PHYS-200: Physique III

Part 2: Electricity and Magnetism and Special Relativity

Lecture notes

Prof. Hugo Dil

École Polytechnique Fédérale de Lausanne (EPFL)

January 11, 2023

Contents

Introduction

Historically the study of electricity and magnetism marks the transition from classical to modern physics and has formed the basis for most of our current physical approaches and models. The revolutionary step forward was to introduce the concept of fields to describe the interactions between two or more charged objects. This abstract concept of fields is also one of the reasons why many students initially find the topic of electromagnetism difficult as it moves away from things we can grasp and visualise or in a general sense relate to in every day life. Although we only learn the physics behind it in Physique générale I, we grow up developing a feeling for kinematics and mechanics; we know that something falls down if we knock it over and that "objects in motion remain in steady motion unless an external force acts upon them", although we never thought about formulating it that way before college.

The absence of an innate feeling for the laws of electromagnetism can be considered surprising given that it is a force many orders of magnitude stronger than gravity. One can regard electromagnetism as the strongest and most universal force. The Coulomb force between two electrons is about 10^{42} times as large as their gravitational attraction. Electrostatic forces give matter its consistency and allow us to interact with it. Furthermore, the study of electromagnetism will help understand many basic physical phenomena ranging from the obvious examples of electronic equipment and electro motors to copiers, the interaction between light and matter, and how to survive being struck by lightning.

The final goal of this lecture will be the understanding of Maxwell's equations. These equations describe and predict all electromagnetic phenomena and depending on the context they will appear in different forms throughout this lecture. One of the most encountered forms is the so-called

differential form for fields in vacuum:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{1}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{2}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{3}$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{I} + \frac{1}{c^2} \frac{\partial \vec{E}}{\partial t} \tag{4}$$

Here $\vec{E}$ and $\vec{B}$ are the electric and magnetic field, respectively. The $\vec{\nabla}$ vector is called the *nabla* vector and it can be expressed as follows in Cartesian, cylindrical, and spherical coordinates respectively:

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \tag{5}$$

$$\vec{\nabla} = \left(\frac{\partial}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \theta}, \frac{\partial}{\partial z} \right) \tag{6}$$

$$\vec{\nabla} = \left(\frac{\partial}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \theta}, \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \tag{7}$$

Already the shape of the equations and the use of $\vec{\nabla}$ indicates that we will encounter more advanced mathematics with regard to vector calculus in this lecture. Furthermore we will need to use concepts such as (closed) line integrals and (closed) surface integrals. The lecture is too short to explain all these mathematical concepts in detail, but they will be shortly introduced when they are first needed.

Now let's return to the above expression of Maxwell's equations. They represent some of the most important findings of this lecture and before making the first step on our journey it is helpful to realise what some of the destinations are. The first equation states that charges are the source of electric fields and we will encounter this as Gauss's law. From the second equation it follows that magnetic field lines have no beginning and no end; they always close on themselves. Equation (3) represents that a magnetic field which changes with time creates an electric field. This is known as induction and forms the basis for many applications in electronics. The last equations has two parts on the right side. The first part, $\mu_0 \vec{I}$, states that currents, or moving charges, are the source of magnetic fields (in vacuo). The second part shows that also changing electric fields induce a magnetic field, but with much smaller magnitude as c is the speed of light. This is the equation which describes electromagnetic waves such as visible light, but also radio waves and X-rays.

Chapter 1

Charge and current

In this chapter we will introduce two basic entities that will return throughout this lecture: electric charge and current. At first this might seem trivial, but a good definition will save us a lot of confusion later.

1.1 Electric charge

Based on experiments with static electricity shown during the lecture we can conclude that only two types of charge exist: positive (+) and negative (−). When two (or more) charges are put together we see that like charges repel each other and opposite charges attract each other. This will later help us assign a sign to the force between charges and define the direction of electric field lines. The attraction (repulsion) of opposite (like) charges also leads to many electrostatic phenomena that we encounter in daily life. If the air is dry, people, or animals like cats, can become positively charged due to the interaction with the environment whereby negative charge is passed on. The famous examples are rubbing a balloon on hair or taking off a sweater on a dry winter day. Afterwards the excess positive charges in for example our hair repel each other and the hair stands out. But it can also be that we attract neutrally charged light objects because our positive charge attracts the negative charge in these objects and repels its positive charge. Due to the difference in distance the forces do not cancel and the object is attracted as is the case for the cat and the packaging material shown in the lecture. This creation of a charge imbalance in an otherwise neutral object is called electrostatic induction and this forms the basis for many old fashioned high voltage generators such as the Wimshurst generator. A beautiful example is also the Kelvin generator shown in Figure 1.1 where a small charge imbalance is enhanced further and further by letting tap water

flow until it creates a spark.

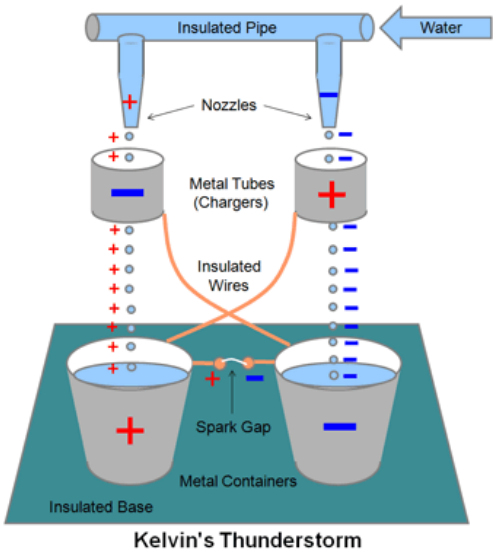

Figure 1.1: Kelvin generator (from www.mpoweruk.com/homebrew.htm)

In the above we have implicitly assumed another important property, namely the conservation of charge. It means that charge can neither be made nor destroyed, it can only be moved around. If we call an object neutral it only means that there is the same amount of positive and negative charge. Now if we bring another charge close we separate the positive from the negative charge. Depending on the properties of the material it is more or less difficult to separate the charges. We will come back to this in more detail in the discussion of *dielectrics*. For the moment we will just distinguish *metals*, where charges move very easily, and *insulators*, where charges can only be moved over very small distances.

Early experiments from the 1820s based on electrolysis have already shown that charge is discrete. Later it was realised that the smallest unit of charge is the electron (e^-) which has a charge of -1.6×10^{-19} coulomb [C], whereby the coulomb is the SI unit for charge. Throughout this lecture Q will be used to identify charge. However, when talking about individual charges we sometimes use q for the single charges and Q for the total charge to avoid confusion. In order to handle mathematical problems that will arise in later chapters we here define some special cases of charge:

- **Point charge**: this is a mathematic concept to simplify problems. It considers that all charge Q is located on a point without spatial

extension.

- **Line charge density** λ: used when only one coordinate is important. It does not mean that all charge is necessarily located on a line. It can also be used for wires and cylinders where for simplicity we can sometimes consider that all charge is located on a line in the centre.

- **Surface charge density** σ: used when two coordinates are important to describe the charge. This can be for example for a plane or the surface of a sphere or cylinder. In a practical sense this can often be used for metals because as we will see later all charge in that case is located on the surface.

- **Volume charge density** ρ: this is the most realistic situation where the amount of charge depends on three coordinates and it is therefore used for charged volumes. Also when no details are specified, such as for general laws, the volume charge density is used.

In all cases the total charge can be calculated by summation (for the point charges) or integration. The integration of course only has to be performed along the coordinate(s) that are relevant. For example to obtain the total charge Q for a system described by surface charge density σ which is a function of x and y:

$$Q = \int_x \int_y \sigma(x, y) dx dy = \iint_A \sigma dA$$

Here dA is an infinitesimally small part of the surface. Throughout this lecture dA will be typically used for open surfaces and dS for closed surfaces. However, this should not be an issue, as also $\iint_{\clubsuit} \sigma d\clubsuit$ yields the same result (but is more annoying to write or say out loud).

1.2 Current as moving charge

Let's start by considering two separated regions, one is charged with $+Q$ and the other one with $-Q$. If we now connect these two regions by a metallic wire, heat will be produced in the wire as an indication that a current is flowing. After a while both regions are neutral and the current stops. As we will see later this *transient current* is typical for a discharge of charged regions. If we keep on adding positive charge to one side or negative charge to the other side the current will not stop and we speak of a *steady current*. These experiments show that current is related to the

change in time of charge, but we don't know how. Around 1875 Rowland's experiments with a rotating charged disk showed that moving charge and current (I) are equivalent. Because we are free to choose the units we can therefore define:

$$I = \frac{dQ}{dt} \tag{1.1}$$

and thus the total charge Q that has flown over time T is

$$Q = \int_T I(t)dt \tag{1.2}$$

The SI unit of current is the ampere [A] which is equivalent to one coulomb per second [C/s].

One of the consequences of current being equivalent to moving charge is that current has a magnitude and a direction; how much charge is moving which way. Therefore to be exact current should always be expressed as a vector ($\vec{I}$) unless there is no way that it could cause confusion, such as is the case for a current through a wire.

Similar to what we did for charge we can also define a **current density** $\vec{j}$, whereby we can restrict ourselves to the volume current density. Due to the vectorial nature we have to be a bit more careful now and define it as *the current per unit area normal to the flow*. If $d\vec{A}$ is the vector normal to a small surface area dA we have the following situation:

$$\begin{aligned} \text{if } d\vec{A} \parallel \vec{j} &\rightarrow \ dI = jdA \\ \text{if } d\vec{A} \perp \vec{j} &\rightarrow \ dI = 0 \end{aligned}$$

Using the scalar product the total current passing through a surface A can be more generally expressed as:

$$I = \iint_A \vec{j} \cdot d\vec{A} \tag{1.3}$$

In this case we can still picture what we are integrating, namely the amount of charge passing through a surface per unit of time. In future chapters we will encounter similar integrals where the physical picture is less obvious, so this is a good point to become familiar with it.

Chapter 2

Coulomb's law

Coulomb's law describes the interaction between charges. It is the central law for electrostatics and plays there a similar role as Newton's gravitational law does for classical mechanics. We will see that it has a similar form as the gravitational force, but with some important differences. Firstly it is many orders of magnitude stronger. More importantly in contrast to gravity it can be both attractive and repulsive. At the moment we will restrict ourselves to charges in vacuo, but in later chapters we will see that the general form will not change.

2.1 Derivation of Coulomb's law

In order to derive a general law we have to consider the following points: the direction of the force, the dependence on the magnitude and sign of the charges, and the dependence on the distance. In order to provide a better understanding of the background of the Coulomb force, and how it was derived in the eighteenth century, these three aspects will be shortly explained here.

Direction

From Newton's third law we know that for every action there is an equal and opposite reaction; the force of charge 1 on charge 2 is the opposite of the force of charge 2 on 1 : $\vec{F}_{12} = -\vec{F}_{21}$. From symmetry considerations we then see that the only option is that the force is directed along the line connecting the two charges. This means that the Coulomb force is a *central force*. In later chapters we will also see examples of forces that are not central.

Dependence on magnitude and sign of charges

In the previous chapter we determined that same charges repel each other and that opposites charges attract. Mathematically this can be represented by multiplication: $- \times - = +$, $+ \times + = +$, but $- \times + = -$. Furthermore we have seen that the force increases when one (or both) of the charges increases, which also hints at a multiplication. Therefore we can say that the Coulomb force is proportional to the product of the charges:

$$F \propto Q_1 Q_2$$

Dependence of distance

From experiments we know that the force between charged objects becomes smaller when the distance between them increases. Without losing generality we can thus say that the force is inversely proportional to the distance r to the power n: $F \propto \frac{1}{r^n}$. Historically much of the effort went into deriving the value of n by means of a variety of experiments. Using a torsion balance Coulomb determined that it must be very close to 2, but the most elegant experiment came from Cavendish. It is based on placing a charge inside a homogeneously charged sphere. If $n \neq 2$ then this charge will experience a force, if $n = 2$ then it will experience no force. By ever increasing precision of the measurements it is now possible to state that $n = 2$ with a possible deviation of 3×10^{-16}. We can thus confidently state that:

$$F \propto \frac{1}{r^2}$$

Putting the previous findings together we obtain that the Coulomb force is proportional to the product of the charges, inversely proportional to the square of the distance, and is oriented along the vector connecting the two charges. In SI units (the charge in coulomb [C] and the distance in meter [m]) the proportionality constant is

$$\frac{1}{4\pi\epsilon_0} \approx 9 \times 10^9$$

in units of Newton times meter squared divided by coulomb square [$\mathrm{Nm^2C^{-2}}$]. Here ϵ_0 ($\approx 8.8 \times 10^{-12}$ $\mathrm{C^2N^{-1}m^{-2}}$) is the *permittivity of free space*, sometimes also called the dielectric constant.

In order to write the Coulomb force in a single formula it is useful to repeat the concept of a unit vector. A unit vector is a vector with length 1 (no units) that is pointing along the direction we are interested in. For

example the unit vector $\hat{r}_{12}$ between charges Q_1 and Q_2 connected by a vector $\vec{r}_{12}$ is

$$\hat{r}_{12} = \frac{\vec{r}_{12}}{|\vec{r}_{12}|} \tag{2.1}$$

Here $|\vec{r}_{12}|$ is the length of the vector, or in other words the distance between Q_1 and Q_2, which we will refer to as r_{12}. Note that the vector $\vec{r}_{12}$ points from charge 2 to charge 1.

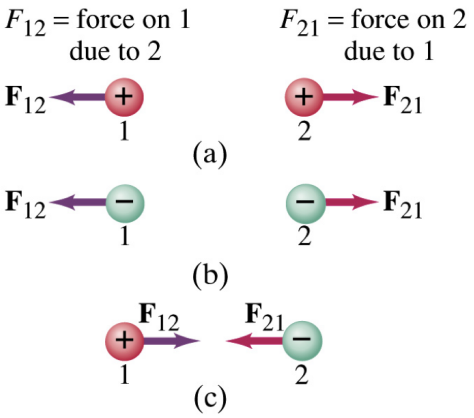

Figure 2.1: Basic illustration of the Coulomb force between two charges

Now we can write **Coulomb's law** between two charges Q_1 and Q_2 in vectorial form as

$$\vec{F} = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r_{12}^2}\hat{r}_{12} \tag{2.2}$$

In Figure 2.1 a basic representation of the Coulomb force between two charges of varying polarity is shown.

To obtain a feeling for the strength of the Coulomb force in everyday life we can consider kitchen salt. This is made up of sodium (Na) and chlorine (Cl) atoms arranged in a 3D checkerboard as shown in Figure 2.2 with a distance between the atoms of $r = 0.28$ nm. The Na atoms each transfer one electron to the Cl atoms and the bonding between them is then due to the attractive Coulomb force. The charge of a single electron is 1.6×10^{-19} C and if we enter all values in Equation 2.2 we obtain a force between two atoms of $F \approx 3 \times 10^{-9}$ N. In 1 mm^2 of salt there are about 10^{13} atomic bonds, yielding that the force needed to break the bonds in 1

mm^2 of kitchen salt is 3×10^4 N. Or in other words it holds about 300 kg! *(This is a simplified example of the use of the Coulomb force. For a more exact calculation one needs to also consider the repulsive force from the next atoms with the same charge, and the attractive force from the next opposite charge after that. If one does this using the Coulomb force one gets a very good comparison to the experimentally determined bonding strength. Which indicates that the Coulomb force is also valid at atomic scales.)*

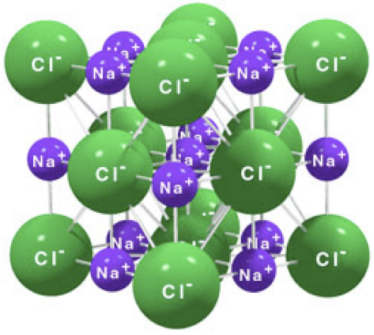

Figure 2.2: Atomic structure of kitchen salt (NaCl)

2.2 Superposition principle

In the previous section we have considered the force between just two charges, but in more realistic situations one would encounter a variety of charges that are either positive or negative, or one has a charge distribution. In these cases the vectorial form of the Coulomb equation has significant advantage as it allows us to use the superposition principle. This states that the total force on charge Q in the environment of N other charges labelled Q_i, is simply the vectorial sum of all individual forces: $\vec{F} = \sum_i \vec{F}_i$. The total Coulomb force can thus be expressed as

$$\vec{F} = \frac{Q}{4\pi\epsilon_0} \sum_{i=1}^{N} \frac{Q_i \hat{r}_i}{r_i^2} \tag{2.3}$$

Here $\vec{r}_i$ are the vectors connecting charge Q with charge Q_i. Depending on the arrangement of the charges and on their sign and magnitude this can also yield a zero force on the charge under investigation if we consider all other charges as fixed.

We can extend the superposition principle to continuous charge distributions. In this case the summation becomes an integration over the charge density which is defined as in Section 1.1. The most useful situation to consider, also with respect to later chapters, is that of a point charge Q in proximity to an object τ with a continuous charge density ρ. (The same arguments are valid for surface and line charge densities, see also the examples in the lecture and exercises.) Because in the charged volume we can define small parts $dq = \rho d\tau$, the Coulomb force on Q from every volume element $d\tau$ connected by $\vec{r}$ is:

$$d\vec{F} = \frac{Q\rho\hat{r}d\tau}{4\pi\epsilon_0 r^2}$$

The total force on point charge Q from the charged object τ then becomes:

$$\vec{F} = \int_\tau d\vec{F} = Q\int_\tau \frac{\rho\hat{r}d\tau}{4\pi\epsilon_0 r^2} \tag{2.4}$$

This equation fails to reveal the full complexity of the situation, namely that during the integration also $\vec{r}$ changes and that this is thus a variable in the integral. Even if τ has a very simple shape, such as a rod or disk, the integral can typically only be solved by smart substitutions and considering the symmetry of the system. To simplify the calculation it is sometimes useful to consider further steps of the superposition principle. For example a negatively charged disk with a hole in it can be seen as a superposition of the negatively charged disk without the hole and a smaller positively charged disk. The focus of this course does not lie on the mathematics, but students are expected to follow and understand the examples given in the lecture and exercises. In a further generalisation ρ can also be inhomogeneous and thus be a further variable in the integration.

2.3 Mutual potential energy of charges

Similar to other forces, such as gravity, we can also associate a potential energy to the Coulomb force. This potential energy can directly be compared with, or summed to, other potential energies. In the next chapter we will derive the more powerful concept of potential as a field quantity and many electrostatic courses skip the current part. However, it is instructive to include it, also to extend the similarities to previous physics courses as far as possible.

The change in potential energy ΔU_{12} is the work done by an external force solely in changing the configuration from state 1 to 2. As illustrated

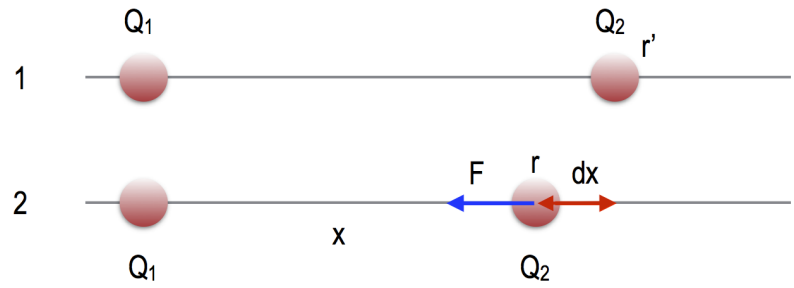

Figure 2.3: Illustration of the mutual potential energy between two charges.

in Figure 2.3 configuration 1 can for example be two positive point charges Q_1 and Q_2 at a distance r'. Configuration 2 is the same two charges placed at a distance r. In order to move charge Q_2 by the distance dx a force F which is opposite to the Coulomb force has to be applied as indicated in the figure. The work done in this infinitesimal movement is

$$dW = Fdx = \frac{-Q_1 Q_2 dx}{4\pi\epsilon_0 x^2}$$

and the total work done in moving the charges from a distance r' to r is equal to the change in potential energy and can be expressed as

$$\Delta U_{12} = W = \int_{r'}^{r} \frac{-Q_1 Q_2 dx}{4\pi\epsilon_0 x^2} = \frac{Q_1 Q_2}{4\pi\epsilon_0}\left(\frac{1}{r} - \frac{1}{r'}\right) \tag{2.5}$$

Instead of working with a potential energy difference it is often easier to define a more absolute scale. This can be done by considering configuration 1 as with the two charges infinitely far apart and call this zero; $U = 0$ for $r' = \infty$. Then Equation 2.5 reduces to the following

$$U = \frac{Q_1 Q_2}{4\pi\epsilon_0 r} \tag{2.6}$$

Two important points should be noted about this equation. Firstly because it was derived by integration it is purely a scalar function; i.e. all vectors are reduced to their magnitude. And second, the potential energy is inversely proportional to the distance and not to the distance squared.

In the previous derivation we only considered two charges, but we can easily expand the concept to more charges. Because the Coulomb force is a central force and thus conservative the potential difference is independent of the path taken and thus valid for any two points. We can thus sum the

mutual potential energies of any pair of charges to obtain the total potential energy. For example for 3 charges the equation becomes

$$U = \frac{Q_1 Q_2}{4\pi\epsilon_0 r_{12}} + \frac{Q_2 Q_3}{4\pi\epsilon_0 r_{23}} + \frac{Q_1 Q_3}{4\pi\epsilon_0 r_{13}} \tag{2.7}$$

This can be expanded for any number of charges and used to calculate the potential energy of a collection of charges.

From potential energy the force on a charged particle can in turn be obtained by taking the partial derivative along the spatial direction of interest; i.e.

$$F_x = -\frac{\partial U}{\partial x},\ F_y = -\frac{\partial U}{\partial y},\ F_z = -\frac{\partial U}{\partial z}$$

This is especially useful for situations where the potential energy landscape is given and the force on a charge needs to be determined.

Similar to what was explained for mechanics, the system will always aim to reduce the potential energy. Thus particles with the same charge will aim to attain an infinite separation and particles with opposite charge will cluster together. The system will be in equilibrium if the total force is zero, or in other words if the partial derivatives of the potential energy are all zero. This equilibrium will be a stable one if this is also a minimum of the potential energy landscape.

Chapter 3

Electrostatic field quantities

Although Coulomb's law describes all of electrostatics it becomes very tedious to work with if the systems become larger or more complex. It will be much easier to work with fields to describe the system and predict what will happen. At a certain point it will even become practically impossible to work with anything but field properties. At first the concept of fields will appear very abstract and grasping this concept will be one of the most difficult parts of this course. It might sound contradictory, but the best way to come to understand fields is by not trying to completely understand and grasp it, but to initially just accept it and then let it grow on you. This is actually a useful approach to many abstract physical concepts. Niels Bohr, who is considered as the "father" of quantum mechanics, once famously said that whoever says they understand quantum mechanics shows that he or she clearly doesn't understand it.

3.1 The electric field (E-field)

The electric field, which is often abbreviated to E-field, can be defined as the (Coulomb) force that *would be* exerted on a charge if it is placed at any point relative to other charges or charged objects. Because the force is a vector it means that the E-field is a so-called vector field; it is a vector defined for every point in space. To be able to calculate the E-field of different objects we consider a *positive* test charge Q_t. This test charge is so small that it doesn't have influence on the other charges and therefore its contribution to the electric field is zero. We can now define the E-field as

$$\vec{E} = \frac{F(\vec{Q}_t)}{Q_t}, \tag{3.1}$$

The E-field is thus defined as force per unit positive charge and the units are Newton per Coulomb [NC^{-1}]. However, for reasons that will become clear later, typically the unit Volt per meter [Vm^{-1}] is used.

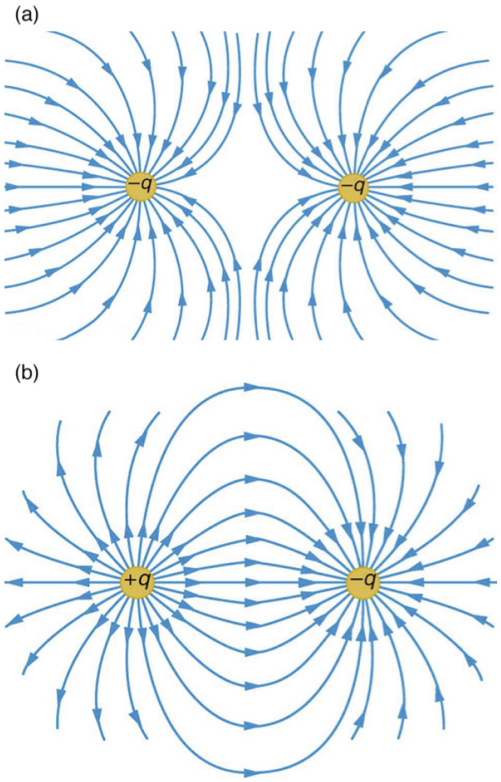

Figure 3.1: Illustration of the E-field by field lines for (a) two negative charges and (b) a positive and negative charge.

We can now use Equation 3.1 to calculate the E-field. The simplest case is the electric field of a point charge Q. From Equation 2.2 we know the Coulomb force of a point charge experienced by a test charge at a distance $\vec{r}$ and can then directly calculate the E-field:

$$\vec{E} = \frac{\vec{F}(Q_t)}{Q_t} = \frac{Q}{4\pi\epsilon_0 r^2}\hat{r} \tag{3.2}$$

Along similar lines we can obtain the E-field for a collection of point charges from the superposition principle in Equation 2.3 which yields:

$$\vec{E} = \frac{Q_1}{4\pi\epsilon_0 r_1^2}\hat{r}_1 + \frac{Q_2}{4\pi\epsilon_0 r_2^2}\hat{r}_2 + ... = \frac{1}{4\pi\epsilon_0}\sum_i \frac{Q_i \hat{r}_i}{r_i^2} \tag{3.3}$$

For a charge distribution ρ of an object τ the E-field can similarly be calculated by integrating over all small contributions to the Coulomb force for a test charge as described in Equation 2.4 and then dividing by this test charge, thus yielding:

$$\vec{E} = \int d\vec{E} = \int_{\tau} \frac{\rho \hat{r} d\tau}{4\pi\epsilon_0 r^2} \tag{3.4}$$

With similar equations for surface or line charge distributions.

An important aspect of the E-field is that it can be graphically expressed by so-called *field lines*. In a primitive sense these can be thought of as all the force vectors placed behind each other to form continuous lines with a direction. More formally the field lines of an E-field originate from charges or charged objects and end at other charged objects or infinity. Even more precisely, the field lines are the path that a positive test charge would follow when allowed to move freely in the vicinity of other charges.

In Figure 3.1 the field lines of two negative charges (a) and a positive and negative charge (b) are shown for a region in the vicinity of the two charges. From this figure we can learn some important aspects of field lines.

- Field lines start at positive charges and end at negative charges.
- The density of field lines is representative for the field strength.
- All the field lines in (a) originate from infinity, and
- All the filed lines in (b) originate from the positive charge and end at the negative charge.

Whereby the last two points are of course only valid if there are no other charged objects around. The picture in Figure 3.1 only shows a two-dimensional cut through the E-field, the field lines expand in three-dimensional space and given the symmetry this can be visualised by rotating the figure around the axis connecting the two charges.

An important example is to calculate the E-field at a distance a along the centre axis of a plane circular sheet with surface charge density σ and radius b as illustrated in Figure 3.2. Before applying Equation 3.4 with brute force it is helpful to first consider the symmetry of the system. For every contribution to the E-field from a region at distance r from the centre there is an equal contribution from a region with $-r$ (i.e. from the opposite half of the disk). All the contributions which are not along the centre axis cancel and only the projection by $\cos\theta$ on the axis has to be considered. Therefore the tedious vectorial summation reduces to a normal integration

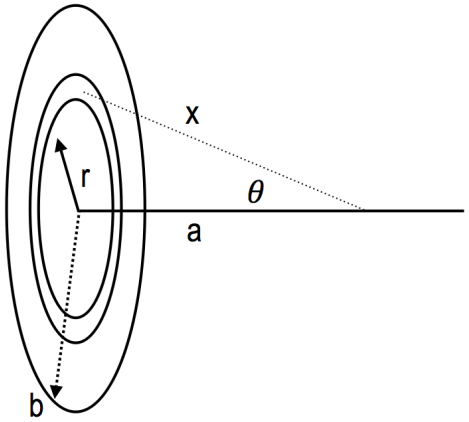

Figure 3.2: Plane circular sheet with surface charge density σ.

as long as we stay on the central axis. A small ring at radius r from the centre and with width dr has a surface area $2\pi r dr$ and contributes to the E-field:

$$dE = \frac{\sigma 2\pi r dr \cos\theta}{4\pi\epsilon_0 x^2}$$

with $\cos\theta = \frac{a}{x}$ this becomes

$$dE = \frac{\sigma a r dr}{2\epsilon_0 x^3}$$

If we no use Pythagoras, $x^2 = a^2 + r^2$, and Eq. 3.4 we obtain

$$E = \frac{\sigma a}{2\epsilon_0} \int_0^b \frac{r dr}{\left(a^2 + r^2\right)^{3/2}}$$

With the consideration that $rdr = \frac{1}{2} d\left(a^2 + r^2\right)$ we can now solve the integral and obtain

$$E = \frac{\sigma a}{2\epsilon_0} \left(\frac{1}{a} - \frac{1}{\sqrt{a^2 + b^2}} \right)$$

The importance of this result is realised when we make the disk very big; $b \to \infty$ or in general $b \gg a$. In this case the E-field becomes

$$E = \frac{\sigma}{2\epsilon_0} \tag{3.5}$$

Note that this is independent of a and because of $b \gg 0$ also valid away from the centre axis (which anyway looses meaning for $b \to \infty$). Equation 3.5 is the expression for a **homogenous E-field**, and we will encounter this expression many times in future chapters.

3.2 Electric potential

In the previous section we defined the electric field as the force that would be exerted on a (test) charge when placed at a certain point in space. Similarly we can define the electric potential as the potential energy a test charge would have if placed at a certain point in space with respect to other charged objects. Formally one can only consider potential energy differences and to be entirely correct we have to name it the electric potential difference, but for simplicity this last part of the name is normally dropped. In practice we typically consider the potential of our planet as zero. In the treatment of many problems here we will set the potential at infinity as zero.

The potential energy is a scalar; i.e. only a magnitude and no direction. The electric potential is therefore a scalar field and not a vector field like the E-field. In everyday life we encounter scalar fields more often as vector fields. In weather forecasts the temperature for different places forms a scalar field map. Often scalar fields are graphically displayed by means of iso-lines connecting points with the same properties. For temperature these are isotherms, and we encounter them commonly as isobars (connecting points with the same air pressure) in overview weather maps. Anybody who ever went to a mountainous area knows the topographic maps with lines that indicate the height above sea level. These lines are iso-altitude lines, but can also be regarded as iso-potentials with regard to the potential energy of the gravitational force of the earth. Also the electric potential, and as we will see later indirectly the electric field, is typically displayed by iso-potential lines which are typically called *equipotential lines or surfaces*.

The electric potential difference Φ between point A and B can be defined as the work done to bring a positive test charge Q_t from point A to B, divided by this test charge

$$\Phi_{AB} = \frac{W_{AB}(Q_t)}{Q_t} \tag{3.6}$$

Based on this definition the units are Joule per Coulomb [JC^{-1}], but typically the unit Volt [V] is used. The Coulomb force is conservative and the potential difference is therefore independent on the path taken, which also means that in calculations we can use any path that suits us best.

To explore the connection between electric potential and electric field it is illustrative to calculate the work done to move a test charge along the path $d\vec{L}$:

$$W_{AB} = \int_A^B \vec{F} \cdot d\vec{L}$$

Here the force is the external force needed and thus opposite to the Coulomb force on a test charge:

$$W_{AB} = \int_A^B -Q_t \vec{E} \cdot d\vec{L}$$

To obtain the potential at B with reference to the potential at A (which we can later define as zero) we can use Eq. 3.6 and get

$$\Phi_{AB} = -\int_A^B \vec{E} \cdot d\vec{L} \tag{3.7}$$

This is a path integral of the E-field along the path $d\vec{L}$. For a path that is always perpendicular to the E-field the scalar product is always zero and all the points along this path are thus at the same potential. We thus see that equipotential lines and surfaces are perpendicular to the E-field. For any other path we have to integrate the local projection of the E-field on this path. Before considering an example it should be remarked that the potential energy of a charge Q now can simply be obtained from

$$U = Q\Phi \tag{3.8}$$

Whereby this potential energy is with respect to where we defined the zero of the potential. *The potential (and also the E-field) is a property of the source charge, whereas the potential energy (and Coulomb force) depends on both charges.*

From Equation 3.2 we know the E-field of a point charge and we can then use Eq. 3.7 to calculate the electric potential at a point P at a distance r_P from a point charge, with respect to the potential at infinity (which we can set to zero):

$$\Phi_P = -\int_\infty^P \frac{Q\hat{r} \cdot d\vec{L}}{4\pi\epsilon_0 r^2}$$

If θ is the angle between $d\vec{L}$ and $\hat{r}$ we find that

$$\hat{r} \cdot d\vec{L} = dL\cos\theta = dr$$

which means we project our $d\vec{L}$ on $\hat{r}$ and call this dr. This is valid for any path but easiest to imagine if we follow a radial path from infinity. Note that there is no path coming from infinity that is always perpendicular to $\hat{r}$. With r_p the distance from the point charge to P we can now rewrite and solve the integral to obtain the **electric potential of a point charge**:

$$\Phi_P = -\frac{Q}{4\pi\epsilon_0}\int_\infty^{r_p} \frac{dr}{r^2} = \frac{Q}{4\pi\epsilon_0 r_p} \tag{3.9}$$

The potential difference between two points at distances r_1 and r_2 from a point charge then becomes

$$\Phi_1 - \Phi_2 = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

The electric potential changes sign when considering a negative point charge instead of a positive one and it asymptotically approaches zero as a function of distance. All points at the same distance (radius) from the point charge have the same potential and the equipotential surfaces thus are spheres centred at the point charge as illustrated in Figure 3.3(a).

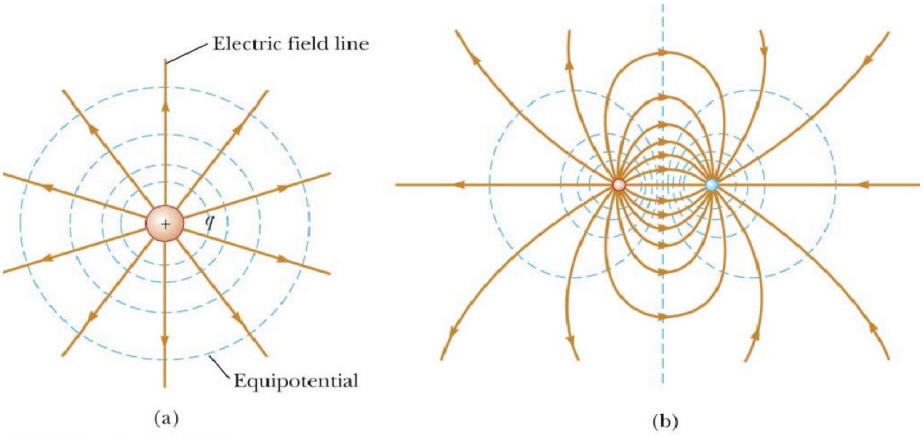

Figure 3.3: Equipotential lines and E-field of (a) a single point charge and (b) two point charges. Note that the E-field lines are always perpendicular to the equipotential lines.

From the superposition principle and Equation 3.3 we can now derive the **electric potential for a collection of point charges**:

$$\Phi = \frac{Q_1}{4\pi\epsilon_0 r_1} + \frac{Q_2}{4\pi\epsilon_0 r_2} + ... = \frac{1}{4\pi\epsilon_0}\sum_i \frac{Q_i}{r_i} \tag{3.10}$$

And along similar lines we can determine that the **electric potential of a continuous charge distribution** is

$$\Phi = \int_\tau \frac{\rho d\tau}{4\pi\epsilon_0 r} \tag{3.11}$$

Whereby again it has to be taken into account that r changes with the integration, and similar expressions can be found for surface and line charge distributions. In both cases the zero of the potential is set at infinity.

Now let's return to the example of the charged disk at the end of the previous section. Using similar arguments as we did there we can determine the contributions of rings of charge to the potential on the axis as

$$d\Phi = \frac{2\pi\sigma r dr}{4\pi\epsilon_0 x}$$

After making similar substitutions the integral can be solved and we obtain

$$\Phi = \frac{\sigma}{2\epsilon_0}\left(\sqrt{a^2+b^2} - a\right)$$

along a. In the previous section we could obtain an expression for a uniform E-field by setting b to infinity, however, if try the same thing here then the potential Φ also becomes infinite. The reason for this is that we actually defined $\Phi = 0$ at infinity and thus we can't use the same expression anymore for $b \to \infty$. In this case it is easier to directly apply Equation 3.7 to the uniform E-field we know results from increasing b, as expressed in Equation 3.5. The only relevant coordinate is a and the term $\vec{E} \cdot d\vec{L}$ reduces to Eda and the integral is easily solved to obtain the **potential for an uniform field**:

$$\Phi = \int -Eda = C - Ea \tag{3.12}$$

Here C is the potential of the charged disk. This result also explains why the E-field is expressed in [Vm^{-1}].

If the electric potential is given it is also possible to use this to obtain the E-field. Reciprocal to the previous description the E-field is obtained by taking the vectorial partial derivative of the potential:

$$\vec{E} = -\nabla\Phi \tag{3.13}$$

whereby ∇ is defined as in Eq. 5. Written out for cartesian coordinates this becomes:

$$\vec{E} = \left(-\frac{\partial\Phi}{\partial x}, -\frac{\partial\Phi}{\partial y}, -\frac{\partial\Phi}{\partial z}\right)$$

From this we see again that the E-field is always perpendicular to equipotential lines.

3.3 Charges in E-fields

We have defined the E-field as the force that would be exerted on a test charge, now it is time to look at what happens if we place an actual charge

in an E-field. An important assumption or boundary condition that we will assume is valid for all cases from now on, is that the charge has no influence on the E-field. This can either be the case if the charge is very small or if the E-field is kept constant by external sources.

The force that a charge Q in an E-field experiences is

$$\vec{F} = Q\vec{E} \tag{3.14}$$

which is according to Newton equal to mass times acceleration:

$$\vec{F} = Q\vec{E} = m\vec{a}.$$

For ions and smaller particles the gravitational force is more than 10^9 times smaller as the electrical force and the former can thus typically be neglected. Instead of the E-field we can of course also consider a potential difference in which case the expression becomes:

$$Q\frac{\Delta\Phi}{x} = ma$$

in which case the acceleration is oriented along the direction the potential difference occurs.

The main point to be learned from this is that an E-field can be used to accelerate charged particles along the field lines, or perpendicular to the equipotential surfaces. It also allows us to use a new energy scale, the electronvolt [eV]. This is the energy gained by an electron when passing a potential difference of 1 V. This makes 1 eV $= 1.6 \times 10^{-19}$ J.

For sake of simplicity we will in this lecture only consider two possible cases and both are for uniform, and stationary, E-fields: 1) the initial velocity of the charge is zero or parallel to the field lines ($\vec{v}_i || \vec{E}$), and 2) the initial velocity is perpendicular to the field lines ($\vec{v}_i \perp \vec{E}$). In the first case the particle is accelerated along the direction of the initial velocity. All the gained energy is transformed in kinetic energy (if we are far from relativistic velocities) and we get

$$Q\Delta\Phi = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$$

which can be used to calculate the final velocity v_f. Thus the particle is simply accelerated or slowed down along the initial direction of travel. This is a very important aspect and is used in a variety of applications. In mechanics this can be compared to throwing an object straight up or down, or letting it drop from a given height.

If the initial velocity is perpendicular to the E-field, the charged particle will not experience any force along the direction of travel. The only acceleration is along the perpendicular direction and is given by:

$$a_\perp = \frac{QE}{m}.$$

If the E-field has a length l along the initial velocity direction, the charge takes a time $t = \frac{l}{v_i}$ to travel through the field. In this time the velocity along the perpendicular direction becomes

$$v_\perp = a_\perp t = \frac{QEl}{mv_i}.$$

By travelling through the field the charge is deflected and the angle of deflection α can be obtained from

$$\tan\alpha = \frac{v_\perp}{v_i} = \frac{QEl}{mv_i^2}.$$

Such a deflector can be used to steer a beam of charged particles to a target and is used in many applications. In mechanics one can compare it to the situation of throwing an object horizontally from a cliff.

In the more general case of where the velocity is neither parallel nor perpendicular to the E-field the problem can be relatively straightforward solved by separating it into a parallel and perpendicular part. However, when the E-field is not uniform, both the velocity of the charge and its direction are changed. In this case beams of charged particles can be collimated or focussed, or manipulated in any required way. This is the field of electron optics which is not part of this lecture, but students should be aware that such possibilities exist.

3.4 Electric dipole

Some highly symmetric charge arrangements occur regularly in treatments and are thus worth to consider separately. The simplest, and most important, of such arrangements is the electric dipole. This consists of two equal but opposite charges $+q$ and $-q$ fixed at a distance a from each other as illustrated in Figure 3.4. The total charge of the dipole is zero. Furthermore we only consider the **ideal dipole** where the distance a is small compared to other distances.

In this case we can use a trick to calculate the electric potential of the whole dipole. In Figure 3.4 the potential at point P having polar

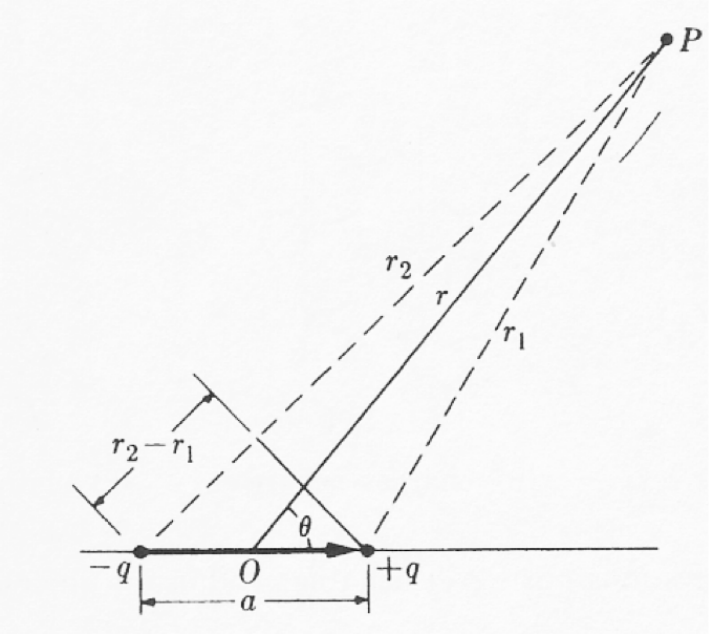

Figure 3.4: Electric dipole and arrangement to calculate its potential.

coordinates (r, θ) due to $+q$ is given by the electric potential of a point charge (Eq. 3.9) as

$$\Phi_P(+q): \ \Phi_q = \frac{q}{4\pi\epsilon_0 r_1} = \frac{q}{4\pi\epsilon_0 r}$$

whereby $r_1 \approx r_2 \approx r$ because $r \gg a$. The potential due to $-q$ is almost equal but opposite and the small difference is due to the fact that $-q$ is displaced by a distance a along the x-axis

$$\Phi_P(-q): \ \Phi_{-q} = -\Phi_q - d\Phi_q.$$

The total potential at point P then becomes

$$\Phi_P = -d\Phi_q = -a\frac{\partial\left(q/4\pi\epsilon_0 r\right)}{\partial x}.$$

Which can be rewritten and solved using that $x = r\cos\theta$ and $\frac{\partial r}{\partial x} = \frac{x}{r} = \cos\theta$ as

$$\Phi_P = \frac{-qa}{4\pi\epsilon_0}\frac{\partial\left(\frac{1}{r}\right)}{\partial x} = \frac{qa}{4\pi\epsilon_0 r^2}\frac{\partial r}{\partial x} = \frac{qa\cos\theta}{4\pi\epsilon_0 r^2}.$$

This can be rewritten in a more general way by introducing the **electric dipole moment** defined as:

$$\vec{p} = q\vec{a} \tag{3.15}$$

This is a vector quantity pointing in the direction from $-q$ to $+q$. Using the scalar product $\vec{p} \cdot \vec{r} = pr\cos\theta$ we can now obtain an expression for the **potential of an electric dipole**

$$\Phi_P = \frac{\vec{p} \cdot \vec{r}}{4\pi\epsilon_0 r^3} = \frac{\vec{p} \cdot \hat{r}}{4\pi\epsilon_0 r^2} \tag{3.16}$$

It should be noted that whereas for a point charge (or electric monopole) the potential is inversely proportional to r, for the dipole it is inversely proportional to r^2. We will later see that this can be expanded for higher order multipoles. Another important point is that the potential is zero for $\vec{p} \cdot \vec{r} = 0$ which is the case for the plane exactly half way between the two charges.

We can now apply Equation 3.13 to Eq. 3.16 and use the definition of ∇ in polar coordinates in Eq. 6 to determine the **E-field of an electric dipole**:

$$E_\theta = -\frac{1}{r}\frac{\partial \Phi}{\partial \theta} = \frac{p\sin\theta}{4\pi\epsilon_0 r^3} \tag{3.17}$$

$$E_r = -\frac{\partial \Phi}{\partial r} = \frac{2p\cos\theta}{4\pi\epsilon_0 r^3} \tag{3.18}$$

When plotted in two dimensions this E-field looks like the one shown in Figure 3.3(b).

In a uniform E-field the forces on both charges are equal but opposite and the *total force on the dipole is thus zero.*

$$\sum \vec{F} = 0$$

In a non-uniform E-field this is generally not the case and the dipole will experience a force proportional to the gradient of the field. Taking into account the changes of the E-field over the length of the dipole, it is possible to derive the total force on the dipole. For example for the x-component this gives

$$F_x = p_x\frac{\partial E_x}{\partial x} + p_y\frac{\partial E_x}{\partial y} + p_z\frac{\partial E_x}{\partial z} \tag{3.19}$$

With similar expressions for the y and z components. For the total force this can be rewritten as

$$\vec{F} = (\vec{p} \cdot \nabla)\vec{E}. \tag{3.20}$$

In a uniform E-field the forces on each of the charges are equal but opposite, but they act on different ends of the dipole. This induces a **torque** which will align the dipole to the E-field in such a way that the dipole moment is aligned parallel to the field. The torque $\vec{\tau}$ on a dipole in an electric field can be expressed as

$$\vec{\tau} = \vec{p} \times \vec{E} \tag{3.21}$$

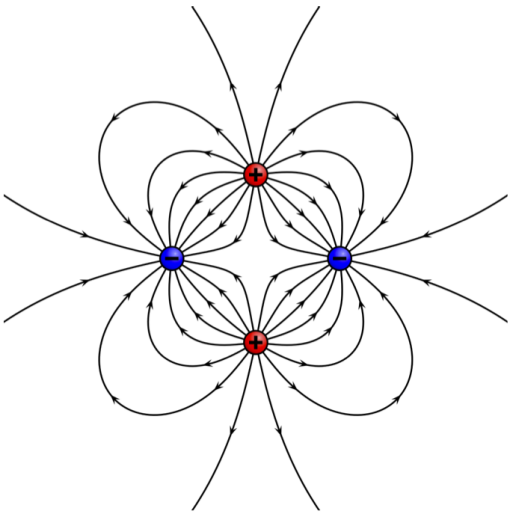

Figure 3.5: E-field of a planer quadrupole.

(Dipole approximation for arbitrary charge distribution: to be included)

Besides the electric dipole we can also define **higher order multipoles**. Just like in a dipole the sum of the monopoles (charges) is zero and we have a dipole moment $\vec{p}$, for a *quadrupole* the sum of both the monopoles and dipole moments equals zero

$$\sum Q = 0, \ \sum \vec{p} = 0.$$

In this case we can define the quadrupole moment $\mathbf{q}$ which is a tensor. For a quadrupole the potential is proportional to $\frac{\mathbf{q}}{r^3}$ and the E-field to $\frac{\mathbf{q}}{r^4}$. The latter is shown for a planar quadrupole in Figure 3.5.

Expanding on this we can define the *octupole* where the sum of the monopoles, dipole moments, and quadrupole moments is zero. In this case the potential is proportional to $\frac{1}{r^4}$ and the E-field to $\frac{1}{r^5}$.

Chapter 4

Gauss's law and its consequences

Carl Friedrich Gauss (1777-1855) is considered to be the greatest mathematician of modern times. Much of modern science, and thus also society, relies heavily on the mathematical concepts he introduced. In this chapter Gauss's law on electrostatics will be derived, showing how the flux of the E-field through any closed surface is directly related to the enclosed charge. It will be shown that this is a powerful tool to find the E-field of objects. In a related sense it will be shown that for static electric fields no closed field lines can exist, which is the so-called circuital law.

There are a variety of important consequences of Gauss's law and the circuital law, which will be explained in detail below. The correct implementation of these laws and especially a thorough consideration of their consequences will help get a more intuitive feeling of how E-fields behave and how situations that initially appear very complex can be greatly simplified. It will even help us understand how Rutherford could determine the size of the atom and how electric eels probe their environment.

4.1 Derivation of Gauss's law

A central ingredient of Gauss's law is the **flux of a vector field through a surface**. It is relatively straightforward to consider the flux, or in this case flow, of water in a river through some surface. Also in Section 1.2 we considered the flux of charge through some surface to define the current, or the current density. In both cases we can still picture something moving through a surface and we can imagine somehow counting the amount of water or charge that passes to get an idea of the flux. However, we can also

consider the flux of a (static) vector field through a surface. In the previous chapter we looked at the E-field as the force that would be exerted on a test charge. Similarly we can think of the flux of the E-field through a surface as the amount of such test charges that would be forced through it, or as a type of current that would flow. This is not a formal definition, but just meant to help make the step from actual things moving to the flux of a field easier.

The formal definition of the flux of a vector field, in this case the E-field, φ through a surface area dA is

$$\varphi_{dA} = \vec{E} \cdot d\vec{A}. \tag{4.1}$$

Here the vector $d\vec{A}$ is again the surface normal whereby the magnitude represents the size of the area and the direction the orientation of the surface in space. This means that the flux is maximal when the surface normal is parallel to the field and the surface itself thus perpendicular. The flux goes to zero when the surface normal and field are perpendicular. In this case no field lines pass through the surface. Because for a closed surface the surface normal always points out of the object, the flux is positive when the field points outward, and negative when it points inward.

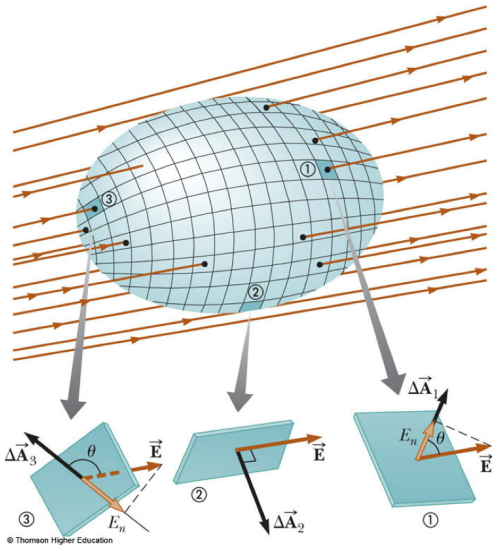

Figure 4.1: Flux of a vector field through a surface.

The total flux of a vector field through any surface A can now be ob-

tained by integrating over that surface

$$\varphi_{tot} = \int_A d\varphi = \int_A \vec{E} \cdot d\vec{A}. \tag{4.2}$$

In Figure 4.1 the above definition of flux of a vector field is illustrated and summarised.

As a first example for the flux of an electrostatic field, we can now consider the flux of the electric field generated by a point charge, through an enclosing sphere with radius r. We choose the sphere to be centred at the point charge and from symmetry we know that the E-field is now always parallel to the surface normal and Eq. 4.1 thus reduces to a normal multiplication. From Eq. 3.2 we know the E-field of a point charge and further we know that the surface area of a sphere is $4\pi r^2$. Using Eq. 4.2 we obtain:

$$\varphi = \int_A E dA = \int_A \frac{Q}{4\pi\epsilon_0 r^2} dA = \frac{Q}{4\pi\epsilon_0 r^2} 4\pi r^2 = \frac{Q}{\epsilon_0}$$

That this rather simple result is not an exception due to the choice of enclosing surface, but a general rule will be the main result of Gauss's law.

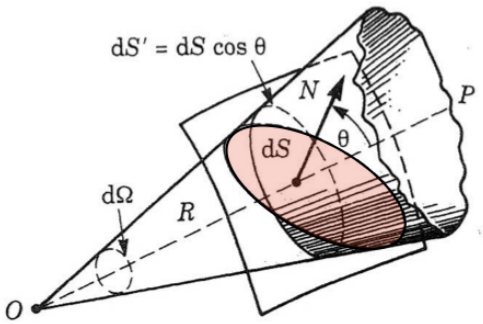

Figure 4.2: Illustration of solid angle $d\Omega$ for a surface element $d\vec{S}$

We consider a randomly shaped *closed* surface S enclosing a charge Q. For simplicity only a single charge is considered, but the argument also holds for a collection of charges or a charge distribution. The flux through a surface element dS can be obtained from Eq. 4.1:

$$\varphi_{dS} = \vec{E} \cdot d\vec{S} = \frac{Q\hat{r} \cdot d\vec{S}}{4\pi\epsilon_0 r^2} = \frac{Q dS \cos\theta}{4\pi\epsilon_0 r^2} \tag{4.3}$$

Here θ is the angle between the surface normal and the radial vector from the charge $\hat{r}$. We can now use the concept of the solid angle $d\Omega$ to simplify this expression. The solid angle is illustrated in Figure 4.2 and can be viewed as a dimensionless area enclosed by a cone. It is accordingly defined as

$$d\Omega = \frac{dS\cos\theta}{r^2} \tag{4.4}$$

Of importance to our derivation is that the integral of the solid angle over any closed surface (hence the $\oiint$) becomes independent of the shape of that surface and always yields

$$\oiint d\Omega = 4\pi \tag{4.5}$$

We can now use Eq. 4.4 and 4.5 to further simplify 4.3:

$$\varphi_{dS} = \frac{Qd\Omega}{4\pi\epsilon_0}$$

and from integrating it follows that:

$$\varphi = \oiint \frac{Qd\Omega}{4\pi\epsilon_0} = \frac{Q}{\epsilon_0}$$

We can now summarise this and apply the superposition theorem to obtain **Gauss's law**:

$$\oiint \vec{E}\cdot d\vec{S} = \sum_i \frac{Q_i}{\epsilon_0} \tag{4.6}$$

Often the summation will be implicitly assumed. Expressed in words, this means that **the flux of an E-field over any closed surface is equal to the total enclosed charge divided by ϵ_0.**

This statement is valid regardless of the chosen surface as long as it is closed, which is often called a *Gaussian surface*. This not only includes nice geometric shapes, but also irregular shapes such as a human body or car. This means we can choose the closed surface that fits our problem best at will. Another consequence is that, if no charge is enclosed, the total flux passing through the surface will be zero; as much flux enters as leaves.

Equation 4.6 is the so-called *integral form* of Gauss's law. Because the closed surface always has some spatial extension, it is also always applied in an extended fashion. Sometimes it is easier to have a local expression of the law and this can be obtained from **Gauss's divergence theorem**. The

derivation will be shown during the lecture, but it comes down to applying Eq. 4.6 to an infinitesimal small cube with charge density ρ. This then yields

$$\left(\frac{\partial E}{\partial x} + \frac{\partial E}{\partial y} + \frac{\partial E}{\partial z}\right) = \frac{\rho}{\epsilon_0}$$

Using the nabla vector ∇ we obtain the *differential form* of Gauss's law:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{4.7}$$

Which is the same as the first Maxwell equation in the introduction.

4.2 Using Gauss's law

Before showing how Gauss's law can be used to find E-fields it is useful to consider some of its general consequences.

All excess charge is located on surface of a conductor: As mentioned before, there is no E-field inside a conductor. If there would be an E-field, this would move the free charge which in turn would compensate the field. If the field is zero, then its flux is also zero. This means that for any Gaussian surface chosen within the conductor the left part of Eq. 4.6 is zero and thus also the total charge enclosed by this surface is zero. Consequentially all the excess charge is located on the surface of the conductor.

Charge inside hollow conductor is screened: We consider a conductor with some cavity in which a charge is placed. Again we know that inside the conductor the E-field is zero. If we choose a Gaussian surface inside the conductor enclosing the cavity the total charge has to be zero. Therefore an equal amount of charge is now located on the inner surface of the conductor to screen the charge.

No E-field exists inside a hollow conductor: We consider the same conductor with a cavity, but without a charge placed inside. If we now assume an E-field can exist inside this cavity, then there are regions with higher and lower potential. If we choose a Gaussian surface enclosing one of these regions then there is a flux across this surface, which according to Gauss's law would mean there is a charge inside. But we know there is no charge, so the assumption of an E-field leads to a contradiction. This absence of a E-field inside a hollow conductor is often referred to as electrostatic shielding or screening, or as the *Faraday cage*.

Besides these general consequences the main use of Gauss's law is to **find E-fields** for a given charge configuration. The strategy is in all cases the same and can be summarised as follows:

- Choose a smart and easy closed surface based on the symmetry of the system under consideration
- Write down expression or solve integral for flux through Gaussian surface and charge contained inside
- Simplify to obtain an expression $E = ...$
- Integrate E to obtain the potential if required

During the lecture and exercises many examples will be given and they will also be included here when time permits.

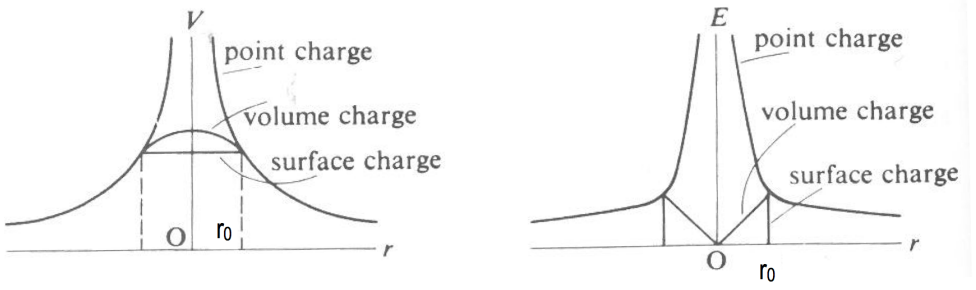

Figure 4.3: Potential and E-field as a function of distance away from the centre for different types of spherical systems. (From Duffin ©McGraw-Hill)

For a charged or conducting sphere we see that the electrostatic potential is inversely proportional to the radius of the sphere for the same amount of charge. This is also illustrated in the summary of potential and electric field behaviour as a function of distance from the centre of a sphere for different objects displayed in Figure 4.3. This means that if we know its charge we can *determine the size of an object from its potential*. This effect was used by Rutherford to determine the size of the nucleus of an atom by looking at the backscattering of α-particles from a gold foil.

We can also consider what happens if we keep the potential the same, but vary the radius of an object. To illustrate this, we look at the E-field and potential just outside a sphere with radius r_0 and surface charge density

$\sigma = Q/4\pi r_0^2$. From Gauss's law we derived that the E-field is given by

$$E_s = \frac{Q}{4\pi\epsilon_0 r_0^2} = \frac{\sigma}{\epsilon_0}$$

and the potential by

$$\Phi_s = \frac{Q}{4\pi\epsilon_0 r_0} = \frac{r_0\sigma}{\epsilon_0}.$$

If we keep the potential fixed and regard the radius as a variable this means that the charge density and local E-field show the following dependence on the radius:

$$\sigma = \frac{\epsilon_0\Phi}{r_0} \tag{4.8}$$

$$E_s = \frac{\Phi}{r_0} \tag{4.9}$$

It should be realised that these equations are only valid under the assumptions expressed above and should only be regarded as locally valid. These expressions mean that both the surface charge density and the E-field just outside the object increase if the radius decreases, or if the curvature increases. This is referred to as the **point effect** because, for a sharp point, the E-field will be the highest, as also illustrated in Figure 4.4. Because at these sharp points the field is the largest it will also be the first point where the air will be ionised and as a consequence a so-called *corona discharge* will occur.

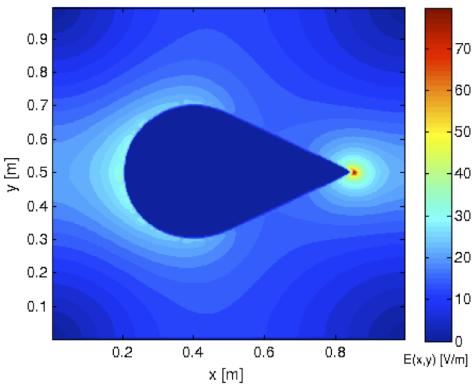

Figure 4.4: Electric field just outside an object with different local curvature.

4.3 The circuital law

The circuital law in a simple picture deals with the question of whether there exist closed E-field lines. To answer this we consider the integral of the E-field along a closed loop L:

$$\oint_L \vec{E} \cdot d\vec{L} \tag{4.10}$$

In Figure 4.5 such a possible loop is indicated which we can always separate into two paths I and II between points A and B. The integral $\int \vec{E} \cdot d\vec{L}$ between A and B is the potential difference between those points. From the fact that the Coulomb force is conservative we know that this potential difference is independent of the path taken and furthermore that the sign of the integral reverses if we reverse the path:

$$\int_{A,I}^{B} \vec{E} \cdot d\vec{L} = \int_{A,II}^{B} \vec{E} \cdot d\vec{L}$$
$$\text{and} \int_{B,II}^{A} \vec{E} \cdot d\vec{L} = -\int_{A,I}^{B} \vec{E} \cdot d\vec{L}$$

If we add all this together we obtain

$$\oint \vec{E} \cdot d\vec{L} = \int_{A,I}^{B} \vec{E} \cdot d\vec{L} + \int_{B,II}^{A} \vec{E} \cdot d\vec{L} = 0$$

Or summarised in the circuital law:

$$\oint \vec{E} \cdot d\vec{L} = 0 \tag{4.11}$$

Equation 4.11 represents the integral form of the circuital law. It means that no closed E-field lines exist for electrostatic fields and that, as we will see later, a power source is needed to create a current. In later chapters, we will see examples going beyond electrostatics where the right hand side of this equation is not equal to zero.

Also the circuital law can be expressed in a local differential form. The derivation of this goes beyond the scope of this course and relies of Stoke's theorem in mathematics. Here only the result of this is relevant and the circuital law can also be expressed as:

$$\nabla \times \vec{E} = 0 \tag{4.12}$$

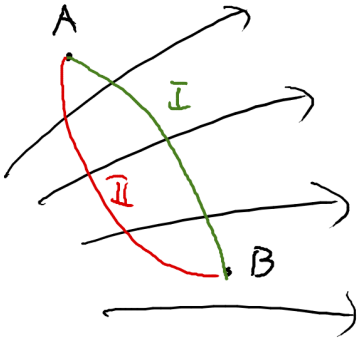

Figure 4.5: A general closed path through an E-field (arrows).

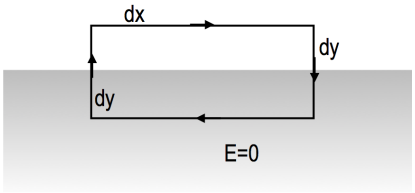

Figure 4.6: Surface of conductor with closed path

Although the impact at this point of the circuital law appears to be limited, the fact that we can treat it as a general law has some further consequences. Let's consider the rectangular closed loop around the surface of a conductor in Figure 4.6. If we decompose the E-field into components perpendicular ($E_\perp$) and parallel to the surface ($E_\parallel$) and realise that the E-field inside the conductor is zero, we can decompose Eq. 4.11 in parts as follows starting at the top right

$$-E_\perp dy + 0 + E_\perp dy + E_\parallel dx = 0$$

From this it directly follows that $E_\parallel = 0$ and that **E-field lines are always perpendicular to the surface of a conductor.** This means that, if a conductor is placed in an electric field, the field lines will be bent in such a way as to ensure they connect to the conducting surface perpendicularly. This effect is of obvious importance when determining how an E-field will look, but it is also used by some animals to determine the conductivity, and thus edibility, of objects close by. The most famous example of this is the electric eel who uses the change in shape of the field lines to determine

whether something is a fish and to observe how it moves without having to identify it visually.

4.4 Uniqueness theorem

We can combine the differential form of Gauss's law (4.7) and the relationship between E-field and potential (Eq. 3.13) to obtain

$$\nabla \cdot \nabla\Phi = -\frac{\rho}{\epsilon_0}$$

Because Φ is a scalar field this can be rewritten as

$$\nabla^2\Phi = -\frac{\rho}{\epsilon_0} \tag{4.13}$$

This is known as **Poisson's equation** in the presence of charge, and as the **Laplace equation** if no charge is present; i.e.

$$\nabla^2\Phi = 0 \tag{4.14}$$

with

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \tag{4.15}$$

Although Eq. 4.13 and 4.14 look simple they can form an extremely complex set of differential equations and an unique solution can only be found if the boundary conditions are specified.

The students in this course will not be required to solve Poisson's or the Laplace equation, but are expected to be aware of their use. Here we will use those equations to derive a very powerful tool for the determination of electric fields; the **uniqueness theorem**.

Given is a certain charge configuration with well defined boundary conditions. Let's now assume that there are two possible solutions for the potential (and E-field) Φ_1 and Φ_2. Then we can define their difference as $\Phi_3 = \Phi_1 - \Phi_2$. Because both Φ_1 and Φ_2 are solutions we know from Eq. 4.13 that:

$$\nabla^2\Phi_1 = -\frac{\rho}{\epsilon_0} \text{ and } \nabla^2\Phi_2 = -\frac{\rho}{\epsilon_0}$$

From their difference it directly follows that

$$\nabla^2\Phi_3 = 0$$

Which is the Laplace equation for Φ_3. In this equation there is no charge and thus no sources or sinks, so the maxima and minima can only occur at the boundary. However, at the boundary, the boundary conditions are the same for Φ_1 and Φ_2 meaning that

$$\Phi_1 = \Phi_2 \ \rightarrow \ \Phi_3 = 0 \text{ (at boundary)}$$

This implies that both the maximum and the minimum of Φ_3 are zero and that Φ_3 is thus zero everywhere. Consequently this means that Φ_1 and Φ_2 are equal everywhere and no two different solutions can exist.

The fact that no two solutions of the potential and E-field are possible for a given charge distribution and boundary conditions is known as the uniqueness theorem. It can be formulated as follows: *If a solution can be found, then it is the only correct one.* The main implication is that in many cases we don't have to calculate the E-field, because we can "replace" the system by a charge configuration with the "same" boundary conditions that we do know the solution to.

This still sounds rather complex, so let's consider an example. Given is a positive point charge $+q$ placed at a distance d in front of a grounded plane conductor, giving the boundary condition $\Phi = 0$ at this surface. Even for this simple system, solving Eq. 4.13 is already rather complex, so instead we consider a system we know with the "same" charge distribution and boundary condition. This would be two opposite but equal charges $+q$ and $-q$ placed at a distance $2d$ from each other. We know that half way between the charges there is a plane where $\Phi = 0$ creating the same boundary condition. If we now only consider the left half of Figure 4.7, we have found a solution to the posed problem. According to the uniqueness theorem we also know that this is the only possible correct solution. This way of solving a problem is known as the **image charge solution**.

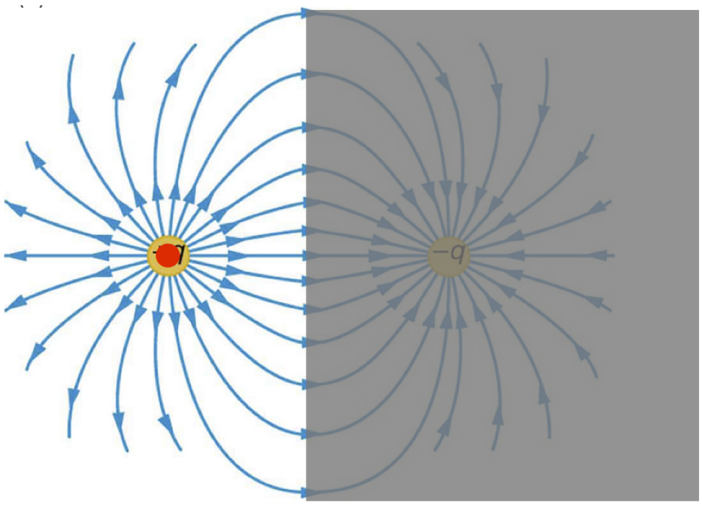

Figure 4.7: Image charge solution to the problem of a charge in front of a plane conductor.

Chapter 5

Capacitance and dielectrics

In the previous chapters we developed the terminology and methods to determine the electric field and potential for model systems. In this chapter we will make the step toward real systems and applications. The first part will be concerned with capacitance and capacitors, which are important components for a wide range of applications. Furthermore, the use of vacuum as a medium is, of course, highly unrealistic compared to the real situation. In the second part about dielectrics it will be shown how non-metallic materials influence the E-field and that, with only small corrections, we can continue using the laws which were derived for vacuum.

5.1 Capacitance of a conductor

We have seen that the electrostatic potential of an object is changed if we place charge on it. This means that, for a given object, there is an intrinsic relationship between its potential and charge. In other words, there is a certain amount of charge Q a conductor can hold for a given potential Φ. This "capacity for charge" is called the **capacitance** (C) of the object and it is defined as

$$C = \frac{Q}{\Phi} \tag{5.1}$$

The units of capacitance are coulomb per volt [CV^{-1}] or farad [F]. Because the coulomb is a large unit compared to everyday life, also the farad is large. Typically values for capacitance are in the pico- to micro-farad range. By definition capacitance is always positive.

Using the uniqueness theorem, it can be shown that *capacitance is constant, is independent of given voltages or charge densities, and only depends*

on the geometry. The proof itself is not of importance and will be omitted here. The result, however, is of utmost importance and will be used implicitly throughout this lecture.

We will consider the capacitance of a conducting sphere as an example. From the previous chapter we know that the potential of a sphere depends on charge Q and radius r_0 as follows

$$\Phi = \frac{Q}{4\pi\epsilon_0 r_0}$$

Using Eq. 5.1 we can now calculate the *capacitance of a sphere*:

$$C = 4\pi\epsilon_0 r_0 \tag{5.2}$$

If we now consider a sphere of radius $r_0 = 10$ mm to obtain a capacitance of $C \approx 10^{-12}$ F or 1 picofarad. Using the same equation we can also calculate the capacitance of the earth ($r_0 = 6.3 \times 10^6$ m) and obtain $C \approx 7 \times 10^{-4}$ F.

5.2 Capacitors

In the previous section it was ignored where the charge on the object came from or what the environment looked like. In principle it was assumed that there was only a single object surrounded by vacuum. This is obviously not a realistic situation and, instead of the capacitance of a single object, we will here consider the capacitance of a combination of objects. Together these are termed capacitors.

An **ideal capacitor** is formed by two charged objects (or more, but we restrict ourselves to the simple case) whereby all the field lines from the positively charged object (A) end at the negatively charged object (B). This means that if we now draw a closed surface enclosing both objects the total flux through this surface is zero. According to Gauss's law this means that the total charge of the two objects is also zero. Therefore the situation can be regarded as if the amount of charge Q has moved from object B to object A. Due to this charge transfer, both objects are at a different potential with a potential difference $\Delta\Phi = \Phi_A - \Phi_B$. We can now define the capacitance of this ideal capacitor as

$$C = \frac{Q}{\Phi_A - \Phi_B} \tag{5.3}$$

Which is again by definition positive and if needed the absolute value of the potential difference can be used because the transferred charge Q is always positive.

The name might suggest that one will almost never encounter such an ideal capacitor, but this is luckily not the case. Most real capacitors are extremely close to being ideal capacitors and any deviations can typically be neglected. This allows us to apply the simple description derived above in a variety of cases. In all examples in this lecture it will thus be assumed that the negative charge on one object is the same as the positive charge on the other and it will just be called Q.

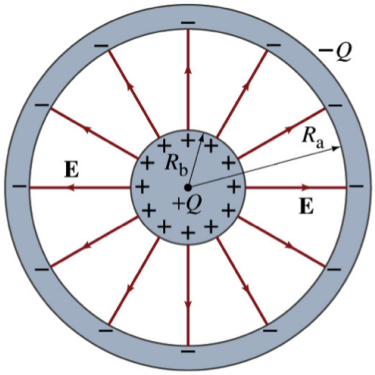

Figure 5.1: Illustration of a spherical capacitor or cut through of a cylindrical capacitor.

The most obvious example of a (ideal) capacitor is the **spherical capacitor** shown in Figure 5.1. This consist of two concentric spherical electrodes with radius R_a and R_b as illustrated in the figure. All the field lines point radially from the inner electrode to the outer one. The voltage difference between the two electrodes can be calculated from their respective potentials and the fact that both have the same but opposite charge:

$$\Delta\Phi = \frac{Q}{4\pi\epsilon_0 R_b} - \frac{Q}{4\pi\epsilon_0 R_a} = \frac{Q\,(R_a - R_b)}{4\pi\epsilon_0 R_a R_b}$$

From Eq. 5.3 we can now calculate the capacitance and obtain

$$C = 4\pi\epsilon_0 \frac{R_a R_b}{(R_a - R_b)} \tag{5.4}$$

It is easy to picture why a spherical capacitor is not the first choice in any applications; the manufacturing is difficult and connecting the electrodes even more so. More commonly used is the **parallel plate capacitor**, which consists of two charged plates of surface area A located at a distance

d from each other. If $d \ll A$ this can be considered as an ideal capacitor and the capacitance can be obtained as follows. The surface charge density on the electrodes is $\sigma = \frac{Q}{A}$ and we know from the examples in the lecture that in this case the homogeneous E-field of an "infinite" plane is

$$E = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A}$$

From this we can obtain the potential difference as

$$\Delta\Phi = Ed = \frac{Qd}{\epsilon_0 A}$$

and the capacitance thus becomes according to 5.3

$$C = \epsilon_0 \frac{A}{d} \tag{5.5}$$

A further important example is the **cylindrical capacitor**, which forms the basis for the co-axial cable used for signal transmission. We again refer to Figure 5.1, but now consider it as the cross-section through a cylinder with length $l \gg R_a$ running perpendicular to the paper. The line charge density now becomes $\lambda = \frac{Q}{l}$. From the lectures related to previous chapter we know the E-field is thus

$$E = \frac{\lambda}{2\pi\epsilon_0 r} = \frac{Q}{2\pi\epsilon_0 r l}$$

To obtain the potential difference we integrate the E-field between the core and outer cylinder:

$$\Delta\Phi = -\int E dr = -\int_{R_a}^{R_b} \frac{Q dr}{2\pi\epsilon_0 r l} = \frac{Q}{2\pi\epsilon_0 l} \ln \frac{R_a}{R_b}$$

From Eq. 5.3 we can now calculate the capacitance

$$C = \frac{2\pi\epsilon_0 l}{\ln \frac{R_a}{R_b}} \tag{5.6}$$

In all cases the capacitance only depends on the geometry and objects can be scaled (can be made larger or smaller) without loss of generality, as long as the assumptions are met. Furthermore, to increase the surface area, and thus capacitance, of a parallel plate capacitor it can even be rolled up in a spiral, without significantly changing its properties.

The properties of **combinations of capacitors** are presented here without derivation. The background will be discussed during the lecture. For capacitors *in series* the total capacitance C' becomes

$$\frac{1}{C'} = \sum_i \frac{1}{C_i} \tag{5.7}$$

And for capacitors *in parallel* it becomes

$$C' = \sum_i C_i \tag{5.8}$$

Note that these equations are exactly the opposite of what is obtained for resistors in parallel or in series.

5.3 Dielectrics

There are many applications for capacitors in everyday life, ranging from energy storage to signal filtering. Practically all electronic equipment contains a large number of capacitors and modern touch screens are also based on capacitance. It is obvious that all such technologies are not based only on capacitors that rely on a vacuum between the electrodes. Therefore it is important to consider what happens to matter when it is placed in an E-field, for example, between the plates of a parallel plate capacitor. We already know for the case of a metal that all charges will flow according to the electric field and as a result will cancel it, giving zero field inside. Here we will consider all materials where the charge can't move that easily and that are thus not a metal. All these non-metallic systems, thus including insulators and semiconductors, are called *dielectrics*.

The first thing that should be realised is that (besides vacuum) a true insulator does not exist. All matter is made up of a collection of positive and negative charges, even though they can't move long distances, and the material can thus be polarised. One example of the **polarisation of matter** is when it is made up of polar molecules, such as water. Similar to NaCl discussed in Chapter 2, the bonding between the hydrogen and oxygen atoms in water is because an electron from each hydrogen goes to the oxygen. As illustrated in Figure 5.2 on the left this results in a negative charge on the oxygen and positive on the hydrogen. Because the hydrogen atoms are arranged in an angle of 104° with respect to each other, the two dipole moments $\vec{p}_1$ and $\vec{p}_2$ don't cancel and the molecule has a net dipole moment. In an external electric field the water molecules will thus tend to align and, as will be explained below, reduce the E-field.

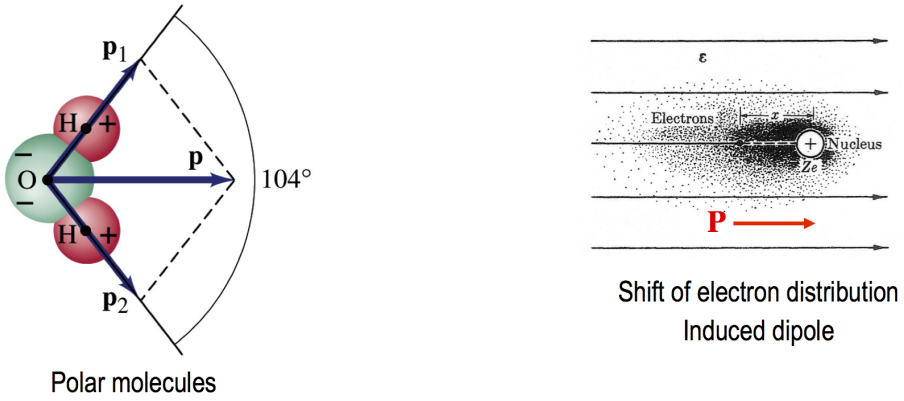

Figure 5.2: Examples of the polarisation of matter.

Another possibility of how matter can be polarised is illustrated in Figure 5.2 on the right. This is representative for any material where the atoms are close enough together to form a solid. As we know, each atom is made up of a positively charged nucleus and a negatively charged electron cloud. These electrons are bound to the nucleus by the Coulomb force, but in an external electric field their distribution can shift a bit, which creates a dipole moment pointing from the centre of mass of the electron cloud to the nucleus. One can thus regard this as a bad metal; the electrons want to move due to the E-field, but they are still too strongly bound by the nucleus to really move away. Therefore just like a metal completely screens the E-field, *in a dielectric the E-field will be partially screened* and thus reduce the field inside the material.

To be able to quantify this picture we need to introduce two new concepts: the **polarisation of a material** $\vec{P}$ and the **polarisation charge** Q_p or polarisation surface charge density σ_p. The polarisation of a material is the summation of all the dipole moments mentioned in the previous paragraph. In principle, all these individual dipoles don't have to point exactly in the same direction and if they all point in random directions $\vec{P} = 0$, otherwise $\vec{P}$ will have some finite value. In a very simple sense we can regard the continues arrangement of small dipoles as a vector field, and we will see that we can almost directly add or subtract $\vec{P}$ from the E-field.

The polarisation charge is the charge that is associated with the polarisation, and here we will mostly use it as a helpful mathematical concept. More formally it can be derived by considering the amount of charge that has moved through a surface S in the dielectric when the polarisation was

induced. Thus $dQ_p = \vec{P} \cdot d\vec{S}$ and if we integrate this out we obtain

$$Q_P = -\oiint \vec{P} \cdot d\vec{S} \tag{5.9}$$

whereby the minus sign is due to the definition of the direction of a dipole. Again, this definition is not of importance for this lecture, but it should be realised that the charge density of polarisation charge at the surface is the aforementioned σ_p. For a homogeneous polarisation this will be positive on one side of the material and negative on the other side. What we referred to as charges in the previous chapters will be called **conduction charges** (sub-index c) in the presence of dielectrics.

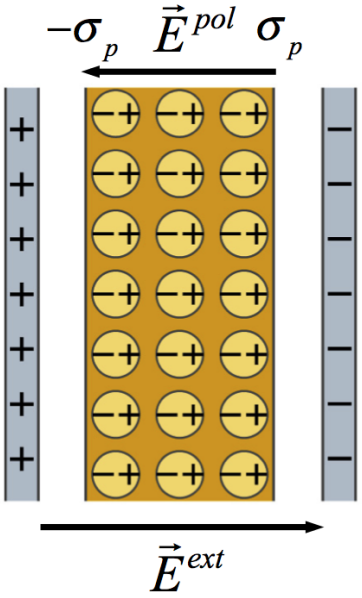

Figure 5.3: Dielectric in an external E-field.

Now let's consider what happens if we place a dielectric in a (homogeneous) external electric field. This situation is illustrated in Figure 5.3 for a constant external E-field $\vec{E}^{ext}$. Besides this external field, we see that also the surface charge density on each side of the dielectric will cause a field, which is opposite to the external field and not larger than $\vec{E}^{ext}$, this is termed $\vec{E}^{pol}$. Now we can consider the average E-field in the dielectric $\left\langle \vec{E} \right\rangle$ by taking the difference of the external field and the polarisation field:

$$\left\langle \vec{E} \right\rangle = |\vec{E}^{ext}| - |\vec{E}^{pol}| \tag{5.10}$$

To illustrate that the field is reduced here the difference of magnitudes is taken, but to be precise one should take a vectorial sum. The reason the average E-field in the dielectric is considered is to take possible inhomogeneities or an anisotropic response into account. However, this goes beyond the scope of this lecture and in the following we consider $\vec{E}$ as the field inside the dielectric instead of $\left\langle \vec{E} \right\rangle$.

From the above considerations we can now define a connection between the polarisation of the dielectric and the (average) field inside

$$\vec{P} = \epsilon_0 \chi_e \vec{E} \tag{5.11}$$

Here χ_e is the electric susceptibility, or how easy it is to polarise a material. Large values indicate a dielectric that is easily polarised and the χ_e of vacuum is zero.

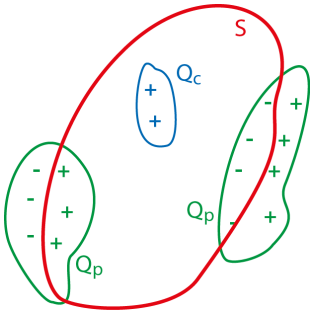

Figure 5.4: Gauss's law with dielectrics.

In the previous chapter we have seen that Gauss's law is a powerful tool, thus it is important to consider how it can be used in the presence of a dielectric. In Figure 5.4 the very general case is illustrated of conduction charges and a part of a dielectric enclosed in a Gaussian surface S. We can now write down Gauss's law (Eq. 4.6) and get

$$\oiint_S \vec{E} \cdot d\vec{S} = \sum \frac{Q_c}{\epsilon_0} + \sum \frac{Q_p}{\epsilon_0}$$

Combining this with Eq. 5.9 we obtain:

$$\oiint_S \vec{E} \cdot d\vec{S} = \sum \frac{Q_c}{\epsilon_0} - \frac{1}{\epsilon_0} \oiint_S \vec{P} \cdot d\vec{S}$$

And after rewriting

$$\oiint_S \left(\epsilon_0 \vec{E} + \vec{P} \right) \cdot d\vec{S} = \sum Q_c \tag{5.12}$$

This looks very similar to the original Gauss's law and can be even further simplified by the definition of the so-called D-field

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P} \tag{5.13}$$

Based on this Gauss's law for D-fields is obtained

$$\oiint \vec{D} \cdot d\vec{S} = \sum Q_c \tag{5.14}$$

$$\nabla \cdot \vec{D} = \rho_c \tag{5.15}$$

Two things should be noticed. The first is that *only conduction charges are a source of D-fields.* The second is that also in vacuum we can use the D-field and *it is then equal to the E-field times* ϵ_0.

The D-field is historically termed the displacement field, but this terminology has no real physical meaning and will not be used in this lecture. Here it will just be called D-field and its use lies in the fact that we can define a field quantity that does not depend on the material or medium.

If we combine Equations 5.13 and 5.11 we obtain

$$\vec{D} = \epsilon_0 \left(1 + \chi_e\right) \vec{E} \tag{5.16}$$

We now define the **relative permittivity** ϵ_r as follows

$$\epsilon_r = 1 + \chi_e \tag{5.17}$$

To obtain the relationship between the E-field and D-field

$$\vec{D} = \epsilon_0 \epsilon_r \vec{E} \tag{5.18}$$

This relationship is valid for any material (in vacuum $\epsilon_r = 1$). In a general sense the relative permittivity is a tensor; it can be anisotropic, non-linear, and in-homogenous. For simplicity we will consider ϵ_r here as a material-specific scalar, which has values ranging from 1.0006 (air), to 7 in rubber, 80 in water, 300 in strontium titanate, or even more than 100'000 in some complex oxides.

Because only conduction charges are sources of the D-field, it only depends on the external parameters and not on the material which is present. From Eq. 5.18 it then directly follows that **in a dielectric the E-field is reduced by a factor** ϵ_r. This effect is directly measurable and the E-field of a point charge in a dielectric becomes

$$\vec{E} = \frac{Q\hat{r}}{4\pi\epsilon_r\epsilon_0 r^2}$$

and the D-field of the same point charge is

$$\vec{D} = \frac{Q\hat{r}}{4\pi r^2}$$

regardless of the medium. This reduction of the E-field is often referred to as the *screening of charges in dielectrics*. This effect is responsible for many of the functional properties of doped semiconductors, but this goes beyond the scope of this lecture and is more fitting for a lecture in solid state physics.

Let's now consider a capacitor with dielectric placed between the plates, similar to the situation shown in Figure 5.3, but without vacuum between the dielectric and the plates. Here a definition of ϵ_r which is used to determine the value in practice becomes important. The relative permittivity is given by the ratio of capacitance of the empty (vacuum) capacitor C_0 and the identical capacitor with the dielectric inserted:

$$\epsilon_r = \frac{C}{C_0} \tag{5.19}$$

Based on the definition in Eq. 5.3 the empty capacitor will have charge Q_0 and potential difference Φ_0 given by

$$\Phi_0 = \frac{Q_0}{C_0}$$

In the empty capacitor the surface charge density σ_0 causes an electric field $E_0 = \frac{\sigma_0}{\epsilon_0}$.

We now have to distinguish between two cases: 1) where the capacitor is isolated and the charge is thus kept constant, and 2) where the capacitor is connected to a power supply and the potential difference is kept constant.

In the first case the surface charge density will also be the same, thus the D-field will be the same and the E-field will be reduced by a factor ϵ_r:

$$E = \frac{E_0}{\epsilon_r} = \frac{\sigma_0}{\epsilon_0 \epsilon_r}$$

The new voltage drop can be calculated from Eq. 5.3 and 5.19 and becomes

$$\Phi = \frac{Q}{C} = \frac{Q_0}{C_0 \epsilon_r} = \frac{\Phi_0}{\epsilon_r}$$

Thus also the potential difference is reduced by a factor ϵ_r, which should come as no surprise given that fact that the geometry is the same. It should

be stressed again that the above equations are **only valid for an isolated capacitor.**

For the second case, the power supply provides the charge needed to keep the potential difference constant at Φ_0. The charge on the capacitor after inserting the dielectric then becomes

$$Q = \Phi_0 C = Q_0 \frac{C}{C_0} = Q_0 \epsilon_r$$

The amount of charge thus increases by a factor ϵ_r and also the surface charge density increases by the same factor: $\sigma = \sigma_0 \epsilon_r$. We can now calculate the E-field in the dielectric as

$$E = \frac{\sigma}{\epsilon_0 \epsilon_r} = \frac{\sigma_0 \epsilon_r}{\epsilon_0 \epsilon_r} = \frac{\sigma_0}{\epsilon_0} = E_0$$

The E-field inside the dielectric is thus the same as in the empty capacitor. It is easily verified that the D-field on the other hand increases by a factor ϵ_r in this case. Again, it should be noted that these last derivation is **only valid for a capacitor connected to a power supply.**

5.4 Electric energy

For applications it is useful to consider the energy that can be stored in a capacitor or collection of charged objects. This will also allow us to calculate the force exerted on the dielectric or between charged conductors.

To calculate the electric energy stored in a capacitor it is most straightforward to consider the work done in charging the capacitor. Let's consider a parallel plate capacitor as illustrated in Figure 5.5 with capacitance C and distance d. An amount of charge q corresponds to a potential difference Φ according to $q = \Phi C$ and a E-field $E = \frac{\Phi}{d}$. The force on a small additional amount of charge dq between the plates of the capacitor is

$$F = dqE = \frac{\Phi dq}{d}$$

The work done to move this charge against the potential Φ from one plate to the other is

$$dW = Fdx = \frac{\Phi dq}{d} d = \Phi dq$$

The total work done to charge the capacitor with a charge Q then becomes

$$W = \int_0^Q \Phi dq = \int_0^Q \frac{q dq}{C} = \frac{Q^2}{2C}$$

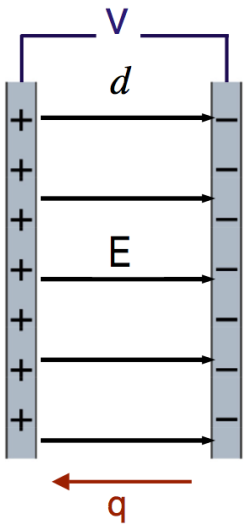

Figure 5.5: Charging a capacitor

The energy stored in the capacitor is equal to the work done in charging it:

$$U_E = \frac{Q^2}{2C} = \frac{1}{2}C\Phi^2 = \frac{1}{2}Q\Phi \tag{5.20}$$

Depending on what parameters are known any of the three expressions can be used.

Without further evidence it will be stated that this result in Eq. 5.20 can in general be used for any charged object, or collection of objects. As an example we consider the collection of three charges for which we derived the mutual potential energy in Eq. 2.7. This equation can be rewritten as

$$U = \frac{1}{2}Q_1\left(\frac{Q_2}{4\pi\varepsilon_0 r_{12}} + \frac{Q_3}{4\pi\varepsilon_0 r_{31}}\right) + \frac{1}{2}Q_2\left(\frac{Q_1}{4\pi\varepsilon_0 r_{12}} + \frac{Q_3}{4\pi\varepsilon_0 r_{32}}\right) + \frac{1}{2}Q_3\left(\frac{Q_2}{4\pi\varepsilon_0 r_{23}} + \frac{Q_1}{4\pi\varepsilon_0 r_{13}}\right)$$

The term between the brackets is the potential Φ_i at the respective charge outside the brackets Q_i due to all other charges. The expression can thus be summarised as

$$U = \frac{1}{2}Q_1\Phi_1 + \frac{1}{2}Q_2\Phi_2 + \frac{1}{2}Q_3\Phi_3$$

or

$$U_E = \sum_i \frac{1}{2}Q_i\Phi_i \tag{5.21}$$

This has the same form as Eq. 5.20 and shows that the total electric energy can be obtained by summation.

We now return to the capacitor with a dielectric between the plates to determine how the dielectric changes the amount of stored energy. The electric energy of the empty capacitor is according to 5.20: $U_E^0 = \frac{1}{2}C_0\Phi_0^2$. Similar to the example described above we again have to consider the two different cases: 1) the capacitor is isolated when the dielectric is inserted, and 2) it is connected.

In the first case the charge is constant at Q_0 and the electric energy after inserting the dielectric becomes (5.20)

$$U_E = \frac{Q_0^2}{2C} = \frac{Q_0^2}{2\epsilon_r C_0} = \frac{1}{\epsilon_r}U_E^0$$

The stored electric energy is thus reduced by a factor ϵ_r.

In the second case the potential is kept constant at Φ_0 and the electric energy with dielectric is

$$U_E = \frac{1}{2}C\Phi_0^2 = \epsilon_r\frac{1}{2}C_0\Phi_0^2 = \epsilon_r U_E^0$$

The stored energy thus increases by a factor ϵ_r. This additional energy is provided by the voltage source.

An isolated system will try to reduce the energy and the force between charged objects can thus be determined from

$$\vec{F} = -(\nabla U_E)_Q \tag{5.22}$$

Where the subscript Q indicates that the derivatives are taken for constant charge. We expand this now for the x direction, but similar expansions hold for other coordinates and coordinate systems. Using $U_E = \frac{Q^2}{2C}$ from Eq. 5.20 and realising Q is constant we obtain

$$F_x = -\left(\frac{\partial U_E}{\partial x}\right) = -\frac{1}{2}Q^2\partial\left(\frac{1}{C}\right)\frac{1}{\partial x} = \frac{Q^2}{2C^2}\frac{\partial C}{\partial x}$$

Thus

$$F_x = \frac{1}{2}\Phi^2\frac{\partial C}{\partial x} \tag{5.23}$$

Applied to the parallel plate capacitor we can now use this to calculate the force between the plates. The capacitance is given by Eq. 5.5 and if

we put this in Eq. 5.23 considering the spacing of the plates along the x-direction, we obtain

$$F_x = -\frac{1}{2}\epsilon_0 A\frac{\Phi^2}{x^2} = -\frac{1}{2}\epsilon_0 E^2 A = \frac{-\sigma^2 A}{2\epsilon_0}$$

It should come as no surprise that the plates *attract* each other with a force that is proportional to the surface charge density squared multiplied by the surface area.

For a connected system with constant potential difference the force can be expressed as follows

$$\vec{F} = +(\nabla U_E)_\Phi \tag{5.24}$$

The positive sign comes from the extra work done by the power supply and the subscript Φ indicates the potential is kept constant. Using $U_E = \frac{1}{2}C\Phi^2$ from 5.20 we obtain along the x-direction

$$F_x = \left(\frac{\partial U_E}{\partial x}\right) = \frac{1}{2}\Phi^2\frac{\partial C}{\partial x}$$

Which is identical to Eq. 5.23. Which makes sense because regardless of how a situation is reached, the forces in the steady state situation should be the same. Thus in both cases the plates attract each other. It is left up to the reader to calculate the force with which a dielectric is pulled into a capacitor from the change of capacitance.

In some cases it is more useful or exact to **calculate the electric energy from the E-field**. Here only the final equations will be stated without further proof, in the lecture the derivation was shown

$$U_E = \frac{1}{2}\epsilon_0 \int_\tau E^2 d\tau \tag{5.25}$$

Here τ is the region of interest in space. In a dielectric Eq. 5.25 changes to

$$U_E = \frac{1}{2}\epsilon_0 \int_\tau \epsilon_r E^2 d\tau \tag{5.26}$$

or more formally

$$U_E = \frac{1}{2}\int_\tau \vec{D}\cdot\vec{E} d\tau \tag{5.27}$$

Chapter 6

Electrical networks

In the previous chapters the static situation was considered on how electric fields look and what their consequences are. In everyday life and for many useful applications a current typically needs to flow through the respective elements and we want to be able to manipulate this current. In this chapter some of the basic physical concepts for creating an electric network are considered. Many aspects are treated in more detail in other lectures.

The main idea of this chapter is that we want to have a useful current. However, we have seen in Equation 4.11 that $\oint \vec{E} \cdot d\vec{L} = 0$, thus somewhere works needs to be done to provide a useful current. This work is called the **electromotive force** with units of Volt [V] and it is what we typically consider when we connect a circuit to a power supply or battery. In the rest of the lecture such a source will be assumed to be connected if needed.

6.1 Resistivity and conductivity

From experiments it follows that the current density flowing through a material is directly proportional to the electric field applied across it. The material specific proportionality constant is called the *conductivity* σ (not to be confused with the surface charge density). Thus $\vec{j} = \sigma\vec{E}$, whereby in the following we assume everything is collinear and we thus omit the vectors for simplicity.

From Eq. 1.3 we know that in this case the total current is $I = jA$ with A being the cross-sectional area of the material. For this homogeneous E-field we can also rewrite it in terms of potential difference $\Delta\Phi$ according to Eq. 3.12: $E = \frac{\Delta\Phi}{l}$ whereby l the length of the material is. Combining this we obtain

$$I = \frac{\sigma A}{l}\Delta\Phi$$

We now define the (macroscopic) *conductance* Σ of a material as

$$\Sigma = \frac{\sigma A}{l} \tag{6.1}$$

Meaning that $I = \Sigma \Delta\Phi$. For most applications we are more interested in the electrical *resistance* (units Ohm [Ω]) of a material, which is the reciprocal value to the conductance: $R = \frac{1}{\Sigma}$. We then obtain a, probably familiar, expression

$$I = \frac{\Delta\Phi}{R} \text{ or } \Delta\Phi = RI \tag{6.2}$$

Which is known as **Ohm's law**. For simplicity $\Delta\Phi$ is often also replaced by just Φ

Based on the definition of the resistance above, we can also define a material specific parameter the *resistivity* ρ (not to be confused with charge density).

$$\rho = \frac{1}{\sigma} = \frac{RA}{l}$$

From this it is clear that the units of resistivity are Ωm.

More common is the reverse expression

$$R = \frac{\rho l}{A} \tag{6.3}$$

This allows us to calculate the resistance of a piece of material if we know the area and length. It fits directly in with the experience that short, thick cables have a lower resistance as long, thin ones.

Ohm's law presented above is based on macroscopic observations and does not explain the microscopic picture of why a material has a certain conductivity, and thus resistivity. The first step in this direction is provided by the **Drude model**. Further refinements are based on quantum mechanical considerations and are one of the topics of condensed matter physics.

The Drude model is based on the distinction of the electron velocities in a material in two different regimes: the "real" or average velocity v, and the drift velocity v_d. The average velocity is the speed the electrons achieve between collisions with each other or with defects in the crystal lattice. This is on the order of 10^6 m/s, material dependent, and independent of whether an E-field is applied or not. As indicated in Figure 6.1(a) it is oriented in all directions and thus induces no net electron flow. The drift velocity on

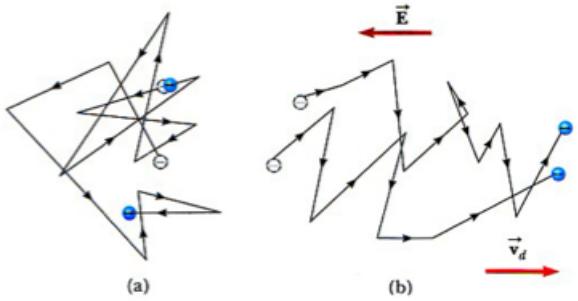

Figure 6.1: Drude model. a) Electron trajectories without applied field. b) Electron trajectories with E-field applied.

the other hand is induced by the E-field and also has a direction dictated by the field, as indicated in Figure 6.1(b). It is on the order of 10^{-4} m/s.

There is a material dependent average mean free path between collisions λ (has nothing to do with line charge density), which leads to an average time τ between collisions of

$$\tau = \frac{\lambda}{v}$$

During this time the E-field can accelerate the electron according to

$$a = \frac{eE}{m}$$

Where m the effective mass of the electron in the given material. The drift velocity after a time τ becomes

$$v_d = a\tau = \frac{\lambda eE}{vm}$$

The current flowing through the material due to the applied E-field is the density of electrons n with this drift velocity times their charge and velocity

$$j = nev_d = \frac{ne^2E\lambda}{2mv}$$

Where the factor $\frac{1}{2}$ comes from the fact that the average drift velocity is half the maximum drift velocity.

We also know that $j = \sigma E$ which leads to the following expression for the conductivity

$$\sigma = \frac{ne^2\lambda}{2mv} \tag{6.4}$$

In this expression, n, λ, m, and v are all material dependent. This thus directly explains why the conductivity of different materials is different, and it even explains the temperature dependence of conductivity through the temperature dependence of λ. However, to explain why different materials have different parameters and how they can be determined, goes far beyond the scope of this course.

6.2 Kirchhoff's laws

The Kirchhoff laws are probably familiar to any student of this course. Here only their physical background will be shortly explained.

The first Kirchhoff law that the sum of all currents at a given point equals zero, or

$$\sum_i I_i = 0 \tag{6.5}$$

is based on the conservation of charge. Whatever amount of charge goes into a given point also has to exit again.

The second Kirchhoff law states that the sum of all voltage sources is equal to all voltage drops over elements. Or that the sum of all potential drops is equal if we consider the electromotive force being a potential drop in the opposite direction. This can be expressed as

$$\sum_i \Delta\Phi_i = 0 \tag{6.6}$$

This law is based on the path independence of potential difference.

These two laws can be used to derive the current, voltage, or resistance at any point in a network of resistors. One of the main results is the determination of the equivalent resistance R' for a combination of resistors. For resistors in series this leads to

$$R' = \sum_i R_i \tag{6.7}$$

For resistors in parallel one obtains

$$\frac{1}{R'} = \sum_i \frac{1}{R_i} \tag{6.8}$$

It should be noted that this is opposite to the situation for capacitors in Eq. 5.7 and 5.8.

6.3 Combination of capacitor and resistor

The combination of a capacitor and resistor is an important element in many electrical networks. Here we will only consider the DC case, but also in AC networks this combination plays a crucial role. The main use is in signal shaping and in making filters.

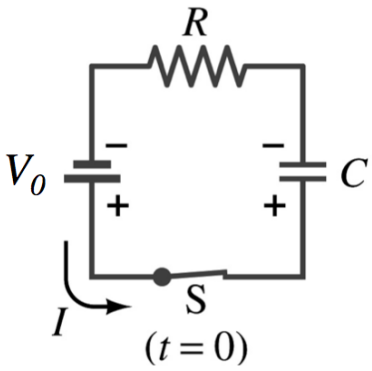

Figure 6.2: Simplest scheme of a resistor and capacitor. The situation for charging a capacitor is shown.

In Figure 6.2 the simplest possible set-up of a resistor-capacitor (RC) combination is shown. We will first consider the case of **charging the capacitor**. To this means the switch S will be closed at time $t = 0$ and a current can flow to charge the capacitor.

To determine how the charge on the capacitor changes as a function of time we start with Kirchhoff's second law applied to the circuit, leading to

$$\Phi_0 - \Phi_C - RI = 0$$

Where Φ_C is the potential drop over the capacitor. It also applies that $\Phi_C = \frac{Q}{C}$ and thus

$$\frac{d\Phi_C}{dt} = \frac{\frac{dQ}{dt}}{C} = \frac{I(t)}{C}$$

Combined this leads to the following expression

$$\Phi_0 - \Phi_C - RC\frac{d\Phi_C}{dt} = 0$$

Which can be rewritten as a separable differential equation

$$\frac{d\left(\Phi_0 - \Phi_C\right)}{\Phi_0 - \Phi_C} = -\frac{dt}{RC}$$

Integration yields the solution $\ln(\Phi_0 - \Phi_C) = \frac{-t}{RC} + K$ with K a constant. By considering that for $t = 0$, $\Phi_C = 0$ this constant can be determined as $K = \ln \Phi_0$. Put together this gives the following behaviour of the voltage over capacitor as a function of time

$$\Phi_C(t) = \Phi_0 \left(1 - e^{\frac{-t}{\tau}}\right), \ \tau = RC \tag{6.9}$$

With $Q(t) = C\Phi_C(t)$ we can now determine the charge on the capacitor as function of time. This shows a trend as displayed in Figure 6.3 at the top, reaching a maximum value of $C\Phi_0$.

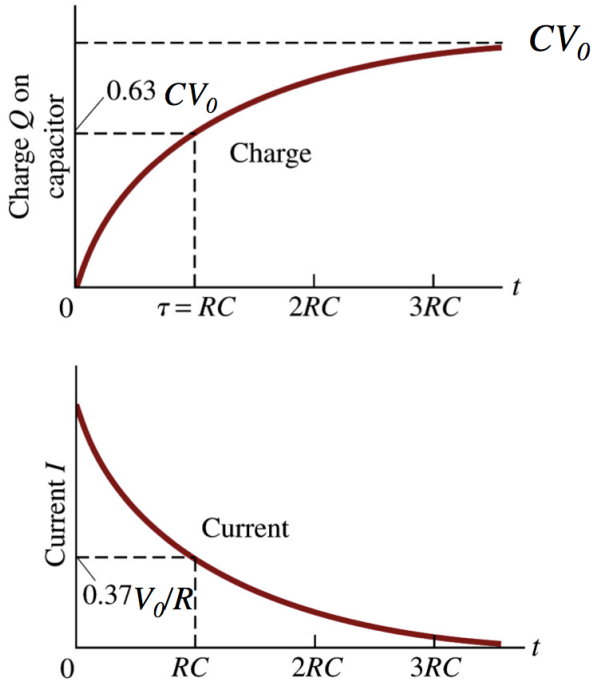

Figure 6.3: Behaviour of the charge (top) and current (bottom) when charging a capacitor.

From the voltage also the current needed to charge the capacitor as a function of time can be determined:

$$I(t) = C\frac{d\Phi_C}{dt} = \frac{\Phi_0}{R} e^{\frac{-t}{\tau}} \tag{6.10}$$

This shows an exponential decay as displayed in Figure 6.3 at the bottom.

In Equations 6.9 and 6.10 $\tau = RC$ is the characteristic time constant for the circuit. It is a measure of how long it takes to charge the capacitor. It should be noted however that a truly fully charged capacitor is obtained only after infinite time.

The behaviour when **discharging a capacitor** can be obtained along similar lines. In this case the power supply is removed from Figure 6.2, the capacitor is charged with an initial charge Q_0, and at $t = 0$ the switch is closed to let the capacitor discharge over the resistor. Equation 6.6 now yields that $RI + \Phi_C = 0$, which, with the same substitutions as above, can be rewritten as

$$\frac{d\Phi_C}{\Phi_C} = \frac{-dt}{RC}$$

This can also be solved by integration. Taking into account the boundary condition that the voltage at $t = 0$ is $\Phi_0 = \frac{Q_0}{C}$ the potential difference of the capacitor as function of time becomes

$$\Phi_C(t) = \Phi_0 e^{\frac{-t}{\tau}},\ \tau = RC \tag{6.11}$$

The current shows the same exponential decay as when charging a capacitor and is

$$I(t) = \frac{\Phi_0}{R} e^{\frac{-t}{\tau}}$$

The time constant τ of discharging a well calibrated capacitor can be used to precisely determine the value of the resistor. Furthermore, cyclical charging and discharging a capacitor can be used to create a saw-tooth shaped current signal.

Chapter 7

Magnetostatics

All the previous chapters dealt with the properties of electrostatic interactions. The Coulomb law served as a fundamental law from which we could derive other properties, such as the E-field and potential. Static charges are the source, or basic element, of electric fields and all fields can, in principle, be obtained by a superposition of charges. We will now turn to magnetic interactions and see that there are many similarities, but also some crucial differences.

The main difference is that there is no magnetic equivalent to the point charge. This would be a magnetic monopole, just like a point charge is half a dipole, but magnetic monopoles have never been identified and physicists still argue whether they can exist at all. Consequently, there is also no magnetic equivalent to a conductor.

The most important similarity is that both electric and magnetic interactions can be described by vector fields; the E-field and the B-field. We will therefore start from the B-field and expand from there. The most straightforward way is now to look at different objects or units and determine the magnetic field they produce. Further we will consider how they react to a B-field and which force they experience.

7.1 Current element as basis for B-field

Hans Christian Oersted was in 1820 the first to observe that a current carrying wire influences the needle of a compass. Based on this initial observation Andre Ampère performed more detailed experiments. He found that the needle of a compass indeed circles around a current carrying wire as illustrated in Figure 7.1, following a right hand rule type of pointing direction. More importantly he also determined the force on a small current

carrying test wire as a function of relative orientation and position with respect to the main wire. This showed that a current element $I\vec{dl}$ is a source of a magnetic field, but also responds to it, just like in electrostatics a point charge is both a source of, and influenced by, a E-field.

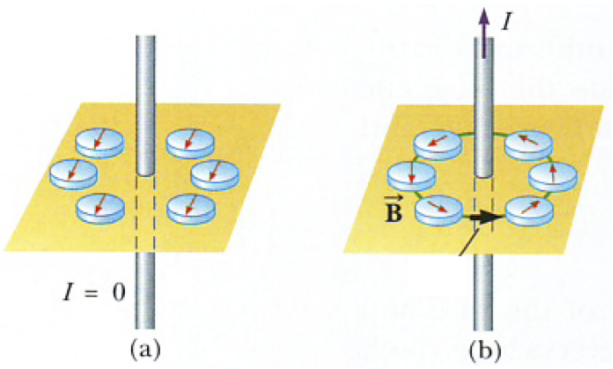

Figure 7.1: Charge of compass needle pointing direction around a current carrying wire.

The findings of Ampère can be summarised as follows. The force the test wire experiences is proportional to the current running through it. The force is perpendicular to the current element: $\vec{F} \perp I\vec{dl}$. It is also perpendicular to the magnetic field produced by the big wire (as measured by the compass): $\vec{F} \perp \vec{B}$. And it scales with sine of the angle between current and field. Putting all this together yields

$$d\vec{F} = I\vec{dl} \times \vec{B} \tag{7.1}$$

Whether the B-field originates from another wire or from any other source is not important. From the vector product it is directly clear that *this is not a central force and thus also can't be conservative.*

We can now use this result to calculate the force on a current circuit in a homogeneous B-field. For simplicity we consider the square circuit shown in Figure 7.2, but the result is valid for a "closed" circuit of any shape. Based on Eq. 7.1 the total force on the circuit is given by $\oint I\vec{dl} \times \vec{B}$. In sections 1 and 3 the current element is (anti)parallel to the B-field and the force is thus zero: $F_1 = -F_3 = 0$. The force on section 2 is $F_2 = IaB$ and the force on part 4 is $F_4 = -IaB$. The total force is thus zero. However, similar to a dipole in an E-field, the opposite forces F_2 and F_4 induce a **torque on the circuit**. This torque is given by

$$\vec{T} = I\vec{A} \times \vec{B} \tag{7.2}$$

Here $\vec{A}$ is the vector normal to the circuit with a magnitude given by the circuit area. In the square example this thus becomes $ab\vec{e}_n$ whereby the direction of $\vec{e}_n$ is given by applying the right hand rule to the current circulation.

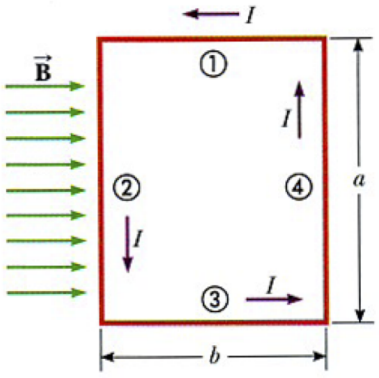

Figure 7.2: Square current circuit in a magnetic field.

This torque on a current carrying circuit in a B-field can be used for a variety of applications. The most prominent ones are a galvanometer, and similar measurement devices, where the torque as function of the current directly allows to let a needle indicate the magnitude of the current. The other example is a stepper motor, which is often used when precise movements are required. In this case the current special periodic shape to allow the current carrying circuit to make full rotations.

In the discussion above we considered the influence of a B-field on a current element, here we will now look at the B-field created by a current element. The central law in this respect is the **law of Biot-Savart** which states

$$d\vec{B} = \frac{\mu_0 I d\vec{l} \times \hat{r}}{4\pi r^2} \tag{7.3}$$

Here μ_0 is the *permeability of free space* and $\mu_0 = 4\pi \times 10^{-7}$.

During the lecture it will be explained how Equation 7.3 can be used to determine the B-field generated by several current carrying objects. Here only the final results will be given.

A **single loop** with radius R at a distance x from the centre along the axis of the loop:

$$B_x = \frac{\mu_0 I R^2}{2\left(x^2 + R^2\right)^{\frac{3}{2}}} \text{ and } B_y = B_z = 0 \tag{7.4}$$

Inside a **solenoid** with n windings per unit length:

$$B_x = \mu_0 n I \text{ and } B_y = B_z = 0 \tag{7.5}$$

The B-field inside the solenoid is homogeneous and independent on the radius of the windings.

Due to a **straight wire**:

$$\vec{B} = \frac{\mu_0 \vec{I} \times \hat{r}}{2\pi r} \tag{7.6}$$

This is the B-field as characterised by Oersted and Ampère.

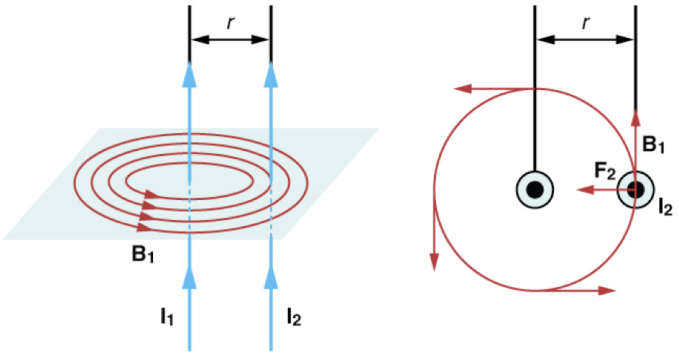

Figure 7.3: Force between two wires

We can now use these results to calculate the force between two straight wires carrying currents I_1 and I_2 as illustrated in Figure 7.3. When substituting Eq. 7.6 in 7.1 and integrating over the length l we obtain

$$\vec{F} = \frac{\mu_0 l \vec{I}_2 \times \left(\vec{I}_1 \times \hat{r}\right)}{2\pi r} \tag{7.7}$$

If we consider only parallel wires and consider the direction of the force from the right part of Figure 7.3 this can be simplified to

$$F = \frac{\mu_0 I_1 I_2 l}{2\pi r} \tag{7.8}$$

Here it should be realised that *opposite currents repel and like currents attract each other.*

7.2 Moving charges and B-fields

An electrical current is, as we know, composed of moving charges. We can thus use the results derived above to determine the force of the B-field on a moving charge, and the B-field created by a moving charge. In order to do so we can express the current element above as $I\vec{dl} = nq\vec{v}$. Here n is the number of charges and in the following we want to derive expressions for a single charge and thus set $n = 1$.

If we enter this substitution in the law of Biot-Savart in Eq. 7.3 we obtain the following **B-field due to a single moving charge**

$$d\vec{B} = \frac{\mu_0 q\vec{v} \times \hat{r}}{4\pi r^2} \tag{7.9}$$

Thus any moving charge creates a B-field which moves along with the charge. As we will see below every moving charge also experiences a force due a magnetic field and this is what counteracts the (expansive) Coulomb interaction in a beam of moving electrons.

To determine the **force on a moving charge in a B-field** we substitute $I\vec{dl} = q\vec{v}$ in Ampère's law in Eq. 7.1 to obtain

$$\vec{F} = q\vec{v} \times \vec{B} \tag{7.10}$$

This is known as the **Lorentz force**.

For a charged particle moving parallel to the magnetic field, the force, and thus also the acceleration, is zero. If the velocity is perpendicular to a homogeneous magnetic field, the Lorentz force is perpendicular to both velocity and B-field. Because also the resulting acceleration is perpendicular to the velocity, the magnitude of the velocity will be constant. Only the direction of the velocity will change. In a large enough field the charge will describe a circular motion as illustrated in Figure 7.4, where the acceleration is given by $a = \frac{v^2}{r}$, with r the radius of the circle. Using that $F = qvB = ma$ we can rewrite this obtain the radius of the circle

$$r = \frac{mv}{qB} \tag{7.11}$$

If the extension of the field is smaller as the diameter of the circle, this radius represents the local curvature of the path.

This radius can be used to identify, or select, particles based on their charge to mass ratio. One example of this is the bubble chamber for the identification of subatomic particles. Further the separation of radioactive isotopes can also be performed by a magnetic field and formed the basis for the uranium enrichment plant in Oakridge in 1944.

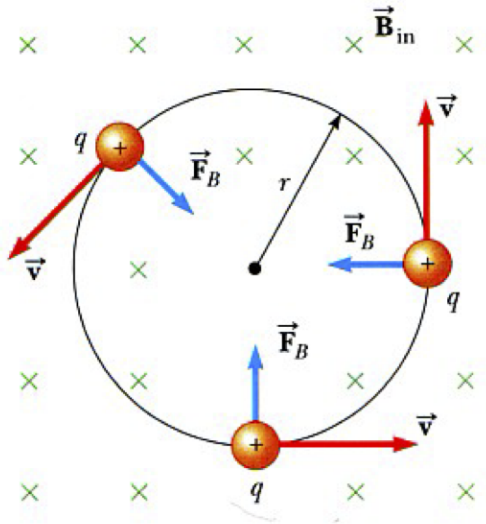

Figure 7.4: Charged particle with velocity perpendicular to homogeneous B-field.

In the general case of a charged particle with non-zero velocity in a homogeneous B-field the velocity component parallel to the field is not affected, whereas the perpendicular component will form a circle. The particle will thus describe a spiral path. The time for one circle is according to Eq. 7.11

$$t = \frac{2\pi r}{v} = \frac{2\pi m}{qB}$$

The pitch of the spiral $x_{\parallel}$ then becomes

$$x_{\parallel} = v_{\parallel} t = \frac{2\pi m v_{\parallel}}{qB}$$

In non-homogenous B-fields the trajectory of a charged particle becomes a distorted spiral in a very general sense. Well shaped magnetic fields can actually be used to trap particles with a given initial velocity and charge to mass ratio. In the earth's magnetic field cosmic charged particles are deflected to the poles and some of them are trapped going back and forth between the north and south pole, forming the *Van Allen belt*.

7.3 Combinations of B- and E-fields

To determine the force on a charged particle with velocity $\vec{v}$ in a combination of an E- and B-field we can simply sum Eq. 3.14 and 7.10 to obtain

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{7.12}$$

This expression is generally valid for any combination of velocity and fields. For simplicity we will here limit ourselves to the case that the electric field and the magnetic field are perpendicular to each other.

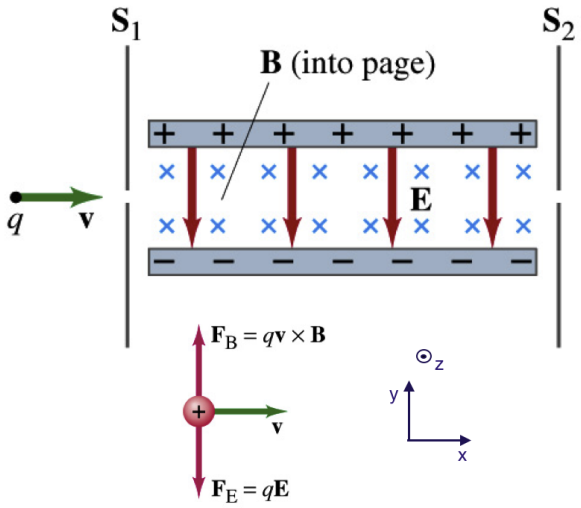

Figure 7.5: Combination of E- and B-field to form a velocity filter.

Consider the situation illustrated in Figure 7.5, with the coordinate system as indicated. Both the E- and B-field only induce a force along the y direction, and with the indicated relative orientation their contributions to the force are opposite. The total force along the y direction then becomes

$$F_y = q\left(E - v_x B\right)$$

A special case arises when

$$v_x = \frac{E}{B}$$

In this case also $F_y = 0$ and thus the total force on the charge is zero and it will travel straight independent of the charge q. This is the working principle of a velocity filter for charged particles.

For charged particles in free space the influence of E- and B-fields can in a very general sense be summarised as follows. An E-field determines the kinetic energy, a B-field determines the momentum, and a combination of both determines the velocity.

An important application of combined fields is the **Hall effect**. Here a magnetic field is applied perpendicular to a slab of material with thickness t through which a current I flows. The charges q will have drift velocity v_d either in the direction of the current or in the opposite direction, depending on whether q is positive or negative. Because also the resulting force $\vec{F} = q\vec{v}_d \times \vec{B}$ depends on the sign of q the charge carriers will always be pushed in the same direction. This creates an charge imbalance, and thus potential difference, between the two sides of the sample. The sign of this potential difference will depend on sign of q and can be used to determine whether the current flows as (negative) electrons, or as (positive) missing electrons which are termed holes.

Furthermore, the relationship between the current, the applied B-field, the thickness of the sample, and the resulting potential difference can be derived to be

$$\Delta\Phi = \frac{BI}{nqt}$$

Here n is the carrier density according to the Drude model in Eq. 6.1. Because the charge can only be e^- or e^+ and the other parameters are known, the Hall effect can also be used to determine the carrier density in a material.

7.4 Magnetic dipole

In section 3.4 we encountered the electric dipole, consisting of opposite charges at a small distance, as an important building block for electrostatics. Here the magnetic dipole will be introduced. Their behaviour in an E-field, respectively B-field will be very similar and also the shape of the field they produce is identical.

An important example of a magnetic dipole is a **small current loop**. Based on the use of current elements as the basis of a B-field, a dipole would be two opposite current elements, and a loop is exactly such a continuous collection of opposite current elements. The shape of the loop is not important, but typically one considers it as circular.

In the discussion surrounding Figure 7.2 we saw that the force on a current loop in a homogenous magnetic field is zero and that the loop experiences a torque as described in Eq. 7.2. If we now define the **magnetic**

dipole moment $\vec{m}$ as

$$\vec{m} = I\vec{A} \tag{7.13}$$

with I and $\vec{A}$ as defined for Eq. 7.2. With this definition the similarity with the expressions for the electric dipole becomes directly clear. All the properties for the magnetic dipole can be derived by replacing $\vec{B}$ for $\vec{E}$, $\vec{m}$ for $\vec{p}$, and $\frac{1}{\mu_0}$ for ϵ_0.

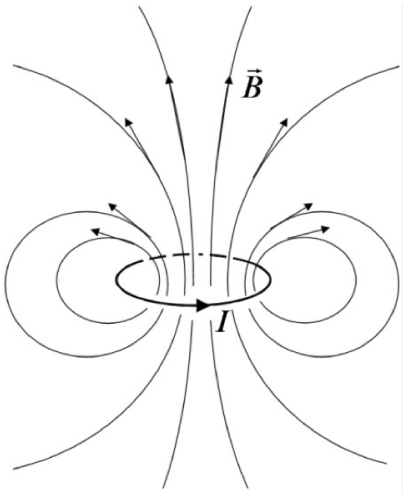

Figure 7.6: B-field of a magnetic dipole.

The torque becomes

$$\vec{T} = \vec{m} \times \vec{B} \tag{7.14}$$

The force on a dipole in a B-field is

$$\vec{F} = (\vec{m} \cdot \nabla)\, \vec{B} \tag{7.15}$$

Consequently the potential energy of a magnetic dipole in a magnetic field becomes

$$U = -\vec{m} \cdot \vec{B} \tag{7.16}$$

The B-field due to a magnetic dipole, decomposed in a radial and tangential component, is given by

$$B_r = \frac{2\mu_0 m \cos\theta}{4\pi r^3} \tag{7.17}$$

$$B_\theta = \frac{\mu_0 m \sin\theta}{4\pi r^3} \tag{7.18}$$

This has exactly the same shape as the E-field of an electric dipole as shown in Figure 7.6.

Also a **small permanent magnet** is a magnetic dipole and can't be distinguished from a small current loop based on the behaviour in a external field, nor on the field it creates. In this case the dipole moment is related to the magnetisation of the material. More on this topic will be discussed in a later chapter.

7.5 Ampère's circuital law

In the treatment of electrostatics we saw the usefulness of general laws such as Gauss's law. In magnetostatics this role is taken by Ampère's circuital law.

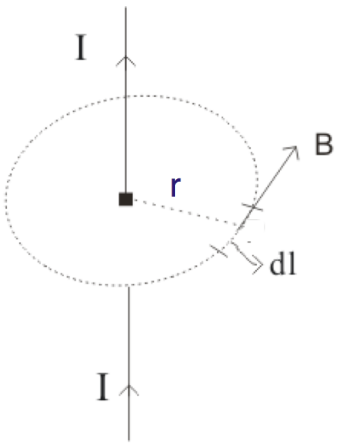

Figure 7.7: Illustration of Ampère's circuital law.

Consider a circuital path around a current carrying wire as illustrated in Figure 7.7. We know from Eq. 7.6 that the tangential magnetic field is given by

$$B_\theta = \frac{\mu_0 I}{2\pi r}$$

If we now integrate this B-field along the path L we obtain

$$\oint_L \vec{B} \cdot d\vec{L} = B_\theta L = \frac{\mu_0 I}{2\pi r} 2\pi r = \mu_0 I$$

Although this appears to be a special case, this result is *valid for any closed path* of whatever shape and angle relative to the wire. Based on the

superposition principle of fields, it is also valid if several wires are encircled. This leads us to formulate **Ampère's circuital law** as follows

$$\oint \vec{B} \cdot d\vec{L} = \sum_i \mu_0 I_i \tag{7.19}$$

As we will see later, we have to be careful with what encircled current means. If we draw the closed path around some current carrying wires, the idea is clear, but there are situations where a more formal definition is required. In this case we should integrate the current density $\vec{j}$ over the surface S of which the closed path forms the boundary. This means that

$$\oint \vec{B} \cdot d\vec{L} = \mu_0 \iint_S \vec{j} \cdot d\vec{S} \tag{7.20}$$

Which can also be expressed in differential form as

$$\nabla \times \vec{B} = \mu_0 \vec{j} \tag{7.21}$$

This law again shows that currents are the source of magnetic fields.

In the lecture it will be shown how Equations 7.19 and 7.20 can be used to easily determine the magnetic field of current carrying objects.

It is also possible to formulate **Gauss's law for B-fields**, although it has much less importance as for E-fields. Independent of whether the source is a current element, a dipole, or a permanent magnet the total flux of the B-field over any closed surface is always zero. Thus

$$\oiint \vec{B} \cdot d\vec{S} = 0 \tag{7.22}$$

$$\nabla \cdot \vec{B} = 0 \tag{7.23}$$

This represents the absence of sources and sinks for magnetic field lines; they always close on themselves. Or in other words it represents the absence of magnetic monopoles.

7.6 Vector potential

Because the force due to a magnetic field is not conservative, the work done does depend on the path taken. It is therefore of only very limited use to define a scalar potential for the magnetic field in a similar way as we did for the E-field. Another problem becomes clear in the presence of currents

if we do consider a scalar potential given by $\vec{B} = -\nabla\Phi_B$. When we take the curl of both sides of this expression we get $\nabla \times \vec{B} = \mu_0\vec{j}$ for the left side and $-\nabla \times \nabla\Phi_B = 0$ for the right side. This shows that Φ_B becomes almost meaningless in the presence of a current, which we know to be the source of $\vec{B}$ and thus of very limited (or no) use.

The reason we could define a scalar potential for the E-field was that $\nabla \times \vec{E} = 0$. For the B-field we have a similar situation with regard to the divergence ($\nabla \cdot \vec{B} = 0$) and we can make use of this to define a potential given by a vector field, the so-called vector potential $\vec{\mathcal{A}}$. The definition of the **vector potential** is

$$\vec{B} = \nabla \times \vec{A} \tag{7.24}$$

Whereas for E-fields any constant added to the scalar potential field gives the same $\vec{E}$, we can add any gradient of a scalar field to $\vec{\mathcal{A}}$ and obtain the same B-field. The definition of the vector potential in Eq. 7.24 gives us freedom in the choice of both $\vec{\mathcal{A}}$ and $\nabla \cdot \vec{\mathcal{A}}$. One typical choice for the latter is $\nabla \cdot \vec{\mathcal{A}} = 0$, but other choices can be made.

As an example, let's consider what could be a suitable vector potential for a current element. Starting with Eq. 7.3 and rewriting $\frac{-\hat{r}}{r^2} = \nabla\frac{1}{r}$ we obtain for the B-field of a circuit

$$\vec{B} = -\oint \frac{\mu_0 I}{4\pi} d\vec{l} \times \nabla\frac{1}{r}$$

With

$$-d\vec{l} \times \nabla\frac{1}{r} = \nabla \times \frac{d\vec{l}}{r} - \frac{1}{r}\nabla \times d\vec{l}$$

and that $\nabla \times d\vec{l} = 0$ this becomes

$$\vec{B} = -\oint \frac{\mu_0 I}{4\pi} \nabla \times \frac{d\vec{l}}{r}$$

after reversing the order of differentiation and integration we get

$$\vec{B} = \nabla \times \oint \frac{\mu_0 I d\vec{l}}{4\pi r} \tag{7.25}$$

Now we can compare Equations 7.25 and 7.24 to get an expression for the vector potential, yielding

$$d\vec{\mathcal{A}} = \frac{\mu_0 I d\vec{l}}{4\pi r} \tag{7.26}$$

The main thing we learn from Eq. 7.26 is that the field lines of the vector potential are typically parallel to $d\vec{l}$ and that this is a suitable choice of $\vec{\mathcal{A}}$.

7.7 Angular momentum and precession of a dipole

One important difference between magnetic dipoles and electric dipoles is that all *magnetic dipoles also have angular momentum.* This is most clearly illustrated by a current loop with radius r. There are particle (electrons) with mass m_e and angular velocity ω circling around, yielding an angular momentum of $\vec{L} = m_e r^2 \omega \hat{e}_n$. We can also describe the (circular) current as number of charges that pass by per second and thus as $I = \frac{e\omega}{2\pi}$. If we combine this expressions with the definition of the dipole moment in Eq. 7.13 we obtain

$$\vec{m} = I\vec{A} = \frac{e\omega}{2\pi}\pi r^2 \hat{e}_n = \frac{1}{2} e r^2 \omega \hat{e}_n$$

Which can be rewritten based on the angular momentum as

$$\vec{m} = \frac{e}{2m_e}\vec{L}$$

Using the **gyromagnetic ratio** γ we obtain the following expression for the general case as

$$\vec{m} = \gamma \vec{L} \tag{7.27}$$

For a current loop $\gamma = \frac{e}{2m_e}$. Which is also valid for an electron circling around an atom.

Most elementary particles also have an intrinsic magnetic dipole moment, and thus also angular momentum, which is called the spin. In this case it is common to use the **Landé g-factor** instead of the gyromagnetic ratio. The relationship between them is

$$g = \frac{2m_e}{e}\gamma$$

For an electron around an atomic core $g = 1$, but for a free electron the intrinsic magnetic dipole moment creates the situation that $g = 2$. For other elementary particles g takes other values.

Whereas an electric dipole will turn to align to an external E-field, the angular momentum of a magnetic dipole will make it *precess* around a B-field. This is similar to the recession of a spinning top in the earth gravitational field. The frequency of precession is called the Larmor frequency ν_L and it depends on the applied magnetic field as follows

$$\nu_L = \frac{-\gamma}{2\pi} B$$

In magnetic resonance imaging (MRI) techniques this frequency is measured. From this the gyromagnetic ratio can be determined, which in turn gives insight in the atomic composition.

Chapter 8

Induction

In the previous chapters we have derived the most important properties of static electric and magnetic fields. In the central equations representing the circuital laws and Gauss's laws we see that for this steady situation the E- and B-fields are decoupled. One of the main results of this chapter will be that this decoupling is lifted for fields that vary as a function of time. Especially, a changing B-field will induce an electrical field which can be used for applications.

8.1 Electromagnetic induction

Central to the understanding of induction is the concept of the **flux of a magnetic field**. Analogous to the situation for an E-field illustrated in Figure 4.1 and Eq. 4.2, we can define the flux of a B-field as

$$\varphi_m = \int_A \vec{B} \cdot d\vec{A}. \tag{8.1}$$

From this equation it is clear that φ_m changes either when the B-field changes, the area changes, or the relative orientation of $\vec{B}$ and $d\vec{A}$ changes.

Let us start with the example illustrated in Figure 8.1 of a conductor with length $\vec{l}$ moving with velocity $\vec{v}$ in a constant homogeneous B-field. All the charges in the conductor will feel a Lorentz force according to Eq. 7.10 which will separate the charges according to their charge in different directions. This charge imbalance will induce a potential difference, which we can calculate from the E-field. According to the definition of the E-field in Eq. 3.1 the field due to the Lorentz force will be the following

$$\vec{E} = \frac{\vec{F}}{q} = \frac{q\vec{v} \times \vec{B}}{q} = \vec{v} \times \vec{B}$$

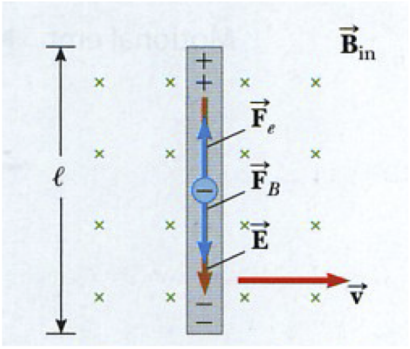

Figure 8.1: Moving conductor in a B-field.

If we now take the situation illustrated in Figure 8.1 where $\vec{B} \perp \vec{v} \perp \vec{l}$ then $E = vB$ and the potential difference between the two ends of the conductor becomes

$$\Delta\Phi = El = Bvl$$

Note that the E-field and potential difference in this case are not due to static charge, but more closely related to the electromotive force in Chapter 6.

Another way of looking at this problem is by realising that in a time dt the conductor has moved a distance vdt and given the length of the conductor this is equivalent to an area $A = vldt$. Thus the magnetic flux cut by the conductor in a time dt is

$$d\varphi_m = BA = Bvldt$$

From the previous result we know that $\Delta\Phi = El = Bvl$ and thus that

$$\Delta\Phi = -\frac{d\varphi_m}{dt} \tag{8.2}$$

Here the minus sign comes from the relative orientation of the conductor, velocity, and B-field. This equation is generally valid, although we only derived it for a specific case, and can be used to determine the induced voltage due to any change in magnetic flux as a function of time. It is often referred to as **Faraday's law of electromagnetic induction**.

We can now apply this expression to the circuit of changing size illustrated in Figure 8.2. An external force $\vec{F}_{app}$ moves the (green) sliding part of the circuit to the right with a velocity v. The flux of the constant magnetic

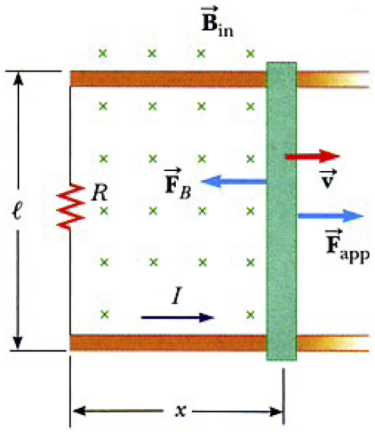

Figure 8.2: A circuit which changes in size in a homogeneous and constant magnetic field.

field through the circuit is $\varphi_m = Blx$ and the change is flux as a function of time becomes

$$\frac{d\varphi_m}{dt} = Bl\frac{dx}{dt} = Blv$$

According to Eq. 8.2 the induced potential difference, or electromotive force is thus

$$\Delta\Phi = \frac{d\varphi_m}{dt} = Blv$$

This will potential difference will let a current flow through the circuit in the direction as indicated in the figure. In the B-field this current will experience a force $\vec{F}_B$ trying to move the sliding conductor against the original direction.

That the system with induction forms a negative feedback loop where the induced current opposes the original motion is a very general result and valid for any magnetic induction effect. It is often referred to as **Lenz's law** which states that: *the direction of any magnetic induction effect is such as to oppose the cause of the effect.*

From this it directly follows that one of the many applications of magnetic induction is to stop or damp motion. This is called a **Eddy current break**. Furthermore, almost any generator is based on magnetic induction. An external force, such as water power, is used to rotate a circuit in a fixed B-field, resulting in an alternating current. Further important applications of induction are the microphone and magnetic tape read head.

Here we used Eq. 8.2 to consider moving conductors, but the same expression can be used for any other reason why the flux changes. This can

be a varying B-field, or when the relative orientation of circuit and B-field changes.

In the previous discussion we have focussed on the potential difference induced by a changing magnetic flux, here we will make the link to the electric field. From Eq. 3.7 we know that the E-field and the potential difference are related and if we now consider the potential difference going around a closed circuit, as induced by a changing magnetic flux, we obtain

$$\Delta\Phi = \oint_L \vec{E} \cdot d\vec{L} = -\frac{d\varphi_m}{dt}$$

Using the definition of the magnetic flux in Eq. 8.1 we can rewrite this as

$$\oint_L \vec{E} \cdot d\vec{L} = -\frac{d}{dt} \iint_A \vec{B} \cdot d\vec{A} \tag{8.3}$$

This extension of the circuital law is called **Faraday's law**. It is generally valid, both in vacuum and in a material. If no time varying magnetic flux is present then it reduces back to Eq. 4.11. It should be noted that in the presence of a varying magnetic flux the E-field is no longer conservative. Many expressions and conclusions that were based on this assumption can thus not necessarily be used in this case.

During the lecture it will be shown how the **differential form** of Equation 8.3 can be derived. Here it suffices to just give the final result

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{8.4}$$

8.2 Inductance

In the previous chapter we have seen that every current creates a magnetic field. For a current carrying circuit it is easy to realise that this magnetic field has a flux through the circuit which is also proportional to the current: $\varphi_m \propto \vec{B} \propto I$. We now define the **self-inductance** L as the proportionality constant between current and flux:

$$\varphi_m = LI \tag{8.5}$$

By definition self-inductance is always positive. It has the unit Henry [H] and is just like capacitance primarily dependent on the geometry. This is the same L as encountered in AC networks, where in combination with a resistor and capacitor it becomes a band pass filter.

If we now apply Eq. 8.2 and take into account that the self-inductance is not time dependent we obtain that

$$\Delta\Phi = -L\frac{dI}{dt} \tag{8.6}$$

This expression represents that a varying current will cause a changing magnetic flux and thus induce a potential difference. As a result of Lenz's law, this voltage drop is opposite to the applied voltage difference to let the current flow. In other words, the self-inductance will create a current which opposes the original current and can thus be regarded as an extra resistance.

To summarise, *self-inductance is an internal resistance against changes in the current*. Because every current carrying element causes a magnetic field and also partly picks up the flux of this field, this resistance **can't be avoided**. Even a simple single wire of length x has a self-inductance of $L = \frac{\mu_0 x}{8\pi}$ or $\frac{1}{20}$ μH per meter.

The most obvious example of an object to calculate the self-inductance is the solenoid. We consider a solenoid or coil with n windings per meter, and area A, and length l. Using the B-field from Eq. 7.5 we can calculate the flux cut per winding

$$\varphi_n = BA = \mu_0 nIA$$

There are a total of nl windings, thus the total flux becomes $\varphi_{tot} = \mu_0 n^2 lAI$. From the definition in Eq. 8.5 we can now calculate the self-inductance of a solenoid as

$$L = \frac{\varphi_{tot}}{I} = \mu_0 n^2 lA \tag{8.7}$$

During the lecture it will be shown how to calculate the self-inductance of a co-axial cable.

Along similar lines as a circuit will induce a magnetic flux through itself, it can also cause a flux through another circuit. The relation between the current in one circuit and the flux trough another is called the **mutual inductance** M. In general this is only determined by the geometry and thus the mutual inductance is the same whichever way we look at the problem. It is thus defined as

$$\varphi_2 = MI_1 \quad \text{and} \quad \varphi_1 = MI_2 \tag{8.8}$$

If we now want to determine the potential difference induced in circuit number 2 due a changing current in circuit 1 we obtain

$$\Delta\Phi_2 = -\frac{d\varphi_m}{dt} = -M\frac{dI_1}{dt} \tag{8.9}$$

And the same is valid in reverse. It should be noted that also the mutual inductance has the unit Henry.

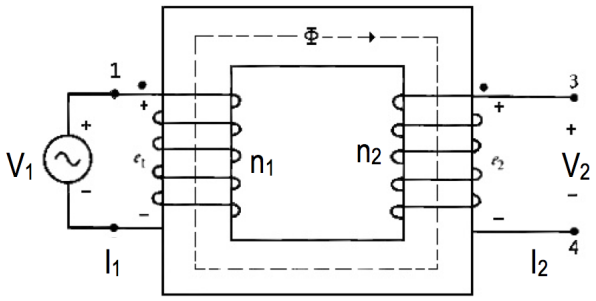

Figure 8.3: Ideal transformer where the same flux passes through both coils.

For the general case of coupled circuits we have to consider how strong the coupling between the circuits is, based on how much of the flux from one circuit can pass through the other. This is referred to as the *coefficient of coupling* k which has a maximum value of 1, and is defined as

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{8.10}$$

Here we will only consider the ideal situation of $k = 1$ for simplicity, but the general argument will not change.

In Figure 8.3 and ideal **transformer** is shown consisting of a primary coil with n_1 windings and a secondary coil with n_2 windings. The working mechanism is that an alternating current and voltage in the primary coil will create a varying magnetic flux in the secondary coil, which will induce a potential difference and current there. Through using a magnetic core, as will be explained in the next chapter, $k = 1$ and thus the same flux φ_m passes through both coils. The total flux passing through coil 1 is $\varphi_1 = n_1\varphi_m$ and through coil 2 $\varphi_2 = n_2\varphi_m$. Thus the potential difference $\Phi_1 = -n_1\frac{d\varphi_m}{dt}$, but through the other coil it passes in the other direction and $\Phi_2 = n_2\frac{d\varphi_m}{dt}$. If the coupling were not perfect, this would have to be

multiplied by k. The ratio of the voltages on the primary and secondary coil now becomes

$$\frac{\Phi_2}{\Phi_1} = -\frac{n_2}{n_1}$$

Here the minus sign indicates the phase difference between the two signals. Because the energy can't increase, the current has to decrease in a similar fashion and

$$\frac{I_2}{I_1} = \frac{n_1}{n_2}$$

Transformers are used to reduce or increase the voltage of an AC signal and are encountered when making the step down from high voyage power lines to the house grid, but also from the grid to, for example, a laptop computer. The ideal transformer for a given application depends on factors such as the operating frequency and expected load. However, these issues are beyond the scope of this course.

Chapter 9

Magnetism in materials

In the previous chapters we have only considered the magnetic field in vacuo, here we will consider what happens to the magnetic field in a medium. There will be some similarities to the discussion of the electric field in dielectric materials in Chapter 5, but there will also be some important differences. The first is that due to the non-existence of magnetic monopoles in nature, there is no magnetic equivalence to a conductor, and the presented theory will have to be valid for all types of materials. Furthermore, in the discussion of dielectrics we could, for sake of simplicity, ignore any non linear or temperature dependent effects. If we would do the same for magnetic materials we would bypass the most prominent member of the family; the ferromagnets, which are commonly referred to as magnets.

9.1 Magnetic fields in materials

In general the B-field is zero in materials in the absence of an external magnetic field, except in the case of ferromagnets which will be discussed in more detail later. We also know that currents are the source of an external B-field, for example in a solenoid. Now the question becomes how the magnetic field is influenced by the presence of a material.

The most straightforward approach is to consider how the inductance of a coil is changed when we insert a material, a so-called core, in it. The inductance without a core is L_0 and the inductance with a core is changed to $L_m = \mu_r L_0$. Here μ_r is the **relative permeability** which is this defined as

$$\mu_r = \frac{L_m}{L_0} \tag{9.1}$$

By definition $\mu_r = 1$ for vacuum. In typical materials it take values ranging from 0.9 to 10^7. In dielectrics the equivalent parameter is the relative permittivity as can be seen from the definition in Eq. 5.19.

From Eq. 8.5 we know that the flux $\varphi_m = LI$, and if we now consider the solenoid with the same current before and after inserting the core we obtain

$$\frac{\varphi_m}{\varphi_0} = \frac{L_m I}{L_0 I} = \mu_r$$

If the magnetic flux is changed and the area stays the same, it means that also the mean magnetic field is changed accordingly: $\frac{B_m}{B_0} = \mu_r$. Or more formally

$$\vec{B}_m = \mu_r \vec{B}_0 \tag{9.2}$$

In section 5.3 we saw that in a dielectric the electric field is reduced by a factor ϵ_r and that all equations change accordingly. For a magnetic field in a medium the field and all related equations are *increased* by a factor μ_r.

In everyday language we can for example say that "iron becomes magnetised", but what do we mean by this? Here we can again make the link to dielectrics where we defined the polarisation of a material through the collection of electric dipole moments in the material. We can now define the **magnetisation** $\vec{M}$ of a material as magnetic dipole moment per volume

$$\vec{M} = \frac{d\vec{m}}{d\tau}$$

We will expand on the nature of magnetic dipoles in a material in the next section. This magnetisation causes a magnetic field $\vec{B}_M$ by itself, just like the polarisation causes a E-field, and the total magnetic field inside a material becomes

$$\vec{B} = \vec{B}_0 + \vec{B}_M$$

Similar to the introduction of the D-field in section 5.3, we can now introduce the **magnetic field strength** or the H-field.

$$\vec{H} = \frac{\vec{B}}{\mu_0} - \vec{M} \tag{9.3}$$

The currents that we considered in the previous chapters are the sources of the H-field and Ampere's circuital law thus becomes

$$\oint \vec{H} \cdot d\vec{L} = I \tag{9.4}$$

$$\nabla \times \vec{H} = \vec{j} \tag{9.5}$$

The B-field can be referred to as the **magnetic flux density** and this name already implies that Gauss's law for magnetic fields in Eq. 7.23 applies to the B-field, also in materials. On the other hand, when using Ampere's circuital law in the presence of a material one has to use the H-field and thus Eq. 9.5. The reason for this goes beyond the scope of this lecture, but when carefully applied this provides the procedure to solve many magnetic problems involving materials.

To clarify the situation further let us consider a solenoid with a fixed current. The H-field inside the solenoid does not depend on the material which is inserted and is always nI (follows from circuital law Eq. 9.5). However, the B-field changes with μ_r depending on the material. This is similar to a capacitor with fixed voltage; the E-field will be independent of the material inserted, but the D field will change with the factor ϵ_r. Now we consider the same solenoid but with materials with different μ_r (or partly air), which is the problem of the **electromagnet**, but also generally valid for interfaces. Because no flux leaves the system, the magnitude of the B-field is the same in all parts (but different from the empty solenoid), and the magnitude of the H-field thus changes across interfaces. Similarly, if we take a capacitor with different dielectrics the magnitude of the D-field will be the same everywhere (but different from the empty capacitor) and the magnitude of the E-field changes at the interface according to the material.

As will be explained below, the important point is thus to distinguish interfaces between materials (B is constant and H changes), and what happens when you insert a material in a solenoid (H is constant and B changes). A handwaving way to describe it is that once the solenoid system is defined by the current and all materials present, the magnitude of the B-field is the same everywhere in space, but this value of the B-field depends on what materials are present.

The relationship between field and magnetisation, or how easily the material can become magnetised, is given by the **magnetic susceptibility** χ_m according to

$$\vec{M} = \chi_m \vec{H} \tag{9.6}$$

If we now rewrite Eq. 9.3 and insert the above expression we obtain

$$\vec{B} = \mu_0 \left(\vec{H} + \vec{M} \right) = \mu_0 \left(1 + \chi_m \right) \vec{H}$$

An alternative definition for the relative permeability, along the lines of Eq. 5.17, is

$$\mu_r = 1 + \chi_m \tag{9.7}$$

Thus we obtain for the B-field

$$\vec{B} = \mu_0 \mu_r \vec{H} \tag{9.8}$$

Again the similarity, and differences, with respect to the case for dielectrics described in Eq. 5.18 should be noted.

Equation 9.8 is always valid, also in vacuum where it becomes $\vec{B} = \mu_0 \vec{H}$. The relative permeability can be very complex. Apart from orientational dependence and inhomogeneities, which will not be discussed here, it can be a function of temperature, H-field, and even of history. This will be discussed in the next section.

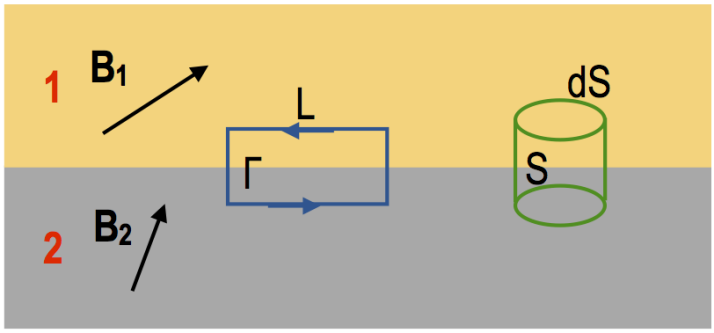

Figure 9.1: Magnetic screening at an interface.

Before giving an explanation of the responses of different types of material in a magnetic field, it is useful to consider what will happen at the interface between two materials with different μ_r. In Figure 9.1 such an interface between material 1 and 2 is displayed. We can first apply Gauss's law for B-fields, Eq. 7.23, on a cylinder with infinitesimal length, and upper and lower area dS, over the interface. Because we can neglect the side surfaces, and the other surfaces are pointing in opposite direction, we obtain

$$B_{\perp 1} dS - B_{\perp 2} dS = 0$$

Which yields

$$B_{\perp 1} = B_{\perp 2} \tag{9.9}$$

For the circuital law we have to consider the H-field. This can be applied to the closed loop indicated in the figure, but with infinitesimal size perpendicular to the interface. There are no free currents at the interface thus we obtain

$$\oint \vec{H} \cdot d\vec{L} = H_{\parallel 2} L - H_{\parallel 1} L = 0$$

Thus $H_{\parallel 1} = H_{\parallel 2}$ and using Eq. 9.8 we obtain

$$B_{\parallel 1} = \frac{\mu_{r1}}{\mu_{r2}} B_{\parallel 2} \tag{9.10}$$

Thus the perpendicular component of the B-field does not change, whereas the parallel component is rescaled according to the ratio of the relative permeabilities. This can be used for **magnetic screening** to isolate something from surrounding magnetic fields. The idea is to take a material with a very high relative permeability, such as mu-metal which has $\mu_r \approx 10^5$. At the interface with, for example air, all the magnetic field lines inside the material will be bent almost almost flat along the interface. If the material is thick enough, typically about a mm is sufficient, the B-field will not be able to pass through it. A closed surface of such a material will thus guide all field lines around its interior. Note that this will also be a Faraday cage and thus also no E-field will pass inside. Inside a mu-metal box is thus an electromagnetically very quite place, as long as no sources are present. Furthermore, this effect will also guide field lines through a magnetic material, which forms the basis for the transformer in the previous chapter. Lastly, because there is no magnetic equivalence to a conductor, the magnetic field lines typically don't, and are certainly not required to, impinge perpendicular to a surface.

9.2 Microscopic picture of magnetism

Based on experiments considering the response to an applied magnetic field, 5 different classes of materials can be distinguished.

- **Diamagnetic**: Materials with a small and negative χ_m, which is independent of the applied H-field, and also does not depend on temperature.
- **Paramagnetic**: Materials with a small and positive χ_m, which is independent of H, but decreases with increasing temperature.
- **Ferromagnetic**: Metallic materials with a large and positive χ_m which strongly depends on H and on history. These materials become paramagnetic above a critical temperature T_C called the Curie temperature
- **Antiferromagnetic**: Materials with a small and positive χ_m, which depends on H and history. Becomes paramagnetic above a critical temperature T_N called the Néel temperature.

- **Ferrimagnetic**: Similar behaviour as ferromagnetic materials, but non-metallic.

The last two classes are given for completeness and will only be shortly explained below.

The first major difference is whether χ_m of a material is positive or negative. Based on magnetic energy considerations, which go beyond the scope of this course, the force on a material with χ_m in an H-field can be determined. Along the x-direction, for example, the force is approximately

$$F_x \approx \frac{1}{2}\mu_0\chi_m\frac{\partial H^2}{\partial x} \tag{9.11}$$

Here H^2 becomes the magnetic field density. This means that diamagnetic materials will seek regions of lower field density and are thus repelled by a magnetic dipole, as illustrated in the lower panel of Figure 9.2. Materials with positive χ_m, such as paramagnetic materials, seek regions with higher field density and are thus attracted by a magnetic dipole, as illustrated in the upper panel of Figure 9.2. Note that in a homogeneous magnetic field, the field density is constant.

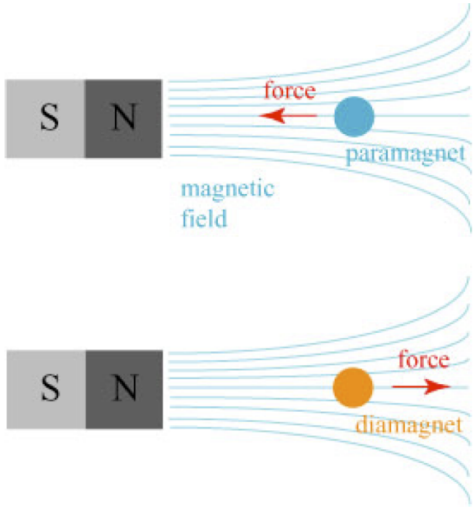

Figure 9.2: Force on a paramagnetic and diamagnetic material in a dipole field.

We will now to turn to the question on *why materials have a different magnetic susceptibility*. In Section 7.6 the concept of intrinsic angular

momentum, and thus magnetic dipole moment, of the electron was introduced. This **electron spin** can only have discrete values, referred to as *up* and *down*, as was shown in the Stern-Gerlach experiment in 1922. These up and down values mean opposite magnetic dipole moments and often they are thus drawn as vectors or arrows. Based on the principle of how atomic levels are filled (Hund's rule) we can now have two possible situations: *i*) per atom there are just as much electrons with spin up as spin down, meaning that all electrons are paired, *ii*) there are **unpaired electrons** meaning there are more of one spin direction then the other, thus not all spin is cancelled and there is a *resultant magnetic dipole moment*.

In **diamagnetic materials** all electrons are paired and there is thus no resultant magnetic moment. Examples of this are copper, gold, graphite, and water, thus almost all living things are diamagnetic. An external magnetic field will cause small circular currents and a Larmor precession of electrons. This will create a magnetic field opposing the external field and thus reducing the field inside the material. All materials have this diamagnetic contribution, but typically it is overshadowed by other effects. Typically the diamagnetic effect is very small, but as will be shown in the lecture, a superconductor is a perfect diamagnet and the effect is very strong.

In **paramagnetic materials** there are unpaired electrons and every atom thus has a magnetic dipole moment. However, these dipole moments are randomly ordered with expect to each other, as illustrated in the first panel in Figure 9.3. When an H-field is applied these dipole moments will align with the field, creating an $\vec{M}$ pointing in the same direction as $\vec{H}$. The field is thus increased by the presence of the material and χ_m is positive. Higher temperatures mean that the magnetic moments will start to fluctuate, reducing their ordering also in an H-field, and thus reducing χ_m.

Most metals in the periodic table are paramagnetic, except for the noble metals and bismuth. Examples are aluminium, tungsten, and platinum. But also some small molecules, such as O_2 are paramagnetic. Some oxidation states of hemoglobin are slightly paramagnetic, but most are diamagnetic.

Ferromagnetic materials are what we commonly refer to as magnets. In this class of materials the atoms have a resultant magnetic moment, but in contrast to paramagnetic materials these moments are ordered also without applying an external field (Figure 9.3B). The reason for this spontaneous ordering goes far beyond the scope of this lecture and can only be explained using quantum mechanical considerations. This magnetisation will greatly enhance the applied magnetic field, thus making $\chi_m \gg 0$. The only pure materials which are ferromagnetic at room temperature are nickel, iron, and

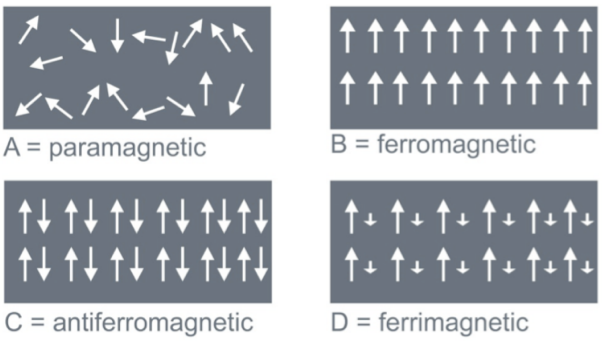

Figure 9.3: Order of the resultant magnetic moment, or spin, in different material classes.

cobalt. Other materials become ferromagnetic at lower temperature, and many composite materials based on these elements are also ferromagnetic.

If we increase the temperature for a ferromagnetic material, thermal fluctuations will compete with the spontaneous ordering, and above a certain temperature the ordering will break down and the material will behave like a paramagnet. This is called the Curie temperature. For iron $T_C = 770°\text{C}$. If we now cool down the material again the spontaneous ordering reappears, but a block of iron as a whole will not be magnetic anymore. The reason for this is that there are many different **magnetic domains** in which all magnetic moments point in the same direction, but between domains the order is random. These domains are of the order of tens of micrometer.

In an external magnetic field the domains will start to align in the direction of the field, which requires significant energy. One can actually listen to cracking noise the alignment makes, which is referred to as the Barkhausen effect. If the H-field becomes large enough all domains are oriented in the same direction, and the material is said to be saturated. This initial magnetisation is indicated by the dashed line from 0 to a in Figure 9.4.

If we now reduce the external field again, some domains will switch back, but many will remain as they were in the saturation field. When the H-field is zero we reach point b in the curve in 9.4, which is called the **residual magnetisation**. How large this is is material dependent, and this is the magnetisation we use in permanent magnets. When switch the H-field, the magnetic domains will start to switch with it and at a certain moment the resultant B-field is zero (point c). This is called the **coercive field** and indicates how difficult it is to flip the magnetisation of a ferromagnet. At a

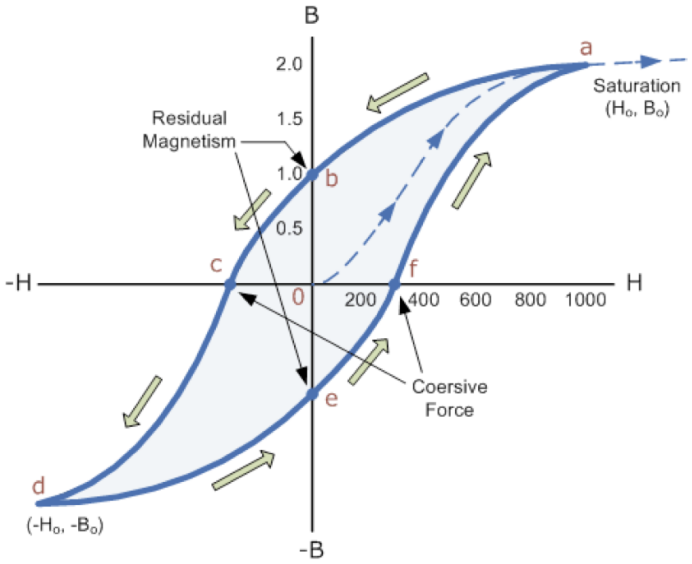

Figure 9.4: Hysteresis curve of a ferromagnetic material.

certain moment we will reach opposite saturation (point d), and if we reduce the H-field from that point the residual magnetisation will be opposite from before (point e).

A full loop as displayed in Figure 9.4 is called a **hysteresis cycle**. It contains the most important properties of a ferromagnetic material, apart from the Curie temperature. When the cycle is narrow, i.e. the coercive force and remanence are small, the material is a soft magnet. If the cycle is broad with large coercive field and and remanence, the material is a hard magnet. The latter are ideal for permanent magnets and memory devices, whereas soft magnets are better for transformers. The reason for this is that the work done per unit volume in one hysteresis cycle is the area enclosed by the curve in the B vs. H diagram:

$$W_V = \oint B dH = \oint H dB \qquad \text{(per unit volume)} \tag{9.12}$$

In a transformer the system is driven through this cycle with the frequency of the AC signal and this energy is lost in the form of noise and heat.

For completeness the situation for **antiferromagnetic materials** is shortly explained. In this case there are also unpaired electrons and a resultant magnetic moment per atom. Also here there is spontaneous ordering, but now the spins on adjacent atoms are arranged exactly opposite to each other, as illustrated in Figure 9.3C. With an applied H-field the

magnetisation is thus zero, but in a field some of the magnetic moments will flip. The only pure antiferromagnetic material is chromium, for the rest it is found in certain oxides and other compounds. Antiferromagnets play an important role in modern read heads, but a further discussion goes far beyond the scope of this lecture.

Chapter 10

Electromagnetic waves

In the previous part of this script we have looked at static and slowly varying E- and B-fields in vacuum and in materials. Here we will look at the coupling between these fields in more detail and find that the solutions of Maxwell's equations can be expressed as a wave equation. One of the main limitations for the discussion in this chapter will be that it has to stay rather superficial. A truly complete and detailed discussion of the encountered phenomena will not only require far more time, but also advanced mathematical knowledge, and concepts that we skipped in the previous sections to keep the material easy to digest.

10.1 The wave equation

In the previous course on physics the harmonic oscillator has been introduced, with a prominent example being the mass on a spring. Here we will extend this to a long chain of N masses (m) and springs (spring constant k) as illustrated in Figure 10.1. The chain runs along the x direction and distance between the masses is b. For the mass at position $x+b$ we can now derive the equation of motion. The spring force experienced by this mass is proportional to the distance to the next mass and thus their displacements u

$$F_{spring} = k[u(x+2b) - u(x+b)] - k[u(x+b) - u(x)]$$

If we put this in Newton's third law $F = ma = m\frac{\partial^2 u(x+b)}{\partial t^2}$ and simplify we obtain

$$m\frac{\partial^2 u(x+b)}{\partial t^2} = k[u(x+2b) - 2u(x+b) + u(x)]$$

Now we can rewrite this in properties of the full chain, the length $L = Nb$, mass $M = Nm$ and spring constant $K = k/N$:

$$m\frac{\partial^2 u(x+b)}{\partial t^2} = \frac{KL^2}{M}\left(\frac{u(x+2b) - 2u(x+b) + u(x)}{b^2}\right)$$

If we now take the limit that $b \to 0$ but keeping the length, mass, and spring constant of the chain constant, we see that the term between brackets on the right side of the equation is the definition of the second derivative of u with respect to x. The equation of motion for a single point on this chain now becomes

$$\frac{\partial^2 u}{\partial t^2} = \frac{KL^2}{M}\frac{\partial^2 u}{\partial x^2}$$

.

Figure 10.1: Long chain of masses and springs.

The equation we have derived describes the motion of a piece of mass in a long spring-like chain, which could also be a rope. We can generalise this expression by replacing the proportionality constants by c^2, for reasons that will become clear below. This yields the **1D wave equation**

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2} \tag{10.1}$$

The solution will be dependent on both space and time: $u(x,t)$.

By substitution into Eq. 10.1 it can be easily verified that any function f of $x - ct$ is a solution to the wave equation, thus $u(x,t) = f(x-ct)$. As illustrated in Figure 10.2 this function represents that anything which is at the position x at time $t = 0$ will have moved a distance ct along the positive x axis during time t. It is thus a signal package moving with velocity c along the x-axis.

Along similar lines it can be shown that any function g of $x + ct$ is also a function of Eq. 10.1. Because the system is linear this means that the general solution of the 1D wave equation is the sum of both functions

$$u(x,t) = f(x-ct) + g(x+ct) \tag{10.2}$$

Whereby either of them can be zero. The general solution is therefore a superposition of wave packages moving along positive x and along negative

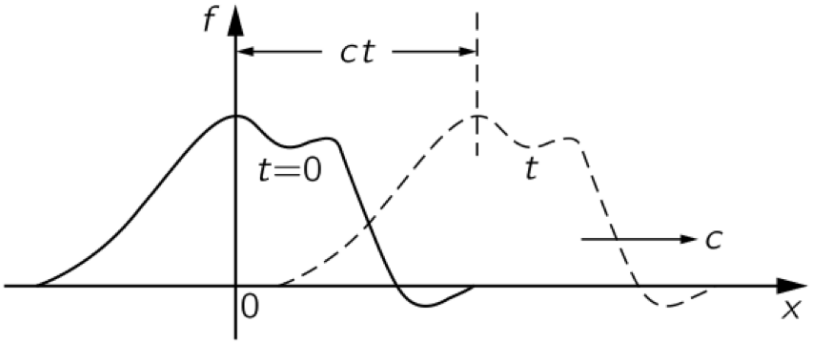

Figure 10.2: General solution to the 1D wave equation.

x both with velocity c. The exact shape of this function will depend on the boundary conditions or what sets things in motion. Under some special conditions the two functions perfectly match and a *standing wave* is formed. In a musical instrument such a standing wave in either a string or air, causes the associated tone.

The above derivation for the wave equation is for a 1D system, but this can be extended to full space including all directions. This gives the **general wave equation in 3D**

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u \tag{10.3}$$

The solutions are again functions representing signals moving in any direction in space with velocity c.

10.2 Extension to Ampere's law

In section 7.5 we have introduced Ampere's circuital law: $\nabla \times \vec{B} = \mu_0 \vec{j}$ (Eq. 7.19 and 7.21), and we have been using it since to solve many problems. However, if we look more closely at this law we will notice that something is wrong with it. We did not have to worry about it before, because we never touched the regime where this incompleteness would play a role.

From a mathematical point of view the problem becomes clear if we take the divergence of both sides of the equation. On the left we get $\nabla \cdot \nabla \times \vec{B}$ which is always zero because the divergence of the curl is zero by definition. On the right hand side we have $\nabla \cdot \mu_0 \vec{j}$, which can be zero, but doesn't have to be. For example it is not zero when a current enters, but doesn't

exit, a closed surface, which happens when we are charging, or discharging, a capacitor.

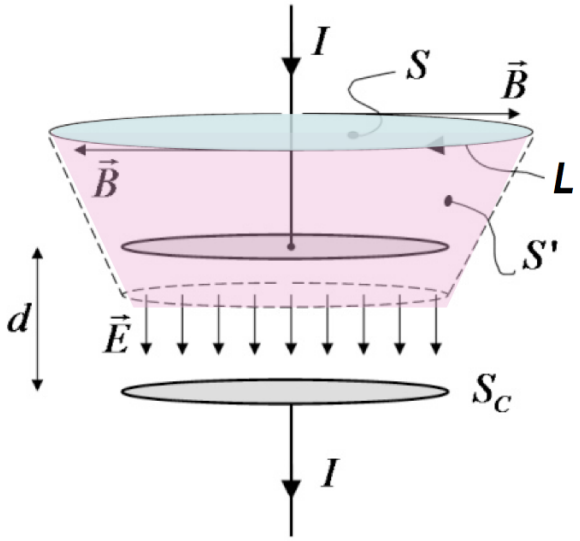

Figure 10.3: Ampere's law when charging a capacitor.

To fix this problem we can consider exactly this situation and look at what happens with Ampere's law when charging a capacitor. The first thing we have to consider is what it means that the closed line integral of the B-field is equal to the current it encircles (times μ_0) in Eq. 7.19. Formally the closed path is the boundary of the surface the current flows through and expressed in Eq. 7.20. For the situation of charging a capacitor illustrated in Figure 10.3 we can define different surfaces S and S' for the closed path L. The current through surface S is I whereas the current through S' is zero. This would mean that the resulting B-field would depend on the surface we choose. This can of course not be the case for a fundamental law.

In order to solve this discrepancy we include an extra term in Ampere's law, the displacement current I_d. Thus

$$\oint_L \vec{B} \cdot d\vec{L} = \mu_0(I_c + I_d)$$

whereby I_c is the current as we considered it before. We see that there is only a paradox when we are charging (or discharging) the capacitor, meaning that it will be related to the change in the E-field in the capacitor. The E-field in the capacitor is $E = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 S_c}$ where Q is the charge on the

capacitor and S_c the surface of the plates. The current I_c brings the charge to the plates and is thus

$$I_c = \frac{dQ}{dt} = \epsilon_0 \frac{d\left(ES_c\right)}{dt}$$

Through surface S' we know that $I_c = 0$ and to make sure that we obtain the same B-field this means that $I_d = \epsilon_0 \frac{d(ES_c)}{dt}$. The E-field times the surface S_c it is perpendicular to, is nothing else than the flux of the E-field, and I_d is thus related to the change in flux through the surface S' between the capacitor plates

$$I_d = \epsilon_0 \frac{d\varphi_E}{dt} = \epsilon_0 \frac{\partial}{\partial t} \iint_{S'} \vec{E} \cdot d\vec{S'}$$

If we combine this with our initial assumption and if we generalise for any surface S that L can be a boundary off, we obtain the **complete version of Ampere's circuital law**

$$\oint_L \vec{B} \cdot d\vec{L} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \iint_S \vec{E} \cdot d\vec{S} \tag{10.4}$$

or, using Eq. 7.20

$$\oint_L \vec{B} \cdot d\vec{L} = \mu_0 \iint_S \vec{j} \cdot d\vec{S} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \iint_S \vec{E} \cdot d\vec{S} \tag{10.5}$$

In differential form this becomes

$$\nabla \times \vec{B} = \mu_0 \vec{j} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \tag{10.6}$$

It is left up to the reader to verify that this also solves the mathematical problem indicated above.

10.3 Electromagnetic waves in vacuo

In vacuo and in the absence of sources ($\rho = 0$ and $j = 0$) the differential form of the Maxwell equations now reduces to

$$\nabla \cdot \vec{E} = 0 \tag{10.7}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{10.8}$$

$$\nabla \cdot \vec{B} = 0 \tag{10.9}$$

$$\nabla \times \vec{B} = \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \tag{10.10}$$

When combining these equations with the vector identity

$$\nabla \times \nabla \times \vec{A} = \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}$$

it is possible to rewrite the equations eliminating either $\vec{B}$ or $\vec{E}$ as shown during the lecture. This results in the following expression for the E-field

$$\frac{\partial^2 \vec{E}}{\partial t^2} = \frac{1}{\epsilon_0 \mu_0} \nabla^2 \vec{E} \tag{10.11}$$

And an equivalent expression for the B-field

$$\frac{\partial^2 \vec{B}}{\partial t^2} = \frac{1}{\epsilon_0 \mu_0} \nabla^2 \vec{B} \tag{10.12}$$

From a comparison with Eq. 10.3 one can directly recognise that these expressions represent 3D waves of E and B fields. The velocity c of these waves is the speed of light and is given by

$$c = \sqrt{\frac{1}{\epsilon_0 \mu_0}} \approx 3 \times 10^8 \text{ m/s} \tag{10.13}$$

Which means that the *speed of light can be determined from electrical measurements!*

The solution of Equations 10.11 and 10.12 depends on the boundary conditions, such as how the wave was created. The general solution will be of the form given in Eq. 10.2 and can represent anything from a short pulse to a nice sinusoidal signal. One could easily write hundreds of pages and give several courses about the properties of such electromagnetic (EM) waves, but this goes far beyond the scope of this course and we will restrict ourselves to some simplified aspects.

Before discussing these properties is it important to consider the **electromagnetic spectrum** in Figure 10.4. This indicates that all electromagnetic radiation, ranging from radio waves, via infrared heat transfer and visible light, to high energy X-rays, all are part of the same family. They are described by exactly the same equations and they all travel with the same velocity c. The difference lies in the frequency ν of the oscillation of the E-field, and in the distance between maxima in the field, which is the wavelength λ. This wavelength can range from hundreds of kilometres in long radio waves, to subatomic distances in γ-rays. Visible light has a range of about 400 to 700 nm.

The product of λ and ν is constant and is the velocity c

$$\lambda \nu = c \tag{10.14}$$

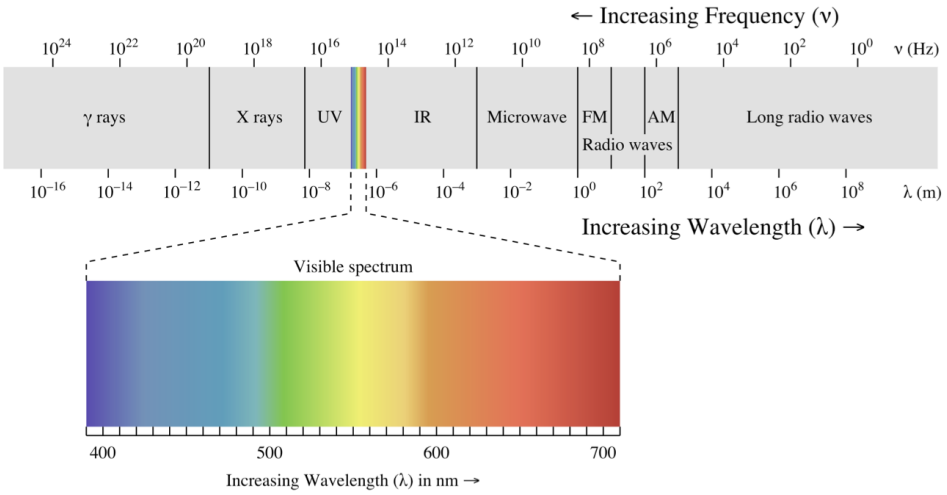

Figure 10.4: The electromagnetic spectrum.

Thus radio waves have a low frequency, megahertz in commercial radio, and X-rays have a high frequency. According to the wave-particle dualism one can also talk about a light particle; the photon. The frequency of the wave is directly linked to the **energy of the photon**, and thus also to the wavelength

$$E = h\nu = \frac{hc}{\lambda} \tag{10.15}$$

Here h is *Planck's constant* given as

$$h = 6.626 \times 10^{-34} \text{ Js}$$

Often the energy is expressed in electronvolt (eV) instead of Joule (J) whereby 1 eV $= 1.6 \times 10^{-19}$ J, which is the energy of an electron accelerated by 1 volt potential difference. The advantage of this scale is that the numbers are more easy to work with and that 1 eV is around the energy of a red photon. X-rays and γ-rays have energies of tens of kiloelectronvolts and are therefore called high energy radiation.

10.4 Monochromatic plane waves

As indicated in the previous section we have to restrict ourselves to a certain class of electromagnetic waves in this lecture. These are *monochromatic*

plane waves. The monochromatic means we are dealing with only one colour, or better only one energy/frequency/wavelength. Plane wave means that the wave has a wavefront only perpendicular to the direction it travels. This is similar to a straight wave at a long beach, and this situation is obtained if we look far away from the source.

The E- and B-fields of the wave still have to be solutions to Maxwell's equations in vacuo without sources. We first assume solutions of the form $\vec{E} = E(x, y, z; t)\hat{y}$ and $\vec{B} = B(x, y, z; t)\hat{z}$ for a wave travelling along the $\hat{x}$-direction. In the lecture it is shown that if we now explicitly write out the differential form of Maxwell's equations and realise which partial derivatives become zero the equations reduce to

$$\vec{E} = E(x; t)\hat{y} \quad \text{and} \quad \frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \tag{10.16}$$

and

$$\vec{B} = B(x; t)\hat{z} \quad \text{and} \quad \frac{\partial B}{\partial x} = -\epsilon_0\mu_0\frac{\partial E}{\partial t}$$

Thus The E- and B-field are perpendicular to each other and the spatial dependence of E is coupled to the time dependence of B, and vice versa.

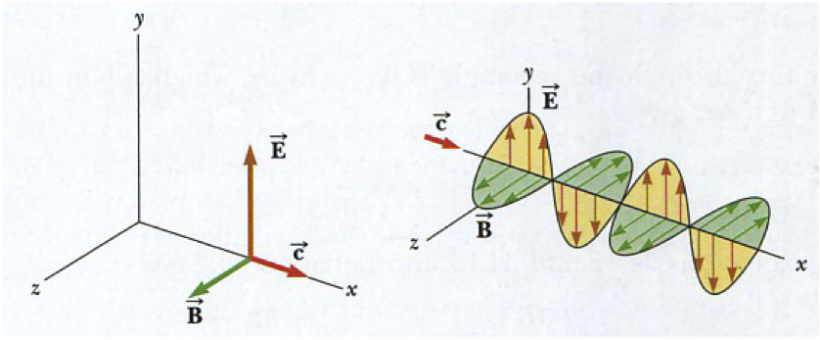

Figure 10.5: Illustration of the E- and B-fields of a monochromatic plane wave travelling along the x-axis

The propagation direction is typically refereed to as the **Poynting vector** $\vec{S}$ or $\vec{c}$, which is also the direction in which the energy transfer occurs. The E- and B-field oscillation directions are related to the Poynting vector as

$$\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B} \tag{10.17}$$

As illustrated in Figure 10.5. More formally it should be the H-field instead of the B-field here, but at this point this would only complicate matters.

We can now try solutions of the form $f(t - \frac{x}{c}) = \sin(-\omega(t - \frac{x}{c})) = \sin(kx - \omega t)$ Whereby $\omega = 2\pi\nu$ is the angular frequency and $k = \frac{2\pi}{\lambda}$ is the wave number. We thus get solutions

$$E_y = E_{y0}\sin(kx - \omega t) B_z = B_{z0}\sin(kx - \omega t + \phi)$$

With ϕ a phase difference between the E- and B-field oscillations. If we enter this in Eq. 10.16 we obtain the following

$$kE_{y0}\cos(kx - \omega t) = \omega B_{z0}\cos(kx - \omega t + \phi)$$

From which it directly follows that ϕ has to be zero and the **E- and B-field oscillate in phase**. Furthermore we can determine the aptitude ratio of the two fields to be

$$\frac{E_{y0}}{B_{z0}} = \frac{\omega}{k} = \lambda\nu = c$$

And thus

$$E_y = cB_z \tag{10.18}$$

Which means that the E-field is c times, or 3×10^8 times larger than the B-field for any electromagnetic wave. Therefore, most of the properties and influence on the environment are related to the E-field.

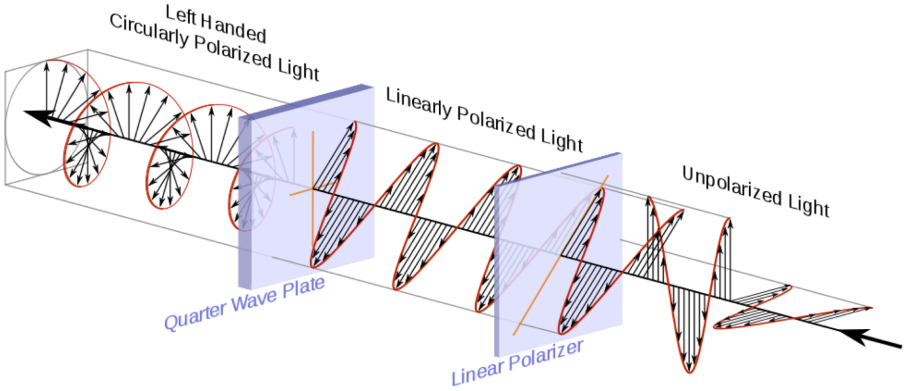

Figure 10.6: Illustration of the E-field oscillation direction for different polarisation directions.

Although the E- and B-field have to be perpendicular to each other, and to the propagation direction, they can change direction together as long as

Eq. 10.17 is satisfied. The oscillation direction of the E-field is referred to as the **polarisation of the electromagnetic wave**. Some possible light polarisations are indicated in Figure 10.6. If the E-field oscillation direction, and consequently also the B-field, randomly changes direction, the wave is unpolarised. After passing the wave through a polariser the E-field only oscillates along one direction and the wave is linearly polarised. By creating a superposition of such a wave with a phase shifted copy, complex polarisation patterns can be obtained, of which a circular polarisation is a good example. Here the E-field oscillation direction continuously changes direction in a screw-like fashion.

10.5 EM waves in non-conducting media

Some of the power of the Maxwell equations comes from the fact that they are independent on the origin of the electric and magnetic field. This means that all the equation we have derived in previous chapters also apply to electromagnetic waves.

We know that in a dielectric the E-field is changed and that we have to replace ϵ_0 by $\epsilon_r\epsilon_0$. Similarly we have to replace μ_0 by $\mu_r\mu_0$ in all equations. Therefore, the velocity of the EM wave in a non-conducting medium c_m changes to

$$c_m = \frac{1}{\sqrt{\epsilon_r\epsilon_0\mu_r\mu_0}} = \frac{c}{\sqrt{\epsilon_r\mu_r}} \tag{10.19}$$

If we now restrict ourselves to the by far larger class of non-ferromagnetic materials, $\mu_r \approx 1$ and thus $c_m \approx \frac{c}{\sqrt{\epsilon_r}}$. We have seen that ϵ_r can range from 1 to several tens of thousands, which will thus significantly alter the velocity of the EM wave.

This change in velocity causes a refraction of the wave at an interface between materials with different ϵ_r. The **refractive index** n, which should be familiar from high-school physics, is determined by the ratio of the velocity with respect to the speed of light in vacuum

$$n = \frac{c}{c_m} \approx \sqrt{\epsilon_r} \tag{10.20}$$

It is thus possible to approximately measure the optical refractive index of a material by placing it in a capacitor and measuring the change in capacitance.

As explained during the lecture, the relative permittivity is typically not a constant, but a non-linear tensor. In particular it will depend on the

wavelength/frequency/energy of an EM wave, which is related to processes happening on an atomic scale and how easily dipoles are induced. As a consequence also the refractive index will change with wavelength, and thus with colour, leading to fascinating phenomena such as the rainbow.

Many "optical" phenomena, including wave guiding and birefringence, can now be explained by the exact form and dependency of ϵ_r, including its anisotropic behaviour. It should be realised that by choosing and combining materials with corresponding ϵ_r an almost complete control over EM wave properties over interfaces can be achieved.

10.6 Generation of electromagnetic waves

In the previous sections we have discussed some of the properties and behaviour of electromagnetic waves assuming that they are there. Here we will consider how they can be generated. This could be the topic for a full course or even a study, so we will only scrape the surface here. Especially sources based on the recombination of electrons and holes and subsequent emission of a photon, such as light-emitting diode (LED) and LASER, require an in-depth understanding of solid state physics.

We have seen that in the absence of sources, $Q = 0$ and $I = 0$, there is only propagation and no generation of EM waves. If the charge is constant, then $I = 0$ and we have a steady E-field. If Q moves at constant speed, then the current is constant and the B-field is steady. However, we need that both the E-field and the B-field change as a function of time, so we need a current that changes with time, and thus **accelerating charges**.

In principle *any acceleration of a charge* causes an EM wave, whereby the properties of the wave directly depend on the form of the acceleration. In bremsstrahlung very fast electrons are suddenly stopped in a target, resulting in the emission of X-rays. But also changes in direction are accelerations, which is the working principle of synchrotron radiation.

In a more formal description we can consider that Coulomb's law (Eq. 3.2) needs to be corrected for the fact that the E-field will only change with the speed of light. The first step is to realise that not the position $\vec{r}$ at this moment is relevant, but the position $\vec{r'}$ a time $\frac{r'}{c}$ ago. Furthermore, we need to take into account what happened with the charge in the mean time. Now the E-field becomes

$$\vec{E} = \frac{q}{4\pi\epsilon_0}\left[\frac{\hat{r}'}{r'^2} + \frac{r'}{c}\frac{d}{dt}\left(\frac{\hat{r}'}{r'^2}\right) + \frac{1}{c^2}\frac{d^2\hat{r}'}{dt^2}\right] \tag{10.21}$$

Which for a steady charge reduces back to Eq. 3.2. In the next chapter a derivation of Eq. 10.21 is presented. This has been done by two students

of the lecture and thus slightly deviates in terminology, but is correct in content.

For the generation of an EM wave only acceleration is relevant so we only look at the last term of Eq. 10.21

$$\vec{E}_{EM} = \frac{q}{4\pi\epsilon_0 c^2}\frac{d^2\hat{r}'}{dt^2}$$

Furthermore, only the component of the acceleration perpendicular to our line of sight, here a_x, can generate waves that reach us. Thus the expression for the **E-field of an accelerating charge** reduces to

$$E_x(t) = \frac{q}{4\pi\epsilon_0 c^2 r}a_x(t - \frac{r}{c}) \tag{10.22}$$

Where it is explicitly indicated that we have to consider the acceleration at a time $\frac{r}{c}$ ago. The corresponding B-field of the EM wave becomes

$$B_y(t) = \frac{q}{4\pi\epsilon_0 c^3 r}a_x(t - \frac{r}{c}) \tag{10.23}$$

In the generation of electromagnetic waves the **oscillating electric dipole** plays a central role. That such an oscillation is automatically accompanied by the acceleration of charges should be clear. Along similar lines as the derivation for a single charge the E-field as a function of time for the dipole can be obtained

$$\vec{E}(r,t) = \frac{1}{4\pi\epsilon_0 r^3 c^2}(\ddot{\vec{p}}(t - \frac{r}{c}) \times \vec{r}) \times \vec{r} \tag{10.24}$$

Where $\ddot{\vec{p}}$ is the second derivative of the dipole moment with time, considered at a time $\frac{r}{c}$ ago. If we look from far away, we only need to consider the radial component which becomes

$$E_{rad}(r,t) = \frac{\ddot{p}(t - \frac{r}{c})\sin\theta}{4\pi\epsilon_0 r c^2} \tag{10.25}$$

Here the electric dipole moment as a function of time can for example be $\vec{p}(t) = q\vec{l}\sin(\omega t)$. The polarisation of the EM wave emitted from an oscillated dipole is along the direction of oscillation.

The reason why the oscillating dipole is important, is because it is the working principle of **antenna**. Here a generator drives a alternating current along different directions of a metallic rod. Together this forms a large

charge oscillation. The emitted electromagnetic waves have the same frequency as with which the current is driven, and thus the corresponding wavelength. The polarisation will be along the direction of the rod. Maximum power is achieved when the length of the antenna is approximately $\lambda/2$. This radiation will cause a similar oscillating current in a second metallic rod placed at a certain distance at any orientation which is not perpendicular to the first antenna. This current is easily detected and a signal is thus transferred.

10.7 Power of electromagnetic waves

For wireless transmission of signals, but also for example to estimate the possible damage, it is important to consider the power that is transmitted by radiation. In the following the average power, thus the average energy transferred per unit of time, of an oscillating electric dipole far away from the source is calculated.

The dipole oscillation can be described by $\vec{p}(t) = \vec{p_0}\sin(\omega t)$. For the power the time offset and the exact orientation are not of importance and from the previous section we obtain the following expressions for the E- and B-field far away from the source

$$E(t) = \frac{\ddot{p}\sin\theta}{4\pi\epsilon_0 r c^2} \tag{10.26}$$

$$B(t) = \frac{\ddot{p}\sin\theta}{4\pi\epsilon_0 r c^3} \tag{10.27}$$

Whereby the B-field is perpendicular to the E-field. As indicated with the introduction in Eq. 10.17, the Poynting vector describes the energy transfer. Inserting the two equations above and using that $c^2 = \frac{1}{\epsilon_0\mu_0}$ the following expression for $\vec{S}$ is obtained

$$\vec{S} = \frac{[\ddot{p}]^2\sin^2\theta}{16\pi^2\epsilon_0 r^2 c^3}\hat{r} \tag{10.28}$$

The power is the energy integrated over space: $P = \oint \vec{S}\cdot d\vec{A}$. And for a sphere ($4\pi r^2$) one thus obtains

$$P = \frac{2\,[\ddot{p}]^2\sin^2\theta}{12\pi\epsilon_0 c^3}. \tag{10.29}$$

For the average power we need the average of the square of the second derivative of the dipole moment

$$\left\langle [\ddot{p}]^2 \right\rangle = \frac{\omega^4 p_0^2}{2} \tag{10.30}$$

The **average power emitted by an oscillating dipole** thus becomes

$$\langle P \rangle = \frac{\omega^4 p_0^2}{12\pi\epsilon_0 c^3}. \tag{10.31}$$

From this is directly follows that the oscillation frequency, and thus also the frequency of the radiation is the most important factor in determining the power. Low frequency radiation, such as radio waves (10^7 Hz), contain much less power compared to high frequency radiation such as X-rays (10^{18} Hz). Also the power of the radiation emitted by the typical 60 Hz of an AC network is negligible. Although Eq. 10.31 has been derived for EM waves emitted by an oscillating dipole the result is similar for any other source of electromagnetic waves.

Chapter 11

Deriving Feynman equation for a retarded electrical field

This chapter is written by Linley Vion, Maxime Nemoz, and colleagues in December 2022. The goal was to provide a background to the rather ad-hoc presentation of Eq. 10.21 during the lecture and in the lecture notes. I thank them sincerely for their nice work and present the text without further alterations.

11.1 Introduction

In the universe, every entity seems to move at a finite speed. Whether it is stars, rockets or even light. Let us consider light as an electromagnetic wave and the speed of light (in vacuum), the highest amongst every particle/wave in the Universe [1]. It wouldn't make sense to assume that an electrical field, which propagates as an electromagnetic wave, would face a different reality. So it is to be considered that electrical fields take time to propagate through space. As presented in Feynman Lectures on Physics Volume I and II, an approximation for the equation of a retarded electrical field is given in simplified and understandable terms (at this point in the book) [2]. But unfortunately, no demonstration of this equation was found and the explanation of each term seemed not precise enough to allow one to derive the equation just from its interpretation. On top of that, the equation for retarded potentials that could have been derived looked too far in their notation to be directly linked to Feynman's equation. Hence, in this constant need for explanations, this equation will be derived and explained.

11.2 The retarded electrical field equation

Presenting the equation To begin with, the equation that was evoked in the introduction is the following:

$$\vec{E} = -\frac{q}{4\pi\epsilon_0}\left[\frac{\hat{r}'}{r'^2} + \frac{r'}{c}\frac{d}{dt}\left(\frac{\hat{r}'}{r'^2}\right) + \frac{1}{c^2}\frac{d^2\hat{r}'}{dt^2}\right]_{t_r} \tag{11.1}$$

This equation describes the electrical field produced by a moving point charge in vacuum, with a speed $v \ll c$. The electrical charge of the point charge is q, ϵ_0 is the vacuum permittivity and $\hat{r}'$ is the unitary vector that points towards the position of the point charge at $t_{r'}$, the time where the field was "emitted", with length r'. Finally, the minus sign in front of the equation can be explained as such: the $\vec{E}$-field is emitted towards the observer but expressed by vectors pointing at the charge (see Fig. 11.1). It is then logical to invert their direction, using a minus.

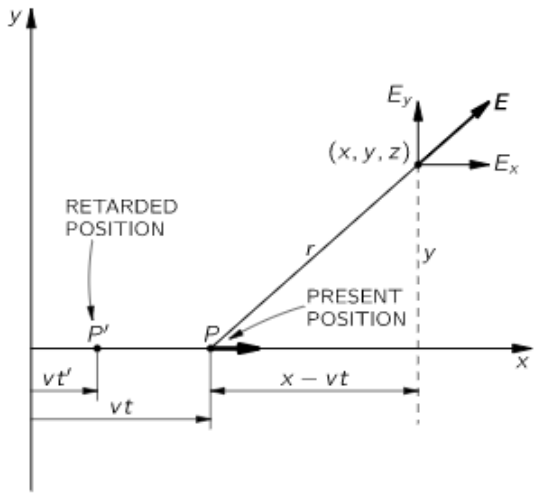

Figure 11.1: For a moving charge, the electric field points radially from the present position of the charge [3].

Deriving the equation Now, the equation will be derived using a limited development to the second order of the $\vec{E}$-field to make the second and the third term appear because the advised eye of the reader will have noticed that the first term corresponds to Coulomb's law. The development is :

$$\vec{E}_{t_r} = \vec{E}_{t_{r'}} + \Delta t\frac{d}{dt}\vec{E}_{t_{r'}} + (\Delta t)^2\frac{d^2}{dt^2}\vec{E}_{t_{r'}} \tag{11.2}$$

with Δt the time the point charge had to move. It corresponds to the time the information emitted from the point charge took to reach the observer. Considering r' the distance at that time between the point charge and the observer and c the speed of propagation of the $\vec{E}$-field in vacuum: $\Delta t = r'/c$. Then the three terms that will be developed separately become :

$$\vec{E}_{t_r} = \vec{E}_{t_{r'}} + \frac{r'}{c}\frac{d}{dt}\vec{E}_{t_{r'}} + \left(\frac{r'}{c}\right)^2 \frac{d^2}{dt^2}\vec{E}_{t_{r'}} \tag{11.3}$$

The first term is quite simple and can be expressed without further calculations :

$$\vec{E}_{t_{r'}} = -\frac{q}{4\pi\epsilon_0}\frac{\hat{r}'}{r'^2} \tag{11.4}$$

Now, the second term just requires to be fully expressed and can be found in Eq.(11.1) without further calculations :

$$\frac{r'}{c}\frac{d}{dt}\vec{E}_{t_{r'}} = -\frac{r'}{c}\frac{d}{dt}\left(\frac{q}{4\pi\epsilon_0}\frac{\hat{r}'}{r'^2}\right) = -\frac{r'}{c}\frac{q}{4\pi\epsilon_0}\frac{d}{dt}\left(\frac{\hat{r}'}{r'^2}\right) \tag{11.5}$$

It follows from the fact that q is a point charge, so it has a constant charge. The third term requires more calculations, which will be realised now. The physical assumptions/arguments will be explained in time. The first one is, once again, considering that the only term considered time-dependent will be the distance to the point charge and the unitary vector. The term containing the permittivity and the charge will be omitted here, to simplify the calculations :

$$\begin{aligned}\frac{d^2}{dt^2}\left(\frac{\hat{r}'}{r'^2}\right) &= \frac{d}{dt}\left(\frac{\frac{d\hat{r}'}{dt}r'^2 - 2r'\frac{dr'}{dt}\hat{r}'}{r'^4}\right) \\ &= \frac{1}{r'^2}\frac{d^2\hat{r}'}{dt^2} - \frac{2}{r'^3}\left(\frac{d^2r'}{dt^2}\hat{r}' + 2\frac{dr'}{dt}\frac{d\hat{r}'}{dt}\right) + \frac{6}{r'^4}\left(\frac{dr'}{dt}\right)^2\hat{r}'\end{aligned}$$

It gives when multiplied by Δt^2 :

$$\frac{1}{c^2}\frac{d^2\hat{r}'}{dt^2} - \frac{2}{r'c^2}\left(\frac{d^2r'}{dt^2}\hat{r}' + 2\frac{dr'}{dt}\frac{d\hat{r}'}{dt}\right) + \frac{6}{r'^2c^2}\left(\frac{dr'}{dt}\right)^2\hat{r}' \tag{11.6}$$

The two terms depending on $1/r'$ and $1/r'^2$ respectively are considered higher terms of development. Indeed, the limit of these terms when divided by r'^2 is 0, not because of the $1/r'$ or $1/r'^2$ but because of the first and second derivatives contained in those terms. When the radius variation

tends to 0 (being equivalent to $\Delta t \to 0$), the derivatives tend to 0 faster, allowing these terms to be contained in $o(\Delta t^2)$. Using Eq.(11.6) the third and last term of Eq.(11.3) becomes :

$$-\frac{1}{c^2}\frac{q}{4\pi\epsilon_0}\frac{d^2\hat{r}'}{dt^2} + o(\Delta t^2) \tag{11.7}$$

The equation for the retarded electrical field is then :

$$\vec{E} = -\frac{q}{4\pi\epsilon_0}\left[\frac{\hat{r}'}{r'^2} + \frac{r'}{c}\frac{d}{dt}\left(\frac{\hat{r}'}{r'^2}\right) + \frac{1}{c^2}\frac{d^2\hat{r}'}{dt^2}\right] + o(\Delta t^2) \tag{11.8}$$

where $o(\Delta t^2)$ is neglected in Eq.(11.1). This demonstrate Eq.(11.1).

Explanation and meaning of each term The first term, as stated before, corresponds to Coulomb's Law. It is the expression of the electric field for a point charge q seen many times before. The second term describes how the field will have changed (or here how much did the source move) during the time it took for the field to reach the observer. It is why this term can be found equal to the variation in time (Δt) times the variation of the field through time. The third term is the contribution to the electric field due to the acceleration of the point charge. This acceleration is creating electromagnetic waves (or electromagnetic radiation also called EMR).

11.3 Conclusion

Please note that the considerations made here are not perfectly rigorous but rather coming from a physical understanding of the situation for a point charge moving way slower than c, allowing us to compose the $o(\Delta t^2)$ as it was done in Eq. (11.7). Nevertheless, the equation, under the considerations stated above, represents a good approximation of the retarded electric field. The notation makes it easier to use and understand when compared to retarded potential derived equation for both the electric field and the magnetic field [4][5].

Chapter 12

Special relativity

There are various accounts on the origin and development of the theory of special relativity. Some give all the credit to Albert Einstein. Others, including Einstein himself, highlight that after the completion and interpretation of Maxwell's equations, it was only a matter of logical deduction to arrive at special relativity. In this lecture we will follow the second approach and thus start with the problems posed by Maxwell's equations and how the resolution of these problems led to the theory of special relativity.

The first problem of Maxwell's equations (Eq. 1-4) is that they are **not invariant under Galilean transformation**, which is in stark contrast to the other physical laws known at this point and especially with respect to Newton's laws. The second problem is that electromagnetic waves, or light, requires no medium to travel in and that the speed of light is independent of the reference frame. Also this was completely different from any other known wave phenomena, like sound waves or waves in fluids. Alterations to Maxwell's equations were proposed to solve these problems, but these changes all predicted new phenomena that were not observed. Furthermore, many experiments and especially the **Michelson-Morley interferometer** confirmed that the speed of light is independent of the reference frame and that either the earth has no velocity relative to the medium (ether) or that no medium is needed.

The Galilean transformations for a reference frame moving with velocity u along the x-direction are

$$x' = x - ut,\ y' = y,\ z' = z,\ t' = t \tag{12.1}$$

where the prime indicates observables in the moving reference frame. Hendrik Lorentz noted that if one instead uses the transformations below, then

Maxwell's equations are invariant

$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}},\ y' = y,\ z' = z,\ t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} \tag{12.2}$$

where c is the speed of light in vacuum. These are known a the **Lorentz transformations** and their peculiarity is that the space and time coordinates become mixed. This was considered a curiosity by many, but Poincaré pointed out that there are no coincidences in physics and this stimulated Einstein to state that **all laws are unchanged under Lorentz transformation**. This means that also Newton's equations have to be adjusted. Similarly, while others considered alternative solutions, Einstein simply postulated that **there is no ether and the speed of light is the same for all inertial observers**.

As a reminder, an **inertial reference frame** is a reference frame which is not accelerating and thus moving with constant velocity. In such a refernce frame it is impossible to perform any experiment or observation to find out that it is is moving.

In the following sections we will discuss some of the consequences of these postulates and how they lead to a new notion of space and time. Here it is interesting to note that already Leibniz disagreed with the Newtonian idea of time as something absolute and universal, and was the opinion that also time should be treated as something relative and on the same footing as the spatial dimensions.

12.1 Time dilation and simultaneity

Let us consider a clock illustrated in Figure 12.1 based on an emitter/detector of a light pulse on one side and a mirror opposite to this at a distance L from the source/detector. A time unit is defined as the time it takes for a light pulse to go from the emitter, be reflected by the mirror, and be detected. For a stationary clock, or in the reference frame S' of a clock moving with velocity u along $\hat{x}$, the total time is $(\Delta t =)2t = \frac{2L}{c}$, whereby the Δ will be partially dropped in the following.

On the other hand, for a stationary observer, the mirror will have moved to the right while the light pulse is travelling from the source to the mirror. If we call the time it takes for the light pulse to reach the mirror t' then the mirror will have moved the distance ut'. Applying Pythagoras to the triangle formed by the source to mirror spacing L, light path ct', and clock movement we obtain

$$(ct')^2 = L^2 + (ut')^2$$

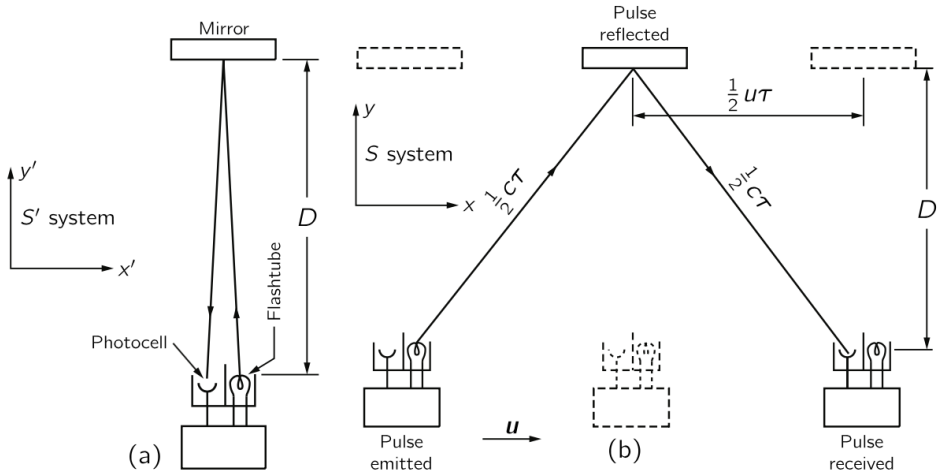

Figure 12.1: (a) A light clock at rest in the S' system. (b) The same clock, moving through the S system. From Feynman lectures.

and thus

$$t' = \frac{L}{\sqrt{c^2 - u^2}}$$

which can be rewritten as

$$t' = \frac{L/c}{\sqrt{1 - u^2/c^2}} \tag{12.3}$$

The same argument can be applied to the path from the mirror to the detector and the total time thus becomes $2t'$. Comparing this to the time in the reference frame of the clock we obtain

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - u^2/c^2}} = \gamma \Delta t \tag{12.4}$$

Thus the stationary observer will claim that the moving clock runs slower as the clock in her reference frame. In other words **moving clocks runs slow**. This is valid for any type of clock and also for biological or fundamental processes. If not, one would be able to detect from the difference in time on difference clocks that one is moving. It is thus really time that is slowing down in a moving frame with respect to the stationary one. Of course, the moving observer will claim that the stationary clocks run slow.

This time dilation is routinely applied to fast moving satellites where it needs to be taken into account that their electronic clock moves slower as

the ones on earth. It can also be observed in the decay of particles such as muons. These have only an average life time of 2×10^{-6}s but can travel several kilometres because their internal clock runs slow.

In Eq. 12.4 the factor γ was introduced because we will encounter this factor many times. It is defined as

$$\gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \tag{12.5}$$

Note that by definition $\gamma \geq 1$ and approaches unity for $u \ll c$.

A rather counterintuitive consequence of the fact that the speed of light is the same for all observers is that **events that are simultaneous in one inertial frame are not necessarily simultaneous in another**. This directly follows from the Lorentz transformations but can also be seen by considering two light pulses that are emitted at the same time from the centre of a box in the direction of motion of the box and the opposite direction. In the inertial frame of the box, the two pulses will hit the edges of the box at the same time. However, for a stationary observer, the front of the box will have slightly moved away from the light pulse and the back of the box will have slightly moved towards the pulse. Thus the distance the pulse needs to travel is shorter for the latter and the event of the light pulse hitting the back of the box takes place before the event of the light pulse hitting the front of the box.

The two observers will thus disagree on the simultaneity of the events. A third observer moving in the same direction of the box, but with higher velocity, will even claim that the order of the events is reversed compared to the stationary observer. It is not the question who is right and who is wrong; all three observers are correct, but the "truth" depends on the reference frame. Although entirely valid in physics, this relativity of truth is often abused in other domains and especially politics. We will come back to this point, not the political one, in our later discussion of space-time intervals.

12.2 Lorentz contraction

Now we consider the same light clock as in the previous section, but turned on its side. Thus the light travels along the same direction as the motion of the clock. The time for a round trip of the light pulse in the reference frame of the moving clock S' is:

$$\Delta t' = 2\frac{L}{c} \tag{12.6}$$

Where L is length in the frame moving with the clock, or when the clock is stationary.

For an external observer in a stationary reference frame S the time to go from the source to the mirror is increased because the latter moves away from the light pulse. Furthermore, we have to use the length as it appears to the external observed L'. The first part of the trajectory thus becomes

$$\Delta t_1 = \frac{L' + u\Delta t_1}{c} = \frac{L'}{c-u}$$

On the other hand, the time between reflection from the mirror and detection is shortened because the detector moves towards the light pulse:

$$\Delta t_2 = \frac{L' - u\Delta t_2}{c} = \frac{L'}{c+u}$$

The total time for a round trip is the sum of both and thus

$$\Delta t = \Delta t_1 + \Delta t_2 = 2\frac{L'}{c}\frac{1}{1-u^2/c^2} = 2\frac{L'}{c}\gamma^2 \tag{12.7}$$

Where γ is defined as in Eq. 12.5.

To be able to compare the expressions for the two inertial frames we have to take the time dilation into account. However, we have to be careful and rewrite Eq. 12.4 to get how the time in S' appears in the stationary reference frame. We thus get that $\Delta t' = \frac{\Delta t}{\gamma}$. Comparing the expressions for Δt and $\Delta t'$ yields

$$\Delta t' = 2\frac{L}{c} = \frac{\Delta t}{\gamma} = 2\frac{L'}{c}\gamma \tag{12.8}$$

And thus for the relation between the stationary length L and how it appears to a stationary observer when the object is moving L':

$$L' = \frac{L}{\gamma} \tag{12.9}$$

This means that moving objects appear shorter, or contracted, in the stationary reference frame. For the observer in S' the length is L and objects in S appear contracted along the direction of travel.

From the Lorentz transformation, here repeated with the γ term,

$$x' = \gamma(x - ut),\ y' = y,\ z' = z,\ t' = \gamma(t - \frac{u}{c^2}x) \tag{12.10}$$

it is directly clear that the dimensions orthogonal to the direction of travel are not affected. This can also be pictured from the idea of drawing a line on a wall from a (very rapidly) moving train. For a stationary train this mark would be at a given hight. If the vertical dimension would be contracted, then from the moving train this mark would be lower. However, from the frame of the train the wall appears to be moving and would thus be contracted and the line would be made higher. The only way to resolve this contradiction is by having **no contraction or expansion in the directions perpendicular to the velocity**.

12.3 Velocity addition

In classical Newtonian physics, velocities are added vectorial. Thus if one throws a ball with 10 m/s in a train that moves with 50 m/s, a stationary observer would say the ball travels with 60 (or 40) m/s, depending on the direction it is thrown. In special relativity this no longer holds as it would lead to objects travelling faster as the speed of light. The velocity addition this becomes slightly more complex.

Let us consider a particle that moves with a velocity $v = \frac{dx}{dt}$ in reference frame S. What would be its velocity v' in a frame S' that moves with velocity u along the x-direction? Using the LT in Eq. 12.10 we can obtain dx' and dt'

$$dx' = \gamma(dx - udt) \tag{12.11}$$

$$dt' = \gamma(dt - \frac{u}{c^2}dx) \tag{12.12}$$

The velocity in S' thus becomes

$$v' = \frac{dx'}{dt'} = \frac{dx - udt}{dt - u/c^2 dx} = \frac{\frac{dx}{dt} - u}{1 - u/c^2 \frac{dx}{dt}} = \frac{v - u}{1 - uv/c^2} \tag{12.13}$$

Which is the **Einstein velocity addition rule**. It can easily be verified that if $v = c$ then also $v' = c$ and thus that the speed of light is the same for all observers.

Due to time dilation also the orthogonal velocity component changes even though $dy' = dy$. In S we have $v_y = \frac{dy}{dt}$ and in S' this becomes

$$v'_y = \frac{dy'}{dt'} = \frac{dy}{\gamma dt} = \frac{v_y}{\gamma} \tag{12.14}$$

Instead of the time dilation in Eq. 12.4 one gets the same result using the LT and setting $dx = 0$, i.e. setting $v_x = 0$. If the parallel component of the velocity is not zero, then we have to take this into account and thus get

$$v'_y = \frac{dy'}{dt'} = \frac{dy}{\gamma(dt - \frac{u}{c^2}dx)} = \frac{v_y}{\gamma(1 - \frac{u}{c^2}v_x)} \tag{12.15}$$

where $v_x = \frac{dx}{dt}$. This is the more general expression often found in textbooks, which with $v_x = 0$ reduces to Eq. 12.14. Note that one should use the γ for the moving frame (u) and not for the particle (v). That the apparent reduction of the speed of light for $v_y = c$ is found in the component parallel to u is left for the reader to verify.

12.4 Geometry of space-time

As mentioned above, one of the peculiarities of the Lorentz transformation is that space and time coordinates are mixed. This is similar to the mixing of the x and y coordinates in a classical rotation by an angle θ:

$$x' = x\cos\theta + y\sin\theta,\ y' = y\cos\theta - x\sin\theta,\ z' = z$$

To understand such a rotation, it is essential to consider all the relevant coordinates (x, y, z) to define a point is space. Along similar lines, the LT drives us to combine space and time in a single coordinate system called **space-time**. Here a point with (x, y, z, t) is called an **event**, which is thus described by a **four-vector**. In order to have the same units for all four coordinates, mostly ct is used instead of t and an event is thus defines by (x, y, z, ct). This is often referred to as Minkowski space-time.

A typical space-time diagram is shown in Figure 12.2. The y and z axes are not drawn for simplicity. They extend orthogonal to both x and ct and each other, forming a rather hard to draw hyper cube. The light line should always have a slope of unity in all such diagrams and this helps us to consider how to draw a moving inertial frame S' in the same diagram. The x' axis should obtain some time contribution and thus be lifted away from the x axis, whereas the time axis should obtain some x component. If we would now just rotate the coordinate system, the light line would no longer have unity slope with regard to x' and ct'. Therefore the time axis ct' has to also rotate towards the light line as illustrated in Figure 12.2.

The coordinates of an event (A) can now be determined in both reference frames, whereby one has to remember to read the coordinates by going orthogonal to the axis in the relevant coordinate system; i.e. parallel to the

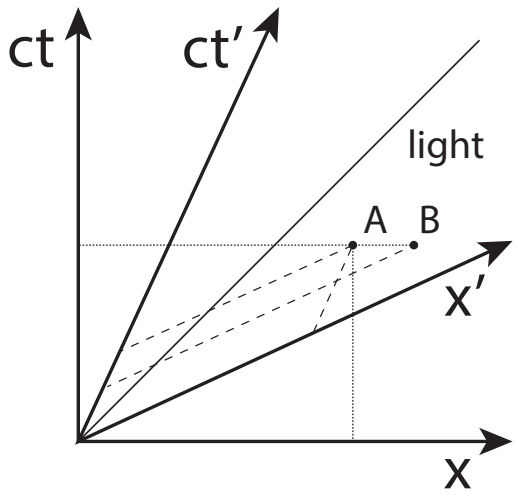

Figure 12.2: Space-time diagram with light line and transformation to moving inertial system S'.

other axis. It also directly becomes visible that two events A and B that are simultaneous in S are not simultaneous in S'.

Although useful to illustrate general ideas, such diagrams are not used much to actually calculate transformations between inertial systems. An important role in such transformations is played by the **space-time interval**. Similar to the distance between two points in "ordinary space" ($\sqrt{x^2+y^2+z^2}$), the space-time interval (or just interval for short) between two events is the same in all coordinate systems. Due to the type of "rotation" illustrated in Figure 12.2 the sign of the time and space coordinates have to be opposite in the summation. We use here the convention of Poincaré where time is positive and space negative, but one can also find the opposite. A space-time interval is then defined as

$$\sqrt{c^2t^2 - x^2 - y^2 - z^2} = \text{constant} \tag{12.16}$$

Here it should be realised that the (x, y, z, ct) coordinates are taken in reference to the first event placed at the origin of the coordinate system. Alternatively all coordinates can be replaced by a displacement; i.e. $t = dt$, $x = dx$ etc.

The sum under the square root in Eq. 12.16 can be positive or negative, and thus the interval can be real or imaginary, or zero. If the sum is negative and the interval imaginary we call it **space-like**. If the interval is real, it is

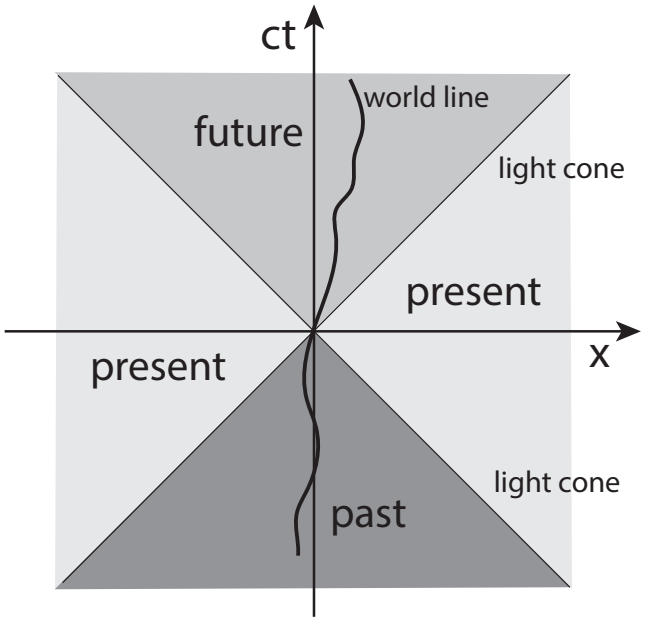

Figure 12.3: Space-time diagram with light cones and world line and different regions indicated.

called **time-like**, and when it is zero then it is **light-like**. The origin of the space-time diagram in Figure 12.3 is the "here and now" and is taken as a reference for the intervals. The time-like intervals are separated in negative and positive time and are called the "past" and "future" respectively. The now can be influenced by the past and can have information about it. The now can be influenced by the now, but it is not possible to have knowledge about it. The space-like intervals are all classified as the "present" and it is neither possible for the now to have information about it, nor to influence it, and vice versa. If an event happens in the present we can only know about it, or be influenced by it, in the future. The zero intervals separate the time- and space-like intervals in form the light cones. That they are cones becomes clear if we also consider the orthogonal space axis.

The **world line** of an object or particle (or person) can move in both spatial directions, but only forward along the time axis. The slope should of course never be less that unity. The world line forms a collection of "nows" and for every point the past, present, and future can be defined.

The classification in time- and space-like intervals has a further important function, namely it allows us to rapidly say how events will look in another reference frame. If the interval is real, and thus time-like, an in-

ertial system exists in which the two events occur at the same point. This is more trivial than it seems and we experience it in every day life. In the reference frame of the train, its departure in one city and its arrival in the next occur at the same point, whereas in the stationary frame of the earth these are events separated in space and time. On the other hand, if an interval is imaginary, and thus space-like, an inertial system exists in which the events happen at the same time. This we already discussed above and is typically not experienced in everyday life.

12.5 Mass and energy

A wide range of experiments, based for example on the relationship between mass and radius in a magnetic field of Eq. 7.11, showed that the mass of a particle increases when it is accelerated to a velocity close to the speed of light. Several early experimental results are shown in Figure 12.4, and more recent experiments allowed to approach velocities around $0.99c$. Thus the **mass increases with velocity** and the relationship is

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = \gamma m_0 \tag{12.17}$$

Here m_0 is the mass at zero velocity, or the **rest mass**.

The same result can also be derived from symmetry considerations in a collision between particles with the same mass and velocity and changing the frame of reference. However, the experimental results leave no doubt about the validity of Eq. 12.17 and are here regarded as sufficient evidence. The main consequence of this increase in mass is that it becomes impossible to accelerate an object with finite rest mass to the speed of light. Almost all energy invested in accelerating the object will go into increasing the mass. Because elementary particles, like electrons and muons, have such low m_0 they can achieve speeds within a fraction of c.

In a first approximation one can now use all the laws of kinematics and mechanics from earlier semesters, but replacing the mass by Eq. 12.17. For example the momentum becomes $\vec{p} = m\vec{v} = \gamma m_0 \vec{v}$. Conservation of momentum still works, but means that both the mass and velocity can change. These changes to Newton's laws don't form the core the of this lecture and will not be treated in detail.

The dependency of mass on velocity also has far reaching consequences for our understanding of energy. For example, an increase of temperature is understood as an increase of velocity of the atoms, which means that their mass increases. To obtain more insight we can make a binomial expansion

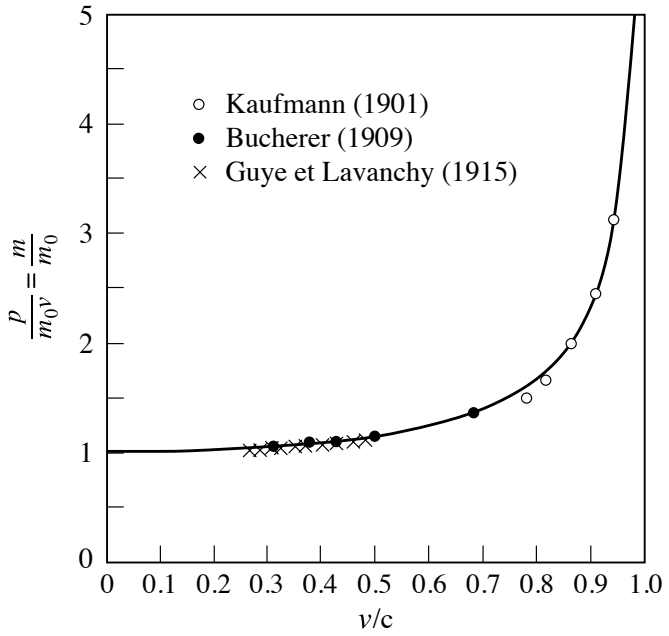

Figure 12.4: Increase of mass versus velocity for several experiments. Solid line according to Eq. 12.17

of Eq. 12.17 and obtain

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = m_0 \left(1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + ...\right) \approx m_0 + \frac{1}{2} m_0 v^2 \left(\frac{1}{c^2}\right) \tag{12.18}$$

Where terms with c^{-4} or higher are ignored in the last step. The last term is the kinetic energy divided by c^2 and we can thus say that an increase in mass is an increase in energy (divided by c^2)

$$\Delta m = \Delta E_{kin} \left(\frac{1}{c^2}\right).$$

Einstein considered it illogical that only the increase in mass should be related to energy and thus postulated his famous expression on the equivalence of mass and energy

$$E = mc^2 \approx m_0 c^2 + \frac{1}{2} m_0 v^2 \tag{12.19}$$

where Eq. 12.18 is used in the last step. The first term $m_0 c^2$ is often referred to as the rest energy, whereas the second one is the well known

kinetic energy. Thus a change in energy can be related to a change in kinetic energy, but also to a change in rest mass m_0. This is experimentally observed in annihilation experiments where all mass is turned into energy in the form of EM waves, but also in nuclear fission where the generated energy is the same as the mass lost (times c^2).

Before making the link back to the previous chapters on electromagnetism, just one last remark on the **difference between invariant and conserved** quantities. Invariant means that the quantity is the same in all inertial systems, whereas conserved means that it is the same before and after some process. For example, rest mass is invariant, but not conserved; energy is conserved, but not invariant; velocity is neither invariant nor conserved; and charge is both invariant and conserved. This last point we will use in the next section.

12.6 Relativistic electrodynamics

From the previous chapters we know that a moving charge creates an E- and B-field for a stationary observer. However, in the reference frame moving with the charge there will only be an E-field. Because Maxwell's equations are Lorentz invariant there is no conflict, nor do we need to add new terms. It just means that $\vec{E}$ and $\vec{B}$ are the same thing, but just with a different name depending on the inertial system. Another way of putting it is that $\vec{E}$ transforms into $\vec{B}$ and vice versa. In the lecture an example will be given how the law of Biot-Savart can be obtained from considering the rest frame of a moving charge and the relativistic transformation of the E-field of a neutral wire. Here we will directly jump to the generalisation of such transformations.

The fundamental idea of using fields is that their origin is irrelevant. We can thus take the simplest example and the obtained results will be valid for all fields. For the E-field the simplest example is the parallel plate capacitor. As expressed in Eq. 3.5 and further discussion, the E-field between two oppositely charged plates is $E = \frac{\sigma}{\epsilon_0}$ and perpendicular to the plates. Thus for the capacitor illustrated in Figure 12.5 the field is

$$\vec{E_0} = \frac{\sigma_0}{\epsilon_0}\hat{y} \tag{12.20}$$

with $\sigma_0 = |\sigma_+| = |\sigma_-|$. This is in the rest frame of the capacitor which we call S_0.

Now we consider the same capacitor, but from an inertial system S moving to the right with v_0. In this case the length L_0 is Lorentz contracted

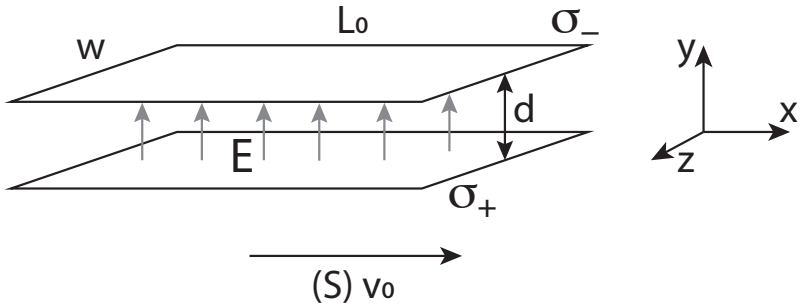

Figure 12.5: Parallel plate capacitor in S_0 and frame S moving to the right with v_0.

and thus $L = \frac{L_0}{\gamma_0}$, with γ as defined in Eq. 12.5, but with v_0 instead of u. Because charge is invariant, the charge density is thus increased when regarded from S and becomes $\sigma = \gamma_0 \sigma_0$. With this the E-field in S becomes

$$\vec{E} = \frac{\gamma_0 \sigma_0}{\epsilon_0} \hat{y} = \gamma_0 \vec{E}_0 \tag{12.21}$$

If we turn the plates in the xy-plane we get the same result for the z component and in general

$$\vec{E}_\perp = \gamma_0 \vec{E}_{0\perp} \tag{12.22}$$

For the parallel, or x, component we turn the plates in the yz-plane. In this case the distance between the plates is Lorentz contracted, but because the E-field is independent of d there is no change. We can thus state that

$$\vec{E}_{||} = \vec{E}_{0||} \tag{12.23}$$

This gives us only information about the transformation of the E-field between different reference frames. If we also want to include the B-field, we have to start from S, where the moving surface charge density causes a $\vec{B}$. Therefore, we introduce a third inertial system S' as illustrated in Figure 12.6 that moves with v with respect to S.

The charge density moving with v_0 to the left from the point of view of S can be considered as a surface current $\vec{J} = \pm \sigma v_0 \hat{x}$. For the top plate we have negative charge moving to the left and thus $\vec{J}$ to the right and for the bottom plate the current is in the opposite direction. Using the right-hand

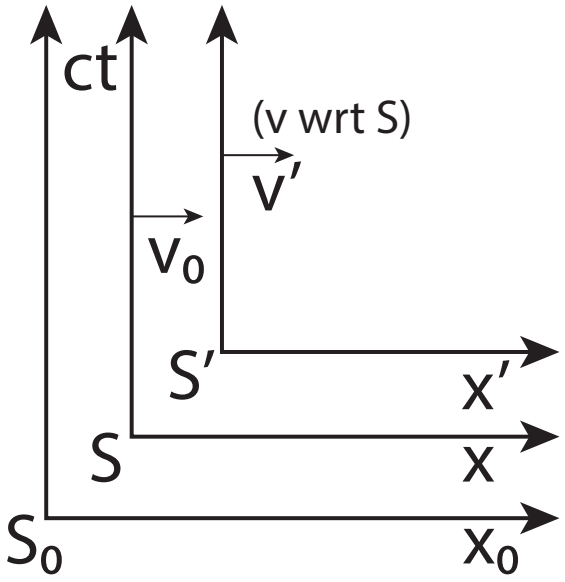

Figure 12.6: Three inertial system: S moves with v_0 with respect to S_0, S' moves with v with respect to S and thus v' with respect to S_0.

rule and the result of one of the exercises we get a magnetic field between the plates along the negative z direction

$$\vec{B} = -\mu_0 \sigma v_0 \hat{z} = B_y \tag{12.24}$$

In the following the vector notation will be simplified by the subscript indicating the direction. We now have the E- and B-field in S.

The next step is to determine the fields in S'. Along the same lines as argued before these are $E'_y = \frac{\sigma'}{\epsilon_0}$ and $B'_z = -\mu_0 \sigma' v'$. Where v' is the velocity relative to S_0 and σ' can be obtained from σ_0 using this velocity. Using Eq. 12.13 we get that

$$v' = \frac{v + v_0}{1 + v v_0 / c^2}$$

and we can use this to define the γ' for this frame

$$\gamma' = \frac{1}{\sqrt{1 - v'^2/c^2}}$$

Thus

$$\sigma' = \gamma' \sigma_0 = \frac{\gamma'}{\gamma_0} \sigma$$

and based on the expressions above

$$E'_y = \frac{\gamma'}{\gamma_0}\frac{\sigma}{\epsilon_0},\ B'_z = -\frac{\gamma'}{\gamma_0}\mu_0\sigma v'$$

The problematic term is $\frac{\gamma'}{\gamma_0}$ because we want to express things in the transition from S to S', which is defined by γ. With some rewriting we obtain that

$$\frac{\gamma'}{\gamma_0} = \gamma\left(1 + \frac{vv_0}{c^2}\right)$$

and keeping in mind that $\frac{1}{c^2} = \epsilon_0\mu_0$ we can write the field transformations for this geometry:

$$E'_y = \gamma\left(1 + \frac{vv_0}{c^2}\right)\frac{\sigma}{\epsilon_0} = \gamma\left(E_y - \frac{v}{c^2\epsilon_0\mu_0}B_z\right) = \gamma(E_y - vB_z) \tag{12.25}$$

$$B'_z = -\gamma\left(1 + \frac{vv_0}{c^2}\right)\mu_0\sigma\left(\frac{v + v_0}{1 + vv_0/c^2}\right) = \gamma(B_z - \mu_0\epsilon_0 vE_y) = \gamma(B_z - \frac{v}{c^2}E_y) \tag{12.26}$$

We thus directly see that the E- and B-fields transform into each other.

For the other perpendicular component we again rotate the plates in the xy-plane. If we rotate such that E_z is positive then also B_y will be along the positive y-direction and $B_y = \mu_0\sigma v_0$. The rest of the derivation is the same as above and we only get a sign change

$$E'_z = \gamma(E_z + vB_y) \tag{12.27}$$

$$B'_y = \gamma(B_z + \frac{v}{c^2}E_z) \tag{12.28}$$

For the E-field we have already derived the transformation of the parallel component. For the parallel B-field we have to use another configuration then the parallel plate capacitor and we consider a solenoid with its axis along $\hat{x}$. With a current I and n windings per unit length we get that $B_x = \mu_0 nI$ in S. In S' the solenoid will be Lorentz contracted and the density of the windings will thus change to $n' = \gamma n$. However, due to time dilation also the current ($\frac{dQ}{dt}$) will change and $I' = \frac{I}{\gamma}$ put together these two contributions cancel. We thus have

$$E'_x = E_x \tag{12.29}$$

$$B'_x = B_x \tag{12.30}$$

This gives us all the field transformations, considering that $\vec{v} = v\hat{x}$.

These transformations can be expressed in various ways and the explicit expressions given above are often the most useful. However sometimes a vector notation can be clearer. Again taking $\vec{v} = v\hat{x}$ this becomes

$$E'_x = E_x \tag{12.31}$$
$$E'_y = \gamma(\vec{E} + \vec{v} \times \vec{B})_y \tag{12.32}$$
$$E'_z = \gamma(\vec{E} + \vec{v} \times \vec{B})_z \tag{12.33}$$
$$B'_x = B_x \tag{12.34}$$
$$B'_y = \gamma(\vec{B} - \frac{\vec{v} \times \vec{E}}{c^2})_y \tag{12.35}$$
$$B'_z = \gamma(\vec{B} - \frac{\vec{v} \times \vec{E}}{c^2})_z \tag{12.36}$$

Where the subscript again indicate the relevant component. This combination of $\vec{E}$ and $\vec{B}$ is the generalised electromagnetic field and together it can be expressed using an antisymmetric second rank tensor. However, this goes beyond the content of this course.

Finally we consider two special cases, namely that either $\vec{B} = 0$ or $\vec{E} = 0$ in S. In the first case we can ignore the B terms in the transformations above and we get

$$\vec{B}' = \gamma\frac{v}{c^2}(E_z\hat{y} - E_y\hat{z}) = \frac{v}{c^2}(E'_z\hat{y} - E'_y\hat{z})$$

and thus with $\vec{v} = v\hat{x}$ this yields

$$\vec{B}' = -\frac{1}{c^2}(\vec{v} \times \vec{E}') \tag{12.37}$$

Similarly, for the second case we can ignore the E terms

$$\vec{E}' = -\gamma v(B_z\hat{y} - B_y\hat{z}) = v(B'_y\hat{z} - B'_z\hat{y})$$

and thus

$$\vec{E}' = \vec{v} \times \vec{B}' \tag{12.38}$$

This last expression we already used in the chapter on induction, but we didn't justify its validity.

Chapter 13

Various end stuff

This script has been written under substantial time pressure and any reader might recognise this. Although care has been taken to avoid any mistakes, I can't guarantee the absence of typos and other errors. Any comments in this respect are more than welcome. At this point I thank Mauro, Stefan, and Andrew for proof-reading parts of the manuscript.

It should be realised that this script is not meant to replace a proper text book and I have no claim with regard to completeness. Many topics would deserve a more in-depth treatment and there are excellent texts that do exactly this. My only aim is to provide an overview of my lecture as given at this level and covering the knowledge that I require from my students.

Lastly, many figures in this script have been taken from a variety of sources and I have not been as strict as I should have been listing these sources. This script is only meant for distribution with the relevant students at the EPFL and not for any further distribution or commercial gain. However, if anyone feels that their copyright has been inflicted I kindly ask them to contact me and the issue will be fixed immediately.

Bibliography

[1] A. Einstein, "On the electrodynamics of moving bodies", p.23 (1905)

[2] R. Feynman, R. Leighton, M. Sands, "The Feynman Lectures on Physics", Vol. I (1963)

[3] R. Feynman, R. Leighton, M. Sands, "The Feynman Lectures on Physics", Vol. II (1963)

[4] Wikipedia page of Retarded Potential

[5] Wikipedia page of Liénard-Wiechert potential