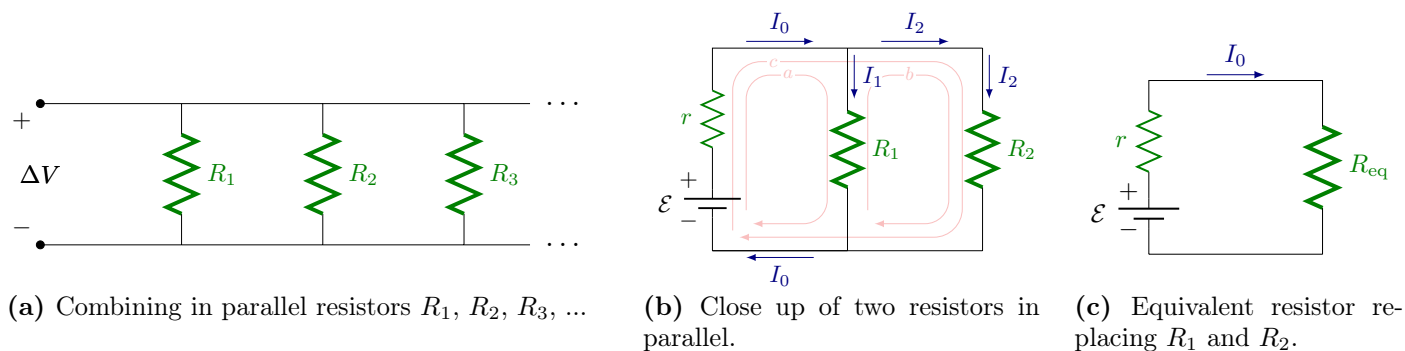


Figure 6.6: Resistors in parallel.

6.4 Resistors in parallel

Consider two resistors, R_1 and R_2 , that are in parallel, as in Fig. 6.6b. Using the second Kirchhoff's rule, we see that

$$I_0 = I_1 + I_2 \quad (6.10)$$

There are three loops in Fig. 6.6b, so applying Kirchhoff's first rule :

$$0 = \mathcal{E} - I_0 r - I_1 R_1, \quad (6.11)$$

$$0 = \mathcal{E} - I_0 r - I_2 R_2, \quad (6.12)$$

$$0 = -I_1 R_1 - I_2 R_2. \quad (6.13)$$

It is clear that $I_1 R_1 = I_2 R_2$. Simultaneously, Kirchhoff's law says that $I_0 = I_1 + I_2$. To replace the two resistors in parallel with one equivalent resistor R_{eq} , we need

$$0 = \mathcal{E} - I_0 r - I_0 R_{\text{eq}}, \quad (6.14)$$

or,

$$I_0 R_{\text{eq}} = I_1 R_1 = I_2 R_2. \quad (6.15)$$

From this, one can easily show that the equivalent resistance for R_1 and R_2 is

$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2}. \quad (6.16)$$

The total resistance of the circuit is then

$$R = r + R_{\text{eq}}, \quad (6.17)$$

where the R_{eq} and the internal r are now in series.

Generally,

Resistors in parallel.

$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \dots \quad (6.18)$$

Notice that the behavior of resistors in series or parallel is reversed with respect to capacitors.

6.5 Measuring circuits

There is always an error when measuring voltages and currents. Suppose you want to measure the circuit in Fig. 6.7. Ideally, the current meter, or *ammeter*, A should have no internal resistance $R_A = 0$, so that it does not change the current. The voltmeter V on the other hand, should have an infinite resistance $R_V = \infty$, such that no current through it, and it does not change the voltage.

If we disconnect voltmeter V, the current measured by the ammeter A goes up. This means that the V actually has a finite resistance. On the other hand, if we bypass ammeter A, by connecting a simple wire to each of its end, we see that the voltage measured by the voltmeter goes up.

6.6 RC circuits

Let us look at circuits that contain a resistor and a capacitor.

6.6.1 RC discharge

Suppose a circuit contains a resistor, a capacitor with charge Q_0 and capacitance C , and an open switch, as in Fig. 6.8a. Once we close the switch, the circuit is complete, and the charge between the capacitor can flow as a current through the switch and resistor. This is called a *RC discharge*. The initial current is

$$I_0 = \frac{V_0}{R} = \frac{Q_0}{CR}, \quad (6.19)$$

but as the charge on the capacitor decreases, such that there is a current:

$$I = -\frac{dQ}{dt}, \quad (6.20)$$

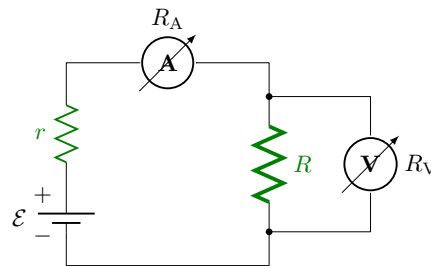
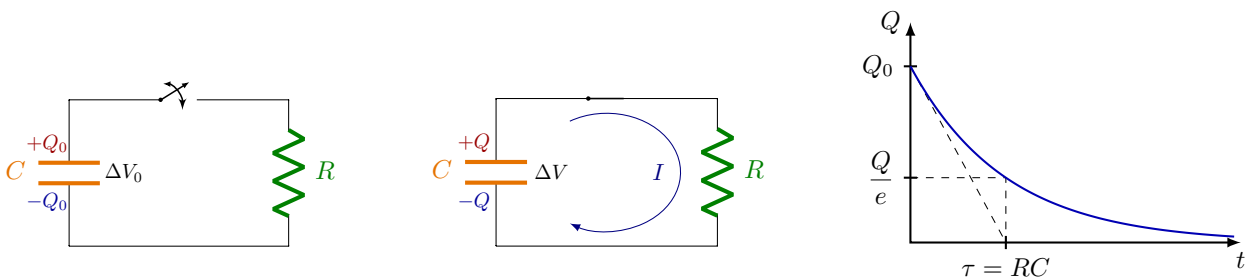


Figure 6.7: Measuring a circuit with an ammeter and voltmeter.



(a) Open circuit: No current and capacitor has initial charge Q_0 .

(b) Closed circuit: Current flows and C discharges.

(c) Plot of Q vs. time once the circuit is closed.

Figure 6.8: Discharging a charged capacitor C in a simple RC circuit.

where the minus sign indicates the current decreases.

Using Kirchhoff's rule on the entire loop, we see that

$$0 = \frac{Q}{C} - IR, \quad (6.21)$$

where both Q and I depend on time, while C and R are constant. Using the definition of current I ,

$$0 = \frac{Q}{C} - R \frac{dQ}{dt}. \quad (6.22)$$

We can write that

$$\frac{dQ}{Q} = -\frac{1}{RC} dt, \quad (6.23)$$

which we can integrate to

$$\ln Q = -\frac{t}{RC} + A, \quad (6.24)$$

Where A is some integration constant. We therefore find an explicitly time-dependent expression for Q :

$$Q(t) = e^{-\frac{t}{RC} + A} \quad (6.25)$$

Because we can isolate A from the argument as a simple multiplicative factor, we can write that

$$Q(t) = Q_0 e^{-\frac{t}{RC}}, \quad (6.26)$$

where we used the initial condition $Q_0 = Q(0)$. The denominator $\tau = RC$ is the *time constant* of the RC circuit. This is the time it takes for charge to decrease to $1/e$ of its original charge, see Fig. 6.8c. This is also the time to discharge if the rate were constant from $t = 0$, indicated in Fig. 6.8c by the straight, dashed line from $(t, Q) = (0, Q_0)$ to $(\tau, 0)$.

How about the current I ? By deriving Eq. (6.26), we see that

$$I = \frac{Q_0}{RC} e^{-\frac{t}{\tau}}, \quad (6.27)$$

or,

$$I = I_0 e^{-\frac{t}{\tau}}, \quad (6.28)$$

with

$$I_0 = \frac{Q_0}{RC} = \frac{V_0}{R}. \quad (6.29)$$

I decreases with time in the exact same way as Q .

6.6.2 Charging a RC circuit

Now imagine we charge the previous RC circuit again, by adding a EMF source as in Fig. 6.9a. If we close the switch, Kirchhoff's first rule gives us

$$0 = \mathcal{E} - IR - \frac{Q}{C}. \quad (6.30)$$

Using again the fact that I is a derivative of Q ,

$$\mathcal{E} = R \frac{dQ}{dt} + \frac{Q}{C}. \quad (6.31)$$

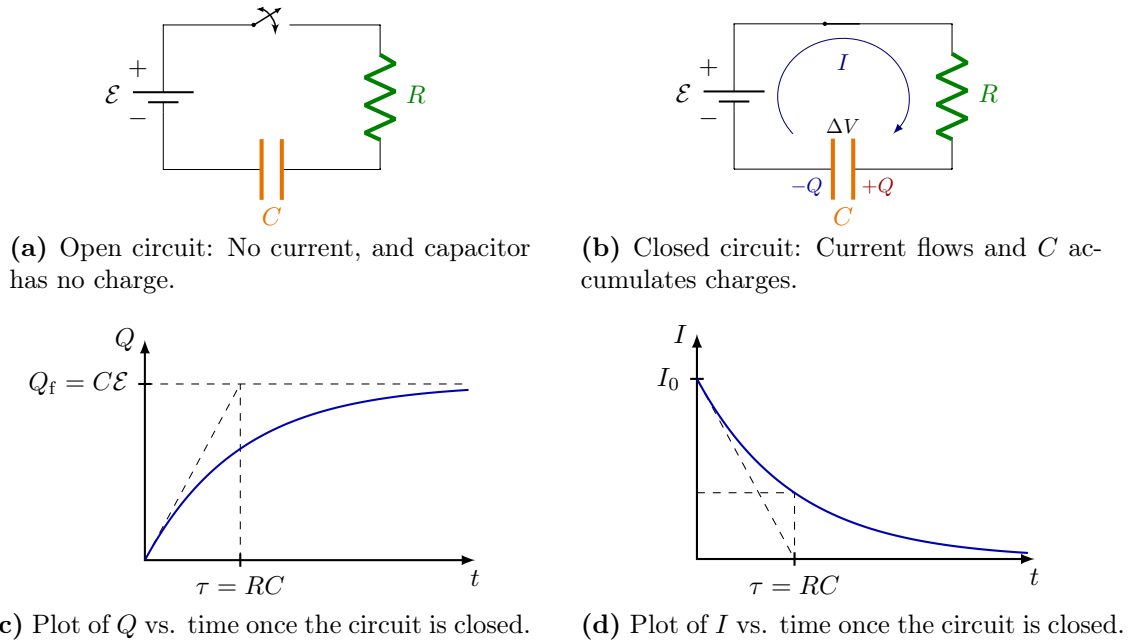


Figure 6.9: Charging a capacitor C with EMF \mathcal{E} .

Rewriting to a differential equation that is easy solve,

$$\frac{dQ}{C\mathcal{E} - Q} = \frac{dt}{RC}. \quad (6.32)$$

The solution is

$$Q(t) = C\mathcal{E} - Be^{-\frac{t}{RC}}. \quad (6.33)$$

At time $t = 0$, capacitor C has charge 0, so

$$Q(t) = C\mathcal{E}(1 - e^{-\frac{t}{RC}}). \quad (6.34)$$

Figure 6.8c shows a plot of this exponential behavior. The capacitor will charge and converge to a final charge of $Q_f = C\mathcal{E}$.

The current's behavior will still look very similar like in the case of discharging, i.e. an exponential decay:

$$I = I_0 e^{-\frac{t}{\tau}}, \quad (6.35)$$

with

$$I_0 = \frac{\mathcal{E}}{R}. \quad (6.36)$$

6.7 Application: Tubular capacitors and Geiger counters

Imagine you have a conductive tube with gas inside. We pull a conducting wire through the tube, which is at a negative potential compared to the tube. The tube is then a cathode (negative potential), while the wire is an anode (positive potential).

Imagine you have some radiation, like a charge particle, going through the tube (“ionizing event”). As it passes through the gas, it ionizes the gas particles, creating free electrons and ions (“ion pairs”). Because there is a potential difference between the tube and wire, there will be a radial electric field between them, going as

$$E \propto \frac{1}{r}. \quad (6.37)$$

Due to this field, the free electrons will move towards the anode. As the field gets stronger, close to the wire, the electrons ionize more gas particles, creating an avalanche of electrons. Once this avalanche hits the wire, it will create a significant pulse on the anode. A single ionizing event can be amplified to 10^6 – 10^8 ion pairs. More things can happen: photons can be emitted by the interaction between the electrons and gas particles. This is the principle of a Geiger counter, which is used to measure the amount of radiation.

The dose of radiation received by some material, is measured as the sum of kinetic energy of charged particles, divided by the material mass. The unit of *radiation dose* is Gray, or

$$1 \text{ Gray} = 1 \frac{\text{J}}{\text{kg}} \quad (6.38)$$

Now consider a non-conducting tube filled with gas, and with an anode and cathode plate on either end. If we apply a strong potential difference between the plate, and lower the pressure, the electrons can acquire enough energy to ionize the gas. This causes a glowing effect from fluorescence of this gas. If we do this experiment with a glass tube, we see many glowing fringes will form. This distance d between the the fringes are related to the *mean free path* of the electron. There is a simple relation between the size and the gas pressure as

$$d \propto \frac{1}{P}. \quad (6.39)$$

The glow voltage V is proportional to the pressure and distance D between the anode and cathode,

$$V \propto PD. \quad (6.40)$$

As the gas pressure in the tube is decreased, the distance must increase, in order to get a glowing effect, so we see the path of the fringes to longer and longer total path lengths as we decrease pressure P .

Chapter 7

Magnetism

Magnetism is a phenomenon that, as we will see, is intimately related to movement of electric charges (i.e. currents). For now, let us start with some basic observations when we entertain ourselves with a some classic bar magnets and some nails, like in Fig. 7.1. Bar magnets are *dipole magnets*: They have a *south* and *north pole*. We observe that the different pole types attract (south-north), while the same types repel (north-north, south-south). Non-magnetic nails, on the other hand, are attracted to both poles.

Just like electric charges create an electric vector field around them, magnets create a *magnetic field*, denoted by \mathbf{B} . The field lines of the field of magnetic dipoles look like in Fig. 7.2a.

Magnetic field lines point from magnetic north poles to south poles, so we see them appear to originate in North and return into South. The north of a compass points in the

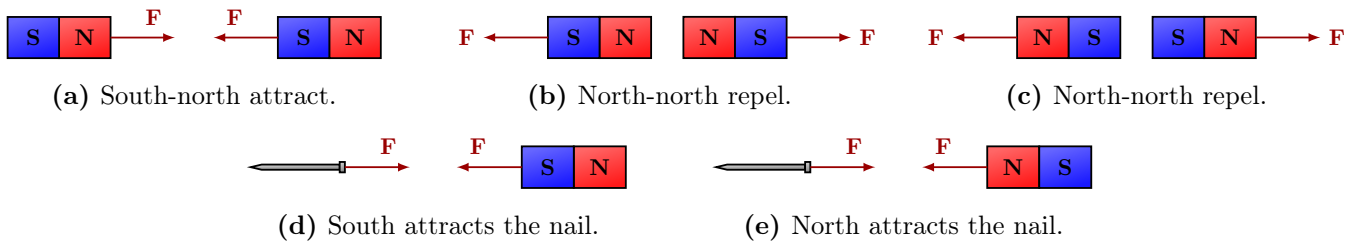
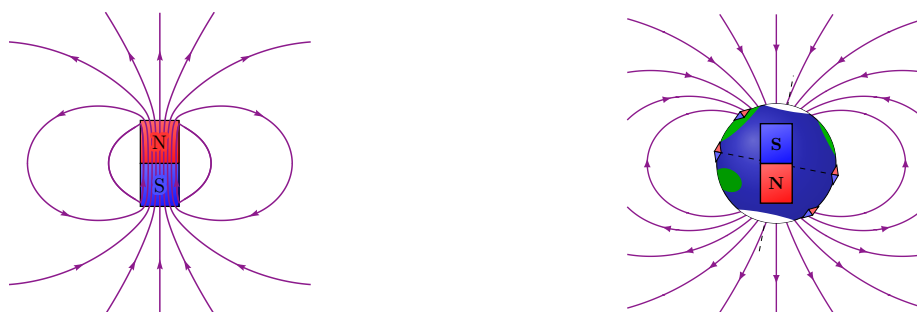


Figure 7.1: The magnetic force between two bar magnet depends on their orientation, but between a non-magnetic nail and bar magnet, orientation does not matter.



(a) The magnetic field of bar magnet looks like the electric field of an electric dipole. The field lines close their loops inside the bar magnet.

(b) Earth's magnetic field looks like that of a bar magnet. Magnetic compasses point to Earth's geographic north pole, the magnetic south pole.

Figure 7.2: Bar magnets and the Earth create a magnetic dipole field (purple).

vector direction of the magnetic field (so from north to south).

The Earth has a magnetic field that is created by convection currents of molten iron in Earth's outer core. It turns out Earth's field is similar to that of a magnetic dipole. The north of a magnetic compass must point to the south pole of another magnet. This is confusing because then why does a compass always point to the Earth's North Pole. It turns out that Earth's *geographical north pole* is actually Earth's *magnetic south pole*, and vice versa.

7.1 Gauss's law for magnetic fields

The magnetic dipole field may remind you of the electric dipole field created by two opposite charges (Fig. 1.11a). This begs the question: Are there "magnetic charges"? However, it turns out that so-called *magnetic monopoles* have never been discovered. North and south of a magnet cannot be separated. If you try to break a magnet in half, you end up with two smaller magnets, each having their own north and south pole. Since there are no magnetic monopoles, the magnetic field lines have no beginning or end: Even though the field lines go from north to south, they always complete a full (i.e. closed) loop inside the magnet. Whereas in an electric dipole, the E-field lines have a start (in positive charges) and an end (in negative charges).

Because Gauss's law (Eq. (2.14) or (2.40)) is a consequence of a more general theorem in vector calculus, it also is valid for magnetic fields, except this time there are no "magnetic charges" which act like sources of the vector field.

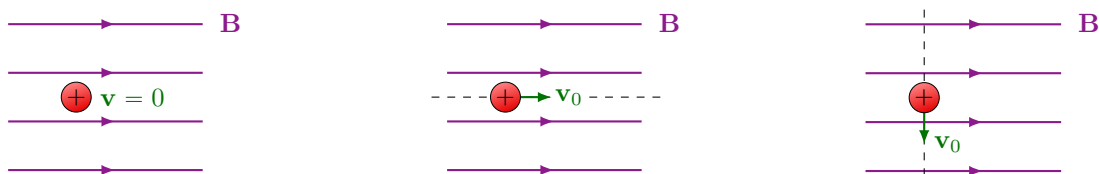
Gauss's law for magnetic fields. *The magnetic flux through a closed surface S is always zero:*

$$\Phi_M = \oint_S \mathbf{B} \cdot \hat{\mathbf{n}} \, dA = 0, \quad (7.1)$$

or in vector notation:

$$\nabla \cdot \mathbf{B} = 0. \quad (7.2)$$

The way to interpret these statements is that magnetic charges do not exist ($\rho_M = 0$), or equivalently, magnetic field lines always form closed loops: Because the magnetic flux through a closed surface is always zero, the same number of magnetic field lines enter as leave any closed surface. This is the second of Maxwell's four equations of electromagnetism, which we will come back to later.



(a) Charge at rest experiences no force. (b) Charge moving along the magnetic field experiences no force. (c) Charge moving across a magnetic field experiences a force.

Figure 7.3: Magnetic forces on a charge depends on the velocity relative to the magnetic field.

7.2 Magnetic force on an electric charge

Another big difference with electric fields is the way a magnetic field exerts a force on an electric charge. The electric force \mathbf{F} is always in the same direction as the electric field \mathbf{E} , and only depends on the field strength and charge. In the case of magnetic fields, the force also depends on the velocity. If a charge is at rest, it experiences no magnetic field (Fig. 7.3a). The same when a charge is moving parallel to the magnetic field (Fig. 7.3b). If it makes some non-zero angle $\theta \neq 0$, however, it does experience a magnetic force. This force depends only on the charge q , velocity \mathbf{v} , magnetic field \mathbf{B} in the following way:

Magnetic force on a charge.

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}, \quad (7.3)$$

The magnetic force has size

$$F_B = |q|vB \sin \theta. \quad (7.4)$$

The force is maximum if the \mathbf{v} is perpendicular to \mathbf{B} , $\theta = \pi/2$:

$$F_B = |q|vB. \quad (7.5)$$

The direction of the force can be found with the right-hand rule, illustrated in Fig. 7.4: Point your flat hand in the direction of velocity \mathbf{v} . Keeping your index finger along \mathbf{v} , point your middle finger along \mathbf{B} , sweeping the smallest angle θ between \mathbf{v} and \mathbf{B} . Your thumb will point along $\mathbf{v} \times \mathbf{B}$. This is the direction of force \mathbf{F}_B for a positive charge ($q > 0$)

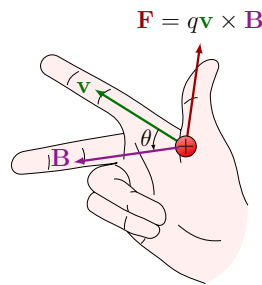
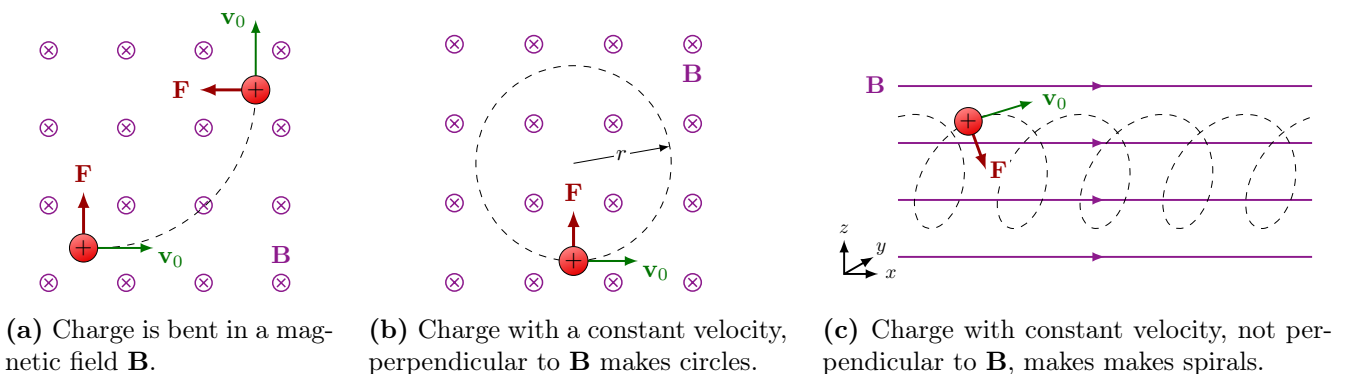


Figure 7.4: Right-hand rule for the magnetic force on a positive charge $q > 0$.



(a) Charge is bent in a magnetic field \mathbf{B} .

(b) Charge with a constant velocity, perpendicular to \mathbf{B} makes circles.

(c) Charge with constant velocity, not perpendicular to \mathbf{B} , makes spirals.

Figure 7.5: Charge with a non-zero velocity, not parallel to a uniform magnetic field \mathbf{B} , experiences a force perpendicular to the velocity and magnetic field.

point. For a negative charge ($q < 0$) the force is flipped, so you could exceptionally use your left hand instead.

By the properties of the cross product, the force \mathbf{F}_B is perpendicular to both \mathbf{v} and \mathbf{B} . So a charge moving not in parallel to a magnetic field, will be bent, and the magnetic force will not change the speed $v = |\mathbf{v}|$. This is a centripetal acceleration, so a charge with a constant velocity \mathbf{v}_0 , perpendicular to a uniform magnetic field will make full circles, as in Fig. 7.5b. Remember from last semester that the centripetal acceleration a due to a centripetal force F_\perp is given by

$$F_\perp = ma = \frac{mv^2}{r}. \quad (7.6)$$

So for a charge with \mathbf{v} perpendicular to \mathbf{B} ,

$$F_B = qvB = \frac{mv^2}{r}, \quad (7.7)$$

and therefore, the radius is

$$r = \frac{mv}{qB}, \quad (7.8)$$

and the angular velocity is

$$\omega = \frac{v}{r} = \frac{qB}{m}, \quad (7.9)$$

If the constant velocity is not completely perpendicular to \mathbf{B} , the charge will spiral in a circular helix instead, see Fig. 7.5c. In this picture, the uniform field $\mathbf{B} = B\hat{\mathbf{x}}$ is only in the x direction, while the velocity has a non-zero component \mathbf{v}_\perp perpendicular to \mathbf{B} field (i.e. in the yz plane) and some non-zero x component, \mathbf{v}_x . The helical path will be of the form

$$\mathbf{r} = r \cos(\omega t)\hat{\mathbf{y}} + r \sin(\omega t)\hat{\mathbf{z}} + v_x t\hat{\mathbf{x}}, \quad (7.10)$$

where r is given by Eq. (7.8), and ω by Eq. (7.9) for the transverse component v_\perp .

The dimension of the magnetic field are force over charge and speed:

$$1 \text{ T} = 1 \frac{\text{N}\cdot\text{s}}{\text{C}\cdot\text{m}}. \quad (7.11)$$

where Tesla (T) are the SI units of magnetic fields.

7.2.1 Lorentz force

In general, an electric charge q can move through an electric *and* magnetic field, \mathbf{E} and \mathbf{B} respectively. This is called the *Lorentz force*:

Lorentz force.

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}. \quad (7.12)$$

7.2.2 Application: Velocity selector

The velocity-dependence of the magnetic force is very useful if you have a beam of particles with charge q (like electrons) and varying velocity, but you only want to keep the ones with some velocity \mathbf{v} . This can be achieved by balancing the electric and magnetic force, but only for that specific velocity \mathbf{v} . An example of such a *velocity selector*, or *Wien filter*, is a plate capacitor in a uniform magnetic field like in Fig. 7.6a. The uniform electric field of

the capacitor is perpendicular to the magnetic field, so any charges q moving between the plates, will experience a Lorentz force, where the electric and magnetic components will point in opposite directions:

$$\sum F = qvB - qE. \quad (7.13)$$

By carefully tuning the two fields, the forces are balanced, but only for one specific velocity \mathbf{v} ,

$$qvB = qE. \quad (7.14)$$

Therefore the selected velocity is given precisely by

$$v = \frac{E}{B}. \quad (7.15)$$

It only depends on the electric and magnetic field strengths. If the velocity is smaller or larger than this, it will bent away from the beam axis, while the charges with exactly the right speed, experience a net zero force, and will follow a straight path.

7.2.3 Application: Mass spectrometer

A *mass spectrometer* is a device that can separate charged particles of different mass, like ions, out of a beam of charged particles. It consist of two stages: a velocity selector, and a particle detector. The velocity selector with electric and magnetic field E and B_1 makes sure the outgoing beam has a uniform velocity $v = E/B_1$. The beam then enters a different magnetic field B_2 that deflects the charged particles backward onto a particle detector. Because the radius given by Eq. (7.8) is proportional to the charge mass m , where the particles land, depend on their mass. The radius is therefore given by

$$r = \frac{mE}{qB_1B_2}, \quad (7.16)$$

such that higher-mass particles have a larger radius. With this technique, it was possible to discover stable isotopes, i.e. atoms of the same chemical element, and therefore the same number of protons, but with different number of neutrons. For example, about 90% of neon atoms in nature are the $^{20}_{10}\text{Ne}$ isotope, while about 9% are $^{22}_{10}\text{Ne}$.

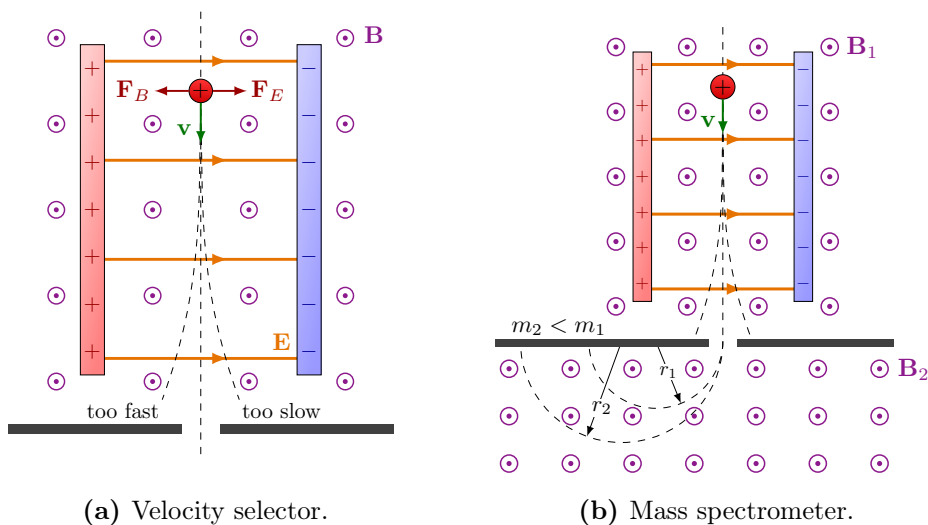


Figure 7.6: By balancing the electric and magnetic force, one can select for a specific velocity. One can also tell mass apart thanks to the different radii of curvature.

7.2.4 Application: Cathode-ray tube

Old television sets produced images with so-called *cathode-ray tubes*. These are devices that accelerate electrons and project them onto screens to create images. By using switching horizontal and vertical electric fields, the electrons can be steered horizontally and vertically to make a trace across the screen. The screen glows where the electrons strike. The electron beam is scanned across the screen so fast that the human eye sees the appearance of one full image.

In class, we do an experiment with a cathode ray, plus a switching electric field, plus a focusing magnetic field. A magnetic field parallel to the electron trajectory from the cathode ray will focus the electron beam back to a single point.

7.3 Currents in magnetic fields

Now consider if we have a bunch of charges moving in an external magnetic field, for example a wire carrying a current I , like in Fig. 7.7. Using the right-hand rule, we find that in this picture, with the magnetic field in the page, and the current to the right, the force indeed points upward. The whole wire experiences therefore a net upward force due to the magnetic field. Notice that this does not depend on the charge sign of the charge carriers.

What is the total force on a straight piece of wire with length ℓ ? If the wire has cross-sectional area A and charge-carrier density n , then the (averaged) net force is

$$\mathbf{F}_{\text{tot}} = (nA\ell)(q\mathbf{v}_d \times \mathbf{B}). \quad (7.17)$$

In Section 5.1, we had already derived that $I = nqv_dA$, so actually, we can write this in terms of the current I .

Magnetic force on an current.

$$\mathbf{F} = I\boldsymbol{\ell} \times \mathbf{B}. \quad (7.18)$$

Here, $\boldsymbol{\ell}$ points along the current. If $\boldsymbol{\ell} \perp \mathbf{B}$, simply $F = BI\ell$, which can be memorized with the name “Bill” as a mnemonic device.

Suppose we have a current-carrying conductor (wire) with some shape in a magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$, like the wire in Fig. 7.8. It consists of three segments, one anti-parallel to \mathbf{B} , one half-circle in the xy plane, and one parallel to \mathbf{B} . What would be the force on it? Segments 1 and 3 are parallel and anti-parallel, so the forces on them are zero. Segment 2 is a half circle in the xy plane, perpendicular to \mathbf{B} . Everywhere along the wire, there is a

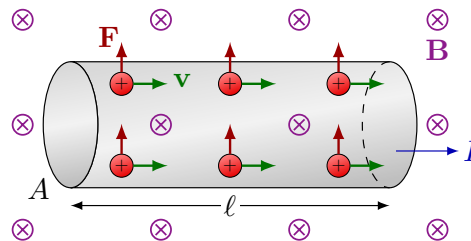


Figure 7.7: Wire carrying an electric current in a magnetic field experiences a force.

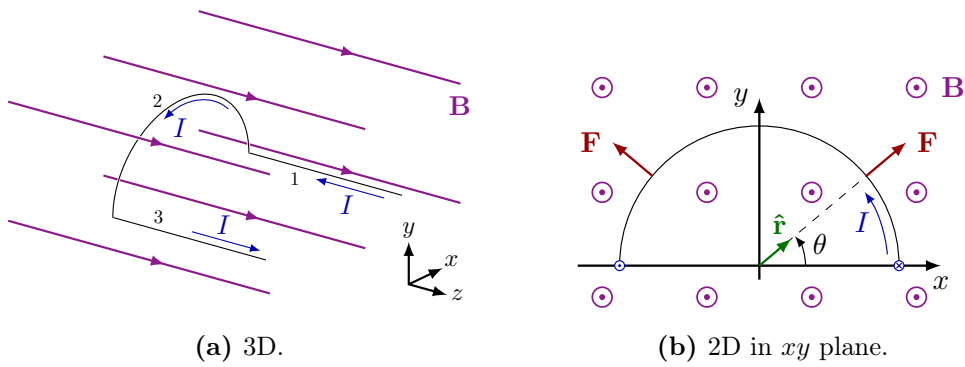


Figure 7.8: Half-circle wire with a current in an external, uniform magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$.

radially outward force, so we need to integrate

$$\mathbf{F}_2 = \int_{\ell_2} I d\boldsymbol{\ell} \times \mathbf{B} \quad (7.19)$$

over its length. Here, $I d\boldsymbol{\ell}$ is the direction a positive charge would go. It is a half-circle, so we can use polar coordinates: $\boldsymbol{\ell} = R d\boldsymbol{\theta}$, where the vector $d\boldsymbol{\theta}$ is along $\hat{\boldsymbol{\theta}}$, i.e. perpendicular to the radial unit vector $\hat{\mathbf{r}}$ and along $d\boldsymbol{\ell}$. The integral along the total length $\ell_2 = R\pi$ of segment 2 is

$$\begin{aligned} \mathbf{F}_2 &= \int_0^\pi (IR d\boldsymbol{\theta} \times \mathbf{B}), \\ &= \int_0^\pi IRB\hat{\mathbf{r}} d\theta. \end{aligned} \quad (7.20)$$

Because $\hat{\mathbf{r}} = \cos\theta\hat{\mathbf{x}} + \sin\theta\hat{\mathbf{y}}$, the x component is

$$F_x = \int_0^\pi IRB \cos\theta d\theta = 0 \quad (7.21)$$

as expected by symmetry. Meanwhile, the y component is non-zero:

$$F_y = \int_0^\pi IRB \sin\theta d\theta = 2IRB. \quad (7.22)$$

The wire will feel a net upward force

$$\mathbf{F} = 2IRB\hat{\mathbf{y}}. \quad (7.23)$$

7.3.1 Force on a current loop in a magnetic field

What if we have a complete loop in a magnetic field? Consider the rectangular loop in Fig. 7.9. Each side will have a force towards the center, and so they all cancel, and consequently, the total force is zero.

This turns out to be true in general:

The force a complete current loop in a constant magnetic field, is zero, regardless of the shape and angle with the magnetic field.

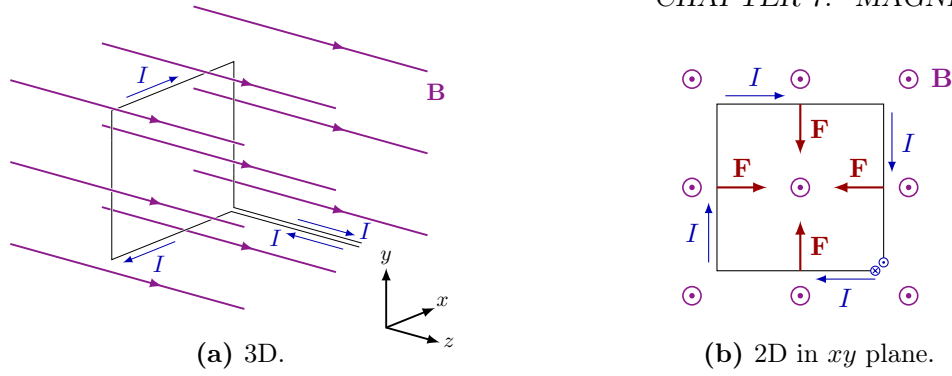


Figure 7.9: Rectangular current loop in an external, uniform magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$.

7.3.2 Torque on a current loop in a magnetic field

Instead of a net, linear force, consider that there still can be a torque $\boldsymbol{\tau} = \sum \mathbf{r} \times \mathbf{F}$ on a current loop.

Consider a rectangular loop again in a uniform magnetic field, as in Fig. 7.10. It has width w and height h . The vertical segments, 1 and 3, experience no force, because they are parallel to the field. The horizontal segment 2 feels a force $\mathbf{F}_2 = -IwB\hat{\mathbf{z}}$ going inside the page, and segment 4 feels a force $\mathbf{F}_2 = IwB\hat{\mathbf{z}}$ going outside the page. The forces on either sides are equal, but in opposite directions, so although the net force is zero, there is a torque. The loop will twist, or if it is rigid, rotate around some axis. In what direction is the torque? Using the right-hand rule, we find that the torque with respect to the center is

$$\begin{aligned}
 \boldsymbol{\tau} &= \mathbf{r}_1 \times \mathbf{F}_1 + \mathbf{r}_2 \times \mathbf{F}_2 \\
 &= -\left(\frac{h}{2}\right)(IwB)(\hat{\mathbf{x}} \times \hat{\mathbf{z}}) - \left(\frac{h}{2}\right)(IwB)(\hat{\mathbf{x}} \times \hat{\mathbf{z}}) \\
 &= -IwhB\hat{\mathbf{x}} \\
 &= -IAB\hat{\mathbf{x}},
 \end{aligned} \tag{7.24}$$

where the total area is the width w times height h , $A = wh$.

7.3.3 Magnetic moment

In general, we have the following.

Torque on a flat current loop. *If a flat current loop makes an angle θ between a uniform magnetic field \mathbf{B} everywhere,*

$$\boldsymbol{\tau} = IA\hat{\mathbf{n}} \times \hat{\mathbf{B}}, \tag{7.25}$$

where $\hat{\mathbf{n}}$ is the normal direction of the loop's plane.

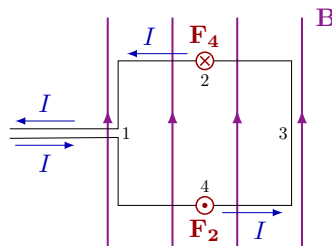
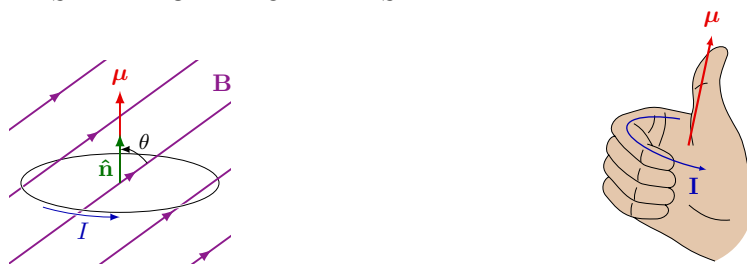


Figure 7.10: Rectangular current loop in an external magnetic field \mathbf{B} .



(a) Magnetic moment of a current loop in a uniform magnetic field.

(b) Right-hand rule for the magnetic moment of a current loop.

Figure 7.11: Magnetic moment.

The size of the torque depends on the angle between the loop's plane and the field, $IAB \sin \theta$. A useful concept for a loop current is the *magnetic moment*,

$$\boldsymbol{\mu} = IA\hat{\mathbf{n}}, \quad (7.26)$$

and shown in Fig. 7.11a. The direction of $\hat{\mathbf{n}}$ or $\boldsymbol{\mu}$ is given by the right-hand rule illustrated in Fig. 7.11b: One curls their fingers of their right hand along the current, and their thumb will point in the direction of $\boldsymbol{\mu}$. So we also have,

Torque on a magnetic moment.

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \hat{\mathbf{B}}. \quad (7.27)$$

Good to remember, is that the torque is always in the direction to align the magnetic moment $\boldsymbol{\mu}$ with \mathbf{B} . The potential energy of a current loop in a magnetic field becomes

Potential energy of a current loop in a magnetic field.

$$U = -\boldsymbol{\mu} \cdot \mathbf{B}. \quad (7.28)$$

Here, a constant was chosen, such that $U = 0$ if $\boldsymbol{\mu}$ is parallel to \mathbf{B} .

Notice that this is analogous to what we saw before for an electric dipole with electric moment \mathbf{p} in an electric field \mathbf{E} , which has a potential energy $U = -\mathbf{p} \cdot \mathbf{E}$ (Eq. (1.26)).

We can create a stronger magnetic moment, by coiling up many loops, like in a *solenoid*, see Fig. 7.12. If the solenoid has N windings (i.e. loops), the total magnetic moment is simply

$$\boldsymbol{\mu} = NIA\hat{\mathbf{n}}. \quad (7.29)$$

7.3.4 Non-uniform field

Up until now, we were assuming the magnetic field was uniform. What if the magnet field is non-uniform? Take for example the flat current loop in the non-uniform field of a simple bar magnet like in Fig. 7.13. If the south points to the loop, the vertical forces cancel out, and so the net force on the loop is towards the magnet. If we flip the magnet, north now pointing to the loop, the total force is now pointing away from the magnet.

**Figure 7.12:** Magnetic moment of a solenoid with N windings.

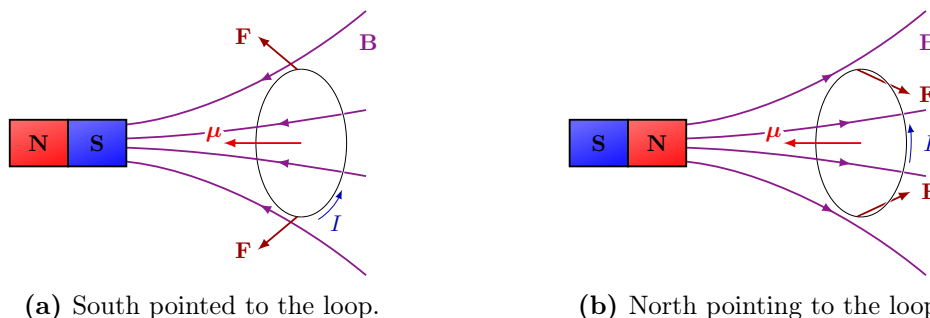


Figure 7.13: Magnetic moment of a current loop due to a non-uniform field.

7.4 Magnetization of matter

Electrons and atoms can be thought of as spinning electrical charges. Because the charged electrons are orbiting the nucleus of an atom, they create a total magnetic moment μ . Each electron itself also possesses a magnetic moment, caused by its *spin*. Spin in classical terms can be thought of as a spinning ball of charge, although this picture has some issues, as you will see in a course on quantum mechanics.

This helps us to understand why a non-magnetic nail is attracted to both the north and south ends of a magnet. This happens in a few stages. When there is no external magnetic field, magnetic momenta of the atoms inside the nail point in random directions. Still, small microscopic groups of atoms, called *domains*, with aligned magnetic momenta tend to form. Once you apply a magnetic field, most of these microscopic magnetic moments will align in one direction over the whole object, creating one macroscopic, net magnetic moment of the nail.

We see this works for both the south and north pole. However, a net force is only created if the nail is in a non-uniform magnetic field: In a uniform magnetic field, the induced magnetic moment will be exactly parallel to the field.

We will see magnetization in more detail in Chapter 11.



(a) Electrons orbiting around the atom form a current, giving the atom a magnetic moment μ .

(b) An electron can (classically) be thought of as a spinning ball of charge.

Figure 7.14: Elementary magnetic moments of matter at the atomic level.



(a) No field: The nail has microscopic domains (not to scale in figure) with magnetic moments pointing in random directions.

(b) A non-zero field causes the microscopic magnetic momenta to align, creating a net magnetic moment.

Figure 7.15: Magnetization of a nail.

7.5 Application: Hall effect

When a current-carrying conductor is placed in an external magnetic field, a voltage difference transverse to the current direction is produced. This is called the *Hall effect*. This phenomenon lets us figure out if charge carriers in a conducting material are positive or negative, and what its density of charge carriers is per unit of volume.

Consider for example the conductor in Fig. 7.16. The current goes horizontally to the right, while the external magnetic field goes into the page. The charge carriers, no matter if they are negative or positive, always drift upward, accumulating on the top.

To figure out the sign of the carriers, all we need to do is measure the potential difference between the top and bottom, to figure out which side has the higher potential. The potential difference is called V_H .

(Small disclaimer: In some semi-conductors, the current is carried by so-called *electron holes*, so the potential difference is actually opposite. However, for conductors like copper or aluminum, electrons are the charge carriers.)

To get the density of charge carriers, we note that the potential difference V_H creates an electric force opposite to the magnetic force \mathbf{F} . Therefore, a non-zero electric force is created inside the conductor. The charge carriers that flow through the conductor with voltage V_H , experience a magnetic and an electric force that balance precisely each other:

$$qv_d B = qE = q \frac{V_H}{w}, \quad (7.30)$$

if the voltage is over the width w of the conductor. If the conductor also has a thickness t , such that the conductor's cross-sectional area is $A = wt$ (perpendicular to the current), we find that the current is

$$I = nqv_d A = nqv_d wt \quad (7.31)$$

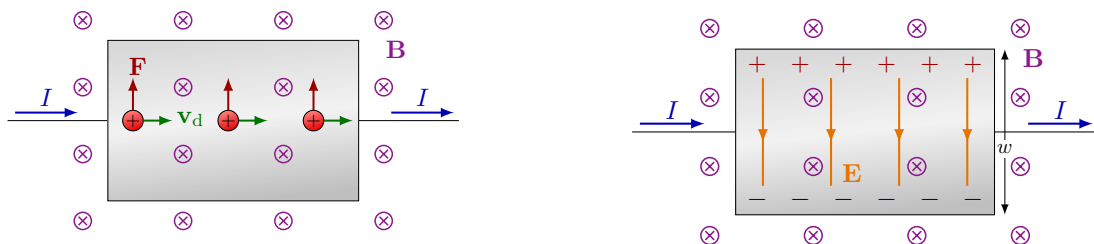
(see Section 5.1). Therefore, the charge density is

$$n = \frac{IB}{eV_H t}, \quad (7.32)$$

If we measure V_H , we can compute the charge carrier density n . Note also that

$$V_H = \frac{IB}{ent}, \quad (7.33)$$

so we could use a conducting strip of known charge density n and thickness t to measure the strength of a magnetic field. This is a *magnetometer*, also called a *Hall probe*.



(a) Positive charge carriers experience an upward force. (Note negative forces do as well). (b) Positive charge carriers accumulate on the top, while the bottom becomes negatively charged.

Figure 7.16: Hall-effect in a current-carrying conductor with positive charge carriers in an external magnetic field, resulting in a **voltage difference** transverse to the magnetic field.

Chapter 8

Laws of Magnetism

In the introduction chapter on magnetism, we had already seen the Lorentz force on a charge in a electromagnetic field, and Gauss's law for magnetic fields. In this chapter, we will see some other useful laws of magnetism, like Biot-Savart's law, Ampère's law Faraday's law and Lenz's law.

8.1 Biot-Savart's law

8.1.1 Single, moving charge

A moving charge q generates its own magnetic field \mathbf{B} . The direction of the field depends on the vectorial distance from the charge \mathbf{r} and its velocity \mathbf{v} : \mathbf{B} is perpendicular to both \mathbf{r} and \mathbf{v} , so the field lines form loops around the charge. Its magnitude decreases as $1/r^2$. The field is given by

Biot-Savart's law for a moving charge.

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{q\mathbf{v} \times \hat{\mathbf{r}}}{r^2}. \quad (8.1)$$

Here, μ_0 is the *permeability of free space*:

$$\mu_0 = 4\pi \times 10^{-7} \frac{\text{T}\cdot\text{m}}{\text{A}} = 1.26 \times 10^{-6} \frac{\text{T}\cdot\text{m}}{\text{A}}. \quad (8.2)$$

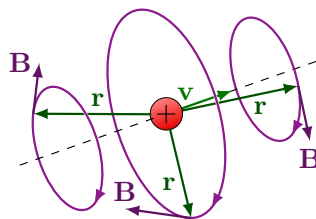


Figure 8.1: Magnetic field of a moving charge. The magnetic field loops around the charge, perpendicular to its motion. Notice the magnetic field in front and behind its position is smaller (but not zero).

8.1.2 Current

A current is just moving charge, so this can be translated to some current flowing over some path given by ℓ . A small amount of $I d\ell$ produces a small amount of $d\mathbf{B}$:

Biot-Savart's law for a current.

$$d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{I d\ell \times \hat{\mathbf{r}}}{r^2}. \quad (8.3)$$

To get the magnetic field at some point, the current has to be integrated over the whole path. The line segment ℓ always points along the current I .

Magnetic field of a current-carrying wire

Let's look at a simple example of a thin, straight wire carrying a current I . Put the x axis along the current, and a point on the y axis, like in Fig. 8.2a. The current is in the $+\hat{\mathbf{x}}$ direction, so it is clear that the magnetic field points out the page on the top ($\hat{\mathbf{x}} \times \hat{\mathbf{r}}$), and in the page on the bottom. This follows the right-hand rule illustrated in Fig. 8.3a. The magnetic field due to a short piece $d\ell$ of the wire at an angle θ is

$$d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{I dx}{r^2} \hat{\mathbf{x}} \times \hat{\mathbf{r}}. \quad (8.4)$$

Solving the triangle,

$$d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{I \cos \theta dx}{r^2} \hat{\mathbf{z}}. \quad (8.5)$$

It will be easier to integrate this over θ , and we want the magnetic field as a function of the distance y from the wire. Because $x = y \tan \theta$,

$$\frac{dx}{d\theta} = y \sec^2 \theta = \frac{r^2}{y}. \quad (8.6)$$

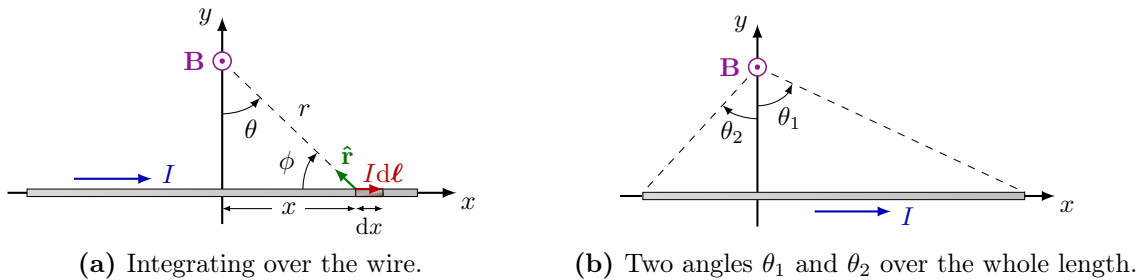


Figure 8.2: Calculating the magnetic field of a thin current-carrying wire.

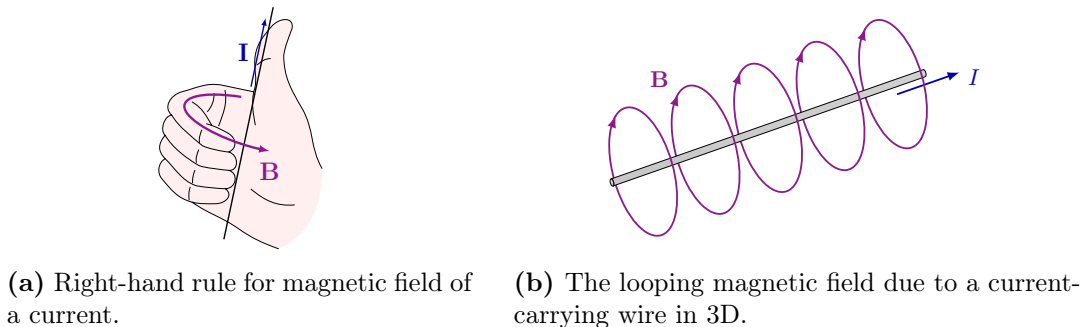


Figure 8.3: Magnetic field lines of a thin current-carrying wire.

We transform the integral to run over $d\theta$, and between the maximal angles θ_1 and θ_2 :

$$\begin{aligned}\mathbf{B} &= \frac{\mu_0 I}{4\pi y} \int_{\theta_2}^{\theta_1} \cos \theta \, d\theta \, \hat{\mathbf{z}}. \\ &= \frac{\mu_0 I}{4\pi y} (\sin \theta_1 - \sin \theta_2) \hat{\mathbf{z}}.\end{aligned}\quad (8.7)$$

Notice that for $y > 0$, $\theta_1 > 0$ (counterclockwise) and $\theta_2 < 0$ (clockwise), so $\sin \theta_1 - \sin \theta_2 > 0$, and \mathbf{B} points in the $+\hat{\mathbf{z}}$ direction.

Now apply our usual trick, where we assume the wire is infinitely long. In that limit, θ_1 and θ_2 approach 90° . The magnetic field is

Magnetic field of a wire.

$$\mathbf{B} = \frac{\mu_0 I}{2\pi y} \hat{\mathbf{z}}, \quad (8.8)$$

which depends on $1/r$ instead of $1/r^2$.

8.2 Ampère's law

There is again a simpler way! To compute the magnetic field for simple configurations with a lot of symmetry, one can use

Ampère's law.

$$\oint_C \mathbf{B} \cdot d\boldsymbol{\ell} = \mu_0 I_C. \quad (8.9)$$

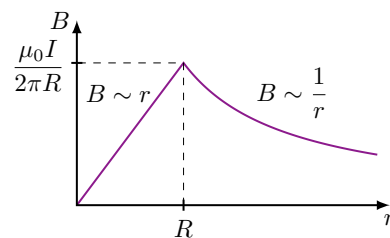
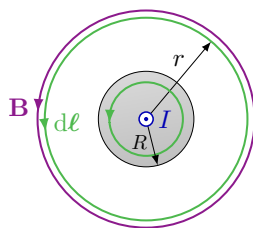
The integrand $\mathbf{B} \cdot d\boldsymbol{\ell}$ is the component of the magnetic field that is tangential to the closed curve C , while $\mu_0 I_C$ is the total current through that curve.

8.2.1 Magnetic field of a cylinder

Consider a current that is distributed uniformly on the cylinder with radius R , like in Fig. 8.4a, with current I coming out of the page. Because we already know that the magnetic field loops concentrically around the cylinder, we make circular curves of radius r along the field lines.

Outside the cylinder, $r > R$, the total current through the loop is $I_C = I$, and the magnetic field is constant and always parallel to the curve. Therefore Eq. (8.9) simply becomes

$$B \oint_C d\ell = 2\pi r B = \mu_0 I, \quad (8.10)$$



(a) A loop inside ($r < R$) and outside ($r > R$). (b) Plot of the magnetic field of a cylinder.

Figure 8.4: Using Ampère's law to calculate the the magnetic field of a current-carrying cylinder.

or,

$$B = \frac{\mu_0}{2\pi r} I, \quad (8.11)$$

exactly the same as what we had found for a thin wire before, Eq. (8.8)!

Inside the cylinder, $r < R$, the total current through the circular curve is

$$I \frac{\pi r^2}{\pi R^2} = I_C. \quad (8.12)$$

So this time, we find

$$B(2\pi r) = \mu_0 \left(I \frac{r^2}{R^2} \right), \quad (8.13)$$

which becomes

$$B = \frac{\mu_0 I r}{2\pi R^2}. \quad (8.14)$$

This is plotted in Fig. 8.4b.

8.2.2 Magnetic field of a loop

What if we bend a current-carrying wire in a loop? Opposite sides will add up in the middle of the loop, while mostly canceling outside, shown in Fig. 8.5a. On the loop's axis the magnetic field will be straight.

A realistic field due to a loop of current is shown in Fig. 8.6b. Notice the similarity to the field due to a bar magnet (Fig. 7.2a).

8.2.3 Magnetic field of a solenoid

We can amplify the straight magnetic field inside our loop by making many loops, like in a coil or "solenoid". A solenoid is made out of N turns of a wire over a length L . Therefore the number of loops per unit of length is

$$n = \frac{N}{L}. \quad (8.15)$$

What is the exact size of the magnetic field inside? We use Ampère's law, by assuming that the field is uniform inside and zero outside. This is a good approximation since the field is spread out azimuthally outside the coil, but concentrated inside. A convenient curve is a rectangle that encircles some of the wires, and is parallel to the field inside, just like in Fig. 8.6a. If sides 1 and 3 have length a , the enclosed current is

$$I_C = naI. \quad (8.16)$$

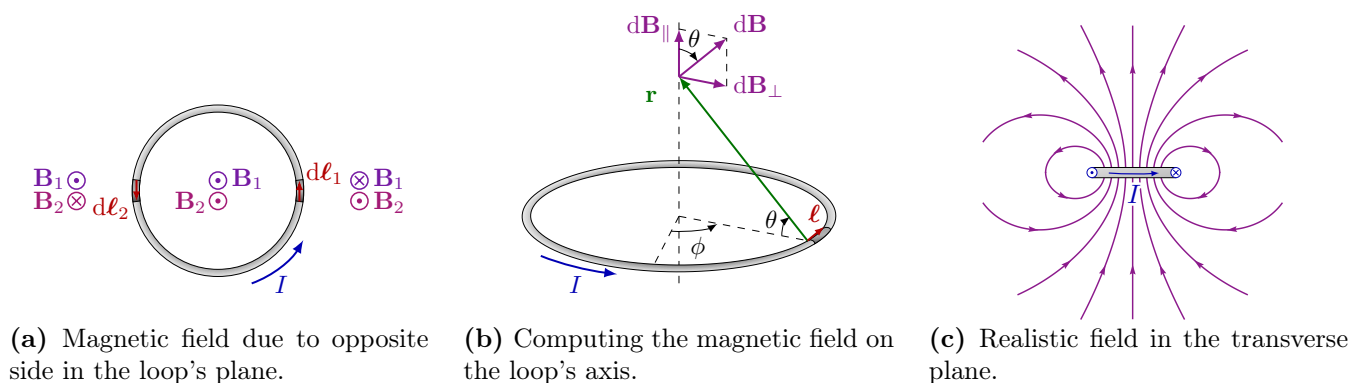
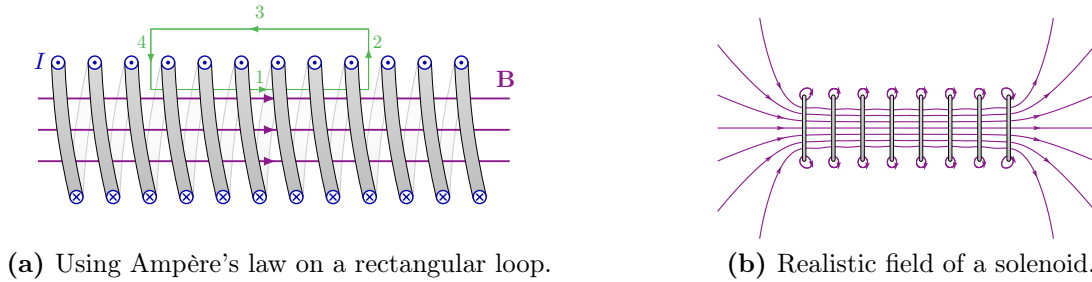


Figure 8.5: Magnetic field due to a current-carrying loop.



(a) Using Ampère's law on a rectangular loop.

(b) Realistic field of a solenoid.

Figure 8.6: Magnetic field due to a solenoid.

The magnetic field is zero along side 3, and either zero or perpendicular to sides 2 and 4, so only the integral over side 1 inside the solenoid remains:

$$Ba = \mu_0 n a I. \quad (8.17)$$

This results in

Magnetic field inside a solenoid.

$$B = \mu_0 n I = \mu_0 \frac{N}{L} I. \quad (8.18)$$

In reality, the field lines outside the coil are not zero, as they close the field lines inside the coil, but are less dense as is shown in Fig. 8.6b.

8.2.4 Permeability

If you put some material inside that can be magnetized, as for example an iron rod, the magnetic field will change:

$$B = \mu n I. \quad (8.19)$$

Here,

$$\mu = \mu_0 \kappa, \quad (8.20)$$

where μ is the *permeability*, which is a constant of the material in question, and κ is the *relative permeability*, as compared to free space. Some values are listed in Table 8.1.

8.2.5 Magnetic force between conducting wires

In Section 7.3, we had already seen that there is a force due to current flowing over a length ℓ , which is

$$F = BI\ell \quad (8.21)$$

if the wire is straight and the magnetic field is perpendicular. But a current will also create its own field. So what are the forces on two current-carrying wires of length ℓ next to each

Table 8.1: Relative permeability κ of several materials.

Material	κ [μ/μ_0]
iron (99.8% pure)	5000
iron (99.95% pure)	200 000
air	1.00000037
water	0.99992
copper	0.999994

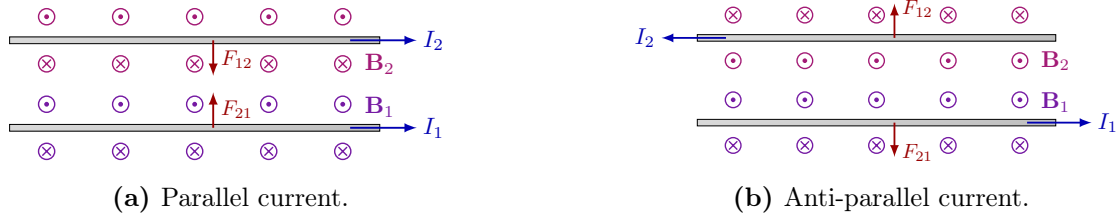


Figure 8.7: Magnetic force between current-carrying wires.

other? The wires will exert an equal but opposite force on each other:

$$\mathbf{F}_{12} = -\mathbf{F}_{21}. \quad (8.22)$$

For the force wire 1 exerts on 2, we have

$$F_{12} = B_1 I_2 \ell, \quad (8.23)$$

if wire 1 has a field B_1 , and wire 2 has a current I_2 . We have derived the expression for B_1 above:

$$F_{12} = \frac{\mu_0 I_1}{2\pi r} I_2 \ell. \quad (8.24)$$

In other words, the force on each wire per unit of length is

$$\frac{\mu_0 I_1 I_2}{2\pi r}. \quad (8.25)$$

If the currents in the wires are parallel (Fig. 8.7a), the wires attract, while if the currents are anti-parallel, the wires repel (Fig. 8.7b).

8.3 Magnetic flux

To quantify how much magnetic field goes through a surface, we define the magnetic flux Φ_M , just like we did for the electric flux Φ in Section 2.1. Suppose you have a wire that is looped forming a circuit C , and is placed in a uniform magnetic field \mathbf{B} . The surface that is bounded by the circuit C has some area A . If the field lines are all perpendicular to a flat surface, the flux is simply

$$\Phi_M = BA. \quad (8.26)$$

This is the case where \mathbf{B} is parallel to the surface's normal vector $\hat{\mathbf{n}}$. If the normal vector $\hat{\mathbf{n}}$ makes some angle θ with \mathbf{B} ,

$$\Phi_M = \mathbf{B} \cdot \mathbf{A} = \mathbf{B} \cdot \hat{\mathbf{n}} A = BA \cos \theta. \quad (8.27)$$

More generally, the surface may not be completely flat, and the magnetic field not uniform, so one has to integrate $\mathbf{B} \cdot d\mathbf{A}$ over the whole surface bounded by the circuit C .

Magnetic flux. *The magnetic flux by a magnetic field \mathbf{B} through an open surface S is given by*

$$\Phi_M = \int_S \mathbf{B} \cdot d\mathbf{A}, \quad (8.28)$$

where $d\mathbf{A} = \hat{\mathbf{n}} dA$.

The magnetic flux has its own units, called Weber,

$$1 \text{ Wb} = 1 \text{ T m}^2. \quad (8.29)$$

8.3.1 Magnetic flux through a solenoid

Should a surface be enclosed by a circuit with N loops of the same area, the total magnetic flux through the surface is multiplied by N :

$$\Phi_M = N \int_S \mathbf{B} \cdot d\mathbf{A}, \quad (8.30)$$

Consider the flux through a solenoid with N turns over a length ℓ , and with radius r . What is the magnetic flux through the whole solenoid? One loop inside has an area $A = \pi r^2$, and therefore it sees a magnetic flux

$$\Phi_{M,1} = BM = (\mu_0 n I)(\pi r^2) = \mu_0 \frac{N}{\ell} I \pi r^2. \quad (8.31)$$

So because the solenoid has N windings, it has a flux

$$\Phi_M = \mu_0 \frac{N^2}{\ell} I \pi r^2 = \mu_0 n^2 \ell I \pi r^2. \quad (8.32)$$

8.4 Faraday's law

It turns out that there is a deep connection between the magnetic flux and the EMF in a circuit.

Remember that the EMF of a circuit is a source of electric potential, representing the work done per electric charge to bring charges from a low potential to a higher potential. For a circuit, it is also the work done to move charge carriers around the loop C :

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{l}. \quad (8.33)$$

If you change the magnetic flux Φ_M , it creates an EMF given by

Faraday's law.

$$\mathcal{E} = \oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{d\Phi_M}{dt}. \quad (8.34)$$

Comparing the units on the left- and right-hand side, we see that

$$1 \text{ V} = 1 \frac{\text{Wb}}{\text{s}}. \quad (8.35)$$

If the magnetic field B increases only in strength, the magnetic flux Φ_M also increases. Faraday's law Eq. (8.34) then implies that this change in flux induces an electric field. This time the induced current is actually in the opposite direction of the right-hand rule! We will see why in the next section.

8.5 Lenz's law

Notice the negative sign in Faraday's law. This means something:

Lenz's law. *The induced EMF and induced current are in a direction as to oppose the change in magnetic flux that produces them.*

This is kept purposefully vague, in order to keep it applicable in many different cases, as will be shown below.

There are two ways you can change the magnetic flux: Either you change the magnetic field's strength or orientation, or you change the area of a circuit. Suppose you have a bar magnet moving with its north side towards a conducting loop with some velocity, as in Fig. 8.8a. Which way will the current go? As we push the magnet closer, the flux of its magnetic field \mathbf{B} increases: The magnetic field lines are more dense near the magnet's poles, so more lines will go through the loop. Lenz's law now says that the change in flux will induce a current in the loop so as to oppose the increase in the magnetic flux. The current will create its own changing magnetic field \mathbf{B}_{ind} in the opposite direction and decreases the total magnetic field to oppose the increase in magnetic flux. If we now move the magnet away from the loop (Fig. 8.8b), the magnetic flux due to the magnet gets smaller. Therefore, the current will now flow in the opposite direction, such that \mathbf{B}_{ind} is in the same direction as \mathbf{B} . In this case, the total magnetic field increases to oppose the decrease in magnetic flux.

Notice that if the magnetic field would be uniform and constant, the flux would not change, and there would be no induced current.

In Section 7.3.4, we had seen that a moving, non-uniform magnetic field induces a current on a loop with some magnetic momentum μ . To help understand this at a deeper level, consider the forces that are at play here. If the magnet's north pole is moving closer to the loop, the loop will create an opposing field \mathbf{B}_{ind} . Suppose that this induced field would be due to some new bar magnet instead of the conducting loop. The new magnet would be pointing with its north pole to the "real" magnet. Moving the real bar magnet towards the new one, we feel a repelling force between them. In the other situation, the real bar magnet moves away from the loop. Here, the magnet's field \mathbf{B} and the induced field \mathbf{B}_{ind} will be in the same direction, which causes a force that pulls them together.

8.5.1 Lenz's law in a changing loop

Suppose you lay a conducting rod on top of two connected rails, as in Fig. 8.9a, such that the rod can freely slide or roll back and forth on the rails, and the rails and rod form a closed loop that allows a current to flow through. If the rails are a length ℓ apart, and the rod lays at some position x from the end of the rails, then the closed loop has an area

$$A = \ell x. \quad (8.36)$$

If the rod moves with some velocity

$$\mathbf{v} = \frac{d\mathbf{x}}{dt}, \quad (8.37)$$

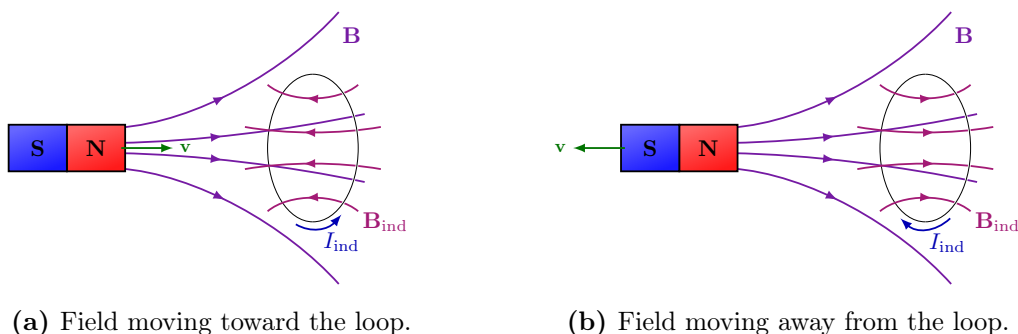
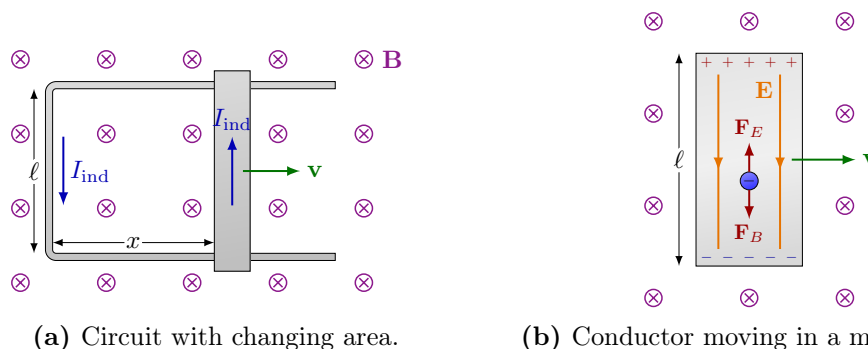


Figure 8.8: The magnetic field \mathbf{B} of a moving bar magnet will induce a current I_{ind} in a conducting loop and therefore a magnetic field \mathbf{B}_{ind} .



(a) Circuit with changing area. (b) Conductor moving in a magnetic field.

Figure 8.9: Lenz's law can be applied to conductors in magnetic fields.

the area will change. If an external magnetic field \mathbf{B} is applied, the flux $\Phi_M = B\ell x$ changes according to Faraday's law:

$$\frac{d\Phi_M}{dt} = B\ell \frac{dx}{dt} = B\ell v \quad (8.38)$$

What is the direction of the EMF? If the area is increasing, Φ_M increases, and so the induced EMF will tend to oppose the external magnetic field. So in Fig. 8.9a, the induced current I_{ind} is counterclockwise.

Once a current starts flowing, the rod will feel a force due to the external magnetic field:

$$F = I_{\text{ind}}\ell B, \quad (8.39)$$

opposite to its velocity, opposing its motion. The induced force has a tendency to stop the motion. This example represents a different formulation of Lenz's law:

Lenz's law for motional EMFs. *A motional EMF is an EMF induced by the relative motion between a magnetic field and a direction of charge (or current).*

8.5.2 Moving conductor in a magnetic field

Consider a conductor moving in a magnetic field \mathbf{B} with some velocity \mathbf{v} as in Fig. 8.9b. Suppose the conductor does not form a loop, so there is no closed circuit. For simplicity, suppose the conductor is a simple rectangular bar with its length ℓ perpendicular to its horizontal direction of motion \mathbf{v} . The conductor has free charge carriers that will feel a force \mathbf{F}_B due to \mathbf{B} that is perpendicular to the direction of motion \mathbf{v} . The charges will redistribute, with the negative charges accelerating to the bottom, inducing positive charges (ions) to accumulate on the top. The charge redistribution will cause an electric field \mathbf{E} inside the conductor such that the charges now feel an electric force \mathbf{F}_E that opposes \mathbf{F}_B . At equilibrium, the force on a charge is

$$\mathbf{F}_E = \mathbf{F}_B. \quad (8.40)$$

We are familiar with the expression for these forces:

$$qE = qvB, \quad (8.41)$$

or,

$$E = vB \quad (8.42)$$

at equilibrium. The potential difference ΔV between the top and bottom of the conductor is

$$\Delta V = E\ell. \quad (8.43)$$

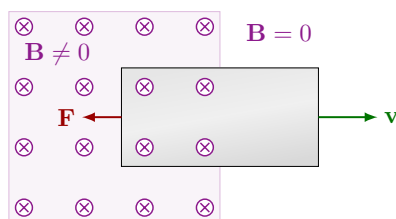


Figure 8.10: A conductor moving out an magnetic field experiences a force pulling it back in.

8.5.3 Application: Eddy currents

What happens if we move a conductor between two magnets that create a strong divergent field, such as two bar magnets. If we move the conductor out of the strong field, the flux decreases, and so the induced current I must be in the direction to make the total magnetic field bigger. Therefore, the induced B field is in the same direction and yields a force that opposes the motion: The conductor is pulled back in, but only when it is moved outward. Likewise, when the conductor is moved in to the magnetic field, the magnetic flux increases, and the induced current creates a magnetic field in the opposite direction, so as to oppose the magnetic flux increase. This will cause a force pushing the conductor away as it enters the magnetic field. Induced currents in conductors due to magnetic fields are called *Eddy currents*, and yield forces.

Generally, they are not wanted, since the forces or currents interfere with the expected behaviors. One can decrease the Eddy currents on a large conductor by creating gaps in the conductor material, such that large current loops are broken up into smaller ones. However, Eddy currents are the working principle of metal detectors. And because current can generate heat, another application is induction stoves.

Another example is protective electronic circuits. Suppose we have a current loop supplying current to a device. We know that the current through the loop should be the same along the loop before the device and after. If not, the current may be electrocuting somebody ! How can we check if $I_{\text{before}} = I_{\text{after}}$? A simple way is to take a donut-shaped piece of iron, and pass both I_{before} and I_{after} through it. If the currents are equal and opposite, they cancel out, and there is no induced magnetic field in the iron. However, if $I_{\text{before}} > I_{\text{after}}$, then a magnetic field will be induced in the donut. Now, suppose we loop a third wire loop around the donut. The induced magnetic field will induce an Eddy current in the third wire loop, and this current can be used in a switch to disconnect power. Such a circuit works within a few milliseconds, and can detect small changes in electric current of a circuit that indicate the current is leaking out unexpectedly. The third circuit, in fact, may function also using Eddy currents: when a current passes through it, it induces a magnetic field in an iron switch, causing the switch to move and disconnect.

Chapter 9

Inductance & RLC circuits

9.1 Inductance

Previously, we have found that for a solenoid,

$$\Phi_M = \mu_0 n^2 A \ell I. \quad (9.1)$$

The first factors, only have to do with the geometry (n , A , ℓ) and material (μ_0 or μ) of the solenoid, so we can group them together in to a quantity called the *self-inductance*,

$$L = \mu_0 n^2 A \ell. \quad (9.2)$$

Inductance has units of Henry,

$$1 \text{ H} = 1 \frac{\text{Wb}}{\text{A}} = 1 \frac{\text{T m}^2}{\text{A}} \quad (9.3)$$

We can therefore obtain any desired value L by changing n , A , or ℓ , or, the permeability μ by inserting some material into the solenoid. Because L is constant with time, Faraday's law, Eq. (8.34) becomes

Faraday's law for inductors.

$$\mathcal{E} = -\frac{d\Phi_M}{dt} = -L \frac{dI}{dt}. \quad (9.4)$$

Notice from the above equation that the units for inductance must also satisfy:

$$1 \text{ H} = 1 \frac{\text{V s}}{\text{A}}, \quad (9.5)$$

which gives us a correspondence between the units of magnetic flux and the units of electric potential.

Remember, because of the negative sign, the voltage through the inductor opposes the voltage of the generated current of the circuit.

If we add a solenoid to a circuit, it is called an *inductor*. When current flows through the inductor, it creates an EMF that opposes the EMF of the generator current, and stores energy in its magnetic field. The inductance L is a measure of how much an inductor opposes the change in current for a given voltage; $-dI/dt = \mathcal{E}/L$. We will see some examples of such circuits in the next sections.

Note that not only solenoids can act as inductors: Since any simple circuit has a loop of current, it has some amount of inductance that opposes the generated current.

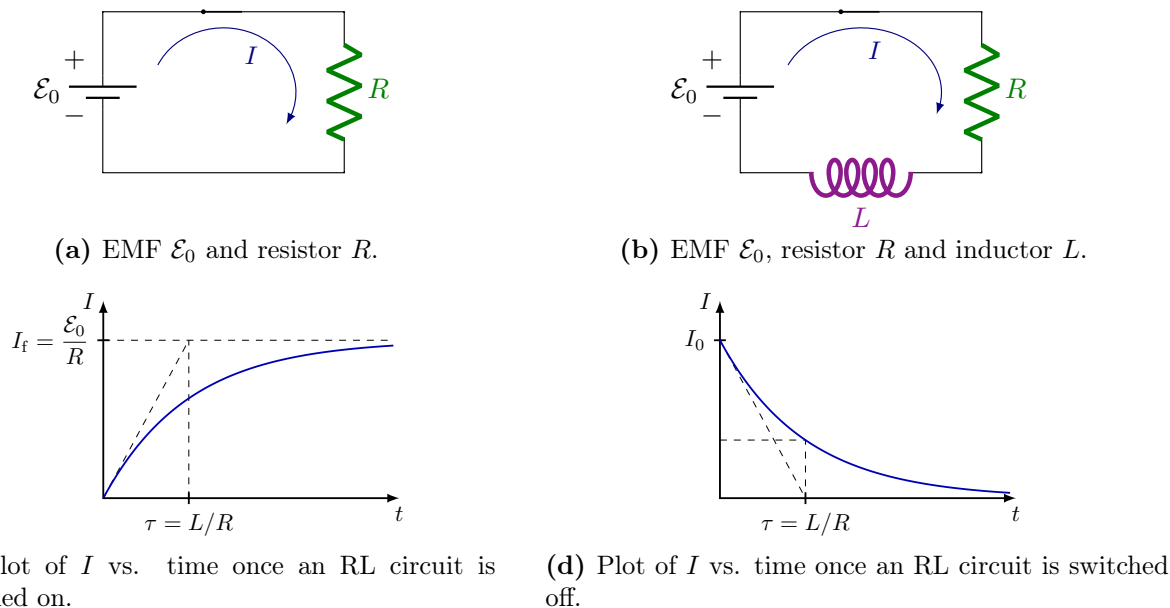


Figure 9.1: Circuits with an EMF \mathcal{E}_0 , resistor R and inductor L .

9.2 RL circuits

If we put L in a circuit, with some constant \mathcal{E}_0 , the inductor creates, or induces, a current opposes the current I generated by \mathcal{E}_0 .

First, consider the simple circuit in Fig. 9.1a with an EMF source \mathcal{E}_0 , a resistor R and a switch. When we close the switch, we can use Kirchoff's loop rule to find that

$$\mathcal{E}_0 - IR = 0, \quad (9.6)$$

or

$$I = \frac{\mathcal{E}_0}{R}. \quad (9.7)$$

Now add an inductor with inductance L after R . If we close the switch this time, current flows, but this time, there is an opposing EMF \mathcal{E} induced in the inductor, and the loop rule becomes

$$\mathcal{E}_0 - IR - L \frac{dI}{dt} = 0. \quad (9.8)$$

This is a differential equation, which we solve in the same way we solved a charging RC circuit in Section 6.6.2.

$$I = \frac{\mathcal{E}_0}{R} \left(1 - e^{-\frac{Rt}{L}} \right). \quad (9.9)$$

Here, the time constant τ is

$$\tau = \frac{L}{R}. \quad (9.10)$$

This is the time when the current I is at $1 - 1/e \sim 63\%$ of its maximum value, and defines the time to charge based on the initial slope of increase. Fig. 9.1d Instead of reaching the current $I = \mathcal{E}_0/R$, the inductor L slows the growth of the current so that the final current $I\mathcal{E}_0/R$ is reached only after some delay in time.

When we open the switch again (switch it off), the current does not immediately drop to zero, as without the inductor L . It is easy to show, using the same type of differential equation that the current slowly decreases, as is shown in Fig. 9.1d.

9.2.1 Energy stored in an RL circuit

Eq. (9.8) can be rearranged, and multiplied by I , to get

$$\mathcal{E}_0 I = I^2 R + LI \frac{dI}{dt}. \quad (9.11)$$

With some simple dimensional analysis, we recognize the left-hand side as the power of the battery (see Eq. (5.28)). The first term on the right-hand side, $I^2 R$, is the power to heat the resistor (see Eq. (5.30)), while the second one can be recognized as the rate of change of energy in the inductor. Let's call U_M the energy stored in the inductor, so:

$$\frac{dU_M}{dt} = LI \frac{dI}{dt}. \quad (9.12)$$

We rewrite to

$$dU_M = LI dI, \quad (9.13)$$

and integrate between 0 and some final current I_F :

$$U_M = \int_0^{I_F} LI dI. \quad (9.14)$$

The result (after substituting I for I_F) is :

Energy stored in an inductor.

$$U_M = \frac{1}{2} LI^2. \quad (9.15)$$

This is the energy stored in the magnetic field of an inductor. Notice the similarity to the electric energy stored in a capacitor (Eq. (4.31)):

$$U_E = \frac{1}{2} CV^2. \quad (9.16)$$

The full expression of U_M for a solenoid with n windings, and length ℓ is

$$U_M = \frac{B^2}{2\mu_0} A\ell. \quad (9.17)$$

Note that $A\ell$ is the total volume of the solenoid. We can therefore define the energy density as

Energy density of an magnetostatic field.

$$\eta_M = \frac{\text{energy}}{\text{volume}} = \frac{1}{2} \frac{B^2}{\mu_0}. \quad (9.18)$$

Again compare to our result of the electric field density for a capacitor:

$$\eta_E = \frac{1}{2} \epsilon_0 E^2. \quad (9.19)$$

In fact, electromagnetic waves carry both electric magnetic energy, as we will see later:

$$\begin{aligned} \eta_{EM} &= \eta_B + \eta_E \\ &= \frac{1}{2} \frac{B^2}{\mu_0} + \frac{1}{2} \epsilon_0 E^2. \end{aligned} \quad (9.20)$$

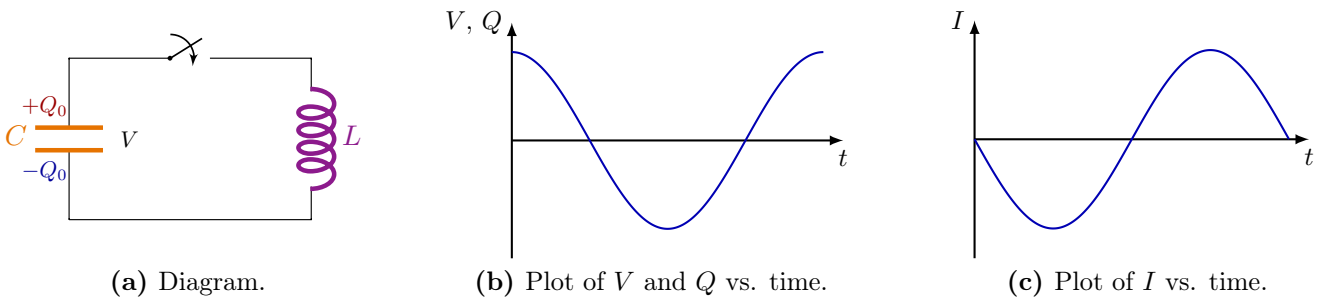


Figure 9.2: An LC circuit with a capacitor C and an inductor L .

9.3 LC circuit

Consider a circuit containing a capacitor with capacitance C and some charge Q_0 , an inductor with inductance L , and a switch. We close the switch, and a current I start flowing through the inductor to get charge from one end of the capacitor to the other, as the capacitor discharges with $dQ/dt < 0$. We write down the loop rule to find what exactly happens:

$$-L \frac{dI}{dt} - \frac{Q}{C} = 0. \quad (9.21)$$

We substitute I as a time-derivative of charge, with $I = dQ/dt < 0$:

$$\frac{d^2Q}{dt^2} + \frac{Q}{LC} = 0. \quad (9.22)$$

We recognize this differential equation as the same as for a spring with angular velocity $\omega = \sqrt{k/m}$:

$$\ddot{x} + \omega^2 x = 0. \quad (9.23)$$

Analogously, the solution for the charge Q in the capacitor becomes

$$Q(t) = Q_0 \cos(\omega t - \phi), \quad (9.24)$$

with

$$\omega = \frac{1}{\sqrt{LC}}. \quad (9.25)$$

We see that when the inductance L is larger, the discharging-recharging cycle takes longer (i.e. ω is smaller, and period $T = 2\pi/\omega$ is longer) because the inductor is offering more resistance to the change in current. Similarly, if we increase the capacitance C for a given Q_0 and L , we decrease the voltage across the inductor, and the change in current is again smaller, $dI/dt = -Q_0/LC$. If $Q(0) = Q_0$ at $t = 0$, then the phase ϕ vanishes. The current through the inductor is simply

$$I(t) = \frac{dQ}{dt} = -Q_0\omega \sin(\omega t - \phi). \quad (9.26)$$

This is a harmonic oscillation, just like the spring! It means that once we close the switch, the capacitor is constantly being charged and discharged from $+Q_0$ to $-Q_0$. Negative charge means that the original polarity of the capacitor is flipped.

If $\phi = 0$, the current initially becomes negative. This means the capacitor is being discharged, $I = dQ/dt < 0$, and the current is flowing away from the initially positive capacitor plate to the negative one. After that the polarity of the capacitor keeps oscillating. The maximum current is given by $I_{\max} = Q_0\omega = Q_0/LC$.

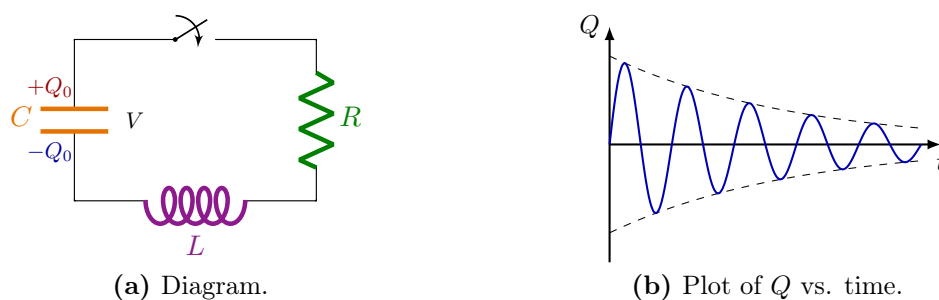


Figure 9.3: An RLC circuit with a capacitor C , a resistor R , and an inductor L .

The voltage through the capacitor will be in phase with its charge:

$$V(t) = V_0 \cos(\omega t - \phi), \quad (9.27)$$

but is $\pi/2$ out of phase with the current I through the inductor. The inductor current is ahead of the voltage through the capacitor.

The total energy in the circuit flows back and forth between the electric field of the capacitor, and the magnetic field of the inductor. The total energy is constant,

$$E = \frac{1}{2}CV^2 + \frac{1}{2}LI^2, \quad (9.28)$$

but the energy terms on the left-hand side oscillate between each other.

9.3.1 RLC circuit

If we add a resistor to the circuit of the previous section, we get a more realistic situation since every circuit has resistance, and therefore energy loss through heat. Now the loop rule is

$$-L \frac{dI}{dt} - IR - \frac{Q}{C} = 0. \quad (9.29)$$

The differential equation is

$$\frac{d^2Q}{dt^2} + \frac{R}{L} \frac{dQ}{dt} + \frac{Q}{LC} = 0. \quad (9.30)$$

The new term with R will cause the system to lose energy through heat. The RLC circuit therefore behaves just like a damped oscillator, which we previously solved in the last semester. It has a solution

$$Q(t) = Q_0 e^{-\frac{Rt}{2L}} \cos(\omega' t - \phi), \quad (9.31)$$

with

$$\omega' = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}, \quad (9.32)$$

and where $e^{-\frac{Rt}{2L}}$ is the damping term, which causes the oscillation to die out exponentially. The energy conservation in the circuit implies that at any moment,

$$E = \frac{1}{2}LI^2 + \frac{1}{2}CV^2, \quad (9.33)$$

where $V = Q/C$ is still the voltage over C . It is easy to prove using Eqs. (9.29) and (9.33) that the rate of energy loss is

$$\frac{dE}{dt} = -I^2 R. \quad (9.34)$$

This is the power of heat loss through the resistor R , which was not there in the LC circuit.

Chapter 10

AC Circuits & Phasors

In Chapter 9, we had seen some of the interesting properties when combining resistors, capacitors and inductors in circuits. However, these were each time with a direct current (DC) due to a battery with an EMF \mathcal{E} .

In this chapter, we will have a closer look at circuits with an alternating current (ACs). These also have some interesting effects when introducing capacitors and inductors to the circuit. Namely, capacitors and inductors act as resistors, but they introduce phase shifts of the current and voltage. They have a wide range of real-life applications.

This is only a supplementary chapter, and is not covered on the exam.

10.1 Simple resistor

First, let's look at a trivial example as reference: Consider the circuit with an AC generator in series with a single resistor R as in Fig. 10.1a. Assume that the AC oscillate harmonically as

$$\mathcal{E}(t) = \mathcal{E}_0 \cos(\omega t), \quad (10.1)$$

where \mathcal{E}_0 is the maximum EMF created, and ω is the angular frequency. The “normal” frequency is of course $f = \omega/2\pi$. For simplicity we have set the phase to zero, $\phi = 0$, such that the oscillation starts at its maximum. If V_R is the voltage over R , then according to Kirchhoff's loop rule,

$$\mathcal{E} - V_R = 0, \quad (10.2)$$

so simply, $V_R = \mathcal{E}$. The current I over R still follows Ohm's law,

$$I(t) = \frac{V_R}{R} = \frac{\mathcal{E}_0}{R} \cos(\omega t). \quad (10.3)$$

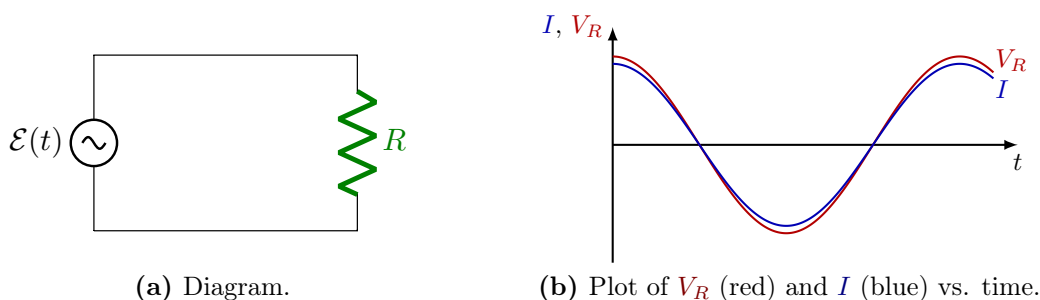


Figure 10.1: An AC circuit with a resistor R with voltage difference V_R .

We immediately read off that the the maximum current is

$$I_{\max} = \frac{\mathcal{E}_0}{R}, \quad (10.4)$$

and that the current follows the same oscillation. The current I is in phase with the voltage V_R (Fig. 10.1b).

This was trivial. Let's look at some more interesting cases.

10.2 Capacitive reactance

Now replace the resistor R with a capacitor C , as in Fig. 10.2a. The loop rule implies again that the voltage over C is

$$V_C = \mathcal{E} = \mathcal{E}_0 \cos(\omega t). \quad (10.5)$$

We also know that by the definition of capacitance C ,

$$Q = CV_C = C\mathcal{E}_0 \cos(\omega t). \quad (10.6)$$

There is therefore an implicit assumption that the capacitor is already charged by

$$Q(0) = Q_{\max} = CV_C(0) = C\mathcal{E}_0 \quad (10.7)$$

at time $t = 0$. This gives us the current as

$$I = \frac{dQ}{dt} = -\omega C\mathcal{E}_0 \sin(\omega t), \quad (10.8)$$

or converting back to a cosine by introducing a phase shift $-\pi/2$:

$$I = \omega C\mathcal{E}_0 \cos\left(\omega t + \frac{\pi}{2}\right). \quad (10.9)$$

In other words, the voltage drop across the capacitor lags 90° (or $\pi/2$) *behind* the current. This is illustrated in Fig. 10.2b. The maximum current this time is

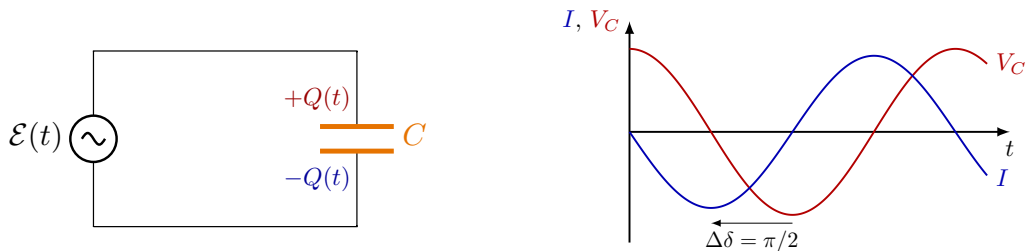
$$I_{\max} = \omega C\mathcal{E}_0. \quad (10.10)$$

Notice that this looks very similar to the maximum current for the AC circuit with a resistor R , Eq. (10.4). The factor ωC has the inverse units of resistance (Ω^{-1}). We can therefore define the *capacitive reactance*

Capacitive reactance.

$$X_C = \frac{1}{\omega C}, \quad (10.11)$$

which acts like an effective resistor $R = X_C$ in an AC circuit, but also introduces a $+\pi/2$ phase shift of the current I with respect to the voltage V_C . This can be understood by the fact that the current is zero, $I = 0$ when the capacitor is fully charged, $Q_{\max} = C\mathcal{E}_0$, and carries a maximum voltage \mathcal{E}_0 .



(a) Diagram.

(b) Plot of V_C (red) and I (blue) vs. time.

Figure 10.2: An AC circuit with a capacitor C with voltage difference V_C .

10.3 Inductive reactance

A inductor L in series with an AC generator (Fig. 10.3a) also acts as a resistor. To see why, remember Faraday's law for inductors (9.4):

$$V_L = L \frac{dI}{dt}, \quad (10.12)$$

. Applying the loop rule, $\mathcal{E} - V_L = 0$, and where again that due to the loop rule, the voltage across L is

$$V_L = \mathcal{E} = \mathcal{E}_0 \cos(\omega t). \quad (10.13)$$

Therefore, we have to integrate this time to obtain I :

$$I = \int \frac{V_L}{L} dt = \frac{\mathcal{E}_0}{L} \int \cos(\omega t) dt, \quad (10.14)$$

so assuming that $I(0) = 0$ at $t = 0$,

$$I = \frac{\mathcal{E}_0}{\omega L} \sin(\omega t). \quad (10.15)$$

This is again a phase shift with respect to the driving cosine:

$$I = \frac{\mathcal{E}_0}{\omega L} \cos(\omega t - \pi/2), \quad (10.16)$$

where the maximum is

$$I_{\max} = \frac{\mathcal{E}_0}{\omega L}. \quad (10.17)$$

Again we notice the similarity to a resistor and define the *inductive reactance*

Inductive reactance.

$$X_L = \omega L. \quad (10.18)$$

We conclude that an inductor L acts as an effective resistor $R = X_L$ in an AC circuit, introducing a $-\pi/2$ phase shift of the current I with respect to the voltage V_L . This phase shift is due to the inductor building up or winding down its induced magnetic field. When the current is maximal, its magnetic field is maximal and completely cancels the voltage in the solenoid.

10.4 Phasors

A very useful tool to represent the phase relations in AC currents is with *phasors*. These are 2D vectors that rotate in the xy plane.

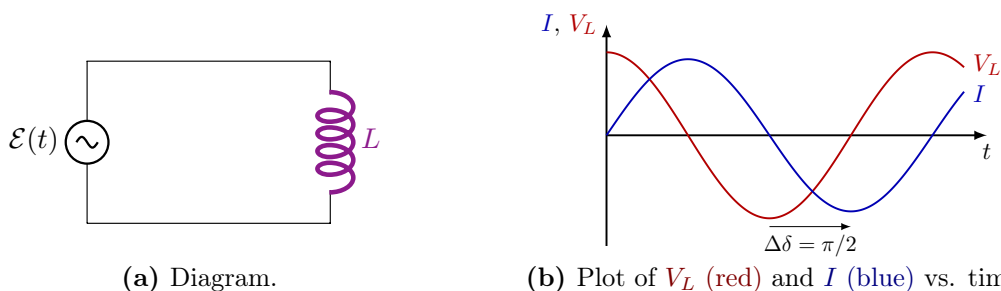


Figure 10.3: An AC circuit with an inductor L with voltage difference V_L .

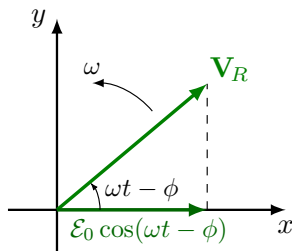


Figure 10.4: A phasor rotating with a fixed magnitude in the xy plane.

Consider again the simple resistor R in an AC circuit. The driving EMF \mathcal{E} and voltage V_R across R are

$$\mathcal{E} = V_R = \mathcal{E}_0 \cos(\omega t - \phi). \quad (10.19)$$

This can be thought of as the projection of a rotating vector

$$\mathbf{V}_R = \mathcal{E}_0 \cos(\omega t - \phi) \hat{\mathbf{x}} + \mathcal{E}_0 \sin(\omega t - \phi) \hat{\mathbf{y}} \quad (10.20)$$

with fixed magnitude $|\mathbf{V}_R|$ onto the x axis. This is illustrated in Fig. 10.4. We typically do not really care about the y component, but phasors are useful illustrations of phase differences between reactances (R, X_L, X_C), represented by rotating 2D vectors. Let's look at some examples.

10.4.1 RLC circuit in series

Let look at an RLC circuit with each of component in series with an AC source. We have already seen an example without an EMF in Section 9.3.1. We also already know that the inductor L , capacitor C and resistor R are out of phase with respect to each other. This illustrated by the phasors in Fig. 10.6a: The inductive reactance \mathbf{V}_L is $\pi/2$ ahead of the resistor \mathbf{V}_R , while the capacitive reactance \mathbf{V}_C is $\pi/2$ behind. They all rotate counterclockwise the angular velocity ω of the AC source, and have constant magnitude. At any given time t , the component have a different voltage, which is given by the components are simply the x components of these vectors.

Let's look at their relation to the driving EMF \mathcal{E} from the AC source. The loop rule implies

$$\mathcal{E} - L \frac{dI}{dt} - RI - \frac{Q}{C} = 0. \quad (10.21)$$

Substituting in the current as a derivative of I , and writing the EMF explicitly, we find the differential equation

$$L \frac{d^2 Q}{dt^2} + R \frac{dQ}{dt} + \frac{Q}{C} = \mathcal{E}_0 \cos(\omega t) \quad (10.22)$$

This is a forced harmonic oscillator, like we saw before with springs! It looks really hard to solve. But fret not, there is an easier, visual way: phasors. Notice that the loop rule

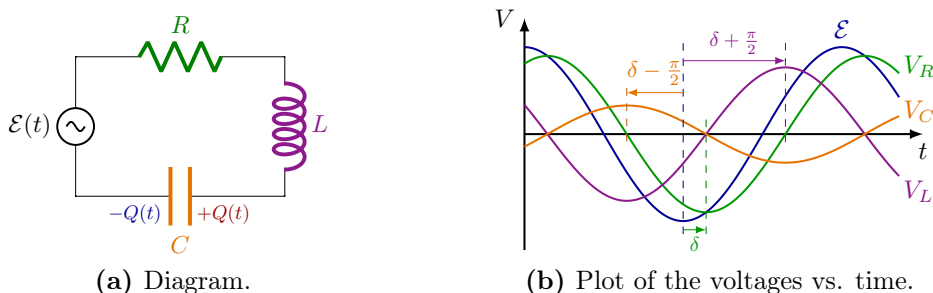


Figure 10.5: An RLC circuit with a driving EMF \mathcal{E} .



(a) Each reactance adds a rotating phasors. (b) The vectorial sum of the reactances is the EMF \mathcal{E} .

Figure 10.6: Phasors for a RLC in series.

basically implies that the sum of the voltage across the components in series, is the same as the EMF:

$$\mathcal{E} = V_R + V_L + V_C, \quad (10.23)$$

If we assume that each term in this equation is the x component of a phasors, this still holds, and we can even pose

$$\mathcal{E} = \mathbf{V}_R + \mathbf{V}_L + \mathbf{V}_C, \quad (10.24)$$

where \mathcal{E} must be

$$\mathcal{E} = \mathcal{E}_0 \cos(\omega t)\hat{\mathbf{x}} + \mathcal{E}_0 \sin(\omega t)\hat{\mathbf{y}}, \quad (10.25)$$

because it must oscillate with angular frequency ω and constant magnitude \mathcal{E}_0 . Because the vectorial sum of the reactance phasors in Eq. (10.24) is \mathcal{E} , we can now reverse-engineer each of them.

The magnitude $|\mathcal{E}|$ is

$$\mathcal{E}_0 = |\mathbf{V}_R + \mathbf{V}_L + \mathbf{V}_C| \quad (10.26)$$

$$= \sqrt{V_{R,\max}^2 + (V_{L,\max} - V_{C,\max})^2}, \quad (10.27)$$

where we noticed that \mathbf{V}_L and \mathbf{V}_R are anti-parallel and both perpendicular to \mathbf{V}_C , so we get the triangle shown in Fig. 10.6b. We actually know the maximum voltage over each individual component from the sections at the beginning of this chapter, so

$$\mathcal{E}_0 = I_{\max} \sqrt{R^2 + (X_L - X_C)^2}. \quad (10.28)$$

We refer to $X_L - X_C$ as the *total reactance*, while

Impedance for a RLC circuit in series.

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \quad (10.29)$$

is called the *impedance*, such that

$$\mathcal{E}_0 = I_{\max} Z, \quad (10.30)$$

which again reminds us of Ohm's law $V = IR$.

Next thing is the phase. If we know the phase ϕ of \mathbf{V}_R , we automatically know those of \mathbf{V}_L and \mathbf{V}_C , because they only differ by $\pi/2$. Notice that ϕ is simply the angle between \mathcal{E} and \mathbf{V}_R as in Fig. 10.6b:

$$\begin{aligned} \tan \phi &= \frac{|\mathbf{V}_L + \mathbf{V}_C|}{|\mathbf{V}_R|} \\ &= \frac{X_L - X_C}{R}. \end{aligned} \quad (10.31)$$

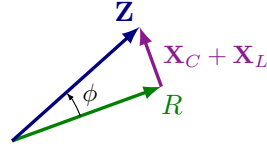


Figure 10.7: Triangle to remember for a RLC circuit.

This means that if $X_L = X_C$ exactly, the voltage over R , V_R , is in phase with \mathcal{E} because the inductor and capacitor cancel each other's effect. This means that the voltage over the whole circuit is just $V_R = \mathcal{E}$. Otherwise V_R is out of phase,

$$V_R = \frac{\mathcal{E}_0 R}{Z} \cos(\omega t - \phi) \quad (10.32)$$

and experience a current

$$I = \frac{\mathcal{E}}{Z} \cos(\omega t - \phi). \quad (10.33)$$

The maximum voltage can be rewritten as

$$\frac{\mathcal{E}_0 R}{Z} = \mathcal{E}_0 \cos \phi. \quad (10.34)$$

So instead of solving a difficult differential equation, Eq. (10.22), we used our physical insight and the powerful visual trick of phasors as an ansatz to derive the voltages across the components. It is helpful to remember the triangle shown in Fig. 10.7, so we can immediately read off the relations

Useful relations for RLC components.

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \quad (10.35)$$

$$\tan \phi = \frac{X_L - X_C}{R} \quad (10.36)$$

$$\cos \phi = \frac{R}{Z}. \quad (10.37)$$

Power

In terms of power delivered by the EMF then,

$$P = I^2 R = I_{\max}^2 R \cos^2(\omega t). \quad (10.38)$$

So when $X_C = X_L$, the circuit will deliver the most power, which is dissipated by R . The average of the power is

$$\langle P \rangle = \frac{1}{2} I_{\max}^2 R = \frac{1}{2} \mathcal{E}_0^2 \frac{R}{Z^2} = \frac{1}{2} I_{\max} \mathcal{E}_0, \quad (10.39)$$

because the average of $f = \cos^2$ is

$$\langle \cos^2 \rangle = \frac{1}{2}. \quad (10.40)$$

You can formally derive this by integrating over one period, $\langle f \rangle = \frac{1}{b-a} \int_a^b f(x) dx$. Now using Eq. (10.34),

$$\langle P \rangle = \frac{1}{2} I_{\max} \mathcal{E}_0 \cos \phi. \quad (10.41)$$

In terms of the *root mean square* of the current and voltage,

$$\mathcal{E}_{\text{RMS}} = \sqrt{\langle \mathcal{E}^2 \rangle} = \frac{\mathcal{E}_0}{\sqrt{2}} \quad (10.42)$$

and

$$I_{\text{RMS}} = \sqrt{\langle I^2 \rangle} = \frac{I_{\text{max}}}{\sqrt{2}}, \quad (10.43)$$

this becomes

$$\langle P \rangle = \mathcal{E}_{\text{RMS}} I_{\text{RMS}} \cos \phi. \quad (10.44)$$

This is why the value $\cos \phi$ is sometimes called the *power factor*.

Resonance

Lets look closer at the current in Eq. (10.33) again

$$I = \frac{\mathcal{E}}{\sqrt{R^2 + (X_L - X_C)^2}} \cos(\omega t - \phi), \quad (10.45)$$

with $\tan \phi = (X_L - X_C)/R$. We had already pointed out that if $X_C = X_L$, $\phi = 0$ and the current is in phase with the AC generator. But there is a second thing we notice: The current is maximal because the impedance in the numerator is minimal:

$$I_{\text{max}} = \frac{\mathcal{E}}{R}. \quad (10.46)$$

Simultaneously,

$$\omega L = \frac{1}{\omega C}, \quad (10.47)$$

must be true, and thus,

$$\omega = \frac{1}{\sqrt{LC}}. \quad (10.48)$$

This turns out to be a resonatory behavior, with the resonance frequency

$$\omega_0 = \frac{1}{\sqrt{LC}} \quad (10.49)$$

of the circuit.

The average power becomes

$$\langle P \rangle = \frac{1}{2} \mathcal{E}_0^2 \frac{R}{Z^2}, \quad (10.50)$$

where Z^2 now is

$$\begin{aligned} Z^2 &= R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \\ &= R^2 + \frac{L^2}{\omega^2} (\omega^2 - \omega_0)^2. \end{aligned}$$

In other words,

$$\langle P \rangle = \frac{\mathcal{E}_0^2 R \omega^2}{L^2 (\omega^2 - \omega_0)^2 + \omega^2 R^2}. \quad (10.51)$$

When you plot this as a function angular frequency ω , you get Fig. 10.8. It clearly shows a resonance around $\omega_0 = 1/\sqrt{LC}$.

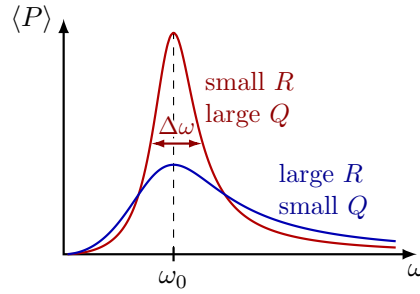


Figure 10.8: Plots of the average power $\langle P \rangle$ vs. angular frequency ω for different values of R and Q .

We can define a so-called Q factor

$$Q = \frac{\omega_0 L}{R}, \quad (10.52)$$

which relates to the full width $\Delta\omega$ of the resonance at half maximum (FWHM),

$$Q = \frac{\omega_0}{\Delta\omega} = \frac{f_0}{\Delta f}. \quad (10.53)$$

The higher the Q factor, the higher the power output. Radio receivers use the RLC circuits: By tuning the capacitance, different frequencies resonate, and so different radio emissions can be picked up.

10.4.2 RLC circuit in parallel

Now let's consider an RLC circuit in parallel, as in Fig. 10.9a. This time the voltage $V = \mathcal{E}$ is the same across each component, but the total current I branches off:

$$I = I_R + I_L + I_C, \quad (10.54)$$

where we have

$$I_R = \frac{V}{R}, \quad I_L = \frac{V}{X_L}, \quad I_C = \frac{V}{X_C}. \quad (10.55)$$

The phase differences are still there, but they manifest in the currents. We can therefore use our phasor trick and construct \mathbf{I}_R , \mathbf{I}_L and \mathbf{I}_C like in Fig. 10.9b:

$$\mathbf{I} = \mathbf{I}_R + \mathbf{I}_L + \mathbf{I}_C. \quad (10.56)$$

The x components again give us the real currents in the circuit. The total current is therefore

$$I = \sqrt{I_R^2 + (I_L - I_C)^2}, \quad (10.57)$$

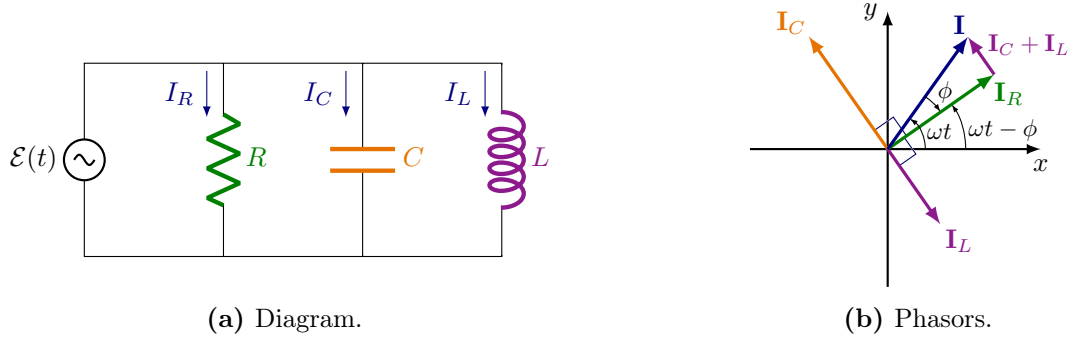
or in terms of the voltage,

$$I = \sqrt{\frac{V^2}{R^2} + \left(\frac{V}{X_L} - \frac{V}{X_C}\right)^2}. \quad (10.58)$$

This again allows us to rewrite it in a familiar form

$$I = \frac{V}{Z}, \quad (10.59)$$

where



(a) Diagram.

(b) Phasors.

Figure 10.9: An RLC circuit with a driving EMF \mathcal{E} .

Impedance for a RLC circuit in parallel.

$$\frac{1}{Z} = \sqrt{\frac{1}{R^2} + \left(\frac{1}{X_L} - \frac{1}{X_C}\right)^2}. \quad (10.60)$$

If $X_C = X_L$ this time, the currents between the inductor and capacitor cancel, and $Z = R$. In this case, the current is minimal for any given R :

$$I = \frac{V}{R}. \quad (10.61)$$

Notice, that while currents in the inductor and capacitors on their own are not identically zero, they oscillate back and forth, 180° out of phase, so their total current $I_R + I_C$ is exactly zero if $X_C = X_L$.

10.5 Complex impedance

Often, the math becomes *simpler* if one uses *complex numbers*. Remember from last semester that any complex number z can be written as an exponent, or some combination of a real cosine and imaginary sine:

$$Z = Re^{i\theta} \quad (10.62)$$

$$= R \cos \theta + R \sin \theta, \quad (10.63)$$

where the R is the modulus $|Z| = R$ and θ is an angle with respect to the real axis in the complex plane. Notice that we are interested in the real part. So we can take the EMF

$$\tilde{\mathcal{E}}(t) = \mathcal{E}_0 e^{i\omega t}, \quad (10.64)$$

with real part

$$\mathcal{E}(t) = \text{Re} \left[\tilde{\mathcal{E}}(t) \right] = \mathcal{E}_0 \cos(\omega t). \quad (10.65)$$

To demonstrate this is indeed easier, start with our differential equation, Eq. (10.22), for a RLC circuit in series:

$$L \frac{d^2 Q}{dt^2} + R \frac{dQ}{dt} + \frac{Q}{C} = \mathcal{E}. \quad (10.66)$$

Let's convert it to a complex form,

$$L \frac{d^2 Q}{dt^2} + R \frac{dQ}{dt} + \frac{Q}{C} = \mathcal{E}_0 e^{i\omega t}, \quad (10.67)$$

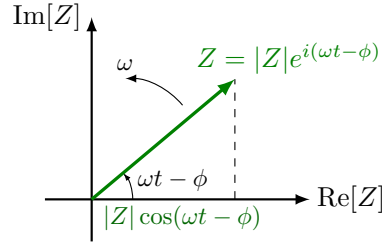


Figure 10.10: A phasor rotating with a fixed magnitude in the xy plane.

where we can promote the charge Q to a complex ansatz,

$$Q(t) = Q_0 e^{i\omega t}. \quad (10.68)$$

So we assumed the charge (and therefore the current) oscillated with the same angular frequency as \mathcal{E} . We will see below that if Q_0 is a complex number constant in time, there is actually an additional phase hidden in there. Substituting,

$$-LQ_0\omega^2 e^{i\omega t} + iRQ_0\omega e^{i\omega t} + \frac{Q_0}{C} e^{i\omega t} = \mathcal{E}_0 e^{i\omega t}, \quad (10.69)$$

and getting rid of the time-dependent factor $e^{i\omega t}$,

$$-LQ_0\omega^2 + iRQ_0\omega + \frac{Q_0}{C} = \mathcal{E}_0. \quad (10.70)$$

So it seem our ansatz worked as a solution! Namely, we can find Q_0 :

$$Q_0 = \frac{\mathcal{E}_0}{-L\omega^2 + iR\omega + \frac{1}{C}}. \quad (10.71)$$

The current I is simple the first derivative:

$$I = I_0 e^{i\omega t}, \quad (10.72)$$

with

$$I_0 = i\omega Q_0 = \frac{\mathcal{E}_0}{iL\omega + iR + \frac{1}{i\omega C}}, \quad (10.73)$$

where we used the fact that $i = -1/i$. In a more familiar form,

$$I_0 = \frac{\mathcal{E}_0}{Z}, \quad (10.74)$$

where Z is our *complex impedance*:

Complex impedance for a RLC circuit in series.

$$Z = iL\omega + R + \frac{1}{i\omega C} \quad (10.75)$$

Now notice that the impedance is a mix of purely imaginary terms $L\omega$ and $1/\omega C$, and one purely real term R . We can regroup these terms the impedance Z as

$$Z = R + iX, \quad (10.76)$$

where R and X are real quantities, and X is the total reactance

$$X = \omega L - \frac{1}{\omega C}. \quad (10.77)$$

It turns out that the complex impedance has the same behavior as a simple resistor:

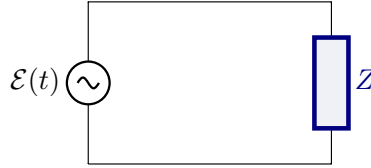


Figure 10.11: An AC circuit with a single impedance Z .

Sum of complex impedance. *The equivalent impedance of impedances in series is*

$$Z_{\text{eq}} = Z_1 + Z_2 + \dots \quad (10.78)$$

The equivalent impedance of impedances in parallel is

$$\frac{1}{Z_{\text{eq}}} = \frac{1}{Z_1} + \frac{1}{Z_2} + \dots \quad (10.79)$$

The impedance is represented in a circuit diagram by replacing the RLC components with one rectangular box as in Fig. 10.11.

We have found fixed formulas for the impedance due to each component:

Complex impedance of RLC components.

$$Z_R = R \quad \text{for a resistor} \quad (10.80)$$

$$Z_C = \frac{1}{i\omega C} = \frac{1}{\omega C} e^{-i\frac{\pi}{2}} \quad \text{for a capacitor} \quad (10.81)$$

$$Z_L = i\omega L = L\omega e^{i\frac{\pi}{2}} \quad \text{for an inductor} \quad (10.82)$$

What is cool about complex impedances, is that they actually encode the phase shift! You can see this by rewriting

$$Z_L = iL\omega = L\omega e^{i\frac{\pi}{2}} \quad (10.83)$$

and

$$Z_C = \frac{1}{i\omega C} = \frac{1}{\omega C} e^{-i\frac{\pi}{2}}. \quad (10.84)$$

In words, they are shifted by a phase $\pi/2$ with respect to Z_R . These phases will add to the time-dependent factor $e^{i\omega t}$. So when you add impedances as complex numbers using Eqs. 10.78 and 10.79, you also carry around the phase shifts in a compact way.

Let's look at some concrete examples.

10.5.1 RC circuit

Suppose there is no inductor ($L = 0$), only a resistor R capacitor C and capacitor C in series with an AC source \mathcal{E} . The complex impedance is given by

$$Z = R + iZ_C = R + \frac{1}{i\omega C} \quad (10.85)$$

This can be rewritten as

$$Z = |Z|e^{i\phi}, \quad (10.86)$$

where

$$|Z| = \sqrt{R^2 + \frac{1}{\omega^2 C^2}} \quad (10.87)$$

$$\tan \phi = \frac{1}{\omega RC}. \quad (10.88)$$

Therefore, the complex current is

$$I = \frac{\mathcal{E}}{Z} = \frac{\mathcal{E}_0}{|Z|} e^{i(\omega t - \phi)}. \quad (10.89)$$

So the real, physical current is

$$\text{Re}[I] = \frac{\mathcal{E}_0}{|Z|} \cos(\omega t - \phi). \quad (10.90)$$

The voltage drop over R is

$$\begin{aligned} V_R &= RI \\ &= R\mathcal{E}_0 e^{i(\omega t - \phi)}, \end{aligned} \quad (10.91)$$

while over V_C , there appears an extra phase shift:

$$\begin{aligned} V_C &= Z_C I \\ &= \left(\frac{1}{\omega C} e^{-i\frac{\pi}{2}} \right) \left(\frac{\mathcal{E}_0}{|Z|} e^{i(\omega t - \phi)} \right) \\ &= \frac{\mathcal{E}_0}{\sqrt{\omega^2 R^2 C^2 + 1}} e^{i(\omega t - \phi - \frac{\pi}{2})}. \end{aligned} \quad (10.92)$$

10.5.2 RLC circuit in series

Now, go back to a RLC circuit in series. The complex impedance is now given by

$$Z = R + i \left(\omega L - \frac{1}{\omega C} \right). \quad (10.93)$$

This has a magnitude

$$|Z| = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2}, \quad (10.94)$$

and a complex phase ϕ given by

$$\tan \phi = \frac{\omega^2 LC - 1}{\omega RC}. \quad (10.95)$$

From this we can immediately compute the complex phasors V_R , V_L and V_C . This is left as an exercise.

10.5.3 RLC circuit in parallel

For a RLC circuit in parallel,

$$\frac{1}{Z} = \frac{1}{R} + \frac{1}{i\omega L} + i\omega C \quad (10.96)$$

What are the modulus $|Z|$ and complex angle ϕ of Z ?

Chapter 11

Magnetization

11.1 Magnetic moment

In Section 7.4, we had seen that atoms have magnetic moments due to the electrons orbiting the nucleus. Consider a hydrogen atom with just one electron. It has a charge $q = e$ and mass m_e . Classically, the electron moves with some velocity v around the nucleus in a circle with a radius r , just like a planet orbits the sun. It therefore has an angular momentum

$$L = m_e v r, \quad (11.1)$$

and for a circle, the magnetic moment is

$$m = IA = I\pi r^2. \quad (11.2)$$

The current of a charge with orbital period T is $I = q/T$. To go in a full circle around the nucleus is $T = 2\pi r/v$. So the electrical current due to the orbiting electron is

$$I = \frac{qvr}{2\pi r}. \quad (11.3)$$

Then, rewriting the magnetic momentum with Eq. (11.3) and Eq. (11.1):

$$m = \frac{qL}{2m_e}. \quad (11.4)$$

Let's put a box around it.

Classical magnetic moment of a charge moving in a circle. For a general charge q with mass m_q ,

$$\mathbf{m} = \frac{q\mathbf{L}}{2m_q}. \quad (11.5)$$

For $q > 0$, \mathbf{m} and \mathbf{L} are in the same direction, while for $q < 0$, they oppose each other.

11.2 Quantization of magnetic moment

Equation (11.5) is a general relation that hold in classical mechanics and electrodynamics. It also holds for quantum theory actually. However, in the quantum theory for the atom,

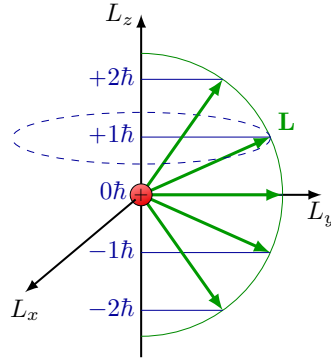


Figure 11.1: Orbital angular momentum with magnetic quantum number $m = 0, \pm 1, \pm 2$.

the orbital angular momentum is *quantized*. The orbital angular momentum always has is a discrete multiple of the constant

$$\hbar = \frac{h}{2\pi}, \quad (11.6)$$

where Plank's and Dirac's constant are, respectively,

$$h = 6.67 \times 10^{-34} \text{ J s}, \quad (11.7)$$

$$\hbar = 1.05 \times 10^{-34} \text{ J s}. \quad (11.8)$$

Customarily, the z component of the angular momentum \mathbf{m} along the external B field is considered, and turns out to be an integer multiple of \hbar :

$$L_z = n\hbar, \quad (11.9)$$

for any integer $n = 0, \pm 1, \pm 2, \pm 3, \dots$ This integer is called the *magnetic quantum number* and is often denoted with m (not to be confused with the magnetic moment itself). The quantization of L_z is illustrated in Fig. 11.1. Correspondingly, the z component of the magnetic momentum for a hydrogen atom becomes

$$m_z = -\frac{en\hbar}{2m_e} = -\frac{e\hbar}{2m_e} \frac{L_z}{\hbar} = -m_B \frac{L_z}{\hbar}. \quad (11.10)$$

In the last equation, we recognize the magnetic moment for the electron,

$$m_B = \frac{e\hbar}{2m_e} = 9.27 \times 10^{-24} \text{ A m}^2. \quad (11.11)$$

It has been measured to great precision.

11.3 Magnetization

Magnetization of a material is due to the combined effect of the magnetic moments of the atoms. In any material, you have many atoms, each with their own electrons orbiting the nucleus. As discussed in the previous section, each minuscule loop of current causes a small magnetic moment. For most materials, however, the magnetic moments rotate due to heat, and are all pointing in random directions, such that in total, the magnetic moments cancel.

What if these magnetic moments all align in the same direction? The adjacent currents will cancel in between the atoms, but on the outside, they will add up and cause a macroscopic *magnetization*.



(a) Each atom has its own small current loops, and their own magnetic moment. (b) One can think of the microscopic currents adding up to one big one.

Figure 11.2: From microscopic to macroscopic magnetic moment in a magnetized material, in which the atomic magnetic moments are aligned.

Magnetization.

$$\mathbf{M} = \frac{d\mathbf{m}}{dV} \quad (11.12)$$

where V denotes the volume. This effect is illustrated in Fig. 11.2. We know that for a small current i with loop area A ,

$$d\mathbf{m} = A d\mathbf{i}. \quad (11.13)$$

Since $dV = A d\ell$, where ℓ is the length of the magnet, Eq. (11.12) becomes

$$\mathbf{M} = \frac{d\mathbf{i}}{d\ell}. \quad (11.14)$$

Consequently, we can think of \mathbf{M} as the current per unit of length.

For a hollow solenoid we have $B = \mu_0 n I$, such that we get

$$M = \frac{di}{d\ell} = nI. \quad (11.15)$$

Therefore, we recognize that

$$\mathbf{B} = \mu_0 \mathbf{M}, \quad (11.16)$$

where \mathbf{M} is the magnetization per volume.

11.3.1 Magnetization in an external field

Magnetization depends on the material *and* the external magnetic field:

$$\mathbf{M} = \chi_M \frac{\mathbf{B}_{\text{ext}}}{\mu_0}, \quad (11.17)$$

where χ_M is the *magnetic susceptibility*. It satisfies the relation

$$\frac{\mu}{\mu_0} = 1 + \chi_M \quad (11.18)$$

Values for some materials are shown in Table 11.1.

The total magnetic field then, is a combination of the external magnetic field and the magnetization.

$$\mathbf{B} = \mathbf{B}_{\text{ext}} + \mu_0 \mathbf{M}. \quad (11.19)$$

Substituting Eq. (11.17):

$$\mathbf{B} = \mathbf{B}_{\text{ext}} (1 + \chi_M). \quad (11.20)$$

This formula works for *paramagnetic* and *diamagnetic* materials, but not well for *ferromagnetic* materials like iron.

Table 11.1: Susceptibility χ_M of several materials.

Material	χ_M
Aluminium	2.5×10^{-5}
Gold	-3.6×10^{-5}
Bismuth	-1.7×10^{-5}
Nickel	600
Pure iron	200 000
Copper	-9.6×10^{-6}
Water	-9.0×10^{-6}
Graphite	1.5×10^{-5} or $1.3 \times 10^{-3*}$

(*) Depending on orientation.

11.3.2 Paramagnetism

Paramagnetic materials have a small, positive value χ_M . The atom magnetic momenta are randomly aligned. They tend to align in an external field. (This behavior is explained by the open-shell structure of the electron orbitals.)

But thermal motion counteracts this tendency. The two effects of thermal excitations and alignment with the magnetic field compete: Flipping the magnetic moment of an atom requires energy (Fig. 11.3). Remember that $U = -\mathbf{m}_e \cdot \mathbf{B}$. To flip the \mathbf{m} , you need

$$\Delta U = 2m_e B. \quad (11.21)$$

For a field of $B = 1$ T, you would need

$$\Delta U = 2(9.27 \times 10^{-24} \text{ J/T})(1 \text{ T}) \sim 2 \times 10^{-23} \text{ J} \quad (11.22)$$

Now assume a room temperature of $T = 300$ K. The atoms have a thermal energy

$$k_B T \approx 4 \times 10^{-21} \text{ J}. \quad (11.23)$$

where k_B is the Boltzmann constant, which you will learn about in the next chapters on thermodynamics. The thermal energy is typically much more than the magnetic potential energy of an atom. So the magnetization is stronger at lower temperatures

Curie's law.

$$M = \frac{m B_{\text{ext}}}{3k_B T} M_S. \quad (11.24)$$

If you look at the realistic plot of M vs. B_{ext} in Fig. 11.4a, you will see there is actually a plateau. This is the saturation value M_S , at which all the atomic magnetic momenta are fully aligned, and thus the material is maximally magnetized. However, Curie's law holds approximately for smaller values of external fields (the straight line in Fig. 11.4a).

On the other hand, the magnetization will decrease with temperature as $M \sim T^{-1}$.

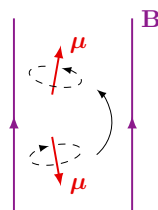
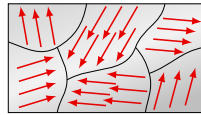


Figure 11.3: It costs some energy to flip a magnetic moment.

(a) Plot of M vs. B_{ext} . Curie's law holds for low B_{ext} .(b) Plot of M vs. temperature T .**Figure 11.4:** Plots of magnetization M of paramagnetic material.**Figure 11.5:** The atomic magnetic moments of most non-magnetized materials point in different directions, although they are grouped in “domains”.

11.3.3 Ferromagnetism

Ferromagnetic materials have large positive values of χ_M . Examples are iron, cobalt and nickel. Atoms exert strong forces on their neighbors, causing alignment in groups called *domains*, illustrated in Fig. 11.5. But if we increase \mathbf{B}_{ext} , the domains tend to align, causing a macroscopic field.

We can magnetize an iron rod inside a solenoid. From Eq. (11.19),

$$\mathbf{B} = \mathbf{B}_{\text{ext}} + \mu_0 \mathbf{M} = \mu_0 n I (1 + \chi_M) = \mu n I. \quad (11.25)$$

We now see where the peculiar definition Eq. (11.18) for magnetic susceptibility χ_M originates from. Like before in Section 8.2.4,

$$\kappa_M = \frac{\mu}{\mu_0} = 1 + \chi_M \quad (11.26)$$

Iron, for example, $\kappa_M \sim 5000$. More values are listed in Table 8.1.

Hysteresis

We study what happens to un-magnetized iron when we apply an external field \mathbf{B}_{ext} in several steps, plotted Fig. 11.6:

1. Start with un-magnetized iron, and then apply an increasing external field \mathbf{B}_{ext} . The magnetization is along \mathbf{B}_{ext} .
2. Decrease \mathbf{B}_{ext} to zero. There is still a residual magnetic field left over, \mathbf{B}_r , from aligned domains.
3. Increase \mathbf{B}_{ext} in the negative direction. The magnetization lags \mathbf{B}_{ext} , but eventually also points in the negative direction.
4. Increase \mathbf{B}_{ext} again to zero, and we see the same residual magnetization \mathbf{B}_r in the opposite direction.
5. We can repeat it again, (not using path 1).

This is known as the *hysteresis curve*: The magnetization depends on the history of the external field B_{ext} . When the external field goes back to $B_{\text{ext}} = 0$, the ferromagnetic material obtains a “remnant” magnetic field, \mathbf{B}_r , and stays magnetized on its own.

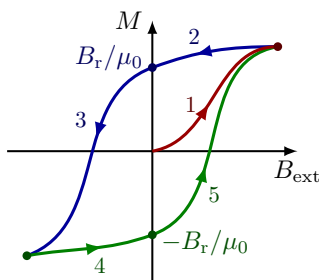


Figure 11.6: Hysteresis: The magnetization depends on the history.

11.3.4 Diamagnetism

Diamagnetic materials have small negative values of $\chi_M < 0$. The magnetic moments do not align with an external magnetic field; they are “stuck”. But the materials are repelled by either side of a magnet! This strange behavior was discovered in bismuth by Faraday in 1846.

Diamagnetic materials typically have atoms with a pair of electrons orbiting in the opposite direction. This should mean that their magnetic moments oppose each other,

$$\mathbf{m}_{\text{atom}} = \mathbf{m}_1 + \mathbf{m}_2 = 0. \quad (11.27)$$

To understand how this causes diamagnetic materials to repel from magnets, we need to invoke Lenz’s law. If we turn on an external magnetic field, Lenz’s law says that the current should oppose an increasing external magnetic field. If we assume the radius r of the electron orbit does not change, then the velocity of the electron has to. The current of a current loop is

$$I = \frac{qv}{2\pi r}. \quad (11.28)$$

To increase the current by ΔI , there has to be some increase of velocity Δv . Assume that the magnetic momentum \mathbf{m}_1 due to one electron is aligned with \mathbf{B}_{ext} , then the velocity *decreases* $\Delta v < 0$. So while the \mathbf{B}_{ext} increases, Lenz’s law implies that \mathbf{m}_1 becomes smaller. Meanwhile, the other electron has a magnetic moment \mathbf{m}_2 anti-parallel to \mathbf{B}_{ext} , and the velocity *increases* $\Delta v > 0$. This time, Lenz’s law tells us that when \mathbf{B}_{ext} increases, \mathbf{m}_2 becomes bigger. So now, $\mathbf{m}_1 + \mathbf{m}_2 \neq 0$ and opposes \mathbf{B}_{ext} .

Notice that this magnetic field is induced because \mathbf{B}_{ext} increases, and therefore, the magnetic flux changes. However, this magnetization will actually persist as long as \mathbf{B}_{ext} is kept turned on. This is not obvious, but it is because atoms do not lose energy (e.g. due to heat) caused by the moving charge, as we will learn later in quantum mechanics. So until \mathbf{B}_{ext} changes again, the velocity v will stay changed.

11.3.5 Superconductor

A superconductor is a perfect diagnet. It creates a magnetic field that cancels the external magnetic field. Therefore, if

$$B = B_{\text{ext}}(1 + \chi_M) = 0, \quad (11.29)$$

then χ_M must be -1 . So a superconducting material is always repelled by an external magnetic field in any orientation. This leads to cool experiments, where you can levitate small objects, or even frogs!

Noteworthy, Alex Müller from the the University of Zürich won the Nobel Prize in 1986 for creating a superconductor at a temperature of 35 K.

Chapter 12

Maxwell's Equations

We have now come to a point where we can summarize the fundamental laws that with the Lorentz force describe all of classical electromagnetism, optics and electric circuits. These are the *four Maxwell equations*.

Remember that in Section 2.4 we had seen how to convert Gauss's law from a surface integral to a volume integral, invoking the fundamental integral theorem from calculus:

$$\oint_S \mathbf{v} \cdot d\mathbf{A} = \int_V (\nabla \cdot \mathbf{v}) dV, \quad (12.1)$$

where \mathbf{v} is some vector field. With this theorem, we could write a purely differential form of Gauss's law, without any integrals. The same can be done for magnetic fields:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (12.2)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (12.3)$$

These are two of Maxwell's equations.

We also convert the remaining two Maxwell equations, Ampère's law (Eq. (8.9)) and Faraday's law (Eq. (8.34)), from an integral to differential form. But first we need to borrow another theorem from calculus.

12.1 Differential form of Ampère's law & Faraday's law

12.1.1 Interlude: Kelvin-Stokes theorem

Now we will also need Kelvin-Stokes theorem to convert the line integrals in Ampère's law (Eq. (8.9)) and Faraday's law (Eq. (8.34)).

$$\oint_C \mathbf{v} \cdot d\boldsymbol{\ell} = \int_S (\nabla \times \mathbf{v}) \cdot d\mathbf{A}. \quad (12.4)$$

Remember that the curl of a vector field \mathbf{v} is given by

$$\nabla \times \mathbf{v} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_x & v_y & v_z \end{vmatrix} \quad (12.5)$$

$$= \left(\frac{\partial v_z}{\partial x} - \frac{\partial v_y}{\partial y} \right) \hat{\mathbf{x}} + \left(\frac{\partial v_x}{\partial x} - \frac{\partial v_z}{\partial y} \right) \hat{\mathbf{y}} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right) \hat{\mathbf{z}} \quad (12.6)$$

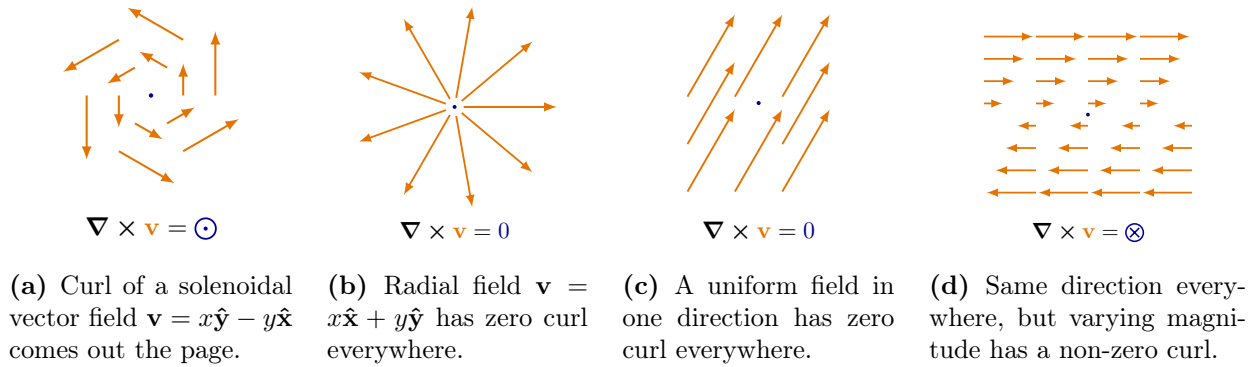


Figure 12.1: Intuitive examples of curl in a point of a simple vector fields \mathbf{v} . The curl of a vector field follows the right-hand rule. Compare this to the examples of divergence in Fig. 2.8.

So unlike the divergence of a vector field, which is a scalar field (see Eq. (2.35)), the curl of a vector field is again a vector field. Intuitively, the curl in a given point can in most cases be thought of as a measure of how much the field's vectors rotate around it. The direction of rotation at a point is given by imagining how a paddle wheel (\otimes) you drop there will rotate. The direction of the curl follows again the right-hand rule: Curl your fingers along the arrows of \mathbf{v} , and your thumb points along $\nabla \times \mathbf{v}$. Some examples of this are shown in Fig. 12.1.

12.2 Ampère's law

Ampère's law is easy. It already has a line integral, which can now be rewritten as a surface integral

$$\mu_0 I_C = \oint_C \mathbf{B} \cdot d\boldsymbol{\ell} = \int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{A}. \quad (12.7)$$

But we also need to rewrite the current as a surface integral. This can be done by using the current density \mathbf{j} , which we defined back in Section 5.2:

$$I = \int_S \mathbf{j} \cdot d\mathbf{A}. \quad (12.8)$$

This is Maxwell's fourth law:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \quad (12.9)$$

We will see later that there is actually a term missing.

12.3 Faraday's law

Faraday's law also has a line integral, namely over the circuit in which an EMF is induced:

$$-\frac{d\Phi_M}{dt} = \mathcal{E} = \oint_C \mathbf{E} \cdot d\boldsymbol{\ell} = \int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{A}. \quad (12.10)$$

Meanwhile, the magnetic flux is simply

$$\Phi_M = \int_S \mathbf{B} \cdot d\mathbf{A}. \quad (12.11)$$

Therefore, we can write

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (12.12)$$

12.4 Maxwell's equations

We are now ready to formulate an [integral](#) and a [differential form](#) of the Maxwell's equations.

Maxwell's equations (1861). *The four equations of Maxwell are:*

1. *Gauss's law:*

$$\oint_S \mathbf{E} \cdot d\mathbf{A} = \frac{Q}{\epsilon_0} \longleftrightarrow \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (12.13)$$

2. *Faraday's law*

$$\oint_C \mathbf{E} \cdot d\boldsymbol{\ell} = \frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{A} \longleftrightarrow \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (12.14)$$

3. *No magnetic monopoles*

$$\oint_S \mathbf{B} \cdot d\mathbf{A} = 0 \longleftrightarrow \nabla \cdot \mathbf{B} = 0 \quad (12.15)$$

4. *Ampere's law*

$$\oint_C \mathbf{B} \cdot d\boldsymbol{\ell} = \mu_0 I + \mu_0 \epsilon_0 \frac{d}{dt} \int_S \mathbf{E} \cdot d\mathbf{A} \longleftrightarrow \nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad (12.16)$$

The differential forms are basically differential equations, that can be solved for some configuration of charge and current distribution ρ and \mathbf{j} . In fact, in vacuum, $\rho = 0$ and $\mathbf{j} = 0$, and one can obtain coupled electric and magnetic fields \mathbf{E} and \mathbf{B} that oscillate harmonically. These vacuum solutions are exactly what we know to be electromagnetic waves! They propagate with a velocity

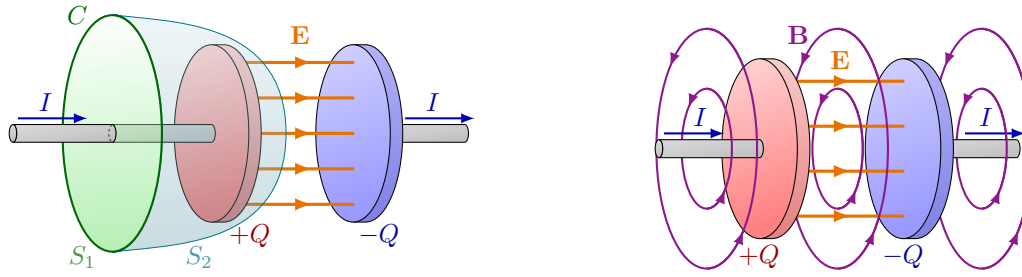
$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx 3 \times 10^8 \text{ m/s}. \quad (12.17)$$

We will come back to this in a later chapter.

Maxwell's first equation can be interpreted as the fact that electric field lines always start ("diverge") from positive charges (sources $\rho > 0$) and end in negative ones (sinks $\rho < 0$), while the third equation basically means that there are no magnetic monopoles in Nature, and therefore magnetic field lines always form closed loops.

We had already seen that Faraday's law, Maxwell's second equation, tells you that a changing magnetic field ($\partial \mathbf{B} / \partial t$) gives rise to an EMF. This can also be seen intuitively from the differential form, because an EMF is basically a circulating electric field, meaning $\nabla \times \mathbf{E} \neq 0$.

Finally, Ampère's law, Maxwell's fourth equation, can be interpreted as the fact that a flowing electric current \mathbf{j} gives rise to a magnetic field \mathbf{B} that circles around the current \mathbf{j} , because of the curl $\nabla \times \mathbf{B}$. Maxwell had realized that there needed to be an extra term for this equation to be complete, $\partial \mathbf{E} / \partial t$. We will explain in the next section.



(a) Two possible open surfaces, S_1 and S_2 , to compute the current I_C through the Ampère curve C . (b) Magnetic field due to a changing electric field in the capacitor is similar to that of a wire carrying an electric current.

Figure 12.2: The magnetic field due to a charging capacitor follows $\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \partial \mathbf{E} / \partial t$.

12.4.1 Displacement current

In the last formula, Eq. (12.16), we have added an extra term that can be identified with the electric flux

$$\Phi_E = \int_S \mathbf{E} \cdot d\mathbf{A}. \quad (12.18)$$

This term is called *Maxwell's correction*, or *Maxwell's displacement current*. To see why we need a correction, consider a capacitor that is being charged. If we want to use Ampère's law to compute the magnetic field due to the current I , we need to integrate over a closed curve C , like in Fig. 12.2. This is the difficulty: What is the enclosed current I_C ? Is it the current flowing through the flat surface S_1 ($I_C = I$), or through the bulging S_2 ($I_C = 0$)? There is no reason to prefer S_1 above S_2 . This shows that I_C is ill-defined in the case where the current is discontinuous. However, notice there is an electric field \mathbf{E} between the capacitor plates. As the capacitor is charging over time, the amount of charge increases

$$\frac{dQ}{dt} = \epsilon_0 \frac{d}{dt} \int_S \mathbf{E} \cdot d\mathbf{A} = \epsilon_0 \frac{d\Phi_E}{dt}, \quad (12.19)$$

where we used Gauss's law on $S = S_1 + S_2$. The left-hand side is the so-called *displacement current*,

$$I_d = \frac{dQ}{dt}. \quad (12.20)$$

This should give us a hint that there is a correlation between the current I_C and changing electric flux.

There is a different way to see why Maxwell's fourth equation is incomplete and see where the correction term originates from. Assume—wrongly for a moment—that the term was not there:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}. \quad (12.21)$$

Taking the divergence of both sides,

$$\nabla \cdot (\nabla \times \mathbf{B}) = \mu_0 \nabla \cdot \mathbf{j}, \quad (12.22)$$

even if the electric field changes in time. Using the fact that the cross product is orthogonal to both vectors ($a \cdot (a \times b) = 0$ because $a \perp (a \times b)$),

$$0 = \nabla \cdot (\nabla \times \mathbf{B}) = \mu_0 \nabla \cdot \mathbf{j} \quad (12.23)$$

But this is a problem: In general, the divergence of the current is not exactly zero. But wait! We recognize this term from our discussion on the continuity equation in Section 5.3:

$$\nabla \cdot \mathbf{j} = -\frac{\partial \rho}{\partial t}. \quad (12.24)$$

In fact, using Maxwell's first law,

$$\nabla \cdot \mathbf{j} = -\epsilon_0 \frac{\partial}{\partial t} (\nabla \cdot \mathbf{E}) = -\epsilon_0 \nabla \cdot \frac{\partial \mathbf{E}}{\partial t}. \quad (12.25)$$

So we conclude that we can prevent a breakdown of Maxwell's fourth equation, by including the term

$$\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad (12.26)$$

or,

$$\mu_0 \epsilon_0 \frac{d}{dt} \int_S \mathbf{E} \cdot d\mathbf{A} \quad (12.27)$$

in the integral form.

This extra term suggests that a magnetic field is created, not only by a current, but by a changing electric flux at locations where there is no current from moving charges.

Going back to the charging capacitor, even though there is no current flowing between its terminals in the form of moving charges, there is a displacement current due to the changing electric field. So a capacitor placed in a circuit will not interrupt the flow of charge carriers (i.e. current) in that circuit, as long as it is being charged, and additionally, it will create a magnetic field, just like the current-carrying wires.

Chapter 13

Electromagnetic Waves

In Chapter 12, we saw the Maxwell equation are a system of differential equations. They can be solved for given charge and current distributions, ρ and \mathbf{j} . Their solutions are dynamic function $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$. “Dynamic”, because they vary with time.

If one solves the Maxwell equations in vacuum, where $\rho = 0$ and $\mathbf{j} = 0$, one finds a set of wave equations that describe the propagation of *electromagnetic waves*, or *electromagnetic radiation*. The derivation is discussed in more detail in the course *Vertiefung zu Physik* (PHY120).

An electromagnetic wave are depicted in Fig. 13.1. It is composed of an electric and magnetic field \mathbf{E} and \mathbf{B} that travel in the $\mathbf{v} \propto \mathbf{E} \times \mathbf{B}$ direction, and oscillate perpendicularly to the direction of motion \mathbf{v} , and perpendicular to each other. The oscillation is harmonic, so ignoring a phase, they must be of the form

$$\mathbf{E} = E_0 \sin(kz - \omega t) \hat{\mathbf{x}} \quad (13.1)$$

$$\mathbf{B} = B_0 \sin(kz - \omega t) \hat{\mathbf{y}}, \quad (13.2)$$

where always $E_y = 0$ and $B_x = 0$, the angular frequency is given by the frequency f as

$$\omega = 2\pi f, \quad (13.3)$$

in units of radians per second, while the *wave number* k is given by

$$k = \frac{2\pi}{\lambda} \quad (13.4)$$

in units of radians per meter. If the velocity is c , remember that the wavelength is related to frequency as

$$\lambda = \frac{c}{f} = \frac{2\pi c}{\omega} \quad (13.5)$$

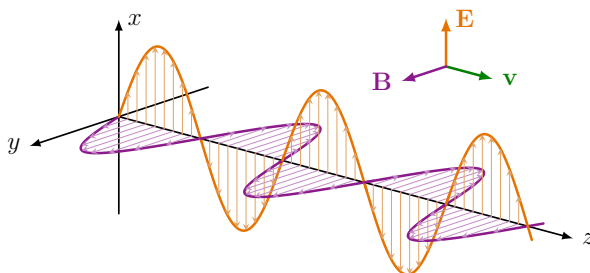


Figure 13.1: An electromagnetic wave.

Therefore, k is related to frequency as

$$k = \frac{\omega}{\lambda} = \frac{2\pi f}{c}. \quad (13.6)$$

Another interesting relationship, which follows from the Maxwell equations, is

$$E = cB \quad (13.7)$$

in any point of an electromagnetic wave. It's easy to show that the units work out.

13.1 Electromagnetic spectrum

Electromagnetic radiation is everywhere around us; radio waves, infrared (IR), visible light, ultraviolet (UV), X rays (Röntgen), gamma rays. Their wavelengths, frequencies and energies are compared in Fig. 13.2 and Table 13.1.

Our eyes can only perceive the *optical* or *visible spectrum*, which ranges from violet ($\lambda \sim 400$ nm) to red ($\lambda \sim 700$ nm), but some animals can also see into the IR, or UV.

While hiding from the Great Plague during his “year of wonders”, 1666, Newton had used a prism to demonstrate that white light is actually a combination of all the colors of the rainbow. He divided the rainbow into the seven ROYGBIV colors: red, orange, green, blue (cyan), indigo (dark blue), and violet. However, this division of a continuous spectrum is somewhat arbitrary and has varied over time and between cultures. Newton aimed for seven to match the number of musical notes in the western major scale, days in the week and planets known at his time.

Table 13.1: Rough classification of electromagnetic spectrum and its applications.

Class	Wavelength λ	Frequency f	Energy E	Application
Radio waves	> 1 m	< 300 MHz	$< 1.24 \mu\text{eV}$	Radio & TV broadcast, telecommunications maritime navigation
Microwaves	1 mm – 1 m	300 MHz – 300 GHz	1.24 meV – 1.24 μeV	Microwave oven, radar, mobile phones, 4G, Wi-Fi, satellite communications, GPS, cosmic microwave background
Infrared	750 nm – 1 mm	300 GHz – 400 THz	1.24 meV – 1.7 eV	Thermal imaging, TV remote control, night vision, bio imaging, optical fibers
Visible	400 nm – 750 nm	400 THz – 750 THz	1.7 eV – 3 eV	Human vision, illumination, photography, microscopes, lasers
Ultraviolet	10 nm – 400 nm	750 THz – 30 PHz	3 eV – 124 eV	Disinfection, dental curing, black lights, sun tanning, counterfeit detector
X rays	0.01 nm – 10 nm	30 PHz – 30 EHz	124 eV – 124 keV	Crystallography, radiation therapy, medical imaging, security scans
Gamma rays	< 0.01 nm	> 30 EHz	> 124 keV	Radioactive sources, cancer treatments, PET scans, cargo container screening

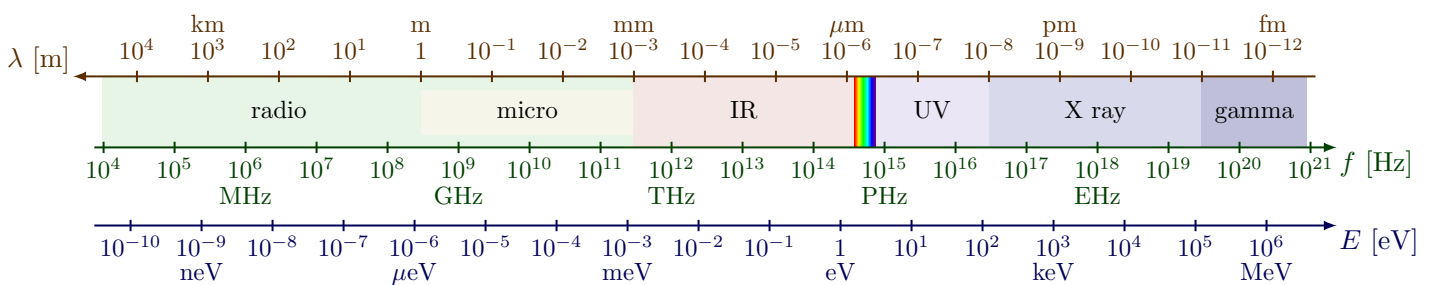


Figure 13.2: Electromagnetic spectrum with rough classification.

There is a wide range of applications and natural phenomena of electromagnetic radiation. Some examples are given in Table 13.1 and below. The classification of different *classes* or *bands* of the spectrum are also somewhat arbitrary and varies between sources.

UV light with energies roughly above 10 eV and any radiation with larger energies (smaller wavelengths) have enough energy to knock electrons loose from their atoms. For this reason, the short-wavelength part of the spectrum is called *ionizing*, while longer wavelengths are called *non-ionizing*. Ionizing radiation can be dangerous, as it can cause damage to biological tissues and DNA. This is why the sun can give you a sunburn and an increased probability of skin cancer, but also why exposure to X ray machines at the hospital and airport, or to radioactive sources with gamma decay should be limited. On the flip side, it can be used for sterilization to kill most bacteria, but also viruses, for example in unfiltered water from a lake. Melanine, which is a biological pigment that gives color to the skin, but also to hair and dark eyes, can very effectively absorb UV radiation, and therefore protect the cells. Glass and sun screen also absorb a large fraction of UV, and so they are not transparent in UV and both protect you from a sunburn. To protect from ionizing radiation with higher energies, one needs enough metal shielding that absorb the radiation. Typically lead is used, because it has a high density. Non-ionizing radiation can still excite the electrons in atoms, which give rise to emission (like in gas-discharge lamps) and absorption spectral lines (like in stellar spectra). But such radiation can also induce currents and subsequently heating of objects whose size is similar to that of the wavelength. Microwave ovens on the other hand, work by making water molecules oscillate with the electric field (see electric dipoles in Section 1.4).

Micro waves and radio waves are useful for telecommunication because they can travel long distances. Thanks to their long wavelengths, they are less likely to be stopped by obstacles. Another advantage is that the atmosphere does not tend to absorb a large band of radio waves, like it mostly does for IR, UV, X rays and gamma radiation. This is called the *radio window* and allows for satellite communications and radio telescopes on earth, while astronomical telescopes detecting IR, UV and X ray have to be in space, outside the atmosphere. Similarly, an *optical window* allow us to gaze up to the moon and stars, with or without a telescope, while the atmosphere offers us protection from most harmful (ionising) radiation. However, short-wavelength radio waves can be reflected or refracted by the ionosphere, the upper part of the atmosphere, overcoming the curvature of the earth and extending their range to other continents.

Microwave ovens operate around 2.45 GHz, which is around 12.2 cm. It can therefore interfere with Wi-Fi, which uses bands around 2.4 GHz (12 cm) and 5 GHz (6 cm).

13.2 Polarization

If the \mathbf{E} and \mathbf{B} fields only oscillate in one direction, the wave is said to be *linearly polarized*. However, electromagnetic waves with shapes more complicated than sine waves exist, but as we know from Fourier analysis, any shape can be decomposed into a sum of sine waves of different amplitude and phases. One can make almost any electromagnetic wave by varying the amplitude and wave with wavelength and time. This will be discussed in more detail in Section 14.4.

13.3 Speed of light

As a solution of the Maxwell's equation, the speed of light in vacuum is given by the only constants involved in these equations:

Speed of light in vacuum.

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}. \quad (13.8)$$

Due to interactions with a medium, the speed of light can decrease:

Speed of light in a medium.

$$c' = \frac{1}{\sqrt{\epsilon \mu}}, \quad (13.9)$$

where ϵ is the permittivity, and μ is the permeability of the medium (see Sections 4.2 and 8.2.4). This leads to a ratio that is useful to study refraction:

Refraction index.

$$n = \frac{c}{c'} = \sqrt{\frac{\epsilon \mu}{\epsilon_0 \mu_0}}. \quad (13.10)$$

Here, n is the *refraction index* of the speed of light between vacuum and the medium. For air, this is $n = 1.000029$. Because $n > 1$ in all media, $c' < c$.

13.3.1 Measurement of speed of light

The earliest measurement of the speed of light was in 1676 by the Danish physicist Ole Rømer. He noticed that the time between eclipses of the moon Io by Jupiter was not constant. It depended on whether the Earth moved towards or away from Jupiter.

Other early measurement light speed used a “toothed” wheel that rotates such that it would block a beam of light periodically. Light that was let through a gap of the wheel would travel some long distance d and bounce off a mirror. The mirror is slightly angled such that reflected light beam returns can enter through a gap that directly neighbours the tooth blocking the incoming beam. If the wheel is rotating at the exact right speed, the incoming and reflected light beams are let through by neighbouring gaps at the same time. By measuring the rotation speed at which both the incoming and reflected beams pass through the hole, one can calculate the speed of light.

13.3.2 Special relativity

One of the insights that lead Einstein to the formulation of special relativity in 1905, was that the speed of light is constant in any reference frame. This was observed by the *Michelson–Morley experiment* earlier in 1887. As a consequence of *Einstein’s theory of special relativity*, the speed of light in vacuum is a sort of speed limit on any type of matter or radiation imposed by the universe: Nothing can travel faster than the speed of light.

Not only was the constancy of the speed of light observed, it also seems to follow from Maxwell’s equations. Namely, Eq. (13.8) only depends on vacuum constants ϵ_0 and μ_0 , which do not change in a reference frame that is moving with some constant velocity to the inertial lab frame.

One curious thing about the fact that the speed of light in a medium, c' , can be less than the speed of light in vacuum, c , is that some massive particles like electrons could actually travel faster than light in that medium. This leads to interesting phenomena such

as *Cherenkov light*, which is exploited by some detectors to measure the speed of those particles. Still, massive particles can never go faster than light in the vacuum.

13.3.3 Light year

For astronomical distances, it is useful to define a *light year*, the distance light travels in one year,

$$1 \text{ ly} = c \cdot (1 \text{ y}) = 9.5 \times 10^{15} \text{ m}, \quad (13.11)$$

which unit of distance, even though “year” appears in the name.

13.4 Intensity & energy density

So the electromagnetic fields must have some energy stored. In Sections 4.36 and 9.2.1 we had already derived the total energy stored in electric and magnetic fields due to capacitors and solenoids in a circuit, but these equations hold for electromagnetic waves as a simple sum:

Energy density of electromagnetic waves.

$$\eta = \eta_E + \eta_M \quad (13.12)$$

$$= \frac{1}{2} \epsilon_0 E^2 + \frac{1}{2} \frac{B^2}{\mu_0} \quad (13.13)$$

$$= \epsilon_0 E^2 = \frac{B^2}{\mu_0} = \frac{EB}{\mu_0 c} = EBc\epsilon_0. \quad (13.14)$$

In the last line, Eqs. (13.7) and (13.8) were used, because

$$\frac{1}{2} \epsilon_0 E^2 = \frac{1}{2} \epsilon_0 (cB)^2 = \frac{1}{2} \frac{B^2}{\mu_0}, \quad (13.15)$$

$$\frac{1}{2} \frac{B^2}{\mu_0} = \frac{1}{2} \frac{(E/c)^2}{\mu_0} = \frac{1}{2} \epsilon_0 E^2. \quad (13.16)$$

The *instantaneous intensity* is defined as the power per unit area. This can therefore be expressed as the energy density times the wave speed,

$$I = \eta c = \frac{EB}{\mu_0}, \quad (13.17)$$

which varies instantaneously because E and B vary with time. The intensity has a direction

$$\mathbf{S} = \frac{\mathbf{E} \times \mathbf{B}}{\mu_0}. \quad (13.18)$$

This intensity vector is called the *Poynting vector* of the wave. Notice its size is the intensity $|\mathbf{S}| = S = I$. In the case of a simple sine wave case:

$$\mathbf{S} = \frac{E_0 B_0}{\mu_0} \sin^2(kz - \omega t) \hat{\mathbf{z}}. \quad (13.19)$$

This points in the direction of travel. It also varies with time, so what is the average intensity? We need to average over an discrete (i.e. integer) number of waves:

$$\langle I \rangle = \langle S \rangle = \frac{E_0 B_0}{2\mu_0} = \frac{c B_0^2}{2\mu_0} = \frac{\epsilon_0 c E_0^2}{2} \quad (13.20)$$

because $\langle \sin^2 x \rangle = 1/2$ over one wavelength, similar to Eq. (10.40).

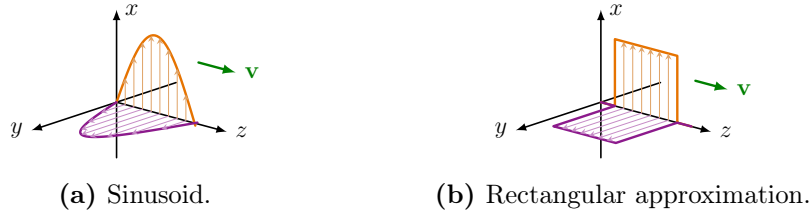


Figure 13.3: Electromagnetic pulse.

13.5 Momentum

Electromagnetic waves also carry momentum as well as energy. One way to see this is to consider what would happen when an electric charge is hit with an electromagnetic wave pulse. There will be a Lorentz force due to the electromagnetic force

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}. \quad (13.21)$$

For simplicity, let's make a rough approximation, assuming the pulse has a constant amplitude E and B for a short interval Δt , as in Fig. 13.3b. If the charge have initial velocity $v = 0$ and gets hit at $t = 0$, the force for $t < \Delta t$ will be given by

$$F_x = qE = \frac{mv_x}{t} F_y = qv_x B = \frac{q^2 E B t}{m}, \quad (13.22)$$

or, after Δt

$$v_x = \frac{qE}{m} \Delta t. \quad (13.23)$$

So the kinetic energy is

$$K = \frac{mv^2}{2} = \frac{q^2 E^2}{2m} \Delta t^2. \quad (13.24)$$

To figure out the how much momentum the particle gained from the magnetic field, we need to find the impulse, which is the integral of the force:

$$\Delta p = \int_0^{\Delta t} F_y dt = \frac{q^2 E B}{2m} \Delta t^2 = \frac{q^2 c^2 E^2}{2m} \Delta t^2. \quad (13.25)$$

Comparing these two equations, we see that

Total energy of an electromagnetic wave.

$$E = pc. \quad (13.26)$$

Here $E = K$, is the total energy, not to be confused with the electric field E . This holds for electromagnetic waves in general.

13.5.1 Radiation pressure

Because the waves carry energy and momentum, they also exert a *radiation pressure*. This is simply the force per unit area, or the intensity per velocity

$$P = \frac{I}{c} = \frac{B_0^2}{2\mu_0}. \quad (13.27)$$

Example 13.1: To get a feel, suppose you radiate an area of 1 m^2 with a power of 50 W . The radiation pressure will be $P = 1.3 \times 10^{-9} \text{ Pa}$. Remember, atmospheric pressure is about 1030 hPa . Still, a spacecraft traveling to Mars will be displaced thousands of kilometers due to the radiation pressure of the sun, which has to be taken into account in its trajectory. Using mirrors as solar sails and lasers or the sun has been seriously proposed as a method of spacecraft propulsion.

13.6 Creating & detecting electromagnetic waves

13.6.1 Dipoles & antennas

We can accelerate electric charges with an oscillating potential. Suppose you have two conductors with an alternating current between them. An electric field line will connect the two rods. As the field grows, the electric field lines will travel outward. And when the field goes back to zero, the field lines close. As the current keeps oscillating, new waves of electric fields leave the conductors. This is called an oscillating dipole, which emit electromagnetic waves.

Detecting an electromagnetic wave is the opposite of making one. With the same conductors, an incoming wave will induce an alternating current, which can be measured. This is the working principle of an antenna, which can send and receive electromagnetic signals.

13.6.2 EMF of an electromagnetic wave

To detect an electromagnetic wave, we can use a loop of radius r , and make use of the fact that changing magnetic field induces an EMF \mathcal{E} . In particular, the magnetic field B of an electromagnetic wave will induce alternating currents in a loop perpendicular to the magnetic field and direction of its motion. The EMF will be

$$|\mathcal{E}| = \frac{d\Phi_M}{dt} = A \frac{dB}{dt}, \quad (13.28)$$

for an area $A = \pi r^2$. But the electromagnetic wave is oscillating, so the average EMF is

$$\langle \mathcal{E} \rangle = \pi r^2 \left\langle \frac{dB}{dt} \right\rangle. \quad (13.29)$$

If

$$B = B_0 \sin(kz - \omega t), \quad (13.30)$$

then

$$\frac{dB}{dt} = -\omega B_0 \cos(kz - \omega t). \quad (13.31)$$

Therefore, if we forget about the sign and use the fact that $\langle |\cos x| \rangle = 1/\sqrt{2}$ over one wavelength, the average EMF is

$$\langle |\mathcal{E}| \rangle = \frac{\pi r^2 \omega}{2} B_0. \quad (13.32)$$

Chapter 14

Optics

In this chapter, we will study several optical phenomena observed in Nature. Some of them are shown in the overview of Fig. 14.1. We will mostly focus on the geometrical description, and a useful model to help us understand most of these phenomena is Huygens' principle.

The interaction of light and matter is a more complicated subject that falls outside the scope of this course.

14.1 Huygens' principle

A single *point source* of light emits waves in all directions. If all the emitted light is in phase, there will be a growing sphere of a wave moving radially outward from the source. The sphere can be connected where the light waves are at their maximum (or minimum) as in Fig. 14.2a. This is called the *wavefront*. Each wavefront is separated by a wavelength λ from the previous and next wavefront.

A useful concept in optics is *Huygens' principle*, which states that

Huygens' principle. *Each points on a wavefront of light can be regarded as new point source that emits waves (wavelets) in all directions.*

Figure 14.2b illustrates Huygens' principle for a single point source. If you have more than one source, there will be constructive and destructive interference, which will be discussed on more detail below. In the case of many point sources sitting next to each other on one

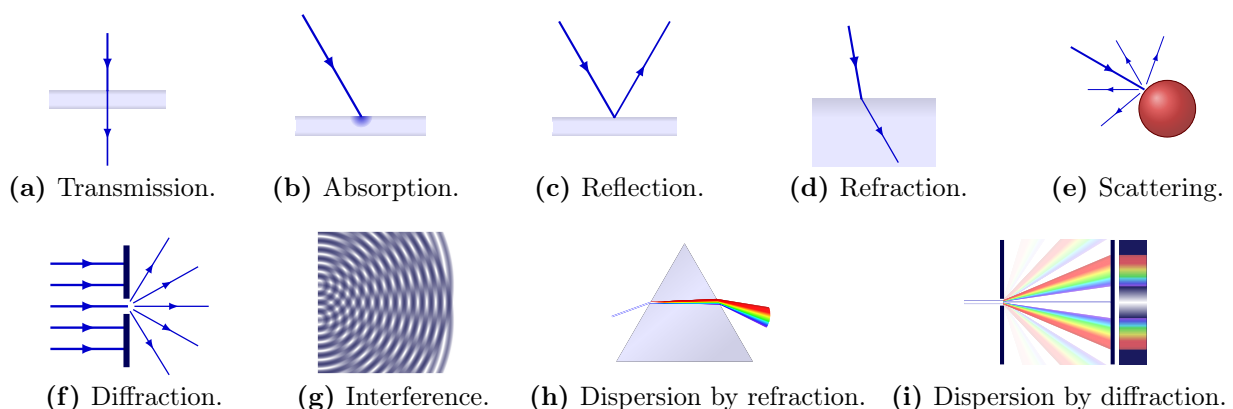
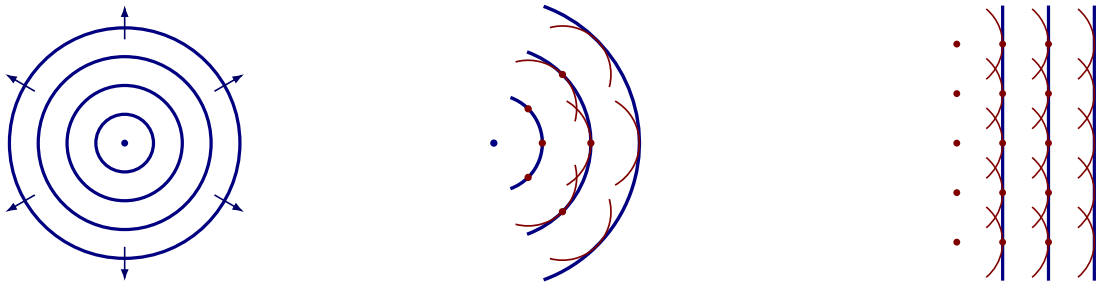


Figure 14.1: Overview of different phenomena of light.



(a) A point source emits spherical waves spreading in all directions. (b) Each point on a wavefront is a new point source. (c) Many point sources on a straight line will emit a plane wave.

Figure 14.2: Huygens' principle: Wavefronts (thick blue lines) are separated by wavelength λ . Each point on a wavefront is a new point source (red) spherically emitting wavelets.

line, and emitting light radially and in phase, the waves will add up to form a single plane wave moving away from the line as a new straight line as in Fig. 14.2c.

An improved version of this model for wave propagation is *Huygens-Fresnel principle*, which explicitly allows the wavelets to have different amplitudes and phases in different directions. Namely, if the wavelets have the same amplitude in all directions, it would imply that the point sources in Fig. 14.2c also create a plane wave in the backward direction.

14.2 Reflection and refraction of light

If light meets a different medium it can be *reflected* and *refracted* (or *bent*) as shown in Fig. 14.3a. *Refraction* (or *bending*) means that light will change directions once entering the new medium. Reflection and refraction are phenomena that occur generally for any type of wave.

Let's look more closely at what happens. If the *incident* light beam has an *angle of incidence* θ_1 with the normal of the *interface* between two media, some fraction of the light gets reflected with the same angle $\theta_r = \theta_1$, and other fraction gets refracted with *angle of refraction* θ_2 . If the refraction index of the second medium is larger, $n_2 > n_1$, then $\theta_2 < \theta_1$, while if $n_2 < n_1$, $\theta_2 > \theta_1$.

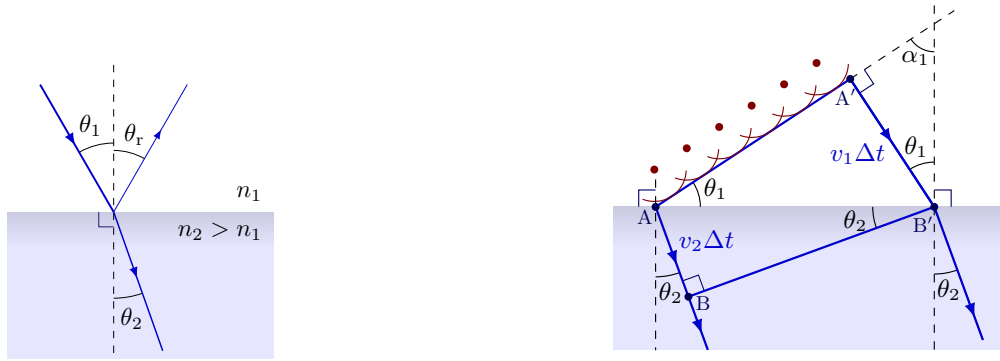
Remember from Section 13.3 that the index of refraction of a medium depends on the speed of light in that medium, v :

$$n = \frac{c}{v} \quad (14.1)$$

Basically, the refraction of light between two media happens because the change of light speed. To understand this better, and find what θ_2 is, it is useful to analyze this problem with Huygens' principle. Imagine the incident light beam has some wavefront like AA' in Fig. 14.3a. Say that one end (A) hits the interface first at $t = 0$. It will take the other end some time Δt to also reach the interface (B') after travelling a distance $v_1 \Delta t$. In that same time, light from end A will have propagated a distance $v_2 \Delta t$ to a new point B. If we assume $n_2 > n_1$, then $v_2 \Delta t < v_1 \Delta t$ because $v_2 < v_1$. The new wave front is BB' , but it will be in a different angle. For a given θ_1 , we can solve the two right triangle, ABA' and ABB' for θ_2 by noticing that

$$|AB| = \frac{v_1 \Delta t}{\sin \theta_1} = \frac{v_2 \Delta t}{\sin \theta_2}. \quad (14.2)$$

Using Eq. (14.1), we therefore find



(a) Incident light gets partially reflected at $\theta_1 = \theta_r$ and refracted at an angle $\theta_1 < \theta_2$ if $n_1 > n_2$. (b) Huygens' principle explains refraction, as points on the wavefront hit the interface at a different time.

Figure 14.3: Refraction and reflection of light.

Law of refraction (Snell's law). *The incident and refracted angles in two media with refractive indices n_1 and n_2 are given by*

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{14.3}$$

It should be noted, however, that while it is useful as a heuristic that gives you the correct answer, there are some problems with using Huygens' principle to explain refraction. For a more correct and rigorous explanation, one needs to invoke Maxwell's laws and decompose the electric and magnetic fields, but this will not be covered here.

For reflection we had already had

Law of reflection.

$$\theta_1 = \theta_r \tag{14.4}$$

If $\theta_1 = 0^\circ$, the reflected intensity is

$$I_r = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 I_0 \tag{14.5}$$

The rest of the light is transmitted (and refracted). For air, $n \sim 1$, while for glass $n \sim 1.5$. Therefore,

$$\frac{I_r}{I_0} = 4\%. \tag{14.6}$$

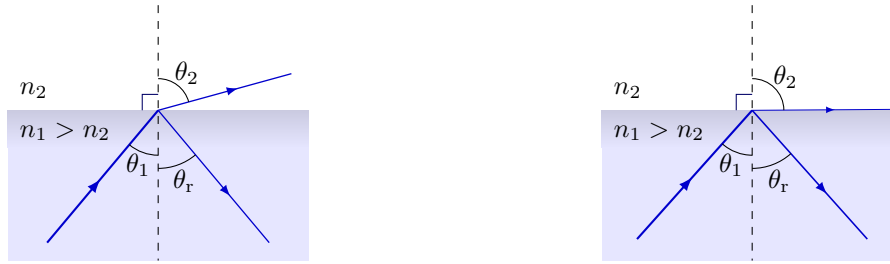
The closer $n_1 - n_2$, the less is reflected.

14.2.1 Total internal reflection

If $n_1 < n_2$, then the angle of refraction is always larger, $\theta_2 > \theta_1$, so at some *critical angle* θ_c , the angle of refraction becomes perpendicular to the interface, $\theta_2 = 90^\circ$, like in Fig. 14.4. At what point does this happen? We simply set $\theta_2 = 90^\circ$ in Snell's law Eq. (14.3):

$$n_1 \sin \theta_c = n_2 \sin \theta_2 = n_2. \tag{14.7}$$

So, as long as $n_2 > n_1$,



(a) Refraction angle θ_2 quickly approaches 90° . (b) When $\theta_1 = \theta_c$, the light is reflected totally.

Figure 14.4: Total internal reflection if $n_1 > n_2$.

Critical angle for total reflection.

$$\sin \theta_c = \frac{n_2}{n_1}. \quad (14.8)$$

This is the volume of the critical angle for θ_1 , where all the light is internally reflected. In fact, it turns out that if then $\theta_1 > \theta_c$, there is no refraction any more at all. The reflection is total.

In a *wave guide* like optical fiber, light always bounces off the inner surface at an incident angle greater than the critical angle (as long as the guide is not bent too much). In this way, a wave guide can efficiently and ultra-fast guide light in a non-straight path and with little loss due to refraction. Optical fiber cables are useful for transfer of digital data (like internet) because they allow transmission of data over longer distances, higher bandwidths (data transfer rates) and with less interference than the usual electrical, metal cables.

14.2.2 Mirage

The index of refraction also depends on the density of a medium. This can lead to some interesting phenomena.

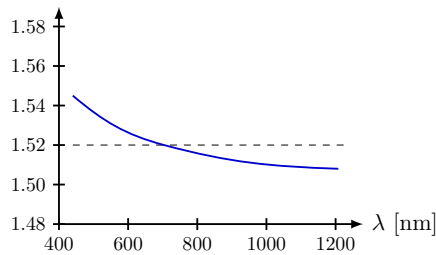
Take for example the atmosphere on a hot day, where the ground is very hot, and the temperature varies with height close to the ground. The density will vary continuously with height from the ground, and so light will be bent differently, depending on the height. Suppose you are looking at a tall object like a tree, which is standing very far from you. Light from top of the tree that would otherwise hit the ground, can be bent upward again to hit your eye. This means you will see an image of the top of the tree appearing on the ground. This effect is called a *mirage*, and is typically seen when the ground is very hot and the horizon is very flat, like hot roads or deserts.

14.2.3 Refraction wavelength

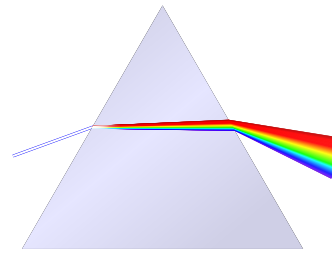
In refraction, the speed changes from $v_1 = c/n_1$ to $v_2 = c/n_2$. However, the frequency of the light stays the same, because the atoms in the medium still absorb the light, vibrate, and emit the light at the same frequency. If the frequencies are $f_1 = f_2$, but $v_1 \neq v_2$, the wavelengths must change:

$$f_1 = \frac{v_1}{\lambda_1} = f_2 = \frac{v_2}{\lambda_2}, \quad (14.9)$$

or



(a) Refraction index vs. wavelength.

(b) Prism like on *Dark Side of the Moon*.**Figure 14.5:** Dispersion: Wavelength-dependence of refraction in glass.

Change in wavelength.

$$\lambda_2 = \lambda_1 \frac{n_1}{n_2}. \quad (14.10)$$

For air to water, $n_2 > n_1$, so the wavelength gets smaller, which means light becomes “more blue” underwater.

14.2.4 Dispersion & prisms

The index of refraction has a slight dependence on the wavelength, like illustrated in Fig. 14.5a. Blue and violet light, with the shortest wavelength in the visible spectrum, is refracted more than red, with the longest wavelength. So a prism like in Fig. 14.5b, which refracts light twice, will spread a white light beam into a rainbow. This wavelength-dependency is called *dispersion*. We will see another example of dispersion in our discussion of diffraction below.

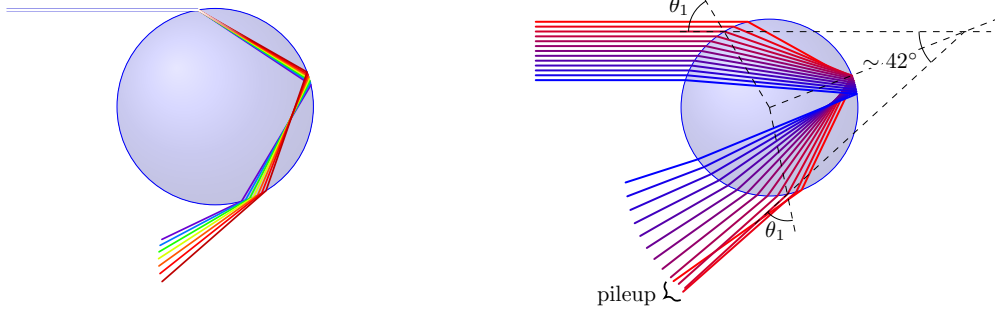
14.2.5 Rainbow

To understand the rainbow, one needs to combine two concepts:

1. Dispersion: Angle of refraction depends on the wavelength.
2. Internal reflection on the back of a water droplet.

Rainbows appear when light is internally reflected inside water droplets. As white light enters a droplet, the lights gets refracted, (partially) reflected from the back of the droplet, and then refracted again in the front of the droplet. Each wavelength was refracted twice at a slightly different angle, creating a spread of colors, just like the prism. To an observer this will appear as a rainbow.

Because the light is internally reflected, the light source (like the sun) should always be opposite of the rainbow image, i.e. to the back of the observer. Furthermore, the light refracting back out the droplet will always make a characteristic angle between 40° (blue) and 42° (red), which will form a continuous circle (as long as there are rain droplets, and so far the view is not blocked by the horizon), with the observer on its central axis. Red will be on the bottom, blue on the top. This is the *primary rainbow*. However, there is also a *secondary rainbow* due to double internal reflection, which will make an angle of around $50\text{--}54^\circ$, this time from blue on the bottom to red on the top.



(a) A white light beam is spread in a rainbow due to dispersion (not to scale).

(b) At an effective reflection angle of about 42° , light is more concentrated (color unrelated to wavelength).

Figure 14.6: Explaining rainbows with dispersion and internal reflection in water droplets.

14.3 Scattering

Air molecules will also absorb light, vibrate and re-emit them in some direction. This is called *scattering*.

Scattering can explain why the sky is blue: Blue light with a shorter wavelength has a much higher probability of scattering off an air molecule than red with long wavelengths. This type of scattering, where the wavelength of light is much larger than the scattering particles is called *Rayleigh scattering*. The amount of scattering is inversely proportional to the fourth power of the wavelength:

Rayleigh scattering.

$$I \propto \frac{1}{\lambda^4}. \quad (14.11)$$

When the sun sets, the light that reaches us has traveled through a larger column of air, and most of the blue has already been scattered away, so the sky will turn more orange and red. Similarly, if the sky is filled with large dust particles from for example volcano eruptions or human pollution, even more scattering will happen. In this case, so-called *Mie scattering* dominates. The sky, or the sunlight reflected off the moon at night, will turn red, even without a sunset. You can keep this in mind next time you have a romantic moment under a red moon.

Scattering also explains why some eyes or veins appear blue. A blue or green iris do not get their color from a blue or green pigment; it even has a lower concentration of the brownish-black melanine. Longer wavelengths of light still tend to be absorbed in the back, while shorter wavelengths are reflected back and undergo Rayleigh scattering in the transparent layer. This is also something you can also explain to your lover while looking into their eyes.

14.4 Polarization

Polarization can happen with any kind of wave. In case of electromagnetic waves, polarization means that the oscillation of the field is happening in on particular direction. As an electromagnetic wave can be a combination of waves of different phases and amplitudes differ, many types of polarization can happen:

1. In *linear polarization*, the fields are pointing in one direction only, e.g. up and down, left and right, like in Fig. 13.1.
2. In *circular polarization*, the direction of the fields rotate around the direction of travel, like in Fig. 14.10.
3. If a light beam is *unpolarized*, the direction of oscillation is randomly pointing in all direction.

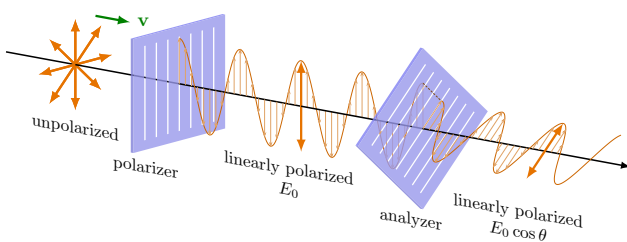
Polarization can be produced by

1. *absorption*,
2. *scattering*,
3. *reflection* and
4. *birefringence*.

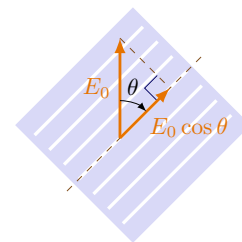
14.4.1 Polarization by absorption

Polarization by absorption happens in natural crystals or polarizing films. Polarizing films like polaroid, typically have long chains of aligned molecules called polymers. If the electric field of the electromagnetic wave is aligned with the direction the chain, it gets absorbed. In the other extreme case, where the electric field of the wave is perpendicular to the length of the chain, the wave is transmitted. For any angle in between, the wave is only partially absorbed. For most polarizers the transmitted wave is linearly polarized in some direction, like in Fig. 14.7a. This direction is called the *transmission axis* or *polarizing axis*. In case of polarizing films, this axis is actually perpendicular to the chains.

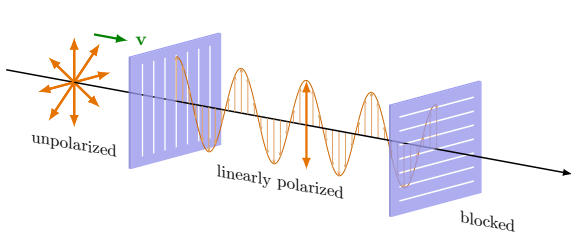
Part of the incoming wave will be absorbed, so the electric field of the outgoing wave will be decreased by some amount. If the wave is initially linearly polarized, and whose



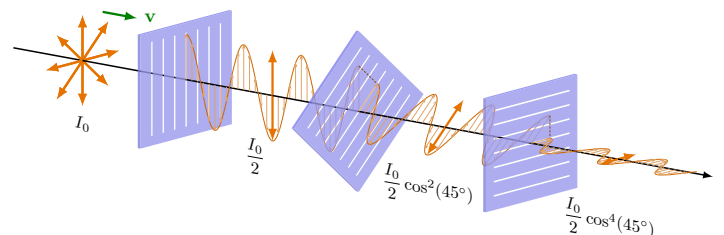
(a) Initially unpolarized light beam gets linearly polarized. The transmitted electric field is reduced to $E_0 \cos \theta$.



(b) Polarizer only lets through the component parallel to its polarizing axis.



(c) If the axes of two subsequent polarizer differ by 90° , all light gets blocked.



(d) Inserting one new polarizer in the middle at an angle in between, allows transmission again, but at reduced intensity.

Figure 14.7: Principles of polarizers.

direction of polarization makes an angle θ with the polarizing axis, the new wave will have an electric field

$$E = E_0 \cos \theta. \quad (14.12)$$

This is the projection of \mathbf{E} onto the polarizing axis Fig. 14.7b. Notice that if \mathbf{E} is perpendicular, $\theta = 90^\circ$, the light is let through without any reduction of its field. If $\theta = 0^\circ$ like in Fig. 14.7c, all light is blocked. Weirdly however, one can insert a new polarizer inbetween to get a new direction as in Fig. 14.7d.

Remember from Section 13.4 that the intensity of electromagnetic waves are proportional to the electric field squared,

$$I \propto E^2 \quad (14.13)$$

So if the incoming beam of electromagnetic waves has an intensity I_0 , the transmitted intensity is given by

Malus's law.

$$I = I_0 \cos^2 \theta. \quad (14.14)$$

If the incident electromagnetic wave is initially unpolarized, it means all directions of polarization are as just likely, so one can integrate Eq. (14.14) over all directions between 0 and 180° :

$$I = I_0 \int_0^\pi \cos^2 \theta \, d\theta = \frac{I_0}{2}. \quad (14.15)$$

So the intensity of a unpolarized light beam is halved due to a linear polarizer.

14.4.2 Polarization by reflection

When light is reflected, it becomes partially polarized, where the direction of polarization is parallel to the interface.

There is again a critical angle called the *polarization angle*, or *Brewster's angle* $\theta_1 = \theta_p$, where the reflected light is completely polarized. It turns out this happens when the reflected and refracted rays make a 90° angle between each other as in Fig. 14.8. The condition is therefore

$$\theta_2 = 90^\circ - \theta_p. \quad (14.16)$$

Snell's law Eq. (14.3) becomes

$$n_1 \sin \theta_p = n_2 \sin(90^\circ - \theta_p) = n_2 \cos \theta_p. \quad (14.17)$$

From this we formulate

Brewster's law.

$$\tan \theta_p = \frac{n_2}{n_1}. \quad (14.18)$$

In the special case the incident light is initially polarized perpendicularly to the interface, there will be no reflected light at all.

Often you will see a blinding glare on surface of large waters or on a wet road when the sun is in front of you. This can be dangerous if you are driving. But a simple solution to this is to wear sun glasses which are polarized (perpendicular to the ground). Similarly, if you want to photograph undeep waters, you can get rid of most reflection on the surface with a polarizer filter to better see the bottom.

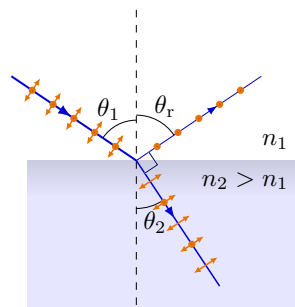
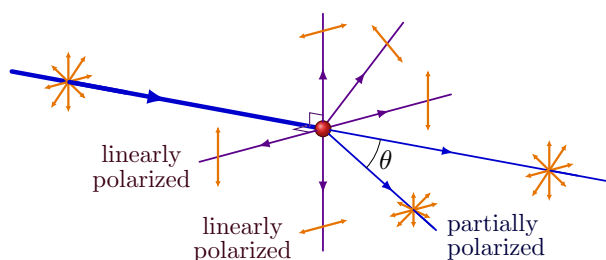
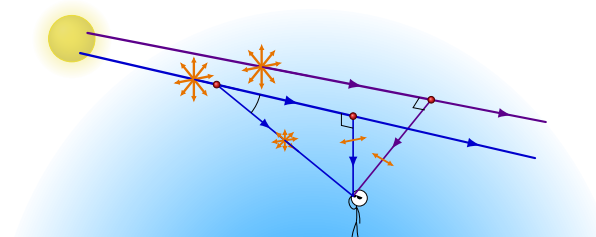


Figure 14.8: At Brewster's angle $\theta_1 = \theta_p$, all reflected light is polarized parallel to the surface.



(a) Polarization by scattering on a molecule.



(b) Scattered sun light in the atmosphere gets polarized.

Figure 14.9: Polarization of light by scattering. Light is maximally polarized when it is scattered perpendicular to the original direction of propagation.

14.4.3 Polarization by scattering

Light that is scattered off molecules will also be fully polarized perpendicular to the original travel of motion, as shown in Fig. 14.9a. If it makes some other angle θ , it will be partially polarized.

This means that sun light scattered by air molecules in the atmosphere, will also be polarized, depending on its path. If you look at the blue sky with a polarizer, it will indeed look like different shades, depending on the angle compared to the sun rays as shown in Fig. 14.9b.

14.4.4 Circular polarization

Consider that you add two electromagnetic waves which are perpendicular to each other and have a phase difference, for example

$$\mathbf{E}_1 = E_1 \sin(kx - \omega t) \hat{\mathbf{x}} \quad (14.19)$$

$$\mathbf{E}_2 = E_2 \sin(kx - \omega t - \phi) \hat{\mathbf{y}}. \quad (14.20)$$

The sum will be

$$\mathbf{E} = E_1 \sin(kx - \omega t) \hat{\mathbf{x}} + E_2 \sin(kx - \omega t - \phi) \hat{\mathbf{y}}. \quad (14.21)$$

If $E_1 = E_2$ and $\phi = \pi/2$, we can rewrite this as

$$\mathbf{E} = E \sin(kx - \omega t) \hat{\mathbf{x}} + E \cos(kx - \omega t) \hat{\mathbf{y}}. \quad (14.22)$$

This looks like a phasor from Chapter 10: For some fixed point x , the electric field will turn around with an angular frequency ω . In other words, we get something that describes a spiral, as we can see in Fig. 14.10.

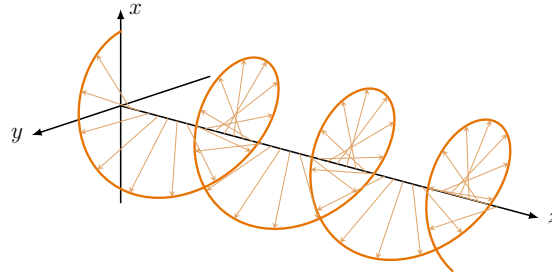


Figure 14.10: Circular polarization. This one is right-handed: Point your right thumb along the direction of propagation. Your fingers should curl along the rotation of the helix.

If $\phi = \pi/2$, the field rotates counter-clockwise. This is called *left circularly polarized*. If on the other hand $\phi = -\pi/2$, it rotates clockwise and is said to be *right circularly polarized*. If $\phi \neq 0$ and $\phi \neq \pi/2$, or $E_1 \neq E_2$, one will get an ellipse instead, and we speak of *elliptical polarization*.

How do you get circular polarization in real life? Some materials have a refractive index that depends on the direction due to its internal structure. Calcite (CaCO_3), for example, has $n_x = 1.658$ in the x and $n_y = 1.486$ in the y -direction. Two waves that initially have the the same phase and are polarized perpendicularly to each other, will have different speeds inside such a medium. They will therefore get a phase difference with respect to each other when they exit the medium again.

14.5 Interference

Last semester we have already seen that waves can be combined according to the *superposition principle*. If two waves of the same wavelength have the same phase, $\Delta\phi = 0$, there will be *constructive interference* because the maxima (*crests* or *peaks*) and minima (*troughs*) align like in Fig. 14.11c. If they, on the other hand, have a phase difference of $\Delta\phi = 90^\circ$, there will be *destructive interference* like in Fig. 14.11a.

14.5.1 Interference pattern

The same happens with light. Imagine you have two point sources that emit coherent light like S_1 and S_2 in Fig. 14.12a. Here, *coherent* means that the light they emit are with same phase and wavelength λ . However, in any point the light coming from S_1 and S_2 will have travelled a different path (Fig. 14.11c), which causes a phase difference. Each of their phases in that point will be depend on the distances r_1 and r_2 of their respective point sources:

$$\phi_1 = 2\pi \frac{r_1}{\lambda} \quad (14.23)$$

$$\phi_2 = 2\pi \frac{r_2}{\lambda}. \quad (14.24)$$

So the phase difference will depend on the path difference $\Delta r = |r_2 - r_1|$:

$$\Delta\phi = 2\pi \frac{\Delta r}{\lambda}. \quad (14.25)$$

We now realize the following.

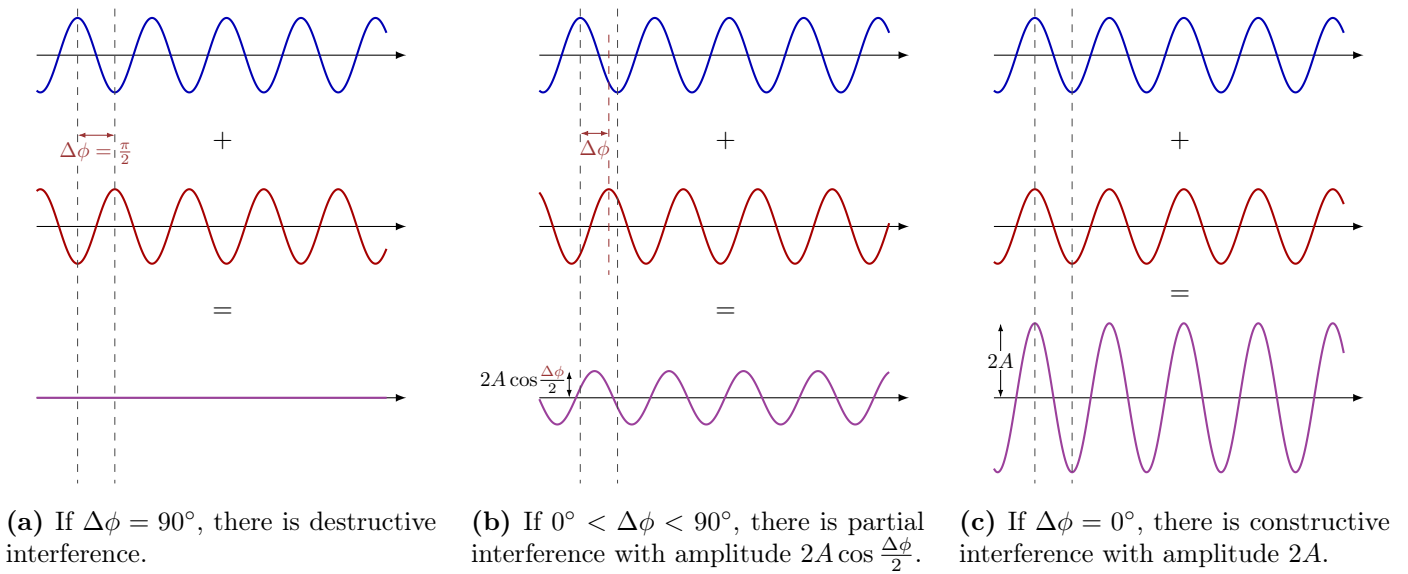


Figure 14.11: Interference between two waves with the same wavelength λ and amplitude A , but phase difference $\Delta\phi$.

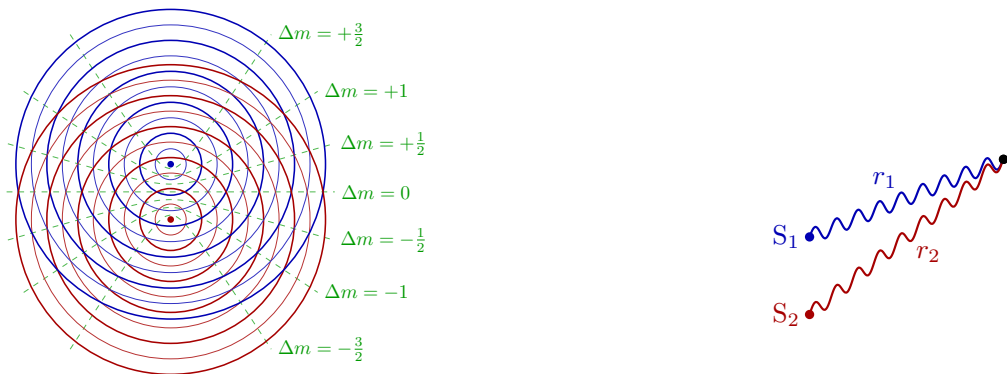


Figure 14.12: Two point sources create an interference pattern. There is destructive interference if the path difference $|r_2 - r_1| = \Delta m\lambda$ where Δm is half-integer, constructive if Δm is an integer.

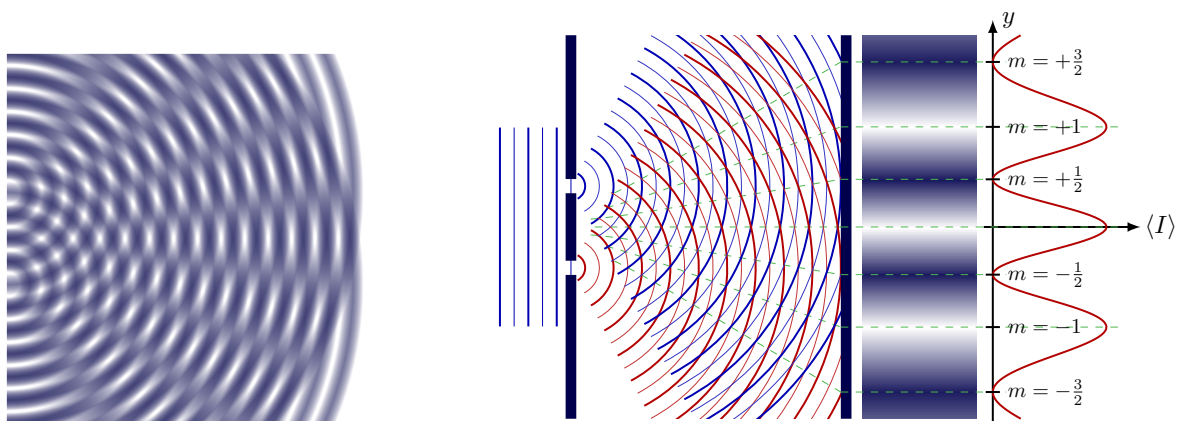


Figure 14.13: Interference pattern of light. There is a dark fringe if the path difference $|r_2 - r_1| = m\lambda$ where m is half-integer, light fringe if m is an integer.

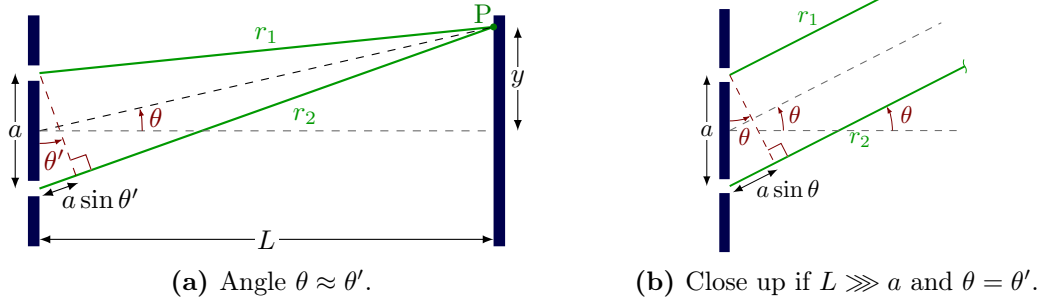


Figure 14.14: Path difference of two slit separated by a depends on the angle θ .

Conditions for destructive and constructive interference. *If the path length differs by an integer number m of wavelengths λ , there will be constructive interference:*

$$\Delta r = m\lambda \quad \text{where } m = 0, \pm 1, \pm 2, \dots \quad (14.26)$$

If the path length differs by a half-integer number $m + \frac{1}{2}$ of wavelengths λ , there will be destructive interference:

$$\Delta r = \left(m + \frac{1}{2}\right) \lambda \quad \text{where } m = 0, \pm 1, \pm 2, \dots \quad (14.27)$$

If we connect all the points where there is constructive or destructive interference, we get the green *nodal lines* in Fig. 14.12a.

What is the intensity of interference? Adding two waves with the same intensity E_0 and angular frequency ω that have some phase difference $\Delta\phi = \phi_2 - \phi_1$ in a point, will add up to

$$E_1 + E_2 = E_0 \sin(\omega t - \phi_1) + E_0 \sin(\omega t - \phi_2) \quad (14.28)$$

$$= 2E_0 \cos \frac{\Delta\phi}{2} \sin \left(\omega t - \frac{\phi_1 + \phi_2}{2} \right). \quad (14.29)$$

See Fig. 14.11. The (average) intensity is this squared and averaged over one period.

Intensity of interference.

$$I(\Delta\phi) = I_0 \cos^2 \frac{\Delta\phi}{2}. \quad (14.30)$$

14.5.2 Double-slit interference pattern

A famous experiment in physics is *Young's experiment* (1801), where light is passed through two narrow slits, caused an interference pattern with light and dark fringes to be project on a screen. This is illustrated in Fig. 14.13b.

To describe where the dark and light fringes will appear, we have to make an approximation that the screen is very far way from the slits, such that $L \gg a$, and we can assume $\theta' = \theta$ in Fig. 14.14a, and so the paths are nearly parallel like in Fig. 14.14b. In that case, the path will be

$$\Delta r = a \sin \theta, \quad (14.31)$$

and the phase difference will be

$$\Delta\phi = 2\pi \frac{a \sin \theta}{\lambda}. \quad (14.32)$$

Let's write it in terms of the position y on the screen. We see that

$$\sin \theta = \frac{y}{R}, \quad (14.33)$$

where $R = \sqrt{L^2 + y^2}$. If the angle θ is very small, we can use the *small-angle approximation* (which is easy to prove with Taylor expansion of $\sin \theta$ and $\tan \theta$) that

$$\sin \theta \approx \theta \approx \tan \theta = \frac{y}{L} \quad (14.34)$$

Equivalently, we see that if $L \gg y$, $R \sim L$. So,

$$\Delta r = \frac{ay}{L}. \quad (14.35)$$

Using our previous conditions (14.26) and (14.26), we find the following conditions.

Conditions for double-slit interference pattern. Assuming $L \gg a$ and $L \gg y$, there will be constructive interference in $y = y_m$, or $\theta = \theta_m$ if

$$\frac{y_m}{L} = \sin \theta_m = \frac{m\lambda}{a} \quad \text{where } m = 0, \pm 1, \pm 2, \dots \quad (14.36)$$

There will be destructive interference if

$$\frac{y_m}{L} = \sin \theta_m = \left(m + \frac{1}{2}\right) \frac{\lambda}{a} \quad \text{where } m = 0, \pm 1, \pm 2, \dots \quad (14.37)$$

So the light fringes will appear on the screen at

$$y_m = \Delta m \frac{L\lambda}{d} \quad (14.38)$$

if Δm is an integer, and dark ones will appear if Δm is a half-integer. The number m is called the *order*. For example, the first maximum in the middle with $m = 0$ is called the *0th order maximum*, the maxima with $m = \pm 2$ are called *second order maxima*.

The intensity will be given by Eq. (14.30):

Intensity of interference of a double-slit.

$$I(\theta) = I_0 \cos^2 \frac{\pi a \sin \theta}{\lambda}. \quad (14.39)$$

Quantum double-slit experiment

As another tangent: The double-slit experiment can be done for all kinds of waves. In particular, important quantum experiments show that massive particles like electrons also behave as waves in certain circumstances. A beam of electrons will also make an interference pattern when meeting two slits. But what is more, proving that quantum mechanics is definitely counter-intuitive and weird: Repeatedly shooting a single electron on two slits will also eventually create an interference pattern over time, as if each electron wave went through both slits at the same time and interfered with itself!

14.6 Diffraction

Diffraction is the bending of light around an obstacle. Several cases are illustrated in Fig. 14.15. For example, for a slit that is much larger than the wavelength, $a \gg \lambda$, the middle of the plane wave is still straight, but the edges are diffracted (Fig. 14.15a). This can be understood with Huygens' principle, where only the point sources in the middle add up to a straight wavefront, but those on the edges spread sideways.

The same happens when a plane wave strikes an obstacle as in Fig. 14.15c. This explains why you can still hear sound coming from around the corner of a large building. This fact also sets a fundamental limit on the resolution of a camera, telescope, or microscope.

If the slit becomes comparable in size to the wavelength, $a \lesssim \lambda$, it will spread as a point source (Fig. 14.15b), as we have already seen with the double-slit experiment, but again an interesting interference pattern emerges. The diffracted light will interfere with itself as in Fig. 14.16a. The derivation of the location of the dark fringes follows the same logic in Section 14.5.2. Consider the top and middle path in Fig. 14.17a. If the screen is sufficiently far away, $L \gg a$, their path lengths will differ by

$$\Delta r = \frac{a}{2} \sin \theta. \quad (14.40)$$

Destructive interference in a point P should happen when

$$\frac{a}{2} \sin \theta = \frac{m}{2} \lambda \quad (14.41)$$

for $m = \pm 1, \pm 2, \dots$. And this is true for the other rays in the slit that P. In a nice box:

Conditions for a single-slit diffraction pattern. Assuming $L \gg a$, there will be destructive interference if

$$\sin \theta_m = \frac{y_m}{L} = m \frac{\lambda}{a} \quad \text{where } m = 0, \pm 1, \pm 2, \dots \quad (14.42)$$

Here we again used the small-angle approximation $\sin \theta \approx y/L$.

The full diffraction pattern at a long distance is best described by

Fraunhofer diffraction equation.

$$I(\theta) = I_0 \left(\frac{\sin \left(\frac{a\pi}{\lambda} \sin \theta \right)}{\frac{a\pi}{\lambda} \sin \theta} \right)^2. \quad (14.43)$$

For close-distance diffraction, the *Fresnel diffraction equation* gives a better description, but this falls out of the scope of these lectures notes.

Because the diffraction pattern depends on the wavelength λ , diffraction will also cause dispersion, as illustrated in Fig. 14.16b.

One thing that was not mentioned in Section 14.5.2, is that each of the two slits also causes diffraction, so the actual double-slit interference pattern will have two simultaneous effects, as shown in Fig. 14.18.

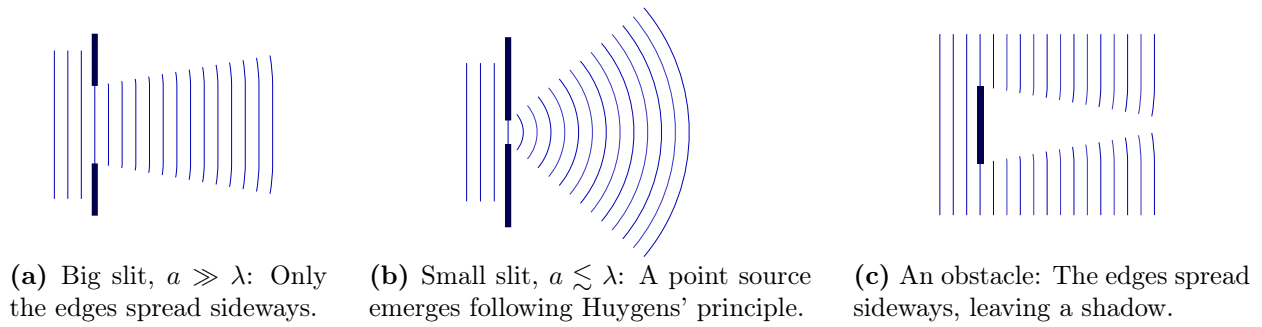


Figure 14.15: Diffraction around slits and obstacles.

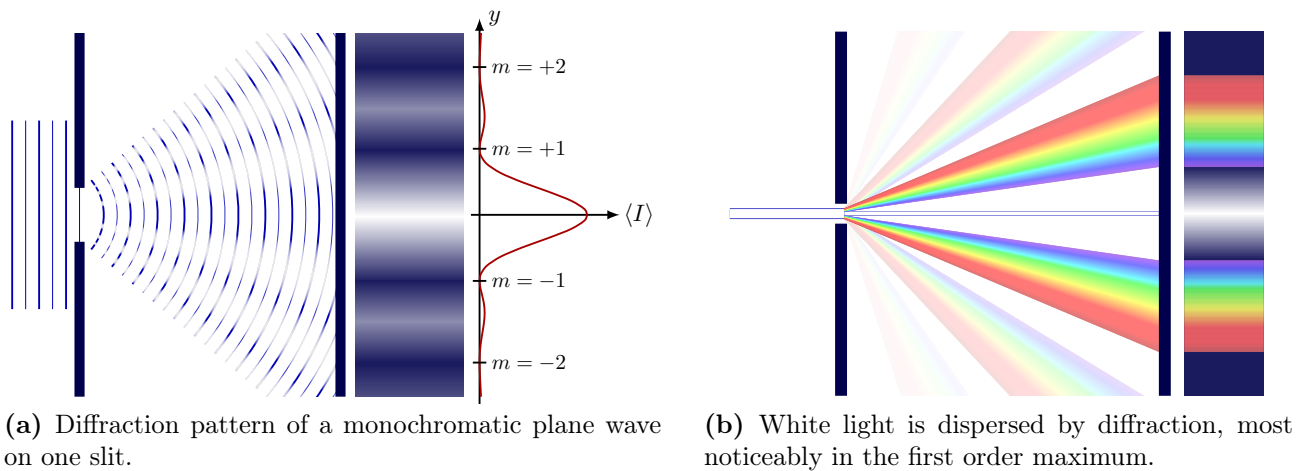


Figure 14.16: Path difference of one slit of size a depends on the angle θ . The interference pattern Eq. (14.43) in a given direction θ will depend only on a and the wavelength λ .

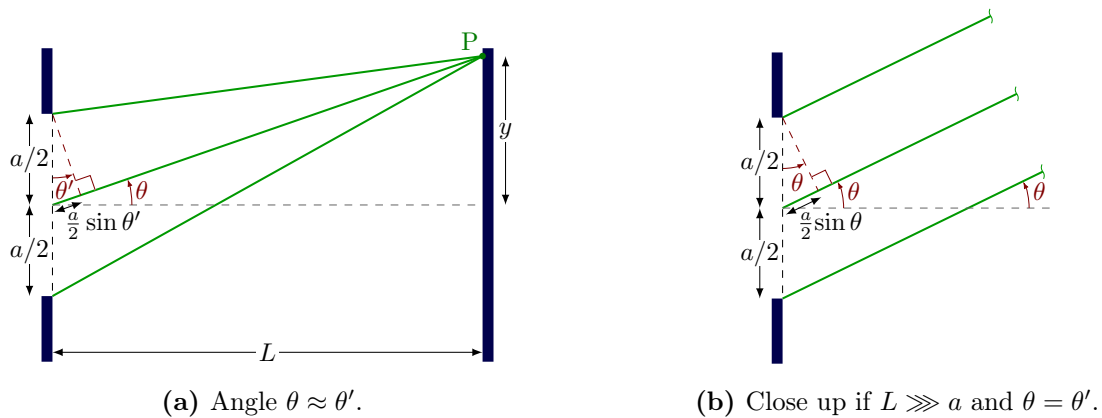


Figure 14.17: Path difference $\Delta r = a \sin(\theta)/2$ of one slit of size a depends on the angle θ .

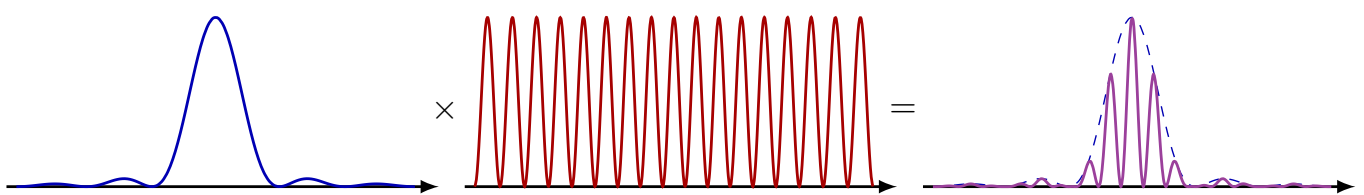


Figure 14.18: The actual interference pattern of two slit (purple) also includes a diffraction effect (blue, Eq. (14.43)), which acts as an envelope on the two-slit interference (red, Eq. (14.39)).

14.6.1 Application: Crystallography

Diffraction has several applications. One that is noteworthy, is *X-ray diffraction crystallography*. Because X-rays have a wavelength that is typically of the same order of magnitude as the typical spacing between atoms in a crystal, they are useful to probe the structure of crystal. As X rays enter the regular arrangement of atoms of a crystal, it will be diffracted many times, and Fourier analysis of the intricate interference pattern will allow to reverse-engineer the geometric structure of the crystal. However, neutron or electron also behave as waves at the quantum level, and can therefore be used for diffraction crystallography.

It can be used to study chemical bonds and structure of crystals from graphite and diamonds to salt and sugar. However, it is also possible to use with less regular molecules like proteins. In fact, Francis Crick and James Watson in 1953 managed to deduce from X-ray diffraction studies, made by Maurice Wilkins and Rosalind Franklin, the molecular structure of DNA, which is a double helix of nucleic acid polymers held together by nucleotides.

14.7 Iridescence

Iridescence happens when the apparent color of something changes depending on the angle you are looking at it. The word comes from the Greek word for “rainbow”, *iris*. This typically happens when light is reflected several times by a thin layer, like in a thin film of soap in a soap bubble, and causes interference. But it can also happen by diffraction off the grooves of a CD, or by the micro structures of feathers and butterfly wings. While the latter do not contain any blue or green pigment, they can appear those colors by *structural coloration*.

Part II

Thermodynamics

Chapter 15

Temperature & The Zeroth Law

Thermodynamics is the study of temperature, heat and the exchange of energy. *Temperature* has the symbol T and we typically use the units of $^{\circ}\text{C}$ or K . This is not to be confused with the concept of *heat*, denoted as Q , which is the energy transfer caused by a *temperature difference* ΔT between two systems, and therefore carries the units of energy, Joules.

At the *macroscopic level*, the temperature of a gas in a container is related to the volume of the container, V , and its pressure, P . For a lot of gases, this relation can be approximated by the *ideal gas law*,

$$PV = nRT, \quad (15.1)$$

which will be discussed in the next chapter. At a *microscopic scale*, you can think of a gas as an *ensemble* of N particles, each with some mass m and velocity v . They therefore have a kinetic energy of

$$K = \frac{1}{2}mv^2. \quad (15.2)$$

These different representations of gas are shown in Figs. 15.1a and 15.1b. Because the particles can move in the three spatial dimensions, they have three *degrees of freedom* (d.o.f.). But most particles are not just point particles, or perfectly spherical symmetric. Often they are molecules with some atoms bound together. In this case, the molecule can also make rotations, and there there are additional degrees of freedoms. In the example of a diatomic gas molecule, like H_2 or CO , the molecule is cylindrical symmetric, and can rotate in two independent directions, as is shown in Fig. 15.1c. These molecules therefore have two *rotational degrees of freedom*, in addition to their three *translational degrees of freedom*; five in total.

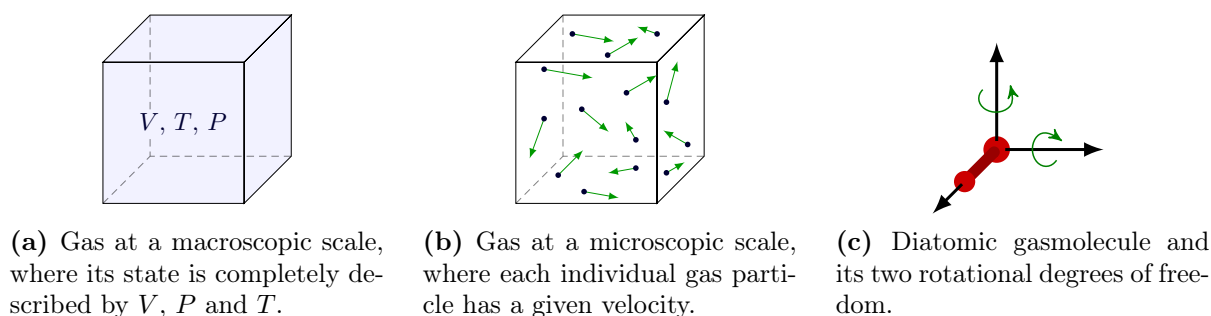


Figure 15.1: Different descriptions of gas.

But what is temperature? Temperature measures “hotness” or “coldness”. This definition is a bit vague... We measure temperature by studying the changes in the properties of matter. As temperature increases, materials can expand, for examples a gas in a balloon, liquids and even solids. Metals conduct electricity worse at higher temperatures, as the thermal velocity competes against the drift velocity, as we compared previously. For ohmic metals, the current is given by

$$I = \frac{V}{R}, \quad (15.3)$$

where V is the voltage, and the resistance $R = R(T)$ is temperature dependent.

15.1 Thermometers and temperature scales

A *thermometer* measures its own temperature. However, once two bodies interact, they reach a thermal equilibrium, and there is no further change in the system.

Consider a system with three bodies, A, B and C, that are not in direct contact. The bodies have different temperatures at first. Now an insulator is placed between A and B so no heat can flow, but two heat conductors are placed between A and C, as well as B and C, like in Fig. 15.2a. Now A-C, B-C are in *thermal contact*, and because they have different temperatures, heat will transfer through the conductor until A and B become in *thermal equilibrium* with C. However, is A in thermal equilibrium with B? To test this, we swap the insulator with the conductors, such that A and B are in thermal contact with each other, but not with C, as is shown in Fig. 15.2b. Experimentally it has been proven that A and B are always in thermal equilibrium. This fact is encapsulated in the zeroth law of thermodynamics:

0th law of thermodynamics. *If C is initially in thermal equilibrium with both A and B, then A and B are in thermal equilibrium with each other.*

Historically, this law was added after the first three laws of thermodynamics, but people thought it so important, they called it the zeroth law. This is because this principle is needed to define what temperature is in the first place.

Now suppose that C is a thermometer. It is in stable equilibrium with A and B, so its temperature must be stable. Therefore, we are able to define temperature:

Two systems are in thermal equilibrium if and only if they have the same temperature.



(a) A and B are each in thermal contact with C through conductors, but isolated from each other by insulator I. (b) A and B are each in thermal contact, but both isolated from C by I.

Figure 15.2: Illustration of the 0th law of thermodynamics with three objects A, B and C.

We can define a temperature scale by choosing two fixed reference points. The metric system uses the *Celsius scale*, 0°C is defined as the point where water freezes, while 100°C is based on the point where water starts to boil, both at standard atmospheric pressure.

The imperial system uses the *Fahrenheit scale*, whose 0°F is historically based on the freezing temperature of a solution of water, ice and ammonium chloride, and the human body temperature for 100°F (about 36°C). This scale has been redefined since, such that water freezes at 32°F , and water boils at 212°F at standard atmospheric pressure. This is why the average body temperature is now 98°F . A quick conversion can be approximated with

$$T_{\text{F}} = \frac{\Delta T_{\text{C}}}{\Delta T_{\text{F}}} T_{\text{C}} + 32^\circ\text{C}, \quad \text{using} \quad \frac{\Delta T_{\text{C}}}{\Delta T_{\text{F}}} \sim \frac{100^\circ\text{C}}{180^\circ\text{F}}. \quad (15.4)$$

The *Kelvin scale* is defined with the absolute zero temperature -273.15°C , or 0 K , and the triple point of water 0.01°C , or 273.16 K , at a pressure of $P_{\text{triple}} = 610\text{ Pa}$. Absolute zero is the point where all molecular motion comes to a halt. The triple point of water is where all three phases (ice, liquid water and vapor) can exist. To produce water in its triple point, add water, ice, and vapor to an evacuated container with no air. Equilibrium will be reached at 0.01°C , at a pressure of 610 Pa . The Kelvin scale uses the same size of unit as Celsius: $\Delta T_{\text{K}} = \Delta T_{\text{C}}$. This means:

$$T_{\text{K}} = \frac{\text{K}}{^\circ\text{C}} T_{\text{C}} + 273.15\text{ K} \quad (15.5)$$

At 0 K , the pressure of all gases becomes 0 Pa , and in this scale the pressures P_1 and P_2 of an ideal gas at two different temperatures T_1 and T_2 , respectively, are related by

$$\frac{T_2}{T_1} = \frac{P_2}{P_1}, \quad \text{or} \quad T_2 = T_1 \left(\frac{P_2}{P_1} \right). \quad (15.6)$$

For example, the Kelvin scale for water vapor with pressure P is given by

$$T_{\text{K}} = 273.15 \frac{P}{P_{\text{triple}}}, \quad (15.7)$$

Because the pressure and temperature are proportional to each other, the Kelvin scale is convenient for writing down the ideal gas law.

Often *room temperature* is used as an easy reference temperature. This can be anywhere between 20 (293.15) and 25°C (298.15 K), or even 300°K (26.85°C) to have a nice round number.

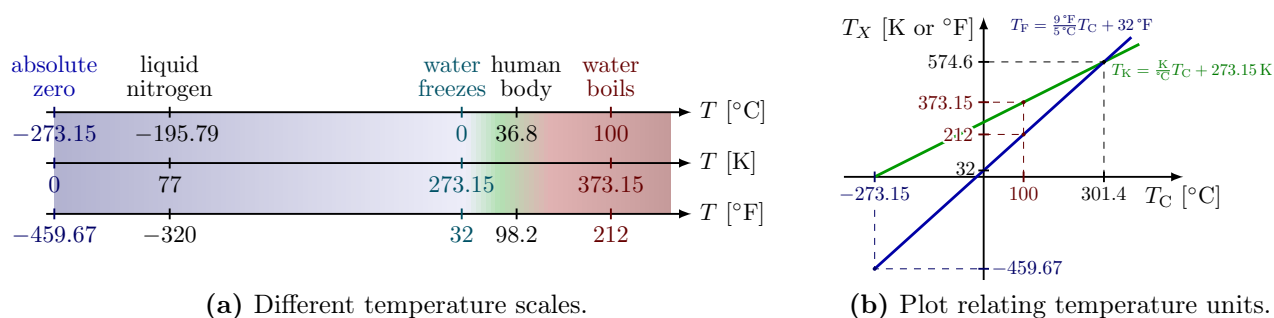


Figure 15.3: Comparing temperature units of Celsius, Kelvin and Fahrenheit.

Chapter 16

Ideal Gas Law

The *ideal gas law* is so important that it deserves its own chapter. It is given by

Ideal gas law.

$$PV = nRT, \quad (16.1)$$

with the gas pressure P (N/m²), volume V (m³ or L), the amount of moles n , $R = 0.8206$ L atm/K, and the temperature T in Kelvin. It can also be written as

$$PV = Nk_{\text{B}}T, \quad (16.2)$$

with the number of gas molecules, N and the Boltzmann constant k_{B} . The number of molecules is given by

$$N = nN_{\text{A}}, \quad (16.3)$$

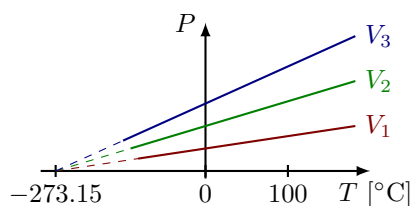
with Avogadro's number $N_{\text{A}} = 6.02 \times 10^{23}$ molecules per mole.

The ideal gas law is a good first-order approximation of the behavior of most gasses under many conditions. It is a simple equation that is a combination of several empirical laws discovered in experiments with gas. For example, the *Boyle–Mariotte law* (1662) states that

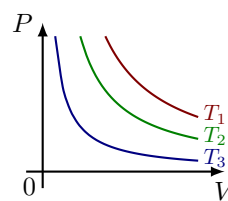
Boyle–Mariotte law. *The pressure P of an ideal gas with fixed mass and temperature T is inversely proportional to the volume V ,*

$$P \propto \frac{1}{V}, \quad (16.4)$$

or equivalently, PV is a constant.



(a) Isochore: Fixed volumes $V_1 > V_2 > V_3$.



(b) Isotherm: Fixed temperatures $T_1 > T_2 > T_3$.

Figure 16.1: Pressure P according to the ideal gas law $P = nRT/V$.

In the next chapter, we will derive the ideal gas law from of some idealized assumptions of gas. One important assumption is that the gas exist out of small solid spheres that do not interact with each other, except for elastic collisions (meaning total kinetic energy and linear momentum are conserved). Small modifications can be made to improve upon this ideal model. In Section 17.5, we will see that the *van der Waals equation* takes into account intermolecular forces between the gas particles, and their finite size.

Experiment 16.1 (Hot air balloon): Why does a balloon filled with hot air fly? The balloon is opened from below, and so the pressure P inside equalizes with the atmospheric pressure outside, P_0 . Instead of expanding further and stretching the balloon, some air escapes, and the hot air inside becomes less dense. This is easy to see: One can write the total mass as $M = nm$ and the density as $\rho = M/V$. Plugging this in the ideal gas law, (16.1), one finds

$$\rho = \frac{mP}{RT}, \quad (16.5)$$

which is proportional to the inverse of the temperature. So as the air inside heats up, the air gets less dense, and denser air exerts a buoyancy force on the the balloon.

Example 16.1: What volume is occupied by 1 mole of gas at 0°C and 1 atm of pressure?

$$V = \frac{nRT}{P} = \frac{(1 \text{ mol}) (0.0821 \frac{\text{L} \cdot \text{atm}}{\text{mol} \cdot \text{K}}) (273 \text{ K})}{1 \text{ atm}},$$

or $V = 22.4 \text{ L}$. If we double the temperature by heating, and we reduce the volume to half, what is the new pressure? Assume n stays the same, then

$$\frac{P_1 V_1}{T_1} = \frac{P_2 V_2}{T_2} = \frac{P_2 (\frac{1}{2} V_1)}{(2T_1)},$$

and thus $P_2 = 4P_1$.

Chapter 17

Statistical Mechanics

Statistical mechanics is an important field in modern physics which aims to understand macroscopic behavior from the statistical behavior of microscopic things. Because gas is actually a large ensemble of microscopic behavior, we can actually derive the ideal gas law from first principles, only using Newton's law and averaging. Because there are too many degrees of freedom, i.e. too many particles, to keep track of, we need tools from statistics, like probability density functions. It turns out that a lot of thermodynamic concepts and laws emerge from statistical behavior of microscopic processes.

17.1 Equipartition theorem

Consider a gas in a volume, and each gas particle bouncing off the walls of the container without interacting with other particles. A gas particle has a well-defined velocity vector

$$\mathbf{v} = v_x \hat{x} + v_y \hat{y} + v_z \hat{z}. \quad (17.1)$$

When a particle bounces off a wall in the x direction,

$$\Delta p_x = mv_{fx} - mv_{ix} \quad (17.2)$$

We can safely assume an elastic collision because the gas particle's mass is very small compared to the wall, so $v_{fx} = v_{ix} \equiv v_x$, and

$$\Delta p_x = 2mv_x. \quad (17.3)$$

In some time Δt , molecules hit the wall, with area A . The number of molecules hitting the wall should be within a distance $d = v_x \Delta t$, and moving to the right. We can construct an expression for this number as

$$\underbrace{\frac{N}{V}}_{\text{density of molecules}} \underbrace{v_x \Delta t A}_{\text{volume of molecules with } v_x} \underbrace{\frac{1}{2}}_{\text{moving to the right}}. \quad (17.4)$$

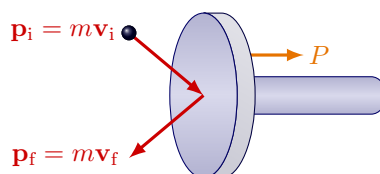


Figure 17.1: Elastic collision of a gas particle off the piston: Both momentum and kinetic energy are conserved.

So the total momentum is

$$\Delta p_x = \left(\frac{1}{2} \frac{N}{V} v_x \Delta t A \right) (2mv_x), \quad (17.5)$$

or,

$$\Delta p_x = \frac{N}{V} m v_x^2 A \Delta t. \quad (17.6)$$

Then, the total force on the wall is

$$F = \frac{\Delta p_x}{\Delta t} = \frac{N}{V} m v_x^2 A, \quad (17.7)$$

and the pressure

$$P = \frac{F}{A} = \frac{N}{V} m v_x^2, \quad (17.8)$$

or,

$$PV = N m v_x^2. \quad (17.9)$$

We can rewrite this in terms of kinetic energy of one molecule in the x direction:

$$PV = 2N \left(\frac{m v_x^2}{2} \right). \quad (17.10)$$

This expression looks suspiciously familiar. If we compare it to the ideal gas law Eq. (16.1), we can find a deeper connection between the temperature of a gas and the kinetic energy of an individual molecule:

$$K = \frac{m v_x^2}{2} = \frac{k_B T}{2}. \quad (17.11)$$

But we made a big assumption: Not all particles have the same velocities. Rather, their velocities are distributed, and the important quantity is the *average velocity* $\langle v_x \rangle$, and the *average kinetic energy*

$$\frac{m \langle v_x \rangle^2}{2} = \frac{k_B T}{2}. \quad (17.12)$$

In general, we should consider all three directions:

$$K = \frac{m \langle v_x \rangle^2}{2} + \frac{m \langle v_y \rangle^2}{2} + \frac{m \langle v_z \rangle^2}{2} \quad (17.13)$$

$$= \frac{3}{2} m \langle v_x \rangle^2. \quad (17.14)$$

The direction does not matter, so $\langle v_x \rangle = \langle v_y \rangle = \langle v_z \rangle$. Or in terms of temperature,

$$K = \frac{3}{2} k_B T. \quad (17.15)$$

For N molecules, then, the total kinetic energy of the gas is

$$K = \frac{3}{2} N k_B T \quad (17.16)$$

$$= \frac{3}{2} n R T \quad (17.17)$$

This leads us to:

Equipartition theorem. *There is an energy of $\frac{1}{2} k_B T$ per molecule (or $\frac{1}{2} R T$ per mole) associated with each degree or freedom.*

17.2 Interlude: Random walks

We learned that the temperature of a material is related to the movement of individual molecules. For an ideal gas, there are three degrees of freedom

$$K = \frac{1}{2} \langle v^2 \rangle = \frac{3}{2} k_B T. \quad (17.18)$$

This process has a randomness to it.

Example 17.1 (Coin flip): If we have a coin, and we flip it three times, what are the chances of getting two heads (H), and one tails (T) in a row? There are nine possible outcomes:

HHH, THH, HTH, HHT, TTH, THT, HTT, TTT,

three of which have two heads. Both outcomes of a single coin flip have a probability

$$P(H) = \frac{1}{2} = P(T).$$

Therefore,

$$P(2H) = \frac{3}{8}.$$

Another example is the *random walk*, sometimes also called the *drunken man's walk*. Imagine you have a particle that after each time interval Δt makes a decision to go left (L) or right (R) like the ball the Galton board shown in Fig. 17.2. Assume each direction has an equal probability:

$$P(L) = \frac{1}{2} = P(R).$$

What if we have $n = 10$ rows of rods? We can use the *binomial distribution*. The probability of a ball getting in bin k is after n steps is

Binomial distribution.

$$P(n, k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (17.19)$$

For example, the probability a ball gets in the third bin is

$$P(10, 3) = \binom{10}{3} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{10-3}.$$

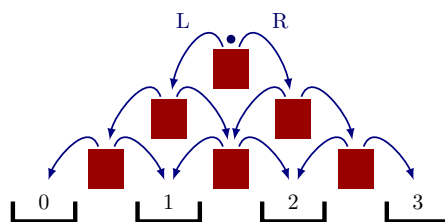


Figure 17.2: Random walk of a ball in a Galton Board with four bins.

17.3 Interlude: Gaussian distributions

What about the probability of a ball getting in any of the first 10 left bins, for $n = 50$ bins. For large n , we can approximate the binomial distribution by the *normal* or *Gaussian distribution*:

Normal distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (17.20)$$

with the mean μ , or *expected value of x* also written as

$$\mu = \text{E}[x] = \langle x \rangle, \quad (17.21)$$

and the standard deviation σ , related to the variance as

$$\text{Var}[X] = \sigma^2. \quad (17.22)$$

The gaussian $f(x)$ is shown in Fig. 17.3. It is a *probability density function* (PDF), which means it is normalized to unity:

$$1 = \int_{-\infty}^{+\infty} f(x) dx, \quad (17.23)$$

and the probability that $x \leq a$ is given by

Cumulative distribution function.

$$\text{P}(X \leq a) = \int_{-\infty}^a f(x) dx. \quad (17.24)$$

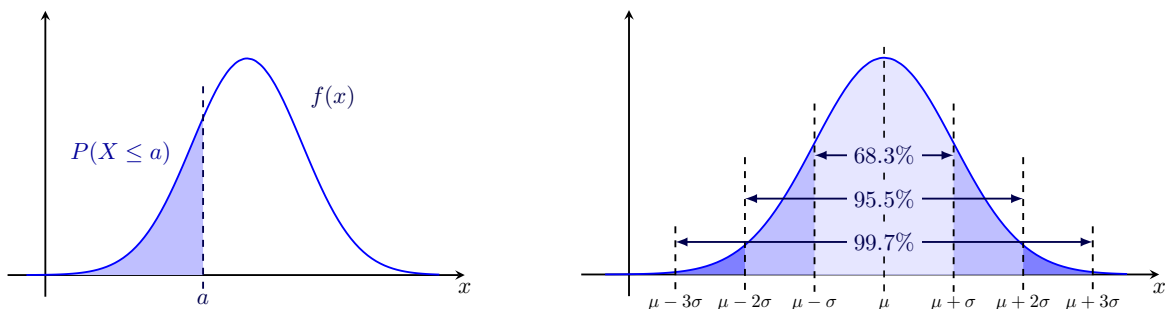
The average is the integral

$$\langle x \rangle = \int_{-\infty}^{+\infty} x f(x) dx. \quad (17.25)$$

The probability $\text{P}(X \leq a)$ corresponds to the colored area in Fig. 17.3b. Some common values that are used for *confidence intervals* are given by Table 17.1, where for example

$$\text{P}(-\sigma \leq X \leq \sigma) = \int_{-\sigma}^{\sigma} f(x) dx = 68.3\% \quad (17.26)$$

This is illustrated with the differently colored areas in Fig. 17.3b.



(a) The colored area under the distribution represents the probability $\text{P}(X \leq a)$.

(b) 65-95-99 rule for the interval with standard deviation σ from the mean μ .

Figure 17.3: Gaussian distributions and their integrals.

Table 17.1: 65-95-99 rule.

Standard deviations	Probability [%]
1σ	68.3
2σ	95.5
3σ	99.7
4σ	99.994
5σ	99.99994

To approximate a binomial distribution, one can make the correspondence

$$\mu = np \quad (17.27)$$

$$\sigma = \sqrt{npq}, \quad (17.28)$$

where $q = (1 - p)$. So for $n = 50$, $\mu = 25$, $\sigma = 3.6$. We can express this in numbers of standard deviations (σ 's) with the z -score for $x = 10$ is

$$z = \frac{x - \mu}{\sigma} \quad (17.29)$$

$$\sim -4\sigma.$$

Therefore,

$$\begin{aligned} P(x \leq 10) &\approx P(z \leq -4\sigma) \\ &= \frac{1}{2}(1 - P(z \leq 4\sigma)) \\ &= 0.00003. \end{aligned}$$

17.4 Maxwell-Boltzmann distribution

The velocities of particles in a gas follow the *Maxwell-Boltzmann distribution*.

Maxwell-Boltzmann distribution. *The number of gas particles, dN , with velocity between v and $v + dv$ is given by*

$$dN = Nf(v) dv, \quad (17.30)$$

with the probability density function

$$f(v) = \frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_{\text{B}}T} \right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_{\text{B}}T}}, \quad (17.31)$$

where T is the temperature T of the gas, and m is the mass of the gas particle.

We can easily rewrite $f(v)$ in terms of energy, using $E = \frac{1}{2}mv^2$:

$$f'(E) = \frac{2}{\sqrt{\pi}} \left(\frac{1}{2k_{\text{B}}T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_{\text{B}}T}}, \quad (17.32)$$

such that there are

$$dN = Nf'(E) dE, \quad (17.33)$$

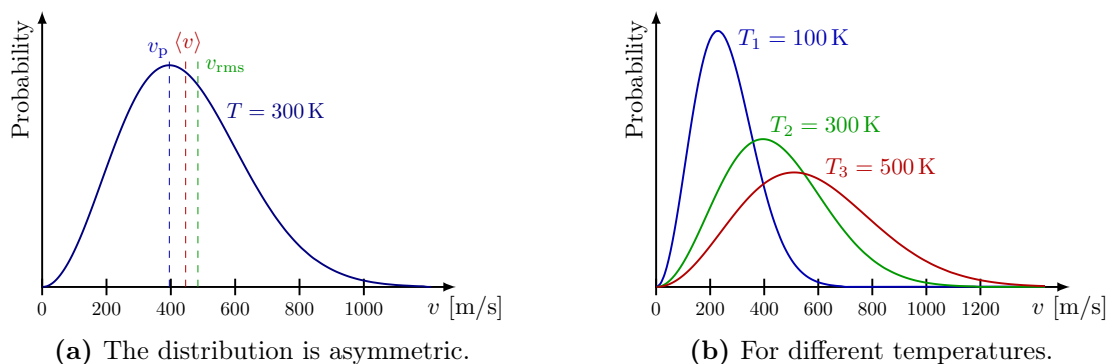


Figure 17.4: Maxwell-Boltzmann distribution for oxygen gas O_2 with atomic weight ~ 16 per atom.

particles with energies between E and $E + dE$.

As we saw above, the average of the velocity squared of molecules in a gas is given by

$$\frac{m \langle v^2 \rangle}{2} = \frac{3}{2} k_B T. \quad (17.34)$$

In terms of the *root mean square*

$$v_{\text{rms}} = \sqrt{\langle v^2 \rangle} = \sqrt{\frac{3k_B T}{m}}, \quad (17.35)$$

The velocity v_p for which f is maximum can be found via derivation:

$$v_p = \sqrt{2k_B T m}. \quad (17.36)$$

This is the most probable velocity a randomly sampled gas particle can have. The average $\langle v \rangle$ can be found with the integral Eq. (17.25), and is

$$\langle v \rangle = \frac{2}{\pi} v_p. \quad (17.37)$$

Clearly, $v_p < \langle v \rangle < v_{\text{rms}}$. This is because the Maxwell-Boltzmann distribution is asymmetric, as shown in Fig. 17.4.

The distribution is similar for liquids, but liquids can evaporate (without boiling). If the velocity of a particle is above some escape velocity, it can overcome gravity and leave the liquid. Because the liquid is losing particles with higher velocity, and thus higher energy, the liquid as a whole cools off.

17.5 Van der Waals equation

The ideal gas law is an approximation that assumes molecules are point particles (i.e. no size), and there are no forces between molecules.

However, there is a more accurate extension of the ideal gas law, that better describes real gasses, by assuming there are attractive forces between particles, and the particles have some finite size. It adds a correction factor to the pressure to take into account molecular attractions, and a correction factor to the volume to take into account the volume of molecules:

Van der Waals equation.

$$\left(P + \frac{an^2}{V^2}\right)(V - bn) = nRT. \quad (17.38)$$

Here, the effective volume V is smaller by an amount nb , where n is the amount of moles, and b is the volume of one moles of the gas particle in question. The constant a , on the other hand, depends on the attractive forces between gas molecules. The gas particles can now attract each other, and as when they are in the center of the bulk of the gas particles, they will experience no net force, but if they are at close to the wall of the container and away from the center, the bulk will exert a net force. This cause them to hit the wall less frequently and with less momentum. This net force is proportional to the density n/V of the gas as a whole, and the number of particles close to the walls will also be proportional to the density n/V . Therefore, the measured pressure P is decrease by $a(n/V)^2$:

$$P = \frac{nRT}{V - bn} - \frac{an^2}{V^2}, \quad (17.39)$$

where a acts as a constant of proportionality that is a measure of the intermolecular attraction. Some exemplary values for typical gasses are given in Table 17.2.

Table 17.2: Van der Waals constants for some typical gasses, and their corresponding inversion temperature (see Section 21.3.1).

Gas	a [Nm ⁴ /mol]	b [Nm ⁴ /mol]	T_i [K]
Helium	0.0033	24	30
N ₂	0.136	38.5	620
N ₂	0.137	31.6	765
Air	–	–	650
CO ₂	0.365	42.5	> 1000

Chapter 18

Heat & The First Law

Heat is a form of energy that is transferred to a thermodynamic system, other than work. When we add heat Q to something, usually the temperature increases by some ΔT according to

Specific heat.

$$Q = mc\Delta T, \quad (18.1)$$

where m is the mass and c is the *specific heat* of the object being heated. The larger c , the more energy an object can absorb per unit of mass for a given temperature increase ΔT . Sometimes it's easier to use the *heat capacity* $C = mc$ of a given mass m instead: s

$$Q = C\Delta T. \quad (18.2)$$

To avoid confusion, make sure your units match. Some actual values of common materials are listed in Table 18.1.

Table 18.1: Specific heat c of some common materials.

Substance	c [kJ/kg K]
Water	4.18
Wood	2–3
Aluminium	0.900
Copper	0.386

Example 18.1: What is the ΔT for 1 kg of aluminium or water if it absorbs 1 kJ? Water has a specific heat $c = 4.18$ kJ/kg K, and therefore, its temperature increases by $\Delta T = 0.24^\circ\text{C}$. For aluminium with an heat capacity of $c = 0.900$ kJ/kg K, $\Delta T = 1.1^\circ\text{C}$.

Example 18.2: What is the difference of temperature ΔT for $m = 1$ kg of water or aluminium if it absorbs $Q = 1$ kJ of heat? For water,

$$\Delta T = \frac{1 \text{ kJ}}{(1 \text{ kg}) \left(4.18 \frac{\text{kJ}}{\text{kg K}}\right)} = 0.24 \text{ K},$$

while for aluminium $\Delta T = 1.1$ K. Water is good at storing energy. It can absorb or release large amounts of energy, while its temperature changes only slightly. This is why a big lake will moderate the temperature of its surrounding area.

18.1 Calorimeters

Imagine you want to heat up or cool off an object placed in a quantity of water in an insulated container, but you do not know its heat capacity c_o . The object is heated from T_{io} to T_{fo} . The water and container have known capacities c_w and c_c respectively. We assume they have been in thermal equilibrium from the start, so their temperature must be the same: $T_{ic} = T_{iw}$. Once the object is heated, it reaches thermal equilibrium with the water and container, so the final temperature must be $T_f = T_{fo} = T_{fw} = T_{fc}$. Energy conservation says that the heat going into the object, Q_{in} must be equal to the heat leaving the water and container Q_{out} , therefore we can write:

$$Q_{in} = Q_{out}, \quad (18.3)$$

$$m_o c_o (T_f - T_{io}) = m_w c_w (T_f - T_{iw}) + m_c c_c (T_f - T_{ic}), \quad (18.4)$$

where m_o , m_w and m_c are the masses of the object, water and container, respectively. If the masses, c_w , c_c , and temperatures before are known, then one can measure the heat capacity c_o of the object by measuring the final temperature T_f . This is the principle of *calorimeters*, which are used to measure heat capacity or heat by absorption.

18.2 Heat transfer

There are three main processes that transfer thermal energy:

- *Conduction*: If two object are in direct thermal contact, heat can be transferred by interactions between atoms and molecules.
- *Convection*: Heat is transported by some mass of material moving, like hot air, which rises because it is less dense than the surrounding cold air.
- *Radiation*: Energy is absorbed and emitted by electromagnetic radiation (e.g. visible, infrared light). At equilibrium, an object will emit and absorb the same amount of radiation. If it is hotter than its environment, it will emit more.

Consider a solid bar connecting a steam bath to an ice bath, like in Fig. 18.1. Heat Q will flow uniformly across the bar and the temperature will vary linearly along the bar. The steam bath functions as a *heat source*, and the ice bath as a *heat sink*. Then one can define a *thermal current*

$$I = \frac{\Delta Q}{\Delta t}, \quad (18.5)$$

with an amount ΔQ of heat transferred within a time Δt . The current is proportional to the cross-sectional area A and the temperature difference ΔT , and inversely proportional to the bar's length Δx

$$\frac{\Delta Q}{\Delta t} = \kappa A \frac{\Delta T}{\Delta x}. \quad (18.6)$$

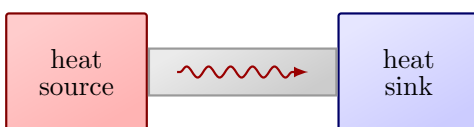


Figure 18.1: Heat conduction between a heat source and heat sink.

Table 18.2: Coefficient of thermal conductivity κ .

material	κ [W/m K]
air (27°C)	0.026
ice	0.592
copper	401
wood	0.11–0.15
glass	~ 0.8
aluminium	237

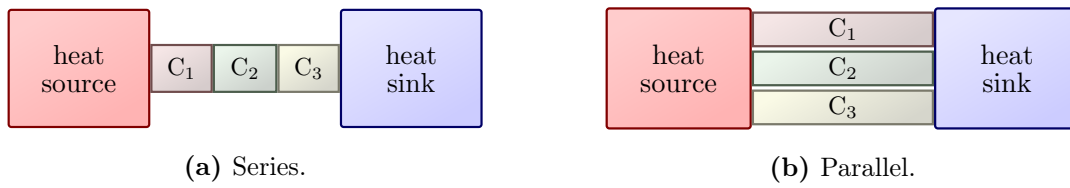


Figure 18.2: Multiple heat conductors.

The proportionality constant is called the *coefficient of thermal conductivity* κ with units W/mK. In analogue to electric currents (Eq. 15.3), one can define a *thermal resistance* R as

$$R = \frac{\Delta x}{\kappa A}, \quad (18.7)$$

such that

$$\Delta T = RI. \quad (18.8)$$

Example 18.3: What is the thermal resistance R for a cylinder of copper with a length $\Delta x = 1$ m and a radius $r = 1$ cm?

$$R = \frac{1 \text{ m}}{\left(400 \frac{\text{W}}{\text{mK}}\right) \left(\pi(0.01 \text{ m})^2\right)} = 8 \frac{\text{K}}{\text{W}}$$

For a wooden cylinder, with smaller conductivity, this is 4000 times bigger.

If the cylinder between the source and sink is replaced by three materials with different thermal resistances R_i in series as is shown in Fig. 18.2a, one can show that the equivalent thermal resistance is

$$R_{\text{eq}} = R_1 + R_2 + \dots \quad (18.9)$$

such that $\Delta T = R_{\text{eq}}I$. On the other hand, if the conductors are placed in parallel, like in Fig. 18.2b, then

$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \dots \quad (18.10)$$

This is in analogue with serial and parallel electric circuits.

18.3 1st law of thermodynamics

Heat can be used to do work W , but it can also increase or decrease the internal energy ΔU . Energy in total, however, needs to be conserved:

1st law of thermodynamics (Energy conservation). *The total energy of a closed system is conserved.*

$$Q = \Delta U + W \quad (18.11)$$

Be careful of the sign conventions, shown in Fig. 18.3: Q is positive if it's *added to* the system, negative if it's *extracted*. Similarly, W is positive if the system *does work* W , and negative if work is done on the system.

Example 18.4: Consider a volume that is filled with gas, and is closed off by a moveable piston (Fig. 18.4). Some time-dependent heat $Q(t)$ is added, and work $W(t)$ will be done by the system.

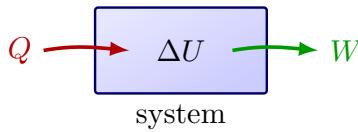


Figure 18.3: 1st law of thermodynamics: The change in internal energy in $\Delta U = Q - W$.

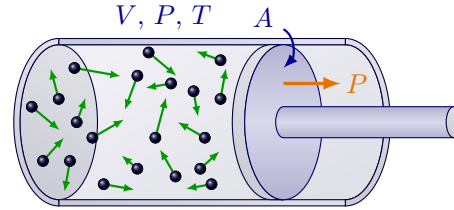


Figure 18.4: Gas with volume V , pressure P , temperature T , and a piston with area A .

Example 18.5: Suppose you have a system of water. We stir it with 15 kJ of energy. We remove 50 kJ of heat. What happens to the temperature? The stirring is work $W = -15$ kJ done on the system. And the heat $Q = -50$ kJ is removed from the system. Therefore,

$$\Delta U = Q - W$$

The internal energy is decreased, and therefore the temperature is decreased as well.

18.4 Heat expansion of solids

Most solids will expand when heated. Consider a thin rod with initial length L_0 that is heated from its initial temperature T_0 to $T_0 + \Delta T$. Then the rod will expand along its length by an amount

$$\Delta L = \alpha L_0 \Delta T, \tag{18.12}$$

with the coefficient of linear expansion α , which carries units K^{-1} or $^\circ\text{C}^{-1}$. This is illustrated in Fig. 18.5. This can also be written infinitesimally as

$$dL = \alpha L_0 dT, \tag{18.13}$$

This can be easily generalized to volumes. Consider a cube of the same material, and with sides of initial length L_0 . Because $V = L^3$,

$$\frac{dV}{dL} = 3L^2. \tag{18.14}$$

Therefore, after heating, the cube will expand in all directions and increase by a volume

$$dV = \alpha(3L_0^2)(L_0 dT), \tag{18.15}$$

or

$$\Delta V = 3\alpha V_0 \Delta T. \tag{18.16}$$

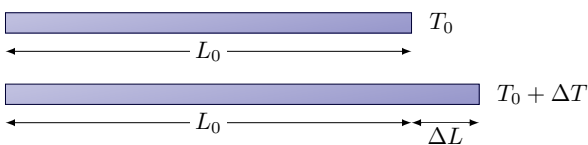
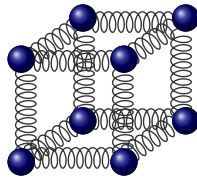


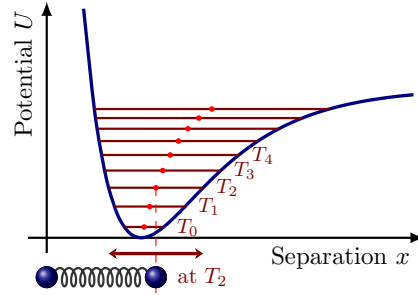
Figure 18.5: Linear expansion of a rod after heating.

Table 18.3: Coefficient of linear expansion α .

Material	α [$1/^\circ\text{C}$]
Aluminium	2.4×10^{-5}
Copper	1.7×10^{-5}
steel	1.2×10^{-5}
Quartz	0.04×10^{-5}



(a) Representation of solid, with bounds between atoms shown as springs.



(b) Morse potential describing molecular vibrations, for different temperatures T_i and average molecular separation $\langle r \rangle$.

Figure 18.6: Illustrations of molecular vibrations.

18.4.1 Thermal expansion at atomic level

Bimetals are fabricated by affixing two thin strips of different type of metals to each other. Each metal has its own coefficient of linear expansion, α . Upon heating, the bimetal will bend in the direction of the metal with the smaller coefficient.

At the atomic level, solids are held together by bonds, which act like springs, as shown in Fig. 18.6a. When the temperature increases, the energy and the amplitude of vibrations increases, as well as the average separation $\langle r \rangle$ between atoms. These inter-molecular bounds are described by the *Morse potential*, shown in Fig. 18.6b.

18.4.2 Strain & Young’s Modulus

The *strain* on a material is defined by

$$\frac{\Delta L}{L}. \tag{18.17}$$

The force exerted to create the strain, is called the *tensile stress*

$$\frac{F}{A}, \tag{18.18}$$

which has the same dimensions as pressure with units N/m^2 or Pa. The Young’s modulus Υ of a material gives the stress needed for an exerted stress.

$$\Upsilon = \frac{F/A}{\Delta L/L}, \tag{18.19}$$

We can create a strain with a thermal expansion, and this will create a stress according to Υ . We bring this mechanical stress-strain relation in this chapter, since the most important (and often neglected) source of strain on a material is temperature changes. Clearly, one can see that if a temperature change yields a length change, it will also create a mechanical stress, that is a force per area that can damage a structure. Take, for example, a steel bridge that will contract in the winter and expand in the summer, and potentially damage the concrete it is attached to.

Table 18.4: Young’s modulus Υ of some common materials.

Substance	Υ [GN/m ²]
Steel	200
Gold	78
Lead	16

Chapter 19

Thermal Radiation

19.1 Black body

Thermal energy can be radiated in the form of electromagnetic waves. An object radiates according to its temperature T ,

Stefan-Boltzmann law.

$$P = e\sigma AT^4, \quad (19.1)$$

with the *emissivity* $0 < e < 1$ of the object and Stefan's constant

$$\sigma = 5.670373 \times 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4} \quad (19.2)$$

At the extreme $e = 1$, the object is a perfect absorber, called a *black body*. On the other end, if $e = 0$, the object is a perfect reflector; it emits nor absorbs radiation.

Object will also absorb radiation from its surroundings,

$$P_{\text{net}} = e\sigma AT_{\text{S}}^4, \quad (19.3)$$

where the temperature T_{S} of the surrounding. If an object is hotter than its environment, it emits more radiation than it absorbs.

$$P_{\text{net}} = e\sigma A (T^4 - T_{\text{S}}^4), \quad (19.4)$$

If $T = T_{\text{S}}$, then $P_{\text{net}} = 0$, and the object is in equilibrium with the surrounding.

Example 19.1 (Human body temperature): Estimate the heat loss of a naked person at room temperature, if the room temperature is T_0 and the skin temperature is T . Human beings have an emissivity of $e = 0.98 \sim 1$, which means we can be well approximated as a black body. To get the area A , we use the approximate a human being to a rectangle of 0.4 by 1.7 m, such that the total area, front and back, is $A = 1.36 \text{ m}^2$. Putting everything together:

$$P_{\text{net}} = (1) \left(5.67 \times 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4} \right) (1.4 \text{ m}^2) (306^4 \text{ K}^4 - 293^4 \text{ K}^4)$$

So $P_{\text{net}} = 110 \text{ W}$. How much food do we need to compensate for this loss of heat after one day? The amount of lost energy is given by $\Delta E = P_{\text{net}} \Delta t$, or

$$\Delta E = \left(110 \frac{\text{J}}{\text{s}} \right) \left(3600 \frac{\text{s}}{\text{h}} \right) \left(24 \frac{\text{h}}{\text{day}} \right),$$

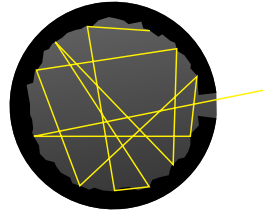


Figure 19.1: Approximation of a black body. All light captured by hole is absorbed by the internal wall as heat, and finally radiated again as thermal radiation.

or, 9.5 MJ. This can be converted into more familiar units of calories to $\Delta E = 2270$ kcal, using $1 \text{ kcal} = 4184 \text{ J}$.

We can find the rate of cooling by deriving Eq. (19.1) with respect to T :

$$dP = e\sigma AT^3 dT. \quad (19.5)$$

This implies that the loss of heat depends on temperature difference between object and its surroundings.

$$\Delta P = e\sigma AT^3 \Delta T \quad (19.6)$$

This is encapsulated by

Newton's Law of cooling. *The rate of cooling is proportional to the temperature difference between an object and its surroundings.*

A *black body* is a perfect absorber and radiator. It can be approximated by a box of any material with a small hole, as is shown in Fig. 19.1. All light (electromagnetic radiation) entering the cavity through the hole is completely absorbed as heat. The object then radiates this added heat as thermal radiation. If the object is in thermal equilibrium, radiation leaving the hole will approximate that of an ideal black body. This *black body radiation* follows a very particular distribution of wavelength, as we will discuss next.

19.2 Rayleigh-Jeans law

Radiation has a wavelength λ characteristic of temperature. The power radiated by a black body is well approximated by

Rayleigh-Jeans law.

$$P(\lambda, T) = \frac{8\pi k_B T}{\lambda^4}. \quad (19.7)$$

For $T < 600 \text{ K}$, light has a wavelength $\lambda > 700 \text{ nm}$, which is in the infrared part of the spectrum. This is invisible to the human eye, which can only see the optical band $400 \text{ nm} < \lambda < 700 \text{ nm}$. Some animals like the mosquito can see in the part of the spectrum. If an object reaches a temperature of about $T \sim 900\text{--}1000 \text{ K}$, the radiation starts to appear as visible red light. At even higher temperatures, the power distribution becomes more intense in the optical spectrum, and more colors start to appear to the point that the object emits is a continuous mix of all optical wavelengths, which appears as white. The visible color of a black body is shown in Fig. 19.2. This is why lava and burning charcoal glow red, while hotter objects glow white like the sun, and even hotter objects turn blue.

The power spectrum peaks at some wavelength λ_{max} , which depends on the temperature, and is given by

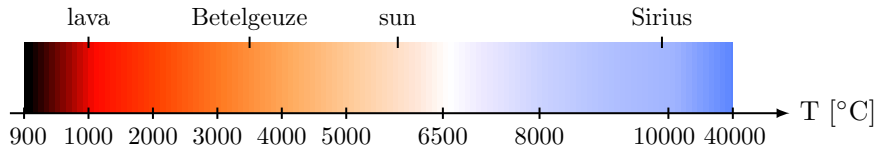


Figure 19.2: Color of a black body for different temperatures T in the optical spectrum.

Wien's displacement law.

$$\lambda_{\max} = \frac{2.898 \text{ mm K}}{T}. \tag{19.8}$$

We can use this relationship to measure the temperature, as is done by a thermal camera.

19.3 Plank's law

Rayleigh-Jeans formula is derived from classical theory of electromagnetism, and is a very good approximation of the black body power spectrum at large λ . However, it breaks spectacularly at small λ . As $\lambda \rightarrow 0$, $P \rightarrow \infty$. This is called the *ultraviolet catastrophe*. This deviation from the actual power spectrum of black body radiation is shown in Fig. 19.3.

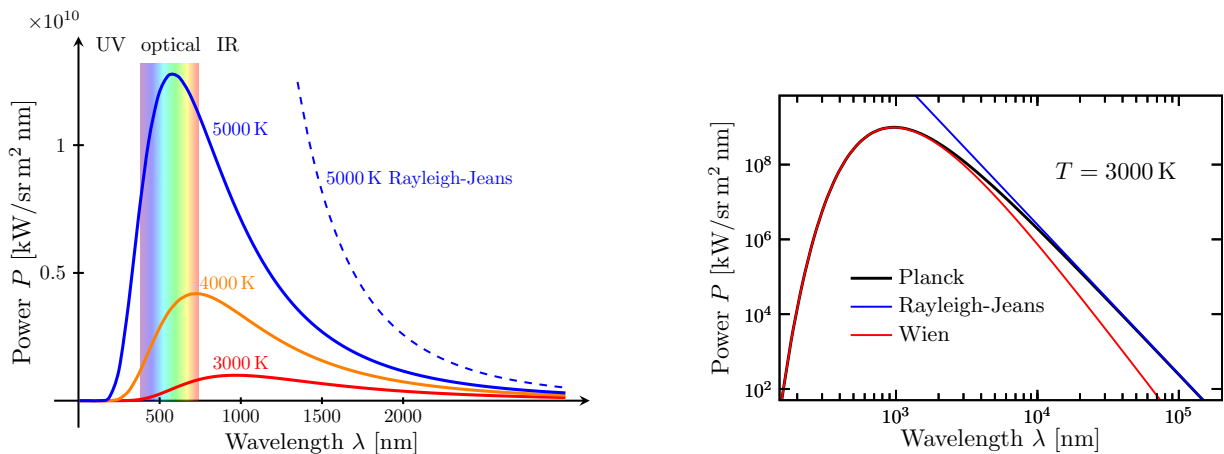
In 1900, Planck found that he could fit the data very well with

Planck's law.

$$B(\lambda, T) = \frac{2\pi c^2 h}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}. \tag{19.9}$$

Here, $B(\lambda, T)$ is the power (W) per unit area (m^2) of the body, and per unit of wavelength (typically nm). The maximum of these distribution are given by Wien's displacement law Eq. (19.8).

The sun, like most stars, follows the black body distribution in the ideal black body spectrum shown in Fig. 19.3a very well to a first approximation, with an effective temperature of about 5800 K. Incandescent light bulbs also emit black body radiation: A current heats their tungsten filament to temperatures somewhere between 2000 and 3300 K, still well below the melting point of tungsten, about 3700 K. This is enough to light a room, but most of the power will be lost as heat through infrared radiation. This is why they are being replaced with more cost-effective fluorescent lamps.



(a) Linear plot for different temperatures. Rayleigh-Jeans deviates at small λ .

(b) Log-log plot for $T = 3000 \text{ K}$. Wien's approximation $P \sim e^{-\frac{1}{\lambda}}/\lambda^5$ (red) agrees well at low λ .

Figure 19.3: Power spectrum of black body radiation.

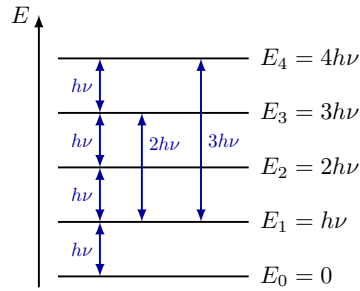


Figure 19.4: Energy levels of a Planck oscillator in the wall of a black body’s cavity, with a frequency ν . The energy levels are equidistant and given by $E = nh\nu$. Energy can be absorbed or radiated by a photon with energy $\Delta E = \Delta nh\nu$.

In any direction of in space you look, you will find a static microwave radiation. This background actually follows a black body distribution with a temperature of about $T = 2.725$ K, and is called the *cosmic microwave background* (CMB). It is thermal radiation left over from the hot plasma in early universe. Because of the universe’s expansion, the matter in the universe has cooled off (and formed stars, planets, dust or gas clouds, etc.), while the original thermal radiation has shifted to the microwave wavelengths due to so-called *red shift*.

19.3.1 Quantum mechanical model

To derive his formula, Planck assumed that the black body radiation was caused by oscillators on the walls of the cavity. We now identify these with vibrating atoms and molecules. Each oscillator is assumed to have some frequency ν and an energy of

Planck’s postulate: oscillator energy.

$$E = nh\nu, \quad (19.10)$$

where n is some integer n , as is illustrated in Fig. 19.4, and h is some constant. Planck did not think these oscillators with discrete, equidistant energy levels really existed, and intended a mathematical trick where one takes $h \rightarrow 0$ to get a continuous energy spectrum. The trick did not work, however. He actually needed to set

Planck’s constant.

$$h = 6.6261 \times 10^{-34} \text{ J s}, \quad (19.11)$$

to describe the observed black body spectrum. This means energy of vibrating atoms and molecules are indeed quantized: Each oscillator can absorb or spontaneously emit a *quantum* of energy $\Delta E = \Delta nh\nu$, which is the difference between two energy levels. The oscillator cannot take on any other energy value than those discrete energy levels.

This is considered to be the birth of quantum mechanics. To explain these discrete energy levels, and the *photoelectric effect*, Einstein later theorized in 1905 that the quantized energy absorption and emissions are carried by electromagnetic waves as discrete packets, or *quanta*. These were later named *photons*. This implied light is made out of particles, which contradicted the long-settled idea that light is an electromagnetic wave. This apparent paradox was later accepted as the *wave-particle duality* in quantum mechanics. A photon has an energy E , which is associated to a frequency $\nu = E/h$ (not to be confused with the frequency of Planck’s oscillator), or equivalently, a wavelength $\lambda = hc/E$.

Chapter 20

Phase Changes

20.1 Isotherms, critical temperatures & vapor pressures

If the temperature of an ideal gas law is kept constant, then the pressure is a simple hyperbolic function in volume:

$$P = \frac{nRT}{V}, \quad (20.1)$$

for any temperature T . These are called *isotherms*, and are shown in Fig. 20.1.

If a gas follows the more realistic van der Waals equation Eq. (17.38), things become a bit more complicated. Fig. 20.2a shows the relation between P and V for different chosen values of T . There are two interesting behaviors to point out. Firstly, as you go to lower volumes, the pressure steeply rises. This means the substance is incompressible, and in fact, it has changed phases from a gas into a liquid. Secondly, for higher temperatures, the curves look similar to that of an ideal gas, but at some *critical temperature* T_C , there appears an inflection point at the *critical volume* V_C and *critical pressure* P_C . This *critical point* is different for each gas. For temperatures $T < T_C$, a local minimum and maximum appear. In reality, it turns out this local minimum and maximum are not physical. Instead, the pressure and temperature stays constant at a specific pressure, as the volume decreases and the gas slowly undergoes a phase transition to liquid. This is drawn in Fig. 20.2b, and how to calculate this is prescribed by the *Maxwell construction*, which falls outside the scope of these lectures. The pressure at which this happens, turns out to be the *vapor pressure* P_{eq} . At P_{eq} , gas and liquid are in equilibrium and can coexist. Adding a small amount of heat will cause the liquid to boil. The vapor pressure depends on the temperature. For water, $T_C = 647 \text{ K}$, or $374 \text{ }^\circ\text{C}$. Below this temperature, water liquid and gas phase can coexist.

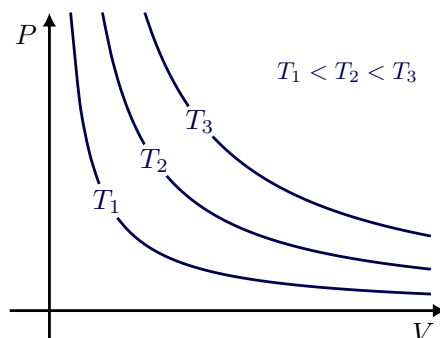
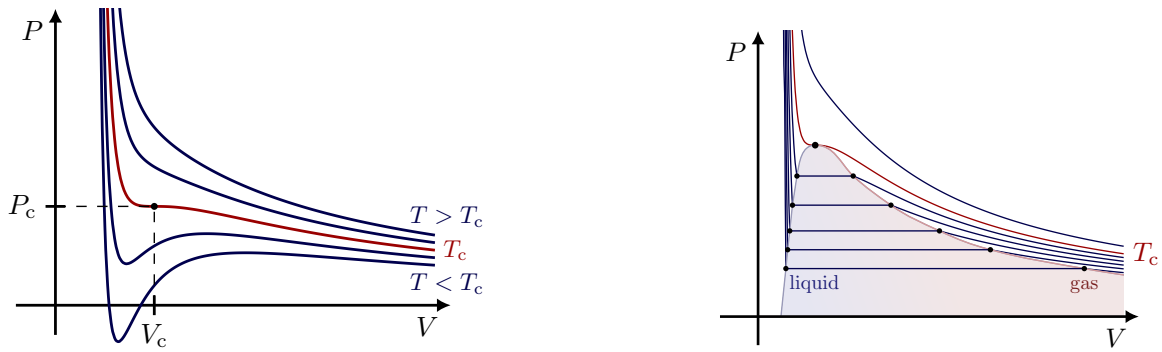


Figure 20.1: Isotherms for an ideal gas, using Eq. (16.1).



(a) PV diagram of a gas described by the van der Waals equation (17.38) for different constant temperatures. At the gas's **critical temperature** T_C , an inflection point appears at the critical point (V_C, P_C) .

(b) Maxwell construction for physical isotherms. The straight lines are states where liquid and gas coexist at a constant vapor pressure and temperature, while the volume changes and the gas or liquid changes phase.

Figure 20.2: Isotherms for van der Waals equation (17.38). Below a gas's **critical temperature** T_C , it will turn into a liquid at high pressures.

Table 20.1: Vapor pressures of water.

T [°C]	Vapor pressure [bar]
0	0.00611
100	10130

At very high temperatures $T > T_C$, however, a gas will liquify at any pressure, as will be described in the next lecture.

20.2 Phase changes & latent heat

Sometimes heat is absorbed or emitted with no temperature change. This is a *phase change* or *phase transition* between different *states of matter*. As heat is added, molecules move apart, which requires work. The extra energy becomes the potential energy of molecules, and there is no temperature change.

Fig. 20.3 summarizes the three main phase changes we observe in daily live and the different possible transitions between them. Some energy is needed to *melt* a liquid to a solid, as the molecules need extra energy to overcome forces of attraction. For melting, this given by the *latent heat of fusion*, L_f , and its total mass m :

$$Q = mL_f. \quad (20.2)$$

Similarly, the energy needed to evaporate a liquid, is called the *latent heat of evaporation*, L_v .

Table 20.2 presents the latent heat of some materials.

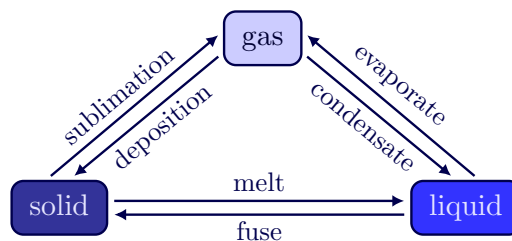


Figure 20.3: Phase transitions.

Table 20.2: Characteristic temperatures, latent heats L and specific heat c of some common materials.

Substance	Melting point [K]	L_f [kJ/kg]	Boiling point [K]	L_v [kJ/kg]	c [kJ/kg K]
Gold	1336	62.8	3081	1701	0.126
Lead	600	42.7	2023	858	0.128
Water	273.15	333.5	373.15	2257	4.18
Ice	273.15				2.05
Alcohol	159	109	351	879	2.4

Example 20.1 (Ice to steam): What is the total amount of heat needed to change 1 kg of ice at -20°C to steam? This problem involves several steps, shown in Fig. 20.4: There are two phase transitions that need latent heat, and in between heat needs to be added to raise the temperature:

1. Heat ice from -20 to 0°C , which needs $Q_1 = mc\Delta T_{12}$.
2. Melt ice, which needs $Q_2 = mL_f$.
3. Heat water from 0 to 100°C , which needs $Q_3 = mc\Delta T_{23}$.
4. Evaporate water, which needs $Q_4 = mL_v$.

So, summing, the total heat needed is $Q = Q_1 + Q_2 + Q_3 + Q_4$.

20.2.1 Phase diagrams

Depending on the pressure P and temperature T , a material will be either a solid, liquid or gas. This can be summarized in a phase diagram, like Fig. 20.5 for water. One can read several things from this diagram:

1. The pressure-temperature points for which matter changes phase are given by a line.
2. Along lines, two phases can coexist, while away from the line, only one phase exists.
3. At the triple point all three phases can exist.
4. Above the critical point pressure P_C , there is no distinction between liquid and gas. At P_C , T_C , density of liquid and gas phases are the same. This is sometimes called a *supercritical fluid*.

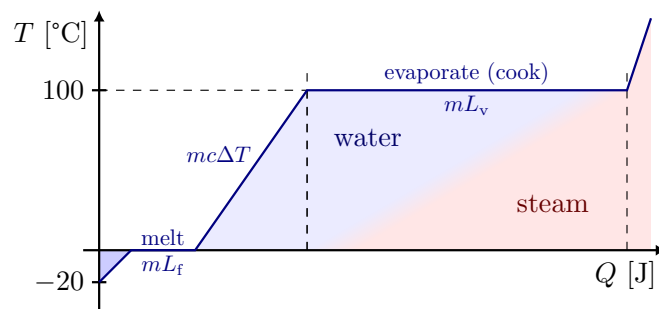


Figure 20.4: Heating up ice to steam (not to scale). During phase changes, the temperature stays constant, as heat is needed to overcome attractive forces.

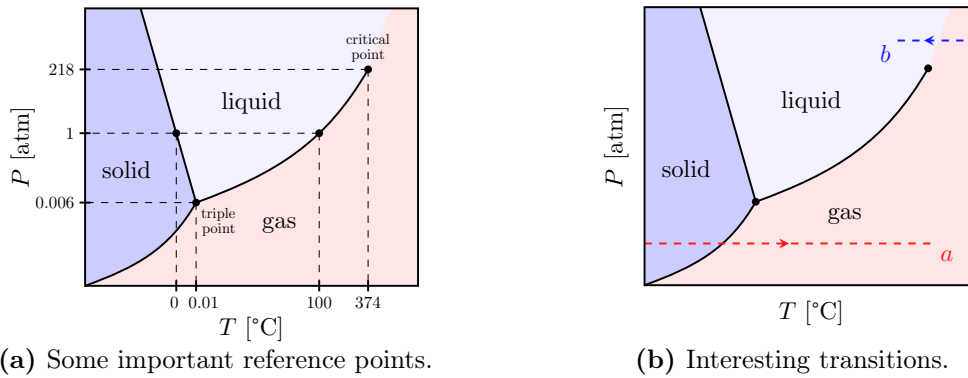


Figure 20.5: Phase diagrams of water.

Let's move along **line a** on phase diagram Fig. 20.5b. We heat, while keeping the pressure constant. We are below the triple point pressure, we go from solid to gas. For CO_2 for example, the triple point is 5.1 atm, so we always have either solid or gas at normal atmospheric pressure. Along **line b** , there are no clear and sudden phase transitions anymore.

Example 20.2 (Opalescence): We heat up SF_6 , and see a continuous or second-order phase transition. Once we cooled it again, the mixture becomes completely opaque, which is called *critical opalescence*. This was explained by Einstein in 1910, when he was a professor at UZH, who realized that the mixture of gas and liquid has fluctuations of a size like the wavelength of visible light.

Chapter 21

Thermodynamic Processes

21.1 PV diagrams

Examine the work done by changing the volume V and or pressure P :

Work by a gas.

$$W = \int P dV \quad (21.1)$$

One can create work in different processes through heating, cooling, expanding and compressing gasses. Several processes are summarized in Fig. 21.1. These are so-called PV diagrams, and Eq. (21.1) implies that one can know the amount of work done from the area under the line, which is the integral in Eq. (21.1).

In Fig. 21.1a, a gas is first expanded at a constant pressure P_1 from a volume V_1 to V_2 . Then, the pressure is lowered from P_1 to P_2 . The amount of work done is simply the rectangle

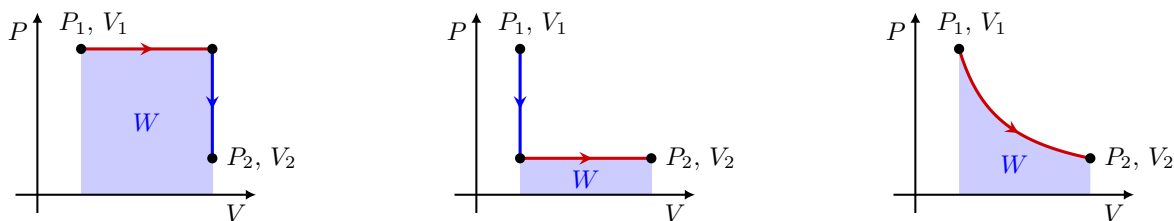
$$W = P_1(V_2 - V_1). \quad (21.2)$$

Notice that by expanding the gas, for example by heating it and allowing a piston to move out, work was done, while the last step did not cost, nor deliver any work. The pressure can be lowered at fixed volume by cooling off the gas, while fixing the piston.

In Fig. 21.1b, the order of the two steps is switched, and we recover a smaller area:

$$W = P_2(V_2 - V_1). \quad (21.3)$$

Another typical process is an *isothermal expansion*, shown in Fig. 21.1c, where the temperature is kept constant, while expanding the volume. Using the ideal gas law, we see



(a) Expanding at constant P , compressing at constant V . (b) Compressing at constant V , expanding at constant P . (c) Isothermal expansion.

Figure 21.1: PV diagrams for different processes where work W is done.

that we can use

$$dW = PdV = \frac{nRT}{V}dV, \quad (21.4)$$

which we can integrate to

$$W = \int_{V_1}^{V_2} P \frac{nRT}{V} dV. \quad (21.5)$$

This results in

Work for isothermal expansion.

$$W = \frac{nRT}{V} \ln \frac{V_2}{V_1}. \quad (21.6)$$

If one goes back to the initial state, one can close the curves in the PV diagrams and create a *cycle* that can be repeated. Cycles in PV diagrams will be the basis for our study of engines later on.

21.2 Relation between heat capacities

Let's examine how we can heat and cool things by an amount ΔT through adding heat Q , or increasing the internal energy with ΔU a bit closer. There are different paths you could take, like shown in Fig. 21.2a.

Previously we already saw the definition of the heat capacity as $C = mc$ in $Q = C\Delta T$. For gases, the heat capacity is *path dependent*. From now on, we will distinguish between C_V and C_P , where

$$C_V = \left(\frac{\partial Q}{\partial T} \right)_V \quad (21.7)$$

is the heat capacity if the gas is being held at a constant volume, while

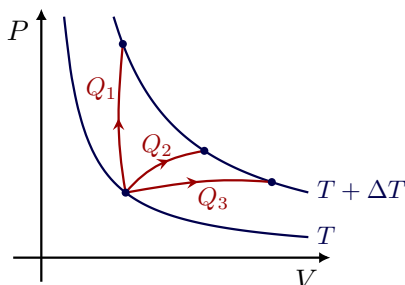
$$C_P = \left(\frac{\partial Q}{\partial T} \right)_P \quad (21.8)$$

is the heat capacity if the gas is being held at a constant pressure. These two paths are illustrated in Fig. 21.2b. C_V and C_P are not the same, but how do C_P and C_V relate for an ideal gas? We remember from the 1st law of thermodynamics (18.11) that

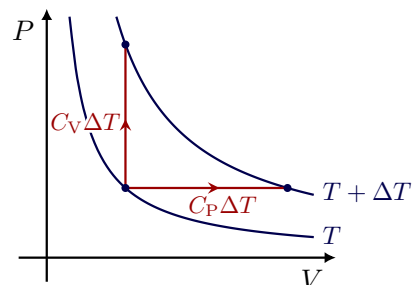
$$Q = \Delta U + W. \quad (21.9)$$

Therefore, at constant volume, we add heat.

$$Q = \Delta U, \quad (21.10)$$



(a) Different paths between isotherms.



(b) Constant volume or pressure.

Figure 21.2: Adding heat Q to a gas to increase its temperature by ΔT .

because no work is done ($W = 0$) while the volume stays constant ($dP = 0$).

$$Q = \Delta U = C_V \Delta T. \quad (21.11)$$

Now taking the infinitesimal of $\Delta T \rightarrow 0$, we can write

$$dU = C_V dT, \quad (21.12)$$

or,

$$C_V = \frac{dU}{dT}. \quad (21.13)$$

Now assume, the pressure is kept constant while the volume is allowed to change:

$$Q = C_P \Delta T \quad (21.14)$$

On the other hand, the gas expands with added heat:

$$\begin{aligned} Q &= \Delta U + \Delta W \\ &= \Delta U + P \Delta V. \end{aligned} \quad (21.15)$$

For small changes,

$$C_P dT = dU + p dV. \quad (21.16)$$

Using Eq. (21.13), and because we assumed an ideal gas, we can write

$$C_P dT = C_V dT + nR dT. \quad (21.17)$$

We find that

Ideal gas heat capacities.

$$C_P = C_V + nR. \quad (21.18)$$

This means that for an ideal gas, C_P is always larger than C_V by an amount nR . For a liquid or solid, $C_P \sim C_V$ because they do not expand or compress ($dV = 0$).

If you change the temperature at a constant volume, no work is done because $dQ = C_V dT$. If you change the volume, but keep the pressure constant, work will be done, and heat will be exchanged:

$$dQ = C_V dT + nR dT. \quad (21.19)$$

We saw earlier that the translational kinetic energy of an ideal monatomic gas is

$$K = \frac{3}{2} nRT. \quad (21.20)$$

If the internal energy was purely translational energy, then,

$$U = \frac{3}{2} nRT, \quad (21.21)$$

or,

$$C_V = \frac{dU}{dT} = \frac{3}{2} nR. \quad (21.22)$$

From Eq. (21.18), it follows that

$$C_P = \frac{5}{2} nR. \quad (21.23)$$

If we have a diatomic gas, like H_2 , there are two rotational degrees of freedom in addition to our three rotational ones (see Fig. 15.1c), so now

$$U = \frac{5}{2}nRT, \quad (21.24)$$

from which we find that

$$C_P = \frac{7}{2}nR. \quad (21.25)$$

The ratio of the heat capacities is dimensionless:

Heat capacity ratio.

$$\gamma = \frac{C_P}{C_V}. \quad (21.26)$$

For an ideal gas we found Eqs. (21.22) and (21.23), such that

$$\gamma = \frac{5}{3}. \quad (21.27)$$

Meanwhile for a diatomic gas,

$$\gamma = \frac{7}{5}. \quad (21.28)$$

A comparison is shown in Table 21.1.

Table 21.1: Comparison of formula for heat capacity C of idealized gasses.

Ideal gas	C_V [J/K]	C_P [J/K]	$C_P - C_V$ [J/K]	$\gamma = C_P/C_V$
Monoatomic	$\frac{3}{2}nR$	$\frac{5}{2}nR$	nR	$\frac{5}{3}$
Diatomic	$\frac{5}{2}nR$	$\frac{7}{2}nR$	nR	$\frac{7}{5}$

21.3 Adiabatic expansion

An *adiabatic* process is one in which no heat enters or escapes the system. It can occur in two primary ways with gases. A quasi-static, adiabatic process occurs in a reversible way. That is, one could represent the change from initial to final state as a series of small reversible steps, with continuous variations of pressure, volume, and temperature according to gas laws.

In an adiabatic expansion, the volume expands more rapidly than the time scale for heat to enter or exit the system. Therefore, there is no exchange of heat, $Q = 0$. Owing to the 1st law of thermodynamics (18.11), then

$$\Delta U = -W. \quad (21.29)$$

As the system expands, it does work. The work is time-dependent, so the internal energy ΔU is as well, and decreases, as is indicated by the minus sign in Eq. (21.29).

Adiabatic expansion is depicted in the PV diagram Fig. 21.3a, where we move from point (V_1, P_1) on an isotherm corresponding to a temperature T_1 , to (V_2, P_2) on an isotherm with T_2 . Note that this is a quasi-static process in which the gas does work in order to

expand, in a reversible way. But what is kept constant in this process, and what work is done? We know that the heat exchange is zero:

$$\begin{aligned} 0 &= Q \\ &= dU + dW \end{aligned} \quad (21.30)$$

$$\begin{aligned} &= C_V dT + PdV \\ &= C_V dT + \frac{nRT}{V} dV, \end{aligned} \quad (21.31)$$

which we can rearrange as

$$0 = \frac{dT}{T} + \frac{nR}{C_V} \frac{dV}{V}. \quad (21.32)$$

Since $C_P = C_V + nR$,

$$\frac{C_P - C_V}{C_V} = \frac{nR}{C_V}, \quad (21.33)$$

or in terms of γ from Eq. (21.26):

$$\gamma - 1 = \frac{nR}{C_V}. \quad (21.34)$$

Substituting this into Eq. (21.32):

$$0 = \frac{dT}{T} + (\gamma - 1) \frac{dV}{V}. \quad (21.35)$$

Integrating, we get

$$\ln T + (\gamma - 1) \ln V = \text{constant}, \quad (21.36)$$

or,

$$TV^{\gamma-1} = \text{constant}. \quad (21.37)$$

Using the ideal gas law, we find

Adiabatic process.

$$PV^\gamma = \text{constant}. \quad (21.38)$$

This gives us a description of the curve in Fig. 21.3a. Notice that this curve depends on the degrees of freedom through γ . In our original example,

$$P_1 V_1^\gamma = P_2 V_2^\gamma. \quad (21.39)$$

So what work is done? We have $Q = 0$, but

$$dU = C_V dT, \quad (21.40)$$



(a) Adiabatic expansion (red) between two isotherms. (b) Work done by an adiabatic expansion (red).

Figure 21.3: PV diagrams of adiabatic expansion.

and therefore, using Eq. (21.30),

$$dW = -C_V dT. \quad (21.41)$$

Integrating, we find that

$$W = -C_V \Delta T \quad (21.42)$$

$$= C_V(T_1 - T_2), \quad (21.43)$$

because $\Delta T = (T_2 - T_1)$. Using the ideal gas law once again,

$$W = C_V \left(\frac{P_1 V_1}{nR} - \frac{P_2 V_2}{nR} \right). \quad (21.44)$$

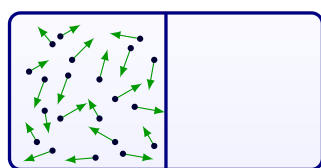
We know that $nR = C_P - C_V$, and therefore,

$$\begin{aligned} W &= \frac{C_V}{C_P - C_V} (P_1 V_1 - P_2 V_2) \\ &= \frac{1}{C_P/C_V - 1} (P_1 V_1 - P_2 V_2) \\ &= \frac{1}{\gamma - 1} (P_1 V_1 - P_2 V_2). \end{aligned} \quad (21.45)$$

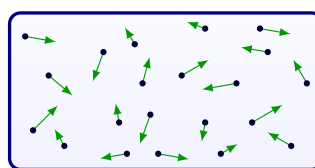
21.3.1 Adiabatic free expansion & enthalpy

A second type of adiabatic process, referred to as adiabatic free expansion, is one in which a gas expands freely without doing work. For this second type, one can imagine an insulated box of an ideal gas with a breakable partition inside like in Fig. 21.4. The ideal gas is contained on one side of the breakable partition. When the partition is broken, the gas expands into the full volume of the box, but without doing work on the system. In this case, there is no heat transferred into or out of the system due to this almost instantaneous change, but also no work is done by or on the ideal gas. Therefore, $Q = 0$ and $W = 0$. From the first law of thermodynamics, Eq. (18.11), this means that U also does not change ($dU = 0$). We also notice that this process is irreversible. It cannot be undone since the gas will not move back into a smaller volume without work being done on it. For an ideal gas, which has no attractive forces between molecules, the internal energy of the gas directly relates to the temperature. Since U does not change, the temperature of an ideal gas does not change during adiabatic free expansion.

However, if the gas undergoing the adiabatic free expansion was a real gas, then it would have attractive forces and a non-zero size of molecules, and the internal energy would then include both kinetic energy and potential energy. In this adiabatic process, it is still true that $Q = 0$ and $W = 0$, so $dU = 0$. However, for a real gas, U of the gas includes both kinetic energy and potential energy due to the attractive forces. Therefore, during the expansion, a real gas may increase or decrease its temperature, depending on whether the potential energy increases or decreases. If work is done by the gas against



(a) Separated by a partition.



(b) After removing the partition.

Figure 21.4: Gas in a box.

the attractive forces, then the gas loses kinetic energy, and its temperature decreases. The opposite can happen also. This depends on the initial and final conditions of the system, as well as the values of attractive force a and volume b , in the van der Waals equation. The temperature of inversion is

$$T_i = \frac{2a}{bR}, \quad (21.46)$$

where R is the ideal gas constant, and b and a are found in Eq. (17.38).

This can be determined with the following steps. In the adiabatic free expansion of a real gas, $dQ = 0$. What is constant in this process is a quantity called *enthalpy* H , where

Enthalpy.

$$H = U + PV \quad (21.47)$$

We can solve Eq. (17.38) for P , and we can calculate the internal energy U of the real gas is the sum of potential energy and kinetic energy, $U = K + E_p$. The kinetic energy is $K = (f/2)nRT$, where f is the number of degrees of freedom, and E_p can be determined from integrating an^2/V^2 from an infinite volume to a finite volume:

$$E_p = \int_{\infty}^V \frac{an^2}{V^2} dV = -\frac{an^2}{V} \quad (21.48)$$

Then the enthalpy can be found as

$$H = K + E_p + PV, \quad (21.49)$$

which is

$$H = \frac{f}{2}nRT - \frac{an^2}{V} + \left(\frac{nRT}{V - bn} - \frac{an^2}{V} \right) V \quad (21.50)$$

and then

$$H = nRT \left(\frac{f}{2} + \frac{V}{V - bn} \right) - \frac{2an^2}{V}. \quad (21.51)$$

Since H is constant, $dH = 0$. Taking partial derivatives,

$$dH = \frac{\partial H}{\partial V} dV + \frac{\partial H}{\partial T} dT. \quad (21.52)$$

Solving for dT , and making some approximations to remove small terms, we find that

$$dT \approx \frac{bRTn^2 - 2an^2}{nR/V^2(f/2 + 1)} \quad (21.53)$$

From the numerator, we see that dT can either be positive or negative depending on the sign of $bRTn^2 - 2an^2$. Therefore, for a free expansion of a real gas, the temperature may increase or decrease as above. For gases expanding at room temperature into atmospheric pressure, most gases including nitrogen, oxygen, carbon dioxide, and argon decrease in temperature, while helium increases in temperature. This is called the *Joule-Thomson effect*. The temperature at which gasses change their behavior, is called the *inversion temperature*,

$$T_i = \frac{2a}{bR}, \quad (21.54)$$

and the values of some typical gasses is given in Table 17.2.

Notice that if we had an ideal gas, a and b would be zero, and dT would be zero, so an ideal gas would not change temperature in free expansion. (However, for quasi-static adiabatic expansion as above, there is a decreases in temperature of an ideal gas as in the formula above.)

Chapter 22

Engines

22.1 A simple engine

Suppose we need some work to lift a mass with weight m . To do this we use a piston filled with gas that we can heat, like in Fig. 22.1. Initially, our setup is at equilibrium. We then perform the following steps:

1. We fix the gas's volume at V_1 , and we start heating the gas in the piston. Because the volume is constant, the pressure will increase to some P_2 .
2. We put a weight on on the platform, which is connected to the piston. We then unfix the piston, such that it expands to a new volume V_2 . The pressure, however, will stay the same. As the piston moves outwards, the platform with the weight will be pulled upward.
3. Now, we fix the volume at V_2 , and we slide the weight off the platform.
4. We remove the heat source. As gas cools off, the pressure will decrease at a constant volume V_2 to P_1 .
5. Once we unfix the piston again, the volume will decrease at constant pressure P_1 down to V_1 .

22.2 Steam engine

By cooling of the gas of the simple engine, we can return to the initial state, and repeat the process again for another mass. This is an example of a simple *steam engine*. Because

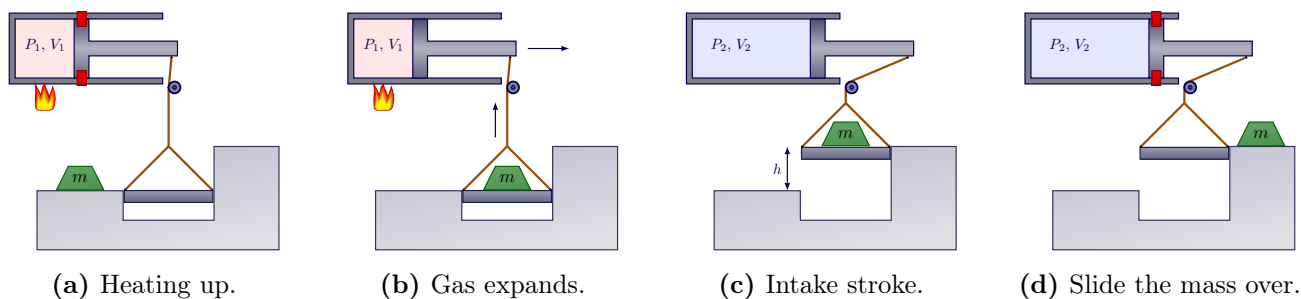


Figure 22.1: Simple engine to lift a mass.

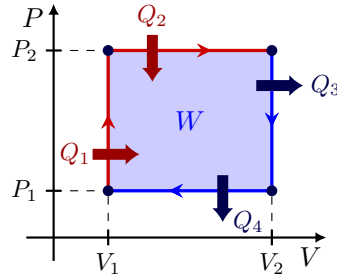


Figure 22.2: PV diagram for the simple process shown in Fig. 22.1.

the initial and final state are the same, this is a *cycle*, and it is summarized in summarized in the PV diagram Fig. 22.2. The total work done is the area:

$$W = P_2(V_2 - V_1) - P_1(V_2 - V_1). \quad (22.1)$$

We can sum up the energies for the total heat in,

$$Q_{\text{in}} = Q_1 + Q_2, \quad (22.2)$$

and the total heat out,

$$Q_{\text{out}} = Q_3 + Q_4, \quad (22.3)$$

with each Q_i as indicated in Fig. 22.2. The work done, then, is

$$W = Q_{\text{in}} - Q_{\text{out}}. \quad (22.4)$$

If you want to lift a mass m , like the simple engine before, the work needs to be

$$W = mgh. \quad (22.5)$$

22.3 Internal combustion engine

Consider the *internal combustion engine* shown in Fig. 22.3. As the name suggest, this engine makes use of combustion of some gas, like oxygen. It is sometimes also called the *four-stroke engine*, because of the different steps in its cycle:

1. *Intake stroke:* We close the exhaust valve, and open the intake valve. As the piston moves down, cool gas and air are sucked in at a constant pressure.
2. *Compression stroke:* We close the intake valve. Now the piston moves up, and compresses the gas.
3. *Ignition:* A spark ignites the gas at a fixed volume, causing the reaction



4. *Power stroke:* The piston is pushed down, as the volume is no longer fixed. This is where the engine actually delivers the power.
5. *Exhaust stroke:* We open the exhaust valves, and hot gas can now escape. The piston moves back to the intial position, and the cycle can be repeated.

This is an example of the *Otto cycle*, and its PV diagram is shown in Fig. 22.4.

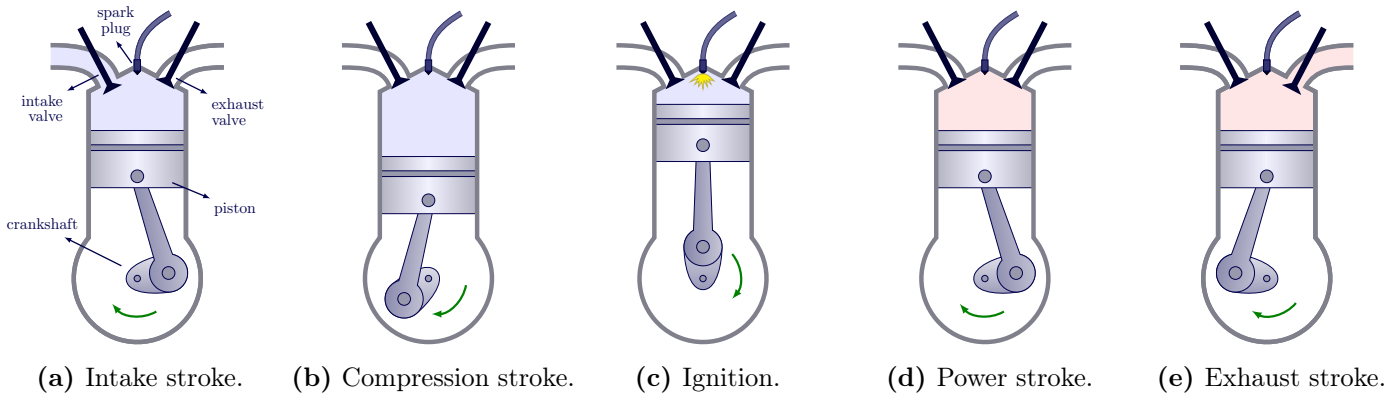


Figure 22.3: Internal-combustion engine.

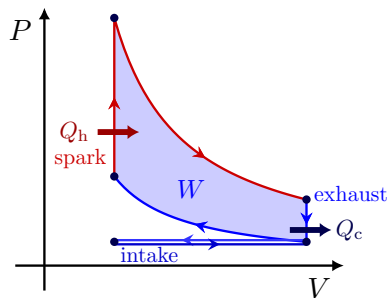


Figure 22.4: PV diagram for the Otto cycle of a internal combustion engine shown in Fig. 22.3.

22.4 Alpha-type Stirling engine

Consider the *single-chamber version of a Stirling engine*, shown in Fig. 22.5. It consists of a fly wheel that connected to two pistons. “Piston 1” can compress the air in a cold chamber, while “piston 2” can the air in a hot chamber. The flywheel has some inertial momentum to keep the process smoothly operating.

1. At the beginning of the cycle, the hot cylinder has the maximum volume $V_2(t_0)$, so the expansion can only happen into the cold cylinder. This forces piston 1 to go up.
2. When the system is at the maximum volume $V_1(t_1) + V_2(t_1)$, it has more contact with the cold cylinder by design. This drives the compression of gas in the cold cylinder.
3. The cold cylinder is now at the maximum volume $V_1(t_2)$. As the wheel spins further, volume $V_2(t_2)$ must increase again, such that more gas will start heating. (Here, the fly wheel’s rotational inertia does work to keep things moving.)
4. When the system is at the minimum volume $V_1(t_3) + V_2(t_3)$, it has more contact with the hot cylinder, driving the expansion of the gas in the hot cylinder.

This works because if the temperature increases, either P or V must increase as well, and the opposite if the temperature decreases.

22.5 Beta-type Stirling engine

There is also a *single-chamber version of a Stirling engine*, shown in Fig. 22.6. It has a fly wheel driving two pistons in the same cylinder. One loosely fitting piston can push away

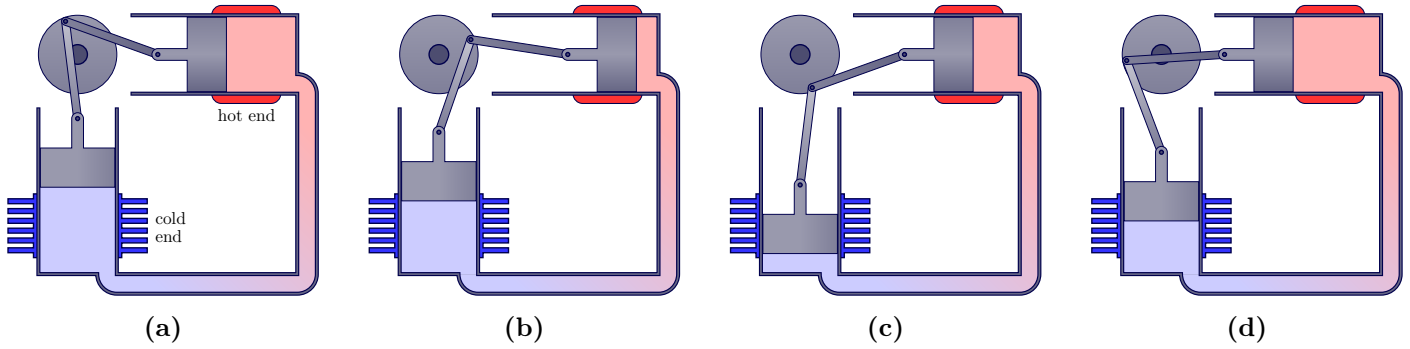


Figure 22.5: Alpha-type Stirling engine with two cylinders.

hot air from the hot end, while one tightly fitting piston will compress the air from the cold end.

1. Most gas is in the hot end. So the gas increases in pressure, and expands into the cold area.
2. The power stroke: The wheel is spinning, which pushes the loose piston down.
3. The loose piston compresses the volume at the hot end, such that the hot gas moves into the cold end. The wheel spins, which pushes the tight piston down again, just as the loose piston moves up.
4. This is easy because the cold gas is easy to compress. So the gas is mostly in the hot end. As we move back to step (1).

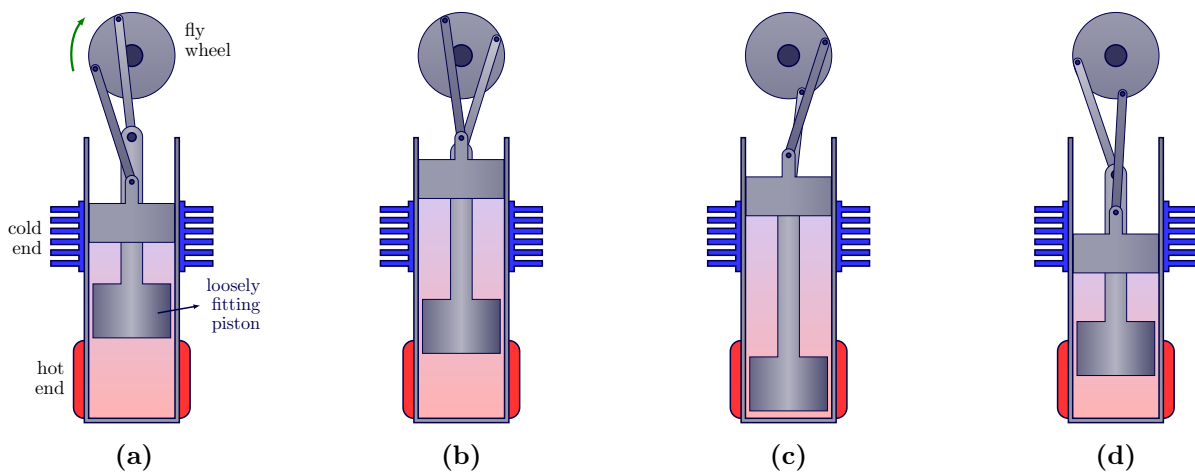


Figure 22.6: Beta-type Stirling engine with two cylinders.

Chapter 23

Entropy & The Second Law

23.1 Efficiency of engines

In all the engines discussed so far, there is a cycle in which a substance absorbs a quantity of heat $Q_{\text{in}} = Q_{\text{h}}$, then it does work, after which it expels a quantity of heat, $Q_{\text{out}} = Q_{\text{c}}$. The temperature increases and decreases in this cycle. A heat engine's purpose is to convert as much Q_{h} into work as possible. A general schematic of a heat engine is shown in Fig. 23.1a.

In a cycle, the initial and final state are the same: There is no change in internal energy ($\Delta U = 0$). From the 1st law of thermodynamics,

First law for heat engines.

$$W = Q_{\text{h}} - |Q_{\text{c}}|, \quad (23.1)$$

as is schematically illustrated in Fig. 23.1a. The absolute value of Q_{c} is used to avoid confusion with signs. The efficiency is defined as how much work W we can get out of an input Q_{h}

$$\epsilon = \frac{W}{Q_{\text{h}}} \quad (23.2)$$

$$= \frac{Q_{\text{h}} - |Q_{\text{c}}|}{Q_{\text{h}}} \quad (23.3)$$

$$= 1 - \frac{|Q_{\text{c}}|}{Q_{\text{h}}} \quad (23.4)$$

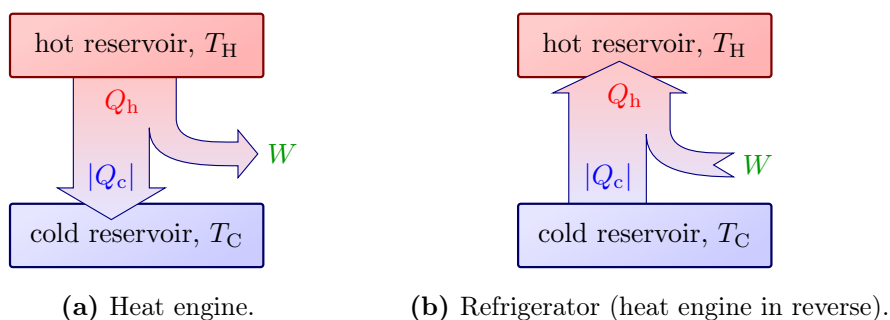


Figure 23.1: Schematic diagrams of first law for engines.

Typical efficiencies are:

- steam engine: $\epsilon \sim 40\%$,
- internal combustion engine: $\epsilon \sim 25\%$ (Formula One car engine: $\epsilon \sim 47\%$),
- rocket engine: $\epsilon \sim 70\%$.

This leads us to formulate the second law:

2nd law of thermodynamics for heat engines. *It is impossible for a heat engine to convert 100% of heat from a heat source at constant temperature into work energy.*

It is possible to do so during isothermal expansion, but not in a cycle, where there will always be heat dissipated, that cannot be used to do work anymore.

23.2 Refrigerators

A refrigerator is like a heat engine, but running “backwards”. Work is added to extract heat from a cold reservoir. This is shown schematically in Fig. 23.1b. A piston (*expansion valve*) is used lower the pressure, and thus the temperature, of refrigeration fluid.

The second law actually implies that one always has need to do some work to move heat from cold to hot reservoirs:

2nd law of thermodynamics for refrigerators. *It is impossible for a refrigerator cycle to only transfer heat from a cold object to a hot object. One needs to do work.*

The measure of performance for cooling is the *coefficient of performance*

Coefficient of performance.

$$\text{COP} = \frac{Q_c}{W}. \quad (23.5)$$

This is understood as the ratio between the energy transferred to the hot reservoir Q_c and the work W required to transfer that energy. The 2nd law of thermodynamics implies that the COP cannot be infinite. Typical COP for refrigerator is ~ 5 – 6 .

Example 23.1: For a refrigerator of COP = 5.5, how much work is needed to make ice cubes from 1 L of water at 10°C ? How much power consumed if we do it in 10 seconds? How much heat is produced into hot reservoir to do this? How much heat needs to be removed.

$$\begin{aligned} Q_c &= Q_1 + Q_2 \\ &= mc\Delta T + mL_F = 375 \text{ kJ}. \end{aligned}$$

So the work is

$$W = \frac{Q_c}{\text{COP}} = 68 \text{ kJ}.$$

The the power is

$$P = \frac{68 \text{ kJ}}{10 \text{ s}} = 6.8 \text{ kW}.$$

A typical fridge uses about 25 W.

$$Q_h = Q_c + W = 443 \text{ kJ}.$$

What if we leave open the fridge door, to try to cool the whole room? We would actually heat the room by $Q = -375 \text{ kJ} + 443 \text{ kJ} = 68 \text{ kJ}$.

23.3 2nd law of thermodynamics for a reversible process

The 2nd law of thermodynamics says efficiency cannot be perfect. So what is the maximum possible efficiency? One can calculate this for a reversible process, with no energy lost to friction, heat conduction, or radiation. All reversible engines have some efficiency, so an ideal gas is used to make the calculations more easy. This problem was solved by Carnot in 1824.

Consider Fig. 23.2 with two isotherms corresponding to temperature T_c and T_h .

- $1 \rightarrow 2$ is an isothermal expansion at T_h ;
- $2 \rightarrow 3$ is an adiabatic expansion from T_h to at T_c ;
- $3 \rightarrow 4$ is an isothermal compression at T_c ;
- $4 \rightarrow 1$ is an adiabatic compression from T_c to at T_h again.

We calculate the efficiency as

$$\epsilon = 1 - \frac{|Q_c|}{Q_h}. \quad (23.6)$$

So what are Q_c and Q_h ? There are several steps. There is no work done in the adiabatic steps, as no heat is exchanged. But work is done during the isothermal expansion in $1 \rightarrow 2$:

$$Q_h = W = \int_{V_1}^{V_2} P dV = nRT_h \ln \frac{V_2}{V_1}, \quad (23.7)$$

and likewise, in the isothermal compression in $3 \rightarrow 4$:

$$|Q_c| = nRT_c \ln \frac{V_3}{V_4}. \quad (23.8)$$

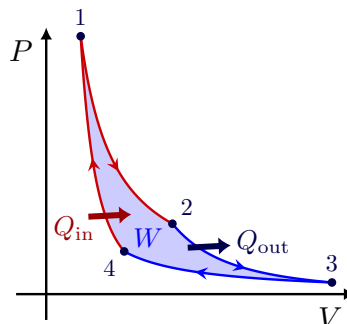


Figure 23.2: PV diagram for a Carnot cycle.

Remember that for adiabatic processes, $TV^{\gamma-1}$ is constant. This allows us to relate V_2 and V_3 in $2 \rightarrow 3$:

$$\frac{T_h}{T_c} = \frac{V_3^{\gamma-1}}{V_2^{\gamma-1}}. \quad (23.9)$$

And similarly for $1 \rightarrow 4$:

$$\frac{T_h}{T_c} = \frac{V_4^{\gamma-1}}{V_1^{\gamma-1}}. \quad (23.10)$$

Therefore, we must have

$$\frac{V_3}{V_4} = \frac{V_2}{V_1}, \quad (23.11)$$

and so the efficiency is

$$\epsilon = 1 - \frac{T_c \ln \frac{V_3}{V_4}}{T_h \ln \frac{V_2}{V_1}} = 1 - \frac{T_c}{T_h}. \quad (23.12)$$

This leads us to yet another way to formulate the 2nd law:

2nd law of thermodynamics (Carnot's formulation). *The efficiency for a perfect, reversible engine is*

$$\epsilon_C = 1 - \frac{T_c}{T_h} = 1 - \frac{|Q_c|}{Q_h}. \quad (23.13)$$

This is the *Carnot efficiency*, and this number cannot be beat by any engine. An efficiency higher than this would violate the 2nd law.

Notice the handy relation for reversible engines:

$$\frac{T_c}{T_h} = \frac{Q_c}{Q_h}. \quad (23.14)$$

We see that the efficiency depends only depends on the temperature difference. Engines are defined with respect to

$$\epsilon_{SL} = \frac{\epsilon}{\epsilon_C}, \quad (23.15)$$

where ϵ is the actual efficiency, and ϵ_C is the ideal Carnot efficiency.

Example 23.2: To get some feelings of these numbers, consider an engine running between room temperature and 500°C ,

$$\epsilon_C = 1 - \frac{296\text{ K}}{773\text{ K}} \approx 0.6. \quad (23.16)$$

Example 23.3: A power plant burning fossil fuels to generate electricity. This has a Carnot efficiency of $\epsilon_C \sim 50\%$, while the actual efficiency is $\epsilon \sim 40\%$. Thus, the second law efficiency is $\epsilon_{SL} = 80\%$. If all the electricity were to be converted directly to heat, $\epsilon = 40\%$. So if the goal is heating, we can in general reach higher efficiencies.

23.3.1 Heat pump

A heat pump is a refrigerator that removes an amount of Q_c of heat from a cold reservoir and puts Q_h into a hot reservoir, using work W to do so. Using the first law Eq. (23.1), we can rewrite the performance as

$$\text{COP} = \frac{Q_c}{W} = \frac{Q_c}{|Q_h| - Q_c} \quad (23.17)$$

What is the best performance we can achieve? Using our Carnot conditions in Eq. (23.13),

$$\text{COP}_{\text{max}} = \frac{T_c/T_h}{1 - T_c/T_h} \quad (23.18)$$

For a typical heat pump with $\text{COP} \sim 3.5$, the efficiency is

$$\frac{|Q_h|}{W} = 1 + \text{COP} \sim 450\%.$$

In words, for some amount of work, we get a heat that is 4.5 times this work, $Q_h = 4.5W$.

23.4 Entropy, reversibility and disorder

The 2nd law of thermodynamics is related to the fact that some processes are irreversible. Irreversible processes move toward a less ordered state. It is impossible to move a system back to a more ordered state without the input of work from outside the system, which will inadvertently increase disorder elsewhere. Conversely, the lower the disorder, or entropy, of a system, the more work can be extracted.

Consider a box of gas, which has a total mass M and a velocity v . The box is moving and therefore has a kinetic energy $K = \frac{1}{2}Mv^2$ that could be used for work. We realize that inside this object there are molecules, each with their own kinetic energy as well. Imagine we collide this box with a wall and stays stuck. This is an inelastic collision where all the initial kinetic energy of the box is lost. This energy was converted into internal energy, which increase the temperature of gas inside. The gas now has random energy, disordered, and lost the ability to do work. This can be described by the *entropy* S , which is a measure of the disorder of a system. Together with the other thermodynamic state parameters P , V , T and U , S describes the state of a system.

The change in entropy for a given change in temperature is defined as

Change of entropy.

$$\Delta S = \int \frac{dQ_{\text{rev}}}{T}, \quad (23.19)$$

where Q_{rev} is the heat that must be added in any reversible process to bring it from its initial state to the final one. The differential dQ_{rev} can be negative if the heat is removed. In that case, ΔS would be negative, and the system becomes more ordered. A reversible process is not needed to actually obtain a difference in entropy ΔS , but we need to choose one to calculate ΔS . This is because the entropy is only a state variable: It describes the current state of the system, and does not care how it got there.

23.4.1 Entropy for a reversible cycle

Let's start with the special case of the reversible Carnot cycle discussed in Section 23.3. The heat exchange only happens during the isothermal processes, so the integrand is constant in temperature, and the total entropy for one full cycle is simply

$$\Delta S = \frac{|Q_h|}{T_h} - \frac{|Q_c|}{T_c}. \quad (23.20)$$

From our previous relation Eq. (23.14), we find that actually

$$\Delta S = 0. \quad (23.21)$$

Since the entropy only depends on the state and not the process, this result holds for any reversible process:

Total entropy of a reversible process. *The total entropy of one full cycle of a reversible process is*

$$\Delta S = \oint \frac{dQ_{\text{rev}}}{T} = 0. \quad (23.22)$$

23.4.2 Entropy for an ideal gas

What is a reversible process where an ideal gas absorbs heat dQ ? We already know that

$$dQ = dU + PdV, \quad (23.23)$$

and for an ideal gas

$$dQ = C_V dT + \frac{nRT}{V} dV. \quad (23.24)$$

We cannot integrate this yet, as V depends on T . But if we divide by T ,

$$\frac{dQ}{T} = C_V \frac{dT}{T} + nR \frac{dV}{V}. \quad (23.25)$$

We can naively integrate by not caring about the process to obtain

$$\Delta S = C_V \ln \frac{T_2}{T_1} + nR \ln \frac{V_2}{V_1}. \quad (23.26)$$

To calculate what the exact value of ΔS is with Eq. (23.19), we need to specify a reversible process. Let's look at several examples of how one would calculate ΔS for an ideal gas.

Isobaric heating

The first example would be heating the gas under constant pressure from T_1 to T_2 . What is ΔS ? Because we are dealing with an ideal gas, we can write

$$dS = C_P \frac{dT}{T}. \quad (23.27)$$

Integrating between the initial and final state,

$$\Delta S = C_P \ln \frac{T_2}{T_1}. \quad (23.28)$$

This result is valid as long as the pressure does not change, and even if the process happened in an irreversible way: The change in entropy ΔS only describes the change of state of the gas, not how it got there.

Isothermal expansion

Another example would be an isothermal expansion of an ideal gas. This means T does not change, but V does. So from Eq. (23.26):

$$\Delta S = nR \frac{V_2}{V_1}. \quad (23.29)$$

The heat added becomes work done:

$$|Q| = W = \int_{V_1}^{V_2} PdV = nRT \ln \frac{V_2}{V_1}. \quad (23.30)$$

Comparing Q to ΔS , we see that

$$\Delta S_{\text{gas}} = +\frac{|Q|}{T}. \quad (23.31)$$

In this process, heat comes from some hot reservoir that is used to heat the gas. The entropy change of the hot reservoir must be

$$\Delta S_{\text{h}} = -\frac{|Q|}{T}. \quad (23.32)$$

So,

$$\Delta S_{\text{gas}} + \Delta S_{\text{h}} = 0. \quad (23.33)$$

This means that in the case of a reversible process, the entropy change of the whole “universe” is zero. Here, by “universe” we mean the system plus its surroundings, which has a total change in entropy

$$\Delta S_{\text{U}} = \Delta S_{\text{gas}} + \Delta S_{\text{h}}. \quad (23.34)$$

Adiabatic free expansion

As a third example, we consider a freely expanding gas in a perfectly insulated box with a removable partition, like in Section 21.3.1. At first, all the gas is on one side of the partition, while the other side is empty. After we remove the partition, the gas uniformly fills up the whole volume of the box. The gas is at temperature T , and the volume increases from V_1 to V_2 . No work was done, and no heat was transferred during this free expansion. We had seen before initial and final temperature are the same. However, this is not a reversible process, because there is no way the gas molecules spontaneously move back to one side of the partition. But we can use another process that has the same initial and final state, but is reversible, to compute ΔS from Eq. (23.19). A good choice is the isothermal expansion in the second example, where you can imagine a piston slowly being pushed by the gas to increase the volume. Therefore, Eq. (23.29) applies.

In our third example, the gas is completely isolated from the surroundings, and this introduces an entropy change of the universe.

$$\Delta S_{\text{U}} = \Delta S_{\text{gas}}. \quad (23.35)$$

This brings us to formulate the 2nd law in terms of entropy:

2nd law of thermodynamics for entropy. *For any process, the total entropy of the universe never decreases:*

$$\Delta S \geq 0. \quad (23.36)$$

Consider a block falling from a height h , and hitting the ground without bouncing. Initially, it has a potential energy mgh . Potential energy is absorbed by the ground and atmosphere, increasing the temperature slightly. If we assume all the energy was converted into heat:

$$\Delta S = \frac{mgh}{T} \quad (23.37)$$

And so the universe's entropy increased by $\Delta S_U = \Delta S$. Energy is "wasted": The useful energy available for work is converted into disordered energy. The energy wasted is $T\Delta S_U$. In an irreversible process, energy equal to the entropy change of the universe times the temperature of the coldest available reservoir becomes unavailable for doing work. We call this work lost

$$W_{\text{lost}} = T_c \Delta S_U. \quad (23.38)$$

In the third example,

$$W_{\text{lost}} = T_c \left(nR \frac{V_2}{V_1} \right), \quad (23.39)$$

while in the second example with isothermal expansion, there was no work lost.

Thermal conduction

Now consider a fourth example, where heat is being conducted from a hot to a cold reservoir, where both reservoirs are assumed to stay at a constant temperatures. What are ΔS_h , ΔS_c , ΔS_U and W_{lost} ? Like in Fig. 23.1a. In principle, this is an irreversible process, as heat would never flow spontaneously from the cold to the hot reservoir, but we can imagine one reversible, isothermal process extracting heat from the hot reservoir, and another dumping it in the cold reservoir. Because the temperatures are assumed constant, integral (23.19) for these isothermal processes become

$$\Delta S_h = -\frac{|Q|}{T_h}, \quad (23.40)$$

$$\Delta S_c = -\frac{|Q|}{T_c}. \quad (23.41)$$

The total entropy for our original thermal conduction between the reservoirs is simply the sum

$$\Delta S_U = \frac{|Q|}{T_c} - \frac{|Q|}{T_h} > 0, \quad (23.42)$$

which is always strictly positive as long as $T_c < T_h$. The lost work becomes

$$W_{\text{lost}} = |Q| \left(1 - \frac{T_c}{T_h} \right). \quad (23.43)$$

Notice that this is the same as the work provided by the Carnot engine with efficiency

$$\epsilon_C = \frac{W}{|Q|} = 1 - \frac{T_c}{T_h}. \quad (23.44)$$

23.5 Entropy and probability

Entropy turns out to be related to probability. This is an important concept explored in statistical mechanics. The state of a gas at macroscopic level (i.e. the *macrostate*) is given by its temperature, volume, pressure, and so on. But as we saw in Chapter 17, we can derive thermodynamic laws with statistical analysis by considering the individual particles, each with their own position and momentum. This is the *microstate* of a gas.

Consider again a freely expanding gas with ΔS given by Eq. (23.29), like in the third example with a partition dividing the box into two equal volumes, such that $V_2 = 2V_1$. Suppose the gas has $N = 10$ molecules. The probability for one molecule to be found on the left side is $P(L) = 50\%$. So the probability for all 10 particles to be on the left is

$$\left(\frac{1}{2}\right)^{10} = \frac{1}{1024}. \quad (23.45)$$

So we can check the gas once per second for 1024 seconds, or equivalently, 17 minutes long. On average, we would see all 10 particles on the left once during this time. At the exact moment all the molecules are on one side, the volume of the gas is then $V_1 = V_2/2$ again, and the entropy would have been decreased by

$$\Delta S_U = \Delta S_{\text{box}} \quad (23.46)$$

$$= nR \ln \frac{V_2}{V_1} \quad (23.47)$$

$$= Nk_B \ln \frac{1}{2}, \quad (23.48)$$

where $nR = Nk_B$, and $\ln 1/2 \sim 0.693$. We have decreased the entropy of the universe, which *seems* to break the 2nd law of thermodynamics! But this only happens 1 out of 1024 seconds. We can think of entropy as a statement of probability, and the second law only emergence from a large number of states. For example, if we take a larger number of particles, $N = 50$, the probability to find all particles on the left side is many orders of magnitude smaller:

$$\left(\frac{1}{2}\right)^{50} \sim 10^{-15}, \quad (23.49)$$

or, once every 36 million years. And for the usual macroscopic systems with $N \sim 10^{23}$ molecules, the probability P is vanishingly small. We see that the probability is

$$P = \left(\frac{V_2}{V_1}\right)^N, \quad (23.50)$$

and we see that we can rewrite

$$\Delta S = Nk_B \ln \frac{V_2}{V_1} = k_B \ln \left(\frac{V_2}{V_1}\right)^N \quad (23.51)$$

It follows that entropy is a measure of the probability of a given state or configuration:

Boltzmann's equation for entropy change. *The change in entropy between to states is given by the probability P of finding the final state, starting from the initial state:*

$$\Delta S = k_B \ln P. \quad (23.52)$$

In general, we can make a stronger statement by defining the *number of microstates* Ω that are available for a given macrostate. We immediately see that the probability P of observing one macrostate consisting of Ω microstates is simply

$$P = \frac{1}{\Omega}. \quad (23.53)$$

This leads to

Boltzmann-Planck equation for entropy. *The entropy for a given macrostate is given by the number of possible microstates Ω :*

$$S = k_B \ln \Omega = -k_B \ln P. \quad (23.54)$$

This equation was developed by Ludwig Boltzmann in the early 1870s. This relation between the macroscopic and microscopic state of a thermodynamic system was so significant, that Eq. (23.54) was engraved on Boltzmann's tomb stone.

At the macro-level, thermodynamic systems are defined by several *state functions*, like temperature T , volume V , pressure P , energy E , etc. But there are many ways an ensemble of microscopic particles can be arranged to have the same set of state functions. Entropy measures in some sense how many *microstates* (Ω) can result in the same *macrostate*. So entropy encodes how much microscopic information is “hidden” if you only know the macrostate from the thermodynamic state functions.

So what is again the connection with order and disorder? Generally speaking, macrostates that are intuitively considered more “orderly”, have fewer possible arrangements of microstates. Vice versa, there are many possible arrangements for disordered states. For example, there are fewer ways to arrange gas particles on one side of the box, while leaving the other side more empty (Fig. 21.4a), than letting the gas fill the whole box (Fig. 21.4b). So a gas that occupies a smaller volume will have a lower entropy.

Example: Gas in a box

Let's make this more concrete. Consider a real gas with N molecules that have some non-zero volume, but for which we neglect the attractive forces. Call the volume of a molecule λ^3 . What is the gas's entropy? Say the gas is in a box with volume V . Then there are

$$M = \frac{V}{\lambda^3} \quad (23.55)$$

positions available to arrange the individual molecules in a non-overlapping way. So the number of possible microstates in a volume V is

$$\Omega = \frac{M!}{N!(M - N)!} \quad (23.56)$$

So the entropy is

$$S = k_B \ln \frac{\frac{V}{\lambda^3}!}{N! \left(\frac{V}{\lambda^3} - N \right)!}, \quad (23.57)$$

which can be simplified with Stirling's approximation, $\ln n! \approx n \ln n - n$ for large n , to

$$S = k_B \ln \frac{V}{N\lambda^3}. \quad (23.58)$$

This is consistent with the thermodynamic identity Eq. (23.26) and Eq. (23.51) for ΔS .