Electrodynamics for teachers

Lecture given by Frank Krauss, winter term 2004/2005,
at Dresden university of technology

Disclaimer:

This script is, to a large extent, influenced by a number of excellent books on electrodynamics. Specifically, the following books have been used:

- R. P. Feynman: “The Feynman lectures II”
- J. D. Jackson: “Electrodynamics”

Contents

Chapter 1

Mathematical tools

1.1 Scalar and vector fields

This section deals with some basic mathematical tools which will be of central importance throughout this lecture, namely with the notion of scalar and vector-valued quantities, their differentiation and integration. It is important to understand that scalars and vectors in physics incorporate a bit more than in mathematics, namely a certain way they behave under coordinate transformations. The basic idea is that in physics, scalar quantities like energy and vector quantities like velocity or momentum depend in a specific way on the choice of coordinate system. This is because of the agreement underlying all discussions that physical systems and their dynamics should be described in an "objective" manner, i.e. in a way that is independent on the particular observer.

The transformations in question, under which this independence should be guaranteed, are the Galilee-transformations (a superposition of spatial rotations and spatial and temporal translations) in the non-relativistic case and the Poincare-transformations (a superposition of spatial rotations, boosts and translations) in the relativistic case. Neglecting translations for a moment, the former are usually described by either real-valued 3×3 matrices, rotating the three vectors describing coordinates in space. Such transformations, when leaving the length of vectors invariant, are characterised by the transformation matrices having determinant of 1 or -1, they belong to the group $SO(3)$ or $O(3)$ depending on whether reflections with determinant -1 are excluded or not.

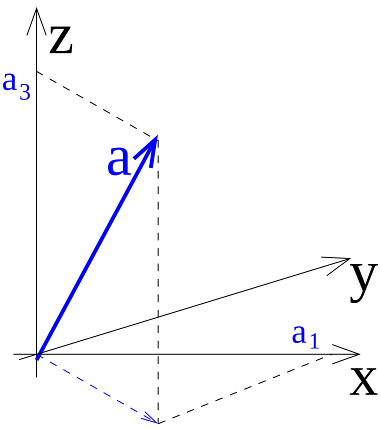

Figure 1.1: . A right-handed orthonormal system and a vector inside. The individual components can be obtained through some parallel projection.

1.1.1 Vector algebra

Vectors

Mathematically speaking, vectors are objects with different components. In physics, relevant vectors usually have three or four components; such vectors are then often called three- or four-vectors. For the moment, consider three vectors, like, e.g.

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = (a_1, a_2, a_3)^T \,. \tag{1.1}$$

The three components refer to three directions spanning an orthonormal system, cf. Fig.1.1. Therefore, $\vec{a}$ may be represented as a linear combination of the base vectors of this system,

$$\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3 = \sum_{i=1}^{3} a_i\vec{e}_i \,. \tag{1.2}$$

Its length is usually denoted by a and given by

$$a = |\vec{a}| = \sqrt{a_1^2 + q_2^2 + a_3^2} \,. \tag{1.3}$$

At this point, however, it is important to stress that physical vectors remain the same objects independent of their representation although their representation in terms of their components depends on the base system $\vec{e}_i$ chosen. This immediately translates into certain relations between their components in different coordinate systems, which are connected through transformations such as rotation. In addition, it also implies that certain characteristic properties such as their length remain invariant in different coordinate systems as long as they have the same scale.

Scalar product

The scalar product of two vectors maps the two vectors on a scalar quantity, i.e. , a number,

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{3} a_i b_i = |\vec{a}||\vec{b}| \cos\theta_{ab} \,. \tag{1.4}$$

The following rules apply for the scalar product:

- Commutativity:

$$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a} \,. \tag{1.5}$$

- Distributivity:

$$\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c} \,. \tag{1.6}$$

- Special cases:

$$\begin{array}{rclcrcl} \vec{a} \cdot \vec{a} & = & a^2 \,, & & & & \\ \vec{a} \cdot \vec{b} & = & 0 & \text{if} & \theta_{ab} & = & \frac{\pi}{2}, \frac{3\pi}{2}, \ldots \,, \\ \vec{a} \cdot \vec{b} & = & ab & \text{if} & \theta_{ab} & = & 0 \,, \\ \vec{a} \cdot \vec{b} & = & -ab & \text{if} & \theta_{ab} & = & \pi \,. \end{array} \tag{1.7}$$

Furthermore, the following inequality is valid for sums of vectors:

- Triangle inequality:

$$|\vec{a} + \vec{b}| \leq |\vec{a}| + |\vec{b}| \,. \tag{1.8}$$

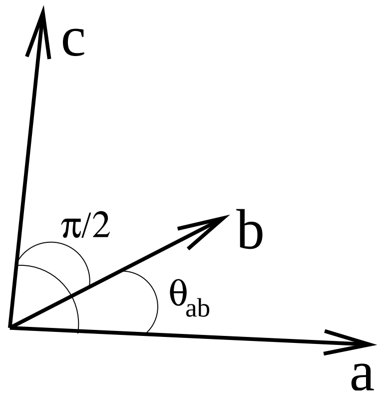

Figure 1.2: Pictorial representation of the vector product $\vec{c} = \vec{a} \times \vec{b}$. The "right hand rule" means that $\vec{a}$, $\vec{b}$, and $\vec{c}$ can be ordered according to thumb...

Vector product

The vector product of two vectors maps the two vectors on another vector,

$$\vec{a} \times \vec{b} = \vec{c} = \det \begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} , \tag{1.9}$$

("right-hand rule") where

$$c = |\vec{c}| = |\vec{a}||\vec{b}| \sin\theta_{ab} \tag{1.10}$$

and

$$\vec{c} \perp \vec{a} \quad \text{and} \quad \vec{c} \perp \vec{b}\,, \tag{1.11}$$

cf. , Fig.1.2. In other words, the vector product results in a vector perpendicular to the plane spanned by the two vectors with a length equal to the surface of the rhomboid given by them. One may also write the vector product by applying Einstein's convention of summing over all repeated indices and with help of the anti-symmetric tensor ϵ_{ijk}. The indices of this tensor run from 1 to 3; it is defined such that it yields a zero if two indices are equal, otherwise it is a ± 1, depending on whether the indices are ordered in a cyclic permutation of 123 (+) or not (-). Then,

$$\vec{a} \times \vec{b} = \epsilon_{ijk} a_i b_j \vec{e}_k\,. \tag{1.12}$$

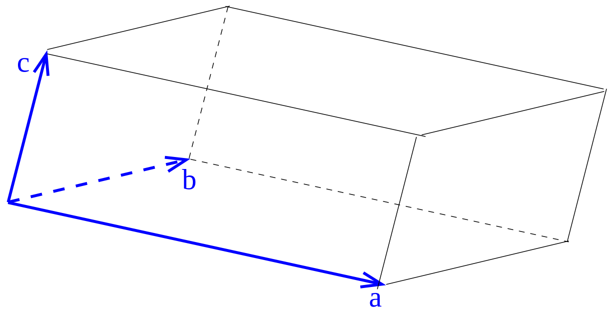

Figure 1.3: Pictorial representation of the mixed product $(\vec{a}\times\vec{b})\cdot\vec{c}$. The volume in this example is positive, since the three vectors form a right-system.

Directly from the definition above, it is clear that

$$\begin{aligned}
\vec{a}\times\vec{b} &= -\vec{b}\times\vec{a}\,,\\
\vec{a}\times\vec{a} &= 0\,,\\
\vec{a}\cdot(\vec{a}\times\vec{b}) &= 0\,,\\
\vec{a}\cdot(\vec{b}\times\vec{c}) &= (\vec{a}\times\vec{b})\cdot\vec{c}\,,\\
\vec{a}\times(\vec{b}\times\vec{c}) &= \vec{b}(\vec{a}\cdot\vec{c})-\vec{c}(\vec{a}\cdot\vec{b})\,,\\
\vec{a}\times(\vec{b}+\vec{c}) &= \vec{a}\times\vec{b}+\vec{a}\times\vec{c}\,.
\end{aligned} \tag{1.13}$$

The mixed product

$$(\vec{a}\times\vec{b})\cdot\vec{c} = \det\begin{vmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix} \tag{1.14}$$

yields the (signed) volume of the parallelepiped spanned by the three vectors, cf. Fig.1.3. The sign depends on whether the three vectors in their sequence $\vec{a}$, $\vec{b}$, $\vec{c}$ yield a right- or a left system. In the former case the sign is positive, otherwise it is negative. To exemplify this, consider the special case of

$$\vec{a}=\vec{e}_1\quad \vec{b}=\vec{e}_2\quad \vec{c}=\pm\vec{e}_3\,. \tag{1.15}$$

1.1.2 Polar vs. axial vectors

Vectors can be classified according to their behaviour under inversions $\mathcal{P}$, also dubbed parity transformations. Such transformations merely are reflections of space[1] and can be represented by 3×3-matrices with all diagonal elements equal to -1. Polar vectors, in the following denoted by $\vec{p}$, are defined to obtain an extra sign, when experiencing such a transformations, whereas axial vectors, denoted by $\vec{a}$, remain untouched by reflections:

$$\begin{aligned} \mathcal{P} \otimes \vec{p} &= -\vec{p}\,, \\ \mathcal{P} \otimes \vec{a} &= \vec{a}\,. \end{aligned} \tag{1.16}$$

In general, vector products of two vectors result in

$$\begin{aligned} \vec{p}_i \times \vec{p}_j &= \vec{a}_i \times \vec{a}_j = \vec{p}\,, \\ \vec{p} \times \vec{a} &= \vec{a} \times \vec{p} = \vec{a}\,. \end{aligned} \tag{1.17}$$

Examples for polar vectors include space $\vec{r}$, velocity $\vec{v}$ and acceleration $\vec{a}$, momentum $\vec{p}$ and force $\vec{F}$, the electrical field $\vec{E}$ and the current density $\vec{j}$. Axial vectors for instance are angular velocity $\vec{\omega}$ angular momentum $\vec{L}$ and torsional momentum $\vec{T}$, and the magnetic field $\vec{B}$.

1.1.3 Scalar and vector fields

Mathematically speaking, fields are nothing but functions, which may be real-valued (scalar fields), vector-valued (vector fields), or have even higher dimensions, such as tensor-valued functions. In physics, fields usually depend on space and time coordinates [2]. The simple most fields are scalar fields, they represent a mapping of space and, eventually, time on a single value:

$$\Phi = \Phi(t,\, \vec{r}) = \Phi(t,\, x,\, y,\, z) \in \mathbf{R}\,. \tag{1.18}$$

In contrast, vector fields represent a mapping of the coordinates onto a vector, i.e.

$$\vec{A} = \vec{A}(t,\, \vec{r}) \in \mathbf{R}^3\,. \tag{1.19}$$

[1] As an illustrative example consider a reflection around the x-y plane, resulting in a mapping $z \to -z$.

[2] What makes fields special is that a change of coordinate system should not change their value - in other words, two different observers with two different coordinate systems should experience the same field at the same position.

It should be clear that such vector fields can be rewritten as a vector of three scalar fields, namely

$$\vec{A} = \vec{A}(t,\, \vec{r}) = \begin{pmatrix} A_1(t,\, x,\, y,\, z) \\ A_2(t,\, x,\, y,\, z) \\ A_3(t,\, x,\, y,\, z) \end{pmatrix}, \tag{1.20}$$

where each of the A_i is a field like Φ.

Infinitesimal changes

However, going back to the simpler case of a scalar field, it is worth to remember that the infinitesimal change of $f(x,\, y,\, z)$ under infinitesimal shifts of its coordinates is given by the relation

$$\Delta f(x,\, y,\, z) = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \frac{\partial f}{\partial z}\Delta z\,. \tag{1.21}$$

It must be stressed that this is valid in the limit of

$$\Delta_x \to 0, \quad \Delta_z \to 0, \quad \Delta_y \to 0 \tag{1.22}$$

only. Note that the derivatives $\partial f/\partial x$ denote partial derivatives - i.e. they "take account of the explicit dependence of f on x." Then Δf is represented by the local slope in each direction times the distance along this direction. Clearly, if $\Delta f = 0$ for arbitrary Δx, Δy, and Δz, then these slopes are all zero and some local minimum or maximum is found.
Also, for "well-behaved" functions, the following relation holds true

$$\frac{\partial^2}{\partial x \partial y} = \frac{\partial^2}{\partial y \partial x}\,, \tag{1.23}$$

i.e. the sequence of derivation does not matter.

Examples for fields

Consider, for the moment, fields with no time-dependence. Such a field can be exemplified with the position-dependent temperature within some block of material, which has been heated or cooled at different places. Then the temperature reads

$$T = T(x,\, y,\, z)\,. \tag{1.24}$$

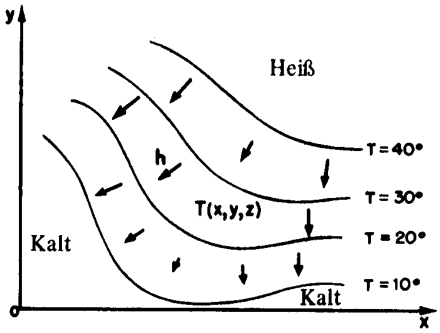

Figure 1.4: Isothermes in a block of material.

It is a scalar quantity - one number suffices to describe it for each position. To represent the temperature in the x-y plane with, say, $z = 0$, one could for instance draw all lines of constant T, which are dubbed isotherms for a temperature field. This is exemplified in Fig.1.4.
To illustrate what a vector field is, assume now that heat is flowing in the block of material above from warmer places to colder ones. Of course, at different positions the heat is flowing into different directions. This heat-flow therefore is a directed quantity, $\vec{h}$, whose absolute value signifies how much heat is transported. An example can be found in Fig.1.5. Now, a more precise definition of $\vec{h}$ can be given: The size of this vector is given by the infinitesimal amount of thermic energy per infinitesimal time ΔJ, which flows through an infinitesimal area Δa perpendicular to the flow direction:

$$\vec{h} = \frac{\Delta J}{\Delta a}\vec{e}_J\,, \tag{1.25}$$

where $\vec{e}_J$ denotes a unit vector pointing into the direction of the energy flow.

Of course, the vector $\vec{h}$ can also be defined through its components. To do so, consider the heat flow through any small surface Δs at an arbitrary angle θ with respect to the direction of the heat flow. The orientation of the surface can be represented by an unit vector perpendicular to it, usually called the normal vector $\vec{n}_s$. Consider a similar surface, Δa, which again is perpendicular to the heat flow, i.e. its normal vector $\vec{n}_a \parallel \vec{h}$. How can the heat flow through Δs be determined, if the areas of both surfaces are identical? The answer by now should be clear: The heat flow per unit area

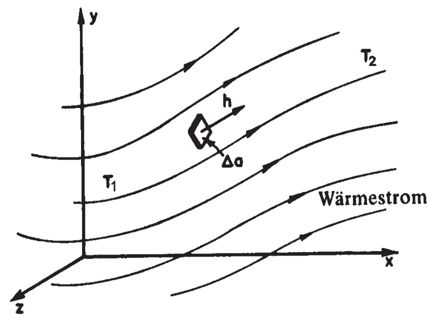

Figure 1.5: Heat flow in a block of material.

going through Δa is given by

$$\left|\vec{h}\right| = \frac{\Delta J}{\Delta a}\,. \tag{1.26}$$

The relevant area of Δs is, of course, given by

$$\Delta s_{\text{rel.}} = \Delta a\, \vec{n}_s \cdot \vec{n}_a = \Delta a \cos\theta\,, \tag{1.27}$$

leading to

$$\frac{\Delta J}{\Delta s} = \frac{\Delta J}{\Delta a}\cos\theta = \frac{\Delta J}{\Delta a}\vec{e}_J \cdot \vec{n}_s = \vec{h}\cdot\vec{n}_s\,. \tag{1.28}$$

1.2 Derivatives of fields

Similar to any other well-behaved function in mathematics, derivatives of fields can be formed. As an example consider the derivative w.r.t. time of the temperature field Φ above. It is possible to form to different derivatives, partial and total. Whereas the first one, symbolised by

$$\frac{\partial\Phi}{\partial t}\,, \tag{1.29}$$

takes care of the explicit time-dependence only, the second one also takes into account the implicit time-dependence that may be hidden in the space coordinates, viz.

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\partial\Phi}{\partial t} + \frac{\partial\Phi}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial\Phi}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial\Phi}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}\,. \tag{1.30}$$

In this expression, partial derivatives w.r.t. different spatial components show up. Obviously, they describe the connection between the value of the field at two distinct but infinitesimally close points. But what exactly is $\partial\Phi/\partial x$? Is it a scalar or a vector? In fact it is neither a scalar nor a vector! This is quite easy to understand by just considering the fact that different people may have a different notion about how the various axes are oriented in space. And, of course, a different orientation will most likely result in a different value for $\partial\Phi/\partial x$. But wait! On the other hand there are three such expressions, $\partial\Phi/\partial x$, $\partial\Phi/\partial y$, and $\partial\Phi/\partial z$. Maybe they form a vector? Of course, it is not true that three arbitrary numbers arranged in a row just form a vector. After all, the components of a vector must satisfy certain relations under transformations such as rotation. Therefore, it is important to analyse the behaviour of $(\partial\Phi/\partial x,\, \partial\Phi/\partial y,\, \partial\Phi/\partial z)^T$.
To check whether $(\partial\Phi/\partial x,\, \partial\Phi/\partial y,\, \partial\Phi/\partial z)^T$ indeed forms a vector, the following strategy may be exploited: Ask a question, whose answer is independent of the coordinate system that has been used, and try to answer in an "invariant" way. A simple example - which, not quite coincidentally, will be used - is the fact that $S = \vec{A}\cdot\vec{B}$ is a scalar (i.e. the **same in all coordinate systems**), if, as indicated, $\vec{A}$ and $\vec{B}$ are vectors. In turn, if $\vec{A}$ is a vector and S is **known** to be a scalar, then $\vec{B}$ **must** be a vector. This will be used now. For equal times the difference of the value of Φ in two close points, $\Delta\Phi$, clearly should be independent of the coordinate system and just depend on the points and the field itself. Then $\Delta\Phi$ may be calculated as

$$\begin{aligned}\Delta\Phi &= \frac{\partial\Phi}{\partial x}\Delta x + \frac{\partial\Phi}{\partial y}\Delta y + \frac{\partial\Phi}{\partial z}\Delta z \\ &= \left(\frac{\partial\Phi}{\partial x}, \frac{\partial\Phi}{\partial y}, \frac{\partial\Phi}{\partial z}\right)\cdot\begin{pmatrix}\Delta x\\ \Delta y\\ \Delta z\end{pmatrix}.\end{aligned} \tag{1.31}$$

Coming back to the example of the previous section, Φ can be readily identified as the temperature field T, and ΔT then represents the temperature difference in two points.
Following from the considerations above, it is clear that $\mathrm{d}\Phi/\mathrm{d}t$ in Eq.(1.30) represents a scalar quantity - it is just a sum of a scalar, the time explicit derivative of Φ plus another scalar quantity. It would be nice, however, to rewrite this expression in a more catchy way, like,

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\partial\Phi}{\partial t} + \frac{\partial\Phi}{\partial\vec{r}}\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \frac{\partial\Phi}{\partial t} + \frac{\partial\Phi}{\partial\vec{r}}\cdot\vec{v}\,. \tag{1.32}$$

In order to retain this as a real-valued expression equal to the one displayed in Eq.(1.30), the term $\partial\Phi/\partial\vec{r}$ must be a vector. Comparing coefficients, it is clear that it has to assume the form

$$\frac{\partial\Phi}{\partial\vec{r}} = \begin{pmatrix} \partial\Phi/\partial x \\ \partial\Phi/\partial y \\ \partial\Phi/\partial z \end{pmatrix} = \frac{\partial\Phi}{\partial x}\vec{e}_x + \frac{\partial\Phi}{\partial y}\vec{e}_y + \frac{\partial\Phi}{\partial z}\vec{e}_z = \sum_{i=1}^{3} \frac{\partial\Phi}{\partial x_i}\vec{e}_i\,. \tag{1.33}$$

1.2.1 The gradient

In fact, the operation $\partial/\partial\vec{r}$ applied on a scalar field (a scalar function) is called the gradient of this function,

$$\text{grad }\Phi \equiv \partial\Phi/\partial\vec{r} = \sum_{i=1}^{3} \frac{\partial\Phi}{\partial x_i}\vec{e}_i = \frac{\partial\Phi}{\partial x_i}\vec{e}_i\,. \tag{1.34}$$

Consider, for the moment, again the example of the temperature field above, $T = T(x,\,y,\,z)$. Assume the temperatures at two nearby points P_1 and P_2, separated by a (directed) distance $\Delta\vec{R}$, are given by T_1 and T_2, respectively. Clearly, the temperatures at these physical points do not depend on any convention of any observer of how to choose the axes of his or her favourite coordinate system. Therefore, also the temperature difference ΔT between these two points is independent of any choice of coordinates and, hence, a scalar quantity. Using a simple coordinate system (the Cartesian one), one may write

$$T_1 = T(x,\,y,\,z) \quad \text{and} \quad T_2 = T(x+\Delta x,\,y+\Delta y,\,z+\Delta z)\,, \tag{1.35}$$

such that in turn

$$\Delta\vec{R} = (\Delta x,\,\Delta y,\,\Delta z)^T\,. \tag{1.36}$$

Therefore, for infinitesimal distances,

$$\Delta T = T_2 - T_1 = \text{grad }T\cdot\Delta\vec{R}\,. \tag{1.37}$$

1.2.2 The operator $\vec{\nabla}$

Consider, once more, the operation $\partial/\partial\vec{r}$. So far, it has been applied to a scalar field, resulting in a vector, the gradient of this field. On a more abstract level, i.e. omitting the function it is acting on, it is clear from Eq.(1.34) that one can represent the corresponding operator as

$$\frac{\partial}{\partial \vec{r}} = \sum_{i=1}^{3} \vec{e}_i \frac{\partial}{\partial x_i} = \vec{e}_i \frac{\partial}{\partial x_i} \equiv \vec{\nabla} \,. \tag{1.38}$$

This operator is usually denoted with the symbol $\vec{\nabla}$, or ∇ alone. This symbol is called "nabla". The fact that this abstraction is possible and meaningful is one of the beauties hidden in mathematics. Remember that it was irrelevant which scalar field Φ was chosen for the proof that $\vec{\nabla}\Phi$ indeed is a vector. Since the equations that transform differentiation under, say rotations around the z-axis, are independent of the field Φ, one may easily identify

$$\frac{\partial}{\partial x} \longrightarrow \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y} \,, \tag{1.39}$$

when θ was the angle of rotation. $\partial/\partial x$ now represents an operator: It is "hungry" for something to be attached to the **right** of it, in order to differentiate it. Similarly, $\vec{\nabla}$ waits for something it can act on, the corresponding object, by convention, is put to the right of it in order to indicate the action. However, it is important to keep in mind that operators remain meaningless as long as they do not act on something; in such a case they just represent a mathematical operation, without any specification of which object is subject to the respective operation. If this operator is meaningless alone, the question of what it relates to when multiplied with, say, a scalar function, has to be answered. The answer is simple: $\Phi\vec{\nabla}$ still is some kind of operator - the individual components are just scaled by a common factor Φ - and it is still a vector-valued operator. In contrast, $\vec{\nabla}\Phi$ now becomes something meaningful, the derivatives in the operator are supplied with a function to act on. In fact, it is a physical vector, incorporating the rate with which Φ changes in space. The x-component for instance manifestly gives the change of Φ in x-direction. The last question to be answered here is in which direction this vector $\vec{\nabla}\Phi$ points. The answer should be clear by now: It is this direction in space, in which Φ changes most (for a given point, i.e. for a given reference vector $(x_0,\, y_0,\, z_0)^T$ at which it is taken). **The gradient of Φ has the direction of the steepest rise or fall of Φ.**

Fun with $\vec{\nabla}$: grad, div, and rot

With help of this operator, the gradient can be rewritten as

$$\text{grad}\ \Phi = \vec{\nabla}\Phi\,, \tag{1.40}$$

making explicit the vector character of the gradient. In fact, the $\vec{\nabla}$ may also be applied on a vector field through a dot-product (scalar product), giving rise to the **divergence**,

$$\vec{\nabla}\cdot\vec{A} \equiv \text{div}\ \vec{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} = \sum_{i=1}^{3}\frac{\partial A_i}{\partial x_i} = \frac{\partial A_i}{\partial x_i}\,. \tag{1.41}$$

Writing the divergence with the nabla-operator, and going through the same steps as before,it is quite obvious that the result is a scalar. This immediately translates into its invariance under transformations of the coordinate system such as rotations. The physical meaning of this quantity will be discussed later on.
What else can be done with $\vec{\nabla}$? How about a vector (cross) product? It can be expected that

$$\vec{\nabla}\times\vec{A} \equiv \text{rot}\ \vec{A} = \text{avector.} \tag{1.42}$$

Its components can be written as

$$\vec{\nabla}\times\vec{A} = \begin{pmatrix} \nabla_y A_z & - & \nabla_z A_y \\ \nabla_z A_x & - & \nabla_x A_z \\ \nabla_x A_y & - & \nabla_y A_x \end{pmatrix} = \begin{pmatrix} \frac{\partial A_z}{\partial y} & - & \frac{\partial A_y}{\partial z} \\ \frac{\partial A_x}{\partial z} & - & \frac{\partial A_z}{\partial x} \\ \frac{\partial A_y}{\partial x} & - & \frac{\partial A_x}{\partial y} \end{pmatrix} \tag{1.43}$$

This object is dubbed the **rotor** of the vector field. The physical meaning of it as well as its interpretation will be discussed later.
Similarly to the findings in case of the gradient, the ∇-operator remains to be a "hungry" operator when multiplied from the right with a vector field, e.g. for

$$\vec{A}\cdot\vec{\nabla} \quad \text{and} \quad \vec{A}\times\vec{\nabla}\,. \tag{1.44}$$

1.2.3 Example: The differential equation of heat flow

To gain some first insight into the formulation of physical laws with help of the ∇ operator, the example of the heat flow will be considered. In fact, all deliberations made now are quite rough, the law discussed here is not very precise; at best it constitutes quite an approximation, which, nevertheless, works surprisingly well.
Consider a regularly formed volume of a material, say, a block, which is heated on one side to the temperature T_2 and cooled on the other side to the temperature T_1. It is clear that in this situation heat flows through the material, from the regions of the higher to those of the lower temperature. In its simplest approximation, the heat flow is proportional to the area A it flows through and it is also proportional to the temperature difference. In addition, it is inversely proportional to the thickness of the block, i.e. to the distance of the two sides d, cf. Fig.1.6a). The total energy J floating through the material per time is then

$$J = \kappa(T_2 - T_1)\frac{A}{d}\,, \tag{1.45}$$

where the constant of proportionality κ is known as the heat conductance.
Well, this was the simple case; what happens in a more complicated case, when the geometry of the volume is not just a simple block? The answer to this question can now be formulated with help of the mathematical tools that have been developed so far. First of all, a very small slice of the material is considered. It will be oriented such that its boundary are parallel to the isotherms inside the material, cf. Fig.1.6b). This allows to use Eq.(1.45) in a form that it is correct for a small section of the slice,

$$\Delta J = \kappa \Delta T \frac{\Delta A}{\Delta d}\,, \tag{1.46}$$

where ΔT is the temperature difference between the two boundary layers of the slice, Δd is its thickness, and ΔA is its area. Remember now that previously $\Delta J/\Delta A$ has been defined as the norm of $\vec{h}$, which in turn has the direction of the heat flow. It should be clear that the heat flows from $T + \Delta T$ to T, i.e. perpendicular to the isotherms . Also, $\Delta T/\Delta d$ is the ratio with which T changes with position between the isotherms. Of course, the maximal change is perpendicular to the isotherms, hence $\Delta T/\Delta d$ is the maximal rate of temperature change. Taking now the limit $\Delta d \to 0$, it is

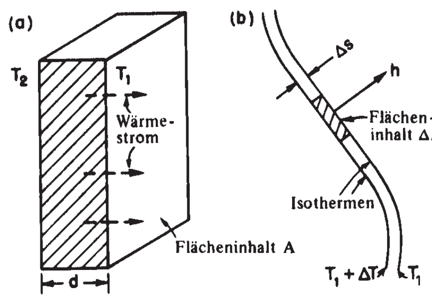

Figure 1.6: Heat flow in an infinitesimal block of material.

clear that the ratio $\Delta T/\Delta d$ becomes $\vec{\nabla}T$, taking into account the direction. This direction, however, is opposite to $\vec{h}$, therefore Eq.1.46 can be written as

$$\vec{h} = -\kappa\vec{\nabla}T\,. \tag{1.47}$$

This is, in the crude approximation made, the differential equation for heat flow in media. It is a real vector equation if κ is just a number; its nice feature is that it does not depend on any notion of a coordinate system. It is thus independent of any choice of coordinates and it allows to show the physical meaning of a physical law in quite a catchy way.

1.2.4 Second derivatives

So far, only first derivatives involving the ∇ operator have been discussed. Let's check things out when second derivatives are involved (in fact, this will become quite important for the case of electrodynamics, too). What combinations can be formed? It is easy to check that only the following ones are possible:

1. $\vec{\nabla}\cdot(\vec{\nabla}\Phi)$
2. $\vec{\nabla}\times(\vec{\nabla}\Phi)$
3. $\vec{\nabla}(\vec{\nabla}\cdot\vec{A})$
4. $\vec{\nabla}\cdot(\vec{\nabla}\times\vec{A})$

5. $\vec{\nabla} \times (\vec{\nabla} \times \vec{A})$

Let's start the discussion with expression (2.). It has the form

$$\vec{a} \times (\vec{a}\phi) = (\vec{a} \times \vec{a})\phi = 0\,. \tag{1.48}$$

This is because $\vec{a} \times \vec{a}$ is always zero, as is already known from Eqs.(1.13). Therefore, one may guess that also

$$\vec{\nabla} \times (\vec{\nabla}\Phi) = \text{rot grad } \Phi = 0\,. \tag{1.49}$$

This is good! Maybe there are other expressions which immediately reduce to zero? From Eqs.(1.13)

$$\vec{a} \cdot (\vec{a} \times \vec{b}) = 0\,, \tag{1.50}$$

since $\vec{a} \times \vec{b}$ is perpendicular to $\vec{a}$. Therefore, expression (4.) above reduces to

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = \text{div (rot } \vec{A}) = 0\,. \tag{1.51}$$

Let us turn now to those expression which do not yield a zero. Consider expression (1.) first. It reads

$$\begin{aligned} \vec{\nabla} \cdot (\vec{\nabla}\Phi) &= \nabla_x(\nabla_x\Phi) + \nabla_y(\nabla_y\Phi) + \nabla_z(\nabla_z\Phi) \\ &= \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2}\,, \end{aligned} \tag{1.52}$$

which should be a number. In fact, it is another scalar field. Obviously, the brackets are not necessary, instead one could equally well write

$$\vec{\nabla} \cdot (\vec{\nabla}\Phi) = \vec{\nabla} \cdot \vec{\nabla}\Phi = (\vec{\nabla} \cdot \vec{\nabla})\Phi = \nabla^2\Phi\,, \tag{1.53}$$

where ∇^2 can be viewed as a new operator. It is coined the Laplace operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\,. \tag{1.54}$$

Since this is a scalar operator, one may easily apply it to a vector field. This however, can be written down in a straightforward manner only in Cartesian coordinates. There it reads

$$\nabla^2\vec{A} = \begin{pmatrix} \nabla^2 A_x \\ \nabla^2 A_y \\ \nabla^2 A_z \end{pmatrix}\,, \tag{1.55}$$

i.e. the same operation is applied to all three components. Writing the field, for instance, in polar coordinates results in a slightly more complicated expression.
The next expression to be examined is the one of (5.), the rotor of a rotor of a vector field. Again, Eqs.(1.13) may be invoked, in particular

$$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})\,. \tag{1.56}$$

Translating this literally to expression (5.) results in

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \vec{A}(\vec{\nabla} \cdot \vec{\nabla})\,, \tag{1.57}$$

clearly unwanted, since the Laplace operator has nothing to act on. This situation can be simply resolved; unlike the case of matrices, the ordering of vectors in the template equation, Eq.(1.56), is to some extent arbitrary and could equally well be put like

$$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - (\vec{a} \cdot \vec{b})\vec{c}. \tag{1.58}$$

Of course, then

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A}. \tag{1.59}$$

The last expression on the list is (3.), and in fact there is nothing much to say about it. It is just a vector field. Taken together, the reasoning so far yields

1. $\vec{\nabla} \cdot (\vec{\nabla}\Phi) = \nabla^2\Phi =$ a scalar field.
2. $\vec{\nabla} \times (\vec{\nabla}\Phi) = 0$.
3. $\vec{\nabla}(\vec{\nabla} \cdot \vec{A}) =$ a vector field.
4. $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0$.
5. $\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2\vec{A}$.
6. $(\vec{\nabla} \cdot \vec{\nabla})\vec{A} =$ a vector field.

Two useful theorems

Before turning to the issue of integrating fields, two useful theorems will be stated without proof. These theorems will be of major importance in this course.

Theorem 1:

If the rotor of a vector field is zero, then the vector field can be written as the gradient of a scalar field.

$$\text{rot}\,\vec{A} = \vec{\nabla}\times\vec{A} = 0 \quad \Longleftrightarrow \quad \exists\Phi\ :\ \vec{A} = \text{grad}\,\Phi = \vec{\nabla}\Phi\,. \tag{1.60}$$

Theorem 2:

If the divergence of a vector field is zero, then the vector field can be written as the rotor of another vector field.

$$\text{div}\,\vec{A} = \vec{\nabla}\cdot\vec{A} = 0 \quad \Longleftrightarrow \quad \exists\vec{B}\ :\ \vec{A} = \text{rot}\,\vec{B} = \vec{\nabla}\times\vec{B}\,. \tag{1.61}$$

1.3 Integrals of fields

The last section can be summarised as follows: There exists a set of three operators $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$, which can be organised as the three components of a vector operator $\vec{\nabla}$. This operator can be applied in various combinations on scalar or vector fields.

Applying it on a scalar field is synonymous to calculating its gradient. The interpretation of the gradient as a vector pointing into the direction of the steepest rise or fall of the scalar field with a length giving the rate of change was quite straightforward to obtain. To gain similar interpretations of this operator when acting on a vector field through a dot- or a cross-product - synonymous to calculating the divergence or rotor of this field - is not so simple. To this end, integral relations are much better suited, and this will be the subject of the next few pages.

The integral equations obtained in these pages are, in principle, nothing but mathematical theorems. They do, however, provide a deeper understanding of the structure of the theory of electromagnetism. This does not necessarily render them particularly useful for real, physical applications or for the solution of problems. Nevertheless, for simple applications (and this is at least a good starting point) they can and they will provide solutions.

1.3.1 The line integral of $\vec{\nabla}\Psi$

The first integral contains the gradient. Since the gradient is nothing else but the rate of change of the corresponding field, the overall change between two points $(B) = (x_B,\, y_B,\, z_B)^T$ and $(E) = (x_E,\, y_E,\, z_E)^T$ can be obtained through integration. Using the shorthand notation $\Psi(B)$ and $\Psi(E)$ for the value at Ψ at those two points, one is tempted to write

$$\Psi(E) - \Psi(B) = \int\limits_{(B)}^{(E)} \vec{\nabla}\Psi \cdot \mathrm{d}\vec{l}, \tag{1.62}$$

where the integral is along any path connecting the two points. Here, $\mathrm{d}\vec{l}$ is any infinitesimal line element along the path, pointing tangentially to the path at any given point.

To understand a bit better the meaning of this line integral, consider first any scalar function $f(x,\, y,\, z)$ and a path L connecting two points, for simplicity again (B) and (E). Now, let's divide the path L into some segments by specifying points along the path. In the limit of small segments, the length of any segment can be approximated by connecting its two boundaries with their respective distance, c.f. Fig.1.7. Thus,

$$\int\limits_S \mathrm{d}l f \Longleftarrow \sum_i f_i \Delta l_i\,, \tag{1.63}$$

where taking the limit of infinitesimal lengths Δl_i and infinite segmentation of L is implicit. By identifying the function f in the above equation with the tangential component of $\vec{\nabla}\Psi$, i.e. the component of $\vec{\nabla}\Psi$ that is parallel to the respective segment Δl_i, the identification of Eq.(1.62) with the limit of an infinite sum becomes transparent.

But how can Eq.(1.62) seen to be correct? Coming back to the interpretation of the gradient and considering a point (1) close to (B) one may approximate

$$\Delta\Psi(1) \equiv \Psi(1) - \Psi(B) = (\vec{\nabla}\Psi)_{\|}(B)\Delta l_1 = (\vec{\nabla}\Psi)(B) \cdot \Delta\vec{l}_1\,, \tag{1.64}$$

essentially the first non-vanishing term in a Taylor expansion around (B). This may be repeated for a subsequent point,

$$\Delta\Psi(2) \equiv \Psi(2) - \Psi(1) = (\vec{\nabla}\Psi)_{\|}(1)\Delta l_2 = (\vec{\nabla}\Psi)(1) \cdot \Delta\vec{l}_2\,, \tag{1.65}$$

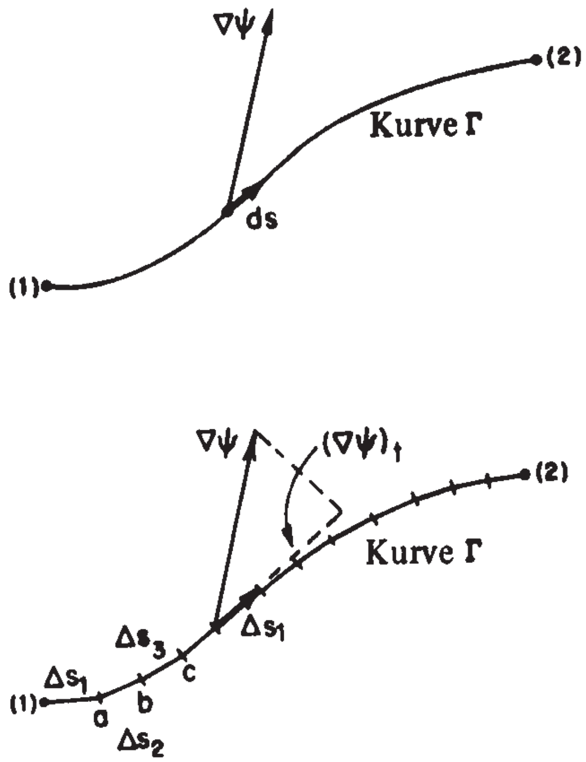

Figure 1.7: Construction of a line integral of Ψ along the curve Γ.

and generalised to any point n,

$$\Delta\Psi(n+1) \equiv \Psi(n+1) - \Psi(n) = (\vec{\nabla}\Psi)_{\parallel}(n)\Delta l_{n+1} = (\vec{\nabla}\Psi)(n) \cdot \Delta\vec{l}_{n+1} \quad (1.66)$$

In all cases $(\vec{\nabla}\Psi)(i)$ denotes the gradient taken at point (i), and $(\vec{\nabla}\Psi)_{\parallel}(i)$ is the corresponding component tangential w.r.t. the path L. Adding up the individual pieces yields

$$\Psi(E) - \Psi(B) = \lim_{N\to\infty} \sum_{i=0}^{N} (\vec{\nabla}\Psi)(i)\Delta\vec{l}_i \,, \quad (1.67)$$

where $i = 0 \equiv B$ and $i = N+1 \equiv E$. Obviously the left side does not depend at all on the way the intervals have been chosen, implying that taking the limit on the r.h.s. is a meaningful operation. From these considerations it is clear that, as long as the function remains smooth, i.e. as long as its gradient is finite everywhere, the following theorem holds:
<u>Theorem 3:</u>

$$\Psi(E) - \Psi(B) = \int_L \vec{\nabla}\Psi \cdot \mathrm{d}\vec{l} \quad \forall \text{ paths } L \text{ between } (B) \text{ and } (E)\,. \quad (1.68)$$

1.3.2 Flux of a vector field

Flux through a surface

The integral theorem of Gauss is closely connected with a certain physical picture. This picture has a very clear notion when applied to the case of heat fluxes, the prime example of this chapter. There, the vector $\vec{h}$ has been defined as the heat floating through a unit area per time interval. Let's assume now that within a block of material, there is a closed surface S around given volume V. Mathematicians tend to display this connection by calling $S = \delta V$. How much heat is floating out of the volume? The answer is clear: the total heat entering or leaving V equals the total amount of heat passing S, where, of course, signs matter. Again this looks like integrals have to be evaluated. This time however, it is not about any integral along a line, but rather about integrals over areas or volumes. In Cartesian coordinates for instance

$$\mathrm{d}^2 s = \mathrm{d}x\,\mathrm{d}y \quad (1.69)$$

for areas in the x-y-plane and

$$\mathrm{d}^3V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,. \tag{1.70}$$

The exponents in the infinite area and volume element serve as reminders of the dimensionality of the integration. However, the heat floating out of the volume through d^2s is given by the product of d^2s times the component of $\vec{h}$ orthogonal to this area element. Thus the overall heat flux is given by

$$\text{heat flux out of the volume } V = \int\limits_{S=\delta V} \mathrm{d}^2s\,\vec{h}\cdot\vec{n}_a\,, \tag{1.71}$$

where $\vec{n}_a$ is the vector orthogonal on each area element. This is usually also called "the heat flux through the surface". Thus $\vec{h}$ can be defined as the current density corresponding to the heat flux, and its integral is the total heat current through the respective surface. This integral has units of energy per time (J/s).
This reasoning can be easily generalised to those cases where some general vector function "floats" through a surface. In particular, one may apply this to the electric field, finding expression for the flux of $\vec{E}$ through surfaces. This translates into identifying, in this context, flux with the normal component of a vector field w.r.t. a surface, irrespective of whether the surface is closed or not. In other words "flux through an area" translates into the integral of the normal component of the corresponding vector field through the area.
Returning to the example of the heat flux above, consider a case where the total amount of heat is conserved. A good example for this is some material which has been heated for some time and which is then left without having any influx of energy or heat. If in this situation heat floats out of a closed surface inside this material, the heat inside the volume enclosed by the surface has to decrease. In such a case

$$\int\limits_{S=\delta V} \mathrm{d}^2s\,\vec{h}\cdot\vec{n}_a = -\frac{\mathrm{d}Q_V}{\mathrm{d}t}\,, \tag{1.72}$$

where Q_V is the total amount of heat inside the volume V. In other words: the heat floating out of the surface equals the change in the total amount of heat inside the volume enclosed by the surface. This is a tautology to the statement "conserved heat". Of course, this identity would break down, if any heat was produced or absorbed inside the volume.

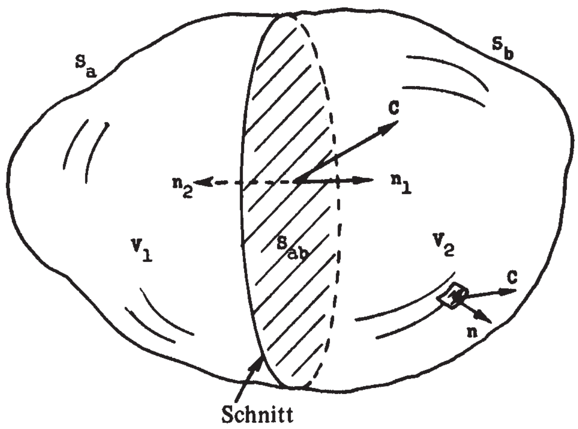

Figure 1.8: Partitioning the volume through a cut.

Partitioning the volume

In the following an interesting and important fact will be stressed. Assume that the volume V is separated into two sub-volumes V_1 and V_2 by a "cut" through it: $V \to V_1 + V_2$. Of course, the cut will be nothing but an area, for simplicity called S_{ab}. This area is the common boundary of the two sub-volumes and, therefore, an integral part of the two enclosing surface S_1 and S_2. In fact, $S = S_1 + S_2 - 2S_{ab}$. It is easy to imagine that the two surfaces consist of one piece of the original surface, say S_a and S_b, respectively, and of the boundary area S_{ab}, $S_{1,2} = S_{a,b} + S_{ab}$, c.f. Fig. 1.8.
It is natural to ask the following question: Is the sum of the two fluxes through S_1 and S_2 identical to the one through S ? The answer is yes, since the flux going through S_{ab}, common to both S_1 and S_2 cancels due to the different orientation (at the boundary, flux out of V_1 translates into flux into V_2 and vice versa, giving rise to an extra minus sign). Formally,

$$\begin{aligned}
\text{flux through } S_1 &= \int\limits_{S_a} \mathrm{d}^2 s_a\, \vec{h} \cdot \vec{n}_a + \int\limits_{S_a b} \mathrm{d}^2 s_a\, \vec{h} \cdot \vec{n}_a \\
\text{flux through } S_2 &= \int\limits_{S_b} \mathrm{d}^2 s_b\, \vec{h} \cdot \vec{n}_b + \int\limits_{S_a b} \mathrm{d}^2 s_b\, \vec{h} \cdot \vec{n}_b \qquad (1.73)
\end{aligned}$$

and, since $n_a = -n_b$,

$$\int\limits_{S_a b} \mathrm{d}^2 s_a\, \vec{h} \cdot \vec{n}_a = - \int\limits_{S_a b} \mathrm{d}^2 s_b\, \vec{h} \cdot \vec{n}_b\,. \tag{1.74}$$

This partitioning could be repeated, giving rise to increasingly smaller sub-volumes. But independent of how the partitioning is performed, the overall flux out of the original volume remains the same and can be written as the sum of the partial fluxes through the surfaces of the sub-volumes. The cancellation of fluxes through boundary areas is independent of their exact form.

1.3.3 Theorems of Gauss and Stokes

Theorem of Gauss: Flux out of a cube

This idea of partitioning a volume can be repeated until very small cubes (or rectangular parallelepipeds) constitute the sub-volumes. What is the flux of a vector field, say $\vec{A}$, out of such a cube? To simplify matters, let's assume the cube is oriented such that its edges are parallel to the x-, y-, and z-axis, see Fig. 1.9. Then the corner closest to the origin may have the coordinates $(x,\, y,\, z)^T$ and its dimensions are small and denoted by Δx, Δy, and Δz. To calculate the flux out of this cube, it suffices to calculate the flux through each of the six sides. Let's start with the two sides oriented parallel to the y-z plane. Then the component of A normal to these sides is A_x. In the following these two sides will be denoted by (1) and (2). Since the two sides are small, it suffices for the integration to take the value of A_x at the centre of each side:

$$\begin{aligned} \text{flux through(1)} &= -A_x(1)\Delta y \Delta z \\ \text{flux through(2)} &= A_x(2)\Delta y \Delta z\,, \end{aligned} \tag{1.75}$$

where the minus sign reflects the fact that the "out"-directions of the two sides are opposite w.r.t. each other. Now, using the fact that the sides are close to each other, the following approximation can be employed

$$A_x(2) = A_x(1) + \frac{\partial A_x(1)}{\partial x} \Delta x\,, \tag{1.76}$$

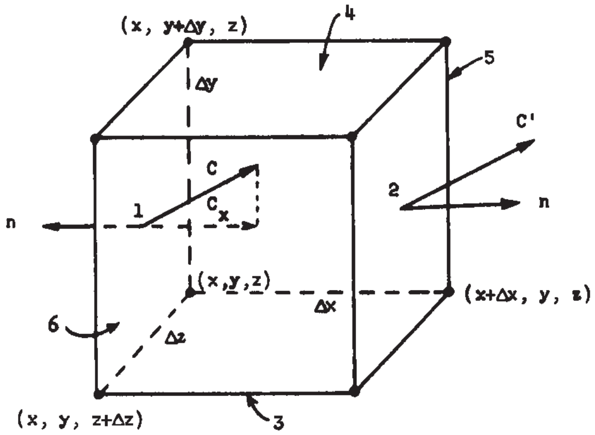

Figure 1.9: Flux through an infinitesimal cube.

resulting in

$$
\begin{aligned}
\text{flux through(1)and(2)} &= \left[-A_x(1) + A_x(2)\right] \Delta y \Delta z \\
&= \frac{\partial A_x(1)}{\partial x} \Delta x \Delta y \Delta z \,. \qquad (1.77)
\end{aligned}
$$

The derivative should be calculated at the center of the side, i.e. at $(x,\, y + \Delta y/2, z + \Delta z/2)^T$, calculating it at the corner $(x,\, y,\, z)^T$ instead results in a small and thus negligible deviation. Similar reasoning also holds true for the other two pairs of sides. Therefore, the total flux out of the cube can be written as

$$
\int\limits_{\text{cube}} \vec{A} \cdot \vec{n} \mathrm{d}^2 s = \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) \Delta x \Delta y \Delta z \qquad (1.78)
$$

in the limit of vanishing size of the cube. The sum of the derivatives inside the bracket, however, is nothing but $\vec{\nabla} \cdot \vec{A}$ and $\Delta x \Delta y \Delta z$ is just the volume element $\mathrm{d}^3 V$. Hence, for an infinitesimally small cube

$$
\int\limits_{\text{surface}} \vec{A} \cdot \vec{n} \mathrm{d}^2 s = \int\limits_{\text{cube}} \vec{\nabla} \cdot \vec{A} \mathrm{d}^3 V \,. \qquad (1.79)
$$

This identity allows for an interpretation of the divergence: The divergence of a vector $\vec{A}$ at a point P equals the outward flux of $\vec{A}$ per unit volume around P.

This links the divergence of $\vec{A}$ with the flux of $\vec{A}$ out of a small cube. Using the partitioning of volumes from the previous section in a backward fashion, i.e. building larger and larger volumes out of infinitesimally small cubes, allows to integrate the divergence of a vector field over a finite volume. This yields the following theorem:
Theorem 4 (Gauss' theorem):
The integral of the normal component of a vector field over a closed surface equals the integral over the divergence of this vector field over the enclosed volume

$$\int\limits_{S=\delta V} \vec{A} \cdot \vec{n} \mathrm{d}^2 s = \int\limits_{V} \vec{\nabla} \cdot \vec{A} \mathrm{d}^3 V \,. \tag{1.80}$$

Heat conductance

To become acquainted with this theorem let's consider an example for its application. Of course, this example again will be related to the heat flux within some material. The situation to be considered is simple: Heat has been supplied before, and the material is left now to cool down. There are no sources of heat around, hence the total amount of heat is conserved. The question to be tackled is how much heat can be found for any given time inside a certain volume. Well, the answer should be clear: The heat inside a volume decreases by the amount floating out of the surface. For any small cube this would amount to

$$\text{heat flux} = \int_{\text{cube}} \vec{h} \cdot \vec{n}_s \mathrm{d}^2 s = \vec{\nabla} \cdot \vec{h} \Delta V \,. \tag{1.81}$$

Identifying the amount of heat per unit volume with q, the heat inside this volume is $q\Delta V$ and its rate of loss is $-\mathrm{d}q/\mathrm{d}t$. Comparing coefficients, it is easy to see that

$$-\frac{\mathrm{d}q}{\mathrm{d}t} = \vec{\nabla} \cdot \vec{h} \,. \tag{1.82}$$

This is a prototypical equation; in most cases such equations express a conserved quantity, in this case it is the conservation of heat. It has the same meaning as Eq.(1.72), but whereas there the conservation law is formulated in integral form, in Eq.(1.82) it is formulated in its differential form. It is Gauss' theorem that allows to connect these two formulations.

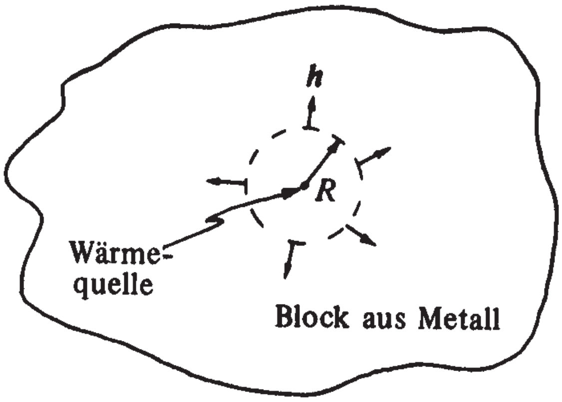

Figure 1.10: Heat source inside some material.

Let's consider now another scenario. Inside the block of material there is a small source of heat, maybe some small are where wires generate heat through resistance. This is sketched in Fig. 1.10. In any case, assume that the heat is generated in one point, where W is the energy which is generated per second. Still, heat conservation is implicit, and a case is considered where the energy has been generated and distributed for such a long time that temperature does not change any longer. This can be achieved only, if there is a constant flux of heat through the material. The natural question to ask now is: what is the form of $\vec{h}$ at every point inside the material, how much heat is floating? Due to the constance of temperature, each integration of the normal component of $\vec{h}$ over any closed surface including the heat source yields W. Therefore, all heat generated in the source has to float to the outside of such a surface as long as the flux is continuous. This translates into the problem to find a vector field ($\vec{h}$), which always yields W when its normal component is integrated over a surface including the source. This sounds complicated, but by choosing a simple surface S the solution is trivial. Just take a sphere with radius R and with the source in its centre. Due to the symmetry there is no preferred direction for the heat flux through the sphere - hence, the only reasonable direction for $\vec{h}$ is radial, and, for the same reason, it should have the same value all over the sphere. Well, if $\vec{h}$ is directed radial and has the same value everywhere, the integral of its normal component over the sphere is simple to calculate. The normal component of the sphere is radial and hence

$$\int\limits_S \vec{h} \cdot \vec{n} \mathrm{d}^2 s = 4\pi R^2 h \,, \tag{1.83}$$

where $h = |\vec{h}|$, the value of $\vec{h}$. This integral must equal the amount of heat W produced at the source per time interval. Therefore

$$h = \frac{W}{4\pi R^2} \quad \text{and} \quad \vec{h} = \frac{W}{4\pi R^2} \vec{e}_r \,, \tag{1.84}$$

where $\vec{e}_r$ is a unit vector pointing into the radial direction. This result shows that $\vec{h}$ is proportional to W and that it decreases with the inverse square of the distance to the source.
The result obtained so far is valid for heat fluxes close to point-like heat sources. But what happens in the general case, i.e. where the only condition maintained is that the amount of heat is conserved, but the sources may have any shape? The differential equation for the heat conductance has been derived in a crude approximation before, Eq.(1.47),

$$\vec{h} = -\kappa \vec{\nabla} T \,. \tag{1.85}$$

Combining this with Eq.(1.82) yields

$$-\frac{\mathrm{d}q}{\mathrm{d}t} = \vec{\nabla} \cdot \vec{h} = -\vec{\nabla} \cdot (\kappa \vec{\nabla} T) = -\kappa \nabla^2 T \,, \tag{1.86}$$

where κ is assumed to be constant. So far so good; let us add yet another assumption, namely that the temperature of the material is proportional to the heat energy per unit volume. Introducing the specific heat constant per unit volume c_v – often a good approximation – as a specific property of the material,

$$\Delta q = c_v \Delta T \quad \text{or, in differential form} \quad \frac{\mathrm{d}q}{\mathrm{d}t} = \frac{c_v \mathrm{d}T}{\mathrm{d}t} \,, \tag{1.87}$$

and combining these two equations yields

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\kappa}{c_v} \nabla^2 T \,. \tag{1.88}$$

In other words, the temporal change in the temperature field is proportional to the Laplace-operator of T, the second derivative of the field w.r.t. the spatial coordinates.
This differential equation is a diffusion equation, in that case the law of heat diffusion, where $D = \kappa/c_v$ denotes the corresponding diffusion constant. In fact, this type of diffusion can be found in many different contexts: the diffusion of gases, of neutrons inside materials etc..

Theorem of Stokes: Circulation

In analogy to the divergence in the following the rotor will be investigated. As a first step the circulation of a vector field will be calculated; this is the line integral of the vector field around a closed loop. If L is a closed loop enclosing an area S, then the line integral of $\vec{A}$ around L is given by

$$\oint\limits_{L=\delta S} \mathrm{d}l A_{\|} = \oint\limits_{L=\delta S} \vec{A} \cdot \mathrm{d}\vec{l}, \tag{1.89}$$

where the integral symbol hints at the fact that the integration path is closed. The name circulation for such an object stems from its first appearance in fluid mechanics, but, similar to the case of the flux discussed before, it has been extended to all vector fields.
Now, the partitioning of the area inside loop will be investigated, i.e. cutting the area S into two smaller areas S_1 and S_2, such that $S = S_1 + S_2$. The loop itself can then be written as $L = L_1 + L_2 - 2L_{ab}$, where L_{ab} is the extra line that had to be added in order to cut the area. Again, L_1 and L_2 can be written as a part belonging to the original line and the extra line, i.e. $L_{1,2} = L_{a,b} + L_{ab}$, cf. Fig. 1.11. What happens with the circulation of the vector field? Can the circulation around the full loop also be written as the sum of the circulations around the two sub-loops? The answer is yes! And it is quite easy to understand why this is the case: In order to define the line integral one has to define the direction of integration, for a closed loop this is tautological to defining a direction of circulation/rotation. In mathematics, the choice has been made to define the rotational direction to be counter-clockwise. When the two circulations are added, the extra piece L_{ab} contributes twice, but in opposite direction. Thus, these contributions cancel and it is clear that the circulation of a vector field around a loop can be partitioned. The sum of all circulations around tiny loops building a

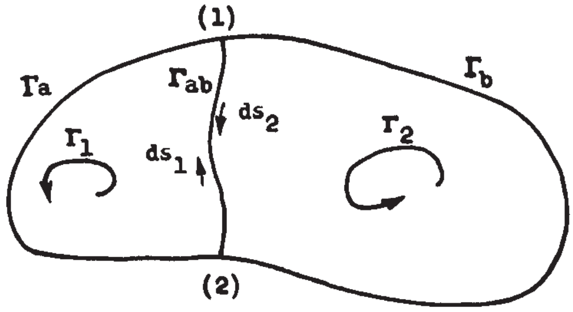

Figure 1.11: Cutting the loop.

larger one is just identical to the circulation around the larger loop. This is indicated in Fig. 1.12.
In particular, the small loops can now be taken as a little square (similar to the little cube that has been considered for the flux). For simplicity it will be oriented along the x- and y-axes within the x-y-plane, cf. Fig.1.13. The corner closest to the z-axis will have coordinates $(x,\, y)^T$ in the plane and the edges will have the lengths Δx and Δy respectively. To calculate the circulation around this square the tangential components along the four edges may be approximated with their x- or y-components at the centre of the respective edge. Then

$$\oint \vec{A} \cdot \mathrm{d}\vec{l} = A_x(1)\Delta x + A_y(2)\Delta y - A_x(3)\Delta x - A_y(4)\Delta y\,. \tag{1.90}$$

The minus signs arise due to the different orientation of the individual pieces. Again, the different components may be related through Taylor expansion, yielding for instance

$$\begin{aligned} A_x(3) &= A_x(1) + \frac{\partial A_x}{\partial y}\Delta y \\ A_y(4) &= A_y(2) - \frac{\partial A_y}{\partial x}\Delta x\,, \end{aligned} \tag{1.91}$$

where higher terms in Δx and Δy can sagely be neglected. Combining terms results in

$$\oint \vec{A} \cdot \mathrm{d}\vec{l} = \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\Delta x \Delta y\,. \tag{1.92}$$

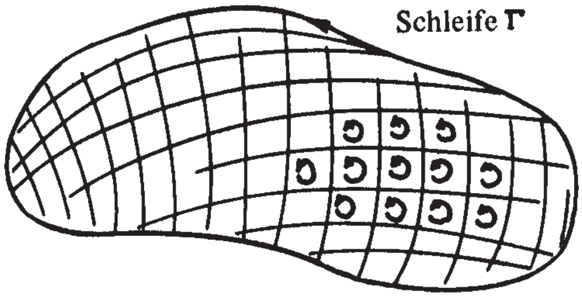

Figure 1.12: Many circulations around mini-loops form the circulation around the large one.

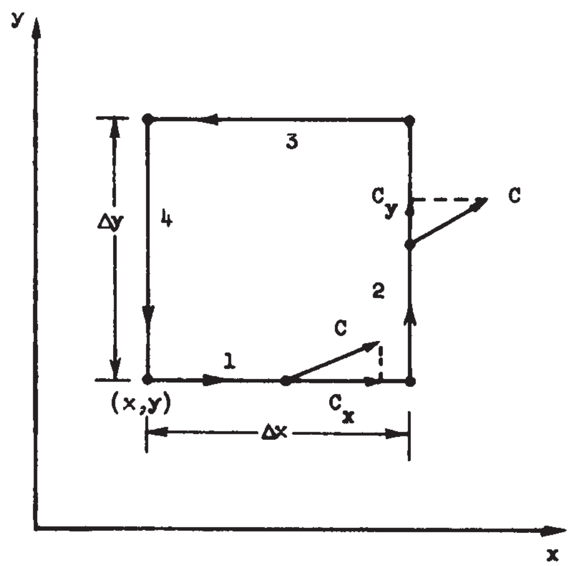

Figure 1.13: Circulation around an infinitesimal square

This is nice! The term in the bracket is just the z-component of the rotor of $\vec{A}$. But the z-component in this case is just normal to the plane, in which the square is located and therefore

$$\oint \vec{A} \cdot \mathrm{d}\vec{l} = \left(\vec{\nabla} \times \vec{A}\right)_z \Delta a = \left(\vec{\nabla} \times \vec{A}\right) \cdot \vec{n}_a \Delta a = , \tag{1.93}$$

where Δa is just the area of the square. The area may be given a direction such that

$$\vec{n}_a a =: \vec{a} , \tag{1.94}$$

this allows to write this as

$$\oint \vec{A} \cdot \mathrm{d}\vec{l} = \left(\vec{\nabla} \times \vec{A}\right) \cdot \Delta \vec{a} = \ . \tag{1.95}$$

The orientation is such that if you let the finger of your right hand follow the sense direction of the line integral, your thumb will point into the normal direction.
The result that the circulation of a vector field around an infinitesimally small square is just the component of the rotor of the field that is normal to the square times its area. Using again the partitioning of loops in a backward way, it is clear that large areas can be constructed out of small squares. The sum of all circulations around the squares can then be replaced with the circulation around the emerging area. As a result the following useful theorem emerges:
Theorem 5 (Stokes' theorem):
The integral of the normal component of the rotor of a vector field over a closed surface equals the circulation of this vector field around the surface

$$\int\limits_{L=\delta S} \vec{A} \cdot \mathrm{d}\vec{s} = \int\limits_{S} (\vec{\nabla} \times \vec{A}) \cdot \mathrm{d}^2\vec{a} \, . \tag{1.96}$$

1.4 Miscellaneous

In the remainder of this section, a number of barely connected, but, nevertheless, important mathematical tools will be presented.

1.4.1 Coordinate systems

Sometimes, physical problems exhibit a specific symmetry, like, for instance, spherical symmetry etc.. Owing to this it is often advantageous to choose suitable an orthogonal curvi-linear coordinate system for the description of the system in question. In three dimensions, such a coordinate system, denoted by coordinates ξ_i, is connected to the Cartesian coordinates x_i used so far through

$$\begin{aligned} x_1 &= x_1(\xi_1, \xi_2, \xi_3) \\ x_2 &= x_2(\xi_1, \xi_2, \xi_3) \\ x_3 &= x_3(\xi_1, \xi_2, \xi_3)\,. \end{aligned} \tag{1.97}$$

Consider now a vector $\vec{x}$ in Cartesian coordinates

$$\vec{x} = \sum_{i=1}^{3} x_i \vec{e}_i^{(x)} \tag{1.98}$$

with the orthonormal Cartesian base vectors $\vec{e}_i^{(x)}$. Usually, they form a right-handed coordinate system, pointing into the x-, y-, and z-direction. Forming the infinitesimal vector $\vec{x}$ in two different fashions,

$$\mathrm{d}\vec{x} = \sum_{i=1}^{3} \mathrm{d}x_i \vec{e}_i^{(x)} = \sum_{i=1}^{3} \mathrm{d}x_i \frac{\partial \vec{x}}{\partial x_i} \tag{1.99}$$

and comparing coefficients yields

$$\vec{e}_i^{(x)} = \frac{\partial \vec{x}}{\partial x_i}\,. \tag{1.100}$$

For a general coordinate system, labeled with the superscript (ξ) this translates to

$$\vec{e}_i(\xi) = b_{\xi_i}^{-1} \frac{\partial \vec{x}}{\partial \xi_i}\,, \tag{1.101}$$

where multiplying with the inverse of the scale factor

$$b_{\xi_i} = \left| \frac{\partial \vec{x}}{\partial \xi_i} \right| \tag{1.102}$$

guarantees that the new base vectors are normalised to one,

$$|\vec{e}_i(\xi)| = 1\,. \tag{1.103}$$

note that in the general case the *Dreibein* is not fixed like in the Cartesian case with its three axes, instead it rather “moves” with the vector.

Gradient and ∇-operator

In terms of these new base vectors and coordinates

$$\mathrm{d}\vec{x} = \sum_{i=1}^{3} \mathrm{d}\xi_i \frac{\partial \vec{x}}{\partial \xi_i} = \sum_{i=1}^{3} b_{\xi_i} \vec{e}_i^{(\xi)} \mathrm{d}\xi_i \,, \tag{1.104}$$

where the scale factors now act as kind of "metric".
The gradient, forming a vector, can be easily projected on components of the curvi-linear coordinates through a scalar product. For instance, the ith component of grad ϕ in the curvi-linear coordinates ξ reads

$$\begin{aligned} \left|\vec{\nabla}\phi\right|_{\xi_i} &= \vec{e}_i^{(\xi)} \cdot \vec{\nabla}_x \phi = b_{\xi_i}^{-1} \frac{\partial \vec{x}}{\partial \xi_i} \cdot \vec{\nabla}_x \phi \\ &= b_{\xi_i}^{-1} \left(\frac{\partial x_1}{\partial \xi_i} \frac{\partial \phi}{\partial x_1} + \frac{\partial x_2}{\partial \xi_i} \frac{\partial \phi}{\partial x_2} + \frac{\partial x_3}{\partial \xi_i} \frac{\partial \phi}{\partial x_3} \right) = b_{\xi_i}^{-1} \frac{\partial \phi}{\partial \xi_i} \,. \end{aligned} \tag{1.105}$$

Therefore, the gradient in curvi-linear coordinates reads

$$\vec{\nabla}_\xi = \left(b_{\xi_1}^{-1} \frac{\partial \phi}{\partial \xi_1}, \, b_{\xi_2}^{-1} \frac{\partial \phi}{\partial \xi_2}, \, b_{\xi_3}^{-1} \frac{\partial \phi}{\partial \xi_3} \right)^T = \sum_{i=1}^{3} b_{\xi_i}^{-1} \vec{e}_i^{(\xi)} \frac{\partial}{\partial \xi_i} \,. \tag{1.106}$$

Obviously, the difference w.r.t. Cartesian coordinates is in the appearance of the scale factors.

Divergence

First, write the divergence with help of the sum convention as

$$\vec{\nabla} \cdot \vec{a} = b_{\xi_i}^{-1} \vec{e}_i^{(\xi)} \frac{\partial}{\partial \xi_i} \cdot \left(a_j \vec{e}_j^{(\xi)} \right) = b_{\xi_i}^{-1} \frac{\partial a_i}{\partial \xi_i} + b_{\xi_i}^{-1} \vec{e}_i^{(\xi)} a_j \frac{\partial \vec{e}_j^{(\xi)}}{\partial \xi_i} \,, \tag{1.107}$$

where the two terms arise from differentiating both the coordinates and the base vectors with respect to the components of ξ and where

$$\vec{e}_i^{(\xi)} \cdot \vec{e}_j^{(\xi)} = \delta_{ij} \,, \tag{1.108}$$

i.e. the implicit assumption of an orthonormal coordinate system, has been used. To simplify the expression obtained for the divergence so far, consider

$$\frac{\partial^2 \vec{r}}{\partial \xi_i \partial \xi_j} = \frac{\partial^2 \vec{r}}{\partial \xi_j \partial \xi_i} \tag{1.109}$$

leading to

$$\frac{\partial}{\partial \xi_i}\left(b_{\xi_j}\vec{e}_j^{(\xi)}\right) = \frac{\partial}{\partial \xi_j}\left(b_{\xi_i}\vec{e}_i^{(\xi)}\right) . \tag{1.110}$$

Performing the differentiation and multiplying both sides with $\vec{e}_i^{(\xi)}$ yields

$$b_{\xi_j}\vec{e}_i^{(\xi)} \cdot \frac{\partial \vec{e}_j^{(\xi)}}{\partial \xi_i} + \delta_{ij}\frac{\partial b_{\xi_j}}{\partial \xi_i} = b_{\xi_i}\vec{e}_i^{(\xi)} \cdot \frac{\partial \vec{e}_i^{(\xi)}}{\partial \xi_j} + \frac{\partial b_{\xi_i}}{\partial \xi_j} \tag{1.111}$$

The first term on the r.h.s. vanishes due to

$$\vec{e}_i^{(\xi)} \cdot \frac{\partial \vec{e}_i^{(\xi)}}{\partial \xi_j} = \frac{1}{2}\frac{\partial}{\partial \xi_j}\left(\vec{e}_i^{(\xi)}\right)^2 = \frac{1}{2}\frac{\partial 1}{\partial \xi_j} = 0 . \tag{1.112}$$

Hence

$$b_{\xi_j}\vec{e}_i^{(\xi)} \cdot \frac{\partial \vec{e}_j^{(\xi)}}{\partial \xi_i} = \frac{\partial b_{\xi_i}}{\partial \xi_j} - \delta_{ij}\frac{\partial b_{\xi_j}}{\partial \xi_i} = (1-\delta_{ij})\frac{\partial b_{\xi_i}}{\partial \xi_j} . \tag{1.113}$$

Inserting this into Eq.(1.107) and exchanging some indices results in

$$\text{div }\vec{a} = b_{\xi_i}^{-1}\left[\frac{\partial a_i}{\partial \xi_i} + \frac{a_i}{b_{\xi_j}}(1-\delta_{ij})\frac{\partial b_{\xi_j}}{\partial \xi_i}\right] . \tag{1.114}$$

For the $i = 1$ piece this reads

$$\begin{aligned} b_{\xi_1}^{-1}\left[\frac{\partial a_1}{\partial \xi_1} + \frac{a_1}{b_{\xi_j}}(1-\delta_{1j})\frac{\partial b_{\xi_j}}{\partial \xi_1}\right] &= \frac{1}{b_{\xi_1}}\left[\frac{\partial a_1}{\partial \xi_1} + \frac{a_1}{b_{\xi_2}}\frac{\partial b_{\xi_2}}{\partial \xi_1} + \frac{a_1}{b_{\xi_3}}\frac{\partial b_{\xi_3}}{\partial \xi_1}\right] \\ = \frac{1}{b_{\xi_1}b_{\xi_2}b_{\xi_3}}\left[b_{\xi_2}b_{\xi_3}\frac{\partial a_1}{\partial \xi_1} + a_1 b_{\xi_3}\frac{\partial b_{\xi_2}}{\partial \xi_1} + a_1 b_{\xi_2}\frac{\partial b_{\xi_3}}{\partial \xi_1}\right] &= \frac{1}{b_{\xi_1}b_{\xi_2}b_{\xi_3}}\frac{\partial(a_1 b_{\xi_2}b_{\xi_3})}{\partial \xi_1} . \end{aligned} \tag{1.115}$$

Similar equations hold true for the other tow terms, i.e. for $i = 2$ and $i = 3$. The final result reads

$$\vec{\nabla} \cdot \vec{a} = \frac{1}{b_{\xi_1}b_{\xi_2}b_{\xi_3}}\left[\frac{\partial(a_1 b_{\xi_2}b_{\xi_3})}{\partial \xi_1} + \frac{\partial(a_2 b_{\xi_1}b_{\xi_3})}{\partial \xi_2} + \frac{\partial(a_3 b_{\xi_1}b_{\xi_2})}{\partial \xi_3}\right] . \tag{1.116}$$

Rotor

For the construction of the rotor in curvi-linear orthonormal coordinates it is convenient to start with a little detour. From

$$\text{grad}\ \xi_i = \vec{\nabla}_\xi \xi_i = b_{1_i}^{-1} \vec{e}_i^{(\xi)} \quad \Longrightarrow \quad \vec{e}_i^{(\xi)} = b_{1_i} \vec{\nabla}_\xi \xi_i\,, \tag{1.117}$$

which will be used when rewriting

$$\text{rot}\ \vec{a} = \vec{\nabla} \times \vec{a} = \vec{\nabla} \times \left(a_i \vec{e}_i^{(\xi)} \right) \tag{1.118}$$

by replacing the base vectors. Concentrating on the first term only, it can be cast into the following form

$$\begin{aligned} \vec{\nabla} \times \left(a_1 \vec{e}_1^{(\xi)} \right) &= \vec{\nabla} \times \left(a_1 b_{\xi_1} \vec{\nabla}_\xi \xi_1 \right) = \vec{\nabla} \left(a_1 b_{\xi_1} \right) \times \vec{\nabla}_\xi \xi_1 + \left(a_1 b_{\xi_1} \right) \vec{\nabla} \times \vec{\nabla}_\xi \xi_1 \\ &= \vec{\nabla} \left(a_1 b_{\xi_1} \right) \times \vec{\nabla}_\xi \xi_1 = \vec{\nabla} \left(a_1 b_{\xi_1} \right) \times \left(b_{\xi_1}^{-1} \vec{e}_1^{(\xi)} \right) \\ &= \left[\frac{\vec{e}_i^{(\xi)}}{b_{\xi_i}} \frac{\partial}{\partial \xi_i} \left(a_1 b_{\xi_1} \right) \right] \times \left(b_{\xi_1}^{-1} \vec{e}_1^{(\xi)} \right) \\ &= \frac{1}{b_{\xi_1} b_{\xi_2} b_{\xi_3}} \left[\vec{e}_2^{(\xi)} b_{\xi_2} \frac{\partial \left(a_1 b_{\xi_1} \right)}{\partial \xi_2} - \vec{e}_3^{(\xi)} b_{\xi_3} \frac{\partial \left(a_1 b_{\xi_1} \right)}{\partial \xi_3} \right] . \end{aligned} \tag{1.119}$$

Repeating the same exercise for the other components and arranging terms yields the final result for the rotor in curvi-linear orthonormal systems, written as a determinant,

$$\text{rot}\ \vec{a} = \frac{1}{b_{\xi_1} b_{\xi_2} b_{\xi_3}} \begin{vmatrix} \vec{e}_1^{(\xi)} b_{\xi_1} & \vec{e}_2^{(\xi)} b_{\xi_2} & \vec{e}_3^{(\xi)} b_{\xi_3} \\ \partial/\partial \xi_1 & \partial/\partial \xi_2 & \partial/\partial \xi_3 \\ a_1 b_{\xi_1} & a_2 b_{\xi_2} & a_3 b_{\xi_3} \end{vmatrix} \tag{1.120}$$

Integration

The differential volume element $\mathrm{d}^3 V$ becomes

$$\mathrm{d}^3 V = \mathrm{d}^3 x = \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 = J(x_i,\, \xi_j)\, \mathrm{d}\xi_1 \mathrm{d}\xi_2 \mathrm{d}\xi_3 \tag{1.121}$$

where the Jacobian reads

$$J(x_i,\, \xi_j) = \det\frac{\partial x_i}{\partial \xi_j} = \begin{vmatrix} \dfrac{\partial x_1}{\partial \xi_1} & \dfrac{\partial x_1}{\partial \xi_2} & \dfrac{\partial x_1}{\partial \xi_3} \\ \dfrac{\partial x_2}{\partial \xi_1} & \dfrac{\partial x_2}{\partial \xi_2} & \dfrac{\partial x_2}{\partial \xi_3} \\ \dfrac{\partial x_3}{\partial \xi_1} & \dfrac{\partial x_3}{\partial \xi_2} & \dfrac{\partial x_3}{\partial \xi_3} \end{vmatrix} . \tag{1.122}$$

Examples: Cylinder and spherical coordinates

Two typical orthogonal curvi-linear coordinate systems in three dimensions with frequent application in physics are:

- Cylinder coordinates:
  They are given through polar coordinates in one plane, say the x-y-plane, supplemented by the z-coordinate.

$$\begin{aligned} x &\to \rho\cos\phi \\ y &\to \rho\sin\phi \\ z &\to z\,, \end{aligned} \tag{1.123}$$

  where

$$\rho = \sqrt{x^2+y^2} \quad \text{and} \quad \phi = \arctan\frac{x}{y}\,. \tag{1.124}$$

  The volume element is given through the Jacobean of the transformation,

$$\frac{\partial(x,\, y,\, z)}{\partial(\rho,\, \phi,\, z)} = \begin{vmatrix} \cos\phi & -\rho\sin\phi & 0 \\ \sin\phi & \rho\cos\phi & 0 \\ 0 & 0 & 1 \end{vmatrix} = \rho\,. \tag{1.125}$$

  leading to

$$\mathrm{d}^3V = \mathrm{d}^3x = \frac{\partial(x,\, y,\, z)}{\partial(\rho,\, \phi,\, z)}\mathrm{d}\rho\,\mathrm{d}\phi\,\mathrm{d}z = \rho\mathrm{d}\rho\,\mathrm{d}\phi\,\mathrm{d}z\,. \tag{1.126}$$

  The lines described by two coordinates being constant are:

- ρ-line: Radial line in the x-y-plane, leaving from the z-axis.
- ϕ-line: Circle in the x-y-plane.
- z-line: Line parallel to the z-axis.

The corresponding unit base vectors are

$$\begin{aligned}
b_\rho \vec{e}_\rho &= \frac{\partial \vec{r}}{\partial \rho} = \frac{\partial}{\partial \rho} \begin{pmatrix} \rho\cos\phi \\ \rho\sin\phi \\ z \end{pmatrix} = \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} \qquad \Longrightarrow \quad b_\rho = 1 \\
b_\phi \vec{e}_\phi &= \frac{\partial \vec{r}}{\partial \phi} = \frac{\partial}{\partial \phi} \begin{pmatrix} \rho\cos\phi \\ \rho\sin\phi \\ z \end{pmatrix} = \begin{pmatrix} -\rho\sin\phi \\ \rho\cos\phi \\ 0 \end{pmatrix} \quad \Longrightarrow \quad b_\phi = \rho \\
b_z \vec{e}_z &= \frac{\partial \vec{r}}{\partial z} = \frac{\partial}{\partial z} \begin{pmatrix} \rho\cos\phi \\ \rho\sin\phi \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \Longrightarrow \quad b_z = 1
\end{aligned} \tag{1.127}$$

It is simple to check that these three vectors are orthogonal.
Thus the differential reads

$$\mathrm{d}\vec{r} = \vec{e}_\rho \, \mathrm{d}\rho + \vec{e}_\phi \, \rho \, \mathrm{d}\phi + \vec{e}_z \, \mathrm{d}z \tag{1.128}$$

and the gradient becomes

$$\vec{\nabla} = \vec{e}_\rho \frac{\partial}{\partial \rho} + \vec{e}_\phi \frac{\partial}{\rho \partial \phi} + \vec{e}_z \frac{\partial}{\partial z} \tag{1.129}$$

The divergence reads

$$\operatorname{div} \vec{A} = \frac{1}{\rho} \frac{\partial(\rho A_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial A_\phi}{\partial \phi} + \frac{\partial (A_z)}{\partial z} \,, \tag{1.130}$$

and the rotor is given by

$$\begin{aligned}
&\operatorname{rot} \vec{A} = \\
&\quad \vec{e}_\rho \left(\frac{1}{\rho} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right) + \vec{e}_\phi \left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho} \right) + \vec{e}_z \frac{1}{\rho} \left(\frac{\partial (\rho A_\phi)}{\partial \rho} - \frac{\partial A_\rho}{\partial \phi} \right) .
\end{aligned} \tag{1.131}$$

Finally, the Laplace-operator reads

$$\nabla^2 = \frac{\partial}{\rho \partial \rho} \left(\frac{\rho \partial}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \,. \tag{1.132}$$

- Spherical coordinates:
  They are given through

$$\begin{aligned} x &\rightarrow r\sin\theta\cos\phi \\ y &\rightarrow r\sin\theta\sin\phi \\ z &\rightarrow r\cos\theta\,, \end{aligned} \tag{1.133}$$

  where

$$r = \sqrt{x^2+y^2+z^2}\,. \tag{1.134}$$

  The construction of the base vectors and the derivation of the gradient, divergence, rotor, Laplace operator and volume element are left as an exercise.

1.4.2 δ-function

For this lecture it will be useful to introduce Dirac's δ-function. Mathematically speaking, however, it is a distribution rather than a function. However, in one dimension it is defined through

$$\begin{aligned} \delta(x-a) &= 0 \quad \text{for } x \neq a \\ \int \mathrm{d}x\delta(x-a) &= \begin{cases} 1 & \text{if } x = a \in \text{region of integration} \\ 0 & \text{else} \end{cases} \end{aligned} \tag{1.135}$$

[3]

[3]The δ-function can be understood through some simple consideration: Assume it to be the limit of a function with a strongly developed, unique maximum, such as Gauss' function, where the limit is such that the maximum becomes increasingly pronounced but the area under the function is fixed to equal one. In fact, the δ-function can be defined as the limit of a series of functions $\delta_n(x)$ which are normalised to one,

$$\delta(x) = \lim_{n\to\infty} \delta_n(x)\,, \tag{1.136}$$

where also

$$\lim_{n\to\infty} \int \mathrm{d}x\delta_n(x-a)f(x) = f(a)\,. \tag{1.137}$$

From this definition, the following properties of δ-functions emerge:

$$\begin{aligned} \int \mathrm{d}x\delta(x-a)f(x) &= f(a) \\ \int \mathrm{d}x f(x)\frac{\mathrm{d}\delta(x-a)}{\mathrm{d}x} &= -\frac{\mathrm{d}f}{\mathrm{d}x}(a)\,, \end{aligned} \tag{1.139}$$

where the latter follows from partial integration. Also

$$\delta\left[f(x)\right] = \sum_i \frac{\delta(x-x_i)}{\left|\frac{\mathrm{d}f}{\mathrm{d}x}(x_i)\right|}\,, \tag{1.140}$$

where x_i are denote the zeroes of f, i.e. $f(x_i) = 0$. In particular,

$$\delta(x^2-a^2) = \frac{\delta(x-a)+\delta(x+a)}{2|a|}\,. \tag{1.141}$$

This definition can be extended to more than one dimensions. For instance,

$$\delta^3(\vec{x}-\vec{y}) = \delta(x_1-y_1)\delta(x_2-y_2)\delta(x_3-y_3) \tag{1.142}$$

in Cartesian coordinates, where the individual components are numbered. Consequently,

$$\int\limits_V \mathrm{d}^3x\delta^3(\vec{x}) \;=\; \left\{\begin{array}{ll} 1 & \text{if} \quad x\in V \\ 0 & \text{else} \end{array}\right. \tag{1.143}$$

may serve as a definition for a higher-dimensional δ-function, resulting in

$$\int \mathrm{d}^3x\delta^3(\vec{x}-\vec{a})f(\vec{x}) \;=\; f(\vec{a})\,. \tag{1.144}$$

When going from Cartesian coordinates $(x_1,\, x_2,\, x_3)^T$ to another set of coordinates $(\xi_1\,\xi_2\,\xi_3)^T$ the volume element transforms like

$$\mathrm{d}^3x = J(x_i,\,\xi_j)\mathrm{d}^3\xi\,, \tag{1.145}$$

Examples for such series include

$$\begin{aligned} \delta_n(x) &= ne^{-\pi n^2x^2} \\ \delta_n(x) &= \frac{n}{\pi}\left(\frac{\sin(nx)}{nx}\right)^2\,. \end{aligned} \tag{1.138}$$

where J is the Jacobian of the transformation linking both. From the integral definition of the δ-function it is straightforward to obtain

$$\delta^3(\vec{x}) = J^{-1}(x_i,\, \xi_i)\delta(\xi_1)\delta(\xi_2)\delta(\xi_3)\,. \tag{1.146}$$

The n-dimensional δ-function always has the dimension of an inverse n-dimensional volume (in corresponding coordinates).

Chapter 2

Basic equations

There are two groups of objects playing a role in the description of electrodynamics:

- Electrical charges and electrical currents, where the latter emerges through moving electrical charges. “Charge” can be understood as a quantitative way of describing how an object (like, e.g. a particle) can participate in an interaction transmitted through fields. This participation can be understood as a process of emitting fields or of experiencing forces that are communicated through already existing fields (which in turn may have been emitted by other “charges” of the same kind). This is quite analogous to the concept of “mass” generating gravitational fields or enabling a body to experience the force brokered through gravity.

- Electric and magnetic field. In static problems the fields appear to be purely auxiliary: they merely transmit forces between charges, currents, etc.. The forces obey Newton’s mechanics (and, thus, the “actio=reactio”-principle) and they all have the form $\propto r^{-2}$, similar to gravity. Only in dynamical problems the fields start to become the important quantities: a body produces a field which in turn spreads out. In its propagation the fields have a maximal velocity (the speed of light), under which they causally connect the emission of the field and their effect in terms of a force. The propagation of the fields can then be described independent of their emitter: electromagnetic waves emerge.

In this chapter the interplay of these object will be presented. It is organised through a set of equations known as the Maxwell equations. At first glance they have a complicated structure; but in fact they incorporate the first unification of forces that have been connected with distinct phenomena, thus leading to a far simpler interpretation. They also helped to establish that light is nothing else but an electromagnetic wave.

2.1 Charges, currents, fields and forces

2.1.1 Charge and current density

Charge density

In most cases, it is not sufficient to know that there is a specific amount of charge in a volume, instead it is rather important to know its exact distribution as a function of space, and, eventually, time. This gives rise to the notion of a charge density, defined as

$$\rho(t,\,\vec{x}) = \frac{\mathrm{d}^3Q(t)}{\mathrm{d}^3V} = \frac{\mathrm{d}^3Q(t)}{\mathrm{d}^3\vec{x}} = \frac{\mathrm{d}^3Q(t)}{\mathrm{d}x^3}\,. \tag{2.1}$$

The implication of this should be clear: the charge density is the amount of charge in an infinitesimally small volume. Since charges can be positive or negative, the charge density also may have both signs. And, of course, the density may fluctuate with position. For a simple, point-like charge q at position $\vec{x}_0$ however, the charge density can be written by means of the δ-function as

$$\rho_{\mathrm{point}}(t,\,\vec{x}) = q\delta^3(\vec{x}-\vec{x}_0)\,. \tag{2.2}$$

The total charge per volume V in contrast results from an integration of the charge density over this volume,

$$Q(t) = \int\limits_V \mathrm{d}Q(t) = \int\limits_V \mathrm{d}^3V\rho(t,\,\vec{x})\,, \tag{2.3}$$

for the point-like charge distribution of Eq.(2.2) one immediately finds

$$Q(t) = \int\limits_V \mathrm{d}^3V\rho(t,\,\vec{x}) = q\int\limits_V \mathrm{d}^3x\delta^3(x-x_0) = \left\{\begin{array}{ll} q & \text{if } x_0\in V \\ 0 & \text{else} \end{array}\right. \tag{2.4}$$

If ρ is homogeneously distributed over a certain volume, then $\rho = Q/V$, where Q is the total charge inside this volume.[1] Since charges are ultimately connected to elementary particles, such as electrons, which are not densely packed, the "infinitesimal" volumes in question should not be too small, because at sub-atomic scales the charge densities become δ functions.

Current density

The electric current through an area A is defined as

$$I(t) = \frac{\mathrm{d}Q_A(t)}{\mathrm{d}t}\,, \tag{2.6}$$

where Q_A denotes the charge passing per time interval through the area. Positive currents emerge from positive charges passing the area in positive direction or from negative charges passing in negative direction. This leaves some arbitrariness due to the definition of positive and negative direction. Charges that do not pass the area do not contribute. Again, time-intervals $\mathrm{d}t$ should not be too small, because then only individual particles would pass through the area, inducing some type of shot-noise.
For many applications however, it is not sufficient just to know the integral electric current $I(t)$; rather its spatial distribution must be known. Phenomena where this is the case include gas discharges, eddy currents in extended conductors, etc.. In contrast, when dealing with thin conductors only, usually the integral current is sufficient, since then the distribution can be sagely averaged over. In the former case, however, it may be important to consider the distribution as ratio of the current and the (oriented) area $\vec{A}$,

$$\vec{j} = \frac{\mathrm{d}^2 I(t)}{\mathrm{d}^2\vec{A}}\,. \tag{2.7}$$

To gain a better understanding of this, assume that there is a flux of charged particles passing through the area A with velocity v, parallel to the z-axis,

---

[1] To connect charge densities with point-like charges, consider a sphere of radius r_0 with a uniform charge distribution $\rho = \rho_0$ inside the sphere and no charges in the outside, $\rho = 0$. When the total charge is constant, say q_0, then the value of the constant charge density, ρ_0, in dependence on the radius reads

$$\rho_0(r) = q_0/V_{\text{sphere}} = q_0/(4\pi r_0^3/3)\,. \tag{2.5}$$

Going to smaller and smaller spheres leaves the total charge constant, it just shifts the value of the charge density ρ_0. For $r_0 \to 0$ this object becomes a δ-function.

$v = v\vec{e}_z$. Take the charge density to be spatially homogenous, $\rho = \rho_0$. Then, at any moment, the total amount of charge in front of A passing through it within the next time interval $\mathrm{d}t$ is

$$\mathrm{d}Q(t) = \mathrm{d}^3x\rho(t,\, \vec{x}) = \mathrm{d}^2Av\mathrm{d}t\rho_0\,, \tag{2.8}$$

This leads to

$$\mathrm{d}I(t) = \frac{\mathrm{d}Q}{\mathrm{d}t} = \mathrm{d}^2\vec{A}\cdot\vec{v}\rho_0\,. \tag{2.9}$$

If the flux consists of different particles with different velocities, and hence, the charge density fluctuates the overall result can be written as a superposition of currents stemming from the individual particles. However, it is easy to see that the current density per particle of charge q reads

$$\vec{j} = q\vec{v}\,. \tag{2.10}$$

2.1.2 Connecting fields with force

As discussed before, the charge quantifies the way an object experiences fields, inducing forces acting on it. Experimentally it is known that in the non-relativistic case

$$\vec{F} = q\vec{E} + q\vec{v}\times\vec{B} = q\vec{E} + \vec{j}\times\vec{B}\,. \tag{2.11}$$

Recalling that a force induces a change in the momentum of an object,

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{m\vec{v}}{\sqrt{1-v^2/c^2}}\right] \tag{2.12}$$

allows to construct an Equation of Motion (E.o.M.) for relativistic particles,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{m\vec{v}}{\sqrt{1-v^2/c^2}}\right] = q\vec{E} + \vec{j}\times\vec{B}\,. \tag{2.13}$$

2.2 Maxwell equations

2.2.1 Differential and integral forms

The focus will shift now on the presentation of the Maxwell equations. It is interesting to note that their structure is deeply connected with the behaviour of vectors and scalars under transformations of space-time, in particular with spatial inversions (reflections). Remember that there was a difference between polar and axial vectors, related to their behaviour under reflections. For axial vectors some variation of a right-hand rule is necessary, which allows them to decide whether they actually live in a right- or a left-handed coordinate system. In any physically sensible equation, however, things on the left hand side and on the right hand side have to have the same behaviour. Thus, one may start with a simple classification like:

$$\begin{array}{lcl} \text{Scalar quantities} & : & \rho,\, \vec{\nabla}\cdot\vec{E},\, \vec{\nabla}\cdot\vec{B},\, \ldots \\ \text{Vector quantities} & : & \vec{j},\, \vec{E},\, \dot{\vec{E}},\, \vec{B},\, \dot{\vec{B}},\, \vec{\nabla}\times\vec{E},\, \vec{\nabla}\times\vec{B},\, \ldots \end{array}$$

But this can be refined, when taking into account the polar or axial character of the vectors:
$\vec{j}$ and $\vec{E}$ are polar vectors; $\vec{j}$ will remain the same independent of the orientation of the coordinate system. The force $\vec{F}$ is also a polar vector for the same reason, due to $\vec{F} = q\vec{E}$ in the absence of magnetic field, also $\vec{E}$ is a polar vector. Since differentiating with respect to a scalar does not change the behaviour under reflection, also $\partial\vec{E}/\partial t$ is a polar vector. Scalar or dot-products do not alter the parity under reflections either, hence $\vec{\nabla}\cdot\vec{E} = \text{div } \vec{E}$ is a proper scalar. Due to the vector or cross-product, implying an orientation, $\vec{\nabla}\times\vec{E} = \text{rot } \vec{E}$ constitutes an axial vector.
In contrast, $\vec{B}$ is an axial vector. This can be seen from two simple considerations. First of all, in the absence of electric fields $\vec{F} = q\vec{v}\times\vec{B}$ where $\vec{F}$ and $\vec{v}$ are polar vectors. Since the cross-product induces an orientation, $\vec{B}$ has to be axial. Also, since the angular velocity $\vec{\omega}$ can be defined through a vector product, it is axial. Remember that for particles in a homogenous magnetic field $\vec{\omega} = q/m\vec{B}$, thus $\vec{B}$ has to be axial as well. Of course then $\partial\vec{B}/\partial t$ is also axial, and div $\vec{B}$ is a pseudo-scalar, changing its sign under reflections. In turn, rot $\vec{B}$ is a polar vector, because both the cross product and $\vec{B}$ itself need an orientation.
The refinement reads:

Scalars	:	ρ, $\vec{\nabla}\cdot\vec{E}$
Pseudo-scalars	:	$\vec{\nabla}\cdot\vec{B}$
Polar vectors	:	$\vec{j}$, $\vec{E}$, $\dot{\vec{E}}$, $\vec{\nabla}\times\vec{B}$
Axial vectors	:	$\vec{B}$, $\dot{\vec{B}}$, $\vec{\nabla}\times\vec{E}$

Differential form

According to the classification above, one may be tempted to guess the most general form of linear equations, where only the charge and current densities and the first derivatives of the fields w.r.t. either space or time appear.
Let us start with the scalars. Since there are only two scalars it is not difficult to guess that they should be proportional, $\rho \propto \vec{\nabla}\cdot\vec{E}$. The only pseudo-scalar can be zero only, because this is the only scalar at hand for which a change of sign is inconsequential, $\vec{\nabla}\cdot\vec{B} = 0$. The polar vectors should be connected, defining an a priori unknown constant η one may guess that $\vec{\nabla}\times\vec{B} + \eta\partial\vec{E}/\partial t \propto \vec{j}$. Finally, the axial vectors may be connected through $\vec{\nabla}\times\vec{E} \propto \partial\vec{B}/\partial t$.
For some obvious reason, the constants of proportionality depend on the measure in which physical objects will be quantified. Throughout this lecture the **MKSA** units will be used, which has its merits especially for macroscopic and non-relativistic applications. In these units the Maxwell equations in differential form read:

$$\begin{array}{rclcrcl} \vec{\nabla}\cdot\vec{E} & = & \dfrac{\rho}{\epsilon_0} & \qquad & \vec{\nabla}\cdot\vec{B} & = & 0 \\ \vec{\nabla}\times\vec{E} + \dfrac{\partial\vec{B}}{\partial t} & = & 0 & \qquad & \vec{\nabla}\times\vec{B} - \epsilon_0\mu_0\dfrac{\partial\vec{E}}{\partial t} & = & \mu_0\vec{j}\,, \end{array} \tag{2.14}$$

where ϵ_0 and μ_0 are two constants defined through the system charges and currents are measured in. In fact, due to these equations describing electromagnetic waves, i.e. light, these two constants are related through the speed of light c

$$\epsilon_0\mu_0 = \frac{1}{c^2}\,, \quad c \approx 3\cdot 10^8\frac{m}{s}\,. \tag{2.15}$$

This allows to eliminate the magnetic induction constant μ_0. From measurements of the force between two charged condensator plates the electric induction constant ϵ_0 is determined as

$$\frac{1}{4\pi\epsilon_0} \approx 9 \cdot 10^9 \frac{V s}{Am} \,. \tag{2.16}$$

Potentials

From theorems 1 and 2, Eqs.(1.60) and (1.61), some first consequences can be drawn with respect to the electric and magnetic field. Let's start with the latter. Maxwell's equations assert that the divergence of the magnetic field is always zero,

$$\text{div } \vec{B} = 0 \,; \tag{2.17}$$

hence, according to theorem 2, Eq.(1.61), there must be a vector field $\vec{A}$ such that

$$\vec{B} = \text{rot } \vec{A} = \vec{\nabla} \times \vec{A} \,. \tag{2.18}$$

This field is usually referred to as vector potential. Employing this vector potential to identify

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = \vec{\nabla} \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t} \right) = 0 \tag{2.19}$$

and using theorem 1, Eq.(1.61), it is straightforward to see that the combination $\vec{E} + \partial\vec{A}/\partial t$ can be written as the gradient of a scalar function,

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} = \text{grad } \phi = -\vec{\nabla}\phi \,, \tag{2.20}$$

where the minus sign is a free choice, for some technical simplicity later on. However, ϕ is usually known as the scalar potential. In the static case, it reduces to the electric potential ϕ with

$$\vec{E} = -\text{grad } \phi \,. \tag{2.21}$$

Let's return to the vector potential. From

$$\vec{\nabla} \times \vec{\nabla}\phi = 0 \tag{2.22}$$

it is clear that the magnetic fields $\vec{B}$ won't change when adding the gradient of a scalar function to the vector potential $\vec{A}$, i.e. $\vec{B}$ remains invariant under a transformation

$$\vec{A} \to \vec{A}' = \vec{A} + \vec{\nabla}\psi \,. \tag{2.23}$$

In general, however, under such a transformation, $\vec{E}$ would change:

$$\vec{E}' = -\vec{\nabla}\phi - \frac{\partial(\vec{A} - \vec{\nabla}\psi)}{\partial t} \neq \vec{E} \,. \tag{2.24}$$

The difference

$$\vec{E}' - \vec{E} = \vec{\nabla}\frac{\partial(\psi)}{\partial t} \tag{2.25}$$

can be cast into the form of a gradient. This means that $\vec{E}$ remains invariant if the scalar potential is changed together with the vector potential such that

$$\vec{A} \to \vec{A}' = \vec{A} + \vec{\nabla}\psi \quad \text{and} \quad \phi \to \phi' = \phi - \frac{\partial(\psi)}{\partial t} \,. \tag{2.26}$$

Such a combined transformation is called gauge transformation. It changes the, in principle unmeasurable, potentials, and leaves the physical quantities, the fields, invariant.
As it will turn out later it is in many cases easier and more practical to find solutions for these potentials rather than for the fields themselves. In addition the formulation in terms of these potentials lends itself to a manifestly covariant formulation of electrodynamics in terms of a vector-potential

$$A^\mu = (\phi,\, \vec{A})^T \,. \tag{2.27}$$

Introducing, in full analogy, a four-vector current density

$$j^\mu = (\rho,\, \vec{j})^T \tag{2.28}$$

allows a Lagrangian formulation of electrodynamics. From this, the full machinery of deriving the E.o.M. as Euler-Lagrange equations through a variational principle, a systematic discussion of symmetries, and of quantising the electromagnetic field, becomes available.

Integral form

Using Gauss' and Stokes theorem it is easy to obtain the integral form of the Maxwell equations. They read

$$\begin{aligned}
\vec{\nabla}\cdot\vec{E} &= \frac{\rho}{\epsilon_0} &\Longleftrightarrow\quad& \oint\limits_{\delta V}\mathrm{d}^2\vec{s}\cdot\vec{E} = \frac{1}{\epsilon_0}\int\limits_V \mathrm{d}^3x\,\rho &=& \frac{Q}{\epsilon_0}\\
\vec{\nabla}\cdot\vec{B} &= 0 &\Longleftrightarrow\quad& \oint\limits_{\delta V}\mathrm{d}^2\vec{s}\cdot\vec{B} &=& 0\\
\vec{\nabla}\times\vec{E}+\frac{\partial\vec{B}}{\partial t} &= 0 &\Longleftrightarrow\quad& \oint\limits_{\delta S}\mathrm{d}\vec{l}\cdot\vec{E}+\int\limits_S \mathrm{d}^2\vec{s}\cdot\frac{\partial\vec{B}}{\partial t} &=& 0\\
c^2\vec{\nabla}\times\vec{B}-\frac{\partial\vec{E}}{\partial t} &= \frac{1}{\epsilon_0}\vec{j} &\Longleftrightarrow\quad& c^2\oint\limits_{\delta S}\mathrm{d}\vec{l}\cdot\vec{B}-\int\limits_S \mathrm{d}^2\vec{s}\cdot\frac{\partial\vec{E}}{\partial t} &=& \frac{1}{\epsilon_0}\int\limits_S \mathrm{d}^2\vec{s}\cdot\vec{j}\,.
\end{aligned} \tag{2.29}$$

Interpretation

All laws have in common that they are linear in the first derivatives of the fields. Hence, the evolution of the fields is known for all times, once the system of fields, sources and currents has been fixed for an arbitrary time t_0 and when the evolution of the currents and sources with time is known. Since they are usually connected to some mechanical equations of motion, the full system of fields and their sources behaves fully deterministic.
Also, the meaning of the individual equations is quite clear:

- The first law states that the electric field has sources. The flux of the electric field out of a volume is directly proportional to the total charge inside this volume. The electric field is a source-field.

- The second law states that the magnetic field has no sources. Stated the other way around: There are no magnetic mono-poles - no free, “single-poled” charges. Pictorially speaking, magnets always have a north- and a south-pole, whereas electric charges can be either positively or negatively charged.

- The third law is also called the induction law: any time-dependent magnetic field induces an usually also time-dependent electric curl-field. Note however that the induction phenomena related to a moving conducting loop are **not** incorporated in this law.

- The fourth law, also known as the Oerstedt-Ampere law, states that the curls of the magnetic field are produced through temporal changes of an electric field or through an electric current. Of course, in the static case, the former term, the temporal change of an electric field vanishes. Nevertheless, if it exists it acts like any current, resulting in a magnetic field. This formal equivalence has been re-interpreted as physical equivalence, and the integral of the term $\epsilon_0 \, \partial\vec{E}/\partial t$ has been called the displacement current, I_V. This naming has become custom, although it is quite misleading. After all, the displacement current is nothing else but the derivative of the electric field w.r.t. time.

The reasoning so far may provoke the perception that Maxwell's equations are a result of hard thinking only. In fact, it would be quite easy to take the following point of view: Well, I'd like to describe a system of two vector fields, which are intimately connected with each other, but still different. To leave them different, I'll make one a polar and the other one an axial vector. I'd like to describe their dynamics, i.e. their changes with space and/or time, therefore I need at least their first derivatives. Since I'm lazy I'll try to keep the equations linear in all ingredient, and, for greater simplicity, I'll keep only derivatives of the fields. Such a reasoning would indeed lead to a similar system of basic equations. Such a reasoning however, is quite short-sighted. After all, ideas like the linearity of fields, the absence of magnetic mono-poles, etc. are related to experimental findings. There is a superposition principle of fields, tied to the linearity of the equations, and after all, no magnetic mono-poles have been found yet. This proves that experimental findings, empirical facts, are the basis and the ultimate test of any theory - considerations like the catalogue done here just help to select a suitable mathematical structure, a model, that can be tested.

2.2.2 Conserved quantities

A continuity equation

One of the first consequences of the Maxwell equations is that they incorporate a conserved quantity: electrical charge. To see this, consider the

Oerstedt-Ampere law:

$$\frac{1}{\mu_0}\vec{\nabla} \times \vec{B} = \epsilon_0 \frac{\partial \vec{E}}{\partial t} + \vec{j}\,, \tag{2.30}$$

form the divergence on both sides and apply that div $\vec{E} = \epsilon_0 \rho$,

$$\frac{1}{\mu_0}\vec{\nabla} \cdot \left(\vec{\nabla} \times \vec{B}\right) = \vec{\nabla} \cdot \left[\epsilon_0 \frac{\partial \vec{E}}{\partial t} + \vec{j}\right] = \epsilon_0 \frac{\partial(\vec{\nabla} \cdot \vec{E})}{\partial t} + \vec{\nabla} \cdot \vec{j} = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j}\,. \tag{2.31}$$

The l.h.s. of this equation equals zero for all configurations of $\vec{B}$. This results in a continuity equation for the charges:

$$0 = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j}\,, \tag{2.32}$$

in other words: electrical charges are conserved. Any change in the electrical charge density has to accompanied with a corresponding current density.

Some history

At this point it is worthwhile to follow some historical detour. Before Maxwell accumulated the laws of electrodynamics in their differential form, the Oerstedt-Ampere law described the magnetic field of a stationary current density alone, without the term $\epsilon_0\,\partial \vec{E}/\partial t$,

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j}\,. \tag{2.33}$$

When examining this equation, Maxwell realised that something is wrong with it. Taking the divergence on both sides, he found that div $\vec{j} = 0$, i.e. that the electric current density has no sources. Hence, the total current floating out of a closed surface would be zero! This is clearly wrong, since it already had been established that the total current out of a closed surface is related to a change of charge inside the enclosed volume. This change could by no means be zero for all configurations, especially in view of the fact that currents are related with the transport of charges. Thus Maxwell, based on these theoretical considerations, **proposed** to add the term $\epsilon_0\,\partial \vec{E}/\partial t$ to this equation. At his time, it was unknown to argue on the base of abstract fields, and he made a model which was considering the vacuum to be similar to some elastic material. His contemporaries, however, were pretty sceptical, because of the model, because of the reasoning in terms of abstract fields, and because at the beginning there was no experimental justification whatsoever. This of course has changed by now, and this is why the equations are called after Maxwell who assembled them and extended them into a wonderful framework, the theory of electrodynamics.

Consequences of Maxwell's term

As a first example a spherical symmetric radial charge density will be considered; imagine to deal with something like a small sphere of radioactive material spitting out charged particles. In such a case there is a current of charges, directed radially to the outside. For simplicity it is assumed that the current is isotropic, in other words it has the same magnitude in all directions. When $Q(r)$ is the total charge inside a sphere of volume r, then the radial current density at this radius is $j(r)$. The two objects, charge and current, are connected through Eq.(2.32). Q has to decrease with the rate

$$\frac{\partial Q(r)}{\partial t} = -4\pi r^2\, j(r)\,. \tag{2.34}$$

Which magnetic field will be induced by the current? To answer this, imagine a closed loop L on the surface of a sphere with radius r. Obviously, there is a fair amount of the current j passing through the area enclosed by the loop, and it is quite convincing to expect a magnetic field circulating in the proper (right-handed) direction.

But this already means trouble. Drawing any other closed loop L' results in another direction of $\vec{B}$. Or, phrased differently: Since the current density is spherically symmetric, how can any specific direction of the magnetic field be expected on this sphere? How can there be any circulation of $\vec{B}$ around the current?

This is where Maxwell's additional term comes to a rescue. Let us check this. The electric field strength at radius r equals

$$E(r) = \frac{Q(r)}{4\pi\epsilon_0 r^2} \tag{2.35}$$

as will be seen explicitly later on. Therefore, the derivative of E w.r.t. time yields

$$\frac{\partial E(r)}{\partial t} = \frac{1}{4\pi\epsilon_0 r^2}\frac{\partial Q(r)}{\partial t} = -\frac{j}{\epsilon_0}\,. \tag{2.36}$$

Hence, in Eq.(2.30) both terms compensate each other, leading to a vanishing magnetic field around such a configuration.
This is cute! By adding the new term, the resulting equation lives up to what could be expected. This is true when focusing on Eq.(2.30) alone. It is only in connection with the other equations that something new emerges.

2.3 A propagating field

Consider a huge (infinite) charged plane in the y-z-plane at $x = 0$ and assume that at $t = 0$ it was instantaneously brought to a constant velocity u in the y direction. The result is a current $\vec{J}$ with density J per unit along the z-axis. What will happen?
Having a current in, say, positive y-direction translates into a magnetic field around it. For positive $x > 0$ the magnetic field will point into negative z-direction and for negative $x < 0$ it will point into positive z-direction. The value of $\vec{B}$ can be obtained by equating the line integral with the current, divided by $\epsilon_0 c^2$. The result reads

$$B = |\vec{B}| = \frac{|\vec{J}|}{2\epsilon_0 c^2} \tag{2.37}$$

since per unit length in z-direction there is one unit of $\vec{J}$ but two pieces of the line integral around it, one at positive, and one at negative values of x. This is the field close to the plane, for small values of x. The same reasoning results in the same field also for larger values of x. Immediately with switching on the current there would be a finite magnetic field accompanying it. But any change in the magnetic field induces an electric field, doesn't it? By moving the charged plane, a changing magnetic field emerges, which in turn gives rise to a changing electric field, these changes induce a derivative w.r.t. time, $\partial\vec{E}/\partial t$, which together with J influences the magnetic field. The two fields are therefore deeply connected and one has to find a simultaneous solution for both.
From the Maxwell equations alone it is not easy to see how to construct such a solution. Therefore, let's first start with a description of the solution on

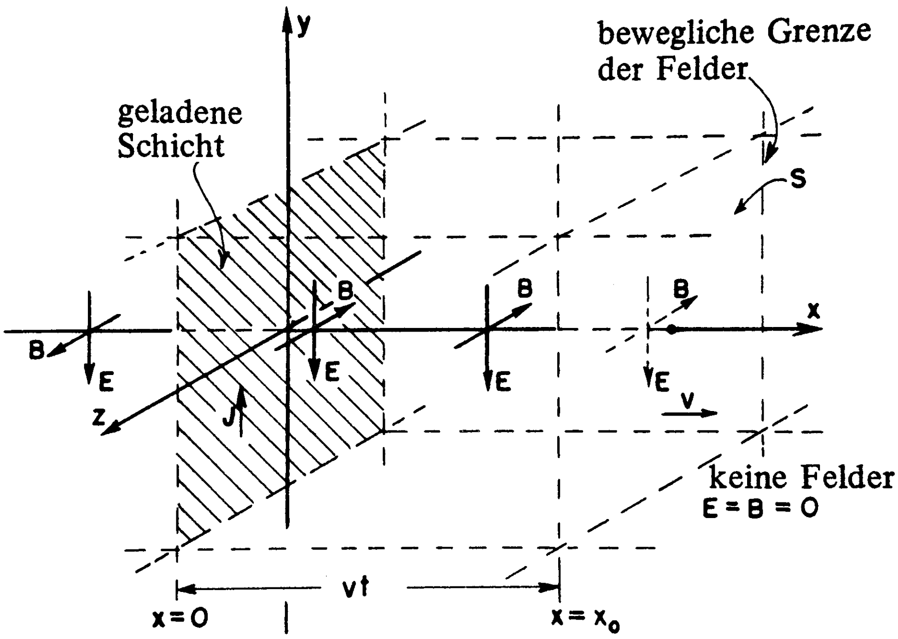

Figure 2.1: How the fields expand.

physical terms before checking its validity. In fact, the answer is as follows: With switching on the current, the magnetic field B calculated before will emerge close to the plane. This is, strictly speaking, true only for distances from the plane which are small enough such that no electric current fits in between, i.e. for distance basically zero. For larger values of x the magnetic field strength B is zero at the beginning. After remaining zero for some time it suddenly starts rising before reaching a finite value. This rising B field expands out of the region in the vicinity of the current. After some time, there is a homogenous B field up to some distance x, for larger distances the magnetic field strength is still zero. Of course this reasoning holds true in both directions, for positive and for negative x.
The electric field does the same. At $t = 0$ the E field is zero everywhere. After some time t both the electric and the magnetic field are homogenous up to a distance $x = vt$ and zero for larger distances. The fields expand like a wave with constant velocity v, see also Fig. 2.1 [2]. It should be stressed here

[2] In fact, this velocity will turn out to be c, the speed of light; for the moment, one may stick to the idea of the fields expanding with some velocity v to be determined later on.

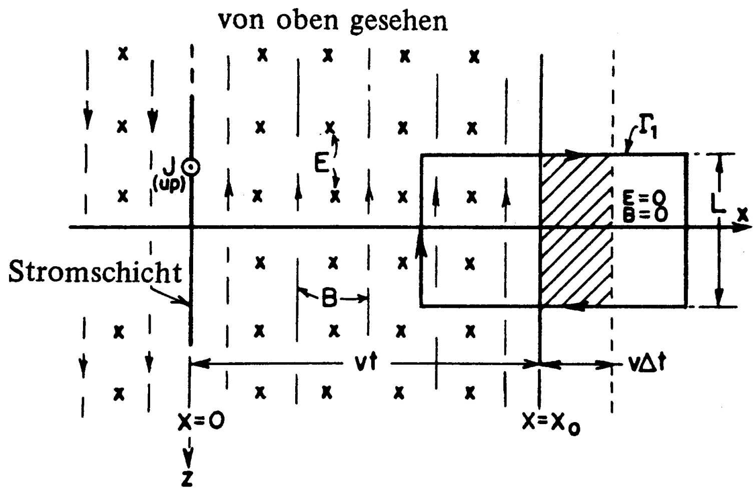

Figure 2.2: The closed loop L_2, connecting $\vec{\nabla} \times \vec{B}$ with $\partial \vec{E}/\partial t$.

that according to our assumption of an infinitely large plane the two fields are stretching out infinitely into both directions, y and z.

2.3.1 The speed of light

Let's dwell a little longer on this example and let's try to be quantitative now. To do so, consider two closed loops: one, dubbed L_1, arranged in the x-z-plane such that it stretches over the boundary $x = vt$ up to which the fields managed to expand so far, the other one, dubbed L_2, in contrast is arranged in the x-y plane, again such that it stretches out over x.
First of all, one may check whether the field configurations are consistent with the Maxwell equations. To do so, consider L_2 first and use

$$\oint\limits_{\delta S} \mathrm{d}\vec{l} \cdot \vec{E} + \int\limits_{S} \mathrm{d}^2\vec{s} \cdot \frac{\partial \vec{B}}{\partial t} = 0 \tag{2.38}$$

In L_2, cf. Fig. 2.2, the $\vec{E}$ field is oriented in the direction of negative y for all distances between $x = \pm vt$ from the moving charged plane. The electric field therefore is in parallel with L_2. The magnetic field, in contrast, points into direction perpendicular to L_2. The line integral of the electric field around

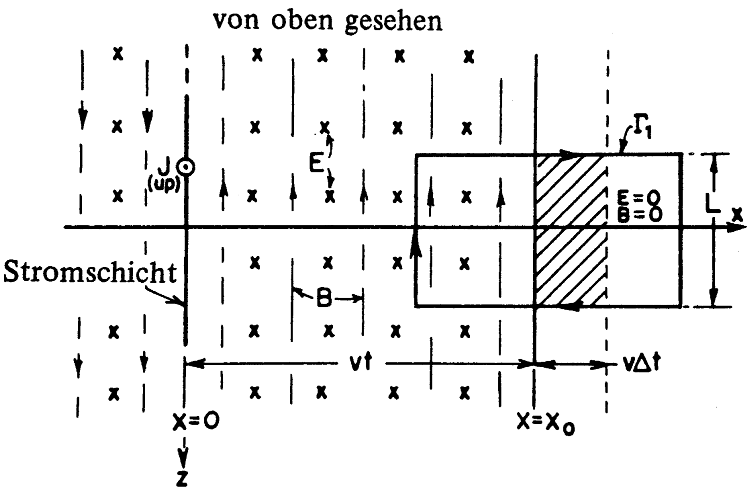

Figure 2.3: The closed loop L_1, connecting $\vec{\nabla} \times \vec{E}$ with $\partial \vec{B}/\partial t$.

L_2 is constant. It is just the part proportional to one of its edges, $EL_{2,y}$, until the full loop is inside the region of non-vanishing fields. The B-field, however, changes with time, since the magnetic flux through the loop changes with time. This is because the fields expand. The rate of change is given by $BvL_{2,y}$. Equating both yields

$$E = vB\,. \tag{2.39}$$

But Eq.(2.38) is not the only of Maxwell's equations that connect temporal changes of fields with the other field. In addition,

$$c^2 \oint_{\delta S} \mathrm{d}\vec{l} \cdot \vec{B} - \int_S \mathrm{d}^2\vec{s} \cdot \frac{\partial \vec{E}}{\partial t} = \frac{1}{\epsilon_0} \int_S \mathrm{d}^2\vec{s} \cdot \vec{j} \tag{2.40}$$

is valid. Arranging L_1 such that it does not enclose any piece of the moving charged plane, the r.h.s. of the equation above vanishes, see Fig. 2.3. Similar reasoning as for the loop L_2 yields as line integral of the magnetic field around L_1 the value $BL_{1,x}$. The rate of change of the electric flux through the loop is $EvL_{1,x}$. Plugging these findings into Eq.(2.40) yields

$$c^2 B = vE\,. \tag{2.41}$$

Therefore, there is a solution for the two fields behind the wave front. It is characterised by a constant $\vec{E}$ and a constant $\vec{B}$ field, which are orthogonal w.r.t. each other and are both perpendicular to the direction into which the wave front travels. From the two constraints it also becomes clear that the wave front **must** travel with velocity $v = c$, with the speed of light.
What happens when the moving charged plane is stopped after some time T? The answer can be seen with help of the superposition principle, which will be discussed im more detail in a later section. The solution for the opposite case, i.e. for switching on the current, is known. A net current of zero can be reached when taking another charged plane and moving it into the opposite direction with the same speed u. Then, after some time T, when the second plane starts moving, the net current equals zero. The new, added current, produces the same fields as the original one, just oriented in the opposite direction. With some delay, namely $\Delta t = T$ the new fields will arrive and add to the original fields. Due to their opposite direction, the net resulting field strengths will be zero. In other words, moving a system of charges with uniform speed u for some time results in a "block" of electric and magnetic currents which travel with the speed of light. Their respective field strengths are related through $E = cB$. Inside this block, the fields are perpendicular w.r.t. each other and both are orthogonal to the direction of expansion. The fields have "taken off", they propagate freely. It is the combined effect of the two free Maxwell equations $c^2\text{rot}\,\vec{B} = \partial\vec{E}/\partial t$ and rot $\vec{E} = -\partial\vec{B}/\partial t$ that keeps them up and running. Maxwell's proposed term ultimately produced a theory of light!

2.3.2 First solutions: Potentials and wave equation

Deriving the wave equation

First of all, Eq.(2.20) is employed to rewrite

$$\vec{\nabla} \cdot \vec{E} = \vec{\nabla} \cdot \left(-\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \right) = -\nabla^2\phi - \frac{\partial}{\partial t}\vec{\nabla} \cdot \vec{A} = \frac{\rho}{\epsilon_0}\,. \tag{2.42}$$

In a similar fashion, Eqs.(2.18) and (2.20) may be used to identify

$$
\begin{aligned}
c^2\vec{\nabla}\times\vec{B}-\frac{\partial\vec{E}}{\partial t} &= c^2\vec{\nabla}\times\vec{\nabla}\times\vec{A}-\frac{\partial}{\partial t}\left(-\vec{\nabla}\phi-\frac{\partial\vec{A}}{\partial t}\right)\\
&= -c^2\nabla^2\vec{A}+c^2\vec{\nabla}\cdot(\vec{\nabla}\cdot\vec{A})+\frac{\partial}{\partial t}\vec{\nabla}\phi+\frac{\partial^2\vec{A}}{\partial^2 t}=\frac{\vec{j}}{\epsilon_0}\,.
\end{aligned}
\tag{2.43}
$$

From Eq.(2.26) it is known that the divergence of the vector potential can be chosen at will. To simplify things, the Lorentz gauge

$$
\vec{\nabla}\cdot\vec{A}=-\frac{1}{c^2}\frac{\partial\phi}{\partial t}
\tag{2.44}
$$

is chosen. Plugging this into the equation above, the two middle terms cancel and the two equations for the potentials read

$$
\begin{aligned}
\nabla^2\phi-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} &= -\frac{\rho}{\epsilon_0}\\
\nabla^2\vec{A}-\frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} &= -\frac{\vec{j}}{\epsilon_0 c^2}\,.
\end{aligned}
\tag{2.45}
$$

What a beautiful system! It describes waves, expanding with c, the speed of light, originating from and coupling to electric charges and currents. Writing out the Laplace operator, e.g. for the scalar potential,

$$
\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}=-\frac{\rho}{\epsilon_0}
\tag{2.46}
$$

the nice symmetry between the spatial and temporal coordinates comes into full visibility. Of course, the term $1/c^2$ remains; it is related to space and time being different. The minus-sign is a quiet reminder of Minkowski's metric, the geometric structure of space-time.

Free solution

For a first solution, consider the free case, i.e. where the sources of the wave, charges and currents, have long vanished or a far away. Then, for the scalar potential the wave equation reads

$$
\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}=0\,.
\tag{2.47}
$$

It is solved by the ansatz

$$\phi(t,\, \vec{x}) = \phi_0 e^{\pm i(\omega t + \vec{k}\cdot\vec{x})} \tag{2.48}$$

yielding the relation

$$k_x^2 + k_y^2 + k_z^2 - \frac{\omega^2}{c^2} = k^2 - \frac{\omega^2}{c^2} = 0\,. \tag{2.49}$$

Similarly, free solutions for $\vec{A}$ read

$$\vec{A}(t,\, \vec{x}) = \vec{A}_0 e^{\pm i(\omega t + \vec{k}\cdot\vec{x})}\,. \tag{2.50}$$

Having constructed a solution for the waves it is neccessary to keep in mind that full solutions emerge as superpositions of these solutions. Due to the structure of the wave equation, it is possible to construct solutions that remain invariant when propagating. In general, a solution for the wave equation in one dimension,

$$\frac{\partial^2\phi(x,\, t)}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\phi(x,\, t)}{\partial t^2} = 0\,, \tag{2.51}$$

looks like

$$\phi(x,\, t) = f(t - x/c) + \bar{f}(t + x/c)\,. \tag{2.52}$$

This is a superposition of functions of the only variable $(x - ct)$. Hence, the function $f(x - ct)$ for instance describes a "pattern" in x moving with velocity c in positive x-direction. If, for instance, f has a maximum where its argument vanishes, this maximum is at $x = 0$ for $t = 0$ and at $x = cT$ for $t = T$. The correctness of the combined solution is easy to check. Just consider the first and second derivative of, say, f w.r.t. either t or x:

$$\begin{aligned}
\frac{\partial f(t-x/c)}{\partial t} &= \frac{\partial f(z=t-x/c)}{\partial z}\frac{\partial(z=t-x/c)}{\partial t} = f' \\
\frac{\partial^2 f(t-x/c)}{\partial t^2} &= \frac{\partial f'(z)}{\partial z}\frac{\partial(z=t-x/c)}{\partial t} = f'' \\
\frac{\partial f(t-x/c)}{\partial x} &= \frac{\partial f(z=t-x/c)}{\partial z}\frac{\partial(z=t-x/c)}{\partial x} = -\frac{1}{c}\,f' \\
\frac{\partial^2 f(t-x/c)}{\partial x^2} &= -\frac{1}{c}\frac{\partial f'(z)}{\partial z}\frac{\partial(z=t-x/c)}{\partial x} = \frac{1}{c^2}\,f''\,.
\end{aligned} \tag{2.53}$$

Similar reasoning holds true for $\bar{f}$; the sign in front of x/c in the argument does not matter, since only quadratic derivatives are of importance.

Chapter 3

Electrostatics

In this chapter some problems of electrostatics will be discussed. Static problems are characterised by the fact that there is no time-dependence, hence all derivatives w.r.t. time disappear. In such cases the Maxwell equations separate into two subsets:

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{\rho}{\epsilon_0} \\ \vec{\nabla} \times \vec{E} &= 0\,, \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \cdot \vec{B} &= \frac{\vec{j}}{\epsilon_0 c^2}\,. \end{aligned} \tag{3.2}$$

These two subsets describe electrostatics and magnetostatics, respectively. As it turns out, the electric field then has no curls - it is a pure source-field. In contrast, the magnetic field has no sources, it is a pure curl-field. In the following, electrostatics will be studied in more detail, a discussion of magnetostatics is left for the next chapter. Situations, where all charges are fixed at some positions will be the starting point. For such situations, the Coulomb law, supplemented with some wizardry in integrations will suffice to come up with solutions. In most cases, however, and these are the ones that are really interesting, the respective position of the charges and the correct form of the electric field are intimately tied, rendering solutions more complicated.

3.1 Coulomb's law and electrostatic potential

3.1.1 Principle of superposition

It would seem logical to start with Eqs.(3.1) and just use them. Nevertheless, it may be simpler to start with an equation that is better known and then reanalyse things. The results will be fully equivalent, hopefully convincing reason to accept Eqs.(3.1) as correct. Let's start with a law that is already known to be correct, Coulomb's law. It states that the force between two charges is proportional to each of them (and, hence, proportional to their product) and proportional to the inverse square of their distance. The force is directed along the line connecting the two charges and it is attractive (repulsive) if both charges have different (same) sign. Denoting the force acting on charge i as $\vec{F}_i$ the Coulomb law reads

$$\vec{F}_1 = -\vec{F}_2 = \frac{q_1 q_2}{4\pi\epsilon_0} \frac{\vec{r}_{12}}{r_{12}^3}\,, \tag{3.3}$$

where the q_i are the charge and $\vec{r}_{12}$ is the vector between them.
The interesting question is: What happens if more than two charges are present? In this case, Coulomb's law has to be supplemented with another law, the law of superposition. It states that the force on any charge is given as the vector sum of all Coulomb forces between it and any other charge. This is, basically, the full point about electrostatics; the difficulty in it is not to state the principle but to come up with solutions. When looking more closely at Eqs.(3.1), it will turn out that they just state Coulomb's law and the principle of superposition.

Electric field and Coulomb's law

When applying Coulomb's law it is convenient to introduce the electric field. The (directed) field strength at, say, position (1) is then given by the force per unit charge acting on q_1, i.e. it is

$$\vec{E}_1 = \frac{\vec{F}_1}{q_1} = \frac{q_2}{4\pi\epsilon_0} \frac{\vec{r}_{12}}{r_{12}^3}\,. \tag{3.4}$$

This may be interpreted as some information concerning (1), even if q_1 is not present at all. Of course, $\vec{E}_1$ is a vector with three components, where, e.g., the x-component reads

$$E_{1,x} = \frac{q_2}{4\pi\epsilon_0} \frac{x_1 - x_2}{[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2]^{3/2}} \,. \tag{3.5}$$

In the presence of more charges, the field strength at some point (1) is given as the sum of contributions from all charges,

$$\vec{E}_1 = \sum_j \frac{q_j}{4\pi\epsilon_0} \frac{\vec{r}_{1j}}{r_{1j}^3} \,. \tag{3.6}$$

If the charges are distributed continously it is simple to imagine a charge density ρ, a scalar function. It is defined through the limiting process

$$\rho(\vec{r}_2) = \lim_{\Delta V_2 \to 0} \frac{\Delta q_2}{\Delta V_2} \,, \tag{3.7}$$

where the charge Δq_2 represents the charge at point $(2) = \vec{r}_2$ and ΔV_2 is a small volume around this point. Using charge densities, in which discrete charges may be represented by δ-functions, the electric field strength of Eq.(3.6) becomes

$$\vec{E}_1 = \sum_j \frac{q_j}{4\pi\epsilon_0} \frac{\vec{r}_{1j}}{r_{1j}^3} \Longrightarrow \frac{1}{4\pi\epsilon_0} \int \mathrm{d}^3 r_2 \frac{\rho(\vec{r}_2)(\vec{r}_1 - \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|^3} \,. \tag{3.8}$$

This equation yields the field strength in situations when all charges are fixed. So, if the question is: What are the electric fields with given charges at given positions? Then the answer is: Calculate this integral! Using a computer, this integration is simple and boring. Therefore this will not be discussed in two much detail. It is important to note that in many cases discussed later on, the answer is already there. Sometimes, however, and this will be in the focus of forthcoming discussions, it is much more fun and much more instructive to gain the answer from physical reasoning rather than from tedious, but straightforward calculations.

3.1.2 (Re-)construction of the potential

First of all, the idea of having a potential connected to the field will be dwelled on. The potential relates to the work needed to move a (test-) charge from one point to another when other charges are present. This is in full analogy to the gravitational potential. The work that has to be delivered **against** the other charges when moving the test charge along a trajectory L is given by the line integral of the **negative** parallel component of the force acting on it:

$$W^q_{ab} = -\int_a^b \mathrm{d}\vec{l} \cdot \vec{F} \,. \tag{3.9}$$

Since the force acting on a unit charge is identical to the field strength this can be rewritten as

$$W^{\mathrm{unit}}_{ab} = -\int_a^b \mathrm{d}\vec{l} \cdot \vec{E} \,. \tag{3.10}$$

At this point it seems that the work depends on the actual trajectory. Therefore, by choosing one trajectory for $a \to b$ and another one for the way back $b \to a$ with $W_{ba} < W_{ab}$ it seems that it may be possible to gain more work on the way back than what had been invested before. This, however, is impossible. Of course there are configurations where the moving test charge applies forces on the other charges, such that they move and loose energy; in electrostatics such a scenario is ruled out. This is because in electrostatics all charges are fixed, therefore the moving test charge does not act on the other charges such that any amount of work is released from them. Consequently, assuming the conservation of energy to be valid (there is no reason to doubt it here!) the test charge cannot absorb any energy along a closed path. In the following, using Coulomb's law this will be proven.
Let's check first, what happens in a field produced by a single charge q. Obviously, the $\vec{E}$ field points radially towards the charge, with a field strength proportional to the inverse square of the distance to the charge. Consider the first path, cf.Fig.**??**. The test charge would first be moved on a circular arc around the charge ($a \to a'$, with $r_a = r_{a'}$). On this path no work has to be done, since the field and the respective line element are orthogonal w.r.t. each other. On the second piece of the path, the movement is in radial direction, such that the line element and the field strength are parallel. Hence the total

work along this path is

$$\begin{aligned} W^{\text{unit}}_{ab} &= W^{\text{unit}}_{aa'} + W^{\text{unit}}_{a'b} = -\int\limits_{a'}^{b} \mathrm{d}\vec{l} \cdot \vec{E} \\ &= -\frac{q}{4\pi\epsilon_0} \int\limits_{a'}^{b} \frac{\mathrm{d}r}{r^2} = -\frac{q}{4\pi\epsilon_0} \left(\frac{1}{r_a} - \frac{1}{r_b} \right) . \end{aligned} \tag{3.11}$$

Any other path can be thought to be composed of, possibly infinitesimal, pieces that are either radial or along a circular arc. The former will yield contributions, which finally add up to the result above, whereas the pieces along the arcs will not contribute at all. Thus, the result of Eq.(3.11) remains independent on the particular path chosen. This result depends on the endpoints of the path, it can be understood as a difference of two numbers

$$W^{\text{unit}}_{ab} = -\frac{q}{4\pi\epsilon_0} \left(\frac{1}{r_a} - \frac{1}{r_b} \right) = \phi(b) - \phi(a) . \tag{3.12}$$

Since this is a difference, it is independent of some common offset, which could be fixed by the – irrelevant – choice of a reference value of ϕ taken at the point $\vec{r}_0$. Having fixed this point, the scalar function ϕ is fully specified. It is called the electrostatic potential, defined as

$$\phi(\vec{r}) = -\int\limits_{\vec{r}_0}^{\vec{r}} \mathrm{d}\vec{r'} \cdot \vec{E} , \tag{3.13}$$

in full analogy with the gravitational potential already known from Newton's mechanics. For simplicity $\vec{r}_0$ is often chosen in some infinite distance, yielding

$$\phi(\vec{r}) = \frac{q}{4\pi\epsilon_0} \frac{1}{|\vec{r}|} = \frac{q}{4\pi\epsilon_0} \frac{1}{r} . \tag{3.14}$$

Again, the principle of superposition applies for the potential, if more than one charge is present. In particular, for any charge distribution,

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\mathrm{d}^3 r' \rho(\vec{r'})}{|\vec{r} - \vec{r'}|} . \tag{3.15}$$

It should be stressed that the potential in fact has a physical meaning: It equals the potential energy of an unit charge which has been brought from the reference point to $\vec{r}$.

$\vec{E} = \text{grad}\ \phi$ and direct consequences

Well, what can be done with this potential, apart from its nice physical interpretation? The answer is that it is quite simple to obtain the electric field out of the potential by just differentiating it.

$$\vec{E} = \text{grad}\ \phi \tag{3.16}$$

is the differential form of the definition of the potential, Eq.(3.13). This equivalence is easily understood by recalling Eq.(1.62),

$$\phi(\vec{r}) - \phi(\vec{r}_0) = \int\limits_{(\vec{r}_0)}^{(\vec{r})} \vec{\nabla}\phi \cdot \mathrm{d}\vec{r'}\,.$$

But is it any advantage to use ϕ for calculational purposes rather than $\vec{E}$? The answer is yes: To calculate $\vec{E}$, three integrals for the three components must be evaluated, whereas ϕ, being a scalar function, necessitates only one integral to be done. Apart from merely numbers of integration it is worth to note that $1/r$ is simpler to integrate than x/r^3 or similar.
Is there anything else about this formulation of the electric field $\vec{E}$ being the gradient of a scalar potential? Again, the answer is yes. Just remember that rot grad $\phi = 0$, cf.Eq.(1.49). It follows immediately that

$$\text{rot}\ \vec{E} = \vec{\nabla} \times \vec{E} = 0\,. \tag{3.17}$$

This shows that the electric field stemming from Coulomb's equation satisfies the condition of being curl-free. In fact, this has been known before, since it had been shown that the work done along a closed path vanishes,

$$\oint \vec{E} \cdot \vec{s} = 0 \Longleftrightarrow \vec{\nabla} \times \vec{E} = 0\,. \tag{3.18}$$

3.1.3 Flux of $\vec{E}$ and Gauss' law

Recovering Gauss' law for a point charge

Now it is time to check that, at least for point charges, Gauss' law is correct and that it is intimately tied to the fact that the Coulomb- force behaves like $1/r^2$. This $1/r^2$-dependence seems so natural, doesn't it? This is just

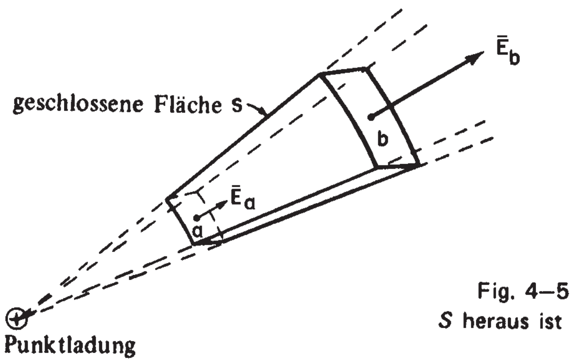

Figure 3.1: Simple surface close to a point charge.

the way, things like light expand. Just take a source of light; the amount of light floating through a cone with apex at the source is independent of the radius at which the cone is taken. Of course, this **must** be the case as long as the energy is conserved. Since the area of the cone grows with the square of the radius, the intensity, i.e. the amount of light per unit area, must decrease with the square of the radius to conserve the overall amount of light inside the cone. However appealing, it is not clear how this reasoning should apply to the electric field strength as well, unless it is accepted that field strength and intensity of light are proportional. If the electric field would behave like a continuous flux of little balls or so, where the number of balls would be conserved, then such a reasoning would apply. In other words: If the electric field would be connected to a current with a corresponding conserved quantity, like, e.g. the heat flow $\vec{h}$ discussed before, then $\vec{E}$ would behave like $1/r^2$ near a point source. In the following this behaviour of $\vec{E}$ similar to a "current of something" will be investigated and it will become clear how the law of the inverse square of the distance is connected with it. To this end, let's try to calculate the flux of $\vec{E}$ out of an arbitrary closed surface in the neighbourhood of the point charge responsible for it. As a starting point, consider a simple surface, like the one depicted in Fig.3.1. If $\vec{E}$ behaves similar to a current, then the total out of this surface is zero. On the radial side surfaces, the normal component of $\vec{E}$ and hence the flux is zero: $\vec{E}$ has radial direction, and the normal vectors of the area are orthogonal to it. Therefore, there remain only the two non-radial pieces a and b. Well, their surface increases quadratically with the distance from the point source, but the field strength decreases with the same rate. The overall product is independent of the distance, yielding identical influx and efflux of $\vec{E}$. This difference in sign ensures that both cancel; taken together

$$\oint_S \mathrm{d}^2\vec{s} \cdot \vec{E} = 0\,. \tag{3.19}$$

The two non-radial end-pieces can be slanted with respect to the radial direction. This can be seen directly by taking infinitesimally small areas. There, the size of $\vec{E}$ does not change significantly and can thus be approximated by its value on the centre of the area. The size of the area increases with $1/\cos\theta$, where θ is the angle under which the area has been slanted. The normal component of $\vec{E}$ however will be multiplied with $\cos\theta$, the net effect being that the product of normal component and area remains the same. If this is true for any small surface, any larger surface can be composed out of the small ones. In particular, this proves that the flux of $\vec{E}$ out of any closed surface is zero.

But wait! This is true only for those surfaces S which do not enclose the charge itself. What happens if the charge is inside such a surface? In this case, one may subdivide the surface in pairs of surface which are connected by radial lines going through the point charge. The flux of of these pairs does not vanish any longer, since now the orientation of the non-radial areas and $\vec{E}$ is the same, there is no cancellation, and therefore there is a net flux out of this pair of sub-surfaces. The question now is: how large is the flux through the full original surface? To gain an answer, let's employ a simple trick. First of all, one may cut out the charge of the volume enclosed by S, for instance by having a smaller surface S' close around the charge. This surface may have any form, in particular, one may choose a sphere. The flux out of the volume enclosed by S and S' is zero - there are no charges left in it. Therefore, the flux out of S equals the flux out of S'. If the radius of the sphere S' is r the value of $\vec{E}$ on its surface is

$$E = \frac{q}{4\pi\epsilon_0 r^2}\,, \tag{3.20}$$

when q is the charge inside. The direction of $\vec{E}$ is radial, and, hence, orthogonal to the surface. Therefore the overall flux through S' can be obtained by multiplying with the surface of the sphere, it yields

$$\text{Flux through S'} \; = \frac{q}{4\pi\epsilon_0 r^2} \cdot (4\pi r^2) = \frac{q}{\epsilon_0} \,. \tag{3.21}$$

This is independent on the radius of the sphere. Therefore, the flux out of S is also q/ϵ_0, independent on the form of S. In other words:

$$\int\limits_{\text{any surface}\, s} \mathrm{d}^2\vec{s} \cdot \vec{E} = \begin{cases} 0 & \text{for} \quad \text{q outside of S} \\ \dfrac{q}{\epsilon_0} & \text{for} \quad \text{q inside of S} \end{cases} \tag{3.22}$$

From the reasoning above it is clear that any other dependence of $\vec{E}$ on r would result in some residual dependence on the shape of S.

Gauss law: $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_0$

The result above can easily be generalised to a distribution of charges, the corresponding law is known as Gauss' law. It reads, for an arbitrary volume V with surface S

$$\int\limits_{S=\delta V} \mathrm{d}^2\vec{s} \cdot \vec{E} = \int\limits_{V} \mathrm{d}^3x \frac{\rho(x)}{\epsilon_0} \,. \tag{3.23}$$

Its differential form reads, according to theorem 1, Eq.(1.80),

$$\text{div}\, \vec{E} = \vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \,. \tag{3.24}$$

From the derivation above it is clear that this law is intimately connected with the fact that the exponent of r in Coulomb's law is -2. Any other exponent and any other dependence on r would result in a different law. Therefore, Gauss' law just expresses Coulomb's law, the two laws are equivalent.

3.2 First applications

3.2.1 Electric flux lines and equipotential surfaces

To gain a geometric description of the electrostatic field, the two fundamental laws (the flux is proportional to the enclosed charge and the electric field strength is the gradient of a potential) need to be represented through geometric concepts. Let's consider two illustrative examples, the field of a point charge and the field between two equal charges of different sign.

Point charge

Consider first the field of a point charge. First, draw some lines in the direction of the field; the field lines will be directed, by convention, they start at positive and end at negative charges. These lines will take off radially from the point charge, with suitable arrow indicating their direction. But, apart from their direction, it would be nice to visualise the field strength. This could be done by the density of lines. Now, check this out: Since the density of lines is proportional to the number of lines per surface orthogonal w.r.t. the lines, and since the surface of a sphere goes with the square of its radius, the density will decrease with the inverse square of the radius if the number of lines is kept constant. Therefore, it is sufficient to keep the number of lines constant in order to have this visualisation: The field strength is proportional to the density of lines, the number of lines stretching out from a charge being constant. This can be realized easily by demanding that the lines are produced at charges only and expand continously, without any splitting. If the number of lines emerging or ending at a charge is proportional to q/ϵ_0, contact to Gauss' law is established.
The potential can be best visualised through the equipotential surfaces. The geometric connection to the flux lines is clear: Since the electric field is related to the gradient of the potential, and since the gradient is perpendicular to equipotential surfaces, these surfaces can be constructed by demanding that they are orthogonal w.r.t. the flux lines constructed before. For an example of this, see Fig. 3.2.

Two point charges

As a first guess, one may argue that for two point charges the flux lines etc. can be gained through the superposition principle. Well, let's try by just taking two of the pictures emerging for a point charge and overlay them. The result is disappointing! The flux lines intersect, clearly not the pattern which is called for. In real life, the flux lines are not supposed to intersect, for the simple reason that at one point $\vec{E}$ must not have two directions. Although it is hard to accept, this is a case, where only calculation helps. A result is

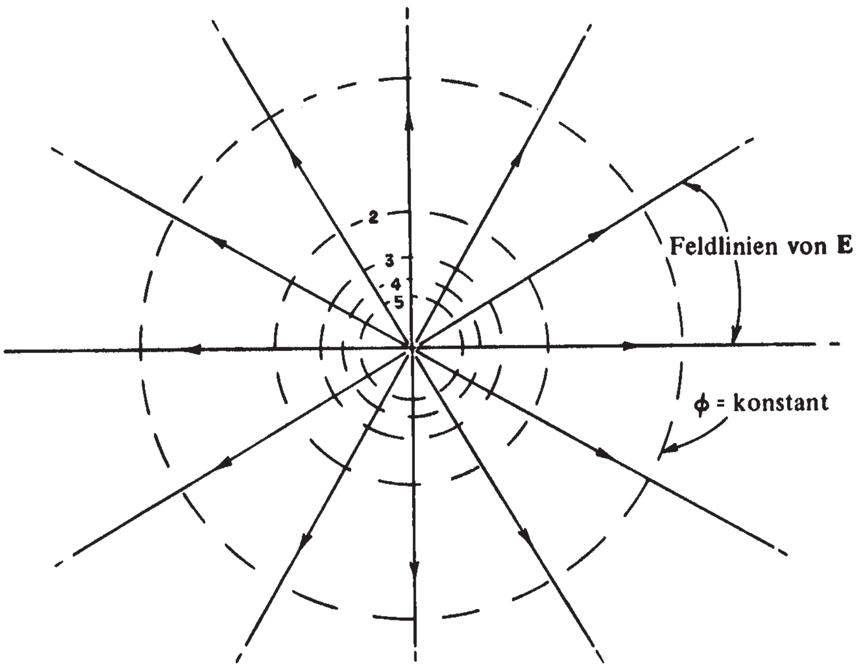

Figure 3.2: Visualisation of the electric field and potential of a point charge.

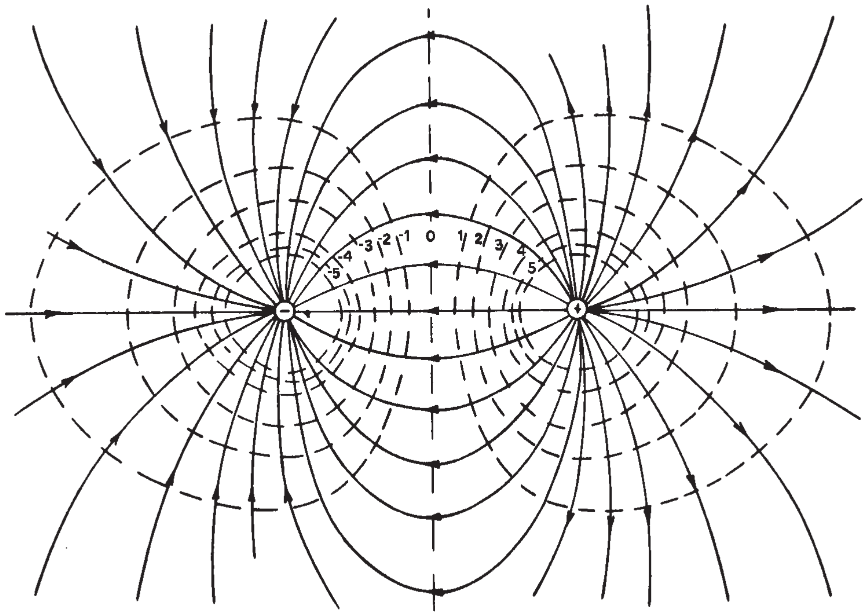

Figure 3.3: Visualisation of the electric field and potential of two point charges.

depicted in Fig. 3.3.

3.2.2 Fields of some charge configurations

Field of a charged sphere

One of the complicated problems in the theory of gravity was to prove that a solid sphere of material generates the same force on its surface as if all mass was concentrated in its center. Newton did not publish his theory for a long time since he did not know whether this theorem is correct. Now it's time to prove it in a very simple and straightforward way. But, instead of doing the calculation for gravitation it will be done for the electric field. However, since the laws of electrostatics, i.e. Coulomb's law, are so similar to those of gravity, the same reasoning holds true for both.
In particular, the task is to answer the following question: What is the electric field strength at any point $\vec{r}_P$ outside of a homogeneously charged sphere with

radius R? Since there is no preferred direction apart from pointing radially away from the sphere, it is clear that $\vec{E}$ points away from its centre. Consider a concentric sphere going through $\vec{r}_P$; for this surface S the flux of $\vec{E}$ reads

$$\oint_S \mathrm{d}^2\vec{s}\cdot\vec{E} = \oint_S \mathrm{d}^2 s E = 4\pi r_P^2\, E = \frac{Q}{\epsilon_0}\,, \tag{3.25}$$

where Q is the total charge of the sphere. Thus, if $r > R$,

$$\vec{E}(\vec{r}) = \frac{Q}{4\pi\epsilon_0}\frac{\vec{r}}{r^3} = \frac{Q}{4\pi\epsilon_0}\frac{\vec{e}_r}{r^2}\,. \tag{3.26}$$

This is exactly what could have been expected from a point charge located at the centre of the sphere with charge Q.
What happens inside this sphere? Well, consider any point $\vec{r}_P$ inside, i.e. with $r_P < R$. Again, for the surface of the corresponding concentric sphere inside the charged ball, the electric field points in radial direction and its flux through S reads

$$\oint_{S=\delta V} \mathrm{d}^2\vec{s}\cdot\vec{E} = 4\pi r_P^2\, E = \int_V \mathrm{d}^3 x\frac{\rho_0}{\epsilon_0} = \frac{4\pi r_P^3\rho_0}{3\epsilon_0}\,, \tag{3.27}$$

where the constant charge density is given by

$$\rho_0 = \frac{Q}{V_{\text{sphere}}} = \frac{3Q}{4\pi R^3}\,. \tag{3.28}$$

Inserting this yields for the electric field for $r < R$

$$\vec{E}(\vec{r}) = \frac{Q}{4\pi\epsilon_0}\frac{r^3}{R^3}\frac{\vec{e}_r}{r^2} = \frac{Q}{4\pi\epsilon_0}\frac{r\vec{e}_r}{R^3}\,. \tag{3.29}$$

Field of a charged line

As next example, consider a system with cylindric instead of spherical symmetry, like e.g. a very long, homogeneously charged rod. This can be approximated through a charged, infinitely long line. Due to the cylindric symmetry, the $\vec{E}$ field points radially away from the line. Therefore it is oriented perpendicular to a cylinder with radius R with the charged line as axis. Then

$$2\pi R E = \frac{\lambda}{\epsilon_0}\,, \tag{3.30}$$

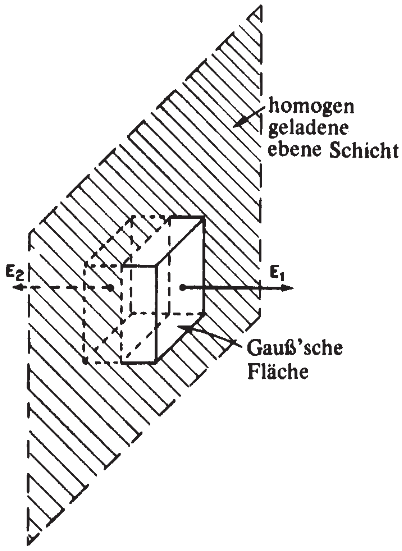

Figure 3.4: Homogeneously charged layer and the corresponding Gauss' surface.

where λ is the charge per unit length along the line. Therefore,

$$\vec{E}(\rho) = \frac{\lambda \vec{e}_\rho}{2\pi\epsilon_0\rho}\,, \tag{3.31}$$

with ρ the corresponding radial coordinate of the cylinder.

Field of charged layers

The last example is the field of a homogeneously charged, infinitely large plane. Assume that its charge per unit area is σ. Then, if there is no other charge, the fields on bot sides of the plane must be equal in size but different in direction: pointing orthogonally away from the plane. By considering a box with area A parallel to the plane, cutting out some section of the plane, cf. Fig. 3.4, the total flux out of the box is given by the flux through the pieces parallel to the plane. Therefore,

$$2|\vec{E}|A = 2EA = \frac{\sigma A}{\epsilon_0}\,, \tag{3.32}$$

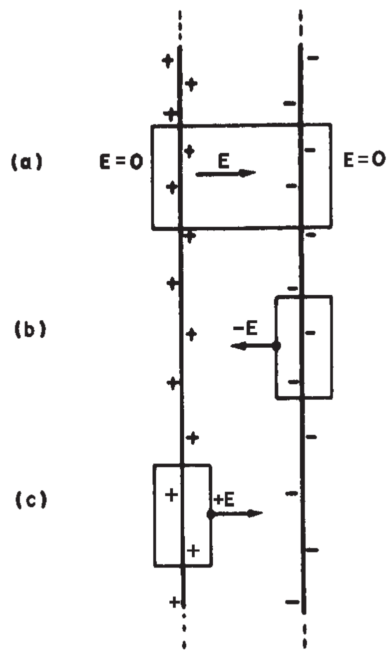

Figure 3.5: Two homogeneously charged layers and the corresponding Gauss' surfaces.

and

$$E = \frac{\sigma}{2\epsilon_0} \,. \tag{3.33}$$

This is the field evoked through charges on the plane, when no other charges are around. If other charges are around, then the two fields E_1 and E_2 on both sides of the plane add up to yield

$$E_1 + E_2 = \frac{\sigma}{\epsilon_0} \,, \tag{3.34}$$

where the relative size of the two contributions depend on the external charges.

Consider now two planes such charged planes, where the charges have opposite sign, see Fig. 3.5 When using a box, similar to the one above, closing around both planes, one finds

$$E = 0 \quad \text{outside the planes}\,. \tag{3.35}$$

Then, by taking boxes around one plane only, the E field between the them can be found to be

$$E = \frac{\sigma}{\epsilon_0} \quad \text{between the planes}\,. \tag{3.36}$$

3.2.3 Field inside a conductor

An electric conductor is a solid body which includes many free electrons. They can freely move inside the material, but doe to electric forces they won't leave it. In a metal (the prototypical conductor material), there are so many electrons that any electric field makes many of them move. When there are no external energy sources, the resulting current will eventually discharge the sources/charge disparities that triggered it. In electrostatics there are no external sources, therefore the electrons keep moving until the electric field inside the medium vanishes everywhere. Of course, if there was an electric field, the electrons would carry on moving. Thus, the only electrostatic solution inside a conductor is to have no electric field inside it. Therefore, the potential inside a conductor must be constant (in order to guarantee a vanishing gradient), and conductors are volumes or areas of equipotential. In particular, the surface of a conductor must be an equipotential surface. Also, the electric field at the surface of a conductor must be perpendicular to it, if this was not the case, there would be a tangential component, moving the electrons inside.
But how is talking about a charged conductor possible at all, where are the charges, if there is no field? After all, the field is proportional to the charge density, which, therefore, must vanish with the field. The answer is that the charges are concentrated at the surface of the metal, usually in the first one or two atomic layers. For the purposes of this lecture it suffices to say that when extra charges are brought to a conductor, they will all sit on its surface. This allows to calculate the field strength at the surface. Consider a tiny box such that it is around the surface of the conductor. By the same reasoning applied to the charged planes before, the field can be evaluated to yield

$$E = \frac{\sigma}{\epsilon_0}\,, \tag{3.37}$$

where again σ is the charge density per unit area at the point in question.

3.3 Electrical dipole

From Maxwell's equations for electrostatics,

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \quad \text{and} \quad \vec{\nabla} \times \vec{E} = 0$$

it is known that the electric field can be understood as the gradient of a potential

$$\vec{E} = -\vec{\nabla}\phi\,,$$

where

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\mathrm{d}^3 r' \rho(\vec{r'})}{|\vec{r} - \vec{r'}|}\,.$$

Thus, it is determined through the Poisson equation,

$$\vec{\nabla} \cdot \vec{\nabla}\phi = \nabla^2 \phi = \frac{\rho}{\epsilon_0}\,. \tag{3.38}$$

In some sense, electrostatics is nothing else but a discussion of solutions of this equation.
First of all, let's consider those problems where ρ is known as a function of position. The simplest case not dealt with so far are two point charges $+q$ and $-q$ located in a distance d from each other: when the distance becomes small, this configuration is called a dipole. Orienting the two charges along the z-axis, the charge density can be written as

$$\rho(\vec{r}) = q\left[\delta^3\left(\vec{r} - \frac{d}{2}\vec{e}_z\right) - \delta^3\left(\vec{r} + \frac{d}{2}\vec{e}_z\right)\right]\,. \tag{3.39}$$

Thus the dipole potential reads

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\mathrm{d}^3 r' \rho(\vec{r'})}{|\vec{r} - \vec{r'}|} = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{\sqrt{\left(\vec{r} - \frac{d}{2}\vec{e}_z\right)^2}} - \frac{q}{\sqrt{\left(\vec{r} + \frac{d}{2}\vec{e}_z\right)^2}}\right]\,. \tag{3.40}$$

The $\vec{E}$ field can gained from the potential through differentiation.

Quite frequently, dipoles occur in the discussion of physical systems. As an illustrative example take "dipole antennas". Even though these antennas usually are more interesting for moving charges, the limit of fixed charges often helps to gain some first insight into the structure of the corresponding field. Maybe even more important are atomic dipoles. As already discussed before, electrons tend to move in material when electric fields are applied. In case of conducting materials, some electrons in such a case will move to the surface such that there is no field inside the material. In non-conducting materials, the electrons are not so mobile. There they are attracted and pulled back by the field of the atomic remnants. Nevertheless, the shift a bit. Even though such atoms remain neutral in an electric field, there is a small separation of positive and negative charges, which acts like a microscopic dipole when large distances (large meaning large compared to the typical scale of the shift) are considered. It should be noted that there are - electrically neutral - molecules such as water which exhibit this behaviour even when no field is present. Of course, the geometry is not exactly two point charges, after all it's two hydrogen and one oxygen atom, the former having a positive net charge and the latter having a net negative charge. Nevertheless, as will be seen later on, in large distances such details become less and less important.

3.3.1 Playing with dipoles

The dipole potential

Let's return to Eq.(3.40) and take the limit $d \to 0$ [1]. One has

$$\left(\vec{r} - \frac{d}{2}\vec{e}_z\right)^2 = r^2 - \vec{r}\cdot\vec{e}_z d + \frac{d^2}{4} \approx r^2\left(1 - \frac{\vec{r}\cdot\vec{e}_z d}{r^2}\right) , \tag{3.41}$$

leading to

$$\frac{1}{\sqrt{\left(\vec{r} - \frac{d}{2}\vec{e}_z\right)^2}} \approx \frac{1}{r}\left(1 - \frac{\vec{r}\cdot\vec{e}_z d}{r^2}\right)^{-1/2} \approx \frac{1}{r}\left(1 + \frac{1}{2}\frac{\vec{r}\cdot\vec{e}_z d}{r^2}\right) . \tag{3.42}$$

Similarly,

$$\frac{1}{\sqrt{\left(\vec{r} + \frac{d}{2}\vec{e}_z\right)^2}} \approx \frac{1}{r}\left(1 - \frac{1}{2}\frac{\vec{r}\cdot\vec{e}_z d}{r^2}\right) , \tag{3.43}$$

[1] Of course, for $d = 0$ both charges are on top of each other, exactly compensating each other.

such that the potential reads

$$\phi(\vec{r}) = \frac{q}{4\pi\epsilon_0}\frac{\vec{r}\cdot\vec{e}_z d}{r^3}\,. \tag{3.44}$$

This potential is proportional to the product of charges and distance, called the dipole moment $\vec{p}$. In the case considered here,

$$\vec{p} = qd\vec{e}_z\,, \tag{3.45}$$

and the potential of Eq.(3.40) finally becomes

$$\phi_{\text{dipole}}(\vec{r}) = \frac{1}{4\pi\epsilon_0}\frac{\vec{r}\cdot\vec{p}}{r^3}\,. \tag{3.46}$$

This specific form is called the dipole potential.
It is easy to obtain the electric field of a dipole through differentiation of the potential. Using the specific orientation above,

$$\begin{aligned} E_z &= -\frac{\partial\phi}{\partial z} = -\frac{p}{4\pi\epsilon_0}\frac{\partial}{\partial z}\frac{z}{r^3} = -\frac{p}{4\pi\epsilon_0}\left(\frac{1}{r^3}-\frac{3z^2}{r^5}\right)\\ E_{\{x,y\}} &= -\frac{\partial\phi}{\partial\{x,y\}} = -\frac{p}{4\pi\epsilon_0}\frac{\partial}{\partial\{x,y\}}\frac{z}{r^3} = \frac{p}{4\pi\epsilon_0}\frac{3z\{x,y\}}{r^5}\,. \end{aligned} \tag{3.47}$$

The total electric field strength is, of course, given by

$$E = \sqrt{E_x^2+E_y^2+E_z^2}\,. \tag{3.48}$$

It falls like the third power of the distance from the dipole.[2] Schematically it looks like the one depicted in Fig. 3.6.

[2]A comment is in order at this point: Often, using compact notation seems wonderfully elegant and extremely beautiful. For many practical applications and calculations, however, it is in many cases more practical to make a smart choice of coordinate system. Of course it is not particularly sensible to go through all the pain of using arbitrary coordinates and complicated expressions for angles etc., when a simple choice delivers simple results that can be cast into a form that is independent of the choice of coordinate system. For instance, calculating explicitly div $\vec{E}$ in a certain coordinate system, may seem little elegant, but in many cases it is a helpful to gain a first idea about what is going on.

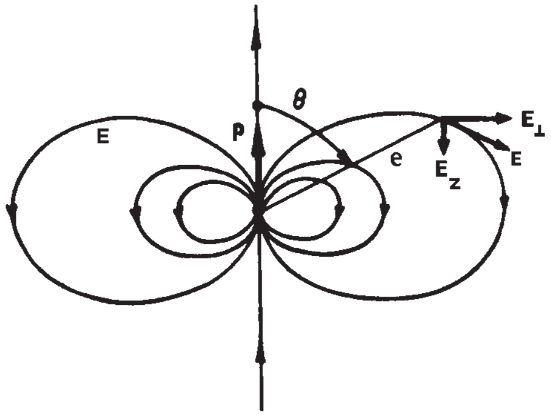

Figure 3.6: Schematic view of the field of a dipole.

The dipole potential as gradient

The dipole potential, Eq.(3.46) can also be written as

$$\phi_{\text{dipole}}(\vec{r}) = -\frac{1}{4\pi\epsilon_0}\vec{p}\cdot\vec{\nabla}\frac{1}{r}\,. \tag{3.49}$$

How can this be seen? That's quite simple! After all, it is well known by now that the field of a point charge is proportional to $\vec{e}_r/r^2 = \vec{r}/r^3$, and this field emerges as a gradient of $1/r$. For the dipole potential, this can be used. But apart from coincidence, there is another, physical reason for it. Omitting factors of $1/(4\pi\epsilon_0)$ in the following, the potential of a point source at distance r is

$$\phi_0 = \frac{q}{r}\,. \tag{3.50}$$

How does the potential change when taking a positive charge and shifting it by a distance Δz? This change is identical to the one when leaving the charge at its position, but shifting the distance by $-\Delta z$:

$$\phi_+ = \phi_0 - \frac{\partial\phi_0}{\partial z}\Delta z = \frac{q}{r} - \frac{\partial}{\partial z}\left(\frac{q}{r}\right)\Delta z\,. \tag{3.51}$$

The same reasoning can be applied to a negative charge being shifted in the opposite direction, the corresponding potential reads

$$\phi_- = \phi_0 + \frac{\partial \phi_0}{\partial z}\Delta z = \frac{-q}{r} + \frac{\partial}{\partial z}\left(\frac{-q}{r}\right)\Delta z\,. \tag{3.52}$$

Adding both yields

$$\phi = \phi_+ + \phi_- = -2\frac{\partial}{\partial z}\left(\frac{q}{r}\right)\Delta z - \frac{\partial}{\partial z}\left(\frac{q}{r}\right) d\,, \tag{3.53}$$

after identifying $d = 2\Delta z$. Writing this in an invariant form yields, after reinserting $1/(4\pi\epsilon_0)$, the potential

$$\phi = -\frac{1}{4\pi\epsilon_0}\vec{p}\cdot\vec{\nabla}\left(\frac{1}{r}\right)\,. \tag{3.54}$$

The lesson from this little exercise is that although the potential of any charge distribution can be obtained from straightforward integration, it is sometimes possible to save time by using the principle of superposition. Whenever there is a charge distribution that can be understood as the superposition of two charge distribution with a known potential, the potential of the resulting summed charge distribution is just the sum of the known potentials.

3.3.2 Dipole approximation

The electric field of dipoles plays an important role in many physical considerations. As an illustrative example take objects with a complicated charge distribution, like e.g. the water molecule. Although it is electrically neutral, it still has, due to its polarised structure, some remnant fields. In large distances, i.e. when the distance is much larger than the scale of the internal structure, such a field can then be approximated by a dipole field.
To understand this a bit better, consider a set of point charges q_i, which are distributed with distances $\vec{d}_i$ around the origin, which marks the "centre of gravity" of the charges. Then the potential at a point $\vec{R}$ is given by

$$\phi(\vec{R}) = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{\left|\vec{R}-\vec{d}_i\right|} \stackrel{R\gg d_i}{\longrightarrow} \frac{1}{4\pi\epsilon_0 R}\sum_i q_i = \frac{Q}{4\pi\epsilon_0 R}\,, \tag{3.55}$$

where Q is the total charge, see Fig. 3.7. This is not a big surprise: The

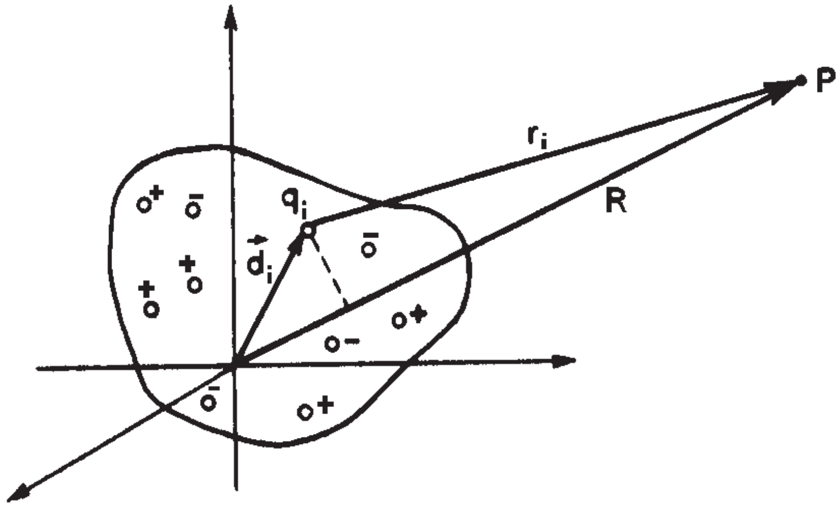

Figure 3.7: Towards dipole approximation.

potential far away from a charge distribution can be approximated by the potential of an equivalent point charge at the corresponding distance. But what happens if this approximation is not good enough, for instance if $Q = 0$? Then this approximation needs improvement, doesn't it? The only way to gain such an improved approximation is to improve the limit

$$\frac{1}{\left|\vec{R} - \vec{d}_i\right|} \overset{R \gg d_i}{\longrightarrow} \frac{1}{R} \tag{3.56}$$

by taking

$$\frac{1}{\left|\vec{R} - \vec{d}_i\right|} \overset{R \gg d_i}{\longrightarrow} \frac{1}{R}\left(1 + \frac{\vec{d}_i \cdot \vec{R}}{R^2}\right) \tag{3.57}$$

instead. Identifying

$$\vec{p} = \sum_i q_i \vec{d}_i \tag{3.58}$$

results in

$$\phi = \frac{1}{4\pi\epsilon_0}\left[\frac{Q}{R} + \frac{\vec{p}\cdot\vec{R}}{R^3}\right], \tag{3.59}$$

where the first term looks like the potential of an equivalent point charge (a monopole) and the second term looks like an equivalent dipole potential. In this framework, $\vec{p}$ is called the dipole moment of the distribution. In particular, when the overall charge of an object is zero, it is this equivalent dipole that defines the potential,and, hence, the electric field of the object. The water molecule, for instance, has quite a large dipole moment. It is responsible for the electric fields in water and hence for some of the important properties of water. In contrast, other molecules, such as carbon dioxide (CO_2), do not posses a strong dipole moment. In such cases the approximation above, Eq.(3.57), must be evolved further, to even higher terms. The next term of such a multipole evolution would be the quadrupole term, vanishing like $1/R^3$ with distance R. [3]

3.4 Methods

From all investigations so far, the impression may arise that the calculation of potentials and electric fields of various charge configurations is not too difficult - straightforward integrations are the only real difficulties that may arise. The situation changes, however, if conductors appear in the game. An illustrative example is the field closed to a charged, wildly shaped conducting object, because one needs to distribute the charges over the conductor first. The distribution of charges is subject to the constraint that the surface of this object is an equipotential surface. This problem can be solved by "guesstimating": Guess a distribution and check, whether the object's surface is at equipotential. If this is not the case, refine the guess. This may be quite time-consuming; nature is much faster in doing the guesswork - if the guess is wrong, the charges will continue moving inside the conductor. Of course, the problem may also be solved numerically, but apart from this ultima ratio, there is a number of little problems of that kind which can be solved analytically by using little tricks. Since the tricks are so pretty, and because their presentation may be quite instructive, some of them will be discussed in the following.

[3] This is due to the numerator, similar to the dipole term, which vanishes like $1/R^2$.

3.4.1 Method of mirror charges

Consider the field evoked by two point charges; of course it has various field lines and equipotential surfaces. Take one of the equipotential surfaces and arrange a thin metal layer such that it follows exactly this surface. When the metal layers potential is carefully set to the potential at the equipotential surface, the presence of the metal layer does not change any characteristic of the field around. In fact, from an electrostatic point of view, the full situation did not change. But wait! In fact this is also marks the solution to another problem, namely the calculation of the field of a conductor with corresponding shape and potential near in the vicinity of a point charge. This is especially true if the metal layer is large enough or, eventually, even closed such that it divides the volume into some subvolumes inside and outside a conducting shell. Previously it was shown that the fields inside and outside such a conducting shell are completely independent from each other, and, hence the field outside is independent of what is inside the shell. Therefore, one may calculate the field strength of a point charge in front of the conducting layer by just adding the second point charge and superimposing their potentials. This fictitious second charge is called the "mirror charge" and the method may be called mapping or the "method of mirror charges".

Point charge in front of a conducting plane

Let us illustrate this method by the simple most example, a positive point charge q in front of an infinitely large plane, with distance d. What would the mirror charge look like? For symmetry reasons, the answer is that, without the plane, a second, negative point charge in distance $-d$ from the plane would do the job. Then, all field lines going through the plane would be perpendicular to it. Hence without the second charge, but with the conducting plane the pattern looks identical (at least on the side pointing towards the positive charge), see Fig.3.8. Apart from the positive charge, its field will also induce the shift of some electrons on the metal plane, inducing some effective negative charge density on the side of the plane pointing towards the charge. These charges have a density per unit area, denoted by σ. The normal component of the field at the surface of the plane is given, through Gauss' law, by σ divided by ϵ_0.
consider a point on the surface of the plane with distance ρ to the point on the plane that is directly opposite to the charge. The electric field at this

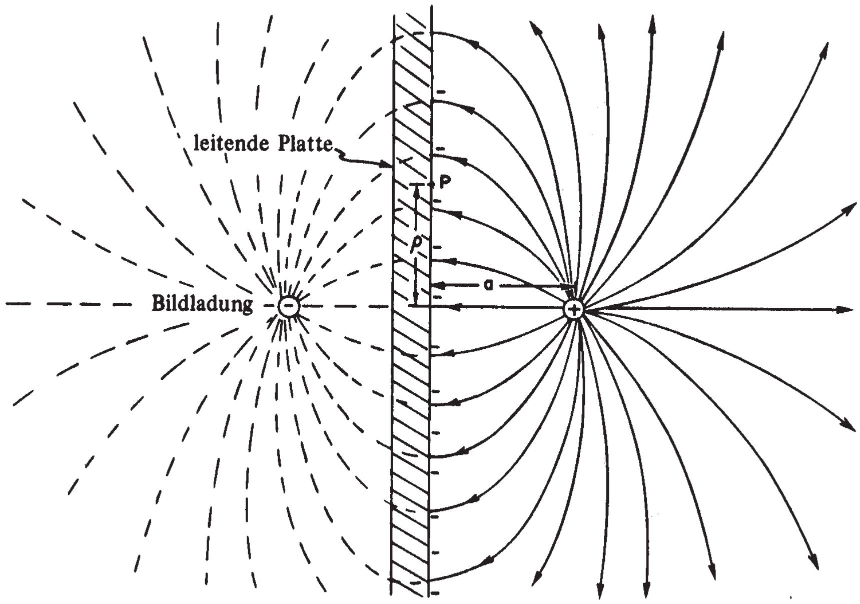

Figure 3.8: Point charge in front of a conducting plane (plus mirror charge indicated on the left of the plane).

position points into the plane and is normal to it. The normal component produced by the point charge reads

$$E_{n+} = \frac{q}{4\pi\epsilon_0}\frac{(\vec{r})_\perp}{r^3} = -\frac{q}{4\pi\epsilon_0}\frac{d}{(d^2+\rho^2)^{3/2}}\,, \tag{3.60}$$

where $\vec{r}$ was the vector denoting the distance from the point on the surface to the charge. The field of the negative mirror charge has the strength and is oriented such that the normal component is doubled and all other components are cancelled. Therefore the charge density reads

$$\sigma(\rho) = \epsilon_0 E(\rho) = -\frac{2dq}{4\pi(d^2+\rho^2)^{3/2}}\,. \tag{3.61}$$

Integrating it over the full plane yields

$$\int_0^{2\pi} \mathrm{d}\phi \int_0^{\infty} \mathrm{d}\rho\, \rho\sigma(\rho) = -qd\int_0^{\infty} \frac{\mathrm{d}\rho^2}{2}(d^2+\rho^2)^{-3/2} = -q\,, \tag{3.62}$$

as expected. Also, the force that the induced charges apply on the point charge is easy to calculate. It just equals the force the mirror charge would apply, hence

$$F = \frac{q^2}{4\pi\epsilon_0(2d)^2}\,, \tag{3.63}$$

pointing towards the plane.

Point charge in front of a conducting sphere

The next conducting surface that will be investigated is a grounded conducting sphere with a point charge in its vicinity. Assume that the sphere has radius a and the point charge is in distance $|\vec{y}|$ from its centre, which is located at the origin. From symmetry reasons it is clear that the mirror charge q' has to be on the line connecting the centre of the sphere with the charge, i.e. its distance vector $\vec{y}' \parallel \vec{y}$, cf. Fig.3.9. At any point $\vec{r}$ the potential of the two charges is

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{|\vec{r}-\vec{y}|} + \frac{q'}{|\vec{r}-\vec{y}'|}\right]\,. \tag{3.64}$$

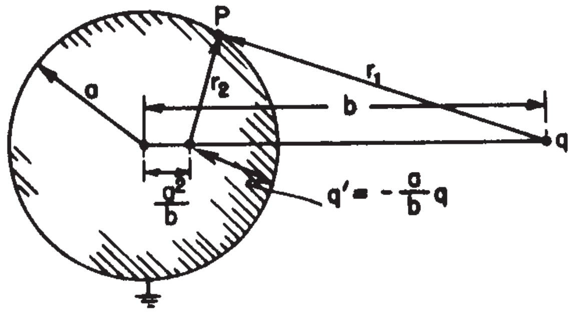

Figure 3.9: Point charge in front of a conducting sphere (plus mirror charge indicated inside the sphere).

For any $\vec{r}$ on the surface of the sphere, factor out suitable terms,

$$4\pi\epsilon_0\phi(\vec{r} = a\vec{e}_r) = \frac{q}{a\left|\vec{e}_r - \frac{y}{a}\vec{e}_y\right|} + \frac{q}{y'\left|\vec{e}_y - \frac{a}{y'}\vec{e}_r\right|} \,. \tag{3.65}$$

In order to have zero potential on the surface of the sphere for all values of $\vec{e}_r \cdot \vec{e}_y$, i.e. under all angles with respect to the line connecting its centre with the charge, the following relations have to hold

$$\frac{q}{a} = -\frac{q'}{y'} \quad \text{and} \quad \frac{y}{a} = \frac{a}{y'} \,, \tag{3.66}$$

leading to the following value and position of the mirror charge

$$q' = -\frac{a}{y}q \quad \text{and} \quad y' = \frac{a^2}{y} \,. \tag{3.67}$$

Hence, moving the charge towards the sphere makes the mirror charge larger and closer to the surface of the sphere. If the charge is nearly touching the sphere, the mirror charge is $-q$ and exactly opposite, on the other side of the sphere[4].

[4]Of course, the same reasoning applies for a charge inside a sphere - then the mirror charge is just outside.

3.4.2 Separation of variables : Cartesian coordinates

Partial differential equations can sometimes be solved through separation of variables. By using this method, one often encounters sets of orthogonal functions, which are of great importance by themselves. In the following this method will be applied to the Laplace equation

$$\nabla^2\phi = 0\,. \tag{3.68}$$

If the potential $\phi(\vec{x})$ can be written in Cartesian coordinates as a product of three functions, one for each coordinate,

$$\phi(\vec{x}) = X(x)Y(y)Z(z) \tag{3.69}$$

then the Laplace equation reads

$$\frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}^2x} + \frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}^2y} + \frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}^2z} = 0\,. \tag{3.70}$$

Here the partial differentials became total ones, since the functions depend on one variable only. In order to guarantee a solution for all $\vec{x}$, the following system has to be solved

$$\begin{aligned}\frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}^2x} &= \alpha^2\\ \frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}^2y} &= \beta^2\\ \frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}^2z} &= \gamma^2\,,\end{aligned} \tag{3.71}$$

with the constraint that

$$\alpha^2 + \beta^2 + \gamma^2 = 0\,. \tag{3.72}$$

A general solution therefore reads

$$\phi(\vec{x}) = \phi_0 e^{\pm\alpha x} e^{\pm\beta y} e^{\pm i\sqrt{\alpha^2+\beta^2}z}\,, \tag{3.73}$$

with ϕ_0 the potential at $\vec{x} = 0$. In order to determine α and β, boundary conditions need to be specified. As an example consider an orthogonal parallelepiped with dimensions $a \times b \times c$ in x-, y, and z-direction, respectively.

Assume now that all sides are on potential zero, apart from the one at $z = c$, which has a fixed potential $V_{z=c}(x,\, y)$.
What is the potential inside the parallelepiped? The constraint that $\phi = 0$ for $x = 0$, $y = 0$, $z = 0$ implies

$$X = \sin(\alpha x)\ , \quad Y = \sin(\beta y)\ , \quad Z = \sinh(\sqrt{\alpha^2 + \beta^2} z)\,. \tag{3.74}$$

The additional constraint that $\phi = 0$ for $x = a$, $y = b$ yields

$$\alpha a = n\pi \quad \text{and} \quad \beta b = m\pi\,, \tag{3.75}$$

where n and m are integers. From these identification of the frequencies, the expansion coefficients of the Z function can be inferred. They read

$$\gamma_{nm} = \sqrt{\alpha^2 + \beta^2} = \pi\sqrt{\frac{n^2}{a^2} + \frac{m^2}{b^2}}\,. \tag{3.76}$$

Thus, the potential can be written as a sum of components, where each of which is given by

$$\phi_{nm} = A_{nm} \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right) \sinh\left(\gamma_{nm} z\right)\ . \tag{3.77}$$

By construction this potential vanishes for $x = 0,\, a$, $y = 0,\, b$, and for $z = 0$. The coefficients A_{nm} can now be fixed such that

$$\sum_{nm} \phi_{nm}(x,\, y,\, z = c) = V_{z=c}(x,\, y)\,. \tag{3.78}$$

This seems horrible at first, by taking a closer look, however, it becomes clear that it is just a double Fourier series for the function $V_{z=c}$. Thus, the coefficients A_{nm} are given by

$$A_{nm} = \frac{1}{\sinh\left(\gamma_{nm} c\right)} \cdot \frac{4}{ab} \int\limits_0^a \mathrm{d}x \int\limits_0^b \mathrm{d}y V_{z=c}(x,\, y) \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right)\ . \tag{3.79}$$

If the potential is different from zero on all sides of the parallelepiped, then the overall potential can be constructed from the superposition principle: Just repeat this procedure for the other five sides, and add the partial results.

3.5 Physical examples

3.5.1 Capacitor

Capacity

The focus will shift now on another problem involving conductors, namely capacitors. In its simplest case, a capacitor consists of two parallel metal slabs whose distance d is small compared to their size. Opposite charges of equal size are brought on the two slabs, the charges distribute evenly on the inner surfaces. The result are area charge densities $\pm\sigma$. Due to their opposite signs, the charges on the two slabs attract each other, the field strength between the two slabs is σ/ϵ_0, cf. Eq.(3.34), outside the system the field strength is, in simplest approximation, zero. The two slabs then have different potentials, their difference

$$V = \phi_1 - \phi_2 \tag{3.80}$$

is called the voltage. The potential difference (or voltage) is the work which is necessary to transport a unit charge from one slab to the other, i.e.

$$V = Ed = \frac{\sigma d}{\epsilon_0} = \frac{dQ}{\epsilon A}\,, \tag{3.81}$$

where Q is the absolute value of the charge brought onto one of the slabs, whose surface is A. Obviously, the voltage is proportional to the charge. This in fact is true for all pairs of conductors with charges of the same size but different sign.
The proportionality is yet another result of the principle of superposition. To understand this, assume that the solution for one pair of charges is known. Adding two such solution yields the result for another situation: The charges are twice as large, hence the field strengths are doubled and the work needed to transport a unit charge has increased correspondingly. Formally this can be expressed through the capacity C:

$$Q = CV\,, \tag{3.82}$$

from which it is easy to deduce the capacity of the two slabs as

$$C_\parallel = \frac{\epsilon_0 A}{d}\,. \tag{3.83}$$

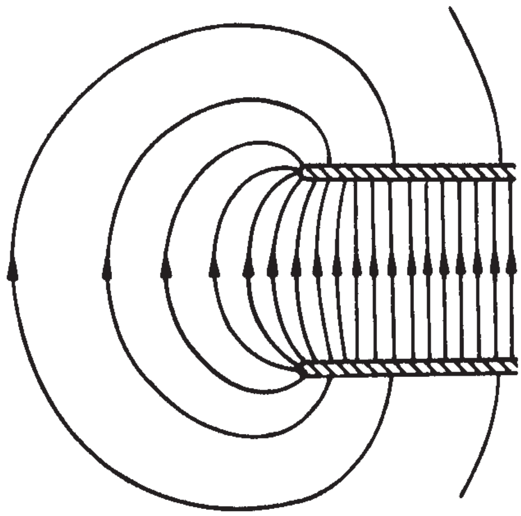

Figure 3.10: Field lines at the edge of a capacitor made of two planes.

This is only approximately true, since the slabs are not infinitely large. At their boundaries the field strength E is not homogenous, but rather looks like in Fig. 3.10. Hence, the total charge is not quite σA; there are small corrections, which are complicated to calculate [5].

So far, capacity was defined as a property of two charged objects. Sometimes, however, it is convenient to talk about the capacity of a single object, like, e.g., a single sphere with radius R. In such a case the assumption is that there is a second sphere enclosing it, with infinite radius. This leads to a capacity

$$C_o = 4\pi\epsilon_0 R \tag{3.84}$$

of the single sphere.

From the discussion so far, it is clear that the unit of the capacity is Coulomb/Volt,

[5] These calculations yield the following result: The charge density increases a little bit towards the boundary of the slabs. Hence the capacity is slightly increased - this increase can be approximated by increasing the area A of the slabs by roughly 3/8 of their distance.

also known as Farad. This leads to

$$\epsilon_0 \approx \frac{1}{36\pi \times 10^9} \text{Farad/Meter}\,. \tag{3.85}$$

Sparks at high voltages

Before continuing with other physical systems, let's discuss qualitatively another property of electric fields in the vicinity of charged conductors. In particular, consider the electric field close to a sharp edge of a charged conductor. There, close to the edge, the field is much stronger than anywhere else on the otherwise smooth conductor. What is the reason for this behaviour? Well, in principle this is quite simple. Inside a conductor, the charges tend to maximise their distance - on the tip of such an edge, the charges are well separated from the other charges inside the rest of the conductor; for an illustration see Fig.3.11 A good fraction of the total charge is therefore forced close to the tip of an edge. Even a small fraction can then cause a large surface density on a small area, leading to strong fields.
The electric field is thus the strongest close to those regions of a conductor which have the largest curvature. To see this consider two conducting balls, connected by a wire, cf. Fig.3.12 Obviously, the wire has only limited influence on the field, it is there to keep both balls on the same potential value. Now, which ball has the larger field on its surface? Well, if the balls have radii $R_1 \gg R_2$ and charges Q_1 and Q_2 respectively, the potentials are approximately

$$\phi_i = \frac{1}{4\pi\epsilon_0} \frac{Q_i}{R_i} \tag{3.86}$$

on the surfaces. Of course the presence of charges on one ball influences the distribution of charges on the other ball, but this effect will be neglected here. After all, the present reasoning is meant to yield some orders of magnitude only. However, since the two potentials must be equal by construction (the wire!), $Q_1/R_1 = Q_2/R_2$. On the other hand, the electric field on the surfaces is proportional to the area charge density, which scales like the inverse square of the radius. Thus

$$\frac{E_1}{E_2} = \frac{Q_1/R_1^2}{Q_2/R_2^2} = \frac{R_2}{R_1}\,. \tag{3.87}$$

Hence the field on the smaller ball is larger than the field on the larger ball, with the ratio roughly scaling like the inverse ratio of the radii.

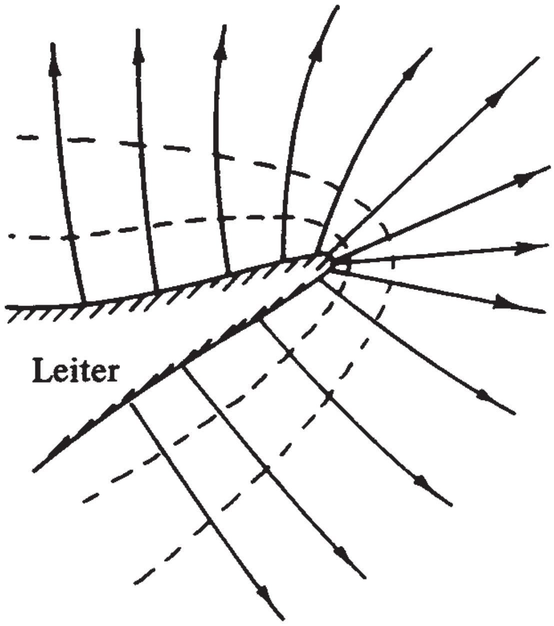

Figure 3.11: Field lines at the edge of a irregularly shaped conductor. At the edge there is a higher density of lines emerging, indicating a stronger field.

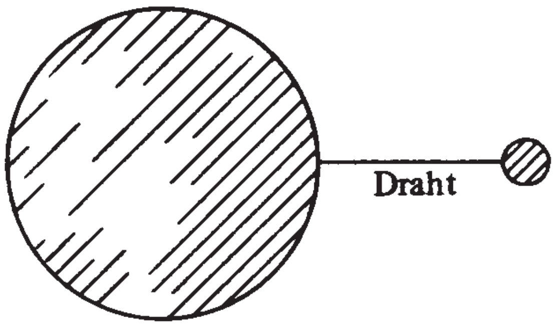

Figure 3.12: Two conducting balls connected by a wire are a good approximation for the conductor of Fig.3.11.

This result is quite important, especially since materials, like, e.g., air, tend to stop isolating when the electric field becomes too strong. When the field becomes too strong, somewhere in the medium a charge gets freed and will be accelerated until it hits another atom or molecule. If the impact is violent enough, more charges get freed, leading to a cascade of ions. This phenomenon is called discharge or spark. Therefore, to keep a piece of metal on a high potential/high voltage, its surface should be plane.

3.5.2 Plasma oscillations

Let us turn now to a physical problem that connects two types of equations, namely, first, the equations of electrostatics, connecting fields with charge distributions, and, second, the equations that govern the movement of charges in the presence of fields.

A simple example is the phenomenon of plasma oscillations. A plasma is nothing but an ionised gas, consisting of ions and free electrons, distributed over some volume. One of the atmospheric layers, the ionosphere, is a realisation of such a system. There, the ultraviolet radiation of the sun is responsible for the process of ionisation. Anyways, in such a plasma, the ions tend to be much heavier than the electrons (a proton is roughly 2000 times more massive than an electron), hence the movement of the ions can be neglected with respect to the movement of the electrons.

Assume now that the equilibrium density of the electrons is n_0. Of course, due to being electrically neutral in total, the ion density must be the same. What happens if for some reason, the density of electrons increases in some area? The electrons there will start to repel each other (remember, the ions are less mobile, hence it is assumed that their density remains constant). The repulsion aims at bringing the electrons back to their equilibrium density. But, on the way back, the electrons gain kinetic energy, such that when they arrive at equilibrium they "overshoot" it. This triggers a oscillation, called "plasma oscillation". It is quite analogous to acoustic waves, where the gaseous pressure produces the force - in the plasma waves, it is the electric repulsion, acting like the pressure.

To simplify the discussion, consider the one-dimensional case. Assume that the electrons located at equilibrium position x at some time t are shifted by some little amount $s(x, t)$. Due to the shift, their density changes. This change is simple to calculate. Consider an interval Δx; its left border is changed by s, its right border is changed by $s + \Delta s$. The number of electrons in Δx before the shift happened equals $n_0 \Delta x$, they are moved into the interval $\Delta x + \Delta s$, the number $n(\Delta x + \Delta s)$ being conserved. Thus the new density n after the shift reads

$$n = \frac{n_0 \Delta x}{\Delta x + \Delta s} = \frac{n_0}{1 + (\Delta s / \Delta x)} = n_0 \left(1 - \frac{\Delta s}{\Delta x} \right) + \mathcal{O}(\Delta s^2) \,. \tag{3.88}$$

The average charge density at each point is therefore given by

$$\rho = -(n - n_0) q_e = n_0 q_e \frac{\mathrm{d}s}{\mathrm{d}x} \,. \tag{3.89}$$

This density is related to a field through the Maxwell equations,

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \,. \tag{3.90}$$

For one-dimensional problems and in the absence of any other field, the electric field has only one component in x-direction, E_x. It is described by

$$\frac{\partial E_x}{\partial x} = \frac{n_0 q_e}{\epsilon_0} \frac{\partial s}{\partial x} \tag{3.91}$$

leading to

$$E_x = \frac{n_0 q_e}{\epsilon_0} s \,. \tag{3.92}$$

Because of $E_x = 0$ for $s = 0$ the constant of integration must equal zero. The force (with a component in x-direction only) acting on each shifted electron reads

$$F_x = -\frac{n_0 q_e^2}{\epsilon_0} s \,. \tag{3.93}$$

It acts as a repulsive force. The corresponding equation of motion of the electrons is

$$m_e \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -\frac{n_0 q_e^2}{\epsilon_0} s \,. \tag{3.94}$$

Its solution in exponential form is something like $\exp(i\omega_p t)$ with the plasma frequency

$$\omega_p^2 = \frac{n_0 q_e^2}{\epsilon_0 m_e} \,. \tag{3.95}$$

When talking about electrical charges, often the entity e^2 is used; it is given by

$$e^2 = \frac{q_e^2}{4\pi\epsilon_0} = 2.3068 \times 20^{-28} \text{Newton} \cdot \text{Meter}^2 \,. \tag{3.96}$$

Then the plasma frequency reads

$$\omega_p^2 = \frac{4\pi n_0 e^2}{m_e} \,, \tag{3.97}$$

a common form in textbooks.
This shows that a perturbation of a plasma results in free oscillations of the electrons around their equilibrium state with a frequency proportional to the square root of their equilibrium density. The plasma electrons behave like a resonant system. This triggers some interesting effects. For instance, when trying to send a radio wave through the ionosphere, one experiences that its frequency must be larger than the plasma frequency. Otherwise the signal is reflected. Therefore, to communicate with a satellite in space, high frequency signals have to be used - in contrast, the signals of a radio station beyond the horizon can be received through reflection as long as their frequency is lower than the plasma frequency.

More details

To understand this in a little bit more detail, consider photons invading a metal (or the ionosphere) and employ some knowledge about electrodynamics inside materials. The force acting on the (free) electrons is given by Newton's law:

$$m[\ddot{\vec{x}}(t) + \omega_0^2 \vec{x}(t)] = -e\vec{E}(t) \,, \tag{3.98}$$

where any friction terms linear in $\dot{\vec{x}}$ have been neglected. Fourier transforming ($t \to \omega$) this differential equation results in

$$m[\omega_0^2 - \omega^2]\tilde{\vec{x}}(\omega) = -e\tilde{\vec{E}}(\omega) \,. \tag{3.99}$$

The dipole moment resulting from shifting the electrons out of their rest position is given by

$$\begin{aligned} \vec{p}(t) &= -e\vec{x}(t) \,, \\ \tilde{\vec{p}}(\omega) &= -e\tilde{\vec{x}}(\omega) = \frac{e^2}{m[\omega_0^2 - \omega^2]}\tilde{\vec{E}}(\omega) \end{aligned} \tag{3.100}$$

The dielectric number $\epsilon = \epsilon_0(1 + \chi_e)$ is connected with the susceptibility which is subject to the constraint that

$$\begin{aligned} n\vec{p}(t) &= \epsilon_0 \chi_e \vec{E}(t)\,, \\ n\tilde{\vec{p}}(\omega) &= \epsilon_0 \chi_e \tilde{\vec{E}}(\omega)\,, \end{aligned} \tag{3.101}$$

where n is the number of dipoles per unit volume.
Consequently, Eq.(3.100) yields

$$\frac{\epsilon_0 \chi_e}{n} = \frac{e^2}{m[\omega_0^2 - \omega^2]} \tag{3.102}$$

As a result, the dielectricity is a function of ω (the frequency of the incoming wave)

$$\frac{\epsilon}{\epsilon_0} = 1 + \chi_e = 1 + \frac{ne^2}{\epsilon_0 m[\omega_0^2 - \omega^2]}\,. \tag{3.103}$$

As a result

$$\epsilon_r(\omega) = \frac{\epsilon(\omega)}{\epsilon_0} = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} \tag{3.104}$$

Using the dispersion relation in medium $c^2k^2 = \omega^2 \epsilon_r(\omega)$ it follows

$$\omega_0 < \omega < \omega_p \quad \Rightarrow \quad \epsilon_r(\omega) < 0 \quad \Rightarrow \quad k^2 < 0\,. \tag{3.105}$$

This means that a wave with wave vector $\vec{k}$ cannot travel in the corresponding medium: it gets reflected.

3.5.3 Electrostatic field of a lattice

As a last example consider the electric field close to an array of charged wires. To keep things simple consider infinitely long parallel wires with distance D in a plane. When investigating the field far away from the array, it will be homogenous, coming closer to the plane of the array, homogeneity is lost, moving in parallel to the array, the field changes periodically. For such periodically changing behaviour a Fourier expansion is the method of choice. Let's try to find a suitable function fulfilling the field equations.

For this, assume that the wires are in the x-y plane and parallel to the y-axis. A good guess is to test expressions like

$$\phi(x,\, y) = F_n(z) \cos \frac{2\pi n x}{D}\,, \tag{3.106}$$

where a sum over n - the harmonic order - is implicit. To render this expression valid, it must satisfy the Laplace equation in the volume well above or below the array of wires. There, there are no charges, hence

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} = 0\,. \tag{3.107}$$

Testing this equation yields an equation for the F_n, namely

$$-\frac{4\pi^2 n^2}{D^2} F_n(z) \cos \frac{2\pi n x}{D} + \frac{\mathrm{d}^2 F_n(z)}{\mathrm{d}z^2} \cos \frac{2\pi n x}{D} = 0\,, \tag{3.108}$$

leading to

$$\frac{\mathrm{d}^2 F_n(z)}{\mathrm{d}z^2} = \frac{4\pi^2 n^2}{D^2} F_n(z)\,. \tag{3.109}$$

A solution therefore is

$$F_n(z) = A_n \exp\left(-\frac{2\pi n z}{D}\right)\,. \tag{3.110}$$

If there is an nth harmonic Fourier component of the field, it must vanish exponentially with a characteristic distance $D/2\pi n$. For the first harmonic ($n = 1$) the field is reduced by a factor $\exp(-2\pi) \approx 1/500$, each time the distance z is increased by a lattice parameter D. After a couple of such steps, therefore all oscillating terms practically vanish, and the only term left is the "zeroth harmonic",

$$\phi_0 = -E_0 z\,, \tag{3.111}$$

yielding the homogenous field at large z. The full result therefore is a combination of all harmonics, including the homogenous term. The coefficients A_n then have to be adjusted such that they produce an electric field that coincides with the charge density on the wires. However, this reasoning explains, why the electrostatic shielding with a grid is often as efficient as the one obtained by a solid metal layer.

Chapter 4

Magnetostatics

In this chapter, the second set of static equations, namely

$$\begin{aligned} \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{B} &= \frac{\vec{j}}{\epsilon_0 c^2}\,, \end{aligned} \tag{4.1}$$

cf. Eq.(3.2), will be discussed. These equations are valid only if all charge densities are constant and if all currents are stationary, i.e. if all fields are constant in time. Before entering this discussion, however, some general comments are in order.
The most important statement in this context is that in many cases it is quite a bit too simplistic to assume that static magnetic situations exist at all - after all, there must be currents to have a magnetic field, and these fields are nothing but moving charges. Therefore "magnetostatics" is just an approximation. The foremost approximation here is that the number of moving charges is large enough that they can be replaced by a stationary current without significant errors. It is only in such a position that there is a current density $\vec{j}$ which does not change with time. The assumption of stationary fields implies that all terms $\partial\vec{E}/\partial t$ and $\partial\vec{B}/\partial t$ vanish. Also, the second equation,

$$\vec{\nabla} \times \vec{B} \propto \vec{j}$$

implies that

$$\vec{\nabla} \cdot \vec{j} = 0\,, \tag{4.2}$$

since the divergence of a rotor vanishes. This also implies that the charge density is a constant, $\partial\rho/\partial t = 0$, which is the case as long as the fields are constant. Thus, the assumptions made so far are consistent with each other. In addition, the condition above, namely that the current density has no divergence, translates into currents along closed lines only (remember that the divergence of a vector field is a measure for its source density). They can float in closed circuits, wires forming a closed loop, maybe attached to a battery to guarantee it being constant. These circuits, however, should not contain capacitors which would act like sources or sinks when charging or discharging.

4.1 Lorentz force and Ampere's law

4.1.1 Lorentz force

The force acting on a charge does not depend on its position only, but also on its velocity. Thus, each position in space is characterised by two vectors, describing the force acting on an arbitrary charge. Oner of them is the electric force, which is independent of the charges velocity. It is described by the electric field strength $\vec{E}$. The second one is the magnetic force, which does depend on the velocity of the charge. It has a pronounced character: Both its direction and its strength do depend on the direction of the charges velocity, the force is always orthogonal with respect to the velocity, in addition it is always orthogonal to a position dependent but fixed reference direction. Finally the strength of the magnetic force is proportional to the velocity component orthogonal to the reference direction. This behaviour can be described by introducing a magnetic field strength $\vec{B}$, fixing both the reference direction and the constant of proportionality for the force, which in turn may be written as $q\vec{v}\times\vec{B}$. Then the total electromagnetic force, known as the Lorentz-force, reads

$$\vec{F} = q(\vec{E} + \vec{v}\times\vec{B})\,. \tag{4.3}$$

The effect of this force is quite simple to test: Just bring a magnet close to a TV set (or any other cathode ray indicator). The deflection of the electron beam proves the presence of a magnetic field. The units of the magnetic field is

$$\frac{\text{Newton}\cdot\text{Second}}{\text{Coulomb}\cdot\text{Meter}} = \frac{\text{Volt}\cdot\text{Second}}{\text{Meter}^2} = \frac{\text{Weber}}{\text{Meter}^2}\,. \tag{4.4}$$

Force on a current

This allows to calculate the force acting on an electric current going through a wire, and, hence, on the wire itself. The current consists of electrons that move with velocity v along the wire, each of them experiencing a transversal force

$$\vec{F} = q_e \vec{v} \times \vec{B} \,. \tag{4.5}$$

For N electrons per unit volume, the number of electrons per small volume ΔV inside the wire is $N\Delta V$, and the magnetic force ΔF acting on them is

$$\Delta\vec{F} = (N\Delta V) q_e \vec{v} \times \vec{B} \,. \tag{4.6}$$

But $Nq_e v$ is just the electric current density j, therefore

$$\frac{\Delta\vec{F}}{\Delta V} = \vec{j} \times \vec{B} \,, \tag{4.7}$$

the force per unit volume is just $\vec{j} \times \vec{B}$. If the wire has a constant cross-section A, then the total electric current I through the wire is jA and the force acting on the wire per unit length is

$$\frac{\Delta\vec{F}}{\Delta L} = \vec{I} \times \vec{B} \,. \tag{4.8}$$

In other words: The force acting on wire with an electric current though it depends on the current. It does not depend on the size or sign of the elementary charges producing the current.

4.1.2 Ampere's law

Since a magnetic field acting on a current results in a force and because of the principle of *actio = reactio* it can be expected that a force acts on the source of a magnetic field when a current is brought nearby. This is a real effect, just take the effect of a current on a compass needle. It is also known that magnets induce forces on each other, hence a current through a wire results in a magnetic field around it. The natural question to ask now is: Which magnetic field is produced by a given current? The answer was found empirically by Ampere through a combination of theoretical reasoning and three experiments.

However, the approach in this context will be to accept that Maxwell's laws have been validated by a huge number of experiments. Therefore, it suffices to inspect equations Eq.(4.1),

$$\begin{aligned}\vec{\nabla}\cdot\vec{B} &= 0\\ \vec{\nabla}\times\vec{B} &= \frac{\vec{j}}{\epsilon_0 c^2}\,.\end{aligned} \tag{4.9}$$

Comparing the first one with its electrostatic counterpart,

$$\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\epsilon_0} \tag{4.10}$$

translates into the statement that there is no magnetic analogue to electric charges: There are no magnetic monopoles. This absence means, pictorially, that there are no charges from which magnetic field lines could start: they must begin and end nowhere, or, in other words, they must be either closed or they must be infinitely long.

The second equation above explains then, how these closed lines emerge. Whenever currents are present, they come into being. Their rotor is proportional to the current density. Currents are surrounded by magnetic field lines forming closed lines around them. This connection between the $\vec{B}$-field and currents is contained in the second equation. In electrostatics its analogue read

$$\vec{\nabla}\times\vec{E} = 0\,, \tag{4.11}$$

which, by the theorem of Stokes, means that the line integral of $\vec{E}$ around a closed line L equals zero,

$$\oint\limits_L \mathrm{d}\vec{l}\cdot\vec{E} = 0\,. \tag{4.12}$$

Applying Stokes theorem on the magnetic equation results in

$$\oint\limits_{L=\delta S} \mathrm{d}\vec{l}\cdot\vec{B} = \int\limits_S \mathrm{d}^2\vec{s}\cdot\left(\vec{\nabla}\times\vec{B}\right) = \frac{1}{\epsilon_0 c^2}\int\limits_S \mathrm{d}^2\vec{s}\cdot\vec{j}\,, \tag{4.13}$$

where the line L encloses the surface S. It should be noted here that the integral of $\vec{j}$ on the r.h.s. of the equation above is nothing but the current I

through the surface S. In magnetostatics, this law, called Ampere's law, plays the same role as Gauss' law in electrostatics. However, Ampere's law alone does not fix the $\vec{B}$ field, in the general case also its property of being source-free ($\vec{\nabla} \cdot \vec{B} = 0$) must be invoked. But, as will be seen in the following, there are some special cases, where symmetry arguments allow to fix $\vec{B}$ without this further condition.

Magnetic field of a straight conductor

As a first example for the results of Ampere's law let us determine the magnetic field close to a wire. More specifically, let us try to answer the question: What is the magnetic field close to a cylindrical wire with a current I going through it. To solve this assume that the field lines of $\vec{B}$ form concentric circles around the wire. With this assumption, and due to the implicit symmetry, $\vec{B}$ has the same strength on all points of such a circle. Therefore the line integral of $\vec{B}$ along such a concentric circle with radius r reads

$$\oint\limits_{L=\delta S} \mathrm{d}\vec{l} \cdot \vec{B} = 2\pi r |\vec{B}| = 2\pi r B \,. \tag{4.14}$$

On the other hand, the current through this circle is nothing but the current I on the wire, hence

$$2\pi r B = \frac{I}{\epsilon_0 c^2} \,, \tag{4.15}$$

leading to

$$B = \frac{1}{4\pi\epsilon_0 c^2} \frac{2I}{r} \,. \tag{4.16}$$

Therefore, the magnetic field B vanishes with the inverse distance from the wire. Writing the equation above in vector form results in

$$\vec{B} = \frac{1}{4\pi\epsilon_0 c^2} \frac{2\vec{I} \times \vec{e}_r}{r} \,. \tag{4.17}$$

here, the factor $1/4\pi\epsilon_0 c^2$ has been pulled out, since it has a frequent appearance. In the MKS-system it equals exactly 10^{-7} - in fact with this equation the unit of the electrical current, the Ampere, is defined: In one meter distance from a wire with current one Ampere the magnetic field equals $2 \cdot 10^{-7}$ Weber per meter squared[1].

Magnetic field inside a coil (solenoid)

As another application exemplifying the virtues of Ampere's law consider a long coil on which a wire is wound up, a solenoid. If the length of the solenoid is much larger than its diameter, it is found experimentally that its magnetic field is much larger in the inside than in the outside. Using this fact, it is quite straightforward to calculate the field inside the coil, which will be approximated to be infinitely long.
Since in the infinite-length approximation the field remains confined inside the coil and since it has divergence zero, the field lines must be parallel to the axis of the solenoid. For simplicity this axis will be taken to be the x-axis. In this case one may apply Ampere's law to the rectangular line exhibited in Fig. **??**. It proceeds in parallel to the field B_0 along a distance L inside the coil, all other parts of the line are either perpendicular on the field or in areas where the field vanishes. The total current enclosed by this line equals the current I times the number of windings N of the wire around the coil. Therefore

$$B_0 L = \frac{NI}{\epsilon_0 c^2}, . \tag{4.18}$$

Introducing the number of winding n per unit length this can be rewritten as

$$B_0 = \frac{nI}{\epsilon_0 c^2}, . \tag{4.19}$$

Leaving the approximation of an infinite coil, one may ask what the field looks like at its ends. Well, at each end they distribute somehow and, after bending back, they return to the other end of the coil. This is exactly the field of a bar magnet. But, what actually is such a bar magnet? How does a piece of material become magnetic? Maxwell's equations state that any

[1]Since a current in a wire produces a magnetic field it will impose a force on any wire nearby that carries a current itself. If the two wires are parallel, each is normal w.r.t. the magnetic field of the other one. Thus, the wires must attract or repel each other, dependent on whether the current are in the same or in opposite directions, respectively.

magnetic field is the result of a current - how does this relate to ordinary iron bars producing a magnetic field? The answer is quite straightforward: Inside magnetic materials there are internal currents, permanent magnets therefore have permanent internal currents. This is a complex issue. At this point it should suffice to say that these internal currents are related to the spin of the electrons. The analogue of spin in classical physics is that the electrons rotate around their own axis. In most materials, the sense and direction of this rotation is distributed randomly, resulting in a zero net effect. In magnetic materials however, there is quite a strong ordering of the spins, and although their individual current is small their overall effect is sizable due to the large number of electrons inside matter.

4.2 Connection of electric and magnetic fields

4.2.1 Forces

When discussing the force on a moving particle induced by a magnetic field you may have asked: Well, if this force is proportional to the velocity, in which system do you measure the particle's speed? In fact, from the definition of the Lorentz force, cf. Eq.(4.3), it seems as if the force crucially depends on the choice of reference system. After some closer inspection, it will turn out that any inertial system will do the job; even more, as a consequence of this finding a deep connection of electric and magnetic fields will emerge - as will be seen, they are just two sides of the same coin. In fact, they must always be understood as a unified and complete electromagnetic field. Although in the stationary case the Maxwell equations seem to separate into two disjoint sets of equations, without any explicit connection between them, there is this deep link provided by the principle of relativity[2]. Let us now apply this principle on electromagnetism.

Qualitative discussion

As an example system consider a negative point charge q moving with velocity v_0 in parallel to a wire with current I on it. What happens in the inertial

[2]Historically, however, this principle of relativity was discovered after the Maxwell equations have been established as the "true" theory of electromagnetism. In fact, the Maxwell equations triggered this discovery.

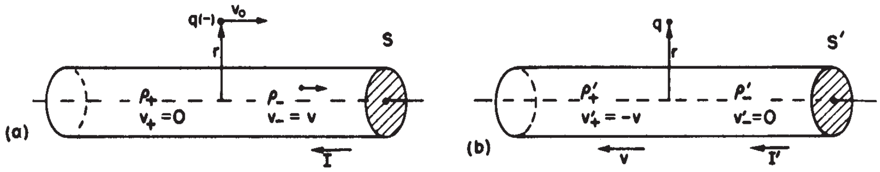

Figure 4.1: The two systems S and S'. S is the inertial system of the wire, S' is the inertial system of the moving point charge.

systems (see Fig.4.1) of the wire (system S) and of the charge (system S')? In the S-system there is, obviously, a magnetic force acting on the particle. This force is directed towards or away from the wire; if the particle was to move freely, its trajectory would therefore bend towards or away from the wire. In contrast, in the S' system, there is no magnetic force acting on the point charge, since there is no velocity. Does it remain where it is? This would imply that two different things happen in the two systems. According to the principle of relativity, also in the system S' the particle should feel a force accelerating it towards or away from the wire.
To understand how this works, let us return to an atomic description of currents. In any conductor, currents emerge as the cooperative movement of a few (negatively charged, of course) electrons. The nuclei and the other electrons, usually from the inner shells, remain more or less fixed and do not move. A current density is then given as the density of "conductor electrons" times their velocity. In the system S, the "conductor-electron" density is given as ρ_- and their velocity is v. This results in a current

$$\vec{I} = A\rho_- \vec{v}\,, \tag{4.20}$$

if the wire has a cross section A. Since the wire has no net charge, the density of fixed positively charged atom remnants is $\rho_+ = -\rho_-$, such that their sum vanishes. Outside the wire there is no electric field, the magnetic field in the charges distance r from the wire is given by

$$\vec{B} = \frac{1}{4\pi\epsilon_0 c^2}\frac{2A\rho_- \vec{v}}{r}\,, \tag{4.21}$$

cf. Eq.(4.16). As a result, the Lorentz force is

$$F = |q\vec{v}_0 \times \vec{B}| = \frac{1}{4\pi\epsilon_0 c^2} \frac{2qA\rho_- v_0 v}{r} \,. \tag{4.22}$$

Of course, one could choose any velocity v_0 for the moving charge, but to keep things simple, let's choose the same velocity as the velocity of the conducting electrons inside the wire, v. Then

$$F = \frac{q}{2\pi\epsilon_0} \frac{A\rho_-}{r} \frac{v^2}{c^2} \,. \tag{4.23}$$

What happens now in S'? Well, the conducting electrons now seem to be in rest, but the positively charged atomic remnants induce a current and, thus, a magnetic field B'. But since the particle is at rest, therefore, no force is acting on it. Hence, any force must be due to an electric field. Consequently, in order to fulfil the principle of relativity, the moving wire must produce an electric field, which is possible only, if it was charged. This translates into the statement that any neutral wire at rest with a current through it seems to be charged when moving.
Let us therefore try to calculate the charge density of the wire in system S' based on what is known from it in system S. Naively, it would seem as if the two charge densities $\rho_\pm$ are identical. But it is known that lengths and hence volumes change when going from one inertial system to another one; since the charge densities depend on the volumes, they may therefore change. Before going into this, a few words about the charges themselves are in order. It is known from relativity that relativistic masses change like $1/\sqrt{1-v^2/c^2}$ when moving. This is not true for charges. They remain the same independent of whether they move or not. Otherwise, charge conservation would be violated.

Quantitative results

Thus the change of volume when going from S to S' must be calculated. To do so, notice first that lengths change relativistically only in the direction of transformation, length contraction only occurs in the direction of a boost. From special relativity it is known that any piece of wire of length L in system S, will have the length L' in system S' given by

$$L' = L\sqrt{1-v^2/c^2} \,. \tag{4.24}$$

The cross section will not change, since the boost direction being proportional to the velocity vector of the point charge (and, according to a previous assumption, to the current inside the wire) is perpendicular to the area. Thus $A' = A$. If ρ is a charge density in S, the total charge is $Q = \rho LA$. But because the total charges are independent of the system, $Q' = Q$, and

$$\rho' L' A' = \rho LA \quad \Longrightarrow \quad \rho' L' = \rho L\,. \tag{4.25}$$

Thus

$$\rho' = \frac{\rho}{\sqrt{1 - v^2/c^2}}\,. \tag{4.26}$$

This is nice! The charge density of a distribution of charges transforms under boosts like the relativistic mass of a particle. Applying this on the positive charge density results in

$$\rho'_+ = \frac{\rho_+}{\sqrt{1 - v^2/c^2}}\,. \tag{4.27}$$

This is because in S the positive charges are at rest, in S' they therefore move with velocity v. In contrast, in S the negatively charged conducting electrons are moving with velocity v and they are at rest in S', leading to

$$\rho_- = \frac{\rho'_-}{\sqrt{1 - v^2/c^2}}\,. \tag{4.28}$$

It is now simple to understand why there are electric fields in S': The wire has a net charge given by

$$\rho' = \rho'_+ + \rho'_- = \frac{\rho_+}{\sqrt{1 - v^2/c^2}} + \rho_- \sqrt{1 - v^2/c^2} = \rho_+ \frac{v^2/c^2}{\sqrt{1 - v^2/c^2}}\,, \tag{4.29}$$

where $\rho_+ = -\rho_-$ has been used. Thus the moving wire is positively charged and therefore produces an electric field at the point where the point charges resides. According to Eq.(3.31) the electric field of a charged line with charge density λ per unit length reads

$$E = \frac{\lambda}{2\pi\epsilon_0 r}\,, \tag{4.30}$$

resulting in

$$E' = \frac{\rho' A}{2\pi\epsilon_0 r} = \frac{\rho_+ A v^2/c^2}{2\pi\epsilon_0 r\sqrt{1-v^2/c^2}}\,. \tag{4.31}$$

Consequently the force acting on the point charge in S' reads

$$F' = \frac{q}{2\pi\epsilon_0}\frac{\rho_+ A}{r}\frac{v^2/c^2}{\sqrt{1-v^2/c^2}} \tag{4.32}$$

and it is directed towards or away from the wire. Comparing this with the force in system S, Eq.(4.23), yields

$$F' = \frac{F}{\sqrt{1-v^2/c^2}}\,, \tag{4.33}$$

which is for sure identical for small velocities v. However, physics here is even more beautiful. Still the transformation of forces under boosts has to be included in the reasoning above. A simple way to do so is to use the identity

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t}\,. \tag{4.34}$$

The momentum of a particle changes with time, when forces are acting on it. Now, the change of momentum for sure is in transversal direction, therefore, $\mathrm{d}\vec{p}$ is the same in both systems, i.e. $\mathrm{d}\vec{p} = \mathrm{d}\vec{p}'$. But time changes under boosts - an effect known as time dilation. Time intervals appear to be longer on a moving particle than on the same particle at rest. More specifically

$$\mathrm{d}t = \frac{\mathrm{d}t'}{\sqrt{1-v^2/c^2}}\,, \tag{4.35}$$

since the particle is at rest in S'. Combining these findings shows that physics is the same in both systems; the forces are identical when properly transformed.

Choosing any other inertial system to analyse the situation would just result in yet another combination of electric and magnetic fields, yielding the same force. This shows that electric and magnetic forces and fields are parts of the same physical phenomenon, the electromagnetic interaction[3]. The separation

[3]This in fact was the first "unification" of forces in physics.

into electric and magnetic components depends on the choice of system, a complete description is invariant, taking properly into account both electricity and magnetism is consistent with Einstein's theory of special relativity. This reasoning means that the identification with field lines has to be used with care: In system S for instance, there are magnetic field lines, but no electric ones, and in system S' there are electric field lines.

4.2.2 Transformation of charges and currents

Now the behaviour of charge and current density under Lorentz transformations will be formalised. It will turn out that they can be understood as the temporal and spatial components of a four-vector. If ρ is the charge density of a charge distribution in its rest system S, then in a system S' moving with velocity $\vec{v}$ with respect to S the charge density is

$$\rho' = \frac{\rho}{\sqrt{1 - v^2/c^2}} . \tag{4.36}$$

In S' the resulting current density is

$$\vec{j}' = \rho' \vec{v} = \frac{\rho \vec{v}}{\sqrt{1 - v^2/c^2}} . \tag{4.37}$$

From special relativity it is known that energy U and momentum p of a particle of rest mass m_0 moving with velocity v is given by

$$U = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} \quad \text{and} \quad \vec{p} = \frac{m_0 \vec{v}}{\sqrt{1 - v^2/c^2}} , \tag{4.38}$$

respectively. It is also known that U and $\vec{p}$ form a four-vector, implying a specific transformation behaviour under boosts. Consider, for instance, the four-vector $(t, \vec{x}) = (t, x, y, z)$ (in S) in a system S' moving with velocity $\vec{u} = u\vec{e}_x$ with respect to S'. According to special relativity,

$$t' = \frac{t - \vec{u} \cdot \vec{x}/c^2}{\sqrt{1 - u^2/c^2}} = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} \quad \text{and} \quad \vec{x}' = \frac{\vec{x} - \vec{u}t}{\sqrt{1 - u^2/c^2}} . \tag{4.39}$$

The four-vector $(\rho, \vec{j})$ transforms in full analogy - this is what four-vector basically boils down to: Identical behaviour under Lorentz transformations,

$$\rho' = \frac{\rho - \vec{u} \cdot \vec{j}/c^2}{\sqrt{1 - u^2/c^2}} = \frac{\rho - uj_x/c^2}{\sqrt{1 - u^2/c^2}} \quad \text{and} \quad \vec{j}' = \frac{\vec{j} - \vec{u}\rho}{\sqrt{1 - u^2/c^2}} . \tag{4.40}$$

This is obtained by just replacing $t \to \rho$ and $\vec{x} \to \vec{j}$. With this charges and currents in different inertial systems can be related allowing for solutions of the electromagnetic problem as given by the Maxwell equations in different systems. Later on the lecture will return to this by discussing the relativistic transformation properties of the fields and by formulating the Maxwell equations in a relativistically invariant manner.

4.3 The vector potential

4.3.1 Towards a constructive solution

General remarks

So far magnetic fields produced by stationary currents have been guessed, in the following methods will be discussed to solve the equations describing magnetostatics systematically, i.e. without extra symmetries or estimates. In electrostatics a method for such a systematic solution has been provided by the introduction of the scalar potential ϕ which is related to the electric field through its gradient; ϕ itself was found to be related to an integral over the charge distribution, see Eq.(3.15). In the following a similar method for the constructive solution of the general magnetostatic problem will be discussed. In electrostatics, the potential ϕ could be constructed since the rotor of the electric field vanishes, rendering it possible to write it as the gradient of a scalar potential. In magnetostatics in contrast, the rotor if $\vec{B}$ in general is not zero, instead the divergence of the magnetic field vanishes. Therefore $\vec{B}$ can always be written as the rotor of a vector potential, see also the reasoning that lead to Eq.(2.18). There are even more similarities to the electrostatic case. There, any potential ϕ yielding a solution for a given problem could be replaced by another one, $\phi + C$, with a constant C added to it. This is due to the fact that grad $C = 0$ for all constants C, resulting in the same electric fields and, hence, to the same physics. This means that one may even add a function ψ to any potential without changing the fields as long as its gradient vanishes, grad $\psi = 0$. In fact it is this freedom of gauge that hints at the fact that the potentials are physically not observable. Similarly to the electrical potential, one may add the gradient of any function ψ to any vector potential solving a magnetostatic problem. In other words, if $\vec{A}$ is the solution to a magnetostatic problem, the magnetic field $\vec{B} = \vec{\nabla} \times \vec{A}$

remains the same when replacing $\vec{A}$ by

$$\vec{A}' = \vec{A} + \vec{\nabla}\psi\,, \tag{4.41}$$

cf. Eq.(2.26). This degree of freedom can be used. In the electric case, for instance, it proved to be advantageous to demand that the potential vanishes at infinity, for the vector potential in magnetostatics the condition that its divergence vanishes, i.e. that

$$\vec{\nabla}\cdot\vec{A} = 0 \tag{4.42}$$

imposes a useful constraint[4].

A first, simple example

To highlight this reasoning let us investigate the vector potential of a homogenous magnetic field $\vec{B}_0 = B_0\vec{e}_z$ pointing, as indicated, in the direction of the z-axis. From $\vec{B} = \vec{\nabla}\times\vec{A}$ one has

$$\begin{aligned}
B_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0\,,\\
B_y &= \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = 0\,,\\
B_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B_0\,.
\end{aligned} \tag{4.43}$$

A possible solution is

$$\vec{A} = (0,\, xB_0,\, 0)^T\,, \tag{4.44}$$

but also

$$\vec{A} = (-yB_0,\, 0,\, 0)^T\,, \tag{4.45}$$

or any linear combination of both leads to the same magnetic field. In particular, there is one interesting linear combination ,namely

$$\vec{A} = \frac{1}{2}(-yB_0,\, xB_0,\, 0)^T = \frac{1}{2}\vec{B}_0\times\vec{r}\,, \tag{4.46}$$

[4]This however, does not yet fully fix the vector potential, usually it is also subject to the condition that the emerging field vanishes at infinity.

with $\vec{r} = (x,\, y,\, z)^T$. This is interesting; it seems as if the vector potential itself is circulating. In fact, if, for instance, $\vec{B}_0$ is the homogenous field inside a long coil, the corresponding vector potential in this form circulates in the same direction as the current in the coil.
The vector potential of a (homogenous) magnetic field can be obtained in yet another way. Since the circulation of $\vec{A}$ around a closed line can be connected to the integral of its rotation over the enclosed area (Stokes' theorem),

$$\oint\limits_{L=\delta S} \mathrm{d}\vec{l} \cdot \vec{A} = \int\limits_{S} \mathrm{d}^2\vec{s} \cdot (\vec{\nabla} \times \vec{A})\,,$$

and due to the definition of the vector potential one finds

$$\oint\limits_{L=\delta S} \mathrm{d}\vec{l} \cdot \vec{A} = \int\limits_{S} \mathrm{d}^2\vec{s} \cdot \vec{B}\,. \tag{4.47}$$

The circulation of $\vec{A}$ around an arbitrary closed line equals the flux of $\vec{B}$ through the enclosed area. By choosing a circle with radius r, one recovers the above result.

Master formula for the calculation of the vector potential

So far, the vector potential has been obtained from the magnetic field. Usually, however, the procedure is the other way around, and it will turn out that especially for more complicated situations the calculation of the vector potential before obtaining the magnetic field from it is the simpler option.
Starting from

$$\vec{\nabla} \times \vec{B} = \frac{\vec{j}}{\epsilon_0 c^2} \tag{4.48}$$

one immediately has (after applying the gauge condition $\vec{\nabla} \cdot \vec{A} = 0$)

$$\vec{\nabla} \times \vec{\nabla} \times \vec{A} = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2\vec{A} = -\nabla^2\vec{A} = \frac{\vec{j}}{\epsilon_0 c^2}\,, \tag{4.49}$$

which plays the same role in magnetostatics as

$$\vec{\nabla} \cdot \vec{\nabla}\phi = -\frac{\rho}{\epsilon_0} \tag{4.50}$$

in electrostatics. The only difference of both is that in the electrostatic case there was just one component, and, hence, one integral to be carried out, whereas in magnetostatics there are three (vector-) components, and, hence, three integrals to be evaluated. Replacing $\phi \to \vec{A}$ and $\rho \to \vec{j}/c^2$ in Eq.(3.15) one has for the vector potential

$$\vec{A}(\vec{r}) = \frac{1}{4\pi\epsilon_0 c^2} \int \frac{\mathrm{d}^3 r' \vec{j}(\vec{r'})}{|\vec{r} - \vec{r'}|} \,. \tag{4.51}$$

4.3.2 Some applications

Straight conductor, once again

As a first example of how to use the master formula, Eq.(4.51), let us recalculate the magnetic field of a long straight wire with radius a and a current I through it. After orienting the wire along the z-axis, the current density has a z-component only, given by

$$j_z = \frac{I}{\pi a^2} \,. \tag{4.52}$$

Due to $j_x = j_y = 0$ one immediately has $A_x = A_y = 0$. In order to calculate A_z one may look up the solution for the electric field of a wire with charge density λ, which is, mathematically, the analogous problem. From Eq.(3.31) one has

$$\vec{E}(\rho) = \frac{\lambda \vec{e}_\rho}{2\pi\epsilon_0 \rho} \,,$$

and integrating it over ρ to gain the potential yields

$$\phi(\rho) = -\frac{\lambda}{2\pi\epsilon_0} \ln \rho \,. \tag{4.53}$$

The same procedure works out for the vector potential of a current through a wire. Identifying the charge density per unit length λ with the current density per unit length $j_z \pi a^2$ and including the extra factor $1/c^2$ the vector potential at distance ρ from the wire reads

$$A_z(\rho) = -\frac{\pi a^2 j_z}{2\pi\epsilon_0 c^2} \ln \rho = -\frac{I}{2\pi\epsilon_0 c^2} \ln \rho \,. \tag{4.54}$$

The magnetic field can now be obtained through differentiation. Only two out of the six terms are different from zero, namely

$$\begin{aligned} B_x &= -\frac{I}{2\pi\epsilon_0 c^2}\frac{\partial \ln\rho}{\partial y} &= -\frac{I}{2\pi\epsilon_0 c^2}\frac{y}{\rho^2}\,, \\ B_y &= \frac{I}{2\pi\epsilon_0 c^2}\frac{\partial \ln\rho}{\partial x} &= \frac{I}{2\pi\epsilon_0 c^2}\frac{x}{\rho^2}\,, \\ B_z &= 0\,. \end{aligned} \tag{4.55}$$

Writing it in a more catchy way, and employing cylindrical coordinates consistently one finds

$$\vec{B} = -\frac{I}{2\pi\epsilon_0 c^2}\frac{1}{\rho}\,\vec{e}_\phi\,. \tag{4.56}$$

Solenoid, revisited

The next example to be tackled is the long solenoid. Through its surface there is a circular current with density nI per unit length[5]. Similar to the charge density per area now a current density $\vec{J}$ per area can be defined; it is equal to the current per unit area on the surface of the coil. In the current case, $\vec{J}$ is given as

$$\vec{J} = -nI\vec{e}_\phi = nI\begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix} \tag{4.57}$$

in cylindrical coordinates.
Let's try first to calculate the x-component of the vector potential A_x outside the coil. The result is the same as the one for the electrical potential outside a cylinder with area charge density

$$\sigma = \sigma_0\sin\phi = \frac{nI}{c^2}\,\sin\phi\,. \tag{4.58}$$

This looks complicated. But in fact it can be solved after a little detour. In Section 3.3 it has been found out that the potential of a charge distribution that can be mapped onto a superposition of known charge distributions is

[5] Imagine a thin wire with n windings per unit length around the coil and neglect the little angle due to the winding as well as the small diameter of the wire.

just the superposition of the individual potentials. To gain such a charge distribution proportional to $\sin\phi$, consider two uniformly charged circles with equal radii R, where one is positively and the other one is negatively charged. Overlay them and move their centers apart in the y direction by a distance much smaller than R. If both circles carry charges of the same size (but different sign), the net charge in the overlap region vanishes. At one end, where the positively charged circle exceeds the negative one, there is a positive net charge and vice versa. The charge distribution is equal to a charge density on the surface of the circle, which is proportional to $\sin\phi$. What is the common electrical potential of the two charge distributions? Well, the answer is simple; it looks like a dipole potential, which according to Eq.(3.54) can be written as being proportional to the gradient of the potential of one circle times the dipole moment (pointing in y-direction). Therefore,

$$\phi \propto \frac{\partial \ln \rho}{\partial y} = \frac{y}{\rho^2} \,. \tag{4.59}$$

The same reasoning also applies for the y component of the vector potential, which is the same as the one for the electrical potential outside a cylinder with area charge density

$$\sigma = -\sigma_0 \cos\phi = -\frac{nI}{c^2} \cos\phi \,. \tag{4.60}$$

There, however, the two circles are shifted in x-direction. Taken together, the vector potential outside the solenoid should look like

$$A_x = -K\frac{y}{\rho^2} \,, \quad A_y = K\frac{x}{\rho^2} \,, \tag{4.61}$$

where K is some constant of proportionality to be fixed later on. This is strange. Although the magnetic field outside the solenoid should vanish, the vector potential does not. Simple calculation show that trivially $B_x = B_y = 0$ and

$$B_z = \frac{\partial}{\partial x}A_y - \frac{\partial}{\partial y}A_x = K\left(\frac{1}{\rho^2} - \frac{2x^2}{\rho^4} + \frac{1}{\rho^2} - \frac{2y^2}{\rho^4}\right) = 0 \,, \tag{4.62}$$

as it should be. Hence the magnetic field outside a long solenoid is in fact zero, although the vector potential does not vanish. The constant K of proportionality can now be fixed by employing the fact that the magnetic field

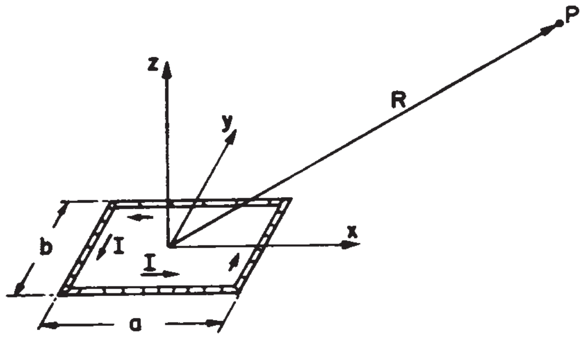

Figure 4.2: Configuration of a magnetic dipole: A small rectangular wire with homogenous current around it from a large distance.

inside the solenoid is homogenous. The circulation of the vector potential around any circle with radius ρ larger than the radius R of the solenoid is $2\pi\rho A = 2\pi K$, independent of the radius of the circle. This must be the case, as long as there is no B-field outside the solenoid, because then only the constant B-field inside the solenoid contributes with a flux of $B\pi R^2$. Equating this and inserting the result of Eq.(4.19) yields

$$K = \frac{B\pi R^2}{2\pi} = \frac{\pi R^2}{2\pi}\frac{nI}{\epsilon_0 c^2}\,. \tag{4.63}$$

Therefore the vector potential outside the solenoid reads

$$\vec{A}(\rho) = \frac{nIR^2}{2\epsilon_0 c^2}\frac{1}{\rho}\,\vec{e}_\phi\,. \tag{4.64}$$

A small loop, magnetic dipole

Now the vector potential will be employed to determine the magnetic field of a small rectangular wire with a homogenous current on it. As usual "small" translates into the interest in fields far away from the loop, cf.Fig.4.2. As it will turn out, any small circular current produces a magnetic dipole, a field configuration similar to the electric field of an electric dipole. The rectangle is oriented in the x-y plane with dimensions a and b in x- and y-direction, respectively. Since there are no currents in z-direction, A_z equals zero. Let us

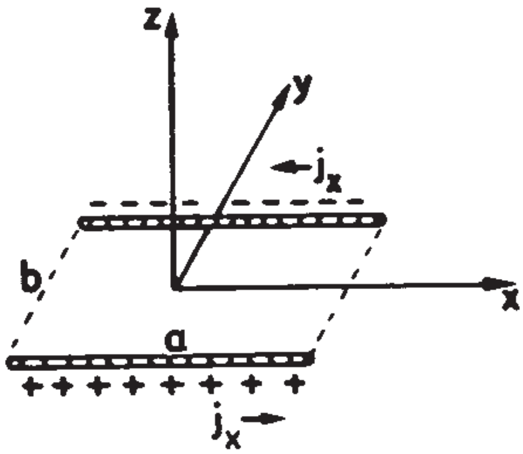

Figure 4.3: The electrical analogue of two wires with current in opposite direction - two thin slabs with opposite charge.

consider first the two pieces of wire parallel to the x-axis. Since the currents have opposite direction, the electrostatic analogue is that of two thin slabs with homogenous but opposite charge density in a relative distance b, see Fig.4.3. Due to the opposite sign of the charges, the electrical potential at large distance equals the dipole potential,

$$\phi = \frac{1}{4\pi\epsilon_0}\frac{\vec{p}\cdot\vec{r}}{r^3}, \tag{4.65}$$

where $\vec{p}$ is the dipole moment of the charge distribution. In this case the dipole moment is just the total charge on one slab times the distance of both slabs, i.e.

$$\vec{p} = -\lambda ab\vec{e}_y\,. \tag{4.66}$$

Adjusting parameters and replacing the charge on the slab by the current through it ($\lambda \to I/c^2$), yields

$$A_x = -\frac{Iab}{4\pi\epsilon_0 c^2}\frac{\vec{e}_y\cdot\vec{r}}{r^3} \tag{4.67}$$

for the x-component of the vector potential. Similarly

$$A_y = \frac{Iab}{4\pi\epsilon_0 c^2} \frac{\vec{e}_x \cdot \vec{r}}{r^3} . \tag{4.68}$$

Clearly, the vector potential rotates around around the z-axis with the same orientation as the current. The strength of A is proportional to Iab, the product of the current through and the area enclosed by the wire. This is called the magnetic dipole moment μ, in vector form it reads

$$\vec{\mu} = I\vec{a} \, , \tag{4.69}$$

where $\vec{a}$ is the area oriented according to the right-hand rule w.r.t. the current through its edge. Then

$$\vec{A} = \frac{1}{4\pi\epsilon_0 c^2} \frac{\vec{\mu} \times \vec{r}}{r^3} . \tag{4.70}$$

Finally the magnetic field remains to be calculated. It reads

$$\begin{aligned}
B_x &= -\frac{\partial}{\partial z} A_y \sim \frac{3xz}{r^5} \, , \\
B_y &= \frac{\partial}{\partial z} A_x \sim \frac{3yz}{r^5} \, , \\
B_z &= \frac{\partial}{\partial x} A_y - \frac{\partial}{\partial y} A_x \sim -\left(\frac{1}{r^3} - 3z^2 r^5 \right) .
\end{aligned} \tag{4.71}$$

Obviously, the components of the $\vec{B}$-field behave like the ones of the $\vec{E}$-field of a dipole pointing in z-direction. That's why the magnetic field is dubbed a "magnetic dipole field", although this naming is somewhat misleading (after all, without monopoles, you cannot have a dipole), since it is produced by a closed current.

Nevertheless, it is surprising that starting with different laws ($\vec{\nabla} \times \vec{B} \propto \vec{j}$ and $\vec{\nabla} \cdot \vec{E} \propto \rho$) leads to nearly identical results. Why is this the case? The answer is simple: It is not because of the specific form of the equation but because of the approximation - fields far away from the well-located charges or currents. For the dominant part of the volume therefore both fields obey the same relations. For instance, both the electric and the magnetic field have divergence and rotor equal to zero in most or all of the volume considered. This motivates the similarity of the solution although the charges or currents are so different.

The law of Biot and Savart

In electrostatics the electric field at a position $\vec{r}_1$ can be obtained directly from a charge distribution through Eq.(3.8),

$$\vec{E}(\vec{r}_1) = \frac{1}{4\pi\epsilon_0} \int \mathrm{d}^3 r_2 \frac{\rho(\vec{r}_2)(\vec{r}_1 - \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|^3}\,. \tag{4.72}$$

Of course, in most cases it is more convenient to make the detour of determining the potential and calculating the field as its gradient, which leaves one integral and some differentials to be evaluated instead of three integrals as in the equation above. However, there is a analogue to the equation above in magnetostatics, namely

$$\vec{B}(\vec{r}_1) = \vec{\nabla} \times \vec{A}(\vec{r}_1) = \vec{\nabla} \times \left[\frac{1}{4\pi\epsilon_0 c^2} \int \mathrm{d}^3 r_2 \frac{\vec{j}(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|}\right]. \tag{4.73}$$

Note that this expression has to be treated carefully. This is because in order to arrive at the $\vec{B}$-field at coordinates 1 one needs to form the rotor of $\vec{A}$ at position 1, implying that all differentials are to be taken w.r.t. the coordinates 1. Since the integration is over coordinates 2 it is therefore allowed to pull the rotor into the integral as long as it involves only differentiation w.r.t. 1. After some calculation, it is clear that

$$\vec{B}(\vec{r}_1) = \frac{1}{4\pi\epsilon_0 c^2} \int \mathrm{d}^3 r_2 \frac{\vec{j}(\vec{r}_2) \times (\vec{r}_1 - \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|^3}\,. \tag{4.74}$$

As long as the currents are confined to a thin wire along a path S_2, this can be rewritten as

$$\vec{B}(\vec{r}_1) = \frac{1}{4\pi\epsilon_0 c^2} \int\limits_{S_2} \frac{I \mathrm{d}\vec{s}_2 \times (\vec{r}_1 - \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|^3}\,, \tag{4.75}$$

where $\mathrm{d}\vec{s}_2$ is a line element along the wire. This is called the law of Biot and Savart.

Chapter 5

Dynamical situations

Having discussed static situations in the past chapters let us turn now to the discussion of some dynamic situations. They are described by adding in the time-dependent pieces of Maxwell's equations - this will couple the electric and magnetic fields. As a first example of how this works out, the law of induction will be analysed. It will turn out that temporal changes of a magnetic field induce an electric field, and, hence, a corresponding potential. Such a potential then leads to induced currents, when conductors are present. The induced quantities also change with time. This basically is the working principle of machines like alternating current generators. Of course, in turn, the induced currents may lead to magnetic fields, which then also change in time, leading to complicated situations when a number of coils are present. The second example for dynamical situations concerns electromagnetic radiation, light. A sketchy discussion of how light is emitted can be found in Sec. 2.3. In this chapter, the previous discussion will be formulated in a more formal way and some simple solutions for the propagation of light will be found. Also, a simple model for antennae will be discussed.

5.1 Electromagnetic induction

5.1.1 The electromotive force

Integral form of induction law - time-dependent geometry

The induction law (a.k.a. Faraday's law) in its differential and integral form reads

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \quad \text{and} \quad \oint\limits_{\delta S} \mathrm{d}\vec{l} \cdot \vec{E} + \int\limits_{S} \mathrm{d}^2\vec{s} \cdot \frac{\partial \vec{B}}{\partial t} = 0\,, \tag{5.1}$$

respectively, cf. Eq.(2.29). The integral form, however, is valid only under the (implicit) assumption that the area does not change with time. What does it look like, when the area changes with time? To see this we start by making all dependencies on time etc. in the fields explicit:

$$\oint\limits_{L(t)=\delta S(t)} \mathrm{d}\vec{l} \cdot \vec{E}(\vec{l}, t) = -\int\limits_{S(t)} \mathrm{d}^2\vec{s} \cdot \frac{\partial \vec{B}(\vec{r}, t)}{\partial t}\,. \tag{5.2}$$

Let's consider first the temporal derivative of the magnetic flux through the surface, namely

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int\limits_{S(t)} \mathrm{d}^2\vec{s} \cdot \vec{B} &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\{ \int\limits_{S(t+\Delta t)} \mathrm{d}^2\vec{s} \cdot \vec{B}(t+\Delta t) - \int\limits_{S(t)} \mathrm{d}^2\vec{s} \cdot \vec{B}(t) \right\} \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\{ \int\limits_{S(t+\Delta t)} \mathrm{d}^2\vec{s} \cdot \vec{B}(t) + \Delta t \int\limits_{S(t+\Delta t)} \mathrm{d}^2\vec{s} \cdot \frac{\partial \vec{B}(t)}{\partial t} - \int\limits_{S(t)} \mathrm{d}^2\vec{s} \cdot \vec{B}(t) \right\} \\
&= \int\limits_{S(t)} \mathrm{d}^2\vec{s} \cdot \frac{\partial \vec{B}(t)}{\partial t} + \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\{ \int\limits_{S(t+\Delta t)} \mathrm{d}^2\vec{s} \cdot \vec{B}(t) - \int\limits_{S(t)} \mathrm{d}^2\vec{s} \cdot \vec{B}(t) \right\},
\end{aligned} \tag{5.3}$$

where in going from the first to the second line the Taylor expansion of $\vec{B}(t+\Delta t)$ up to first order in Δt,

$$\vec{B}(t+\Delta t) = \vec{B} + \Delta t\, \frac{\partial \vec{B}}{\partial t} + \mathcal{O}((\Delta t)^2) \tag{5.4}$$

has been employed. Of course, there is the limit left in Eq.(5.3), which still waits to be evaluated. To this end, it is important to note that the (moving) surface $S(t)$ sweeps over a volume ΔV during any time interval Δt. The integral over the surface of V is then given by

$$\begin{aligned}\oint\limits_{\partial(\Delta V)} \mathrm{d}^2\vec{s}\cdot\vec{B} &= \int\limits_{S(t+\Delta t)} \mathrm{d}^2\vec{s}\cdot\vec{B} - \int\limits_{S(t)} \mathrm{d}^2\vec{s}\cdot\vec{B} + \int\limits_{\Delta S} \mathrm{d}^2\vec{s}\cdot\vec{B} \qquad (5.5)\\ &= \int\limits_{\Delta V} \mathrm{d}^3V\,\vec{\nabla}\cdot\vec{B} = 0\,,\end{aligned}$$

where ΔS is the part of the surface not covered by either $S(t)$ and $S(t+\Delta t)$. The edge of $S(t)$, $\delta S(t)$, produces it while going from $S(t)$ to $S(t+\Delta t)$. Thus, the surface elements on ΔS reads

$$\mathrm{d}^2\vec{s} = \mathrm{d}\vec{r}\times\vec{v}\Delta t\,, \tag{5.6}$$

where, of course, $\vec{v} = \vec{v}(\vec{r},\,t)$. Note that in the expression above, Eq.(5.5), the minus sign compensates for the "wrong" orientation of the flux through $S(t)$: The normal vector of any piece has to be directed to the outside of V.

$$\begin{aligned}&\int\limits_{S(t+\Delta t)} \mathrm{d}^2\vec{s}\cdot\vec{B} - \int\limits_{S(t)} \mathrm{d}^2\vec{s}\cdot\vec{B}\\ &\quad= -\int\limits_{\Delta S} \mathrm{d}^2\vec{s}\cdot\vec{B} = -\Delta t \oint\limits_{L(t)=\delta S(t)} (\mathrm{d}\vec{l}\times\vec{v}\,)\cdot\vec{B} \qquad (5.7)\end{aligned}$$

Employing now Eq.(5.2) yields for Eq.(5.3):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{S(t)} \mathrm{d}^2\vec{s}\cdot\vec{B} &= \int\limits_{S(t)} \mathrm{d}^2\vec{s}\cdot\frac{\partial\vec{B}}{\partial t} - \oint\limits_{L(t)=\delta S(t)} \mathrm{d}\vec{l}\cdot(\vec{v}\times\vec{B}\,)\\ &= -\oint\limits_{L(t)=\delta S(t)} \mathrm{d}\vec{l}\cdot\vec{E} - \oint\limits_{L(t)=\delta S(t)} \mathrm{d}\vec{l}\cdot(\vec{v}\times\vec{B}\,) = -\oint\limits_{L(t)=\delta S(t)} \mathrm{d}\vec{l}\cdot\left[\vec{E}+\vec{v}\times\vec{B}\right]\,. \qquad (5.8)\end{aligned}$$

A few comments are in order here: First of all, the sense of orientation in the line integral matters, it should be counter-clockwise. The $\vec{E}$ and $\vec{B}$ fields are to be taken in the laboratory system. Anyways, the r.h.s. above is the line integral over the Lorentz force and, hence, represents a potential. It can thus be interpreted as voltage. This is in particular true, when $L(t)$ is not just a fictitious line but a thin conductor that can be approximated by a line. In this case, the equation above states then under which conditions a current can result from the fields. Such a current is commonly known as "induced current", the voltage therefore is also induced and often it is denoted by U_{ind}.

$$U_{\text{ind}} = -\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{S(t)} \mathrm{d}^2\vec{s}\cdot\vec{B} = \oint\limits_{L(t)=\delta S(t)} \mathrm{d}\vec{l}\cdot\left[\vec{E}+\vec{v}\times\vec{B}\right]\,. \tag{5.9}$$

Interpretation of Eq.(5.9)

Let's examine this form, Eq.(5.9), of the induction equation a little bit closer in cases, where the line L stands for a thin conductor. If the electric field is independent of time, the closed line integral $\oint \mathrm{d}\vec{l}\cdot\vec{E}$ vanishes. This is because a static field, stemming from a potential, does not do any work along a closed path, which can be seen from $\vec{\nabla}\times\vec{E} = 0$ for static cases. Therefore, any voltage is a consequence of a movement of the conductor in a magnetic field. Then the velocity $\vec{v}$ together with the magnetic field $\vec{B}$ induces a Lorentz force on the mobile charges inside the conductor and thus moves them around.
This can be illustrated by the following example: Consider a rectangular, closed wire with dimensions $a\times b$, where one of the sides can be shifted. The rectangular is in the x-y-plane. Orthogonal to this rectangular, there is a homogenous magnetic field $\vec{B} = B\vec{e}_z$. The flux through the wire at every instant is

$$\int\limits_{\text{rec.}} \mathrm{d}^2\vec{s}\cdot\vec{B} = abB \tag{5.10}$$

and it changes with the rate with which the side is moving. Hence

$$-\frac{\mathrm{d}}{\mathrm{d}t}\int \mathrm{d}^2\vec{s}\cdot\vec{B} = -aB\frac{\mathrm{d}b}{\mathrm{d}t} = \oint\limits_{\delta\text{rec.}} \mathrm{d}\vec{l}\cdot(\vec{v}\times\vec{B}) = avB\,, \tag{5.11}$$

where v is the velocity with which the piece of wire moves, i.e. $v = -\mathrm{d}b/\mathrm{d}t$, cf. Fig. 5.1. The two results coincide and yield the induced potential or electromotive force

$$U_{\text{ind}} = avB\,. \tag{5.12}$$

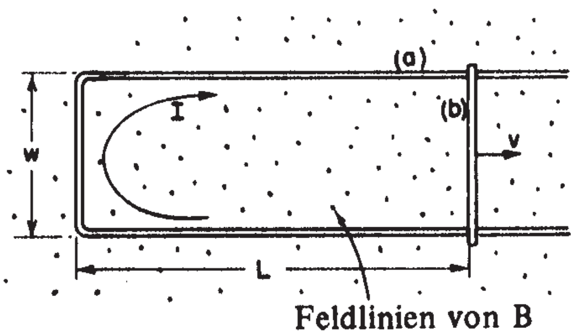

Figure 5.1: A change of magnetic flux, triggered by changing the area of a closed current loop.

If, in contrast, the fields are changing with time, things change a little bit. To underline this consider the case of a fixed conductor configuration but a time-dependent magnetic field $\vec{B} = \vec{B}(t)$. In such a case, the magnetic piece of the electromotive force vanishes; after all, the conductor does not move. But now the fluctuating $\vec{B}$-field yields a non-vanishing rotor of the $\vec{E}$-field through $\vec{\nabla} \times \vec{E} = -\partial\vec{B}/\partial t$. This non-vanishing rotor yields a non-vanishing line integral. Careful inspection yields again the same result for both the l.h.s. and the r.h.s. of Eq.(5.8).

To put this differently: There are no forces acting on the charges inside a conductor that are due to changing magnetic fields. All forces that act on the charges in a conductor are either a consequence of the $\vec{B}$-field (if the conductor or pieces thereof move) or of the $\vec{E}$-field (if the conductor is stationary and the charges within are at rest). If the magnetic field changes with time, it produces an electric field. In both cases, the work per charge is determined by the temporal change of the magnetic flux through the surface. The negative sign in Eq.(5.9) connecting the electromotive force and the change of flux reflects the fact that the induced voltage drives a current which acts against its cause. In other words: The induced current tries to reduce the change of flux. This is Lenz' law.

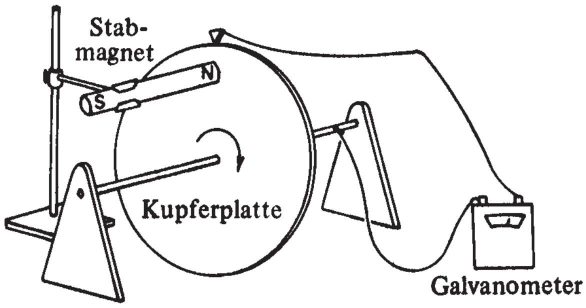

Figure 5.2: When the plane rotates, there is an induced potential due to $\vec{v} \times \vec{B}$, although there is no change of flux.

Exceptions

When applying the induction law in the form of Eq.(5.9) to real, extended conductors, some care has to be taken since in such a case neither the surfaces nor the enclosing lines are well-defined. In such a case the induced currents move inside the conductors in some way that has to be taken care of by a potentially non-homogenous current density $\vec{j}$ inside the conductor with all consequences. This automatically implies that, e.g., the corresponding fields have to be calculated. In such a case details of the measurement, like, for instance, the position of contact points for voltage measurements, start to matter. In the following a situation will be highlighted, where Eq.(5.9) cannot be applied naively.
Fig. 5.2 shows a situation, in which an electromotive force appears although the flux through a circuit does not change. Consider a conducting plane, which, in the presence of a homogeneous magnetic field, can be rotated around its centre with $\vec{\omega} \parallel \vec{B}$. One contact is located at the rotation axis, the other one drags along the plane. The two contacts are connected to a galvanometer, and an electric circuit is thus closed. Of course, the circuit does not change its area, and the flux through it is constant. Nevertheless, since the plane rotates, a $\vec{v} \times \vec{B}$ force acts and produces an electromotive force, measurable through the galvanometer. Hence, although the magnetic flux through the circuit is constant, although the magnetic field does not alternate, and although the circuit does not change its area, there is an electromotive force, due to $\vec{v} \times \vec{B}$ acting in some extended, moving material.

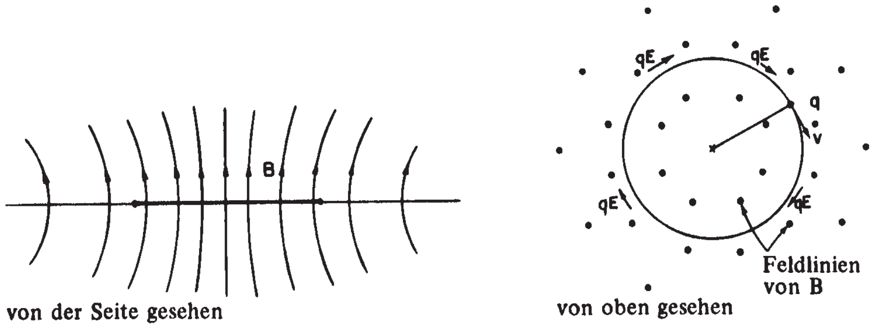

Figure 5.3: Working principle of a betatron: Acceleration of charged particles in an axially symmetric, but alternating magnetic field.

5.1.2 Some simple applications

Betatron: Accelerating particles

As has been seen previously in this section, the electromotive force generated by a alternating magnetic field, does not depend on the presence of conductors. In other words there may be magnetic induction without wires. But of course, one may still talk about the electromotive force along a arbitrary line. Then it is the tangential component of $\vec{E}$ integrated around the closed line. According to Faraday's law, the induction law, this integral equals the change of the magnetic flux through the enclosed surface. As an application let us consider the movement of, say, an electron in a varying magnetic field. This field is orthogonal to a plane, cf. Fig. 5.3, and has a cylindrical symmetry on the plane. In other words its strength on the plane depends on the distance from some axis only. It is generated from some electromagnet, but such details should not matter in the discussion. Imagine now an electron moving in the plane on a circular trajectory with constant radius around the symmetry axis. Because of the varying magnetic field, there is an electric field. Due the way the magnetic field is arranged, the $\vec{E}$ field is tangential to the trajectory of the electron, thus accelerating it. Because of the cylindrical symmetry of the $\vec{B}$ field the strength of the electric field is equal all along the circle and given by

$$\oint \mathrm{d}\vec{l} \cdot \vec{E} - \int \mathrm{d}^2\vec{s} \frac{\partial \vec{B}}{\partial t} = 0 \,. \tag{5.13}$$

When averaging over the magnetic field strength inside such a circular trajectory with radius, say, R, one therefore has

$$2\pi R|\vec{E}| = \frac{\partial}{\partial t}\left(\bar{B}\pi R^2\right) \,, \tag{5.14}$$

where $\bar{B}$ is the average value of the magnetic field strength. Assuming R to be constant yields

$$E = \frac{R}{2}\frac{\mathrm{d}\bar{B}}{\mathrm{d}t} \,. \tag{5.15}$$

The electron feels the magnetic field strength and is thus accelerated, hence (neglecting for a moment the vector character of $\vec{E}$ and $\vec{p}$)

$$qE = \frac{\mathrm{d}p}{\mathrm{d}t} \,. \tag{5.16}$$

Equating the last two equations and integrating over t results in

$$p = p_0 + \frac{qr}{2}\Delta\bar{B} \,. \tag{5.17}$$

Here p_0 is the initial momentum of the electron and $\Delta\bar{B}$ is the subsequent change of the average magnetic field strength. Basically, this is the working principle of the betatron.

Betatrons are machines that allow to accelerate electrons or other charged particles to high energies. To see how this works out it needs to be investigated how the electrons can be kept on a circular trajectory. In fact, as is already known, the magnetic force $q\vec{v} \times \vec{B}$ tends to bend charged particles without further accelerating them. Inside a betatron this transversal force results from the magnetic field and bends the particles. Let's calculate how strong the magnetic field has to be at each point, i.e. at each distance from the axis of symmetry.

In a betatron, the particles move orthogonal w.r.t. the magnetic field, hence the transverse force reads

$$qvB = \frac{\mathrm{d}p_\perp}{\mathrm{d}t} = \omega p = \frac{v}{r}p \,, \tag{5.18}$$

where the fact has been used that for circular trajectories $\mathrm{d}\vec{p}_\perp/\mathrm{d}t = \vec{\omega} \times \vec{p}$ with $\omega \perp p$ and $\omega = v/r$. Equating this yields

$$B = \frac{p}{qr}\,. \tag{5.19}$$

While the betatron is running, the momentum of the particles increases proportional to $\bar{B}$, the average magnetic field inside the trajectory. The magnetic at the trajectory must increase also proportional to p, in order to keep the particles on circles. Comparing Eqs.(5.17) and (5.19) yields

$$\Delta\bar{B} = 2\Delta B\,. \tag{5.20}$$

Hence in order to have a working betatron the average magnetic field inside the trajectory has to increase twice as fast as the magnetic field on the trajectory. If this is the case, the magnetic field on the orbit is just strong enough to keep the particle on it. However, this does not work up to infinite energies. The reason for this is that accelerated charges (even when they are "just" bend on a circle) radiate and thus loose energy, an effect known as synchrotron radiation. At some point the rate of energy loss is as large as the rate of energy gain through the magnetic field, an effect which has a strong dependence on the radius of the orbit.

Alternating current generator

As a next example, consider the alternating current generator. This machine consists of a wire wrapped around a coil with diameter R such that $S = R^2\pi$ that spins in a homogenous magnetic field $\vec{B}$, cf. Fig. 5.5. Since the coil is spinning, the angle θ between the surface it encloses and the magnetic field is changing. The flux through the coil is given by $BS\cos\theta$. If the coil has N windings and rotates with angular velocity ω the induced potential is

$$U_{\text{ind}} = NBS\omega\sin(\omega t) = U_0\sin(\omega t)\,. \tag{5.21}$$

Hence, an alternating current (AC) voltage is induced. This potential drives electrons along the wire, until the electric field resulting from the emerging charge imbalance compensates the force stemming from the induced potential.

When the current is then closed, for instance by adding a resistor, the free charges can follow the potential difference, and no particular electric field

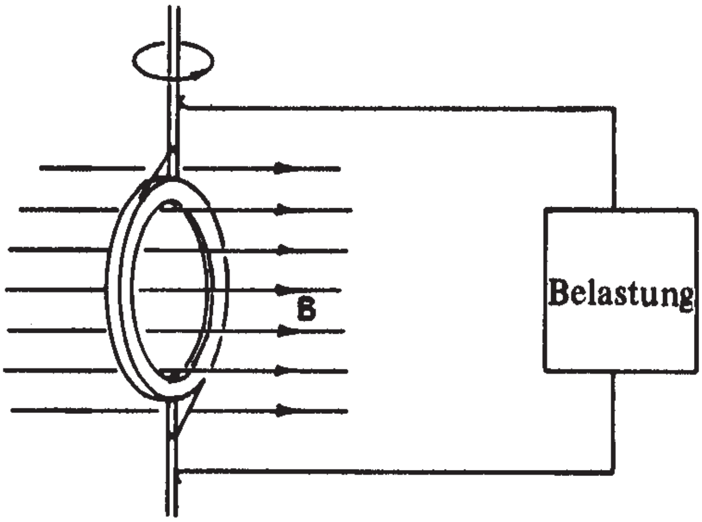

Figure 5.4: Working principle of an AC generator: A coil rotating in a homogenous magnetic field

balances the electromotive force. The generator will "try" to keep the wires on their potential difference. The current is thus proportional to the induced induced and inversely proportional to the total resistance R of the circuit,

$$I = \frac{U_{\text{ind}}}{R} = \frac{U_0}{R} \sin(\omega t)\,. \tag{5.22}$$

An alternating current emerges.
From this analysis the amount of energy delivered by the generator can also be deduced. Each charge inside the wire gains energy with a rate $\vec{F} \cdot \vec{v}$, where $\vec{F}$ is the force acting on the charge, and $\vec{v}$ denotes its velocity. If n is the density of charges per unit length of the wire, the power added to an element ds of the wire reads

$$P = \vec{F} \cdot \vec{v}\, n \mathrm{d}s = nv\vec{F} \cdot \mathrm{d}\vec{s}, \tag{5.23}$$

since the charges always move in parallel with the wire. The total power in the overall circuit then reads

$$P = \oint nv\vec{F} \cdot \mathrm{d}\vec{s}, \tag{5.24}$$

with the integral along the circuit. Remember now that the current I equals qnv and that the electromotive force (induced potential) is defined as the integral of F/q along an electric circuit. Then

$$P = U_{\text{ind}} I = NISB\omega \sin(\omega t)\,. \tag{5.25}$$

If there is a current going through the coil, mechanical forces are acting on it. The torsional moment $\vec{M}$ on the coil equals the vector product of its magnetic moment and the magnetic field strength $\vec{B}$. The magnetic moment of the coil is given by the current through it times the area enclosed, thus

$$|\vec{M}| = NISB \sin(\omega t)\,. \tag{5.26}$$

The rate of mechanical work necessary in order to keep the coil rotating equals the product of angular velocity and torsional moment,

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \vec{\omega} \cdot \vec{M} = \omega NISB \sin(\omega t) = U_{\text{ind}} I\,. \tag{5.27}$$

Comparing Eqs. (5.25) and (5.27) shows that the rate of mechanical work necessary to keep the generator going equals $U_{\text{ind}} I$, which is exactly the rate with which electric energy is distributed along the wire. The total mechanical energy keeping the generator going emerges as electrical energy inside the circuit.

5.1.3 Counter- and self-induction

Counter-induction: transformers

In contrast to the previous example of an AC generator let's discuss now what happens in the case of fixed wire windings but changing magnetic fields going through them. When discussing the generation of magnetic fields through currents, stationary situations have been investigated; as long as the currents are only slowly changing, the emerging magnetic field at each time is pretty much the same as in the stationary case. Hence, in the following, it is assumed that the currents are only slowly changing.

Fig. **??** schematically exhibits the working principle of a transformer: Coil 1 consists of a wire arranged to form a long solenoid, while coil 2 is isolated form it and consists of a few wire windings only. If a current is sent through coil 1, a magnetic field emerges. This magnetic field is also inside coil 2. Changing

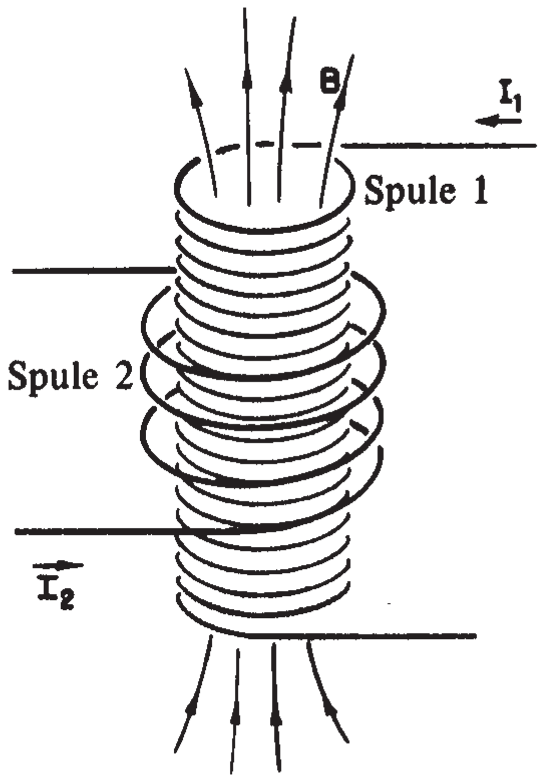

Figure 5.5: Working principle of a transformer: Two coils, a current in one coil results in a magnetic field and, hence, an induced current in the other coil.

the current through coil 1 results in a change of magnetic flux through coil 2, and an electromotive force is induced, which will be calculated now.
From Eq. (4.19) and the corresponding discussion it is known that the magnetic field inside a long solenoid is homogenous, and its value reads

$$B = \frac{1}{\epsilon_0 c^2} \frac{N_1 I_1(t)}{L_1}, \tag{5.28}$$

where N_1 and L_1 are the number of windings on coil 1 and L_1 is its length, whereas $I_1(t)$ is the (time-dependent) current sent through it. The magnetic flux through coil 2 induces an electromotive force, namely

$$U_{\text{ind},\,2} = -N_2 S_1 \frac{\mathrm{d}B}{\mathrm{d}t} = -\frac{N_1 N_2 S_1}{\epsilon_0 c^2 L_1} \frac{\mathrm{d}I_1(t)}{\mathrm{d}t}, \tag{5.29}$$

where N_2 is the number of windings of coil 2 and S_1 is the area of coil 1, through which the magnetic field produces the magnetic flux. The induced potential in coil 2 thus changes with the changing electrical current in coil 1; the constant of proportionality is called the counter-inductivity. It is mainly a geometrical object, commonly denoted by $\mathcal{M}_{21}$, such that

$$U_{\text{ind},\,2} = \mathcal{M}_{21} \frac{\mathrm{d}I_1(t)}{\mathrm{d}t}. \tag{5.30}$$

The reverse calculation, i.e. the determination of $\mathcal{M}_{12}$ would have been more complicated, since coil 2 has less windings resulting in a magnetic field with more pronounced inhomogeneities; anyway, as will be seen later, $\mathcal{M}_{12} = \mathcal{M}_{21}$. Because the field strength is proportional to the current for each type of coil, the factor of proportionality between induced potential and electrical current depends on the geometry of the system only. To calculate the counter-inductivity of two arbitrary coils let's start with the general expression for the electromotive force, Eq. (5.9),

$$U_{\text{ind},1} = -\frac{\mathrm{d}}{\mathrm{d}t} \int\limits_{S(1)} \mathrm{d}^2\vec{s} \cdot \vec{B}, \tag{5.31}$$

where, again, $\vec{B}$ is the magnetic field strength going through the surface S enclosed by the circuit 1. Using now Stokes theorem and connecting the magnetic field strength with the vector potential yields

$$U_{\text{ind},1} = -\frac{\mathrm{d}}{\mathrm{d}t} \oint\limits_{L(1)=\delta S(1)} \mathrm{d}\vec{l} \cdot \vec{A}, \tag{5.32}$$

where $\mathrm{d}\vec{l}$ denotes a line element along circuit 1. Let's assume now that the vector potential along 1 is entirely due to the currents running in circuit 2; then

$$\vec{A}(1) = \frac{1}{4\pi\epsilon_0 c^2} \oint\limits_{L(2)=\delta S(2)} \frac{I_2 \mathrm{d}\vec{l}}{r_{12}}\,, \tag{5.33}$$

where $r_{12} = |\vec{r}_1 - \vec{r}_2|$ is the distance between the line elements $\mathrm{d}\vec{l}$ and the point 1 where the $\vec{A}$ is to be taken, and I_2 is the current running through circuit 2. Hence, the induced potential can be written as a double closed line integral,

$$U_{\mathrm{ind},1} = -\frac{1}{4\pi\epsilon_0 c^2} \frac{\mathrm{d}}{\mathrm{d}t} \oint\limits_{L(1)} \oint\limits_{L(2)} \frac{I_2 \mathrm{d}\vec{l}_1 \mathrm{d}\vec{l}_2}{r_{12}}\,. \tag{5.34}$$

Here, all integrals are over stationary circuits and, thus, over a fixed geometry. The only object which varies with time is the current I_2, which, in turn, does not depend on geometry. This allows to write

$$U_{\mathrm{ind},1} = \left\{ -\frac{1}{4\pi\epsilon_0 c^2} \oint\limits_{L(1)} \oint\limits_{L(2)} \frac{\mathrm{d}\vec{l}_1 \mathrm{d}\vec{l}_2}{r_{12}} \right\} \frac{\mathrm{d}I_2}{\mathrm{d}t} = \mathcal{M}_{12} \frac{\mathrm{d}I_2}{\mathrm{d}t}\,. \tag{5.35}$$

This expression shows that the counter inductivity depends on the geometry only and that it is symmetric under $1 \leftrightarrow 2$, i.e.

$$\mathcal{M}_{12} = \mathcal{M}_{21}\,. \tag{5.36}$$

It depends on something like the "average distance" of the two circuits, such that those pieces of the two coils have the largest contribution that are parallel with respect to another.

Self-induction

So far, only such cases have been considered, where the current is through one of the two coils. In cases where the current is through both coils, the magnetic flux at each point is the sum of the magnetic fluxes induced by the individual coils. This is yet another consequence of the principle of superposition, which of course is also valid for the magnetic field in dynamical situations. However, following this reasoning the induced potential in one coil is due to changes in the current in both coils, schematically,

$$U_{\mathrm{ind},2} = \mathcal{M}_{21}\,\frac{\mathrm{d}I_1}{\mathrm{d}t} + \mathcal{M}_{22}\,\frac{\mathrm{d}I_2}{\mathrm{d}t}\,, \tag{5.37}$$

and, similarly,

$$U_{\mathrm{ind},1} = \mathcal{M}_{12}\,\frac{\mathrm{d}I_2}{\mathrm{d}t} + \mathcal{M}_{11}\,\frac{\mathrm{d}I_1}{\mathrm{d}t}\,. \tag{5.38}$$

The coefficients $\mathcal{M}_{11}$ and $\mathcal{M}_{22}$ are always negative. Usually they are denoted as the self-inductivities $\mathcal{L}$, defined by

$$\mathcal{M}_{11} = -\mathcal{L}_1 \quad \text{and} \quad \mathcal{M}_{22} = -\mathcal{L}_2\,, \tag{5.39}$$

Of course the self-inductivities exist also when only one coil is present: Each coil has its own, specific self-inductivity which only depends on the coil itself. The electromotive force is proportional to the rate of change of the current, for a single coil the signs are by convention arranged such that the induced potential and the current are positive when they go into the same direction. Then

$$U_{\mathrm{ind}} = -\mathcal{L}\frac{\mathrm{d}I}{\mathrm{d}t}\,. \tag{5.40}$$

The negative sign indicates that the induced potential mediated by the self-inductivity works against the current. Hence, a current inside a coil has a certain "inertia". To change the current inside a coil, this "inertia" has to be overcome. When a source of current, like, e.g. a battery, is added, the relation of voltage and current inside such a circuit reads

$$\mathcal{V} = \mathcal{L}\frac{\mathrm{d}I}{\mathrm{d}t}\,. \tag{5.41}$$

This equation has some similarity with Newton's equation connecting force and velocity of a particle. The analogy would read

Particle	Coil
F (force)	$\mathcal{V}$ (potential difference, voltage)
v (velocity)	I (current)
x (shift)	q (charge)
m (mass)	$\mathcal{L}$ (self-inductivity)
$F = m\mathrm{d}v/\mathrm{d}t$	$\mathcal{V} = \mathcal{L}\mathrm{d}I/\mathrm{d}t$
mv (momentum)	$\mathcal{L}I$
$mv^2/2$ (kinetic energy)	$\mathcal{L}I^2/2$ (magnetic energy)

Inductivity

Exploiting the analogies in the table above, one is tempted to identify the product of self-inductivity $\mathcal{L}$ and current, $\mathcal{L}I$, as the "magnetic momentum" of the current inside the circuit. This, however, is not true. The electrical circuit as a whole may be at rest - then it has no momentum. Hence, the analogy is merely such that $\mathcal{L}I$ fulfils the analogous equations as mv in classical mechanics. The same holds also true for $\mathcal{L}I^2/2$ - it fulfils the analogous equations as $mv^2/2$, the kinetic energy. But, as will be seen in the next chapter, $\mathcal{L}I^2/2$ in fact is the magnetic energy. This is due to the fact that work is performed on the inductivity with the rate $\mathcal{V}I$, analogous to Fv in the mechanical case. Therefore, in this case, the quantities are not only analogous in the mathematical sense, they also correspond in terms of physics.
According to Eq. 5.25 and identifying $\mathcal{V} = U_{\mathrm{ind}}$,

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} = U_{\mathrm{ind}}I = -\mathcal{L}\frac{\mathrm{d}I}{\mathrm{d}t}I\,, \tag{5.42}$$

where W is the work done through the inductive force. Integrating over time yields the anticipated result,

$$-W = \frac{1}{2}\mathcal{L}I^2\,. \tag{5.43}$$

This is the energy needed to overcome the electromotive force induced by the self-inductivity of the single coil. Hence the energy stored in an inductive object (e.g. a coil) equals $\mathcal{L}I^2/2$.
Using similar reasoning it can be shown that the total energy U stored in a pair of coils equals

$$U = -W = \frac{1}{2}\mathcal{L}_1 I_1^2 + \frac{1}{2}\mathcal{L}_2 I_2^2 + \mathcal{M}I_1 I_2\,. \tag{5.44}$$

How can this be seen? Well, let's start with both coils at $I = 0$ and let's switch on I_1 in coil 1, leaving $I_2 = 0$. The work done in this coil then is $\mathcal{L}_1 I_1^2/2$. Switching on I_2 in coil 2, the work done in coil 2 is, obviously, $\mathcal{L}_2 I_2^2/2$. In addition, this current also induces an electromotive force in coil 1, its integral times the constant current I_1 yields the work performed on coil 1. [1]

[1]To determine the force between a pair of arbitrary coils with arbitrary currents on

With Eq. 5.44 also a nice inequality can be deduced. It is clear that the energy stored in both coils has to be positive - if this was not the case, i.e. if switching on the currents would not increase the energy of the system, currents in the two coils would spontaneously arise and lead to an energy transfer to the environment. Anyways, Eq. 5.44 can be cast into

$$U = \frac{1}{2}\mathcal{L}_1\left(I_1 + \frac{\mathcal{M}}{\mathcal{L}_1}I_2\right)^2 + \frac{1}{2}\left(\mathcal{L}_2 - \frac{\mathcal{M}}{\mathcal{L}_1}\right)I_2^2\,. \tag{5.47}$$

This expression has to be positive for all values of $I_{1,2}$, including cases where $I_2 = -I_1\mathcal{L}_1/\mathcal{M}$. In such cases the first term of the expression above yields zero, demanding the energy to be positive results in

$$\mathcal{L}_1\mathcal{L}_2 \geq \mathcal{M}^2\,. \tag{5.48}$$

5.2 Electromagnetic waves and radiation

Let's return for a moment to the reasoning of Sec. 2.3. There, a special case has been considered: An infinite "surface" of current was switched on. After some time t, homogenous electric and magnetic fields exist, stretching up to a distance ct away from the source, the current. As before, assume the current layer to be in the y-z plane, where the area charge density J moves into positive y-direction. Then the electric field has just a y-component and the magnetic field is along the z-direction only. The respective components are given by

$$E_y = cB_z = -\frac{J}{2\epsilon_0 c}\,. \tag{5.49}$$

---

them, one may use the principle of virtual displacements, well-known from classical mechanics. The change of energy can be taken from Eq. 5.44. When changing the relative configuration of the two coils, it is clear that the only quantity that changes is $\mathcal{M}$, hence the change of energy reads

$$\Delta U = I_1 I_2 \Delta\mathcal{M} \tag{5.45}$$

and the force acting between the two coils is given by

$$F\Delta x = -\Delta U_{\text{mech.}} = \Delta U\,, \tag{5.46}$$

where $\Delta U_{\text{mech.}}$ is the "mechanical energy" stored in the two coils.

for all positive $x < ct$. For distances x larger than ct the fields vanish. With increasing time therefore a "wave front" travels in positive (and, of course, also negative) x-direction, leaving constant fields behind.
By applying the principle of superposition one may then discuss sequences of events such as: Switch on a current at $t = 0$, triple its strength at $t = t_1$ and switch it off at $t = t_2$. What do the fields look like, for instance at some time $t > t_2$? The answer is simple: For distances larger than ct, the fields vanish, between $x = c(t - t_1)$ and $x = ct$ they have the simple strength, see above, between $x = c(t - t_2)$ and $x = c(t - t_1)$ they have the triple strength and for $x < c(t - t_2)$ they vanish. This "double box" configuration travels with speed of light to larger and larger distances x away from the moving current plane. As seen before, this propagation continues even after all currents ceased to exist. A more complete mathematical formulation of this effect can be found by writing that the electric and magnetic fields are proportional to the current at the source, but not at time t but at time $t - x/c$:

$$E_y(t) = cB_z(t) = -\frac{J(t - x/c)}{2\epsilon_0 c} \,. \tag{5.50}$$

5.2.1 Homogenous solutions of the wave equations

Generalities (one-dimensional)

Let us discuss now the general form of how electric and magnetic fields behave in large distances from any sources, charges and currents. Close to these sources, i.e. in those cases where the sources had not enough time to change during the propagation of the fields, the fields can be determined like in the static cases, presented in Chapters 3 and 4. At very large distances, however, where the propagation time does play a role, the behaviour changes fundamentally. The basic question to be answered first is: Which type of fields can exist at all in those areas where charge and current density ρ and $\vec{j}$ vanish? In other words, which solutions are allowed for

$$\begin{aligned} \nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} &= 0 \\ \nabla^2\vec{A} - \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} &= 0\,, \end{aligned} \tag{5.51}$$

cf. Eq. (2.45), where the sources have been set to zero and where ϕ and $\vec{A}$ are the scalar and the vector potential, respectively. Obviously, the equations for the three components of the vector potential have exactly the same form as the equation for the scalar potential, in general, equations of this type, i.e.

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} \quad = \quad 0 \tag{5.52}$$

are called wave equation in three dimensions, since usually ψ depends on all three space dimensions x, y, and z.
To study the behaviour of the fields, the connections

$$\vec{B} = \vec{\nabla}\times\vec{A} \quad \text{and} \quad \vec{E} = -\vec{\nabla}\phi - \frac{\partial\vec{A}}{\partial t} \tag{5.53}$$

of the fields with the potentials have to be included. This, however, is pretty straight-forward. Let's turn to the case of the magnetic field first. Since $\vec{B} = \vec{\nabla}\times\vec{A}$, a differential equation for $\vec{B}$ emerges by taking the rotor of the second of the equations 5.51 above,

$$\begin{aligned}\vec{\nabla}\times\left[\nabla^2\vec{A} - \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2}\right] &= \nabla^2\left[\vec{\nabla}\times\vec{A}\right] - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left[\vec{\nabla}\times\vec{A}\right] \\ &= \nabla^2\vec{B} - \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = 0\,,\end{aligned} \tag{5.54}$$

since the Laplace operator and $\partial^2/\partial t^2$ are scalar operators. Hence, each component of the magnetic field also has to satisfy the wave equation. In a similar fashion, the same can be found for the electric field.
However, before trying to construct the most general solution of this set of equations, it is worthwhile to consider the wave equation in one dimension. A possible solution for

$$\frac{\partial^2\psi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} \quad = \quad 0 \tag{5.55}$$

is a function of the form

$$\psi(x,\,t) = f(x-ct)\,. \tag{5.56}$$

This is a function of the only variable $(x-ct)$. Hence, the function $f(x-ct)$ describes a "pattern" in x moving with velocity c in positive x-direction. If,

for instance, f has a maximum where its argument vanishes, this maximum is at $x = 0$ for $t = 0$ and at $x = cT$ for $t = T$. For some purposes it is more convenient to define the solutions as functions of $t - x/c$; these two ways of writing $\psi(x,\, t)$ as either $\psi(x,\, t) = f(x - ct)$ or $\psi(x,\, t) = F(t - x/c)$ are synonymous.
Anyway, in addition to the solution of ψ being a function of $x - ct$, there is also another solution, namely

$$\psi(x,\, t) = g(x + ct)\,. \tag{5.57}$$

This is not a big surprise, since the wave equation contains c^2 only, which is insensitive to the sign of or in front of c. In fact, the most general solution of the wave equation is

$$\psi(x,\, t) = f(x - ct) + g(x + ct)\,. \tag{5.58}$$

The first piece describes a wave travelling in positive x-direction, whereas the second piece describes a wave travelling in negative x-direction, the general solution is a superposition of both.

Properties of electromagnetic waves

Being equipped now with the solution for the wave equation in one dimension, one is tempted to construct solutions for the electromagnetic waves. But instead, let us start with the Maxwell equations in the vacuum to see how closely they are related with the waves. In vacuum, Maxwell's equations read

$$\begin{aligned} \vec{\nabla}\cdot\vec{E} &= 0 & \qquad \vec{\nabla}\cdot\vec{B} &= 0 \\ \vec{\nabla}\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t} & \qquad c^2\vec{\nabla}\times\vec{B} &= \frac{\partial\vec{E}}{\partial t}\,, \end{aligned} \tag{5.59}$$

cf. Eq. 2.14, where we have set ρ and $\vec{j}$ to zero.
Consider now a "wave front" travelling in x direction, such that the fields are independent of y and z. The first equation then boils down to

$$\frac{\partial E_x}{\partial x} = 0\,, \tag{5.60}$$

equivalent to the statement that E_x does not depend on the position. Adding in the last equation results in the statement that E_x does not depend on

time either, it is just a constant number. Such a field could, for instance, be produced by two large charged capacitor planes at a large distance. Such a static case, however, is of no interest at the moment, hence one may set $E_x = 0$.
An immediate consequence of this finding is that during the propagation of planar electromagnetic waves in arbitrary direction the electric field is always perpendicular w.r.t. the direction of propagation. It is transverse. This field can always be separated in two components, here the y and the z component. Assume now that only one of these components matters, the other one is taken to be zero. Then the only component of $\vec{\nabla} \times \vec{E}$ different from zero is in z-direction and reads

$$\vec{\nabla} \times \vec{E}\bigg|_z = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = \frac{\partial E_y}{\partial x} \,. \tag{5.61}$$

Hence,

$$\frac{\partial B_z}{\partial t} = -\frac{\partial E_y}{\partial x} \tag{5.62}$$

and the other two components of $\vec{B}$ are just constants in accordance to some magnetostatic solutions. It is just like someone left some permanent magnets leading to some stationary magnetic fields, which will also be neglected in the present discussion. It should be noted that this reasoning is in perfect agreement to what has been done for the electric field: Again, also in the magnetic case, the divergence of the field is zero, due to the assumption that the fields depend on x only, B_x has to be a constant, taken to be zero. This results in the finding that both the electric and the magnetic field are transverse, i.e. perpendicular w.r.t. the direction of wave propagation. In addition, they are perpendicular to each other.
When releasing the constraint that the fields have an y- or z-component only, they can be written as

$$\begin{aligned}
\vec{E} &= \begin{pmatrix} 0 \\ E_y \\ E_z \end{pmatrix} = \begin{pmatrix} 0 \\ f(x-ct) + \bar{f}(x+ct) \\ g(x-ct) + \bar{g}(x+ct) \end{pmatrix} \\
c\vec{B} &= \begin{pmatrix} 0 \\ cB_y \\ cB_z \end{pmatrix} = \begin{pmatrix} 0 \\ f(x-ct) - \bar{f}(x+ct) \\ -g(x-ct) + \bar{g}(x+ct) \end{pmatrix} \,.
\end{aligned} \tag{5.63}$$

Such waves, travelling in x-direction, have an $\vec{E}$-vector, that is not constant, but rather rotates in an arbitrary manner in the y-z-plane. At each instant, the $\vec{B}$ field is perpendicular to the electric field an the direction of propagation. In three dimensions, planar waves can be described as a superposition of the one-dimensional planar waves discussed so far. If the waves are travelling in one direction only, there is a simple rule for the relative orientation of $\vec{E}$ and $\vec{B}$. This rule states that $\vec{E} \times \vec{B}$ points into the direction of wave propagation. Later on it will be seen that this Poynting vector has an additional meaning: It is the vector describing the energy flux of an electromagnetic wave.

Spherical waves

So far only those solutions of the wave equation have been discussed that correspond to planar waves. In some cases, however, it is more convenient to describe waves in another mathematical form. One of these forms are spherical waves. They correspond to waves spreading out from a common centre (or running into a common centre). A simple example for such waves in two-dimensions are those waves that emerge when throwing a stone into water. In the following therefore, spherically symmetric solutions for the wave equation will be investigated. The resulting solutions will then depend only on the radial distance r to the centre, which will be at the origin. To this end, the wave equation has to be cast into polar coordinates; for a scalar field $\psi = \psi(r,\, t)$ it reads

$$\begin{aligned} \nabla^2\psi(r,\, t) - \frac{1}{c^2}\frac{\partial^2\psi(r,\, t)}{\partial t^2} &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi(r,\, t)}{\partial r}\right) - \frac{1}{c^2}\frac{\partial^2\psi(r,\, t)}{\partial t^2} \\ &= = \frac{1}{r}\frac{\partial^2}{\partial r^2}\left(r\psi(r,\, t)\right) - \frac{1}{c^2}\frac{\partial^2\psi(r,\, t)}{\partial t^2} = 0\,. \end{aligned} \tag{5.64}$$

This looks considerably more complicated than the planar case. Multiplying it with r, however, results in

$$\frac{\partial^2}{\partial r^2}\left[r\psi(r,\, t)\right] - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}[r\psi(r,\, t)] = 0\,, \tag{5.65}$$

and it becomes clear that this time the function $[r\psi]$ has to satisfy the one-dimensional wave equation. Thus, any general solution of Eq. (5.65) should look like

$$r\psi(r,\, t) = f(t - r/c) + \bar{f}(t + r/c) \tag{5.66}$$

leading to

$$\psi(r,\, t) = \frac{f(t - r/c) + \bar{f}(t + r/c)}{r}\,. \tag{5.67}$$

This is the most general form of a spherical wave spreading out or falling into the origin in a spherically symmetric fashion. The denominator $1/r$ is the clear and visible difference between spherical and planar waves: Whereas the latter remain invariant both in shape and amplitude when propagating, the former become smaller (or larger) amplitude-wise when propagating away (or towards) the origin[2]. It should be noted that in order to fully fix the actual form of the functions f and $\bar{f}$, it is necessary to specify suitable boundary values, like, e.g. the absolute value and first derivative of the wave at a given position and time.
A further comment is in order here: The solution above becomes infinite at the origin of the wave, i.e. for $r \to 0$. This is not so nice; as a solution rather a finite and smooth wave would be desirable. The solution found here instead represents a situation with a source actually present. Accidentally the "wrong" solution has been found, with some matter, the source, at the origin. The free wave equation has not been solved everywhere, the error happened because some steps in the derivation are illegal for $r = 0$. This type of error, however, may happen already in the static case. Just try to solve the equation for the electrostatic potential in vacuum, $\nabla^2\phi = 0$. Of course, this should yield zero, since there are no charges around. But how about spherically symmetric solutions $\phi = hi(r)$ of this equation? Employing the form of the Laplace operator found previously, one has

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}\left[r\phi(r)\right] = 0 \quad \Leftrightarrow \quad \frac{\partial^2}{\partial r^2}\left[r\phi(r)\right] = 0\,. \tag{5.68}$$

[2]This effect is easy to understand with some simple physical reasoning: As will be seen later, the energy density of a wave is proportional to the square of the amplitude. When the wave spreads out (the f-piece, while $\bar{f}$ is neglected here), the energy spreads over larger and larger areas, proportional to the radial distance squared from the source. If the total energy is to be conserved, the energy density therefore has to decrease like $1/r^2$, hence the amplitude has to diminish like $1/r$. Thus, this factor $1/r$ keeps the spherical wave physically reasonable.

Integrating over r yields, with an arbitrary constant a,

$$\frac{\partial}{\partial r}\left[r\phi(r)\right] = a\,, \tag{5.69}$$

integrating once more results in

$$\left[r\phi(r)\right] = ar + b\,, \tag{5.70}$$

where b is another constant of integration. Hence the solution for the potential is

$$\phi = a + \frac{b}{r}\,. \tag{5.71}$$

This is clearly wrong. If no charges are present, the potential should be a constant, but there is another term in the solution here, representing the potential of a point charge. Although a solution in the vacuum was to be constructed, this extra term kicked in. This is quite similar to what has been found also for the wave. However, this is not too bad (in fact, it'll be beneficial): After all, there would be no spherically symmetric wave spreading out from the origin, if there was no source for it!

5.2.2 Adding sources

Let us now turn to solutions for the Maxwell equations when charges and currents are present. From chapter 2 it is known that the fields can be obtained from the potentials ϕ and $\vec{A}$, which in turn have to satisfy Eqs. (2.45)

$$\begin{aligned} \nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} &= -\frac{\rho}{\epsilon_0} \\ \nabla^2\vec{A} - \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} &= -\frac{\vec{j}}{\epsilon_0 c^2} \end{aligned} \tag{5.72}$$

and the condition

$$\vec{\nabla}\cdot\vec{A} = -\frac{1}{c^2}\frac{\partial\phi}{\partial t}\,. \tag{5.73}$$

In other words, a solution for the equation

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -s \tag{5.74}$$

has to be found, where ψ can be either ϕ or any of the components of $\vec{A}$, and the source term s is either ρ/ϵ_0 or the corresponding component of $\vec{j}/\epsilon_0 c^2$, respectively.

Point charge

To obtain such a general solution, the principle of superposition can be employed, once again. In this context it states that a solution for some complicated source distribution $s(\vec{r},\, t)$ can be obtained as a superposition of solutions for a source term of the form $s = s(t)\delta(\vec{r})$, supplemented with some shifts in space. In other words, the solution for a complicated charge distribution can be obtained as a superposition of solutions for suitable point charges.
As noted before, the solution Eq. (5.67) for spherically symmetric waves, implies some charge or other source at the origin. This looks promising! The question then is, which kind of charge would produce such a wave. To answer this question, re-consider this solution and check what happens for small r. To keep things simple, in the following it will be assumed that only waves produced at the origin and travelling "outwards" will be investigated then

$$\psi(r,\, t) = \frac{f(t - r/c)}{r} \quad \overset{r\to 0}{\longrightarrow} \quad \frac{f(t)}{r}\,, \tag{5.75}$$

as long as f is sufficiently smooth. Hence, ψ resembles a Coulomb field for a charge situated at the origin that changes with time. For a small amount of charge Q, constrained to a very small volume V around the origin, it is known that

$$\phi = \frac{Q}{4\pi\epsilon_0 r} \tag{5.76}$$

satisfies the equation

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0} \tag{5.77}$$

if

$$Q = \int\limits_V \mathrm{d}^3V\rho\,. \tag{5.78}$$

This immediately implies that ψ satisfies

$$\nabla^2\psi = -s \quad \text{for} \quad r \to 0\,, \tag{5.79}$$

connecting f with s through

$$f = \frac{1}{4\pi}\int\limits_V \mathrm{d}^3V s\,. \tag{5.80}$$

This is true as long as the source s is also distributed over a very tiny volume around the origin. The only difference to the static case is that s now is a function of time. The trick now is that if ψ satisfies Eq. (5.79) for small distances from the origin, it also satisfies Eq. (5.75) in the same limit. Coming close to the origin, the $1/r$-dependence of ψ results in a huge spatial derivative, such that the temporal derivative can be neglected. In this limit of $r \to 0$ the two equations just become equivalent.
The sole effect of the temporal derivative in the equation is that it introduces a retardation $s(t) \to s(t - r/c)$ into the solution. Hence, the solution for ψ reads

$$\psi(r,\,t) = \frac{1}{4\pi r}\int\limits_V \mathrm{d}^3V s(t - r/c) = \frac{1}{4\pi r}S(t - r/c)\,. \tag{5.81}$$

Here s and S relate to each other like ρ/ϵ_0 and Q in the static case.

General solution

Having found the solution for a "point" source, it is quite straightforward to construct the more general solution for an extended source. This is simple: Each extended source $s = s(\vec{r},\,t)$ can be understood as a superposition of many "point"-sources, one for each volume element d^3V. This works perfectly well, since Eq. (5.67) is linear in ψ. Using the result from the previous discussion, in particular Eq. (5.81), it is clear that

$$d\psi(\vec{r}_1,\,t) = \frac{1}{4\pi|\vec{r}_1 - \vec{r}_2|}\mathrm{d}^3V_2 s(\vec{r}_2,\,t - |\vec{r}_1 - \vec{r}_2|/c) \tag{5.82}$$

is the contribution of the field at position $\vec{r}_1$, that stems from the source element $\mathrm{d}^3V_2 s(\vec{r}_2,\,t)$ at position $\vec{r}_2$. This translates into the fact that the

field ψ at position $\vec{r}_1$ and time t is the sum of all spherical waves that left the source elements at some position $\vec{r}_2$ at time $t - |\vec{r}_1 - \vec{r}_2|/c$.
From this finding it is straightforward to find expressions for the solutions of Eqs. (5.72). These solutions are

$$\begin{aligned}\phi(\vec{r}_1,\,t) &= \int \mathrm{d}^3V_2 \frac{\rho(\vec{r}_2,\,t - |\vec{r}_1 - \vec{r}_2|/c)}{4\pi\epsilon_0|\vec{r}_1 - \vec{r}_2|} \\ \vec{A}(\vec{r}_1,\,t) &= \int \mathrm{d}^3V_2 \frac{\vec{j}(\vec{r}_2,\,t - |\vec{r}_1 - \vec{r}_2|/c)}{4\pi\epsilon_0 c^2|\vec{r}_1 - \vec{r}_2|}\,.\end{aligned} \tag{5.83}$$

From the potentials the field strengths $\vec{E}$ and $\vec{B}$ can be obtained through the gradient or rotor, respectively. This is the solution of Maxwell's equations: If charge and current densities are known at arbitrary times, the equations above connect them with the potentials and, finally, with the fields.

5.2.3 Moving charges

In what follows the fields emitted by moving charges will be investigated. The simple-most case is the case of a dipole that oscillates non-relativistically. This is synonymous to the fields emitted by a distribution of charges moving within a small volume that is far away.

Oscillating dipoles: Some first considerations

To keep things simple let's assume that the light is emitted by a moving charge q in the vicinity of a fixed charge $-q$. If the distance between the two charges is $\vec{d}$, the associated dipole moment is given by $\vec{p} = q\vec{d}$, cf. Eq. (3.45). Since the charges move relatively w.r.t. each other, the dipole moment is a function of time, $\vec{p} = \vec{p}(t)$. As long as the fields close enough to the dipole are considered, retardation effects do not matter (in other words, terms of the form x/c can be neglected in arguments of the form $t - x/c$) and the field looks like the stationary dipole field, taken for $\vec{p}(t)$. At large distances, however, there will be terms proportional to $1/r$ that depend on the acceleration of charges orthogonal to the direction towards the dipole.
The starting point of the derivation is the second of Eqs. (5.83), the one for the vector potential. Let's assume the moving charge is in fact distributed over a little volume with density ρ, moving with velocity $\vec{v}$. To leave things easy, $\vec{v} = v\vec{e}_z$ with non-relativistic velocities v. The current density then

reads $\vec{j} = \rho\vec{v}$. Then the determination of the vector potential at position $\vec{r}_1$ at time t boils down to solving

$$\vec{A}(\vec{r}_1,\, t) = \int \mathrm{d}^3V_2 \frac{\vec{j}(\vec{r}_2,\, t - r_{12}/c)}{4\pi\epsilon_0 c^2 r_{12}}\,, \tag{5.84}$$

where $r_{12} = |\vec{r}_1 - \vec{r}_2|$. To make explicit that the distance r_{12} between the actual position $\vec{r}_2$ inside the distribution and $\vec{r}_1$ does not change significantly, if the distribution is well-located, $\vec{r}_{12}$ can be replaced by $\vec{r}$, the distance of $\vec{r}_1$ from the centre of the distribution. Then

$$\vec{A}(\vec{r}_1,\, t) = \int \mathrm{d}^3V_2 \frac{\vec{v}\rho(\vec{r},\, t - r/c)}{4\pi\epsilon_0 c^2 r}\,. \tag{5.85}$$

Since the charge distribution has the same velocity all over it, the integral is just $q\vec{v}/r$. But this is nothing but the temporal derivative of the dipole moment, divided by r. Hence

$$\vec{A}(\vec{r}_1,\, t) = \frac{1}{4\pi\epsilon_0 c^2} \frac{\dot{\vec{p}}(t - r/c)}{r}\,. \tag{5.86}$$

In other words, the current of a changing dipole just produces a vector potential in the form of a spherical wave, where the source is, up to constants, given by the first temporal derivative of the dipole moment. From the vector potential the magnetic field can be obtained. Since, according to the assumption, $\vec{p}$ and its changes point into the z-direction, $\vec{A}$ has just a z-component. The rotor has just two non-vanishing terms, namely $B_x = \partial A_z/\partial y$ and $B_y = -\partial A_z/\partial x$. Let's take a closer look at the former term. It reads

$$\begin{aligned} B_x = \frac{\partial A_z}{\partial y} &= \frac{1}{4\pi\epsilon_0 c^2} \frac{\partial}{\partial y}\left(\frac{\dot{\vec{p}}(t - r/c)}{r}\right) \\ &= \frac{1}{4\pi\epsilon_0 c^2}\left[\dot{\vec{p}}(t - r/c)\frac{\partial}{\partial y}\left(\frac{1}{r}\right) + \left(\frac{1}{r}\right)\frac{\partial\dot{\vec{p}}(t - r/c)}{\partial y}\right] \\ &= \frac{1}{4\pi\epsilon_0 c^2}\left[\dot{\vec{p}}\left(\frac{-y}{r^3}\right) - \left(\frac{1}{r}\right)\ddot{\vec{p}}\frac{y}{r}\right] = -\frac{1}{4\pi\epsilon_0 c^2}\frac{y}{r^2}\left[\frac{\dot{\vec{p}}}{r} + \frac{\ddot{\vec{p}}}{c}\right]\,. \end{aligned} \tag{5.87}$$

The first term is the field component of a corresponding static dipole, decreasing like $1/r^2$, and it is the second term that is responsible for the radiation.

It describes a field decreasing like $1/r$, depending on the acceleration of the charge.
Written in compact form, the magnetic field reads

$$\vec{B} = \frac{1}{4\pi\epsilon_0 c^2}\left[\frac{\left(\dot{\vec{p}} + \frac{r}{c}\ddot{\vec{p}}\right)_{t-r/c} \times \vec{r}}{r^3}\right] . \tag{5.88}$$

Closer inspection reveals that for large r, it is the $\ddot{p}$ term that matters. The direction of the field is determined by $\ddot{\vec{p}} \times \vec{r}$, which is perpendicular to both the accelerated charge and $\vec{r}$.
But, beyond formal considerations, how does this term emerge? To see this, consider Eq. (5.86). The vector potential decreases with $1/r$ and, up to some retardation effects, it looks exactly like a Coulomb potential with the charge replaced by $\dot{p}$. Naively, one would expect just a $1/r^2$ field, but obviously this is not the full truth. The reason for this can be understood by considering a case where the dipole bounces back and forth in a harmonic oscillation,

$$\vec{p} = p_0 e_z \sin\omega t\, . \tag{5.89}$$

Then

$$A_z = \frac{1}{4\pi\epsilon_0 c^2}\,\frac{\omega p_0 \cos[\omega(t - r/c)]}{r} . \tag{5.90}$$

When analysing A_z as a function of r at a given time, then it becomes apparent that the maximal value of the amplitudes decreases with $1/r$. Inside an envelope $1/r$ there exists an oscillation in space. This is due to the fact that in the process of continuous emission, the temporal changes of the dipole are translated into spatial changes at a given time (like in a snapshot). When calculating the spatial derivatives, it is clear that the maximal values are proportional to the envelope; this explains the functional form of the radiation term.

Oscillating dipoles: The magnetic field nearby

Let's turn our attention now to the magnetic field closer to the dipole. In chapter 4 the law of Biot and Savart has been discussed, that deals with the magnetic field of a given current element. This law states that a current element $\vec{j}\mathrm{d}^3V$ yields the magnetic field

$$\mathrm{d}\vec{B} = \frac{1}{4\pi\epsilon_0 c^2}\frac{\vec{j}\times\vec{r}}{r^3}\mathrm{d}^3V\,, \tag{5.91}$$

in distance r away from it. This looks pretty much like the first term in Eq. (5.88), especially after realising that $\dot{\vec{p}}$ is just a current. A difference, however, still exists: The current in Eq. (5.88) has to be taken at some time $t - r/c$, a retardation that does not appear in the instantaneous equation above. Nevertheless, the equation above is still more exact that anticipated now for small r, due to the fact that the second term of Eq. (5.88) tends to cancel the effect of retardation. Both together yield, at small distances, a result quite similar to the law of Biot and Savart.
To see this, $\dot{\vec{p}}$ may be Taylor-expanded around t for small r,

$$\dot{\vec{p}}(t - r/c) = \dot{\vec{p}}(t) - \frac{r}{c}\ddot{\vec{p}}(t) + \mathcal{O}\left(\frac{r^2}{c^2}\right)\,, \tag{5.92}$$

and up to the same order in r/c,

$$\ddot{\vec{p}}(t) = \ddot{\vec{p}}(t - r/c)\,. \tag{5.93}$$

When summing the first and the second term in Eq. (5.88) thus, at small distances, only the un-retarded current $\dot{p}$ remains, plus some terms of order r^2/c^2. Hence, Eq. (5.88) yields a result that is very similar to the one obtained in instantaneous theory - in fact, the result is more similar than those obtained from instantaneous theory plus a retardation. The stationary formulae are quite exact for small distances, much more precise than what could be anticipated from the start. It is only for large distances, where the retardation effects become important.

Oscillating dipoles: The electric field

To gain the electric field for all distances, a solution for the electrostatic potential has to be found. When the vector potential was deduced, the approximation was made to neglect the effect of the spatial distribution on the distances and to replace all r_{12} by r. This approximation would lead to wrong results in the determination of ϕ, since this approximation would lead to $1/r$ times the integral over a charge density. This is not precise enough.

But instead of calculating the potential directly at higher orders the scalar potential can be taken from the constraint Eq. (5.73),

$$\vec{\nabla} \cdot \vec{A} = -\frac{1}{c^2}\frac{\partial \phi}{\partial t}\,, \tag{5.94}$$

by employing the solution for the vector potential that has been found already. For the case under consideration the divergence of $\vec{A}$ is just

$$\vec{\nabla} \cdot \vec{A} = \frac{\partial A_z}{\partial z} = -\frac{1}{4\pi\epsilon_0 c^2}\left[\frac{z\dot{p}(t-r/c)}{r^3} + \frac{z\ddot{p}(t-r/c)}{cr^2}\right]. \tag{5.95}$$

Written in vectorial form this is

$$\vec{\nabla} \cdot \vec{A} = -\frac{1}{4\pi\epsilon_0 c^2}\left[\frac{\left(\dot{\vec{p}} + \frac{r}{c}\ddot{\vec{p}}\right)_{t-r/c} \cdot \vec{r}}{r^3}\right]. \tag{5.96}$$

This translates into

$$\frac{\partial \phi}{\partial t} = \frac{1}{4\pi\epsilon_0}\left[\frac{\left(\dot{\vec{p}} + \frac{r}{c}\ddot{\vec{p}}\right)_{t-r/c} \cdot \vec{r}}{r^3}\right], \tag{5.97}$$

integrating over t just reduces the number of temporal derivatives by one, hence

$$\phi(\vec{r},\, t) = \frac{1}{4\pi\epsilon_0}\left[\frac{\left(\vec{p} + \frac{r}{c}\dot{\vec{p}}\right)_{t-r/c} \cdot \vec{r}}{r^3}\right]. \tag{5.98}$$

Any integration constant neglected here would just correspond to another static field, which will not be taken into account here.
This allows to determine the electric field. From

$$\vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \tag{5.99}$$

one finds after taking some derivatives and taking properly into account the retardation effects

$$\vec{E}(\vec{r},\, t) = -\frac{1}{4\pi\epsilon_0 r^3}\left[-\vec{p}^* - 3\frac{(\vec{p}^* \cdot \vec{r})\vec{r}}{r^2} + \frac{(\ddot{\vec{p}} \times \vec{r}) \times \vec{r}}{c^2}\right], \tag{5.100}$$

where

$$\vec{p}^* = \vec{p} + \frac{r}{c}\dot{\vec{p}}, \tag{5.101}$$

and where all dipole momenta or there derivatives are taken at $t - r/c$. This looks complicated but it is not too complicated to understand: $\vec{p}^*$ is something like the retarded dipole moment, corrected such that the two expressions involving $\vec{p}^*$ just boil down to the static dipole field at small distances. If in contrast, $\vec{r}$ is large, i.e. at large distances, it is the last term that dominates, like in the case of the magnetic field. Again, at large distances, the field is proportional to the acceleration of the charges and perpendicular w.r.t. it. The field lines and their evolution while the dipole is oscillating are exhibited in Fig. 5.6 [3].

Potentials of a moving charge: Solution of Lienard and Wiechert

The reasoning so far was constrained to the case of non-relativistic velocities; the discussion now turns to the more general case of a point charge moving in arbitrary manner - even relativistically. To start with, let's try to calculate the scalar potential at position $\vec{r}_1$ of a point charge located at $\vec{r}_2$. Again the point charge will be "smeared out" in a volume d^3V_2 around $\vec{r}_2$, then ϕ reads

$$\phi(r_1,\, t) = \frac{1}{4\pi\epsilon_0}\int \mathrm{d}^3V_2\,\frac{\rho(\vec{r}_2,\, t - r_{12}/c)}{r_{12}}\,. \tag{5.102}$$

It looks tempting to do the integral over the charge distribution and just to take the average distance $\bar{r}_{12}$ of the distribution from $\vec{r}_1$, doesn't it? Something like

$$\phi(r_1,\, t) = \frac{1}{4\pi\epsilon_0}\frac{q}{\bar{r}_{12}} \quad \text{(wrong!)} \tag{5.103}$$

would emerge. But this is wrong! The correct answer is

$$\phi(r_1,\, t) = \frac{1}{4\pi\epsilon_0}\frac{q}{\bar{r}_{12}}\,\frac{1}{1 - \vec{v}\cdot\vec{n}_{12}/c}\,, \tag{5.104}$$

where $\vec{n}_{12} = \vec{r}_{12}/r_{12}$ projects out the component of the charge's velocity parallel to $\vec{r}_{12}$. In the following it will be shown why the result above is the correct one.

[3] For a nice animation, see, e.g.,
`http://physics.usask.ca/ hirose/ep225/radiation.htm`

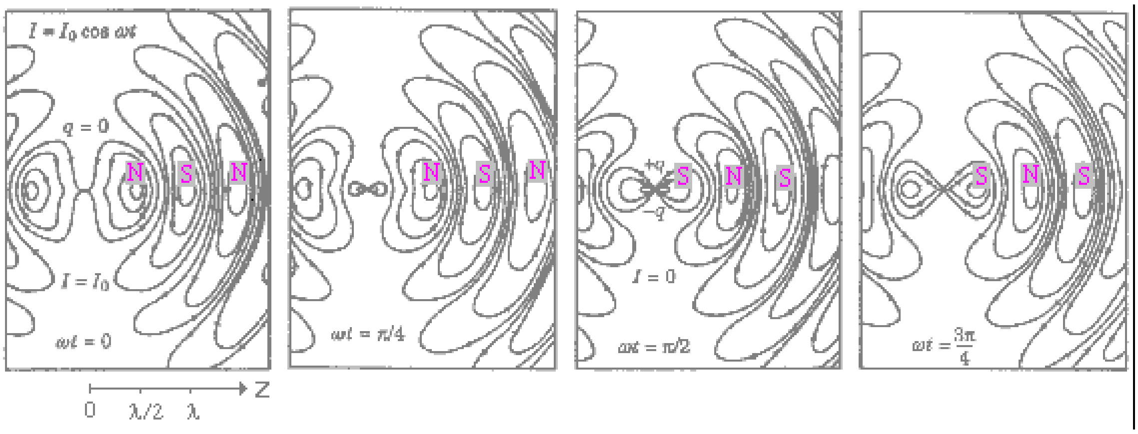

Figure 5.6: Schematic view of the evolution of fields during the oscillation of a dipole. An oscillating dipole (oscillation ω) radiates an electric field as well as a magnetic field as shown in the figure. Magnetic polarities associated with the wave-fronts are also depicted. At a distance $z > \lambda/2$ the magnetic/electric field detaches itself from the dipole and propagates outward. However for $z < \lambda/2$ the magnetic field is still associated with the dipole. If there was an external magnetic field, at $z < \lambda/2$ the dipole will experience a force against the external field. If however an external magnetic field exists only for $z > \lambda/2$ the dipole will not experience any force against the external field, but there will be a force between a particular wave-front and the external field (Picture from `http://www.dip.ee.uct.ac.za/ mdevill/emthrust.html`).

To this end, the point charge will be replaced by a small charged box of dimension a, moving towards $\vec{r}_1$. The distance of the centre of the box and $\vec{r}_1$, $\vec{r}_{12}$ is assumed to be much larger than a. In order to evaluate the integral over the volume element d^3V_2, the box is divided into thin slices, orthogonal to the direction of movement. Then the integral can be replaced by the sum over the slices i,

$$\sum_i \frac{\rho_i \Delta V_i}{r_i}\,, \tag{5.105}$$

where ΔV_i is the volume of the ith slice, ρ_i is the charge density inside the slice and r_i is the corresponding distance to $\vec{r}_1$. The latter two, charge density and distance are taken at the time $t_i = t - r_i/c$.
The width of the volume elements ΔV_i, taken to be a constant w, is assumed to be very small compared to a, by construction each ΔV_i has a volume wa^2. What is the charge inside? Well, for each volume element ΔV_i the charge density ρ is to be taken at $t_i = t - r_i/c$, but, since the charge is moving, its position is different for each volume element. To see this, consider Fig. 5.7. There, more slices have been drawn in part (a) than there are slices in the little box. Start with the slice 1, located at the far edge of the box to the left. The far edge of the charge, ρ_1 is in ΔV_1 at the time $t_1 = t - r_1/c$. But when evaluating the next slice of charge, ρ_2 at ΔV_2, it has to be taken at a somewhat later time $t_2 = t - r_2/c$ (it's later, because $r_2 < r_1$). This allows to calculate the sum. Each slice of the volume overlapping some charge contributes with $wa^2\rho$, where ρ is the uniform charge density inside the box. The distance r_i in the denominator can be approximated by r_{12}, the average distance of the charge; taken together, the sum is

$$\sum_{i=2}^{N} \frac{wa^2\rho}{r_{12}} = \frac{Nwa^2\rho}{r_{12}} = \frac{a^3\rho}{r_{12}}\frac{Nw}{a}\,. \tag{5.106}$$

Here ΔV_N is the last slice, that overlaps the moving charge, Nw is the length b of the overlap box. But, how large is b? Well, it is the length of the box plus the distance the charge travels between t_1 and t_N,

$$\Delta t = (t_N - t_1) = \frac{(r_1 - r_N)}{c} = \frac{b}{c}\,. \tag{5.107}$$

Thus

$$b = a + \frac{vb}{c} \quad \Longrightarrow \quad b = \frac{a}{1 - v/c}\,. \tag{5.108}$$

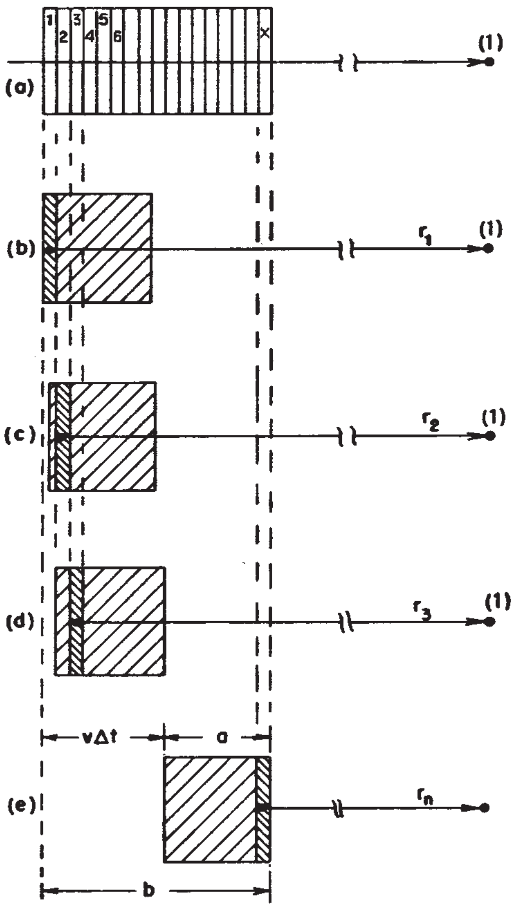

Figure 5.7: Change of integration volume when the charge is moving.

Plugging this into the expression for the potential results in

$$\phi(\vec{r}_1,\, t) = \frac{q}{4\pi\epsilon_0 \bar{r}_{12}} \frac{1}{[1-(v/c)]_{\text{ret.}}}\,, \tag{5.109}$$

where the subscript "ret." indicates that v is to be taken at the retarded time $t - \bar{r}_{12}/c$. This coincides with the result given previously: A correction term $1/[1-(v/c)]$ emerges, which is due to the fact that the charge moves "while the integral is taken". If the charges moves towards $\vec{r}_1$, its contribution to the integral is enhanced by b/a, which is just the correction term. For other directions it is clear that only the component of the velocity pointing towards (or away) from the spectator at $\vec{r}_1$ is of any consequence. Written in a nicer form, and suppressing the bar in the average, the general form of the scalar potential of a moving point-like charge q reads

$$\phi(\vec{r}_1,\, t) = \frac{q}{4\pi\epsilon_0 [r_{12} - (\vec{v}\cdot\vec{r}_{12}/c)]_{\text{ret.}}}\,, \tag{5.110}$$

where the subscript of the bracket is a reminder that all quantities in the bracket have to be evaluated at $t' = t - r_{12}/c$.
The same happens also, when calculating the vector potential $\vec{A}$ related to the point charge. The current density is ρv, the integral over ρ remains the same, and the final result reads

$$\vec{A}(\vec{r}_1,\, t) = \frac{qv}{4\pi\epsilon_0 c^2 [r_{12} - (\vec{v}\cdot\vec{r}_{12}/c)]_{\text{ret.}}}\,. \tag{5.111}$$

To gain the fields from the potentials, $\vec{\nabla}\phi$ and $\vec{\nabla}\times\vec{A}$ have to be evaluated. This is a tedious job, which will be left out here.

5.3 Energy and momentum of the fields

5.3.1 Local conservation of energy

Charge conservation as an example

Clearly, when objects like, e.g. the moving point charges of the last sections, emit electromagnetic waves, they emit light, and thus loose energy. Their energy is not conserved. As a physicist, one is accustomed to the idea of energy conservation, triggering the idea that the "lost" energy has just been transformed into another form of energy. From Lorentz-invariance, this immediately implies that the charges did not only loose energy when emitting waves, but also momentum; thus, electromagnetic waves also have momentum. To understand in more detail how this conservation of energy and momentum (or of four-momentum) works out, let's go back to the law that describes the conservation of charge, Eq. (2.32),

$$0 = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j}\,, \tag{5.112}$$

This law states that there is a charge and a current density, ρ and $\vec{j}$, respectively; when the charge density decreases (increases) at some point, it must flow away from (towards) the charge, hence there is a current density at that point. This law states in differential manner that the total charge of the world is constant.

However, the conservation of the world's total charge could be described in yet another way: Assume there is some charge Q_1 at some poi stion (1) while there is no charge at a point (2) nearby. Assume now that, while time goes by, Q_1 decreases and some increasing charge Q_2 emerges at (2), in such a way that $Q_1 + Q_2 = \text{const.}$. In this fashion, the total charge would remain constant - this is the global conservation of charge. The difference to Eq. (5.112) lies in the fact that the global version of charge conservation does not give any information about how the charge moved from (1) to (2). This lack of information boils down to some instantaneous transfer of charge: In the very moment it "gets lost" at (1) it reappears at (2). Let's think about this a bit deeper. What is wrong with such a concept? Well, the weak point in such a version of charge conservation is the fact that charges disappear and reappear at some other position instantaneously. Such instantaneous processes cannot be described in the framework of relativity, since there the notion of "instantaneous" is not equivalent in different systems. Two events that happen simultaneously in one system happen in sequence in yet another system, moving w.r.t. the first one. The only way to guarantee both charge conservation and Lorentz-invariance is to formulate it as a local law.

This local version of charge conservation immediately implies the notion of something happening between two points when a charge travels from one position to the next. This "something" is described by the current density $\vec{j}$. It turns out that energy conservation is a local conservation, too. Hence there is not only a scalar quantity - the energy density - but also a flux of energy, an energy current density, represented by a vector. This vector describes the rate of energy transfer through some surface.

Energy conservation

To formulate this in a quantitative form, the energy density, i.e. the energy in an infinitesimally small volume, and the rate of energy flux must be known. Let's start with the energy of the electromagnetic fields, denoted by u, and the corresponding flux, represented by $\vec{S}$. $\vec{S}$ describes the energy flux, i.e. the amount of energy per unit time and surface, through a surface element perpendicular w.r.t. it. Then, in full analogy to Eq. (2.32),

$$0 = \frac{\partial u}{\partial t} + \vec{\nabla} \cdot \vec{S} \,. \tag{5.113}$$

Of course this law is not valid in all situations, but only in those where the fields alone matter. This is quite a restricted set of situations. As a counter example consider yourself in a dark room and switch on the light. When the light flashes, suddenly there is energy in the electromagnetic fields, which has to come from somewhere (most probably from some bulbs). In such a situation the coupling of the fields to objects around matters in order to describe the full conservation of energy. In other words: In realistic situations, the fields apply work on matter and vice versa, changing the respective energies. This energy transfer has to be taken into proper account. After all, this is not too complicated. Each particle has energy

$$E_{\text{part.}} = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} \,, \tag{5.114}$$

where m_0 is the rest mass of the particle. The total energy of a block of matter therefore is the sum of particle energies, and the flux of this energy through the block's surface is just the sum of all particles' energies going through the surface. To write down a balance equation for the electromagnetic field energy boils down to stating that the total field energy inside a volume decreases (increases), either because energy evades (enters) the volume or because the field losses energy to (gains energy from) the particles inside. So, before fixing the balance, it is necessary to find an expression for the rate with which the fields supply work to the particles. The rate with which work is performed on a particle is given by $\vec{F} \cdot \vec{v} = q\vec{E} \cdot \vec{v}$, where the force is given by the Lorentz force

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})\,.$$

Assuming N particles per unit volume, the rate of work equals $q\vec{E} \cdot \vec{v} = \vec{E} \cdot \vec{j}$. Hence, this quantity $\vec{E} \cdot \vec{j}$ equals the energy loss of the field towards matter per unit time and volume. Thus,

$$-\frac{\partial}{\partial t} \int\limits_V u \mathrm{d}^3 V = \int\limits_{A=\delta V} \vec{S} \cdot \mathrm{d}^2\vec{a} + \int\limits_V \vec{E} \cdot \vec{j} \mathrm{d}^3 V\,. \tag{5.115}$$

This is the conservation law for the energy of electromagnetic fields in integral form. By applying Gauss' law its differential form can be found. It reads

$$-\frac{\partial u}{\partial t} = \vec{\nabla} \cdot \vec{S} + \vec{E} \cdot \vec{j}\,. \tag{5.116}$$

Well, this looks great, but it is not so great until the quantities u and $\vec{S}$ have been determined. This will be done now.

Energy density and flux

The full idea of this chapter boils down to the conviction that there is an energy density u and an energy flux $\vec{S}$ of electromagnetic fields which can be - up to constants - expressed through the fields alone. To see how this works out, let's try to rewrite the term $\vec{E} \cdot \vec{j}$ as a sum of two terms: One should be the time derivative of something, whereas the other term should be the divergence of something else. Both terms should be expressed by the fields alone, this would immediately yield u and $\vec{S}$. How is this possible? Well, the answer should be clear: Let's check Maxwell's equations! In fact, employing them results in

$$\vec{j} = \epsilon_0 c^2 \vec{\nabla} \times \vec{B} - \epsilon_0 \frac{\partial \vec{E}}{\partial t} \tag{5.117}$$

and leads to

$$\vec{E} \cdot \vec{j} = \epsilon_0 c^2 \vec{E} \cdot \left(\vec{\nabla} \times \vec{B}\right) - \epsilon_0 \vec{E} \cdot \frac{\partial \vec{E}}{\partial t} = \epsilon_0 c^2 \vec{E} \cdot \left(\vec{\nabla} \times \vec{B}\right) - \frac{\epsilon_0}{2} \frac{\partial E^2}{\partial t}\,. \tag{5.118}$$

The second term is the derivative w.r.t. time - or at least a part of it - that was part of the game. At first sight, also the first term looks good; employing a little vector algebra one is tempted to write

$$\vec{E}\cdot\left(\vec{\nabla}\times\vec{B}\right)=\vec{\nabla}\cdot\left(\vec{B}\times\vec{E}\right)\quad\text{wrong!}\,, \tag{5.119}$$

but unfortunately this is wrong, since, by convention, the ∇ operator acts on everything to the right of it. But this can be compensated. After some algebra it turns out that

$$\vec{\nabla}\cdot\left(\vec{B}\times\vec{E}\right)=\vec{E}\cdot\left(\vec{\nabla}\times\vec{B}\right)-\vec{B}\cdot\left(\vec{\nabla}\times\vec{E}\right)\,. \tag{5.120}$$

Plugging this into the equation above, one realises immediately that two terms already have the right behaviour, being either a time-derivative or a divergence; the only term left that does not fit is the last one. But, never mind, it can also be cast into a suitable form. Using Maxwell's equations leads to

$$\vec{B}\cdot\left(\vec{\nabla}\times\vec{E}\right)=\vec{B}\cdot\left(-\frac{\partial\vec{B}}{\partial t}\right)=-\frac{1}{2}\frac{\partial B^2}{\partial t}\,. \tag{5.121}$$

That's it. The equation for the energy is

$$\vec{E}\cdot\vec{j}=\vec{\nabla}\cdot\left(\epsilon_0c^2\vec{B}\times\vec{E}\right)-\frac{\partial}{\partial t}\left(\frac{\epsilon_0c^2}{2}\,B^2+\frac{\epsilon_0}{2}\,E^2\right)\,, \tag{5.122}$$

identifying terms leads to the definitions

$$\begin{aligned}u&=\frac{\epsilon_0c^2}{2}\,B^2+\frac{\epsilon_0}{2}\,E^2\\ \vec{S}&=\epsilon_0c^2\vec{B}\times\vec{E}\end{aligned} \tag{5.123}$$

for the energy density and flux, respectively. It is remarkable that these considerations lead, up to constants, to the same energy flux $\vec{S}$ that has been anticipated before (remember, this was the direction of plane wave propagation !). The vector is called the Poynting vector.
The first example is the one of the plane electromagnetic wave, cf. Sec. 5.2.1. In such a wave, the $\vec{E}$ and the $\vec{B}$ field are orthogonal w.r.t. each other and both are also orthogonal to the direction of wave propagation. As has been

discussed, the field strengths are proportional to each other, with constant of proportionality being the speed of light c. Therefore, for such waves the energy flux per unit area and second is given by

$$\vec{S} = \epsilon_0 c^2 \vec{B} \times \vec{E}\,, \quad S = \epsilon_0 c E^2\,. \tag{5.124}$$

This generalises to light, since it can always be locally approximated by plane waves.

Example for energy density and flow

Let's investigate another, more involved example for energy density and flow related to electromagnetic fields. The example is a capacitor that is being slowly charged by some AC voltage. The AC frequency should be low enough to ensure that the capacitor does not behave like a cavity. Let's take a cylindrical capacitor with circular plates of radius a, in a distance h. There will be an $\vec{E}$ field between the planes. Neglecting inhomogeneities translates into a homogenous electric field filling the volume $V = \pi a^2 h$ between the planes; thus the total energy between the plates reads

$$U = \int\limits_V \mathrm{d}^3 V u = \frac{\epsilon_0}{2} E^2 \, \pi a^2 h\,. \tag{5.125}$$

This energy changes with $\vec{E}$. Charging the capacitor implies that energy is added to the volume with a rate

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \epsilon_0 \pi a^2 h \vec{E} \cdot \dot{\vec{E}}\,. \tag{5.126}$$

Since the energy changes, there must be some energy flux entering the space between the plates. This seems simple, it's the wires and the planes that lives up for the energy flux, right? Well, this is wrong. After all, $\vec{E}$ is orthogonal to the planes, hence $\vec{S} \sim \vec{E} \times \vec{B}$ is parallel to them - there is no energy flow through the planes.
The correct answer concerning the energy flow is a bit more complicated. While the capacitor is charged, the $\vec{E}$ field changes, inducing a $\vec{B}$ field. This field circulates around the axis of the capacitor and around the $\vec{E}$ field. Its strength is proportional to the rate of change of the electric field, doing the maths results in

$$2\pi a c^2 B = \cdot E \pi a^2 \quad \Longrightarrow \quad B = \frac{a}{2c^2} \dot{E}\,. \tag{5.127}$$

This leads to an energy flow proportional to $\vec{E}\times\vec{B}$ (they are in fact orthogonal to each other, $\vec{B}$ around the capacitor and $\vec{E}$ along its axis). The disturbing thing is that the energy entering the volume between the capacitor planes seemingly comes from the space around the capacitor. Let's check whether this influx is enough to compensate for the energy change inside the volume between the plates. The absolute value of $\vec{S}$ is

$$|\vec{S}| = \epsilon_0 c^2 |\vec{B}||\vec{E}| = \epsilon_0 \pi a^2 h E \dot{E}\,, \tag{5.128}$$

in perfect agreement with $\mathrm{d}U/\mathrm{d}t$. This settles the case.

5.3.2 Momentum of the fields

Momentum balance

In full analogy to the energy density of the fields, a momentum density $\vec{g}$ may be defined. Similar reasoning as above results in the demand that such a momentum density needs to fulfil a continuity equation, too. Ignoring the coupling to external sources for a moment, one may therefore write

$$\frac{\partial \vec{g}}{\partial t} + \vec{\nabla}\text{something} = 0\,. \tag{5.129}$$

The question now arises: What is this “something”? Due to the vector character of the momentum density $\vec{g}$ and, hence, its temporal derivative, it is clear that applying $\vec{\nabla}$ on this something must yield a vector. Thus, “something” cannot be a vector, because then, the application of $\vec{\nabla}$ is synonymous to taking the divergence of this vector, resulting in a scalar quantity. Similarly, the $\vec{\nabla}$ operation should not be done on a scalar (i.e., taking this to be the gradient of a scalar), since then the two quantities would have different behaviour under coordinate transformations. The only option left is that “something” is a tensor; it will turn out that it is Maxwell's tension tensor, denoted here by $\hat{\Pi}$.
Now the fields, or, more precisely, their momentum density, will be coupled to external particles and sources. This is not too complicated, after all, it is known that for the momentum of a point particle

$$\dot{\vec{p}} = \vec{F}\,, \tag{5.130}$$

where $\vec{p}$ is the particle's momentum and $\vec{F}$ is the external force acting on it. Analogous reasoning yields

$$\frac{\partial \vec{g}}{\partial t} + \vec{\nabla}\text{something} = \vec{f}, \tag{5.131}$$

where $\vec{f}$ is the force density. The Lorentz force is given by

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B} \tag{5.132}$$

the corresponding Lorentz force density is given by

$$\vec{f} = \rho\vec{E} + \vec{j} \times \vec{B} \tag{5.133}$$

with ρ and $\vec{j}$ the charge and current density, respectively. Again, Maxwell's equations can be used to replace them, resulting in

$$\begin{aligned}
\vec{f} &= \left(\epsilon_0 \vec{\nabla} \cdot \vec{E}\right)\vec{E} + \left(\epsilon_0 c^2 \vec{\nabla} \times \vec{B} - \epsilon_0 \frac{\partial \vec{E}}{\partial t}\right) \times \vec{B} \\
&= \frac{\epsilon_0}{2}\vec{\nabla}E^2 + \left(\epsilon_0 c^2 \vec{\nabla} \times \vec{B}\right) \times \vec{B} - \epsilon_0 \frac{\partial}{\partial t}\left(\vec{E} \times \vec{B}\right) + \epsilon_0 \vec{E} \times \left(\frac{\partial \vec{B}}{\partial t}\right) \\
&= \frac{\epsilon_0}{c^2}\frac{\partial \vec{S}}{\partial t} + \frac{\epsilon_0}{2}\vec{\nabla}E^2 + \left(\epsilon_0 c^2 \vec{\nabla} \times \vec{B}\right) \times \vec{B} - \epsilon_0 \vec{E} \times \left(\vec{\nabla} \times \vec{E}\right) .
\end{aligned} \tag{5.134}$$

Comparing with Eq. 5.131 motivates that the momentum density of the fields equals the Poynting vector, divided by c^2. The Poynting vector therefore does not only yield the energy flow but also the momentum density of the electromagnetic fields.

Interpretation

The following interpretation of the fields' momentum density dates back to Einstein and goes as follows: Assume that there is a heavy waggon of mass M, rolling without friction on the tracks. On one end inside there is a mechanism, emitting light (maybe a laser or similar), that will be absorbed at the other end. By emitting and absorbing light some amount U of energy has been transported over a distance L, from one end of the waggon to the other. Now, according to Einstein, energy relates to some mass according to $E = mc^2$, thus this energy transfer relates to a mass transfer of mass U/c^2.

If the waggon remains at rest, its centre of mass has been shifted by some internal action - this is bad. This shift of centre of mass must be compensated by a shift the full waggon experienced. Let's call it x, it relates to the other quantities like (conservation of the centre of mass)

$$Mx = \frac{U}{c^2}L\,. \tag{5.135}$$

How does this work? The explanation Einstein found is that in the moment the light was emitted, the waggon recoiled with momentum p, driving the waggon back. The velocity of the waggon is

$$v = \frac{p}{M}\,, \tag{5.136}$$

clearly not relativistic, since the waggon is so heavy. Anyways, the cart moves, until the light reaches the other end of the waggon where it is absorbed. The distance the cart travels is

$$x = vt = \frac{p}{M}\frac{L}{c}\,. \tag{5.137}$$

Plugging this into the conservation of the centre of mass results in

$$p = \frac{U}{c}\,. \tag{5.138}$$

This is the momentum of the light, to gain its momentum density, one needs to divide by c and finds

$$g = \frac{p}{c} = \frac{U}{c^2}\,, \tag{5.139}$$

as anticipated.

5.3.3 Electromagnetic mass

Energy and momentum of the field of a point charge

Let's calculate now the energy of a charged particle. To keep things easy, a simple model for, say, an electron will be used: It is just assumed that its total charge q is distributed over the surface of a sphere with radius a.

If the electron is at rest, there is an electric field only, its absolute value is $q/4\pi\epsilon_0 r^2$. Hence, the energy density of the electric field reads

$$u = \frac{\epsilon_0}{2}E^2 = \frac{q^2}{32\pi^2\epsilon_0 r^4}\,. \tag{5.140}$$

The total energy of the electric field results from integration over the full space. Using spherical coordinates this boils down to

$$U = \int\limits_a^\infty u \cdot 4\pi^2 \mathrm{d}r = \frac{1}{2}\frac{q^2}{4\pi\epsilon_0}\frac{1}{a} \tag{5.141}$$

Everything's fine, until a point charge is assumed, i.e. until $a \to 0$. This is the big problem: The field surrounding a point charge contains an infinite amount of energy. Is this a problem? As long as the energy is just there, doing nothing, it is just a nuisance, it becomes a problem only, if observable physical effects are triggered. To answer this, let's check what happens if the charge moves. If the changes are infinite, the theory has a problem.
Assume now that the electron moves with uniform velocity $v \ll c$. This electron has a momentum (even if it was massless before "being charged") because of the momentum of the accompanying electromagnetic field. It can be shown that this momentum is in the same direction as v, and that it is proportional to v for small velocities. For an arbitrary point P in distance r to the centre of the charge and with angle θ to the direction of motion the electrical field is radially oriented. The magnetic field is given by $\vec{v} \times \vec{E}/c^2$, as seen before. Then the momentum density is

$$\vec{g} = \epsilon_0 \vec{E} \times \vec{B} \quad \Longrightarrow \quad |\vec{g}| = \frac{\epsilon_0 v}{c^2}E^2 \sin\theta \tag{5.142}$$

and it is tilted w.r.t. the direction of motion.
The fields are symmetric w.r.t. the direction of motion, thus their integrated transverse components are zero. Hence, the resulting momentum is parallel to $\vec{v}$. The component of $\vec{g}$ in that direction is $|\vec{g}|\sin\theta$, which has to be integrated over the full volume. As volume element a ring of radius r is most suitably chosen; its plane is perpendicular w.r.t. $\vec{v}$, and its volume element

is given by $2\pi r^2 \mathrm{d}r \sin\theta \mathrm{d}\theta$. The full momentum then reads

$$
\begin{aligned}
|\vec{p}| &= \int |\vec{g}| \sin\theta \, 2\pi r^2 \mathrm{d}r \sin\theta \mathrm{d}\theta \\
&= \frac{2\pi\epsilon_0 v}{c^2} E^2 \int\limits_a^\infty r^2 \mathrm{d}r \int\limits_0^\pi \sin^3\theta \mathrm{d}\theta \\
&= \frac{2\pi\epsilon_0 v}{c^2} E^2 \left(\frac{q^2}{16\pi^2\epsilon_0^2 a}\right)\left(\frac{4}{3}\right) = \frac{2}{3}\frac{q^2}{4\pi\epsilon_0}\frac{v}{ac^2} \,. \qquad (5.143)
\end{aligned}
$$

This looks good: The momentum changes with v, this allows to define an electromagnetic mass.

Electromagnetic mass

This momentum, as said before, is parallel to v and proportional to it, therefore one may interpret the coefficient of v as the electromagnetic mass $m_{\mathrm{el.}}$,

$$
m_{\mathrm{el.}} = \frac{2}{3}\frac{q^2}{4\pi\epsilon_0 a c^2} \,. \qquad (5.144)
$$

But where does this mass come from? So far the custom was that particles have properties, one of them being the mass. These properties, to some extent were just there and characterised the objects. Now these properties are somewhat changed in their origin[4], charged particles, when moving, have a momentum proportional to their velocity, even if they were massless at rest. This opens the door for speculations, like, e.g., understanding masses, and even rest masses, dynamically as the consequence of some interaction. But to be on the safe side one may also assume that the momentum of a charged particle is given as the sum of its "ordinary" mechanical momentum plus the electromagnetic momentum that has just been discussed. In that sense, its momentum $\vec{p} = (m_{\mathrm{mech.}} + m_{\mathrm{el.}})\vec{v}$. Having said that, let's check now, which size an electron should have in order not to have a mechanical mass. This can be found by equating $m_{\mathrm{el.}}$ with the measured mass of the electron, m_e. Introducing $e^2 = q^2/4\pi\epsilon_0$ one finds

$$
a = \frac{2}{3}\frac{e^2}{m_e^2 c^2} \,. \qquad (5.145)
$$

[4]Remember, so far, the electron could have very well been massless without changing the reasoning leading to the electromagnetic mass.

The quantity

$$r_0 = \frac{e^2}{m_e^2 c^2} = 2.82 \cdot 10^{-13} cm \tag{5.146}$$

is known as the "classical electron radius". The missing factor of 2/3 can be explained by hinting at the fact that in this derivation a specific charge distribution has been assumed, of course one may have chosen just a charged sphere or anything else. But independent of these details, the characteristic scales are just $e^2/m_e^2 + c^2$.

But let's follow the track of a moving electron further on, by lifting the constraint $v \ll c$. What happens thus for large velocities? First attempts in this direction ended badly, until Lorentz realised that the sphere becomes flatter and flatter in the direction of motion (yet another effect of length contraction of relativistic objects), until it becomes a "pancake". He also realised that the fields change according to the Lienard-Wiechert potentials. When calculating the integrals, it turns out that the momentum is changed by some factor $1/\sqrt{1 - v^2/c^2}$ for arbitrary velocities. Thus

$$\vec{p} = \frac{2}{3} \frac{e^2}{ac^2} \frac{\vec{v}}{\sqrt{1 - v^2/c^2}} \,. \tag{5.147}$$

In other words, the electromagnetic mass increases with this very same factor of $1/\sqrt{1 - v^2/c^2}$. This discovery has been made before the advent of the theory of special relativity. A period of frantic measurements began - people tried to measure this rise of the mass in order to determine the fraction of mechanical and electrodynamic masses - until it was realised that also mechanical masses rise by the same factor.

But let's return finally to the calculation of the field energy, Eq. (5.141). Using the relativistic connection of energy and mass, one may find another electrodynamic mass, namely

$$m'_{\text{el.}} = \frac{U}{c^2} = \frac{1}{2} \frac{e^2}{ac^2} \,, \tag{5.148}$$

which does not coincide with what has been found so far. Combining things the other way around, one finds instead

$$U = \frac{3}{4} m_{\text{el.}} c^2 \,. \tag{5.149}$$

Chapter 6

Matter

6.1 Dielectrics

6.1.1 Dielectric materials

A phenomenon involving dielectrics

Consider a capacitor made of two slabs with area A and distance d. In Eq. (3.83) it has been shown that the relation between charges on the slabs Q and voltage V in between is given by the capacity $C = Q/V$. For a simple capacitor of the type described here, C is given by

$$C_{\parallel} = \frac{\epsilon_0 A}{d} \,. \tag{6.1}$$

According to experimental findings the capacity is increasing when some isolating material, like e.g. glass, is brought in the volume between the slabs. This increase of capacity implies that the voltage is reduced although the amount of charge on each slab remains constant. But since the voltage is nothing but the integral of the electrical field strength $|\vec{E}|$ inside the capacitor, the electrical field is reduced. How is this possible? After all, Gauss' law states that the electrical field strength is related to the included charge. As an example consider the Gauss' surface S in Fig. 6.1. Since $\vec{E}$ is decreased by the presence of the dielectric, the conclusion is that the total charge inside S is diminished. There is one reason for this only: There must be some positive charge on the surface of the dielectric. Since the electrical field strength is not zero, but just smaller, it is clear that the negative charges on the inside

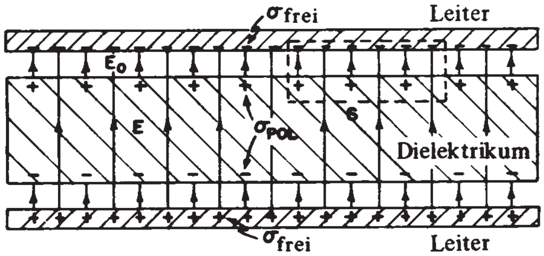

Figure 6.1: A slab capacitor with a dielectric inside. A Gaussian surface S is indicated by the dashed lines, in addition, the field lines of $\vec{E}$ are shown.

of the capacitor slab outnumber the positive ones on the dielectric. This phenomenon thus may be explained when it is clear why in dielectrics, when brought into such a configuration, positive and negative charges are moved on the surfaces in a way that they partially compensate the field around.
For conductors, the case is clear: There, charges are moved around on the surface in such a way that the interior of the conductor is totally shielded, i.e. such that the charges "push" the external fields out of the conductor. To illustrate this, assume that a conductor of thickness b is brought into the capacitor. The charges on the conductor move around in a fashion that the fields outside the conductor remain unchanged, but there is no field inside. Thus, when integrating the field strength over the volume, the effective size is $d-b$ rather than d, the potential is given by

$$V = \frac{\sigma(d-b)}{\epsilon_0}\,, \tag{6.2}$$

where σ is the area charge density on the slabs, and the capacity turns out to be

$$C = \frac{\epsilon_0 A}{d(1-b/d)}\,. \tag{6.3}$$

It is enhanced by a factor $1/(1-b/d)$, this is exactly the part occupied by the conductor.
From this reminder, a simple model for dielectrics can be developed: A dielectric consists of little conducting balls inside some isolating material.

Electrical fields induce charges on the balls, hence each of them behaves like a little dipole with a moment induced by the external field. It is easy to think of these balls as atoms or molecules with some polar structure that gets oriented. When the external field is not too strong, it seems reasonable to assume that there is a linear connection between the induced dipole moment and the field strength. Throughout this chapter, linearity will be taken for granted. Assuming the dipole moment to be due to some charge q shifted by a tiny distance δ, N dipoles per unit volume produce a dipole moment P per unit volume given by

$$P = Nq\delta\,. \tag{6.4}$$

Usually P is a function of position inside the dielectric; however, at each point it is proportional (linearity!) to E. The constant of proportionality depends, of course, on the material.

Susceptibility

Let's check now what the model above has to say about capacitors with a dielectric inside. For a start consider a layer of this dielectric with a certain dipole moment $\vec{P}$ per unit volume. Does this dipole moment generate a charge density? The answer clearly is no as long as $\vec{P}$ is homogenous [1] But if $\vec{P}$ is larger or smaller at some points, then the charges are not just shifted in some "parallel" way; in such a case, more charges would be moved into a region than charges were taken out. This would clearly result in a charge density different from zero. Thus, to keep things simple, in the following it will always be assumed that $\vec{P}$ is homogenous. Then the only effects due to this dipole moment manifest themselves on the surface of the layers. In fact, a homogenous $\vec{P}$ over each layer results in a number of electrons (negative charges) moving out of the layer by some distance δ and a number of holes (positive charges) moving into the layer on the other side by the same distance. This pattern reproduces over all layers, finally some area charge densities emerge on the surface of the dielectric. It can be calculated as follows: If A is the area of the layer - oriented parallel to the capacitor, of course - then the number of electrons moving out of it is $NA\delta$, where N is

[1] As long as the negative and positive charges, that are shifted w.r.t. each other, have the same density, the net effect cancels out: The shift doesn't matter since it does not generate a net charge somewhere.

the density per unit volume. Hence the total area charge is

$$\sigma_{\text{pol}} = \pm N q_e \delta \,, \tag{6.5}$$

where q_e is the electron charge and the sign reflects the two surface in question. Obviously then $P = \sigma_{\text{pol}}$.
On the inner surfaces of a capacitor, there are charges which can move freely on them, leading to an area charge density σ_{free}. These charges are the ones responsible for the field inside the capacitor. It is clear that if a dielectric is brought into such a capacitor, the charges inside the dielectric will move such that negative charges on the dielectric are closer to the positively charged capacitor plate and vice versa. Of course, the so emerging polarisation appears when the capacitor is charged and it vanishes when the capacitor is discharged. Both effects are not due to charges moving along some wire attached to the dielectric - they are just the emergence of a polarisation.
Now, Gauss' law can be applied to the surface indicated in Fig. 6.1. The electric field strength inside the dielectric equals the total area charge density, divided by ϵ_0. Taking into account both charge densities with their proper sign leads to

$$|\vec{E}| = \frac{\sigma_{\text{free}} - \sigma_{\text{pol}}}{\epsilon_0} = \frac{\sigma_{\text{free}} - P}{\epsilon_0} \,. \tag{6.6}$$

Similar reasoning implies that the field in the volume between capacitor slab and dielectric is larger than E, but this will be omitted for the time being. Assuming now a simple linear dependence of $\vec{P}$ on $\vec{E}$ in the form

$$\vec{P} = \epsilon_0 \chi \vec{E} \tag{6.7}$$

leads to

$$|\vec{E}| = \frac{\sigma_{\text{free}}}{\epsilon_0} \frac{1}{1+\chi} \,, \tag{6.8}$$

where the constant χ is the electrical susceptibility of the dielectric, clearly a property of its material. In a similar fashion, the potential (voltage) on the capacitor is reduced by $1/(1+\chi)$, thus the capacity is enhanced by $1+\chi$,

$$\begin{aligned} V &= Ed = \frac{\sigma_{\text{free}} d}{\epsilon_0} \frac{1}{1+\chi} \\ C &= \frac{Q}{V} = \frac{\epsilon_0 A (1+\chi)}{d} \,. \end{aligned} \tag{6.9}$$

Of course, situations can be anticipated, where $\vec{P} \not\parallel \vec{E}$, but the connection is still linear. Then the susceptibility can be written as a tensor of rank two (represented as a matrix).
Let's complicate the situation a bit now by assuming that the polarisation $\vec{P}$ is not the same all over the dielectric. Then charge densities result inside the dielectric, and the exact amount of charge entering and leaving some volume elements inside the dielectric have to be determined. It is clear that the amount of charge entering through a surface S is given by the component of $\vec{P}$ orthogonal to it, hence

$$\sigma_{\mathrm{pol}}\big|_S = \vec{P} \cdot \vec{n}_S \,. \tag{6.10}$$

Of course this reasoning does not lead to a charge density on the surface element in question, but combining some surface elements may easily to some charge density in the volume enclosed by them. The total charge leaving a volume through some polarisation $\vec{P}$ equals the integral over the surface enclosing it times the normal component of $\vec{P}$ pointing out of the volume. Inside the same amount of charge, but with reverse sign remains,

$$\Delta Q_{\mathrm{pol}} = \oint\limits_S \mathrm{d}^2\vec{s} \cdot \vec{P} = \int\limits_V \mathrm{d}^3V \rho_{\mathrm{pol}} \,. \tag{6.11}$$

This immediately connects the charge density inside a dielectric with a polarisation vector $\vec{P}$. Applying Gauss' theorem yields

$$\rho_{\mathrm{pol}} = -\vec{\nabla} \cdot \vec{P} \,. \tag{6.12}$$

6.1.2 Electrostatics in the presence of dielectrics

Basic equations

Combining this finding with the basic equations of electrostatic leads to

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} = \frac{\rho_{\mathrm{free}} + \rho \mathrm{pol}}{\epsilon_0} = \frac{\rho_{\mathrm{free}} - \vec{\nabla} \cdot \vec{P}}{\epsilon_0} \tag{6.13}$$

or, written in different form,

$$\vec{\nabla} \cdot \left(\vec{E} + \frac{\vec{P}}{\epsilon_0} \right) = \frac{\rho_{\mathrm{free}}}{\epsilon_0} \,. \tag{6.14}$$

Defining the term in brackets as

$$\vec{D} = \vec{E} + \frac{\vec{P}}{\epsilon_0} \tag{6.15}$$

then yields as basic equations for electrostatics in the presence of dielectrics

$$\vec{\nabla} \cdot \vec{D} \;=\; \rho_{\text{free}} \quad \text{and} \quad \vec{\nabla} \times \vec{E} = 0\,. \tag{6.16}$$

Can this system be solved? Well, a solution can only be constructed when $\vec{D}$ and $\vec{E}$ are connected. Assuming again a linear dependence of the polarisation on the external field produces this connection, formalised as

$$\vec{D} = \epsilon_0(1+\chi)\vec{E} = \epsilon\vec{E}\,, \tag{6.17}$$

where ϵ is called the permeability. This explains the frequent occurrence of ϵ_0: It is the permeability of the vacuum. This way of representing electrostatics in the presence of media is partly a historical "accident". In earlier times, no-one had a clue about the atomic mechanism producing the polarisation of materials, and nobody thought about a charge density inside dielectrics, ρ_{pol} was just not a concept and ρ_{free} was thought to represent the total charge present at any given point. In order to cast Maxwell's equations in vacuum in a similar (and similarly simple) form, when materials where present, $\vec{D}$ was introduced. Todays perspective is more economic. Our equations in vacuum are much simpler - after all, only one electrical field strength is needed - and in order to compensate for material effects, just all charges need to be taken into account. Separating some charges from this balance implies a different form of some equations, one of them being the one with the $\vec{D}$ field.
Another point is worth stressing: Equations like Eq. (6.17) provide an attempt to parametrise the properties of matter. But they are complicated, and in most cases such an equation might be tempting but somewhat incomplete or even wrong. For instance, if $\vec{E}$ becomes too large, the linear dependence breaks down. In this respect, Eq. (6.17) is an approximation, just like Hooke's law is an approximation for the connection of the force acting on a spring and its strain. In other words, such a law does not represent a fundamental concept of physics. In contrast, Maxwell's equations do represent such a fundamental concept.

Forces and fields

Let's consider some rather general properties of the equations above, Eq. (6.16). It is clear that, if the connection between $\vec{E}$ and $\vec{D}$ is linear with constant factor, the constant factor can be multiplied to the field in the rotor. Then

$$\vec{\nabla} \times \vec{D} = 0\,. \tag{6.18}$$

Hence, the same equations are valid for $\vec{E}$ in vacuum and for $\vec{D}$ inside a dielectric. There, the field is smaller by a factor $\kappa = \epsilon/\epsilon_0$. Since potential differences are line integrals of the field strength, the voltages inside a dielectric are reduced by the same factor. In turn, capacities are enhanced by this factor.

But what is the force between two charged conductors inside such a dielectric? To illustrate this, consider a homogenous dielectric fluid. It is known that one possibility to gain such a force is to differentiate the energy w.r.t. the corresponding distance. If the two conductors are charged with the same charge Q but with different sign, the energy of this system is

$$U = \frac{Q^2}{2C}\,, \tag{6.19}$$

where C is the capacity of them. Thus

$$\vec{F} = -\vec{\nabla}U = -\frac{Q^2}{2}\vec{\nabla}\frac{1}{C}\,. \tag{6.20}$$

Since the capacity is enhanced by κ due to the dielectric, all forces are reduced by this factor. In this reasoning, however, it needs to be stressed, that it is valid for fluid dielectrics only; in solid dielectrics each movement of conductors applies mechanical stress and hence the electrical properties change. This change also alters the energy balance.

So far, things look quite simple, but this is, in fact, not quite the case. To illustrate this consider the following, surprisingly difficult problem in the theory of dielectrics. Why does a charged body attract little pieces of dielectric? The answer has to do with the polarisation a dielectric forms when brought into an electric field. From the reasoning above it should be clear that the polarisation works in such a way that those charges opposite to the charged body are moved closest to it, and those with the same sign are moved away from it. Since the dielectrics have some spatial extension this results in an attraction that is slightly stronger than the corresponding repulsion. The inhomogeneity of the electrical field of a charged body (decreasing with increasing distance) is an important part of this mechanism. But how strong is

the force? Some reasoning shows that the attractive force is always proportional to the square of the electrical field strength. This looks surprising first; however, it is not so complicated. From the reasoning above the polarisation is proportional to the field strength, resulting in a charge inhomogeneity proportional to the field. Since the force on a charge applied by an electrical field is proportional to the product of charge and field, the force on the dielectric goes with the square of the field.
There is a similar problem where the force acting on a dielectric can be calculated exactly. This is a slab capacitor, where a dielectric is only partially inserted. As will be seen in the following, the dielectric is pulled into the capacitor. To calculate the force directly, complicated reasoning is needed, boiling down to investigating the effect of inhomogeneities of the field at the edge of the dielectric. But when these details are ignored and the principle of energy conservation is employed instead, the force can be easily determined. Writing

$$U = \frac{Q^2}{2C} = \frac{V^2 C}{2} \tag{6.21}$$

results in

$$\vec{F} = -\vec{\nabla} U = \frac{V^2}{2} \vec{\nabla} C = \frac{V^2}{2} \frac{\partial C}{\partial x} \vec{e}_x \,, \tag{6.22}$$

where the capacitor is oriented such that the dielectric moves parallel to the x-axis. The only thing left then is to find out how the capacity changes with the moving dielectric. In the following L is the total length of the slabs and W is their width and if d is the thickness of the dielectric, which fits exactly into the volume between the slabs. The capacity is the ratio of free charges on the inner surfaces of the slabs and the voltage. These charges are given by

$$Q = \frac{\kappa \epsilon_0 V}{d} x W + \frac{\epsilon_0 V}{d} (L - x) W \tag{6.23}$$

leading to

$$C = \frac{\kappa \epsilon_0}{d} x W + \frac{\epsilon_0}{d} (L - x) W \tag{6.24}$$

resulting in

$$F_x = \frac{V^2}{2} \frac{\epsilon_0 W}{d} (\kappa - 1) \,. \tag{6.25}$$

6.2 Electromagnetic waves in matter

6.2.1 Wave propagation in matter

Polarisation of matter

In the following a simple model for the reaction of matter on an incoming electromagnetic wave (light) will be presented. It will serve as the starting point for a discussion of material specific properties such as refraction index or reflexion at surfaces. Assume a particle density of N particles per unit volume and assume electrons to be bound in atoms or molecules through a force that is proportional to its shift. This is, of course, simplistic, but it serves the point. The equation of motion for such a bound electron is

$$\vec{F} = q_e\vec{E} \equiv m_e(\ddot{\vec{x}} + \gamma\dot{\vec{x}} + \omega_0^2\vec{x})\,, \tag{6.26}$$

where $\vec{x}$ is the displacement of the electron in the direction of the field and where some friction γ has been added[2]. For simplicity a linearly polarised wave has been taken - in such a case, $\vec{E}$ does not change its direction during a wave length. If the field depends sinusoidally on time,

$$\vec{E} = \vec{E}_0 e^{i\omega t} \tag{6.27}$$

the displacement oscillates with the same frequency,

$$\vec{x} = \vec{x}_0 e^{i\omega t}\,, \tag{6.28}$$

and by replacing $\dot{\vec{x}} = i\omega\vec{x}$ and $\ddot{\vec{x}} = -\omega^2\vec{x}$ one finds

$$\vec{x} = \frac{q_e}{m_e}\,\frac{\vec{E}}{(\omega_0^2 - \omega^2) + i\gamma\omega}\,. \tag{6.29}$$

This oscillation induces a dipole moment $\vec{p}$ of the atom or molecule in question, namely

$$\vec{p} = q_e\vec{x} = \frac{q_e^2}{m_e}\,\frac{\vec{E}}{(\omega_0^2 - \omega^2) + i\gamma\omega} = \epsilon_0\alpha(\omega)\vec{E}\,, \tag{6.30}$$

[2] Also, an isotropic harmonic oscillator has been taken as a model for the bound electron, i.e. a three-dimensional oscillator, with no no preferred direction.

where $\alpha(\omega)$ is the atomic polarisability. It depends on the incident frequency and reads

$$\alpha(\omega) = \frac{q_e^2}{m_e} \frac{1}{(\omega_0^2 - \omega^2) + i\gamma\omega} \,. \tag{6.31}$$

In a more sophisticated model of the atom or molecule, there's a number of eigenfrequencies ω_k (and not only one, ω_0), each of which has its own dissipation constant γ_k and relative strength f_k. In such a model, α reads

$$\alpha(\omega) = \frac{q_e^2}{m_e} \sum_k \frac{f_k}{(\omega_k^2 - \omega^2) + i\gamma_k\omega} \,. \tag{6.32}$$

From this the dipole moment per unit volume $\vec{P}$ can be determined. It is given by

$$\vec{P} = \epsilon_0 N \alpha(\omega) \vec{E} \,. \tag{6.33}$$

It should be stressed that the constant of proportionality between the incident electrical field and the resulting dipole moment depends on ω, in most cases, α is small for very large frequencies ω (the electrons do not have enough time to react) and it may be large for comparably small frequencies. In addition, there might be a complex part (the $i\gamma$ piece), leading to a corresponding phase shift. However, this oscillating dipole moment leads to an emission of another wave which interferes with the incident one and thus leads to refraction.

Maxwell's equation in a dielectric

The existence of this polarisation leads to corresponding charges and currents that have to be taken into account when calculating the electromagnetic fields through Maxwell's equations. In the following solutions for Maxwell's equations will be constructed, where the charges and currents do not vanish like in the vacuum case but where the polarisation $\vec{P}$ is responsible for them. Thus, as a first step, the average value of the charge density ρ and the current density $\vec{j}$ will be calculated in a small volume. For simplicity, this volume will be of the same size as the one used for the definition of $\vec{P}$. Remember that previously the polarisation-induced charge density $\rho_{\rm pol}$ has been defined by

$$\rho_{\rm pol} = -\vec{\nabla} \cdot \vec{P} \tag{6.34}$$

for static polarisations. However, the same connection is valid for temporally changing fields and polarisations as well. But if $\vec{P}$ and thus ρ_{pol} changes with time, there must be corresponding current densities. Each of the oscillating charges discussed above adds its share to a current density equal to the product of its charge q_e and velocity v. Assuming, as before, N charges per unit volume, the current density reads

$$\vec{j} = N q_e \vec{v}\,, \tag{6.35}$$

identifying $\vec{v} = \mathrm{d}\vec{x}/\mathrm{d}t$ and using $\vec{p} = q_e \vec{x}$ for each charge results in

$$\vec{j}_{\text{pol}} = N q_e \frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = N \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \frac{\mathrm{d}\vec{P}}{\mathrm{d}t}\,. \tag{6.36}$$

This renders the solution for Maxwell's equations a straightforward problem: Using the relation of $\vec{E}$ and $\vec{P}$ solutions can be constructed, including ρ_{pol} and $\vec{j}_{\text{pol}}$. In particular, solutions for electromagnetic waves can be found. Maxwell's equations in a dielectric without any other, additional currents or charges read

$$\begin{aligned} \vec{\nabla}\cdot\vec{E} &= \frac{\rho_{\text{pol}}}{\epsilon_0} = -\frac{\vec{\nabla}\cdot\vec{P}}{\epsilon_0} & \vec{\nabla}\cdot\vec{B} &= 0 \\ \vec{\nabla}\times\vec{B} - \epsilon_0\mu_0\frac{\partial\vec{E}}{\partial t} &= \mu_0\vec{j} = \frac{\partial\vec{P}}{\epsilon_0\partial t} & \vec{\nabla}\times\vec{E} + \frac{\partial\vec{B}}{\partial t} &= 0\,. \end{aligned} \tag{6.37}$$

The solution can be constructed in perfect analogy to the vacuum case. First of all, the rotor of the last equation reads

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

which can be cast into

$$\vec{\nabla}\left(\vec{\nabla}\cdot\vec{E}\right) - \nabla^2\vec{E} = -\frac{\partial}{\partial t}\left(\vec{\nabla}\times\vec{B}\right)\,. \tag{6.38}$$

Using the equations for $\vec{\nabla}\times\vec{B}$ and for $\vec{\nabla}\cdot\vec{E}$ results in

$$\nabla^2\vec{E} - \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = -\frac{1}{\epsilon_0}\vec{\nabla}\left(\vec{\nabla}\cdot\vec{P}\right) + \frac{1}{\epsilon_0 c^2}\frac{\partial^2\vec{P}}{\partial t^2}\,. \tag{6.39}$$

Instead of the usual wave equation this means that the d'Alembert operator acting on $\vec{E}$ equals two expressions containing the polarisation. Since $\vec{P}$ depends on the electrical field, however, wave solutions are still possible. Concentrating on isotropic dielectrics, i.e. on such material where $\vec{P} \parallel \vec{E}$, let's try to find a solution for a wave propagating in z direction, polarised in x direction. In such a case, the electrical field has an x component only which behaves like $E_x \sim \exp(i\omega t - ik_z z)$. It is well-known that each function of $(z - vt)$ describes a wave propagating in z-direction with velocity v. Similarly, the exponent of

$$E_x = E_0 e^{i(wt - k_z z)} \tag{6.40}$$

can be written as $-ik(z - \omega/kt)$, establishing that E_x is described by a wave with phase velocity

$$v_{\mathrm{ph}} = \frac{\omega}{k} \,. \tag{6.41}$$

Refraction index

The connection of phase velocity and refraction n index is

$$n = \frac{c}{v_{\mathrm{ph}}} \,; \tag{6.42}$$

hence,

$$E_x = E_0 e^{iw(t - nz/c)} \,. \tag{6.43}$$

Hence, the refraction index of the dielectric can be found by investigating which value of k_z allows for a solution of the wave equation above, Eq. (6.39). In an isotropic material, where $\vec{E} \parallel \vec{P}$, there is only an x-component of $\vec{P}$, since there is only an x component of $\vec{E}$. But the wave is propagating in z direction, therefore $\vec{P}$ does not depend on x and $\vec{\nabla} \cdot \vec{P} = 0$. Assuming a linear dielectric, P_x changes with $\exp(i\omega t)$ and

$$\frac{\partial^2 \vec{P}_x}{\partial t^2} = -\omega^2 P_x \,. \tag{6.44}$$

Plugging in this behaviour, and taking also into account the form of $\vec{E}$ results in

$$-k^2 E_x + \frac{\omega^2}{c^2} E_x = -\frac{\omega^2}{\epsilon_0 c^2} P_x \,. \tag{6.45}$$

Let's assume for the moment that Eq. (6.33) is valid. Then

$$P_x = N\epsilon_0\alpha(\omega)E_x\,, \tag{6.46}$$

and E_x drops out of the equation above, which in turn becomes

$$k^2 = \frac{\omega^2}{c^2}\left[1 + N\alpha(\omega)\right]\,. \tag{6.47}$$

This determines the wave number and the - frequency-dependent - refraction index becomes

$$n^2 = 1 + N\alpha(\omega)\,. \tag{6.48}$$

Plugging in the equation for α, Eq. (6.31) results in

$$n^2 = 1 + \frac{Nq_e^2}{m\epsilon_0}\frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega}\,, \tag{6.49}$$

a complex refraction index. Now it's time to resist the temptation to believe that this is the full truth. A couple of things have been omitted so far. First of all, in dense materials, the mobile charges (electrons) do not feel E_x as the polarising field. In fact, E_x has to be supplemented with the polarisation acting in the neighbourhood of the atom or molecule in question. Depending on the density, this polarisation may be of the same order as E_x, and thus cannot be neglected. However, assuming an isotropic material, it can be shown that the average total field at a given point is given by

$$\vec{E}_{\text{local}} = \vec{E} + \frac{\vec{P}}{3\epsilon_0}\,. \tag{6.50}$$

In the polarisation vector the $\vec{E}$ field has to be replaced by this local field and therefore reads

$$\vec{P} = \epsilon_0 N\alpha\vec{E}_{\text{local}} = \epsilon_0 N\alpha\left(\vec{E} + \frac{\vec{P}}{3\epsilon_0}\right) = \frac{N\alpha}{1 - (N\alpha/3)}\epsilon_0\vec{E}\,. \tag{6.51}$$

hence, for sinusoidal electrical fields, the polarisation is proportional to $\vec{E}$ also in dense material, where the average polarisation of neighbours matters.

But in such cases, the constant of proportionality is altered, leading to the following refraction index

$$n^2 = 1 + \frac{N\alpha}{1 - (N\alpha/3)} \tag{6.52}$$

or, written in a more commonly used form,

$$3\frac{n^2 - 1}{n^2 + 2} = N\alpha\,. \tag{6.53}$$

Employing this form, yet another difficulty can be mastered. This additional complication is due to the fact that the linear combination of Eq. (6.32) rather than the simple expression of Eq. (6.31) describes the atomic polarisability. Then

$$3\frac{n^2 - 1}{n^2 + 2} = \frac{Nq_e^2}{m_e}\sum_k \frac{f_k}{(\omega_k^2 - \omega^2) + i\gamma_k\omega}\,. \tag{6.54}$$

Finally, the dielectric could consist of different materials; then a further specification on the different components becomes mandatory.
Let's try to understand how a complex refraction index can be understood. To this end, it is written as

$$n(\omega) = n_R(\omega) - in_I(\omega)\,, \tag{6.55}$$

where n_R and n_I are real functions. The minus sign in front of the imaginary part allows to keep n_I and n_R both positive in ordinary matter. Then, the planar wave above takes the form

$$E_x = E_0 e^{i\omega(t - n_R z/c)} e^{-\omega n_I z/c}\,. \tag{6.56}$$

The first exponential describes a wave propagating with the velocity c/n_R through the medium, hence n_R is what is commonly understood as the refraction index. The amplitude, however, decreases exponentially with z like $E_0 \exp(-\omega n_I z/c)$. Thus, the imaginary part describes the attenuation of the wave due to energy loss. This is clear, since the intensity I of the wave is proportional to the square of the amplitude,

$$I \sim e^{-2\omega n_I z/c} \tag{6.57}$$

such that $\beta = 2\omega n_I/c$ can be understood as the absorption coefficient of the material in question. Thus, in transparent material, such as glass, $c/\omega n_I$ is large compared to its thickness.

Waves in metals

The reasoning so far can be extended also to conductors, like, e.g., metals. There, some of the electrons do not experience a binding force, instead they are "freely" moving inside some specific bands. Of course, the atoms also have some bound electrons beneath, their effect beyond reducing the effective charge of the atomic remnant, however, is not so important for the overall behaviour of metals and will thus be neglected here. When no binding force is present for the electrons, their corresponding equation of motion is the same as before but with $\omega_0 = 0$,

$$\vec{F} = q_e\vec{E} \equiv m_e(\ddot{\vec{x}} + \gamma\dot{\vec{x}}) \,. \tag{6.58}$$

In metals there is yet another difference w.r.t. dielectrics. In dielectrics, there was a difference of the average and the local field due to the fact that in dielectrics isolating material separates fixed dipoles. In contrast, in metals the charges are freely moving and thus not longer fixed, hence the field acting on them is the same both locally and averaged. Therefore, the refraction index is

$$n^2 = 1 + \frac{Nq_e^2}{m\epsilon_0}\frac{1}{-\omega^2 + i\gamma\omega} \,, \tag{6.59}$$

where N is the density of the conducting electrons. But in this case, also γ can be fixed. It is known that the conductivity of metals is limited due to the fact that the conducting electrons perform a kind of Brownian motion, between scatters with the atomic remnants they move on straight paths, accelerated by the electrical field only. The average drift velocity equals the acceleration q_eE/m_e times the average time τ between two scatters, thus

$$v_{\mathrm{drift}} = \frac{q_eE}{m_e}\tau \,. \tag{6.60}$$

This implies that the force and hence E is constant. Since there is no extra acceleration acting on the full ensemble, the dissipative force γmv can be identified with the accelerating force between two scatters, leading to

$$\gamma = \frac{1}{\tau} \,. \tag{6.61}$$

Introducing the conductivity as the connection of current density and electrical field, and using the drift velocity for the determination of the current

$$\vec{j} = \sigma\vec{E} = Nq_e v_{\mathrm{drift}} \tag{6.62}$$

yields

$$\sigma = \frac{Nq_e^2}{m_e}\tau\,. \tag{6.63}$$

Taking everything together allows to express the refraction index in metals through their conductivity, namely

$$n^2 = 1 + \frac{\sigma}{i\epsilon_0\omega(1+\omega\tau)}\,. \tag{6.64}$$

This result implies two different behaviours for very low and very high frequencies. For low frequencies

$$n^2 \approx -i\frac{\sigma}{\epsilon_0\omega}\,, \tag{6.65}$$

thus

$$n = \sqrt{\frac{\sigma}{2\epsilon_0\omega}}\,(1-i)\,. \tag{6.66}$$

The real and imaginary part of the refraction index have the same size, hence the incoming wave is severely attenuated, its amplitude decreases like

$$\exp\left[-\sqrt{\frac{\sigma\omega}{2\epsilon_0 c^2}}\,z\right] = \exp\left[-\frac{z}{\delta}\right]\,, \tag{6.67}$$

where

$$\delta = \sqrt{\frac{2\epsilon_0 c^2}{\sigma\omega}} \tag{6.68}$$

can be interpreted as the frequency-dependent penetration depth.
The approximation above, i.e. neglecting the term $\sim \omega\tau$, also gives some idea of how to interpret the "low frequencies". Apparently, the approximations made break down, unless $\omega \ll 1/\tau$ and $\omega \ll \sigma/\epsilon_0$.
Let's turn the focus now on the behaviour of the refraction index of a metal under high frequencies. Then,

$$n^2 \approx 1 - \frac{\sigma}{\epsilon_0\tau\omega^2}\,. \tag{6.69}$$

The refraction index is real and smaller than one. This can already be seen when checking Eq. (6.59). For large ω, the dissipation γ can be neglected, resulting in

$$n^2 = 1 - \frac{Nq_e^2}{m_e\epsilon_0}\frac{1}{\omega^2}\,. \tag{6.70}$$

Defining the plasma frequency as

$$\omega_P^2 = \frac{Nq_e^2}{m_e\epsilon_0}\,, \tag{6.71}$$

allows to cast the refraction index into the form

$$n^2 = 1 - \frac{\omega_P^2}{\omega^2}\,. \tag{6.72}$$

The plasma frequency thus serves as a critical frequency: For $\omega < \omega_P$ the refraction index has an imaginary and, thus, absorptive part, for $\omega \gg \omega_P$ the metal becomes transparent.

6.2.2 Reflection at surfaces

Boundary conditions

Before entering the discussion let's summarise what has been learned about sinusoidal waves in matter. Each component of the wave, for instance some E, can be written as

$$E(t, \vec{r}) = E_0 e^{i(\omega t - \vec{k}\cdot\vec{r})}\,, \tag{6.73}$$

where E is the amplitude for a given time t and distance $\vec{r}$ from the origin. The vector $\vec{k}$ indicates the direction of propagation of the wave, its absolute value $k = 2\pi/\lambda$ is the wave number. The phase velocity $v_{\rm ph} = \omega/k$, for light in matter with refraction index n, $v_{\rm ph} = c/n$. For sinusoidal waves as above, taking the derivative w.r.t. time of taking the gradient boils down to

$$\frac{\partial}{\partial t} \longrightarrow i\omega \quad \text{and} \quad \vec{\nabla} \longrightarrow -i\vec{k}\,. \tag{6.74}$$

In fact, the latter is valid for each operation with $\vec{\nabla}$. For instance, the z-component of $\vec{\nabla}\times\vec{E}$ reads

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\,. \tag{6.75}$$

if both components are described by sinusoidal waves, then

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -ik_x E_y + ik_y E_x \,, \tag{6.76}$$

as anticipated. For such waves the Faraday equation

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{6.77}$$

becomes

$$-i\vec{k} \times \vec{E} = -i\omega \vec{B} \quad \Longrightarrow \quad \vec{B} = \frac{\vec{k} \times \vec{E}}{\omega} \,. \tag{6.78}$$

This is exactly what has been found for waves in vacuum before.
Let's discuss now, what happens, when a plane wave reaches the boundary between two different materials, with different refraction indices. According to experience, there will be a reflected wave and a wave passing through the boundary and entering the second material. This will be investigated in a bit more detail. Let's choose the orientation of our coordinates such that the boundary is in the yz-plane and the wave travels in the xy-plane. The electric vectors of the incoming, the reflected and the transient waves then read

$$\begin{aligned} \vec{E}_i &= \vec{E}_0 e^{i(\omega t - \vec{k}\cdot\vec{r})} \\ \vec{E}_r &= \vec{E}_0' e^{i(\omega' t - \vec{k}'\cdot\vec{r})} \\ \vec{E}_t &= \vec{E}_0'' e^{i(\omega'' t - \vec{k}''\cdot\vec{r})} \,. \end{aligned} \tag{6.79}$$

For each of the waves the Maxwell equations are valid, hence for each wave the magnetic vector reads

$$\begin{aligned} \vec{B}_i &= \frac{\vec{k} \times \vec{E}_i}{\omega} \\ \vec{B}_r &= \frac{\vec{k}' \times \vec{E}_r}{\omega'} \\ \vec{B}_t &= \frac{\vec{k}'' \times \vec{E}_t}{\omega''} \,. \end{aligned} \tag{6.80}$$

Including the refraction indices for the two materials, n_1 and n_2, establishes the relations of frequency and wave number as

$$\begin{aligned} k^2 &= k_x^2 + k_y^2 = \frac{\omega^2 n_1^2}{c^2} \\ k'^2 &= \frac{\omega^2 n_1^2}{c^2} \\ k''^2 &= \frac{\omega^2 n_2^2}{c^2}\,. \end{aligned} \tag{6.81}$$

So far so good. These are descriptions for the three waves, the only question is, how do they match? Or, stated in other words, how are the various parameters connected? This is the subject of formulating appropriate boundary conditions. Let's start by reconsidering the Maxwell equations in the presence of a dielectric, Eq. (6.37). They read

$$\begin{aligned} \vec{\nabla}\cdot\vec{E} &= -\frac{\vec{\nabla}\cdot\vec{P}}{\epsilon_0} & \vec{\nabla}\cdot\vec{B} &= 0 \\ \vec{\nabla}\times\vec{B} - \epsilon_0\mu_0\frac{\partial\vec{E}}{\partial t} &= \frac{\partial\vec{P}}{\epsilon_0\partial t} & \vec{\nabla}\times\vec{E} + \frac{\partial\vec{B}}{\partial t} &= 0\,. \end{aligned} \tag{6.82}$$

These equations now need to be analysed with focus on what happens in x-direction (this is the direction orthogonal to the boundary of the two materials), more specifically for derivatives w.r.t. x at $x = 0$. What can be said about, say, the x-component of the $\vec{E}$-field? The first equation yields

$$\frac{\partial E_x}{\partial x} = -\frac{1}{\epsilon_0}\,\frac{\partial P_x}{\partial x}\,. \tag{6.83}$$

When integrating over some small x region, including the boundary layer and assuming E and P to be constant in both regions left and right from the boundary results in

$$\epsilon_0(E_{x2} - E_{x1}) = -(P_{x2} - P_{x1}) \quad \text{or} \quad \epsilon_0 E_{x2} + P_{x2} = \epsilon_0 E_{x1} + P_{x1}\,. \tag{6.84}$$

In other words: Any discontinuity in the electric field must be compensated by a corresponding discontinuity in the polarisation vector. The respective sum must be constant: $E_x + P_x$ must be continuous over the boundary, yielding the first boundary condition. The last equation yields

$$E_{z2} = E_{z1} \quad \text{and} \quad E_{y2} = E_{y1}\,. \tag{6.85}$$

The equation $\vec{\nabla} \cdot \vec{B}$ demands continuity of B_x over the boundary,

$$B_{x2} = B_{x1} \, . \tag{6.86}$$

The remaining equation yields the same for the other components of the magnetic field,

$$B_{z2} = B_{z1} \quad \text{and} \quad B_{y2} = B_{y1} \, . \tag{6.87}$$

It should be stressed, however, that this finding $\vec{B}_1 = \vec{B}_2$ is valid only if both materials are not magnetic.

Constructing the waves

Having discussed the formulation of boundary conditions in a kind of "for pedestrians"-manner, let's try to do the same thing a bit more formal. Since results depend on the direction of the electric vector, one may choose two different, independent orientations and then just obtain each other polarisation through linear combination of both. Physically, this means that the ratio of reflected and transient waves depends on the polarisation of the incoming wave.
First, let's analyse the case, where the incoming $\vec{E}$-field has just a z-component [3]. As long as both materials are isotropical, also the induced little oscillations will be in z-direction, and hence $P_x = P_y = 0$. This implies that all electric fields - those of the incoming, the reflected, and the transient wave - have a z-component only; hence in the following vector signs will be omitted. The boundary conditions demand some relations between the E and the B field in both regions, for the electric field this boils down to

$$E_i + E_r = E_t \tag{6.88}$$

at the boundary, i.e. for $x = 0$. Hence

$$E_0 e^{i(\omega t - k_y y)} + E_0' e^{i(\omega' t - k_y' y)} = E_0'' e^{i(\omega'' t - k_y y)} \, , \tag{6.89}$$

valid for all t and y. Specialising on $y = 0$ implies that the frequencies are the same (some thought shows that this is the only way to combine some oscillations),

$$\omega'' = \omega' = \omega \, . \tag{6.90}$$

[3] Remember that each polarisation is perpendicular to the direction of propagation.

Therefore,

$$\frac{k''^2}{n_2^2} = \frac{k'^2}{n_1^2} = \frac{k}{n_1^2}\,. \tag{6.91}$$

Specialising Eq. (6.89) on $t = 0$ implies

$$k''_y = k'_y = k_y\,. \tag{6.92}$$

The equation above also demands that $k'^2 = {k'_x}^2 + {k'_y}^2 = k^2$, hence

$${k'_x}^2 = k_x^2 \quad \longrightarrow \quad k'_x = -k_x\,. \tag{6.93}$$

The other solution, $k'_x = k_x$ would not describe a reflected wave. Taking together, this means that

$$E_r = E'_0 e^{i(\omega t + k_x x - k_y y)}\,. \tag{6.94}$$

For the transient wave

$$k''_y = k_y \quad \text{and} \quad \frac{k''^2}{n_2^2} = \frac{k}{n_1^2}\,, \tag{6.95}$$

which results in

$$k''^2_x = \frac{n_2^2}{n_1^2} k^2 - k_y^2\,. \tag{6.96}$$

assuming for a moment that the refraction indices are real numbers, then all $k's$ are real numbers and from the definition of angles,

$$\frac{k_y}{k} = \sin\theta_i \quad \text{and} \quad \frac{k''_y}{k''} = \sin\theta_t \tag{6.97}$$

follows Snellius' law,

$$n_2 \sin\theta_t = n_1 \sin\theta_1\,. \tag{6.98}$$

Let's turn to the determination of the amplitudes now. From the findings for the ω and $\vec{k}$, the exponential functions can be eliminated and

$$E_0 + E'_0 = E''_0\,. \tag{6.99}$$

But both E_0' and E_0'' are unknown, necessitating yet another relation. After some trying, it becomes apparent that the equation in question is

$$B_{y1} = B_{y2} \tag{6.100}$$

implying

$$B_{yi} = -\frac{k_x E_i}{\omega}\ , \quad B_{yr} = -\frac{k_x' E_r}{\omega'}\ , \quad B_{yt} = -\frac{k_x'' E_t}{\omega''}\ . \tag{6.101}$$

Plugging in the expression for the electric vectors at $x = 0$ yields

$$\frac{k_x}{\omega} E_0 e^{i(\omega t - k_y y)} + \frac{k_x'}{\omega'} E_0' e^{i(\omega' t - k_y' y)} = \frac{k_x''}{\omega''} E_0'' e^{i(\omega'' t - k_y'' y)} \tag{6.102}$$

reducing to

$$k_x E_0 + k_x' E_0' = k_x'' E_0'' \tag{6.103}$$

or

$$\begin{aligned} E_0' &= \frac{k_x - k_x''}{k_x + k_x''} E_0 \\ E_0'' &= \frac{2k_x}{k_x + k_x''} E_0\ . \end{aligned} \tag{6.104}$$

This fixes the amplitudes.
Similar reasoning can be applied for the polarisation, which is in the xy plane, but orthogonal to $\vec{k}$. To keep things simple, it is better to use the magnetic rather than the electric fields. However, after some algebra one finds

$$\begin{aligned} |E_0'| &= \frac{n_2^2 k_x - n_1^2 k_x''}{n_2^2 k_x + n_1^2 k_x''} |E_0| \\ |E_0''| &= \frac{2 n_1 n_2 k_x}{n_2^2 k_x + n_1^2 k_x''} |E_0|\ . \end{aligned} \tag{6.105}$$

Let's investigate the consequences of these findings. For real indices and k's,

$$\begin{aligned} k_x &= k \cos\theta_i = \frac{\omega n_1}{c} \cos\theta_i \\ k_x'' &= k'' \cos\theta_t = \frac{\omega n_2}{c} \cos\theta_t\ . \end{aligned} \tag{6.106}$$

Plugging this into Eq. (6.104) results in

$$\frac{E_0'}{E_0} = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t}\,, \tag{6.107}$$

and inserting Snellius' law yields

$$\frac{E_0'}{E_0} = \frac{\sin\theta_t \cos\theta_i - \sin\theta_i \cos\theta_t}{\sin\theta_t \cos\theta_i + \sin\theta_i \cos\theta_t} = \frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}\,. \tag{6.108}$$

Since the intensities are proportional to the squared field, this implies, for the z-polarisation

$$R = \frac{I_r}{I_i} = \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)}\,. \tag{6.109}$$

Similarly for the polarisation in the xy-plane,

$$R = \frac{I_r}{I_i} = \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)}\,. \tag{6.110}$$

For a wave hitting the boundary perpendicularly,

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2\,. \tag{6.111}$$