ETH Zürich · Institute for Theoretical Physics

# Electrodynamics

Lecture notes

**Babis Anastasiou**

Institute for Theoretical Physics, ETH Zurich,
8093 Zurich, Switzerland
E-mail: babis@phys.ethz.ch

**Abstract**

These are lecture notes for an introductory course on classical electrodynamics. Starting from the qualitative picture of the electromagnetic force, we develop the theory in the following steps:

- electrostatics and boundary-value problems;
- magnetostatics;
- the laws of electrodynamics and electromagnetic waves;
- retarded potentials and the general solution of Maxwell's equations;
- special relativity and the relativistic formulation of electrodynamics;
- the Lagrangian formulation of the electromagnetic field;
- radiation from moving charges and scattering;
- electrodynamics in matter;
- a first glimpse of quantum electrodynamics.

The notes are self-contained, but they are best read alongside a textbook. Recommended companions are the second volume of the Feynman Lectures [1], the book of Jackson [2] and, for the relativistic formulation, the book of Landau and Lifshitz [3]; the book of Weinberg [6] gives an elegant account of special relativity and its tensor methods.

# Contents


1 **Electromagnetic force** 3

2 **Electrostatics** 6
- 2.1 Coulomb's law . . . 6
- 2.2 Gauss' law from Coulomb's law . . . 8
- 2.3 Scalar potential . . . 11
- 2.4 Potential energy of a charge distribution . . . 12
- 2.5 Charged conductors . . . 15

3 **Boundary condition problems in electrostatics** 18
- 3.1 Dirichlet and Neumann boundary conditions . . . 18
- 3.2 Green's functions . . . 20
- 3.3 Explicit solutions of boundary condition problems . . . 22
- 3.4 Green's functions from Laplacian eigenfunctions . . . 28
- 3.5 Laplace operator and spherical symmetry . . . 32
- 3.6 Multipole expansion . . . 39

4 **Magnetic field** 42
- 4.1 Currents . . . 42
- 4.2 Magnetic field of steady currents . . . 44
- 4.3 Applications of Ampère's law . . . 45
- 4.4 Vector potential . . . 46
- 4.5 Wires of a small thickness . . . 48
- 4.6 Magnetic dipole . . . 50
- 4.7 Relativity of the electric and magnetic fields . . . 54

5 **Time-varying electromagnetic fields** 57
- 5.1 Charge conservation . . . 57
- 5.2 Vector and scalar potentials . . . 57
- 5.3 Gauge invariance. . . . 58
- 5.4 Electromagnetic waves in empty space . . . 60

6 **General solutions of Maxwell's equations with sources** 64
- 6.1 Green's functions . . . 64
- 6.2 Fourier transformations . . . 65
- 6.3 Fourier transformations and Green's functions . . . 65
- 6.4 Potential of a moving charge with a constant velocity . . . 70

7 **Special relativity** 72
- 7.1 Proper time . . . 72
- 7.2 Subgroups of Lorentz transformations . . . 74

7.3 Time dilation . . . 76
7.4 Doppler effect . . . 76
7.5 Particle dynamics . . . 77
7.6 Energy and momentum . . . 79
7.7 The inverse of a Lorentz transformation . . . 80
7.8 Vectors and tensors . . . 81
7.9 Currents and densities . . . 83
7.10 Energy-momentum tensor . . . 84

**8 Relativistic formulation of electrodynamics 88**
8.1 Energy-momentum tensor in the presence of an electromagnetic field . . . 92
8.2 Lagrangian formalism of electrodynamics . . . 94
8.3 Maxwell's equations as Euler–Lagrange equations . . . 95

**9 Radiation from moving charges 97**
9.1 The vector potential from a moving charge . . . 97
9.2 The electromagnetic field tensor from a moving charge . . . 101
9.3 Radiation from an accelerated charge in its rest frame: Larmor's formula 104
9.4 Radiation from an accelerated charge with a relativistic velocity . . . 105
9.5 Angular distribution of radiation from a linearly accelerated relativistic charge . . . 107

**10 Scattering 109**
10.1 Thomson scattering . . . 109
10.2 Rayleigh scattering . . . 111

**11 Electrodynamics in a medium 113**
11.1 Average of the atomic charge density . . . 114
11.2 Average of the atomic current density . . . 115
11.3 Maxwell's equations in a medium . . . 117
11.4 Maxwell's equations inside a dielectric material . . . 118
11.5 Waves in a dielectric medium . . . 119
11.6 A model for the dielectric susceptibility $\chi$ . . . 120
11.7 The complex index of refraction . . . 121
11.8 Waves in metals . . . 122
11.9 Reflection and refraction . . . 124
11.10 Cherenkov radiation . . . 128

**12 A glimpse of quantum electrodynamics 132**
12.1 Probability amplitudes . . . 132
12.2 The three basic actions . . . 133
12.3 Thomson and Rayleigh scattering . . . 135
12.4 Partial reflection of light by glass . . . 136
12.5 Summing over the surface: stationary phase . . . 137
12.6 Front and back surfaces . . . 140
12.7 Transmission . . . 141
12.8 Conclusion . . . 142

# 1

# Electromagnetic force

The electromagnetic force has a simple qualitative description. Electrically charged matter is divided into two types of positive and negative charges. Charges of the same type repel each other while charges of different types attract each other. The force between two charges grows when they are brought closer together.

Matter contains enormous numbers of positive (protons) and negative (electrons) charges. The forces exerted by these charges are enormous. However, protons and electrons are found to be very close together in equal numbers and these forces cancel each other almost entirely.

**Exercise:** *Calculate the force exerted on a single electron from only the protons in a gram of sugar placed at a distance of one metre. How does this force compare to the weight of a typical human?*

While positive and negative charges are perfectly balanced at a macroscopic level, charge imbalances are evident at an atomic level. The electric force is responsible for protons and electrons binding into atoms as well as molecular chemical bonds.

One easily realises that the classical qualitative description of the electric force as an attraction or repulsion of electric charges does not work when applied to atoms. Protons and electrons should be collapsing on top of each other due to their attraction. While in orbit, electrons are getting accelerated, and according to classical electrodynamics they should lose energy in the form of radiation, thus falling to lower and lower orbits. To explain the stability of atoms, we shall need to combine the laws of classical electrodynamics with the laws of quantum mechanics. The latter impose a minimum energy for electrons orbiting around protons which cannot be reduced any further. In this course, we will not apply electrodynamics to atomic systems. Nevertheless, there are many interesting macroscopic phenomena which we can understand within the classical theory, without using quantum laws.

The force acting on an electric charge $q$ at a position $\vec{x}$ depends on the relative position and relative motion of all other electric charges. We can sum up the effects of all other charges into two vectors:

- $\vec{E}(\vec{x})$, the electric field and
- $\vec{B}(\vec{x})$, the magnetic field.

The force is then given by:

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right), \tag{1.1}$$

where $\vec{v}$ is the velocity of the charge $q$. The electric and magnetic fields are determined

from the **equations of Maxwell**:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0}, \tag{1.2}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{1.3}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{1.4}$$

$$\vec{\nabla} \times \vec{B} = \frac{\vec{j}}{c^2 \epsilon_0} + \frac{1}{c^2} \frac{\partial \vec{E}}{\partial t}, \tag{1.5}$$

where $\rho(\vec{x})$ is the electric charge density and $\vec{j}$ is the electric current density. $\epsilon_0$ is a constant, the so-called vacuum permittivity, and has the value

$$\epsilon_0 = 8.854\,187\,8128(13) \times 10^{-12} \frac{A \cdot s}{\text{Volt} \cdot m}. \tag{1.6}$$

$c$ is the speed of light,

$$c = 2.997\,924\,58 \times 10^8 \frac{m}{s} \text{ (exact)}. \tag{1.7}$$

It is striking that a property of light, its speed, appears in the equations which govern the forces among charges in wires and capacitors on a laboratory table. The very same equations describe light which was emitted billions of years ago in distant regions of the universe. Maxwell's equations must therefore hold over an enormous range of distances and situations, and by testing them in extreme conditions we will learn where the laws of physics need to be modified. We recall the definition of the differential operator

$$\vec{\nabla} \equiv \left( \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right). \tag{1.8}$$

The inner product of two vectors is

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^{3} A_i B_i, \tag{1.9}$$

while the components of the cross (vector) product of two vectors are

$$\left( \vec{A} \times \vec{B} \right)_i \equiv \sum_{j,k=1}^{3} \epsilon_{ijk} A_j B_k. \tag{1.10}$$

In the above, $\epsilon_{ijk}$ is the fully antisymmetric tensor with $\epsilon_{123} = +1$.

What is a vector? It is not any collection of three numbers. In physics, a vector is an object whose components change under a rotation of our coordinate system exactly as the components of the position vector,

$$x_i' = \sum_{j=1}^{3} R_{ij} x_j, \qquad A_i' = \sum_{j=1}^{3} R_{ij} A_j, \tag{1.11}$$

with the same matrix $R$ for all vectors. A rotation does not change lengths, $\left|\vec{x}'\right|^2 = \left|\vec{x}\right|^2$, which requires that $R$ is an orthogonal matrix, $R^T R = 1$. Taking the determinant of this equation, we find $(\det R)^2 = 1$. A physical rotation can be reached continuously starting

from no rotation at all, $R = 1$, and therefore $\det R = +1$. It is easy to show that the inner product of two vectors does not change under rotations (it is a scalar), while the cross product of two vectors transforms as a vector (**exercise**). Assuming that the charge density is a scalar, the first and the third Maxwell's equations keep exactly the same form in a rotated frame, while both sides of the second and of the fourth equation transform as vectors. Maxwell's equations are thus the same for an experimenter who rotates the table on which he works.

The solutions of Maxwell's equations describe all macroscopic electromagnetic phenomena. First, we will explore these solutions in mathematically simple situations, assuming steady currents, fixed charge distributions and static electric and magnetic fields. In such setups, the electric and magnetic fields decouple from each other in Maxwell's equations. After gaining experience with solving the equations in electrostatics and magnetostatics, we will examine phenomena such as induction and electromagnetic radiation, where the fields vary with time. Finally, we will examine the structure of Maxwell's equations from a purely aesthetic, or rather theoretical, point of view. We will compare the equations in different inertial frames of reference and will discover how they transform under relativistic transformations. We will also explore another symmetry of Maxwell's equations, the so-called gauge symmetry. Strikingly, this symmetry governs the laws of physics at the smallest distances ($\sim 10^{-19}$ m) that we have been able to explore experimentally.

Let us anticipate the outcome. Maxwell's equations keep their form if we move our experiment to another table (translations) or if we rotate it (rotations). However, they change their form if we place the experiment on a moving car and relate the coordinates of the two observers with the Galilean transformation $\vec{x}' = \vec{x} - \vec{v}t$, $t' = t$. All inertial observers must find the same laws of physics. Therefore, either Maxwell's equations or the Galilean transformations must be wrong. Psychologically, it would have been easier to doubt the newly discovered laws of electromagnetism. It turns out, however, that it is the Galilean transformations which are wrong; they must be replaced by the Lorentz transformations of special relativity.

# 2

# Electrostatics

Consider fixed charge distributions and static electric fields. In this situation, the first two of Maxwell's equations take the form:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \qquad \text{(Gauss' law)}, \tag{2.1}$$

$$\vec{\nabla} \times \vec{E} = 0. \tag{2.2}$$

Notice that these equations imply the superposition principle. If two charge distributions $\rho_1$ and $\rho_2$ yield electric fields $\vec{E}_1$ and $\vec{E}_2$ respectively, a charge distribution $\rho = \rho_1 + \rho_2$ produces an electric field $\vec{E} = \vec{E}_1 + \vec{E}_2$.

## 2.1 Coulomb's law

These equations are equivalent to the very familiar **Coulomb's law**. *Remark.* The equivalence requires a boundary condition: $\vec{\nabla} \cdot \vec{E}$ and $\vec{\nabla} \times \vec{E}$ fix $\vec{E}$ only up to a field $-\vec{\nabla}\chi$ with $\nabla^2\chi = 0$ (e.g. a constant external field). Coulomb's law follows from the differential equations only if $\vec{E} \to 0$ at infinity (Helmholtz theorem; Ref. [4], App. B). It states that for two charges $q$ and $q_1$ at positions $\vec{x}$ and $\vec{y}_1$ respectively, the force acting on the charge $q$ is

$$\vec{F} = \frac{q}{4\pi\epsilon_0} \frac{q_1}{\left|\vec{x} - \vec{y}_1\right|^3} (\vec{x} - \vec{y}_1). \tag{2.3}$$

According to Coulomb's law, the electric field due to the charge $q_1$ at a point $\vec{x}$ is

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \frac{q_1}{\left|\vec{x} - \vec{y}_1\right|^3} (\vec{x} - \vec{y}_1). \tag{2.4}$$

Conversely, Coulomb's law follows directly from Gauss' law. For a single point charge $q_1$ at $\vec{y}_1$ the problem has a spherical symmetry around $\vec{y}_1$: the electric field points radially and its magnitude depends only on the distance $r = \left|\vec{x} - \vec{y}_1\right|$. Applying the integral form of Gauss' law (Eq. 2.25, derived below) to a sphere of radius $r$ centred at the charge, we find $4\pi r^2 E(r) = q_1/\epsilon_0$, which is Eq. 2.4. With the superposition principle we can then build up the field of any charge distribution. In order to compute the electric field at a point $\vec{x}$ due to a distribution of $N$ charges $q_i$ at positions $\vec{y}_i$, we invoke the superposition principle [1]:

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N} \frac{q_i}{\left|\vec{x} - \vec{y}_i\right|^3} (\vec{x} - \vec{y}_i). \tag{2.5}$$

[1] We know that the superposition principle is valid down to very small distances. However, at subatomic distances where quantum physics is also at play, we know that it breaks down. The electric field of one charge interacts with the electric field of another charge and the combined electric field is different from the

It is extremely cumbersome to keep track of the numerous charges which are contained in very small volumes of matter. It is then appropriate to define macroscopic continuous charge distributions $\rho(\vec{y})$. The charge $dq$ in a small volume $d^3\vec{y}$ is given by

$$dq = \rho(\vec{y})d^3\vec{y}. \tag{2.6}$$

The electric field is then given by an integral over the charge density:

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\int d^3\vec{y}\frac{\rho(\vec{y})}{\left|\vec{x}-\vec{y}\right|^3}(\vec{x}-\vec{y}). \tag{2.7}$$

---

### 2.1.1 Mathematical interlude: the delta-function

Is there an elegant way to transition from the continuous integral of Eq. 2.7 to the discrete sum of Eq. 2.5 in the case of a distribution of point-like well separated charges? Assume for simplicity the case of only one point-like charge $q$ at a point $\vec{y}$. The charge distribution should be zero at every point in space except $\vec{y}$. On the contrary, at the point $\vec{y}$ where the charge stands, the distribution is infinite, since the volume of a single point is zero ($\rho = q/0$). Dirac introduced a function, the so-called $\delta$-function, which is

$$\delta(x-y) = \begin{cases} 0, & x \neq y, \\ \infty, & x = y. \end{cases} \tag{2.8}$$

In addition, the integrals with a delta function kernel are defined as

$$\int_{-\infty}^{\infty} dx\, f(x)\delta(x-y) = f(y). \tag{2.9}$$

In many ($D$) dimensions, the delta-function is defined as

$$\delta(\vec{x}-\vec{y}) = \prod_{i=1}^{D}\delta(x_i - y_i). \tag{2.10}$$

With the help of the delta-function, the charge density of a single charge $q$ at a position $\vec{y}$ is expressed as:

$$\rho(\vec{x}) = q\delta(\vec{x}-\vec{y}). \tag{2.11}$$

The volume integral over this charge distribution correctly gives

$$\int d^3\vec{x}\rho(\vec{x}) = \int d^3\vec{x}q\delta(\vec{x}-\vec{y}) = q. \tag{2.12}$$

The charge distribution of many point-like charges $q_i$ at positions $\vec{y}_i$ is

$$\rho(\vec{x}) = \sum_i q_i\delta(\vec{x}-\vec{y}_i). \tag{2.13}$$

---

sum. The effect ("light by light scattering") is tiny but very well measured in precision experiments. *Remark.* The QED non-linearity is controlled by field strength rather than distance alone: it becomes of order one near the Schwinger field $E_c = m_e^2c^3/(e\hbar) \approx 1.3\times10^{18}$ V/m, reached at distances comparable to the electron Compton wavelength (Euler–Heisenberg Lagrangian, Ref. [2], Sec. I.7). Light-by-light scattering was observed directly only in 2017 (ATLAS, Nature Phys. 13 (2017) 852) with tens-of-percent precision; the precision tests are indirect, through its loop contributions to the lepton $g-2$.

Substituting the charge density of Eq. 2.13 into Eq. 2.7 we recover the expression of Eq. 2.5.

The transition from discrete charges to a continuous charge density, and back with $\delta$-functions, deserves a physics comment. A smooth density is an approximation: matter is made of atoms, and classical physics cannot tell us how the charge of an electron or a proton is distributed at very short distances, where quantum physics takes over. We can imagine that a "point-like" charge is a tiny cloud of charge of size $a$ around $\vec{y}$, for example with a Gaussian profile, whose exact shape we do not know. All such models must agree on one thing, the total charge $q$. For any physical quantity $f(\vec{x})$ which hardly varies over the size $a$ of the cloud, we have $\int d^3\vec{x}\, f(\vec{x})\rho(\vec{x}) \approx q f(\vec{y})$. The $\delta$-function is the idealisation of such a cloud for $a \to 0$. Using it, we assume that the unknown short-distance physics does not matter for what we compute at larger distances. We will soon meet a case, the self-energy, where this assumption fails.

Integrating with an infinite kernel may be tricky. A $\delta$-function is a distribution, mapping well-behaved test functions into real numbers. Its properties can be derived by a limiting procedure on suitable representation functions. For example, we can think of a delta function as the limit:

$$\delta(x-y) = \lim_{a\to 0} \delta_a(x-y), \tag{2.14}$$

where

$$\delta_a(x-y) = \begin{cases} \frac{1}{a}, & x \in \left[y-\frac{a}{2}, y+\frac{a}{2}\right], \\ 0, & x \notin \left[y-\frac{a}{2}, y+\frac{a}{2}\right]. \end{cases} \tag{2.15}$$

**Exercise:** What is the meaning of

- the derivative of a $\delta$ function?
- $\delta(g(x))$?

**Exercise:** Prove that the derivative of a step function is a delta function.

---

## 2.2 Gauss' law from Coulomb's law

In this section we will show that Gauss' law can be derived from Coulomb's law. First we make the mathematical observation that

$$\vec{\nabla}_{\vec{x}} \frac{1}{|\vec{x}-\vec{y}|} = -\frac{\vec{x}-\vec{y}}{|\vec{x}-\vec{y}|^3}. \tag{2.16}$$

The subscript in the nabla operator denotes that it acts on the components of the vector $\vec{x}$, i.e. $\vec{\nabla}_{\vec{x}} = (\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$. One can easily prove the above by differentiating the left-hand side. Therefore, the electric field is the gradient of a scalar function:

$$\vec{E} = -\vec{\nabla}\Phi(\vec{x}), \tag{2.17}$$

with the so-called scalar potential

$$\Phi(\vec{x}) = \frac{1}{4\pi\epsilon_0} \int d^3\vec{y} \frac{\rho(\vec{y})}{|\vec{x}-\vec{y}|}. \tag{2.18}$$

We can now compute the divergence of the electric field:

$$\vec{\nabla}\cdot\vec{E} = -\nabla^2\Phi = \frac{1}{4\pi\epsilon_0}\int d^3\vec{y}\,\rho(\vec{y})\left(-\nabla^2_{\vec{x}}\frac{1}{|\vec{x}-\vec{y}|}\right). \tag{2.19}$$

We recall the definition of the Laplacian operator,

$$\nabla^2 \equiv \vec{\nabla}\cdot\vec{\nabla} = \sum_{i=1}^{3}\frac{\partial^2}{\partial x_i^2}. \tag{2.20}$$

We will prove shortly that

$$\nabla^2_{\vec{x}}\frac{1}{|\vec{x}-\vec{y}|} = -4\pi\delta(\vec{x}-\vec{y}). \tag{2.21}$$

Notice the pitfall here. A direct differentiation gives $\nabla^2_{\vec{x}}\left|\vec{x}-\vec{y}\right|^{-1} = 0$ for every $\vec{x}\neq\vec{y}$. One could then conclude that the integral vanishes and that $\nabla^2\Phi = 0$ for any charge density, which contradicts Gauss' law. The entire contribution comes from the single point $\vec{x} = \vec{y}$, where the inverse distance is singular, and it must be exactly a $\delta$-function. Thus, we have that

$$\vec{\nabla}\cdot\vec{E} = -\nabla^2\Phi = \frac{\rho}{\epsilon_0}. \tag{2.22}$$

This is the differential form of Gauss' law. **Further reading:** the derivation of Gauss' law from Coulomb's law in Ref. [1], Sections 4-5 and 4-6.

### 2.2.1 Integral form of Gauss' law

Consider a volume $V(S)$ bounded by a surface $S(V)$. We can integrate both sides of Gauss' law over this volume,

$$\int_{V(S)} d^3\vec{x}\vec{\nabla}\cdot\vec{E} = \frac{1}{\epsilon_0}\int_{V(S)} d^3\vec{x}\rho(\vec{x}). \tag{2.23}$$

The integral on the right-hand side is the total charge $Q_{S(V)}$ enclosed in the surface $S(V)$. The left-hand side can be written as a surface integral, by applying the divergence theorem, which states that the flux of a vector $\vec{A}$ through a closed surface $S(V)$ is equal to the divergence of the vector integrated over the volume $V(S)$ enclosed by the surface,

$$\int_{S(V)} d\vec{S}\cdot\vec{A} = \int_{V(S)} d^3\vec{x}\,\vec{\nabla}\cdot\vec{A}. \tag{2.24}$$

Combining the above, we obtain Gauss' law in an integral form, stating that *the flux of the electric field through any closed surface $S(V)$ is proportional to the electric charge*:

$$\int_{S(V)} d\vec{S}\cdot\vec{E} = \frac{1}{\epsilon_0}\int_{V(S)} d^3\vec{x}\rho(\vec{x}). \tag{2.25}$$

Given that this is valid for any closed surface, all the steps that we have made in our derivation are reversible. Therefore, the differential and integral forms of Gauss' law are equivalent.

### 2.2.2 Mathematical interlude: the Laplacian of inverse distance

It is now time to prove the identity of Eq. 2.21 which we used to derive Gauss' law from Coulomb's law. For reasons to become soon apparent, we shall assume a generic number $N$ of space dimensions. At the end, we will specialise to $N = 3$, but keeping the dimension as a generic parameter will elucidate mathematically how to proceed in intermediate steps of our derivation. With a direct differentiation we find that

$$\nabla^2 \frac{1}{|\vec{x} - \vec{y}|} = \frac{3 - N}{|\vec{x} - \vec{y}|^3}. \tag{2.26}$$

In $N = 3$ and for $\vec{x} \neq \vec{y}$, we find that the right-hand side is zero. For $\vec{x} = \vec{y}$ and arbitrary $N$ the result is infinity. What is the result for $N = 3$? It will still be an infinity, as we would expect for a delta function. But let us carefully resolve this $\frac{0}{0}$ conundrum.

Consider the integral in $N$ dimensions:

$$I_N[f] \equiv \int d^N \vec{y}\, f(\vec{y}) \nabla^2_{\vec{x}} \frac{1}{|\vec{x} - \vec{y}|}. \tag{2.27}$$

$f$ is an arbitrary function. Let us now split the integration volume into two regions:

- $V_{in}$ is a sphere (in $N$ dimensions) drawn around the point $\vec{x}$ with a very small radius $R$. Inside the sphere, we can approximate:

$$f(\vec{y}) \approx f(\vec{x}), \quad \forall\, \vec{y} \text{ in } V_{in}. \tag{2.28}$$

- $V_{out}$ is the rest of the integration volume, not included in $V_{in}$.

We first compute the integral $I_N[f]$ inside $V_{in}$. We have

$$\begin{aligned}
I_{N,in}[f] &= \int_{V_{in}} d^N \vec{y}\, f(\vec{y}) \nabla^2_{\vec{x}} \frac{1}{|\vec{x} - \vec{y}|} \\
&= f(\vec{x}) \int_{V_{in}} d^N \vec{y}\, \nabla^2_{\vec{x}} \frac{1}{|\vec{x} - \vec{y}|} + \mathcal{O}(\mathcal{R}) \\
&= (3 - N) f(\vec{x}) \int_{V_{in}} d^N \vec{y}\, \frac{1}{|\vec{x} - \vec{y}|^3} + \mathcal{O}(\mathcal{R}) \\
&= (3 - N) \Omega_N f(\vec{x}) \int_0^R dr\, r^{N-1} \frac{1}{r^3} + \mathcal{O}(\mathcal{R}) \\
&= (3 - N) \Omega_N f(\vec{x}) \frac{R^{N-3} - 0^{N-3}}{N - 3} + \mathcal{O}(\mathcal{R}) \\
&= -\Omega_N f(\vec{x}) R^{N-3} + \mathcal{O}(\mathcal{R}).
\end{aligned} \tag{2.29}$$

$\Omega_N$ is the solid angle in $N$ dimensions. In the above, we have used spherical coordinates in a system centred at $\vec{x}$. For $N = 3$ and setting the radius of $V_{in}$ to zero, $R \to 0$, we obtain:

$$I_{3,in}[f] = -4\pi f(\vec{x}). \tag{2.30}$$

The integral in the rest of the integration volume is:

$$\begin{aligned}
I_{N,out}[f] &= \int_{V_{out}} d^N \vec{y}\, f(\vec{y}) \nabla^2_{\vec{x}} \frac{1}{|\vec{x} - \vec{y}|} \\
&= (3 - N) \int_{V_{out}} d^N \vec{y}\, \frac{f(\vec{y})}{|\vec{x} - \vec{y}|^3}.
\end{aligned} \tag{2.31}$$

The integrand has no singularity, since outside the sphere $V_{in}$ we always guarantee that $\vec{x} \neq \vec{y}$ and it is convergent for the typical smooth functions $f(\vec{x})$ that we are interested in. For $N = 3$, we then have:

$$I_{3,out}[f] = 0. \tag{2.32}$$

We have then found that

$$I_3[f] = I_{3,in}[f] + I_{3,out}[f] = -4\pi f(\vec{x}), \tag{2.33}$$

or, explicitly,

$$\int d^3\vec{y}\, f(\vec{y}) \nabla^2_{\vec{x}} \frac{1}{|\vec{x} - \vec{y}|} = -4\pi f(\vec{x}). \tag{2.34}$$

The above is valid for every smooth function $f(\vec{x})$. We have therefore proven the desired identity:

$$\nabla^2_{\vec{x}} \frac{1}{|\vec{x} - \vec{y}|} = -4\pi\delta(\vec{y} - \vec{x}). \tag{2.35}$$

Keeping the number of dimensions $N$ as a continuous parameter and taking the limit $N \to 3$ at the end, as we did above, is the idea behind *dimensional regularisation*, the standard method for treating divergent integrals in quantum field theory. If this physicist's proof does not satisfy you, there are other ways. For example, the vector field $\vec{\nabla}_{\vec{x}} |\vec{x} - \vec{y}|^{-1} = -(\vec{x} - \vec{y})/|\vec{x} - \vec{y}|^3$ has a flux $-4\pi$ through every sphere centred at $\vec{y}$, independently of its radius. By the divergence theorem, this flux is the integral of $\nabla^2_{\vec{x}} |\vec{x} - \vec{y}|^{-1}$ over the ball (**exercise:** complete this proof).

---

## 2.3 Scalar potential

We have found earlier (Eq. 2.17) that the electric field can be derived from the gradient of the scalar potential. We conclude immediately that the curl of a static electric field is zero:

$$\begin{aligned} \left(\vec{\nabla} \times \vec{E}\right)_i &= -\left(\vec{\nabla} \times \vec{\nabla}\Phi\right)_i = -\sum_{jk} \epsilon_{ijk} \frac{\partial^2 \Phi}{\partial x_j \partial x_k} \\ &= -\frac{1}{2} \sum_{jk} \left(\epsilon_{ijk} + \epsilon_{ikj}\right) \frac{\partial^2 \Phi}{\partial x_j \partial x_k} = 0. \end{aligned} \tag{2.36}$$

We have thus proven that the second law of electrostatics also holds,

$$\vec{\nabla} \times \vec{E} = 0. \tag{2.37}$$

Using Stokes' theorem,

$$\int_S d\vec{S} \cdot \vec{\nabla} \times \vec{A} = \oint_{\partial S} d\vec{l} \cdot \vec{A}, \tag{2.38}$$

we find that the circulation of the electrostatic field in any closed loop $\Gamma$ is zero:

$$\oint_\Gamma d\vec{l} \cdot \vec{E} = 0. \tag{2.39}$$

The scalar potential has an immediate physics interpretation. Consider an electric field $\vec{E}$ and compute the work needed to transport a test charge from a position $\vec{x}_A$ to a position $\vec{x}_B$. The field exerts a force $\vec{F} = q\vec{E} = -q\vec{\nabla}\Phi$ which we need to counteract with an external force $\vec{F}_{\text{ext}} = -\vec{F} = q\vec{\nabla}\Phi$ while transferring the charge from the starting to the finishing point over a given path. The work done is

$$W_{A\to B} = q\int_{\vec{x}_A}^{\vec{x}_B} d\vec{l}\cdot\vec{\nabla}\Phi = q\left(\Phi(\vec{x}_B) - \Phi(\vec{x}_A)\right). \tag{2.40}$$

Nicely, the work done is independent of the path chosen and depends through the potential only on the starting and ending points.

## 2.4 Potential energy of a charge distribution

In this section, we will compute the energy stored in a charge distribution. Assume $N$ charges $q_i$ at positions $\vec{x}_i$. The potential energy of the system is equal to the energy required to bring the charges one by one to their positions from an infinite distance (where the electric field vanishes),

$$W = \sum_{i=1}^{N} W_i, \tag{2.41}$$

where $W_i$ is the work needed to bring the charge $q_i$ to $\vec{x}_i$. It costs no energy to bring the first charge $q_1$ to $\vec{x}_1$. Once at $\vec{x}_1$, $q_1$ creates a potential

$$\Phi_1(\vec{x}) = \frac{1}{4\pi\epsilon_0}\frac{q_1}{|\vec{x}-\vec{x}_1|}. \tag{2.42}$$

To bring $q_2$ to $\vec{x}_2$ requires work

$$W_2 = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{|\vec{x}_1-\vec{x}_2|}. \tag{2.43}$$

Consequently, the new potential is

$$\Phi_{12} = \frac{1}{4\pi\epsilon_0}\left[\frac{q_1}{|\vec{x}-\vec{x}_1|} + \frac{q_2}{|\vec{x}-\vec{x}_2|}\right], \tag{2.44}$$

and the work needed to bring the charge $q_3$ to $\vec{x}_3$ is

$$W_3 = \frac{q_3}{4\pi\epsilon_0}\left[\frac{q_1}{|\vec{x}_3-\vec{x}_1|} + \frac{q_2}{|\vec{x}_3-\vec{x}_2|}\right]. \tag{2.45}$$

For the $i$-th charge the work needed is

$$W_i = \frac{q_i}{4\pi\epsilon_0}\sum_{j<i}\frac{q_j}{|\vec{x}_i-\vec{x}_j|}. \tag{2.46}$$

The potential energy is then

$$W = \frac{1}{4\pi\epsilon_0}\sum_{i=2}^{N}\sum_{j<i}\frac{q_i q_j}{|\vec{x}_i-\vec{x}_j|}. \tag{2.47}$$

Alternatively, we can write a more symmetric form

$$W = \frac{1}{8\pi\epsilon_0} \sum_{i \neq j} \frac{q_i q_j}{\left|\vec{x}_i - \vec{x}_j\right|}. \tag{2.48}$$

**Further reading:** the electrostatic energy of a salt crystal in Ref. [1], Section 8-3.

### 2.4.1 Potential energy of a continuous charge distribution

The expression of Eq. 2.48 in the continuous charge distribution limit becomes

$$W = \frac{1}{8\pi\epsilon_0} \int d^3\vec{x} d^3\vec{y}\, \frac{\rho(\vec{x})\rho(\vec{y})}{\left|\vec{x} - \vec{y}\right|}. \tag{2.49}$$

This can also be written as

$$W = \frac{1}{2} \int d^3\vec{x}\, \rho(\vec{x})\Phi(\vec{x}), \tag{2.50}$$

where we have recognised the scalar potential in the kernel of the $\vec{x}$ integral. Using Gauss' law,

$$\nabla^2\Phi = -\frac{\rho}{\epsilon_0} \tag{2.51}$$

(which in this form is known as the Poisson equation), we have

$$W = -\frac{\epsilon_0}{2} \int d^3\vec{x}\, \Phi(\vec{x})\nabla^2\Phi(\vec{x}). \tag{2.52}$$

We now use the integration-by-parts identity (**exercise**)

$$\Phi\nabla^2\Phi = \vec{\nabla} \cdot \left(\Phi\vec{\nabla}\Phi\right) - \left|\vec{\nabla}\Phi\right|^2 = \vec{\nabla} \cdot \left(\Phi\vec{\nabla}\Phi\right) - \left|\vec{E}\right|^2. \tag{2.53}$$

The first term gives a surface integral (using the divergence theorem) and vanishes if we take the boundary at infinity, where the electric field vanishes. We are then left with the result

$$W = \frac{\epsilon_0}{2} \int d^3\vec{x}\, \left|\vec{E}\right|^2, \tag{2.54}$$

which expresses the electrostatic energy as an integral whose kernel is the squared magnitude of the electric field. We can interpret the positive definite quantity

$$w = \frac{\epsilon_0}{2} \left|\vec{E}\right|^2 \tag{2.55}$$

as the energy density of the electric field. This is also intuitive, since we associate more energy with regions of space where the electric field is stronger.

### 2.4.2 Self-energy

There is a problem with the expression of Eq. 2.49 when applied to discrete charge distributions with a density

$$\rho(\vec{x}) = \sum_i q_i \delta(\vec{x} - \vec{x}_i). \tag{2.56}$$

We obtain a potential energy

$$W = \frac{1}{8\pi\epsilon_0} \sum_{i,j=1}^{N} \frac{q_i q_j}{\left|\vec{x}_i - \vec{x}_j\right|}, \tag{2.57}$$

where in the above summation we do not exclude the cases of $i = j$. These terms yield an infinity. The problem is not overcome either when we apply Eq. 2.54 to discrete cases. For example, if we consider only one point-like charge $q$ the energy density is

$$w = \frac{\epsilon_0}{2} \left|\vec{E}\right|^2 = \frac{q^2}{32\pi^2\epsilon_0} \frac{1}{r^4}, \tag{2.58}$$

which becomes infinite for $r = 0$. This infinity is not integrable. The total energy associated with the existence of a single charge is

$$W = 4\pi \int_0^\infty r^2 dr\, w = -\frac{q^2}{8\pi\epsilon_0} \int_0^\infty d\left(\frac{1}{r}\right) = \infty. \tag{2.59}$$

Obviously, our classical electrodynamics results are not valid when we try to compute the energy associated with point-like particles. The physics laws need to be modified. The problem also persists in quantum electrodynamics, although there we have a technical way to remove the singularities due to the self-energy of point-like charges and absorb them into a renormalisation of the mass of the charged particle. *Remark.* In QED the electron self-energy diverges only logarithmically, $\delta m = \frac{3\alpha}{2\pi} m \ln(\Lambda/m)$ (Weisskopf 1939), instead of linearly as the classical $q^2/(8\pi\epsilon_0 a)$; the electric charge is renormalised separately by vacuum polarisation (Ref. [5], Ch. 7).

The self-energy also resolves a puzzle. The expression of Eq. 2.54 is positive definite. Yet, for a pair of opposite charges $q$ and $-q$, Eq. 2.48 gives a negative energy, as it should: opposite charges attract each other and we gain energy by bringing them together. The two expressions differ exactly by the terms with $i = j$. The continuum expression contains, in addition to the interaction energy, the self-energies of the charges, which are positive and infinite and cannot be compensated by any finite negative contribution. However, the self-energies are present also when the charges are infinitely far apart from each other, where we have set the energy to zero. The physical potential energy is therefore

$$W_{\text{phys}} = W - \sum_i W_{\text{self},i}, \tag{2.60}$$

which agrees with Eq. 2.48.

Where does the infinity come from? It comes from assuming that a charge sits exactly at a point and stays there at rest, a notion which quantum mechanics does not permit. If some unknown short-distance physics smears the charge over a tiny size $\delta$, the self-energy becomes finite, of order $q^2/(4\pi\epsilon_0\delta)$. We do not need to know this physics: the self-energies do not change when we move the charges around, and they drop out of every energy difference. Physical observables depend on energy differences, as we will see in the calculation of the pressure on a conductor below.

In quantum field theory, the Coulomb interaction of two charges arises from the emission of a photon by one charge and its absorption by the other. The self-energy corresponds to a charge emitting and re-absorbing its own photon. This contribution is also divergent. The self-energy problem troubled Feynman a lot, and it was one of his motivations in developing quantum electrodynamics.

**Exercise:** Calculate the electrostatic energy of a uniformly charged sphere in a couple of ways. What happens when you take the limit of a zero radius? You can compare your answer with Ref. [1], Section 8-1.

## 2.5 Charged conductors

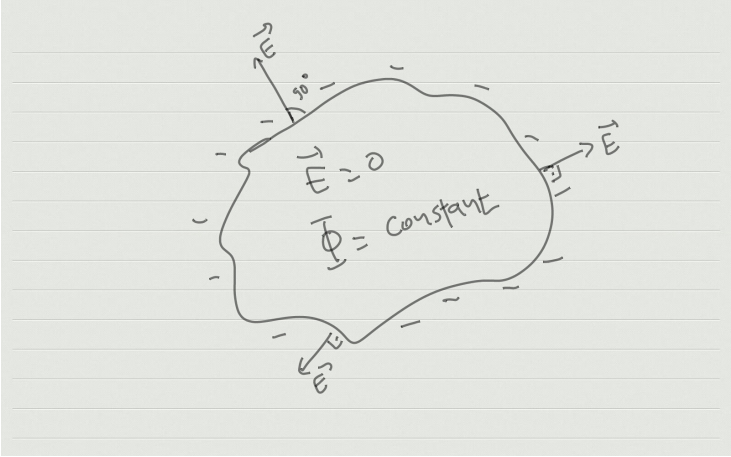



**Figure 2.1.** A conductor charged with some excess charge. The excess charge distributes itself at the surface. The electric field vanishes in the interior and the potential is a constant.

Conductors are materials which allow electrons to move freely within their mass. Commonly, conductors are neutral, containing the same amount of positive and negative charge. However, it is possible to charge them by bringing some excess charge onto them (adding or removing electrons). After we stop charging the conductor, the excess charges will settle in such a way that the electrostatic energy is minimal. This happens when their relative distances are maximal. Therefore, the excess charges distribute themselves on the surface of the conductor. *Remark.* The argument of maximal relative distances is only a heuristic: the equilibrium surface density is not uniform and minimises the energy (Thomson's theorem) rather than the distances. The rigorous argument goes the other way round: $\vec{E}_{\text{inside}} = 0$ (below) and Gauss' law give $\rho = \epsilon_0 \vec{\nabla} \cdot \vec{E} = 0$ in the interior, so any excess charge sits on the surface (Ref. [4], Sec. 2.5.1). Inside a settled charged conductor the electric field vanishes,

$$\vec{E}_{\text{inside}} = 0. \tag{2.61}$$

If it did not, it would push and set in motion the free electrons in the interior (which are balanced by positive ions in the crystal structure) contrary to the requirement that the charges have settled to fixed positions. Why can such a field not persist? A field inside the conductor drives a current. In a real conductor the moving electrons meet resistance and the current dissipates energy as heat. A field which sustained a current forever in an isolated piece of metal would be a perpetual motion machine. Therefore, the free charges rearrange themselves, very quickly in practice, until the field inside vanishes. Notice that the interior of the conductor still contains electrons and positive ions; they balance each other, and only the excess charge resides on the surface. Consequently, the scalar potential is everywhere the same in the conductor:

$$\vec{E}_{\text{inside}} = 0 \rightsquigarrow \vec{\nabla}\Phi = 0 \rightsquigarrow \Phi = \text{constant}. \tag{2.62}$$

The electric field is non-zero on the surface of the conductor and it is perpendicular to the surface. Indeed, a parallel component would set in motion the charges on the surface of the conductor, again contradicting the electrostatic assumption. A perpendicular component,

on the other hand, is allowed: it pushes the surface charges outwards, but they cannot leave the conductor.

We can easily compute the magnitude of the electric field on the surface by applying Gauss' law in the integral form for the surface of Fig. 2.2. The flux of the electric field

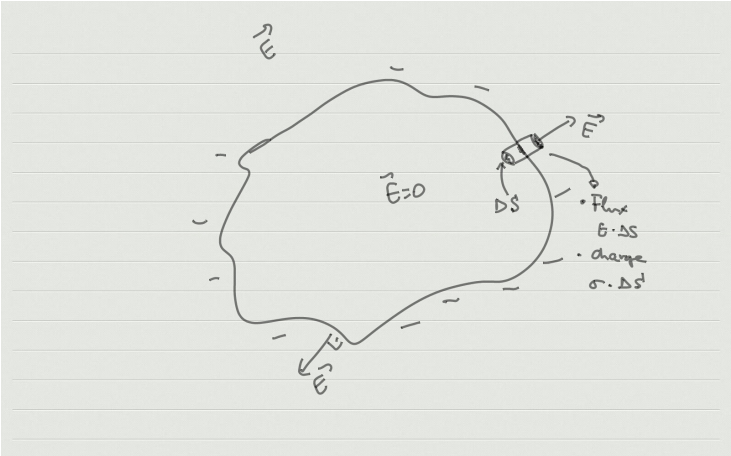



**Figure 2.2.** Using the integral form of Gauss' law in order to compute the electric field on the surface of the conductor.

through the Gaussian surface is

$$\text{Flux} = \int d\vec{S} \cdot \vec{E} = E\Delta S, \tag{2.63}$$

and the charge enclosed is

$$\text{Charge} = \sigma(\vec{x})\Delta S, \tag{2.64}$$

where $\sigma(\vec{x})$ is the charge surface density (charge per unit surface) of the conductor. The electric field on the surface is then

$$E = \frac{\sigma(\vec{x})}{\epsilon_0}. \tag{2.65}$$

The energy density on the surface of the conductor is

$$w = \frac{\epsilon_0}{2}\left|\vec{E}\right|^2 = \frac{\sigma^2(\vec{x})}{2\epsilon_0}. \tag{2.66}$$

The calculation of the pressure is a good example of how to use the energy formula of Eq. 2.54 in spite of the infinite self-energies which it contains. We compare the energies of two configurations with exactly the same charges, so that the self-energies cancel in the difference. Indeed, only the field in the thin volume $\Delta x \Delta S$, which is different in the two configurations, contributes. The excess charges on the surface of a conductor exert a pressure (which we can calculate). If the charges on the surface of the conductor were not constrained and were allowed to increase their separation further, the electrostatic energy would decrease. Consider the original conductor and an otherwise identical one with a small deformation $\Delta x \Delta S$ in its volume, as in Fig. 2.3. The difference in electrostatic energy for the two configurations is

$$\Delta W = \int_{V_{\text{out}}-\Delta x \Delta S} d^3\vec{x} w - \int_{V_{\text{out}}} d^3\vec{x} w = -\Delta S \Delta x \frac{\sigma^2}{2\epsilon_0}. \tag{2.67}$$

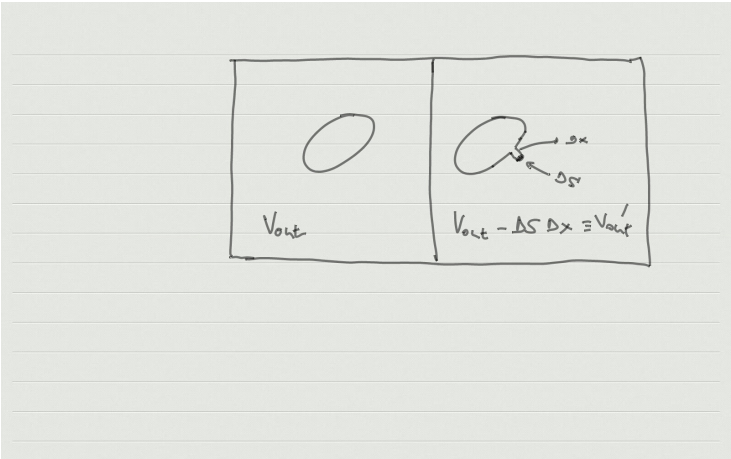



**Figure 2.3.** Two almost identical conductors, with an infinitesimal $\Delta x \Delta S$ volume deformation. The conductors have the same excess charge and charge density. The electrostatic energy density of the two systems is also identical.

The force which is needed to undo such a deformation should provide the same amount of work:

$$\Delta W = F \Delta x. \tag{2.68}$$

Thus, the pressure on the surface of the conductor is

$$\frac{|F|}{\Delta S} = \frac{|\Delta W|}{\Delta x \Delta S} = \frac{\sigma^2}{2\epsilon_0}. \tag{2.69}$$

# 3

# Boundary condition problems in electrostatics

We can claim that we have solved an electrostatics problem if we are able to compute the scalar potential. From it, we can deduce the electric field, the forces on electric charges and everything else we may need to know. The electrostatic potential is given by an integral over all space:

$$\Phi(\vec{x}) = \frac{1}{4\pi\epsilon_0} \int \int \int_{-\infty}^{\infty} d^3\vec{y}\, \frac{\rho(\vec{y})}{|\vec{x} - \vec{y}|}. \tag{3.1}$$

This equation may be impractical for computing the potential in a plethora of situations. It requires an integration over an infinite volume in which we are supposed to have full knowledge of the charge density. It is more likely, however, that we are able to have partial information about the charge density, restricted within the boundaries of a smaller finite volume $V$ (such as the boundaries of a laboratory or an electric device). Is it possible to account for the missing information on the charge density outside the small volume that we control with something equivalent obtained/measured on the boundaries of our volume?

We have expressed Gauss' law in the form of a differential equation,

$$\nabla^2\Phi = -\frac{\rho}{\epsilon_0}. \tag{3.2}$$

Instead of using the infinite integration in order to compute the potential within the volume $V$, we can solve the differential equation. Differential equations yield multiple solutions. To obtain the unique physical solution, we will need to supply additional information, in lieu of the charge density outside $V$. What boundary conditions suffice in order to obtain this unique solution?

## 3.1 Dirichlet and Neumann boundary conditions

Assume that we know two different solutions $\Phi_1$ and $\Phi_2$ of Poisson's equation, Eq. 3.2, inside a volume $V$ with boundary $S(V)$. The difference $U = \Phi_1 - \Phi_2$ satisfies the Laplace equation,

$$\nabla^2 U = 0. \tag{3.3}$$

Therefore,

$$\begin{aligned} U\nabla^2 U &= 0 \\ \rightsquigarrow \vec{\nabla}\cdot\left(U\vec{\nabla}U\right) - \left|\vec{\nabla}U\right|^2 &= 0 \\ \rightsquigarrow \int_V d^3\vec{x}\,\left|\vec{\nabla}U\right|^2 &= \int_V d^3\vec{x}\,\vec{\nabla}\cdot(U\vec{\nabla}U) \\ \rightsquigarrow \int_V d^3\vec{x}\,\left|\vec{\nabla}U\right|^2 &= \int_{S(V)} d\vec{S}\cdot(U\vec{\nabla}U). \end{aligned} \tag{3.4}$$

The first integral is over the entire volume $V$ in which we want to compute the potential. The second integral is over the surface of the boundary.

Let us assume that the potential is fully known, or it can be measured, at the boundary (but not inside the volume):

$$\Phi_1(\vec{x}) = \Phi_2(\vec{x}) = \Phi(\vec{x}) \quad \forall\vec{x} \in S(V). \tag{3.5}$$

Then the right-hand side of Eq. 3.4 vanishes. The left-hand side is a volume integral with a positive definite integrand. In order for this integral to vanish, the integrand must vanish at all points inside $V$. Hence,

$$\vec{\nabla}U = 0 \rightsquigarrow \vec{\nabla}(\Phi_1 - \Phi_2) = 0 \rightsquigarrow \vec{E}_1(\vec{x}) = \vec{E}_2(\vec{x}), \quad \forall\vec{x} \in V. \tag{3.6}$$

We conclude that if we know the scalar potential on the boundary $S(V)$ (Dirichlet boundary condition), the solution of the Poisson equation is unique for the electric field inside the volume $V$. The potential $\Phi(\vec{x})$ inside the volume $V$ is then unique as well: $U$ is constant in $V$ and vanishes on $S(V)$, hence $U = 0$.

Let us assume now that the information that we have on the boundary is the value of the electric field $\vec{E} = -\vec{\nabla}\Phi$ (Neumann boundary condition), thus

$$\vec{\nabla}\Phi_1(\vec{x}) = \vec{\nabla}\Phi_2(\vec{x}) = \vec{\nabla}\Phi(\vec{x}) \quad \forall\vec{x} \in S(V). \tag{3.7}$$

*Remark.* The Neumann condition proper fixes only the normal derivative $\hat{n}\cdot\vec{\nabla}\Phi = \partial\Phi/\partial n$ (the normal component of $\vec{E}$) on $S(V)$; this already makes the surface term $U\,\partial U/\partial n$ in Eq. 3.4 vanish. Prescribing the full gradient, as above, is sufficient but over-determines the problem: the tangential components of $\vec{E}$ on $S$ already fix $\Phi$ on $S$ up to a constant (a Dirichlet condition), and Dirichlet and Neumann data cannot in general be prescribed independently (see Ref. [2], Sec. 1.9). As for a Dirichlet boundary condition, we conclude that $\vec{\nabla}U = 0$ for all points inside the volume $V$ and the solution for the electric field $\vec{E}$ inside $V$ is unique.

A unique solution for $\vec{E}$ inside $V$ is also guaranteed due to Eq. 3.4 for a mixed boundary condition, where for part of the surface $S(V)$ we know the potential $\Phi$ and for the rest of the surface we know the electric field $\vec{E}$.

We note that all three types of boundary conditions (Dirichlet, Neumann, mixed) determine uniquely the solution of Poisson's equation for the electric field $-\vec{\nabla}\Phi$. With Dirichlet or mixed conditions, $U$ is constant in $V$ and vanishes on (part of) $S(V)$, so that $U \equiv 0$ and $\Phi$ itself is unique. With pure Neumann conditions, the scalar potential is determined only up to a physically unimportant constant (see Ref. [2], Sec. 1.9).

## 3.2 Green's functions

From the divergence theorem,

$$\int_V d^3\vec{x}\,\vec{\nabla}\cdot\vec{F} = \int_{S(V)} d\vec{S}\cdot\vec{F}, \tag{3.8}$$

and by choosing

$$\vec{F} = \phi\vec{\nabla}\psi, \tag{3.9}$$

we obtain

$$\int_V d^3\vec{x}\,\vec{\nabla}\cdot\phi\vec{\nabla}\psi = \int_{S(V)} d\vec{S}\cdot\phi\vec{\nabla}\psi. \tag{3.10}$$

Interchanging $\phi$ and $\psi$ in the above and taking the difference, we have

$$\int_V d^3\vec{x}\,\vec{\nabla}\cdot\left(\phi\vec{\nabla}\psi - \psi\vec{\nabla}\phi\right) = \int_{S(V)} d\vec{S}\cdot\left(\phi\vec{\nabla}\psi - \psi\vec{\nabla}\phi\right), \tag{3.11}$$

and performing the differentiations on the left-hand side we obtain

$$\int_V d^3\vec{x}\,\left(\phi\nabla^2\psi - \psi\nabla^2\phi\right) = \int_{S(V)} d\vec{S}\cdot\left(\phi\vec{\nabla}\psi - \psi\vec{\nabla}\phi\right). \tag{3.12}$$

Equation 3.12 will be our main tool for solving Poisson's differential equation with Dirichlet or Neumann boundary conditions.

Let $\phi$ in Eq. 3.12 be a potential $\phi = \Phi(\vec{x})$ satisfying Poisson's equation

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

and $\psi$ a so-called Green's function $G(\vec{x},\vec{y})$ defined to satisfy

$$\nabla^2_{\vec{x}} G(\vec{x},\vec{y}) = -4\pi\delta(\vec{x}-\vec{y}). \tag{3.13}$$

It is of the form

$$G(\vec{x},\vec{y}) = \frac{1}{|\vec{x}-\vec{y}|} + F(\vec{x},\vec{y}), \tag{3.14}$$

where the function $F$ satisfies the Laplace equation:

$$\nabla^2_{\vec{x}} F(\vec{x},\vec{y}) = 0. \tag{3.15}$$

Substituting for $\phi$ and $\psi$ into Eq. 3.12, we obtain

$$\begin{aligned}
&\int_V d^3\vec{x}\,\left[\Phi(\vec{x})\nabla^2_{\vec{x}} G(\vec{x},\vec{y}) - G(\vec{x},\vec{y})\nabla^2\Phi(\vec{x})\right] \\
&\quad = \int_{S(V)} d\vec{S}\cdot\left[\Phi(\vec{x})\vec{\nabla}_{\vec{x}} G(\vec{x},\vec{y}) - G(\vec{x},\vec{y})\vec{\nabla}\Phi(\vec{x})\right] \\
&\rightsquigarrow \int_V d^3\vec{x}\,\left[-4\pi\delta(\vec{x}-\vec{y})\Phi(\vec{x}) + \frac{\rho(\vec{x})}{\epsilon_0} G(\vec{x},\vec{y})\right] \\
&\quad = \int_{S(V)} d\vec{S}\cdot\left[\Phi(\vec{x})\vec{\nabla}_{\vec{x}} G(\vec{x},\vec{y}) - G(\vec{x},\vec{y})\vec{\nabla}\Phi(\vec{x})\right].
\end{aligned}$$

Performing the integration over the $\delta$-function, we finally obtain

$$\begin{aligned}\Phi(\vec{y}) &= \frac{1}{4\pi\epsilon_0}\int_V d^3\vec{x}\, G(\vec{x},\vec{y})\rho(\vec{x}) \\ &\quad -\frac{1}{4\pi}\int_{S(V)} d\vec{S}\cdot\left[\Phi(\vec{x})\vec{\nabla}_{\vec{x}}G(\vec{x},\vec{y}) - G(\vec{x},\vec{y})\vec{\nabla}\Phi(\vec{x})\right].\end{aligned} \tag{3.16}$$

This expression is valid for every point $\vec{y} \in V$.

Let us interpret this result. If we take the simplest Green's function, $G = 1/\left|\vec{x}-\vec{y}\right|$ (i.e. $F = 0$), the first term is Coulomb's law, restricted to the charges inside $V$. The effect of all charges outside $V$ is replaced by the surface integral. This surface term requires both the potential and the electric field on $S(V)$. However, we have proven that only one of the two is needed for a unique solution; knowing both over-constrains the problem. This suggests that we can exploit the freedom in choosing $F$ in order to remove one of the two surface terms. This is what we will do next. Notice that $F$ needs to satisfy the Laplace equation only for $\vec{x}$ inside $V$; outside $V$ it can be anything.

As a sanity check, let $V$ be all space and assume that the charges are localised, so that $\Phi \sim 1/r$ and $\left|\vec{\nabla}\Phi\right| \sim 1/r^2$ at large distances $r$. With $G = 1/\left|\vec{x}-\vec{y}\right|$, both surface terms on a sphere of radius $r$ decrease as $1/r$, since the area of the sphere grows only as $r^2$. For $r \to \infty$ we recover Coulomb's law, Eq. 2.18.

### 3.2.1 Dirichlet boundary conditions

For Dirichlet boundary conditions, we search for a Green's function $G_D(\vec{x},\vec{y})$ which vanishes on the surface $S(V)$:

$$G_D(\vec{x},\vec{y}) = 0, \quad \forall\, \vec{x} \text{ on } S(V), \tag{3.17}$$

such that the second term of the surface integral of Eq. 3.16 vanishes. Then we have

$$\begin{aligned}\Phi(\vec{y}) &= \frac{1}{4\pi\epsilon_0}\int_V d^3\vec{x}\, G_D(\vec{x},\vec{y})\rho(\vec{x}) \\ &\quad -\frac{1}{4\pi}\int_{S(V)} d\vec{S}\cdot\Phi(\vec{x})\vec{\nabla}_{\vec{x}}G_D(\vec{x},\vec{y}).\end{aligned} \tag{3.18}$$

Is $G_D$ unique? Its part $F$ satisfies the Laplace equation inside $V$ with the Dirichlet boundary condition $F(\vec{x},\vec{y}) = -1/\left|\vec{x}-\vec{y}\right|$ for $\vec{x}$ on $S(V)$. This is a special case ($\rho = 0$) of the boundary condition problem that we have studied above, and its solution is unique. We may still add to $G$ some other solution of the Laplace equation, but then the surface term also requires the electric field on the boundary, i.e. more information than necessary.

### 3.2.2 Neumann boundary conditions

With Neumann boundary conditions, one is tempted to use a Green's function with

$$\vec{\nabla}G(\vec{x},\vec{y})\cdot\hat{n} = 0,$$

where $\hat{n}$ is a unit vector perpendicular to $S(V)$. However, such a Green's function cannot exist. From the definition of a Green's function we expect:

$$\begin{aligned}
&\nabla^2 G(\vec{x},\vec{y}) = -4\pi\delta(\vec{x}-\vec{y}) \\
&\rightsquigarrow \int_V d^3\vec{x}\,\nabla^2 G(\vec{x},\vec{y}) = -4\pi \int_V d^3\vec{x}\,\delta(\vec{x}-\vec{y}) \\
&\rightsquigarrow \int_{S(V)} d\vec{S}\cdot\vec{\nabla}G(\vec{x},\vec{y}) = -4\pi \neq 0 \\
&\rightsquigarrow \vec{\nabla}G(\vec{x},\vec{y})\cdot\hat{n} \neq 0.
\end{aligned} \tag{3.19}$$

We can choose, however, a Green's function which has a constant component for its gradient in the direction perpendicular to the surface, such as

$$\hat{n}\cdot\vec{\nabla}_{\vec{x}} G_N(\vec{x},\vec{y}) = -\frac{4\pi}{S}, \quad \forall\,\vec{x}\in S(V), \tag{3.20}$$

where $S$ is the total area of the boundary $S(V)$,

$$S = \int_{S(V)} d\vec{S}\cdot\hat{n}. \tag{3.21}$$

Then, Eq. 3.16 yields:

$$\begin{aligned}
\Phi(\vec{y}) \;=\; & \frac{1}{4\pi\epsilon_0}\int_V d^3\vec{x}\,G_N(\vec{x},\vec{y})\rho(\vec{x}) \\
& + \frac{1}{4\pi}\int_{S(V)} d\vec{S}\cdot G_N(\vec{x},\vec{y})\vec{\nabla}_{\vec{x}}\Phi(\vec{x}) + \langle\Phi\rangle_{S(V)}\,,
\end{aligned} \tag{3.22}$$

where the last term is the average of the potential on the boundary,

$$\langle\Phi\rangle_{S(V)} = \frac{\int_{S(V)} d\vec{S}\cdot\Phi\hat{n}}{S}, \tag{3.23}$$

and it is an unimportant constant for physical observables.

## 3.3 Explicit solutions of boundary condition problems

We have identified required properties of Green's functions for Dirichlet and Neumann boundary conditions. Finding these Green's functions is not easy. In the following, we will consider some simple examples and develop techniques for such a purpose.

What do we gain by trading the Poisson equation for the problem of finding a Green's function? The Green's function knows nothing about the charges; it depends only on the geometry of the boundary. Once we have found it for a given geometry, for example for our laboratory, we can compute by mere integrations the potential for any charge distribution inside it and for any values of the potential on its walls. We solve one geometrical problem and we obtain the solution of a whole class of electrostatic problems. The same Green's functions serve every physics problem which is governed by the Laplace operator, for example Newtonian gravity.

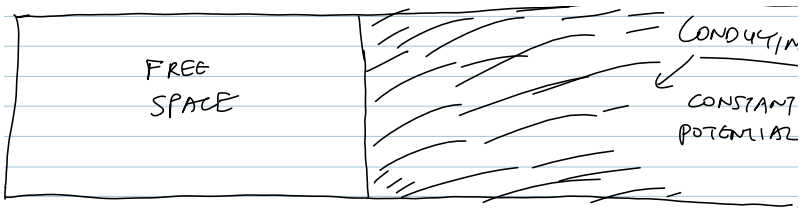



**Figure 3.1.** Half of space filled with a conductor. The potential on the conductor is constant.

### 3.3.1 Example 1: A conductor filling half of space

Suppose that we fill half of space with a conducting material, as in Fig. 3.1. We live in the other half of space where we can have a charge density $\rho(\vec{x})$. The potential of the conductor is constant throughout its mass. The boundary condition for this problem is of Dirichlet type. We require a Green's function which vanishes on the surface $S$ separating the conductor from the half-space we live in:

$$G(\vec{x}, \vec{y}) = 0 \quad \forall\, \vec{x} \text{ on } S. \tag{3.24}$$

$G(\vec{x}, \vec{y})$ has the form

$$G(\vec{x}, \vec{y}) = \frac{1}{|\vec{x} - \vec{y}|} + F(\vec{x}, \vec{y}), \tag{3.25}$$

where

$$\nabla^2 F(\vec{x}, \vec{y}) = 0. \tag{3.26}$$

We choose a coordinate system as in Fig. 3.2. For each vector $\vec{y} = (y_1, y_2, y_3)$, we define a

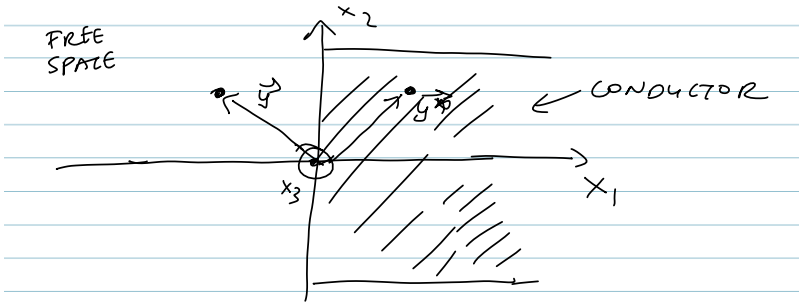



**Figure 3.2.** Half of space filled with a conductor: choice of the coordinate system.

dual vector $\vec{y}^* = (-y_1, y_2, y_3)$. We take $\vec{y}$ to belong to the free space, so that $\vec{y}^*$ belongs to the space occupied by the conductor. The two vectors $\vec{y}$ and $\vec{y}^*$ are identical for $y_1 = 0$, on the surface $S$ separating the free space and the conductor. With some guesswork, we find that

$$F(\vec{x}, \vec{y}) = -\frac{1}{|\vec{x} - \vec{y}^*|} \tag{3.27}$$

satisfies all the criteria we have required for our Green's function. Indeed,

$$G(\vec{x}, \vec{y}) = \frac{1}{|\vec{x} - \vec{y}|} - \frac{1}{|\vec{x} - \vec{y}^*|} \tag{3.28}$$

vanishes on the boundary $\vec{x} = (0, x_2, x_3)$. In addition,

$$\nabla^2 F(\vec{x}, \vec{y}) = -\nabla^2 \frac{1}{\left|\vec{x} - \vec{y}^*\right|} = 4\pi\delta(\vec{x} - \vec{y}^*) = 0. \tag{3.29}$$

The delta function above is always zero since $\vec{x}$ belongs to the free half-space, while $\vec{y}^*$ always belongs to the half-space occupied by the conductor.

Eq. 3.18 yields for the potential:

$$\begin{aligned}\Phi(\vec{y}) &= \frac{1}{4\pi\epsilon_0}\int_{\text{free-space}} d^3\vec{x}\,\frac{\rho(\vec{x})}{\left|\vec{x} - \vec{y}\right|} \\ &\quad -\frac{1}{4\pi\epsilon_0}\int_{\text{free-space}} d^3\vec{x}\,\frac{\rho(\vec{x})}{\left|\vec{x} - \vec{y}^*\right|} \\ &\quad -\frac{1}{4\pi}\int_S d\vec{S}\cdot\Phi(\vec{x})\vec{\nabla}\left[\frac{1}{\left|\vec{x} - \vec{y}\right|} - \frac{1}{\left|\vec{x} - \vec{y}^*\right|}\right].\end{aligned} \tag{3.30}$$

We now make the observation that

$$\left|\vec{x} - \vec{y}^*\right| = \sqrt{(x_1 + y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} = \left|\vec{y} - \vec{x}^*\right|, \tag{3.31}$$

where $\vec{x}^* = (-x_1, x_2, x_3)$. We can then write:

$$\begin{aligned}\Phi(\vec{y}) &= \frac{1}{4\pi\epsilon_0}\int_{\text{free-space}} d^3\vec{x}\,\frac{\rho(\vec{x})}{\left|\vec{x} - \vec{y}\right|} \\ &\quad +\frac{1}{4\pi\epsilon_0}\int_{\text{free-space}} d^3\vec{x}\,\frac{\left(-\rho(\vec{x})\right)}{\left|\vec{x}^* - \vec{y}\right|} \\ &\quad -\frac{1}{4\pi}\int_S d\vec{S}\cdot\Phi(\vec{x})\vec{\nabla}\left[\frac{1}{\left|\vec{x} - \vec{y}\right|} - \frac{1}{\left|\vec{x}^* - \vec{y}\right|}\right].\end{aligned} \tag{3.32}$$

The potential on the surface $S$ is a constant, $\Phi(\vec{x}) = V$. It is left as an exercise to calculate the surface integral in the last term. One finds

$$-\frac{1}{4\pi}\int_S d\vec{S}\cdot\Phi(\vec{x})\vec{\nabla}\left[\frac{1}{\left|\vec{x} - \vec{y}\right|} - \frac{1}{\left|\vec{x}^* - \vec{y}\right|}\right] = V, \tag{3.33}$$

and we have

$$\begin{aligned}\Phi(\vec{y}) &= V + \frac{1}{4\pi\epsilon_0}\int_{\text{free-space}} d^3\vec{x}\,\frac{\rho(\vec{x})}{\left|\vec{x} - \vec{y}\right|} \\ &\quad +\frac{1}{4\pi\epsilon_0}\int_{\text{free-space}} d^3\vec{x}\,\frac{\left(-\rho(\vec{x})\right)}{\left|\vec{x}^* - \vec{y}\right|}.\end{aligned} \tag{3.34}$$

Eq. 3.34 can be used to calculate the potential for arbitrary charge distributions for this geometry. If analytic solutions are not possible, the integrals can be performed with numerical methods.

In the special case of a discrete charge distribution,

$$\rho(\vec{x}) = \sum_i q_i\delta(\vec{x} - \vec{x}_i), \tag{3.35}$$

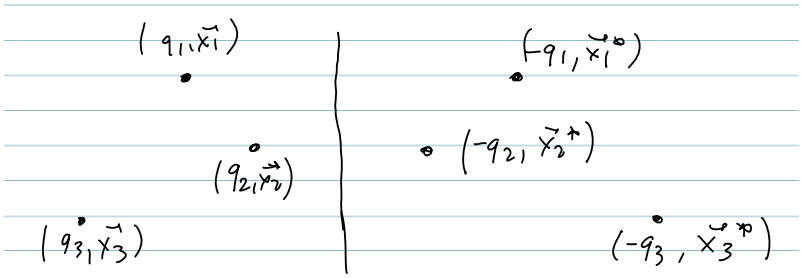



**Figure 3.3.** The potential for discrete charges $q_i$ and an infinite conducting half-space is the same in the free half-space as the potential of the charges $q_i$ and their mirror charges $-q_i$.

where $\vec{x}_i$ lie inside the free volume, we obtain

$$\Phi(\vec{y}) = \frac{1}{4\pi\epsilon_0} \sum_i \left[ \frac{q_i}{|\vec{y} - \vec{x}_i|} + \frac{(-q_i)}{|\vec{y} - \vec{x}_i^*|} \right]. \tag{3.36}$$

This result is familiar! It is the solution of an electrostatic problem different from the one that we have considered here. It is the electrostatic potential due to the charges $q_i$ at positions $\vec{x}_i$ and $-q_i$ charges at reflected (mirrored) positions $\vec{x}_i^*$. The problem of finding the potential for charges $q_i$ at $\vec{x}_i$ in the presence of an infinite plane conductor has exactly the same solution in the free half-volume not occupied by the conductor as the problem of finding the potential for the charges $q_i$ at $\vec{x}_i$ and the "mirror charges" $-q_i$ at $\vec{x}_i^*$ (see Fig. 3.3).

Imagine a physicist who has spent all his life in the free half-space, unaware that the other half is filled with a conductor. Studying the forces among electric charges, he would not discover Coulomb's law but Eq. 3.36: every charge $q_i$ appears to be accompanied by a hidden charge $-q_i$ at the mirror position. Boundaries change what we perceive as the force among electric charges.

The situation is realistic. For a thundercloud, the surface of the earth can be considered as an infinite flat conductor. The electric field above the ground is the field of the charges of the cloud together with the field of their mirror images below the ground. The mirror charges are, however, only a mathematical device. For a cloud at a height of one kilometre, there are no charges one kilometre below the ground. The real charges induced by the cloud sit on the surface of the conductor, distributed exactly so that the potential is constant there; in the free half-space, the mirror charges have the same effect. (Conventionally, the earth is at zero potential; "grounding" a conductor means fixing its potential to zero.) Even if a conducting wall is not infinite, mirror charges give a quick estimate of the electric field close to it.

### 3.3.2 Example 2: Reverse engineering and the method of images (mirror charges)

It is often possible to figure out a clever configuration of charges which has the same effect within some volume $V$ as boundary conditions on the surface $S(V)$ of the volume. As an example (Fig. 3.4), we place a charge $+q$ at a position $\vec{d}$ and a charge of opposite sign $-\frac{R}{d}q$ at a position $\frac{R^2}{d^2}\vec{d}$. The contributions of the two charges to the scalar potential at a position

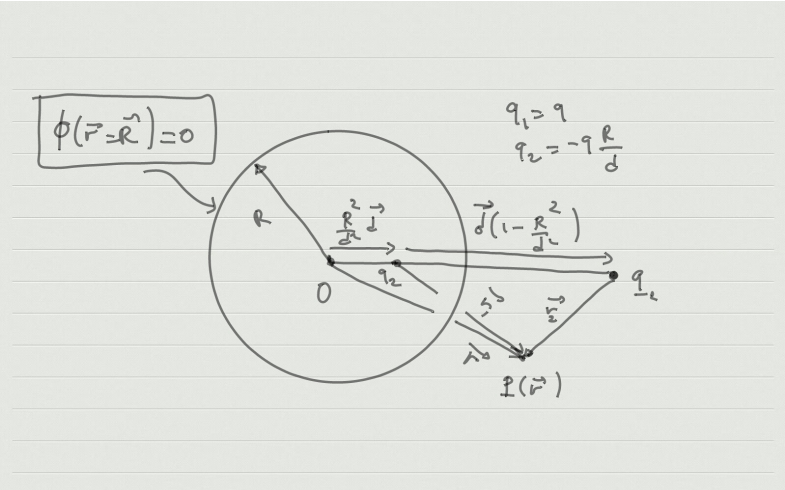



**Figure 3.4.** The potential of a positive and a negative charge. There is a sphere on which the potential is zero.

$\vec{r}$ are of opposite sign:

$$\Phi(\vec{r}) = \frac{q}{4\pi\epsilon_0}\left[\frac{1}{\left|\vec{r}-\vec{d}\right|} - \frac{\frac{R}{d}}{\left|\vec{r}-\frac{R^2}{d^2}\vec{d}\right|}\right]. \tag{3.37}$$

There is a surface on which the two contributions cancel against each other. Indeed, Eq. 3.37 gives zero for all points $\vec{r} = \vec{R}$ on the surface of a sphere with radius $R = \left|\vec{R}\right|$:

$$\begin{aligned}
\Phi(\vec{R}) &= \frac{q}{4\pi\epsilon_0}\left[\frac{1}{\left|\vec{R}-\vec{d}\right|} - \frac{\frac{R}{d}}{\left|\vec{R}-\frac{R^2}{d^2}\vec{d}\right|}\right] \\
&= \frac{q}{4\pi\epsilon_0}\left[\frac{1}{\left|\vec{R}-\vec{d}\right|} - \frac{\frac{R}{d}}{\left[R^2+\frac{R^4}{d^2}-2\frac{R^2}{d^2}\vec{R}\cdot\vec{d}\right]^{\frac{1}{2}}}\right] \\
&= \frac{q}{4\pi\epsilon_0}\left[\frac{1}{\left|\vec{R}-\vec{d}\right|} - \frac{1}{\left[d^2+R^2-2\vec{R}\cdot\vec{d}\right]^{\frac{1}{2}}}\right] \\
&= \frac{q}{4\pi\epsilon_0}\left[\frac{1}{\left|\vec{R}-\vec{d}\right|} - \frac{1}{\left|\vec{R}-\vec{d}\right|}\right] \\
&= 0.
\end{aligned} \tag{3.38}$$

How did we know where to place the second charge and how large to make it? We need a charge of opposite sign, otherwise the potential cannot vanish anywhere at a finite distance. By symmetry, it must lie on the line of $\vec{d}$. Let it be $-\lambda q$ at the position $\varrho\vec{d}$, with $\lambda, \varrho$ unknown. Requiring that the potential vanishes at the two points $\vec{r} = \pm R\,\vec{d}/d$ of the

sphere which lie on this line, we obtain (for $\varrho d < R < d$)

$$\frac{1}{d-R} = \frac{\lambda}{R-\varrho d}, \qquad \frac{1}{d+R} = \frac{\lambda}{R+\varrho d} \quad \rightsquigarrow \quad \lambda = \frac{R}{d}, \quad \varrho = \frac{R^2}{d^2}. \tag{3.39}$$

The calculation above shows that, with these values, the potential vanishes on the whole sphere and not only at these two points.

Now, consider a different electrostatic problem of a conducting sphere of a radius $R$ at zero potential and a charge $+q$ at a distance $d$ from the centre of the sphere. The potential inside the conducting sphere is everywhere zero. What is the potential outside the sphere? The theorem of Eq. 3.18 tells us that once we know the potential at the boundary of a volume the solution of Poisson's equation for the potential is unique within that volume. In our case, the volume we are interested in is bounded by the surface of the sphere of the conductor and the surface of all space at infinity. On both surfaces (the sphere and the surface at infinity) the potential is zero. However, recall that we have identified zero-potential surfaces with exactly the same geometry in our first problem of the two charges in free space. Therefore, the solution of the two problems should be identical in the volume enclosed by the surfaces. Hence, the solution for the problem with the charge $+q$ at a distance $d$ from the centre of a conductor of radius $R$ is

$$\Phi(\vec{r}) = \begin{cases} \frac{q}{4\pi\epsilon_0}\left[\frac{1}{\left|\vec{r}-\vec{d}\right|} - \frac{\frac{R}{d}}{\left|\vec{r}-\frac{R^2}{d^2}\vec{d}\right|}\right], & \forall\,\vec{r}\,:\, r \geq R \\ \\ 0, \quad \forall\,\vec{r}\,:\, r < R. \end{cases} \tag{3.40}$$

Let us compare this solution with the general solution of Eq. 3.18. For the charge density of the point charge, it reads

$$\begin{aligned} \Phi(\vec{r}) \;\; = \;\; & \frac{1}{4\pi\epsilon_0}\int_V d^3\vec{x}\, G_D(\vec{x},\vec{r}) q\delta(\vec{x}-\vec{d}) \\ & -\frac{1}{4\pi}\int_{S(\vec{x}=\vec{R})} d\vec{S}\cdot\Phi(\vec{x})\vec{\nabla}_{\vec{x}} G_D(\vec{x},\vec{r}), \end{aligned} \tag{3.41}$$

or, equivalently,

$$\Phi(\vec{r}) = \frac{q}{4\pi\epsilon_0} G_D(\vec{d},\vec{r}). \tag{3.42}$$

Therefore, the simple solution of the two-charges problem gives us the Green's function for all Dirichlet boundary condition problems outside a sphere of radius $R$. We have

$$G_D(\vec{d},\vec{r}) = \frac{4\pi\epsilon_0}{q}\Phi(\vec{r}) = \frac{1}{\left|\vec{r}-\vec{d}\right|} - \frac{\frac{R}{d}}{\left|\vec{r}-\frac{R^2}{d^2}\vec{d}\right|}. \tag{3.43}$$

With the Green's function and Eq. 3.18 at hand we can solve any other problem with Dirichlet boundary conditions on the same geometry. For example, one could find the solution for the potential when the conductor is kept at a potential $V \neq 0$ and a single charge $+q$ outside the conductor, or the potential in the presence of a spherical conductor and an arbitrary charge distribution inside the free volume (**exercise:** try it for a spherical segment distribution with the same centre as the conductor).

## 3.4 Green's functions from Laplacian eigenfunctions

In this section, we will present a systematic way to obtain Green's functions for Dirichlet boundary conditions on a surface $S(V)$ surrounding a volume $V$. Recall that a Dirichlet Green's function vanishes on $S(V)$.

Why should this work? A physicist's argument comes from quantum mechanics. The momentum operator is $-i\hbar\vec{\nabla}$ and the Hamiltonian of a free particle is $-\frac{\hbar^2}{2m}\nabla^2$. The eigenfunctions of the Laplace operator are, therefore, the energy eigenstates of a free particle; with the condition that they vanish on $S(V)$, they are the energy eigenstates of a particle trapped in a box with impenetrable walls. Every state of the particle is a superposition of energy eigenstates, and we expect that the eigenfunctions of the Laplace operator form a complete basis of functions. The eigenvalues turn out to be real (see below), as energies should. We also gain something in simplicity: the Green's function depends on two points, $\vec{x}$ and $\vec{y}$, while each eigenfunction depends on one point only.

Consider the eigenfunctions $\psi_n(\vec{x})$ of the Laplace operator:

$$\nabla^2\psi_n = \lambda_n\psi_n, \tag{3.44}$$

and impose on them that they vanish on the boundary:

$$\psi_n(\vec{x}) = 0, \quad \forall\vec{x} \in S(V). \tag{3.45}$$

Often, the eigenvalues $\lambda_n$ turn out to be discrete. Let us assume that this is the case for the rest of the analysis. In the continuous cases, our formulae will need to be modified trivially, changing summations into integrations. In addition, we will assume that the eigenvalues are not degenerate (all $\lambda_n$ are different from each other). *Remark.* Non-degeneracy is not needed: eigenfunctions belonging to the same eigenvalue can be chosen orthonormal (Gram–Schmidt), and the construction below goes through unchanged. Degeneracies do occur, e.g. for the parallelepiped below with $a = b$ ($\lambda_{lmn} = \lambda_{mln}$). For Dirichlet conditions all $\lambda_n < 0$, so the division by $\lambda_n$ below is safe.

We now make use of the theorem in Eq. 3.12. We have

$$\int_V d^3\vec{x}\,\left[\psi_m^*\nabla^2\psi_n - \psi_n\nabla^2\psi_m^*\right] = \int_{S(V)} d\vec{S}\cdot\left[\psi_m^*\vec{\nabla}\psi_n - \psi_n\vec{\nabla}\psi_m^*\right]. \tag{3.46}$$

The right-hand side of Eq. 3.46 vanishes since the eigenfunctions $\psi_n$ vanish on the boundary $S(V)$. Thus, we obtain

$$(\lambda_m^* - \lambda_n)\int_V d^3\vec{x}\,\psi_m^*(\vec{x})\psi_n(\vec{x}) = 0. \tag{3.47}$$

For $m = n$, we have that

$$(\lambda_n^* - \lambda_n)\int_V d^3\vec{x}\,\left|\psi_n(\vec{x})\right|^2 = 0. \tag{3.48}$$

Since the above integral is positive definite, we conclude that

$$\lambda_n^* = \lambda_n, \tag{3.49}$$

stating that the eigenvalues are real. For $m \neq n$, we conclude that

$$\int_V d^3\vec{x}\,\psi_m^*(\vec{x})\psi_n(\vec{x}) = 0. \tag{3.50}$$

Choosing appropriately the normalisation of the eigenfunctions, we have the orthonormality condition

$$\int_V d^3\vec{x}\, \psi_m^*(\vec{x})\psi_n(\vec{x}) = \delta_{nm}. \tag{3.51}$$

It is often the case that the eigenfunctions form a complete basis, meaning that any other function $f(\vec{x})$ which vanishes on the boundary $S(V)$ can be written as a linear superposition of the Laplace eigenfunctions:

$$f(\vec{x}) = \sum_n c_n \psi_n(\vec{x}). \tag{3.52}$$

Multiplying with $\psi_m^*(\vec{x})$ and integrating over the volume $V$, we obtain

$$\int_V d^3\vec{x}\, \psi_m^*(\vec{x}) f(\vec{x}) = \sum_n c_n \int_V d^3\vec{x}\, \psi_m^*(\vec{x})\psi_n(\vec{x}) = \sum_n c_n \delta_{nm} = c_m. \tag{3.53}$$

Substituting into Eq. 3.52, we obtain

$$f(\vec{x}) = \int d^3\vec{y}\, f(\vec{y}) \sum_n \psi_n^*(\vec{y})\psi_n(\vec{x}), \tag{3.54}$$

which leads to the completeness condition

$$\sum_n \psi_n^*(\vec{y})\psi_n(\vec{x}) = \delta(\vec{x} - \vec{y}). \tag{3.55}$$

We are now ready to apply the above to a Dirichlet Green's function $G_D(\vec{x}, \vec{y})$, which we can write as a linear superposition of Laplace eigenfunctions:

$$G_D(\vec{x}, \vec{y}) = \sum_n c_n(\vec{y})\psi_n(\vec{x}). \tag{3.56}$$

Applying the Laplace operator on both sides we obtain

$$\begin{aligned}
\nabla^2 G_D(\vec{x}, \vec{y}) &= \sum_n c_n(\vec{y})\nabla^2\psi_n(\vec{x}) \\
\rightsquigarrow -4\pi\delta(\vec{x} - \vec{y}) &= \sum_n c_n(\vec{y})\lambda_n\psi_n(\vec{x}) \\
\rightsquigarrow -4\pi \sum_n \psi_n^*(\vec{y})\psi_n(\vec{x}) &= \sum_n c_n(\vec{y})\lambda_n\psi_n(\vec{x}) \\
\rightsquigarrow c_n(\vec{y}) &= -\frac{4\pi}{\lambda_n}\psi_n^*(\vec{y}).
\end{aligned} \tag{3.57}$$

Knowing the eigenvalues and eigenfunctions of the Laplace operator, we can construct the Green's function via

$$G_D(\vec{x}, \vec{y}) = -4\pi \sum_n \frac{\psi_n^*(\vec{y})\psi_n(\vec{x})}{\lambda_n}. \tag{3.58}$$

### 3.4.1 Example: all (infinite) space

The Green's function with boundaries at infinity is the inverse of the distance of the two vectors in the arguments of the function:

$$G_D(\vec{x},\vec{y}) = \frac{1}{\left|\vec{x}-\vec{y}\right|}. \tag{3.59}$$

We will derive this result from the eigenfunctions of the Laplacian:

$$\nabla^2\psi(\vec{x}) = \lambda\psi(\vec{x}). \tag{3.60}$$

The solutions of this equation (eigenfunctions) are:

$$\psi_{\vec{k}}(\vec{x}) = \frac{1}{(2\pi)^{\frac{3}{2}}}e^{i\vec{k}\cdot\vec{x}}, \tag{3.61}$$

with eigenvalues

$$\lambda_{\vec{k}} = -\left|\vec{k}\right|^2. \tag{3.62}$$

Notice that these eigenvalues are degenerate: all $\vec{k}$ with the same magnitude give the same $\lambda_{\vec{k}}$, in the same way as the kinetic energy of a free particle is the same for all momenta of the same magnitude. The degeneracy is lifted if we solve first the eigenvalue problem of the gradient, $-i\vec{\nabla}\psi_{\vec{k}} = \vec{k}\,\psi_{\vec{k}}$ (the momentum operator, up to a factor $\hbar$). Its eigenfunctions are the plane waves above, labelled by distinct eigenvalues $\vec{k}$, and they are automatically eigenfunctions of $\nabla^2 = -\left(-i\vec{\nabla}\right)^2$. **Exercise:** *Prove that*

$$\int_{-\infty}^{\infty} dx e^{ixa} = 2\pi\delta(a). \tag{3.63}$$

*Hint: think about the Fourier transform of a delta function and the inverse Fourier transform of* 1. These solutions are orthonormal. Indeed,

$$\int d^3\vec{x}\,\psi_{\vec{k}}(\vec{x})^*\psi_{\vec{k}'}(\vec{x}) = \int \frac{d^3\vec{x}}{(2\pi)^3}\,e^{i\vec{x}\cdot\left(\vec{k}'-\vec{k}\right)} = \delta(\vec{k}'-\vec{k}). \tag{3.64}$$

They are also complete:

$$\int d^3\vec{k}\,\psi_{\vec{k}}(\vec{x})^*\psi_{\vec{k}}(\vec{y}) = \int \frac{d^3\vec{k}}{(2\pi)^3}\,e^{i\vec{k}\cdot(\vec{y}-\vec{x})} = \delta(\vec{x}-\vec{y}). \tag{3.65}$$

We have established all conditions required for applying Eq. 3.58 in order to calculate the Green's function from the eigenfunctions of the Laplace operator. We have

$$G(\vec{x},\vec{y}) = -4\pi\int \frac{d^3\vec{k}}{(2\pi)^3\left(-\left|\vec{k}\right|^2\right)}e^{i\vec{k}\cdot(\vec{x}-\vec{y})}. \tag{3.66}$$

We can perform the integrations by writing

$$\vec{k}\cdot\left(\vec{x}-\vec{y}\right) = \left|\vec{k}\right|\left|\vec{x}-\vec{y}\right|\cos\theta, \tag{3.67}$$

using spherical coordinates:

$$d^3\vec{k} = d\left|\vec{k}\right| \left|\vec{k}\right|^2 d\cos\theta\, d\phi, \tag{3.68}$$

and Eq. 3.63. We find

$$G(\vec{x},\vec{y}) = \frac{1}{\left|\vec{x}-\vec{y}\right|}, \tag{3.69}$$

which is the anticipated result. *Remark.* After the angular integrations one is left with $\frac{2}{\pi}\int_0^\infty dk\, \frac{\sin(kr)}{kr}$, $r = |\vec{x}-\vec{y}|$; this is the Dirichlet integral $\int_0^\infty du\ \sin u/u = \pi/2$, which follows from Eq. 3.63 only with some care (e.g. write $\sin u/u = \frac{1}{2}\int_{-1}^{1} dt\, e^{iut}$ and integrate over $u \in (-\infty,\infty)$ first). Equivalently, we have shown that $4\pi/\left|\vec{k}\right|^2$ is the Fourier transform of $1/\left|\vec{x}\right|$. This has been a lot of work in order to recover a result which Coulomb knew more than two centuries ago. The method pays off in geometries for which no simple guess is available, as in the next example.

### 3.4.2 Example: Inside an orthogonal parallelepiped

We will now calculate the Green's function for Dirichlet boundary conditions on the sides of an orthogonal parallelepiped:

$$V:\ x \in [0,a],\ y \in [0,b],\ z \in [0,c]. \tag{3.70}$$

The eigenfunctions of the Laplace operator which vanish on the boundary of the orthogonal parallelepiped are:

$$\psi_{lmn} = \sqrt{\frac{8}{abc}} \sin\left(\frac{l\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right) \sin\left(\frac{n\pi z}{c}\right), \tag{3.71}$$

where $l,m,n = 1,2,\ldots$, with eigenvalues

$$\lambda_{lmn} = -\pi^2\left(\frac{l^2}{a^2}+\frac{m^2}{b^2}+\frac{n^2}{c^2}\right). \tag{3.72}$$

You can check that they are orthonormal and satisfy the completeness condition (**exercise**). The Green's function is then an infinite series:

$$G(\vec{x},\vec{y}) = -4\pi \sum_{l,m,n=1}^{\infty} \frac{\psi_{lmn}(\vec{x})\psi_{lmn}(\vec{y})}{\lambda_{lmn}}. \tag{3.73}$$

This is not an academic example: laboratories, and most modern buildings, are boxes. Notice that already this simplest realistic geometry does not give a Green's function in a simple closed form, but an infinite series. The series is, nevertheless, well defined, and the potential for any charge distribution in the box can be evaluated from it numerically. Exercises are designed so that the answers fit on a page; real-life geometries lead us to more complicated functions.

## 3.5 Laplace operator and spherical symmetry

Let us now consider boundary condition problems with spherical symmetry. Spherical coordinates are the natural choice in many more situations than those with an exact spherical symmetry. Think of a star which has expelled part of its charge, not necessarily isotropically, or of a molecule in which the positive and negative charges are slightly displaced from each other. Seen from far away, such charge distributions look almost point-like, and it would be unnatural to describe them in Cartesian coordinates. We will find that the angular part of the problem has a universal solution, which we need to derive only once. The same angular functions appear in quantum mechanics, in the theory of heat diffusion, and in every problem with the Laplace operator in spherical coordinates.

To find the eigenfunctions of the Laplace operator,

$$\nabla^2\psi = \lambda\psi, \tag{3.74}$$

it is best to work in spherical coordinates $(r, \theta, \phi)$. The Laplace operator takes the form

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2}r + \frac{\hat{A}}{r^2}, \tag{3.75}$$

where $\hat{A}$ is a differential operator acting only on the angles $\theta, \phi$:

$$\hat{A} = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{3.76}$$

Setting $x = \cos\theta$, the angular differential operator takes the form:

$$\hat{A} = \frac{\partial}{\partial x}(1-x^2)\frac{\partial}{\partial x} + \frac{1}{1-x^2}\frac{\partial^2}{\partial\phi^2}. \tag{3.77}$$

We can solve this differential equation by means of the method of separation of variables. We seek factorisable solutions:

$$\psi(r, \theta, \phi) = \frac{R(r)}{r}Y(\theta, \phi), \tag{3.78}$$

where the radial and angular parts are factorised. Notice that not every solution is a product of a radial and an angular function. The factorised ansatz is justified by completeness: if the angular solutions $Y$ that we will find form a complete set of functions of the angles, then every solution can be written as a superposition $\sum R(r)Y(\theta, \phi)/r$ of such products. We must check, at the end, that this is indeed the case. Substituting in Eq. 3.74, we obtain:

$$r^2\left[\frac{1}{R}\frac{\partial^2 R}{\partial r^2} - \lambda\right] = -\frac{1}{Y}\hat{A}Y. \tag{3.79}$$

Since the rhs depends only on angular coordinates and the lhs depends only on the radial coordinate, for them to be equal both must be the same constant $l(l+1)$ (written in such a way for future convenience). We thus have two separate differential equations:

$$\frac{1}{R}\frac{\partial^2 R}{\partial r^2} - \frac{l(l+1)}{r^2} - \lambda = 0, \tag{3.80}$$

and

$$\hat{A}Y = -l(l+1)Y. \tag{3.81}$$

In the following, we solve the above differential equations separately.

### 3.5.1 Radial differential equation

We defer the general solution of the radial differential equation to the exercise classes. Here, we consider the special case of a zero eigenvalue: $\lambda = 0$. Notice that this corresponds to solving $\nabla^2\Psi = 0$, that is, to calculating the potential in empty space. The radial equation becomes:

$$\frac{1}{R}\frac{\partial^2 R}{\partial r^2} = \frac{l(l+1)}{r^2}. \tag{3.82}$$

For

$$R = r^a,$$

we obtain:

$$a(a-1) = l(l+1), \tag{3.83}$$

with solutions:

$$a = -l, 1+l. \tag{3.84}$$

Thus, the general solution of the Laplace equation in spherical coordinates is:

$$\Psi(r,\theta,\phi) = \sum_l \frac{1}{r}\left(A_l r^{-l} + B_l r^{l+1}\right) Y_l(\theta,\phi), \tag{3.85}$$

where the sum could represent an integral if $l$ turns out to be a continuous constant (it will not!). For $\lambda \neq 0$ the radial equation is harder, but it is a differential equation in one variable, which we can always solve, if not in closed form then numerically. Boundary conditions translate directly into conditions on $R(r)$. For example, for Dirichlet boundary conditions on two concentric conducting spheres of radii $r_1 < r_2$, we require $R(r_1) = R(r_2) = 0$. Notice also that the result for $\lambda = 0$ gives the behaviour near the origin for every $\lambda$: for $r \to 0$ the term $\lambda R$ in Eq. 3.80 is suppressed by a factor $r^2$ with respect to the others, so that $R \sim r^{-l}$ or $R \sim r^{l+1}$. If the origin belongs to our volume, only the second behaviour, $\psi \sim r^l$, is regular.

### 3.5.2 Angular differential equation

Eq. 3.81 can also be solved with the method of separation of variables. We write:

$$Y(\theta,\phi) = \Theta(\theta)\Phi(\phi). \tag{3.86}$$

Then, it becomes:

$$\frac{1-x^2}{\Theta}\left[\frac{\partial}{\partial x}(1-x^2)\frac{\partial}{\partial x} + l(l+1)\right]\Theta = -\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2}. \tag{3.87}$$

As before, the two sides must be equal to a constant, which we call $m^2$. This leads to the two differential equations:

$$\frac{\partial^2\Phi}{\partial\phi^2} = -m^2\Phi, \tag{3.88}$$

and

$$\left[\frac{\partial}{\partial x}(1-x^2)\frac{\partial}{\partial x} + l(l+1) - \frac{m^2}{1-x^2}\right]\Theta = 0. \tag{3.89}$$

The differential equation for the azimuthal angle is easily solved, yielding

$$\Phi(\phi) = e^{im\phi}. \tag{3.90}$$

Demanding that this function is single-valued,

$$\Phi(\phi) = \Phi(\phi + 2\pi), \tag{3.91}$$

we obtain that $m$ must be an integer,

$$m = 0, \pm 1, \pm 2, \ldots . \tag{3.92}$$

We first solve the differential equation of Eq. 3.89 for the polar angle $\theta$ in the special case $m = 0$, where it becomes

$$\left[\frac{\partial}{\partial x}(1-x^2)\frac{\partial}{\partial x} + l(l+1)\right]\Theta = 0. \tag{3.93}$$

Consider a polynomial of degree $2l$:

$$p_l(x) = (x^2-1)^l. \tag{3.94}$$

Let us introduce an abbreviation for the differential operator:

$$D \equiv \frac{d}{dx}, \quad D^m \equiv \frac{d^m}{dx^m}.$$

The product rule for multiple derivatives is:

$$D^m(fg) = \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} D^k f D^{m-k} g. \tag{3.95}$$

We start with the identity:

$$(x^2-1)Dp_l(x) = 2lxp_l(x). \tag{3.96}$$

Applying the differential operator $D^{l+1}$ on the above and using the product rule of Eq. 3.95, we arrive at the identity

$$D(1-x^2)D\left(D^l p_l(x)\right) + l(l+1)\left(D^l p_l(x)\right) = 0. \tag{3.97}$$

Therefore, the polynomials $D^l p_l(x)$ are solutions of the differential equation of Eq. 3.93. These polynomials, normalised as

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l, \tag{3.98}$$

are known as the Legendre polynomials. The first few of them are:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= 3/2\,x^2 - 1/2 \\ P_3(x) &= 5/2\,x^3 - 3/2\,x \\ P_4(x) &= \frac{35}{8}x^4 - \frac{15}{4}x^2 + 3/8 \\ &\ldots \end{aligned} \tag{3.99}$$

Why must $l$ be a non-negative integer? A systematic way to solve Eq. 3.93 is the power series ansatz $\Theta = \sum_n \beta_n x^n$. Substituting it, we find the recursion relation

$$\beta_{n+2} = \frac{n(n+1) - l(l+1)}{(n+1)(n+2)} \beta_n. \tag{3.100}$$

The even and the odd powers form two independent series. For large $n$, $\beta_{n+2}/\beta_n \approx 1 - 2/n$, so that $\beta_n \sim 1/n$, and an infinite series diverges logarithmically at $x = \pm 1$, i.e. at the poles $\theta = 0, \pi$. A solution which is finite everywhere on the sphere requires that the series terminates. This happens only if $l(l+1) = n(n+1)$ for some non-negative integer $n$, i.e. for $l = 0, 1, 2, \ldots$ (the other root, $l = -n - 1$, gives the same equation), and only for the series with the same parity as $l$; the other series must be absent. The terminating series are the Legendre polynomials. The Legendre polynomials can be cast as the sum:

$$P_l(x) = \frac{1}{2^l} \sum_{k=0}^{l} \left( \begin{array}{c} l \\ k \end{array} \right)^2 (x-1)^{l-k} (x+1)^k. \tag{3.101}$$

Notice that they are normalised so that

$$P_l(1) = 1. \tag{3.102}$$

Legendre polynomials vanish when integrated with any polynomial of lower degree in the range $[-1, 1]$:

$$\int_{-1}^{1} dx x^k P_l(x) = 0, \quad \forall\, k = 0, 1, \ldots (l-1). \tag{3.103}$$

Indeed,

$$\begin{aligned} \int_{-1}^{1} dx x^k D^l (x^2-1)^l &= \int_{-1}^{1} dx x^k D D^{l-1} (x^2-1)^l \\ &= x^k D^{l-1} (x^2-1)^l \Big|_{-1}^{1} - k \int_{-1}^{1} dx x^{k-1} D^{l-1} (x^2-1)^l. \end{aligned} \tag{3.104}$$

The first term vanishes at the boundary, since

$$D^{l-1} (x^2-1)^l \propto (x^2-1).$$

We can integrate by parts $k < l$ times, always ending up with vanishing boundary terms proportional to $(x^2 - 1)$. It is then obvious that two different Legendre polynomials are necessarily orthogonal:

$$\int_{-1}^{1} dx P_l(x) P_m(x) = 0, \quad \forall\, l \neq m. \tag{3.105}$$

With the above procedure, it is also easy to prove that:

$$\int_{-1}^{1} dx x^l P_l(x) = \frac{l!^2 2^{l+1}}{(1+2l)!}. \tag{3.106}$$

To calculate the normalisation of a Legendre polynomial, we first note that the highest-order term is

$$P_l(x) = \frac{(2l)!}{2^l l!^2} x^l + O(x^{l-1}) \tag{3.107}$$

and therefore

$$\int_{-1}^{1} dx P_l(x)^2 = \frac{(2l)!}{2^l l!^2} \int_{-1}^{1} dx x^l P_l(x) = \frac{2}{1+2l}. \tag{3.108}$$

We then write the orthogonality relation

$$\int_{-1}^{1} dx P_l(x) P_m(x) = \frac{2}{1+2l} \delta_{lm}. \tag{3.109}$$

Legendre polynomials form a basis for all continuous functions $f(x)$ in $[-1, 1]$:

$$f(x) = \sum_{n=0}^{\infty} c_n P_n(x), \tag{3.110}$$

where the coefficients are found as:

$$\begin{aligned} \int_{-1}^{1} dx P_m(x) f(x) &= \sum_{n=0}^{\infty} c_n \int_{-1}^{1} dx P_n(x) P_m(x) \\ \rightsquigarrow \int_{-1}^{1} dx P_m(x) f(x) &= c_m \frac{2}{1+2m}, \end{aligned} \tag{3.111}$$

and finally,

$$c_n = \frac{2n+1}{2} \int_{-1}^{1} dx P_n(x) f(x). \tag{3.112}$$

Inserting this coefficient into Eq. 3.110, we find:

$$f(x) = \int_{-1}^{1} dy f(y) \sum_{n=0}^{\infty} P_n(x) P_n(y) \frac{2n+1}{2}, \tag{3.113}$$

which gives the completeness relation:

$$\sum_{n=0}^{\infty} P_n(x) P_n(y) \frac{2n+1}{2} = \delta(x-y). \tag{3.114}$$

The completeness of the Legendre polynomials is easy to believe. $P_l$ is a polynomial of degree $l$, and we can invert the relations: $1 = P_0$, $x = P_1$, $x^2 = \frac{2}{3}P_2 + \frac{1}{3}P_0$, and in general every monomial $x^n$ is a linear combination of $P_0, \ldots, P_n$. Therefore, every function which can be expanded in a Taylor series in $x$ can also be written, reshuffling the coefficients, as a series of Legendre polynomials. The Legendre polynomials are "as complete as the monomials", with the advantage that they are also orthogonal.

We now turn to the more complicated differential equation of Eq. 3.89 for arbitrary values of $m$. Solutions which are finite in the range $x \in [-1, 1]$ are

$$P_l^m(x) = (-1)^m (1-x^2)^{\frac{m}{2}} \frac{\partial^m}{\partial x^m} P_l(x), \tag{3.115}$$

or,

$$P_l^m(x) = \frac{(-1)^m}{2^l l!} (1-x^2)^{\frac{m}{2}} \frac{\partial^{l+m}}{\partial x^{l+m}} (x^2-1)^l, \tag{3.116}$$

where $m$ is an integer in the range

$$m = -l, -l+1, \ldots, 0, \ldots, l-1, l. \tag{3.117}$$

The polynomials $P_l^m(x)$ are the associated Legendre polynomials. Polynomials with negative values of the integer $m$ are related to polynomials with positive values of the integer $m$ via

$$P_l^{-m} = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x). \tag{3.118}$$

*Remark.* The first form, with $\partial^m P_l / \partial x^m$, is valid only for $m \geq 0$; the second (Rodrigues) form, with $\partial^{l+m}$, holds for all $-l \leq m \leq l$, and the relation for $P_l^{-m}$ follows from it. Also, $P_l^m$ is a polynomial in $x$ only for even $m$ (for odd $m$ it contains $\sqrt{1-x^2}$); for this reason the $P_l^m$ are also called "associated Legendre functions" (see Ref. [2], Eqs. 3.49–3.51). They satisfy the orthogonality relation

$$\int_{-1}^{1} dx P_k^m(x) P_l^m(x) = \frac{2(m+l)!}{(2l+1)(l-m)!} \delta_{kl}. \tag{3.119}$$

**Exercise:** Prove that

$$\int_{-1}^{1} dx \frac{P_l^m(x) P_l^n(x)}{1-x^2} = \frac{(l+m)!}{m(l-m)!} \delta_{m,n}, \quad \text{if } m \neq 0. \tag{3.120}$$

Putting together the polar and azimuthal solutions, we find that the functions:

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} e^{im\phi} P_l^m(\cos\theta) \tag{3.121}$$

are solutions of Eq. 3.81. These functions are called **spherical harmonics** and appear in almost every physics problem with spherical symmetry.

Spherical harmonics are orthogonal. The orthogonality condition reads

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \, \sin\theta Y_{l'm'}^*(\theta, \phi) Y_{lm}(\theta, \phi) = \delta_{l'l} \delta_{m'm}, \tag{3.122}$$

and follows directly from the separate orthogonality conditions of the associated Legendre polynomials (Eq. 3.119) and the azimuthal solutions $e^{im\phi}$:

$$\int_0^{2\pi} \frac{d\phi}{2\pi} e^{-im\phi} e^{im'\phi} = \delta_{m'm}. \tag{3.123}$$

Every function of the polar and azimuthal angles can be written as a linear superposition of spherical harmonics:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_{lm}(\theta, \phi), \tag{3.124}$$

where the coefficients are

$$c_{lm} = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \, \sin\theta Y_{lm}^*(\theta, \phi) f(\theta, \phi). \tag{3.125}$$

Inserting this expression back into Eq. 3.124 we obtain the completeness identity for spherical harmonics:

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_{lm}^*(\theta', \phi') Y_{lm}(\theta, \phi) = \delta(\phi' - \phi) \delta(\cos\theta' - \cos\theta). \tag{3.126}$$

The explicit form of the spherical harmonics can always be looked up in tables. What we must remember is that they are orthonormal and complete. These two properties take us very far. Because of completeness, every function can be written as

$$\psi(r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{R_{lm}(r)}{r} Y_{lm}(\theta,\phi). \tag{3.127}$$

Acting with the Laplace operator of Eq. 3.75, $\hat{A}$ is replaced by its eigenvalue $-l(l+1)$. With the orthogonality of the spherical harmonics, the eigenvalue equation of Eq. 3.74 then yields, for every $l, m$, the one-dimensional differential equation of Eq. 3.80 for $R_{lm}(r)$. The angular dependence is solved once and for all, and every three-dimensional problem becomes a one-dimensional problem in the radial variable. This is why spherical harmonics are so important, here and later in quantum mechanics.

We note that for $m = 0$, the spherical harmonics collapse to the Legendre polynomials:

$$Y_{l0}(\theta,\phi) = \sqrt{\frac{1+2l}{4\pi}} P_l(\cos\theta), \tag{3.128}$$

and are independent of the angle $\phi$. In addition, spherical harmonics with $m \neq 0$ vanish at $\theta = 0$. To see this, recall the factor $(1-x^2)^{m/2} \sim \sin^m\theta$ in the definition of the associated Legendre polynomials $P_l^m(\cos\theta)$.

Finally, we note that spherical harmonics with the same "polar" integer $l$ and opposite "azimuthal" integer $m$ are related by:

$$Y_{l,-m}(\theta,\phi) = (-1)^m Y_{lm}^*(\theta,\phi). \tag{3.129}$$

**Exercise:** Find the eigenfunctions of the Laplace operator in terms of spherical harmonics,

$$\nabla^2 \Psi_\lambda(r,\theta,\phi) = \lambda \Psi_\lambda(r,\theta,\phi) \tag{3.130}$$

with the boundary conditions $\Psi_\lambda(0,\theta,\phi) = \Psi_\lambda(a,\theta,\phi) = 0$. *Remark.* Taken literally, the condition $\Psi_\lambda(0,\theta,\phi) = 0$ excludes the $l = 0$ modes, since $j_0(kr) = \sin(kr)/(kr) \to 1$ at the origin. The appropriate condition at the origin is regularity, i.e. $R(0) = 0$ for $\Psi = R(r)Y/r$; the solutions are then $j_l(kr)Y_{lm}$ with $\lambda = -k^2$ and $j_l(ka) = 0$.

### 3.5.3 Expansion of inverse distance in Legendre polynomials

Consider two vectors:

$$\vec{r}_L, \vec{r}_S, \quad r_L > r_S.$$

Their distance is

$$\left|\vec{r}_L - \vec{r}_S\right| = \left[r_L^2 + r_S^2 - 2r_L r_S \cos\theta\right]^{\frac{1}{2}}, \tag{3.131}$$

where $\theta$ is the angle between the two vectors. Any function of $\cos\theta$ can be expanded as a series in Legendre polynomials. In particular, for the inverse of a distance we have

$$\frac{1}{\left|\vec{r}_L - \vec{r}_S\right|} = \sum_{l=0}^{\infty} c_l P_l(\cos\theta). \tag{3.132}$$

To compute the coefficients, we consider the special case of two parallel vectors, $\theta = 0$. Recall that Legendre polynomials are normalised to be one for a unit argument:

$$P_l(\cos 0) = P_l(1) = 1.$$

Then

$$\sum_{l=0}^{\infty} c_l = \frac{1}{r_L - r_S} = \sum_{l=0}^{\infty} \frac{r_S^l}{r_L^{l+1}} \rightsquigarrow c_l = \frac{r_S^l}{r_L^{l+1}}. \tag{3.133}$$

Strictly speaking, the value at $\theta = 0$ fixes only the sum $\sum_l c_l$, not the individual $c_l$. The identification is justified because $1/|\vec{r}_L - \vec{r}_S|$ solves the Laplace equation in $\vec{r}_S$: by the general solution above, $c_l = A_l\, r_S^l$ (regularity at $r_S = 0$), and the $\theta = 0$ series in powers of $r_S$ then determines $A_l = r_L^{-l-1}$ uniquely (see Ref. [2], Sec. 3.3). We then have the result

$$\frac{1}{\left|\vec{r}_L - \vec{r}_S\right|} = \sum_{l=0}^{\infty} \frac{r_S^l}{r_L^{l+1}} P_l(\cos\theta). \tag{3.134}$$

For a general power $a$ the expansion is not in Legendre polynomials but in Gegenbauer polynomials $C_l^{(a/2)}$:

$$\frac{1}{\left|\vec{r}_L - \vec{r}_S\right|^a} = \sum_{l=0}^{\infty} \frac{r_S^l}{r_L^{l+a}} C_l^{(a/2)}(\cos\theta), \qquad C_l^{(1/2)} = P_l, \tag{3.135}$$

which at $\theta = 0$ reduces to the binomial series with $C_l^{(a/2)}(1) = (a,l)/l!$, where the Pochhammer symbol is defined as:

$$(a,l) = \frac{\Gamma(a+l)}{\Gamma(a)}. \tag{3.136}$$

## 3.6 Multipole expansion

Consider a charge distribution $\rho(\vec{x})$ which occupies a small volume $V'$. We are interested in the potential that this distribution creates at a point $\vec{r}$ outside the region of the charge distribution. Intuitively, we expect that if we are sufficiently far from the charge distribution the potential will resemble a Coulomb potential. This is correct to a first approximation, but not exact; there are corrections, which become more important the closer we approach the charges of the distribution.

The potential is given by

$$\Phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \int_{V'} d^3\vec{x} \frac{\rho(\vec{x})}{\left|\vec{x} - \vec{r}\right|}. \tag{3.137}$$

Since $r > x$, we can expand the inverse distance in the integrand in $\frac{x}{r}$, obtaining:

$$\Phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \sum_{l=0}^{\infty} \frac{1}{r^{l+1}} \int_{V'} d^3\vec{x}\, \rho(\vec{x}) x^l P_l(\cos\gamma), \tag{3.138}$$

where $\gamma$ is the angle formed by the two vectors $\vec{x}$ and $\vec{r}$. We write the two vectors as:

$$\vec{x} \equiv (x, \theta_x, \phi_x), \quad \vec{r} \equiv (r, \theta, \phi). \tag{3.139}$$

It is an easy geometry exercise to prove that:

$$\cos\gamma = \frac{\vec{x}\cdot\vec{r}}{xr} = \cos\theta\cos\theta_x + \sin\theta\sin\theta_x\cos(\phi - \phi_x). \tag{3.140}$$

We would like to perform the $d^3\vec{x} = x^2 dx d\Omega_x$ integration in spherical coordinates. However, the expression for $\cos\gamma$ in terms of $\theta_x, \phi_x$ seems overly complicated. To our rescue comes the following identity, called the "addition theorem":

$$P_l(\cos\gamma) = \frac{4\pi}{1+2l} \sum_{m=-l}^{l} Y^*_{lm}(\theta_x, \phi_x) Y_{lm}(\theta, \phi). \tag{3.141}$$

It is a consequence of the fact that the lhs is a function of the angular coordinates and can therefore be expressed as a linear superposition of spherical harmonics. It is also manifest that the expression is symmetric under the exchange: $(\theta, \phi) \leftrightarrow (\theta_x, \phi_x)$. Why does the sum contain only spherical harmonics with the same $l$? The angular operator $\hat{A}$ does not change under rotations of the coordinate system. In a rotated system whose $z$-axis points along $\vec{x}$, the function $P_l(\cos\gamma)$ is a Legendre polynomial of the polar angle of $\vec{r}$, i.e. an eigenfunction of $\hat{A}$ with eigenvalue $-l(l+1)$. It remains an eigenfunction with the same eigenvalue in the original coordinates, and therefore it can only be a superposition of spherical harmonics $Y_{lm}(\theta, \phi)$ with the same $l$. In quantum mechanics, $-\hbar^2\hat{A}$ is the square of the angular momentum; a rotation does not change the total angular momentum $l$, it only mixes its projections $m$ on the $z$-axis. We will not prove this theorem in full here; it is proven more easily with the tools of quantum mechanics. Substituting into Eq. 3.138, we obtain:

$$\Phi(\vec{r}) = \frac{1}{\epsilon_0} \sum_{l=0}^{\infty} \frac{1}{1+2l} \sum_{m=-l}^{l} \left[ \int_{V'} x^2 dx d\Omega_x Y^*_{lm}(\theta_x, \phi_x) \rho(\vec{x}) x^l \right] \frac{Y_{lm}(\theta, \phi)}{r^{1+l}}, \tag{3.142}$$

or, in a more compact form:

$$\Phi(\vec{r}) = \frac{1}{\epsilon_0} \sum_{l=0}^{\infty} \frac{1}{1+2l} \sum_{m=-l}^{l} \frac{q_{lm} Y_{lm}(\theta, \phi)}{r^{1+l}}, \tag{3.143}$$

where we have defined the quantities:

$$q_{lm} = \int_{V'} d^3\vec{x}\, Y^*_{lm}(\theta_x, \phi_x) \rho(\vec{x}) x^l, \tag{3.144}$$

which are called multipole moments. The moments characterise the geometry of the charge distribution. For example,

$$q_{00} = \int d^3\vec{x} \rho(\vec{x}) Y_{00}(\theta, \phi) = \frac{1}{\sqrt{4\pi}} \int d^3\vec{x} \rho(\vec{x}) = \frac{Q}{\sqrt{4\pi}}, \tag{3.145}$$

which is proportional to the total charge in the distribution. The $l = 0$ term is the Coulomb potential of the total charge placed at the origin. At very large distances every localised charge distribution looks like a point charge, which limits our ability to tell distributions apart. A neutral distribution, such as a molecule with its positive and negative charges slightly displaced from each other, has $q_{00} = 0$. Its leading term is then the dipole term ($l = 1$), which falls as $1/r^2$, faster than the Coulomb potential, and depends on the direction of observation relative to the orientation of the dipole. For the higher moments, we find:

$$q_{11} = -\sqrt{\frac{3}{8\pi}}(p_1 - ip_2), \tag{3.146}$$

$$q_{10} = \sqrt{\frac{3}{4\pi}}(p_3), \tag{3.147}$$

with

$$\vec{p} = (p_1, p_2, p_3) = \int d^3\vec{x}\, \vec{x}\rho(\vec{x}) \tag{3.148}$$

the dipole moment. For the next moments, we find

$$q_{22} = \frac{1}{12}\sqrt{\frac{15}{2\pi}}(Q_{11} - 2iQ_{12} - Q_{22}), \tag{3.149}$$

$$q_{21} = -\frac{1}{3}\sqrt{\frac{15}{8\pi}}(Q_{13} - iQ_{23}), \tag{3.150}$$

$$q_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}Q_{33}, \tag{3.151}$$

where

$$Q_{ij} = \int d^3\vec{x}\,(3x_i x_j - x^2\delta_{ij})\rho(\vec{x}) \tag{3.152}$$

is the quadrupole tensor, and so on for higher $l$. The most important moments are the first ones, since the higher moments are suppressed by powers of $1/r$ in their contribution to the potential. A measurement of the angular distribution of the potential due to some unknown charge distribution can be used for the extraction of the multipole moments of the distribution. This is our best way to learn something about the geometrical characteristics of the distribution.

This "inverse problem" is not specific to electrostatics. In cosmology, the temperature of the cosmic microwave background radiation, about 2.7 K, is measured in every direction of the sky. Its tiny variations from one direction to another (of the order of $10^{-5}$, after removing the dipole due to our own motion) are expanded in spherical harmonics, and the multipole coefficients are compared with theory. A similar analysis is performed for the distribution of galaxies. These variations originate from small inhomogeneities in the density of the very early, hot and dense universe, which grew under the action of gravity, a force which in the weak-field limit obeys a law analogous to Coulomb's law. The primordial inhomogeneities are believed to be quantum fluctuations. By measuring multipole moments in the sky, we learn about the quantum physics at the very beginning of the universe.

**Exercise:** *Compute the multipole moments $q_{lm}$ of (i) a point charge $q$ at the position $a\hat{z}$ and (ii) two charges $\pm q$ at the positions $\pm\frac{a}{2}\hat{z}$. Which moments vanish? Write the leading term of the potential at large distances in each case.*

# 4

# Magnetic field

The force acting on a charge $q$ which moves with velocity $\vec{v}$ inside an electromagnetic field $(\vec{E}, \vec{B})$ is:

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right). \tag{4.1}$$

The direction of the electric component of the force is along or against the direction of the electric field $\vec{E}$. The magnetic component of the force is at right angles to both the magnetic field and the direction of motion of the charge.

## 4.1 Currents

Macroscopically, it is easy to observe the magnetic force acting on a large number of charges which move together inside a magnetic field $\vec{B}$. Currents can be realised inside conductors, which allow electrons to move freely within their body.

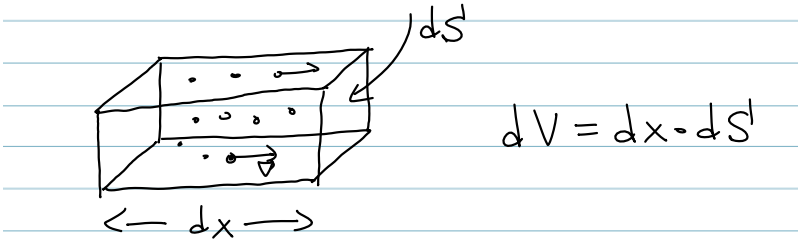



**Figure 4.1.** A small volume element $dV = dx\, dS$ containing moving charges.

Let us consider a volume element $dV = dx\, dS$ (Fig. 4.1) with a charge density

$$\rho = \frac{\Delta Q_0}{dV}$$

at an initial time $t_0$. $\Delta Q_0$ is the total charge contained inside $dV$ at $t_0$. Assume that all elementary charges are identical, with charge $q$. Then

$$\Delta Q_0 = N_0 q, \tag{4.2}$$

where $N_0$ is their number inside $dV$ at $t_0$. Let us assume that all charges are moving with a velocity $\vec{v}$ (which we have aligned with the $x$-axis). In a small time $dt$, a number of charges move out of the volume $dV$ (Fig. 4.2). The charge they carry is

$$qN_{\text{escaping}} = \rho(\vec{v}dt) \cdot d\vec{S}. \tag{4.3}$$

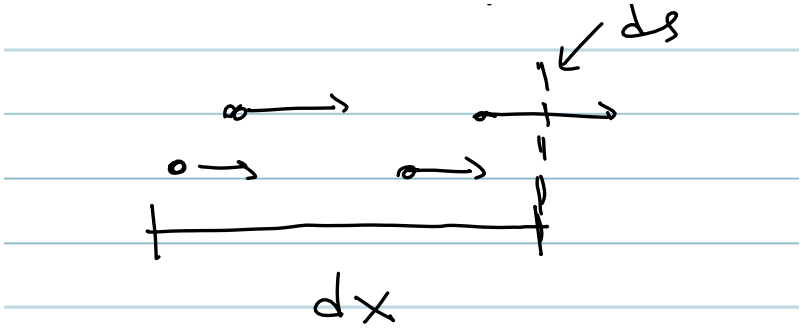



**Figure 4.2.** Charges moving with velocity $\vec{v}$ along the $x$-axis. In a small time $dt$, some of them cross the surface $d\vec{S}$ and leave the volume.

We define the current density $\vec{J}$ as the charge traversing a surface $d\vec{S}$ per unit time and unit area. Specifically,

$$qN_{\text{escaping}} = \vec{J} \cdot d\vec{S}\, dt. \tag{4.4}$$

By comparing the last two equations we conclude that:

$$\vec{J} = \rho\vec{v}. \tag{4.5}$$

After time $dt$, the remaining charge inside the volume $dV$ is

$$\begin{aligned} \Delta Q_{t_0+dt} &= \Delta Q_0 - qN_{\text{escaping}} \\ \rightsquigarrow \frac{\Delta Q_{t_0+dt} - \Delta Q_0}{dt} &= -\vec{J} \cdot d\vec{S} \\ \rightsquigarrow dV\frac{\partial \rho}{\partial t} &= -\vec{J} \cdot d\vec{S}. \end{aligned} \tag{4.6}$$

Eq. 4.6 states that the change in the charge density during $dt$ is equal to minus the flux of the current density through the surface $d\vec{S}$. *Remark.* The balance above counts only the charge leaving through the front face $d\vec{S}$. For a uniform flow, the same amount enters through the back face, so the correct statement for the cell is $dV\,\partial_t\rho = -\oint_{\partial(dV)} \vec{J} \cdot d\vec{S}$, the *net* outward flux through all its faces. This is what the integration over $S(V)$ below uses (see Ref. [2], Sec. 5.2). Integrating over the surface $S(V)$ of an arbitrary volume $V$, we have:

$$\begin{aligned} \int_V dV\frac{\partial \rho}{\partial t} &= -\int_{S(V)} \vec{J} \cdot d\vec{S} \\ \rightsquigarrow \int_V d^3\vec{x}\,\frac{\partial \rho}{\partial t} &= -\int_V d^3\vec{x}\,\vec{\nabla} \cdot \vec{J} \\ \rightsquigarrow \int_V d^3\vec{x}\,\left[\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{J}\right] &= 0. \end{aligned} \tag{4.7}$$

Eq. 4.7 must hold for **any** volume $V$. For this to happen, the integrand must vanish. Therefore, we arrive at the continuity equation:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{J} = 0. \tag{4.8}$$

The above equation is a consequence of charge conservation.

## 4.2 Magnetic field of steady currents

We will now study a very common situation where the charge density is constant:

$$\frac{\partial \rho}{\partial t} = 0.$$

This is realised in simple circuits, where charge flows into and out of any volume element at the same rate. The divergence of the electric current density is therefore:

$$\vec{\nabla} \cdot \vec{J} = 0. \tag{4.9}$$

In such situations, the electric and magnetic fields are constant in time. The magnetic field is determined by two of Maxwell's equations, which become

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{4.10}$$

and

$$c^2 \vec{\nabla} \times \vec{B} = \frac{\vec{J}}{\epsilon_0}. \tag{4.11}$$

We can derive integral forms of these equations. From Eq. 4.10, we obtain

$$\int_V d^3\vec{x}\, \vec{\nabla} \cdot \vec{B} = 0 \rightsquigarrow \int_{S(V)} \vec{B} \cdot d\vec{S} = 0. \tag{4.12}$$

Therefore, the flux of the magnetic field out of any volume is zero. Recall that in electrostatics,

$$\int_{S(V)} \vec{E} \cdot d\vec{S} = \frac{Q_{\text{inside } V}}{\epsilon_0}.$$

The analogous Eq. 4.12 states that there is no magnetic charge to be found anywhere.

Integrating Eq. 4.11 over an open surface $S(\Gamma)$ bounded by a closed loop $\Gamma$ (Fig. 4.3), we

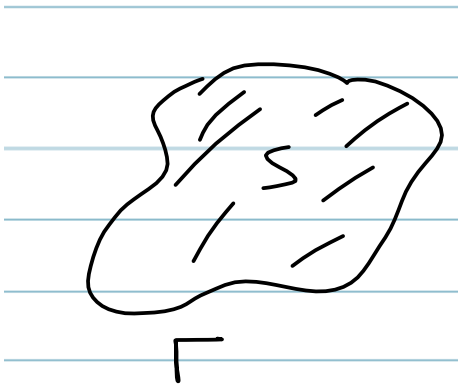



**Figure 4.3.** An open surface $S(\Gamma)$ bounded by a closed loop $\Gamma$.

obtain

$$\begin{aligned} & c^2 \int_{S(\Gamma)} d\vec{S} \cdot \left(\vec{\nabla} \times \vec{B}\right) = \frac{\int_{S(\Gamma)} \vec{J} \cdot d\vec{S}}{\epsilon_0} \\ \rightsquigarrow \quad & c^2 \oint_\Gamma \vec{B} \cdot d\vec{l} = \frac{\int_{S(\Gamma)} \vec{J} \cdot d\vec{S}}{\epsilon_0}. \end{aligned} \tag{4.13}$$

The rhs is proportional to the current intensity passing through the loop $\Gamma$:

$$I_{\text{through }\Gamma} \equiv \int_{S(\Gamma)} \vec{J} \cdot d\vec{S}, \tag{4.14}$$

which is the total charge passing through the closed loop $\Gamma$ per unit time. We then arrive at:

$$\oint_{\Gamma} \vec{B} \cdot d\vec{l} = \frac{I_{\text{through }\Gamma}}{\epsilon_0 c^2}, \tag{4.15}$$

which is known as Ampère's law.

## 4.3 Applications of Ampère's law

For problems with a manifest symmetry, Ampère's law can be sufficient to determine the magnetic field. Let us review two well-known applications.

### 4.3.1 Long straight wire

Consider a long straight wire with a current $I$ flowing through it (Fig. 4.4). Close to the

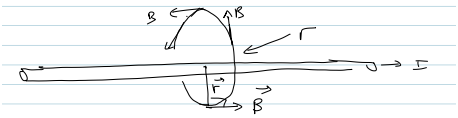


**Figure 4.4.** A long straight wire carrying a current $I$, and a circular loop $\Gamma$ of radius $r$ around it.

wire, at a distance $r$ which is much smaller than the length of the wire, the magnitude of the magnetic field is, by symmetry, the same at all points at the same distance $r$. Applying Ampère's law to a loop which is a circle of radius $r$ around the wire, we obtain

$$\begin{aligned} \oint \vec{B} \cdot d\vec{l} &= \frac{I}{\epsilon_0 c^2} \\ \rightsquigarrow B 2\pi r &= \frac{I}{\epsilon_0 c^2} \\ \rightsquigarrow B &= \frac{1}{4\pi\epsilon_0 c^2} \frac{2I}{r}. \end{aligned} \tag{4.16}$$

### 4.3.2 Solenoid

Consider a long solenoid with a current $I$ through it (Fig. 4.5). To a good approximation,

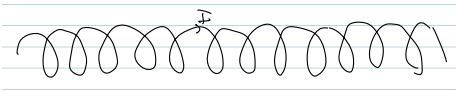


**Figure 4.5.** A long solenoid carrying a current $I$.

very long solenoids have a homogeneous magnetic field inside them and a zero magnetic field outside them. These two assumptions allow us to estimate the magnetic field inside the solenoid. Let us apply Ampère's law to the loop $\Gamma$ of Fig. 4.6. The total current through

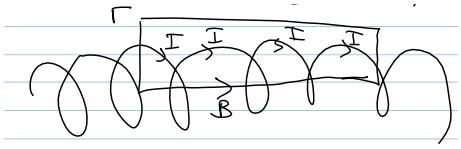



**Figure 4.6.** A rectangular loop Γ, with one side of length $L$ inside the solenoid, used to apply Ampère's law.

Γ is the number $N_{\text{in }\Gamma}$ of turns inside it times the current $I$ through one turn:

$$I_{\text{through }\Gamma} = IN_{\text{in }\Gamma}. \tag{4.17}$$

Then, from Ampère's law,

$$BL = \frac{IN_{\text{in }\Gamma}}{\epsilon_0 c^2}, \tag{4.18}$$

which gives the magnetic field

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

where $n$ is the number of turns per unit length in the solenoid.

## 4.4 Vector potential

Let us examine further the basic equations of magnetostatics (Eqs. 4.10–4.11). Eq. 4.10 is automatically satisfied if we cast the magnetic field as the curl of a vector field $\vec{A}$:

$$\vec{B} = \vec{\nabla} \times \vec{A}. \tag{4.20}$$

We introduce the abbreviation:

$$\partial_i \equiv \frac{\partial}{\partial x_i}, \tag{4.21}$$

and Einstein's summation convention:

$$A_i B_i \equiv \sum_{i=1}^{3} A_i B_i = \vec{A} \cdot \vec{B}. \tag{4.22}$$

Then

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = \partial_i (\vec{\nabla} \times \vec{A})_i = \partial_i \left(\epsilon_{ijk} \partial_j A_k\right) = \epsilon_{ijk} \partial_i \partial_j A_k = 0. \tag{4.23}$$

The tensor $\epsilon_{ijk}$ is fully antisymmetric, while the second derivative is symmetric and therefore their contraction vanishes.

From Eq. 4.11 we obtain:

$$\begin{aligned}
&(\vec{\nabla} \times \vec{B})_i = \frac{J_i}{\epsilon_0 c^2} \\
&\rightsquigarrow \epsilon_{ijk} \partial_j B_k = \frac{J_i}{\epsilon_0 c^2} \\
&\rightsquigarrow \epsilon_{ijk} \partial_j \epsilon_{klm} \partial_l A_m = \frac{J_i}{\epsilon_0 c^2} \\
&\rightsquigarrow \epsilon_{ijk} \epsilon_{klm} \partial_j \partial_l A_m = \frac{J_i}{\epsilon_0 c^2}.
\end{aligned} \tag{4.24}$$

The contraction of two Levi-Civita tensors yields

$$\epsilon_{ijk}\epsilon_{klm} = \begin{vmatrix} \delta_{il} & \delta_{im} \\ \delta_{jl} & \delta_{jm} \end{vmatrix} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}. \tag{4.25}$$

Thus, we have:

$$\begin{aligned} &\left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right)\partial_j\partial_l A_m = \frac{J_i}{\epsilon_0 c^2} \\ &\rightsquigarrow \partial_i\partial_j A_j - \partial_j^2 A_i = \frac{J_i}{\epsilon_0 c^2}. \end{aligned} \tag{4.26}$$

In full vector form:

$$\nabla^2\vec{A} - \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) = -\frac{\vec{J}}{c^2\epsilon_0}. \tag{4.27}$$

This is a second-order differential equation for the components of the vector potential. Had it not been for the second term on the lhs, this equation would be our familiar Poisson differential equation from electrostatics. In fact, we can choose to eliminate this term by exploiting a property of the vector potential, known as gauge invariance. Consider two vector potential functions $\vec{A}$ and $\vec{A}'$ which differ from each other by the gradient of a scalar function $f$:

$$\vec{A}' = \vec{A} + \vec{\nabla}f. \tag{4.28}$$

These two vector potentials are physically equivalent and give rise to the same magnetic field,

$$\vec{\nabla}\times\vec{A}' = \vec{\nabla}\times\vec{A} = \vec{B}. \tag{4.29}$$

Indeed:

$$(\vec{\nabla}\times\vec{\nabla}f)_i = \epsilon_{ijk}\partial_j\partial_k f = 0. \tag{4.30}$$

It is easy to verify that $\vec{A}'$ satisfies the same differential equation, Eq. 4.27. Indeed, under a "gauge transformation":

$$\vec{A} \to \vec{A}' = \vec{A} + \vec{\nabla}f, \tag{4.31}$$

we have

$$\begin{aligned} &\nabla^2\vec{A}' - \vec{\nabla}(\vec{\nabla}\cdot\vec{A}') \\ &= \nabla^2(\vec{A} + \vec{\nabla}f) - \vec{\nabla}[\vec{\nabla}\cdot(\vec{A} + \vec{\nabla}f)] \\ &= \nabla^2\vec{A} - \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) \\ &= -\frac{\vec{J}}{c^2\epsilon_0}. \end{aligned} \tag{4.32}$$

We therefore have the liberty to add any gradient to the vector potential without changing the physics. Exercising this freedom is called a gauge choice. Gauge symmetry differs from the symmetries we have met so far, such as rotations. It does not act on the space coordinates but on the field $\vec{A}$ itself, and it is therefore called an "internal" symmetry. It is also *local*: the function $f(\vec{x})$ can be different at every point. Two experimenters, one at Hönggerberg and one in downtown Zurich, may choose their gauge independently of each other, while a rotation acts in the same way everywhere. Nature makes use of such local symmetries beyond electrodynamics: the weak interactions, which are responsible for radioactivity, and the strong nuclear interactions are also described by gauge theories. We

can use this freedom to simplify the differential equation for the potential. For example, assume a potential $\vec{A}'$ with

$$\vec{\nabla} \cdot \vec{A}' \neq 0. \tag{4.33}$$

Now, let us find a solution of the Poisson equation:

$$\nabla^2 f = -\vec{\nabla} \cdot \vec{A}'. \tag{4.34}$$

Its solution is the familiar

$$f(\vec{r}) = \frac{1}{4\pi} \int d^3\vec{x}\, \frac{(\vec{\nabla} \cdot \vec{A}')(\vec{x})}{|\vec{r} - \vec{x}|}. \tag{4.35}$$

Then for

$$\vec{A} = \vec{A}' + \vec{\nabla} f = \vec{A}' + \frac{1}{4\pi} \vec{\nabla} \int d^3\vec{x}\, \frac{(\vec{\nabla} \cdot \vec{A}')(\vec{x})}{|\vec{r} - \vec{x}|}$$

we obtain

$$\vec{\nabla} \cdot \vec{A} = 0. \tag{4.36}$$

The "gauge" for which the above happens is called the "Coulomb" gauge. In that gauge, the vector potential satisfies a Poisson equation:

$$\nabla^2 \vec{A} = -\frac{\vec{J}}{c^2 \epsilon_0}, \tag{4.37}$$

which can be solved with the same techniques as we have developed in electrostatics. If we know all currents in all space, we can write the solution as:

$$\vec{A}(\vec{x}) = \frac{1}{4\pi\epsilon_0 c^2} \int d^3\vec{y}\, \frac{\vec{J}(\vec{y})}{|\vec{x} - \vec{y}|}. \tag{4.38}$$

## 4.5 Wires of a small thickness

A common situation is the calculation of the vector potential $\vec{A}$ and the corresponding magnetic field $\vec{B} = \vec{\nabla} \times \vec{A}$ due to electric currents circulating in thin wires (Fig. 4.7). The

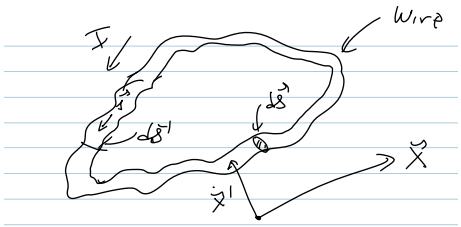



**Figure 4.7.** A current $I$ circulating in a thin wire. Only points $\vec{x}'$ inside the wire contribute to the vector potential at $\vec{x}$.

integral of Eq. 4.38,

$$\vec{A}(\vec{x}) = \frac{1}{4\pi\epsilon_0 c^2} \int d^3\vec{x}'\, \frac{\vec{J}(\vec{x}')}{|\vec{x} - \vec{x}'|}, \tag{4.39}$$

receives zero contributions from points outside the wires, where the current density is zero. We therefore have

$$\vec{A}(\vec{x}) = \frac{1}{4\pi\epsilon_0 c^2}\int_{\text{in wires}} d^3\vec{x}'\,\frac{\vec{J}(\vec{x}')}{\left|\vec{x}-\vec{x}'\right|}. \tag{4.40}$$

Let us assume that the thickness of the wires is small and that the density $\vec{J}(\vec{x}')$ of the current as well as the distance $\left|\vec{x}-\vec{x}'\right|$ are practically constant on a cross-section $\vec{S}$ of the wire. We can then integrate over the transverse directions $d\vec{S}$, obtaining:

$$\vec{A}(\vec{x}) = \frac{1}{4\pi\epsilon_0 c^2}\oint_{\text{wire}} dl(\vec{x}')\,\frac{\vec{I}(\vec{x}')}{\left|\vec{x}-\vec{x}'\right|}. \tag{4.41}$$

In the above, $l(\vec{x}')$ parametrises the curve of the wire. $\vec{I}$ is a vector with a magnitude equal to the current intensity

$$I(\vec{x}) = \int d\vec{S}\cdot\vec{J}(\vec{x}) \tag{4.42}$$

and with the direction of the current. We can now define a vector $d\vec{l}$ which is tangential to the loop. Usually, we take $d\vec{l}(\vec{x})$ to circulate anti-clockwise. The current intensity vector $\vec{I}(\vec{x})$ at a point $\vec{x}$ is either parallel or anti-parallel to $d\vec{l}(\vec{x})$:

$$d\hat{l}(\vec{x}) = \pm\hat{I}(\vec{x}). \tag{4.43}$$

Assuming an anti-clockwise steady current, we have

$$\vec{A}(\vec{r}) = \frac{I}{4\pi\epsilon_0 c^2}\oint_{\text{wire}} d\vec{l}(\vec{x})\,\frac{1}{\left|\vec{r}-\vec{x}\right|}. \tag{4.44}$$

For a clockwise current, one should remember to include a minus sign or to change the direction convention for $d\vec{l}$.

Let us now see an application of the above result. Consider a very long wire along the $x$-axis with current intensity $I$ (Fig. 4.8). Since the current flows only along the $x$-axis, the $y$ and $z$ components of the vector potential are zero:

$$A_y = A_z = 0. \tag{4.45}$$

The $x$-component of the vector potential is:

$$\begin{aligned} A_x &= \frac{I}{4\pi\epsilon_0 c^2}\int_{-\Delta}^{\Delta}\frac{dx}{\sqrt{r^2+x^2}} \\ &= \frac{I}{4\pi\epsilon_0 c^2}\log\left[\frac{\sqrt{1+\frac{r^2}{\Delta^2}}+1}{\sqrt{1+\frac{r^2}{\Delta^2}}-1}\right] \approx \frac{I}{4\pi\epsilon_0 c^2}\left[-\log r^2+\log(4\Delta^2)\right], \end{aligned} \tag{4.46}$$

where the wire extends from $-\Delta$ to $\Delta$ and, in the last step, we have taken the limit $\Delta\to\infty$. Notice that the expression has a logarithmic singularity in this limit. However, this has no physical consequence: the magnetic field is computed from the curl of the vector potential, and this infinite constant drops out upon differentiation. We have

$$B_i = \epsilon_{ijk}\partial_j A_k, \tag{4.47}$$

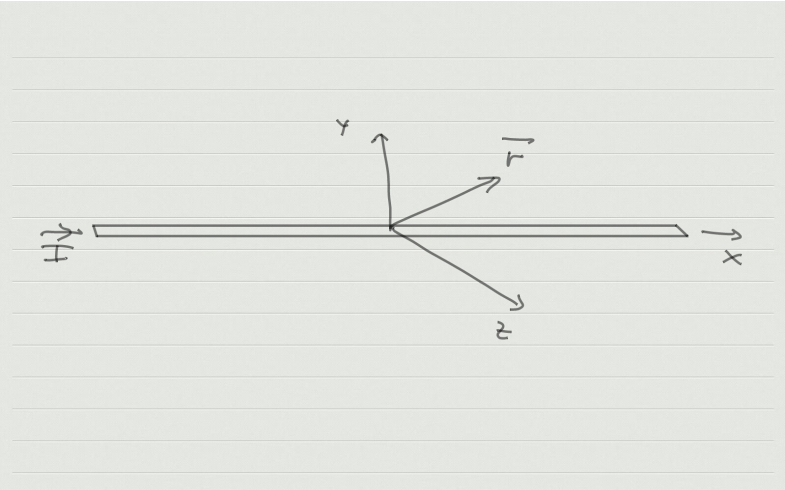



**Figure 4.8.** A long straight wire along the $x$-axis carrying a current $I$. The vector potential is computed at the point $\vec{r}$.

which gives for the components:

$$B_x = 0, \tag{4.48}$$

$$B_y = -\frac{I}{2\pi\epsilon_0 c^2}\frac{z}{r^2}, \tag{4.49}$$

$$B_z = +\frac{I}{2\pi\epsilon_0 c^2}\frac{y}{r^2}. \tag{4.50}$$

The magnitude of the magnetic field is

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} = \frac{I}{2\pi\epsilon_0 c^2}\frac{1}{r}, \tag{4.51}$$

which is in agreement with the result obtained from Ampère's law.

## 4.6 Magnetic dipole

Consider a steady current $\vec{J}$ circulating in a small region of space. We are interested in computing the vector potential and the magnetic field at a position $\vec{r}$ which is far from the current.

The vector potential is given by Eq. 4.38:

$$\vec{A}(\vec{r}) = \frac{1}{4\pi\epsilon_0 c^2}\int d^3\vec{x}\,\frac{\vec{J}(\vec{x})}{|\vec{r}-\vec{x}|}. \tag{4.52}$$

We could expand the inverse distance in the ratio $\frac{x}{r}$ by means of spherical harmonics. Since, here, we are interested only in the leading terms of the expansion, we use a simpler method, a Taylor series:

$$\frac{1}{|\vec{r}-\vec{x}|} = \frac{1}{\left[r^2 + x^2 - 2\vec{x}\cdot\vec{r}\right]^{\frac{1}{2}}} = \frac{1}{r} + \frac{\vec{x}\cdot\vec{r}}{r^3} + O\left(\frac{x^2}{r^3}\right). \tag{4.53}$$

The first term of the rhs in Eq. 4.53 gives a zero contribution to the vector potential of Eq. 4.52. Indeed, total derivatives of a localised current density integrated over all space vanish (by the divergence theorem), since the current density vanishes at infinity. We therefore have

$$\begin{aligned} 0 &= \int d^3\vec{x}\,\partial_i\left(x_k J_i\right) = \int d^3\vec{x}\,\left[\delta_{ik}J_i + x_k\partial_i J_i\right] = \int d^3\vec{x}\,J_k \\ &\rightsquigarrow \int d^3\vec{x}\,\vec{J} = 0. \end{aligned} \tag{4.54}$$

In the above, we have used that

$$\vec{\nabla}\cdot\vec{J} = \partial_i J_i = -\frac{\partial\rho}{\partial t} = 0.$$

Physically, a steady current which stays inside a bounded region cannot flow away forever; it must return. For every contribution to $\int d^3\vec{x}\,\vec{J}$ in one direction there is an opposite contribution from the returning current. There is thus no magnetic analogue of the total charge, and the $1/r$ term of the potential is absent. The vector potential of a localised steady current starts at order $1/r^2$, and the magnetic field at order $1/r^3$, as we will see below. The equations for $\vec{A}$ and $\phi$ have the same form and we can borrow the solutions from electrostatics. However, there is a qualitative difference: far from a bounded circuit the magnetic field falls off faster than the Coulomb field $\sim 1/r^2$ of a net charge. Similarly, we can prove that:

$$\begin{aligned} 0 &= \int d^3\vec{x}\,\partial_i\left(x_k J_i x_l\right) \\ &\rightsquigarrow \int d^3\vec{x}\,\left[J_k x_l + J_l x_k\right] = 0. \end{aligned} \tag{4.55}$$

Now we are ready to calculate the contribution of the second term in the inverse distance expansion to the vector potential:

$$\begin{aligned} A_i &\approx \frac{1}{4\pi\epsilon_0 c^2\, r^3}\int d^3\vec{x}\,J_i x_j r_j \\ &= \frac{r_j}{4\pi\epsilon_0 c^2\, r^3}\int d^3\vec{x}\,\left[\frac{J_i x_j + J_j x_i}{2} + \frac{J_i x_j - J_j x_i}{2}\right] \\ &= \frac{1}{8\pi\epsilon_0 c^2\, r^3}\int d^3\vec{x}\,\left[J_i(x_j r_j) - (J_j r_j)x_i\right]. \end{aligned} \tag{4.56}$$

In vector notation, we have

$$\vec{A} \approx \frac{1}{8\pi\epsilon_0 c^2\, r^3}\int d^3\vec{x}\,\left[\vec{J}(\vec{x}\cdot\vec{r}) - (\vec{J}\cdot\vec{r})\vec{x}\right]. \tag{4.57}$$

The bracket in the integral is a double cross product:

$$\vec{r}\times\left(\vec{x}\times\vec{J}\right) = \vec{x}\left(\vec{r}\cdot\vec{J}\right) - \vec{J}\left(\vec{r}\cdot\vec{x}\right). \tag{4.58}$$

Therefore,

$$\vec{A} = \frac{1}{4\pi\epsilon_0 c^2}\frac{\vec{m}\times\vec{r}}{r^3}, \tag{4.59}$$

where the vector $\vec{m}$ depends on the geometric characteristics of the distribution and is known as the magnetic moment:

$$\vec{m} = \frac{1}{2}\int d^3\vec{x}\,(\vec{x}\times\vec{J}). \tag{4.60}$$

Having computed the vector potential of a given "magnetic dipole" with magnetic moment $\vec{m}$, we can compute the magnetic field as:

$$\vec{B} = \vec{\nabla}\times\vec{A} = \frac{1}{4\pi\epsilon_0 c^2}\vec{\nabla}\times\frac{\vec{m}\times\vec{r}}{r^3} = -\frac{1}{4\pi\epsilon_0 c^2}\vec{\nabla}\times\left[\vec{m}\times\vec{\nabla}\frac{1}{r}\right]. \tag{4.61}$$

Performing the differentiations and using the identity

$$\nabla^2\frac{1}{r} = -4\pi\delta(\vec{r}) = 0 \quad \forall\, r > 0, \tag{4.62}$$

we find:

$$\vec{B} = \frac{1}{4\pi\epsilon_0 c^2}\left(\vec{m}\cdot\vec{\nabla}\right)\vec{\nabla}\frac{1}{r} = -\frac{1}{4\pi\epsilon_0 c^2}\frac{\vec{m} - 3\hat{r}\left(\hat{r}\cdot\vec{m}\right)}{r^3}. \tag{4.63}$$

As an example of a magnetic dipole, consider the special case of a steady current circulating anti-clockwise in a wire which lies in a plane. The magnetic moment becomes

$$\vec{m} = \oint \frac{I}{2}\vec{x}\times d\vec{l}, \tag{4.64}$$

where $I$ is the intensity of the current. Notice that the integrand is a surface element (Fig. 4.9),

$$d\vec{S} = \frac{1}{2}\vec{x}\times d\vec{l}. \tag{4.65}$$

Therefore, the magnetic moment is simply:

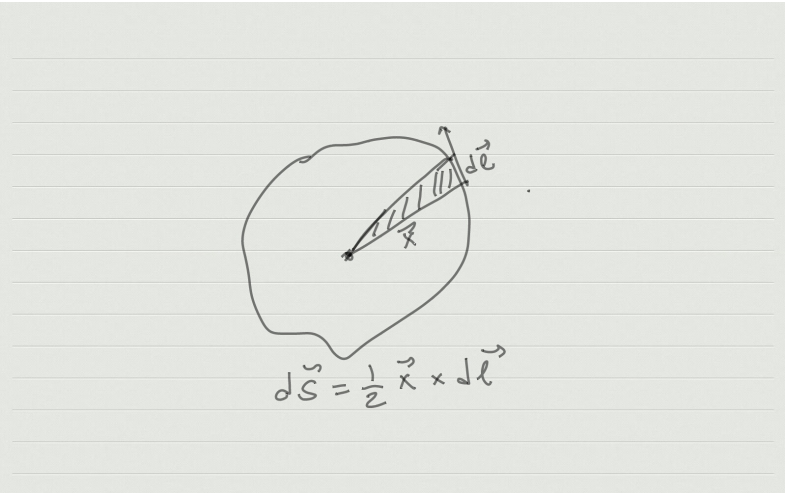



**Figure 4.9.** The surface element $d\vec{S} = \frac{1}{2}\vec{x}\times d\vec{l}$ swept by the position vector $\vec{x}$ along the loop.

$$\vec{m} = I\vec{S}, \tag{4.66}$$

where

$$\vec{S} = \int d\vec{S}$$

is the (vector) area of the loop.

As a second example, consider the current created by a single charge $q$ and mass $M$ moving along a closed loop. The current density is:

$$\vec{J} = \rho\vec{v} = q\delta(\vec{x} - \vec{r}(t))\vec{v}, \tag{4.67}$$

where

$$\vec{v} = \frac{d\vec{r}}{dt}$$

is the velocity of the charge. The magnetic moment is

$$\vec{m} = \frac{1}{2}\int d^3\vec{x}\,\vec{x}\times\vec{J} = \frac{q}{2}\vec{r}\times\vec{v} = \frac{q}{2M}\vec{L}, \tag{4.68}$$

where

$$\vec{L} = \vec{r}\times(M\vec{v}) \tag{4.69}$$

is the angular momentum of the charge. *Remark.* For a single moving charge, $\vec{\nabla}\cdot\vec{J} = -\partial_t\rho \neq 0$, so the magnetostatic derivation does not apply instantaneously; $\vec{m} = \frac{q}{2M}\vec{L}$ holds for the time-averaged current of a periodic orbit, where $\langle\vec{r}\times\vec{v}\rangle$ enters (see Ref. [2], Sec. 5.6).

### 4.6.1 Force on a "magnetic dipole" inside a magnetic field

Let us place a "magnetic dipole" inside a magnetic field $\vec{B}$. The force acting on it is

$$\vec{F} = \int d^3\vec{x}\,\rho(\vec{x})\vec{v}\times\vec{B}(\vec{x}) = \int d^3\vec{x}\,\vec{J}(\vec{x})\times\vec{B}(\vec{x}). \tag{4.70}$$

If the current is contained in a small region of space, we can assume that the magnetic field varies slowly within this region and approximate it with the first few terms of a Taylor expansion:

$$\vec{B}(\vec{x}) \approx \vec{B}(0) + \left(\vec{x}\cdot\vec{\nabla}\right)\vec{B}(0) + \dots . \tag{4.71}$$

The first term vanishes upon integration, since $\int d^3\vec{x}\,\vec{J} = 0$. For the second term, we find

$$\begin{aligned} F_i &\approx \int d^3\vec{x}\,\epsilon_{ilm}J_l(\vec{x})\left[B_m(\vec{0}) + x_\rho\partial_\rho B_m(0) + \dots\right] \\ &= \epsilon_{ilm}\partial_\rho B_m(0)\int d^3\vec{x}\,x_\rho J_l(\vec{x}) = \frac{1}{2}\epsilon_{ilm}\partial_\rho B_m(0)\int d^3\vec{x}\,\left[x_\rho\,J_l(\vec{x}) - x_l\,J_\rho(\vec{x})\right] \\ &= \frac{1}{2}\epsilon_{ilm}\partial_\rho B_m(0)\epsilon_{n\rho l}\int d^3\vec{x}\,\left[\vec{x}\times\vec{J}(\vec{x})\right]_n = \epsilon_{ilm}\epsilon_{n\rho l}\partial_\rho B_m(0)m_n \\ &= m_n\partial_i B_n(0) - m_i\,\vec{\nabla}\cdot\vec{B}(0) = m_n\partial_i B_n(0), \end{aligned} \tag{4.72}$$

which yields

$$\vec{F} \approx \vec{\nabla}(\vec{m}\cdot\vec{B}). \tag{4.73}$$

The force can be written as

$$\vec{F} = -\vec{\nabla}U, \tag{4.74}$$

where

$$U = -\vec{m} \cdot \vec{B} \tag{4.75}$$

is the potential energy of the magnetic dipole.

**Exercise:** Prove that the torque is

$$\vec{\tau} = \int \vec{x} \times d\vec{F} = \vec{m} \times \vec{B}(0). \tag{4.76}$$

## 4.7 Relativity of the electric and magnetic fields

Consider a wire on the $x$-axis producing a magnetic field

$$B = \frac{1}{2\pi\epsilon_0 c^2} \frac{I}{r} \tag{4.77}$$

at a distance $r$ due to a current with an intensity $I$. Let us assume that the conduction electrons inside the wire have a velocity $\vec{v}$. We now imagine another electron outside the wire, moving parallel to the wire at a distance $r$ from it with the same velocity $\vec{v}$ (Fig. 4.10). The force acting on the electron outside the wire is

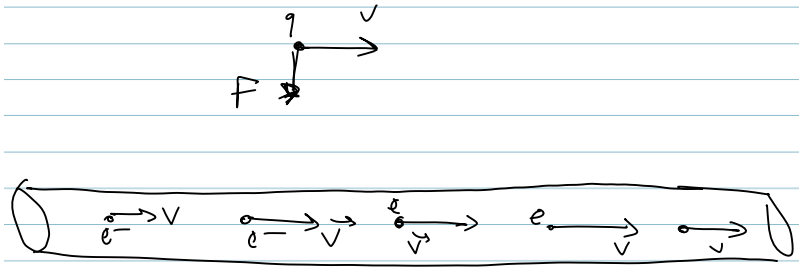


**Figure 4.10.** A charge outside a current-carrying wire, moving parallel to it with the same velocity $\vec{v}$ as the conduction electrons.

$$F = qvB = \frac{1}{2\pi\epsilon_0} \frac{qvI}{c^2 r}. \tag{4.78}$$

The intensity $I$ of the current is:

$$I = S\rho_- v, \tag{4.79}$$

where $S$ is the cross-section of the wire and $\rho_-$ is the electron charge density inside the wire. The force is then:

$$F = \frac{qS}{2\pi\epsilon_0} \frac{\rho_-}{r} \frac{v^2}{c^2}. \tag{4.80}$$

We now change reference frame and choose one which moves with a relative velocity $\vec{v}$ along the wire. In the new frame the electron outside is static and has zero velocity:

$$v' = 0. \tag{4.81}$$

In this frame, the magnetic component of the force must vanish! This is a paradox: it appears that the charge is accelerated in one frame and free in another. The paradox is resolved by Einstein's special relativity: an equivalent electric force develops in the new frame.

In the original frame, the positive charge density inside the wire cancels the negative charge density of the free electrons:

$$\rho = \rho_+ + \rho_- = 0. \tag{4.82}$$

The total charge density in the moving frame is

$$\rho' = \rho'_+ + \rho'_-. \tag{4.83}$$

It turns out that it is not zero!

Consider $N$ charges inside a volume $V = SL$ which are at rest in a certain reference frame (Fig. 4.11). The total charge is

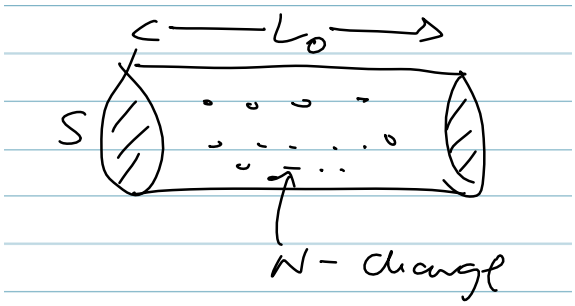



**Figure 4.11.** $N$ charges at rest in a piece of material with cross-section $S$.

$$qN = \rho_{\text{rest}} V = \rho_{\text{rest}} LS. \tag{4.84}$$

In a reference frame where the charges are moving with velocity $\vec{v}$ (Fig. 4.12), the same number of charges is included in the same piece of material, which, however, appears contracted due to the laws of special relativity:

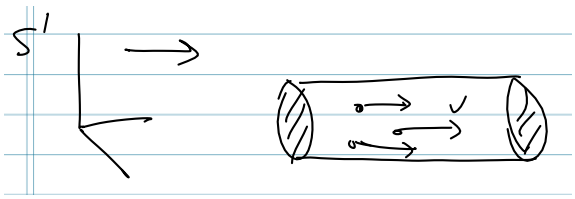



**Figure 4.12.** The same charges seen from a reference frame $S'$ in which they move with velocity $\vec{v}$.

$$Nq = \rho_{\text{moving}} SL_{\text{moving}} = \rho_{\text{moving}} SL\sqrt{1 - \frac{v^2}{c^2}}. \tag{4.85}$$

We therefore obtain a relation between the charge densities in a frame where the charges are at rest and in a frame where they move:

$$\rho_{\text{moving}} = \frac{\rho_{\text{rest}}}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{4.86}$$

The density of a set of charges is therefore smallest in their rest frame. In every other frame it is larger, compensating for the contraction of the length that contains them. In

the original frame the electrons move, so their density is not their rest-frame density. In the frame moving with them their density decreases, while the positive ions, which are now moving, appear denser. This is why the balance $\rho_+ + \rho_- = 0$ cannot hold in both frames. The force on the outside charge has not disappeared by changing frame; it has only changed name, from magnetic to electric.

For the electron and positive charge densities inside the wire we obtain:

$$\rho'_+ = \frac{\rho_+}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{4.87}$$

and

$$\rho_- = \frac{\rho'_-}{\sqrt{1-\frac{v^2}{c^2}}} \rightsquigarrow \rho'_- = \rho_- \sqrt{1-\frac{v^2}{c^2}}. \tag{4.88}$$

Given that $\rho_+ = -\rho_-$ we have:

$$\rho' = -\rho_- \frac{\frac{v^2}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{4.89}$$

This charge distribution spreads over an infinitely long wire and creates an electric field:

$$E' = \frac{1}{2\pi\epsilon_0} S \frac{\rho'}{r}. \tag{4.90}$$

The force is:

$$F' = qE' = -\frac{qS}{2\pi\epsilon_0}\frac{\rho_-}{r}\frac{\frac{v^2}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}} = -\frac{F}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad |F'| = \frac{|F|}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{4.91}$$

Notice that these two forces cause the same physical effect in their respective reference frames. Both are perpendicular to the wire and point towards it. The momentum change of the charge $q$ in the direction transverse to the wire is the same:

$$\Delta p_T = F\Delta t = F'\sqrt{1-\frac{v^2}{c^2}}\frac{\Delta t'}{\sqrt{1-\frac{v^2}{c^2}}} = F'\Delta t' = \Delta p'_T. \tag{4.92}$$

From this discussion, we conclude that the magnetic and electric fields transform into each other when we change reference frame. In view of this observation, it is desirable to find a unified description of the two fields as components of a single electromagnetic field.

# 5

# Time-varying electromagnetic fields

It is now time to consider the full Maxwell's equations.

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{5.1}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{5.2}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{5.3}$$

$$c^2 \vec{\nabla} \times \vec{B} = \frac{\vec{j}}{\epsilon_0} + \frac{\partial \vec{E}}{\partial t} \tag{5.4}$$

## 5.1 Charge conservation

These equations imply charge conservation. Taking the divergence of Eq. 5.4 we obtain:

$$c^2 \epsilon_0 \vec{\nabla} \cdot \left( \vec{\nabla} \times \vec{B} \right) = \vec{\nabla} \cdot \vec{J} + \frac{\partial}{\partial t} \left( \epsilon_0 \vec{\nabla} \cdot \vec{E} \right) \rightsquigarrow 0 = \vec{\nabla} \cdot \vec{J} + \frac{\partial \rho}{\partial t}, \tag{5.5}$$

which is the continuity equation expressing charge conservation. Integrating over all space,

$$0 = \int d^3 \vec{x} \left[ \vec{\nabla} \cdot \vec{J} + \frac{\partial \rho}{\partial t} \right] = \frac{\partial Q_{\text{universe}}}{\partial t}, \tag{5.6}$$

we obtain that the total charge in the universe is a constant. The divergence term becomes, by the divergence theorem, a surface integral at the edge of the universe, where we assume that there are no currents. This is a majestic statement: charge conservation does not need to be imposed separately, since it is built into Maxwell's equations. Observations are consistent with the constant value of the total charge being zero.

## 5.2 Vector and scalar potentials

It is now time to find the properties of the scalar and vector potentials in the general case of time-dependent electromagnetic fields. The equations

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad \nabla^2 \vec{A} = -\frac{\vec{J}}{c^2 \epsilon_0} \tag{5.7}$$

are not general and they are correct only for static fields.

The equation $\vec{\nabla} \cdot \vec{B} = 0$ is generally valid and it can be satisfied by introducing a vector potential as before:

$$\vec{B} = \vec{\nabla} \times \vec{A}. \tag{5.8}$$

The Maxwell equation 5.3 becomes

$$\vec{\nabla} \times \vec{E} = -\frac{\partial}{\partial t}\vec{B} \rightsquigarrow \vec{\nabla} \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t}\right) = 0. \tag{5.9}$$

This is satisfied if we introduce a scalar potential $\phi$ through

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi. \tag{5.10}$$

Indeed,

$$\vec{\nabla} \times \vec{\nabla}\phi = \epsilon_{ijk}\partial_j\partial_k\phi = 0. \tag{5.11}$$

Substituting Eqs. 5.10 and 5.8 into the Maxwell equation 5.4, we obtain

$$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right]\vec{A} + \vec{\nabla}\left[\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t}\right] = \frac{\vec{J}}{c^2\epsilon_0}. \tag{5.12}$$

In this derivation, we have used the identity

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{A}\right) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2\vec{A}. \tag{5.13}$$

With the same substitutions, the Maxwell equation 5.1 becomes

$$-\nabla^2\phi - \frac{\partial}{\partial t}\vec{\nabla} \cdot \vec{A} = \frac{\rho}{\epsilon_0}. \tag{5.14}$$

Equivalently,

$$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right]\phi - \frac{\partial}{\partial t}\left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t}\right) = \frac{\rho}{\epsilon_0}. \tag{5.15}$$

The differential operator in square brackets in Eqs. 5.12 and 5.15 is known as the d'Alembert or "box" operator:

$$\Box \equiv \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2. \tag{5.16}$$

## 5.3 Gauge invariance

The electric and magnetic fields remain invariant under a simultaneous transformation of the scalar and vector potentials, known as a gauge transformation:

$$\phi \rightarrow \phi' = \phi - \frac{\partial}{\partial t}f, \tag{5.17}$$

$$\vec{A} \rightarrow \vec{A}' = \vec{A} + \vec{\nabla}f, \tag{5.18}$$

where

$$f = f(\vec{x}, t)$$

is an arbitrary function of space-time. Indeed,

$$\begin{aligned}\vec{E}' &= -\partial_t \vec{A}' - \vec{\nabla}\phi' = -\partial_t \vec{A} - \partial_t \vec{\nabla} f - \vec{\nabla}\phi + \vec{\nabla}\partial_t f \\ &= -\partial_t \vec{A} - \vec{\nabla}\phi = \vec{E}. \end{aligned} \tag{5.19}$$

For the magnetic field,

$$\vec{B}' = \vec{\nabla} \times \vec{A}' = \vec{\nabla} \times \vec{A} + \vec{\nabla} \times \vec{\nabla} f = \vec{\nabla} \times \vec{A} = \vec{B}. \tag{5.20}$$

How could one guess this form of the transformation? The scalar potential is a function, so we can only add to it a scalar quantity made out of $f$, such as $\partial_t f$. The vector potential is a vector, so we add to it the gradient $\vec{\nabla} f$. The overall sign is a convention (we can replace $f \to -f$), but the relative sign between the two transformations is essential and at this stage looks mysterious. Special relativity will explain it. With the four-vectors $A^\mu = (\phi/c, \vec{A})$ and $\partial^\mu = (\frac{1}{c}\partial_t, -\vec{\nabla})$ of the following chapter, the gauge transformation takes the simple form $A^\mu \to A^\mu - \partial^\mu f$. The only requirement on $f$ is that the derivatives we take exist. For a time-independent $f$ we recover the gauge transformation of magnetostatics. Equations 5.12–5.15 also remain invariant under gauge transformations. This is easily verified by observing that the combination

$$\begin{aligned}&\vec{\nabla} \cdot \vec{A}' + \frac{1}{c^2}\partial_t \phi' = \vec{\nabla} \cdot \left(\vec{A} + \vec{\nabla} f\right) + \frac{1}{c^2}\partial_t \left(\phi - \partial_t f\right) \\ = \;& \vec{\nabla} \cdot \vec{A} + \frac{1}{c^2}\partial_t \phi - \Box f \end{aligned} \tag{5.21}$$

changes only by a term $-\Box f$. This change is cancelled by the change of the terms with the box operator,

$$\Box \vec{A} \to \Box \vec{A} + \vec{\nabla} \Box f, \quad \Box \phi \to \Box \phi - \partial_t \Box f.$$

Given our freedom to modify the vector and scalar potentials by choosing the function $f(\vec{x}, t)$, we have an opportunity to make a choice which simplifies their differential equations. For example, we can choose $f$ so that the potentials $\vec{A}_L, \phi_L$ in this gauge satisfy

$$\vec{\nabla} \cdot \vec{A}_L + \frac{1}{c^2}\partial_t \phi_L = 0. \tag{5.22}$$

This is called the "Lorentz gauge". The subscript $_L$ is to remind us that this equation is valid only in this specific gauge. *Remark.* The condition is due to L. V. Lorenz (1867), not H. A. Lorentz, and is nowadays often called the Lorenz gauge (see the footnote in Sec. 6.3 of Ref. [2]). It can always be reached: starting from arbitrary potentials, choose $f$ with $\Box f = \vec{\nabla} \cdot \vec{A} + \frac{1}{c^2}\partial_t \phi$. It does not fix $f$ completely: any further $f$ with $\Box f = 0$ (restricted gauge transformation) preserves the condition. In the Lorentz gauge, Maxwell's equations give the following differential equations for the potentials:

$$\Box \phi_L = \frac{\rho}{\epsilon_0} \tag{5.23}$$

and

$$\Box \vec{A}_L = \frac{\vec{J}}{c^2 \epsilon_0}. \tag{5.24}$$

**Exercise:** Show that the d'Alembert operator is not invariant under the Galilean transformations of Newtonian mechanics, but it is invariant under the Lorentz transformations

of special relativity. The minus sign in the d'Alembert operator is important. If it were a plus sign, $\Box$ would be a Laplacian in four dimensions, with $ct$ as a fourth coordinate. We could then solve the equations for the potentials by repeating our electrostatics in four dimensions. The minus sign cannot be changed, because it is dictated by the consistency with special relativity, which the exercise above demonstrates. Potentials which solve the equations in one inertial frame are again solutions in another frame only if the two frames are related by a Lorentz transformation and not by a Galilean transformation. This is the strongest hint we have met so far that special relativity is the correct way to change between inertial frames. The techniques to solve the equations (Green's functions) will nevertheless be very similar to those of electrostatics.

## 5.4 Electromagnetic waves in empty space

Maxwell's equations admit non-zero solutions for the electric and magnetic field in places where there exist no currents and charges. This is of course not a surprise since already in magnetostatics and electrostatics we have found that the scalar and vector potentials extend over infinite distances from their sources. However, now we are set to find more exciting solutions of Maxwell's equations, which correspond to ripples of electromagnetic energy propagating through empty space with a constant speed, the speed of light.

In empty space and in the Lorentz gauge, Maxwell's equations take the form

$$\Box \vec{A} = \Box \phi = 0. \tag{5.25}$$

The electric and magnetic fields are given by

$$\vec{E} = -\vec{\nabla}\phi - \partial_t \vec{A}, \quad \vec{B} = \vec{\nabla} \times \vec{A}. \tag{5.26}$$

Acting on the above equations with the d'Alembert operator, we obtain

$$\Box \vec{E} = \Box \vec{B} = 0. \tag{5.27}$$

In empty space, all fields satisfy the same equation

$$\Box f = 0, \quad f \in \{\phi, \vec{A}, \vec{E}, \vec{B}\}. \tag{5.28}$$

A solution of this equation is

$$f(\vec{x}, t) = f(\hat{\eta} \cdot \vec{x} - ct), \tag{5.29}$$

with

$$\hat{\eta}^2 = 1. \tag{5.30}$$

In this solution, $f$ depends on a single combination

$$u = \hat{\eta} \cdot \vec{x} - ct. \tag{5.31}$$

We can easily verify that $f(u)$ is indeed a solution. We have

$$\partial_t f(u) = (\partial_t u)\partial_u f(u) = -c\partial_u f. \tag{5.32}$$

For the second time derivative, we find

$$\partial_t^2 f = c^2 \partial_u^2 f. \tag{5.33}$$

Similarly,

$$\vec{\nabla} f(u) = (\vec{\nabla} u)\partial_u f = \hat{\eta}\partial_u f, \tag{5.34}$$

and

$$\nabla^2 f(u) = \hat{\eta}^2 \partial_u^2 f = \partial_u^2 f. \tag{5.35}$$

Therefore

$$\left[\frac{1}{c^2}\partial_t^2 - \nabla^2\right] f = \partial_u^2 f - \partial_u^2 f = 0.$$

The solution $f(\hat{\eta}\cdot\vec{x} - ct)$ is a travelling wave with a speed equal to the speed of light $c$, moving along the direction of the unit vector $\hat{\eta}$ (Fig. 5.1). The greatest success of Maxwell's

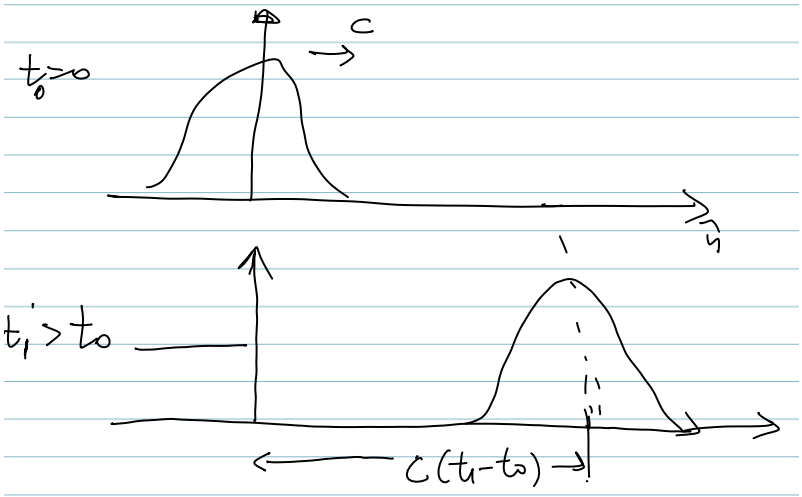



**Figure 5.1.** A travelling wave $f(\hat{\eta}\cdot\vec{x} - ct)$: its profile moves rigidly with speed $c$ along $\hat{\eta}$.

theory of electrodynamics has been to realise that electromagnetic energy (light) can propagate as electromagnetic waves.

For the electric and magnetic fields, we write

$$\vec{E} = \hat{e}E(\hat{\eta}\cdot\vec{x} - ct), \tag{5.36}$$

$$\vec{B} = \hat{b}B(\hat{\eta}\cdot\vec{x} - ct). \tag{5.37}$$

Then

$$0 = \vec{\nabla}\cdot\vec{E} = \partial_i \hat{e}_i E(\hat{\eta}_k x_k - ct) = \hat{e}_i\hat{\eta}_i\partial_u E, \tag{5.38}$$

which leads to

$$\hat{e}\cdot\hat{\eta} = 0. \tag{5.39}$$

Similarly, from

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

we conclude that also

$$\hat{b}\cdot\hat{\eta} = 0. \tag{5.40}$$

Therefore, the electric and magnetic fields in an electromagnetic wave propagating in empty space are vectors which are transverse to the direction $\hat{\eta}$ of the propagation of the wave:

$$\vec{E}, \vec{B} \perp \hat{\eta}.$$

We now turn to the remaining Maxwell's equations. From Faraday's law,

$$\vec{\nabla} \times \vec{E} = -\partial_t \vec{B} \rightsquigarrow \left(\hat{\eta} \times \hat{e}\right) \partial_u E = c \partial_u B \hat{b}. \tag{5.41}$$

The magnetic field is therefore perpendicular to both the electric field and the direction of the wave propagation (Fig. 5.2). For the magnitudes of the electric and magnetic fields we

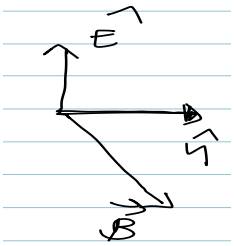



**Figure 5.2.** The electric field, the magnetic field and the direction of propagation $\hat{\eta}$ of a plane electromagnetic wave are mutually perpendicular.

have the equation

$$\partial_u E = c \partial_u B. \tag{5.42}$$

For dynamical fields (time-varying with no constant component), the electric and magnetic fields are proportional to each other:

$$E = cB. \tag{5.43}$$

*Remark.* The ansatz $\vec{E} = \hat{e}\, E(u)$ with a fixed $\hat{e}$ describes a linearly polarised wave. The general plane wave is $\vec{E} = \vec{\mathcal{E}}(u)$ with an arbitrary transverse vector function, $\hat{\eta} \cdot \vec{\mathcal{E}} = 0$, and $\vec{B} = \frac{1}{c}\, \hat{\eta} \times \vec{E}$ (see Secs. 7.1–7.2 of Ref. [2]); circular and elliptic polarisation are superpositions of two linear ones.

The last of Maxwell's equations

$$c^2 \vec{\nabla} \times \vec{B} = \partial_t \vec{E} \tag{5.44}$$

yields a relation equivalent to Eq. 5.41.

### 5.4.1 Spherical waves

In the last section we found plane-wave solutions of Maxwell's equations in empty space. These are not the only wave solutions that exist. One can find a diversity of solutions, depending on the sources and boundary conditions that generate them. For example, if we are able to change the charge and current density at only one point in the entire space, we will generate a wave with spherical symmetry, i.e. with no preferred direction (Fig. 5.3). Maxwell's equations give

$$\Box f = 0, \quad \forall f \in \{\phi, \vec{A}, \vec{B}, \vec{E}\}. \tag{5.45}$$

A wave solution with spherical symmetry is of the form

$$f(\vec{r}, t) = f(r, t) \tag{5.46}$$

and is independent of the azimuthal and polar angles $\phi, \theta$. In spherical coordinates, we have

$$\Box f(r,t) = \frac{1}{c^2} \partial_t^2 f - \frac{1}{r} \partial_r^2 (r f) = 0, \tag{5.47}$$

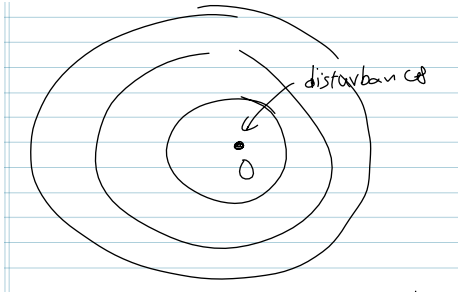



**Figure 5.3.** A spherical wave emitted from a point source.

which gives

$$\left[\frac{1}{c^2}\partial_t^2 - \partial_r^2\right](rf) = 0. \tag{5.48}$$

A solution of this equation is

$$rf = A(r - ct) + B(r + ct), \tag{5.49}$$

and, equivalently,

$$f(r,t) = \frac{A(r-ct)}{r} + \frac{B(r+ct)}{r}. \tag{5.50}$$

The first term corresponds to a spherical wave which propagates outwards. The second term is a spherical wave that propagates inwards. It is very hard to realise boundary conditions where a disturbance arises simultaneously on a distant sphere at a time $t_0$ and propagates towards the centre. Typically, outward propagation is the realistic case, where a disturbance takes place at $t_0$ at one point only. *Remark.* This holds for a scalar wave (or for the Cartesian components of $\vec{A}$ in the Lorentz gauge); for a vector field, the Laplacian in spherical coordinates does not reduce to the radial Laplacian acting on $|\vec{E}|$. There is no spherically symmetric electromagnetic wave: a radial field $\vec{E} = E(r,t)\hat{r}$ has $\vec{\nabla}\times\vec{E} = 0$, so $\vec{B}$ is static, and by Gauss' law $E(r,t) = Q/(4\pi\epsilon_0 r^2)$ with $Q$ conserved, i.e. the field is pure Coulomb (no monopole radiation, cf. Chapter 9 of Ref. [2]). Also, $A(r-ct)/r$ solves $\Box f = 0$ only for $r \neq 0$; at the origin $\Box\,[A(r-ct)/r]$ produces a source proportional to $\delta^3(\vec{r})$. This is how the retarded Green's function of Eq. 6.30 arises.

# 6

# General solutions of Maxwell's equations with sources

We will now develop tools for finding solutions of Maxwell's equations in the presence of sources ($\vec{J}, \rho \neq 0$). We have seen that in the Lorentz gauge Maxwell's equations for the scalar and vector potentials become

$$\Box\phi = \frac{\rho}{\epsilon_0}, \quad \Box\vec{A} = \frac{\vec{J}}{c^2\epsilon_0}. \tag{6.1}$$

## 6.1 Green's functions

To solve the above differential equations, we will use the technique of Green's functions. We seek functions

$$G\left(\vec{r}, t; \vec{r}', t'\right) \tag{6.2}$$

which satisfy

$$\Box_{\vec{r},t} G(\vec{r}, t; \vec{r}', t') = \delta(\vec{r} - \vec{r}')\delta(t - t'). \tag{6.3}$$

Then a solution of the Maxwell equation for the scalar potential is

$$\phi(\vec{r}, t) = \int_{-\infty}^{\infty} d^3\vec{r}' dt' G(\vec{r}, t; \vec{r}', t') \frac{\rho(\vec{r}', t')}{\epsilon_0}. \tag{6.4}$$

This is easy to verify by acting with the d'Alembert operator on both sides of Eq. 6.4. To the above expression, we can add free-space solutions $\phi_{\text{free}}(\vec{r}, t)$, such as the electromagnetic waves found in the last section, which satisfy

$$\Box\phi_{\text{free}}(\vec{r}, t) = 0. \tag{6.5}$$

Our general solution is then

$$\phi(\vec{r}, t) = \phi_{\text{free}}(\vec{r}, t) + \int_{-\infty}^{\infty} d^3\vec{r}' dt' G(\vec{r}, t; \vec{r}', t') \frac{\rho(\vec{r}', t')}{\epsilon_0}. \tag{6.6}$$

The solution is fixed by boundary conditions.

Similarly, the general solution for the vector potential is

$$\vec{A}(\vec{r}, t) = \vec{A}_{\text{free}}(\vec{r}, t) + \int_{-\infty}^{\infty} d^3\vec{r}' dt' G(\vec{r}, t; \vec{r}', t') \frac{\vec{J}(\vec{r}', t')}{c^2\epsilon_0}. \tag{6.7}$$

## 6.2 Fourier transformations

In this section, we recall the salient properties of Fourier transformations. As an exercise, prove that

$$\int_{-\infty}^{\infty} dw e^{-iwt} = 2\pi\delta(t). \tag{6.8}$$

(Hint: use that

$$\partial_t \Theta(t) = \delta(t)$$

and that the integral

$$-\frac{1}{2\pi i}\lim_{\epsilon\to 0^+}\int_{-\infty}^{\infty} dw \frac{e^{-iwt}}{w+i\epsilon}$$

is a representation of the Heaviside $\Theta$ function.) For a general smooth function $f(t)$ we have

$$f(t) = \int_{-\infty}^{\infty} dt' \delta(t-t') f(t') = \int_{-\infty}^{\infty} dt' \frac{1}{2\pi}\int_{-\infty}^{\infty} dk e^{-ik(t-t')} f(t'), \tag{6.9}$$

which leads to

$$f(t) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}} e^{-ikt} \cdot \int_{-\infty}^{\infty} \frac{dt'}{\sqrt{2\pi}} e^{ikt'} f(t'). \tag{6.10}$$

We define the function

$$\tilde{f}(k) = \int_{-\infty}^{\infty} \frac{dt'}{\sqrt{2\pi}} e^{ikt'} f(t'). \tag{6.11}$$

Then, we have

$$f(t) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}} e^{-ikt} \tilde{f}(k). \tag{6.12}$$

Equations 6.11–6.12 define the Fourier transform of the function $f$ and its inverse. In solving differential equations, it is often beneficial to work with $\tilde{f}(k)$ rather than $f(t)$. We will make good use of Fourier transforms in finding Green's functions.

## 6.3 Fourier transformations and Green's functions

We seek Green's functions of the form

$$G(\vec{x},\vec{x}';t,t') = G(\vec{x}-\vec{x}',t-t'), \tag{6.13}$$

anticipating that they depend on space-time differences rather than on absolute positions and times. Writing $\vec{x}-\vec{x}' = \Delta\vec{x}$ and $t-t' = \Delta t$, the Green's function must satisfy

$$\Box G(\Delta\vec{x},\Delta t) = \delta(\Delta\vec{x})\delta(\Delta t). \tag{6.14}$$

We substitute into the above the Fourier transform

$$G(\Delta\vec{x},\Delta t) = \int \frac{d^3\vec{k}dE}{(2\pi)^4} e^{-i\left(Ec\Delta t - \vec{k}\cdot\Delta\vec{x}\right)} \tilde{G}(\vec{k},E). \tag{6.15}$$

We then find

$$\int \frac{d^3\vec{k}dE}{(2\pi)^4} e^{-i\left(Ect-\vec{k}\cdot\vec{x}\right)} \left[-E^2+\vec{k}^2\right] \tilde{G}(\vec{k},E) = c\int \frac{d^3\vec{k}dE}{(2\pi)^4} e^{-i\left(Ect-\vec{k}\cdot\vec{x}\right)}. \tag{6.16}$$

This gives for the Fourier transform of the Green's function

$$\tilde{G}(\vec{k}, E) = -\frac{c}{E^2 - \vec{k}^2}. \tag{6.17}$$

This solution is a bit naive, since it does not impose any physical boundary condition on the Green's function. In physical situations, we require that the effect on the vector and scalar potentials $\phi(\vec{x}, t), \vec{A}(\vec{x}, t)$ from the sources $\rho(\vec{x}', t'), \vec{j}(\vec{x}', t')$ takes place at times $t > t'$. In other words, if we switch on a source at a given time $t'$, we expect physically that this source will affect the potentials only at later times and not earlier. A physical Green's function must then satisfy the boundary condition

$$G(\Delta\vec{x}, \Delta t) = 0, \quad \text{for } \Delta t < 0. \tag{6.18}$$

This requirement is not a mathematical detail. In Eq. 6.6 the $t'$ integration runs over all times, including times later than $t$. With a generic Green's function, a change of the sources in the future would affect the potential now: causality would be violated. Maxwell's equations alone do not exclude this. For example, the "advanced" Green's function, with $t'$ later than $t$ by $|\vec{x} - \vec{x}'|/c$, also solves them. Causality is an additional physical input. We cannot simply drop the unwanted part of the solution. Instead, we slightly deform the theory, introducing a small parameter $\delta > 0$ as below, and take $\delta \to 0^+$ only at the level of the solutions. Limits of this kind depend on how they are taken. For example, $\log(-1)$ is $+i\pi$ or $-i\pi$ depending on whether $-1$ is approached from above or from below the real axis. The limiting prescription is therefore part of the definition of the theory. In quantum field theory we will meet a different prescription (the Feynman prescription), which yields the physically correct propagators there. We implement this requirement by modifying the Fourier transformation of the Green's function. We write

$$G(\Delta\vec{x}, \Delta t) = \int \frac{d^3\vec{k}dE}{(2\pi)^4} e^{-i\left((E+i\delta)c\Delta t - \vec{k}\cdot\Delta\vec{x}\right)} \tilde{G}(\vec{k}, E), \tag{6.19}$$

where $\delta$ is an infinitesimally small positive parameter. This Fourier transformation differs from the one of Eq. 6.15 in that the "energy" $E$ integration variable has a small positive imaginary part in the exponent. For $\Delta t \to -\infty$ the Green's function vanishes:

$$G(\Delta\vec{x}, \Delta t) \propto e^{\Delta t\delta} \to e^{-\infty\delta} \to 0, \tag{6.20}$$

which is in accordance with the physical requirement for the Green's function to vanish before the sources are turned on. Up to corrections of $O(\delta)$, we now find for the Fourier transform of the Green's function

$$\tilde{G}(\vec{k}, E) = \frac{-c}{(E + i\delta)^2 - \vec{k}^2} = \frac{c}{2k}\left[\frac{1}{E + k + i\delta} - \frac{1}{E - k + i\delta}\right], \tag{6.21}$$

where $k \equiv \left|\vec{k}\right|$. We can now compute the Green's function as

$$G(\Delta\vec{x}, \Delta t) = \frac{c}{2(2\pi)^4} \int \frac{d^3\vec{k}dE}{k} e^{-i\left((E+i\delta)c\Delta t - \vec{k}\cdot\Delta\vec{x}\right)} \left[\frac{1}{E + k + i\delta} - \frac{1}{E - k + i\delta}\right]. \tag{6.22}$$

To perform the integrations we use spherical coordinates,

$$\begin{aligned} d^3\vec{k} &= k^2 dk d\phi d\cos\theta \\ \vec{k}\cdot\Delta\vec{x} &= k\Delta x\cos\theta. \end{aligned} \tag{6.23}$$

After the angular $\phi, \theta$ integrations we arrive at

$$\begin{aligned} G(\Delta\vec{x}, \Delta t) &= \frac{ic}{2(2\pi)^3\,\Delta x}\int_0^\infty dk\ \left[e^{-ik\Delta x} - e^{+ik\Delta x}\right] \\ &\quad \int_{-\infty}^{\infty} dE\ e^{-i(E+i\delta)c\Delta t}\left[\frac{1}{E+k+i\delta} - \frac{1}{E-k+i\delta}\right]. \end{aligned} \tag{6.24}$$

We perform the "energy" $dE$ integration using Cauchy's theorem. First, we promote $E$ to a complex variable:

$$E = E_r + iE_i.$$

The integration runs along the horizontal axis of the $(E_r, E_i)$ plane. The time-dependent phase factor behaves as

$$e^{-ic\Delta t(E_r+iE_i)} = e^{-ic\Delta t E_r}e^{E_i c\Delta t} \to 0 \quad \text{for} \begin{cases} \Delta t < 0, E_i \to +\infty, \\ \quad \text{or} \\ \Delta t > 0, E_i \to -\infty. \end{cases} \tag{6.25}$$

Therefore, we can close the contour of integration in the upper $(E_r, E_i)$ half-plane for $\Delta t < 0$ and in the lower half-plane for $\Delta t > 0$. The upper half-plane contains no poles and therefore the integral is zero for negative $\Delta t$. We then have

$$G(\Delta\vec{x}, \Delta t) = 0, \quad \text{if } \Delta t < 0. \tag{6.26}$$

This is in agreement with what we have required for our Green's function, i.e. to vanish at times earlier than the time at which the sources are turned on. Notice that the vanishing is exact for **every** negative $\Delta t$, however small, and not only in the limit $\Delta t \to -\infty$ suggested by the damping factor $e^{\delta\Delta t}$. For any fixed $\Delta t < 0$, the contribution of the large semicircle $|E| = R$ vanishes as $R \to \infty$. Where $E_i > 0$ the exponential suppresses it. Near the real axis, where the exponential gives no suppression, the integrand still falls off as $1/E^2$, so the arc contributes $O(1/R)$. Causality is respected at arbitrarily short times. The integral for $\Delta t > 0$ does not vanish, since the lower half-plane contains two poles at $E = \pm k - i\delta$. With the residue theorem, we obtain

$$\begin{aligned} G(\Delta\vec{x}, \Delta t > 0) &= \frac{c}{2(2\pi)^2\Delta x}\int_0^\infty dk\,\left\{e^{-ik(c\Delta t-\Delta x)} + e^{+ik(c\Delta t-\Delta x)}\right. \\ &\quad \left. -e^{-ik(c\Delta t+\Delta x)} - e^{+ik(c\Delta t+\Delta x)}\right\} \\ &= \frac{c}{2(2\pi)^2\Delta x}\int_{-\infty}^\infty dk\,\left\{e^{-ik(c\Delta t-\Delta x)} - e^{+ik(c\Delta t+\Delta x)}\right\}. \end{aligned} \tag{6.27}$$

Finally, the last integral yields $\delta$-functions, according to Eq. 6.8. We then find

$$G(\Delta\vec{x}, \Delta t > 0) = \frac{c}{4\pi\left|\vec{x}-\vec{x}'\right|}\left[\delta(c\Delta t - \Delta x) - \delta(c\Delta t + \Delta x)\right]. \tag{6.28}$$

Since $\Delta t$ and $\Delta x$ are both positive, the argument of the second delta function never vanishes and

$$\delta(c\Delta t + \Delta x) = 0, \quad \text{for } \Delta x, \Delta t > 0. \tag{6.29}$$

Thus the Green's function becomes

$$G(\vec{x}-\vec{x}', t-t') = \frac{1}{4\pi\left|\vec{x}-\vec{x}'\right|}\delta\left(t - t' - \frac{\left|\vec{x}-\vec{x}'\right|}{c}\right)\Theta(t > t'), \tag{6.30}$$

where the $\Theta$-function is +1 for $t > t'$ and 0 otherwise.

Substituting into Eq. 6.6 and assuming no free electromagnetic fields $\phi_{\text{free}}, \vec{A}_{\text{free}}$ that are present in the absence of sources, we find

$$\phi(\vec{x},t) = \frac{1}{4\pi\epsilon_0}\int d^3\vec{x}'dt'\rho(\vec{x}',t')\frac{\delta\left(t-t'-\frac{|\vec{x}-\vec{x}'|}{c}\right)}{|\vec{x}-\vec{x}'|}\Theta(t>t'). \tag{6.31}$$

Performing the $t'$ integration yields

$$\phi(\vec{x},t) = \frac{1}{4\pi\epsilon_0}\int d^3\vec{x}'\frac{\rho\left(\vec{x}',t-\frac{|\vec{x}-\vec{x}'|}{c}\right)}{|\vec{x}-\vec{x}'|}. \tag{6.32}$$

We observe that for time-independent charge densities,

$$\rho(\vec{x}',t') = \rho(\vec{x}'),$$

Eq. 6.32 reproduces our known result from electrostatics. In the limit of an infinite speed of light $c \to \infty$, the potential becomes once again the known potential from electrostatics:

$$\lim_{c\to\infty}\phi(\vec{x},t) = \frac{1}{4\pi\epsilon_0}\int d^3\vec{x}'\frac{\rho\left(\vec{x}',t\right)}{|\vec{x}-\vec{x}'|}. \tag{6.33}$$

If $c$ were infinite, we would simply need Coulomb's law in order to compute the potential $\phi(\vec{x},t)$ at a time $t$, integrating over the contributions of the charge density $\rho(\vec{x},t)$ as it appears *at the same time* $t$. For a finite value of $c$, the potential $\phi(\vec{x},t)$ is due to the charge density $\rho(\vec{x}',t')$ as it appears at an earlier time $t - \frac{|\vec{x}-\vec{x}'|}{c}$. We interpret the time delay (retardation)

$$\delta t = \frac{|\vec{x}-\vec{x}'|}{c}$$

as the time required for the action of the source to propagate with the speed of light, in the form of an electromagnetic wave, up to the point $\vec{x}$ where we measure the potential. Two consequences are worth noting. First, the further away the sources are, the older the distribution that we perceive. Looking at large distances, as in astronomy, means looking back in time. Second, the formula tells us how to create electromagnetic waves. Suppose that we separate some positive from negative charges and then bring them back together. The potential due to this disturbance is not felt everywhere at once. It reaches a distance $r$ after a time $r/c$, with an amplitude suppressed by $1/r$. The response of the fields to the sources is linear. If we let the charges or currents oscillate, the potentials oscillate in the same way and the oscillation propagates outwards with the speed of light: this is an electromagnetic wave.

Similarly, we can calculate the vector potential $\vec{A}(\vec{x},t)$. We find

$$\vec{A}(\vec{x},t) = \frac{1}{4\pi\epsilon_0 c^2}\int d^3\vec{x}'\frac{\vec{J}\left(\vec{x}',t-\frac{|\vec{x}-\vec{x}'|}{c}\right)}{|\vec{x}-\vec{x}'|}. \tag{6.34}$$

The Green's function of Eq. 6.30 is known as the *retarded Green's function*, since it accounts for the retardation caused by the propagation of the electromagnetic signals from

one point to another with a finite speed. It is useful to derive an alternative form for the Green's function. We first prove that for real $a, b$

$$\begin{aligned} & \delta(a^2 - |b|^2)\Theta(a > 0) = \delta((a - |b|)(a + |b|))\Theta(a > 0) \\ = & \frac{1}{2|b|}\left[\delta(a - |b|) + \delta(a + |b|)\right]\Theta(a > 0) \\ = & \frac{1}{2|b|}\delta(a - |b|)\Theta(a > 0). \end{aligned} \tag{6.35}$$

Applying the above for $a = t - t'$ and $b = \frac{|\vec{x} - \vec{x}'|}{c}$, we can cast the retarded Green's function as

$$G(\vec{x} - \vec{x}', t - t') = \frac{1}{2\pi c}\delta\left((t - t')^2 - \frac{\left|\vec{x} - \vec{x}'\right|^2}{c^2}\right)\Theta(t > t'). \tag{6.36}$$

## 6.4 Potential of a moving charge with a constant velocity

As a first application of Eqs. 6.32–6.34, we compute the scalar and vector potentials of a point-like charge $q$ moving with a constant velocity $\vec{v}$. The charge density corresponding to the moving charge is

$$\rho(\vec{x}', t') = q\delta(\vec{x}' - \vec{v}t'). \tag{6.37}$$

Using the above charge density and the retarded Green's function in the form of Eq. 6.36, we obtain for the scalar potential

$$\begin{aligned}\Phi(\vec{x}, t) &= \frac{q}{2\pi\epsilon_0 c}\int d^3\vec{x}'dt'\delta(\vec{x}' - \vec{v}t')\delta\left((t-t')^2 - \frac{\left|\vec{x}-\vec{x}'\right|^2}{c^2}\right)\Theta(t > t') \\ &= \frac{q}{2\pi\epsilon_0 c}\int dt'\delta\left((t-t')^2 - \frac{\left|\vec{x}-\vec{v}t'\right|^2}{c^2}\right)\Theta(t > t').\end{aligned} \tag{6.38}$$

We now need to find the zeros of the argument of the delta function. We can write

$$0 = (t-t')^2 - \frac{\left|\vec{x}-\vec{v}t'\right|^2}{c^2} = t'^2 - 2t't + t^2 - \frac{(x_\parallel - vt')^2 + x_\perp^2}{c^2}, \tag{6.39}$$

where we have decomposed

$$\vec{x} = \vec{x}_\parallel + \vec{x}_\perp \tag{6.40}$$

into its components parallel and perpendicular to the velocity $\vec{v}$. We now define the "boosted" variables

$$x_b = \gamma\left(x_\parallel - vt\right), \quad t_b = \gamma\left(t - \frac{x_\parallel v}{c^2}\right), \tag{6.41}$$

where

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{6.42}$$

and we introduce the quantity

$$\tau^2 = c^2t^2 - (x_\parallel^2 + x_\perp^2) = c^2t_b^2 - (x_b^2 + x_\perp^2). \tag{6.43}$$

One can find the zeros of the quadratic form in $t'$ directly, but the algebra is lengthy and error-prone. The choice of variables above is motivated by physics. The argument of the $\delta$-function, $c^2(t-t')^2 - |\vec{x}-\vec{v}t'|^2$, is the space-time interval between the emission of a signal by the charge and its arrival at $\vec{x}$, which is the same for all inertial observers. Notice that at this stage $x_b, t_b$ are merely new names for the point $(\vec{x}, t)$ where we compute the potential. The integration variable $t'$ is not transformed, and we do not yet perform a change of reference frame. Equation 6.39 can then be cast in the form

$$\frac{t'^2}{\gamma^2} - 2\frac{t't_b}{\gamma} + \frac{\tau^2}{c^2} = 0. \tag{6.44}$$

The discriminant is

$$\Delta = \frac{4(t_b^2 - \tau^2/c^2)}{\gamma^2} = \frac{4r_b^2}{\gamma^2c^2}, \tag{6.45}$$

where we have defined the "boosted" distance

$$r_b^2 = x_b^2 + x_\perp^2. \tag{6.46}$$

The solutions are

$$t_\pm = \gamma\left(t_b \pm \frac{r_b}{c}\right). \tag{6.47}$$

The delta function becomes

$$\delta\left((t-t')^2 - \frac{\left|\vec{x}-\vec{v}t'\right|^2}{c^2}\right)\Theta(t>t') = \frac{c\gamma}{2r_b}\Theta(t>t')\delta(t'-t_-). \tag{6.48}$$

It is now trivial to perform the $t'$ integration in Eq. 6.38, obtaining

$$\Phi(\vec{x},t) = \frac{q}{4\pi\epsilon_0}\frac{\gamma}{r_b}. \tag{6.49}$$

Explicitly,

$$\Phi(\vec{x},t) = \frac{q}{4\pi\epsilon_0}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\frac{1}{\left[\left(\frac{x_\parallel - vt}{\sqrt{1-\frac{v^2}{c^2}}}\right)^2 + x_\perp^2\right]^{\frac{1}{2}}}. \tag{6.50}$$

Following the same steps, we find that the vector potential of a moving charge is (**exercise**):

$$\vec{A}(\vec{x},t) = \frac{\vec{v}}{c^2}\Phi(\vec{x},t). \tag{6.51}$$

In this derivation of the scalar and vector potentials of a moving charge, we see the "rise" of special relativity. In a reference frame where the charge is at rest, the scalar and vector potentials are

$$\Phi(\vec{x},t)\Big|_{\text{rest}} = \frac{q}{4\pi\epsilon_0}\frac{1}{\left[x_\parallel^2 + x_\perp^2\right]^{\frac{1}{2}}}, \quad \vec{A}(\vec{x},t)\Big|_{\text{rest}} = 0. \tag{6.52}$$

Comparing with the corresponding expressions in a frame where the charge is moving, we obtain an excellent hint that space coordinates transform under boosts according to Lorentz transformations,

$$x_\parallel \to \gamma\left(x_\parallel - vt\right), \quad x_\perp \to x_\perp, \tag{6.53}$$

and not according to Galilean transformations.

What about the emergence of the factor $\gamma$ and of a non-vanishing vector potential in the frame where the charge is moving? These are hints that the vector and scalar potentials are components of a single relativistic "four-vector":

$$\begin{pmatrix} \Phi \\ \vec{A} \end{pmatrix} \equiv A^\mu.$$

*Remark.* In the SI units used here, $\Phi$ and $\vec{A}$ have different dimensions, and the four-vector is $A^\mu = (\Phi/c, \vec{A})$. Indeed $\vec{A} = \frac{\vec{v}}{c^2}\Phi$ means $(\Phi/c, \vec{A}) = \gamma\,\frac{\Phi_{\text{rest}}}{c}\,(1, \vec{v}/c)$ at the corresponding point, i.e. $\Lambda^\mu_{\ 0}$ times the rest-frame vector $(\Phi_{\text{rest}}/c, \vec{0})$. The form $(\Phi, \vec{A})$ holds in Gaussian units (see Sec. 11.9 of Ref. [2]) or Heaviside–Lorentz units with $c = 1$, as used in the later chapters.

# 7

# Special relativity

Special relativity is based on the assumption that the laws of nature are the same for all inertial observers whose coordinates are related by Lorentz transformations,

$$x^\mu \to x^{\mu\prime} = \Lambda^\mu_{\ \nu} x^\nu + \rho^\mu, \tag{7.1}$$

where

$$x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{7.2}$$

is a "four-vector" comprising the space-time coordinates $x^0 = ct, x^1 = x, x^2 = y, x^3 = z$. Here $\rho^\mu$ is a constant four-vector and $\Lambda^\mu_{\ \nu}$ satisfies

$$\Lambda^\mu_{\ \rho} \Lambda^\nu_{\ \sigma} g_{\mu\nu} = g_{\rho\sigma}. \tag{7.3}$$

The $4 \times 4$ matrix $g_{\mu\nu}$ is the so-called *metric*, defined as

$$g_{\mu\nu} = \begin{cases} +1, & \mu = \nu = 0, \\ -1, & \mu = \nu = 1, 2, 3, \\ 0, & \mu \neq \nu. \end{cases} \tag{7.4}$$

Two equivalent conventions for the metric are in use. Our choice, with diagonal entries $(+1, -1, -1, -1)$, is common in particle physics, while the choice with diagonal entries $(-1, +1, +1, +1)$ is common in general relativity and cosmology. With our choice the invariant square of the four-momentum of a particle (defined later) is positive, $p^\mu p_\mu = m^2 c^2$. The physics does not depend on the convention. It is a good exercise to translate results from one convention to the other; one can then read every book in the literature. In the above we have used Einstein's summation convention. For example, one would write explicitly

$$\Lambda^\mu_{\ \nu} x^\nu = \Lambda^\mu_{\ 0} x^0 + \Lambda^\mu_{\ 1} x^1 + \Lambda^\mu_{\ 2} x^2 + \Lambda^\mu_{\ 3} x^3. \tag{7.5}$$

This is a convention that we will use extensively from now on.

## 7.1 Proper time

Lorentz transformations leave "proper-time" intervals invariant. These are defined as

$$d\tau^2 \equiv c^2 dt^2 - d\vec{x}^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{7.6}$$

*Remark.* With this definition, $\tau$ has the dimension of a length: it is $c$ times the proper time $d\tau_J = dt/\gamma$ of Ref. [2] (Eq. (11.26)). The formulae of this chapter ($d\tau = c\,dt/\gamma$, $p^\mu = mc\,dx^\mu/d\tau$, $f^\mu = mc^2\,d^2x^\mu/d\tau^2$) consistently use this convention. Indeed, in a different reference frame we have from Eq. 7.1,

$$dx^{\mu\prime} = \Lambda^\mu_{\ \nu} dx^\nu. \tag{7.7}$$

A proper-time interval in the new frame is

$$\begin{aligned} d\tau'^2 &= g_{\mu\nu} dx'^\mu dx'^\nu \\ &= g_{\mu\nu} \left(\Lambda^\mu_{\ \rho} dx^\rho\right) \left(\Lambda^\nu_{\ \sigma} dx^\sigma\right) \\ &= \left(g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma}\right) dx^\rho dx^\sigma \\ &= g_{\rho\sigma} dx^\rho dx^\sigma = d\tau^2. \end{aligned} \tag{7.8}$$

As a consequence of the invariance of proper-time intervals, *the speed of light is the same in all inertial frames*. Indeed, for light we have

$$\left|\frac{d\vec{x}}{dt}\right| = c \rightsquigarrow d\tau^2 = c^2 dt^2 - d\vec{x}^2 = 0. \tag{7.9}$$

In a new frame,

$$d\tau^{2\prime} = d\tau^2 = 0 \rightsquigarrow \left|\frac{d\vec{x}'}{dt'}\right| = c. \tag{7.10}$$

Lorentz transformations are the only non-singular transformations which preserve proper-time intervals. Indeed,

$$\begin{aligned} d\tau^2 &= d\tau'^2 \\ \rightsquigarrow g_{\rho\sigma} dx^\rho dx^\sigma &= g_{\mu\nu} dx'^\mu dx'^\nu \\ \rightsquigarrow g_{\rho\sigma} dx^\rho dx^\sigma &= g_{\mu\nu} \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial x'^\nu}{\partial x^\sigma} dx^\rho dx^\sigma, \end{aligned} \tag{7.11}$$

we conclude that

$$g_{\rho\sigma} = g_{\mu\nu} \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial x'^\nu}{\partial x^\sigma}. \tag{7.12}$$

Differentiating with respect to $x^\epsilon$, we obtain

$$0 = g_{\mu\nu} \left[\frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho} \frac{\partial x'^\nu}{\partial x^\sigma} + \frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\sigma} \frac{\partial x'^\nu}{\partial x^\rho}\right]. \tag{7.13}$$

To this, we add the same equation with $\epsilon \leftrightarrow \rho$ and subtract the same equation with $\epsilon \leftrightarrow \sigma$. We obtain

$$0 = 2 g_{\mu\nu} \frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho} \frac{\partial x'^\nu}{\partial x^\sigma}. \tag{7.14}$$

Assuming that the transformation $x^\mu \longrightarrow x'^\mu$ is a well-behaved differentiable function and that the inverse of the transformation also exists,

$$\frac{\partial x'^\mu}{\partial x^\sigma} \frac{\partial x^\sigma}{\partial x'^\nu} = \delta^\mu_{\ \nu}, \tag{7.15}$$

we obtain that

$$\frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho} = 0. \tag{7.16}$$

Therefore, the transformation $x^\mu \to x'^\mu$ ought to be linear:

$$x'^\mu = \Lambda^\mu_{\ \nu} x^\nu + \rho^\mu. \tag{7.17}$$

The two inputs of this derivation have a physical meaning. The inverse transformation must exist because we must always be able to return to the original frame. The metric $g_{\mu\nu}$ is the same constant matrix at every point of space-time. Linearity, i.e. $\Lambda$ and $\rho$ independent of $x$, means that no point of space-time is treated differently from any other. In general relativity, massive bodies make the regions of space-time around them special. There the metric depends on the position and the transformations between coordinate systems are no longer linear.

## 7.2 Subgroups of Lorentz transformations

The set of all Lorentz transformations

$$x'^\mu = \Lambda^\mu_{\ \nu} x^\nu + \rho^\mu, \quad g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma}, \tag{7.18}$$

form a group (**exercise:** prove it), which is known as the *inhomogeneous Lorentz group* or the Poincaré group. The subset of transformations with $\rho^\mu = 0$ is known as the *homogeneous Lorentz group*.

From

$$g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma}$$

and for $\rho = \sigma = 0$, we have:

$$(\Lambda^0_{\ 0})^2 - \sum_{i=1}^{3} (\Lambda^i_{\ 0})^2 = 1 \rightsquigarrow (\Lambda^0_{\ 0})^2 \geq 1. \tag{7.19}$$

Also, in matrix form the definition of the Lorentz transformation becomes

$$g = \Lambda^T g \Lambda \rightsquigarrow \det g = \det(\Lambda^T g \Lambda) \rightsquigarrow (\det \Lambda)^2 = 1. \tag{7.20}$$

The subgroup of transformations with

$$\det \Lambda = 1, \ \Lambda^0_{\ 0} \geq 1,$$

which contains the unit element $\mathbf{1} = \delta^\mu_{\ \nu}$, is known as the group of *proper* Lorentz transformations. All other transformations are known as *improper Lorentz transformations*. It is impossible to go from

$$\det \Lambda = 1 \to \det \Lambda = -1 \text{ or } \Lambda^0_{\ 0} \geq 1 \to \Lambda^0_{\ 0} \leq -1$$

by a continuous change of parameters. Improper Lorentz transformations involve either space-reflection ($\det \Lambda = -1, \Lambda^0_{\ 0} \geq 1$) or time reversal ($\det \Lambda = -1, \Lambda^0_{\ 0} \leq -1$) or both ($\det \Lambda = 1, \Lambda^0_{\ 0} \leq -1$). The transformations with $\det \Lambda = -1$ do not form a subgroup by themselves, since they do not contain the unit element. The condition $\Lambda^0_{\ 0} \geq 1$ has a simple meaning. For two ticks of a clock at rest in one frame, $dx^\mu = (c\, dt, \vec{0})$, the time interval in another frame is $dt' = \Lambda^0_{\ 0}\, dt$. With $\Lambda^0_{\ 0} \geq 1$ time flows forward for all observers. Notice that the bound is $\Lambda^0_{\ 0} \geq 1$ and not just $\Lambda^0_{\ 0} > 0$: we will find below that $\Lambda^0_{\ 0} = \gamma$. An inertial observer cannot reverse the flow of time, nor reflect space as in a mirror, by changing

velocity or orientation. The proper Lorentz transformations are therefore those which relate physical observers.

Proper homogeneous or inhomogeneous Lorentz transformations have a further subgroup: the group of rotations,

$$\Lambda^0_{\ 0} = 1,\ \Lambda^i_{\ 0} = \Lambda^0_{\ i} = 0,\ \Lambda^i_{\ j} = R_{ij}, \tag{7.21}$$

with

$$\det R = 1, \quad R^T R = 1. \tag{7.22}$$

Thus, for rotations and translations ($x^\mu \to x'^\mu = x^\mu + \rho^\mu$), Lorentz transformations are no different from Galilean transformations.

A difference from Galilean transformations arises for *boosts*. Consider a reference frame $O$ in which a certain particle appears at rest, and a reference frame $O'$ in which the particle appears to move with a velocity $\vec{v}$. Space-time intervals in the two frames are related via

$$dx'^\mu = \Lambda^\mu_{\ \nu} dx^\nu = \Lambda^\mu_{\ 0} c\,dt, \tag{7.23}$$

given that $d\vec{x} = 0$ in the frame $O$. For $\mu = 0$, this equation gives

$$dt' = \Lambda^0_{\ 0} dt. \tag{7.24}$$

For $\mu = i = 1, 2, 3$ we have

$$dx'^i = \Lambda^i_{\ 0} c\,dt. \tag{7.25}$$

Dividing the two, we have

$$v^i \equiv \frac{dx'^i}{dt'} = c\frac{\Lambda^i_{\ 0}}{\Lambda^0_{\ 0}} \rightsquigarrow \Lambda^i_{\ 0} = \frac{v^i}{c}\Lambda^0_{\ 0}. \tag{7.26}$$

From

$$g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma}$$

and for $\rho = \sigma = 0$, we have

$$\begin{aligned} & (\Lambda^0_{\ 0})^2 - (\Lambda^i_{\ 0})^2 = 1 \\ \rightsquigarrow \quad & \Lambda^0_{\ 0} = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \end{aligned} \tag{7.27}$$

and thus

$$\Lambda^i_{\ 0} = \gamma\frac{v^i}{c}. \tag{7.28}$$

The remaining components are not determined uniquely by knowing the velocity $\vec{v}$ of the particle. Indeed, two Lorentz transformations

$$\Lambda^\mu_{\ \nu} \quad \text{and} \quad \Lambda^\mu_{\ \rho} R^\rho_{\ \nu},$$

where $R$ is a rotation, boost a particle to the same velocity. For coordinate systems $O$ and $O'$ with parallel axes we find that (**exercise**)

$$\Lambda^i_{\ j} = \delta^i_j + \frac{v^i v^j}{v^2}(\gamma - 1) \tag{7.29}$$

and

$$\Lambda^0_{\ j} = \gamma\frac{v^j}{c}. \tag{7.30}$$

## 7.3 Time dilation

Consider an inertial observer $O$ who looks at a clock at rest. Two ticks of the clock correspond to a space-time interval

$$d\vec{x} = 0, \quad dt = \Delta t. \tag{7.31}$$

The proper-time interval is

$$d\tau = (c^2 dt^2 - d\vec{x}^2)^{\frac{1}{2}} = c\Delta t. \tag{7.32}$$

A second observer sees the clock moving with velocity $\vec{v}$. Two ticks of the clock define a space-time interval

$$dt' = \Delta t', \quad d\vec{x}' = \vec{v} dt'. \tag{7.33}$$

The proper-time interval in the new frame is

$$d\tau' = (c^2 dt'^2 - d\vec{x}'^2)^{\frac{1}{2}} = c\Delta t' \sqrt{1 - \left|\frac{d\vec{x}'}{c dt'}\right|^2} = c\Delta t' \sqrt{1 - \frac{\vec{v}^2}{c^2}}. \tag{7.34}$$

The proper time is invariant under a change of inertial reference frame. Thus we conclude that

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} = \gamma \Delta t. \tag{7.35}$$

## 7.4 Doppler effect

Take our clock to be a source of light with a frequency

$$\omega = \frac{2\pi}{\Delta t}.$$

For an observer who sees the light source moving with velocity $\vec{v}$, this time interval is measured to be

$$dt' = \gamma \Delta t.$$

In the same period, the distance of the observer from the light source increases by

$$v_r dt',$$

where $v_r$ is the component of the velocity of the light source along the line of sight of the observer. The time $dt_0$ elapsing between the reception of two successive light wave fronts by the observer is given by

$$c dt_0 = c dt' + v_r dt'. \tag{7.36}$$

The frequency measured by the observer is

$$\omega' = \frac{2\pi}{dt_0} = \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v_r}{c}} \omega. \tag{7.37}$$

If the light source is moving along the line of sight, $v_r = v$, we have

$$\omega' = \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \omega. \tag{7.38}$$

If the light source moves away from the observer, $v_r > 0$, the frequency decreases and the light appears redder (red shift). If the source moves towards the observer, the frequency increases (blue shift). Two effects are combined in the above formula. The denominator $1 + \frac{v_r}{c}$ is the classical Doppler effect, which is also familiar from the siren of a passing ambulance. Between two wave crests the source moves away by $v_r dt'$. This factor exists for any wave, e.g. sound, with $c$ the speed of the wave. The numerator $\sqrt{1 - \frac{v^2}{c^2}}$ is due to time dilation and is a purely relativistic effect. It depends on the magnitude of the velocity and not only on its component along the line of sight. Even for a source moving perpendicularly to the line of sight, $v_r = 0$, the frequency is shifted, $\omega' = \omega\sqrt{1 - \frac{v^2}{c^2}}$ (transverse Doppler effect), while no shift is expected classically. The Doppler effect is a powerful instrument of astronomy. Atoms such as hydrogen emit light at sharp frequencies, fixed by transitions between their discrete energy levels. Comparing the frequencies of these lines in the light of a star with those measured in the laboratory, we find whether, and how fast, the star moves towards or away from us. Measuring in this way the velocities of the stars in a galaxy, one finds that far from the centre they move faster than the gravitational attraction of the visible matter can account for. This is one of the indications for dark matter.
**Exercise:** Calculate the angle between the direction of motion of the light source and the line of sight of the observer for which there is no shift in the frequency.

For an application of the Doppler effect in cosmology, read about Hubble's law.

## 7.5 Particle dynamics

How can we compute the force on a particle which moves with a relativistic velocity $\vec{v}$? We should expect that our classical formulae from Newtonian mechanics need to be modified. Nevertheless, Newtonian expressions for the force should be valid if a particle is at rest. We can always change reference frame with a Lorentz transformation to bring a particle to rest, and calculate the change in its velocity for a small time interval using Newtonian mechanics. However, we would need to perform such a change of reference frame at every small increase of the velocity of the particle during its acceleration due to the force.

For a more elegant solution of the problem, we define a relativistic force acting on a particle as

$$f^\mu = mc^2 \frac{d^2 x^\mu}{d\tau^2}, \tag{7.39}$$

where $m$ is the mass of the particle.[1] If the particle is at rest, the proper-time interval $d\tau$ coincides (up to the factor $c$) with the ordinary time interval $dt$,

$$d\tau = c dt.$$

Therefore, in the rest frame of the particle, the "space" components of the force four-vector become

$$f^i_{\text{rest}} = m \frac{d^2 x^i}{dt^2} = F^i_{\text{Newton}}, \quad \text{for } i = 1, 2, 3, \tag{7.40}$$

where $\vec{F}_{\text{Newton}}$ is the force vector as we know it from Newtonian mechanics. The "time" component of the force four-vector vanishes:

$$f^0_{\text{rest}} = mc \frac{d^2 t}{dt^2} = 0. \tag{7.41}$$

[1] By mass we mean the mass of a particle as it is measured in its rest frame. We refrain from using the "relativistic", velocity-dependent, mass.

Under a Lorentz transformation, $f^\mu$ transforms as

$$f^\mu = mc^2 \frac{d^2 x^\mu}{d\tau^2} \to mc^2 \frac{d^2 x'^\mu}{d\tau^2} = mc^2 \frac{d^2 \left(\Lambda^\mu_{\ \nu} x^\nu + \rho^\mu\right)}{d\tau^2} = \Lambda^\mu_{\ \nu}\, mc^2 \frac{d^2 x^\nu}{d\tau^2}. \tag{7.42}$$

Therefore,

$$f'^\mu = \Lambda^\mu_{\ \nu} f^\nu. \tag{7.43}$$

The components of $f^\mu$ transform under Lorentz transformations in exactly the same way as the space-time coordinates. It is therefore a four-vector as well.

For a specific transformation from the rest frame of a particle to a frame where the particle moves with a velocity $\vec{v}$, we have

$$f^\mu = \Lambda^\mu_{\ \nu}(\vec{v}) f^\nu_{\text{rest}}, \tag{7.44}$$

where we have found that

$$\begin{aligned} \Lambda^0_{\ 0}(\vec{v}) = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \quad \Lambda^i_{\ 0}(\vec{v}) = \Lambda^0_{\ i}(\vec{v}) = \gamma \frac{v^i}{c}, \\ \Lambda^i_{\ j}(\vec{v}) = \delta^i_j + \frac{v^i v^j}{v^2}(\gamma - 1). \end{aligned} \tag{7.45}$$

Therefore, the force on a moving particle is

$$\vec{f} = \vec{F}_{\text{Newton}} + (\gamma - 1) \frac{\vec{v}\left(\vec{F}_{\text{Newton}} \cdot \vec{v}\right)}{v^2}, \tag{7.46}$$

and

$$f^0 = \gamma \frac{\vec{v} \cdot \vec{F}_{\text{Newton}}}{c} = \frac{\vec{v}}{c} \cdot \vec{f}. \tag{7.47}$$

The Newtonian limit is not manifest at first sight, since $\frac{\vec{v}(\vec{F}_{\text{Newton}} \cdot \vec{v})}{v^2}$ does not vanish as $v \to 0$. However, $\frac{\gamma - 1}{v^2} \to \frac{1}{2c^2}$, so that the second term of $\vec{f}$ is of order $v^2/c^2$, and $f^0$ also vanishes for small velocities. We recover Newton's law. Notice also that the particle is allowed to accelerate. Special relativity requires only that the **observer** is inertial. The particles that the observer studies may experience forces and accelerate, and the formalism above describes exactly that situation.

In Newtonian mechanics, if the force $\vec{F}$ is given, we can compute the trajectory $\vec{x}(t)$ by solving the second-order differential equation

$$\frac{d^2 \vec{x}}{dt^2} = \frac{\vec{F}(\vec{x}, t)}{m}. \tag{7.48}$$

Similarly, in special relativity, when the relativistic force $f^\mu$ is known, the differential equation 7.39 can, in principle, be solved to give the space-time coordinates as a function of the proper time $\tau$,

$$x^\mu = x^\mu(\tau). \tag{7.49}$$

To calculate the trajectory, we then need to express the proper time in terms of the time coordinate by inverting

$$x^0 = x^0(\tau) \rightsquigarrow \tau = \tau(x^0), \tag{7.50}$$

which we can use to cast the space components directly as functions of the time coordinate.

The four equations (7.39) are, however, not four independent equations for four unknown functions. In the rest frame the $\mu = 0$ equation reads $0 = 0$ and carries no information on how $t$ depends on $\tau$. Equivalently, $f^0 = \frac{\vec{v}}{c} \cdot \vec{f}$ is not an independent input: it is fixed by the space components. One equation is missing, and it is provided by a constraint. This is an additional condition that must be satisfied by our solutions $x^\mu(\tau)$, namely

$$\Omega \equiv g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 1. \tag{7.51}$$

For the derivative of $\Omega$ with respect to the proper time we have

$$\frac{d\Omega}{d\tau} = 2 g_{\mu\nu} \frac{d^2x^\mu}{d\tau^2} \frac{dx^\nu}{d\tau} = \frac{2}{mc^2} g_{\mu\nu} f^\mu \frac{dx^\nu}{d\tau}. \tag{7.52}$$

The right-hand side is a Lorentz-invariant quantity. In a new frame,

$$g_{\mu\nu} f'^\mu \frac{dx'^\nu}{d\tau} = g_{\mu\nu} \left( \Lambda^\mu_{\ \rho} f^\rho \right) \frac{(\Lambda^\nu_{\ \sigma} dx^\sigma)}{d\tau} = \left( g_{\mu\nu} \Lambda^\mu_{\ \rho} \Lambda^\nu_{\ \sigma} \right) f^\rho \frac{dx^\sigma}{d\tau} = g_{\rho\sigma} f^\rho \frac{dx^\sigma}{d\tau}.$$

We are therefore allowed to compute $\frac{d\Omega}{d\tau}$ in any reference frame we wish. Let us choose the rest frame of the particle, where

$$dx^\mu = (cdt, \vec{0}), \quad f^\mu = (0, \vec{F}_{\text{Newton}}).$$

We obtain

$$\frac{d\Omega}{d\tau} = \frac{2}{mc^2} \left( f^0 \frac{dx^0}{d\tau} - \vec{f} \cdot \frac{d\vec{x}}{d\tau} \right) = 0. \tag{7.53}$$

Therefore, the quantity $\Omega$ is a constant of motion:

$$\Omega(\tau) = \text{constant}. \tag{7.54}$$

If for some initial value $\tau_0$ we choose the constant to be one, we will have

$$\Omega(\tau) = \Omega(\tau_0) = 1, \qquad \forall\ \tau. \tag{7.55}$$

**Exercise:** Calculate the trajectory of a particle on which the exerted four-vector force is $f^\mu = 0$.

## 7.6 Energy and momentum

We can define a relativistic four-vector analogue of momentum as

$$p^\mu = mc \frac{dx^\mu}{d\tau}. \tag{7.56}$$

We have that

$$d\tau = \left( c^2 dt^2 - d\vec{x}^2 \right)^{\frac{1}{2}} = cdt \left[ 1 - \left( \frac{d\vec{x}}{cdt} \right)^2 \right]^{\frac{1}{2}} = cdt \left[ 1 - \frac{\vec{v}^2}{c^2} \right]^{\frac{1}{2}} = \frac{cdt}{\gamma}. \tag{7.57}$$

Thus, for the time component ($\mu = 0$) of the four-momentum we have

$$p^0 = mc \frac{dx^0}{d\tau} = m\gamma c. \tag{7.58}$$

For the space components ($\mu = i = 1, 2, 3$) we have

$$p^i = mc\frac{dx^i}{d\tau} = m\gamma\frac{dx^i}{dt} = m\gamma v^i. \tag{7.59}$$

For small velocities, we can expand the factor $\gamma$ as

$$\gamma = \left[1 - \frac{v^2}{c^2}\right]^{-\frac{1}{2}} \approx 1 + \frac{1}{2}\frac{v^2}{c^2} + \mathcal{O}\left(\frac{v^4}{c^4}\right). \tag{7.60}$$

Therefore, for small velocities the space components of the four-momentum become the classical momentum,

$$p^i \approx mv^i + \ldots, \tag{7.61}$$

while the time component becomes

$$p^0 \approx mc + \frac{1}{2c}mv^2 + \ldots. \tag{7.62}$$

In the second term of the above expansion we recognise the kinetic energy $\frac{1}{2}mv^2$ of the particle. We then identify the relativistic energy of a particle with

$$E = cp^0 = m\gamma c^2. \tag{7.63}$$

Eliminating the velocity $\vec{v}$ from Eqs. 7.59 and 7.63, we obtain the relation

$$E = \sqrt{\vec{p}^2c^2 + m^2c^4}. \tag{7.64}$$

The identification $E = cp^0$ is not a mathematical proof. The Taylor expansion gives us only a hint, and it was a bold step of Einstein to generalise the theory in this way. Such a step must be tested by experiment. It makes a new prediction: a particle at rest has an energy $E = mc^2$, which never appeared in classical mechanics. This is confirmed in the decays of unstable particles. A particle decaying at rest produces particles whose energies, including their kinetic energies, add up to the rest energy $mc^2$ of the decaying particle. The energy-momentum relation above, in place of the non-relativistic $E = mc^2 + \frac{\vec{p}^2}{2m}$, is tested daily at particle accelerators, where the momenta and the energies of fast particles are measured independently. These relations hold for free particles.

## 7.7 The inverse of a Lorentz transformation

Recall the metric matrix

$$g_{\mu\nu} = \text{diag}(1, -1, -1, -1). \tag{7.65}$$

We define its inverse

$$g^{\mu\nu}: \quad g^{\mu\nu}g_{\nu\rho} = \delta^\mu_\rho, \tag{7.66}$$

where $\delta^\mu_\nu$ is the Kronecker delta. It is easy to verify that the inverse of the metric is the metric itself:

$$g^{\mu\nu} = g_{\mu\nu} = \text{diag}(1, -1, -1, -1). \tag{7.67}$$

Now consider a Lorentz transformation $\Lambda^\mu_\nu$, which satisfies the identity

$$\Lambda^\mu_\rho\Lambda^\nu{}_\sigma g_{\mu\nu} = g_{\rho\sigma}. \tag{7.68}$$

We can prove that the matrix

$$\Lambda_{\mu}^{\ \nu} \equiv g_{\mu\rho} g^{\nu\sigma} \Lambda^{\rho}_{\ \sigma} \tag{7.69}$$

is the inverse of $\Lambda^{\mu}_{\ \nu}$. Indeed,

$$\Lambda^{\mu}_{\ \lambda} \Lambda_{\mu}^{\ \nu} = g_{\mu\rho} g^{\nu\sigma} \Lambda^{\rho}_{\ \sigma} \Lambda^{\mu}_{\ \lambda} = g_{\sigma\lambda} g^{\nu\sigma} = \delta^{\nu}_{\ \lambda}. \tag{7.70}$$

If $\Lambda^{\mu}_{\ \nu}$ is the boost with velocity $\vec{v}$ of Eq. 7.45, then

$$\begin{aligned} \Lambda_0^{\ 0}(\vec{v}) = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \quad \Lambda_i^{\ 0}(\vec{v}) = \Lambda_0^{\ i}(\vec{v}) = -\gamma \frac{v^i}{c}, \\ \Lambda_i^{\ j}(\vec{v}) = \delta_i^j + \frac{v^i v^j}{v^2}(\gamma - 1). \end{aligned} \tag{7.71}$$

We therefore have that the inverse of a boost is

$$\Lambda_{\mu}^{\ \nu}(\vec{v}) = \Lambda^{\mu}_{\ \nu}(-\vec{v}), \tag{7.72}$$

as we also expect physically.

## 7.8 Vectors and tensors

It is now time to give a formal definition of vectors in special relativity. We call any set of four components which transform according to the rule

$$V^{\mu} \longrightarrow V'^{\mu} = \Lambda^{\mu}_{\ \nu} V^{\nu} \tag{7.73}$$

a *contravariant* vector. Contravariant vectors transform in the same way as space-time coordinates $x^{\mu}$ do under homogeneous Lorentz transformations.

Not all vectors transform as contravariant vectors. Consider the derivative $\frac{\partial}{\partial x^{\mu}}$. Under a Lorentz transformation, it transforms as

$$\frac{\partial}{\partial x^{\mu}} \longrightarrow \frac{\partial}{\partial x'^{\mu}} = \frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial}{\partial x^{\rho}}. \tag{7.74}$$

We have that

$$\left(\frac{\partial x^{\rho}}{\partial x'^{\mu}}\right)\left(\frac{\partial x'^{\mu}}{\partial x^{\nu}}\right) = \delta^{\rho}_{\ \nu} \rightsquigarrow \left(\frac{\partial x^{\rho}}{\partial x'^{\mu}}\right) \Lambda^{\mu}_{\ \nu} = \delta^{\rho}_{\ \nu}. \tag{7.75}$$

Therefore, $\left(\frac{\partial x^{\rho}}{\partial x'^{\mu}}\right)$ is the inverse of a Lorentz transformation $\Lambda^{\mu}_{\ \nu}$,

$$\frac{\partial x^{\nu}}{\partial x'^{\mu}} = \Lambda_{\mu}^{\ \nu}. \tag{7.76}$$

Substituting into Eq. 7.74, we find

$$\frac{\partial}{\partial x^{\mu}} \longrightarrow \frac{\partial}{\partial x'^{\mu}} = \Lambda_{\mu}^{\ \rho} \frac{\partial}{\partial x^{\rho}}. \tag{7.77}$$

We find that the derivative does not transform according to the Lorentz transformation but according to its inverse. All vectors which transform with the inverse Lorentz transformation,

$$U_{\mu} \longrightarrow U'_{\mu} = \Lambda_{\mu}^{\ \nu} U_{\nu}, \tag{7.78}$$

are called *covariant* vectors.

For every contravariant vector $U^\mu$ there is a dual vector

$$U_\mu = g_{\mu\nu} U^\nu. \tag{7.79}$$

We can invert the above equation by multiplying with $g^{\rho\mu}$,

$$g^{\rho\mu} U_\mu = g^{\rho\mu} g_{\mu\nu} U^\nu = \delta^\rho_{\ \nu} U^\nu = U^\rho. \tag{7.80}$$

The dual vector $U_\mu$ is a covariant vector. Indeed, under a Lorentz transformation we have

$$U_\mu \to U'_\mu = g_{\mu\nu} U'^\nu = g_{\mu\nu} \Lambda^\nu_{\ \rho} U^\rho = g_{\mu\nu} \Lambda^\nu_{\ \rho} g^{\rho\sigma} U_\sigma = \Lambda_\mu^{\ \sigma} U_\sigma. \tag{7.81}$$

The scalar product of a contravariant and a covariant vector

$$A \cdot B \equiv A^\mu B_\mu = A_\mu B^\mu = g_{\mu\nu} A^\mu B^\nu = g^{\mu\nu} A_\mu B_\nu, \tag{7.82}$$

is invariant under Lorentz transformations. Indeed,

$$A \cdot B \to A' \cdot B' = A'^\mu B'_\mu = \Lambda^\mu_{\ \rho} A^\rho \Lambda_\mu^{\ \sigma} B_\sigma = \delta^\sigma_{\ \rho} A^\rho B_\sigma = A^\rho B_\rho = A \cdot B. \tag{7.83}$$

Invariants such as the scalar product are our main tool for building theories which respect special relativity. The laws of a system follow from the variation of its action. If the action is written in terms of quantities which are invariant under Lorentz transformations, the resulting laws are automatically the same in every inertial frame. This is how one constructs new models, e.g. for physics at high energies, without the risk of violating special relativity. In practical calculations, notice that in a correct expression every summed index appears exactly twice, once up and once down. An index which appears three times signals a mistake.

Let us define for short

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \tag{7.84}$$

and the dual contravariant vector

$$\partial^\mu = \frac{\partial}{\partial x_\mu} = g^{\mu\nu} \partial_\nu. \tag{7.85}$$

The d'Alembert operator is the scalar product

$$\Box \equiv \partial^2 \equiv \partial_\mu \partial^\mu = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2. \tag{7.86}$$

Being a scalar product, the d'Alembert operator is invariant under Lorentz transformations.

Finally, we define a tensor with multiple "up" and/or "down" indices to be an object

$$T^{\mu_1 \mu_2 \cdots}_{\nu_1 \nu_2 \ldots} \tag{7.87}$$

which transforms as

$$T^{\mu_1 \mu_2 \cdots}_{\nu_1 \nu_2 \ldots} \to \Lambda^{\mu_1}_{\ \rho_1} \Lambda^{\mu_2}_{\ \rho_2} \ldots \Lambda_{\nu_1}^{\ \sigma_1} \Lambda_{\nu_2}^{\ \sigma_2} \ldots T^{\rho_1 \rho_2 \cdots}_{\sigma_1 \sigma_2 \ldots}. \tag{7.88}$$

## 7.9 Currents and densities

Consider a set of particles $\{n\}$ with charges $q_n$ at positions $\vec{r}_n(t)$. The charge and current densities are

$$\rho(\vec{x},t) = \sum_n q_n \delta(\vec{x} - \vec{r}_n(t)), \tag{7.89}$$

$$\vec{j}(\vec{x},t) = \sum_n q_n \frac{d\vec{r}_n(t)}{dt} \delta(\vec{x} - \vec{r}_n(t)) = \sum_n q_n \frac{d\vec{x}}{dt} \delta(\vec{x} - \vec{r}_n(t)). \tag{7.90}$$

In the above, we can set $\vec{r}_n = \vec{x}$ because of the delta function, which vanishes for $\vec{x} \neq \vec{r}_n$. This is only a notational shorthand: $\vec{x}$ is the (independent) field point, so $d\vec{x}/dt$ means the velocity $d\vec{r}_n/dt$ of the particle located at $\vec{x}$. Below, $dx^\mu/dt$ and $dx^\mu/d\tau$ inside the sums should be read as $dr_n^\mu/dt$ and $dr_n^\mu/d\tau$ (cf. Sec. 2.8 of Ref. [6], with $c = 1$). Recall now that for the vector $x^\mu = (ct, \vec{x})$, we have

$$\frac{dx^\mu}{dt} = \frac{d}{dt}(ct, \vec{x}) = \left(c, \frac{d\vec{x}}{dt}\right). \tag{7.91}$$

We can then combine the charge and current densities into one object,

$$j^\mu \equiv \left(c\rho, \vec{j}\right), \tag{7.92}$$

with

$$j^\mu(\vec{x},t) = \sum_n q_n \frac{dx^\mu}{dt} \delta(\vec{x} - \vec{r}_n(t)). \tag{7.93}$$

We can now cast $j^\mu$ in a form which manifestly shows that it is a contravariant four-vector. First, we define a delta function in four dimensions as

$$\delta\left(x^\mu - y^\mu\right) = \delta(x^0 - y^0)\delta(\vec{x} - \vec{y}) = \frac{1}{c}\delta(t_x - t_y)\delta(\vec{x} - \vec{y}). \tag{7.94}$$

Notice that the $\delta$-function of a four-vector is a scalar. Under a Lorentz transformation,

$$\delta\left(U^\mu\right) \to \delta\left(U'^\mu\right) = \delta\left(\Lambda^\mu_{\ \nu} U^\nu\right) = \frac{\delta(U^\nu)}{|\det \Lambda|} = \delta(U^\nu). \tag{7.95}$$

With the use of the delta function, we can write the current-density four-vector of Eq. 7.93 as an integral

$$\begin{aligned} j^\mu(\vec{x},t) &= \sum_n q_n \int dt' \frac{dr_n^\mu(t')}{dt'} \delta(\vec{x} - \vec{r}_n(t'))\delta(t' - t) \\ &= c \sum_n q_n \int dt' \frac{dr_n^\mu}{dt'} \delta^{(4)}(x - r_n(t')), \end{aligned} \tag{7.96}$$

where

$$x^\mu \equiv (ct, \vec{x}), \quad r_n^\mu(t') \equiv (ct', \vec{r}_n(t')).$$

Now, we change the integration variable from $t'$ to $\tau$. Recall that

$$dt' = \frac{d\tau}{c}\gamma. \tag{7.97}$$

Then, we have

$$j^\mu(\vec{x},t) = c\sum_n q_n \int d\tau \frac{dr_n^\mu}{d\tau}\delta^{(4)}(x - r_n(\tau)), \tag{7.98}$$

which shows manifestly that $j^\mu$ transforms as $\frac{dx^\mu}{d\tau}$ and is therefore a contravariant four-vector. The integral form is obtained by inserting the unity $\int dt'\,\delta(t'-t) = 1$. This completes the three-dimensional $\delta$-function into a four-dimensional one at the price of an integration. Its only purpose is to show the four-vector character of $j^\mu$. In practical calculations we use Eq. 7.93: we add up the charges at the positions where they are at time $t$, multiplied by $dx^\mu/dt$.

We have already shown that charge conservation implies the continuity equation

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot\vec{j} = 0. \tag{7.99}$$

In relativistic notation, the continuity equation takes the more elegant form

$$\partial_\mu j^\mu = 0. \tag{7.100}$$

## 7.10 Energy-momentum tensor

Consider a collection of particles $\{n\}$ at positions $\vec{r}_n(t)$. The energy density is

$$\text{energy density} = \sum_n E_n(t)\delta(\vec{x} - \vec{r}_n(t)). \tag{7.101}$$

Changes in the energy density are associated with an "energy-current density",

$$\text{energy-current density} = \sum_n E_n(t)\frac{d\vec{r}_n}{dt}\delta(\vec{x} - \vec{r}_n(t)). \tag{7.102}$$

As with the charge density and its current density in the last section, we can combine the two into a single object,

$$\sum_n E_n(t)\frac{dr_n^\nu}{dt}\delta(\vec{x} - \vec{r}_n(t)). \tag{7.103}$$

Similarly to the energy, the densities and current densities of the components of the three-dimensional momentum are

$$\sum_n p_n^i(t)\frac{dr_n^\nu}{dt}\delta(\vec{x} - \vec{r}_n(t)), \quad i = 1,2,3. \tag{7.104}$$

Collectively, we can form the so-called "energy-momentum tensor", which encompasses the density and current density of all components of the four-momentum:

$$T^{\mu\nu} \equiv \sum_n p_n^\mu \frac{dr_n^\nu}{dt}\delta(\vec{x} - \vec{r}_n(t)), \tag{7.105}$$

or, in the equivalent form,

$$T^{\mu\nu} = c\sum_n \int d\tau p_n^\mu \frac{dr_n^\nu}{d\tau}\delta^{(4)}(x - r_n(\tau)). \tag{7.106}$$

From the last equation, we see manifestly that this object is justifiably called a tensor, since it transforms as the product of two four-vectors under Lorentz transformations:

$$T'^{\mu\nu} = \Lambda^{\mu}_{\ \rho}\Lambda^{\nu}_{\ \sigma}T^{\rho\sigma}. \tag{7.107}$$

Why do we need a tensor and not simply an energy density with its current? In special relativity, energy in one frame becomes partly momentum in another frame, so we must treat the densities and currents of all four components of $p^\mu$ together. The first index of $T^{\mu\nu}$ labels the component of the four-momentum and the second index labels density ($\nu = 0$) or current ($\nu = j$). Thus $T^{00}$ is the energy density, $c\,T^{0j}$ is the energy current density in the direction $j$, and $T^{i0}/c$ is the density of the momentum component $p^i$. The space components $T^{ij}$ measure how the momentum component $p^i$ flows in the direction $j$. They are most easily pictured in a fluid, where they contain e.g. the pressure. The energy-momentum tensor becomes indispensable when we replace particles by fields. It will tell us where the energy and momentum of the electromagnetic field are located and how they propagate.

The energy-momentum tensor is symmetric:

$$T^{\mu\nu} = T^{\nu\mu}. \tag{7.108}$$

To verify this, we recall that

$$p^{\nu}_n = m_n c\frac{dr^{\nu}_n}{d\tau} = m_n\gamma_n\frac{dr^{\nu}_n}{dt} = \frac{E_n}{c^2}\frac{dr^{\nu}_n}{dt}. \tag{7.109}$$

The energy-momentum tensor then takes the explicitly symmetric form

$$T^{\mu\nu} \equiv \sum_n c^2\frac{p^{\mu}_n p^{\nu}_n}{E_n}\delta(\vec{x} - \vec{r}_n(t)). \tag{7.110}$$

For the charge-density four-vector, we have found a continuity identity,

$$\partial_\mu j^\mu = 0.$$

This was a consequence of the conservation of charge. If the total energy and momentum of the system of particles are conserved, we anticipate a similar continuity identity for the energy-momentum tensor,

$$\partial_\nu T^{\mu\nu} = 0.$$

We have

$$\begin{aligned}
\partial_i T^{\mu i} &= \sum_n p^{\mu}_n\frac{dr^i_n}{dt}\frac{\partial}{\partial x^i}\delta\left(\vec{x} - \vec{r}_n\right) \\
&= -\sum_n p^{\mu}_n\frac{dr^i_n}{dt}\frac{\partial}{\partial r^i_n}\delta\left(\vec{x} - \vec{r}_n\right) \\
&= -\sum_n p^{\mu}_n\frac{\partial}{\partial t}\delta\left(\vec{x} - \vec{r}_n\right) \\
&= -\frac{\partial}{\partial t}\sum_n p^{\mu}_n\delta\left(\vec{x} - \vec{r}_n\right) + \sum_n\frac{\partial p^{\mu}_n}{\partial t}\delta\left(\vec{x} - \vec{r}_n\right) \\
&= -\partial_0 T^{\mu 0} + \sum_n\frac{\partial p^{\mu}_n}{\partial t}\delta\left(\vec{x} - \vec{r}_n\right).
\end{aligned} \tag{7.111}$$

In the above we have used that

$$\begin{aligned}\partial_x \delta(x-y) &= \frac{1}{2\pi}\partial_x \int_{-\infty}^{\infty} dw e^{-iw(x-y)} = -\frac{i}{2\pi}\int_{-\infty}^{\infty} dw\, w e^{-iw(x-y)} \\ &= -\frac{1}{2\pi}\partial_y \int_{-\infty}^{\infty} dw e^{-iw(x-y)} = -\partial_y \delta(x-y). \end{aligned} \tag{7.112}$$

We have therefore arrived at the equation

$$\partial_\nu T^{\mu\nu} = G^\mu, \tag{7.113}$$

where

$$G^\mu = \sum_n \frac{\partial p_n^\mu}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right) = \sum_n \frac{1}{c}\frac{d\tau}{dt} f_n^\mu(t)\delta\left(\vec{x}-\vec{r}_n\right) \tag{7.114}$$

is the "force density".

For free particles, where the energy and momentum of every particle are separately conserved, $p_n^\mu$ = constant, the energy-momentum tensor satisfies the continuity equation

$$\partial_\nu T^{\mu\nu} = 0. \tag{7.115}$$

The energy-momentum tensor is also conserved if the particles interact only at the points where they collide with each other. In that case, the force density is

$$\begin{aligned} G^\mu &= \sum_n \frac{\partial p_n^\mu}{\partial t}\delta\left(\vec{x}-\vec{r}_n\right) \\ &= \sum_{\text{coll.}} \delta(\vec{x}-\vec{x}_{\text{coll}}(t))\frac{d}{dt}\sum_{n\in\text{coll.}} p_n^\mu(t). \end{aligned} \tag{7.116}$$

We have grouped the sum over all particles contributing to the force density according to the collision points at which they meet. At each collision point, the sum of the momenta of the colliding particles is conserved:

$$\frac{d}{dt}\sum_{n\in\text{coll.}} p_n^\mu(t) = 0 \rightsquigarrow \partial_\nu T^{\mu\nu} = 0. \tag{7.117}$$

Point-like interactions are not the picture of classical electrodynamics, where a charge feels the fields of the others at a distance. However, this is how interactions are described in the quantum theory. An electron emits a photon at an interaction point and is deflected, with the momentum conserved there. The photon travels and is absorbed by a second electron at another interaction point, where the momentum is conserved again. If we count the photons, i.e. the electromagnetic field, among the particles, the energy-momentum tensor of the full system is conserved. In classical electrodynamics we will need to account explicitly for the energy and momentum carried by the field.

If the continuity equation is satisfied, then

$$\begin{aligned} 0 &= \partial_\nu T^{\mu\nu} \\ \rightsquigarrow 0 &= \partial_0 T^{\mu 0} + \partial_i T^{\mu i} \\ \rightsquigarrow 0 &= \partial_0 \int d^3\vec{x} T^{\mu 0} + \int d^3\vec{x}\partial_i T^{\mu i} \\ 0 &= \partial_0 \int d^3\vec{x} T^{\mu 0}, \end{aligned} \tag{7.118}$$

where the surface term from $\partial_i T^{\mu i}$ vanishes for a localised distribution of particles. This means that the vector

$$P^\mu \equiv \frac{1}{c}\int d^3\vec{x} T^{\mu 0} = \text{ constant} \tag{7.119}$$

is conserved. We find that the conserved vector is the total four-momentum of the particles in the distribution:

$$P^\mu = \frac{1}{c}\sum_n \int d^3\vec{x}\, p_n^\mu \frac{dr_n^0}{dt}\delta(\vec{x} - \vec{r}_n(t)) = \sum_n p_n^\mu. \tag{7.120}$$

# 8

# Relativistic formulation of electrodynamics

From now on we will set $\epsilon_0 = c = 1$. It will be easy to restore the full dependence on these constants with dimensional analysis when necessary. Maxwell's equations are

$$\vec{\nabla} \cdot \vec{E} = \rho, \tag{8.1}$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0, \tag{8.2}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{8.3}$$

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

*Remark.* With $\epsilon_0 = c = 1$ also $\mu_0 = 1/(\epsilon_0 c^2) = 1$, i.e. these are rationalised Heaviside–Lorentz units (no $4\pi$ in Maxwell's equations, Coulomb potential $q/(4\pi r)$). The metric is $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$ of Eq. 7.4; cf. Ref. [2], Appendix on units, and Sec. 11.9. We construct an "electromagnetic field tensor" $F^{\mu\nu}$ from the components of the electric $\vec{E} \equiv (E^1, E^2, E^3)$ and magnetic field $\vec{B} \equiv (B^1, B^2, B^3)$ as:

$$F^{0i} = -E^i, \quad F^{ij} = -\epsilon_{ijk} B^k, \quad F^{\mu\nu} = -F^{\nu\mu}. \tag{8.5}$$

(we use $\epsilon_{123} = +1$). Explicitly,

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{8.6}$$

Conversely, if the tensor $F^{\mu\nu}$ is given, we can obtain the magnetic field via (**exercise**):

$$B^i = -\frac{1}{2} \epsilon_{ijk} F^{jk}. \tag{8.7}$$

**Exercise:** Write in matrix form the dual tensor

$$\tilde{F}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}. \tag{8.8}$$

**Exercise:** Prove that:

$$-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = \frac{\vec{E}^2 - \vec{B}^2}{2}, \quad \epsilon_{\mu\nu\rho\sigma} F^{\mu\nu} F^{\rho\sigma} = 8 \vec{E} \cdot \vec{B}. \tag{8.9}$$

(Here $\epsilon_{0123} = +1$, hence $\epsilon^{0123} = -1$; with the opposite convention the right-hand side is $-8\,\vec{E}\cdot\vec{B}$. The dual tensor is often defined with an extra factor $\frac{1}{2}$, $\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$.)

As we have already noted, the charge and current densities form a four-vector:

$$j^\mu = (\rho, \vec{j}).$$

We observe that two of Maxwell's equations can be combined into one:

$$\partial_\nu F^{\nu\mu} = j^\mu. \tag{8.10}$$

Indeed, for $\mu = 0$, we have

$$\partial_0 F^{00} + \partial_i F^{i0} = j^0 \rightsquigarrow \vec{\nabla}\cdot\vec{E} = \rho.$$

For $\mu = i = 1,2,3$ we have

$$\partial_0 F^{0i} + \partial_j F^{ji} = j^i \rightsquigarrow -\partial_0 E^i + \epsilon_{ijk}\partial_j B^k = j^i \rightsquigarrow \vec{\nabla}\times\vec{B} - \frac{\partial\vec{E}}{\partial t} = \vec{j}.$$

The compact form makes the consistency of Maxwell's equations with special relativity evident by mere inspection. Under a Lorentz transformation the current transforms as a four-vector, $j'^\mu = \Lambda^\mu_{\ \sigma} j^\sigma$, the field strength as a second-rank contravariant tensor, $F'^{\nu\mu} = \Lambda^\nu_{\ \rho}\Lambda^\mu_{\ \sigma}F^{\rho\sigma}$, and the derivative with the inverse transformation, $\partial'_\nu = (\Lambda^{-1})^\kappa_{\ \nu}\partial_\kappa$. In $\partial'_\nu F'^{\nu\mu}$ the inverse matrix cancels one of the two $\Lambda$'s and we are left with $\Lambda^\mu_{\ \sigma}\partial_\rho F^{\rho\sigma} = \Lambda^\mu_{\ \sigma} j^\sigma = j'^\mu$: the equation has the same form in every inertial frame. Equations which looked rather complicated have become short, simply because we understood how their ingredients transform. One gets the feeling that special relativity could have been discovered from the structure of Maxwell's equations alone; indeed, it was the electrodynamics of moving bodies which led Einstein to it.

The remaining two of Maxwell's equations tell us, as we have found earlier, that we can derive the electric and magnetic fields by means of the scalar and vector potential $\phi, \vec{A}$, via

$$\vec{E} = -\vec{\nabla}\phi - \frac{\partial\vec{A}}{\partial t}, \quad \vec{B} = \vec{\nabla}\times\vec{A}. \tag{8.11}$$

We can combine the scalar and vector potential into a single four-vector:

$$A^\mu \equiv (\phi, \vec{A}) = (\phi, A^1, A^2, A^3). \tag{8.12}$$

Then, the above equations take the elegant form:

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{8.13}$$

Indeed, for $\mu = 0, \nu = i = 1,2,3$ we have

$$F^{0i} = \partial^0 A^i - \partial^i A^0 \rightsquigarrow -\vec{E} = \frac{\partial\vec{A}}{\partial t} + \vec{\nabla}\phi.$$

For $\mu = i, \nu = j,\ i,j = 1,2,3$ we have

$$F^{ij} = \partial^i A^j - \partial^j A^i \rightsquigarrow -\frac{1}{2}\epsilon_{ijk}F^{ij} = \epsilon_{ijk}\partial_i A^j \rightsquigarrow \vec{B} = \vec{\nabla}\times\vec{A}.$$

It is now straightforward to prove the following identities (**exercise**):

$$\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0, \tag{8.14}$$

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0. \tag{8.15}$$

Substituting Eq. 8.13 into Eq. 8.10 we find

$$\partial^2 A^\nu - \partial^\nu(\partial_\mu A^\mu) = j^\nu. \tag{8.16}$$

Explicitly, for $\nu = 0$ and $\nu = i = 1, 2, 3$ we recover the known differential equations for the scalar and vector potentials respectively:

$$\Box\phi - \frac{\partial}{\partial t}\left(\vec{\nabla}\cdot\vec{A} + \frac{\partial\phi}{\partial t}\right) = \rho, \tag{8.17}$$

$$\Box\vec{A} + \vec{\nabla}\left[\vec{\nabla}\cdot\vec{A} + \frac{\partial\phi}{\partial t}\right] = \vec{j}. \tag{8.18}$$

The property of gauge invariance becomes more elegant as well in relativistic notation. The gauge transformations of the vector and scalar potentials which leave Maxwell's equations invariant are written as:

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu\chi, \tag{8.19}$$

where $\chi$ is a scalar function. *Remark.* Compared with Eqs. 5.17–5.18 ($\phi \to \phi - \partial_t f$, $\vec{A} \to \vec{A} + \vec{\nabla} f$, i.e. $A^\mu \to A^\mu - \partial^\mu f$, $A_\mu \to A_\mu - \partial_\mu f$), this corresponds to $\chi = -f$; only the sign convention of the gauge function differs. Indeed, under a gauge transformation the electromagnetic field tensor remains invariant:

$$F_{\mu\nu} \to \partial_\mu(A_\nu + \partial_\nu\chi) - \partial_\nu(A_\mu + \partial_\mu\chi) = \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}. \tag{8.20}$$

The Lorentz gauge-fixing condition

$$\vec{\nabla}\cdot\vec{A} + \frac{\partial\phi}{\partial t} = 0 \tag{8.21}$$

can be written simply as

$$\partial_\mu A^\mu = 0. \tag{8.22}$$

In the Lorentz gauge, Maxwell's equations for the four-vector potential become:

$$\partial^2 A^\mu = j^\mu. \tag{8.23}$$

Earlier, we have found that the vector and scalar potential can be computed via

$$\phi(\vec{r}, t) = \int_{-\infty}^{\infty} d^3\vec{r}'dt' G\left(\vec{r} - \vec{r}'; t - t'\right)\rho(\vec{r}', t') \tag{8.24}$$

and

$$\vec{A}(\vec{r}, t) = \int_{-\infty}^{\infty} d^3\vec{r}'dt' G\left(\vec{r} - \vec{r}'; t - t'\right)\vec{j}(\vec{r}', t'), \tag{8.25}$$

where the Green's function can be written in the form:

$$G(\Delta\vec{r}, \Delta t) = \frac{1}{2\pi}\delta\left(\Delta t^2 - \Delta\vec{r}^2\right)\Theta(\Delta t > 0). \tag{8.26}$$

In relativistic notation, the solutions for the four-vector potential take the form:

$$A^\nu(x^\mu) = \frac{1}{2\pi}\int d^4x' j^\nu(x'^\mu)\delta\left((x'^\mu - x^\mu)^2\right)\Theta\left(x^0 > x'^0\right). \tag{8.27}$$

The electromagnetic force acting on a particle with a charge $q$ is:

$$f^\mu = qF^{\mu\nu}\frac{dx_\nu}{d\tau}. \tag{8.28}$$

Indeed, in the rest frame of the particle, where $d\tau = dt$ and $d\vec{x} = 0$, we get

$$f^\mu_{\text{rest}} = qF^{\mu 0} \rightsquigarrow f^0_{\text{rest}} = 0,\ \vec{f}_{\text{rest}} = q\vec{E}. \tag{8.29}$$

Since our expression for the electromagnetic force is correct in one reference frame, it should hold in every inertial reference frame since it is written as an equation between four-vectors. In a frame where the charge is moving with a velocity $\vec{v}$, the three-dimensional force is:

$$\begin{aligned} f^i &= qF^{i0}\frac{dx^0}{d\tau} - qF^{ij}\frac{dx^j}{d\tau} = q\gamma\left(E^i + \epsilon_{ijk}B^k v^j\right) \\ \rightsquigarrow \vec{f} &= q\gamma\left(\vec{E} + \vec{v}\times\vec{B}\right). \end{aligned} \tag{8.30}$$

Recall that

$$\vec{f} = \frac{d\vec{p}}{d\tau} = \gamma\frac{d\vec{p}}{dt}. \tag{8.31}$$

We therefore have recovered our familiar expression for the Lorentz force:

$$\frac{d\vec{p}}{dt} = q\left(\vec{E} + \vec{v}\times\vec{B}\right). \tag{8.32}$$

The time component of the force carries information too. With $F^{0i} = -E^i$ and $dx_i/d\tau = -\gamma v^i$ we find $f^0 = q\gamma\,\vec{E}\cdot\vec{v}$. Since $f^0 = dp^0/d\tau = \gamma\,dE/dt$, this is

$$\frac{dE}{dt} = q\,\vec{E}\cdot\vec{v},$$

the rate at which the field does work on the charge. Only the electric field does work; the magnetic force is perpendicular to the velocity. In the rest frame of the charge $f^0$ vanishes, as we found above.

It is worth separating what, in this formulation, is a definition, what is a derivation and what is a physical law. We postulate a four-potential $A^\mu$ which transforms as a contravariant four-vector. That $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is a second-rank contravariant tensor is then a mathematical consequence, since we know how $\partial^\mu$ transforms. That $\partial_\nu F^{\nu\mu} = j^\mu$ are the field equations, and that the field acts on a charge with the force $f^\mu = qF^{\mu\nu}dx_\nu/d\tau$, are physical laws: they are not derived, and they are justified by experiment. The form of the force is nevertheless strongly suggested by relativity. We need a four-vector (one free upper index), linear in the field and proportional to the charge, built out of $F^{\mu\nu}$ and the motion of the particle. The natural candidate contracts the second index of $F^{\mu\nu}$ with the *covariant* four-velocity $dx_\nu/d\tau$: the transformation of the upper index $\nu$ is then cancelled by the inverse transformation of the lower one, and only $\mu$ transforms.

## 8.1 Energy-momentum tensor in the presence of an electromagnetic field

Consider a number of charges $q_n$ which interact via the electromagnetic field. The energy-momentum tensor $T^{\mu\nu}$ of the charges is not conserved:

$$\partial_\nu T^{\mu\nu} = G^\mu, \tag{8.33}$$

where the force density is given by

$$\begin{aligned} G^\mu &= \sum_n \frac{\partial\tau}{\partial t} f_n^\mu(t)\delta\left(\vec{x}-\vec{r}_n\right) \\ &= \sum_n \frac{\partial\tau}{\partial t} q_n F^\mu{}_\nu \frac{dr_n^\nu}{d\tau}\delta\left(\vec{x}-\vec{r}_n\right) \\ &= F^\mu{}_\nu \sum_n q_n \frac{dr_n^\nu}{dt}\delta\left(\vec{x}-\vec{r}_n\right) \\ &= F^{\mu\nu} j_\nu. \end{aligned} \tag{8.34}$$

Consider the tensor

$$T_{em}^{\mu\nu} \equiv F^\mu{}_\rho F^{\rho\nu} + \frac{1}{4} g^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}. \tag{8.35}$$

This tensor is explicitly

- symmetric,
- gauge invariant.

The components of the tensor are (**exercise**):

$$T_{em}^{00} = \frac{\vec{E}^2 + \vec{B}^2}{2}, \quad T_{em}^{0i} = T_{em}^{i0} = (\vec{E}\times\vec{B})_i. \tag{8.36}$$

**Exercise:** Find the remaining components.

We find

$$\begin{aligned} \partial_\nu T_{em}^{\mu\nu} &= \partial_\nu \left\{ F^{\mu\rho} F_\rho{}^\nu + \frac{g^{\mu\nu}}{4} F_{\rho\sigma} F^{\rho\sigma} \right\} \\ &= F^{\mu\rho}\partial_\nu F_\rho{}^\nu + \left(\partial^\nu F^{\mu\rho}\right) F_{\rho\nu} + \frac{1}{2} F_{\rho\sigma}\partial^\mu F^{\rho\sigma} \\ &= -F^{\mu\rho} j_\rho + \frac{1}{2}\left(\partial^\nu F^{\mu\rho} - \partial^\rho F^{\mu\nu}\right) F_{\rho\nu} + \frac{1}{2} F_{\rho\sigma}\partial^\mu F^{\rho\sigma} \\ &= -F^{\mu\rho} j_\rho + \frac{1}{2}\left(\partial^\sigma F^{\mu\rho} + \partial^\mu F^{\rho\sigma} + \partial^\rho F^{\sigma\mu}\right) F_{\rho\sigma}, \end{aligned} \tag{8.37}$$

which, due to Eq. 8.15, yields:

$$\partial_\nu T_{em}^{\mu\nu} = -F^{\mu\nu} j_\nu. \tag{8.38}$$

$T_{em}^{\mu\nu}$ is purely electromagnetic. Neither $T^{\mu\nu}$ nor $T_{em}^{\mu\nu}$ satisfies a continuity equation on its own, but their sum

$$\Theta^{\mu\nu} \equiv T^{\mu\nu} + T_{em}^{\mu\nu} = \sum_n \frac{p_n^\mu p_n^\nu}{E_n}\delta(\vec{x}-\vec{r}_n(t)) + F^{\mu\rho} F_\rho{}^\nu + \frac{g^{\mu\nu}}{4} F_{\rho\sigma} F^{\rho\sigma} \tag{8.39}$$

does:

$$\partial_\nu \Theta^{\mu\nu} = 0. \tag{8.40}$$

From the continuity equation, we obtain a conserved four-vector:

$$\partial_t P^\mu \equiv \partial_t \int d^3\vec{x}\Theta^{\mu 0} = -\int d^3\vec{x}\partial_i \Theta^{\mu i} = -\left.\Theta^{\mu i}\right|_\infty = 0. \tag{8.41}$$

The conserved vector is:

$$P^\mu = \int d^3\vec{x}\Theta^{\mu 0} = \sum_n p_n^\mu + \int d^3\vec{x} T_{em}^{\mu 0}. \tag{8.42}$$

The four-momentum of the charges

$$P^\mu_{\text{charges}} = \sum_n p_n^\mu \tag{8.43}$$

is not conserved on its own. Some momentum

$$P^\mu_{em} = \int d^3\vec{x} T_{em}^{\mu 0} \tag{8.44}$$

is also carried by the electromagnetic field itself. This is not conserved either. Momentum can be exchanged between the charges and the field; however, this happens in such a way that the total momentum is always the same:

$$P^\mu = P^\mu_{\text{charges}} + P^\mu_{em} = \text{constant}. \tag{8.45}$$

The time component of the four-vector is the total energy. The energy stored in the electromagnetic field is:

$$E_{em} = \int d^3\vec{x} T_{em}^{00}. \tag{8.46}$$

Therefore, the energy density $w$ of the electromagnetic field is:

$$w = T_{em}^{00} = \frac{\vec{E}^2 + \vec{B}^2}{2}, \quad \text{EM field energy density.} \tag{8.47}$$

This justifies the interpretation of $T_{em}^{00}$ as an energy density: it is positive definite, and for static fields ($\vec{B} = 0$) it reduces to $\vec{E}^2/2$, the energy density which we found in electrostatics by computing the energy of a charge distribution. Notice that $T_{em}^{\mu\nu}$ depends only on $\vec{E}$ and $\vec{B}$ and not on any charges: an electromagnetic wave far away from all charges carries energy and momentum. When the field is quantised, this energy and momentum are carried by the photons.

Similarly, we find that the three-momentum density $\vec{S}$ of the electromagnetic field is:

$$\vec{S} = T_{em}^{0i} = T_{em}^{i0} = (\vec{E} \times \vec{B})^i, \quad \text{EM field momentum density.} \tag{8.48}$$

The vector

$$\vec{S} \equiv \vec{E} \times \vec{B}, \tag{8.49}$$

is known as the *Poynting vector*. *Remark.* $T_{em}^{0i}$ plays two roles: $T_{em}^{i0}$ is the energy-flux density (Poynting vector, entering $\partial_t w + \vec{\nabla} \cdot \vec{S} = \ldots$ below) and $T_{em}^{0i}$ is the momentum density. They

coincide only because $c = 1$; in SI units $\vec{S} = \vec{E} \times \vec{B}/\mu_0$ (energy flux) while the momentum density is $\vec{g} = \epsilon_0 \vec{E} \times \vec{B} = \vec{S}/c^2$ (see Ref. [2], Sec. 6.7).

Setting $\mu = 0$ in Eq. 8.38 we have

$$\partial_0 T^{00}_{em} + \partial_i T^{i0}_{em} = -F^{0i} j_i, \tag{8.50}$$

and equivalently,

$$\frac{\partial w}{\partial t} + \vec{\nabla} \cdot \vec{S} = -\vec{E} \cdot \vec{j}. \tag{8.51}$$

**Exercise:** Forget how we derived this equation. Insert $w = (\vec{E}^2 + \vec{B}^2)/2$ and $\vec{S} = \vec{E} \times \vec{B}$ into its left-hand side and use Maxwell's equations in their three-dimensional form to verify it. (This is how it is usually derived in textbooks.) The electromagnetic energy in a region decreases either because it flows out through the boundary ($\vec{\nabla} \cdot \vec{S}$) or because the field does work on the charges ($\vec{E} \cdot \vec{j}$).

## 8.2 Lagrangian formalism of electrodynamics

In classical mechanics, we can obtain equations of motion by requiring that an action remains stationary under small variations of the physical degrees of freedom around their physical values. This is a convenient way to cast the laws of physics, since the action is a scalar quantity and is guaranteed to be the same in all reference frames. We have

$$\delta S = 0, \tag{8.52}$$

where for a system with a finite number of degrees of freedom,

$$S = \int L dt, \tag{8.53}$$

and the Lagrangian is

$$L = \sum_i L_i, \tag{8.54}$$

where $L_i$ corresponds (for typical systems) to the difference of the kinetic and potential energy.

For the electromagnetic field, energy is distributed continuously in space. We should then introduce a *Lagrangian density* $\mathcal{L}$ where

$$L_i \to d^3\vec{x}\mathcal{L}. \tag{8.55}$$

The action becomes a four-dimensional integral over the Lagrangian density:

$$S = \int dt \int d^3\vec{x}\mathcal{L} = \int d^4x \mathcal{L}. \tag{8.56}$$

The principle of stationary action reads:

$$\delta \int d^4x \mathcal{L} = 0. \tag{8.57}$$

## 8.3 Maxwell's equations as Euler–Lagrange equations

For the electromagnetic field, we assume a Lagrangian density which is a function of the four-vector potential $A^\mu$ and its first derivatives[1]:

$$\mathcal{L} = \mathcal{L}(A^\mu, \partial^\nu A^\mu). \tag{8.58}$$

Then

$$\delta S = 0 \rightsquigarrow \int d^4x \left\{ \frac{\partial \mathcal{L}}{\partial A^\mu} \delta A^\mu(x) + \frac{\partial \mathcal{L}}{\partial(\partial^\nu A^\mu)} \delta \partial^\nu A^\mu(x) \right\} = 0. \tag{8.59}$$

Recall that

$$\frac{\delta f(x)}{\delta f(y)} = \delta(x-y),, \tag{8.60}$$

and

$$\frac{\delta G(f(x))}{\delta f(y)} = \frac{\partial G}{\partial f(x)} \delta(x-y). \tag{8.61}$$

Thus,

$$\begin{aligned} 0 &= \delta S \\ \rightsquigarrow 0 &= \int d^4x \left[ \frac{\partial \mathcal{L}}{\partial A^\mu} \delta A^\mu(x) + \frac{\partial \mathcal{L}}{\partial(\partial^\nu A^\mu)} \delta(\partial^\nu A^\mu(x)) \right] \\ \rightsquigarrow 0 &= \int d^4x \left[ \frac{\partial \mathcal{L}}{\partial A^\mu} \delta A^\mu(x) + \frac{\partial \mathcal{L}}{\partial(\partial^\nu A^\mu)} \partial^\nu(\delta A^\mu(x)) \right]. \end{aligned} \tag{8.62}$$

Performing integration by parts and dropping the total derivative term, we obtain:

$$\int d^4x \left[ \frac{\partial \mathcal{L}}{\partial A^\mu} - \partial^\nu \frac{\partial \mathcal{L}}{\partial(\partial^\nu A^\mu)} \right] \delta A^\mu(x) = 0. \tag{8.63}$$

The above must be satisfied for arbitrary small variations $\delta A^\mu(x)$. This requires

$$\frac{\partial \mathcal{L}}{\partial A^\mu} - \partial^\nu \frac{\partial \mathcal{L}}{\partial(\partial^\nu A^\mu)} = 0. \tag{8.64}$$

Consider now the Lagrangian:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J_\nu A^\nu. \tag{8.65}$$

The Euler–Lagrange equations, Eq. 8.64, give (**exercise**) Maxwell's equations:

$$\partial_\mu F^{\mu\nu} = J^\nu. \tag{8.66}$$

**Mass as frozen potential energy.** The Lagrangian formalism lets us see how a field can acquire a mass. The mass of a particle is read off from the propagator, the amplitude (or Green's function) for propagating from $x$ to $y$: it is the position of its pole in momentum space. For the photon the denominator is $k^2$ (see Chapter 12), with a pole at $k^2 = 0$: the photon is massless. Consider now photons inside a superconductor. There the electrons

[1] We do not need a Lagrangian density which depends on second or higher derivatives, since Maxwell's equations are second-order differential equations. From our experience with classical mechanics, we know that higher derivatives in $\mathcal{L}$ lead to equations of motion of third or higher order.

bind into Cooper pairs, which we can describe by a complex field $\phi$ of charge $q = -2e$, with the Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) - V(\phi), \qquad D_\mu = \partial_\mu + iqA_\mu.$$

The potential $V$ has its minimum at a value $|\phi| = \phi_0 \neq 0$, so that in the ground state the field is frozen at the constant value $\phi_0$ everywhere in the material. Setting $\phi = \phi_0$, the term $(D_\mu\phi)^*(D^\mu\phi)$ becomes $q^2\phi_0^2 A_\mu A^\mu \equiv \frac{m^2}{2}A_\mu A^\mu$ and the Euler–Lagrange equations (**exercise**) become

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = J^\nu.$$

Taking the divergence of both sides, current conservation now forces $\partial_\mu A^\mu = 0$, and the equations take the form $(\partial^2 + m^2)A^\nu = J^\nu$. The pole has moved to $k^2 = m^2$: inside the superconductor the photon behaves as a particle of mass $m$, and a static magnetic field is screened over a distance $1/m$ (Meissner effect). What looks like a mass is the energy of the interaction of the photon with a field which is frozen everywhere at the minimum of its potential. In particle physics this is the Higgs mechanism: we believe that the masses of the elementary particles originate from their interaction with the Higgs field, whose potential energy is minimal at a non-zero value which fills all of space-time.

### 8.3.1 Gauge fixing

We can include a gauge choice within the Lagrangian formalism for the electromagnetic field. Consider the Lagrangian density

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J_\nu A^\nu - \frac{\xi}{2}(\partial_\mu A^\mu)^2. \tag{8.67}$$

The Euler–Lagrange equations yield:

$$\partial^2 A^\mu - (1-\xi)\partial^\mu(\partial\cdot A) = J^\mu. \tag{8.68}$$

Choosing $\xi = 1$ produces Maxwell's equations in the Lorentz gauge. *Remark.* Strictly, $\xi = 1$ gives the wave equation $\partial^2 A^\mu = J^\mu$ without imposing $\partial\cdot A = 0$. Taking the divergence and using $\partial_\mu J^\mu = 0$ gives $\partial^2(\partial\cdot A) = 0$, so the Lorentz condition holds at all times only if $\partial\cdot A$ and $\partial_t(\partial\cdot A)$ vanish initially. In the usual normalisation $-\frac{1}{2\xi}(\partial\cdot A)^2$ (Ref. [5], Sec. 9.4) the case $\xi = 1$ is the Feynman gauge.

If we promote $\xi$ to a field, the corresponding Euler–Lagrange equation yields:

$$\frac{\partial\mathcal{L}}{\partial\xi} - \partial^\nu\frac{\partial\mathcal{L}}{\partial(\partial^\nu\xi)} = 0 \rightsquigarrow (\partial\cdot A)^2 = 0 \rightsquigarrow \partial\cdot A = 0, \tag{8.69}$$

which is the Lorentz gauge-fixing condition.

# 9

# Radiation from moving charges

In this section we will study the electromagnetic field which is created by moving charges.

By radiation we mean energy which is emitted outwards from the charges and which can reach arbitrarily large distances. This is different from what we know from electrostatics, where the effect of a charge distribution diminishes as $1/r^2$ as we move away from it. A charge at rest does not radiate, and neither does a charge moving with a constant velocity, which is just a charge at rest seen from another inertial frame. To radiate, a charge must be accelerated or decelerated, as, for example, the charges which are driven up and down in an antenna. Our plan is: take a charge on an arbitrary trajectory $\vec{r}(t)$, compute the four-potential which it creates, then the field tensor $F^{\mu\nu}$, and finally the Poynting vector, which tells us how much energy flows to large distances. For many charges we use the superposition principle.

## 9.1 The vector potential from a moving charge

A moving charge $q$ has a current density

$$j^\mu(x) = \left(q\delta(\vec{x} - \vec{r}(t)), q\frac{d\vec{r}}{dt}\delta(\vec{x} - \vec{r}(t))\right). \tag{9.1}$$

In an explicitly covariant form, we have found:

$$j^\mu = q\int d\tau v^\mu \delta^4\left(x - r(\tau)\right), \tag{9.2}$$

where $\tau$ is the proper time and

$$v^\mu \equiv \frac{dr^\mu}{d\tau}, \tag{9.3}$$

the four-velocity of the charge.

The solution of Maxwell's equations for the vector potential at a position $x^\mu$ is:

$$A^\mu(x) = \int d^4x' G_{\text{ret}}(x - x')j^\mu(x'), \tag{9.4}$$

with the "retarded" Green's function given by

$$G_{\text{ret}}(x - x') = \frac{1}{2\pi}\delta((x - x')^2)\Theta(x^0 > x'^0). \tag{9.5}$$

Substituting the expression of the Green's function into the potential integral solution and performing the $d^4x'$ integration, we obtain:

$$A^\mu(x) = \frac{q}{2\pi}\int d\tau v^\mu(\tau)\delta((x-r(\tau))^2)\Theta(x^0 > r^0(\tau)). \tag{9.6}$$

To perform the $\tau$ integration, we need to solve the constraint imposed by the delta function:

$$f(\tau) = (x-r(\tau))^2 = 0, \tag{9.7}$$

with the additional constraint

$$x^0 > r^0(\tau). \tag{9.8}$$

We have

$$\begin{aligned} 0 &= (x-r(\tau))^2 = (x^0-r^0(\tau))^2 - \left|\vec{x}-\vec{r}(\tau)\right|^2 \\ \rightsquigarrow x^0 - r^0(\tau) &= \pm\left|\vec{x}-\vec{r}(\tau)\right| = \pm R(\tau), \end{aligned} \tag{9.9}$$

where $R(\tau)$ is the distance of the charge from the observation point and it is a positive quantity. Due to the constraint $x^0 > r^0(\tau)$, we keep only the solution originating from:

$$x^0 = r^0(\tau) + \left|\vec{x}-\vec{r}(\tau)\right|. \tag{9.10}$$

Physically, the charge influences the potential at $(\vec{x}, x^0)$ only from the position at which it was at the time $r^0$ such that a light signal leaving it has exactly the time $x^0 - r^0$ needed to cover the distance $|\vec{x}-\vec{r}|$; where the charge is *now* is irrelevant. As a thought experiment (ignoring charge conservation), imagine a charge which suddenly appears at rest at the origin at $t = 0$, $\rho = q\,\delta(\vec{x})\Theta(t)$. The retarded solution $\phi(\vec{x},t) = \int d^3\vec{x}'\,\rho(\vec{x}', t-|\vec{x}-\vec{x}'|)/(4\pi|\vec{x}-\vec{x}'|)$ gives, at a distance $r$,

$$\phi(\vec{x},t) = \frac{q}{4\pi r}\,\Theta(t-r)\,:$$

nothing until the light signal emitted at the birth of the charge arrives, and the Coulomb potential from then on.

Once we know the roots $\tau_i$ of Eq. 9.7, $f(\tau_i) = 0$, which satisfy Eq. 9.8, we can use that

$$\delta\left[f(\tau)\right] = \sum_i \frac{\delta(\tau-\tau_i)}{\left|\frac{df}{d\tau}\right|_{\tau=\tau_i}}. \tag{9.11}$$

For particles with velocity $v < c = 1$ we find only one such solution $\tau_0$. Later, we will demonstrate how this is done geometrically for a charge moving with a constant velocity. Let us now proceed to perform the $\tau$ integration assuming that this solution $\tau_0$ has been found.

The derivative is

$$\begin{aligned} \frac{df(\tau)}{d\tau} &= \frac{d}{d\tau}(x-r(\tau))^2 = \frac{d}{d\tau}\left[g_{\mu\nu}(x-r(\tau))^\mu(x-r(\tau))^\nu\right] \\ &= 2g_{\mu\nu}(x-r(\tau))^\mu\frac{d}{d\tau}(x-r(\tau))^\nu = -2g_{\mu\nu}(x-r(\tau))^\mu\frac{dr^\nu(\tau)}{d\tau} \\ &= -2g_{\mu\nu}(x-r(\tau))^\mu v^\nu(\tau) = -2v\cdot(x-r(\tau)). \end{aligned} \tag{9.12}$$

The vector potential becomes:

$$A^\mu(x) = \frac{q}{4\pi}\int d\tau\frac{\delta(\tau-\tau_0)v^\mu(\tau)}{|(x-r(\tau))\cdot v(\tau)|}. \tag{9.13}$$

As can be seen in the rest frame of the charge, the quantity $(x - r(\tau)) \cdot v(\tau)$ is positive definite and we can therefore drop the absolute value from the denominator. (By "rest frame" we mean the inertial frame in which the charge is momentarily at rest at the retarded time $\tau_0$. An accelerated charge does not stay at rest in it, but this does not matter: a scalar product needs to be evaluated in one inertial frame only. There $v^\mu = (1, \vec{0})$ and $(x - r) \cdot v = x^0 - r^0 = |\vec{x} - \vec{r}| > 0$.) Performing the $\tau$ integration is trivial, yielding

$$A^\mu(x) = \frac{q}{4\pi} \left. \frac{v^\mu}{v \cdot (x - r(\tau))} \right|_{\text{retarded}} . \tag{9.14}$$

The subscript denotes that all quantities in this expression (distance and velocity of the charge) must be computed at a retarded proper time $\tau = \tau_0$ and not at the current time of the measurement. The expression of Eq. 9.14 is known as the Liénard–Wiechert potential. Notice that for a charge at rest, $v^\mu = (1, \vec{0})$, it reproduces our known Coulomb potential:

$$A^\mu(x) = \frac{q}{4\pi} \frac{(1, \vec{0})}{x^0 - r^0} = \frac{q}{4\pi} \frac{(1, \vec{0})}{|\vec{x} - \vec{r}|} . \tag{9.15}$$

### 9.1.1 Potential from a charge moving with constant velocity

Let us now see how we can use Eq. 9.14 in order to recover the results of Section 6.4 for the scalar and vector potential due to a charged particle moving with a constant velocity $\vec{v}$. At a time $t$, we assume the particle to be at a position $P$ which is at a distance

$$(PS) = |\vec{v}|t ,$$

from the origin $S$. We are interested in the potential $A^\mu(\vec{x}, t)$ at a position $O(\vec{x})$ at the time $t$ (see Fig. 9.1). The electromagnetic wave which is measured at $(\vec{x}, t)$ has been emitted at

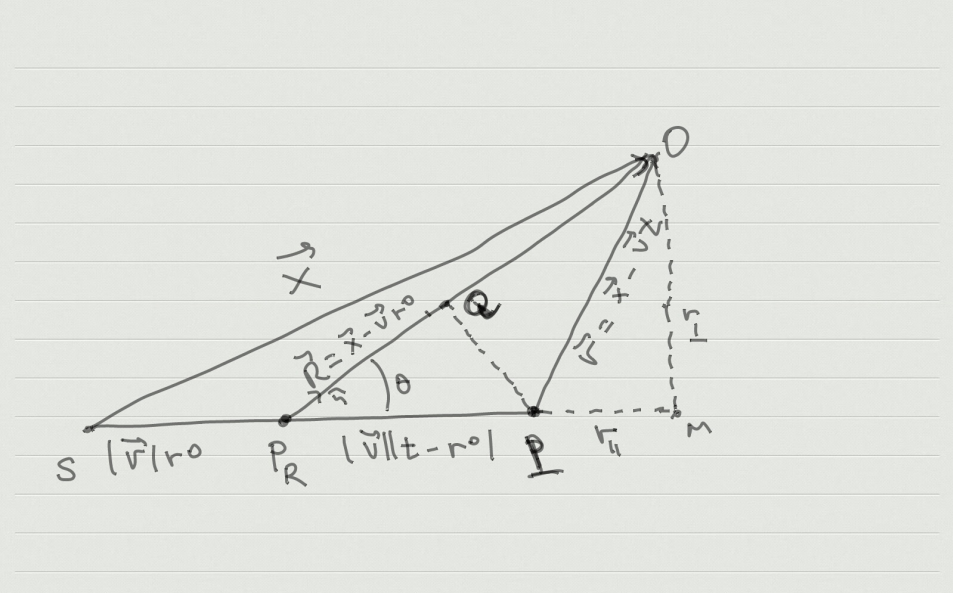



**Figure 9.1.** A charge moving with constant velocity $\vec{v}$ is at the position $P$ at the time $t$ of the observation at $O$. The signal measured at $O$ was emitted from the retarded position $P_R$ at the time $r^0$. $Q$ is the projection of $P$ on the line $P_R O$ and $M$ the projection of $O$ on the trajectory of the charge; $S$ is the origin.

an earlier time $r^0$ when the charge was at a retarded position $P_R(\vec{v}r^0)$. The time needed for the signal to travel is

$$t - r^0, \tag{9.16}$$

and (in units of $c = 1$) the distance covered is

$$(P_R O) = R = \left|\vec{x} - \vec{v}r^0\right| = t - r^0. \tag{9.17}$$

In the same time, the charge has advanced by a distance:

$$(P_R P) = |\vec{v}|(t - r^0) = |\vec{v}|R. \tag{9.18}$$

Let $Q$ be the projection of the position $P$ of the charge on the trajectory of the light signal. Then,

$$(P_R Q) = |\vec{v}|R\cos\theta = R\vec{v}\cdot\hat{n}, \tag{9.19}$$

where $\hat{n}$ points to the direction of travel of the light signal. Then,

$$(QO) = R(1 - \vec{v}\cdot\hat{n}). \tag{9.20}$$

Now consider the projection $M$ of the observation point $O$ on the trajectory line of the charge. From Pythagoras' theorem we have:

$$(PQ)^2 + (QO)^2 = (PM)^2 + (MO)^2 = r^2. \tag{9.21}$$

The distance of the observation point from the trajectory of the charge is:

$$(MO) \equiv r_\perp, \tag{9.22}$$

and the parallel distance is

$$(MP) \equiv r_\parallel = x_\parallel - vt. \tag{9.23}$$

From Eq. 9.19 and a bit of trigonometry, we obtain (**exercise**)

$$R(1 - \vec{n}\cdot\vec{v}) = \sqrt{1 - v^2}\left[\left(\frac{x_\parallel - vt}{\sqrt{1 - v^2}}\right)^2 + r_\perp^2\right]^{\frac{1}{2}}. \tag{9.24}$$

On the left-hand side, we recognise the scalar product:

$$(x - r)\cdot v|_{\text{retarded}} = R(1, \hat{n})\cdot\gamma(1, \vec{v}) = \gamma R(1 - \vec{v}\cdot\hat{n}). \tag{9.25}$$

Thus, the potential is:

$$A^\mu = \frac{q}{4\pi}\frac{v^\mu}{(x - r)\cdot v} = \frac{q}{4\pi}\frac{\gamma(1, \vec{v})}{\gamma R(1 - \vec{v}\cdot\hat{n})} = \frac{q}{4\pi}\frac{\gamma(1, \vec{v})}{\left[\left(\frac{x_\parallel - vt}{\sqrt{1 - v^2}}\right)^2 + r_\perp^2\right]^{\frac{1}{2}}}. \tag{9.26}$$

This is in agreement with the result found in Section 6.4.

## 9.2 The electromagnetic field tensor from a moving charge

The electromagnetic field tensor

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu, \tag{9.27}$$

for a moving charge can be derived by differentiating the expression for the Liénard–Wiechert potential. Because the terms of Eq. 9.14 must be evaluated at retarded positions, it is easier to work with the integral of Eq. 9.6 and perform the integration after we have carried out the differentiations. We define the four-vector:

$$R^\mu \equiv x^\mu - r^\mu(\tau). \tag{9.28}$$

The potential is then:

$$A^\nu(x) = \frac{q}{2\pi}\int d\tau v^\nu(\tau)\delta(R^2)\Theta(R^0 > 0). \tag{9.29}$$

Its derivative is

$$\begin{aligned}
\partial^\mu A^\nu &= \frac{q}{2\pi}\int d\tau v^\nu(\tau)\partial^\mu\left[\delta(R^2)\Theta(R^0>0)\right] \\
&= \frac{q}{2\pi}\int d\tau v^\nu(\tau)\left[\Theta(R^0>0)\partial^\mu\delta(R^2) + \delta(R^2)\partial^\mu\Theta(R^0>0)\right] \\
&= \frac{q}{2\pi}\int d\tau v^\nu(\tau)\left[\Theta(R^0>0)\partial^\mu\delta(R^2) + \delta(R^2)\delta(R^0)g^{\mu 0}\right].
\end{aligned} \tag{9.30}$$

The second term in the bracket vanishes,

$$\delta(R^2)\delta(R^0) = \delta((R^0)^2 - \vec{R}^2)\delta(R^0) = \delta(-\vec{R}^2)\delta(R^0) = 0, \tag{9.31}$$

unless we are interested in calculating the electromagnetic field exactly at a space-time point occupied by the electric charge $q$. For all other points, where in fact classical electrodynamics should be valid, we have

$$\partial^\mu A^\nu = \frac{q}{2\pi}\int d\tau v^\nu(\tau)\Theta(R^0>0)\partial^\mu\delta(R^2). \tag{9.32}$$

The derivative of the delta function is:

$$\begin{aligned}
\partial^\mu\delta(F) &= \partial^\mu\delta(R^2) = \partial^\mu F\frac{\partial\delta(F)}{\partial F} \\
&= \frac{\partial^\mu F}{\frac{\partial F}{\partial\tau}}\frac{\partial\delta(F)}{\partial\tau} = \frac{\partial^\mu R^2}{\frac{\partial R^2}{\partial\tau}}\frac{\partial\delta(R^2)}{\partial\tau} \\
&= \frac{2R^\mu}{2R\cdot\frac{\partial R}{\partial\tau}}\frac{\partial\delta(R^2)}{\partial\tau} = \frac{-R^\mu}{R\cdot v(\tau)}\frac{\partial\delta(R^2)}{\partial\tau}.
\end{aligned} \tag{9.33}$$

Thus, the derivative of the potential becomes:

$$\partial^\mu A^\nu = -\frac{q}{2\pi}\int d\tau\Theta(R^0>0)\frac{R^\mu v^\nu}{R\cdot v}\frac{\partial\delta(R^2)}{\partial\tau}. \tag{9.34}$$

The field tensor is then

$$\begin{aligned} F^{\mu\nu} &= \partial^\mu A^\nu - \partial^\nu A^\mu \\ &= -\frac{q}{2\pi}\int d\tau \Theta(R^0>0)\frac{R^\mu v^\nu - R^\nu v^\mu}{R\cdot v}\frac{\partial\delta(R^2)}{\partial\tau} \\ &= -\frac{q}{2\pi}\int d\tau \frac{\partial}{\partial\tau}\left[\Theta(R^0>0)\frac{R^\mu v^\nu - R^\nu v^\mu}{R\cdot v}\delta(R^2)\right] \\ &\quad +\frac{q}{2\pi}\int d\tau \delta(R^2)\frac{\partial}{\partial\tau}\left[\Theta(R^0>0)\frac{R^\mu v^\nu - R^\nu v^\mu}{R\cdot v}\right]. \end{aligned} \tag{9.35}$$

The first term requires the evaluation of the integrand at the boundaries of integration, and it is zero: the argument of $\delta(R^2)$ vanishes inside the integration region and not at the edges. Also, the term which is produced by differentiating $\Theta(R^0)$ vanishes for the same reason as we have seen earlier. Thus, the field tensor becomes:

$$F^{\mu\nu} = +\frac{q}{2\pi}\int d\tau \delta(R^2)\Theta(R^0>0)\frac{\partial}{\partial\tau}\left[\frac{R^\mu v^\nu - R^\nu v^\mu}{R\cdot v}\right]. \tag{9.36}$$

We can now perform the $\tau$ integration in exactly the same manner as in Section 9.1. We obtain:

$$F^{\mu\nu} = +\frac{q}{4\pi}\left.\frac{\frac{\partial}{\partial\tau}\left[\frac{R^\mu v^\nu - R^\nu v^\mu}{R\cdot v}\right]}{R\cdot v}\right|_{\text{retarded}}. \tag{9.37}$$

Performing the differentiations, and using that:

$$v_\mu v^\mu = \frac{dr_\mu}{d\tau}\frac{dr^\mu}{d\tau} = \frac{d\tau^2}{d\tau^2} = 1, \tag{9.38}$$

as well as the definition of the acceleration four-vector:

$$a^\mu \equiv \frac{dv^\mu}{d\tau}, \tag{9.39}$$

we find that

$$F^{\mu\nu} = \frac{q}{4\pi}\left.\frac{R^\mu Q^\nu - R^\nu Q^\mu}{(R\cdot v)^3}\right|_{\text{retarded}}, \tag{9.40}$$

where

$$Q^\nu = v^\nu + a^\nu R\cdot v - v^\nu R\cdot a. \tag{9.41}$$

The acceleration four-vector is:

$$\begin{aligned} a^\nu &= \frac{dv^\nu}{d\tau} = \frac{d}{d\tau}\gamma(1,\vec{v}) \\ &= \frac{d\gamma}{d\tau}(1,\vec{v}) + \gamma\left(0,\frac{d\vec{v}}{d\tau}\right) \\ &= \frac{1}{\gamma}\frac{d\gamma}{d\tau}v^\nu + a_T^\nu, \end{aligned} \tag{9.42}$$

with

$$a_T^\mu = \gamma\left(0,\frac{d\vec{v}}{d\tau}\right) = \gamma^2\left(0,\frac{d\vec{v}}{dt}\right), \tag{9.43}$$

and

$$\frac{d\vec{v}}{dt} \equiv \frac{d\vec{v}}{dt}. \tag{9.44}$$

*Question: Is $a_T^\mu$ a four-vector?*
Substituting into Eq. 9.41, we find that

$$Q^\nu = v^\nu + a_T^\nu R \cdot v - v^\nu R \cdot a_T. \tag{9.45}$$

### 9.2.1 The electric and magnetic fields of a moving charge

We can read the components of the electric and magnetic field from the electromagnetic field tensor of Eq. 9.40. We now write the distance vector $R^\mu$ as

$$R^\mu = |\vec{R}|(1, \hat{n}). \tag{9.46}$$

Then, the components of the electric field are

$$\begin{aligned} E^i = F^{i0} &= \frac{q}{4\pi} \frac{R^i Q^0 - R^0 Q^i}{(R \cdot v)^3}\bigg|_{\text{retarded}} \\ \rightsquigarrow \vec{E} &= \frac{q|\vec{R}|}{4\pi} \frac{Q^0 \hat{n} - \vec{Q}}{(R \cdot v)^3}\bigg|_{\text{retarded}}. \end{aligned} \tag{9.47}$$

The components of the magnetic field are

$$\begin{aligned} B^i = -\frac{1}{2}\epsilon_{ijk}F^{jk} &= -\frac{1}{2}\epsilon_{ijk}\frac{q}{4\pi}\frac{R^j Q^k - R^k Q^j}{(R \cdot v)^3}\bigg|_{\text{retarded}} \\ &= -\frac{q|\vec{R}|}{4\pi}\frac{\epsilon_{ijk}\hat{n}^j Q^k}{(R \cdot v)^3}\bigg|_{\text{retarded}} \\ \rightsquigarrow \vec{B} &= -\frac{q|\vec{R}|}{4\pi}\frac{\hat{n} \times \vec{Q}}{(R \cdot v)^3}\bigg|_{\text{retarded}}. \end{aligned} \tag{9.48}$$

We have therefore found explicitly that

$$\vec{B} = \hat{n} \times \vec{E}. \tag{9.49}$$

To compute the magnitude of the electric field, we need the products:

$$R \cdot v = |\vec{R}|\gamma(1 - \hat{n} \cdot \vec{v}), \tag{9.50}$$

and

$$R \cdot a_T = -|\vec{R}|\gamma^2 \frac{d\vec{v}}{dt} \cdot \hat{n}. \tag{9.51}$$

Finally, we arrive at the result (**exercise**):

$$\vec{E} = \frac{q}{4\pi(1 - \hat{n} \cdot \vec{v})^3}\left\{\frac{(1 - \vec{v}^2)}{|\vec{R}|^2}(\hat{n} - \vec{v}) + \frac{1}{|\vec{R}|}\hat{n} \times \left[(\hat{n} - \vec{v}) \times \frac{d\vec{v}}{dt}\right]\right\}\bigg|_{\text{retarded}}. \tag{9.52}$$

Unlike Coulomb's law, the field depends not only on the (retarded) position of the charge but also on its velocity and its acceleration. For $\vec{v} = \frac{d\vec{v}}{dt} = 0$ we recover the Coulomb field $q\hat{n}/(4\pi|\vec{R}|^2)$. The first term falls off as $1/|\vec{R}|^2$, as in electrostatics. The second term is new: it exists only if the charge is accelerated, and it falls off only as $1/|\vec{R}|$. It is responsible for radiation. The relation $\vec{B} = \hat{n} \times \vec{E}$ is the one we know from electromagnetic waves: here we see such waves being created by an accelerated charge.

## 9.3 Radiation from an accelerated charge in its rest frame: Larmor's formula

For a charge $q$ which is accelerated but momentarily at rest, the electric field is:

$$\vec{E} = \frac{q}{4\pi}\left\{\frac{\hat{n}}{|\vec{R}|^2} + \frac{1}{|\vec{R}|}\hat{n}\times\left[\hat{n}\times\frac{d\vec{v}}{dt}\right]\right\}\Bigg|_{\text{retarded}}. \tag{9.53}$$

A vanishing velocity together with a large acceleration is not an exotic situation. Think of a mass on a spring at its turning point: it stops, while the force on it is maximal. The charges in an antenna are driven up and down but cannot leave it; their velocities remain small while their accelerations can be very large.

To study the flow of energy (radiation) we compute the Poynting vector:

$$\vec{S} = \vec{E}\times\vec{B} = \vec{E}\times\left(\hat{n}\times\vec{E}\right) = \left|\vec{E}\right|^2\hat{n} - (\vec{E}\cdot\hat{n})\vec{E}. \tag{9.54}$$

Expanding in $1/|\vec{R}|$, we have:

$$\vec{S} = \hat{n}\frac{q^2}{16\pi^2|\vec{R}|^2}\left|\hat{n}\times(\hat{n}\times\frac{d\vec{v}}{dt})\right|^2 + O\left(\frac{1}{|\vec{R}|^3}\right). \tag{9.55}$$

Equivalently,

$$\vec{S} = \hat{n}\frac{q^2}{16\pi^2|\vec{R}|^2}|\frac{d\vec{v}}{dt}|^2\sin^2\Theta + O\left(\frac{1}{|\vec{R}|^3}\right), \tag{9.56}$$

where $\Theta$ is the angle between $\hat{n}$ and $\frac{d\vec{v}}{dt}$:

$$\frac{d\vec{v}}{dt}\cdot\hat{n} = |\frac{d\vec{v}}{dt}|\cos\Theta. \tag{9.57}$$

The power (energy per unit time) $dP$ emitted through a segment $d\vec{A}$ of a closed surface around the retarded position of the charge $q$ is given by

$$dP \equiv \frac{dW}{dt} = d\vec{A}\cdot\vec{S}. \tag{9.58}$$

For a segment of a sphere with radius $|\vec{R}|$ centred around the retarded position of the charge, we have:

$$d\vec{A} = \hat{n}|\vec{R}|^2 d\Omega, \tag{9.59}$$

and therefore

$$\frac{dP}{d\Omega} = \frac{q^2}{16\pi^2}|\frac{d\vec{v}}{dt}|^2\sin^2\Theta + O\left(\frac{1}{|\vec{R}|}\right). \tag{9.60}$$

In $\vec{S}\propto\vec{E}^2$, the square of the acceleration term gives a contribution $\propto 1/|\vec{R}|^2$, its product with the Coulomb term one $\propto 1/|\vec{R}|^3$, and the square of the Coulomb term one $\propto 1/|\vec{R}|^4$. The area of a sphere grows as $|\vec{R}|^2$. Therefore, only the acceleration term transports a non-vanishing power through a sphere of arbitrarily large radius. Once emitted, this energy keeps travelling; we receive light which was emitted billions of years ago at the other side of the universe.

The radiation power, i.e. the power which is radiated at infinitely large distances, per solid angle $d\Omega$ is:

$$\frac{dP_{\text{rad.}}}{d\Omega} = \frac{q^2}{16\pi^2}|\frac{d\vec{v}}{dt}|^2 \sin^2\Theta. \tag{9.61}$$

The total power radiated at all solid angles surrounding the retarded position of the charge is:

$$P_{\text{rad.}} = \int d\Omega \frac{dP_{\text{rad.}}}{d\Omega}. \tag{9.62}$$

Writing

$$d\Omega = \sin\Theta d\Theta d\phi, \tag{9.63}$$

and performing the angular integrations we obtain:

$$P_{\text{rad.}} = \frac{q^2}{4\pi}\frac{2}{3}|\frac{d\vec{v}}{dt}|^2. \tag{9.64}$$

Restoring the $\epsilon_0, c$ constants, we have:

$$P_{\text{rad.}} = \frac{q^2}{4\pi\epsilon_0}\frac{2}{3}\frac{|\frac{d\vec{v}}{dt}|^2}{c^3}. \tag{9.65}$$

This is known as Larmor's formula. It gives the total power of radiation of an accelerated charge when it has zero velocity. However, it can also be used as a good approximation for the radiation power emitted by accelerated charges with a small (non-relativistic) velocity $|\vec{v}| \ll c$.

Two features of the result are worth remembering. The radiated power grows quadratically with the acceleration (and with the charge). And radiation is not emitted uniformly: according to Eq. 9.61, nothing is emitted in the direction of the acceleration, and most is emitted perpendicular to it.

## 9.4 Radiation from an accelerated charge with a relativistic velocity

The power of radiation

$$\text{Power} = \frac{\text{Energy}}{\text{Time}}$$

is invariant under Lorentz transformations. *Remark.* Energy alone is not invariant; the argument is that in the instantaneous rest frame the radiated four-momentum is $dP^\mu_{\text{rad}} = (dW, \vec{0})$ (the dipole pattern is symmetric under $\hat{n} \to -\hat{n}$), and the corresponding time interval is $(dt, \vec{0})$, so $dW$ and $dt$ both transform as time components and $dW/dt$ is invariant (see Ref. [2], Sec. 14.2). The ratio refers to emission (retarded) time of the charge. If a charge $+q$ is accelerated and moving with a velocity $\vec{v}$, we can make a Lorentz transformation to its rest frame and use Larmor's formula. Alternatively, we can write Larmor's formula in a manifestly Lorentz-invariant form, expressing the three-dimensional vectors of the rest frame in terms of the corresponding four-vectors. Larmor's formula reads:

$$P_{\text{rad.}} = \frac{q^2}{4\pi}\frac{2}{3}|\frac{d\vec{v}}{dt}|^2 = \frac{q^2}{4\pi}\frac{2}{3m^2}\left(\frac{d\vec{p}}{dt}\right)^2. \tag{9.66}$$

The relativistic force in the rest frame is:

$$\frac{dp^\mu}{d\tau} = \left(0, \frac{d\vec{p}}{dt}\right), \tag{9.67}$$

and

$$\frac{dp_\mu}{d\tau}\frac{dp^\mu}{d\tau} = -\left(\frac{d\vec{p}}{dt}\right)^2. \tag{9.68}$$

We can then write:

$$P_{\text{rad.}} = -\frac{q^2}{4\pi}\frac{2}{3m^2}\frac{dp_\mu}{d\tau}\frac{dp^\mu}{d\tau}. \tag{9.69}$$

In a reference frame where the particle moves with velocity $\vec{v}$, the above expression gives:

$$P_{\text{rad.}} = \frac{q^2}{4\pi}\frac{2}{3}\gamma^6\left[|\frac{d\vec{v}}{dt}|^2 - \left|\vec{v}\times\frac{d\vec{v}}{dt}\right|^2\right]. \tag{9.70}$$

The way we obtained Eq. 9.69 is a trick which we will use again and again: compute a Lorentz-invariant quantity in the frame where it is simplest, write the result in terms of scalar products of four-vectors, and the result holds in every frame. Notice also that both terms in Eq. 9.70 are proportional to the acceleration (since $|\vec{v}\times\frac{d\vec{v}}{dt}| \leq |\vec{v}||\frac{d\vec{v}}{dt}| < |\frac{d\vec{v}}{dt}|$, the bracket vanishes only for $\frac{d\vec{v}}{dt} = 0$): no matter how fast a charge moves, it does not radiate unless its velocity changes.

### 9.4.1 Circular motion

Let us examine the case of a charge in circular motion. The acceleration in that case is perpendicular to the velocity of the charge:

$$\vec{v} \perp \frac{d\vec{v}}{dt}. \tag{9.71}$$

Eq. 9.70 then gives

$$P_{\text{rad.}} = \frac{q^2}{4\pi}\frac{2}{3}\gamma^4|\frac{d\vec{v}}{dt}|^2. \tag{9.72}$$

A charged particle in a circular orbit will radiate constantly. This observation tells us that a planetary-type model for atoms does not work. According to classical electrodynamics, electrons orbiting a nucleus would continuously lose energy by emitting radiation, falling to lower and lower orbits and eventually collapsing onto the nucleus. This does not happen because of quantum mechanics and the uncertainty principle. The electrons of an atom occupy discrete energy levels. In the ground state, which is the state of minimum allowed energy, they never radiate.

We also notice in Eq. 9.72 that the radiated power

$$P_{\text{rad.}} \sim \gamma^4$$

is larger for particles with larger velocity (larger $\gamma$).

In a circular collider, such as the Large Hadron Collider, one has to compensate for the radiation loss (synchrotron radiation) due to the acceleration of the charged particles which keeps them in a circular orbit.

The radiation loss depends strongly on the mass of the particle. For a given energy $E$ the Lorentz factor is $\gamma = E/(mc^2)$; at fixed energy and radius $\rho$ of the orbit ($|\frac{d\vec{v}}{dt}| = v^2/\rho \approx c^2/\rho$), Eq. 9.72 scales as $1/m^4$. An electron radiates $(m_p/m_e)^4 \approx 1.1\times10^{13}$ times more than a proton of the same energy. This is why the highest energies in circular colliders are reached with protons (LHC), while for electrons and positrons linear colliders are an alternative. Electron-positron collisions are attractive despite the cost: electrons are elementary particles without substructure, and the collisions are much cleaner than those of the composite protons.

### 9.4.2 Linear accelerators

Let us now assume that a charge is accelerated in a straight line:

$$\vec{v} \parallel \frac{d\vec{v}}{dt}.$$

Eq. 9.70 gives that the radiated power is:

$$P_{\text{rad.}} = \frac{q^2}{4\pi}\frac{2}{3}\gamma^6 |\frac{d\vec{v}}{dt}|^2. \tag{9.73}$$

For a relativistic charge the energy loss is

$$P_{\text{rad.}} \sim \gamma^6.$$

*Remark.* The $\gamma^6$ is at fixed $|\frac{d\vec{v}}{dt}|$. For a given applied force, $d|\vec{p}|/dt = \gamma^3 m|\frac{d\vec{v}}{dt}|$, so Eq. 9.66 gives $P_{\text{rad.}} = \frac{q^2}{4\pi}\frac{2}{3m^2}(d\vec{p}/dt)^2$, independent of $\gamma$: radiation losses in linear accelerators are negligible. In circular motion $d\vec{p}/dt = \gamma m\frac{d\vec{v}}{dt}$ and $P_{\text{rad.}} \propto \gamma^2(d\vec{p}/dt)^2$, i.e. $\propto \gamma^4/\rho^2$ at fixed radius $\rho$, which is why synchrotron losses dominate for electron rings (see Ref. [2], Sec. 14.2).

## 9.5 Angular distribution of radiation from a linearly accelerated relativistic charge

In this section, we will study the angular dependence of radiation from a charge which is accelerated in a straight line. The electric field of an accelerated charge is:

$$\vec{E} = \frac{q}{4\pi(1-\hat{n}\cdot\vec{v})^3}\frac{1}{|\vec{R}|}\hat{n}\times\left[(\hat{n}-\vec{v})\times\frac{d\vec{v}}{dt}\right] + O\left(\frac{1}{|\vec{R}|^2}\right)\Bigg|_{\text{retarded}}, \tag{9.74}$$

and for

$$\vec{v} \parallel \frac{d\vec{v}}{dt},$$

it becomes

$$\vec{E} = \frac{q}{4\pi(1-\hat{n}\cdot\vec{v})^3}\frac{1}{|\vec{R}|}\hat{n}\times\left[\hat{n}\times\frac{d\vec{v}}{dt}\right] + O\left(\frac{1}{|\vec{R}|^2}\right)\Bigg|_{\text{retarded}}. \tag{9.75}$$

The Poynting vector is:

$$\vec{S} = \hat{n}\frac{q^2}{16\pi^2|\vec{R}|^2}\frac{\left|\hat{n}\times(\hat{n}\times\frac{d\vec{v}}{dt})\right|^2}{(1-\hat{n}\cdot\vec{v})^6} + O\left(\frac{1}{|\vec{R}|^3}\right). \tag{9.76}$$

The power of radiation through a solid angle $d\Omega$ at a retarded distance $|\vec{R}|$ is:

$$\frac{dP_{\text{rad.}}}{d\Omega} = \frac{q^2}{16\pi^2}\left|\frac{d\vec{v}}{dt}\right|^2\frac{\sin^2\Theta}{(1-v\cos\Theta)^6}, \tag{9.77}$$

where $\Theta$ is the angle between the direction of the emitted radiation and the velocity (or, equivalently, the acceleration) of the particle. *Remark.* $|\vec{R}|^2\,\hat{n}\cdot\vec{S}$ is the energy per unit *observer* time $t$. The energy per unit *emission* time $t' = r^0$ of the charge carries one extra factor $dt/dt' = 1 - \hat{n}\cdot\vec{v}$, i.e. the power $(1 - v\cos\Theta)^{-5}$ (see Ref. [2], Eq. (14.39)); only that distribution integrates to the Liénard result Eq. 9.73, $\int d\Omega\,\sin^2\Theta/(1-v\cos\Theta)^5 = \frac{8\pi}{3}\gamma^6$, while the $(\cdot)^{-6}$ form integrates to a larger value. With the exponent 5 the peak is at $\gamma\Theta = 1/2$ instead of $1/\sqrt{5}$; the conclusion $\Theta \sim 1/\gamma$ is unchanged.

Let us now examine the limit where the velocity of the particle is very close to the speed of light:

$$v \approx c = 1.$$

In that limit, the denominator of Eq. 9.77 becomes large for radiation emitted collinearly to the direction of motion:

$$\frac{1}{(1-v\cos\Theta)^6} \overset{v\approx 1}{\longrightarrow} \frac{1}{(1-\cos\Theta)^6} \overset{\Theta\to 0}{\longrightarrow} \infty.$$

Therefore, radiation tends to be collinear. To study this double limit better, $\Theta \to 0, v \to 1$, we observe that (**exercise**):

$$1 - v\cos\Theta \approx \frac{1}{2\gamma^2}(1+\gamma^2\Theta^2), \tag{9.78}$$

and

$$\frac{dP_{\text{rad.}}}{d\Omega} \approx \frac{4q^2}{\pi^2}\gamma^{10}\left|\frac{d\vec{v}}{dt}\right|^2 \frac{(\gamma\Theta)^2}{(1+(\gamma\Theta)^2)^6}. \tag{9.79}$$

The distribution (**exercise:** plot it) vanishes for small and large values of $\gamma\Theta$. The maximum of the distribution is for values of

$$\gamma\Theta = \frac{1}{\sqrt{5}} = 0.4472\ldots.$$

Therefore, we conclude that the radiation is emitted within a characteristic angle:

$$\Theta \sim \frac{1}{\gamma}. \tag{9.80}$$

# 10

# Scattering

Consider an electric charge $+q$ which is struck by an incident electromagnetic field with a certain frequency $\omega$. An electromagnetic force is then exerted on the charge which accelerates it. Consequently, the electric charge will emit radiation. In this section, we will study the characteristics of the radiation scattered by the accelerated charge. We will distinguish two cases. In one (Thomson scattering), the charge is free, as happens inside a conductor or a plasma. In the second (Rayleigh scattering), the charge is bound inside an atom or a molecule.

Scattering of light on charges happens everywhere around us: sunlight scatters on the electrons of the molecules of the air, or on the free electrons of the ionosphere; in the laboratory we shine a laser on a target and place detectors around it. The two cases differ in how the incident wave accelerates the charge: a bound charge is subject also to the forces of its atom, so its acceleration, and hence its radiation, is different. Notice the logic of our classical description: the incident field accelerates the charge, and then we compute the radiation of the accelerated charge, as if the incident field had been absorbed, without any further cross-talk between the incident and the emitted field. This is the best we can do classically. Remarkably, it is also the quantum picture: the electron absorbs the incident photon and emits a new one (Chapter 12).

## 10.1 Thomson scattering

Consider a free charge $q$ at rest, $\vec{v} = 0$ (or moving with a very small, non-relativistic velocity $v \approx 0$), struck by an electromagnetic field:

$$\vec{E} = \hat{e} E \Re e^{i(\omega t - \vec{k}\cdot\vec{x})}. \tag{10.1}$$

The force acting on the charge, placed at the position $\vec{x} = 0$, is

$$\vec{F} = q\vec{E}, \tag{10.2}$$

and the acceleration of the charge will be:

$$\frac{d\vec{v}}{dt} = \frac{\vec{F}}{m} = \hat{e}\frac{qE}{m}\Re e^{i\omega t} = \hat{e}\frac{qE}{m}\cos(\omega t). \tag{10.3}$$

Due to its acceleration, the charge emits radiation with a power angular distribution given by Larmor's formula:

$$\frac{dP_{\text{rad.}}}{d\Omega} = \frac{q^2}{16\pi^2}\frac{d\vec{v}}{dt}^2 \sin^2\Theta = \frac{q^4}{16\pi^2 m^2}\sin^2\Theta E^2\cos^2(\omega t), \tag{10.4}$$

where $\Theta$ is the angle between the direction of the acceleration $\hat{e}$ and the direction $\hat{n}$ of the emitted radiation. Averaging over time, we have

$$\frac{1}{T}\int_0^T dt\,\cos^2(\omega t) = \frac{1}{2}, \quad T = \frac{2\pi}{\omega}, \tag{10.5}$$

and thus

$$\left\langle \frac{dP_{\text{rad.}}}{d\Omega} \right\rangle = \frac{q^4}{32\pi^2 m^2} E^2 \sin^2\Theta. \tag{10.6}$$

The average incoming flux of radiation is the time-average of the incoming Poynting vector:

$$\langle S \rangle = \left\langle E^2 \cos^2(\omega t) \right\rangle = \frac{E^2}{2}. \tag{10.7}$$

The differential cross-section is defined as the ratio of the outgoing radiation power per unit solid angle to the incoming flux of energy:

$$\frac{d\sigma}{d\Omega} \equiv \frac{\left\langle \frac{dP_{\text{rad.}}}{d\Omega} \right\rangle}{\langle S \rangle} = \frac{q^4}{16\pi^2 m^2} \sin^2\Theta. \tag{10.8}$$

It has units of area. Experimentally, it is more convenient to express the cross-section in terms of the angle $\theta$ formed by the propagation direction $\hat{k}$ of the incident electromagnetic field and the direction of the outgoing radiation (see Fig. 10.1). We can write:

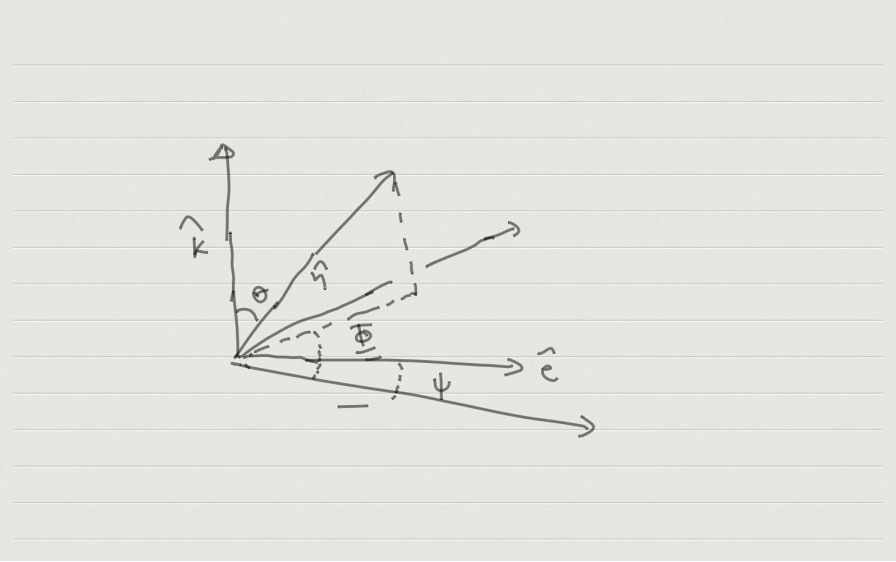



**Figure 10.1.** Angles in Thomson scattering: $\theta$ is the angle between the direction $\hat{k}$ of the incident wave and the direction $\hat{n}$ of the scattered radiation, and $\Theta$ the angle between $\hat{n}$ and the polarisation vector $\hat{e}$; $\phi$ and $\psi$ are the azimuthal angles of $\hat{n}$ and $\hat{e}$ around $\hat{k}$.

$$\sin^2\Theta = 1 - \sin^2\theta\cos^2(\phi - \psi). \tag{10.9}$$

Averaging over the azimuthal angle $\psi$ of the polarisation vector, we have

$$\frac{1}{2\pi}\int_0^{2\pi} d\psi\,\cos^2(\phi - \psi) = \frac{1}{2}. \tag{10.10}$$

Thus, the differential cross-section for an unpolarised incoming beam is:

$$\frac{d\sigma}{d\Omega} = \frac{q^4}{16\pi^2 m^2}\frac{1+\cos^2\theta}{2}. \tag{10.11}$$

Radiation is scattered in all directions, but mostly forward ($\theta = 0$) and backward ($\theta = \pi$); perpendicular to the beam, at $\theta = \pi/2$, the cross-section is half as large.

The total cross-section is obtained by integrating the differential cross-section over all solid angles:

$$\sigma = \int d\Omega \frac{d\sigma}{d\Omega} = \frac{q^4}{16\pi^2 m^2}\frac{8\pi}{3} = \frac{8\pi}{3} r_q^2, \tag{10.12}$$

where

$$r_q \equiv \frac{q^2}{4\pi\epsilon_0 m c^2}. \tag{10.13}$$

For an electron, the Thomson radius is:

$$r_e \approx 2.82 \times 10^{-13}\,\text{cm}. \tag{10.14}$$

Why "radius"? A classical billiard ball of radius $r$ struck by a tiny projectile has a cross-section equal to the area $\pi r^2$ of its silhouette. If we insist on picturing the electron as a small ball, Eq. 10.12 assigns to it an effective size $r_e$; the total cross-section is $\sigma_{\text{Thomson}} = \frac{8\pi}{3} r_e^2 \approx 6.65 \times 10^{-29}\,\text{m}^2 = 0.665\,\text{barn}$. This is not a true size: as far as we can tell experimentally, the electron is point-like. Still, $r_e$ is a very useful number for back-of-the-envelope estimates of how often light scatters on electrons.

Notice that the cross-section is independent of the frequency of the incoming electromagnetic wave. Our result for the cross-section receives significant quantum corrections for frequencies $\hbar\omega \sim mc^2$.

## 10.2 Rayleigh scattering

We now examine the scattering of an electromagnetic wave on an electric charge which is bound in an atom or molecule. The position of the charge is then obtained by solving

$$\frac{qE}{m} = \ddot{x} + \gamma \dot{x} + \omega_0^2 x, \tag{10.15}$$

with a solution:

$$\vec{x} = \frac{\frac{q\vec{E}}{m}}{\omega_0^2 - \omega^2 + i\gamma\omega}. \tag{10.16}$$

The acceleration is then:

$$\frac{d\vec{v}}{dt} = \ddot{\vec{x}} = \frac{\frac{\ddot{q\vec{E}}}{m}}{\omega_0^2 - \omega^2 + i\gamma\omega} = \frac{q\vec{E}}{m}\frac{-\omega^2}{\omega_0^2 - \omega^2 + i\gamma\omega}. \tag{10.17}$$

This is a crude, but useful, model of an atom. When the incident wave pulls the charge away, the atom pulls it back; the simplest such restoring force is that of a harmonic oscillator with frequency $\omega_0$, where $\hbar\omega_0$ is of the order of the binding energy of the charge. The friction constant $\gamma$ (not to be confused with the Lorentz factor) is phenomenological: it represents losses, for example radiation which is absorbed by other electrons of the atom

instead of reaching our detector. We also want the atom to stay intact: for $\hbar\omega \sim \hbar\omega_0$ it can be excited or ionised. With $e^{-i\omega t}$ instead of $e^{i\omega t}$ the sign of $i\gamma\omega$ flips. This is only a convention: physical quantities are real parts or absolute values squared, and in Eq. 10.20 $\gamma$ appears squared.

Unlike in Thomson scattering, where the charge is free, the acceleration is now frequency dependent. We can write

$$\frac{d\vec{v}}{dt} = \frac{d\vec{v}}{dt}_{\text{Thomson}} \frac{-\omega^2}{\omega_0^2 - \omega^2 + i\gamma\omega}, \tag{10.18}$$

where

$$\frac{d\vec{v}}{dt}_{\text{Thomson}} = \frac{q\vec{E}}{m}, \tag{10.19}$$

the acceleration of the charge if it were free. The calculation of the cross-section proceeds exactly as in Thomson scattering. We find

$$\sigma = \sigma_{\text{Thomson}} \frac{\omega^4}{(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2}, \tag{10.20}$$

with

$$\sigma_{\text{Thomson}} \equiv \frac{8\pi}{3} r_q^2. \tag{10.21}$$

For $\omega \ll \omega_0$, the process is known as Rayleigh scattering and the cross-section has a characteristic $\omega^4$ frequency dependence:

$$\sigma_{\text{Rayleigh}} = \sigma_{\text{Thomson}} \frac{\omega^4}{\omega_0^4}. \tag{10.22}$$

High frequencies are scattered more than lower frequencies. Rayleigh scattering takes place in the atmosphere. Sunlight contains all visible frequencies; the highest ones, at the blue end of the spectrum, scatter the most. This explains why the light we see away from the line of sight to the sun is bluish, while it is more yellow or red (at sunset and sunrise) when we look in the direction of the sun (see a video on the blue sky). *Remark.* The highest visible frequencies are in fact violet; the sky appears blue rather than violet because the solar spectrum falls off in the violet, part of it is absorbed in the upper atmosphere, and the eye is much less sensitive there. The scatterers are the $N_2$, $O_2$ molecules, whose resonances $\omega_0$ lie in the ultraviolet, so $\omega \ll \omega_0$ holds for visible light (see Ref. [2], Sec. 10.2).

# 11

# Electrodynamics in a medium

In a macroscopic amount of matter we can identify two types of motion of charged particles:

- fast currents within small distances, due to the motion of charges in atoms and molecules;
- slow currents, due to free electrons or ions, which extend over distances much larger than the size of atoms.

Electrodynamics in a medium is not a new theory: Maxwell's equations hold as before. The difficulty is that a macroscopic piece of matter contains a huge number of charges and we cannot solve their equations of motion one by one. We need an *effective theory*: instead of solving the full theory and then taking the limit which is relevant to us, we replace it from the start by a simpler theory which is valid directly in that limit. The limit here is the hierarchy between atomic and macroscopic scales. Atomic distances are much smaller than the distances over which we measure the fields (light, for example, has a wavelength much larger than an atom and sees its charges only on average), and atoms as a whole move slowly while the charges inside them move fast. What we derive will be adequate for an engineer who uses materials in a device, but not for atomic physics, where these assumptions fail.

In our macroscopic measurements of the currents and the electromagnetic field, we are sensitive only to the slow variations and insensitive to the fast ones taking place at the atomic level. We can then average over the currents within the atoms. We can write

$$j^\mu \approx j^\mu_{\text{slow}} + \left\langle j^\mu_{\text{atomic}} \right\rangle , \tag{11.1}$$

where $\langle X \rangle$ is the macroscopic average of the quantity $X$. It is not clear how one should average over the atomic distances. However, we can parametrise our ignorance by means of macroscopic properties of the material, which reflect its atomic structure and which we can measure experimentally.

Averaging is not only a matter of convenience. Quantum mechanically, it makes no sense to ask for both the position and the velocity of a charge inside an atom; and even classically, the charges are too many and too fast to be followed. We can, however, tell atoms apart: our ignorance concerns distances of the order of a few atoms and not beyond.

We introduce an averaging of a function over some distances via:

$$\left\langle F(\vec{x}, t) \right\rangle = \int d^3\vec{y}\, f(\vec{y}) F(\vec{x} - \vec{y}, t). \tag{11.2}$$

The weighting factor $f(\vec{y})$ is unknown, but let's assume that it is a well-behaved smooth positive function with unit norm (for a probabilistic interpretation),

$$\int d^3\vec{y}\, f(\vec{y}) = 1. \tag{11.3}$$

The function $f(\vec{y})$ should be peaked at $\vec{y} = 0$ and fall off to zero very fast for distances $|\vec{y}| > r_{\text{atom}}$.

## 11.1 Average of the atomic charge density

The charge density $j^0_{\text{atomic}} = \rho_{\text{atomic}}$ corresponds to the charge density of the charges inside molecules/atoms. Characteristically, the charges are constrained within regions of a small atomic size. We write

$$\rho_{\text{atomic}} = \sum_{n\in \text{ atoms}} \rho_{(n)}, \tag{11.4}$$

where $\rho_{(n)}$ is the charge density of the $n$-th atom/molecule. Let's assume that the $n$-th atom/molecule consists of charges $q_j$:

$$\rho_{(n)} = \sum_{j\in\,(n)} q_j \delta(\vec{x} - \vec{x}_n - \vec{x}_j), \tag{11.5}$$

where $\vec{x}_n$ is the position of the atom/molecule and $\vec{x}_j$ is the position of the charge with respect to the "centre" of the molecule. The average is:

$$\begin{aligned} \langle \rho_{(n)} \rangle &= \sum_{j\in(n)} \langle q_j \delta(\vec{x} - \vec{x}_n - \vec{x}_j) \rangle \\ &= \sum_{j\in(n)} q_j \int d^3\vec{y}\, f(\vec{y}) \delta(\vec{x} - \vec{x}_n - \vec{x}_j - \vec{y}) \\ &= \sum_{j\in(n)} q_j f(\vec{x} - \vec{x}_n - \vec{x}_j). \end{aligned} \tag{11.6}$$

We can now perform a Taylor expansion in

$$\frac{|\vec{x}_j|}{|\vec{x} - \vec{x}_n|}, \tag{11.7}$$

the atomic scale over the macroscopic scale of observation. We obtain:

$$\begin{aligned} \langle \rho_{(n)} \rangle &= \sum_{j\in(n)} \left[ q_j f(\vec{x} - \vec{x}_n) - (q_j \vec{x}_j) \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) + \ldots \right] \\ \rightsquigarrow \langle \rho_{(n)} \rangle &= \left( \sum_{j\in(n)} q_j \right) f(\vec{x} - \vec{x}_n) - \left( \sum_{j\in(n)} q_j \vec{x}_j \right) \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) + \ldots . \end{aligned} \tag{11.8}$$

On the right-hand side, we recognise the total charge

$$q_n = \sum_{j\in(n)} q_j \tag{11.9}$$

and the dipole moment

$$\vec{p}_n = \sum_{j\in(n)} q_j \vec{x}_j \tag{11.10}$$

of the $n$-th molecule. Thus:

$$\left\langle \rho_{(n)} \right\rangle = q_n f(\vec{x} - \vec{x}_n) - \vec{p}_n \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) + \dots . \tag{11.11}$$

The gradient operator $\vec{\nabla}$ differentiates with respect to the components of $\vec{x}$ (it does not act on $\vec{x}_n$ or $\vec{x}_j$). We can then write the above in the form

$$\left\langle \rho_{(n)} \right\rangle = q_n f(\vec{x} - \vec{x}_n) - \vec{\nabla} \cdot \left(\vec{p}_n f(\vec{x} - \vec{x}_n)\right) + \dots . \tag{11.12}$$

Summing over all atoms/molecules, we find that the average charge density in atoms/molecules can be written as:

$$\left\langle \rho_{\text{atomic}} \right\rangle \equiv \sum_n \left\langle \rho_{(n)} \right\rangle = \left\langle \rho_{\text{eff/atom}} \right\rangle - \vec{\nabla} \cdot \vec{P} + \dots , \tag{11.13}$$

where

$$\left\langle \rho_{\text{eff/atom}} \right\rangle = \sum_{n\in\text{atoms}} q_n f\left(\vec{x} - \vec{x}_n\right) \tag{11.14}$$

is an effective charge distribution computed by treating the atoms as point-like with a total charge $q_n = \sum_{j\in(n)} q_j$, and

$$\vec{P} \equiv \sum_{n\in\text{atoms}} \vec{p}_n f(\vec{x} - \vec{x}_n) \tag{11.15}$$

is the so-called "polarisation of the medium". The polarisation $\vec{P}$ sums together the average dipole moments of all atoms and it is the leading contribution due to the non-point-like geometry of the atoms.

## 11.2 Average of the atomic current density

We will restrict our computation to currents with motions which are not relativistic.

The current density in an atom/molecule can be written as

$$\vec{j}_{(n)}(\vec{x}, t) = \sum_{k\in n} q_k(\vec{v}_n + \vec{v}_k)\delta(\vec{x} - \vec{x}_n - \vec{x}_k), \tag{11.16}$$

where the velocity of the $k$-th charge in the $n$-th atom is expressed as the sum $\vec{v}_n + \vec{v}_k$ of the velocity $\vec{v}_n$ of the atom as a whole and the velocity $\vec{v}_k$ of the charge relative to the centre of the atom.[1].

For the average, we have:

$$\left\langle \vec{j}_n \right\rangle = \sum_k q_k(\vec{v}_n + \vec{v}_k) f(\vec{x} - \vec{x}_n - \vec{x}_k). \tag{11.17}$$

For what follows, it is important to remember that $|\vec{x}_k| \ll |\vec{x} - \vec{x}_n|$ and that the velocity of the atom as a whole system is much smaller than the relative velocities of its constituent charges $|\vec{v}_n| \ll |\vec{v}_k|$. Once again, we expand

$$f(\vec{x} - \vec{x}_n - \vec{x}_k) \approx f(\vec{x} - \vec{x}_n) - \vec{x}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) + \dots . \tag{11.18}$$

[1] Relativistically, the velocity of the charge is not a simple sum as we have assumed here.

This yields

$$\begin{aligned}\left\langle \vec{j}_n \right\rangle &= \sum_k q_k \vec{v}_k f(\vec{x} - \vec{x}_n) \\ &\quad + \sum_k q_k \vec{v}_n f(\vec{x} - \vec{x}_n) - \sum_k q_k \vec{v}_k \, \vec{x}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) \\ &\quad + O\left(x_k^2, x_k v_n, v_n^2\right). \end{aligned} \tag{11.19}$$

The term in the first line is parametrically the largest, as it depends on large velocities within an atom. The terms in the second line are smaller as they are suppressed by either a small distance within an atom $x_k$ or a slow velocity of a whole atom $v_n$. Finally, we omit higher-order terms suppressed by the product of two such small parameters.

We now rewrite

$$\begin{aligned}\sum_k q_k \vec{v}_k f(\vec{x} - \vec{x}_n) &= \frac{d}{dt}\left(\sum_k q_k \vec{x}_k f(\vec{x} - \vec{x}_n)\right) - \sum_k q_k \vec{x}_k \frac{d}{dt} f(\vec{x} - \vec{x}_n) \\ &= \frac{d}{dt}\left(\vec{p}_n f(\vec{x} - \vec{x}_n)\right) + \sum_k q_k \vec{x}_k \left(\vec{v}_n \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n)\right) \\ &\approx \frac{d}{dt}\left(\vec{p}_n f(\vec{x} - \vec{x}_n)\right). \end{aligned} \tag{11.20}$$

We can also write

$$\begin{aligned}&\sum_k q_k \vec{v}_k \, \vec{x}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) \\ &= \sum_k \frac{q_k}{2} \vec{v}_k \, \vec{x}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) - \sum_k \frac{q_k}{2} \vec{x}_k \, \vec{v}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) \\ &\quad + \frac{d}{dt}\left(\sum_k \frac{q_k}{2} \vec{x}_k \, \vec{x}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n)\right) \\ &\approx \sum_k \frac{q_k}{2} \vec{v}_k \, \vec{x}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) - \sum_k \frac{q_k}{2} \vec{x}_k \, \vec{v}_k \cdot \vec{\nabla} f(\vec{x} - \vec{x}_n) \\ &\approx -\vec{\nabla} \times \left(\vec{m}_n f(\vec{x} - \vec{x}_n)\right), \end{aligned} \tag{11.21}$$

where we have used that

$$\vec{a} \times \left(\vec{b} \times \vec{c}\right) = \vec{b}\left(\vec{a} \cdot \vec{c}\right) - \vec{c}\left(\vec{a} \cdot \vec{b}\right) \tag{11.22}$$

and recognised the magnetic moment of an atom,

$$\vec{m}_n = \sum_j \frac{q_j}{2} (\vec{x}_j \times \vec{v}_j). \tag{11.23}$$

*Remark.* The total time derivative $\frac{d}{dt} \sum_k \frac{q_k}{2} \vec{x}_k \left(\vec{x}_k \cdot \vec{\nabla} f\right)$ which we have dropped is of the same order $(x_k^2/T)$ as the magnetic-moment term that we keep. It is the current partner of the electric quadrupole term in $\left\langle \rho_{(n)} \right\rangle$, which we have also dropped (the "…"). Dropping both together is consistent and conserves charge (see Ref. [2], Sec. 6.6). Summing the averaged contributions from all atoms, we have

$$\left\langle \vec{j}_{\text{atomic}} \right\rangle = \sum \left\langle \vec{j}_{(n)} \right\rangle = \vec{j}_{\text{eff/atomic}} + \frac{d\vec{P}}{dt} + \vec{\nabla} \times \vec{M}, \tag{11.24}$$

where

$$\vec{M} = \sum_n \left(\vec{m}_n f(\vec{x} - \vec{x}_n)\right) \tag{11.25}$$

is the magnetization of the material and

$$\vec{j}_{\text{eff/atomic}} = \sum_n q_n \vec{v}_n f(\vec{x} - \vec{x}_n). \tag{11.26}$$

## 11.3 Maxwell's equations in a medium

We can now approximate the charge and current density in the medium by

$$j^\mu = j^\mu_{\text{free}} + j^\mu_{\text{atomic}} \approx j^\mu_{\text{free}} + \left\langle j^\mu_{\text{atomic}} \right\rangle . \tag{11.27}$$

By direct substitution into Maxwell's equations

$$\partial_\mu F^{\mu\nu} \approx j^\nu_{\text{free}} + \left\langle j^\nu_{\text{atomic}} \right\rangle , \tag{11.28}$$

we obtain

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{11.29}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{11.30}$$

$$\vec{\nabla} \cdot \left(\vec{E} + \frac{\vec{P}}{\epsilon_0}\right) = \frac{\rho_{\text{eff}}}{\epsilon_0}, \tag{11.31}$$

where

$$\rho_{\text{eff}} = \rho_{\text{free}} + \rho_{\text{eff/atomic}}, \tag{11.32}$$

and

$$\vec{\nabla} \times \left(\vec{B} - \frac{\vec{M}}{c^2 \epsilon_0}\right) = \frac{\vec{j}_{\text{eff}}}{c^2 \epsilon_0} + \frac{1}{c^2}\frac{\partial}{\partial t}\left(\vec{E} + \frac{\vec{P}}{\epsilon_0}\right), \tag{11.33}$$

with

$$\vec{j}_{\text{eff}} = \vec{j}_{\text{free}} + \vec{j}_{\text{eff/atomic}}. \tag{11.34}$$

### 11.3.1 The $\vec{D}$ and $\vec{H}$ fields

Historically, Maxwell's equations in matter have often been presented in terms of the combinations:

$$\vec{D} \equiv \epsilon_0 \vec{E} + \vec{P} \tag{11.35}$$

and

$$\vec{H} \equiv \vec{B} - \frac{\vec{M}}{c^2 \epsilon_0}. \tag{11.36}$$

In this notation, and writing $\rho, \vec{J}$ for $\rho_{\text{eff}}, \vec{j}_{\text{eff}}$, Maxwell's equations in the medium take the form:

$$\vec{\nabla} \cdot \vec{D} = \rho \tag{11.37}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{11.38}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{11.39}$$

$$\vec{\nabla} \times \vec{H} = \frac{\vec{J}}{c^2 \epsilon_0} + \frac{1}{\epsilon_0 c^2} \frac{\partial \vec{D}}{\partial t}. \tag{11.40}$$

*Remark.* With $1/(c^2\epsilon_0) = \mu_0$, our field $\vec{H} = \vec{B} - \mu_0 \vec{M}$ is $\mu_0$ times the field $\vec{H}_{\text{SI}} = \vec{B}/\mu_0 - \vec{M}$ used in most textbooks (see, e.g., Ref. [2], Eq. (6.81)). In terms of $\vec{H}_{\text{SI}}$ the last equation reads $\vec{\nabla} \times \vec{H}_{\text{SI}} = \vec{J} + \partial \vec{D}/\partial t$.

The fields $\vec{D}$ and $\vec{H}$ are nothing mysterious. They absorb into a redefinition of the fields the fact that, seen from large distances, atoms look like small electric dipoles and small magnets. Maxwell himself wrote his equations in terms of such quantities, since experiments were always done in the presence of matter. Keep in mind that, because of the averaging and the Taylor expansion, all these equations hold only approximately.

## 11.4 Maxwell's equations inside a dielectric material

Assume a dielectric material, such as air or water, with no magnetisation $\vec{M} = 0$ but with the ability to acquire a polarisation $\vec{P}$. Maxwell's equations take the form:

$$\vec{\nabla} \cdot \left( \vec{E} + \frac{\vec{P}}{\epsilon_0} \right) = \frac{\rho}{\epsilon_0}$$

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

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{11.41}$$

$$\vec{\nabla} \times \vec{B} = \frac{\vec{J}}{\epsilon_0 c^2} + \frac{1}{c^2} \frac{\partial}{\partial t} \left( \vec{E} + \frac{\vec{P}}{\epsilon_0} \right). \tag{11.42}$$

To solve these equations, we need information about the polarisation $\vec{P}$ of the medium. $\vec{P}$ and $\vec{E}$ are not truly independent. In fact, the larger the electric field $\vec{E}$ the more it will stretch the atoms and molecules in the medium inducing a larger polarisation $\vec{P}$. Ignoring non-linear effects, we can write a phenomenological proportionality relation between the electric field and the polarisation,

$$\vec{P} = \chi \epsilon_0 \vec{E}, \tag{11.43}$$

where $\chi$ is known as the "electric susceptibility" of the dielectric medium. Why linear? Expand $\vec{P}$ in powers of $\vec{E}$. There is no constant term: the atoms of stable matter are neutral, and without an external field there is no reason for their charges to be displaced. In an isotropic material (more precisely, one which looks the same after an inversion $\vec{x} \to -\vec{x}$) the polarisation must change sign when the field does, so that quadratic terms $P_i = \beta_{ijk} E_j E_k$ are absent as well; they would need a preferred direction. Non-linear terms arise when the binding of the charges deviates from a harmonic oscillator, and they are suppressed

by powers of the displacement of the charges over the size of the atom. They become important only in very strong fields, for example of intense lasers, which distort the atoms substantially. In the end, the linear approximation is justified by experiment.

Maxwell's equations then become:

$$\begin{aligned}
\vec{\nabla}\cdot\vec{E} &= \frac{\rho}{\epsilon_0(1+\chi)} \\
\vec{\nabla}\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t} \\
\vec{\nabla}\cdot\vec{B} &= 0 \qquad (11.44)\\
\vec{\nabla}\times\vec{B} &= \frac{\vec{J}}{\epsilon_0 c^2}+\frac{1+\chi}{c^2}\frac{\partial\vec{E}}{\partial t}. \qquad (11.45)
\end{aligned}$$

Defining

$$c_m = \frac{c}{\sqrt{1+\chi}} \qquad (11.46)$$

and

$$\epsilon = (1+\chi)\epsilon_0, \qquad (11.47)$$

Maxwell's equations take the same form as in the vacuum,

$$\begin{aligned}
\vec{\nabla}\cdot\vec{E} &= \frac{\rho}{\epsilon} \\
\vec{\nabla}\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t} \\
\vec{\nabla}\cdot\vec{B} &= 0 \qquad (11.48)\\
\vec{\nabla}\times\vec{B} &= \frac{\vec{J}}{\epsilon c_m^2}+\frac{1}{c_m^2}\frac{\partial\vec{E}}{\partial t}, \qquad (11.49)
\end{aligned}$$

with the electric permittivity and the speed of light replaced by

$$\epsilon_0 \to \epsilon, \quad c \to c_m. \qquad (11.50)$$

The effect of the medium is therefore twofold. For static fields $\chi > 0$ (the dipoles align along the field which induces them), so that charges and currents appear weaker by a factor $1+\chi$, as if partially screened by the polarised atoms around them; and electromagnetic disturbances propagate with a smaller speed $c_m < c$. We will see shortly that $\chi$ depends on the frequency and is in general complex, so this interpretation must be taken with some care.

## 11.5 Waves in a dielectric medium

Obviously, the solutions that we have found for Maxwell's equations in the vacuum are also solutions of Maxwell's equations in a dielectric, as long as we make the replacement of Eq. 11.50. Away from macroscopic charges and currents, we have (as in the vacuum) that the electric and magnetic field satisfy differential equations of the type:

$$\left[\frac{1}{c_m^2}\frac{\partial^2}{\partial t^2}-\nabla^2\right]\vec{E}(\vec{x},t)=0, \qquad (11.51)$$

which admit solutions

$$\vec{E} = \vec{E}_0 e^{i(\omega t - \vec{k}\cdot\vec{x})}. \tag{11.52}$$

This represents a wave with frequency

$$\omega = 2\pi f \tag{11.53}$$

and phase velocity

$$v_{\text{phase}} = \frac{\omega}{\left|\vec{k}\right|}. \tag{11.54}$$

Substituting the solution into the differential equation, we obtain

$$k^2 = \frac{\omega^2}{c_m^2} = \frac{\omega^2}{c^2}(1+\chi). \tag{11.55}$$

For the phase velocity, we have

$$v_{\text{phase}} = \frac{c}{n}, \tag{11.56}$$

where

$$n = \sqrt{1+\chi} \tag{11.57}$$

is the so-called index of refraction, which characterises the dielectric material.

## 11.6 A model for the dielectric susceptibility $\chi$

To make further progress in understanding the properties of electromagnetic waves inside dielectrics, we need to calculate $\chi = n^2 - 1$ which characterises the material. Equivalently, we need to calculate the polarisation $\vec{P}$ of the material when we subject it to an electric field $\vec{E}$.

In a dielectric, positive and negative charges are bound together inside atoms or molecules. To a good approximation, we can view the dielectric as a collection of dipoles each with a positive and a negative charge $(+q, -q)$. The charges in a dipole are bound together by their electric attraction, but they cannot come arbitrarily close to each other because of the laws of quantum mechanics. A full account of the quantum effects which are at play is beyond our scope. However, we can approximate phenomenologically the resulting binding force as a harmonic oscillator. If a charge separates away from the dipole, a force will pull it back, and if it moves closer to the other charge a force will push it away. As we have seen when studying radiation, electric charges which accelerate radiate and lose energy. To account for this effect, we can also include a friction or drag force. Let's assume that the electric field is in the $\hat{x}$ direction. The force acting on a charge $+q$ of a dipole is then

$$F = qE = m(\ddot{x} + \gamma\,\dot{x} + \omega_0^2 x). \tag{11.58}$$

For a time-varying electric field

$$E = E_0 e^{i\omega t}, \tag{11.59}$$

we find solutions of the form:

$$x = x_0 e^{i\omega t}, \tag{11.60}$$

with

$$x_0 = \frac{\frac{qE_0}{m}}{\omega_0^2 - \omega^2 + i\omega\gamma}. \tag{11.61}$$

Therefore,

$$\vec{x} = \frac{\frac{q}{m}}{\omega_0^2 - \omega^2 + i\omega\gamma}\vec{E}, \tag{11.62}$$

and the electric dipole moment is

$$\vec{p} = q\vec{x} = \frac{\frac{q^2}{m}}{\omega_0^2 - \omega^2 + i\omega\gamma}\vec{E}. \tag{11.63}$$

Assuming a constant density $N$ of $+q$ charges, the polarisation of the dielectric is:

$$\vec{P} = N\vec{p} = Nq\vec{x} = \frac{Nq^2}{m}\frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma}\vec{E}. \tag{11.64}$$

The dielectric susceptibility is then

$$\chi = n^2 - 1 = \frac{Nq^2}{m\epsilon_0}\frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma}. \tag{11.65}$$

Equivalently, the square of the index of refraction is

$$n^2 = 1 + \frac{Nq^2}{m\epsilon_0}\frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma}. \tag{11.66}$$

*Remark.* We have taken the field acting on each charge to be the macroscopic $\vec{E}$. This is adequate for dilute media (gases). In dense media the local field $\vec{E} + \vec{P}/(3\epsilon_0)$ acts instead, giving the Clausius–Mossotti (Lorentz–Lorenz) form $(n^2 - 1)/(n^2 + 2) = Nq^2/(3m\epsilon_0(\omega_0^2 - \omega^2 + i\omega\gamma))$ (see Ref. [2], Secs. 4.5 and 7.5). Real media also have several resonances $\omega_j$, each with its own oscillator strength $f_j$.

## 11.7 The complex index of refraction

We have found that the index of refraction is a complex number:

$$n = n_R - in_I. \tag{11.67}$$

A plane wave propagating in the dielectric is:

$$\vec{E} = \vec{E}_0 e^{i\omega\left[t - \frac{n}{c}\hat{k}\cdot\vec{x}\right]} = \vec{E}_0 e^{i\omega\left[t - \frac{n_R}{c}\hat{k}\cdot\vec{x}\right]} e^{-\omega\frac{n_I}{c}\hat{k}\cdot\vec{x}}. \tag{11.68}$$

The amplitude of the plane wave is

$$\left|\vec{E}\right| = \left|\vec{E}_0\right| e^{-\omega\frac{n_I}{c}\hat{k}\cdot\vec{x}}, \tag{11.69}$$

and it decreases (for the physical case $n_I > 0$) as the wave penetrates further inside the material.

This is why light does not reach the bottom of the ocean. A warning about complex quantities: the physical field is the real part of the complex solution, but the real part may only be taken at the very end. Taking the real part of $\chi$ or $n$ first would be wrong, since, e.g., the real part of $e^{i\omega t}/(a + ib)$ depends on $b$. The imaginary part of $n$ describes a genuine physical effect, the attenuation of the wave.

## 11.8 Waves in metals

Metals are conductors permitting electrons to move freely over large distances. We can compute their index of refraction as a special limit of Eq. 11.66 by setting

$$\omega_0 = 0.$$

This corresponds to the absence of the binding force which, in dielectrics, holds the charges near a fixed position. The index of refraction is then:

$$n^2 = 1 + \frac{Nq^2}{m\epsilon_0}\frac{1}{-\omega^2 + i\omega\gamma}. \tag{11.70}$$

The density $N$ can be obtained from macroscopic properties of the metal. The constant $\gamma$ is an intrinsic parameter of our model. To make contact with reality and test our theory of wave propagation in metals, we must find a way to relate $\gamma$ to an experimentally measured quantity.

In our model, the equation of motion for the electrons in the metal can be obtained by solving the differential equation:

$$q\vec{E} = m(\ddot{\vec{x}} + \gamma\,\dot{\vec{x}}), \tag{11.71}$$

which is a first-order differential equation for the velocity $v$ of the electrons:

$$q\vec{E} = m(\dot{\vec{v}} + \gamma\vec{v}). \tag{11.72}$$

Let us now consider a constant electric field $\vec{E}$. So far, we have considered waves, $\vec{E} = \vec{E}_0\cos(\omega t)$; we will return to waves shortly, but the case of a constant electric field teaches us something useful too. The general solution of the differential equation for a constant electric field is

$$\vec{v} = \frac{q\vec{E}}{m\gamma} + \vec{v}_0 e^{-\gamma t}. \tag{11.73}$$

The second term is exponentially decreasing with time. Therefore, the electrons after some time drift in the metal with a practically constant velocity:

$$\vec{v}_{\text{drift}} = \frac{q\vec{E}}{m\gamma}. \tag{11.74}$$

This corresponds to a current density

$$\vec{J} = Nq\vec{v}_{\text{drift}} = \frac{Nq^2}{m\gamma}\vec{E}. \tag{11.75}$$

Our model has led us to Ohm's law,

$$\vec{J} = \sigma\vec{E}, \tag{11.76}$$

where $\sigma$ is the conductivity of the metal.

**Exercise:** What is the relation between $\sigma$ and the resistance $R$ of a wire in $V = IR$?

We therefore obtain

$$\gamma = \frac{Nq^2}{m\sigma}. \tag{11.77}$$

To get a physical picture of the propagation of waves in metals, let us look at the low- and high-frequency limits.

### 11.8.1 Low-frequency approximation

In the low-frequency limit, $\omega \to 0$, we will assume that $\omega\gamma \gg \omega^2$. Eq. 11.70 gives:

$$n^2 = -i\frac{\sigma}{\epsilon_0\omega}. \tag{11.78}$$

The index of refraction is then

$$n = \sqrt{\frac{\sigma}{2\epsilon_0\omega}}(1-i). \tag{11.79}$$

The index of refraction has a large imaginary part (as large as the real part). The amplitude of a wave reduces exponentially with the distance $x$ as it propagates inside the conductor:

$$\left|\vec{E}\right| = \left|\vec{E}_0\right| e^{-\frac{x}{\delta}}, \tag{11.80}$$

where the penetration distance is:

$$\delta = \sqrt{\frac{2\epsilon_0 c^2}{\sigma\omega}}. \tag{11.81}$$

For copper and a microwave frequency $\nu = \frac{\omega}{2\pi} = 10^4\,\mathrm{MHz}$, the penetration length is

$$\delta \approx 6.7 \times 10^{-5}\,\mathrm{cm} \tag{11.82}$$

(see Ref. [1], Section 32-7).

The damping refers to a wave moving *into* the metal; a wave is free to travel back. A metal does not let low-frequency waves through: it reflects them.

### 11.8.2 High-frequency approximation

In the high-frequency limit, $\omega^2 \gg \omega\gamma$, the index of refraction becomes

$$n^2 \approx 1 - \frac{Nq^2}{m\epsilon_0}\frac{1}{\omega^2} = 1 - \frac{\omega_P^2}{\omega^2}, \tag{11.83}$$

where

$$\omega_P^2 = \frac{Nq^2}{m\epsilon_0}. \tag{11.84}$$

For frequencies which are smaller than the "plasma frequency" $\omega < \omega_P$, the index of refraction is imaginary and the wave is damped after some distance inside the metal. For frequencies larger than the plasma frequency, $\omega > \omega_P$, the index of refraction is real and the metal becomes transparent to the electromagnetic wave. Nothing in this derivation is specific to metals: it holds for any system of freely moving charges, in particular for a plasma. An example is the ionosphere, the upper layer of the atmosphere, which is ionised by solar radiation. With $q = e$ and $m = m_e$, $f_P = \omega_P/(2\pi) \approx 9\,\mathrm{Hz} \times \sqrt{N/\mathrm{m}^{-3}}$, and for electron densities $N \sim 10^{11}$–$10^{12}\,\mathrm{m}^{-3}$ this is a few MHz. Radio waves of lower frequency are reflected by the ionosphere (which lets short-wave radio reach around the globe), while communication with a satellite above the ionosphere requires frequencies well above $f_P$ (in practice, in the GHz range). Conversely, the frequency above which a material becomes transparent tells us the density of its free charges.

**Exercise:** Calculate the wavelengths $\lambda_P = \frac{2\pi c}{\omega_P}$ for Li, Na, K and Rb (see Ref. [1], Section 32-7).

## 11.9 Reflection and refraction

Consider two materials with indices of refraction $n_1$ and $n_2$, separated by a boundary surface in the $y$–$z$ plane, as in Fig. 11.1. Maxwell's equations should govern both regions,

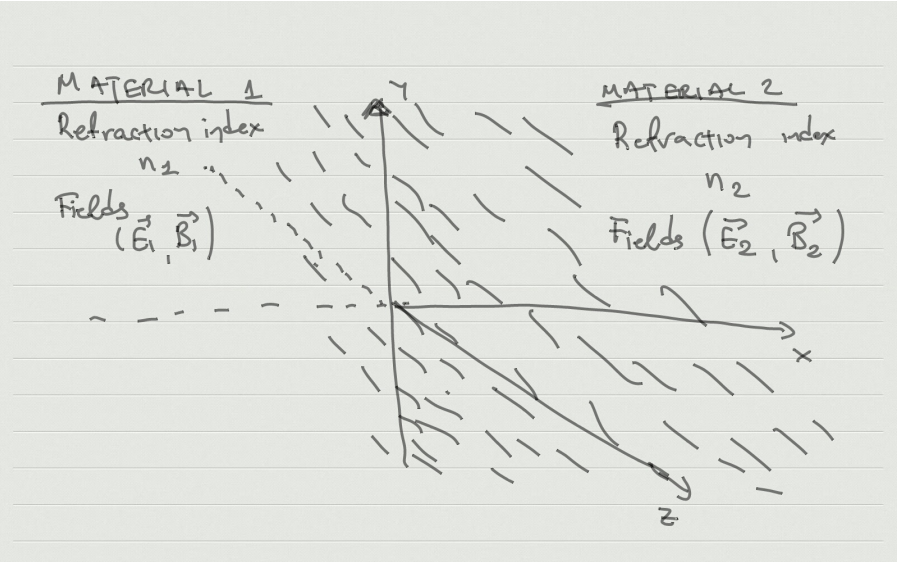



**Figure 11.1.** Two media with indices of refraction $n_1$ and $n_2$, separated by a boundary in the $y$–$z$ plane.

as well as the "intermediate" boundary region. Let's assume no macroscopic charges and currents in the two media. The first Maxwell equation is

$$\vec{\nabla} \cdot \vec{E} = -\frac{1}{\epsilon_0} \vec{\nabla} \cdot \vec{P}. \tag{11.85}$$

Written explicitly in terms of components, we have:

$$\partial_x E_x + \partial_y E_y + \partial_z E_z = -\frac{1}{\epsilon_0} \left( \partial_x P_x + \partial_y P_y + \partial_z P_z \right). \tag{11.86}$$

Derivatives with respect to the $x$-variable in the boundary regions compare the fields on the two sides of the boundary. We anticipate them to be larger than derivatives with respect to $y$ or $z$ or $t$, which compare the fields on the same side of the boundary or at different times. The reason is that the two materials polarise very differently: across the thin transition layer $\vec{P}$ jumps from its value in one medium to its value in the other, while along the boundary, or in time for a not too rapidly varying wave, the fields change smoothly. Notice that "no macroscopic charges" means $\rho_{\text{eff}} = 0$; the bound charges $-\vec{\nabla} \cdot \vec{P}$ of the atomic dipoles do exist, and they sit precisely in the boundary layer. (The same boundary conditions follow, more rigorously, from the integral form of Maxwell's equations applied to a thin box and a thin loop straddling the boundary.)

If this is the case, in the boundary region the above Maxwell equation becomes

$$\partial_x E_x = -\frac{1}{\epsilon_0} \partial_x P_x \tag{11.87}$$

or, equivalently,

$$\frac{E_{x2} - E_{x1}}{\Delta x} = -\frac{1}{\epsilon_0} \frac{P_{x2} - P_{x1}}{\Delta x}, \tag{11.88}$$

which yields our first boundary condition:

$$\epsilon_0 E_{x2} + P_{x2} = \epsilon_0 E_{x1} + P_{x1}. \tag{11.89}$$

Let us now work similarly with another Maxwell equation:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}. \tag{11.90}$$

Decomposing in components, the above encompasses three equations. For example,

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t} \tag{11.91}$$

contains no large $\frac{\partial}{\partial x}$ derivatives and it yields no boundary condition. However,

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\partial B_y}{\partial t} \tag{11.92}$$

yields

$$\frac{\partial E_z}{\partial x} = 0 \rightsquigarrow E_{z1} = E_{z2}. \tag{11.93}$$

Similarly, we obtain

$$E_{y1} = E_{y2}. \tag{11.94}$$

From the remaining Maxwell's equations, we obtain (**exercise**) the boundary condition:

$$\vec{B}_2 = \vec{B}_1. \tag{11.95}$$

In summary, we have obtained the following boundary conditions for the electromagnetic field:

$$\vec{B}_1 = \vec{B}_2 \tag{11.96}$$

$$\vec{E}_{1,\parallel} = \vec{E}_{2,\parallel} \tag{11.97}$$

$$(\epsilon_0 \vec{E}_1 + \vec{P}_1)_\perp = (\epsilon_0 \vec{E}_2 + \vec{P}_2)_\perp, \tag{11.98}$$

where $\perp$ denotes the component perpendicular to the boundary plane and $\parallel$ the components parallel to it. *Remark.* All components of $\vec{B}$ are continuous here because $\vec{M} = 0$ and there are no surface currents. For magnetic media only $B_\perp$ and $\vec{H}_\parallel$ are continuous, and in the presence of free surface charges or currents $D_\perp$ and $\vec{H}_\parallel$ jump (see Ref. [2], Sec. I.5).

### 11.9.1 Snell's law

Let us now assume that an electromagnetic plane wave $(\vec{E}_I, \vec{B}_I)$ approaches the boundary from medium $n_1$. Experience tells us that there will be a reflected $(\vec{E}_R, \vec{B}_R)$ and a transmitted electromagnetic field $(\vec{E}_T, \vec{B}_T)$. The electric and magnetic fields on the two sides of the boundary will be:

$$\vec{E}_1 = \vec{E}_I + \vec{E}_R, \quad \vec{E}_2 = \vec{E}_T, \tag{11.99}$$

and

$$\vec{B}_1 = \vec{B}_I + \vec{B}_R, \quad \vec{B}_2 = \vec{B}_T, \tag{11.100}$$

with

$$\vec{E}_I = \hat{e}_I E_I e^{i\left(\omega_I t - \vec{k}_I \cdot \vec{x}\right)}, \tag{11.101}$$

$$\vec{E}_R = \hat{e}_R E_R e^{i\left(\omega_R t - \vec{k}_R \cdot \vec{x}\right)}, \tag{11.102}$$

$$\vec{E}_T = \hat{e}_T E_T e^{i\left(\omega_T t - \vec{k}_T \cdot \vec{x}\right)}, \tag{11.103}$$

and

$$\vec{B}_I = \frac{\vec{k}_I \times \vec{E}_I}{\omega_I}, \quad \vec{B}_R = \frac{\vec{k}_R \times \vec{E}_R}{\omega_R}, \quad \vec{B}_T = \frac{\vec{k}_T \times \vec{E}_T}{\omega_T}. \tag{11.104}$$

The electric and magnetic fields are perpendicular to the direction of propagation:

$$\vec{k}_I \cdot \hat{e}_I = \vec{k}_R \cdot \hat{e}_R = \vec{k}_T \cdot \hat{e}_T = 0. \tag{11.105}$$

For the magnitudes of the wave vectors $\vec{k}$ we have

$$\frac{k_I}{\omega_I} = \frac{k_R}{\omega_R} = \frac{n_1}{c}, \quad \frac{k_T}{\omega_T} = \frac{n_2}{c}. \tag{11.106}$$

From the boundary condition,

$$\vec{E}_{1,\parallel} = \vec{E}_{2,\parallel},$$

we have that on the boundary

$$\vec{x} = (0, y, z) \tag{11.107}$$

the electric fields satisfy

$$\hat{e}_{I,\parallel} E_I e^{i\left(\omega_I t - \vec{k}_{I,\parallel} \cdot \vec{x}\right)} + \hat{e}_{R,\parallel} E_R e^{i\left(\omega_R t - \vec{k}_{R,\parallel} \cdot \vec{x}\right)} = \hat{e}_{T,\parallel} E_T e^{i\left(\omega_T t - \vec{k}_{T,\parallel} \cdot \vec{x}\right)}, \tag{11.108}$$

where $\vec{k}_\parallel = (0, k_y, k_z)$. Let us choose to evaluate Eq. 11.108 at

$$\vec{x} = (0, 0, 0). \tag{11.109}$$

Then, we obtain:

$$\hat{e}_{I,\parallel} E_I e^{i\omega_I t} + \hat{e}_{R,\parallel} E_R e^{i\omega_R t} = \hat{e}_{T,\parallel} E_T e^{i\omega_T t}. \tag{11.110}$$

Let's assume that we arrange for the above equation to be satisfied at $t = 0$. If the frequencies $\omega_I, \omega_R, \omega_T$ are different from each other, then the three terms of the equation will change differently at an arbitrary later time $t > 0$ and the equation will not be satisfied any longer. We thus conclude that the frequencies of the transmitted and reflected light are the same as the frequency of the incident light:

$$\omega_I = \omega_R = \omega_T = \omega. \tag{11.111}$$

This is a physical prediction: reflection and refraction in a linear medium do not change the colour of monochromatic light.

Similarly, requiring Eq. 11.108 to hold at $t = 0, \vec{x} = (0, 0, z)$ and at $t = 0, \vec{x} = (0, y, 0)$ yields

$$\vec{k}_{I,\parallel} = \vec{k}_{R,\parallel} = \vec{k}_{T,\parallel} = \vec{k}_\parallel. \tag{11.112}$$

From the above, we conclude that the incident, the reflected and the transmitted waves propagate in the same plane (see Fig. 11.2). Assuming that the indices of refraction $n_1, n_2$

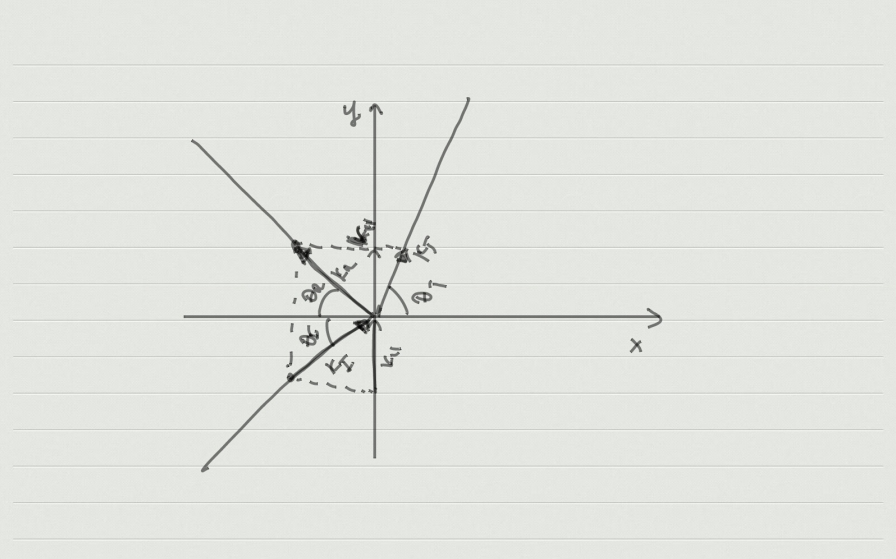



**Figure 11.2.** The incident, reflected and transmitted waves and their angles $\theta_I$, $\theta_R$, $\theta_T$ with respect to the normal to the boundary.

are real, Eq. 11.112 yields a relation for the incident, reflected and transmitted wave angles:

$$k_I \sin\theta_I = k_R \sin\theta_R = k_T \sin\theta_T. \tag{11.113}$$

Recalling Eq. 11.106, we obtain:

$$n_1 \sin\theta_I = n_1 \sin\theta_R = n_2 \sin\theta_T. \tag{11.114}$$

We conclude that the incident and reflected wave angles are equal,

$$\theta_I = \theta_R, \tag{11.115}$$

and that the transmitted wave angle is:

$$\sin\theta_T = \frac{n_1}{n_2} \sin\theta_I. \tag{11.116}$$

Eq. 11.116 is the familiar law of Snell in optics. Geometrical optics obtains the same laws from Fermat's principle: light travels between two points along the path of least (more precisely, stationary) time. Here they came out of Maxwell's equations. Both descriptions are limits of a more fundamental theory; in Chapter 12 we will see how quantum electrodynamics explains why light appears to take the path of least time.

### 11.9.2 Polarisation

Let us align the $y$-axis so that

$$\vec{k}_{\|} = k_{\|}\hat{y}. \tag{11.117}$$

The electric and magnetic fields are perpendicular to the wave vector $\vec{k}$. We distinguish two possibilities:

- $\vec{E}_I \parallel \hat{z}$ and $\vec{B}_I$ lies in the $x$–$y$ plane;
- $\vec{B}_I \parallel \hat{z}$ and $\vec{E}_I$ lies in the $x$–$y$ plane.

All other polarisations can be calculated as a superposition of these two cases.

For $\hat{e}_I = \hat{z}$, Eq. 11.108 gives

$$E_I + E_R = E_T. \tag{11.118}$$

Also, from

$$\vec{B}_1 = \vec{B}_2 \rightsquigarrow B_{I,y} + B_{R,y} = B_{T,y} \rightsquigarrow k_{I,x}E_I + k_{R,x}E_R = k_{T,x}E_T. \tag{11.119}$$

The two equations yield for the reflected and transmitted waves:

$$E_T = \frac{2k_{I,x}}{k_{I,x} + k_{T,x}} E_I, \tag{11.120}$$

and

$$E_R = \frac{k_{I,x} - k_{T,x}}{k_{I,x} + k_{T,x}} E_I, \tag{11.121}$$

where we have used that

$$k_{R,x} = -k_{I,x}.$$

**Exercise:** Calculate $E_T, E_R$ in terms of $\theta_I, \theta_T$.

**Exercise:** Calculate $E_T, E_R$ for the polarisation $\vec{B} \parallel \hat{z}$.

## 11.10 Cherenkov radiation

In a dielectric medium, like air or water, Maxwell's equations take approximately the same form as in the vacuum where we substitute

$$c \to c_m = \frac{c}{\sqrt{1+\chi}}, \quad \epsilon_0 \to \epsilon = \epsilon_0(1+\chi). \tag{11.122}$$

Therefore, the solutions of Maxwell's equations for the four-vector potential $A^\mu$ in the vacuum are also solutions of Maxwell's equations in a dielectric, as long as we adjust the constants $c, \epsilon_0$ appropriately.

In the vacuum, all particles propagate with a velocity smaller than the speed of light. However, in a medium electrically charged particles can have a velocity which is larger than the speed of light in that medium:

$$c_m < v < c.$$

Our solutions of Maxwell's equations have assumed so far that $v < c_m$. It is interesting to see what $A^\mu(x)$ is in the case where some currents $j^\mu$ contain charged particles with

$$v > c_m.$$

Consider a single charge $q$ travelling with a uniform velocity $v > c_m$. In units of $c_m = 1, \epsilon = 1$, the solution of Maxwell's equations takes the form:

$$A^\mu(x) = \int d^4x' G(x-x') J^\mu(x'), \tag{11.123}$$

where

$$G(x-x') = \frac{1}{2\pi}\delta\left((x-x')^2\right)\Theta(x^0 - x'^0), \tag{11.124}$$

and

$$J^\mu(x') = q\frac{dx'^\mu}{dt'}\delta(\vec{x}' - \vec{v}t'). \tag{11.125}$$

Writing

$$x^\mu \equiv (t, \vec{x}), \quad x'^\mu \equiv (t', \vec{x}'),$$

we have

$$A^\mu(x) = \frac{(q, q\vec{v})}{2\pi} \int dt'\, \delta\left((t-t')^2 - \left|\vec{x} - \vec{v}t'\right|^2\right) \Theta(t > t'). \tag{11.126}$$

Let us set

$$\vec{R} \equiv \vec{x} - \vec{v}t,$$

and

$$\delta t \equiv t - t'.$$

As usual, we now find the roots of the equation obtained by requiring the argument of the delta function to vanish:

$$\begin{aligned} 0 &= (t-t')^2 - \left|\vec{x} - \vec{v}t'\right|^2 \\ &= \delta t^2 - \left|\vec{R} + \vec{v}\delta t\right|^2 \\ &= \delta t^2(1-v^2) - 2Rv\cos\theta\delta t - R^2, \end{aligned} \tag{11.127}$$

where $\theta$ is the angle formed by $\vec{R}$ and $\vec{v}$:

$$\vec{R}\cdot\vec{v} = Rv\cos\theta.$$

The discriminant of the quadratic polynomial in Eq. 11.127 is

$$\Delta = 4R^2v^2\cos^2\theta + 4(1-v^2)R^2 = 4R^2v^2\left[\frac{1}{v^2} - \sin^2\theta\right]. \tag{11.128}$$

For $v < 1$, the discriminant is positive and the delta-function constraint always has a real solution. In our case, $v > 1$, we have a solution only when

$$-\frac{1}{v} < \sin\theta < \frac{1}{v} \tag{11.129}$$

(see Fig. 11.3). Outside this interval, there are no solutions for the delta-function constraint and the potential is zero. The solutions of the quadratic equation are

$$\delta t = \frac{Rv}{v^2-1}\left[-\cos\theta \pm \sqrt{\frac{1}{v^2} - \sin^2\theta}\right]. \tag{11.130}$$

For $v > 1$, the square root in the bracket is smaller in magnitude than the first term $-\cos\theta$ in the bracket. The Theta-function constraint $\Theta(x^0 > x'^0)$ requires that $\delta t > 0$. This can only occur for

$$\cos\theta < 0 \rightsquigarrow \theta > \frac{\pi}{2}. \tag{11.131}$$

In summary, the vector potential is non-zero only in a cone trailing the charged particle, defined by the angle

$$\theta > \frac{\pi}{2}, \quad -\frac{1}{v} < \sin\theta < \frac{1}{v} \tag{11.132}$$

(see Fig. 11.4, and this visualisation of Cherenkov radiation). Computing the value of

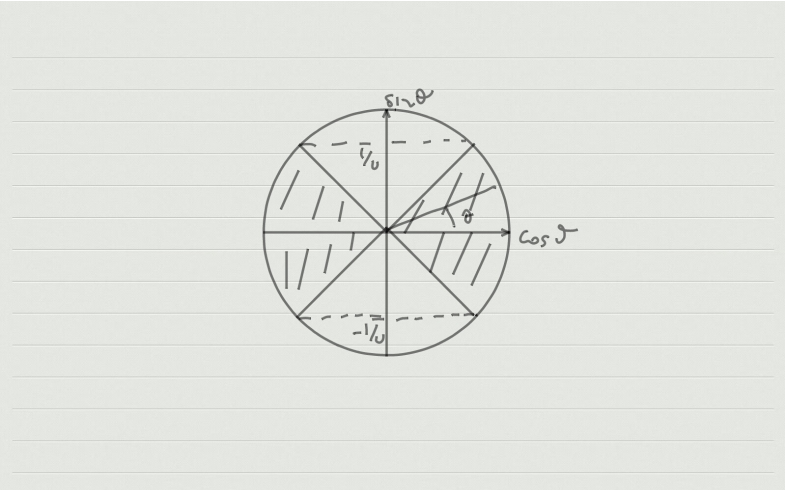



**Figure 11.3.** A charge moving with velocity $\vec{v}$, $v > 1$ (in units of $c_m$), and the angle $\theta$ between $\vec{R}$ and $\vec{v}$.

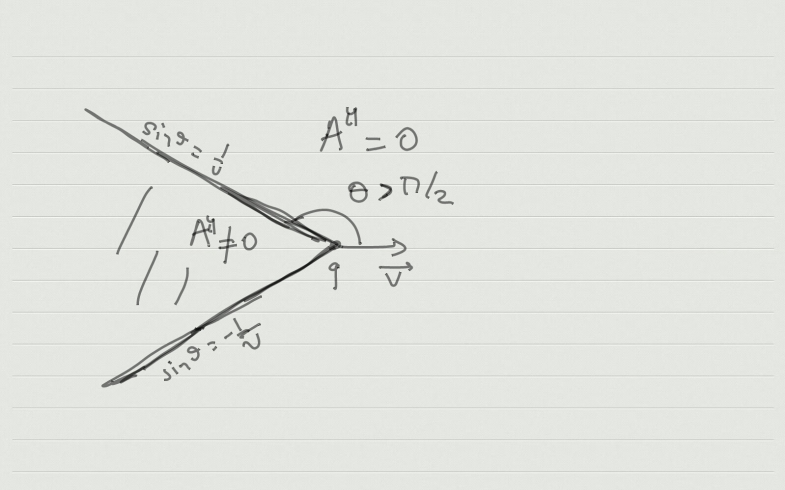



**Figure 11.4.** The Cherenkov cone trailing a charge which moves faster than the speed of light in the medium. The potential is non-zero only inside the cone.

the potential inside the cone is straightforward (both roots $\delta t_\pm > 0$ contribute, each with weight $1/|2R\sqrt{1 - v^2 \sin^2\theta}|$), yielding:

$$A^\mu = \frac{(q, q\vec{v})}{2\pi R\sqrt{1 - v^2 \sin^2\theta}}. \tag{11.133}$$

The potential changes abruptly as we move from the inner side of the cone, where it is non-zero, to the outside of the cone, where it is zero. Its gradient, and hence the electromagnetic field, is therefore very large across the boundary of the cone. This is the Cherenkov radiation associated with the superluminal motion of charged particles in a dielectric material. *Remark.* With a frequency-independent $n$ the field on the cone is singular (the potential diverges as $v \sin\theta \to 1$). The finite Cherenkov spectrum (Frank–Tamm formula) needs dispersion: radiation occurs only at frequencies with $n(\omega)\, v > c$, at angle $\cos\theta_C = c/(nv)$ to the velocity (see Ref. [2], Sec. 13.4).

In 2011, a lot of excitement was created by a mistaken experimental measurement of the OPERA collaboration, which gave a velocity larger than the speed of light in the vacuum for neutrinos travelling from CERN in Geneva to their neutrino detector in Gran Sasso in Italy. Theorists were sceptical about this measurement, which did not fit the expectations from Cherenkov radiation phenomena. Although neutrinos do not have electric charge, they do have "weak-force" charge and they are expected to slow down as they travel by emitting Cherenkov "weak-radiation". The neutrinos detected in Gran Sasso were measured to be more energetic, as if they were not losing energy during their travel in contradiction to expectations. The measurement turned out to be wrong, and when it was repeated correctly the velocity of the neutrinos was consistent with $v < c$.
*Remark.* More precisely, Cohen and Glashow (arXiv:1109.6562) showed that superluminal neutrinos would lose energy mainly through $\nu \to \nu e^+ e^-$ pair emission (a Cherenkov-like process), depleting the high-energy part of the beam. The energy spectrum seen at Gran Sasso showed no such depletion, and ICARUS found no excess of pairs (arXiv:1110.3763). The OPERA anomaly was eventually traced to a faulty fibre-optic connection.

# 12

# A glimpse of quantum electrodynamics

We close the course with a look at what electrodynamics becomes when we combine it with quantum mechanics. We will not derive anything in full here; we follow the spirit of Feynman's little book *QED: The Strange Theory of Light and Matter*, and try to see how the classical phenomena of the previous sections (reflection, refraction, scattering of light) emerge from a few simple quantum rules.

We have obtained the laws of reflection and refraction in two ways: from Maxwell's equations, and from Fermat's principle of least time on which geometrical optics is based. Which is the right description of light? Neither is exact. Both are limits of quantum electrodynamics, and they work so well because, in the conditions in which we usually observe light, the quantum theory reduces to them. We will also see why.

## 12.1 Probability amplitudes

Electrodynamics at a fundamental level is simple. There are three simple rules for how electrons and photons move and interact. The rules are *probabilistic*: we do not know what those particles do. We can only compute the probability that something happens,

$$P = |\mathcal{M}(\text{event})|^2 \,. \tag{12.1}$$

The complex number $\mathcal{M}$ is the *probability amplitude* of the event. Feynman draws it as an arrow in the plane: its length squared is the probability, its direction is the phase.

How is $\mathcal{M}$ constructed? With simple "composition" rules. We define an event as a transition from an "initial state" to a "final state". For example, a photon goes from $A$ to $B$; or a photon and two electrons transform into three electrons and one positron; etc.

In classical physics, if an event can happen in two alternative ways, $A \xrightarrow{(1)} B$ or $A \xrightarrow{(2)} B$, we calculate the probability as

$$P = P_1 + P_2. \tag{12.2}$$

Also, an event can happen in a succession of steps, first $(A \to C)$ and then $(C \to B)$, and then

$$P = P_1 \cdot P_2. \tag{12.3}$$

In quantum mechanics we have instead:

$$\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 \qquad \text{for alternative ways,} \tag{12.4}$$

$$\mathcal{M} = \mathcal{M}_1 \cdot \mathcal{M}_2 \qquad \text{for successive steps.} \tag{12.5}$$

For alternatives,

$$P = |\mathcal{M}_1 + \mathcal{M}_2|^2 = P_1 + P_2 + \underbrace{2\,\mathrm{Re}\left(\mathcal{M}_1^* \mathcal{M}_2\right)}_{\text{interference}} \neq P_1 + P_2. \tag{12.6}$$

The interference term can be positive or negative. Two alternatives can reinforce each other, or cancel each other completely. Everything that follows is an application of this simple observation.




alternatives: add head to tail successive steps: multiply

**Figure 12.1.** Amplitudes as arrows. Alternatives are added as vectors in the complex plane; for successive steps the lengths are multiplied and the phases (angles) are added.

## 12.2 The three basic actions

What are the simplest events whose composition gives all the varied types of events we observe in nature? There are three types of basic events involving electrons and photons. Both are particles and exhibit similar quantum behaviour.

### 12.2.1 A photon goes from one space-time point to another

A photon goes from $x \to y$ if $x^0 > y^0$, or a photon goes from $y$ to $x$ if $y^0 > x^0$. (Here $x, y$ are space-time points: a place and a time.) We draw it as a wavy line, and its amplitude is

$$x,\mu \sim\sim\sim\sim\sim\sim\sim\sim y,\nu \;=\; \int \frac{d^4k}{(2\pi)^4}\,\frac{e^{-ik\cdot(x-y)}\left(-g_{\mu\nu}\right)}{k^2 + i\delta}, \qquad \delta \to 0. \tag{12.7}$$

The factor $-g_{\mu\nu}$ accounts for the photon polarisation (we will ignore this). The largest contribution comes from $k^2 = 0$, i.e.

$$E_{\vec{k}}^2 - \vec{k}^2 = 0 \rightsquigarrow E_{\vec{k}} = |\vec{k}| \qquad \text{(on-shell | massless)}. \tag{12.8}$$

Notice that

$$\Box_x \bullet\sim\sim\sim\sim\sim\sim\bullet \;=\; g_{\mu\nu}\,\delta^{(4)}(x-y), \tag{12.9}$$

since $\Box\, e^{-ik\cdot x} = -k^2 e^{-ik\cdot x}$ and $\int d^4k\, e^{-ik\cdot(x-y)}/(2\pi)^4 = \delta^{(4)}(x-y)$. The photon amplitude is a Green's function for solving Maxwell's equations! Indeed, in the Lorentz gauge ($\xi = 1$ in Eq. 8.67), Maxwell's equations read $\Box A_\mu = J_\mu$, and the solution is $A_\mu(x) = \int d^4y\, D_{\mu\nu}(x-y) J^\nu(y)$, with $D_{\mu\nu}$ the amplitude of Eq. 12.7; compare with the Green's functions of Eq. 6.15.

*Remark.* In the usual conventions of quantum field theory the propagators carry an extra factor of $i$ (e.g. $-ig_{\mu\nu}/(k^2 + i\delta)$ for the photon). Such overall constant factors do not matter for the qualitative discussion of this section.

Let's now observe what happens when $k^2 \approx 0$ in the exponent. With $E_{\vec{k}} = |\vec{k}|$ and $\theta$ the angle between $\vec{k}$ and $\vec{x} - \vec{y}$,

$$e^{-ik\cdot(x-y)} \approx e^{-iE_{\vec{k}}\left[(x^0-y^0)-|\vec{x}-\vec{y}|\cos\theta\right]}. \tag{12.10}$$

The bracket describes propagation with the speed of light. Photons travel mostly with the speed of light. There is a probability that they do not, but this is small. The major contribution to the amplitude comes from $\theta \approx 0$, $x^0 - y^0 = |\vec{x} - \vec{y}|$. For all other values the phase oscillates rapidly as we integrate over $\vec{k}$ and the contributions cancel; we will meet the same mechanism again, in a very concrete setting, when we discuss reflection.

### 12.2.2 An electron goes from one space-time point to another

The electron propagates in a similar way:

$$x \longrightarrow y \quad = \int \frac{d^4k}{(2\pi)^4} \frac{(\not{k}+m_e\mathbf{1})_{4\times4}}{k^2 - m_e^2 + i\delta} e^{-ik\cdot(x-y)}, \tag{12.11}$$

where the $4\times4$ matrix in the numerator ($\not{k} \equiv \gamma^\mu k_\mu$, with $\gamma^\mu$ the Dirac matrices) is due to the electron spin. Otherwise, it is very similar. Now most of the amplitude originates from

$$k^2 \approx m_e^2 \qquad \text{(massive electron)}. \tag{12.12}$$

This is the relation $E^2 = \vec{p}^{\,2} + m_e^2$ of a free relativistic particle (with $c = 1$); the electron, too, mostly moves as a classical free particle would, but not always.

### 12.2.3 An electron absorbs or emits a photon

The third basic event is that electrons absorb or emit photons (Fig. 12.2):

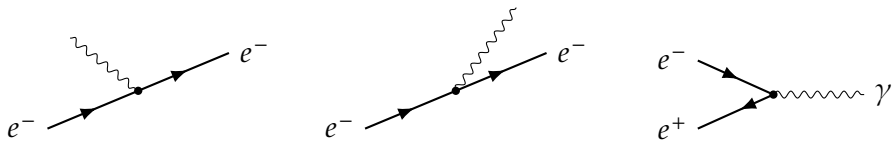



**Figure 12.2.** Absorption and emission of a photon by an electron, and electron–positron annihilation into a photon.

(One more twist of the same rule: electrons and positrons annihilate and give photons.) The amplitude for the electron-photon interaction is a constant:

$$\longrightarrow\!\!\!\!\!\!\sim \quad = -ie\gamma^\mu, \tag{12.13}$$

where the matrix $\gamma^\mu$ takes care of spin and polarisation. In the following qualitative description, I will ignore spin and polarisations. Here $e$ is the electric charge of the electron. In appropriate natural units,

$$\frac{e^2}{4\pi} = \alpha \approx \frac{1}{137.036} \approx 0.7\% \ll 1, \tag{12.14}$$

the *fine-structure constant*. Restoring $\hbar$ and $c$ (in Heaviside–Lorentz units), $\alpha = e^2/(4\pi\hbar c)$. The smallness of $\alpha$ is the reason why diagrams with few vertices dominate.

Let's see how these rules explain partial reflection and diffraction of light.

## 12.3 Thomson and Rayleigh scattering

Consider photon-electron scattering. With time running upwards and space horizontally, the amplitude is a sum of the alternatives shown in Fig. 12.3:

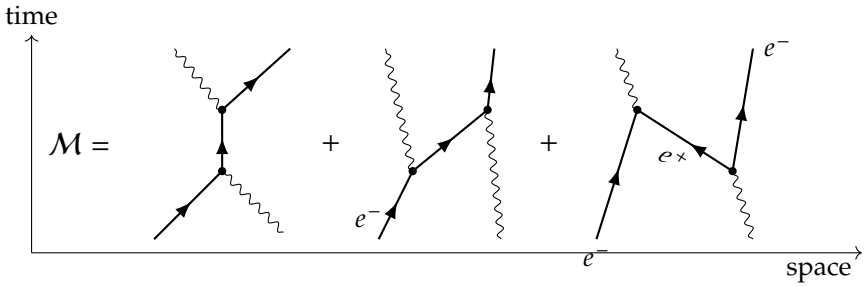



**Figure 12.3.** The three alternatives contributing to photon–electron (Thomson) scattering. Time runs upwards.

This is Thomson scattering. In the first diagram the electron absorbs the incoming photon and later emits the outgoing one; in the second it first emits the outgoing photon and later absorbs the incoming one. In the third diagram we meet the positron: the incoming photon creates an electron-positron pair, and the positron later annihilates with the incoming electron into the outgoing photon. The middle line describes either an $e^-$ travelling from the future to the past or an $e^+$ travelling forward in time. Both alternatives are part of the single amplitude of Eq. 12.11, which does not ask which time is earlier. For photon energies $\hbar\omega \ll m_e c^2$, the sum of these diagrams reproduces the classical, frequency-independent Thomson cross-section of Eq. 10.12.

In the first diagram, the intermediate electron cannot be an ordinary free electron. For an electron at rest, $p^\mu = (m_e, \vec{0})$, which absorbs a photon $k^\mu = (E_\gamma, \vec{k})$ with $|\vec{k}| = E_\gamma$, momentum conservation gives for the intermediate electron $p'^\mu = (m_e + E_\gamma, \vec{k})$ and

$$p'^2 = m_e^2 + 2m_e E_\gamma \neq m_e^2.$$

It is "off its mass shell", a *virtual* particle: this is the region $k^2 \neq m_e^2$ of the propagator of Eq. 12.11, which contributes less than the classical region but does not vanish. Virtual particles are never observed directly. They exist only between interactions, for a time short enough to be compatible with the uncertainty principle; the particles which reach our detectors satisfy $p^2 = m^2$.

In an atom the photon scatters in a more complicated process, Rayleigh scattering (Fig. 12.4):

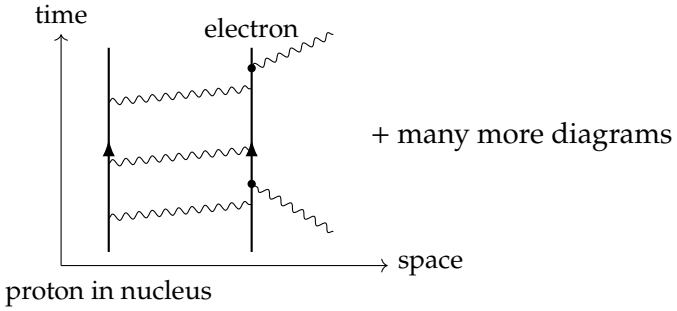



**Figure 12.4.** Rayleigh scattering of a photon on an electron bound in an atom.

The amplitude for Rayleigh scattering depends on the atom or molecule, and it is different in various materials. We do not attempt to compute it; we collect all these diagrams into a single blob, a "Rayleigh scattering amplitude",

$$\mathcal{M}_R = ? \tag{12.15}$$

It determines the reflection/refraction properties of the material! Classically, this is the bound-electron model of Eq. 10.20: for $\omega \ll \omega_0$ the cross-section grows as $\omega^4$, Eq. 10.22. In what follows we only need that $\mathcal{M}_R$ is some (small) complex number, the same for every atom of the glass.

## 12.4 Partial reflection of light by glass

Consider a layer of glass. A photon which arrives at the glass is either reflected or transmitted (Fig. 12.5):

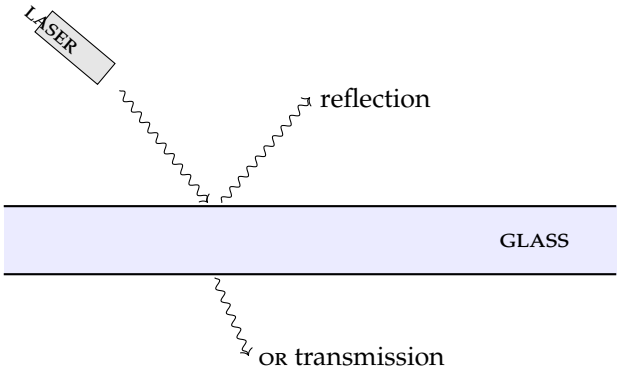



**Figure 12.5.** A photon from a laser is either reflected by or transmitted through a layer of glass.

Only one of the two (transmission or reflection) happens. Let us analyse all steps of this process.

This is where the quantum and the classical descriptions differ observably. For a classical wave there is always a reflected and a transmitted wave, and detectors on both sides of the glass respond. Now dim the laser until it emits single photons. Each photon is registered either by the detector of the reflected light or by that of the transmitted light, never by both; the probabilities only tell us how often each of the two happens. Single photons can be detected (the dark-adapted human eye responds to a handful of them), and such experiments confirm the quantum picture.

Let us consider monochromatic light (a laser): $E_\gamma = \hbar\omega$, so $\omega$ (the colour) is fixed. But $(\Delta E)(\Delta t) \sim \hbar$; we cannot know the emission time precisely. The amplitude

$$\mathcal{M}(\hbar\omega \text{ being emitted at time } t \text{ before detection}) \sim e^{i\omega t} \tag{12.16}$$

depends on the time $t$. (*Remark.* Whether we write $e^{i\omega t}$ or $e^{-i\omega t}$ is a convention, compare with the factor $e^{-iE(x^0-y^0)}$ above; only the relative phases between alternatives are measurable. In Feynman's language, the arrow turns like the hand of a stopwatch, $\omega/2\pi$ times per second, while the photon travels.)

### 12.4.1 The steps of the process

First, light travels to the glass, from the source $S$ to a point $G$ of the glass:

$$\mathcal{M}(S \to G) \sim \mathcal{M}(S \to G)_{\text{on-shell}} \sim 1, \tag{12.17}$$

since the photon travels (mostly) with the speed of light. Then, at $G$, Rayleigh scattering takes place on an electron, with amplitude $\mathcal{M}_R$ (the sum of the electron-photon diagrams, collected into one blob). Finally, the photon travels from $G$ to the detector $D$,

$$\mathcal{M}(G \to D) \sim \mathcal{M}(v_{\text{light}} \sim c) \sim 1 \tag{12.18}$$

(see Fig. 12.6).

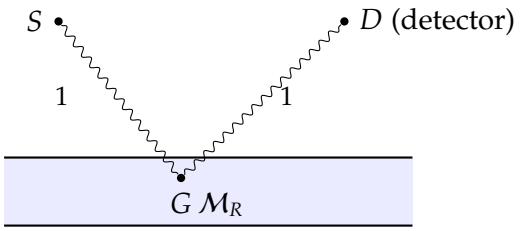



**Figure 12.6.** Reflection at a point $G$ of the glass: the photon travels from the source $S$ to $G$, scatters there, and travels on to the detector $D$. The amplitude of each step is indicated.

Notice that the reflected photon is not the incident photon bouncing off the glass: the incident photon is absorbed by an electron and a new photon is emitted. The photon may also scatter two or more times inside the glass before reaching the detector, but every additional scattering costs a further small factor, and we neglect such processes.

### 12.4.2 Total amplitude

The amplitude for reflection at a point $G$ in the glass is the product of the amplitudes of the successive steps,

$$\begin{aligned} \mathcal{M}_{\text{REFLECTION}}(\text{at a point } G) &= \mathcal{M}_{\text{emission at S}}(t) \cdot \mathcal{M}(S \to G) \\ &\quad \cdot \mathcal{M}_R(\text{at } G) \cdot \mathcal{M}(G \to D) \\ &\sim \mathcal{M}_R \, e^{i\omega t}, \end{aligned} \tag{12.19}$$

where $t$ is the time that light needed in total to travel along its path. All the information about the geometry is in the phase: the magnitude is the same for every point $G$, but the phase depends on the length of the path $S \to G \to D$.

The photon could have been scattered anywhere. Alternatives add, so

$$\mathcal{M}_{\text{REFLECTION (COMPLETE)}} = \sum_{\text{all points}} \mathcal{M}_{\text{REFLECTION}}(\text{point in glass}), \tag{12.20}$$

where the sum runs over all points where Rayleigh scattering may have occurred.

## 12.5 Summing over the surface: stationary phase

Let's first sum up the contributions from the surface of the glass. We place the source $S$ and the detector $D$ at a height $h$ above the surface, at a horizontal distance $2d$ from each other, and let $x$ be the position of the scattering point on the surface, measured from the point midway between them (Fig. 12.7).

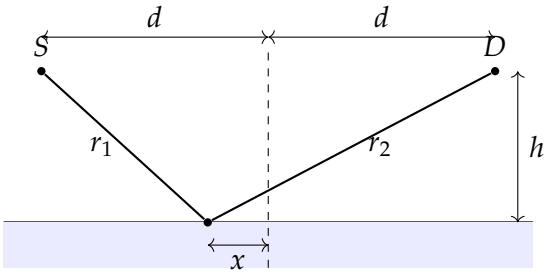



**Figure 12.7.** Geometry of the reflection from the surface of the glass.

Then

$$\begin{aligned}\mathcal{M}_{\text{SURF}} &= \int_{-\infty}^{+\infty} dx\, \mathcal{M}(x) = \int_{-\infty}^{+\infty} dx\, \mathcal{M}_R\, e^{i\omega t(x)} \\ &= \mathcal{M}_R \int_{-\infty}^{+\infty} dx\, e^{\frac{i\omega}{c}\left(\sqrt{(d-x)^2+h^2}+\sqrt{(d+x)^2+h^2}\right)}. \end{aligned} \tag{12.21}$$

(The surface is two-dimensional; the integration in the direction perpendicular to the plane of the figure works in exactly the same way, and we suppress it.)

The real part of the integrand is shown in Fig. 12.8.

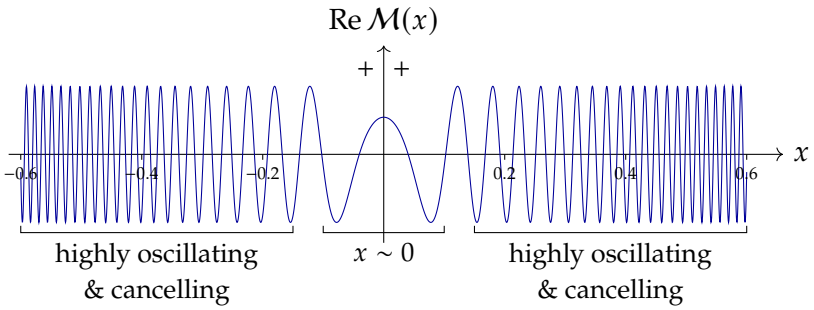



**Figure 12.8.** Real part of the integrand of Eq. 12.21 for $\omega/c = 1000$ and $d = h = 1$, in arbitrary units of length.

Away from $x \approx 0$ the integrand is highly oscillating and the positive and negative contributions cancel. Let us make a steepest-descent (here: stationary-phase) approximation of this integral. Expanding the path length around $x = 0$, with $r_0 \equiv \sqrt{h^2 + d^2}$,

$$\sqrt{(d-x)^2+h^2} + \sqrt{(d+x)^2+h^2} = 2r_0 + \frac{h^2}{r_0^3}\, x^2 + O(x^4),$$

the linear terms cancel: $x = 0$ is the path of *least time*, Fermat's principle, and the angle of incidence equals the angle of reflection. Using $\int_{-\infty}^{\infty} dx\, e^{iax^2} = \sqrt{i\pi/a}$, we find

$$\begin{aligned}\mathcal{M}_{\text{SURF}} &\approx \mathcal{M}_R\, e^{2i\frac{\omega}{c}\sqrt{h^2+d^2}} \int_{-\infty}^{+\infty} dx\, e^{i\frac{\omega}{c}\frac{h^2}{(h^2+d^2)^{3/2}}x^2+\ldots} \\ &= \mathcal{M}_R\, e^{2i\frac{\omega}{c}\sqrt{h^2+d^2}} \sqrt{\frac{i\pi\,(h^2+d^2)^{3/2}}{(\omega/c)\,h^2}}. \end{aligned} \tag{12.22}$$

This gives most of the amplitude. We can check this numerically. We integrate Eq. 12.21 only over a piece $-X < x < X$ of the surface,

$$\mathcal{M}_{\text{SURF}}(X) = \mathcal{M}_R \int_{-X}^{X} dx\, e^{\frac{i\omega}{c}\left(\sqrt{(d-x)^2+h^2}+\sqrt{(d+x)^2+h^2}\right)}, \tag{12.23}$$

and compare with the steepest descent result of Eq. 12.22 (Fig. 12.9). Once $X$ is larger than the stationary region, $\mathcal{M}_{\text{SURF}}(X)$ only oscillates, with a shrinking amplitude, around the steepest descent value: adding more and more of the glass surface changes almost nothing.

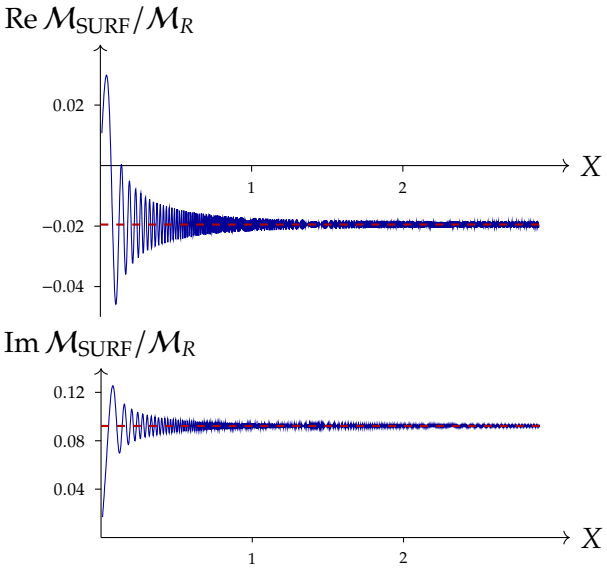



**Figure 12.9.** Real (left) and imaginary (right) part of the amplitude for reflection from the piece $-X < x < X$ of the surface (solid), compared with the steepest descent approximation of Eq. 12.22 (dashed), for $\omega/c = 1000$, $d = h = 1$.

Indeed, we see that most of the contribution originates from the region $|x| \lesssim \sqrt{\frac{c}{\omega}\frac{(h^2+d^2)^{3/2}}{h^2}}$, which for $d \lesssim h$ is of the order of $\sqrt{\lambda h}$! This region is much larger than a wavelength, but tiny compared to the macroscopic distances $h, d$: for visible light and $h = 1\,\text{m}$ it is less than a millimetre. Only a small piece of the mirror around the point predicted by geometric optics contributes; the rest cancels.

Feynman visualises the integral of Eq. 12.21 by adding the little arrows of all points $x$ head to tail. Fig. 12.10 shows the result: the arrows from the far left and far right curl up in tight spirals and contribute almost nothing; the straight middle part, from the region around $x = 0$, gives the total arrow $\mathcal{M}_{\text{SURF}}$. Its length $\sqrt{\pi}$ (in units of $a = 1$) is exactly the magnitude of Eq. 12.22.

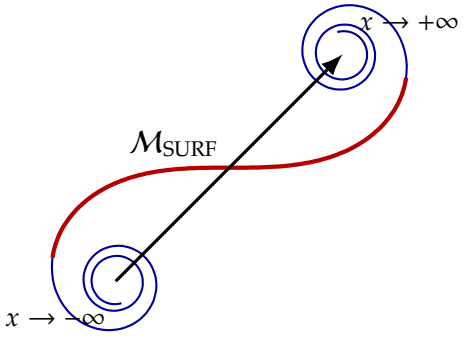



**Figure 12.10.** Adding the arrows $e^{iax^2}dx$ of all points of the surface head to tail (Cornu spiral, here $a = 1$ and $-4 < x < 4$). The thick red part comes from $|x| < 1.2$, around the stationary point; the spirals at the two ends come from the rapidly oscillating regions and cancel.

The far regions of the mirror do contribute, with arrows as long as those of the central region; we do not see their light only because neighbouring strips cancel each other. We can spoil this cancellation. If we paint black the strips whose arrows point in one direction and keep those whose arrows point in the opposite direction, the remaining arrows add up, and we observe light reflected from parts of the mirror far from the geometric reflection point, at angles forbidden by geometrical optics. Such a mirror is a *diffraction grating*. Since the width of the strips depends on the wavelength, a grating sends different colours into different directions. In a crystal, where the scattering electrons sit on a lattice, the integral over $x$ is replaced by a sum over the lattice sites; for a lattice spacing much smaller than the wavelength the result is the same.

## 12.6 Front and back surfaces

Now let us integrate vertically across the layers of the glass. Doing so, we find that it suffices to consider reflection from the front and the back surface of the glass (Fig. 12.11). (Each thin layer inside the glass gives an arrow; the arrows of successive layers turn by a small angle and, just as for the surface, they add up to a spiral whose net contribution comes only from its two ends: the front and the back surface.)

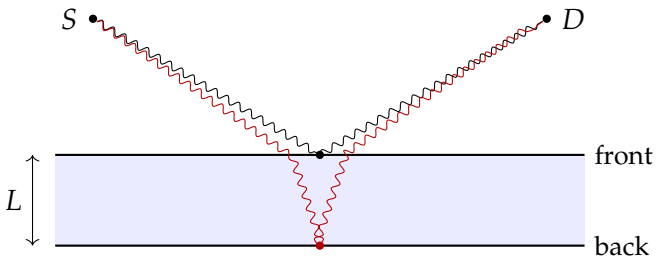



**Figure 12.11.** Reflection from the front and from the back surface of a layer of glass of thickness $L$.

With $L$ the thickness of the glass,

$$\begin{aligned} \mathcal{M} &= \mathcal{M}_{\text{FRONT}} - \mathcal{M}_{\text{BACK}} = e^{it_{\text{front}}\omega} - e^{it_{\text{back}}\omega} \\ &= e^{it_{\text{front}}\omega}\left(1 - e^{i\omega(t_{\text{back}}-t_{\text{front}})}\right). \end{aligned} \tag{12.24}$$

The relative minus sign is a property of $\mathcal{M}_R$ summed over the layers: the front surface (air to glass) reflects with the opposite sign to the back surface (glass to air). Classically this is the reflection coefficient $E_R/E_I = (k_{I,x} - k_{T,x})/(k_{I,x} + k_{T,x})$ of the section on reflection and refraction, which at normal incidence is $(1-n)/(1+n)$ at the front surface and $(n-1)/(n+1)$ at the back. Both arrows have the same length (up to a common factor, which we have dropped), so that

$$\begin{aligned} P(\text{REFL}) &\sim 2\left(1 - \cos\left(\omega\left(t_{\text{back}} - t_{\text{front}}\right)\right)\right) P_{\text{SINGLE SURFACE}}, \\ \omega\left(t_{\text{back}} - t_{\text{front}}\right) &= \frac{2n\omega L}{c}, \end{aligned} \tag{12.25}$$

at normal incidence, where $n$ is the refractive index of the glass (the light travels the extra distance $2L$, down and up, with speed $c/n$). The reflection probability varies with the thickness. For $L \approx 0$ there is no reflection. But for $\cos\left(\frac{2n\omega}{c}L\right) = -1$,

$$P_{\text{REFL}} = 4\,P_{\text{SINGLE SURFACE}}. \tag{12.26}$$

For glass with $n = 1.5$, a single surface reflects $P_{\text{SINGLE SURFACE}} = \left(\frac{n-1}{n+1}\right)^2 = 4\%$ of the light. A sheet of glass therefore reflects anything between 0 and 16%, depending on its thickness (Fig. 12.12); this is the origin of the colours of soap bubbles and oil films, since the condition $\cos(2n\omega L/c) = -1$ depends on $\omega$. Averaged over many thicknesses (or colours) we get 8%, twice the single-surface value, as if the two surfaces reflected independently. (We neglect multiple reflections inside the glass, a small effect of relative size 4%.)

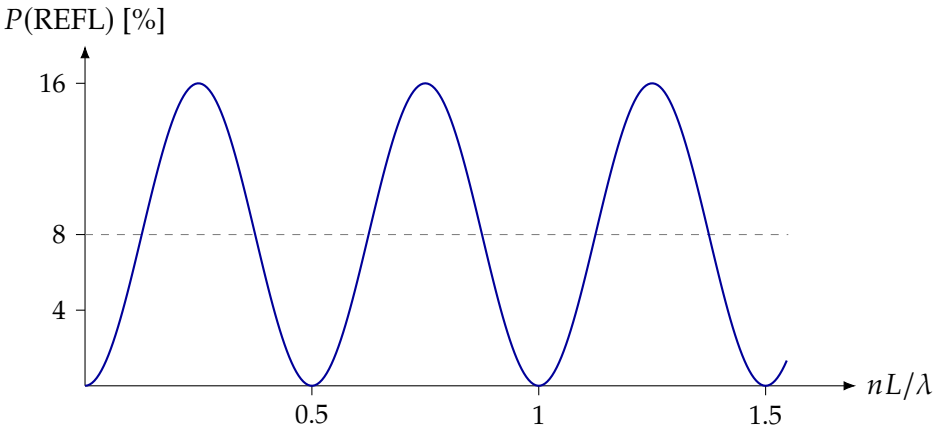



**Figure 12.12.** Reflection probability of a sheet of glass ($n = 1.5$) as a function of its thickness, $P = 2(1 - \cos(4\pi nL/\lambda)) \times 4\%$, at normal incidence.

## 12.7 Transmission

How about transmission? The photon can go through the glass without scattering, or it can scatter in the glass and continue (Fig. 12.13):

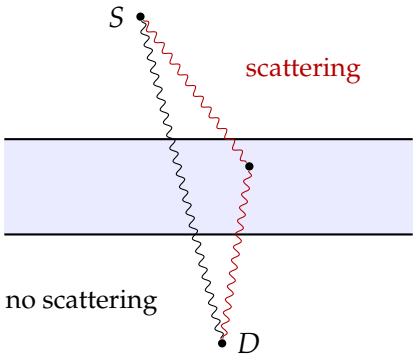



**Figure 12.13.** Transmission through the glass, without and with scattering.

$$P_{\text{TRANSM}} = \left| \mathcal{M}_{0\text{-scattering}} + \int \mathcal{M}_{1\text{-scattering}} \right|^2 , \tag{12.27}$$

where the integral runs over all points of the glass where the scattering may have occurred. This is what gives Snell's law, Eq. 11.116. The denser the glass, the bigger the change in the phase of $\mathcal{M}$ (change of direction). The scattered arrows, added to the unscattered one, turn the total arrow: the light appears to travel more slowly in the glass, with speed $c/n$, and the path of stationary phase between $S$ and $D$ is no longer a straight line but the refracted path of Snell's law. This is the quantum version of the classical picture of Eq. 11.66, where the refractive index arose from the waves re-emitted by the atoms of the medium.

The same rules explain many more phenomena, for example the laser (stimulated emission) and, together with the spin of the electron, the exclusion principle. We do not discuss them here.

## 12.8 Conclusion

Three basic actions (a photon goes from place to place, an electron goes from place to place, an electron emits or absorbs a photon) and two rules (add the arrows of alternatives, multiply the arrows of successive steps) are all we needed. Partial reflection, its oscillation with the thickness of the glass, the law of reflection, Fermat's principle and Snell's law all came out of adding arrows and finding where they do not cancel. The classical wave optics of the previous sections is the limit in which very many photons follow these rules. The same rules, with the precise amplitudes of Eqs. 12.7, 12.11 and 12.13 and the small coupling $\alpha$, give the most precisely tested predictions in all of physics, for example the magnetic moment of the electron.

# Bibliography


[1] R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman Lectures on Physics, Vol. II: Mainly Electromagnetism and Matter*, www.feynmanlectures.caltech.edu.

[2] J. D. Jackson, *Classical Electrodynamics*, 3rd edition, Wiley, New York, 1999.

[3] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Course of Theoretical Physics, Vol. 2, 4th edition, Pergamon Press, Oxford, 1975.

[4] D. J. Griffiths, *Introduction to Electrodynamics*, 4th edition, Cambridge University Press, Cambridge, 2017.

[5] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley, Reading, MA, 1995.

[6] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, Wiley, New York, 1972.