Classical Field Theory

Uwe-Jens Wiese
Institute for Theoretical Physics
Bern University

May 25, 2009

Contents

Chapter 1

Introduction

In this chapter we give a brief introduction to classical field theory and we relate it to current problems in modern physics.

1.1 The Cube of Physics

In order to orient ourselves in the space of physics theories let us consider what one might call the "cube of physics". Classical field theory — the subject of this course — has its well-deserved place in the space of all theories. Theory space can be spanned by three axes labelled with the three most fundamental constants of Nature: Newton's gravitational constant

$$G = 6.6720 \times 10^{-11} \text{kg}^{-1} \text{m}^3 \text{sec}^{-2}, \tag{1.1.1}$$

the velocity of light

$$c = 2.99792456 \times 10^8 \text{m sec}^{-1}, \tag{1.1.2}$$

(which would deserve the name Einstein's constant), and Planck's quantum

$$h = 6.6205 \times 10^{-34} \text{kg m}^2 \text{sec}^{-1}. \tag{1.1.3}$$

Actually, it is most convenient to label the three axes by G, $1/c$, and h. For a long time it was not known that light travels at a finite speed or that there is quantum mechanics. In particular, Newton's classical mechanics corresponds to $c = \infty \;\Rightarrow\; 1/c = 0$ and $h = 0$, i.e. it is non-relativistic and non-quantum, and it thus takes place along the G-axis. If we also ignore gravity and put

$G = 0$ we are at the origin of theory space doing Newton's classical mechanics but only with non-gravitational forces. Of course, Newton realized that in Nature $G \neq 0$, but he couldn't take into account $1/c \neq 0$ or $h \neq 0$. Maxwell and the other fathers of electrodynamics had c built into their theory as a fundamental constant, and Einstein realized that Newton's classical mechanics needs to be modified to his theory of special relativity in order to take into account that $1/c \neq 0$. This opened a new dimension in theory space which extends Newton's line of classical mechanics to the plane of relativistic theories. When we take the limit $c \to \infty \;\; \Rightarrow \;\; 1/c \to 0$ special relativity reduces to Newton's non-relativistic classical mechanics. The fact that special relativity replaced non-relativistic classical mechanics does not mean that Newton was wrong. Indeed, his theory emerges from Einstein's in the limit $c \to \infty$, i.e. if light would travel with infinite speed. As far as our everyday experience is concerned this is practically the case, and hence for these purposes Newton's theory is sufficient. There is no need to convince a mechanical engineer to use Einstein's theory because her airplanes are not designed to reach speeds anywhere close to c. Once special relativity was discovered, it became obvious that there must be a theory that takes into account $1/c \neq 0$ and $G \neq 0$ at the same time. After years of hard work, Einstein was able to construct this theory — general relativity — which is a relativistic theory of gravity. The G-$1/c$-plane contains classical (i.e. non-quantum) relativistic and non-relativistic physics. Classical electrodynamics and general relativity are perfectly consistent with one another. They are the most fundamental classical field theories and the main subject of this course.

A third dimension in theory space was discovered by Planck who started quantum mechanics and introduced the fundamental action quantum h. When we put $h = 0$ quantum physics reduces to classical physics. Again, the existence of quantum mechanics does not mean that classical mechanics is wrong. It is, however, incomplete and should not be applied to the microscopic quantum world. In fact, classical mechanics is entirely contained within quantum mechanics as the limit $h \to 0$, just like it is the $c \to \infty$ limit of special relativity. Quantum mechanics was first constructed non-relativistically (i.e. by putting $1/c = 0$). When we allow $h \neq 0$ as well as $1/c \neq 0$ (but put $G = 0$) we are doing relativistic quantum physics. This is where the quantum version of classical electrodynamics — quantum electrodynamics (QED) — is located in theory space. Also the entire standard model of elementary particle physics which includes QED as well as its analog for the strong force — quantum chromodynamics (QCD) — is located there. Today we know that there must be a consistent physical theory that allows $G \neq 0$, $1/c \neq 0$, and $h \neq 0$ all at the same time. However, this theory of relativistic quantum gravity has not yet been found, although there are some promising attempts using string theory.

1.2 Field Theory

Unlike the weak and strong nuclear forces which play a role only at distances as short as 1 fm = 10^{-15} m, gravity and electromagnetism manifest themselves at macroscopic scales. This implies that, while the weak and strong nuclear forces must be treated quantum mechanically, gravity and electromagnetism can already be investigated using classical (i.e. non-quantum) physics. The investigation of classical electrodynamics has led to the introduction of a powerful fundamental concept — the concept of fields. It has been discovered experimentally that "empty" space is endowed with physical entities that do not manifest themselves as material particles. Such entities are e.g. the electric and magnetic field. Each point in space and time has two vectors $\vec{E}(\vec{x},t)$ and $\vec{B}(\vec{x},t)$ attached to it. The electric and magnetic fields $\vec{E}(\vec{x},t)$ and $\vec{B}(\vec{x},t)$ represent physical degrees of freedom, 3+3 = 6 for each spatial point $\vec{x}$. The physical reality of these abstract fields has been verified in numerous experiments. We should hence think of "empty" space as a medium with nontrivial physical properties. Maxwell's equations take the form

$$\begin{aligned}
&\vec{\nabla} \cdot \vec{E}(\vec{x},t) = 4\pi\rho(\vec{x},t), \\
&\vec{\nabla} \times \vec{E}(\vec{x},t) + \frac{1}{c}\partial_t \vec{B}(\vec{x},t) = 0, \\
&\vec{\nabla} \cdot \vec{B}(\vec{x},t) = 0, \\
&\vec{\nabla} \times \vec{B}(\vec{x},t) - \frac{1}{c}\partial_t \vec{E}(\vec{x},t) = \frac{4\pi}{c}\vec{j}(\vec{x},t).
\end{aligned} \tag{1.2.1}$$

They are four coupled partial differential equations that allow us to determine the electromagnetic fields $\vec{E}(\vec{x},t)$ and $\vec{B}(\vec{x},t)$ once the electric charge and current densities $\rho(\vec{x},t)$ and $\vec{j}(\vec{x},t)$ have been specified and some appropriate boundary and initial conditions have been imposed. These four equations describe all electromagnetic phenomena at the classical level, ranging from Coulomb's $1/r^2$ law to the origin of light. They summarize all results ever obtained over centuries of intense experimentation with electromagnetism, and they are the basis of countless electronic devices that we use every day. Maxwell's equations are an enormous achievement of nineteenth century physics. Numerous discoveries of the twentieth century would have been completely impossible to make without that prior knowledge.

The field concept has turned out to be of central importance far beyond classical electrodynamics. Also general relativity — Einstein's relativistic generalization of Newton's theory of gravity — is a field theory, and the quantum mechanical standard model of elementary particles is a so-called quantum field theory. In fact, the quantum mechanical version of electrodynamics — quantum electrodynamics (QED) — is an integral part of the standard model. Applying the field concept to the classical dynamics of the electromagnetic field as well as to gravity will allow us to understand numerous phenomena ranging from the origin of light to the cosmological evolution. Furthermore, it will be a first step towards reaching the frontier of knowledge in fundamental physics which is currently defined by the standard model of particle physics.

1.3 Electrodynamics and Relativity

Electrodynamics, the theory of electric and magnetic fields, is also a theory of light. Light waves, excitations of the electromagnetic field, travel with the velocity c. Remarkably, as was found experimentally by Michelson and Morley in 1895, light always travels with the speed c, independent of the speed of an observer. For example, if a space ship leaves the earth at a speed $c/2$ and shines light back to the earth, we see the light approaching us with the velocity c, not with $c/2$. This was correctly described by Einstein in his theory of special relativity in 1905. Nevertheless, as a theory of light, Maxwell's equations of 1865 had relativity built in from the start. It was soon realized, e.g. by Lorentz and Poincaré, that there is a fundamental conflict between electrodynamics and Newton's classical mechanics. Both theories assume different fundamental structures of space and time and are thus inconsistent. Only after Einstein generalized Newton's classical mechanics to his theory of special relativity, this conflict was resolved. In Einstein's theory of special relativity space and time are unified to Minkowski's space-time. In particular, Maxwell's equations are invariant against Lorentz transformations, rotations between spatial and temporal directions in space-time.

A point in space-time $x = (\vec{x}, t)$ can be described by a 4-vector

$$x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z). \tag{1.3.1}$$

Remarkably, time (multiplied by the velocity of light) acts as a fourth coordinate in space-time. There are two space-time versions of the Nabla operator

$$\partial_\mu = (\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}), \; \partial^\mu = (\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}). \tag{1.3.2}$$

In the relativistic formulation of classical electrodynamics, the charge density $\rho(\vec{x}, t)$ and the current density $\vec{j}(\vec{x}, t)$ are combined to a 4-vector

$$j^\mu(x) = (c\rho(\vec{x}, t), j_x(\vec{x}, t), j_y(\vec{x}, t), j_z(\vec{x}, t)). \tag{1.3.3}$$

Similarly, another 4-vector

$$A^\mu(x) = (\Phi(\vec{x},t), A_x(\vec{x},t), A_y(\vec{x},t), A_z(\vec{x},t)) \tag{1.3.4}$$

is constructed from the scalar potential $\Phi(\vec{x},t)$ and the vector potential $\vec{A}(\vec{x},t)$. It can be used to construct the so-called field strength tensor

$$F^{\mu\nu}(x) = \partial^\mu A^\nu(x) - \partial^\nu A^\mu(x) \tag{1.3.5}$$

whose components will turn out to be the electric and magnetic fields. The relativistic formulation of classical electrodynamics gives rise to a very compact form of Maxwell's equations

$$\partial_\mu F^{\mu\nu}(x) = \frac{4\pi}{c} j^\nu(x). \tag{1.3.6}$$

Here we have used Einstein's summation convention according to which repeated indices are summed over.

1.4 General Relativity, Cosmology, and Inflation

Another fundamental classical field theory is general relativity — Einstein's relativistic theory of gravity. In this theory space-time itself has non-trivial dynamical properties. In particular, it can be curved by matter, radiation, or other forms of energy such as e.g. vacuum energy. General relativity is governed by Einstein's equation of motion

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.4.1}$$

Here $G_{\mu\nu}$ is the so-called Einstein tensor, G is Newton's constant, and $T_{\mu\nu}$ is the energy-momentum tensor. The essence of this equation is that mass (and more generally any form of energy) is a source of space-time curvature. Applied to the Universe on the largest scales, general relativity predicts the cosmic evolution. Starting from the big bang, depending on its energy content, the Universe expands forever or will eventually recollapse. Recent data on supernova explosions and on the cosmic microwave background radiation indicate that "empty" space contains a uniformly distributed form of energy — so-called vacuum energy (also known as a cosmological constant or in a more dynamical variant as quintessence). If the Universe is dominated by vacuum energy it will expand forever.

Vacuum energy may also have played an important role immediately after the big bang. The idea of the inflationary Universe is based on a phase of exponential expansion at very early times. The expansion is caused by a scalar field whose value is slowly rolling to the classical minimum of its potential. The idea of inflation solves several problems of the standard big bang cosmology. For example, it explains (more or less naturally) why the Universe is old and flat. The course ends with a study of classical scalar field theory in the context of inflation.

Chapter 2

Relativistic Form of Electrodynamics

The relativistic nature of Maxwell's equations is not manifest in their original form. Here we reformulate electrodynamics such that its invariance under Lorentz transformations — i.e. under rotations in Minkowski space-time — becomes manifest. This will lead us to a deeper understanding of electromagnetism and will also pave the way to the discussion of general relativity.

2.1 Euclidean Space and Minkowski Space-Time

Everyday experience confirms that we live in a 3-d Euclidean space. This means that space itself is flat (i.e. not curved), homogeneous (i.e. translation invariant), and isotropic (i.e. rotation invariant). The length squared of a vector $\vec{r} = (x, y, z)$ is given by

$$r^2 = |\vec{r}|^2 = x^2 + y^2 + z^2. \quad (2.1.1)$$

Obviously, the length is invariant under spatial rotations or reflections, i.e.

$$r'^2 = |\vec{r}\,'|^2 = |\vec{r}|^2 = r^2, \quad (2.1.2)$$

where

$$\vec{r}\,' = O\vec{r}. \quad (2.1.3)$$

Here O is a 3×3 orthogonal matrix describing rotations or refections. The orthogonality condition takes the form

$$O^T O = O O^T = \mathbb{1}, \quad (2.1.4)$$

i.e. the transposed matrix O^T is the inverse of the matrix O. A simple example is a rotation by an angle α around the z-axis. In that case the rotation or reflection matrix takes the form

$$O = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.1.5}$$

The Euclidean distance squared between two points $\vec{r}_a = (x_a, y_a, z_a)$ and $\vec{r}_b = (x_b, y_b, z_b)$ is given by

$$(\Delta r)^2 = |\vec{r}_a - \vec{r}_b|^2 = (x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2. \tag{2.1.6}$$

This distance is invariant under spatial translations and rotations or reflections, i.e.

$$(\Delta r')^2 = |\vec{r}_a\,' - \vec{r}_b\,'|^2 = |\vec{r}_a - \vec{r}_b|^2 = (\Delta r)^2, \tag{2.1.7}$$

where

$$\vec{r}_a\,' = O\vec{r}_a + \vec{d},\ \vec{r}_b\,' = O\vec{r}_b + \vec{d}, \tag{2.1.8}$$

Here $\vec{d}$ is a displacement vector and O is again an orthogonal rotation or reflection matrix.

Relativistic theories such as Einstein's special relativity or Maxwell's electrodynamics are invariant not only under spatial translations and rotations but also under Lorentz transformations that rotate spatial and temporal coordinates into one another. The resulting enlarged invariance is known as Lorentz invariance (against space-time rotations) or as Poincaré invariance against space-time rotations and translations. Minkowski was first to realize that in relativistic theories space and time (which are separate entities in Newtonian mechanics) are naturally united to space-time. A point in Minkowski space-time is described by four coordinates — one for time and three for space — which form a so-called 4-vector

$$x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z). \tag{2.1.9}$$

In particular, the time t (multiplied by the velocity of light c) plays the role of the zeroth component of the 4-vector. Minkowski's space-time does not have Euclidean geometry. In particular, the length squared of the 4-vector x^μ is given by

$$s^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2, \tag{2.1.10}$$

and may thus be negative. Besides the so-called contra-variant vector x^μ it is useful to introduce an equivalent form

$$x_\mu = (x_0, x_1, x_2, x_3) = (ct, -x, -y, -z), \tag{2.1.11}$$

the so-called co-variant 4-vector. Both the co- and the contra-variant 4-vectors contain the same physical information. Their components are simply related by

$$x_0 = x^0 = ct,\ x_1 = -x^1 = -x,\ x_2 = -x^2 = -y,\ x_3 = -x^3 = -z. \tag{2.1.12}$$

The length squared of the 4-vector can then be written as

$$\sum_{\mu=0}^{3} x_\mu x^\mu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = s^2. \tag{2.1.13}$$

Instead of always writing sums over space-time indices μ explicitly, Einstein has introduced a summation convention according to which repeated indices (one co- and one contra-variant index) will automatically be summed. Using Einstein's summation convention the above equation simply takes the form

$$x_\mu x^\mu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = s^2. \tag{2.1.14}$$

The summation over μ is no longer written explicitly, but is still implicitly understood, because the index μ occurs twice (once as a co- and once as a contra-variant one).

The norm of a 4-vector induces a corresponding metric via

$$(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} g_{\mu\nu} x^\mu x^\nu = g_{\mu\nu} x^\mu x^\nu. \tag{2.1.15}$$

In the last step, we have again used Einstein's summation convention and have dropped the explicit sums over the repeated indices μ and ν. The metric tensor g with the elements $g_{\mu\nu}$ is a 4×4 matrix given by

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.1.16}$$

The metric tensor can also be used to relate co- and contra-variant 4-vectors by lowering a contra-variant index, i.e.

$$x_\mu = g_{\mu\nu} x^\nu = \sum_{\nu=0}^{3} g_{\mu\nu} x^\nu. \tag{2.1.17}$$

Again, the repeated index ν is summed over, while the unrepeated index μ is not summed. Let us also introduce the inverse metric g^{-1} with the components $g^{\mu\nu}$

which is given by

$$g^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{2.1.18}$$

and which obeys

$$gg^{-1} = \mathbb{1}. \tag{2.1.19}$$

In components this relation takes the form

$$g_{\mu\nu}g^{\nu\rho} = \sum_{\nu=0}^{3} g_{\mu\nu}g^{\nu\rho} = \delta_\mu{}^\rho, \tag{2.1.20}$$

where $\delta_\mu{}^\rho$ is just the Kronecker symbol, i.e. it represents the matrix elements of the unit matrix $\mathbb{1}$. The inverse metric can now be used to raise co-variant indices, e.g.

$$x^\mu = g^{\mu\nu}x_\nu = \sum_{\nu=0}^{3} g^{\mu\nu}x_\nu. \tag{2.1.21}$$

Let us ask under what kind of rotations the length squared of a 4-vector is invariant. The rotated 4-vector can be written as

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu, \tag{2.1.22}$$

where Λ is a 4×4 space-time rotation matrix. Similarly, we obtain

$$x'_\mu = g_{\mu\nu}x'^\nu = g_{\mu\nu}\Lambda^\nu{}_\rho x^\rho. \tag{2.1.23}$$

The length squared of x'^μ is then given by

$$s'^2 = x'_\mu x'^\mu = g_{\mu\nu}\Lambda^\nu{}_\rho x^\rho \Lambda^\mu{}_\sigma x^\sigma. \tag{2.1.24}$$

It is invariant under space-time rotations, i.e. $s'^2 = s^2$, only if

$$g_{\mu\nu}\Lambda^\nu{}_\rho \Lambda^\mu{}_\sigma = g_{\rho\sigma}. \tag{2.1.25}$$

This can be rewritten as

$$\Lambda^T_\sigma{}^\mu g_{\mu\nu}\Lambda^\nu{}_\rho = g^T_{\sigma\rho}, \tag{2.1.26}$$

or equivalently as the matrix multiplication

$$\Lambda^T g \Lambda = g^T = g. \tag{2.1.27}$$

This condition is the Minkowski space-time analog of the Euclidean space condition $O^T O = \mathbb{1}$ for orthogonal spatial rotations. One now obtains

$$\begin{aligned} x'_\mu &= g_{\mu\nu}\Lambda^\nu{}_\rho x^\rho = g_{\mu\nu}\Lambda^\nu{}_\rho g^{\rho\sigma} x_\sigma = [g\Lambda g^{-1}]_\mu{}^\sigma x_\sigma = x_\sigma [g\Lambda g^{-1}]^{T\sigma}{}_\mu \\ &= x_\sigma \Lambda^{-1\sigma}{}_\mu. \end{aligned} \quad (2.1.28)$$

Here we have used eq.(2.1.27) which leads to

$$[g\Lambda g^{-1}]^T = g^{-1}\Lambda^T g = \Lambda^{-1}. \quad (2.1.29)$$

Finally, as a consequence of eq.(2.1.28) we obtain

$$x_\nu = x'_\mu \Lambda^\mu{}_\nu. \quad (2.1.30)$$

Space-time rotations which obey eq.(2.1.27) are known as Lorentz transformations. A simple example of a Lorentz transformation is a rotation between time and the x-direction, which leaves the y- and z-coordinates unchanged. This transformation is given by

$$\Lambda = \begin{pmatrix} \cosh\alpha & -\sinh\alpha & 0 & 0 \\ -\sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.1.31)$$

In components, this takes the form

$$ct' = ct\cosh\alpha - x\sinh\alpha, \; x' = x\cosh\alpha - ct\sinh\alpha, \; y' = y, \; z' = z. \quad (2.1.32)$$

Introducing

$$\beta = \frac{v}{c}, \; \gamma = \frac{1}{\sqrt{1-\beta^2}} = \frac{1}{\sqrt{1-v^2/c^2}} \quad (2.1.33)$$

and identifying

$$\cosh\alpha = \gamma, \; \sinh\alpha = \sqrt{\cosh^2\alpha - 1} = \frac{v/c}{\sqrt{1-v^2/c^2}} = \beta\gamma, \quad (2.1.34)$$

one indeed arrives at the familiar Lorentz transformation

$$t' = \frac{t - xv/c^2}{\sqrt{1-v^2/c^2}}, \; x' = \frac{x - vt}{\sqrt{1-v^2/c^2}}, \; y' = y, \; z' = z. \quad (2.1.35)$$

Lorentz transformations form a group known as the Lorentz group.

The distance squared in space-time between two 4-vectors x_a^μ and x_b^μ is given by

$$(\Delta s)^2 = (x_a^0 - x_b^0)^2 - (x_a^1 - x_b^1)^2 - (x_a^2 - x_b^2)^2 - (x_a^3 - x_b^3)^2, \quad (2.1.36)$$

and may again be negative. This distance is invariant under both Lorentz transformations Λ and space-time translations d^μ, i.e. $(\Delta s')^2 = (\Delta s)^2$, with

$$x_a'^{\,\mu} = \Lambda^\mu{}_\nu x_a^\nu + d^\mu, \; x_b'^{\,\mu} = \Lambda^\mu{}_\nu x_b^\nu + d^\mu. \quad (2.1.37)$$

Lorentz transformations and space-time translations again form a group — the so-called Poincaré group — which contains the Lorentz group as a subgroup.

2.2 Gradient as a 4-Vector and d'Alembert Operator

In order to do field theory in manifestly Lorentz-invariant form, we also need to combine temporal and spatial derivatives to a 4-vector. Let us introduce

$$\partial^\mu = (\partial^0, \partial^1, \partial^2, \partial^3) = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right) = \left(\frac{\partial}{\partial x_0}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right). \quad (2.2.1)$$

How does this object transform under Lorentz transformations? Using eq.(2.1.30) one obtains

$$\partial'^\mu = \frac{\partial}{\partial x'_\mu} = \frac{\partial x_\nu}{\partial x'_\mu}\frac{\partial}{\partial x_\nu} = \Lambda^\mu{}_\nu \partial^\nu. \quad (2.2.2)$$

This shows that ∂^μ indeed transforms as a contra-variant 4-vector. Similarly, one can define the co-variant 4-vector

$$\partial_\mu = (\partial_0, \partial_1, \partial_2, \partial_3) = \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right). \quad (2.2.3)$$

One can form the scalar product of the co- and contra-variant derivative 4-vectors. In this way one obtains a second derivative operator which transforms as a space-time scalar, i.e. it is invariant under Lorentz transformations. This Minkowski space-time analog of the Laplace operator in Euclidean space is known as the d'Alembert operator and is given by

$$\Box = \partial_\mu \partial^\mu = \frac{1}{c^2}\frac{\partial}{\partial t^2} - \frac{\partial}{\partial x^2} - \frac{\partial}{\partial y^2} - \frac{\partial}{\partial z^2}. \quad (2.2.4)$$

2.3 4-Vector Current

Let us now begin to express electrodynamics in manifestly Lorentz co-variant form. We proceed step by step and begin with the charge and current densities $\rho(\vec{r},t)$ and $\vec{j}(\vec{r},t)$. The corresponding continuity equation which expresses charge conservation takes the form

$$\partial_t \rho(\vec{r},t) + \vec{\nabla} \cdot \vec{j}(\vec{r},t) = 0. \tag{2.3.1}$$

Since temporal and spatial derivatives are combined to a gradient 4-vector, it is natural to combine $\rho(\vec{r},t)$ and $\vec{j}(\vec{r},t)$ to the current 4-vector

$$j^\mu(x) = (c\rho(\vec{r},t), j_x(\vec{r},t), j_y(\vec{r},t), j_z(\vec{r},t)). \tag{2.3.2}$$

Here we have introduced $x = (\vec{r},t)$ as a short-hand notation for a point in space-time. It goes without saying that this should not be confused with the x-component of the spatial vector $\vec{r}$. In Lorentz-invariant form the continuity equation thus takes the form

$$\partial_\mu j^\mu(x) = \frac{1}{c}\partial_t c\rho(\vec{r},t) + \partial_x j_x(\vec{r},t) + \partial_y j_y(\vec{r},t) + \partial_z j_z(\vec{r},t) = 0. \tag{2.3.3}$$

Here we have combined the co-variant 4-vector ∂_μ and the contra-variant 4-vector $j^\mu(x)$ to the Lorentz-scalar zero. The Lorentz invariance of the continuity equation implies that charge conservation is valid in any inertial frame, independent of the motion of an observer.

Of course, the charge and current densities themselves are dependent on the reference frame in which they are considered. For example, let us consider a uniform static charge distribution $\rho(\vec{r},t) = \rho$. How is this charge distribution described by an observer in another reference frame moving with velocity v in the x-direction? In order to answer this question, we simply perform the corresponding Lorentz transformation

$$j'^\mu = \Lambda^\mu{}_\nu j^\nu. \tag{2.3.4}$$

In components this equation takes the form

$$\begin{pmatrix} c\rho' \\ j'_x \\ j'_y \\ j'_z \end{pmatrix} = \begin{pmatrix} \cosh\alpha & -\sinh\alpha & 0 & 0 \\ -\sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\rho \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{2.3.5}$$

such that

$$\rho' = \rho \cosh \alpha = \frac{\rho}{\sqrt{1 - v^2/c^2}}, \ j'_x = -c\rho \sinh \alpha = -\frac{v\rho}{\sqrt{1 - v^2/c^2}} = -v\rho',$$
$$j'_y = j'_z = 0. \tag{2.3.6}$$

Of course, it is no surprise that the moving observer sees the static charge distribution as a current flowing opposite to his direction of motion. It may be less obvious why he sees an enhanced charge density $\rho' = \gamma\rho$. Doesn't this mean that the moving observer sees a larger total charge, which would violate charge conservation? We should not forget that a moving observer sees a given length L_x contracted to

$$L'_x = \sqrt{1 - \frac{v^2}{c^2}} L_x = \frac{L_x}{\gamma} \tag{2.3.7}$$

in the direction of motion, while transverse lengths remain the same, i.e. $L'_y = L_y$, $L'_z = L_z$. Hence, the total charge contained in the volume $V' = L'_x L'_y L'_z$ is given by

$$Q' = \rho' V' = \gamma\rho L'_x L'_y L'_z = \rho L_x L_y L_z = \rho V = Q, \tag{2.3.8}$$

and is hence indeed independent of the reference frame.

If a general (non-uniform and non-static) charge and current density is transformed into another reference frame, one must also transform the space-time point x at which the density is evaluated, i.e.

$$j'^{\mu}(x') = \Lambda^{\mu}{}_{\nu} j^{\nu}(x) = \Lambda^{\mu}{}_{\nu} j^{\nu}(\Lambda^{-1}x'). \tag{2.3.9}$$

2.4 4-Vector Potential

From the scalar potential $\Phi(\vec{r}, t)$ and the vector potential $\vec{A}(\vec{r}, t)$ one can construct another 4-vector field

$$A^{\mu}(x) = (\Phi(\vec{r}, t), A_x(\vec{r}, t), A_y(\vec{r}, t), A_z(\vec{r}, t)). \tag{2.4.1}$$

Under Lorentz transformations it again transforms as

$$A'^{\mu}(x') = \Lambda^{\mu}{}_{\nu} A^{\nu}(\Lambda^{-1}x'). \tag{2.4.2}$$

Scalar and vector potentials transform non-trivially under gauge transformations

$${}^{\varphi}\Phi(\vec{r}, t) = \Phi(\vec{r}, t) + \frac{1}{c}\partial_t \varphi(\vec{r}, t), \ {}^{\varphi}\vec{A}(\vec{r}, t) = \vec{A}(\vec{r}, t) - \vec{\nabla}\varphi(\vec{r}, t). \tag{2.4.3}$$

In 4-vector notation this relation takes the form

$$^{\varphi}A^{\mu}(x) = A^{\mu}(x) + \partial^{\mu}\varphi(x). \tag{2.4.4}$$

Here $\varphi(x)$ is an arbitrary space-time-dependent gauge transformation function. This function is a space-time scalar, i.e. under Lorentz transformations it transforms as

$$\varphi'(x') = \varphi(\Lambda^{-1}x'). \tag{2.4.5}$$

The Lorenz gauge fixing condition

$$\frac{1}{c}\partial_t\Phi(\vec{r},t) + \vec{\nabla}\cdot\vec{A}(\vec{r},t) = 0 \tag{2.4.6}$$

can be re-expressed as

$$\partial_{\mu}A^{\mu}(x) = 0. \tag{2.4.7}$$

In the Lorenz gauge the wave equations take the form

$$\frac{1}{c^2}\partial_t^2\Phi(\vec{r},t) - \Delta\Phi(\vec{r},t) = 4\pi\rho(\vec{r},t), \tag{2.4.8}$$

$$\frac{1}{c^2}\partial_t^2\vec{A}(\vec{r},t) - \Delta\vec{A}(\vec{r},t) = \frac{4\pi}{c}\vec{j}(\vec{r},t), \tag{2.4.9}$$

which can be combined to

$$\Box A^{\mu}(x) = \frac{4\pi}{c}j^{\mu}(x). \tag{2.4.10}$$

A plane wave solution of the wave equation in the vacuum takes the form

$$A^{\mu}(x) = C^{\mu}\exp(ik_{\nu}x^{\nu}) = C^{\mu}\exp(i(\omega t - \vec{k}\cdot\vec{r})), \tag{2.4.11}$$

with the co- and contra-variant wave 4-vectors

$$k_{\mu} = (k_0, k_1, k_2, k_3) = (\frac{\omega}{c}, -k_x, -k_y, -k_z),$$
$$k^{\mu} = (k^0, k^1, k^2, k^3) = (\frac{\omega}{c}, k_x, k_y, k_z). \tag{2.4.12}$$

Inserting the plane wave ansatz into the Lorenz gauge condition one obtains

$$\partial_{\mu}A^{\mu}(x) = ik_{\mu}C^{\mu}\exp(ik_{\nu}x^{\nu}) = 0 \;\Rightarrow k_{\mu}C^{\mu} = 0. \tag{2.4.13}$$

Similarly, inserting the ansatz in the wave equation in the vacuum one finds

$$\Box A^{\mu}(x) = -k_{\rho}k^{\rho}C^{\mu}\exp(ik_{\nu}x^{\nu}) = 0 \;\Rightarrow\; k_{\rho}k^{\rho} = \frac{\omega^2}{c^2} - |\vec{k}|^2 = 0 \;\Rightarrow\; \omega = |\vec{k}|c. \tag{2.4.14}$$

The 4-vector k^{μ} has norm zero, i.e. it is a so-called null-vector.

2.5 Field Strength Tensor

It may not be entirely obvious how to express the electric and magnetic fields $\vec{E}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$ in relativistic form. We need to use

$$\vec{E}(\vec{r},t) = -\vec{\nabla}\Phi(\vec{r},t) - \frac{1}{c}\partial_t \vec{A}(\vec{r},t), \ \vec{B}(\vec{r},t) = \vec{\nabla} \times \vec{A}(\vec{r},t). \tag{2.5.1}$$

Obviously, $\vec{E}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$ are constructed form the 4-vectors ∂^μ and $A^\mu(x)$. The scalar product of these two 4-vectors

$$\partial_\mu A^\mu(x) = \frac{1}{c}\partial_t \Phi(\vec{r},t) + \vec{\nabla} \cdot \vec{A}(\vec{r},t) \tag{2.5.2}$$

appears in the Lorenz gauge fixing condition but does not yield the electric or magnetic field. The 4-vectors ∂^μ and $A^\mu(x)$ can also be combined to the symmetric tensor

$$D^{\mu\nu} = \partial^\mu A^\nu(x) + \partial^\nu A^\mu(x), \tag{2.5.3}$$

as well as to the anti-symmetric tensor

$$F^{\mu\nu} = \partial^\mu A^\nu(x) - \partial^\nu A^\mu(x). \tag{2.5.4}$$

Under gauge transformations these tensors transform as

$$\begin{aligned}
{}^\varphi D^{\mu\nu} &= \partial^\mu \, {}^\varphi A^\nu(x) + \partial^\nu \, {}^\varphi A^\mu(x) \\
&= \partial^\mu A^\nu(x) + \partial^\mu \partial^\nu \varphi(x) + \partial^\nu A^\mu(x) + \partial^\nu \partial^\mu \varphi(x) \\
&= D^{\mu\nu} + 2\partial^\mu \partial^\nu \varphi(x), \\
{}^\varphi F^{\mu\nu} &= \partial^\mu \, {}^\varphi A^\nu(x) - \partial^\nu \, {}^\varphi A^\mu(x) \\
&= \partial^\mu A^\nu(x) + \partial^\mu \partial^\nu \varphi(x) - \partial^\nu A^\mu(x) - \partial^\nu \partial^\mu \varphi(x) \\
&= F^{\mu\nu},
\end{aligned} \tag{2.5.5}$$

i.e. the anti-symmetric tensor $F^{\mu\nu}$ is gauge invariant, while the symmetric tensor $D^{\mu\nu}$ is not. As a consequence, it does not play any particular role in electrodynamics. Since the electromagnetic fields are gauge invariant, we expect them to be related to $F^{\mu\nu}$. Let us consider the various components of this tensor

$$\begin{aligned}
F^{01}(x) &= \partial^0 A^1(x) - \partial^1 A^0(x) = \frac{1}{c}\partial_t A_x(\vec{r},t) + \partial_x \Phi(\vec{r},t) = -E_x(\vec{r},t), \\
F^{02}(x) &= \partial^0 A^2(x) - \partial^2 A^0(x) = \frac{1}{c}\partial_t A_y(\vec{r},t) + \partial_y \Phi(\vec{r},t) = -E_y(\vec{r},t), \\
F^{03}(x) &= \partial^0 A^3(x) - \partial^3 A^0(x) = \frac{1}{c}\partial_t A_z(\vec{r},t) + \partial_z \Phi(\vec{r},t) = -E_z(\vec{r},t), \\
F^{12}(x) &= \partial^1 A^2(x) - \partial^2 A^1(x) = -\partial_x A_y(\vec{r},t) + \partial_y A_x(\vec{r},t) = -B_z(\vec{r},t), \\
F^{23}(x) &= \partial^2 A^3(x) - \partial^3 A^2(x) = -\partial_y A_z(\vec{r},t) + \partial_z A_y(\vec{r},t) = -B_x(\vec{r},t), \\
F^{31}(x) &= \partial^3 A^1(x) - \partial^1 A^3(x) = -\partial_z A_x(\vec{r},t) + \partial_x A_z(\vec{r},t) = -B_y(\vec{r},t).
\end{aligned} \tag{2.5.6}$$

Hence, the anti-symmetric tensor indeed contains the electric and magnetic fields as

$$F^{\mu\nu}(x) = \begin{pmatrix} 0 & -E_x(\vec{r},t) & -E_y(\vec{r},t) & -E_z(\vec{r},t) \\ E_x(\vec{r},t) & 0 & -B_z(\vec{r},t) & B_y(\vec{r},t) \\ E_y(\vec{r},t) & B_z(\vec{r},t) & 0 & -B_x(\vec{r},t) \\ E_z(\vec{r},t) & -B_y(\vec{r},t) & B_x(\vec{r},t) & 0 \end{pmatrix}. \tag{2.5.7}$$

The co-variant components of this tensor are given by

$$F_{\mu\nu}(x) = \begin{pmatrix} 0 & E_x(\vec{r},t) & E_y(\vec{r},t) & E_z(\vec{r},t) \\ -E_x(\vec{r},t) & 0 & -B_z(\vec{r},t) & B_y(\vec{r},t) \\ -E_y(\vec{r},t) & B_z(\vec{r},t) & 0 & -B_x(\vec{r},t) \\ -E_z(\vec{r},t) & -B_y(\vec{r},t) & B_x(\vec{r},t) & 0 \end{pmatrix}. \tag{2.5.8}$$

2.6 Inhomogeneous Maxwell Equations

Let us first consider the inhomogeneous Maxwell equations

$$\vec{\nabla} \cdot \vec{E}(\vec{r},t) = 4\pi\rho(\vec{r},t), \ \vec{\nabla} \times \vec{B}(\vec{r},t) - \frac{1}{c}\partial_t \vec{E}(\vec{r},t) = \frac{4\pi}{c}\vec{j}(\vec{r},t). \tag{2.6.1}$$

These are four equations with the components of the 4-vector current $j^\mu(x)$ on the right-hand side. Hence, on the left-hand side there must also be a 4-vector. The left-hand side consists of derivatives, i.e. of components of the gradient 4-vectors ∂_μ, and of the electromagnetic fields, i.e. of the components of the field strength tensor $F^{\mu\nu}$. Hence, the 4-vector ∂_μ and the tensor $F^{\mu\nu}$ on the left-hand side must be combined to another 4-vector. This can be achieved by forming $\partial_\mu F^{\mu\nu}(x)$ and thus by contracting (i.e. by summing) one co- and one contra-variant index. The various components of this object take the form

$$
\begin{aligned}
\partial_\mu F^{\mu 0}(x) &= \partial_x E_x(\vec{r},t) + \partial_y E_y(\vec{r},t) + \partial_z E_z(\vec{r},t) \\
&= \vec{\nabla} \cdot \vec{E}(\vec{r},t) = 4\pi\rho(\vec{r},t), \\
\partial_\mu F^{\mu 1}(x) &= -\frac{1}{c}\partial_t E_x(\vec{r},t) + \partial_y B_z(\vec{r},t) - \partial_z B_y(\vec{r},t) \\
&= [\vec{\nabla} \times \vec{B}]_x(\vec{r},t) - \frac{1}{c}\partial_t E_x(\vec{r},t) = \frac{4\pi}{c} j_x(\vec{r},t), \\
\partial_\mu F^{\mu 2}(x) &= -\frac{1}{c}\partial_t E_y(\vec{r},t) - \partial_x B_z(\vec{r},t) + \partial_z B_x(\vec{r},t) \\
&= [\vec{\nabla} \times \vec{B}]_y(\vec{r},t) - \frac{1}{c}\partial_t E_y(\vec{r},t) = \frac{4\pi}{c} j_y(\vec{r},t), \\
\partial_\mu F^{\mu 3}(x) &= -\frac{1}{c}\partial_t E_z(\vec{r},t) + \partial_x B_y(\vec{r},t) - \partial_y B_x(\vec{r},t) \\
&= [\vec{\nabla} \times \vec{B}]_z(\vec{r},t) - \frac{1}{c}\partial_t E_z(\vec{r},t) = \frac{4\pi}{c} j_z(\vec{r},t). \qquad (2.6.2)
\end{aligned}
$$

Indeed, these equations can be summarized as

$$\partial_\mu F^{\mu\nu}(x) = \frac{4\pi}{c} j^\nu(x). \qquad (2.6.3)$$

At this point the usefulness of the compact 4-dimensional notation should be obvious.

Inserting eq.(2.5.4) into the inhomogeneous Maxwell equations we obtain

$$\partial_\mu F^{\mu\nu}(x) = \partial_\mu(\partial^\mu A^\mu(x) - \partial^\nu A^\mu(x)) = \Box A^\nu(x) - \partial^\nu \partial_\mu A^\mu(x) = \frac{4\pi}{c} j^\nu(x). \quad (2.6.4)$$

If the 4-vector potential obeys the Lorenz gauge fixing condition $\partial_\mu A^\mu(x) = 0$, this is nothing but the wave equation eq.(2.4.10).

2.7 Homogeneous Maxwell Equations

How can we express the homogeneous Maxwell equations

$$\vec{\nabla} \cdot \vec{B}(\vec{r},t) = 0, \ \vec{\nabla} \times \vec{E}(\vec{r},t) + \frac{1}{c}\partial_t \vec{B}(\vec{r},t) = 0 \qquad (2.7.1)$$

in 4-dimensional form? Except for the vanishing right-hand side, they look very similar to the inhomogeneous equations. All we need to do is to substitute $\vec{E}(\vec{r},t)$

by $-\vec{B}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$ by $\vec{E}(\vec{r},t)$. Such a substitution is known as a duality transformation. Under this operation the field strength tensor turns into the dual tensor

$$\widetilde{F}^{\mu\nu}(x) = \begin{pmatrix} 0 & B_x(\vec{r},t) & B_y(\vec{r},t) & B_z(\vec{r},t) \\ -B_x(\vec{r},t) & 0 & -E_z(\vec{r},t) & E_y(\vec{r},t) \\ -B_y(\vec{r},t) & E_z(\vec{r},t) & 0 & -E_x(\vec{r},t) \\ -B_z(\vec{r},t) & -E_y(\vec{r},t) & E_x(\vec{r},t) & 0 \end{pmatrix}, \tag{2.7.2}$$

and the homogeneous Maxwell equations can thus be expressed as

$$\partial_\mu \widetilde{F}^{\mu\nu}(x) = 0. \tag{2.7.3}$$

The co-variant components of the dual field strength tensor take the form

$$\widetilde{F}_{\mu\nu}(x) = \begin{pmatrix} 0 & -B_x(\vec{r},t) & -B_y(\vec{r},t) & -B_z(\vec{r},t) \\ B_x(\vec{r},t) & 0 & -E_z(\vec{r},t) & E_y(\vec{r},t) \\ B_y(\vec{r},t) & E_z(\vec{r},t) & 0 & -E_x(\vec{r},t) \\ B_z(\vec{r},t) & -E_y(\vec{r},t) & E_x(\vec{r},t) & 0 \end{pmatrix}. \tag{2.7.4}$$

It is obvious that the field strength tensor $F^{\mu\nu}(x)$ and its dual $\widetilde{F}_{\mu\nu}(x)$ consist of the same components, namely of the electric and magnetic fields $\vec{E}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$. Hence, there must be a relation between the two tensors. This relation takes the form

$$\widetilde{F}_{\mu\nu}(x) = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}(x). \tag{2.7.5}$$

Here $\epsilon_{\mu\nu\rho\sigma}$ is a totally anti-symmetric tensor with components 0 or ± 1. If any of the indices μ, ν, ρ, and σ are equal, the value of $\epsilon_{\mu\nu\rho\sigma}$ is zero. Only if all indices are different, the value of $\epsilon_{\mu\nu\rho\sigma}$ is non-zero. The value is $\epsilon_{\mu\nu\rho\sigma} = 1$ if $\mu\nu\rho\sigma$ is an even permutation of 0123 (i.e. it requires an even number of index pair permutations to turn $\mu\nu\rho\sigma$ into 0123). Similarly, $\epsilon_{\mu\nu\rho\sigma} = -1$ if $\mu\nu\rho\sigma$ is an odd permutation of 0123. For example, we obtain

$$\begin{aligned} \widetilde{F}_{01}(x) &= \frac{1}{2}\epsilon_{01\rho\sigma}F^{\rho\sigma}(x) = \frac{1}{2}\left(\epsilon_{0123}F^{23}(x) + \epsilon_{0132}F^{32}(x)\right) \\ &= F^{23}(x) = -B_x(\vec{r},t), \end{aligned} \tag{2.7.6}$$

as well as

$$\begin{aligned} \widetilde{F}_{12}(x) &= \frac{1}{2}\epsilon_{12\rho\sigma}F^{\rho\sigma}(x) = \frac{1}{2}\left(\epsilon_{1203}F^{03}(x) + \epsilon_{1230}F^{30}(x)\right) \\ &= F^{03}(x) = -E_z(\vec{r},t), \end{aligned} \tag{2.7.7}$$

Inserting eq.(2.7.5) into the homogeneous Maxwell equations (2.7.3) one obtains

$$\partial^\mu \widetilde{F}_{\mu\nu}(x) = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\partial^\mu F^{\rho\sigma}(x) = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\partial^\mu \left(\partial^\rho A^\sigma(x) - \partial^\sigma A^\rho(x)\right) = 0. \tag{2.7.8}$$

Due to the anti-symmetry of $\epsilon_{\mu\nu\rho\sigma}$ and the commutativity of the derivatives ∂^μ and ∂^ρ, this equation is automatically satisfied. This is no surprise, because the original Maxwell equations were also automatically satisfied by the introduction of the scalar and vector potentials $\Phi(\vec{r}, t)$ and $\vec{A}(\vec{r}, t)$. The homogeneous Maxwell equations can alternatively be expressed as

$$\partial^\mu F^{\rho\sigma} + \partial^\rho F^{\sigma\mu} + \partial^\sigma F^{\mu\rho} = 0. \tag{2.7.9}$$

Indeed, multiplying this relation with $\varepsilon_{\mu\nu\rho\sigma}$ and applying cyclic permutations to the indices μ, ρ, and σ, one again arrives at eq.(2.7.8).

2.8 Space-Time Scalars from Field Strength Tensors

Which scalar quantities can be formed by combining the field strength tensors $F^{\mu\nu}(x)$ and $\widetilde{F}^{\mu\nu}(x)$? First, we can construct the combination

$$F_{\mu\nu}(x)F^{\mu\nu}(x) = 2\left(\vec{B}(\vec{r}, t)^2 - \vec{E}(\vec{r}, t)^2\right), \tag{2.8.1}$$

which will later turn out to be the Lagrange density of electrodynamics. Then we can construct

$$\widetilde{F}_{\mu\nu}(x)\widetilde{F}^{\mu\nu}(x) = 2\left(\vec{E}(\vec{r}, t)^2 - \vec{B}(\vec{r}, t)^2\right), \tag{2.8.2}$$

which is thus the same up to a minus-sign. While the electromagnetic fields themselves are obviously not Lorentz-invariant, the difference of their magnitudes squared is.

One can also mix the field strength tensor with its dual and one then obtains

$$F_{\mu\nu}(x)\widetilde{F}^{\mu\nu}(x) = 4\vec{E}(\vec{r}, t) \cdot \vec{B}(\vec{r}, t). \tag{2.8.3}$$

Similarly, one obtains

$$\widetilde{F}_{\mu\nu}(x)F^{\mu\nu}(x) = 4\vec{E}(\vec{r}, t) \cdot \vec{B}(\vec{r}, t), \tag{2.8.4}$$

which is thus equivalent. Interestingly, the projection of the electric on the magnetic field $\vec{E}(\vec{r}, t) \cdot \vec{B}(\vec{r}, t)$ is also Lorentz invariant, i.e. it has the same value in all inertial frames.

2.9 Transformation of Electromagnetic Fields

Since they form the components of the field strength tensors, it is obvious that the electromagnetic fields $\vec{E}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$ are not Lorentz-invariant, i.e. they depend on the motion of the observer. Under a Lorentz transformation the field strength tensor transforms as

$$F'^{\mu\nu}(x') = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma}(x) = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma}(\Lambda^{-1}x'). \tag{2.9.1}$$

This can be rewritten in matrix form as

$$F'^{\mu\nu}(x') = \Lambda^\mu{}_\rho F^{\rho\sigma}(\Lambda^{-1}x') \Lambda^T{}_\sigma{}^\nu \ \Rightarrow \ F'(x') = \Lambda F(\Lambda^{-1}x')\Lambda^T. \tag{2.9.2}$$

Under the particular Lorentz transformation of eq.(2.1.31) the fields thus transform as

$$\begin{pmatrix} 0 & -E'_x & -E'_y & -E'_z \\ E'_x & 0 & -B'_z & B'_y \\ E'_y & B'_z & 0 & -B'_x \\ E'_z & -B'_y & B'_x & 0 \end{pmatrix} =$$

$$\begin{pmatrix} C & -S & 0 & 0 \\ -S & C & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix} \begin{pmatrix} C & -S & 0 & 0 \\ -S & C & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} C & -S & 0 & 0 \\ -S & C & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E_x S & -E_x C & -E_y & -E_z \\ E_x C & -E_x S & -B_z & B_y \\ E_y C - B_z S & -E_y S + B_z C & 0 & -B_x \\ E_z C + B_y S & -E_z S - B_y C & B_x & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & -E_x & -CE_y + SB_x & -CE_z - SB_y \\ E_x & 0 & SE_y - CB_z & SE_z + CB_y \\ E_y C - B_z S & -E_y S + B_z C & 0 & -B_x \\ E_z C + B_y S & -E_z S - B_y C & B_x & 0 \end{pmatrix}. \tag{2.9.3}$$

Here $C = \cosh\alpha = \gamma$ and $S = \sinh\alpha = \beta\gamma$ and we have used $C^2 - S^2 = 1$. In components we thus obtain

$$E'_x = E_x, \ E'_y = \frac{E_y - B_z v/c}{\sqrt{1 - v^2/c^2}}, \ E'_z = \frac{E_z + B_y v/c}{\sqrt{1 - v^2/c^2}},$$
$$B'_x = B_x, \ B'_y = \frac{B_y + E_z v/c}{\sqrt{1 - v^2/c^2}}, \ B'_z = \frac{B_z - E_y v/c}{\sqrt{1 - v^2/c^2}}. \tag{2.9.4}$$

Interestingly, the longitudinal components E_x and B_x remain the same in an inertial frame moving with velocity $\vec{v} = (v, 0, 0)$ in the x-direction, while the transverse components change.

2.10 Electromagnetic Field of a Moving Charge

Let us consider a static point charge q in its rest frame. The corresponding electromagnetic field is then given by

$$\vec{E}(\vec{r},t) = q\frac{\vec{r}}{|\vec{r}|^3},\ \vec{B}(\vec{r},t) = 0. \tag{2.10.1}$$

What is the electromagnetic field seen by a moving observer? Or equivalently, what is the field of a charge moving in the opposite direction? Let us perform a Lorentz transformation in order to determine the field seen by the moving observer. First of all, we have

$$t = \frac{t' + x'v/c^2}{\sqrt{1-v^2/c^2}},\ x = \frac{x' + vt'}{\sqrt{1-v^2/c^2}},\ y = y',\ z = z', \tag{2.10.2}$$

such that

$$|\vec{r}|^2 = x^2 + y^2 + z^2 = \frac{(x' + vt')^2}{1 - v^2/c^2} + y'^2 + z'^2. \tag{2.10.3}$$

The moving observer sees the charge at the position $(-vt', 0, 0)$ such that the distance vector from the charge to the point $\vec{r}\,' = (x', y', z')$ is given by

$$\vec{R}\,' = (x' + vt', y', z'), \tag{2.10.4}$$

with the length squared

$$R'^2 = (x' + vt')^2 + y'^2 + z'^2, \tag{2.10.5}$$

such that

$$|\vec{r}|^2 = \frac{R'^2 - d^2v^2/c^2}{1 - v^2/c^2}, \tag{2.10.6}$$

where $d^2 = y'^2 + z'^2 = y^2 + z^2$ is the transverse distance squared. Let us now look at the various components of the electromagnetic field. One obtains

$$\begin{aligned}
E'_x(\vec{r}\,', t') &= E_x(\vec{r}, t) = q\frac{x' + vt'}{\sqrt{1 - v^2/c^2}|\vec{r}|^3} = q\frac{(x' + vt')(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}, \\
E'_y(\vec{r}\,', t') &= \frac{E_y(\vec{r}, t)}{\sqrt{1 - v^2/c^2}} = q\frac{y'}{\sqrt{1 - v^2/c^2}|\vec{r}|^3} = q\frac{y'(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}, \\
E'_z(\vec{r}\,', t') &= \frac{E_z(\vec{r}, t)}{\sqrt{1 - v^2/c^2}} = q\frac{z'}{\sqrt{1 - v^2/c^2}|\vec{r}|^3} = q\frac{z'(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}, \\
B'_x(\vec{r}\,', t') &= B_x(\vec{r}, t) = 0, \\
B'_y(\vec{r}\,', t') &= \frac{E_z(\vec{r}, t)v/c}{\sqrt{1 - v^2/c^2}} = q\frac{z'v/c}{\sqrt{1 - v^2/c^2}|\vec{r}|^3} = \frac{qv}{c}\frac{z'(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}, \\
B'_z(\vec{r}\,', t') &= -\frac{E_y(\vec{r}, t)v/c}{\sqrt{1 - v^2/c^2}} = -q\frac{y'v/c}{\sqrt{1 - v^2/c^2}|\vec{r}|^3} = -\frac{qv}{c}\frac{y'(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}.
\end{aligned} \tag{2.10.7}$$

Hence, using $\vec{v}\,' = (-v, 0, 0)$, we can write

$$\vec{E}'(\vec{r}\,', t') = q\frac{\vec{R}'(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}, \ \vec{B}'(\vec{r}\,', t') = \frac{q\vec{v}\,'}{c} \times \frac{\vec{R}'(1 - v^2/c^2)}{\sqrt{R'^2 - d^2v^2/c^2}^3}. \tag{2.10.8}$$

Since a moving charge generates a current, it should not be surprising that the moving observer also sees a magnetic field. To the leading order of the non-relativistic expansion, i.e. for sufficiently small velocity v such that v^2/c^2 is negligible, one finds

$$\vec{E}'(\vec{r}\,', t') = q\frac{\vec{R}\,'}{R'^3}, \ \vec{B}'(\vec{r}\,', t') = \frac{q\vec{v}\,'}{c} \times \frac{\vec{R}\,'}{R'^3}. \tag{2.10.9}$$

2.11 The Relativistic Doppler Effect

Let us now consider the special case of a plane wave propagating in the x-direction described by the wave 4-vector

$$k^\mu = (\frac{\omega}{c}, k, 0, 0) = (k, k, 0, 0). \tag{2.11.1}$$

Let us assume that the corresponding electric field is pointing in the y-direction, i.e. $\vec{E} = (0, E, 0)$. The magnetic field is then pointing in the z-direction and is given by $\vec{B} = (0, 0, E)$. After the Lorentz transformation we have

$$\begin{aligned} k'^\mu &= \Lambda^\mu{}_\nu k^\nu = ((C - S)k, (C - S)k, 0, 0) \\ &= \frac{1 - v/c}{\sqrt{1 - v^2/c^2}}(k, k, 0, 0) = \sqrt{\frac{c - v}{c + v}}(k, k, 0, 0). \end{aligned} \tag{2.11.2}$$

Similarly, the electric field is now given by

$$\begin{aligned} &E'_x = 0,\ E'_y = \frac{E - Ev/c}{\sqrt{1 - v^2/c^2}} = E\sqrt{\frac{c - v}{c + v}},\ E'_z = 0, \\ &B'_x = 0,\ B'_y = 0,\ B'_z = \frac{E - Ev/c}{\sqrt{1 - v^2/c^2}} = E\sqrt{\frac{c - v}{c + v}}. \end{aligned} \tag{2.11.3}$$

Up to an overall rescaling by a factor of $\sqrt{(c - v)/(c + v)}$ the vectors $\vec{k}$, $\vec{E}$, and $\vec{B}$ remain unchanged. The scaling factor is related to the relativistic Doppler effect. In particular, the observer following the electromagnetic wave with velocity v sees the frequency

$$\omega' = |\vec{k}\,'|c = \sqrt{\frac{c - v}{c + v}}|\vec{k}|c = \sqrt{\frac{c - v}{c + v}}\omega, \tag{2.11.4}$$

which is thus smaller than the frequency ω observed in the rest frame.

In contrast to non-relativistic physics, in relativistic physics there is also a transverse Doppler effect. In order to understand this, let us now consider a plane wave propagating in the z-direction described by the wave 4-vector

$$k^\mu = (\frac{\omega}{c}, 0, 0, k) = (k, 0, 0, k). \tag{2.11.5}$$

After the Lorentz transformation we now have

$$k'^\mu = \Lambda^\mu{}_\nu k^\nu = (Ck, -Sk, 0, k), \tag{2.11.6}$$

such that

$$\omega' = C\omega = \frac{\omega}{\sqrt{1 - v^2/c^2}}. \tag{2.11.7}$$

Hence, if one moves perpendicular to the direction of propagation of an electromagnetic wave, one sees a larger frequency than an observer at rest. It should be noted that angles are also frame-dependent. When we say that the observer moves perpendicular to the propagation direction of the electromagnetic wave, we make this statement in the frame of the observer at rest.

2.12 Relativistic Particle in an Electromagnetic Field

In non-relativistic physics Netwon's equation and the Lorentz force determine the motion of a charged particle in an electromagnetic field through Newton's equation of motion

$$m\vec{a}(t) = q\left[\vec{E}(\vec{r}(t),t) + \frac{\vec{v}(t)}{c} \times \vec{B}(\vec{r}(t),t)\right]. \quad (2.12.1)$$

What is the correct relativistic version of the equation of motion? In order to answer this question, let us first consider a free relativistic particle of mass m. The corresponding action is given by

$$S[\vec{r}(t)] = -\int dt\ mc^2\sqrt{1-\vec{v}(t)^2/c^2}, \quad (2.12.2)$$

which is just the rest energy mc^2 times the length of the particle's world line. The Lagrange function hence takes the form

$$L = -mc^2\sqrt{1-\vec{v}(t)^2/c^2}, \quad (2.12.3)$$

and the momentum canonically conjugate to the coordinate $\vec{r}(t)$ is given by

$$p_i = \frac{\partial L}{\partial v_i(t)} = \frac{mv_i(t)}{\sqrt{1-\vec{v}(t)^2/c^2}}. \quad (2.12.4)$$

The corresponding Hamilton function is then given by

$$\begin{aligned} H &= \vec{p}\cdot\vec{v} - L = \frac{m\vec{v}(t)^2}{\sqrt{1-\vec{v}(t)^2/c^2}} + mc^2\sqrt{1-\vec{v}(t)^2/c^2} \\ &= \frac{mc^2}{\sqrt{1-\vec{v}(t)^2/c^2}} = \sqrt{(mc^2)^2+(\vec{p}c)^2} \end{aligned} \quad (2.12.5)$$

in agreement with the energy-momentum relation for a free relativistic particle.

The action of a charge and current distribution in an external electromagnetic field is given by

$$\begin{aligned} S[j(x),A(x)] &= -\frac{1}{c}\int d^4x\ A_\mu(x)j^\mu(x) \\ &= -\int dtd^3x\ [\rho(\vec{x},t)\Phi(\vec{x},t) - \frac{1}{c}\vec{j}(\vec{x},t)\cdot\vec{A}(\vec{x},t)]. \end{aligned} \quad (2.12.6)$$

The charge and current densities of a single charged particle are given by

$$\rho(\vec{x},t) = q\delta(\vec{x}-\vec{r}(t)),\ \vec{j}(\vec{x},t) = q\vec{v}(t)\delta(\vec{x}-\vec{r}(t)). \quad (2.12.7)$$

It is easy to show that charge is conserved, i.e.

$$\partial_t \rho(\vec{x},t) + \vec{\nabla} \cdot \vec{j}(\vec{x},t) = 0. \tag{2.12.8}$$

Inserting eq.(2.12.7) into eq.(2.12.6), and integrating over space, for the action one obtains

$$S[\vec{r}(t), A(\vec{r}(t))] = -\int dt \, \left[q\Phi(\vec{r}(t),t) - q\frac{\vec{v}(t)}{c} \cdot \vec{A}(\vec{r}(t),t) \right]. \tag{2.12.9}$$

This action is invariant under gauge transformations because

$$\begin{aligned}
&\int dt \, \left[{}^{\varphi}\Phi(\vec{r}(t),t) - \frac{\vec{v}(t)}{c} \cdot {}^{\varphi} \vec{A}(\vec{r}(t),t) \right] = \\
&\int dt \, \left[\Phi(\vec{r}(t),t) + \frac{1}{c}\partial_t \varphi(\vec{r}(t),t) - \frac{\vec{v}(t)}{c} \cdot (\vec{A}(\vec{r}(t),t) - \vec{\nabla}\varphi(\vec{r}(t),t)) \right] = \\
&\int dt \, \left[\Phi(\vec{r}(t),t) - \frac{\vec{v}(t)}{c} \cdot \vec{A}(\vec{r}(t),t) + \frac{1}{c}\frac{d}{dt}\varphi(\vec{r}(t),t) \right],
\end{aligned} \tag{2.12.10}$$

and because the total derivative

$$\frac{d}{dt}\varphi(\vec{r}(t),t) = \partial_t \varphi(\vec{r}(t),t) + \vec{v} \cdot \vec{\nabla}\varphi(\vec{r}(t),t), \tag{2.12.11}$$

integrates to zero as long as $\varphi(\vec{r}(t),t)$ vanishes in the infinite past and future. Identifying the total Lagrange function as

$$L = -mc^2\sqrt{1 - \vec{v}(t)^2/c^2} - q\Phi(\vec{r}(t),t) + q\frac{\vec{v}(t)}{c} \cdot \vec{A}(\vec{r}(t),t), \tag{2.12.12}$$

it is straightforward to derive the equation as the Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial v_i(t)} - \frac{\partial L}{\partial r_i} = 0, \tag{2.12.13}$$

and one obtains

$$\frac{d}{dt}\left(\frac{m\vec{v}(t)}{\sqrt{1 - \vec{v}(t)^2/c^2}} \right) = q\left[\vec{E}(\vec{r}(t),t) + \frac{\vec{v}(t)}{c} \times \vec{B}(\vec{r}(t),t) \right]. \tag{2.12.14}$$

The theory can also be formulated in terms of a classical Hamilton function

$$H = \vec{p}(t) \cdot \vec{v}(t) - L, \tag{2.12.15}$$

where $\vec{p}$ again is the momentum canonically conjugate to the coordinate $\vec{r}$. One now finds

$$\frac{m\vec{v}(t)}{\sqrt{1 - \vec{v}(t)^2/c^2}} = \vec{p}(t) - \frac{q}{c}\vec{A}(\vec{r}(t),t), \tag{2.12.16}$$

and thus one obtains

$$H = \sqrt{(mc^2)^2 + \left[\vec{p}(t)c - q\vec{A}(\vec{r}(t), t)\right]^2} + q\Phi(\vec{r}(t), t). \quad (2.12.17)$$

This is indeed consistent because

$$v_i(t) = \frac{dr_i(t)}{dt} = \frac{\partial H}{\partial p_i(t)} = \frac{[p_i(t)c - qA_i(\vec{r}(t), t)]\, c}{\sqrt{(mc^2)^2 + \left[\vec{p}(t)c - q\vec{A}(\vec{r}(t), t)\right]^2}}. \quad (2.12.18)$$

The other equation of motion is

$$\frac{dp_i(t)}{dt} = -\frac{\partial H}{\partial r_i(t)}. \quad (2.12.19)$$

It is straightforward to show that these equations of motion are again equivalent to eq.(2.12.14).

2.13 From Point Mechanics to Classical Field Theory

Before we formulate electrodynamics in the Lagrangian formalism, let us compare a generic classical field theory with point mechanics. Point mechanics describes the dynamics of classical non-relativistic point particles. The coordinates of the particles represent a finite number of degrees of freedom. In the simplest case — a single particle moving in one spatial dimension — we are dealing with a single degree of freedom: the x-coordinate of the particle. The dynamics of a particle of mass m moving in an external potential $V(x)$ is described by Newton's equation

$$m\partial_t^2 x = ma = F(x) = -\frac{dV(x)}{dx}. \quad (2.13.1)$$

Once the initial conditions are specified, this ordinary second order differential equation determines the particle's path $x(t)$, i.e. its position as a function of time. Newton's equation results from the variational principle to minimize the action

$$S[x] = \int dt\ L(x, \partial_t x), \quad (2.13.2)$$

over the space of all paths $x(t)$. The action is a functional (a function whose argument is itself a function) that results from the time integral of the Lagrange function

$$L(x, \partial_t x) = \frac{m}{2}(\partial_t x)^2 - V(x). \quad (2.13.3)$$

The Euler-Lagrange equation

$$\partial_t \frac{\partial L}{\partial(\partial_t x)} - \frac{\partial L}{\partial x} = 0, \tag{2.13.4}$$

is nothing but Newton's equation.

Classical field theories are a generalization of point mechanics to systems with infinitely many degrees of freedom — a given number for each space point $\vec{x}$. In this case, the degrees of freedom are the field values $\phi(\vec{x})$, where ϕ is some generic field. In case of a neutral scalar field, ϕ is simply a real number representing one degree of freedom per space point. A charged scalar field, on the other hand, is described by a complex number and hence represents two degrees of freedom per space point. The scalar Higgs field $\phi^a(\vec{x})$ (with $a \in \{1, 2\}$) in the standard model of particle physics, for example, is a complex doublet, i.e. it has four real degrees of freedom per space point. An Abelian gauge field $A_i(\vec{x})$ (with a spatial direction index $i \in \{1, 2, 3\}$) — for example, the vector potential in electrodynamics — is a neutral vector field with 3 real degrees of freedom per space point. One of these degrees of freedom is redundant due to the $U(1)_{em}$ gauge symmetry. Hence, an Abelian gauge field has two physical degrees of freedom per space point which correspond to the two polarization states of the massless photon. The time-component $A_0(\vec{x})$ does not represent a physical degree of freedom. It is just a Lagrange multiplier field that enforces the Gauss law. A non-Abelian gauge field $A_i^a(\vec{x})$ is charged and has an additional index a. For example, the gluon field in chromodynamics with a color index $a \in \{1, 2, ..., 8\}$ represents $2 \times 8 = 16$ physical degrees of freedom per space point, again because of some redundancy due to the $SU(3)_c$ color gauge symmetry. The field that represents the W- and Z-bosons in the standard model has an index $a \in \{1, 2, 3\}$ and transforms under the gauge group $SU(2)_L$. Thus, it represents $2 \times 3 = 6$ physical degrees of freedom. However, in contrast to the photon, the W- and Z-bosons are massive due to the so-called Higgs mechanism and have three (not just two) polarization states. The extra degree of freedom is provided by the Higgs field.

The analog of Newton's equation in field theory is the classical field equation of motion. For example, for a neutral scalar field this is the Klein-Gordon equation

$$\partial_\mu \partial^\mu \phi = -\frac{dV(\phi)}{d\phi}. \tag{2.13.5}$$

Again, after specifying appropriate initial conditions it determines the classical field configuration $\phi(x)$, i.e. the values of the field ϕ at all space-time points $x = (ct, \vec{x})$. Hence, the role of time in point mechanics is played by space-time in field theory, and the role of the point particle coordinates is now played by the field values. As before, the classical equation of motion results from minimizing the action

$$S[\phi] = \int d^4x \ \mathcal{L}(\phi, \partial_\mu \phi). \tag{2.13.6}$$

The integral over time in eq.(2.13.2) is now replaced by an integral over space-time and the Lagrange function of point mechanics gets replaced by the Lagrange density function (or Lagrangian)

$$\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - V(\phi). \tag{2.13.7}$$

A simple massive free field theory has the potential

$$V(\phi) = \frac{M^2}{2}\phi^2. \tag{2.13.8}$$

Here M is the mass of the scalar field. Note that the mass term corresponds to a harmonic oscillator potential in the point mechanics analog. As before, the Euler-Lagrange equation

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} - \frac{\partial \mathcal{L}}{\partial \phi} = 0, \tag{2.13.9}$$

is the classical equation of motion, in this case the Klein-Gordon equation

$$(\partial_\mu \partial^\mu + M^2)\phi = 0. \tag{2.13.10}$$

In point mechanics the momentum canonically conjugate to the coordinate x is given by

$$p_x = \frac{\partial L}{\partial(\partial_t x)}, \tag{2.13.11}$$

and the corresponding Hamilton function is given by

$$H = p_x \partial_t x - L. \tag{2.13.12}$$

Similarly, in classical field theory the momentum canonically conjugate to the field ϕ is

$$p_\phi = \frac{\partial \mathcal{L}}{\partial(\partial_t \phi)} = \frac{1}{c^2}\partial_t \phi. \tag{2.13.13}$$

The energy density is described by the Hamilton density or Hamiltonian

$$\begin{aligned} \mathcal{H} &= p_\phi \partial_t \phi - \mathcal{L} = \frac{1}{2}\left(\frac{1}{c^2}\partial_t \phi \partial_t \phi + \vec{\nabla}\phi \cdot \vec{\nabla}\phi\right) + V(\phi) \\ &= \frac{1}{2}\left(c^2 p_\phi p_\phi + \vec{\nabla}\phi \cdot \vec{\nabla}\phi\right) + V(\phi). \end{aligned} \tag{2.13.14}$$

It is interesting to consider the so-called energy-momentum tensor (also known as the stress-energy tensor)

$$T^{\mu\nu} = \partial^\mu \phi \partial^\nu \phi - \mathcal{L} g^{\mu\nu}, \tag{2.13.15}$$

which represents the currents corresponding to a set of four conserved charges, i.e.

$$\partial_\mu T^{\mu\nu} = 0. \tag{2.13.16}$$

The time-time-component of the energy-momentum tensor is the energy density

$$T^{00} = \partial^0 \phi \partial^0 \phi - \mathcal{L} = \mathcal{H}. \tag{2.13.17}$$

Similarly, the space-time components of the energy-momentum tensor represent the momentum density

$$T^{i0} = \partial^i \phi \partial^0 \phi. \tag{2.13.18}$$

The continuity equation that describes energy conservation is given by

$$\partial_0 T^{00} + \partial_i T^{i0} = 0. \tag{2.13.19}$$

The analogies between point mechanics and field theory are summarized in table 2.1.

2.14 Action and Euler-Lagrange Equation

The Lagrangian for the electromagnetic field interacting with a charge and current distribution j^μ is given by

$$\mathcal{L} = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} - \frac{1}{c} A_\mu j^\mu. \tag{2.14.1}$$

The corresponding action is obtained by integrating the Lagrangian over space-time, i.e.

$$S[A] = \int dt\ d^3x\ \mathcal{L} = \int d^4x\ \left(-\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} - \frac{1}{c} A_\mu j^\mu \right). \tag{2.14.2}$$

The action can be viewed as a functional (i.e. a function of a function) of the electromagnetic 4-vector potential A^μ. The Euler-Lagrange equation of motion resulting from the principle of least action now takes the form

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial \partial_\mu A_\nu} - \frac{\partial \mathcal{L}}{\partial A_\nu} = \frac{1}{4\pi} \partial_\mu F^{\mu\nu} - \frac{1}{c} j^\nu = 0. \tag{2.14.3}$$

Point Mechanics	Field Theory
time t	space-time $x = (ct, \vec{x})$
particle coordinate x	field value ϕ
particle path $x(t)$	field configuration $\phi(x)$
action $S[x] = \int dt\ L(x, \partial_t x)$	action $S[\phi] = \int d^4x\ \mathcal{L}(\phi, \partial_\mu \phi)$
Lagrange function $L(x, \partial_t x) = \frac{m}{2}(\partial_t x)^2 - V(x)$	Lagrangian $\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - V(\phi)$
equation of motion $\partial_t \frac{\partial L}{\partial(\partial_t x)} - \frac{\partial L}{\partial x} = 0$	field equation $\partial_\mu \frac{\partial L}{\partial(\partial_\mu \phi)} - \frac{\partial L}{\partial \phi} = 0$
Newton's equation $\partial_t^2 x = -\frac{dV(x)}{dx}$	Klein-Gordon equation $\partial_\mu \partial^\mu \phi = -\frac{dV(\phi)}{d\phi}$
kinetic energy $\frac{m}{2}(\partial_t x)^2$	kinetic energy $\frac{1}{2}\partial_\mu \phi \partial^\mu \phi$
harmonic oscillator potential $\frac{m}{2}\omega^2 x^2$	mass term $\frac{M^2}{2}\phi^2$
conjugate momentum $p_x = \frac{\partial L}{\partial(\partial_t x)}$	conjugate momentum $p_\phi = \frac{\partial \mathcal{L}}{\partial(\partial_t \phi)}$
Hamilton function $H = p_x \partial_t x - L$	Hamiltonian $\mathcal{H} = p_\phi \partial_t \phi - \mathcal{L}$

Table 2.1: *The dictionary that translates point mechanics into the language of field theory.*

Indeed, this yields just the inhomogeneous Maxwell equations

$$\partial_\mu F^{\mu\nu} = \frac{4\pi}{c} j^\nu. \tag{2.14.4}$$

The homogeneous Maxwell equations are automatically satisfied as a consequence of $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$.

2.15 Energy-Momentum Tensor

Let us consider the energy-momentum tensor of the free electromagnetic field (i.e. in the absence of charges and currents)

$$T^{\mu\nu} = -\frac{1}{4\pi} F^{\mu\rho} F^{\nu}{}_{\rho} - \mathcal{L} g^{\mu\nu}, \tag{2.15.1}$$

which obeys the continuity equation

$$\partial_\mu T^{\mu\nu} = 0. \tag{2.15.2}$$

The time-time component of the energy-momentum tensor is given by

$$\begin{aligned} T^{00} &= -\frac{1}{4\pi}F^{0\rho}F^{0}{}_{\rho} + \frac{1}{16\pi}F_{\rho\sigma}F^{\rho\sigma}g^{00} = \frac{1}{4\pi}\vec{E}^2 + \frac{1}{8\pi}(\vec{B}^2 - \vec{E}^2) \\ &= \frac{1}{8\pi}(\vec{E}^2 + \vec{B}^2) = u. \end{aligned} \tag{2.15.3}$$

Here u is the well-known energy density of the electromagnetic field. The space-time components of the energy-momentum tensor take the form

$$T^{i0} = -\frac{1}{4\pi}F^{i\rho}F^{0}{}_{\rho} + \frac{1}{16\pi}F_{\rho\sigma}F^{\rho\sigma}g^{i0} = \frac{1}{4\pi}\varepsilon_{ijk}E_jB_k = \frac{1}{4\pi}(\vec{E}\times\vec{B})_i = \frac{1}{c}S_i. \tag{2.15.4}$$

Here ε_{ijk} is the completely anti-symmetric Levi-Civita tensor and

$$\vec{S} = \frac{c}{4\pi}\vec{E}\times\vec{B} \tag{2.15.5}$$

is the Poynting vector which is known to represent the momentum density of the electromagnetic field. The continuity equation that represents energy conservation takes the form

$$\partial_\mu T^{\mu 0} = \partial_0 T^{00} + \partial_i T^{i0} = \frac{1}{c}(\partial_t u + \vec{\nabla}\cdot\vec{S}) = 0. \tag{2.15.6}$$

Chapter 3

Basics of General Relativity

General relativity is a theory of gravity, a generalization of Newton's theory. For weak gravitational fields and for processes slow compared to the velocity of light, it reduces to Newton's theory of gravity. General relativity is a classical theory, which has strongly resisted quantization, just like Einstein himself had problems with quantum mechanics. General relativity is physics in curved space-time. The gravitational "field" is represented by the metric of space-time, and by related quantities like the connection, Riemann's curvature tensor, the Ricci tensor and the curvature scalar. Matter fields are the sources that curve space-time, and in turn the dynamics of matter is influenced by the curvature itself. In order to understand the cosmological evolution we don't need too much of general relativity, but we must understand its basic principles, and we must get used to Riemann's geometry. First, we will talk just about space (not yet about space-time), and indeed curved spaces play a role in the physics of the early Universe, although our space is rather flat.

3.1 Spaces of Constant Curvature

In 1854 in Göttingen (Germany) Riemann formulated the theory that is used to describe the geometry of curved spaces. Half a century later, Einstein used Riemann's mathematical framework to describe curved space-time.

Let us begin with something very simple, namely with two-dimensional spaces, and let us first discuss the simplest 2-d space, the plane $\mathbb{R}^2$. We use Cartesian coordinates and characterize a point by $x = (x^1, x^2)$. The distance between two infinitesimally close points is computed following Pythagoras

$$(ds)^2 = (dx^1)^2 + (dx^2)^2 = g_{ij}(x)dx^i dx^j. \tag{3.1.1}$$

We have just introduced a metric

$$g(x) = \begin{pmatrix} g_{11}(x) & g_{12}(x) \\ g_{21}(x) & g_{22}(x) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.1.2}$$

The metric is independent of x, which is a special case, because $\mathbb{R}^2$ is flat. Also the metric is symmetric, i.e.

$$g_{ji}(x) = g_{ij}(x), \tag{3.1.3}$$

which is generally the case.

Coordinates, however, are not physical. We can choose other ones if we want to. Sometimes it is useful to work with curve-linear coordinates, for example with polar coordinates. We then have

$$x^1 = r\cos\varphi,\ x^2 = r\sin\varphi, \tag{3.1.4}$$

such that

$$dx^1 = \cos\varphi\ dr - r\sin\varphi\ d\varphi,\ dx^2 = \sin\varphi\ dr + r\cos\varphi\ d\varphi, \tag{3.1.5}$$

and hence

$$(ds)^2 = (dr)^2 + r^2(d\varphi)^2. \tag{3.1.6}$$

Now we have another metric

$$g(x) = \begin{pmatrix} g_{rr}(r,\varphi) & g_{r\varphi}(r,\varphi) \\ g_{\varphi r}(r,\varphi) & g_{\varphi\varphi}(r,\varphi) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}, \tag{3.1.7}$$

which does depend on x (namely on r). The metric depends on the choice of coordinates. Of course, the space is still the same. In particular, even a flat space may have an x-dependent metric. Hence, curve-linear coordinates certainly do not imply a curved space.

Let us consider something a bit more interesting, a curved space — the sphere S^2. We can easily visualize the sphere S^2 by embedding it in the flat space $\mathbb{R}^3$. However, two-dimensional inhabitants of the surface of the sphere could calculate just as we do — embeddings are not necessary. For 3-d people like ourselves it is particularly easy, because we just write

$$x^1 = R\sin\theta\cos\varphi,\ x^2 = R\sin\theta\sin\varphi,\ x^3 = R\cos\theta, \tag{3.1.8}$$

and therefore

$$\begin{aligned}(ds)^2 &= (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \\ &= R^2(\cos\theta\cos\varphi\ d\theta - \sin\theta\sin\varphi\ d\varphi)^2 \\ &+ R^2(\cos\theta\sin\varphi\ d\theta - \sin\theta\cos\varphi\ d\varphi)^2 + R^2\sin^2\theta\ d\theta^2 \\ &= R^2(d\theta^2 + \sin^2\theta d\varphi^2).\end{aligned} \tag{3.1.9}$$

The metric is thus given by

$$g(x) = R^2 \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}. \tag{3.1.10}$$

Of course, we can again introduce other coordinates. The following choice may seem unmotivated, but it will be useful later. We introduce

$$\rho = \sin\theta,\ d\rho = \cos\theta\ d\theta, \tag{3.1.11}$$

and obtain

$$(ds)^2 = R^2(\frac{d\rho^2}{1-\rho^2} + \rho^2 d\varphi^2). \tag{3.1.12}$$

Thus the new metric takes the form

$$g(x) = R^2 \begin{pmatrix} \frac{1}{1-\rho^2} & 0 \\ 0 & \rho^2 \end{pmatrix}. \tag{3.1.13}$$

The sphere has a constant positive curvature. Let us now consider a space with constant negative curvature — the hyperbolically curved surface H^2. We can no longer embed this surface in $\mathbb{R}^3$, and we may thus feel a bit uncomfortable. For the 2-d people on the surface of S^2 it does not make a difference, and from now on also we must rely just on the mathematics. A possible metric of H^2 is

$$(ds)^2 = R^2(d\theta^2 + \sinh^2\theta\ d\varphi^2), \tag{3.1.14}$$

where now $\theta \in [0, \infty]$. We can embed H^2 in a 3-d Minkowski-space by writing

$$(ds)^2 = (dx^1)^2 + (dx^2)^2 - (dx^3)^2, \tag{3.1.15}$$

if we identify

$$x^1 = R\sinh\theta\cos\varphi,\ x^2 = R\sinh\theta\sin\varphi,\ x^3 = R\cosh\theta, \tag{3.1.16}$$

because we then have

$$\begin{aligned}(ds)^2 &= R^2(\cosh\theta\cos\varphi\ d\theta - \sinh\theta\sin\varphi\ d\varphi)^2 \\ &+ R^2(\cosh\theta\sin\varphi\ d\theta - \sinh\theta\cos\varphi\ d\varphi)^2 - R^2\sinh^2\theta\ d\theta^2 \\ &= R^2(d\theta^2 + \sinh^2\theta\ d\varphi^2).\end{aligned} \tag{3.1.17}$$

We can still try to visualize H^2, when we realize that

$$(x^1)^2 + (x^2)^2 - (x^3)^2 = R^2(\sinh^2\theta - \cosh^2\theta) = -R^2. \tag{3.1.18}$$

Interpreting

$$(x^3)^2 = (x^1)^2 + (x^2)^2 + R^2 \tag{3.1.19}$$

correctly is, however, delicate, because we still are in Minkowski-space, and we may hence still feel a bit uncomfortable. Let us again introduce new coordinates

$$\rho = \sinh\theta,\ d\rho = \cosh\theta\ d\theta, \tag{3.1.20}$$

such that the metric then is

$$(ds)^2 = R^2(\frac{d\rho^2}{1+\rho^2} + \rho^2 d\varphi^2). \tag{3.1.21}$$

We can now write the metrics for $\mathbb{R}^2$, S^2 and H^2 in a common form

$$(ds)^2 = R^2(\frac{d\rho^2}{1-k\rho^2} + \rho^2 d\varphi^2). \tag{3.1.22}$$

Here $k = 0, 1, -1$ for $\mathbb{R}^2$, S^2 and H^2, respectively. For $\mathbb{R}^2$ the introduction of $r = R\rho$ is a bit artificial. We have introduced a scale factor R, while ρ is dimensionless.

We now have a list of all two-dimensional spaces of constant curvature. This list is complete, at least as long as we restrict ourselves to local properties. However, we can still have various global topologies. For example, we can roll up the plane to a cylinder by identifying appropriate points. Our 2-d people don't notice the difference (the cylinder is as flat as the plane), unless they travel around their entire Universe (global topology). We can also roll up the cylinder to a torus T^2, even though we cannot visualize this in $\mathbb{R}^3$ without extra curvature. The torus is finite, just like the sphere, but as flat as the plane. Similarly, there are spaces that look like H^2 locally, but are finite. Spaces that are locally like S^2 are always finite. Still, we can also change the global topology of S^2, for example by identifying antipodal points.

All we said about two-dimensional spaces of constant curvature generalizes trivially to three-dimensional spaces of constant curvature. The general form of the metric then is

$$(ds)^2 = R^2(\frac{d\rho^2}{1-k\rho^2} + \rho^2(d\theta^2 + \sin^2\theta\ d\varphi^2)). \tag{3.1.23}$$

When we investigate the Universe on the largest scales, we find that it is rather homogeneous and isotropic (galaxy distribution, cosmic background radiation). Since the observable Universe is homogeneous and isotropic, and since we assume that we don't inhabit a special place in it, we conclude that the whole Universe is a three-dimensional space of constant curvature. Note that we do not talk about space-time yet — just about 3-d space. We have investigated various metrics, and we found that the metric depends on the coordinate system. Coordinates, however, are not physical. When we want to answer physical questions (like "what is the shortest path from A to B ?", or "how strongly curved is space ?") we need more of Riemann's beautiful mathematics.

3.2 Geodesics and Metric Connection

What is the shortest path from A to B ? We are looking for a geodesic (and that is exactly what particles in space-time do as well). Let us consider an arbitrary (d-dimensional) space with a metric, and a curve $x(\lambda)$ connecting two points A and B, i.e. $x(0) = A$, $x(1) = B$. The length S of the curve is then given by

$$S = \int_0^1 d\lambda \ (g_{ij}(x)\dot{x}^i\dot{x}^j)^{1/2}, \ \dot{x} = \frac{dx}{d\lambda}. \tag{3.2.1}$$

The integral is reparameterization invariant, i.e. defining

$$x' = \frac{dx}{d\lambda'} = \frac{dx}{d\lambda}\frac{d\lambda}{d\lambda'} = \dot{x}\frac{d\lambda}{d\lambda'}, \tag{3.2.2}$$

implies

$$S' = \int_0^1 d\lambda' \ (g_{ij}(x)x'^i x'^j)^{1/2} = \int_0^1 d\lambda \ \frac{d\lambda'}{d\lambda}(g_{ij}(x)\dot{x}^i\frac{d\lambda}{d\lambda'}\dot{x}^j\frac{d\lambda}{d\lambda'})^{1/2} = S. \tag{3.2.3}$$

We are looking for a curve of shortest length — i.e. we are trying to solve a variational problem — and we can use what we know from classical mechanics. The classical action

$$S = \int_{t_0}^{t_1} dt \ L(x, \dot{x}) \tag{3.2.4}$$

is minimal for the classical path that solves the Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0. \tag{3.2.5}$$

Applying this to our geodesics problem implies

$$\frac{d}{d\lambda}\frac{1}{2}(g_{ij}(x)\dot{x}^i\dot{x}^j)^{-1/2} \ 2g_{kl}\dot{x}^l - \frac{1}{2}(g_{ij}(x)\dot{x}^i\dot{x}^j)^{-1/2} \ \partial_k g_{lm}(x)\dot{x}^l\dot{x}^m = 0. \tag{3.2.6}$$

We now use the reparameterization invariance, and parameterize the curve by its arc-length s

$$ds = (g_{ij}(x)\dot{x}^i\dot{x}^j)^{1/2}d\lambda, \tag{3.2.7}$$

such that

$$\begin{aligned}
&\frac{d}{ds}(g_{kl}\frac{dx^l}{ds}) - \frac{1}{2}\partial_k g_{lm}\frac{dx^l}{ds}\frac{dx^m}{ds} = 0 \ \Rightarrow \\
&g_{kl}\frac{d^2x^l}{ds^2} + \partial_i g_{kl}\frac{dx^i}{ds}\frac{dx^l}{ds} - \frac{1}{2}\partial_k g_{lm}\frac{dx^l}{ds}\frac{dx^m}{ds} = 0.
\end{aligned} \tag{3.2.8}$$

We introduce the inverse metric

$$g^{ij}g_{jk} = \delta^i_k, \tag{3.2.9}$$

and we use the symmetry of the metric to write

$$\frac{d^2x^l}{ds^2} + \frac{1}{2}g^{lk}(\partial_i g_{kj} + \partial_j g_{ki} - \partial_k g_{ij})\frac{dx^i}{ds}\frac{dx^j}{ds} = 0. \tag{3.2.10}$$

The new object deserves its own name. It is called the metric connection, or just the connection

$$\Gamma^l_{ij} = \frac{1}{2}g^{lk}(\partial_i g_{kj} + \partial_j g_{ki} - \partial_k g_{ij}). \tag{3.2.11}$$

The Γ^l_{ij} are also known as Christoffel symbols. Using the connection, the equation for the geodesic takes the form

$$\frac{d^2x^l}{ds^2} + \Gamma^l_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 0. \tag{3.2.12}$$

Let us consider a trivial example — geodesics in $\mathbb{R}^2$. We know that these are straight lines

$$x^l = a^l s + b^l \ \Rightarrow \ \frac{d^2x^l}{ds^2} = 0 \ \Rightarrow \ \Gamma^l_{ij} = 0. \tag{3.2.13}$$

All coefficients of the connection vanish. Now we go to polar coordinates

$$\begin{aligned}
&x^1 = r\cos\varphi,\ x^2 = r\sin\varphi \ \Rightarrow \\
&\frac{d^2r}{ds^2}\cos\varphi - 2\frac{dr}{ds}\sin\varphi\frac{d\varphi}{ds} - r\cos\varphi(\frac{d\varphi}{ds})^2 - r\sin\varphi\frac{d^2\varphi}{ds^2} = 0, \\
&\frac{d^2r}{ds^2}\sin\varphi + 2\frac{dr}{ds}\cos\varphi\frac{d\varphi}{ds} - r\sin\varphi(\frac{d\varphi}{ds})^2 + r\cos\varphi\frac{d^2\varphi}{ds^2} = 0 \ \Rightarrow \\
&\frac{d^2r}{ds^2} - r(\frac{d\varphi}{ds})^2 = 0,\ 2\frac{dr}{ds}\frac{d\varphi}{ds} + r\frac{d^2\varphi}{ds^2} = 0 \ \Rightarrow \\
&\Gamma^r_{\varphi\varphi} = -r,\ \Gamma^\varphi_{r\varphi} = \frac{1}{r}.
\end{aligned} \tag{3.2.14}$$

All other connection coefficients vanish. As expected, the connection depends on the choice of coordinates. Hence, it is useful to look for a "good" coordinate system, before one begins to compute the connection coefficients. In our example, the calculation in Cartesian coordinates was much simpler.

3.3 Parallel Transport and Riemann Tensor

The connection is also used to define parallel transport of a vector along some curve. The equation

$$\dot{v}^l(\lambda) + \Gamma^l_{ij} v^i(\lambda)\dot{x}^j = 0 \tag{3.3.1}$$

yields the vector $v^l(\lambda)$ parallel transported along the curve $x(\lambda)$ for a given initial $v^l(0)$. Next we introduce a covariant derivative matrix

$$D^l_{ij} = \delta^l_i \partial_j + \Gamma^l_{ij}, \tag{3.3.2}$$

and we write

$$\begin{aligned} \dot{x}^j D_j v(x(\lambda)) &= \dot{x}^j(\delta^l_i \partial_j + \Gamma^l_{ij}) v^i(x(\lambda)) \\ &= \dot{x}^j \partial_j v^l(x(\lambda)) + \Gamma^l_{ij} v^i(x(\lambda))\dot{x}^j \\ &= \dot{v}^l(\lambda) + \Gamma^l_{ij} v^i(\lambda)\dot{x}^j = 0. \end{aligned} \tag{3.3.3}$$

In particular, we can view the equation for a geodesic as a parallel transport equation for the tangent unit vector $t^i(s) = dx^i/ds$

$$\frac{dt^l(s)}{ds} + \Gamma^l_{ij} t^i(s)\frac{dx^j}{ds} = 0. \tag{3.3.4}$$

Parallel transport in curved space is qualitatively different from the one in flat space. Let us consider any closed curve in $\mathbb{R}^2$. When we choose Cartesian coordinates, all connection coefficients vanish, and a vector parallel transported along the curve returns to its initial orientation. On the other hand, when we do the same on the surface of S^2, the vector does not return to its initial orientation.

Let us consider parallel transport around an infinitesimal surface element. We compare the results along two paths

$$\begin{aligned}
[D_j, D_k] &= D_j D_k - D_k D_j \\
&= D^l_{mj} D^m_{ik} - D^l_{mk} D^m_{ij} \\
&= (\delta^l_m \partial_j + \Gamma^l_{mj})(\delta^m_i \partial_k + \Gamma^m_{ik}) - (\delta^l_m \partial_k + \Gamma^l_{mk})(\delta^m_i \partial_j + \Gamma^m_{ij}) \\
&= \delta^l_i \partial_j \partial_k + \partial_j \Gamma^l_{ik} + \Gamma^l_{ik} \partial_j + \Gamma^l_{ij} \partial_k + \Gamma^l_{mj} \Gamma^m_{ik} \\
&- \delta^l_i \partial_k \partial_j - \partial_k \Gamma^l_{ij} - \Gamma^l_{ij} \partial_k - \Gamma^l_{ik} \partial_j - \Gamma^l_{mk} \Gamma^m_{ij} \\
&= \partial_j \Gamma^l_{ik} - \partial_k \Gamma^l_{ij} + \Gamma^l_{mj} \Gamma^m_{ik} - \Gamma^l_{mk} \Gamma^m_{ij} = R^l_{ijk}. \qquad (3.3.5)
\end{aligned}$$

We have just introduced Riemann's curvature tensor

$$R^l_{ijk} = \partial_j \Gamma^l_{ik} - \partial_k \Gamma^l_{ij} + \Gamma^l_{mj} \Gamma^m_{ik} - \Gamma^l_{mk} \Gamma^m_{ij}. \qquad (3.3.6)$$

In flat space the curvature tensor vanishes. This follows immediately when we choose Cartesian coordinates, for which all connection coefficients are zero.

3.4 Ricci and Einstein Tensors and Curvature Scalar

Via index contraction we can obtain other curvature tensors from the Riemann tensor. The Ricci tensor, for example, is given by

$$R_{ik} = R^j_{ijk}, \qquad (3.4.1)$$

and the curvature scalar is defined as

$$\mathcal{R} = R^i_i = g^{ki} R_{ik}. \qquad (3.4.2)$$

Finally, the Einstein tensor is given by

$$G_{ik} = R_{ik} - \frac{1}{2} g_{ik} \mathcal{R}. \qquad (3.4.3)$$

We now have discussed enough Riemann geometry in order to attack general relativity. What was geometry of space so far, will then turn into the geometrodynamics of space-time.

3.5 Curved Space-Time

We have learned something about curved spaces, and we have discussed the basics of Riemann's geometry. Let us now apply these tools to curved space-time. From special relativity we know Minkowski-space — flat space-time. When we choose Cartesian coordinates its metric takes the form

$$(ds)^2 = (dt)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \tag{3.5.1}$$

From now on we will work in natural units in which time is measured in units of length by setting the velocity of light to $c = 1$. Denoting $t = x_0$ we can write

$$(ds)^2 = g_{\mu\nu} dx^\mu dx^\nu, \ g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3.5.2}$$

We can discuss curved space-times when we allow more general (space-time-dependent) metrics. We can then use Riemann's geometry, define the connection

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\sigma} (\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}), \tag{3.5.3}$$

and write down the equation for a geodesic

$$\frac{d^2 x^\rho}{ds^2} + \Gamma^\rho_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0. \tag{3.5.4}$$

3.6 Newton-Cartan Curved Space-Time

Before we turn to Einstein's theory of gravity, let us look back at Newton's good old theory. There space and time are well separated, space is flat and time is absolute. Let us take a mass distribution $\rho(y)$ and let us compute its gravitational potential

$$\Phi(x) = G \int dy \ \frac{\rho(y)}{|x - y|}, \ \Delta\Phi(x) = 4\pi G \rho(x). \tag{3.6.1}$$

A particle of mass m moving in this potential follows Newton's equation of motion

$$m \frac{d^2 x^l}{dt^2} = F^l = -m \frac{\partial \Phi}{\partial x^l} \ \Rightarrow \ \frac{d^2 x^l}{dt^2} + \frac{\partial \Phi}{\partial x^l} = 0. \tag{3.6.2}$$

The particle's path is a curved trajectory in flat space, with the particle moving along it as absolute time passes. Let us reinterpret this in terms of a curved Newton space-time (as it was first done by Cartan in 1923). Then absolute time would be another coordinate. We also introduce the proper time s of the particle which it reads off from a comoving watch. Since Newton's time is absolute, we simply have $s = t$. We can therefore write Newton's equations of motion (with $t = x^0$) in the form

$$\frac{d^2x^l}{ds^2} + \frac{\partial\Phi}{dx^l}\frac{dx^0}{ds}\frac{dx^0}{ds} = 0, \tag{3.6.3}$$

which we can identify as the equation for a geodesic in a curved space-time. The connection coefficients then are

$$\Gamma^l_{00} = \frac{\partial\Phi}{\partial x^l},\ l \in \{1,2,3\}, \tag{3.6.4}$$

and all other coefficients vanish. This connection cannot be derived from an underlying metric, however, it still defines a meaningful equation for geodesics in a curved Newton-Cartan space-time. We can also write down the Riemann tensor

$$R^\sigma_{\mu\nu\rho} = \partial_\nu\Gamma^\sigma_{\mu\rho} - \partial_\rho\Gamma^\sigma_{\mu\nu} + \Gamma^\sigma_{\lambda\nu}\Gamma^\lambda_{\mu\rho} - \Gamma^\sigma_{\lambda\rho}\Gamma^\lambda_{\mu\nu}. \tag{3.6.5}$$

The only non-vanishing components are

$$R^l_{0i0} = -R^l_{00i} = \partial_i\Gamma^l_{00} = \frac{\partial^2\Phi}{\partial x^i\partial x^l}. \tag{3.6.6}$$

The Ricci tensor is then given by

$$R_{\mu\rho} = R^\nu_{\mu\nu\rho} \ \Rightarrow\ R_{00} = R^i_{0i0} = \frac{\partial^2\Phi}{\partial x^{i2}} = \Delta\Phi = 4\pi G\rho, \tag{3.6.7}$$

and again all other components are zero. We cannot build the curvature scalar or the Einstein tensor in this case, because we don't have a metric. The above results can be interpreted as follows. The ordinary curved trajectories in flat space are "straight" lines (geodesics) in the curved Newton-Cartan space-time, and the mass density ρ is responsible for the curvature (Ricci tensor). From this point it is not far to go to Einstein's general relativity.

3.7 Einstein's Curved Space-Time

The existence of a metric is essential for Einstein's theory. The equation of motion for a massive particle that feels gravity only (free falling particle) is the equation for a geodesic

$$\frac{d^2x^\rho}{ds^2} + \Gamma^\rho_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0. \tag{3.7.1}$$

Here we have parameterized the particle's trajectory by its arc-length — the proper time s. Introducing the four-velocity

$$u^\mu = \frac{dx^\mu}{ds} \tag{3.7.2}$$

we can write

$$\frac{du^\rho}{ds} + \Gamma^\rho_{\mu\nu} u^\mu u^\nu = 0. \tag{3.7.3}$$

The four-velocity is a tangent unit-vector

$$u^2 = u^\mu u_\mu = g_{\mu\nu} u^\mu u^\nu = g_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = \frac{(ds)^2}{(ds)^2} = 1. \tag{3.7.4}$$

The equation of motion for massless particles (photons, light rays) is simply

$$(ds)^2 = 0, \tag{3.7.5}$$

which is also known as a null-geodesic.

3.8 Einstein's Gravitational Field Equation

We still need an equation that determines the curvature of space-time based on the mass distribution. This is Einstein's field equation

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} = 8\pi G T_{\mu\nu}. \tag{3.8.1}$$

Here $T_{\mu\nu}$ is the energy-momentum tensor of matter (particles and fields). Hence, it is not just mass, but any form of energy and momentum, that determines the curvature of space-time. In order to achieve a static Universe, Einstein had modified his equation to

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} + g_{\mu\nu}\Lambda. \tag{3.8.2}$$

After he learned of Edwin Hubble's observation of the expanding Universe, Einstein has been quoted as considering the introduction of the cosmological constant Λ his biggest blunder. Today there is indeed observational evidence for an extremely small but non-zero cosmological constant, which causes the cosmic expansion to accelerate. A non-zero value of Λ implies that "empty" space produces a gravitational effect due to vacuum energy. Since Λ represents a form of energy, it should indeed be considered part of the energy-momentum tensor. It is presently not understood why the cosmological constant is so incredibly small. This is the so-called cosmological constant problem, one of the greatest puzzles in modern physics.

3.9 The Einstein-Hilbert Action

In 1915 there was a fierce competition between Einstein and Hilbert about finding the correct equations for general relativity, i.e. the relativistic theory of gravity. Five days before Einstein had published his field equations, Hilbert had published an action from which the field equations follow by a variational principle. Although the field equations that Hilbert had published turned out to be incorrect, his action indeed yields Einstein's correct field equations. The Einstein-Hilbert action is a functional of the metric g given by

$$S[g] = \int d^4x \sqrt{-\mathrm{det}g}\ \mathcal{L}. \tag{3.9.1}$$

Here

$$dV = d^4x \sqrt{-\mathrm{det}g} \tag{3.9.2}$$

is the invariant volume element of space-time and $\mathrm{det}g < 0$ is the determinant of the metric. The Lagrangian takes the form

$$\mathcal{L} = \frac{1}{16\pi G}\mathcal{R} + \mathcal{L}_m, \tag{3.9.3}$$

where $\mathcal{R}$ is the scalar curvature and $\mathcal{L}_m$ is the Lagrangian of "matter", i.e. any non-gravitational form of energy (including fields of all kinds). Einstein's field equations result as the Euler-Lagrange equations associated with the Einstein-Hilbert action upon variation with respect to the metric. The corresponding derivation is somewhat tedious and goes beyond the scope of this course. The variation of $\sqrt{-\mathrm{det}g}\ \mathcal{R}$ with respect to $g^{\mu\nu}$ yields $\frac{1}{16\pi G}G_{\mu\nu}$ while

$$\frac{1}{\sqrt{-\mathrm{det}g}}\frac{\partial\left(\sqrt{-\mathrm{det}g}\ \mathcal{L}_m\right)}{\partial g^{\mu\nu}} = \frac{\partial\mathcal{L}_m}{\partial g^{\mu\nu}} - \frac{1}{2}g_{\mu\nu}\mathcal{L} = -\frac{1}{2}T_{\mu\nu}. \tag{3.9.4}$$

In the last step we have used

$$\frac{1}{\sqrt{-\mathrm{det}g}}\frac{\partial\sqrt{-\mathrm{det}g}}{\partial g^{\mu\nu}} = -\frac{1}{2}g_{\mu\nu}. \tag{3.9.5}$$

Chapter 4

Standard Big Bang Cosmology

We now have the tools that we need to deal with curved space-time, and we have the dynamical equations for the motion of massive particles and of light rays. Also we have Einstein's field equations, which determine the space-time curvature from the energy-momentum distribution of particles and fields. Now we want to apply general relativity to the Universe as a whole. We are interested only in the largest scales, not in local structures like galaxy-clusters, galaxies, or even single stars, or planetary systems. Of course, all these objects curve space-time locally (in case of black holes even very strongly). However, we will average over all the local structures, and we assume that the whole Universe is filled homogeneously and isotropically by a gas of matter, whose "molecules" are, for example, the galaxies. At first sight, this idealization may seem very drastic. On the other hand, we know from hydrodynamics that a continuum description of gases works very well, although they have a very discontinuous structure at molecular scales. At the end it is a question for observations, how homogeneous and isotropic our Universe really is. The observed galaxy distribution is indeed rather homogeneous and isotropic when one averages out the structures of galaxy-clusters and individual galaxies. Also the cosmic background radiation is extremely isotropic, which indicates that space was homogeneous and isotropic immediately after the Big Bang.

The standard cosmological model assumes that the Universe is a homogeneous and isotropic three-dimensional space of constant curvature, however, with a time-dependent scale parameter R. The corresponding metric is the so-called Friedmann-Lemaitre-Robertson-Walker (FLRW) metric. The Einstein field equations then determine the dynamics of the Universe (expansion), when one assumes a certain energy-momentum tensor of matter. Before we study the dynamics of space-time in detail, we will investigate the kinematics of the FLRW metric, i.e. we will play free falling observer and we will investigate the horizon.

4.1 The Friedmann-Lemaitre-Robertson-Walker Metric

Let us use some observational facts to obtain a useful ansatz for the metric of our Universe on the largest scales. Spatial homogeneity and isotropy lead us to the spaces of constant curvature. An experiment that we repeat every day tells us that we are interested in three-dimensional spaces. Thus we have the candidates $\mathbb{R}^3$, S^3 and H^3 eventually endowed with some non-trivial global topology. The spatial part of the metric is then given by

$$R^2(\frac{d\rho^2}{1-k\rho^2} + \rho^2(d\theta^2 + \sin^2\theta d\varphi^2)), \; k = 0, \pm 1, \tag{4.1.1}$$

where R is a scale parameter (of dimension length). We now use Hubble's observation that our Universe is expanding, and thus allow R to be time-dependent. Altogether, we make the FLRW ansatz

$$ds^2 = dt^2 - R(t)^2(\frac{d\rho^2}{1-k\rho^2} + \rho^2(d\theta^2 + \sin^2\theta d\varphi^2)). \tag{4.1.2}$$

We have used a special space-time coordinate system, which clearly manifests the symmetries of our ansatz. Indeed, we are in the comoving coordinate system of the "galaxy-gas", and only in this system the Universe appears homogeneous and isotropic.

We know how to compute the connection coefficients, the Riemann tensor, the Ricci tensor, the curvature scalar and the Einstein tensor. It is quite tedious to compute these quantities for the three-dimensional spaces of constant curvature. Since we have done this exercise in two-dimensions (in a homework), we simply quote the results

$$\begin{aligned}
\Gamma^l_{ij} &= \frac{1}{2} g^{lk}(\partial_j g_{ki} + \partial_i g_{kj} - \partial_k g_{ij}), \\
\Gamma^0_{ij} &= -\frac{\dot{R}}{R} g_{ij}, \ \Gamma^i_{0j} = \frac{\dot{R}}{R}\delta^i_j, \\
R_{ij} &= -(\frac{\ddot{R}}{R} + 2\frac{\dot{R}^2}{R^2} + 2\frac{k}{R^2}) g_{ij}, \ R_{00} = -3\frac{\ddot{R}}{R}, \\
\mathcal{R} &= -6(\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2}), \\
G_{ij} &= (2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) g_{ij}, \ G_{00} = 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}).
\end{aligned} \tag{4.1.3}$$

All other components vanish.

4.2 Cosmological Red-Shift and Hubble Law

Let us now follow a light ray (photon) through the Universe. We ignore all interactions with matter (interstellar gas, plasma) and consider only gravitational effects. Hence, we are looking for a null-geodesic

$$ds^2 = 0. \tag{4.2.1}$$

Since the Universe is homogeneous and isotropic, we can limit ourselves to motion along the ρ-direction, and put $\theta = \varphi = 0$. We then have

$$ds^2 = dt^2 - R(t)^2 \frac{d\rho^2}{1 - k\rho^2} = 0 \ \Rightarrow \ \frac{dt}{R(t)} = \pm \frac{d\rho}{\sqrt{1 - k\rho^2}}. \tag{4.2.2}$$

We consider a light signal, that has been emitted at time $t = 0$ (say in the moment of the big bang) at the space-point with dimensionless coordinate $\rho = \rho_H$. We observe the signal today (at time t) at the point $\rho = 0$, such that

$$\int_0^t \frac{dt'}{R(t')} = \int_0^{\rho_H} \frac{d\rho}{\sqrt{1 - k\rho^2}}. \tag{4.2.3}$$

Today, what is the distance to the point, at which the light signal was emitted originally? Even though the Universe has expanded in the mean time ($R(t)$ has increased) the dimensionless coordinate of that point still is ρ_H. The distance to that point is given by

$$\begin{aligned} d_H(t) &= \int_0^1 d\lambda\ (-g_{ij}\dot{x}^i\dot{x}^j)^{1/2} = \int_0^{\rho_H} d\rho\ \sqrt{-g_{\rho\rho}} = \int_0^{\rho_H} d\rho\ \frac{R(t)}{\sqrt{1-k\rho^2}} \\ &= R(t)\int_0^t \frac{dt'}{R(t')}. \end{aligned} \tag{4.2.4}$$

The distance to the horizon is determined by the behavior of $R(t)$ close to the big bang $t = 0$. Let us assume that $R(t) \propto t^n$. Then, for $n < 1$,

$$d_H(t) \propto t^n \int_0^t \frac{dt'}{t'^n} = [\frac{t^n}{(1-n)t'^{n-1}}]_0^t = \frac{t}{1-n}, \tag{4.2.5}$$

i.e. the distance is finite. For $n \geq 1$, on the other hand, the distance to the horizon is infinite. In a radiation dominated Universe, as it existed when the cosmic background radiation was generated, one has $n = 1/2$. Now the Universe is dominated by non-relativistic matter and we have $n = 2/3$. In both cases, the distance to the horizon is indeed finite.

Now let us consider the motion of a massive particle along a geodesic

$$\frac{du^\rho}{ds} + \Gamma^\rho_{\mu\nu}u^\mu u^\nu = 0, \tag{4.2.6}$$

and let us concentrate on the zero-component of this equation

$$\frac{du^0}{ds} + \Gamma^0_{\mu\nu}u^\mu u^\nu = 0, \tag{4.2.7}$$

In the FLRW metric one has

$$\Gamma^0_{ij} = -\frac{\dot{R}}{R}g_{ij}, \tag{4.2.8}$$

and we obtain

$$\frac{du^0}{ds} - \frac{\dot{R}}{R}g_{ij}u^i u^j = \frac{du^0}{ds} + \frac{\dot{R}}{R}|\vec{u}|^2 = 0. \tag{4.2.9}$$

The four-velocity is a tangent unit-vector

$$\begin{aligned} &(u^0)^2 - |\vec{u}|^2 = 1 \ \Rightarrow\ u^0 du^0 - |\vec{u}|d|\vec{u}| = 0 \ \Rightarrow \\ &\frac{1}{u^0}\frac{d|\vec{u}|}{ds} + \frac{\dot{R}}{R}|\vec{u}| = \frac{ds}{dx^0}\frac{d|\vec{u}|}{ds} + \frac{\dot{R}}{R}|\vec{u}| = |\dot{\vec{u}}| + \frac{\dot{R}}{R}|\vec{u}| = 0 \ \Rightarrow \\ &\frac{|\dot{\vec{u}}|}{|\vec{u}|} = -\frac{\dot{R}}{R}, \end{aligned} \tag{4.2.10}$$

such that

$$|\vec{u}| \propto R^{-1}. \tag{4.2.11}$$

The four-momentum is given by

$$p^\mu = m u^\mu, \tag{4.2.12}$$

which implies that the three-momentum $|\vec{p}| \propto R^{-1}$ is red-shifted. The ordinary three-velocity is

$$v^i = \frac{dx^i}{dt} = \frac{dx^i}{ds}\frac{ds}{dx^0} = \frac{u^i}{u^0} \ \Rightarrow \ |\vec{v}|^2 = \frac{|\vec{u}|^2}{(u^0)^2} = \frac{|\vec{u}|^2}{1+|\vec{u}|^2} \ \Rightarrow \ |\vec{u}|^2 = \frac{|\vec{v}|^2}{1-|\vec{v}|^2}. \tag{4.2.13}$$

Hence, for small $|\vec{v}|$ we have

$$|\vec{v}| \propto R^{-1}, \tag{4.2.14}$$

i.e. in an expanding Universe a free falling observer will ultimately come to rest in the cosmic rest frame of matter.

Let us also investigate the red-shift of light signals. We consider a signal emitted at time t_1 and coordinate ρ_1, and observed at time t_0 at coordinate $\rho = 0$, such that

$$ds^2 = 0 \ \Rightarrow \ \frac{dt}{R(t)} = \pm \frac{d\rho}{\sqrt{1-k\rho^2}} \ \Rightarrow \ \int_{t_1}^{t_0} \frac{dt}{R(t)} = \int_0^{\rho_1} \frac{d\rho}{\sqrt{1-k\rho^2}}. \tag{4.2.15}$$

Let us assume that a wave-maximum was emitted at t_1, and that the next wave-maximum was emitted at $t_1 + \delta t_1$. That maximum will then be observed at time $t_0 + \delta t_0$, such that

$$\int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{R(t)} = \int_0^{\rho_1} \frac{d\rho}{\sqrt{1-k\rho^2}}. \tag{4.2.16}$$

This immediately implies

$$\int_{t_0}^{t_0+\delta t_0} \frac{dt}{R(t)} = \int_{t_1}^{t_1+\delta t_1} \frac{dt}{R(t)} \ \Rightarrow \ \frac{\delta t_0}{R(t_0)} = \frac{\delta t_1}{R(t_1)}. \tag{4.2.17}$$

The wave-length λ is inversely proportional to the frequency, and hence proportional to δt, such that

$$\frac{\lambda(t_0)}{R(t_0)} = \frac{\lambda(t_1)}{R(t_1)} \ \Rightarrow \ \lambda(t) \propto R(t). \tag{4.2.18}$$

The wave-lengths of light in the Universe are stretched together with space itself. In an expanding Universe one hence finds a red-shift z with

$$1 + z = \frac{\lambda(t_0)}{\lambda(t_1)} = \frac{R(t_0)}{R(t_1)}. \tag{4.2.19}$$

Let us introduce the Hubble parameter

$$H(t) = \frac{\dot{R}(t)}{R(t)}, \tag{4.2.20}$$

which measures the strength of the expansion, and let us Taylor expand about the present epoch

$$R(t) = R(t_0) + (t - t_0)\dot{R}(t_0) \ \Rightarrow \ \frac{R(t)}{R(t_0)} = 1 + (t - t_0)H(t_0), \tag{4.2.21}$$

and hence

$$1 + z = \frac{R(t_0)}{R(t_1)} = 1 - (t_1 - t_0)H(t_0) \ \Rightarrow \ z = (t_0 - t_1)H(t_0). \tag{4.2.22}$$

What is the distance to the emitting galaxy? We find

$$\begin{aligned} d &= \int_0^1 d\lambda \, (-g_{ij}\dot{x}^i\dot{x}^j)^{1/2} = \int_0^{\rho_1} d\rho \, \sqrt{-g_{\rho\rho}} = R(t_0) \int_0^{\rho_1} \frac{d\rho}{\sqrt{1 - k\rho^2}} \\ &= R(t_0) \int_{t_1}^{t_0} \frac{dt'}{R(t')} = t_0 - t_1, \end{aligned} \tag{4.2.23}$$

such that

$$z = dH(t_0). \tag{4.2.24}$$

This is Hubble's law: the observed red-shift of light emitted from a distant galaxy is proportional to the distance of the galaxy. To derive Hubble's law we have made a power series expansion about the present epoch. Consequently, there will be corrections to Hubble's law at early times. It is interesting that the leading order calculation is independent of the details of the dynamics of the Universe (radiation- or matter-domination, curved or flat, open or closed).

4.3 Solutions of the Field Equations

The Einstein field equation takes the form

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{4.3.1}$$

In the FLRW metric, the non-vanishing components of the Einstein tensor are

$$G_{ii} = (2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2})g_{ii}, \ G_{00} = 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}). \tag{4.3.2}$$

Consequently, the energy-momentum tensor of matter must be diagonal as well. Let us consider the energy-momentum tensor of an ideal gas

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - p g_{\mu\nu}. \tag{4.3.3}$$

Here ρ is the energy density and p is the pressure of the gas, and u_μ is the four-velocity field. The tensor $T_{\mu\nu}$ is diagonal only if u_μ has a non-vanishing zeroth component only, i.e. if $u_0 = 1$. This means that, in the coordinate system of the FLRW metric, the gas of matter is at rest, and simply follows the expansion of the Universe (cosmic rest frame). Thus, we have

$$T_{ii} = -p g_{ii},\ T_{00} = \rho. \tag{4.3.4}$$

Density and pressure are temperature-dependent functions that are connected via an equation of state. From the space and time components we obtain two equations, the Friedmann equation

$$3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) = 8\pi G\rho, \tag{4.3.5}$$

as well as

$$2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = -8\pi G p. \tag{4.3.6}$$

Hence, one obtains

$$\begin{aligned}
&\frac{d}{dt}(8\pi G\rho R^3) + 8\pi G p\frac{d}{dt}R^3 = \\
&\frac{d}{dt}(3R\dot{R}^2 + 3kR) - (2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2})3R^2\dot{R} = \\
&3\dot{R}^3 + 6R\dot{R}\ddot{R} + 3k\dot{R} - 6\ddot{R}R\dot{R} - 3\dot{R}^3 - 3k\dot{R} = 0 \ \Rightarrow \\
&\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}R^3.
\end{aligned} \tag{4.3.7}$$

This equation can be identified as the first law of thermodynamics (energy conservation)

$$dE = -p dV, \tag{4.3.8}$$

where $E = \rho R^3$ is the rest energy of a given volume R^3. Let us form a linear combination of the two original equations

$$\begin{aligned}
&6\frac{\ddot{R}}{R} + 3\frac{\dot{R}^2}{R^2} + 3\frac{k}{R^2} - 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) = -8\pi G(3p + \rho) \ \Rightarrow \\
&\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(3p + \rho).
\end{aligned} \tag{4.3.9}$$

For the matter that we find on earth, one has $3p + \rho > 0$ and hence $\ddot{R} < 0$. Observations of the rotation curves of galaxies indicate that also so-called dark matter must exist in the Universe. One speculates that the dark matter may consist of heavy non-relativistic weakly interacting massive elementary particles (so-called WIMPs). Ordinary matter together with non-relativistic (so-called cold) dark matter would lead to a decelerated expansion. If $3p+\rho > 0$ was realized also in the past, there necessarily must have been a big bang, not earlier than the Hubble time $H(t_0)^{-1}$. Observations of very distant supernovae have revealed that the expansion of the Universe is actually accelerating. This implies that $3p+\rho < 0$. No ordinary form of matter obeys such an equation of state, which would require a negative pressure. However, vacuum energy (e.g. Einstein's cosmological constant or quintessence — a dynamical form of so-called dark energy) indeed has a negative pressure $p = -\rho$ such that $3p + \rho = -2\rho < 0$.

We can read the Friedmann equation as an equation for the Hubble parameter

$$H^2 + \frac{k}{R^2} = \frac{8\pi G}{3}\rho \ \Rightarrow \ \frac{k}{H^2 R^2} = \frac{8\pi G\rho}{3H^2} - 1 = \Omega - 1. \tag{4.3.10}$$

We have introduced the parameter

$$\Omega = \frac{\rho}{\rho_c}, \ \rho_c = \frac{3H^2}{8\pi G}. \tag{4.3.11}$$

Of course, $H^2 R^2 \geq 0$, such that we obtain a relation between the critical density of matter ρ_c (that depends on time via H) and the curvature k of space. For $k = 1$ space has positive curvature (S^3) and $\rho > \rho_c$. For $k = -1$ (H^3), on the other hand, $\rho < \rho_c$. Finally, if the density is critical ($\rho = \rho_c$) — i.e. if $\Omega = 1$ — space is flat ($\mathbb{R}^3$).

First, let us consider a matter dominated Universe — a system of non-relativistic matter clustered in galaxies, which are far away from each other. Then $p/\rho \leq 10^{-6}$, such that we can neglect the pressure and put $p = 0$. Using the first law of thermodynamics one finds

$$\rho R^3 = \frac{3}{4\pi} M, \tag{4.3.12}$$

where M is the conserved total mass contained in a sphere of radius R. Inserting this in the Friedmann equation one obtains

$$3\left(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}\right) = 6\frac{GM}{R^3} \ \Rightarrow \ \frac{\dot{R}^2}{2} - \frac{GM}{R} = -\frac{k}{2}. \tag{4.3.13}$$

The last equation can be interpreted as the equation of motion of a "point particle" (of unit mass) at position $R(t)$ in an attractive external potential $-GM/R$ with total "energy" $-k/2$. We then obtain

$$\dot{R} = \sqrt{\frac{2MG}{R} - k} \ \Rightarrow \ \int_0^t dt' = \int_0^{R(t)} \frac{dR}{\sqrt{\frac{2MG}{R} - k}}. \tag{4.3.14}$$

In flat space ($k = 0$) we find

$$t = \int_0^{R(t)} dR \sqrt{\frac{R}{2MG}} = \frac{2}{3}\sqrt{\frac{R(t)^3}{2MG}} \ \Rightarrow \ R(t) = (\frac{9}{2}MGt^2)^{1/3} \propto t^{2/3}. \tag{4.3.15}$$

The same behavior follows for small R, even in the cases $k = \pm 1$.

Let us now consider a negatively curved space ($k = -1$) for large values of R

$$t = \int_0^{R(t)} dR \ \Rightarrow \ R(t) = t. \tag{4.3.16}$$

Such a Universe expands forever, faster than a flat space. For a positively curved space, on the other hand, one has

$$\dot{R}^2 = \frac{2MG}{R} - 1 = 0 \ \Rightarrow \ R_{max} = 2MG, \tag{4.3.17}$$

i.e. the scale parameter reaches a maximal value R_{max}, and then the Universe contracts ending in a big crunch.

At very early times the Universe was filled with a gas of hot relativistic particles — it was radiation dominated. Then the equation of state is

$$p = \frac{1}{3}\rho, \tag{4.3.18}$$

and we obtain

$$\begin{aligned} &\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}R^3 = -\frac{1}{3}\rho 3R^2\dot{R} = -\rho R^2\dot{R} \ \Rightarrow \\ &\dot{\rho}R^3 + 3\rho R^2\dot{R} + \rho R^2\dot{R} = \dot{\rho}R^3 + 4\rho R^2\dot{R} = 0, \end{aligned} \tag{4.3.19}$$

such that

$$\frac{d}{dt}(\rho R^4) = \dot{\rho}R^4 + 4\rho R^3\dot{R} = 0. \tag{4.3.20}$$

Hence, we can write

$$\rho R^4 = N, \tag{4.3.21}$$

with N being a constant which is proportional to the number of particles in the relativistic gas. By again inserting this in the Friedmann equation we now obtain

$$3\left(\frac{\dot{R}^2}{R^2}+\frac{k}{R^2}\right)=\frac{8\pi GN}{R^4} \;\Rightarrow\; \frac{\dot{R}^2}{2}-\frac{4\pi GN}{3R^2}=-\frac{k}{2}, \tag{4.3.22}$$

which can be interpreted in the "point particle" analogy as motion in the attractive potential $-4\pi GN/3R^2$ such that

$$\dot{R}=\sqrt{\frac{8\pi GN}{3R^2}-k}. \tag{4.3.23}$$

The Universe is radiation dominated at early times. Hence, we can assume that R is small, such that we can neglect k and obtain

$$t=\int_0^{R(t)} dR\; R\sqrt{\frac{3}{8\pi GN}}=\frac{1}{2}R(t)^2\frac{3}{8\pi GN} \;\Rightarrow\; R(t)\propto t^{1/2}. \tag{4.3.24}$$

Let us also consider an empty Universe, however, with a cosmological constant Λ. Then we have $T_{\mu\nu}=0$ and thus

$$G_{\mu\nu}=8\pi GT_{\mu\nu}+g_{\mu\nu}\Lambda=g_{\mu\nu}\Lambda. \tag{4.3.25}$$

In the ideal gas parametrization of the energy-momentum tensor this corresponds to

$$\rho=\frac{\Lambda}{8\pi G},\; p=-\rho=-\frac{\Lambda}{8\pi G}. \tag{4.3.26}$$

Remarkably, in contrast to a relativistic or non-relativistic gas, a positive cosmological constant (i.e. vacuum energy) exerts a negative pressure. As we will see, this leads to an accelerated expansion of the Universe that counteracts the deceleration due to matter. Again, by insertion into the Friedmann equation we now obtain

$$3\left(\frac{\dot{R}^2}{R^2}+\frac{k}{R^2}\right)=\Lambda \;\Rightarrow\; \frac{\dot{R}^2}{2}-\frac{\Lambda R^2}{6}=-\frac{k}{2}, \tag{4.3.27}$$

which implies

$$\dot{R}=\sqrt{\frac{\Lambda}{3}R^2-k} \;\Rightarrow\; t=\int_{R(0)}^{R(t)}\frac{dR}{\sqrt{\frac{\Lambda}{3}R^2-k}}. \tag{4.3.28}$$

First, let us consider a flat Universe ($k=0$), such that

$$t=\int_{R(0)}^{R(t)}\frac{dR}{\sqrt{\frac{\Lambda}{3}}R}=\sqrt{\frac{3}{\Lambda}}[\log R]_{R(0)}^{R(t)} \;\Rightarrow\; R(t)=R(0)\exp(\sqrt{\frac{\Lambda}{3}}t). \tag{4.3.29}$$

In this case, the Hubble parameter is

$$H(t) = \frac{\dot{R}(t)}{R(t)} = \sqrt{\frac{\Lambda}{3}}, \tag{4.3.30}$$

and thus constant in time. Exponential expansion due to vacuum energy is exactly what happens in the inflationary Universe.

How does the cosmological constant affect the expansion of the Universe today? For a matter dominated Universe with cosmological constant we write

$$\begin{aligned}
&2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = \Lambda, \ 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) = 8\pi G\rho + \Lambda \ \Rightarrow \\
&\frac{d}{dt}(8\pi G\rho R^3) = 3\frac{d}{dt}(\dot{R}^2 R + kR) - \Lambda\frac{d}{dt}R^3 \\
&= 3(2\ddot{R}\dot{R}R + \dot{R}^3 + k\dot{R}) - 3\Lambda R^2\dot{R} = 0 \ \Rightarrow \\
&\rho R^3 = \frac{3}{4\pi}M \ \Rightarrow \ 3\left(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}\right) = \frac{6GM}{R^3} + \Lambda \ \Rightarrow \\
&\frac{1}{2}\dot{R}^2 - \frac{\Lambda}{6}R^2 - \frac{GM}{R} = -\frac{k}{2}.
\end{aligned} \tag{4.3.31}$$

To get an overview over all possible solutions, we can again interpret the equation in the point particle analogy now with the potential

$$V(R) = -\frac{\Lambda}{6}R^2 - \frac{GM}{R}, \tag{4.3.32}$$

and with total energy $-k/2$. For $k = 0, -1$ the Universe expands forever. At the end the $(\Lambda/3)R^2$ term dominates, and the Universe expands exponentially. For $k = 1$ the expansion comes to an end ($\dot{R} = 0$) if

$$\frac{\Lambda}{3}R^2 + \frac{2GM}{R} = 1. \tag{4.3.33}$$

We obtain a static Universe, if at the same time $\ddot{R} = 0$, such that

$$\frac{\Lambda}{3} = \frac{MG}{R^3} \ \Rightarrow \ \frac{\Lambda}{3}R^2 + \frac{2\Lambda}{3}R^2 = 1 \ \Rightarrow \ R = \frac{1}{\sqrt{\Lambda}}. \tag{4.3.34}$$

The density in such an Einstein Universe is

$$\rho = \frac{3}{4\pi}\frac{M}{R^3} = \frac{1}{4\pi G}\Lambda. \tag{4.3.35}$$

The static Einstein Universe exists for a critical cosmological constant $\Lambda_c = 1/(3GM)^2$. For $\Lambda > \Lambda_c$ the Universe expands forever, if it does so at present.

For $\Lambda < \Lambda_c$ there are two cases. For $\dot{R} > 0$ and large R the Universe expands forever, and there was a minimal radius in the past. For $\dot{R} > 0$ and small R there is also a solution with a maximal radius and a subsequent contraction that ends in a big crunch.

4.4 Four Possible Eras in the Late Universe

It is interesting to ask what may happen to the Universe in the distant future, even if we won't be able to test these ideas in observations. Whatever we predict will be based on our incomplete knowledge about particle physics (and physics in general) today. Thus, many of the following considerations should be considered highly speculative and may have to be revised as our knowledge progresses.

Following Adams and Laughlin, we may divide the future somewhat arbitrarily into four possible eras. During the stelliferous era that started with galaxy formation, stars dominate the Universe. Stars burn out after a while, and star formation will come to an end at about 10^{14} years after the bang. By then, the stellar objects have turned into brown dwarfs, white dwarfs, neutron stars, and an occasional black hole. White dwarfs and neutron stars are supported by degeneracy pressure of electrons and neutrons, respectively. When these objects dominate the Universe, we enter the degenerate era, which may last until about 10^{40} years after the bang (depending on appropriate assumptions about the proton life-time). At that time, all protons may have decayed, and thus all stellar objects have disappeared. As a result, we enter the black hole era, which ends when the last biggest black hole in the Universe has evaporated at about 10^{100} years. After that, in the dark era, the Universe consists of photons, as well as some electrons and positrons, which ultimately annihilate into photons at about 10^{140} years after the bang. If nothing else happens by then, the Universe will be a very dull place, unpleasant to live in for even the most patient creatures.

Let us start discussing the stelliferous era, which we have just entered. The crucial observation is that, while massive stars dominate the Universe now, they will soon run out of fuel. Depending on their mass, they will either blow up in a supernova, leaving behind a neutron star, or they will become a red giant, and then end as a white dwarf. Less massive stars, on the other hand, burn the nuclear fuel at a much more moderate rate. These objects therefore live much longer, up to 10^{13} years, and finally also become white dwarfs (without going through a red giant phase). There are also brown dwarfs, objects which are not sufficiently massive to ignite a nuclear fire in their cores. They remain unchanged during the stelliferous era.

The fate of the earth is decided when our sun enters the red giant phase in about five billion years. Most likely the earth will be swallowed and grilled within a time of about 50 years. Still, there is a chance that the sun looses enough material before going into the red giant phase, thus shifting the planets to further away orbits. This may save the earth for a long time to come.

Star formation happens in the gaseous clouds within galaxies. However, the supply of material is finite, and after a while there will be nothing left to make stars. One can estimate that the last star will form at about 10^{14} years after the bang, and it will soon be burned out. This is when the stelliferous era comes to an end.

At the end of the stelliferous era we are left with a system of many brown and white dwarfs, and an occasional neutron star or black hole, and we now enter the degenerate era. The above stellar objects are still bound inside galaxies, but galaxies will now die, because over time the stellar objects evaporate at a time scale of 10^{20} years, or fall into the black hole at the center of the galaxy. Still, during this era, occasionally a luminous star may be formed in a collision of brown or white dwarfs. One can estimate the collision rate for such processes, and one finds that a typical galaxy will then contain about 50 luminous stars, compared to 10^{10} today. At about 10^{30} years after the bang, all galaxies will have turned into a system of black holes and single evaporated stars.

The degenerate era comes to an end when the evaporated stellar objects outside black holes disappear due to proton decay. Depending on one's favorite GUT theory, this may happen at different times, e.g. 10^{37} years after the bang. When a proton decays, it may turn into a pion and a positron. The positron annihilates with an electron and produces two photons, and the pion also decays into two photons. At the end we loose a proton and an electron, and we get four photons. This means that the stars still shine, however, with a luminosity of about 10^{-24} of the sun. The stellar temperature then is about 0.06 K. One such star can power no more than a couple of light bulbs. Before a white dwarf disappears completely, its electron degeneracy is lifted, because the mass gets too small. Similarly, a neutron star will turn into a white dwarf, when sufficiently many neutrons have decayed. Also planets — e.g. the earth if it survived the sun's red giant phase — will disappear via proton decay and shine by emitting photons. The luminosity, however, is ridiculously small, about 0.4 mW. Depending on the proton life-time, the degenerate era may come to an end at about 10^{40} years after the bang. Based on the standard model of particle physics alone (i.e. without using GUT theories) the proton would live much longer, about 10^{172} years.

At the end of the degenerate era we enter the black hole era, during which black holes coalesce and finally disappear by emitting Hawking radiation at about 10^{100} years after the big bang. We then enter the dark era, in which the universe consists of just photons, and a few electrons and positrons. Electrons and positrons form positronium states — typically of the horizon scale, and cascade down to the ground state, from which they soon annihilate and become photons. Then, at an estimated 10^{140} years after the big bang, not much can happen, because there is nothing left that would interact with anything else.

Of course, for the Universe to survive until 10^{140} years after the bang, we have assumed that the Universe is not closed. Otherwise, the fun might end in a big crunch as early as 10^{11} years after the bang. Even if the Universe is flat in the part that we can observe now, as it expands a large density fluctuation may enter the horizon later, and close it. One can use inflation to make the Universe homogeneous on scales even larger than the ones we care about today. In order to guarantee survival of the Universe for about 10^{100} years, one needs about 130 e-foldings, which may be possible with sufficiently elaborate inflationary model building.

Also, any amount of vacuum energy today will dominate the Universe in the future. This should lead to another inflationary epoch. Thus, in order to go through the previously discussed eras, we must assume an even tinier cosmological constant than what is actually observed. Otherwise, stellar objects loose causal contact before they disappear via proton decay. One would see stars disappear through the horizon, and it gets dark much before the dark era.

An even more drastic effect may be due to vacuum tunneling. What if we still live in the wrong vacuum? This is quite possible, since today the vacuum energy is not exactly zero. Then a bubble of true vacuum may nucleate any time anywhere in the Universe. Such a bubble will grow with the velocity of light, turning the whole Universe into the true vacuum. For us this would be desasterous, because nothing in the old Universe will survive this change. Also such an event is impossible to predict, and could basically happen tomorrow. I guess we should all feel much saver, when some clever string theorist finally proves that the true vacuum is $SU(3) \otimes SU(2) \otimes U(1)$ invariant.

Much of this last section is speculative, and — in any case — not testable in observations. Many of the above ideas may have to be revised, as our knowledge increases. Still, whatever we may learn in the future, it is likely that the Universe will remain an interesting and lively place for a very long time. We should do our best to make sure that the same remains true for the earth.

Chapter 5

The Inflationary Universe

The standard model of cosmology, i.e. a Friedmann Universe with Robertson-Walker metric which is first radiation and then matter or vacuum dominated, is very successful in describing phenomena in the early Universe. The synthesis of light nuclei at about 1 sec to 1 min after the big bang provides the earliest test of the standard model of cosmology. Also the explanation of the cosmic background radiation should be viewed as a big success. The standard model of cosmology together with the standard model of particle physics and its extensions (e.g. grand unified theories (GUT)) lead to interesting phenomena in the extremely early Universe, like phase transitions in the strong and electroweak sector, or the creation of magnetic monopoles at a GUT phase transition. Although the standard model of particle physics agrees very well with experiments, from a theoretical point of view it is not completely satisfactory due to its large number of parameters. Also the standard model of cosmology has some problems. On the one hand, we know that due to the classical treatment of gravity it cannot be valid around the Planck time. On the other hand, it contains several parameters in the form of initial conditions, which must be fine-tuned in order to explain the large age of our Universe. Such fine-tuning appears unnatural to most cosmologists. The Robertson-Walker ansatz for the metric was motivated by observation (e.g. by the observed isotropy of the cosmic background radiation). In the framework of the standard model, it remains an open question why our Universe is isotropic and homogeneous on such large scales. On the other hand, on smaller (but still very large) scales we observe a lot of structure, like galaxies, galaxy clusters, super-clusters as well as large voids. In the framework of the standard model it is unclear how these deviations from homogeneity have evolved from small initial fluctuations. Finally, GUTs — which can at least qualitatively account for the baryon asymmetry — unavoidably lead to monopole creation, but monopoles have not been observed. All these puzzles find an answer in Alan Guth's idea of the inflationary Universe.

One then assumes that e.g. at a first order GUT phase transition a scalar field remained in the symmetric phase, i.e. at an unstable minimum of the temperature-dependent effective potential, while the true minimum corresponds to the broken phase. The energy of the scalar field then acts as a cosmological constant, which leads to an exponential, inflationary expansion of the Universe. In this way any effect of earlier initial conditions is eliminated, and the Universe becomes homogeneous and isotropic on the largest scales. Eventually present magnetic monopoles get extremely diluted, and are effectively eliminated from the Universe. Furthermore, the quantum fluctuations of the scalar field give rise to initial inhomogeneities, which may have evolved into the structures observed today. The exponential expansion leads to a tremendous increase of the scale factor, and hence to a flat space. This implies $\Omega = 1$, and hence requires a sufficient amount of vacuum energy.

When $\Omega = 1$, this also explains the large age of our Universe (compared to the Planck time) and hence an important condition for our own existence. Hence, the inflationary Universe allows us to avoid the anthropic principle, which explains the large age of the Universe by the mere fact that we exist: we could not have developed in a short-lived Universe. Alan Guth's original idea has been modified by Linde who has introduced the scenario of chaotic inflation. One then assumes that initially the scalar field assumes some chaotic random values, and hence only certain regions of space undergo inflation. In this way different parts of the Universe with possibly different vacuum structure evolve. It should then be viewed as a historical fact, that in our local vacuum bubble the symmetry was broken to the $SU(3) \otimes SU(2) \otimes U(1)$ symmetry of the standard model of particle physics and not to something else. Linde still uses the anthropic principle to explain why this is so: in another vacuum we could not have evolved.

Since at present the Universe does no longer expand exponentially, inflation (if it ever took place) must have come to an end. This must have been associated with the scalar field "rolling" down to the stable minimum of broken symmetry. In this process an enormous amount of latent heat is released, which manifests itself in an enormous entropy generation. At present that entropy resides in the cosmic background radiation. The idea of inflation thus naturally leads to a Universe similar to the one we live in: old, flat and homogeneous and isotropic. Still, inflation should not yet be viewed as a completely established fact. It is a very attractive idea that still needs to be formulated in a realistic particle physics framework, and it needs to be tested further in observations. A strong indication for an inflationary epoch is the verification of $\Omega = 1$. Still, inflation does not solve all problems of cosmology. In particular, the question of the cosmological constant remains unsolved: why is the vacuum energy so small? To answer this question one probably needs a quantum theory of gravity, one of the great challenges in theoretical physics.

5.1 Deficiencies of Standard Cosmology

Although the standard model of cosmology is very successful in many respects, it leaves some important questions unanswered. First, there is the question of the age of the Universe and, related to that, its flatness, which is also connected with the enormous entropy in the cosmic background radiation. To understand the age problem, let us consider the Friedmann equation

$$\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = \frac{8\pi}{3}G\rho \;\Rightarrow\; \frac{k}{H^2R^2} = \frac{8\pi G\rho}{3H^2} - 1 = \frac{\rho}{\rho_c} - 1 = \Omega - 1, \tag{5.1.1}$$

where $H = \dot{R}/R$ is the Hubble parameter, and

$$\rho_c = \frac{3H^2}{8\pi G} \tag{5.1.2}$$

is the critical density. It is an observational fact, that our Universe has an energy density close to the critical one, i.e.

$$\Omega \approx 1. \tag{5.1.3}$$

What does this mean for earlier times? As long as the Universe was matter dominated, one has

$$R(t) \propto t^{2/3} \;\Rightarrow\; H(t) = \frac{\dot{R}(t)}{R(t)} = \frac{2}{3t}, \tag{5.1.4}$$

and thus

$$\Omega(t) - 1 = \frac{k}{H(t)^2R(t)^2} \propto \frac{t^2}{t^{4/3}} = t^{2/3}. \tag{5.1.5}$$

The Universe started to be matter dominated at an age of 10^5 years, and today it is about 10^{10} years old. Hence, when the cosmic background radiation decoupled, we had

$$\Omega(t) - 1 = (\frac{10^5}{10^{10}})^{2/3} \approx 10^{-3}, \tag{5.1.6}$$

i.e. the density was even closer to the critical one than it is today. At earlier times the Universe was radiation dominated with

$$R(t) \propto t^{1/2} \;\Rightarrow\; H(t) = \frac{\dot{R}(t)}{R(t)} = \frac{1}{2t}, \tag{5.1.7}$$

and hence

$$\Omega(t) - 1 = \frac{k}{H(t)^2 R(t)^2} \propto \frac{t^2}{t} = t. \tag{5.1.8}$$

At the time when nucleosynthesis started, i.e. about 1 sec after the bang, one finds

$$\Omega(t) - 1 = 10^{-3}\frac{10^{-7}}{10^5} = 10^{-15}, \tag{5.1.9}$$

which again means that the density was extremely close to critical. Going further back in time, e.g. to a GUT phase transition at about 10^{-34} sec after the bang, one obtains

$$\Omega(t) - 1 = 10^{-15}10^{-34} = 10^{-49}, \tag{5.1.10}$$

and if one goes back to the Planck time, 10^{-44} sec, one even finds

$$\Omega(t) - 1 = 10^{-15}10^{-44} = 10^{-59}. \tag{5.1.11}$$

At first sight, today a value of Ω close to one seems not unnatural, but who has fine-tuned the energy density to sixty decimal places when the Universe was created at the Planck time?

This problem can also be expressed in terms of the entropy. In the standard cosmology the total entropy S in a comoving volume of size $R(t)^3$ is conserved. In the radiation dominated epoch the energy density is

$$\rho = \frac{\pi^2}{30} g_* T^4, \tag{5.1.12}$$

and the entropy density is

$$s = \frac{2\pi^2}{45} g_* T^3 = \frac{S}{R^3}. \tag{5.1.13}$$

At the Planck time the temperature is equal to the Planck mass, i.e. $T = m_P \approx 10^{19}$ GeV, such that then

$$\begin{aligned} \frac{\Omega - 1}{\Omega} &= \frac{k}{H^2 R^2}\frac{3H^2}{8\pi G\rho} = \frac{3k}{8\pi G\rho R^2} = \frac{3k}{8\pi G\rho}(\frac{s}{S})^{2/3} \\ &= \frac{3k}{8\pi}\frac{30}{\pi^2 g_*}(\frac{2\pi^2}{45}g_*)^{2/3}\frac{1}{G m_P^2} S^{-2/3} \approx 0.1 S^{-2/3}. \end{aligned} \tag{5.1.14}$$

Using $\Omega - 1 \approx 10^{-59}$ one obtains the enormous entropy $S \approx 10^{87}$ per comoving volume. Today this entropy is contained in the cosmic background radiation.

Not only the enormous entropy of the cosmic background radiation cannot be understood in the framework of the standard cosmology. Also its isotropy presents a puzzle. The observed isotropy of the cosmic background radiation motivated the Robertson-Walker ansatz for the metric, but one does not understand where the isotropy comes from. In a radiation or matter dominated Robertson-Walker space-time the distance to the horizon

$$d_H(t) = R(t) \int_0^t \frac{dt'}{R(t')} \tag{5.1.15}$$

is finite. In a matter dominated Universe, e.g. one has $d_H(t) \propto 3t$. This is the maximal distance between two events that are causally connected. When matter and radiation decoupled at about $t_d = 10^5$ years after the bang, physical processes could establish thermal equilibrium over scales of about

$$d_H(t_d) \propto 3t_d. \tag{5.1.16}$$

Today, photons of the cosmic background radiation reach us from different directions. Their regions of generation at time t_d cannot have been in causal contact, and still the photons have almost exactly the same temperature.

Let us consider a photon that was emitted at time t_d at the point ρ, and that reaches us today at time $t_0 = 10^{10}$ years at $\rho = 0$. We then have

$$ds^2 = dt^2 - R(t)^2 \frac{d\rho^2}{1 - k\rho^2} = 0 \;\Rightarrow\; \frac{dt}{R(t)} = \pm \frac{d\rho}{\sqrt{1 - k\rho^2}}, \tag{5.1.17}$$

and thus

$$\int_{t_d}^{t_0} \frac{dt'}{R(t')} = \int_0^{\rho} \frac{d\rho'}{\sqrt{1 - k\rho^2}}. \tag{5.1.18}$$

A second photon reaches us today from exactly the opposite direction. The coordinate difference between the two points of creation of the photons is therefore 2ρ. At the time t_d of their creation, this corresponded to a physical distance

$$\begin{aligned} d(t_d) &= 2\int_0^{\rho} d\rho \sqrt{-g_{\rho\rho}} = 2\int_0^{\rho} d\rho \frac{R(t_d)}{\sqrt{1 - k\rho^2}} = 2R(t_d) \int_{t_d}^{t_0} \frac{dt'}{R(t')} \\ &= 2d_H(t_d) \int_{t_d}^{t_0} \frac{dt'}{R(t')} \Big/ \int_0^{t_d} \frac{dt'}{R(t')}. \end{aligned} \tag{5.1.19}$$

Since the decoupling of the photons the Universe was matter dominated, i.e. $R(t) \propto t^{2/3}$. Thus we obtain

$$\begin{aligned}\frac{d(t_d)}{d_H(t_d)} &= 2(3t_0^{1/3} - 3t_d^{1/3})/3t_d^{1/3} \\ &= 2(\frac{t_0}{t_d})^{1/3} - 2 \approx 2(\frac{10^{10}}{10^5})^{1/3} \approx 40. \qquad (5.1.20)\end{aligned}$$

This implies that the points at which the two photons were created were separated by 40 causality-lengths, and still the photons have the same temperature. Who has thermalized the two causally disconnected regions at extremely similar temperatures?

A GUT phase transition would unavoidably lead to the creation of magnetic monopoles via the so-called Kibble mechanism. The contribution of monopoles to the energy density leads to $\Omega \approx 10^{11}$ in complete disagreement with observation. How can we get rid of the unwanted monopoles, without discarding GUTs altogether? After all, GUT are quite useful, because they may explain the baryon asymmetry of the Universe?

5.2 The Idea of Inflation

The idea of the inflationary Universe is to use vacuum energy (a non-zero cosmological constant) to blow up the size of the Universe exponentially. When vacuum energy dominates the Friedmann equation takes the form

$$\frac{\dot{R}(t)^2}{R(t)^2} = \frac{\Lambda}{3} \Rightarrow R(t) = R(t_0)\exp(\sqrt{\frac{\Lambda}{3}}(t - t_0)). \qquad (5.2.1)$$

When the inflationary period lasts for a time Δt the scale factor of the Universe increases by

$$Z = \frac{R(t_0 + \Delta t)}{R(t_0)} = \exp(\sqrt{\frac{\Lambda}{3}}\Delta t). \qquad (5.2.2)$$

The inflationary period ends when the vacuum energy goes to zero, because some scalar field (e.g. of a GUT) rolls down to the stable minimum of a low energy broken phase. In that moment latent heat is released, and the Universe (now in the low energy broken phase) is reheated up to T_c. This process is accompanied by entropy and particle creation. The entropy increase is given by

$$\frac{S(t_0 + \Delta t)}{S(t_0)} = Z^3 = \exp(\sqrt{3\Lambda}\Delta t). \qquad (5.2.3)$$

Assuming that the entropy before inflation was not unnaturally large (e.g. $S(t_0) \approx 1$), in order to explain $S \approx 10^{87}$ we need an inflation factor

$$Z = 10^{29} \;\Rightarrow\; \sqrt{\frac{\Lambda}{3}}\Delta t = 67. \tag{5.2.4}$$

An exponential expansion can also solve the horizon problem, because we then have

$$\begin{aligned} R(t_0+\Delta t)\int_{t_0}^{t_0+\Delta t}\frac{dt'}{R(t')} &= \frac{R(t_0+\Delta t)}{R(t_0)}\frac{3}{\Lambda}[1-\exp(-\sqrt{\frac{\Lambda}{3}}\Delta t)] \\ &\approx Z\sqrt{\frac{3}{\Lambda}} = \frac{Z}{\ln Z}\Delta t. \end{aligned} \tag{5.2.5}$$

For comparison, a radiation dominated Universe (with $R(t) \propto t^{1/2}$) has

$$R(\Delta t)\int_0^{\Delta t}\frac{dt'}{R(t')} = 2\Delta t. \tag{5.2.6}$$

Thus, the horizon problem is improved by a factor

$$\frac{Z}{2\ln Z} = 10^{27}, \tag{5.2.7}$$

although a factor 40 would be sufficient to solve it.

Also the monopole problem is solved by inflation, because monopoles created before or during inflation are diluted by a factor Z^3. In this way the monopole contribution to the critical density is reduced to

$$\Omega_M = 10^{11} Z^{-3} = 10^{-76}. \tag{5.2.8}$$

Hence, both the monopole and the horizon problem automatically disappear, once the flatness (or entropy) problem is solved.

5.3 The Dynamics of Inflation

Let us consider a scalar field ϕ (e.g. in a GUT) that obtains a vacuum expectation value. The potential of the scalar field may be given by

$$V(\phi) = \frac{M^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4. \tag{5.3.1}$$

Let us assume that at time t_0 the system is in the symmetric phase with $\phi = 0$. Afterwards it slowly rolls down to the broken symmetry minimum with $|\phi| = v$. The Lagrange density of the scalar field is given by

$$\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2}\partial^\mu \phi \partial_\mu \phi - V(\phi), \tag{5.3.2}$$

and the corresponding energy momentum tensor is

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \mathcal{L} g_{\mu\nu}. \tag{5.3.3}$$

Assuming that ϕ is spatially constant, but time-dependent, for a Robertson-Walker metric one obtains

$$\begin{aligned} T_{00} &= \frac{1}{2}\dot{\phi}^2 + V(\phi) = \rho, \\ T_{ii} &= -(\frac{1}{2}\dot{\phi}^2 - V(\phi))g_{ii} = -p g_{ii} \ \Rightarrow \ p = \frac{1}{2}\dot{\phi}^2 - V(\phi). \end{aligned} \tag{5.3.4}$$

As usual, we have identified density and pressure. The Einstein field equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{5.3.5}$$

then takes the following form

$$\begin{aligned} 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) &= 8\pi G(\frac{1}{2}\dot{\phi}^2 + V(\phi)), \\ 2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} &= -8\pi G(\frac{1}{2}\dot{\phi}^2 - V(\phi)). \end{aligned} \tag{5.3.6}$$

We now identify the effective cosmological constant as

$$\Lambda = 8\pi G V(\phi). \tag{5.3.7}$$

The typical scale for the potential $V(\phi)$ is the GUT scale 10^{14} GeV. Hence, we can assume $V(\phi) = (10^{14}\text{GeV})^4$ and thus

$$\sqrt{\frac{\Lambda}{3}} = \sqrt{\frac{8\pi}{3}\frac{V(0)}{m_P^2}} = \sqrt{\frac{8\pi}{3}\frac{10^{14}}{10^{19}}} 10^{14}\text{GeV} \approx 10^9\text{GeV}. \tag{5.3.8}$$

In order to reach a sufficient amount of inflation, we need it to last for at least

$$\Delta t = 67\sqrt{\frac{3}{\Lambda}} \approx 10^{-7}\text{GeV}^{-1} \approx 10^{-32}\text{sec}. \tag{5.3.9}$$

This is about 100-1000 times the age of the Universe at the time of inflation.

Is it possible to delay the transition to the broken phase with $\phi = v$ for such a long time? To answer that question, we consider the equation of motion for the scalar field

$$\ddot{\phi} + 3\frac{\dot{R}}{R}\dot{\phi} + \frac{\partial V}{\partial \phi} = 0. \tag{5.3.10}$$

The term $3(\dot{R}/R)\dot{\phi}$, which deviates from the ordinary Klein-Gordon equation, results from the red-shift of the momentum of the scalar field. For a long-lasting inflationary period we need a small $\dot{\phi}$. Ignoring $\ddot{\phi}$ results in

$$\dot{\phi} = -\frac{R}{3\dot{R}}\frac{\partial V}{\partial \phi}. \tag{5.3.11}$$

In the Friedmann equation (with $k = 0$) we can furthermore neglect $\dot{\phi}^2$ compared to $V(\phi)$ such that

$$\frac{\dot{R}}{R} = \sqrt{\frac{8\pi G}{3}V(\phi)} \;\Rightarrow\; \dot{\phi} = -\frac{1}{\sqrt{24\pi G V(\phi)}}\frac{\partial V}{\partial \phi}. \tag{5.3.12}$$

The duration of inflation is now given by

$$\begin{aligned}\Delta t &= \int_{t_0}^{t_0+\Delta t} dt' = \int_{\phi(t_0)}^{\phi(t+\Delta t)} d\phi\ \frac{1}{\dot{\phi}} = \int_0^v d\phi\ \sqrt{24\pi G V(\phi)}(\frac{\partial V}{\partial \phi})^{-1} \\ &= \sqrt{24\pi}t_P \int_0^v d\phi\ \sqrt{V(\phi)}(\frac{\partial V}{\partial \phi})^{-1}.\end{aligned} \tag{5.3.13}$$

We parameterize the potential as

$$V(\phi) = \lambda(\phi^2 - v^2)^2, \tag{5.3.14}$$

and we obtain

$$\int_0^v d\phi\ \sqrt{V(\phi)}(\frac{\partial V}{\partial \phi})^{-1} = \int_0^v d\phi\ \sqrt{\lambda}\frac{\phi^2 - v^2}{2\lambda(\phi^2 - v^2)2\phi} = \frac{1}{4\sqrt{\lambda}}\int_0^v \frac{d\phi}{\phi} = \infty. \tag{5.3.15}$$

We need to distort ϕ slightly from its unstable value $\phi = 0$. For example, we can put it to $\phi = \phi_0$ initially, and obtain

$$\Delta t = \sqrt{\frac{3\pi}{2\lambda}}t_P \ln(\frac{v}{\phi_0}). \tag{5.3.16}$$

The value of v is determined, for example, by the GUT scale, and $\lambda \approx 1$. In order to reach

$$\Delta t \approx 10^{-32}\text{sec} \approx 10^{12} t_P, \tag{5.3.17}$$

one needs an extremely small initial scalar field value

$$\phi_0 \approx v \exp(-10^{12}). \tag{5.3.18}$$

In order to last sufficiently long, the initial conditions for inflation must be adjusted extremely carefully. Hence, unfortunately, we have traded the fine-tuning problem for Ω for another one — the fine-tuning of ϕ_0. For other forms of the scalar potential the new fine-tuning problem may be less severe, but this particular scenario of inflation does not look very natural.

At the end of the inflationary period the scalar field approaches its vacuum value by oscillating around it. The oscillations are damped by the couplings to other fields. This leads to a release of the latent heat stored in the scalar field. The generated entropy reheats the Universe to the critical temperature T_c of the GUT transition.

New data to be obtained by the Planck satellite are expected to further test the intriguing idea of cosmic inflation. The chances seem reasonably good that inflation will become part of a future standard model of cosmology.