Chapter 1

Theoretical Foundations

1.1 Units

1.1.1 Lengths, masses, times, and temperatures

We use Gaussian centimetre-gram-second (cgs) units throughout. Lengths are measured in cm, masses in grams and time in seconds. The derived units of force, energy and power are listed in Table 1.1. Temperatures are unvariedly measured in Kelvin (K).

Table 1.1 The units of force, energy and power are listed here in the cgs system together with their relations to SI units.

	quantity	cgs unit	alternatives	
force	mass · acceleration	$\frac{\mathrm{g\,cm}}{\mathrm{s}^2}$	dyn	10^{-5} N
energy	mass · velocity^2	$\frac{\mathrm{g\,cm}^2}{\mathrm{s}^2}$	erg	10^{-7} J
power	energy / time	$\frac{\mathrm{erg}}{\mathrm{s}}$		10^{-7} W

The main reason for using these rather than SI units is they allow electromagnetic relations to be expressed in a much easier way, as we shall now discuss.

1.1.2 Charges and electromagnetic fields

The unit of charge is chosen such that the Coulomb force between two charges q separated by the distance r is

$$F_{\mathrm{Coulomb}} = \frac{q^2}{r^2} \; . \tag{1.1}$$

With this choice, the dielectric constant of the vacuum, ε_0, becomes dimensionless and unity. Electric and magnetic fields are defined to have the same unit. This is most sensible in view of the fact that they are both related, and can be converted into each other, by Lorentz transforms. Their unit is chosen such that the force caused by an electric field E on a charge q is

$$F_{\text{electric}} = qE \ . \tag{1.2}$$

This implies that charge, electric and magnetic fields must have the units given in Table 1.2. The squared electric or magnetic field strengths then have the dimension of an energy density.

?

Confirm the cgs units of charge and electric or magnetic fields listed in Tab. 1.2.

Table 1.2 This table lists the units of charge, electric and magnetic field in the Gaussian cgs system, their physical dimensions, and alternative units.

quantity		cgs unit	alternative
charge	force$^{1/2}\cdot$ length	$\dfrac{\mathrm{g}^{1/2}\mathrm{cm}^{3/2}}{\mathrm{s}}$	esu
electric or magnetic field	force / charge	$\dfrac{\mathrm{g}^{1/2}}{\mathrm{cm}^{1/2}\,\mathrm{s}}$	Gauss

By definition, the units of charge in the SI and the Gaussian cgs systems are related by

$$1\,\text{Coulomb} = 2.9979\cdot 10^{9}\,\text{esu} \ . \tag{1.3}$$

Electrostatic potential differences, or electrostatic potential energy changes per unit charge, are measured in Volts in SI units. Consequently, we must have

$$1\,\text{Volt} = 1\,\frac{\text{Joule}}{\text{Coulomb}} = \frac{10^{7}\,\text{erg}}{2.9979\cdot 10^{9}\,\text{esu}} = \frac{1}{299.79}\,\frac{\mathrm{g}^{1/2}\,\mathrm{cm}^{1/2}}{\mathrm{s}} \ . \tag{1.4}$$

The energy gained by a unit charge moving through an electrostatic potential difference of 1 Volt, defined as the electron-Volt, must then be

$$1\,\text{eV} = 1.6022\cdot 10^{-12}\,\text{erg} \ . \tag{1.5}$$

?

Use the Boltzmann constant k_{B} to convert 1 eV to an equivalent temperature.

1.1.3 Natural constants

The most frequently used natural constants in cgs units are tabulated in Table 1.3.

In addition, some units used in astronomy and astrophysics are listed in Table 1.4.

Caution Note that the light speed is exact by definition of the metre. ◂

1.1.4 Conventions and notation

For the Minkowski metric, we use the signature

$$\eta = \operatorname{diag}(-1,+1,+1,+1) \ . \tag{1.6}$$

We adopt the convention

$$\tilde{f}(k) = \mathcal{F}[f](k) = \int \frac{\mathrm{d}^d k}{(2\pi)^d}\, f(x)\,\mathrm{e}^{-\mathrm{i}k\cdot x} \tag{1.7}$$

Table 1.3 The most frequently used natural constants are tabulated here with their common symbols and their values in cgs units. The values are taken from the Particle Data Group (http://pdg.lbl.gov/, last accessed on Nov. 22, 2020).

quantity	symbol	value in cgs units
light speed	c	$2.9979 \cdot 10^{10}$
elementary charge	e	$4.8032 \cdot 10^{-10}$
electron mass	m_e	$9.1094 \cdot 10^{-28}$
proton mass	m_p	$1.6726 \cdot 10^{-24}$
Boltzmann's constant	k_B	$1.3806 \cdot 10^{-16}$
Newton's constant	G	$6.6743 \cdot 10^{-8}$
Planck's constant	$\hbar$	$1.0546 \cdot 10^{-27}$

Table 1.4 Some units common in astronomy and astrophysics are listed here.

unit	symbol	type	value in cgs units
Solar radius	$R_\odot$	length	$6.9634 \cdot 10^{10}$
astronomical unit	AU	length	$1.4960 \cdot 10^{13}$
light year	ly	length	$9.4607 \cdot 10^{17}$
parsec	pc	length	$3.0857 \cdot 10^{18}$
Earth mass	$M_\oplus$	mass	$5.9724 \cdot 10^{27}$
Jupiter mass	$M_♃$	mass	$1.8990 \cdot 10^{30}$
Solar mass	$M_\odot$	mass	$1.9884 \cdot 10^{33}$
tropical year	y	time	$3.1557 \cdot 10^{7}$
sidereal year	y	time	$3.1558 \cdot 10^{7}$
Solar luminosity	$L_\odot$	energy/time	$3.8460 \cdot 10^{33}$
Jansky	Jy	specific intensity	10^{-23}

for the Fourier transform in d dimensions, and

$$f(x) = \mathcal{F}^{-1}(x) = \int \mathrm{d}^d x\, \tilde{f}(k)\, \mathrm{e}^{\mathrm{i}k\cdot x} \tag{1.8}$$

for its inverse. We use the short-hand notation

$$\int_x := \int \mathrm{d}^d x \quad \text{and} \quad \int_k := \int \frac{\mathrm{d}^d k}{(2\pi)^d} \tag{1.9}$$

for the integrals over coordinates $x \in \mathbb{R}^d$ and wave vectors $k \in \mathbb{R}^d$.

1.2 Lorentz Invariance

This section summarises the concepts of special relativity and their consequences for the structure of space-time and for the dynamics of a particle. Its most important results are the relativistic time dilation (1.36) and the Lorentz contraction (1.40), the addition theorem for velocities (1.42) and the transformation of angles (1.45), the combination of energy and momentum into the momentum four vector (1.63) and the relativistic relations (1.66) and (1.67) between energy, momentum and velocity.

Perhaps it is helpful to begin with the statement that classical physics aims to quantify the behaviour of physical entities in space with time. Point mechanics, for example, studies the trajectories of particles with negligible extension. A trajectory can be quantified by a vector-valued function $\vec{x}(t)$ which assigns a spatial vector $\vec{x}$ to any instant t from a finite or infinite time interval. Field theory describes forces as the effect of fields, which are functions of space and time obeying their own dynamics. Immediately, we are led to the question how we want to identify points in space and instants in time in a quantifiable manner.

This is achieved by a reference frame or a coordinate system. In Newtonian physics, space and time were both assumed to be absolute. A rigid reference frame was assumed to exist which identified each point in space by a triple $\vec{x}$ of real-valued, spatial coordinates, and by a real number t for the time. Having formulated the laws of physics in this absolute frame, the immediate further question arises as to how other frames of reference, or coordinate systems, could be chosen such that those laws would remain valid without changing their form. The answer of Newtonian physics was that the laws of physics are the same in all so-called inertial frames. In slightly different words, the laws of physics were claimed to be invariant under all transformations leading from one inertial frame to another.

A clarifying remark should be in order here before we move on. Notice the perhaps trivial point that not the physical *quantities* are generally assumed to be unchanged under transformations from one inertial frame to another, but the *form of the physical laws* relating them. For example, Newton's second axiom, force is mass times acceleration, is expected to hold in all inertial frames, irrespective of the specific values of the acceleration and the force. In another inertial frame, the values of force and acceleration may and generally will be different, but the statement of the law, force equals mass times acceleration, is expected to remain valid. Valid physical laws are expected to be invariant in this sense. If, in addition, physical quantities can be identified that remain invariant under transformations from one inertial frame to another, such conserved quantities play an important role in the analysis of specific physical systems under consideration. It is thus of central importance for any part of theoretical physics to clearly state which type of transformation should lead from one inertial frame to another.

In a more mathematical language, transformations between inertial frames form groups. Admissible physical laws are those which are invariant under the operation of those groups. The identification of the invariance group of a physical theory is perhaps the most fundamental step in its foundation.

1.2.1 The Special Lorentz Transform

In Newtonian mechanics, inertial frames are related by Galilei transformations. If one inertial frame is given, any Galilei transform turns it into another one. The Galilei transforms form a ten-parameter group of transformations. They contain shifts of the origin in space and time (four parameters), translations with constant velocity (three parameters), and rotations in space (further three parameters, e.g. the Euler angles). Consequences of the Galilei invariance of Newtonian mechanics are the existence of an absolute time and the Galilean addition theorem for velocities.

However, the Galilei invariance of Newtonian mechanics leads to contradictions with experience. The decay of muons sets a prominent example. Myons are leptons comparable to the electron, but with a mass of 105.6 MeV instead of 0.511 MeV. They decay according to

$$\mu \to e^- + \bar{\nu}_e + \nu_\mu \tag{1.10}$$

into electrons and (anti-) neutrinos with a half-life of $\tau_\mu = 1.5 \cdot 10^{-6}\,\mathrm{s}$. Experiments show, however, that the lifetime increases if the muon moves in the laboratory frame with velocities near the speed of light. The electron emitted in the decay has almost light speed, but never exceeds it even if the muon had already moved with almost the speed of light. Clearly, the muon seems to live longer in the laboratory rest frame than in its own rest frame, and the Galilean theorem for adding velocities does not longer apply.

Einstein's theory of Special Relativity replaced the Galilei invariance of Newtonian mechanics by the Lorentz invariance of relativistic physics. Special Relativity grew from the problem that the speed of light c appears as an absolute velocity in Maxwell's vacuum equations of electrodynamics. Einstein radically solved this problem by elevating the postulate to a principle that the speed of light c is a universal constant, independent of the state of motion of the light source relative to the observer. Interestingly, the concepts of absolute space and time underlying Newtonian physics were thus replaced by the concept of an absolute, observer-independent maximal velocity.

Consider now two inertial frames, S and S', moving relative to each other at an arbitrary, constant speed (Figure 1.1). Imaginge a flash of light going off. By the principle of the constant light speed, the wave front of the flash must

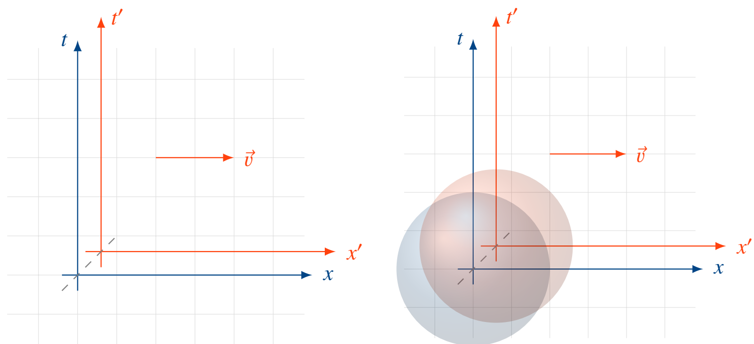

Figure 1.1 *Left*: Two inertial frames are shown moving with constant velocity $\vec{v}$ relative to each other. They are synchronised such that their origins coincide at $t = 0 = t'$. *Right*: A light signal emerging from a source at the common origin of both frames, illustrated by the coloured spheres, propagates in the same way in both frames, despite the relative motion of the two frames.

propagate in the same way in both frames irrespective of their relative velocity and therefore obey the condition

$$\mathrm{d}\vec{x}^2 - c^2\mathrm{d}t^2 = \mathrm{d}\vec{x}'^2 - c^2\mathrm{d}t'^2 \; . \tag{1.11}$$

For definiteness and without loss of generality, we now rotate the coordinate frames S and S' such that they move with respect to each other along their common $\hat{e}_z$ axis, and further set the origin of time such that both frames coincide at $t = 0 = t'$. Requiring further that the transformation between S and S' be linear leads directly to the *special Lorentz transform*

$$x'^3 = \gamma\left(x^3 + \beta ct\right) \; , \quad ct' = \gamma\left(ct + \beta x^3\right) \; , \tag{1.12}$$

where $\beta = v/c$ is the relative velocity in units of the light speed, and the *Lorentz factor*

$$\gamma := \left(1 - \beta^2\right)^{-1/2} \tag{1.13}$$

appears. In the limit of low velocities, $\beta \ll 1$, the Lorentz factor is $\gamma \approx 1 + \beta^2/2$ to second order in β, or $\gamma \approx 1$ to first order. Note that we write the vector indices in (1.12) as superscripts. This may appear arbitrary here, but has a deeper mathematical sense that will shortly be explained.

As (1.12) shows, the time t and the spatial coordinates x^i cannot be uniquely or invariantly separated under special Lorentz transforms. They lose their independent identity and become coupled to each other, depending on the relative motion of the frames in which they are measured. Instead of the rigid Newtonian, Euclidean space-time with its uniquely defined, absolute time axis, we thus need to adopt a four-dimensional space-time with a different structure. We introduce $ct := x^0$ as a further coordinate and combine the coordinate quadruples to *four-vectors* $x = (x^\mu) = (x^0, x^1, x^2, x^3)^\top$. This four-dimensional space with a structure to be clarified below is called Minkowski space $\mathbb{M} = \mathbb{R}^{3+1}$.

The Lorentz transform connects any two inertial frames in the four-dimensional Minkowski space. General Lorentz transforms are composed of special Lorentz transforms in all spatial directions, the so-called Lorentz boosts, plus the orthogonal three-dimensional spatial rotations. Poincaré transforms are general Lorentz transforms combined with arbitrary translations in space and time. Like the Galilei transformations, Poincaré transformations have ten parameters: the three Euler angles for the orthogonal three-dimensional rotations, the four translations, and one velocity for the Lorentz boosts in all three independent spatial directions. In relativistic mechanics, the Poincaré transformations replace the Galilei transformations of Newtonian mechanics.

1.2.2 Minkowski Space

Since Lorentz transforms leave the expression $-(x^0)^2 + \vec{x}^2$ invariant by construction, we define the Minkowskian scalar product between two four-vectors as

$$\langle x, y\rangle = -x^0 y^0 + \vec{x} \cdot \vec{y} = \eta(x, y) \ , \tag{1.14}$$

where $\vec{x} \cdot \vec{y}$ is the ordinary scalar product between two vectors in Euclidean space. The product $\langle \cdot, \cdot \rangle$ is a pseudo-scalar product because it is not positive semi-definite. Based on this scalar product, the Lorentz group as the invariance group of relativistic physics, abbreviated by $O(3, 1)$, can now formally be defined as the set of all linear transforms represented by real-valued, square, 4×4 matrices $\mathcal{M}(4, \mathbb{R})$ that leave the scalar product (1.14) unchanged,

$$O(3, 1) = \{\Lambda \in \mathcal{M}(4, \mathbb{R}) : \langle \Lambda x, \Lambda y\rangle = \langle x, y\rangle \ \forall \ x, y \in \mathbb{M}\} \ . \tag{1.15}$$

This clearly repeats as a mathematical statement that Lorentz transforms are defined as those linear transforms leaving the speed of light invariant.

?

Show that the condition

$$\Lambda^\top \eta \Lambda = \eta$$

is equivalent to $\langle \Lambda x, \Lambda y\rangle = \langle x, y\rangle$.

The object η introduced in (1.14) satisfies the definition of a second-rank tensor, as it is a bilinear map of two vectors from Minkowski space $\mathbb{M}$ into the real numbers,

$$\eta : \mathbb{M} \times \mathbb{M} \to \mathbb{R} \ , \quad (x, y) \mapsto \eta(x, y) = \langle x, y\rangle \ . \tag{1.16}$$

This tensor is the metric tensor of Minkowski space, or the Minkowski metric. Generally, a metric is a second-rank, symmetric tensor which is non-degenerate. This means that if $\langle x, y\rangle = 0$ for all $x \in \mathbb{M}$, then $y = 0$. Once a Cartesian coordinate basis is introduced for Minkowski space, the metric can be represented by the diagonal matrix

$$\left(\eta_{\mu\nu}\right) = \mathrm{diag}(-1, 1, 1, 1) \ , \tag{1.17}$$

which allows us to write the scalar product (1.14) as

$$\langle x, y\rangle = \sum_{\mu,\nu} \eta_{\mu\nu} x^\mu y^\nu \ . \tag{1.18}$$

The subscripted indices introduced here are again not arbitrarily set and will be further illustrated below. By means of the metric, the linear map x^* defined by

$$x^* : \mathbb{M} \to \mathbb{R} \ , \quad y \mapsto x^*(y) = \eta(x, y) = \langle x, y\rangle \tag{1.19}$$

can be introduced on Minkowski space. It maps vectors into the real numbers as shown. The set of all such linear maps forms the dual vector space $\mathbb{M}^*$ to Minkowski space $\mathbb{M}$.

While vector components are identified by upper indices, dual-vector components are written with lower indices. Then, according to

$$\langle x, y\rangle = \sum_{\mu,\nu} \eta_{\mu\nu} x^\mu y^\nu = \sum_{\nu=0}^{4} \left(\sum_{\mu=0}^{4} \eta_{\mu\nu} x^\mu \right) y^\nu \ , \tag{1.20}$$

the dual vector x^* of a four-vector x has the components

$$x_\nu = \sum_{\mu=0}^{4} \eta_{\mu\nu} x^\mu = \left(-x^0, x^1, x^2, x^3\right) \ . \tag{1.21}$$

In Euclidean space, the distinction between vectors and dual vectors is irrelevant because its metric can be represented by the unit matrix. In Minkowski space, it becomes vitally important because of the minus sign of the time-time (or 0-0) component in the metric.

We now introduce Einstein's sum convention in the following form. If an index appears twice in a product and at different levels (i.e. one sub- and one superscripted), a sum over the repeated index is implied. Thus, for example,

$$x_\mu y^\mu = \sum_{\mu=0}^{3} x_\mu y^\mu \ . \tag{1.22}$$

This notation simplifies the previous expressions considerably. Written in components, the scalar product between two vectors x and y simply becomes

$$\langle x, y\rangle = x_\mu y^\mu \ . \tag{1.23}$$

The notation of four-vectors and their dual vectors is made consistent by writing the inverse Minkowski metric η^{-1} with superscripted indices, since then

$$x^\mu = \eta^{\mu\alpha} x_\alpha = \eta^{\mu\alpha} \eta_{\alpha\nu} x^\nu = \delta^\mu_\nu x^\nu \ . \tag{1.24}$$

Thus, we must have

$$\eta^{\mu\alpha} \eta_{\alpha\nu} = \delta^\mu_\nu \ , \tag{1.25}$$

from which we conclude that the matrix representations of the Minkowski metric as well as of its inverse can be brought into the diagonal form

$$(\eta^{\mu\nu}) = \left(\eta_{\mu\nu}\right) = \mathrm{diag}(-1, 1, 1, 1) \ . \tag{1.26}$$

In the notation developed so far, the special Lorentz transform (1.12) can be written as

$$x'^\mu = \Lambda^\mu{}_\nu \, x^\nu \quad \text{with} \quad (\Lambda^\mu{}_\nu) = \begin{pmatrix} \gamma & 0 & 0 & \gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \gamma\beta & 0 & 0 & \gamma \end{pmatrix} . \tag{1.27}$$

Since the Lorentz transform is constructed to leave the Minkowski scalar product invariant, recall (1.15), we must have

$$\eta_{\alpha\beta} x^\alpha x^\beta = \langle x, x\rangle = \langle x', x'\rangle = \eta_{\mu\nu} \Lambda^\mu{}_\alpha \, x^\alpha \Lambda^\nu{}_\beta \, x^\beta = \left(\eta_{\mu\nu} \Lambda^\mu{}_\alpha \, \Lambda^\nu{}_\beta\right) x^\alpha x^\beta \ , \tag{1.28}$$

showing that the Lorentz transform also leaves the Minkowski metric invariant,

$$\eta_{\alpha\beta} = \eta_{\mu\nu}\Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta} \;. \tag{1.29}$$

This relation replaces the perhaps more familiar orthonormality relation in Euclidean space. There, orthonormal transformations R need to satisfy the condition

$$(R\vec{x}) \cdot (R\vec{y}) = \vec{x} \cdot \vec{y} \;, \tag{1.30}$$

which implies the condition $R^\top = R^{-1}$ on matrix representations of R.

?

Compare (1.29) with the condition

$$\Lambda^\top \eta \Lambda = \eta \;.$$

The Minkowskian orthonormality relation (1.29) implies that dual-vector components must transform under Lorentz transformations as

$$x'_{\mu} = \Lambda_{\mu}{}^{\nu} x_{\nu} \;, \tag{1.31}$$

which differs from the transformation (1.27) of vector components. Quantities transforming like vector or dual-vector components under Lorentz transforms are called Lorentz contravariant or covariant, respectively. Quantities unchanged by Lorentz transforms are Lorentz invariant. Vectors are consequently sometimes addressed as contravariant vectors, dual vectors as covariant vectors, which is a terminology which we avoid here because it hides the more fundamental mathematical distinction between vectors and dual vectors (which is also decisively important elsewhere, e.g. in quantum mechanics).

Since the coordinate time becomes largely arbitrary in Special Relativity as it loses any invariant meaning, it needs to be replaced by an invariant measure of time. The only Lorentz-invariant quantity that can be defined to characterise the separation between two space-time points x^μ and $x^\mu + \mathrm{d}x^\mu$ is the so-called line element of the Minkowski metric (1.14),

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \;. \tag{1.32}$$

This line element is interpreted as the so-called proper time $\mathrm{d}\tau$ by the identification

$$\mathrm{d}s^2 = -c^2\mathrm{d}\tau^2 \;. \tag{1.33}$$

This definition is meaningful since the proper time equals the time measured by an observer in his or her own rest frame. In that frame, an observer arbitrarily placed at the spatial coordinate origin has the Minkowski coordinates $(x^0, 0, 0, 0)^\top$. Two subsequent events experienced by that observer at instants of coordinate time separated by $\mathrm{d}x^0$ in the rest frame have the invariant distance

$$c^2\mathrm{d}\tau^2 = \left(\mathrm{d}x^0\right)^2 = c^2\mathrm{d}t^2 \;, \tag{1.34}$$

which shows that the proper time agrees with the coordinate time in any observer's rest frame.

1.2.3 Some Properties of the Minkowski World

We briefly summarise some essential conclusions from the Lorentz covariance of the Minkowski world (see also Figure 1.2). First, let two events happen in the unprimed system S at the same location $\vec{x} = 0$, but with a time difference

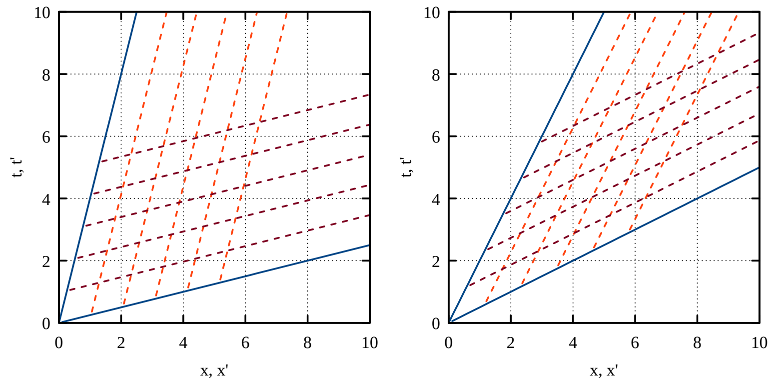

Figure 1.2 Lines of constant t' (dark red) and x' (light red) are shown in the unprimed system S for $\beta = 0.25$ (*left*) and $\beta = 0.5$ (*right*). The lines are inclined with an angle $\arctan(\beta)$ relative to the unprimed axes.

δt or $\delta x^0 = c\delta t$. These events have the four-vectors $x_1 = (0, 0, 0, 0)^\top$ and $x_2 = (\delta x^0, 0, 0, 0)^\top$. By the special Lorentz transform (1.27), they are transformed into the events

$$x_1' = (0, 0, 0, 0)^\top \quad , \quad x_2' = \left(\gamma\delta x^0, 0, 0, \beta\gamma\delta x^0\right)^\top \ . \tag{1.35}$$

Thus, in the primed system S', they are separated by the larger time interval

$$\delta x'^0 = \gamma\delta x^0 \quad \text{or} \quad \boxed{\delta t' = c^{-1}\delta x'^0 = \gamma\delta t} \ . \tag{1.36}$$

This is the relativistic time dilation: *Moving clocks go slow*.

Next, we consider a unit rule oriented in the direction of the relative motion of the two frames and resting in the unprimed system S. Its end points, measured at an arbitrary time $ct = x^0$ in S, are marked by the four-vectors $x_1 = (x_1^0, 0, 0, 0)$ and $x_2 = (x_2^0, 0, 0, 1)$. Now, an observer in S' measures its end points. It is important that he does so at one fixed instant of his coordinate time, which we arbitrarily and without loss of generality set to be $x'^0 = 0$. By (1.27), this requires

$$0 = x'^0 = \gamma x^0 + \beta\gamma x^3 \quad \text{or} \quad x^0 = -\beta x^3 \ . \tag{1.37}$$

For the two end points of our unit rule, this simultaneity condition implies that

$$x_1^0 = 0 \quad \text{and} \quad x_2^0 = -\beta \tag{1.38}$$

since $x_1^3 = 0$ and $x_2^3 = 1$ by construction. The unit rule's end points $x_{1,2}$ appear at

$$x_1' = \left(\gamma x_1^0, 0, 0, \beta\gamma x_1^0\right)^\top \quad , \quad x_2' = \left(\gamma x_2^0 + \beta\gamma, 0, 0, \beta\gamma x_2^0 + \gamma\right)^\top \tag{1.39}$$

in the primed observer's rest frame S'. Inserting (1.38) here gives

$$x_1'^3 = 0 \quad \text{and} \quad \boxed{x_2'^3 = \left(1 - \beta^2\right)\gamma = \gamma^{-1}} \ . \tag{1.40}$$

Thus, in the primed system S', the unit rule turns out to have the length $x_2'^3 - x_1'^3 = \gamma^{-1}$, which is smaller than its unit length in the rest frame. This is the relativistic length contraction: *Moving rods are shorter*.

Let us now consider a particle moving with velocity $\vec{u} = (u_x, u_y, u_z)^\top$ in the unprimed system. Its four vector in S, $x = (x^0, u_x t, u_y t, u_z t)^\top = x^0(1, u_x/c, u_y/c, u_z/c)^\top$, is transformed into

$$x' = x_0 \left(\gamma + \beta\gamma\frac{u_z}{c}, \frac{u_x}{c}, \frac{u_y}{c}, \beta\gamma + \gamma\frac{u_z}{c}\right)^\top \ . \tag{1.41}$$

The velocity components of the particle in the primed system S' are then found to be

$$u'_{x,y} = c\frac{x'^{1,2}}{x'^0} = \frac{u_{x,y}}{\gamma(1 + \beta u_{x,y}/c)} \ , \quad u'_z = c\frac{x'^3}{x'^0} = \frac{v + u_z}{1 + v u_z/c^2} \ . \tag{1.42}$$

The last equation is the relativistic law for the addition of velocities. While the velocity components perpendicular to the relative motion of the two frames S and S' are reduced by the Lorentz factor γ, the velocity component parallel to the motion adds to the relative velocity of the two frames in such a way that the sum of the two velocities u_z and v never exceeds c.

Let the particle now fly with the speed of light into a direction enclosing the angle θ with the $\hat{e}_z$ axis along which the two frames move relative to each other. For simplicity, but without loss of generality, we further rotate both coordinate frames about their common $\hat{e}_z$ axis such that the particle moves in the x-z coordinate plane. Then,

$$u_x = c\sin\theta \ , \quad u_y = 0 \ , \quad u_z = c\cos\theta \ , \tag{1.43}$$

in the unprimed system, and

$$u'_x = c\sin\theta \ , \quad u'_y = 0 \ , \quad u'_z = \frac{v + c\cos\theta}{1 + \beta\cos\theta} \tag{1.44}$$

in the primed system. Since the absolute velocity must also remain $|\vec{u}'| = c$ in the primed frame, the direction of motion in S' is

$$\cos\theta' = \frac{u'_z}{c} = \frac{\beta + \cos\theta}{1 + \beta\cos\theta} \ . \tag{1.45}$$

This is the relativistic aberration of light: Light rays propagating perpendicularly to $\hat{e}_z$ in S enclose an angle $\theta' = \arccos\beta$ with the $\hat{e}'_z$ axis in S'. For non-relativistic velocities, $\beta \ll 1$ and $\cos\theta' \approx \beta + \cos\theta$ to first order in β.

?

Confirm the non-relativistic limit of the relation (1.45).

Consequently, the solid-angle element spanned by light rays also changes due to the relative motion of S' relative to S. As the velocity components perpendicular to the direction of motion are unchanged, so is the azimuthal angle, $\phi' = \phi$ and $\mathrm{d}\phi' = \mathrm{d}\phi$. From the aberration formula (1.45), we have

$$\mathrm{d}\cos\theta' = \frac{\mathrm{d}\cos\theta}{1 + \beta\cos\theta} - \frac{(\beta + \cos\theta)\beta\mathrm{d}\cos\theta}{(1 + \beta\cos\theta)^2} = \frac{\mathrm{d}\cos\theta}{\gamma^2(1 + \beta\cos\theta)^2} \ , \tag{1.46}$$

which implies that the solid-angle element spanned by a light bundle transforms as

$$\mathrm{d}\Omega' = \mathrm{d}\phi'\mathrm{d}\cos\theta' = \frac{\mathrm{d}\phi\mathrm{d}\cos\theta}{\gamma^2(1 + \beta\cos\theta)^2} = \frac{\mathrm{d}\Omega}{\gamma^2(1 + \beta\cos\theta)^2} \ . \tag{1.47}$$

This is relativistic beaming: Isotropic radiation in the unprimed system S attains a highly anisotropic angular distribution in S', pointing strongly into the forward direction (Figures 1.3 and 1.4).

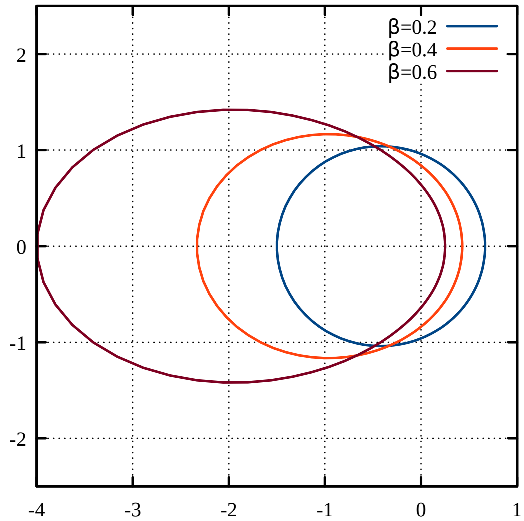

Figure 1.3 Illustration of the relativistic deformation of the solid-angle element $\mathrm{d}\Omega'/\mathrm{d}\Omega$ for the three different velocities $\beta = 0.2, 0.4, 0.6$ as indicated. The curves illustrate how isotropic radiation emitted by a point source resting in the unprimed system S would appear focussed into the direction of motion in the primed system S'.

1.2.4 Relativistic Dynamics

Since the coordinate time has no invariant meaning any more in relativity, the definition of velocity must be changed. The four-velocity is introduced as the derivative of a position four-vector with respect to the invariant proper time τ,

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \ . \tag{1.48}$$

By definition of the proper time in (1.32),

$$\begin{aligned} \mathrm{d}\tau &= c^{-1}\sqrt{-\mathrm{d}s^2} = c^{-1}\sqrt{-\mathrm{d}x^\mu \mathrm{d}x_\mu} = c^{-1}\sqrt{c^2\mathrm{d}t^2 - \mathrm{d}\vec{x}^2} = \mathrm{d}t\sqrt{1-\beta^2} \\ &= \gamma^{-1}\mathrm{d}t \ . \end{aligned} \tag{1.49}$$

Accordingly, the components of the four-velocity are

$$u^\mu = \gamma\left(c, \vec{v}\right)^\top = c\gamma\left(1, \vec{\beta}\right)^\top \ , \tag{1.50}$$

hence its (Minkowski) square is

$$u^2 = \langle u, u\rangle = u^\mu u_\mu = -c^2\gamma^2\left(1-\beta^2\right) = -c^2 \ , \tag{1.51}$$

which is obviously and by construction invariant. Since $\mathrm{d}\tau$ is also invariant, u^μ transforms like the four-vector x^μ under Lorentz transformations, and is thus also a four-vector.

Similarly, the four-momentum of a particle with mass m is defined as

$$p^\mu = m u^\mu = \gamma m c\left(1, \vec{\beta}\right) \ . \tag{1.52}$$

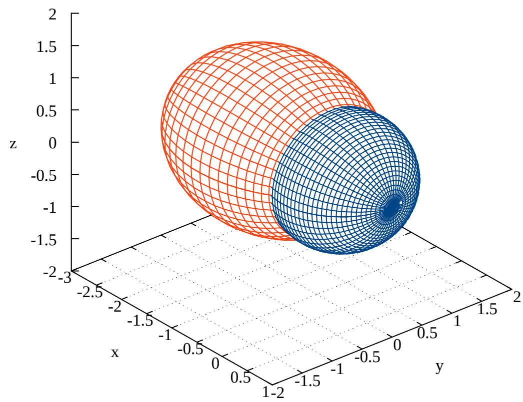

Figure 1.4 The relativistic deformation of the solid angle is shown here in a pseudo-three-dimensional representation. The blue sphere around a source at rest illustrates isotropy. When the source is moving, the sphere surrounding it in its rest frame appears strongly distorted into its forward direction.

Up to second order in β, the zero (time) component of the four-momentum is

$$p^0 = \gamma mc \approx mc\left(1 + \frac{\beta^2}{2}\right) = c^{-1}\left(mc^2 + \frac{m}{2}v^2\right) . \qquad (1.53)$$

Here, the non-relativistic kinetic energy $mv^2/2$ appears together with the rest energy mc^2.

In analogy to classical mechanics, we now search for the action S of a free, relativistic particle, i.e. a particle moving relativistically in absence of external forces. The action must be Lorentz invariant since it must not depend on the arbitrary state of motion of any observer. Therefore, it must only depend on Lorentz scalars characterising a free particle. For a free particle, the only such scalar is the proper time τ, scaled with a constant α to be determined later,

$$S = \alpha \int_a^b \mathrm{d}\tau , \qquad (1.54)$$

where a and b mark the fixed four-dimensional start and end points of the particle's trajectory. The action must have the dimension [energy]·[time]. Since τ has the dimension [time], the constant α must be a constant energy, which we shall determine later.

Writing the action as a function of the coordinate time t, we find

$$S = \alpha \int_{t_a}^{t_b} \mathrm{d}t \sqrt{1-\beta^2} , \qquad (1.55)$$

from which we can identify the Lagrange function

$$L(\vec{x}, \vec{v}, t) = \alpha \sqrt{1-\beta^2} \qquad (1.56)$$

for the free relativistic particle. For non-relativistic motion, $\beta \ll 1$, this must reproduce the Lagrange function of a free particle in Newtonian mechanics,

$$\alpha\sqrt{1-\beta^2} \approx \alpha\left(1 - \frac{\beta^2}{2}\right) = \alpha - \frac{\alpha v^2}{c^2} . \qquad (1.57)$$

Ignoring the irrelevant constant α left as a first term on the right-hand side, the second term shows that $\alpha = -mc^2$ must be chosen to satisfy the limit of non-relativistic mechanics. Accordingly, the action of the free relativistic particle is

$$S = -mc^2 \int_a^b \mathrm{d}\tau \ , \tag{1.58}$$

and its Lagrange function is

$$L = -mc^2 \sqrt{1-\beta^2} \ . \tag{1.59}$$

The Euler-Lagrange equation requires

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \vec{v}} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\gamma mc^2 \frac{\vec{\beta}}{c}\right) = mc\frac{\mathrm{d}}{\mathrm{d}t}\left(\gamma\vec{\beta}\right) = 0 \ , \tag{1.60}$$

which implies $\dot{\vec{\beta}} = 0 = \dot{\vec{v}}$: The free particle moves on a straight line, as expected.

The momentum conjugate to the three-dimensional position vector $\vec{x}$ is

$$\vec{p} = \frac{\partial L}{\partial \vec{v}} = \frac{m\vec{v}}{\sqrt{1-\beta^2}} = \gamma m\vec{v} \ . \tag{1.61}$$

The particle's Hamilton function follows from the Legendre transform

$$H = \vec{v}\cdot\vec{p} - L = \gamma m v^2 + mc^2\sqrt{1-\beta^2} = \gamma mc^2\left(\beta^2 + \frac{1}{\gamma^2}\right) = \gamma mc^2 \ . \tag{1.62}$$

This is to be interpreted as the energy E of the particle. Taking the results (1.61) and (1.62) together and comparing them with the momentum four-vector shows that we can write the latter in the form

$$\boxed{p^\mu = (E/c, \vec{p})^\top} \ . \tag{1.63}$$

This identifies the momentum four-vector with the energy-momentum vector of a relativistic particle. Its Minkowski square is

$$\langle p, p\rangle = p_\mu p^\mu = -\frac{E^2}{c^2} + \vec{p}^2 \ , \tag{1.64}$$

while the equivalent definition $p^\mu = mu^\mu$ implies

$$\langle p, p\rangle = m^2\langle u, u\rangle = -m^2c^2 \ . \tag{1.65}$$

Together, (1.64) and (1.65) form the relativistic energy-momentum relation

$$\boxed{E^2 = c^2\vec{p}^2 + m^2c^4} \ . \tag{1.66}$$

Combining (1.61) and (1.62) finally gives the very useful relation

$$\boxed{\vec{p} = \frac{E}{c^2}\,\vec{v} = \frac{E}{c}\vec{\beta}} \ . \tag{1.67}$$

Let us conclude this section with a remark on energy, momentum and their conservation in relativity. Energy and momentum are conserved if the Lagrange- or Hamilton functions of a system are invariant under translations in time and space, respectively. In relativity, time and space lose their independent existence. Time intervals and spatial distances can at least partially be transformed into each other, depending on the observer's state of motion relative to the system considered. Therefore, separate energy-momentum conservation cannot retain an invariant meaning in relativistic mechanics, and must be combined to the joint energy-momentum conservation.

Problems

1. Recall the mathematical definitions of a group, a field, a vector space, a scalar product, a dual vector space, and a tensor.

2. Write down the transformations of time $t \to t'$ and position $\vec{x} \to \vec{x}'$ under Galilei transformations.

3. Which of the following quantities are Lorentz invariant?

$$\vec{x}^2\,, \quad x_\mu x^\mu\,, \quad x^\mu x^\nu\,, \quad \eta_{\mu\nu}\,, \quad \mathrm{d}s^2\,, \quad \left(\mathrm{d}x^0\right)^2\,, \quad \gamma\,, \quad \mathrm{d}\tau^2 \tag{1.68}$$

4. Compute the following expressions:

$$\partial_\alpha x^\mu\,, \quad \partial_\alpha x_\mu\,, \quad \partial_\alpha \langle x, x\rangle = \partial_\alpha\left(x^\mu x_\mu\right)\,. \tag{1.69}$$

5. Light rays are described by their wave vector $k^\mu = (\omega/c, \vec{k})$, where $\vec{k}$ is the three-dimensional wave vector pointing into the propagation direction of the light ray and satisfying the vacuum dispersion relation $\omega = ck$ with the frequency ω.

   (a) Compute the (Lorentz-invariant) scalar product of the wave vector k^μ and an arbitrary four-velocity u^μ. Explain why the frequency measured by an observer moving with four-velocity u^μ is

$$\omega_{\mathrm{obs}} = -\langle u, k\rangle = -u_\mu k^\mu\,. \tag{1.70}$$

   (b) Comparing two observers, one at rest and one moving with respect to the first with velocity $\vec{v}$, derive the relativistic Doppler relation

$$\frac{\omega'}{\omega} = \frac{1 - \vec{n}\cdot\vec{\beta}}{\sqrt{1-\beta^2}}\,, \tag{1.71}$$

   where $\vec{\beta} = \vec{v}/c$ and $\vec{n} = \vec{k}/k$.

   (c) The four-momentum of a particle is $p^\mu = mu^\mu$, where the four-velocity

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \tag{1.72}$$

   is the derivative of the coordinates x^μ with respect to the proper time τ. Starting from the relativistic Hamilton function

$$H = \frac{1}{2m} p_\mu p^\mu\,, \tag{1.73}$$

   of a free particle, derive the equations of motion and show that its Lagrange function is

$$L = \frac{m}{2} u_\mu u^\mu\,. \tag{1.74}$$

6. Beginning with the defining condition

$$\Lambda^\top \eta \Lambda = \eta \quad \text{with} \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.75}$$

   for the Lorentz transform in two dimensions,

(a) argue why an angle ψ must exist such that

$$\Lambda(\psi) = \begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix} . \tag{1.76}$$

(b) Define $\beta = \tanh\psi$ and show that $\cosh\psi = \gamma$ and $\sinh\psi = \beta\gamma$.

(c) Show that $\Lambda(\psi_1)\Lambda(\psi_2) = \Lambda(\psi_1 + \psi_2)$. Use this result to derive the relativistic law for adding velocities.

1.3 Electromagnetism

This section summarises the foundations and some important results of classical electrodynamics. The theory is motivated as the only Lorentz invariant, linear theory for six field components that satisfies Coulomb's force law. Maxwell's equations are derived in covariant form from the appropriate action and solved by means of the retarded Greens function. The general formalism for the energy-momentum tensor of a field theory is introduced and applied to the electromagnetic field. From the Liénard-Wiechert potentials, Larmor's formula is derived in relativistic form, and the covariant expression for the Lorentz force is derived from the action. The main results are Maxwell's equations themselves, most compactly expressed in Lorenz gauge by the wave equation (1.100), the energy-momentum tensor (1.110) for the electromagnetic field, the Liénard-Wiechert potentials (1.117), the relativistic Larmor formula (1.138), its solid-angle integrated version (1.141) and its non-relativistic approximation (1.143), and finally the relativistic expression (1.147) for the Lorentz force.

1.3.1 Field Tensor and Sources

Electromagnetism is a classical field theory with six degrees of freedom, namely the three components each of the electric and magnetic fields $\vec{E}$ and $\vec{B}$. Fields are functions of space and time. Since special relativity teaches us that space and time are not independent, any field theory must explicitly be constructed to agree with the space-time structure of special relativity. The electromagnetic field must thus be expressed as a four-vector or a tensor field. Obviously, a four vector is not sufficient to describe six degrees of freedom. The simplest object available is a rank-2 tensor, which offers 16 independent components in its most general form. A symmetric rank-2 tensor in four dimensions still has ten independent components, while an antisymmetric rank-2 tensor has exactly the required six degrees of freedom. The simplest possibility to describe six degrees of freedom with a Lorentz-covariant object in four dimensions is thus provided by an antisymmetric field tensor F of rank two, whose components must satisfy

$$F^{\mu\nu} = -F^{\nu\mu} , \quad F_{\mu\nu} = -F_{\nu\mu} . \tag{1.77}$$

The antisymmetry is most conveniently ensured expressing the components of F as derivatives of a four-potential A with components

$$A^{\mu} = \begin{pmatrix} \Phi \\ \vec{A} \end{pmatrix} , \tag{1.78}$$

where Φ is the ordinary scalar potential and $\vec{A}$ is the three-dimensional vector potential. The components of the rank-$(2,0)$ field tensor are then written in the manifestly antisymmetric form

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \ . \tag{1.79}$$

They can be conveniently summarised as

$$(F^{\mu\nu}) = \begin{pmatrix} 0 & \vec{E}^\top \\ -\vec{E} & \mathcal{B} \end{pmatrix} \tag{1.80}$$

where the matrix

$$\mathcal{B}_{ij} = \varepsilon_{ija} B^a \tag{1.81}$$

is formed from the components of the magnetic field. The fields themselves are thus given by

$$\vec{E} = -\frac{1}{c}\dot{\vec{A}} - \vec{\nabla}\Phi \ , \quad \vec{B} = \vec{\nabla} \times \vec{A} \ . \tag{1.82}$$

Caution As usual, ε_{ijk} is the totally antisymmetric Levi-Civita symbol, defined such that $\varepsilon_{ijk} = 0$ if any two of its indices are equal and ε_{ijk} is the signature of the permutation of the indices (ijk). ◀

Given our signature $(-,+,+,+)$ of the Minkowski metric, the associated rank-$(0,2)$ tensor has the components

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -\vec{E}^\top \\ \vec{E} & \mathcal{B} \end{pmatrix} \ . \tag{1.83}$$

?

Convince yourself that (1.80) and (1.83) are correct.

The source of the electromagnetic field is the four-current density j which has the components

$$(j^\mu) = \begin{pmatrix} \rho c \\ \vec{j} \end{pmatrix} \ , \tag{1.84}$$

where ρ is the charge density and $\vec{j}$ is the three-dimensional current density. Charge conservation is expressed by the vanishing four-divergence of the four-current,

$$\partial_\mu j^\mu = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0 \ . \tag{1.85}$$

1.3.2 Lorentz transform of the electromagnetic field

Changing from one inertial frame to another moving with a velocity $\vec{v} = c\vec{\beta}$ with respect to the original frame, the field tensor is Lorentz transformed according to

$$F'^{\mu\nu} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta F^{\alpha\beta} \ . \tag{1.86}$$

Orienting both coordinate frames such that their $\hat{e}_z$ axes coincide with the direction of relative motion, the special Lorentz transform is represented by the matrix given in (1.26), and (1.86) gives the following transformation rules for the electric and magnetic field components:

$$\begin{aligned} E'_x &= \gamma(E_x + \beta B_y) \ , \quad E'_y = \gamma(E_y - \beta B_x) \ , \quad E'_z = E_z \ , \\ B'_x &= \gamma(B_x - \beta E_y) \ , \quad B'_y = \gamma(B_y + \beta E_x) \ , \quad B'_z = B_z \ . \end{aligned} \tag{1.87}$$

?

Confirm the transformation equations (1.87) for the electric- and magnetic-field components

While the field components in the direction of motion remain unchanged, the transverse components are enhanced by the Lorentz factor γ. In particular, a purely electric or magnetic field in one frame obtains a magnetic or electric component in the other, moving frame, repectively. It is, however, not possible to transform a purely electric field into a purely magnetic field or vice versa. This is easily understood because the Lorentz transform must keep all Lorentz invariants unchanged that can be formed from the field tensor. These invariants can be written as

$$F_{\mu\nu}F^{\mu\nu} = 2\left(\vec{B}^2 - \vec{E}^2\right) \;, \quad (*F)_{\mu\nu}F^{\mu\nu} = -4\vec{E}\cdot\vec{B} \;, \tag{1.88}$$

Caution The Hodge dual field tensor is obtained from the field tensor by replacing $\vec{E} \to \vec{B}$ and $\vec{B} \to -\vec{E}$,

$$(*F^{\mu\nu}) = \begin{pmatrix} 0 & \vec{B}^\top \\ -\vec{B} & -\mathcal{E} \end{pmatrix}$$

with $\mathcal{E}_{ij} = \varepsilon_{ija}E^a$. ◀

?

Can you confirm Eqs. (1.88)?

where $*F$ is the (Hodge-) dual field tensor. Any Lorentz transform must thus conserve $(\vec{E}^2 - \vec{B}^2)$ and $\vec{E}\cdot\vec{B}$. Starting with $\vec{B} = 0$ in one inertial frame first of all demands that $\vec{E}'$ and $\vec{B}'$ must remain perpendicular to each other in any inertial frame. By the invariance of $(\vec{E}^2 - \vec{B}^2)$, a complete conversion of a purely electric to a purely magnetic field would require

$$\vec{E}^2 = -\vec{B}'^2 \;, \tag{1.89}$$

which is only possible in the trivial case $\vec{E} = 0 = \vec{B}'$ because $\vec{E}^2$ and $\vec{B}'^2$ are positive definite otherwise.

One remark on the transformation formula (1.86) may be in order to avoid confusion. In Euclidean space, a transformation R from one coordinate frame to another changes the matrix representation of a tensor T according to

$$T' = RTR^{-1} = RTR^\top \tag{1.90}$$

if R is orthogonal, $R^{-1} = R^\top$. Although the matrix representation (1.26) of the Lorentz transform does not satisfy this relation, the Lorentz transform is still orthogonal in the sense that it leaves (Minkowski) scalar products invariant, just as orthogonal transformations in Euclidean space leave the Euclidean scalar product unchanged; see also the discussion of this issue in Sect. 1.1.2 above. For this reason, (1.86) remains valid for Lorentz transformations.

1.3.3 Maxwell's Equations

The dynamical equations of a field theory are the Euler-Lagrange equations applied to a Lagrange density which, for a linear theory like electrodynamics, must satisfy three conditions: It must be Lorentz invariant, it must contain at most quadratic terms in the field quantities to ensure a linear theory, and it must reproduce the Coulomb force law in the case of electrodynamics. The only Lagrangian that satisfies these criteria is

$$\mathcal{L} = \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} - \frac{1}{c}j_\mu A^\mu \;, \tag{1.91}$$

where the constants must be chosen such as to reproduce the measured coupling strength of the electromagnetic field to matter. The otherwise perfectly legitimate term $A_\mu A^\mu$ is excluded because it would give the electromagnetic field an effective mass and thus violate the Coulomb force law.

Since the field tensor depends on A^μ only through derivatives, it is invariant under the gauge transformation

$$A^\mu \to A^\mu + \partial^\mu\chi \;, \tag{1.92}$$

where χ is an arbitrary function of all four coordinates x^μ. At first sight, the Lagrangian (1.91) appears to violate gauge invariance, but Gauss' law applied to the action

$$S = \int \mathrm{d}^4x\,\mathcal{L} \tag{1.93}$$

shows that charge conservation (1.85) ensures gauge invariance.

Maxwell's equations are now the Euler-Lagrange equations

$$\partial^\nu \frac{\partial\mathcal{L}}{\partial(\partial^\nu A^\mu)} - \frac{\partial\mathcal{L}}{\partial A^\mu} = 0 \tag{1.94}$$

of the Lagrangian (1.91). They turn out to be

$$\partial_\nu F^{\mu\nu} = \frac{4\pi}{c} j^\mu\ , \tag{1.95}$$

which are four inhomogeneous equations. Since the field tensor is antisymmetric, it identically satisfies the equation

$$\partial_{[\alpha}F_{\beta\gamma]} = 0\ , \tag{1.96}$$

which represents the homogeneous Maxwell equations. For $\alpha = 0$, $(\beta,\gamma) = (1,2)$, $(1,3)$ and $(2,3)$, the homogeneous equations (1.96) give

$$\boxed{\dot{\vec{B}} + c\vec{\nabla}\times\vec{E} = 0}\ , \tag{1.97}$$

while we find

$$\boxed{\vec{\nabla}\cdot\vec{B} = 0} \tag{1.98}$$

for $\alpha = 1$, $(\beta,\gamma) = (2,3)$. Setting $\mu = 0$ and $\mu = i$, the inhomogeneous equations (1.95) give

$$\boxed{\vec{\nabla}\cdot\vec{E} = 4\pi\rho\ , \quad c\vec{\nabla}\times\vec{B} - \dot{\vec{E}} = 4\pi\vec{j}}\ , \tag{1.99}$$

respectively.

With the definition (1.79) of the field tensor in terms of the four-potential and with the Lorenz gauge condition $\partial_\mu A^\mu = 0$, the inhomogeneous equations (1.95) can be cast into the form

$$\boxed{\Box A^\mu = -\frac{4\pi}{c} j^\mu}\ , \tag{1.100}$$

where $\Box = -\partial_0^2 + \vec{\nabla}^2$ is the d'Alembert operator. The particular solution of this inhomogeneous wave equation is given by the convolution of the source with the retarded Greens function

$$G(t-t', \vec{x}-\vec{x}') = \frac{1}{\left|\vec{x}-\vec{x}'\right|}\delta_\mathrm{D}\left(t - t' - \frac{\left|\vec{x}-\vec{x}'\right|}{c}\right)\ , \tag{1.101}$$

i.e. by

$$A^\mu(t,\vec{x}) = \frac{1}{c}\int \mathrm{d}^3x' \int \mathrm{d}t' G\left(t-t', \vec{x}-\vec{x}'\right)\, j^\mu\left(t', \vec{x}'\right)\ . \tag{1.102}$$

The Greens function (1.101) has an intuitive meaning (Figure 1.5). Its first factor, proportional to the inverse distance between the observer and the source,

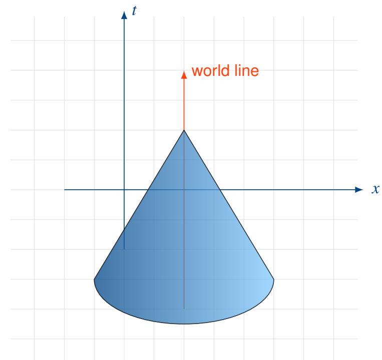

Figure 1.5 Illustration of the geometrical meaning of the retarded Green's function: All signals received from the observer on the red world line at a given instant of time must originate from the backward light cone ending at that time.

expresses Coulomb's force law, which is an immediate consequence of photons being massless. If photons had a mass, the Greens function would have a Yukawa shape with an exponential cut-off. The second factor, the delta function, shows that only such sources can influence the potential at the observer whose world lines intersect with the observer's backward light cone.

Since the Greens function is defined as

$$\Box G\left(t-t', \vec{x}-\vec{x}'\right) = -4\pi\delta_{\mathrm{D}}\left(t-t', \vec{x}-\vec{x}'\right) \;, \tag{1.103}$$

it represents any component of the four-potential created by a point source on the backward light cone of the observer. The convolution (1.102) assembles the complete four-potential by superposition of all contributing sources. This is possible only because electromagnetism is a linear field theory.

1.3.4 Energy-Momentum Conservation

A field theory with a Lagrangian $\mathcal{L}(q, \partial_\nu q)$ for a single field q and its derivatives $\partial_\nu q$ has the energy-momentum tensor

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu q)}\partial_\nu q - \mathcal{L}\delta^\mu_\nu \;, \tag{1.104}$$

which simply corresponds to the Legendre transformation leading from the Lagrange to the Hamilton function in classical mechanics. Should the expression (1.104) turn out to be asymmetric, it needs to be symmetrised to ensure the symmetry of the energy-momentum tensor. For the electromagnetic field, any A^γ can take the role of q, thus

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\gamma)}\partial_\nu A_\gamma - \mathcal{L}\delta^\mu_\nu \;. \tag{1.105}$$

With the Lagrange density of the free electromagnetic field,

$$\mathcal{L} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} \;, \tag{1.106}$$

this implies the energy-momentum tensor

$$T^{\mu}{}_{\nu} = \frac{1}{4\pi}\left(F^{\mu\lambda}F_{\nu\lambda} - \frac{1}{4}F^{\alpha\beta}F_{\alpha\beta}\delta^{\mu}_{\nu}\right) \tag{1.107}$$

of the electromagnetic field. From the representations (1.80) and (1.83) of the field tensor, we find first

$$F^{\mu\lambda}F_{\nu\lambda} = \begin{pmatrix} -\vec{E}^2 & -\left(\vec{E}\times\vec{B}\right)^{\top} \\ \vec{E}\times\vec{B} & -E_iE_j + \delta_{ij}\vec{B}^2 - B_iB_j \end{pmatrix} \tag{1.108}$$

and confirm

$$F^{\alpha\beta}F_{\alpha\beta} = 2\left(\vec{B}^2 - \vec{E}^2\right) . \tag{1.109}$$

Thus, the energy-momentum tensor can be written as

$$T^{\mu}{}_{\nu} = \frac{1}{4\pi}\begin{pmatrix} -(\vec{E}^2+\vec{B}^2)/2 & \left(\vec{E}\times\vec{B}\right)^{\top} \\ -\vec{E}\times\vec{B} & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \bar{T}_{ij} \end{pmatrix} , \tag{1.110}$$

where

$$\bar{T}_{ij} = \frac{1}{4\pi}\left[\left(\frac{1}{2}\vec{E}^2\delta_{ij} - E_iE_j\right) + \left(\frac{1}{2}\vec{B}^2\delta_{ij} - B_iB_j\right)\right] \tag{1.111}$$

are thew components of Maxwell's stress tensor, whose magnetic part will become important in magnetohydrodynamics. The energy density of the electromagnetic field is

$$\varepsilon = T_{00} = \frac{\vec{E}^2 + \vec{B}^2}{8\pi} . \tag{1.112}$$

The energy-momentum tensor satisfies the conservation equation

$$\partial_{\nu}T^{\mu\nu} = 0 \tag{1.113}$$

which, for $\mu = 0$, returns the continuity equation

$$\frac{\partial\varepsilon}{\partial t} + \vec{\nabla}\cdot\vec{S} = 0 \tag{1.114}$$

for the energy density, where the Poynting vector

$$\vec{S} = \frac{c}{4\pi}\vec{E}\times\vec{B} \tag{1.115}$$

represents the energy-current density of the electromagnetic field.

?

Carry out all calculations leading to the results (1.110) and (1.111) for the energy-momentum tensor of the electromagnetic field.

1.3.5 Liénard-Wiechert Potentials and the Larmor Formula

A particle with charge q on a trajectory $\vec{r}_0(t)$ has the current density

$$j^{\mu} = q\begin{pmatrix} c \\ \vec{v} \end{pmatrix}\delta_{\mathrm{D}}\left[\vec{r} - \vec{r}_0(t)\right] . \tag{1.116}$$

When inserted into the convolution (1.102) with the retarded Greens function, this yields the Liénard-Wiechert potentials

$$\Phi(\vec{r}) = \frac{q}{R\left(1 - \hat{e}\cdot\vec{\beta}\right)} , \quad \vec{A}(\vec{r}) = \frac{q\vec{\beta}}{R\left(1 - \hat{e}\cdot\vec{\beta}\right)} = \Phi\vec{\beta} , \tag{1.117}$$

where the right-hand sides have to be evaluated at the retarded time

$$t' = t - \frac{R}{c} \,. \tag{1.118}$$

The vector $\vec{R} \equiv \vec{r} - \vec{r}_0(t')$ points from the retarded particle position to the observer, $R = |\vec{R}|$, and $\hat{e}$ is the unit vector in $\vec{R}$ direction,

$$\hat{e} = \frac{\vec{R}}{R} \,. \tag{1.119}$$

?

Can you confirm expressions (1.117) for the Liénard-Wiechert potentials of a point charge?

The fields $\vec{E}$ and $\vec{B}$ are obtained as the usual derivatives of Φ and $\vec{A}$, but it must be taken into account that the potentials are expressed in retarded coordinates, while we need the derivatives with respect to the observer's coordinates. The spatial derivatives of Φ are

$$\partial_i \Phi = -\frac{q}{\left(R - \vec{R}\cdot\vec{\beta}\right)^2}\left(\partial_i R - \beta_j \partial_i R_j - R_j \partial_i \beta_j\right) \,. \tag{1.120}$$

While the first two terms decrease $\propto R^{-2}$, the third decreases $\propto R^{-1}$. Aiming at the fields far away from any source, we retain only the latter, thus

$$(\partial_i \Phi)_{\mathrm{far}} = \frac{q R_j \partial_i \beta_j}{\left(R - \vec{R}\cdot\vec{\beta}\right)^2} = \frac{q\left(\hat{e}\cdot\dot{\vec{\beta}}\right)\partial_i t'}{R\left(1 - \hat{e}\cdot\vec{\beta}\right)^2} \,. \tag{1.121}$$

The remaining spatial derivative of the retarded time is

$$\partial_i t' = -\frac{\partial_i R}{c} = -\frac{R_j}{R}\partial_i R_j = -e_j\left(\frac{\delta_{ij}}{c} - \beta_j \partial_i t'\right) = -\frac{e_i}{c} + \left(\hat{e}\cdot\vec{\beta}\right)\partial_i t' \,. \tag{1.122}$$

This equation gives

$$\partial_i t' = -\frac{e_i}{c\left(1 - \hat{e}\cdot\vec{\beta}\right)} \,, \tag{1.123}$$

which implies with (1.121)

$$(\vec{\nabla}\Phi)_{\mathrm{far}} = -\frac{q\left(\hat{e}\cdot\dot{\vec{\beta}}\right)\hat{e}}{Rc\left(1 - \hat{e}\cdot\vec{\beta}\right)^3} \tag{1.124}$$

for the gradient of Φ in the far-field. The time derivative of $\vec{A}$ is

$$(\partial_t \vec{A})_{\mathrm{far}} = \Phi\partial_t\vec{\beta} + \vec{\beta}\partial_t\Phi = \frac{q\dot{\vec{\beta}}\partial_t t'}{R\left(1 - \hat{e}\cdot\vec{\beta}\right)} + \frac{q\vec{\beta}\left(\hat{e}\cdot\dot{\vec{\beta}}\right)\partial_t t'}{R\left(1 - \hat{e}\cdot\vec{\beta}\right)^2} \tag{1.125}$$

if we again drop all terms with a steeper R dependence than R^{-1} to isolate the far-field. Now, the time derivative of t' is given by

$$\partial_t t' = 1 - \frac{\partial_t R}{c} = 1 - \frac{R_j}{Rc}\partial_t R_j = 1 + \hat{e}\cdot\vec{\beta}\partial_t t' \,, \tag{1.126}$$

thus

$$\partial_t t' = \frac{1}{1 - \hat{e}\cdot\vec{\beta}} \,. \tag{1.127}$$

The far-field time derivative of $\vec{A}$ then turns into

$$\left(\partial_t\vec{A}\right)_{\text{far}} = \frac{q}{R\left(1-\hat{e}\cdot\vec{\beta}\right)^3}\left[\left(1-\hat{e}\cdot\vec{\beta}\right)\dot{\vec{\beta}}+\vec{\beta}\left(\hat{e}\cdot\dot{\vec{\beta}}\right)\right] . \tag{1.128}$$

From this, together with (1.124), and using the identity $\vec{a}\times(\vec{b}\times\vec{c}) = (\vec{a}\cdot\vec{c})\vec{b} - (\vec{a}\cdot\vec{b})\vec{c}$ twice, we find the electric field far from the source,

$$\vec{E}_{\text{far}} = \frac{q}{Rc\left(1-\hat{e}\cdot\vec{\beta}\right)^3}\hat{e}\times\left[\left(\hat{e}-\vec{\beta}\right)\times\dot{\vec{\beta}}\right] . \tag{1.129}$$

The magnetic field is

$$\vec{B} = \vec{\nabla}\times\vec{A} = \Phi\vec{\nabla}\times\vec{\beta}-\vec{\beta}\times\vec{\nabla}\Phi . \tag{1.130}$$

Taking the curl of the velocity $\vec{\beta}$, we must be aware that $\vec{\beta}$ depends on position through the retarded time t'. In components, we have

$$\left(\vec{\nabla}\times\vec{\beta}\right)_i = \varepsilon_{ijk}\partial_j\beta_k = \varepsilon_{ijk}\dot{\beta}_k\partial_j t' . \tag{1.131}$$

With the help of (1.123), we then find

$$\vec{\nabla}\times\vec{\beta} = -\frac{\hat{e}\times\dot{\vec{\beta}}}{c\left(1-\hat{e}\cdot\vec{\beta}\right)} , \tag{1.132}$$

which, together with (1.124), allows us to write

$$\vec{B}_{\text{far}} = -\frac{q}{Rc\left(1-\hat{e}\cdot\vec{\beta}\right)^3}\hat{e}\times\left[\dot{\vec{\beta}}+\hat{e}\times\left(\vec{\beta}\times\dot{\vec{\beta}}\right)\right] . \tag{1.133}$$

?

Convince yourself by your own calculation that expressions (1.129) and (1.133) for the electric and magnetic fields far from the source are correct.

Comparing to (1.129), it is straightforward to confirm that

$$\vec{B}_{\text{far}} = \hat{e}\times\vec{E}_{\text{far}} . \tag{1.134}$$

Using this result, the Poynting vector far away from the source is

$$\vec{S} = \frac{q^2}{4\pi R^2 c\left(1-\hat{e}\cdot\vec{\beta}\right)^6}\left|\hat{e}\times\left[\left(\hat{e}-\vec{\beta}\right)\times\dot{\vec{\beta}}\right]\right|^2\hat{e} . \tag{1.135}$$

This quantifies the energy received per unit area per unit time by the observer. We now need to distinguish between a time interval $\mathrm{d}t$ measured by the observer and the corresponding interval $\mathrm{d}t'$ of the retarded time. The latter is the time interval during which the source needs to emit for the observer to see its radiation for the time interval $\mathrm{d}t$. Since, according to (1.127), the retarded time interval $\mathrm{d}t'$ is related to the time interval $\mathrm{d}t$ measured by the observer through

$$\mathrm{d}t = \left(1-\hat{e}\cdot\vec{\beta}\right)\mathrm{d}t' , \tag{1.136}$$

the energy emitted per the observer's unit time $\mathrm{d}t$ into the solid angle element $\mathrm{d}\Omega$ is

$$\mathrm{d}E = \vec{S}\cdot\hat{e}R^2\mathrm{d}\Omega\mathrm{d}t = \frac{q^2}{4\pi c\left(1-\hat{e}\cdot\vec{\beta}\right)^5}\left|\hat{e}\times\left[\left(\hat{e}-\vec{\beta}\right)\times\dot{\vec{\beta}}\right]\right|^2\mathrm{d}\Omega\mathrm{d}t' , \tag{1.137}$$

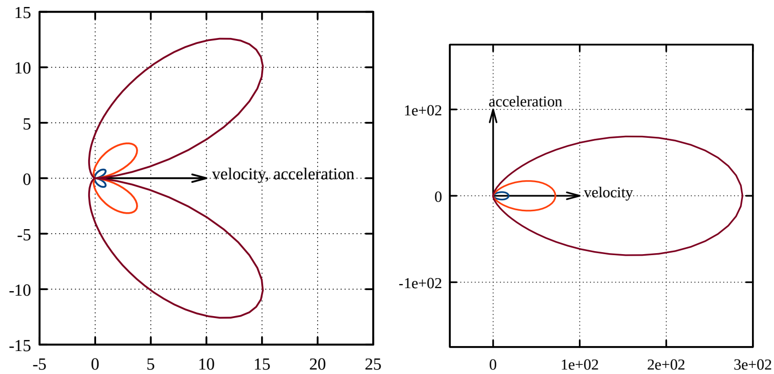

Figure 1.6 The radiation power according to the relativistic Larmor formula is illustrated for a charge accelerated parallel (left) and perpendicular (right) to its relativistic velocity. Three curves are given for three arbitrary values of the acceleration. Notice the different scales of the two plots!

and thus the power emitted per unit solid angle and per unit retarded time $\mathrm{d}t'$ is

$$\frac{\mathrm{d}P}{\mathrm{d}\Omega} = \frac{q^2}{4\pi c\left(1 - \hat{e}\cdot\vec{\beta}\right)^5}\left|\hat{e}\times\left[\left(\hat{e}-\vec{\beta}\right)\times\dot{\vec{\beta}}\right]\right|^2 . \tag{1.138}$$

This is the relativistic Larmor formula which describes the power radiated by a source per unit solid angle (Figures 1.6 and 1.7).

The total emitted power is the solid-angle integral of (1.138). This calculation is not difficult to carry out, but lengthy. Perhaps the most straightforward way begins by expanding the double vector product using the identity

$$\vec{a}\times\left(\vec{b}\times\vec{c}\right) = (\vec{a}\cdot\vec{c})\,\vec{b} - \left(\vec{a}\cdot\vec{b}\right)\vec{c}\,, \tag{1.139}$$

followed by squaring the result. Then, it is useful to introduce coordinates such that the velocity $\vec{\beta}$ points into the $\hat{e}_z$ direction, $\vec{\beta} = \beta\hat{e}_z$, the acceleration $\dot{\vec{\beta}}$ falls into the x-z plane, $\dot{\vec{\beta}} = \dot{\beta}(\sin\alpha\hat{e}_x + \cos\alpha\hat{e}_z)$, and

$$\hat{e} = \begin{pmatrix}\cos\phi\sin\theta\\ \sin\phi\sin\theta\\ \cos\theta\end{pmatrix} . \tag{1.140}$$

Then, the ϕ and θ integrations can be carried out in this order, giving the result

$$P = \frac{2e^2}{3c}\gamma^6\left[\dot{\beta}^2 - \left(\vec{\beta}\times\dot{\vec{\beta}}\right)^2\right] . \tag{1.141}$$

?

Can you confirm that integrating the Larmor formula (1.138) over the solid angle results in (1.141)?

The factor γ^6 is most remarkable: A relativistically moving charge with a high Lorentz factor radiates with an enormous power. For non-relativistically moving charges, equations (1.138) and (1.141) simplify to

$$\frac{\mathrm{d}P}{\mathrm{d}\Omega} = \frac{q^2}{4\pi c}\left|\hat{e}\times\left(\hat{e}\times\dot{\vec{\beta}}\right)\right|^2 = \frac{q^2}{4\pi c}\left|\dot{\vec{\beta}} - \left(\dot{\vec{\beta}}\cdot\hat{e}\right)\hat{e}\right|^2 = \frac{q^2}{4\pi c}\dot{\beta}_\perp^2\,, \tag{1.142}$$

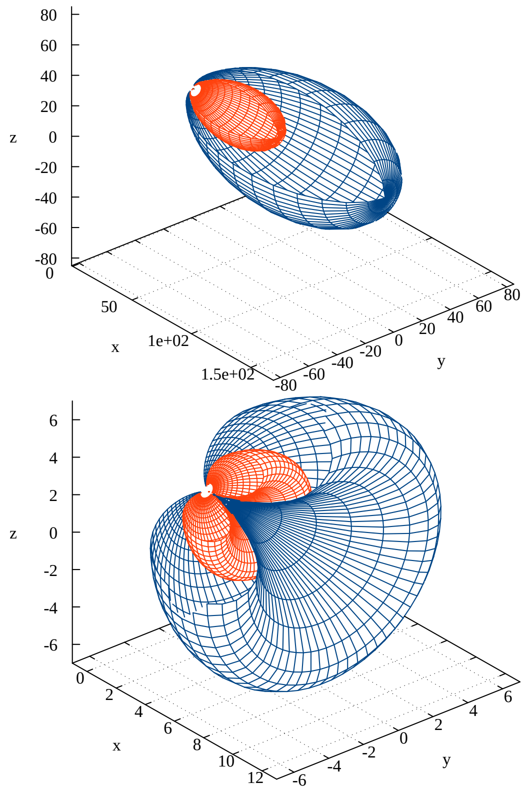

Figure 1.7 Three-dimensional illustrations of the radiation power of two accelerated charges with $\beta = 0.5$. A charge accelerated perpendicular to its direction of motion has its emission peaked strongly into the forward direction of its motion (top panel), while the radiation of a charge accelerated parallel to its direction of motion is emitted into a collar surrounding its trajectory (bottom panel).

where $\dot{\vec{\beta}}_\perp$ is the acceleration perpendicular to $\hat{e}$, and

$$P = \frac{2q^2}{3c}\dot{\beta}^2 \; . \qquad (1.143)$$

1.3.6 The Lorentz Force

The action for a relativistic particle with mass m and charge q in an electromagnetic field with vector potential A^μ is

$$S = -mc^2 \int \mathrm{d}\tau + \frac{q}{c} \int A_\mu \mathrm{d}x^\mu \; . \qquad (1.144)$$

This is the simplest Lorentz-invariant expression that can be formed from the only Lorentz-invariant quantity of a free particle, i.e. its proper time τ, the four potential A^μ and the coordinates x^μ of the particle trajectory. Variation of the action (1.144) with respect to the particle trajectory under a fixed four-potential A^μ and equating the result to zero leads to the equation of motion

$$m\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = \frac{q}{c}F^\mu{}_\nu u^\nu \ . \tag{1.145}$$

With $u^0 = \gamma c$, $u^i = \gamma v^i$ and $\mathrm{d}\tau = \gamma^{-1}\mathrm{d}t$, the 0-component of this equation means

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\gamma mc^2\right) = q\vec{E}\cdot\vec{v} \,, \tag{1.146}$$

showing that the work done by the electric field changes the energy γmc^2 of the particle. The spatial components give

$$\boxed{m\frac{\mathrm{d}\left(\gamma\vec{v}\right)}{\mathrm{d}t} = q\vec{E} + \frac{q}{c}\vec{v}\times\vec{B}} \ . \tag{1.147}$$

For non-relativistic motion, $\gamma = 1$, and (1.147) reproduces the common equation of motion under the Lorentz force.

?

Derive the equation of motion (1.145) as the Euler-Lagrange equation of the action (1.144).

Problems

1. Starting from the Lorentz transform (1.86) of the electromagnetic field tensor, expressly derive the Lorentz transform (1.87) of the field components.

2. Show by explicit calculation that Maxwell's equation in three-dimensional form follow from their relativistic forms (1.95) and (1.96).

3. Convince yourself that the components of the energy-momentum tensor (1.110) have the appropriate physical units.

4. From the invariants (1.88) of the electromagnetic field tensor, derive the following statements:

   (a) If $\vec{E}$ and $\vec{B}$ have the same amplitude $|\vec{E}| = |\vec{B}|$ in one inertial frame, then also in all other inertial frames.
   (b) If $\vec{E}$ and $\vec{B}$ are orthogonal in one inertial frame, then also in all other inertial frames.

5. Apply the Larmor formula to the classical picture of an electron in a hydrogen atom.

   (a) Decide whether the non-relativistic approximation of the Larmor formula can be applied.
   (b) Estimate the classical lifetime of a hydrogen atom.

6. Derive the electromagnetic field of a point charge q uniformly moving with the velocity $\vec{v}_0$.

(a) Calculate the Liénard-Wiechert potentials (1.117) for a point charge moving with constant velocity along a straight line and compute the electromagnetic fields from them. *Hint*: Since the velocity is constant, we never need the retarded time itself, but only the separation R between the charge and any point $\vec{x}$ at the retarded time. Introducing the vector $\vec{\omega} = \vec{x} - \vec{v}_0 t$ helps greatly.

(b) Find the fields by a suitable Lorentz transform and compare the two results.

1.4 Elementary kinetic theory

This section serves a dual purpose. The discussion of the Boltzmann equation and the BBGKY hierarchy prepares the derivation of the hydrodynamical equations later in this book. The Fokker-Planck equation derived thereafter from a diffusion approximation of the collision terms occurs under a variety of circumstances in astrophysics, from radiation transport to stellar dynamics. The main results are the Boltzmann equation (1.156), the master equation (1.161), the Fokker-Planck equation in its original form (1.163), its form (1.172) with one of the diffusion coefficients eliminated by equilibrium considerations, and its form (1.180) for small changes in absolute momentum.

1.4.1 The BBGKY hierarchy and the Boltzmann equation

Kinetic theory describes how *ensembles* of particles change in time, in absence or in presence of mutual collisions. In classical mechanics, generalised coordinates q_i are assigned to the degrees of freedom that the system under consideration has. The number of degrees of freedom d depends on the number of components of the system and their mutual relations to each other. If the system consists of N independent point particles in three-dimensional space, $d = 3N$. If those particles are linked to form a solid body, $d = 6$, because only three degrees of translational and three degrees of rotational freedom remain. By Newton's second law, two initial conditions must be given for each degree of freedom, which can be chosen to be the generalised coordinates q_i and the associated velocities, $\dot{q}_i$, at some initial time.

If the system can be described by a Lagrange function $L(q_i, \dot{q}_i, t)$, the canonically conjugated momenta

$$p_i = \frac{\partial L(q_i, \dot{q}_i, t)}{\partial \dot{q}_i} \tag{1.148}$$

can be substituted for the velocities $\dot{q}_i$ by the Legendre transform

$$H(q_i, p_i, t) = \sum_{i=1}^{d} \dot{q}_i p_i - L\left[q_i, \dot{q}_i(p_i), t\right] , \tag{1.149}$$

leading to the Hamilton function $H(q_i, p_i, t)$. The equations of motion for all degrees of freedom are then Hamilton's equations,

$$\dot{q}_i = \frac{\partial H(q_i, p_i, t)}{\partial p_i} , \quad \dot{p}_i = -\frac{\partial H(q_i, p_i, t)}{\partial q_i} . \tag{1.150}$$

The physical state of such a system is fully characterised by the d generalised coordinates $\vec{q} = (q_1, \ldots, q_d)$ and their d canonically conjugated momenta $\vec{p} = (p_1, \ldots, p_d)$. The generalised coordinates $\vec{q}$ span the configuration space of the system. Together with their conjugate momenta, they span the $2d$-dimensional phase space.

It is by no means unique how the $2d$ phase-space coordinates are to be divided into generalised coordinates and their conjugate momenta. Canonical transformations applied to phase space leave Hamilton's equations invariant, but can turn coordinates into momenta and vice versa.

The classical physical state of the system is given by the system's location in the $2d$-dimensional phase space. Statistical mechanics is not interested in the phase-space coordinates of all particles in an ensemble. Rather, it divides phase space into cells of small, but finite size, sums the number of particles in each cell and studies the time evolution of this number instead of the time evolution of each individual pair (q_i, p_i) of phase-space coordinates. We thus introduce a distribution function $f^{(d)}(t, \vec{q}, \vec{p}\,)$ such that the probability for finding the system in a small phase-space cell around to the phase-space point $(\vec{q}, \vec{p}\,)$ at time t is

$$\mathrm{d}P^{(d)}\,(t, \vec{q}, \vec{p}\,) = f^{(d)}\,(t, \vec{q}, \vec{p}\,)\;\mathrm{d}^d q\,\mathrm{d}^d p\ . \tag{1.151}$$

For systems with very many degrees of freedom, the full phase-space distribution function $f^{(d)}$ becomes utterly unmanageable, apart from the fact that the complete knowledge of the evolution of all d degrees of freedom is then neither desired nor necessary. Rather, we are then interested in the reduced phase-space distribution function $f^{(k)}$, obtained by integrating $f^{(d)}$ over $d-k$ coordinates and momenta,

$$f^{(k)}(t, q_1, \ldots, q_k, p_1, \ldots p_k) = \int \mathrm{d}q_{k+1} \ldots \mathrm{d}q_d \int \mathrm{d}p_{k+1} \ldots \mathrm{d}p_d\, f^{(d)}(t, \vec{q}, \vec{p}\,)\ . \tag{1.152}$$

By Liouville's theorem and Hamilton's equations, the time evolution of the full phase-space distribution function $f^{(d)}$ is determined by Liouville's equation

$$\frac{\partial f^{(d)}}{\partial t} + \dot{q}_i \frac{\partial f^{(d)}}{\partial q_i} + \dot{p}_j \frac{\partial f^{(d)}}{\partial p_j} = \frac{\partial f^{(d)}}{\partial t} + \frac{\partial H}{\partial p_i}\frac{\partial f^{(d)}}{\partial q_i} - \frac{\partial H}{\partial q_j}\frac{\partial f^{(d)}}{\partial p_j} = 0\ . \tag{1.153}$$

Searching for an evolution equation for any of the reduced phase-space distribution functions $f^{(k)}$, we have to integrate Liouville's equation over $d-k$ degrees of freedom and sort terms accordingly. It then appears that the evolution of the reduced distribution function $f^{(k)}$ depends on the reduced distribution function at the next higher level, $f^{(k+1)}$. This establishes the so-called BBGKY hierarchy of equations of motion for the reduced distribution functions, where the acronym stands for the authors Born, Bogoliubov, Green, Kirkwood and Yvon.

To see what the BBGKY hierarchy means, let us begin with the reduced phase-space distribution $f^{(1)}$ for a single degree of freedom. It will depend on the distribution function $f^{(2)}$ for two degrees of freedom, which expresses the notion that individual degrees of freedom do not evolve in isolation, but in correlation with others. In an ensemble of particles, the motion of a single particle is determined by two-body correlations with other particles, which in turn are affected by three-body correlations, and so forth. Clearly, the BBGKY hierarchy needs to be terminated somewhere, or closed, for us to make any progress. This closure is typically set by ignoring any correlations higher than a certain order.

We are particularly interested in the evolution of the distribution function for *single particles*. Let us therefore imagine that we have an ensemble of N point particles with $d = 3N$ degrees of freedom. We then integrate out all $3N - 3 = 3(N - 1)$ degrees of freedom belonging to $N - 1$ of the N particles and arrive at an evolution equation for the one-particle distribution function. According to the BBGKY hierarchy, this evolution equation will contain two-particle correlations. Closure can now be achieved by assuming that any two particles are *statistically uncorrelated*. The joint probability for finding a pair of particles at two positions in phase space is then simply the product of the probabilities for finding one of the particles at one position and the other at the other position. The two-particle distribution function can then be written as a product of one-particle distribution functions.

This closure condition means that any two particles affect each other's motion exclusively by direct two-body collisions. They move independently until they collide, and continue moving independently after the collision. This is possible if the interaction potential between any two particles is short-ranged compared to the mean inter-particle distance.

Following these considerations, we introduce a one-particle distribution function $f(t, \vec{q}, \vec{p}\,)$ by integrating $f^{(d)}$ over all but those degrees of freedom that belong to a single particle. For an ensemble of point particles in three-dimensional space, $f(t, \vec{q}, \vec{p}\,)$ is then defined on an effective, six-dimensional phase space. Moreover, we normalize the distribution $f(t, \vec{q}, \vec{p}\,)$ such that

$$f(t, \vec{q}, \vec{p}\,)\, \mathrm{d}^3 q\, \mathrm{d}^3 p = \mathrm{d}N \tag{1.154}$$

is the number of particles expected to be found within the infinitesimal phase-space volume $\mathrm{d}\Gamma = \mathrm{d}^3 q \mathrm{d}^3 p$ around the phase-space position $(\vec{q}, \vec{p}\,)$. For this one-particle phase-space distribution function $f(t, \vec{q}, \vec{p}\,)$, Liouville's equation reduces to Boltzmann's equation,

$$\frac{\partial f}{\partial t} + \dot{\vec{q}} \cdot \frac{\partial f}{\partial \vec{q}} + \dot{\vec{p}} \cdot \frac{\partial f}{\partial \vec{p}} = C[f]\ , \tag{1.155}$$

where the term $C[f]$ is called *collision term*: According to our closure condition for the BBGKY hierarchy, particle interactions are determined by direct particle collisions only and thus by the one-particle distribution function itself. The collision term must then be a functional of f. For a Hamiltonian system with Hamilton function $H = H(t, \vec{q}, \vec{p}\,)$, Boltzmann's equation reads

$$\boxed{\frac{\partial f}{\partial t} + \frac{\partial H}{\partial \vec{p}} \cdot \frac{\partial f}{\partial \vec{q}} - \frac{\partial H}{\partial \vec{q}} \cdot \frac{\partial f}{\partial \vec{p}} = C[f]}\ . \tag{1.156}$$

If, as usual, the Hamilton function can be written as $H = T + V$, with the kinetic energy T depending on the conjugate momenta $\vec{p}$ only and a potential energy V depending only on the generalised coordinates $\vec{q}$, and if further $T = \vec{p}^{\,2}/(2m)$, then we can write

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial \vec{q}} \cdot \frac{\vec{p}}{m} - \frac{\partial f}{\partial \vec{p}} \cdot \frac{\partial V}{\partial \vec{q}} = C[f]\ . \tag{1.157}$$

1.4.2 Collision terms

In presence of collisions, the phase-space density changes schematically according to

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \text{gain} - \text{loss} \,, \tag{1.158}$$

where the gain and loss terms are due to scattering into and out of the phase-space element $\mathrm{d}\vec{w}$ under consideration. Let $\psi(\vec{w}, \delta\vec{w})\mathrm{d}\delta\vec{w}\mathrm{d}t$ be the transition probability due to scattering by an amount $\delta\vec{w}$ from $\vec{w}$ to $\vec{w} + \delta\vec{w}$ within the time interval $\mathrm{d}t$. Typically, ψ would be quantified by a scattering cross section. Then, the gain term is

$$\text{gain} = \int \mathrm{d}\delta\vec{w}\, \psi(\vec{w} - \delta\vec{w}, \delta\vec{w}) f(t, \vec{w} - \delta\vec{w}) \tag{1.159}$$

since the integral quantifies the expected number of particles moving per unit time from the phase-space coordinates $\vec{w} - \delta\vec{w}$ to the phase-space coordinates $\vec{w}$: It multiplies the number of particles at the original phase-space point with their transition probability per unit time and integrates over all possible changes $\delta\vec{w}$. Similarly, the loss term is

$$\text{loss} = \int \mathrm{d}\delta\vec{w}\, \psi(\vec{w}, \delta\vec{w}) f(t, \vec{w}) \,. \tag{1.160}$$

Inserting these gain and loss terms (1.159) and (1.160) into (1.158) yields the so-called master equation

$$\frac{\mathrm{d}f(t, \vec{w})}{\mathrm{d}t} = \int \mathrm{d}\delta\vec{w} \left[\psi(\vec{w} - \delta\vec{w}, \delta\vec{w}) f(t, \vec{w} - \delta\vec{w}) - \psi(\vec{w}, \delta\vec{w}) f(t, \vec{w})\right] \,, \tag{1.161}$$

describing the change of the phase-space density due to the collisions causing the transition probability ψ in phase space.

1.4.3 Diffusion in phase space: The Fokker-Planck approximation

We study the time evolution of the phase-space density f here under the quite relevant assumption that the phase-space coordinates of particles change only by small amounts in individual collisions. Then, the particles diffuse in phase space and their phase-space density changes gradually in a way that can be described with two diffusion coefficients. As we shall see in the course of this treatment, it is sufficient for this approximation if the absolute values of the phase-space coordinates change only very little in each collision, while the scattering angles can even be large. Under these circumstances, this diffusion approximation is most useful to describe all kinds of particle ensembles which either have low mass or low energy and interact with another particle ensemble of high mass or high energy. The equation describing how the phase-space density f changes with time under this approximation is called the Fokker-Planck equation. Its derivation, and general methods for its solution, are the main subject of the following treatment.

Specifically, let us assume that conditions are such that it is permissible to assume that the change $\Delta\vec{w}$ in the phase-space coordinates is small enough for the transition probability and the phase-space density at $\vec{w} - \Delta\vec{w}$ to be approximated by Taylor expansions up to second order,

$$\begin{aligned} \psi(\vec{w} - \delta\vec{w}, \delta\vec{w})f(t, \vec{w} - \delta\vec{w}) \approx\ & \psi(\vec{w}, \delta\vec{w})f(t, \vec{w}) \\ & - \frac{\partial}{\partial w_i}\left[\psi(\vec{w}, \delta\vec{w})f(t, \vec{w})\right]\delta w_i \\ & + \frac{1}{2}\frac{\partial^2}{\partial w_i \partial w_j}\left[\psi(\vec{w}, \delta\vec{w})f(t, \vec{w})\right]\delta w_i \delta w_j\ . \end{aligned} \tag{1.162}$$

Inserting this into the master equation (1.161) leads us already to the Fokker-Planck equation

$$\frac{\mathrm{d}f(\vec{w})}{\mathrm{d}t} = -\frac{\partial}{\partial w_i}\left[f(\vec{w})\, D_1^i(\vec{w})\right] + \frac{\partial^2}{\partial w_i \partial w_j}\left[f(\vec{w})\, D_2^{ij}(\vec{w})\right]\ , \tag{1.163}$$

which approximates scattering as a second-order diffusion process in phase space. The first- and second-order diffusion coefficients are

$$\begin{aligned} D_1^i(\vec{w}) &= \int \mathrm{d}\delta\vec{w}\, \psi(\vec{w}, \delta\vec{w})\, \delta w_i\ , \\ D_2^{ij}(\vec{w}) &= \frac{1}{2}\int \mathrm{d}\delta\vec{w}\, \psi(\vec{w}, \delta\vec{w})\, \delta w_i \delta w_j\ . \end{aligned} \tag{1.164}$$

The first-order coefficient D_1^i integrates the change δw_i in the phase-space coordinate w_i over the transition probability per unit time and thus quantifies the mean change of w_i per unit time. Similarly, the second-order coefficient D_2^{ij} quantifies the variances D_2^{ii} of the changes in w_i, and the covariances D_2^{ij} of different phase-space coordinates w_i and w_j for $i \neq j$. Thus, the combined vector with components D_1^i is the mean change per unit time of the position vector $\vec{w}$ in phase space, while D_2^{ij} is the covariance matrix of all individual changes.

?

Verify the Fokker-Planck equation (1.163) and the expressions (1.164) for the diffusion coefficients by your own derivation.

Suppose now that any change in the spatial coordinates is irrelevant, for example because all relevant particle species are homogeneously distributed in space. In fact, this assumption is much less restrictive than it might seem. It can also be satisfied statistically in the sense that although particles may move in space, the number of particles moving away from a specific point in space is compensated by an equal number moving there. In other words, what we set out to consider now is a dynamical spatial equilibrium. Then, we can concentrate on the d-dimensional momentum subspace of phase space, restrict $\vec{w} = \vec{p}$ and $\delta\vec{w} = \delta\vec{p}$ and consider the phase-space distribution function f as a function of $(t, \vec{p})$ only. The total time derivative of $f(t, \vec{p})$ then equals its partial time derivative, because

$$\frac{\partial f(t, \vec{p})}{\partial \vec{q}} = 0 \quad \text{and} \quad \frac{\partial f(t, \vec{p})}{\partial \vec{p}} \cdot \dot{\vec{p}} = 0 \quad \text{for} \quad \dot{\vec{p}} = -\frac{\partial H}{\partial \vec{q}} = 0\ . \tag{1.165}$$

Then, the Fokker-Planck equation (1.163) simplifies to a partial differential equation in time and momentum only,

$$\begin{aligned} \frac{\partial f(t, \vec{p})}{\partial t} &= -\frac{\partial}{\partial p_i}\left[f(t, \vec{p})D_1^i(\vec{p})\right] + \frac{\partial^2}{\partial p_i \partial p_j}\left[f(t, \vec{p})D_2^{ij}(\vec{p})\right] \\ &= -\frac{\partial}{\partial p_i}\left[\left(D_1^i(\vec{p}) - \frac{\partial}{\partial p_j}D_2^{ij}(\vec{p})\right)f(t, \vec{p}) - D_2^{ij}(\vec{p})\frac{\partial f(t, \vec{p})}{\partial p_j}\right]\ . \end{aligned} \tag{1.166}$$

This equation manifestly has the form of a continuity equation, where the term in brackets represents the current density $\vec{j}_p$ in momentum space,

$$\frac{\partial f(t,\vec{p}\,)}{\partial t} + \vec{\nabla}_p \cdot \vec{j}_p = 0 \ , \qquad j_p^i = \left(D_1^i(\vec{p}\,) - \frac{\partial}{\partial p_j} D_2^{ij}(\vec{p}\,)\right) f(t,\vec{p}\,) - D_2^{ij}(\vec{p}\,)\frac{\partial f(t,\vec{p}\,)}{\partial p_j} \ . \tag{1.167}$$

At this point, it is important to note that the two diffusion coefficients D_1^i and D_2^{ij} are generally not independent. In an equilibrium situation, the momentum current $\vec{j}_p$ must vanish. Setting the components $j_p^i = 0$ in (1.167) for an equilibrium phase-space distribution $\bar{f}(t,\vec{p}\,)$ implies that then the coefficient D_1^i can be expressed by D_2^{ij} and the derivative of $\bar{f}(t,\vec{p}\,)$ with respect to the momentum,

$$D_1^i(\vec{p}\,) = \frac{\partial D_2^{ij}(\vec{p}\,)}{\partial p_j} + D_2^{ij}(\vec{p}\,)\frac{\partial \ln \bar{f}(t,\vec{p}\,)}{\partial p_j} \ . \tag{1.168}$$

However, since both coefficients do not depend on the specific form of f, we can now use them in the more general situation of an arbitrary phase-space distribution. Inserting the relation (1.168) into (1.167), the derivative of D_2^{ij} with respect to the momenta cancels, and the momentum current

$$j_p^i = -D_2^{ij}(\vec{p}\,) f(t,\vec{p}\,) \frac{\partial}{\partial p_j}\left[\ln f(t,\vec{p}\,) - \ln \bar{f}(t,\vec{p}\,)\right] \tag{1.169}$$

is shown to be driven by the momentum gradient of the ratio between the actual and the equilibrium phase-space distributions.

Example: Maxwellian momentum distribution

Suppose, for example, that the equilibrium distribution of the particle species under consideration can be described as a Maxwellian momentum distribution with a temperature $\bar{T}$. Then,

$$\bar{f}(t,\vec{p}\,) \propto \exp\left(-\frac{p^2}{2mk\bar{T}}\right) \ , \qquad \frac{\partial \ln \bar{f}(t,\vec{p}\,)}{\partial p_j} = -\frac{p_j}{mk\bar{T}} \ , \tag{1.170}$$

the components of the momentum current simplify to

$$j_p^i = -D_2^{ij}(\vec{p}\,) f(t,\vec{p}\,) \left[\frac{\partial \ln f(t,\vec{p}\,)}{\partial p_j} + \frac{p_j}{mk\bar{T}}\right] \ , \tag{1.171}$$

and the Fokker-Planck equation becomes

$$\frac{\partial f(t,\vec{p}\,)}{\partial t} - \frac{\partial}{\partial p_i}\left[D_2^{ij}(\vec{p}\,) f(t,\vec{p}\,)\left(\frac{\partial f(t,\vec{p}\,)}{\partial p_j} + \frac{p_j}{mk\bar{T}}\right)\right] = 0 \ . \tag{1.172}$$

◂

1.4.4 Diffusion in absolute momentum

Quite frequently, the scattering process changes the absolute value of the momentum by a small amount only, while the scattering angle may be large. Then, the diffusion approximation is still valid in terms of the absolute momentum, but not in the full three-dimensional momentum space any more. In other words, momentum can then be considered as slowly diffusing between spherical shells in momentum space, while its direction angles may be vastly redistributed from one shell to another. Instead of the phase-space density $f(t,\vec{p}\,)$, we must then consider the density $f(t,p)p^2$ of particles in absolute momentum, irrespective of its direction. The Fokker-Planck approximation then still applies between momentum shells, and the Fokker-Planck equation becomes

$$\frac{\partial\left(fp^2\right)}{\partial t} = \frac{\partial}{\partial p}\left[\left(D_1 + \frac{\partial D_2}{\partial p}\right)\left(fp^2\right) + D_2\frac{\partial\left(fp^2\right)}{\partial p}\right] , \tag{1.173}$$

with the diffusion coefficients

$$D_1(p) = \int \mathrm{d}\delta p\,\psi(p,\delta p)\delta p\ , \quad D_2(p) = \frac{1}{2}\int \mathrm{d}\delta p\,\psi(p,\delta p)\delta p^2\ . \tag{1.174}$$

Both coefficients are now one-dimensional. The first-order coefficient D_1 is the mean momentum change per unit time, while the second-order coefficient D_2 is its mean-square.

We can now express the Fokker-Planck equation as a radial diffusion equation in momentum space,

$$\frac{\partial f}{\partial t} + \frac{1}{p^2}\frac{\partial\left(j_p p^2\right)}{\partial p} = 0\ , \quad j_p = \left(D_1 + \frac{\partial D_2}{\partial p}\right)f + \frac{D_2}{p^2}\frac{\partial\left(fp^2\right)}{\partial p}\ , \tag{1.175}$$

where now j_p is the radial component of the momentum current. Notice that the operator applied to the momentum current is the divergence in spherical polar coordinates, so the meaning of the equation has not changed: It remains a conservation equation, expressing that any change in phase-space density is caused by a momentum current.

Again, j_p must vanish in an equilibrium situation, expressed by an equilibrium phase-space density $\bar{f}$. This requirement establishes the relation

$$D_1 = -\left(\frac{2D_2}{p} + \frac{\partial D_2}{\partial p}\right) - D_2\frac{\partial\ln\bar{f}}{\partial p} \tag{1.176}$$

between D_1 and D_2. Inserting this result into the current density in (1.175) gives, after some straightforward rearrangement,

$$j_p = D_2 f\frac{\partial}{\partial p}\left(\ln f - \ln\bar{f}\right) = D_2 f\frac{\partial}{\partial p}\ln\frac{f}{\bar{f}}\ . \tag{1.177}$$

?

Convince yourself of the relations (1.176) and (1.177) under the conditions discussed here.

1.4.5 Calculation of the diffusion coefficient D_2

For an actual calculation of the diffusion coefficient D_2, we return to its definition in (1.164) or the more specialised form (1.174) and recall that the physical meaning of D_2 is (one half) the mean-squared momentum change per unit time of the particle species considered,

$$D_2 = \frac{1}{2}\left\langle\delta p^2\right\rangle\ . \tag{1.181}$$

Example: Maxwellians with different temperatures

To give an example, let us assume that both the actual and the equilibrium phase-space distributions, f and $\bar{f}$, are Maxwellians characterised by two different temperatures T and $\bar{T}$, respectively. Then,

$$\frac{\partial \ln f}{\partial p} - \frac{\partial \ln \bar{f}}{\partial p} = -\frac{p}{mkT}\left(1 - \frac{T}{\bar{T}}\right) , \tag{1.178}$$

the momentum current density becomes

$$j_p = -D_2 \frac{pf}{mkT}\left(1 - \frac{T}{\bar{T}}\right) , \tag{1.179}$$

and the Fokker-Planck equation reduces to

$$\frac{\partial f}{\partial t} = \frac{1}{p^2 mkT}\left(1 - \frac{T}{\bar{T}}\right)\frac{\partial}{\partial p}\left(D_2 p^3 f\right) . \tag{1.180}$$

◀

To illustrate this, consider a species of heavy particles with mass M embedded in a sea of light particles with mass $m \ll M$. Then, the energy of the heavy particles is almost unchanged by the collisions with the light particles, while momentum conservation implies a small change δp in absolute momentum determined by

$$\delta p^2 = 2q^2(1 - \cos\theta) \tag{1.182}$$

per collision, if q and θ are the momentum and the scattering angle of the light particle. The probability of a light particle with velocity $v = q/m$ scattering off a heavy particle per unit time into the solid-angle element $\mathrm{d}\Omega$ is

$$nv\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\mathrm{d}\Omega = \frac{nq}{m}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\mathrm{d}\Omega , \tag{1.183}$$

where n is the number density of light particles. Thus, the mean-squared momentum change per unit time of a heavy particle is

$$\left\langle \delta p^2 \right\rangle = \frac{2n}{m}\left\langle \int q^3(1 - \cos\theta)\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\mathrm{d}\Omega \right\rangle , \tag{1.184}$$

where the average has to be taken over the momentum distribution of the light particles.

?

Can you confirm that the differential cross section for light point particles scattered by a hard sphere is given by (1.185)?

Suppose that the heavy particles can be considered as hard spheres with radius R, while the light particles approximate point masses. Then, in the idealised situation of light particles bouncing off heavy, hard spheres,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{R^2}{4} , \quad \int (1 - \cos\theta)\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\mathrm{d}\Omega = \frac{\pi R^2}{2}\int_{-1}^{1}(1 - \cos\theta)\mathrm{d}(\cos\theta) = \pi R^2 , \tag{1.185}$$

and the diffusion coefficient D_2 becomes

$$D_2 = \frac{1}{2}\left\langle \delta p^2 \right\rangle = \frac{\pi n R^2}{m}\left\langle q^3 \right\rangle , \tag{1.186}$$

where the average over the cubed momentum of the light particles remains.

If their velocity distribution is of Maxwellian form with temperature $\bar{T}$,

$$\left\langle q^3 \right\rangle = \frac{8\sqrt{2}}{\sqrt{\pi}}(mk\bar{T})^{3/2} , \tag{1.187}$$

and the diffusion coefficient finally assumes the form

$$D_2 = 8nR^2 \left[2\pi m \left(k\bar{T}\right)^3\right]^{1/2} \tag{1.188}$$

which is even independent of the momentum p. This result can now be used with the Fokker-Planck equation (1.180) to calculate how a non-equilibrium phase-space distribution f evolves in time towards its equilibrium by collisions with heavier particles.

Suggested further reading: [1, 2, 3, 4]